Privately Learning Smooth Distributions on the Hypercube by Projections

Clément Lalanne, Sébastien Gadat

WORKING PAPERS

N° 1505

February 2024

"Privately Learning Smooth Distributions on the Hypercube by Projections"

Clément Lalanne and Sébastien Gadat

Toulouse School of Economics

Privately Learning Smooth Distributions on the Hypercube by Projections

Clément Lalanne [1] **Sébastien Gadat** [1]

Abstract

Fueled by the ever-increasing need for statistics that guarantee the privacy of their training sets, this article studies the centrally-private estimation of Sobolev-smooth densities of probability over the hypercube in dimension d. The contributions of this article are two-fold : Firstly, it generalizes the one-dimensional results of (Lalanne et al., 2023b) to non-integer levels of smoothness and to a high-dimensional setting, which is important for two reasons : it is more suited for modern learning tasks, and it allows understanding the relations between privacy, dimensionality and smoothness, which is a central question with differential privacy. Secondly, this article presents a private strategy of estimation that is *data-driven* (usually referred to as *adaptive* in Statistics) in order to privately choose an estimator that achieves a good bias-variance trade-off among a finite family of private projection estimators *without prior knowledge of the ground-truth smoothness* β. This is achieved by adapting the Lepskii method for private selection, by adding a new penalization term that makes the estimation privacy-aware.

1. Introduction

Multiple experimental pieces of work have demonstrated that the unrestricted use of data for various learning tasks may cause privacy concerns (Narayanan & Shmatikov, 2006; Backstrom et al., 2007; Fredrikson et al., 2015; Dinur & Nissim, 2003; Homer et al., 2008; Loukides et al., 2010; Narayanan & Shmatikov, 2008; Sweeney, 2000; Gonon et al., 2023; Wagner & Eckhoff, 2018; Sweeney, 2002; Carlini et al., 2022). As a result, formal guarantees have been developed through *differential privacy* (Dwork et al., 2006) in order to guarantee that a quantity built on users' data does not leak more information than a given threshold. It is now considered as the gold standard in terms of privacy protection, and it is notably used by Apple (Thakurta et al., 2017), Google (Erlingsson et al., 2014; Bittau et al., 2017), Microsoft (Ding et al., 2017) and the US Census Bureau (Machanavajjhala et al., 2008; Haney et al., 2017; Abowd, 2018) among many others.

Let f be a density of probability on $[0,1]^d$ w.r.t. Lebesgue's measure, and let $X_1, \dots, X_n$ be n i.i.d. random variables with a distribution of probability that admits f as density on $[0,1]^d$. In this article, we will study the estimation of f with a quantity $\hat{f}$ that privately builds on $X_1, \dots, X_n$. The notion of privacy that is adopted in this article is the notion of *central zero-concentrated differential privacy* (Dwork & Rothblum, 2016; Bun & Steinke, 2016) (see Section 2).

This problem is statistically difficult (in the sense that it requires a lot of data) and suffers from the curse of dimensionality, which means that even without privacy considerations, one must expect an exponential number (in the dimensionality) of data points in order to solve it. Yet, its interest lies in its generality, and in its expressivity. Exploring the effects of privacy on this statistical problem is interesting on a theoretical standpoint, in order to better understand differential privacy, and for the practitioner in order to better decide between this general approach and a different one that incorporates more prior information about the distribution to estimate.

The privacy constraint naturally has a cost on the utility of estimators for this task, as with other forms of communication constraints (Barnes et al., 2019; 2020; Acharya et al., 2021a;c;d;b). An important question with differential privacy is to precisely characterize this cost, and to compare it to the incompressible error due to the estimation from samples. In this article, we quantify this trade-off when the density f has a certain level of smoothness β. Furthermore, we also explain how to privately estimate f when this smoothness level is not accessible to the practitioner, a property of the estimator referred-to as *adaptivity*.

1.1. Related work

Statistics and differential privacy. Estimating various quantities under differential privacy has received an in-

[1]Toulouse School of Ecolomics, Université Toulouse 1 Capitole, Toulouse, France. Correspondence to: Clément Lalanne <clement.lalanne@tse-fr.eu>.

Table 1. Comparison with concurrent work.

WORK	PRIVACY	DIMENSIONALITY	SMOOTHNESS	ADAPTIVITY	ESTIMATION RATE
(WASSERMAN & ZHOU, 2010)	FIXED	$d = 1$	$\beta \in \left(\frac{1}{2}, +\infty\right)$	$\times$	$\Theta\left(n^{-\frac{2\beta}{2\beta+1}}\right)$
(BARBER & DUCHI, 2014)	VARIABLE	$d \in \mathbb{N} \setminus \{0\}$	$\beta = 1$	$\times$	$\Theta\left(n^{-\frac{2}{2+d}} + (n\sqrt{\rho})^{-\frac{2}{1+d}}\right)$
(LALANNE ET AL., 2023B)	VARIABLE	$d = 1$	$\beta \in \mathbb{N} \setminus \{0\}$	$\times$	$\Theta\left(n^{-\frac{2\beta}{2\beta+1}} + (n\sqrt{\rho})^{-\frac{2\beta}{\beta+1}}\right)$
THIS WORK	VARIABLE	$d \in \mathbb{N} \setminus \{0\}$	$\beta \in (0, +\infty)$	$\surd$	$\Theta\left(n^{-\frac{2\beta}{2\beta+d}} + (n\sqrt{\rho})^{-\frac{2\beta}{\beta+d}}\right)$

In (Wasserman & Zhou, 2010), the smoothness is defined in terms of Sobolev ellipsoids. The results are presented under pure differential privacy, which implies concentrated differential privacy. In (Barber & Duchi, 2014), the smoothness is expressed in therms of Lipschitz continuity, which is usually assimilated heuristically to $\beta = 1$ in terms of Sobolev spaces. Again, the authors worked under ϵ pure differential privacy, but we took the liberty to express the results with $\rho = \epsilon^2$ in order to simplify comparisons.

creasing amount of attention during the last decade. A non-exhaustive list of references include (Wasserman & Zhou, 2010; Barber & Duchi, 2014; Diakonikolas et al., 2015; Karwa & Vadhan, 2018; Bun et al., 2019; 2021; Kamath et al., 2019; Biswas et al., 2020; Kamath et al., 2020; Acharya et al., 2021e; Lalanne, 2023; Aden-Ali et al., 2021; Cai et al., 2019; Brown et al., 2021; Cai et al., 2019; Kamath et al., 2022a; Lalanne et al., 2023c;d; Singhal, 2023; Kamath et al., 2023; 2022b). Most of those references study parametric estimation problems (i.e. estimating a quantity living in a finite-dimensional space), and observe (at a meta level) that the error of estimation can usually be expressed as a function of the sample size (n), the dimensionality (d), the level of privacy (ρ), and various quantities that characterize the regularity of the distribution class (sub-Gaussian, moments, smoothness, ...). Besides, the interesting effects of the privacy can be observed when the level of privacy (ρ) is considered as a free variable of the problem. Conversely, fixing the level of privacy usually results in rates of estimation that are the same as in the non-private case. In this article, we will consider the privacy budget as free, thus allowing to investigate some interesting trade-offs between the sample size n and the level of privacy ρ.

Unconstrained density estimation. The problem of estimating the density f is known as a *nonparametric statistical problem*. It differs from some more usual problems in the sense that the quantity to estimate (f) lives in an infinite-dimensional vector space. Specific techniques thus have to be used to estimate it. One of those techniques consists in approximating f by learning its projections on subspaces of growing dimension, and it is being used in this article. Without privacy concerns, this problem has been extensively studied for multiple decades. Without trying to be exhaustive, some important monographs include (Conover, 1999; Györfi et al., 2002; Tsybakov, 2009; Wasserman, 2006).

Density estimation with differential privacy. With differential privacy, the problem of nonparametric density estimation has been studied in a few articles. Before continuing, it is important to note that there are two main privacy attack models in the literature (depending on whether an aggregator can be trusted or not), leading to two distinct definitions of privacy : *central* differential privacy or *local* differential privacy (Evfimievski et al., 2003; Kasiviswanathan et al., 2008). This article studies the *central* model, and *local* differential privacy is outside its scope. This paragraph only covers the literature in the central model. An important early piece of work (Wasserman & Zhou, 2010) has paved the way for private non-parametric density estimation, presenting general private projection and histogram estimators. However, it only studied the case where the level of privacy ρ is kept constant, leading to the rather anticlimactic conclusion that privacy had no effect on the optimal rate of estimation for the problem at hand. In (Barber & Duchi, 2014), the authors were the first to consider ρ as a variable, and to study rates of convergence that are not privacy-agnostic. A shortcoming of their study is that they only study the estimation of Lipschitz-continuous densities, which imposes a fixed level of smoothness. More recently, (Lalanne et al., 2023b) studied the estimation of one-dimensional densities of general integer-valued Sobolev-smoothness β in a non privacy-agnostic way. This is the piece of work that is the closest to our article. However, three problems are that the authors only tackle the case of one-dimensional data, that the smoothness parameter only takes discrete values, and that their optimal estimation procedure needs to know the ground-truth smoothness β beforehand. This article solves all of these issues. A comparison between our article and this body of literature is summarized in Table 1.

Adaptive estimation. Classical frameworks for adaptive estimators build estimators of the bias of each model and select the model with the lowest estimated squared bias penalized by the variance (Akaike, 1998; Mallows, 1973;

Birgé & Massart, 1993; Barron et al., 1999; Laurent & Massart, 2000; Massart, 2007). The Lepskii method (Lepskii, 1991; 1992; 1993; Goldenhsluger & Lepski, 2007; Goldenshluger & Lepski, 2008; 2011; 2013) is similar, except that the bias is replaced by a comparative bias (within the model class), which is in itself defined as the extremum of a penalized expression. For instance, it has been studied in the context of non-private projection estimators in (Comte & Johannes, 2012; Chagny, 2013; Bertin et al., 2016). However, to the best of our knowledge, it has never been used as a privacy-aware selection mechanism in the context of central differential privacy before. A nice overview of non-private adaptive methods is presented in (Chagny, 2016).

In the literature of differential privacy, there are clever ways to perform model selection (which is here used as a synonym of adaptivity) without having to split the privacy budget (with composition theorems like Lemma 2.3) between all the models to choose from (e.g. the *Exponential Mechanism* (McSherry & Talwar, 2007), *Report Noisy Max* (Dwork & Roth, 2014) or the *Permute-And-Flip* mechanism (McKenna & Sheldon, 2020; Ding et al., 2021)). Such methods have found their way in multiple applications (Hardt et al., 2012; Blocki et al., 2016; Smith, 2011; Bhaskar et al., 2010; Liu & Talwar, 2019). Unfortunately, the adaptive estimation procedure that we adopt here does not adequately fit in any of those frameworks, and we will thus resort to using composition theorems for the model selection. A blessing of the procedure that is presented here, however, is that it only needs to select between very few models (typically of the order of a polynomial of $\log(n)$), and the degradation of utility will hence be small.

Under local privacy. For completeness, we include references for related problems in the *local* model of privacy (that we recall is different to the model of this article). In this setup, nonparametric density estimation was studied in (Duchi et al., 2013; 2016; Butucea et al., 2019; Kroll, 2021; Schluttenhofer & Johannes, 2022; Györfi & Kroll, 2023). In (Butucea et al., 2019), adaptivity is obtained by leveraging the properties of the wavelet basis that is used for the estimation. (Kroll, 2021) uses a variant of the Lepskii method for adaptivity, with the twist that the level of privacy is fixed beforehand. In (Schluttenhofer & Johannes, 2022), the authors modify the latter to be adaptive to the level of privacy as well. Our results differ from theirs by the model of privacy, and by the fact that they look at the estimation of the density at a single point whereas we look at the estimation of the density on the whole support. In particular, the rates of estimation are different. Finally, nonparametric regression was studied in (Berrett et al., 2021; Györfi & Kroll, 2022), nonparametric tests were studied in (Lam-Weil et al., 2022), and recently, nonparametric locally-private Bayesian modeling was proposed in (Beraha et al., 2023).

1.2. Contributions

The main contributions of this article could be summarized as follows :

Non-integer levels of smoothness. While the results of (Lalanne et al., 2023b) coincide with the ones presented in this article in the case of *integer-valued* β's (in dimension 1), the authors did not mention the eventuality of more fine-grained levels of smoothness. A usual trick for generalizing consists in defining the class of densities of interest in terms of their Fourier coefficient instead of their derivatives (which was the reason for the integer-valued smoothness in the first place). However, such definition does not lead to provably good lower-bounds under differential privacy. Instead, we circumvent that difficulty by considering an extended definition of Sobolev spaces via Höderian remainders (see Appendix E). This definition is a bit harder to work with, yet it has the advantage of leading to tight lower and upper-bounds for any non-negative β.

Arbitrary dimension. Concurrent work (Lalanne et al., 2023b) focuses on the estimation of the density of univariate data. In this article, we focus on the more general setup of arbitrary dimension d. In particular, the effects of dimensionality on the estimation and on the privacy-utility tradeoff are discussed in Section 5.2.

Adaptivity. The last main contribution of this article is to propose an adaptive estimator based on the Lepskii method that almost matches the performance of the optimal estimator, without prior knowledge of the smoothness of the density of interest. Adaptivity is an important property in statistics and in particular with density estimation, and to the best of our knowledge, no concurrent work for density estimation in the context of central differential privacy has presented such adaptive procedure before.

1.3. Notations

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{C}$ are respectively used to refer to the sets of natural numbers (including 0), relative numbers, real numbers, and complex numbers. In order to avoid confusion with indexes, we note $i_{\mathbb{C}}$ the canonical complex square root of -1. If $x \in \mathbb{C}$, $\bar{x}$ is used to refer to its conjugate complex number, $|z|$ to its modulus, $R(z)$ to its real part and $I(z)$ to its imaginary part. We equip $\mathbb{C}^d$ with its standard Hermitian product $\langle \cdot, \cdot \rangle$, and its associated norm is noted $\| \cdot \|$. We note $B(x, r)$ the open ball or radius r centered in x for $\| \cdot \|$. For $p \in \mathbb{N} \setminus \{0\} \cup \{+\infty\}$, $\| \cdot \|_p$ refers to the usual l_p norm for complex-valued vectors (in particular $\| \cdot \| = \| \cdot \|_2$), and to the usual L^p norm for complex-valued measurable functions. For any $k \in \mathbb{N}$, $\mathcal{C}^k(\mathcal{S})$ is used to refer to the set of functions from a space $\mathcal{S}$ to $\mathbb{C}$ that are k times continuously differentiable. $\mathcal{C}^\infty(\mathcal{S})$ is used to refer to $\cap_{k \in \mathbb{N}} \mathcal{C}^k(\mathcal{S})$. For a

multi-index $a = (a^1, \dots, a^d) \in \mathbb{N}^d$, $|a|$ is used to refer to the *length* of a, which is $\sum_{i=1}^d a^i$.

For a multi-index $a = (a^1, \dots, a^d) \in \mathbb{N}^d$ and $b \in \mathbb{C}$, we define $ba := (ba^1, \dots, ba^d)$, $b^a := b^{|a|}$, and $a^b := ((a^1)^b, \dots, (a^d)^b)$. Furthermore, if $b = (b^1, \dots, b^d) \in \mathbb{C}^d$, $b^{\times a} := (b^1)^{a^1} \times \cdots \times (b^d)^{a^d}$. Given a $k \in \mathbb{N}$, $f \in \mathcal{C}^k(\mathbb{R}^d)$, and a multi-index $a = (a^1, \dots, a^d) \in \mathbb{N}^d$ such that $|a| \leq k$, we use the notation

$$\partial^a f := \frac{\partial^{|a|} f}{\partial_1^{a^1} \partial_2^{a^2} \dots \partial_d^{a^d}} ,$$

where ∂/∂_i is used to refer to the derivation w.r.t. the i^{th} component in the canonical basis of $\mathbb{R}^d$. Alternatively, we may also note $f^{(a)}$ as a short for $\partial^a f$. $\mathcal{N}(\mu, \Sigma)$ refers to the multivariate normal distribution of mean vector μ and of covariance matrix Σ. When a distribution is used in vector calculus (e.g. $a + \mathcal{N}(\mu, \Sigma)$), the distribution has to be understood as a random variable with the desired distribution. Without further specification, it is taken independent of the rest of the stochastic quantities of the article. For a density of probability f, we may simply refer by f the probability distribution associated with it. The rest of the notations are introduced within the article directly.

2. Differential privacy

This section presents some basic background on differential privacy that will be needed for the rest of the article.

Given two datasets $\mathbf{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$ and $\mathbf{Y} = (Y_1, \dots, Y_n) \in \mathcal{X}^n$ where $\mathcal{X}$ is the feature space ($[0,1]^d$ in this article), the *Hamming* distance between $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$d_{\text{ham}}(\mathbf{X}, \mathbf{Y}) := \sum_{i=1}^n \mathbb{1}_{X_i \neq Y_i} .$$

Definition 2.1 (ρ-zCDP (Dwork & Rothblum, 2016; Bun & Steinke, 2016))**.** Given an output space $\mathcal{O}$ and $\rho \in (0, +\infty)$, a randomized mechanism (i.e. a conditional kernel of probabilities) $M : \mathcal{X}^n \to \mathcal{O}$ is ρ-zero concentrated differentially private (ρ-zCDP) if $\forall \mathbf{X}, \mathbf{Y} \in \mathcal{X}^n$, $d_{\text{ham}}(\mathbf{X}, \mathbf{Y}) \leq 1 \implies$

$$\forall 1 < \alpha < +\infty : \mathrm{D}_\alpha(M(\mathbf{X}) \| M(\mathbf{Y})) \leq \rho\alpha,$$

where $\mathrm{D}_\alpha(\cdot \| \cdot)$ denotes the Renyi divergence of level α, defined when $\alpha > 1$ as:

$$\mathrm{D}_\alpha(\mathbb{P} \| \mathbb{Q}) := \frac{1}{\alpha - 1} \log \int \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)^{\alpha - 1} d\mathbb{Q} .$$

For more details on this measure of divergence, please refer to (van Erven & Harremoës, 2014).

Lemma 2.2 (Privacy of the Gaussian mechanism (Proposition 6 with Lemma 7 in (Bun & Steinke, 2016)))**.** *Given a deterministic function h mapping a dataset to a quantity in $\mathbb{R}^{d'}$, one can define the l_2-sensitivity of h as*

$$\Delta_2 h := \sup_{\mathbf{X}, \mathbf{Y} \in \mathcal{X}^n : d_{\text{ham}}(\mathbf{X}, \mathbf{Y}) \leq 1} \left\| h(\mathbf{X}) - h(\mathbf{Y}) \right\|_2 .$$

When this quantity is finite, for any $\rho > 0$, the Gaussian mechanism defined as

$$\mathbf{X} \mapsto h(\mathbf{X}) + \frac{\Delta_2 h}{\sqrt{2\rho}} \mathcal{N}(0, I_{d'}) ,$$

is ρ-zCDP.

Lemma 2.3 (Adaptive composition of private mechanisms (Lemma 7 in (Bun & Steinke, 2016)))**.** *If the private mechanisms $M_1(\cdot), M_2(\cdot, z)$ are respectively ρ_1-zCDP and ρ_2-zCDP for any context z, then the private mechanism $M_2(\cdot, M_1(\cdot))$ is $(\rho_1 + \rho_2)$-zCDP.*

The last result can easily be generalized to a finite family of mechanisms by induction.

Finally, the last property of private mechanisms that we will use implicitly throughout this article is the data-processing inequality (or post-processing lemma in the language of differential privacy (Lemma 8 in (Bun & Steinke, 2016))), which states that if M satisfies ρ-zCDP, then for any conditional kernel of probabilities g, $g \circ M$ also satisfies ρ-zCDP.

3. (Private) projection estimators

In Statistics, when the quantity to estimate f belongs to some Hilbert space that admits a countable Hilbert basis $(\phi_k)_k$, projection estimators (Tsybakov, 2009) usually refer to estimators of the form

$$\hat{f} = \sum_k \hat{\theta}_k \phi_k ,$$

where the sum is usually truncated with a spectral cut-off of frequencies, and where $(\hat{\theta}_k)_k$ is a sequence of estimators of the true coefficients of the decomposition in the Hilbert basis. The name comes from the fact that such estimator mimics the orthogonal projection of f onto the space spanned by the first vectors of this Hilbert basis. To the best of our knowledge, their first appearance in the context of differential privacy is in (Wasserman & Zhou, 2010).

3.1. Explicit construction

We detail in Section 4 the exact functional spaces in which we assume the unknown density f to be. For now, we only need to know that f is in $L^2([0,1]^d)$ equipped with Lebesgue's measure and its standard Hermitian product

$$\langle f, g \rangle := \int_{[0,1]^d} f \bar{g} ,$$

and its standard inherited norm $\|\cdot\|$. We further fix the Hilbert basis $(\phi_k)_k$ of $L^2([0,1]^d)$ as the one associated to the following Fourier basis :

$$\forall k \in \mathbb{Z}^d, \quad \phi_k(x) := e^{i_{\mathbb{C}} 2\pi \langle k, x \rangle} = e^{i_{\mathbb{C}} 2\pi (k_1 x^1 + \cdots + k_d x^d)} . \tag{1}$$

We also define $S_k := \text{Span}\,(\phi_k)_{k \in \{-M,\ldots,M\}^d}$ the finite-dimensional vector space spanned by the ϕ_k's with every index in k lower than M, and we define f_M as the orthogonal projection of f onto S_M.

