# Nonparametric reconstruction of a multifractal function from noisy data

Arnaud Gloter and Marc Hoffmann

*Université Paris-Est and CNRS (UMR 8050)*

**Abstract**

We estimate a real-valued function $f$ of $d$ variables, subject to additive Gaussian perturbation at noise level $\varepsilon > 0$, under $L^\pi$-loss, for $\pi \geq 1$. The main novelty is that $f$ can have an extremely varying local smoothness, exhibiting a so-called multifractal behaviour. The results of Jaffard on the Frisch-Parisi conjecture suggest to link the singularity spectrum of $f$ to Besov properties of the signal that can be handled by wavelet thresholding for denoising purposes.

We prove that the optimal (minimax) rate of estimation of multifractal functions with singularity spectrum $d(H)$ has explicit representation $\varepsilon^{2v(d(\bullet),\pi)}$, with

$$v(d(\bullet),\pi) = \min_H \frac{H + \big(d - d(H)\big)/\pi}{2H+d}.$$

The minimum is taken over a specific domain and the rate is corrected by logarithmic factors in some cases. In particular, the usual rate $\varepsilon^{2s/(2s+d)}$ is retrieved for monofractal functions (with spectrum reduced to a single value $s$) irrespectively of $\pi$. More interestingly, the sparse case of estimation over single Besov balls has a new interpretation in terms of multifractal analysis.

*Keywords:* multifractal analysis; Frisch-Parsisi conjecture; wavelet threshold algorithm; minimax rate of convergence.
*Mathematical Subject Classification:* 62G05; 62G20.
*e-mail address:* arnaud.gloter@univ-mlv.fr, marc.hoffmann@univ-mlv.fr
Laboratoire d'Analyse et de Mathématiques Appliquées, 5 boulevard Descartes, 77454 Marne–la–Vallée, Cedex 2, France.

# 1 Introduction

We consider the usual formulation of signal denoising in nonparametric estimation: we want to recover a real-valued function $f$ defined on a regular bounded domain $\mathcal{D} \subset \mathbb{R}^d$. We can make linear measurements, but each measurement is contaminated by systematic noise: we observe

$$Y_\varepsilon = f + \varepsilon \dot{W}, \tag{1.1}$$

where $\dot{W}$ is a Gaussian white noise on $L^2(\mathcal{D})$ and $\varepsilon > 0$ a noise level. Asymptotics are taken as $\varepsilon \to 0$. Observable quantities take the form

$$Y_\varepsilon(\varphi) := \langle \varphi, f \rangle + \varepsilon\, \xi(\varphi),$$

where $\varphi \in L^2(\mathcal{D})$ is a test function and $\langle \bullet, \bullet \rangle$ denotes the inner product on $L^2(\mathcal{D})$. The random process $\xi(\varphi)$ is centred Gaussian, with covariance $\mathbb{E}[\xi(\varphi)\xi(\psi)] = \langle \varphi, \psi \rangle$ for $\varphi, \psi \in L^2(\mathcal{D})$. The symbol $\mathbb{E}[\bullet]$ denotes mathematical expectation. This setting is meaningless without further smoothness properties on the signal $f$. A commonly used assumption is that $f$ belongs to a Besov ball

$$B^s_{p,\infty}(r) := \big\{ f \in B^s_{p,\infty},\ \|f\|_{B^s_{p,\infty}} \le r \big\}, \tag{1.2}$$

for some $r > 0$, with the additional condition $s - d/p > 0$ so that the functions in $B^s_{p,\infty}(r)$ are all continuous. Here, $B^s_{p,\infty}$ denotes the Besov space on $\mathcal{D}$, appended with boundary conditions and $p$ ranges in $(0, +\infty)$; more in Section 3.1 below.

In this paper, we are interested in signals that possess local smoothness in a Hölder sense that *vary extremely* from one point to the other and that we shall informally refer to as *multifractal* before getting to a rigorous definition. In this context, the classical approach of single Besov balls (1.2) needs to be generalized.

## 1.1 Multifractal analysis

Let $x_0 \in \mathcal{D}$, $\alpha > 1$. Following Jaffard [23], we say that $f : \mathcal{D} \to \mathbb{R}$ is $\mathcal{C}^\alpha(x_0)$ if there exists $c > 0$ and a polynomial $P_{x_0}$ of degree at most $[\alpha]$ such that in a neighbourhood of $x_0$:

$$|f(x) - P_{x_0}(x)| \le c|x - x_0|^\alpha. \tag{1.3}$$

If $\alpha \in (0,1]$, we simply replace $P_{x_0}(x)$ by $f(x_0)$. The Hölder exponent of $f$ at $x_0$ is

$$h_f(x_0) := \sup\big\{\alpha > 0,\ f \in \mathcal{C}^\alpha(x_0)\big\}. \tag{1.4}$$

The level sets

$$S_f(H) := \{x \in \mathcal{D},\ h_f(x) = H\} \tag{1.5}$$

with maximal Hölder regularity $H$ of the functions we want to consider have typical Lebesgue measure zero, see [23], and in this setting, it seems more appropriate to consider the Hausdorff dimension[1] of $S_f(H)$. The function $d(H) := \dim S_f(H)$ is called the *Hölder* or *singularity spectrum of* $f$ and is extended to the whole line by setting $\dim(H) = -\infty$ if $H$ is nowhere the Hölder exponent of $f$.

**Definition 1.1.** *A function $f : \mathcal{D} \to \mathbb{R}$ is multifractal if its spectrum of singularity $d(H) \neq -\infty$ at least on an interval of non-empty interior.*

## 1.2 Empirical multifractal evidence

Empirical evidence of multifractal behaviour in signal modelling was first obtained in velocity fields of fully developed turbulent flows [13, 14, 15, 16] around 1980, and lays its roots in the theoretical founding papers of Kolmogorov and Obukov [27, 32] in 1962. More recently, the multifractal approach has been introduced in traffic networks [33], coding sequences in genome analysis [1, 34, 36], financial data [19, 2, 3, 28, 29], and Bayesian statistical analysis [17, 18]. Clearly however, the spectrum of singularity $d(H)$ defined by (1.4) and (1.5) is an asymptotic notion that cannot be related to quantities that are measured with limited accuracy or in presence of noise. The link between $d(H)$ and related observable objects can be given by the Frisch-Parisi conjecture which reads[2]

$$d(H) = \inf_p \big\{pH - ps(1/p) + d\big\}, \tag{1.6}$$

where the exponent $s(\bullet)$ is defined pointwise by

$$s(1/p) := \sup\big\{s \geq 0,\ f \in B^s_{p,\infty}\big\}.$$

The use of correspondence (1.6) suggests a strategy to define a consistent statistical setup since Besov norms are tractable functionals that can be

[1] We recall the definition of the Hausdorff dimension in the appendix for sake of completeness.

[2] A classical heuristic derivation of (1.6) is proposed in Appendix 6.1.

estimated in presence of noise by wavelet thresholding [9, 10, 11, 12, 26, 6]. Of course, the range of $H$ and $p$ for which (1.6) can be valid must be assessed precisely, see Section 3.4. When (1.6) holds, we informally say that $f$ satisfies *a multifractal formalism.* It is noteworthy that in all the applied fields cited above, it is a Besov-type related quantity, the so-called *structure function*, that measures statistical evidence of multifractality. The structure function can be defined as

$$M_j(f, 1/p) := 2^{-j} \sum_{k=1}^{2^j} \left| f(k2^{-j}) - f\big((k-1)2^{-j}\big) \right|^p, \tag{1.7}$$

in dimension $d = 1$ with $\mathcal{D} = [0, 1]$ for notational simplicity. Multifractal empirical evidence corresponds to a scaling law of the type

$$M_j(f, 1/p) \approx 2^{-jps(1/p)}, \ \ j \to \infty \tag{1.8}$$

where the function $1/p \rightsquigarrow s(1/p)$ is *not constant.* Indeed, interpreting the local fluctuation of $f$ around $k2^{-j}$ as a wavelet coefficient up to rescaling, we have

$$\sup_j 2^{js(1/p)} M_j(f, 1/p)^{1/p} \approx \|f\|_{B^{s(1/p)}_{p,\infty}}, \tag{1.9}$$

so having $s(\bullet)$ not being constant and putting together (1.9)–(1.6) yields the notion of multifractality of Definition 1. The precise meaning of (1.9) together with the link to Besov norms will become transparent in Sections 2 and 3 below, at least for small values of $s(\bullet)$.

### 1.3 Organisation and results of the paper

In Section 2, we define rigorously our statistical setting and the corresponding multifractal formalism by means of so-called *Besov domains* inspired by Jaffard [23] and that enables us to consider multifractal signals without loosing the standard minimax approach of recovering functions in Besov spaces.

An upper bound for estimating signals within a prescribed Besov domain is given in $L^\pi$-loss error in Theorem 3.4, for $\pi \geq 1$. It is achieved by the wavelet threshold algorithm, and the proof heavily relies on the modern formulation of wavelet estimation over atomic spaces, as introduced by Cohen *et al.* [6]. We make a systematic use of embeddings properties of strong Besov spaces into weak Besov spaces, thanks to the results of Kerkyacharian and Picard [26], a key reference for the paper. These powerful analytical tools render the proof of Theorem 3.4 quite simple, yet technical. This result is optimal by Theorem 3.5, up to logarithmic factors. This is actually the most delicate part of the paper: in order to prove a lower bound, an appropriate prior has to be chosen over functions which are genuinely multifractal with a prescribed singularity spectrum. A nontrivial construction is proposed, using tools developed in Jaffard [23], another central reference to the paper.

The translation in terms of multifractal analysis and Hölder spectrum whenever the Frisch-Parisi conjecture holds is given in Theorem 3.8. We show that the minimax rate of estimation of multifractal functions with singularity spectrum $d(H)$ has explicit representation $\varepsilon^{2v(d(\bullet),\pi)}$, with

$$v(d(\bullet),\pi) = \min_H \frac{H + \big(d - d(H)\big)/\pi}{2H + d}.$$

The minimum is taken over a specific domain and the rate is corrected by logarithmic factors in $\varepsilon$ in some cases. As a consequence, the classical rate $\varepsilon^{2s/(2s+d)}$ is retrieved for monofractal functions having singularity spectrum $d(s) = d$ and $d(H) = -\infty$ for $H \neq s$, irrespectively of the $L^\pi$-loss. More interestingly, the sparse case of estimation over single Besov balls has a new interpretation in terms of multifractal analysis, as well as the critical case that separates dense and sparse regimes. Examples and applications are derived in Section 4. In particular, we revisit and somewhat improve former results of Hall, Kerkyacharian and Picard [20] on estimation of noisy signals exhibiting abberations of chirp or Doppler type, thanks to the multifractal approach. The proofs are delayed until Section 5 and auxiliary technical results are given in an appendix (Section 6).

# 2 Multifractal formalism and signal estimation

## 2.1 Besov domains and function classes

We generalize the scale of Besov classes (1.2) of the Introduction by describing the approach of Besov domains. For $f \in L^2$, the minimal assumption so that the statistical model (1.1) is well defined, we have the following:

**Definition 2.1.** *The scaling function of $f$ is*

$$1/p \rightsquigarrow s_f(1/p) := \sup\{s \geq 0,\ f \in B^s_{p,\infty}\}.$$

The function $s_f(\bullet)$ is defined pointwise over a domain of $1/p \subset [0,+\infty)$ that contains at least $[1/2,+\infty)$ since $f \in L^2 = B^0_{2,2} \subset B^0_{2,\infty}$. This domain may contain 0 if we allow for $p = \infty$. The function $s_f(\bullet)$ can take the value $+\infty$, in which case it becomes trivially equal to $+\infty$ over $[0,+\infty)$. This is a consequence of the following simple lemma:

**Lemma 2.2.** *The function $s_f(\bullet)$ is increasing, concave and satisfies $s'_f(\bullet) \leq d$.*

Here, $s'_f(\bullet)$ denotes the left-derivative of $s_f(\bullet)$. We adopt the same convention for any concave function in the sequel.

