On Spatial Adaptive Nonparametric Estimation of Functions Satisfying Differential Inequalities

A. Goldenshluger*, A.Nemirovski*

September 1996

*Faculty of Industrial Engineering & Managament at the Technion – Israel Institute of Technology, Haifa, Israel; e–mail: goldensh@techunix.technion.ac.il, nemirovs@techunix.technion.ac.il

Abstract

The paper is devoted to developing spatial adaptive estimates of the signals satisfying linear differential inequalities with unknown differential operator of a given order. The classes of signals under consideration cover a wide variety of classes usual in the non-parametric regression problem; moreover, they contain the signals whose parameters of smoothness are not uniformly bounded, even locally. We develop the estimate which is optimal in order over these classes and wide range of "discrete" global accuracy measures.

1 Intorduction

The goal of this paper is to develop a *spatial adaptive* estimate of nonparametric signals satisfying differential inequalities. Our basic problem is as follows:

given observations

$$y(x) = f(x) + \xi(x) \tag{1}$$

of a function $f : \mathbf{R} \to \mathbf{C}$ *along a regular grid* $X_h = \{th \mid t = 0, 1, ..., \lfloor 1/h \rfloor\}$, $\xi(x)$ *being independent Gaussian random vectors from* $\mathbf{C} = \mathbf{R}^2$ *with the zero mean and the unit covariance matrix, restore the restriction* f^h *of* f *onto the grid.*

Note that for us it is convenient to deal with complex–valued signals and, consequently, with complex–valued Gaussian noise. This clearly does not restrict generality: in the case when the signal is real–valued and the noise in observations is the usual discrete Gaussian white noise, one can think of the signal as of being complex–valued and add to the actual (real) observations artifical discrete Gaussian white noise as their imaginary part. It is also convenient to assume that the signals are defined on the entire real axis; let us, anyhow, stress that we observe the signals only on the grid $X_h \subset [0, 1]$ and are interested in restoring the signal only within the bounds of this segment.

The essence of the matter is, of course, the class of signals the estimate is oriented to. The usual classes here are those of *smooth* signals f of certain fixed smoothness (like the Sobolev and the Besov spaces); these classes form the subject of the majority of the papers on nonparametric regression (see [12, 5, 10, 3] and references therein). Note that the classes of smooth functions are associated with the standard differential operators $q_l(d/dx) = d^l/dx^l$; e.g., the Sobolev class

$$W_l^p(L) = \{f \mid \|f^{(l)}\|_p \le L\}$$

is defined by imposing an upper bound on the L_p-norm of the image of a signal f under the action of the operator q_l. A natural way to generalize the family of signals in question is to impose similar bounds on the image of f under action of another differential operator $q(d/dx) = d^l/dx^l + \sum_{i=0}^{l-1} q_i d^i/dx^i$, thus coming to the classes

$$W_{q(\cdot)}^p(L) = \{f \mid \|q(d/dx)f\|_p \le L\};$$

the class $W_{q(\cdot)}^p(L)$ is comprised of all solutions to the differential inequalities

$$|q(d/dx)f| \le g \tag{2}$$

with the right hand side g satisfying the restriction $\|g\|_p \le L$. The indicated extension is of rather restricted practical interest, since in practice we hardly could specify the underlying differential operator. What seems to be more attractive, is to consider the classes comprised of *all* solutions to *all* differential inequalities (2) associated with differential operators of a given order l and right hand sides of a given L_p-norm, i.e., the classes

$$Q_l^p(L) = \bigcup_{q(\cdot)} W_{q(\cdot)}^p(L),$$

the union being taken over all differential operators $q(\cdot)$ ($\equiv$ all polynomials $q(\cdot)$) of a given order l with the unit principal coefficient. Thus, our hypothesis on the signal in question now is that the signal is an output of an *unknown* linear filter of a given order l, the input to the filter being not too large, while in the traditional case the filter is known in advance and is simply d^l/dx^l. The problem of nonparametric regression estimation for the classes Q was studied (for the case of $p = \infty$) in [11].

The conceptual drawback of the estimate developed in [11] (same as of the majority of known estimates for the usual Sobolev classes) is that to use the estimate, one should know in advance the parameters p, l, L specifying the class in question. Recent progress in nonparametric estimation resulted in developing *spatial adaptive estimates* of the signals from the Sobolev/Besov spaces. These estimates do not require a priori knowledge of the smoothness parameters of the signal; moreover, the latter should not necessarily be smooth on the entire segment where it is observed. What is guaranteed by spatial adaptive estimates is, roughly speaking, that if the signal is smooth, with certain (unknown) smoothness parameters, on certain (unknown) segment, then the quality of its restoring on this segment will be as good as if we knew in advance the segment and the smoothness parameters (for precise formulations, see the papers of Donoho et al. [1, 2, 4], Lepskii et al. [9] and Juditsky [7], where the first estimates of this type were suggested and studied).

The goal of this paper is to develop spatial adaptive estimates of the signals satisfying differential inequalities with unknown differential operators of a given order. We analyze quality of the adaptive estimate under two different assumptions reflecting local and global behaviour of f. More exactly, assume that there exists a segment $D \subset [0,1]$ where the signal f in question can be decomposed into a sum

$$f = \sum_{j=1}^{k} f^j$$

of signals f^j such that

$$\sum_{j=1}^{k} \|q^j(d/dx)f^j\|_{p,D} \le L$$

for some differential operators q^j of order l. The indicated assumption describes local behaviour of the signals f we are interested in; the assumption (and the corresponding family of the signals f) is denoted by $\mathbf{A}[k,l;p,L;D]$. Now, assume that there exists real $d(f) \ge d > 0$ such that for any x from a segment $D' \subset intD$ assumption $\mathbf{A}[k,l;p,L_f(x);[x-d(f),x+d(f)]]$ holds with $\|L_f(\cdot)\|_{p,D'} \le L$. This assumption describes global behaviour of the signals f and is denoted by $\mathbf{F}_d[k,l;p,L;D]$. Let us stress that all quantities involved in the descriptions of $\mathbf{A}[k,l;p,L;D]$ and $\mathbf{F}_d[k,l;p,L;D]$ *are not assumed to be known in advance.* All what we know is an upper bound m on the product lk; this upper bound is the only "design parameter" of our estimate $\hat{f}$. And what is guaranteed is that the quality of restoring f at the grid X_h of the segment D' measured in the "discrete L_q-norm"

$$\mathcal{R}_{q,D'}(\hat{f},f) \equiv \mathcal{E}\left[\left(h \sum_{x \in X_h \cap D'} |\hat{f}(x) - f(x)|^q\right)^{1/q}\right]$$

is optimal in order in the minmax sense:

$$\sup_{f\in\mathbf{F}_d[k,l;p,L;D]} \mathcal{R}_{q,D'}(\hat{f},f) \le O(1)\inf_{\tilde{f}} \sup_{f\in\mathbf{F}_d[k,l;p,L;D]} \mathcal{R}_{q,D'}(\tilde{f},f);$$

inf in the right hand side being taken over the set of all possible estimates $\tilde{f}$ and $O(1)$ being constant factor independent of h and of the parameters L, D.

Let us make some comments on the formulation of the problem and the outlined result.

1) Note that classes $\mathbf{F}_d[\cdot]$ are wider than those $\mathbf{A}[\cdot]$; besides this, both cover a wide variety of classes usual in the nonparametric regression problem (for details, see Section 3).

2) Note that the developed estimate is optimal in order uniformly over all polynomials $q^j(\cdot)$, $j = 1,\dots,k$ of order l satisfying $kl \le m$. In particular, the class $\mathbf{A}[2,l;p,L;D]$ includes the signals which are obtained by modulation of a smooth function by a sine of arbitrary frequency, namely, all signals of the type

$$f(x) = g(x)\sin(\omega t+\phi), \quad \|g^{(l)}\|_{p,D} \le L.$$

(see Section 3). It seems surprising that the quality of our estimate does not depend on the frequency of the modulator; e.g., the risk $\mathcal{R}_{2,D'}$ of our estimate on the signals in question is bounded, for small h, *uniformly in* ω by the quantity

$$O(1)|D|^{1/2}L^{1/(2l+1)}(h\ln\frac{1}{h})^{l/(2l+1)},$$

with $O(1)$ depending on the design parameter m ($\ge 2l$) of the estimate.

It seems to be impossible to obtain the latter result for the “traditional” estimates, like the kernel or the wavelet-based ones; these estimates are oriented to smooth, at least locally, signals, while the parameters of smoothness of the signals in question are not uniformly bounded, even locally.

3) In what follows we are interested in restoring f *only on the grid* X_h, since as far as the uniform with respect to the polynomials $q^j(\cdot)$ risk is concerned, only this goal is achievable. Indeed, the function $f_*(x) = \sin 2\pi h^{-1}x$ clearly satisfies assumption $\mathbf{A}[1,2;p,L;D]$ for any $L>0, p\in[1,\infty], D\subseteq[0,1]$ and vanishes on X_h. Therefore, observations (1) do not allow to distinguish between $f_*(x)$ and $f\equiv 0$.

The rest of the paper is organized as follows. Section 2 is devoted to constituting the estimate. Section 3 establishes the upper bounds on the accuracy of our estimate. The concluding Section 4 contains the lower bounds and some comments.

2 The estimate

In this section we construct our adaptive estimate. The scheme combines the approach developed in [11] for estimating signals satisfying differential inequalities with unknown operator of known order and known bounds on the right hand side of the inequality, and the general adaptation technique of Lepskii [8].

2.1 Preliminaries

Space of sequences. Let $\mathcal{P}$ be the space of two-sided complex-valued sequences $\{\phi_t\}_{t\in\mathbf{Z}}$ with finitely many nonzero entries. In what follows we identify a sequence $\{\phi_t\}\in\mathcal{P}$ with the rational function

$$\phi(z)=\sum_t \phi_t z^t.$$

We equip $\mathcal{P}$ with the natural linear operations - addition and multiplication by scalars from $\mathbf{C}$ - and with multiplication

$$(\phi\psi)(z)=\phi(z)\psi(z)$$

(which corresponds to the convolution in the initial "sequence" representation of the elements of $\mathcal{P}$).

For $\phi(\cdot)\in\mathcal{P}$ we denote by $d(\phi)$ the minimum of those $\tau\geq 0$ for which $\phi_t=0$, $|t|>\tau$, so that

$$\phi(z)=\sum_{|t|\leq d(\phi)}\phi_t z^t.$$

We denote by $\mathcal{P}_N$, N being a nonzero integer, the subspace of $\mathcal{P}$ comprised of all ϕ with $d(\phi)\leq N$.

Norms on $\mathcal{P}$. For $\infty\geq N\geq 0$ and $p\in[1,\infty]$ let

$$\|\phi\|_{N,p}=\left(\sum_{t=-N}^{N}|\phi_t|^p\right)^{1/p}$$

(if $p=\infty$, then the right hand side, as usual, is $\max_{|t|\leq N}|\phi_t|$) be the standard p-(semi)norm on $\mathcal{P}$; restricted on $\mathcal{P}_N$, this is an actual norm. We shall omit explicit indicating N in the notation of norm in the case of $N=\infty$; thus, $\|\phi\|_p$ is the same as $\|\phi\|_{\infty,p}$.

Dicrete Fourier transformation. Let N be positive integer, and let Γ_N be the set of all roots

$$\mu_k=\exp\left\{\frac{2\pi k}{2N+1}\mathbf{i}\right\},\ k=0,1,...,2N,$$

of the unit of the degree $2N+1$.

Let $\mathbf{C}_N$ be the space of $\mathbf{C}$-valued functions of Γ_N, i.e., the space $\mathbf{C}^{2N+1}$ with the entries of the vectors indiced by the elements of Γ_N. We define the Fourier transformation $F_N:\mathcal{P}\to\mathbf{C}(\Gamma_N)$ by the usual formula

$$(F_N\phi)(\mu)=\frac{1}{\sqrt{2N+1}}\sum_{|t|\leq N}\phi_t\mu^t\quad[=\frac{1}{\sqrt{2N+1}}\phi(\mu)\text{ for }\phi\in\mathcal{P}_N],\ \ \mu\in\Gamma_N,\qquad(3)$$

which is equivalent to

$$\phi_t=\frac{1}{\sqrt{2N+1}}\sum_{\mu\in\Gamma_N}(F_N\phi)(\mu)\mu^{-t},\ \ |t|\leq N.$$

Being restricted on $\mathcal{P}_N$, F_N is, as it is well-known, an isometry in 2-norms:

$$\langle \phi, \psi \rangle_N \equiv \sum_{|t| \le N} \phi_t \overline{\psi}_t = \langle F_N\phi, F_N\psi\rangle \equiv \sum_{\mu \in \Gamma_N} (F_N\phi)(\mu)\overline{(F_N\psi)(\mu)}, \quad \phi, \psi \in \mathcal{P}, \tag{4}$$

where $\bar{a}$ denotes the conjugate of $a \in \mathbf{C}$.