From this, we define the natural estimators of the coefficients in the Fourier basis

$$\tilde{\theta}_k := \frac{1}{n} \sum_{j=1}^n \bar{\phi}_k(X_j) = \frac{1}{n} \sum_{j=1}^n e^{-i_{\mathbb{C}} 2\pi (k_1 X_j^d + \cdots + k_d X_j^d)} , \tag{2}$$

and their noisy estimates

$$\hat{\theta}_k := \tilde{\theta}_k + \sigma_M \xi_k , \tag{3}$$

where σ_M will be a variance factor that will be tuned later on to obtain the desired level of privacy, and $(\xi_k)_{k \in \mathbb{Z}^d}$ is an *i.i.d.* complex Gaussian noise

$$\xi_k \sim (\mathcal{N}(0,1) + i_{\mathbb{C}} \mathcal{N}(0,1)) . \tag{4}$$

Finally, we define the projection estimator at rank M as

$$\tilde{f}_M := \sum_{k \in \{-M,\ldots,M\}^d} \tilde{\theta}_k \phi_k , \tag{5}$$

and its private counterpart as

$$\hat{f}_M := \sum_{k \in \{-M,\ldots,M\}^d} \hat{\theta}_k \phi_k . \tag{6}$$

3.2. General utility

The general utility of the previous estimator is given by the following result :

Lemma 3.1 (General bias-variance decomposition of $\hat{f}_M$)**.** *For any M, the estimator $\hat{f}_M$ satisfies*

$$\mathbb{E}\left(\|f - \hat{f}_M\|^2\right) \leq \underbrace{\|f - f_M\|^2}_{\text{Squared Bias}} + \underbrace{\frac{(2M+1)^d}{n}}_{\text{Sampling Variance UB}} + \underbrace{2(2M+1)^d \sigma_M^2}_{\text{Privacy Noise Variance}} .$$

Proof. See Appendix A.1. □

The bias term $\|f - f_M\|$ simply characterizes how well f is approximated in S_M. Controlling this term requires regularity assumptions on f, which is done in Section 4.

3.3. Privacy guarantees

The privacy of this estimation procedure is given by the following theorem :

Theorem 3.2 (Privacy of $\hat{f}_M$)**.** *For any M, the mechanism $(X_1, \ldots, X_n) \mapsto \hat{f}_M$ (or equivalently the mechanism that releases the computed $\hat{\theta}_k$'s for $k \in \{-M, \ldots, M\}^d$) is ρ-zCDP if $\sigma_M = \frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho}}$.*

Proof. See Appendix A.2. □

If follows from the application of the classical privacy guarantees of the Gaussian mechanism.

4. Upper-Bounds for different smoothness levels

As explained in the last section, controlling the bias term $\|f - f_M\|$ requires regularity assumptions on f. This section solves this issue by imposing Sobolev-smoothness.

4.1. Sobolev spaces in high dimension

In order to simplify the reading flow of the article, its main body only presents spaces of *integer* smoothness $\beta \in \mathbb{N} \setminus \{0\}$. All the results can be generalized to spaces of *real* smoothness $\beta > 0$. With every result that we present for an integer β in the main body of the article, we will talk about its counterpart in the case of real β, and we will link to the technical details in the appendix.

For $\beta \in \mathbb{N} \setminus \{0\}$ and $L > 0$, the isotropic Sobolev space $\mathcal{S}_L(\beta)$ is defined as the subset of $\mathcal{C}^k([0,1]^d)$ of functions of which the energy of the β^{th} derivative is bounded by L^2. Namely, $f \in \mathcal{S}_L(\beta)$ if $f \in \mathcal{C}^k([0,1]^d)$ and if

$$\sum_{\alpha \in \mathbb{N}^d : |\alpha| = \beta} \int_{[0,1]^d} |\partial^\alpha f|^2 \leq L^2 .$$

β is referred to as the smoothness parameter of the functional space $\mathcal{S}_L(\beta)$. For real β's, Sobolev spaces are defined similarly, except that non-integer derivatives are handled via Hölderian remainders (see Appendix E).

As it is often the case when dealing with Fourier coefficients, it is convenient to define the *periodic* Sobolev space $\mathcal{S}_L^p(\beta)$ by making sure that the functions and their derivatives are compatible with the typical periodicity of the Fourier basis. A function $f \in \mathcal{S}_L(\beta)$ is in $\mathcal{S}_L^p(\beta)$ if for any multi-index $\alpha \in \mathbb{N}^d$ of length at most β (strict) and any $x = (x_1, \ldots, x_d) \in [0,1]^d$, $x_i \in \{0,1\} \implies$

$$\partial^\alpha f(x) = \partial^\alpha f(x_1, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_d) . \tag{7}$$

The definition of periodic spaces is identical in the case of real-valued β's.

4.2. Implications on the bias

The Sobolev-smoothness of f imposes that its Fourier coefficient have a polynomial decrease (see Lemma B.1). This property may in turn be used to control the bias of with the following lemma :

Lemma 4.1 (Bias of $\hat{f}_M$ with Sobolev assumption)**.** *For any M, if $f \in \mathcal{S}_L^p(\beta)$, then the bias of f_M satisfies*

$$\|f - f_M\|^2 \leq \frac{L^2}{(2\pi)^{2\beta}} \frac{1}{(M+1)^{2\beta}} .$$

Proof. See Appendix B.1. □

In the case of real-valued β's, a similar control on the bias is given in Proposition E.1. Its main conceptual difference with Lemma 4.1 is that it adds a linear dependence in the dimension.

4.3. Estimation upper-bound in Sobolev spaces

Combining Lemma 4.1 and Lemma 3.1, and then optimizing over M yields the following upper-bound for the private statistical estimation in $\mathcal{S}_L^p(\beta)$:

Theorem 4.2 (Upper-bound in $\mathcal{S}_L^p(\beta)$)**.** *There exists a positive C that depends on β and L only such that, if $f \in \mathcal{S}_L^p(\beta)$, and if the values M and σ_M are tuned as*

$$M + 1 = \min\left\{ \left\lfloor \left(n/2^d\right)^{\frac{1}{2\beta+d}} \right\rfloor, \left\lfloor \left(n\sqrt{\rho}/2^d\right)^{\frac{1}{\beta+d}} \right\rfloor \right\} ,$$

and $\sigma_M = \frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho}}$, then the mechanism that returns $\hat{f}_M$ is ρ-zCDP and its error is bounded as

$$\mathbb{E}\left(\|f - \hat{f}_M\|^2\right) \leq C(M+1)^{-2\beta}.$$

Proof. See Appendix B.2. □

Lemma 4.1 and Proposition E.1 are similar enough that the only adaptation to Theorem 4.2 needed to make it work for integer-valued β's is to add that C also depends linearly on d. In particular, the scaling in n and ρ remains the same.

5. Lower-bounds and minimax optimality

This section presents lower-bounds on the private estimation in $\mathcal{S}_L^p(\beta)$, and discusses on the role of the different parameters on the difficulty of estimation.

5.1. Quantitative lower-bound

We have the following lower-bound, which generalizes the results of (Lalanne et al., 2023b) in general dimension d :

Theorem 5.1 (Lower-bound in $\mathcal{S}_L^p(\beta)$)**.** *There exist two positive constants C_1 and C_2 depending on L, β and d only such that, for any n and ρ, if $\hat{f}$ satisfies ρ-zCDP, then there exists $f \in \mathcal{S}_L^p(\beta)$ such that*

$$\mathbb{E}_f\left(\|f - \hat{f}\|^2\right) \geq C_1 \max\left\{ n^{-\frac{2\beta}{2\beta+d}}, (n\sqrt{\rho})^{-\frac{2\beta}{\beta+d}}) \right\}$$

as soon as $\min\left\{n, n\sqrt{\rho}\right\} \geq C_2$.

Proof. See Appendix C.1. □

For real-valued β's, this result also holds. Appendix E.3 discusses the adaptation of the proof of Theorem 5.1 to this more general case.

Theorem 5.1, when compared to the upper-bound given in Theorem 4.2 allows concluding that private projection estimators converge at the minimax-optimal rate

$$r_{n,\rho}(\beta) := \max\left\{ n^{-\frac{2\beta}{2\beta+d}}, (n\sqrt{\rho})^{-\frac{2\beta}{\beta+d}}) \right\}, \quad \forall \beta > 0 , \tag{8}$$

up to a multiplicative constant depending on L, β and d only. While the dependence in those quantities is easily explained in the upper-bounds, a caveat of the proof of Theorem 5.1 is that the dependence is *implicit* by construction, and that no closed-form formula may easily be obtained.

5.2. Qualitative implications

From this optimal rate of estimation, we may describe the effects of the different parameters of the privacy-utility trade-off.

- The privacy parameter ρ : The two important regimes of estimation are $\rho \gtrsim n^{-\frac{2\beta}{2\beta+d}}$ where $\gtrsim$ should be understood as "greater up to a multiplicative constant" and its complement $\rho \ll n^{-\frac{2\beta}{2\beta+d}}$. In the first regime, when the level of privacy is not too high compared to the amount of data, privacy comes at a negligible cost on the estimation. On the other hand, in the complementary regime, the utility can be arbitrarily degraded by making ρ arbitrarily small.

- The smoothness β : The higher β, the smaller the cut-off rate $n^{-\frac{2\beta}{2\beta+d}}$. In other words, the smoother the density to estimate, the more private the estimation can be with no significant degradation of utility.

- The dimensionality d : Dimensionality has the converse effect on the cut-off rate. The higher the dimension, the more data will be needed to make the effects of privacy negligible. Furthermore, the cut-off itself is affected by the curse of dimensionality.

6. Adaptivity

As seen previously, it is possible to design a private mechanism via projection estimators that is minimax optimal for the class of densities in $\mathcal{S}_L^p(\beta)$ in dimension d.

However, to do so, the optimal cut-off frequency:

$$M_{n,\rho}(\beta) := \min\left\{\left\lfloor n^{\frac{1}{2\beta+d}}\right\rfloor, \left\lfloor (n\sqrt{\rho})^{\frac{1}{\beta+d}}\right\rfloor\right\}$$

is chosen based on the knowledge on n, ρ, d and β (see Theorem 4.2). For the practitioner, the knowledge of n, ρ and d is not difficult. The knowledge of β on the other hand is a much stronger hypothesis, and it already implies a strong prior knowledge on f. This section presents a private estimation strategy that is *adaptive* in the sense that it does not require the prior knowledge of β, while almost achieving the utility of Theorem 4.2 (up to polylogarithmic factors and negligible terms).

6.1. A first candidate for private selection

At first, an idea for private adaptive estimation could be to :

(i) Compute a non-private adaptive estimator of the density with classical methods (like for instance the non-private Lepskii method (Lepskii, 1991)).

(ii) Then add noise to its Fourier coefficients in order to make it private.

However, there is a trap with this method that one must not fall into : the adaptive truncation rank $\hat{M}$ that is selected by the non-private adaptive method is a quantity that is built from the data, and it may leak user's information. It is thus not possible to simply add noise to the Fourier coefficients of the non-private Fourier coefficients up to truncation rank $\hat{M}$ with magnitude $\sigma_{\hat{M}}$ calibrated as is Theorem 3.2 and to call the result differentially private. Instead, one must add noise to the Fourier coefficients up to a truncation rank that is either fixed in advance, or that builds on the data, in which case the privacy budget of such will have to be accounted for. The problem with such method is that classical adaptive methods will only try to balance the bias and the sampling variance, but won't account for the privacy variance. In particular, when ρ is small, it is unclear if this method may have the optimal rate of convergence. In the next subsection, we detail the alternative method that we chose, that balances the bias, the sampling variance and the privacy variance *at the same time*, leading to private and adaptive near-optimal estimation.

6.2. Private and privacy-aware Lepskii method

Multiple flavors of the Lepskii method exist in the literature. Here, we present our adaptations of the main two ones to the context of private model selection. We discuss the advantages and the drawbacks of each method.

6.2.1. Risk penalization

We introduce the penalized risk (up to a useful log term):

$$r_{n,\rho}(\beta)^* := C(\log n)^a r_{n,\rho}(\beta), \tag{9}$$

where $C > 1$ and $a > 0$ are some constants independent from n and ρ that will be specified later on. We introduce a grid $\mathbb{B}$ on the possible values of β that ranges between 0 and $\log n$, defined by:

$$\mathbb{B}_n := \left\{\beta_0 = \frac{k_n\epsilon}{\log(n)}, \beta_1 = \beta_0 - \frac{\varepsilon}{\log(n)}, \right. \\ \left. \beta_2 = \beta_1 - \frac{\varepsilon}{\log(n)}, \ldots, \beta_{k_n-1} \geq 0\right\}. \tag{10}$$

The number of possible values for β in $\mathbb{B}_n$ is then denoted by k_n and $k_n = \lfloor \varepsilon^{-1}\log^2 n\rfloor$.

Our Lepskii decision rule is built upon the estimation of the smoothness parameter with the computation of a collection of estimators for several values of $\beta \in \mathbb{B}_n$ and then with a clever selection among theses values with the help of a trade-off criterion. Thanks to Lemma 2.3, to ensure a desired level of privacy of our final estimator, we introduce

$$\rho'_n = \rho\varepsilon\log^{-2} n. \tag{11}$$

We are ready to define our *adaptive* selection rule as:

$$\hat{m}_n := \inf\left\{m \geq 0 : \quad \forall \ell \geq m, \right. \\ \left. \|\hat{f}_{M_{n,\rho'_n}(\beta_m)} - \hat{f}_{M_{n,\rho'_n}(\beta_\ell)}\|_2^2 \leq r_{n,\rho'_n}(\beta_\ell)^*\right\} \tag{12}$$

For the sake of clarity, we will use the following shortcut of notations to improve the readability of our paper:

$$\hat{f}_{\hat{M}} := \hat{f}_{M_{n,\rho'_n}(\beta_{\hat{m}_n})} \quad \text{and} \quad \hat{M} := M_{n,\rho'_n}(\beta_{\hat{m}_n}).$$

We establish the following result.

Theorem 6.1. *Assume that $a \geq 1$, $C \geq 8L^2 \vee 2^{2d+9}$ and $n \geq 3$, if $\hat{f}_{\hat{M}}$ is the adaptive estimator selected with the Lepskii rule, then $\hat{f}_{\hat{M}}$ is ρ-zCDP and it satisfies the risk upper bound*

$$\mathbb{E}\left(\|\hat{f}_{\hat{M}} - f\|_2\right) \leq 2\sqrt{r_{n,\rho'_n}(\beta)^*}\exp\left(\frac{\varepsilon}{\beta+d}\right) \\ + \sqrt{8(2+d)}\varepsilon^{-3/2}\left(1+\rho_n'^{-\frac{1}{2(1+d)}}\right)\log^2 n n^{-2}$$

Proof. See Appendix D.1 □

Comments. This result shows that out of the box (i.e. without additional assumptions on f), $\hat{f}_{\hat{M}}$ nearly matches the optimal speed of estimation up to negligible terms, and by excluding the fact that we did not take the error squared, but simply the error in L^2 distance.

6.2.2. Penalization of the estimated bias

Let $\mathcal{M}$ be the collection of spectral cut-offs. The following method describes how to choose $\hat{M}$, and the associated $\hat{f}_{\hat{M}}$. We start by estimating $\tilde{f}_M$ and $\hat{f}_M$ for any $M \in \mathcal{M}$ with $\sigma_M = \frac{\sqrt{2(2M+1)^d}}{n\sqrt{\rho'}}$ where ρ' is tuned to obtain ρ-zCDP in the end as $\rho' = \frac{\rho}{|\mathcal{M}|}$.

Then for any M, we define the following estimator of the squared bias of $\hat{f}_M$:

$$B^2(K) := \max_{M' \in \mathcal{K}} \left\{ \| \operatorname{Proj}_{S_{M'}}(\hat{f}_M) - \hat{f}_{M'} \|^2 - \Lambda^{(1)}(M') \right\} \tag{13}$$

where $\Lambda^{(1)}(\cdot)$ is a penalization term that is fixed later. Then, $\hat{M}$ is chosen as the minimizer of the penalized estimated squared bias.

$$\hat{M} := \operatorname{argmin}_{M \in \mathcal{M}} \left\{ B^2(K) + \Lambda^{(2)}(K) \right\} . \tag{14}$$

Again, $\Lambda^{(2)}(\cdot)$ is a penalization term that is fixed later on.

$\hat{f}_{\hat{M}}$ satisfies the following oracle inequality :

Theorem 6.2. *When computed with $\sigma_M = \frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho/|\mathcal{M}|}}$ for any $M \in \mathcal{M}$, the mechanism that releases $\hat{f}_{\hat{M}}$ satisfies ρ-zCDP. Furthermore, there exist two absolute constants $C_1 > 0$ and $C_3 > 0$ and a quantity $C_2 > 0$ depending only on $\|f\|_\infty$[1] such that, if for any M,*

$$\Lambda^{(1)}(M) = \frac{96(2M+1)^d}{n} + \frac{96(2M+1)^{2d}}{n^2\rho/|\mathcal{M}|}$$

and

$$\Lambda^{(2)}(M) = \Lambda^{(1)}(M) + \frac{16(2M+1)^{2d}}{n^2\rho/|\mathcal{M}|} ,$$

and if $(2\max\mathcal{M}+1)^d \le n$, then $\hat{f}_{\hat{M}}$ satisfies

$$\mathbb{E}\left(\|f - \hat{f}_{\hat{M}}\|^2 \right) \le C_1 \min_{M \in \mathcal{M}} \underbrace{\left\{ \|f - f_M\|^2 + \frac{(2M+1)^d}{n} + 2(2M+1)^d \sigma_M^2 \right\}}_{\text{Best bias-variance tradeoff in } \mathcal{M} \text{ with privacy budget } \frac{\rho}{|\mathcal{M}|}} + \underbrace{\frac{C_2}{n}}_{\text{Sampling residual}} + \underbrace{\frac{C_3|\mathcal{M}|}{n^2\rho}}_{\text{Privacy residual}} . \tag{15}$$

Proof. See Appendix D.2. □

Since the bias is left uncontrolled in this result, it remains true in the case of real-valued β's.

When the collection of spectral cut-offs $\mathcal{M}$ is adequately chosen, this oracle inequality may be used to prove near-optimal convergence speed.

Theorem 6.3. *There exist a $C_1 > 0$ depending on β and L, and a $C_2 > 0$ depending on β, d and $\|f\|_\infty$ such that, if $\min\{n, n\sqrt{\rho/\log_2(n)}\} \ge C_2$, then $\hat{f}_{\hat{M}}$ computed with*

$$\mathcal{M} = \left\{ 1, 2, 4, \dots, 2^{\left\lfloor \log_2\left(\frac{n^{1/d}-1}{2} \right) \right\rfloor} \right\}$$

and all the other hyperparameters set as in Theorem 6.2 is ρ-zCDP and its utility satisfies

$$\mathbb{E}\left(\|f - \hat{f}_{\hat{M}}\|^2 \right) \le C_1 \max\left\{ n^{-\frac{2\beta}{2\beta+d}}, \left(\frac{n\sqrt{\rho}}{\sqrt{\log_2(n)}} \right)^{-\frac{2\beta}{\beta+d}} \right\} . \tag{16}$$

Proof. See Appendix D.3. □

Because of the extra dimensionality term in the control of the bias in the case of real-valued β's, this result remains true in this case if one adds that C_1 also depends on d.

Comments. Contrary tho the last procedure, this new one is near-optimal in terms of *squared* error, at the cost of the control of $\|f\|_\infty$. As explained before, this requirement comes from a technical detail in the proof, and it might be an artifact of a suboptimal analysis from us. Also, the polylogarithmic degradation only affects the privacy term.

[1] This dependence arises because of a technical argument in the proof.

Acknowledgements

Clément Lalanne acknowledges the help of Aurélien Garivier and Rémi Gribonval on previous work on which this article builds on.

References

Abowd, J. M. The us census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2867–2867, 2018.

Acharya, J., Canonne, C., Singh, A. V., and Tyagi, H. Optimal rates for nonparametric density estimation under communication constraints. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 26754–26766. Curran Associates, Inc., 2021a. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/e1021d43911ca2c1845910d84f40aeae-Paper.pdf.

Acharya, J., Canonne, C. L., Freitag, C., Sun, Z., and Tyagi, H. Inference under information constraints iii: Local privacy constraints. *IEEE Journal on Selected Areas in Information Theory*, 2(1):253–267, 2021b. doi: 10.1109/JSAIT.2021.3053569. URL https://doi.org/10.1109/JSAIT.2021.3053569.

Acharya, J., Canonne, C. L., Mayekar, P., and Tyagi, H. Information-constrained optimization: can adaptive processing of gradients help? *CoRR*, abs/2104.00979, 2021c. URL https://arxiv.org/abs/2104.00979.

Acharya, J., Canonne, C. L., Sun, Z., and Tyagi, H. Unified lower bounds for interactive high-dimensional estimation under information constraints. *CoRR*, abs/2010.06562, 2021d. URL https://arxiv.org/abs/2010.06562.

Acharya, J., Sun, Z., and Zhang, H. Differentially private Assouad, Fano, and Le Cam. In Feldman, V., Ligett, K., and Sabato, S. (eds.), *Algorithmic Learning Theory, 16-19 March 2021, Virtual Conference, Worldwide*, volume 132 of *Proceedings of Machine Learning Research*, pp. 48–78. PMLR, 2021e. URL http://proceedings.mlr.press/v132/acharya21a.html.

Aden-Ali, I., Ashtiani, H., and Kamath, G. On the sample complexity of privately learning unbounded high-dimensional gaussians. In Feldman, V., Ligett, K., and Sabato, S. (eds.), *Algorithmic Learning Theory, 16-19 March 2021, Virtual Conference, Worldwide*, volume 132 of *Proceedings of Machine Learning Research*, pp. 185–216. PMLR, 2021. URL http://proceedings.mlr.press/v132/aden-ali21a.html.

Akaike, H. *Information Theory and an Extension of the Maximum Likelihood Principle*, pp. 199–213. Springer New York, New York, NY, 1998. ISBN 978-1-4612-1694-0. doi: 10.1007/978-1-4612-1694-0_15. URL https://doi.org/10.1007/978-1-4612-1694-0_15.

Backstrom, L., Dwork, C., and Kleinberg, J. M. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In Williamson, C. L., Zurko, M. E., Patel-Schneider, P. F., and Shenoy, P. J. (eds.), *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pp. 181–190. ACM, 2007. doi: 10.1145/1242572.1242598. URL https://doi.org/10.1145/1242572.1242598.

Barber, R. F. and Duchi, J. C. Privacy and statistical risk: Formalisms and minimax bounds. *CoRR*, abs/1412.4451, 2014. URL http://arxiv.org/abs/1412.4451.

Barnes, L. P., Han, Y., and Ozgur, A. Fisher information for distributed estimation under a blackboard communication protocol. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 2704–2708, 2019. doi: 10.1109/ISIT.2019.8849821.

Barnes, L. P., Han, Y., and Özgür, A. Lower bounds for learning distributions under communication constraints via fisher information. *Journal of Machine Learning Research*, 21:Paper No. 236, 30, 2020. ISSN 1532-4435. URL https://jmlr.csail.mit.edu/papers/volume21/19-737/19-737.pdf.

