Fisher Information Contraction and van Trees Inequalities for Fully Interactive Private Federated Estimation

T. Tony Cai and Yicheng Li

Abstract

Federated differentially private (DP) protocols may involve many adaptive rounds and reuse the same local samples. Existing fully interactive lower bounds treat each client's private input as a single observation (local DP) and do not characterize the dependence on the number of locally held samples under sample-level privacy. We establish a clientwise matrix Fisher information contraction and the resulting matrix federated van Trees bound for parameter estimation under squared loss based on any complete public transcript satisfying a clientwise sample-level zero-concentrated differential privacy constraint. The matrix bound preserves the directional Fisher geometry of heterogeneous local experiments by constraining each client's transcript contribution both by its raw Fisher information and by a whitened privacy budget. In homogeneous experiments, trace compression yields a simpler federated van Trees bound based on the trace. Together with existing upper bounds for the corresponding problems, these results identify minimax rates for several statistical problems, including mean estimation, linear regression, Bradley–Terry learning on comparison graphs, and nonparametric regression, over the full class of interactive public transcript protocols. For these problems, arbitrary public interaction and repeated sample reuse do not improve the rate over simpler restricted protocols.

Index Terms

Differential privacy, federated learning, Fisher information, minimax lower bounds, van Trees inequality.

I. Introduction

Vast quantities of data are now generated by individuals, businesses, and governments, making the protection of individuals represented in statistical analyses and machine learning systems increasingly important. Differential privacy (DP) [1, 2] provides a rigorous framework for quantifying the privacy guarantees of randomized algorithms. The canonical formulation is the *central* DP model, in which a trusted curator has access to the full dataset and releases an output satisfying a prescribed differential privacy guarantee.

In many modern applications, however, data are distributed across multiple clients, institutions, or devices, and the underlying local datasets cannot be pooled directly because of privacy, legal, ownership, communication, or infrastructure constraints. Differentially private analysis in this decentralized setting is commonly referred to as *federated DP* or *differentially private federated learning*. A fundamental statistical question is then to characterize the privacy–utility tradeoff [3]: for a prescribed level of privacy protection, how much statistical accuracy must necessarily be sacrificed?

Federated learning has become an important paradigm for collaborative statistical learning from decentralized data. Examples include biomedical studies across hospitals, mobile and edge-device learning, financial and commercial data analysis across organizations, multi-institution scientific studies, and privacy-sensitive public-sector applications. In these problems, the goal is not merely to protect a centrally pooled dataset, but to enable learning across many data holders while respecting client-level constraints, heterogeneous sample sizes, different privacy requirements, and limited communication. These features make federated learning both practically attractive and statistically distinct from classical centralized analysis.

From a theoretical and methodological perspective, federated learning introduces several new phenomena. First, information is distributed unevenly across clients: some clients may have larger samples, stronger privacy constraints, better design distributions, or more relevant data for the target task. Second, learning protocols can be interactive: the server may adapt later queries to earlier public messages, and each client may reuse its local samples across many rounds. Third, the privacy constraint is naturally imposed clientwise, often on the complete public transcript rather than on a single release. These features create statistical questions that do not appear in the same form under central DP or local DP. In particular, one must understand how privacy budgets, sample sizes, client heterogeneity, interaction, and sample reuse jointly determine the amount of statistical information that can be extracted.

A line of work has been devoted to establishing lower bounds for the minimax risk of central DP estimation. Early minimax formulations and testing reductions were followed by private analogues of Le Cam's method, Fano's inequality, and Assouad's lemma [4, 5]. For high-dimensional problems, fingerprinting and tracing arguments provide another powerful route [6–9], and

T. Tony Cai is with the Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: tcai@wharton.upenn.edu).

Yicheng Li is with KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, China (e-mail: ycli@sfs.ecnu.edu.cn).

The authors are listed in alphabetical order. Corresponding author: T. Tony Cai.

the score attack framework turns these ideas into a general likelihood score method [3, 10]. More recently, Fisher information contraction and private van Trees inequalities have yielded smooth Bayesian lower bound tools under zCDP [11]. These results are all focused on the central DP model.

Federated learning creates additional challenges for lower bound theory because data and communication are both distributed. Sharp federated lower bounds for several statistical problems use model-specific noninteractive, sequential, or fresh batch settings [12–18]. Meanwhile, fully interactive lower bounds are formulated under local DP [19–21] and do not simply apply to federated clients with $n_l > 1$ local samples. Moreover, although some federated results allow heterogeneous sample sizes or privacy budgets, they do not provide a general theory for heterogeneous local experiments whose information may lie in different parameter directions.

The resulting question is to determine the information-theoretic limits of federated DP estimation when multi-sample, heterogeneous clients participate in fully interactive protocols. We study this question under clientwise sample-level zCDP imposed on the complete public transcript. Each client message may depend on its entire local dataset and all previous public messages, and the same local samples may be reused across adaptive rounds. Our aim is a general lower bound that holds uniformly over this full protocol class while retaining the directional information carried by the heterogeneous local experiments.

A. Our contributions

We address this question by developing a federated van Trees lower bound for estimating finite-dimensional parameters under squared ℓ_2 loss from complete public transcripts. In the formal model, the privacy constraint is imposed on the complete ordered public transcript. This transcript includes public query choices, schedules, participation decisions, stopping rules, and public randomness, and the analysis does not require a round-by-round privacy allocation or a fresh-sample decomposition. Equivalently, all public information that can affect later communication is part of the transcript.

As our main contribution, we establish a federated van Trees lower bound in matrix form through a clientwise matrix Fisher information contraction. The novelty lies in the clientwise matrix decomposition for multi-sample clients under one zCDP guarantee for the complete public transcript: it preserves the heterogeneous information geometry that a trace bound loses. The contraction decomposes the transcript Fisher information into clientwise positive semidefinite contributions, each constrained jointly by the client's raw Fisher information and a whitened trace budget determined by its zCDP parameter. Crucially, this matrix result is not a direct clientwise extension of the central DP trace contraction. Assigning a separate trace budget to each client would still control only the total amount of transcript information and would discard where that information is located in the parameter space. The matrix contraction instead preserves the information subspace and directional geometry contributed by each local experiment.

The resulting matrix van Trees lower bound takes a variational form over feasible clientwise information allocations. We derive an exact finite-dimensional dual for this variational problem and characterize its primal optimizer through a spectral water-filling rule. The dual retains the full matrix geometry without requiring the client Fisher information matrices to be simultaneously diagonalizable. At the variational optimizer, the water-filling rule allocates each client's privacy-limited information not simply to its largest local Fisher directions, but to the directions in which that information most reduces the remaining aggregate uncertainty. This characterization is especially informative for heterogeneous local experiments, where different clients may carry information in distinct and nonaligned parameter subspaces, as in missing coordinate models, anisotropic random designs, and target domain prediction.

By trace relaxation, the matrix lower bound further yields a scalar lower bound that is convenient in homogeneous or isotropic settings. This scalar relaxation also recovers known central and local DP lower bounds as special cases.

Our lower bound is powerful yet easy to use, reducing lower bounds to simple calculation of Fisher information. It enables us to identify or recover the minimax rates in extensive problems. See Table I for a summary of the examples. The applications include Gaussian mean estimation, homogeneous and heterogeneous linear regression, coverage and target-relevance problems, Bradley–Terry learning on comparison graphs, and nonparametric regression problems.

B. Related work

a) Lower bounds in central DP: The central DP lower bound literature adapts several classical minimax techniques to privacy constrained estimation. Private versions of Le Cam, Fano, and Assouad inequalities give general tools based on testing reductions for central DP statistical estimation [5]. These reductions are conceptually close to classical packing and hypercube arguments, and they are especially useful when the private difficulty can be encoded through pairwise or coordinatewise testing instances.

A second line of work is based on fingerprinting, tracing, and score attacks. Fingerprinting codes were first used to prove lower bounds for private query release and empirical risk problems [6, 23], and robust tracing arguments later removed some distributional restrictions [7]. For statistical estimation, generalized fingerprinting lemmas have produced sharp private lower bounds for Gaussian covariance and other exponential family problems [8, 9]. The score attack of Cai et al. [10] abstracts this approach through likelihood scores, recovering and extending sharp lower bounds for generalized linear models, ranking, sparse models, and nonparametric regression. The earlier "cost of privacy" analysis of Cai et al. [3] used related tracing ideas for

Problem	Minimax rate	Matching upper bound	Earlier lower-bound scope
Gaussian mean estimation over $\mathbb{R}^d$	$\frac{d}{mn} + \frac{d^2}{\rho mn^2}$	private weighted mean	non-interactive
Linear regression, $x \in \mathbb{R}^d$	$\frac{d}{mn} + \frac{d^2}{\rho mn^2}$	private weighted GD	sample splitting
Missing features and coordinate coverage	$\frac{d}{qn} + \frac{ds}{q\rho n^2}$	one noninteractive vector message	–
Target domain prediction with q relevant clients, $r = 1$	$\frac{1}{qn} + \frac{1}{q\rho n^2}$	[13]	non-interactive
Target domain prediction with q relevant clients, $r = d$	$\frac{d}{qn} + \frac{d^2}{q\rho n^2}$	[13]	sample splitting
Bradley–Terry preference learning on a graph $\mathcal{G}$	$\frac{\mathrm{Tr}(L_{\mathcal{G}}^{\dagger})}{q}\left(\frac{1}{n} + \frac{1}{\rho n^2}\right)$	Gaussian-perturbed empirical win rates	central DP
Nonparametric regression over H^α with uniform design on $[0,1]$	$(mn)^{-\frac{2\alpha}{2\alpha+1}} + (\rho mn^2)^{-\frac{\alpha}{\alpha+1}}$	[14]	non-interactive
Functional mean: k independent measurements	$\frac{1}{mn} + (mkn)^{-\frac{2\alpha}{2\alpha+1}} + (\rho mkn^2)^{-\frac{\alpha}{\alpha+1}} + \frac{1}{\rho mn^2}$	[16, 22]	sample splitting
Functional mean: k common measurements	$\frac{1}{mn} + k^{-2\alpha} + (\rho mn^2)^{-\frac{2\alpha}{2\alpha+1}}$	[22]	non-interactive

TABLE I

MINIMAX RATES (UP TO LOGARITHMIC FACTORS) OBTAINED BY COMBINING THE LOWER BOUNDS PROVED HERE WITH THE MATCHING CONSTRUCTIONS OR REFERENCES. RATES ARE SHOWN IN THE HOMOGENEOUS CASE $n_l = n$ AND $\rho_l = \rho$. LITERATURE UPPER BOUNDS ARE STATED ON THE CORRESPONDING ZCDP SCALE. HERE m IS THE NUMBER OF CLIENTS, n IS THE SAMPLE SIZE PER CLIENT.

mean estimation and linear regression. More recently, a Fisher information contraction was developed in Cai and Li [11] to give a general van Trees lower bound for central DP estimation, which is especially well suited to dependent priors and matrix parameters. All of these results are central DP results: the statistic is a single private output of a pooled sample, so they do not address clientwise privacy budgets or adaptive public transcripts.

b) Lower bounds in local DP: The local DP model is more restrictive than central DP: each observation, user, or client must privatize its own message before it is observed by any curator. The minimax theory of local privacy was initiated by Duchi et al. [24, 25, 26], who developed private data processing inequalities and sharp minimax rates for locally private estimation.

Subsequent work has refined the information-theoretic and interactive aspects of local privacy. Acharya et al. [19] derive lower bounds for interactive high-dimensional estimation under local privacy and communication constraints, while Barnes et al. [20, 21] derive minimax lower bounds from Fisher information contractions in the blackboard model. In their fully interactive formulations, each player holds a single observation, so these results do not provide a sample-level information bound with explicit dependence on a local sample size $n_l > 1$. Related Fisher information inequalities for local privacy are studied by Chen and Özgür [27]. Results for user-level local DP with multiple observations instead protect each user's entire local collection [28, 29], which corresponds to a different adjacency relation from the sample-level privacy considered here. Other work studies testing and communication limited local privacy [30], nonparametric density and functional estimation under local privacy [31–34], and high-dimensional regression under local privacy [35].

However, the techniques used in local DP lower bounds are not directly applicable to the central or federated DP setting here. If applied naively to the central DP setting, these approaches often lose a factor of n_l in the effective sample size through a rough conversion by group privacy.

c) Lower bounds in federated DP: A growing literature studies statistical and optimization lower bounds under federated or distributed DP constraints. For convex federated learning without a trusted server, Lowy and Razaviyayn [36] establish nearly tight excess risk bounds under inter-silo record level privacy. Zhang et al. [37] study how server trust affects private federated estimation and inference, while related distributed learning work considers robustness and privacy tradeoffs under distributed mechanisms [38]. Their lower bound arguments are tied to convex optimization or specific inferential tasks and do not give a clientwise Fisher information bound for general parametric experiments.

In nonparametric and transfer learning problems, recent papers have shown that heterogeneous client sample sizes and privacy budgets change both the optimal aggregation rule and the minimax rate. Results include federated nonparametric regression and testing under heterogeneous distributed DP [12, 14], federated transfer learning and nonparametric classification under distributed DP [13, 15], adaptive federated density estimation [18], private distributed functional data analysis [16], federated Cox regression and cumulative hazard estimation [17], and federated PCA/spiked covariance estimation [39].

Van Trees lower bound arguments in this literature have been developed for heterogeneous federated nonparametric and

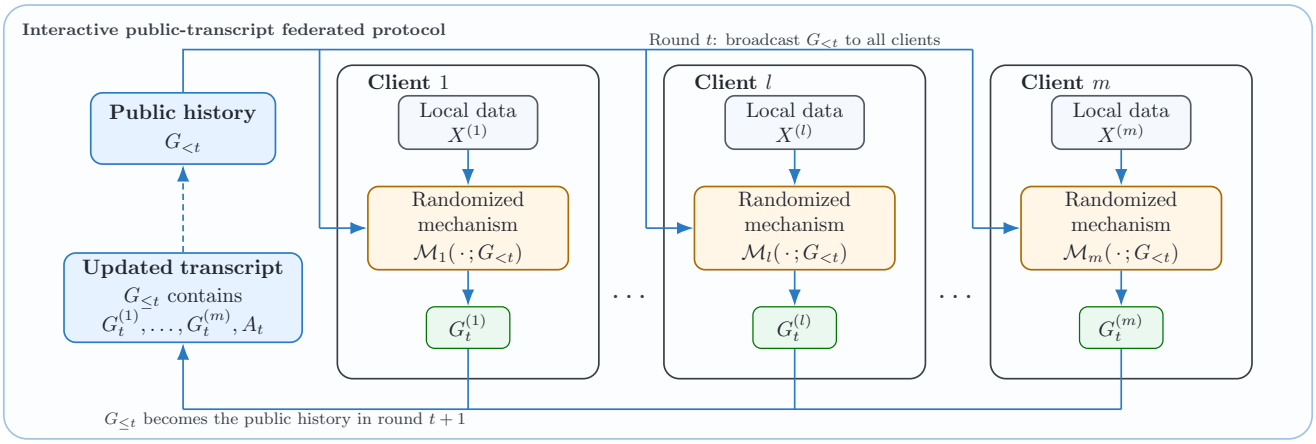

Fig. 1. One round of an interactive federated protocol with a public transcript. At round t, each client receives the public history $G_{<t}$, applies a local randomized algorithm to its local data $X^{(l)}$, and contributes $G_t^{(l)}$; the public environment then appends the public record A_t at the end of the round to form the updated transcript $G_{\le t}$.

matrix estimation [14, 16, 39]. These model-specific proofs use noninteractive local releases, one-pass or chained sequential transcripts, fresh batch decompositions, or reductions adapted to the target problem. These arguments do not cover protocols in which a local dataset is repeatedly queried across adaptive rounds under a single final privacy guarantee. The missing ingredient is therefore not another version of the classical van Trees inequality, but a clientwise contraction for multi-sample clients under one zCDP guarantee on the complete public transcript that retains heterogeneous information directions. We close this gap with a clientwise matrix contraction that depends only on each client's total sample-level zCDP budget, rather than on the number or ordering of communication rounds.

C. Notation

For nonnegative quantities a and b, we write $a \lesssim b$ if $a \le Cb$ for a constant C independent of the main asymptotic parameters, such as dimensions, sample sizes, privacy budgets, and the privacy mechanism. We write $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. For symmetric matrices A and B, $A \preceq B$ means that $B - A$ is positive semidefinite, and $A \succeq B$ is defined analogously. For a matrix J, $J^\dagger$ denotes its Moore–Penrose pseudoinverse. Let P_A denote the orthogonal projection onto the range of a matrix A.

