

# Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids

CRISTINA BUTUCEA<sup>1,\*</sup>, AMANDINE DUBOIS<sup>2,‡</sup>, MARTIN KROLL<sup>1,†</sup> and ADRIEN SAUMARD<sup>2,§</sup>

<sup>1</sup>CREST, ENSAE, Institut Polytechnique de Paris, 5 avenue Henry Le Chatelier, F-91120 Palaiseau, France. E-mail: \*[cristina.butucea@ensae.fr](mailto:cristina.butucea@ensae.fr); †[martin.kroll@ensae.fr](mailto:martin.kroll@ensae.fr)

<sup>2</sup>CREST-ENSAI, Campus de Ker-Lann, Rue Blaise Pascal, BP 37203 – 35172 Bruz Cedex, France. E-mail: ‡[amandine.dubois@ensai.fr](mailto:amandine.dubois@ensai.fr); §[adrien.saumard@ensai.fr](mailto:adrien.saumard@ensai.fr)

We address the problem of non-parametric density estimation under the additional constraint that only privatised data are allowed to be published and available for inference. For this purpose, we adopt a recent generalisation of classical minimax theory to the framework of local  $\alpha$ -differential privacy and provide a lower bound on the rate of convergence over Besov spaces  $\mathcal{B}_{pq}^s$  under mean integrated  $\mathbb{L}^r$ -risk. This lower bound is deteriorated compared to the standard setup without privacy, and reveals a twofold elbow effect. In order to fulfill the privacy requirement, we suggest adding suitably scaled Laplace noise to empirical wavelet coefficients. Upper bounds within (at most) a logarithmic factor are derived under the assumption that  $\alpha$  stays bounded as  $n$  increases: A linear but non-adaptive wavelet estimator is shown to attain the lower bound whenever  $p \geq r$  but provides a slower rate of convergence otherwise. An adaptive non-linear wavelet estimator with appropriately chosen smoothing parameters and thresholding is shown to attain the lower bound within a logarithmic factor for all cases.

*Keywords:* adaptive estimation; Besov classes of functions; density estimation; local differential privacy; lower bounds; minimax rates; wavelet thresholding

## 1. Introduction

### Problem statement

In the modern information age, increasingly more institutions are collecting and storing data. Provided that a certain amount of privacy is guaranteed, some of these institutions might be willing to provide access to selected data sets. Examples of such data may include information about participants in a medical study, clients of a web service, or persons interviewed in a scientific survey. In this framework, the following questions arise naturally: How can data be sufficiently anonymised, given a rigorous definition of privacy, and what are the consequences for subsequent data analyses resulting from the chosen anonymisation procedure? The answer to these questions depends on several interacting parameters, namely the privacy definition at hand, the potential extent of collaboration of the involved data holding entities, and the kind of data mining tasks that should be feasible based on the private data.

In this paper, we consider the problem of non-parametric density estimation under local differential privacy as a special instance of the general problem sketched in the previous paragraph: For  $i = 1, \dots, n$ , the  $i$ th data holder observes a real-valued random variable  $X_i$  distributed according to a probability density function  $f$ . The aim is that every data holder releases an anonymised view  $Z_i$  of  $X_i$  such that the privacy notion of *local differential privacy*, that is introduced next, is satisfied and that the density  $f$  can be estimated from the data  $Z_1, \dots, Z_n$  in an optimal way.

## Local differential private estimation

The notion of *local differential privacy* aggregates two different concepts, namely *local* privacy and *differential* privacy, that we explain in the sequel.

The qualitative notion of *local* privacy characterises how the different entities holding the data  $X_1, \dots, X_n$  might interact to generate a private release  $Z$ . It is opposed to the concept of *global* privacy where the respective data holders share confidence in a common curator who has access to the ensemble of non-masked data  $X_1, \dots, X_n$  and generates the releasable data from this complete information. In the *local* setup, such an authority that is trusted by all the parties, does not exist. However, some amount of interaction between the different parties is still allowed. The releasable data  $Z_1, \dots, Z_n$  are obtained by successively applying suitable Markov kernels. Given  $X_i = x_i$  and  $Z_1 = z_1, \dots, Z_{i-1} = z_{i-1}$ , the  $i$ -th dataholder draws

$$Z_i \sim Q_i(\cdot \mid X_i = x_i, Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})$$

for some Markov kernel  $Q_i : \mathcal{Z} \times \mathcal{X} \times \mathcal{Z}^{i-1} \rightarrow [0, 1]$  where the measure spaces of the non-private and private data are denoted with  $(\mathcal{X}, \mathcal{X}^c)$  and  $(\mathcal{Z}, \mathcal{Z}^c)$ , respectively. An important special case is that of *non-interactive* local privacy where the random value of  $Z_i$  depends on  $X_i$  only and must not depend on preceding values of  $Z$ . More precisely, in the non-interactive case we have

$$Z_i \sim Q(\cdot \mid X_i = x_i)$$

for some Markov kernel  $Q$  that does no longer depend on the index  $i$ . The non-interactive scenario comes along with some advantages in practice since it is balanced in the sense that the data holders play a symmetric role in the privatisation process, that can also be parallelized in this case. One should notice however that a restriction to non-interactive scenarios might result in slower rates of convergence for statistical inference. But as will be presented below, this is not the case in our density estimation framework, where rates are already optimal for estimators based on non-interactively privatised data.

From a mathematical point of view, however, allowing also interactive procedures does not lead to more technical proofs. Thus, we potentially allow non-interactive methods in our minimax analysis, although the anonymisation techniques proposed in this paper are exclusively non-interactive. Let us mention that for some tasks, however, interactive mechanisms provide natural and attractive alternatives (for instance, for private estimation in generalized linear models; see [6], Section 5.2.1).

The notion of *differential* privacy is a quantitative one and introduces a condition that makes the problem at hand mathematically tractable. We provide its definition for the locally private case only and refer the reader to [14] for a definition in the global case.

**Definition 1.1.** A sequence of Markov kernels  $Q_i : \mathcal{Z} \times \mathcal{X} \times \mathcal{Z}^{i-1} \rightarrow [0, 1]$  provides  $\alpha$ -differential privacy if

$$\sup_{A \in \mathcal{Z}} \frac{Q_i(A | X_i = x, Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})}{Q_i(A | X_i = x', Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})} \leq \exp(\alpha) \quad \text{for all } x, x' \in \mathcal{X}.$$

In the non-interactive case, this condition is replaced with

$$\sup_{A \in \mathcal{Z}} \frac{Q(A | X_i = x)}{Q(A | X_i = x')} \leq \exp(\alpha) \quad \text{for all } x, x' \in \mathcal{X}.$$

We denote with  $\mathcal{Q}_\alpha$  the set of all local  $\alpha$ -differential private Markov kernels.

Thus, the parameter  $\alpha$  quantifies the amount of privacy that is guaranteed: setting  $\alpha = 0$  ensures perfect privacy whereas letting  $\alpha$  tend to infinity softens the privacy restriction. In the non-interactive case, the defining property of  $\alpha$ -differential privacy above ensures that all the probability measures  $Q(dz | X_i = x)$ ,  $x \in \mathcal{X}$  are equivalent. Hence, they admit densities  $q(\cdot | X_i = x)$  with respect to a dominating measure (that can be chosen to be equal to  $Q(dz | X_i = x^*)$ , for any  $x^* \in \mathcal{X}$ ). In case of existence of such densities, we say that the property of  $\alpha$ -differential privacy is satisfied if

$$\sup_{z \in \mathcal{Z}} \frac{q(z | X_i = x)}{q(z | X_i = x')} \leq \exp(\alpha) \quad \text{for all } x, x' \in \mathcal{X}. \tag{1.1}$$

A consequence from the definition of  $\alpha$ -differential privacy is *plausible deniability* of the data in the following sense: Given the private view  $Z_i$  only, the power of any test of the null hypothesis  $H_0 : X_i = x$  against the alternative  $H_1 : X_i = x'$  with prescribed first error probability  $\gamma$  has power bounded from above by  $\gamma \exp(\alpha)$  (see [14], Theorem 2.4).

### Rate optimal density estimation over Besov ellipsoids

Let us briefly review some well-known results on non-parametric density estimation in the *non-private* setup where  $X_1, \dots, X_n$  can be observed. This classical model provides a natural benchmark for the model where additional privacy restrictions are imposed, and having in mind the results for this benchmark model turns out to be useful for understanding the ones for the model with privacy.

Density estimation from a sample  $X_1, \dots, X_n$  of observations is one of the paradigmatic problems in non-parametric statistics. A popular framework is that of minimax optimal estimation: Given a loss function  $\ell$  (that is, a function mapping a pair of density functions  $(f, g)$  to some non-negative real number) and any class  $\mathcal{F}$  of candidate density functions, the quantity of interest is the minimax risk

$$\mathcal{R}_n(\ell, \mathcal{F}) = \inf_{\tilde{f}} \sup_{f \in \mathcal{F}} \mathbb{E}[\ell(\tilde{f}, f)], \tag{1.2}$$

where the infimum is taken over all estimators (that is,  $\sigma(X_1, \dots, X_n)$ -measurable functions). In this setup, an estimator  $\widehat{f}$  is called *rate optimal* if

$$\sup_{f \in \mathcal{F}} \mathbb{E}[\ell(\widehat{f}, f)] \leq C(\ell, \mathcal{F}) \mathcal{R}_n(\ell, \mathcal{F}).$$

Several function classes, loss functions and types of estimators have been intensively studied for the density estimation problem (see [13] and [7] for comprehensive overviews of the topic). Throughout this paper, we consider the integrated risk associated to  $\mathbb{L}^r$ -loss defined by  $\ell(f, g) = \|f - g\|_r^r$  for  $r \geq 1$ . For the Besov spaces to be considered in the sequel, wavelet methods have turned out particularly convenient. Given a father wavelet  $\varphi$  and a mother wavelet  $\psi$  associated to it, verifying some sufficient conditions (see conditions (5.10)–(5.12) in [9]), and an integer  $j_0 \in \mathbb{Z}$ , a wavelet basis of  $\mathbb{L}^2(\mathbb{R})$  is given by

$$\{\varphi_{j_0 k} = 2^{j_0/2} \varphi(2^{j_0}(\cdot) - k) : k \in \mathbb{Z}\} \cup \{\psi_{jk} = 2^{j/2} \psi(2^j(\cdot) - k) : j \geq j_0, k \in \mathbb{Z}\}. \tag{1.3}$$

Given such a basis, the probability density  $f$  admits the following formal expansion (in  $\mathbb{L}^2$  sense):

$$f = \sum_{k \in \mathbb{Z}} \alpha_{j_0 k} \varphi_{j_0 k} + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} \beta_{jk} \psi_{jk}, \tag{1.4}$$

where the wavelet coefficients are defined as

$$\alpha_{j_0 k} = \int_{\mathbb{R}} f(x) \varphi_{j_0 k}(x) \, dx \quad \text{and} \quad \beta_{jk} = \int_{\mathbb{R}} f(x) \psi_{jk}(x) \, dx.$$

An attractive property of wavelet expansions as (1.4) is that the membership of Besov spaces can be characterised in terms of its wavelet coefficients with respect to a well chosen wavelet basis. In the sequel, we will work under the following assumption on the father wavelet  $\varphi$ .

**Assumption 1.2.** Following [9], we assume that the father wavelet function  $\varphi$  generates a multiresolution analysis of  $\mathbb{L}^2(\mathbb{R})$ , that it is  $N + 1$  times weakly differentiable for some integer  $N$ , and that its derivative satisfies  $\sup_x \sum_k |\varphi^{(N+1)}(x - k)| < \infty$  a.e. Moreover, we assume that there exists a bounded, non-increasing function  $\Phi$  on  $\mathbb{R}_+$  such that  $|\varphi(u)| \leq \Phi(|u|)$  and that both  $\int \Phi(|u|) \, du < \infty$  and  $\int \Phi(|u|) |u|^N \, du < \infty$ .

Note that no assumption is needed on the mother wavelet  $\psi$  since it is defined using the father wavelet. If the father wavelet function  $\varphi$  verifies Assumption 1.2 then, given parameters  $0 < s < N + 1$  and  $1 \leq p, q \leq \infty$ , the fact that  $f$  belongs to the Besov space  $\mathcal{B}_{pq}^s$  is equivalent to  $J_{spq}(f) < \infty$  where

$$J_{spq}(f) := \|\alpha_0\|_p + \left( \sum_{j \geq 0} (2^{j(s+1/2-1/p)} \|\beta_{j\cdot}\|_p)^q \right)^{1/q}$$

for  $1 \leq q < \infty$  and the usual modification if  $q = \infty$ . Fixing such a wavelet basis, we consider Besov ellipsoids defined as

$$\mathcal{B}_{pq}^s(L) = \{f : \mathbb{R} \rightarrow \mathbb{R} : J_{spq}(f) \leq L\}.$$

Since our interest is in density estimation, a quite natural class to consider is

$$\mathcal{D}_{pq}^s = \mathcal{D}_{pq}^s(L, T) = \left\{ f : f \in \mathcal{B}_{pq}^s(L), f \geq 0, \int_{\mathbb{R}} f(x) dx = 1 \text{ and } \text{supp}(f) \subseteq [-T, T] \right\},$$

where  $\text{supp}(f)$  denotes the support of the function  $f$ . Note that we consider here the Besov smoothness of  $f$  as a function defined on the whole real line, or, equivalently, that  $f$  belongs to a periodic Besov class. It would equally be possible to define Besov smoothness over the support  $[-T, T]$ . Then the wavelet basis has to be boundary corrected so that it detects the smoothness on this interval only and not the potential lack of smoothness of  $f$  at its boundary. We refer the reader to [7] for boundary corrected wavelets, that also dispose of all the properties that we need in the sequel.

It is well known [2,7,9] that

$$\mathcal{R}_n(\|\cdot\|_r^r, \mathcal{D}_{pq}^s) \gtrsim \tau_n,$$

$$\text{where } \tau_n = \begin{cases} n^{-\frac{rs}{2s+1}} & \text{if } p > \frac{r}{2s+1}, \\ (n/\log n)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+1}} & \text{if } p \leq \frac{r}{2s+1} \text{ and } s \geq \frac{1}{p}, \end{cases} \quad (1.5)$$

and these rates are optimal or suboptimal by a logarithmic factor only (see [9] for an extensive discussion). The structural change of the rate between *dense zone* (where  $p > r/(2s + 1)$ ) and *sparse zone* (where  $p \leq r/(2s + 1)$ ) is sometimes called an *elbow effect*.

Moreover, in the dense case, we can distinguish the *homogeneous zone* when  $p \geq r$  and the *non-homogeneous zone* where  $r/(2s + 1) < p < r$ . In the homogeneous case, *linear* wavelet estimators of the form

$$\sum_{k \in \mathbb{Z}} \alpha'_{j_0 k} \varphi_{j_0 k}(x) + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \beta'_{jk} \psi_{jk}(x)$$

with  $\alpha'_{j_0 k} = \frac{1}{n} \sum_{i=1}^n \varphi_{j_0 k}(X_i)$ ,  $\beta'_{jk} = \frac{1}{n} \sum_{i=1}^n \psi_{jk}(X_i)$ , and appropriately chosen  $j_0, j_1$  are rate optimal whereas linear procedures are necessarily sub-optimal in the non-homogeneous case (see [9] and references therein). In this latter scenario as well as in the sparse case, non-linear estimators based on wavelet thresholding turn out to be optimal at least up to logarithmic factors.

