Minimax rate for multivariate data under componentwise local differential privacy constraints

Chiara Amorino* Arnaud Gloter †

May 19, 2023

Abstract

Our research delves into the balance between maintaining privacy and preserving statistical accuracy when dealing with multivariate data that is subject to *componentwise local differential privacy* (CLDP). With CLDP, each component of the private data is made public through a separate privacy channel. This allows for varying levels of privacy protection for different components or for the privatization of each component by different entities, each with their own distinct privacy policies. We develop general techniques for establishing minimax bounds that shed light on the statistical cost of privacy in this context, as a function of the privacy levels $\alpha_1, ..., \alpha_d$ of the d components.

We demonstrate the versatility and efficiency of these techniques by presenting various statistical applications. Specifically, we examine nonparametric density and covariance estimation under CLDP, providing upper and lower bounds that match up to constant factors, as well as an associated data-driven adaptive procedure. Furthermore, we quantify the probability of extracting sensitive information from one component by exploiting the fact that, on another component which may be correlated with the first, a smaller degree of privacy protection is guaranteed.

Keywords: local differential privacy, minimax optimality, rate of convergence, non parametric estimation, density estimation

AMS 2010 subject classifications: Primary 62C20, secondary 62G07, 62H05

1 Introduction

In the current era of information technology, protecting data privacy has become a significant challenge for statistical inference. With the widespread collection and storage of massive amounts of data, including medical records, social media activity and smartphone user behavior, individuals are increasingly reluctant to share sensitive information with companies or state officials. To address this issue, researchers in computer science and related fields have produced a vast literature on constructing privacy-preserving data release mechanisms. Real-world applications have also emerged, with companies such as Apple [35], Google [21] and Microsoft [12] developing data analysis methodologies that achieve strong statistical performance while maintaining individuals' privacy. This interest has been driven by regulatory pressure and the need to comply with privacy laws (see for example [23, 2]).

A highly effective method of protecting data from privacy breaches consists in differential privacy (see the landmarks [19, 20] as well as [18, 22]). It involves the use of randomized data perturbation, where the original data is replaced with a modified version that maintains the overall statistical properties of the original data, but is different enough to prevent individual

*Université du Luxembourg, L-4364 Esch-Sur-Alzette, Luxembourg. The author gratefully acknowledges financial support of ERC Consolidator Grant 815703 "STAMFORD: Statistical Methods for High Dimensional Diffusions".

†Laboratoire de Mathématiques et Modélisation d'Evry, CNRS, Univ Evry, Université Paris-Saclay, 91037, Evry, France.

data points from being identifiable. This approach offers a high degree of plausible deniability to data providers, as they can argue that their true answer was different from the one provided.

There are two main types of differential privacy: local privacy and central privacy. Local privacy involves privatizing data before sharing it with a data collector, while central privacy involves a centralized curator who maintains the sample and guarantees that any information it releases is appropriately private. While the local model provides stronger privacy protections, it also involves some loss of statistical efficiency. Nevertheless, major technology companies such as Apple and Google (see [3] and [1], respectively) have adopted local differential privacy protections in their data collection and machine learning tools to comply with regulations, protect sensitive data, and maintain transparency. In this paper, the focus is on the local version of differential privacy, which is formally defined in Section 2.

Recently, there has been growing interest in studying differential privacy from a statistical inference perspective. Traditional approaches for locally private analysis have been limited to estimating parameters of a binomial distribution [37]. However, modern research has produced mechanisms for a wide range of statistical problems, including mean and median estimation in [17], hypothesis testing (see for example [27, 26, 4, 29]), robustness [31], change point analysis [6, 30] and nonparametric estimation [11, 10, 28, 5], among others. In light of the increasing significance of data protection, it is crucial to find a balance between statistical utility and privacy: it is essential to ensure that data remains protected from privacy breaches while also allowing for the extraction of useful information and insights. Therefore, finding the optimal balance between these two aspects has become increasingly important.

This paper examines n independent and identically distributed multivariate datasets with law $\boldsymbol{X} = (X^1, ..., X^d)$, subject to what we call *componentwise local differential privacy constraints*. *Componentwise local differential privacy (CLDP)* is a term here introduced and refers to the method of separately making each component public through different privacy channels. This approach can be beneficial as different components may require varying levels of privacy protection or can not be privatized jointly. We denote by α_j the amount of privacy ensured to the component X^j. Intuitively, $\alpha_j = 0$ guarantees perfect privacy while as α_j increases towards infinity, the privacy constraints become less strict. The reader can refer to Equation (2.2) for a precise definition. The study focuses on exploring the trade-off between privacy protection and efficient statistical inference and aims to determine the optimal mechanisms for preserving privacy in this context.

Let us take the example where data is collected from n individuals, comprising d different aspects of their life. For instance, one component could represent data related to sport practice, which is widely available on phone applications, while another component could concern medical expenses. It is evident that disclosing information about the first component has a different impact than disclosing the same about the second, given that maintaining high confidentiality for medical bills is more important. Since some of the components may be correlated, it is necessary to work within a framework that takes this into account. One could wonder if it is possible to extract sensitive information on one component by taking advantage of the fact that on another component, possibly correlated to the first, a smaller amount of privacy is guaranteed.

We give a first answer to this question in Proposition 4.1, where we quantify the amount of private information X^1 carried by the privatized views of the other components in term of the dependence of $(X^2, ..., X^d)$ on X^1 and of privacy levels α_1, $\alpha_{\max}$, where $\alpha_{\max}$ is an upper bound for the levels of privacy ensured on the components $X^2, ..., X^d$.

Our research is also motivated by situations where different components can not be privatized jointly. It may occur when practical aspects prevent the joint components to be gathered by the same organism prior to the privatisation mechanism. Consider the situation where two different entities have collected data on different aspects of the life of n individuals. These two entities could be, for instance, a health insurance company and a tax office having respectively collected health and income data on a population. Let denote by $(X_i^j)_{i=1,\dots,n}$ the data set belonging to the entity j, with $j \in \{1,2\}$ and assume that none of the two entities is disposed to publicly reveal their data. A statistician interested in inferring the joint law of $\boldsymbol{X} = (X^1, X^2)$ would face a componentwise privatisation. Indeed, each entity can still privatize its own data on the individual i, independently of the knowledge of the data owned by the other entity. Such mechanism yields to the privatization of the vector $\boldsymbol{X}_i = (X_i^1, X_i^2)$ component by component. These examples are encouraging us to explore balancing statistical utility and individual privacy (on a componentwise basis) for the people from whom data is obtained. By using a framework that considers componentwise local differential privacy constraints, we are able to identify optimal privacy mechanisms for some statistical problems and characterize how the optimal rate of estimation varies as a function of the privacy levels α_j's. Our approach involves developing general techniques for deriving minimax bounds that not only provide insights into the "statistical price of privacy," but also enable us to compare different privacy mechanisms for producing private data.

With a similar goal in mind but in the case where all the components of one vector are made public through the same privacy channel, Duchi, Jordan and Wainwright proposed the private version of the Le Cam, Fano and Assouad lemmas (see [14, 15, 16, 17]). They provide minimax rates of convergence for specific estimation problems under privacy constraints through a case-by-case study. Rohde and Steinberger's research [33], published in 2020, takes a different approach by developing a general theory, similar to that of Donoho and Liu in [13], to characterize the differentially private minimax rate of convergence using the moduli of continuity.

It is worth highlighting that the minimax approach developed under local differential privacy constraints in [17] enables the authors to examine the private minimax rate of estimation for various classical problems, including mean, median, and density estimation. These bounds have proven to be vital in other research studies that analyze the impact of privacy constraints on the convergence rate of various estimation problems, such as those discussed in [28], [10], [34], or [25], to name a few. The broad range of applications and their diversity demonstrate the significant impact of such research. However, despite the existence of multivariate data, there is currently no work that considers separate components made public with varying levels of privacy. Our goal is to address this research gap.

The novel contribution of this article is to develop bounds to quantify the contraction in Kullback-Liebler divergence that arises from passing multivariate data through d different private channels. These bounds enable us to understand how the optimal convergence rate varies as a function of the privacy levels α_1, ..., α_d, which characterizes the statistical price of privacy. We present statistical applications of such bounds to demonstrate their efficiency and versatility comprehensively. Specifically, we detail the estimation of density and covariance under componentwise local differential privacy constraints. Although our main results are proven under the general framework of sequentially interactive privacy mechanism, we simplify the notation by considering non-interactive algorithms for the two statistical problems.

To elaborate, we have two sets of raw data samples $\boldsymbol{X} = (X^1, ..., X^d)$ and $\tilde{\boldsymbol{X}} = (\tilde{X}^1, ..., \tilde{X}^d)$, each drawn from a probability distribution P and $\tilde{P}$ respectively. We also have two corresponding sets of privatized samples $\boldsymbol{Z} = (Z^1, ..., Z^d)$ and $\tilde{\boldsymbol{Z}} = (\tilde{Z}^1, ..., \tilde{Z}^d)$, where Z^j and $\tilde{Z}^j$ are the α_j-local differential privatized views of X^j and $\tilde{X}^j$, respectively. The following equation explains the closeness of the laws of the privatized samples based on the proximity of the original laws:

$$d_{KL}(L_{\boldsymbol{Z}}, L_{\tilde{\boldsymbol{Z}}}) \leq \Big(\sum_{k=1}^{d} \sum_{(j_1,\dots,j_k)} \prod_{i=1}^{k} (e^{\alpha_{j_i}} - 1) d_{TV}(L_{(X^{j_1},\dots,X^{j_k})}, L_{(\tilde{X}^{j_1},\dots,\tilde{X}^{j_k})}) \Big)^2. \tag{1.1}$$

Here, $L_{(X^{j_1},\dots,X^{j_k})}$ represents the law of the marginals $X^{j_1}, ..., X^{j_k}$ of $\boldsymbol{X}$, and the inner summation is over $1 \leq j_1 < \cdots < j_k \leq d$ where any k distinct indexes in $1, \dots, d$ are considered.
This bound is useful for proving lower bounds for statistical problems and so it will be often used

with two specific priors that are chosen by statisticians. In this case, it is helpful to use priors that have equal $d-1$ marginals, which simplifies the bound above to the following expression

$$d_{KL}(L_{\boldsymbol{Z}}, L_{\tilde{\boldsymbol{Z}}}) \leq \left(\prod_{j=1}^{d} (e^{\alpha_j} - 1) d_{TV}(P, \tilde{P}) \right)^2 .$$

We can compare our result with Theorem 1 of [17], which assumes that only one privacy channel has been used (so $\alpha_1 = ... = \alpha_d = \alpha$). It provides the result $d_{KL}(L_{\boldsymbol{Z}}, L_{\tilde{\boldsymbol{Z}}}) \leq \min(4, e^{2\alpha})(e^{\alpha} - 1)^2 d_{TV}(P, \tilde{P})^2$. Our bound is different because it recovers a factor of $(e^{\alpha} - 1)^{2d}$ instead of $(e^{\alpha} - 1)^2$, when comparing the contribution of the law of the whole vector to the contribution of its one-dimensional marginals. This entails a faster decay in α, for small α.

Using Equation (1.1), we can analyze the rate of convergence for nonparametric density estimation of a vector $\boldsymbol{X}$ belonging to an Hölder class $\mathcal{H}(\beta, \mathcal{L})$. We propose a kernel density estimator based on the observation of privatized variables Z_i^j, where $i = 1, ..., n$ and $j = 1, ..., d$ (refer to (4.20) for details). By imposing the conditions $\alpha_j \leq 1$ and $n \prod_{i=1}^{d} \alpha_j^2 \to \infty$, we demonstrate that the L^2 pointwise error of this estimator reaches the convergence rate $(\frac{1}{n \prod_{i=1}^{d} \alpha_j^2})^{\frac{\beta}{\beta+d}}$. This rate is optimal in a minimax sense for small α (refer to Theorems 4.14, 4.17 below).
It is natural to compare the convergence rate of our kernel density estimator with that of non-componentwise local privacy constraints. According to [10] the latter achieves, for $\alpha < 1$, a convergence rate of $(n(e^{\alpha} - 1)^2)^{-\frac{2\beta}{2\beta+2}} \approx (n\alpha^2)^{-\frac{\beta}{\beta+1}}$ for estimating the density of a vector $\boldsymbol{X}$ belonging to an Hölder class $\mathcal{H}(\beta, \mathcal{L})$ (see Remark 4.19 below for more details).
Our results are consistent with those in [10] when $d = 1$, and they provide some extensions for $d > 1$. In particular, when $\alpha_1 = ... = \alpha_d = \alpha$, the role of α^2 in [10] is replaced by α^{2d} in our analysis.

Furthermore, we provide a detailed analysis of the estimation of the covariance matrix of a two-dimensional vector (X^1, X^2) under componentwise privacy constraints, in addition to the density estimation discussed above. Here, we again find that under componentwise privacy mechanism, the quality of the estimation of the covariance is degraded as α becomes small, compared to a joint privacy mechanism (see Remark 4.6).

The paper is organized as follows. In Section 2, we provide an introduction to differential privacy and present our notation for componentwise local differential privacy and minimax risk. Our main results are presented in Section 3, where we derive bounds on divergence between pairs in Section 3.1 and extend them to the case of interactive privatization of independent sampling in Section 3.2. We demonstrate the practical application of our results in statistical problems in Section 4. Firstly, in Section 4.1, we use our techniques to investigate the precision of revealing one marginal of $\boldsymbol{X}$ by observing $\boldsymbol{Z}$. Next, in Section 4.2, we focus on the problem of estimating the covariance: we propose a private estimator and establish upper and lower bounds for its L^2 risk in Sections 4.2.1 and 4.2.2, respectively. In Section 4.2.3, we suggest an adaptive procedure for the estimation of the covariance. Then, we examine the problem of nonparametric density estimation in Section 4.3, using a private kernel density estimator. The convergence rate of the estimator is studied in Section 4.3.1, while in Section 4.3.2 we establish the minimax optimality of such rate. We conclude the density estimation section by proposing a data-driven procedure for bandwidth selection in Section 4.3.3. Finally, all proofs are collected in the appendix.

2 Problem formulation

We consider $\boldsymbol{X}_1$, ..., $\boldsymbol{X}_n$ iid data whose law is $\boldsymbol{X} = (X^1, ..., X^d) \in \boldsymbol{\mathcal{X}} = \prod_{j=1}^{d} \mathcal{X}^j$. It can represent the information coming from n different individuals, about d different aspect of their life. For each individual the information is privatized in a different way. Compared to the literature, where all the components relative to the same person are made public through the same channel,

we now consider the case where each component is made public separately, that is why we talk of "*componentwise local differential privacy*" (CLDP).
Let us formalize the framework discussed before. The act of privatizing the raw samples $(\boldsymbol{X}_i)_{i=1,...,n}$ and transforming them into the public set of samples $(\boldsymbol{Z}_i)_{i=1,...,n}$ is modeled by a conditional distribution, called privacy mechanism or channel distribution. We assume that each component of a disclosed observation, denoted by Z_i^j, is privatized separately and belongs to some space $\mathcal{Z}^j$, which may vary depending on the component j. This implies that the observation $\boldsymbol{Z}_i$ is an element of the product space $\boldsymbol{\mathcal{Z}} := \prod_{j=1}^d \mathcal{Z}^j$.
We also assume that the spaces $\mathcal{X}^j$ and $\mathcal{Z}^j$ are separable complete metric spaces, with their Borel sigma-fields defining measurable spaces $(\mathcal{X}^j, \Xi_{\mathcal{X}^j})$ and $(\mathcal{Z}^j, \Xi_{\mathcal{Z}^j})$, respectively, for all $j \in \{1, \dots, d\}$.

The privacy mechanism is allowed to be sequentially interactive, meaning that during the privatization of the j-th component of the i-th observation X_i^j, all previously privatized values $(\boldsymbol{Z}_m)_{m=1,...,i-1}$ are publicly available. This leads to the following conditional independence structure, for any $j \in \{1, ..., d\}$:

$$\{X_i^j, \boldsymbol{Z}_1, ..., \boldsymbol{Z}_{i-1}\} \to Z_i^j, \qquad Z_i^j \perp X_k^j \mid \{X_i^j, \boldsymbol{Z}_1, ..., \boldsymbol{Z}_{i-1}\} \text{ for } k \neq i.$$

More precisely, for $j = 1, ..., d$ and $i = 1, ..., n$, given $X_i^j = x_i^j \in \mathcal{X}^j$ and $\boldsymbol{Z}_m = \boldsymbol{z}_m \in \boldsymbol{\mathcal{Z}}$ for $m = 1, ..., i-1$; the i-th privatized output $Z_i^j \in \mathcal{Z}^j$ is drawn as

$$Z_i^j \sim Q_i^j(\cdot | X_i^j = x_i^j, \boldsymbol{Z}_1 = \boldsymbol{z}_1, ..., \boldsymbol{Z}_{i-1} = \boldsymbol{z}_{i-1}) \tag{2.1}$$

for Markov kernels $Q_i^j : \Xi_{\mathcal{Z}^j} \times (\mathcal{X}^j \times (\boldsymbol{\mathcal{Z}})^{i-1}) \to [0, 1]$. The notation $(\boldsymbol{\mathcal{Z}}, \Xi_{\boldsymbol{\mathcal{Z}}}) = (\prod_{j=1}^d \mathcal{Z}^j, \otimes_{j=1}^d \Xi_{\mathcal{Z}^j})$ refers to the measurable space of non-private data while $(\boldsymbol{\mathcal{X}}, \Xi_{\boldsymbol{\mathcal{X}}}) = (\prod_{j=1}^d \mathcal{X}^j, \otimes_{j=1}^d \Xi_{\mathcal{X}^j})$ is the space of private or raw data.
All of the examples presented in Section 4 have raw data that take values in $\boldsymbol{\mathcal{X}} = \mathbb{R}^d$. The space of privatized data, denoted by $\boldsymbol{\mathcal{Z}}$, can be quite general, as it is selected by the statistician based on a specific privatization mechanism. Nonetheless, in all of the practical examples of privatization that are discussed in Section 4, the privatized data will also be valued in $\boldsymbol{\mathcal{Z}} = \mathbb{R}^d$.

A specific example of the privacy mechanism described earlier is the non-interactive algorithm, where the value of Z_i^j is solely dependent on X_i^j. Therefore, Equation (2.1) no longer contains any correlation with the previously generated $\boldsymbol{Z}$ values. In this scenario, we eliminate any dependence of the Markov kernels on the observation i. However, when different components represent diverse encrypted information associated with the same individual, there is no justification for the distinct components to follow the same distribution. Therefore, it is necessary to consider that different components may have different laws. In the non-interactive case for any $j = 1, ..., d$ and for any $i = 1, ..., n$ the privatized output is given by

$$Z_i^j \sim Q^j(\cdot | X_i^j = x_i^j).$$

Although it is usually easier to consider non interactive algorithms, as they lead to iid privatized sample, in some situations it is useful for the channel's output to rely on previous computations. Stochastic approximation schemes, for instance, necessitate this kind of dependency (see [32]).

It is possible to quantify the privacy through the notion of local differential privacy. For a given parameter $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_d)$, $\alpha_j \geq 0$, for any $j \in \{1, ..., d\}$, the random variable Z_i^j is an α_j-differentially locally privatized view of X_i^j if for all $\boldsymbol{z}_1, ..., \boldsymbol{z}_{i-1} \in \boldsymbol{\mathcal{Z}}$ and $x, x' \in \mathcal{X}^j$ we have

$$\sup_{A \in \Xi_{\mathcal{Z}^j}} \frac{Q_i^j(A | X_i^j = x, \boldsymbol{Z}_1 = \boldsymbol{z}_1, ..., \boldsymbol{Z}_{i-1} = \boldsymbol{z}_{i-1})}{Q_i^j(A | X_i^j = x', \boldsymbol{Z}_1 = \boldsymbol{z}_1, ..., \boldsymbol{Z}_{i-1} = \boldsymbol{z}_{i-1})} \leq \exp(\alpha_j). \tag{2.2}$$

We say that the privacy mechanism $\boldsymbol{Q}_i = (Q_i^1, \dots, Q_i^d)$ for $i = 1, \dots, n$ is $\boldsymbol{\alpha}$-differentially locally private if each variable Z_i^j is an α_j- differentially locally private view of X_i^j. We denote by

$\mathcal{Q}_{\boldsymbol{\alpha}}^{(n)}$ the set of all local $\boldsymbol{\alpha}$-differential private Markov kernels $(Q_i^j)_{\substack{1\le i\le n\\ 1\le j\le d}}$.

The parameter α_j quantifies the amount of privacy that is guaranteed to the variable X_i^j: setting $\alpha_j = 0$ ensures perfect privacy for recovering X_i^j from the view of $Z_i^j, \boldsymbol{Z}_1, \dots, \boldsymbol{Z}_{i-1}$, whereas letting α_j tend to infinity softens the privacy restriction.

In the non-interactive case, $\boldsymbol{Q}_i = \boldsymbol{Q} = (Q^1, \dots, Q^d)$ does not depend on i and the bound (2.2) becomes

$$\sup_{A\in\Xi_{\mathcal{Z}^j}} \frac{Q^j(A|X_i^j = x)}{Q^j(A|X_i^j = x')} \le \exp(\alpha_j). \tag{2.3}$$

Under componentwise local differential privacy the kernels $Q^j(\cdot|X_i^j = x)$ are mutually absolutely continuous for different x. Hence, we can suppose that there exists a dominating measure μ^j on $(\mathcal{Z}^j, \Xi_{\mathcal{Z}^j})$ such that the kernel Q^j admits a density with respect to μ^j. We denote by q^j this density. Then, the property of $\boldsymbol{\alpha}$-CLDP defined in (2.3) is equivalent to the following. For all $x, x' \in \mathcal{X}^j$

$$\sup_{z\in\mathcal{Z}^j} \frac{q^j(z|X^j = x)}{q^j(z|X^j = x')} \le \exp(\alpha_j). \tag{2.4}$$

In this framework, we want to characterize the tradeoff between local differential privacy and statistical utility. In particular, we want to characterize how, for several canonical estimation problems, the optimal rate of convergence changes as a function of the privacy. For this reason, we develop some bounds on pairwise divergences which lead us to the derivation of minimax bounds under CLDP constraints.

2.1 Minimax framework

Before we keep proceeding, we introduce the minimax risk in the classical framework. It will be useful to present the notion of multivariate $\boldsymbol{\alpha}$-private minimax rate, which is defined starting from the observation of the privatized outputs Z_i^j, for $i \in \{1, ..., n\}$ and $j \in \{1, ..., d\}$.

Suppose we have a set of probability distributions $\mathcal{P}$ defined on a sample space $\mathcal{X}$, and let $\theta(P)$ be a function that maps each distribution in $\mathcal{P}$ to a value in a set of parameters Θ. The specific set Θ depends on the statistical model being used. For instance, if we are estimating the mean of a single variable, Θ will be a subset of the real numbers. On the other hand, if we are estimating a probability density function, Θ can be a subset of the space of all possible density functions over $\mathcal{X}$. Suppose moreover we have a function ρ that measures the distance between two points in the set of parameters Θ and which is a semi-metric (i.e. it does not necessarily satisfy the triangle inequality). We use this function to evaluate the performance of an estimator for the parameter θ. Additionally, we consider a non-decreasing function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\Phi(0) = 0$. The classical example consists in taking $\rho(x, y) = |x - y|$ and $\Phi(t) = t^2$.

In a scenario without privacy, a statistician has access to iid observations $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ that are drawn from a probability distribution $P \in \mathcal{P}$. The goal is to estimate an unknown parameter $\theta(P)$ that belongs to a set of parameters Θ. To achieve this goal, the statistician uses a measurable function $\hat{\theta} : \mathcal{X}^n \to \Theta$. The quality of the estimator $\hat{\theta}(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ is evaluated in terms of its minimax risk, defined as

$$\inf_{\hat{\theta}} \sup_{P\in\mathcal{P}} \mathbb{E}_P[\Phi(\rho(\hat{\theta}(\boldsymbol{X}_1, ..., \boldsymbol{X}_n), \theta(P)))], \tag{2.5}$$

where the inf is taken over all the possible estimators $\hat{\theta}$.

A vast body of statistical literature is dedicated to the development of methods for determining upper and lower bounds on the minimax risk for different types of estimation problems.

In this paper we want to consider the private analogous of the minimax risk described above, which takes into account the privacy constraints in the multivariate context, where the components are made public separately. Its definition is a straightforward consequence of the $\boldsymbol{\alpha}$-CLDP mechanism as in (2.2). Indeed, for any given privacy level $\alpha_j > 0$ we have $\mathcal{Q}_{\boldsymbol{\alpha}}$ denoting the set of all the privacy mechanisms having the $\boldsymbol{\alpha}$-CLDP property. Then, for any sample $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$, any distribution $\boldsymbol{Q}^n := (\boldsymbol{Q}_1, ..., \boldsymbol{Q}_n) \in \mathcal{Q}_{\boldsymbol{\alpha}}^{(n)}$ produces a set of privatized observations which have been made public separately, i.e. $Z_1^1, ..., Z_1^d, ..., Z_n^1, ..., Z_n^d$. We can focus on estimators $\hat{\theta}$ which depend exclusively on the privatized sample and we can therefore write $\hat{\theta} = \hat{\theta}(Z_1^1, ..., Z_1^d, ..., Z_n^1, ..., Z_n^d)$. Then, it seems natural to look for the privacy mechanism $\boldsymbol{Q}^n \in \mathcal{Q}_{\boldsymbol{\alpha}}^{(n)}$ for which the estimator $\hat{\theta}(Z_1^1, ..., Z_1^d, ..., Z_n^1, ..., Z_n^d)$ performs as good as possible. The performance of the estimator is even in this case judged in term of the minimax risk, which leads us to the following definition.

Definition 1. *Given a privacy parameter $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_d)$, $\alpha_j > 0$ and a family of distributions $\theta(P)$, the componentwise $\boldsymbol{\alpha}$ private minimax risk in the metric ρ is*

$$\inf_{\boldsymbol{Q}^n \in \mathcal{Q}_{\boldsymbol{\alpha}}^{(n)}} \inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,\boldsymbol{Q}}[\Phi(\rho(\hat{\theta}(Z_1^1, ..., Z_n^d), \theta(P)))],$$

where the inf is taken over all the estimators $\hat{\theta}$ and all the choices $(\boldsymbol{Q}_1, ..., \boldsymbol{Q}_n) \in \mathcal{Q}_{\boldsymbol{\alpha}}^{(n)}$ such that the data $Z_1^1, ..., Z_n^d$ are $\boldsymbol{\alpha}$-CLDP views of $X_1^1, ..., X_n^d$ in the sense of (2.2).

Our main goal consists in proving some sharp bounds on pairwise divergences, as in Section 3.1. From there, it will be possible to derive sharp lower bounds on the $\boldsymbol{\alpha}$ private minimax risk for the statistical estimation of manifolds canonical problems, see Section 4 for some examples of applications.

3 Main results

In this section, we establish a connection between the proximity of two laws for the private individual variable $\boldsymbol{X}$ and the proximity of their corresponding public views under the $\boldsymbol{\alpha}$-CLDP property. Then, we explore the usefulness of this result for the privatization of independent samplings.

