Submitted to the Annals of Statistics

THE COST OF ADAPTATION UNDER DIFFERENTIAL PRIVACY: OPTIMAL ADAPTIVE FEDERATED DENSITY ESTIMATION

BY T. TONY CAI[1,a], ABHINAV CHAKRABORTY[2,b] AND LASSE VUURSTEEN[3,c]

[1]*Department of Statistics and Data Science,*
The Wharton School, University of Pennsylvania, [a]*tcai@wharton.upenn.edu*

[2]*Department of Statistics,*
Columbia University, [b]*ac4662@columbia.edu*

[3]*Department of Statistical Science,*
Duke University, [c]*lv121@duke.edu*

Privacy-preserving data analysis has become a central challenge in modern statistics. At the same time, a long-standing goal in statistics is the development of adaptive procedures—methods that achieve near-optimal performance across diverse function classes without prior knowledge of underlying smoothness or complexity. While adaptation is often achievable at no extra cost in the classical non-private setting, this naturally raises a fundamental question: to what extent is adaptation still possible under privacy constraints?

We address this question in the context of density estimation under federated differential privacy (FDP), a framework that encompasses both central and local DP models. We establish sharp results that characterize the cost of adaptation under FDP for both global and pointwise estimation, revealing fundamental differences from the non-private case. We then propose an adaptive FDP estimator that achieves explicit performance guarantees by introducing a new noise mechanism, enabling one-shot adaptation via post-processing. This approach strictly improves upon existing adaptive DP methods. Finally, we develop new lower bound techniques that capture the limits of adaptive inference under privacy and may be of independent interest beyond this problem.

Our findings reveal a sharp contrast between private and non-private settings. For global estimation, where adaptation can be achieved for free in the classical non-private setting, we prove that under FDP an intrinsic adaptation cost is unavoidable. For pointwise estimation, where a logarithmic penalty is already known to arise in the non-private setting, we show that FDP introduces an additional logarithmic factor, thereby compounding the cost of adaptation. Taken together, these results provide the first rigorous characterization of the adaptive privacy-accuracy trade-off.

1. Introduction. Privacy protection has become a critical concern in modern data analysis. Differential privacy (DP), introduced by [31], provides a rigorous mathematical framework that guarantees statistical analyses can be published without compromising the privacy of individual data subjects. Many differentially private statistical methods have since been developed, see for example [32], [1], and [33]. Among the various privacy protection frameworks available today, DP has emerged as particularly popular both for its theoretical foundations and its many applications across academic disciplines and industrial applications, see for example [35, 51, 40], and is finding applications within major technology companies such as Google, Amazon, Microsoft, and Apple, and governmental institutions such as the U.S. Census Bureau. Estimation and inference under differential privacy become a major focus

MSC2020 subject classifications: Primary 62G07, 62G20; secondary 62C20.
Keywords and phrases: Adaptation, Differential Privacy, Density Estimation, Minimax Rates.

in statistics and machine learning. Rigorous studies of privacy-utility trade-offs have established fundamental performance limits in diverse statistical settings. These include work on estimation [60, 38, 29, 30, 54, 22, 48, 15, 61], testing [2, 7, 50, 49], classification [25] and uncertainty quantification [42, 57].

For many complicated statistical problems, the performance of methods depends on unknown hyperparameters, which need to be tuned to unknown underlying regularity parameters. This is known as the problem of *adaptation*. For example, high-dimensional regression procedures typically require tuning to the sparsity level, while in nonparametric regression or density estimation, the smoothness of the underlying function is often unknown and methods must adapt to this unknown regularity. Such adaptation problems have been extensively studied and are well understood in the classical non-private setting, see for example [46, 27, 56, 47, 59, 20, 8].

In the context of differential privacy, however, adaptation poses a significant challenge and remains poorly understood. While adaptive DP procedures have been developed for various statistical problems -— including density estimation [10, 12, 43, 53], goodness-of-fit testing [45, 16], classification [5], hyperparameter tuning for stochastic gradient descent [52] and adaptation to sparsity in linear regression [62] -— all of these methods exhibit performance gaps between the optimal non-adaptive DP rate and the rate achieved by the adaptive DP procedure. To date, no specific adaptive lower bounds have been established to determine whether these performance gaps are an intrinsic consequence of adaptation under differential privacy constraints, or merely artifacts of current state-of-the-art adaptive procedures.

In this work, we address the fundamental question of adaptation under differential privacy constraints in the context of federated nonparametric density estimation. Density estimation serves as a canonical nonparametric problem with well-understood optimality theory in the non-private setting, that allows us to isolate the intrinsic cost of adaptation under privacy constraints. We consider both global risk (mean squared error) and pointwise risk, which capture fundamentally different aspects of adaptation and exhibit distinct phenomena. Moreover, density estimation has particular relevance for privacy applications, enabling synthetic data generation that preserves statistical properties while protecting individual privacy.

While most existing work on differential privacy focuses on either *central* differential privacy (CDP), where a trusted curator holds all data, or *local* differential privacy (LDP), where privacy is enforced at the individual level before data collection, our analysis uses the more general *federated differential privacy* (FDP) framework. FDP distributes data across multiple servers with privacy guarantees enforced at each server, generalizing both CDP (a single server with multiple observations) and LDP (many servers with one observation each). This unified framework makes our results applicable across the entire spectrum of privacy frameworks.

1.1. *Our contributions and related work.*

Our work addresses a critical gap in the literature by providing the first complete characterization of the fundamental cost of adaptation in differentially private estimation. While adaptation has been extensively studied in classical nonparametric statistics, the additional constraints imposed by differential privacy introduce new complexities that have remained theoretically unresolved.

The main contributions of our paper are as follows:

- We establish the first sharp minimax rates for adaptive density estimation under differential privacy constraints for both global and pointwise risk, fully characterizing the fundamental cost of adaptation under privacy for both LDP and CDP settings.
- We develop a new privacy method that enables optimal adaptive estimation. This mechanism provides DP guarantees while avoiding the variance inflation that typically plagues adaptive DP methods, allowing us to achieve the theoretical limits.

- We derive novel lower bound techniques that precisely quantify the fundamental cost of adaptation under differential privacy. Our results demonstrate that this cost manifests as an unavoidable logarithmic penalty relative to non-private adaptive estimation, resolving an open question in the literature.

Our findings reveal fundamental differences between private and non-private adaptation. In the classical setting, global risk adaptation can be achieved without cost, while our results show that differential privacy introduces an inherent penalty. For pointwise risk, we show that adaptation can incur additional costs beyond those already present in the non-private case. In particular, this closes the open question raised by [12] and [53] on whether such log-factors can be circumvented.

Existing work on adaptive density estimation under LDP includes methods based on Lepski-type procedures [43, 53, 55] and a wavelet thresholding approach [10]. However, these methods suffer from suboptimal rates because they either compute estimators for the worst-case regularity level or maintain multiple estimators for each regularity level considered. In both cases, the required privacy noise variance scales proportionally with the number of regularity levels, leading to rate-suboptimal performance. Our noise mechanism overcomes these limitations by carefully accounting for dependencies across different regularity levels, allowing us to achieve optimal adaptive rates with a single pass through the data while maintaining the same privacy guarantees.

Various minimax lower bound techniques in private settings have been developed using modifications of classical methods, such as Fano's inequality, Assouad's lemma, and Le Cam's method [44, 4], as well as van Trees-based bounds [6, 17]. Alternative approaches, including fingerprinting methods, tracing attack, and score attack, have also been explored in private learning theory [34, 23, 41]. Despite these advances, none of these methods have successfully characterized the exact cost of adaptation under differential privacy constraints. To the best of our knowledge, our work presents the first lower bound techniques that precisely quantifies the adaptation cost in the differentially private setting, establishing a fundamental limit on adaptive private estimation. While our optimal adaptive procedure is one-shot, our lower bound techniques are more broadly applicable to sequential protocols, showing that the cost of adaptation is unavoidable even in such interactive settings.

1.2. *Problem formulation.* We consider a federated setting with m servers, where server $j \in \{1, \dots, m\}$ holds n i.i.d. observations $X^{(j)} = (X_1^{(j)}, \dots, X_n^{(j)})$ drawn from an unknown density f on $[0,1]$, yielding $N = mn$ total samples. Each server computes a local transcript $T^{(j)}$ based on its data, and these transcripts are aggregated centrally to form estimators $\hat{f}$ or $\hat{T}$. We consider two estimation objectives: global risk $\mathbb{E}_f \|\hat{f} - f\|_2^2$ for estimating the entire density function, and pointwise risk $\mathbb{E}_f[(\hat{T} - f(t_0))^2]$ for estimating the density at a specific point $t_0 \in (0,1)$.

We shall assume that the true density f belongs to a Besov space $\mathcal{B}^{\alpha}_{p,q}$, which provides a flexible framework capturing diverse smoothness classes including Sobolev and Hölder spaces. Loosely speaking, $\mathcal{B}^{\alpha}_{p,q}$ contains functions with α derivatives, bounded in L_p-space, where the parameters (α, p, q) together control different aspects of regularity. We give a formal definition of Besov spaces in Section 3.1.

Crucially, optimal estimation rates depend on these smoothness parameters, but in practice they are unknown. The adaptive estimation challenge is to develop data driven procedures achieve optimal performance across a range of smoothness parameters, without requiring prior knowledge of (α, p, q).

Privacy constraints are formalized through federated differential privacy: Each server's transcript must satisfy differential privacy with respect to changes in its local dataset.

DEFINITION 1.1. *A distributed estimation protocol is* (ε, δ)-FDP *if, for each server* $j = 1, \ldots, m$*, its local transcript* $T^{(j)}$ *satisfies*

$$\mathbb{P}\left(T^{(j)} \in A \mid X^{(j)} = x\right) \leq e^{\varepsilon}\mathbb{P}\left(T^{(j)} \in A \mid X^{(j)} = x'\right) + \delta,$$

for any pair of local datasets $x, x' \in [0,1]^n$ *differing in at most one observation, and for any measurable subset* A *in the range of* $T^{(j)}$*, where the probability is over the randomness of the local mechanism.*

The parameters $\varepsilon > 0$ and $\delta \in [0, 1)$ control privacy strength, with smaller values providing stronger guarantees. The FDP framework encompasses CDP (which is recovered when $m = 1$) and LDP (when $n = 1$) as special cases. Let $\mathcal{F}^{\varepsilon,\delta}$ and $\mathcal{T}^{\varepsilon,\delta}$ denote the classes of all (ε, δ)-FDP protocols producing estimators in $L_2[0,1]$ and $\mathbb{R}$, respectively. The FDP framework for density estimation is illustrated in Figure 1.

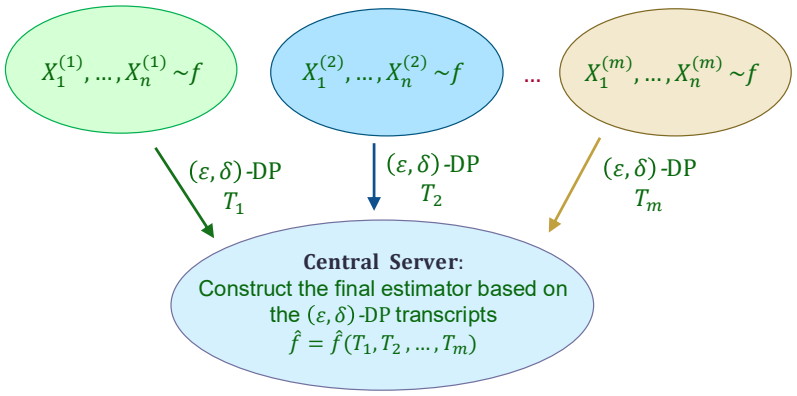

FIG 1. *An illustration of federated density estimation under DP.*

To understand the challenge of adaptation under privacy constraints, it is instructive to first consider the non-adaptive case where smoothness parameters are assumed to be known. When (α, p, q) are given, [17] established the minimax estimation rates for both risk types:

$$\inf_{\hat{f} \in \mathcal{F}^{\varepsilon,\delta}} \sup_{f \in \bar{\mathcal{B}}_{p,q}^{\alpha,R}} \mathbb{E}_f \|\hat{f} - f\|_2^2 \asymp N^{-\frac{2\alpha}{2\alpha+1}} + \left(mn^2\varepsilon^2\right)^{-\frac{2\alpha}{2\alpha+2}} =: \rho^2_{\text{MSE,NON-ADAPTIVE}}(\alpha), \tag{1.1}$$

$$\inf_{\hat{T} \in \mathcal{T}^{\varepsilon,\delta}} \sup_{f \in \bar{\mathcal{B}}_{p,q}^{\alpha,R}} \mathbb{E}_f[(\hat{T} - f(t_0))^2] \asymp N^{-\frac{2\nu}{2\nu+1}} + \left(mn^2\varepsilon^2\right)^{-\frac{2\nu}{2\nu+2}} =: \rho^2_{\text{POINT,NON-ADAPTIVE}}(\alpha, p),$$

where $\nu = \alpha - 1/p$ and $\bar{\mathcal{B}}_{p,q}^{\alpha,R}$ denotes the Besov ball of radius $R > 0$ intersected with the space of probability densities on $[0,1]$.

