Appl. Comput. Harmon. Anal. 27 (2009) 157–179

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

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 30332-0205, USA

ARTICLE INFO

Article history:
Received 18 March 2008
Revised 12 December 2008
Accepted 2 February 2009
Available online 14 February 2009
Communicated by Dominique Picard

Keywords:
Nonparametric denoising
Oracle inequalities
Adaptive filtering

ABSTRACT

We consider the problem of pointwise estimation of multi-dimensional signals s, from noisy observations (y_τ) on the regular grid $\mathbb{Z}^d$. 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 can be summarized as follows: suppose that the signal s is "well-filtered", i.e. there exists an adapted 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 the value 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.

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 $(s_\tau=s_\tau(\xi))_{\tau\in\mathbb{Z}^d,\,\xi\in\Omega}$ over $\mathbb{Z}^d$ 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. Our focus here is at estimating the value s_t of signal at a given location $t\in\mathbb{Z}^d$.

* Corresponding author.
E-mail addresses: anatoli.juditsky@imag.fr (A. Juditsky), nemirovs@isye.gatech.edu (A. Nemirovski).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2009.02.001

The above problem is "classical" in statistical estimation and signal processing, and as such, has received much attention. In particular, linear estimators (referred as linear filters in the signal processing community) are widely used in the statistical literature. To be more precise, suppose that our aim is to recover the value s_0 of the signal at zero given observations (y_τ) on the box $\mathbf{O}_T = \{\tau \in \mathbb{Z}^d\colon\ |\tau_j| \leqslant T,\ 1 \leqslant j \leqslant d\}$. We call the estimation $\hat{s}$ of s_0 linear if it is of the form

$$\hat{s}_\ell = \sum_{\tau \in \mathbf{O}_T} q_\tau y_\tau$$

for some $q \in C(\mathbf{O}_T)$, where $C(\mathbf{O}_T)$ is the set of complex-valued fields $q = \{q_\tau,\ \tau \in \mathbf{O}_T\}$ over $\mathbf{O}_T$.

The simplicity of linear estimators is responsible for their popularity in statistical signal processing. Another outstanding feature of such estimators is their minimax property. Suppose that the a priori information resumes to the fact that (s_τ) belongs to some convex compact set which is symmetric with respect to zero, let us call it $\mathbf{S}$. One of the most renown results of estimation theory (see, for instance, [8,10,15]) states that the *linear minimax estimator* is, in a certain sense, an optimal estimator of s_0 in our problem. Indeed, consider the following *linear minimax estimation strategy*: let $q_*^{(T)}$ be the optimal solution[1] to the problem

$$\min_{q \in C(\mathbf{O}_T)} \max_{s \in \mathbf{S}} E_s\left(s_0 - \sum_{\tau \in \mathbf{O}_T} q_\tau y_\tau\right)^2$$

(here E_s stands for the expectation with respect to the distribution of (y_τ) which corresponds to the underlying signal s). The linear minimax estimator $\hat{s}_\ell^*$ of s_0 is defined by

$$\hat{s}_\ell^* = \sum_{\tau \in \mathbf{O}_\tau} q_{*,\tau}^{(T)} y_\tau.$$

Then

$$\max_{s \in \mathbf{S}} E_s\big(s_0 - \hat{s}_\ell^*\big)^2 \leqslant C \inf_{\hat{s}} \max_{s \in \mathbf{S}} E_s(s_0 - \hat{s})^2,$$

where the infimum in the right-hand side is taken over all possible estimators of s_0 from observations (y_τ) and C is a moderate absolute constant (e.g., $C \leqslant 1.25$). In other words, the linear estimator $\hat{s}_\ell$ is a (almost) minimax estimator of s_0. We would like to stress the exceptional power of the above result – we only need $\mathbf{S}$ to be convex and compact for the linear estimator to be minimax optimal. The evident downside of using linear minimax estimators is that the *a priori* information about the set $\mathbf{S}$ of signals should be as precise as possible to achieve descent estimation accuracy. There was a significant research on *adaptive estimation* in the above setting (cf. [6,7]). Those techniques allow to choose the "best" in a certains sense set which contains the signal from special finite families of convex sets. Another "classical" approach to adaptation for linear estimators has been developed in [20–23,28]. In the latter approach the "form" of the filter $q^{(T)}$ is considered as given in advance (no information about sets of signals is used in this case), and the parameter T (the "window width") is selected adaptively to achieve the best bias/variance tradeoff. Recently, more general adaptation techniques has been studied in [14,24], which allow to choose the best estimator from special finite families of available linear estimators.

The problem we are interested in here, *when posed informally*, is as follows: if we consider the form of the filter 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 hypothetic 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 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* $(s_\tau)_{\tau \in \mathbb{Z}^d} \equiv (s_\tau(\xi))_{\tau \in \mathbb{Z}^d,\, \xi \in \Omega}$ *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* $q_*^{(T)}$ *which recovers* s_0 *via* $O(T^d)$ *observations around zero such that*

$$E\left\{\left|s_0 - \sum_{\tau \in \mathbf{O}_T} q_{*,\tau}^{(T)} y_\tau\right|^2\right\} \leqslant O\big(\sigma^2 T^{-d}\big). \tag{2}$$

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

[1] For evident reasons such a solution exists in the situation we are interested in.

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 $E(\sum_{\tau\in\mathbf{O}_T} q_{*,\tau}^{(T)} e_\tau)^2$ 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 quite wide. As we shall see later, this family contains also "highly oscillating" sampled modulated smooth signals, sampled harmonic functions, etc.

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 $t\in\mathbb{Z}^d$ via noisy observations taken at the points $\{\tau\in\mathbb{Z}^d\colon\ t_j - T\leqslant\tau_j\leqslant t_j-\kappa\}$ "preceding" the point t, with a given in advance "forecast horizon" $\kappa\geqslant 0$.

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 $\mathbb{Z}^d$), 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 [25] and [26]. 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 Appendix A.

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 the companion paper [18]. The results of [18] extend to the wide classes of modulated signals the results of [12,13,16,27] on spatial adaptive estimates of signals with inhomogeneous smoothness.

2. Problem statement

In order to proceed we need some notations.

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

- Given nonnegative integer T and $p\in[1,\infty]$, we define semi-norms $|\cdot|_{T,p}$ on $C(\mathbb{Z}^d)$ by $|r|_{T,p}=(\sum_{|\tau|\leqslant T}|r_\tau|^p)^{1/p}$, $|\tau|=\max\{|\tau_1|,\dots,|\tau_d|\}$, with the standard interpretation of the right-hand side when $p=\infty$, and we set $|r|_p=\lim_{T\to\infty}|r|_{T,p}\in\mathbb{R}\cup\{+\infty\}$. A field $r\in C(\mathbb{Z}^d)$ with finitely many nonzero entries r_τ is called *a filter*, and the smallest T such that $r_\tau=0$ whenever $|\tau|>T$, is called the *order* $\operatorname{ord}(r)$ of a filter r; we write $C_T(\mathbb{Z}^d)=\{r\in C(\mathbb{Z}^d)\mid \operatorname{ord}(r)\leqslant T\}$. 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 any of $\tau_j<0$, $j=1,\dots,d$). The set of all polynomials is denoted $P(\mathbb{Z}^d)$. For integers k,T, $0\leqslant k\leqslant T$, we denote by $P_T^k(\mathbb{Z}^d)$ the subspace of $P(\mathbb{Z}^d)$ formed by polynomials r for which the entries r_τ vanish outside the set $k\leqslant\tau_j\leqslant T$, $j=1,\dots,d$.
- We denote by Δ_j, $j=1,\dots,d$, the "basic shift operators" on $C(\mathbb{Z}^d)$:

$$(\Delta_j r)_{\tau_1,\dots,\tau_d}=r_{\tau_1,\dots,\tau_{j-1},\tau_j-1,\tau_{j+1},\dots,\tau_d}.$$