*Proof.* Since $\mathcal{D}$ is bounded, the spaces $B^s_{p,\infty}$ are decreasing in $p$ thus $1/p \rightsquigarrow s_f(1/p)$ is increasing. If $f$ belongs to $B^{s_1}_{p_1,\infty} \cap B^{s_2}_{p_2,\infty}$, then $f \in B^{s_3}_{p_3,\infty}$ for $s_3 = us_1 + (1-u)s_2$ and $1/p_3 = u/p_1 + (1-u)/p_2$ for all $u \in [0,1]$ by interpolation hence $s_f(\bullet)$ must be concave. Finally, the Sobolev embedding $B^{s_1}_{p_1,\infty} \subset B^{s_2}_{p_2,\infty}$ if $s_1 - d/p_1 = s_2 - d/p_2$, $p_2 \geqslant p_1$ yields the bound $s'_f(\bullet) \leq d$. $\square$

Conversely, we adopt the following:

**Definition 2.3.** *A non-decreasing concave function $s(\bullet) : [0,+\infty) \to \mathbb{R}$ such that $s(0) > 0$ is called admissible. The Besov domain of an admissible function $s(\bullet)$ is the set of functions defined by*

$$\mathcal{M}\big(s(\bullet)\big) := \big\{ f \in L^2,\ \ \forall\ 1/p \in [0,+\infty),\ \ s_f(1/p) \geq s(1/p) \big\}.$$

**Remark 2.4.** *The assumption $s(0) > 0$ guarantees some uniform Hölder regularity since $\mathcal{M}\big(s(\bullet)\big) \subset B^{s(0)-\varepsilon}_{\infty,\infty}$ for all $\varepsilon > 0$. In particular, the functions in $\mathcal{M}\big(s(\bullet)\big)$ are continuous over $\mathcal{D}$. This assumption is crucial for the interpretation of Jaffard's theorem in Section 2.2 in terms of multifractal analysis, but not essential as far as statistical estimation is concerned. We intend to describe extensions beyond the continuous case in a forthcoming work.*

The Besov domain $\mathcal{M}\big(s(\bullet)\big)$ coincides with the space

$$\bigcap_{p \in (0,+\infty)} \bigcap_{\varepsilon > 0} B^{s(1/p)-\varepsilon}_{p,\infty}$$

that consists of functions that saturate their smoothness in $L^p$ with the exponent $s(1/p)$ simultaneously for all $p \in (0,+\infty)$. In particular, we cover the case of single Besov spaces in the following sense: let $s_0, p_0$ satisfy $s_0 - d/p_0 > 0$ and define

$$s_{s_0,p_0}(1/p) := s_0 + d\big(1/p - 1/p_0\big) \ \text{ if } \ 0 \leq 1/p \leq 1/p_0,$$

and $s_{s_0,p_0}(1/p) := s_0$ otherwise. Clearly, $s_{s_0,p_0}(\bullet)$ is admissible and defines the Besov domain $\mathcal{M}\big(s_{s_0,p_0}(\bullet)\big)$ with the following property

$$\mathcal{M}\big(s_{s_0,p_0}(\bullet)\big) = \bigcap_{\varepsilon>0} B^{s_0-\varepsilon}_{p_0,\infty}.$$

We also introduce the restriction of functions of $\mathcal{M}\big(s(\bullet)\big)$ with prescribed radius in all $B^s_{p,\infty}$ (quasi)-norms: for $r > 0$, the Besov domain with radius $r > 0$ of an admissible function $s(\bullet)$ is defined by

$$\mathcal{M}\big(s(\bullet), r\big) := \Big\{f \in \mathcal{M}\big(s(\bullet)\big), \quad \sup_{p\in(0,+\infty]} \|f\|_{B^{s(1/p)}_{p,\infty}} \leq r\Big\}.$$

## 2.2 Besov domains and multifractal functions

Let us first recall Jaffard's theorem in our context: if $s(\bullet)$ is admissible, define

$$1/p_c := \inf\big\{t > 0, \ \ s(t) \leq dt\big\}, \tag{2.1}$$

which equals $+\infty$ in the extremal case where $s'(\bullet) = d$ in a neighbourhood of $+\infty$.

**Proposition 2.5.** *(Theorem 1 in [23]). If $s(\bullet)$ is admissible, the spectrum of singularity of quasi-all function of $\mathcal{M}\big(s(\bullet)\big)$ is $[s(0), d/p_c]$ and is given by*

$$d(H) = \inf_{p\geq p_c} \big\{Hp - ps(1/p) + d\big\}. \tag{2.2}$$

So we interpret $\mathcal{M}\big(s(\bullet)\big)$ as containing multifractal functions with spectrum of singularity satisfying (2.2). The class $\mathcal{M}\big(s(\bullet)\big)$ is however too big and contains functions $g$ with smoother scaling function, in the sense that

$$s_g(1/p) \geq s(1/p), \ \ p \in (0, +\infty).$$

This includes in particular monofractal functions for which $s_g(\bullet)$ is constant, see Section 4.1. Nevertheless, the minimax methodology forces optimal rates of convergence to be governed by multifractal functions that sit at the "boundary" of $\mathcal{M}\big(s(\bullet)\big)$ and for which (2.2) holds exactly. This will become transparent in Section 3.4 below.

# 3 Main result

An estimator $\widehat{f}$ of $f$ is a measurable function of the observation $Y_\varepsilon$ defined in (1.1). We measure its performance in $L^\pi$-loss error simultaneously for all $\pi \geq 1$ over the class $\mathcal{M}\big(s(\bullet), r\big)$ by setting

$$\mathcal{E}_\pi\big(\widehat{f}\big) := \sup_{f \in \mathcal{M}(s(\bullet), r)} \mathbb{E}\left[\|\widehat{f} - f\|_{L^\pi}^\pi\right]^{1/\pi}. \tag{3.1}$$

We look for an estimator $\widehat{f}$ independent of $s(\bullet)$ and $r > 0$ with minimal error $\mathcal{E}_\pi\big(\bullet\big)$.

## 3.1 Wavelet bases and superconcentration

Wavelets are documented in numerous textbooks[3]. We use regular wavelet bases $(\psi_\lambda)_\lambda$ adapted to the domain $\mathcal{D}$. The multi-index $\lambda$ concatenates the spatial index and the resolution level $j = |\lambda|$. We set $\Lambda_j := \{\lambda,\ |\lambda| = j\}$ and $\Lambda := \cup_{j \geq -1} \Lambda_j$. Thus, for $f \in L^p$, we have

$$f = \sum_{j \geq -1} \sum_{\lambda \in \Lambda_j} f_\lambda \psi_\lambda = \sum_{\lambda \in \Lambda} f_\lambda \psi_\lambda, \quad \text{with } \ f_\lambda := \langle f, \psi_\lambda \rangle,$$

where we have set $j := -1$ in order to incorporate the low frequency part of the decomposition. From now on the basis $(\psi_\lambda)_\lambda$ is fixed. Let $c(\mathcal{D})$ denote a constant such that $\operatorname{Card} \Lambda_j \leqslant c(\mathcal{D})^d 2^{jd}$.

**Definition 3.1.** *For $s > 0$ and $p \in (0, \infty]$, $f$ belongs to $B^s_{p,\infty}$ if the following norm is finite:*

$$\|f\|_{B^s_{p,\infty}} := c(\mathcal{D})^{s-d/p} \sup_{j \geq -1} 2^{j\left(s + d(\frac{1}{2} - \frac{1}{p})\right)} \Big( \sum_{\lambda \in \Lambda_j} |\langle f, \psi_\lambda \rangle|^p \Big)^{1/p} \tag{3.2}$$

*with the usual modification if $p = \infty$.*

Precise connection between this definition of Besov norm and more standard ones can be found in [5]. Given a basis $(\psi_\lambda)_\lambda$, there exists $\sigma > 0$ such that for $p \geqslant 1$ and $s \leqslant \sigma$ the Besov space defined by (3.2) exactly matches the usual definition in terms of modulus of smoothness for $f$. The index $\sigma$ can be taken arbitrarily large. Taking (3.2) as a definition

[3] We follow closely the notation of Cohen [5].

is technically convenient in the sequel. In particular, by incorporating the factor $c(\mathcal{D})^{s-d/p}$ into the definition, the norm $\|f\|_{B^s_{p,\infty}}$ decreases as $p$ decreases. Moreover the following Sobolev embedding and interpolation inequalities hold without inflating the norm:

$$\|f\|_{B^{s_2}_{p_2,\infty}} \leqslant \|f\|_{B^{s_1}_{p_1,\infty}} \text{ for } s_1 - d/p_1 = s_2 - d/p_2, p_2 \geqslant p_1 \tag{3.3}$$

$$\|f\|_{B^{s_3}_{p_3,\infty}} \leqslant \|f\|^u_{B^{s_1}_{p_1,\infty}} \|f\|^{1-u}_{B^{s_2}_{p_2,\infty}} \tag{3.4}$$

for $s_3 = us_1 + (1-u)s_2$, with $1/p_3 = u/p_1 + (1-u)/p_2$, for $u \in [0,1]$. The additional properties of the wavelet basis $(\psi_\lambda)_\lambda$ that we need are summarized in the next assumption.

**Assumption 3.2.** *For $\pi \geq 1$:*

- *We have*

$$\|\psi_\lambda\|^\pi_{L^\pi} \sim 2^{|\lambda| d(\pi/2-1)}. \tag{3.5}$$

- *For some $\sigma > 0$ and for all $s \leqslant \sigma$, $j_0 \geqslant 0$, we have*

$$\|f - \sum_{j \leqslant j_0} \sum_{\lambda \in \Lambda_j} f_\lambda \psi_\lambda\|_{L^\pi} \lesssim 2^{-j_0 s} \|f\|_{B^s_{\pi,\infty}}. \tag{3.6}$$

- *For any subset $\Lambda_0 \subset \Lambda$*

$$\int_{\mathcal{D}} \Big( \sum_{\lambda \in \Lambda_0} |\psi_\lambda(x)|^2 \Big)^{\pi/2} dx \sim \sum_{\lambda \in \Lambda_0} \|\psi_\lambda\|^\pi_{L^\pi}. \tag{3.7}$$

- *If $\pi > 1$, for any sequence $(u_\lambda)_{\lambda \in \Lambda}$*

$$\big\| \big( \sum_{\lambda \in \Lambda} |u_\lambda \psi_\lambda|^2 \big)^{1/2} \big\|_{L^\pi} \sim \| \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda \|_{L^\pi}. \tag{3.8}$$

The symbol $\sim$ means inequality in both ways, up to a constant depending on $\pi$ and $\mathcal{D}$ only. The property (3.6) reflects that our definition (3.2) of Besov spaces matches the definition in term of linear approximation.

Property (3.8) reflects an unconditional basis property, see [26, 7] and (3.7) is referred to as a superconcentration inequality, or Temlyakov property [26]. The formulation of (3.7)-(3.8) in the context of statistical estimation is posterior to the original papers of Donoho and Johnstone [9, 10] and Donoho *et al.* [11, 12] and is due to Kerkyacharian and Picard

[26]. The existence of compactly supported wavelet bases satisfying Assumption 3.2 is discussed in [30], see also [5].

**The threshold algorithm**. We consider the classical hard threshold estimator. For $x \in \mathbb{R}$ and $\pi \geq 1$, introduce

$$T_{\varepsilon,\pi}(x) := x\, 1_{\left\{|x| \geq \kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right\}}, \quad \kappa(\pi) := 4\sqrt{\max\{\pi, 2\}}. \tag{3.9}$$

We consider estimators of the form

$$\widehat{f}_{\varepsilon,\pi} := \sum_{|\lambda| \leq J_\varepsilon} T_{\varepsilon,\pi}(\hat{f}_\lambda)\psi_\lambda, \;\; 2^{-J_\varepsilon} := \left(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\right)^{2/d}.$$

The empirical wavelet coefficients $\hat{f}_\lambda$ are defined by

$$\hat{f}_\lambda := Y_\varepsilon(\psi_\lambda) = f_\lambda + \varepsilon\, \xi(\psi) \tag{3.10}$$

with $\xi(\psi)$ a standard normal random variable by (1.1) since $\|\psi_\lambda\|_{L^2} = 1$. Thus $\widehat{f}_{\varepsilon,\pi}$ is specified by $\pi$ and the choice of the basis $(\psi_\lambda)_\lambda$ only. The choice of $\kappa(\pi)$ in (3.9) is motivated by the following estimates

**Lemma 3.3.** *For all $\pi \geq 1$, we have*

$$\mathbb{E}\left[|\hat{f}_\lambda - f_\lambda|^\pi\right] \lesssim \varepsilon^\pi \tag{3.11}$$

*and*

$$\mathbb{P}\left[|\hat{f}_\lambda - f_\lambda| \geq \tfrac{\kappa(\pi)}{2}\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\right] \lesssim \varepsilon^{2\max\{\pi,2\}}. \tag{3.12}$$

*where $\lesssim$ means inequality up to constants depending on $\pi$ only.*

*Proof.* Inequality (3.11) readily follows from (3.10). If $\xi$ is standard normal, we have $\mathbb{P}[|\xi| \geq t] \leq \exp(-t^2/2)$ for $t > 0$ so the left-hand side in (3.12) is less than $\varepsilon^{\kappa(\pi)^2/2} = \varepsilon^{2\max\{\pi,2\}}$ thanks to the choice of $\kappa(\pi)$. □

## 3.2 Upper bound

We introduce the fundamental equation

$$s(1/p) = \frac{d}{2}\left(\frac{\pi}{p} - 1\right). \tag{3.13}$$

We define $p^\star$ as the necessarily unique solution of (3.13) if it exists. Notice that the index $s(1/p^\star)$ depends on $d$ and $\pi$.

**Theorem 3.4.** *Grant Assumption 3.2 for some $\sigma > 0$. Let $\pi \geq 1$ and assume that $s(\bullet)$ is admissible.*

- *If $s'(\infty) < d\pi/2$ (which is always true if $\pi > 2$), the solution $p^\star$ to (3.13) exists.*
- *For $\sigma \geqslant s(1/\pi)$, we have*
$$\mathcal{E}_\pi\big(\widehat{f}_{\varepsilon,\pi}\big) \lesssim \Big(\varepsilon\sqrt{\log \tfrac{1}{\varepsilon}}\Big)^{2s(1/p^\star)/(2s(1/p^\star)+d)} \Big(\log \frac{1}{\varepsilon}\Big)^{1/\pi}, \tag{3.14}$$
*where $\lesssim$ means up to a constant depending on $\pi$, $s(\bullet)$ and $r$ only.*
- *Extremal case: If $s'(\bullet) = 0$ in a neighbourhood of $1/p^\star$ then we have the refinement*
$$\mathcal{E}_\pi\big(\widehat{f}_{\varepsilon,\pi}\big) \lesssim \Big(\varepsilon\sqrt{\log \tfrac{1}{\varepsilon}}\Big)^{2s(1/p^\star)/(2s(1/p^\star)+d)}.$$

## 3.3 Lower bound

The next result states that the rate obtained in (3.14) is the best one, up to a logarithmic correction.