### Minimax framework under privacy constraints

Let us now describe how to extend the classical minimax setup in order to encompass the framework of local differential privacy. Since not only the estimation procedure but also the Markov

kernels guaranteeing local  $\alpha$ -differential privacy can freely be chosen, it is natural to replace (1.2) with the local  $\alpha$ -differential minimax risk defined as

$$\mathcal{R}_{n,\alpha}^*(\ell, \mathcal{F}) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} \sup_{Q \in \mathcal{Q}_\alpha} \mathbb{E}_{f,Q}[\ell(f, \hat{f})].$$

Here the infimum is taken both over all  $(\mathcal{Z}, \mathcal{Z})$ -measurable estimators of  $f$  and all Markov kernels guaranteeing local  $\alpha$ -differential privacy. A tuple  $(\hat{Q}, \hat{f})$  consisting of a privacy mechanism and an estimator  $\hat{f}$  is rate optimal (with respect to the local  $\alpha$ -differential private risk) if

$$\sup_{f \in \mathcal{F}} \mathbb{E}_{f, \hat{Q}}[\ell(f, \hat{f})] \leq C(\ell, \mathcal{F}) \mathcal{R}_{n,\alpha}^*(\ell, \mathcal{F}).$$

The quantity  $\mathcal{R}_{n,\alpha}^*(\ell, \mathcal{F})$  as well as the construction of optimal privacy mechanism and estimators represent the principal interest of the rest of the paper.

## Related work

Research on statistical estimation under privacy constraints is rather recent. A landmark paper is [14] where research on the subject has been initiated and density estimation via histograms and orthogonal series in the global privacy setup have been discussed. In the same global framework, the article [8] considers anonymisation of functional data and discusses kernel density estimators as the main example. Local  $\alpha$ -differential privacy was intensively studied in [5] and the companion article [6]. In [5] the authors show that the well-known technique of randomized response from survey statistics can be interpreted under the umbrella of local  $\alpha$ -differential privacy. In the context of density estimation, [5] established minimax rates of convergence for the mean integrated squared error over Sobolev classes with arbitrary smoothness parameter  $\beta \geq 1$ . They establish the minimax rate of order  $n^{-\beta/(\beta+1)}$  for the mean integrated squared error over Sobolev classes with  $\beta = 1$  and show that this optimal rate can be attained by Laplace perturbation of empirical histogram coefficients. The papers [5,6] provide also results for Sobolev classes with higher degrees of smoothness ( $\beta > 1$ ) but in this case a mere perturbation of the empirical Fourier coefficients does not lead to a rate optimal method (see [5], Observation 1 for the non-optimality of this approach). By means of a more sophisticated sampling technique (see [5], p. 11 or [6], Section 5.2.2), however, the authors derive the minimax rate of convergence that is  $(n\alpha^2)^{-\beta/(\beta+1)}$  also in the general case. Furthermore, [5] provides private versions of classical information-theoretical bounds that allow to apply standard lower bound techniques also in the private setup. In [12], the estimation of linear functionals in the framework of local privacy is considered and a characterisation of the rates of convergence in terms of moduli of continuity is obtained which is in parallel to well-known results for the non-private setup [3]. This general analysis contains the private estimation of a probability density at a fixed point under mean squared error as a special case.

### Main results

In Section 2, in addition and in formal analogy to (1.5), we derive the following lower bound on the private minimax risk:

$$\mathcal{R}_{n,\alpha}^*(\|\cdot\|_r^r, \mathcal{D}_{pq}^s) \gtrsim \mathfrak{r}_{n,\alpha}^*,$$

$$\text{where } \mathfrak{r}_{n,\alpha}^* = \begin{cases} (n(e^\alpha - 1)^2)^{-\frac{rs}{2s+2}} & \text{if } p > \frac{r}{s+1}, \\ \left( \frac{n(e^\alpha - 1)^2}{\log(n(e^\alpha - 1)^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}} & \text{if } p \leq \frac{r}{s+1} \text{ and } s \geq \frac{1}{p}. \end{cases} \quad (1.6)$$

This lower bound is complemented by corresponding upper bound results: The anonymisation technique used to create the private views of the non-releasable data  $X_1, \dots, X_n$  consists in an appropriately scaled version of the classical Laplace mechanism applied on the empirical wavelet functions (Section 3). The wavelet estimators considered in Sections 4 and 5 are based on the availability of the privatised data  $Z_1, \dots, Z_n$  only. As in the non-private case, a linear wavelet estimator attains the given rate in the homogeneous case, that is, whenever  $p \geq r$  (Section 4). In Section 5, we study non-linear estimators and show that an estimator using hard thresholding can nearly attain the lower bounds both in the dense and in the sparse zone.

### Notational conventions

For real numbers  $a, b$  we write  $\llbracket a, b \rrbracket = [a, b] \cap \mathbb{Z}$ . We denote with  $C$  a generic constant that might change with every appearance. For two sequences  $\{a_n\}_n, \{b_n\}_n$ , we denote by  $a_n \lesssim b_n$  that there exist some constant  $C > 0$  and a fixed integer number  $N$  such that  $a_n \leq Cb_n$ , for all  $n \geq N$ . We say that  $a_n \asymp b_n$ , if both  $a_n \lesssim b_n$  and  $b_n \lesssim a_n$ . If  $b_n > 0$ , we denote by  $a_n \simeq b_n$  the fact that  $a_n/b_n \rightarrow 1$  as  $n \rightarrow \infty$ . We recall that a centred Laplace distribution with parameter  $\lambda > 0$  has the probability density function  $p_\lambda(x) = \frac{1}{2\lambda} \exp(-\frac{|x|}{\lambda})$ , for all real number  $x$ . In particular, if  $X \sim p_\lambda$ , then  $\mathbb{E}|X|^k = k!\lambda^k$  for all  $k \in \mathbb{N}$ .

## 2. Lower bounds

The purpose of this section is to derive (1.6) and hence providing an analogue of (1.5) under local  $\alpha$ -differential privacy. To this purpose, we proceed in two steps. The first lower bound, given in Proposition 2.1, is stronger in the private dense zone ( $p > r/(s+1)$ ), whereas the second one, given in Proposition 2.2, dominates in the private sparse zone where  $p \leq r/(s+1)$ . An essential tool for both proofs is a strong information theoretical inequality (our Lemma A.1) proved in [6], which states a bound for the Kullback–Leibler divergence between any distributions that have been processed through an arbitrary channel guaranteeing local  $\alpha$ -differential privacy. We begin with the lower bound that is dominating in the dense zone.

**Proposition 2.1.** *Let  $\alpha \in (0, \infty)$  and let  $L, T > 0$ . Then,*

$$\inf_{\substack{\tilde{f} \\ Q \in \mathcal{Q}_\alpha}} \sup_{f \in \mathcal{D}_{pq}^s(L, T)} \mathbb{E}_{f, Q} \|\tilde{f} - f\|_r^r \gtrsim (n(e^\alpha - 1)^2)^{-\frac{rs}{2s+2}},$$

where the infimum is taken over all estimators  $\tilde{f}$  based on the private views  $Z_1, \dots, Z_n$  and all Markov kernels  $Q \in \mathcal{Q}_\alpha$  guaranteeing local  $\alpha$ -differential privacy.

The proof of Proposition 2.1 is based on a reduction of the class  $\mathcal{D}_{pq}^s$  to a finite number of hypotheses indexed by the vertices of a hypercube of suitable dimension. It is given in Section A.1 in the Appendix.

The following proposition complements Proposition 2.1 in stating a lower bound that is stronger in the private sparse zone.

**Proposition 2.2.** *Let  $\alpha \in (0, \infty)$ . Let  $p \geq 1, s \geq 1/p$  and let  $L, T > 0$ . Then,*

$$\inf_{\tilde{f}} \sup_{f \in \mathcal{D}_{pq}^s(L, T)} \mathbb{E}_{f, Q} \|\tilde{f} - f\|_r^r \gtrsim \left( \frac{\log(n(e^\alpha - 1)^2)}{n(e^\alpha - 1)^2} \right)^r, \frac{s-1/p+1/r}{2(s-1/p)+2},$$

where the infimum is taken over all estimators  $\tilde{f}$  based on the private views  $Z_1, \dots, Z_n$  and all channels  $Q \in \mathcal{Q}_\alpha$  providing local  $\alpha$ -differential privacy.

The proof of Proposition 2.2 is given in Section A.2 in the Appendix.

Taking the maximum of the lower bounds obtained in Propositions 2.1 and 2.2 yields (1.6). In addition to our novel lower bounds, the known bounds (1.5) from the non-private framework still hold true under local  $\alpha$ -differential privacy since processing the original data  $X_1, \dots, X_n$  through a privacy mechanism can be interpreted equivalently as imposing a restriction on the set of admissible estimators in (1.2). More precisely, the constraint of local  $\alpha$ -differential privacy confines the set of potential estimators to those of the form  $\tilde{f} = f \circ Q$  where  $Q \in \mathcal{Q}_\alpha$  and  $f$  is any measurable function. Thus,

$$\mathcal{R}_{n,\alpha}^* \geq \mathcal{R}_n \vee \tau_{n,\alpha}^* \geq \tau_n \vee \tau_{n,\alpha}^*,$$

where the quantity  $\tau_n$  is defined in (1.5). Hence, the following corollary holds.

**Corollary 2.3.** *Let the assumptions of Propositions 2.1 and 2.2 hold true. Then,*

$$\mathcal{R}_{n,\alpha}^* (\|\cdot\|_r^r, \mathcal{D}_{pq}^s) \gtrsim \begin{cases} n^{-\frac{rs}{2s+1}} \vee (n(e^\alpha - 1)^2)^{-\frac{rs}{2s+2}} & \text{if } p > \frac{r}{s+1}, \\ n^{-\frac{rs}{2s+1}} \vee \left( \frac{n(e^\alpha - 1)^2}{\log(n(e^\alpha - 1)^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}} & \text{if } \frac{r}{2s+1} < p \leq \frac{r}{s+1}, \\ \left( \frac{n}{\log n} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+1}} \vee \left( \frac{n(e^\alpha - 1)^2}{\log(n(e^\alpha - 1)^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}} & \text{if } p \leq \frac{r}{2s+1}. \end{cases}$$

Note that the frontier between the dense and the sparse zone in the private framework is different from the one in the non-private framework leading to a partition into three regimes for the lower bound and a twofold elbow effect. Note that these lower bounds match the upper bounds derived in Sections 4 and 5 at most up to logarithmic factors whenever  $\alpha$  stays bounded as  $n$  increases. In addition, the bounds from the non-private setup dominate provided that  $\alpha$  increases sufficiently fast in terms of  $n$ .

### 3. Privacy mechanisms

Let us denote with  $X_1, \dots, X_n$  the real-valued random variables that represent the non-private observations held by the different data holders. We assume that  $X_1, \dots, X_n \sim f$  for  $f \in \mathcal{D}_{pq}^s$ . In particular, the support of the density  $f$  is contained in the interval  $[-T, T]$ . In this section, we introduce a non-interactive privacy mechanism creating a private release  $Z_1, \dots, Z_n$  based on the non-private sample that satisfies the defining property of  $\alpha$ -differential privacy. For this purpose, we consider a wavelet basis as in (1.3). We assume in the sequel that the following condition on the parent wavelets is satisfied:

$$\varphi \text{ and } \psi \text{ are compactly supported on an interval } [-A, A]. \tag{W1}$$

The idea of the proposed anonymisation technique is to mask the empirical wavelet coefficients  $\alpha'_{j_0k}$  and  $\beta'_{jk}$  for certain values of  $j$ . A consequence of (W1) and the compact support of  $f$  is that for any  $j_0 \in \mathbb{Z}$  and any fixed resolution level  $j \in \mathbb{Z}$ , the corresponding  $\alpha_{j_0k}$  and  $\beta_{jk}$  can *a priori* be non-zero for a finite number of  $k$  only. We denote the set of  $k$  with potentially non-zero  $\alpha_{j_0k}$  by  $\mathcal{N}_{j_0-1}$ . Analogously, for  $j \geq j_0$ , the set of  $k$  with potentially non-zero  $\beta_{jk}$  is denoted with  $\mathcal{N}_j$ .

Let us now define two privacy mechanisms that will turn out to be convenient for the purposes of this paper. It will be sufficient to consider  $j_0, j_1 \in \mathbb{N}$  from now on. Note that since the wavelets coefficients  $\alpha'_{j_0k}$  and  $\beta'_{jk}$  are in any case zero for  $k \notin \mathcal{N}_{j_0-1}$  and  $k \notin \mathcal{N}_j$ , respectively, we do not have to consider any privatisation of these quantities.

#### First privacy mechanism

For  $i \in \llbracket 1, n \rrbracket$ ,  $j \in \llbracket j_0 - 1, j_1 \rrbracket$ , define

$$Z_{ijk} = \begin{cases} \varphi_{j_0k}(X_i) + \sigma_{j_0-1} W_{i,j_0-1,k} & \text{if } j = j_0 - 1, k \in \mathcal{N}_{j_0-1}, \\ \psi_{jk}(X_i) + \tilde{\sigma}_j W_{ijk} & \text{if } j \in \llbracket j_0, j_1 \rrbracket, k \in \mathcal{N}_j, \end{cases} \tag{3.1}$$

where  $W_{ijk}$  are independent Laplace distributed random variables with parameter 1,

$$\sigma_{j_0-1} = \frac{4c_A \|\varphi\|_\infty}{\alpha} \cdot 2^{j_0/2} \quad \text{and} \quad \tilde{\sigma}_j = \frac{4c_A \|\psi\|_\infty}{\alpha} \cdot \frac{\sqrt{2}}{\sqrt{2}-1} \cdot 2^{j_1/2}$$

for  $j \in \llbracket j_0, j_1 \rrbracket$  with  $c_A = 2\lceil A \rceil + 1$ .

### Second privacy mechanism

For  $i \in \llbracket 1, n \rrbracket$ ,  $j \in \llbracket j_0 - 1, j_1 \rrbracket$ , define

$$Z_{ijk} = \begin{cases} \varphi_{j_0k}(X_i) + \sigma_{j_0-1} W_{i,j_0-1,k} & \text{if } j = j_0 - 1, k \in \mathcal{N}_{j_0-1}, \\ \psi_{jk}(X_i) + \sigma_j W_{ijk} & \text{if } j \in \llbracket j_0, j_1 \rrbracket, k \in \mathcal{N}_j, \end{cases} \tag{3.2}$$

where  $W_{ijk}$  are independent Laplace distributed random variables with parameter 1,

$$\sigma_{j_0-1} = \frac{4c_A \|\varphi\|_\infty}{\alpha} \cdot 2^{j_0/2} \quad \text{and} \quad \sigma_j = \frac{4c_A \|\psi\|_\infty}{\alpha} \cdot \frac{2\nu - 1}{\nu - 1} \cdot (j \vee 1)^\nu \cdot 2^{j/2}$$

for  $j \in \llbracket j_0, j_1 \rrbracket$  with  $c_A = 2\lceil A \rceil + 1$  and some  $\nu > 1$ .

Note that both privacy mechanisms in (3.1) and (3.2) are non-interactive because  $Z_{ijk}$  does only depend on  $X_i$  and not on  $Z_{i'jk}$  for  $i' \neq i$ . The use of the Laplace distribution is convenient to provide  $\alpha$ -differential privacy whereas the Gaussian distribution is suitable for weaker notions of privacy like approximate differential privacy and KL-divergence differential privacy [4]. The following proposition shows that both privacy mechanisms,  $Z_i = (Z_{ijk})_{j \in \llbracket j_0-1, j_1 \rrbracket, k \in \mathcal{N}_j}$  satisfy the condition of  $\alpha$ -differential privacy.