For both risk types, the rate consists of a non-private term and a privacy cost that dominates when ε is small. The term representing the non-private rate – which is of the form $N^{-\frac{2r}{2r+1}}$ – is the rate achievable when privacy is not a concern (e.g., $\varepsilon = \infty$). The second term – which is of the form $(mn^2\varepsilon^2)^{-\frac{2r}{2r+2}}$ – represents the additional cost of adaptation due to privacy constraints, which becomes substantial whenever $\varepsilon \ll n^{-\frac{r}{2r+1}} m^{\frac{1}{2(2r+1)}}$; where the privacy cost begins to dominate the classical statistical estimation error. Notably, when a fixed number of data points $N = mn$ is distributed across sufficiently many servers m, the privacy cost always dominates.

The rates depend critically on Besov parameters α and p: smaller α (less smooth densities) makes estimation harder. The parameter p affects the pointwise problem through $\nu = \alpha - 1/p$. The pointwise problem is potentially more difficult than the global problem for small values of p. For example, in the the Sobolev space $\mathcal{B}^{\alpha}_{2,2}$, where $\nu = \alpha - 1/2$. For Hölder smooth densities – corresponding to $\mathcal{B}^{\alpha}_{\infty,\infty}$ – $\nu = \alpha$ and the problems of global and pointwise estimation are equally difficult in terms of α.

Since these rates and optimal procedures of [17] depend on the (in practice) unknown parameters (α, p, q), adaptation becomes essential. This leads to a fundamental question: *When (α, p, q) are unknown, to what extent can the minimax rates in (1.1) still be attained?*

Our objective is to characterize the optimal rates under federated privacy constraints in the adaptive setting, where the regularity of the underlying function is unknown. Formally, given a collection $\mathcal{I}$ of smoothness parameter values, we seek to understand for which functions $(\alpha, p) \mapsto \rho^2_{\alpha,p;m,n,\varepsilon}$ the following adaptive minimax risks can be attained:

$$\inf_{\hat{f}\in\mathcal{F}^{\varepsilon,\delta}} \sup_{(\alpha,p,q)\in\mathcal{I}} \sup_{f\in\bar{\mathcal{B}}^{\alpha,R}_{p,q}} \frac{\mathbb{E}_f\|\hat{f}-f\|_2^2}{\rho^2_{\text{MSE}}(\alpha;m,n,\varepsilon)} \quad \text{and} \quad \inf_{\hat{T}\in\mathcal{T}^{\varepsilon,\delta}} \sup_{(\alpha,p,q)\in\mathcal{I}} \sup_{f\in\bar{\mathcal{B}}^{\alpha,R}_{p,q}} \frac{\mathbb{E}_f[(\hat{T}-f(t_0))^2]}{\rho^2_{\text{POINT}}(\alpha,p;m,n,\varepsilon)}.$$

That is, we seek estimators that automatically adapt to unknown smoothness, achieving (near-)optimal performance across all parameter values in $\mathcal{I}$. In the non-private setting, adaptation to unknown smoothness incurs no additional cost for global risk, but pointwise estimation pays a logarithmic penalty: the optimal adaptive rate becomes $(\frac{\log N}{N})^{\frac{2\nu}{2\nu+1}}$, see e.g. [13, 14]. Our central question is whether differential privacy fundamentally alters these adaptation costs.

1.3. *Organization of the paper.* The remainder of this paper is organized as follows. Section 2 presents our main theoretical results, including sharp minimax rates for both global and pointwise risks under adaptive differential privacy constraints. Section 3 develops our optimally adaptive estimation procedures, introducing the novel noise mechanism and wavelet thresholding estimator that achieve the theoretical limits. Section 4 derives the fundamental lower bounds that characterize the unavoidable cost of adaptation under differential privacy. Section 5 concludes with a discussion of our findings and directions for future research. Proofs and auxiliary results are collected in the Supplementary Material [18].

1.4. *Notation, definitions and assumptions.* For positive sequences a_k, b_k, we write $a_k \lesssim b_k$ if $a_k \le C b_k$ for some universal constant C, and $a_k \asymp b_k$ if $a_k \lesssim b_k$ and $b_k \lesssim a_k$. We denote $a_k \ll b_k$ when $a_k/b_k = o(1)$. We use $a \vee b$ and $a \wedge b$ for the maximum and minimum of a and b, respectively. The ℓ_p-norm of $v \in \mathbb{R}^d$ is $\|v\|_p$. Throughout, c and C are universal constants that may change from line to line.

2. Main results. Our central finding is that differential privacy fundamentally alters the cost of adaptation in nonparametric estimation. While classical statistics shows that adaptation to unknown smoothness can be achieved without penalty for global risk, we prove that differential privacy introduces an unavoidable logarithmic deterioration in the privacy-dominated regime.

This is formalized in our first main result, which characterizes the global risk when smoothness parameters are unknown. Recall that $N = mn$ is the total number of data points.

THEOREM 2.1. *Let $\delta \ll \varepsilon^2/m$ and define*

$$\rho^2_{\text{MSE}}(\alpha) \equiv \rho^2_{\text{MSE}}(\alpha, m, n, \varepsilon) = N^{-\frac{2\alpha}{2\alpha+1}} + \left(\frac{mn^2\varepsilon^2}{\log(N)}\right)^{-\frac{2\alpha}{2\alpha+2}}. \tag{2.1}$$

For any $p \geq 2$, $q \geq 1$, $\alpha_{\max} > \alpha_{\min} > 1/2$, it holds that

$$\inf_{\hat{f} \in \mathcal{F}^{\varepsilon,\delta}} \sup_{\alpha \in (\alpha_{\min}, \alpha_{\max})} \sup_{f \in \bar{\mathcal{B}}_{pq}^{\alpha,R}} \mathbb{E}_f \|\hat{f} - f\|_2^2 \, \rho_{\text{MSE}}^{-2}(\alpha) \asymp 1. \tag{2.2}$$

The theorem establishes that $\rho_{\text{MSE}}^2(\alpha)$ in (4.6) is the sharp adaptive rate. Comparing to the non-adaptive rate $N^{-2\alpha/(2\alpha+1)} + (mn^2\varepsilon^2)^{-2\alpha/(2\alpha+2)}$, adaptation introduces the logarithmic factor $\log N$ in the denominator of the privacy term, making the privacy cost strictly worse. This penalty is unavoidable: no estimator can achieve better than constant normalized risk across smoothness levels in any interval $(\alpha_{\min}, \alpha_{\max})$.

The result follows from matching upper and lower bounds. The lower bound is established in Section 4, while our constructive upper bound (Theorem 3.1 in Section 3) provides an explicit $(\varepsilon, 0)$-FDP estimator that achieves the rate in (4.6). This demonstrates that the larger class of (ε, δ)-FDP protocols offers no improvement over $\delta = 0$ protocols for the adaptive problem.

Our second main result characterizes the pointwise estimation problem, for which the situation is more complex than the global case. Even in the non-private setting, adaptation to unknown smoothness incurs a logarithmic penalty, yielding the rate $(\log N/N)^{2\nu/(2\nu+1)}$ compared to the non-adaptive rate $N^{-2\nu/(2\nu+1)}$. Under differential privacy, this adaptation cost is either 'equal' or compounded by additional logarithmic deterioration in the privacy-dominated regime, depending on the number of servers m in relation to the total number of samples N.

Our characterization, formally given in the theorem below, shows that the sharp adaptive rate for pointwise estimation is given by:

$$\rho_{\text{POINT}}^2(\nu) := \rho_{\text{POINT}}(\nu; m, n, \varepsilon) = \left(\frac{N}{\log N}\right)^{-\frac{2\nu}{2\nu+1}} + \left(\frac{mn^2\varepsilon^2}{L_{m,N}}\right)^{-\frac{2\nu}{2\nu+2}}, \tag{2.3}$$

where $\nu = \alpha - 1/p$ and

$$L_{m,N} = \begin{cases} \frac{\log^2 N}{m} & \text{if } m \leq \log N, \\ \log N & \text{if } m > \log N. \end{cases} \tag{2.4}$$

Compared to the non-adaptive pointwise rate $N^{-2\nu/(2\nu+1)} + (mn^2\varepsilon^2)^{-2\nu/(2\nu+2)}$, adaptation introduces logarithmic penalties in both terms: $\log N$ appears in the denominator of the first term (extending the non-private adaptation cost), while $L_{m,N}$ creates additional deterioration in the privacy term that depends on the server configuration.

THEOREM 2.2. *Let $\delta \ll \varepsilon/(N \log N)$. Consider any collection $\mathcal{I} \subset (0, \infty)^2$ with $\alpha - 1/p > 1/2$ for all $(\alpha, p) \in \mathcal{I}$ and $(\alpha, p), (\alpha', p') \in \mathcal{I}$ such that $\alpha - 1/p > \alpha' - 1/p'$.*

It holds that

$$\inf_{\hat{T} \in \mathcal{T}^{\varepsilon,\delta}} \sup_{(\alpha,p) \in \mathcal{I}} \sup_{f \in \bar{\mathcal{B}}_{p,q}^{\alpha,R}} \mathbb{E}_f (\hat{T} - f(t_0))^2 \, \rho_{\text{POINT}}^{-2}(\alpha - 1/p) \asymp 1.$$

Contrasting the above theorem with Theorem 2.1 reveals a fundamental difference between global and pointwise adaptation under privacy. While global risk suffers only from the logarithmic deterioration in the privacy term, pointwise estimation incurs logarithmic penalties in both the statistical and privacy components. The factor $L_{m,N}$ captures an interesting elbow-effect: when $m \leq \log N$ (few servers), the privacy cost degrades by an additional logarithmic factor, but when $m > \log N$ (many servers), this extra penalty disappears. In particular, this implies that the adaptation cost for the pointwise problem under LDP is (in relative terms) less severe than for the CDP setting, as highlighted in Corollaries 2.1 and 2.2 below.

As with the global risk case, this result follows from matching upper and lower bounds established in Sections 3 and 4, showing that the rate in (2.3) is both necessary and achievable. Furthermore, also the pointwise adaptive rate remains valid for the larger class of chained sequentially interactive DP protocols, eventhough the rate is attained by a one-shot $(\varepsilon, 0)$-FDP protocol.

2.1. *Special cases: Central and local differential privacy.* Our general federated framework encompasses the classical central and local DP settings as special cases. To better understand the implications of our results for each of these respective extremes of the federated spectrum, we now examine them separately, which reveal different behaviors for the cost of adaptation.

The first corollary considers the cost of adaptation in the CDP setting, where a single trusted server holds all N samples.

