

Nonparametric Denoising of Signals with Unknown Local Structure, I: Oracle Inequalities

Anatoli Juditsky^{a,*}, Arkadi Nemirovski^{b,**}

^a*LJK, B.P. 53, 38041 Grenoble Cedex 9, France*

^b*ISyE, Georgia Institute of Technology, 765 Ferst Drive, Atlanta GA 30332 USA*

Abstract

We consider the problem of pointwise estimation of multi-dimensional signals s , from noisy observations (y_τ) on the regular grid \mathbf{Z}^d . Our focus is on the adaptive estimation in the case when the signal can be well recovered using a (hypothetical) linear filter, which can depend on the unknown signal itself.

The basic setting of the problem we address here is as follows: suppose that the signal s is “well-filtered”, i.e. there exists an adopted time-invariant linear filter q_T^* with the coefficients which vanish outside the “cube” $\{0, \dots, T\}^d$ which recovers s_0 from observations with small mean-squared error. We suppose that we do not know the filter q^* , although, we do know that such a filter exists. We give partial answers to the following questions:

- is it possible to construct an adaptive estimator of s_0 , which relies upon observations and recovers s_0 with basically the same estimation error as the unknown filter q_T^* ?
- how rich is the family of well-filtered (in the above sense) signals?

We show that the answer to the first question is affirmative and provide a numerically efficient construction of a nonlinear adaptive filter. Further, we establish a simple calculus of “well-filtered” signals, and show that their family is quite large: it contains, for instance, sampled smooth signals, sampled modulated smooth signals and sampled harmonic functions.

Key words: Nonparametric denoising, oracle inequalities, adaptive filtering.

* Corresponding author.

**Research of this author was partly supported by the NSF Award DMI # 0619977
Email addresses: anatoli.juditsky@imag.fr (Anatoli Juditsky),
nemirovs@isye.gatech.edu (Arkadi Nemirovski).

1 Introduction

In this paper, we focus on the problem of denoising of multi-dimensional signals. 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

$$y_\tau = s_\tau + e_\tau. \tag{1}$$

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.

This “classical” problem is a model example in statistical estimation and signal processing, and as such, has received much attention from statistical and signal processing community. Consequently, numerous approaches to the problem has been proposed in the literature, which differ by the a priori assumptions, imposed on the signal. The study, presented in this paper, fits the nonparametric regression estimation.

When transposed to our case, the “traditional” setting of nonparametric estimation is as follows:

- A. The objective is to recover unknown values $s_i = f(x_\tau)$ of a smooth function $f : [0, 1]^d \rightarrow \mathbb{R}$, which are sampled on the observation grid $\Gamma_n = \{x_\tau = m^{-1}\tau : 0 \leq \tau_1, \dots, \tau_d \leq m\}$ with $(m + 1)^d = n$. The risk of recovery \hat{f} of $f(x_t)$, $t \in \Gamma$ is measured by the expected squared error: $E_f |\hat{f} - f(x_t)|^2$;
- 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 or Sobolev balls;
- C. The focus is on the asymptotic, as the volume of observations n goes to infinity, behavior of the estimation routines, with emphasis on *asymptotically minimax (nearly) optimal estimates* – those which attain (nearly) best possible rates of convergence of the risks to 0 as $n \rightarrow \infty$.

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., [32,49,56,33,4,45,58,31,51,30,37]). Later on, there

was a significant research on *adaptive estimation*. Adaptive estimation methods are free of a priori assumptions on the smoothness parameters of the signal to be recovered, and the primary goal is to develop the routines which exhibit asymptotically (nearly) optimal behavior on a wide variety of families of smooth functions (cf. [50,39–43,12,16,17,34,5,13,28,54,57]). For a more complete overview of results on smooth nonparametric regression estimation see, for instance, [47].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 r and the dimensionality d of the polynomial. The corresponding estimate $\hat{p}^T(x_t)$ 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 $\mathbf{O}_T = \{\tau \in \mathbf{Z}^d : |\tau_j| \leq T, 1 \leq j \leq d\}$:

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

then the estimation \hat{f} of $f(x_t)$ is taken as $\hat{f} = \hat{p}(x_t)$.

Note that the kernel $q^{(T)}$ is readily given by the degree r of the approximating polynomial, T and dimension d . The “classical” adaptation routines takes care of choosing “good” values of the approximation parameters (namely, T and r). On the other hand, the polynomial approximation “mechanism” is supposed to be fixed once for ever. Thus, in those procedures the “form” of the kernel is considered as given in advance.

When posed informally, the problem we are interested in, is as follows: if we consider the form of the kernel as a “free parameter”, is it possible to provide an estimation procedure which is adaptive with respect to this parameter? In other words, suppose that a “good” filter $q_*^{(T)}$, with a small estimation error exists. Then, is it possible to construct a data-driven estimation method which has (almost) the same accuracy as the “oracle” – a hypothetical optimal estimation method which uses the “good” filter $q_*^{(T)}$. It is natural, as it is common in adaptive nonparametric estimation, to measure the quality of an

³ Our “brief outline” of adaptive approach to nonparametric regression would be severely incomplete without mentioning a novel approach aimed at recovering *non-smooth* signals possessing *sparse representations* in properly constructed functional systems [11,18,10,19–23,25,9,55,24,27]. This promising approach is completely beyond the scope of our paper.

adaptive estimation routine with the factor by which the risk of the adaptive procedure is greater than that of the “oracle” estimator. What we look for is the estimation method for which this factor is not too large. Let us consider, for instance, the following question:

(?) Suppose that know that the (deterministic or random) signal $(s_\tau)_{\tau \in \mathbf{Z}^d} \equiv (s_\tau(\xi))_{\substack{\tau \in \mathbf{Z}^d \\ \xi \in \Omega}}$ underlying observations (1) 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 $q^{(T)} \in C_T(\mathbf{Z}^d)$ (i.e., a filter which recovers s_t via $O(T^d)$ observations around the point t) such that

$$E \left\{ \left| s_t - \sum_{\tau \in \mathbf{O}_T} q_\tau^{(T)} y_{t-\tau} \right|^2 \right\} \leq O(\sigma^2 T^{-d}). \quad (2)$$

Can we mimic this filter?

We show that the answer to the question (?) is positive. Namely, 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*.

Several points should be stressed in the above claim. First, we are able to mimic only ideal filters $q_*^{(T)}$ of small l_2 -norm. Indeed, the relation (2) implies that the stochastic term of the error $E \left(\sum_{\tau \in \mathbf{O}_T} q_\tau^{(T)} e_{t-\tau} \right)^2$ is bounded with $O(\sigma^2 T^{-d})$, which is conceivable only if $\|q_*^{(T)}\|_2 = O(T^{-d/2})$. This constraint is crucial, as the price for adaptation becomes prohibitive when the l_2 -norm of the ideal filter is much larger than $O(T^{-d/2})$. Though this assumption seems quite restrictive, the family of well-filtered signals is much wider than that of sampled regular signals. As we shall see later, this family contains also “highly oscillating” sampled modulated smooth signals, sampled harmonic functions, etc. Note also that in our setting, in contrast to the traditional one aimed at recovering smooth signals, *all we are interested in is to recover the values $s_\tau = f(x_\tau)$ on the observation grid only*. This is quite natural, since we intend to handle, as we shall see in the sequel, along with others, also highly oscillating signals. Such a signal, in general, cannot be recovered outside the observation grid. Indeed, a highly oscillating signal can merely vanish on the observation grid and be arbitrarily large outside it (which is clearly impossible for smooth signals).

In this paper we also treat the problem of adaptive prediction, when 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” $\kappa \geq 0$.

The rest of our paper is organized as follows. In Section 2 we give a formal definition of a well-filtered (well-predicted) signal on a d -dimensional regular grid (the latter, w.l.o.g., is normalized to be \mathbf{Z}^d), and then show in Section 3 demonstrate that such a signal can be recovered at a nearly parametric rate *without a priori knowledge of the corresponding “good filter”* (Theorems 4 and 5). The underlying estimation routines (i.e., “Algorithm A” of Section 3.1 and “Algorithm B” of Section 3.2) constitute a substantial extension of the procedures proposed in [44] and of those presented in [48]. In Section 4.1, 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.

To make the exposition more readable, all proofs are collected in the appendix.

The denoising procedures, described in this paper constitute the basic bricks of the construction of adaptive estimators of *locally well-filtered* signals, which we describe in [36]. The results of [36] extend to the wide classes of modulated signals the results of [50,12,16,17,42,34,28] on spatial adaptive estimates of smooth signals.

2 Problem statement

In order to proceed we need some notations.

Fields over \mathbf{Z}^d . Let $C(\mathbf{Z}^d)$ 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 $C(\mathbf{Z}^d)$ by $|r|_{T,p} = \left(\sum_{|\tau| \leq T} |r_\tau|^p \right)^{1/p}$, $|\tau| = \max\{|\tau_1|, \dots, |\tau_d|\}$, with the standard interpretation of the right hand side when $p = \infty$, and we set $|r|_p = \lim_{T \rightarrow \infty} |r|_{T,p} \in \mathbb{R} \cup \{+\infty\}$. A field $r \in C(\mathbf{Z}^d)$ 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* $\text{ord}(r)$ of a filter r ; we write $C_T(\mathbf{Z}^d) = \{r \in C(\mathbf{Z}^d) \mid \text{ord}(r) \leq T\}$. We identify a filter r with the multivariate Laurent sum $r(z_1, \dots, z_d) = \sum_{\tau} r_\tau z_1^{\tau_1} \dots z_d^{\tau_d}$.

- We call a filter r *polynomial*, if the corresponding Laurent sum is a polynomial (i.e., if the entries r_τ vanish when $\tau < 0$). 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 entries r_τ vanish outside the set $k \leq \tau_j \leq T$, $j = 1, \dots, d$.

- We denote by $\Delta_j, j = 1, \dots, d$, the “basic shift operators” on $C(\mathbb{Z}^d)$: $(\Delta_j r)_{\tau_1, \dots, \tau_d} = r_{\tau_1, \dots, \tau_{j-1}, \tau_{j+1}, \dots, \tau_d}$.

- 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 $r(\Delta)x \equiv r(\Delta_1, \Delta_2, \dots, \Delta_d)x$, 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 $F_T : C(\mathbb{Z}^d) \rightarrow C(\Gamma_T^d)$ as $(F_T r)(\mu) = \frac{1}{(2T+1)^{d/2}} \sum_{|\tau| \leq T} r_{\tau} \mu_1^{\tau_1} \dots \mu_d^{\tau_d} \equiv \frac{1}{(2T+1)^{d/2}} r(\mu)$, $r \in C_T(\mathbf{Z}^d)$, where $\mu \in \Gamma_T^d$. Note that $r_{\tau} = \frac{1}{(2T+1)^{d/2}} \sum_{\mu \in \Gamma_T^d} (F_T r)(\mu) \mu_1^{-\tau_1} \dots \mu_d^{-\tau_d}$, $\forall (\tau : |\tau| \leq T)$. 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 $p = \infty$.

Now it is time to give a precise meaning to the basic question (?) of Introduction. In order to do this, 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 (2) holds true.