  Further, we use the notation Δ_j^{-1} for the inverse of Δ_j:

$$\left(\Delta_j^{-1}r\right)_{\tau_1,\dots,\tau_d}=r_{\tau_1,\dots,\tau_{j-1},\tau_j+1,\tau_{j+1},\dots,\tau_d}.$$

- Finally, we define the output of a filter r, the input to the filter being a field $x\in C(\mathbb{Z}^d)$, 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 $F_T: C(\mathbb{Z}^d)\to C(\Gamma_T^d)$ as $(F_T r)(\mu)=\frac{1}{(2T+1)^{d/2}}\sum_{|\tau|\leqslant T} r_\tau\mu_1^{\tau_1}\dots\mu_d^{\tau_d}\equiv\frac{1}{(2T+1)^{d/2}}r(\mu)$, $r\in C_T(\mathbb{Z}^d)$, where $\mu\in\Gamma_T^d$. Note that $r_\tau=\frac{1}{(2T+1)^{d/2}}\sum_{\mu\in\Gamma_T^d}(F_T r)(\mu)\mu_1^{-\tau_1}\dots\mu_d^{-\tau_d}$, $\forall(\tau\colon\ |\tau|\leqslant T)$. The Fourier transform allows to equip $C(\mathbb{Z}^d)$ with semi-norms coming from the standard p-norms on $C(\Gamma_T^d)$:

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

  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

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

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 $\theta \geqslant 0$, $\rho \geqslant 1$ be reals, let L be a nonnegative integer or $+\infty$, and let $t \in \mathbb{Z}^d$. Finally, let $(s_\tau)_{\tau\in\mathbb{Z}^d} \equiv (s_\tau(\xi))_{\tau\in\mathbb{Z}^d,\,\xi\in\Omega}$ be a random field on $\mathbb{Z}^d$.

(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 $q = q^{(T)} \in C_T(\mathbb{Z}^d)$, $|q|_2 \leqslant \frac{\rho}{(2T+1)^{d/2}}$, which reproduces (s_τ) in the box $\{\tau\colon |\tau - t| \leqslant L\}$ with the mean square error not exceeding $\theta(2T+1)^{-d/2}$:

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

(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, $0 \leqslant T \leqslant L$, (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 $t \in \mathbb{Z}^d$ via a given number observations "around" this point.[2] Another interesting problem is that of *prediction*, where the goal is to recover s_t via observations y_τ "preceding by a given horizon $\kappa \in \mathbb{Z}_+$" the point t, i.e., observations with $\tau_j \leqslant t_j - \kappa$, $j = 1,\dots,d$.

Definition 2. Let $\theta \geqslant 0$, $\rho \geqslant 1$ be reals, let $T_0 \geqslant \kappa$ be nonnegative integers, L be a nonnegative integer or $+\infty$, and let $t \in \mathbb{Z}^d$. Finally, let $(s_\tau)_{\tau\in\mathbb{Z}^d} \equiv (s_\tau(\xi))_{\tau\in\mathbb{Z}^d,\,\xi\in\Omega}$ be a random field on $\mathbb{Z}^d$.

(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 $q = q^{(T)} \in P_T^\kappa(\mathbb{Z}^d)$, $|q|_2 \leqslant \frac{\rho}{(2T+1)^{d/2}}$, which reproduces (s_τ) in the box $\{\tau\colon |\tau - t| \leqslant L\}$ with the mean square error not exceeding $\theta(2T+1)^{-d/2}$:

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

(2) [*Well-predicted signals*] We say that (s_τ) is well-predicted, with the parameters θ, ρ, κ, T_0, L, at the point t (notation: $(s_\tau) \in \mathbf{P}_{\kappa,T_0,L}^t(\theta,\rho)$), if, for every integer T, $T_0 \leqslant T \leqslant L$, (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 fulfills the requirements of Definition 2 or 1 above. The filters $q^{(T)}$ associated, in the sense of the above 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.

---

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

3.1. Adaptive filtering

The estimator we intend to use is as follows:

Algorithm A. *Given a setup* $(\rho \geqslant 1, T)$ *and a point* $t \in \mathbb{Z}^d$, *we build an estimation* $\hat{s}_t[T, y]$ *of* s_t *via observations* (y_τ), $|\tau - t| \leqslant 4T$, *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* $\hat{\phi}^t \in C_{2T}(\mathbb{Z}^2)$ *is an optimal solution to the following optimization problem*:

$$\min_{\phi \in C_{2T}(\mathbb{Z}^d)} \big\{ \underbrace{\big|\Delta_1^{-t_1} \dots \Delta_d^{-t_d}\big(1 - \phi(\Delta)\big)y\big|^*_{2T,\infty}}_{J(\phi, y^t_{4T})} : \ |\phi|^*_{2T,1} \leqslant 2^{d/2}\rho^2(2T+1)^{-d/2} \big\}, \tag{6}$$

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

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

Theorem 4. *Assume that the signal* (s_τ) *underlying observations* (1) *is* T*-well-filtered, with parameters* θ, ρ, $L \geqslant 3T$: $(s_\tau) \in \mathbf{S}^t_L(\theta, \rho, T)$ *with* $L \geqslant 3T$. *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:*

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

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

$$\big|s_t - \hat{s}_t[T, y]\big| \leqslant c(d)\rho^3\big[\theta + \sigma\rho\Theta^t_T\big](2T+1)^{-d/2},$$
$$\Theta^t_T = \sigma^{-1} \max_{\tau: \ |\tau| \leqslant 2T} \big|\Delta_1^{\tau_1 - t_1} \dots \Delta_d^{\tau_d - t_d} e\big|^*_{2T,\infty}. \tag{8}$$

Comments: note that Theorem 4 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 $q^* \in C_T(\mathbb{Z}^d)$) and with the mean square error which, in an $O(T)$-neighborhood of a point t, is of the "parametric" order $O(\sigma(2T+1)^{-d/2})$:

