Math. Stat. Learn. 6 (2023), 1–50
DOI 10.4171/MSL/37

Phase transitions for support recovery under local differential privacy

Cristina Butucea, Amandine Dubois, and Adrien Saumard

Abstract. We address the problem of variable selection in a high-dimensional but sparse mean model, under the additional constraint that only privatized data are available for inference. The original data are vectors with independent entries having a symmetric, strongly log-concave distribution on $\mathbb{R}$. For this purpose, we adopt a recent generalization of classical minimax theory to the framework of local α-differential privacy. We provide lower and upper bounds on the rate of convergence for the expected Hamming loss over classes of at most s-sparse vectors whose non-zero coordinates are separated from 0 by a constant $a > 0$. As corollaries, we derive necessary and sufficient conditions (up to log factors) for exact recovery and for almost full recovery. When we restrict our attention to non-interactive mechanisms that act independently on each coordinate our lower bound shows that, contrary to the non-private setting, both exact and almost full recovery are impossible whatever the value of a in the high-dimensional regime such that $n\alpha^2/d^2 \lesssim 1$. However, in the regime $n\alpha^2/d^2 \gg \log(d)$ we can exhibit a critical value a^* (up to a logarithmic factor) such that exact and almost full recovery are possible for all $a \gg a^*$ and impossible for $a \leq a^*$. We show that these results can be improved when allowing for all non-interactive (that act globally on all coordinates) locally α-differentially private mechanisms in the sense that phase transitions occur at lower levels.

1. Introduction

We consider the problem of distributed support recovery of the sparse mean of n independent, identically distributed (i.i.d.) random vectors. Precisely, for $i = 1, \dots, n$, the ith data holder observes a random vector $X^i = (X^i_j)_{j=1,\dots,d} \in \mathbb{R}^d$ issued from a rescaled and shifted vector ξ^i: $X^i = \theta + \sigma\xi^i$. The noise is supposed to have independent coordinates ξ^i_j, $j = 1, \dots, d$ identically distributed with a symmetric and strongly log-concave distribution of variance 1 (see Section 1.3 below for definition and details). Note that the standard Gaussian distribution belongs to our model, but

2020 Mathematics Subject Classification. Primary 62G05; Secondary 62G20.
Keywords. Local differential privacy, minimax rates, phase transition, support recovery, variable selection, strong log-concavity.

the symmetric and strongly log-concave probability density functions form a large non-parametric class of functions.

The mean vector θ is assumed to be (s, a)-sparse in the sense that θ belongs to one of the following sets:

$$\begin{aligned}\Theta_d^+(s, a) = \{\theta \in \mathbb{R}^d :\ & \text{there exists a set } S \subseteq \{1, \ldots, d\} \text{ with at most } s \text{ elements} \\ & \text{such that } \theta_j \geq a \text{ for all } j \in S, \text{ and } \theta_j = 0 \text{ for all } j \notin S\},\end{aligned}$$

or

$$\begin{aligned}\Theta_d(s, a) = \{\theta \in \mathbb{R}^d :\ & \text{there exists a set } S \subseteq \{1, \ldots, d\} \text{ with at most } s \text{ elements} \\ & \text{such that } |\theta_j| \geq a \text{ for all } j \in S, \text{ and } \theta_j = 0 \text{ for all } j \notin S\}.\end{aligned}$$

1.1. Differential privacy

Nowadays, a large amount of data, such as internet browsing history, social media activity, location information from smart phones, or medical records, are collected and stored. On the one hand, the analysis of these data can benefit to individuals, companies, or communities such as the scientific one. For instance, companies can use data to improve their products and services, or health data can be used for medical research. On the other hand, people are more and more concerned with the protection of their privacy and may be reluctant to share their sensitive data. In this context, it seems essential to be able to understand the trade-offs between the statistical utility of the collected data and the privacy of individuals from whom these data are obtained. This requires a formal definition of privacy and differential privacy has been adopted by researchers in the computer science, machine learning, and statistics communities as a natural one.

Two kinds of differential privacy are discussed in the literature: central differential privacy which has been introduced by Dwork et al. in [22], and local differential privacy. We will focus in this paper on the second setting but we briefly discuss the difference between central and local privacy. In both settings, n individuals want their privacy to be preserved while their data, which will be denoted $X_1, \ldots, X_n$, are used for statistical analyses. In the central setting, the n data-holders share confidence in a common curator who has access to the original data $X_1, \ldots, X_n$ and use them to generate a private release Z. In a nutshell, central differential privacy ensures that the probability of observing an output does not change much when a single data point of the original database is modified. We refer to [38] for the formal definition of differential privacy in the central setting. In the local setting, data is privatized before it is shared with a data collector: for all $i \in [\![1, n]\!]$, X_i is transformed into a private data Z_i directly on the ith individual's machine and the data collector or the statistician only have access to the private sample $Z_1, \dots, Z_n$. However, some interaction between the different data-holders is allowed. Formally, the privatized data $Z_1, \dots, Z_n$ are obtained by successively applying suitable Markov kernels: given $X_i = x_i$ and $Z_1 = z_1, \dots, Z_{i-1} = z_{i-1}$, the ith data holder draws

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

for some Markov kernel $Q_i \colon \mathscr{Z} \times \mathcal{X} \times \mathcal{Z}^{i-1} \to [0,1]$ where the measure spaces of the non-private and private data are denoted with $(\mathcal{X}, \mathscr{X})$ and $(\mathcal{Z}, \mathscr{Z})$, respectively. Such randomizations are known as sequentially interactive. We say that the sequence of Markov kernels $(Q_i)_{i=1,\dots,n}$ provides α-local differential privacy or that $Z_1, \dots, Z_n$ are α-local differentially private views of $X_1, \dots, X_n$ if

$$\sup_{A \in \mathscr{Z}} \frac{Q_i(A \mid X_i = x, Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})}{Q_i(A \mid X_i = x', Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})} \leq \exp(\alpha) \tag{1}$$

for all $i \in [\![1, n]\!]$, and for all $x, x' \in \mathcal{X}$. In this paper, we will focus on the special case of non-interactive local differential privacy where Z_i depends only on X_i but not on Z_k for $k < i$. In this scenario, we have

$$Z_i \sim Q_i(\cdot \mid X_i = x_i),$$

and condition (1) becomes

$$\sup_{A \in \mathscr{Z}} \frac{Q_i(A \mid X_i = x)}{Q_i(A \mid X_i = x')} \leq \exp(\alpha)$$

for all $i \in [\![1, n]\!]$, and for all $x, x' \in \mathcal{X}$.

The aim is that every data holder releases a private view Z^i of X^i such that the notion of local differential privacy is satisfied and that the support of θ can be estimated from the data $Z^1, \dots, Z^n$ in an optimal way.

Notation. For two sequences $\{a_d\}_d$ and $\{b_d\}_d$ of non-negative real numbers, we write $a_d \lesssim b_d$ if there exists some constant $C > 0$ such that $a_d \leq C b_d$. If $b_d > 0$, we write $a_d \sim b_d$ if $a_d / b_d \to 1$ as $d \to \infty$, and we write $a_d \gg b_d$ if $a_d / b_d \to \infty$ as $d \to \infty$. We recall that a centered Laplace distribution with parameter $\lambda > 0$ has the probability density function defined by $f_\lambda(x) = \frac{1}{2\lambda} \exp(-\frac{|x|}{\lambda})$ on $\mathbb{R}$.

1.2. Motivation

The problem of high-dimensional sparse vectors estimation has recently been studied in the framework of local differential privacy in [19]. For the 1-sparse mean estimation problem, the authors considered the set of distributions P supported on $\mathbb{B}_\infty(r)$, i.e., the ball of radius r in $\mathbb{R}^d$ with respect to the sup norm $\|\cdot\|_\infty$, and having $\|\mathbb{E}_P[X]\|_0 \leq 1$. They proved that the private minimax mean squared error for non-interactive α-locally differentially private mechanisms is bounded from below by

$$\min\left\{r^2, \frac{r^2 d \log(2d)}{n(e^\alpha - 1)^2}\right\},$$

proving that high-dimensional 1-sparse mean estimation is impossible in this setting when both $r^2 \gtrsim 1$ and $r^2 d \log(2d) \gtrsim n(e^\alpha - 1)^2$. This result can be related to selecting the support of a 1-sparse mean vector of such a distribution P, under the same constraints. We generalize these results to symmetric and strongly log-concave distributions on the whole $\mathbb{R}^d$ and to arbitrary sparsity.

Obvious applications of variable selection are the estimation of the set that supports the non-null coefficients in the mean vector, or the estimation of its size. We propose to use our procedure to build a private mean estimator of s-sparse vectors in two steps: use one part of the sample to recover the support and the other part to estimate the mean values of the selected variables, that is a vector of reduced size. Moreover, these results are a benchmark for working on more realistic models such as high-dimensional linear regression and clustering of high-dimensional vectors, see [31] and [30].

1.3. Strongly log-concave distributions

Log-concave measures play a significant role in many areas of pure and applied mathematics, such as convex geometry [23], functional inequalities [7], optimal transport theory [13, 14], random matrix theory [1], Monte-Carlo sampling [17, 20], Bayesian inference [32] or non-parametric estimation [16, 18, 24]. The log-concavity assumption arises also naturally in various modelization contexts, such as survival and reliability analysis [27] or econometrics [4], since it possesses many interesting properties subject to interpretation, such as monotone likelihood ratio or non-decreasing hazard rate function for instance. For further applications and references, see [4, 34].

Let us now state the definitions related to log-concavity that will be in force in this article. A probability distribution P on $\mathbb{R}$ is log-concave if it admits a density p with respect to the Lebesgue measure, that writes $p = \exp(-\phi)$, with ϕ a convex function on $\mathbb{R}$. The function ϕ is called the potential of the density p and of the probability measure P.

Furthermore, a function $\phi\colon \mathbb{R} \to \mathbb{R}$ is c-strongly convex for some constant $c > 0$ if, for all $(x, y) \in \mathbb{R}^2$ and $t \in (0, 1)$, we have

$$\phi(tx + (1-t)y) - \left[t\phi(x) + (1-t)\phi(y)\right] \leq -\frac{c}{2}t(1-t)(x-y)^2. \qquad (2)$$

Note that the parameter c in (2) gives a positive lower bound on the *curvature* of the convex function ϕ. In the case where the function ϕ is two times differentiable, condition (2) indeed corresponds to a lower bound on the second derivative:

$$\inf_{x\in\mathbb{R}}\{\phi''(x)\} \geq c > 0.$$

A probability measure P is said to be c-strongly log-concave if it admits a density function $p\colon \mathbb{R} \to (0, +\infty)$ which is c-strongly log-concave with potential ϕ, in the sense that $p = \exp(-\phi)$ and ϕ is a c-strongly convex potential. This is equivalent to assuming that

$$p(x) = \exp(-\phi_0(x))\exp(-cx^2/2)$$

for all $x \in \mathbb{R}$, with ϕ_0 being a finite convex function.

We consider the problem of support recovery of the sparse mean θ of a random vector $X = \theta + \sigma\xi$ of distribution P_θ, where ξ has i.i.d. coordinates ξ_j, $j = 1, \dots, d$, distributed according to a c-strongly log-concave distribution P^{ξ_1} for some constant $c > 0$, with unit variance and that is symmetric around zero. As ξ_1 is assumed to be symmetric, this amounts to require that the c-strongly convex potential ϕ of p is even, or again that $p(x) = \exp(-\phi_0(x))\exp(-cx^2/2)$ for all $x \in \mathbb{R}$, where ϕ_0 is a finite even convex function.

When dealing with some minimax lower bounds in the sequel, we will need to assume that the normalized noise distribution p is not too peaked around its mean, in the sense that its curvature is bounded from above. More precisely, we will assume in this case that $p = \exp(-\phi)$, where ϕ is a finite convex potential, for a constant $c_+ > 0$, for all $(x, y) \in \mathbb{R}^2$ and $t \in (0, 1)$, satisfying

$$\phi(tx + (1-t)y) - \left[t\phi(x) + (1-t)\phi(y)\right] \geq -\frac{c_+}{2}t(1-t)(x-y)^2. \tag{3}$$

When the potential ϕ is two times differentiable, condition (3) can be equivalently formulated as an upper bound on the second derivative of ϕ:

$$\sup_{x\in\mathbb{R}}\{\phi''\} \leq c_+.$$

Such framework provides a non-parametric generalization of the Gaussian assumption, where ϕ_0 would be assumed to be a constant function and the unit variance of ξ would correspond to the value $c = c_+ = 1$. Note that when ξ is only assumed to be centered and strongly log-concave, with unit variance and scaling parameter c, we have in general $c \leq 1$ and the equality case $c = 1$ characterizes the normal distribution $\mathcal{N}(0, 1)$, see [25]. Informally speaking, this means that the Gaussian distribution is the most peaked among strongly log-concave distributions with a fixed variance and thus, it corresponds to the easiest estimation case for support recovery. Finally, let us denote Φ the cumulative distribution function of the normal distribution.

1.4. Minimax framework

Let X^i, $i = 1, \ldots, n$ be i.i.d. random vectors of $\mathbb{R}^d$ with distribution P_θ. We assume that the vectors $X^i = (X^i_j)_{j=1,\ldots,d}$ for $i = 1, \ldots, n$ are observed by n distinct data holders who refuse to share their respective observations. The statistician does not have access to these data but only to α-locally differentially private views $Z^1, \ldots, Z^n$. We assume that θ belongs to one of the sets $\Theta_d^+(s, a)$ or $\Theta_d(s, a)$ introduced in Section 1.1 and we study the problem of selecting the relevant components of θ, that is, of estimating the vector

$$\eta = \eta(P_\theta) = \big(I(\theta_j \neq 0)\big)_{j=1,\ldots,d},$$

where $I(\cdot)$ is the indicator function. Our goal is to estimate the vector η by a *selector* $\widehat{\eta}$, that is a measurable function $\widehat{\eta} = \widehat{\eta}(Z^1, \ldots, Z^n)$ taking values in $\{0, 1\}^d$, where $Z^1, \ldots, Z^n$ are α-locally differentially private views of $X^1, \ldots, X^n$. We judge the quality of a selector $\widehat{\eta}$ as an estimator of η by the Hamming loss between $\widehat{\eta}$ and η which counts the number of positions at which $\widehat{\eta}$ and η differ:

$$|\widehat{\eta} - \eta| := \sum_{j=1}^{d} |\widehat{\eta}_j - \eta_j| = \sum_{j=1}^{d} I(\widehat{\eta}_j \neq \eta_j).$$

For the support recovery problem, we consider only α-locally differentially private mechanisms which transform each $X^i \in \mathbb{R}^d$ into a private release Z^i taking also values in $\mathbb{R}^d$, that are known as non-interactive privacy mechanisms. However, we distinguish between privacy mechanisms that act on each coordinate of X^i either separately, locally or globally. More specifically, we will consider the two following scenarios:

Coordinate Local (CL) privacy mechanisms. There is a sequence $Q = (Q^i)_{i=1,\ldots,n}$ of Markov kernels providing α-local differential privacy such that

$$Z^i \sim Q^i(\cdot \mid X^i = x^i)$$

for all $i \in [\![1, n]\!]$, and Q^i is obtained as product of coordinate-wise kernels as follows:

$$\forall i \in [\![1, n]\!],\ j \in [\![1, d]\!], \quad Z^i_j \sim Q^i_j(\cdot \mid X^i_j = x)$$

for some (α/d)-differentially private mechanism Q^i_j. We denote by $\mathcal{Q}^{CL}_\alpha$ the set of all privacy mechanisms $Q = (Q^1, \ldots, Q^n)$ satisfying these assumptions.

Coordinate Global (CG) privacy mechanisms. There is a sequence $Q = (Q^i)_{i=1,\ldots,n}$ of Markov kernels providing α-local differential privacy such that

$$Z^i \sim Q^i(\cdot \mid X^i = x^i)$$

for all $i \in [\![1,n]\!]$. We denote by $\mathcal{Q}_\alpha$ the set of all privacy mechanisms $Q = (Q^1, \dots, Q^n)$ satisfying this assumption.

In other words, in the Coordinate Local case, we consider only non-interactive α-locally differentially private mechanisms that act coordinates by coordinates. This scenario is easier to study than the second one for which any non-interactive α-locally differentially private mechanism is allowed to be used.

For both scenarios, if P_θ denotes the distribution of X^i then we denote by $Q^i P_\theta$ the distribution of Z^i. Since the distribution of $(X^1, \dots, X^n)$ is $P_\theta^{\otimes n}$, the distribution of $(Z^1, \dots, Z^n)$ will be denoted by $Q(P_\theta^{\otimes n})$. In the Coordinate Local case, we denote by P_{θ_j} the distribution of X_j^i and by $Q_j^i P_{\theta_j}$ the distribution of Z_j^i.

We say that a selector $\widehat{\eta} = (\widehat{\eta}_1, \dots, \widehat{\eta}_d)$ is *separable* if for all $j = 1, \dots, d$ its jth component $\widehat{\eta}_j$ depends only on $(Z_j^i)_{i=1,\dots,n}$. We denote by $\mathcal{T}$ the set of all separable selectors. We are interested in the study of the following private minimax risks

$$\mathcal{R}_n^{CL}(\alpha, \Theta) = \inf_{Q \in \mathcal{Q}_\alpha^{CL}} \inf_{\widehat{\eta} = \widehat{\eta}(Z^1,\dots,Z^n) \in \mathcal{T}} \sup_{\theta \in \Theta} \frac{1}{s} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\widehat{\eta}(Z^1, \dots, Z^n) - \eta|, \qquad (4)$$

in the coordinate local case, and

$$\mathcal{R}_n(\alpha, \Theta) = \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\widehat{\eta} = \widehat{\eta}(Z^1,\dots,Z^n) \in \mathcal{T}} \sup_{\theta \in \Theta} \frac{1}{s} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\widehat{\eta}(Z^1, \dots, Z^n) - \eta|, \qquad (5)$$

in the coordinate global case, for $\Theta = \Theta_d^+(s, a)$ and $\Theta = \Theta_d(s, a)$.

We are interested in the study of two asymptotic properties: *almost full recovery* and *exact recovery*, that we define here. Let $(\Theta_d^+(s_d, a_d))_{d \geq 1}$ be a sequence of classes of sparse vectors. We will say that *almost full recovery is possible* for $(\Theta_d^+(s_d, a_d))_{d \geq 1}$ in the Coordinate Local case if there exists $Q \in \mathcal{Q}_\alpha^{CL}$ and a selector $\widehat{\eta}$ such that

$$\lim_{d \to \infty} \sup_{\theta \in \Theta_d^+(s_d, a_d)} \frac{1}{s_d} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\widehat{\eta} - \eta| = 0.$$

We will say that *almost full recovery is impossible* for $(\Theta_d^+(s_d, a_d))_{d \geq 1}$ in the Coordinate Local case if

$$\liminf_{d \to +\infty} \inf_{Q \in \mathcal{Q}_\alpha^{CL}} \inf_{\widehat{\eta} = \widehat{\eta}(Z^1,\dots,Z^n) \in \mathcal{T}} \sup_{\theta \in \Theta_d^+(s,a)} \frac{1}{s_d} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\widehat{\eta} - \eta| > 0.$$

We will say that *exact recovery is possible* for $(\Theta_d^+(s_d, a_d))_{d \geq 1}$ in the Coordinate Local case if there exists $Q \in \mathcal{Q}_\alpha^{CL}$ and a selector $\widehat{\eta}$ such that

$$\lim_{d \to \infty} \sup_{\theta \in \Theta_d^+(s_d, a_d)} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\widehat{\eta} - \eta| = 0.$$

We will say that *exact recovery is impossible* for $(\Theta_d^+(s_d, a_d))_{d\geq 1}$ in the Coordinate Local case if

$$\liminf_{d\to+\infty} \inf_{Q\in\mathcal{Q}_\alpha^{CL}} \inf_{\hat{\eta}=\hat{\eta}(Z^1,\ldots,Z^n)\in\mathcal{T}} \sup_{\theta\in\Theta_d^+(s,a)} \mathbb{E}_{Q(P_\theta^{\otimes n})}|\hat{\eta}-\eta| > 0.$$

We use similar definitions in the Coordinate Global case with $\mathcal{Q}_\alpha^{CL}$ replaced by $\mathcal{Q}_\alpha$.

