Denoising Signals of Unknown Local Structure

Anatoli Juditsky [a,*], Arkadi Nemirovski [b]

[a] *LMC, B.P. 53, 38041 Grenoble Cedex 9, France*

[b] *School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive Atlanta GA 30332-0205 USA*

1 Introduction

In this paper, we focus on the *nonparametric regression problem* as follows: *Given noisy observations*

$$y_t = f(x_t) + \sigma e_t, \quad t = (t_1, ..., t_d) \in \mathbf{Z}^d,\ 0 \leq t_j \leq m \tag{1}$$

of a "signal" $f : [0,1]^d \to \mathbb{C}$ taken along the equidistant grid $\Gamma_n = \{x_t = m^{-1}t : 0 \leq t_1, t_2, ..., t_d \leq m\}$ on the unit d-dimensional cube $[0,1]^d$ with $n = (m+1)^d$ observation points, we are interested to recover f on the observation grid.

In (1), $\{e_t\}$ are independent of each other standard complex-valued Gaussian noises; the adjective "standard" means that $\Re(e_t)$ and $\Im(e_t)$ are independent of each other $\mathbf{N}(0,1)$ random variables.

The problem we are interested in is investigated in a huge number of studies. The traditional setting of the problem, which forms the frame of a vast majority of these studies, is as follows:

A. We intend to recover the signal both *on* and *outside* the observation grid and measure the risk of recovering f on a cube $B \subset [0,1]^d$ by the standard integral L_q-norms;

B. The estimation routines are aimed at recovering *smooth* signals, and their quality is measured by their maximal risks, the maximum being taken over f running through natural families of smooth signals, e.g., Hölder and Sobolev balls;

C. The focus is on the asymptotic, as the volume of observations n goes to

* Corresponding author.
Email addresses: `Anatoli.Juditsky@inrialpes.fr` (Anatoli Juditsky), `nemirovs@ie.technion.ac.il` (Arkadi Nemirovski).

infinity, behavior of the estimation routines, with emphasis on *asymptotically minimax (nearly) optimal estimates* – those with the (nearly) best possible rate of convergence of the risks to 0 as $n \to \infty$, the risks being taken on the classes mentioned in B.

Initially, the research was focused on recovering smooth signals with *a priori known smoothness parameters* and the estimation routines were tuned to these parameters (see, e.g., [31,43,48,32,4,45,40,49,30,46,29,34]). Later on, there was a significant research on *adaptive estimates* [44,11,16,17,38,33,5,12,27]. Adaptive estimation routines are free of a priori assumptions on the smoothness parameters of the signal to be recovered, and the emphasis is on developing routines which exhibit asymptotically optimal behavior on a wide variety of families of smooth signals. For a comprehensive overview of results on smooth nonparametric regression estimation, see [42] [3] .

The traditional focus on recovering smooth signals ultimately comes from the fact that such a signal *locally* can be well-approximated by a polynomial of a fixed order ν, and such a polynomial is an "easy to estimate" entity. Specifically, for every integer $T \geq 0$, the value of a polynomial p at an observation point x_t can be recovered via $(2T+1)^d$ neighboring observations $\{x_\tau : |\tau_j - t_j| \leq T, 1 \leq j \leq d\}$ "at a parametric rate" – with the expected squared error $C\sigma^2(2T+1)^{-d}$ inverse proportional to the amount $(2T+1)^d$ of the observations used by the estimate; the coefficient C depends solely on the order and the dimensionality d of the polynomial. The corresponding estimate $\widehat{p}^T(x_t)$ of $p(x_t)$ is pretty simple: it is given by a "time-invariant filter", that is, by convolution of observations with an appropriate discrete kernel $q^{(T)} = (q_\tau^{(T)})_{\tau \in \mathbf{Z}^d}$ vanishing outside the box $\mathbf{O}_T = \{\tau \in \mathbf{Z}^d : |\tau_j| \leq T, 1 \leq j \leq d\}$:

$$\widehat{p}(x_t) = \sum_{\tau \in \mathbf{O}_T} q_\tau^{(T)} y_{t-\tau}.$$

On the other hand, one can observe that the space of algebraic polynomials $p(x) = \sum\limits_{0 \leq \tau_j \leq \nu} c_\tau x_1^{\tau_1} ... x_d^{\tau_d}$ of a given order ν is not the only "easy to estimate" parametric family. For example, replacing algebraic polynomials with *exponential polynomials* $p(x) = \sum\limits_{\substack{\nu=(\nu_1,...\nu_d)\\ 0 \leq \nu_1,...,\nu_d \leq m}} c_\nu x_1^{\nu_1} x_2^{\nu_2} ... x_d^{\nu_d} \exp\{i\omega^T(\nu)x\}$ with fixed "frequencies" $\omega(\nu) \in \mathbf{C}^d$, we still have a possibility to recover $p(x_t)$ at a parametric rate from noisy observations of $p(\cdot)$ taken at $(2T+1)^d$ neighboring to x_t observation points x_τ, $\tau \in t - \mathbf{O}_T$, and the recovering routine is still given by a time-invariant filter. Consequently, *in principle* we have essentially the same possibilities to recover smooth signals $f(x)$ – those which locally can be well-

[3] Our "super-short outline" of Nonparametric Regression would be severely incomplete without mentioning a novel approach aimed at recovering *nonsmooth* signals possessing *sparse representations* in properly constructed functional systems [10,12,13,18–24,26,9,47,25,15]. This promising approach is completely beyond the scope of our paper.

approximated by specific "easy to estimate" signals (algebraic polynomials), as the possibilities to recover "modulated signals" $f(x)\cos(\omega^T x)$ with smooth f and $\omega \in \mathbf{R}^d$. Indeed, a modulated signal locally can be well-approximated by an exponential polynomial, and the latter is as easy to estimate as an algebraic polynomial of the same order. At the same time, the family of modulated signals is much wider than the family of smooth signals and contains "highly oscillating" signals which, from the traditional viewpoint, are difficult to recover.

In light of these observations, where the unique role played in the traditional non-parametric regression by smooth signals comes from? The answer is as follows: what is crucial in the traditional context is not the *existence* of time-invariant filters which recover algebraic polynomials at a parametric rate, but rather the fact that *we know these filters in advance* (they are readily given by the order of polynomial, d and T). By contrast, in most of the cases where it would be natural to work with "modulated" signals, the corresponding frequencies are not known in advance. Since time-invariant filters which recover well exponential polynomials depend on the corresponding frequencies, the nonparametric recovering routines based on these filters are essentially meaningless – in order to use them, one should first identify the frequencies, which by itself is a pretty difficult problem. The main goal of this paper is to demonstrate that, in a sense, *there is no necessity to solve this difficult problem at all*. Essentially, we show that
(i) whenever a discrete time signal (that is, a signal defined on a regular discrete grid) is *well-filtered*, i.e., can be recovered from its noisy observations *at a parametric rate* by a *linear time-invariant filter*, we can recover this signal at a "nearly parametric" rate *without a priori knowledge of the associated filter*;
(ii) whenever a "continuous time" signal $f : [0, 1]^d \to \mathbf{C}$, restricted onto the observation grid, can be locally well-approximated by a well-filtered signal, f can be recovered *on the observation grid* basically as well as if the well-filtered approximation were known to be an algebraic polynomial of a given order (i.e., as well as if f were smooth with known in advance smoothness parameters). In particular, we demonstrate that a modulated signal $f(x)\cos(\omega^T x)$ with smooth f and *unknown in advance* (and perhaps high) "frequency" $\omega \in \mathbf{R}^d$ can be recovered *on the observation grid* with basically the same quality as the smooth signal f itself.
In connection with (ii), it should be stressed that in our setting, in contrast to the traditional one aimed at recovering smooth signals, *all we are interested in is to recover f on the observation grid only* (and therefore we measure risks by discrete analogues of the integral L_q-norms). This is quite natural, since we intend to handle, along with others, also highly oscillating signals, and such a signal, in general, cannot be recovered well outside the observation grid. Indeed, a highly oscillating signal can merely vanish on the observation grid and be arbitrarily large outside it (which is impossible for smooth signals).
The rest of our paper is organized as follows. In Sections 3, 4 we focus on item

(i). Specifically, in Section 3 we give a formal definition of a well-filtered signal on a d-dimensional regular grid (the latter, w.l.o.g., is normalized to be $\mathbf{Z}^d$) and demonstrate that such a signal can be recovered at a nearly parametric rate *without a priori knowledge of the corresponding "good filter"* (Theorem 2). The estimation routine underlying Theorem 2 ("Algorithm A", Section 3.2) is a substantial extension of the procedure proposed in [39]. In Section 3.4, we demonstrate that the family of well-filtered signals is pretty wide – it contains a wide spectrum of "basic functions" (for example, exponential polynomials) and is closed with respect to a number of basic operations, including modulation, taking linear combinations and tensor products. In Section 4 we present the "prediction" versions of the results of Section 3; now we are interested in recovering of a discrete time signal at a point $t \in \mathbf{Z}^d$ via noisy observations taken at the points $\{\tau \in \mathbf{Z}^d : t_j - T \leq \tau_j \leq t_j - \kappa\}$ "preceding" the point t, with a given in advance "forecast horizon" $\kappa \geq 0$. In Section 5 we develop *adaptive* versions of estimates given by Algorithms A, B. Finally, in Section 6 we focus on item (ii) – recovering continuous time signals which can be locally well-approximated by well-filtered signals. Our main results here are stated in Theorem 22. This theorem extends onto wide classes of *locally well-filtered signals* the results of [44,11,16,17,38,33,27] on spatial adaptive estimates of smooth signals and can be treated as a substantial generalization of the results of [41,28] on estimating univariate signals satisfying differential inequalities with unknown in advance linear differential operators.
To make the exposition more readable, all proofs are collected in the appendix.

2 Preliminaries

Fields over $\mathbf{Z}^d$. Let $C(\mathbb{Z}^d)$ be the linear space of complex-valued fields $r = \{r_\tau : \tau \in \mathbf{Z}^d\}$ over $\mathbf{Z}^d$.

• Given nonnegative integer T and $p \in [1, \infty]$, we define semi-norms $|\cdot|_{T,p}$ on $C(\mathbb{Z}^d)$ by $|r|_{T,p} = \left(\sum_{|\tau| \leq T} |r_\tau|^p\right)^{1/p}$, $|\tau| = \max\{|\tau_1|, ..., |\tau_d|\}$, with the standard interpretation of the right hand side when $p = \infty$, and we set $|r|_p = \lim_{T \to \infty} |r|_{T,p} \in \mathbb{R} \cup \{+\infty\}$. A field $r \in C(\mathbb{Z}^d)$ with finitely many nonzero entries r_τ is called *a filter*, and the smallest T such that $r_\tau = 0$ whenever $|\tau| > T$, is called the *order* $\mathrm{ord}(r)$ of a filter r; we write $C_T(\mathbf{Z}^d) = \{r \in C(\mathbb{Z}^d) \mid \mathrm{ord}(r) \leq T\}$. We identify a filter r with the multivariate Laurent sum $r(z_1, ..., z_d) = \sum_\tau r_\tau z_1^{\tau_1} ... z_d^{\tau_d}$.

• We call a filter r *polynomial*, if the corresponding Laurent sum is a polynomial (i.e., if the only entries r_τ which can be nonzero are those with $\tau \geq 0$). The set of all polynomials is denoted $P(\mathbf{Z}^d)$. For integers k, T, $0 \leq k \leq T$, we denote by $P_T^k(\mathbf{Z}^d)$ the subspace of $P(\mathbf{Z}^d)$ formed by polynomials r for which the only nonzero entries r_τ can be those for which $k \leq \tau_j \leq T$, $j = 1, ..., d$.

• We denote by Δ_j, $j = 1, ..., d$, the "basic shift operators" on $C(\mathbb{Z}^d)$: $(\Delta_j r)_{\tau_1,...,\tau_d} = r_{\tau_1,...,\tau_{j-1},\tau_j-1,\tau_{j+1},...,\tau_d}$.
• Finally, we define the output of a filter r, the input to the filter being a field $x \in C(\mathbf{Z}^d)$, as the field $r(\Delta)x \equiv r(\Delta_1, \Delta_2, ..., \Delta_d)x$, so that $(r(\Delta)x)_t = \sum_\tau r_\tau x_{t-\tau}$.
Fourier transform. Let T be a nonnegative integer, let Γ_T be the set of roots of 1 of the degree $2T+1$, and let $C(\Gamma_T^d)$ be the space of complex-valued functions on $\Gamma_T^d \equiv (\Gamma_T)^d$.
• We define the Fourier transform $F_T : C(\mathbb{Z}^d) \to C(\Gamma_T^d)$ as $(F_T r)(\mu) = \frac{1}{(2T+1)^{d/2}} \sum_{|\tau|\leq T} r_\tau \mu_1^{\tau_1}...\mu_d^{\tau_d} \equiv \frac{1}{(2T+1)^{d/2}} r(\mu), r \in C_T(\mathbf{Z}^d)$, where $\mu \in \Gamma_T^d$. Note that $r_\tau = \frac{1}{(2T+1)^{d/2}} \sum_{\mu\in\Gamma_T^d} (F_T r)(\mu)\mu_1^{-\tau_1}...\mu_d^{-\tau_d}$, $\forall(\tau : |\tau| \leq T)$. The Fourier transform allows to equip $C(\mathbb{Z}^d)$ with semi-norms coming from the standard p-norms on $C(\Gamma_T^d)$:

$$|r|_{T,p}^* = |F_T r|_p \equiv \left(\sum_{\mu\in\Gamma_T^d} |(F_T r)(\mu)|^p\right)^{1/p},$$

with the standard interpretation of the right hand side for $p = \infty$.

3 Main result

Let $\mathbf{F} = (\Omega, \Sigma, P)$ be a probability space. We consider the problem of recovering unknown random field $(s_\tau = s_\tau(\xi))_{\substack{\tau\in\mathbf{Z}^d \\ \xi\in\Omega}}$ over $\mathbf{Z}^d$ from noisy observations

$$y_\tau = s_\tau + e_\tau. \tag{2}$$

It is convenient for us to assume that both the *signal* (s_τ) and the noises are complex-valued. Besides this, we assume that the field (e_τ) of observation noises is independent of (s_τ) and is of the form $e_\tau = \sigma\epsilon_\tau$, where (ϵ_τ) are independent of each other *standard* Gaussian complex-valued variables; the adjective "standard" means that $\Re(\epsilon_\tau)$, $\Im(\epsilon_\tau)$ are independent of each other $\mathbf{N}(0, 1)$ random variables.
Posed informally, the question we are interested in is as follows:
(?) *We know that the random field* $(s_\tau)_{\tau\in\mathbf{Z}^d} \equiv (s_\tau(\xi))_{\substack{\tau\in\mathbf{Z}^d \\ \xi\in\Omega}}$ *underlying observations* (2) *can be recovered from these observations "at a parametric rate" by "linear time-invariant filtering": for a given* T*, there exists (unknown in advance) filter* $q \in C_T(\mathbf{Z}^d)$ *(i.e., a filter which recovers* s_τ *via* $O(T^d)$ *observations around the point* τ*) such that*

$$E\left\{|s_\tau - (q(\Delta)y)_\tau|^2\right\} \leq O(\sigma^2 T^{-d}). \tag{3}$$

Can we mimic this filter?
To make this question precise, we should specify our a priori knowledge of the constant factor hidden in $O(\cdot)$ and on the ranges on values of T and τ where (3) holds true.

3.1 Well-filtered signals

Since the observation noises are independent of (s_τ), we have

$$E\left\{|s_\tau - (q(\Delta)y)_\tau|^2\right\} = 2\sigma^2|q|_2^2 + E_\xi\left\{|s_\tau(\xi) - (q(\Delta)s(\xi))_\tau|^2\right\}; \tag{4}$$

therefore in order to ensure (3), both terms in the right hand side of the latter inequality should be of order of T^{-d}. This observation motivates the following

Definition 1 *Let $\theta \geq 0$, $\rho \geq 1$ be reals, let L be a nonnegative integer or $+\infty$, and let $t \in \mathbf{Z}^d$. Finally, let $(s_\tau)_{\tau\in\mathbf{Z}^d} \equiv (s_\tau(\xi))_{\substack{\tau\in\mathbf{Z}^d\\ \xi\in\Omega}}$ be a random field on $\mathbf{Z}^d$.*
(1) [T-well-filtered signals] *Let T be a nonnegative integer. We say that (s_τ) is T-well-filtered, with the parameters θ, ρ, L, at the point t* (notation: $(s_\tau) \in \mathbf{S}_L^t(\theta,\rho,T)$)*, if there exists a filter $q = q^{(T)} \in C_T(\mathbf{Z}^d)$, $|q|_2 \leq \frac{\rho}{(2T+1)^{d/2}}$, which reproduces (s_τ) in the box $\{\tau : |\tau - t| \leq L\}$ with the mean square error not exceeding $\theta(2T+1)^{-d/2}$:*

$$\max_{\tau:|\tau-t|\leq L}\left[E\left\{|s_\tau - (q(\Delta)s)_\tau|^2\right\}\right]^{1/2} \leq \theta(2T+1)^{-d/2}. \tag{5}$$

(2) [well-filtered signals] *We say that (s_τ) is well-filtered, with the parameters θ, ρ, L, at the point t* (we use the notation: $(s_\tau) \in \mathbf{F}_L^t(\theta,\rho)$)*, if, for every integer T, $0 \leq T \leq L$, (s_τ) is T-well-filtered, with the parameters θ, ρ, L, at t.*

In the sequel, we refer to filters $q^{(T)}$ associated, in the sense of the above definition, with a well-filtered signal (s_τ) as to filters *certifying* the "well-filterability" of the signal.
We are about to demonstrate that with the interpretation of **(?)** suggested by Definition 1, the answer to the question is affirmative: for a T-well-filtered, with parameters $\theta, \rho, L = 3T$, at a point t signal can be recovered at this point "at a nearly parametric rate" with *no a priory knowledge of the corresponding "good filter"*; all we should know in advance is ρ and T. We start with presenting the recovering routine.

3.2 The estimator

The estimator we intend to use is as follows:
Algorithm A: *Given a setup* $(\rho \geq 1, T)$ *and a point* $t \in \mathbf{Z}^d$*, we build an estimation* $\widehat{s}_t[T,y]$ *of* s_t *via observations* (y_τ), $|\tau - t| \leq 4T$*, as follows:*
(1) When $T = 0$*, we merely set* $\widehat{s}_t[0,y] = y_t$
(2) When $T > 0$*, we set* $\widehat{s}_t[T,y] = (\widehat{\phi}^t(\Delta)y)_t$*, where* $\widehat{\phi}^t \in C_{2T}(\mathbf{Z}^2)$ *is an optimal solution to the following optimization problem:*

$$\min_{\phi \in C_{2T}(\mathbf{Z}^d)} \Big\{ \underbrace{|\Delta_1^{-t_1} ... \Delta_d^{-t_d}(1 - \phi(\Delta))y|^*_{2T,\infty}}_{J(\phi, y^t_{4T})} : |\phi|^*_{2T,1} \leq 2^{d/2}\rho^2(2T+1)^{-d/2} \Big\}, \quad (6)$$

where $y^t_L = \{y_\tau : |t - \tau| \leq L\}$.

Note that the objective in (6) is affected only by observations y^t_{4T}, so that our algorithm recovers s_t via $(8T+1)^d$ observations "around" the point t.

3.3 Main result: de-noising

Theorem 2 *Assume that the signal* (s_τ) *underlying observations* (2) *is* T*-well-filtered, with parameters* θ, ρ, $L \geq 3T$*:* $(s_\tau) \in \mathbf{S}^t_L(\theta, \rho, T)$ *with* $L \geq 3T$*. Then the mean square error of the estimate* $\widehat{s}_t[T,\cdot]$ *of* s_t *yielded by Algorithm A with setup* (ρ, T) *can be bounded from above as follows:*

$$\begin{array}{c} \left(E\left\{|\widehat{s}_t[T,y] - s_t|^2\right\}\right)^{1/2} \leq c(d)\rho^3 \dfrac{\theta + \sigma\rho\sqrt{\ln(2T+1)+1}}{(2T+1)^{d/2}}, \\ c(d) = 3(2^d + 2^{3d-1}). \end{array} \quad (7)$$

In particular, if (s_τ) *is well-filtered, with the parameters* θ, ρ, L*, at a point* t*, then for every integer* T, $0 \leq T \leq \lfloor L/3 \rfloor$*, the accuracy of the estimate* $\widehat{s}_t[T,y]$ *of* s_t *yielded by Algorithm A can be bounded by* (7)*. Finally, in the case of deterministic* (s)*, we have*

$$\begin{array}{l} |s_t - \widehat{s}_t[T,y]| \leq c(d)\rho^3 \left[\theta + \sigma\rho\Theta^t_T\right](2T+1)^{-d/2}, \\ \Theta^t_T = \sigma^{-1} \max\limits_{\tau:|\tau|\leq 2T} |\Delta_1^{\tau_1 - t_1} ... \Delta_d^{\tau_d - t_d} e|^*_{2T,\infty}. \end{array} \quad (8)$$

Comments: note that Theorem 2 indeed gives an affirmative answer to the question **(?)**. Indeed, let a signal (s_τ) admit, for some T, a filter-type estimate

$\bar{s}_\tau = (q^*(\Delta)y)_\tau$ with "window width" T (i.e., with $q^* \in C_T(\mathbf{Z}^d)$) and with the mean square error which, in an $O(T)$-neighborhood of a point t, is of the "parametric" order $O\left(\sigma(2T+1)^{-d/2}\right)$:

$$\max_{\tau:|\tau-t|\leq 3T} E\left\{|s_\tau - \bar{s}_\tau|^2\right\} \leq \kappa^2 \equiv \frac{\sigma^2\mu^2}{(2T+1)^{d/2}} \tag{9}$$

with some known $\mu \geq 1$. We do not know what is this estimate, although do know that it exists (i.e., know the associated T, μ), and we want to recover s_t from observations y_{4T}^t nearly as well as if we were using our hypothetic estimate $\bar{s}_t$. Theorem 2 says that Algorithm A basically achieves this goal. Indeed, from (4), (9) it follows that $|q^*|_2 \leq \frac{\mu}{(2T+1)^{d/2}}$ and $(s_\tau) \in \mathbf{S}_{3T}^t(\sigma\mu, \mu, T)$. Applying Theorem 2 with $\rho = \mu$, $\theta = \sigma\mu$, $L = 3T$, we conclude that with the estimate yielded by Algorithm A, the mean square error of recovering s_t does not exceed $O(1)\mu^3\left[1+\sqrt{\ln(2T+1)}\right]\kappa$. We see that *as far as the dependence on "observation time" T^d is concerned*, the estimate yielded by Algorithm A is just by a logarithmic in T factor worse than the estimate $\bar{s}_t$ we wish to mimic. In the literature on nonparametric estimation the bounds as in Theorem 2 are often referred to as *oracle inequalities*. Since the pioneering work [1] a number of oracle inequalities have been established for a wide variety of estimation problems (cf. the papers [35], [2], [6], [7], [11], [14], [8] among many others). In that context one refer to the filter q, which certifies the well-filterability of the signal, as the oracle, and the bound (7) describes the ability of a particular adaptive method (Algorithm A above) to reproduce the oracle. We complete this upper bound with the following result:

Theorem 3 *For any $m \in \mathbb{N}^+$, positive σ and $T \in \mathbb{N}$ large enough, one can point out a family $\mathbf{F}_m^T$ of real signals on $[-4T-1, 4T+2]$ such that*
- *for each signal $s \in \mathbf{F}_m^T$ there exists a filter $q^* \in C_T(\mathbf{Z})$ with $|q^*|_2 = \frac{\rho}{\sqrt{2T+1}}$, $\rho = \sqrt{m}$, such that* $\max_{-3T-1\leq\tau\leq 3T+2}\left(E|(q^*(\Delta)y)_- s_\tau|^2\right)^{1/2} = \frac{\theta}{\sqrt{2T+1}}$ *with $\theta = \sigma\rho$.*
- *There is $c_0 > 0$ such that for any estimate $\widehat{s}$ of s_0 from the observations (2) it holds*

$$\inf_{\widehat{s}} \sup_{s\in\mathbf{F}_m^T} \left(E|\widehat{s} - s_0|^2\right)^{1/2} \geq c_0\theta\rho\sqrt{\ln(2T+1)}. \tag{10}$$

The lower bound (10) states that the factor $\rho\sqrt{\ln(2T+1)}$ is an unavoidable "price" for adaptation. When comparing the result of Theorem 2 to that of Theorem 3, we observe an extra factor $\rho^2 \geq 1$ in the corresponding upper bound (7). By now we do not know if this extra factor can be completely eliminated. Nevertheless, in light of these results, we can claim that recovering of signals with certifying filter of large l_2-norm is a rather desperate task

– the price for adaptation is then proportional to $\rho \gg 1$ in this case.
When applying Algorithm A and Theorem 2, the crucial question is how to recognize well-filterability. We are about to give a partial answer to this question.