Barron, A. R., Birgé, L., and Massart, P. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113:301–413, 1999. doi: https://doi.org/10.1007/s004400050210.

Beraha, M., Favaro, S., and Rao, V. Mcmc for bayesian nonparametric mixture modeling under differential privacy. *arXiv preprint arXiv:2310.09818*, 2023.

Berrett, T. B., Györfi, L., and Walk, H. Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 15 (1):2430 – 2453, 2021. doi: 10.1214/21-EJS1845. URL https://doi.org/10.1214/21-EJS1845.

Bertin, K., Lacour, C., and Rivoirard, V. Adaptive pointwise estimation of conditional density function. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 52

(2):939 – 980, 2016. doi: 10.1214/14-AIHP665. URL https://doi.org/10.1214/14-AIHP665.

Bhaskar, R., Laxman, S., Smith, A. D., and Thakurta, A. Discovering frequent patterns in sensitive data. In Rao, B., Krishnapuram, B., Tomkins, A., and Yang, Q. (eds.), *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, pp. 503–512. ACM, 2010. doi: 10.1145/1835804.1835869. URL https://doi.org/10.1145/1835804.1835869.

Birgé, L. and Massart, P. Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields*, 97:113–150, 1993. doi: https://doi.org/10.1007/BF01199316.

Biswas, S., Dong, Y., Kamath, G., and Ullman, J. R. Coinpress: Practical private mean and covariance estimation. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/a684eceee76fc522773286a895bc8436-Abstract.html.

Bittau, A., Úlfar Erlingsson, Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., and Seefeld, B. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the Symposium on Operating Systems Principles (SOSP)*, pp. 441–459, 2017. URL https://arxiv.org/abs/1710.00901.

Blocki, J., Datta, A., and Bonneau, J. Differentially private password frequency lists. In *23rd Annual Network and Distributed System Security Symposium, NDSS 2016, San Diego, California, USA, February 21-24, 2016*. The Internet Society, 2016. URL http://wp.internetsociety.org/ndss/wp-content/uploads/sites/25/2017/09/differentially-private-password-frequency-lists.pdf.

Brown, G., Gaboardi, M., Smith, A. D., Ullman, J. R., and Zakynthinou, L. Covariance-aware private mean estimation without private covariance estimation. In Ranzato, M., Beygelzimer, A., Dauphin, Y. N., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 7950–7964, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/42778ef0b5805a96f9511e20b5611fce-Abstract.html.

Bun, M. and Steinke, T. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In Hirt, M. and Smith, A. D. (eds.), *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part I*, volume 9985 of *Lecture Notes in Computer Science*, pp. 635–658, 2016. doi: 10.1007/978-3-662-53641-4_24. URL https://doi.org/10.1007/978-3-662-53641-4_24.

Bun, M., Kamath, G., Steinke, T., and Wu, Z. S. Private hypothesis selection. In Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. B., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 156–167, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/9778d5d219c5080b9a6a17bef029331c-Abstract.html.

Bun, M., Kamath, G., Steinke, T., and Wu, Z. S. Private hypothesis selection. *IEEE Trans. Inf. Theory*, 67(3):1981–2000, 2021. doi: 10.1109/TIT.2021.3049802. URL https://doi.org/10.1109/TIT.2021.3049802.

Butucea, C., Dubois, A., Kroll, M., and Saumard, A. Local differential privacy: Elbow effect in optimal density estimation and adaptation over besov ellipsoids. *CoRR*, abs/1903.01927, 2019. URL http://arxiv.org/abs/1903.01927.

Cai, T. T., Wang, Y., and Zhang, L. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *CoRR*, abs/1902.04495, 2019. URL http://arxiv.org/abs/1902.04495.

Carlini, N., Chien, S., Nasr, M., Song, S., Terzis, A., and Tramèr, F. Membership inference attacks from first principles. In *43rd IEEE Symposium on Security and Privacy, SP 2022, San Francisco, CA, USA, May 22-26, 2022*, pp. 1897–1914. IEEE, 2022. doi: 10.1109/SP46214.2022.9833649. URL https://doi.org/10.1109/SP46214.2022.9833649.

Chagny, G. Penalization *versus Goldenshluger-Lepski* strategies in warped bases regression. *ESAIM: Probability and Statistics*, 17:328–358, 2013. doi: 10.1051/ps/2011165. URL http://www.numdam.org/articles/10.1051/ps/2011165/.

Chagny, G. AN INTRODUCTION TO NONPARAMETRIC ADAPTIVE ESTIMATION. *The Graduate Journal of Mathematics*, 2016(2):105–120, December 2016. URL https://hal.science/hal-02132884.

Comte, F. *Nonparametric Estimation*. Spartacus-Idh, 2017. ISBN 978-2-36693-30-6. URL https://spartacus-idh.com/liseuse/030/.

Comte, F. and Johannes, J. Adaptive functional linear regression. *The Annals of Statistics*, 40(6):2765 – 2797, 2012. doi: 10.1214/12-AOS1050. URL https://doi.org/10.1214/12-AOS1050.

Conover, W. *Practical nonparametric statistics*. Wiley series in probability and statistics. Wiley, New York, NY [u.a.], 3. ed edition, 1999. ISBN 0471160687. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+24551600X&sourceid=fbw_bibsonomy.

Diakonikolas, I., Hardt, M., and Schmidt, L. Differentially private learning of structured discrete distributions. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pp. 2566–2574, 2015. URL https://proceedings.neurips.cc/paper/2015/hash/2b3bf3eee2475e03885a110e9acaab61-Abstract.html.

Ding, B., Kulkarni, J., and Yekhanin, S. Collecting telemetry data privately. In Guyon, I., von Luxburg, U., Bengio, S., Wallach, H. M., Fergus, R., Vishwanathan, S. V. N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 3571–3580, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/253614bbac999b38b5b60cae531c4969-Abstract.html.

Ding, Z., Kifer, D., E., S. M. S. N., Steinke, T., Wang, Y., Xiao, Y., and Zhang, D. The permute-and-flip mechanism is identical to report-noisy-max with exponential noise. *CoRR*, abs/2105.07260, 2021. URL https://arxiv.org/abs/2105.07260.

Dinur, I. and Nissim, K. Revealing information while preserving privacy. In Neven, F., Beeri, C., and Milo, T. (eds.), *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 9-12, 2003, San Diego, CA, USA*, pp. 202–210. ACM, 2003. doi: 10.1145/773153.773173. URL https://doi.org/10.1145/773153.773173.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy and statistical minimax rates. In *51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013, Allerton Park & Retreat Center, Monticello, IL, USA, October 2-4, 2013*, pp. 1592. IEEE, 2013. doi: 10.1109/Allerton.2013.6736718. URL https://doi.org/10.1109/Allerton.2013.6736718.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy, data processing inequalities, and statistical minimax rates, 2014. URL https://arxiv.org/abs/1302.3203.

Duchi, J. C., Wainwright, M. J., and Jordan, M. I. Minimax optimal procedures for locally private estimation. *CoRR*, abs/1604.02390, 2016. URL http://arxiv.org/abs/1604.02390.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

Dwork, C. and Rothblum, G. N. Concentrated differential privacy. *CoRR*, abs/1603.01887, 2016. URL http://arxiv.org/abs/1603.01887.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. D. Calibrating noise to sensitivity in private data analysis. In Halevi, S. and Rabin, T. (eds.), *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pp. 265–284. Springer, 2006. doi: 10.1007/11681878_14. URL https://doi.org/10.1007/11681878_14.

Erlingsson, Ú., Pihur, V., and Korolova, A. RAPPOR: randomized aggregatable privacy-preserving ordinal response. In Ahn, G., Yung, M., and Li, N. (eds.), *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, pp. 1054–1067. ACM, 2014. doi: 10.1145/2660267.2660348. URL https://doi.org/10.1145/2660267.2660348.

Evfimievski, A., Gehrke, J., and Srikant, R. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '03, pp. 211–222, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136706. doi: 10.1145/773153.773174. URL https://doi.org/10.1145/773153.773174.

Fredrikson, M., Jha, S., and Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Ray, I., Li, N., and Kruegel, C. (eds.), *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, October 12-16, 2015*, pp. 1322–1333. ACM, 2015. doi: 10.1145/2810103.2813677. URL https://doi.org/10.1145/2810103.2813677.

Goldenhsluger, A. and Lepski, O. Structural adaptation via l_p-norm oracle inequalities, 2007.

Goldenshluger, A. and Lepski, O. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150 – 1190, 2008. doi: 10.3150/08-BEJ144. URL https://doi.org/10.3150/08-BEJ144.

Goldenshluger, A. and Lepski, O. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39 (3):1608 – 1632, 2011. doi: 10.1214/11-AOS883. URL https://doi.org/10.1214/11-AOS883.

Goldenshluger, A. V. and Lepski, O. V. General selection rule from a family of linear estimators. *Theory of Probability & Its Applications*, 57(2):209–226, 2013. doi: 10.1137/S0040585X97985923. URL https://doi.org/10.1137/S0040585X97985923.

Gonon, A., Zheng, L., Lalanne, C., Le, Q.-T., Lauga, G., and Pouliquen, C. Sparsity in neural networks can improve their privacy, 2023.

Györfi, L. and Kroll, M. On rate optimal private regression under local differential privacy. *arXiv preprint arXiv:2206.00114*, 2022.

Györfi, L., Kohler, M., Krzyzak, A., and Walk, H. *A Distribution-Free Theory of Nonparametric Regression*. Springer series in statistics. Springer, 2002. ISBN 978-0-387-95441-7. doi: 10.1007/b97848. URL https://doi.org/10.1007/b97848.

Györfi, L. and Kroll, M. Multivariate density estimation from privatised data: universal consistency and minimax rates. *Journal of Nonparametric Statistics*, 0(0):1–23, 2023. doi: 10.1080/10485252.2022.2163634. URL https://doi.org/10.1080/10485252.2022.2163634.

Haney, S., Machanavajjhala, A., Abowd, J. M., Graham, M., Kutzbach, M., and Vilhuber, L. Utility cost of formal privacy for releasing national employer-employee statistics. In Salihoglu, S., Zhou, W., Chirkova, R., Yang, J., and Suciu, D. (eds.), *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD Conference 2017, Chicago, IL, USA, May 14-19, 2017*, pp. 1339–1354. ACM, 2017. doi: 10.1145/3035918.3035940. URL https://doi.org/10.1145/3035918.3035940.

Hardt, M., Ligett, K., and McSherry, F. A simple and practical algorithm for differentially private data release. In Bartlett, P. L., Pereira, F. C. N., Burges, C. J. C., Bottou, L., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pp. 2348–2356, 2012. URL https://proceedings.neurips.cc/paper/2012/hash/208e43f0e45c4c78cafadb83d2888cb6-Abstract.html.

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V., Stephan, D. A., Nelson, S. F., and Craig, D. W. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS Genet*, 4 (8):e1000167, 2008.

Kallenberg, O. Lectures on the coupling method (torgny lindvall). *SIAM Review*, 35(3):525–527, 1993. doi: 10.1137/1035121. URL https://doi.org/10.1137/1035121.

Kamath, G., Li, J., Singhal, V., and Ullman, J. R. Privately learning high-dimensional distributions. In Beygelzimer, A. and Hsu, D. (eds.), *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1853–1902. PMLR, 2019. URL http://proceedings.mlr.press/v99/kamath19a.html.

Kamath, G., Singhal, V., and Ullman, J. R. Private mean estimation of heavy-tailed distributions. In Abernethy, J. D. and Agarwal, S. (eds.), *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria]*, volume 125 of *Proceedings of Machine Learning Research*, pp. 2204–2235. PMLR, 2020. URL http://proceedings.mlr.press/v125/kamath20a.html.

Kamath, G., Liu, X., and Zhang, H. Improved rates for differentially private stochastic convex optimization with heavy-tailed data. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvári, C., Niu, G., and Sabato, S. (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 10633–10660. PMLR, 2022a. URL https://proceedings.mlr.press/v162/kamath22a.html.

Kamath, G., Mouzakis, A., and Singhal, V. New lower bounds for private estimation and a generalized fingerprinting lemma. In *NeurIPS*, 2022b. URL http://papers.nips.cc/paper_files/paper/2022/hash/9a6b27821896649194491f55ccf8b75-Abstract-Conference.html.

Kamath, G., Mouzakis, A., Regehr, M., Singhal, V., Steinke, T., and Ullman, J. R. A bias-variance-privacy trilemma for statistical estimation. *CoRR*, abs/2301.13334, 2023. doi: 10.48550/ARXIV.2301.13334. URL https://doi.org/10.48550/arXiv.2301.13334.

Karwa, V. and Vadhan, S. P. Finite sample differentially private confidence intervals. In Karlin, A. R. (ed.), *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, volume 94 of *LIPIcs*, pp. 44:1–44:9. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi: 10.4230/LIPIcs.ITCS.2018.44. URL https://doi.org/10.4230/LIPIcs.ITCS.2018.44.

Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. What can we learn privately? In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pp. 531–540, 2008. doi: 10.1109/FOCS.2008.27.

Klein, T. and Rio, E. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060 – 1077, 2005. doi: 10.1214/009117905000000044. URL https://doi.org/10.1214/009117905000000044.

Kroll, M. On density estimation at a fixed point under local differential privacy. *Electronic Journal of Statistics*, 15 (1):1783 – 1813, 2021. doi: 10.1214/21-EJS1830. URL https://doi.org/10.1214/21-EJS1830.

Lalanne, C. *On the tradeoffs of statistical learning with privacy*. Theses, Ecole normale supérieure de lyon - ENS LYON, October 2023. URL https://theses.hal.science/tel-04379624.

Lalanne, C., Garivier, A., and Gribonval, R. On the Statistical Complexity of Estimation and Testing under Privacy Constraints. *Transactions on Machine Learning Research Journal*, April 2023a. URL https://hal.science/hal-03794374.

Lalanne, C., Garivier, A., and Gribonval, R. About the cost of central privacy in density estimation. *Transactions on Machine Learning Research*, 2023b. ISSN 2835-8856. URL https://openreview.net/forum?id=uq29MIWvIV.

Lalanne, C., Garivier, A., and Gribonval, R. Private Statistical Estimation of Many Quantiles. In *ICML 2023 - 40th International Conference on Machine Learning*, Honolulu, United States, July 2023c. URL https://hal.science/hal-03986170.

Lalanne, C., Gastaud, C., Grislain, N., Garivier, A., and Gribonval, R. Private Quantiles Estimation in the Presence of Atoms. *Information and Inference*, August 2023d. doi: 10.1093/imaiai/iaad030. URL https://hal.science/hal-03572701.

Lam-Weil, J., Laurent, B., and Loubes, J.-M. Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. *Bernoulli*, 28(1):579–600, 2022.

Laurent, B. and Massart, P. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000. ISSN 00905364. URL http://www.jstor.org/stable/2674095.

Ledoux, M. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87, 1997. URL https://www.esaim-ps.org/articles/ps/abs/1997/01/ps-Vol1.4/ps-Vol1.4.html.

Lepskii, O. V. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability & Its Applications*, 35(3):454–466, 1991. doi: 10.1137/1135065. URL https://doi.org/10.1137/1135065.

Lepskii, O. V. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992. doi: 10.1137/1136085. URL https://doi.org/10.1137/1136085.

Lepskii, O. V. Asymptotically minimax adaptive estimation. ii. schemes without optimal adaptation: Adaptive estimators. *Theory of Probability & Its Applications*, 37(3):433–448, 1993. doi: 10.1137/1137095. URL https://doi.org/10.1137/1137095.

Liu, J. and Talwar, K. Private selection from private candidates. In Charikar, M. and Cohen, E. (eds.), *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pp. 298–309. ACM, 2019. doi: 10.1145/3313276.3316377. URL https://doi.org/10.1145/3313276.3316377.

Loukides, G., Denny, J. C., and Malin, B. A. The disclosure of diagnosis codes can breach research participants' privacy. *J. Am. Medical Informatics Assoc.*, 17(3):322–327, 2010. doi: 10.1136/jamia.2009.002725. URL https://doi.org/10.1136/jamia.2009.002725.

Machanavajjhala, A., Kifer, D., Abowd, J. M., Gehrke, J., and Vilhuber, L. Privacy: Theory meets practice on the map. In Alonso, G., Blakeley, J. A., and Chen, A. L. P. (eds.), *Proceedings of the 24th International Conference on Data Engineering, ICDE 2008, April 7-12, 2008, Cancún, Mexico*, pp. 277–286. IEEE Computer Society, 2008. doi: 10.1109/ICDE.2008.4497436. URL https://doi.org/10.1109/ICDE.2008.4497436.

Mallows, C. L. Some comments on cp. *Technometrics*, 15(4):661–675, 1973. ISSN 00401706. URL http://www.jstor.org/stable/1267380.

Massart, P. Concentration inequalities and model selection, ecole d'eté de probabilités de saint-flour xxxiii - 2003. *Lecture Notes in Mathematics -Springer-verlag-*, 1896, 01 2007. doi: 10.1007/978-3-540-48503-2.

McKenna, R. and Sheldon, D. Permute-and-flip: A new mechanism for differentially private selection. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/01e00f2f4bfcbb7505cb641066f2859b-Abstract.html.

McSherry, F. and Talwar, K. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pp. 94–103. IEEE Computer Society, 2007. doi: 10.1109/FOCS.2007.41. URL https://doi.org/10.1109/FOCS.2007.41.

Narayanan, A. and Shmatikov, V. How to break anonymity of the netflix prize dataset. *CoRR*, abs/cs/0610105, 2006. URL http://arxiv.org/abs/cs/0610105.

Narayanan, A. and Shmatikov, V. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (S&P 2008), 18-21 May 2008, Oakland, California, USA*, pp. 111–125. IEEE Computer Society, 2008. doi: 10.1109/SP.2008.33. URL https://doi.org/10.1109/SP.2008.33.

Rigollet, P. and Hütter, J.-C. High dimensional statistics. *MIT lecture notes for course 18S997*, 2015. URL https://math.mit.edu/~rigollet/PDFs/RigNotes17.pdf.

Schluttenhofer, S. and Johannes, J. Adaptive pointwise density estimation under local differential privacy, 2022.

Singhal, V. A polynomial time, pure differentially private estimator for binary product distributions. *CoRR*, abs/2304.06787, 2023. doi: 10.48550/ARXIV.2304.06787. URL https://doi.org/10.48550/arXiv.2304.06787.

Smith, A. D. Privacy-preserving statistical estimation with optimal convergence rates. In Fortnow, L. and Vadhan, S. P. (eds.), *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pp. 813–822. ACM, 2011. doi: 10.1145/1993636.1993743. URL https://doi.org/10.1145/1993636.1993743.

Sweeney, L. Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000):1–34, 2000.

Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, 10(5):557–570, 2002. doi: 10.1142/S0218488502001648. URL https://doi.org/10.1142/S0218488502001648.

Talagrand, M. New concentration inequalities in product spaces. *Inventiones Mathematicae*, 126(3):505–563, November 1996. doi: 10.1007/s002220050108. URL https://link.springer.com/article/10.1007/s002220050108.

Thakurta, A. G., Vyrros, A. H., Vaishampayan, U. S., Kapoor, G., Freudiger, J., Sridhar, V. R., and Davidson, D. Learning new words. *Granted US Patents*, 9594741, 2017.

Tsybakov, A. B. *Introduction to Nonparametric Estimation*. Springer series in statistics. Springer, 2009. ISBN 978-0-387-79051-0. doi: 10.1007/b13794. URL https://doi.org/10.1007/b13794.

van Erven, T. and Harremoës, P. Rényi divergence and kullback-leibler divergence. *IEEE Trans. Inf. Theory*, 60(7):3797–3820, 2014. doi: 10.1109/TIT.2014.2320500. URL https://doi.org/10.1109/TIT.2014.2320500.

Wagner, I. and Eckhoff, D. Technical privacy metrics: A systematic survey. *ACM Comput. Surv.*, 51(3):57:1–57:38, 2018. doi: 10.1145/3168389. URL https://doi.org/10.1145/3168389.

Wasserman, L. *All of Nonparametric Statistics (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387251456.

Wasserman, L. A. and Zhou, S. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010. doi: 10.1198/jasa.2009.tm08651. URL https://doi.org/10.1198/jasa.2009.tm08651.

A. Proofs of Section 3

A.1. Proof of Lemma 3.1

Our starting point is the Parseval equality, that leads to:

$$\mathbb{E}\left(\|f-\hat{f}_M\|^2\right) \overset{\text{Parseval}}{=} \mathbb{E}\left(\sum_{k\in\mathbb{Z}^d\setminus\{-M,\dots,M\}^d}|\theta_k|^2+\sum_{k\in\{-M,\dots,M\}^d}|\theta_k-\hat{\theta}_k|^2\right), \tag{17}$$

where the family $(\theta_k)_k$ refers to the Fourier coefficients of f in the basis defined in (1), and where the noisy Fourier coefficient estimators $\left(\hat{\theta}_k\right)_{k\in\{-M,\dots,M\}^d}$ are defined in (3). First, we may notice that :

$$\sum_{k\in\mathbb{Z}^d\setminus\{-M,\dots,M\}^d}|\theta_k|^2=\|f-f_M\|^2 \tag{18}$$

deterministically. This leads to the bias term in the error decomposition.