COROLLARY 2.1. *Consider the CDP setting, where a single server holds all $n = N$ samples ($m = 1$). Let $\delta \ll \varepsilon/(N \log N)$.*

It holds that

(2.5)
$$\inf_{\hat{f} \in \mathcal{F}^{\varepsilon,\delta}} \sup_{\alpha \in (\alpha_{\min}, \alpha_{\max})} \sup_{f \in \bar{\mathcal{B}}_{p,q}^{\alpha,R}} \mathbb{E}_f \|\hat{f} - f\|_2^2 \left[N^{-\frac{2\alpha}{2\alpha+1}} + \left(\frac{N^2\varepsilon^2}{\log N} \right)^{-\frac{2\alpha}{2\alpha+2}} \right]^{-1} \asymp 1,$$

(2.6)
$$\inf_{\hat{T} \in \mathcal{T}^{\varepsilon,\delta}} \sup_{(\alpha,p) \in \mathcal{I}} \sup_{f \in \bar{\mathcal{B}}_{p,q}^{\alpha,R}} \mathbb{E}_f (\hat{T} - f(t_0))^2 \left[\left(\frac{N}{\log N} \right)^{-\frac{2\nu}{2\nu+1}} + \left(\frac{N^2\varepsilon^2}{\log^2 N} \right)^{-\frac{2\nu}{2\nu+2}} \right]^{-1} \asymp 1,$$

for any $\alpha_{\max} > \alpha_{\min} > 0$ and collection $\mathcal{I} \subset (0,\infty)^2$ for which there exists $(\alpha, p), (\alpha', p') \in \mathcal{I}$ such that $\alpha - 1/p \neq \alpha' - 1/p'$ and $\alpha - 1/p > 1/2$ for all $(\alpha, p) \in \mathcal{I}$.

In the central case, adaptation costs manifest as an additional logarithmic factor in the denominator of privacy term, while the non-private adaptive statistical rates remain unchanged. Specifically, in the regime where privacy dominates, the minimax global risk incurs a logarithmic penalty relative to the non-adaptive private rate, while the pointwise risk suffers logarithmic penalties in both the statistical and privacy components, but the private penalty is a squared logarithm.

Turning to the other extreme of the federated framework: LDP, where each server holds a single observation. We have the following corollary characterizing the cost of adaptation in this setting.

COROLLARY 2.2. *Consider the LDP setting ($n = 1$ sample per server and $m = N$ total servers). Let $\delta \ll \varepsilon/(N \log N)$.*

The adaptation risks satisfy

(2.7)
$$\inf_{\hat{f} \in \mathcal{F}^{\varepsilon,\delta}} \sup_{\alpha \in (\alpha_{\min}, \alpha_{\max})} \sup_{f \in \bar{\mathcal{B}}_{p,q}^{\alpha,R}} \mathbb{E}_f \|\hat{f} - f\|_2^2 \left(\frac{N\varepsilon^2}{\log N} \right)^{-\frac{2\alpha}{2\alpha+2}} \asymp 1,$$

(2.8)
$$\inf_{\hat{T} \in \mathcal{T}^{\varepsilon,\delta}} \sup_{(\alpha,p) \in \mathcal{I}} \sup_{f \in \bar{\mathcal{B}}_{p,q}^{\alpha,R}} \mathbb{E}_f (\hat{T} - f(t_0))^2 \left(\frac{N\varepsilon^2}{\log N} \right)^{\frac{2\nu}{2\nu+2}} \asymp 1,$$

for any $\alpha_{\max} > \alpha_{\min} > 0$ and collection $\mathcal{I} \subset (0,\infty)^2$ for which there exists $(\alpha, p), (\alpha', p') \in \mathcal{I}$ such that $\alpha - 1/p \neq \alpha' - 1/p'$ and $\alpha - 1/p > 1/2$ for all $(\alpha, p) \in \mathcal{I}$.

In the case of LDP, the privacy constraints are so stringent that they completely dominate the estimation error, eliminating the non-private rate terms entirely, as was found in e.g. [17]. However, regarding adaptation costs, we observe another striking asymmetry when comparing to the CDP case: for pointwise risk, the adaptation cost matches that of the non-private case with no additional privacy-induced penalty. This behavior contrasts sharply with the central case, where privacy consistently worsens adaptation costs for both risk types.

The relative adaptation costs – measured as the factor by which rates deteriorate compared to their non-adaptive private counterparts – are thus less severe under LDP than CDP, despite LDP yielding uniformly worse absolute rates. This counterintuitive phenomenon arises from the distributed nature of local privacy mechanisms. To sketch an intuition for why this happens: DP procedures add some form of noise to (some statistic of) the data. To ensure DP, the noise needs to be, in an appropriate sense, 'sufficiently heavy-tailed'. When adapting to unknown smoothness, each of the m servers can add independent noise to its transcript. Aggregating these m noisy transcripts effectively produces a distribution for the 'aggregated noise' with lighter tails than the original noise distributions. This improved tail behavior can be exploited to obtain a performance gain when methods have to be adaptive. We note that this gain is relative to the otherwise more severe privacy costs inherent to the LDP model. In contrast, the CDP setting offers no such aggregation benefits. All N samples are held by a single server, and the privacy noise added to the transcript cannot be averaged across multiple sources. This forces CDP mechanisms to either accept heavier-tailed noise or maintain lighter tails at the cost of increased variance, both of which exacerbate adaptation costs relative to the local setting.

3. Optimally adaptive and differentially private procedures. In this section, we develop a differentially private estimator that achieves the optimal adaptive rate established in Theorems 2.1 and 2.2. We begin by introducing the Besov space framework and the wavelet thresholding estimator, which is a standard technique for adaptive estimation in non-private settings. We then describe the limitations of the Laplace and Gaussian mechanisms for thresholding in the private setting, motivating the need for a novel noise distribution that balances tail decay and sensitivity in an optimal way. This novel noise distribution and the corresponding optimal thresholding estimator is then introduced in Section 3.3. In order to contextualize the method properly, we start with a recap introducing wavelet estimation in Besov space in Section 3.1 and cover private estimation when smoothness is known in Section 3.2.

3.1. *Wavelet thresholding for estimation in Besov space.* We start by giving a formal definition of a Besov space. For a function $f \in L_p[0,1]$, the K-th order difference operator $\Delta_h^{(K)}$ with step size $h > 0$ and integer $K > \alpha$ maps f to a function that equals $\sum_{k=0}^{K}(-1)^{K-k}\binom{K}{k} f(t+kh)$ when $t \in [0, 1-Kh]$ and zero elsewhere. The Besov space $\mathcal{B}_{p,q}^{\alpha}$ is defined as the set of all $f \in L_p[0,1]$ such that the *Besov norm*, defined as

$$\|f\|_{\alpha,p,q} := \begin{cases} \|f\|_p + \left(\int \left[h^{-\alpha} \|\Delta_h^{(K)} f\|_p \right]^q \frac{dh}{h} \right)^{1/q} & \text{if } q < \infty, \\ \|f\|_p + \sup_{h>0} h^{-\alpha} \|\Delta_h^{(K)} f\|_p & \text{if } q = \infty, \end{cases} \tag{3.1}$$

is finite. Loosely speaking, α measures the degree of smoothness and $2 \le p \le \infty$ and $1 \le q \le \infty$ specify the norm used to measure the size of its derivatives. Besov spaces contain many important function classes, such as the Sobolev spaces ($p = q = 2$), the Hölder spaces ($p = q = \infty$) and the space of functions with bounded variation.

Wavelets are known to have many favourable properties when using them for function estimation in classical settings, see for example [28, 37, 13]. Under DP constraints, wavelet constructions have other desirable properties: they allow for exact control of the estimator's

sensitivity to changes in the data. Loosely speaking, this allows us to control the "influence" each individual observation has on the outcome of the estimator, whilst retaining the information the full sample has to a large extent.

Consider compactly supported, A-regular wavelets. Wavelet bases allow characterization of Besov spaces, with α, p, and q capturing the decay of wavelet coefficients.

Following the Daubechies' construction, for any $A \in \mathbb{N}$ one can obtain an A-regular 'father' wavelet $\phi(\cdot)$ with support on $[0, 2A-1]$, and a 'mother' $\psi(\cdot)$ wavelet with A vanishing moments and support on $[-A+1, A]$. We refer to [26] for details. The basis functions are then obtained as dilations and translations of these functions:

$$\left\{\phi_r, \psi_{lk} : r \in \{1, ..., 2^{l_0}\}, \quad l \geq l_0, \quad k \in \{1, ..., 2^l\}\right\},$$

with $\psi_{lk}(x) = 2^{l/2}\psi(2^l x - k)$, and $\phi_k(x) = 2^{l_0/2}\phi(2^{l_0}x - k)$. For a large enough primary resolution level l_0 and appropriate treatment of the boundary, the wavelet basis forms an orthonormal basis of $L_2[0,1]$. Hence, any function $f \in L_2[0,1]$ can be represented as

$$f = \sum_{k=1}^{2^{l_0}} f_{0k}\phi_{lk} + \sum_{l=l_0}^{\infty}\sum_{k=1}^{2^l} f_{lk}\psi_{lk}, \tag{3.2}$$

where $f_{0k} := \int f(x)\phi_{lk}(x)dx$ and $f_{lk} := \int f(x)\psi_{lk}(x)dx$ are the wavelet coefficients. Roughly speaking, the part of the wavelet decomposition corresponding to ϕ_k approximates the 'low-frequency' or 'global' part of the function, while the part corresponding to ψ_{lk} captures the 'high-frequency' or 'local' part.

We describe non-private wavelet based estimation first. Define $\hat{f}_{0k} = N^{-1}\sum_{i=1}^N \phi_{lk}(X_i)$ and $\hat{f}_{lk} = N^{-1}\sum_{i=1}^N \psi_{lk}(X_i)$. A *truncated wavelet estimator* $\hat{f}$ is then defined as follows:

$$\hat{f}(t) = \sum_{k=1}^{2^{l_0}} \hat{f}_{0k}\,\phi_k(t) + \sum_{l=l_0}^{L}\sum_{k=1}^{2^l} \hat{f}_{lk}\,\psi_{lk}(t).$$

It is a common knowledge that the above estimator achieves the non-private minimax rate for the global risk if L is chosen as $L = \lceil 1/(2\alpha+1)\log(N)\rceil$, which requires knowledge of α. Similarly, for the pointwise risk, the estimator $\hat{f}(t_0)$ achieves the minimax rate if L is chosen as $L = \lceil 1/(2\nu+1)\log(N)\rceil$, requireing knowledge of both α and p.

In the non-private setting with α and p unknown, a successful approach to constructing rate optimal adaptive estimators is through *wavelet thresholding*. Here, the wavelet coefficients are grouped together in blocks and thresholded at an appropriate level.

More specifically, consider blocks $B_{lj} = k : jb_l \leq k \leq (j+1)b_l$ with $b_l \in \mathbb{N}$ such that the B_{lj}'s partition $1, \ldots, 2^l$ for $j \in \mathcal{J}_l := 1, \ldots, 1 \wedge (2^l/b_l)$ and b_l an integer such that $2^l/b_l \in \mathbb{N}$. Consider the *b_l-block soft-thresholding operator* $\eta_\tau^{b_l} : \mathbb{R}^{b_l} \to \mathbb{R}^{b_l}$ defined as

$$\eta_\tau^{b_l}(y) = \left(1 - \frac{\tau}{\|y\|_2}\right)_+ y, \tag{3.3}$$

and write $v_{lB_{lj}} = (v_{lk})_{k\in B_{lj}}$ for the elements of a vector $v = \{v_{lk} : l = 1, \ldots, L^*, k = 1, \ldots, 2^l\}$ in the block B_{lj}. Wavelet (block) soft-thresholding estimators are of the form

$$\hat{f}^{\mathrm{WT}}(t) = \sum_{k=1}^{2^{l_0}} \hat{f}_{0k}\,\phi_k(t) + \sum_{l=l_0}^{L^*}\sum_{j\in\mathcal{J}_l} \left(\psi_{lB_{lj}}(t)\right)^\top \eta_\tau^{b_l}\left(\hat{f}_{lB_j}\right). \tag{3.4}$$

For $b_l = 1$ for all l, this corresponds to term-by-term thresholding estimator of [28], and are known to be minimax adaptive for the pointwise risk [14] for the choice of threshold

$\tau = \sqrt{2\log(N)/N}$. For the global risk, block thresholding with $b_l \asymp \log(N)$ is known to be minimax adaptive [13].