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

with some known $\mu \geqslant 1$. 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^t_{4T} 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 $|q^*|_2 \leqslant \frac{\mu}{(2T+1)^{d/2}}$ and $(s_\tau) \in \mathbf{S}^t_{3T}(\sigma\mu, \mu, T)$. 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 $O(1)\mu^3[1 + \sqrt{\ln(2T+1)}]\kappa$. 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 [2–5,9,11,19] 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 [17] 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 $\rho^2 \geqslant 1$ 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* $(\rho \geqslant 1, \kappa, T)$ *and a point* $t \in \mathbb{Z}^d$, *we build a prediction* $\hat{s}_t[\kappa, T, y]$ *of* s_t *via observations* (y_τ), $\kappa \leqslant t_j - \tau_j \leqslant 4T$, $j = 1, \dots, d$, *as* $\hat{s}_t[\kappa, T, y] = (\hat{\psi}^t(\Delta)y)_t$, *where* $\hat{\psi}^t \in P^{\kappa}_{2T}(\mathbb{Z}^2)$ *is an optimal solution to the following optimization problem*:

$$\min_{\psi \in C^{\kappa}_{2T}(\mathbb{Z}^d)} \left\{ \underbrace{\left|\Delta_1^{-t_1} \dots \Delta_d^{-t_d}\big(1 - \psi(\Delta)\big)y\right|^*_{2T,\infty}}_{J(\psi, y^t_{\kappa,4T})} :\ |\psi|^*_{2T,1} \leqslant \frac{2^{d/2}\rho^2}{(2T+1)^{d/2}} \right\}; \tag{10}$$

where $y^t_{\kappa,L} = \{y_\tau\colon \kappa \leqslant t_j - \tau_j \leqslant L,\ j = 1, \dots, d\}$.

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

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

$$\begin{aligned} &\left(E\left\{\left|\hat{s}_t[\kappa, T, y] - s_t\right|^2\right\}\right)^{1/2} \leqslant c(d)\rho^3 \frac{\theta + \sigma\rho\sqrt{\ln(2T+1)+1}}{(2T+1)^{d/2}}, \\ &c(d) = 3\big(2^d + 2^{3d-1}\big). \end{aligned} \tag{11}$$

In particular, if (s_τ) *is well-predicted, with the parameters* θ, ρ, κ, T_0, L, *at a point* t, *then for every integer* T, $T_0 \leqslant T \leqslant \lfloor L/3 \rfloor$, *the accuracy of the estimate* $\hat{s}_t[\kappa, T, y]$ *of* s_t *yielded by Algorithm* B *can be bounded by* (11).

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

$$\begin{aligned} &\left|s_t - \hat{s}_t[T, y]\right| \leqslant c(d)\rho^3\big[\theta + \sigma\rho\Theta^t_T\big](2T+1)^{-d/2}, \\ &\Theta^t_T = \sigma^{-1} \max_{\tau\colon |\tau| \leqslant 2T} \left|\Delta_1^{\tau_1 - t_1} \dots \Delta_d^{\tau_d - t_d} e\right|^*_{2T,\infty}. \end{aligned} \tag{12}$$

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: $\rho' \geqslant \rho$, $\theta' \geqslant \theta$, $L' \leqslant L \Rightarrow \mathbf{F}^t_L(\theta, \rho) \subset \mathbf{F}^t_{L'}(\theta', \rho')$ and $\rho' \geqslant \rho$, $\theta' \geqslant \theta$, $\kappa' \leqslant \kappa$, $T_0' \geqslant T_0$, $L' \leqslant L \Rightarrow \mathbf{P}^t_{\kappa,T_0,L}(\theta, \rho) \subset \mathbf{P}^t_{\kappa',T_0',L'}(\theta', \rho')$.

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

Proposition 6.

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

$$\left(s_\tau \equiv \sum_{j=1}^m \lambda_j s_\tau^j\right) \in \mathbf{F}^t_{L^+}(\theta^+, \rho^+),$$
$$\theta^+ = (2m-1)^{d/2} 2^{m-1} \rho_1 \ldots \rho_m \sum_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j},$$
$$\rho^+ = (2m-1)^{d/2} 2^m \rho_1 \ldots \rho_m, \qquad L^+ = \lfloor L/2 \rfloor. \tag{13}$$

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

$$\left(s_\tau \equiv \sum_{j=1}^m \lambda_j s_\tau^j\right) \in \mathbf{P}^t_{\kappa^+, L^+}(\theta^+, \rho^+),$$
$$\theta^+ = (2m-1)^{d/2} 2^{m-1} \rho_1 \ldots \rho_m \sum_{j=1}^m \frac{\theta_j [E\{|\lambda_j|^2\}]^{1/2}}{\rho_j},$$
$$\rho^+ = (2m-1)^{d/2} 2^m \rho_1 \ldots \rho_m, \qquad \kappa^+ = \min_{1 \leqslant j \leqslant m} \kappa_j, \qquad T_0^+ = m \max_{1 \leqslant j \leqslant m} T_0^j,$$
$$L^+ = \lfloor L/2 \rfloor. \tag{14}$$

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 $(s_\tau) \in \mathbf{F}^t_L(\theta, \rho)$, and let $\omega \in \mathbb{R}^d$, $\phi \in \mathbb{R}$ be deterministic. Then the signal $(\hat{s}_\tau = \exp\{i[\omega^T \tau + \phi]\} s_\tau)_{\tau \in \mathbb{Z}^d}$ belongs to $\mathbf{F}^t_L(\theta, \rho)$ 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 $\mathbf{F}^t_L(\theta, \rho)$.*

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

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

Proposition 8.

(i) *Let $1 \leqslant d \leqslant d^+$, and let $(s_\tau)_{\tau \in \mathbb{Z}^d}$ be a signal which is well-filtered, with parameters θ, ρ, L, at a point $t \in \mathbb{Z}^d$. Then the signal $(s^+_{\tau_1, \ldots, \tau_{d^+}} = s_{\tau_1, \ldots, \tau_d})$ is well-filtered, with the parameters $\theta^+ = (2L+1)^{(d^+ - d)/2}\theta$, $\rho^+ = \rho$, $L^+ = L$ at every point $t^+ \in \mathbb{Z}^{d^+}$ such that $(t_1^+, \ldots, t_d^+) = t$.*

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

V. "Tensor product". Let $d = d' + d''$ with positive integers d', d'', so that $\mathbb{Z}^d = \mathbb{Z}^{d'} \times \mathbb{Z}^{d''}$. Given random fields $(s'_{\tau'}(\xi))_{\substack{\tau' \in \mathbb{Z}^{d'} \\ \xi}}$, $(s''_{\tau''}(\xi))_{\substack{\tau'' \in \mathbb{Z}^{d''} \\ \xi}}$, we define their tensor product as the field $(s_\tau(\xi) = s'_{\tau'}(\xi) s''_{\tau''}(\xi))_{\substack{\tau = (\tau', \tau'') \in \mathbb{Z}^d \\ \xi}}$.

Proposition 9.

(i) *Let $(s'_{\tau'}(\xi))_{\substack{\tau' \in \mathbb{Z}^{d'} \\ \xi}} \in \mathbf{F}^{t'}_L(0, \rho')$, $(s''_{\tau''}(\xi))_{\substack{\tau'' \in \mathbb{Z}^{d''} \\ \xi}} \in \mathbf{F}^{t''}_L(0, \rho'')$. Then $(s_\tau) \in \mathbf{F}^{(t', t'')}_L(0, \rho'\rho'')$.*

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

4.2. Examples of nice signals

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

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

where $\omega(\ell)$ and $\alpha(\ell)$ are deterministic, and c_ℓ may be random. Given an exponential polynomial (s_τ) on $\mathbb{Z}^d$, we define its *partial sizes* N_j, $j = 1,\ldots,d$, as follows: let m_j be the maximum of the degrees $\alpha_j(\ell)$, $\ell = 1,\ldots,M$, 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" $\omega_j(\ell)$: $M_j = \operatorname{Card}\mathbf{O}_j$, $\mathbf{O}_j = \{\omega_j(\ell)\colon 1 \leqslant \ell \leqslant M\}$. The *jth 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 jth size is by 1 larger than the degree of the polynomial w.r.t. jth variable τ_j.