**Theorem 3.5.** *Let $\pi \geqslant 1$, assume that $s(\bullet)$ is admissible and $s'(\infty) < d\pi/2$.*

- *If $s'(1/p^\star) > 0$ then*
$$\inf_{\widehat{f}} \mathcal{E}_\pi\big(\widehat{f}\big) \gtrsim \Big(\varepsilon\sqrt{\log \tfrac{1}{\varepsilon}}\Big)^{2s(1/p^\star)/(2s(1/p^\star)+d)}, \tag{3.15}$$
*where the infimum is taken over all estimators and $\gtrsim$ means up to constant depending on $\pi$, $s(\bullet)$ and $r$ only.*
- *Extremal case: If $s'(1/p^\star) = 0$ we have*
$$\inf_{\widehat{f}} \mathcal{E}_\pi\big(\widehat{f}\big) \gtrsim \varepsilon^{2s(1/p^\star)/(2s(1/p^\star)+d)}.$$

**Remark 3.6.** *The distinction $s'(1/p^\star) = 0$ versus $s'(1/p^\star) > 0$ corresponds to the standard distinction between 'dense' case and 'sparse' case, see Section 4.2 below.*

**Remark 3.7.** *If $s'(\infty) \geqslant d\pi/2$ then the equation* (3.13) *has no solution. However we can choose $\pi' > \pi$ in a way that the equation* (3.13)*, where $\pi'$ replaces $\pi$, has now a solution $p^\star$ as large as we want. Applying the upper bound* (3.14) *with $\pi'$, and using that necessarily $s(\infty) = \infty$, we deduce that the $L^\pi$ risk is bounded by $\varepsilon^\gamma$ with $\gamma$ arbitrarily close to* 1*. Hence, we might conjecture that, when $s'(\infty) \geqslant d\pi/2$, the rate of the statistical problem is essentially the parametric rate $\varepsilon$. In this paper, we do not consider this situation.*

### 3.4 Minimax rates under the Frisch-Parisi conjecture

We may now interpret Theorem 3.4 and 3.5 in the light of the Frisch-Parisi conjecture, thanks to Jaffard's theorem given in Section 2.2 above. Let us denote by $\mathcal{FP}\big(s(\bullet), r\big)$ the subset of $\mathcal{M}\big(s(\bullet), r\big)$ consisting of functions satisfying

$$d(H) = \inf_{p \geq p_c} \big\{Hp - ps(1/p) + d\big\}, \quad \forall H \in [s(0), d/p_c],$$

recall (2.2) and the definition of $p_c$ in (2.1). We interpret $\mathcal{FP}\big(s(\bullet), r\big)$ as the class of genuine multifractal functions as soon as $s(\bullet)$ is not constant. A consequence of Theorems 3.4 and 3.5 is the following expression for the minimax risk over $\mathcal{FP}\big(s(\bullet), r\big)$:

**Theorem 3.8.** *In the same setting as Theorems 3.4 and 3.5 and for $\pi \geq p_c + 2$, we have*

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{FP}(s(\bullet), r)} \mathbb{E}\left[\|\widehat{f} - f\|_{L^\pi}^\pi\right]^{1/\pi} \approx \varepsilon^{2v(d(\bullet),\pi)}, \tag{3.16}$$

*where the infimum is taken over all estimators and*

$$v(d(\bullet), \pi) = \min_{s(0) \leq H \leq d/p_c} \frac{H + \big(d - d(H)\big)/\pi}{2H + d}, \tag{3.17}$$

*with $d(\bullet)$ given by formula* (2.2)*. The notation $\approx$ loosely means equivalence up to constants, possibly corrected by logarithmic factors in $\varepsilon$, subject to the same restrictions as in Theorems 3.4 and 3.5.*

**Remark 3.9.** *Formula* (3.17) *quantifies the connection between local smoothness of a signal and its rate of estimation: it shows how the effect of points with regularity $H$ is balanced by the frequency of such points, assessed by $d(H)$. In the 'dense' case $s'(1/p^\star) = 0$, the infimum in* (3.17) *is attained at some $H$ satisfying $d(H) = d$, whereas in the 'sparse' case it is attained at some $H$ with $d(H) < d$.*

**Remark 3.10.** *It is clear by* (2.2) *that the singularity spectrum* $d(\bullet)$ *contains no information about* $s(1/p)$ *for* $p < p_c$*. Thus it is impossible to relate the results of Theorem 3.4 and 3.5 with the spectrum in the case* $p^\star < p_c$*. Actually the condition* $\pi \geq p_c + 2$ *ensures that this situation does not happen.*

**Remark 3.11.** *By a result of Jaffard [22], if a function* $f \in \mathcal{M}(s(\bullet), r)$ *does not satisfy the Frisch Parisi conjecture, we still have the information on its spectrum* $d(H) \leqslant \inf_{p \geq p_c} \{Hp - ps(1/p) + d\}$*. Thus the right hand side of* (3.17) *always provides a upper bound for the exact rate of estimation of the signal. Remark that the functions of* $\mathcal{FP}(s(\bullet), r)$ *appear as the functions maximizing the frequency of points with a given smoothness* $H$*, for any* $H \in [s(0), d/p_c]$*. However, the violation of the Frisch Parisi conjecture might be severe. For instance Jaffard [22] constructs a function* $f$ *with a linear scaling function* $s_f$*, and which is* $\mathcal{C}^\infty$ *everywhere except at one point. In this situation* $d(H) = -\infty$ *for all* $H$*, and the right hand side of* (3.17) *is equal to* $+\infty$*.*

# 4 Examples and applications

## 4.1 Monofractal functions

Monofractal functions satisfy $s(1/p) = H_0$ for all $p > 0$. A canonical example is given by the sample paths of a fractional Brownian motion with Hurst parameter $H_0 \in (0, 1)$ in dimension $d = 1$, see for instance [35]. In this case, we find the minimax rate $\big(\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\big)^{2H_0/(2H_0+d)}$ for all loss functions $\pi \geq 1$. In particular, the sample path of a noisy 1-dimensional Brownian motion ($H_0 = 1/2$) is recovered with optimal rate $\big(\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\big)^{1/2}$, irrespectively of the loss function.

## 4.2 Besov balls

As already remarked in Section 2.1, single Besov spaces are related to a Besov domain of a certain admissible function $s_{s_0,p_0}(\bullet)$. Using monotonicity of the Besov norm with respect to $1/p$ and (3.3)–(3.4) the relation with Besov balls becomes exact: $B^{s_0}_{p_0,\infty}(r) = \mathcal{M}\big(s_{s_0,p_0}(\bullet), r\big)$. Theorem 1 in Jaffard [23] shows that among functions $f$ in $B^{s_0}_{p_0,\infty}(r)$ that saturate their Besov domain precisely for $s_{s_0,p_0}(\bullet)$, in other words that belong to the set:

$$\big\{f \in B^{s_0}_{p_0,\infty}(r),\ s_f(\bullet) = s_{s_0,p_0}(\bullet)\big\},$$

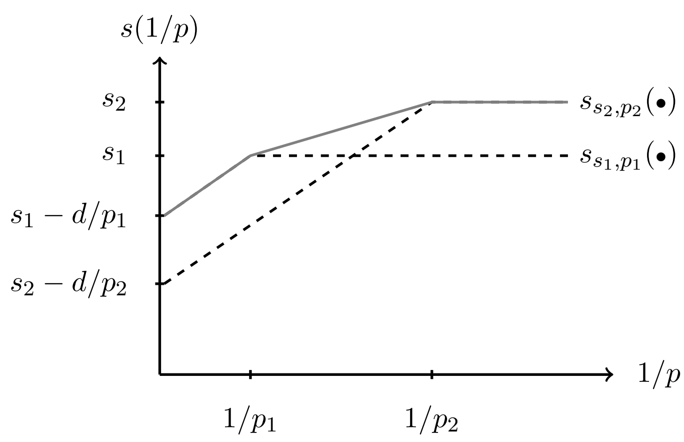
then the Frisch-Parisi holds for quasi-all functions. By Theorems 3.4 and 3.5, we retrieve the classical nonparametric reconstruction results over Besov balls as developed in the mid-1990 [9, 10, 11, 12] and the earlier results of the Russian school [21]. In light of our result, we can reinterpret the classical theory in terms of the multifractal approach: For Besov balls, the minimax rates of convergence are governed by extremal functions $f$ such that $s_f(\bullet) = s_{s_0,p_0}(\bullet)$ and which are generically multifractal. However, this multifractality is very particular, in the sense that $s_{s_0,p_0}(\bullet)$ is either constant or with maximal slope $d$, except in the viscinity of the so-called critical case where $p^\star = p_0$. This point separates the so-called dense and sparse cases (according to the classical terminology [11, 12, 26]).

### 4.3 Intersection of two Besov balls

As an exercice, we can compute the minimax rate of convergence up to logarithmic factors by Theorem 3.5 over the intersection of two Besov balls

$$\mathcal{C} := B^{s_1}_{p_1,\infty}(r) \cap B^{s_2}_{p_2,\infty}(r),\ \ p_1 > p_2,\ \ 0 < s_1 < s_2,$$

with $s_2 - d/p_2 + d/p_1 < s_1$, so that no Besov ball is included into the other. It is easily seen that $\mathcal{C} = \mathcal{M}\big(s(\bullet), r\big)$ where the graph of $s(\bullet)$ is the concave envelope of the graph of $s_{s_1,p_1}(\bullet)$ and $s_{s_2,p_2}(\bullet)$.




In the region for which loss functions $\pi$ are such that $1/p^\star \geq 1/p_2$ the minimax rates of convergence are governed by the dense regime of the space $B^{s_2}_{p_2,\infty}$. Likewise, in the region for which $1/p^\star \geq 1/p_2$, the sparse $B^{s_1}_{p_1,\infty}$ regime dominates. A new intermediate regime appears for $1/p_1 \leq 1/p^\star \leq 1/p_2$:

**Corollary 4.1.** *In the setting of Theorem 3.4 and 3.5, the minimax rate of convergence for the class $\mathcal{C}$ is given (up to logarithmic fators) by $\varepsilon^{2v(s_1,s_2,p_1,p_2,\pi)}$, where*

$$v(s_1,s_2,p_1,p_2,\pi) = \begin{cases} s_2/(2s_2+d) & \text{if} \quad d\pi \leq p_2(d+2s_2) \\ \dfrac{s_1 + d(1/\pi - 1/p_1)}{2(s_1 - d/p_1) + d} & \text{if} \quad d\pi \geq p_1(d+2s_1) \end{cases}$$

*in the classical regimes, and*

$$v(s_1,s_2,p_1,p_2,\pi) = \frac{(s_2-s_1)/\pi + s_1/p_2 - s_2/p_1}{2(s_1/p_2 - s_2/p_1) + d(1/p_2 - 1/p_1)}$$

*in the non-classical regime $p_2(d+2s_2) \leq d\pi \leq p_1(d+2s_1)$.*

### 4.4 A multifractal model for chirps and Dopplers

In [20], Hall, Kerkyacharian and Picard (abbreviated by HKP in the following) develop block threshold methods in the case $\pi = 2$ for wavelet estimators which are adaptive to many variations of signal abberations including those of chirp and Doppler type, which are of the form $x \rightsquigarrow |x-x_0|^\beta \cos(|x-x_0|^{-\alpha})$ for $\alpha, \beta \geq 0$. In dimension 1, HKP introduce the class $\mathcal{H}$ that can be described as follows: $g \in \mathcal{H}$ if for any $j \geq 0$, there exists a set of integers $S_j$ with $\operatorname{Card} S_j \lesssim 2^{j\gamma}$ such that:

- For each $k \in S_j$ there exist constants $a_0 = g(k2^{-j}), a_1, \ldots, a_{N-1}$ such that

  $$\Big| g(x) - \sum_{\ell=0}^{N-1} a_\ell (x - k2^{-j})^\ell \Big| \lesssim 2^{-js_1} \quad \text{for all} \quad x \in [k2^{-j}, (k+v)2^{-j}],$$

  where $v > 0$ is a given constant and

- For each $k \notin S_j$ there exist constants $a_0, a_1, \ldots, a_{N-1}$ such that

  $$\Big| g(x) - \sum_{\ell=0}^{N-1} a_\ell (x - k2^{-j})^\ell \Big| \lesssim 2^{-js_2} \quad \text{for all} \quad x \in [k2^{-j}, (k+v)2^{-j}].$$

The class is parametrized by $0 \leqslant \gamma \leqslant 1$ and $s_1 < s_2$. Proposition 3.2. in [20] shows that if the analyzing wavelet $\psi$ has compact support included in $[0, v]$, then

$$|d_\lambda| \lesssim 2^{-|\lambda|(s_1+1/2)} 1_{\{k \in S_j\}} + 2^{-|\lambda|(s_2+1/2)} 1_{\{k \notin S_j\}}. \tag{4.1}$$

In particular, if $s_1 > 0$, the class $\mathcal{H}$ is embedded into continuous functions, a restriction that HKP do not have, but which is important if we use the interpretation in terms of multifractal analysis. The characterization (4.1) enables to show easily that if $g \in \mathcal{H}$, then $s_g(\bullet) \geq s_{\mathcal{H}}(\bullet)$, with

$$s_{\mathcal{H}}(1/p) := \begin{cases} s_1 + (1-\gamma)/p & \text{for} \quad 1/p < (s_2 - s_1)/(1-\gamma) \\ \\ s_2 & \text{for} \quad 1/p \geq (s_2 - s_1)/(1-\gamma). \end{cases}$$

This reveals the non-trivial Besov domain $\mathcal{M}\big(s_{\mathcal{H}}(\bullet)\big) \supseteq \mathcal{H}$ as soon as $\gamma \neq 1$. In particular, HKP show in their Theorem 4.1. that the minimax rate exponent $\varepsilon^{2s_2/(1+2s_2)}$ is achievable for $\pi = 2$ if the following condition[4] holds:

$$0 \leq \gamma \leq \frac{2s_1 + 1}{2s_2 + 1}.$$

In our formalism, this corresponds exactly to the critical case when the line $1/p \rightsquigarrow -d/2 + d\pi/(2p)$ (with $d = 1$ and $\pi = 2$) intersects $s_{\mathcal{H}}(\bullet)$ for $1/p \geqslant (s_2 - s_1)/(1-\gamma)$ and our approach shows that the result of HKP is sharp. Beyond this critical point, our Theorems 3.4 and 3.5 completes their result and reveals the following non-classical minimax rate of convergence:

$$\varepsilon^{\frac{2s_1 + (1-\gamma)}{2s_1+1}} \quad \text{if} \quad \gamma > \frac{2s_1 + 1}{2s_2 + 1},$$

within a logarithmic factor[5].