The idea behind thresholding, is that the wavelet coefficients decay at a rate that depends on the smoothness of the function. By thresholding the estimated wavelet coefficients, roughly speaking, one can remove the coefficient estimates that are dominated by noise, and retain the large coefficients, for which the coefficient estimates surpass the threshold. In addition, by grouping the coefficients in blocks, a borrowing strength effect is achieved, which allows for optimal performance in the global risk. We note that for pointwise risk, such grouping in blocks is not needed [13, 14].

3.2. *Optimal FDP wavelet estimators when smoothness is known.* When the smoothness parameters α and p are known, optimal private density estimation under federated differential privacy can be achieved using the Laplace mechanism applied to wavelet coefficients, by using truncated wavelet estimator approach (without a thresholding step). In the FDP setting, each server $j = 1, \dots, m$ computes empirical wavelet coefficients from its local data $X^{(j)} = (X_1^{(j)}, \dots, X_n^{(j)})$:

$$\hat{f}_{0r}^{(j)} = \frac{1}{n}\sum_{i=1}^{n} \phi_r(X_i^{(j)}), \quad \hat{f}_{lk}^{(j)} = \frac{1}{n}\sum_{i=1}^{n} \psi_{lk}(X_i^{(j)}), \tag{3.5}$$

for $r = 1, \dots, 2^{l_0}$ and (l,k) with $l_0 < l \le L$, $k = 1, \dots, 2^l$.

Each server then adds calibrated Laplace noise to its local coefficients before transmitting them to the aggregator. The noise scale is determined by the L_1-*sensitivity* of the wavelet coefficients: the maximum change in L_1-norm when altering one observation. For compactly supported, A-regular wavelets, this sensitivity is bounded by $c_{\phi,\psi} 2^{L/2}/n$ for some wavelet-dependent constant $c_{\phi,\psi} > 0$. Server j transmits the noisy transcript:

$$T^{(j)} = \left\{ \hat{f}_{0r}^{(j)} + W_{0r}^{(j)}, \hat{f}_{lk}^{(j)} + W_{lk}^{(j)} : r = 1, \dots, 2^{l_0},\ l_0 < l \le L,\ k = 1, \dots, 2^l \right\}, \tag{3.6}$$

where $W_{0r}^{(j)}, W_{lk}^{(j)} \sim \text{Laplace}(c_{\phi,\psi} 2^{l/2}/(n\varepsilon))$ are independent. The final estimator aggregates these noisy transcripts:

$$\hat{f}^{\text{Lap}}(t) = \sum_{r=1}^{2^{l_0}} \left(\hat{f}_{0r} + \frac{1}{m}\sum_{j=1}^{m} W_{0r}^{(j)} \right) \phi_r(t) + \sum_{l=l_0}^{L} \sum_{k=1}^{2^l} \left(\hat{f}_{lk} + \frac{1}{m}\sum_{j=1}^{m} W_{lk}^{(j)} \right) \psi_{lk}(t). \tag{3.7}$$

With appropriate choice of truncation level L (again depending on the smoothness), this $(\varepsilon, 0)$-FDP estimator achieves the optimal non-adaptive rates established in [17].

However, similarly to the non-private setting, this approach fails when smoothness is unknown. Adaptive estimation requires data-driven selection of the truncation level L, typically through thresholding of the aggregated noisy coefficients. Naively applying soft-thresholding to the Laplace noise-peturbed aggregated coefficients leads to suboptimal performance. Even sophisticated approaches such as the L_1-norm rescaling methods of [10] achieve suboptimal rates.

The fundamental issue, loosely speaking, lies in the fact that current privacy mechanisms do not account for dependencies present across the resolution levels. Privacy mechanisms such as the Laplace mechanism add effectively independent noise to each wavelet coefficient. However, this independent treatment ignores the joint structure present in the wavelet coefficients. The resulting noise, while individually calibrated, creates an unfavorable trade-off between variance and tail behavior, hampering the concentration properties essential for adaptive procedures such as thresholding.

To achieve optimal adaptive rates under FDP constraints, we introduce a novel noise mechanism that accounts for dependence accross resolution levels, whilst aiming at optimal tail decay to facilitate thresholding.

3.3. *Private thresholding: finding the right noise geometry.* The oracle inequality for soft-thresholding reveals why standard privacy mechanisms struggle with adaptive estimation.

We motivate the choice of noise distribution by considering the following oracle inequality for the (block) soft-thresholding error. For a proof, see [24], Lemma 6.

LEMMA 3.1. *Consider $f \in \mathbb{R}^d$, a random vector Z taking values in $\mathbb{R}^d$, and let $Y = f + Z$. Let $\eta_\tau^d : \mathbb{R}^d \to \mathbb{R}^d$ be the soft-thresholding operator for threshold $\tau > 0$. Then,*

$$\mathbb{E}\|\eta_\tau^d(Y) - f\|_2^2 \le \min\{\|f\|_2^2, 4\tau^2\} + 4\mathbb{E}\|Z\|_2^2 \mathbb{1}\{\|Z\|_2 > \tau\}. \tag{3.8}$$

The oracle inequality decomposes the thresholding error into bias and variance components. The term $\min\{\|f\|_2^2, 4\tau^2\}$ represents a bias-variance trade-off: smaller thresholds τ reduce bias for small signals but increase sensitivity to noise. The tail term $4\mathbb{E}\|Z\|_2^2\mathbb{1}\{\|Z\|_2 > \tau\}$ heavily depends on the noise distribution of Z, in particular its tail decay around the threshold τ. Ignoring the presence of statistical noise, this term captures the contribution of privacy noise to the thresholding error.

For optimal thresholding, we want the smallest possible threshold τ such that the tail term remains controlled. This requires noise Z with two key properties: (1) variance scaling appropriately with the sensitivity of the underlying statistic, and (2) rapid tail decay to ensure $\mathbb{P}(\|Z\|_2 > \tau)$ decreases quickly as τ increases. Standard privacy mechanisms fail to achieve both properties simultaneously. The Laplace mechanism does not achieve this balance (even after applying sophisticated rescaling techniques as in [10]), and while the Gaussian mechanism has good tail decay properties, it requires larger variance to achieve the same privacy guarantees for small values of δ.

Our solution is based on using a carefully designed norm that exploits the multiscale structure of wavelets, and employing the exponential mechanism with respect to that norm, akin to the K-norm mechanism of [39]. This approach allows us to achieve the desired balance between sensitivity and tail decay.

Define the *oscillation* of a function $g \in L_2[0,1]$ as $\operatorname{osc}(g) := \sup_{t\in[0,1]} g(t) - \inf_{t\in[0,1]} g(t)$. The *multiscale-oscillation norm* of the wavelet coefficient vectors is defined as

$$\|u\|_{V_L} = \sup_{\substack{g\in\operatorname{span}\{\psi_{lk}:(l,k)\in V_L\} \\ \operatorname{osc}(g)\le 1}} \left| \left\langle \sum_{(l,k)\in V_L} u_{lk}\psi_{lk}, g \right\rangle_{L^2[0,1]} \right|, \tag{3.9}$$

where for $l_0, L \in \mathbb{N}$, $V_L = \{(l,k) : l = l_0, \dots, L, k = 1, \dots, 2^l\}$. Let $s_L = \sum_{l=l_0}^{L} 2^l$. This norm has the crucial property that its sensitivity with respect to changing one data point is exactly $1/n$, independent of the resolution levels considered.

LEMMA 3.2. *For any datasets $x, x' \in [0,1]^n$ that differ in at most one coordinate, say $x_i \neq x_i'$, and let $\Delta_{x,x'}$ denote the coordinate-wise difference of the corresponding empirical wavelet coefficients in (3.5). In multi-scale-oscillation norm it satisfies the following bound:*

$$\|\Delta_{x,x'}\|_{V_L} \le \frac{1}{n}.$$

PROOF. Let f_Δ denote the inverse wavelet transform of the vector $\Delta_{x,x'}$. We compute

$$\|\Delta_{x,x'}\|_{V_L} = \sup_{\substack{g\in\operatorname{span}\{\psi_{lk}\} \\ \operatorname{osc}(g)\le 1}} |\langle f_\Delta, g\rangle_{L^2}|.$$

Consider the reproducing kernel $K_L(s,t) = \sum_{(l,k)\in V_L} \psi_{lk}(s)\psi_{lk}(t)$, such that for $g \in \operatorname{span}\{\psi_{lk} : (l,k) \in V_L\}$, it holds that

$$g(s) = \sum_{(l,k)\in V_L} \psi_{lk}(s) \langle \psi_{lk}, g\rangle_{L^2} = \langle K_L(s,\cdot), g\rangle_{L^2}.$$

By linearity of the integral,

$$\begin{aligned}\langle f_\Delta, g\rangle_{L^2} &= \frac{1}{n}\int_0^1 \left(\sum_{(l,k)\in V_L} (\psi_{lk}(x_i) - \psi_{lk}(x_i'))\psi_{lk}(t)\right) g(t)\,dt \\ &= \frac{1}{n}\left[\langle K_L(x_i,\cdot), g\rangle_{L^2} + \left\langle K_L(x_i',\cdot), g\right\rangle_{L^2}\right] = \frac{1}{n}\left[g(x_i) - g(x_i')\right].\end{aligned}$$

As $|g(x_i) - g(x_i')| \le \operatorname{osc}(g) \le 1$, it follows that $\|\Delta_{x,x'}\|_{V_L} \le \frac{1}{n}$. □

The multiscale-oscillation norm can be used to generate noise through the exponential mechanism: $V^{(j)} = (V_{lk}^{(j)})_{l,k}$ with density proportional to the map $v \mapsto \exp(-\varepsilon n\|v\|_{V_L})$. This yields noisy coefficients

$$T_{lk}^{(j)} = \hat{f}_{lk}^{(j)} + V_{lk}^{(j)}, \quad (l,k) \in V_L, \tag{3.10}$$

that are $(\varepsilon, 0)$-differentially private. Averaging these noisy coefficients across servers yields a private estimate of the wavelet coefficients:

$$\hat{f}_{lk}^{\mathrm{PW}} = \frac{1}{m}\sum_{j=1}^m T_{lk}^{(j)} = \hat{f}_{lk} + \bar{V}_{lk},$$

where $\bar{V}_{lk} = m^{-1}\sum_{j=1}^m V_{lk}^{(j)}$. As with thresholding in the non-private case, we choose a a sufficiently large truncation level $L^* = \lceil \log(N)\rceil$, and then apply (block) soft-thresholding to the wavelet coefficients as a post-processing step to obtain the final estimator. This final post-processing differs depending on whether we consider pointwise or global risk.

3.3.1. *Adaptive global risk post-processing step.* For global risk estimation, we employ block thresholding to leverage borrowing-of-strength effects across wavelet coefficients. We partition the coefficients at each resolution level l into blocks $B_{lj} = \{k : jb_l \le k \le (j+1)b_l\}$ for $j \in \mathcal{J}_l := \{1, \dots, \lfloor 2^l/b_l\rfloor\}$, where b_l is the integer closest to $\lceil \log N\rceil$ such that $2^l/b_l$ is an integer.

The adaptive estimator applies soft-thresholding to each block of noisy wavelet coefficients:

$$\hat{f}^{\mathrm{BTPW}}(t) = \sum_{r=1}^{2^{l_0}} \hat{f}_{0r}^{\mathrm{PW}}\phi_r(t) + \sum_{l=l_0}^{L^*}\sum_{j\in\mathcal{J}_l} \psi_{lB_j}^T(t)\,\eta_{\tau_l}^{b_l}\left(\hat{f}_{lB_{lj}}^{\mathrm{PW}}\right), \tag{3.11}$$

where $\psi_{lB_j}(t) = (\psi_{lk}(t))_{k\in B_{lj}}$ and $\hat{f}_{lB_{lj}}^{\mathrm{PW}} = (\hat{f}_{lk}^{\mathrm{PW}})_{k\in B_{lj}}$ denotes the vector of aggregated noisy coefficients in block B_{lj}.