3.4 Calculus of well-filtered signals

Our current goal is to understand how wide is the family of well-filtered signals, and our plan is as follows: (a) we list a number of operations which preserve the property of well-filterability, and (b) we present a list of examples of well-filtered signals. Applying to "raw materials" from (b) operations from (a), one can produce a wide variety of well-filtered signals.

3.4.1 Operations preserving well-filterability

I. "Scale" of well-filtered signals. We start with the following evident observation: $\rho' \geq \rho, \theta' \geq \theta, L' \leq L \Rightarrow \mathbf{F}^t_L(\theta,\rho) \subset \mathbf{F}^t_{L'}(\theta',\rho')$.
II. Taking linear combinations. Our next observation is that a linear combination of well-filtered signals is again well-filtered, with properly updated parameters:

Proposition 4 *Let $(s^j_\tau) \in \mathbf{F}^t_L(\theta_j,\rho_j)$, and let $\lambda^j \in \mathbb{C}$ be random variables independent of (s^j) and such that $E\{|\lambda_j|^2\} < \infty$, $j = 1,...,m$. Then*

$$\begin{array}{l} (s_\tau \equiv \sum\limits_{j=1}^m \lambda_j s^j_\tau) \in \mathbf{F}^t_{L^+}(\theta^+,\rho^+), \\ \theta^+ = (2m-1)^{d/2} 2^{m-1}\rho_1 ... \rho_m \sum\limits_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j}, \\ \rho^+ = (2m-1)^{d/2} 2^m \rho_1 ... \rho_m, L^+ = \lfloor L/2 \rfloor. \end{array} \tag{11}$$

In the case of $m = 1$, one can set $\rho^+ = \rho_1$, $\theta^+ = |\lambda_1|\theta_1$, $L^+ = L$.
The filters certifying the well-filterability of (s_τ) can be chosen to be independent of the coefficients λ_j.

III. Modulation and conjugation. Our next observation is that the families of well-filtered signals are closed w.r.t. "modulation" and conjugation:

Proposition 5 *Let $(s_\tau) \in \mathbf{F}^t_L(\theta,\rho)$.*
(i) Let $\omega \in \mathbb{R}^d$, $\phi \in \mathbb{R}$ be deterministic. Then the signal $(\widehat{s}_\tau = \exp\{i[\omega^T\tau + \phi]\} s_\tau)_{\tau\in\mathbf{Z}^d}$ belongs to $\mathbf{F}^t_L(\theta,\rho)$ along with (s_τ).
(ii) The signal $(\bar{s}_\tau = \overline{s_\tau})_\tau$ ($\bar{a}$ is the complex conjugate of $a \in \mathbb{C}$) belongs to $\mathbf{F}^t_L(\theta,\rho)$.

IV. Lifting. We are about to demonstrate that a well-filtered signal in a dimension $d \leq d^+$ can be viewed as a well-filtered signal, with properly updated parameters, in a dimension $d^+ > d$:

Proposition 6 . *Let $1 \leq d \leq d^+$, and let $(s_\tau)_{\tau\in\mathbf{Z}^d}$ be a signal which is well-filtered, with parameters θ, ρ, L, at a point $t \in \mathbf{Z}^d$. Then the signal $(s^+_{\tau_1,...,\tau_{d^+}} = s_{\tau_1,...,\tau_d})$ is well-filtered, with the parameters $\theta^+ = (2L+1)^{(d^+-d)/2}\theta$, $\rho^+ = \rho$, $L^+ = L$ at every point $t^+ \in \mathbf{Z}^{d^+}$ such that $(t^+_1, ..., t^+_d) = t$.*

V. "Tensor product". Let $d = d' + d''$ with positive integers d', d'', so that $\mathbf{Z}^d = \mathbf{Z}^{d'} \times \mathbf{Z}^{d''}$. Given random fields $(s'_{\tau'}(\xi))_{\substack{\tau'\in\mathbf{Z}^{d'}\\ \xi}} \in \mathbf{F}^{t'}_L(0, \rho')$, $(s''_{\tau''}(\xi))_{\substack{\tau''\in\mathbf{Z}^{d''}\\ \xi}} \in \mathbf{F}^{t''}_L(0, \rho'')$, we define their tensor product as the field

$$(s_\tau(\xi) = s'_{\tau'}(\xi)s''_{\tau''}(\xi))_{\substack{\tau=(\tau',\tau'')\in\mathbf{Z}^d\\ \xi}}.$$

Proposition 7 *In the situation in question, one has $(s_\tau) \in \mathbf{F}^{(t',t'')}_L(0, \rho'\rho'')$.*

3.5 Examples of well-filtered signals

I. Exponential and algebraic polynomials. Let us define an *exponential polynomial* (s_τ) on $\mathbf{Z}^d$ as a *finite* sum of *exponential monomials* $c\tau^\alpha \exp\{\omega^T\tau\} \equiv c\tau_1^{\alpha_1}...\tau_d^{\alpha_d} \exp\{\omega^T\tau\}$ with *nonnegative* multi-indices α and $\omega \in \mathbb{C}^d$:

$$s_\tau = \sum_{\ell=1}^{M} c_\ell \tau^{\alpha(\ell)} \exp\{\omega^T(\ell)\tau\}, \tag{12}$$

where $\omega(\ell)$ and $\alpha(\ell)$ are deterministic, and c_ℓ may be random. Given an exponential polynomial (s_τ) on $\mathbf{Z}^d$, we define its *partial sizes* N_j, $j = 1, ..., d$, as follows: let m_j be the maximum of the degrees $\alpha_j(\ell)$, $\ell = 1, ..., M$, of the variable τ_j in the monomials of the sum (12), and M_j be the number of *distinct from each other* complex numbers among the "partial frequencies" $\omega_j(\ell)$: $M_j = \mathrm{Card}\,\mathbf{O}_j$, $\mathbf{O}_j = \{\omega_j(\ell) : 1 \leq \ell \leq M\}$. The *$j$-th partial size* $N_j(s)$ of exponential polynomial (12) is, by definition, the integer $(m_j + 1)M_j$. For example, with all frequencies equal to 0, an exponential polynomial becomes an algebraic polynomial, and its j-th size is by 1 larger than the degree of the polynomial w.r.t. j-th variable τ_j.

Proposition 8 *Let (s_τ) be an exponential polynomial on $\mathbf{Z}^d$ of partial sizes $N_1, ..., N_d$. Then for all $t \in \mathbf{Z}^d$ one has*

$$(s_\tau) \in \mathbf{F}^t_\infty(0, \rho_d(N_1, ..., N_d)),\ \rho_d(N_1, ..., N_d) = \prod_{j=1}^{d}[(2N_j - 1)^{1/2}2^{3N_j/2}], \tag{13}$$

and the filters $q^{(T)}$ certifying this inclusion can be chosen to be depending solely on T and on the collection of d sets $\mathbf{O}_j = \{\omega_j(\ell) : 1 \leq \ell \leq M\}$ of partial frequencies.

Remark 9 *A shortcoming of* (13) *is in a dramatic growth of $\rho_d(N, N, ..., N)$ with N and d. In several important cases, better bounds for ρ can be found. For example, an algebraic polynomial of degree m in every variable*

$$p_\tau = \sum_{\alpha \geq 0, |\alpha| \leq m} c_\alpha \tau^\alpha \tag{14}$$

belongs to $\mathbf{F}^t_\infty(0, (16m)^d)$ for every t, and the filters $q^{(T)}$ certifying this inclusion can be chosen to depend solely on T, d, m.

II. Solutions to homogeneous difference equations and harmonic functions. Consider a difference operator $\mathbf{D}$:

$$(\mathbf{D}f)_\tau = \sum_{\ell=1}^{k} w_\ell f_{\tau - \alpha(\ell)}; \tag{15}$$

here $\alpha(1), ..., \alpha(k) \in \mathbf{Z}^d$ and $w_1, ..., w_k \in \mathbb{C}$. For a positive integer N and $t \in \mathbf{Z}^d$, let

$$B^t_N = \{\tau \in \mathbf{Z}^d \mid |\tau - t| \leq N\},\ B^t_N(\mathbf{D}) = \{\tau \in B^t_N \mid \tau + \alpha(\ell) \in B^t_N,\ \ell = 1, ..., k\},$$
$$\mathbf{H}^t_N(\mathbf{D}) = \{(s) \in C(\mathbf{Z}^d) \mid s_\tau = (\mathbf{D}s)_\tau \quad \forall \tau \in B^t_N(\mathbf{D})\}.$$

For example, with

$$(\mathbf{D}f)_\tau = \frac{1}{2d} \sum_{\substack{i=1,...,d \\ \epsilon=\pm 1}} f_{\tau_1, ..., \tau_{i-1}, \tau_i + \epsilon, \tau_{i+1}, ..., \tau_d}, \tag{16}$$

the linear space $\mathbf{H}^t_N(\mathbf{D})$ is the space of fields which are "discrete harmonic" on B^t_N, that is, $s_\tau = \frac{1}{2d} \sum_{\substack{i=1,...,d \\ \epsilon=\pm 1}} s_{\tau_1, ..., \tau_{i-1}, \tau_i + \epsilon, \tau_{i+1}, ..., \tau_d}$ for all τ with $|\tau - t| \leq N - 1$.
Let us call a difference operator $\mathbf{D}$ *regular*, if it possesses the following properties:
R.1 The vectors $\{\alpha(\ell)\}_{1 \leq \ell \leq k}$ span the entire $\mathbb{R}^d$;
R.2 The coefficients $w_\ell = \rho_\ell \exp\{i\phi_\ell\}$ ($\rho_\ell \geq 0$, $\phi_\ell \in \mathbb{R}$) are nonzero, and

$$(a)\ \sum_{\ell=1}^{k} \rho_\ell \leq 1; \qquad (b)\ \sum_{\ell=1}^{k} \rho_\ell \alpha(\ell) = 0. \tag{17}$$

For example, the averaging operator (16) and its degrees are regular.
It turns out that the solutions of homogeneous difference equations with reg-

ular difference operators are well-filtered:

Proposition 10 *Let* $\mathbf{D}$ *be a regular difference operator. Then there exists a constant* $c = c(\mathbf{D}) > 0$ *such that*

$$\forall N > 0: \quad \mathbf{H}^t_N(\mathbf{D}) \subset \mathbf{F}^t_{\lfloor cN \rfloor}(0, c^{-1}). \tag{18}$$

As a nontrivial application example for Proposition 10, consider the families of random fields defined as follows. Let $d \le 4$, M be a positive integer, and R be a positive real. Consider the family $\mathbf{H}^+(M)$ of all deterministic continuous functions f on $\mathbb{R}^d$ which are harmonic in the interior of the box $D^0_{2M} = \{x \in \mathbb{R}^d : |x_j| \le 2M, j \le d\}$: $\left(\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}\right) f(x) = 0$, $x \in \mathrm{int} D^0_{2M}$. Now let $\mathbf{H}^+(M, R)$ be the family of random functions f such that all realizations of a function belong to $\mathbf{H}^+(M)$ and, besides this, $E\{\|f\|^2_{\infty,2M}\} \le R^2$, where $\|f\|_{\infty,2M}$ is the uniform norm on D^0_{2M}. Restricting functions f from $\mathbf{H}^+(M, R)$ on $\mathbf{Z}^d$, we get a family of random fields $\mathbf{H}(M, R)$ on $\mathbf{Z}^d$.

Proposition 11 *Let* $d \le 4$, M *be a positive integer and* $R > 0$ *be a real. For an appropriately chosen absolute constant* $c > 0$, *for all deterministic fields* $(s_\tau) \in \mathbf{H}(M, R)$ *one has*

$$|t| \le cM, L \le cM \Rightarrow (s_\tau) \in \mathbf{F}^t_L(c^{-1}R, c^{-1}), \tag{19}$$

and the filters $q^{(T)}$ *certifying the above inclusion can be chosen depending solely on* d, T.

4 Prediction

Above, we were focusing on the case of *de-noising* – recovering a well-filtered signal (s) at a point $t \in \mathbf{Z}^d$ via a given number observations "around" this point. Now we intend to focus on the case of *prediction*, where the goal is to recover s_t via observations y_τ "preceding by a given horizon $\kappa \in \mathbf{Z}_+$" the point t, i.e., observations with $\tau_j \le t_j - \kappa$, $j = 1, ..., d$. It turns out that the above theory admits straightforward prediction version, which is as follows.

Definition 12 *Let* $\theta \ge 0$, $\rho \ge 1$ *be reals, let* $T_0 \ge \kappa$ *be nonnegative integers,* L *be a nonnegative integer or* $+\infty$, *and let* $t \in \mathbf{Z}^d$. *Finally, let* $(s_\tau)_{\tau \in \mathbf{Z}^d} \equiv (s_\tau(\xi))_{\substack{\tau \in \mathbf{Z}^d \\ \xi \in \Omega}}$ *be a random field on* $\mathbf{Z}^d$.
(1) [T-well-predicted signals] *Let* T *be a nonnegative integer. We say that* (s_τ) *is* T*-well predicted with the parameters* θ, ρ, κ, L, *at the point* t (notation: $(s_\tau) \in \mathbf{Q}^t_{\kappa,L}(\theta, \rho, T)$), *if there exists a filter* $q = q^{(T)} \in P^\kappa_T(\mathbf{Z}^d)$, $|q|_2 \le \frac{\rho}{(2T+1)^{d/2}}$,

which reproduces (s_τ) in the box $\{\tau : |\tau - t| \leq L\}$ with the mean square error not exceeding $\theta(2T+1)^{-d/2}$:

$$\max_{\tau:|\tau-t|\leq L} \left[E\left\{|s_\tau - (q(\Delta)s)_\tau|^2\right\}\right]^{1/2} \leq \theta(2T+1)^{-d/2}. \tag{20}$$

(2) [well-predicted signals] *We say that (s_τ) is well-predicted, with the parameters θ, ρ, κ, T_0, L, at the point t* (notation: $(s_\tau) \in \mathbf{P}^t_{\kappa,T_0,L}(\theta,\rho)$)*, if, for every integer T, $T_0 \leq T \leq L$, (s_τ) is T-well-predicted, with the parameters θ, ρ, κ, L, at t.*

Remark 13 *Note that the quantitative description of a well-predicted field, as compared with the description of a well-filtered field, involves an extra parameter T_0 – the smallest "window width" starting with which a possibility to predict s_t is postulated. In the case of well-filtered fields, this width is just 0, in full accordance with the fact that in the de-noising problem every signal is 0-well-filtered, at every point t, with parameters $\theta = 0$, $\rho = 1$, $L = \infty$ due to the existence of the trivial "single-point" filter $q(z) \equiv 1$.*

4.1 The predictor

The predictor we intend to use is as follows:
Algorithm B: *Given a setup $(\rho \geq 1, \kappa, T)$ and a point $t \in \mathbf{Z}^d$, we build a prediction $\widehat{s}_t[\kappa, T, y]$ of s_t via observations (y_τ), $\kappa \leq t_j - \tau_j \leq 4T$, $j = 1, ..., d$, as $\widehat{s}_t[\kappa, T, y] = (\widehat{\psi}^t(\Delta)y)_t$, where $\widehat{\psi}^t \in P^\kappa_{2T}(\mathbf{Z}^2)$ is an optimal solution to the following optimization problem:*

$$\min_{\psi \in C^\kappa_{2T}(\mathbf{Z}^d)} \Big\{ \underbrace{|\Delta_1^{-t_1} ... \Delta_d^{-t_d}(1-\psi(\Delta))y|^*_{2T,\infty}}_{J(\psi, y^t_{\kappa,4T})} : |\psi|^*_{2T,1} \leq \frac{2^{d/2}\rho^2}{(2T+1)^{d/2}} \Big\}; \tag{21}$$

where $y^t_{\kappa,L} = \{y_\tau : \kappa \leq t_j - \tau_j \leq L,\ j = 1, ..., d\}$.
Note that the objective in (21) is affected only by observations $y^t_{\kappa,4T}$, so that our algorithm recovers s_t via $(4T - \kappa + 1)^d$ observations "around" the point t.

4.2 Main result: prediction

Theorem 14 *Assume that the signal (s_τ) underlying observations (2) is T-well-predicted, with parameters θ, ρ, κ, $L \geq 3T$: $(s_\tau) \in \mathbf{Q}^t_{\kappa,L}(\theta, \rho, T)$ with $L \geq 3T$. Then the mean square error of the estimate $\widehat{s}_t[\kappa, T, \cdot]$ of s_t yielded by*

Algorithm B with setup (ρ, κ, T) can be bounded from above as follows:

$$\left(E\left\{|\widehat{s}_t[\kappa, T, y] - s_t|^2\right\}\right)^{1/2} \leq c(d)\rho^3 \frac{\theta + \sigma\rho\sqrt{\ln(2T+1)+1}}{(2T+1)^{d/2}}, \qquad (22)$$
$$c(d) = 3(2^d + 2^{3d-1}).$$

In particular, if (s_τ) is well-predicted, with the parameters θ, ρ, κ, T_0, L, at a point t, then for every integer T, $T_0 \leq T \leq \lfloor L/3 \rfloor$, the accuracy of the estimate $\widehat{s}_t[\kappa, T, y]$ of s_t yielded by Algorithm B can be bounded by (22). Finally, in the case of deterministic (s), we have

$$|s_t - \widehat{s}_t[T, y]| \leq c(d)\rho^3 \left[\theta + \sigma\rho\Theta_T^t\right](2T+1)^{-d/2}, \qquad (23)$$
$$\Theta_T^t = \sigma^{-1} \max_{\tau:|\tau|\leq 2T} |\Delta_1^{\tau_1 - t_1} ... \Delta_d^{\tau_d - t_d} e|^*_{2T,\infty}.$$

Proof is identical to that of Theorem 2.

4.3 Calculus of well-predicted signals

The "calculus" of well-predicted signals (and its justification) is completely similar to those for well-filtered signals.

4.3.1 Operations preserving well-predictability

I. "Scale" of well-predicted signals.

$$\rho' \geq \rho, \theta' \geq \theta, \kappa' \leq \kappa, T_0' \geq T_0, L' \leq L \Rightarrow \mathbf{P}^t_{\kappa, T_0, L}(\theta, \rho) \subset \mathbf{P}^t_{\kappa', T_0', L'}(\theta', \rho').$$

II. Taking linear combinations.

Proposition 15 *Let $(s^j_\tau) \in \mathbf{P}^t_{\kappa_j, T_0^j, L}(\theta_j, \rho_j)$, $j = 1, ..., m$, and let $\lambda^j \in \mathbb{C}$ be random variable independent of (s^j) and such that $E\{|\lambda_j|^2\} < \infty$, $j = 1, ..., m$. Then*

$$\begin{aligned} &(s_\tau \equiv \sum_{j=1}^m \lambda_j s^j_\tau) \in \mathbf{P}^t_{\kappa^+, L^+}(\theta^+, \rho^+), \\ &\theta^+ = (2m-1)^{d/2} 2^{m-1} \rho_1 ... \rho_m \sum_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j}, \\ &\rho^+ = (2m-1)^{d/2} 2^m \rho_1 ... \rho_m, \kappa^+ = \min_{1\leq j\leq m} \kappa_j, \ T_0^+ = m \max_{1\leq j\leq m} T_0^j, \\ &L^+ = \lfloor L/2 \rfloor. \end{aligned} \qquad (24)$$

In the case of $m = 1$, one can set $\rho^+ = \rho_1$, $\theta^+ = |\lambda_1|\theta_1$, $\kappa^+ = \kappa$, $T_0^+ = T_0$, $L^+ = L$. The filters certifying the well-predictability of (s_τ) can be chosen to be independent of the coefficients λ_j.

III. Modulation and conjugation.

Proposition 16 *Let $(s_\tau) \in \mathbf{P}^t_{\kappa,T_0,L}(\theta,\rho)$.*
(i) Let $\omega \in \mathbb{R}^d$, $\phi \in \mathbb{R}$ be deterministic. Then the signal $(\widehat{s}_\tau = \exp\{i[\omega^T\tau + \phi]\}s_\tau)_{\tau\in\mathbf{Z}^d}$ also belongs to $\mathbf{P}^t_{\kappa,T_0,L}(\theta,\rho)$.
(ii) The field $(\bar{s}_\tau = \overline{s_\tau})_\tau$ ($\overline{a}$ is the complex conjugate of $a \in \mathbb{C}$) belongs to $\mathbf{P}^t_{\kappa,T_0,L}(\theta,\rho)$.

IV. Lifting.

Proposition 17 *. Let $1 \le d \le d^+$, and let $(s_\tau)_{\tau\in\mathbf{Z}^d}$ be a signal which is well-predictable, with parameters θ,ρ,κ,T_0,L, at a point $t \in \mathbf{Z}^d$. Then the signal $(s^+_{\tau_1,...,\tau_{d^+}} = s_{\tau_1,...,\tau_d})$ is well-predictable, with the parameters*

$$\theta^+ = (2L+1)^{(d^+-d)/2}\theta,\ \rho^+ = (2\kappa+1)^{(d_+-d)/2}\rho,\ \kappa^+ = \kappa,\ T_0^+ = T_0,\ L^+ = L,$$

at every point $t^+ \in \mathbf{Z}^{d^+}$ such that $(t_1^+,...,t_d^+) = t$.

V. "Tensor product". Let $d = d' + d''$ with positive integers d', d'', so that $\mathbf{Z}^d = \mathbf{Z}^{d'} \times \mathbf{Z}^{d''}$, and let $(s'_{\tau'}(\xi))_{\substack{\tau'\in\mathbf{Z}^{d'}\\ \xi}} \in \mathbf{P}^{t'}_{\kappa,T_0,L}(0,\rho')$, $(s''_{\tau''}(\xi))_{\substack{\tau''\in\mathbf{Z}^{d''}\\ \xi}} \in \mathbf{P}^{t''}_{\kappa,T_0,L}(0,\rho'')$.

Proposition 18 *In the situation in question, the tensor product $(s_\tau(\xi) = s'_{\tau'}(\xi)s''_{\tau''}(\xi))_{\substack{\tau=(\tau',\tau'')\in\mathbf{Z}^d\\ \xi}}$. of the fields (s'), (s'') belongs to $\mathbf{P}^{(t',t'')}_{\kappa,T_0,L}(0,\rho'\rho'')$.*

4.4 Basic example of well-predicted signal: quasi-stable exponential polynomial

Let us define a *quasi-stable* exponential polynomial (s_τ) on $\mathbf{Z}^d$ as an exponential polynomial

$$s_\tau = \sum_{\ell=1}^{M} c_\ell \tau^{\alpha(\ell)} \exp\{\omega^T(\ell)\tau\} \tag{25}$$

where all partial frequencies $\omega_j(\ell)$ satisfy the restriction $\Re(\omega_j(\ell)) \le 0$. For example, an algebraic polynomial (partial frequencies are zero) and a trigonometric polynomial (partial frequencies are imaginary) are quasi-stable.

Proposition 19 *Let (s_τ) be a quasi-stable exponential polynomial on $\mathbf{Z}^d$ of partial sizes $N_1, ..., N_d$. Then for every integer $\kappa \geq 0$ and all $t \in \mathbf{Z}^d$ one has*

$$
\begin{array}{l}
(s_\tau) \in \mathbf{P}^t_{\kappa,T_0,\infty}(0, \rho_{\kappa,d}(N_1, ..., N_d)),\\
\rho_{\kappa,d}(N_1, ..., N_d) = \prod\limits_{j=1}^{d} [(2N_j - 1)^{1/2} 2^{N_j} \left(\max[2, 2\kappa + 1]\right)^{N_j/2}],,\\
T_0 = \kappa \max\limits_{1\leq j\leq d} N_j
\end{array}
\tag{26}
$$

and the filters $q^{(T)}$ certifying this inclusion can be chosen to be depending solely on T, κ and on the collection of d sets $\mathbf{O}_j = \{\omega_j(\ell) : 1 \leq \ell \leq M\}$ of partial frequencies.