2.1 Nice signals

Since the observation noises are independent of (s_{τ}) , we have

$$E \left\{ |s_{\tau} - (q(\Delta)y)_{\tau}|^2 \right\} = 2\sigma^2 |q|_2^2 + E_{\xi} \left\{ |s_{\tau}(\xi) - (q(\Delta)s(\xi))_{\tau}|^2 \right\}; \quad (3)$$

therefore in order to ensure (2), 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 $\theta(2T + 1)^{-d/2}$:

$$\max_{\tau:|\tau-t|\leq L} \left[E \left\{ |s_\tau - (q(\Delta)s)_\tau|^2 \right\} \right]^{1/2} \leq \theta(2T + 1)^{-d/2}. \quad (4)$$

(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 above definition 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.¹ Another interesting problem is that 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 $\tau_j \leq t_j - \kappa, j = 1, \dots, d$.

Definition 2 Let $\theta \geq 0, \rho \geq 1$ be reals, let $T_0 \geq \kappa$ 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: $(s_\tau) \in \mathbf{Q}_{\kappa, L}^t(\theta, \rho, T)$), if there exists a filter $q = q^{(T)} \in P_T^\kappa(\mathbf{Z}^d)$, $|q|_2 \leq \frac{\rho}{(2T+1)^{d/2}}$, which reproduces (s_τ) in the box $\{\tau : |\tau - t| \leq L\}$ with the mean square error not exceeding $\theta(2T + 1)^{-d/2}$:

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

(2) [well-predicted signals] We say that (s_τ) is well-predicted, with the parameters $\theta, \rho, \kappa, T_0, L$, at the point t (notation: $(s_\tau) \in \mathbf{P}_{\kappa, T_0, L}^t(\theta, \rho)$), 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 3 Note that the quantitative description of a well-predicted field, when 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 $q(z) \equiv 1$.

In the sequel, we qualify as *nice* a signal which fulfils the requirements of Definition 2 or 1 above. The filters $q^{(T)}$ associated, in the sense of the above

¹ To be more precise, in the filtering literature this case is referred to as *interpolation*.

definitions, with a nice signal (s_τ) as to filters *certifying* the “niceness” (“well-filterability” of “well-predictability”) of the signal.

We are about to demonstrate that in the framework, suggested by the above definitions, the answer to the question (?) is affirmative. I.e., a signal which is nice (T -well-filtered or T -well-predicted, with parameters $\theta, \rho, L = 3T$) at a point t can be recovered at this point “at a nearly parametric rate” with *no a priori knowledge of the corresponding “good filter”*; all we should know in advance are the parameters ρ and T .

3 Main result

We start the recovering routine for the adaptive filtering problem.

3.1 Adaptive filtering

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 $\hat{s}_t[T, y]$ of s_t via observations (y_τ), $|\tau - t| \leq 4T$, as follows:

- (1) When $T = 0$, we merely set $\hat{s}_t[0, y] = y_t$
- (2) When $T > 0$, we set $\hat{s}_t[T, y] = (\hat{\phi}^t(\Delta)y)_t$, where $\hat{\phi}^t \in C_{2T}(\mathbf{Z}^d)$ is an optimal solution to the following optimization problem:

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

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

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

Theorem 4 Assume that the signal (s_τ) underlying observations (1) is T -well-filtered, with parameters $\theta, \rho, L \geq 3T$: $(s_\tau) \in \mathbf{S}_L^t(\theta, \rho, T)$ with $L \geq 3T$. Then the mean square error of the estimate $\hat{s}_t[T, \cdot]$ of s_t yielded by Algorithm A with setup (ρ, T) can be bounded from above as follows:

$$\left(E \left\{ |\hat{s}_t[T, y] - s_t|^2 \right\} \right)^{1/2} \leq c(d) \rho^3 \frac{\theta + \sigma \rho \sqrt{\ln(2T + 1) + 1}}{(2T + 1)^{d/2}}, (7)$$

$$c(d) = 3(2^d + 2^{3d-1}).$$

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

$$\begin{aligned} |s_t - \widehat{s}_t[T, y]| &\leq c(d)\rho^3 [\theta + \sigma\rho\Theta_T^t] (2T + 1)^{-d/2}, \\ \Theta_T^t &= \sigma^{-1} \max_{\tau: |\tau| \leq 2T} |\Delta_1^{\tau_1 - t_1} \dots \Delta_d^{\tau_d - t_d} e|_{2T, \infty}^*. \end{aligned} \quad (8)$$

Comments: note that Theorem 4 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 $O(\sigma(2T + 1)^{-d/2})$:

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

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 4 says that Algorithm A basically achieves this goal. Indeed, from (3), (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 4 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 4 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 [38], [2], [6], [7], [12], [15], [8] among many others). In that context one refer to the filter q , which certifies the niceness 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.

Note that the “upper bound” of Theorem 4 may be compared to the lower bound of Theorem 2 of [35] for the 1-dimensional situation. The latter result states that one can exhibit a family of signals which 1) each member of the family can be recovered with the rate $O(\frac{\sigma\rho}{\sqrt{T}})$ using the corresponding certifying filter; 2) the rate of estimation of signals from the family using the observation (1) is at best $O(\sigma\rho^2 \sqrt{\frac{\ln T}{T}})$. In other words, it states that the factor $\rho\sqrt{\ln(2T + 1)}$ is an unavoidable “price” for adaptation. When comparing the

result of Theorem 4 to that lower bound, 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.

3.2 Adaptive prediction

We now turn to the problem of adaptive prediction. 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) , $\kappa \leq t_j - \tau_j \leq 4T$, $j = 1, \dots, d$, as $\widehat{s}_t[\kappa, T, y] = (\widehat{\psi}^t(\Delta)y)_t$, where $\widehat{\psi}^t \in P_{2T}^\kappa(\mathbf{Z}^2)$ is an optimal solution to the following optimization problem:

$$\min_{\psi \in C_{2T}^\kappa(\mathbf{Z}^d)} \left\{ \underbrace{|\Delta_1^{-t_1} \dots \Delta_d^{-t_d} (1 - \psi(\Delta)) y|_{2T, \infty}^*}_{J(\psi, y_{\kappa, 4T}^t)} : |\psi|_{2T, 1}^* \leq \frac{2^{d/2} \rho^2}{(2T + 1)^{d/2}} \right\}; \quad (10)$$

where $y_{\kappa, L}^t = \{y_\tau : \kappa \leq t_j - \tau_j \leq L, j = 1, \dots, d\}$.

Note that the objective in (10) is affected only by observations $y_{\kappa, 4T}^t$, so that our algorithm recovers s_t via $(4T - \kappa + 1)^d$ observations “around” the point t .

Theorem 5 Assume that the signal (s_τ) underlying observations (1) is T -well-predicted, with parameters $\theta, \rho, \kappa, L \geq 3T$: $(s_\tau) \in \mathbf{Q}_{\kappa, L}^t(\theta, \rho, T)$ with $L \geq 3T$. Then the mean square error of the estimate $\widehat{s}_t[\kappa, T, \cdot]$ of s_t , provided by Algorithm B with setup (ρ, κ, T) , can be bounded from above as follows:

$$\left(E \left\{ |\widehat{s}_t[\kappa, T, y] - s_t|^2 \right\} \right)^{1/2} \leq c(d) \rho^3 \frac{\theta + \sigma \rho \sqrt{\ln(2T + 1) + 1}}{(2T + 1)^{d/2}}, \quad (11)$$

$$c(d) = 3(2^d + 2^{3d-1}).$$

In particular, if (s_τ) is well-predicted, with the parameters $\theta, \rho, \kappa, T_0, L$, 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 (11).

Finally, in the case of deterministic (s) , we have

$$|s_t - \widehat{s}_t[T, y]| \leq c(d) \rho^3 [\theta + \sigma \rho \Theta_T^t] (2T + 1)^{-d/2}, \quad (12)$$

$$\Theta_T^t = \sigma^{-1} \max_{\tau: |\tau| \leq 2T} |\Delta_1^{\tau_1 - t_1} \dots \Delta_d^{\tau_d - t_d} e|_{2T, \infty}^*.$$

The proof of Theorem 5 is identical to that of Theorem 4.

4 Families of nice signals

When applying Algorithms A, B and Theorems 4, 5, the crucial question is how to recognize niceness. We are about to give a partial answer to this question.

4.1 Calculus of nice signals

Our current goal is to understand how wide are the families of nice signals, and our plan is as follows: (a) we list a number of operations which preserve the property in question, and (b) we present a list of examples of signals possessing the property. Applying to “raw materials” from (b) operations from (a), one can produce a wide variety of nice signals. Here is a sample of operations preserving niceness of signals.

I. “Scale” of nice signals. We start with the following evident observation: $\rho' \geq \rho, \theta' \geq \theta, L' \leq L \Rightarrow \mathbf{F}_L^t(\theta, \rho) \subset \mathbf{F}_{L'}^t(\theta', \rho')$ and $\rho' \geq \rho, \theta' \geq \theta, \kappa' \leq \kappa, T'_0 \geq T_0, L' \leq L \Rightarrow \mathbf{P}_{\kappa, T_0, L}^t(\theta, \rho) \subset \mathbf{P}_{\kappa', T'_0, L'}^t(\theta', \rho')$.

II. Taking linear combinations. Our next observation is that a linear combination of “good” signals is again good, with properly updated parameters:

Proposition 6 (i) Let $(s_\tau^j) \in \mathbf{F}_L^t(\theta_j, \rho_j)$, and let $\lambda^j \in \mathbb{C}$ be random variables independent of (s_τ^j) and such that $E\{|\lambda_j|^2\} < \infty$, $j = 1, \dots, m$. Then

$$\begin{aligned} (s_\tau &\equiv \sum_{j=1}^m \lambda_j s_\tau^j) \in \mathbf{F}_{L^+}^t(\theta^+, \rho^+), \\ \theta^+ &= (2m-1)^{d/2} 2^{m-1} \rho_1 \dots \rho_m \sum_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j}, \\ \rho^+ &= (2m-1)^{d/2} 2^m \rho_1 \dots \rho_m, L^+ = \lfloor L/2 \rfloor. \end{aligned} \tag{13}$$

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 .

(ii) Let $(s_\tau^j) \in \mathbf{P}_{\kappa_j, T_0^j, L}^t(\theta_j, \rho_j)$, $j = 1, \dots, m$, and let $\lambda^j \in \mathbb{C}$ be random variable

independent of (s^j) and such that $E\{|\lambda_j|^2\} < \infty$, $j = 1, \dots, m$. Then

$$\begin{aligned}
(s_\tau &\equiv \sum_{j=1}^m \lambda_j s_\tau^j) \in \mathbf{P}_{\kappa^+, L^+}^t(\theta^+, \rho^+), \\
\theta^+ &= (2m-1)^{d/2} 2^{m-1} \rho_1 \dots \rho_m \sum_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j}, \\
\rho^+ &= (2m-1)^{d/2} 2^m \rho_1 \dots \rho_m, \kappa^+ = \min_{1 \leq j \leq m} \kappa_j, T_0^+ = m \max_{1 \leq j \leq m} T_0^j, \\
L^+ &= \lfloor L/2 \rfloor.
\end{aligned} \tag{14}$$

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. Next we notice that the families of nice signals are closed w.r.t. “modulation” and conjugation:

Proposition 7 (i) Let $(s_\tau) \in \mathbf{F}_L^t(\theta, \rho)$, and let $\omega \in \mathbb{R}^d$, $\phi \in \mathbb{R}$ be deterministic. Then the signal $(\hat{s}_\tau = \exp\{i[\omega^T \tau + \phi]\} s_\tau)_{\tau \in \mathbf{Z}^d}$ belongs to $\mathbf{F}_L^t(\theta, \rho)$ along with (s_τ) , and the signal $(\bar{s}_\tau = \overline{s_\tau})_\tau$ (\bar{a} is the complex conjugate of $a \in \mathbb{C}$) belongs to $\mathbf{F}_L^t(\theta, \rho)$.