## 4.5 Lacunary wavelet series

An example of multifractal signal $f$ on $\mathcal{D} = [0,1]^d$ is provided by the sample path of a lacunary wavelet series as defined in Jaffard (2000b) [24]. This random process $f$ is defined by its wavelet coefficients as follows: let $\alpha \in (0, d)$ and for each level $j$, choose randomly $\lfloor c(\mathcal{D}) 2^{j(d-\alpha)} \rfloor$ locations among the $\mathrm{Card}(\Lambda_j) \sim c(\mathcal{D}2^{jd})$ locations corresponding to this level. The

[4]The result of HKP is slightly more general, since it allows $2^{-j\gamma} \mathrm{Card}\, S_j$ to grow as $\varepsilon \to 0$ at a certain rate, a situation we discard here for simplicity.

[5]whereas HKP results are sharp up to logarithmic terms.

chosen coefficients are set to the value $2^{-j(\beta+d/2)}$ for some $\beta > 0$, all the other coefficients are set to 0. It is shown in [24] that $s_f(1/p) = \alpha/p + \beta$ and that almost surely the Frisch Parisi conjecture holds with a singularity spectrum given by $d(H) = \frac{H(d-\alpha)}{\beta}$ for $H \in [\beta, \frac{d\beta}{d-\alpha}]$.

Applying Theorem 3.4, an upperbound for the rate of estimation of the signal, with $\mathbf{L}^\pi$ loss, is (up to a logarithmic factor)

$$\varepsilon^{\frac{2\beta+2\alpha/\pi}{2\beta+d}}$$

as soon as $2\alpha/\pi < d$. This rate is increasing with the sparsity of the wavelet series, and the restriction $2\alpha/\pi < d$ corresponds to the restriction by the parametric rate $\varepsilon$. Remark that in this example, the infimum in the formula (3.17) is attained at $H = s(0) = \beta$, which is the minimal degree of smoothness for a singularity appearing in the signal.

## 4.6 Cascade processes

Multiplicative cascade processes were introduced in the initial work by Mandelbrot [28] and mathematical properties were studied further in Kahane and Peyriere [25]. The objective was to provide models which describe the statistical behavior of turbulent flows. More recently, they were applied for modeling many signals which exhibit high intermittency such as the internet traffic, financial data or DNA sequences.

The law of the cascade process is fully characterized by some non negative random variable $W$, which expectation is equal to one, and usually referred to as the 'cascade generator'. This variable determines the evolution, through change of scales, of the cascade process $(X_t)_{t\in[0,1]}$, as can be seen in the relation:

$$X_{t/2} - X_0 \overset{law}{=} (X_t - X_0)W/2, \qquad \text{for all dyadic } t.$$

A large literature is devoted to the study of the properties of the sample paths of multiplicative cascades (see references in [4]). For instance, Molchan [31] proves that the sample paths of these processes are almost surely multifractal and computes the spectrum of singularities as a function of $W$. He shows that the Frisch Parisi conjecture holds and essentially relates the scaling exponent $s_X(\bullet)$ of the typical sample path to the generator $W$:

$$s_X(1/p) = \begin{cases} 1 - \frac{1}{p}\log_2(E(W^p)) & \text{for } p \leqslant p_W \\ \frac{1}{p} + c_W & \text{for } p \geqslant p_W \end{cases}$$

where $p_W$ and $c_W$ are the two unique values that make the function $s_X$ of class $\mathcal{C}^1$ ($p_W$ is thus determined by the condition that $s'_X(1/p_W) = 1$ in both expressions above, and the value of $c_W$ follows by the continuity of $s_X$). Since the choice of the generator $W$ is almost arbitrary, we see that cascade processes provide examples of stochastic signals with a large variety of scaling exponents $s(\bullet)$.

# 5 Proofs

## 5.1 Proof of Theorem 3.4

**Preliminaries from [26].** Weak $\ell_{q,\infty}(\pi)$ sequence spaces are defined by means of the atomic measure $\mu_\pi$ defined over multiindices $\lambda$ by setting

$$\mu_\pi(\{\lambda\}) := \|\psi_\lambda\|_{L^\pi}^\pi, \quad \text{for } 1 \leq \pi < \infty.$$

For $0 < q < \pi$, a function $f = \sum_\lambda f_\lambda \psi_\lambda$ belongs to $\ell_{q,\infty}(\pi)$ if

$$\|f\|_{\ell_{q,\infty}(\pi)}^q := \sup_{t>0} t^q \mu_\pi\big(\lambda,\ |f_\lambda| \geq t\big) < +\infty.$$

The spaces $\ell_{q,\infty}(\pi)$ are linked to the choice of the basis $(\psi_\lambda)_\lambda$ and related to classical Besov spaces $B^s_{p,\infty}$ by the following embedding properties:

**Proposition 5.1.** *(Theorem 6.2 in [26]). Let $1 \leq \pi < \infty$ and $s \geq 0$. Define $p_s := \frac{d\pi}{2s+d}$.*

- *If $p > p_s$, then $B^s_{p,\infty} \subset \ell_{p_s,\infty}(\pi)$.*
- *If $p = p_s$ and $f \in B^s_{p,\infty} \cap B^\delta_{\pi,\infty}$ for some $\delta > 0$, then for $t < 1/2$:*

$$\mu_\pi\big(\lambda,\ |f_\lambda| \geq t\big) \lesssim t^{-p} \log\left(\frac{1}{t}\right).$$

- *If $p < p_s$ and $\frac{2d}{2s+d} < p < p_s$, then $B^s_{p,\infty} \subset \ell_{q_s,\infty}(\pi)$, with $q_s = \frac{d(p/2-1)}{s+d(1/2-1/p)}$.*

The following proposition was obtained in [26], but in order to prove Theorem 3.4, we will actually need a slight extension of it. Therefore and in order to be self-contained, we give a condensed proof of this proposition.

**Proposition 5.2.** *(Theorem 5.1 in [26]). In the setting of Theorem 3.4, we have*

$$\mathbb{E}\left[\|\widehat{f}_{\varepsilon,\pi} - f\|_{L^\pi}^\pi\right] \lesssim \left(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\right)^{\pi-q} \|f\|_{\ell_{q,\infty}(\pi)}^q + \|\sum_{|\lambda|>J_\varepsilon} f_\lambda\psi_\lambda\|_{L^\pi}^\pi,$$

*for all* $0 < q < \pi$ *such that the right-hand side is meaningful, and where* $\lesssim$ *means up to a constant depending on* $\pi$ *and* $\|f\|_{L^\pi}$ *only.*

*Proof.* We have

$$\mathbb{E}\left[\|\widehat{f}_{\varepsilon,\pi} - f\|_{L^\pi}^\pi\right] \lesssim \mathbb{E}\left[\|\sum_{|\lambda|\le J_\varepsilon} (T_{\varepsilon,\pi}(\hat{f}_\lambda) - f_\lambda)\psi_\lambda\|_{L^\pi}^\pi\right] + \|\sum_{|\lambda|>J_\varepsilon} f_\lambda\psi_\lambda\|_{L^\pi}^\pi,$$

and in view of the result, only an inspection of the first term in the right-hand side is needed. Indeed, this first term is, up to a constant, less than $I + II$, with

$$I := \mathbb{E}\left[\|\sum_{|\lambda|\le J_\varepsilon} (\hat{f}_\lambda - f_\lambda)1_{\{|\hat{f}_\lambda|\ge\kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\}}\psi_\lambda\|_{L^\pi}^\pi\right],$$

and

$$II := \mathbb{E}\left[\|\sum_{|\lambda|\le J_\varepsilon} 1_{\{|\hat{f}_\lambda|\le\kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\}} f_\lambda\psi_\lambda\|_{L^\pi}^\pi\right].$$

A fairly classical concentration argument based on (3.12) enables then to ignore the random part of $\hat{f}_\lambda$ in the indicator in I and II, up to modifying the threshold level by a factor 1/2 (see [12], [26]). It follows that $I$ and $II$ can be replaced by

$$III := \mathbb{E}\left[\|\sum_{|\lambda|\le J_\varepsilon} (\widehat{f}_\lambda - f_\lambda)1_{\{2|f_\lambda|\ge\kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\}}\psi_\lambda\|_{L^\pi}^\pi\right]$$

and

$$IV := \|\sum_{|\lambda|\le J_\varepsilon} 1_{\{2|f_\lambda|\le\kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\}} f_\lambda\psi_\lambda\|_{L^\pi}^\pi$$

respectively, without affecting the rates of convergence, inflating the error by a multiplicative factor depending on $\pi$ and $\|f\|_{L^\pi}$ only.

**Step 1: The case $\pi \le 2$.** For any sequence $(u_\lambda)_{\lambda\in\Lambda}$, we have in that case

$$\|\sum_{\lambda\in\Lambda} u_\lambda\psi_\lambda\|_{L^\pi}^\pi \lesssim \sum_{\lambda\in\Lambda} |u_\lambda|^\pi \|\psi_\lambda\|_{L^\pi}^\pi \tag{5.1}$$

as follows from (3.8) and the comparison between $\ell^\pi$-norms in the case $\pi > 1$, and from triangle inequality if $\pi = 1$. Applying successively (5.1) and (3.11), we have

$$
\begin{aligned}
III &\lesssim \sum_{|\lambda|\leq J_\varepsilon} \mathbb{E}\left[|\widehat{f}_\lambda - f_\lambda|^\pi\right] 1_{\left\{2|f_\lambda|\geq \kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right\}} \|\psi_\lambda\|_{L^\pi}^\pi \\
&\lesssim \varepsilon^\pi \sum_{|\lambda|\leq J_\varepsilon} 1_{\left\{2|f_\lambda|\geq \kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right\}} \|\psi_\lambda\|_{L^\pi}^\pi \qquad (5.2) \\
&\lesssim \left(\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right)^{\pi-q} \|f\|_{\ell_{q,\infty}(\pi)}^q
\end{aligned}
$$

by definition of the weak-$\ell_{q,\infty}(\pi)$ space. By (5.1) again,

$$
IV \lesssim \sum_{|\lambda|\leq J_\varepsilon} 1_{\{2|f_\lambda|\leq \kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\}} |f_\lambda|^\pi \|\psi_\lambda\|_{L^\pi}^\pi. \tag{5.3}
$$

Next, we use the fact that for all $q < \pi$,

$$
\sup_{t>0} t^{q-\pi} \sum_\lambda |f_\lambda|^\pi 1_{|f_\lambda|\leq t} \|\psi\|_{L^\pi}^\pi \lesssim \|f\|_{\ell_{q,\infty}(\pi)}^q, \tag{5.4}
$$

up to a constant depending on $\pi$. The characterization (5.4) of $\ell_{q,\infty}(\pi)$ spaces relies on simple calculation (see [26] or section 3 in [6]). So the right-hand side of (5.3) is further bounded by a constant times $\left(\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right)^{\pi-q}\|f\|_{\ell_{q,\infty}(\pi)}^q$ and Proposition 5.2 follows.