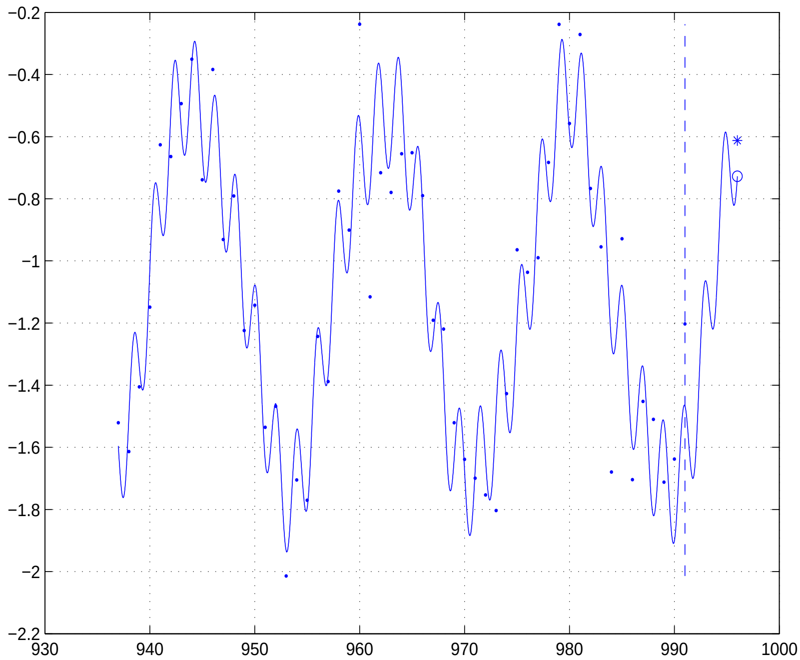
4.5 Numerical illustration

Here we present a numerical illustration for Algorithm B as applied to prediction/filtration of a time series which is a linear combination of a given number of harmonic oscillations; both amplitudes and frequencies of these oscillations are *not* known in advance. The (randomly generated) signal to be considered in our illustration is

$$
\begin{array}{rl}
s_t = & -0.7191 \cos(0.0017\, t) + 1.5200 \cos(0.0025\, t)\\
& +0.6041 \cos(0.3526\, t) - 0.0676 \cos(0.8303\, t)\\
& -0.1892 \cos(3.2371\, t).
\end{array}
\tag{27}
$$

We used Algorithm B to predict $s_{t+\Delta t}$ via $2T = 512$ noisy observations $y_\tau = s_\tau + 0.2\xi_\tau$, $\xi_\tau \in \mathbf{N}(0, 1)$, $\tau = t - 511, t - 510, ..., t$. On Fig. 1 we display the observations and the prediction made at instant $t = 991$ for the prediction time $\Delta t = 5$. On Fig. 2, we compare the quality of prediction given by Algorithm B with the quality of the "ideal least squares prediction" corresponding to the case where the frequencies ω_j, $j = 1, ..., 5$ participating in (27) are known in advance:

$$
\begin{array}{l}
f^*_{t+\Delta t} = \sum\limits_{\tau=0}^{512} q^*_\tau y_{t-\tau},\\
q^* = \operatorname*{argmin}\limits_{q} \left\{ \sum\limits_{\tau=0}^{512} q_\tau^2 : \begin{array}{c} \sum\limits_{\tau=0}^{5} 12 q_\tau \cos(\omega_j(t-\tau)) = \cos(\omega_j(t+\Delta t)) \\ \forall t \in \mathbf{Z}\ \forall j = 1, ..., 5 \end{array} \right\}
\end{array}
$$



dots: observations; **o**: s_{996}; $*$: the prediction

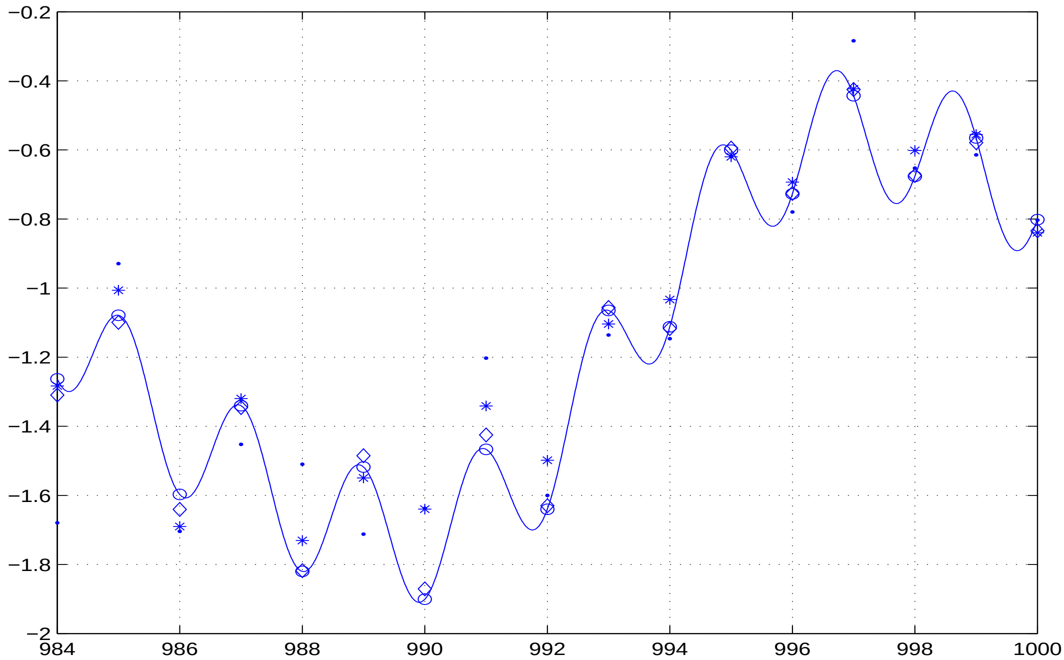
Fig. 1. Prediction of s_{996} via 512 observations $y_{480}, y_{481}, ..., y_{991}$. The curve on the picture is the "continuous time" extension of (s_τ)

5 Adaptive versions of the estimates

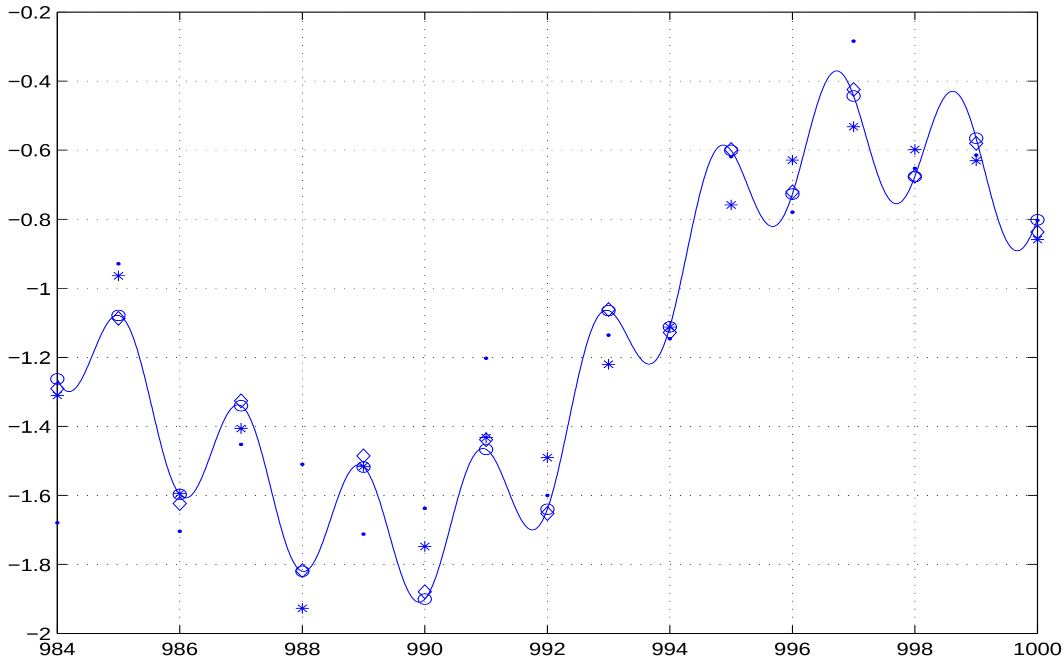
5.1 The goal

In the estimates we have considered so far, the "window width" T was given in advance, and we were able to recover s_t at a "nearly parametric rate", provided that f is well-filtered with parameters ρ, θ, $L \geq 3T$ at the point t and ρ is known in advance. Under the latter assumption, the best way to recover s_t would be to use the estimate associated with the *largest possible* window width T compatible with the assumption, that is, with $T = T_* \equiv \lfloor L/3 \rfloor$. In order to act so, we, however, should know in advance the value of L. What to do if this value is unknown, and all we have is an upper bound $L_\infty < \infty$ on L? Can we still recover s_t as well as if we knew L? It turns out that *in the case of deterministic signal* f, in order to get an affirmative answer to the above question it suffices to know (upper bounds on) *both* θ and ρ. Under this assumption, the desired estimate can be obtained by applying the famous *Lepskii's adaptation routine* [36,37] to the estimates given by Algorithms A, B. Our goal in this section is to present and to analyze the resulting *adaptive* (to L) estimate.

Below, the signal to be recovered is assumed to be deterministic (however, the construction and the results can be easily extended to the case when this signal



$\Delta t = 0$, MSE(B) $= 0.09 = 0.45\sigma$, MSE(I) $= 0.024 = 0.12\sigma$



$\Delta t = 5$, MSE(B) $= 0.066 = 0.33\sigma$, MSE(I) $= 0.036 = 0.18\sigma$

dots: observations; **o**: signal; $*$: prediction, Algorithm B; $\diamondsuit$: ideal prediction

Fig. 2. "Ideal" prediction ($\diamondsuit$) vs. Algorithm B ($*$). MSE(I), MSE(B) – mean square errors of the forecasts taken over 17 predictions shown on the picture.

is Gaussian (i.e., all s_τ, $\tau \in \mathbf{Z}^d$, are Gaussian)). For the sake of definiteness, we restrict ourselves with the de-noising problem and Algorithm A; the case of the prediction problem is completely similar.

5.2 The adaptive estimate

The "design parameters" of the estimate we intend to build are $\theta \geq 0$, $\rho \geq 1$ and a positive integer L_∞ (and, of course, a point $t \in \mathbf{Z}^d$ where we want to recover f).

5.2.1 Preliminaries

Let

$$T_\infty = \lfloor L_\infty/3 \rfloor, \; \Theta^t = \max_{0 \leq T \leq T_\infty} \Theta^t_T \tag{28}$$

(see (8)). Note that by its origin, Θ^t is the maximum of $O(L_\infty^d)$ standard Gaussian random variables; applying Lemma 23, we have

$$\forall w \geq 1 : \text{Prob}\left\{\Theta^t > Cw\sqrt{\ln(2L_\infty + 1)}\right\} \leq \exp\left\{-\frac{w^2 \ln(2L_\infty + 1)}{2}\right\}, \tag{29}$$

where $C > 0$ depends solely on d.
Let us fix a "safety factor" ω in such a way that the event $\Theta^t > \omega\sqrt{\ln(2L_\infty + 1)}$ is "highly un-probable", namely,

$$\text{Prob}\left\{\Theta^t > \omega\sqrt{\ln(2L_\infty + 1)}\right\} \leq (2L_\infty + 1)^{-2d}; \tag{30}$$

by (29), the required ω may be chosen as a function of d only.
"Good" realizations of noise. Let us define the set of "good realizations of noise" as

$$\Xi = \{\epsilon \mid \Theta^t \leq C\omega\sqrt{\ln(2L_\infty + 1)}\} \tag{31}$$

(see (29)).
The "ideal" window width. Let us define the *ideal* window width T_* as the largest $T \geq 0$ such that $3T \leq L_\infty$ and $(s_\tau) \in \mathbf{F}^t_{3T}(\theta, \rho)$. Note that these requirements are clearly satisfied for $T = 0$, so that T_* is well-defined.
Recall that the goal stated at the beginning of this section is to recover s_t as if we knew T_* and were using the estimate given by Algorithm A with the setup

(ρ, T_*). Note that (8) implies the following "conditional bound" on the quality of estimates $\widehat{s}_t[T, y]$ yielded by Algorithm A with the setup $(\rho, T \le T_*)$:

$$0 \le T \le T_*, \epsilon \in \Xi \Rightarrow |s_t - \widehat{s}_t[T, y]| \le \overbrace{c(d)\rho^3 \left[\theta + \sigma\rho\kappa\right](2T+1)^{-d/2}}^{\gamma(T)}, \quad (32)$$
$$\kappa = C\omega\sqrt{\ln(2L_\infty + 1)}.$$

Normal window widths. Given observations, let us say that a window width T is *normal*, if $0 \le T \le L_\infty/3$ and the estimate $\widehat{s}_t[T', y]$, $0 \le T' \le T$, in the sense that

$$0 \le T' \le T \Rightarrow |\widehat{s}_t[T, y] - \widehat{s}_t[T', y]| \le \gamma(T) + \gamma(T'). \quad (33)$$

Note that normal window widths do exist (e.g., $T = 0$) and that the property of a window width to be normal "is observable" – it is expressed solely in terms of the observations. It is also clear that among the normal window widths, there exists the largest one $T_{\max} = T_{\max}(y)$. *By construction*, the adaptive estimate $\widehat{s}_t(y)$ of s_t is the estimate $\widehat{s}_t[T_{\max}, y]$. Note that the only design parameters of this estimate are θ, ρ (these quantities specify the function $\gamma(\cdot)$) and L_∞.
Quality of the adaptive estimate.

Theorem 20 *Let (s_τ) be a deterministic signal, and let T_* be the corresponding ideal window width. The risk of the adaptive estimate $\widehat{s}_t(y)$ can be bounded as follows:*

$$\left(E\left\{|s_t - \widehat{s}_t|^2\right\}\right)^{1/2} \le C(d)\rho^3 \left[\theta + \sigma\rho\sqrt{\ln(2L_\infty + 1)}\right](2T_* + 1)^{-d/2} \quad (34)$$

with $C(d)$ depending solely on d.

Note that bound (34) differs from our "target" bound, namely, the bound (7) with $T = T_*$, only in the logarithmic term.

6 Applications to nonparametric regression estimation

6.1 Recovering locally well-filtered regression functions of unknown local structure: motivation

Now let us pass to the Nonparametric Regression problem. As it was explained in Introduction, in this problem observations (2) come from a function ("signal") f of continuous argument (which we assume to vary in the d-dimensional

unit cube $[0,1]^d$); this function is observed in noise along an n-point equidistant grid in $[0,1]^d$, and the problem is to recover f via these observations.
We are about to demonstrate that the results of Section 3 on recovering well-filtered signals of unknown structure can be applied to recovering nonparametric signals which admit well-filtered *local* approximations. The universe of these signals is much wider than the one of smooth signals (the latter correspond to the very particular case when the local approximations in question are algebraic polynomials of fixed order). In particular, we can handle "modulated smooth signals" – sums of a fixed number of products of smooth functions and multivariate harmonic oscillations of unknown (and arbitrarily high) frequencies. As it was explained in Introduction, such an extension has an unavoidable price – now we cannot hope to recover the signal well outside of the observation grid. As a result, *in what follows we are interested in recovering the signals along the observation grid only* and, consequently, replace the integral L_q risk measures by their grid analogies, see Section 6.4.
The estimates to be developed will be "double adaptive", that is, adaptive with respect to both the unknown in advance structures of well-filtered approximations of our signals and to the unknown in advance "approximation rate" – the dependence between the size of a neighborhood of a point where the signal in question is approximated and the quality of approximation in this neighborhood (in the case of smooth signals, this approximation rate is exactly what is given by the smoothness parameters). The results to follow can be seen as extensions of the results of [41,28] (see also [42]) dealing with the particular case of univariate signals satisfying differential inequalities with unknown differential operators.
Now let us pass to a formal treatment of the regression problem.

6.2 The regression problem

Let m be a positive integer, $n=(m+1)^d$, and let $\Gamma_n=\{x=m^{-1}\alpha:\alpha\in\mathbf{Z}^d, 0\le\alpha, |\alpha|\le m\}$. We associate with a signal $f\in C([0,1]^d)$ its observations along Γ_n:

$$y\equiv y^n_f(\epsilon)=\{y_\tau\equiv y^n_\tau(f,\epsilon)=f(m^{-1}\tau)+e_\tau, e_\tau=\sigma\epsilon_\tau\}_{0\le\tau,|\tau|\le m}, \qquad (35)$$

where $\{\epsilon_\tau\}_{\tau\in\mathbf{Z}^d}$ are independent standard Gaussian complex-valued random noises. Our goal is to recover $f|_{\Gamma_n}$ from observations (35). In what follows, we write

$$f_\tau=f(m^{-1}\tau) \qquad [\tau\in\mathbf{Z}^d, m^{-1}\tau\in[0,1]^d]$$

Notation. Below we use the following notation. We set

$$\Gamma_n^o = \Gamma_n \cap (0,1)^d = \{m^{-1}t : t \in \mathbf{Z}^d, t > 0, |t| < m\}.$$

For a set $B \subset [0,1]^d$, we denote by $\mathbf{Z}(B)$ the set of all $t \in \mathbf{Z}^d$ such that $m^{-1}t \in B$. Let $x = m^{-1}t \in \Gamma_n^o$. We say that a nonempty *open* cube

$$B_h(x) = \{u \mid |u_i - x_i| < h/2, \ i = 1, ..., d\}$$

centered at x is *admissible* for x, if $B_h(x) \subset [0,1]^n$. For such a cube, $T_h(x)$ denotes the largest nonnegative integer T such that $\mathbf{Z}(B) \supset \{\tau \in \mathbf{Z}^d : |\tau - t| \leq 4T\}$. For a cube $B = \{x \in \mathbf{R}^d : |x_i - c_i| \leq h/2, \ i = 1, ..., d\}$, $D(B) = h$ stands for the edge of B. For $\gamma \in (0,1)$ and a cube $B = \{x \in \mathbf{R}^d : |x_i - c_i| \leq h/2, \ i = 1, ..., d\}$,

$$B_\gamma = \{x \in \mathbf{R}^d : |x_i - c_i| \leq \gamma h/2, i = 1, ..., d\}$$

is the γ-shrinkage of B to the center of B.

Families of locally well-filtered signals. Let $B \subset [0,1]^d$ be a cube, k be a positive integer, $\rho \geq 1$, $R \geq 0$ be reals, and let $p \in (d, \infty]$. The collection B, k, ρ, R, p specifies the family $\mathbf{F}^{k,\rho,p}(B,R)$ of *locally well-filtered on* B signals f defined by the following requirements:

(1) $f \in C([0,1]^d)$;

(2) There exists a nonnegative function $F \in L_p(B)$, $\|F\|_{p,B} \leq R$, such that for every $x = m^{-1}t \in \Gamma_n \cap \mathrm{int} B$ and for every admissible for x cube $B_h(x)$ contained in B there exists a field $\phi \in C(\mathbf{Z}^d)$ such that $\phi \in \mathbf{S}^t_{3T_h(x)}(0, \rho, T_h(x))$ and

$$\forall \tau \in \mathbf{Z}(B_h(x)) : |\phi_\tau - f_\tau| \leq h^{k-d/p} \|F\|_{p,B_h(x)}. \tag{36}$$

In the sequel, we use for $\mathbf{F}^{k,\rho,p}(B;R)$ also the shortened notation $\mathbf{F}[\psi]$, where ψ stands for the collection of "parameters" (k, ρ, p, B, R).

Motivating example: Modulated smooth signals. Let a cube $B \subset [0,1]^d$, $p \in (d, \infty]$, positive integers k, ν and a real $R \geq 0$ be given. Consider a collection of ν functions $g_1, ..., g_\nu \in C([0,1]^d)$ which are k times continuously differentiable and satisfy the requirement $\sum\limits_{\ell=1}^{\nu} \|D^k g_\ell\|_{p,B} \leq R$. Let $\omega(\ell) \in \mathbf{R}^d$, and let $f(x) = \sum\limits_{\ell=1}^{\nu} g_\ell(x) \exp\{i\omega^T(\ell)x\}$. By standard argument [3], whenever $x = m^{-1}t \in \Gamma_n \cap \mathrm{int} B$ and $B_h(x)$ is admissible for x, the Taylor polynomial $\Phi_\ell^x(\cdot)$, of order $k-1$, taken at x, of f_ℓ satisfies the inequality $u \in B_h(x) \Rightarrow |\Phi_\ell^x(u) - f_\ell(u)| \leq c_1 h^{k-d/p} \|F_\ell\|_{p,B_h(x)}, \quad F_\ell(u) = |D^k f_\ell(u)|$; here and in what follows, c_i are positive constants depending solely on d, k and ν. It follows that if $\Phi(u) = \sum\limits_{\ell=1}^{\mu} \Phi_\ell^x(u)\exp\{i\omega^T(\ell)u\}$ then

$$
\begin{array}{l}
u \in B_h(x) \Rightarrow |\Phi(u) - f(u)| \leq h^{k-d/p}\|F\|_{p,B_h(x)}, \\
F = c_2 \sum\limits_{\ell=1}^{\nu} F_\ell \qquad [\Rightarrow \|F\|_{p,B} \leq c_3 R].
\end{array} \tag{37}
$$

Now observe that the exponential polynomial $\phi(\tau) = \Phi(m^{-1}\tau)$ belongs to $\mathbf{F}^t_\infty(0, c_4)$ (Proposition 8). Combining this fact with (37), we conclude that $f \in \mathbf{F}^{k,\rho(\nu,k,d),p}(B, c(\nu,k,d)R)$.

6.4 Accuracy measures

Let us fix $\gamma \in (0,1)$. Given an estimate $\widehat{f}_n$ of $f|_{\Gamma_n}$ based on observations (35) (i.e., a Borel real-valued function of $x \in \Gamma_n$ and $y \in \mathbb{C}^n$) and $\psi = (k, \rho, p, B, R)$, let us characterize the quality of the estimate on the set $\mathbf{F}[\psi]$ by the worst-case risks

$$
\widehat{\mathbf{R}}_q\left(\widehat{f}_n; \mathbf{F}[\psi]\right) = \sup_{f \in \mathbf{F}[\psi]} \left(E\left\{\|\widehat{f}_n(\cdot; y_f(\epsilon)) - f|_{\Gamma_n}(\cdot)\|^2_{q,B_\gamma}\right\}\right)^{1/2},
$$

where $\|\cdot\|_{q,B'}$ is the discrete analogy of the standard L_q-norm on a cube B', so that

$$
\|g\|_{q,B_\gamma} = m^{-d/q}\left(\sum_{\tau \in \mathbf{Z}(B_\gamma)} |g_\tau|^q\right)^{1/q}.
$$

6.5 The estimate

The recovering routine we are about to build is aimed at estimating functions from classes $\mathbf{F}^{k,\rho,p}(B,R)$ with *unknown in advance parameters* k, ρ, p, B, R. The only design parameters of the routine is an a priori upper bound μ on the parameter ρ and a $\gamma \in (0,1)$.

6.5.1 Preliminaries

From now on, we denote by $\Theta = \Theta_{(n)}$ the deterministic function of observation noises defined as follows. For every cube $B \subset [0,1]^d$ with vertices in Γ_n, we consider the Fourier transform of the observation noises reduced to $B \cap \Gamma_n$, and take the maximum of modules of the resulting Fourier coefficients, let it be denoted $\theta_B(e)$. By definition,

$$
\Theta \equiv \Theta_{(n)} = \sigma^{-1} \max_B \theta_B(e), \tag{38}
$$

where the maximum is taken over all cubes B of the indicated type. Note that by the origin of $\Theta_{(n)}$ and Lemma 23, we have

$$\forall w \geq 1: \quad \mathrm{Prob}\left\{\Theta_{(n)} > Cw\sqrt{\ln n}\right\} \leq \exp\left\{-\frac{w^2 \ln n}{2}\right\}, \tag{39}$$

where $C > 0$ depends solely on d.

6.5.2 Building blocks: window estimates

To recover a signal f via $n = m^d$ observations (35), we use point-wise window estimates of f defined as follows. Let us fix a point $x = m^{-1}t \in \Gamma_n^o$; our goal is to build an estimate of $f(x)$. Let $B_h(x)$ be an admissible window for x. We associate with this window an estimate $\widehat{f}_n^h = \widehat{f}_n^h(x; y_f^n(\epsilon))$ of $f(x)$ defined as follows. If the window is "very small", specifically, $h \leq m^{-1}$, so that x is the only point from the observation grid Γ_n in $B_h(x)$, we set $T_h(x) = 0$ and $\widehat{f}_n^h = y_t$. For a larger window, we choose the largest nonnegative integer $T = T_h(x)$ such that $\mathbf{Z}(B_h(x)) \supset \{\tau : |\tau - t| \leq 4T\}$ and apply Algorithm A, the design parameters of the algorithm being $(\mu, T_h(x))$, to build the estimate of $f_t = f(x)$, let the resulting estimate be denoted by $\widehat{f}_n^h = \widehat{f}_n^h(x; y_f^n(\epsilon))$.
To characterize the quality of the estimate $\widehat{f}_n^h = \widehat{f}_n^h(x; y_f^n(\epsilon))$, let us set

$$\Phi_\mu(f, B_h(x)) = \min_p \left\{ \max_{\tau \in \mathbf{Z}(B_h(x))} |p_\tau - f_\tau| : p \in \mathbf{S}^t_{3T_h(x)}(0, \mu, T_h(x)) \right\}.$$

Lemma 21 *One has*

$$(f_\tau) \in \mathbf{S}^t_{3T_h(x)}(\theta, \mu, T_h(x)), \quad \theta = \frac{\Phi_\mu(f, B_h(x))(1+\mu)}{(2T+1)^{d/2}}. \tag{40}$$

Assuming that $h > m^{-1}$ and combining (40) with (8), we come to the following upper bound on the error of estimating $f(x)$ by the estimate $\widehat{f}_n^h(x; \cdot)$:

$$|f(x) - \widehat{f}_n^h(x; y_f(\epsilon))| \leq C_1 \left[\Phi_\mu(f, B_h(x)) + \frac{\sigma}{\sqrt{nh^d}} \Theta_{(n)}\right] \tag{41}$$

(note that $(2T_h(x)+1)^{-d/2} \leq C_0(nh^d)^{-1/2}$); from now on, C (perhaps with sub- or superscripts) are positive quantities depending on d, μ, γ only. Note that (41) by evident reasons holds true for "very small windows" (those with $h \leq m^{-1}$) as well.