Proposition 10. *Let (s_τ) be an exponential polynomial on $\mathbb{Z}^d$ of partial sizes $N_1,\ldots,N_d$. Then for all $t \in \mathbb{Z}^d$ one has*

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

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

Remark 11. A major shortcoming of (16) is a dramatic growth of $\rho_d(N,N,\ldots,N)$ 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

$$p_\tau = \sum_{\alpha \geqslant 0,\, |\alpha| \leqslant m} c_\alpha \tau^\alpha \tag{17}$$

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

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

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

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

$$B^t_N = \big\{\tau \in \mathbb{Z}^d \bigm| |\tau - t| \leqslant N\big\}, \qquad B^t_N(\mathbf{D}) = \big\{\tau \in B^t_N \bigm| \tau + \alpha(\ell) \in B^t_N,\ \ell = 1,\ldots,k\big\},$$
$$\mathbf{H}^t_N(\mathbf{D}) = \big\{(s) \in C\big(\mathbb{Z}^d\big) \bigm| s_\tau = (\mathbf{D}s)_\tau\ \ \forall \tau \in B^t_N(\mathbf{D})\big\}.$$

For example, with

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

the linear space $\mathbf{H}^t_N(\mathbf{D})$ is the space of fields which are "discrete harmonic" on B^t_N, that is,

$$s_\tau = \frac{1}{2d}\sum_{\substack{i=1,\ldots,d\\ \epsilon=\pm 1}} s_{\tau_1,\ldots,\tau_{i-1},\tau_i+\epsilon,\tau_{i+1},\ldots,\tau_d}$$

for all τ with $|\tau - t| \leqslant N-1$. Let us call a difference operator $\mathbf{D}$ *regular*, if it possesses the following properties:

R.1 The vectors $\{\alpha(\ell)\}_{1\leqslant\ell\leqslant k}$ span the entire $\mathbb{R}^d$;
R.2 The coefficients $w_\ell = \rho_\ell \exp\{i\phi_\ell\}$ $(\rho_\ell \geqslant 0,\ \phi_\ell \in \mathbb{R})$ are nonzero, and

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

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*

$$\forall N > 0: \quad \mathbf{H}^t_N(\mathbf{D}) \subset \mathbf{F}^t_{\lfloor cN \rfloor}\big(0, c^{-1}\big). \tag{21}$$

As a nontrivial application example for Proposition 12, consider the families of random fields defined as follows. Let $d \leqslant 4$, M be a positive integer, and R be a positive real. Consider the family $\mathbf{H}^+(M)$ of all deterministic continuous functions f on $\mathbb{R}^d$ which are harmonic in the interior of the box $D^0_{2M} = \{x \in \mathbb{R}^d\colon\ |x_j| \leqslant 2M,\ j \leqslant d\}$: $(\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}) f(x) = 0,\ x \in \operatorname{int} D^0_{2M}$. Now let $\mathbf{H}^+(M, R)$ be the family of random functions f such that all realizations of a function belong to $\mathbf{H}^+(M)$ and, besides this, $E\{\|f\|^2_{\infty,2M}\} \leqslant R^2$, where $\|f\|_{\infty,2M}$ is the uniform norm on D^0_{2M}. Restricting functions f from $\mathbf{H}^+(M, R)$ on $\mathbb{Z}^d$, we get a family of random fields $\mathbf{H}(M, R)$ on $\mathbb{Z}^d$.

Proposition 13. *Let* $d \leqslant 4$, 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*

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

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 $\mathbb{Z}^d$ as an exponential polynomial

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

where all partial frequencies $\omega_j(\ell)$ satisfy the restriction $\Re(\omega_j(\ell)) \leqslant 0$. 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* $\mathbb{Z}^d$ *of partial sizes* $N_1, \ldots, N_d$. *Then for every integer* $\kappa \geqslant 0$ *and all* $t \in \mathbb{Z}^d$ *one has*

$$\begin{aligned}
&(s_\tau) \in \mathbf{P}^t_{\kappa,T_0,\infty}\big(0, \rho_{\kappa,d}(N_1, \ldots, N_d)\big),\\
&\rho_{\kappa,d}(N_1, \ldots, N_d) = \prod_{j=1}^d \big[(2N_j - 1)^{1/2} 2^{N_j} \big(\max[2, 2\kappa + 1]\big)^{N_j/2}\big],\\
&T_0 = \kappa \max_{1 \leqslant j \leqslant d} N_j
\end{aligned} \tag{24}$$

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* $\mathbf{O}_j = \{\omega_j(\ell)\colon 1 \leqslant \ell \leqslant M\}$ *of partial frequencies.*

Appendix A

A.1. Preliminaries

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

Parseval equality:

$$(r, s)_T \equiv \sum_{t:\ |t| \leqslant T} r_t \overline{s_t} = \sum_{\mu \in \Gamma^d_T} (F_T r)(\mu) \overline{(F_T s)(\mu)} \equiv \langle F_T r, F_T s \rangle_T, \tag{25}$$

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

$$|r|_{T,2} = |r|^*_{T,2}. \tag{26}$$

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

$$\Big|\sum_{|t|\leqslant T} a_t b_t\Big| \leqslant |a|^*_{T,1}|b|^*_{T,\infty}. \tag{27}$$

Norms of convolutions of filters:

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

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

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

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

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

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

$$\Big[E\Big\{\max_{0\leqslant j<N}|f_j|^2\Big\}\Big]^{1/2} \leqslant \sqrt{2\ln N + 2};$$
$$P\Big\{\max_{0\leqslant j<N}|f_j| > u + \sqrt{2\ln N}\Big\} \leqslant \exp\{-u^2/2\} \quad \forall u \geqslant 0. \tag{31}$$

Proof. We have

$$\psi(r) \equiv P\Big\{\max_{0\leqslant j<N}|f_j| > r\Big\} \leqslant \min\big[1, N\exp\{-r^2/2\}\big]$$
$$\Rightarrow \quad P\Big\{\max_{0\leqslant j<N}|f_j| > u + \sqrt{2\ln N}\Big\} \leqslant N\exp\big\{-(u+\sqrt{2\ln N})^2/2\big\} \leqslant \exp\{-u^2/2\};$$
$$E\Big\{\max_{0\leqslant j<N}|f_j|^2\Big\} = -\int_0^\infty r^2\,d\psi(r) = 2\int_0^\infty r\psi(r)\,dr \leqslant 2\int_0^{\sqrt{2\ln N}} r\,dr + 2N\int_{\sqrt{2\ln N}}^\infty r\exp\{-r^2/2\}\,dr = 2\ln N + 2. \qquad \square$$

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

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

so that

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

Finally, let

$$\Theta_T = \max_{\tau:\ |\tau|\leqslant 2T}\big|\Delta_1^{\tau_1}\dots\Delta_d^{\tau_d}e\big|^*_{2T,\infty}, \tag{34}$$

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

It makes sense to explain here the ideas behind the proof of Theorem 4. Let $\hat s = (\hat\phi(\Delta)y)_0$, consider the error decomposition (cf. (51)):

$$s_0 - \hat s_0 = s_0 - \big(\hat\phi(\Delta)y\big)_0 = \big(\big(1-\hat\phi(\Delta)\big)s\big)_0 - \big(\hat\phi(\Delta)e\big)_0.$$

One can easily obtain the bound for that the second (stochastic) term of this decomposition – it suffices to use the bound (31) on the norm $|e|^*_{2T,\infty}$ and the norm relation (27). To prove the desired bound for the first (bias) term requires a bit of work.