3.1 Bounds on pairwise divergences

We assume that we are given a pair of distributions P and $\tilde{P}$ defined on a common space $\boldsymbol{\mathcal{X}} = (\mathcal{X}^1, \dots, \mathcal{X}^d)$, and a privatization kernel $\boldsymbol{Q} = (Q^1, \dots, Q^d)$ where Q^j is the privatization channel from $\mathcal{X}^j$ to $\mathcal{Z}^j$. We denote by M and $\tilde{M}$ the law of the images of P and $\tilde{P}$ through the operation of privatization. It means that we consider a couple of raw samples $\boldsymbol{X}$, $\tilde{\boldsymbol{X}}$ with distribution P, $\tilde{P}$, and that the associated privatized samples $\boldsymbol{Z}$, $\tilde{\boldsymbol{Z}}$ have distribution denoted by M and $\tilde{M}$. Consistently with the description in Section 2, each channel Q^j acts on its associated component X^j independently of the other channels. More formally, we can write the correspondence between P and M as

$$M\left(\prod_{j=1}^{d} A_j\right) = \int_{\boldsymbol{\mathcal{X}}} \prod_{j=1}^{d} Q^j(A_j \mid X^j = x^j) P(dx^1, \dots, dx^d),$$

for any $A_j \in \Xi_{\mathcal{Z}^j}$.
Before we keep proceeding, let us introduce some notation. We denote as $d_{TV}(P_1, P_2)$ the total variation distance between the two measures P_1 and P_2:

$$d_{TV}(P_1, P_2) = \int |dP_1 - dP_2| = \int |\frac{dP_1}{dP_1 + dP_2}(x) - \frac{dP_2}{dP_1 + dP_2}(x)|(P_1 + P_2)(dx).$$

Moreover, we denote as $d_{KL}(P_1, P_2)$ the Kullback divergence between the two measures P_1 and P_2, $d_{KL}(P_1, P_2) = \int \log\big(\frac{dP_2}{dP_1}\big) dP_2$ for P_2 absolutely continuous to P_1.
Finally, we denote as $L_{(X^{j_1},...,X^{j_k})}$ the law of the marginals $X^{j_1}, ..., X^{j_k}$ of $\boldsymbol{X}$, where k and the indexes j_1, ... , j_k belong to $\{1, ..., d\}$. According to this notation it is clearly $L_{(X^1,...,X^d)} = P$ and $L_{(\tilde{X}^1,...,\tilde{X}^d)} = \tilde{P}$.

Our main result gives an intuition on how close the two output distributions shall be, depending on how close the laws of the anonymised data were. Its proof can be found at the end of this section.

Theorem 3.1. *Let $\alpha_j \geq 0$ and assume that $\boldsymbol{Q} = (Q^1, \dots, Q^d)$ guarantees the $\boldsymbol{\alpha}$-CLDP constraint as defined by the condition* (2.3)*. Then,*

$$d_{KL}(M, \tilde{M}) \leq \Big(\sum_{k=1}^{d} \sum_{(j_1,...,j_k)} \prod_{i=1}^{k} (e^{\alpha_{j_i}} - 1) d_{TV}(L_{(X^{j_1},...,X^{j_k})}, L_{(\tilde{X}^{j_1},...,\tilde{X}^{j_k})})\Big)^2,$$

where the inner summation is on $1 \leq j_1 < j_2 \dots < j_k \leq d$ any k distinct ordered indexes in $\{1, \dots, d\}$.
In the case where $\alpha_1 = ... = \alpha_d =: \alpha$, it reduces to

$$d_{KL}(M, \tilde{M}) \leq \Big(\sum_{k=1}^{d} (e^{\alpha} - 1)^k \sum_{(j_1,...,j_k)} d_{TV}(L_{(X^{j_1},...,X^{j_k})}, L_{(\tilde{X}^{j_1},...,\tilde{X}^{j_k})})\Big)^2. \tag{3.1}$$

Remark 3.2. *To better understand the formula in the statement of the theorem the reader may observe that, for $d = 2$, the left hand side of our main bound is*

$$\Big[(e^{\alpha_1} - 1) d_{TV}(L_{X^1}, L_{\tilde{X}^1}) + (e^{\alpha_2} - 1) d_{TV}(L_{X^2}, L_{\tilde{X}^2}) + (e^{\alpha_1} - 1)(e^{\alpha_2} - 1) d_{TV}(L_{(X^1,X^2)}, L_{(\tilde{X}^1,\tilde{X}^2)})\Big]^2.$$

For $d = 3$ it is instead

$$\Big[\sum_{i=1}^{3} (e^{\alpha_i} - 1) d_{TV}(L_{X^i}, L_{\tilde{X}^i}) + \sum_{1 \leq i < j \leq 3} (e^{\alpha_i} - 1)(e^{\alpha_j} - 1) d_{TV}(L_{(X^i,X^j)}, L_{(\tilde{X}^i,\tilde{X}^j)}) + (e^{\alpha_1} - 1)(e^{\alpha_2} - 1)(e^{\alpha_3} - 1) d_{TV}(L_{(X^1,X^2,X^3)}, L_{(\tilde{X}^1,\tilde{X}^2,\tilde{X}^3)})\Big]^2.$$

Remark 3.3. *In the mono-dimensional case, where $\boldsymbol{\mathcal{X}} = \mathcal{X}^1$ and $\boldsymbol{\alpha} = \alpha_1$, we recover a bound similar to the one in Theorem 1 of [17], which is*

$$d_{KL}(M, \tilde{M}) \leq \min(4, e^{2\alpha})(e^{\alpha} - 1)^2 d_{TV}(P, \tilde{P})^2. \tag{3.2}$$

In the multidimensional setting with $\alpha_1 = \dots = \alpha_d = \alpha$, if we use in (3.1) *the crude bound*

$$d_{TV}(L_{(X^{j_1},...,X^{j_k})}, L_{(\tilde{X}^{j_1},...,\tilde{X}^{j_k})}) \leq d_{TV}(P, \tilde{P}),$$

we obtain

$$d_{KL}(M, \tilde{M}) \leq \left(\sum_{k=1}^{d} (e^{\alpha} - 1)^k \binom{n}{k} d_{TV}(P, \tilde{P})\right)^2 = (e^{\alpha d} - 1)^2 d_{TV}(P, \tilde{P})^2, \tag{3.3}$$

where we have used Newton's binomial formula. It is important to note that when $d > 1$, the inequality stated in Theorem 1 of [17] (referred to as (3.2)*) is still valid. This bound is comparable to the one stated in* (3.3) *when α is small. However, our result is generally more precise. Indeed, our analysis takes into account the fact that the individual components of the vector have been made publicly available, which allows us to recover a factor of $(e^{\alpha\, d} - 1)^2$ instead of $(e^{\alpha} - 1)^2$ when comparing the contribution of the law of the whole vector to the contribution of its one-dimensional marginals.*

As we will see in next section, the bound on pairwise divergences gathered in Theorem 3.1 is particularly helpful when one wants to show lower bound on the minimax risk, in order to illustrate the optimality of a proposed estimator, in a minimax sense.
In this case one can propose two priors whose marginal laws are all the same but for the last term, where the whole vector is considered. Then, our main result reduces to the bound below.

Corollary 3.4. *Let us consider a couple of raw samples $\boldsymbol{X}$, $\tilde{\boldsymbol{X}}$ with distributions P, $\tilde{P}$ and the associated couple of privatized samples $\boldsymbol{Z}$, $\tilde{\boldsymbol{Z}}$ with distributions M, $\tilde{M}$. Assume moreover that, for any $k \in \{1, ..., d-1\}$, $1 \le j_1 < \cdots < j_k \le d$, it is $L_{(X^{j_1},...,X^{j_k})} = L_{(\tilde{X}^{j_1},...,\tilde{X}^{j_k})}$. Then,*

$$d_{KL}(M, \tilde{M}) \le \Big(\prod_{i=1}^{d} (e^{\alpha_i} - 1)^2 \Big) d_{TV}(P, \tilde{P})^2.$$

As we will see in next section, the bound stated in the corollary is extremely useful to assess minimax risks under the α-CLDP property.
The rest of this section is devoted to the proof of Theorem 3.1.

Proof of Theorem 3.1. To prove our main theorem we introduce some notation. We first recall that $q^j(z^j \mid x^j)$ is the density of the the law of Z^j conditional to $X^j = x^j$ with respect to a dominating measure $\mu^j(dz^j)$. We denote by $q(z^1, \ldots, z^d)$ the density of the law of $(Z^1, \ldots, Z^d)$, which exists with respect to the reference measure $\boldsymbol{\mu}(d\boldsymbol{z}) := \mu^1(dz^1) \times \cdots \times \mu^d(dz^d)$. In a more general way, we examine a collection of d symbols $\zeta^1, \ldots, \zeta^d$, which can take one of three possible values: $\zeta^j = dx^j$, $\zeta^j = z^j$, or $\zeta^j = \emptyset$. We define a vector $\boldsymbol{W}$ such that the j-th component of $\boldsymbol{W}$, denoted W^j, takes the value of X^j if $\zeta^j = dx^j$, takes the value of Z^j if $\zeta^j = z^j$, and is removed entirely if $\zeta^j = \emptyset$. We denote as $q(\zeta^1, ..., \zeta^d)$ the Markovian kernel such that

$$q(\zeta^1, ..., \zeta^d) \times \prod_{j:\zeta^j = z^j} \mu^j(dz^j)$$

is the law of $\boldsymbol{W}$. For example,

$$q(dx^1, \ldots, dx^i, \emptyset, \ldots, \emptyset, z^{i+j+1}, ..., z^d)\mu^{i+j+1}(dz^{i+j+1}) \times \cdots \times \mu^d(dz^d)$$

is the law of $(X^1, \ldots, X^i, Z^{i+j+1}, \ldots, Z^d)$ and we thus have

$$\mathbb{E}\left[f(X^1, \ldots, X^i, Z^{i+j+1}, \ldots, Z^d)\right] = \int_{\mathcal{X}^1 \times \cdots \times \mathcal{X}^i \times \mathcal{Z}^{i+j+1} \times \cdots \times \mathcal{Z}^d} f(x^1, \ldots, x^i, z^{i+j+1}, \ldots, z^d)$$
$$q(dx^1, \ldots, dx^i, \emptyset, \ldots, \emptyset, z^{i+j+1}, ..., z^d)\mu^{i+j+1}(dz^{i+j+1}) \times \cdots \times \mu^d(dz^d),$$

for any positive measurable function f. Such Markovian kernels $q(\zeta^1, ..., \zeta^d)$ exist for all choices of symbols ζ^j. Indeed, it is possible to disintegrate the law of $(W^j)_{j:\zeta_j \neq \emptyset}$ with respect to the law of $(W^j)_{j:\zeta^j = z^j}$ and use that the law of $(W^j)_{j:\zeta^j = z^j}$ admits a density with respect to $\prod_{j:\zeta^j = z^j} \mu^j(dz^j)$. With a slight abuse of notation we consider $q(\emptyset, \ldots, \emptyset) = 1$. It is consistent with the fact that, when removing one marginal $W^j = X^j$ (or $W^j = Z^j$) from a random vector, the corresponding probability measure is integrated with respect to the variable x_j (or z_j). Hence, when removing ultimately all the variables the probability integrates to 1, yielding to the notation $q(\emptyset, \ldots, \emptyset) = 1$. Let us stress that these notations are cumbersome as we are dealing with general variables $\boldsymbol{X}$ and $\boldsymbol{Z}$. For instance, in the simple case where $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{Z}} = \mathbb{R}^d$, $\boldsymbol{X}$ with density on $\mathbb{R}^d$, and privacy channels having densities $q^j(z^j \mid x^j)$ with respect to the Lebesgue measure, we would have simply defined $q(\zeta^1, \ldots, \zeta^d)$ as the density of the variables $(W^j)_{j:\zeta^j \neq \emptyset}$. We introduce analogously $\tilde{q}(\zeta_1, ..., \zeta_d)$, which corresponds to the law of $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{Z}}$ in the same way as above.

Then, using these notations and the fact that the law of Z^1 conditional to $(X^1, Z^2, \ldots, Z^d)$ is given, from the definition of the privacy mechanism, by $Q^1(dz^1 \mid X^1 = x^1) = q^1(z^1 \mid x^1)\mu^1(dz^1)$, it is

$$q(z^1, ..., z^d) = \int_{\mathcal{X}^1} q^1(z^1|x^1) q(dx^1, z^2, ..., z^d). \tag{3.4}$$

Finally, we introduce the function $l[\zeta^1, \ldots, \zeta^d] : \mathcal{Z}^1 \times \cdots \times \mathcal{Z}^d \to \mathbb{R}_+$ as below:

$$l[\zeta^1, \ldots, \zeta^d] := |q(z^1, \ldots, z^d) - \tilde{q}(z^1, \ldots, z^d)|, \quad \text{for } \boldsymbol{\zeta} = \boldsymbol{z}, \tag{3.5}$$

$$l[\zeta^1, \ldots, \zeta^d] := \prod_{j:\zeta^j \neq z^j} q^j(z^j|x^j_*) \prod_{j:\zeta^j = dx^j} |e^{\alpha_j} - 1| \times \int_{\prod_{j:\zeta^j = dx^j} \mathcal{X}^j} |q(\zeta^1, ..., \zeta^d) - \tilde{q}(\zeta^1, ..., \zeta^d)|, \quad \text{for } \boldsymbol{\zeta} \neq \boldsymbol{z}, \tag{3.6}$$

where $q^j(z^j|x^j_*) := \inf_{x^j} q^j(z^j|x^j)$ and $\boldsymbol{\zeta} = (\zeta^1, \ldots, \zeta^d)$. To clarify the notation, let us stress that the integration variables in (3.6) are the ζ^j such that $\zeta^j = dx^j$. Moreover, $l[\zeta^1, \ldots, \zeta^d]$ is a function of $(z^1, \ldots, z^d)$ whatever is the choice $(\zeta^1, \ldots, \zeta^d)$. Indeed, the variable z^j appears either in the product $\prod_{j:\zeta^j \neq z^j} q^j(z^j|x^j_*)$ when j is such that $\zeta^j \neq z^j$, or in the integral when $\zeta^j = z^j$. To give an example, the quantity $l[dx^1, \ldots, dx^i, \emptyset, \ldots, \emptyset, z^{i+j+1}, \ldots, z^d]$ is equal to

$$\prod_{l=1}^{i+j} q^l(z^l|x^l_*) \prod_{l=1}^{i} |e^{\alpha_l} - 1| \times \int_{\prod_{l=1}^{i} \mathcal{X}^l} |q(dx^1, \ldots, dx^i, \emptyset, \ldots, \emptyset, z^{i+j+1}, \ldots, z^d) - \tilde{q}(dx^1, \ldots, dx^i, \emptyset, \ldots, \emptyset, z^{i+j+1}, \ldots, z^d)|,$$

which is clearly a function of $(z^1, \ldots, z^d)$. In the scenario where $\boldsymbol{\zeta} = \boldsymbol{z}$, the equation (3.6) aligns with (3.5), except that the integration variables disappear. Then, in the right-hand side of (3.6) the integral no longer appears. We also specify that $l[\emptyset, \ldots, \emptyset] = 0$, which results from the fact that, abusing the notations, we have $q(\emptyset, \ldots, \emptyset) = \tilde{q}(\emptyset, \ldots, \emptyset) = 1$.

Our main result heavily relies on the following proposition. Its proof, based on a recurrence argument, can be found in the appendix.

Proposition 3.5. *Let the function $l[\zeta^1, \ldots, \zeta^d]$ be defined according to* (3.5) *and* (3.6)*. Then, under the hypothesis of Theorem 3.1, the following inequality holds true.*

$$l[z^1, ..., z^d] \le \sum_{(\zeta^1, ..., \zeta^d) \in \{\emptyset, dx^j\}^d} l[\zeta^1, ..., \zeta^d]. \tag{3.7}$$

Recalling (3.5), we remark that the left hand side of (3.7) assesses the difference between the densities of $\boldsymbol{Z}$ and $\tilde{\boldsymbol{Z}}$, while the right hand side only relies on the laws of $\boldsymbol{X}$, $\tilde{\boldsymbol{X}}$ as the symbol $\zeta^j = z^j$ disappears in the sum.
From Proposition 3.5 we have

$$|q(\boldsymbol{z}) - \tilde{q}(\boldsymbol{z})| = l[z^1, ..., z^d] \le \sum_{\boldsymbol{\zeta}:\zeta^j \in \{\emptyset, dx^j\}} l(\boldsymbol{\zeta}) \tag{3.8}$$

$$= \sum_{\boldsymbol{\zeta}:\zeta^j \in \{\emptyset, dx^j\}} \prod_{j=1}^{d} q^j(z^j|x^j_*) \prod_{j:\zeta^j = dx^j} (e^{\alpha_j} - 1) \int_{\prod_{j:\zeta^j = dx^j} \mathcal{X}^j} |q(\boldsymbol{\zeta}) - \tilde{q}(\boldsymbol{\zeta})|. \tag{3.9}$$

By the definition of total variation distance, the quantity above can be seen as

$$\prod_{j=1}^{d} q^j(z^j|x_*^j) \sum_{\boldsymbol{\zeta}:\zeta^j\in\{\emptyset,x^j\}} \Big\{ \big(\prod_{j:\zeta^j=dx^j} (e^{\alpha_j}-1)\big) d_{TV}(L_{(X^j)_{j:\zeta^j=dx^j}}, L_{(\tilde{X}^j)_{j:\zeta^j=dx^j}}) \Big\}$$
$$= \prod_{j=1}^{d} q^j(z^j|x_*^j) \sum_{k=1}^{d} \sum_{\substack{j_1<\ldots<j_k \\ k \text{ different indexes} \\ \text{in } \{1\ldots,d\}}} \Big\{ \big(\prod_{i=1}^{k}(e^{\alpha_{j_i}}-1)\big)\ d_{TV}(L_{(X^{j_1},\ldots,X^{j_k})}, L_{(\tilde{X}^{j_1},\ldots,\tilde{X}^{j_k})}) \Big\}. \tag{3.10}$$

To conclude the proof we observe it is

$$d_{KL}(M,\tilde{M}) = \int_{\mathcal{Z}^1\times\cdots\times\mathcal{Z}^d} q(\boldsymbol{z})\log(\frac{q(\boldsymbol{z})}{\tilde{q}(\boldsymbol{z})})d\boldsymbol{\mu}(\boldsymbol{z}) + \int_{\mathcal{Z}^1\times\cdots\times\mathcal{Z}^d} \tilde{q}(\boldsymbol{z})\log(\frac{\tilde{q}(\boldsymbol{z})}{q(\boldsymbol{z})})d\boldsymbol{\mu}(\boldsymbol{z})$$
$$= \int_{\mathcal{Z}^1\times\cdots\times\mathcal{Z}^d} (q(\boldsymbol{z})-\tilde{q}(\boldsymbol{z}))\log(\frac{q(\boldsymbol{z})}{\tilde{q}(\boldsymbol{z})})d\boldsymbol{\mu}(\boldsymbol{z}). \tag{3.11}$$

Then, Lemma 4 in [17] entails $|\log\frac{q(\boldsymbol{z})}{\tilde{q}(\boldsymbol{z})}| \le \frac{|q(\boldsymbol{z})-\tilde{q}(\boldsymbol{z})|}{\min(q(\boldsymbol{z}),\tilde{q}(\boldsymbol{z}))}$. In order to study the denominator, we write

$$q(\boldsymbol{z}) = \int_{x^1\in\mathcal{X}^1} q^1(z^1|x^1)q(dx^1,z^2,...,z^d) \ge q^1(z^1|x_*^1)q(\emptyset,z^2,...,z^d).$$

We iterate the arguing above, to recover

$$q(\boldsymbol{z}) \ge \prod_{j=1}^{d-1} q^j(z^j|x_*^j) \int_{x^d\in\mathcal{X}^d} q^d(z^d|x^d)q(\emptyset,\ldots,\emptyset,dx^d) \ge \prod_{j=1}^{d} q^j(z^j|x_*^j),$$

where we used $\int_{x^d\in\mathcal{X}^d} q(\emptyset,\ldots,\emptyset,dx^d)=1$. An analogous lower bound holds true for $\tilde{q}(\boldsymbol{z})$. Thus, using also the bound in (3.8)–(3.10), we obtain

$$|\log(\frac{q(\boldsymbol{z})}{\tilde{q}(\boldsymbol{z})})| \le \frac{|q(\boldsymbol{z})-\tilde{q}(\boldsymbol{z})|}{\min(q(\boldsymbol{z}),\tilde{q}(\boldsymbol{z}))}$$
$$\le \sum_{k=1}^{d}\sum_{(j_1,\ldots,j_k)} \big(\prod_{i=1}^{k}(e^{\alpha_{j_i}}-1)\big) d_{TV}(L_{(X^{j_1},\ldots,X^{j_k})}, L_{(\tilde{X}^{j_1},\ldots,\tilde{X}^{j_k})}). \tag{3.12}$$

We replace it in (3.11) which, together with (3.8)–(3.10), implies

$$d_{KL}(M,\tilde{M}) \le \Big(\sum_{k=1}^{d}\sum_{(j_1,\ldots,j_k)}\prod_{i=1}^{k}(e^{\alpha_{j_i}}-1)d_{TV}(L_{(X^{j_1},\ldots,X^{j_k})}, L_{(\tilde{X}^{j_1},\ldots,\tilde{X}^{j_k})})\Big)^2$$
$$\times \int_{\mathcal{Z}^1\times\cdots\times\mathcal{Z}^d}\prod_{j=1}^{d} q^j(z^j|x_*^j)d\boldsymbol{\mu}(\boldsymbol{z}).$$

The proof of Theorem 3.1 is then complete, as the last integral can not be larger than one. □

Remark 3.6. *Let us stress that our proof also provides a control on the difference between the densities of* $\boldsymbol{Z}$ *and* $\tilde{\boldsymbol{Z}}$ *given by* (3.12).

3.2 Application to privatization of independent sampling

This section applies the previously proven results to a scenario where the original samples $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ composed by independent vectors distributed according to $\boldsymbol{X}$, are transformed into privatized samples $\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n$. The notation used in this section is consistent with that used in Section 2 and will be used throughout the rest of the paper. Specifically, X_i^j refers to the j-th component of the i-th individual $\boldsymbol{X}_i$.

Assume we sample a random vector $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ with product measure of the form $P^n(d\boldsymbol{x_1}, ..., d\boldsymbol{x_n}) := \prod_{i=1}^n P_i(d\boldsymbol{x_i})$. We draw then an $\boldsymbol{\alpha}$ componentwise local differential private view of the sample $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ through the privacy mechanism $\boldsymbol{Q}^n = (\boldsymbol{Q}_1, ..., \boldsymbol{Q}_n)$, where $\boldsymbol{Q}_i = (Q_i^1, ..., Q_i^d)$. The privatized samples $(\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n)$ is distributed according to some measure M^n. As we consider also the case where the algorithm is interactive, in general the measure M^n is not in a product form (with respect to i). However, the proposition on tensorization inequality that follows will yield a result similar to that provided by independence. It will prove especially useful for applications.

Proposition 3.7. *Let $\alpha_j \geq 0$ and assume that $\boldsymbol{Q}^n$ guarantees the $\boldsymbol{\alpha}$- CLDP constraint as defined by the condition* (2.2)*. Then, for any paired sequences of independent vectors $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ and $(\tilde{\boldsymbol{X}}_1, ..., \tilde{\boldsymbol{X}}_n)$ of distributions $P^n = \prod_{i=1}^n P_i$ and $\tilde{P}^n = \prod_{i=1}^n \tilde{P}_i$ respectively, we have*

$$d_{KL}(M^n, \tilde{M}^n) \leq \sum_{h=1}^n \Big(\sum_{k=1}^d \sum_{(j_1,...,j_k)} \prod_{i=1}^k (e^{\alpha_{j_i}} - 1) d_{TV}(L_{(X_h^{j_1},...,X_h^{j_k})}, L_{(\tilde{X}_h^{j_1},...,\tilde{X}_h^{j_k})}) \Big)^2, \tag{3.13}$$

where the inner summation is on $j_1 < \cdots < j_k$ any k distinct indexes in $\{1, \ldots, d\}$.

Assume now that the samples $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ and $(\tilde{\boldsymbol{X}}_1, ..., \tilde{\boldsymbol{X}}_n)$, in addition to being independent, are identically distributed and thus with laws $P^n = P^{\otimes n}$ and $\tilde{P}^n = \tilde{P}^{\otimes n}$. Moreover, we suppose that all the marginal laws of $\boldsymbol{X_i}$ and $\tilde{\boldsymbol{X}}_{\boldsymbol{i}}$ are equals, but when the whole vector is considered. Then, in analogy to Corollary 3.4, the proposition above leads to the following corollary.

Corollary 3.8. *Let $\alpha_j \geq 0$ and assume that $\boldsymbol{Q}^n$ guarantees the $\boldsymbol{\alpha}$- CLDP constraint as defined by the condition* (2.2)*. Then, for any paired sequences of iid vectors $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ and $(\tilde{\boldsymbol{X}}_1, ..., \tilde{\boldsymbol{X}}_n)$ of distributions $\tilde{P}^n = \tilde{P}^{\otimes n}$ and $\tilde{P}^n = \tilde{P}^{\otimes n}$ which are such that for any $k \in \{1, ..., d-1\}$ $L_{(X^{j_1},...,X^{j_k})} = L_{(\tilde{X}^{j_1},...,\tilde{X}^{j_k})}$, we have*

$$\begin{aligned} d_{KL}(M^n, \tilde{M}^n) &\leq n \Big(\prod_{j=1}^d (e^{\alpha_j} - 1) \Big)^2 d_{TV}^2(L_{(X^1,...,X^d)}, L_{(\tilde{X}^1,...,\tilde{X}^d)}) \\ &= n \Big(\prod_{j=1}^d (e^{\alpha_j} - 1) \Big)^2 d_{TV}^2(P, \tilde{P}). \end{aligned}$$

The proof of Proposition 3.7 follows next, while Corollary 3.8 is a direct consequence of the aforementioned proposition, remarking that in the two inner sums of (3.13) the only non-zero term is for $k = d$, $(j_1, \ldots, j_k) = (1, \ldots, d)$.