II. Statistical Model and Federated Privacy

Differential privacy requires the output distribution of a randomized algorithm to remain stable when a single sample is modified. Throughout the paper, adjacency is imposed at the sample level within a client: two local datasets $X^{(l)}$ and $\tilde{X}^{(l)}$ are adjacent if they differ in exactly one observation and have the same size. Thus the privacy unit is one sample held by a client, not the whole client dataset. We work mainly with zero-concentrated differential privacy (zCDP) [40, 41], which is particularly convenient for interactive protocols. Let $D_\alpha(P\|Q)$ denote the Rényi divergence of order α between distributions P and Q, given by $D_\alpha(P\|Q) = \frac{1}{\alpha-1} \log \int \left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)^\alpha \mathrm{d}Q$; we use the same notation for random variables.

Let $\mathcal{D}$ denote a data domain, $\mathcal{H}$ an auxiliary input space, and $\mathcal{Y}$ an output space. We begin with the definition of zCDP for a single mechanism [40, 41], which is also standard in the central DP model.

Definition II.1 (ρ-zCDP)**.** A randomized mechanism $M : \mathcal{D} \times \mathcal{H} \to \mathcal{Y}$ is called ρ-zero-concentrated differentially private if, for every fixed pair of adjacent datasets $X, \tilde{X}$, every auxiliary input $h \in \mathcal{H}$,

$$D_\alpha\left(M(X;h) \middle\| M(\tilde{X};h)\right) \le \rho\alpha, \quad \forall \alpha > 1.$$

In federated learning, the data are distributed across multiple clients, and the algorithm may consist of multiple rounds of interaction. Formally, suppose there are m clients and client l holds a local dataset of size n_l:

$$X^{(l)} = (X_1^{(l)}, \ldots, X_{n_l}^{(l)}), \qquad l = 1, \ldots, m.$$

The federation is coordinated through an untrusted public server. We call the public coordination layer, comprising the server, public randomization, and public post-processing records, the public environment. Each client must protect its raw data from both the public environment and the other clients. Each client therefore releases only a privatized *transcript*, which may depend on the raw data, its own private history, and other public information. These transcripts are public and accessible to the server and all clients.

The learning algorithm is an interactive protocol between the clients and the server and may adapt fully through the public transcript. At a high level, the protocol starts from an initial public record. In each round, the current public history is broadcast to the clients, the clients release local public messages, and the public environment appends a public record before the next round begins. Figure 1 illustrates one such round.

Write G_0 for the initial public record. Client l may also maintain a private local history $H_t^{(l)}$, initialized at $H_0^{(l)}$, which is not released as part of the public transcript. In communication round t, after observing the public history $G_{<t}$, client l produces a public local message and updates its private history through a θ-independent randomized algorithm

$$\left(G_t^{(l)}, H_t^{(l)}\right) = \mathcal{M}_l\left(X^{(l)}; G_{<t}, H_{t-1}^{(l)}\right). \tag{1}$$

Once the local public messages in round t are available, the public environment may generate a public post-processing record A_t through a θ-independent public randomized algorithm

$$A_t = \mathcal{A}\left(G_{<t}, G_t^{(1)}, \ldots, G_t^{(m)}\right). \tag{2}$$

The public record in round t is

$$G_t = (G_t^{(1)}, \ldots, G_t^{(m)}, A_t), \qquad G_{<t} = G_{\le t-1} = (G_0, G_1, \ldots, G_{t-1}).$$

Let T be the total number of communication rounds. The complete ordered transcript is $G = G_{\le T}$, and a federated learning estimator is a function $\hat{\theta}(G)$.

We now define the zCDP federated protocol used throughout the paper.

Definition II.2 (zCDP federated protocol)**.** An interactive federated protocol with a public transcript is called a $\boldsymbol{\rho}$-zCDP federated protocol, with $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_m)$, if the complete public transcript G is ρ_l-zCDP as a mechanism of client l's data $X^{(l)}$ conditional on the other clients' data $X^{(-l)}$. Namely, for every client $l \in [m]$, every fixed value of the other clients' data $X^{(-l)}$, every adjacent pair $X^{(l)}, \widetilde{X}^{(l)}$,

$$D_\alpha\left(G(X^{(l)}, X^{(-l)}), G(\widetilde{X}^{(l)}, X^{(-l)})\right) \le \alpha\rho_l, \quad \forall \alpha > 1. \tag{3}$$

The privacy unit in Definition II.2 is one sample within a client, not the entire client dataset. Client-level privacy, where two adjacent datasets may differ in all records held by one client, is a different and generally stronger requirement; it falls within the present sample-level formulation only in the special case $n_l = 1$.

The following conventions specify the scope of the transcript model.

Complete transcript. The privacy guarantee is imposed on the complete public transcript G, not on any single message in isolation. It therefore allows arbitrary adaptivity and sample reuse across rounds. The unreleased local states $H_t^{(l)}$ themselves are not required to satisfy a privacy guarantee; privacy is imposed only on what becomes public. Inactive clients are represented by deterministic null messages. If a nominal communication round has an internal order of local transmissions, we refine the indexing so that t ranges over the resulting ordered public transmissions.

Public records. The public records contain all public control information used by the protocol. The initial record G_0 may include initial public randomness and any initial public query, schedule, or broadcast information. The record A_t appended after round t contains the public computation produced after that round, including server broadcasts, query choices, scheduling and participation decisions, stopping indicators, public randomness, and any other public information used in later rounds. If the protocol uses random stopping times, data dependent schedules, or adaptive participation rules, those decisions are part of G.

Protocol kernels. The local algorithms $\mathcal{M}_l$ depend only on client l's local dataset, its own private history, and the public history. The public algorithm $\mathcal{A}$ depends only on already public information. These randomized algorithms do not explicitly depend on the unknown parameter θ, which enters only through the sampling distribution of the raw data. If a protocol uses explicitly time indexed local or public algorithms, the round counter and all time indexed instructions are included in the public history $G_{<t}$. Private randomizations and local state updates are independent across clients conditional on the full data and public history. No shared private randomness, secure aggregation, aggregation across clients, or trusted server state may affect future communication unless the relevant information is recorded in the public transcript.

Nonexamples. The definition does not cover a secure aggregation protocol whose hidden partial sums or masks affect later communication without being recorded in G. It also excludes a trusted server that stores a private state derived from client messages and uses that state to choose later queries without adding it to the transcript. Similarly, clients may not rely on shared private randomness that is unobserved in G but influences future public messages.

The following examples satisfy this definition. For simplicity, these examples take no unreleased local history, so $\mathcal{M}_l$ outputs only the public message $G_t^{(l)}$.

Example II.1 (One pass sequential protocol). Consider a one pass sequential protocol in which clients $1, \ldots, m$ speak once in order. At step l, client l releases

$$G_l^{(l)} = \mathcal{M}_l(X^{(l)}; G_{<l}), \qquad l = 1, \ldots, m,$$

while inactive clients release deterministic null messages and the public record A_l at the end of the round may be empty. Suppose that, for every realized public history $g_{<l}$, the local mechanism $\mathcal{M}_l(\cdot; g_{<l})$ is ρ_l-zCDP as a mechanism of $X^{(l)}$. Then the complete transcript $G_{\le m}$ is a $\boldsymbol{\rho}$-zCDP federated protocol.

Example II.2 (Roundwise protocol). Consider a roundwise zCDP protocol in which the local mechanism $\mathcal{M}_l(\cdot; g_{<t})$ at round t is $\rho_{l,t}$-zCDP as a mechanism of $X^{(l)}$, conditional on every realized public history $g_{<t}$. Then, by adaptive composition of zCDP, the complete transcript is a $\boldsymbol{\rho}$-zCDP federated protocol with $\rho_l = \sum_{t=1}^{T} \rho_{l,t}$.

Example II.3 (General sequential protocol). Consider a general sequential protocol. Let $a_t \in [m]$ be the active client at transmission t, and write

$$G_t^{(a_t)} = \mathcal{M}_{a_t}(X^{(a_t)}; G_{<t}), \qquad t = 1, \dots, T,$$

with deterministic null messages from inactive clients and with any public post-processing at the end of the round included in A_t. Suppose that the local mechanism $\mathcal{M}_{a_t}(\cdot; g_{<t})$ is $\rho_{a_t,t}$-zCDP as a mechanism of $X^{(a_t)}$, conditional on every realized public history $g_{<t}$. This is a special case of Example II.2: in each transmission, all inactive clients send deterministic null messages, equivalently $\rho_{l,t} = 0$ for $l \neq a_t$. Hence the complete ordered transcript is a $\boldsymbol{\rho}$-zCDP federated protocol with $\rho_l = \sum_{t:a_t=l} \rho_{l,t}$.

III. Clientwise Fisher Information Geometry and Federated van Trees Bounds

This section establishes the paper's general lower bounds for fully interactive federated zCDP protocols. The clientwise matrix Fisher information contraction retains the full geometry of the local experiments and leads to the associated matrix van Trees inequality. We first record the homogeneous trace compression, and then retain the full clientwise geometry for heterogeneous experiments and derive an exact dual for the resulting matrix envelope. This separation makes the privacy argument reusable: for a new regular statistical model, only the raw client Fisher information must be recalculated. For presentation, we defer the regularity conditions to Assumption 1, including the standard differentiability, integrability, boundary, and transcript regularity conditions needed to apply the classical van Trees inequality.

A. Homogeneous Models and Trace Compression

We start with the homogeneous experiment, where all clients sample from the same parametric model but may have different sample sizes and privacy budgets. Client l holds independent local observations

$$X_i^{(l)} \overset{\text{i.i.d.}}{\sim} P_\theta, \qquad i = 1, \dots, n_l, \quad l = 1, \dots, m, \quad \theta \in \Theta \subseteq \mathbb{R}^p. \tag{4}$$

The local datasets are independent across clients. Let p_θ be a density of P_θ with respect to a θ-independent dominating measure. Define

$$s(x;\theta) = \nabla_\theta \log p_\theta(x), \qquad I_x(\theta) = \mathbb{E}_\theta[s(X;\theta)s(X;\theta)^\top]$$

to be the score and Fisher information matrix of one observation. To match the heterogeneous notation used below, this is the case $s_{li} = s$, with

$$S_l(\theta) = \sum_{i=1}^{n_l} s(X_i^{(l)};\theta), \qquad J_l(\theta) = n_l I_x(\theta).$$

The complete public transcript is denoted by G, and all estimators in this section are functions of G. The notation $\mathbb{E}_\pi \mathbb{E}_\theta$ means that $\theta \sim \pi$ is drawn first, after which the data and transcript are generated under P_θ.

For a clientwise zCDP budget $\boldsymbol{\rho} = (\rho_1, \dots, \rho_m)$, write

$$\kappa_l = e^{2\rho_l} - 1, \qquad l = 1, \dots, m.$$

Under the privacy regime where ρ_l is upper bounded by a constant, the factor $\kappa_l \asymp \rho_l$ is equivalent to ρ_l up to constants. The first ingredient is an information contraction inequality for the complete public transcript. It is the only point at which the interactive protocol and the clientwise zCDP constraint enter the trace lower bound.

Proposition III.1 (Trace contraction of the public transcript)**.** *Under model (4) and Assumption 1, let G be the complete public transcript of a $\boldsymbol{\rho}$-zCDP federated protocol. Then, for every fixed θ,*

$$\operatorname{Tr} I_G(\theta) \le \sum_{l=1}^{m} \left[\kappa_l n_l^2 \left\| I_x(\theta) \right\|_{\mathrm{op}} \wedge n_l \operatorname{Tr} I_x(\theta) \right]. \tag{5}$$

The two terms in each clientwise summand have different origins. The term $n_l \operatorname{Tr} I_x(\theta)$ is the ordinary Fisher information available in client l's raw sample. The term $\kappa_l n_l^2 \left\| I_x(\theta) \right\|_{\mathrm{op}}$ places an upper bound on the Fisher information that can pass through the complete public transcript under client l's sample-level zCDP budget. The proposition charges the single complete transcript budget ρ_l. It is therefore compatible with arbitrary adaptivity and repeated reuse of the same local observations across public rounds.

Combining this contraction with the multivariate van Trees inequality gives the main trace lower bound.

Theorem III.2. *Under model (4) and Assumption 1, every estimator $\hat{\theta} = \hat{\theta}(G)$ based on a $\boldsymbol{\rho}$-zCDP federated protocol satisfies*

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|_2^2 \geq \frac{p^2}{\displaystyle\int_\Theta \sum_{l=1}^{m} \left[\kappa_l n_l^2 \left\| I_x(\theta) \right\|_{\mathrm{op}} \wedge n_l \operatorname{Tr} I_x(\theta) \right] \mathrm{d}\pi(\theta) + \operatorname{Tr} J_\pi}.$$

The limitation of this form appears when the loss or the local experiments are directional. A client may carry substantial Fisher information entirely in directions irrelevant to the target, yet the scalar denominator above counts that information together with useful information. In Proposition IV.5, for example, only the q clients whose designs overlap the target subspace can reduce the target risk, whereas a trace calculation alone cannot exclude the remaining clients. The matrix formulation below retains the supporting subspace of each client contribution and therefore gives a sharper bound in such settings; see also Subsection III-D4.

B. Heterogeneous Models and Clientwise Matrix Contraction

The preceding trace bound retains only total transcript Fisher information. This compression is appropriate when the coordinates are statistically comparable and the loss treats all directions symmetrically. In heterogeneous experiments, however, different clients may observe different subspaces or have anisotropic designs. Then total Fisher information is not enough: the lower bound must also record where that information is located. The matrix bound retains this clientwise directional geometry.

Consider independent, not necessarily identically distributed, local observations

$$X_i^{(l)} \sim P_{\theta,li}, \qquad i = 1, \ldots, n_l, \quad l = 1, \ldots, m, \quad \theta \in \Theta \subseteq \mathbb{R}^p. \tag{6}$$

Let

$$s_{li}(X_i^{(l)}; \theta) = \nabla_\theta \log p_{\theta,li}(X_i^{(l)})$$

be the local score, and define

$$S_l(\theta) = \sum_{i=1}^{n_l} s_{li}(X_i^{(l)}; \theta), \qquad J_l(\theta) = \mathbb{E}_\theta[S_l(\theta) S_l(\theta)^\top].$$

If the observations are i.i.d. within client l, then $J_l(\theta) = n_l I_l(\theta)$, where $I_l(\theta)$ is the Fisher information of one sample from that client. Thus $J_l(\theta)$ is the raw Fisher information available at client l before privacy is imposed.

The next theorem decomposes the transcript Fisher information into clientwise positive semidefinite contributions. Each contribution is bounded by the client's raw information matrix and by a whitened trace budget determined by the zCDP constraint. The proof is given in Subsection A-D.

Theorem III.3 (Clientwise matrix contraction)**.** *Under the heterogeneous client model in (6) and Assumption 1, let G be the complete public transcript of a $\boldsymbol{\rho}$-zCDP federated protocol. For every fixed θ, there exist positive semidefinite matrices $A_l(\theta)$, $l = 1, \ldots, m$, such that*

$$I_G(\theta) = \sum_{l=1}^{m} A_l(\theta),$$

and

$$0 \preceq A_l(\theta) \preceq J_l(\theta), \qquad \operatorname{Tr}\left[J_l(\theta)^\dagger A_l(\theta) \right] \leq \kappa_l n_l.$$

Equivalently,

$$A_l(\theta) = J_l(\theta)^{1/2} B_l(\theta) J_l(\theta)^{1/2},$$

where

$$0 \preceq B_l(\theta) \preceq P_{J_l(\theta)}, \qquad \operatorname{Tr} B_l(\theta) \leq \kappa_l n_l.$$

Proof idea. Conditioning on the complete transcript preserves the factorization of the local data, so $I_G(\theta) = \sum_l A_l(\theta)$. For each client, on one hand, the conditional expectation cannot increase its raw Fisher information, which gives $A_l(\theta) \preceq J_l(\theta)$. On the other hand, the trace $\operatorname{Tr}\left[J_l(\theta)^\dagger A_l(\theta) \right]$ represents a sum of samplewise score–transcript correlations. Replacing one sample by an independent copy removes its correlation with the new transcript, while zCDP controls the resulting change, and summing these bounds gives the trace constraint. The representation in terms of $B_l(\theta)$ is then given by whitening with $J_l(\theta)$.