**Proposition 3.1.** *The privacy mechanisms given in (3.1) and (3.2) are local  $\alpha$ -differential private.*

**Proof.** By definition of the privacy mechanism in (3.1), the conditional density of  $Z_i$  given  $X_i = x_i$  can be written as

$$q^{Z_i|X_i=x_i}(z_i) = \prod_{k \in \mathcal{N}_{j_0-1}} \frac{1}{2\sigma_{j_0-1}} \exp\left(-\frac{|z_{i,j_0-1,k} - \varphi_{j_0k}(x_i)|}{\sigma_{j_0-1}}\right) \cdot \prod_{j=j_0}^{j_1} \prod_{k \in \mathcal{N}_j} \frac{1}{2\tilde{\sigma}_j} \exp\left(-\frac{|z_{ijk} - \psi_{jk}(x_i)|}{\tilde{\sigma}_j}\right).$$

Thus, by the reverse and the ordinary triangle inequality,

$$\begin{aligned} \frac{q^{Z_i|X_i=x_i}(z_i)}{q^{Z_i|X_i=x'_i}(z_i)} &= \prod_{k \in \mathcal{N}_{j_0-1}} \exp\left(\frac{|z_{i,j_0-1,k} - \varphi_{j_0k}(x'_i)| - |z_{i,j_0-1,k} - \varphi_{j_0k}(x_i)|}{\sigma_{j_0-1}}\right) \\ &\quad \cdot \prod_{j=j_0}^{j_1} \prod_{k \in \mathcal{N}_j} \exp\left(\frac{|z_{ijk} - \psi_{jk}(x'_i)| - |z_{ijk} - \psi_{jk}(x_i)|}{\tilde{\sigma}_j}\right) \\ &\leq \exp\left(\sum_{k \in \mathcal{N}_{j_0-1}} \frac{|\varphi_{j_0k}(x_i)| + |\varphi_{j_0k}(x'_i)|}{\sigma_{j_0-1}}\right) \cdot \exp\left(\sum_{j=j_0}^{j_1} \sum_{k \in \mathcal{N}_j} \frac{|\psi_{jk}(x_i)| + |\psi_{jk}(x'_i)|}{\tilde{\sigma}_j}\right). \end{aligned}$$

Note that for any fixed  $x_i$  and arbitrary  $j$ ,  $\psi_{jk}(x_i) \neq 0$  holds only for at most  $c_A = 2\lceil A \rceil + 1$  different  $k$ , and the same argument is valid for  $\varphi_{j_0k}(x_i)$ . Thus,

$$\begin{aligned} \frac{q^{Z_i|X_i=x_i}(z_i)}{q^{Z_i|X_i=x'_i}(z_i)} &\leq \exp\left(\frac{2 \cdot 2^{j_0/2} c_A \|\varphi\|_\infty}{\sigma_{j_0-1}}\right) \cdot \exp\left(2\|\psi\|_\infty c_A \cdot \sum_{j=j_0}^{j_1} \frac{2^{j/2}}{\tilde{\sigma}_j}\right) \\ &\leq \exp\left(\frac{\alpha}{2} + \frac{\alpha(\sqrt{2}-1)}{2\sqrt{2}} \sum_{j=j_0}^{j_1} \frac{2^{j/2}}{2^{j_1/2}}\right) \leq \exp(\alpha). \end{aligned}$$

For the privacy mechanism (3.2), analogous calculations yield for the conditional density of  $Z_i$  given  $X_i = x_i$  that

$$\begin{aligned} \frac{q^{Z_i|X_i=x_i}(z_i)}{q^{Z_i|X_i=x'_i}(z_i)} &\leq \exp\left(\frac{2 \cdot 2^{j_0/2} c_A \|\varphi\|_\infty}{\sigma_{j_0-1}}\right) \cdot \exp\left(2\|\psi\|_\infty c_A \cdot \sum_{j=j_0}^{j_1} \frac{2^{j/2}}{\sigma_j}\right) \\ &\leq \exp\left(\frac{\alpha}{2} + \frac{\alpha}{2} \cdot \frac{\nu-1}{2\nu-1} \left(2 + \sum_{j=2}^{\infty} j^{-\nu}\right)\right) \leq \exp(\alpha), \end{aligned}$$

where we used that  $\sum_{j=j_0}^{j_1} (j \vee 1)^{-\nu} \leq \sum_{j=0}^{\infty} (j \vee 1)^{-\nu}$  and  $\sum_{j=2}^{\infty} j^{-\nu} \leq (\nu-1)^{-1}$ . □

### 4. Upper bound for linear wavelet estimators

The expansion (1.4) suggests to consider estimators of the form

$$\hat{f}(x) = \sum_{k \in \mathcal{N}_{j_0-1}} \hat{\alpha}_{j_0k} \varphi_{j_0k}(x) + \sum_{j=j_0}^{j_1} \sum_{k \in \mathcal{N}_j} \hat{\beta}_{jk} \psi_{jk}(x)$$

with appropriate estimators  $\hat{\alpha}_{j_0k}$  and  $\hat{\beta}_{jk}$  of  $\alpha_{j_0k}$  and  $\beta_{jk}$ , respectively. Note that in the local private framework, estimators of the wavelet coefficients are allowed to depend on the private views  $Z_{ijk}$  only but not on the hidden  $X_i$ . For the results concerning the linear estimator in this section, it suffices to consider the case  $j_0 = 0$ . In this case, we put  $\psi_{-1,k} = \varphi_{0,k}$ ,  $\hat{\beta}_{-1,k} = \hat{\alpha}_{0k}$  and define a linear wavelet estimator through

$$\hat{f}_{\text{lin}}(x) = \sum_{j=-1}^{j_1} \sum_{k \in \mathcal{N}_j} \hat{\beta}_{jk} \psi_{jk}(x) \quad \text{with } \hat{\beta}_{jk} = \frac{1}{n} \sum_{i=1}^n Z_{ijk}.$$

Since  $\mathbb{E}W_{ijk} = 0$ , the definition of  $\hat{\beta}_{jk}$  is natural and provides an unbiased estimate of the true wavelet coefficient  $\beta_{jk}$ .

Note that in the global differential privacy setting, a curator has access to the ensemble of original data and can release a privatised version of the estimator  $\hat{f}_{\text{lin}}$  where  $\hat{\alpha}_{j_0k}$  and  $\hat{\beta}_{jk}$  are

allowed to depend on the whole non-masked sample  $X_1, \dots, X_n$ . However, in the local setting, privatisation of data must precede estimation. Therefore, in this later setup, we may only release for any  $i = 1, \dots, n$  the vector containing  $Z_{ijk}$  for all  $j \in \llbracket j_0, j_1 \rrbracket$  and  $k \in \mathcal{N}_j$ . These vectors may be calculated by  $n$  distinct entities and neither the corresponding averages  $\widehat{\alpha}_{j_0k}, \widehat{\beta}_{jk}$  nor the final private estimator  $\widehat{f}_{\text{lin}}$  can be directly released by any of these entities alone. We also underline the fact that the statistician needs to know the wavelet basis used during privatisation and that (s)he can only reconstruct the wavelet estimator by using the given privatised sample. However, this additional information does not diminish the privacy as defined in this context.

The following proposition provides an upper bound for the estimator  $\widehat{f}_{\text{lin}}$  in the so-called *matched case* when  $r = p$ . Its proof is given in Appendix B.

**Proposition 4.1.** *Assume that the father wavelet  $\varphi$  satisfies Assumption 1.2. Let  $1 \leq p < \infty$  and  $Z_{ijk}$  defined as in (3.1). Then*

$$\sup_{f \in \mathcal{D}_{pq}^s} \mathbb{E} \|\widehat{f}_{\text{lin}} - f\|_p^p \lesssim 2^{-j_1 ps} + \left(\frac{2^{2j_1}}{n\alpha^2}\right)^{p/2} + \left(\frac{2^{j_1}}{n}\right)^{p/2}. \tag{4.1}$$

In particular, choosing  $j_1 = j_1(n, \alpha)$  such that

$$2^{j_1} \asymp (n\alpha^2)^{\frac{1}{2s+2}} \wedge n^{\frac{1}{2s+1}}, \tag{4.2}$$

we obtain

$$\sup_{f \in \mathcal{D}_{pq}^s} \mathbb{E} \|\widehat{f}_{\text{lin}} - f\|_p^p \lesssim (n\alpha^2)^{-\frac{ps}{2s+2}} \vee n^{-\frac{ps}{2s+1}}. \tag{4.3}$$

The upper bound (4.3) suggests the following interpretation: As long as  $\alpha^2 \geq n^{1/(2s+1)}$ , the estimator  $\widehat{f}_{\text{lin}}$  attains the rate  $n^{-ps/(2s+1)}$  known to be optimal when the sample  $X_1, \dots, X_n$  is available. However, as soon as  $\alpha^2 < n^{1/(2s+1)}$ , this standard rate is deteriorated and the slower rate  $(n\alpha^2)^{-ps/(2s+2)}$  is attained. As in [6], the alteration of the rate in comparison to the non-private framework concerns both the effective sample size (that changes from  $n$  to  $n\alpha^2$ ) and the exponent appearing in the rate. We emphasize that the privacy mechanism (3.1) consists in a mere additive perturbation of the values  $\varphi_{j_0k}(X_i)$  and  $\psi_{jk}(X_i)$  by Laplace noise. This procedure is in notable contrast to the privacy mechanism suggested in [6] where a more complicated two-step procedure is used to release privatised coefficients in a Fourier series framework. In this Fourier series framework, the authors of [6] show that rate optimal estimation can be achieved by their two-step procedure whereas an additive perturbation of the Fourier coefficients by Laplace noise necessarily leads to non-optimal rates (see [6], Section 5.2.2). In our case, however, Proposition 4.1 together with Proposition 2.1 shows that rate optimal estimation can be achieved by means of additive Laplace perturbation only.

Although the risk bound of Proposition 4.1 is valid only in the matched case, it can be extended to the case  $r \neq p$  by means of the following proposition. Its proof is given in Appendix B.

**Corollary 4.2.** Assume that the father wavelet  $\varphi$  satisfies Assumption 1.2. Let  $1 \leq p, r < \infty$  and  $Z_{ijk}$  defined as in (3.1), and put by  $s' = s - (1/p - 1/r)_+$ . Then, choosing  $j_1$  as in (4.2) yields

$$\sup_{f \in \mathcal{D}_{pq}^s} \mathbb{E} \|\widehat{f}_{\text{lin}} - f\|_r^r \lesssim (n\alpha^2)^{-\frac{rs'}{2s'+2}} \vee n^{-\frac{rs'}{2s'+1}}.$$

Corollary 4.2 together with Proposition 2.1 shows that the estimator  $\widehat{f}_{\text{lin}}$  is of optimal order in the dense homogeneous zone where  $p \geq r$  (which is equivalent to  $s = s'$ ) and for  $\alpha$  bounded from above. In analogy to [2], it would be possible to suggest a non-linear estimation procedure depending on  $s$  that is optimal (up to logarithmic factors in some cases) in the non-homogeneous dense case and in the sparse case as well. However, in Section 5, we directly propose a non-linear estimator that is adaptive to the smoothness  $s$  of the underlying density (as well as to the other parameters  $p$  and  $q$  of the Besov space).

### 5. Upper bounds for the non-linear adaptive estimator

In this section, the privacy mechanism is given by (3.2) in Section 3. We study the theoretical properties of the *non-linear* wavelet estimators of the form

$$\widetilde{f}_n(x) = \sum_k \widehat{\alpha}_{j_0k} \varphi_{j_0k}(x) + \sum_{j=j_0}^{j_1} \sum_k \widetilde{\beta}_{jk} \psi_{jk}(x), \tag{5.1}$$

where

$$\widehat{\alpha}_{j_0k} = \frac{1}{n} \sum_{i=1}^n Z_{i,j_0-1,k} \quad \text{and} \quad \widetilde{\beta}_{jk} = \widehat{\beta}_{jk} \cdot \mathbb{1}_{\{|\widehat{\beta}_{jk}| \geq Kt\}},$$

and  $\widehat{\beta}_{jk} = \frac{1}{n} \sum_{i=1}^n Z_{ijk}$  as in Section 4 (the choice of  $t$  and the value of the numerical constant  $K$  are specified in Theorem 5.1 and its proof below). Thus, non-linearity enters only with respect to the estimation of the detail coefficients  $\beta_{jk}$ .

**Theorem 5.1.** Let the father wavelet  $\varphi$  satisfy Assumption 1.2 for some integer  $N > 0$ . Let the private views  $Z_1, \dots, Z_n$  of the sample  $X_1, \dots, X_n$  be generated with the privacy mechanism in (3.2). Consider the estimator  $\widetilde{f}_n$  defined in (5.1) with

- $j_0 \in \mathbb{N}$  such that  $2^{j_0} \asymp (n\alpha^2)^{\frac{1}{2(N+1)+2}} \wedge n^{\frac{1}{2(N+1)+1}}$ ,
- $j_1 = j'_1 \wedge j''_1$  where  $j'_1, j''_1 \in \mathbb{N}$  are such that

$$2^{j'_1} \asymp \frac{n}{\log n} \quad \text{and} \quad 2^{2j''_1} \asymp \frac{n\alpha^2}{\log(n\alpha^2)},$$

- $K = 4(\overline{L} + \sigma)$  for some  $\overline{L} > 0$  and  $\sigma = 4c_A \|\psi\|_\infty \cdot \frac{2v-1}{v-1}$  with  $v$  introduced in the definition of the second privacy mechanism,

- $t = t_{j,n,\alpha} = \gamma \cdot \frac{j^{\nu+1/2}}{\sqrt{n}} \cdot (1 \vee \frac{2j/2}{\alpha})$  for  $j \in \llbracket j_0, j_1 \rrbracket$  and some sufficiently large constant  $\gamma$  (for instance,  $\gamma \geq r(N + 1)$  works).

Then, the risk bound

$$\sup_{(s,p,q,L) \in \Theta} \sup_{f \in \mathcal{D}_{pq}^s(L,T)} \mathbb{E} \|\tilde{f}_n - f\|_r^r \lesssim (\log n)^C \cdot \mathfrak{R}_{n,\alpha}^*$$

where

$$\mathfrak{R}_{n,\alpha}^* = \begin{cases} n^{-\frac{rs}{2s+1}} \vee (n\alpha^2)^{-\frac{rs}{2s+2}} & \text{if } p > \frac{r}{s+1}, \\ n^{-\frac{rs}{2s+1}} \vee \left( \frac{n\alpha^2}{\log(n\alpha^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}} & \text{if } \frac{r}{2s+1} < p \leq \frac{r}{s+1}, \\ \left( \frac{n}{\log n} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+1}} \vee \left( \frac{n\alpha^2}{\log(n\alpha^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}} & \text{if } p \leq \frac{r}{2s+1}, \end{cases}$$

and where

$$\Theta = (1/p, N + 1) \times [1, \infty) \times [1, \infty) \times [\underline{L}, \bar{L}]$$

for some  $0 < \underline{L} \leq \bar{L} < \infty$ .