We can act as follows: we start with constructing the filter $r = (q^*)^2 \in C_{2T}(\mathbb{Z}^d)$ and showing (cf. Lemma 16) that for the "good" filter q^*, the filter r is "good" as well. Further, the ℓ_1-norm $|r|^*_{2T,1}$ of the Fourier transform of r is small enough for r to satisfy the constraints of the optimization problems (6). Then we decompose

$$\big((1-\hat\phi(\Delta))s\big)_0=\big(r(\Delta)\big(1-\hat\phi(\Delta)\big)s\big)_0+\big(\big(1-r(\Delta)\big)\big(1-\hat\phi(\Delta)\big)s\big)_0. \tag{35}$$

Our next objective is to show that the maximal bias $|(1-\hat\phi(\Delta))s|^*_{2T,\infty}$ in the Fourier domain is small. To this end we use the fact that $\hat\phi$ is an optimal solution to (6), so $J(\hat\phi, y^0_{4T})\leqslant J(r, y^0_{4T})$, and that the maximal stochastic error $|(1-\hat\phi(\Delta))e|^*_{2T,\infty}$ in the Fourier domain is bounded due to (31). Using these results we conclude that

$$\big|\big(1-\hat\phi(\Delta)\big)s\big|^*_{2T,\infty}\leqslant\big|\big(1-\hat\phi(\Delta)\big)y\big|^*_{2T,\infty}+\big|\big(1-\hat\phi(\Delta)\big)e\big|^*_{2T,\infty}$$

is also bounded (see Lemma 17).

Using the bound on $|(1-\hat\phi(\Delta))s|^*_{2T,\infty}$ we can bound the first term in the right-hand side of (35): we use again the norm relation (27). To compute the bound for the second term in (35) we use the commutativity of the convolution and the "goodness" of the filter r:

$$\big|\big(\big(1-r(\Delta)\big)\big(1-\hat\phi(\Delta)\big)s\big)_0\big|\leqslant\big|\big(\big(1-r(\Delta)\big)s\big)_0\big|+|\hat\phi|_{2T,2}\big|\big(1-r(\Delta)\big)s\big|_{2T,2}.$$

Observe that the proof of Theorem 4, as presented in the above outline, heavily relies on the properties of the basic blocks of our construction, i.e. the convolution and the Fourier transform. In particular, the relation (27) for the norms in the Fourier and the time domains and the commutative property of the convolution are crucial.

Now let us put the above plan into practice.

1^0. We start the proof with the following technical lemma:

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

$$|r|_2\leqslant|r|^*_{2T,1}\leqslant 2^{d/2}\hat\rho^2(2T+1)^{-d/2}; \tag{36}$$

$$|r|_1\leqslant\hat\rho^2; \tag{37}$$

$$\big[E\big\{\big|\big(1-r(\Delta)\big)s\big|^2_{2T,2}\big\}\big]^{1/2}\leqslant\kappa(\hat\rho+1)(4T+1)^{d/2}; \tag{38}$$

$$\big|\big(1-r(\Delta)\big)y\big|^*_{2T,\infty}\leqslant\big|\big(1-r(\Delta)\big)s\big|_{2T,2}+\big(1+\hat\rho^2\big)\Theta_T; \tag{39}$$

$$\big[E\big\{\big(\big|\big(1-r(\Delta)\big)y\big|^*_{2T,\infty}\big)^2\big\}\big]^{1/2}\leqslant\sigma\big(1+\hat\rho^2\big)\sqrt{4d\ln(4T+1)+2}+\kappa(\hat\rho+1)(4T+1)^{d/2}. \tag{40}$$

Proof. (36): We have

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

Since $|r|_2=|r|_{2T,2}=|r|^*_{2T,2}\leqslant|r|^*_{2T,1}$, (36) follows.

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

(38): 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

$$\max_{\tau:\ |\tau|\leqslant 3T}E\big\{|h_\tau|^2\big\}\leqslant\kappa^2. \tag{41}$$

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

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

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

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

and (38) follows.

(39), (40): We have

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

The resulting inequality combines with (37) to yield (39). Further, from the resulting inequality and (38) it follows that

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

(we have used (37)). To derive (40) from the resulting inequality, it remains to note that

$$\big(E\big\{\Theta_T^2\big\}\big)^{1/2}\leqslant\sigma\sqrt{4d\ln(4T+1)+2}. \tag{43}$$

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 $E\{\Theta_T^2\}\leqslant\sigma^2(4d\ln(4T+1)+2)$ by Lemma 15. □

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

Lemma 17. *One has*

$$|\hat\phi|_{2T,2}\leqslant 2^{d/2}\rho^2(2T+1)^{-d/2}; \tag{44}$$
$$\big|(1-\hat\phi(\Delta))e\big|^*_{2T,\infty}\leqslant\big(1+2^d\rho^2\big)\Theta_T; \tag{45}$$
$$\big[E\big\{\big(\big|(1-\hat\phi(\Delta))e\big|^*_{2T,\infty}\big)^2\big\}\big]^{1/2}\leqslant\sigma\big(1+2^d\rho^2\big)\sqrt{4d\ln(4T+1)+2}; \tag{46}$$
$$\big|(1-\hat\phi(\Delta))s\big|^*_{2T,\infty}\leqslant\big|(1-r(\Delta))s\big|_{2T,2}+2\big(1+2^d\rho^2\big)\Theta_T; \tag{47}$$
$$\big[E\big\{\big(\big|(1-\hat\phi(\Delta))s\big|^*_{2T,\infty}\big)^2\big\}\big]^{1/2}\leqslant 2\sigma\big(1+2^d\rho^2\big)\sqrt{4d\ln(4T+1)+2}+\kappa(\hat\rho+1)(4T+1)^{d/2}. \tag{48}$$

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

(45), (46): We have

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

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

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

(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 (36), so that the optimal value in the problem does not exceed $J(r,y^0_{4T})$. It follows that

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

Relation (47) follows from (c) combined with (39) and (45) (recall that $\hat\rho \leqslant \rho$). Relation (48) follows from (d) combined with (46) and (40). □

3^0. Our next step is to prove

Lemma 18. *One has*

$$\big|\big((1-r(\Delta))(1-\hat\phi(\Delta))s\big)_0\big| \leqslant \big|\big((1-r(\Delta))s\big)_0\big| + 2^{d/2}\rho^2(2T+1)^{-d/2}\big|(1-r(\Delta))s\big|_{2T,2}; \tag{49}$$

$$\big[E\big\{\big|\big((1-r(\Delta))(1-\hat\phi(\Delta))s\big)_0\big|^2\big\}\big]^{1/2} \leqslant \kappa(\hat\rho+1)\big(2^d\rho^2+1\big). \tag{50}$$