The theorem retains the clientwise geometry of the transcript information. The trace bound above is obtained by taking traces in these clientwise constraints and thereby discarding the location of the information. Here $A_l(\theta)$ is the part of the transcript Fisher information attributable to client l. The Loewner bound $A_l(\theta) \preceq J_l(\theta)$ says that privacy and interaction cannot create more information than the client's raw sample. The whitened trace bound says that, after measuring information in the geometry of $J_l(\theta)$, the public transcript can pass through at most $\kappa_l n_l$ Fisher information units from client l.

Taking traces in the matrix contraction gives the corresponding heterogeneous trace consequence. Since

$$\operatorname{Tr} A_l = \operatorname{Tr}(J_l B_l) \le \|J_l\|_{\mathrm{op}} \operatorname{Tr} B_l \le \kappa_l n_l \|J_l\|_{\mathrm{op}}$$

and $A_l \preceq J_l$ gives $\operatorname{Tr} A_l \le \operatorname{Tr} J_l$, we obtain

$$\operatorname{Tr} I_G(\theta) \le \sum_{l=1}^{m} \left[\kappa_l n_l \|J_l(\theta)\|_{\mathrm{op}} \wedge \operatorname{Tr} J_l(\theta)\right].$$

In the homogeneous model, $J_l(\theta) = n_l I_x(\theta)$, so this reduces to (5).

Combining the matrix van Trees inequality with Theorem III.3 gives the following matrix federated lower bound. In the statement, $P_{J_l(\theta)}$ denotes the orthogonal projection onto $\operatorname{range}(J_l(\theta))$. Thus, if $J_l(\theta)$ is singular, $B_l(\theta)$ is constrained to act only on the directions in which client l's raw experiment has Fisher information. The final inverse is interpreted in the extended sense: zero eigenvalues contribute $+\infty$.

Theorem III.4. *Under the heterogeneous client model in (6) and Assumption 1, for a measurable family $B = \{B_l(\theta)\}_{l=1}^m$, define*

$$M(B) = J_\pi + \int_\Theta \sum_{l=1}^{m} J_l(\theta)^{1/2} B_l(\theta) J_l(\theta)^{1/2} \, \mathrm{d}\pi(\theta). \tag{7}$$

Then every estimator $\hat\theta = \hat\theta(G)$ satisfies

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\|\hat\theta - \theta\right\|_2^2 \ge \inf_{\{B_l(\theta)\}} \operatorname{Tr}\left[M(B)^{-1}\right], \tag{8}$$

where the infimum is over measurable choices satisfying

$$0 \preceq B_l(\theta) \preceq P_{J_l(\theta)}, \qquad \operatorname{Tr} B_l(\theta) \le \kappa_l n_l.$$

The trace inverse is interpreted in the extended sense: zero eigenvalues contribute $+\infty$.

Proof idea. For a fixed protocol, the matrix van Trees inequality lower bounds the Bayes risk by $\operatorname{Tr}[(J_\pi + \int_\Theta I_G(\theta)\,\mathrm{d}\pi(\theta))^{-1}]$. Theorem III.3 supplies a feasible clientwise allocation B^G for which the matrix in this expression is exactly $M(B^G)$. Thus every protocol gives a feasible allocation, and taking the infimum over the stated feasible families yields the variational lower bound.

For heterogeneous experiments, the matrix expression is generally more complex than the scalar trace bound and need not reduce to a simpler closed form. To apply it, one first computes the raw client Fisher information $J_l(\theta)$. The matrix $B_l(\theta)$ then describes the information from client l that passes through the transcript after whitening in the geometry of $J_l(\theta)$. The trace constraint $\operatorname{Tr} B_l(\theta) \le \kappa_l n_l$ says that client l can contribute at most this many effective whitened directions. After combining these clientwise contributions, one inserts the resulting information matrix into the matrix van Trees expression, or restricts it to the subspace relevant to a projected loss.

C. Dual Representation and Spectral Allocation

The matrix lower bound in Theorem III.4 involves a minimization over matrix-valued information allocations that can obscure the relevant directions. The following dual converts this optimization to a finite-dimensional concave maximization and identifies the corresponding clientwise spectral allocation. For a positive semidefinite matrix C, let $\lambda_1(C) \ge \lambda_2(C) \ge \cdots$ be its ordered eigenvalues, extended by $\lambda_r(C) = 0$ for $r > \operatorname{rank}(C)$. For $t \ge 0$, define the fractional Ky–Fan functional

$$\Gamma_t(C) = \sum_{1 \le r \le t} \lambda_r(C) + (t - \lfloor t \rfloor)\lambda_{\lfloor t \rfloor + 1}(C). \tag{9}$$

Thus

$$\sup_{0 \preceq B \preceq P,\ \operatorname{Tr} B \le t} \operatorname{Tr}(BC) = \Gamma_t(PCP) \tag{10}$$

for every orthogonal projection P: the budget fills the largest eigendirections first, with at most one fractional direction away from ties. Write $\tau_l = \kappa_l n_l$.

Theorem III.5 (Dual representation)**.** *Under the conditions of Theorem III.4 and $J_\pi \succ 0$, the matrix value $\mathcal{V} = \inf_B \operatorname{Tr}(M(B)^{-1})$ satisfies*

$$\mathcal{V} = \sup_{X \in \mathbb{R}^{p \times p}} \left\{ 2 \operatorname{Tr} X - \operatorname{Tr}(X^\top J_\pi X) - \int_\Theta \sum_{l=1}^{m} \Gamma_{\tau_l}\left(J_l(\theta)^{1/2} X X^\top J_l(\theta)^{1/2}\right) \mathrm{d}\pi(\theta) \right\}. \tag{11}$$

The dual objective is concave, the primal infimum is attained, and the equality is valid without simultaneous diagonalization of the client Fisher matrices.

Corollary III.6 (Optimal spectral allocation)**.** *If B^* minimizes the primal and $M^* = M(B^*)$, then for almost every θ, client l's matrix $B_l^*(\theta)$ solves the spectral allocation problem for*

$$C_l^*(\theta) = J_l(\theta)^{1/2}(M^*)^{-2}J_l(\theta)^{1/2}.$$

Equivalently, it allocates up to total mass $\tau_l \wedge \operatorname{rank} J_l(\theta)$ to the leading eigendirections of $C_l^*(\theta)$.

The exact dual and Corollary III.6 give complementary views of the primal matrix bound. The primal infimum over measurable matrix-valued information allocations is exact, but leaves the relevant directions implicit. Without relaxation, the exact dual replaces this infimum with a finite-dimensional concave maximization; for each dual variable X, its Ky–Fan terms record the largest contribution allowed by each client's whitened information budget and local Fisher geometry, without requiring the client Fisher matrices to be simultaneously diagonalizable. The water-filling characterization identifies the corresponding primal geometry. At a primal optimizer, $(M^*)^{-2}$ weights directions that remain globally uncertain, so the rule allocates each client's budget not simply to its largest local Fisher directions but to the directions where its local information most reduces aggregate uncertainty. It is self-consistent because all clientwise allocations determine M^*, which in turn determines the matrices $C_l^*(\theta)$ used to select those allocations; it is noncommutative because each client can make this spectral choice in its own eigenbasis. The proof is given in Subsection B-B.

D. Consequences and Discussions

The preceding results identify the exact geometry of the general federated lower bound. We now draw out its rate, interaction, and directional consequences.

1) Homogeneous Rates: Theorem III.2 reduces the homogeneous lower bound calculation to the Fisher information $I_x(\theta)$ of one observation. Suppose, for example, that $\|I_x(\theta)\|_{\mathrm{op}} \le \sigma^{-2}$ uniformly and the prior Fisher information is dominated by the transcript information term. Then the bounded privacy form gives

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|_2^2 \gtrsim \sigma^2 \left[\sum_{l=1}^{m} \left(\frac{p}{n_l} \vee \frac{p^2}{\rho_l n_l^2}\right)^{-1}\right]^{-1}.$$

In the homogeneous case $n_l = n$ and $\rho_l = \rho$, this becomes

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|_2^2 \gtrsim \sigma^2 \left(\frac{p}{mn} \vee \frac{p^2}{\rho m n^2}\right).$$

When $m = 1$, the expression reduces to the central DP van Trees form based on Fisher information contraction [11]. When $n_l = 1$ for all l, the privacy limited term matches the usual high privacy local DP Fisher information scaling studied in local privacy lower bounds [19, 26].

2) Interaction and Matched Rates: Theorems III.2 and III.4 are uniform over the full class of interactive protocols with public transcripts in Definition II.2. They apply to protocols with a single local release, ordered sequential protocols, and protocols with multiple adaptive rounds in which the same local data may be reused across rounds. The bounds depend on the protocol only through the clientwise zCDP budgets ρ_l for the complete transcript, not through the number of rounds, the order of messages, or the way in which a budget is split across rounds. For an explicitly roundwise protocol, this means that the lower bound applies after the usual zCDP composition $\rho_l = \sum_t \rho_{l,t}$; more generally, no roundwise decomposition is required once the complete transcript satisfies Definition II.2.

This uniformity clarifies what a matching upper bound establishes. If a one round or noninteractive procedure attains a rate matching the trace or matrix bound, then allowing arbitrary interaction over multiple rounds cannot improve that rate. In this sense, a simple protocol that matches the bound is rate optimal not only within its own restricted communication structure but also within the larger class of fully interactive protocols with public transcripts. Interaction may still matter for constants, communication, numerical stability, or adaptation to unknown tuning parameters, but it cannot overcome the transcript Fisher information limits in Theorems III.2 and III.4.

This uniformity rests on the clientwise score decomposition for the complete transcript. Existing lower bounds are proved only for noninteractive releases, fresh batch protocols, or sample splitting reductions, and they do not by themselves rule out a fully adaptive algorithm that repeatedly queries the same local datasets. Here the entire ordered public transcript G, including all adaptive messages and public post-processing records, is the statistic to which the information contraction is applied. Conditioning on a realized complete transcript fixes the public adaptive choices, while the unreleased private histories remain local to their clients. This gives posterior factorization and hence the clientwise score projections used in Proposition III.1 and Theorem III.3. The proof therefore accommodates arbitrary interaction and sample reuse directly, rather than reducing to fresh batches or noninteractive releases.

3) Heterogeneous Information Geometry: The exact dual in Subsection III-C gives a finite-dimensional concave representation of the matrix optimization. The fully isotropic case gives a closed form expression for the matrix bound before client Fisher geometries introduce directional distinctions. Suppose that $J_\pi = \lambda I_p$ for some $\lambda > 0$ and that $J_l(\theta) \equiv j_l I_p$ for every client l. Writing $\tau_l = \kappa_l n_l$, the matrix value in (8) reduces to

$$\inf_{\substack{0 \preceq B_l \preceq I_p \\ \operatorname{Tr} B_l \le \tau_l}} \operatorname{Tr}\left[\lambda I_p + \sum_{l=1}^m j_l B_l\right]^{-1} = \frac{p}{\lambda + \sum_{l=1}^m j_l \left(1 \wedge \frac{\tau_l}{p}\right)}.$$

Indeed, rotational averaging makes each optimal B_l a scalar multiple of I_p, and monotonicity uses its full feasible trace budget. The scalar privacy penalty therefore arises by distributing each client's whitened budget symmetrically across statistically indistinguishable directions. Once the client Fisher geometries are no longer jointly isotropic, this compression is unavailable.

The allocation geometry in Corollary III.6 describes the corresponding general case. Although B^* is the information allocation that makes the lower bound weakest rather than an actual protocol, a candidate mechanism approaching the geometry of the lower bound should induce transcript information aligned with these globally weighted directions. Realizing this allocation through a concrete release or a roundwise zCDP budget split depends on the mechanism and may require handling the unknown parameter θ.

For applications, one may either solve the matrix optimization directly or lower bound its value using simpler consequences of the constraints. The measurability condition in Theorem III.4 permits such bounds to be derived from sharp integrated upper bounds on possible transcript information matrices. Enlarging the feasible class gives a weaker but still valid lower bound, whereas evaluating a single feasible family gives only an upper bound on the infimum. Another common route is to replace each feasible contribution $J_l^{1/2} B_l J_l^{1/2}$ by an explicit Loewner upper bound or to project the expression to the subspace appearing in the loss. These relaxations preserve validity because the trace of the inverse is order reversing on the positive semidefinite cone.

The allocation becomes coordinatewise when the relevant Fisher information matrices commute. Suppose, for instance, that J_π and $J_l(\theta)$ are independent of θ and are diagonal in the same orthonormal basis:

$$J_\pi = \operatorname{diag}(j_{\pi 1}, \dots, j_{\pi p}), \qquad J_l = \operatorname{diag}(j_{l1}, \dots, j_{lp}).$$

Then Theorem III.4 yields

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|_2^2 \ge \inf_{\{b_{lj}\}} \sum_{j=1}^p \frac{1}{j_{\pi j} + \sum_{l=1}^m j_{lj} b_{lj}},$$

where

$$0 \le b_{lj} \le \mathbf{1}\{j_{lj} > 0\}, \qquad \sum_{j=1}^p b_{lj} \le \kappa_l n_l \quad \text{for each } l.$$

Thus b_{lj} is the fraction of client l's available information budget allocated to direction j, and $\sum_j b_{lj} \le \kappa_l n_l$ is the corresponding clientwise privacy budget.

This coordinatewise interpretation fails for noncommuting Fisher geometries. For a simple example, take $p = 3$, $J_\pi = \lambda I_3$, and two clients with

$$J_1 = n_1 \sigma^{-2} P_1, \qquad J_2 = n_2 \sigma^{-2} P_2,$$

where P_1 is the projection onto $\operatorname{span}\{e_1, e_2\}$ and P_2 is the projection onto $\operatorname{span}\{(e_1 + e_3)/\sqrt{2}, e_2\}$. Then $P_1 P_2 \ne P_2 P_1$, so there is no common orthonormal basis in which the two clients' allocations reduce to independent scalars b_{lj}. If $\kappa_l n_l \le 1$, each client can pass through at most one effective whitened direction, but the two available planes overlap obliquely. In this case the matrix bound records how information subject to privacy constraints can be placed across nonaligned subspaces, while a trace calculation records only total Fisher mass. The contrast between these two summaries is illustrated in Figure 2.

4) Information Relevant to the Target: The matrix bound records information in the directions selected by the loss, rather than only the total transcript Fisher information. For a linear target $R\theta$, $R \in \mathbb{R}^{q \times p}$, and loss $\left\| \widehat{\psi} - R\theta \right\|_2^2$, it suffices to replace $\operatorname{Tr}[M(B)^{-1}]$ with $\operatorname{Tr}[RM(B)^{-1}R^\top]$ in the lower bound. Thus, R identifies the target directions, while the client Fisher matrices identify the directions from which information can be supplied.

A coordinate target makes this alignment explicit. Let $\mathcal{T} \subseteq [p]$, and let $R_\mathcal{T} \in \mathbb{R}^{|\mathcal{T}| \times p}$ select the coordinates in $\mathcal{T}$. Suppose that client l has Fisher information $J_l = n_l \sigma_l^{-2} P_l$, where P_l is the orthogonal projection onto a coordinate subset $\mathcal{S}_l \subseteq [p]$, and that $J_\pi = \operatorname{diag}(j_{\pi 1}, \dots, j_{\pi p})$. Coordinate sign averaging reduces the matrix allocation to diagonal entries, so every estimator $\widehat{\psi}_\mathcal{T}$ of $R_\mathcal{T}\theta$ satisfies

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\| \widehat{\psi}_\mathcal{T} - R_\mathcal{T}\theta \right\|_2^2 \ge \inf_{\{b_{lj}\}} \sum_{j \in \mathcal{T}} \frac{1}{j_{\pi j} + \sum_{l:\, j \in \mathcal{S}_l} n_l \sigma_l^{-2} b_{lj}},$$

where

$$0 \le b_{lj} \le 1 \quad \text{for } j \in \mathcal{S}_l, \qquad \sum_{j \in \mathcal{S}_l} b_{lj} \le \kappa_l n_l.$$

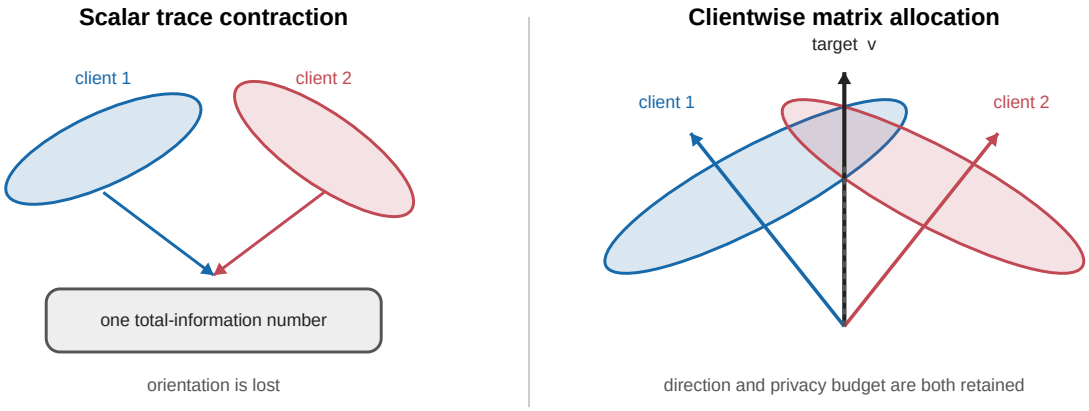

Fig. 2. Why a matrix contraction is needed. A scalar trace constraint retains total Fisher mass but discards its orientation. The clientwise matrix allocation preserves the raw information subspaces and the privacy-limited contribution in the target direction.