The space $\mathbf{C}(\Gamma_N)$ also can be equipped with p-norms

$$\|\zeta(\cdot)\|_p = \left(\sum_{\mu \in \Gamma_N} |\zeta(\mu)|^p \right)^{1/p}$$

with the already indicated standard interpretation of the right hand side for the case of $p = \infty$. Via Fourier transformation, these norms can be translated to $\mathcal{P}$, and we set

$$\|\phi\|^*_{N,p} = \|F_N\phi\|_p;$$

these are seminorms on $\mathcal{P}$, and their restrictions on $\mathcal{P}_N$ are norms on the latter subspace.

2.2 The adaptive estimate

We are about to define an estimate of a sequence $\{f_t \in \mathbf{C}\}$, $t \in \mathbf{Z}$, at a given point $\tau \in I_h = \{0, 1, ..., \lfloor 1/h \rfloor\}$ via noisy observations

$$y = \{y_t = f_t + \xi_t\}_{t \in I_h},$$

where

$$\xi = \{\xi_t\}$$

is a sequence of independent $\mathcal{N}(0, I)$ random elements of $\mathbf{C}$, i.e., of vectors $\xi_t \in \mathbf{C} = \mathbf{R}^2$ with zero mean and the unit covariance matrix I. Thus, we assume, for the sake of convenience, that f_t and ξ_t are defined for all integer t, while y_t are given only for $t \in I_h$.

Our estimate will depend on two "design" parameters - the *threshold* $\kappa \ge 1$ and the *order* m, which is a positive integer (κ will be in the meantime specified as certain function of h and m, so that in fact we have a single design parameter - the order of the estimate).

The estimate will be defined for all τ which are not too close to the enpoints of the range I_h of the index t, namely, for τ satisfying

$$4m \le \tau \le 1/h - 4m.$$

We set

$$T_{max}(\tau) = \max\{T \in \mathbf{N} \mid 4mT \le \tau \le 1/h - 4mT\};$$

note that T_{max} is well-defined due to our restrictions on τ.

After these preliminary conventions, we pass to constructing the estimate.

2.2.1 Estimates $\hat{f}_\tau[T, y]$

To save notation, in this section we omit when possible explicit indicating τ (which for the time being is fixed); thus, we write T_{max} instead of $T_{max}(\tau)$, etc.

Let T be a positive integer such that $T \le T_{max}$. For $\psi \in \mathcal{P}_{2mT}$ and a finite sequence $\{u_t\}_{t\in I_h}$ we set

$$g[\psi, u] = \{g_t[\psi, u]\},$$

where

$$g_t[\psi, u] = \begin{cases} u_{\tau+t} + \sum_{|s|\le 2mT} \psi_s u_{\tau+t-s}, & |t| \le 2mT \\ 0, & |t| > 2mT \end{cases}$$

(since $t \le T_{max}$, $g[\psi, u]$ is well-defined).

For "design" parameter $\kappa \ge 1$ let us introduce two extremal problems with the control vector $\psi \in \mathcal{P}_{2mT}$:

$(P_T[y])$:

$$\|\psi\|^*_{2mT,1} \to \min$$

$$s.t. \quad \psi \in \mathcal{P}_{2mT}, \quad \|g[\psi, y]\|^*_{2mT,\infty} \le 2\kappa;$$

$(P^*_T[y])$:

$$\|g[\psi, y]\|^*_{2mT,\infty} \to \min$$

$$s.t. \qquad \psi \in \mathcal{P}_{2mT}, \quad \|\psi\|^*_{2mT,1} \le \gamma(T) \equiv 2^{2m+2} m^{1/2} T^{-1/2}.$$

Now, we define $\hat{\psi}[T] = \hat{\psi}[T, y]$ as follows:

- if $(P_T[y])$ is feasible and the optimal value in the problem is $\le \gamma(T)$, we take as $\hat{\psi}[T]$ the optimal solution to $(P_T[y])$;
- in the opposite case (i.e., if $(P_T[y])$ is either unfeasible, or feasible with the optimal value greater than $\gamma(T)$) we take as $\hat{\psi}[T]$ the optimal solution to the problem $(P^*_T[y])$.

Finally, the estimate $\hat{f}_\tau[T, y]$ (let us call it the estimate associated with the *window width* T) of the quantity f_τ is given by

$$\hat{f}_\tau[T, y] = -\sum_{|t|\le 2mT} \hat{\psi}_t[T, y] y_{\tau-t}. \tag{5}$$

We also set

$$\hat{f}_\tau[0, y] = y_\tau.$$

2.2.2 Adaptive estimate $\hat{f}_\tau$

To define the estimate $\hat{f}_\tau$ we actually are interested in, we proceed as follows. Let us call a nonnegative integer $T \le T_{max}$ *good* for a given y, if all the segments

$$D_i = [\hat{f}_\tau[i, y] - \frac{\theta_m \kappa}{\sqrt{i}}, \hat{f}_\tau[i, y] + \frac{\theta_m \kappa}{\sqrt{i}}], \quad i = 1, ..., T,$$

and the segment $D_0 = [y_\tau - \theta_m \kappa, y_\tau + \theta_m \kappa]$ have a point in common; here

$$\theta_m = 2^{4m+9} m^{3/2}. \tag{6}$$

Clearly, good T's exist (e.g., $T = 0$). We define $T^*(y)$ as the largest of these good, for the given y, T's, and set

$$\hat{f}_\tau \equiv \hat{f}_\tau[y] = \hat{f}_\tau[T^*(y), y].$$

Our current goal is to investigate the quality of the introduced estimate.

2.3 Auxiliary results

The quality of the estimate will be closely related to certain random variable Θ, and we start with defining this random variable.

2.3.1 Random variable Θ

Consider a segment $\delta = \delta_{t',t''} = \{t', t'+1, ..., t''\} \subset I_h$ containing an odd number $2N(\delta)+1$ of integers, and let $\bar{t}$ be the midpoint of the segment. Consider the random sequence

$$\zeta_t[\delta] = \xi_{\bar{t}-t}$$

and its Fourier transform

$$\zeta^*[\delta] = F_{N(\delta)} \zeta[\delta].$$

Note that entries $\zeta^*[\delta](\mu)$ of this vector are the standard random $\mathcal{N}(0, I)$ vectors from $\mathbf{C} = \mathbf{R}^2$.

We set

$$\Theta = \max_\delta \max_{\mu \in \Gamma_{N(\delta)}} |\zeta^*[\delta](\mu)|$$

(the outer maximum is taken over all segments $\delta \subset I_h$ containing an odd number of integers). We clearly have the following

Lemma 1 *For any $r > 0$ one has*

$$P\{\Theta > r\} \le h^{-3} \exp\{-\frac{r^2}{2}\}.$$

2.3.2 Main result on adaptive estimation of sequences

We say that $\psi \in \mathcal{P}$ is *normalized of degree* l, l being a positive integer, if

- $\|\psi\|_{l,\infty} = 1$;
- the smallest of t's with nonzero ψ_t is zero, and the largest is l.

Let

$$\Delta : \quad (\Delta\phi)_t = \phi_{t-1}$$

be the shift operator on the space of two-side sequences. Given $\psi \in \mathcal{P}$, we can associate with ψ the finite difference operator $\psi(\Delta)$ on $\mathcal{P}$:

$$(\psi(\Delta)g)_t = \sum_s \psi_s g_{t-s}.$$

Our main intermediate result is the following

Theorem 1 *Assume that $k > 0$ and $l > 0$ are two integers, p is a real from $[1, \infty]$ with*

$$pl > 1, \quad kl \leq m;$$

let also T_+ be a positive integer $\leq T_{max}(\tau)$.

Assume, further, that the sequence f we are estimating satisfies the following condition

> $\mathbf{S}[k, l, m; p; T_+]$: *there exists a decomposition $f = \sum_{j=1}^k f^j$ of the sequence f into the sum of k sequences $f^j = \{f_t^j\}_{t \in \mathbf{Z}}$ with the following property:*
>
> *for every sequence f^j there exists $\phi^j \in \mathcal{P}$, normalized of degree l, such that for the sequences*
>
> $$s^j \equiv \phi^j(\Delta)(\Delta^{-\tau} f^j),$$
>
> *one has*
>
> $$\sum_{j=1}^k \|s^j\|_{4mT_+,p} \leq R \equiv 4m^{-1/2} T_+^{1/p-l-1/2} \kappa. \tag{7}$$

Under these assumptions the inaccuracy of our estimate is bounded from above as

$$|\hat{f}_\tau[y] - f_\tau| \leq 4\theta_m \kappa T_+^{-1/2} \chi_{\{\Theta \leq \kappa\}} + 3\theta_m \kappa \Theta \chi_{\{\Theta > \kappa\}}, \tag{8}$$

where χ_A denotes the characteristic function of event A.

Besides this, independently of assumtion $\mathbf{S}[k, l, m; p; T_+]$ one has

$$|\hat{f}_\tau[y] - f_\tau| \leq 4\theta_m \kappa \chi_{\{\Theta < \kappa\}} + 3\theta_m \kappa \Theta \chi_{\{\Theta > \kappa\}}. \tag{9}$$

Proofs of all statements are presented in Appendix.

3 Nonparametric regression

3.1 The problem

Now we come to our main problem of restoring a function $f : \mathbf{R} \to \mathbf{C}$ at the points of the regular grid $X_h \equiv \{th, t \in I_h\}$ via observations

$$Y_h = \{y(x) = f(x) + \xi(x) \mid x \in X_h\}, \tag{10}$$

where $\{\xi(x)\}_{x\in X_h}$ is a sequence of independent complex–valued normal variables with zero mean and unit covariance matrix.

Given observations (10), we can regard them as the observations of the sequence $\{f_t^h = f(th)\}_{t\in\mathbf{Z}}$ at the moments $t \in I_h$; consequently, the estimate presented in the previous section can be treated as certain estimate $\hat{f}(x) \equiv \hat{f}(x, Y_h)$, $x \in X_h$, of the values of f along the grid X_h. The goal of this section is to evaluate the quality of this estimate.

For any segment $D \subseteq [0,1]$ we denote by D_h the set of observation points contained in D, i.e. $D_h \equiv D \cap X_h$.

We characterize the quality of the estimate $\hat{f}$ at a segment $D \in \mathcal{D}$ by its risk

$$\mathcal{R}_{q,D}(\hat{f}, f) = \mathcal{E}\left[\hat{\mathcal{R}}_{q,D}(\hat{f}, f)\right] \equiv \mathcal{E}\left\{\left[h \sum_{x\in D_h} |\hat{f}(x, Y_h) - f(x)|^q\right]^{1/q}\right\}, \quad q \in [1,\infty], \qquad (11)$$

where the expectation $\mathcal{E}$ is taken over distribution of $\xi(x), x \in X_h$.

3.2 Classes of signals

Recall that our adaptive estimate does not require any a priori knowledge on the family of signals it is applied to; the estimate is completely identified by the pair of design parameters m (the order) and κ (the threshold; in the meantime we shall specify the threshold as an explicit function of m and h, so that in fact the estimate is specified by the design parameter m only). Anyhow, in order to evaluate the quality of the estimate, we should look at its behaviour at certain classes of signals, and we start with specifying the classes we are interested in.

3.2.1 Local behaviour of the signal: classes $\mathbf{A}[\cdot]$

Let k, l be positive integers, $p \in [1,\infty]$ and $R > 0$ be reals such that

$$pl > 1,$$

and let $D = [\hat{x} - \hat{\delta}, \hat{x} + \hat{\delta}] \subset [0,1]$. We say that a function $f : \mathbf{R} \to \mathbf{C}$ satisfies the assumption $\mathbf{A}[k, l; p, R; D]$, if there exists a decomposition

$$f = \sum_{j=1}^{k} f^j$$

of f into a sum of k functions f^j such that f^j are l times continuously differentiable on D, and for every $j \le k$ there exists polynomial

$$q^j(z) = z^l + \sum_{i=0}^{l-1} q_i^j z^i$$

such that

$$\sum_{j=1}^{k} \|q^j(d/dx) f^j\|_{p,D} \le R.$$

From now on, $\|\cdot\|_{p,D'}$ denotes the standard L_p-norm of a $\mathbf{C}$-valued function on a segment D' and $|D'|$ is the length of the segment.