The proof of the Theorem is given in Appendix C. Note that both the privacy mechanism and the estimator in Theorem 5.1 are independent of the quantities  $s$ ,  $p$ ,  $q$ , and  $L$  (only an upper bound  $\bar{L}$  on  $L$  and an arbitrary value of  $\nu > 1$  that should be chosen close to 1 are needed in order to specify the value of the constant  $K$ ). Hence, the proposed procedure is adaptive over the collection of Besov spaces parametrized by the set  $\Theta$ . Proposition C.2 and Remark C.3 show that the value  $\bar{L}$  in the definition of  $K$  can be replaced with an upper bound on  $\|f\|_\infty$ . If such an a priori bound of  $\|f\|_\infty$  is not available, it might be replaced by some estimator of this quantity. The proposal of an appropriate estimator and its detailed analysis are outside the scope of our presentation and might be investigated in future work. The actual choice of the parameter  $\nu$  is of secondary importance for our analysis: it should be larger than 1 in order to ensure convergence of the series  $\sum_{j=2}^\infty j^{-\nu}$  in the proof of Proposition 3.1; however, it should not be too large since it appears in the final rates in the exponents of the additional logarithmic factors. We emphasize that neither the necessity nor the optimal expression of these logarithmic factors is not yet known in the framework of differential privacy.

## 6. Discussion

In this article, we have suggested refined methods for density estimation under the constraint of local  $\alpha$ -differential privacy. By the use of estimators based on wavelet expansions, we have been able to obtain adaptive procedures that obtain the minimax rate of convergence up to an additional logarithmic factor only. To the best of our knowledge, adaptation to smoothness has not been

considered in the framework of private estimation so far. Moreover, in allowing for general  $\mathbb{L}^r$ -risk and Besov ellipsoids we have widened the range of results in the privacy framework that has merely focused on  $\mathbb{L}^2$ -risk and Sobolev ellipsoids until now. We emphasize that in our minimax approach a careful coupling of a privacy mechanism associated with a corresponding estimator is provided. In the same spirit, one may produce alternative couplings and we think it would be useful to further compare the various privacy mechanisms from different perspectives.

A significant difference between our approach and the one suggested in Section 5.2.2 of [6] concerns the privacy mechanism: Whereas the procedure in [6] is built on a rather sophisticated sampling strategy aiming at the perturbation of empirical Fourier coefficients, our privacy mechanism consists in a simple Laplace perturbation of empirical wavelet coefficients. In [6] it has been observed (see the last paragraph of Section 5.2.2 in that paper) that such an approach is not feasible for the Fourier basis since it would lead to a suboptimal rate (under  $\mathbb{L}^2$ -risk) of order  $(n\alpha^2)^{-2s/(2s+3)}$  over Sobolev ellipsoids of smoothness  $s$  instead of the optimal rate  $(n\alpha^2)^{-s/(s+1)}$ . A heuristic explanation for the easier accessibility of the problem by means of wavelet bases is given by their well-known *localisation* properties in contrast to the *global* Fourier basis.

Note that wavelet methods in the non-private framework do not necessarily suffer from a logarithmic loss in the rate (see, for instance, [2] where an additional logarithmic loss only appears in the dense zone). The fact that we encounter this type of loss in our private scenario is caused by the term  $j^v$  in the definition of the privacy mechanism (3.2). The problem whether and if so, how such logarithmic losses might be circumvented remains open and provides an interesting direction for future research.

Finally, let us sketch the connection between local private estimation in the non-interactive setup and statistical inverse problems, in particular, density deconvolution: On the one hand, in density deconvolution, the statistician is given a noisy sample  $Z_1, \dots, Z_n$  where  $Z_i = X_i + \varepsilon_i$  for  $X_i \sim f$  and  $\varepsilon_i \sim q$ . Here, the density  $f$  is the quantity of interest and  $q$  an error density which is (at least in the overwhelming part of the literature) supposed to be known. In this setup, the  $Z_i$  are distributed according to the density  $g$  where

$$g(\cdot) = (K_q f)(\cdot) := \int q(\cdot - x) f(x) dx \tag{6.1}$$

is the convolution of  $f$  with the error density  $q$ . It is well known that the difficulty of reconstructing  $f$  from the sample  $Z_1, \dots, Z_n$  is linked with the *degree of ill-posedness* of the inverse problem  $g = K_q f$ . The latter can be described either in terms of the sequence  $(\lambda_k^2)$  of eigenvalues of  $K_q^* K_q$  ( $K_q^*$  denotes the adjoint operator of the linear operator  $K_q$ ) or in terms of the decay of the Fourier transform of the error density  $q$ . General inverse problems of the form  $Kf = g$  have been thoroughly investigated in [10] in the framework of a Gaussian white noise model. For Besov smooth signals  $f$  and  $|\lambda_k| \asymp k^{-\rho}$  for some  $\rho > 0$ , [10] derived adaptive rates of estimation of  $f$  proportional to

$$\begin{cases} (\log n) C n^{-\frac{rs}{2(s+\rho)+1}} & \text{if } s > \left(\rho + \frac{1}{2}\right) \left(\frac{r}{p} - 1\right), \\ (\log n) C n^{-\frac{r(s-1/p+1/r)}{2(s-1/p+\rho)+1}} & \text{if } s \leq \left(\rho + \frac{1}{2}\right) \left(\frac{r}{p} - 1\right) \text{ and } s \geq \frac{1}{p}. \end{cases}$$

On the other hand, the statistician who is given the non-interactive privatised sample  $Z_1, \dots, Z_n$  is confronted with the problem of recovering  $f$  from a sample from the mixture density

$$g(\cdot) = (Kf)(\cdot) := \int q^{Z|X=x}(\cdot) f(x) dx,$$

which is a special instance of an inverse problem and strongly resembles (6.1). In contrast to (6.1), however, the operator  $K$  is now not *a priori* given as a component of the problem but constitutes rather a part of its solution. In the local differential privacy framework, the statistician should select the operator  $K$ , corresponding to the choice of a privacy mechanism, subject to the two following constraints. First, the condition (1.1) concerning  $\alpha$ -differential privacy must hold. Second, the least possible amount of information should be smoothed out by the operator  $K$ . More precisely, denoting with  $\rho$  the degree of ill-posedness as above, the proofs of the lower bounds suggest that the least admissible value for  $\rho$  is  $1/2$ . Our privacy mechanisms, that is, our choices of  $K$  satisfy both constraints by leading to an overall estimation procedure that is nearly minimax.

We emphasize that the above interpretation of the locally differential private estimation problem does not rule out privacy mechanisms that add noise directly to the random variables  $X_1, \dots, X_n$  in principle. In this case, the probability density function  $q$  in (6.1) should satisfy the local  $\alpha$ -differential privacy condition and have smoothness equal to  $1/2$ . An explicit density  $q$  satisfying (1.1) and having smoothness  $\rho = 1/2$  does not seem trivial to find. As already mentioned, [6] have noted that adding Laplace noise directly to the observations cannot lead to an optimal procedure. Indeed, the convolution operator in this case has degree of ill-posedness corresponding to  $\rho = 2$  which yields a suboptimal rate.

## Appendix A: Proofs of Section 2

We distinguish in the sequel the dense case and the sparse case that require different explicit constructions. However, for both proofs of the lower bounds we need the existence of a function  $f_0$  with the following properties (see [9]):

- $f_0$  is a probability density,
- $J_{spq}(f_0) \leq L/2$ ,
- $\text{supp}(f_0) \subseteq [-T, T]$ ,
- $f_0 \equiv c_0 > 0$  on some interval  $[a, b]$ .

In particular,  $f_0 \in \mathcal{D}_{pq}^s(L/2, T)$ .

The main tool in the proof of the lower bounds is adapted from [6]. It allows to reduce the problem to the study of the likelihoods of the non-privatised data and quantifies the loss of information in the process.

Suppose that we are given a finite indexed family of distributions  $\{P_\nu, \nu \in \mathcal{V}\}$ . Let  $V$  denote a random variable that is uniformly distributed over  $\mathcal{V}$ . Conditionally on  $V = \nu$ , suppose we sample a random vector  $(X_1, \dots, X_n)$  according to the product measure  $P_\nu^{\otimes n} = P_\nu \otimes \dots \otimes P_\nu$ . Suppose

that we draw an  $\alpha$ -locally private sample  $Z_1, \dots, Z_n$  according to a channel  $Q$ . Conditioned on  $V = v$ ,  $(Z_1, \dots, Z_n)$  is distributed according to the measure  $M_v^n$  given by

$$M_v^n(S) := \int Q^n(S \mid x_1, \dots, x_n) dP_v^{\otimes n}(x_1, \dots, x_n) \quad \text{for } S \in \sigma(Z^n),$$

where  $Q^n(\cdot \mid x_1, \dots, x_n)$  denotes the joint distribution on  $Z^n$  of the private sample  $Z_{1:n}$  conditioned on  $X_{1:n} = x_{1:n}$ . In this setup, we have the following inequality.

**Lemma A.1 (Based on [6], Theorem 1).** *Let  $\alpha \geq 0$ . For any  $\alpha$ -locally differentially private conditional distribution  $Q$  and any  $v, v' \in \mathcal{V}$ ,  $v \neq v'$ , we have in the above setting*

$$\text{KL}(M_v^n, M_{v'}^n) + \text{KL}(M_{v'}^n, M_v^n) \leq 4n(e^\alpha - 1)^2 \text{TV}^2(P_v, P_{v'}).$$

Lemma A.1 quantifies the property that  $\alpha$ -differential privacy acts as a contraction on the space of probability measures.

### A.1. Proof of Proposition 2.1

It is sufficient to prove the lower bound for  $n$  sufficiently large (the remaining finitely many  $n$  might merely further reduce the value of the numerical constant  $C$ ). Let  $f_0$  be the function introduced above. For fixed  $j$  (the choice of which will be specified later) define  $\mathcal{I}_j$  as a maximal subset of  $\mathbb{Z}$  such that  $\text{supp}(\psi_{jk}) \subset [a, b]$  and  $\text{supp}(\psi_{jk}) \cap \text{supp}(\psi_{jk'}) = \emptyset$  if  $k, k' \in \mathcal{I}_j$  with  $k \neq k'$ . Note that  $N_j := |\mathcal{I}_j| \asymp 2^j$ . Define

$$\mathcal{F} = \left\{ f_\theta : f_\theta = f_0 + \gamma \sum_{k \in \mathcal{I}_j} \theta_k \psi_{jk} \text{ and } \theta = (\theta_k) \in \Theta := \{0, 1\}^{N_j} \right\},$$

where  $\gamma = c(n(e^\alpha - 1)^2)^{-\frac{2s+1}{2(2s+2)}}$  for  $c$  sufficiently small and  $2^j \asymp (n(e^\alpha - 1)^2)^{\frac{1}{2s+2}}$ . For  $c$  sufficiently small, it holds  $\gamma 2^{j/2} \|\psi\|_\infty \leq c_0$ , which ensures that  $f_\theta$  is non-negative for all  $\theta \in \Theta$ . One can easily check that  $\int f_\theta = 1$  and  $\text{supp}(f_\theta) \subseteq [-T, T]$  for all  $\theta \in \Theta$ . Moreover, by the definition of  $\gamma$ , the choice of  $j$  and the equivalence of norms, we have

$$\begin{aligned} \|f_\theta\|_{spq} &\leq \|f_0\|_{spq} + c_1 \gamma 2^{j(s+1/2-1/p)} \left( \sum_{k \in \mathcal{I}_j} |\theta_k|^p \right)^{1/p} \\ &\leq \frac{L}{2} + c_1 \gamma 2^{j(s+1/2)} \leq \frac{L}{2} + C c c_1 \leq L, \end{aligned}$$

where the last inequality holds for  $c$  sufficiently small. Hence,  $\mathcal{F} \subset \mathcal{D}_{pq}^s(L, T)$  and

$$\sup_{f \in \mathcal{D}_{pq}^s(L, T)} \mathbb{E}_f \|\tilde{f} - f\|_r \geq \sup_{f \in \mathcal{F}} \mathbb{E}_f \|\tilde{f} - f\|_r = \max_{\theta \in \Theta} \mathbb{E}_\theta \|\tilde{f} - f_\theta\|_r.$$

Denoting by  $\Delta_{jk}$  the support of  $\psi_{jk}$ , it holds for any estimator  $\tilde{f}$  of  $f$  that

$$\begin{aligned} \mathbb{E}_\theta \|\tilde{f} - f_\theta\|_r^r &= \mathbb{E}_\theta \int |\tilde{f}(x) - f_\theta(x)|^r dx \\ &\geq \sum_{k \in \mathcal{I}_j} \mathbb{E}_\theta \int_{\Delta_{jk}} |\tilde{f}(x) - f_\theta(x)|^r dx \\ &= \sum_{k \in \mathcal{I}_j} \mathbb{E}_\theta \int_{\Delta_{jk}} |\tilde{f}(x) - f_0(x) - \gamma \theta_k \psi_{jk}(x)|^r dx \end{aligned}$$

since  $f_\theta \equiv g_{\theta_k} := f_0 + \gamma \theta_k \psi_{jk}$  on  $\Delta_{jk}$ . Set

$$\|\tilde{f} - g_{\check{\theta}_k}\|_{r, \Delta_{jk}}^r = \int_{\Delta_{jk}} |\tilde{f}(x) - g_{\check{\theta}_k}(x)|^r dx = \int_{\Delta_{jk}} |\tilde{f}(x) - f_0(x) - \gamma \check{\theta}_k \psi_{jk}(x)|^r dx,$$

and  $\check{\theta}_k = \operatorname{argmin}_{\theta \in \{0,1\}} \|\tilde{f} - g_\theta\|_{r, \Delta_{jk}}$ . It follows from the triangle inequality that

$$\begin{aligned} 2\|\tilde{f} - g_{\check{\theta}_k}\|_{r, \Delta_{jk}} &\geq \|\tilde{f} - g_{\theta_k}\|_{r, \Delta_{jk}} + \|\tilde{f} - g_{\check{\theta}_k}\|_{r, \Delta_{jk}} \\ &\geq \|g_{\theta_k} - g_{\check{\theta}_k}\|_{r, \Delta_{jk}} \\ &= \gamma |\theta_k - \check{\theta}_k| \|\psi_{jk}\|_r. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}_\theta \|\tilde{f} - f_\theta\|_r^r &\geq \frac{\gamma^r}{2^r} \sum_{k \in \mathcal{I}_j} \mathbb{E}_\theta [|\check{\theta}_k - \theta_k|^r] \|\psi_{jk}\|_r^r \\ &= \frac{\gamma^r}{2^r} \|\psi_{j1}\|_r^r \cdot \mathbb{E}_\theta [d_H(\check{\theta}, \theta)], \end{aligned}$$

where  $d_H$  denotes the Hamming distance. Therefore,

$$\sup_{f \in \mathcal{D}_{pq}^s(L, T)} \mathbb{E}_f \|\tilde{f} - f\|_r^r \geq \max_{\theta \in \Theta} \mathbb{E}_\theta \|\tilde{f} - f_\theta\|_r^r \geq \frac{\gamma^r}{2^r} \|\psi_{j1}\|_r^r \cdot \inf_{\tilde{\theta}} \max_{\theta \in \Theta} \mathbb{E}_\theta [d_H(\tilde{\theta}, \theta)].$$

In order to apply Lemma A.2, we need to bound the Kullback–Leibler divergence between two different distributions  $M_\theta^n$  and  $M_{\theta'}^n$  of the private sample  $(Z_1, \dots, Z_n)$  resulting from the sample  $X_1, \dots, X_n$  if, for all  $i \in \llbracket 1, n \rrbracket$ ,  $X_i$  is distributed according to  $f_\theta, f_{\theta'}$  with  $d_H(\theta, \theta') = 1$ . We write  $X_i \sim \mathbb{P}_\theta$  if  $X_i$  has density  $f_\theta$ . Using Lemma A.1 we obtain for any channel providing local  $\alpha$ -differential privacy that

$$\text{KL}(M_\theta^n, M_{\theta'}^n) \leq 4(e^\alpha - 1)^2 n \text{TV}^2(\mathbb{P}_\theta, \mathbb{P}_{\theta'}).$$

Now, since  $d_H(\theta, \theta') = 1$  and  $\theta, \theta' \in \Theta$ , there exists  $k_0 \in \mathcal{I}_j$  such that

$$\begin{aligned} \text{TV}(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) &= \frac{1}{2} \int |f_\theta(x) - f_{\theta'}(x)| \, dx = \frac{1}{2} \int \left| \gamma \sum_{k \in \mathcal{I}_j} (\theta_k - \theta'_k) \psi_{jk}(x) \right| \, dx \\ &= \frac{\gamma}{2} \int |\psi_{jk_0}(x)| \, dx = \frac{1}{2} 2^{-j/2} \gamma \|\psi\|_1, \end{aligned}$$

which implies that

$$\text{KL}(M_\theta^n, M_{\theta'}^n) \leq (e^\alpha - 1)^2 \|\psi\|_1^2 n 2^{-j} \gamma^2 \leq c^2 \|\psi\|_1^2 C < \infty.$$

Applying Lemma A.2 from the appendix with  $N = N_j \gtrsim 2^j$  implies

$$\begin{aligned} \sup_{f \in \mathcal{D}_{pq}^s(L, T)} \mathbb{E}_f \|\tilde{f} - f\|_r^r &\gtrsim \frac{\gamma^r}{2^r} 2^{j(r/2-1)} \|\psi\|_r^r \cdot 2^j \\ &\gtrsim (n(e^\alpha - 1))^2 \cdot \frac{\gamma^r}{2^{rs+2}}. \end{aligned}$$

This implies the statement of the proposition since  $\tilde{f}$  and the channel distribution were arbitrary.