Thus, client l contributes to target coordinate j only when $j \in \mathcal{T} \cap \mathcal{S}_l$. The next corollary gives a closed form consequence when the target coordinates have balanced coverage.

Corollary III.7 (Balanced target coordinate coverage)**.** *Under the conditions of Theorem III.4, suppose $n_l = n$, $\rho_l = \rho$, and $J_l = n\sigma^{-2}P_l$, where each P_l is a coordinate projection onto a coordinate subset $\mathcal{S}_l \subseteq [p]$. Let $\varnothing \neq \mathcal{T} \subseteq [p]$, and suppose that, for some $q, r_{\mathcal{T}} \geq 1$, every client intersecting $\mathcal{T}$ observes exactly $r_{\mathcal{T}}$ target coordinates and every target coordinate is observed by q clients:*

$$|\mathcal{S}_l \cap \mathcal{T}| = r_{\mathcal{T}} \quad \text{whenever } \mathcal{S}_l \cap \mathcal{T} \neq \varnothing, \qquad \#\{l : j \in \mathcal{S}_l\} = q \quad \text{for every } j \in \mathcal{T}.$$

If $0 \prec J_\pi \preceq \lambda_\pi I_p$ for some $\lambda_\pi > 0$, then every estimator $\widehat{\psi}_{\mathcal{T}}$ of $R_{\mathcal{T}}\theta$ satisfies

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\| \widehat{\psi}_{\mathcal{T}} - R_{\mathcal{T}}\theta \right\|_2^2 \geq \frac{|\mathcal{T}|}{\lambda_\pi + \sigma^{-2} qn \left(1 \wedge \frac{\kappa n}{r_{\mathcal{T}}}\right)}, \qquad \kappa = e^{2\rho} - 1.$$

In particular, when the prior information is no larger than the transcript information term,

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\| \widehat{\psi}_{\mathcal{T}} - R_{\mathcal{T}}\theta \right\|_2^2 \gtrsim \sigma^2 \left(\frac{|\mathcal{T}|}{qn} \vee \frac{|\mathcal{T}|\, r_{\mathcal{T}}}{q\kappa n^2} \right).$$

The proof is given in Subsection A-D. Here q is the number of clients covering each target coordinate, while $r_{\mathcal{T}}$ is the number of target coordinates over which each target-relevant client must spend its privacy information budget. The privacy contribution therefore produces the additional term $|\mathcal{T}|\, r_{\mathcal{T}}/(q\kappa n^2)$. Unlike the matrix bound, a scalar trace denominator cannot distinguish information in $\mathcal{T}$ from information in its complement. For example, if every J_l is supported on $\mathrm{span}(e_1)$ and $R = e_2^\top$, then the matrix bound leaves the target direction controlled only by the prior, whereas a trace bound still sees positive total Fisher information.

The target subspace result in Proposition IV.5 applies the same principle to a noncoordinate target subspace. The rank-one binary regression result in Proposition IV.6 shows that the corresponding global and fixed linear target risks can be attained in a concrete model.

IV. Applications

This section illustrates the application of Theorems III.2 and III.4 to several canonical estimation problems. The examples progress systematically from settings with isotropic Fisher information to those with directional coverage and target relevance, followed by rank-one binary regression, graph geometry, and infinite-dimensional reductions. Let $\mathcal{M}_{\boldsymbol{\rho}}$ denote the class of estimators generated by $\boldsymbol{\rho}$-zCDP federated protocols, as formalized in Definition II.2. Throughout this section, we assume that the clientwise privacy budgets are uniformly bounded, $0 \leq \rho_l \leq C$, for a fixed constant $C > 0$. For each problem, we establish a minimax lower bound over $\mathcal{M}_{\boldsymbol{\rho}}$. In the homogeneous setting where $n_l = n$ and $\rho_l = \rho$ for all clients l, Table I summarizes the resulting rate comparisons. Detailed derivations for all applications are deferred to Section C.

A. Gaussian Mean Estimation and Homogeneous Linear Regression

The Gaussian mean model and homogeneous linear regression have Fisher information per observation comparable to $\sigma^{-2} I_d$. Applying the trace bound gives lower bounds of the same order in the two settings.

1) Gaussian Mean Estimation: Suppose client l observes

$$X_i^{(l)} \sim \mathcal{N}(\theta, \sigma^2 I_d), \qquad i = 1, \dots, n_l, \tag{12}$$

independently across i and l.

Proposition IV.1 (Gaussian mean estimation)**.** *Under the model (12), we have*

$$\inf_{\hat{\theta} \in \mathcal{M}_{\boldsymbol{\rho}}} \sup_{\theta \in [-1,1]^d} \mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|_2^2 \gtrsim d \wedge \sigma^2 \left[\sum_{l=1}^m \left(\frac{d}{n_l} \vee \frac{d^2}{\rho_l n_l^2} \right)^{-1} \right]^{-1}. \tag{13}$$

The two terms inside the clientwise maximum represent ordinary sampling difficulty and privacy contraction difficulty, respectively. The calculation appears in Subsection C-A1.

2) Homogeneous Linear Regression: To connect mean estimation to regression, consider the random design linear model

$$Y_i^{(l)} = (Z_i^{(l)})^\top \theta + \xi_i^{(l)}, \qquad \xi_i^{(l)} \sim \mathcal{N}(0, \sigma^2), \tag{14}$$

where $Z_i^{(l)}$ is independent of the noise and

$$c_0 I_d \preceq \mathbb{E} Z_i^{(l)} (Z_i^{(l)})^\top \preceq c_1 I_d \tag{15}$$

uniformly in l, for fixed constants $0 < c_0 \le c_1 < \infty$.

Proposition IV.2 (Homogeneous linear regression)**.** *Under the model (14) and design condition (15), we have*

$$\inf_{\hat{\theta} \in \mathcal{M}_{\boldsymbol{\rho}}} \sup_{\theta \in [-1,1]^d} \mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|_2^2 \gtrsim d \wedge \sigma^2 \left[\sum_{l=1}^m \left(\frac{d}{n_l} \vee \frac{d^2}{\rho_l n_l^2} \right)^{-1} \right]^{-1}. \tag{16}$$

The rate has the same form as for Gaussian mean estimation because the Fisher information of one observation is uniformly comparable to $\sigma^{-2} I_d$ in the well conditioned regime. The derivation appears in Subsection C-A2.

B. Coverage Geometry and Target Relevance

The matrix bound records where local information is placed, not merely how much information each client carries. We begin with an example with missing features and then turn to general heterogeneous designs and target loss.

1) Estimation with Missing Features: Let $\mathcal{S}_l \subseteq [d]$ have cardinality s, and suppose client l observes n independent vectors

$$X_i^{(l)} = (X_{ij}^{(l)} : j \in \mathcal{S}_l) \in \{-1, 1\}^s, \qquad \mathbb{P}_\theta(X_{ij}^{(l)} = 1) = \frac{1 + \theta_j}{2},$$

with independent coordinates and $\theta \in [-1/2, 1/2]^d$. Denote by P_l the projection onto the coordinates in $\mathcal{S}_l$. Assume the balanced coverage condition for simplicity:

$$\sum_{l=1}^m P_l = q I_d, \qquad ms = dq.$$

Proposition IV.3 (Estimation with missing features)**.** *In the preceding experiment, let every client have the same zCDP budget $0 \le \rho \le C$. Then*

$$\inf_{\widehat{\theta} \in \mathcal{M}_{\boldsymbol{\rho}}} \sup_{\theta \in [-1/2, 1/2]^d} \mathbb{E}_\theta \left\| \widehat{\theta} - \theta \right\|_2^2 \asymp d \wedge \left(\frac{d}{qn} + \frac{ds}{q\rho n^2} \right),$$

with the usual convention at $\rho = 0$. One noninteractive vector message per client attains the upper bound.

Thus each coordinate receives information from q clients, while each client spends its privacy information budget over s observed coordinates. The proof is given in Subsection C-B1.

2) Heterogeneous Linear Regression: For heterogeneous designs, the matrix bound preserves the orientation of local Fisher information. Suppose client l observes

$$Y_i^{(l)} = (Z_i^{(l)})^\top \theta + \xi_i^{(l)}, \qquad \xi_i^{(l)} \sim \mathcal{N}(0, \sigma_l^2), \qquad \mathbb{E} Z_i^{(l)} (Z_i^{(l)})^\top = \Sigma_l. \tag{17}$$

Proposition IV.4 (Heterogeneous linear regression)**.** *Under the model (17), we have*

$$\inf_{\hat{\theta} \in \mathcal{M}_{\boldsymbol{\rho}}} \sup_{\theta \in [-1,1]^d} \mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|_2^2 \ge \inf_{\{B_l\}} \operatorname{Tr} \left[c I_d + \sum_{l=1}^m n_l \sigma_l^{-2} \Sigma_l^{1/2} B_l \Sigma_l^{1/2} \right]^{-1}, \tag{18}$$

where

$$0 \preceq B_l \preceq P_{\Sigma_l}, \qquad \operatorname{Tr} B_l \le (e^{2\rho_l} - 1) n_l. \tag{19}$$

Here $c > 0$ is a universal constant and P_{Σ_l} denotes the orthogonal projection onto $\operatorname{range}(\Sigma_l)$. The transcript contributes no information in directions outside $\operatorname{span}\{\operatorname{range}(\Sigma_l) : 1 \le l \le m\}$. This directional obstruction is not captured by a scalar trace denominator. The derivation appears in Subsection C-B2.

3) Prediction on a Target Subspace: The target subspace problem is a direct specialization of the matrix bound for linear targets. Let $R_0 \in \mathbb{R}^{r\times d}$ have orthonormal rows spanning the target subspace, and write $P_0 = R_0^\top R_0$. In the heterogeneous linear model (17), the target is $R_0\theta$ and the loss is

$$L_0(\hat{\theta},\theta) = \left\|R_0(\hat{\theta}-\theta)\right\|_2^2 = (\hat{\theta}-\theta)^\top P_0(\hat{\theta}-\theta). \tag{20}$$

Suppose exactly q clients are target relevant with $n_l = n$, $\rho_l = \rho$, $\sigma_l = \sigma$, and $\Sigma_l = P_0$, while all other clients satisfy $P_0\Sigma_l P_0 = 0$:

$$\#\{l : \Sigma_l = P_0,\ n_l = n,\ \rho_l = \rho,\ \sigma_l = \sigma\} = q, \qquad P_0\Sigma_l P_0 = 0 \quad \text{for all other clients.} \tag{21}$$

Proposition IV.5 (Prediction on a target subspace)**.** *Under the model (17), target loss (20), and relevance condition (21), we have*

$$\inf_{\hat{\theta}\in\mathcal{M}_{\boldsymbol{\rho}}} \sup_{\theta\in B_2(0,1)} \mathbb{E}_\theta L_0(\hat{\theta},\theta) \gtrsim 1 \wedge \sigma^2\left(\frac{r}{qn} \vee \frac{r^2}{q\rho n^2}\right). \tag{22}$$

This specialization isolates the target relevance obstruction: only clients relevant to the target contribute to the lower bound, while information in orthogonal source directions is irrelevant for L_0. A trace calculation alone cannot make this distinction, because it summarizes source clients by total Fisher mass rather than by Fisher mass in the target subspace. The matrix bound removes clients whose information is supported in $P_0^\perp$ from the denominator of the lower bound for the target risk. The derivation appears in Subsection C-B3.

C. Rank-One Binary Regression

Rank-one binary regression isolates the directional geometry captured by the matrix bound: client l observes signs only along a public design direction u_l. Let client l have a public unit-norm design profile $u_l \in \mathbb{R}^d$ and observe n_l independent signs $X_i^{(l)} \in \{-1,1\}$ such that

$$\mathbb{P}_\theta(X_i^{(l)} = 1) = \frac{1+u_l^\top\theta}{2}, \qquad \theta \in \Theta_h = [-h,h]^d.$$

Put $P_l = u_l u_l^\top$, let $L_u = \max_l \|u_l\|_1$, and assume $hL_u \le 1/2$. For $0 < \rho_l \le C$, set

$$\omega_l = \left(\frac{1}{n_l} + \frac{2}{\rho_l n_l^2}\right)^{-1}, \qquad H = \sum_{l=1}^m \omega_l P_l,$$

and set $\omega_l = 0$ when $\rho_l = 0$. The matrix H is the privacy-weighted design matrix: ω_l is the precision scale available after client l averages its sample and adds privacy noise, while P_l specifies the direction carrying that precision.

Proposition IV.6 (Rank-One Binary Regression)**.** *Under the preceding model, suppose $H \succeq h^{-2} I_d$. Then*

$$\inf_{\widehat{\theta}\in\mathcal{M}_{\boldsymbol{\rho}}} \sup_{\theta\in\Theta_h} \mathbb{E}_\theta \left\|\widehat{\theta}-\theta\right\|_2^2 \asymp \operatorname{Tr}(H^{-1}).$$

More generally, for every fixed $R \in \mathbb{R}^{q\times d}$,

$$\inf_{\widehat{\psi}_R\in\mathcal{M}_{\boldsymbol{\rho}}} \sup_{\theta\in\Theta_h} \mathbb{E}_\theta \left\|\widehat{\psi}_R - R\theta\right\|_2^2 \asymp \operatorname{Tr}(RH^{-1}R^\top).$$

One noninteractive scalar message per client attains these upper bounds. The implicit constants depend only on C.

The case $R = v^\top$ gives the risk of a contrast in direction v. The following two-dimensional comparison makes the directional role of the matrix bound explicit. Even in dimension two, $H_\perp = WI_2$ and

$$H_\beta = W(e_1e_1^\top + u_\beta u_\beta^\top), \qquad u_\beta = (\cos\beta, \sin\beta)^\top,$$

have the same trace $2W$. If $0 < \beta \le \pi/4$ and $W\sin^2\beta \ge 4h^{-2}$, their contrast risks in direction e_2 have respective orders

$$e_2^\top H_\perp^{-1} e_2 = W^{-1}, \qquad e_2^\top H_\beta^{-1} e_2 = \frac{1+\cos^2\beta}{W\sin^2\beta}.$$

Thus equal total Fisher information can yield contrast risks differing by order β^{-2}: the information must be placed in directions that identify the target. The proof is given in Subsection C-C.