In what follows we use the notation $\mathbf{A}[k,l;p,R;D]$ for both the assumption itself and the family of all signals f satisfying the assumption.

Let us look at typical examples of functions from the classes $\mathbf{A}$.

Example 1 [Sobolev classes] *The standard Sobolev class*

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

is contained in $\mathbf{A}[1,l;p,R;D]$ [set $f^1 = f$, $q^1(z) = z^l$].

Example 2 [Functions satisfying differential inequalities] *Assume that f is an output of an unknown linear filter of order l:*

$$q(d/dx)f = g, \quad [q(z) = z^l + \sum_{i=0}^{l-1} q_i z^i]$$

and the input g is not too big on D: $\|g\|_{p,D} \le R$. Then $f \in \mathbf{A}[1,l;p,R;D]$ [set $f^1 = f$, $q^1 = q$].

Example 3 [Modulated signal] *Let*

$$f(x) = g(x)\cos(\omega x + \phi)$$

be a smooth function $g(x) \in W_p^l(R,D)$ modulated by a sine of certain (unknown) frequency ω. Then $f \in \mathbf{A}[2,l;p,R;D]$ [set $f^1(x) = \frac{1}{2}g(x)\exp\{\mathbf{i}[\omega x + \phi]\}$, $q^1(z) = (z - i\omega)^l$, and $f^2(x) = \frac{1}{2}g(x)\exp\{-\mathbf{i}[\omega x + \phi]\}$, $q^2(z) = (z + i\omega)^l$].

The latter example admits immediate generalization: if

$$f(x) = \sum_{j=1}^{k} g_j(x)\psi_j(x), \quad \psi_j(x) = \frac{\exp\{\lambda_j x\}}{\max_{x\in D}|\exp\{\lambda_j x\}|},$$

with $g_j \in W_p^l(R,D)$, then $f \in \mathbf{A}[k,l;p,kR;D]$ (set $f_j = g_j\psi_j$ and $q^j(z) = (z - \lambda_j)^l$).

The classes $\mathbf{A}$ in what follows play an intermediate role: they describe local behaviour of the signals we are interested in. Now we come to specifying the global behaviour of the signals.

3.2.2 Global behaviour of the signal: classes $\mathbf{F}_d[\cdot]$

Let us fix positive integers k, l, reals $p \in [1,\infty]$, $L > 0$, $d \in (0, 1/6]$, and a segment $D = [\hat{x} - \hat{\delta}, \hat{x} + \hat{\delta}] \subset [d, 1-d]$ with $|D| \ge 4d$.

We say that a function $f : \mathbf{R} \to \mathbf{C}$ belongs to the class $\mathbf{F}_d[k,l;p,L;D]$, if there exist $d(f)$, $|D| \ge d(f) \ge d$, with the following property:

- for any $x \in D_d \equiv [\hat{x} - \hat{\delta} + d, \hat{x} + \hat{\delta} - d]$ the function f on the segment $D_f(x) = [x - d(f), x + d(f)]$ possesses the property $\mathbf{A}[k,l;p,L_f(x)|D_f(x)|^{1/p};D_f(x)]$;

- $L_f(x) \in L_p(D_d)$ is such that $\|L_f(\cdot)\|_{p,D^d} \leq L|D|^{1/p}$.

Our new classes extend the previous ones:

$$\mathbf{A}[k,l;p,L|D|^{1/p};D] \subset \mathbf{F}_d[k,l;p,L';D], \quad L' = k^{1-1/p}L. \tag{12}$$

Indeed, let $f \in \mathbf{A}[k,l;p,L|D|^{1/p};D]$, let $f = \sum_{j=1}^k f^j$ be the decomposition of f given by the definition of the class $\mathbf{A}$, and let q^j be the corresponding polynomials. Setting $d(f) = d$, $x \in D^d$,

$$\rho(x) = \sum_{j=1}^k |(q^j(d/dx)f^j)(x)|,$$

we have

$$\|\rho\|_{p,D} \leq L|D|^{1/p}$$

by definition of the class $\mathbf{A}$. It is immediately seen that for any $x \in D^d$ one has

$$f \in \mathbf{A}[k,l;p,L_f(x)|D_f(x)|^{1/p};D_f(x)]$$

with

$$L_f(x) = (2d)^{-1/p}\sum_{j=1}^k \|q^j(d/dx)f^j\|_{p,[x-d,x+d]} \leq (2d)^{-1/p}k^{1-1/p}\left[\int_{x-d}^{x+d}\rho^p(s)ds\right]^{1/p},$$

so that

$$f \in \mathbf{F}_d[k,l;p,L';D], \;\; |D|^{1/p}L' = (2d)^{-1/p}k^{1-1/p}\left[\int_{D_d}\{\int_{x-d}^{x+d}\rho^p(s)ds\}dx\right]^{1/p} \leq$$

$$\leq k^{1-1/p}\left[\int_D \rho^p(x)dx\right]^{1/p} \leq k^{1-1/p}L|D|^{1/p},$$

as claimed.

Thus, the classes $\mathbf{F}_d$ are at least as wide as the classes $\mathbf{A}$; in fact the former classes are wider, as it is seen from the following

Example 4 ["frequency modulation"]

Let the segment $[0,1]$ be splitted into non-overlapping segments $D_1,...,D_N$ of the length $2d$ each ($2d = 1/N$), and let $x_t = 2td$, $t = 1,...,N$, be the right endpoints of the segments D_i. Let $g(x) \in W_p^l(L,[0,1])$ be an "amplitude" which is assumed to vanish, with its derivatives of orders $\leq l$, at every "switching point" $x_1,...,x_N$. Consider the family of signals f obtained by "frequency modulation" of the amplitude g, namely, such that $f(x) = g(x)\cos(\omega_j x + \phi_j)$, $x \in D_j$, where the "local frequencies" ω_j and "local phases" ϕ_j are arbitrary reals. A signal f from the indicated family belongs to the class $\mathbf{F}_d[4,l;p,2L;[0,1]]$.

Indeed, in the d-neighbourhood $[x-d, x+d]$ of an arbitrary point $x \in [d, 1-d]$ we have representation of the type

$$f(s) = \sum_{\nu=1}^{4} g_\nu(s) \exp\{\mathbf{i}[\omega^\nu s + \phi^\nu]\}$$

with some (depending on x) ω^ν and ϕ^ν and $\sum_{\nu=1}^{4} \|g_\nu^{(l)}\|_{p,[x-d,x+d]} \le 2\|g^{(l)}\|_{p,[x-d,x+d]}$. It follows (cf. Example 3) that $f \in \mathbf{A}[4, l; p, L(x)(2d)^{-1}; [x-d, x+d]]$ with

$$(2d)^{1/p} L(x) = 2\|g^{(l)}\|_{p,[x-d,x+d]};$$

the latter relation implies that $\|L(\cdot)\|_{p,[d,1-d]} \le 2\|g^{(l)}\|_{p,[0,1]}$, so that f indeed belongs to $\mathbf{F}_d[4, l; p, 2L; [0,1]]$.

3.3 Adaptive estimate: quality over classes $\mathbf{A}[\cdot]$

Our current goal is to investigate the quality of the adaptive estimate on a function

$$f \in \mathbf{A}[k, l; p, L|D|^{1/p}; D],$$

where $D = [\hat{x} - \hat{\delta}, \hat{x} + \hat{\delta}] \subset [0,1]$. We are interested in the risk of the estimate at the "narrowed" segment

$$D' = [\hat{x} - \hat{\delta}/2, \hat{x} + \hat{\delta}/2].$$

Theorem 2 *Let f satisfy assumption $\mathbf{A}[k, l; p, L|D|^{1/p}; D]$. Let $m \ge kl$ and let h be small enough, namely, such that*

$$h \le \frac{\hat{\delta}}{40m}. \tag{13}$$

Let c_m be the smallest real $c \ge 1$ such that

$$3\theta_m h^{c^2/2-2} \le 1$$

for all $h < 1/2$, and let the parameter κ underlying the estimate $\hat{f}$ be chosen as

$$\kappa = c_m \sqrt{\ln \frac{1}{h}} \tag{14}$$

Then for the indicated estimate one has

$$\hat{\mathcal{R}}_{q,D'}(\hat{f}, f) \le O(1) \left[\Omega \chi_{\{\Theta \le \kappa\}} + \kappa \Theta |D|^{1/q} \chi_{\{\Theta > \kappa\}}\right], \tag{15}$$

where $O(1)$ depend on m only, and

$$\Omega = |D|^{1/q-1/2} \sqrt{h \ln \frac{1}{h}} + |D|^\sigma L^\phi \left(h \ln \frac{1}{h}\right)^\psi + |D|^{1/q} \left(\frac{L|D|^{1/p}}{\eta_m}\right)^\epsilon \left(h \ln \frac{1}{h}\right)^{l/(2l+1)}, \tag{16}$$

where

$$\eta_m = 2m^{-5/2}(m+1)^{-1}(1+\sqrt{m})^{-m}, \quad \epsilon = \begin{cases} 1, & 2pl^2 \ge 2l+1 \\ pl(pl-1)^{-1}(2l+1)^{-1}, & \text{otherwise} \end{cases} \tag{17}$$

and

- *in the case of* $q \geq p(2l+1)$*:* $\sigma = \dfrac{1-2/q}{2pl+p-2}$; $\phi = \dfrac{p(1-2/q)}{2pl+p-2}$; $\psi = \dfrac{pl+p/q-1}{2pl+p-2}$;
- *in the case of* $q \leq p(2l+1)$*:* $\sigma = 1/q$; $\phi = 1/(2l+1)$; $\psi = l/(2l+1)$.

Besides this,

$$\mathcal{E}\left[\kappa\Theta\chi_{\{\Theta>\kappa\}}\right] \leq h^{1/2}. \tag{18}$$

3.4 Adaptive estimate: quality over classes $\mathbf{F}_d[\cdot]$

Now we are ready to formulate our main result - to evaluate the quality of the adaptive estimate $\hat{f}$ over the classes $\mathbf{F}_d$.

Theorem 3 *Let $\hat{f}$ be the adaptive estimate of order m with the threshold κ chosen according to (14), and let*

$$\mathcal{R}^+(h) \equiv \sup\{\mathcal{R}_{q,D_d}(\hat{f}, f) \mid f \in \mathbf{F}_d[k,l;p,L;D]\}$$

be the uniform risk of the adaptive estimate $\hat{f}$ on the class $\mathbf{F}_d[k,l;p,L;D]$.
Assume that the parameters of the class $\mathbf{F}_d[\cdot]$ are such that $kl \leq m$ and, besides this, that

(i) *h is not too large for the parameters in question, namely,*

$$h\ln\frac{1}{h} \leq \min\left\{\frac{1}{3};\left(\frac{1}{4}|D|\right)^{2l+1} L^2; (160m)^{-1-1/(2l)}; (160m)^{-1-1/(2l)}L^{-\alpha}\right\}, \tag{19}$$

with

$$\alpha = \begin{cases} 2(l(2l+1)p-1), & 2pl^2 \geq 2l+1 \\ (p+2)/(p-1), & \textit{otherwise} \end{cases} \tag{20}$$

(note that α is well defined, since the "otherwise" case may occur only when $l=1$, and here $p>1$ due to the assumption $pl>1$*),*

and

(ii) *d is not too small for a given h, namely,*

$$d \geq \left[L^{-2}h\ln\frac{1}{h}\right]^{1/(2l+1)}. \tag{21}$$

Then

$$\mathcal{R}^+(h) \leq [O(1)]^{\epsilon}|D|^{\sigma}L^{\phi}\left(h\ln\frac{1}{h}\right)^{\psi} \tag{22}$$

with the same $\sigma, \phi, \psi, \epsilon$ as in (16) and with $O(1)$ depending on m only.