1.5. Related work

Variable selection with Hamming loss in the Gaussian mean model in $\mathbb{R}^d$ has been studied in the non-private setting in [11]. The authors provide non-asymptotic lower and upper bounds on the non-private version of minimax risk (4). As corollaries, they derive necessary and sufficient conditions for almost full recovery and exact recovery to be possible. If s, $d \to \infty$ such that $s/d \to 0$, they highlight a critical value

$$a^* = (\sigma/\sqrt{n})\sqrt{2\log(d/s-1)}(1+\delta)$$

for a specific sequence $\delta = \delta(d,s) \to 0$ such that almost full recovery is possible for $a \geq a^*$ and impossible for $a < a^*$. Similar results have been obtained for exact recovery with the greater critical value

$$a^* = (\sigma/\sqrt{n})\big(\sqrt{2\log(d-s)} + \sqrt{2\log s}\big).$$

In the present paper, we will see how these results are affected by the privacy constraints.

For estimating the 1-sparse mean of high-dimensional vectors with distribution supported on a compact support it is known that the rates are deteriorates by a factor d under local differential privacy, see [19]. Under a relaxation of central differential privacy called (α,δ)-approximate differential privacy; see, for instance, [2,6,21] have provided estimators of the mean and the covariance of high-dimensional Gaussian vectors and theoretical guarantees that do not require additional assumptions on the parameters. In some regimes the rates are not deteriorated and it is therefore difficult to anticipate the role of privacy on each particular problem.

A few papers tackle a slightly different selection problem under privacy constraints mostly under central differential privacy constraints. They are interested in the largest sum of k coordinates of the common mean value θ of a vector supported on $\{0,1\}^d$. We are mainly interested in recovering the position of significant coordinates in the s-sparse mean vector θ.

In [35], the authors study top$-k$ selection under a relaxation of central differential privacy called (α,δ)-approximate differential privacy. However, they use a weighted Hamming loss as described below. Precisely, if $X_1,\ldots,X_n$ are drawn i.i.d. from some distribution P on $\{0,1\}^d$, they want to find the k greatest coordinates of the mean vector $\theta = \mathbb{E}_P[X_1]$ while respecting (α,δ)-differential privacy constraints. They prove the existence of a $(1, 1/(nd))$-differentially private mechanism that outputs $Z \in \{0,1\}^d$ with k non-zero coordinates such that

$$\mathbb{E}\left[\sum_{j=1}^{d} \theta_j I(Z_j = 1)\right] \geq \max_{\eta \in \{0,1\}^d : \|\eta\|_1 = k} \sum_{j=1}^{d} \theta_j I(\eta_j = 1) - \beta$$

requires $n \gtrsim \sqrt{k} \log d$ samples in the low accuracy regime where $\beta = k/10$. Moreover, repeated use of the classical exponential mechanism solves this problem with $n = O(\sqrt{k} \log d)$ samples. In [3], the authors study an empirical version of the problem studied in [35]: they want to find the top-k coordinates of the vector $q \in \mathbb{R}^d$ defined by

$$q_j = \frac{1}{n} \sum_{i=1}^{n} X_{i,j}, \quad j = 1, \ldots, d,$$

while respecting (α,δ)-differential privacy constraints. Let τ be the kth largest value among the coordinates $\{q_1, \ldots, q_k\}$. They prove the existence of a (α,δ)-differentially private mechanism that outputs a set $S \subset [\![1,d]\!]$ of k elements such that $q_j \geq \tau - \beta$ for all $j \in S$ requires $n \gtrsim k \log(d)$ samples in the high-accuracy regime where $\beta \asymp \sqrt{\log d/n}$. In [37], the author studies the same problem as [35] for $k = 1$ under non-interactive α-local differential privacy constraints. If we consider the low-accuracy regime considered by [35], this result shows that estimating the largest coordinate of a 1-sparse mean θ under non-interactive α-local differential privacy requires $n \gtrsim d \log d/\alpha^2$ samples, which is by a factor d larger than in the central model of (α,δ)-approximate differential privacy.

1.6. Description of results

We address the problem of variable selection in a symmetric, strongly log-concave model in $\mathbb{R}^d$ under local differential privacy constraints. We provide lower and upper bounds on the rate of convergence for the expected Hamming loss over classes of at most s-sparse vectors whose non-zero coordinates are separated from 0 by a constant $a > 0$.

When we restrict our attention to non-interactive mechanisms that act independently on each coordinate (*coordinate local privacy mechanisms*) we have proved that, contrary to the non-private setting, almost full recovery and exact recovery are impossible whatever the value of a in the high-dimensional regime when $n\alpha^2 \lesssim d^2$. This is due to the fact that the loss of information due to privacy may reduce the effective sample size $N := n\alpha^2/d^2$ under the value 1, and this does not allow support

	$a \lesssim \frac{\sigma}{\sqrt{N}}$	$\frac{\sigma}{\sqrt{N}} \ll a \leq 2\sigma$	$a \geq 2\sigma$
$N := \frac{n\alpha^2}{d^2} \lesssim 1$	impossible	impossible	impossible
$N := \frac{n\alpha^2}{d^2} \gg 1$	impossible	possible, as soon as $a \gg \frac{\sigma}{\sqrt{N}}\sqrt{\log(d)}$ if, moreover, $N \gg \log(d)$	possible if $\frac{\log(d)}{N} \lesssim 1$

Table 1. Exact recovery of θ in either $\Theta_d^+(s, a)$ or $\Theta_d(s, a)$ in the Coordinate Local case. Similar results hold for almost full recovery with $\log(d)$ replaced by $\log(d/s)$.

recovery neither exact nor almost full. This result is significantly different from the non-private case where [11] shows that variable selection is always possible, even for $n = 1$ observation for significant enough mean value a.

However, in the regime $n\alpha^2/d^2 \gg \log(d)$ we exhibit a critical value a^* (up to a logarithmic factor) such that exact recovery is possible for all $a \gg a^*$ and impossible for all $a \leq a^*$. We also prove that these results can be improved when allowing for all non-interactive locally differentially private mechanisms, that we also call *coordinate global*. The effective sample size is Nd in this case and it is larger than N.

Let us note that the separable selectors that we propose are free of the sparsity parameter s. They depend on a and methods could be made adaptive to a, but this is beyond the scope of this work.

For many estimation problems, allowing for sequentially interactive privacy mechanisms, that randomize each vector X_i by using also the publicly available information $Z_1, \dots, Z_{i-1}, i = 2, \dots, n$, does not improve substantially over non-interactive minimax rates. This includes, for instance, density estimation [10], one-dimensional mean estimation [19], and estimation of a linear functional of the true distribution [33]. However, for some estimation problems (see for instance the estimation of the integrated square of a density, [12]) and some testing problems (see [5] and [12]) allowing for sequentially interaction between data-holders can substantially improve over non-interactive minimax rates of estimation or non-interactive minimax rates of testing. We consider here only non-interactive privacy mechanisms for each vector X_i, but we conjecture that the exact and almost full recovery would be improved for interactive privacy mechanisms. It is left for future work to study whether that is indeed the case.

The paper is organized as follows. In Section 2, we study the minimax risk (4). We first provide a lower bound which enables us to derive necessary conditions for almost full recovery and exact recovery to be possible in the case where only coordinate local privacy mechanisms are used. In particular, we prove that almost full recovery

	$a \lesssim \sigma\sqrt{\frac{\log d}{Nd}}$	$\sigma\sqrt{\frac{\log d}{Nd}} \ll a \le 2\sigma$	$a \ge 2\sigma$
$\frac{Nd}{\log d} \lesssim 1$	impossible	impossible	impossible if $a \le \sigma\sqrt{\log\left(1 + \frac{\log d}{16Nd}\right)}$
$\frac{Nd}{\log d} \gg 1$	impossible	possible, as soon as $a \gg \sigma\sqrt{\frac{\log d}{Nd}}$ if, moreover, $Nd \gg \log(d)$	possible

Table 2. Exact recovery of θ in either $\Theta_d^+(s,a)$ or $\Theta_d(s,a)$ in the Coordinate Global case. We have set $N = n\alpha^2/d^2$ for a better comparison with the Coordinate Local case.

is impossible in this case as soon as the quantity $n\alpha^2/d^2$ is bounded from above. We then provide non-asymptotic upper bounds on the minimax risks in propositions and state more explicit asymptotic sufficient conditions for almost full recovery and exact recovery to be possible in our corollaries. These conditions and associated results are summarized in Table 1. In Section 3, we study the minimax risk (5) and prove that the results of Section 2 can be improved when any non-interactive (coordinate global) α-locally differentially private mechanism is allowed. See Table 2 for a summary of these results. Detailed proofs can be found in the appendix.

2. Coordinate local non-interactive privacy mechanisms

In this section, we provide a lower bound on the private minimax risk (4). This enables us to obtain necessary conditions for almost full recovery and exact recovery to be possible in the Coordinate Local scenario. In particular, we prove that almost full recovery is impossible in the private setting of the Coordinate Local case if the quantity $N := n\alpha^2/d^2$ is bounded from above. We then provide upper bounds on the minimax risk that entail sufficient conditions for almost full recovery and exact recovery to be possible.

2.1. Lower bound

We first state our lower bound.

Theorem 2.1. *Assume that the measure P^{ξ_1} of the noise coordinates, is log-concave with a density $p = \exp(-\phi)$, where the potential ϕ has a curvature bounded from* *above by a constant $c_+ > 0$, that satisfies inequality* (3). *Then for any $a > 0$, $\alpha > 0$, $1 \leq s \leq d$, $n \geq 1$, we have*

$$\mathcal{R}_n^{CL}(\alpha, \Theta_d^+(s,a)) \geq \left(1 - \frac{s}{d}\right) \exp\left(-4n(e^{\alpha/d} - 1)^2 \min\left\{\frac{c_+ a^2}{4\sigma^2}, 1\right\}\right). \tag{6}$$

The proof of Theorem 2.1 can be found in Appendix A.2. Some auxiliary results used for the proof of Theorem 2.1 can be found in Appendix A.1. Note that since $\Theta_d^+(s,a) \subset \Theta_d(s,a)$, we have

$$\mathcal{R}_n^{CL}(\alpha, \Theta_d^+(s,a)) \leq \mathcal{R}_n^{CL}(\alpha, \Theta_d(s,a)),$$

thus the right-hand side of (6) is also a lower bound for $\mathcal{R}_n^{CL}(\alpha, \Theta_d(s,a))$.

A careful look at the proof of Theorem 2.1 shows that log-concavity is in fact not needed in the previous result, if we assume the existence of a positive density, converging to zero at infinity, and with a two times continuously differentiable potential achieving (3).

For better confidentiality in practice, the parameter α must not be too large. In particular, we assume that $\alpha/d \to 0$ when $d \to +\infty$. We thus have

$$n(e^{\alpha/d} - 1)^2 \sim n\alpha^2/d^2$$

and Theorem 2.1 immediately shows the following.

Corollary 2.2. *Grant assumptions of Theorem 2.4. Let $\alpha > 0$, $1 \leq s \leq d$, $n \geq 1$ be such that $s/d \leq C_0$ for some constant $C_0 \in (0,1)$, and $\alpha/d \to 0$ when $d \to \infty$. Then, if $n\alpha^2/d^2 \leq C_1$ for some constant $C_1 > 0$ or if $n\alpha^2/d^2 \to \infty$ as $d \to \infty$, and $a^2 \leq C_2\sigma^2 d^2/n\alpha^2$ for some constant $C_2 > 0$ depending only on c_+, it holds that*

$$\mathcal{R}_n^{CL}(\alpha, \Theta) \geq C$$

for some constant $C > 0$, where $\Theta = \Theta_d^+(s,a)$ or $\Theta = \Theta_d(s,a)$.

Corollary 2.2 shows that almost full recovery is impossible under local differential privacy constraints if the quantity $n\alpha^2/d^2$ is bounded from above. In particular, almost full recovery is impossible under local differential privacy constraints in the high-dimensional setting, that is when $n \leq d$, whatever the value of a. Corollary 2.2 also proves that if $n\alpha^2/d^2 \to +\infty$ then almost full recovery is impossible if

$$a \lesssim \sigma d/\sqrt{n\alpha^2}.$$

This underlines a strong difference between the private setting and the classical setting, since [11] proved that in the non-private setting almost full recovery is possible for values of $|a|$ large enough, even if $n = 1$. However, both almost full and exact recovery are impossible for any signal value a when the effective size $N = n\alpha^2/d^2 \lesssim 1$ under privacy constraints.

2.2. Privacy mechanism

In this section, we introduce a non-interactive privacy mechanism creating private views $Z^1, \dots, Z^n$ of the original data $X^1, \dots, X^n$ that satisfy the local differential privacy constraint of level α. These privatized data will then be used to define a private selector whose risk will be studied in Section 2.3.

To obtain the privatized data, we first censor the unbounded random variables X_j^i, for $i = 1, \dots, n$, $j = 1, \dots, d$, and then make use of an appropriately scaled version of the classical Laplace mechanism. For all $i \in [\![1, n]\!]$ and $j \in [\![1, d]\!]$, define

$$Z_j^i = \mathrm{sgn}[X_j^i] + \frac{2d}{\alpha} W_j^i, \tag{7}$$

where $\mathrm{sgn}[x] = 1$, for $x \geq 0$, and 0, for $x < 0$, the W_j^i's are i.i.d. Laplace(1) random variables, and W_j^i is independent from X_j^i.

Note that the privacy mechanism defining $(Z^i)_{i=1,\dots,n}$ is non-interactive since Z^i does only depend on X^i and not on Z^k for $k \neq i$. This is also a coordinate local mechanism since Z_j^i depends on X_j^i but not on the X_l^i for $l \neq j$. The following proposition shows that it satisfies the condition of α-local differential privacy.

Proposition 2.3. *For all $i \in [\![1, n]\!]$ and $j \in [\![1, d]\!]$, Z_j^i is an α/d-differentially private view of X_j^i. Consequently, for all $i \in [\![1, n]\!]$, $Z^i = (Z_j^i)_{j=1,\dots,d}$ is an α-differentially private view of X^i.*

Proof. Set $r = 2d/\alpha$. By definition of the privacy mechanism (7), the conditional density of Z_j^i given $X_j^i = x$ can be written as

$$q^{Z_j^i | X_j^i = x}(z) = \frac{1}{2r} \exp\left(-\frac{|z - \mathrm{sgn}[x]|}{r} \right).$$

Thus, by the reverse and the ordinary triangle inequality, for all $i \in [\![1, n]\!]$, $j \in [\![1, d]\!]$ and all $x, x', z \in \mathbb{R}$, it holds that

$$\begin{aligned} \frac{q^{Z_j^i | X_j^i = x}(z)}{q^{Z_j^i | X_j^i = x'}(z)} &= \exp\left(\frac{|z - \mathrm{sgn}[x']|}{r} - \frac{|z - \mathrm{sgn}[x]|}{r} \right) \\ &\leq \exp\left(\frac{|\mathrm{sgn}[x'] - \mathrm{sgn}[x]|}{r} \right) \leq \exp\left(\frac{2}{r} \right) \leq \exp\left(\frac{\alpha}{d} \right). \end{aligned}$$

This proves that Z_j^i is an α/d-differentially private view of X_j^i. Let us check that Z^i is an α-differentially private view of X^i. Denote by $q^{Z^i | X^i = x}$ the conditional density of Z^i given $X^i = x$ and note, for all $x, x', z \in \mathbb{R}^d$, it holds that

$$\frac{q^{Z^i | X^i = x}(z)}{q^{Z^i | X^i = x'}(z)} = \prod_{j=1}^{d} \frac{q^{Z_j^i | X_j^i = x_j}(z_j)}{q^{Z_j^i | X_j^i = x'_j}(z_j)} \leq e^{\alpha},$$

using the independence of the coordinates $X_1^i, \dots, X_d^i$ and the conditional independence of $Z_1^i, \dots, Z_d^i$ given X^i. ■

2.3. Upper bounds

Using these privatized data, we define two selectors that will provide upper bounds on the minimax risk (4). For the class $\Theta_d^+(s, a)$, we will use the selector $\widehat{\eta}^+$ with the components

$$\widehat{\eta}_j^+ = I\left(\frac{1}{n}\sum_{i=1}^{n} Z_j^i \geq \tau\right), \quad j = 1, \dots, d, \tag{8}$$

where the threshold τ has to be properly chosen, later on. For the class $\Theta_d(s, a)$, we will use the selector $\widehat{\eta}$ with the components

$$\widehat{\eta}_j = I\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_j^i\right| \geq \tau\right), \quad j = 1, \dots, d,$$

where τ to be defined later on. Note that $\widehat{\eta}^+$ and $\widehat{\eta}$ are separable selectors since $\widehat{\eta}_j^+$ and $\widehat{\eta}_j$ depend only on $(Z_j^i)_{i=1,\dots,n}$ and not on the Z_k^i for $k \neq j$. We now study the performances of these selectors. Recall that Φ is a cumulative distribution function of the normal distribution.

Proposition 2.4. *Assume that $a \geq 2\sigma$. Set $C_1 := 2\Phi(2\sqrt{c}) - 1 > 0$. If τ is chosen such that*

$$C_1 - \tau > 0, \quad \tau\alpha/(8d) \leq 1, \quad \text{and} \quad \alpha(C_1 - \tau)/(8d) \leq 1,$$

then it holds, for all $\theta \in \Theta_d^+(s, a)$, that

$$\begin{aligned} \mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+ - \eta|\right] \leq \frac{d - |S|}{s}&\left[\exp\Big(-\frac{n\tau^2}{2^3}\Big) + \exp\Big(-\frac{\tau^2 n\alpha^2}{2^7 d^2}\Big)\right] \\ &+ \frac{|S|}{s}\left[\exp\left(-\frac{n(C_1 - \tau)^2}{2^3}\right) + \exp\left(-\frac{(C_1 - \tau)^2 n\alpha^2}{2^7 d^2}\right)\right], \end{aligned} \tag{9}$$

and, for all $\theta \in \Theta_d(s, a)$, it holds that

$$\begin{aligned} \mathbb{E}\left[\frac{1}{s}|\widehat{\eta} - \eta|\right] \leq 2\frac{d - |S|}{s}&\left[\exp\Big(-\frac{n\tau^2}{2^3}\Big) + \exp\Big(-\frac{\tau^2 n\alpha^2}{2^7 d^2}\Big)\right] \\ &+ 2\frac{|S|}{s}\left[\exp\left(-\frac{n(C_1 - \tau)^2}{2^3}\right) + \exp\left(-\frac{(C_1 - \tau)^2 n\alpha^2}{2^7 d^2}\right)\right], \end{aligned} \tag{10}$$

where S denotes the support of θ.