(ii) Let $(s_\tau) \in \mathbf{P}_{\kappa, T_0, L}^t(\theta, \rho)$, and let $\omega \in \mathbb{R}^d$, $\phi \in \mathbb{R}$ be deterministic. Then the signal $(\hat{s}_\tau = \exp\{i[\omega^T \tau + \phi]\} s_\tau)_{\tau \in \mathbf{Z}^d}$ also belongs to $\mathbf{P}_{\kappa, T_0, L}^t(\theta, \rho)$, and the signal $(\bar{s}_\tau = \overline{s_\tau})_\tau$ belongs to $\mathbf{P}_{\kappa, T_0, L}^t(\theta, \rho)$.

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

Proposition 8 (i) 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 $(s_{\tau_1, \dots, \tau_{d^+}}^+ = s_{\tau_1, \dots, \tau_d})$ 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 $(t_1^+, \dots, t_d^+) = t$.

(ii) Let $1 \leq d \leq d^+$, and let $(s_\tau)_{\tau \in \mathbf{Z}^d}$ be a signal which is well-predictable, with parameters $\theta, \rho, \kappa, T_0, L$, at a point $t \in \mathbf{Z}^d$. Then the signal $(s_{\tau_1, \dots, \tau_{d^+}}^+ = s_{\tau_1, \dots, \tau_d})$ 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 $(t_1^+, \dots, t_d^+) = t$.

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 $(s'_{\tau'}(\xi))_{\tau' \in \mathbf{Z}^{d'}, \xi}$, $(s''_{\tau''}(\xi))_{\tau'' \in \mathbf{Z}^{d''}, \xi}$, we define their tensor product as the field $(s_\tau(\xi) = s'_{\tau'}(\xi) s''_{\tau''}(\xi))_{\tau = (\tau', \tau''), \xi \in \mathbf{Z}^d}$.

Proposition 9 (i) Let $(s'_{\tau'}(\xi))_{\tau' \in \mathbf{Z}^{d'}, \xi} \in \mathbf{F}_L^{t'}(0, \rho')$, $(s''_{\tau''}(\xi))_{\tau'' \in \mathbf{Z}^{d''}, \xi} \in \mathbf{F}_L^{t''}(0, \rho'')$.

Then $(s_\tau) \in \mathbf{F}_L^{(t', t'')}(0, \rho' \rho'')$.

(ii) Let $(s'_{\tau'}(\xi))_{\tau' \in \mathbf{Z}^{d'}} \in \mathbf{P}_{\kappa, T_0, L}^{t'}(0, \rho')$, $(s''_{\tau''}(\xi))_{\tau'' \in \mathbf{Z}^{d''}} \in \mathbf{P}_{\kappa, T_0, L}^{t''}(0, \rho'')$. Then

$(s_\tau) \in \mathbf{P}_{\kappa, T_0, L}^{(t', t'')}(0, \rho' \rho'')$.

4.2 Examples of nice 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 $c\tau^\alpha \exp\{\omega^T \tau\} \equiv c\tau_1^{\alpha_1} \dots \tau_d^{\alpha_d} \exp\{\omega^T \tau\}$ with nonnegative multi-indices α and $\omega \in \mathbb{C}^d$:

$$s_\tau = \sum_{\ell=1}^M c_\ell \tau^{\alpha(\ell)} \exp\{\omega^T(\ell) \tau\}, \quad (15)$$

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 , $j = 1, \dots, d$, as follows: let m_j be the maximum of the degrees $\alpha_j(\ell)$, $\ell = 1, \dots, M$, of the variable τ_j in the monomials of the sum (15), and M_j be the number of distinct from each other complex numbers among the “partial frequencies” $\omega_j(\ell)$: $M_j = \text{Card } \mathbf{O}_j$, $\mathbf{O}_j = \{\omega_j(\ell) : 1 \leq \ell \leq M\}$. The *j-th partial size* $N_j(s)$ of exponential polynomial (15) 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 10 *Let (s_τ) be an exponential polynomial on \mathbf{Z}^d of partial sizes N_1, \dots, N_d . Then for all $t \in \mathbf{Z}^d$ one has*

$$(s_\tau) \in \mathbf{F}_\infty^t(0, \rho_d(N_1, \dots, N_d)), \quad \rho_d(N_1, \dots, N_d) = \prod_{j=1}^d [(2N_j - 1)^{1/2} 2^{3N_j/2}], \quad (16)$$

and the filters $q^{(T)}$ certifying this inclusion can be chosen to be dependent 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 11 *A major shortcoming of (16) is a dramatic growth of $\rho_d(N, N, \dots, N)$ 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*

$$p_\tau = \sum_{\alpha \geq 0, |\alpha| \leq m} c_\alpha \tau^\alpha \quad (17)$$

belongs to $\mathbf{F}_\infty^t(0, (16m)^d)$ 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} :

$$(\mathbf{D}f)_\tau = \sum_{\ell=1}^k w_\ell f_{\tau-\alpha(\ell)}; \quad (18)$$

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

$$B_N^t = \{\tau \in \mathbf{Z}^d \mid |\tau - t| \leq N\}, \quad B_N^t(\mathbf{D}) = \{\tau \in B_N^t \mid \tau + \alpha(\ell) \in B_N^t, \ell = 1, \dots, k\},$$

$$\mathbf{H}_N^t(\mathbf{D}) = \{(s) \in C(\mathbf{Z}^d) \mid s_\tau = (\mathbf{D}s)_\tau \quad \forall \tau \in B_N^t(\mathbf{D})\}.$$

For example, with

$$(\mathbf{D}f)_\tau = \frac{1}{2d} \sum_{\substack{i=1, \dots, d \\ \epsilon=\pm 1}} f_{\tau_1, \dots, \tau_{i-1}, \tau_i+\epsilon, \tau_{i+1}, \dots, \tau_d}, \quad (19)$$

the linear space $\mathbf{H}_N^t(\mathbf{D})$ is the space of fields which are “discrete harmonic” on B_N^t , that is, $s_\tau = \frac{1}{2d} \sum_{\substack{i=1, \dots, d \\ \epsilon=\pm 1}} s_{\tau_1, \dots, \tau_{i-1}, \tau_i+\epsilon, \tau_{i+1}, \dots, \tau_d}$ 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

$$(a) \sum_{\ell=1}^k \rho_\ell \leq 1; \quad (b) \sum_{\ell=1}^k \rho_\ell \alpha(\ell) = 0. \quad (20)$$

For example, the averaging operator (19) and its degrees are regular.

It turns out that the solutions of homogeneous difference equations with regular difference operators are well-filtered:

Proposition 12 *Let \mathbf{D} be a regular difference operator. Then there exists a constant $c = c(\mathbf{D}) > 0$ such that*

$$\forall N > 0: \quad \mathbf{H}_N^t(\mathbf{D}) \subset \mathbf{F}_{\lfloor cN \rfloor}^t(0, c^{-1}). \quad (21)$$

As a nontrivial application example for Proposition 12, consider the families of random fields defined as follows. Let $d \leq 4$, 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 $D_{2M}^0 = \{x \in \mathbb{R}^d : |x_j| \leq 2M, j \leq d\}$: $\left(\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}\right) f(x) = 0, x \in \text{int}D_{2M}^0$. 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, $E\{\|f\|_{\infty, 2M}^2\} \leq R^2$, where $\|f\|_{\infty, 2M}$ is the uniform norm on D_{2M}^0 . 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 13 *Let $d \leq 4$, 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*

$$|t| \leq cM, L \leq cM \Rightarrow (s_\tau) \in \mathbf{F}_L^t(c^{-1}R, c^{-1}), \quad (22)$$

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

4.3 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

$$s_\tau = \sum_{\ell=1}^M c_\ell \tau^{\alpha(\ell)} \exp\{\omega^T(\ell)\tau\} \quad (23)$$

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

Proposition 14 *Let (s_τ) be a quasi-stable exponential polynomial on \mathbf{Z}^d of partial sizes N_1, \dots, N_d . Then for every integer $\kappa \geq 0$ and all $t \in \mathbf{Z}^d$ one has*

$$\begin{aligned} (s_\tau) &\in \mathbf{P}_{\kappa, T_0, \infty}^t(0, \rho_{\kappa, d}(N_1, \dots, N_d)), \\ \rho_{\kappa, d}(N_1, \dots, N_d) &= \prod_{j=1}^d [(2N_j - 1)^{1/2} 2^{N_j} (\max[2, 2\kappa + 1])^{N_j/2}], \\ T_0 &= \kappa \max_{1 \leq j \leq d} N_j \end{aligned} \quad (24)$$

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.

5 Appendix

5.1 Preliminaries

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

- [Parseval equality]:

$$(r, s)_T \equiv \sum_{t:|t|\leq T} r_t \bar{s}_t = \sum_{\mu \in \Gamma_T^d} (F_T r)(\mu) \overline{(F_T s)(\mu)} \equiv \langle F_T r, F_T s \rangle_T, \quad (25)$$

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

$$|r|_{T,2} = |r|_{T,2}^*; \quad (26)$$

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

$$\left| \sum_{|t|\leq T} a_t b_t \right| \leq |a|_{T,1}^* |b|_{T,\infty}^*. \quad (27)$$

- [Norms of convolutions of filters]

$$r, s \in C(\mathbb{Z}^d) \Rightarrow |r(z_1, \dots, z_d) s(z_1, \dots, z_d)|_p \leq |r|_1 |s|_p; \quad (28)$$

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

$$|r|_{T,p}^* \leq (2T+1)^{d[(1/p-1/2)_+ + (1/2-1/q)_+]} |r|_{T,q}, \quad a_+ = \max[a, 0]; \quad (29)$$

$$\text{ord}(r) + \text{ord}(s) \leq T \Rightarrow |r(z_1, \dots, z_d) s(z_1, \dots, z_d)|_{T,p}^* \leq |r|_1 |s|_{T,p}^*. \quad (30)$$

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

Lemma 15 *Let $f_j = \xi_j + i\eta_j$, $0 \leq j < N$, be a sequence of N standard Gaussian complex-valued random variables, not necessarily independent of each other. Then*

$$\begin{aligned} [E\{\max_{0 \leq j < N} |f_j|^2\}]^{1/2} &\leq \sqrt{2 \ln N + 2}; \\ P\{\max_{0 \leq j < N} |f_j| > u + \sqrt{2 \ln N}\} &\leq \exp\{-u^2/2\} \quad \forall u \geq 0. \end{aligned} \quad (31)$$