**Step 2: The case $\pi \geq 2$.** Using successively property (3.8) and Minkowski's generalized inequality, we have

$$
\begin{aligned}
III &\lesssim \mathbb{E}\Big[\int_{\mathcal{D}} \Big(\sum_{|\lambda|\leq J_\varepsilon} 1_{\left\{2|f_\lambda|\geq \kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right\}} |\widehat{f}_\lambda - f_\lambda|^2 |\psi_\lambda(x)|^2\Big)^{\pi/2} dx\Big] \\
&\lesssim \int_{\mathcal{D}} \Big(\sum_{|\lambda|\leq J_\varepsilon} 1_{\left\{2|f_\lambda|\geq \kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right\}} \left(\mathbb{E}[|\widehat{f}_\lambda - f_\lambda|^\pi]\right)^{2/\pi} |\psi_\lambda(x)|^2\Big)^{\pi/2} dx
\end{aligned}
$$

By (3.11), this last quantity is less than

$$\varepsilon^\pi \int_{\mathcal{D}} \Big( \sum_{|\lambda|\le J_\varepsilon} 1_{\left\{2|f_\lambda|\ge\kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right\}} |\psi_\lambda(x)|^2\Big)^{\pi/2} dx$$

$$\lesssim \varepsilon^\pi \sum_{|\lambda|\le J_\varepsilon} 1_{\left\{2|f_\lambda|\ge\kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right\}} \|\psi_\lambda\|_{L^\pi}^\pi, \tag{5.5}$$

where we have used (3.7) for the last line. Thus $III$ has the right order by the definition of weak $\ell_{q,\infty}(\pi)$ spaces as before. Write $\rho_\varepsilon := 2\kappa(\pi)\varepsilon\sqrt{\log\frac{1}{\varepsilon}}$ for notational simplicity. We have

$$\begin{aligned} IV &\lesssim \Big\| \sum_{k\ge1} \sum_{|\lambda|\le J_\varepsilon} 1_{\{2^{-k}\rho_\varepsilon\le|f_\lambda|\le 2^{-k+1}\rho_\varepsilon\}} f_\lambda\,\psi_\lambda \Big\|_{L^\pi}^\pi \\ &\lesssim \Big(\sum_{k\ge1} \Big\| \sum_{|\lambda|\le J_\varepsilon} 1_{\{2^{-k}\rho_\varepsilon\le|f_\lambda|\le 2^{-k+1}\rho_\varepsilon\}} f_\lambda\,\psi_\lambda \Big\|_{L^\pi}\Big)^\pi \end{aligned}$$

by the triangle inequality. Next, using (3.8) and $|f_\lambda| 1_{\{2^{-k}\rho_\varepsilon\le|f_\lambda|\le 2^{-k+1}\rho_\varepsilon\}} \le 2^{-k+1}\rho_\varepsilon 1_{\{|f_\lambda|\ge 2^{-k}\rho_\varepsilon\}}$, the above quantity is less than

$$\rho_\varepsilon^\pi \Big( \sum_{k\ge1} 2^{-k+1} \Big\| \Big( \sum_{|\lambda|\le J_\varepsilon} 1_{\{|f_\lambda|\ge 2^{-k}\rho_\varepsilon\}} |\psi_\lambda|^2\Big)^{1/2} \Big\|_{L^\pi}\Big)^\pi.$$

By applying first (3.7) and then the definition of weak $\ell_{q,\infty}(\pi)$ spaces, this quantity is less than

$$\rho_\varepsilon^\pi \Big( \sum_{k\ge1} 2^{-k+1} \Big( \sum_{|\lambda|\le J_\varepsilon} 1_{\{|f_\lambda|\ge 2^{-k}\rho_\varepsilon\}} \|\psi_\lambda\|_{L^\pi}^\pi\Big)^{1/\pi}\Big)^\pi \tag{5.6}$$

$$\lesssim \Big(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\Big)^{\pi-q} \Big( \sum_{k\ge1} 2^{-k} 2^{(k+1)q/\pi} \|f\|_{\ell_{q,\infty}(\pi)}^{q/\pi}\Big)^\pi.$$

Since $\pi > q$, the geometric series is convergent which ends the proof. $\square$

We are now ready to prove Theorem 3.4 in four steps.

**Step 1: Definiteness of $p^\star$.** It suffices to check that the function

$$t \rightsquigarrow \varphi(t) := s(t) - \tfrac{d}{2}(t\pi - 1)$$

hits 0. We have $\varphi(0) > 0$ since $s(0) > 0$ and $\varphi'(t) = s'(t) - d\pi/2$. Moreover, $s(\bullet)$ is concave so having $\varphi'(t) < 0$ for some $t > 0$ is sufficient, but that follows from the bound $s'(\infty) < d\pi/2$.

**Step 2: Linear term.** As soon as $f \in B^{\delta}_{\pi,\infty}$ for some $\delta > 0$, we have

$$\Big\| \sum_{|\lambda|>J_\varepsilon} f_\lambda \psi_\lambda \Big\|^{\pi}_{L^\pi} \lesssim \|f\|^{\pi}_{B^{\delta}_{\pi,\infty}} 2^{-J_\varepsilon \delta\pi} \lesssim \left(\varepsilon^2 \log \tfrac{1}{\varepsilon}\right)^{\delta\pi/d}.$$

We now prove that we can take $\delta = ds(1/p^\star)/(2s(1/p^\star)+d)$, which gives the desired order. We claim that

$$s(1/\pi) \geq \frac{ds(1/p^\star)}{2s(1/p^\star)+d}, \tag{5.7}$$

from which the result simply follows since $f \in B^{s(1/\pi)}_{\pi,\infty}$ and by the embedding $B^{s}_{\pi,\infty} \subset B^{s'}_{\pi,\infty}$ if $s \geq s'$. By definition of $p^\star$, we have $p^\star = \pi d/(2s(1/p^\star)+d)$ so (5.7) is equivalent to

$$\pi s(1/\pi) \geq \frac{\pi d s(1/p^\star)}{2s(1/p^\star)+d} = p^\star s(1/p^\star).$$

It is enough to prove that $t \rightsquigarrow \tilde{\varphi}(t) := s(t)/t$ is decreasing between $1/\pi$ and $1/p^\star$ since, by definition of $p^\star$, we always have

$$\frac{1}{p^\star} = \frac{2s(1/p^\star)}{\pi d} + \frac{1}{\pi} > \frac{1}{\pi}.$$

Now, $\tilde{\varphi}'(t) = \big(s'(t)t - s(t)\big)/t^2 \leq -s(0)/t^2 \leq 0$ since $s(\bullet)$ is concave and $s(0) > 0$.

**Step 3: Nonlinear terms, large $p$.** Proposition 5.1 provides two regions of embedding, separated by the critical case $p = p_s$. Let us first consider the set $(p^\star, \infty)$, of large $p$'s, for which the following condition holds

$$p > \frac{d\pi}{2s(1/p)+d}.$$

By Proposition 5.1 case 1, we have the embedding

$$B^{s(1/p)}_{p,\infty} \subset \ell_{d\pi/(2s(1/p)+d),\infty}(\pi),$$

therefore the nonlinear term $\left(\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right)^{\pi-q}\|f\|^q_{\ell_{q,\infty}(\pi)}$ in Proposition 5.2 is for the choice $q = d\pi/(2s(1/p)+d)$ of order

$$\left(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\right)^{\pi-d\pi/(2s(1/p)+d)} = \left(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\right)^{\frac{2s(1/p)\pi}{2s(1/p)+d}}.$$

If $s'(1/p_0) = 0$ for some $1/p_0 < 1/p^\star$, $s(\bullet)$ is constant for $1/p \geq 1/p_0$ by concavity hence $s(1/p_0) = s(1/p^\star)$ and we obtain Theorem 3.4 in that case.

**Step 4: Nonlinear terms, critical $p$.** Let us then turn to the critical case

$$p^\star = \frac{d\pi}{2s(1/p^\star)+d}.$$

For this constraint, the decomposition of Proposition 5.2 is of no use since no embedding of $B^{s(1/p)}_{p,\infty}$ in any space $\ell_{q,\infty}(\pi)$ is valid, so we need to refine Proposition 5.2. By Proposition 5.1 case 2, we have

$$\forall t < 1/2, \quad \mu_\pi\big(\lambda,\ |f_\lambda| \geq t\big) \lesssim t^{-p^\star}\log\left(\frac{1}{t}\right) \tag{5.8}$$

if, in addition, $f \in B^\delta_{\pi,\infty}$ for some $\delta > 0$, a case we always have with $\delta = s(1/\pi)$ for instance. This enables us to revisit the terms $III$ and $IV$ in Proposition 5.2. Inspecting (5.2), (5.5) in the proof, we readily have, for all $\pi \geq 1$

$$\begin{aligned} III &\lesssim \varepsilon^\pi \mu_\pi\big(|\lambda| \leq J_\varepsilon,\ |f_\lambda| \geq \tfrac{\kappa(\pi)}{2}\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\big) \\ &\lesssim \varepsilon^\pi\big(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\big)^{-p^\star}\log\frac{1}{\varepsilon} \\ &\lesssim \big(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\big)^{2s(1/p^\star)\pi/(2s(1/p^\star)+d)}\log\tfrac{1}{\varepsilon} \end{aligned}$$

where we have successively used (5.8) and the definition of $p^\star$. We now turn to the more delicate term $IV$. Define, for notational simplicity $\rho_\varepsilon := \frac{\kappa(\pi)}{2}\varepsilon\sqrt{\log\frac{1}{\varepsilon}}$. Inspecting the proof of Proposition 5.2, equation (5.6) yields in the case $\pi \geqslant 2$,

$$IV \lesssim \rho_\varepsilon^\pi\Big(\sum_{k\geqslant 1} 2^{-k+1}\mu_\pi\big(|\lambda| \leq J_\varepsilon,\ |f_\lambda| \geq 2^{-k}\rho_\varepsilon\big)^{1/\pi}\Big)^\pi.$$

Using again (5.8), we deduce that $IV$ is of magnitude less than $\left(\rho_\varepsilon\right)^{\pi-p^\star}\log\frac{1}{\rho_\varepsilon}$ and thus Theorem 3.4 is proved in the case $\pi \geqslant 2$. For $\pi \in [1,2]$, the term $IV$ can be bounded quite similarly by using (5.3), we omit the details here. Theorem 3.4 follows.

**Remark 5.3.** *It is instructive to inspect the behaviour of the threshold algorithm beyond the critical case, i.e. for $p$ satisfying*

$$p < \frac{d\pi}{2s(1/p)+d}.$$

*Proposition 5.1 case 3 tells us that the embedding*

$$B^{s(1/p)}_{p,\infty} \subset \ell_{d(\pi/2-1)/(s(1/p)+d(1/2-1/p)),\infty}(\pi)$$

*holds, provided further that $p > 2d/(2s(1/p)+d)$ which implies in particular the restriction $\pi > 2$. Moreover, no other inclusion of any Besov space into a weak $\ell_{q,\infty}(\pi)$ exists otherwise. In that setting, it is readily checked that the threshold algorithm attains in this region the rate $\left(\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right)^{q^\star\pi}$, with*

$$q^\star = \frac{s(1/p)+d\left(\frac{1}{\pi}-\frac{1}{p}\right)}{s(1/p)+d\left(\frac{1}{2}-\frac{1}{p}\right)},$$

*but a closer inspection of the properties of $s(\bullet)$, namely the fact that $s'(1/p) \leq d$ shows that this rate is suboptimal.*

## 5.2 Proof of Theorem 3.5

The proof is divided into three steps. We first focus on the case of a linear admissible function $s(\bullet)$ and solve a Bayesian problem instead of the initial minimax problem. In a second step, we compare the Bayes risk with the minimax risk. Finally, we extend the result to an arbitrary Besov domain $s(\bullet)$. The extremal case $s'(1/p^\star) = d$ is more delicate and requires a separate proof.

**Step 1: A Bayes risk.** A first delicate issue is to construct a prior on $L^2$ which concentrates on functions with exact scaling function $s(\bullet)$. The construction of such functions is a fairly complex problem solved in [23]. However, there is a simple expression if the scaling function is linear. Let

$$\tilde{s}(1/p) := \beta + \alpha/p, \ \ 0 < 1/p < +\infty$$

which is admissible for $\alpha \in [0, d]$ and $\beta \geqslant 0$. The following lemma provides a simple condition that ensures that a function is in $\mathcal{M}\big(\tilde{s}(\bullet)\big)$.

**Lemma 5.4.** *Let $g = \sum_\lambda c_\lambda \psi_\lambda = \sum_{j \geqslant 0} \sum_{|\lambda|=j} c_\lambda \psi_\lambda$ be such that:*

$$|c_\lambda| \leqslant c(\mathcal{D})^{-\beta} 2^{-j(\beta+d/2)} \textit{for all } \lambda \textit{ with } |\lambda| = j,$$

*and the number of non-zero coefficients on each level satisfies:*

$$\textit{Card}\{\lambda,\, |\lambda| = j \textit{ and } c_\lambda \neq 0\} \leqslant c(\mathcal{D})^{d-\alpha} 2^{j(d-\alpha)}.$$

*Then, $\|g\|_{B^{\tilde{s}(1/p)}_{p,\infty}} \leqslant 1$ for all $p > 0$.*

*Proof.* We have $\Big(\sum_{|\lambda|=j} |c_\lambda|^p\Big)^{1/p} \leqslant c(\mathcal{D})^{-\beta} 2^{j(d-\alpha)/p} c(\mathcal{D})^{(d-\alpha)/p} 2^{-j(\beta+d/2)}$ hence the result by (3.2). □

Note that a function that saturates the conditions of Lemma 5.4 has scaling function $\tilde{s}(\bullet)$. We now choose a level $j$ and define a prior $\mu_j(\mathrm{d}f)$ on $L^2$ by picking at random the wavelet coefficients according to the following distribution: if $\lambda \neq j$ we set $\langle f, \psi_\lambda \rangle = 0$ and if $\lambda = j$, the coefficients $\langle f, \psi_\lambda \rangle = c_\lambda$ are independent Bernoulli variables,

$$\langle f, \psi_\lambda \rangle = c_\lambda = \begin{cases} r c(\mathcal{D})^{-\beta} 2^{-j(\beta+d/2)} & \text{with probability} \quad q \\ 0 & \text{with probability} \quad 1-q \end{cases},$$

with $q = c(\mathcal{D})^{-\alpha} 2^{-j\alpha-1}$. We define the associated Bayes $L^\pi$-error for any estimator $\widehat{f}$:

$$\mathcal{E}_{B,j,\pi}(\widehat{f}) : \left( \int_{L^2} \mathbb{E}_f \left[ \|\widehat{f} - f\|^\pi_{L^\pi} \right] \mu_j(\mathrm{d}f) \right)^{1/\pi},$$

where $\mathbb{P}_f$ denotes the law of $Y_\varepsilon$ with parameter[6] $f$. The following proposition gives a lower bound for this Bayes error. Let us stress that the result depends on $\alpha$ which characterizes in some way the sparsity of the prior. According to the usual terminology, for $\alpha = 0$ the prior is dense, whereas it is sparse in the other cases.