Proof. We have

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

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

$$\begin{aligned}\big[E\big\{\big|\big((1-r(\Delta))(1-\hat\phi(\Delta))s\big)_0\big|^2\big\}\big]^{1/2} &\leqslant \big[E\big\{\big|\big((1-r(\Delta))s\big)_0\big|^2\big\}\big]^{1/2} + 2^{d/2}\rho^2(2T+1)^{-d/2}\big[E\big\{\big|(1-r(\Delta))s\big|^2_{2T,2}\big\}\big]^{1/2} \\ &\leqslant \kappa(\hat\rho+1) + 2^{d/2}\rho^2(2T+1)^{-d/2}\big[E\big\{\big|(1-r(\Delta))s\big|^2_{2T,2}\big\}\big]^{1/2} \quad [\text{see (42)}] \\ &\leqslant \kappa(\hat\rho+1) + 2^{d/2}\rho^2(2T+1)^{-d/2}\kappa(\hat\rho+1)(4T+1)^{d/2} \quad [\text{see (38)}]\end{aligned}$$

and (50) follows. □

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

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

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

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

whence, by definition of Θ_T,

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

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

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

We further have

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

(the inequality a is given by (27), and b follows from the feasibility of $\hat\phi$ for (6)), whence

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

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

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

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

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

$$\left|\epsilon_0^{(1)}\right|\leqslant 2^{1+d/2}\rho^2(2T+1)^{-d/2}\left|\left(1-r(\Delta)\right)s\right|_{2T,2}+2^{1+d/2}\rho^2\left(1+2^d\rho^2\right)(2T+1)^{-d/2}\Theta_T+\left|\left(\left(1-r(\Delta)\right)s\right)_0\right|,\tag{57}$$

while from (38), (42) it follows that

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

Therefore (57) implies that

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

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

A.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 $m\geqslant 2$, let T^+ be an integer, $0\leqslant T^+\leqslant L^+$, and let $T=\lfloor m^{-1}T^+\rfloor$. Since $s^j\in\mathbf{F}_L^t(\theta,\rho)$ and clearly $T\leqslant L$, there exist filters q^j such that

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

Now let filter q be defined by

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

Observe that

$$\operatorname{ord}(q)\leqslant mT\leqslant T^+.\tag{61}$$

Note that

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

Indeed, we clearly have

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

((a) is by (60)(b)–(c) since $|u(z)v(z)|_2\leqslant|u|_1|v|_2$, $|u(z)v(z)|_1\leqslant|u|_1|v|_1$, b is due to $\rho_j\geqslant 1$), as required in (62). Further, by (60)(c), for the filters

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

one has

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

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

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

where a is due to independence of λ_j and (s^j) and b follows from (63), (60)(d), and since

$$|\tau'-\tau| \leqslant (m-1)T, \quad |\tau - t| \leqslant L^+ \quad \Rightarrow \quad |\tau' - t| \leqslant L^+ + T^+ \leqslant L.$$

Combining the resulting inequality, (61), (62) and taking into account that $T^+ \in \{0, 1, \dots, L^+\}$ is arbitrary, we conclude that $s \in \mathbf{F}^t_{L^+}(\theta^+, \rho^+)$. Note that by construction, the filters certifying the latter inclusion are independent of λ_j. □

A.2.2. Proof of Proposition 7

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

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

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

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

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

A.2.3. Proof of Proposition 8

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

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

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

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

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

A.2.4. Proof of Proposition 9

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

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

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

$$q \in C_T(\mathbb{Z}^d), \qquad |q|_2 = |q'|_2|q''|_2 \leqslant \rho'\rho''(2T+1)^{-d/2} \tag{66}$$

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

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

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

A.2.5. Proof of Proposition 10

We start with the following two evident facts:

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

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

Lemma 20. *For every $t \in \mathbb{Z}$, the univariate exponential field $(s_\tau = \exp\{\omega\tau\})$, $\omega \in \mathbb{C}$, belongs to $\mathbf{F}_\infty^t(0, \sqrt{2})$.*

Indeed, assuming $\Re(\omega) \geqslant 0$ and given $T \geqslant 0$, 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 $q \in C_T(\mathbb{Z})$, $|q|_2 = (T+1)^{-1/2} \leqslant 2^{1/2}(2T+1)^{-1/2}$, while clearly $q(\Delta)s \equiv s$. In the case of $\Re(\omega) < 0$, the same reasoning holds true for $q(z) = \frac{1}{T+1}[1+\exp\{\omega\}z+\exp\{2\omega\}z^2+\dots+\exp\{T\omega\}z^T]$.

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*

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

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 $\mathbf{O}_j = \{\omega_j(\ell)\colon \ell = 1,\dots,M\}$.

Lemma 21 ⇒ *Proposition* 10: Assume first that the coefficients c_ℓ in (15) are deterministic. Since every one of the univariate functions $f(t) = t^k$, $0 \leqslant k \leqslant m$, is, uniformly on compact sets, the limit, as $\epsilon \to +0$, of appropriate linear combinations of the $m+1$ exponents $\exp\{-k\epsilon\}$, the exponential polynomial (15) is the pointwise, on $\mathbb{Z}^d$, limit, as $i \to \infty$, of simple exponential polynomials (s^i_τ) with extended sets of "frequencies" $\{\omega_j(\ell)\}_{j,\ell}$: in the approximating polynomials, every one of these frequencies is replaced by (m_j+1) frequencies $\omega_j(\ell) - k\epsilon_i$, $0 \leqslant k \leqslant m_j$. 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^i_τ) can be chosen to depend solely on T and the sets of partial frequencies of these approximating polynomials, from the proof of Lemma 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. □

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

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

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

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

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

$$q^{(T)} \in C_T\big(\mathbb{Z}^{d+1}\big), \qquad \left|q^{(T)}\right|_2 = \left|g^{(T)}\right|_2 \left|h^{(T)}\right|_2 \leqslant \rho_d(N_1,\dots,N_d)\rho_1(N_{d+1}) = \rho_{d+1}(N_1,\dots,N_{d+1}) \tag{69}$$

(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 $\tau' \in \mathbb{Z}^d$, $\tau'' \in \mathbb{Z}$:

$$\sum_{\nu\in\mathbb{Z}^{d+1}} q^{(T)}_\nu s_{\tau-\nu} = \sum_{\nu'\in\mathbb{Z}^d} g^{(T)}_{\nu'} \left(\sum_{\nu''\in\mathbb{Z}} h^{(T)}_{\nu''} s_{\tau'-\nu',\tau''-\nu''} \right) \underbrace{=}_{a} \sum_{\nu'\in\mathbb{Z}^d} g^{(T)}_{\nu'} s_{\tau'-\nu',\tau''} \underbrace{=}_{b} s_{\tau',\tau''}$$

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

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

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

A.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 $q^{(T)}$, $\mathrm{ord}(q^{(T)}) \leqslant T$, depending solely on T,m,d, such that