Proof of Proposition 3.7. We can introduce the marginal distribution of $\boldsymbol{Z}_h$ conditioned on $\boldsymbol{Z}_1 = \boldsymbol{z}_1$, ... , $\boldsymbol{Z}_{h-1} = \boldsymbol{z}_{h-1}$. We denote it as

$$M_h(\cdot | \boldsymbol{Z}_1 = \boldsymbol{z}_1, ..., \boldsymbol{Z}_{h-1} = \boldsymbol{z}_{h-1}) =: M_h(\cdot | \boldsymbol{z}_{1:h-1}).$$

Observe that for any $A_j \in \Xi_{\mathcal{Z}^j}$ and $x \in \mathbb{R}$, $\boldsymbol{z}_1, ..., \boldsymbol{z}_{h-1} \in \mathbb{R}^d$ it is

$$\begin{aligned} M_h(\prod_{j=1}^d A_j | \boldsymbol{z}_{1:h-1}) &= \int_{\boldsymbol{\mathcal{X}}} \prod_{j=1}^d Q_h^j(A_j | X_h^j = x, \boldsymbol{Z}_1 = \boldsymbol{z}_1, ..., \boldsymbol{Z}_{h-1} = \boldsymbol{z}_{h-1}) P_h(dx^1, ..., dx^d) \\ &=: \int_{\boldsymbol{\mathcal{X}}} \prod_{j=1}^d Q_h^j(A_j | X_h^j, \boldsymbol{Z}_{1:h-1}) P_h(dx^1, ..., dx^d). \end{aligned}$$

Moreover, we introduce the notation $d_{KL}(M_h, \tilde{M}_h)$ for the integrated Kullback divergence of the conditional distributions on the $\boldsymbol{Z}_h$, which is

$$\int_{\boldsymbol{\mathcal{Z}}^{h-1}} d_{KL}\Big(M_h(\cdot | \boldsymbol{z}_{1:h-1}), \tilde{M}_h(\cdot | \boldsymbol{z}_{1:h-1})\Big) dM^{h-1}(\boldsymbol{z}_1, ..., \boldsymbol{z}_{h-1}).$$

Then, the chain rule for Kullback-Lieber divergences as gathered in Chapter 5.3 of [24] provides

$$d_{KL}(M^n, \tilde{M}^n) = \sum_{h=1}^n d_{KL}(M_h, \tilde{M}_h).$$

By the definition of $\boldsymbol{\alpha}$-CLDP for sequentially interactive privacy mechanism provided in (2.2), the distribution $Q_h^j(A_j | X_h^j, \boldsymbol{Z}_{1:h-1})$ is α_j-differentially private for X_h^j. We can therefore apply Theorem 3.1 on $d_{KL}\Big(M_h(\cdot | \boldsymbol{z}_{1:h-1}), \tilde{M}_h(\cdot | \boldsymbol{z}_{1:h-1})\Big)$ which entails, together with the chain rule above,

$$\begin{aligned} d_{KL}(M^n, \tilde{M}^n) &= \sum_{h=1}^n \int_{\boldsymbol{\mathcal{Z}}^{h-1}} d_{KL}\Big(M_h(\cdot | \boldsymbol{z}_{1:h-1}), \tilde{M}_h(\cdot | \boldsymbol{z}_{1:h-1})\Big) dM^{h-1}(\boldsymbol{z}_1, ..., \boldsymbol{z}_{h-1}) \\ &= \sum_{h=1}^n \int_{\boldsymbol{\mathcal{Z}}^{h-1}} \Big(\sum_{k=1}^d \sum_{(j_1, ..., j_k)} \prod_{i=1}^k (e^{\alpha_{j_i}} - 1) \\ &\times d_{TV}(L_{(X_h^{j_1}, ..., X_h^{j_k} | \boldsymbol{z}_{1:h-1})}, L_{(\tilde{X}_h^{j_1}, ..., \tilde{X}_h^{j_k} | \boldsymbol{z}_{1:h-1})}) \Big)^2 dM^{h-1}(\boldsymbol{z}_1, ..., \boldsymbol{z}_{h-1}), \end{aligned}$$

where we have denoted as $L_{(X_h^{j_1}, ..., X_h^{j_k} | \boldsymbol{z}_{1:h-1})}$ the conditional distribution of $X_h^{j_1}, ..., X_h^{j_k}$ given the first $h-1$ values $\boldsymbol{Z}_1, ..., \boldsymbol{Z}_{h-1}$. Clearly in an analogous way $L_{(\tilde{X}_h^{j_1}, ..., \tilde{X}_h^{j_k} | \boldsymbol{z}_{1:h-1})}$ is the conditional distribution of $\tilde{X}_h^{j_1}, ..., \tilde{X}_h^{j_k}$ given the first $h-1$ values $\boldsymbol{Z}_1, ..., \boldsymbol{Z}_{h-1}$. However, by construction, the random variables $\boldsymbol{X}_h$ are conditionally independent, which implies that $L_{(X_h^{j_1}, ..., X_h^{j_k} | \boldsymbol{z}_{1:h-1})} = L_{(X_h^{j_1}, ..., X_h^{j_k})}$ and $L_{(\tilde{X}_h^{j_1}, ..., \tilde{X}_h^{j_k} | \boldsymbol{z}_{1:h-1})} = L_{(\tilde{X}_h^{j_1}, ..., \tilde{X}_h^{j_k})}$. It yields

$$\begin{aligned} d_{KL}(M^n, \tilde{M}^n) \leq \sum_{h=1}^n & \Big(\sum_{k=1}^d \sum_{(j_1, ..., j_k)} \prod_{i=1}^k (e^{\alpha_{j_i}} - 1) d_{TV}(L_{(X_h^{j_1}, ..., X_h^{j_k})}, L_{(\tilde{X}_h^{j_1}, ..., \tilde{X}_h^{j_k})}) \Big)^2 \\ & \times \int_{\boldsymbol{\mathcal{Z}}^{h-1}} dM^{h-1}(\boldsymbol{z}_1, ..., \boldsymbol{z}_{h-1}). \end{aligned}$$

The proof is then concluded once we remark that the integral in $dM^{h-1}(\boldsymbol{z}_1, ..., \boldsymbol{z}_{h-1})$ is equal to 1. □

4 Applications to statistical inference

In this upcoming section, one objective is to demonstrate the usefulness of the bounds on divergences between distributions that have been accumulated in Section 3. We will show versatile applications of these bounds in different statistical problems. First, we study how information about a private characteristic of an individual can be revealed by the public views of other characteristics of the same individual. For this problem, the results obtained in Section 3 are insightful tools that lead us to introduce the quantity (4.2) as the main parameter for measuring information leakage. In the following, we will also provide details on the estimation of covariance and density in a locally private and multivariate context. For these statistical problems, we construct explicit estimators, and the results of Section 3 can be used to derive the rate optimality of these estimators. Finally, we propose adaptive versions of our estimators.

4.1 Effective privacy level

When the data $X^1, \dots, X^d$ are disclosed by independent channels and with different privacy levels $\alpha_1, \dots, \alpha_d$, a natural question is how precisely the value of one marginal, say X^1, could be revealed by the observations of $Z^1, \dots, Z^d$, which are publicly available.

The case where some variable $X^1 \in \mathcal{X}^1$ is privatized using a Markov kernel into the public data $Z^1 \in \mathcal{Z}^1$ is the situation studied in [38] and [17]. It is known from [17] that if the privacy channel is α-LDP, then for all $x^1, x^{1\prime} \in \mathcal{X}^1$ and $\psi : \mathcal{Z}^1 \to \{x^1, x^{1\prime}\}$ we have

$$\frac{1}{2}\mathbb{P}\left(\psi(Z^1) \neq x^1 \mid X = x^1\right) + \frac{1}{2}\mathbb{P}\left(\psi(Z^1) \neq x^{1\prime} \mid X = x^{1\prime}\right) \geq \frac{1}{1+e^{\alpha}}, \tag{4.1}$$

This means that even if someone accesses two values, x^1 and $x^{1\prime}$, from the raw data set, it will be impossible for them to determine with a high level of certainty which of the values corresponds to a specific observation, denoted as Z^1. Any attempt to make a decision in this regard would result in an error, albeit with minimal probability.

If a vector $\boldsymbol{X}$ is privatized with independent channel for each components ant the components of $\boldsymbol{X}$ are independent, then the result of [17] applies componentwise. Indeed, the observations of $Z^2, \dots, Z^d$ carry no information about the value of X^1 and thus recovering information on X^1 from $\boldsymbol{Z}$ or from Z^1 is equivalent and a result like (4.1) applies with $\alpha = \alpha_1$.

The situation is more intricate if the components of $\boldsymbol{X}$ are dependent, as the observation of $Z^2, \dots, Z^d$ brings informations on X^1. In the extreme situation where all the components of $\boldsymbol{X}$ are almost surely equal $X^1 = \dots = X^d$, it is clear that the observation of $\boldsymbol{Z} = (Z^1, \dots, Z^d)$ is a repetition of d independent views of the same raw data X^1 with different privacy level. Thus, this mechanism is equivalent to the privatization of the single variable $X^1 \in \mathcal{X}^1$ through a channel taking the value $\boldsymbol{Z} = (Z^1, \dots, Z^d) \in \boldsymbol{\mathcal{Z}}$. By independence of the Z^j's conditional to X^1 and (2.3), we can check that this mechanism is a non componentwise α-LDP view of X^1 with $\alpha = \sum_{j=1}^d \alpha_j$. In turn, the lower bound (4.1) for deciding the value of X^1 from the observation of $\boldsymbol{Z}$ holds true with $\alpha = \sum_{j=1}^d \alpha_j \geq \alpha_1$. This evaluates how the privacy of X^1 is deteriorated by observation of the side-channels $Z^2, \dots, Z^d$ in the worst case scenario $X^1 = X^2 = \dots = X^d$.

In intermediate situation, we need to introduce some quantity which assesses the independence of $(X^2, \dots, X^d)$ on X^1. Let us denote by $q(dx^2, \dots, dx^d \mid X^1 = x^1)$ the conditional distribution of $X^2, \dots, X^d$ conditional to $X^1 = x^1$. We let

$$\Delta_{\text{ind}} := \sup_{x^1, x^{1\prime} \in \mathcal{X}^1} d_{TV}\left(q(dx^2, \dots, dx^d \mid X^1 = x^1), q(dx^2, \dots, dx^d \mid X^1 = x^{1\prime})\right), \tag{4.2}$$

which quantifies how close $(X^2, \dots, X^d)$ are from being independent from X^1. We have indeed $\Delta_{\text{ind}} \in [0, 2]$ and $\Delta_{\text{ind}} = 0$ when X^1 is independent from $(X^2, \dots, X^d)$. We let $m(z_1, \dots, z_d \mid X^1 = x^1)$ the density with respect to $\boldsymbol{\mu}$ of the law of $(Z^1, \dots, Z^d)$ conditional to $X^1 = x^1$. We have, using the notation of Section 3

$$m(z^1, \dots, z^d \mid X^1 = x^1) = \int_{\prod_{j=2}^d \mathcal{X}^j} \prod_{j=1}^d q^j(z^j \mid x^j) q(dx^2, \dots, dx^d \mid X^1 = x^1). \tag{4.3}$$

To elaborate, the function $\boldsymbol{z} \mapsto m(\boldsymbol{z} \mid X^1 = x^1)$ is the density of the channel which gives $\boldsymbol{Z}$ as a public view of the marginal X^1, gathering the information directly revealed by the channel q^1 and indirectly by the channels q^j, $j \geq 2$. The following proposition gives an upper bound for the privacy level of this channel. It is essentially a consequence of Theorem 3.1.

Proposition 4.1. *Let $\alpha_j \geq 0$, and assume that $\boldsymbol{Q} = (Q^1, \dots, Q^d)$ guarantees the $\boldsymbol{\alpha}$-CLDP constraint as defined by the condition (2.3). Assume that there exists α_{max} such that $\alpha_j \leq \alpha_{max}$ for $j \in \{2, \dots, d\}$. Then, we have*

$$\sup_{x^1, x^{1\prime} \in \mathcal{X}^1} \frac{m(z^1, \dots, z^d \mid X^1 = x^1)}{m(z^1, \dots, z^d \mid X^1 = x^{1\prime})} \leq \exp\left(\alpha_1 + \alpha_{max} \times (d-1)\Delta_{ind}\right). \tag{4.4}$$

Remark 4.2. *If $x^1, x^{1\prime} \in \mathcal{X}^1$ and $\psi : \boldsymbol{\mathcal{Z}} \to \{x^1, x^{1\prime}\}$ is any measurable application, then the average probability of mispredicting X^1 from $\boldsymbol{Z}$, $\frac{1}{2}\mathbb{P}\left(\psi(\boldsymbol{Z}) \neq x^1 \mid X^1 = x^1\right) + \frac{1}{2}\mathbb{P}\left(\psi(\boldsymbol{Z}) \neq x^{1\prime} \mid X^1 = x^{1\prime}\right)$ is lower bounded by the same quantity as in Equation (4.1) where α is replaced by $\alpha_1 + \alpha_{max} \times (d-1)\Delta_{ind}$.*

Proof of Proposition 4.1. We will apply the results of Section 3 with two well chosen probabilities $\widetilde{P}$, $\widetilde{P}'$ on $\widetilde{\boldsymbol{\mathcal{X}}} = \prod_{j=2}^d \mathcal{X}^j$. We fix $x^1, x^{1\prime} \in \mathcal{X}^1$ and let P be the measure on $\widetilde{\boldsymbol{\mathcal{X}}}$ given by

$$\widetilde{P}(dx^2, \dots, dx^d) = q(dx^2, \dots, dx^d \mid X^1 = x^1).$$

We define $\widetilde{P}'$ analogously with $x^{1\prime}$ in place of x^1. We denote by $\widetilde{M}$ the measure on $\widetilde{\boldsymbol{\mathcal{Z}}} = \prod_{j=2}^d \mathcal{Z}^j$ of the privatized view of P through the kernel $\widetilde{Q} = (Q^2, \dots, Q^d)$. In an analogous way, $\widetilde{M}'$ is the law of a privatized version of $\widetilde{P}'$. Let us denote by $\widetilde{m}(z^2, \dots, z^d)$ and $\widetilde{m}'(z^2, \dots, z^d)$ the densities of $\widetilde{M}$ and $\widetilde{M}'$. As emphasized in Remark 3.6, the equation (3.12) provides a control on the difference between $\widetilde{m}$ and $\widetilde{m}'$, which yields to

$$\frac{|\widetilde{m}(z^2, \dots, z^d) - \widetilde{m}'(z^2, \dots, z^d)|}{\widetilde{m}'(z^2, \dots, z^d)} \leq \sum_{k=1}^{d-1} \sum_{i_1, \dots, i_k} \prod_{u=1}^{k} (e^{\alpha_{i_u}} - 1) d_{TV}\left(\tilde{P}_{|(X^{i_1}, \dots, X^{i_k})}, \tilde{P}'_{|(X^{i_1}, \dots, X^{i_k})}\right)$$

where the inner sum is on $2 \leq i_1 < \cdots < i_k \leq d$ and $\tilde{P}_{|(X^{i_1}, \dots, X^{i_k})}$ is the restriction of the measure $\tilde{P}$ on $\prod_{u=1}^k \mathcal{X}^{i_u}$. We use the bound $d_{TV}\left(\tilde{P}_{|(X^{i_1}, \dots, X^{i_k})}, \tilde{P}'_{|(X^{i_1}, \dots, X^{i_k})}\right) \leq d_{TV}\left(\tilde{P}, \tilde{P}'\right)$ to deduce,

$$\begin{aligned}
\frac{|\widetilde{m}(z^2, \dots, z^d) - \widetilde{m}'(z^2, \dots, z^d)|}{\widetilde{m}'(z^2, \dots, z^d)} &\leq \sum_{k=1}^{d-1} \sum_{i_1, \dots, i_k} \prod_{u=1}^{k} (e^{\alpha_{i_u}} - 1) d_{TV}\left(\tilde{P}, \tilde{P}'\right) \\
&\leq \sum_{k=1}^{d-1} \binom{p-1}{k} (e^{\alpha_{\max}} - 1)^k d_{TV}\left(\tilde{P}, \tilde{P}'\right), \text{ using } \alpha_j \leq \alpha_{\max} \text{ for } j \geq 2, \\
&\leq \left[e^{\alpha_{\max} \times (p-1)} - 1\right] d_{TV}\left(\tilde{P}, \tilde{P}'\right), \text{ from the binomial formula.}
\end{aligned}$$

The definitions of $\widetilde{P}$ and $\widetilde{P}'$ as conditional distributions imply that $d_{TV}\left(\tilde{P}, \tilde{P}'\right) \leq \Delta_{\text{ind}}$, and thus we deduce

$$\frac{\widetilde{m}(z^2, \dots, z^d)}{\widetilde{m}'(z^2, \dots, z^d)} \leq 1 + \left[e^{\alpha_{\max} \times (p-1)} - 1\right] \Delta_{\text{ind}}.$$

Using the simple inequality $1 + (e^\alpha - 1)q \leq e^{\alpha q}$ for $\alpha, q \geq 0$, we get

$$\frac{\widetilde{m}(z^2, \dots, z^d)}{\widetilde{m}'(z^2, \dots, z^d)} \leq e^{\alpha_{\max} \times (p-1)\Delta_{\text{ind}}}. \tag{4.5}$$

Recalling that $\widetilde{m}(z^2, \dots, z^d)$ is the density of the privatized view of $\tilde{P}$ through $\widetilde{Q} = (Q^2, \dots, Q^d)$, we have

$$\widetilde{m}(z^2, \dots, z^d) = \int_{\prod_{j=2}^{d} \mathcal{X}^j} \prod_{j=2}^{d} q^j(z^j \mid x^j) q(dx^2, \dots, dx^d \mid X^1 = x^1),$$

and thus by comparison with (4.3)

$$m(z^1, \dots, z^d \mid X^1 = x^1) = q^1(z^1 \mid x^1)\widetilde{m}(z_2, \dots, z_d).$$

An analogous relation holds true for m' and in turn,

$$\frac{m(z^1, \dots, z^d \mid X^1 = x^1)}{m(z^1, \dots, z^d \mid X^1 = x^{1\prime})} = \frac{q(z^1 \mid x^1)}{q(z^1 \mid x^{1\prime})} \frac{\widetilde{m}(z^2, \dots, z^d)}{\widetilde{m}'(z^2, \dots, z^d)}.$$

Now, the proposition is a consequence of (2.4) and (4.5). □

4.2 Locally private covariance estimation

In this section we assume that $\boldsymbol{X} = (X^1, X^2)$ is a two-dimensional vector, for which we want to estimate the covariance under local differential privacy constraints. Again, the components are made public separately.

4.2.1 Local differential private estimator

We assume that $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ are n iid copies of $\boldsymbol{X} = (X^1, X^2)$. As in this paper we stick to the framework of local differential privacy, we want to introduce an anonymization procedure to transform the X_i^j to some

$$Z_i^j \sim q_j(dz|X_i^j) = q_j(z|X_i^j)dz$$

which satisfies the condition of local differential privacy, as in (2.4). In particular, the privacy mechanism we consider in this example is non-interactive.
It is well-known that adding centered Laplace distributed noise on bounded random variables provides α differential privacy (cfr [17], [28], [33]). This motivates our choice for the anonymization procedure, which consists in constructing the public version of the $\boldsymbol{X}_i$ by using a Laplace mechanism with an independent channel for each component. Let us denote $\mathcal{E}_j^i$ for $(i, j) \in \{1, \dots, n\} \times \{1, 2\}$ a family of independent random variables, such that $(\mathcal{E}_j^i)_i$ are iid sequences with law $\mathcal{L}(\frac{2T^{(j)}}{\alpha_j})$ for $j = 1, 2$. The truncation $T^{(j)} > 0$ will be specified later. We assume that the variables $\mathcal{E}_i^j$ are independent from the data $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ and we set

$$Z_i^j = \left[X_i^j\right]_{T^{(j)}} + \mathcal{E}_i^j, \quad \forall (i, j) \in \{1, \dots, n\} \times \{1, 2\}, \tag{4.6}$$

where $[x]_T = \max(\min(x, T), -T)$.
Denoting by $z \mapsto q^j(z \mid X_i^j = x)$ the density of the privatized data Z_i^j conditional to $X_i^j = x$ for $j \in \{1, 2\}$, it is easy to check that the local differential privacy control (2.4) holds true, as proven in the following lemma.

Lemma 4.3. *For any $i \in \{1, ..., n\}$ and $j = 1, 2$, the random variables $Z_i^j = \left[X_i^j\right]_{T^{(j)}} + \mathcal{E}_i^j$, with $\mathcal{E}_i^j$ iid $\sim \mathcal{L}(\frac{2T^{(j)}}{\alpha_j})$, are α_j differential private views of the original X_i^j.*

Proof. As $\mathcal{E}_i^j$ is distributed as a centered Laplace random variable with scale parameter $\frac{2T^{(j)}}{\alpha_j}$, its density at the point $x \in \mathbb{R}$ is given by $\frac{1}{2T^{(j)}}\alpha_j \exp(-\frac{1}{2T^{(j)}}\alpha_j|x|)$. Then, the reverse triangle inequality and the fact that $\left[X_i^j\right]_{T^{(j)}}$ is bounded by $T^{(j)}$ provide

$$\begin{aligned}\sup_{z\in\mathcal{Z}} \frac{q^j(z|X_i^j=x)}{q^j(z|X_i^j=x')} &\le \sup_{z\in\mathcal{Z}} \exp\Big(-\frac{1}{2T^{(j)}}\alpha_j(z-[x]_{T^{(j)}})+\frac{1}{2T^{(j)}}\alpha_j(z-\left[x'\right]_{T^{(j)}})\Big)\\ &\le \exp\Big(\frac{1}{2T^{(j)}}\alpha_j([x]_{T^{(j)}}-\left[x'\right]_{T^{(j)}})\Big)\\ &\le \exp(\alpha_j),\end{aligned}$$

as we wanted. □

Assume that $X^1 \in \mathbf{L}^{k_1}$, $X^2 \in \mathbf{L}^{k_2}$ for $k_1 > 1$, $k_2 > 1$ and with $k_1^{-1} + k_2^{-1} < 1$. It ensures that $\mathbb{E}[|X^1X^2|] < \infty$, by Hölder's inequality.
The goal is to estimate $\theta := \mathrm{cov}(X^1, X^2) = \mathbb{E}[X^1X^2] - \mathbb{E}[X^1]\mathbb{E}[X^2]$.
The estimation of $m^{(1)} := \mathbb{E}[X^1]$ and $m^{(2)} := \mathbb{E}[X^2]$ is discussed in [17], from which we recall the result. We will state later the result on the estimation of the cross term $\gamma := \mathbb{E}[X^1X^2]$.
Let

$$\hat{m}_n^{(j)} := \frac{1}{n}\sum_{i=1}^n Z_i^j, \text{for } j \in \{1,2\}, \tag{4.7}$$

$$\hat{\gamma}_n := \frac{1}{n}\sum_{i=1}^n Z_i^1 Z_i^2, \quad \hat{\theta}_n := \hat{\gamma}_n - \hat{m}_n^{(1)}\hat{m}_n^{(2)}. \tag{4.8}$$

Theorem 4.4 (Corollary 1 in [17])**.** *Let $0 < \alpha_j \le 1$. Assume $k_j > 1$ and set $\widetilde{T}^{(j)} = (n\alpha_j^2)^{1/(2k_j)}$ for $j \in \{1,2\}$. Then, there exists $c > 0$ such that for all $n \ge 1$, $j \in \{1,2\}$,*

$$\mathbb{E}\left[\left|m_n^{(j)} - m^{(j)}\right|^2\right] \le c(n\alpha_j^2)^{-\frac{k_j-1}{k_j}}.$$

In Corollary 1 in [17], the privacy level α is the same for all components. However, the result is useful also in our context, as we plan to apply Corollary 1 in [17] with $d = 1$ separately to each components of $\boldsymbol{X}$. By [17], the choice of truncation $\widetilde{T}^{(j)} = (n\alpha_j^2)^{1/(2k_j)}$ is optimal when estimating $m^{(j)}$. The result for the estimation of the covariance is the following.

Theorem 4.5. *Let $\alpha_j \le 1$. Assume $k_1 > 1, k_2 > 1, 1/k_1 + 1/k_2 < 1$ and set $T^{(j)} = (n\alpha_j^4)^{1/(2k_j)}$ for $j \in \{1,2\}$. Then, there exists $c > 0$, such that for all $n \ge 1$,*

$$\mathbb{E}\left[|\hat{\gamma}_n - \gamma|^2\right] \le c(n\alpha_1^2\alpha_2^2)^{-\frac{\overline{k}-2}{\overline{k}}}, \tag{4.9}$$

$$\mathbb{E}\left[\left|\hat{\theta}_n - \theta\right|^2\right] \le c(n\alpha_1^2\alpha_2^2)^{-\frac{\overline{k}-2}{\overline{k}}}, \tag{4.10}$$

where $\overline{k} = 2\left(\frac{1}{k_1}+\frac{1}{k_2}\right)^{-1} > 2$ is the harmonic mean of k_1, k_2. The constant c does not depend on α, n, as soon as $n\alpha_1^2\alpha_2^2 \ge 1$.

The proof of this theorem is in the appendix. It relies on a bias-variance trade-off, and the choice for $T^{(j)}$, given in the statement, is in this regard optimal.