[6]We assume that all the probability measures $\mathbb{P}_f$ are defined simultaneously on the canonical space $L^2$. Such a construction is always possible.

**Proposition 5.5.** • *If* $\alpha \neq 0$, *choose* $M$ *with* $0 < M < r^{-2}c(\mathcal{D})^{2\beta}2\log 2$ *and let* $j = j(\varepsilon)$ *satisfy*

$$M\alpha(j-1)2^{(j-1)(d+2\beta)} \leqslant \varepsilon^{-2} < M\alpha j 2^{j(d+2\beta)}. \tag{5.9}$$

*Then, we have:*

$$\inf_{\widehat{f}} \mathcal{E}_{B,j(\varepsilon),\pi}\big(\widehat{f}\big)^{\pi} \gtrsim \left(\varepsilon\sqrt{\log\tfrac{1}{\varepsilon}}\right)^{2\frac{\alpha+\pi\beta}{d+2\beta}}.$$

- *If* $\alpha = 0$, *let* $j = j(\varepsilon)$ *satisfy*

$$2^{(j-1)(d+2\beta)} \leqslant \varepsilon^{-2} < 2^{j(d+2\beta)}. \tag{5.10}$$

*We have:*

$$\inf_{\widehat{f}} \mathcal{E}_{B,j(\varepsilon),\pi}\big(\widehat{f}\big)^{\pi} \gtrsim \varepsilon^{2\frac{\pi\beta}{d+2\beta}}.$$

*In both cases,* $\gtrsim$ *means up to constant depending on* $\pi$, $r$, $\alpha$, $\beta$.

*Proof.* If $\widehat{f}$ is an arbitrary estimator, denote by $\widehat{c}_\lambda = \langle \widehat{f}, \psi_\lambda\rangle$ the associated estimated wavelet coefficients. Using that the projection on the space generated by $(\psi_\lambda)_{|\lambda|=j}$ is continuous in $L^\pi$ and Bernstein inequality, we have for all function $f$: $\|\widehat{f} - f\|_{L^\pi}^\pi \gtrsim 2^{jd(\pi/2-1)} \sum_{|\lambda|=j} |\widehat{c}_\lambda - \langle f, \psi_\lambda\rangle|^\pi$. We derive:

$$\mathcal{E}_{B,j,\pi}\big(\widehat{f}\big)^{\pi} \gtrsim 2^{jd(\pi/2-1)} \int_{L^2} \sum_{|\lambda|=j} \mathbb{E}_f\left[|\widehat{c}_\lambda - c_\lambda|^\pi\right]\mu_j(\mathrm{d}f). \tag{5.11}$$

Let us consider temporarily a multi-index $\lambda_0 \in \Lambda$ with $|\lambda_0| = j$. The estimated coefficient $\widehat{c}_{\lambda_0}$ is some function of $(Y_\varepsilon)$ and thus can be seen as a function of the collection of all the observable coefficients $Y_\varepsilon(\psi_\lambda) = c_\lambda + \varepsilon\xi(\psi_\lambda)$ for $\lambda \in \Lambda$. Since the variables $\xi(\psi_\lambda)$ for $\lambda \in \Lambda$ are i.i.d. and by our choice of prior, $c_{\lambda_0}$ and $\xi(\psi_{\lambda_0})$ are independent of $\big(c_\lambda, \xi(\psi_\lambda)\big)_{\lambda\neq\lambda_0}$. Hence, conditioning with respect to $\big(c_\lambda, \xi(\psi_\lambda)\big)_{\lambda\neq\lambda_0}$, for any choice of $\widehat{c}_{\lambda_0}$, the minimum of

$$\int_{L^2} \mathbb{E}_f\left[|\widehat{c}_{\lambda_0} - c_{\lambda_0}|^\pi\right]\mu_j(\mathrm{d}f),$$

is obtained when $\widehat{c}_{\lambda_0}$ is a function of $Y_\varepsilon(\psi_{\lambda_0}) = c_{\lambda_0} + \varepsilon\xi(\psi_{\lambda_0})$ only. We derive from (5.11):

$$\mathcal{E}_{B,j,\pi}\big(\widehat{f}\big)^{\pi} \geqslant 2^{jd(\pi/2-1)}2^{jd}\rho_{j,\varepsilon,\pi} = 2^{jd\pi/2}\rho_{j,\varepsilon,\pi}, \tag{5.12}$$

where $\rho_{j,\varepsilon,\pi}$ is the one dimensional Bayes risk:

$$\rho_{j,\varepsilon,\pi} = \inf_{g:\mathbb{R}\to\mathbb{R}} \mathbb{E}\left[|g(c_j + \varepsilon Z) - c_j|^\pi\right]$$

where $Z$ is a standard Gaussian variable and $c_j$ is an independent Bernoulli random variable distributed taking values $K = rc(\mathcal{D})^{-\beta}2^{-j(\beta+d/2)}$ and 0 with probability $q$ and $1-q$ respectively. Similar 1-dimensional problems are studied in [8]. In lemma 6.1 of the Appendix, we prove that under condition (5.9) if $\alpha \neq 0$ (or condition (5.10) if $\alpha = 0$), we have $\rho_{j(\varepsilon),\varepsilon,\pi} \gtrsim qK^\pi$. This implies that

$$\begin{aligned}\mathcal{E}_{B,j,\pi}\big(\widehat{f}\big)^\pi &\gtrsim 2^{-j(\varepsilon)d\pi/2} qK^\pi \\ &\gtrsim 2^{-j(\varepsilon)d\pi/2} 2^{-j(\varepsilon)\alpha} 2^{-j(\varepsilon)\pi(\beta+d/2)} \gtrsim 2^{-j(\varepsilon)(\alpha+\pi\beta)}.\end{aligned}$$

Then using again condition (5.9) (or (5.10) if $\alpha = 0$) we obtain the proposition. $\square$

**Step 2: The minimax risk.** The next result shows that if $\alpha \neq d$ then the minimax risk over $\mathcal{M}(\tilde{s}(\bullet), r)$ and the Bayes risk with prior $\mu_j$ are comparable.

**Proposition 5.6.** *If $\alpha \neq d$, then*

$$\inf_{\widehat{f}} \sup_{f\in\mathcal{M}(\tilde{s}(\bullet),r)} \mathbb{E}_f\left[\|\widehat{f} - f\|^\pi_{L^\pi} \gtrsim \inf_{\widehat{f}} \mathcal{E}_{B,j(\varepsilon),\pi}\big(\widehat{f}\big)^\pi,\right.$$

*where $j(\varepsilon)$ is defined in Proposition 5.5.*

*Proof.* For $j \geqslant 0$ and $f$ in $L^2(\mathcal{D})$, let $N_j(f) := \mathrm{Card}\{\lambda, \langle f, \psi_\lambda\rangle \neq 0,\ |\lambda| = j\}$. By our choice of prior and Lemma 5.4 we have

$$\mu_j\left(f \in \mathcal{M}(\tilde{s}(\bullet), r)\right) \geqslant \mu_j\left(N_j \leqslant c(\mathcal{D})^{d-\alpha}2^{j(d-\alpha)}\right).$$

Under $\mu_j$, $N_j$ is a Binomial random variables with parameters $\mathrm{Card}\{\Lambda_j\}$ and $c(\mathcal{D})^{-\alpha}2^{-j\alpha-1}$. We deduce that its expectation under $\mu_j$ satisfies

$$\mathbb{E}_{\mu_j}[N_j] = \mathrm{Card}\,\Lambda_j\, c(\mathcal{D})^{-\alpha}2^{-j\alpha-1} \leqslant \frac{1}{2}c(\mathcal{D})^{d-\alpha}2^{j(d-\alpha)}.$$

Moreover, simple computation on Binomial laws shows that the centred moment of order $\kappa \geqslant 1$ satisfies

$$\mathbb{E}_{\mu_j}\left[|N_j - \mathbb{E}(N_j)|^\kappa\right] \leqslant c(\kappa)2^{j\kappa/2(d-\alpha)}.$$

By Markov inequality,

$$\mu_j\big[|N_j - \mathbb{E}(N_j)| \geqslant \mathbb{E}(N_j)/4\big] \leqslant 4^\kappa\, \mathbb{E}_{\mu_j}[N_j]^{-\kappa}\, \mathbb{E}_{\mu_j}[(N_j - \mathbb{E}(N_j))^\kappa]$$
$$\leqslant 4^\kappa c(\kappa) 2^{j(d-\alpha)\kappa/2}.$$

Since $d-\alpha > 0$ and $\kappa$ is arbitrary, if $j = j(\varepsilon)$ is given by either (5.9) or (5.10), we deduce that $\mu_{j(\varepsilon)}[N_{j(\varepsilon)} \geqslant c(\mathcal{D})^{d-\alpha} 2^{j(\varepsilon)(d-\alpha)}]$ is negligible versus any power of $\varepsilon$ as $\varepsilon \to 0$. We have thus shown at this stage that for any $c > 0$:

$$\mu_{j(\varepsilon)}\left[f \notin \mathcal{M}(\tilde{s}(\bullet), r)\right] = o(\varepsilon^c). \tag{5.13}$$

Next, we pick an arbitrary estimator $\widehat{f}$. Since for any $f \in \mathcal{M}\left(\tilde{s}(\bullet), r\right)$ we have $\|f\|_{L^\pi} \leqslant \|f\|_{B^{s(1/\pi)}_{\pi,\infty}} \leqslant r$, we can assume that $\|\widehat{f}\|_{L^\pi} \leq 2r$, say, without increasing the minimax risk. Now,

$$\sup_{f\in\mathcal{M}(\tilde{s}(\bullet),r)} \mathbb{E}_f\left[\|\widehat{f} - f\|^\pi_{L^\pi}\right] \geqslant \frac{\int \mathbb{E}_f\left[\|\widehat{f} - f\|^\pi_{L^\pi}\right] 1_{\{f\in\mathcal{M}(\tilde{s}(\bullet),r)\}} \mu_j(\mathrm{d}f)}{\mu_j\big[f \in \mathcal{M}(\tilde{s}(\bullet), r)\big]}$$
$$= \frac{\mathcal{E}_{B,j,\pi}\big(\widehat{f}\big)^\pi}{\mu_j\big[f \in \mathcal{M}(\tilde{s}(\bullet), r)\big]} - r_j,$$

where

$$r_j = \frac{\int \mathbb{E}_f\left[\|\widehat{f} - f\|^\pi_{L^\pi}\right] 1_{\{f\notin\mathcal{M}(\tilde{s}(\bullet),r)\}} \mu_j(\mathrm{d}f)}{\mu_j\big[f \in \mathcal{M}(\tilde{s}(\bullet), r)\big]}.$$

Since $\mu_{j(\varepsilon)}\big[f \in \mathcal{M}(\tilde{s}(\bullet), r)\big] \to 1$ as $\varepsilon \to 0$ by (5.13), using Proposition 5.5, the result follows if we show that $r_{j(\varepsilon)} = o(\varepsilon^c)$ for $c$ large enough. From the definition of the prior, the wavelet coefficients $c_\lambda$ of $f$ are $\mu_j(\mathrm{d}f)$-a.s. bounded by $rc(\mathcal{D})^{-\beta} 2^{-|\lambda|(\beta+d/2)}$. This implies that with full $\mu_j(\mathrm{d}f)$-probability the function $f$ is bounded by some constant (independent of $f$) in $L^\infty$-norm and thus in $L^\pi$-norm too. Since $\|\widehat{f}\|^\pi_{L^\pi} \leqslant 2r$, $r_j(\varepsilon)$ is bounded by some constant times

$$\mu_{j(\varepsilon)}\big[f \notin \mathcal{M}(\tilde{s}(\bullet), r)\big] / \mu_{j(\varepsilon)}\big[f \in \mathcal{M}(\tilde{s}(\bullet), r)\big]$$

and is thus of right order by (5.13) again. □

**Step 3: Arbitrary Besov domains.** Let $s(\bullet)$ be an admissible function and recall that $p^\star$ is the unique solution of (3.13). We choose for $\tilde{s}(\bullet)$ any affine function which is tangent to $s(\bullet)$ at the point $(1/p^*, s(1/p^*))$. For instance, set $\tilde{s}(1/p) = \beta^\star + \alpha^\star/p$ with $\alpha^\star = s'(1/p^\star)$ and $\beta(1/p^\star) - s'(1/p^\star)/p^\star$. By the concavity of $s(\bullet)$ we have $\tilde{s}(\bullet) \geqslant s(\bullet)$ and thus:

$$\mathcal{M}\left(\tilde{s}(\bullet), r\right) \subset \mathcal{M}\left(s(\bullet), r\right), \tag{5.14}$$

so we may prove the lower bound with $\tilde{s}(\bullet)$ in place of $s(\bullet)$. We assume first that $s'(1/p^\star) \neq d$. We have