The proof of Proposition 2.4 is given in Section A.4 in the Appendix. Some auxiliary results used in the proof of Proposition 2.4 can be found in Appendix A.3. The following Corollary gives sufficient conditions so that almost full recovery and exact recovery are possible under local differential privacy in the Coordinate Local case when $a \geq 2\sigma$.

Corollary 2.5. *Set* $C_1 = 2\Phi(2\sqrt{c}) - 1 > 0$. *Assume that*

$$\alpha/d \to 0, \quad n\alpha^2/d^2 \to +\infty, \quad \textit{and} \quad \limsup \frac{\log(d/s)}{n\alpha^2/d^2} < \frac{C_1^2}{2^9}.$$

Then the selector $\widehat{\eta}^+$ *defined by* (8) *with* $\tau = C_1/2$ *satisfies*

$$\sup_{\theta\in\Theta} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1,\dots,Z^d) - \eta| \to 0 \tag{11}$$

for all $a \geq 2\sigma$, *where* $\Theta = \Theta_d^+(s,a)$ *or* $\Theta = \Theta_d(s,a)$. *If, in addition,*

$$\limsup \frac{\log(d)}{n\alpha^2/d^2} < \frac{C_1^2}{2^9},$$

then

$$\sup_{\theta\in\Theta} \mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1,\dots,Z^d) - \eta| \to 0 \tag{12}$$

for all $a \geq 2\sigma$.

The proof of Corollary 2.5 is given in Section A.5 in the appendix. Since we have seen that almost full recovery is impossible when $n\alpha^2/d^2$ is bounded from above or when $n\alpha^2/d^2 \to +\infty$ and $a \lesssim (\sigma d)/(\sqrt{n}\alpha)$, it remains to study the case where $n\alpha^2/d^2 \to +\infty$ and $\sigma d/(\sqrt{n}\alpha) \ll a \leq 2\sigma$. This is done below.

Proposition 2.6. *Let* $a > 0$. *If* τ *is chosen such that* $\tau < 2a/\sigma p(2)$, $\tau\alpha/(8d) < 1$ *and* $\alpha(a/\sigma p(2) - \tau/2)/(4d) \leq 1$, *then, for all* $\theta \in \Theta_d^+(s,a)$, *it holds that*

$$\mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+ - \eta|\right] \leq \frac{d-|S|}{s}\left[\exp\left(-\frac{n\tau^2}{2^3}\right) + \exp\left(-\frac{\tau^2 n\alpha^2}{2^7 d^2}\right)\right] + \frac{|S|}{s}\left[\exp\left(-\frac{n(a/\sigma p(2) - \tau/2)^2}{2^3}\right) + \exp\left(-\frac{(a/\sigma p(2) - \tau/2)^2 n\alpha^2}{2^5 d^2}\right)\right],$$

where S *denotes the support of* θ.

The proof of Proposition 2.6 can be found in Section A.6 in the appendix. Note that as for the case $a \geq 2\sigma$, if $\theta \in \Theta_d(s,a)$ we use $\widehat{\eta}$ instead of $\widehat{\eta}^+$ and we can prove the same result with an extra multiplicative factor 2. The next corollary gives new sufficient conditions so that almost full recovery and exact recovery are possible.

Corollary 2.7. *Assume that $\alpha/d \to 0$, $n\alpha^2/d^2 \to +\infty$ and $\sigma d/(\sqrt{n}\alpha) \ll a \le 2\sigma$. The selector $\widehat{\eta}^+$ defined by (8) with $\tau = p(2)a/\sigma$ satisfies for d large enough*

$$\sup_{\theta \in \Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1,\dots,Z^d) - \eta| \le 2\exp\left(\log\left(\frac{d}{s}\right) - \frac{p^2(2)a^2n\alpha^2}{2^9\sigma^2 d^2}\right).$$

In particular, if $a \gg \frac{\sigma d}{\alpha\sqrt{n}}\log^{1/2}(\frac{d}{s})$, it holds that

$$\sup_{\theta \in \Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1,\dots,Z^d) - \eta| \to 0. \tag{13}$$

Moreover, if $a \gg \frac{\sigma d}{\alpha\sqrt{n}}\log^{1/2}(d)$, then

$$\sup_{\theta \in \Theta_d^+(s,a)} \mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1,\dots,Z^d)) - \eta| \to 0. \tag{14}$$

If $n\alpha^2/d^2 \to \infty$ with $(n\alpha^2/d^2) \gg \log(d/s)$, then Corollary 2.7 combined with Corollary 2.5 and with the lower bound (6) prove a phase transition result (up to log factors) at the value $a^* = a^*(n,\alpha,d,\sigma) = \sigma d/(\alpha\sqrt{n})$. Indeed, we get that almost full recovery is impossible in the Coordinate Local case for all $a \le Ca^*$ and is possible for all $a \gg a^* \log^{1/2}(d/s)$.

3. Coordinate global non-interactive privacy mechanisms

In this section, we study the minimax risk (5). We prove that in the Coordinate Global case, almost full recovery and exact recovery are possible under weaker assumptions than the one we obtained for the Coordinate Local case.

3.1. Privacy mechanism

We describe in this section the privacy mechanism we use to obtain private data that will be used to design a private selector and to obtain upper bounds on the minimax risk (5) in the Coordinate Global case.

For all $i \in [\![1,n]\!]$, the private view Z^i of X^i is obtained using the following steps:

(1) Compute $f(X^i) = (\mathrm{sgn}[X^i_j])_{j=1,\dots,d}$. For short, let us denote $\widetilde{X}^i = f(X^i)$.

(2) Sample $Y^i \sim \mathcal{B}(\pi_\alpha)$ where $\pi_\alpha = e^\alpha/(e^\alpha + 1)$ and generate $\widetilde{Z}^i$ uniformly distributed on the set

$$\{\widetilde{z} \in \{-B,B\}^d \mid \langle \widetilde{z}, \widetilde{X}^i\rangle > 0 \text{ or } (\langle \widetilde{z}, \widetilde{X}^i\rangle = 0 \text{ and } \widetilde{z}_1 = B\widetilde{X}^i_1)\}$$

if $Y^i = 1$, respectively on the set

$$\left\{\widetilde{z} \in \{-B, B\}^d \mid \langle \widetilde{z}, \widetilde{X}^i \rangle < 0 \text{ or } (\langle \widetilde{z}, \widetilde{X}^i \rangle = 0 \text{ and } \widetilde{z}_1 = -B\widetilde{X}_1^i)\right\}$$

if $Y^i = 0$, with

$$B = \frac{e^\alpha + 1}{e^\alpha - 1} K_d, \quad \text{where } \frac{1}{K_d} = \begin{cases} \frac{1}{2^{d-1}} \binom{d-1}{\frac{d-1}{2}} & \text{if } d \text{ is odd,} \\ \frac{(d-2)!(d-2)}{2^{d-1}(\frac{d}{2}-1)!\frac{d}{2}!} & \text{if } d \text{ is even.} \end{cases} \tag{15}$$

(3) Define the vector Z^i by $Z^i = \widetilde{Z}^i$ if d is odd, and by its components

$$Z_j^i = \begin{cases} \frac{d-2}{2(d-1)} \widetilde{Z}_1^i & \text{if } j = 1, \\ \widetilde{Z}_j^i & \forall j \in [\![2, d]\!], \end{cases}$$

if d is even.

This mechanism is strongly inspired by the one proposed by Duchi et al. [19] for mean estimation on the set of distributions P supported on $\mathbb{B}_\infty(r) \subset \mathbb{R}^d$ with $\|\mathbb{E}[X]\|_0 \leq s$. In particular, if d is odd, the event $\{\langle \widetilde{z}, \widetilde{X}^i \rangle = 0\}$ has probability zero for all $\widetilde{z} \in \{-B, B\}^d$ and our mechanism coincides in this case with the one proposed by Duchi et al. [19] applied to $\operatorname{sgn}(X^i)$ instead of X^i.

Proposition 3.1. *For all $i \in [\![1, n]\!]$, Z^i is an α-differentially private view of X^i.*

The following proposition will be useful in the analysis of the selector proposed in Section 3.2.

Proposition 3.2. *For all $i \in [\![1, n]\!]$, it holds that*

$$\mathbb{E}[Z^i \mid X^i] = f(X^i).$$

The proofs of Proposition 3.1 and Proposition 3.2 can be found respectively in Section B.1 and B.2 of the appendix. Note it also holds that $\mathbb{E}[Z^i \mid X^i] = f(X_i)$ when Z_i is produced via the Laplace mechanism described in Section 2.2. However, the variance $\operatorname{Var}(Z_j^i \mid X^i)$ is slower by a multiplicative factor d when Z^i is produced with the Laplace mechanism than when it is obtained with the above coordinate global mechanism. Indeed, if Z^i is produced with the above mechanism, then we have

$$\operatorname{Var}(Z_j^i \mid X^i) \leq B^2.$$

Stirling's approximation yields $K_d^2 \lesssim d$ for d large enough, see Lemma B.1 in Appendix B.3 for details. Thus, if α is bounded, we obtain $\operatorname{Var}(Z_j^i \mid X^i) \lesssim d/\alpha^2$. Now, if Z^i is produced with the Laplace mechanism then it holds that

$$\operatorname{Var}(Z_j^i \mid X^i) = 8d^2/\alpha^2.$$

This faster variance explains that we will obtain better results when allowing for coordinate global mechanisms.

3.2. Upper bounds

Using the privatized data of the previous subsection, we define two selectors that will enable us to obtain upper bounds on the minimax risk (5). As in the Coordinate Local case, for the class $\Theta_d^+(s,a)$, we will use the selector $\widehat{\eta}^+$ with the components

$$\widehat{\eta}_j^+ = I\left(\frac{1}{n}\sum_{i=1}^n Z_j^i \geq \tau\right), \quad j = 1,\dots,d, \tag{16}$$

where the threshold τ has to be chosen. For the class $\Theta_d(s,a)$, we will use the selector $\widehat{\eta}$ with the components

$$\widehat{\eta}_j = I\left(\left|\frac{1}{n}\sum_{i=1}^n Z_j^i\right| \geq \tau\right), \quad j = 1,\dots,d.$$

We now study the performances of these selectors.

The following result gives an upper bound on the risk of selector (16) when $a \geq C\sigma$ and will enable us to obtain sufficient conditions so that almost full recovery is possible when $a \geq 2\sigma$ in the Coordinate Global case.

Proposition 3.3. *Assume that $a > 2\sigma$ and set $C_1 := 2\Phi(2\sqrt{c}) - 1 > 0$. If τ is chosen such that $C_1 - \tau > 0$, then it holds for all $\theta \in \Theta_d^+(s,a)$ that*

$$\mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+ - \eta|\right] \leq \frac{d-|S|}{s}\exp\left(-\frac{n\tau^2}{2B^2}\right) + \frac{|S|}{s}\exp\left(-\frac{n(C_1-\tau)^2}{2B^2}\right),$$

where S denotes the support of θ. In particular, choosing $\tau = C_1/2$ yields

$$\sup_{\theta\in\Theta_d^+(s,a)} \mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+ - \eta|\right] \leq \frac{d}{s}\exp\left(-\frac{C_1^2 n(e^\alpha-1)^2}{8(e^\alpha+1)^2K_d^2}\right)$$

for all $a \geq 2\sigma$.

The proof of Proposition 3.3 can be found in Section B.4 of the appendix. Note that we can provide similar results on the class $\Theta_d(s,a)$ considering the selector $\widehat{\eta}$. The upper bounds are the same as for the class $\Theta_d^+(s,a)$ up to a multiplicative factor 2 that comes from the use in the proof of the two-sided Hoeffding's inequality instead of the one-sided inequality. Since $K_d \leq C\sqrt{d}$ for d large enough, we obtain that a sufficient condition for almost full recovery to be possible when $a \geq 2\sigma$ is that

$$\frac{n(e^\alpha-1)^2}{(e^\alpha+1)^2 d} \gtrsim \log(d/s).$$

Moreover, using that $(e^\alpha - 1)^2/(e^\alpha + 1)^2 \geq 0.2\alpha^2$ if $\alpha \leq 1$, we obtain that a sufficient condition for almost full recovery to be possible when $a \geq 2\sigma$ and $\alpha \leq 1$ is that $n\alpha^2/d \gtrsim \log(d/s)$. This improves the result we obtained when we considered only privacy mechanisms acting coordinates by coordinates for which we needed $n\alpha^2/d^2 \gtrsim \log(d/s)$. We now deal with the case $a \ll \sigma$.

Proposition 3.4. *Let $a > 0$ and $a \leq 2\sigma$. If τ is chosen such that $\tau < 2p(2)a/\sigma$, then it holds for all $\theta \in \Theta_d^+(s,a)$ that*

$$\mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+ - \eta|\right] \leq \frac{d-|S|}{s}\exp\Big(-\frac{n\tau^2}{2B^2}\Big) + \frac{|S|}{s}\exp\Big(-\frac{n(2p(2)a/\sigma - \tau)^2}{2B^2}\Big),$$

where S denotes the support of θ.

The proof of Proposition 3.4 can be found in Appendix B.5.

Corollary 3.5. *Assume that $\alpha/d \to 0$, $n\alpha^2/d \to +\infty$ and $\sigma\sqrt{d}/(\alpha\sqrt{n}) \ll a \leq 2\sigma$. The selector $\widehat{\eta}^+$ defined by (16) with $\tau = p(2)a/\sigma$ satisfies for n, d large enough*

$$\sup_{\theta \in \Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1, \dots, Z^d)) - \eta| \leq \frac{d}{s}\exp\Big(-\frac{n(e^\alpha-1)^2p^2(2)a^2}{2\sigma^2(e^\alpha+1)^2K_d^2}\Big).$$

In particular, if $\alpha \in (0,1]$, if $n\alpha^2/d \to +\infty$ with $n\alpha^2/d \gg \log(d)$ then it holds that

$$\sup_{\theta \in \Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1, \dots, Z^d)) - \eta| \to 0$$

for all a satisfying $\sigma\sqrt{\frac{d}{n\alpha^2}}\sqrt{\log(d/s)} \ll a \leq 2\sigma$; and also that

$$\sup_{\theta \in \Theta_d^+(s,a)} \mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}^+(Z^1, \dots, Z^d)) - \eta| \to 0$$

for all a satisfying $\sigma\sqrt{\frac{d}{n\alpha^2}}\sqrt{\log(d)} \ll a \leq 2\sigma$.

The first statement in Corollary 3.5 is a direct consequence of Proposition 3.4. The second statement is a direct consequence of the first one where we have used $(e^\alpha - 1)^2/(e^\alpha + 1)^2 \geq 0.2\alpha^2$ for $\alpha \in (0,1]$ and $K_d \leq C\sqrt{d}$ for d large enough. In the next subsection, we complement these results with a lower bound. This will enable us to exhibit a value a^* such that exact recovery is impossible for all $a \leq a^*$ and possible for $a \gg a^*$ under the assumptions $\alpha \in (0,1]$ and $n\alpha^2/d \to \infty$ with $n\alpha^2/d \gg \log(d)$.

3.3. Lower bound

Recall that P_0 denotes the distribution of $\sigma\xi_1$ and P_a that of $a + \sigma\xi_1$ and denote by $\chi^2(P_0, P_a)$ the chi-square discrepancy between the two distributions.

Proposition 3.6. *For any $a > 0$ such that $\chi^2(P_0, P_a) < \infty$, $\alpha > 0$, $d \geq 4$, $1 \leq s \leq d$, $n \geq 1$, we have*

$$\inf_{Q \in \mathcal{Q}_\alpha} \inf_{\hat{\eta} \in \mathcal{T}} \sup_{\theta \in \Theta_d^+(s,a)} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\hat{\eta} - \eta| \geq \frac{1}{4}\left(1 - \frac{2n(e^\alpha - 1)^2}{d\log(d)} \chi^2(P_0, P_a)\right).$$

Assume now that the measure P^{ξ_1} of the noise coordinates is strongly log-concave, with density $p = \exp(-\phi)$, with a potential ϕ that has a curvature bounded from above by a constant c_+ as in (3). Then

$$\chi^2\left(P_0, P_a\right) \leq \exp\left(c_+\left(\frac{a}{\sigma}\right)^2\right) - 1. \tag{17}$$

Note that inequality (17) is sharp in the sense that in the Gaussian case, $c_+ = 1$ holds and inequality (25) turns out to be an equality. Note also that log-concavity is actually not needed in Proposition 3.6, since we only require an upper bound on the curvature of the potential ϕ.

The proof of Proposition 3.6 is based on a private version of Fano's method, see [19, Proposition 2]. It can be found in Section B.6 of the appendix. Using that $(e^\alpha - 1)^2 \leq 4\alpha^2$ for $\alpha \in (0, 1)$ and $\exp(c_+x^2) - 1 \leq Lx^2$ for $0 \leq x \leq 2$ and some constant L only depending on c_+ (e.g., $L = (\exp(2c_+) - 1)/2$), Proposition 3.6 immediately shows the following.

Corollary 3.7. *Let $\alpha \in (0, 1)$. If $n\alpha^2/(d \log d) \leq C/(32L)$ for some constant $C \in (0, 1)$, then it holds that*

$$\inf_{Q \in \mathcal{Q}_\alpha} \inf_{\hat{\eta} \in \mathcal{T}} \sup_{\theta \in \Theta_d^+(s,a)} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\hat{\eta} - \eta| \geq \frac{1}{4}(1 - C) > 0,$$

for all $a \leq 2\sigma$.

This shows that exact recovery is impossible for all $a \leq 2\sigma$ if

$$n\alpha^2/(d \log d) \leq C/(32L)$$

for some constant $C \in (0, 1)$. Proposition 3.6 also implies that exact recovery is impossible if

$$a \leq \sigma\sqrt{\log(1 + Cd\log d/(8n\alpha^2))/c_+}$$

for some constant $C \in (0,1)$. However, unlike the coordinate local case, the lower bound provided by Proposition 3.6 does not allow us to say that exact recovery is also impossible for

$$a \geq \max\{2\sigma, \sigma\sqrt{\log(1 + Cd\log d/(8n\alpha^2))/c_+}\}$$

when $n\alpha^2/(d\log d)$ is bounded from above. The following corollary is also a direct consequence of Proposition 3.6. It shows that when $n\alpha^2/(d\log d) \to \infty$, exact recovery is still impossible if a is too small.