Remark 4.6. *The upper bound provided by Theorem 4.5 can be compared with the case where the private data $\boldsymbol{X}_i = (X_i^1, X_i^2)$ is disclosed using a single channel that accesses both components X_i^1 and X_i^2. In this scenario, we can apply the results of Section 3.2.1 in [17] to estimate the mean of the iid private data $(\Gamma_i)_{i=1,\dots,n}$, where $\Gamma_i = X_i^1X_i^2$, using a locally differentially* *privatized version of* Γ_i. *By applying Hölder's inequality, we can see that* $\Gamma_1 \in \mathbf{L}^{\overline{k}/2}$, *and hence by Corollary 1 in [17], there exists an estimator* $\tilde{\gamma}_n$ *such that*

$$\mathbb{E}\left[(\tilde{\gamma}_n - \gamma)^2\right] \leq c(n\alpha^2)^{-\frac{\overline{k}/2-1}{\overline{k}/2}} = c(n\alpha^2)^{-\frac{\overline{k}-2}{\overline{k}}},$$

where α *is the LDP level when disclosing the* $\boldsymbol{X}_i$*'s.*
We can conclude that the rate exponent for estimating γ *is unchanged when the data are disclosed using independent channels for each component, compared to a situation where both components can be accessed before publicly releasing the data. However, the effective number of data is reduced from* $n\alpha^2$ *to* $n\alpha_1^2\alpha_2^2$. *The same holds true for estimating the covariance* θ. *If we consider the special case where* $\alpha_1 = \alpha_2 = \alpha$, *it is evident that the loss is significant for small values of* α. *On the other hand, the loss can be moderate if one of the two* α_j *values is close to* 1. *We will see in Section 4.2.2 that this loss is unavoidable.*

Remark 4.7. *By definitions* (4.6)–(4.8), *when estimating* $\theta = \gamma - m^{(1)}m^{(2)}$, *we use the same truncation levels* $T^{(j)}$ *for the estimation of* γ, $m^{(1)}$ *and* $m^{(2)}$. *It would be possible to use the optimal levels* $\widetilde{T}^{(j)} = (n\alpha_j^2)^{1/(2k_j)}$ *for the estimation of* $m^{(1)}$, $m^{(2)}$ *and the optimal levels* $T^{(j)} = (n\alpha_1^2\alpha_2^2)^{1/(2k_j)}$ *for the estimation of* γ. *However, this approach would necessitate publicly disclosing two values for each private data point: one corresponding to the truncation level* $\widetilde{T}^{(j)}$ *and one with level* $T^{(j)}$. *As a result, the overall privacy of the procedure would be reduced. In Section 4.2.3, we will explore another scenario where we must disclose multiple public values for each private data point.*

4.2.2 Lower bound for the covariance estimation

For $k_1 > 1, k_2 > 1$ with $1/k_1 + 1/k_2 < 1$, we introduce the notation

$$\mathcal{P}_{k_1,k_2} = \left\{P, \quad \text{probability on } \mathbb{R}^2 \text{ such that } \mathbb{E}_P[|X^1|^{k_1}] \leq 1, \mathbb{E}_P[|X^2|^{k_2}] \leq 1\right\},$$

where $\boldsymbol{X} = (X^1, X^2)$ is the canonical random variable on $\mathbb{R}^2$. For $P \in \mathcal{P}_{k_1,k_2}$, we set

$$\gamma(P) = \mathbb{E}_P\left[X^1X^2\right],\ m^{(l)}(P) = \mathbb{E}_P[X^l] \text{ for } l = 1,2,\ \theta(P) = \gamma(P) - m^{(1)}(P)m^{(2)}(P).$$

We denote by $\mathcal{Q}_{\boldsymbol{\alpha}}$ the set of privacy mechanisms, where for simplicity we restrict ourself to non interactive kernels. Thus, $\boldsymbol{Q} = (Q^1, Q^2) \in \mathcal{Q}_{\boldsymbol{\alpha}}$ is such that Q^j is a Markov kernel from $\mathcal{X}^j = \mathbb{R}$ to some measurable space $(\mathcal{Z}^j, \Xi_{\mathcal{Z}^j})$, and the condition (2.3) is satisfied for $j = 1, 2$.
The private data are given by the iid sequence $(\boldsymbol{X}_i)_{i=1,\dots,n}$. We assume that the public data $(\boldsymbol{Z}_i)_{i=1,\dots,n}$ are given by the non interactive mechanism where the variable Z_i^j is drawn according to the law $Q^j(dz \mid X_i^j)$.
We introduce the minimax risk

$$\mathcal{M}_n(\gamma(\mathcal{P}_{k_1,k_2}), \boldsymbol{\alpha}) = \inf_{\boldsymbol{Q}\in\mathcal{Q}_{\boldsymbol{\alpha}}} \inf_{\hat{\gamma}_n} \sup_{P\in\mathcal{P}_{k_1,k_2}} \mathbb{E}_P\left[(\hat{\gamma}_n - \gamma(P))^2\right],$$

where $\hat{\gamma}_n$ is any $\hat{\gamma}_n((Z_i^j)_{\substack{1\leq i\leq n\\1\leq j\leq 2}})$ measurable function from $(\mathcal{Z}^1)^n \times (\mathcal{Z}^2)^n$ taking values in $\mathbb{R}$, with finite second moment. We define analogously

$$\mathcal{M}_n(\theta(\mathcal{P}_{k_1,k_2}), \boldsymbol{\alpha}) = \inf_{\boldsymbol{Q}\in\mathcal{Q}_{\boldsymbol{\alpha}}} \inf_{\hat{\theta}_n} \sup_{P\in\mathcal{P}_{k_1,k_2}} \mathbb{E}_P\left[\left(\hat{\theta}_n - \theta(P)\right)^2\right].$$

Theorem 4.8. *There exists some constant c such that,*

$$\mathcal{M}_n(\gamma(\mathcal{P}_{k_1,k_2}), \boldsymbol{\alpha}) \geq c(n|e^{\alpha_1}-1|^2|e^{\alpha_2}-1|^2)^{-\frac{\overline{k}-2}{\overline{k}}}, \quad \mathcal{M}_n(\theta(\mathcal{P}_{k_1,k_2}), \boldsymbol{\alpha}) \geq c(n|e^{\alpha_1}-1|^2|e^{\alpha_2}-1|^2)^{-\frac{\overline{k}-2}{\overline{k}}},$$

for all $n \geq 1$, $n|e^{\alpha_1} - 1|^2|e^{\alpha_2} - 1|^2 \geq 1$.

Remark 4.9. *Comparing with Theorem 4.5, we see that when $\alpha_1 \leq 1$, $\alpha_2 \leq 1$ the rate $(n\alpha_1^2\alpha_2^2)^{\frac{\overline{k}-2}{\overline{k}}}$ achieved by the estimator of Section 4.2.1 can not be improved.*

Proof. • We first focus on a lower bound for $\mathcal{M}_n(\gamma(\mathcal{P}_{k_1,k_2}),\boldsymbol{\alpha})$. Let $\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\alpha}}$. As we want to apply the two hypothesis method (see for example Section 2.3 in [36]), we need to construct P and P^* such that

1. P, P^* are elements of $\mathcal{P}_{k_1,k_2}$,
2. $\exists c > 0$ with $|\gamma(P) - \gamma(P^*)| \geq c(n|e^{\alpha_1} - 1|^2|e^{\alpha_2} - 1|^2)^{-\frac{\overline{k}-2}{2\overline{k}}}$,
3. $\exists \epsilon_0 > 0$ such that $d_{KL}\Big(Law((\boldsymbol{Z}_i)_{i=1,\dots,n}), Law((\boldsymbol{Z}_i^*)_{i=1,\dots,n})\Big) < \epsilon_0 < 2$,

where $\boldsymbol{Z}_i = (Z_i^1, Z_i^2)$ for $i = 1,\dots,n$ are the public views of $\boldsymbol{X}_i = (X_i^1, X_i^2)$, $i = 1,\dots,n$ a iid sequence of random variables with law P, and $\boldsymbol{Z}_i^* = (Z_i^1, Z_i^2)$ are the public views of $\boldsymbol{X}_i^* = (X_i^{*,1}, X_i^{*,2})$, $i = 1,\dots,n$ a iid sequence of random variables with law P^*.

Let $0 < \delta < 1$ be a parameter whose value will be calibrated later. We denote by P the probability on $\mathbb{R}^2$ which makes $\boldsymbol{X} = (X^1, X^2)$ a discrete random variable taking values in $\{-\delta^{1/k_1}, 0, -\delta^{1/k_1}\} \times \{-\delta^{1/k_2}, 0, -\delta^{1/k_2}\}$ with the following joint distribution,

X^2 \ X^1	$-\delta^{-1/k_1}$	0	δ^{-1/k_1}	Law of X^2
$-\delta^{-1/k_2}$	$\frac{\delta^2}{4}$	$\frac{\delta}{2}(1-\delta)$	$\frac{\delta^2}{4}$	$\delta/2$
0	$\frac{\delta}{2}(1-\delta)$	$(1-\delta)^2$	$\frac{\delta}{2}(1-\delta)$	$1-\delta$
δ^{-1/k_2}	$\frac{\delta^2}{4}$	$\frac{\delta}{2}(1-\delta)$	$\frac{\delta^2}{4}$	$\delta/2$
Law of X^1	$\delta/2$	$1-\delta$	$\delta/2$	

We denote P^* the probability on the same space as P but with the different weights

$X^{*,2}$ \ $X^{*,1}$	$-\delta^{-1/k_1}$	0	δ^{-1/k_1}	Law of $X^{*,2}$
$-\delta^{-1/k_2}$	$\frac{\delta}{2}(1-\delta)$	$\frac{\delta^2}{4}$	$\frac{\delta^2}{4}$	$\delta/2$
0	$\frac{\delta^2}{4}$	$1-\delta-\frac{\delta^2}{2}$	$\frac{\delta^2}{4}$	$1-\delta$
δ^{-1/k_2}	$\frac{\delta^2}{4}$	$\frac{\delta^2}{4}$	$\frac{\delta}{2}(1-\delta)$	$\delta/2$
Law of $X^{*,1}$	$\delta/2$	$1-\delta$	$\delta/2$	

The joint laws of $\boldsymbol{X}$ and $\boldsymbol{X}^*$ are different as the components of $\boldsymbol{X}$ are independent and those of $\boldsymbol{X}^*$ have a strong positive correlation. On the other hand, the marginal laws are the same, which is a crucial fact for the application of Corollary 3.8.

We have $\mathbb{E}_P(|X^j|^{k_j}) = \mathbb{E}_{P^*}(|X^j|^{k_j}) = 1$, for all $j \in \{1,2\}$, and thus $P,\ P^* \in \mathcal{P}_{k_1,k_2}$. Moreover, we have $\theta(P^*) = \gamma(P^*) = \delta^{-\frac{1}{k_1}-\frac{1}{k_2}}[\delta + \delta^2/2]$ and $\theta(P) = \gamma(P) = 0$. Hence, it is $|\gamma(P) - \gamma(P^*)| \geq \delta^{1-\frac{1}{k_1}-\frac{1}{k_2}}$.

We apply Corollary 3.8 to the sequences of raw samples $(\boldsymbol{X}_i)_{i=1,\dots,n}$, $(\boldsymbol{X}_i^*)_{i=1,\dots,n}$ whose distributions are $P^{\otimes n}$ and $(P^*)^{\otimes n}$. Indeed, this is permitted as the marginal laws of the two dimensional vectors $\boldsymbol{X}_i$ and $\boldsymbol{X}_i^*$ coincide. We deduce

$$d_{KL}\Big(Law((\boldsymbol{Z}_i)_{i=1,\dots,n}), Law((\boldsymbol{Z}_i^*)_{i=1,\dots,n})\Big) \leq n \times |e^{\alpha_1} - 1|^2|e^{\alpha_2} - 1|^2 d_{TV}(P, P^*)^2$$

Furthermore, $d_{TV}(P,P^*) = \sum_{(u,v)\in\{-\delta^{1/k_1},0,-\delta^{1/k_1}\}\times\{-\delta^{1/k_2},0,-\delta^{1/k_2}\}} |P\{(u,v)\} - P^*\{(u,v)\}| \leq 9\delta$, and so we get

$$d_{KL}\Big(Law((\boldsymbol{Z}_i)_{i=1,\dots,n}), Law((\boldsymbol{Z}_i^*)_{i=1,\dots,n})\Big) \leq n \times |e^{\alpha_1}-1|^2|e^{\alpha_2}-1|^2 81\delta^2.$$

We now set $\delta = \left(81|e^{\alpha_1}-1|^2|e^{\alpha_2}-1|^2 n\right)^{-1/2}$ which is strictly smaller than 1 by the assumption $n|e^{\alpha_1}-1|^2|e^{\alpha_2}-1|^2 \geq 1$. Then, we get, $d_{KL}\Big(Law((\boldsymbol{Z}_i)_{i=1,\dots,n}), Law((\boldsymbol{Z}_i^*)_{i=1,\dots,n})\Big) = 1 < 2$. Moreover,

$$|\gamma(P)-\gamma(P^*)| \geq \delta^{1-\frac{1}{k_1}-\frac{1}{k_2}} \geq c\left(n|e^{\alpha_1}-1|^2|e^{\alpha_2}-1|^2\right)^{-(\frac{1}{2}-\overline{k}^{-1})} = c\left(n|e^{\alpha_1}-1|^2|e^{\alpha_2}-1|^2\right)^{-\frac{\overline{k}-2}{2\overline{k}}}.$$

We have obtained the points 1–3 stated at the beginning of the proof and the lower bound on $\mathcal{M}_n(\gamma(\mathcal{P}_{k_1,k_2}),\boldsymbol{\alpha})$ follows.
• The lower bound for $\mathcal{M}_n(\theta(\mathcal{P}_{k_1,k_2}),\boldsymbol{\alpha})$ is shown in the same way, after remarking that $\theta(P)-\theta(P^*) = \gamma(P)-\gamma(P^*)$, since $\mathbb{E}_P[X^j] = \mathbb{E}_{P^*}[X^j] = 0$ for $j=1,2$. □

4.2.3 Adpative estimation of the covariance

As discussed in Section 4.2.1, the privacy procedure proposed in this study requires selecting the optimal truncation levels $T^{(j)}$, which depend on the number of finite moments k_j for the variables. However, in practice, it is unrealistic to assume that the number of finite moments is known in all situations. To address this issue, we propose an adaptive method to estimate the covariance that does not necessitate prior knowledge of k_1 and k_2 and conforms to the privacy constraint.
The main idea is to send a collection of public data with different truncation levels via the privatization channel, and then let the statistician decide on the optimal truncation level using a penalization method.
For simplicity we focus on the estimation of $\gamma = \mathbb{E}[X^1X^2]$.
We introduce the following set of truncations :

$$\begin{aligned}&\mathcal{T} := \mathcal{T}^{(1)} \times \mathcal{T}^{(2)}, \text{where for } l=1,2,\\ &\mathcal{T}^{(l)} := \left\{T^{(l)} \in (0,\infty) \mid T^{(l)} = \frac{n}{2^r}, \quad \text{for some } r \in \{1,\dots,\lfloor \log_2(n)\rfloor\}\right\}.\end{aligned} \tag{4.11}$$

Let $\beta_n^1 > 0$, $\beta_n^2 > 0$ be two parameters that we will specify later. For all $i \in \{1,\dots,n\}$ we are given $card(\mathcal{T}^{(1)}) + card(\mathcal{T}^{(2)}) = 2\lfloor\log_2(n)\rfloor$ independent variables, $\mathcal{E}_i^{1,T^{(1)}}$ and $\mathcal{E}_i^{2,T^{(2)}}$, for $T^{(1)} \in \mathcal{T}^{(1)}$ and $T^{(2)} \in \mathcal{T}^{(2)}$, respectively. We assume that each of the variables $\mathcal{E}_i^{1,T^{(j)}}$ follows a Laplace distribution with parameter $2\frac{T^{(j)}}{\beta_n^j}$, where $T^{(j)} \in \mathcal{T}^{(j)}$ and j takes values 1 and 2. We further assume that these variables are independent for different values of i ranging from 1 to n.
We define the privatized data $\boldsymbol{Z}_i = (Z_i^1, Z_i^2) \in \mathbb{R}^{\mathcal{T}^{(1)}} \times \mathbb{R}^{\mathcal{T}^{(2)}}$, $i \in \{1,\dots,n\}$, by

$$\begin{aligned}Z_i^1 &= (Z_i^{1,T})_{T\in\mathcal{T}^{(1)}}, \quad Z_i^{1,T} = \left[X_i^1\right]_T + \mathcal{E}_i^{(1),T}, \quad \text{for } T \in \mathcal{T}^{(1)},\ i \in \{1,\dots,n\},\\ Z_i^2 &= (Z_i^{2,T})_{T\in\mathcal{T}^{(2)}}, \quad Z_i^{2,T} = \left[X_i^2\right]_T + \mathcal{E}_i^{(2),T}, \quad \text{for } T \in \mathcal{T}^{(2)},\ i \in \{1,\dots,n\}.\end{aligned} \tag{4.12}$$

Lemma 4.10. *Assume that $\beta_n^j = \frac{\alpha_j}{card(\mathcal{T}^{(j)})} = \frac{\alpha_j}{\lfloor\log_2(n)\rfloor}$. Then, the privacy procedure satisfies the $\boldsymbol{\alpha}$-CLDP constraint as in* (2.4).

Proof. For $j=1,2$, let us denote by $q^j((z^{j,T})_{T\in\mathcal{T}^{(j)}} \mid X^j = x)$ the density of the law of $Z_i^j = (Z_i^{j,T})_{T\in\mathcal{T}^{(j)}}$ conditional to $X_i^j = x \in \mathbb{R}$. Then, using the independence property of the variables $(Z_i^{j,T})_{T\in\mathcal{T}^{(j)}}$, we have

$$\begin{aligned}\frac{q^j((z^{j,T})_{T\in\mathcal{T}^{(j)}} \mid X_i^j = x)}{q^j((z^{j,T})_{T\in\mathcal{T}^{(j)}} \mid X_i^j = x')} &= \frac{\prod_{T\in\mathcal{T}^{(j)}} \exp\left(|z^{j,T} - [x]_T|\frac{\beta_n^j}{2T}\right)}{\prod_{T\in\mathcal{T}^{(j)}} \exp\left(|z^{j,T} - [x']_T|\frac{\beta_n^j}{2T}\right)} \\ &= \prod_{T\in\mathcal{T}^{(j)}} \exp\left(\frac{\beta_n^j}{2T}\left\{|z^{j,T} - [x]_T| - |z^{j,T} - [x']_T|\right\}\right) \\ &\leq \prod_{T\in\mathcal{T}^{(j)}} \exp\left(\frac{\beta_n^j}{2T}|[x]_T - [x']_T|\right)\end{aligned}$$

where we used the inverse triangular inequality in the last line. As $|[x]_T - [x']_T| \leq 2T$ we deduce,

$$\frac{q^j((z^{j,T})_{T\in\mathcal{T}^{(j)}} \mid X_i^j = x)}{q^j((z^{j,T})_{T\in\mathcal{T}^{(j)}} \mid X_i^j = x')} \leq \prod_{T\in\mathcal{T}^{(j)}} \exp\left(\beta_n^j\right) = \exp(card(\mathcal{T}^{(j)})\beta_n^j) \leq \exp(\alpha_j),$$

by the choice of β_n^j. □

We construct our adaptive estimator, following Goldenshluger-Lepski method. For $T = (T^{(1)}, T^{(2)}) \in \mathcal{T}$, we set

$$\hat{\gamma}_n^{(T)} = \frac{1}{n}\sum_{i=1}^n Z_i^{1,T^{(1)}} Z_i^{2,T^{(2)}}, \tag{4.13}$$

and for $T = (T^{(1)}, T^{(2)}) \in \mathcal{T}$, $T' = (T'^{(1)}, T'^{(2)}) \in \mathcal{T}$

$$\hat{\gamma}_n^{(T,T')} = \frac{1}{n}\sum_{i=1}^n Z_i^{1,T^{(1)}\wedge T'^{(1)}} Z_i^{2,T^{(2)}\wedge T'^{(2)}}. \tag{4.14}$$

Let us remark that the following commutativity relation hold true: $\hat{\gamma}_n^{(T,T')} = \hat{\gamma}_n^{(T',T)}$. Based on the upper bound (A.9) given in the Appendix for the variance of the estimator, we introduce the penalization term for $T \in \mathcal{T}$,

$$\mathbb{V}_T = \mathbb{V}_{(T^{(1)},T^{(2)})} = \kappa_n \frac{|T^{(1)}|^2 |T^{(2)}|^2}{n|\beta_n^1|^2 |\beta_n^2|^2}, \tag{4.15}$$

for $\kappa_n \geq 1$ some sequence tending slowly to ∞, which will be specified in Theorem 4.11.
For $T \in \mathcal{T}$, we define

$$\mathbb{B}_T = \sup_{T'\in\mathcal{T}}\left\{\left(\left|\hat{\gamma}_n^{(T,T')} - \hat{\gamma}_n^{(T')}\right|^2 - \mathbb{V}_{T'}\right)_+\right\}, \tag{4.16}$$

and set

$$\widehat{T} = \underset{T\in\mathcal{T}}{\text{argmin}} \left\{\mathbb{B}_T + \mathbb{V}_T\right\}. \tag{4.17}$$

Our adaptive estimator is $\hat{\gamma}_n^{(\widehat{T})}$.

Theorem 4.11. *Assume that $k_1^{-1} + k_2^{-1} < 1$, $\beta_n^j = \frac{\alpha_j}{\lfloor \log_2(n) \rfloor}$, for $j = 1, 2$ and $\kappa_n = c_0 \log(n)$ for some $c_0 > 0$. If c_0 is large enough, there exist $c > 0$, $\overline{c}_0 > 0$, such that*

$$\mathbb{E}\left[(\hat{\gamma}_n^{\widehat{T}} - \gamma)^2\right] \leq c\left(\frac{n\alpha_1^2\alpha_2^2}{(\log(n))^5}\right)^{-\frac{\overline{k}-2}{\overline{k}}} + \frac{c}{\alpha_1^2\alpha_2^2 n^{\overline{c}_0}},$$

for all $n \geq 1$, $\alpha_j \leq 1$, $(n\alpha_1^2\alpha_2^2)/(\log(n))^5 \geq 1$. Moreover, the constant $\overline{c}_0$ can be chosen arbitrarily large by choosing c_0 large enough.

Remark 4.12. *Comparing with Theorem 4.5, we observe that the rate of the adaptive version of the estimator worsens by a factor of* $\log(n)^5$*. The loss of a* $\log(n)$ *factor is a well-known characteristic of adaptive methods and is sometimes unavoidable, as mentioned in [9]. The additional loss of a* $\log(n)^4$ *term arises from the disclosure of* $\mathrm{card}\,\mathcal{T}^{(j)} \asymp \log_2(n)$ *observations for each raw data, which increases the variance of the privatization mechanism while maintaining a constant level of privacy, as demonstrated in Lemma 4.10. This is one reason why, in defining the sets* $\mathcal{T}^{(j)}$*, we have attempted to minimize their cardinality.*

The proof of the adaptive procedure gathered in Theorem 4.11 can be found in the appendix.

4.3 Locally private multivariate density estimation

In this section we consider the non-parametric estimation of the density of the vector $\boldsymbol{X} = (X^1, ..., X^d)$, under α-CLDP. We will see that, similarly to the case where the components become public jointly, this implies a deterioration on the convergence rate depending on $\boldsymbol{\alpha}$ (see for example Section 5.2.2 of [17]).
Consider $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$, n iid copies of $\boldsymbol{X}$. We will assume that the density π of $\boldsymbol{X}$ belongs to an Hölder class $\mathcal{H}(\beta, \mathcal{L})$ (see for example Definition 1.2 in [36]). We aim at estimating such density under componentwise local differential privacy. We recall we reduce to consider the non-interactive privacy mechanism for the statistical applications, in order to lighten the notation.

4.3.1 Local differential private estimator

In absence of privacy constraints, a well-studied estimator for density estimation consists in the kernel density estimator (see for example Section 1.2 of [36] and Part III of [7]). It achieves the convergence rate $n^{-\frac{2\beta}{2\beta+d}}$, which has been shown to be optimal in a minimax sense (see Theorem 1.1 in [36] for the monodimensional case).
We therefore introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfying, for all $l \in \{1, ..., \beta\}$,

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \|K\|_\infty < \kappa, \quad \mathrm{supp}(K) \subset [-1, 1], \quad \int_{\mathbb{R}} K(x)x^l dx = 0. \tag{4.18}$$

Then, as for the estimation of the covariance, we add centered Laplace distributed noise on bounded random variables to obtain α-CLDP.

Lemma 4.13. *For any* $i \in \{1, ..., n\}$*,* $j \in \{1, ..., d\}$ *and any* $\boldsymbol{x_0} \in \mathbb{R}^d$*, the random variables*

$$Z_i^j := \frac{1}{h} K\Big(\frac{X_i^j - x_0^j}{h}\Big) + \mathcal{E}_i^j, \tag{4.19}$$

with $\mathcal{E}_i^j$ *iid* $\sim \mathcal{L}(\frac{2\kappa}{\alpha_j h})$*, are* α_j*-differentially private views of the original* X_i^j*.*

The index h introduced in (4.19) is small. In particular, we assume $h < 1$. The proof of Lemma 4.13 consists in checking property (2.4), similarly as in Lemma 4.3.

Proof. The density of $\mathcal{E}_i^j$ at the point $x \in \mathbb{R}$ is given by $\frac{1}{2\kappa}\alpha_j h \exp(-\frac{1}{2\kappa}\alpha_j h|x|)$. Then, the reverse triangle inequality and the fact the infinity norm of K is bounded by κ provide

$$\begin{aligned}
\sup_{z \in \mathcal{Z}} \frac{q^j(z|X_i^j = x)}{q^j(z|X_i^j = x')} &\le \sup_{z \in \mathcal{Z}} \exp\Big(-\frac{1}{2\kappa}\alpha_j h(z - \frac{1}{h}K(\frac{x - x_0^j}{h})) + \frac{1}{2\kappa}\alpha_j h(z - \frac{1}{h}K(\frac{x' - x_0^j}{h}))\Big) \\
&\le \exp\Big(\frac{1}{2\kappa}\alpha_j h \frac{1}{h}(K(\frac{x - x_0^j}{h}) - K(\frac{x' - x_0^j}{h}))\Big) \\
&\le \exp(\alpha_j).
\end{aligned}$$

□

We introduce the kernel density estimator $\hat{\pi}_h^Z$ based on the discrete observations Z_i^j, for $i \in \{1, ..., n\}$ and $j \in \{1, ..., d\}$. We define, for any $\boldsymbol{x_0} \in \mathbb{R}^d$,

$$\hat{\pi}_h^Z(\boldsymbol{x_0}) := \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} Z_i^j = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\Big(\frac{1}{h}K(\frac{X_i^j - x_0^j}{h}) + \mathcal{E}_i^j\Big). \tag{4.20}$$

We now prove an upper bound the L^2 pointwise risk, showing that $\hat{\pi}_h^Z$ achieves the convergence rate $(\frac{1}{n\prod_{j=1}^d \alpha_j^2})^{\frac{\beta}{\beta+d}}$.

Theorem 4.14. *Assume $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid copies of an $\mathbb{R}^d$ vector $\boldsymbol{X}$ whose density π belongs to the Hölder class $\mathcal{H}(\beta, \mathcal{L})$ and $\boldsymbol{x_0} \in \mathbb{R}^d$. Let $0 < \alpha_j \le 1$ for any $j \in \{1, ..., d\}$. If $n\prod_{j=1}^d \alpha_j^2 \to \infty$, then there exist $c > 0$ and $n_0 > 0$ such that for any $n \ge n_0$,*

$$\mathbb{E}[|\hat{\pi}_h^Z(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|^2] \le c(\frac{1}{n\prod_{j=1}^d \alpha_j^2})^{\frac{\beta}{\beta+d}}.$$

This shows that the effects of local differential privacy constraints are severe for non-parametric density estimation, as they lead to a different convergence rate.
In the case where $\alpha_1 = ... = \alpha_d$ it is possible to obtain the following result, which provides the threshold which dictates the behaviour of the estimator with respect to the privacy mechanism. Indeed, for $\alpha \ge n^{\frac{1}{2(2\beta+d)}}$, we recover the same convergence rate as in absence of privacy.

Theorem 4.15. *Assume $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid copies of an $\mathbb{R}^d$ vector $\boldsymbol{X}$ whose density π belongs to the Hölder class $\mathcal{H}(\beta, \mathcal{L})$, and $\boldsymbol{x_0} \in \mathbb{R}^d$. Then, the following inequalities hold true*

1. *If $\alpha \ge n^{\frac{1}{2(2\beta+d)}}$, then there exist $c > 0$ and $n_0 > 0$ such that, for any $n \ge n_0$,*

$$\mathbb{E}[|\hat{\pi}_h^Z(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|^2] \le c(\frac{1}{n})^{\frac{2\beta}{2\beta+d}}.$$

2. *If otherwise $\alpha < n^{\frac{1}{2(2\beta+d)}}$ and $n\alpha^{2d} \to \infty$, then there exist $c > 0$ and $n_0 > 0$ such that, for any $n \ge n_0$,*

$$\mathbb{E}[|\hat{\pi}_h^Z(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|^2] \le c(\frac{1}{n\alpha^{2d}})^{\frac{\beta}{\beta+d}}.$$

The proof of these two results can be found in the appendix.

Remark 4.16. *The above result indicates that a threshold for the behavior of a system with and without privacy is determined by $n^{\frac{1}{2(2\beta+d)}}$. If α is greater than this value, it means that the level of privacy provided is not significant enough to degrade the convergence rate of our estimator compared to the case without privacy. However, if α is smaller than $n^{\frac{1}{2(2\beta+d)}}$, then the level of privacy provided is sufficient to reduce the statistical utility, leading to a deterioration of the convergence rate as a function of α. It is essential to note that it is impossible to achieve perfect privacy even in this context ($\alpha = 0$). The condition that $n\alpha^{2d} \to \infty$ must indeed be satisfied, which is the price to pay for allowing statistical inference.*

4.3.2 Lower bound for density estimation

We can now derive minimax lower bound, based on the key result gathered in Theorem 3.1 and its consequences.