Then, for any $k\in\{-M,\dots,M\}^d$,

$$\mathbb{E}\left(|\theta_k-\hat{\theta}_k|^2\right)\le\left|\mathbb{E}\left(\hat{\theta}_k\right)-\theta_k\right|^2+\mathbb{V}\left(\hat{\theta}_k\right). \tag{19}$$

Furthermore,

$$\begin{aligned}\mathbb{E}\left(\hat{\theta}_k\right)&\overset{(3)}{=}\mathbb{E}\left(\tilde{\theta}_k+\sigma_K\left(\mathcal{N}(0,1)+i_{\mathbb{C}}\mathcal{N}(0,1)\right)\right)=\mathbb{E}\left(\tilde{\theta}_k\right)\overset{(2)}{=}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n\bar{\phi}_k(X_i)\right)\\&=\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left(\bar{\phi}_k(X_i)\right)=\frac{1}{n}\sum_{i=1}^n\theta_k=\theta_k\,.\end{aligned} \tag{20}$$

Finally,

$$\begin{aligned}\mathbb{V}\left(\hat{\theta}_k\right)&\overset{(3)}{=}\mathbb{V}\left(\tilde{\theta}_k+\sigma_K\left(\mathcal{N}(0,1)+i_{\mathbb{C}}\mathcal{N}(0,1)\right)\right)\overset{\text{Indep.}}{=}\mathbb{V}\left(\tilde{\theta}_k\right)+\mathbb{V}\left(\sigma_K\left(\mathcal{N}(0,1)+i_{\mathbb{C}}\mathcal{N}(0,1)\right)\right)\\&\overset{(2)}{=}\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n\bar{\phi}_k(X_i)\right)+\mathbb{V}\left(\sigma_M\left(\mathcal{N}(0,1)+i_{\mathbb{C}}\mathcal{N}(0,1)\right)\right)\overset{\text{Indep.}}{=}\frac{1}{n^2}\sum_{i=1}^n\mathbb{V}\left(\bar{\phi}_k(X_i)\right)+2\sigma_M^2\\&\overset{|\phi_k(\cdot)|\le1\text{ \& Lemma F.1}}{\le}\frac{1}{n^2}\sum_{i=1}^n1+2\sigma_M^2=\frac{1}{n}+2\sigma_M^2\end{aligned} \tag{21}$$

A.2. Proof of Theorem 3.2

The mechanism $(X_1,\dots,X_n)\mapsto\hat{f}_M$ may equivalently be seen as the mechanism that releases the vector in $\mathbb{C}^{(2M+1)^d}$ of the privatized Fourier coefficient estimates, or as the mechanism that releases the vector in $\mathbb{R}^{2(2M+1)^d}$ of the real and imaginary parts (respectively noted $R(\cdot)$ and $I(\cdot)$) of the privatized Fourier coefficient estimates.

We aim to apply Lemma 2.2: for this purpose, consider any multi-index k, and any $(X_1,\dots,X_n),(X'_1,\dots,X'_n)\in[0,1]^d$,

$$\begin{aligned}\left|\tilde{\theta}_k(X_1,\dots,X_n)-\tilde{\theta}_k(X'_1,\dots,X'_n)\right|&=\left|\frac{1}{n}\sum_{j=1}^n\bar{\phi}_k(X_j)-\frac{1}{n}\sum_{j=1}^n\bar{\phi}_k(X'_j)\right|\\&\le\frac{1}{n}\sum_{j=1}^n\left|\bar{\phi}_k(X_j)-\bar{\phi}_k(X'_j)\right|\\&\overset{|\phi_k(\cdot)|\le1}{\le}\frac{2d_{\text{ham}}\left((X_1,\dots,X_n),(X'_1,\dots,X'_n)\right)}{n}\,,\end{aligned} \tag{22}$$

Hence, for any k, $\tilde{\theta}_k$ is of l_2 sensitivity $\frac{2}{n}$. Hence, for any k, $R(\tilde{\theta}_k)$ and $I(\tilde{\theta}_k)$ are both of sensitivity at most $\frac{2}{n}$ (because $R(\cdot)$ and $I(\cdot)$ are orthogonal projections and are hence contraction linear mappings).

The l_2 sensitivity of computing the $2(2M+1)^d$ approximate real and imaginary parts of the Fourier coefficients is thus $\frac{2}{n}\sqrt{2(2M+1)^d}$. Then, the application of Lemma 2.2 guarantees that the mechanism that releases $\hat{f}_M$, when computed with $\sigma_M = \frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho}}$ satisfies ρ-zCDP.

B. Proofs of Section 4

B.1. Proof of Lemma 4.1

We will need the following lemma :

Lemma B.1 (Fourier tail in Sobolev spaces)**.** *If $f \in \mathcal{S}_L^p(\beta)$, then*

$$\sum_{k\in\mathbb{Z}^d}\left(\sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta}(2\pi k)^{\times 2\alpha}\right)|\theta_k|^2 \leq L^2 \ , \tag{23}$$

where $(\theta_k)_{k\in\mathbb{Z}^d}$ are the Fourier coefficients of f w.r.t. the basis $(\phi_k)_{k\in\mathbb{Z}^d}$.

Proof. Let $\alpha \in \mathbb{N}^d$ such that $|\alpha| = \beta$ and let $k \in \mathbb{Z}^d$. Let us look $\theta_k^{(\alpha)}$ at the k^{th} Fourier coefficient of $\partial^\alpha f$. Since $\beta \geq 1$, there exists $i_0 \in \mathbb{N}$ such that $\alpha_i \geq 1$. We note α_{-i_0} the multi-index with the same values as α except for its i_0^{th} coordinate which has been decremented by 1. Furthermore, for any $x \in [0,1]^d$, any $i \in \{1,\dots,d\}$, and any $y \in [0,1]$ we note $x^{(i=y)}$ the vector with the same components as x but with y as its i^{th} component.

We have that

$$\begin{aligned}
\theta_k^{(\alpha)} &= \int_{[0,1]^d}\partial^\alpha f(x_1,\dots,x_d)\bar{\phi_k}(x_1,\dots,x_d)dx_1\dots dx_d \\
&\overset{\text{Fubini}}{=} \int_{[0,1]^{d-1}}\left(\int_{[0,1]}\partial^\alpha f(x_1,\dots,x_d)\bar{\phi_k}(x_1,\dots,x_d)dx_{i_0}\right)dx_1\dots dx_{i_0-1}dx_{i_0+1}\dots dx_d \\
&= \int_{[0,1]^{d-1}}\left(\int_{[0,1]}\partial^\alpha f(x_1,\dots,x_d)e^{-i_{\mathbb{C}}2\pi(k_1x_1+\dots+k_dx_d)}dx_{i_0}\right)dx_1\dots dx_{i_0-1}dx_{i_0+1}\dots dx_d \\
&\overset{\text{I.B.P.}}{=} \int_{[0,1]^{d-1}}\Bigg(\partial^{\alpha_{-i_0}}f(x^{(i_0=1)})e^{-i_{\mathbb{C}}2\pi\langle k,x^{(i_0=1)}\rangle} - \partial^{\alpha_{-i_0}}f(x^{(i_0=0)})e^{-i_{\mathbb{C}}2\pi\langle k,x^{(i_0=0)}\rangle} \\
&\qquad + i_{\mathbb{C}}2\pi k_{i_0}\int_{[0,1]}\partial^{\alpha_{-i_0}}f(x_1,\dots,x_d)e^{-i_{\mathbb{C}}2\pi(k_1x_1+\dots+k_dx_d)}dx_{i_0}\Bigg)dx_1\dots dx_{i_0-1}dx_{i_0+1}\dots dx_d \\
&= i_{\mathbb{C}}2\pi k_{i_0}\theta_k^{(\alpha_{-i_0})} \ .
\end{aligned} \tag{24}$$

Thus, by induction, we get that

$$\theta_k^{(\alpha)} = (i_{\mathbb{C}}2\pi k)^{\times\alpha}\theta_k \ . \tag{25}$$

Next, since it holds for any k, we may write

$$\begin{aligned}
\int_{[0,1]^d}|\partial^\alpha f|^2 &\overset{\text{Parseval}}{=} \sum_{k\in\mathbb{Z}^d}\langle\theta_k^{(\alpha)},\theta_k^{(\alpha)}\rangle \\
&\overset{(25)}{=} \sum_{k\in\mathbb{Z}^d}(2\pi k)^{\times 2\alpha}\langle\theta_k,\theta_k\rangle \ .
\end{aligned} \tag{26}$$

Finally, since this holds for any α, we may sum over α and use (4.1) to get that

$$\begin{aligned} L^2 &\geq \sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta} \int_{[0,1]^d} |\partial^\alpha f|^2 \\ &\overset{(26)}{=} \sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta} \sum_{k\in\mathbb{Z}^d} (2\pi k)^{\times 2\alpha} \langle \theta_k, \theta_k \rangle \\ &= \sum_{k\in\mathbb{Z}^d} \left(\sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta} (2\pi k)^{\times 2\alpha} \right) |\theta_k|^2 \ . \end{aligned} \tag{27}$$

□

If $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d \setminus \{-M, \ldots, M\}^d$, then there exists i_0 such that

$$|k_{i_0}| \geq M + 1 \ . \tag{28}$$

By considering the multi-index α_0 composed with only 0's except at the index i_0 to which we assign the value β, we thus get

$$(2\pi(M+1))^{2\beta} \leq (2\pi k)^{\times 2\alpha_0} \ , \tag{29}$$

which allows writing

$$(2\pi(M+1))^{2\beta} \leq \sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta} (2\pi k)^{\times 2\alpha} \tag{30}$$

since α_0 is part of the summation indexes of the right-hand side.

Combining the last inequality with Lemma B.1 yields

$$\begin{aligned} \sum_{k\in\mathbb{Z}^d\setminus\{-M,\ldots,M\}} (2\pi(M+1))^{2\beta} |\theta_k|^2 &\overset{(30)}{\leq} \sum_{k\in\mathbb{Z}^d\setminus\{-M,\ldots,M\}} \left(\sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta} (2\pi k)^{\times 2\alpha} \right) |\theta_k|^2 \\ &\leq \sum_{k\in\mathbb{Z}^d} \left(\sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta} (2\pi k)^{\times 2\alpha} \right) |\theta_k|^2 \\ &\overset{Lemma\ B.1}{\leq} L^2 \ . \end{aligned} \tag{31}$$

Hence,

$$\begin{aligned} \frac{L^2}{(2\pi)^{2\beta}} \frac{1}{(M+1)^{2\beta}} &\geq \sum_{k\in\mathbb{Z}^d\setminus\{-M,\ldots,M\}} |\theta_k|^2 \\ &\overset{\text{Parseval}}{=} \|f - f_M\|^2 \ . \end{aligned} \tag{32}$$

B.2. Proof of Theorem 4.2

The privacy of this mechanism is a direct consequence of Theorem 3.2. Below, □ will refer to a constant that depends on β and L, whose value may change from line to line, and that is independent from n and ρ.

By combining Lemma 3.1 and Lemma 4.1 with the value of the variance factor $\sigma_M = \frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho}}$, we get that:

$$\begin{aligned}
\mathbb{E}\left(\|f-\hat{f}_M\|^2\right) &\overset{Lemma\ 3.1}{\leq} \|f-f_M\|^2 + \frac{(2M+1)^d}{n} + 2(2M+1)^d\left(\frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho}}\right)^2 \\
&\overset{Lemma\ 4.1}{\leq} \frac{L^2}{(2\pi)^{2\beta}}\frac{1}{(M+1)^{2\beta}} + \frac{(2M+1)^d}{n} + 2(2M+1)^d\left(\frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho}}\right)^2 \\
&\overset{2K+1\leq 2(K+1)}{\leq} \frac{L^2}{(2\pi)^{2\beta}}\frac{1}{(M+1)^{2\beta}} + \frac{2^d(M+1)^d}{n} + 2^{d+1}(M+1)^d\left(\frac{2\sqrt{2^d(M+1)^d}}{n\sqrt{\rho}}\right)^2 \\
&\leq \square\left(\frac{1}{(M+1)^{2\beta}} + \frac{2^d(M+1)^d}{n} + \frac{2^{2d}(M+1)^{2d}}{n^2\rho}\right) ,
\end{aligned} \tag{33}$$

We then find the optimal trade-off for M by separating the regimes where the variance is dominated by the sampling noise or by the privacy noise.

- **Bias - Sampling variance equilibrium :** We may first observe that

$$\frac{1}{(M+1)^{2\beta}} \geq \frac{2^d(M+1)^d}{n} \iff M+1 \leq \left(n/2^d\right)^{\frac{1}{2\beta+d}} . \tag{34}$$

- **Bias - Privacy variance equilibrium :** In the meantime, we get:

$$\frac{1}{(M+1)^{2\beta}} \geq \frac{2^{2d}(M+1)^{2d}}{n^2\rho} \iff M+1 \leq \left(n\sqrt{\rho}/2^d\right)^{\frac{1}{\beta+d}} . \tag{35}$$

Hence, by taking

$$M+1 = \min\left\{\left\lfloor\left(n/2^d\right)^{\frac{1}{2\beta+d}}\right\rfloor, \left\lfloor\left(n\sqrt{\rho}/2^d\right)^{\frac{1}{\beta+d}}\right\rfloor\right\} , \tag{36}$$

we have that

$$\max\left\{\frac{2^d(M+1)^d}{n}, \frac{2^{2d}(M+1)^{2d}}{n^2\rho}\right\} \leq \frac{1}{(M+1)^{2\beta}} , \tag{37}$$

and Equation (33) yields

$$\mathbb{E}\left(\|f-\hat{f}_M\|^2\right) \leq C\frac{1}{(M+1)^{2\beta}} . \tag{38}$$

C. Proofs of Section 5

C.1. Proof of Theorem 5.1

Let m be an integer that will be specified later on in the proof. We consider the grid

$$\underbrace{\left\{\frac{1}{m+1}, \frac{2}{m+1}, \ldots, \frac{m}{m+1}\right\} \times \cdots \times \left\{\frac{1}{m+1}, \frac{2}{m+1}, \ldots, \frac{m}{m+1}\right\}}_{d \text{ times}} . \tag{39}$$

It has m^d points, and is hence in bijection with $\{1, \ldots, m^d\}$. For any $i \in \{1, \ldots, m^d\}$, we identify p_i with a unique point on this grid. By construction, we have that

$$\forall i, j \in \{1, \ldots, m^d\}, \quad i \neq j \implies \|p_i - p_j\| \geq \frac{1}{m+1} . \tag{40}$$

Now, let us consider the function Ψ given by Lemma F.2 in dimension d. We note $\psi(\cdot) = a\Psi\left(\frac{\cdot}{2}\right)$ where $a > 0$ is fixed to a small enough value such that

$$\sum_{\alpha\in\mathbb{N}^d:|\alpha|=\beta} \int_{[0,1]^d} |\partial^\alpha \psi|^2 \leq L^2 . \tag{41}$$

We also define $\gamma = \int \psi$ and $\delta = \int \psi^2$.

Let $1 \geq h > 0$. For any $\theta \in \{0,1\}^{m^d}$, we define

$$f_\theta(\cdot) := 1 + h^\beta \sum_{i=1}^{m^d} \theta_i \psi\left(\frac{\cdot - p_i}{h}\right) - \|\theta\|_1 \gamma h^{\beta+d} \,. \tag{42}$$

Let us investigate the conditions under which f_θ is a density of probability w.r.t. Lebesgue's measure on $[0,1]^d$.

- For any θ, f_θ is continuous and hence measurable.
- f_θ has to be positive for any θ. This is for instance the case when for any θ, $\|\theta\|_1 \gamma h^{\beta+d} \leq 1$. Since $\|\theta\|_1 \leq m^d$ for any θ, fixing $h = \min\left\{\frac{1}{\gamma(m+1)}, \frac{1}{4(m+1)}\right\}$ is enough to ensure that condition. The reason why we added the term $\frac{1}{4(m+1)}$ in the minimum and why we took $m+1$ instead of m is that we also have that for any i, $\psi\left(\frac{\cdot - p_i}{h}\right)$ has its support in $(0,1)^d$ and that $i \neq j \implies \psi\left(\frac{\cdot - p_i}{h}\right)$ and $\psi\left(\frac{\cdot - p_j}{h}\right)$ have disjoint supports.
- For any θ, we need $\int f_\theta = 1$, which is immediate by construction with a simple variable swap of inverse Jacobian h^d :

$$\begin{aligned}
\int_{[0,1]^d} f_\theta &= \int_{[0,1]^d} \left(1 + h^\beta \sum_{i=1}^{m^d} \theta_i \psi\left(\frac{x - p_i}{h}\right) - \|\theta\|_1 \gamma h^{\beta+d}\right) dx \\
&= 1 + h^\beta \sum_{i=1}^{m^d} \theta_i \int_{[0,1]^d} \psi\left(\frac{x - p_i}{h}\right) dx - \|\theta\|_1 \gamma h^{\beta+d} \\
&\overset{u_i = \frac{x-p_i}{h}}{=} 1 + \|\theta\|_1 \gamma h^{\beta+d} - \|\theta\|_1 \gamma h^{\beta+d} \\
&= 1
\end{aligned} \tag{43}$$

Furthermore, we may also check that for any θ, $f_\theta \in \mathcal{S}_L^p(\beta)$.

- For any θ, by construction, the support of $\partial^\alpha f_\theta$ is included in $(0,1)^d$ for any multi-index α such that $|\alpha| \geq 1$. Hence, the periodicity argument holds trivially since $\partial^\alpha f_\theta = 0$ on the boundary of $[0,1]^d$. Furthermore, since f_θ is constant on the boundary of $[0,1]^d$, the periodicity argument also holds for f_θ.
- Furthermore, let us fix θ and let α be a multi-index such that $|\alpha| = \beta$. We have

$$\begin{aligned}
\int_{[0,1]^d} \left(f_\theta^{(\alpha)}\right)^2 &= \int_{[0,1]^d} \left(h^\beta \sum_{i=1}^{m^d} \theta_i \left(x \mapsto \psi\left(\frac{x - p_i}{h}\right)\right)^{(\alpha)}\right)^2 \\
&= \int_{[0,1]^d} \left(\sum_{i=1}^{m^d} \theta_i \psi^{(\alpha)}\left(\frac{\cdot - p_i}{h}\right)\right)^2 \\
&\overset{\text{disjoint supports}}{=} \sum_{i=1}^{m^d} \theta_i \int_{[0,1]^d} \left(\psi^{(\alpha)}\left(\frac{\cdot - p_i}{h}\right)\right)^2 \\
&\overset{\|\theta\|_1 \leq m^d \,\&\, \text{variable swap}}{\leq} m^d h^d \int_{[0,1]^d} \left(\psi^{(\alpha)}\right)^2 \\
&\overset{m^d h^d \leq 1}{\leq} \int_{[0,1]^d} \left(\psi^{(\alpha)}\right)^2 \,,
\end{aligned} \tag{44}$$

  Consequently, summing over α yields

$$\sum_{\alpha \in \mathbb{N}^d : |\alpha| = \beta} \int_{[0,1]^d} \left(f_\theta^{(\alpha)}\right)^2 \leq \sum_{\alpha \in \mathbb{N}^d : |\alpha| = \beta} \int_{[0,1]^d} \left(\psi^{(\alpha)}\right)^2 \overset{(41)}{\leq} L^2 \,. \tag{45}$$

Now we will used what is usually referred to as Assouad's lemma, and that has been successfully used to prove lower-bounds under differential privacy in (Duchi et al., 2013; 2014; 2016; Acharya et al., 2021e). The following result is a minor reformulation to match the notations of the article of the version that can be found in (Acharya et al., 2021e).

Fact C.1 (Assouad's Lemma). *If (f_θ) is a family of densities of probability that is parametrized by $\theta \in \{0,1\}^N$, and if there exists a $\tau > 0$ such that*

$$\forall(\theta_1,\theta_2) \ : \quad \|f_{\theta_1} - f_{\theta_2}\|^2 \geq C\tau d_{\mathrm{ham}}(\theta_1,\theta_2) \ , \tag{46}$$

then there exists an absolute constant $C > 0$ such that for any estimator $\hat{f}$, by noting $\hat{\theta}$ the parameter of the closest f_θ in the family $(f_\theta)_{\theta\in\{0,1\}^N}$ for the norm $\|\cdot\|$, then

$$\sup_{\theta\in\{0,1\}^N} \mathbb{E}_{f_\theta^{\otimes n}}\left(\|f_\theta - \hat{f}\|^2\right) \geq C\tau \sum_{i=1}^N \left(\mathbb{P}_{\theta_{-i}}(\hat{\theta}^i \neq 0) + \mathbb{P}_{\theta_{+i}}(\hat{\theta}^i \neq 1)\right) \tag{47}$$

where $\mathbb{P}_{\theta_{+i}}$ and $\mathbb{P}_{\theta_{-i}}$ are the mixture distributions

$$\mathbb{P}_{\theta_{+i}} := \frac{1}{2^{N-1}} \sum_{\theta:\theta^i=1} f_\theta^{\otimes n} \quad \mathbb{P}_{\theta_{-i}} := \frac{1}{2^{N-1}} \sum_{\theta:\theta^i=0} f_\theta^{\otimes n} \ . \tag{48}$$

Notice that in (47) there is a second layer or randomness that is implicit, and that is w.r.t. the estimator itself (for privacy for instance).

Proof. The proof can be found in (Acharya et al., 2021e). □

We will apply this result with $N = m^d$. First, we will check that (46) holds.

Let θ_1, θ_2 be two parametrizations. We have that

$$\begin{aligned}
&\int_{[0,1]^d} (f_{\theta_1} - f_{\theta_2})^2 \\
&\quad \geq \sum_{i=1}^{m^d} \mathbb{1}_{\theta_1^i\neq\theta_2^i} \int_{B(p_i,h/2)} \left(h^{\beta+d}\left(\|\theta_2\|_1 - \|\theta_1\|_1\right)\gamma + (\theta_1^i - \theta_2^i)h^\beta\psi\left(\frac{t-p_i}{h}\right)\right)^2 dt \\
&\quad \geq \sum_{i=1}^{m^d} \mathbb{1}_{\theta_1^i\neq\theta_2^i} \int_{B(p_i,h/2)} \left\{\left(h^\beta\psi\left(\frac{t-p_i}{h}\right)\right)^2 \right. \\
&\qquad\qquad \left. -2\gamma h^{2\beta+d}\left|\|\theta_1\|_1 - \|\theta_2\|_1\right|\psi\left(\frac{t-p_i}{h}\right)\right\} dt \\
&\quad \overset{\text{variable swap}}{\geq} d_{\mathrm{ham}}(\theta_1,\theta_2)\, h^{2\beta+d}\left(\delta - 2m^d h^d\gamma^2\right) \\
&\quad \overset{h=\min\left\{h,\frac{1}{m+1}\left(\delta/(4\gamma^2)^{1/d}\right)\right\}}{\geq} d_{\mathrm{ham}}(\theta_1,\theta_2)\, h^{2\beta+d}\delta/2 \ ,
\end{aligned} \tag{49}$$

where we took the liberty to take a smaller h if needed, with still a scaling proportional to $\frac{1}{m+1}$.

Then, we need to control the term $\mathbb{P}_{\theta_{-i}}(\hat{\theta}^i \neq 0) + \mathbb{P}_{\theta_{+i}}(\hat{\theta}^i \neq 1)$.