### A.2. Proof of Proposition 2.2

We consider  $f_0, \psi, \mathcal{I}_j$  and  $N_j$  as in the proof of Proposition 2.1, but consider now the set

$$\mathcal{F} = \{f_k = f_0 + \gamma \cdot \psi_{jk}, k \in \mathcal{I}_j\} \cup \{f_0\},$$

where  $j$  is chosen such that  $2^j \simeq \left(\frac{n(e^\alpha-1)^2}{\log(n(e^\alpha-1)^2)}\right)^{\frac{1}{2(s+1-1/p)}}$  and  $\gamma = c2^{-j(s+1/2-1/p)}$  for  $c$  sufficiently small. Let us first check that this choice of  $j$  and  $\gamma$  guarantees that  $\mathcal{F} \subset \mathcal{D}_{pq}^s(L, T)$ . First, we have  $f_0 \in \mathcal{D}_{pq}^s(L, T)$  and one can easily check that  $\int f_k = 1$  and  $\text{supp}(f_k) \subseteq [-T, T]$  for all  $k \in \mathcal{I}_j$ . Then, for any  $k \in \mathcal{I}_j$ , we have on  $[a, b]$  that

$$f_k \geq c_0 - \gamma \|\psi_{jk}\|_\infty \geq c_0 - c2^{-j(s+1/2-1/p)} 2^{j/2} \|\psi\|_\infty \geq c_0 - c \|\psi\|_\infty \geq 0$$

for  $c$  sufficiently small, and outside of  $[a, b]$  it holds  $f_k = f_0 \geq 0$ . Furthermore, for any  $k \in \mathcal{I}_j$ ,

$$\|f_k\|_{spq} \leq \|f_0\|_{spq} + \gamma \|\psi_{jk}\|_{spq} \leq L/2 + c2^{-j(s+1/2-1/p)} \|\psi_{jk}\|_{spq} \leq L/2 + cc_1 \leq L$$

for  $c$  sufficiently small. Hence,  $\mathcal{F} \subset \mathcal{D}_{pq}^s(L, T)$  and

$$\sup_{f \in \mathcal{D}_{pq}^s(L, T)} \mathbb{E}_f \|\tilde{f} - f\|_r^r \geq \sup_{f \in \mathcal{F}} \mathbb{E}_f \|\tilde{f} - f\|_r^r.$$

Now, we show that for  $k, k' \in \mathcal{I}_j, k \neq k'$ , the hypotheses  $f_k$  and  $f_{k'}$ , as well as the hypotheses  $f_k$  and  $f_0$ , are sufficiently separated in the sense of Lemma A.3. For such  $k, k'$  we have:

$$\begin{aligned} \|f_k - f_{k'}\|_r^r &\geq \|f_k - f_0\|_r^r = \gamma^r 2^{rj(1/2-1/r)} \cdot \|\psi\|_r^r = c^r 2^{-rj(s+1/2-1/p)} 2^{rj(1/2-1/r)} \cdot \|\psi\|_r^r \\ &= c^r \|\psi\|_r^r 2^{-jr(s-1/p+1/r)} \\ &\geq C \left( \frac{\log(n(e^\alpha - 1)^2)}{n(e^\alpha - 1)^2} \right)^{r \cdot \frac{s-1/p+1/r}{2(s+1-1/p)}}. \end{aligned}$$

For  $k \in \{0\} \cup \mathcal{I}_j$ , let  $M_k^n$  be the distribution of the private sample  $(Z_1, \dots, Z_n)$  resulting from the sample  $X_1, \dots, X_n$  if for all  $i \in \llbracket 1, n \rrbracket$   $X_i$  is distributed according to  $f_k$ . For all  $k \in \mathcal{I}_j$  we have  $M_k^n \ll M_0^n$ . It remains to bound the quantity  $\frac{1}{N_j} \sum_{k \in \mathcal{I}_j} \text{KL}(M_k^n, M_0^n)$ . We write  $X_i \sim \mathbb{P}_k$  if  $X_i$  has density  $f_k, k \in \{0\} \cup \mathcal{I}_j$ . First consider the total variation distance between  $\mathbb{P}_k$  and  $\mathbb{P}_0$  for  $k \in \mathcal{I}_j$ :

$$\begin{aligned} \text{TV}(\mathbb{P}_k, \mathbb{P}_0) &= \frac{1}{2} \int |f_k - f_0| = \frac{\gamma}{2} \int |\psi_{jk}| = \frac{\gamma}{2} 2^{-j/2} \|\psi\|_1 \\ &= \frac{c}{2} \|\psi\|_1 2^{-j(s-1/p+1)}, \end{aligned}$$

and thus

$$\text{TV}^2(\mathbb{P}_k, \mathbb{P}_0) \leq \frac{c^2}{4} \|\psi\|_1^2 C \cdot \frac{\log(n(e^\alpha - 1)^2)}{n(e^\alpha - 1)^2}.$$

Applying Lemma A.1 gives

$$\frac{1}{N_j} \sum_{k \in \mathcal{I}_j} \text{KL}(M_k^n, M_0^n) \leq c^2 \|\psi\|_1^2 C \cdot \log(n(e^\alpha - 1)^2). \tag{A.1}$$

Now,  $\log(N_j) \geq \log(C2^j)$  and

$$\log(C2^j) > \frac{\log(n(e^\alpha - 1)^2)}{2(s-1/p+1)} (1 + o(1)) \geq \frac{1}{2} \frac{\log(n(e^\alpha - 1)^2)}{2(s-1/p+1)}$$

for  $n$  sufficiently large, say  $n \geq n_0$ . Putting this estimate into (A.1) yields

$$\frac{1}{N_j} \sum_{k \in \mathcal{I}_j} \text{KL}(M_k^n, M_0^n) \leq C \log(N_j)$$

for  $n \geq n_0$  and  $C < 1/8$  for  $c$  sufficiently small. We can then apply Lemma A.3, which yields for  $n \geq n_0$  that

$$\sup_{f \in \mathcal{D}_{pq}^s} \mathbb{E}_f \|\tilde{f} - f\|_r^r \geq C \left( \frac{\log(n(e^\alpha - 1)^2)}{n(e^\alpha - 1)^2} \right)^{r \cdot \frac{s-1/p+1/r}{2(s-1/p)+2}}.$$

The statement of the proposition follows since both the estimator  $\tilde{f}$  and the privacy mechanism considered were arbitrary.

### A.3. Further auxiliary results for the lower bound proofs

The following lemma is a Kullback–Leibler version of Assouad’s lemma. As above, we denote by  $d_H$  the Hamming distance, that is,  $d_H(\theta, \theta') = \sum_{i=1}^d \mathbb{1}_{\{\theta_i \neq \theta'_i\}}$  for  $\theta, \theta' \in \mathbb{R}^d$ .

**Lemma A.2 ([13], p. 118, Theorem 2.12).** *Denote with  $\Theta = \{0, 1\}^N$  the set of all binary sequences of length  $N$ . Let  $\{\mathbb{P}_\theta : \theta \in \Theta\}$  be a set of  $2^N$  probability measures on some measurable space  $(\mathcal{X}, \mathcal{A})$  and let the corresponding expectations be denoted by  $\mathbb{E}_\theta$ . Then*

$$\inf_{\tilde{\theta}} \max_{\theta \in \Theta} \mathbb{E}_\theta [d_H(\theta, \tilde{\theta})] \geq \frac{N}{2} \max\{\exp(-\beta)/2, 1 - \sqrt{\beta/2}\}$$

provided that  $\text{KL}(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) \leq \beta < \infty$  for all  $\theta, \theta' \in \Theta$  with  $d_H(\theta, \theta') = 1$ .

For the lower bound in the sparse case we need the following lemma taken from [13].

**Lemma A.3 ([13], p. 101, Theorem 2.7).** *Assume that  $M \geq 1$  and suppose that  $\Theta$  contains elements  $\theta_0, \theta_1, \dots, \theta_M$  such that:*

- (i)  $d(\theta_j, \theta_k) \geq 2\Psi > 0$ , for all  $0 \leq j < k \leq M$ ,
- (ii)  $\mathbb{P}_j \ll \mathbb{P}_0$ , for all  $j = 1, \dots, M$ , and

$$\frac{1}{M} \sum_{j=1}^M \text{KL}(\mathbb{P}_j, \mathbb{P}_0) \leq \beta \log M$$

with  $0 < \beta < 1/8$  and  $\mathbb{P}_j = \mathbb{P}_{\theta_j}$ ,  $j = 0, 1, \dots, M$ . Then

$$\inf_{\tilde{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta (d^r(\tilde{\theta}, \theta)) \geq c(\beta)\Psi^r.$$

## Appendix B: Proofs of Section 4

### B.1. Proof of Proposition 4.1

We give the proof for  $p > 2$  only, which is based on Statement (ii) from Lemma B.1. The proof for  $1 \leq p \leq 2$  follows similarly using (i) instead. We decompose the risk of the estimator  $\widehat{f}_{\text{lin}}$  into approximation and stochastic error:

$$\mathbb{E} \|\widehat{f}_{\text{lin}} - f\|_p^p \leq 2^{p-1} \{ \mathbb{E} \|\widehat{f}_{\text{lin}} - \mathbb{E}[\widehat{f}_{\text{lin}}]\|_p^p + \|\mathbb{E}[\widehat{f}_{\text{lin}}] - f\|_p^p \}.$$

The approximation term can be dealt with exactly as in the case of non-private data (see [9], p. 130),

$$\|\mathbb{E}[\widehat{f}_{\text{lin}}] - f\|_p^p \leq C2^{-spj_1},$$

and it remains to consider the stochastic term. Putting  $\beta'_{-1,k} = \frac{1}{n} \sum_{i=1}^n \varphi(X_i - k)$  and  $\beta'_{jk} = \frac{1}{n} \sum_{i=1}^n \psi_{jk}(X_i)$ , we have

$$\begin{aligned} \widehat{f}_{\text{lin}} - \mathbb{E}[\widehat{f}_{\text{lin}}] &= \sum_{j=-1}^{j_1} \sum_{k \in \mathcal{N}_j} \beta'_{jk} \psi_{jk}(x) - \sum_{j=-1}^{j_1} \sum_{k \in \mathcal{N}_j} \beta_{jk} \psi_{jk}(x) \\ &\quad + \sum_{k \in \mathcal{N}_{-1}} \left( \frac{1}{n} \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k}(x) + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \frac{1}{n} \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk}(x), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \widehat{f}_{\text{lin}} - \mathbb{E}[\widehat{f}_{\text{lin}}] &= \frac{1}{n} \sum_{i=1}^n K_{j_1+1}(x, X_i) - \mathbb{E}[K_{j_1+1}(x, X_1)] \\ &\quad + \sum_{k \in \mathcal{N}_{-1}} \left( \frac{1}{n} \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k}(x) + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \frac{1}{n} \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk}(x), \end{aligned}$$

where  $K_j(x, y) = 2^j \sum_k \varphi(2^j x - k) \tilde{\varphi}(2^j y - k)$ . We further decompose

$$\begin{aligned} &\mathbb{E} \left\| \widehat{f}_{\text{lin}} - \mathbb{E}[\widehat{f}_{\text{lin}}] \right\|_p^p \\ &\leq 2^{p-1} \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n K_{j_1+1}(\cdot, X_i) - \mathbb{E}(K_{j_1+1}(\cdot, X_1)) \right\|_p^p \\ &\quad + 2^{p-1} \mathbb{E} \left\| \sum_{k \in \mathcal{N}_{-1}} \left( \frac{1}{n} \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k} + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \frac{1}{n} \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk} \right\|_p^p. \end{aligned}$$

The first term on the right-hand side is analysed as in the non-private setup (see [9], p. 130) leading to the bound

$$\mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n K_{j_1+1}(\cdot, X_i) - \mathbb{E}[K_{j_1+1}(\cdot, X_1)] \right\|_p^p \leq C \left( \frac{2^{j_1}}{n} \right)^{p/2}. \tag{B.1}$$

For the remaining term, we have by Tonelli's theorem

$$\begin{aligned} & \mathbb{E} \int \left| \sum_{k \in \mathcal{N}_{-1}} \left( \frac{1}{n} \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k}(x) + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \frac{1}{n} \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk}(x) \right|^p dx \\ &= \frac{1}{n^p} \int_{\Delta} \mathbb{E} \left| \sum_{k \in \mathcal{N}_{-1}} \left( \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k}(x) + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk}(x) \right|^p dx, \end{aligned}$$

where  $\Delta$  is some compact set the length of which depends on  $A$  and  $T$  only. The expectation inside the integral is bounded from above by Rosenthal's inequality (statement (ii) of Lemma B.1):

$$\begin{aligned} & \mathbb{E} \left| \sum_{k \in \mathcal{N}_{-1}} \left( \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k}(x) + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk}(x) \right|^p \\ & \lesssim \sum_{k \in \mathcal{N}_{-1}} \sum_{i=1}^n \mathbb{E} |\sigma_{-1} W_{i,-1,k} \psi_{-1,k}(x)|^p + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \sum_{i=1}^n \mathbb{E} |\tilde{\sigma}_j W_{ijk} \psi_{jk}(x)|^p \\ & \quad + \left( \sum_{k \in \mathcal{N}_{-1}} \sum_{i=1}^n \mathbb{E} |\sigma_{-1} W_{i,-1,k} \psi_{-1,k}(x)|^2 + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \sum_{i=1}^n \mathbb{E} |\tilde{\sigma}_j W_{ijk} \psi_{jk}(x)|^2 \right)^{p/2} \\ & = n \sum_{k \in \mathcal{N}_{-1}} \sigma_{-1}^p |\psi_{-1,k}(x)|^p \mathbb{E} |W_{1,-1,k}|^p + n \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \tilde{\sigma}_j^p |\psi_{jk}(x)|^p \mathbb{E} |W_{1jk}|^p \\ & \quad + n^{p/2} \left( \sum_{k \in \mathcal{N}_{-1}} \sigma_{-1}^2 |\psi_{-1,k}(x)|^2 \mathbb{E} |W_{1,-1,k}|^2 + \sum_{j=0}^{j_1} \tilde{\sigma}_j^2 \sum_{k \in \mathcal{N}_j} |\psi_{jk}(x)|^2 \mathbb{E} |W_{1jk}|^2 \right)^{p/2} \\ & \asymp n \sum_{k \in \mathcal{N}_{-1}} \sigma_{-1}^p |\psi_{-1,k}(x)|^p + n \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \tilde{\sigma}_j^p |\psi_{jk}(x)|^p \\ & \quad + n^{p/2} \left( \sum_{k \in \mathcal{N}_{-1}} \sigma_{-1}^2 |\psi_{-1,k}(x)|^2 + \sum_{j=0}^{j_1} \tilde{\sigma}_j^2 \sum_{k \in \mathcal{N}_j} |\psi_{jk}(x)|^2 \right)^{p/2} \\ & \asymp n \sum_{k \in \mathcal{N}_{-1}} |\psi_{-1,k}(x)|^p \frac{1}{\alpha^p} + n \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} 2^{j_1 p/2} |\psi_{jk}(x)|^p \frac{1}{\alpha^p} \\ & \quad + n^{p/2} \left( \sum_{k \in \mathcal{N}_{-1}} |\psi_{-1,k}(x)|^2 \frac{1}{\alpha^2} + \sum_{j=0}^{j_1} 2^{j_1} \sum_{k \in \mathcal{N}_j} |\psi_{jk}(x)|^2 \frac{1}{\alpha^2} \right)^{p/2}. \end{aligned}$$