Theorem 4.17. *Let $\alpha_j \in (0, \infty)$ for $j \in \{1, ..., d\}$ and let $\beta, \mathcal{L} > 0$. Then, there exists a constant $c > 0$ such that*

$$\inf_{\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\alpha}}} \inf_{\tilde{\pi}} \sup_{\pi \in \mathcal{H}(\beta, \mathcal{L})} \mathbb{E}[|\tilde{\pi}(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|^2] \geq c \Big(n \prod_{j=1}^{d} (e^{\alpha_j} - 1)^2 \Big)^{-\frac{\beta}{\beta + d}},$$

for all $n \geq 1$, $n \prod_{j=1}^{d} |e^{\alpha_j} - 1| \to \infty$. The infimum is taken over all the estimators $\tilde{\pi}$ based on the privatized vectors $\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n$ and all the non-interactive Markov kernels in $\mathcal{Q}_{\boldsymbol{\alpha}}$ guaranteeing $\boldsymbol{\alpha}$-CLDP.

Remark 4.18. *When the privacy parameters α_j are small, it is clear that the upper and lower bounds in Theorems 4.14 and 4.17 match each other. This suggests that the proposed privacy mechanism is optimal (in the minimax sense) as long as a reasonable amount of privacy is ensured for all the components (i.e., $\alpha_j < 1$ for any $j \in \{1, ..., d\}$).*

Remark 4.19. *One can compare the deterioration of the convergence rate gathered in our Theorem 4.17 (and the corresponding upper bound in Theorem 4.15) with the results in [10], which focuses on estimating the density under privacy constraints using n independent and identically distributed random variables $X_1, \ldots, X_n$. Their analysis on Besov spaces $\mathcal{B}^s_{pq}$ under mean integrated L^r-risk revealed an elbow effect that led to the optimal (in the minimax sense) convergence rate of $(n(e^\alpha - 1)^2)^{-\frac{rs}{2s+2}}$ whenever $p > \frac{r}{s+1}$ (see Equation (1.2) in [10]). Using our notation, this rate corresponds to $(n(e^\alpha - 1)^2)^{-\frac{2\beta}{2\beta+2}} = (n(e^\alpha - 1)^2)^{-\frac{\beta}{\beta+1}}$ and the condition on p reduces to $2 > \frac{2}{\beta+1}$, which is always true. Therefore, it is evident that our results match those of [10] when considering $d = 1$, but they are in general different as in the case where $\alpha_1 = ... = \alpha_d = \alpha$ the size $(e^\alpha - 1)^2$ in [10] is now replaced by $(e^\alpha - 1)^{2d}$.*

Proof of Theorem 4.17. We can assume without loss of generality that $\boldsymbol{x_0} = \boldsymbol{0}$, the general case can be deduced by translation.
The proof of the lower bound relies on the two hypothesis method, as in Section 2.3 of [36]. It consists in proposing π and π^*, densities of $\boldsymbol{X}$ and $\boldsymbol{X}^*$ and with privatized views $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$, such that the following three conditions hold true:

1. π and π^* belong to $\mathcal{H}(\beta, \mathcal{L})$,
2. $|\pi(\boldsymbol{0}) - \pi^*(\boldsymbol{0})| \geq \frac{1}{M_n}$,
3. $\exists c > 0$ such that $d_{KL}\Big(Law((\boldsymbol{Z}_i)_{i=1,\ldots,n}), Law((\boldsymbol{Z}^*_i)_{i=1,\ldots,n})\Big) < \epsilon_0 < 2$,

where $\frac{1}{M_n}$ is a calibration parameter which will be chosen later, in order to obtain the wanted convergence rate. If the constraints above are satisfied, then in the same way as in the proof of Theorem 4.8 it follows there exists $c > 0$ such that

$$\inf_{Q \in \mathcal{Q}_{\boldsymbol{\alpha}}} \inf_{\tilde{\pi}} \sup_{\pi \in \mathcal{H}(\beta, \mathcal{L})} \mathbb{E}[|\tilde{\pi}(\boldsymbol{0}) - \pi(\boldsymbol{0})|^2] \geq c(\frac{1}{M_n})^2. \tag{4.21}$$

Let us define, for any $\boldsymbol{x} \in \mathbb{R}^d$, $\pi(\boldsymbol{x}) := c_\pi e^{-\eta |x|^2}$. The constant η can be chosen as small as we want, while c_π is a normalization constant added in order to get $\int_{\mathbb{R}^d} \pi(\boldsymbol{x}) d\boldsymbol{x} = 1$. Regarding π^*, we give it as π to which we add a bump. Let $\tilde{\psi} : \mathbb{R} \to \mathbb{R}$ be a C^∞ function with support on $[-1, 1]$ and such that $\tilde{\psi}(0) = 1$, $\int_{-1}^{1} \tilde{\psi}(z) dz = 0$. Then, we set $\pi^*(\boldsymbol{x}) := \pi(\boldsymbol{x}) + \frac{1}{M_n} \prod_{l=1}^{d} \tilde{\psi}(\frac{x^l}{h_n}) =: \pi(\boldsymbol{x}) + \frac{1}{M_n} \psi_{h_n}(\boldsymbol{x})$. As $\frac{1}{M_n}$, h_n will be calibrated later. The two calibration constants satisfy $M_n \to \infty$ for $n \to \infty$ and $h_n \to 0$ for $n \to \infty$.

It is easy to check Condition 1 holds true. Indeed, we can choose η small enough to obtain

$\pi \in \mathcal{H}(\beta, \mathcal{L})$ and a similar reasoning ensures also that $\pi^* \in \mathcal{H}(\beta, \mathcal{L})$. However, $\left\|\psi_{h_n}^{(k)}\right\|_\infty \le \frac{c}{h_n^k}$ and so in the k-derivative of π^* an extra $\frac{1}{M_n h_n^k}$ appears. It implies we have to ask the existence of some constant $c > 0$ such that $\frac{1}{M_n h_n^k} < c$ for any $k \in \{0, ..., \lfloor \beta \rfloor\}$, in order to obtain $\pi^* \in \mathcal{H}(\beta, \mathcal{L})$. Thus, for some $c > 0$ it naturally arises the condition $\frac{1}{M_n h_n^\beta} < c$.
Concerning Condition 2, by construction and from the properties of the function $\tilde{\psi}$ it is

$$|\pi(\mathbf{0}) - \pi^*(\mathbf{0})| = |\frac{1}{M_n}\psi_{h_n}(\mathbf{0})| = \frac{1}{M_n}\prod_{l=1}^d |\tilde{\psi}(0)| = \frac{1}{M_n}.$$

We are left to prove Condition 3. We observe that, for any $k \in \{1, ..., d-1\}$, it is $Law(X^{i_1}, ..., X^{i_k}) = Law(X^{*,i_1}, ..., X^{*,i_k})$. Indeed, the law of $(X^{*,i_1}, ..., X^{*,i_k})$ is given by

$$\begin{aligned}
&\int_{\mathbb{R}^{d-k}} \Big(\pi(x^1, ..., x^d) + \frac{1}{M_n}\psi_{h_n}(x^1, ..., x^d)\Big) \prod_{j:\, j \notin \{i_1, ..., i_k\}} dx^j \\
&= \int_{\mathbb{R}^{d-k}} \pi(x^1, ..., x^d) \prod_{j:\, j \notin \{i_1, ..., i_k\}} dx^j + \frac{1}{M_n}\int_{\mathbb{R}^{d-k}} \prod_{l=1}^d \tilde{\psi}(\frac{x^l}{h_n}) \prod_{j:\, j \notin \{i_1, ..., i_k\}} dx^j \\
&= \int_{\mathbb{R}^{d-k}} \pi(x^1, ..., x^d) \prod_{j:\, j \notin \{i_1, ..., i_k\}} dx^j = Law(X^{i_1}, ..., X^{i_k}),
\end{aligned}$$

where we have used that the integrals of $\tilde{\psi}$ are 0 by construction.
Thus, we can use Corollary 3.8. It yields

$$d_{KL}\Big(Law((\boldsymbol{Z}_i)_{i=1,...,n}), Law((\boldsymbol{Z}_i^*)_{i=1,...,n})\Big) \le n \times \prod_{j=1}^d |e^{\alpha_j} - 1|^2 \Big(d_{TV}(Law(\boldsymbol{X}), Law(\boldsymbol{X}^*))\Big)^2. \tag{4.22}$$

To conclude, we observe it is

$$\Big(d_{TV}(Law(\boldsymbol{X}), Law(\boldsymbol{X}^*))\Big)^2 \le \Big(\int_{\mathbb{R}^d} \frac{1}{M_n}\psi_{h_n}(x^1, ..., x^d)dx^1, ..., dx^d\Big)^2 \le c\frac{h_n^{2d}}{M_n^2}. \tag{4.23}$$

From (4.22) and (4.23) we get there exists some constant $c_k > 0$ such that

$$d_{KL}\Big(Law((\boldsymbol{Z}_i)_{i=1,...,n}), Law((\boldsymbol{Z}_i^*)_{i=1,...,n})\Big) \le c_k n \prod_{j=1}^d |e^{\alpha_j} - 1|^2 \frac{h_n^{2d}}{M_n^2}.$$

Hence, Condition 3 holds true up to say that $c_k n \prod_{j=1}^d |e^{\alpha_j} - 1|^2 \frac{h_n^{2d}}{M_n^2}$ is bounded by some $\epsilon_0 < 2$. The constraint given by Condition 1 leads us to the choice $h_n = (\frac{1}{M_n})^{\frac{1}{\beta}}$, which entails $c_k n \prod_{j=1}^d |e^{\alpha_j} - 1|^2 \Big(\frac{1}{M_n}\Big)^{\frac{2d}{\beta}+2} < \epsilon_0$. It holds true if and only if $\Big(\frac{1}{M_n}\Big)^{\frac{2d+2\beta}{\beta}} < \frac{\epsilon_0}{c_k n \prod_{j=1}^d |e^{\alpha_j} - 1|^2}$. We therefore choose $\frac{1}{M_n} = \Big(\frac{\epsilon_0}{c_k n \prod_{j=1}^d |e^{\alpha_j} - 1|^2}\Big)^{\frac{\beta}{2(d+\beta)}}$.
Equation (4.21) concludes then the proof.

□

4.3.3 Adaptive density estimation

As seen in previous subsection, the proposed procedure leads us to the choice of a bandwidth which depends on the regularity β, that is in general unknown. This motivates a data-driven

procedure for the choice of h.
We introduce the set of candidate bandwidths

$$H_n := \left\{ h \in (0,1] : \text{ such that } \frac{1}{h} = \frac{n}{2^r} \text{ for some } r \in \{1, ..., \lfloor \log_2(n) \rfloor\} \right\}. \tag{4.24}$$

In a similar way as in Section 4.2.3 we introduce, for $j \in \{1, ..., d\}$, some parameters $\beta_n^j > 0$ that will be better specified later. For any $i \in \{1, ..., n\}$ and $j \in \{1, ..., d\}$ the independent variables $\mathcal{E}_i^{j,h}$ are distributed as Laplace random variables with law $\mathcal{L}(\frac{2\kappa}{h\beta_n^j})$ for all $h \in H_n$, where κ is defined in (4.18). We therefore define the privatized data $\boldsymbol{Z}_i = (Z_i^1, ..., Z_i^d) \in \mathbb{R}^{H_n} \times ... \times \mathbb{R}^{H_n}$ by $Z_i^j = (Z_i^{j,h})_{h \in H_n}$;

$$Z_i^{j,h} := \frac{1}{h} K(\frac{X_i^j - x_0^j}{h}) + \mathcal{E}_i^{j,h} \tag{4.25}$$

for $h \in H_n$, $i \in \{1, ..., n\}$ and $j \in \{1, ..., d\}$. The set of potential estimators is defined accordingly:

$$\mathcal{F}(H_n) := \left\{ \hat{\pi}_h^z(\boldsymbol{x_0}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d Z_i^{j,h} \quad h \in H_n \right\}.$$

By following closely the proof of Lemma 4.13 it is easy to check the following lemma, whose proof is in the appendix.

Lemma 4.20. *Assume that $\beta_n^j = \frac{\alpha_j}{\lfloor \log_2 n \rfloor}$ for any $j \in \{1, ..., d\}$. Then, the privatized variables described in (4.25) are $\boldsymbol{\alpha}$-local differential private views of the original X_i^j.*

As our adaptive procedure is based on Goldenshluger-Lepski method, we want to introduce an auxiliary estimator. For any $h, \eta \in H_n$ we set

$$\hat{\pi}_{h,\eta}^z(\boldsymbol{x_0}) := \hat{\pi}_{h \wedge \eta}^z(\boldsymbol{x_0}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \frac{1}{h \wedge \eta} K(\frac{X_i^j - x_0^j}{h \wedge \eta}) + \mathcal{E}_i^{j, h \wedge \eta}.$$

Clearly it is $\hat{\pi}_{h,\eta}^z(\boldsymbol{x_0}) = \hat{\pi}_{\eta,h}^z(\boldsymbol{x_0})$. In the proof of Theorem 4.14 given in the appendix, we obtain the bound (A.29) for the variance of the kernel estimator. This leads us to the introduction of the penalization term $\mathbb{V}_h := a_n \frac{1}{nh^{2d}} \frac{1}{\prod_{j=1}^d (\beta_n^j)^2}$, for $a_n \geq 1$ some sequence tending slowly to ∞, which will be specified in Theorem 4.21 below. We also define, for any $h \in H_n$, $\mathbb{B}_h := \sup_{\eta \in H_n} \left\{ (|\hat{\pi}_{h,\eta}^z(\boldsymbol{x_0}) - \hat{\pi}_\eta^z(\boldsymbol{x_0})|^2 - \mathbb{V}_\eta)_+ \right\}$. Heuristically, $\mathbb{V}_h$ plays the role of the variance while $\mathbb{B}_h$ plays the role of the bias; the chosen bandwidth for the adaptive procedure is the one that minimizes their sum, i.e. $\hat{h} := argmin_{h \in H_n} \{\mathbb{B}_h + \mathbb{V}_h\}$. The associated adaptive estimator is $\hat{\pi}_{\hat{h}}^z(\boldsymbol{x_0})$, for which the following theorem holds true.

Theorem 4.21. *Assume that $\pi \in \mathcal{H}(\beta, \mathcal{L})$ for some β and $\mathcal{L} \geq 1$. Moreover, $\beta_n^j = \frac{\alpha_j}{\lfloor \log_2 n \rfloor}$ for any $j \in \{1, ..., d\}$ and $a_n = c_0 \log n$ for some $c_0 > 0$. If c_0 is large enough, there exist $c > 0$ and $\bar{c} > 0$ such that*

$$\mathbb{E}[(\hat{\pi}_{\hat{h}}^z(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0}))^2] \leq c(\frac{\log n^{1+2d}}{n \prod_{j=1}^d \alpha_j^2})^{\frac{\beta}{\beta+d}} + \frac{c}{n^{\bar{c}} \prod_{j=1}^d \alpha_j^2}$$

for all $n \geq 1$, $\alpha_j \leq 1$ and $\frac{n \prod_{j=1}^d \alpha_j^2}{\log n^{1+2d}} \geq 1$. Moreover, the constant $\bar{c}$ can be chosen arbitrarily large, taking the constant c_0 large enough.

Remark 4.22. *For $d = 2$ the conditions in the theorem above reduce to the ones in Theorem 4.11. This is indeed the generalization to the d-dimensional case of the adaptive procedure proposed in Section 4.2.3, for the data-driven choice of the two truncation levels $T^{(1)}$ and $T^{(2)}$.*

Remark 4.23. *Just as in the case of estimating the covariance, the adaptive version of the density estimation algorithm has a slower convergence rate than the one presented in Theorem 4.14, by a logarithmic factor. Specifically, in the case of estimating the covariance, the loss was a factor of* $(\log n)^5$ *(as shown in Theorem 4.11), while in the current context, we lose a factor of* $(\log n)^{1+2d}$*. This observation is consistent with the earlier remark.*

The proof of the data-driven selection of the bandwidth as proposed in Theorem 4.21 can be found in the appendix.

References

[1] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308-318).

[2] Acharya, J., Sun, Z., & Zhang, H. (2021, July). Robust testing and estimation under manipulation attacks. In International Conference on Machine Learning (pp. 43-53). PMLR.

[3] Apple Differential Privacy Team. Learning with privacy at scale, 2017. Available at https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

[4] Berrett, T., & Butucea, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. Advances in Neural Information Processing Systems, 33, 3164-3173.

[5] Berrett, T. B., Györfi, L., & Walk, H. (2021). Strongly universally consistent nonparametric regression and classification with privatized data.

[6] Berrett, T., & Yu, Y. (2021). Locally private online change point detection. Advances in Neural Information Processing Systems, 34, 3425-3437.

[7] Bosq, D., & Blanke, D. (2008). Inference and prediction in large dimensions. John Wiley & Sons.

[8] Boucheron, S., Lugosi, G., & Massart, P. (2013). Concentration inequalities: A nonasymptotic theory of independence. Oxford university press.

[9] Brown, L. D., & Low, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. The annals of Statistics, 24(6), 2524-2535.

[10] Butucea, C., Dubois, A., Kroll, M., & Saumard, A. (2020). Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids.

[11] Butucea, C., Rohde, A., & Steinberger, L. (2020). Interactive versus non-interactive locally differentially private estimation: Two elbows for the quadratic functional. arXiv preprint arXiv:2003.04773.

[12] Ding, B., Kulkarni, J., & Yekhanin, S. (2017). Collecting telemetry data privately. Advances in Neural Information Processing Systems, 30.

[13] Donoho, D. L., & Liu, R. C. (1991). Geometrizing rates of convergence, II. The Annals of Statistics, 633-667.

[14] Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013, October). Local privacy and statistical minimax rates. In 2013 IEEE 54th Annual Symposium on Foundations of Computer Science (pp. 429-438). IEEE.

[15] Duchi, J., Wainwright, M. J., & Jordan, M. I. (2013). Local privacy and minimax bounds: Sharp rates for probability estimation. Advances in Neural Information Processing Systems, 26.

[16] Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy, data processing inequalities, and statistical minimax rates. arXiv preprint arXiv:1302.3203.

[17] Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2018). Minimax optimal procedures for locally private estimation. Journal of the American Statistical Association, 113(521), 182-201.

[18] Dwork, C., & Nissim, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In Advances in Cryptology–CRYPTO 2004: 24th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 2004. Proceedings 24 (pp. 528-544). Springer Berlin Heidelberg.

[19] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3 (pp. 265-284). Springer Berlin Heidelberg.

[20] Dwork, C. (2008). Differential privacy: A survey of results. In Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings 5 (pp. 1-19). Springer Berlin Heidelberg.

[21] Erlingsson, U., Pihur, V., & Korolova, A. (2014, November). Rappor: Randomized aggregatable privacy-preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (pp. 1054-1067).

[22] Evfimievski, A., Gehrke, J., & Srikant, R. (2003, June). Limiting privacy breaches in privacy preserving data mining. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (pp. 211-222).

[23] Forti, M. (2021). The deployment of artificial intelligence tools in the health sector: privacy concerns and regulatory answers within the GDPR. Eur. J. Legal Stud., 13, 29.

[24] Gray, R. M. (1990). Entropy and information. Entropy and information theory, 21-55.

[25] Györfi, L., & Kroll, M. (2022). On rate optimal private regression under local differential privacy. arXiv preprint arXiv:2206.00114. Chicago

[26] Joseph, M., Mao, J., Neel, S., & Roth, A. (2019, November). The role of interactivity in local differential privacy. In 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS) (pp. 94-105). IEEE.

[27] Kairouz, P., Oh, S., & Viswanath, P. (2014). Extremal mechanisms for local differential privacy. Advances in neural information processing systems, 27.

[28] Kroll, M. (2020). Adaptive spectral density estimation by model selection under local differential privacy. arXiv preprint arXiv:2010.04218.

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

[30] Li, M., Berrett, T., & Yu, Y. (2022a). Network change point localisation under local differential privacy. Advances in Neural Information Processing Systems, 35, 15013-15026.

[31] Li, M., Berrett, T. B., & Yu, Y. (2022b). On robustness and local differential privacy. arXiv preprint arXiv:2201.00751.

[32] Polyak, B. T., & Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. SIAM journal on control and optimization, 30(4), 838-855.

[33] Rohde, A., & Steinberger, L. (2020). Geometrizing rates of convergence under local differential privacy constraints.

[34] Sart, M. (2021). Density estimation under local differential privacy and Hellinger loss.

[35] Tang, J., Korolova, A., Bai, X., Wang, X., & Wang, X. (2017). Privacy loss in apple's implementation of differential privacy on macos 10.12. arXiv preprint arXiv:1709.02753.

[36] Tsybakov, A. B. (2004). Introduction to nonparametric estimation, 2009. URL https://doi.org/10.1007/b13794. Revised and extended from the, 9(10).

[37] Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309), 63-69.

[38] Wasserman, L., & Zhou, S. (2010). A statistical framework for differential privacy. Journal of the American Statistical Association, 105(489), 375-389.

A Appendix

In this section we provide all the technical proofs. We will start by proving Proposition 3.5. After that, we will give the proofs for the locally private estimation of the covariance and of the density, as presented in Sections 4.2 and 4.3, respectively.

A.1 Proof of Proposition 3.5

Proof. Proposition 3.5 is proven by recurrence. The proof is split in two steps.

Step 1
The aim of this step is to show that

$$l[z^1, ..., z^d] \leq l[\emptyset, z^2, ..., z^d] + l[dx^1, z^2, ..., z^d].$$

From the definition of l it is

$$l[z^1, \dots, z^d] = |q(z^1, \dots, z^d) - \tilde{q}(z^1, \dots, z^d)|.$$

As observed in (3.4), we have

$$q(z^1, \dots, z^d) = \int_{\mathcal{X}^1} q^1(z^1|x^1) q(dx^1, z^2, \dots, z^d)$$

and, in the same way,

$$\tilde{q}(z^1, \dots, z^d) = \int_{\mathcal{X}^1} q^1(z^1|x^1) \tilde{q}(dx^1, z^2, \dots, z^d).$$

It follows that

$$l[z^1, \dots, z^d] = \Big| \int_{\mathcal{X}^1} q^1(z^1|x^1)[q(dx^1, z^2, \dots, z^d) - \tilde{q}(dx^1, z^2, \dots, z^d)] \Big|. \tag{A.1}$$

We split the signed measure $q(dx^1, z^2, \dots, z^d) - \tilde{q}(dx^1, z^2, \dots, z^d)$ into its positive and negative part:

$$\begin{aligned}
l[z^1, ..., z^d] &= \Big| \int_{\mathcal{X}^1} q^1(z^1|x^1)[q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)]_+ \\
&\quad + \int_{\mathcal{X}^1} q^1(z^1|x^1)[q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)]_- \Big| \\
&\leq \Big| \sup_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) \int_{\mathcal{X}^1} [q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)]_+ \\
&\quad + \inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) \int_{\mathcal{X}^1} [q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)]_- \Big|,
\end{aligned}$$

using the notation $[A]_+$ and $[A]_-$ for the positive and negative part of A, respectively. We now have

$$\begin{aligned}
&[q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)]_- \\
&= [q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)] - [q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)]_+.
\end{aligned}$$

We remark that, if we integrate the last equation with respect to $x^1 \in \mathcal{X}^1$, the middle term gives a contribution when $d \geq 2$, since $\int_{\mathcal{X}^1} q(dx_1, z_2, ..., z_d) = q(\emptyset, z_2, ..., z_d)$ is not simply 1 as in the mono-dimensional case. Hence, it provides

$$\begin{aligned}
l[z^1, ..., z^d] \leq \Big| \Big(\sup_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) - \inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) \Big) \int_{\mathcal{X}^1} [q(x^1, z^2, ..., z^d) - \tilde{q}(x^1, z^2, ..., z^d)]_+ | \\
+ \inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) | q(\emptyset, z^2, ..., z^d) - \tilde{q}(\emptyset, z^2, ..., z^d) \Big|.
\end{aligned} \tag{A.2}$$

We now observe that

$$\begin{aligned}
\sup_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) - \inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) &= \inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1) \left(\frac{\sup_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1)}{\inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1)} - 1 \right) \\
&\leq \inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1)(e^{\alpha_1} - 1) = q^1(z^1|x^1_*)(e^{\alpha_1} - 1),
\end{aligned} \tag{A.3}$$

where we used (2.4) and the definition $q^1(z^1|x^1_*) = \inf_{x^1 \in \mathcal{X}^1} q^1(z^1|x^1)$. We replace it in (A.2), which entails

$$\begin{aligned}
l[z^1, ..., z^d] &\leq q^1(z^1|x^1_*)(e^{\alpha_1} - 1) \int_{\mathcal{X}^1} |q(dx^1, z^2, ..., z^d) - \tilde{q}(dx^1, z^2, ..., z^d)| \\
&\quad + q^1(z^1|x^1_*)|q(\emptyset, z^2, ..., z^d) - \tilde{q}(\emptyset, z^2, ..., z^d)| \\
&= l[dx^1, z^2, ..., z^d] + l[\emptyset, z^2, ..., z^d],
\end{aligned}$$

which concludes the proof of Step 1.

Step 2

Assume now that $\zeta^i \in \{\emptyset, dx^i\}$ for any i smaller than some j. Then, this step is devoted to the proof of

$$l[\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d] \leq l[\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d] + l[\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d].$$

By definition it is indeed

$$l[\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d] = \prod_{i=1}^{j-1} q^i(z^i|x_*^i) \prod_{i<j:\zeta^i=dx^i} (e^{\alpha_i} - 1) \tag{A.4}$$
$$\times \int_{\prod_{i<j:\zeta^i=dx^i} \mathcal{X}^i} |q(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d)|$$

We observe that, similarly to (3.4), it is

$$q(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d) = \int_{x^j \in \mathcal{X}^j} q^j(z^j|x^j) q(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d)$$

and also,

$$\tilde{q}(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d) = \int_{x^j \in \mathcal{X}^j} q^j(z_j|x_j) \tilde{q}(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d).$$

Then,

$$|q(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d)| \tag{A.5}$$
$$= \Big| \int_{x^j \in \mathcal{X}^j} q^j(z^j|x^j) [q(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d)] \Big|.$$

Acting as above (A.2), remarking also that

$$\int_{x^j \in \mathcal{X}^j} q(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d) = q(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d),$$

it follows that the quantity (A.5) is upper bounded by

$$\Big|(\sup_{x^j \in \mathcal{X}^j} q^j(z^j|x^j) - \inf_{x^j \in \mathcal{X}^j} q^j(z^j|x^j)) \int_{q^j \in \mathcal{X}^j} [q(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d)]_+ \Big|$$
$$+ \inf_{x^j \in \mathcal{X}^j} q^j(z^j|x^j) |q(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d)|.$$

Thus, Equation (A.3) with q^j in place of q^1, entails

$$|q(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d)|$$
$$\leq (e^{\alpha_j} - 1) q^j(z^j|x_*^j) \int_{x_j \in \mathcal{X}^j} |q(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d)|$$
$$+ q^j(z^j|x_*^j) |q(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d)|.$$

We replace it in (A.4), obtaining

$$\begin{aligned}
&l[\zeta^1, ..., \zeta^{j-1}, z^j, ..., z^d] \\
&\leq \prod_{i=1}^{j-1} q^i(z^i|x^i_*) \prod_{i<j:\zeta^i=dx^i} (e^{\alpha_i}-1) \int_{\prod_{i<j:\zeta^i=dx^i} \mathcal{X}^i} \Big[(e^{\alpha_j}-1) q^j(z^j|x^j_*) \\
&\times \int_{x^j \in \mathcal{X}^j} |q(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d)| \\
&+ q^j(z^j|x^j_*)|q(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d)|\Big] \\
&= \prod_{i=1}^{j} q^i(z^i|x^i_*) \Big(\prod_{i<j:\zeta^i=dx^i} (e^{\alpha_i}-1) \Big) (e^{\alpha_j}-1) \\
&\times \int_{\left(\prod_{i<j:\zeta^i=dx^i} \mathcal{X}^i\right) \times \mathcal{X}^j} |q(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d)| \\
&+ \prod_{i=1}^{j} q^i(z^i|x^i_*) \prod_{i<j:\zeta^i=dx^i} (e^{\alpha_i}-1) \\
&\times \int_{\prod_{i<j:\zeta^i=dx^i} \mathcal{X}^i} |q(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d) - \tilde{q}(\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d)| \\
&= l[\zeta^1, ..., \zeta^{j-1}, dx^j, z^{j+1}, ..., z^d] + l[\zeta^1, ..., \zeta^{j-1}, \emptyset, z^{j+1}, ..., z^d].
\end{aligned}$$

Thus, the proof of Step 2 is completed.