3.5 Lower bounds

Here we prove that the risk given by Theorem 3 is optimal in order on the classes $\mathbf{F}_d$, provided that the parameter d of the class in question "fits" the step size h of the observation grid. Namely, let us fix k, l, p, L, D (as always, $pl > 1$ and $D \subset [0,1]$), and let

$$d = d(h) = \left[L^{-2} h \ln \frac{1}{h}\right]^{1/(2l+1)} \tag{23}$$

(cf. (21)). For a fixed $q \in [1,\infty]$ let

$$\mathcal{R}^*(h) = \inf_{f^+} \sup\{\mathcal{R}_{q,D_{d(h)}}(f^+, f) \mid f \in \mathbf{F}_{d(h)}[k,l;p,L,D]\}$$

(inf is taken over all estimates $f^+(\cdot)$) be the minmax risk associated with the class $\mathbf{F}_{d(h)}[k,l;p,L;D]$ and the inaccuracy measure given by q. The announced "optimality in order" of the adaptive estimate $\hat{f}$ is given by the following

Proposition 1 *Let $k \geq 4$, $m \geq kl$ and let h be small enough, namely, let it satisfy (19) and be such that*

$$d(h) \leq |D|/40.$$

Then

$$\mathcal{R}^*(h) \geq O(1)|D|^{\sigma} L^{\phi} \left(h \ln \frac{1}{h}\right)^{\psi} \tag{24}$$

with the same σ, ϕ, ψ as in (16) and with positive $O(1)$ depending on p, q, l only. In particular,

$$\mathcal{R}^*(h) \geq C(p,q,k,l,m)\mathcal{R}^+(h).$$

Proof. Let $\phi(x)$ be a once for ever fixed C^∞ function on the axis which is equal to 1 for $|x| \leq 1/2$ and is identically zero for $|x| \geq 1$. In what follows $d = d(h)$, and C_i denote positive quantities depending on l only.

1^0. Consider first the case of $q < p(2l+1)$. Let Δ_i $i \in \mathbf{Z}$, be the sequence of segments of the length $2d$ each, with Δ_0 starting at the left endpoint of the segment D_d and Δ_{i+1} starting at the right endpoint of Δ_i, and let N be the number of those of Δ_i which are covered by D_d. Let Δ_i', $0 \leq i \leq N-1$, be the concentric to Δ_i twice smaller segments, let X_i be the sets of observation points contained in Δ_i' and let M be the smallest, over $i = 0, ..., N-1$, of cardinalities of X_i. Note that

$$N \geq C_1 |D| d^{-1}; \quad M \geq C_2 d/h \geq 1. \tag{25}$$

Now, let $C_3 = \|\phi^{(l)}(\cdot)\|_\infty$, and let $\mathbf{F}^0$ be the family of signals f defined as follows: f vanishes outside $\cup_{i=0}^N \Delta_i$, and in every Δ_i, $0 \leq i \leq N-1$ f is of the form

$$f(x) = \vartheta \frac{d^l L}{2C_3} \phi\left(\frac{x - x^i}{d}\right) \cos(\omega x),$$

where $\vartheta \in (0,1)$ will be specified later, x^i is the midpoint of Δ_i and $\omega = \omega_{i,f}$ is a frequency of the type $2\pi M^{-1}h^{-1}\nu$, ν being an integer from $\{0, 1, ..., M-1\}$.

Note that $\mathbf{F}^0 \subset \mathbf{F} \equiv \mathbf{F}_{d(h)}[k, l; p, L; D]$ (cf. Example 4). Besides this, we clearly have the following:

(*) *for any $f, g \in \mathbf{F}^0$ and any i either $f \equiv g$ on Δ_i, or*

$$\rho_i(f,g) \equiv h \sum_{x \in X_i} |f(x) - g(x)|^q \geq \vartheta^q C_4^q d^{1+lq} L^q.$$

Now, by the standard reasons there exists a family comprised of

$$S \geq M^{C_5|D|/d} \tag{26}$$

functions $f_1, ..., f_S \in \mathbf{F}^0$ with the following property:

(**) *if $j \neq j'$, then the functions f_j and $f_{j'}$ differ from each other at at least $N/4$ of the segments Δ_1,...,Δ_N.*

Combining (25), (*) and (**), we conclude that

$$h \sum_{x \in X} |f_j(x) - f_{j'}(x)|^q \geq \delta^q \equiv \vartheta^q C_5^q d^{lq} L^q |D| \tag{27}$$

whenever $j \neq j'$, $1 \leq j, j' \leq S$; here $X = X_h \cap D_d$ is the set of observation points in D_d.

Let us prove that for appropriately chosen (as a function of l only) ϑ one has

$$\mathcal{R}^*(h) \geq \delta/16. \tag{28}$$

Indeed, assume that the latter relation is not valid for a given ϑ. Then, due to (27), there clearly exists a routine for distinguishing between S hypotheses H_j on observations, j-th of them being that the observed signal is f_j, with the following property: if H_j is valid, then the probability of wrong answer of the routine is $\leq 1/4$. Applying the Fano lemma (see, e.g., [6]), we come to

$$\mathcal{K} \geq \frac{1}{2} \ln S, \tag{29}$$

where $\mathcal{K}$ is the largest of the Kullback information distances between the distributions of the observations related to our hypotheses. It is immediately seen that

$$\mathcal{K} = \max_{j \neq j', i, j' \leq S} \sum_{x \in X} |f_j(x) - f_{j'}(x)|^2 \leq C_6 \vartheta^2 |D| h^{-1} d^{2l} L^2. \tag{30}$$

Combining (29), (30), (26) and (25), we come to the inequality

$$C_6 |D| d^{-1} \ln \frac{1}{h} \leq C_7 \vartheta^2 |D| h^{-1} d^{2l} L^2,$$

or, which is the same,

$$\vartheta^2 \geq C_8 d^{-(2l+1)} L^{-2} h \ln \frac{1}{h} \geq C_9$$

(the concluding inequality follows from the definition of $d = d(h)$, see (23)). Setting $\vartheta = C_9^{1/2}/2$, we make the latter relation false; consequently, for this choice of ϑ (28) indeed holds true. Relation (28) immediately implies the lower bound announced in (24) for the case of $q < p(2l+1)$.

2^0. In the case of $q \geq p(2l+1)$ (24) was in fact established in [10], where it is proved that the required lower bound is valid already for the Sobolev classes. ■

4 Concluding remarks

In conclusion it should be noticed that the aforementioned results remain valid under weaker assumptions on the noise sequence $\{\xi(x)\}_{x \in X_h}$. Indeed, all we need from the noise is the exponentially decreasing tails for distribution of the random variable Θ (see, Lemma 1). It can be easily checked that the statement like Lemma 1 holds provided the noise sequence is defined as

$$\xi(\tau h) = \sum_{t=-\infty}^{\infty} \nu_t \eta_{\tau - t},$$

where $\{\eta_t\}_{t=-\infty}^{t=\infty}$ is a sequence of normal complex–valued random variables with zero mean and unit covariance matrix, and $\sum_t |\nu_t| \leq 1$.

5 Appendix

5.1 Proof of Theorem 1

0^0. Note the evident relations

$$\|\phi\psi\|_{N,p} \leq \|\phi\|_1 \|\psi\|_{N+d(\phi),p},$$

$$\|\phi\|_{N,2} = \|\phi\|^*_{N,2} \tag{31}$$

(the Parseval equality),

$$\|\phi\|_{N,1} \leq \|\phi\|^*_{N,1}\sqrt{2N+1},$$

$$\|\phi\|^*_{N,\infty} \leq (2N+1)^{1/2-1/p}\|\phi\|_{N,p}, \tag{32}$$

$$d(\phi) + d(\psi) \leq N \Rightarrow \|\phi\psi\|^*_{N,1} \leq \|\phi\|_{N,1}\|\psi\|^*_{N,1}.$$

Further, let

$$(\Delta\phi)_t = \phi_{t-1}$$

be the shifting operator on $\mathcal{P}$. We clearly have

$$\|\phi\psi\|^*_{N,\infty} \leq \|\phi\|_1 \max_{|t| \leq d(\phi)} \|\Delta^t \psi\|^*_{N,\infty} \leq \|\phi\|_1 \|\psi\|_{N+d(\phi),p}(2N+1)^{1/2-1/p} \tag{33}$$

(the concluding inequality follows from (32)).

1^0. We start with the following

Lemma 2 [11] *Let $\phi^j \in \mathcal{P}$ be normalized of degree l, $j = 1, ..., k$. Then, for any $T \geq 1$, there exists $\phi \in \mathcal{P}_{2klT}$ with the following properties*

(i) $\phi(z) = \delta(z) + \omega(z)$, $\delta(z) \equiv 1$ being the convolution unit, with

$$\|\phi\|_1 \leq 2^{kl} \tag{34}$$

and

$$\|\omega\|^*_{N,1} \leq 2^{2kl} \frac{\sqrt{2N+1}}{T}, \;\; \forall N \geq 2klT; \tag{35}$$

(ii) for any $j \leq k$ there exists representation

$$\phi(z) = \phi^j(z)\rho^j(z)$$

with $\rho^j \in \mathcal{P}_{2klT}$ such that

$$\|\rho^j\|_\infty \leq 2^{2kl-1} T^{l-1}. \tag{36}$$

2^0. We proceed as follows.

Lemma 3 *For any T, $1 \leq T \leq T_+$, one has*

$$|\hat{f}_\tau[T, y] - f_\tau| \leq 2^{4m+8} m^{3/2} \frac{\max\{\kappa, \Theta\}}{\sqrt{T}} + 2^{4m+6} m^2 T^{l-1/p} R. \tag{37}$$

Proof a) Let ϕ be the element of $\mathcal{P}$ given, for the value of T in question, by Lemma 2. We have

$$\phi(\Delta)\Delta^{-\tau} f = \sum_{j=1}^k \phi(\Delta)\Delta^{-\tau} f^j = \sum_{j=1}^k \rho^j(\Delta)\phi^j(\Delta)\Delta^{-\tau} f^j = \sum_{j=1}^k \rho^j(\Delta) s^j.$$

Consequently,

$$\|\phi(\Delta)\Delta^{-\tau} f\|^*_{2mT,\infty} \leq$$

[since $\rho^j \in \mathcal{P}_{2klT} \subset \mathcal{P}_{2mT}$ due to $kl \leq m$ and in view of (33) applied with $N = 2mT$]

$$\leq \sum_{j=1}^k \|\rho^j\|_1 \|s^j\|_{4mT,p} (2mT+1)^{1/2-1/p} \leq$$

[due to (36) and the evident relation $\|\rho^j\|_1 \leq (4klT+1)\|\rho^j\|_\infty$; recall that $\rho^j \in \mathcal{P}_{2klT}$]

$$\leq (4mT+1) 2^{2kl-1} T^{l-1} (2mT+1)^{1/2-1/p} \sum_{j=1}^k \|s^j\|_{4mT,p} \leq 2^{2m+3} m^{3/2} T^{l+1/2-1/p} R.$$

(the concluding inequality follows from (7) due to $T \leq T_+$).

By construction, the very first expression in the latter chain is exactly $\|g[\omega, f]\|^*_{2mT,\infty}$, so that we have proved that

$$\|g[\omega, f]\|^*_{2mT,\infty} \leq 2^{2m+3} m^{3/2} T^{l+1/2-1/p} R. \tag{38}$$

b) We have

$$\|g[\omega,\xi]\|^*_{2mT,\infty} = \|\phi(\Delta)\Delta^{-\tau}\xi\|^*_{2mT,\infty} \leq$$

[since $\phi \in \mathcal{P}_{2lkT} \subset \mathcal{P}_{2mT}$]

$$\leq \sum_{|t|\leq 2mT} |\phi_t| \|\Delta^{t-\tau}\xi\|^*_{2mT,\infty} \leq$$

[since $T \leq T_+ \leq T_{max}(\tau)$ and by definition of Θ]

$$\leq \|\phi\|_1 \Theta,$$

whence, in view of (34),

$$\|g[\omega,\xi]\|^*_{2mT,\infty} \leq 2^m \Theta. \tag{39}$$

Note that a byproduct of our reasoning is as follows:

Lemma 4 *For any $\psi \in \mathcal{P}_{2mT}$ one has*

$$\|g[\psi,\xi]\|^*_{2mT,\infty} \leq \|\delta + \psi\|_1 \Theta.$$

c) Combining (38) and (39), we come to the following conclusion:

A. *There exists $\omega \in \mathcal{P}_{2mT}$ such that*

$$\|g[\omega,y]\|^*_{2mT,\infty} \leq \Omega \equiv 2^{2m+3} m^{3/2} T^{l+1/2-1/p} R + 2^m \max\{\kappa, \Theta\}, \tag{40}$$

$$\|\omega\|^*_{2mT,1} \leq 2^{2m+2} m^{1/2} T^{-1/2} \equiv \gamma(T)$$

(the latter relation follows from (35) applied with $N = 2mT$; $\gamma(T)$ was defined in formulation of the problem $(P^*_T[y])$).