D. Bradley–Terry Learning on a Comparison Graph

Pairwise preference data are used to rank products, treatments, players, and candidate model outputs. The Bradley–Terry model is the classical parametric model for such comparisons [42]; modern ranking procedures also exploit the geometry of the comparison graph [43]. We consider a graph setting in which this geometry is explicit. Let $\mathcal{G} = ([K], E)$ be a connected graph with Laplacian $L_{\mathcal{G}}$. For each edge $e = \{a, b\} \in E$, there are q clients, and each client observes n independent outcomes from the oriented comparison

$$\mathbb{P}_\tau(Y_i^{(e,r)} = 1) = \sigma(\tau_a - \tau_b), \qquad \sigma(t) = \frac{e^t}{1 + e^t}. \tag{23}$$

The score vector is centered, $\tau \in \mathbf{1}_K^\perp$, and the privacy unit is one comparison outcome within a client. A bounded score range is necessary for a uniform squared-error statement: along an unbounded score ray with a strict ordering, the event that every finite comparison favors the higher-scored item has probability tending to one, so even the nonprivate data do not identify the scale. For a dynamic range $B > 0$, let

$$\Theta_{\mathcal{G}}(B) = \left\{\tau \in \mathbf{1}_K^\perp : |\tau_a - \tau_b| \le B \text{ for every } \{a, b\} \in E\right\}, \tag{24}$$

For client r on edge e, let $0 \le \rho_{e,r} \le C$ be its zCDP budget, and write

$$\omega_{e,r} = \begin{cases} \left(\dfrac{1}{n} + \dfrac{1}{\rho_{e,r} n^2}\right)^{-1}, & \rho_{e,r} > 0, \\ 0, & \rho_{e,r} = 0, \end{cases} \qquad w_e = \sum_{r=1}^{q} \omega_{e,r}.$$

The privacy-weighted comparison graph has Laplacian

$$L_{\mathcal{G},\boldsymbol{\rho}} = \sum_{e=\{a,b\}\in E} w_e (e_a - e_b)(e_a - e_b)^\top.$$

For a pair of items, write

$$\mathcal{R}_{\mathcal{G},\boldsymbol{\rho}}(a, b) = (e_a - e_b)^\top L_{\mathcal{G},\boldsymbol{\rho}}^\dagger (e_a - e_b), \qquad r_{\mathcal{G},\boldsymbol{\rho}} = \max_{\{a,b\}\in E} \mathcal{R}_{\mathcal{G},\boldsymbol{\rho}}(a, b).$$

Proposition IV.7 (Bradley–Terry model on a graph)**.** *Under the model (23), suppose that $0 < B \le B_0$, that client r on edge e satisfies $\rho_{e,r}$-zCDP, and that $L_{\mathcal{G},\boldsymbol{\rho}}$ is connected. Then*

$$\inf_{\widehat{\tau}\in\mathcal{M}_{\boldsymbol{\rho}}} \sup_{\tau\in\Theta_{\mathcal{G}}(B)} \mathbb{E}_\tau \|\widehat{\tau} - \tau\|_2^2 \gtrsim \frac{\mathrm{Tr}(L_{\mathcal{G},\boldsymbol{\rho}}^\dagger)}{1 + (K-1) r_{\mathcal{G},\boldsymbol{\rho}} / B^2}. \tag{25}$$

If, in addition, for fixed constants $c_0, \kappa_0 > 0$,

$$(K-1) r_{\mathcal{G},\boldsymbol{\rho}} \le c_0 B^2, \qquad \frac{\lambda_{\max}(L_{\mathcal{G},\boldsymbol{\rho}})}{\lambda_2(L_{\mathcal{G},\boldsymbol{\rho}})} \le \kappa_0, \tag{26}$$

then the global score risk satisfies

$$\inf_{\widehat{\tau}\in\mathcal{M}_{\boldsymbol{\rho}}} \sup_{\tau\in\Theta_{\mathcal{G}}(B)} \mathbb{E}_\tau \|\widehat{\tau} - \tau\|_2^2 \asymp \mathrm{Tr}(L_{\mathcal{G},\boldsymbol{\rho}}^\dagger). \tag{27}$$

Under (26), empirical win rates perturbed with Gaussian noise yield a matching upper bound. For a centered error vector z,

$$\|z\|_2^2 = \frac{1}{K} \sum_{a<b} (z_a - z_b)^2, \qquad \mathrm{Tr}(L_{\mathcal{G},\boldsymbol{\rho}}^\dagger) = \frac{1}{K} \sum_{a<b} \mathcal{R}_{\mathcal{G},\boldsymbol{\rho}}(a, b).$$

Thus $\mathrm{Tr}(L_{\mathcal{G},\boldsymbol{\rho}}^\dagger)$ is the average effective resistance over all item pairs for the graph with edge weights w_e. It reflects not only the number of comparisons, but also their placement in the graph and the privacy budgets of the clients making them: poorly connected pairs, or pairs supported mainly by clients with small privacy budgets, contribute more to the global score risk. The proof is given in Subsection C-D.

E. Infinite-Dimensional Models via Finite-Dimensional Reductions

Both examples apply the trace bound to a finite-dimensional submodel, with the resolution p balancing approximation against transcript information.

1) Nonparametric Regression: The trace bound yields nonparametric rates after projection onto finite-dimensional Sobolev submodels. Consider the nonparametric regression model on $[0,1]^d$,

$$Y_i^{(l)} = f(U_i^{(l)}) + \xi_i^{(l)}, \qquad \xi_i^{(l)} \sim \mathcal{N}(0, \sigma^2), \tag{28}$$

where the design density is bounded above and below and $f \in H^\alpha(R)$.

Proposition IV.8 (Nonparametric regression)**.** *Under the model (28), we have*

$$\inf_{\hat{f} \in \mathcal{M}_\rho} \sup_{f \in H^\alpha(R)} \mathbb{E}_f \left\| \hat{f} - f \right\|_{L^2}^2 \gtrsim \sup_{p \geq 1} \left[\frac{p^{2\alpha/d}}{R^2} + \frac{1}{\sigma^2} \sum_{l=1}^{m} \left(\frac{p}{n_l} \vee \frac{p^2}{\rho_l n_l^2} \right)^{-1} \right]^{-1}. \tag{29}$$

In the homogeneous case $n_l = n$ and $\rho_l = \rho$, this implies

$$\inf_{\hat{f} \in \mathcal{M}_\rho} \sup_{f \in H^\alpha(R)} \mathbb{E}_f \left\| \hat{f} - f \right\|_{L^2}^2 \gtrsim R^2 \wedge \left[R^{\frac{2d}{2\alpha+d}} \left(\frac{\sigma^2}{mn} \right)^{\frac{2\alpha}{2\alpha+d}} + R^{\frac{2d}{\alpha+d}} \left(\frac{\sigma^2}{\rho m n^2} \right)^{\frac{\alpha}{\alpha+d}} \right]. \tag{30}$$

The supremum over p applies the lower bound to each p-dimensional Sobolev submodel and retains the strongest result. Here p indexes the finite-dimensional reduction: the first term is the Sobolev truncation contribution, and the sum is the transcript information available at that resolution. The first component is the ordinary nonprivate sampling term, while the second is due to privacy. The derivation appears in Subsection C-E1.

2) Functional Mean Estimation: We next consider functional mean estimation from discretely observed curves. On client l, the n_l private samples are independent curves. The i-th curve is observed through

$$Y_{ij}^{(l)} = X_i^{(l)}(T_{ij}^{(l)}) + \xi_{ij}^{(l)}, \qquad j = 1, \ldots, k_l, \tag{31}$$

where $\xi_{ij}^{(l)} \sim \mathcal{N}(0,1)$, and the mean function is $f = \mathbb{E}X_i^{(l)} \in H^\alpha(R)$ with $\alpha > 1/2$. Let $\mathcal{P}_\alpha(R, C_X)$ denote the class of distributions P of random curves X such that the mean function $f_P = \mathbb{E}_P X$ belongs to $H^\alpha(R)$ and $\|X - f_P\|_{L^2} \leq C_X$ almost surely. We treat the radii R and C_X as fixed.

Under Definition II.2, the privacy unit is one entire discretely observed curve $\{(T_{ij}^{(l)}, Y_{ij}^{(l)}) : 1 \leq j \leq k_l\}$, not an individual grid measurement. Thus adjacent local datasets differ by one curve, and the lower bounds combine measurement sparsity with variation at the curve level.

We consider two design settings. Under independent design, the measurement locations are private and independently sampled for each curve. Under common design, the grid is public and shared across curves. The resulting lower bounds have the same form as those for noninteractive algorithms in Cai et al. [22].

a) Independent design: Suppose the measurement locations are private and independently sampled for each curve:

$$T_{ij}^{(l)} \overset{\text{ind}}{\sim} \mu_l, \qquad 0 < c_0 \leq \frac{\mathrm{d}\mu_l}{\mathrm{d}t} \leq c_1 < \infty. \tag{32}$$

To apply the van Trees lower bound, we use finite-dimensional submodels $f_\theta = \sum_{r=1}^{p} \theta_r \phi_r$, where p is the resolution parameter and $\{\phi_r\}$ is a localized wavelet system. Let $I_{l,p}^{\text{ind}}(\theta)$ be the Fisher information matrix for the coefficient parameter from one discretely observed curve at client l. Its information contribution is limited by both the number of measurements k_l and variation at the curve level:

$$\left\| I_{l,p}^{\text{ind}}(\theta) \right\|_{\text{op}} \lesssim k_l \wedge p^{2\alpha+1}, \qquad \operatorname{Tr} I_{l,p}^{\text{ind}}(\theta) \lesssim k_l p \wedge p^{2\alpha+2}.$$

The k_l-terms reflect the fact that only k_l random measurement locations are observed from each curve. The $p^{2\alpha+1}$-terms reflect variation at the curve level, which limits the information that one curve can carry about the p mean coefficients. Substituting these bounds into the van Trees inequality gives the following result. The detailed derivation appears in the appendix with the application proofs.

Proposition IV.9 (Functional mean estimation with independent design)**.** *Under the model (31) and independent design condition (32), we have*

$$\inf_{\hat{f} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha(R, C_X)} \mathbb{E}_P \left\| \hat{f} - f_P \right\|_{L^2}^2 \gtrsim \sup_{p \geq 1} \left[p^{2\alpha} + \sum_{l=1}^{m} \left(\frac{k_l n_l}{p} \wedge \frac{\rho_l k_l n_l^2}{p^2} \wedge p^{2\alpha} n_l \wedge p^{2\alpha-1} \rho_l n_l^2 \right) \right]^{-1}. \tag{33}$$

In the homogeneous case $n_l = n$, $k_l = k$, and $\rho_l = \rho$, this gives

$$\inf_{\hat{f} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha(R, C_X)} \mathbb{E}_P \left\| \hat{f} - f_P \right\|_{L^2}^2 \gtrsim 1 \wedge \left[\frac{1}{mn} + (mkn)^{-\frac{2\alpha}{2\alpha+1}} + (\rho m k n^2)^{-\frac{\alpha}{\alpha+1}} + \frac{1}{\rho m n^2} \right]. \tag{34}$$

In the homogeneous independent-design case, the privacy term follows by balancing the Sobolev contribution $p^{2\alpha}$ with $\rho m k n^2 / p^2$, while the low-frequency privacy bottleneck is obtained at $p = 1$. The appendix with the application proofs carries out the full optimization over p.

b) Common design: Suppose the grid is public and common across curves, with $k_l = k$:

$$T_{ij}^{(l)} = t_j, \qquad j = 1, \dots, k, \quad l = 1, \dots, m. \tag{35}$$

Fix a sufficiently small constant $c_{\mathrm{grid}} \in (0, 1)$. We again use finite-dimensional submodels $f_\theta = \sum_{r=1}^{p} \theta_r \phi_r$, now choosing ϕ_r as localized bump functions aligned with the public grid and taking $1 \le p \le c_{\mathrm{grid}} k$. Let $I_p^{\mathrm{com}}(\theta)$ be the Fisher information matrix for the coefficient parameter from one curve observed on the common grid. In this case,

$$\left\| I_p^{\mathrm{com}}(\theta) \right\|_{\mathrm{op}} \lesssim p, \quad \operatorname{Tr} I_p^{\mathrm{com}}(\theta) \lesssim p^2.$$

Together with the usual discretization obstruction term $k^{-2\alpha}$, this yields the following lower bound; its derivation appears in the appendix with the application proofs.

Proposition IV.10 (Functional mean estimation with common design)**.** *Under the model (31) and common design condition (35), suppose $k \ge c_{\mathrm{grid}}^{-1}$. Then*

$$\inf_{\hat f \in \mathcal{M}_{\boldsymbol{\rho}}} \sup_{P \in \mathcal{P}_\alpha(R, C_X)} \mathbb{E}_P \left\| \hat f - f_P \right\|_{L^2}^2 \gtrsim k^{-2\alpha} + \sup_{1 \le p \le c_{\mathrm{grid}} k} \left[p^{2\alpha} + \sum_{l=1}^{m} \left(n_l \wedge \frac{\rho_l n_l^2}{p} \right) \right]^{-1}. \tag{36}$$

In the homogeneous case $n_l = n$ and $\rho_l = \rho$,

$$\inf_{\hat f \in \mathcal{M}_{\boldsymbol{\rho}}} \sup_{P \in \mathcal{P}_\alpha(R, C_X)} \mathbb{E}_P \left\| \hat f - f_P \right\|_{L^2}^2 \gtrsim 1 \wedge \left[\frac{1}{mn} + k^{-2\alpha} + (\rho m n^2)^{-\frac{2\alpha}{2\alpha+1}} \right], \tag{37}$$

For common design, the privacy exponent comes from balancing $p^{2\alpha}$ with $\rho m n^2 / p$, subject to $p \le c_{\mathrm{grid}} k$. When this resolution exceeds the public grid, the discretization term $k^{-2\alpha}$ is active. The appendix with the application proofs carries out the full optimization over p.

V. Conclusion

We developed van Trees lower bounds for federated learning under sample-level zCDP with complete public transcripts. Our framework accommodates fully interactive protocols: clients may communicate over multiple adaptive rounds, the server's subsequent queries may depend on past public messages, and local samples may be reused across rounds. At the technical core of our work is a clientwise matrix Fisher information contraction for the complete transcript under local privacy constraints. This result decomposes the total transcript information into positive semidefinite contributions, each bounded by the client's raw Fisher information and a whitened privacy budget. By preserving the directional Fisher geometry of heterogeneous local experiments, this matrix bound proves essential when clients observe distinct subspaces or when the loss depends on specific target directions. For homogeneous models, trace compression yields a simplified scalar bound that reduces the lower bound calculation directly to the single-observation Fisher information and clientwise privacy budgets.

A key advantage of these inequalities is that they decouple the model-specific Fisher information calculation from the privacy and interaction mechanism. For the applications considered herein, this separation yields minimax lower bounds across the full class of interactive protocols with public transcripts for mean estimation, linear regression, nonparametric regression, and functional mean estimation. When paired with existing upper bounds, our results establish the exact minimax rates for these problems. Consequently, we show that simpler one-round or restricted protocols achieving these rates remain rate-optimal even against arbitrary multi-round interaction and repeated sample reuse.

Several open directions remain. For pure ϵ_l-DP, the lower bounds follow immediately because any ϵ_l-DP client mechanism is $\epsilon_l^2/2$-zCDP, so our results apply with $\rho_l = \epsilon_l^2/2$. However, approximate (ϵ_l, δ_l)-DP lacks a direct conversion to zCDP. Extending these results would require replacing the single-client zCDP contraction with an approximate-DP comparison that controls the exceptional δ_l-events, which remains an avenue for future work. Other promising extensions include lower bounds under secure aggregation, shuffle privacy, and protocols with trusted private server state. Furthermore, one could combine the complete transcript privacy contraction with explicit communication or computational constraints. Finally, the matrix-valued bound points toward new applications in heterogeneous learning and transfer learning, where minimax performance is governed by the specific subspace directions spanned by the clients rather than overall scalar Fisher information.

Appendix A
Proofs of Information Contraction and Federated van Trees Bounds

The following sufficient conditions are standard for Bayesian Cramér–Rao or van Trees inequalities; see Gill and Levit [44], Gassiat and Stoltz [45] and Lemma C.1 of Cai and Li [11]. We state them for the heterogeneous product model. For each θ, the raw data law is

$$P_\theta^X = \bigotimes_{l=1}^{m} \bigotimes_{i=1}^{n_l} P_{\theta,li}, \qquad X_i^{(l)} \sim P_{\theta,li}.$$

Let $p_{\theta,li}$ be the density of $P_{\theta,li}$, and define the local scores and information matrices by

$$s_{li}(X_i^{(l)};\theta) = \nabla_\theta \log p_{\theta,li}(X_i^{(l)}), \qquad I_{li}(\theta) = \mathbb{E}_\theta[s_{li}(X_i^{(l)};\theta)s_{li}(X_i^{(l)};\theta)^\top].$$

For the client and full data scores, write

$$S_l(\theta) = \sum_{i=1}^{n_l} s_{li}(X_i^{(l)};\theta), \qquad S(\theta) = \sum_{l=1}^{m} S_l(\theta).$$

Regularity of the raw data gives

$$\begin{aligned} \mathbb{E}_\theta s_{li}(X_i^{(l)};\theta) = 0, \qquad l = 1,\dots,m, \quad i = 1,\dots,n_l, \\ I_{li}(\theta) = \mathrm{Cov}_\theta(s_{li}(X_i^{(l)};\theta)), \\ J_l(\theta) := \mathbb{E}_\theta[S_l(\theta)S_l(\theta)^\top] = \sum_{i=1}^{n_l} I_{li}(\theta). \end{aligned} \tag{38}$$

Consequently, $\mathbb{E}_\theta S_l(\theta) = 0$ for each l. For a possibly randomized statistic T generated from the raw data, let $p_{\theta,T}$ be the density of its induced law, $I_T(\theta)$ its Fisher information matrix, and $S_X(\theta)$ the score of the raw data entering T. The corresponding score projection identities are

$$\begin{aligned} \nabla_\theta \log p_{\theta,T}(T) = \mathbb{E}_\theta[S_X(\theta) \mid T], \\ I_T(\theta) = \mathbb{E}_\theta\left[\mathbb{E}_\theta[S_X(\theta) \mid T]\mathbb{E}_\theta[S_X(\theta) \mid T]^\top\right]. \end{aligned} \tag{39}$$

For the complete public transcript G, write $I_G(\theta)$ for the Fisher information matrix of its induced law.