To conclude, we want to show that Steps 1 and 2 imply the stated result of the proposition. From the two steps above we have, by recurrence,

$$\begin{aligned}
l[z^1, ..., z^d] &\leq l[\emptyset, z^2 ..., z^d] + l[dx^1, z^2 ..., z^d] \\
&\leq l[\emptyset, \emptyset, z^3 ..., z^d] + l[\emptyset, dx^2, z^3 ..., z^d] + l[dx^1, \emptyset, z^3 ..., z^d] + l[dx^1, dx^2, z^3 ..., z^d] \\
&\leq \sum_{(\zeta^1, ..., \zeta^d) \in \{\emptyset, dx^i\}^d} l[\zeta^1, ..., \zeta^d],
\end{aligned}$$

as we wanted. ☐

A.2 Locally private covariance estimation

In this subsection we prove all the technical results related to the estimation of the covariance under local differential privacy constraints.

A.2.1 Proof of Theorem 4.5

Proof. • We first prove the result for $\hat{\gamma}_n$ as stated in (4.9). This is based on a bias-variance decomposition for the L^2 risk. Let us denote the bias term by

$$b_{T^{(1)},T^{(2)}} := \mathbb{E}[Z_i^1 Z_i^2] - \mathbb{E}[X_i^1 X_i^2] = \mathbb{E}[Z^1 Z^2] - \mathbb{E}[X^1 X^2]. \tag{A.6}$$

We first state as lemma a control on the magnitude of this bias term, its proof can be found at the end of this section.

Lemma A.1. *We have*

$$\left|b_{T^{(1)},T^{(2)}}\right| \le \frac{3}{2}(T^{(1)})^{-k_1(1-\frac{2}{k})}\|X^1\|_{k_1}^{1+k_1(1-\frac{2}{k})}\|X^2\|_{k_2} + \frac{3}{2}(T^{(2)})^{-k_2(1-\frac{2}{k})}\|X^1\|_{k_1}\|X^2\|_{k_2}^{1+k_2(1-\frac{2}{k})}. \quad \text{(A.7)}$$

To get (4.9), we write the bias variance decomposition, $\mathbb{E}\left[(\hat{\gamma}_n-\gamma)^2\right] = (b_{T^{(1)},T^{(2)}})^2 + \mathrm{var}(\hat{\gamma}_n)$. By independence and Cauchy-Schwarz's inequality,

$$\mathrm{var}(\hat{\gamma}_n) = n^{-2}\sum_{i=1}^{n}\mathrm{var}(Z_i^1Z_i^2) \le n^{-1}\mathbb{E}(|Z^1Z^2|^2) \le n^{-1}\mathbb{E}(|Z^1|^4)^{\frac{1}{2}}\mathbb{E}(|Z^2|^4)^{\frac{1}{2}}. \quad \text{(A.8)}$$

We have for $j=1,2$,

$$\mathbb{E}(|Z^j|^4) = \mathbb{E}(|\left[X^j\right]_{T^{(j)}} + \mathcal{E}^j|^4) \le 8\mathbb{E}[\left[X^j\right]_{T^{(j)}}^4] + 8\mathbb{E}[|\mathcal{E}^j|^4] \le 8|T^{(j)}|^4\left(1+\left(\frac{2}{\alpha_j}\right)^4\mathbb{E}\left[|\mathcal{L}(1)|^4\right]\right),$$

as $\mathcal{E}^j$ is equal in law to a $\frac{2T^{(j)}}{\alpha_j}\times\mathcal{L}(1)$ variable. We deduce $\mathbb{E}(|Z^j|^4) \le 8|T^{(j)}|^4(1+\frac{2^4 4!}{\alpha_j^4}) \le C|T^{(j)}|^4/\alpha_j^4$ for some universal constant C, using $\alpha\le 1$. From (A.8), it entails

$$\mathrm{var}(\hat{\gamma}_n) \le Cn^{-1}\frac{|T^{(1)}|^2|T^{(2)}|^2}{\alpha_1^2\alpha_2^2}, \quad \text{(A.9)}$$

for some universal constant $C>0$. In turn, recalling Lemma A.1, we get

$$\mathbb{E}\left[(\hat{\gamma}_n-\gamma)^2\right] \le c\left((T^{(1)})^{-2k_1(1-\frac{2}{k})} + (T^{(2)})^{-2k_2(1-\frac{2}{k})} + n^{-1}\frac{|T^{(1)}|^2|T^{(2)}|^2}{\alpha_1^2\alpha_2^2}\right). \quad \text{(A.10)}$$

The calibration given in the statement of the theorem $(T^{(1)})^{k_1} = (T^{(2)})^{k_2} = (n\alpha_1^2\alpha_2^2)^{1/2}$ is such that the three terms in the right hand side of (A.10) equilibrates and $\mathbb{E}\left[(\hat{\gamma}_n-\gamma)^2\right] \le c(n\alpha_1^2\alpha_2^2)^{-(1-\frac{2}{k})} = c(n\alpha_1^2\alpha_2^2)^{-\frac{k-2}{k}}$, which is (4.9).

• We now prove the result (4.10) on $\hat{\theta}_n$. An additional error appears in the estimation of θ due to the inference of both means m_1, m_2. We will see that these additional errors are at most of the same magnitude as the estimation error of the cross term $\gamma = E[X^1X^2]$. We need to recall the bias-variance decomposition given in the proof of Corollary 1 in [17] :

$$\mathbb{E}\left[(\hat{m}_n^{(j)} - m^{(j)})^2\right] \le c\left((T^{(j)})^{-(k_j-1)} + n^{-1}\frac{|T^{(j)}|^2}{\alpha_j^2}\right), \quad \text{for } j\in\{1,2\}, \quad \text{(A.11)}$$

where the constant c does not depend on $n\ge 1$, $\alpha_j\in(0,1]$. Let us emphasize that the optimal trade-off in (A.11) is given by the choices of $T^{(j)}$ appearing in the statement of Theorem 4.4. However, our choice $T^{(j)} = (n\alpha_1^2\alpha_2^2)^{\frac{1}{2k_j}}$ is tailored to get the optimal trade-off while estimating γ with $\alpha_1,\alpha_2<1$, and yields to a $\mathbf{L}^2$-risk for the estimation of the means which is suboptimal in α_j. Replacing in (A.11), we get,

$$\begin{aligned}\mathbb{E}\left[(\hat{m}_n^{(j)} - m^{(j)})^2\right] &\le c\left((n\alpha_1^2\alpha_2^2)^{-\frac{k_j-1}{k_j}} + n^{-1}\frac{(n\alpha_1^2\alpha_2^2)^2}{\alpha_j^2}\right) \le c(n\alpha_1^2\alpha_2^2)^{-\frac{k_j-1}{k_j}}\left(1+\alpha_{3-j}^2\right)\\ &\le c(n\alpha_1^2\alpha_2^2)^{-\frac{k_j-1}{k_j}} \quad \text{for } j\in\{1,2\}. \quad \text{(A.12)}\end{aligned}$$

Now, we split

$$\hat{\theta}_n - \theta = \hat{\gamma}_n - \gamma - \left[(\hat{m}_n^{(1)} - m^{(1)})m^{(2)} + \hat{m}_n^{(1)}(\hat{m}_n^{(2)} - m^{(2)})\right], \tag{A.13}$$

and denote by $\sum_{l=1}^{2} e_n^{(l)}$ the two terms in the bracket of the above equation. By (A.12), we have,

$$\begin{aligned}\mathbb{E}\left[|e_n^{(1)}|^2\right] = \mathbb{E}\left[(\hat{m}_n^{(1)} - m^{(1)})^2\right]|m^{(2)}|^2 \le c(n\alpha_1^2\alpha_2^2)^{-\frac{k_1-1}{k_1}} \\ \le c(n\alpha_1^2\alpha_2^2)^{-[1-\frac{1}{k_1}-\frac{1}{k_2}]} = c(n\alpha_1^2\alpha_2^2)^{-\frac{\overline{k}-2}{\overline{k}}} \end{aligned} \tag{A.14}$$

where we used $n\alpha_1^2\alpha_2^2 \ge 1$ and $X^2 \in \mathbf{L}^{k_2} \subset \mathbf{L}^1$. The control of the second term necessitates more care. Using (4.6), we have, recalling the definition $\hat{m}_n^{(1)} = n^{-1}\sum_{i=1}^{n} Z_i^1$, as given in (4.7),

$$\begin{aligned}\mathbb{E}\left[|e_n^{(2)}|^2\right] &= \mathbb{E}\left[|\hat{m}_n^{(1)}|^2(\hat{m}_n^{(2)} - m^{(2)})^2\right] \\ &\le 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left[X_i^1\right]_{T^{(1)}}\right)^2(\hat{m}_n^{(2)} - m^{(2)})^2\right] + 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{E}_i^1\right)^2(\hat{m}_n^{(2)} - m^{(2)})^2\right] \\ &\le 2|T^{(1)}|^2\mathbb{E}\left[(\hat{m}_n^{(2)} - m^{(2)})^2\right] + 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{E}_i^1\right)^2\right]\mathbb{E}\left[(\hat{m}_n^{(2)} - m^{(2)})^2\right],\end{aligned}$$

where we used $\left|\left[X_i^1\right]_{T^{(1)}}\right| \le T^{(1)}$ for the first expectation, and the independence of $(\mathcal{E}_i^1)_i$ and $\hat{m}_n^{(2)}$ for the second one. Recalling $T^{(1)} = (n\alpha_1^2\alpha_2^2)^{\frac{1}{2k_1}}$ and that $\mathcal{E}_i^1$ are iid centered variable with variance $\frac{8|T^{(1)}|^2}{\alpha_1^2}$, we get

$$\begin{aligned}\mathbb{E}\left[|e_n^{(2)}|^2\right] &\le c\left(|T^{(1)}|^2 + \frac{|T^{(1)}|^2}{n\alpha_1^2}\right)\mathbb{E}\left[(\hat{m}_n^{(2)} - m^{(2)})^2\right] \\ &\le c\left((n\alpha_1^2\alpha_2^2)^{\frac{1}{k_1}} + \alpha_1^2(n\alpha_1^2\alpha_2^2)^{\frac{1}{k_1}-1}\right)\mathbb{E}\left[(\hat{m}_n^{(2)} - m^{(2)})^2\right] \\ &\le c\left((n\alpha_1^2\alpha_2^2)^{\frac{1}{k_1}} + \alpha_1^2(\alpha_1^2\alpha_2^2)^{\frac{1}{k_1}-1}\right)(n\alpha_1^2\alpha_2^2)^{-\frac{k_2-1}{k_2}} \\ &\le c(n\alpha_1^2\alpha_2^2)^{\frac{1}{k_1}}(n\alpha_1^2\alpha_2^2)^{-\frac{k_2-1}{k_2}} = c(n\alpha_1^2\alpha_2^2)^{-\frac{\overline{k}-2}{\overline{k}}},\end{aligned} \tag{A.15}$$

where we used $\alpha_1^2 \le 1$, $n\alpha_1^2\alpha_2^2 \ge 1$.
Collecting (4.9) with (A.13)– (A.15), we deduce (4.10). □

Proof of Lemma A.1. From (4.6) and (A.6), we have

$$\begin{aligned}b_{T^{(1)},T^{(2)}} &= \mathbb{E}\left[\left(\left[X^1\right]_{T^{(1)}} + \mathcal{E}^1\right)\left(\left[X^2\right]_{T^{(2)}} + \mathcal{E}^2\right)\right] - \mathbb{E}\left[X^1X^2\right] \\ &= \mathbb{E}\left[\left[X^1\right]_{T^{(1)}}\left[X^2\right]_{T^{(2)}}\right] - \mathbb{E}\left[X^1X^2\right],\end{aligned} \tag{A.16}$$

where we used that $\mathcal{E}^1$, $\mathcal{E}^2$ are centered and independent of X^1, X^2. We set $\Delta^{(j)} = \left[X^j\right]_{T^{(j)}} - X^j$ for $j = 1, 2$ and get, adding and removing $\mathbb{E}\left[\left[X^1\right]_{T^{(1)}}X^2\right]$ and $E\left[\Delta^{(1)}\left[X^2\right]_{T^{(2)}}\right]$,

$$b_{T^{(1)},T^{(2)}} = \mathbb{E}\left[\Delta^{(1)}\left[X^2\right]_{T^{(2)}}\right] + \mathbb{E}\left[\left[X^1\right]_{T^{(1)}}\Delta^{(2)}\right] + \mathbb{E}\left[\Delta^{(1)}\Delta^{(2)}\right] =: \sum_{l=1}^{3} b_{T^{(1)},T^{(2)}}^{(l)}.$$

We assess the magnitude of each term in the sum above. Let $r \geq 1$ be such that $\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{r} = 1$. Then, remarking that $\Delta^{(1)}$ is different from 0 for $|X^1| \geq T^{(1)}$, we have

$$\begin{aligned}\left|b^{(1)}_{T^{(1)},T^{(2)}}\right| &\leq \mathbb{E}\left[\left|\Delta^{(1)}\right|\mathbb{1}_{\{|X^1|\geq T^{(1)}\}}|\left[X^2\right]_{T^{(2)}}|\right] \\ &\leq \|\Delta^{(1)}\|_{k_1}\|X^2\|_{k_2}\mathbb{P}\left(|X^1| \geq T^{(1)}\right)^{\frac{1}{r}}, \\ &\leq \|\Delta^{(1)}\|_{k_1}\|X^2\|_{k_2}\left(\frac{\mathbb{E}\left(|X^1|^{k_1}\right)}{(T^{(1)})^{k_1}}\right)^{\frac{1}{r}},\end{aligned}$$

where we used Hölder's inequality in the second line and Markov's inequality in the last line. Using $|\Delta^{(1)}| \leq |X^1|$, we deduce,

$$\left|b^{(1)}_{T^{(1)},T^{(2)}}\right| \leq \|X^1\|_{k_1}^{1+\frac{k_1}{r}}\|X^2\|_{k_2}\left(T^{(1)}\right)^{-\frac{k_1}{r}} \leq \|X^1\|_{k_1}^{1+k_1(1-\frac{2}{\bar{k}})}\|X^2\|_{k_2}\left(T^{(1)}\right)^{-k_1(1-\frac{2}{\bar{k}})} \tag{A.17}$$

where we used $\frac{1}{r} = 1 - \frac{1}{k_1} - \frac{1}{k_2} = 1 - \frac{2}{\bar{k}}$. Similarly, we get

$$\left|b^{(2)}_{T^{(1)},T^{(2)}}\right| \leq \|X^1\|_{k_1}\|X^2\|_{k_2}^{1+k_2(1-\frac{2}{\bar{k}})}\left(T^{(2)}\right)^{-k_2(1-\frac{2}{\bar{k}})}. \tag{A.18}$$

For the term $b^{(3)}_{T^{(1)},T^{(2)}}$, we write

$$\begin{aligned}\left|b^{(3)}_{T^{(1)},T^{(2)}}\right| &\leq \mathbb{E}\left[|\Delta^{(1)}\Delta^{(2)}|\right] \leq \mathbb{E}\left[|X^1||X^2|\mathbb{1}_{\{|X^1|\geq T^{(1)}\}}\mathbb{1}_{\{|X^2|\geq T^{(2)}\}}\right] \\ &\leq \|X^1\|_{k_1}\|X^2\|_{k_2}\mathbb{E}\left[\mathbb{1}_{\{|X^1|\geq T^{(1)}\}}\mathbb{1}_{\{|X^2|\geq T^{(2)}\}}\right]^{\frac{1}{r}},\end{aligned}$$

where $\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{r} = 1$. Then,

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{\{|X^1|\geq T^{(1)}\}}\mathbb{1}_{\{|X^2|\geq T^{(2)}\}}\right]^{\frac{1}{r}} &\leq \mathbb{P}\left(|X^1| \geq T^{(1)}\right)^{1/2r}\mathbb{P}\left(|X^1| \geq T^{(1)}\right)^{1/2r} \\ &\leq \|X^1\|_{k_1}^{\frac{k_1}{2r}}\left(T^{(1)}\right)^{-\frac{k_1}{2r}}\|X^2\|_{k_2}^{\frac{k_2}{2r}}\left(T^{(2)}\right)^{-\frac{k_2}{2r}} \\ &\leq \frac{1}{2}\|X^1\|_{k_1}^{\frac{k_1}{r}}\left(T^{(1)}\right)^{-\frac{k_1}{r}} + \frac{1}{2}\|X^2\|_{k_2}^{\frac{k_2}{2r}}\left(T^{(2)}\right)^{-\frac{k_1}{r}},\end{aligned}$$

where we successively used Cauchy–Schwarz's inequality, Markov's inequality and the simple upper bound $ab \leq a^2/2 + b^2/2$. We deduce,

$$\left|b^{(3)}_{T^{(1)},T^{(2)}}\right| \leq \frac{1}{2}\|X^1\|_{k_1}^{1+\frac{k_1}{r}}\|X^2\|_{k_2}\left(T^{(1)}\right)^{-\frac{k_1}{r}} + \frac{1}{2}\|X^1\|_{k_1}\|X^2\|_{k_2}^{1+\frac{k_2}{r}}\left(T^{(2)}\right)^{-\frac{k_2}{r}}. \tag{A.19}$$

Collecting (A.17)–(A.19) with $1/r = 1 - 2/\bar{k}$ proves the lemma. □

A.2.2 Proof adaptive procedure

Before proving Theorem 4.11, we introduce several notations and state some auxiliary lemmas. We set for $T \in \mathcal{T}$,

$$\mathbb{D}_T = \left(\sup_{T'\in\mathcal{T}}\left|\mathbb{E}\left[\hat{\gamma}_n^{(T,T')} - \hat{\gamma}_n^{(T')}\right]\right|\right) \vee \left|\mathbb{E}\left[\hat{\gamma}_n^{(T)}\right] - \gamma\right|, \tag{A.20}$$

$$\begin{aligned}\bar{\boldsymbol{b}}_T = \bar{\boldsymbol{b}}_{(T^{(1)},T^{(2)})} = \frac{3}{2}(T^{(1)})^{-k_1(1-\frac{2}{\bar{k}})}\|X^1\|_{k_1}^{1+k_1(1-\frac{2}{\bar{k}})}\|X^2\|_{k_2} + \\ \frac{3}{2}(T^{(2)})^{-k_2(1-\frac{2}{\bar{k}})}\|X^1\|_{k_1}\|X^2\|_{k_2}^{1+k_2(1-\frac{2}{\bar{k}})}.\end{aligned}$$

The quantity $\bar{\boldsymbol{b}}_T$ is some upper bound for the bias term according to Lemma A.1. We show in the next lemma that it also controls $\mathbb{D}_T$.

Lemma A.2. *We have* $\mathbb{D}_T \leq 2\overline{\boldsymbol{b}}_T$.

Proof. By Lemma A.1 we have $\left|\mathbb{E}\left[\hat{\gamma}_n^{(T)}\right] - \gamma\right| \leq \overline{\boldsymbol{b}}_T$ for $T \in \mathcal{T}$. For $(T, T') \in \mathcal{T}^2$, recalling (4.13)–(4.14), we have

$$\begin{aligned}\mathbb{E}\left[\hat{\gamma}_n^{(T,T')} - \hat{\gamma}_n^{(T')}\right] &= \mathbb{E}\left[Z_1^{1,T^{(1)}\wedge T'^{(1)}} Z_1^{2,T^{(2)}\wedge T'^{(2)}}\right] - \mathbb{E}\left[Z_1^{1,T'^{(1)}} Z_1^{2,T'^{(2)}}\right] \\ &= \mathbb{E}\left[\left[X_1^1\right]_{T^{(1)}\wedge T'^{(1)}} \left[X_1^2\right]_{T^{(2)}\wedge T'^{(2)}}\right] - \mathbb{E}\left[\left[X_1^1\right]_{T'^{(1)}} \left[X_1^2\right]_{T'^{(2)}}\right] \\ &=: \boldsymbol{b}_{(T^{(1)},T^{(2)},T'^{(1)},T'^{(2)})},\end{aligned}$$

where we used definitions (4.12) and the centering of the Laplace variables in the second line. Now we show $\left|\boldsymbol{b}_{(T^{(1)},T^{(2)},T'^{(1)},T'^{(2)})}\right| \leq 2\overline{\boldsymbol{b}}_{(T^{(1)},T^{(2)})}$. We discuss different cases according to the relative positions of $T^{(j)}$ and $T'^{(j)}$, $j = 1, 2$.

- *Case 1 :* $T^{(1)} \geq T'^{(1)},\ T^{(2)} \geq T'^{(2)}$. Then, $|\boldsymbol{b}_{(T^{(1)},T^{(2)},T'^{(1)},T'^{(2)})}| = 0 \leq \overline{\boldsymbol{b}}_{(T^{(1)},T^{(2)})}$.
- *Case 2 :* $T^{(1)} \leq T'^{(1)},\ T^{(2)} \leq T'^{(2)}$. Then, by triangular inequality,

$$\begin{aligned}|\boldsymbol{b}_{(T^{(1)},T^{(2)},T'^{(1)},T'^{(2)})}| &= \left|\mathbb{E}\left[\left[X_1^1\right]_{T^{(1)}} \left[X_1^2\right]_{T^{(2)}}\right] - \mathbb{E}\left[\left[X_1^1\right]_{T'^{(1)}} \left[X_1^2\right]_{T'^{(2)}}\right]\right| \\ &\leq \left|\mathbb{E}\left[\left[X_1^1\right]_{T^{(1)}} \left[X_1^2\right]_{T^{(2)}}\right] - \gamma\right| + \left|\mathbb{E}\left[\left[X_1^1\right]_{T'^{(1)}} \left[X_1^2\right]_{T'^{(2)}}\right] - \gamma\right|.\end{aligned}$$

  Then using Lemma A.1 and (A.16) and the definition of $\overline{\boldsymbol{b}}_{(T^{(1)},T^{(2)})}$, we deduce

$$\begin{aligned}|\boldsymbol{b}_{(T^{(1)},T^{(2)},T'^{(1)},T'^{(2)})}| &\leq \overline{\boldsymbol{b}}_{(T^{(1)},T^{(2)})} + \overline{\boldsymbol{b}}_{(T'^{(1)},T'^{(2)})} \\ &\leq 2\overline{\boldsymbol{b}}_{(T^{(1)},T^{(2)})},\end{aligned}$$

  where in the last line we used that $T^{(1)} \leq T'^{(1)}$, $T^{(2)} \leq T'^{(2)}$ implies $\overline{\boldsymbol{b}}_{(T'^{(1)},T'^{(2)})} \leq \overline{\boldsymbol{b}}_{(T^{(1)},T^{(2)})}$.
- *Case 3 :* $T^{(1)} \leq T'^{(1)},\ T^{(2)} \geq T'^{(2)}$. Then, we have

$$\begin{aligned}\boldsymbol{b}_{(T^{(1)},T^{(2)},T'^{(1)},T'^{(2)})} &= \mathbb{E}\left[\left[X_1^1\right]_{T^{(1)}} \left[X_1^2\right]_{T'^{(2)}}\right] - \mathbb{E}\left[\left[X_1^1\right]_{T'^{(1)}} \left[X_1^2\right]_{T'^{(2)}}\right] \\ &= \mathbb{E}\left[\left(\left[X_1^1\right]_{T^{(1)}} - X_1^1\right)\left[X_1^2\right]_{T'^{(2)}}\right] - \mathbb{E}\left[\left(\left[X_1^1\right]_{T'^{(1)}} - X_1^1\right)\left[X_1^2\right]_{T'^{(2)}}\right].\end{aligned}$$

  Using the notation $\Delta^{(1)} = \left[X_1^1\right]_{T^{(1)}} - X_1^1$ and following the same computation as those yielding to (A.17), we can show

$$\begin{aligned}\left|\mathbb{E}\left[\left(\left[X_1^1\right]_{T^{(1)}} - X_1^1\right)\left[X_1^2\right]_{T'^{(2)}}\right]\right| &= \left|\mathbb{E}\left[\Delta^{(1)}\left[X_1^2\right]_{T'^{(2)}}\right]\right| \\ &\leq \|X^1\|_{k_1}^{1+k_1(1-\frac{2}{k})} \|X^2\|_{k_2} \left(T^{(1)}\right)^{-k_1(1-\frac{2}{k})}.\end{aligned}$$

  In the same way, we also have,

$$\left|\mathbb{E}\left[\left(\left[X_1^1\right]_{T'^{(1)}} - X_1^1\right)\left[X_1^2\right]_{T'^{(2)}}\right]\right| \leq \|X^1\|_{k_1}^{1+k_1(1-\frac{2}{k})} \|X^2\|_{k_2} \left(T'^{(1)}\right)^{-k_1(1-\frac{2}{k})}.$$

  As $T^{(1)} \leq T'^{(1)}$, we deduce $\left|\boldsymbol{b}_{(T^{(1)},T^{(2)},T'^{(1)},T'^{(2)})}\right| \leq 2\overline{\boldsymbol{b}}_{(T^{(1)},T^{(2)})}$.
- *Case 4 :* $T^{(1)} \geq T'^{(1)},\ T^{(2)} \leq T'^{(2)}$, is treated similarly to Case 3.

□

The following proposition shows that $\mathbb{B}_T$, defined in (4.16), can be compared to the bias and heuristically justifies the choice (4.17).