Privacy cost. First, we do so by exploiting the constraint of ρ-zCDP. Let us give the following lemma, which is borrowed from (Lalanne et al., 2023b).

Lemma C.2. *If $\hat{f}$ satisfies ρ-zCDP, then for any i,*

$$\begin{aligned}
&\mathbb{P}_{\theta_{-i}}(\hat{\theta}^i \neq 0) + \mathbb{P}_{\theta_{+i}}(\hat{\theta}^i \neq 1) \geq \\
&\quad \frac{1}{2}\left(1 - n\sqrt{\rho/2}\frac{1}{2^{N-1}} \sum_{\theta^1,\dots,\theta^{i-1},\theta^{i+1}\dots,\theta^N\in\{0,1\}} \mathrm{TV}\left(f_{(\theta^1,\dots,\theta^{i-1},0,\theta^{i+1}\dots,\theta^N)}, f_{(\theta^1,\dots,\theta^{i-1},1,\theta^{i+1}\dots,\theta^N)}\right)\right) \ ,
\end{aligned}$$

where $\mathrm{TV}(\cdot,\cdot)$ *denotes the* total variation *distance between probability measures defined as*

$$\mathrm{TV}(\mathbb{P}_1,\mathbb{P}_2) := \sup_{S \text{ measurable}} \mathbb{P}_1(S) - \mathbb{P}_2(S) \ .$$

Proof. Let us consider the coupling $\mathcal{C}$ that selects $\theta^1,\dots,\theta^{i-1},\theta^{i+1}\dots,\theta^N \in \{0,1\}$ uniformly at random, and then returns a random variable that follows a conditional distribution $\mathbb{Q}^{\otimes n}_{\theta^1,\dots,\theta^{i-1},\theta^{i+1}\dots,\theta^N}$ where $\mathbb{Q}_{\theta^1,\dots,\theta^{i-1},\theta^{i+1}\dots,\theta^N}$ is a maximal coupling between $f_{(\theta^1,\dots,\theta^{i-1},0,\theta^{i+1}\dots,\theta^N)}$ and $f_{(\theta^1,\dots,\theta^{i-1},1,\theta^{i+1}\dots,\theta^N)}$, in the sense that if $X,Y \sim \mathbb{Q}_{\theta^1,\dots,\theta^{i-1},\theta^{i+1}\dots,\theta^N}$, then $\mathbb{P}(X=Y) = 1-\mathrm{TV}\left(f_{(\theta^1,\dots,\theta^{i-1},0,\theta^{i+1}\dots,\theta^N)}, f_{(\theta^1,\dots,\theta^{i-1},1,\theta^{i+1}\dots,\theta^N)}\right)$. The existence of such coupling is well known (see, *e.g.* (Kallenberg, 1993)).

Then, the similarity function given by Lemma 8 in (Lalanne et al., 2023a) leads to:

$$\mathbb{P}_{\theta_{-i}}(\hat{\theta}^i \neq 0) + \mathbb{P}_{\theta_{+i}}(\hat{\theta}^i \neq 1) \geq \frac{1}{2}\left(1 - \sqrt{\rho/2}\mathbb{E}_{\mathbf{X},\mathbf{Y}\sim\mathcal{C}}\left(d_{\mathrm{ham}}(\mathbf{X},\mathbf{Y})\right)\right) \ ,$$

which reduces to the advertised result. □

Let us fix $\theta^1,\dots,\theta^{i-1},\theta^{i+1}\dots,\theta^{m^d} \in \{0,1\}$, we have that, by the classical rewriting of the total variation distance $\mathrm{TV}(f,g) = \frac{1}{2}\int|f-g|$,

$$\begin{aligned}
&\mathrm{TV}\left(f_{(\theta^1,\dots,\theta^{i-1},0,\theta^{i+1}\dots,\theta^{m^d})}, f_{(\theta^1,\dots,\theta^{i-1},1,\theta^{i+1}\dots,\theta^{m^d})}\right) \\
&\quad= \frac{1}{2}\int_{[0,1]^d}\left|f_{(\theta^1,\dots,\theta^{i-1},0,\theta^{i+1}\dots,\theta^{m^d})} - f_{(\theta^1,\dots,\theta^{i-1},1,\theta^{i+1}\dots,\theta^{m^d})}\right| \\
&\quad\leq \frac{1}{2}\int_{[0,1]^d}\left(\gamma h^{\beta+d} + h^\beta\psi\left(\frac{\cdot - p_i}{h}\right)\right) \\
&\quad\overset{\text{variable swap}}{=} \gamma h^{\beta+d}
\end{aligned} \tag{50}$$

All in all, by combining (50), Lemma C.2, (49) and Fact C.1, there exist two absolute constants $C_1 > 0$ and $C_2 > 0$ such that, if $\hat{f}$ satisfies ρ-zCDP, then:

$$\sup_{\theta\in\{0,1\}^N} \mathbb{E}_{f_\theta{}^{\otimes n}}\left(\|f_\theta - \hat{f}\|^2\right) \geq C_1 h^{2\beta+d} m^d \delta\left(1 - C_2\gamma n\sqrt{\rho}h^{\beta+d}\right) \ . \tag{51}$$

Finally, choosing h of the order of $\left(\gamma n\sqrt{\rho}\right)^{-\frac{1}{\beta+d}}$, and $m+1$ of the order of $\frac{\min\left\{1/\gamma, 1/4, \left(\delta/(4\gamma^2)^{1/d}\right)\right\}}{h}$ complies with all the requirements on h for the calculus to be valid, and allows writing that there are two quantities $C_1 > 0$ and $C_2 > 0$ depending on L, β and d such that, if $n\sqrt{\rho} > C_2$, then

$$\sup_{\theta\in\{0,1\}^N} \mathbb{E}_{f_\theta{}^{\otimes n}}\left(\|f_\theta - \hat{f}\|^2\right) \geq C_1\left(n\sqrt{\rho}\right)^{-\frac{2\beta}{\beta+d}} \ . \tag{52}$$

Usual sampling cost. Without trying to exploit the private nature of the estimation, we may adopt more usual lower-bounding inequalities.

Let us fix $\hat{f}$ and i. Neyman-Pearson-Le Cam's inequality (Of which the proof can be found in (Rigollet & Hütter, 2015)) allows writing

$$\mathbb{P}_{\theta_{-i}}(\hat{\theta}^i \neq 0) + \mathbb{P}_{\theta_{+i}}(\hat{\theta}^i \neq 1) \geq 1 - \mathrm{TV}\left(\mathbb{P}_{\theta_{+i}}, \mathbb{P}_{\theta_{-i}}\right) \ . \tag{53}$$

Then, Pinsker's inequality (see for instance (Tsybakov, 2009)) gives

$$\mathbb{P}_{\theta_{-i}}(\hat{\theta}^i \neq 0) + \mathbb{P}_{\theta_{+i}}(\hat{\theta}^i \neq 1) \geq 1 - \sqrt{\mathrm{KL}\left(\mathbb{P}_{\theta_{+i}} \middle\| \mathbb{P}_{\theta_{-i}}\right)} \ , \tag{54}$$

where $\mathrm{KL}(\cdot\|\cdot)$ is the Kullback-Leibler (KL) divergence which is defined for any two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ such that $\mathbb{P} \ll \mathbb{Q}$ (absolute continuity) as

$$\mathrm{KL}(\mathbb{P}\|\mathbb{Q}) = \int \log\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{P} \ .$$

Then, Theorem 11 in (van Erven & Harremoës, 2014) gives that

$$\mathrm{KL}\left(\frac{1}{2^{N-1}}\sum_{\theta:\theta^i=1} f_\theta^{\otimes n}\middle\|\frac{1}{2^{N-1}}\sum_{\theta:\theta^i=0} f_\theta^{\otimes n}\right) \le \frac{1}{2^{N-1}}\sum_{\theta:\theta^i=0}\mathrm{KL}\left(f_{\theta^{(i\leftarrow 1)}}^{\otimes n}\middle\| f_{\theta^{(i\leftarrow 0)}}^{\otimes n}\right) ,$$

where $\theta^{(i\leftarrow j)}$ means that we assign j as the value of the i^{th} component in θ.

Finally, by the tensorization property of the KL divergence (van Erven & Harremoës, 2014)

$$\mathbb{P}_{\theta_{-i}}(\hat{\theta}^i \neq 0) + \mathbb{P}_{\theta_{+i}}(\hat{\theta}^i \neq 1) \ge 1 - \sqrt{\frac{1}{2^{N-1}}\sum_{\theta:\theta^i=0} n\mathrm{KL}\left(f_{\theta^{(i\leftarrow 1)}}\middle\| f_{\theta^{(i\leftarrow 0)}}\right)} . \tag{55}$$

Let us fix a θ. We will upper-bound $\mathrm{KL}\left(f_{\theta^{(i\leftarrow 1)}}\middle\| f_{\theta^{(i\leftarrow 0)}}\right)$ uniformly in θ. By definition,

$$\mathrm{KL}\left(f_{\theta^{(i\leftarrow 1)}}\middle\| f_{\theta^{(i\leftarrow 0)}}\right) = \int_{[0,1]^d}\log\left(\frac{f_{\theta^{(i\leftarrow 1)}}}{f_{\theta^{(i\leftarrow 0)}}}\right) f_{\theta^{(i\leftarrow 1)}} , \tag{56}$$

and a classical upper bound of the KL divergence by the χ^2-divergence which follows from $\log(\cdot) \le \cdot - 1$ gives

$$\mathrm{KL}\left(f_{\theta^{(i\leftarrow 1)}}\middle\| f_{\theta^{(i\leftarrow 0)}}\right) = \int_{[0,1]^d}\frac{(f_{\theta^{(i\leftarrow 1)}} - f_{\theta^{(i\leftarrow 0)}})^2}{f_{\theta^{(i\leftarrow 0)}}} . \tag{57}$$

Notice that we took the liberty to divide by various densities of probability without justifying why they were different from 0. We will solve this issue right now, and also control the denominator $f_{\theta^{(i\leftarrow 0)}}$ at the same time.

When we made sure that for any θ, f_θ was always positive, we imposed that $m^d\gamma h^{\beta+d} \le 1$. We can be more aggressive and impose that $m^d\gamma h^{\beta+d} \le 1/2$, for instance by taking $h \le \frac{1}{2\gamma(m+1)}$. This way, we have that for any θ, $f_\theta \ge 1/2$.

As a consequence,

$$\begin{aligned}\mathrm{KL}\left(f_{\theta^{(i\leftarrow 1)}}\middle\| f_{\theta^{(i\leftarrow 0)}}\right) &\le 2\int_{[0,1]^d}(f_{\theta^{(i\leftarrow 1)}} - f_{\theta^{(i\leftarrow 0)}})^2 \\ &\le 2\int_{[0,1]^d}\left(\gamma h^{\beta+d} + h^\beta\psi\left(\frac{x-p_i}{h}\right)\right)^2 \\ &= 2\left(\gamma^2 h^{2\beta+2d} + 2\gamma^2 h^{2\beta+2d} + \delta h^{2\beta+d}\right) .\end{aligned} \tag{58}$$

So, there exist $C_1 > 0$ and $C_2 > 0$ that depend on L, β and d such that when $h < C_2$, then

$$\mathrm{KL}\left(f_{\theta^{(i\leftarrow 1)}}\middle\| f_{\theta^{(i\leftarrow 0)}}\right) \le C_1 h^{2\beta+d} . \tag{59}$$

Furthermore, we can note that C_1 and C_2 are uniform in θ.

Combining this last result with (55), Fact C.1 and (49), we obtain that there exists an absolute $C_3 > 0$ such that, for any estimator $\hat{f}$,

$$\sup_{\theta\in\{0,1\}^N}\mathbb{E}_{f_\theta^{\otimes n}}\left(\|f_\theta - \hat{f}\|^2\right) \ge C_3 h^{2\beta+d} m^d\delta\left(1 - \sqrt{C_1 n h^{2\beta+d}}\right) , \tag{60}$$

as soon as $h < C_2$.

In the end, choosing h of the order of $(n)^{-\frac{1}{2\beta+d}}$, and $m+1$ of the order of $\frac{\min\left\{1/(2\gamma), 1/4, \left(\delta/(4\gamma^2)^{1/d}\right)\right\}}{h}$ complies with all the requirements on h for the calculus to be valid, and allows writing that there are two quantities $C_1 > 0$ and $C_2 > 0$ depending on L, β and d such that, if $n > C_2$, then

$$\sup_{\theta\in\{0,1\}^N}\mathbb{E}_{f_\theta^{\otimes n}}\left(\|f_\theta - \hat{f}\|^2\right) \ge C_1 n^{-\frac{2\beta}{2\beta+d}} . \tag{61}$$

The two lower-bounds being valid for ρ-zCDP estimators, their maximum is also a lower-bound, yielding the result.

D. Proofs of Section 6

D.1. Proof of Theorem 6.1

We also define m^* the integer that is associated to the closest point (from below) of the grid $\mathbb{B}_n$ to the unknown smoothness parameter β:

$$m^* = \min\{m \le k_n \,:\, \beta_m \le \beta\} \quad \text{and} \quad \beta^* = \beta_{m^*}. \tag{62}$$

We emphasize that m^* is a theoretical object, which is purely deterministic and not used in our adaptative procedure. We nevertheless need m^* for our mathematical analysis of the Lepskii method. For the sake of clarity, we will use the following shortcut of notations to improve the readability of our paper:

$$\hat{f}_{M_{n,\rho'_n}(\beta_{\hat{m}_n})} = \hat{f}_{\hat{M}} \quad \text{and} \quad \hat{f}_{M_{n,\rho'_n}(\beta_{m^*})} = \hat{f}_{M^*} \quad \text{and} \quad \hat{f}_{M_{n,\rho'_n}(\beta_\ell)} = \hat{f}_{M(\ell)},$$

and the associated shortcut indices as well:

$$\hat{M} = M_{n,\rho'_n}(\beta_{\hat{m}_n}) \quad \text{and} \quad M^* = M_{n,\rho'_n}(\beta_{m^*}) \quad \text{and} \quad M(\ell) = M_{n,\rho'_n}(\beta_\ell).$$

To establish our adaptive result stated in Theorem 6.1, we need the next cornerstone result.

Proposition D.1. *Assume that $f \in S_L^p(\beta)$ with $n \ge e^\beta$, then $\hat{f}_{M_{n,\rho'_n}(\beta_{\hat{m}_n})} = \hat{f}_{\hat{M}}$ satisfies:*

$$\mathbb{E}[\|\hat{f}_{\hat{M}} - f\|_2] \le 2\sqrt{r_{n,\rho'_n}(\beta)^*}\exp\left(\frac{\varepsilon}{\beta+d}\right) \tag{63}$$

$$+ \sqrt{\sum_{\ell=0}^{k_n} r_{n,\rho'_n}(\beta_\ell)}\sqrt{\sum_{\ell>m^*}\mathbb{P}\left[\|\hat{f}_{M(\ell)} - f\|_2^2 > \frac{1}{4}r_{n,\rho'_n}(\beta_\ell)^*\right]} \tag{64}$$

Proof. We observe that the elementary decomposition holds:

$$\mathbb{E}[\|\hat{f}_{\hat{M}} - f\|_2] = \mathbb{E}[\|\hat{f}_{\hat{M}} - f\|_2 \mathbf{1}_{\hat{m}_n \le m^*}] + \mathbb{E}[\|\hat{f}_{\hat{M}} - f\|_2 \mathbf{1}_{\hat{m}_n > m^*}]. \tag{65}$$

We then consider the two terms separately.

On the event $\hat{m}_n \le m^*$: We apply the triangle inequality and obtain:

$$\mathbb{E}[\|\hat{f}_{\hat{M}} - f\|_2 \mathbf{1}_{\hat{m}_n \le m^*}] \le \mathbb{E}\left[\left(\|\hat{f}_{\hat{M}} - \hat{f}_{M^*}\|_2 + \|\hat{f}_{M^*} - f\|_2\right)\mathbf{1}_{\hat{m}_n \le m^*}\right]$$

Using the definition of $\hat{m}_n$ and $\hat{f}_{\hat{M}}$, we observe that *almost surely*:

$$\begin{aligned}
\|\hat{f}_{\hat{M}} - \hat{f}_{M^*}\|_2 \mathbf{1}_{\hat{m}_n \le m^*} &\le \sqrt{r_{n,\rho'_n}(\beta^*)^*} \\
&\le \sqrt{C(\log n)^a r_{n,\rho'_n}(\beta^*)} \\
&\le \sqrt{C(\log n)^a r_{n,\rho'_n}(\beta)}\exp\left(\left(\frac{\beta}{2\beta+d} - \frac{\beta^*}{2\beta^*+d}\right)\log n\right) \vee \exp\left(\left(\frac{\beta}{\beta+d} - \frac{\beta^*}{\beta^*+d}\right)\frac{\log(n\sqrt{\rho'_n})}{2}\right) \\
&\le \sqrt{C(\log n)^a r_{n,\rho'_n}(\beta)}\left(\exp\left(\frac{(\beta-\beta^*)d}{(2\beta+d)(2\beta^*+d)}\right)\log n\right) \vee \exp\left(\frac{(\beta-\beta^*)d}{(\beta+d)(\beta^*+d)}\frac{\log(n\sqrt{\rho'_n})}{2}\right) \\
&\le \sqrt{C(\log n)^a r_{n,\rho'_n}(\beta)}\exp\left(\frac{\varepsilon}{2(2\beta+d)}\right) \vee \exp\left(\frac{\varepsilon}{\beta+d}\right) \\
&= \sqrt{C(\log n)^a r_{n,\rho'_n}(\beta)}\exp\left(\frac{\varepsilon}{\beta+d}\right),
\end{aligned}$$

where we used above

$$\frac{(\beta-\beta^*)d}{(\beta+d)(\beta^*+d)}\frac{\log(n\sqrt{\rho'_n})}{2} \le \varepsilon \log^{-1} n \frac{d}{(\beta+d)(\beta^*+d)}\left(\log n + \frac{1}{2}\log\rho - \frac{1}{2}\log\log n\right) \le \frac{\varepsilon}{\beta+d}.$$

Obviously, the same upper bound applies when considering the expectation and we deduce that

$$\mathbb{E}\left[\|\hat{f}_{\hat{M}} - \hat{f}_{M^*}\|_2 \mathbf{1}_{\hat{m}_n \leq m^*}\right] \leq \sqrt{C(\log n)^a r_{n,\rho'_n}(\beta)} \exp\left(\frac{\varepsilon}{\beta + d}\right) = \sqrt{r_{n,\rho'_n}(\beta)^*} \exp\left(\frac{\varepsilon}{\beta + d}\right). \tag{66}$$

The second term is dealt easily using the non-adaptive rate of convergence of $\hat{f}_{M^*}$, regardless the value of m^* with respect to $\hat{m}_n$, and the Cauchy-Schwarz inequality:

$$\mathbb{E}\left[\|\hat{f}_{M^*} - f\|_2 \mathbf{1}_{\hat{m}_n \leq m^*}\right] \leq \sqrt{\mathbb{E}\left[\|\hat{f}_{M^*} - f\|_2^2\right]} \leq \sqrt{r_{n,\rho'_n}(\beta^*)}$$

Using the same arguments as above, we obtain similarly:

$$\mathbb{E}\left[\|\hat{f}_{M^*} - f\|_2 \mathbf{1}_{\hat{m}_n \leq m^*}\right] \leq \sqrt{r_{n,\rho'_n}(\beta)} \exp\left(\frac{\varepsilon}{\beta + d}\right). \tag{67}$$

We now gather Equations (66) and (67) and obtain that:

$$\mathbb{E}\left[\|\hat{f}_{\hat{M}} - f\|_2 \mathbf{1}_{\hat{m}_n \leq m^*}\right] \leq 2\sqrt{r_{n,\rho'_n}(\beta)^*} \exp\left(\frac{\varepsilon}{\beta + d}\right). \tag{68}$$

On the event $\hat{m}_n > m^*$: We still apply the triangle inequality and observe that for any pair (M, M'):

$$\|\hat{f}_M - \hat{f}_{M'}\|_2 \leq \|\hat{f}_M - f\|_2 + \|\hat{f}_{M'} - f\|_2.$$

Consequently, we have

$$\begin{aligned}
\{\hat{m}_n > m^*\} &= \left\{\exists \ell > m^* \,:\, \|\hat{f}_{M(\ell)} - \hat{f}_{M^*}\|_2 > \sqrt{r_{n,\rho'_n}(\beta_\ell)^*}\right\} \\
&\subset \left\{\exists \ell > m^* \,:\, \|\hat{f}_{M(\ell)} - f\|_2 + \|\hat{f}_{M^*} - f\|_2 > \sqrt{r_{n,\rho'_n}(\beta_\ell)^*}\right\} \\
&\subset \left\{\exists \ell > m^* \,:\, \|\hat{f}_{M(\ell)} - f\|_2 > \frac{1}{2}\sqrt{r_{n,\rho'_n}(\beta_\ell)^*}\right\} \cup \left\{\exists \ell > m^* \,:\, \|\hat{f}_{M^*} - f\|_2 > \frac{1}{2}\sqrt{r_{n,\rho'_n}(\beta_\ell)^*}\right\} \\
&\subset \left\{\exists \ell > m^* \,:\, \|\hat{f}_{M(\ell)} - f\|_2 > \frac{1}{2}\sqrt{r_{n,\rho'_n}(\beta_\ell)^*}\right\} \cup \left\{\|\hat{f}_{M^*} - f\|_2 > \frac{1}{2}\sqrt{r_{n,\rho'_n}(\beta_{m^*})^*}\right\},
\end{aligned}$$

where the last inequality comes from the monotonicity (decreasing function) of $\beta \longmapsto r_{n,\rho'_n}(\beta)$. We then deduce with a union bound that:

$$\mathbb{E}\left[\mathbf{1}_{\{\hat{m}_n > m^*\}}\right] \leq \sum_{\ell > m^*} \mathbb{P}\left[\|\hat{f}_{M(\ell)} - f\|_2^2 > \frac{1}{4} r_{n,\rho'_n}(\beta_\ell)^*\right] \tag{69}$$

We then use the Cauchy-Schwarz inequality and (69) to obtain:

$$\begin{aligned}
\mathbb{E}[\|\hat{f}_{\hat{M}} - f\|_2 \mathbf{1}_{\hat{m}_n > m^*}] &\leq \sqrt{\mathbb{E}[\|\hat{f}_{\hat{M}} - f\|_2^2]} \sqrt{\mathbb{E}[\mathbf{1}_{\hat{m}_n > m^*}]} \\
&\leq \sqrt{\mathbb{E}\left[\sum_{\ell=0}^{k_n} \|\hat{f}_{M(\ell)} - f\|_2^2\right]} \sqrt{\sum_{\ell > m^*} \mathbb{P}\left[\|\hat{f}_{M(\ell)} - f\|_2^2 > \frac{1}{4} r_{n,\rho'_n}(\beta_\ell)^*\right]} \\
&\leq \sqrt{\sum_{\ell=0}^{k_n} r_{n,\rho'_n}(\beta_\ell)} \sqrt{\sum_{\ell > m^*} \mathbb{P}\left[\|\hat{f}_{M(\ell)} - f\|_2^2 > \frac{1}{4} r_{n,\rho'_n}(\beta_\ell)^*\right]}
\end{aligned}$$

□

From Proposition D.1, we observe that the upper bound of the risk of our adaptive procedure depends on two terms. The first one involves the risk $r_{n,\rho}(\beta)$, up to some multiplicative $\log n$ term, while the second term will be shown to be negligible with respect to the first one as soon as a and C are suitably chosen (see Definition (9)).