Assumption 1 (Regularity conditions for the van Trees bounds)**.** For the heterogeneous product model above, we use the following standard sufficient conditions for the van Trees inequality. They are not intended to be minimal.

(i) For each l and i, the density $p_{\theta,li}$ is defined with respect to a θ-independent dominating measure and is continuously differentiable in θ.

(ii) For each l and i, the local likelihood satisfies the conditions for differentiating under the integral sign that yield the local identities in (38).

(iii) The client Fisher information matrices $J_l(\theta)$ are well defined and

$$\int_\Theta \|J_l(\theta)\|_{\mathrm{op}} \, \mathrm{d}\pi(\theta) < \infty, \qquad l = 1,\dots,m.$$

(iv) The parameter space Θ is a compact subset of $\mathbb{R}^p$ with piecewise smooth boundary.

(v) The prior distribution π has a continuously differentiable density on $\mathbb{R}^p$ and vanishes on the boundary of Θ.

(vi) The prior Fisher information matrix

$$J_\pi = \int_\Theta \nabla_\theta \log \pi(\theta) \nabla_\theta \log \pi(\theta)^\top \pi(\theta) \, \mathrm{d}\theta$$

is well defined and finite.

(vii) The data, private history, and transcript spaces are standard Borel, so regular conditional distributions for randomized transcripts exist. The induced transcript laws are dominated and differentiable in θ and satisfy (39).

A. Information Contraction for One Client

The trace argument starts from the following zCDP information contraction for one client, due to Cai and Li [11].

Lemma A.1 (zCDP information contraction for one client)**.** *Let $X_1,\dots,X_n$ be i.i.d. samples from P_θ, with score $s(X;\theta)$ and one-sample Fisher information $I_x(\theta)$. Let T be a statistic generated from $X = (X_1,\dots,X_n)$ by a θ-independent ρ-zCDP mechanism at the sample level. Write $S_X(\theta) = \sum_{i=1}^n s(X_i;\theta)$. Then*

$$\mathbb{E}_\theta \, \|\mathbb{E}_\theta[S_X(\theta) \mid T]\|_2^2 = \mathrm{Tr}\, I_T(\theta) \le (e^{2\rho}-1)n^2 \, \|I_x(\theta)\|_{\mathrm{op}} .$$

The matrix argument uses a strengthening invariant under affine transformations. Its proof adapts the zCDP information contraction argument of Lemmas C.2 and C.3 of Cai and Li [11] after whitening the score.

Lemma A.2 (Matrix zCDP information contraction for one client)**.** *Let $X_1,\dots,X_n$ be independent observations from models $P_{\theta,i}$ that are not necessarily identical. Let $s_i(X_i;\theta)$ and $I_i(\theta)$ be the score and Fisher information matrix of X_i. Let*

$$S(\theta) = \sum_{i=1}^n s_i(X_i;\theta), \qquad J(\theta) = \mathbb{E}_\theta[S(\theta)S(\theta)^\top] = \sum_{i=1}^n I_i(\theta).$$

Let $T = M(X)$ be generated by a θ-independent ρ-zCDP mechanism at the sample level, and define

$$U(T;\theta) = \mathbb{E}_\theta[S(\theta) \mid T], \qquad A(\theta) = \mathbb{E}_\theta[U(T;\theta)U(T;\theta)^\top].$$

Then

$$0 \preceq A(\theta) \preceq J(\theta), \qquad \mathrm{Tr}\left[J(\theta)^\dagger A(\theta)\right] \le (e^{2\rho}-1)n.$$

Proof. Fix θ, take all expectations in this proof under P_θ, and write S, J, U, A for the corresponding quantities. We first assume $J \succ 0$. The law of total covariance gives

$$J = \mathbb{E}[\mathrm{Cov}(S \mid T)] + \mathrm{Cov}(\mathbb{E}[S \mid T]),$$

and since $\mathbb{E}S = 0$, the second term is A. Hence $0 \preceq A \preceq J$.

Set $W(T) = J^{-1}U(T)$ and

$$a = \mathrm{Tr}(J^{-1}A) = \mathbb{E}[U^\top J^{-1} U] = \mathbb{E}\left\langle W(T), U(T)\right\rangle.$$

Since $U(T) = \mathbb{E}[S \mid T]$,

$$a = \mathbb{E}\left\langle W(T), S\right\rangle = \sum_{i=1}^n \mathbb{E}\left\langle W(T), s_i(X_i)\right\rangle.$$

Fix i. Let X_i' be an independent copy of X_i, define

$$X^{(i)} = (X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n), \qquad T^{(i)} = M(X^{(i)}),$$

and write $W^{(i)} = W(T^{(i)})$. Since $X^{(i)}$ is independent of X_i and $\mathbb{E}s_i(X_i) = 0$,

$$\mathbb{E}\left\langle W^{(i)}, s_i(X_i)\right\rangle = 0.$$

Therefore

$$\mathbb{E}\left\langle W(T), s_i(X_i)\right\rangle = \mathbb{E}\left\langle W(T) - W^{(i)}, s_i(X_i)\right\rangle.$$

The datasets X and $X^{(i)}$ are adjacent, so zCDP gives, conditional on $(X, X^{(i)})$,

$$D_2(M(X)\|M(X^{(i)})) \le 2\rho.$$

Hence, using the relation between order-2 Rényi divergence and χ^2 divergence [46],

$$\chi^2(T\|T^{(i)} \mid X, X^{(i)}) = \chi^2(M(X)\|M(X^{(i)})) \le e^{2\rho} - 1 =: \kappa.$$

We use the chi-square mean difference inequality in Lemma F.1 of Cai and Li [11],

$$|\mathbb{E}_P f - \mathbb{E}_Q f|^2 \le \chi^2(P\|Q)\mathbb{E}_Q[f^2],$$

Applying it conditionally on $(X, X^{(i)})$, with $P = \mathcal{L}(T \mid X, X^{(i)})$, $Q = \mathcal{L}(T^{(i)} \mid X, X^{(i)})$, and $f(t) = \langle W(t), s_i(X_i)\rangle$, gives

$$\Big(\mathbb{E}[f(T) \mid X, X^{(i)}] - \mathbb{E}[f(T^{(i)}) \mid X, X^{(i)}]\Big)^2 \le \kappa\mathbb{E}[f(T^{(i)})^2 \mid X, X^{(i)}].$$

Averaging and applying Jensen's inequality yields

$$|\mathbb{E}\left\langle W(T), s_i(X_i)\right\rangle|^2 \le \kappa\,\mathbb{E}\left\langle W(T^{(i)}), s_i(X_i)\right\rangle^2.$$

Since $T^{(i)}$ is independent of X_i and has the same marginal distribution as T,

$$\mathbb{E}\left\langle W(T^{(i)}), s_i(X_i)\right\rangle^2 = \mathrm{Tr}\left(\mathbb{E}[W(T)W(T)^\top]I_i\right) = \mathrm{Tr}(J^{-1}AJ^{-1}I_i).$$

Denote the last quantity by b_i. Then by the triangle inequality and Cauchy–Schwarz,

$$a \le \sqrt{\kappa}\sum_{i=1}^n \sqrt{b_i} \le \sqrt{\kappa}\sqrt{n\sum_{i=1}^n b_i}.$$

But

$$\sum_{i=1}^n b_i = \mathrm{Tr}\left(J^{-1}AJ^{-1}\sum_{i=1}^n I_i\right) = \mathrm{Tr}(J^{-1}A) = a.$$

Thus $a \le \sqrt{\kappa n a}$. If $a > 0$, dividing by $\sqrt{a}$ gives $a \le \kappa n$; if $a = 0$, the same conclusion is trivial.

If J is singular, let $R = \mathrm{range}(J)$. For $v \in R^\perp = \ker(J)$,

$$0 = v^\top J v = \mathbb{E}(v^\top S)^2,$$

so $v^\top S = 0$ almost surely. Hence $v^\top U = \mathbb{E}[v^\top S \mid T] = 0$ almost surely, and $v^\top A v = \mathbb{E}(v^\top U)^2 = 0$. Since $A \succeq 0$, this implies $Av = 0$ for all $v \in R^\perp$, or equivalently $\mathrm{range}(A) \subseteq R$. If $R = \{0\}$, then $A = 0$ and the claim is trivial. Otherwise, choose an orthonormal matrix Q whose columns span R and apply the preceding positive definite argument to $\widetilde{S} = Q^\top S$, $\widetilde{U} = Q^\top U$, $\widetilde{J} = Q^\top J Q$, and $\widetilde{A} = Q^\top A Q$. Then $\widetilde{J} \succ 0$, and

$$\mathrm{Tr}(\widetilde{J}^{-1}\widetilde{A}) \le \kappa n.$$

Because $\mathrm{range}(A) \subseteq R$, the last trace equals $\mathrm{Tr}(J^\dagger A)$. In both the nonsingular and singular cases, set $B = (J^\dagger)^{1/2} A (J^\dagger)^{1/2}$. The range inclusion gives $A = J^{1/2} B J^{1/2}$, while $A \preceq J$ gives $B \preceq P_J$; moreover $\mathrm{Tr}\, B = \mathrm{Tr}(J^\dagger A)$. This gives the representation needed below. □

B. Posterior Factorization

This subsection supplies the factorization step that allows transcript Fisher information to be charged client by client. Once a complete transcript value g is fixed, all public adaptive choices and post-processing records are fixed, and each unreleased private history enters only through the likelihood factor of its own client. This yields posterior factorization of the local datasets given $G = g$, after which the matrix score decomposition follows by projecting the raw client scores onto the sigma-field generated by G.

Lemma A.3 (Posterior factorization)**.** *Under the public transcript protocol and independent client sampling model, for $P_{\theta,G}$-almost every transcript value g,*

$$\mathcal{L}_\theta(X^{(1)}, \dots, X^{(m)} \mid G = g) = \bigotimes_{l=1}^{m} \mathcal{L}_\theta(X^{(l)} \mid G = g).$$

Proof. Let θ be fixed. To make the core idea transparent, we first prove the result in a dominated setting where the local randomizations, local private history updates, initial randomizations, and public post-processing kernels admit densities. We then give the general argument with probability kernels. Let $\nu_0(g_0)$ be the density of the initial public record G_0, and write client l's private history path as $h_{0:T}^{(l)} = (h_0^{(l)}, \dots, h_T^{(l)})$. Let

$$\lambda_l(h_0^{(l)} \mid x^{(l)}, g_0)$$

be the conditional density of the initial private history $H_0^{(l)}$ given local data $x^{(l)}$ and the initial public record g_0, and let

$$q_{t,l}(g_t^{(l)}, h_t^{(l)} \mid x^{(l)}, g_{<t}, h_{t-1}^{(l)}),$$

be the conditional density of the pair $(G_t^{(l)}, H_t^{(l)})$ given local data $x^{(l)}$, public history $g_{<t}$, and previous private history $h_{t-1}^{(l)}$. Finally, let

$$\alpha_t(a_t \mid g_{<t}, g_t^{(1)}, \dots, g_t^{(m)})$$

be the density of the public post-processing record A_t at the end of the round given the already public information in that round. Deterministic messages, deterministic private history updates, and deterministic public post-processing are interpreted as point masses in this notation.

For a realized complete transcript

$$g = \big(g_0, (g_1^{(1)}, \dots, g_1^{(m)}, a_1), \dots, (g_T^{(1)}, \dots, g_T^{(m)}, a_T)\big),$$

the conditional density of $G = g$ given all datasets is

$$K(g \mid x^{(1)}, \dots, x^{(m)}) = P_{\mathrm{pub}}(g) \prod_{l=1}^{m} \int \lambda_l(h_0^{(l)} \mid x^{(l)}, g_0) \times \prod_{t=1}^{T} q_{t,l}(g_t^{(l)}, h_t^{(l)} \mid x^{(l)}, g_{<t}, h_{t-1}^{(l)})\, \mathrm{d}h_{0:T}^{(l)},$$

where the public factor is

$$P_{\mathrm{pub}}(g) = \nu_0(g_0) \prod_{t=1}^{T} \alpha_t(a_t \mid g_{<t}, g_t^{(1)}, \dots, g_t^{(m)}).$$

Since g is fixed, all adaptive public histories and all A_t-factors in this display are constants. The private histories remain inside client l's own integral and are not shared across clients. Define

$$L_l(g; x^{(l)}) = \int \lambda_l(h_0^{(l)} \mid x^{(l)}, g_0) \prod_{t=1}^{T} q_{t,l}(g_t^{(l)}, h_t^{(l)} \mid x^{(l)}, g_{<t}, h_{t-1}^{(l)}) \, \mathrm{d}h_{0:T}^{(l)}.$$

Once the transcript value is fixed, the public factor $P_{\mathrm{pub}}(g)$ is independent of the data and may be absorbed into one local factor, for instance by setting $K_1(g \mid x^{(1)}) = P_{\mathrm{pub}}(g) L_1(g; x^{(1)})$ and $K_l(g \mid x^{(l)}) = L_l(g; x^{(l)})$ for $l \geq 2$. Hence

$$K(g \mid x^{(1)}, \ldots, x^{(m)}) = \prod_{l=1}^{m} K_l(g \mid x^{(l)}).$$

By independence across clients,

$$p_\theta(x^{(1)}, \ldots, x^{(m)}) = \prod_{l=1}^{m} p_{\theta,l}(x^{(l)}),$$

where $p_{\theta,l}$ is the local joint density of client l's data. Bayes' formula gives

$$p_\theta(x^{(1)}, \ldots, x^{(m)} \mid g) = \prod_{l=1}^{m} \frac{p_{\theta,l}(x^{(l)}) K_l(g \mid x^{(l)})}{Z_l(g)},$$

with $Z_l(g) = \int p_{\theta,l}(u) K_l(g \mid u) \, \mathrm{d}u$. This proves posterior independence of the local datasets given G.

The same result follows without densities by working with probability kernels. Assume the data, private history, and transcript spaces are standard Borel, so regular conditional distributions exist. Fix θ. The protocol consists of the initial public kernel $\nu_0(\, \mathrm{d}g_0)$, the initial local private history kernels $\Lambda_l(\, \mathrm{d}h_0^{(l)} \mid x^{(l)}, g_0)$, the local round kernels

$$Q_{t,l}(\, \mathrm{d}g_t^{(l)}, \, \mathrm{d}h_t^{(l)} \mid x^{(l)}, g_{<t}, h_{t-1}^{(l)}),$$

and the public post-processing kernels

$$\mathsf{A}_t(\, \mathrm{d}a_t \mid g_{<t}, g_t^{(1)}, \ldots, g_t^{(m)}).$$

Deterministic messages and deterministic post-processing are simply degenerate kernels.

We use the following elementary disintegration fact. If, conditional on a public variable C, the pairs (Y_l, M_l), $l = 1, \ldots, m$, are independent with product conditional law $\otimes_l R_l(\, \mathrm{d}y_l, \, \mathrm{d}m_l \mid C)$, then, for almost every realized $m = (m_1, \ldots, m_m)$,

$$\mathcal{L}(Y_1, \ldots, Y_m \mid C, M = m) = \bigotimes_{l=1}^{m} \mathcal{L}(Y_l \mid C, M_l = m_l).$$

This is just disintegration of a product kernel with respect to its M-marginal.