Proposition A.3. *Assume that $k_1^{-1}+k_2^{-1}<1$, $\beta_n^j=\frac{\alpha_j}{\lfloor \log_2(n)\rfloor}$, for $j=1,2$ and $\kappa_n=c_0\log_2(n)$ for some $c_0>0$. Then, if c_0 is large enough, there exists $c>0$, $\overline{c}>0$, such that for all $T\in\mathcal{T}$, $\forall n\geq 1$, $(n\alpha_1^2\alpha_2^2)/(\log(n))^5\geq 1$,*

$$\mathbb{E}\left[\mathbb{B}_T\right]\leq c\overline{\boldsymbol{b}}_T^2+\frac{c}{(n^{\overline{c}}\alpha_1^2\alpha_2^2)}.$$

The constant $\overline{c}$ can be arbitrarily large if c_0 is chosen large enough.

Proof. For $(T,T')\in\mathcal{T}^2$, we write,

$$|\hat{\gamma}_n^{(T,T')}-\hat{\gamma}_n^{(T')}|^2\leq 8\left|\hat{\gamma}_n^{(T,T')}-\mathbb{E}\left[\hat{\gamma}_n^{(T,T')}\right]\right|^2+8\left|\hat{\gamma}_n^{(T')}-\mathbb{E}\left[\hat{\gamma}_n^{(T')}\right]\right|^2+8\left|\mathbb{E}\left[\hat{\gamma}_n^{(T')}\right]-\mathbb{E}\left[\hat{\gamma}_n^{(T,T')}\right]\right|^2,$$

and deduce from (4.16) and (A.20) that

$$\begin{aligned}\mathbb{B}_T\leq\ & 8\sum_{T'\in\mathcal{T}}\left\{\left(\left|\hat{\gamma}_n^{(T,T')}-\mathbb{E}[\hat{\gamma}_n^{(T,T')}]\right|^2-\frac{1}{16}\mathbb{V}_{T'}\right)_+\right\}\\ &+8\sum_{T'\in\mathcal{T}}\left\{\left(\left|\hat{\gamma}_n^{(T')}-\mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2-\frac{1}{16}\mathbb{V}_{T'}\right)_+\right\}+8\mathbb{D}_T^2\\ =:\ & 8\left[\mathbb{B}_T^{(1)}+\mathbb{B}_T^{(2)}+\mathbb{D}_T^2\right]. \end{aligned}\tag{A.21}$$

Using Lemma A.2, we see that the proposition will be proved as soon as we show,

$$\mathbb{E}\left[\mathbb{B}_T^{(l)}\right]\leq\frac{c}{n^{\overline{c}}},\ \text{for } T\in\mathcal{T},\ \text{and } l=1,2.\tag{A.22}$$

First, we focus on $\mathbb{E}\left[\mathbb{B}_T^{(2)}\right]$. For $T'\in\mathcal{T}$, we denote by $g_{T'}:\mathbb{R}^2\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ the function defined as $g_{T'}(\boldsymbol{x},e_1,e_2)=([x_1]_{T'^{(1)}}+e_1)([x_2]_{T'^{(2)}}+e_2)$, which is such that

$$g_{T'}(\boldsymbol{X}_i,\mathcal{E}_i^{(1),T'^{(1)}},\mathcal{E}_i^{(2),T'^{(2)}})=\left(\left[X_i^1\right]_{T'^{(1)}}+\mathcal{E}_i^{(1),T'^{(1)}}\right)\left(\left[X_i^2\right]_{T'^{(2)}}+\mathcal{E}_i^{(2),T'^{(2)}}\right)=Z_i^{1,T'^{(1)}}Z_i^{2,T'^{(2)}},$$

by (4.12). Thus,

$$\hat{\gamma}_n^{(T')}-\mathbb{E}\left[\hat{\gamma}_n^{(T')}\right]=\frac{1}{n}\sum_{i=1}^n\left\{g_{T'}(\boldsymbol{X}_i,\mathcal{E}_i^{(1),T'^{(1)}},\mathcal{E}_i^{(2),T'^{(2)}})-\mathbb{E}\left[g_{T'}(\boldsymbol{X}_i,\mathcal{E}_i^{(1),T'^{(1)}},\mathcal{E}_i^{(2),T'^{(2)}})\right]\right\},$$

recalling (4.13). We intend to apply Bernstein's inequality and introduce a set on which the random variables we consider are bounded. Let

$$\Omega_n=\left\{\omega\in\Omega\mid\forall j\in\{1,2\},\ \forall T\in\mathcal{T}^{(l)},\ \forall i\in\{1,\dots,n\},\text{we have }|\mathcal{E}_i^{(j),T}|\leq T\tilde{\kappa}_n^{(j)}\right\}$$

where $\tilde{\kappa}_n^{(j)}=\frac{(2c_0+16)\log(n)}{\beta_n^j}$ for $j=1,2$, and c_0 is the constant given in the statement of the proposition. We introduce the following lemma, its proof is postponed until after the current proposition.

Lemma A.4. *We have $\mathbb{P}(\Omega_n^c)\leq 2\frac{\lfloor log_2(n)\rfloor}{n^{8+c_0}}$.*

We introduce a bounded modification of $g_{T'}$ defined as

$$\tilde{g}_{T'}(\boldsymbol{X}_i,\mathcal{E}_i^{1,T'^{(1)}},\mathcal{E}_i^{2,T'^{(2)}})=\left(\left[X_i^1\right]_{T'^{(1)}}+\left[\mathcal{E}_i^{(1),T'^{(1)}}\right]_{T'^{(1)}\tilde{\kappa}_n}\right)\left(\left[X_i^2\right]_{T'^{(2)}}+\left[\mathcal{E}_i^{(2),T'^{(2)}}\right]_{T'^{(2)}\tilde{\kappa}_n}\right).$$

We have $\|\tilde{g}_{T'}\|_\infty\leq|T'^{(1)}||T'^{(2)}|(1+\tilde{\kappa}_n^{(1)})(1+\tilde{\kappa}_n^{(2)})=:M_{T'}$, and with computations analogous to the ones giving (A.9), we can prove

$$var\left(\tilde{g}_{T'}(\boldsymbol{X}_i,\mathcal{E}_i^{1,T'^{(1)}},\mathcal{E}_i^{2,T'^{(2)}})\right)\leq C\frac{|T'^{(1)}|^2|T'^{(2)}|^2}{|\beta_n^1|^2|\beta_n^2|^2}=:v_{T'},$$

for some universal constant C. We recall the Bernstein's inequality (see e.g. [8]) : for $(G_i)_{i=1,\dots,n}$ a iid sequence with $\|G_i\|_\infty \leq M$ and $var(G_i) \leq v$,

$$\mathbb{P}\left(|\frac{1}{n}\sum_{i=1}^{n} G_i - \mathbb{E}[G_i]| \geq t\right) \leq 2\exp\left(-\frac{nt^2}{2(v+Mt)}\right) \leq 2\exp\left(-\frac{nt^2}{4v}\right) + 2\exp\left(-\frac{nt}{4M}\right).$$

As on Ω_n, the random variables $g_{T'}(\boldsymbol{X}_i, \mathcal{E}_i^{1,T'^{(1)}}, \mathcal{E}_i^{2,T'^{(2)}})$ and $\tilde{g}_{T'}(\boldsymbol{X}_i, \mathcal{E}_i^{1,T'^{(1)}}, \mathcal{E}_i^{2,T'^{(2)}})$ are almost surely equal, we deduce for $t \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left\{\left|\hat{\gamma}_n^{(T')} - \mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2 - \frac{1}{16}\mathbb{V}_{T'} \geq t;\ \Omega_n\right\}\right) \\
&= \mathbb{P}\left(\left\{|n^{-1}\sum_{i=1}^{n}\tilde{g}_{T'}(\boldsymbol{X}_i, \mathcal{E}_i^{1,T'^{(1)}}, \mathcal{E}_i^{2,T'^{(2)}}) - \mathbb{E}[\tilde{g}_{T'}(\boldsymbol{X}_i, \mathcal{E}_i^{1,T'^{(1)}}, \mathcal{E}_i^{2,T'^{(2)}})]| \geq \sqrt{\frac{1}{16}\mathbb{V}_{T'} + t};\ \Omega_n\right\}\right) \\
&\qquad \leq 2\exp\left(-\frac{n\left(\frac{1}{16}\mathbb{V}_{T'} + t\right)}{4v_{T'}}\right) + 2\exp\left(-\frac{n\sqrt{\frac{1}{16}\mathbb{V}_{T'} + t}}{4M_{T'}}\right) \\
&\qquad \leq 2\exp\left(-\frac{n\mathbb{V}_{T'}}{64v_{T'}}\right)\exp\left(-\frac{nt}{4v_{T'}}\right) + 2\exp\left(\frac{-n\sqrt{\mathbb{V}_{T'}}}{32M_{T'}}\right)\exp\left(\frac{-n\sqrt{t}}{8M_{T'}}\right). \qquad (A.23)
\end{aligned}$$

By (4.15), we have $\frac{n\mathbb{V}_{T'}}{v_{T'}} = \frac{\kappa_n}{C} = \frac{c_0\ln(n)}{C}$ for some universal constant C, and $\frac{n\sqrt{\mathbb{V}_{T'}}}{M_{T'}} = \frac{\sqrt{n}\sqrt{\kappa_n}}{\beta_n^1\beta_n^2(1+\tilde{k}_n^{(1)})(1+\tilde{k}_n^{(2)})}$ is equal to $\frac{\sqrt{n}\sqrt{c_0\log(n)}\lfloor\log_2(n)\rfloor^2}{\alpha_1\alpha_2\left(1+(2c_0+16)\frac{\log(n)\lfloor\log_2(n)\rfloor}{\alpha_1}\right)\left(1+(2c_0+16)\frac{\log(n)\lfloor\log_2(n)\rfloor}{\alpha_2}\right)} \geq \frac{C'\sqrt{n}}{(c_0+8)^{3/2}\log(n)^{3/2}}$ for some constant C'. Hence,

$$\begin{aligned}
&\mathbb{P}\left(\left\{\left|\hat{\gamma}_n^{(T')} - \mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2 - \frac{1}{16}\mathbb{V}_{T'} \geq t;\ \Omega_n\right\}\right) \\
&\quad \leq 2\exp\left(-\frac{c_0\ln(n)}{C64}\right)\exp\left(-\frac{nt}{4v_{T'}}\right) + 2\exp\left(-\frac{C'\sqrt{n}}{(c_0+8)^{3/2}\log(n)^{3/2}}\right)\exp\left(\frac{-n\sqrt{t}}{8M_{T'}}\right).
\end{aligned}$$

By choosing c_0 large enough, we have

$$\mathbb{P}\left(\left\{\left|\hat{\gamma}_n^{(T')} - \mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2 - \frac{1}{16}\mathbb{V}_{T'} \geq t;\ \Omega_n\right\}\right) \leq Cn^{-\bar{c}}\left(\exp\left(-\frac{nt}{4v_{T'}}\right) + \exp\left(\frac{-n\sqrt{t}}{8M_{T'}}\right)\right),$$

where $\bar{c} > 0$ is any arbitrary constant. Since the previous control is valid for any $T' \in \mathcal{T}$, we deduce that,

$$\begin{aligned}
&\mathbb{P}\left(\left\{\sup_{T'\in\mathcal{T}}\left(\left|\hat{\gamma}_n^{(T')} - \mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2 - \frac{1}{16}\mathbb{V}_{T'}\right)_+ \geq t;\ \Omega_n\right\}\right) \\
&\qquad\qquad \leq Cn^{-\bar{c}}\sum_{T'\in\mathcal{T}}\left(\exp\left(-\frac{nt}{4v_{T'}}\right) + \exp\left(\frac{-n\sqrt{t}}{8M_{T'}}\right)\right).
\end{aligned}$$

Integrating with respect to t on $[0,\infty)$, we get

$$\mathbb{E}\left[\sup_{T'\in\mathcal{T}}\left(\left|\hat{\gamma}_n^{(T')} - \mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2 - \frac{1}{16}\mathbb{V}_{T'}\right)_+ \mathbb{1}_{\Omega_n}\right] \leq Cn^{-\bar{c}}\sum_{T'\in\mathcal{T}}\left(\frac{v_{T'}}{n} + \frac{M_{T'}^2}{n^2}\right).$$

Using (4.11), $v_{T'} = C\frac{T'^{(1)}T'^{(2)}}{|\beta_n^1|^2|\beta_n^2|^2} \leq C\frac{n^4\log_2(n)^4}{\alpha_1^2\alpha_2^2}$ and $M_{T'} = |T'^{(1)}||T'^{(2)}|(1+\tilde{\kappa}_n^{(1)})(1+\tilde{\kappa}_n^{(2)})$ is upper bounded by $n^2(1+\frac{C\log(n)^2}{\alpha_1})(1+\frac{C\log(n)^2}{\alpha_2})$, we deduce

$$\mathbb{E}\left[\sup_{T'\in\mathcal{T}}\left(\left|\hat{\gamma}_n^{(T')} - \mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2 - \frac{1}{16}\mathbb{V}_{T'}\right)_+ \mathbb{1}_{\Omega_n}\right] \leq Cn^{-\bar{c}}\frac{n^3\log(n)^5}{\alpha_1^2\alpha_2^2}. \qquad (A.24)$$

We now study the contribution coming from the event Ω_n^c. We have, using the simple inequality $(a-b)_+ \le (a)_+$ and Jensen's inequality

$$\mathbb{E}\left[\sup_{T'\in\mathcal{T}}\left(\left|\hat{\gamma}_n^{(T')}-\mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2-\frac{1}{16}\mathbb{V}_{T'}\right)_+\mathbb{1}_{\Omega_n^c}\right]\le 2\sum_{T'\in\mathcal{T}}\mathbb{E}\left[\left|\hat{\gamma}_n^{(T')}\right|^2\mathbb{1}_{\Omega_n^c}\right].$$

By using Cauchy–Schwarz's inequality, it comes

$$\begin{aligned}\mathbb{E}\left[\sup_{T'\in\mathcal{T}}\left(\left|\hat{\gamma}_n^{(T')}-\mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2-\frac{1}{16}\mathbb{V}_{T'}\right)_+\mathbb{1}_{\Omega_n^c}\right]&\le 2\sum_{T'\in\mathcal{T}}\mathbb{E}\left[\left|\hat{\gamma}_n^{(T')}\right|^4\right]^{1/2}\mathbb{P}(\Omega_n^c)^{1/2}\\ &\le 2\sum_{T'\in\mathcal{T}}\left[n^{-1}\sum_{i=1}^n\mathbb{E}(|Z^{1,T'^{(1)}}Z^{2,T'^{(2)}}|^4)\right]^{1/2}\mathbb{P}(\Omega_n^c)^{1/2},\end{aligned}$$

where we used again Jensen's inequality. Since $\mathbb{E}(|Z^{l,T'^{(l)}}|)^q \le C_q\frac{|T'^{(l)}|^q}{\beta_n^q}$ for all $q\ge 1$ and $l\in\{1,2\}$, we deduce,

$$\begin{aligned}\mathbb{E}\left[\sup_{T'\in\mathcal{T}}\left(\left|\hat{\gamma}_n^{(T')}-\mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2-\frac{1}{16}\mathbb{V}_{T'}\right)_+\mathbb{1}_{\Omega_n^c}\right]&\le C\sum_{T'\in\mathcal{T}}|T'^{(1)}|^2|T'^{(2)}|^2\mathbb{P}(\Omega_n^c)^{1/2}\\ &\le C\,card(\mathcal{T})\times\frac{n^4}{|\beta_n^1|^2|\beta_n^2|^2}\times\frac{\sqrt{\log(n)}}{n^{4+c_0/2}}\\ &\le C\frac{\log(n)^{11/2}}{\alpha_1^2\alpha_2^2n^{c_0/2}}\end{aligned}\tag{A.25}$$

by Lemma A.4. Collecting (A.24) and (A.25), with the fact that c_0 can be chosen arbitrarily large, and $\alpha\le 1$, we have

$$\mathbb{E}\left[\sup_{T'\in\mathcal{T}}\left(\left|\hat{\gamma}_n^{(T')}-\mathbb{E}[\hat{\gamma}_n^{(T')}]\right|^2-\frac{1}{16}\mathbb{V}_{T'}\right)_+\right]\le\frac{C}{\alpha_1^2\alpha_2^2n^{\bar{c}}}.$$

This is exactly (A.22) with $l=2$. The proof of (A.22) with $l=1$ is obtained similarly, remarking that the application of the Bernstein's inequality, with the same constants $M_{T'}$, $v_{T'}$ still yields to the upper bound (A.23) for $\mathbb{P}\left(\left\{\left|\hat{\gamma}_n^{(T,T')}-\mathbb{E}[\hat{\gamma}_n^{(T,T')}]\right|^2-\frac{1}{16}\mathbb{V}_{T'}\ge t;\ \Omega_n\right\}\right)$. □

Proof of Lemma A.4. The set Ω_n^c is included in

$$\bigcup_{j=1}^2\bigcup_{T^{(j)}\in\mathcal{T}^{(j)}}\{|\mathcal{E}^{(j),T^{(j)}}|\ge T^{(j)}\tilde{\kappa}_n^{(j)}\}.$$

But $\mathcal{E}^{(j),T^{(j)}}/T^{(j)}$ is distributed as a Laplace $\frac{2}{\beta_n^j}\times\mathcal{L}(1)$ variable, and we deduce

$$\begin{aligned}\mathbb{P}(\Omega_n^c)&\le card\left(\mathcal{T}^{(1)}\right)\mathbb{P}\left(\frac{2}{\beta_n^1}|\mathcal{L}(1)|\ge\tilde{\kappa}_n^{(1)}\right)+card\left(\mathcal{T}^{(2)}\right)\mathbb{P}\left(\frac{2}{\beta_n^2}|\mathcal{L}(1)|\ge\tilde{\kappa}_n^{(2)}\right)\\ &\le 2\lfloor\log_2(n)\rfloor e^{-\tilde{\kappa}_n\beta_n/2}=2\lfloor\log_2(n)\rfloor e^{-\ln(n)(2c_0+16)/2}=2\lfloor\log_2(n)\rfloor n^{-8-c_0},\end{aligned}$$

as we wanted. □

Proposition A.5. *Assume that $k_1^{-1}+k_2^{-1}<1$, and $\beta_n^j=\frac{\alpha_j}{\lfloor\log_2(n)\rfloor}$, $\kappa_n=c_0\log(n)$ for some $c_0>0$. If c_0 is large enough, there exist $c>0$, $\bar{c}_0>0$, such that*

$$\mathbb{E}\left[(\hat{\gamma}_n^{\widehat{T}}-\gamma)^2\right]\le c\inf_{T\in\mathcal{T}}[\bar{\boldsymbol{b}}_T^2+\mathbb{V}_T]+\frac{c}{\alpha_1^2\alpha_2^2n^{\bar{c}_0}},$$

for all $n\ge 1$, $\alpha_j\le 1$, $(n\alpha_1^2\alpha_2^2)/(\log(n))^5\ge 1$. Moreover, the constant $\bar{c}_0$ can be chosen arbitrarily large by choosing c_0 large enough.

Proof. Let $T \in \mathcal{T}$, we have

$$|\hat{\gamma}^{(\widehat{T})} - \gamma| \le |\hat{\gamma}^{(\widehat{T})} - \hat{\gamma}^{(\widehat{T},T)}| + |\hat{\gamma}^{(\widehat{T},T)} - \hat{\gamma}^{(T)}| + |\hat{\gamma}^{(T)} - \gamma|.$$

From the definition (A.20), we have $|\hat{\gamma}^{(T)} - \gamma| \le |\hat{\gamma}^{(T)} - \mathbb{E}[\hat{\gamma}^{(T)}]| + |\mathbb{E}[\hat{\gamma}^{(T)} - \gamma]| \le |\hat{\gamma}^{(T)} - \mathbb{E}[\hat{\gamma}^{(T)}]| + \mathbb{D}_T$, and it follows

$$|\hat{\gamma}^{(\widehat{T})} - \gamma| \le |\hat{\gamma}^{(\widehat{T})} - \hat{\gamma}^{(\widehat{T},T)}| + |\hat{\gamma}^{(\widehat{T},T)} - \hat{\gamma}^{(T)}| + |\hat{\gamma}^{(T)} - \mathbb{E}[\hat{\gamma}^{(T)}]| + \mathbb{D}_T.$$

By (4.16), we have $|\hat{\gamma}^{(\widehat{T},T)} - \hat{\gamma}^{(T)}|^2 \le \mathbb{B}_{\widehat{T}} + \mathbb{V}_{\widehat{T}}$, and recalling $\hat{\gamma}^{(\widehat{T},T)} = \hat{\gamma}^{(T,\widehat{T})}$, we also get $|\hat{\gamma}^{(\widehat{T})} - \hat{\gamma}^{(\widehat{T},T)}|^2 \le \mathbb{B}_T + \mathbb{V}_T$. Thus,

$$|\hat{\gamma}^{(\widehat{T})} - \gamma|^2 \le 16\left[\mathbb{B}_{\widehat{T}} + \mathbb{V}_{\widehat{T}} + \mathbb{B}_T + \mathbb{V}_T + |\hat{\gamma}^{(T)} - \mathbb{E}[\hat{\gamma}^{(T)}]|^2 + \mathbb{D}_T^2\right].$$

Using (4.17), we deduce

$$|\hat{\gamma}^{(\widehat{T})} - \gamma|^2 \le 16\left[2\mathbb{B}_T + 2\mathbb{V}_T + |\hat{\gamma}^{(T)} - \mathbb{E}[\hat{\gamma}^{(T)}]|^2 + \mathbb{D}_T^2\right].$$

From the study of the variance of $\hat{\gamma}^{(T)}$ as in (A.9), we have $\mathbb{E}[|\hat{\gamma}^{(T)} - \mathbb{E}[\hat{\gamma}^{(T)}]|^2] \le Cn^{-1}\frac{|T^{(1)}|^2|T^{(2)}|^2}{|\beta_n^1|^2|\beta_n^2|^2}$, which is smaller than $\mathbb{V}_T$, if c_0 in (4.15) is large enough. Thus, we deduce

$$\mathbb{E}\left[|\hat{\gamma}^{(\widehat{T})} - \gamma|^2\right] \le 16\left[2\mathbb{E}[\mathbb{B}_T] + 3\mathbb{V}_T + \mathbb{D}_T^2\right].$$

Now, Lemma A.2 and Proposition A.3 yield to

$$\mathbb{E}\left[|\hat{\gamma}^{(\widehat{T})} - \gamma|^2\right] \le c\left[\overline{\boldsymbol{b}}_T^2 + \mathbb{V}_T\right] + \frac{c}{\alpha_1^2\alpha_2^2 n^{\overline{c}_0}},$$

for any $T \in \mathcal{T}$. This proves the proposition. □

We end this section by providing the proof of Theorem 4.11

Proof of Theorem 4.11. By Proposition A.3, it is sufficient to evaluate $\inf_{T\in\mathcal{T}}\left[\overline{\boldsymbol{b}}_T^2 + \mathbb{V}_T\right]$ which is, up to a constant, the infimum over $T \in \mathcal{T}$ of

$$\begin{aligned}
&(T^{(1)})^{-2k_1(1-\frac{2}{\overline{k}})} + (T^{(2)})^{-2k_2(1-\frac{2}{\overline{k}})} + n^{-1}\kappa_n\frac{|T^{(1)}|^2|T^{(2)}|^2}{|\beta_n^1|^2|\beta_n^2|^2} \\
&\le (T^{(1)})^{-2k_1(1-\frac{2}{\overline{k}})} + (T^{(2)})^{-2k_2(1-\frac{2}{\overline{k}})} + Cn^{-1}\log(n)^5\frac{|T^{(1)}|^2|T^{(2)}|^2}{\alpha_1^2\alpha_2^2}, \qquad (A.26)
\end{aligned}$$

for some $C > 0$. If we set $T^{*(j)} = \left(\frac{n\alpha_1^2\alpha_2^2}{\log(n)^5}\right)^{1/(2k_l)}$ for $j \in \{1,2\}$, the above quantity is smaller than some constant time

$$\left(\frac{\log(n)^5}{n\alpha_1^2\alpha_2^2}\right)^{\frac{\overline{k}-2}{\overline{k}}},$$

which is the expected rate. It remains to check that the same rate can be obtained by restricting T in $\mathcal{T}$. As $(n\alpha_1^2\alpha_2^2)/(\log(n))^5 \ge 1$, and $\alpha_j \le 1$, we see that for $n \ge 3$, $T^{*(1)} \in [1,n]$ and $T^{*(2)} \in [1,n]$. By the definition (4.11) of $\mathcal{T}^{(j)}$, we see that $\mathcal{T}^{(j)}$ is a grid of $[1,n]$ such that for any $t^* \in [1,n]$, there exists $t^{(j)} \in \mathcal{T}^{(j)}$ with $\frac{1}{2}t^* \le t^{(j)} \le 2t^*$. Hence, by replacing the $T^{*(j)}$ by their closest values in $\mathcal{T}^j$ we only increase the value of (A.26) by a multiplicative constant. We deduce

$$\inf_{T\in\mathcal{T}}\left[\overline{\boldsymbol{b}}_T^2 + \mathbb{V}_T\right] \le c\left(\frac{\log(n)^5}{n\alpha_1^2\alpha_2^2}\right)^{\frac{\overline{k}-2}{\overline{k}}},$$

and Theorem 4.11 follows from Proposition A.3. □

A.3 Proof locally private density estimation

Proof of Theorem 4.14. The proof is based on the usual bias-variance decomposition. We have indeed $\mathbb{E}[|\hat{\pi}_h^Z(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|^2] = |\mathbb{E}[\hat{\pi}_h^Z(\boldsymbol{x_0})] - \pi(\boldsymbol{x_0})|^2 + var(\hat{\pi}_h^Z(\boldsymbol{x_0}))$. One can easily bound the bias part (see for example Proposition 1.2 in [36]), obtaining for any $\boldsymbol{x_0} \in \mathbb{R}^d$ and any $h > 0$

$$|\mathbb{E}[\hat{\pi}_h^Z(\boldsymbol{x_0})] - \pi(\boldsymbol{x_0})|^2 \leq ch^{2\beta}, \tag{A.27}$$

for some $c > 0$. Regarding the variance of $\hat{\pi}_h^Z(\boldsymbol{x_0})$, we use its explicit form and the fact that the vectors $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ and the Laplace random variables are independent to get

$$\begin{aligned} var(\hat{\pi}_h^Z(\boldsymbol{x_0})) &= \frac{1}{n^2}\sum_{i=1}^n var\Big(\prod_{j=1}^d (\frac{1}{h}K(\frac{X_i^j - x_0^j}{h}) + \mathcal{E}_i^j)\Big) \\ &= \frac{1}{n^2}\sum_{i=1}^n var\Big(\sum_{I_k}\prod_{j\in I_k}\frac{1}{h}K(\frac{X_i^j - x_0^j}{h})\prod_{j\in (I_k)^c}\mathcal{E}_i^j\Big) \end{aligned}$$

where I_k is a set of index such that $|I_k| = k$, for $k \in \{1, ..., d\}$. Then, it is well known that $var(\frac{1}{h^d}\prod_{j=1}^d K(\frac{X_i^j - x_0^j}{h})) \leq \frac{c}{h^d}$ for some positive c (one can easily see that by adapting Proposition 1.1 in [36] to the multidimensional context). Moreover, by construction, $\mathcal{E}_i^j$ are iid $\sim \mathcal{L}(\frac{2\kappa}{\alpha_j h})$, which guarantees that $var(\prod_{j=1}^d \mathcal{E}_i^j) \leq \frac{c}{\prod_{j=1}^d (\alpha_j h)^2}$. One can then readily check that, for any set of index I_k such that $|I_k| = k$, it is

$$var(\prod_{j\in I_k}\frac{1}{h}K(\frac{X_i^j - x_0^j}{h})\prod_{j\in (I_k)^c}\mathcal{E}_i^j) \leq \frac{c}{h^k}\frac{1}{h^{2(d-k)}}\prod_{j\in(I_k)^c}\frac{1}{\alpha_j^2}.$$

It implies

$$var(\hat{\pi}_h^Z(\boldsymbol{x_0})) \leq \frac{1}{n}\sum_{k=0}^d \frac{c}{h^k}\frac{1}{h^{2(d-k)}}\sum_{I_k}\prod_{j\in(I_k)^c}\frac{1}{\alpha_j^2} = \frac{c}{nh^{2d}\prod_{j=1}^d \alpha_j^2}\sum_{k=0}^d\sum_{I_k}\prod_{j\in I_k}(h\alpha_j^2). \tag{A.28}$$

We now recall that h is a bandwidth we have assumed being smaller than 1. We have also required $\alpha_j \leq 1$ for any $j \in \{1, ..., d\}$, which yields $h\alpha_j^2 < 1$ for any $j \in \{1, ..., d\}$. Then, the largest term in the sum above is the one for which $k = 0$, which means that $I_k = \emptyset$. We derive

$$var(\hat{\pi}_h^Z(\boldsymbol{x_0})) \leq \frac{c}{nh^{2d}\prod_{j=1}^d \alpha_j^2}. \tag{A.29}$$

We then look for the choice of h that realizes the trade-off between the bound on the variance here above and the bound on the bias term gathered in (A.27). This is achieved by the rate optimal bandwidth $h^* := (\frac{1}{n\prod_{j=1}^d \alpha_j^2})^{\frac{1}{2(\beta+d)}}$. We remark that, as by hypothesis it is $n\prod_{j=1}^d \alpha_j^2 \to \infty$ for $n \to \infty$, it clearly follows $h^* = h_n^* \to 0$ for $n \to \infty$. Replacing it in (A.27) and (A.29) we obtain $\mathbb{E}[|\hat{\pi}_h^Z(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|^2] \leq c\Big(\frac{1}{n\prod_{j=1}^d \alpha_j^2}\Big)^{\frac{\beta}{\beta+d}}$, which concludes the proof. □

Proof of Theorem 4.15. We observe that, for $\alpha_1 = ... = \alpha_d = \alpha$, (A.28) translates to $var(\hat{\pi}_h^Z(\boldsymbol{x_0})) \leq \frac{c}{n(h\alpha)^{2d}}\sum_{k=0}^d (h\alpha^2)^k$. Then, we consider two different cases.