The next proposition is purely technical and does not involve any statistical insight.

Proposition D.2. *Assume that $\varepsilon \leq 1/2$, then for any $\rho > 0, n \geq 1$ and $d \geq 1$:*

$$\sum_{\ell=0}^{k_n} r_{n,\rho'_n}(\beta_\ell) \leq 4(2+d)\varepsilon^{-1}\log n^2 \left({\rho'_n}^{-\frac{1}{1+d}} + 2\right).$$

Proof. We observe from our definition of $r_{n,\rho'_n}(\beta)$ that:

$$\begin{aligned}
\sum_{\ell=0}^{k_n} r_{n,\rho'_n}(\beta_\ell) &= \sum_{\ell=0}^{k_n}\left(n^{-\frac{2\beta_\ell}{2\beta_\ell+d}} + (n\sqrt{\rho'_n})^{-\frac{2\beta_\ell}{\beta_\ell+d}})\right) \\
&= \sum_{\ell=0}^{\lfloor \varepsilon^{-1}\log^2 n\rfloor} n^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d/2}} + (n\sqrt{\rho'_n})^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n+d/2}} \\
&= \sum_{\ell\varepsilon<\log n} n^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d/2}} + (n\sqrt{\rho'_n})^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n+d/2}} + \sum_{\ell\geq\lfloor\varepsilon^{-1}\log n\rfloor}^{\lfloor\varepsilon^{-1}\log^2 n\rfloor} n^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d/2}} + (n\sqrt{\rho'_n})^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n+d/2}}
\end{aligned}$$

We focus on the first sum and observe that when $\ell\varepsilon < \log n$:

$$n^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d/2}} = e^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d/2}\log n} = e^{-\frac{\ell\varepsilon}{\ell\varepsilon/\log n+d/2}} \leq e^{-\frac{\ell\varepsilon}{1+d/2}},$$

and similarly:

$$(n\sqrt{\rho'_n})^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n+d/2}} = e^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n+d/2}\log n}{\rho'_n}^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d}} \leq e^{-\frac{2\ell\varepsilon}{1+d}}{\rho'_n}^{-\frac{1}{1+d}}.$$

Hence, using a geometric series, we get:

$$\begin{aligned}
\sum_{\ell\varepsilon<\log n} n^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d/2}} + (n\sqrt{\rho'_n})^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n+d/2}} &\leq \sum_{\ell=0}^{+\infty} e^{-\frac{\ell\varepsilon}{1+d/2}} + e^{-\frac{2\ell\varepsilon}{1+d}}{\rho'_n}^{-\frac{1}{1+d}} \\
&= \frac{1}{1-e^{-\frac{\varepsilon}{1+d/2}}} + \frac{{\rho'_n}^{-\frac{1}{1+d}}}{1-e^{-\frac{2\varepsilon}{1+d}}} \\
&\leq 4(2+d)\varepsilon^{-1}{\rho'_n}^{-\frac{1}{1+d}}. \qquad (70)
\end{aligned}$$

where the last line comes from the bound $e^{-t} \leq 1 - t/2$ when $t \in [0, 1/2)$.

Concerning now the second sum, when $\ell \geq \varepsilon^{-1}\log n$, we verify that:

$$\ell \geq \varepsilon^{-1}\log n \Longrightarrow \frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n + d/2} > \frac{2}{2+d} \quad \text{and} \quad \frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n + d/2} > \frac{2}{1+d},$$

which in turn implies that

$$\sum_{\ell\geq\varepsilon^{-1}\log n}^{k_n} n^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/\log n+d/2}} + (n\sqrt{\rho'_n})^{-\frac{\ell\varepsilon/\log n}{\ell\varepsilon/2\log n+d/2}} < \varepsilon^{-1}\log^2 n\left(n^{-\frac{2}{2+d}} + (n\sqrt{\rho'_n})^{-\frac{2}{1+d}}\right) \qquad (71)$$

Gathering Equations (70) and (71) yields the bound independent from n and d as soon as $\varepsilon < 1/2$:

$$\sum_{\ell=0}^{k_n} r_{n,\rho'_n}(\beta_\ell) \leq 4(2+d)\varepsilon^{-1}\log n^2\left({\rho'_n}^{-\frac{1}{1+d}} + 2\right)$$

□

We finally upper bound the second term of (63) that involves $\mathbb{P}\left[\|\hat{f}_{M(\ell)} - f\|_2^2 > \frac{1}{4} r_{n,\rho}(\beta_\ell)^*\right]$, to be studied when $\ell > m^*$. We obtain the next result.

Proposition D.3. *Assume that $C > 8L^2 \vee 2^{2d+10}$, that $a \geq 1$ and $n \geq 3$, then*

$$\sqrt{\sum_{l > m^*} \mathbb{P}\left[\|\hat{f}_{M(\ell)} - f\|_2^2 > \frac{1}{4} r_{n,\rho'_n}(\beta_\ell)^*\right]} \leq \sqrt{2\varepsilon^{-1}}\, \log n\, n^{-2}.$$

Proof. We first consider any integer $\ell > m^*$ and our starting point is the Parseval equality: we decompose the loss between $\hat{f}_{M(\ell)}$ and f as follows:

$$\begin{aligned}\|\hat{f}_{M(\ell)} - f\|_2^2 &= \|\hat{f}_{M(\ell)} - f_{M(\ell)}\|_2^2 + \|f_{M(\ell)} - f\|_2^2 \\ &\leq 2\left(\sum_{k \in \{-M(\ell),\dots,M(\ell)\}^d} |\theta_k - \tilde{\theta}_k|^2 + \sigma_{M(\ell)}^2 \sum_{k \in \{-M(\ell),\dots,M(\ell)\}^d} |\xi_k|^2\right) + \frac{L^2}{(2\pi)^{2\beta}} (M(\ell)+1)^{-2\beta},\end{aligned}$$

where in the last line we used the tail upper bound of the Fourier series on Sobolev spaces stated in Lemma 4.1.

We observe with our alleviated notations, we obtain that:

$$\frac{1}{4} r_{n,\rho'_n}(\beta_\ell)^* = \frac{C}{4} (\log n)^a M_{n,\rho'_n}(\beta_\ell)^{-2\beta_\ell} = \frac{C}{4} (\log n)^a M(\ell)^{-2\beta_\ell} > \frac{C}{4} (\log n)^a (M(\ell)+1)^{-2\beta_\ell}.$$

Hence, when $\ell > m^*$, we get $\beta_\ell < \beta_{m^*} < \beta$, which implies $(M(\ell)+1)^{-2\beta_\ell} > (M(\ell)+1)^{-2\beta}$. Therefore, as soon as $\frac{C}{4} > 2L^2$, we have:

$$\frac{L^2}{(2\pi)^{2\beta}} (M(\ell)+1)^{-2\beta} < \frac{1}{8} r_{n,\rho'_n}(\beta_\ell)^*.$$

For a such choice of C, we then obtain that for any $a > 0$ and any $n \geq 3$:

$$\begin{aligned}\left\{\|\hat{f}_{M(\ell)} - f\|_2^2 > \frac{1}{4} r_{n,\rho'_n}(\beta_\ell)^*\right\} &\subset \left\{\sum_{k \in \{-M(\ell),\dots,M(\ell)\}^d} |\theta_k - \tilde{\theta}_k|^2 + \sigma_{M(\ell)}^2 \sum_{k \in \{-M(\ell),\dots,M(\ell)\}^d} |\xi_k|^2 > \frac{1}{16} r_{n,\rho'_n}(\beta_\ell)^*\right\} \\ &\subset \underbrace{\left\{\sum_{k \in \{-M(\ell),\dots,M(\ell)\}^d} |\theta_k - \tilde{\theta}_k|^2 > \frac{1}{32} r_{n,\rho'_n}(\beta_\ell)^*\right\}}_{:=E_1} \\ &\quad \cup \underbrace{\left\{\sigma_{M(\ell)}^2 \sum_{k \in \{-M(\ell),\dots,M(\ell)\}^d} |\xi_k|^2 > \frac{1}{32} r_{n,\rho'_n}(\beta_\ell)^*\right\}}_{:=E_2}.\end{aligned}$$

We now consider E_1 and E_2 separately.

Study of E_1: concentration of the sequence $(\tilde{\theta}_k)_{k \in \mathbb{Z}^d}$. We use a simple union bound:

$$E_1 \subset \bigcup_{k \in \{-M(\ell),\dots,M(\ell)\}^d} \left\{|\theta_k - \tilde{\theta}_k|^2 \geq \frac{r_{n,\rho'_n}(\beta_\ell)^*}{32(2M(\ell)+1)^d}\right\}.$$

The Hoeffding inequality applied to the (complex) bounded sequence $(\tilde{\theta}_k)_{k \in \mathbb{Z}^d}$ yields

$$\forall t > 0 \qquad \mathbb{P}(|\theta_k - \tilde{\theta}_k|^2 \geq t) \leq 4e^{-nt^2/4}.$$

Applying this previous inequality in the union bound above leads to

$$\begin{aligned}\mathbb{P}(E_1) &\leq 4(2M(\ell)+1)^d e^{-n\frac{r_{n,\rho'_n}(\beta_\ell)^*}{128(2M(\ell)+1)^d}} \\ &= 4(2M(\ell)+1)^d e^{-n\frac{C(\log n)^a M(\ell)^{-2\beta_\ell}}{128(2M(\ell)+1)^d}} \\ &\leq 4(2M(\ell)+1)^d e^{-n\frac{C(\log n)^a M(\ell)^{-(2\beta_\ell+d)}}{2^d 128}}.\end{aligned}$$

Using Equation (**??**), we observe that $nM(\ell)^{2\beta_\ell+d} \geq 1$, which entails:

$$\mathbb{P}(E_1) \leq 4(2M(\ell)+1)^d e^{-\frac{C(\log n)^a}{2^d 128}}.$$

Then, using that $a > 1$ and remarking from Equation (**??**) that $M(\ell)^d \leq n$, we deduce thanks to our choice of C that:

$$\mathbb{P}(E_1) \leq 2^{d+2} n^{1-\frac{C}{2^{d+6}}} \leq 2^{d+2} n^{-2^d-4} \leq n^{-4}. \tag{72}$$

Study of E_2: concentration of the χ^2 noise of privacy. From the definition of $(\xi_k)_{k\in\mathbb{Z}^d}$ as a complex Gaussian random variable, we now that

$$\sum_{k\in\{-M(\ell),\dots,M(\ell)\}^d} |\xi_k|^2 \sim \chi^2(2(2M(\ell)+1)^d),$$

and centering the chi square distribution yields:

$$\begin{aligned}\mathbb{P}(E_2) &= \mathbb{P}\left(\sigma^2_{M(\ell)}\chi^2(2(2M(\ell)+1)^d) > \frac{r_{n,\rho'_n}(\beta_\ell)^*}{32}\right) \\ &= \mathbb{P}\left(\chi^2(2(2M(\ell)+1)^d) - 2(2M(\ell)+1)^d) > \frac{r_{n,\rho'_n}(\beta_\ell)^*}{32\sigma^2_{M(\ell)}} - 2(2M(\ell)+1)^d)\right).\end{aligned}$$

Using that the variance factor needs to be tuned as $\sigma_{M(\ell)} = \frac{2\sqrt{(2M(\ell)+1)^d}}{n\sqrt{\rho'_n}}$ to ensure a $\rho - zCDP$ and the value of $r_{n,\rho'_n}(\beta_\ell)^*$ stated in (9), we can expand the right hand side of the last inequality as:

$$\begin{aligned}\frac{r_{n,\rho'_n}(\beta_\ell)^*}{32\sigma^2_{M(\ell)}} - 2(2M(\ell)+1)^d) &= 2(2M(\ell)+1)^d)\left(\frac{C(\log n)^a r_{n,\rho'_n}(\beta_\ell) n^2\rho'_n}{256(2M(\ell)+1)^{2d})} - 1\right) \\ &= 2(2M(\ell)+1)^d)\left(\frac{C}{4^d 256}(\log n)^a M(\ell)^{-2(\beta+d)} n^2\rho'_n - 1\right) \\ &\geq 2(2M(\ell)+1)^d)\left(\frac{C}{4^d 256}(\log n)^a - 1\right),\end{aligned}$$

where the last line comes from the definition of $M(\ell)$ that guarantees

$$M(\ell)^{2(\beta_\ell+d)} \leq n^2\rho'_n.$$

We may choose $C \geq 4^d 512$, define $D = 2(2M(\ell)+1)^d$ and we observe that the probability of E_2 is upper bounded by:

$$\mathbb{P}(E_2) \leq \mathbb{P}\left(\chi^2(D) - D \geq \frac{C}{2} D(\log n)^a\right).$$

We now use the χ^2 concentration upper bound stated in Equation (F.4) with $\sigma = 1$ and $\delta = \frac{C}{2}(\log n)^a$ and obtain that:

$$\mathbb{P}(E_2) \leq e^{-D\frac{C^2(\log n)^{2a}}{16}} \vee e^{-D\frac{C(\log n)^a}{4}} \leq e^{-\frac{C(\log n)^a}{2}} \leq n^{-C/2} \leq n^{-5}, \tag{73}$$

according to $a \geq 1$, $D \geq 2$ and our choice of C in the statement of the proposition. □

D.2. Proof of Theorem 6.2

First, we can notice that the claim about the privacy of the whole estimation procedure is a direct consequence of Lemma 2.3. The rest of this proof only focuses on the utility claim.

Let us note $\rho' = \rho/|\mathcal{M}|$ We start by writing $\Lambda^{(1)}(\cdot)$ as a sum of two terms: a sampling one and a privacy one:

$$\Lambda^{(1)}(M) := \Lambda^{(1)}_{\text{samp}}(M) + \Lambda^{(1)}_{\text{priv}}(M) \quad \forall M \in \mathcal{M}\,, \tag{74}$$

and $\Lambda^{(2)}(\cdot)$ as the sum of $\Lambda^{(1)}(\cdot)$ and of a privacy term

$$\Lambda^{(2)}(M) := \Lambda^{(1)}(M) + \Delta_{\text{priv}}(M) \quad \forall M \in \mathcal{M}\,. \tag{75}$$

The values of $\Lambda^{(1)}_{\text{samp}}(\cdot)$, $\Lambda^{(1)}_{\text{priv}}(\cdot)$ and $\Delta_{\text{priv}}(\cdot)$ will be fixed later in the proof.

Then, for any M,

$$\begin{aligned}\|\hat{f}_{\hat{M}} - f\|^2 &\leq 3\left(\|\hat{f}_{\hat{M}} - \text{Proj}_{S_M}(\hat{f}_{\hat{M}})\|^2 + \|\text{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \hat{f}_M\|^2 + \|\hat{f}_M - f\|^2\right)\\ &\leq 6\left(\|\hat{f}_{\hat{M}} - \text{Proj}_{S_{\hat{M}}}(\hat{f}_M)\|^2 + \|\text{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \hat{f}_M\|^2 + \|\text{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \text{Proj}_{S_{\hat{M}}}(\hat{f}_M)\|^2 + \|\hat{f}_M - f\|^2\right)\,.\end{aligned} \tag{76}$$

Because of the definition of $B^2(\cdot)$, we may write that

$$\|\hat{f}_{\hat{M}} - f\|^2 \leq 6\left(B^2(M) + \Lambda^{(1)}(\hat{M}) + B^2(\hat{M}) + \Lambda^{(1)}(M) + \|\text{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \text{Proj}_{S_{\hat{M}}}(\hat{f}_M)\|^2 + \|\hat{f}_M - f\|^2\right)\,, \tag{77}$$

which gives, because of the relation linking $\Lambda^{(1)}(\cdot)$ and $\Lambda^{(2)}(\cdot)$,

$$\begin{aligned}\|\hat{f}_{\hat{M}} - f\|^2 \leq 6\Bigg(&B^2(M) + \Lambda^{(2)}(\hat{M}) + B^2(\hat{M}) + \Lambda^{(2)}(M)\\ &+ \left(\|\text{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \text{Proj}_{S_{\hat{M}}}(\hat{f}_M)\|^2 - (\Delta_{\text{priv}}(M) + \Delta_{\text{priv}}(\hat{M}))\right) + \|\hat{f}_M - f\|^2\Bigg)\,.\end{aligned} \tag{78}$$

Finally, because of the selection rule of $\hat{M}$,

$$\|\hat{f}_{\hat{M}} - f\|^2 \leq 6\Bigg(2(B^2(M) + \Lambda^{(2)}(M)) + \underbrace{\left(\|\text{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \text{Proj}_{S_{\hat{M}}}(\hat{f}_M)\|^2 - (\Delta_{\text{priv}}(M) + \Delta_{\text{priv}}(\hat{M}))\right)_+}_{\text{Extra term 1}} + \|\hat{f}_M - f\|^2\Bigg)\,. \tag{79}$$

We recall that this holds for any $M \in \mathcal{M}$. Furthermore, in order to have control on $B^2(\cdot)$, we may write that for any model $\mathcal{M}' \in \mathcal{M}$,

$$\begin{aligned}&\|\text{Proj}_{S_{M'}}(\hat{f}_M) - \hat{f}_{M'}\|^2 - \Lambda^{(1)}(M')\\ &\quad\leq 2\left(\|\text{Proj}_{S_{M'}}(\tilde{f}_M) - \tilde{f}_{M'}\|^2 + \|\text{Proj}_{S_{M'}}((\hat{f}_M - \tilde{f}_M)) - (\hat{f}_{M'} - \tilde{f}_{M'}))\|^2\right) - \Lambda^{(1)}(M')\\ &\quad\leq 6\Bigg(\|\tilde{f}_{M'} - f_{M'}\|^2 + \|\text{Proj}_{S_{M'}}(\tilde{f}_M) - f_{M\wedge M'}\|^2 + \|f_{M'} - f_{M\wedge M'}\|^2\\ &\qquad\qquad + \|\text{Proj}_{S_{M'}}((\hat{f}_M - \tilde{f}_M)) - (\hat{f}_{M'} - \tilde{f}_{M'}))\|^2\Bigg) - \Lambda^{(1)}(M')\,.\end{aligned} \tag{80}$$

Then using that $\text{Proj}_{S_{M'}}(\tilde{f}_M) = \tilde{f}_{M\wedge M'}$ and that $\|f_{M'} - f_{M\wedge M'}\|^2 \leq \|f - f_M\|^2$ (which is easily seen using the Parseval

formula),

$$\begin{aligned}
&\|\operatorname{Proj}_{S_{M'}}(\hat{f}_M) - \hat{f}_{M'}\|^2 - \Lambda^{(1)}(M') \\
&\qquad \le 6\Bigg(\|\tilde{f}_{M'} - f_{M'}\|^2 + \|\tilde{f}_{M\wedge M'} - f_{M\wedge M'}\|^2 + \|f - f_M\|^2 \\
&\qquad\qquad + \|\operatorname{Proj}_{S_{M'}}((\hat{f}_M - \tilde{f}_M)) - (\hat{f}_{M'} - \tilde{f}_{M'}))\|^2\Bigg) - \Lambda^{(1)}(M') \\
&\qquad \le 6\Bigg(2\|\tilde{f}_{M'} - f_{M'}\|^2 + \|f - f_K\|^2 + 2\|\operatorname{Proj}_{S_{M'}}((\hat{f}_M - \tilde{f}_M))\|^2 + 2\|(\hat{f}_{M'} - \tilde{f}_{M'}))\|^2\Bigg) - \Lambda^{(1)}(M') \,.
\end{aligned} \tag{81}$$

Finally, the decomposition of $\Lambda^{(1)}(\cdot)$ yields

$$\begin{aligned}
&\|\operatorname{Proj}_{S_{M'}}(\hat{f}_M) - \hat{f}_{M'}\|^2 - \Lambda^{(1)}(M') \\
&\qquad = 6\Bigg(2\left(\|\tilde{f}_{M'} - f_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{samp}}(M')}{12}\right) + \|f - f_M\|^2 + 2\|\operatorname{Proj}_{S_{M'}}((\hat{f}_M - \tilde{f}_M))\|^2 \\
&\qquad\qquad + 2\left(\|\hat{f}_{M'} - \tilde{f}_{M'})\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)\Bigg) \\
&\qquad \le 6\Bigg(\underbrace{2\left(\|\tilde{f}_{M'} - f_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{samp}}(M')}{12}\right)_+}_{\text{Extra term 2}} + \|f - f_M\|^2 + 2\|\hat{f}_M - \tilde{f}_M\|^2 \\
&\qquad\qquad + \underbrace{2\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)_+}_{\text{Extra term 3}}\Bigg) \,.
\end{aligned} \tag{82}$$

We thus have decomposed the problem in quantities that we can perfectly control, and with two extra terms that we have to control. This is where the penalization terms are useful in order to force the exponential convergence.

Control of the extra term 2. This term is handled with the help of the Talagrand inequality (Talagrand, 1996; Ledoux, 1997; Klein & Rio, 2005), using a strategy close to the one presented in (Comte, 2017).