Corollary 3.8. *If $\alpha \in (0,1)$, $n\alpha^2/d \to +\infty$ with $n\alpha^2/d \gg \log d$ and*

$$a \leq (\sigma/(16L))\sqrt{d\log d/(n\alpha^2)}$$

it holds that

$$\liminf_{d\to+\infty} \inf_{Q\in\mathcal{Q}_\alpha} \inf_{\hat{\eta}\in\mathcal{T}} \sup_{\theta\in\Theta_d^+(s,a)} \mathbb{E}_{Q(P_\theta^{\otimes n})}|\hat{\eta}-\eta| \geq \frac{1}{8}.$$

The lower bound of Proposition 3.6 combined with the upper bounds of Subsection 3.2 exhibit a phase transition at the value a^* (up to a logarithmic factor) such that exact recovery is impossible for all $a \leq a^*$ and possible for $a \gg a^*$ under the assumptions $\alpha \in (0,1]$ and $n\alpha^2/d \to \infty$ with $n\alpha^2/d \gg \log(d)$. Precisely, set

$$a^* = a^*(n,\alpha,d,\sigma) = \frac{\sigma}{16L}\sqrt{\frac{d\log d}{n\alpha^2}},$$

where we recall that $L = (\exp(2c_+) - 1)/2$. Proposition 3.6 combined with Corollary 3.5 and Proposition 3.3 give the following result.

Corollary 3.9. *Assume that $\alpha \in (0,1]$ and $n\alpha^2/d \to +\infty$ with $n\alpha^2/d \gg \log(d)$. Then, exact recovery is impossible for all $a \leq a^*$ and is possible for all $a \gg a^*$.*

Note that Proposition 3.6 does not allow us to obtain impossibility results for almost full recovery in the regime $n\alpha^2/(d\log d) \gg 1$. Its proof relies on a private Fano's method ([19, Proposition 2]) applied with the family of distributions

$$\{\mathcal{N}(a\omega_i, \sigma^2 I_d), i = 1,\dots,d\},$$

where $\omega_i \in \{0,1\}^d$ is defined by $\omega_{ij} = \delta_{ij}$ and δ is the Kronecker delta. The same proof with ω_i defined by $\omega_{ij} = 1$ if $j \in [\![(i-1)s+1, is]\!]$ and $\omega_{ij} = 0$ otherwise for $i = 1,\dots,\lfloor d/s\rfloor$, provides the following lower bound.

Proposition 3.10. *Let $a > 0$ be such that $\chi^2(P_0, P_a) < \infty$, $\alpha > 0$, $n \geq 1$. If $d/s \leq 4$, then we have*

$$\inf_{Q\in\mathcal{Q}_\alpha} \inf_{\hat{\eta}\in\mathcal{T}} \sup_{\theta\in\Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\hat{\eta}-\eta| \geq \frac{1}{4}\left(1 - \frac{2n(e^\alpha-1)^2}{\lfloor d/s\rfloor\log(\lfloor d/s\rfloor)}\chi^2(P_0^{\otimes s}, P_a^{\otimes s})\right).$$

	$a \lesssim \frac{\sigma}{s}\sqrt{\frac{\log(d/s)}{Nd}}$	$\frac{\sigma}{s}\sqrt{\frac{\log(d/s)}{Nd}} \ll a \lesssim \frac{\sigma}{\sqrt{s}}$	$a \gg \frac{\sigma}{\sqrt{s}}$
$s\frac{Nd}{\log(d/s)} \lesssim 1$	impossible	impossible	?
$s\frac{Nd}{\log(d/s)} \gg 1$	impossible	possible, as soon as $a \gg \sigma\sqrt{\frac{\log(d/s)}{Nd}}$ if, moreover, $Nd \gg \log(d/s)$	possible

Table 3. Almost full recovery of θ in either $\Theta_d^+(s,a)$ or $\Theta_d(s,a)$ in the Coordinate Global case. We have set $N = n\alpha^2/d^2$.

If the noise has a potential ϕ that is two times continuously differentiable, with curvature bounded from above by a constant c_+ as in (3)*, then it holds that*

$$\chi^2(P_0^{\otimes s}, P_a^{\otimes s}) = \exp(s \cdot c_+ a^2/\sigma^2) - 1.$$

Note that $\chi^2(P_0^{\otimes s}, P_a^{\otimes s}) = (\chi^2(P_0, P_a) + 1)^s - 1$.

However, this bound turns out to be suboptimal in the sense that when $n\alpha^2/d \gg \log(d/s)$ holds, the combination of this bound with upper bounds in Proposition 3.3 and Corollary 3.5 allows us to exhibit the critical value a^* for almost full recovery only up to a logarithmic factor times the sparsity s. Indeed, on the one hand, Proposition 3.3 and Corollary 3.5 prove that almost full recovery is possible for all

$$a \gg \sigma\sqrt{d/(n\alpha^2)}\sqrt{\log(d/s)}$$

in the regime $n\alpha^2/d \gg \log(d/s)$. On the other hand Proposition 3.10 proves that, in the same regime, almost full recovery is impossible for

$$a \lesssim (\sigma/s)\sqrt{d/(n\alpha^2)}\sqrt{\log(d/s)},$$

but does not allow us to say what happens for

$$(\sigma/s)\sqrt{d/(n\alpha^2)}\sqrt{\log(d/s)} \ll a \lesssim \sigma\sqrt{d/(n\alpha^2)}\sqrt{\log(d/s)}.$$

A. Proofs of Section 2

A.1. Some auxiliary results for the proof of the lower bound

The proof of Theorem 2.1 strongly relies on the following result known as the Bayesian version of the Neyman–Pearson lemma.

Theorem A.1 ([29, Problem 3.10]). *Let P_0 and P_1 be probability distributions possessing densities p_0 and p_1 with respect to a measure μ. Consider the problem of testing H_0: $P = P_0$ against H_1: $P = P_1$, and suppose that known probabilities π and $1 - \pi$ can be assigned to H_0 and H_1 prior to the experiment. Then the test T^* given by*

$$T^*(X) = I\big((1-\pi)p_1(X) > \pi p_0(X)\big)$$

is a minimizer of the overall probability of error resulting from the use of a test T,

$$\pi \mathbb{E}_0[T(X)] + (1-\pi)\mathbb{E}_1[1 - T(X)].$$

The following lemmas are also useful to prove the lower bound.

Lemma A.2. *Let $b, c > 0$. Let P and Q be two probability measures having densities p and q with respect to some measure μ. It holds that*

$$\int \min\{bp(x), cq(x)\}\,\mathrm{d}\mu(x) \geq \frac{bc}{b+c}\bigg(\int \sqrt{p(x)q(x)}\,\mathrm{d}\mu(x)\bigg)^2.$$

The case $b = c = 1$ can be found in [36, Lemma 2.3]. We generalize the proof for any $b, c > 0$.

Proof. The Cauchy–Schwarz inequality yields

$$\begin{aligned} bc\bigg(\int \sqrt{p(x)q(x)}\,\mathrm{d}\mu(x)\bigg)^2 &= \bigg(\int \sqrt{bp(x)\cdot cq(x)}\,\mathrm{d}\mu(x)\bigg)^2 \\ &= \bigg(\int \sqrt{\min\{bp(x), cq(x)\}}\sqrt{\max\{bp(x), cq(x)\}}\,\mathrm{d}\mu(x)\bigg)^2 \\ &\leq \int \min\{bp(x), cq(x)\}\,\mathrm{d}\mu(x) \int \max\{bp(x), cq(x)\}\,\mathrm{d}\mu(x). \end{aligned}$$

In order to finish, we use that $\max\{u, v\} \leq u + v$ and get

$$\int \max\{bp(x), cq(x)\}\,\mathrm{d}\mu(x) \leq b\int p(x)\,\mathrm{d}\mu(x) + c\int q(x)\,\mathrm{d}\mu(x) = b + c. \qquad \blacksquare$$

In the proof of the lower bound, Lemma A.2 will be combined with the following result whose proof can be found in [36].

Lemma A.3. *Let P and Q be two probability measures having densities p and q with respect to some measure μ. It holds that*

$$\bigg(\int \sqrt{p(x)q(x)}\,\mathrm{d}\mu(x)\bigg)^2 \geq \exp(-\mathrm{KL}(P, Q)).$$

A.2. Proof of Theorem 2.1

Let $Q \in \mathcal{Q}_\alpha^{CL}$ and let $\widehat{\eta}$ be a separable selector. Since $\widehat{\eta}_j$ depends only on $(Z_j^i)_{i=1,\dots,n}$, it holds that

$$\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}(Z)-\eta| = \sum_{j=1}^{d} \mathbb{E}_{\otimes_{i=1}^n Q_j^i P_{\theta_j}} |\widehat{\eta}_j(Z_j^1,\dots,Z_j^n)-\eta_j|.$$

Following the proof of [11, Theorem 2.2], we denote by Θ' the set of all θ in $\Theta_d^+(s,a)$ such that exactly s components of θ are equal to a and the remaining $d-s$ components are equal to 0. Since Θ' is a subset of $\Theta_d^+(s,a)$, it holds that

$$\begin{aligned}
&\sup_{\theta\in\Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}(Z)-\eta| \\
&\quad\geq \frac{1}{s|\Theta'|}\sum_{\theta\in\Theta'}\sum_{j=1}^{d} \mathbb{E}_{\otimes_{i=1}^n Q_j^i P_{\theta_j}} |\widehat{\eta}_j(Z_j^1,\dots,Z_j^n)-\eta_j| \\
&\quad= \frac{1}{s|\Theta'|}\sum_{j=1}^{d}\Bigg(\sum_{\theta\in\Theta':\theta_j=0} \mathbb{E}_{\otimes_{i=1}^n Q_j^i P_0}(\widehat{\eta}_j) + \sum_{\theta\in\Theta':\theta_j=a} \mathbb{E}_{\otimes_{i=1}^n Q_j^i P_a}(1-\widehat{\eta}_j)\Bigg) \\
&\quad= \frac{1}{s}\sum_{j=1}^{d}\Big(\Big(1-\frac{s}{d}\Big)\mathbb{E}_{\otimes_{i=1}^n Q_j^i P_0}(\widehat{\eta}_j) + \frac{s}{d}\mathbb{E}_{\otimes_{i=1}^n Q_j^i P_a}(1-\widehat{\eta}_j)\Big) \\
&\quad\geq \frac{1}{s}\sum_{j=1}^{d}\inf_{T\in[0,1]}\Big(\Big(1-\frac{s}{d}\Big)\mathbb{E}_{\otimes_{i=1}^n Q_j^i P_0}(T) + \frac{s}{d}\mathbb{E}_{\otimes_{i=1}^n Q_j^i P_a}(1-T)\Big).
\end{aligned}$$

Set

$$L_j^* = \inf_{T\in[0,1]}\Big(\Big(1-\frac{s}{d}\Big)\mathbb{E}_{\otimes_{i=1}^n Q_j^i P_0}(T) + \frac{s}{d}\mathbb{E}_{\otimes_{i=1}^n Q_j^i P_a}(1-T)\Big).$$

Since Q_j^i provides α_j-differential privacy, the channel probabilities $Q_j^i(\cdot \mid x)$ have densities $z \mapsto q_j^i(z \mid x)$ with respect to some measure μ_j^i. Therefore,

$$dQ_j^i P_0(z) = m_{j,0}^i(z)\,\mathrm{d}\mu_j^i(z) \quad \text{and} \quad dQ_j^i P_a(z) = m_{j,a}^i(z)\,\mathrm{d}\mu_j^i(z),$$

where

$$m_{j,b}^i(z) = \int_{\mathbb{R}} q_j^i(z \mid x)dP_b(x), \quad b\in\{0,a\}.$$

Thus, for $b\in\{0,a\}$, it holds that

$$d(\otimes_{i=1}^n Q_j^i P_b)(y_1,\dots,y_n) = \left[\prod_{i=1}^n m_{j,b}^i(y_i)\right]\mathrm{d}\mu_j(y_1,\dots,y_n),$$

where $\mu_j = \mu_j^1 \otimes \cdots \otimes \mu_j^n$. According to Theorem A.1, the infimum L_j^* is thus attained for $T = T_j^*$ given by

$$T_j^*(Y_1, \dots, Y_n) = I\left(\frac{s}{d}\prod_{i=1}^n m_{j,a}^i(Y_i) > \left(1 - \frac{s}{d}\right)\prod_{i=1}^n m_{j,0}^i(Y_i)\right).$$

Set

$$A_j = \left\{(y_1, \dots, y_n) \in \mathbb{R}^n : \frac{s}{d}\prod_{i=1}^n m_{j,a}^i(y_i) > \left(1 - \frac{s}{d}\right)\prod_{i=1}^n m_{j,0}^i(y_i)\right\}.$$

Then

$$\begin{aligned}
&\sup_{\theta\in\Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}(Z) - \eta| \\
&\geq \frac{1}{s}\sum_{j=1}^d \Bigg[\left(1 - \frac{s}{d}\right)\int_{A_j}\left[\prod_{i=1}^n m_{j,0}^i(y_i)\right]\mathrm{d}\mu_j(y_1, \dots, y_n) \\
&\qquad + \frac{s}{d}\int_{A_j^C}\left[\prod_{i=1}^n m_{j,a}^i(y_i)\right]\mathrm{d}\mu_j(y_1, \dots, y_n)\Bigg] \\
&= \frac{1}{s}\sum_{j=1}^d \int_{\mathbb{R}^n} \min\left\{\left(1 - \frac{s}{d}\right)\prod_{i=1}^n m_{j,0}^i(y_i), \frac{s}{d}\prod_{i=1}^n m_{j,a}^i(y_i)\right\}\mathrm{d}\mu_j(y_1, \dots, y_n) \\
&\geq \left(1 - \frac{s}{d}\right)\cdot\frac{1}{d}\sum_{j=1}^d\left(\int_{\mathbb{R}^n}\sqrt{\left(\prod_{i=1}^n m_{j,0}^i(y_i)\right)\left(\prod_{i=1}^n m_{j,a}^i(y_i)\right)}\,\mathrm{d}\mu_j(y_1, \dots, y_n)\right)^2 \\
&\geq \left(1 - \frac{s}{d}\right)\cdot\frac{1}{d}\sum_{j=1}^d \exp\left(-\operatorname{KL}\left(\otimes_{i=1}^n Q_j^i P_0, \otimes_{i=1}^n Q_j^i P_a\right)\right),
\end{aligned}$$

where the two last inequalities follow from Lemmas A.2 and A.3. Using the tensorization property of the Kullback–Leibler divergence, we obtain

$$\begin{aligned}
&\sup_{\theta\in\Theta_d^+(s,a)} \frac{1}{s}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}(Z) - \eta| \\
&\qquad\geq \left(1 - \frac{s}{d}\right)\cdot\frac{1}{d}\sum_{j=1}^d \exp\left(-\sum_{i=1}^n \operatorname{KL}\left(Q_j^i P_0, Q_j^i P_a\right)\right) \\
&\qquad\geq \left(1 - \frac{s}{d}\right)\cdot\frac{1}{d}\sum_{j=1}^d \exp\left(-4n(e^{\alpha/d} - 1)^2\operatorname{TV}(P_0, P_a)^2\right) \\
&\qquad= \left(1 - \frac{s}{d}\right)\exp\left(-4n(e^{\alpha/d} - 1)^2\operatorname{TV}(P_0, P_a)^2\right),
\end{aligned}$$

where the second inequality is a direct consequence of the strong data processing inequality in [19, Theorem 1] showing the contraction property of privacy:

$$\mathrm{KL}(QP_0, QP_a) + \mathrm{KL}(QP_a, QP_0) \le (4 \wedge e^{2\alpha/d})(e^{\alpha/d} - 1)^2 TV(P_0, P_a)^2$$

if Q is α/d-DP. Since this result holds for all $Q \in \mathcal{Q}_\alpha^{CL}$ and all separable selector $\widehat{\eta}$, we obtain

$$\inf_{Q \in \mathcal{Q}_\alpha^{CL}} \inf_{\widehat{\eta} \in \mathcal{T}} \sup_{\theta \in \Theta_d^+(s,a)} \frac{1}{s} \mathbb{E}_{Q(P_\theta^{\otimes n})} |\widehat{\eta} - \eta| \\ \ge \left(1 - \frac{s}{d}\right) \exp\left(-4n(e^{\alpha/d} - 1)^2 \mathrm{TV}(P_0, P_a)^2\right).$$

Note that the TV distance is invariant to a scale parameter, thus $TV(P_0, P_a)$ can be calculated as the TV distance between the distribution of ξ_1 and the same one shifted by a/σ. The inequality $TV(P_0, P_a) \le 1$, Pinsker's inequality and inequality (19) of Lemma A.4 below, give

$$TV(P_0, P_a) \le \sqrt{\frac{\mathrm{KL}(P_0, P_a)}{2}} \le \frac{a\sqrt{c_+}}{2\sigma},$$

which implies the statement of Theorem 2.1.

Lemma A.4. *Consider that the measure P^{ξ_1} of the noise coordinates is c-strongly log-concave on $\mathbb{R}$, with density $p = \exp(-\phi)$, the convex function ϕ thus being c-strongly convex for some constant $c > 0$. Recall that the measure P_0 is the distribution of the scaled noise coordinate $\sigma\xi_1$ and that P_a is the image of P_0 by the translation of a. It holds that*

$$\mathrm{KL}(P_0, P_a) \ge \frac{ca^2}{2\sigma^2}. \tag{18}$$

If, on the other hand, we assume that the measure P^{ξ_1} has a density $p = \exp(-\phi)$ converging to zero at infinity, with ϕ being two times continuously differentiable and satisfying inequality (3) for a constant $c_+ > 0$, that gives a uniform upper bound of the curvature of ϕ by the constant c_+, then it holds

$$\mathrm{KL}(P_0, P_a) \le \frac{c_+ a^2}{2\sigma^2}. \tag{19}$$

If P^{ξ_1} is c-strongly log-concave and its potential achieves (3) for a positive constant c_+, then inequality (19) holds true.

Note that Lemma A.4 is tight in the sense that if P_0 is Gaussian with variance σ^2, then $\mathrm{KL}(P_0, P_a) = a^2/(2\sigma^2)$ and we have equality in both bounds (18) and (19), with $c = c_+ = 1$.

Proof. Consider first the case of a c-strongly log-concave density p.By standard approximation arguments (see, for instance, [34, Proposition 5.5]), we can assume without loss of generality that the potential ϕ is two times continuously differentiable. Consequently, c-strong convexity of ϕ is equivalent to $\phi''(x) \geq c$ for all $x \in \mathbb{R}$. Using Taylor expansion, this gives for any $x \in \mathbb{R}$ that

$$\phi(x-a) \geq \phi(x) - a\phi'(x) + c\frac{a^2}{2}.$$

This gives

$$\begin{aligned}\mathrm{KL}(P_0, P_a) &= \frac{1}{\sigma^2}\int_{\mathbb{R}} [\phi(x-a) - \phi(x)] \exp(-\phi(x))\,\mathrm{d}x \\ &\geq -\frac{a}{\sigma^2}\int_{\mathbb{R}} \phi'(x) \exp(-\phi(x))\,\mathrm{d}x + c_- \frac{a^2}{2\sigma^2}\int_{\mathbb{R}} \exp(-\phi(x))\,\mathrm{d}x \\ &= c_- \frac{a^2}{2\sigma^2}.\end{aligned}$$

Hence, (18) is proved. In order to prove (19), just remark that the regularity of ϕ and the upper bound on its curvature yield $\sup_{x\in\mathbb{R}} \phi''(x) \leq c_+$. Hence, by Taylor expansion again,

$$\phi(x-a) \leq \phi(x) - a\phi'(x) + c_+\frac{a^2}{2}.$$

Analogous computations now give $\mathrm{KL}(P_0, P_a) \leq c_+ a^2/(2\sigma^2)$, which is (19).