6.5.3 The adaptive estimate

We are about to "aggregate" the window estimates $\widehat{f}_n^h$ into an *adaptive* estimate, applying Lepskii's adaptation scheme (cf. Section 5) in the same fashion as in [38,27,28].
Let us fix a "safety factor" ω in such a way that the event $\Theta_{(n)} > \omega\sqrt{\ln n}$ is "highly un-probable", namely,

$$\mathrm{Prob}\left\{\Theta_{(n)} > \omega\sqrt{\ln n}\right\} \le n^{-4(\mu+1)}; \tag{42}$$

by (39), the required ω may be chosen as a function of μ, d only.
"Good" realizations of noise. Let us define the set of "good realizations of noise" as

$$\Xi_n = \{\epsilon \mid \Theta_{(n)} \le \omega\sqrt{\ln n}\}. \tag{43}$$

Note that (41) implies the "conditional" error bound

$$\begin{aligned} &\epsilon \in \Xi_n \Rightarrow |f(x) - \widehat{f}_n^h(x; y_f(\epsilon))| \le C_1\left[\Phi_\mu(f, B_h(x)) + S_n(h)\right], \\ &S_n(h) = \frac{\sigma}{\sqrt{nh^d}}\omega\sqrt{\ln n}. \end{aligned} \tag{44}$$

Observe that

(*) *As h grows, the "deterministic term" $\Phi_\mu(f, B_h(x))$ does not decrease, while the "stochastic term" $S_n(h)$ decreases.*

The "ideal" window. Let us define the *ideal* window $B_*(x)$ as the largest admissible window for which the stochastic term dominates the deterministic one:

$$\begin{aligned} &B_*(x) = B_{h_*(x)}(x), \\ &h_*(x) = \max\{h \mid h > 0, B_h(x) \subset [0,1]^d, \Phi_\mu(f, B_h(x)) \le S_n(h)\}. \end{aligned} \tag{45}$$

Note that such a window does exist, since $S_n(h) \to \infty$ as $h \to +0$. Besides this, since the cubes $B_h(x)$ are open, the quantity $\Phi_\mu(f, B_h(x))$ is continuous from the left, so that

$$0 < h \le h_*(x) \Rightarrow \Phi_\mu(f, B_h(x)) \le S_n(h). \tag{46}$$

Thus, the ideal window $B_*(x)$ is well-defined for every x possessing admissible windows, i.e., for every $x = \Gamma_n^o = \{m^{-1}t : t \in \mathbf{Z}^d, 0 < t, |t| < m\}$.

Normal windows. Assume that $\epsilon \in \Xi_n$. Then the errors of all estimates $\widehat{f}_n^h(x;y)$ associated with admissible windows smaller than the ideal one are dominated by the corresponding stochastic terms:

$$\epsilon \in \Xi_n, 0 < h \leq h_*(x) \Rightarrow |f(x) - \widehat{f}_n^h(x; y_f(\epsilon))| \leq 2C_1 S_n(h) \tag{47}$$

(by (44) and (46)). Let us fix an $\epsilon \in \Xi_n$ (and thus – a realization y of the observations) and let us call an admissible for x window $B_h(x)$ *normal*, if the associated estimate $\widehat{f}_n^h(x;y)$ differs from every estimate associated with a smaller window by no more than $4C_1$ times the stochastic term of the latter estimate, i.e.

$$\begin{array}{c}\text{Window } B_h(x) \text{ is normal}\\ \Updownarrow \\ \left\{\begin{array}{l} B_h(x) \text{ is admissible} \\ \forall h', 0 < h' \leq h: \quad |\widehat{f}_n^{h'}(x;y) - \widehat{f}_n^h(x;y)| \leq 4C_1 S_n(h') \quad [y = y_f(\epsilon)] \end{array}\right.\end{array} \tag{48}$$

Note that if $x \in \Gamma_n^o$, then x possesses a normal window, specifically, the window $B_{m^{-1}}(x)$. Indeed, this window contains a single observation point, namely, x itself, so that the corresponding estimate, same as every estimate corresponding to a smaller window, by construction coincides with the observation at x, so that all the estimates $\widehat{f}_n^{h'}(x;y)$, $0 < h' \leq m^{-1}$, are the same. Note also that (47) says that

(!) *If $\epsilon \in \Xi_n$, then the ideal window $B_*(x)$ is normal.*

The adaptive estimate $\widehat{f}_n(x;y)$. The property of an admissible window to be normal is "observable" – given observations y, we can say whether a given window is or is not normal. Besides this, it is clear that among all normal windows there exists the largest one $B^+(x) = B_{h^+(x)}(x)$. *The adaptive estimate $\widehat{f}_n(x;y)$ is exactly the window estimate associated with the window $B^+(x)$.*
Note that from (!) it follows that

(!!) *If $\epsilon \in \Xi_n$, then the largest normal window $B^+(x)$ contains the ideal window $B_*(x)$.*

By definition of a normal window, under the premise of (!!) we have

$$|\widehat{f}_n^{h^+(x)}(x;y) - \widehat{f}_n^{h_*(x)}(x;y)| \leq 4C_1 S_n(h_*(x)),$$

and we come to the conclusion as follows:
(*) *If $\epsilon \in \Xi_n$, then the error of the estimate $\widehat{f}_n(x;y) \equiv \widehat{f}_n^{h^+(x)}(x;y)$ is dominated*

by the error bound (44) *associated with the ideal window:*

$$\epsilon \in \Xi_n \Rightarrow |\widehat{f}_n(x;y) - f(x)| \le 5C_1 \left[\Phi_\mu(f, B_{h_*(x)}(x)) + S_n(h_*(x))\right]. \qquad (49)$$

Thus, the estimate $\widehat{f}_n(\cdot;\cdot)$ – which is based solely on observations and does not require any a priori knowledge of the "parameters of well-filterability of f" – possesses basically the same accuracy as the "ideal" estimate associated with the ideal window (provided, of course, that the realization of noises is not "pathological": $\epsilon \in \Xi_n$).
Note that the adaptive estimate $\widehat{f}_n(x;y)$ we have built depends solely on "design parameters" μ, γ (recall that C_1 depends on μ, γ), the volume of observations n and the dimension d.

6.6 Quality of estimation

Our main result is as follows:

Theorem 22 *Let $\gamma \in (0,1)$, $\mu \ge 1$ be an integer, let $\mathbf{F} = \mathbf{F}^{k,\rho,p}(B;R)$ be a family of locally well-filtered signals associated with a cube $B \subset [0,1]^d$ with $mD(B) \ge 1$, $\rho \le \mu$ and $p > d$. For properly chosen $P \ge 1$ depending solely on μ, d, p, γ and nonincreasing in $p > d$ the following statement takes place:*
If the volume $n = m^d$ of observations (35) *is large enough, namely,*

$$P^{-1} n^{\frac{2k-2d\pi+d}{2d}} \ge \frac{R}{\widehat{\sigma}_n} \ge P D^{-\frac{2k-2d\pi+d}{2}}(B) \qquad \left[\widehat{\sigma}_n = \sigma\sqrt{\frac{\ln n}{n}}, \quad \pi = \frac{1}{p}\right] \qquad (50)$$

($D(B)$ is the edge of the cube B), then for every $q \in [1,\infty]$ the worst case, with respect to $\mathbf{F}$, q-risk of the adaptive estimate $\widehat{f}_n(\cdot,\cdot)$ associated with the parameter μ can be bounded as follows:

$$\begin{aligned}
\widehat{\mathbf{R}}_q\left(\widehat{f}_n;\mathbf{F}\right) &\equiv \sup_{f\in\mathbf{F}} \left(E\left\{\|\widehat{f}_n(\cdot;y_f(\epsilon)) - f(\cdot)\|^2_{q,B_\gamma}\right\}\right)^{1/2} \\
&\le PR\left(\tfrac{\widehat{\sigma}_n}{R}\right)^{2\beta(p,k,d,q)} D^{d\lambda(p,k,d,q)}(B), \\
\beta(p,k,d,q) &= \begin{cases} \frac{k}{2k+d}, & \theta \ge \pi\frac{d}{2k+d} \\ \frac{k+d\theta-d\pi}{2k-2d\pi+d}, & \theta \le \pi\frac{d}{2k+d} \end{cases}, \ \theta = \tfrac{1}{q}, \\
\lambda(p,k,d,q) &= \begin{cases} \theta - \frac{d\pi}{2k+d}, & \theta \ge \pi\frac{d}{2k+d} \\ 0, & \theta \le \pi\frac{d}{2k+d} \end{cases};
\end{aligned} \qquad (51)$$

here B_γ is the concentric to B γ times smaller in linear sizes cube.

Note that the rates of convergence to 0, as $n \to \infty$, of the risks $\widehat{\mathbf{R}}_q\left(\widehat{f}_n; \mathbf{F}\right)$ of our adaptive estimate on the families $\mathbf{F} = \mathbf{F}^{k,\rho,p}(B;R)$ are exactly the same as those stated by Theorem 3.3.1 from [42] in the case of recovering non-parametric *smooth* regression functions from local Sobolev balls. It is well-known (see, e.g., [42]) that in the smooth case the latter rates are optimal in order, up to logarithmic in n factors. Since the families of locally well-filtered signals are much wider than local Sobolev balls (smooth signals are trivial examples of modulated smooth signals!), it follows that the rates of convergence stated by Theorem 22 also are nearly optimal.

7 Appendix: proofs

7.1 Preliminaries

Norm relations. Let us list several evident relations between the introduced semi-norms on $C(\mathbb{Z}^d)$.

• [Parseval equality]:

$$(r,s)_T \equiv \sum_{t:|t|\le T} r_t \overline{s_t} = \sum_{\mu\in\Gamma_T^d} (F_T r)(\mu)\overline{(F_T s)(\mu)} \equiv \langle F_T r, F_T s\rangle_T, \tag{52}$$

where $\overline{a}$ is the complex conjugate of $a \in \mathbb{C}$; in particular,

$$|r|_{T,2} = |r|^*_{T,2}; \tag{53}$$

A useful corollary of Parseval's equality combined with the fact that $|q|^*_{T,p} = |\bar{q}|^*_{T,p}$ is the relation

$$\left|\sum_{|t|\le T} a_t b_t\right| \le |a|^*_{T,1}|b|^*_{T,\infty}. \tag{54}$$

• [Norms of convolutions of filters]

$$r,s \in C(\mathbb{Z}^d) \Rightarrow |r(z_1,...,z_d)s(z_1,,...,z_d)|_p \le |r|_1|s|_p; \tag{55}$$

• [Relations between $|\cdot|$ and $|\cdot|^*$]: for $p,q \in [1,\infty]$ one has

$$|r|^*_{T,p} \le (2T+1)^{d[(1/p-1/2)_+ + (1/2-1/q)_+]}|r|_{T,q}, \quad a_+ = \max[a,0]; \tag{56}$$

$$\mathrm{ord}(r) + \mathrm{ord}(s) \le T \Rightarrow |r(z_1,...,z_d)s(z_1,...,z_d)|^*_{T,p} \le |r|_1|s|^*_{T,p}. \tag{57}$$

Useful fact. In the sequel, we need the following simple and well-known fact:

Lemma 23 *Let $f_j = \xi_j + i\eta_j$, $0 \le j < N$, be a sequence of N standard Gaussian complex-valued random variables, not necessarily independent of each other. Then*

$$\begin{array}{l} [E\{\max\limits_{0\le j<N} |f_j|^2\}]^{1/2} \le \sqrt{2\ln N + 2}; \\ P\{\max\limits_{0\le j<N} |f_j| > u + \sqrt{2\ln N}\} \le \exp\{-u^2/2\} \quad \forall u \ge 0. \end{array} \tag{58}$$

We have

$$\begin{array}{l} \psi(r) \equiv P\{\max\limits_{0\le j<N} |f_j| > r\} \le \min[1, N\exp\{-r^2/2\}] \Rightarrow \\ P\{\max\limits_{0\le j<N} |f_j| > u + \sqrt{2\ln N}\} \le N\exp\{-(u+\sqrt{2\ln N})^2/2\} \le \exp\{-u^2/2\}; \\ E\{\max\limits_{0\le j<N} |f_j|^2\} = -\int\limits_0^\infty r^2 d\psi(r) = 2\int\limits_0^\infty r\psi(r)dr \le 2\int\limits_0^{\sqrt{2\ln N}} rdr \\ +2N\int\limits_{\sqrt{2\ln N}} r\exp\{-r^2/2\}dr = 2\ln N + 2. \end{array}$$

7.2 Proof of Theorem 2

Preliminaries W.l.o.g., we may assume that $t = 0$. We denote by q^* the filter associated with (s_τ) via the description of the inclusion $(s_\tau) \in \mathbf{S}^0_{3T}(\theta, \rho, T)$. Let us set

$$|q^*|_2 = \widehat{\rho}(2T+1)^{-d/2}; \quad \kappa = \theta(2T+1)^{-d/2} \qquad [\widehat{\rho} \le \rho], \tag{59}$$

so that

$$\bar{s} = q^*(\Delta)s \Rightarrow \max_{\tau:|\tau|\le 3T} E\{|s_\tau - \bar{s}_\tau|^2\} \le \kappa^2. \tag{60}$$

Finally, let

$$\Theta_T = \max_{\tau:|\tau|\le 2T} |\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty}, \tag{61}$$

and let $\widehat{\phi}$ be the optimal solution, used in Algorithm A, of the optimization problem (6).

1^0. We start with simple technical lemma:

Lemma 24 *Let* $r(z_1,...,z_d) = (q^*(z_1,...,z_d))^2$*. Then* $r \in C_{2T}(\mathbf{Z}^d)$ *possesses the following properties:*

$$
\begin{aligned}
&|r|_2 \le |r|^*_{2T,1} \le 2^{d/2}\widehat{\rho}^2(2T+1)^{-d/2}; && (62)\\
&|r|_1 \le \widehat{\rho}^2; && (63)\\
&\Big[E\{|(1-r(\Delta))s|^2_{2T,2}\}\Big]^{1/2} \le \kappa(\widehat{\rho}+1)(4T+1)^{d/2}; && (64)\\
&|(1-r(\Delta))y|^*_{2T,\infty} \le |(1-r(\Delta))s|_{2T,2} + (1+\widehat{\rho}^2)\Theta_T && (65)\\
&\left[E\left\{\left(|(1-r(\Delta))y|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \le \sigma(1+\widehat{\rho}^2)\sqrt{4d\ln(4T+1)+2} && (66)\\
&\qquad\qquad +\kappa(\widehat{\rho}+1)(4T+1)^{d/2}. && \\
& && (67)
\end{aligned}
$$

Proof. (62): We have

$$
\begin{aligned}
|r|^*_{2T,1} &= \sum_{\mu\in\Gamma^d_{2T}} \tfrac{|r(\mu)|}{(4T+1)^{d/2}} = \sum_{\mu\in\Gamma^d_{2T}} \tfrac{|q^*(\mu)|^2}{(4T+1)^{d/2}} = (4T+1)^{d/2}\sum_{\mu\in\Gamma^d_{2T}} \left|\tfrac{q^*(\mu)}{(4T+1)^{d/2}}\right|^2\\
&= (4T+1)^{d/2}(|q^*|^*_{2T,2})^2 = (4T+1)^{d/2}|q^*|^2_{2T,2} \le 2^{d/2}\widehat{\rho}^2(2T+1)^{-d/2}.
\end{aligned}
$$

Since $|r|_2 = |r|_{2T,2} = |r|^*_{2T,2} \le |r|^*_{2T,1}$, (62) follows.
(63): We clearly have $|r|_1 \le |q^*|^2_1 \le ((2T+1)^{d/2}|q^*|_2)^2 = \widehat{\rho}^2$.
(64): Let $h = (1-q^*(\Delta))s$, so that by virtue of $(s_\tau) \in \mathbf{S}^0(\theta,\rho,T)$ and in view of the origin of q^* we have

$$\max_{\tau:|\tau|\le 3T} E\{|h_\tau|^2\} \le \kappa^2. \tag{68}$$

Setting $g = (1-r(\Delta))s$, we have

$$
\begin{aligned}
&g_\tau = ((1+q^*(\Delta))(1-q^*(\Delta))s)_\tau = ((1+q^*(\Delta))h)_\tau = h_\tau + (q^*(\Delta)h)_\tau\\
&\Rightarrow |g_\tau| \le |h_\tau| + |q^*|_2|\Delta_1^{-\tau_1}...\Delta_d^{-\tau_d}h|_{T,2}\\
&\Rightarrow (E\{|g_\tau|^2\})^{1/2} \le (E\{|h_\tau|^2\})^{1/2} + |q^*|_2\left(\sum_{\tau':|\tau'-\tau|\le T} E\{|h_{\tau-\tau'}|^2\}\right)^{1/2};
\end{aligned}
$$

applying (68) and taking into account that $|q^*|_2 = \widehat{\rho}(2T+1)^{-d/2}$, we come to

$$\max_{\tau:|\tau|\le 3T} E\{|((1-r(\Delta))s)_\tau|^2\} \le [\kappa(\widehat{\rho}+1)]^2, \tag{69}$$

and (64) follows.
(65), (66): We have

$$
\begin{aligned}
&|(1-r(\Delta))y|^*_{2T,\infty} \leq |(1-r(\Delta))s|^*_{2T,\infty} + |(1-r(\Delta))e|^*_{2T,\infty}\\
&\leq |(1-r(\Delta))s|^*_{2T,2} + |(1-r(\Delta))e|^*_{2T,\infty} = |(1-r(\Delta))s|_{2T,2} + |(1-r(\Delta))e|^*_{2T,\infty}\\
&\leq |(1-r(\Delta))s|_{2T,2} + |e|^*_{2T,\infty} + \sum_{\tau:|\tau|\leq 2T} |r_\tau||\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty}\\
&\leq |(1-r(\Delta))s|_{2T,2} + (1+|r|_1) \max_{\tau:|\tau|\leq 2T} |\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty}.
\end{aligned}
$$

The resulting inequality combines with (63) to yield (65). Further, from the resulting inequality and (64) it follows that

$$
\begin{aligned}
&\left(E\left\{\left(|(1-r(\Delta))y|^*_{2T,\infty}\right)^2\right\}\right)^{1/2}\\
&\leq \kappa(\widehat{\rho}+1)(4T+1)^{d/2} + (1+|r|_1)\Big(E\Big\{\underbrace{\left(\max_{\tau:|\tau|\leq 2T} |\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty}\right)^2}_{\Theta_T^2}\Big\}\Big)^{1/2}\\
&\leq \kappa(\widehat{\rho}+1)(4T+1)^{d/2} + (1+\widehat{\rho}^2)\left(E\{\Theta_T^2\}\right)^{1/2}
\end{aligned}
$$

(we have used (63)). To derive (66) from the resulting inequality, it remains to note that

$$
\left(E\{\Theta_T^2\}\right)^{1/2} \leq \sigma\sqrt{4d\ln(4T+1)+2}. \tag{70}
$$

Indeed, the coordinates of the Fourier transform of $\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e$ are, up to factor σ, standard complex-valued Gaussian random variables, so that $\sigma^{-2}\Theta_T^2$ is the maximum of squared modulae of $(4T+1)^{2d}$ of these variables; therefore $E\{\Theta_T^2\} \leq \sigma^2(4d\ln(4T+1)+2)$ by Lemma 23. □
2^0. We now study the properties of the solution $\widehat{\phi}$ of problem (6).

Lemma 25 *One has*

$$
|\widehat{\phi}|_{2T,2} \leq 2^{d/2}\rho^2(2T+1)^{-d/2}; \tag{71}
$$

$$
|(1-\widehat{\phi}(\Delta))e|^*_{2T,\infty} \leq (1+2^d\rho^2)\Theta_T; \tag{72}
$$

$$
\left[E\left\{\left(|(1-\widehat{\phi}(\Delta))e|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \leq \sigma(1+2^d\rho^2)\sqrt{4d\ln(4T+1)+2}; \tag{73}
$$

$$
|(1-\widehat{\phi}(\Delta))s|^*_{2T,\infty} \leq |(1-r(\Delta))s|_{2T,2} + 2(1+2^d\rho^2)\Theta_T; \tag{74}
$$

$$
\begin{aligned}
\left[E\left\{\left(|(1-\widehat{\phi}(\Delta))s|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \leq\ & 2\sigma(1+2^d\rho^2)\sqrt{4d\ln(4T+1)+2}\\
&+\kappa(\widehat{\rho}+1)(4T+1)^{d/2}.
\end{aligned} \tag{75}
$$

Proof. (71): $|\widehat{\phi}|_{2T,2} = |\widehat{\phi}|^*_{2T,2} \leq |\widehat{\phi}|^*_{2T,1} \leq 2^{d/2}\rho^2(2T+1)^{-d/2}$, (the concluding inequality comes from the fact that $\widehat{\phi}$ is feasible for (6)).
(72), (73): We have

$$
\begin{aligned}
&|(1-\widehat{\phi}(\Delta))e|^*_{2T,\infty} \leq (1+|\widehat{\phi}|_{2T,1}) \max_{\tau:|\tau|\leq 2T} |\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty} \\
&\leq \left(1+(4T+1)^{d/2}|\widehat{\phi}|_{2T,2}\right) \max_{\tau:|\tau|\leq 2T} |\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty} \\
&\leq (1+2^d\rho^2) \max_{\tau:|\tau|\leq 2T} |\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty}
\end{aligned}
$$

(we have used (71)). The resulting inequality implies that

$$
\begin{aligned}
&\left[E\left\{\left(|(1-\widehat{\phi}(\Delta))e|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \leq (1+2^d\rho^2)\left[E\left\{\max_{\tau:|\tau|\leq 2T}\left(|\Delta_1^{\tau_1}...\Delta_d^{\tau_d} e|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \\
&\leq (1+2^d\rho^2)\sigma\sqrt{4d\ln(4T+1)+2}
\end{aligned}
$$

(we have used (70)).
(75), (75): Note that the polynomial r defined in Lemma 24 is a feasible solution of the optimization problem (6) by the first relation in (62), so that the optimal value in the problem does not exceed $J(r, y^0_{4T})$. It follows that

$$
\begin{aligned}
&(a)\ J(\widehat{\phi}, y^0_{4T}) \leq J(r, y^0_{4T}) \\
\Rightarrow\ &(b)\ |(1-\widehat{\phi}(\Delta))y|^*_{2T,\infty} \leq |(1-r(\Delta))y|^*_{2T,\infty} \\
\Rightarrow\ &(c)\ |(1-\widehat{\phi}(\Delta))s|^*_{2T,\infty} \leq |(1-\widehat{\phi}(\Delta))e|^*_{2T,\infty} + |(1-r(\Delta))y|^*_{2T,\infty} \\
\Rightarrow\ &(d)\ \left[E\left\{\left(|(1-\widehat{\phi}(\Delta))s|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \leq \left[E\left\{\left(|(1-\widehat{\phi}(\Delta))e|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \\
&\qquad + \left[E\left\{\left(|(1-r(\Delta))y|^*_{2T,\infty}\right)^2\right\}\right]^{1/2}
\end{aligned}
$$

Relation (74) follows from (c) combined with (65) and (72) (recall that $\widehat{\rho} \leq \rho$). Relation (75) follows from (d) combined with (73) and (66). □
$\mathbf{3^0}$**.** Our next step is to prove

Lemma 26 *One has*

$$
\begin{aligned}
\left|\left((1-r(\Delta))(1-\widehat{\phi}(\Delta))s\right)_0\right| &\leq |((1-r(\Delta))s)_0| \\
&\quad + 2^{d/2}\rho^2(2T+1)^{-d/2}\,|(1-r(\Delta))s|_{2T,2}\,;
\end{aligned} \tag{76}
$$

$$
\left[E\left\{\left|\left((1-r(\Delta))(1-\widehat{\phi}(\Delta))s\right)_0\right|^2\right\}\right]^{1/2} \leq \kappa(\widehat{\rho}+1)(2^d\rho^2+1). \tag{77}
$$