Let $\mathcal{M}' \in \mathfrak{M}$, we have that

$$\|\tilde{f}_{M'} - f_{M'}\|^2 = \sup_{g:\|g\|\le 1} |\langle \tilde{f}_{M'} - f_{M'}, g\rangle|^2 \,.$$

Furthermore, by separability of L^2 and the fact that $g \mapsto |\langle \tilde{f}_{M'} - f_{M'}, g\rangle|^2$ is continuous, we may consider this supremum over a *countable* family of functions only (for applying Lemma F.5).

For any g such that $\|g\| \le 1$,

$$\begin{aligned}
\langle \tilde{f}_{M'} - f_{M'}, g\rangle &= \left\langle \frac{1}{n}\sum_{i=1}^n \left(\sum_{k\in\{-M,\dots,M\}^d} \bar{\phi}_k(X_i)\phi_k - f_{K'}\right), g\right\rangle \\
&= \frac{1}{n}\sum_{i=1}^n \left(\underbrace{\sum_{k\in\{-M,\dots,M\}^d} \langle \phi_k, g\rangle\, \bar{\phi}_k(X_i)}_{=:T_g^{(K')}(X_i)} - \underbrace{\langle f_{K'}, g\rangle}_{=\mathbb{E}\left(T_g^{(M')}(X_i)\right)}\right) \\
&= \nu_n(T_g^{(M')}) \,,
\end{aligned} \tag{83}$$

where $\nu_n(T_g^{(M')})$ is defined in Lemma F.5.

We may thus rewrite

$$\|\tilde{f}_{M'} - f_{M'}\|^2 = \sup_{g:\|g\|\leq 1} |\nu_n(T_g^{(M')})|^2 \,,$$

where the sup may be restricted to a countable family. However, $\nu_n(T_g^{(M')})$ is *not* a real-valued quantity, and we cannot apply Lemma F.5 directly. We will have to resort to decompose the quantities of interest and to add an extra factor 2 at the end since

$$|\nu_n(T_g^{(M')})|^2 = R(\nu_n(T_g^{(M')}))^2 + I(\nu_n(T_g^{(M')}))^2 = \nu_n(R(T_g^{(M')}))^2 + \nu_n(I(T_g^{(M')}))^2 \,,$$

and since taking the real part or the imaginary part are contractive projections, and hence reduce the quantities such as the modulus and the variance.

We may first see that

$$\begin{aligned} \|T_g^{(M')}\|^2 &= \sum_{k\in\{-M',\dots,M'\}^d} |\langle \phi_k, g\rangle|^2 \\ &\leq \|g\|^2 \\ &\leq 1 \end{aligned} \tag{84}$$

because $\|g\| \leq 1$.

Hence, we may write

$$T_g^{(M')} = \sum_{k\in\{-M',\dots,M'\}^d} \alpha_k \phi_k \tag{85}$$

where

$$\sum_{k\in\{-M',\dots,M'\}^d} |\alpha_k|^2 \leq 1. \tag{86}$$

Then,

$$\begin{aligned} |\nu_n(T_g^{(M')})|^2 &\leq \left| \sum_{k\in\{-M',\dots,M'\}^d} \alpha_k \nu_n(\phi_k) \right|^2 \\ &\overset{\text{Cauchy-Schwarz}}{\leq} \left(\sum_{k\in\{-M',\dots,M'\}^d} |\alpha_k|^2 \right) \left(\sum_{k\in\{-M',\dots,M'\}^d} |\nu_n(\phi_k)|^2 \right) \\ &= \sum_{k\in\{-M',\dots,M'\}^d} |\nu_n(\phi_k)|^2 \,, \end{aligned} \tag{87}$$

which in turn gives that

$$\begin{aligned} \max_{P(\cdot)=R(\cdot) \text{ or } I(\cdot)} \left\{ \left(\mathbb{E}\left(\sup_{T_g^{(M')}:\|g\|\leq 1} |\nu_n\left(P(T_g^{(M')})\right)| \right) \right)^2 \right\} &\leq \left(\mathbb{E}\left(\sup_{T_g^{(M')}:\|g\|\leq 1} |\nu_n\left(T_g^{(M')}\right)| \right) \right)^2 \\ &\overset{\text{Jensen}}{\leq} \mathbb{E}\left(\sup_{T_g^{(K')}:\|g\|\leq 1} \left(\mu_n(t)\right)^2 \right) \\ &\leq \mathbb{E}\left(\sum_{k\in\{-M',\dots,M'\}^d} |\nu_n(\phi_k)|^2 \right) \\ &= \frac{1}{n} \sum_{k\in\{-M',\dots,M'\}^d} \mathbb{V}\left(\phi_k(X_1)\right) \\ &\overset{\textit{Lemma F.1}}{\leq} \frac{(2M'+1)^d}{n} \,. \end{aligned} \tag{88}$$

This last value may thus be used as H^2 in the application of Lemma F.5.

Furthermore, for any g such that $\|g\| \leq 1$,

$$\begin{aligned}\max\{\|R(T_g^{(M')})\|_\infty, \|I(T_g^{(M')})\|_\infty\} &\leq \|T_g^{(M')}\|_\infty \\ &= \sup_t \left| \sum_{k\in\{-M',\dots,M'\}^d} \langle \phi_k, g\rangle \, \bar{\phi}_k(t) \right| \\ &\overset{\text{Cauchy-Schwarz}}{\leq} \sup_t \sqrt{\sum_{k\in\{-M',\dots,M'\}^d} |\langle \phi_k, g\rangle|^2} \sqrt{\sum_{k\in\{-M',\dots,M'\}^d} |\bar{\phi}_k(t)|^2} \\ &\overset{\|g\|\leq 1 \& |\bar{\phi}_k(\cdot)|\leq 1}{\leq} \sqrt{(2M'+1)^d} \\ &\leq \sqrt{n}\end{aligned} \tag{89}$$

where the last inequality comes from the fact that $(2\max\mathcal{M}+1)^d \leq n$. This gives the value of M_1 for Lemma F.5.

Finally, for any g such that $\|g\| \leq 1$,

$$\begin{aligned}\max\{\mathbb{V}(R(T_g^{(M')}(X_1))), \mathbb{V}(I(T_g^{(M')}(X_1)))\} &\leq \mathbb{V}(T_g^{(M')}(X_1)) \\ &\leq \mathbb{E}\left(\left|T_g^{(M')}(X_1)\right|^2\right) \\ &= \int \left|T_g^{(M')}(x)\right|^2 f(x)dx \\ &\overset{|f(\cdot)|\|\leq \|f\|_\infty \text{ a.s.}}{\leq} \|f\|_\infty \int \left|T_g^{(M')}\right|^2 dx \\ &\overset{\|T_g^{(K')}\|\leq 1}{\leq} \|f\|_\infty\end{aligned} \tag{90}$$

which gives the value of v for Lemma F.5.

So in the end, Lemma F.5 tells us that there exists absolute constants $C_1, C_2, C_3 > 0$ such that, when tuned with $\Lambda^{(1)}_{\text{samp}}(K) \geq 96\frac{(2K+1)^d}{n}$,

$$\mathbb{E}\left(\sum_{K'\in\mathcal{K}} \left(\|\tilde{f}_{M'} - f_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{samp}}(K')}{12}\right)_+\right) \leq \sum_{M'\in\mathcal{K}} \frac{C_1}{n}\left(\|f\|_\infty e^{-C_2\frac{(2M'+1)^d}{\|f\|_\infty}} + e^{-C_2\sqrt{(2M'+1)^d}}\right) \tag{91}$$

Hence, since the series $\sum_n e^{-n}$ and $\sum_n e^{-\sqrt{n}}$ converge, there exists a constant C depending only on $\|f\|_\infty$ such that

$$\mathbb{E}\left(\sum_{M'\in\mathcal{M}} \left(\|\tilde{f}_{M'} - f_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{samp}}(M')}{12}\right)_+\right) \leq \frac{C}{n}\,. \tag{92}$$

Control of the extra term 3. For any $\mathcal{M} \in \mathcal{M}$,

$$\begin{aligned}\mathbb{E}\left(\max_{M'\in\mathcal{M}} \left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)_+\right) &\leq \mathbb{E}\left(\sum_{M'\in\mathcal{M}} \left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)_+\right) \\ &= \sum_{M'\in\mathcal{M}} \mathbb{E}\left(\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)_+\right)\end{aligned} \tag{93}$$

For any M', we may notice that $\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2$ has a χ^2 distribution scaled by $\sigma_{M'}$ and with $2(2M'+1)^d$ degrees of freedom. Lemma F.4 using $\delta = 1$ thus yields:

$$\mathbb{E}\left(\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - (1+1)\sigma^2_{M'}2(2M'+1)^d\right)_+\right) \leq \frac{2\sigma^2_{M'}}{1}e^{-\frac{2(2M'+1)^d}{4}} + 2\sigma^2_{M'}e^{-\frac{2(2M'+1)^d}{2}} \tag{94}$$

Furthermore, since $\sigma_{M'} = \frac{2\sqrt{(2M'+1)^d}}{n\sqrt{\rho'}}$, we have that

$$\mathbb{E}\left(\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{8(2M'+1)^{2d}}{n^2\rho'}\right)_+\right) \leq C\frac{(2M'+1)^d}{n^2\rho'}\left(e^{-\frac{(2M'+1)^d}{2}} + e^{-\frac{(2M'+1)^d}{1}}\right), \tag{95}$$

where C is a non-negative absolute constant.

In the end, using that from our statement $\Lambda^{(1)}_{\text{priv}}(M') \geq \frac{96(2M'+1)^{2d}}{n^2\rho'}$, we may write that

$$\begin{aligned}
\mathbb{E}\left(\max_{M'\in\mathcal{M}}\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)_+\right) &\leq \sum_{M'\in\mathcal{M}}\mathbb{E}\left(\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{8(2M'+1)^{2d}}{n^2\rho'}\right)_+\right) \\
&\leq \sum_{M'\in\mathcal{M}} C\frac{(2M'+1)^d}{n^2\rho'}\left(e^{-\frac{(2M'+1)^d}{2}} + e^{-\frac{(2M'+1)^d}{1}}\right) \\
&\leq \frac{C}{n^2\rho'}\sum_{j\in\mathbb{N}} j\left(e^{-\frac{j}{2}} + e^{-\frac{j}{1}}\right) \\
&\leq \frac{C'}{n^2\rho'}
\end{aligned} \tag{96}$$

where C' is a non-negative absolute constant since $\sum_{j\in\mathbb{N}} j\left(e^{-\frac{j}{2}} + e^{-\frac{j}{1}}\right)$ is finite.

Control of the extra term 1. For a fixed $\mathcal{M} \in \mathfrak{M}$,

$$\begin{aligned}
&\mathbb{E}\left(\left(\|\operatorname{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \operatorname{Proj}_{S_{\hat{M}}}(\hat{f}_M)\|^2 - (\Delta_{\text{priv}}(M) + \Delta_{\text{priv}}(\hat{M}))\right)_+\right) \\
&\leq \mathbb{E}\left(\left(2\|\operatorname{Proj}_{S_M}(\hat{f}_{\hat{M}}) - \tilde{f}_{\hat{M}\wedge M}\|^2 + 2\|\operatorname{Proj}_{S_{\hat{M}}}(\hat{f}_M) - \tilde{f}_{\hat{M}\wedge M}\|^2 - (\Delta_{\text{priv}}(M) + \Delta_{\text{priv}}(\hat{M}))\right)_+\right) \\
&\leq \mathbb{E}\left(\left(2\|\hat{f}_{\hat{M}} - \tilde{f}_{\hat{M}}\|^2 + 2\|\hat{f}_M - \tilde{f}_M\|^2 - (\Delta_{\text{priv}}(M) + \Delta_{\text{priv}}(\hat{M}))\right)_+\right) \\
&\leq \mathbb{E}\left(\left(2\|\hat{f}_{\hat{M}} - \tilde{f}_{\hat{M}}\|^2 - \Delta_{\text{priv}}(\hat{M})\right)_+ + 2\|\hat{f}_M - \tilde{f}_M\|^2\right) \\
&\leq \mathbb{E}\left(\sum_{K'\in\mathcal{M}}\left(2\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \Delta_{\text{priv}}(M')\right)_+ + 2\|\hat{f}_M - \tilde{f}_M\|^2\right)
\end{aligned} \tag{97}$$

Furthermore, following a roadmap similar as the one used in the control of the extra term 3 (see (96)), we observe that if $\Delta_{\text{priv}}(M') \geq \frac{16(2M'+1)^{2d}}{n^2\rho'}$, there exists an absolute constant $C > 0$ such that

$$\mathbb{E}\left(\sum_{M'\in\mathcal{M}}\left(2\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \Delta_{\text{priv}}(M')\right)_+ + 2\|\hat{f}_M - \tilde{f}_M\|^2\right) \leq C\left(\frac{1}{n^2\rho'} + \|\hat{f}_M - \tilde{f}_M\|^2\right) \tag{98}$$

Putting the pieces together. All in all, by taking the expectation, we have proved that for any $M \in \mathcal{M}$,

$$\begin{aligned}
\mathbb{E}\left(\|\hat{f}_{\hat{M}} - f\|^2\right)/C \leq \|f - f_M\|^2 &+ \mathbb{E}\left(\|\hat{f}_M - \tilde{f}_M\|^2\right) + \mathbb{E}\left(\sum_{M'\in\mathcal{M}}\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)_+\right) \\
&+ \mathbb{E}\left(\sum_{M'\in\mathcal{M}}\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Delta_{\text{priv}}(M')}{2}\right)_+\right) + \mathbb{E}\left(\sum_{M'\in\mathcal{M}}\left(\|\tilde{f}_{M'} - f_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{samp}}(M')}{12}\right)_+\right) \\
&+ \Lambda^{(2)}(M)
\end{aligned} \tag{99}$$

where $C > 0$ is an absolute constant

Furthermore, $\mathbb{E}\left(\|\hat{f}_M - \tilde{f}_M\|^2\right) = 2(2M+1)^d\sigma_M^2$, and with the values of $\Lambda^{(1)}_{\text{samp}}(\cdot)$, $\Lambda^{(1)}_{\text{priv}}(\cdot)$ and $\Delta_{\text{priv}}(\cdot)$ that were taken within the proof, the other expectations are controlled, yielding

$$\begin{aligned}\mathbb{E}\left(\|\hat{f}_{\hat{M}} - f\|^2\right)/C' \leq \|f - f_M\|^2 + 2(2M+1)^d\sigma_M^2 + \mathbb{E}\left(\sum_{M'\in\mathcal{M}}\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{priv}}(M')}{12}\right)_+\right) \\ + \mathbb{E}\left(\sum_{M'\in\mathcal{M}}\left(\|\hat{f}_{M'} - \tilde{f}_{M'}\|^2 - \frac{\Delta_{\text{priv}}(M')}{2}\right)_+\right) + \mathbb{E}\left(\sum_{M'\in\mathcal{M}}\left(\|\tilde{f}_{M'} - f_{M'}\|^2 - \frac{\Lambda^{(1)}_{\text{samp}}(M')}{12}\right)_+\right) \\ + \Lambda^{(2)}(M)\end{aligned} \tag{100}$$

D.3. Proof of Theorem 6.3

We recall that for any $M \in \mathcal{M}$, the bias-variance tradeoff $BV(M)$ in Theorem 6.2 reads

$$BV(M) \leq \|f - f_M\|^2 + \frac{(2M+1)^d}{n} + 2(2M+1)^d\sigma_M^2\,, \tag{101}$$

where $\sigma_M = \frac{2\sqrt{(2M+1)^d}}{n\sqrt{\rho/|\mathcal{M}|}}$.

As in the proof of Theorem 4.2, the dichotomy of having the variance dominated by sampling or privacy leads to the the introduction of the optimal cut-off

$$M^* + 1 := \min\left\{(n/2^d)^{\frac{1}{2\beta+d}}, \left(n\sqrt{\rho/\mathcal{M}}/2^d\right)^{\frac{1}{\beta+d}}\right\}\,.$$

If one could guarantee that $M^* + 1$ belongs to $\mathcal{M}$, then Theorem 6.2 would guarantee the advertised result. However, this is not the case.

Even if one cannot guarantee that $M^* + 1 \in \mathcal{M}$, with the construction rule for $\mathcal{M}$, we can always guarantee that for n big enough (the "big enough" depends on β and d), there will exist $M' + 1 \in \mathcal{M}$ such that $(M^* + 1)/2 \leq M' + 1 \leq M^* + 1$.

Since the variance terms are non-increasing with M, using M' instead of M^* only decreases the variance.

The bias term on the other hand is non-decreasing with M. However, by looking at the expressions of the bias in Lemma 4.1 or Proposition E.1 shows that in the worst case, being off by a factor at most $1/2$ degrades the estimation bias by a factor $2^{2\beta}$.

Using that the $\min_{M\in\mathcal{M}} BV(M)$ in Theorem 6.2 is upper-bounded by $BV(M')$ and that the residual terms are negligible yields the result.

E. On non-integer multi-dimensional Sobolev spaces

This section presents all the technical details on how to handle Sobolev spaces of non-integer smoothness.

E.1. Definition

Below, we shall discuss on the multi-dimensional Sobolev spaces with a non-integer parameter $\beta \geq 0$.

Our starting point is the space of Hölderian functions with a (fractional) order $s \in (0,1)$ and radius R:

$$\mathcal{H}_R(s) = \left\{f : \mathbb{R}^d \longrightarrow \mathbb{R}\,|\,\|f\|_{\mathcal{H}_s} := \sup_{(x,y)\in[0,1]^d\times[0,1]^d}\frac{|f(x) - f(y)|}{\|x - y\|^s} \leq R\right\}. \tag{102}$$

Then, for any *real* value β, we shall use the decomposition $\beta = \lfloor\beta\rfloor + \nu$ where $\nu = \beta - \lfloor\beta\rfloor \in [0,1)$. In this decomposition, $\lfloor\beta\rfloor$ is then the integer part of the order derivatives and ν the fractional one: $\lfloor\beta\rfloor$ encodes for a number of integer derivatives whereas ν refers to an Holderian smoothness of these derivatives.

For a given $L > 0$, we will say that $f \in S_L(\beta)$ if:

$$S_L(\beta) := \left\{ f : \mathbb{R}^d \longrightarrow \mathbb{R} \,|\, \sum_{|\alpha|=\lfloor\beta\rfloor} \|\partial^\alpha f\|_2^2 + \mathbf{1}_{\nu>0} \sum_{|\alpha|=\lfloor\beta\rfloor} \|\partial^\alpha f\|_{\mathcal{H}_\nu}^2 \leq L^2 \right\}. \tag{103}$$

We observe that when β is an integer, $S_L(\beta)$ synchronises with the standard definition.

E.2. Control of the bias

We establish below the important tail behaviour of the Fourier series in our generalized Holderian Sobolev spaces:

Proposition E.1. *Assume that $f \in S_L^p(\beta)$, then an explicit constant $\square(\beta)$ independent from d exists such that*

$$\sum_{k \notin \{-M,\dots,M\}^d} |\theta_k(f)|^2 \leq \square(\beta) d (M+1)^{-2\beta} L^2$$

We first state an important proposition on the relation between the fractional Holder exponent $s \in (0,1)$ of any function f and the Fourier series associated to f.

Proposition E.2. *Assume that $f \in \mathcal{H}_R(s)$ for $s \in (0,1)$ and that f satisfies the periodicity condition (7) for $\alpha = (0,\dots,0)$, then the Fourier series associated to $(\theta_k(f))_{k\in\mathbb{Z}^d}$ of f satisfies*

$$\sum_{k \notin \{-M,\dots,M\}^d} |\theta_k(f)|^2 \leq C(s) d R^2 (M+1)^{-2s},$$

where $C(s) = \frac{2^{2s}3^{-s}}{1-2^{-2s}}$.

Proof. Below, k refers to a d dimensional vector of integers, and $\max(|k|)$ is the maximal value of the vector that contains the absolute values of the coordinates of k.