Finally, if p is strongly log-concave, then it tends to zero at infinity. By convolution by Gaussians, we can also assume without loss of generality that ϕ is two times continuously differentiable ([34, Proposition 5.5]). Hence, if in addition p satisfies (3), then it achieves the conditions that lead to (19). This concludes the proof of Lemma A.4. ■

A.3. Some auxiliary results for the upper bounds

First recall that since the vector $\xi = (\xi_1, \dots, \xi_d)$ is c-strongly log-concave – as it has independent c-strongly log-concave coordinates – then it achieves the following sub-Gaussian concentration inequality for Lipschitz functions (see [28, Proposition 2.18]): for any L-Lipschitz function $F\colon \mathbb{R}^d \to \mathbb{R}$, and any $r \geq 0$, we have

$$\mathbb{P}(F(\xi) - \mathbb{E}[F(\xi)] \geq r) \leq \exp(-cr^2/(2L^2)).$$

Furthermore, the celebrated Cafarelli's contraction theorem [14] state that the Brenier transport map pushing forward a Gaussian distribution to a strongly log-concave measure with the same Gaussian factor is a contraction. As a result, one can derive the following Mill's type bound for the deviations the coordinate ξ_1 (see [15, Proposition 2]): for any $r > 0$, we have

$$\mathbb{P}(|\xi_1| \geq r) \leq 2(1 - \Phi(\sqrt{c}r)) \leq \sqrt{\frac{2}{\pi}} \frac{\exp(-cr^2/2)}{\sqrt{c}r}, \tag{20}$$

where Φ is the standard Gaussian cumulative distribution function. Another useful fact is that, as ξ_1 is log-concave, it is unimodal (see [34]) and as ξ_1 is also symmetric, the maximum of its density $p(x) = \exp(-\phi_0(x)) \exp(-cx^2/2)$ is attained at its median 0. Note that as ϕ_0 is convex and symmetric, the maximum of the function $\exp(-\phi_0(x))$ is also attained at 0. In addition, by a result of Bobkov and Ledoux [8] (see also [34, Proposition 5.2]), as 0 is the median of ξ_1, it holds that

$$p(0) = \exp(-\phi_0(0)) \leq 1/(\sqrt{2}\sigma_{\xi_1}) \leq 1/\sqrt{2}.$$

Putting things together, for any $x \in \mathbb{R}$, we obtain that

$$p(x) \leq \frac{1}{\sqrt{2}} \exp(-cx^2/2).$$

Lemma A.5. *For all $a \geq 0$ and all $\tau > 0$, by Hoeffding's inequality, we have*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\operatorname{sgn}[a + \sigma\xi_j^i] - \mathbb{E}\left[\operatorname{sgn}[a + \sigma\xi_j^i]\right]\right) \geq \tau\right) \leq \exp\left(-n\frac{\tau^2}{2}\right).$$

We now recall Bernstein's inequality (cf. [9, Corollary 2.11]).

Theorem A.6. *Let $Y_1, \ldots, Y_n$ be independent real valued random variables. Assume that there exist some positive numbers v and u such that*

$$\sum_{i=1}^{n} \mathbb{E}[Y_i^2] \leqslant v,$$

and for all integers $m \geqslant 3$,

$$\sum_{i=1}^{n} \mathbb{E}[|Y_i|^m] \leqslant \frac{m!}{2} v u^{m-2}.$$

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

$$\mathbb{P}(S \geqslant t) \leqslant \exp\left(-\frac{t^2}{2(v + ut)}\right). \tag{21}$$

Note that if $v \leq ut$, then (21) yields

$$\mathbb{P}(S \geq t) \leq \exp(-t/(4u)).$$

If $ut \leq v$, then (21) yields

$$\mathbb{P}(S \geq t) \leq \exp(-t^2/4v).$$

We will apply this to get concentration bounds for the average of i.i.d. Laplace distributed random variables that check the assumptions of the theorem.

A.4. Proof of Proposition 2.4

It holds that

$$\begin{aligned} |\widehat{\eta}^+ - \eta| &= \sum_{j:\eta_j=0} \widehat{\eta}_j^+ + \sum_{j:\eta_j=1} (1 - \widehat{\eta}_j^+) \\ &= \sum_{j:\eta_j=0} I\left(\frac{1}{n}\sum_{i=1}^n \mathrm{sgn}[\sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i \geq \tau\right) \\ &\quad + \sum_{j:\eta_j=1} I\left(\frac{1}{n}\sum_{i=1}^n \mathrm{sgn}[\theta_j + \sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i < \tau\right). \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+ - \eta|\right] &= \frac{1}{s}\sum_{j:\eta_j=0} \underbrace{\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathrm{sgn}[\sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i \geq \tau\right)}_{=T_{1,j}} \\ &\quad + \frac{1}{s}\sum_{j:\eta_j=1} \underbrace{\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathrm{sgn}[\theta_j + \sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i < \tau\right)}_{=T_{2,j}}. \end{aligned}$$

We first study $T_{1,j}$. It holds that

$$T_{1,j} \leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathrm{sgn}[\sigma\xi_j^i] \geq \frac{\tau}{2}\right) + \mathbb{P}\left(\sum_{i=1}^n W_j^i \geq \frac{\tau n\alpha}{4d}\right).$$

Note that $\mathbb{E}[\mathrm{sgn}[\sigma\xi_j^i]] = 0$ by symmetry of ξ_j^i. Using Lemma A.5 to bound from above the first term and Bernstein's inequality (21) with $v = 2n$ and $u = 1$ to bound from above the second term, we obtain if $\tau\alpha/(8d) < 1$

$$T_{1,j} \leq \exp\Big(-\frac{n\tau^2}{2^3}\Big) + \exp\Big(-\frac{\tau^2 n\alpha^2}{2^7 d^2}\Big).$$

Since $x \mapsto \operatorname{sgn}[x]$ is a non-decreasing function and since $\theta_j \geq a$ for all j such that $\eta_j = 1$, it holds that

$$\begin{aligned}
T_{2,j} &\leq \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} \operatorname{sgn}[a+\sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^{n} W_j^i < \tau\Big) \\
&= \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n}\big(\operatorname{sgn}[a+\sigma\xi_j^i] - \mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_j^i]\big]\big) \\
&\qquad + \mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_j^1]\big] + \frac{2d}{n\alpha}\sum_{i=1}^{n} W_j^i < \tau\Big) \\
&= \mathbb{P}\Big(-\frac{1}{n}\sum_{i=1}^{n}\big(\operatorname{sgn}[a+\sigma\xi_j^i] - \mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_j^i]\big]\big) \\
&\qquad - \frac{2d}{n\alpha}\sum_{i=1}^{n} W_j^i > \mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_j^1]\big] - \tau\Big).
\end{aligned}$$

As ξ_1 is symmetric and absolutely continuous, we have

$$\mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_1]\big] = \mathbb{P}\Big(\xi_1 \geq -\frac{a}{\sigma}\Big) - \mathbb{P}\Big(\xi_1 < -\frac{a}{\sigma}\Big) = 1 - 2\mathbb{P}\Big(\xi_1 > \frac{a}{\sigma}\Big).$$

Using (20), we further get

$$\mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_1]\big] \geq 2\Phi\Big(\sqrt{c}\,\frac{a}{\sigma}\Big) - 1 \geq 2\Phi(2\sqrt{c}) - 1 =: C_1,$$

for $a/\sigma \geq 2$, with Φ the cumulative distribution function of the standard Gaussian distribution.

Thus, if $a \geq 2\sigma$, it holds

$$\begin{aligned}
T_{2,j} &\leq \mathbb{P}\Big(-\frac{1}{n}\sum_{i=1}^{n}\big(\operatorname{sgn}[a+\sigma\xi_j^i] - \mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_j^i]\big]\big) - \frac{2d}{n\alpha}\sum_{i=1}^{n} W_j^i > C_1 - \tau\Big) \\
&\leq \mathbb{P}\Big(-\frac{1}{n}\sum_{i=1}^{n}\big(\operatorname{sgn}[a+\sigma\xi_j^i] - \mathbb{E}\big[\operatorname{sgn}[a+\sigma\xi_j^i]\big]\big) > \frac{C_1-\tau}{2}\Big) \\
&\quad + \mathbb{P}\Big(\sum_{i=1}^{n}(-W_j^i) > \frac{n\alpha(C_1-\tau)}{4d}\Big).
\end{aligned}$$

We can now bound from above the first term using Lemma A.5 and the second term using Bernstein's inequality. If $C_1 \geq \tau$ and $\alpha(C_1-\tau)/(8d) \leq 1$, this gives

$$T_{2,j} \leq \exp\Big(-\frac{n(C_1-\tau)^2}{2^3}\Big) + \exp\Big(-\frac{(C_1-\tau)^2 n\alpha^2}{2^7 d^2}\Big).$$

This ends the proof of (9). We now prove (10). If $\theta \in \Theta_d(s,a)$, we use the estimator $\widehat{\eta}$ instead of $\widehat{\eta}^+$ and it holds that

$$\mathbb{E}\Big[\frac{1}{s}|\widehat{\eta}-\eta|\Big] = \frac{1}{s}\sum_{j:\eta_j=0} \underbrace{\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n \operatorname{sgn}[\sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i\Big| \geq \tau\Big)}_{=\widetilde{T}_{1,j}}$$

$$+ \frac{1}{s}\sum_{j:\eta_j=1} \underbrace{\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n \operatorname{sgn}[\theta_j+\sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i\Big| < \tau\Big)}_{=\widetilde{T}_{2,j}}.$$

We first study $\widetilde{T}_{1,j}$. It holds that

$$\begin{aligned}\widetilde{T}_{1,j} &\leq \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n \operatorname{sgn}[\sigma\xi_j^i]\Big| + \frac{2d}{n\alpha}\Big|\sum_{i=1}^n W_j^i\Big| \geq \tau\Big)\\ &\leq \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n \operatorname{sgn}[\sigma\xi_j^i]\Big| \geq \frac{\tau}{2}\Big) + \mathbb{P}\Big(\Big|\sum_{i=1}^n W_j^i\Big| \geq \frac{\tau n\alpha}{4d}\Big)\\ &\leq 2\Big[\exp\Big(-\frac{n\tau^2}{2^3}\Big) + \exp\Big(-\frac{\tau^2 n\alpha^2}{2^7 d^2}\Big)\Big],\end{aligned}$$

if $\tau\alpha/(8d) < 1$, where we have used the two-sided versions of the concentration inequalities we used to prove (9). We now study $\widetilde{T}_{2,j}$. For all j such that $\eta_j = 1$, it holds that

$$\begin{aligned}\widetilde{T}_{2,j} &= \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n (\operatorname{sgn}[\theta_j+\sigma\xi_j^i] - \mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^i]])\\ &\qquad + \mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^1]] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i\Big| < \tau\Big)\\ &\leq \mathbb{P}\Big(\big|\mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^1]]\big|\\ &\qquad - \Big|\frac{1}{n}\sum_{i=1}^n (\operatorname{sgn}[\theta_j+\sigma\xi_j^i] - \mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^i]]) + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i\Big| < \tau\Big)\\ &= \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^n (\operatorname{sgn}[\theta_j+\sigma\xi_j^i] - \mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^i]]) + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i\Big|\\ &\qquad > \big|\mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^1]]\big| - \tau\Big).\end{aligned}$$

Now, observe that

$$\big|\mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^1]]\big| \geq \mathbb{E}[\operatorname{sgn}[\theta_j+\sigma\xi_j^1]] \geq \mathbb{E}[\operatorname{sgn}[a+\sigma\xi_j^1]],$$

if $\theta_j \geq a$ since $x \mapsto \mathrm{sgn}[x]$ is non-decreasing, and if $\theta_j \leq -a$, we have

$$\begin{aligned}\left|\mathbb{E}\big[\mathrm{sgn}[\theta_j+\sigma\xi_j^1]\big]\right| &\geq -\mathbb{E}\big[\mathrm{sgn}[\theta_j+\sigma\xi_j^1]\big] \geq -\mathbb{E}\big[\mathrm{sgn}[-a+\sigma\xi_j^1]\big] \\ &= -\mathbb{E}\big[\mathrm{sgn}[-a-\sigma\xi_j^1]\big] = \mathbb{E}\big[\mathrm{sgn}[a+\sigma\xi_j^1]\big],\end{aligned}$$

where we have used that $x \mapsto \mathrm{sgn}[x]$ is a non-decreasing and odd function and that $-\xi_j^1$ and ξ_j^1 have the same distribution. Moreover, we have seen in the proof of (9) that the following holds:

$$\mathbb{E}\big[\mathrm{sgn}[a+\sigma\xi]\big] \geq 2\Phi(2\sqrt{c})-1 =: C_1,$$

where Φ denotes the cumulative distribution function of the Gaussian distribution. Thus, if $a \geq 2\sigma$, it holds that $\mathbb{E}[\mathrm{sgn}[a+\sigma\xi_j^1]] \geq C_1$ for all j such that $\eta_j = 1$, and

$$\begin{aligned}\widetilde{T}_{2,j} &\leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\Big(\mathrm{sgn}[\theta_j+\sigma\xi_j^i]-\mathbb{E}\big[\mathrm{sgn}[\theta_j+\sigma\xi_j^i]\big]\Big)+\frac{2d}{n\alpha}\sum_{i=1}^n W_j^i\right| > C_1-\tau\right) \\ &\leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\Big(\mathrm{sgn}[\theta_j+\sigma\xi_j^i]-\mathbb{E}\big[\mathrm{sgn}[\theta_j+\sigma\xi_j^i]\big]\Big)\right|+\left|\frac{2d}{n\alpha}\sum_{i=1}^n W_j^i\right| > C_1-\tau\right) \\ &\leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\Big(\mathrm{sgn}[\theta_j+\sigma\xi_j^i]-\mathbb{E}\big[\mathrm{sgn}[\theta_j+\sigma\xi_j^i]\big]\Big)\right| > \frac{C_1-\tau}{2}\right) \\ &\quad+\mathbb{P}\left(\left|\sum_{i=1}^n W_j^i\right| > \frac{n\alpha(C_1-\tau)}{4d}\right).\end{aligned}$$

Using the two-sided version of the concentration inequalities that we used to bound $T_{2,j}$ in the proof of (9), if $C_1 > \tau$ and $\alpha(C_1-\tau)/(8d) \leq 1$, we obtain

$$\widetilde{T}_{2,j} \leq 2\left[\exp\left(-\frac{n(C_1-\tau)^2}{2^3}\right)+\exp\left(-\frac{(C_1-\tau)^2 n\alpha^2}{2^7 d^2}\right)\right].$$

This ends the proof of (10).

A.5. Proof of Corollary 2.5

Let us prove (11). Note that if the assumptions of Corollary 2.5 are satisfied, and if $\tau = C_1/2$ then the assumptions of Proposition 2.4 are also satisfied and for all $a \geq 2\sigma$, we have

$$\begin{aligned}\sup_{\theta\in\Theta}\mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+-\eta|\right] &\leq 2\cdot\frac{d}{s}\left[\exp\left(-\frac{C_1^2 n}{2^5}\right)+\exp\left(-\frac{C_1^2 n\alpha^2}{2^9 d^2}\right)\right] \\ &= 2\left\{\exp\left(\log\left(\frac{d}{s}\right)-\frac{C_1^2 n}{2^5}\right)+\exp\left(\log\left(\frac{d}{s}\right)-\frac{C_1^2 n\alpha^2}{2^9 d^2}\right)\right\}\end{aligned}$$

$$= 2\exp\left(-\frac{n\alpha^2}{d^2}\left[\frac{C_1^2 d^2}{2^5\alpha^2} - \frac{\log(d/s)}{n\alpha^2/d^2}\right]\right)$$
$$+ 2\exp\left(-\frac{n\alpha^2}{d^2}\left[\frac{C_1^2}{2^9} - \frac{\log(d/s)}{n\alpha^2/d^2}\right]\right).$$

The two terms appearing in the last inequality both tend to 0 as $d \to +\infty$ under the assumptions of Corollary 2.5, which gives (11). The proof of (12) is similar.

A.6. Proof of Proposition 2.6

The beginning of the proof is similar to the proof of Proposition 2.4. It holds that

$$\mathbb{E}\left[\frac{1}{s}|\widehat{\eta}^+ - \eta|\right] = \frac{1}{s}\sum_{j:\eta_j=0} \underbrace{\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \operatorname{sgn}[\sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i \geq \tau\right)}_{=T_{1,j}}$$
$$+ \frac{1}{s}\sum_{j:\eta_j=1} \underbrace{\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \operatorname{sgn}[\theta_j + \sigma\xi_j^i] + \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i < \tau\right)}_{=T_{2,j}},$$

and we have

$$T_{1,j} \leq \exp\left(-\frac{n\tau^2}{2^3}\right) + \exp\left(-\frac{\tau^2 n\alpha^2}{2^7 d^2}\right)$$

if $\tau\alpha/(8d) < 1$, and

$$T_{2,j} \leq \mathbb{P}\Bigg(-\frac{1}{n}\sum_{i=1}^n (\operatorname{sgn}[a + \sigma\xi_j^i] - \mathbb{E}\big[\operatorname{sgn}[a + \sigma\xi_j^i]\big])$$
$$- \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i > \mathbb{E}\big[\operatorname{sgn}[a + \sigma\xi_j^1]\big] - \tau\Bigg).$$

Now, we bound from below $\mathbb{E}\big[\operatorname{sgn}[a + \sigma\xi]\big]$ in a different way than in the proof of Proposition 2.4 by the tighter bound

$$\mathbb{E}\big[\operatorname{sgn}[a + \sigma\xi_1]\big] = 2\mathbb{P}\left(0 < \xi_1 \leq \frac{a}{\sigma}\right) \geq 2\frac{a}{\sigma}p\left(\frac{a}{\sigma}\right) \geq 2\frac{a}{\sigma}p(2),$$

for $a/\sigma < 2$, as the probability density function p of ξ_1 is c-strongly log-concave and symmetric and thus uni-modal at 0 and decreasing on $(0, \infty)$. Thus,

$$T_{2,j} \leq \mathbb{P}\Bigg(-\frac{1}{n}\sum_{i=1}^n \Big(\operatorname{sgn}[a + \sigma\xi_j^i] - \mathbb{E}\big[\operatorname{sgn}[a + \sigma\xi_j^i]\big]\Big)$$
$$- \frac{2d}{n\alpha}\sum_{i=1}^n W_j^i > 2\frac{a}{\sigma}p(2) - \tau\Bigg)$$

$$\leq \mathbb{P}\Big(-\frac{1}{n}\sum_{i=1}^{n}(\mathrm{sgn}[a+\sigma\xi_j^i]-\mathbb{E}\big[\mathrm{sgn}[a+\sigma\xi_j^i]\big]) > \frac{a}{\sigma}p(2)-\frac{\tau}{2}\Big) \\ + \mathbb{P}\Big(\sum_{i=1}^{n}(-W_j^i) > \frac{n\alpha(a/\sigma p(2)-\tau/2)}{2d}\Big).$$

We can now bound from above the first term using lemma A.5 and the second term using Bernstein's inequality. If $\tau < 2a/\sigma p(2)$ and $\alpha(a/\sigma p(2)-\tau/2)/(4d) \leq 1$, this gives

$$T_{2,j} \leq \exp\Big(-\frac{n(a/\sigma p(2)-\tau/2)^2}{2^3}\Big) + \exp\Big(-\frac{(a/\sigma p(2)-\tau/2)^2 n\alpha^2}{2^5 d^2}\Big).$$

A.7. Proof of Corollary 2.7

Let prove (13). The chosen value of $\tau = a/\sigma \cdot p(2)$ satisfies the assumptions of Proposition 2.6 for d large enough and yield

$$\sup_{\theta\in\Theta_d^+(s,a)} \mathbb{E}\Big[\frac{1}{s}|\hat{\eta}^+-\eta|\Big] \leq \frac{d}{s}\Big[\exp\Big(-\frac{na^2}{2^3\sigma^2}p^2(2)\Big) + \exp\Big(-\frac{n\alpha^2a^2}{2^7\sigma^2d^2}p^2(2)\Big)\Big] \\ = \exp\Big(\log\Big(\frac{d}{s}\Big) - \frac{na^2}{2^3\sigma^2}p^2(2)\Big) + \exp\Big(\log\Big(\frac{d}{s}\Big) - \frac{n\alpha^2a^2}{2^7\sigma^2d^2}p^2(2)\Big) \\ \leq 2\exp\Big(\log\Big(\frac{d}{s}\Big) - \frac{n\alpha^2a^2}{2^7\sigma^2d^2}\Big).$$

Conclude using that $a \gg \sigma d/\sqrt{n\alpha^2}\sqrt{\log(d/s)}$. The proof of (14) is similar.