We have

$$\begin{aligned}
\psi(r) &\equiv P\{\max_{0 \leq j < N} |f_j| > r\} \leq \min[1, N \exp\{-r^2/2\}] \Rightarrow \\
P\{\max_{0 \leq j < N} |f_j| > u + \sqrt{2 \ln N}\} &\leq N \exp\{-(u + \sqrt{2 \ln N})^2/2\} \leq \exp\{-u^2/2\}; \\
E\{\max_{0 \leq j < N} |f_j|^2\} &= -\int_0^\infty r^2 d\psi(r) = 2 \int_0^\infty r \psi(r) dr \leq 2 \int_0^{\sqrt{2 \ln N}} r dr \\
&+ 2N \int_{\sqrt{2 \ln N}}^\infty r \exp\{-r^2/2\} dr = 2 \ln N + 2.
\end{aligned}$$

5.2 Proof of Theorem 4

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 $(s_\tau) \in \mathbf{S}_{3T}^0(\theta, \rho, T)$. Let us set

$$|q^*|_2 = \hat{\rho}(2T + 1)^{-d/2}; \quad \kappa = \theta(2T + 1)^{-d/2} \quad [\hat{\rho} \leq \rho], \quad (32)$$

so that

$$\bar{s} = q^*(\Delta)s \Rightarrow \max_{\tau: |\tau| \leq 3T} E\{|s_\tau - \bar{s}_\tau|^2\} \leq \kappa^2. \quad (33)$$

Finally, let

$$\Theta_T = \max_{\tau: |\tau| \leq 2T} |\Delta_1^{\tau_1} \dots \Delta_d^{\tau_d} e|_{2T, \infty}^*, \quad (34)$$

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

I^0 . We start with simple technical lemma:

Lemma 16 *Let $r(z_1, \dots, z_d) = (q^*(z_1, \dots, z_d))^2$. Then $r \in C_{2T}(\mathbf{Z}^d)$ possesses the following properties:*

$$|r|_2 \leq |r|_{2T, 1}^* \leq 2^{d/2} \hat{\rho}^2 (2T + 1)^{-d/2}; \quad (35)$$

$$|r|_1 \leq \hat{\rho}^2; \quad (36)$$

$$\left[E\{|(1 - r(\Delta))s|_{2T, 2}^2\} \right]^{1/2} \leq \kappa(\hat{\rho} + 1)(4T + 1)^{d/2}; \quad (37)$$

$$|(1 - r(\Delta))y|_{2T, \infty}^* \leq |(1 - r(\Delta))s|_{2T, 2} + (1 + \hat{\rho}^2)\Theta_T \quad (38)$$

$$\begin{aligned}
\left[E\left\{ \left(|(1 - r(\Delta))y|_{2T, \infty}^* \right)^2 \right\} \right]^{1/2} &\leq \sigma(1 + \hat{\rho}^2) \sqrt{4d \ln(4T + 1) + 2} \\
&+ \kappa(\hat{\rho} + 1)(4T + 1)^{d/2}.
\end{aligned} \quad (39)$$

(40)

Proof. (35): We have

$$\begin{aligned} |r|_{2T,1}^* &= \sum_{\mu \in \Gamma_{2T}^d} \frac{|r(\mu)|}{(4T+1)^{d/2}} = \sum_{\mu \in \Gamma_{2T}^d} \frac{|q^*(\mu)|^2}{(4T+1)^{d/2}} = (4T+1)^{d/2} \sum_{\mu \in \Gamma_{2T}^d} \left| \frac{q^*(\mu)}{(4T+1)^{d/2}} \right|^2 \\ &= (4T+1)^{d/2} (|q^*|_{2T,2}^*)^2 = (4T+1)^{d/2} |q^*|_{2T,2}^2 \leq 2^{d/2} \hat{\rho}^2 (2T+1)^{-d/2}. \end{aligned}$$

Since $|r|_2 = |r|_{2T,2} = |r|_{2T,2}^* \leq |r|_{2T,1}^*$, (35) follows.

(36): We clearly have $|r|_1 \leq |q^*|_1^2 \leq ((2T+1)^{d/2} |q^*|_2)^2 = \hat{\rho}^2$.

(37): 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

$$\max_{\tau: |\tau| \leq 3T} E\{|h_\tau|^2\} \leq \kappa^2. \quad (41)$$

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

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

applying (41) and taking into account that $|q^*|_2 = \hat{\rho}(2T+1)^{-d/2}$, we come to

$$\max_{\tau: |\tau| \leq 3T} E\{|((1 - r(\Delta))s)_\tau|^2\} \leq [\kappa(\hat{\rho} + 1)]^2, \quad (42)$$

and (37) follows.

(38), (39): We have

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

The resulting inequality combines with (36) to yield (38). Further, from the resulting inequality and (37) it follows that

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

(we have used (36)). To derive (39) from the resulting inequality, it remains to note that

$$\left(E\{\Theta_T^2\} \right)^{1/2} \leq \sigma \sqrt{4d \ln(4T + 1) + 2}. \quad (43)$$

Indeed, the coordinates of the Fourier transform of $\Delta_1^{\tau_1} \dots \Delta_d^{\tau_d} e$ 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 15. \square

2^0 . We now study the properties of the solution $\widehat{\phi}$ of problem (6).

Lemma 17 *One has*

$$|\widehat{\phi}|_{2T,2} \leq 2^{d/2} \rho^2 (2T + 1)^{-d/2}; \quad (44)$$

$$|(1 - \widehat{\phi}(\Delta))e|_{2T,\infty}^* \leq (1 + 2^d \rho^2) \Theta_T; \quad (45)$$

$$\left[E \left\{ \left(|(1 - \widehat{\phi}(\Delta))e|_{2T,\infty}^* \right)^2 \right\} \right]^{1/2} \leq \sigma(1 + 2^d \rho^2) \sqrt{4d \ln(4T + 1) + 2}; \quad (46)$$

$$|(1 - \widehat{\phi}(\Delta))s|_{2T,\infty}^* \leq |(1 - r(\Delta))s|_{2T,2} + 2(1 + 2^d \rho^2) \Theta_T; \quad (47)$$

$$\begin{aligned}
\left[E \left\{ \left(|(1 - \widehat{\phi}(\Delta))s|_{2T,\infty}^* \right)^2 \right\} \right]^{1/2} & \leq 2\sigma(1 + 2^d \rho^2) \sqrt{4d \ln(4T + 1) + 2} \\
& \quad + \kappa(\widehat{\rho} + 1)(4T + 1)^{d/2}. \quad (48)
\end{aligned}$$

Proof. (44): $|\widehat{\phi}|_{2T,2} = |\widehat{\phi}|_{2T,2}^* \leq |\widehat{\phi}|_{2T,1}^* \leq 2^{d/2} \rho^2 (2T + 1)^{-d/2}$, (the concluding inequality comes from the fact that $\widehat{\phi}$ is feasible for (6)).

(45), (46): We have

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

(we have used (44)). The resulting inequality implies that

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

(we have used (43)).

(48), (48): Note that the polynomial r defined in Lemma 16 is a feasible solution of the optimization problem (6) by the first relation in (35), so that the optimal value in the problem does not exceed $J(r, y_{4T}^0)$. It follows that

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

Relation (47) follows from (c) combined with (38) and (45) (recall that $\widehat{\rho} \leq \rho$).

Relation (48) follows from (d) combined with (46) and (39). \square

$\mathfrak{3}^0$. Our next step is to prove

Lemma 18 *One has*

$$\begin{aligned} \left| \left((1 - r(\Delta))(1 - \widehat{\phi}(\Delta))s \right)_0 \right| & \leq \left| \left((1 - r(\Delta))s \right)_0 \right| \\ & \quad + 2^{d/2} \rho^2 (2T + 1)^{-d/2} \left| \left((1 - r(\Delta))s \right)_{2T, 2} \right|; \end{aligned} \quad (49)$$

$$\left[E \left\{ \left| \left((1 - r(\Delta))(1 - \widehat{\phi}(\Delta))s \right)_0 \right|^2 \right\} \right]^{1/2} \leq \kappa(\widehat{\rho} + 1)(2^d \rho^2 + 1). \quad (50)$$

Proof. We have

$$\begin{aligned} & \left| \left((1 - r(\Delta))(1 - \widehat{\phi}(\Delta))s \right)_0 \right| \leq \left| \left((1 - r(\Delta))s \right)_0 \right| + \left| \left(\widehat{\phi}(\Delta)(1 - r(\Delta))s \right)_0 \right| \\ & \leq \left| \left((1 - r(\Delta))s \right)_0 \right| + |\widehat{\phi}|_{2T, 2} \left| \left((1 - r(\Delta))s \right)_{2T, 2} \right| \\ & \leq \left| \left((1 - r(\Delta))s \right)_0 \right| + 2^{d/2} \rho^2 (2T + 1)^{-d/2} \left| \left((1 - r(\Delta))s \right)_{2T, 2} \right| \quad [\text{see (44)}] \end{aligned}$$

as required in (49). From the resulting inequality it follows that

$$\begin{aligned}
& \left[E \left\{ \left| \left((1-r(\Delta))(1-\widehat{\phi}(\Delta))s \right)_0 \right|^2 \right\} \right]^{1/2} \leq \left[E \left\{ \left| \left((1-r(\Delta))s \right)_0 \right|^2 \right\} \right]^{1/2} \\
& + 2^{d/2} \rho^2 (2T+1)^{-d/2} \left[E \left\{ \left| (1-r(\Delta))s \right|_{2T,2}^2 \right\} \right]^{1/2} \\
& \leq \kappa(\widehat{\rho}+1) + 2^{d/2} \rho^2 (2T+1)^{-d/2} \left[E \left\{ \left| (1-r(\Delta))s \right|_{2T,2}^2 \right\} \right]^{1/2} \quad [\text{see (42)}] \\
& \leq \kappa(\widehat{\rho}+1) + 2^{d/2} \rho^2 (2T+1)^{-d/2} \kappa(\widehat{\rho}+1) (4T+1)^{d/2} \quad [\text{see (37)}]
\end{aligned}$$

and (50) follows. \square

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

$$\begin{aligned}
s_0 - \widehat{s}_0 &= s_0 - (\widehat{\phi}(\Delta)y)_0 = \left((1-\widehat{\phi}(\Delta))s \right)_0 - \left(\widehat{\phi}(\Delta)e \right)_0 \equiv \epsilon_0^{(1)} + \epsilon_0^{(2)}, \\
\epsilon_\tau^{(1)} &= \left((1-\widehat{\phi}(\Delta))s \right)_\tau, \quad \epsilon_\tau^{(2)} = \left(\widehat{\phi}(\Delta)e \right)_\tau.
\end{aligned} \tag{51}$$

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

$$\begin{aligned}
|\epsilon_0^{(2)}| &= \left| \sum_{\tau:|\tau|\leq 2T} \widehat{\phi}_\tau e_{-\tau} \right| \leq |\widehat{\phi}|_{2T,1}^* |f|_{2T,\infty}^* \quad [\text{see (27)}] \\
&\leq 2^{d/2} \rho^2 (2T+1)^{-d/2} |f|_{2T,\infty}^*, \quad [\text{since } \widehat{\phi} \text{ is feasible for (6)}]
\end{aligned}$$

whence, by definition of Θ_T ,

$$|\epsilon_0^{(2)}| \leq 2^{d/2} \rho^2 (2T+1)^{-d/2} \Theta_T. \tag{52}$$

Applying (43), we derive from the latter inequality that

$$\left[E \left\{ \left| \epsilon_0^{(2)} \right|^2 \right\} \right]^{1/2} \leq 2^{d/2} \sigma \rho^2 (2T+1)^{-d/2} \sqrt{2d \ln(4T+1) + 2}. \tag{53}$$