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

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 $T \geqslant 0$ there exists a filter $q \in C_T(\mathbb{Z})$, depending solely on T,m, with $|q|_2 \leqslant 16m(2T+1)^{-1/2}$ such that $p(t) = \sum_\nu q^{(T)}_\nu p(t-\nu)$ for all $t \in \mathbb{Z}$.*

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

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

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

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

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

$$M \equiv \max_{\phi}\left|\rho(\phi)\right| = \max_{|\phi|\leqslant\frac{\pi}{2}}\left|\rho(\phi)\right| \geqslant \left|\phi(0)\right| = 1.$$

By Bernstein's Theorem on trigonometric polynomials, we have $|\rho'(\phi)| \leqslant mM$. 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 $|\rho(\phi)| \geqslant |\rho(\phi_*)| - \frac{1}{2m}(mM) \geqslant M/2$. 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 $n \geqslant (1-\sin(\pi/2-1/(2m)))T-2 \geqslant \frac{T}{16m^2}-2$, whence

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

When $T \geqslant 64m^2$, the latter quantity is $\geqslant \frac{2T+1}{256m^2}$, and in any case $\sum_{t=-T}^{T}\rho^2(\phi_t) \geqslant \rho^2(\phi_0)=1$. Thus, we always have $\sum_{t=-T}^{T}\rho^2(\phi_t) \geqslant \frac{2T+1}{256m^2}$, as required in (71). □

A.2.7. Proof of Proposition 12

In the proof to follow, c_i stand for positive constants depending solely on **D**.

1^0. We start with the following evident observation:

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

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

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

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

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 $D=\{\zeta\in\mathbb{C} \mid |\zeta|\leqslant 1\}$. Let μ be the distribution of values of $\delta|_{\Omega_N^d}$, so that μ is the measure supported by the finite set $\mathbf{M}=\{\zeta \mid \exists\omega\in\Omega_N^d\colon\ \zeta=\delta(\omega)\}$, and $\mu(\{\zeta\})=\sum_{\omega\in\Omega_N^d\colon\ \delta(\omega)=\zeta}\nu(\{\omega\})$. Let also $F(\alpha)=\mu(\{\zeta \mid \Re(\zeta)\geqslant 1-\alpha\})$, $\alpha\geqslant 0$.

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

$$\mathbf{M}\subset\hat{\mathbf{M}}=\bigl\{\zeta \bigm| |\zeta|\leqslant 1,\ \bigl|\Im(\zeta)\bigr|\leqslant c_2^{-1}\bigl(1-\Re(\zeta)\bigr)^{3/2}\bigr\}, \tag{74}$$

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

Proof. (74), (75) are evident when $\sum_{\ell=1}^k\rho_\ell<1$, since then $|\delta(\omega)|\leqslant 1-c_2$ 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 $\sum_{\ell=1}^k\rho_\ell\leqslant 1$ by R.2).

$2^0(1)$. Let $\mathbf{K}=\{\omega\in[-\pi,\pi]^d\colon\ \delta(\omega)=1\}$. 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

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

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

$2^0(2)$. Let $\omega\in\mathbf{K}$, and let $d\omega\in\mathbb{R}^n$ be such that $|d\omega|\leqslant 1$. Then

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

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

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

Since $\mathrm{Card}(\mathbf{K})\leqslant c_3$ by $2^0(1)$ and $|\delta(\omega)|\leqslant 1$ for all ω, we conclude that

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

$2^0(3)$. Now we can complete the proof of (74), (75). Let $\bar\omega\in\mathbf{K}$, $d\omega\in\mathbb{R}^d$, $|d\omega|\leqslant 1$. We have

$$
\begin{aligned}
\delta(\bar\omega+d\omega) &= \sum_{\ell=1}^{k}\rho_\ell \exp\bigl\{i\bigl[\phi_\ell+\bar\omega^T\alpha(\ell)\bigr]\bigr\}\exp\bigl\{i(d\omega)^T\alpha(\ell)\bigr\}\\
&\underbrace{=}_{a}\sum_{\ell=1}^{k}\rho_\ell \exp\bigl\{i(d\omega)^T\alpha(\ell)\bigr\}\\
&= \sum_{\ell=1}^{k}\rho_\ell\left(1+i(d\omega)^T\alpha(\ell)-\frac{1}{2}\bigl((d\omega)^T\alpha(\ell)\bigr)^2-\frac{i}{6}\bigl((d\omega)^T\alpha(\ell)\bigr)^3+r_\ell(\omega,d\omega)\right),
\end{aligned} \tag{79}
$$

$$\bigl[\bigl|r_\ell(\omega,d\omega)\bigr|\leqslant c_{10}|d\omega|^4\bigr]\underbrace{=}_{b}\sum_{\ell=1}^{k}\rho_\ell\left(1-\frac{1}{2}\bigl((d\omega)^T\alpha(\ell)\bigr)^2-\frac{i}{6}\bigl((d\omega)^T\alpha(\ell)\bigr)^3+r_\ell(\omega,d\omega)\right)$$

(for a, see (76), for b, see (20)). Taking into account that $\sum_\ell\rho_\ell=1$ and $c_{11}|d\omega|^2\leqslant\sum_\ell\rho_\ell((d\omega)^T\alpha(\ell))^2\leqslant c_{12}|d\omega|^2$, we conclude from (77) combined with (79) that for properly chosen c_{13} one has

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

and (74) follows. By (74) one has $|1-\delta(\omega)|\leqslant c_{14}(1-\Re(\delta(\omega)))$, so that (75) follows from (78). $\square$

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

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

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 $T_n(\zeta)=\cos(n\,\mathrm{acos}(\zeta))$, whence $T_n(1)=1$, $T_n'(1)=n^2$. From these relations it follows that the function $P_n(\zeta)=\frac{1-T_n(\zeta)}{n^2(1-\zeta)}$ is a polynomial of degree $n-1$, and $P_n(1)=1$.

Lemma 25. *One has*

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

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

$$|\beta| \leqslant c_{16}\alpha^{3/2}. \tag{82}$$

We have

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

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

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

$3^0(1)$. Assume, first, that $0 \leqslant \alpha \leqslant \frac{1}{n^2}$. In this case from (83) it follows that $|1-w| \leqslant \sqrt{2}n^{-1}$, whence, taking into account (83),

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

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

$3^0(2)$. Now consider the case of $\frac{1}{n^2} \leqslant \alpha \leqslant 2$. From (83), (84) it follows that $|w| \leqslant 1 + c_{21}\alpha$, $|w^{-1}| \leqslant 1 + c_{21}\alpha$, whence $|P_n(\zeta)| = \frac{|\frac{w^n+w^{-n}}{2}-1|}{n^2|\alpha - i\beta|} \leqslant \frac{c_{22}(1+c_{21}\alpha)^n}{n^2\alpha}$, as required in (81). □

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

$$\zeta = 1 - \alpha + i\beta \in \hat{\mathbf{M}} \quad\Rightarrow\quad \big|Q(\zeta)\big| \leqslant 1 - c_{23}\alpha \quad \left[c_{23} < \frac{1}{2}\right]. \tag{85}$$

Now let c_{24} be a positive integer which is $\geqslant \frac{c_{15}}{c_{23}}$ (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*:

(a) $\deg(S_n) \leqslant c_{25}n$;

(b) $S_n(1) = 1$;