Recall the definition of  $\psi_{jk}$  and noting that due to the boundedness of the support of the wavelet parents  $\varphi$  and  $\psi$  we have for any  $x$  and fixed  $j$  that  $\psi_{jk}(x) \neq 0$  only for a finite number of  $k$  that is independent of  $j$ . Thus, using the last expression we bound from above as follows

$$\begin{aligned} & \int_{\Delta} \mathbb{E} \left| \sum_{k \in \mathcal{N}_{-1}} \left( \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k}(x) + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk}(x) \right|^p dx \\ & \leq C \left( \frac{n}{\alpha^p} + n 2^{j_1 p/2} \sum_{j=0}^{j_1} \frac{2^{jp/2}}{\alpha^p} + n^{p/2} \left( \frac{1}{\alpha^2} + 2^{j_1} \sum_{j=0}^{j_1} \frac{2^j}{\alpha^2} \right)^{p/2} \right) \\ & \simeq \frac{n}{\alpha^p} + n \cdot \frac{2^{pj_1}}{\alpha^p} + \frac{n^{p/2}}{\alpha^p} + n^{p/2} \cdot \frac{2^{pj_1}}{\alpha^p}. \end{aligned}$$

Thus,

$$\begin{aligned} & \mathbb{E} \int_{\Delta} \left| \sum_{k \in \mathcal{N}_{-1}} \left( \frac{1}{n} \sum_{i=1}^n \sigma_{-1} W_{i,-1,k} \right) \psi_{-1,k}(x) + \sum_{j=0}^{j_1} \sum_{k \in \mathcal{N}_j} \left( \frac{1}{n} \sum_{i=1}^n \tilde{\sigma}_j W_{ijk} \right) \psi_{jk}(x) \right|^p dx \\ & \lesssim \frac{2^{pj_1}}{\alpha^p n^{p-1}} + \left( \frac{2^{2j_1}}{n \alpha^2} \right)^{p/2}. \end{aligned} \tag{B.2}$$

Combining (B.1) and (B.2) yields

$$\mathbb{E} \|\widehat{f}_{\text{in}} - \mathbb{E}[\widehat{f}_{\text{in}}]\|_p^p \lesssim \left( \frac{2^{2j_1}}{n \alpha^2} \right)^{p/2} + \left( \frac{2^{j_1}}{n} \right)^{p/2},$$

which proves (4.1). Choosing  $j_1 = j_1(n, \alpha)$  as in (4.2) immediately implies (4.3).

### B.2. Proof of Corollary 4.2

We distinguish between the cases  $p \geq r$  and  $p < r$ .

1. *Case:  $p > r$ .* In this case,  $s' = s$ . Let us consider the estimator  $\widehat{f}_{\text{in}}$  with  $j_1$  chosen as in Proposition 4.1. First note that there exists a constant  $\bar{C} > 0$  such that the Lebesgue measure of  $\text{supp}(\widehat{f}_{\text{in}} - f)$  is bounded from above by a constant  $\bar{C} > 0$ . Then, applying Hölder’s inequality and Proposition 4.1 yields

$$\mathbb{E} \|\widehat{f}_{\text{in}} - f\|_r^r \leq \bar{C}^{1-r/p} (\mathbb{E} \|\widehat{f}_{\text{in}} - f\|_p^p)^{r/p} \lesssim (n \alpha^2)^{-\frac{rs}{2s+2}} \vee n^{-\frac{rs}{2s+1}}.$$

2. *Case:  $p \leq r$ .* In this case,  $s' = s - 1/p + 1/r$ . Thanks to the Besov embedding it holds  $\mathcal{B}_{pq}^s \subset \mathcal{B}_{r q'}^{s'}$ , which implies  $\mathcal{D}_{pq}^s \subset \mathcal{D}_{r q'}^{s'}$ . Thus, again using the upper bound for the matched case

from Proposition 4.1,

$$\begin{aligned} \sup_{f \in \mathcal{D}_{pq}^s} \mathbb{E} \|\widehat{f}_{\text{lin}} - f\|_r^r &\leq \sup_{f \in \mathcal{D}_{r'q}^{s'}} \mathbb{E} \|\widehat{f}_{\text{lin}} - f\|_r^r \\ &\lesssim (n\alpha^2)^{-\frac{rs'}{2s'+2}} \vee n^{-\frac{rs'}{2s'+2}}, \end{aligned}$$

which is the desired bound for the case  $p \leq r$ .

### B.3. Inequalities for moments of sums of independent random variables

**Lemma B.1.** *Let  $X_1, \dots, X_n$  be independent centred random variables with  $\mathbb{E}[|X_i|^r] < \infty$ .*

(i) *If  $0 < r \leq 2$ , then*

$$\mathbb{E} \left( \left| \sum_{i=1}^n X_i \right|^r \right) \leq \left( \sum_{i=1}^n \mathbb{E}(X_i^2) \right)^{r/2}.$$

(ii) *If  $r > 2$ , then there exists a constant  $C = C(r)$  such that*

$$\mathbb{E} \left( \left| \sum_{i=1}^n X_i \right|^r \right) \leq C \left\{ \sum_{i=1}^n \mathbb{E}(|X_i|^r) + \left( \sum_{i=1}^n \mathbb{E}(X_i^2) \right)^{r/2} \right\}.$$

Inequality (i) follows directly from Jensen's inequality and concavity of  $x \mapsto x^{r/2}$  for  $0 < r \leq 2$ . For a proof of inequality (ii) we refer to [11], p. 59, Theorem 2.9.

## Appendix C: Proof of Theorem 5.1

This section is devoted to the proof of Theorem 5.1. The main reasoning is given in Section C.1 but some tedious calculations for this proof are deferred to Section C.2. Sections C.3 and C.4 contain auxiliary results used in Section C.2.

### C.1. Proof of Theorem 5.1

As in the proof of the Corollary 4.2, we note that it is sufficient to prove the result for  $p \leq r$  and one can deduce the result for  $p > r$  as in the proof of this corollary.

We consider the upper bound  $\mathbb{E} \|\widehat{f}_n - f\|_r^r \leq 3^{r-1} (\mathbb{E} \|A\|_r^r + \mathbb{E} \|B\|_r^r + \|C\|_r^r)$  where

$$A = \sum_{k \in \mathbb{Z}} (\widehat{\alpha}_{j_0 k} - \alpha_{j_0 k}) \varphi_{j_0 k}, \quad B = \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} (\widetilde{\beta}_{jk} - \beta_{jk}) \psi_{jk}$$

and

$$C = \sum_{k \in \mathbb{Z}} \alpha_{j_1 k} \varphi_{j_1 k} - f.$$

We consider the risk bounds for  $\mathbb{E}\|A\|_r^r$ ,  $\mathbb{E}\|B\|_r^r$ , and  $\|C\|_r^r$  separately.

Upper bound for the term  $\mathbb{E}\|A\|_r^r$ : Putting  $\alpha'_{j_0 k} = \frac{1}{n} \sum_{i=1}^n \varphi_{j_0 k}(X_i)$  it holds

$$\mathbb{E}\|A\|_r^r \leq 2^{r-1} \mathbb{E} \left\| \sum_{k \in \mathbb{Z}} (\alpha'_{j_0 k} - \alpha_{j_0 k}) \varphi_{j_0 k} \right\|_r^r + 2^{r-1} \mathbb{E} \left\| \sum_k \left( \frac{\sigma_{j_0-1}}{n} \sum_{i=1}^n W_{i, j_0-1, k} \right) \varphi_{j_0 k} \right\|_r^r.$$

The first term on the right-hand side is bounded by the compact support assumption on  $\varphi$  and using Lemma 1 from [2] as in the non-private case (see [2], p. 522):

$$2^{r-1} \mathbb{E} \left\| \sum_{k \in \mathbb{Z}} (\alpha'_{j_0 k} - \alpha_{j_0 k}) \varphi_{j_0 k} \right\|_r^r \leq C(r) 2^{j_0(r/2-1)} \sum_k \mathbb{E} |\alpha'_{j_0 k} - \alpha_{j_0 k}|^r \leq C(r) \left( \frac{2^{j_0}}{n} \right)^{r/2}.$$

Concerning the second term, first, by Fubini's theorem

$$\mathbb{E} \left\| \sum_k \left( \frac{\sigma_{j_0-1}}{n} \sum_{i=1}^n W_{i, j_0-1, k} \right) \varphi_{j_0 k} \right\|_r^r = \int \mathbb{E} \left| \sum_k \left( \frac{\sigma_{j_0-1}}{n} \sum_{i=1}^n W_{i, j_0-1, k} \right) \varphi_{j_0 k}(x) \right|^r dx,$$

and the integrand on the right-hand side can be bounded as follows: for  $r > 2$ ,

$$\begin{aligned} & \mathbb{E} \left| \sum_k \left( \frac{\sigma_{j_0-1}}{n} \sum_{i=1}^n W_{i, j_0-1, k} \right) \varphi_{j_0 k}(x) \right|^r \\ & \leq \frac{C(r)}{n^r} \left[ \sigma_{j_0-1}^r \sum_k |\varphi_{j_0 k}(x)|^r \sum_{i=1}^n \mathbb{E} |W_{i, j_0-1, k}|^r \right. \\ & \quad \left. + \left( \sigma_{j_0-1}^2 \sum_k |\varphi_{j_0 k}(x)|^2 \sum_{i=1}^n \mathbb{E} |W_{i, j_0-1, k}|^2 \right)^{r/2} \right] \\ & = \frac{C(r)}{n^r} \left[ \sigma_{j_0-1}^r \sum_k |\varphi_{j_0 k}(x)|^r nr! + \left( 2n\sigma_{j_0-1}^2 \sum_k |\varphi_{j_0 k}(x)|^2 \right)^{r/2} \right] \\ & \lesssim \frac{1}{n^r} \left[ 2^{rj_0} \cdot \frac{n}{\alpha^r} + (2^{2j_0} n \alpha^{-2})^{r/2} \right] \\ & = \frac{2^{rj_0}}{n^{r-1} \alpha^r} + \left( \frac{2^{2j_0}}{n \alpha^2} \right)^{r/2}, \end{aligned}$$

whereas for  $r \leq 2$ ,

$$\mathbb{E} \left| \sum_k \left( \frac{\sigma_{j_0-1}}{n} \sum_{i=1}^n W_{i,j_0-1,k} \right) \varphi_{j_0k}(x) \right|^r \lesssim \left( \frac{2^{2j_0}}{n\alpha^2} \right)^{r/2}.$$

Thus, altogether,

$$\mathbb{E} \|A\|_r^r \lesssim \left( \frac{2^{j_0}}{n} \right)^{r/2} + \left( \frac{2^{2j_0}}{n\alpha^2} \right)^{r/2}.$$

Hence, for our choice of  $j_0$  and grant to  $s < N + 1$  from Assumption 1.2, we obtain

$$\begin{aligned} \mathbb{E} \|A\|_r^r &\lesssim \left( \frac{n^{\frac{1}{2(N+1)+1}}}{n} \right)^{r/2} + \left( \frac{(n\alpha^2)^{\frac{2}{2(N+1)+2}}}{n\alpha^2} \right)^{r/2} \\ &= n^{-\frac{r(N+1)}{2(N+1)+1}} + (n\alpha^2)^{-\frac{r(N+1)}{2(N+1)+2}} \\ &\leq n^{-\frac{rs}{2s+1}} + (n\alpha^2)^{-\frac{rs}{2s+2}} \\ &\lesssim n^{-\frac{rs}{2s+1}} \vee (n\alpha^2)^{-\frac{rs}{2s+2}} \vee \left( \frac{n}{\log n} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+1}} \vee \left( \frac{n\alpha^2}{\log(n\alpha^2)} \right)^{-\frac{r(s-1/p+1/r)}{2(s-1/p)+2}}, \end{aligned}$$

and the bound on the right-hand side is the claimed rate.

*Upper bound for the term  $\mathbb{E} \|B\|_r^r$ :* We consider the sets

$$\begin{aligned} \widehat{B}_j &= \{k : |\widehat{\beta}_{jk}| > Kt_{j,n,\alpha}\}, & \widehat{S}_j &= \widehat{B}_j^c, \\ B_j &= \{k : |\beta_{jk}| > (K/2)t_{j,n,\alpha}\}, & S_j &= B_j^c, \\ B'_j &= \{k : |\beta_{jk}| > 2Kt_{j,n,\alpha}\}, & S'_j &= (B'_j)^c, \end{aligned}$$

and the decomposition

$$\begin{aligned} B &= \sum_{j=j_0}^{j_1} \sum_k (\widehat{\beta}_{jk} - \beta_{jk}) \psi_{jk} [\mathbb{1}_{\widehat{B}_j \cap S_j}(k) + \mathbb{1}_{\widehat{B}_j \cap B_j}(k)] \\ &\quad - \sum_{j=j_0}^{j_1} \sum_k \beta_{jk} \psi_{jk} [\mathbb{1}_{\widehat{S}_j \cap B'_j}(k) + \mathbb{1}_{\widehat{S}_j \cap S'_j}(k)] \\ &=: (e_{bs} + e_{bb}) - (e_{sb} + e_{ss}). \end{aligned}$$

Appropriate bounds for the four terms  $e_{bs}$ ,  $e_{bb}$ ,  $e_{sb}$ ,  $e_{ss}$  are derived in Appendix C.2.