d) From **A** it follows that

B. *$\hat{\psi}[T,y]$ possesses the following properties:*

$$\hat{\psi}[T,y] \in \mathcal{P}_{2mT}; \tag{41}$$

$$\|\hat{\psi}[T,y]\|^*_{2mT,1} \leq \gamma(T) = 2^{2m+2} m^{1/2} T^{-1/2}; \tag{42}$$

$$\|g[\hat{\psi}[T,y],y]\|^*_{2mT,\infty} \leq \Omega. \tag{43}$$

Indeed, (41) - (42) are readily given by the construction, see Section 2.2. To get (43), note that if $\hat{\psi}$ is given by the solution of (P_T), then (43) is valid even with the right hand side replaced with 2κ - this is the constraint in (P_T); if $\hat{\psi}$ is given by the solution of (P^*_T), then (43) follows from the fact that the right hand side in (43) majorates the value of the objective of the problem (P^*_T) at its feasible solution ω, see (40).

e) Now we are ready to evaluate the inaccuracy of the estimate $\hat{f}_\tau[T,y]$. In what follows we write $\hat{f}_\tau$ and $\hat{\psi}$ instead of $\hat{f}_\tau[T,y]$, $\hat{\psi}[T,y]$, respectively. We clearly have (see (5))

$$|\hat{f}_\tau - f_\tau| \leq \epsilon_1 + \epsilon_2, \tag{44}$$

with

$$\epsilon_1 = |g_0[\hat{\psi}, \xi] - \xi_\tau|$$

and

$$\epsilon_2 = |f_\tau + \sum_{|t| \le 2mT} \hat{\psi}_t f_{\tau - t}| = |d_0|,$$

where

$$d = g[\hat{\psi}, f] = (1 + \hat{\psi}(\Delta))\Delta^{-\tau} f. \tag{45}$$

Now, let $\mathcal{J}$ be the standard involution in $\mathcal{P}$:

$$(\mathcal{J}\psi)_t = \bar{\psi}_{-t}.$$

By definition of ϵ_1, we have

$$\epsilon_1 = |\langle \hat{\psi}, \mathcal{J}\Delta^{-\tau}\xi \rangle_{2mT}| =$$

[see (4)]

$$= |\langle F_N \hat{\psi}, F_N \mathcal{J}\Delta^{-\tau}\xi \rangle| \le \|F_N \hat{\psi}\|_1 \|F_N \mathcal{J}\Delta^{-\tau}\xi\|_\infty = \|F_N \hat{\psi}\|_1 \|F_N \Delta^{-\tau}\xi\|_\infty;$$

this inequality combined with (42) and the definition of Θ results in

$$\epsilon_1 \le \gamma(T)\Theta. \tag{46}$$

Now let us evaluate ϵ_2. From (43) we know that

$$\|d + g[\hat{\psi}, \xi]\|^*_{2mT,\infty} = \|g[\hat{\psi}, y]\|^*_{2mT,\infty} \le \Omega \tag{47}$$

(d is given by (45)). Besides this, from Lemma 4

$$\|g[\hat{\psi}, \xi]\|^*_{2mT,\infty} \le \|\delta + \hat{\psi}\|_1 \Theta \le (1 + \|\hat{\psi}\|_1)\Theta. \tag{48}$$

Since $\hat{\psi} \in \mathcal{P}_{2mT}$ (see (41)), we have

$$\|\hat{\psi}\|_1 = \|\hat{\psi}\|_{2mT,1} \le \sqrt{4mT + 1} \|\hat{\psi}\|_{2mT,2} =$$

[The Parseval equality (31)]

$$= \sqrt{4mT+1}\|F_{2mT}\hat{\psi}\|_2 \le \sqrt{4mT+1}\|F_{2mT}\hat{\psi}\|_1 = \sqrt{4mT+1}\|\hat{\psi}\|^*_{2mT,1} \le \sqrt{4mT+1}\gamma(T)$$

(the concluding inequality follows from (42)). Thus,

$$1 + \|\hat{\psi}\|_1 \le 1 + \sqrt{4mT+1}\gamma(T) \le 2^{2m+4} m \tag{49}$$

(see (42)). Combining this result with (48), we come to

$$\|g[\hat{\psi}, \xi]\|^*_{2mT,\infty} \le 2^{2m+4} m\Theta. \tag{50}$$

From (47), (50) and the triangle inequality we get

$$\|d\|^*_{2mT,\infty} \le \Omega + 2^{2m+4} m\Theta. \tag{51}$$

Now comes the crucial point. We have

$$d = (1 + \hat{\psi}(\Delta))\Delta^{-\tau} f,$$

whence for ϕ given by Lemma 2 one has

$$\phi(\Delta)d = (1 + \hat{\psi}(\Delta))\phi(\Delta)\Delta^{-\tau} f = (1 + \hat{\psi}(\Delta))\phi(\Delta)\sum_{j=1}^{k} \Delta^{-\tau} f^j =$$

$$= (1 + \hat{\psi}(\Delta))\sum_{j=1}^{k} \rho^j(\Delta)\phi^j(\Delta)\Delta^{-\tau} f^j = (1 + \hat{\psi}(\Delta))\sum_{j=1}^{k} \rho^j(\Delta)s^j.$$

Looking at the resulting equality between sequences at $t = 0$, we get

$$d_0 + (\omega(\Delta)d)_0 \equiv (\phi(\Delta)d)_0 = \left((1 + \hat{\psi}(\Delta))\sum_{j=1}^{k} \rho^j(\Delta)s^j\right)_0,$$

whence

$$\epsilon_2 = |d_0| \le |(\omega(\Delta)d)_0| + |\left((1 + \hat{\psi}(\Delta))\sum_{j=1}^{k} \rho^j(\Delta)s^j\right)_0| \equiv \epsilon_{2,1} + \epsilon_{2,2}. \tag{52}$$

We have

$$\epsilon_{2,1} = |(\omega(\Delta)d)_0| = |\langle \omega, \mathcal{J}d\rangle_{2mT}| =$$

[see (3)]

$$= |\langle F_{2mT}\omega, F_{2mT}\mathcal{J}d\rangle| \le \|F_{2mT}\omega\|_1 \|F_{2mT}\mathcal{J}d\|_\infty = \|\omega\|^*_{2mT,1}\|d\|^*_{2mT,\infty} \le$$

[Lemma 2 applied with $N = 2mT$ and (51)]

$$\le (2^{2m+2}m^{1/2}T^{-1/2})(\Omega + 2^{2m+4}m\Theta).$$

Thus,

$$\epsilon_{2,1} \le (2^{2m+2}m^{1/2}T^{-1/2})(\Omega + 2^{2m+4}m\Theta). \tag{53}$$

It remains to evaluate $\epsilon_{2,2}$. We have

$$\epsilon_{2,2} \equiv |\left((1 + \hat{\psi}(\Delta))\sum_{j=1}^{k} \rho^j(\Delta)s^j\right)_0| \le$$

[since $\hat{\psi} \in \mathcal{P}_{2mT}$]

$$\le \|\delta + \hat{\psi}\|_1 \sum_{j=1}^{k} \|\rho^j s^j\|_{2mT,\infty} \le$$

[(49)]

$$\le 2^{2m+4}m \sum_{j=1}^{k} \|\rho^j s^j\|_{2mT,\infty} \le$$

[since $\|\rho^j s^j\|_{2mT,\infty} \le \|\rho^j\|_{2mT,\infty}\|s^j\|_{4mT,p}(2mT+1)^{1-1/p}$ by Hölder inequality and due to $\rho^j \in \mathcal{P}_{2mT}$, and because of $\|\rho^j\|_{2mT,\infty} \le 2^{2m-1}T^{l-1}$, see (36)]

$$\le 2^{4m+3}mT^{l-1}(2mT+1)^{1-1/p}\sum_{j=1}^{k}\|s^j\|_{4mT,p} \le 2^{4m+5}m^2T^{l-1/p}R.$$

Thus,

$$\epsilon_{2,2} \le 2^{4m+5}m^2T^{l-1/p}R.$$

Combining (44), (46), (52) and (53), we get

$$|\hat{f}_\tau - f_\tau| \le \epsilon_1 + \epsilon_2 \le \gamma(T)\Theta + (2^{2m+2}m^{1/2}T^{-1/2})(\Omega + 2^{2m+4}m\Theta) + 2^{4m+5}m^2T^{l-1/p}R.$$

Substituting

$$\gamma(T) = 2^{2m+2}m^{1/2}T^{-1/2}$$

and

$$\Omega = 2^m\max\{\kappa;\Theta\} + 2^{2m+3}m^{3/2}T^{l+1/2-1/p}R,$$

we come to

$$|\hat{f}_\tau - f_\tau| \le 2^{4m+8}m^{3/2}\frac{\max\{\kappa,\Theta\}}{\sqrt{T}} + 2^{4m+6}m^2T^{l-1/p}R,$$

as required in (37). Lemma 3 is proved.

3^0. Now let

$$\Xi = \{\xi \mid \Theta \le \kappa\}.$$

For $\xi \in \Xi$ and $T \le T_+$ the right hand side of (37) does not exceed

$$2^{4m+8}m^{3/2}\frac{\kappa}{\sqrt{T}} + 2^{4m+6}m^2T^{l-1/p}R \le 2^{4m+9}m^{3/2}\frac{\kappa}{\sqrt{T}} = \theta_m\frac{\kappa}{\sqrt{T}}$$

(we have used (6) and (7)). It follows that for $\xi \in \Xi$ all segments D_i, $1 \le i \le T^+$, (see Section 2.2.2) have a point in common, namely, f_τ; in particular,

$$|\hat{f}_\tau[T_+, f+\xi] - f_\tau| \le \theta_m T_+^{-1/2}\kappa.$$

The point f_τ belongs also to the segment $D_0 = [y_\tau - \theta_m\kappa, y_\tau + \theta_m\kappa]$, since one clearly has $|y_\tau - f_\tau| = |\xi_\tau| \le \Theta$, and $\Theta \le \kappa$ for $\xi \in \Xi$. Consequently, the estimate $\hat{f}_\tau[f+\xi]$ for these ξ is certain $\hat{f}_\tau[T, f+\xi]$ with $T \ge T_+$, and by construction the segment D_T intersects with D_{T_+}. It follows that

$$|\hat{f}_\tau[f+\xi] - f_\tau| \le |\hat{f}_\tau[f+\xi] - \hat{f}_\tau[T_+, f+\xi]| + |\hat{f}_\tau[T_+, f+\xi] - f_\tau| \le$$

$$\le \theta_m\kappa[T^{-1/2} + T_+^{-1/2} + 2T_+^{-1/2}] \le 4\theta_m\kappa T_+^{-1/2}$$

(note that $T \ge T^+$). Thus,

$$\xi \in \Xi = \{\xi \mid \Theta \le \kappa\} \Rightarrow |\hat{f}_\tau[f+\xi] - f_\tau| \le 4\theta_m\kappa T_+^{-1/2}.$$

Now let $\xi \notin \Xi$. By construction, $\hat{f}_\tau[f+\xi]$ is certain $\hat{f}_\tau[T, f+\xi]$, where D_T intersects with D_0. It follows that in the case in question

$$|\hat{f}_\tau[f+\xi]-f_\tau| \le |\hat{f}_\tau[f+\xi]-y_\tau|+|\xi_\tau| \le \frac{1}{2}[|D_0|+|D_T|]+|y_\tau-f_\tau| \le 2\theta_m\kappa+\Theta \le 3\theta_m\kappa\Theta. \quad (54)$$

The proof of (8) is completed.

4^0. It remains to verify (9). To this end note that the relation (54) is given by the contsruction of our estimate and is independent of any assumptions on f; thus,

$$|\hat{f}_\tau[f+\xi]-f_\tau| \le |D_0| + |\xi_\tau| = 2\theta_m\kappa + |\xi_\tau|.$$

If $\xi \in \Xi$, the right hand side in this inequality is $\le 2\theta_m\kappa + \kappa \le 3\theta_m\kappa$, and for all ξ the right hand side is at most $3\theta_m\kappa\Theta$. ∎

5.2 Proof of Theorem 2

To avoid extra comments, in the below proof we restrict ourselves to the case of $q < \infty$; the announced result for the case $q = \infty$ can be obtained by straightforward passing to limit in (15) – (16) as finite q tends to ∞.

1^0. According to (9), in the case of $\Theta > \kappa$ the left hand side in (15) is bounded from above by the quantity $3\theta_m\kappa\Theta(Nh)^{1/q}$, N being the cardinality of the set D'_h, independently of any hypotheses on f. We clearly have $N \le 2|D|/h$, and we conclude that (15) indeed takes place if $\Theta > \kappa$. Note also that (18) is an immediate consequence of Lemma 1 and our choice of κ. With these remarks we see that all we need is to prove (15) for the case of the event $\Theta \le \kappa$.