We further have

$$\begin{aligned}
|\epsilon_0^{(1)}| &= \left| \left((1-\widehat{\phi}(\Delta))s \right)_0 \right| \leq \left| \left(r(\Delta)(1-\widehat{\phi}(\Delta))s \right)_0 \right| \\
&+ \left| \left((1-r(\Delta))(1-\widehat{\phi}(\Delta))s \right)_0 \right| \underbrace{\leq}_{a} |r|_{2T,1}^* |(1-\widehat{\phi}(\Delta))s|_{2T,\infty}^* \\
&+ \left| \left((1-r(\Delta))(1-\widehat{\phi}(\Delta))s \right)_0 \right| \underbrace{\leq}_{b} 2^{d/2} \rho^2 (2T+1)^{-d/2} |(1-\widehat{\phi}(\Delta))s|_{2T,\infty}^* \\
&+ \left| \left((1-r(\Delta))(1-\widehat{\phi}(\Delta))s \right)_0 \right|
\end{aligned} \tag{54}$$

(a is given by (27), b by feasibility of $\hat{\phi}$ for (6)), whence

$$\begin{aligned}
& \left[E \left\{ |\epsilon_0^{(1)}|^2 \right\} \right]^{1/2} \leq 2^{d/2} \rho^2 (2T+1)^{-d/2} \left[E \left\{ \left(|(1 - \hat{\phi}(\Delta))s|_{2T, \infty}^* \right)^2 \right\} \right]^{1/2} \\
& + \left[E \left\{ \left| (1 - r(\Delta)(1 - \hat{\phi}(\Delta))s \right|_0^2 \right\} \right]^{1/2} \\
& \leq 2^{d/2} \rho^2 (2T+1)^{-d/2} \left[2^d \sigma (1 + 2^d \rho^2) \sqrt{4d \ln(4T+1)} + 2 \right. \\
& \left. + \kappa(\hat{\rho} + 1)(4T+1)^{d/2} \right] + \kappa(\hat{\rho} + 1)(2^d \rho^2 + 1)
\end{aligned} \tag{55}$$

(see (48), (50)). Combining (51), (53), (55), we finally get

$$\begin{aligned}
& \left[E \left\{ |s_0 - \hat{s}_0|^2 \right\} \right]^{1/2} \leq 2^{d/2} \sigma \rho^2 (2T+1)^{-d/2} \sqrt{2d \ln(4T+1)} + 2 \\
& + 2^{d/2} \rho^2 (2T+1)^{-d/2} \left[2^d \sigma (1 + 2^d \rho^2) \sqrt{4d \ln(4T+1)} + 2 \right. \\
& \left. + \kappa(\hat{\rho} + 1)(4T+1)^{d/2} \right] + \kappa(\hat{\rho} + 1)(2^d \rho^2 + 1).
\end{aligned} \tag{56}$$

Recalling that $\hat{\rho} \leq \rho$, $\kappa = \theta(2T+1)^{-d/2}$ and that $\rho \geq 1$, (7) follows.

Now assume that (s) is deterministic. In this case, from (54) combined with (47) and (49) implies that

$$\begin{aligned}
& |\epsilon_0^{(1)}| \leq 2^{1+d/2} \rho^2 (2T+1)^{-d/2} |(1 - r(\Delta))s|_{2T, 2} \\
& + 2^{1+d/2} \rho^2 (1 + 2^d \rho^2) (2T+1)^{-d/2} \Theta_T + |((1 - r(\Delta))s)_0|,
\end{aligned} \tag{57}$$

while from (37), (42) it follows that

$$\begin{aligned}
& |(1 - r(\Delta))s|_{2T, 2} \leq \kappa(\hat{\rho} + 1)(4T+1)^{d/2} \leq 2^{d/2} \theta(1 + \rho), \\
& |((1 - r(\Delta))s)_0| \leq \kappa(1 + \hat{\rho}) \leq \theta(1 + \rho)(2T+1)^{-d/2}.
\end{aligned} \tag{58}$$

Therefore (57) implies that

$$|\epsilon_0^{(1)}| \leq 3^{3+d} \rho^3 [\theta + \rho \Theta_T] (2T+1)^{-d/2}. \tag{59}$$

Combining this relation with (52) and (51), we arrive at (8). \square

5.2.1 Proof of Proposition 6

In the proofs to follow, we focus on the case of well-filtered signals; the reasoning in the case of well-predicted signals is completely similar.

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 $s^j \in \mathbf{F}_L^t(\theta, \rho)$ and clearly $T \leq L$, there exist filters q^j such that

$$\begin{aligned}
(a) &: \text{ord}(q^j) \leq T; \quad (b) : |q^j|_2 \leq \rho_j(2T+1)^{-d/2}; \\
(c) &: |q^j|_1 = |q^j|_{T,1} \leq (2T+1)^{d/2}|q^j|_2 \leq \rho_j; \\
(d) &: [E \{|s_\tau^j - (q^j(\Delta)s_\tau^j)|^2\}]^{1/2} \leq \theta_j(2T+1)^{-d/2} \quad \forall(\tau : |\tau - t| \leq L).
\end{aligned} \tag{60}$$

Now let filter q be defined by

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

Observe that

$$\text{ord}(q) \leq mT \leq T^+. \tag{61}$$

Note that

$$|q|_2 \leq 2^m \rho_1 \dots \rho_m (2T+1)^{-d/2} \leq (2m-1)^{d/2} 2^m \rho_1 \dots \rho_m (2T^++1)^{-d/2}. \tag{62}$$

Indeed, we clearly have

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

(a is by (60.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 (62). Further, by (60.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

$$|Q^j|_1 \leq (1 + \rho_1) \dots (1 + \rho_{j-1}) (1 + \rho_{j+1}) \dots (1 + \rho_m) \leq \frac{2^{m-1} \rho_1 \dots \rho_m}{\rho_j}. \tag{63}$$

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

$$\begin{aligned}
& \left[E \left\{ |(1 - q(\Delta))s_\tau|^2 \right\} \right]^{1/2} = \left[E \left\{ \left| \sum_{j=1}^m \lambda_j (1 - q(\Delta))s^j \right|^2 \right\} \right]^{1/2} \\
& \leq \sum_{j=1}^m \left[E \left\{ |\lambda_j (1 - q(\Delta))s^j|^2 \right\} \right]^{1/2} \stackrel{a}{=} \sum_{j=1}^m [E\{|\lambda_j|^2\}]^{1/2} \left[E \left\{ |(1 - q(\Delta))s^j|^2 \right\} \right]^{1/2} \\
& \leq \sum_{j=1}^m [E\{|\lambda_j|^2\}]^{1/2} \left[E \left\{ |(Q^j(\Delta)(1 - q^j(\Delta))s^j)|^2 \right\} \right]^{1/2} \\
& \leq \underbrace{\sum_{j=1}^m [E\{|\lambda_j|^2\}]^{1/2}}_b |Q^j|_1 \max_{\tau': |\tau' - \tau| \leq (m-1)T} \left[E \left\{ |((1 - q^j(\Delta))s^j)_{\tau'}|^2 \right\} \right]^{1/2} \\
& \leq 2^{m-1} \rho_1 \dots \rho_m (2T + 1)^{-d/2} \sum_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j} \\
& \leq \left[(2m - 1)^{d/2} 2^{m-1} \rho_1 \dots \rho_m \sum_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j} \right] (2T^+ + 1)^{-d/2}
\end{aligned}$$

where a is due to independence of λ_j and (s^j) and b is by (63), (60.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, (61), (62) and taking into account that $T^+ \in \{0, 1, \dots, L^+\}$ is arbitrary, we conclude that $s \in \mathbf{F}_{L^+}^t(\theta^+, \rho^+)$. Note that by construction, the filters certifying the latter inclusion are independent of λ_j . \square

5.2.2 Proof of Proposition 7

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

$$\begin{aligned}
& \text{ord}(q) \leq T, |q|_2 \leq \frac{\rho}{(2T+1)^{d/2}}, \\
& \max_{\tau: |\tau - t| \leq L} \left[E \left\{ |((1 - q(\Delta))s)_\tau|^2 \right\} \right]^{1/2} \leq \frac{\theta}{(2T+1)^{d/2}}.
\end{aligned} \tag{64}$$

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

$$\begin{aligned}
((1 - \hat{q}(\Delta))\hat{s})_\tau &= \exp\{i[\omega^T \tau + \phi]\} s_\tau \\
&\quad - \sum_{\tau'} (\exp\{i\omega^T \tau'\} q_{\tau'}) (\exp\{i[\omega^T (\tau - \tau') + \phi]\} s_{\tau - \tau'}) \\
&= \exp\{i[\omega^T \tau + \phi]\} ((1 - q(\Delta))s)_\tau,
\end{aligned}$$

so that (64) remains valid when $q, (s)$ are replaced with $\hat{q}, (\hat{s})$. Thus, $(\hat{s}) \in \mathbf{F}_L^t(\theta, \rho)$. (i) is proved; (ii) is evident. \square

5.2.3 Proof of Proposition 8

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

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

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

5.2.4 Proof of Proposition 9

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

$$\begin{aligned} (a) : |q'|_2 &\leq \rho'(2T+1)^{-d'/2}, |q''|_2 \leq \rho''(2T+1)^{-d''/2}, \\ s'_{\tau'}(\xi) &= \sum_{\nu'} s'_{\tau'-\nu'} q'_{\nu'}, |\tau' - t'| \leq L, \\ s''_{\tau''}(\xi) &= \sum_{\nu''} s''_{\tau''-\nu''} q''_{\nu''}, |\tau'' - t''| \leq L \end{aligned} \quad (65)$$

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

$$q \in C_T(\mathbf{Z}^d), \quad |q|_2 = |q'|_2 |q''|_2 \leq \rho' \rho'' (2T+1)^{-d/2} \quad (66)$$

(see (65.a)). Now let $\tau = (\tau', \tau'')$ be such that $|\tau - (t', t'')| \leq L$. We have

$$\begin{aligned} (q(\Delta)s(\xi))_\tau &= \sum_{(\nu', \nu'') \in \mathbf{Z}^{d'} \times \mathbf{Z}^{d''}} s'_{\tau'-\nu'}(\xi) s''_{\tau''-\nu''}(\xi) q'_{\nu'} q''_{\nu''} = \sum_{\nu' \in \mathbf{Z}^{d'}} s'_{\tau'-\nu'} q'_{\nu'} s''_{\tau''}(\xi) \\ &= s'_{\tau'}(\xi) s''_{\tau''}(\xi) = s_{(\tau', \tau'')}, \end{aligned}$$

which combines with (66) to yield that $(s_\tau) \in \mathbf{F}_L^{(t', t'')}(0, \rho' \rho'')$. \square

5.2.5 Proof of Proposition 10

We start with the following two evident facts:

Lemma 19 *Let $(s^j) \in C(\mathbf{Z}^d)$ be deterministic fields belonging to $\mathbf{F}_L^t(\theta, \rho)$, $j = 1, 2, \dots$ such that $s_\tau^j \rightarrow s_\tau$, $j \rightarrow \infty$, for every $\tau \in \mathbf{Z}^d$. Then $(s) \in \mathbf{F}_L^t(\theta, \rho)$.*

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

Lemma 20 *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 $\Re(\omega) \geq 0$ and given $T \geq 0$, let us set $q(z) = \frac{1}{T+1}[1 + \exp\{-\omega\}z^{-1} + \exp\{-2\omega\}z^{-2} + \dots + \exp\{-T\omega\}z^{-T}]$. Then $q \in C_T(\mathbf{Z})$, $|q|_2 = (T+1)^{-1/2} \leq 2^{1/2}(2T+1)^{-1/2}$, while clearly $q(\Delta)s \equiv s$. In the case of $\Re(\omega) < 0$, the same reasoning holds true for $q(z) = \frac{1}{T+1}[1 + \exp\{\omega\}z + \exp\{2\omega\}z^2 + \dots + \exp\{T\omega\}z^T]$. \square

To complete the proof, we need the following fact:

Lemma 21 *Let (s_τ) be a “simple” exponential polynomial – a deterministic exponential polynomial of the form $(s_\tau) = \sum_{\ell=1}^M c_\ell \exp\{\omega^T(\ell)\tau\}$. Then*

$$\forall t \in \mathbf{Z}^d : (s_\tau) \in \mathbf{F}_\infty^t(0, \rho_d(N_1, \dots, N_d)), \quad (67)$$

where $\rho_d(\cdot, \dots, \cdot)$ is given by (16) and N_1, \dots, 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 $\mathbf{O}_j = \{\omega_j(\ell) : \ell = 1, \dots, M\}$.

Lemma 21 \Rightarrow *Proposition 10*: Assume first that the coefficients c_ℓ in (15) are deterministic. Since every one of the univariate functions $f(t) = t^k$, $0 \leq k \leq m$, is, uniformly on compact sets, the limit, as $\epsilon \rightarrow +0$, of appropriate linear combinations of the $m+1$ exponents $\exp\{-k\epsilon\}$, the exponential polynomial (15) is the pointwise, on \mathbf{Z}^d , limit, as $i \rightarrow \infty$, of simple exponential polynomials (s_τ^i) 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 $\omega_j(\ell) - k\epsilon_i$, $0 \leq k \leq m_j$. 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 21 and (19), we immediately conclude that the exponential polynomial (15) belongs to $\mathbf{F}_\infty^t(0, \rho_d(N_1, \dots, N_d))$. Since the filters $q^{(T),i}$ certifying well-filterability of the approximating polynomials (s_τ^i) can be chosen to depend solely on T and the sets of partial frequencies of these approximating polynomials, from the proof of Lemma 19 it follows that the filters $q^{(T)}$ certifying the inclusion $(s_\tau) \in \mathbf{F}_\infty^t(0, \rho_d(N_1, \dots, N_d))$ can be chosen to depend solely on T and the sets of partial frequencies of (s_τ) , as required in Proposition 10. We have proved Proposition 10 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. \square

Proof of Lemma 21. Proof is by induction in d .

Base $d = 1$ is readily given by Lemma 20 combined with Proposition 6.

Step $1 \leq d \Rightarrow d + 1$: 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 , $j = 1, \dots, N$. Let $T \geq 0$, 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 $\mathbf{O}_1, \dots, \mathbf{O}_{d+1}$) such that

$$\begin{aligned} (a) : |g^{(T)}|_2 &\leq \rho_d(N_1, \dots, N_d)(2T + 1)^{-d/2}, \\ (a') : |h^{(T)}|_2 &\leq \rho_1(N_{d+1})(2T + 1)^{-1/2}, \\ (b) : r_\tau &= \sum_{\nu \in \mathbf{Z}^d} r_{\tau-\nu} g_\nu^{(T)} \quad \forall \tau \in \mathbf{Z}^d \quad \forall (r_\tau) \in \mathbf{E}(\mathbf{O}_1, \dots, \mathbf{O}_d), \\ (b') : p_\tau &= \sum_{\nu \in \mathbf{Z}} p_{\tau-\nu} h_\nu^{(T)} \quad \forall \tau \in \mathbf{Z} \quad \forall (p_\tau) \in \mathbf{E}(\mathbf{O}_{d+1}), \end{aligned} \tag{68}$$

where $\mathbf{E}(\mathbf{O}^1, \dots, \mathbf{O}^m)$ is the space of all simple exponential polynomials on \mathbf{Z}^m with the sets of partial frequencies $\mathbf{O}^1, \dots, \mathbf{O}^m$. Setting $q_\tau^{(T)} = g_{\tau_1, \dots, \tau_d}^{(T)} h_{\tau_{d+1}}^{(T)}$, $\tau \in \mathbf{Z}^{d+1}$, we clearly have

$$\begin{aligned} q^{(T)} &\in C_T(\mathbf{Z}^{d+1}), \quad |q^{(T)}|_2 = |g^{(T)}|_2 |h^{(T)}|_2 \leq \rho_d(N_1, \dots, N_d) \rho_1(N_{d+1}) \\ &= \rho_{d+1}(N_1, \dots, N_{d+1}) \end{aligned} \tag{69}$$

(see (68.a, a')). Further, for every $(s_\tau) \in \mathbf{E}(\mathbf{O}_1, \dots, \mathbf{O}_{d+1})$ we have, setting $\tau = (\tau', \tau'')$ with $\tau' \in \mathbf{Z}^d$, $\tau'' \in \mathbf{Z}$:

$$\begin{aligned} \sum_{\nu \in \mathbf{Z}^{d+1}} q_\nu^{(T)} s_{\tau-\nu} &= \sum_{\nu' \in \mathbf{Z}^d} g_{\nu'}^{(T)} \left(\sum_{\nu'' \in \mathbf{Z}} h_{\nu''}^{(T)} s_{\tau'-\nu', \tau''-\nu''} \right) \underbrace{=}_a \sum_{\nu' \in \mathbf{Z}^d} g_{\nu'}^{(T)} s_{\tau'-\nu', \tau''} \\ &\underbrace{=}_b s_{\tau', \tau''} \end{aligned}$$

(a is by (68.b') since $(s_{\tau'-\nu', \mu})_{\mu \in \mathbf{Z}} \in \mathbf{E}(\mathbf{O}_{d+1})$, b is by (68.b) since $(s_{\mu, \tau''})_{\mu \in \mathbf{Z}^d} \in \mathbf{E}(\mathbf{O}_1, \dots, \mathbf{O}_d)$), which combines with (69) to imply that

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

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

5.2.6 Proof of statement in Remark 11

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

$$\begin{aligned} (a) \quad & q^{(T)}(\Delta)p = p \text{ for every polynomial (17),} \\ (b) \quad & |q^{(T)}|_2 \leq \left(\frac{16m}{\sqrt{2T+1}}\right)^d \equiv \Theta^d. \end{aligned} \tag{70}$$

This well-known fact can be proved by induction in d completely similar to the one used to prove Lemma 21; the only difference is in the Base, which now should be replaced with the following statement:

Lemma 22 *Let $p(\tau) = \sum_{\ell=0}^m p_\ell \tau^\ell$ be a deterministic univariate algebraic polynomial of degree m . Then for every $T \geq 0$ there exists a filter $q \in C_T(\mathbf{Z})$, depending solely on T, m , with $|q|_2 \leq 16m(2T+1)^{-1/2}$ such that $p(t) = \sum_{\nu} q_\nu^{(T)} p(t-\nu)$ 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

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

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

$$\sum_{t=-T}^T \rho^2(\phi_t) \geq \frac{2T+1}{256m^2}, \quad \phi_t = \text{asin}(t/T). \tag{71}$$

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,

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

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 $\hat{\Delta}$ 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 (71). \square

5.2.7 Proof of Proposition 12

In the proof to follow, c_i stand for positive constants depending solely on \mathbf{D} . I^0 . We start with the following evident observation:

Lemma 23 *There exists c_1 such that for every polynomial $p(t)$ of one variable satisfying the relation $p(1) = 1$ one has*

$$\begin{aligned} M &\leq c_1 N, \quad \deg(p) \leq c_1 N, \\ (s) \in \mathbf{H}_N^t(\mathbf{D}) &\Rightarrow s_\tau = (p(\mathbf{D})s)_\tau \quad \forall (\tau : |\tau - t| \leq M). \end{aligned} \tag{72}$$

2^0 . Let us fix a positive integer N , and let

$$\begin{aligned} \delta(\omega) &= \sum_{\ell=1}^k w_\ell \exp\{i\omega^T \alpha(\ell)\} : [-\pi, \pi]^d \rightarrow \mathbb{C}, \\ \Omega_N^d &= \left\{ \omega \in \mathbb{R}^d \mid \omega_j \in \left\{ \frac{q\pi}{2N+1} \right\}_{|q| \leq N}, j = 1, \dots, d \right\}, \end{aligned} \tag{73}$$

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 $\mu(\{\zeta\}) = \sum_{\omega \in \Omega_N^d : \delta(\omega) = \zeta} \nu(\{\omega\})$. Let also $F(\alpha) = \mu(\{\zeta \mid \Re(\zeta) \geq 1 - \alpha\})$, $\alpha \geq 0$.

Lemma 24 *There exists $c_2 \in (0, 1)$ such that*

$$\mathbf{M} \subset \widehat{\mathbf{M}} = \left\{ \zeta \mid |\zeta| \leq 1, |\Im(\zeta)| \leq c_2^{-1}(1 - \Re(\zeta))^{3/2} \right\}, \tag{74}$$

$$F(\alpha) \leq c_2^{-1}[\alpha^{d/2} + N^{-d}], \quad 0 \leq \alpha \leq 2. \tag{75}$$

Proof. (74), (75) are evident when $\sum_{\ell=1}^k \rho_\ell < 1$, 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 $\sum_{\ell=1}^k \rho_\ell = 1$ (recall that $\sum_{\ell=1}^k \rho_\ell \leq 1$ by R.2).

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

$$\exp\{i[\phi_\ell + \omega^T \alpha(\ell)]\} = 1 \quad \forall (1 \leq \ell \leq k), \quad (76)$$

whence $\phi_\ell + \omega^T \alpha(\ell) \in 2\pi\mathbf{Z} \quad \forall (1 \leq \ell \leq k)$. Since $\text{Rank}\{\alpha(\ell) : 1 \leq \ell \leq k\} = d$, 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, $\text{Card } \mathbf{K} \leq c_3$.