Proof. We have

$$
\begin{aligned}
&\left|\Big((1-r(\Delta))(1-\widehat{\phi}(\Delta))s\Big)_0\right| \le |((1-r(\Delta))s)_0| + \left|\Big(\widehat{\phi}(\Delta)(1-r(\Delta))s\Big)_0\right| \\
&\le |((1-r(\Delta))s)_0| + |\widehat{\phi}|_{2T,2}\,|(1-r(\Delta))s|_{2T,2} \\
&\le |((1-r(\Delta))s)_0| + 2^{d/2}\rho^2(2T+1)^{-d/2}\,|(1-r(\Delta))s|_{2T,2} \qquad \text{[see (71)]}
\end{aligned}
$$

as required in (76). From the resulting inequality it follows that

$$
\begin{aligned}
&\left[E\left\{\left|\Big((1-r(\Delta))(1-\widehat{\phi}(\Delta))s\Big)_0\right|^2\right\}\right]^{1/2} \le \left[E\left\{|((1-r(\Delta))s)_0|^2\right\}\right]^{1/2} \\
&+2^{d/2}\rho^2(2T+1)^{-d/2}\left[E\left\{|(1-r(\Delta))s|^2_{2T,2}\right\}\right]^{1/2} \\
&\le \kappa(\widehat{\rho}+1) + 2^{d/2}\rho^2(2T+1)^{-d/2}\left[E\left\{|(1-r(\Delta))s|^2_{2T,2}\right\}\right]^{1/2} \qquad \text{[see (69)]} \\
&\le \kappa(\widehat{\rho}+1) + 2^{d/2}\rho^2(2T+1)^{-d/2}\kappa(\widehat{\rho}+1)(4T+1)^{d/2} \qquad \text{[see (64)]}
\end{aligned}
$$

and (77) follows. □

4^0. Now we are able to complete the proof of Theorem 2. The error of the estimate $\widehat{s}$ at the point $t=0$ is

$$
\begin{aligned}
&s_0 - \widehat{s}_0 = s_0 - (\widehat{\phi}(\Delta)y)_0 = \Big((1-\widehat{\phi}(\Delta))s\Big)_0 - \Big(\widehat{\phi}(\Delta)e\Big)_0 \equiv \epsilon_0^{(1)} + \epsilon_0^{(2)}, \\
&\epsilon_\tau^{(1)} = \Big((1-\widehat{\phi}(\Delta))s\Big)_\tau, \quad \epsilon_\tau^{(2)} = \Big(\widehat{\phi}(\Delta)e\Big)_\tau.
\end{aligned} \tag{78}
$$

Setting $f_\tau = \overline{e_{-\tau}}$, we have

$$
\begin{aligned}
|\epsilon_0^{(2)}| &= \left|\sum_{\tau:|\tau|\le 2T} \widehat{\phi}_\tau e_{-\tau}\right| \le |\widehat{\phi}|^*_{2T,1}|f|^*_{2T,\infty} && \text{[see (54)]} \\
&\le 2^{d/2}\rho^2(2T+1)^{-d/2}|f|^*_{2T,\infty}, && \text{[since } \widehat{\phi} \text{ is feasible for (6)]}
\end{aligned}
$$

whence, by definition of Θ_T,

$$
|\epsilon_0^{(2)}| \le 2^{d/2}\rho^2(2T+1)^{-d/2}\Theta_T. \tag{79}
$$

Applying (70), we derive from the latter inequality that

$$
\left[E\left\{|\epsilon_0^{(2)}|^2\right\}\right]^{1/2} \le 2^{d/2}\sigma\rho^2(2T+1)^{-d/2}\sqrt{2d\ln(4T+1)+2}. \tag{80}
$$

We further have

$$
\begin{array}{l}
|\epsilon_0^{(1)}| = \left|\Big((1-\widehat{\phi}(\Delta))s\Big)_0\right| \leq \left|\Big(r(\Delta)(1-\widehat{\phi}(\Delta))s\Big)_0\right| \\
+\left|\Big((1-r(\Delta))(1-\widehat{\phi}(\Delta))s\Big)_0\right| \underbrace{\leq}_{a} |r|^*_{2T,1}|(1-\widehat{\phi}(\Delta))s|^*_{2T,\infty} \\
+\left|\Big((1-r(\Delta))(1-\widehat{\phi}(\Delta))s\Big)_0\right| \underbrace{\leq}_{b} 2^{d/2}\rho^2(2T+1)^{-d/2}|(1-\widehat{\phi}(\Delta))s|^*_{2T,\infty} \\
+\left|\Big((1-r(\Delta))(1-\widehat{\phi}(\Delta))s\Big)_0\right|
\end{array} \tag{81}
$$

(a is given by (54), b by feasibility of $\widehat{\phi}$ for (6)), whence

$$
\begin{array}{l}
\left[E\left\{|\epsilon_0^{(1)}|^2\right\}\right]^{1/2} \leq 2^{d/2}\rho^2(2T+1)^{-d/2}\left[E\left\{\left(\left|(1-\widehat{\phi}(\Delta))s\right|^*_{2T,\infty}\right)^2\right\}\right]^{1/2} \\
+\left[E\left\{\left|\Big((1-r(\Delta)(1-\widehat{\phi}(\Delta))s\Big)_0\right|^2\right\}\right]^{1/2} \\
\leq 2^{d/2}\rho^2(2T+1)^{-d/2}\Big[2^d\sigma(1+2^d\rho^2)\sqrt{4d\ln(4T+1)+2} \\
+\kappa(\widehat{\rho}+1)(4T+1)^{d/2}\Big] + \kappa(\widehat{\rho}+1)(2^d\rho^2+1)
\end{array} \tag{82}
$$

(see (75), (77)). Combining (78), (80), (82), we finally get

$$
\begin{array}{l}
\left[E\left\{|s_0-\widehat{s}_0|^2\right\}\right]^{1/2} \leq 2^{d/2}\sigma\rho^2(2T+1)^{-d/2}\sqrt{2d\ln(4T+1)+2} \\
+2^{d/2}\rho^2(2T+1)^{-d/2}\Big[2^d\sigma(1+2^d\rho^2)\sqrt{4d\ln(4T+1)+2} \\
+\kappa(\widehat{\rho}+1)(4T+1)^{d/2}\Big] + \kappa(\widehat{\rho}+1)(2^d\rho^2+1).
\end{array} \tag{83}
$$

Recalling that $\widehat{\rho} \leq \rho$, $\kappa = \theta(2T+1)^{-d/2}$ and that $\rho \geq 1$, (7) follows.
Now assume that (s) is deterministic. In this case, from (81) combined with (74) and (76) implies that

$$
\begin{array}{l}
|\epsilon_0^{(1)}| \leq 2^{1+d/2}\rho^2(2T+1)^{-d/2}|(1-r(\Delta))s|_{2T,2} \\
+2^{1+d/2}\rho^2(1+2^d\rho^2)(2T+1)^{-d/2}\Theta_T + \left|((1-r(\Delta))s)_0\right|,
\end{array} \tag{84}
$$

while from (64), (69) it follows that

$$
\begin{array}{l}
|(1-r(\Delta))s|_{2T,2} \leq \kappa(\widehat{\rho}+1)(4T+1)^{d/2} \leq 2^{d/2}\theta(1+\rho), \\
|((1-r(\Delta))s)_0| \leq \kappa(1+\widehat{\rho}) \leq \theta(1+\rho)(2T+1)^{-d/2}.
\end{array} \tag{85}
$$

Therefore (84) implies that

$$|\epsilon_0^{(1)}| \le 3^{3+d}\rho^3 \left[\theta + \rho\Theta_T\right](2T+1)^{-d/2}. \tag{86}$$

Combining this relation with (79) and (78), we arrive at (8). $\square$

7.3 Proof of Theorem 3

Consider the following construction. Let us fix an integer $m \le (2T+1)^\alpha$, $\alpha < 1/2$. For the sake of simplicity we suppose that $2T+1 = lm$ for some $l \in \mathbb{N}$. Note that $l \ge (2T+1)^{1-\alpha}$. We set

$$\beta = 2\sigma\sqrt{(1-2\alpha)\ln(2T-1)}. \tag{87}$$

Let $r = (r_1, ..., r_m)^T$ be a vector with integer components such that each r_j, $j = 1, .., m$, takes value in $[l(j-1)+1, lj]$. Consider now the signals $s_\tau^{(r)} = \frac{\beta}{\sqrt{2T+1}}\sum_{j=1}^m \exp\left(\frac{2\pi r_j}{2T+1}\tau\right)$, $\tau = -4T-1, ...4T+2$. Note that

$$s^{(r)}(0) = \frac{m\beta}{\sqrt{2T+1}}, \quad \text{for any } r. \tag{88}$$

We consider the family $\mathbf{F}_m^T$ which contains signals $(s_\tau^{(r)})$ along with $(s_\tau^{(0)}) \equiv 0$ (we set $r = 0$ in this case). Suppose now that a signal $s \in \mathbf{F}_m^T$ is observed in the noise:

$$y_\tau = s_\tau + \sigma e_\tau, \quad \tau = -4T-1, ..., 4T+2, \tag{89}$$

where (e_τ) is a sequence of i.i.d. standard complex-valued Gaussian random variables.

Lemma 27 *For each $s^{(r)} \in \mathbf{F}_m^T$ there exists a linear filter $q^* \in C_T(\mathbf{Z})$ such that $|q^*|_2 = \sqrt{\frac{m}{2T+1}}$ and $\left[E\left\{(q^*(\Delta)y)_\tau - s_\tau^{(r)}\right\}^2\right]^{1/2} \le |q^*|_2\sigma$ for $-3T-1 \le \tau \le 3T+2$.*

Proof: One can easily verify that the requirements are satisfied by the filter q_r^* given by $q_r^*(k) = \frac{1}{2T+1}\sum_{j=1}^m \exp\left(\frac{2\pi r_j}{2T+1}k\right)$. $\square$
Our objective now is to show that that given the observations (89), one cannot construct an estimate $\widehat{s}$ of $s_0^{(r)}$ such that its quadratic risk is less than

$$c_0^2\frac{\ln(2T+1)}{2T+1}\sigma^2 m^2 \ge c_1\sigma^2|q^*|_2^4\ln(2T+1).$$

uniformly over $\mathbf{F}_m^T$. To this end we introduce a related Bayesian estimation problem on $\mathbf{F}_m^T$, obtained from (89) as follows: let F_{8T+4} be the Fourier transform on $[-4T-1, 4T+2]$. When applied to $(s_\tau^{(r)})$ it gives

$$
\begin{aligned}
(F_{8T+4}s^{(r)})_k &= \frac{1}{2\sqrt{2T+1}} \sum_{\tau=-4T-1}^{4T+2} s_\tau^{(r)} \exp\left(-\frac{2\pi i k\tau}{8T+4}\right) \\
&= \frac{\beta}{2(2T+1)} \sum_{\tau=-4T-1}^{4T+2} \sum_{j=1}^{m} \exp\left(\frac{2\pi i r_j \tau}{2T+1}\right) \exp\left(-\frac{2\pi i k\tau}{8T+4}\right) = \frac{\beta}{2}\delta_{k,4r_j},
\end{aligned} \tag{90}
$$

where $\delta_{j,k} = \begin{cases} 0, \; j \neq k \\ 1, \; j = k \end{cases}$ is the Kronecker symbol. On the other hand,

$$
z_j = (F_{8T+4}y)_j = (F_{8T+4}s^{(r)})_j + \sigma\xi_j, \tag{91}
$$

where ξ_j are orthogonal (and thus independent) standard complex-valued Gaussian random variables.
Let us now construct an a priori probability measure on $\mathbf{F}_m^T$. Let $\zeta \in \{0,1\}$ be a random variable such that $P(\zeta = 1) = P(\zeta = 0) = 1/2$. We set $r = 0$ if $\zeta = 0$ and if not, the entries r_j, $j = 1, ..., m$ of r are mutually independent (and independent of (ξ_k)) and uniformly distributed over $[\ell(j-1)+1, \ell j]$. We denote P_B the corresponding probability on $\mathbf{F}^T$.
Let the parameter θ take the value $\theta_0 = 0$ if $s \equiv 0$ is realized and $\theta_r = m\beta$ if $s = s^{(r)}$ for some $r \neq 0$. Due to (88) on can easily see that in order to obtain bound (10) for the original problem it suffices to show the following lower bound for the Bayesian risk of estimation of the parameter θ from observations (91):

$$
r(T,m) \equiv \inf_{\widehat{\theta}} E_B\left\{E_r\left\{(\widehat{\theta}-\theta_r)^2\right\}\right\} \geq c_2\sigma^2 m^2 \ln(2T+1) \tag{92}
$$

for some $c_2 > 0$ (here $E_B\{\cdot\}$ stands for the expectation with respect to the probability P_B).
We proceed as follows. Note that

$$
\begin{aligned}
&r(T,m) = \tfrac{1}{2}E_B\left\{E_r\left\{(\widehat{\theta}-\theta_r)^2\right\} + E_0\left\{\widehat{\theta}^2\right\}\right\} \\
&= \tfrac{1}{2}E_0\left\{E_B\left\{(\widehat{\theta}-\theta_r)^2 Z_r\right\} + \widehat{\theta}^2\right\},
\end{aligned} \tag{93}
$$

where $Z_r = \left|\frac{dP_r}{dP_0}\right|$ is the likelihood ratio. Observe now that $\theta^* = \frac{E_B\theta_r Z_r}{1+E_B Z_r}$ is the minimizer of $r(T,m)$ with respect to $\widehat{\theta}$. When substituting $\widehat{\theta} = \theta^*$ into (93) we obtain

$$2r(T,m) \geq E_0 \left\{ E_B\{\theta_r^2 Z_r\} - \frac{(E_B\{\theta_r Z_r\})^2}{1+E_B\{Z_r\}} \right\} \geq E_0 \left\{ \frac{E_B\{\theta_r^2 Z_r\}}{1+E_B\{Z_r\}} \right\}$$
$$= \beta^2 m^2 E_0 \left\{ \frac{E_B\{Z_r\}}{1+E_B\{Z_r\}} \right\}. \qquad (94)$$

Now let us study the properties of the likelihood ratio Z_r. We have the following simple lemma:

Lemma 28 *For T large enough one has $P_0\{\frac{1}{2} \leq E_B\{Z_r\}s \leq 4e\} \geq \frac{1}{16e}$.*

Proof. Let $\eta_k = \Re(\xi_k)$. Note first that $Z_r = \prod_{j=1}^m \exp\left(\frac{\eta_{4r_j}\beta}{2\sigma} + \frac{\beta^2}{8\sigma^2}\right)$ since the Fourier transform of $s^{(r)}$ is real (cf. (90)). Then

$$E_B\{Z_r\} = \prod_{j=1}^m \sum_{k=\ell(j-1)+1}^{\ell j} \exp\left(\frac{\eta_{4k}\beta}{2\sigma} - \frac{\beta^2}{8\sigma^2}\right)\frac{1}{\ell}.$$

By evident reasons, $E_0\{E_B\{Z_r\}\} = 1$. On the other hand,

$$E_0\left\{(E_B\{Z_r\})^2\right\} = \prod_{j=1}^m E_0\left\{\left[\sum_{k=\ell(j-1)+1}^{\ell j} \exp\left(\frac{\eta_{4k}\beta}{2\sigma} - \frac{\beta^2}{8\sigma^2}\right)\frac{1}{\ell}\right]^2\right\}$$
$$= \prod_{j=1}^m I_j. \qquad (95)$$

Let us estimate a term in the product above:

$$I_j = (1-\ell^{-1}) + \ell^{-2} \sum_{k=\ell(j-1)+1}^{\ell j} E_0\left\{\exp\left(\frac{\eta_{4k}\beta}{\sigma} - \frac{\beta^2}{4\sigma^2}\right)\right\}$$
$$= 1 + \ell^{-2} \sum_{k=\ell(j-1)+1}^{\ell j} \left(\exp\left(\frac{\beta^2}{4\sigma}\right) - 1\right) = 1 + \ell^{-1}\left(\exp\left(\frac{\beta^2}{4\sigma^2}\right) - 1\right)$$
$$\leq \exp\left[\ell^{-1}\exp\left(\frac{\beta^2}{4\sigma^2}\right)\right].$$

When substituting this bound into (95) we get

$$E_0\left\{(E_B\{Z_r\})^2\right\} \leq \prod_{j=1}^m \exp\left(\frac{1}{l}\exp\{\frac{\beta^2}{4\sigma^2}\}\right) = \exp\left(\frac{2T+1}{\ell^2}\exp\{\frac{\beta^2}{4\sigma^2}\}\right).$$

Due to the choice of β in (87), $\frac{2T+1}{\ell^2}\exp\{\frac{\beta^2}{4\sigma^2}\} \leq 1$, and so $E_0\left\{(E_B\{Z_r\})^2\right\} \leq e$ for all large enough values of T.
Now let a positive random variable X satisfy $E\{X\} = 1$, $E\{X^2\} < \infty$. Then for every $R > 1$ one has $\int\limits_0^{RE\{X^2\}} x\mu(dx) \geq 1 - 1/R$ and $\int\limits_{1/2}^{RE\{X^2\}} x\mu(dx) \geq$

$1 - \frac{1}{R} - \int\limits_0^{1/2} x\mu(dx) \geq \frac{1}{2} - \frac{1}{R}$. We conclude that $\mu\left\{x : \frac{1}{2} \leq x \leq RE\{X^2\}\right\} \geq \frac{\frac{1}{2}-\frac{1}{R}}{RE\{X^2\}}$. For $R = 4$ we get $\mu\left\{x : \frac{1}{2} \leq x \leq 4E\{X^2\}\right\} \geq \frac{1}{16E\{X^2\}}$. Specifying here $X = E_B\{Z_r\}$, we obtain $P_0\left\{\frac{1}{2} \leq E_B\{Z_r\} \leq 4e\right\} \geq \frac{1}{16e}$. □
Let $A = \{\omega : \ 1/2 \leq E_B\{Z_r\} \leq 4e\}$. By Lemma 28, relation (94) implies that

$$\begin{aligned}
&2r(T,m) \geq \beta^2 m^2 E_0\left\{\frac{E_B\{Z_r\}}{1+E_B Z_r} 1_A\right\} \geq \frac{\beta^2 m^2}{2(1+4e)} P_0\{A\} \geq \frac{\beta^2 m^2}{32e(1+4e)} \\
&\geq c_2 \sigma^2 m^2 \ln(2T+1),
\end{aligned}$$

as required in (92). □

7.4 Proof of Proposition 4

The case of $m = 1$ is evident. Now let $m \geq 2$, let T^+ be an integer, $0 \leq T^+ \leq L^+$, and let $T = \lfloor m^{-1} T^+ \rfloor$. Since $s^j \in \mathbf{F}^t_L(\theta,\rho)$ and clearly $T \leq L$, there exist filters q^j such that

$$\begin{aligned}
&(a) : \mathrm{ord}(q^j) \leq T; \ (b) : |q^j|_2 \leq \rho_j (2T+1)^{-d/2}; \\
&(c) : |q^j|_1 = |q^j|_{T,1} \leq (2T+1)^{d/2} |q^j|_2 \leq \rho_j; \\
&(d) : \left[E\left\{|s^j_\tau - (q^j(\Delta)s^j)_\tau|^2\right\}\right]^{1/2} \leq \theta_j (2T+1)^{-d/2} \quad \forall (\tau : |\tau - t| \leq L).
\end{aligned} \tag{96}$$

Now let filter q be defined by

$$1 - q(z) = \prod_{j=1}^m (1 - q^j(z)), \ z = (z_1, ..., z_d).$$

Observe that

$$\mathrm{ord}(q) \leq mT \leq T^+. \tag{97}$$

Note that

$$|q|_2 \leq 2^m \rho_1 ... \rho_m (2T+1)^{-d/2} \leq (2m-1)^{d/2} 2^m \rho_1 ... \rho_m (2T^+ + 1)^{-d/2}. \tag{98}$$

Indeed, we clearly have

$$
\begin{array}{l}
|q(z)|_2 = \left|\sum\limits_{\ell=1}^m (-1)^{\ell+1} \sum\limits_{1\leq j_1<j_2<...<j_\ell\leq m} q^{j_1}(z)q^{j_2}(z)...q^{j_\ell}(z)\right|_2 \\
\leq \sum\limits_{\ell=1}^m \sum\limits_{1\leq j_1<j_2<...<j_\ell\leq m} |q^{j_1}(z)q^{j_2}(z)...q^{j_\ell}(z)|_2 \underbrace{\leq}_{a} \sum\limits_{\ell=1}^m \sum\limits_{1\leq j_1<j_2<...<j_\ell\leq m} \frac{\rho_{j_1}\rho_{j_2}...\rho_{j_\ell}}{(2T+1)^{d/2}} \\
\leq [(1+\rho_1)...(1+\rho_m)-1](2T+1)^{-d/2} \underbrace{\leq}_{b} 2^m\rho_1...\rho_m(2T+1)^{-d/2}
\end{array}
$$

(a is by $(96.b-c)$ since $|u(z)v(z)|_2 \leq |u|_1|v|_2$, $|u(z)v(z)|_1 \leq |u|_1|v|_1]$, b is due to $\rho_j \geq 1$), as required in (98). Further, by (96.c), for the filters

$$
Q^j(z) = \left(\prod_{\ell=1}^{j-1}(1-q^\ell(z))\right)\left(\prod_{\ell=j+1}^{m}(1-q^\ell(z))\right)
$$

one has

$$
|Q^j|_1 \leq (1+\rho_1)...(1+\rho_{j-1})(1+\rho_{j+1})...(1+\rho_m) \leq \frac{2^{m-1}\rho_1...\rho_m}{\rho_j}. \tag{99}
$$

Now let $\tau \in \mathbf{Z}^d$ be such that $|\tau - t| \leq L^+$. We have

$$
\begin{array}{l}
\left[E\left\{|(1-q(\Delta))s)_\tau|^2\right\}\right]^{1/2} = \left[E\left\{\left|\sum\limits_{j=1}^m \lambda_j\,(1-q(\Delta))s^j)_\tau\right|^2\right\}\right]^{1/2} \\
\leq \sum\limits_{j=1}^m \left[E\left\{|\lambda_j\,(1-q(\Delta))s^j)_\tau|^2\right\}\right]^{1/2} \underbrace{=}_{a} \sum\limits_{j=1}^m [E\{|\lambda_j|^2\}]^{1/2}\left[E\left\{|(1-q(\Delta))s^j)_\tau|^2\right\}\right]^{1/2} \\
\leq \sum\limits_{j=1}^m [E\{|\lambda_j|^2\}]^{1/2}\left[E\left\{|(Q^j(\Delta)(1-q^j(\Delta))s^j)_\tau|^2\right\}\right]^{1/2} \\
\underbrace{\leq}_{b} \sum\limits_{j=1}^m [E\{|\lambda_j|^2\}]^{1/2}|Q^j|_1 \max\limits_{\tau':|\tau'-\tau|\leq(m-1)T}\left[E\left\{|((1-q^j(\Delta))s^j)_{\tau'}|^2\right\}\right]^{1/2} \\
\leq 2^{m-1}\rho_1...\rho_m(2T+1)^{-d/2}\sum\limits_{j=1}^m \frac{\theta_j[E\{|\lambda_j|^2\}]^{1/2}}{\rho_j} \\
\leq \left[(2m-1)^{d/2}2^{m-1}\rho_1...\rho_m\sum\limits_{j=1}^m \frac{\theta_j[E\{|\lambda_j|^2\}]^{1/2}}{\rho_j}\right](2T^++1)^{-d/2}
\end{array}
$$

where a is due to independence of λ_j and (s^j) and b is by (99), (96.d) and since $|\tau'-\tau| \leq (m-1)T, |\tau - t| \leq L^+ \Rightarrow |\tau' - t| \leq L^+ + T^+ \leq L$. Combining the resulting inequality, (97), (98) and taking into account that $T^+ \in \{0, 1, ..., L^+\}$ is arbitrary, we conclude that $s \in \mathbf{F}^t_{L^+}(\theta^+, \rho^+)$. Note that by construction, the filters certifying the latter inclusion are independent of λ_j. $\square$

7.5 Proof of Proposition 5

(i): Let $T \leq L$, and let q be such that

$$\begin{array}{l} \mathrm{ord}(q) \leq T,\ |q|_2 \leq \frac{\rho}{(2T+1)^{d/2}}, \\ \max\limits_{\tau:|\tau-t|\leq L} \left[E\left\{|((1-q(\Delta))s)_\tau|^2\right\}\right]^{1/2} \leq \frac{\theta}{(2T+1)^{d/2}}. \end{array} \tag{100}$$

Let us set $\widehat{q}_\tau = \exp\{i\omega^T\tau\}q_\tau,\ \tau \in \mathbf{Z}^d$. Then $\mathrm{ord}(\widehat{q}) \leq T$, $|\widehat{q}|_2 = |q|_2$ and

$$\begin{array}{rl} ((1-\widehat{q}(\Delta))\widehat{s})_\tau = & \exp\{i[\omega^T\tau + \phi]\}s_\tau \\ & - \sum\limits_{\tau'}(\exp\{i\omega^T\tau'\}q_{\tau'})(\exp\{i[\omega^T(\tau-\tau') + \phi]\}s_{\tau-\tau'}) \\ = & \exp\{i[\omega^T\tau + \phi]\}\left((1-q(\Delta))s\right)_\tau, \end{array}$$

so that (100) remains valid when $q, (s)$ are replaced with $\widehat{q}, (\widehat{s})$. Thus, $(\widehat{s}) \in \mathbf{F}^t_L(\theta,\rho)$. (i) is proved; (ii) is evident. $\square$