What remains is to choose thresholds that balance bias and variance for both statistical and privacy noise. We set:

$$\tau_l = \sqrt{\kappa_1\frac{\log N}{N} + \kappa_2\frac{2^l\log^2 N}{mn^2\varepsilon^2}}, \tag{3.12}$$

where $\kappa_1, \kappa_2 > 0$ are wavelet-dependent constants. The first term corresponds to the statistical noise threshold, while the second term accounts for the privacy noise, with the additional $\log N$ factor reflecting the adaptation cost.

The threshold choice is motivated by the following tight tailbound for our multiscale-oscillation noise mechanism.

LEMMA 3.3. *Let* $\bar{V}_{lk} = \frac{1}{m}\sum_{j=1}^{m} V_{lk}^{(j)}$ *be the aggregated privacy noise and* $\bar{V}_{lB_j} = (\bar{V}_{lk})_{k\in B_{lj}}$. *For sufficiently large* $\kappa > 0$ *and all* $l \in \{l_0, \dots, L^*\}$, $j \in \mathcal{J}_l$,

$$\mathbb{E}\|\bar{V}_{lB_{lj}}\|_2^2 \mathbb{1}\left\{\|\bar{V}_{lB_{lj}}\|_2 \geq \kappa \frac{2^{l/2} b_l}{\sqrt{m} n\varepsilon}\right\} \leq C \frac{2^l b_l^2}{mn^2\varepsilon^2} e^{-b_l}$$

for a constant $C > 0$.

The exponential decay in the block size b_l shown in Lemma 3.3 is crucial for controlling the tail term in the oracle inequality and enables the use of small thresholds necessary for optimal adaptation.

Analysis of the multiscale-oscillation noise is considerably more complex than for standard Laplace or Gaussian noise. An important step in the analysis is that the multiscale-oscillation noise mechanism admits a tractable decomposition as a Gamma-distributed 'radius' and a uniformly distributed 'direction' on the norm's unit ball:

$$(V_{lk}^{(j)})_{(l,k)\in V_{L^*}} \overset{d}{=} D^{(j)} U^{(j)}, \quad D^{(j)} \sim \Gamma\left(s_{L^*} + 1, \frac{1}{n\varepsilon}\right), \quad U^{(j)} \sim \mathcal{U}\left(\{u : \|u\|_{V_{L^*}} \leq 1\}\right),$$

where $s_{L^*} = \sum_{l=l_0}^{L^*} 2^l$ is the total number of detail coefficients, see e.g. [39]. Exploiting this decomposition, the analysis of the tail behavior boils down to of the averaged $U_{lB_j}^{(j)}$. This is possible by combining stochastic domination arguments where we compare with more tractable uniform distributions on convex compacts. Complete technical details of this analysis are provided in Section A.1 in the Supplementary Material.

Combining the oracle inequality with the exponential tail bounds yields the following performance guarantee for our estimator.

THEOREM 3.1. *For any* α, p, q *with* $0 < \alpha < A$, $p \geq 2$, *and* $q \geq 1$,

$$\sup_{f\in\mathcal{B}_{p,q}^{\alpha}(R)} \mathbb{E}_f\|\hat{f}^{\mathrm{BTPW}} - f\|_2^2 \lesssim N^{-\frac{2\alpha}{2\alpha+1}} + \left(\frac{\log N}{mn^2\varepsilon^2}\right)^{\frac{2\alpha}{2\alpha+2}}.$$

Since the estimator is a post-processing of an $(\varepsilon, 0)$-FDP protocol, it inherits the same privacy guarantee. The theorem shows that the estimator achieves the optimal adaptive rate for the global risk simultaneously over all Besov spaces $\mathcal{B}_{pq}^{\alpha}(R)$ with $0 < \alpha < A$, $p \geq 2$ and $q \geq 1$. Taking A large enough, we see that the estimator attains the optimal adaptive rate of Theorem 2.1. We defer the proof to Section A.2 of the Supplementary Material.

REMARK 3.1. As an alternative to soft-thresholding, one can consider the hard-thresholding operator. The estimator is then given by

$$\check{f}^{\mathrm{BTPW}}(t) = \sum_{r=1}^{2^{l_0}} \hat{f}_{0r}^{\mathrm{PW}} \phi_r(t) + \sum_{l=l_0}^{L^*} \sum_{j\in\mathcal{J}_l} \mathbb{1}\left\{\|\hat{f}_{lB_{lj}}^{\mathrm{PW}}\|_2 \geq \tau_l\right\} \psi_{lB_j}^T(t) \hat{f}_{lB_{lj}}^{\mathrm{PW}}. \tag{3.13}$$

A similar proof yields the same upper bound as in Theorem 3.1.

3.3.2. *Adaptive pointwise risk estimation.* For pointwise risk with unknown α and p, we use term-by-term thresholding (setting $b_l = 1$ for all l). While block thresholding can achieve optimal adaptive pointwise rates in the non-private setting [13], term-by-term thresholding offers superior performance under the multiscale-oscillation norm privacy mechanism. We expand on the reason for this below, in Remark 3.2.

The pointwise estimator is given by

$$\hat{f}^{\mathrm{TTPW}}(t_0) = \sum_{r=1}^{2^{l_0}} \hat{f}_{0r}^{\mathrm{PW}} \phi_r(t_0) + \sum_{l=l_0}^{L^*} \sum_{k=1}^{2^l} \eta_{\tau_l}\left(\hat{f}_{lk}^{\mathrm{PW}}\right) \psi_{lk}(t_0), \tag{3.14}$$

where $\eta_{\tau_l}(\cdot)$ denotes the soft-thresholding operator applied coordinate-wise.

The pointwise setting requires different threshold calibration. We set:

$$\tau_l = \sqrt{\kappa_{1\psi} \frac{\log N}{N} + \kappa_{2\psi} \frac{2^l L_{m,N}}{mn^2\varepsilon^2}}, \tag{3.15}$$

where $\kappa_{1\psi}, \kappa_{2\psi} > 0$ are constants depending on the choise of wavelets and

$$L_{m,N} = \begin{cases} \log(N), & m \geq \log(N), \\ \frac{\log^2(N)}{m}, & m < \log(N). \end{cases}$$

The factor $L_{m,N}$ captures how server distribution affects noise tail behavior around the threshold. When $m \geq \log N$, averaging across many servers transforms the tail decay from sub-exponential (for individual servers) to sub-Gaussian rates, reducing the adaptation penalty from up to $\log^2 N$ to $\log N$. When $m < \log N$, averaging effects are insufficient and the sub-exponential tails necessitate larger thresholds. This is captured by the following tail bound for the pointwise setting, for which we defer its proof to Section A.1 of the Supplementary Material.

LEMMA 3.4. *Consider* $\bar{V}_{lk} = \frac{1}{m}\sum_{j=1}^m V_{lk}^{(j)}$ *for* $(l,k) \in V_{L^*}$. *For all* $l \in \{l_0, \ldots, L^*\}$, $k = 1, \ldots, 2^l$, *and* $t \geq c_0 \frac{2^{l/2}}{\sqrt{mn}\varepsilon}$,

$$\mathbb{E}\bar{V}_{lk}^2 \mathbb{1}\left\{|\bar{V}_{lk}| \geq t\right\} \lesssim \left(t^2 + \frac{2^l}{mn^2\varepsilon^2}\right) \exp\left(-c_1 m \min\left\{\frac{t^2 n^2 \varepsilon^2}{2^l}, \frac{tn\varepsilon}{2^{l/2}}\right\}\right)$$

for constants $c_0, c_1 > 0$.

The next theorem confirms that the pointwise estimator achieves optimal adaptive rates. Its proof follows by the oracle inequality in Lemma 3.1 (applied with $d = 1$) combined with the tail bound in Lemma 3.4, and is provided in Section A.2.

THEOREM 3.2. *Let* $t_0 \in (0,1)$ *be given and consider the estimator* $\hat{T} := \hat{f}^{\mathrm{TTPW}}(t_0)$. *For any* $p \in [2, \infty]$, $q \in [1, \infty]$, *and* α *such that* $\nu := \alpha - 1/p > 1/2$ *and* $\alpha < A$, *it holds that*

$$\sup_{f \in \mathcal{B}_{p,q}^{\alpha}(R)} \mathbb{E}_f |\hat{T} - f(t_0)|^2 \lesssim \left(\frac{\log N}{N}\right)^{\frac{2\nu}{2\nu+1}} + \left(\frac{L_{m,N}}{mn^2\varepsilon^2}\right)^{\frac{2\nu}{2\nu+2}}.$$

The theorem confirms that pointwise adaptive estimation incurs logarithmic penalties in both the statistical term (changing $N^{-2\nu/(2\nu+1)}$ to $(N/\log N)^{-2\nu/(2\nu+1)}$, as in the non-private setting) and the privacy term (through $L_{m,N}$).

REMARK 3.2. When $m = 1$ (the CDP setting), the plug-in block thresholding estimator $\hat{f}^{\mathrm{BTPW}}(t_0)$ performs equally well as the term-by-term approach, since averaging across machines provides no reduction in the heaviness of the tails of the privacy noise.

4. Deriving the adaptation lower bounds. This section establishes fundamental lower bounds that characterize the unavoidable cost of adaptation under differential privacy constraints. We prove that any differentially private estimator must pay additional logarithmic penalties beyond the classical non-private adaptation cost, and this penalty is inherent to the privacy requirement rather than an artifact of any particular estimation procedure.

Our analysis separates the pointwise and global estimation settings, which each exhibit different behavior and require different proof strategies. The pointwise case, treated in Section 4.1, requires novel modifications to classical constrained risk inequality techniques. The global case, covered in Section 4.2, is proven through a dimension-free Fisher information bound that captures the interaction between adaptation and privacy across multiple regularity levels.

Our impossibility results extend beyond the one-shot FDP protocols used in our upper bounds. The lower bounds apply to sequential protocols, which are known to show improvement in certain LDP settings (see e.g. [3, 12, 11] and references therein). In sequential protocols, servers communicate in a single round through a chain (server j sends to server $j+1$, who sends to server $j+2$, etc.), allowing each server to condition on message from a previous server. Hence, our lower bounds demonstrate that the logarithmic adaptation penalties are fundamental limitations that cannot be circumvented even when servers can leverage information from earlier steps in the communication chain.

For completeness, we formalize the sequential FDP setting below: servers communicate in a chain where server j receives message $T^{(j-1)}$ from the previous server and computes $T^{(j)}$ based on its local data $X^{(j)}$ and the received message. Each step must satisfy local differential privacy with respect to the server's own data.

DEFINITION 4.1 (Sequential FDP). *A sequential distributed protocol is (ε, δ)-FDP if each server j's transcript $T^{(j)}$, conditioned on any fixed input $T^{(j-1)} = t$, satisfies: for datasets $x, x' \in [0,1]^n$ differing in one observation,*

$$\mathbb{P}\left(T^{(j)} \in A \mid X^{(j)} = x, T^{(j-1)} = t\right) \leq e^{\varepsilon}\mathbb{P}\left(T^{(j)} \in A \mid X^{(j)} = x', T^{(j-1)} = t\right) + \delta.$$

We note that, unlike previous work, we do not require the conditional distributions of the transcripts to be dominated. This condition, which mandates that the distribution of the current transcript—conditional on the data and all previous transcripts—be dominated by a reference measure for every possible realization of the conditioning variables, is typically assumed in the literature studying global risk metrics (e.g. L_2-error) [6, 17, 61], either explicitly or implicitly by restricting the transcript space to take values in a countable space. In Section B.1.1, we demonstrate that our lower bound techniques remain valid without these restrictive assumptions.

In the remainder of this section, for pointwise risk, let $\mathcal{T}(\varepsilon, \delta)$ include sequential (ε, δ)-FDP protocols, and let $\mathcal{F}(\varepsilon, \delta)$ denote the corresponding class for global risk. We note that this forms a strictly larger class than the one-shot protocols considered in Definition 1.1.

4.1. *Adaptation lower bound for pointwise estimation.* Our approach builds upon the classical constrained risk inequality line of thought developed in [9], which establishes adaptation costs by showing that improved performance at one function necessarily creates worse performance elsewhere in the parameter space.