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

$$\begin{aligned} \delta(\omega + d\omega) &= \sum_{\ell=1}^k \rho_\ell \exp\{i[\phi_\ell + \omega^T \alpha(\ell)]\} \exp\{i(d\omega)^T \alpha(\ell)\} \\ &\stackrel{a}{=} \sum_{\ell=1}^k \rho_\ell \exp\{i(d\omega)^T \alpha(\ell)\} \\ \Rightarrow |\delta(\omega + d\omega)| &= \left| \sum_{\ell=1}^k \rho_\ell \exp\{i(d\omega)^T \alpha(\ell)\} \right| \\ &\stackrel{b}{\leq} \left| \sum_{\ell=1}^k \rho_\ell \left(1 + i(d\omega)^T \alpha(\ell) - \frac{1}{2} \left((d\omega)^T \alpha(\ell) \right)^2 \right) \right| + c_4 |d\omega|^3 \\ &\stackrel{b}{=} \left| \sum_{\ell=1}^k \rho_\ell \left(1 - \frac{1}{2} \left((d\omega)^T \alpha(\ell) \right)^2 \right) \right| + c_4 |d\omega|^3 \stackrel{c}{\leq} 1 - c_5 |d\omega|^2 + c_4 |d\omega|^3 \end{aligned}$$

(for a , see (76), b is by (20.b), c is due to $\text{Rank}(\{\alpha(\ell)\}_\ell) = d$). It follows that with properly chosen c_6 one has

$$\forall (\omega \in [-\pi, \pi]^d, |\delta(\omega) - 1| \leq \alpha) \quad \exists \bar{\omega} \in \mathbf{K} : |\omega - \bar{\omega}| \leq c_6^{-1} \sqrt{\alpha}. \quad (77)$$

Since $\text{Card}(\mathbf{K}) \leq c_3$ by 2⁰.1) and $|\delta(\omega)| \leq 1$ for all ω , we conclude that

$$\nu \left(\{\omega \in \Omega_N^d : |\delta(\omega) - 1| \leq \alpha\} \right) \leq c_7 [\alpha^{d/2} + N^{-d}] \quad \forall \alpha \leq 2. \quad (78)$$

2^o.3) Now we can complete the proof of (74), (75). Let $\bar{\omega} \in \mathbf{K}$, $d\omega \in \mathbb{R}^d$, $|d\omega| \leq 1$. We have

$$\begin{aligned}
\delta(\bar{\omega} + d\omega) &= \sum_{\ell=1}^k \rho_\ell \exp\{i[\phi_\ell + \bar{\omega}^T \alpha(\ell)]\} \exp\{i(d\omega)^T \alpha(\ell)\} \\
&\stackrel{a}{=} \sum_{\ell=1}^k \rho_\ell \exp\{i(d\omega)^T \alpha(\ell)\} = \sum_{\ell=1}^k \rho_\ell \left(1 + i(d\omega)^T \alpha(\ell) \right. \\
&\quad \left. - \frac{1}{2} \left((d\omega)^T \alpha(\ell) \right)^2 - \frac{i}{6} \left((d\omega)^T \alpha(\ell) \right)^3 + r_\ell(\omega, d\omega) \right), \tag{79} \\
&\qquad\qquad\qquad [|r_\ell(\omega, d\omega)| \leq c_{10}|d\omega|^4] \\
&\stackrel{b}{=} \sum_{\ell=1}^k \rho_\ell \left(1 - \frac{1}{2} \left((d\omega)^T \alpha(\ell) \right)^2 - \frac{i}{6} \left((d\omega)^T \alpha(\ell) \right)^3 + r_\ell(\omega, d\omega) \right)
\end{aligned}$$

(for a , see (76), for b , see (20)). Taking into account that $\sum_\ell \rho_\ell = 1$ and $c_{11}|d\omega|^2 \leq \sum_\ell \rho_\ell \left((d\omega)^T \alpha(\ell) \right)^2 \leq c_{12}|d\omega|^2$, we conclude from (77) combined with (79) 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 (74) follows. By (74) one has $|1 - \delta(\omega)| \leq c_{14}(1 - \Re(\delta(\omega)))$, so that (75) follows from (78). \square

3^o. 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:

$$T_n(\zeta) = \frac{w^n + w^{-n}}{2}, \text{ where } w = \zeta + i\sqrt{1 - \zeta^2}. \tag{80}$$

In (80), 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 $T_n(\zeta) = \cos(n \arccos(\zeta))$, 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 25 *One has*

$$\begin{aligned}
p_n(\alpha) &\equiv \max_{\zeta} \{|P_n(\zeta)| : \zeta \in \widehat{\mathbf{M}}, \Re(\zeta) = 1 - \alpha\} \\
&\leq q_n(\alpha) = \begin{cases} c_{15}, & 0 \leq \alpha \leq \frac{1}{n^2} \\ \frac{c_{15}(1 + c_{15}\alpha)^n}{n^2\alpha}, & \frac{1}{n^2} \leq \alpha \leq 2 \end{cases}. \tag{81}
\end{aligned}$$

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

$$|\beta| \leq c_{16}\alpha^{3/2}. \quad (82)$$

We have

$$\begin{aligned} w &\equiv \zeta + i\sqrt{1 - \zeta^2} = 1 - \alpha + i\beta + i\sqrt{2\alpha - \alpha^2 - 2i(1 - \alpha)\beta + \beta^2} \\ &= 1 - \alpha + i\beta + i\sqrt{2\alpha}\sqrt{1 - 0.5\alpha + [0.5\beta - i(1 - \alpha)](\beta/\alpha)} \\ &= 1 + i\sqrt{2\alpha} + r_1(\zeta), \quad |r_1(\zeta)| \leq c_{17}\alpha \end{aligned} \quad (83)$$

(since $|\beta/\alpha| \leq c_{16}\sqrt{\alpha}$ by (83)). Note that completely similar considerations demonstrate that

$$w^{-1} = \zeta - i\sqrt{1 - \zeta^2} = 1 - i\sqrt{2\alpha} + r_2(\zeta), \quad |r_2(\zeta)| \leq c_{17}\alpha. \quad (84)$$

3⁰.1) Assume, first, that $0 \leq \alpha \leq \frac{1}{n^2}$. In this case from (83) it follows that $|1 - w| \leq \sqrt{2}n^{-1}$, whence, taking into account (83),

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

Thus, one has $|P_n(\zeta)| = \frac{\left| \frac{w^n + w^{-n}}{2} - 1 \right|}{n^2|\alpha - i\beta|} \leq c_{15}$, as required in (81) for the case of $0 \leq \alpha \leq \frac{1}{n^2}$.

3⁰.2) Now consider the case of $\frac{1}{n^2} \leq \alpha \leq 2$. From (83), (84) it follows that

$$|w| \leq 1 + c_{21}\alpha, \quad |w^{-1}| \leq 1 + c_{21}\alpha, \quad \text{whence } |P_n(\zeta)| = \frac{\left| \frac{w^n + w^{-n}}{2} - 1 \right|}{n^2|\alpha - i\beta|} \leq \frac{c_{22}(1 + c_{21}\alpha)^n}{n^2\alpha},$$

as required in (81). \square

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

$$\zeta = 1 - \alpha + i\beta \in \widehat{\mathbf{M}} \Rightarrow |Q(\zeta)| \leq 1 - c_{23}\alpha \quad [c_{23} < \frac{1}{2}]. \quad (85)$$

Now let c_{24} be a positive integer which is $\geq \frac{c_{15}}{c_{23}}$ (see (81)). Consider the polynomial $S_n(\zeta) = P_n(\zeta)Q^{c_{24}n}(\zeta)$.

Lemma 26 *For every positive integer n , the polynomial $S_n(\zeta)$ possesses the*

following properties:

$$\begin{aligned}
(a) : \deg(S_n) &\leq c_{25}n; \quad (b) : S_n(1) = 1; \\
(c) : \max_{\zeta} \{|S_n(\zeta)| : \zeta \in \widehat{\mathbf{M}}, \Re(\zeta) = 1 - \alpha\} &\leq c_{15} \min \left[\frac{1}{n^2\alpha}; 1 \right].
\end{aligned} \tag{86}$$

Proof. Relations (86.a – b) are evident (take into account that $P_n(1) = 1$ and $\deg(P_n) \leq n$). To verify (86.c), note that if $\zeta = 1 - \alpha + i\beta \in \widehat{\mathbf{M}}$, then in view of (81) one has

$$\begin{aligned}
0 \leq \alpha \leq \frac{1}{n^2} &\Rightarrow |S_n(\zeta)| \leq |P_n(\zeta)| |Q(\zeta)|^{c_{24}n} \leq |P_n(\zeta)| \leq c_{15}; \\
\frac{1}{n^2} \leq \alpha \leq 2 &\Rightarrow |S_n(\zeta)| \leq |P_n(\zeta)| |Q(\zeta)|^{c_{24}n} \stackrel{a}{\leq} c_{15} \frac{(1+c_{15}\alpha)^n}{n^2\alpha} (1 - c_{23}\alpha)^{c_{24}n} \\
&\leq c_{15} \frac{\exp\{c_{15}n\alpha\}}{n^2\alpha} \exp\{-c_{23}c_{24}n\alpha\} \stackrel{b}{\leq} \frac{c_{15}}{n^2\alpha}
\end{aligned}$$

(for a , see (85), b is due to $c_{23}c_{24} \geq c_{15}$). \square

5⁰. Now we are ready to complete the proof of Proposition 12. Given a positive integer n , let us set $R_n(\zeta) = S_n^d(\zeta)$. In view of (86) one has

$$\begin{aligned}
(a) : \deg(R_n) &\leq c_{26}n; \quad (b) : R_n(1) = 1; \\
(c) : \max_{\zeta} \{|R_n(\zeta)| : \zeta \in \widehat{\mathbf{M}}, \Re(\zeta) = 1 - \alpha\} &\leq r_n(\alpha) \\
&\equiv c_{26} \min \left[\frac{1}{n^{2^d}\alpha^d}; 1 \right].
\end{aligned} \tag{87}$$

Consider the filters $q^{(n)}(z)$ given by $q^{(n)}(\Delta) = R_n(\mathbf{D})$, $n = 0, 1, \dots$ By (87.b) and Lemma 23 we have

$$\left. \begin{aligned} T &\leq c_{27}N \\ 1 \leq n(T) &\equiv \lfloor c_{27}T \rfloor \\ (s) \in \mathbf{H}_N^t(\mathbf{D}) & \end{aligned} \right\} \Rightarrow \begin{cases} \text{ord}(q^{(n(T))}) \leq T, \\ s_{\tau} = (q^{(n(T))}(\Delta)s)_{\tau} \quad \forall (\tau : |\tau - t| \leq c_{27}N). \end{cases} \tag{88}$$

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

$$|q^{(n)}|_2^2 = \int_{\Omega_N^d} |R_n(\delta(\omega))|^2 \nu(d\omega) = \int_{\mathbf{M}} |R_n(\zeta)|^2 \mu(d\zeta) \stackrel{a}{\leq} \int_0^2 \underbrace{r_n^2(\alpha)}_{\rho_n(\alpha)} dF(\alpha) \tag{89}$$

with a given by (87.c), (74) 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

$$F(\alpha) \leq G(\alpha) \equiv \gamma([0, \alpha]) \quad \forall \alpha \in [0, 2] \tag{90}$$