B. Proofs of Section 3

B.1. Proof of Proposition 3.1

Note that it is sufficient to prove that $\widetilde{Z}^i$ is an α-LDP view of X^i. Indeed, if $\widetilde{Z}^i$ is an α-LDP view of X^i then it holds for all $z \in \mathcal{Z}$ and $x, x' \in \mathbb{R}^d$ (we omit the superscript i) that

$$\frac{\mathbb{P}(Z=z \mid X=x)}{\mathbb{P}(Z=z \mid X=x')} = \frac{\sum_{\tilde{z}\in\{-B,B\}^d}\mathbb{P}(Z=z \mid \widetilde{Z}=\tilde{z}, X=x)\mathbb{P}(\widetilde{Z}=\tilde{z} \mid X=x)}{\sum_{\tilde{z}\in\{-B,B\}^d}\mathbb{P}(Z=z \mid \widetilde{Z}=\tilde{z}, X=x')\mathbb{P}(\widetilde{Z}=\tilde{z} \mid X=x')} \\ = \frac{\sum_{\tilde{z}\in\{-B,B\}^d}\mathbb{P}(Z=z \mid \widetilde{Z}=\tilde{z})\mathbb{P}(\widetilde{Z}=\tilde{z} \mid X=x)}{\sum_{\tilde{z}\in\{-B,B\}^d}\mathbb{P}(Z=z \mid \widetilde{Z}=\tilde{z})\mathbb{P}(\widetilde{Z}=\tilde{z} \mid X=x')} \leq e^{\alpha},$$

where we have used that Z is independent from X conditionally to $\widetilde{Z}$ and the fact that $\mathbb{P}(\widetilde{Z}=\tilde{z} \mid X=x) \leq e^{\alpha}\mathbb{P}(\widetilde{Z}=\tilde{z} \mid X=x')$ for all $\tilde{z} \in \{-B,B\}^d$ if $\widetilde{Z}$ is an

α-LDP view of X. So, let's prove that $\widetilde{Z}^i$ is an α-LDP view of X^i. In what follows, we omit once again the superscript i. We have to prove that for all $\widetilde{z} \in \{-B, B\}^d$ and all $x, x' \in \mathbb{R}^d$ it holds that

$$\frac{\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid X = x)}{\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid X = x')} \leq e^{\alpha}.$$

Let $\widetilde{z} \in \{-B, B\}^d$ and $x \in \mathbb{R}^d$. It holds that

$$\begin{aligned}\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid X = x) &= \sum_{\widetilde{x} \in \{-1,1\}^d} \mathbb{P}(\widetilde{Z} = \widetilde{z} \mid X = x, \widetilde{X} = \widetilde{x}) \cdot \mathbb{P}(\widetilde{X} = \widetilde{x} \mid X = x) \\ &= \sum_{\widetilde{x} \in \{-1,1\}^d} \mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}) \cdot \mathbb{P}(\widetilde{X} = \widetilde{x} \mid X = x),\end{aligned}$$

and since Y and $\widetilde{X}$ are independent, we have

$$\begin{aligned}\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}) &= \mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}, Y = 1) \cdot \mathbb{P}(Y = 1) \\ &\quad + \mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}, Y = 0) \cdot \mathbb{P}(Y = 0) \\ &= \pi_\alpha \mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}, Y = 1) \\ &\quad + (1 - \pi_\alpha)\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}, Y = 0).\end{aligned}$$

Moreover, since for $\widetilde{x} \in \{-1, 1\}^d$, we have

$$\begin{aligned}&\mathrm{Card}\big(\big\{\widetilde{z} \in \{-B, B\}^d \mid \langle \widetilde{z}, \widetilde{x}\rangle > 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = B\widetilde{x}_1)\big\}\big) \\ &\quad = \mathrm{Card}\big(\big\{\widetilde{z} \in \{-B, B\}^d \mid \langle \widetilde{z}, \widetilde{x}\rangle < 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = -B\widetilde{x}_1)\big\}\big) = 2^{d-1},\end{aligned}$$

it holds that

$$\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}, Y = 1) = \begin{cases} 0 & \text{if } \langle \widetilde{z}, \widetilde{x}\rangle < 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = -B\widetilde{x}_1), \\ \frac{1}{2^{d-1}} & \text{if } \langle \widetilde{z}, \widetilde{x}\rangle > 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = B\widetilde{x}_1), \end{cases}$$

and

$$\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}, Y = 0) = \begin{cases} \frac{1}{2^{d-1}} & \text{if } \langle \widetilde{z}, \widetilde{x}\rangle < 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = -B\widetilde{x}_1), \\ 0 & \text{if } \langle \widetilde{z}, \widetilde{x}\rangle > 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = B\widetilde{x}_1). \end{cases}$$

We thus have

$$\mathbb{P}(\widetilde{Z} = \widetilde{z} \mid \widetilde{X} = \widetilde{x}) = \begin{cases} \frac{1-\pi_\alpha}{2^{d-1}} & \text{if } \langle \widetilde{z}, \widetilde{x}\rangle < 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = -B\widetilde{x}_1), \\ \frac{\pi_\alpha}{2^{d-1}} & \text{if } \langle \widetilde{z}, \widetilde{x}\rangle > 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = B\widetilde{x}_1), \end{cases}$$

and, if we set

$$A_{\widetilde{z}} = \big\{\widetilde{x} \in \{-1, 1\}^d : \langle \widetilde{z}, \widetilde{x}\rangle > 0 \text{ or } (\langle \widetilde{z}, \widetilde{x}\rangle = 0 \text{ and } \widetilde{z}_1 = B\widetilde{x}_1)\big\}$$

and

$$C_{\tilde{z}} = \left\{\tilde{x} \in \{-1,1\}^d : \langle \tilde{z}, \tilde{x}\rangle < 0 \text{ or } (\langle \tilde{z}, \tilde{x}\rangle = 0 \text{ and } \tilde{z}_1 = -B\tilde{x}_1)\right\},$$

we obtain

$$\mathbb{P}(\widetilde{Z} = \tilde{z} \mid X = x) = \frac{\pi_\alpha}{2^{d-1}} \sum_{\tilde{x}\in A_{\tilde{z}}} \mathbb{P}(\widetilde{X} = \tilde{x} \mid X = x) + \frac{1-\pi_\alpha}{2^{d-1}} \sum_{\tilde{x}\in C_{\tilde{z}}} \mathbb{P}(\widetilde{X} = \tilde{x} \mid X = x).$$

Consequently, for all $\tilde{z} \in \{-B, B\}^d$ and all $x \in \mathbb{R}^d$, it holds that

$$\frac{\min\{\pi_\alpha, 1-\pi_\alpha\}}{2^{d-1}} \leq \mathbb{P}(\widetilde{Z} = \tilde{z} \mid X = x) \leq \frac{\max\{\pi_\alpha, 1-\pi_\alpha\}}{2^{d-1}},$$

where we have used that $A_{\tilde{z}} \sqcup C_{\tilde{z}} = \{-1,1\}^d$ and $\sum_{\tilde{x}\in\{-1,1\}^d} \mathbb{P}(\widetilde{X} = \tilde{x} \mid X = x) = 1$. Finally, for all $\tilde{z} \in \{-B, B\}^d$ and all $x, x' \in \mathbb{R}^d$, we obtain

$$\frac{\mathbb{P}(\widetilde{Z} = \tilde{z} \mid X = x)}{\mathbb{P}(\widetilde{Z} = \tilde{z} \mid X = x')} \leq \frac{\max\{\pi_\alpha, 1-\pi_\alpha\}}{\min\{\pi_\alpha, 1-\pi_\alpha\}} = \frac{\pi_\alpha}{1-\pi_\alpha} = e^\alpha.$$

B.2. Proof of Proposition 3.2

Let $x \in \mathbb{R}^d$. We first compute $\mathbb{E}[\widetilde{Z} \mid X = x]$. It holds that

$$\begin{aligned}\mathbb{E}[\widetilde{Z} \mid X = x] &= \sum_{\tilde{x}\in\{-1,1\}^d} \mathbb{P}(\widetilde{X} = \tilde{x} \mid X = x) \cdot \mathbb{E}[\widetilde{Z} \mid X = x, \widetilde{X} = \tilde{x}] \\ &= \sum_{\tilde{x}\in\{-1,1\}^d} \mathbb{P}(\widetilde{X} = \tilde{x} \mid X = x) \cdot \mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \tilde{x}],\end{aligned}$$

and since Y and $\widetilde{X}$ are independent, we have

$$\begin{aligned}\mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \tilde{x}] = \mathbb{P}(Y = 1) \cdot \mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \tilde{x}, Y = 1] \\ + \mathbb{P}(Y = 0) \cdot \mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \tilde{x}, Y = 0] \\ = \pi_\alpha \mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \tilde{x}, Y = 1] + (1-\pi_\alpha)\mathbb{E}[\widetilde{Z} = z \mid \widetilde{X} = \tilde{x}, Y = 0].\end{aligned}$$

Define

$$\begin{aligned}A_{\tilde{x}} &:= \left\{\tilde{z} \in \{-B, B\}^d \mid \langle \tilde{z}, \tilde{x}\rangle > 0 \text{ or } (\langle \tilde{z}, \tilde{x}\rangle = 0 \text{ and } \tilde{z}_1 = B\tilde{x}_1)\right\}, \\ C_{\tilde{x}} &:= \left\{\tilde{z} \in \{-B, B\}^d \mid \langle \tilde{z}, \tilde{x}\rangle < 0 \text{ or } (\langle \tilde{z}, \tilde{x}\rangle = 0 \text{ and } \tilde{z}_1 = -B\tilde{x}_1)\right\}.\end{aligned}$$

Conditionally on $\{\widetilde{X} = \tilde{x}, Y = 1\}$, it holds that $Z \sim \mathcal{U}(A_{\tilde{x}})$. Thus,

$$\mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \tilde{x}, Y = 1] = \sum_{\tilde{z}\in A_{\tilde{x}}} \mathbb{P}(\widetilde{Z} = \tilde{z} \mid \widetilde{X} = \tilde{x}, Y = 1)\tilde{z} = \frac{1}{\mathrm{Card}(A_{\tilde{x}})} \sum_{\tilde{z}\in A_{\tilde{x}}} \tilde{z}.$$

Similarly,

$$\mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \widetilde{x}, Y = 0] = \frac{1}{\mathrm{Card}(C_{\widetilde{x}})} \sum_{\widetilde{z} \in C_{\widetilde{x}}} \widetilde{z} = \frac{1}{\mathrm{Card}(C_{\widetilde{x}})} \sum_{\widetilde{z} \in A_{\widetilde{x}}} (-\widetilde{z})$$
$$= -\mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \widetilde{x}, Y = 1],$$

where we have used that $\mathrm{Card}(C_{\widetilde{x}}) = \mathrm{Card}(A_{\widetilde{x}})$. We thus obtain

$$\mathbb{E}[\widetilde{Z} \mid \widetilde{X} = \widetilde{x}] = \frac{2\pi_\alpha - 1}{\mathrm{Card}(A_{\widetilde{x}})} \sum_{\widetilde{z} \in A_{\widetilde{x}}} \widetilde{z},$$

and, using that $\mathrm{Card}(A_{\widetilde{x}}) = 2^{d-1}$ for all $\widetilde{x} \in \{-1, 1\}^d$, we obtain

$$\mathbb{E}[\widetilde{Z} \mid X = x] = \frac{2\pi_\alpha - 1}{2^{d-1}} \sum_{\widetilde{x} \in \{-1,1\}^d} \left[\mathbb{P}(\widetilde{X} = \widetilde{x} \mid X = x) \cdot \sum_{\widetilde{z} \in A_{\widetilde{x}}} \widetilde{z} \right].$$

We now compute $\sum_{\widetilde{z} \in A_{\widetilde{x}}} \widetilde{z}$ for all $\widetilde{x} \in \{-1, 1\}^d$. Note that for $\widetilde{z} \in \{-B, B\}^d$ and $\widetilde{x} \in \{-1, 1\}^d$, $\langle \widetilde{z}, \widetilde{x} \rangle$ is a sum of d terms, each equal to $-B$ or B. If a denotes the number of elements of this sum equal to B and b denotes the number of elements of this sum equal to $-B$, then it holds $a + b = d$ and $\langle \widetilde{z}, \widetilde{x} \rangle = aB - bB = B(d - 2b)$. Thus we can only have $\langle \widetilde{z}, \widetilde{x} \rangle = kB$, where $k \in [\![-d, d]\!]$ and $|k|$ has the same parity as d. We thus have

$$\sum_{\widetilde{z} \in A_{\widetilde{x}}} \widetilde{z} = \sum_{p=0}^{(d-1)/2} \sum_{\{\widetilde{z} \in \{-B,B\}^d : \langle \widetilde{z}, \widetilde{x} \rangle = (2p+1)B\}} z, \tag{22}$$

if d is odd, and

$$\sum_{\widetilde{z} \in A_{\widetilde{x}}} \widetilde{z} = \sum_{p=1}^{d/2} \sum_{\substack{\widetilde{z} \in \{-B,B\}^d : \\ \langle \widetilde{z}, \widetilde{x} \rangle = 2p \cdot B}} \widetilde{z} + \sum_{\substack{\widetilde{z} \in \{-B,B\}^d : \\ \langle \widetilde{z}, \widetilde{x} \rangle = 0, \\ \widetilde{z}_1 = B\widetilde{x}_1}} \widetilde{z}, \tag{23}$$

if d is even. Now, observe that for all $\widetilde{x} \in \{-1, 1\}^d$, for all $j \in [\![1, d]\!]$ and for all $k \in \{0, \ldots, d\}$ with the same parity as d, it holds that

$$\sum_{\widetilde{z} \in \{-B,B\}^d : \langle \widetilde{z}, \widetilde{x} \rangle = kB} \widetilde{z}_j = B\left[\binom{d-1}{\frac{d+k}{2} - 1} - \binom{d-1}{\frac{d+k}{2}} \right] \widetilde{x}_j. \tag{24}$$

Indeed, for all $\widetilde{z} \in \{-B, B\}^d$, for all $\widetilde{x} \in \{-1, 1\}^d$, and for all $k \in \{0, \ldots, d\}$ with the same parity as d, it holds that

$$\langle \widetilde{z}, \widetilde{x} \rangle = k \cdot B \iff \begin{cases} \widetilde{z}_j = B\widetilde{x}_j & \text{for } \frac{d+k}{2} \text{ elements } j \in [\![1, d]\!], \\ \widetilde{z}_j = -B\widetilde{x}_j & \text{for } \frac{d-k}{2} \text{ elements } j \in [\![1, d]\!]. \end{cases}$$

Setting $D_{k,\tilde{x}} = \{\tilde{z} \in \{-B, B\}^d : \langle \tilde{z}, \tilde{x} \rangle = k \cdot B\}$, it thus holds that

$$\begin{aligned}\sum_{\tilde{z} \in D_{k,\tilde{x}}} \tilde{z}_j &= \sum_{\tilde{z} \in D_{k,\tilde{x}}} B\tilde{x}_j \mathbb{1}(\tilde{z}_j = B\tilde{x}_j) - \sum_{z \in D_{k,\tilde{x}}} B\tilde{x}_j \mathbb{1}(\tilde{z}_j = -B\tilde{x}_j) \\ &= B\big[\mathrm{Card}(\tilde{z} \in D_{k,\tilde{x}} : \tilde{z}_j = B\tilde{x}_j) - \mathrm{Card}(\tilde{z} \in D_{k,\tilde{x}} : \tilde{z}_j = -B\tilde{x}_j)\big]\tilde{x}_j \\ &= B\left[\binom{d-1}{\frac{d+k}{2} - 1} - \binom{d-1}{\frac{d+k}{2}}\right]\tilde{x}_j .\end{aligned}$$

We now end the proof of Proposition 3.2 when d is odd. Combining (24) with (22), we obtain for d odd

$$\sum_{z \in A_{\tilde{x}}} \tilde{z} = B\binom{d-1}{\frac{d-1}{2}}\tilde{x},$$

and the choice of B yields

$$\begin{aligned}\mathbb{E}[\tilde{Z} \mid X = x] &= \frac{2\pi_\alpha - 1}{2^{d-1}} B\binom{d-1}{\frac{d-1}{2}} \sum_{\tilde{x} \in \{-1,1\}^d} \mathbb{P}(\tilde{X} = \tilde{x} \mid X = x) \cdot \tilde{x} \\ &= \mathbb{E}[\tilde{X} \mid X = x].\end{aligned}$$

Since for all $j \in [\![1, d]\!]$ it holds that

$$\mathbb{E}[\tilde{X}_j \mid X = x] = \mathrm{sgn}[x_j],$$

we obtain for d odd

$$\mathbb{E}[Z \mid X = x] = \mathbb{E}[\tilde{Z} \mid X = x] = \mathbb{E}[\tilde{X} \mid X = x] = f(x),$$

which proves Proposition 3.2 when d is odd. From now on, we assume that d is even. Combining (24) with (23), we obtain

$$\sum_{\tilde{z} \in A_{\tilde{x}}} \tilde{z} = B\binom{d-1}{\frac{d}{2}}\tilde{x} + \sum_{\substack{\tilde{z} \in \{-B,B\}^d : \\ \langle \tilde{z}, \tilde{x} \rangle = 0, \\ \tilde{z}_1 = B\tilde{x}_1}} \tilde{z}.$$