Differential privacy fundamentally alters such an analysis in two important ways. First, privacy constraints limit the distinguishability between probability measures, making total variation distance a natural distance to capture distinguishability, as observed by for example [54]. Second, the standard change-of-measure techniques used in classical proofs must be

replaced with privacy-specific methods that capture how the randomness introduced by the privacy mechanism affects distinguishability.

Our key technical contribution is the a constrained risk inequality for federated differential privacy, which we derive in a general setting in Section 4.1.1, in the form of Lemma 4.1. This lemma has general consequences for super-efficient private estimation which might be of independent interest (see for an illustration Example 4.1). The lower bound on the pointwise risk is established in Section 4.1.2 by combining Lemma 4.1 with arguments specific to the adaptive setting.

4.1.1. *A differentially private constrained risk inequality.* We formalize the super-efficiency trade-off between differential privacy and risk in the following lemma, which we state here in the simpler $(\varepsilon, 0)$-FDP case. The general (ε, δ)-FDP version appears in Section B.2 of the Supplementary Material, together with its proof.

LEMMA 4.1. *Consider a model $\{P_f : f \in \Theta\}$ on $(\mathcal{X}, \mathscr{X})$ indexed by a semi-metric space (Θ, d), and $f, g \in \Theta$ such that $\mathrm{d}(f, g) \geq \Delta$ for some $\Delta > 0$. Consider servers $j = 1, \ldots, m$ each with i.i.d. samples $X_1^{(j)}, \ldots, X_n^{(j)}$ with distribution P_h for $h \in \Theta$.*

If an $(\varepsilon, 0)$-FDP estimation protocol $\hat{T}$ on the basis of $(X_i^{(j)})_{i=1,\ldots,n}^{j=1,\ldots,m}$ satisfies

$$\mathbb{E}_f \, \mathrm{d}(\hat{T}, f)^2 \leq \gamma^2 \Delta^2 \quad \textit{for some } \gamma > 0,$$

then

$$\mathbb{E}_g \, \mathrm{d}(\hat{T}, g)^2 \geq \frac{\Delta^2}{4} \left[1 - 2 \exp\left(m(\bar{\varepsilon} \wedge \bar{\varepsilon}^2) + \log \gamma \right) \right],$$

where $\bar{\varepsilon} = 6n\varepsilon \| P_f - P_g \|_{TV}$.

The lemma captures the fundamental trade-off imposed by differential privacy: when an estimator achieves risk $\gamma^2 \Delta^2$ under P_f (where γ quantifies the improvement factor at f), its performance under the alternative measure P_g degrades necessarily; unless the privacy constraint is weak (large ε) or the measures are far apart in total variation. This degradation is proportional to the logarithm of the improvement factor γ.

The lemma hence establishes a private version of a constraint risk inequality: it states that super-efficiency at a particular point in the parameter space comes at a cost in terms of increased risk at other points in the parameter space. We exemplify this in the following simple example.

EXAMPLE 4.1. Consider $(\varepsilon, 0)$-differentially private estimation of a population proportion p over a neighborhood of $1/2$, on the basis of observations $Y_1^{(j)}, \ldots, Y_n^{(j)} \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(p)$ for $j = 1, \ldots, m$, where ε, m, n are allowed to have the asymptotics as introduced in Section 1.2. The minimax rate for the problem is easy to derive: after adding Laplace noise, the server averages can be privately communicated, where averaging over the servers results in an $(\varepsilon, 0)$-FDP estimator attaining the rate $(mn)^{-1} + (mn^2\varepsilon^2)^{-1}$. Known techniques (for instance, via Lemma 4.1) yield a matching lower bound, confirming the optimality of this rate.

However, Lemma 4.1 provides a more nuanced insight: it demonstrates that super-efficiency at a single point incurs a penalty elsewhere. To illustrate this, consider $\hat{T}$ to be an $(\varepsilon, 0)$-FDP estimator that is *super-efficient* at $p = 1/2$; meaning that $\mathbb{E}_{1/2} |\hat{T} - 1/2|^2 \lesssim (mn^2\varepsilon^2)^{-C} + (mn)^{-C}$ for some $C > 1$ (we construct such an estimator in Section B.2.4 of

the Supplementary Material). Lemma 4.1 then implies that for any fixed neighborhood B of $1/2$,

$$\sup_{p\in B} \mathbb{E}_p|\hat{T}-p|^2 \gtrsim \frac{L_{m,N}}{mn^2\varepsilon^2}\left(1-2e^{-c\log(N)}\right)$$

for some constant $c>0$, where $L_{m,N}$ is the elbow-effect factor from (2.4). To see this, one can apply the lemma with $p_{m,n,\varepsilon}=1/2+c'\sqrt{L_{m,N}/(mn^2\varepsilon^2)}$ for sufficiently small $c'>0$, noting that the total variation distance between $\mathrm{Ber}(1/2)$ and $\mathrm{Ber}(p_{m,n,\varepsilon})$ scales as $|p_{m,n,\varepsilon}-1/2|$.

This example shows that super-efficient estimators performs worse than the minimax rate by a logarithmic factor $L_{m,N}$ over the neighborhood B. Crucially, this it mirrors the elbow effect inherent to adaptation under federated privacy constraints: the elbow effect in super-efficiency penalties manifest differently across the privacy spectrum from central to local models. For CDP ($m=1$), the penalty is $\log^2 N$, whereas for LDP ($n=1$), it is only $\log N$.

Armed with Lemma 4.1, we proceed with the lower bound for the pointwise risk.

4.1.2. *The pointwise risk lower bound.* Theorem 4.1 below provides a fundamental super-efficiency lower bound for pointwise density estimation under differential privacy that. Corollary 4.1 distills the result to the adaptive setting, which combined with the upper bound in Theorem 3.2, precisely characterizes the unavoidable cost of adaptation. It generalizes the lower bound result stated earlier in Theorem 2.2 and provides the technical foundation for understanding why the additional logarithmic adaptation penalty is inherent to the privacy constraint.

THEOREM 4.1. *Let $m,n\in\mathbb{N}$ and $N:=mn$. Fix a sequence $A_N\geq e$ and an (ε,δ)-FDP estimator $\hat{T}$ with*

$$\delta \ll \frac{\varepsilon}{mn\,A_N}. \tag{4.1}$$

Suppose there exists $f_0\in\mathcal{B}^\alpha_{p,q}(R')$ with $R'<R$ and $f_0(t_0)>0$ such that

$$\mathbb{E}_{f_0}\left(\hat{T}-f_0(t_0)\right)^2 \lesssim \frac{1}{A_N}\left\{N^{-\frac{2\nu}{2\nu+1}}\vee(mn^2\varepsilon^2)^{-\frac{2\nu}{2\nu+2}}\right\}=o(1). \tag{4.2}$$

Then,

$$\sup_{f\in\mathcal{B}^\alpha_{p,q}(R)}\mathbb{E}_f\left(\hat{T}-f(t_0)\right)^2 \gtrsim \left(\frac{N}{\log A_N}\right)^{-\frac{2\nu}{2\nu+1}}\vee\left(\frac{mn^2\varepsilon^2}{L_{m,N}}\right)^{-\frac{2\nu}{2\nu+2}}, \tag{4.3}$$

where $L_{m,N}:=\log A_N\left(1\vee\frac{\log A_N}{m}\right)$.

We highlight that the theorem is a consequence of Lemma 4.1, combined with arguments specific to the nonparametric setting which we postpone to Section B.2 of the Supplementary Material. The theorem establishes that any differentially private estimator that performs well at a specific function f_0 (better than the minimax rate by a factor of A_N) must necessarily perform worse on some other functions in the Besov ball. The better the performance at f_0, the worse the performance on other functions, where the loss is a logarithmic factor in A_N.

The fundamental trade-off posed by Theorem 4.1 has direct implications for adaptation across smoothness classes. Consider an estimator that achieves near-optimal rates for functions in a smoother Besov class $\mathcal{B}^\alpha_{p,q}$. Since such an estimator necessarily performs much

better than the minimax rate on functions that lie in the interior of a less smooth class $\mathcal{B}^{\alpha'}_{p',q'}$ (corresponding to a large improvement factor A_N), Theorem 4.1 implies it must pay a logarithmic penalty when estimating other functions in the less smooth class. This adaptivity trade-off is formalized in the following corollary, which shows that achieving optimal rates up to a polylogarithmic factor for one smoothness class forces a logarithmic adaptation penalty when estimating functions from any less smooth class.

COROLLARY 4.1. *Consider an estimator $\hat{T} \in \mathcal{T}^{\varepsilon,\delta}$ for $\delta \ll \varepsilon/\log(N)$ in the federated setting with $N = mn$ total samples across m servers. Suppose that, for some (α, p, q) such that $\nu := \alpha - 1/p > 1/2$ and $g \in \mathcal{B}^{\alpha,R}_{p,q}$ it holds that*

$$\mathbb{E}_g(\hat{T} - g(t_0))^2 \lesssim (\log N)^{O(1)} \left(\left(\frac{1}{N}\right)^{\frac{2\nu}{2\nu+1}} + \left(\frac{1}{mn^2\varepsilon^2}\right)^{\frac{2\nu}{2\nu+2}} \right). \tag{4.4}$$

Then, for any (α', p', q') such that $\nu' := \alpha' - 1/p' < \nu$, we have that

$$\sup_{f \in \mathcal{B}^{\alpha',R}_{p',q'}} \mathbb{E}_f(\hat{T} - f(t_0))^2 \gtrsim \left(\frac{N}{\log N}\right)^{-\frac{2\nu'}{2\nu'+1}} + \left(\frac{mn^2\varepsilon^2}{L_{m,N}}\right)^{-\frac{2\nu'}{2\nu'+2}}, \tag{4.5}$$

where $L_{m,N}$ is as defined in (2.4).

REMARK 4.1. Corollary 4.1 shows that the cost of adaptation is at least $\log N$ in the case of the Besov classes $\mathcal{B}^{\alpha_1}_{p_1,q_1}$ and $\mathcal{B}^{\alpha_2}_{p_2,q_2}$ with $\alpha_1 - 1/p > \alpha_2 - 1/p$. Under DP, the phenomenon of pointwise adaptive estimation where one can 'trade-off' regularity versus integrability of derivatives remains like in the non-private case (see Theorem 1 of [14]): Whenever $\alpha_1 - 1/p = \alpha_2 - 1/p$, no adaptive cost is paid. For example, adapting between a 2-smooth Sobolev ball and a $3/2$-smooth Hölder ball is possible without adaptive cost for the pointwise risk.

4.2. *Adaptation lower bound for the global risk.* We now turn to the lower bound for global risk adaptation. The theorem below establishes that any (ε, δ)-FDP estimator attempting to adapt across any range of smoothness levels $(\alpha_{\min}, \alpha_{\max})$ must incur the logarithmic penalty $\log N$ in the privacy term, uniformly across all smoothness levels. Unlike pointwise estimation, global risk exhibits a uniform logarithmic adaptation penalty: there is no benefit from distributing data across many servers.

THEOREM 4.2. *Assume $\delta \ll n\varepsilon^2/N$. Consider any $\alpha_{\min} > 1/2$ and $\alpha_{\max} > \alpha_{\min}$ and let*

$$\rho^2(\alpha) \equiv \rho^2_{\alpha,m,n,\varepsilon} = \left(\frac{1}{N}\right)^{\frac{2\alpha}{2\alpha+1}} + \left(\frac{\log N}{mn^2\varepsilon^2}\right)^{\frac{2\alpha}{2\alpha+2}}. \tag{4.6}$$

Then,

$$\inf_{\hat{f} \in \mathcal{F}(\varepsilon,\delta)} \sup_{\alpha \in (\alpha_{\min}, \alpha_{\max})} \sup_{f \in \mathcal{B}^{\alpha}_{p,q}(R)} \mathbb{E}_f \|\hat{f} - f\|_2^2 \, \rho(\alpha)^{-2} \gtrsim 1. \tag{4.7}$$

Although the estimators for both risk types are similar, the global risk setting exhibits fundamentally different behavior and requires fundamentally different techniques from the pointwise case. In the non-private setting, adaptation for global risk can be achieved without cost. Capturing the cost of adaptation under privacy constraints through a lower bound thus requires a novel argument.