7.6 Proof of Proposition 6

Let $T \leq L$, and let $q = (q_\tau)_{\tau\in\mathbf{Z}^d}$ be such that $\mathrm{ord}(q) \leq T$, $|q|_2 \leq \rho(2T+1)^{-d/2}$,

$$\left[E\left\{|((1-q(\Delta))s)_\tau|^2\right\}\right]^{1/2} \leq \theta(2T+1)^{-d/2} \quad \forall(\tau \in \mathbf{Z}^d : |\tau - t| \leq L).$$

Setting $q^+_{\tau_1,\dots,\tau_{d^+}} = (2T+1)^{-(d^+-d)}q_{\tau_1,\dots,\tau_d}$, we clearly have $\mathrm{ord}(q^+) \leq T$, $|q^+|_2 \leq \rho(2T+1)^{-d^+/2}$ and

$$\left[E\left\{\left|\left((1-q^+(\Delta))s^+\right)_\tau\right|^2\right\}\right]^{1/2} \leq \theta(2T+1)^{-d/2} \quad \forall(\tau \in \mathbf{Z}^{d^+} : |\tau - t^+| \leq L).$$

It remains to note that $\theta(2T+1)^{-d/2} \leq \theta^+(2T+1)^{-d^+/2}$ for $0 \leq T \leq L$. $\square$

7.7 Proof of Proposition 7

Let $T \leq L$, and let $q' \in C_T(\mathbf{Z}^{d'})$, $q'' \in C_T(\mathbf{Z}^{d''})$ be such that

$$\begin{array}{l} (a): |q'|_2 \leq \rho'(2T+1)^{-d'/2},\ |q''|_2 \leq \rho''(2T+1)^{-d''/2}, \\ s'_{\tau'}(\xi) = \sum\limits_{\nu'} s'_{\tau'-\nu'}q'_{\nu'},\ |\tau' - t'| \leq L, \\ s''_{\tau''}(\xi) = \sum\limits_{\nu''} s''_{\tau''-\nu''}q''_{\nu''},\ |\tau'' - t''| \leq L \end{array} \quad . \tag{101}$$

Let $q(z_1,...,z_d)=q'(z_1,...,z_{d'})q''(z_{d'+1},...,z_d)$, so that

$$q\in C_T(\mathbf{Z}^d),\quad |q|_2=|q'|_2|q''|_2\le\rho'\rho''(2T+1)^{-d/2} \tag{102}$$

(see (101.a)). Now let $\tau=(\tau',\tau'')$ be such that $|\tau-(t',t'')|\le L$. We have

$$\begin{array}{l}(q(\Delta)s(\xi))_\tau=\sum\limits_{(\nu',\nu'')\in\mathbf{Z}^{d'}\times\mathbf{Z}^{d''}} s'_{\tau'-\nu'}(\xi)s''_{\tau''-\nu''}(\xi)q'_{\nu'}q''_{\nu''}=\sum\limits_{\nu'\in\mathbf{Z}^{d'}} s'_{\tau'-\nu'}q'_{\nu'}s''_{\tau''}(\xi)\\ =s'_{\tau'}(\xi)s''_{\tau''}(\xi)=s_{(\tau',\tau'')},\end{array}$$

which combines with (102) to yield that $(s_\tau)\in\mathbf{F}_L^{(t',t'')}(0,\rho'\rho'')$. $\square$

7.8 Proof of Proposition 8

We start with the following two evident facts:

Lemma 29 *Let $(s^j)\in C(\mathbf{Z}^d)$ be deterministic fields belonging to $\mathbf{F}_L^t(\theta,\rho)$, $j=1,2,...$ such that $s^j_\tau\to s_\tau$, $j\to\infty$, for every $\tau\in\mathbf{Z}^d$. Then $(s)\in\mathbf{F}_L^t(\theta,\rho)$.*

Indeed, for every T, $0\le T\le L$, the filters $q^{j,T}\in C_T(\mathbf{Z}^d)$ which certify the inclusions $(s^j)\in\mathbf{F}_L^t(\theta,\rho)$ satisfy $|q^{j,T}|_2\le\rho(2T+1)^{-d/2}$ and therefore have a limiting point $q^T\in C_T(\mathbf{Z}^d)$ with $|q^j|_2\le\rho(2T+1)^{-d/2}$. The filters $\{q^T\}_{0\le T\le L}$ clearly certify the inclusion $(s)\in\mathbf{F}_L^t(\theta,\rho)$. $\square$

Lemma 30 *For every $t\in\mathbf{Z}$, the univariate exponential field $(s_\tau=\exp\{\omega\tau\})$, $\omega\in\mathbb{C}$, belongs to $\mathbf{F}_\infty^t(0,\sqrt{2})$.*

Indeed, assuming $\Re(\omega)\ge0$ and given $T\ge0$, let us set $q(z)=\frac{1}{T+1}[1+\exp\{-\omega\}z^{-1}+\exp\{-2\omega\}z^{-2}+...+\exp\{-T\omega\}z^{-T}]$. Then $q\in C_T(\mathbf{Z})$, $|q|_2=(T+1)^{-1/2}\le2^{1/2}(2T+1)^{-1/2}$, while clearly $q(\Delta)s\equiv s$. In the case of $\Re(\omega)<0$, the same reasoning holds true for $q(z)=\frac{1}{T+1}[1+\exp\{\omega\}z+\exp\{2\omega\}z^2+...+\exp\{T\omega\}z^T]$. $\square$
To complete the proof, we need the following fact:

Lemma 31 *Let (s_τ) be a "simple" exponential polynomial – a deterministic exponential polynomial of the form $(s_\tau)=\sum\limits_{\ell=1}^{M}c_\ell\exp\{\omega^T(\ell)\tau\}$. Then*

$$\forall t\in\mathbf{Z}^d:(s_\tau)\in\mathbf{F}_\infty^t(0,\rho_d(N_1,...,N_d)), \tag{103}$$

where $\rho_d(\cdot,...,\cdot)$ is given by (13) and $N_1,...,N_d$ are the partial sizes of the polynomial. Besides this, the filters $q^{(T)}$ certifying the above inclusion can be *chosen to depend solely on T and on the collection of the d sets $\mathbf{O}_j = \{\omega_j(\ell) : \ell = 1, ..., M\}$.*

Lemma 31 ⇒ Proposition 8: Assume first that the coefficients c_ℓ in (12) are deterministic. Since every one of the univariate functions $f(t) = t^k$, $0 \le k \le m$, is, uniformly on compact sets, the limit, as $\epsilon \to +0$, of appropriate linear combinations of the $m+1$ exponents $\exp\{-k\epsilon\}$, the exponential polynomial (12) is the pointwise, on $\mathbf{Z}^d$, limit, as $i \to \infty$, of simple exponential polynomials (s^i_τ) with extended sets of "frequencies" $\{\omega_j(\ell)\}_{j,\ell}$: in the approximating polynomials, every one of these frequencies is replaced by (m_j+1) frequencies $\omega_j(\ell) - k\epsilon_i$, $0 \le k \le m_j$. Note that by the definition of partial sizes of exponential polynomials, the approximating polynomials have exactly the same partial sizes as the original polynomial (s_τ). Combining Lemmas 31 and (29), we immediately conclude that the exponential polynomial (12) belongs to $\mathbf{F}^t_\infty(0, \rho_d(N_1, ..., N_d))$. Since the filters $q^{(T),i}$ certifying well-filterability of the approximating polynomials (s^i_τ) can be chosen to depend solely on T and the sets of partial frequencies of these approximating polynomials, from the proof of Lemma 29 it follows that the filters $q^{(T)}$ certifying the inclusion $(s_\tau) \in \mathbf{F}^t_\infty(0, \rho_d(N_1, ..., N_d))$ can be chosen to depend solely on T and the sets of partial frequencies of (s_τ), as required in Proposition 8. We have proved Proposition 8 for the case of a deterministic exponential polynomial; since the filters certifying well-filterability of such a polynomial are independent of the coefficients c_ℓ, the result is valid for random polynomials as well. □

Proof of Lemma 31. Proof is by induction in d.

Base $d = 1$ is readily given by Lemma 30 combined with Proposition 4.

Step $1 \le d \Rightarrow d+1$: Let $s_\tau = \sum_\ell c_\ell \exp\{\omega^T(\ell)\tau\}$ be a simple exponential polynomial on $\mathbf{Z}^{d+1}$ with partial sizes N_j and the sets of partial frequencies $\mathbf{O}_j$, $j = 1, ..., N$. Let $T \ge 0$, and let $t \in \mathbf{Z}^{d+1}$. By the inductive hypothesis, there exist filters $g^{(T)} \in C_T(\mathbf{Z}^d)$, $h^{(T)} \in C_T(\mathbf{Z})$ (depending solely on T and on $\mathbf{O}_1$,...,$\mathbf{O}_{d+1}$) such that

$$
\begin{aligned}
&(a) : |g^{(T)}|_2 \le \rho_d(N_1, ..., N_d)(2T+1)^{-d/2},\\
&(a') : |h^{(T)}|_2 \le \rho_1(N_{d+1})(2T+1)^{-1/2},\\
&(b) : r_\tau = \sum_{\nu \in \mathbf{Z}^d} r_{\tau-\nu} g^{(T)}_\nu \ \forall \tau \in \mathbf{Z}^d \ \forall (r_\tau) \in \mathbf{E}(\mathbf{O}_1, ..., \mathbf{O}_d),\\
&(b') : p_\tau = \sum_{\nu \in \mathbf{Z}} p_{\tau-\nu} h^{(T)}_\nu \ \forall \tau \in \mathbf{Z} \ \forall (p_\tau) \in \mathbf{E}(\mathbf{O}_{d+1}),
\end{aligned}
\tag{104}
$$

where $\mathbf{E}(\mathbf{O}^1, ..., \mathbf{O}^m)$ is the space of all simple exponential polynomials on $\mathbf{Z}^m$ with the sets of partial frequencies $\mathbf{O}^1, ..., \mathbf{O}^m$. Setting $q^{(T)}_\tau = g^{(T)}_{\tau_1,...,\tau_d} h^{(T)}_{\tau_{d+1}}$, $\tau \in$

$\mathbf{Z}^{d+1}$, we clearly have

$$\begin{array}{l} q^{(T)} \in C_T(\mathbf{Z}^{d+1}), \quad |q^{(T)}|_2 = |g^{(T)}|_2 |h^{(T)}|_2 \le \rho_d(N_1, ..., N_d)\rho_1(N_{d+1}) \\ = \rho_{d+1}(N_1, ..., N_{d+1}) \end{array} \tag{105}$$

(see (104.a, a'). Further, for every $(s_\tau) \in \mathbf{E}(\mathbf{O}_1, ..., \mathbf{O}_{d+1})$ we have, setting $\tau = (\tau', \tau'')$ with $\tau' \in \mathbf{Z}^d$, $\tau'' \in \mathbf{Z}$:

$$\sum_{\nu \in \mathbf{Z}^{d+1}} q_\nu^{(T)} s_{\tau-\nu} = \sum_{\nu' \in \mathbf{Z}^d} g_{\nu'}^{(T)} \left(\sum_{\nu'' \in \mathbf{Z}} h_{\nu''}^{(T)} s_{\tau'-\nu', \tau''-\nu''} \right) \underbrace{=}_{a} \sum_{\nu' \in \mathbf{Z}^d} g_{\nu'}^{(T)} s_{\tau'-\nu', \tau''}$$
$$\underbrace{=}_{b} s_{\tau', \tau''}$$

(a is by (104.b') since $(s_{\tau'-\nu', \mu})_{\mu \in \mathbf{Z}} \in \mathbf{E}(\mathbf{O}_{d+1})$, b is by (104.b) since $(s_{\mu, \tau''})_{\mu \in \mathbf{Z}^d} \in \mathbf{E}(\mathbf{O}_1, ..., \mathbf{O}_d)$), which combines with (105) to imply that

$$(s_\tau) \in \mathbf{S}_\infty^t(0, \rho_{d+1}(N_1, ..., N_{d+1}), T).$$

Thus, the filters $q^{(T)}$ (which depend solely on T and $\mathbf{O}_1$,...,$\mathbf{O}_{d+1}$) certify the inclusion $(s_\tau) \in \mathbf{L}_\infty^t(0, \rho_{d+1}(N_1, ..., N_{d+1})$. The inductive step is completed. □

7.9 Proof of statement in Remark 9

It suffices to prove that for every nonnegative integer T and every m, d there exists a filter $q^{(T)}$, $\mathrm{ord}(q^{(T)}) \le T$, depending solely on T, m, d, such that

$$\begin{array}{ll} (a) & q^{(T)}(\Delta)p = p \text{ for every polynomial (14)}, \\ (b) & |q^{(T)}|_2 \le \left(\frac{16m}{\sqrt{2T+1}} \right)^d \equiv \Theta^d. \end{array} \tag{106}$$

This well-known fact can be proved by induction in d completely similar to the one used to prove Lemma 31; the only difference is in the Base, which now should be replaced with the following statement:

Lemma 32 *Let $p(\tau) = \sum\limits_{\ell=0}^{m} p_\ell \tau^\ell$ be a deterministic univariate algebraic polynomial of degree m. Then for every $T \ge 0$ there exists a filter $q \in C_T(\mathbf{Z})$, depending solely on T, m, with $|q|_2 \le 16m(2T+1)^{-1/2}$ such that $p(t) = \sum\limits_\nu q_\nu^{(T)} p(t-\nu)$ for all $t \in \mathbf{Z}$.*

Proof. By evident reasons, it suffices to prove that for a given $T \geq 0$ there exists a collection of weights q_t, $-T \leq t \leq T$, such that

$$\sum_{t=-T}^{T} q_t = 1, \ \sum_{t=-T}^{T} q_t t^i = 0, \ i = 1, ..., m, \ \sum_{t=-T}^{T} q_t^2 \leq \Theta^2 \equiv \frac{256m^2}{2T+1}.$$

By the standard separation arguments, this is the same as to prove that for every real algebraic polynomial $r(t)$ of degree $\leq m$ such that $r(0) = 1$ one has $\sum_{t=-T}^{T} r^2(t) \geq \frac{2T+1}{256m^2}$, or, which is the same, that for the real trigonometric polynomial $\rho(\phi) = r(T\sin(\phi))$ one has

$$\sum_{t=-T}^{T} \rho^2(\phi_t) \geq \frac{2T+1}{256m^2}, \quad \phi_t = \mathrm{asin}(t/T). \tag{107}$$

Note that the degree of the trigonometric polynomial $\rho(\cdot)$ is $\leq m$ and that $\rho(0) = 1$. Besides this, $\rho(\phi) = \rho(\pi - \phi)$; due to the latter fact,

$$M \equiv \max_{\phi} |\rho(\phi)| = \max_{|\phi| \leq \frac{\pi}{2}} |\rho(\phi)| \geq |\phi(0)| = 1.$$

By Bernstein's Theorem on trigonometric polynomials, we have $|\rho'(\phi)| \leq mM$. Now let $\phi_* \in [-\pi/2, \pi/2]$ be a point such that $|\rho(\phi_*)| = M$, let $\widehat{\Delta}$ be the segment of the length $\frac{1}{m}$ centered at ϕ_*, and Δ be the part of this segment in $[-\pi/2, \pi/2]$. Note that the length of Δ is at least $\frac{1}{2m}$ and that for $\phi \in \Delta$ one has $|\rho(\phi)| \geq |\rho(\phi_*)| - \frac{1}{2m}(mM) \geq M/2$. Let n be the minimum number of points ϕ_t belonging to a segment $\delta \subset [-\pi/2, \pi/2]$ of the length $1/(2m)$, the minimum being taken over all positions of δ in $[-\pi/2, \pi/2]$. It is immediately seen that $n \geq (1 - \sin(\pi/2 - 1/(2m)))T - 2 \geq \frac{T}{16m^2} - 2$, whence

$$\sum_{t=-T}^{T} \rho^2(\phi_t) \geq \sum_{t:\phi_t \in \Delta} \rho^2(\phi_t) \geq \frac{M^2}{4} n \geq \frac{M^2}{4}\left[\frac{T}{16m^2} - 2\right] \geq \frac{1}{4}\left[\frac{T}{16m^2} - 2\right].$$

When $T \geq 64m^2$, the latter quantity is $\geq \frac{2T+1}{256m^2}$, and in any case $\sum_{t=-T}^{T} \rho^2(\phi_t) \geq \rho^2(\phi_0) = 1$. Thus, we always have $\sum_{t=-T}^{T} \rho^2(\phi_t) \geq \frac{2T+1}{256m^2}$, as required in (107). □

7.10 Proof of Proposition 10

In the proof to follow, c_i stand for positive constants depending solely on **D**.
1^0. We start with the following evident observation:

Lemma 33 *There exists* c_1 *such that for every polynomial* $p(t)$ *of one variable satisfying the relation* $p(1) = 1$ *one has*

$$\begin{array}{l} M \leq c_1 N, \deg(p) \leq c_1 N, \\ (s) \in \mathbf{H}_N^t(\mathbf{D}) \Rightarrow s_\tau = (p(\mathbf{D})s)_\tau \ \forall (\tau : |\tau - t| \leq M). \end{array} \tag{108}$$

$\mathbf{2^0.}$ Let us fix a positive integer N, and let

$$\begin{array}{l} \delta(\omega) = \sum\limits_{\ell=1}^{k} w_\ell \exp\{i\omega^T\alpha(\ell)\} : [-\pi,\pi]^d \to \mathbb{C}, \\ \Omega_N^d = \left\{\omega \in \mathbb{R}^d \mid \omega_j \in \left\{\frac{q\pi}{2N+1}\right\}_{|q|\leq N}, j = 1, ..., d\right\}, \end{array} \tag{109}$$

and let ν be the normalized counting measure on Ω_N^d: $\nu(\{\omega\}) = (2N+1)^{-d}$, $\omega \in \Omega_N^d$. Observe that in view of **R.2** the function $\delta(\cdot)$ maps Ω_N^d into the unit disk $D = \{\zeta \in \mathbb{C} \mid |\zeta| \leq 1\}$. Let μ be the distribution of values of $\delta|_{\Omega_N^d}$, so that μ is the measure supported by the finite set $\mathbf{M} = \{\zeta \mid \exists \omega \in \Omega_N^d : \zeta = \delta(\omega)\}$, and $\mu(\{\zeta\}) = \sum\limits_{\omega\in\Omega_N^d:\delta(\omega)=\zeta} \nu(\{\omega\})$. Let also $F(\alpha) = \mu\left(\{\zeta \mid \Re(\zeta) \geq 1-\alpha\}\right),\ \alpha \geq 0$.

Lemma 34 *There exists* $c_2 \in (0,1)$ *such that*

$$\mathbf{M} \subset \widehat{\mathbf{M}} = \left\{\zeta \mid |\zeta| \leq 1, |\Im(\zeta)| \leq c_2^{-1}(1-\Re(\zeta))^{3/2}\right\}, \tag{110}$$

$$F(\alpha) \leq c_2^{-1}[\alpha^{d/2} + N^{-d}],\ 0 \leq \alpha \leq 2. \tag{111}$$

Proof. (110), (111) are evident when $\sum\limits_{\ell=1}^{k} \rho_\ell < 1$, since then $|\delta(\omega)| \leq 1 - c_2$ for properly chosen c_2 and all ω. Thus, in the sequel we focus on the case of $\sum\limits_{\ell=1}^{k} \rho_\ell = 1$ (recall that $\sum\limits_{\ell=1}^{k} \rho_\ell \leq 1$ by **R.2**).

$\mathbf{2^0.1)}$ Let $\mathbf{K} = \{\omega \in [-\pi,\pi]^d : \delta(\omega) = 1\}$. Since $\rho_\ell > 0$, $\sum\limits_\ell \rho_\ell = 1$ and $\delta(\omega) = \sum\limits_\ell \rho_\ell \exp\{i\phi_\ell + \omega^T\alpha(\ell)\}$, a point $\omega \in \mathbf{K}$ must satisfy the equations

$$\exp\{i[\phi_\ell + \omega^T\alpha(\ell)]\} = 1 \quad \forall (1 \leq \ell \leq k), \tag{112}$$

whence $\phi_\ell + \omega^T\alpha(\ell) \in 2\pi\mathbf{Z} \quad \forall (1 \leq \ell \leq k)$. Since $\mathrm{Rank}\{\alpha(\ell) : 1 \leq \ell \leq k\} = d$, the latter system of equations implies that $\mathbf{K}$ belongs to a set of the form $r + A\mathbf{Z}^d$ with certain $d \times d$ nonsingular matrix A (depending solely on $\mathbf{D}$). The cardinality of the intersection of latter set with the cube $[-\pi,\pi]^d$ does not exceed certain c_3. Thus, $\mathrm{Card}\,\mathbf{K} \leq c_3$.

2^0.2) Let $\omega \in \mathbf{K}$, and let $d\omega \in \mathbb{R}^n$ be such that $|d\omega| \leq 1$. Then

$$\begin{array}{l}
\delta(\omega + d\omega) = \sum\limits_{\ell=1}^{k} \rho_\ell \exp\{i[\phi_\ell + \omega^T\alpha(\ell)]\} \exp\{i(d\omega)^T\alpha(\ell)\} \\
\underbrace{=}_{a} \sum\limits_{\ell=1}^{k} \rho_\ell \exp\{i(d\omega)^T\alpha(\ell)\} \\
\Rightarrow |\delta(\omega + d\omega)| = \left| \sum\limits_{\ell=1}^{k} \rho_\ell \exp\{i(d\omega)^T\alpha(\ell)\} \right| \\
\underbrace{\leq}_{b} \left| \sum\limits_{\ell=1}^{k} \rho_\ell \left(1 + i(d\omega)^T\alpha(\ell) - \frac{1}{2}\left((d\omega)^T\alpha(\ell)\right)^2\right) \right| + c_4|d\omega|^3 \\
\underbrace{=}_{b} \left| \sum\limits_{\ell=1}^{k} \rho_\ell \left(1 - \frac{1}{2}\left((d\omega)^T\alpha(\ell)\right)^2\right) \right| + c_4|d\omega|^3 \underbrace{\leq}_{c} 1 - c_5|d\omega|^2 + c_4|d\omega|^3
\end{array}$$

(for a, see (112), b is by (17.b), c is due to Rank $(\{\alpha(\ell)\}_\ell) = d$). It follows that with properly chosen c_6 one has

$$\forall (\omega \in [-\pi, \pi]^d, |\delta(\omega) - 1| \leq \alpha) \quad \exists \bar{\omega} \in \mathbf{K} : |\omega - \bar{\omega}| \leq c_6^{-1}\sqrt{\alpha}. \tag{113}$$

Since Card $(\mathbf{K}) \leq c_3$ by 2^0.1) and $|\delta(\omega)| \leq 1$ for all ω, we conclude that

$$\nu\left(\{\omega \in \Omega_N^d : |\delta(\omega) - 1| \leq \alpha\}\right) \leq c_7[\alpha^{d/2} + N^{-d}] \quad \forall \alpha \leq 2. \tag{114}$$

2^0.3) Now we can complete the proof of (110), (111). Let $\bar{\omega} \in \mathbf{K}$, $d\omega \in \mathbb{R}^d$, $|d\omega| \leq 1$. We have

$$\begin{array}{l}
\delta(\bar{\omega} + d\omega) = \sum\limits_{\ell=1}^{k} \rho_\ell \exp\{i[\phi_\ell + \bar{\omega}^T\alpha(\ell)]\} \exp\{i(d\omega)^T\alpha(\ell)\} \\
\underbrace{=}_{a} \sum\limits_{\ell=1}^{k} \rho_\ell \exp\{i(d\omega)^T\alpha(\ell)\} = \sum\limits_{\ell=1}^{k} \rho_\ell \Big(1 + i(d\omega)^T\alpha(\ell) \\
-\frac{1}{2}\left((d\omega)^T\alpha(\ell)\right)^2 - \frac{i}{6}\left((d\omega)^T\alpha(\ell)\right)^3 + r_\ell(\omega, d\omega)\Big), \qquad (115) \\
\qquad\qquad\qquad\qquad [|r_\ell(\omega, d\omega)| \leq c_{10}|d\omega|^4] \\
\underbrace{=}_{b} \sum\limits_{\ell=1}^{k} \rho_\ell \left(1 - \frac{1}{2}\left((d\omega)^T\alpha(\ell)\right)^2 - \frac{i}{6}\left((d\omega)^T\alpha(\ell)\right)^3 + r_\ell(\omega, d\omega)\right)
\end{array}$$

(for a, see (112), for b, see (17)). Taking into account that $\sum\limits_\ell \rho_\ell = 1$ and $c_{11}|d\omega|^2 \leq \sum\limits_\ell \rho_\ell \left((d\omega)^T\alpha(\ell)\right)^2 \leq c_{12}|d\omega|^2$, we conclude from (113) combined with (115) that for properly chosen c_{13} one has

$$\omega \in [-\pi, \pi]^d \Rightarrow |\Im(\delta(\omega))| \leq c_{13}(1 - \Re(\delta(\omega)))^{3/2},$$

and (110) follows. By (110) one has $|1 - \delta(\omega)| \leq c_{14}(1 - \Re(\delta(\omega)))$, so that (111) follows from (114). $\square$

3^0. Let n be a positive integer, and let $T_n(\zeta)$ be the Tschebyshev polynomial of degree n. Recall that this polynomial is defined as follows:

$$T_n(\zeta) = \frac{w^n + w^{-n}}{2}, \text{ where } w = \zeta + i\sqrt{1-\zeta^2}. \tag{116}$$

In (116), the choice of the branch of $\sqrt{\cdot}$ affects the value of w, but does not affect the value of $w^n + w^{-n}$; since we intend to work with ζ from the unit disk, so that $\Re(1-\zeta^2) > 0$, in the calculations to follow we deal with the main branch of $\sqrt{\cdot}$ in the closed right half-plane. On the segment $[-1, 1]$ of the real axis one has $T_n(\zeta) = \cos(n\,\mathrm{acos}(\zeta))$, whence $T_n(1) = 1$, $T_n'(1) = n^2$. From these relations it follows that the function $P_n(\zeta) = \frac{1-T_n(\zeta)}{n^2(1-\zeta)}$ is a polynomial of degree $n-1$, and $P_n(1) = 1$.