Now, observe that for $\tilde{z} \in \{-B, B\}^d$ and $\tilde{x} \in \{-1, 1\}^d$ it holds that $\langle \tilde{z}, \tilde{x} \rangle = 0$ if and only if $\tilde{z}_j = B\tilde{x}_j$ for exactly $d/2$ subscripts $j \in [\![1, d]\!]$, and $\tilde{z}_j = -B\tilde{x}_j$ for exactly $d/2$ subscripts $j \in [\![1, d]\!]$. We thus have

$$\begin{aligned}\sum_{\substack{\tilde{z} \in \{-B,B\}^d : \\ \langle \tilde{z}, \tilde{x} \rangle = 0 \\ \tilde{z}_1 = B\tilde{x}_1}} \tilde{z}_1 &= B\tilde{x}_1 \cdot \mathrm{Card}\big(\{\tilde{z} \in \{-B, B\}^d : \langle \tilde{z}, \tilde{x} \rangle = 0 \text{ and } \tilde{z}_1 = B\tilde{x}_1\}\big) \\ &= B\binom{d-1}{\frac{d}{2} - 1}\tilde{x}_1,\end{aligned}$$

and for $j \geq 2$ it holds that

$$\begin{aligned}\sum_{\substack{\tilde{z}\in\{-B,B\}^d:\\ \langle\tilde{z},\tilde{x}\rangle=0\\ \tilde{z}_1=B\tilde{x}_1}} \tilde{z}_j &= B\tilde{x}_j\Big[\mathrm{Card}\big(\{\tilde{z}\in\{-B,B\}^d : \langle\tilde{z},\tilde{x}\rangle=0, \tilde{z}_1=B\tilde{x}_1, \tilde{z}_j=B\tilde{x}_j\}\big)\\ &\qquad - \mathrm{Card}\big(\{\tilde{z}\in\{-B,B\}^d : \langle\tilde{z},\tilde{x}\rangle=0, \tilde{z}_1=B\tilde{x}_1, \tilde{z}_j=-B\tilde{x}_j\}\big)\Big]\\ &= B\left[\binom{d-2}{\frac{d}{2}-2}-\binom{d-2}{\frac{d}{2}-1}\right]\tilde{x}_j.\end{aligned}$$

We thus obtain

$$\sum_{\tilde{z}\in A_{\tilde{x}}} \tilde{z}_j = \begin{cases} B\binom{d}{\frac{d}{2}}\tilde{x}_1 & \text{if } j=1,\\ B\left[\binom{d-1}{\frac{d}{2}}+\binom{d-2}{\frac{d}{2}-2}-\binom{d-2}{\frac{d}{2}-1}\right]\tilde{x}_j & \text{if } j\in[\![2,d]\!].\end{cases}$$

The choice

$$B = \frac{2^{d-1}}{2\pi_\alpha - 1}\cdot\frac{(\frac{d}{2}-1)!\frac{d}{2}!}{(d-2)!(d-2)}$$

then yields

$$\begin{aligned}&\mathbb{E}[\tilde{Z}_j \mid X=x]\\ &= \begin{cases} \frac{(2\pi_\alpha-1)B}{2^{d-1}}\binom{d}{\frac{d}{2}}\sum_{\tilde{x}\in\{-1,1\}^d}\tilde{x}_1\mathbb{P}(\tilde{X}=\tilde{x}\mid X=x) & \text{if } j=1,\\ \frac{(2\pi_\alpha-1)B}{2^{d-1}}\cdot\frac{(d-2)!(d-2)}{(\frac{d}{2}-1)!\frac{d}{2}!}\sum_{\tilde{x}\in\{-1,1\}^d}\tilde{x}_j\mathbb{P}(\tilde{X}=\tilde{x}\mid X=x) & \text{if } j\in[\![2,d]\!],\end{cases}\\ &= \begin{cases} \frac{2(d-1)}{d-2}\sum_{\tilde{x}\in\{-1,1\}^d}\tilde{x}_1\mathbb{P}(\tilde{X}=\tilde{x}\mid X=x) & \text{if } j=1,\\ \sum_{\tilde{x}\in\{-1,1\}^d}\tilde{x}_j\mathbb{P}(\tilde{X}=\tilde{x}\mid X=x) & \text{if } j\in[\![2,d]\!].\end{cases}\end{aligned}$$

Thus, it holds $\mathbb{E}[Z_j \mid X=x] = \sum_{\tilde{x}\in\{-1,1\}^d}\mathbb{P}(\tilde{X}=\tilde{x}\mid X=x)\tilde{x}_j$ for all $j\in[\![1,d]\!]$, and

$$\mathbb{E}[Z\mid X=x] = \sum_{\tilde{x}\in\{-1,1\}^d}\mathbb{P}(\tilde{X}=\tilde{x}\mid X=x)\tilde{x} = \mathbb{E}[\tilde{X}\mid X=x] = f(x).$$

B.3. Asymptotic analysis of the value K_d defined in (15)

Lemma B.1. *The value K_d defined in (15) behaves asymptotically in d as*

$$K_d \underset{d\to\infty}{\sim} \sqrt{\frac{\pi}{2}}\sqrt{d}.$$

In particular, it holds $K_d \lesssim \sqrt{d}$ for d large enough.

The proof relies on Stirling's approximation. We first deal with the case where d is odd. In this case, Stirling's approximation yields

$$\begin{aligned} K_d &= 2^{d-1}\frac{[(\frac{d-1}{2})!]^2}{(d-1)!} \\ &\underset{d\to\infty}{\sim} 2^{d-1}\cdot\pi(d-1)\Big(\frac{d-1}{2e}\Big)^{d-1}\cdot\Big[\sqrt{2\pi(d-1)}\Big(\frac{d-1}{e}\Big)^{d-1}\Big]^{-1}. \end{aligned}$$

The right-hand side of this asymptotic equivalence is equal to $\sqrt{\pi/2}\sqrt{d-1}$. We thus obtain

$$K_d \underset{d\to\infty}{\sim} \sqrt{\pi/2}\sqrt{d}.$$

We now assume that d is even. in this case, Stirling's approximation yields

$$\begin{aligned} K_d &= \frac{2^{d-1}(\frac{d}{2}-1)!\frac{d}{2}!}{(d-2)!(d-2)} \\ &\underset{d\to\infty}{\sim} \frac{2^{d-1}}{d-2}\cdot\pi\sqrt{(d-2)d}\Big(\frac{d-2}{2e}\Big)^{\frac{d}{2}-1}\Big(\frac{d}{2e}\Big)^{\frac{d}{2}}\cdot\Big[\sqrt{2\pi(d-2)}\Big(\frac{d-2}{e}\Big)^{d-2}\Big]^{-1}. \end{aligned}$$

The right-hand side of this asymptotic equivalence is equal to

$$\frac{\sqrt{\pi}}{e\sqrt{2}}\sqrt{d}(d-2)^{-\frac{d}{2}}d^{\frac{d}{2}} = \frac{\sqrt{\pi}}{e\sqrt{2}}\sqrt{d}\exp\Big(-\frac{d}{2}\log\Big(1-\frac{2}{d}\Big)\Big) \underset{d\to\infty}{\sim} \sqrt{\frac{\pi}{2}}\sqrt{d},$$

which ends the proof.

B.4. Proof of Proposition 3.3

The proof is similar to the one we made in the Coordinate Local case (Proposition 2.4). However, in the Coordinate Global case, for all $j\in[\![1,d]\!]$ the $(Z_j^i)_i$ are bounded random variables, which will enable us to use Hoeffding's inequality instead of Lemma A.5 and Bernstein's inequality.

Writing

$$|\hat{\eta}^+ - \eta| = \sum_{j:\eta_j=0}\hat{\eta}_j^+ + \sum_{j:\eta_j=1}(1-\hat{\eta}_j^+),$$

we have

$$\mathbb{E}\Big[\frac{1}{s}|\hat{\eta}^+-\eta|\Big] = \frac{1}{s}\sum_{j:\eta_j=0}\underbrace{\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n Z_j^i\geq\tau\Big)}_{=T_{1,j}} + \frac{1}{s}\sum_{j:\eta_j=1}\underbrace{\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n Z_j^i<\tau\Big)}_{=T_{2,j}}.$$

We first study $T_{1,j}$. For j satisfying $\eta_j=0$, it holds that

$$\mathbb{E}[Z_j^i] = \mathbb{E}\big[\mathbb{E}[Z_j^i\mid X^i]\big] = \mathbb{E}\big[\mathrm{sgn}[X_j^i]\big] = \mathbb{E}\big[\mathrm{sgn}[\sigma\xi_j^i]\big] = 0,$$

where we have used Proposition 3.2 and the fact that the distribution of the random variable ξ_j^i is symmetric. Thus, Hoeffding's inequality yields

$$T_{1,j} = \mathbb{P}\bigg(\sum_{i=1}^{n}(Z_j^i - \mathbb{E}[Z_j^i]) \geq n\tau\bigg) \leq \exp\Big(-\frac{n\tau^2}{2B^2}\Big).$$

We now study $T_{2,j}$. Let $j \in [\![1, d]\!]$ such that $\eta_j = 1$. It holds that

$$\begin{aligned} T_{2,j} &= \mathbb{P}\bigg(\frac{1}{n}\sum_{i=1}^{n}\Big(Z_j^i - \mathbb{E}[Z_j^i]\Big) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Z_j^i] < \tau\bigg) \\ &= \mathbb{P}\bigg(\frac{1}{n}\sum_{i=1}^{n}\Big(-Z_j^i - \mathbb{E}[-Z_j^i]\Big) > \mathbb{E}[Z_j^1] - \tau\bigg). \end{aligned}$$

Proposition 3.2 gives

$$\mathbb{E}[Z_j^1] = \mathbb{E}\big[\operatorname{sgn}[X_j^1]\big] = \mathbb{E}\big[\operatorname{sgn}[\theta_j + \sigma\xi_j^1]\big] \geq \mathbb{E}\big[\operatorname{sgn}[a + \sigma\xi_j^1]\big],$$

and we have proved in Appendix A.4 that the following holds:

$$\mathbb{E}\big[\operatorname{sgn}[a + \sigma\xi_1]\big] \geq 2\Phi\Big(\sqrt{c}\,\frac{a}{\sigma}\Big) - 1,$$

where Φ denotes the standard Gaussian cumulative distribution function. Thus, if $a \geq 2\sigma$, it holds that $\mathbb{E}[\operatorname{sgn}[a + \sigma\xi_j^1]] \geq C_1$ with $C_1 = 2\Phi(2\sqrt{c}) - 1$, and

$$T_{2,j} \leq \mathbb{P}\bigg(\frac{1}{n}\sum_{i=1}^{n}(-Z_j^i - \mathbb{E}[-Z_j^i]) > C_1 - \tau\bigg) \leq \exp\bigg(-\frac{n(C_1-\tau)^2}{2B^2}\bigg)$$

according to Hoeffding's inequality if $C_1 - \tau > 0$. This yields

$$\mathbb{E}\bigg[\frac{1}{s}|\widehat{\eta}^+ - \eta|\bigg] \leq \frac{d-|S|}{s}\exp\Big(-\frac{n\tau^2}{2B^2}\Big) + \frac{|S|}{s}\exp\bigg(-\frac{n(C_1-\tau)^2}{2B^2}\bigg).$$

The proof of the second statement of Proposition 3.3 is straightforward.

B.5. Proof of Proposition 3.4

The beginning of the proof is similar to the proof of Proposition 3.3. It holds that

$$\mathbb{E}\bigg[\frac{1}{s}|\widehat{\eta}^+ - \eta|\bigg] = \frac{1}{s}\sum_{j:\eta_j=0}\underbrace{\mathbb{P}\bigg(\frac{1}{n}\sum_{i=1}^{n}Z_j^i \geq \tau\bigg)}_{=T_{1,j}} + \frac{1}{s}\sum_{j:\eta_j=1}\underbrace{\mathbb{P}\bigg(\frac{1}{n}\sum_{i=1}^{n}Z_j^i < \tau\bigg)}_{=T_{2,j}},$$

with

$$T_{1,j} \leq \exp\Big(-\frac{n\tau^2}{2B^2}\Big),$$

and

$$T_{2,j} \leq \mathbb{P}\bigg(\frac{1}{n}\sum_{i=1}^{n}(-Z_j^i - \mathbb{E}[-Z_j^i]) > \mathbb{E}\big[\mathrm{sgn}[a+\sigma\xi_j^i]\big] - \tau\bigg).$$

Moreover, we have proved in Appendix A.6 that $\mathbb{E}\big[\mathrm{sgn}[a+\sigma\xi_1]\big] \geq 2p(2)a/\sigma$ for $a/\sigma < 2$. Thus, if $\tau < 2p(2)a/\sigma$, Hoeffding's inequality yields

$$T_{2,j} \leq \mathbb{P}\bigg(\frac{1}{n}\sum_{i=1}^{n}(-Z_j^i - \mathbb{E}[-Z_j^i]) > \frac{2p(2)a}{\sigma} - \tau\bigg) \leq \exp\bigg(-\frac{n(2p(2)a/\sigma - \tau)^2}{2B^2}\bigg).$$

B.6. Proof of Proposition 3.6

For $i = 1, \dots, d$, define the vector $\omega_i \in \{0,1\}^d$ by $\omega_{i,j} = 1$ if $j = i$, $\omega_{i,j} = 0$ if $j \neq i$ and define P_{ω_i} as the multivariate distribution of the random vector $X = a\omega_i + \sigma\xi$. For $i \neq j$, it holds that

$$|\eta(P_{\omega_i}) - \eta(P_{\omega_j})| = |\omega_i - \omega_j| = 2.$$

The private Fano method ([19, Proposition 2]) thus yields

$$\begin{aligned}
&\inf_{Q\in\mathcal{Q}_\alpha}\inf_{\widehat{\eta}\in\mathcal{T}}\sup_{\theta\in\Theta_d^+(s,a)}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}-\eta| \\
&\qquad \geq \frac{1}{2}\bigg\{1 - \frac{n(e^\alpha-1)^2}{d\log(d)}\bigg[\sup_{\gamma\in\mathbb{B}_\infty(\mathbb{R}^d)}\sum_{i=1}^{d}(\varphi_{\omega_i}(\gamma))^2\bigg] - \frac{\log(2)}{\log(d)}\bigg\},
\end{aligned}$$

with

$$\begin{aligned}
\mathbb{B}_\infty(\mathbb{R}^d) &= \big\{\gamma \in L_\infty(\mathbb{R}^d) \mid \|\gamma\|_\infty \leq 1\big\},\\
\varphi_{\omega_i}(\gamma) &= \int_{\mathcal{X}}\gamma(x)(dP_{\omega_i}(x) - d\bar{P}(x)) = \int_{\mathbb{R}^d}\gamma(x)(f_{\omega_i}(x) - \bar{f}(x))\,\mathrm{d}x,
\end{aligned}$$

where f_{ω_i} is the density of P_{ω_i} and $\bar{f} = (1/d)\sum_{i=1}^{d} f_{\omega_i}$. We have

$$\begin{aligned}
&\sum_{i=1}^{d}(\varphi_{\omega_i}(\gamma))^2 \\
&\quad = \sum_{i=1}^{d}\bigg(\int_{\mathbb{R}^d}\gamma(x)(f_{\omega_i}(x)-\bar{f}(x))\,\mathrm{d}x\bigg)\bigg(\int_{\mathbb{R}^d}\gamma(y)(f_{\omega_i}(y)-\bar{f}(y))\,\mathrm{d}y\bigg)\\
&\quad = \int_{\mathbb{R}^d}\gamma(x)\bigg[\int_{\mathbb{R}^d}\bigg(\sum_{i=1}^{d}(f_{\omega_i}(x)-\bar{f}(x))(f_{\omega_i}(y)-\bar{f}(y))\bigg)\gamma(y)\,\mathrm{d}y\bigg]\mathrm{d}x.
\end{aligned}$$

Let $\overline{p}$ denote the density of the random vector $\sigma\xi$. If γ belongs to $\mathbb{B}_\infty(\mathbb{R}^d)$ then it also belongs to $L_2(\mathbb{R}^d, dq)$ and, moreover, $\|\gamma\|_{L_2(\mathbb{R}^d, d\overline{p})} \leq 1$. We can write

$$\begin{aligned}
&\sum_{i=1}^{d} (\varphi_{\omega_i}(\gamma))^2 \\
&= \int_{\mathbb{R}^d} \gamma(x) \Bigg[\int_{\mathbb{R}^d} \left(\sum_{i=1}^{d} \frac{f_{\omega_i}(x) - \overline{f}(x)}{\overline{p}(x)} \cdot \frac{f_{\omega_i}(y) - \overline{f}(y)}{\overline{p}(y)} \right) \gamma(y) \overline{p}(y)\, \mathrm{d}y \Bigg] \overline{p}(x)\, \mathrm{d}x \\
&= \langle \gamma, K\gamma \rangle_{L_2(\mathbb{R}^d, d\overline{p})},
\end{aligned}$$

where

$$\begin{aligned}
K \colon L_2(\mathbb{R}^d, d\overline{p}) &\to L_2(\mathbb{R}^d, d\overline{p}) \\
\gamma &\mapsto \int_{\mathbb{R}^d} \left(\sum_{i=1}^{d} \frac{f_{\omega_i} - \overline{f}}{\overline{p}}(\cdot) \cdot \frac{f_{\omega_i}(y) - \overline{f}(y)}{\overline{p}(y)} \right) \gamma(y) \overline{p}(y)\, \mathrm{d}y.
\end{aligned}$$

For any $\omega \in \{0,1\}^d$, $f_\omega \in L_2(\mathbb{R}^d, d\overline{p})$. Note that we can rewrite

$$K\gamma = \sum_{i=1}^{d} \left[\left\langle \frac{f_{\omega_i} - \overline{f}}{\overline{p}}, \gamma \right\rangle_{L_2(\mathbb{R}^d, d\overline{p})} \cdot \frac{f_{\omega_i} - \overline{f}}{\overline{p}} \right].$$

This expression implies that K is an operator of finite rank (it is thus a compact operator), K is self-adjoint, and $\langle K\gamma, \gamma \rangle \geq 0$ for all $\gamma \in L_2(\mathbb{R}^d, d\overline{p})$. In particular, the last point implies that the eigenvalues of K are non-negative. We have

$$\begin{aligned}
\sup_{\gamma \in \mathbb{B}_\infty(\mathbb{R}^d)} \sum_{i=1}^{d} (\varphi_{\omega_i}(\gamma))^2 &\leq \sup_{\{\gamma \in L_2(\mathbb{R}^d, d\overline{p}) : \|\gamma\|^2_{L_2(\mathbb{R}^d, d\overline{p})} \leq 1\}} \langle \gamma, K\gamma \rangle_{L_2(\mathbb{R}^d, d\overline{p})} \\
&= \sup_{\{\gamma \in L_2(\mathbb{R}^d, d\overline{p}) : \|\gamma\|^2_{L_2(\mathbb{R}^d, d\overline{p})} = 1\}} \langle \gamma, K\gamma \rangle_{L_2(\mathbb{R}^d, d\overline{p})} \\
&= \sup_{\{\gamma \in L_2(\mathbb{R}^d, d\overline{p}) : \|\gamma\|^2_{L_2(\mathbb{R}^d, d\overline{p})} = 1\}} |\langle \gamma, K\gamma \rangle_{L_2(\mathbb{R}^d, d\overline{p})}| \\
&= \|K\|,
\end{aligned}$$

where the last equality follows from the fact that $(L_2(\mathbb{R}^d, d\overline{p}), \langle \cdot, \cdot \rangle_{L_2(\mathbb{R}^d, d\overline{p})})$ is a Hilbert space and K is self-adjoint. Since K is also compact and since the eigenvalues of K are non-negative it follows that

$$\sup_{\gamma \in \mathbb{B}_\infty(\mathbb{R}^d)} \sum_{i=1}^{d} (\varphi_{\omega_i}(\gamma))^2 \leq \|K\| = \max\{|\lambda| : \lambda \in VP(T)\} = \max\{\lambda : \lambda \in VP(T)\},$$

where $VP(T)$ is the set of all the eigenvalues of K. It remains to compute this maximum. By definition, λ is an eigenvalue of K if $\lambda I - K$ is not injective. For $\lambda \neq 0$, the Fredholm alternative for compact self-adjoint operators (see, for instance, [26]) implies that $\lambda I - K$ is not injective if and only if $\lambda I - K$ is not surjective. Thus, the non-zero eigenvalues of K are the values of $\lambda \in \mathbb{R}^*$ such that the operator $\lambda I - K$ is not surjective. For $\lambda \in \mathbb{R}$, let A_λ be the matrix with coefficients

$$(A_\lambda)_{ij} = \left\langle \frac{f_{\omega_i} - \overline{f}}{\overline{p}}, \frac{f_{\omega_j} - \overline{f}}{\overline{p}} \right\rangle_{L_2(\mathbb{R}^d, d\overline{p})} - \lambda \delta_{ij}, \quad i, j \in [\![1, d]\!],$$

where δ is the Kronecker delta. The following result proves that if λ is a non-zero eigenvalue of K then it holds $\mathrm{Det}(A_\lambda) = 0$.