• If $\alpha \geq n^{\frac{1}{2(2\beta+d)}}$, we choose the optimal bandwidth as in the privacy free-context $h^* := (\frac{1}{n})^{\frac{1}{2\beta+d}}$. We observe we have in this case $h\alpha^2 \geq n^{-\frac{1}{2\beta+d}+\frac{1}{2\beta+d}} = 1$. Hence, in (A.28), the worst term is the one for which $k = d$. It implies the variance of $\hat{\pi}_h^Z(\boldsymbol{x_0})$ is bounded by

$$\frac{c}{n(h\alpha)^{2d}}(h\alpha^2)^d = \frac{c}{nh^d} = c(\frac{1}{n})^{\frac{2\beta}{2\beta+d}}.$$

We observe that the bandwidth h^* is the one that achieves the balance in the decomposition bias-variance, as it is also $(h^*)^{2\beta} = (\frac{1}{n})^{\frac{2\beta}{2\beta+d}}$. The proof in the first case is then concluded.

• Consider now what happens for $\alpha < n^{\frac{1}{2(2\beta+d)}}$. In this case, we will see that the optimal choice in terms of convergence rate will consist in taking $h^* := (\frac{1}{n\,\alpha^{2d}})^{\frac{1}{2(\beta+d)}}$. Remark that we have assumed $n\alpha^{2d} \to \infty$ for $n \to \infty$ so that $h^* \to 0$, for n going to ∞. We observe in this context it is

$$h^*\alpha^2 = (\frac{1}{n\,\alpha^{2d}})^{\frac{1}{2(\beta+d)}}\alpha^2 = (\frac{1}{n})^{\frac{1}{2(\beta+d)}}\alpha^{\frac{2\beta+d}{\beta+d}} \le (\frac{1}{n})^{\frac{1}{2(\beta+d)}} n^{\frac{1}{2(\beta+d)}} = 1.$$

Thus, the largest term in the sum in (A.28) is for $k = 0$, which implies

$$var(\hat{\pi}_h^Z(\boldsymbol{x_0})) \le \frac{c}{n(h\alpha)^{2d}} = \frac{c}{n\alpha^{2d}}(n\alpha^{2d})^{\frac{d}{\beta+d}} = c(\frac{1}{n\alpha^{2d}})^{\frac{\beta}{\beta+d}}.$$

The bandwidth h^* realizes the balance between the variance and the bias, as $(h^*)^{2\beta} = (\frac{1}{n\alpha^{2d}})^{\frac{\beta}{\beta+d}}$. The proof is then complete.

□

A.3.1 Proof adaptive procedure

We start by proving that Z_i^j defined according to (4.12) are α_j local differential private view of the observation X_i^j, as stated in Lemma 4.20.

Proof of Lemma 4.20. From the definition of Laplace, using the independence of the variables $(Z_i^{j,h})_{h\in H_n}$ and denoting as $q^j((z^{j,h})_{h\in H_n}|X_i^j = x)$ the density of the law of $Z_i^{j,h}$ conditional to $X_i^j = x \in \mathbb{R}$ we obtain

$$\begin{aligned}\frac{q^j((z^{j,h})_{h\in H_n}|X_i^j = x)}{q^j((z^{j,h})_{h\in H_n}|X_i^j = x')} &= \frac{\prod_{h\in H_n}\exp(|z^{j,h} - \frac{1}{h}K(\frac{x-x_0^j}{h})|\frac{\beta_n^j h}{2\kappa})}{\prod_{h\in H_n}\exp(|z^{j,h} - \frac{1}{h}K(\frac{x'-x_0^j}{h})|\frac{\beta_n^j h}{2\kappa})} \\ &\le \prod_{h\in H_n}\exp(|K(\frac{x-x_0^j}{h}) - K(\frac{x'-x_0^j}{h})|\frac{\beta_n^j}{2\kappa}) \\ &\le \prod_{h\in H_n}\exp(\beta_n^j) \\ &= \exp(Card(H_n)\beta_n^j) \le \exp(\alpha_j),\end{aligned}$$

being the last a consequence of how we have chosen β_n^j. □

Before proving the main theorem, let us introduce the notation $\pi_h^z(\boldsymbol{x_0})$ for $\mathbb{E}[\hat{\pi}_h^z(\boldsymbol{x_0})]$ and

$$\begin{aligned}\mathbb{D}_h &:= \Big(\sup_{\eta\in H_n}|\mathbb{E}[\hat{\pi}_{\eta\wedge h}^z(\boldsymbol{x_0}) - \hat{\pi}_h^z(\boldsymbol{x_0})]|\Big) \vee |\mathbb{E}[\hat{\pi}_h^z(\boldsymbol{x_0})] - \pi(\boldsymbol{x_0})| \\ &= \Big(\sup_{\eta<h}|\mathbb{E}[\hat{\pi}_\eta^z(\boldsymbol{x_0}) - \hat{\pi}_h^z(\boldsymbol{x_0})]|\Big) \vee |\mathbb{E}[\hat{\pi}_h^z(\boldsymbol{x_0})] - \pi(\boldsymbol{x_0})|.\end{aligned} \tag{A.30}$$

As for (A.27), with classical computations as in Proposition 1.2 of [36] it readily follows, for some $c > 0$,

$$\mathbb{D}_h \le ch^\beta. \tag{A.31}$$

The proof of Theorem 4.21 heavily relies on the following proposition.

Proposition A.6. *Assume that $\pi \in \mathcal{H}(\beta, \mathcal{L})$ for some β and $\mathcal{L} \geq 1$. Moreover, $\beta_n^j = \frac{\alpha_j}{[\log_2 n]}$ for any $j \in \{1, ..., d\}$ and $a_n = c_0 \log n$ for some $c_0 > 0$. If c_0 is large enough, there exist $c > 0$ and $\bar{c} > 0$ such that*

$$\mathbb{E}[(\hat{\pi}^z_{\hat{h}}(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0}))^2] \leq c \inf_{h \in H_n} (\mathbb{V}_h + \mathbb{D}_h^2) + \frac{c}{n^{\bar{c}} \prod_{j=1}^d \alpha_j^2}$$

for all $n \geq 1$, $\alpha_j \leq 1$ and $\frac{n \prod_{j=1}^d \alpha_j^2}{\log n^{1+2d}} \geq 1$. Moreover, the constant $\bar{c}$ can be chosen arbitrarily large, taking the constant c_0 large enough.

Proof of Proposition A.6. Let $h \in H_n$. It is

$$|\hat{\pi}^z_{\hat{h}}(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})| \leq |\hat{\pi}^z_{\hat{h}}(\boldsymbol{x_0}) - \hat{\pi}^z_{\hat{h},h}(\boldsymbol{x_0})| + |\hat{\pi}^z_{\hat{h},h}(\boldsymbol{x_0}) - \hat{\pi}^z_{h}(\boldsymbol{x_0})| + |\hat{\pi}^z_{h}(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|.$$

Following the same computations as in the proof of Proposition A.5 it is then easy to check that

$$\mathbb{E}[|\hat{\pi}^z_{\hat{h}}(\boldsymbol{x_0}) - \pi(\boldsymbol{x_0})|^2] \leq c\Big(\mathbb{E}[\mathbb{B}_h] + \mathbb{V}_h + \mathbb{D}_h^2\Big).$$

Next, we study in detail $\mathbb{E}[\mathbb{B}_h]$. Splitting $\mathbb{B}_h$ in a way analogous to (A.21) in Proposition A.3 we have

$$\begin{aligned}
\mathbb{B}_h \leq 8 \sum_{\eta \in H_n} & \left\{ \left(\left| \hat{\pi}^z_{h,\eta}(\boldsymbol{x_0}) - \mathbb{E}[\hat{\pi}^z_{h,\eta}(\boldsymbol{x_0})] \right|^2 - \frac{1}{16} \mathbb{V}_\eta \right)_+ \right\} \\
& + 8 \sum_{\eta \in H_n} \left\{ \left(\left| \hat{\pi}^z_{\eta}(\boldsymbol{x_0}) - \mathbb{E}[\hat{\pi}^z_{\eta}(\boldsymbol{x_0})] \right|^2 - \frac{1}{16} \mathbb{V}_\eta \right)_+ \right\} + 8\mathbb{D}_h^2 \\
=: 8 & \left[\mathbb{B}_h^{(1)} + \mathbb{B}_h^{(2)} + \mathbb{D}_h^2 \right].
\end{aligned}$$

Hence, Proposition A.6 will be proven once we show that

$$\mathbb{E}[\mathbb{B}_h^{(l)}] \leq \frac{c}{n^{\bar{c}}} \frac{1}{\prod_{j=1}^d \alpha_j^2} \tag{A.32}$$

for $h \in H_n$ and $l = 1, 2$. We start by considering $\mathbb{E}[\mathbb{B}_h^{(2)}]$. Similarly as in Proposition A.6, we introduce for any $\eta \in H_n$ the function $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, defined as

$$g_\eta(x^1, ..., x^d, e^1, ..., e^d) = (\frac{1}{\eta} K(\frac{x^1 - x_0^1}{\eta}) + e^1) \times ... \times (\frac{1}{\eta} K(\frac{x^d - x_0^d}{\eta}) + e^d).$$

It is such that

$$\begin{aligned}
g_\eta({X_i}^1, ..., X_i^d, \mathcal{E}_i^{1,\eta}, ..., \mathcal{E}_i^{d,\eta}) &= (\frac{1}{\eta} K(\frac{{X_i}^1 - x_0^1}{\eta}) + \mathcal{E}_i^{1,\eta}) \times ... \times (\frac{1}{\eta} K(\frac{X_i^d - x_0^d}{\eta}) + \mathcal{E}_i^{d,\eta}) \\
&= Z_i^{1,\eta} \times ... \times Z_i^{d,\eta}.
\end{aligned}$$

Hence, we can write

$$\hat{\pi}^z_\eta(\boldsymbol{x_0}) - \mathbb{E}[\hat{\pi}^z_\eta(\boldsymbol{x_0})] = \frac{1}{n} \sum_{i=1}^n \{g_\eta(\boldsymbol{X}_i, \boldsymbol{\mathcal{E}}_i^\eta) - \mathbb{E}[g_\eta(\boldsymbol{X}_i, \boldsymbol{\mathcal{E}}_i^\eta)]\}.$$

As in the proof of Proposition A.3 we want to apply Bernstein's inequality, for which we need the variables to be bounded. For this reason we introduce

$$\tilde{\Omega}_n = \left\{ \omega \in \Omega \mid \forall j \in \{1, ..., d\} \ \forall l \in \{1, 2\}, \ \forall h \in H_n, \ \forall i \in \{1, \dots, n\}, \text{we have } |\mathcal{E}_i^{j,h}| \leq \frac{\tilde{a}_n^j}{h} \right\},$$

where $\tilde{a}_n^j := \frac{\log n}{\beta_n^j} 2\kappa(c_0 + 4d)$, κ is as in (4.18) and c_0 is the constant given in the statement. Then, we can modify g_η. We set

$$\tilde{g}_\eta(X_i{}^1, ..., X_i^d, \mathcal{E}_i^{1,\eta}, ..., \mathcal{E}_i^{d,\eta}) := (\frac{1}{\eta} K(\frac{X_i{}^1 - x_0^1}{\eta}) + [\mathcal{E}_i^{1,\eta}]_{\frac{\tilde{a}_n^1}{\eta}}) \times ... \times (\frac{1}{\eta} K(\frac{X_i^d - x_0^d}{\eta}) + [\mathcal{E}_i^{d,\eta}]_{\frac{\tilde{a}_n^d}{\eta}}),$$

where we have used the same notation as in Section 4.2 to denote the truncation of the Laplace random variables.
Following the proof of Lemma A.4 it is easy to check that

$$\mathbb{P}(\tilde{\Omega}_n^c) \leq d \frac{\lfloor \log_2 n \rfloor}{n^{4d+c_0}}. \tag{A.33}$$

Indeed, $\tilde{\Omega}_n^c$ is included in $\cup_{j=1}^d \cup_{h \in H_n} \{|\mathcal{E}_i^{j,h}| \geq \frac{\tilde{a}_n^j}{h}\}$, with $h\mathcal{E}_i^{j,h}$ distributed as $\frac{2\kappa}{\beta_n^j} \mathcal{L}(1)$. Hence,

$$\mathbb{P}(\tilde{\Omega}_n^c) \leq Card(H_n) \sum_{j=1}^d \mathbb{P}(\frac{2\kappa}{\beta_n^j} |\mathcal{L}(1)| \geq \tilde{a}_n^j) \leq \lfloor \log_2 n \rfloor \sum_{j=1}^d e^{-\frac{\tilde{a}_n^j \beta_n^j}{2\kappa}} \leq d \lfloor \log_2 n \rfloor n^{-4d-c_0},$$

as we have chosen $\tilde{a}_n^j := \frac{\log n}{\beta_n^j} 2\kappa(c_0 + 4d)$.
We observe that

$$\|\tilde{g}_\eta\|_\infty \leq \frac{1}{\eta^d} (\kappa + \tilde{a}_n^1) \times ... \times (\kappa + \tilde{a}_n^d) =: M_\eta. \tag{A.34}$$

Acting as in the proof of Theorem 4.14 (see in particular (A.28)) it is moreover easy to see that

$$var(\tilde{g}_\eta(X_i{}^1, ..., X_i^d, \mathcal{E}_i^{1,\eta}, ..., \mathcal{E}_i^{d,\eta})) \leq \frac{c}{\eta^{2d}} \frac{1}{\prod_{j=1}^d (\beta_n^j)^2} \sum_{k=0}^d \prod_{j=1}^k \eta(\beta_n^j)^2 \leq \frac{c}{\eta^{2d}} \frac{1}{\prod_{j=1}^d (\beta_n^j)^2} =: v_\eta, \tag{A.35}$$

where we have used that $\eta(\beta_n^j)^2 \leq 1$. We then apply Bernstein inequality (similarly to the proof of Proposition A.3) to the random variables $\tilde{g}_\eta$, on $\tilde{\Omega}_n$. It follows, as in (A.23),

$$\begin{aligned}
&\mathbb{P}\left(\left\{\left|\hat{\pi}_\eta^z(\boldsymbol{x_0}) - \mathbb{E}[\hat{\pi}_\eta^z(\boldsymbol{x_0})]\right|^2 - \frac{1}{16}\mathbb{V}_\eta \geq t;\ \tilde{\Omega}_n\right\}\right) \\
&\quad = \mathbb{P}\left(\left\{|n^{-1} \sum_{i=1}^n \tilde{g}_\eta(\boldsymbol{X}_i, \boldsymbol{\mathcal{E}}_i^\eta) - \mathbb{E}[\tilde{g}_\eta(\boldsymbol{X}_i, \boldsymbol{\mathcal{E}}_i^\eta)]| \geq \sqrt{\frac{1}{16}\mathbb{V}_\eta + t};\ \tilde{\Omega}_n\right\}\right) \\
&\qquad \leq 2\exp\left(-\frac{n\mathbb{V}_\eta}{64 v_\eta}\right)\exp\left(-\frac{nt}{4v_\eta}\right) + 2\exp\left(\frac{-n\sqrt{\mathbb{V}_\eta}}{32 M_\eta}\right)\exp\left(\frac{-n\sqrt{t}}{8M_\eta}\right).
\end{aligned}$$

We now have $n\frac{\mathbb{V}_\eta}{v_\eta} = \frac{a_n}{c} = \frac{c_0 \log n}{c}$ for some universal constant c. Moreover,

$$\begin{aligned}
\frac{n\sqrt{\mathbb{V}_\eta}}{M_\eta} &= n\sqrt{a_n \frac{1}{n\eta^{2d}} \frac{1}{\prod_{j=1}^d (\beta_n^j)^2}}\, \eta^d \frac{1}{\prod_{j=1}^d (\kappa + \tilde{a}_n^j)} \\
&= \sqrt{n}\sqrt{c_0 \log n} \frac{1}{\prod_{j=1}^d (\kappa + \frac{\log n}{\beta_n^j} 2\kappa(c_0 + 4d))} \frac{1}{\prod_{j=1}^d (\beta_n^j)^2} \\
&= \sqrt{n}\sqrt{c_0 \log n} \frac{1}{\prod_{j=1}^d (\kappa + \frac{\log n \lfloor \log_2 n \rfloor}{\alpha_j} 2\kappa(c_0 + 4d))} \frac{(\lfloor \log_2 n \rfloor)^{2d}}{\prod_{j=1}^d \alpha_j^2} \\
&\geq c' \sqrt{n} (\log n)^{\frac{1}{2}}
\end{aligned}$$

for some constant c'. Then, we can follow the arguing in Proposition A.3 and, integrating with respect to t on $[0,\infty)$, we obtain

$$\mathbb{E}\left[\sup_{\eta\in H_n}\left(\left|\hat{\pi}^z_\eta(\boldsymbol{x_0})-\mathbb{E}[\hat{\pi}^z_\eta(\boldsymbol{x_0})]\right|^2-\frac{1}{16}\mathbb{V}_\eta\right)_+ 1_{\tilde{\Omega}_n}\right]\le cn^{-\bar{c}}\sum_{\eta\in H_n}(\frac{v_\eta}{n}+\frac{M_\eta^2}{n^2}). \tag{A.36}$$

From (A.34) and (A.35) and the definition (4.24) it follows

$$M_\eta\le\frac{1}{\eta^d}\prod_{j=1}^d(\kappa+\frac{\log n}{\beta_n^j}2\kappa(c_0+4d))\le cn^d\prod_{j=1}^d(1+\frac{\log n}{\alpha_j}\lfloor\log_2 n\rfloor),$$

$$v_\eta\le\frac{c}{\eta^{2d}}\frac{1}{\prod_{j=1}^d(\beta_n^j)^2}\le n^{2d}(\log_2 n)^{2d}\frac{1}{\prod_{j=1}^d\alpha_j^2}.$$

Replacing it in (A.36) we obtain

$$\begin{aligned}&\mathbb{E}\left[\sup_{\eta\in H_n}\left(\left|\hat{\pi}^z_\eta(\boldsymbol{x_0})-\mathbb{E}[\hat{\pi}^z_\eta(\boldsymbol{x_0})]\right|^2-\frac{1}{16}\mathbb{V}_\eta\right)_+ 1_{\tilde{\Omega}_n}\right] \\ &\le cn^{-\bar{c}}\left(n^{2d-1}(\log_2 n)^{2d+1}\frac{1}{\prod_{j=1}^d\alpha_j^2}+n^{2d-2}\prod_{j=1}^d(1+\frac{\log n}{\alpha_j}\lfloor\log_2 n\rfloor)^2\lfloor\log_2 n\rfloor\right).\end{aligned} \tag{A.37}$$

We then deal with the contribution on $\tilde{\Omega}_n^c$ which, together with (A.33), provides

$$\mathbb{E}\left[\sup_{\eta\in H_n}\left(\left|\hat{\pi}^z_\eta(\boldsymbol{x_0})-\mathbb{E}[\hat{\pi}^z_\eta(\boldsymbol{x_0})]\right|^2-\frac{1}{16}\mathbb{V}_\eta\right)_+ 1_{\tilde{\Omega}_n^c}\right]\le 2\sum_{\eta\in H_n}\mathbb{E}[|\hat{\pi}^z_\eta(\boldsymbol{x_0})|^4]^{\frac{1}{2}}\mathbb{P}(\tilde{\Omega}_n^c)^{\frac{1}{2}}.$$

We observe that, because of Jensen inequality, it is

$$\mathbb{E}[|\hat{\pi}^z_\eta(\boldsymbol{x_0})|^4]\le\frac{1}{n}\sum_{i=1}^n\mathbb{E}[|Z_i^{1,\eta}\times\ldots\times Z_i^{d,\eta}|^4].$$

Moreover, for all $q\ge 1$,

$$\mathbb{E}[|Z_i^{j,\eta}|^q]=\mathbb{E}[|\frac{1}{\eta}K(\frac{X_i^j-x_0^j}{\eta})+\mathcal{E}_i^{j,\eta}|^q]\le\frac{c}{\eta^q}+(\frac{2\kappa}{\eta\beta_n^j})^q\mathbb{E}[|\mathcal{L}(1)|^q]\le(\frac{c}{\eta\beta_n^j})^q.$$

It yields

$$\begin{aligned}\mathbb{E}\left[\sup_{\eta\in H_n}\left(\left|\hat{\pi}^z_\eta(\boldsymbol{x_0})-\mathbb{E}[\hat{\pi}^z_\eta(\boldsymbol{x_0})]\right|^2-\frac{1}{16}\mathbb{V}_\eta\right)_+ 1_{\tilde{\Omega}_n^c}\right]&\le c\sum_{\eta\in H_n}\frac{1}{(\eta\beta_n^1)^2}\times\ldots\times\frac{1}{(\eta\beta_n^d)^2}(d\lfloor\log_2 n\rfloor n^{-4d-c_0})^{\frac{1}{2}} \\ &\le c\; card(H_n)\frac{n^{2d}}{\prod_{j=1}^d(\beta_n^j)^2}\frac{\sqrt{\log n}}{n^{2d+\frac{c_0}{2}}} \\ &\le c\frac{(\log n)^{2d+\frac{3}{2}}}{n^{\frac{c_0}{2}}\prod_{j=1}^d\alpha_j^2}.\end{aligned} \tag{A.38}$$

From (A.37) and (A.38), recalling that c_0 can be chosen arbitrarily large and that $\alpha_j\le 1$ for any $j\in\{1,\ldots,d\}$ we obtain, for some $\bar{c}>0$,

$$\mathbb{E}\left[\sup_{\eta\in H_n}\left(\left|\hat{\pi}^z_\eta(\boldsymbol{x_0})-\mathbb{E}[\hat{\pi}^z_\eta(\boldsymbol{x_0})]\right|^2-\frac{1}{16}\mathbb{V}_\eta\right)_+\right]\le c\frac{1}{n^{\bar{c}}\prod_{j=1}^d\alpha_j^2}.$$

We have therefore proven (A.32) for $l = 2$. For $l = 1$ the proof of the bound in (A.32) is obtained in the same way, applying Bernstein inequality with the same constants M_η and v_η on

$$\mathbb{P}\left(\left\{\left|\hat{\pi}^z_{\eta,h}(\boldsymbol{x_0}) - \mathbb{E}[\hat{\pi}^z_{\eta,h}(\boldsymbol{x_0})]\right|^2 - \frac{1}{16}\mathbb{V}_\eta \geq t;\ \tilde{\Omega}_n\right\}\right).$$

The proof is therefore concluded. □

Proof. Theorem 4.21. From Proposition A.6 above one can remark it is enough to evaluate $\inf_{h\in H_n}(\mathbb{D}_h^2 + \mathbb{V}_h)$. Equation (A.31) entails we want to evaluate, up to a constant, the infimum over $h \in H_n$ of

$$h^{2\beta} + a_n \frac{1}{nh^{2d}\prod_{j=1}^d (\beta_n^j)^2} \leq h^{2\beta} + \frac{c\log n}{nh^{2d}}\frac{(\log n)^{2d}}{\prod_{j=1}^d \alpha_j^2}. \tag{A.39}$$

If we choose

$$h^*(n) := \left(\frac{(\log n)^{2d+1}}{n\prod_{j=1}^d \alpha_j^2}\right)^{\frac{1}{2(\beta+d)}}, \tag{A.40}$$

then we obtain the quantity $\left(\frac{(\log n)^{2d+1}}{n\prod_{j=1}^d \alpha_j^2}\right)^{\frac{2\beta}{2(\beta+d)}}$, which is the wanted rate. To conclude the proof we have to check that $h^*(n)$ as in (A.40) belongs to H_n. It is true as H_n has been constructed in analogy to $\mathcal{T}$, with $\frac{1}{h}$ playing the same role as $T^{(l)}$, for $l = 1, 2$. Indeed, following the same argumentation as in the proof of Theorem 4.11, as $\frac{n\prod_{j=1}^d \alpha_j^2}{(\log n)^{2d+1}} \geq 1$ and $\alpha_j \leq 1$, it is $\frac{1}{h^*(n)} \in [1, n]$. Then, by replacing $\frac{1}{h^*(n)}$ by its closest value in H_n we only modify its value in (A.39) by a constant, which provides

$$\inf_{h\in H_n}(\mathbb{D}_h^2 + \mathbb{V}_h) \leq c\left(\frac{(\log n)^{2d+1}}{n\prod_{j=1}^d \alpha_j^2}\right)^{\frac{2\beta}{2(\beta+d)}}.$$

It concludes the proof of Theorem 4.21.

□