Before giving a formal proof below, we sketch the argument of our technique. The central idea is to exploit the fact that under FDP constraints, the 'Fisher information induced by

an FDP protocol' within a finite-dimensional sub-model is, for well-behaved sub-models, dimension-free. This phenomenon was first observed in a non-adaptive LDP setting by [6], and subsequently in non-adaptive FDP settings in [17, 61]. While in these earlier works this property was used to characterize the cost of privacy for fixed regularity, the adaptive setting requires a more delicate construction: it is precisely this dimension-free structure, hitherto unexploited in this context, that enables our rate-optimal lower bound.

The first part of the construction is familiar (see e.g. [56]): we formulate a sub-model that partitions into further sub-models, each indexed by a resolution level L and consisting of random wavelet perturbations at that level. Due to the multiscale nature of wavelets, we can consider a grid of regularity values with cardinality of order $\log N$, where each regularity value corresponds to a different component of the partition. We then define a prior over the perturbations to ensure that the resulting densities belong to the appropriate Besov balls almost surely.

This construction allows us to invoke the van Trees inequality, which relates the 'adaptive Bayes risk' to the *minimum* transcript-induced Fisher information across the partition (Lemma 4.2). Intuitively, successful adaptation requires the protocol to retain sufficient information for every sub-model. While this relationship holds generally (that is; we have hitherto not invoked privacy specific arguments), the privacy constraint introduces a bottleneck: the *total* induced Fisher information is bounded by a dimension-free quantity (Lemma 4.3). With approximately $\log N$ sub-models sharing this limited information budget, the signal available for any single smoothness level is diluted. This dilution manifests as the $\log N$ penalty—the unavoidable cost of adaptation under privacy constraints.

PROOF OF THEOREM 4.2. Let $\mathcal{A} \subset (\alpha_{\min}, \alpha_{\max})$ such that for all $\alpha \in \mathcal{A}$,

$$\rho(\alpha) \leq 2\left(\frac{\log(N)}{mn^2\varepsilon^2}\right)^{-\frac{\alpha}{2\alpha+2}} \iff \rho(\alpha)^{-2} \geq \frac{1}{4}\left(\frac{\log(N)}{mn^2\varepsilon^2}\right)^{\frac{2\alpha}{2\alpha+2}} =: \tilde{\rho}_\alpha^{-2}. \tag{4.8}$$

If the complement of $\mathcal{A}$ in $(\alpha_{\min}, \alpha_{\max})$ is non-empty,

$$\inf_{\hat{f}\in\mathcal{F}(\varepsilon,\delta)} \sup_{\alpha\in\mathcal{A}^c} \sup_{f\in B^\alpha_{pq}(R)} \mathbb{E}_f\, N^{\frac{2\alpha}{2\alpha+1}} \|\hat{f} - f\|_2^2$$

lower bounds (4.7), and the latter quantity can be further lower bounded by a constant following standard arguments (e.g., using Assouad's lemma or Fano's inequality (ignoring privacy constraints); see [58], Chapter 2). In case $\mathcal{A}$ is empty as $N \to \infty$, the statement of the theorem follows.

Assume next that $\mathcal{A}$ is not empty. Since $\alpha \mapsto \rho(\alpha)$ is continuous, we can take $\mathcal{A}$ such that there exists an open neighborhood $\mathcal{A}_* \subset \mathcal{A}$ for which (4.8) holds for all $\alpha \in \mathcal{A}_*$. Without loss of generality, write $\mathcal{A}_* = (\alpha_{\min}, \alpha_{\max})$ and consider an (approximately) equispaced grid $\tilde{\mathcal{A}}$ of $(\alpha_{\min}, \alpha_{\max})$ of size $\lceil \log N \rceil$; such that for $\alpha_1, \alpha_2 \in \tilde{\mathcal{A}}$ with $\alpha_1 > \alpha_2$, we have $\alpha_1 - \alpha_2 \geq 1/(2\log N)$.

For each $\alpha \in \tilde{\mathcal{A}}$, let L_α be the integer closest to the solution of $2^{L(\alpha+1)} = \tilde{\rho}_\alpha^{-1}$ and consider the set $\mathcal{L} := \{L_\alpha : \alpha \in \tilde{\mathcal{A}}\}$. Recall that $N^{-\omega} \lesssim \varepsilon \lesssim 1$ for some $\omega \in [0, 1)$, which implies that $|\mathcal{L}| \asymp \log N$ and there exists a constant $c > 0$ such that $L \geq c \log N$ for all $L \in \mathcal{L}$.

Consider now the random $L_2[0,1]$-valued elements

$$F^U_{\mathcal{L}}(x) = 1 + \sum_{L\in\mathcal{L}} \sum_{k=1}^{2^L} U_{Lk}\psi_{Lk}(x) \quad x \in [0,1], \tag{4.9}$$

where the collection $U = \{U_L : L \in \mathcal{L}\}$ of coefficient vectors $U_L = (U_{Lk})_{k=1}^{2^L}$ is drawn from a 'prior' distribution supported on the hypercube

$$\mathcal{U} = \cup_{\alpha\in\tilde{\mathcal{A}}} [-C_R 2^{-L_\alpha(\alpha+1/2)}, C_R 2^{-L_\alpha(\alpha+1/2)}]^{2^{L_\alpha}}.$$

Here, ψ_{Lk} are the elements of a wavelet basis that is $A > \alpha_{\max}$-smooth and satisfies $\int \psi_{Lk} = 0$ for all k and $L \in \mathcal{L}$. Since $\alpha > 0$ and $\min_{L\in\mathcal{L}} L \to \infty$, $F_{\mathcal{L}}$ is a probability density over the full support of the prior whenever $mn^2\varepsilon^2$ is large enough. The following lemma establishes a lower bound on adaptive minimax risk in terms of the trace of the Fisher information of the submodels induced by $F_{\mathcal{L}}^U$, $U \in \mathcal{U}$.

LEMMA 4.2. *Consider $\hat{f} \in \mathcal{F}(\varepsilon, \delta)$, let $T = (T^{(j)})_{j=1,\dots,m}$ denote the corresponding FDP transcripts.*

There exists a distribution $\mathbb{P}^U$ of U over $\mathcal{U}$ such that (4.7) is lower bounded by

$$\left(\frac{\log(N)}{mn^2\varepsilon^2} \min_{\alpha\in\tilde{\mathcal{A}}} \mathbb{E}^U \mathit{Tr}(\mathcal{I}_{L_\alpha}(U)) + 1 \right)^{-1},$$

where $\mathcal{I}_{L_\alpha}(U) = \mathbb{E}_{F_{\mathcal{L}}^U} \mathbb{E}_{F_{\mathcal{L}}^U} [S_{L_\alpha}|T] \mathbb{E}_{F_{\mathcal{L}}^U} [S_{L_\alpha}|T]^\top$, with

$$S_L \equiv S_L(X_1, \dots, X_N) := \nabla_{(u_{Lk})_{k=1}^{2^L}} \sum_{j=1}^{m} \sum_{i=1}^{n} \log F_{\mathcal{L}}^u(X_i^{(j)}) \,|_{u=U},$$

and where $\mathbb{E}^U$ denotes the push-forward expectation with respect to the distribution $\mathbb{P}^U$.

A proof of the above lemma can be found in Section B.1 of the Supplementary Material. To see that the object $\mathcal{I}_{L_\alpha}(U)$ deserves the name *transcript-induced Fisher information matrix* and is in fact well-defined, we refer the reader to Section B.1.1.

We have that

$$\min_{\alpha\in\tilde{\mathcal{A}}} \mathbb{E}^U \mathrm{Tr}(\mathcal{I}_{L_\alpha}(U)) \le \frac{1}{|\mathcal{L}|} \sum_{L\in\mathcal{L}} \mathbb{E}^U \mathrm{Tr}(\mathcal{I}_{L_\alpha}(U)) = \frac{1}{|\mathcal{L}|} \mathbb{E}^U \mathrm{Tr}\left(\mathcal{I}_{\mathcal{L}}(U)\right), \tag{4.10}$$

where $\mathcal{I}_{\mathcal{L}}(U)$ denotes the 'full' induced Fisher information matrix corresponding to the model in (4.9):

$$\mathcal{I}_{\mathcal{L}}(U) = \mathbb{E}_{F_{\mathcal{L}}^U} \mathbb{E}_{F_{\mathcal{L}}^U} [S_{\mathcal{L}}|T] \mathbb{E}_{F_{\mathcal{L}}^U} [S_{\mathcal{L}}|T]^\top,$$

where $S_{\mathcal{L}} = \mathrm{vec}(S_L : L \in \mathcal{L})$ is the vector obtained by stacking the vectors S_L for $L \in \mathcal{L}$.

The proof now follows by combining the observation of (4.10) with the following data-processing inequality, which captures the rather striking phenomenon that the trace of the transcript-induced Fisher information matrix is free of dimension in the differential privacy setting.

LEMMA 4.3. *Whenever $\delta \ll n\varepsilon^2/N$, it holds that $\mathit{Tr}(\mathcal{I}_{\mathcal{L}}(u)) \le Cmn^2\varepsilon^2$ for some absolute constant $C > 0$ independent of n, m, δ and ε.*

We provide a proof of this lemma in Section B.1 of the Supplementary Material [18]. □

5. Discussion. The results presented in this paper provide a unified and comprehensive framework for understanding the cost of adaptation in federated differentially private (FDP) density estimation. Our analysis shows that privacy constraints introduce an intrinsic penalty to adaptive estimation—a subtle yet fundamental cost that distinguishes the private setting from its classical, non-private counterpart. Although logarithmic in magnitude (and thus modest compared with polynomial penalties), this adaptation cost is conceptually significant: it reflects a fundamental limitation imposed by privacy protection and has practical implications for valid statistical inference beyond point estimation. These findings open several avenues for further investigation in related statistical models and problem settings. For

instance, understanding the exact cost of adaptation is essential for constructing adaptive confidence intervals or confidence bands with guaranteed coverage, as the adaptation penalty directly determines the required bandwidth [19, 36, 21]. Extending our framework to such inferential tasks would clarify how privacy noise interacts with the uncertainty inherent in adaptation.

Beyond density estimation, the principles and tools developed here—particularly our lower bound techniques, noise mechanism design, and adaptive thresholding strategy—have the potential to inform a wide range of nonparametric and high-dimensional problems. In settings such as sparse regression or function estimation under shape constraints, adaptation to unknown sparsity or structural parameters poses analogous challenges. Applying or extending our methods in these contexts could lead to new privacy-preserving procedures that attain minimax-optimal or near-optimal adaptive performance.

Another promising direction concerns adaptation to more complex functional characteristics, such as spatial inhomogeneity, anisotropy, or locally varying smoothness. These features arise frequently in real-world applications and introduce dependencies that may amplify the privacy-adaptivity trade-off. Understanding whether such structure-dependent adaptation leads to more substantial costs under differential privacy could deepen our insight into statistical efficiency under privacy constraints.

Our findings also have implications for statistical tasks beyond estimation, including hypothesis testing and model assessment. Although recent work has shown that adaptive nonparametric goodness-of-fit testing under differential privacy is feasible [45, 16], the fundamental cost of adaptation in these testing problems remains largely open. Determining whether similar penalties arise—or whether privacy induces different trade-offs—would be an important direction for future research.

Taken together, our framework highlights the broader potential for studying adaptation under privacy constraints across diverse statistical models. By systematically characterizing the interplay between privacy, adaptability, and statistical efficiency, this work deepens our theoretical understanding of privacy-utility trade-offs and lays the groundwork for developing principled, adaptive, and privacy-preserving methodologies for modern data analysis.

Funding. The research of Tony Cai was supported in part by NSF grant NSF DMS-2413106 and NIH grants R01-GM123056 and R01-GM129781.

SUPPLEMENTARY MATERIAL

Supplementary Material to "The Cost of Adaptation under Differential Privacy: Optimal Adaptive Federated Density Estimation"

The supplementary material [18] contains proofs of the main results, technical lemmas, and additional details on the construction of the estimators and lower bounds.