Upper bound for the term  $\|C\|_r^r$ : In the case we consider,  $p \leq r$ , we use the embedding  $\mathcal{B}_{pq}^s \subset \mathcal{B}_{r\infty}^{s'}$ , where we recall that  $s' = s - \frac{1}{p} + \frac{1}{r}$ . Then, it holds

$$\left\| \sum_{k \in \mathbb{Z}} \alpha_{j_1 k} \varphi_{j_1 k} - f \right\|_r^r \leq C \|f\|_{spq}^r \cdot 2^{-j_1 s' r}.$$

Moreover, with our choice of  $j_1$ ,

$$\begin{aligned} 2^{-j_1 s' r} &\leq 2^{-j_1' s' r} + 2^{-j_1'' s' r} \\ &\lesssim \left( \frac{n}{\log n} \right)^{-\frac{rs'}{2(s-1/p)+1}} + \left( \frac{n\alpha^2}{\log(n\alpha^2)} \right)^{-\frac{rs'}{2(s-1/p)+2}}, \end{aligned}$$

and the sum on the right-hand side is bounded from above by the claimed rate.

### C.2. Bounds for the terms $e_{bs}$ , $e_{bb}$ , $e_{sb}$ , and $e_{ss}$

Consider the event  $A_{jk}$  defined via  $A_{jk} = \{|\widehat{\beta}_{jk} - \beta_{jk}| > K/2 \cdot t_{j,n,\alpha}\}$ . The concentration inequality (C.5) for this event as well as the bound (C.6) will be used frequently in the sequel without further reference. In the following, we bound the terms  $\mathbb{E}\|e_{bs}\|_r^r$ ,  $\mathbb{E}\|e_{bb}\|_r^r$ ,  $\mathbb{E}\|e_{sb}\|_r^r$ , and  $\mathbb{E}\|e_{ss}\|_r^r$  separately.

#### C.2.1. Bound for $e_{bs}$

By the Cauchy–Schwarz inequality and the fact that  $\widehat{B}_j \cap S_j \subset A_{jk}$ ,

$$\begin{aligned} \mathbb{E}\|e_{bs}\|_r^r &\lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_{k \in \mathcal{N}_j} \mathbb{E}[|\widehat{\beta}_{jk} - \beta_{jk}|^r \mathbb{1}_{\widehat{B}_j \cap S_j}(k)] \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_{k \in \mathcal{N}_j} (\mathbb{E}[|\widehat{\beta}_{jk} - \beta_{jk}|^{2r}])^{1/2} \cdot \mathbb{P}(|\widehat{\beta}_{jk} - \beta_{jk}| \geq K/2 \cdot t_{j,n,\alpha})^{1/2} \\ &\leq \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_{k \in \mathcal{N}_j} (n^{-r/2} \vee j^{vr/2} 2^{jr/2} (n\alpha^2)^{-r/2}) \cdot 2^{-\gamma j/2} \\ &\leq \sum_{j=j_0}^{j_1} 2^{jr/2} 2^{-\gamma j/2} (n^{-r/2} \vee j^{vr/2} 2^{jr/2} (n\alpha^2)^{-r/2}) \\ &\leq n^{-r/2} \sum_{j=j_0}^{j_1} 2^{jr/2} 2^{-\gamma j/2} + (n\alpha^2)^{-r/2} j_1^{vr/2} \sum_{j=j_0}^{j_1} 2^{jr} 2^{-\gamma j/2} \end{aligned}$$

and this term is bounded from above by the claimed rate provided that  $\gamma \geq 2r$ .

C.2.2. Bound for  $e_{sb}$

Using the relation  $\widehat{S}_j \cap B'_j \subset A_{jk}$ , we obtain

$$\begin{aligned} \mathbb{E}\|e_{sb}\|_r^r &\lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_k |\beta_{jk}|^r \cdot \mathbb{E}[\mathbb{1}_{\widehat{S}_j \cap B'_j}(k)] \\ &\lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_{k \in \mathcal{N}_j} |\beta_{jk}|^r \cdot \mathbb{P}(|\widehat{\beta}_{jk} - \beta_{jk}| \geq K \cdot t_{j,n,\alpha}) \\ &\lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1-\gamma)} \|\beta_j\|_r^r. \end{aligned}$$

In the considered case  $p \leq r$ , we exploit the embedding  $\mathcal{B}_{pq}^s \subseteq \mathcal{B}_{rq}^{s'}$  with  $s' = s - \frac{1}{p} + \frac{1}{r}$  to get the bound

$$\|\beta_j\|_r \lesssim 2^{-j(s'+\frac{1}{2}-\frac{1}{r})}.$$

Hence,

$$\mathbb{E}\|e_{sb}\|_r^r \lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1-\gamma)} 2^{-jr(s'+\frac{1}{2}-\frac{1}{r})} = \sum_{j=j_0}^{j_1} 2^{-j(\gamma+rs')} \lesssim 2^{-j_0(\gamma+rs')}$$

by the definition of  $j_0$ . Noting that

$$\begin{aligned} 2^{-j_0(\gamma+rs')} &\lesssim (n\alpha^2)^{-\frac{\gamma+rs'}{2(N+1)+2}} \vee n^{-\frac{\gamma+rs'}{2(N+1)+1}} \\ &\leq (n\alpha^2)^{-\frac{rs}{2s+2}} \vee n^{-\frac{rs}{2s+1}} \end{aligned}$$

provided that  $\gamma$  is large enough ( $\gamma \geq r(N+1)$  is sufficient), shows that  $\mathbb{E}\|e_{sb}\|_r^r$  is at most of the same order as the claimed rate.

C.2.3. Bound for  $e_{bb}$

Put  $t'_{j,n,\alpha} = \gamma j^{v+1/2} n^{-1/2}$  and  $t''_{j,n,\alpha} = \gamma j^{v+1/2} (n\alpha^2)^{-1/2} 2^{j/2}$ . Note that  $t_{j,n,\alpha} = t'_{j,n,\alpha} \vee t''_{j,n,\alpha}$ . For any  $\rho \geq 0$ , it holds

$$\begin{aligned} \mathbb{E}\|e_{bb}\|_r^r &\lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_k \mathbb{E}[|\widehat{\beta}_{jk} - \beta_{jk}|^r \mathbb{1}_{B_j \cap B_j}(k)] \\ &\lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_k (n^{-r/2} \vee j^{vr/2} 2^{jr/2} (n\alpha^2)^{-r/2}) \mathbb{1}_{B_j}(k) \end{aligned}$$

$$\begin{aligned}
 &\lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_k n^{-r/2} \mathbb{1}_{B_j}(k) \\
 &\quad + \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_k j^{vr/2} 2^{jr/2} \cdot (n\alpha^2)^{-r/2} \mathbb{1}_{B_j}(k) \\
 &\lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} (t'_{j,n,\alpha})^r \cdot \sum_k |\beta_{jk}|^\rho (t'_{j,n,\alpha})^{-\rho} \\
 &\quad + \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} (t''_{j,n,\alpha})^r \sum_k |\beta_{jk}|^\rho (t''_{j,n,\alpha})^{-\rho} \\
 &\lesssim \underbrace{\sum_{j=j_0}^{j_1} 2^{j(r/2-1)} (t'_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho}_{=:S_1} + \underbrace{\sum_{j=j_0}^{j_1} 2^{j(r/2-1)} (t''_{j,n,\alpha})^{r-\rho} \cdot \sum_k |\beta_{jk}|^\rho}_{=:S_2}. \tag{C.1}
 \end{aligned}$$

As this argument shows, one can even choose distinct values of  $\rho$  for different  $j$ , which will be used in the following calculations. Note that

$$\sum_k |\beta_{jk}|^p \lesssim 2^{-jp(s+1/2-1/p)},$$

and, if  $\rho \leq p$ , by Hölder’s inequality

$$\sum_k |\beta_{jk}|^\rho \leq 2^{j(1-\rho/p)} \left( \sum_k |\beta_{jk}|^p \right)^{\rho/p} \leq 2^{j(1-\rho/p)} 2^{-jp(s+1/2-1/p)} = 2^{-j\rho(s+1/2-1/p)}.$$

In the sequel, we consider three different cases corresponding to the three regimes in the statement of Theorem 5.1.

1. Case:  $p > r/(s + 1)$ .

- Bound for  $S_1$ : Set  $q_1 = r/(2s + 1)$  and define  $\kappa_1 \in \mathbb{N}$  such that

$$2^{\kappa_1(r/2-p/2-sp)} \asymp n^{-\frac{p-q_1}{2}}.$$

Choosing  $\rho < q_1 \leq p$  for the indices  $j \in \llbracket j_0, \kappa \rrbracket$ , we obtain (note that  $s + 1/2 = r/(2q_1)$ )

$$\begin{aligned}
 \sum_{j=j_0}^{\kappa_1} 2^{j(r/2-1)} (t'_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho &\lesssim J_1^{(r-\rho)(v+1/2)} n^{-(r-\rho)/2} \sum_{j=j_0}^{\kappa_1} 2^{j(r/2-1)} \sum_k |\beta_{jk}|^\rho \\
 &\lesssim J_1^{(r-\rho)(v+1/2)} n^{-(r-\rho)/2} \sum_{j=j_0}^{\kappa_1} 2^{j(r/2-\rho(s+1/2))}
 \end{aligned}$$

$$\begin{aligned}
&\lesssim (\log n)^C n^{-(r-\rho)/2} 2^{\kappa_1(r/2 - \frac{r\rho}{2q_1})} \\
&\asymp (\log n)^C n^{(q_1-r)/2} \\
&= (\log n)^C n^{-\frac{rs}{2s+1}}.
\end{aligned}$$

Choosing  $\rho = p$  for indices  $j \in \llbracket \kappa_1 + 1, j_1 \rrbracket$ , we obtain

$$\begin{aligned}
\sum_{j=\kappa_1+1}^{j_1} 2^{j(r/2-1)} (t'_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho &\lesssim j_1^{(r-p)(v+1/2)} n^{-(r-p)/2} \sum_{j=\kappa_1+1}^{j_1} 2^{j(r/2-sp-p/2)} \\
&\lesssim j_1^{(r-p)(v+1/2)} n^{-(r-p)/2} 2^{\kappa_1(r/2-sp-p/2)} \\
&\lesssim j_1^{(r-p)(v+1/2)} n^{-(r-q_1)/2} \\
&\asymp (\log n)^C \cdot n^{-\frac{rs}{2s+1}}.
\end{aligned}$$

- Bound for  $S_2$ : Set  $q_2 = r/(s+1)$  and define  $\kappa_2 \in \mathbb{N}$  such that

$$2^{\kappa_2(r-p-sp)} \asymp (n\alpha^2)^{-\frac{p-q_2}{2}}.$$

Choosing  $\rho < q_2 \leq p$  for the indices  $j \in \llbracket j_0, \kappa_2 \rrbracket$ , we obtain (note that  $s+1 = r/q_2$ )

$$\begin{aligned}
\sum_{j=j_0}^{\kappa_2} 2^{j(r/2-1)} (t''_{j,n,\alpha})^{r-\rho} \cdot \sum_k |\beta_{jk}|^\rho &\lesssim j_1^{(r-p)(v+1/2)} (n\alpha^2)^{-(r-\rho)/2} \sum_{j=j_0}^{\kappa_2} 2^{j(r-\rho(s+1))} \\
&\lesssim (\log n)^C (n\alpha^2)^{-(r-\rho)/2} 2^{\kappa_2(r - \frac{r\rho}{q_2})} \\
&\asymp (\log n)^C (n\alpha^2)^{(q_2-r)/2} \\
&= (\log n)^C (n\alpha^2)^{-\frac{rs}{2s+2}}.
\end{aligned}$$

Choosing  $\rho = p$  for indices  $j \in \llbracket \kappa_2 + 1, j_1 \rrbracket$ , we obtain

$$\begin{aligned}
\sum_{j=\kappa_2+1}^{j_1} 2^{j(r/2-1)} (t''_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho &\lesssim j_1^{(r-p)(v+1/2)} (n\alpha^2)^{-(r-p)/2} \sum_{j=\kappa_2+1}^{j_1} 2^{j(r-sp-p)} \\
&\lesssim j_1^{(r-p)(v+1/2)} (n\alpha^2)^{-(r-p)/2} 2^{\kappa_2(r-sp-p)} \\
&\lesssim (\log n)^C \cdot (n\alpha^2)^{(q_2-r)/2} \\
&= (\log n)^C \cdot (n\alpha^2)^{-\frac{rs}{2s+2}}.
\end{aligned}$$

2. Case:  $p \in (r/(2s + 1), r/(s + 1)]$ .

- Bound for  $S_1$ : The sum  $S_1$  can be dealt with as in the first case, since the choices of  $q_1$  and  $\kappa_1$  from that case are still legitimated for  $p \in (r/(2s + 1), r/(s + 1)]$ .
- Bound for  $S_2$ : In order to bound  $S_2$  in the second case, define  $q_2$  and  $\kappa_2$  via the relations

$$q_2 = r \cdot \frac{1 - 1/r}{s - 1/p + 1} \quad \text{and} \quad 2^{\kappa_2} \asymp (n\alpha^2)^{\frac{q_2}{2(r-1)}}.$$

To deal with the sum over  $j \in \llbracket j_0, \kappa_2 \rrbracket$ , we take  $\rho = p$  and obtain

$$\begin{aligned} \sum_{j=j_0}^{\kappa_2} 2^{j(r/2-1)} (t''_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho &\lesssim j_1^{(v+1/2)(r-p)} (n\alpha^2)^{(r-p)/2} \sum_{j=j_0}^{\kappa_2} 2^{j(r-sp-p)} \\ &\lesssim (\log n)^C (n\alpha^2)^{-(r-p)/2} \sum_{j=j_0}^{\kappa_2} 2^{j(r-1)(1-p/q_2)} \\ &\lesssim (\log n)^C (n\alpha^2)^{-(r-p)/2} 2^{\kappa_2(r-1)(1-p/q_2)} \\ &\asymp (\log n)^C (n\alpha^2)^{(q_2-r)/2} \\ &= (\log n)^C (n\alpha^2)^{-\frac{rs'}{2(s-1/p)+2}}. \end{aligned}$$

For the sum over indices  $j \in \llbracket \kappa_2 + 1, j_1 \rrbracket$ , we choose  $\rho > q_2 > p$ , and obtain by monotony of  $\ell^\omega$ -norms in  $\omega$ , and putting  $s'' = s + 1/2 - 1/p$  that

$$\begin{aligned} \sum_{j=\kappa_2+1}^{j_1} 2^{j(r/2-1)} (t''_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho &\lesssim j_1^{(v+1/2)(r-\rho)} (n\alpha^2)^{-\frac{r-\rho}{2}} \sum_{j=\kappa_2+1}^{j_1} 2^{j(r-\rho/2-1-\rho s'')} \\ &\lesssim (\log n)^C (n\alpha^2)^{-\frac{r-\rho}{2}} \sum_{j=\kappa_2+1}^{j_1} 2^{j(r-1-\rho/2-\rho s'')} \\ &= (\log n)^C (n\alpha^2)^{-\frac{r-\rho}{2}} \sum_{j=\kappa_2+1}^{j_1} 2^{j(r-1)(1-\rho/q_2)} \\ &\lesssim (\log n)^C (n\alpha^2)^{-\frac{r-\rho}{2}} 2^{\kappa_2(r-1)(1-\rho/q_2)} \\ &\asymp (\log n)^C (n\alpha^2)^{\frac{q_2-r}{2}} \\ &= (\log n)^C (n\alpha^2)^{-\frac{rs'}{2(s-1/p)+2}}. \end{aligned}$$