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{M}(\tilde{s}(\bullet),r)} \mathbb{E}_f \left[\|\widehat{f} - f\|_{L^\pi}^\pi\right] \gtrsim \inf_{\widehat{f}} \mathcal{E}_{B,j(\varepsilon),\pi}\big(\widehat{f}\big)^\pi$$
$$\gtrsim \varepsilon^{2\frac{\alpha^\star + \pi\beta^\star}{d+2\beta^\star}} \left[\left(\log \tfrac{1}{\varepsilon}\right)^{\frac{\alpha^\star+\pi\beta^\star}{d+2\beta^\star}} 1_{\alpha^\star \neq 0} + 1_{\alpha^\star = 0}\right]$$

where we successively applied Proposition 5.6, the fact that $\alpha^\star \neq d$ and Proposition 5.5. The conclusion follows from the following identity:

$$\frac{\alpha^\star + \pi\beta^\star}{d + 2\beta^\star} = \frac{\pi s(1/p^\star)}{d + 2s(1/p^\star)}. \tag{5.15}$$

Indeed, if $\alpha^\star = s'(1/p^\star) = 0$ this follows from

$$\beta^\star = s(1/p^\star) - s'(1/p^\star)/p^\star = s(1/p^\star).$$

In the general case, replacing $\alpha^\star$ and $\beta^\star$ by their value in function of $s(\bullet)$ we have:

$$\frac{\alpha^\star + \pi\beta^\star}{d + 2\beta^\star} = \frac{\pi s(1/p^\star) + s'(1/p^\star)[1 - \frac{\pi}{p^\star}]}{d + 2s(1/p^\star) - s'(1/p^\star)\frac{2}{p^\star}}.$$

After some computations using (3.13), one checks that the rational function

$$x \rightsquigarrow \frac{\pi s(1/p^\star) + x[1 - \frac{\pi}{p^\star}]}{d + 2s(1/p^\star) - x\frac{2}{p^\star}},$$

is independent of $x$. We end the proof by noting that the right-hand side of (5.15) is obtained for $x = 0$.

**Step 4: The case $s'(1/p^\star) = d$.** We cannot rely on the comparison between the Bayes risk and the minimax one given in Proposition 5.6 anymore. Nevertheless we still consider $\tilde{s}(1/p) = \alpha^\star/p + \beta$ with $\alpha^\star = s'(1/p^\star) = d$ and $\beta = s(1/p^\star) - d/p^\star$. Theorem 3.5 is then a consequence of (5.15) with $\alpha^* = d$ and of the following proposition.

**Proposition 5.7.** *We have*

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{M}(\tilde{s}(\bullet),r)} \mathbb{E}_f \left[\|\widehat{f} - f\|_{L^\pi}^\pi\right] \gtrsim \left(\varepsilon\sqrt{\log \tfrac{1}{\varepsilon}}\right)^{2\frac{d+\pi\beta^\star}{d+2\beta^\star}}.$$

*Proof.* For sake of brevity, we only give the mains steps. Define $j = j(\varepsilon)$ by

$$\frac{d}{16}(j-1)2^{(j-1)(d+2\beta^\star)} \leqslant \varepsilon^{-2} r^2 c(\mathcal{D})^{-2\beta^*} \leqslant \frac{d}{16} j 2^{j(d+2\beta^\star)}, \tag{5.16}$$

and for $\lambda \in \Lambda_j$ set $f^{(\lambda)} := r 2^{-j(d/2+\beta^\star)} c(\mathcal{D})^{-\beta}\psi_\lambda$. By Lemma 5.4 the $f^{(\lambda)}$ are elements of $\mathcal{M}(\tilde{s}(\bullet), r)$. Moreover, the following three classical properties are easily checked:

- For all $\lambda \neq \lambda' \in \Lambda_j$, we have $\|f^{(\lambda)} - f^{(\lambda')}\|_{L^\pi} \geqslant r c(\mathcal{D})^{-\beta^\star} 2^{-j(d/\pi+\beta^\star)}$.
- For all $\lambda \in \Lambda_j$, we have $\mathbb{P}_{f^{(\lambda)}} \ll \mathbb{P}_0$ where $\mathbb{P}_0$ is the law of the observation (1.1) when $f = 0$ and $\ll$ means absolute continuity with respect to probability measures.
- We have a bound control on the Kullback-Leibler divergence $K(\mathbb{P}_{f^{(\lambda)}} \mid \mathbb{P}_0) \leqslant \|f^{(\lambda)}\|_{L^2}^2 \varepsilon^{-2} = r^2 c(\mathcal{D})^{-2\beta^*} 2^{-j(d+2\beta^\star)}\varepsilon^{-2}$. By (5.16) and $\log(\mathrm{Card}\{\Lambda_j\}) \sim dj$, we readily obtain, for large enough $j$:
$$\frac{1}{\mathrm{Card}\,\Lambda_j}\sum_{\lambda\in\Lambda_j} K(\mathbb{P}_{f^{(\lambda)}} \mid \mathbb{P}_0) \leqslant \log(\mathrm{Card}\,\Lambda_j)/8.$$

Then, standard arguments based on Fano's lemma, see e.g. [21], entail:

$$\inf_{\widehat{f}}\sup_{\lambda\in\Lambda_j} \mathbb{P}_{f^{(\lambda)}}\left[\|\widehat{f} - f\|_{L^\pi}^\pi \geqslant r^\pi c(\mathcal{D})^{\pi\beta^\star} 2^{-j(d+\pi\beta^\star)}\right] \geqslant c > 0, \tag{5.17}$$

where $c$ is some constant independent of $\varepsilon$. We conclude by (5.17) and (5.16). □

### 5.3 Proof of Theorem 3.8

In a first step we show that the exponent $s(1/p^\star)/(d+2s(1/p^\star))$ can be reinterpreted as the right hand side of (3.17). Then we prove (3.16).

**Step 1: A new expression for the minimax rate.** For the proof, we use the convenient notation $\alpha \in s'(1/p)$ when $\alpha$ is a real number in the interval $[s'_l(1/p), s'_r(1/p)]$ whose endpoints are the, possibly different, left and right derivatives of $s(\bullet)$ at $1/p$. Then the proof of (3.17) is based on the following key identity:

$$\frac{\pi s(1/p^\star)}{2s(1/p^\star)+d} = \inf_{\substack{0\leq 1/p\leq 1/p_c\\ \alpha\in s'(1/p)}} \frac{\pi s(1/p) + \alpha[1-\frac{\pi}{p}]}{d + 2s(1/p) - \frac{2\alpha}{p}}, \tag{5.18}$$

where the infimum is attained for $1/p = 1/p^\star$, with any choice of $\alpha \in s'(1/p^\star)$. Actually, since $\pi \geq p_c + 2$, we must have $1/p^\star \leq 1/p_c$ and we have seen in Step 3 of the proof of Theorem 3.5 that, for $1/p = 1/p^\star$ the fractional function above does not depend on $\alpha$ and is equal to the left hand side of (5.18). It remains to see that the infimum in the right hand side of (5.18) is attained for $1/p = 1/p^\star$, this can easily be checked by standard computations relying on the concavity of $s$. Define the Legendre transform of $s(\bullet)$ as $\mathcal{L}s(\alpha) := \inf_{q>0} \big\{q\alpha - s(q)\big\}$. If $\alpha_0 \in s'(q_0)$ for some $q_0 > 0$, we easily see that the infimum in the definition of $\mathcal{L}s(\alpha)$ is attained at $q = q_0$ and we deduce the useful relationship

$$\mathcal{L}s(\alpha_0) = q_0\alpha_0 - s(q_0).$$

This enables to transform (5.18) into,

$$\begin{aligned}\frac{\pi s(1/p^\star)}{2s(1/p^\star) + d} &= \inf_{\substack{0 \leqslant 1/p \leq 1/p_c \\ \alpha \in s'(1/p)}} \frac{\alpha - \pi \mathcal{L}s(\alpha)}{d - 2\mathcal{L}s(\alpha)} \\ &= \inf_{s'_r(1/p_c) \leq \alpha \leq s'_r(0)} \frac{\alpha - \pi \mathcal{L}s(\alpha)}{d - 2\mathcal{L}s(\alpha)}, \qquad (5.19)\end{aligned}$$

where $s'_r$ denotes the right derivative of $s(\bullet)$. Now, since by (3.16), $d(H) = \inf_{p \geq p_c} \big\{Hp - ps(1/p) + d\big\}$, we can check that if $\alpha \in s'(q_0)$ for some $q_0 \leq 1/p_c$:

$$\begin{aligned}d\big(-\mathcal{L}s(\alpha)\big) = d\big(s(q_0) - \alpha q_0\big) &= \inf_{p \geq p_c} \big\{(s(q_0) - \alpha q_0)p - ps(1/p) + d\big\} \\ &= (s(q_0) - \alpha q_0)/q_0 + (1/q_0)s(q_0) + d = -\alpha + d.\end{aligned}$$

This identity enables to rewrite (5.19) as:

$$\inf_{s'_r(1/p_c) \leq \alpha \leq s'_r(0)} \frac{d - d(-\mathcal{L}s(\alpha)) - \pi \mathcal{L}s(\alpha)}{d - 2\mathcal{L}s(\alpha)}.$$

The function $-\mathcal{L}s(\bullet)$ maps $[s'_r(0), s'_r(1/p_c)]$ onto

$$[H_0, H_c] := [s(0), s(1/p_c) - s'_r(1/p_c)/p_c],$$

therefore

$$\inf_{s'_r(1/p_c) \leq \alpha \leq s'_r(0)} \frac{d - d(-\mathcal{L}s(\alpha)) - \pi \mathcal{L}s(\alpha)}{d - 2\mathcal{L}s(\alpha)} = \inf_{s(0) \leq H \leq H_c} \frac{d - d(H) + \pi H}{2H + d}.$$

For $H \in [H_c, d/p_c]$, the representation (2.2) yields $d(H) = Hp_c$, and we deduce, by $\pi \geq 2 + p_c$, that the infimum can actually be taken over

$[s(0), d/p_c]$ which is the desired result.

**Step 2: Proof of** (3.16)**.** Since $\mathcal{FP}\big(s(\bullet), r\big)$ is a subset of $\mathcal{M}\big(s(\bullet), r\big)$ it is immediate that the upper bound of Theorem 3.4 holds true over it. To see that the same lower bound still holds on this subclass, we have to slightly modify the proof of the Theorem 3.5 in the following way. Let $g_0$ be some fixed function in $\mathcal{FP}\big(s(\bullet), r\big)$ and then modify the prior $\mu_j$ introduced in Step 1 of Section 5.2 by adding $g_0$ to every realization drawn under the probability $\mu_j$. Denote by $\tilde{\mu}_j$ the corresponding new prior. It is clear that the Bayes risk is unchanged. Moreover, $\tilde{\mu}_j(\mathrm{d}f)$-a.s., the wavelet expansion of $f$ coincides with the one of $g_0$ on low scales. Thus, $\tilde{\mu}_j(\mathrm{d}f)$-a.s, the function $f$ has the same local behavior as $f_0$ and its singularity spectrum is given by (2.2). We then repeat the arguments of Steps 2 and 3, but now $\tilde{\mu}_j$ is supported with large probability on $\mathcal{FP}\big(s(\bullet), 2r\big)$. If $s'(1/p^\star) = d$, we modify Step 4 of Section 5.2 by adding $g_0$ to the $f^{(\lambda)}$ accordingly.