We prove by induction on the public rounds that the local private states are conditionally independent given the public transcript available at that time:

$$\mathcal{L}_\theta\big((X^{(1)}, H_t^{(1)}), \ldots, (X^{(m)}, H_t^{(m)}) \mid G_{\leq t}\big) = \bigotimes_{l=1}^{m} \mathcal{L}_\theta(X^{(l)}, H_t^{(l)} \mid G_{\leq t})$$

for $P_{\theta, G_{\leq t}}$-almost every $G_{\leq t}$. At $t = 0$, the local datasets are independent and, conditional on $(X^{(l)}, G_0)$, the initial private histories are generated by separate local kernels; hence the claim holds conditional on G_0. Suppose it holds before round t, conditional on $G_{<t}$. Given $G_{<t}$, each client applies its own kernel $Q_{t,l}$ only to $(X^{(l)}, H_{t-1}^{(l)})$ and the public history. Therefore the enlarged local blocks

$$\big(X^{(l)}, H_t^{(l)}, G_t^{(l)}\big), \qquad l = 1, \ldots, m,$$

are conditionally independent given $G_{<t}$. Disintegrating this product kernel with respect to the observed public messages $(G_t^{(1)}, \ldots, G_t^{(m)})$ shows that $(X^{(l)}, H_t^{(l)})$, $l = 1, \ldots, m$, remain conditionally independent given $G_{<t}$ and those messages. The additional public record A_t is then drawn from a kernel depending only on already public quantities, so conditioning on A_t does not couple the local private blocks. Hence the induction step gives conditional independence given $G_{\leq t}$.

Taking $t = T$ and then marginalizing out the private histories gives

$$\mathcal{L}_\theta(X^{(1)}, \ldots, X^{(m)} \mid G = g) = \bigotimes_{l=1}^{m} \mathcal{L}_\theta(X^{(l)} \mid G = g)$$

for $P_{\theta,G}$-almost every g, which is the desired factorization. □

Corollary A.4 (Clientwise score decomposition)**.** *Under the conditions of Lemma A.3, if*

$$U_l(G;\theta) = \mathbb{E}_\theta[S_l(\theta) \mid G], \qquad A_l(\theta) = \mathbb{E}_\theta[U_l(G;\theta)U_l(G;\theta)^\top],$$

then

$$I_G(\theta) = \sum_{l=1}^{m} A_l(\theta).$$

Proof. The score projection identity gives

$$\nabla_\theta \log p_{\theta,G}(G) = \mathbb{E}_\theta[S(\theta) \mid G] = \sum_{l=1}^{m} U_l(G;\theta).$$

Hence

$$I_G(\theta) = \sum_{l} \mathbb{E}_\theta[U_l U_l^\top] + \sum_{l\neq k} \mathbb{E}_\theta[U_l U_k^\top].$$

For $l \neq k$, Lemma A.3 gives

$$\mathbb{E}_\theta[S_l(\theta)S_k(\theta)^\top \mid G] = \mathbb{E}_\theta[S_l(\theta) \mid G]\mathbb{E}_\theta[S_k(\theta) \mid G]^\top = U_l U_k^\top.$$

Taking expectations,

$$\mathbb{E}_\theta[U_l U_k^\top] = \mathbb{E}_\theta[S_l(\theta)S_k(\theta)^\top] = 0,$$

because the local datasets are unconditionally independent and the local scores have mean zero. Therefore $I_G(\theta) = \sum_l A_l(\theta)$. □

C. Proof of the Federated van Trees Bound for Homogeneous Models

Lemma A.5 (Trace contraction of the public transcript)**.** *Let G be the complete public transcript of a $\boldsymbol{\rho}$-zCDP federated protocol under the common iid model in Subsection III-A. Then*

$$\operatorname{Tr} I_G(\theta) \le \sum_{l=1}^{m} \left[(e^{2\rho_l} - 1)n_l^2 \left\|I_x(\theta)\right\|_{\mathrm{op}} \wedge n_l \operatorname{Tr} I_x(\theta)\right].$$

Proof. By Corollary A.4,

$$\operatorname{Tr} I_G(\theta) = \sum_{l=1}^{m} \mathbb{E}_\theta \left\|\mathbb{E}_\theta[S_l(\theta) \mid G]\right\|_2^2.$$

The ordinary Fisher bound follows from L^2 contraction of conditional expectation:

$$\mathbb{E}_\theta \left\|\mathbb{E}_\theta[S_l(\theta) \mid G]\right\|_2^2 \le \mathbb{E}_\theta \left\|S_l(\theta)\right\|_2^2 = n_l \operatorname{Tr} I_x(\theta).$$

For the privacy bound, set

$$\widetilde{U}_l(G, X^{(-l)}) = \mathbb{E}_\theta[S_l(\theta) \mid G, X^{(-l)}].$$

Since conditioning on $(G, X^{(-l)})$ gives a finer sigma-field than conditioning on G, Jensen's inequality gives

$$\mathbb{E}_\theta \left\|\mathbb{E}_\theta[S_l(\theta) \mid G]\right\|_2^2 \le \mathbb{E}_\theta \left\|\widetilde{U}_l(G, X^{(-l)})\right\|_2^2.$$

Conditional on $X^{(-l)} = x^{(-l)}$, the complete transcript G is a ρ_l-zCDP statistic of the n_l i.i.d. local samples $X^{(l)}$. Therefore Lemma A.1, applied conditionally and then averaged over $X^{(-l)}$, gives

$$\mathbb{E}_\theta \left\|\mathbb{E}_\theta[S_l(\theta) \mid G]\right\|_2^2 \le (e^{2\rho_l} - 1)n_l^2 \left\|I_x(\theta)\right\|_{\mathrm{op}}.$$

Combining the two clientwise bounds and summing over l proves the lemma. □

Proof of Theorem III.2. The classical multivariate van Trees inequality [44, 45], in the form recalled in Lemma C.1 of Cai and Li [11], applied to the statistic G gives

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|_2^2 \ge \frac{p^2}{\int_\Theta \operatorname{Tr} I_G(\theta)\, \mathrm{d}\pi(\theta) + \operatorname{Tr} J_\pi}.$$

Substituting the information contraction bound from Lemma A.5 proves the displayed bound. □

D. Proof of Clientwise Matrix Contraction and Its Consequences

Proof of Theorem III.3. Let

$$U_l(G;\theta) = \mathbb{E}_\theta[S_l(\theta) \mid G], \qquad A_l(\theta) = \mathbb{E}_\theta[U_l(G;\theta)U_l(G;\theta)^\top].$$

By Corollary A.4,

$$I_G(\theta) = \sum_{l=1}^{m} A_l(\theta).$$

The ordinary matrix contraction follows from the law of total covariance:

$$A_l(\theta) = \mathrm{Cov}_\theta(\mathbb{E}_\theta[S_l(\theta) \mid G]) \preceq \mathrm{Cov}_\theta(S_l(\theta)) = J_l(\theta).$$

It remains to prove the whitened trace bound. Define the finer posterior score

$$\widetilde{U}_l(G, X^{(-l)};\theta) = \mathbb{E}_\theta[S_l(\theta) \mid G, X^{(-l)}],$$

and

$$\widetilde{A}_l(\theta) = \mathbb{E}_\theta[\widetilde{U}_l(G, X^{(-l)};\theta)\widetilde{U}_l(G, X^{(-l)};\theta)^\top].$$

Since $U_l(G;\theta) = \mathbb{E}_\theta[\widetilde{U}_l(G, X^{(-l)};\theta) \mid G]$, Jensen's inequality in Loewner order gives

$$A_l(\theta) \preceq \widetilde{A}_l(\theta).$$

Conditional on $X^{(-l)} = x^{(-l)}$, the complete transcript G is a ρ_l-zCDP mechanism at the sample level for the local dataset $X^{(l)}$. The local score covariance is $J_l(\theta)$, which does not depend on $x^{(-l)}$. Applying Lemma A.2 conditionally gives

$$\mathrm{Tr}\left(J_l(\theta)^\dagger \mathbb{E}_\theta[\widetilde{U}_l\widetilde{U}_l^\top \mid X^{(-l)} = x^{(-l)}]\right) \le (e^{2\rho_l} - 1)n_l.$$

Averaging over $X^{(-l)}$ and using $A_l \preceq \widetilde{A}_l$ gives

$$\mathrm{Tr}\left[J_l(\theta)^\dagger A_l(\theta)\right] \le \mathrm{Tr}\left[J_l(\theta)^\dagger \widetilde{A}_l(\theta)\right] \le (e^{2\rho_l} - 1)n_l.$$

Finally, since $0 \preceq A_l \preceq J_l$, the range of A_l is contained in the range of J_l. Setting

$$B_l(\theta) = (J_l(\theta)^\dagger)^{1/2} A_l(\theta) (J_l(\theta)^\dagger)^{1/2}$$

yields

$$A_l(\theta) = J_l(\theta)^{1/2} B_l(\theta) J_l(\theta)^{1/2}, \qquad 0 \preceq B_l(\theta) \preceq P_{J_l(\theta)},$$

and $\mathrm{Tr}\, B_l(\theta) = \mathrm{Tr}[J_l(\theta)^\dagger A_l(\theta)]$. The jointly measurable score versions make $A_l(\theta)$ measurable. Matrix square root and Moore–Penrose inversion are Borel maps on the positive semidefinite cone, so the displayed choice of $B_l(\theta)$ is measurable as well. □

Proof of Theorem III.4. The matrix form of the van Trees inequality [44, 45] gives

$$\mathbb{E}_\pi \mathbb{E}_\theta[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top] \succeq \left(J_\pi + \int_\Theta I_G(\theta)\, \mathrm{d}\pi(\theta)\right)^{-1},$$

with the usual regularization if the matrix is singular. Now fix an admissible protocol. By Theorem III.3, this protocol induces positive semidefinite matrices $A_l(\theta)$ and hence feasible matrices $B_l(\theta)$ satisfying

$$I_G(\theta) = \sum_{l=1}^{m} J_l(\theta)^{1/2} B_l(\theta) J_l(\theta)^{1/2}.$$

Substituting this feasible family into the matrix van Trees bound gives a lower bound for the fixed protocol. The displayed theorem follows by weakening that bound: the family induced by the protocol is one admissible measurable choice in the feasible class, so its trace inverse value is at least the infimum over all feasible measurable choices $B_l(\theta)$. □

a) Coordinate target reduction.: Suppose that J_π is diagonal and $J_l = n_l \sigma_l^{-2} P_l$, where P_l is a coordinate projection onto $\mathcal{S}_l \subseteq [p]$. Let $R_\mathcal{T}$ select the coordinates in $\mathcal{T} \subseteq [p]$, and set $P_\mathcal{T} = R_\mathcal{T}^\top R_\mathcal{T}$. For a fixed protocol with integrated information matrix K, the matrix van Trees inequality for the linear target $R_\mathcal{T}\theta$ lower bounds its Bayes risk by $\mathrm{Tr}(P_\mathcal{T} K^{-1})$. For any feasible family $B_l(\theta)$, first set

$$\overline{B}_l = \int_\Theta B_l(\theta)\ \mathrm{d}\pi(\theta).$$

The averaged matrices remain feasible, and they give the same integrated information matrix. For a diagonal sign matrix D, the matrices $D\overline{B}_l D$ are feasible as well, while both J_π and every P_l commute with D. Moreover, the map

$$K \longmapsto \mathrm{Tr}(P_\mathcal{T} K^{-1})$$

is convex and invariant under $K \mapsto DKD$, with the usual regularization when needed. Averaging over all diagonal sign matrices therefore yields a diagonal feasible allocation without increasing the target inverse-trace objective. Writing its diagonal entries as b_{lj} gives the coordinate target bound in Subsection III-D4; conversely, every collection of such entries satisfying the displayed constraints defines a feasible diagonal allocation.

Proof of Corollary III.7. Fix any feasible family $B_l(\theta)$ in Theorem III.4, and set

$$\overline{B}_l = \int_\Theta B_l(\theta)\ \mathrm{d}\pi(\theta).$$

Convexity of the Loewner and trace constraints gives $0 \preceq \overline{B}_l \preceq P_l$ and $\mathrm{Tr}\,\overline{B}_l \le \kappa n$. Since $J_l = n\sigma^{-2} P_l$ is constant in θ, the integrated matrix in the theorem is

$$K = J_\pi + \sum_{l=1}^m n\sigma^{-2} P_l \overline{B}_l P_l.$$

Write

$$b_{lj} = e_j^\top \overline{B}_l e_j, \qquad a_j = \sum_{l:\, j \in \mathcal{S}_l} b_{lj}, \qquad j \in \mathcal{T},$$

where $b_{lj} = 0$ if $j \notin \mathcal{S}_l$. Let

$$m_\mathcal{T} = \#\{l : \mathcal{S}_l \cap \mathcal{T} \neq \varnothing\}.$$

For each target-relevant client,

$$\sum_{j \in \mathcal{S}_l \cap \mathcal{T}} b_{lj} \le r_\mathcal{T} \wedge \kappa n.$$

Double counting the target incidences gives

$$m_\mathcal{T} r_\mathcal{T} = \sum_{j \in \mathcal{T}} \#\{l : j \in \mathcal{S}_l\} = |\mathcal{T}|\, q.$$

Therefore

$$\sum_{j \in \mathcal{T}} a_j = \sum_{l=1}^m \sum_{j \in \mathcal{S}_l \cap \mathcal{T}} b_{lj} \le m_\mathcal{T}(r_\mathcal{T} \wedge \kappa n) = |\mathcal{T}|\, q \left(1 \wedge \frac{\kappa n}{r_\mathcal{T}}\right).$$

Also $K_{jj} \le \lambda_\pi + \sigma^{-2} n a_j$. By the Schur complement inequality, $(K^{-1})_{jj} \ge 1/K_{jj}$; the inverse exists because $K \succeq J_\pi \succ 0$. Hence

$$\mathrm{Tr}(P_\mathcal{T} K^{-1}) \ge \sum_{j \in \mathcal{T}} \frac{1}{\lambda_\pi + \sigma^{-2} n a_j}.$$

By convexity of $x \mapsto (\lambda_\pi + \sigma^{-2} n x)^{-1}$,

$$\sum_{j \in \mathcal{T}} \frac{1}{\lambda_\pi + \sigma^{-2} n a_j} \ge \frac{|\mathcal{T}|}{\lambda_\pi + \sigma^{-2} q n \left(1 \wedge \frac{\kappa n}{r_\mathcal{T}}\right)}.$$

Taking the infimum over feasible matrix fields $B_l(\theta)$ in the target form of the matrix van Trees bound proves the first display. The final display follows when λ_π is absorbed into the transcript information term. □

APPENDIX B
MATRIX DUALITY AND FUNCTIONAL TARGETS

This appendix proves the exact matrix dual and water-filling characterization stated in Subsection III-C, and records the consequences for functional targets. For the dual calculations below, we impose the additional nondegeneracy condition $J_\pi \succ 0$.

A. Matrix van Trees Bounds for Functional Targets

For a continuously differentiable target $\psi : \Theta \to \mathbb{R}^q$, define

$$D_{\pi,\psi} = \int_\Theta \nabla\psi(\theta)\ \mathrm{d}\pi(\theta), \qquad K_G = J_\pi + \int_\Theta I_G(\theta)\ \mathrm{d}\pi(\theta).$$

Under the regularity conditions in Assumption 1, the standard matrix-valued van Trees inequality gives

Theorem B.1 (Matrix van Trees bounds for functional targets)**.** *Assume $J_\pi \succ 0$. For every continuously differentiable target ψ with integrable derivative and every transcript estimator $\widehat{\psi}(G)$,*

$$\mathbb{E}_\pi \mathbb{E}_\theta[(\widehat{\psi} - \psi(\theta))(\widehat{\psi} - \psi(\theta))^\top] \succeq D_{\pi,\psi} K_G^{-1} D_{\pi,\psi}^\top.$$

In particular, for the identity target,

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\|\widehat{\theta} - \theta\right\|_2^2 \geq \operatorname{Tr} K_G^{-1}.$$

Proof. This is the standard matrix and functional form of the van Trees inequality under the regularity conditions in Assumption 1; see Gill and Levit [44], Gassiat and Stoltz [45]. The result is used here as the van Trees foundation for the dual calculation below. □

B. Proof of the Dual Representation and Allocation Rule

We restate the local spectral allocation notation used in the proofs. The extension to functional targets follows the two proofs below.