Lemma B.2. *Let $\lambda \in \mathbb{R}$, $\lambda \neq 0$. If $\mathrm{Det}(A_\lambda) \neq 0$, then $\lambda I - K$ is surjective.*

Proof. To lighten the notation, set $\langle \cdot, \cdot \rangle_{2,\overline{p}} = \langle \cdot, \cdot \rangle_{L_2(\mathbb{R}^d, d\overline{p})}$. Let $\lambda \in \mathbb{R}$, $\lambda \neq 0$ and assume that $\mathrm{Det}(A_\lambda) \neq 0$. We prove that for all $g \in L_2(\mathbb{R}^d, d\overline{p})$, there exists $\gamma \in L_2(\mathbb{R}^d, d\overline{p})$ such that $g = (\lambda I - K)\gamma$. Consider $g \in L_2(\mathbb{R}^d, d\overline{p})$. Since $\mathrm{Det}(A_\lambda) \neq 0$, the matrix A_λ is invertible and for all $v \in \mathbb{R}^d$ there exists $\xi \in \mathbb{R}^d$ such that $v = A_\lambda \xi$. In particular, for

$$v = \left(\left\langle \frac{f_{\omega_1} - \overline{f}}{\overline{p}}, g \right\rangle_{2,\overline{p}}, \ldots, \left\langle \frac{f_{\omega_d} - \overline{f}}{\overline{p}}, g \right\rangle_{2,\overline{p}} \right)^T,$$

there exists $\xi \in \mathbb{R}^d$ such that $v = A_\lambda \xi$, that is

$$\left\langle \frac{f_{\omega_i} - \overline{f}}{\overline{p}}, g \right\rangle_{2,\overline{p}} = (A_\lambda \xi)_i = \sum_{j=1}^{d} \left\langle \frac{f_{\omega_i} - \overline{f}}{\overline{p}}, \frac{f_{\omega_j} - \overline{f}}{\overline{p}} \right\rangle_{2,\overline{p}} \xi_j - \lambda \xi_i$$

for all $i \in [\![1, d]\!]$. Define

$$\gamma = \frac{1}{\lambda} g - \frac{1}{\lambda} \sum_{j=1}^{d} \xi_j \frac{f_{\omega_j} - \overline{f}}{\overline{p}}.$$

We have

$$\begin{aligned}
&(\lambda I - K)\gamma = \lambda\gamma - K\gamma \\
&= g - \sum_{i=1}^{d} \xi_i \frac{f_{\omega_i} - \overline{f}}{\overline{p}} - \sum_{i=1}^{d} \left[\left\langle \frac{f_{\omega_i} - \overline{f}}{\overline{p}}, \gamma \right\rangle_{L_2(\mathbb{R}^d, d\overline{p})} \cdot \frac{f_{\omega_i} - \overline{f}}{\overline{p}} \right]
\end{aligned}$$

$$= g - \sum_{i=1}^{d} \underbrace{\left[\xi_i + \frac{1}{\lambda}\left\langle \frac{f_{\omega_i} - \bar{f}}{\bar{p}}, g\right\rangle_{2,\bar{p}} - \frac{1}{\lambda}\sum_{j=1}^{d} \xi_j \left\langle \frac{f_{\omega_i} - \bar{f}}{\bar{p}}, \frac{f_{\omega_j} - \bar{f}}{\bar{p}}\right\rangle_{2,\bar{p}}\right]}_{=0} \frac{f_{\omega_i} - \bar{f}}{\bar{p}}$$

$$= g,$$

which concludes the proof of the lemma. ■

We now find the values of λ for which we have $\mathrm{Det}(A_\lambda) = 0$. To do so, we first make explicit the coefficients of A_λ. It holds that

$$\begin{aligned}
&\left\langle \frac{f_{\omega_i} - \bar{f}}{\bar{p}}, \frac{f_{\omega_j} - \bar{f}}{\bar{p}}\right\rangle_{2,\bar{p}} \\
&\quad = \left\langle \frac{f_{\omega_i}}{\bar{p}}, \frac{f_{\omega_j}}{\bar{p}}\right\rangle_{2,\bar{p}} - \left\langle \frac{f_{\omega_i}}{\bar{p}}, \frac{\bar{f}}{\bar{p}}\right\rangle_{2,\bar{p}} - \left\langle \frac{\bar{f}}{\bar{p}}, \frac{f_{\omega_j}}{\bar{p}}\right\rangle_{2,\bar{p}} + \left\langle \frac{\bar{f}}{\bar{p}}, \frac{\bar{f}}{\bar{p}}\right\rangle_{2,\bar{p}} \\
&\quad = \left\langle \frac{f_{\omega_i}}{\bar{p}}, \frac{f_{\omega_j}}{\bar{p}}\right\rangle_{2,\bar{p}} - \frac{1}{d}\sum_{k=1}^{d} \left\langle \frac{f_{\omega_i}}{\bar{p}}, \frac{f_{\omega_k}}{\bar{p}}\right\rangle_{2,\bar{p}} \\
&\qquad\qquad - \frac{1}{d}\sum_{k=1}^{d} \left\langle \frac{f_{\omega_k}}{\bar{p}}, \frac{f_{\omega_j}}{\bar{p}}\right\rangle_{2,\bar{p}} + \frac{1}{d^2}\sum_{k=1}^{d}\sum_{l=1}^{d} \left\langle \frac{f_{\omega_k}}{\bar{p}}, \frac{f_{\omega_l}}{\bar{p}}\right\rangle_{2,\bar{p}}.
\end{aligned}$$

Furthermore, due to the independence of the coordinates of the vector ξ, the scalar products $\langle \frac{f_{\omega_k}}{\bar{p}}, \frac{f_{\omega_l}}{\bar{p}}\rangle_{2,\bar{p}}$ can only take two values. More precisely, recall that P_0 denotes the distribution of the random variable $\sigma\xi_1$ and P_a the distribution of the random variable $a + \sigma\xi_1$, we get

$$\left\langle \frac{f_{\omega_i}}{\bar{p}}, \frac{f_{\omega_j}}{\bar{p}}\right\rangle_{2,\bar{p}} = \begin{cases} 1 + \chi^2(P_0, P_a) & \text{if } j = i, \\ 1 & \text{if } j \neq i. \end{cases}$$

We thus obtain

$$\left\langle \frac{f_{\omega_i} - \bar{f}}{\bar{p}}, \frac{f_{\omega_j} - \bar{f}}{\bar{p}}\right\rangle_{2,\bar{p}} = \begin{cases} \left(1 - \frac{1}{d}\right)\chi^2(P_0, P_a) & \text{if } j = i, \\ -\frac{1}{d}\chi^2(P_0, P_a) & \text{if } j \neq i. \end{cases}$$

Write

$$C_1 = \left(1 - \frac{1}{d}\right)\chi^2(P_0, P_a) \quad \text{and} \quad C_2 = -\frac{1}{d}\chi^2(P_0, P_a).$$

The matrix A_λ has its diagonal elements equal to $C_1 - \lambda$ and the other coefficients equal to C_2. Operations on the rows and columns of A_λ yield

$$\mathrm{Det}(A_\lambda) = \big(C_1 + (d-1)C_2 - \lambda\big)(C_1 - C_2 - \lambda)^{d-1} = -\lambda\big(\chi^2(P_0, P_a) - \lambda\big)^{d-1}.$$

Thus, the operator K has only one non-zero eigenvalue and it is equal to $\chi^2(P_0, P_a)$. We finally obtain

$$\inf_{Q\in\mathcal{Q}_\alpha}\inf_{\widehat{\eta}\in\mathcal{T}}\sup_{\theta\in\Theta_d^+(s,a)}\mathbb{E}_{Q(P_\theta^{\otimes n})}|\widehat{\eta}-\eta| \geq \frac{1}{2}\left(1-\frac{n(e^\alpha-1)^2}{d\log(d)}\chi^2(P_0,P_a)-\frac{\log(2)}{\log(d)}\right)$$
$$\geq \frac{1}{4}\left(1-\frac{2n(e^\alpha-1)^2}{d\log(d)}\chi^2(P_0,P_a)\right),$$

if $d \geq 4$. To conclude with the proof of Proposition 3.6, just use Lemma B.3 below.

Lemma B.3. *Consider that the measure P^{ξ_1} of the noise coordinates has a density $p = \exp(-\phi)$, where the potential ϕ is two times continuously differentiable and has a curvature that is bounded from above by a constant c_+ as in (3). Then it holds that*

$$\chi^2(P_0,P_a) \leq \exp\left(c_+\left(\frac{a}{\sigma}\right)^2\right)-1. \tag{25}$$

If the density p is log-concave, with a potential with curvature bounded above by c_+ as in (3), then inequality (25) holds without assuming the differentiability of ϕ.

Note that Lemma B.3 is sharp in the sense that in the Gaussian case, $c_+ = 1$ holds and Inequality (25) turns out to be an equality. Note also that log-concavity is actually not needed in Lemma B.3, since we only require an upper bound on the curvature of the potential ϕ.

Proof. Denote $\overline{a} = a/\sigma$. It suffices to show the following inequality,

$$\int_{\mathbb{R}}\frac{p^2(x-\overline{a})}{p(x)}\,\mathrm{d}x \leq \exp(c_+\overline{a}^2), \tag{26}$$

where we recall that p is the density of ξ_1. It holds

$$\int_{\mathbb{R}}\frac{p^2(x-\overline{a})}{p(x)}\,\mathrm{d}x = \int_{\mathbb{R}}\exp\big(\phi(x)-2\phi(x-\overline{a})\big)\,\mathrm{d}x.$$

As ϕ is two times continuously differentiable, for all $x\in\mathbb{R}$, we have by Taylor expansion

$$\phi(x)-\phi(x-\overline{a}) \leq \overline{a}\phi'(x-\overline{a})+c_+\frac{\overline{a}^2}{2}$$

and

$$\phi(x-2\overline{a})-\phi(x-\overline{a}) \leq -\overline{a}\phi'(x-\overline{a})+c_+\frac{\overline{a}^2}{2}.$$

By adding the two previous inequalities, we get

$$\phi(x)-2\phi(x-\overline{a}) \leq -\phi(x-2\overline{a})+c_+\overline{a}^2.$$

This gives

$$\begin{aligned}\int_{\mathbb{R}} \frac{p^2(x-\overline{a})}{p(x)}\,\mathrm{d}x &= \int_{\mathbb{R}} \exp\bigl(\phi(x) - 2\phi(x-\overline{a})\bigr)\,\mathrm{d}x \\ &\le \exp(c_+\overline{a}^2) \int_{\mathbb{R}} \bigl(\exp(-\phi(x-2\overline{a}))\bigr)\,\mathrm{d}x \\ &= \exp(c_+\overline{a}^2).\end{aligned}$$

We proved (26). In the case where p is log-concave, it can be suitably approximated by infinitely differentiable densities, via the use of convolutions with Gaussian random variables, which completes the proof of Lemma B.3. ■

Acknowledgments. The authors are most grateful to the Associate Editor and to the anonymous referees for their suggestions that allowed us to improve some log factors and for challenging us to extend the framework of our results.

Funding. Cristina Butucea was supported by the French National Research Agency (ANR) under the grant Labex Ecodec (ANR-11-LABEX-0047).

References

[1] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.* **23** (2010), no. 2, 535–561 Zbl 1206.60006 MR 2601042

[2] I. Aden-Ali, H. Ashtiani, and G. Kamath, On the sample complexity of privately learning unbounded high-dimensional Gaussians. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, pp. 185–216, Proceedings of Machine Learning Research 132, PMLR, 2021 MR 4227321

[3] M. Bafna and J. Ullman, The price of selection in differential privacy. In *COLT 2017 – Proceedings of the 2017 Conference on Learning Theory*, pp. 151–168, Proceedings of Machine Learning Research 65, PMLR, 2017

[4] M. Bagnoli and T. Bergstrom, Log-concave probability and its applications. *Econom. Theory* **26** (2005), no. 2, 445–469 Zbl 1077.60012 MR 2213177

[5] T. B. Berrett and C. Butucea, Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. In *NeurIPS 2020 – Advances in Neural Information Processing Systems 33*, pp. 3164–3173, Curran Associates, 2020

[6] S. Biswas, Y. Dong, G. Kamath, and J. Ullmann, Coinpress: Practical private mean and covariance estimation. In *NeurIPS 2020 – Advances in Neural Information Processing Systems 33*, pp. 14475–14485, Curran Associates, 2020

[7] S. Bobkov and M. Ledoux, From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.* **10** (2000), no. 5, 1028–1052 Zbl 0969.26019 MR 1800062

[8] S. Bobkov and M. Ledoux, One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** (2019), no. 1259 Zbl 1454.60007 MR 4028181

[9] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*. Oxford University Press, Oxford, 2013 Zbl 1279.60005 MR 3185193

[10] C. Butucea, A. Dubois, M. Kroll, and A. Saumard, Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli* **26** (2020), no. 3, 1727–1764 Zbl 1444.62045 MR 4091090

[11] C. Butucea, M. Ndaoud, N. A. Stepanova, and A. B. Tsybakov, Variable selection with Hamming loss. *Ann. Statist.* **46** (2018), no. 5, 1837–1875 Zbl 1414.62126 MR 3845003

[12] C. Butucea, A. Rohde, and L. Steinberger, Interactive versus non-interactive locally, differentially private estimation: Two elbows for the quadratic functional. *Ann. Statist.*, to appear.

[13] L. A. Caffarelli, The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5** (1992), no. 1, 99–104 Zbl 0753.35031 MR 1124980

[14] L. A. Caffarelli, Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** (2000), no. 3, 547–563 Zbl 0978.60107 MR 1800860

[15] D. Cordero-Erausquin, Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.* **161** (2002), no. 3, 257–269 Zbl 0998.60080 MR 1894593

[16] M. Cule and R. Samworth, Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** (2010), 254–270 Zbl 1329.62183 MR 2645484

[17] A. S. Dalalyan, Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** (2017), no. 3, 651–676 Zbl 1411.62030 MR 3641401

[18] C. R. Doss and J. A. Wellner, Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** (2016), no. 3, 954–981 Zbl 1338.62101 MR 3485950

[19] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** (2018), no. 521, 182–201 Zbl 1398.62021 MR 3803452

[20] A. Durmus and É. Moulines, Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** (2017), no. 3, 1551–1587 Zbl 1377.65007 MR 3678479

[21] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT 2006 – Advances in cryptology*, pp. 486–503, Lecture Notes in Comput. Sci. 4004, Springer, Berlin, 2006 Zbl 1140.94336 MR 2423560

[22] C. Dwork, F. McSherry, K. Nissim, and A. Smith, Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, pp. 265–284, Lecture Notes in Comput. Sci. 3876, Springer, Berlin, 2006 Zbl 1112.94027 MR 2241676

[23] O. Guédon, Concentration phenomena in high dimensional geometry. *Journées MAS (2012)*, pp. 47–60, ESAIM Proc. 44, EDP Sci., Les Ulis, 2014 Zbl 1358.52013 MR 3178607

[24] Q. Han, Set structured global empirical risk minimizers are rate optimal in general dimensions. *Ann. Statist.* **49** (2021), no. 5, 2642–2671 Zbl 1478.62081 MR 4338378

[25] E. Hillion, O. Johnson, and A. Saumard, An extremal property of the normal distribution, with a discrete analog. *Statist. Probab. Lett.* **145** (2019), 181–186 Zbl 1414.62052 MR 3873905

[26] F. Hirsch and G. Lacombe, *Elements of functional analysis*. Grad. Texts in Math. 192, Springer, New York, 1999 Zbl 0924.46001 MR 1678925

[27] M. C. Jones and A. Noufaily, Log-location-scale-log-concave distributions for survival and reliability analysis. *Electron. J. Stat.* **9** (2015), no. 2, 2732–2750 Zbl 1329.62409 MR 3435809

[28] M. Ledoux, *The concentration of measure phenomenon*. Math. Surveys Monogr. 89, American Mathematical Society, Providence, RI, 2001 Zbl 0995.60002 MR 1849347

[29] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*. Springer Texts Statist., Springer, New York, 2005 Zbl 1076.62018 MR 2135927

[30] M. Ndaoud, Sharp optimal recovery in the two component Gaussian mixture model. *Ann. Statist.* **50** (2022), no. 4, 2096–2126 Zbl 07610764 MR 4474484

[31] M. Ndaoud and A. B. Tsybakov, Optimal variable selection and adaptive noisy compressed sensing. *IEEE Trans. Inform. Theory* **66** (2020), no. 4, 2517–2532 Zbl 1448.94081 MR 4087700

[32] L. Paninski, Log-concavity results on gaussian process methods for supervised and unsupervised learning. In *NIPS 2004 – Advances in Neural Information Processing Systems 17*, pp. 1025–1032, MIT Press, Cambridge, MA, 2004

[33] A. Rohde and L. Steinberger, Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.* **48** (2020), no. 5, 2646–2670 Zbl 1457.62396 MR 4152116

[34] A. Saumard and J. A. Wellner, Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** (2014), 45–114 Zbl 1360.62055 MR 3290441

[35] T. Steinke and J. Ullman, Tight lower bounds for differentially private selection. In *FOCS 2017 – 58th Annual IEEE Symposium on Foundations of Computer Science*, pp. 552–563, IEEE Computer Soc., Los Alamitos, CA, 2017 MR 3734260

[36] A. B. Tsybakov, *Introduction to nonparametric estimation*. Springer Ser. Statist., Springer, New York, 2009 Zbl 1176.62032 MR 2724359

[37] J. Ullman, Tight lower bounds for locally differentially private selection. 2018, arXiv:1802.02638

[38] L. Wasserman and S. Zhou, A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** (2010), no. 489, 375–389 Zbl 1364.62011 MR 2656057

Received 24 November 2020; revised 24 June 2022.

Cristina Butucea
CREST, ENSAE Paris, IP Paris, 5, avenue Henry le Chatelier, 91120 Palaiseau, France;
cristina.butucea@ensae.fr

Amandine Dubois
CREST, ENSAI, Campus de Ker-Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz, France;
amandine.dubois@ensai.fr

Adrien Saumard
CREST, ENSAI, Campus de Ker-Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz, France;
adrien.saumard@ensai.fr