2^0. For a function $f : \mathbf{R} \to \mathbf{C}$ let f^h be its restriction onto the grid $\{th\}_{t\in\mathbf{Z}}$: $f^h_t \equiv f_t = f(th)$, $t = 0, \pm1, \pm2, ...$ We need the following lemma (cf. [11], Lemma 6):

Lemma 5 *Let $h > 0$, and let $f \in \mathbf{A}[k,l;p,R;D]$. Assume that the midpoint $\hat{x}$ of D belongs to the grid:*

$$\hat{x} = \tau h \quad [\tau \in \mathbf{Z}],$$

and that the half-length $\hat{\delta}$ of the segment is at least lh. Let $T = \lfloor \hat{\delta} h^{-1} \rfloor$; then there exist decomposition

$$f^h = \sum_{j=1}^{k} f^{h,j}$$

and polynomials ϕ^j, $j = 1, ..., k$, normalized of degree l, such that the sequences

$$s^{h,j} = \phi^j(\Delta)\Delta^{-\tau} f^{h,j}$$

satisfy the relation

$$\sum_{j=1}^{k} \|s^{h,j}\|_{T-l,p} \le kl(l+1)(1+\sqrt{l})^l h^{l-1/p} R \equiv \nu_l h^{l-1/p} R. \quad (55)$$

Proof It clearly suffices to prove the lemma for the case of $k = 1$. Thus, assume that there exists a polynomial $q(z) = z^l + q_{l-1}z^{l-1} + ... + q_0$ such that for $g = q(d/dx)f$ we have $\|g\|_{p,D} \le R$. Let $\lambda_1, ..., \lambda_l$ be the roots of the polynomial q; without loss of generality we may assume that $\lambda_1, ..., \lambda_n$ belong to the closed right half-plane, while $\lambda_{n+1}, ..., \lambda_l$ belong to the open left half-plane. Let $\mu_i = \exp\{-\lambda_i h\}$, and let

$$\bar{\phi}(z) = \left(\prod_{i=1}^{n}(z - \mu_i)\right)\left(\prod_{i=n+1}^{l}(1 - \mu_i^{-1}z)\right) \equiv \phi_1(z)\phi_2(z).$$

Let also

$$\gamma_i(x) = \exp\{\lambda_i x\}\chi_{\{x\ge 0\}},\ i = 1, ..., n, \quad \gamma_i(x) = -\exp\{\lambda_i x\}\chi_{\{x\le 0\}},\ i = n+1, ..., l,$$

where, as always, $\chi_{\{x\in A\}}$ is the characteristic function of the set A; note that γ_i is the Green function of the differential operator $q_i(d/dx) = d/dx - \lambda_i$: $q_i(d/dx)\gamma_i = \delta(x)$, $\delta(x)$ being the standard δ-function. Let us set

$$\gamma = \gamma_1 * \ldots * \gamma_l$$

($*$ denotes the convolution on the axis). Then, by construction, $q(d/dx)\gamma = \delta$, whence

$$\bar{f}(x) \equiv f(x - \hat{x}) = (\gamma * \bar{g})(x) + \psi(x) \equiv r(x) + \psi(x),$$

where $\bar{g}(x) = g(x - \hat{x})\chi_{\{x\in D\}}$ and $\psi(x)$ satisfies the differential equation

$$q(d/dx)\psi = 0,\ |x| \le \hat{\delta}.$$

Let w^h denote the restriction of a function $w(\cdot)$ on the axis onto the grid $\{th\}_{t\in\mathbf{Z}}$. Let us prove that

$$\|\bar{\phi}(\Delta)\bar{f}^h\|_{T-l,p} \le lh^{l-1/p}R. \tag{56}$$

Indeed, due to the origin of ψ^h we have $(\bar{\phi}(\Delta)\psi^h)_t = 0$ when $t \le T - l$, so that the left hand side in (56) is nothing but

$$\|\rho\|_{T-l,p}, \quad \rho = \bar{\phi}(\Delta)r^h.$$

Let $\mathcal{T}$ be the shifting operator $w(\cdot) \mapsto w(\cdot - h)$ on the space of functions on the axis; we have $\Delta(v * w)^h = ((\mathcal{T}v) * w)^h$, and, consequently,

$$\rho_t = \left[(\Delta - \mu_1)...(\Delta - \mu_n)(1 - \frac{1}{\mu_{n+1}}\Delta)...(1 - \frac{1}{\mu_l}\Delta)[\gamma_1 * ...\gamma_l * \bar{g}]^h\right]_t =$$

$$= \left[(\mathcal{T}\gamma_1 - \mu_1\gamma_1) * ... * (\mathcal{T}\gamma_n - \mu_n\gamma_n) * (\gamma_{n+1} - \mu_{n+1}^{-1}\mathcal{T}\gamma_{n+1}) * ... * (\gamma_l - \mu_l^{-1}\mathcal{T}\gamma_l) * \bar{g}\right]_t^h \equiv$$

$$\equiv [v_1 * ... * v_l * \bar{g}]_t^h \equiv [v * \bar{g}]_t^h.$$

It is immediately seen that the functions v_i vanish outside $[0, h]$ for $1 \leq i \leq l$, while the absolute values of the functions do not exceed 1. It follows that function $v = v_1 * ...v_l$ vanishes outside $[0, lh]$ and $\|v\|_\infty \leq h^{l-1}$. Consequently, for $p < \infty$ and $t \leq T - l$ we have

$$|\rho_t|^p = |\int_0^{lh} v(s)\bar{g}(th - s)ds|^p \leq h^{p(l-1)}[\int_0^{lh} |\bar{g}(th - s)|ds]^p \leq$$

$$\leq h^{p(l-1)} \left[(lh)^{p-1} \int_{th}^{th+lh} |\bar{g}(s)|^p ds\right]$$

whence

$$\|\rho\|_{T-l,p} \leq \left[h^{pl-1} l^{p-1} l \int_{-Th}^{Th} |\bar{g}(s)|^p ds\right]^{1/p} \leq l h^{l-1/p} \|g\|_{p,D},$$

as required in (56).

Now let w be the maximum of absoulte values of the coefficients of $\bar{\phi}$, and let $\phi(z) = w^{-1}\bar{\phi}(z)$, so that ϕ is a normalized polynomial of the degree l. From (56) it follows that

$$\|\phi(\Delta)\bar{f}^h\|_{T-l,p} \leq w^{-1} l h^{l-1/p} R; \tag{57}$$

the left hand side in this relation is exactly the quantity $\|s^h\|_{T-l,p}$ associated with the choice $\phi^1 = \phi$ (recall that we are considering the case of $k = 1$). Thus, to complete the proof of the lemma we should verify that the right hand side in our inequality does not exceed that one in the inequality (55). This is immediate: there evidently exists a point z^* in the unit circle which is at the distance at least $d = (1 + \sqrt{l})^{-1}$ to each of the points $\mu_1, ..., \mu_n$ and at least at the same distance from the boundary of the circle. We have, consequently,

$$(l + 1)w \geq |\bar{\phi}(z^*)| \geq d^l,$$

whence

$$w^{-1} \leq (l + 1)(1 + \sqrt{l})^l,$$

which combined with (57) results in (55). ■

3^0. There are two possible cases: that one of "small" h:

$$h^{l-1/p} L |D|^{1/p} \leq \eta_m \kappa, \tag{58}$$

with η_m given by (17), and the opposite case ("big" h). In the case of "big" h (9) implies that

$$\left[h \sum_{x \in D'_h} |\hat{f}(x) - f(x)|^q\right]^{1/q} \leq O(1)|D|^{1/q}\kappa. \tag{59}$$

Now, in the case of "big" h we have, due to $\kappa \geq 1$ (see (13) and (14))

$$\frac{1}{h} \leq (\eta_m^{-1} L |D|^{1/p})^{p/(pl-1)} \tag{60}$$

and

$$\kappa \leq \eta_m^{-1} L |D|^{1/p} h^{l-1/p}. \tag{61}$$

In the case of $2pl^2 \geq 2l+1$ we have $l - 1/p \geq l/(2l+1)$, and (60) and (61) imply that

$$\kappa \leq \eta_m^{-1} L|D|^{1/p} h^{l/(2l+1)};$$

in the case of $2pl^2 < 2l+1$ we have $l - 1/p < l/(2l+1)$ and therefore can rewrite (61) as

$$\kappa \leq \eta_m^{-1} L|D|^{1/p} (1/h)^{p^{-1}-l+l(2l+1)^{-1}} h^{l/(2l+1)} \leq$$

[see (60)]

$$\left(\eta_m^{-1} L|D|^{1/p}\right)^{1+p(pl-1)^{-1}(p^{-1}-l+l(2l+1)^{-1})} h^{l/(2l+1)} = (\eta_m^{-1} L|D|^{1/p})^{pl(pl-1)^{-1}(2l+1)^{-1}} h^{l/(2l+1)}.$$

The resulting upper bounds on κ demonstrate that the right hand side of (59) does not exceed[1]

$$O(1)|D|^{1/q} (\eta_m^{-1} L|D|^{1/p})^{\epsilon} \left(h \ln \frac{1}{h}\right)^{l/(2l+1)}.$$

Thus, (15) – (16) are valid in the case of "big" h.

From now on we can assume that h is "small".

4^0. Let us fix $f \in \mathbf{A}[k, l; p, |D|^{1/p} L; D]$, and let f^j and q^j be the corresponding functions and polynomials.

Consider first the case of $p < \infty$.

Let us associate with f certain partitioning of the segment D'. Denote

$$D = [u, v]; \quad D' = [u', v']; \quad \zeta = \kappa^2 h; \quad \beta = \frac{p}{2pl + p - 2}.$$

4^0.1. Given a point $x \in D'' = [\frac{1}{2}(u+u'), \frac{1}{2}(v'+v)]$, denote by $\delta_+(x)$ the largest $\delta \leq \hat{\delta}/4$ such that

$$\delta \leq 10m(\eta_m^2 \zeta)^{\beta} \left[\int_x^{x+\delta} R^p(s)ds\right]^{-2\beta/p}, \qquad R(s) = \sum_{j=1}^{k} |(q^j(d/ds)f^j)(s)|.$$

Let us set

$$\rho(x) = \int_x^{x+\delta_+(x)} R^p(s)ds;$$

then

$$\delta_+(x) \leq 10m(\eta_m^2 \zeta)^{\beta} [\rho(x)]^{-2\beta/p}, \tag{62}$$

with the inequality being equality for *proper* x, i.e., those with $\delta_+(x) < \hat{\delta}/4$.

Note that

$$\rho(x) \leq \int_D R^p(s)ds \leq |D| L^p \tag{63}$$

due to the inclusion $f \in \mathbf{A}$.

4^0.2. Let us set $u_1 = \frac{1}{2}(u+u')$, $u_2 = u_1 + \delta_+(u_1)$, $u_3 = u_2 + \delta_+(u_2)$,...; we terminate this construction at the step N, when it turns out that $\frac{1}{2}(u_N + u_{N+1}) \geq v'$. Due to (63) and

[1] to simplify forthcoming considerations, we replace h by larger (due to (19)) quantity $h \ln(1/h)$

the fact that (62) is an equality for proper x the quantities $\delta_+(x)$ are bounded away from zero for all $x \in D''$, so that the above N is well defined. Let $D_i = [u_i, u_{i+1}]$, $i = 1, ..., N$, and let x_i and $2d_i$ be the midpoint of the segment D_i and its length, respectively.

$4^0.3$. Note that

$$d_i \geq 5mh. \tag{64}$$

Indeed, by construction for a given i, $1 \leq i \leq N$, there are two possibilities:

a) $2d_i = \hat{\delta}/4$; in this case $d_i = \hat{\delta}/8 \geq 5mh$ due to (13);

b) $2d_i < \hat{\delta}/4$ and (62) is an equality. In this case, due to (63),

$$2d_i = \delta_+(u_i) = 10m(\eta_m^2\zeta)^\beta[\rho(u_i)]^{-2\beta/p} \geq 10m(\eta_m^2\zeta)^\beta(|D|L^p)^{-2\beta/p};$$

in view of (58) the concluding quantity here indeed is $\geq 10mh$.

$4^0.4$. For $1 \leq i \leq N$ let

$$\rho_i = \rho(u_i) = \int_{u_i}^{u_{i+1}} \sum_{j=1}^k |(q^j(d/ds)f^j)(s)|^p ds.$$

Note that

$$\sum_{i=1}^N \rho_i \leq |D|L^p. \tag{65}$$

$4^0.5$. For $1 \leq i < N$ let $X^i = X_h \cap [x_i, x_{i+1}]$. Note that by construction

$$D'_h \equiv X_h \cap D' \subset \bigcup_{i=1}^{N-1} X^i. \tag{66}$$

Let $d_i^* = \min\{d_i, d_{i+1}\}$, $1 \leq i < N$, and let for $x \in X^i$ the quantity $T_+(x)$ be defined as

$$T_+(x) = \lfloor \frac{1}{5mh} \max\{d_i^*; |x - u_{i+1}|\} \rfloor.$$

Note that $T_+(x)$ is positive integer due to (64).