# 6 Appendix

## 6.1 The Frisch-Parisi heuristics

If $E \subset \mathbb{R}^d$ and $\eta > 0$ we let $\mathcal{C}(E, \eta)$ denote the set of countable coverings $c$ of $E$ by open balls $b$ with diameter at most $\eta$. The *Hausdorff dimension* of $E$ is by definition

$$\begin{aligned}\dim(E) : &\inf\big\{q \geq 0,\ \lim_{\eta \to 0} \inf_{c \in \mathcal{C}(E,\eta)} \sum_{b \in c} |b|^q = 0\big\} \\ &= \sup\big\{q \geq 0,\ \lim_{\eta \to 0} \inf_{c \in \mathcal{C}(E,\eta)} \sum_{b \in c} |b|^q = +\infty\big\}.\end{aligned}$$

The *box-dimension* (or Minkowski dimension) of $E$ is defined as follows: let $\mathcal{C}_{\mathrm{exact}}(E, \eta)$ denote the set of countable coverings of $E$ by open balls with diameter exactly equal to $\eta$. Set

$$\Lambda(E, \eta) := \inf_{c \in \mathcal{C}_{\mathrm{exact}}(E,\eta)} \mathrm{Card}(c),$$

*i.e.* the minimal number of balls with diameter $\eta$ that are necessary to cover $E$. The *lower box-dimension* of $E$ is defined as

$$\underline{d}(E) := \liminf_{\eta} \frac{\log\big(\Lambda(E, \eta)\big)}{-\log \eta}$$

and likewise, the *upper box-dimension* of $E$ is

$$\bar{d}(E) := \limsup_{\eta} \frac{\log\big(\Lambda(E,\eta)\big)}{-\log \eta}.$$

We have the following chain of inequality between the Hausdorff dimension and the box-dimension:

$$d(E) \le \underline{d}(E) \le \bar{d}(E).$$

In practice, the Hausdorff dimension is often approximated by the box-dimension. For simplicity, we develop the argument in dimension $d = 1$ and $\mathcal{D} = [0,1]$. We interpret a scaling law of the type $M_j(f, 1/p) \approx 2^{-jps(1/p)}$ for $j \to \infty$ where $M_j(f, 1/p)$ is defined in (1.7). The exponent $s(1/p)$ is the value of the Besov domain of $f$ at $1/p$. The contribution of points with maximal regularity $H > 0$, that is over the set

$$S_f(H) := \{x \in [0,1],\ h_f(x) = H\}$$

will be given by

$$2^{-j} \sum_{k2^{-j} \in S_f(H)}^{2^j} \big|f(k2^{-j}) - f\big((k-1)2^{-j}\big)\big|^p \approx 2^{-j\left(pH - d(H) + 1\right)} \tag{6.1}$$

using that at most $2^{jd(H)}$ boxes are necessary to cover $S_f(H)$. Here, we purposefully make a confusion bewteen the box-dimension of $S_f(H)$ and its Hausdorff dimension. Next, by a geometric series argument, the total contribution in $H$ will be dominated by the maximal exponent in (6.1) so that

$$M_j(f, 1/p) \approx \sup_{H>0} 2^{j\left(Hp - d(H) + 1\right)} \approx 2^{-jps(1/p)}$$

which yields

$$ps(1/p) = \sup_{H>0} \big\{Hp - d(H) + 1\big\}.$$

We recognize the Legendre transfom of $H \rightsquigarrow d(H) - 1$ so if the inversion is meaningful, we obtain the Frisch-Parisi conjecture

$$d(H) = \sup_{p>0} \big\{pH - ps(1/p) + 1\big\}. \tag{6.2}$$

In dimension $d \ge 2$, (6.2) reads like (1.6).

## 6.2 An univariate Bayes risk

**Lemma 6.1.** *Set*

$$\rho_{j,\varepsilon,\pi} := \inf_{g:\mathbb{R}\to\mathbb{R}} \mathbb{E}\left[|g(c_j+\varepsilon Z) - c_j|^\pi\right] \tag{6.3}$$

*where $Z$ is a standard Gaussian variable and $c_j$ is an independent Bernoulli variable taking values $K = rc(\mathcal{D})^{-\beta}2^{-j(\beta+d/2)}$ and $0$, with probability $q = c(\mathcal{D})^{-\alpha}2^{-j\alpha-1}$ and $1-q$ and respectively. If $j = j(\varepsilon)$ and $\varepsilon$ are related by* (5.9) *in the case $\alpha \neq 0$ (and by* (5.10) *otherwise) we have:*

$$\rho_{j,\varepsilon,\pi} \geqslant c(\alpha,\beta,\pi,d,c(\mathcal{D}))qK^\pi.$$

*Proof.* Note first that the infimum in the right-hand-side of (6.3) is obtained for $\widehat{g}(c) := \mathrm{argmin}_{x\in\mathbb{R}} \mathbb{E}[|x-c_j|^\pi \mid \tilde{c}_j = c]$ where $\tilde{c}_j = c_j + \varepsilon Z$. The posterior distribution of $c_j$ conditional on $\tilde{c}_j = c$ has support $\{0, K\}$ with:

$$\mathbb{P}[c_j = K \mid \tilde{c}_j = c] = \frac{q\Phi_K(c)}{q\Phi_K(c) + (1-q)\Phi_0(c)},$$
$$\mathbb{P}[c_j = 0 \mid \tilde{c}_j = c] = \frac{(1-q)\Phi_0(c)}{q\Phi_K(c) + (1-q)\Phi_0(c)},$$

where $\Phi_0$ (respectively $\Phi_K$) is the Gaussian density function with variance $\varepsilon^2$ and mean 0 (respectively $K$). Thus $\widehat{g}(c)$ is obtained as the minimizer of

$$x \rightsquigarrow |K-x|^\pi q\Phi_K(c) + |x|^\pi(1-q)\Phi_0(c).$$

It is clear that the minimizer lies in $[0, K]$. If $\pi > 1$, it is easily checked that $\widehat{g}(c)$ is the unique solution of the equation of the variable $x$

$$\left(\frac{K-x}{x}\right)^{\pi-1} = \frac{(1-q)\Phi_0(c)}{q\Phi_K(c)},$$

and if $\pi = 1$, $\widehat{g}(c) = K1_{\left\{\frac{(1-q)\Phi_0(c)}{q\Phi_K(c)}<1\right\}}$. In both cases, we see that $\widehat{g}(c)$ lies in the interval $[0, K/2]$ as soon as

$$\frac{(1-q)\Phi_0(c)}{q\Phi_K(c)} \geqslant 1. \tag{6.4}$$

Using the identity $\Phi_K(c) = \Phi_0(c)e^{(Kc-K^2/2)\varepsilon^{-2}}$, we can rewrite the condition (6.4) as $c \leqslant \overline{c}_{\varepsilon,K,q}$ with:

$$\overline{c}_{\varepsilon,K,q} = K/2 - \log(q)\varepsilon^2/K + \log(1-q)\varepsilon^2/K. \tag{6.5}$$

We have thus shown at this stage that $c < \overline{c}_{\varepsilon,K,q}$ implies $\widehat{g}(c) \leqslant K/2$. This is sufficient to obtain a lower bound for the Bayes risk. Indeed:

$$\begin{aligned}
\rho_{j,\varepsilon,\pi} &= \mathbb{E}\left[|\widehat{g}(\tilde{c}_j) - c_j|^{\pi}\right] \\
&\geqslant \mathbb{E}\left[|\widehat{g}(\tilde{c}_j) - c_j|^{\pi} 1_{\{c_j = K\}} 1_{\{\tilde{c}_j < \overline{c}_{\varepsilon,K,q}\}}\right] \\
&\geqslant (K/2)^{\pi}\, \mathbb{P}[c_j = K,\ \tilde{c}_j < \overline{c}_{\varepsilon,K,q}] \\
&= (K/2)^{\pi} q\, \mathbb{P}[K + \varepsilon Z < \overline{c}_{\varepsilon,K,q}](K/2)^{\pi} q\, \mathbb{P}[Z < (\overline{c}_{\varepsilon,K,q} - K)/\varepsilon].
\end{aligned}$$

The lemma is proved if we can show that the probability above remains bounded away from zero, or equivalently if

$$\frac{\overline{c}_{\varepsilon,K,q} - K}{\varepsilon} = -\frac{K}{2\varepsilon} - \log q \frac{\varepsilon}{K} + \log(1-q)\frac{\varepsilon}{K} \tag{6.6}$$

remains bounded away from $-\infty$. In the case $\alpha = 0$, we have $q = 1/2$ and if $j = j(\varepsilon)$ is given by the condition (5.10) we get $K/\varepsilon \in [2^{-d-2\beta} r^2 c(\mathcal{D})^{-2\beta}, r^2 c(\mathcal{D})^{-2\beta}]$. This implies that (6.6) remains bounded by below. In the case $\alpha > 0$, we have $\log q \sim -j\alpha \log 2$ and if $j = j(\varepsilon)$ is given by (5.9) with $M < r^{-2} c(\mathcal{D})^{2\beta} 2 \log 2$ one checks that $-K/(2\varepsilon) - (\log q)\varepsilon/K \to +\infty$ as $\varepsilon \to 0$. Since $\log(1-q)\varepsilon/K \to 0$, the quantity (6.6) remains bounded away from $-\infty$ and the lemma is proved. $\square$

**Acknowledgments** The authors wish to thank the anonymous referees and the associated editor for their helpful comments and suggestions.

# References


[1] Arneodo, A., d'Aubenton-Cafara Y., Bacry E., Graves, P.V., Muzy J.F. and Thermes C. (1996). Wavelet-based fractal analysis of DNA sequences. *Physica D*, **96**, 291–320.

[2] Bacry, E. Muzy, J.F. and Delour J. (2000). Modelling fluctuations of financial time series: from cascade processes to stochastic volatility models. *Eur. Phys. J. B*, **17**, 537–548.

[3] Bacry, E. Muzy, J.F. and Delour J. (2001). Multifractal random walks. *Phys. Rev. E* **64**, 026103–026106.

[4] Barral J. and Mandelbrot B.B. (2004) Introduction to infinite products of independent random functions (Random multiplicative multifractal measures. Part I). *Proc. Sympos. Pure Math.*, **72(2)**, *Amer. Math. Soc., Providence*, 3–16.

[5] Cohen, A. (2000) *Wavelets in numerical analysis*, Handbook of Numerical Analysis, vol. VII, P.G. Ciarlet and J.L. Lions, eds., Elsevier, Amsterdam.

[6] Cohen A., DeVore R., Kerkyacharian G. and Picard D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comp. Harm. Anal.* **11**, 167–191.

[7] DeVore, R., Konyagin, S.V. and Temlyakov, V.N. (1998). Hyperbolic wavelet approximation. Constructive approximation, Springer-Verlag.

[8] Donoho D.L., Johnstone I.M. (1994). Minimax risk over $l_p$-balls for $l_q$-error. *Probab. Theory Related Fields* **99(2)**, 277–303.

[9] Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.

[10] Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, **90**, 1200–1224.

[11] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet Shrinkage: Asymptopia? *J. Roy. Statist. Soc.* **57** 301–369.

[12] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539.

[13] Frisch U. (1995) *Turbulence.* Cambridge Univ. Press, Cambridge.

[14] Frisch, U. and Parisi G. (1985). On the singularity structure of fully developed turbulence; appendix to fully developed turbulence and intermittencey, by U. Frisch, in *Proc. Int. Summer School Phys. Enrico Fermi.* North-Holland, 84–88.

[15] Gagne, Y. (1987) Étude expérimentale de l'intermittence et des singularités dans le plan complexe en turbulence pleinement développée. Thèse de l'Université de Grenoble, 1987.

[16] Gagne, Y., Marchand, M. and Castaing, B. (1994). Conditional velocity pdf in 3D Turbulence. *J. Phys. II*, **4**, 1–8.

[17] Gamboa, F. and Loubes, J.M. (2005). Wavelet estimation of a multifractal function. *Bernoulli*, **11**, 221–246.

[18] Gamboa, F. and Loubes, J.M. (2007). Estimation of parameters of a multifractal process. *Test*, **16**, 383–407.

[19] Ghashghaie, S., Breymann, W., Peinke, J., Talkner, P. and Dodge, Y. (1996). Turbulent cascades in foreign exchange markets. *Nature*, **381**, 767–770.

[20] Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block threshold wavelet estimators. *Statist. Sinica*, **9**, 33–49.

[21] Ibragimov I.A. and Has'minskii R.Z. (1981). Statistical estimation: asymptotic theory. *Springer, Berlin.*

[22] Jaffard S. (1997). Multifractal formalism for functions part I: results valid for all functions. *SIAM J. Math. Anal.* **28**, 944–970.

[23] Jaffard, S. (2000a). On the Frisch-Parisi conjecture. *J. Math. Pures Appl.* **79**, 525–552.

[24] Jaffard, S. (2000b). On lacunary wavelet series. *Ann appl probab.* **10**, 313–329.

[25] Kahane J.P. and Peyrière J. (1976). Sur certaines martingales de Benoit Mandelbrot. *Adv. in Mathematics* **22**, 131–145.

[26] Kerkyacharian G. and Picard D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. With comments, and a rejoinder by the authors. *Test* **9**, 283–344.

[27] Kolmogorov, A.N. (1962). A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds numbers. *J. Fluid. Mech.*, **13**, 82–85.

[28] Mandelbrot, B. B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *J. Fluid. Mech.*, **62**, 331–358.

[29] Mandelbrot, B.B. (1997). *Fractals and scaling in finance.* Springer, New-York.

[30] Meyer, Y. (1990). *Ondelettes et opérateurs.* Vol. 1. Hermann, Paris.

[31] Molchan G.M. (1996). Scaling exponents and multifractal dimensions for independent random cascades. *Comm. in Math. Phys.* **179**, 681–702.

[32] Oboukhov,, A.M. (1962). Some specific featuress of atmospheric turbulence. *J. Fluid Mech.*, **13**, 77–81.

[33] Park, K., Willinger W. (1999) Self-similar network traffic: An overview. In K. Park and W. Willinger, editors, Self-Similar Network Traffic and Performance Evaluation. Wiley Interscience.

[34] Peng C.K., Buldyrev S.V., Goldberger A.L. , Havlin S., Sciortino F., Simons M. , Stanley H.E. (1992) Long-range correlations in nucleotide sequences. *Nature* **356**, 168 - 170.

[35] Samorodnitsky, G. and Taqqu, M. (1994) *Stable non Gaussian random processes: stochastic models with infinite variance.* Chapman & Hall, New-York.

[36] Yu, Z.G, Anh, V. and Lau, K.S. (2001). Multifractal characterisation of length sequences of coding and noncoding segments in a complete genome. *Physica A,* **301**, 351–361