(c) $$\max_\zeta\big\{\big|S_n(\zeta)\big|:\ \zeta \in \hat{\mathbf{M}},\ \Re(\zeta) = 1 - \alpha\big\} \leqslant c_{15}\min\left[\frac{1}{n^2\alpha}; 1\right]. \tag{86}$$

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

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

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

5^0. 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

(a) $\deg(R_n) \leqslant c_{26} n$;

(b) $R_n(1) = 1$;

(c) $\max_{\zeta}\{|R_n(\zeta)|:\ \zeta \in \hat{\mathbf{M}},\ \Re(\zeta) = 1-\alpha\} \leqslant r_n(\alpha) \equiv c_{26}\min\left[\frac{1}{n^{2d}\alpha^d}; 1\right]$. (87)

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

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

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

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

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

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

(see (75)). We have

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

(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

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

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

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

(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, $0 \leqslant T \leqslant L$. Thus, $\mathbf{H}_N^t(\mathbf{D}) \subset \mathbf{F}^t_{\lfloor c_{29}L\rfloor}(0, c_{34})$. □

A.2.8. Proof of Proposition 13

Lemma 27. *Let $f \in \mathbf{H}^+(M)$ be a deterministic function, let $N \leqslant M/2$, and let $t \in \mathbb{Z}^d$, $|t| \leqslant N$. Consider the "discrete box" $B_N^t = \{\tau \in \mathbb{Z}^d:\ |\tau - t| \leqslant N\}$, and let ϕ be a deterministic function on B_N^t which coincides with f on the "discrete boundary" $\partial B_N^t \equiv \{\tau \in \mathbb{Z}^d:\ |\tau - t| = N\}$ of B_N^t and is "discrete harmonic": $\tau \in \mathbb{Z}^d$, $|\tau - t| < N \Rightarrow \phi_\tau = \frac{1}{2d}\sum_{\epsilon=(\epsilon_1,\ldots,\epsilon_d),\,|\epsilon_1|=\cdots=|\epsilon_d|=1} \phi_{\tau+\epsilon}$. Then*

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

(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" $\tau \in \operatorname{int} B_N^t \equiv B_N^t \backslash \partial B_N^t$ we make a random step of length 1 along one of the coordinate axes, choosing every one of $2d$ possible steps with probability $1/(2d)$. It is immediately seen that the expected penalty payed at the termination, treated as a function of the initial state, is a discrete harmonic function with the boundary values ψ.

Now, since $|t| \leqslant N$ and $2N \leqslant M$, the function f is harmonic in the "continuous box" $D_{2N}^t = \{\tau \in \mathbb{R}^d \colon |\tau - t| \leqslant 2N\}$, 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

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

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

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

Since f is harmonic, we conclude that

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

Now let $h = f|_{\mathbb{Z}^d} - \phi \in C(B_N^t)$ 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 $\tau \in \operatorname{int} B_N^t$:

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

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

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

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

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 $|\phi_\tau - f_\tau|^2 \leqslant c_6^2 \|f\|_{\infty,2M}^2 L^{-4}$ for $\tau \in B_{2L}^t$. From (97) it now follows that

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

(recall that $d \leqslant 4$). □

The proof of Proposition 14. It 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, P. Massart, Risk bounds for model selection via penalization, Probab. Theory Related Fields 113 (3) (1999) 301–413.
[3] L. Birgé, An alternative point of view on Lepski's method, in: State of the Art in Probability and Statistics, Leiden, 1999, pp. 113–133.
[4] L. Birgé, P. Massart, Gaussian model selection, J. Eur. Math. Soc. 3 (3) (2001) 203–268.
[5] T. Cai, Adaptive wavelet estimation: A block thresholding and oracle inequality approach, Ann. Statist. 27 (3) (1999) 898–924.
[6] T. Cai, M. Low, Adaptive estimation of linear functionals under different performance measures, Bernoulli 11 (2005) 341–358.
[7] T. Cai, M. Low, On adaptive estimation of linear functionals, Ann. Statist. 33 (2005) 2311–2343.

[8] D. Donoho, M. Low, Renormalization exponents and optimal pointwise rates of convergence, Ann. Statist. 20 (2) (1992) 944–970.
[9] D. Donoho, I. Johnstone, Ideal spatial adaptation via wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[10] D. Donoho, Statistical estimation and optimal recovery, Ann. Statist. 22 (1) (1995) 238–270.
[11] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Wavelet shrinkage: Asymptopia? J. Roy. Statist. Soc. Ser. B 57 (2) (1995) 301–369.
[12] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, Wavelet shrinkage: Asymptopia? (with discussion and reply by the authors) J. Roy. Statist. Soc. Ser. B 57 (2) (1995) 301–369.
[13] A. Goldenshluger, A. Nemirovski, On spatially adaptive estimation of nonparametric regression, Math. Methods Statist. 6 (2) (1997) 135–170.
[14] A. Goldenshluger, O. Lepski, Structural adaptation via *Lp*-norm oracle inequalities, Probab. Theory Related Fields, in press.
[15] I. Ibragimov, R. Khasminskii, Nonparametric estimation of the value of a linear functional in Gaussian white noise, Theory Probab. Appl. 29 (1984) 1–32.
[16] A. Juditsky, Wavelet estimators: Adapting to unknown smoothness, Math. Methods Statist. 6 (1) (1997) 1–25.
[17] A. Juditsky, A. Nemirovski, Oracle inequalities for adaptive filtering problem, Ann. Inst. H. Poincare, submitted for publication.
[18] A. Juditsky, A. Nemirovski, Nonparametric denoising of signals of unknown local structure, II: Nonparametric regression estimation, Appl. Comput. Harmon. Anal., submitted for publication.
[19] A. Kneip, Ordered linear smoothers, Ann. Statist. 22 (2) (1994) 835–866.
[20] O. Lepski, On a problem of adaptive estimation in Gaussian white noise, Theory Probab. Appl. 35 (3) (1990) 454–466.
[21] O. Lepski, Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates, Theory Probab. Appl. 36 (4) (1991) 682–697.
[22] O. Lepski, Asymptotically minimax adaptive estimation. II. Statistical model without optimal adaptation. Adaptive estimators, Theory Probab. Appl. 37 (1992) 433–448.
[23] O. Lepski, V. Spokoiny, Optimal pointwise adaptive methods in nonparametric estimation, Ann. Statist. 25 (6) (1997) 2512–2546.
[24] O. Lepski, B. Levit, Adaptive nonparametric estimation of smooth multivariate functions, Math. Methods Statist. 8 (1999) 344–370.
[25] A. Nemirovski, On forecast under uncertainty, Problemy Peredachi Informatsii 17 (4) (1981) 73–83, English transl. in Probl. Inf. Transm. 17 (1981).
[26] A. Nemirovski, Denoising signals of unknown local structure, Medallion Lecturer of IMS, JSM 2003, San Francisco, August 3–5, 2003.
[27] M. Pinsker, S. Efromovitch, Learning algorithm for nonparametric filtering, Autom. Remote Control 45 (11) (1984) 1434–1440.
[28] A. Tsybakov, Pointwise and sup norm sharp adaptive estimation of functions on the Sobolev classes, Ann. Statist. 26 (1998) 2469–2520.