REFERENCES

[1] ABADI, M., CHU, A., GOODFELLOW, I., MCMAHAN, H. B., MIRONOV, I., TALWAR, K. and ZHANG, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* 308–318. https://doi.org/10.1145/2976749.2978318

[2] ACHARYA, J., CANONNE, C., FREITAG, C. and TYAGI, H. (2019). Test without Trust: Optimal Locally Private Distribution Testing. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (K. CHAUDHURI and M. SUGIYAMA, eds.). *Proceedings of Machine Learning Research* **89** 2067–2076. PMLR.

[3] ACHARYA, J., CANONNE, C. L., LIU, Y., SUN, Z. and TYAGI, H. (2022). Interactive Inference Under Information Constraints. *IEEE Transactions on Information Theory* **68** 502–516. https://doi.org/10.1109/TIT.2021.3123905

[4] ACHARYA, J., SUN, Z. and ZHANG, H. (2021). Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory* 48–78. PMLR.

[5] AUDDY, A., CAI, T. T. and CHAKRABORTY, A. (2025). Minimax and adaptive transfer learning for nonparametric classification under distributed differential privacy constraints. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. Advance online publication. https://doi.org/10.1093/jrsssb/qkaf070

[6] BARNES, L. P., CHEN, W.-N. and ÖZGÜR, A. (2020). Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory* **1** 645–659. https://doi.org/10.1109/JSAIT.2020.3039461

[7] BERRETT, T. and BUTUCEA, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. In *Advances in Neural Information Processing Systems* (H. LAROCHELLE, M. RANZATO, R. HADSELL, M. F. BALCAN and H. LIN, eds.) **33** 3164–3173. Curran Associates, Inc.

[8] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous Analysis of Lasso and Dantzig Selector. *Annals of Statistics* **37** 1705–1732. https://doi.org/10.1214/08-AOS620 MR2533469

[9] BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *The annals of Statistics* **24** 2524–2535. https://doi.org/10.1214/aos/1032181166

[10] BUTUCEA, C., DUBOIS, A., KROLL, M. and SAUMARD, A. (2020). Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli* **26** 1727 – 1764. https://doi.org/10.3150/19-BEJ1165

[11] BUTUCEA, C., KLOCKMANN, K. and KRIVOBOKOVA, T. (2025). Nonparametric spectral density estimation using interactive mechanisms under local differential privacy. *arXiv preprint*. arXiv:2504.00919.

[12] BUTUCEA, C., ROHDE, A. and STEINBERGER, L. (2023). Interactive versus noninteractive locally differentially private estimation: Two elbows for the quadratic functional. *The Annals of Statistics* **51** 464 – 486. https://doi.org/10.1214/22-AOS2254

[13] CAI, T. T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *The Annals of Statistics* **27** 898–924. https://doi.org/10.1214/AOS/1018031262

[14] CAI, T. T. (2003). RATES OF CONVERGENCE AND ADAPTATION OVER BESOV SPACES UNDER POINTWISE RISK. *Statistica Sinica* **13** 881–902.

[15] CAI, T. T., CHAKRABORTY, A. and VUURSTEEN, L. (2024). Optimal Federated Learning for Functional Mean Estimation under Heterogeneous Privacy Constraints. *arXiv preprint*. arXiv:2412.18992.

[16] CAI, T. T., CHAKRABORTY, A. and VUURSTEEN, L. (2024). Federated Nonparametric Hypothesis Testing with Differential Privacy Constraints: Optimal Rates and Adaptive Tests. *arXiv preprint*. arXiv:2406.06749.

[17] CAI, T. T., CHAKRABORTY, A. and VUURSTEEN, L. (2024). Optimal Federated Learning for Nonparametric Regression with Heterogeneous Distributed Differential Privacy Constraints. *arXiv preprint*. arXiv:2406.06755.

[18] CAI, T. T., CHAKRABORTY, A. and VUURSTEEN, L. (2025). Supplementary Material to "The Cost of Adaptation under Differential Privacy: Optimal Adaptive Federated Density Estimation". Supplementary Material.

[19] CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *The Annals of statistics* **32** 1805-1840. https://doi.org/10.1214/009053604000000049

[20] CAI, T. T. and LOW, M. G. (2005). Adaptive estimation of linear functionals under different performance measures. *Bernoulli* **11** 341 – 358. https://doi.org/10.3150/bj/1116340298

[21] CAI, T. T., LOW, M. G. and MA, Z. (2014). Adaptive confidence bands for nonparametric regression functions. *Journal of the American Statistical Association* **109** 1054-1070. https://doi.org/10.1080/01621459.2013.879260

[22] CAI, T. T., WANG, Y. and ZHANG, L. (2021). The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics* **49** 2825–2850. https://doi.org/10.1214/21-AOS2058

[23] CAI, T. T., WANG, Y. and ZHANG, L. (2023). Score attack: A lower bound technique for optimal differentially private learning. *arXiv preprint*. arXiv:2303.07152.

[24] CAI, T. T. and ZHOU, H. H. (2010). Nonparametric Regression in Natural Exponential Families. In *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown*, (J. O. Berger, T. T. Cai and I. M. Johnstone, eds.). *IMS Collections* **6** 199–215. Institute of Mathematical Statistics. https://doi.org/10.1214/10-IMSCOLL614

[25] CHAUDHURI, K. and HSU, D. (2011). Sample Complexity Bounds for Differentially Private Learning. In *Proceedings of the 24th Annual Conference on Learning Theory* (S. M. KAKADE and U. VON LUXBURG, eds.). *Proceedings of Machine Learning Research* **19** 155–186. PMLR, Budapest, Hungary.

[26] DAUBECHIES, I. (1992). *Ten lectures on wavelets*. SIAM. https://doi.org/10.1137/1.9781611970104
[27] DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Association* **90** 1200–1224. https://doi.org/10.1080/01621459.1995.10476626
[28] DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *The annals of Statistics* **26** 879–921. https://doi.org/10.1214/aos/1024691081
[29] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local Privacy and Statistical Minimax Rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* 429–438. https://doi.org/10.1109/FOCS.2013.53
[30] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association* **113** 182–201. https://doi.org/10.1080/01621459.2017.1389735
[31] DWORK, C., MCSHERRY, F., NISSIM, K. and 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* 265–284. Springer. https://doi.org/10.1007/11681878_14
[32] DWORK, C. and SMITH, A. (2010). Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality* **1**. https://doi.org/10.29012/jpc.v1i2.570
[33] DWORK, C., SMITH, A., STEINKE, T. and ULLMAN, J. (2017). Exposed! a survey of attacks on private data. *Annual Review of Statistics and Its Application* **4** 61–84. https://doi.org/10.1146/annurev-statistics-060116-054123
[34] DWORK, C., SMITH, A., STEINKE, T., ULLMAN, J. and VADHAN, S. (2015). Robust traceability from trace amounts. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science* 650–669. IEEE.
[35] FICEK, J., WANG, W., CHEN, H., DAGNE, G. and DALEY, E. (2021). Differential privacy in health research: A scoping review. *Journal of the American Medical Informatics Association* **28** 2269–2276. https://doi.org/10.1093/jamia/ocab135
[36] GINÉ, E. and NICKL, R. (2010). Confidence Bands in Density Estimation. *The Annals of Statistics* **38** 1122–1170. https://doi.org/10.1214/09-AOS738
[37] HALL, P., KERKYACHARIAN, G. and PICARD, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statistica Sinica* 33–49.
[38] HALL, R., RINALDO, A. and WASSERMAN, L. (2013). Differential privacy for functions and functional data. *J. Mach. Learn. Res.* **14** 703–727.
[39] HARDT, M. and TALWAR, K. (2010). On the geometry of differential privacy. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*. *STOC '10* 705–714. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/1806689.1806786
[40] JIANG, B., LI, J., YUE, G. and SONG, H. (2021). Differential privacy for industrial internet of things: Opportunities, applications, and challenges. *IEEE Internet of Things Journal* **8** 10430–10451. https://doi.org/10.1109/JIOT.2021.3057419
[41] KAMATH, G., MOUZAKIS, A., REGEHR, M., SINGHAL, V. et al. (2025). A Bias-Accuracy-Privacy Trilemma for Statistical Estimation. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.2024.2443275
[42] KARWA, V. and VADHAN, S. (2018). Finite Sample Differentially Private Confidence Intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)* 44:1–44:9. https://doi.org/10.4230/LIPIcs.ITCS.2018.44
[43] KROLL, M. (2021). On density estimation at a fixed point under local differential privacy. *Electronic Journal of Statistics* **15** 1783 – 1813. https://doi.org/10.1214/21-EJS1830
[44] LALANNE, C., GARIVIER, A. and GRIBONVAL, R. (2023). On the statistical complexity of estimation and testing under privacy constraints. *Transactions on Machine Learning Research Journal*.
[45] LAM-WEIL, J., LAURENT, B. and LOUBES, J.-M. (2022). Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. *Bernoulli* **28** 579–600.
[46] LEPSKI, O. V. (1991). On a Problem of Adaptive Estimation in Gaussian White Noise. *Theory of Probability & Its Applications* **35** 454-466. https://doi.org/10.1137/1135065
[47] LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics* **25** 2512 – 2546. https://doi.org/10.1214/aos/1030741083
[48] LI, M., TIAN, Y., FENG, Y. and YU, Y. (2024). Federated Transfer Learning with Differential Privacy. *arXiv preprint*. arXiv:2403.11343.
[49] NARAYANAN, S. (2022). Private High-Dimensional Hypothesis Testing. In *Proceedings of Thirty Fifth Conference on Learning Theory* (P.-L. LOH and M. RAGINSKY, eds.). *Proceedings of Machine Learning Research* **178** 3979–4027. PMLR.

[50] NARAYANAN, S., MIRROKNI, V. and ESFANDIARI, H. (2022). Tight and robust private mean estimation with few users. In *International Conference on Machine Learning* 16383–16412. PMLR.

[51] PAN, K., ONG, Y.-S., GONG, M., LI, H., QIN, A. K. and GAO, Y. (2024). Differential privacy in deep learning: A literature survey. *Neurocomputing* 127663. https://doi.org/10.1016/j.neucom.2024.127663

[52] PAPERNOT, N. and STEINKE, T. (2022). Hyperparameter Tuning with Renyi Differential Privacy. In *International Conference on Learning Representations*.

[53] RANDRIANARISOA, T., STEINBERGER, L. and SZABÓ, B. (2025). Towards multi-purpose locally differentially-private synthetic data release via spline wavelet plug-in estimation. *arXiv preprint*. arXiv:2508.13969 [math.ST].

[54] ROHDE, A. and STEINBERGER, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *The Annals of Statistics* **48** 2646 – 2670. https://doi.org/10.1214/19-AOS1901

[55] SCHLUTTENHOFER, S. and JOHANNES, J. (2022). Adaptive pointwise density estimation under local differential privacy. *arXiv preprint*. arXiv:2206.07663 [math.ST].

[56] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics* **24**. https://doi.org/10.1214/aos/1032181163

[57] STEINBERGER, L. (2024). Efficiency in local differential privacy. *The Annals of Statistics* **52** 2139–2166. https://doi.org/10.1214/24-AOS2425

[58] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer. https://doi.org/10.1007/978-0-387-79052-7

[59] TSYBAKOV, A. B. (1998). Pointwise and Sup-norm Sharp Adaptive Estimation of Functions on the Sobolev Classes. *Annals of Statistics* **26** 2420–2469. https://doi.org/10.1214/aos/1024691478 MR1700239

[60] WASSERMAN, L. and ZHOU, S. (2010). A Statistical Framework for Differential Privacy. *Journal of the American Statistical Association* **105** 375–389. https://doi.org/10.1198/jasa.2009.tm08651

[61] XUE, G., LIN, Z. and YU, Y. (2024). Optimal Estimation in Private Distributed Functional Data Analysis. *arXiv preprint*. arXiv:2412.06582.

[62] ZHANG, Z., NAKADA, R. and ZHANG, L. (2024). Differentially Private Federated Learning: Servers Trustworthiness, Estimation, and Statistical Inference. *arXiv preprint*. arXiv:2404.16287 [stat.ML].