Lemma 35 *One has*

$$\begin{array}{l} p_n(\alpha) \equiv \max\limits_{\zeta}\{|P_n(\zeta)| : \zeta \in \widehat{\mathbf{M}}, \Re(\zeta) = 1-\alpha\} \\ \leq q_n(\alpha) = \begin{cases} c_{15}, & 0 \leq \alpha \leq \frac{1}{n^2} \\ \frac{c_{15}(1+c_{15}\alpha)^n}{n^2\alpha}, & \frac{1}{n^2} \leq \alpha \leq 2 \end{cases}. \end{array} \tag{117}$$

Proof. Let $\zeta = 1 - \alpha + i\beta \in \widehat{\mathbf{M}}$, so that

$$|\beta| \leq c_{16}\alpha^{3/2}. \tag{118}$$

We have

$$\begin{array}{l} w \equiv \zeta + i\sqrt{1-\zeta^2} = 1 - \alpha + i\beta + i\sqrt{2\alpha - \alpha^2 - 2i(1-\alpha)\beta + \beta^2} \\ = 1 - \alpha + i\beta + i\sqrt{2\alpha}\sqrt{1 - 0.5\alpha + [0.5\beta - i(1-\alpha)](\beta/\alpha)} \\ = 1 + i\sqrt{2\alpha} + r_1(\zeta), \quad |r_1(\zeta)| \leq c_{17}\alpha \end{array} \tag{119}$$

(since $|\beta/\alpha| \leq c_{16}\sqrt{\alpha}$ by (119)). Note that completely similar considerations demonstrate that

$$w^{-1} = \zeta - i\sqrt{1-\zeta^2} = 1 - i\sqrt{2\alpha} + r_2(\zeta), \quad |r_2(\zeta)| \leq c_{17}\alpha. \tag{120}$$

3^0.1) Assume, first, that $0 \le \alpha \le \frac{1}{n^2}$. In this case from (119) it follows that $|1-w| \le \sqrt{2}n^{-1}$, whence, taking into account (119),

$$\begin{array}{l} |w^n - (1 + n(w-1) + \frac{n(n-1)}{2}(w-1)^2)| \le c_{17}(n|w-1|)^3 \le c_{18}n^3\alpha^{3/2}, \\ |w^{-n} - (1 - n(w-1) + \frac{n(n+1)}{2}(w-1)^2)| \le c_{17}(n|w-1|)^3 \le c_{18}n^3\alpha^{3/2} \\ \Rightarrow \left|\frac{w^n+w^{-n}}{2} - 1\right| \le \frac{n^2}{2}|w-1|^2 + c_{18}n^3\alpha^{3/2} \le c_{19}(n^2\alpha + n^3\alpha^{3/2}) \le c_{20}n^2\alpha. \end{array}$$

Thus, one has $|P_n(\zeta)| = \frac{\left|\frac{w^n+w^{-n}}{2}-1\right|}{n^2|\alpha-i\beta|} \le c_{15}$, as required in (117) for the case of $0 \le \alpha \le \frac{1}{n^2}$.

3^0.2) Now consider the case of $\frac{1}{n^2} \le \alpha \le 2$. From (119), (120) it follows that $|w| \le 1 + c_{21}\alpha$, $|w^{-1}| \le 1 + c_{21}\alpha$, whence $|P_n(\zeta)| = \frac{\left|\frac{w^n+w^{-n}}{2}-1\right|}{n^2|\alpha-i\beta|} \le \frac{c_{22}(1+c_{21}\alpha)^n}{n^2\alpha}$, as required in (117). □

4^0. Let $Q(\zeta) = \frac{1+\zeta}{2}$. It is immediately seen that

$$\zeta = 1 - \alpha + i\beta \in \widehat{\mathbf{M}} \Rightarrow |Q(\zeta)| \le 1 - c_{23}\alpha \quad [c_{23} < \frac{1}{2}]. \tag{121}$$

Now let c_{24} be a positive integer which is $\ge \frac{c_{15}}{c_{23}}$ (see (117)). Consider the polynomial $S_n(\zeta) = P_n(\zeta)Q^{c_{24}n}(\zeta)$.

Lemma 36 *For every positive integer n, the polynomial $S_n(\zeta)$ possesses the following properties:*

$$\begin{array}{l} (a) : \deg(S_n) \le c_{25}n;\ (b) : S_n(1) = 1; \\ (c) : \max\limits_{\zeta}\{|S_n(\zeta)| : \zeta \in \widehat{\mathbf{M}},\ \Re(\zeta) = 1-\alpha\} \le c_{15}\min\left[\frac{1}{n^2\alpha}; 1\right]. \end{array} \tag{122}$$

Proof. Relations (122.$a-b$) are evident (take into account that $P_n(1) = 1$ and $\deg(P_n) \le n$). To verify (122.c), note that if $\zeta = 1 - \alpha + i\beta \in \widehat{\mathbf{M}}$, then in view of (117) one has

$$\begin{array}{l} 0 \le \alpha \le \frac{1}{n^2} \Rightarrow |S_n(\zeta)| \le |P_n(\zeta)||Q(\zeta)|^{c_{24}n} \le |P_n(\zeta)| \le c_{15}; \\ \frac{1}{n^2} \le \alpha \le 2 \Rightarrow |S_n(\zeta)| \le |P_n(\zeta)||Q(\zeta)|^{c_{24}n} \underbrace{\le}_{a} c_{15}\frac{(1+c_{15}\alpha)^n}{n^2\alpha}(1-c_{23}\alpha)^{c_{24}n} \\ \le c_{15}\frac{\exp\{c_{15}n\alpha\}}{n^2\alpha}\exp\{-c_{23}c_{24}n\alpha\} \underbrace{\le}_{b} \frac{c_{15}}{n^2\alpha} \end{array}$$

(for a, see (121), b is due to $c_{23}c_{24} \ge c_{15}$). □

5^0. Now we are ready to complete the proof of Proposition 10. Given a positive integer n, let us set $R_n(\zeta) = S_n^d(\zeta)$. In view of (122) one has

$$\begin{array}{l} (a): \deg(R_n) \le c_{26}n;\ (b): R_n(1) = 1; \\ (c): \max\limits_{\zeta}\{|R_n(\zeta)| : \zeta \in \widehat{\mathbf{M}},\ \Re(\zeta) = 1-\alpha\} \le r_n(\alpha) \\ \quad \equiv c_{26} \min\left[\frac{1}{n^{2d}\alpha^d}; 1\right]. \end{array} \tag{123}$$

Consider the filters $q^{(n)}(z)$ given by $q^{(n)}(\Delta) = R_n(\mathbf{D}),\ n = 0, 1, ...$ By (123.b) and Lemma 33 we have

$$\left.\begin{array}{r} T \le c_{27}N \\ 1 \le n(T) \equiv \lfloor c_{27}T \rfloor \\ (s) \in \mathbf{H}_N^t(\mathbf{D}) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \mathrm{ord}(q^{(n(T))}) \le T, \\ s_\tau = (q^{(n(T))}(\Delta)s)_\tau\ \forall(\tau : |\tau - t| \le c_{27}N). \end{array}\right. \tag{124}$$

By Parseval's equality, we have also (in what follows, $n = n(T)$)

$$|q^{(n)}|_2^2 = \int\limits_{\Omega_N^d} |R_n(\delta(\omega))|^2 \nu(d\omega) = \int\limits_{\mathbf{M}} |R_n(\zeta)|^2 \mu(d\zeta) \underbrace{\le}_{a} \int\limits_0^2 \underbrace{r_n^2(\alpha)}_{\rho_n(\alpha)} dF(\alpha) \tag{125}$$

with a given by (123.c), (110) and the definition of $F(\cdot)$. Let γ be the measure on $[0, 2]$ defined by $G(\alpha) \equiv \gamma([0, \alpha]) = c_2^{-1}(\alpha^{d/2} + N^{-d})$, so that

$$F(\alpha) \le G(\alpha) \equiv \gamma([0, \alpha]) \quad \forall \alpha \in [0, 2] \tag{126}$$

(see (111)). We have

$$\begin{array}{l} \int\limits_0^2 \rho_n(\alpha) dF(\alpha) = \rho_n(2) - \int\limits_0^2 \rho_n'(\alpha) F(\alpha) d\alpha \underbrace{\le}_{a} \rho_n(2) \\ -\int\limits_0^2 \rho_n'(\alpha) G(\alpha) d\alpha = \rho_n(2) - \rho_n(2)G(2) + \int\limits_0^2 \rho_n(\alpha)\gamma(d\alpha) \\ \underbrace{\le}_{b} \int\limits_0^2 \rho_n(\alpha)\gamma(d\alpha) \underbrace{=}_{c} c_2^{-1}\Big[c_{28} \int\limits_0^2 \min^2\left[n^{-2d}\alpha^{-d}, 1\right] \alpha^{\frac{d}{2}-1} d\alpha \\ +\rho_n(0)N^{-d}\Big] \underbrace{\le}_{d} c_{30}\left[N^{-d} + n^{-d}\right] \le c_{31}(2T+1)^{-d} \end{array} \tag{127}$$

(a holds since $\rho_n(\cdot)$ is nonincreasing, see (123.c), and by (126), b holds since $c_2 \in (0, 1)$, see Lemma 34, c is by (123.c) and (125), d is due to $n = n(T) =$

$\lfloor c_{27}T\rfloor$). Combining (125) and (127), we conclude that

$$|q^{(n(T))}|_2 \le c_{32}(2T+1)^{-d/2}. \tag{128}$$

From (124) and (128) we conclude that if $L = \lfloor c_{33}N\rfloor$ and $T \le L$ is such that $n(T) \equiv \lfloor c_{27}T\rfloor \ge 1$, then

$$\exists q^{(T)} \in C_T(\mathbf{Z}^d) : \begin{cases} |q^{(T)}|_2 \le c_{32}(2T+1)^{-d/2}, \\ s_\tau = (q^{(T)}(\Delta)s)_\tau \,\forall(\tau, |\tau - t| \le L, (s) \in \mathbf{H}_N^t(\mathbf{D})) \end{cases} \tag{129}$$

(indeed, one can choose, as a required $q^{(T)}$, the filter $q^{(n(T))}$). Setting $q^{(T)}(z) \equiv 1$ for $T < \frac{1}{c_{27}}$, we enforce the validity of (129) for all T, $0 \le T \le L$. Thus, $\mathbf{H}_N^t(\mathbf{D}) \subset \mathbf{F}^t_{\lfloor c_{29}L\rfloor}(0, c_{34})$. $\square$

7.11 Proof of Proposition 11

Lemma 37 *Let $f \in \mathbf{H}^+(M)$ be a deterministic function, let $N \le M/2$, and let $t \in \mathbf{Z}^d$, $|t| \le N$. Consider the "discrete box" $B_N^t = \{\tau \in \mathbf{Z}^d : |\tau - t| \le N\}$, and let ϕ be a deterministic function on B_N^t which coincides with f on the "discrete boundary" $\partial B_N^t \equiv \{\tau \in \mathbf{Z}^d : |\tau - t| = N\}$ of B_N^t and is "discrete harmonic": $\tau \in \mathbf{Z}^d, |\tau - t| < N \Rightarrow \phi_\tau = \frac{1}{2d} \sum\limits_{\substack{\epsilon=(\epsilon_1,\dots,\epsilon_d) \\ |\epsilon_1|=\dots=|\epsilon_d|=1}} \phi_{\tau+\epsilon}$. Then*

$$\tau \in B_N^t \Rightarrow |f(\tau) - \phi_\tau| \le c_1 \|f\|_{\infty,2M} N^{-2} \tag{130}$$

(from now on, c_i are positive absolute constants).

Proof. First, we should prove that the "discrete harmonic" function ϕ on B_N^t which coincides with f on ∂B_N^t does exist. This fact is well known; we present here its proof just for the sake of completeness. Let ψ be a function on ∂B_N^t. Consider the following random walk on B_N^t: arriving for the first time at a point τ from ∂B_N^t, we pay penalty $\psi(\tau)$ and terminate; from an "interior point" $\tau \in \mathrm{int}B_N^t \equiv B_N^t \backslash \partial B_N^t$ we make a random step of length 1 along one of the coordinate axes, choosing every one of $2d$ possible steps with probability $1/(2d)$. It is immediately seen that the expected penalty payed at the termination, treated as a function of the initial state, is a discrete harmonic function with the boundary values ψ.
Now, since $|t| \le N$ and $2N \le M$, the function f is harmonic in the "continuous box" $D_{2N}^t = \{\tau \in \mathbb{R}^d : |\tau - t| \le 2N\}$, and the uniform norm of f in this square does not exceed $\|f\|_{\infty,2M}$. From the standard results on harmonic functions it

follows that

$$\forall(\tau \in D_N^t): \left|\frac{\partial^\kappa}{\partial x_j^\kappa} f(\tau)\right| \le c_2\|f\|_{\infty,2M} N^{-\kappa},\ \kappa = 1,2,3,4,\ j = 1,...,d. \quad (131)$$

Consequently, for the basic orths e_j, $j = 1, ..., d$ we have

$$\tau \in D_N^t, |s| \le 1 \Rightarrow \left| f(\tau + se_j) - \sum_{\kappa=0}^{3} \frac{1}{\kappa!} \frac{\partial^\kappa}{\partial x_j^\kappa} f(\tau) s^\kappa \right| \le c_3 |s|^4 \|f\|_{\infty,2M} N^{-4}.$$

Since f is harmonic, we conclude that

$$|(\mathbf{D}f)_\tau| \le c_4 \|f\|_{\infty,2M} N^{-4},\ \tau \in B_N^t. \quad (132)$$

Now let $h = f|_{\mathbf{Z}^d} - \phi \in C(B_N^t)$ and let $h_\tau^\pm = h_\tau \pm \frac{2c_4\|f\|_{\infty,2M}}{N^4} \sum_{j=1}^{d} (\tau_j - t_j)^2$. Taking into account (132) and the fact that ϕ is discrete harmonic, we have for $\tau \in \mathrm{int} B_N^t$:

$$(\mathbf{D}h^+)_\tau = (\mathbf{D}h)_\tau + \frac{2c_4\|f\|_{\infty,2M}}{N^4} > 0,\ (\mathbf{D}h^-)_\tau = (\mathbf{D}h)_\tau - \frac{2c_4\|f\|_{\infty,2M}}{N^4} < 0,$$

whence both the maximum of h^+ and the minimum of h^- over B_N^t are attained at ∂B_N^t. Since at the discrete boundary of B_N^t we have $f = \phi$ and therefore $h^+ \le 4c_4\|f\|_{\infty,2M} N^{-2}$, we conclude that $\tau \in B_N^t \Rightarrow h_\tau \le h_\tau^+ \le \max_{\tau \in \partial B_N^t} h_\tau^+ \le 2dc_4\|f\|_{\infty,2M} N^{-2}$. By similar reasons, $\tau \in B_N^t \Rightarrow h_\tau \ge h_\tau^- \ge \min_{\tau \in \partial B_N^t} h_\tau^- \ge -2dc_4\|f\|_{\infty,2M} N^{-2}$. $\square$

Now let $|t| \le M/8$ and $L \le M/8$. Given T, $0 \le T \le L$, and applying Proposition 10, we can build filter $q^{(T)} \in C_T(\mathbf{Z}^d)$ such that

$$|q^{(T)}|_2 \le c_5(2T+1)^{-1},\ \phi_\tau = \sum_{|\nu| \le T} \phi_{\tau-\nu} q_\nu^{(T)}\ \forall(\tau : |\tau - t| \le L) \quad (133)$$

for every ϕ which is discrete harmonic in the discrete box B_{2L}^t. Now let $f \in \mathbf{H}(M, R)$. Applying Lemma 37, we can find function ϕ which is discrete harmonic in the box B_{2L}^t and such that $|\phi_\tau - f_\tau|^2 \le c_6^2 \|f\|_{\infty,2M}^2 L^{-4}$ for $\tau \in B_{2L}^t$. From (133) it now follows that

$$\begin{aligned}
&\forall(\tau : |\tau - t| \le L) : \left[E\left\{ |f_\tau - \sum_{|\nu| \le T} f_{\tau-\nu} q_\nu^{(T)}|^2 \right\} \right]^{1/2} \\
&\le c_6 \underbrace{[E\{\|f\|_{\infty,2M}^2\}]^{1/2}}_{\le R} L^{-2}(1 + |q^{(T)}|_1) \le c_6 R L^{-2}(1 + |q^{(T)}|_2 (2T+1)^{d/2}) \\
&\le c_8 R L^{-2} \le c_9 R (2T+1)^{-d/2}
\end{aligned}$$

(recall that $d \leq 4$). $\square$

7.12 Proof of Theorem 20

1^0. We start from the following observation:

$$\epsilon \notin \Xi \Rightarrow |s_t - \widehat{s}_t(y)| \leq 3c(d)\rho^3 \left[\theta + \sigma\rho\Theta^t\right]. \tag{134}$$

Indeed, assume that $\epsilon \notin \Xi$. By construction, $\widehat{s}_t(y)$ is certain estimate $\widehat{s}_t[T, y]$ corresponding to a normal window width T; by the definition of normality, we have $|\widehat{s}_t(y) - \widehat{s}_t[0, y]| = |\widehat{s}_t[T, y] - \widehat{s}_t[0, y]| \leq \gamma(T) + \gamma(0) \leq 2\gamma(0)$. Taking into account that $\widehat{s}_t[0, y] = y_t$ (see the description of Algorithm A), we conclude that $|\widehat{s}_t(y) - s_t| \leq |\widehat{s}_t(y) - \widehat{s}_t[0, y]| + |\widehat{s}_t[0, y] - s_t| \leq 2\gamma(0) + |e_t| \leq 2\gamma(0) + \sigma|\epsilon_t|$, so that

$$|\widehat{s}_t(y) - s_t| \leq 2\gamma(0) + \sigma\Theta^t. \tag{135}$$

Recalling the definition of $\gamma(\cdot)$ (see (32)), we arrive at

$$\gamma(0) \leq c(d)\rho^3 \left[\theta + C\sigma\rho\omega\sqrt{\ln(2L_\infty + 1)}\right]; \tag{136}$$

since $\epsilon \notin \Xi$, we have $C\omega\sqrt{\ln(2L_\infty + 1))} \leq \Theta^t$, so that (135), (136) imply (134).
2^0. Now we are ready to complete the proof of Theorem 20. Assume first that $\epsilon \notin \Xi$. In this case from it follows from (32) that the window width $T = T_*$ is normal, and consequently the largest normal window width $T_+(y)$ is $\geq T_*$. Since $T_+(y)$ is a normal window width, (32) is applicable with $T = T_+(y), T' = T_*$, so that $|\widehat{s}_t(y) - \widehat{s}_t[T_*, y]| \leq \gamma(T_+(y)) + \gamma(T_*) \leq 2\gamma(T_*). It follows that \epsilon \in \Xi \Rightarrow |\widehat{s}_t(y) - s_t| \leq 2\gamma(T_*) + |\widehat{s}_t[T_*, y] - s_t|$. The latter relation combines with (134) to imply that

$$\begin{aligned} &|\widehat{s}_t(y) - s_t| \leq 2\gamma(T_*) + |\widehat{s}_t[T_*, y] - s_t| + 3c(d)\rho^3 \left[\theta + \sigma\rho\Theta^t\right] \chi_{\epsilon\notin\Xi} \\ &= 2c(d)\rho^3 \left[\theta + C\sigma\rho\omega\sqrt{\ln(2L_\infty + 1)}\right] (2T_* + 1)^{-d/2} + |\widehat{s}_t[T_*, y] - s_t| \\ &+3c(d)\rho^3 \left[\theta + \sigma\rho\Theta^t\right] \chi_{\epsilon\notin\Xi} \end{aligned} \tag{137}$$

Denoting by c_i positive quantities depending solely on d (which is the case for C and ω), we conclude from (137) that

$$\begin{array}{l}\left(E\left\{|\widehat{s}_t(y)-s_t|^2\right\}\right)^{1/2} \le c_1\rho^3\left[\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}\right](2T_*+1)^{-d/2}\\ +\overbrace{\left(E\left\{|\widehat{s}_t[T_*,y]-s_t|^2\right\}\right)^{1/2}}^{J}+c_1\rho^3\theta\mathrm{Prob}\{\xi\not\in\Xi\}\\ +c_1\rho^4\sigma\left(E\left\{(\Theta^t)^2\chi_{\epsilon\not\in\Xi}\right\}\right)^{1/2}.\end{array} \tag{138}$$

We have

$$J \le c(d)\rho^3\frac{\theta+\sigma\rho\sqrt{\ln(2T_*+1)}}{(2T_*+1)^{d/2}} \le c(d)\rho^3\frac{\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}}{(2T_*+1)^{d/2}}$$

by Theorem 2 (recall that $(s)\in\mathbf{F}^t_{3T_*}(\theta,\rho)$ and $T_*\le L_\infty$ by the definition of T_*), and therefore (138) implies that

$$\begin{array}{l}\left(E\left\{|\widehat{s}_t(y)-s_t|^2\right\}\right)^{1/2} \le c_2\rho^3\left[\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}\right](2T_*+1)^{-\frac{d}{2}}\\ +c_1\rho^3\theta\mathrm{Prob}\{\xi\not\in\Xi\}+c_1\rho^4\sigma\left(E\left\{(\Theta^t)^2\chi_{\epsilon\not\in\Xi}\right\}\right)^{1/2}\\ \underbrace{\le}_{a} c_2\rho^3\left[\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}\right](2T_*+1)^{-\frac{d}{2}}+c_1\rho^3\theta(2L_\infty+1)^{-2d}\\ +c_1\rho^4\sigma\left(E\left\{(\Theta^t)^2\chi_{\epsilon\not\in\Xi}\right\}\right)^{1/2}\le c_3\rho^3\left[\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}\right](2T_*+1)^{-\frac{d}{2}}\\ +c_1\rho^4\sigma\left(E\left\{(\Theta^t)^2\chi_{\epsilon\not\in\Xi}\right\}\right)^{1/2}\le c_3\rho^3\left[\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}\right](2T_*+1)^{-\frac{d}{2}}\\ +c_1\rho^4\sigma\left[\left(E\left\{(\Theta^t)^4\right\}\right)^{1/2}\left(E\left\{\chi_{\epsilon\not\in\Xi}\right\}\right)^{1/2}\right]^{1/2}\\ \underbrace{\le}_{b} c_3\rho^3\frac{\left[\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}\right]}{(2T_*+1)^{\frac{d}{2}}}+c_4\rho^4\sigma\frac{\sqrt{\ln(2L_\infty+1)}}{(2L_\infty+1)^{-\frac{d}{2}}}\underbrace{\le}_{c} c_5\rho^3\frac{\left[\theta+\sigma\rho\sqrt{\ln(2L_\infty+1)}\right]}{(2T_*+1)^{\frac{d}{2}}}\end{array}$$

(for a, see by (30), for b, see (29), (30), for c, take into account that $T_*\le L_\infty$), as required in (34). □

7.13 Proof of Theorem 22

7.13.1 Preliminaries

Proof of Lemma 21. To save notation, let $B=B_h(x)$ and $T=T_h(x)$. Let $p\in C(\mathbf{Z}^d)$ be such that $p\in\mathbf{S}^t_{3T}(0,\mu,T)$ and $|p_\tau-f_\tau|\le\Phi_\mu(f,B_h(x))$ for all $\tau\in\mathbf{Z}(B_h(x))$. Since $p\in\mathbf{S}^t_{3T}(0,\mu,T)$, there exists a filter $q\in C_T(\mathbf{Z}^d)$ such that $|q|_2\le\mu(2T+1)^{-d/2}$ and $(q(\Delta)p)_\tau=p_\tau$ whenever $|\tau-t|\le 3T$. Setting