Lemma 6 *Let $x \in X^i$, $1 \leq i < N$. Then*

$$f \in \mathbf{A}[k, l; p, R_i; [x - 5mhT_+(x), x + 5mhT_+(x)]] \tag{67}$$

with

$$R_i \leq 2m\eta_m\kappa h^{-l+1/p}[T_+(x)]^{-\frac{1}{2\beta}}; \tag{68}$$

besides this,

$$R_i \leq k^{1-1/p}[\rho_i + \rho_{i+1}]^{1/p}. \tag{69}$$

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

a) Let $D_x = [x - 5mhT_+(x), x + 5mhT_+(x)]$. We clearly have $f \in \mathbf{A}[k, l; p, R^x; D_x]$ with

$$R^x = \sum_{j=1}^{k} \left[\int_{D_x} |(q^j(d/ds)f^j)(s)|^p ds\right]^{1/p} \le$$

$$\le k^{1-1/p} \left[\int_{D_x} \left(\sum_{j=1}^{k} |(q^j(d/ds)f^j)(s)|\right)^p ds\right]^{1/p} = k^{1-1/p} \left[\int_{D_x} R^p(s) ds\right]^{1/p}. \qquad (70)$$

b) Consider the case of

$$d_i^* \ge |x - u_{i+1}|;$$

here $T_+(x) = \lfloor (5mh)^{-1} d_i^* \rfloor$, so that

$$\frac{1}{2} d_i^* \le 5mhT_+(x) \le d_i^*, \qquad (71)$$

whence, in particular,

$$D_x \subset [u_i, x_{i+1}]$$

and consequently

$$R^x \le R_i \equiv k^{1-1/p} [\rho_i + \rho_{i+1}]^{1/p};$$

this inequality combined with (70) proves (67) with the indicated R_i (which satisfies (69)). To prove (68), note that from (62) it follows that

$$\rho_i \le (10m)^{p/(2\beta)} (\eta_m^2 \zeta)^{p/2} (2d_i)^{-p/(2\beta)}, \ i = 1, ..., N,$$

whence

$$R_i = k^{1-1/p} [\rho_i + \rho_{i+1}]^{1/p} \le 2^{1/p} k^{1-1/p} (10m)^{1/(2\beta)} (\eta_m^2 \zeta)^{1/2} (2d_i^*)^{-1/(2\beta)} \le$$

[due to (71), $\beta = p/(2pl + p - 2)$ and $\zeta = \kappa^2 h$]

$$\le 2m\eta_m \kappa h^{-l+1/p} [T_+(x)]^{-1/(2\beta)},$$

as required in (68).

c) Now consider the case when $u_{i+1} - x > d_i^*$. Since $x_i \le x \le u_{i+1}$, we have $|x - u_{i+1}| \le d_i$, so that the case in question is possible only if $d_i > d_i^*$, i.e., when $d_i > d_{i+1}$. Here we have

$$\frac{1}{2}(u_{i+1} - x) \le 5mhT_+(x) \le u_{i+1} - x, \qquad (72)$$

and since $x_i \le x \le u_{i+1}$, we conclude that $D_x \subset D_i$. Consequently,

$$R^x \le R_i \equiv k^{1-1/p} \rho_i^{1/p}.$$

Taking into account (70), we see that the inclusion (67) is valid with the indicated R_i (which satisfies (69)). To prove (68), one should repeat with evident modifications the related part of b), with (72) playing the role of (71) and $u_{i+1} - x$ playing the role of d_i^*. ∎

$4^0.6$. Let $x = \tau h \in X^i$ for certain i, $1 \leq i < N$. By Lemma 6, we have $f \in \mathbf{A}[k,l;p,R_i;D_x]$ with R_i satisfying (68) and (69). Consequently, by Lemma 5, there exist decomposition $f^h = \sum_{j=1}^k f^{h,j}$ of the restriction f^h on the grid $\{th\}_{t\in\mathbf{Z}}$ and normalized polynomials ϕ^j of the degree l such that

$$\|\phi^j(\Delta)\Delta^{-\tau} f^{h,j}\|_{4mT_+(x),p} \leq \nu_l h^{l-1/p} R_i \leq 2m\eta_m\nu_l [T_+(x)]^{-1/(2\beta)}\kappa$$

(we have used (68)). Due to the definition of η_m and ν_l (see, respectively, (17) and (55)), the right hand side in the latter inequality is $\leq 4m^{-1/2}[T_+(x)]^{1/p-l-1/2}\kappa$; applying Theorem 1, we come to

$$|\hat{f}(x) - f(x)| \leq 4\theta_m \kappa [T_+(x)]^{-1/2} \tag{73}$$

(recall that we are considering the case of the event $\Theta \leq \kappa$).

5^0. Now let us estimate from above the quantity

$$\hat{\mathcal{R}}_{q,D'}(\hat{f}, f) = \left[h \sum_{x\in D'_h} |\hat{f}(x) - f(x)|^q\right]^{1/q}.$$

Let us start with the case of

$$q \geq 2.$$

We have in view of (66) and (73)

$$\hat{\mathcal{R}}^q_{q,D'} \leq [4\theta_m]^q \sum_{i=1}^{N-1} \mathcal{J}_i, \qquad \mathcal{J}_i = h \sum_{x\in X^i} \kappa^q [T_+(x)]^{-q/2}.$$

Let I' be the set of those $i \leq N-1$ for which both the segments D_i and D_{i+1} are of the length $\hat{\delta}/4$, and I'' be the set comprised of the remaining $i \leq N-1$.

If $i \in I'$ and $x \in X^i$, then, by construction, $T_+(x) \geq (40mh)^{-1}\hat{\delta}$, so that

$$\mathcal{J}_i \leq C_1^q (\kappa^2 h)^{q/2} \hat{\delta}^{1-q/2}$$

(from now on, C_i are positive quantities depending on m only). Besides this, the cardinality of I' clearly does not exceed $O(1)$, whence

$$\mathcal{I}_{1,q} \equiv \left[\sum_{i\in I'} \mathcal{J}_i\right]^{1/q} \leq C_2 \zeta^{1/2} |D|^{1/q-1/2}. \tag{74}$$

Now let $i \in I''$, and let

$$\hat{\rho}_i \equiv \max\{\rho_i; \rho_{i+1}\} \qquad [\leq \int_{D_i\cup D_{i+1}} R^p(s)ds]$$

Note that from construction of $T_+(x)$ it follows, by evident reasons, that

$$h \sum_{x\in X^i} T_+^{-q/2}(x) \leq C_3^q h^{q/2} (d_i^*)^{1-q/2}. \tag{75}$$

Since one of the segments D_i and D_{i+1} is of the length $< \hat{\delta}/4$, the smaller of the segments, let it be $D_{i'}$, also is of the length $< \hat{\delta}/4$; by construction of the segments D_i it is possible only if

$$2d_i^* = |D_{i'}| = 10m(\eta_m^2\zeta)^\beta \rho_{i'}^{-2\beta/p} \geq 10m(\eta_m^2\zeta)^\beta (\hat{\rho}_i)^{-2\beta/p},$$

whence, in view of (75) and $q \geq 2$,

$$h \sum_{x \in X^i} [T_+(x)]^{-q/2}(x) \leq C_4^q \zeta^{\beta(1-q/2)} (\hat{\rho}_i)^{\beta(q-2)/p} h^{q/2}.$$

Consequently,

$$\mathcal{I}_{2,q}^q \equiv \sum_{i \in I''} \mathcal{J}_i \leq C_4^q \zeta^{(q-\beta(q-2))/2} \sum_{i \in I''} (\hat{\rho}_i)^{\beta(q-2)/p}, \tag{76}$$

and in view of (65) we have

$$\sum_{i=1}^{N-1} \hat{\rho}_i \leq 2|D|L^p. \tag{77}$$

In the case of $q \geq p(2l+1)$ we have $\beta(q-2)/p \geq 1$, so that (76) – (77) imply

$$q \geq p(2l+1) \Rightarrow \mathcal{I}_{2,q}^q \leq C_4^q \zeta^{(q-\beta(q-2))/2} (|D|L^p)^{\beta(q-2)/p},$$

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

$$q \geq p(2l+1) \Rightarrow \mathcal{I}_{2,q} \leq C_5 \zeta^{\frac{l+1/q-1/p}{2l+1-2/p}} (|D|^{1/p}L)^{\frac{1-2/q}{2l+1-2/p}}. \tag{78}$$

Combining (74) and (78), we come to the estimate required in (15) 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

$$\mathcal{I}_{2,q} \leq |D|^{1/q-1/q'} \mathcal{I}_{2,q'}, \ \ 1 \leq q \leq q';$$

setting $q' = p(2l+1)$, we come to

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

combining this estimate with (74), we come to the result announced in (15) for the case of $q < p(2l+1)$.

6^0. We have established (15) for the case of $p < \infty$. To get the result for the case of $p = \infty$, it suffices to pass in (15) to limit as finite p tends to ∞. ■.

5.3 Proof of Theorem 3

From the definition of the classes $\mathbf{F}[\cdot]$ it is clear that these classes increase as d decreases, the remaining parameters of the class being fixed; therefore we without loss of generality can assume that (21) is equality rather than inequality:

$$d = [L^{-2}\zeta]^{1/(2l+1)} \quad [\zeta = h \ln \frac{1}{h}]. \tag{79}$$

Let us fix $f \in \mathbf{F}_d[k, l; p, L; D]$.

1^0. Let us choose $\delta \in [d/4, d/2]$ in such a way that the length $2\hat{\delta} - 2d$ of the segment D_d is an integer multiple of δ. Let us cover D_d by the segments $\Delta_1, ..., \Delta_N$ of the lengths 2δ each as follows: the first segment is centered at the left endpoint of D_d; each subsequent segment is centered at the right endpoint of its predecessor; the last segment is centered at the right endpoint of the segment D_d.

Let us fix $i \leq N$. For every $y \in \Delta_i$ the segment $D_f(y)$ centered at y of the length $2d(f) \geq 2d$ is such that $f \in \mathbf{A}[k, l; p, L_f(y)|D_f(y)|^{1/p}; D_f(y)]$. Let us choose $y_i \in \bar{\Delta}_i = \Delta_i \cap D_d$ in such a way that

$$L_f(y_i) \leq \left[|\bar{\Delta}_i|^{-1} \int_{\bar{\Delta}_i} L_f^p(y)dy\right]^{1/p}. \tag{80}$$

Let $D^i = D_f(y_i)$; since the midpoint of the segment D^i is in the segment Δ_i and the length $2d(f)$ of D^i is at least twice the length $2\delta \leq d$ of the segment Δ_i, we have $D^i \supset \Delta_i$. Let $I_i = \{n \leq N \mid \Delta_n \subset D^i\}$; since $\Delta_i \subset D^i$, we have $\cup_{i=1}^N I_i = I \equiv \{1, ..., N\}$.

Now, for $i \leq N$ let $f = \sum_{j=1}^k f^{i,j}$ be the decomposition of f on the segment D^i given by the inclusion $f \in \mathbf{A}[k, l; p, L_f(y_i)|D^i|^{1/p}; D^i]$, and let $q^{i,j}$ be the corresponding polynomials. Let us set

$$\rho_i(x) = \sum_{j=1}^k |(q^{i,j}(d/dx)f^{i,j})(x)|, \ \ x \in D^i,$$

and

$$L_{n,i} = (2\delta)^{-1/p}\left[\int_{\Delta_n} \rho_i^p(x)dx\right]^{1/p}, \ \ n \in I_i.$$

Let also

$$L_n = \min_{i:n \in I_i} L_{n,i}.$$

We claim that

$$f \in \mathbf{A}[k, l; p, k^{1-1/p} L_n |\Delta_n|^{1/p}; \Delta_n], \ \ n = 1, ..., N, \tag{81}$$

and that

$$\left[2\delta \sum_{n=1}^N L_n^p\right]^{1/p} \leq [40]^{1/p} |D|^{1/p} L. \tag{82}$$

The proof is as follows. Same as in the proof of (12), we have

$$f \in \mathbf{A}[k, l; p, k^{1-1/p} L_{n,i} |\Delta_n|^{1/p}; \Delta_n], \quad \forall i : n \in I_i,$$

and (81) follows.