3. Case:  $p \leq r/(2s + 1)$ .

- Bound for  $S_1$ : Put

$$q_1 = r \cdot \frac{1/2 - 1/r}{s + 1/2 - 1/p},$$

and choose  $\kappa_1 \in \mathbb{N}$  such that

$$2^{\kappa_1} \asymp n^{\frac{1}{2} \frac{q_1}{r/2-1}}.$$

Then, taking  $\rho = p$  for the indices  $j \in \llbracket j_0, \kappa_1 \rrbracket$  in the first sum in (C.1), we obtain

$$\begin{aligned} \sum_{j=j_0}^{\kappa_1} 2^{j(r/2-1)} (t'_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho &\leq j_1^{(v+1/2)(r-\rho)} n^{-(r-p)/2} \sum_{j=j_0}^{\kappa_1} 2^{j(r/2-sp-p/2)} \\ &\lesssim (\log n)^C n^{-(r-p)/2} \sum_{j=j_0}^{\kappa_1} 2^{j(r/2-1)(1-p/q_1)} \\ &\lesssim (\log n)^C n^{-(r-p)/2} 2^{\kappa_1(r/2-1)(1-p/q_1)} \\ &= (\log n)^C n^{\frac{q_1-r}{2}} \\ &= (\log n)^C n^{-\frac{rs'}{2(s-1/p)+1}}. \end{aligned}$$

For the sum over indices  $j \in \llbracket \kappa_1 + 1, j_1 \rrbracket$ , we choose  $\rho > q_1 > p$ , and obtain by monotony of  $\ell^\omega$ -norms in  $\omega$  and putting  $s'' = s + 1/2 - 1/p$  that

$$\begin{aligned} \sum_{j=\kappa_1+1}^{j_1} 2^{j(r/2-1)} (t'_{j,n,\alpha})^{r-\rho} \sum_k |\beta_{jk}|^\rho &\lesssim j_1^{(v+1/2)(r-\rho)} n^{-\frac{r-\rho}{2}} \sum_{j=\kappa_1+1}^{j_1} 2^{j(r/2-1-\rho s'')} \\ &\lesssim (\log n)^C n^{-\frac{r-\rho}{2}} \sum_{j=\kappa_1+1}^{j_1} 2^{j(r/2-1-\rho s'')} \\ &= (\log n)^C n^{-\frac{r-\rho}{2}} \sum_{j=\kappa_1+1}^{j_1} 2^{j(r/2-1)(1-\rho/q_1)} \\ &\lesssim (\log n)^C n^{-\frac{r-\rho}{2}} 2^{\kappa_1(r/2-1)(1-\rho/q_1)} \\ &\asymp (\log n)^C n^{\frac{q_1-r}{2}} \\ &= (\log n)^C n^{-\frac{rs'}{2(s-1/p)+1}}. \end{aligned}$$

- Bound for  $S_2$ :  $S_2$  can be dealt with exactly as in the second case.

C.2.4. Bound for  $e_{ss}$

For any  $0 \leq \rho \leq r$

$$\begin{aligned} \mathbb{E}\|e_{ss}\|_r^r &\lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)} \sum_k |\beta_{jk}|^r \mathbb{E}[\mathbb{1}_{\widehat{S}_j \cap S'_j}(k)] \\ &\lesssim \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)} ((t'_{j,n,\alpha})^{r-\rho} \vee (t''_{j,n,\alpha})^{r-\rho}) \sum_k |\beta_{jk}|^\rho. \end{aligned}$$

This term can be bounded from above by the right-hand side of (C.1), and we conclude in the same way as for the term  $e_{bb}$ .

C.3. A concentration inequality for the  $\widehat{\beta}_{jk}$

For our proof, we need concentration inequalities for the events

$$A_{jk} := \left\{ |\widehat{\beta}_{jk} - \beta_{jk}| \geq (K/2) \frac{j^{\nu+1/2}}{\sqrt{n}} \left( 1 \vee \frac{2^{j/2}}{\alpha} \right) \right\}$$

for  $K > 0$ , where  $j \in \llbracket j_0, j_1 \rrbracket$  and  $k \in \mathcal{N}_j$ . Let recall the two-sided Bernstein's inequality (cf. [1] Theorem 2.10).

**Theorem C.1.** *Let  $Y_1, \dots, Y_n$  be independent real valued random variables. Assume that there exist some positive numbers  $v$  and  $c$  such that*

$$\sum_{i=1}^n \mathbb{E}[Y_i^2] \leq v, \tag{C.2}$$

and for all integers  $m \geq 3$

$$\sum_{i=1}^n \mathbb{E}[|Y_i|^m] \leq \frac{m!}{2} v c^{m-2}. \tag{C.3}$$

Let  $S = \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])$ , then for every positive  $x$

$$\mathbb{P}[|S| \geq \sqrt{2vx} + cx] \leq 2 \exp(-x). \tag{C.4}$$

Using this inequality, we can prove the following result.

**Proposition C.2.** *For all  $j \in \llbracket j_0, j_1 \rrbracket$  satisfying  $j \leq n$ , for all  $k \in \mathcal{N}_j$ , and for all  $\gamma \geq 1$  we have*

$$\mathbb{P}\left( |\widehat{\beta}_{jk} - \beta_{jk}| \geq 4(\bar{c} + \sigma) \gamma \frac{j^{\nu+1/2}}{\sqrt{n}} \left( 1 \vee \frac{2^{j/2}}{\alpha} \right) \right) \leq 2^{-\gamma j}, \tag{C.5}$$

where  $\bar{c}$  is an upper bound for  $\sup_{f \in \mathcal{D}_{pq}^s(L, T)} \|f\|_\infty$  and  $\sigma = 4c_A \|\psi\|_\infty (2\nu - 1)/(\nu - 1)$  appears in the privacy mechanism (3.2).

**Remark C.3.** By Equation (15) in [2], the choice  $\bar{c} = L$  is admissible for  $f \in \mathcal{D}_{pq}^s(L, T)$ .

**Proof of Proposition C.2.** We will apply Bernstein's inequality to the random variables  $\{Z_{ijk}\}_{i=1, \dots, n}$ . Using that  $\psi_{jk}(X_i)$  and  $W_{ijk}$  are independent and that  $\mathbb{E}[W_{ijk}] = 0$ , we get for all  $i \in \llbracket 1, n \rrbracket$

$$\begin{aligned} \mathbb{E}[Z_{ijk}^2] &= \mathbb{E}[\psi_{jk}(X_i)^2] + \sigma_j^2 \mathbb{E}[W_{ijk}^2] + 2\sigma_j \mathbb{E}[\psi_{jk}(X_i)W_{ijk}] \\ &= \mathbb{E}[\psi_{jk}(X_i)^2] + \sigma_j^2 \mathbb{E}[W_{ijk}^2] + 2\sigma_j \mathbb{E}[\psi_{jk}(X_i)] \mathbb{E}[W_{ijk}] \\ &= \mathbb{E}[\psi_{jk}(X_i)^2] + \sigma_j^2 \mathbb{E}[W_{ijk}^2] \\ &\leq \bar{c} + 2\sigma_j^2 \\ &\leq 2(\bar{c} + \sigma_j)^2, \end{aligned}$$

where  $\bar{c} > 0$  depends on  $L$  is such that  $\|f\|_\infty \leq \bar{c}$  for all  $f$  in  $\mathcal{B}_{pq}^s(L)$  with  $s > \frac{1}{p}$ . Let  $m \geq 3$  be an integer. Using again that  $\psi_{jk}(X_i)$  and  $W_{ijk}$  are independent we get for all  $i \in \llbracket 1, n \rrbracket$

$$\begin{aligned} \mathbb{E}[|Z_{ijk}|^m] &\leq \mathbb{E}[ (|\psi_{jk}(X_i)| + \sigma_j |W_{ijk}|)^m ] \\ &= \mathbb{E} \left[ \sum_{l=0}^m \binom{m}{l} \sigma_j^l |W_{ijk}|^l |\psi_{jk}(X_i)|^{m-l} \right] \\ &= \sum_{l=0}^m \binom{m}{l} \sigma_j^l \mathbb{E}[|W_{ijk}|^l] \mathbb{E}[|\psi_{jk}(X_i)|^{m-l}] \\ &= \sum_{l=0}^m \binom{m}{l} \sigma_j^l \mathbb{E}[|\psi_{jk}(X_i)|^{m-l}] l! \\ &\leq m! \sum_{l=0}^m \binom{m}{l} \sigma_j^l (\bar{c})^{m-l} \\ &= m!(\bar{c} + \sigma_j)^m. \end{aligned}$$

Conditions (C.2) and (C.3) are thus satisfied with  $v = 2n(\bar{c} + \sigma_j)^2$  and  $c = \bar{c} + \sigma_j$ , and according to Bernstein's inequality (C.4) we have for all  $x > 0$

$$\mathbb{P} \left( |\widehat{\beta}_{jk} - \beta_{jk}| \geq (\bar{c} + \sigma_j) \left( 2\sqrt{\frac{x}{n}} + \frac{x}{n} \right) \right) \leq 2 \exp(-x).$$

Note that we have for all  $j \in \llbracket j_0, j_1 \rrbracket$ ,

$$\bar{c} + \sigma_j = \bar{c} + \sigma j^v \frac{2^{j/2}}{\alpha} \leq (\bar{c} + \sigma) j^v \left( 1 \vee \frac{2^{j/2}}{\alpha} \right),$$

where  $\sigma = 4c_A \|\psi\|_\infty (2v - 1)/(v - 1)$  appears in the definition of  $\sigma_j$  in (3.2). Take  $x = Cj$ ,  $C > 0$  and note that  $2\sqrt{Cj/n} + Cj/n \leq (2\sqrt{C} + C)\sqrt{j/n}$  if  $j \leq n$ . Consequently, we get for all  $C > 0$ , for all  $j \in \llbracket j_0, j_1 \rrbracket$  satisfying  $j \leq n$  and for all  $k \in \mathcal{N}_j$ ,

$$\mathbb{P} \left( |\widehat{\beta}_{jk} - \beta_{jk}| \geq (\bar{c} + \sigma)(C + 2\sqrt{C}) \frac{j^{v+1/2}}{\sqrt{n}} \left( 1 \vee \frac{2^{j/2}}{\alpha} \right) \right) \leq 2 \exp(-Cj).$$

Then, it suffices to take  $C = 2 \ln(2)\gamma$  to obtain (C.5). □

### C.4. Moment bounds and norm inequalities

Consider an arbitrary random function

$$\widehat{g} = \sum_{j=j_0}^{j_1} \sum_k \widehat{g}_{jk} \psi_{jk}.$$

Putting

$$S(t) = \sum_{j=j_0}^{j_1} 2^{jt} \leq \begin{cases} c_\gamma 2^{\max(j_1 t, j_0 t)} & \text{if } t \neq 0, \\ j_1 - j_0 & \text{if } t = 0, \end{cases}$$

it has been shown in [2] that for arbitrary  $v \in \mathbb{R}$  and  $u = r/(r - 2)$  it holds

$$\mathbb{E} \|\widehat{g}\|_r^r \leq \begin{cases} C^r \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_k \mathbb{E} |\widehat{g}_{jk}|^r & \text{if } 1 \leq r \leq 2, \\ C^r S(uv)^{(r/2-1)_+} \sum_{j=j_0}^{j_1} 2^{j(r/2-1-vr/2)} \sum_{k \in \mathbb{Z}} \mathbb{E} |\widehat{g}_{jk}|^r & \text{if } r > 2. \end{cases}$$

As in [2], adopting the formal convention  $S^0 = 1$ , it suffices to consider the second inequality for all  $r \geq 1$  (setting  $v = 0$  for the case  $r \leq 2$ ). Thus, for any  $r \geq 1$ ,

$$\mathbb{E} \|\widehat{g}\|_r^r \lesssim \sum_{j=j_0}^{j_1} 2^{j(r/2-1)} \sum_k \mathbb{E} |\widehat{g}_{jk}|^r. \tag{C.6}$$

Consider again the decomposition  $\widehat{\beta}_{jk} = \beta'_{jk} + \frac{\sigma_j}{n} \sum_{i=1}^n W_{ijk}$ . We have, for any  $m \geq 1$ ,

$$\mathbb{E} |\widehat{\beta}_{jk} - \beta_{jk}|^m \leq 2^{m-1} \mathbb{E} |\beta'_{jk} - \beta_{jk}|^m + 2^{m-1} \mathbb{E} \left| \frac{\sigma_j}{n} \sum_{i=1}^n W_{ijk} \right|^m.$$

In [2], p. 520, Equation (16) it is shown that

$$\mathbb{E}|\beta'_{jk} - \beta_{jk}|^m \leq cn^{-m/2} \quad (\text{C.7})$$

provided that  $2^j \leq n$  with a constant  $c$  depending only on  $s, p, q, L, \|\psi\|_m$  and  $m$ . In addition, by Rosenthal's inequality, it can be shown for any  $m \geq 1$  that

$$\mathbb{E}\left|\frac{\sigma_j}{n} \sum_{i=1}^n W_{ijk}\right|^m \lesssim j^{vm/2} 2^{jm/2} (n\alpha^2)^{-m/2}. \quad (\text{C.8})$$

Combining (C.7) and (C.8) yields

$$\mathbb{E}|\widehat{\beta}_{jk} - \beta_{jk}|^m \lesssim n^{-m/2} \vee j^{vm/2} 2^{jm/2} (n\alpha^2)^{-m/2}. \quad (\text{C.9})$$

## Acknowledgements

The authors gratefully acknowledge financial support from GENES. Cristina Butucea and Martin Kroll also gratefully acknowledge financial support from the French National Research Agency (ANR) under the grant Labex Ecodec (ANR-11-LABEX-0047).

## References

- [1] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. MR3185193 <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
- [2] Donoho, D.L., Johnstone, I.M., Kerkycharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974 <https://doi.org/10.1214/aos/1032894451>
- [3] Donoho, D.L. and Liu, R.C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667, 668–701. MR1105839 <https://doi.org/10.1214/aos/1176348114>
- [4] Duchi, J. and Rogers, R. (2019). Lower bounds for locally private estimation via communication complexity. In *Proceedings of the Thirty-Second Conference on Learning Theory* (A. Beygelzimer and D. Hsu, eds.). *Proceedings of Machine Learning Research* **99** 1161–1191. Phoenix, AZ: PMLR.
- [5] Duchi, J.C., Jordan, M.I. and Wainwright, M.J. (2013). Local privacy and minimax bounds: Sharp rates for probability estimation. In *Advances in Neural Information Processing Systems*.
- [6] Duchi, J.C., Jordan, M.I. and Wainwright, M.J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** 182–201. MR3803452 <https://doi.org/10.1080/01621459.2017.1389735>
- [7] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics* **40**. New York: Cambridge Univ. Press. MR3588285 <https://doi.org/10.1017/CBO9781107337862>
- [8] Hall, R., Rinaldo, A. and Wasserman, L. (2013). Differential privacy for functions and functional data. *J. Mach. Learn. Res.* **14** 703–727. MR3033345

- [9] Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics* **129**. New York: Springer. MR1618204 <https://doi.org/10.1007/978-1-4612-2222-4>
- [10] Kerkycharian, G., Petrushev, P., Picard, D. and Willer, T. (2007). Needlet algorithms for estimation in inverse problems. *Electron. J. Stat.* **1** 30–76. MR2312145 <https://doi.org/10.1214/07-EJS014>
- [11] Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Oxford Studies in Probability* **4**. New York: The Clarendon Press, Oxford University Press. MR1353441
- [12] Rohde, A. and Steinberger, L. (2018). Geometrizing rates of convergence under differential privacy constraints. Preprint. Available at [arXiv:1805.01422](https://arxiv.org/abs/1805.01422).
- [13] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. New York: Springer. MR2724359 <https://doi.org/10.1007/b13794>
- [14] Wasserman, L. and Zhou, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 <https://doi.org/10.1198/jasa.2009.tm08651>

*Received February 2019 and revised July 2019*