$\delta_\tau = f_\tau - p_\tau$, we have

$$
\begin{array}{l}
|\tau - t| \leq 3T \Rightarrow |f_\tau - (q(\Delta)f)_\tau| \leq |\delta_\tau| + |p_\tau - (q(\Delta)p)_\tau| + |(q(\Delta)\delta)_\tau| \\
\leq \Phi_\mu(f, B_h(x)) + |q|_1 \max\{|\delta_\nu| : |\nu - \tau| \leq T\} \leq \Phi_\mu(f, B_h(x)) \\
+|q|_2(2T+1)^{d/2}\Phi_\mu(f, B_h(x)) \max\{|\delta_\nu| : |\nu - \tau| \leq T\} \\
\qquad\qquad [\text{note that } |\tau - t| \leq 3T,\ |\nu - \tau| \leq T \text{ implies } |\nu - t| \leq 4T] \\
\leq \Phi_\mu(f, B_h(x))(1+\mu) = \frac{\Phi_\mu(f,B_h(x))(1+\mu)}{(2T+1)^{d/2}}(2T+1)^{-d/2}
\end{array}
$$

as required in (40). $\square$

7.13.2 Proof of Theorem 22

1^0. In the main body of the proof, we focus on the case $p, q < \infty$; the case of infinite p and/or q will be considered at the concluding step 4^0.
Let us fix a family of well-filtered signals $\mathbf{F} = \mathbf{F}_d^{k,\rho,p}(B; R)$ with the parameters satisfying the premise of Theorem 22 and a function f from this class.
Recall that by the definition of $\mathbf{F}$ there exists a function $F \geq 0$, $\|F\|_{p,B} \leq R$, such that

$$
\begin{array}{l}
\forall\, (x = m^{-1}t \in (\mathrm{int}B) \cap \Gamma_n)\ \ \forall\, (h, B_h(x) \subset B) : \\
\Phi_\mu(f, B_h(x)) \leq P_1 h^{k-d\pi}\Omega(f, B_h(x)),\ \ \Omega(f, B') = \left(\int\limits_{B'} F^p(u)du\right)^{1/p};
\end{array}
\tag{139}
$$

from now on, P (perhaps with sub- or superscripts) are quantities ≥ 1 depending on μ, d, γ, p only and nonincreasing in $p > d$.
2^0. Our central auxiliary result is as follows:

Lemma 38 *Assume that*

$$
n^{\frac{k-d\pi}{d}}\sqrt{\ln n} \geq P_1(\mu+3)^{k-d\pi+d/2}\frac{R}{\sigma\omega}. \tag{140}
$$

Given a point $x \in \Gamma_n \cap B_\gamma$, let us choose the largest $h = h(x)$ such that

$$
(a) : h \leq (1-\gamma)D(B),\ \ (b) : P_1 h^{k-d\pi}\Omega(f, B_h(x)) \leq S_n(h). \tag{141}
$$

Then $h(x)$ is well-defined and

$$
h(x) \geq m^{-1}. \tag{142}
$$

Besides this, the error at x of the adaptive estimate $\widehat{f}_n$ as applied to f can be bounded as follows:

$$\begin{array}{l} (a): \underline{\textit{in the case of } \epsilon \in \Xi_n}: |\widehat{f}_n(x;y) - f(x)| \le C_2 S_n(h(x)); \\ (b): \underline{\textit{in the case of } \epsilon \notin \Xi_n}: |\widehat{f}_n(x;y) - f(x)| \le C_2 \sigma \Theta_{(n)}. \end{array} \tag{143}$$

Proof of Lemma. a^0. The quantity $h(x)$ is well-defined, since for small positive h the left hand side in (141.b) is close to 0, while the right hand side one is large. Besides this, let $h_- = m^{-1}$. From (140) it follows that h_- satisfies (141.a), so that $B_{h_-}(x) \subset B$. Moreover, (140.b) implies that $P_1 h_-^{k-d\pi} R \le S_n(h_-)$; the latter inequality, in view of $\Omega(f, B_{h_-}(x)) \le R$, says that h_- satisfies (141.b) as well. Thus, $h(x) \ge h_-$, as claimed in (142).
b^0. Consider the window $B_{h(x)}(x)$. By (141.a) it is admissible for x, while from (141.b) combined with (139) we get $\Phi_\mu(f, B_{h(x)}(x)) \le S_n(h)$. It follows that the ideal window $B_*(x)$ of x is not smaller than $B_{h(x)}(x)$.
c^0. Assume that $\epsilon \in \Xi_n$. Then, according to (49), we have

$$|\widehat{f}_n(x;y) - f(x)| \le 5C_1 \left[\Phi_\mu(f, B_{h_*(x)}(x)) + S_n(h_*(x))\right]. \tag{144}$$

Now, by the definition of an ideal window, $\Phi_\mu(f, B_{h_*(x)}(x)) \le S_n(h_*(x))$, and the right hand side in (144) does not exceed $10C_1 S_n(h_*(x)) \le 10C_1 S_n(h(x))$ (recall that, as we have seen, $h_*(x) \ge h(x)$), as required in (143.a).
d^0. Now let $\epsilon \notin \Xi_n$. Note that $\widehat{f}_n(x;y)$ is certain estimate $\widehat{f}^h(x;y)$ associated with a centered at x and admissible for x cube $B_h(x)$ which is normal and such that $h \ge m^{-1}$ (the latter – since the window $B_{m^{-1}}(x)$ always is normal, and $B_h(x)$ is the largest normal window centered at x). Applying (48) with $h' = m^{-1}$ (so that $\widehat{f}_n^{h'}(x;y) = f(x) + \sigma\epsilon_t$), we get $|(f(x) + \sigma\epsilon_t) - \widehat{f}_n(x;y)| \le 4C_1 S_n(m^{-1})$, whence

$$|f(x) - \widehat{f}_n(x;y)| \le \sigma|\epsilon_t| + 4C_1 S_n(m^{-1}) \le \sigma\Theta_{(n)} + 4C_1\sigma\omega\sqrt{\ln n} \le C_2\Theta_{(n)}$$

(recall that we are in the situation $\epsilon \notin \Xi_n$, whence $\omega\sqrt{\ln n} \le \Theta_{(n)}$). We have arrived at (143.b). □
3^0. Now we are ready to complete the proof. Assume that (140) takes place, and let us fix q, $\frac{2k+d}{d} p \le q < \infty$.

3^0.a) Note that for every $x \in \Gamma_n \cap B_\gamma$

$$\begin{array}{lll} \text{either } h(x) = (1-\gamma)D(B), & & (a) \\ \text{or} \quad \underbrace{P_1 h^{k-d\pi}(x)\Omega(f, B_{h(x)}(x)) = S_n(h(x))} & & (b) \\ \quad \Leftrightarrow h(x) = \left(\dfrac{\widehat{\sigma}_n}{P_1\Omega(f, B_{h(x)}(x))}\right)^{\frac{2}{2k+d-2d\pi}} & & \end{array} \tag{145}$$

Let U, V be the sets of those $x \in B_\gamma^n \equiv \Gamma_n \cap B_\gamma$ for which the first, respectively, the second of this possibilities takes place. If V is nonempty, let us partition it as follows.
1) We can choose $x_1 \in V$ (V is finite!) such that $h(x) \geq h(x_1) \quad \forall x \in V$. After x_1 is chosen, we set $V_1 = \{x \in V \mid B_{h(x)}(x) \cap B_{h(x_1)}(x_1) \neq \emptyset\}$.
2) If the set $V \backslash V_1$ is nonempty, we apply the construction from 1) to this set, thus getting $x_2 \in V \backslash V_1$ such that $h(x) \geq h(x_2) \quad \forall x \in V \backslash V_1$, and set $V_2 = \{x \in V \backslash V_1 \mid B_{h(x)}(x) \cap B_{h(x_2)}(x_2) \neq \emptyset\}$. If the set $V \backslash (V_1 \cup V_2)$ still is nonempty, we apply the same construction to this set, thus getting x_3 and V_3, and so on.
The outlined process clearly terminates after certain step (since V is finite). On termination, we get a collection of M points $x_1, ..., x_M \in V$ and a partition $V = V_1 \cup V_2 \cup ... \cup V_M$ with the following properties:
(i) The cubes $B_{h(x_1)}(x_1), ..., B_{h(x_M)}(x_M)$ are mutually disjoint;
(ii) For every $\ell \leq M$ and every $x \in V_\ell$ we have $h(x) \geq h(x_\ell)$ and $B_{h(x)}(x) \cap B_{h(x_\ell)}(x_\ell) \neq \emptyset$.
We claim that also
(iii) For every $\ell \leq M$ and every $x \in V_\ell$ one has

$$h(x) \geq \max\left[h(x_\ell); \|x - x_\ell\|_\infty\right]. \tag{146}$$

Indeed, $h(x) \geq h(x_\ell)$ by (ii), so that it suffices to verify (146) in the case when $\|x - x_\ell\|_\infty \geq h(x_\ell)$. Since $B_{h(x)}(x)$ intersects $B_{h(x_\ell)}(x_\ell)$, we have $\|x - x_\ell\|_\infty \leq \frac{1}{2}(h(x) + h(x_\ell))$, whence $h(x) \geq 2\|x - x_\ell\|_\infty - h(x_\ell) \geq \|x - x_\ell\|_\infty$, which is what we need.
3^0.b) Let us set $B_\gamma^n = \Gamma_n \cap B_\gamma$. Assume that $\epsilon \in \Xi_n$. Then (below $D = D(B)$)

$$\begin{array}{l}
\|\widehat{f}_n(\cdot; y) - f(\cdot)\|_{q,B_\gamma}^q \underbrace{\leq}_{a} C_2^q m^{-\frac{d}{q}} \sum\limits_{x \in B_\gamma^n} S_n^q(h(x)) = C_2^q m^{-\frac{d}{q}} \sum\limits_{x \in U} S_n^q(h(x)) \\
+ C_2^q m^{-\frac{d}{q}} \sum\limits_{\ell=1}^{M} \sum\limits_{x \in V_\ell} S_n^q(h(x)) \underbrace{=}_{b} C_2^q m^{-\frac{d}{q}} \sum\limits_{x \in U} \left[\frac{\widehat{\sigma}_n}{((1-\gamma)D)^{d/2}}\right]^q \\
+ C_2^q m^{-\frac{d}{q}} \sum\limits_{\ell=1}^{M} \sum\limits_{x \in V_\ell} S_n^q(h(x)) \underbrace{\leq}_{c} C_3^q \widehat{\sigma}_n^q m^{-\frac{d}{q}} \sum\limits_{\ell=1}^{M} \sum\limits_{x \in V_\ell} \left(\max\left[h(x_\ell), \|x - x_\ell\|_\infty\right]\right)^{-\frac{dq}{2}} \\
+ C_3^q \widehat{\sigma}_n^q D^{\frac{d(2-q)}{2}} \underbrace{\leq}_{d} C_4^q \widehat{\sigma}_n^q \sum\limits_{\ell=1}^{M} \int \left(\max\left[h(x_\ell), \|x - x_\ell\|_\infty\right]\right)^{-\frac{dq}{2}} dx \\
+ C_3^q \widehat{\sigma}_n^q D^{\frac{d(2-q)}{2}} \leq C_5^q \widehat{\sigma}_n^q \sum\limits_{\ell=1}^{M} \int\limits_0^\infty r^{d-1} \left(\max\left[h(x_\ell), r\right]\right)^{-\frac{dq}{2}} dr + C_3^q \widehat{\sigma}_n^q D^{\frac{d(2-q)}{2}} \\
\underbrace{\leq}_{e} C_6^q \widehat{\sigma}_n^q \sum\limits_{\ell=1}^{M} \left[h(x_\ell)\right]^{\frac{d(2-q)}{2}} + C_3^q \widehat{\sigma}_n^q D^{\frac{d(2-q)}{2}} \underbrace{=}_{f} + C_6^q \widehat{\sigma}_n^q \sum\limits_{\ell=1}^{M} \left[\frac{\widehat{\sigma}_n}{P_1 \Omega(f, B_{h(x_\ell)}(x_\ell))}\right]^{\frac{d(2-q)}{2k-2d\pi+d}} \\
+ C_3^q \widehat{\sigma}_n^q D^{\frac{d(2-q)}{2}} \underbrace{=}_{g} C_3^q \widehat{\sigma}_n^q D^{\frac{d(2-q)}{2}} + C_6^q \widehat{\sigma}_n^{2\beta(p,k,d,q)q} \sum\limits_{\ell=1}^{M} \left[P_1 \Omega(f, B_{h(x_\ell)}(x_\ell))\right]^{\frac{d(q-2)}{2k-2d\pi+d}}
\end{array}$$

In this computation, a is by (143.a), b is since $h(x) = (1-\gamma)D$ for $x \in U$, c is given by (146), d is due to $h(x) \geq m^{-1}$, see (142). To justify e, note that $\frac{dq}{2} - d + 1 \geq \frac{2k+d}{2}p - d + 1 \geq d^2/2 + 1$ in view of $q \geq \frac{2k+d}{d}p$, $k \geq 1$ and $p > d$, f is given by f by (145.b), and g is by definition of $\beta(p,k,d,q)$.
Now note that $\frac{d(q-2)}{2k-2d\pi+d} \geq p$ in view of $q \geq \frac{2k+d}{d}p$, so that

$$\sum_{\ell=1}^{M} \left[P_1\Omega(f, B_{h(x_\ell)}(x_\ell))\right]^{\frac{d(q-2)}{2k-2d\pi+d}} \leq \left[\sum_{\ell=1}^{M} \left(P_1\Omega(f, B_{h(x_\ell)}(x_\ell))\right)^p\right]^{\frac{dq-2d}{p(2k-2d\pi+d)}} \\ \leq [P_1^p R^p]^{\frac{d(q-2)}{p(2k-2d\pi+d)}}$$

(see (139) and take into account that the cubes $B_{h(x_\ell)}(x_\ell)$, $\ell = 1, ..., M$, are mutually disjoint by (i)). We have arrived at

$$\begin{array}{l} \epsilon \in \Xi_n \Rightarrow \|\widehat{f}_n(\cdot; y_f(\epsilon)) - f(\cdot)\|_{q,B_\gamma} \leq C_7\widehat{\sigma}_n D^{\frac{d(2\theta-1)}{2}}(B) \\ +P_2\widehat{\sigma}_n^{2\beta(p,k,d,q)} R^{\frac{d(1-2\theta)}{2k-2d\pi+d}} = C_7\widehat{\sigma}_n D^{\frac{d(2\theta-1)}{2}}(B) + P_2 R \left(\frac{\widehat{\sigma}_n}{R}\right)^{2\beta(p,k,d,q)} \end{array} \tag{147}$$

3^0.c) Now assume that $\epsilon \notin \Xi_n$. In this case, by (143.b), $|\widehat{f}_n(x;y) - f(x)| \leq C_2\sigma\Theta_{(n)} \quad \forall x \in B_\gamma^n$, whence, taking into account that $mD(B) \geq 1$,

$$\|\widehat{f}_n(\cdot;y) - f(\cdot)\|_{q,B_\gamma} \leq C_2\sigma\Theta_{(n)} D^{\frac{d}{q}}(B). \tag{148}$$

3^0.d) Combining (147) and (148), we get

$$\begin{array}{l} \sqrt{E\left\{\|\widehat{f}_n(\cdot;y) - f(\cdot)\|_{q,B_\gamma}^2\right\}} \\ \leq C_8 \max\left[\widehat{\sigma}_n D^{\frac{d(2\theta-1)}{2}}(B); P_4 R\left(\frac{\widehat{\sigma}_n}{R}\right)^{2\beta(p,k,d,q)}, J\right], \\ J = \sqrt{E\left\{\chi_{\epsilon\notin\Xi_n} C_2\sigma^2\Theta_{(n)}^2\right\}} \leq C_2\sigma\sqrt{\mathrm{Prob}^{\frac{1}{2}}\{\epsilon \notin \Xi_n\}\left(E\left\{\Theta_{(n)}^4\right\}\right)^{\frac{1}{2}}} \\ \leq C_2\sigma\mathrm{Prob}^{\frac{1}{4}}\{\epsilon \notin \Xi_n\}\left(E\left\{\Theta_{(n)}^4\right\}\right)^{\frac{1}{4}} \leq C_9\sigma n^{-(\mu+1)}\sqrt{\ln n} \end{array} \tag{149}$$

(we have used (39) and (42)). Thus, under assumptions (140) for all $d < p < \infty$ and all q, $\frac{2k+d}{d}p \leq q < \infty$ we have

$$\begin{array}{l} \sqrt{E\left\{\|\widehat{f}_n(\cdot;y) - f(\cdot)\|_{q,B_\gamma}^2\right\}} \\ \leq C_8 \max\left[\widehat{\sigma}_n D^{\frac{d(2\theta-1)}{2}}(B), P_4 R\left(\frac{\widehat{\sigma}_n}{R}\right)^{2\beta(p,k,d,q)}, C_9\sigma n^{-(\mu+1)}\sqrt{\ln n}\right]. \end{array} \tag{150}$$

Now, it is easily seen that if $P \geq 1$ is a properly chosen function of μ, d, γ, p nonincreasing in $p > d$ and (50) takes place, then, first, assumption (140) is satisfied and, second, the right hand side in (150) does not exceed the quantity

$$PR\left(\frac{\widehat{\sigma}_n}{R}\right)^{2\beta(p,k,d,q)} = PR\left(\frac{\widehat{\sigma}_n}{R}\right)^{2\beta(p,k,d,q)} D^{d\lambda(p,k,d,q)}(B)$$

(see (51) and take into account that we are in the situation $q \geq \frac{2k+d}{d}p$, so that $\lambda(p,k,d,q) = 0$). We have obtained bound (51) for the case of $d < p < \infty$, $\infty > q \geq \frac{2k+d}{d}p$; passing to limit as $q \to \infty$, we get the desired bound for $q = \infty$ as well.

4^0. Now let $d < p < \infty$ and $1 \leq q \leq q_* \equiv \frac{2k+d}{d}p$. By Hölder inequality and in view of $mD(B) \geq 1$ we have $\|g\|_{q,B_\gamma} \leq C_9\|g\|_{q*,B_\gamma}|B_\gamma|^{\frac{1}{q}-\frac{1}{q*}}$, whence $\widehat{\mathbf{R}}_q\left(\widehat{f}_n;\mathbf{F}\right) \leq C_9\widehat{\mathbf{R}}_{q*}\left(\widehat{f}_n;\mathbf{F}\right)D^{d\left(\frac{1}{q}-\frac{1}{q*}\right)}(B)$. Combining this observation with the (already proved) bound (51) associated with $q = q_*$, we see that (51) is valid for all $q \in [1,\infty]$, provided that $d < p < \infty$. Passing in the resulting bound to limit as $p \to \infty$, we conclude that (51) is valid for all $p \in (d,\infty]$, $q \in [1,\infty]$. □

References

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *2nd internat. Sympos. Inform. Theory*, Tsahkadsor, 1973.

[2] A. Barron, L. Birg, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields*, 113(3):301–413, 1999.

[3] O.V. Besov, V.P. Il'in, and S.M. Nikol'ski. *Integral representations of functions and embedding theorems.* Moscow: Nauka Publishers, 1975 (in Russian).

[4] L. Birgé. Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 65:181–237, 1983.

[5] L. Birgé, P. Massart. From model selection to adaptive estimation. In: D. Pollard, E. Torgersen and G. Yang, Eds., *Festschrift for Lucien Le Cam*, Springer 1999, 55–89.

[6] L. Birg. An alternative point of view on Lepski's method. Technical Report 549, Paris 6, december 1999.

[7] L. Birg and P. Massart. Gaussian model selection. *J. Eur. Math. Soc.*, 3(3):203–268, 2001.

[8] T. Tony Cai. Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Stat.*, 27(3):898–924, 1999.

[9] E. Candés, D. Donoho. Recovering edges in ill-posed inverse problems: optimality of curvelet frames. *Ann. Statist.* 30(3):784-842, 2002.

[10] S. Chen, D.L. Donoho, M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* 20(1):33-61, 1998.

[11] D. Donoho, I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425-455, 1994.

[12] D. Donoho, I. Johnstone. Minimax risk over ℓ_p balls for ℓ_q losses. *Probab. Theory Related Fields* 99:277-303, 1994.

[13] D. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmon. Anal.* 1:100–115, 1994.

[14] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *J. R. Stat. Soc., Ser. B*, 57(2):301–369, 1995.

[15] M. Elad, A. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. on Information Theory* (to appear)

[16] D. Donoho, I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* 90(432):1200–1224, 1995.

[17] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard. Wavelet shrinkage: Asymptopia? (with discussion and reply by the authors). *J. Royal Statist. Soc. Series B* 57(2):301–369, 1995.

[18] D. Donoho. Tight frames of k-plane ridgelets and the problem of representing objects that are smooth away from d-dimensional singularities in $\mathbb{R}^n$. *Proc. Natl. Acad. Sci. USA* 96(5):1828-1833, 1999.

[19] E. Candés, D. Donoho. Ridgelets: a key to high-dimensional intermittency? *Philos. Trans. Roy. Soc. London Ser. A* 357:2495-2509, 1999.

[20] D. Donoho. Wedgelets: nearly minimax estimation of edges. *Ann. Statist.* 27:859-897, 1999.

[21] D. Donoho. Orthonormal ridgelets and linear singularities. *SIAM J. Math. Anal.* 31:1062-1099, 2000.

[22] D. Donoho. Ridge functions and orthonormal ridgelets. *J. Approx. Theory* 111(2):143-179, 2001.

[23] D. Donoho. Curvelets and curvilinear integrals. *J. Approx. Theory* 113(1):59-90, 2001.

[24] D. Donoho. Sparse components of images and optimal atomic decompositions. *Constr. Approx.* 17:353-382, 2001.

[25] D. Donoho, X. Huo. Uncertainty principle and ideal atomic decomposition. *IEEE Trans. on Information Theory* 47(7):2845-2862, 2001.

[26] D. Donoho, X. Huo. Beamlets and multiscale image analysis. *Lect. Comput. Sci. Eng.* 20:149-196, Springer, 2002.

[27] A. Goldenshluger, A. Nemirovski. On spatially adaptive estimation of nonparametric regression. *Math. Methods of Statistics* 6(2):135–170, 1997.

[28] A. Goldenshluger, A. Nemirovski. Adaptive de-noising of signals satisfying differential inequalities. *IEEE Trans. on Information Theory* 43, 1997.

[29] Yu. Golubev. Asymptotic minimax estimation of regression function in additive model. *Problemy peredachi informatsii* 28(2):3–15, 1992. (English transl. in *Problems Inform. Transmission* 28, 1992.)

[30] W. Härdle. *Applied Nonparametric Regression*, ES Monograph Series 19, Cambridge, U.K., Cambridge University Press, 1990.

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

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

[33] A. Juditsky. Wavelet estimators: Adapting to unknown smoothness. *Math. Methods of Statistics* 6(1):1–25, 1997.

[34] A. Korostelev, A. Tsybakov. *Minimax theory of image reconstruction.* Lecture Notes in Statistics 82, Springer, New York, 1993.

[35] A. Kneip. Ordered linear smoothers. *Ann. Stat.*, 22(2): 835–866, 1994.

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

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

[38] O. Lepskii, E. Mammen, V. Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* 25(3):929–947, 1997.

[39] A. Nemirovski. On forecast under uncertainty. *Problemy peredachi informatsii* 17(4):73–83, 1981. (English transl. in *Problems Inform. Transmission* 17, 1981.)

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

[41] A. Nemirovski. On nonparametric estimation of functions satisfying differential inequalities. R. Khasminski, Ed. *Advances in Soviet Mathematics* 12:7–43, American Mathematical Society, 1992.

[42] A. Nemirovski. *Topics in Non-parametric Statistics.* In: M. Emery, A. Nemirovski, D. Voiculescu, *Lectures on Probability Theory and Statistics, Ecole d'Eteé de Probabilités de Saint-Flour XXVII – 1998, Ed. P. Bernard.* – Lecture Notes in Mathematics 1738:87–285.

[43] M. Pinsker. Optimal filtration of square-integrable signals in Gaussian noise. *Problemy peredachi informatsii*, 16(2):120–133. 1980. (English transl. in *Problems Inform. Transmission* 16, 1980.)

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

[45] B.L.S. Prakasa Rao. *Nonparametric functional estimation.* Academic Press, Orlando, 1983.

[46] M. Rosenblatt. *Stochastic curve estimation.* Institute of Mathematical Statistics, Hayward, California, 1991.

[47] J.-L. Starck, E. Candés, D. Donoho. The curvelet transform for image denoising. *IEEE Trans. Image Process.* 11(6):670-684, 2002.

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

[49] G. Wahba. *Spline models for observational data.* SIAM, Philadelphia, 1990.