To verify (82), assume first that $p < \infty$. We clearly have

$$\sum_{n \in I_i} L_{n,i}^p |\Delta_l| \leq \int_{D^i} \rho_i^p(x)dx \leq \left(\sum_{j=1}^k \left[\int_{D^i} |q^{i,j}(d/dx)f^{i,j}|^p dx\right]^{1/p}\right)^p \leq L_f^p(y_i)|D^i|,$$

whence

$$2\delta\sum_{n=1}^{N}\sum_{i:n\in I_i}L^p_{n,i}=2\delta\sum_{i=1}^{N}\sum_{n\in I_i}L^p_{n,i}\le 2d(f)\sum_{i=1}^{N}L^p_f(y_i). \qquad (83)$$

Now, every inner sum $S_n=\sum_{i:n\in I_i}L^p_{n,i}$ is at least $L^p_n k_n$, where k_n is the number of those i for which $n\in I_i$. Since the midpoint of D^i belongs to Δ_i and the length of D^i is $2d(f)\ge 2d\ge 4\delta$ and at the same time $d(f)\le|D|$, we have $k_n\ge 0.1d(f)/d$; thus, (83) implies that

$$\sum_{n=1}^{N}L^p_n\le 10\sum_{i=1}^{N}L^p_f(y_i).$$

The right hand side in the latter inequality, due to (80), is

$$10\sum_{i=1}^{N}|\bar{\Delta}_i|^{-1}\int_{\bar{\Delta}_i}L^p_f(y)dy\le 20\delta^{-1}\int_{D_d}L^p_f(y)dy\le 20\delta^{-1}|D|L^p$$

(we have taken into account that almost all points of D_d belong to at most 2 of the segments $\bar{\Delta}_i$), and (82) follows.

The case of $p=\infty$ can be obtained by the standard passing to limit.

2^0. Now we are ready to establish (22). It clearly suffices to verify (22) for the case of $q<\infty$; to get the result for the case of $q=\infty$, it suffices to pass in (22) to limit as $q<\infty$ tends to ∞.

Let us denote by Δ_i' the segment concentric to Δ_i and twice smaller than Δ_i. By construction, the segments Δ_i' form a covering of D_d. Now, for every $i\le N$ we have $f\in\mathbf{A}[k,l;p,L_i|\Delta_i|^{1/p};\Delta_i]$ (see (81)). We are about to apply Theorem 2 with $D=\Delta_i$ to bound from above the inaccuracy of our estimate on the segment Δ_i'; to this end let us verify that the theorem indeed is applicable, i.e., that relation (13) takes place with (with $\hat{delta}$ specified as the half-length δ of the segment Δ_i). Since $\delta\ge d/4$ and $h\le\zeta$ (see (19)), it suffices to verify that

$$\zeta\le\frac{d}{160m}\quad\left[\equiv\frac{1}{160m}[L^{-2}\zeta]^{1/(2l+1)},\ \text{see (79)}\right], \qquad (84)$$

or, which is the same, to demonstrate that

$$\zeta\le(160m)^{-1-1/(2l)}L^{-1/l}. \qquad (85)$$

(85) is an immediate consequence of (19). Indeed,

- since, due to (19), $\zeta\le(160m)^{-1-1/(2l)}$, (85) is valid in the case of $L\le 1$;
- to prove (85) in the case of $L>1$, note that, according to (19),

  $$\zeta\le(160m)^{-1-1/(2l)}L^{-\alpha},$$

  and $L^{-\alpha}\le L^{-1/l}$ due to $L>1$ and $\alpha\ge 1/l$ (see (20)).

3^0. Applying Theorem 2 with $D=\Delta_i$, we get

$$h\sum_{x\in\Delta_i'\cap X_h}|\hat f(x)-f(x)|^q\le C_1^q\left[\Omega_i\chi_{\{\Theta\le\kappa\}}+|\Delta_i|\kappa^q\Theta^q\chi_{\{\Theta>\kappa\}}\right],$$

where

$$\Omega_i=\delta^{1-q/2}\zeta^{q/2}+\delta^{q\sigma}L_i^{q\phi}\zeta^{q\psi}+\delta(\eta_m^{-1}L_i\delta^{1/p})^{q\epsilon}\zeta^{ql/(2l+1)} \tag{86}$$

and the constants $\eta_m,\sigma,\ \phi,\ \psi,\ \epsilon$ are given by Theorem 2. Taking sum over $i=1,...,N$ (recall that $N=|D_d|/\delta+1\le 2|D|/\delta$), we come to

$$\hat{\mathcal{R}}^q_{q,D_d}(\hat f,f)\le C_1^q\left[\Omega\chi_{\{\Theta\le\kappa\}}+|D|\kappa^q\Theta^q\chi_{\{\Theta>\kappa\}}\right], \tag{87}$$

where, due to (86) and relations $N\le 2|D|/\delta$ and $d/2\ge\delta\ge d/4$,

$$\Omega=\sum_{i=1}^N\Omega_i\le C_2^q\left[\mathcal{J}_1+\mathcal{J}_2+\mathcal{J}_3\right], \tag{88}$$

with

$$\mathcal{J}_1=|D|(\zeta/d)^{q/2},$$

$$\mathcal{J}_2=(d^\sigma\zeta^\psi)^qS_{q\phi},\qquad S_r\equiv\sum_{i=1}^N L_i^r, \tag{89}$$

and

$$\mathcal{J}_3=\delta^{1+q\epsilon/p}\zeta^{ql/(2l+1)}C_3^{q\epsilon}S_{q\epsilon}. \tag{90}$$

3^0.1. Due to (79) one has

$$\mathcal{J}_1\le|D|L^{q/(2l+1)}\zeta^{ql/(2l+1)}. \tag{91}$$

It follows from (82) that

$$S_r\le C_4^r\begin{cases}\left[|D|d^{-1}L^p\right]^{r/p}, & r\ge p\\ |D|d^{-1}L^r, & 0\le r\le p\end{cases} \tag{92}$$

3^0.2. To estimate $\mathcal{J}_2$, consider two possible cases:

1) $q\ge p(2l+1)$. In this case $q\phi\ge p$, and (92) implies that

$$S_{q\phi}\le C_4^{q\phi}\left[|D|d^{-1}\right]^{q\phi/p}L^{q\phi}=C_4^{q\phi}\left[|D|d^{-1}\right]^{q\sigma}L^{q\phi}$$

(note that in the case in question $\sigma=\phi/p$). Taking into account (89) and evident relation $\phi\le C_5$, we obtain

$$\mathcal{J}_2\le C_6^q\left[|D|^\sigma L^\phi\zeta^\psi\right]^q. \tag{93}$$

2) $q\le p(2l+1)$. In this case $q\phi\le p$, $\sigma=1/q$, and (89) and (92) again lead to (93).

3^0.3. Let us verify that in fact

$$\mathcal{J}_1+\mathcal{J}_2\le C_7^q\left[|D|^\sigma L^\phi\zeta^\psi\right]^q\equiv C_7^q\mathcal{P}^q. \tag{94}$$

To this end, in view of (91) and (93), it suffices to verify that

$$|D|^{1/q}L^{1/(2l+1)}\zeta^{l/(2l+1)} \le |D|^{\sigma}L^{\phi}\zeta^{\psi}. \tag{95}$$

This relation is evident in the case of $q \le p(2l+1)$, since here $\sigma = 1/q$, $\phi = 1/(2l+1)$, $\psi = l/(2l+1)$. Now consider the case of $q > p(2l+1)$. (95) is equivalent to

$$\zeta^{l/(2l+1)-\psi} \le |D|^{\sigma-1/q}L^{\phi-1/(2l+1)}. \tag{96}$$

Due to the origin of σ, ϕ, ψ, the left hand side in the latter inequality is

$$\zeta^{\beta/\gamma},\ \ \beta = q - p(2l+1),\ \ \gamma = (2l+1)q(2pl+p-2),$$

while the right hand side is

$$|D|^{(2l+1)\beta/\gamma}L^{2\beta/\gamma};$$

in view of these identities, (96) is an immediate consequence of (19).

4^0. We are about to prove that $\mathcal{J}_3$ is majorated, within appropriate factor, by $\mathcal{P}$ (see (94)). To this end let us set

$$\mathcal{F} = [\mathcal{J}_3/\mathcal{P}]^{1/q}$$

and consider two possible cases:

- I. $q\epsilon \ge p$. In this case (94), (79), (92) and (90) result in

$$\mathcal{F} \le C_8^{\epsilon}L^P|D|^Q\zeta^R, \tag{97}$$

  with

$$P = \epsilon - \frac{2}{q(2l+1)} - \phi, \quad Q = \frac{\epsilon}{p} - \sigma, \quad R = \frac{l}{2l+1} + \frac{1}{q(2l+1)} - \psi. \tag{98}$$

- II. $q\epsilon < p$. In this case (94), (79), (92) and (90) result in

$$\mathcal{F} \le C_9^{\epsilon}L^P|D|^Q\zeta^R, \tag{99}$$

  with

$$P = \epsilon\left[1 - \frac{2}{p(2l+1)}\right] - \phi, \quad Q = \frac{1}{q} - \sigma, \quad R = \frac{lp+\epsilon}{p(2l+1)} - \psi. \tag{100}$$

4^0.I. *Case I.* First note that from (17) it is clear that $\epsilon \ge 1$. With this observation, (98) immediately implies that

$$P \ge 0; Q \ge 0; R \ge 0.$$

Since $|D| \le 1$ and $Q \ge 0$, we have

$$|D|^R \le 1; \tag{101}$$

further, from (19) it follows that $L \leq \zeta^{-1/\alpha}$, and since $P \geq 0$, we conclude that

$$L^P \zeta^R \leq \zeta^{S/\alpha}, \quad S = \alpha R - P. \tag{102}$$

We are about to demonstrate that in the case in question $S \geq 0$; note that this inequality combined with (102), (101) and relation $\zeta \leq 1$ (see (19)) would imply that

$$\mathcal{F} \leq C_8. \tag{103}$$

To verify that $S \geq 0$, note that from (98) it is clear that R and $-P$ are nonincreasing in q when q varies over the segment $[p, p(2l+1)]$, so that S also is nonincreasing in this segment. At the right endpoint of the segment $S = 0$, as it is immediately seen from the origin of α (note that the second case in (20) is possible only if $l = 1$). Thus, $S \geq 0$ when $q \in [p, p(2l+1)]$. Since we are in the case I, the only remaining allowed values of q are those $> p(2l+1)$. When $q > p(2l+1)$, from (98) it follows that R and $-P$ are nondecreasing functions of q, so that S grows with q on the ray $[p(2l+1), \infty)$. As we already have mentioned, at the left endpoint of the ray $S = 0$, so that S indeed is ≥ 0 when $q \geq p$; we have established (103) for the case I.

4^0.II. *Case II.* Let us verify that in this case (103) also is valid. Indeed, as it was already mentioned, $\epsilon \geq 1$, whence in the case II $q \leq p$, and consequently (see the definitions of σ, ϕ, ψ and (100))

$$P = \epsilon \left[1 - \frac{2}{p(2l+1)}\right] - \frac{1}{2l+1} \geq 0, \;\; Q = 0, \;\; R = \frac{\epsilon}{p(2l+1)} \geq 0.$$

Same as above, nonnegativity of P, Q, R implies that to establish (103) it suffices to verify that

$$S \equiv \alpha R - P \geq 0,$$

which is given by immediate calculation.

5^0. Combining (94) and (103), we get

$$\mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 \leq C_9^{q\epsilon} \left[|D|^\sigma L^\phi \zeta^\psi\right]^q;$$

$$\mathcal{J}_3 \leq C_5^{q\epsilon} d^{q\epsilon/p} |D| L^{q\epsilon} h^{q/2} \leq C_5^{q\epsilon} |D| (|D|^{1/p} L)^{q\epsilon} h^{q/2}.$$

Combining this inequality, (88), (87) and (18), we come to (22). ■

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 and I. Johnstone. Minimax estimation via wavelet shrinkage. Technical report, Department of Statistics, Stanford University, 1992.

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

[5] I. Ibragimov and R. Khasminskii. On nonparametric estimation of regression. *Soviet Math. Dokl.*, 21:810–814, 1980.

[6] I. Ibragimov and R. Khasminskii. *Statistical Estimation.* Springer-Verlag, New York, 1981.

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

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

[9] 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.

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

[11] A. Nemirovskii. On nonparametric estimation of functions satisfying differential inequalities. *Advances in Soviet Mathematics*, 12:7–43, 1992.

[12] Ch. Stone. Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, 8(3):1348–1360, 1980.