(see (75)). We have

$$\begin{aligned}
\int_0^2 \rho_n(\alpha) dF(\alpha) &= \rho_n(2) - \int_0^2 \rho'_n(\alpha) F(\alpha) d\alpha \stackrel{a}{\leq} \rho_n(2) \\
- \int_0^2 \rho'_n(\alpha) G(\alpha) d\alpha &= \rho_n(2) - \rho_n(2)G(2) + \int_0^2 \rho_n(\alpha) \gamma(d\alpha) \\
\stackrel{b}{\leq} \int_0^2 \rho_n(\alpha) \gamma(d\alpha) &\stackrel{c}{=} c_2^{-1} \left[c_{28} \int_0^2 \min^2 [n^{-2d} \alpha^{-d}, 1] \alpha^{\frac{d}{2}-1} d\alpha \right. \\
&\left. + \rho_n(0) N^{-d} \right] \stackrel{d}{\leq} c_{30} [N^{-d} + n^{-d}] \leq c_{31} (2T+1)^{-d}
\end{aligned} \tag{91}$$

(a holds since $\rho_n(\cdot)$ is nonincreasing, see (87.c), and by (90), b holds since $c_2 \in (0, 1)$, see Lemma 24, c is by (87.c) and (89), d is due to $n = n(T) = \lfloor c_{27} T \rfloor$). Combining (89) and (91), we conclude that

$$|q^{(n(T))}|_2 \leq c_{32} (2T+1)^{-d/2}. \tag{92}$$

From (88) and (92) we conclude that if $L = \lfloor c_{33} N \rfloor$ and $T \leq L$ is such that $n(T) \equiv \lfloor c_{27} T \rfloor \geq 1$, then

$$\exists q^{(T)} \in C_T(\mathbf{Z}^d) : \begin{cases} |q^{(T)}|_2 \leq c_{32} (2T+1)^{-d/2}, \\ s_\tau = (q^{(T)}(\Delta)s)_\tau \forall (\tau, |\tau-t| \leq L, (s) \in \mathbf{H}_N^t(\mathbf{D})) \end{cases} \tag{93}$$

(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 (93) for all T , $0 \leq T \leq L$. Thus, $\mathbf{H}_N^t(\mathbf{D}) \subset \mathbf{F}_{\lfloor c_{29} L \rfloor}^t(0, c_{34})$. \square

5.2.8 Proof of Proposition 13

Lemma 27 *Let $f \in \mathbf{H}^+(M)$ be a deterministic function, let $N \leq M/2$, and let $t \in \mathbf{Z}^d$, $|t| \leq N$. Consider the “discrete box” $B_N^t = \{\tau \in \mathbf{Z}^d : |\tau-t| \leq N\}$, 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”: $\tau \in \mathbf{Z}^d, |\tau-t| < N \Rightarrow \phi_\tau = \frac{1}{2d} \sum_{\substack{\epsilon=(\epsilon_1, \dots, \epsilon_d) \\ |\epsilon_1|=\dots=|\epsilon_d|=1}} \phi_{\tau+\epsilon}$. Then*

$$\tau \in B_N^t \Rightarrow |f(\tau) - \phi_\tau| \leq c_1 \|f\|_{\infty, 2M} N^{-2} \tag{94}$$

(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” $\tau \in \text{int} B_N^t \equiv B_N^t \setminus \partial B_N^t$ 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 $|t| \leq N$ and $2N \leq M$, the function f is harmonic in the “continuous box” $D_{2N}^t = \{\tau \in \mathbb{R}^d : |\tau - t| \leq 2N\}$, 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

$$\forall(\tau \in D_N^t) : \left| \frac{\partial^\kappa}{\partial x_j^\kappa} f(\tau) \right| \leq c_2 \|f\|_{\infty, 2M} N^{-\kappa}, \quad \kappa = 1, 2, 3, 4, \quad j = 1, \dots, d. \quad (95)$$

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

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

Since f is harmonic, we conclude that

$$|(\mathbf{D}f)_\tau| \leq c_4 \|f\|_{\infty, 2M} N^{-4}, \quad \tau \in B_N^t. \quad (96)$$

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 (96) and the fact that ϕ is discrete harmonic, we have for $\tau \in \text{int} B_N^t$:

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

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 $h^+ \leq 4c_4 \|f\|_{\infty, 2M} N^{-2}$, we conclude that $\tau \in B_N^t \Rightarrow h_\tau \leq h_\tau^+ \leq \max_{\tau \in \partial B_N^t} h_\tau^+ \leq 2dc_4 \|f\|_{\infty, 2M} N^{-2}$. By similar reasons, $\tau \in B_N^t \Rightarrow h_\tau \geq h_\tau^- \geq \min_{\tau \in \partial B_N^t} h_\tau^- \geq -2dc_4 \|f\|_{\infty, 2M} N^{-2}$. \square

Now let $|t| \leq M/8$ and $L \leq M/8$. Given $T, 0 \leq T \leq L$, and applying Proposition 12, we can build filter $q^{(T)} \in C_T(\mathbf{Z}^d)$ such that

$$|q^{(T)}|_2 \leq c_5 (2T + 1)^{-1}, \quad \phi_\tau = \sum_{|\nu| \leq T} \phi_{\tau - \nu} q_\nu^{(T)} \quad \forall(\tau : |\tau - t| \leq L) \quad (97)$$

for every ϕ which is discrete harmonic in the discrete box B_{2L}^t . Now let $f \in \mathbf{H}(M, R)$. Applying Lemma 27, we can find function ϕ which is discrete harmonic in the box B_{2L}^t and such that $|\phi_\tau - f_\tau|^2 \leq c_6^2 \|f\|_{\infty, 2M}^2 L^{-4}$ for $\tau \in B_{2L}^t$. From (97) it now follows that

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

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

The proof of Proposition 14 is completely similar to that of Proposition 10.

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. Birgé, 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. Birgé. An alternative point of view on Lepski's method. Technical Report 549, Paris 6, december 1999.
- [7] L. Birgé 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] E. Candés, D. Donoho. Ridgelets: a key to high-dimensional intermittency? *Philos. Trans. Roy. Soc. London Ser. A* 357:2495–2509, 1999.

- [11] S. Chen, D.L. Donoho, M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* 20(1):33-61, 1998.
- [12] D. Donoho, I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425-455, 1994.
- [13] D. Donoho, I. Johnstone. Minimax risk over ℓ_p balls for ℓ_q losses. *Probab. Theory Related Fields* 99:277-303, 1994.
- [14] D. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmon. Anal.* 1:100-115, 1994.
- [15] 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.
- [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] D. Donoho. Wedgelets: nearly minimax estimation of edges. *Ann. Statist.* 27:859-897, 1999.
- [20] D. Donoho. Orthonormal ridgelets and linear singularities. *SIAM J. Math. Anal.* 31:1062-1099, 2000.
- [21] D. Donoho. Ridge functions and orthonormal ridgelets. *J. Approx. Theory* 111(2):143-179, 2001.
- [22] D. Donoho. Curvelets and curvilinear integrals. *J. Approx. Theory* 113(1):59-90, 2001.
- [23] D. Donoho. Sparse components of images and optimal atomic decompositions. *Constr. Approx.* 17:353-382, 2001.
- [24] D. Donoho, X. Huo. Uncertainty principle and ideal atomic decomposition. *IEEE Trans. on Information Theory* 47(7):2845-2862, 2001.
- [25] D. Donoho, X. Huo. Beamlets and multiscale image analysis. *Lect. Comput. Sci. Eng.* 20:149-196, Springer, 2002.
- [26] S. Efromovitch. On global and pointwise adaptive estimation. *Bernoulli* 4(2), 273-282, 1998.
- [27] M. Elad, A. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. on Information Theory* (to appear)

- [28] A. Goldenshluger, A. Nemirovski. On spatially adaptive estimation of nonparametric regression. *Math. Methods of Statistics* 6(2):135–170, 1997.
- [29] A. Goldenshluger, A. Nemirovski. Adaptive de-noising of signals satisfying differential inequalities. *IEEE Trans. on Information Theory* 43, 1997.
- [30] 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.)
- [31] W. Härdle. *Applied Nonparametric Regression*, ES Monograph Series 19, Cambridge, U.K., Cambridge University Press, 1990.
- [32] I. Ibragimov and R. Khasminskii. On nonparametric estimation of regression. *Soviet Math. Dokl.* 21:810–814, 1980.
- [33] I. Ibragimov and R. Khasminskii. *Statistical Estimation*. Springer-Verlag, New York, 1981.
- [34] A. Juditsky. Wavelet estimators: Adapting to unknown smoothness. *Math. Methods of Statistics* 6(1):1–25, 1997.
- [35] A. Juditsky and A. Nemirovski. Oracle Inequalities for adaptive filtering problems submitted to *Annals of I.H.P.*.
- [36] A. Juditsky, A. Nemirovski. Nonparametric denoising of signals of unknown local structure, II: application to nonparametric regression estimation. Submitted to *Appl. Comp. Harm. Anal.*.
- [37] A. Korostelev, A. Tsybakov. *Minimax theory of image reconstruction*. Lecture Notes in Statistics 82, Springer, New York, 1993.
- [38] A. Kneip. Ordered linear smoothers. *Ann. Stat.*, 22(2): 835–866, 1994.
- [39] O. Lepski. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probabi. Appl.* 35(3):454–466, 1990.
- [40] O. Lepski. Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory of Probabi. Appl.*, 36(4):682–697, 1991.
- [41] O. Lepski. Asymptotically minimax adaptive estimation. II. Statistical model without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* 37, 433–448, 1992.
- [42] O. Lepski, 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.
- [43] O. Lepski, V. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* 25(6), 2512–2546, 1997.
- [44] A. Nemirovski. On forecast under uncertainty. *Problemy peredachi informatsii* 17(4):73–83, 1981. (English transl. in *Problems Inform. Transmission* 17, 1981.)

- [45] A. Nemirovskii. On nonparametric estimation of smooth regression functions. *Sov. J. Comput. Syst. Sci.*, 23(6):1–11, 1985.
- [46] A. Nemirovski. On nonparametric estimation of functions satisfying differential inequalities. R. Khasminski, Ed. *Advances in Soviet Mathematics* 12:7–43, American Mathematical Society, 1992.
- [47] 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.
- [48] A. Nemirovski. Denoising signals of unknown local structure Medallion Lecturer of IMS, JSM 2003, San Francisco, August 3-5, 2003.
- [49] 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.)
- [50] M. Pinsker, S. Efroimovich. Learning algorithm for nonparametric filtering. *Automation and Remote Control* 45(11):1434–1440, 1984.
- [51] M. Rosenblatt. *Stochastic curve estimation*. Institute of Mathematical Statistics, Hayward, California, 1991.
- [52] J. Polzehl, V. Spokoiny. Adaptive weights smoothing with applications to image restoration. *J. R. Stat. Soc., Ser. B Stat. Methodol.* 62(2), 335–354, 2000.
- [53] J. Polzehl, V. Spokoiny. Image denoising: pointwise adaptive approach. *Ann. Statist.* 31(1), 3057, 2003.
- [54] V. Spokoiny. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Statist.* 26 13561378,1998.
- [55] J.-L. Starck, E. Candés, D. Donoho. The curvelet transform for image denoising. *IEEE Trans. Image Process.* 11(6):670-684, 2002.
- [56] Ch. Stone. Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, 8(3):1348–1360, 1980.
- [57] A. Tsybakov. Pointwise and sup norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26: 2520-2469, 1998.
- [58] G. Wahba. *Spline models for observational data*. SIAM, Philadelphia, 1990.