We consider f and a translation of f denoted by $f_{\mathfrak{h}}$: $f_{\mathfrak{h}}(x) = f(x-\mathfrak{h})$ where $\mathfrak{h}$ is any vector of $[0,1]^d$. Using the periodicity of f, we have:

$$\forall k \in \mathbb{Z}^d \qquad \theta_k(f) = \int_{[0,1]^d} f(x) e^{-i_{\mathbb{C}} 2\pi\langle k,x\rangle} \mathrm{d}x \qquad \text{and} \qquad \theta_k(f_{\mathfrak{h}}) = \int_{[0,1]^d} f_{\mathfrak{h}}(x) e^{-i_{\mathbb{C}} 2\pi\langle k,x\rangle} \mathrm{d}x = \theta_k(f) e^{-i_{\mathbb{C}} 2\pi\langle k,\mathfrak{h}\rangle},$$

which entails:

$$\theta_k(f)\left(1 - e^{-i_{\mathbb{C}} 2\pi\langle k,\mathfrak{h}\rangle}\right) = \int_{[0,1]^d} (f(x) - f_{\mathfrak{h}}(x)) e^{-i_{\mathbb{C}} 2\pi\langle k,x\rangle} \mathrm{d}x$$

We get from the Parseval equality and the fractional Holder hypothesis on f, for any collection of vectors $\mathfrak{h}^{(j)} \in [0,1]^d$:

$$\forall j \in \{1,\dots,d\} \qquad \sum_{k\in\mathbb{Z}^d} |\theta_k(f)|^2 \left|1 - e^{-i_{\mathbb{C}} 2\pi\langle k,\mathfrak{h}^{(j)}\rangle}\right|^2 = \|f - f_{\mathfrak{h}^{(j)}}\|_2^2 \leq R^2 |\mathfrak{h}^{(j)}|^{2s} \tag{104}$$

We now consider $k = (k_1,\dots,k_d) \in \mathbb{Z}^d$ and assume that for $j \in \{1,\dots,d\} : |k_j| = K \in [2^m, 2^{m+1})$. For this coordinate $j \in \{1,\dots,d\}$, we consider the vector $\mathfrak{h}^{(j)} = \frac{2^{-m}}{3}\delta_j$ and we verify that:

$$|2\pi\langle k, \mathfrak{h}^{(j)}\rangle| = \frac{2\pi}{3} k_i 2^{-m} \in \left[\frac{2\pi}{3}, \frac{4\pi}{3}\right).$$

It implies that:

$$\left|1 - e^{-i_{\mathbb{C}} 2\pi\langle k,\mathfrak{h}^{(j)}\rangle}\right|^2 \geq 1,$$

which in turn leads to

$$\begin{aligned}\sum_{k\in\mathbb{Z}^d : |k_j| \in [2^m, 2^{m+1})} |\theta_k(f)|^2 &\leq \sum_{k\in\mathbb{Z}^d : |k_j| \in [2^m, 2^{m+1})} |\theta_k(f)|^2 \left|1 - e^{-i_{\mathbb{C}} 2\pi\langle k,\mathfrak{h}^{(j)}\rangle}\right|^2 \\ &\leq R^2 |\mathfrak{h}^{(j)}|^{2s}\end{aligned}$$

where we applied Equation (104) in the last line. Using the value of $\mathfrak{h}^{(j)}$, we deduce that:

$$\forall j \in \{1,\dots,d\} \qquad \sum_{k\in\mathbb{Z}^d : |k_j| \in [2^m, 2^{m+1})} |\theta_k(f)|^2 \leq R^2 3^{-2s} 2^{-2ms}. \tag{105}$$

We are now able to conclude the proof: we consider any integer $M \geq 1$ and the dyadic scale, which associates $m_0 \geq 0$ such that $2^{m_0} \leq M < 2^{m_0+1}$, we observe that

$$\begin{aligned}
\sum_{k\in\mathbb{Z}^d \,:\, k\notin\{-M,\dots,M\}^d} |\theta_k(f)|^2 &\leq \sum_{k\in\mathbb{Z}^d \,:\, k\notin\{-2^{m_0},\dots,2^{m_0}\}^d} |\theta_k(f)|^2 \\
&\leq \sum_{k\in\mathbb{Z}^d \,\exists j \,:\, |k_j|\geq 2^{m_0}} |\theta_k(f)|^2 \\
&\leq \sum_{j=1}^{d} \sum_{k\in\mathbb{Z}^d \,:\, |k_j|\geq 2^{m_0}} |\theta_k(f)|^2 \\
&\leq \sum_{j=1}^{d} \sum_{m\geq m_0} \sum_{k\in\mathbb{Z}^d \,:\, 2^m \leq |k_j| < 2^{m+1}} |\theta_k(f)|^2 \\
&\leq R^2 3^{-s} \sum_{j=1}^{d} \sum_{m\geq m_0} 2^{-2ms} \\
&\leq \frac{R^2 3^{-s}}{1-2^{-2s}} d 2^{-2m_0 s} \\
&\leq \frac{R^2 2^{2s} 3^{-s}}{1-2^{-2s}} d (M+1)^{-2s}.
\end{aligned}$$

We obtain the conclusion of the proof with $C(s) = \frac{2^{2s}3^{-s}}{1-2^{-2s}}$. □

Proof of Proposition E.1. We are now ready to extend our estimate stated in Lemma 4.1 from integer Sobolev spaces to fractional ones. Assume that $\beta > 0$: we observe that

- If $\beta \in \mathbb{N}$, then Lemma 4.1 yields
$$\|f - f_M\| \leq \frac{L^2}{(2\pi)^{2\beta}} (M+1)^{-2\beta}.$$
- Oppositely, if $\beta = \lfloor \beta \rfloor + s$ with $s \in (0,1)$ and assume that $f \in S_L^p(\beta)$, we know from Proposition E.2 that:
$$\sum_{|\alpha|=\lfloor\beta\rfloor} \sum_{k\notin\{-M,\dots,M\}^d} |\theta_k(\partial^\alpha(f))|^2 \leq \sum_{|\alpha|=\lfloor\beta\rfloor} C(s) d M^{-2s} \|\partial^\alpha f\|_{\mathcal{H}_s}^2 \leq C(s) d (M+1)^{-2s} L^2.$$
We then conclude following the same guidelines as the ones of Lemma 4.1:
$$\begin{aligned}
((2\pi)(M+1))^{2\lfloor\beta\rfloor} \sum_{k\notin\{-M,\dots,M\}^d} |\theta_k(f)|^2 &\leq \sum_{|\alpha|=\lfloor\beta\rfloor} \sum_{k\notin\{-M,\dots,M\}^d} (2\pi k)^{2\alpha} |\theta_k(f)|^2 \\
&\leq C(s) d (M+1)^{-2s} L^2,
\end{aligned}$$
which implies with $\beta = \lfloor\beta\rfloor + s$ the final bound:
$$\sum_{k\notin\{-M,\dots,M\}^d} |\theta_k(f)|^2 \leq \frac{C(s) L^2}{(2\pi)^{2\lfloor\beta\rfloor}} d (M+1)^{-2\beta}.$$

□

E.3. Lower-bounds : Adaptation of the proof of Theorem 5.1 in the case of real-valued β's

The only adaptation needed to the proof is to handle the new Hölderian part in the definition. In fact, the only adaptation needed is to slightly modify the function $\psi(\cdot)$ in Appendix C.1, and to verify that the subsequent family of functions defined from it is a family of densities of probability in $\mathcal{S}_L^p(\beta)$. We use the decomposition $\beta = \lfloor\beta\rfloor + \nu$ where $\nu = \beta - \lfloor\beta\rfloor \in [0,1)$.

Let $\epsilon > 0$ that will be fixed later. The old ψ of Appendix C.1 is replaced by a new $\psi(\cdot) = a\Psi\left(\frac{\cdot}{2}\right)$ where $a > 0$ is fixed to a small enough value such that

$$\sum_{|\alpha|=\lfloor\beta\rfloor} \|\partial^\alpha \psi\|_2^2 + \mathbf{1}_{\nu>0} \sum_{|\alpha|=\lfloor\beta\rfloor} \|\partial^\alpha \psi\|_{\mathcal{H}_\nu}^2 \leq \epsilon \ . \tag{106}$$

All the other quantities are defined from this new ψ as in Appendix C.1.

The entire proof of Theorem 5.1 remains unchanged except for one detail : we first to check that the new family of densities (f_θ) is in $\mathcal{S}_L^p(\beta)$ for non-integer β's. We separate two cases :

- $\beta \geq 1$: Let $\theta \in \{1,\dots,m^d\}$, and let $|\alpha| = \lfloor\beta\rfloor$. With the same reasoning steps as in (44), we obtain that

$$\begin{aligned}
\|\partial^\alpha f_\theta\|_2^2 &= \int_{[0,1]^d} \left(h^\beta \sum_{i=1}^{m^d} \theta_i \left(x \mapsto \psi\left(\frac{x-p_i}{h}\right)\right)^{(\alpha)}\right)^2 \\
&= \int_{[0,1]^d} h^{2\nu} \left(\sum_{i=1}^{m^d} \theta_i \psi^{(\alpha)}\left(\frac{\cdot - p_i}{h}\right)\right)^2 \\
&\overset{\text{disjoint supports}}{=} h^{2\nu} \sum_{i=1}^{m^d} \theta_i \int_{[0,1]^d} \left(\psi^{(\alpha)}\left(\frac{\cdot - p_i}{h}\right)\right)^2 \\
&\overset{\|\theta\|_1 \leq m^d \,\&\, \text{variable swap}}{\leq} h^{2\nu} m^d h^d \int_{[0,1]^d} \left(\psi^{(\alpha)}\right)^2 \\
&\overset{m^d h^d \leq 1}{\leq} h^{2\nu} \|\partial^\alpha \psi\|_2^2 \overset{h\leq 1}{\leq} \|\partial^\alpha \psi\|_2^2 \ .
\end{aligned} \tag{107}$$

  In order to control $\|\partial^\alpha f_\theta\|_{\mathcal{H}_\nu}^2$, we will need the following lemma :

  Lemma E.3. *If g_1 and g_2 are continuous with compact supports and if their supports are disjoint, then*

$$\|g_1 + g_2\|_{\mathcal{H}_\nu} \leq \max\left\{\|g_1\|_{\mathcal{H}_\nu}, \|g_2\|_{\mathcal{H}_\nu}\right\} \ .$$

  Proof. Let $x \neq y \in [0,1]^d$. We will upper-bound the Hölderian ratio $\frac{|(g_1+g_2)(x)-(g_1+g_2)(y)|}{|x-y|^\nu}$ by a Hölderian ratio depending only on g_1 or g_2. If x and y both live in the support of either g_1 or g_2, then we may rewrite, in the case where it is in the support of g_1,

$$\frac{|(g_1+g_2)(x)-(g_1+g_2)(y)|}{|x-y|^\nu} \leq \frac{|g_1(x)-g_1(y)|}{|x-y|^\nu} \leq \|g_1\|_{\mathcal{H}_\nu} \ . \tag{108}$$

  Alternatively, the case when it is in the support of g_2 gives the majoration by $\|g_2\|_{\mathcal{H}_\nu}$.

  Now let us look at the case where x and y do not both live in the support of either g_1 or g_2. Let us suppose that $g_1(x) \geq g_2(y)$, the other case being treated in the same fashion. Since g_1 and g_2 have disjoint supports, there exists $t \in (0,1)$ such that $g_1(tx+(1-t)y) = g_2(tx+(1-t)y) = 0$ (connexity argument). Now, by the intermediate values theorem (g_1 is continuous), there exists $t' \in [0,t]$ such that $g_1(t'x+(1-t')y) = g_2(y)$. We thus obtain that

$$\begin{aligned}
\frac{|(g_1+g_2)(x)-(g_1+g_2)(y)|}{|x-y|^\nu} &= \frac{|g_1(x)-g_2(y)|}{|x-y|^\nu} \\
&= \frac{|g_1(x)-g_1(t'x+(1-t')y)|}{|x-y|^\nu} \\
&\leq \frac{|g_1(x)-g_1(t'x+(1-t')y)|}{|x-(t'x+(1-t')y)|^\nu} \\
&\leq \|g_1\|_{\mathcal{H}_\nu} \ .
\end{aligned} \tag{109}$$

The other case leads to a majoration by $\|g_2\|_{\mathcal{H}_\nu}$. All in all, this proves that for any $x \neq y$,

$$\frac{|(g_1+g_2)(x)-(g_1+g_2)(y)|}{|x-y|^\nu} \leq \max\left\{\|g_1\|_{\mathcal{H}_\nu}, \|g_2\|_{\mathcal{H}_\nu}\right\} ,$$

and taking the supremum on the left-hand side yields the desired result. □

Back to our problem, we may write that

$$\begin{aligned}
\|\partial^\alpha f_\theta\|_{\mathcal{H}_\nu} &= \left\| h^\beta \sum_{i=1}^{m^d} \theta_i \left(x \mapsto \psi\left(\frac{x-p_i}{h}\right)\right)^{(\alpha)} \right\|_{\mathcal{H}_\nu} \\
&= h^\nu \left\| \sum_{i=1}^{m^d} \theta_i \psi^{(\alpha)}\left(\frac{\cdot - p_i}{h}\right) \right\|_{\mathcal{H}_\nu} \\
&\overset{Lemma\ E.3}{\leq} h^\nu \max_i \left\{ \left\| \psi^{(\alpha)}\left(\frac{\cdot-p_i}{h}\right)\right\|_{\mathcal{H}_\nu} \right\} \\
&= h^\nu \max_i \left\{ \sup_{x\neq y} \frac{\psi^{(\alpha)}\left(\frac{x-p_i}{h}\right) - \psi^{(\alpha)}\left(\frac{y-p_i}{h}\right)}{|x-y|^\nu} \right\} \\
&= h^\nu \max_i \left\{ h^{-\nu} \sup_{x\neq y} \frac{\psi^{(\alpha)}(x) - \psi^{(\alpha)}(y)}{|x-y|^\nu} \right\} \\
&= \|\partial^\alpha \psi\|_{\mathcal{H}_\nu} .
\end{aligned} \tag{110}$$

So all in all, fixing $\epsilon = L^2$ ensures that for any θ,

$$\sum_{|\alpha|=\lfloor\beta\rfloor} \|\partial^\alpha f_\theta\|_2^2 + \mathbf{1}_{\nu>0} \sum_{|\alpha|=\lfloor\beta\rfloor} \|\partial^\alpha f_\theta\|_{\mathcal{H}_\nu}^2 \leq L^2 . \tag{111}$$

- $\beta \in (0,1)$: When $\beta < 1$, there is one extra technical detail to consider : Since no integer derivative is performed, the constant parts in the densities (f_θ) do not vanish. This is not a problem for the Hölderian part since the seminorm $\|\cdot\|_{\mathcal{H}_\nu}$ is unchanged up to the addition or removal of a constant function. For the sobolev par on the other hand, we may use that $\|g_1+g_2\|^2 \leq (1+\eta)\|g_1\|^2 + (1+1/\eta)\|g_2\|^2$ for any $g_1, g_2 \in L^2$ and any $\eta > 0$, which gives that

$$\|f_\theta\|_2^2 + \|f_\theta\|_{\mathcal{H}_\nu}^2 \leq L^2 . \tag{112}$$

when applied with g_1 the constant part of f_θ, g_2 the part with the kernels, $\eta = \frac{L^2-1}{2}$, and ϵ that satisfies $(1+2/\eta)\,\epsilon \leq \frac{L^2-1}{2}$. Obviously, this only holds if $L > 1$. However, since $\beta \in (0,1)$, Jensen's inequality already implies that any density of probability g satisfies $\|g\|_2^2 + \|g\|_{\mathcal{H}_\nu}^2 \geq \|g\|_2^2 \geq 1$, with equality if and only if g is the density of the uniform distribution. So $L > 1$ is not restrictive on non-trivial classes of distributions $\mathcal{S}_L^p(\beta)$.

Now that we have verified that the family of densities (f_θ) is a subset of $\mathcal{S}_L^p(\beta)$, the rest of the proof follows line by line the one of Theorem 5.1 in the case of integer-valued β.

F. Technical results

Lemma F.1 (Popoviciu's inequality for multivariate random variables)**.** *Let X be a random variable in $\mathbb{R}^{d'}$. If there exist μ and σ such that $\|X-\mu\| \leq \sigma$ almost-surely, then one has*

$$\mathbb{V}(X) := \mathbb{E}\left(\|X - \mathbb{E}(X)\|^2\right) \leq \sigma^2 , \tag{113}$$

thus allowing to gain a factor 4 compared to the natural majoration $\mathbb{V}(X) \leq 4\sigma^2$. In particular, with the isometric identification $(\mathbb{C}, |\cdot|) \cong (\mathbb{R}^2, \|\cdot\|)$, this allows bounding the variance of a complex random variable.

Proof. $\mathbb{V}(X)$ minimizes the function $t \mapsto \mathbb{E}\left(\|X-t\|^2\right)$. Thus, $\mathbb{V}(X)$ is upper-bounded by the value of the same function in μ, yielding the result. □

Lemma F.2 (Existence of $\mathcal{C}^\infty$ function with support in unit ball of $\mathbb{R}^d$)**.** *The function Ψ from R^d to $[0,+\infty)$ which is defined by*

$$\Psi(x) := \begin{cases} e^{-\frac{1}{1-\|x\|^2}} & \text{if } \|x\| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{114}$$

is in $\mathcal{C}^\infty(\mathbb{R}^d)$ and takes non-negative values.

Proof. By induction, we get that for any $\alpha \in \mathbb{N}^d$, $\partial^\alpha \phi(x) = \frac{P_\alpha(x)}{Q_\alpha(x)} e^{-\frac{1}{1-\|x\|^2}}$ when $\|x\| < 1$ where P_α and Q_α are polynomial expressions (in the coefficients of their input vector) with $Q_\alpha(x) \neq 0$, and immediately $\partial^\alpha \Psi(x) = 0$ when $\|x\| > 1$. This proves that $\partial^\alpha \Psi$ is continuous on $\mathbb{R}^d$ with $\partial^\alpha \Psi(x) = 0$ when $\|x\| \geq 1$ because the exponential term is dominant near the unit circle. Since this holds for any $\alpha \in \mathbb{N}^d$, the result follows. □

Lemma F.3 (Hoeffding's inequalities)**.** *If $X_1, \ldots, X_n$ are independent real-valued random variables such that for any i, $a_i \leq X_i \leq b_i$, then for any $t > 0$,*

$$\mathbb{P}\left(\left|\sum_i (X_i - \mathbb{E}(X_i))\right| > t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right) .$$

As a consequence, if $X_1, \ldots, X_n$ are independent complex-valued random variables such that for any i, $X_i \in B(c_i, r_i)$,

$$\mathbb{P}\left(\left|\sum_i (X_i - \mathbb{E}(X_i))\right| > t\right) \leq 4\exp\left(-\frac{t^2}{4\sum_i r_i^2}\right) .$$

Proof. The first inequality for real-valued random variables is folklore, and its proof may for instance be found in (Tsybakov, 2009). For the claim about complex random variables, we have

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_i (X_i - \mathbb{E}(X_i))\right| > t\right) \\
&= \mathbb{P}\left(\left|\sum_i (X_i - \mathbb{E}(X_i))\right|^2 > t^2\right) \\
&= \mathbb{P}\left(R\left(\sum_i (X_i - \mathbb{E}(X_i))\right)^2 + I\left(\sum_i (X_i - \mathbb{E}(X_i))\right)^2 > t^2\right) \\
&\leq \mathbb{P}\left(R\left(\sum_i (X_i - \mathbb{E}(X_i))\right)^2 > t^2/2\right) + \mathbb{P}\left(I\left(\sum_i (X_i - \mathbb{E}(X_i))\right)^2 > t^2/2\right) \\
&= \mathbb{P}\left(\left(\sum_i (R(X_i) - \mathbb{E}(R(X_i)))\right)^2 > t^2/2\right) + \mathbb{P}\left(\left(\sum_i (I(X_i) - \mathbb{E}(I(X_i)))\right)^2 > t^2/2\right) \\
&\leq 2\exp\left(-\frac{2(t/\sqrt{2})^2}{\sum_i (2r_i)^2}\right) + 2\exp\left(-\frac{2(t/\sqrt{2})^2}{\sum_i (2r_i)^2}\right) ,
\end{aligned}$$

where the last inequality comes from Hoeffding's inequality for real-valued random variables. □

Lemma F.4 (χ^2 concentration)**.** *Let $X_1, \ldots, X_d$ be i.i.d. random variables with distribution $\mathcal{N}(0, \sigma^2)$. Let us define $Z = X_1^2 + \cdots + X_d^2$. Then, for any $\delta > 0$,*

$$\mathbb{P}\left(Z \geq (1+\delta) d\sigma^2\right) \leq \max\left\{e^{-\frac{d\delta^2}{4}}, e^{-\frac{d\delta}{2}}\right\} . \tag{115}$$

Furthermore, the integrated version gives, for any $\delta > 0$,

$$\mathbb{E}\left(\left(Z - (1+\delta)d\sigma^2\right)_+\right) \leq \frac{2\sigma^2}{\delta}e^{-\frac{d\delta^2}{4}} + 2\sigma^2 e^{-\frac{d\delta}{2}} \ . \tag{116}$$

Proof. According to Lemma 1 in (Laurent & Massart, 2000), for any $x > 0$,

$$\mathbb{P}\left(Z \geq d\sigma^2 + 2\sigma^2\sqrt{dx} + 2\sigma^2 x\right) \leq e^{-x} \ . \tag{117}$$

Furthermore, we have $\delta\sigma^2 d = 2\sigma^2\sqrt{dx_1}$ iff $x_1 = \frac{d\delta^2}{4}$ and $\delta\sigma^2 d = 2\sigma^2 x_2$ iff $x_2 = \frac{d\delta}{2}$. By noting $f(x) = 2\sigma^2\sqrt{dx} + 2\sigma^2 x$, we have

$$\begin{aligned}
\mathbb{P}\left(Z \geq (1+\delta)d\sigma^2\right) &\leq \mathbb{P}\left(Z \geq d\sigma^2 + f(\min\{x_1, x_2\})\right) \\
&\leq e^{-\min\{x_1, x_2\}} \\
&= \max\left\{e^{-\frac{d\delta^2}{4}}, e^{-\frac{d\delta}{2}}\right\} \ .
\end{aligned} \tag{118}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}\left(\left(Z - (1+\delta)d\sigma^2\right)_+\right) &\leq \int_{(1+\delta)d\sigma^2}^{+\infty} \mathbb{P}(Z \geq t)dt \\
&= \int_{\delta}^{+\infty} d\sigma^2 \mathbb{P}(Z \geq (1+u)d\sigma^2)du \\
&\leq \int_{\delta}^{+\infty} d\sigma^2 \left(e^{-\frac{du^2}{4}} + e^{-\frac{du}{2}}\right)du \\
&\leq \int_{\delta}^{+\infty} d\sigma^2 \left(\frac{u}{\delta}e^{-\frac{du^2}{4}} + e^{-\frac{du}{2}}\right)du \\
&\leq \frac{2\sigma^2}{\delta}e^{-\frac{d\delta^2}{4}} + 2\sigma^2 e^{-\frac{d\delta}{2}} \ .
\end{aligned} \tag{119}$$

□

Lemma F.5 (Talagrand's inequality (one of many) (From Appendix A in (Comte, 2017))). *Let $n \in \mathbb{N} \setminus \{0\}$, $\mathcal{F}$ be a countable family of real-valued measurable functions and $(X_i)_{i=1\ldots,n}$ be n* independent *random variables taking values in a common Polish space. By noting, for any $f \in \mathcal{F}$,*

$$\nu_n(f) := \frac{1}{n}\sum_{i=1}^{n}(f(X_i) - \mathbb{E}(f(X_i))) \ , \tag{120}$$

if there exist three positive constants M_1, H and v such that

$$\sup_{f\in\mathcal{F}} \|f\|_\infty \leq M_1 \ , \tag{121}$$

$$\mathbb{E}\left(\sup_{f\in\mathcal{F}} |\nu_n(f)|\right) \leq H \ , \tag{122}$$

$$\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\mathbb{V}(f(X_i)) \leq v \ , \tag{123}$$

then for any $\delta > 0$,

$$\mathbb{E}\left(\left(\sup_{f\in\mathcal{F}} |\nu_n(f)|^2 - 2(1+2\delta)H^2\right)_+\right) \leq \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\delta\frac{nH^2}{v}} + \frac{49M_1^2}{K_1K(\delta)^2n^2}e^{-\frac{\sqrt{2}K_1K(\delta)\sqrt{\delta}}{7}\frac{nH}{M_1}}\right) \ , \tag{124}$$

where $(y)_+ := \max\{y, 0\}$, $K_1 := \frac{1}{6}$ and $K(\delta) := \min\{\sqrt{1+\delta} - 1, 1\}$.