For a positive semidefinite matrix C, let $\lambda_1(C) \geq \cdots \geq \lambda_p(C)$. For $0 \leq t \leq p$, put $r = \lfloor t \rfloor$ and define the fractional Ky–Fan functional

$$\Gamma_t(C) = \sum_{j=1}^{r} \lambda_j(C) + (t - r)\lambda_{r+1}(C), \tag{40}$$

where the last term is absent if $r = p$. If P is an orthogonal projection, diagonalization on $\operatorname{range}(P)$ gives

$$\sup_{0 \preceq B \preceq P,\ \operatorname{Tr} B \leq t} \operatorname{Tr}(BC) = \Gamma_{t \wedge \operatorname{rank}(P)}(PCP). \tag{41}$$

Indeed, if u_j is an ordered eigenbasis of PCP, every feasible B has diagonal entries $b_j = u_j^\top B u_j \in [0, 1]$ with $\sum_j b_j = \operatorname{Tr} B$, and $\operatorname{Tr}(BC) = \sum_j b_j \lambda_j(PCP)$. Conversely, every such vector is realized by $B = \sum_j b_j u_j u_j^\top$. The matrix optimization therefore reduces to the linear program $\sup\{\sum_j b_j \lambda_j : 0 \leq b_j \leq 1, \sum_j b_j \leq t\}$, which fills the largest eigenvalues first. If the cutoff eigenvalue is simple, one may use at most one fractional coefficient; a tied cutoff permits an arbitrary positive contraction of the required trace inside that eigenspace.

Write $\tau_l = \kappa_l n_l$.

Proof of Theorem III.5. For every positive definite M, completing the square proves

$$\operatorname{Tr}(M^{-1}) = \sup_X \big(2 \operatorname{Tr} X - \operatorname{Tr}(X^\top M X)\big), \tag{42}$$

because the difference between the right-hand objective and $\operatorname{Tr}(M^{-1})$ is $-\operatorname{Tr}((X - M^{-1})^\top M (X - M^{-1}))$. Let $F(B, X)$ denote the objective after substituting $M(B)$.

Let $\mathbb{S}^p$ denote the real symmetric $p \times p$ matrices. The feasible families form a convex subset of $\prod_l L^\infty(\pi; \mathbb{S}^p)$. It is weak-star compact: the constraint $0 \preceq B_l \preceq I$ gives a uniformly bounded set, and the order, range, and trace inequalities are weak-star closed because each can be tested by integration against nonnegative L^1 scalar functions and a countable dense set of fixed vectors. For example, $B_l \succeq 0$ is equivalent to $\int h(\theta) v^\top B_l(\theta) v\ \mathrm{d}\pi \geq 0$ for every nonnegative $h \in L^1$ and every rational vector v, and the other constraints have the same form. The integrability in Assumption 1 makes F weak-star continuous and affine in B. It is continuous and concave in X. Moreover, uniformly in B,

$$F(B, X) \leq 2\sqrt{p}\, \|X\|_{\mathrm{F}} - \lambda_{\min}(J_\pi)\, \|X\|_{\mathrm{F}}^2 .$$

Since $F(B, 0) = 0$, every relevant maximizer lies in one fixed compact Frobenius ball. Sion's minimax theorem [47] therefore applies on a compact convex restriction of the X-space and yields

$$\inf_B \sup_X F(B, X) = \sup_X \inf_B F(B, X).$$

For fixed X, the term depending on client l and θ is the negative of

$$\operatorname{Tr}\left(B_l(\theta) J_l(\theta)^{1/2} X X^\top J_l(\theta)^{1/2}\right).$$

Thus its infimum is the negative pointwise spectral supremum in (41). The pointwise optimizers can be chosen measurably: ordered eigenvalues and spectral projections are Borel functions of a symmetric matrix, and inside a tied eigenspace one obtains a measurable ordered orthonormal basis by projecting the fixed standard basis and applying lexicographic Gram–Schmidt. Hence the pointwise supremum may be integrated, giving (11).

Each Ky–Fan term is the supremum of the convex quadratic forms $X \mapsto \operatorname{Tr}(X^\top J_l^{1/2} B J_l^{1/2} X)$ over feasible B. It is therefore convex in X, so the displayed dual objective is concave. Finally, $B \mapsto M(B)$ is continuous on the weak-star compact feasible set and $M(B) \succeq J_\pi \succ 0$; continuity of matrix inversion proves primal attainment. □

Proof of Corollary III.6. For any feasible $\widetilde{B}$, set $B_t = (1-t)B^* + t\widetilde{B}$. Differentiating $f(t) = \operatorname{Tr}(M(B_t)^{-1})$ at zero gives

$$f'(0+) = -\int \sum_l \operatorname{Tr}[(\widetilde{B}_l(\theta) - B_l^*(\theta))C_l^*(\theta)]\ \mathrm{d}\pi(\theta).$$

The map $B \mapsto \operatorname{Tr}(M(B)^{-1})$ is convex, so its first-order condition is necessary and sufficient. Optimality implies $f'(0+) \geq 0$, so B^* maximizes the corresponding linear functional over the feasible class. The constraints separate over clients and θ; applying (41) pointwise gives the asserted rule. Indeed, a failure on a set of positive prior measure could be replaced there by the measurable spectral optimizer constructed in the preceding theorem, strictly increasing the linear functional and contradicting optimality. □

C. Bounds for Functional Targets

Theorem B.2 (Federated van Trees bound for functional targets)**.** *Under the heterogeneous model and public-transcript zCDP conditions of Theorem III.4 and $J_\pi \succ 0$, let $\psi : \Theta \to \mathbb{R}^q$ be continuously differentiable with $D_{\pi,\psi} = \int \nabla\psi\ \mathrm{d}\pi$. Every transcript estimator obeys*

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\| \widehat{\psi} - \psi(\theta) \right\|_2^2 \geq \inf_B \operatorname{Tr}(D_{\pi,\psi} M(B)^{-1} D_{\pi,\psi}^\top). \tag{43}$$

The right side equals

$$\sup_{X \in \mathbb{R}^{p\times q}} \left(2\operatorname{Tr}(D_{\pi,\psi}X) - \operatorname{Tr}(X^\top J_\pi X) - \int_\Theta \sum_{l=1}^m \Gamma_{\tau_l \wedge \operatorname{rank} J_l(\theta)}\big(J_l(\theta)^{1/2} X X^\top J_l(\theta)^{1/2}\big)\ \mathrm{d}\pi(\theta) \right). \tag{44}$$

Proof. Apply Theorem B.1 and the clientwise information representation in Theorem III.3. The family induced by the protocol is feasible, so replacing it by the infimum over all feasible B proves (43). For a positive definite M, completion of the square gives

$$\operatorname{Tr}(D_{\pi,\psi} M^{-1} D_{\pi,\psi}^\top) = \sup_{X \in \mathbb{R}^{p\times q}} \big(2\operatorname{Tr}(D_{\pi,\psi}X) - \operatorname{Tr}(X^\top M X)\big).$$

The compactness, minimax, measurable selection, and pointwise spectral allocation steps in the proof of Theorem III.5 apply verbatim, now with $XX^\top$ of rank at most q. This proves the exact dual. □

Corollary B.3 (Closed form for scalar functionals)**.** *Under the conditions of Theorem B.2 and $J_\pi \succ 0$, let $q = 1$ and write $a_{\pi,\psi} = \int_\Theta \nabla\psi(\theta)\ \mathrm{d}\pi(\theta)$. Then*

$$\mathbb{E}_\pi \mathbb{E}_\theta (\widehat{\psi} - \psi(\theta))^2 \geq a_{\pi,\psi}^\top \left[J_\pi + \int_\Theta \sum_{l=1}^m (1 \wedge \tau_l) J_l(\theta)\ \mathrm{d}\pi(\theta) \right]^{-1} a_{\pi,\psi}. \tag{45}$$

Proof. In the scalar dual, $X = x \in \mathbb{R}^p$, and $J_l^{1/2} x x^\top J_l^{1/2}$ has rank at most one with nonzero eigenvalue $x^\top J_l x$. Its fractional Ky–Fan value is therefore $(1 \wedge \tau_l) x^\top J_l x$. The dual reduces to $\sup_x \big(2a_{\pi,\psi}^\top x - x^\top H x\big)$, where H is the matrix in brackets in (45). Completion of the square gives $a_{\pi,\psi}^\top H^{-1} a_{\pi,\psi}$. □

Appendix C
Proofs of the Applications

This appendix gives the proofs in the order of Section IV. We repeatedly use the following minimax reduction. If a prior π is supported in the parameter space, then

$$\inf_{\hat\theta \in \mathcal{M}_\rho} \sup_{\theta\in\Theta} \mathbb{E}_\theta L(\hat\theta, \theta) \geq \inf_{\hat\theta \in \mathcal{M}_\rho} \mathbb{E}_\pi \mathbb{E}_\theta L(\hat\theta, \theta).$$

For rectangular finite-dimensional submodels we use the cosine prior with density

$$p_0(t) = a^{-1} \cos^2\left(\frac{\pi t}{2a}\right), \qquad t \in [-a, a],$$

and $p_0(t) = 0$ otherwise. It has one-dimensional Fisher information π^2/a^2. Thus the p-fold product prior has Fisher matrix $(\pi^2/a^2) I_p$ and trace $p\pi^2/a^2$.

A. Gaussian Mean Estimation and Homogeneous Linear Regression

1) Gaussian Mean Estimation:

Proof of Proposition IV.1. For the Gaussian mean model in (12), the score of one observation has Fisher information

$$I_x(\theta) = \sigma^{-2} I_d, \qquad \|I_x(\theta)\|_{\mathrm{op}} = \sigma^{-2}, \qquad \operatorname{Tr} I_x(\theta) = d\sigma^{-2}.$$

Applying Theorem III.2 with the d-fold cosine prior supported on $[-1,1]^d$ and using $\kappa_l \asymp \rho_l$ for bounded ρ_l gives

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|_2^2 \gtrsim \frac{d^2}{\sigma^{-2} \sum_{l=1}^m (\rho_l n_l^2 \wedge d n_l) + d}.$$

Since

$$\rho_l n_l^2 \wedge d n_l = d^2 \left(\frac{d}{n_l} \vee \frac{d^2}{\rho_l n_l^2} \right)^{-1},$$

the preceding display is equivalent, up to constants, to

$$\inf_{\hat{\theta} \in \mathcal{M}_\rho} \sup_{\theta \in [-1,1]^d} \mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|_2^2 \gtrsim d \wedge \sigma^2 \left[\sum_{l=1}^m \left(\frac{d}{n_l} \vee \frac{d^2}{\rho_l n_l^2} \right)^{-1} \right]^{-1}.$$

The minimax bound follows because the prior is supported in $[-1,1]^d$. □

2) Homogeneous Linear Regression:

Proof of Proposition IV.2. For client l in random design linear regression,

$$Y = (Z^{(l)})^\top \theta + \xi, \qquad \xi \sim \mathcal{N}(0, \sigma^2),$$

the joint observation $X = (Z^{(l)}, Y)$ has score $\sigma^{-2} Z^{(l)} (Y - (Z^{(l)})^\top \theta)$, and hence

$$I_{x,l}(\theta) = \sigma^{-2} \mathbb{E} Z^{(l)} (Z^{(l)})^\top.$$

If $\mathbb{E} Z^{(l)} (Z^{(l)})^\top \preceq c_1 I_d$ uniformly in l, then

$$\|I_{x,l}(\theta)\|_{\mathrm{op}} \le c_1 \sigma^{-2}, \qquad \operatorname{Tr} I_{x,l}(\theta) \le c_1 d \sigma^{-2}.$$

Applying the trace consequence of the heterogeneous matrix contraction with the d-dimensional cosine prior on $[-1,1]^d$ gives the same denominator as in the Gaussian mean calculation, up to constants depending on c_1. Thus

$$\inf_{\hat{\theta} \in \mathcal{M}_\rho} \sup_{\theta \in [-1,1]^d} \mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|_2^2 \gtrsim d \wedge \sigma^2 \left[\sum_{l=1}^m \left(\frac{d}{n_l} \vee \frac{d^2}{\rho_l n_l^2} \right)^{-1} \right]^{-1}.$$

The lower eigenvalue condition in (15) is not needed for the lower bound itself; it records the usual well conditioned regime in which this lower bound has the same order as the corresponding nonprivate and private upper rates. □

B. Coverage Geometry and Target Relevance

1) Estimation with Missing Features and Target Subspaces: For an r-dimensional target subspace, take $d = s = r$, let each of the q clients relevant to the target observe every target coordinate, and allow any number of additional clients whose observation laws do not depend on the target parameter. This corresponds exactly to Proposition IV.5.

Corollary C.1 (Prediction on a target subspace)**.** *In the target experiment just described, under the common $0 \le \rho \le C$ zCDP budget,*

$$\inf_{\widehat{\theta} \in \mathcal{M}_\rho} \sup_{\theta \in [-1/2,1/2]^r} \mathbb{E}_\theta \left\|\widehat{\theta} - \theta\right\|_2^2 \asymp r \wedge \left(\frac{r}{qn} + \frac{r^2}{q\rho n^2} \right).$$

Here the privacy term is interpreted as $+\infty$ when $\rho = 0$. The same rate holds for squared prediction loss on an orthonormal target subspace after identifying its coefficient vector with θ.

Proof of Proposition IV.3 and Corollary C.1. For one bounded observation from client l, the Fisher information is

$$I_l(\theta) = \sum_{j \in \mathcal{S}_l} \frac{1}{1 - \theta_j^2} e_j e_j^\top.$$

On $[-1/2, 1/2]^d$,

$$P_l \preceq I_l(\theta) \preceq \frac{4}{3} P_l, \qquad \|I_l(\theta)\|_{\mathrm{op}} \le \frac{4}{3}, \qquad \operatorname{Tr} I_l(\theta) \le \frac{4s}{3}.$$

The n-record Fisher information is $J_l(\theta) = nI_l(\theta)$. Let $A_l(\theta)$ denote the clientwise transcript information matrices from Theorem III.3, and put $\kappa = e^{2\rho} - 1$. The matrix constraints imply

$$\operatorname{Tr} A_l(\theta) \le \frac{4}{3}\left(\kappa n^2 \wedge ns\right).$$

The raw information bound gives the second term, while the whitened trace constraint gives $\operatorname{Tr} A_l \le \|J_l\|_{\mathrm{op}} \operatorname{Tr} B_l \le (4n/3)\kappa n$. Thus

$$\operatorname{Tr} I_G(\theta) \le \frac{4m}{3}\left(\kappa n^2 \wedge ns\right).$$

Use the product cosine prior on $[-1/2, 1/2]^d$, for which $J_\pi = 4\pi^2 I_d$. Set $K_G = J_\pi + \int_\Theta I_G(\theta)\, \mathrm{d}\pi(\theta)$. The identity-target case of Theorem B.1 and $\operatorname{Tr}(K_G^{-1}) \ge d^2/\operatorname{Tr} K_G$ give the Bayes, and hence minimax, lower bound

$$\frac{cd^2}{d + m\left(\kappa n^2 \wedge ns\right)}.$$

Since $ms = dq$ and $\kappa \asymp \rho$ for bounded ρ, this is equivalent up to constants to

$$d \wedge \left(\frac{d}{qn} + \frac{ds}{q\rho n^2}\right).$$

For the matching upper bound with $\rho > 0$, client l releases

$$Y_l = \overline{X}_l + Z_l, \qquad Z_l \sim \mathcal{N}\left(0, \frac{2s}{\rho n^2} I_s\right),$$

where $\overline{X}_l$ is its empirical mean over the coordinates in $\mathcal{S}_l$. Replacing one record changes $\overline{X}_l$ by at most $2\sqrt{s}/n$ in Euclidean norm, and the Gaussian mechanism with variance $\Delta_2^2/(2\rho)$ is ρ-zCDP. For each coordinate j, averaging the q releases from clients with $j \in \mathcal{S}_l$ has variance at most

$$\frac{1}{qn} + \frac{2s}{q\rho n^2}.$$

Summing over coordinates, projecting onto $[-1/2, 1/2]^d$, and comparing with the zero estimator prove the upper bound, including $\rho = 0$.

For the target subspace result, take $d = s = r$ and $m = q$ for the relevant clients. Additional clients whose observation laws do not depend on the target parameter have zero score and raw Fisher information for that parameter. The heterogeneous contraction therefore assigns them zero transcript information in the lower bound argument; the upper construction ignores them. This proves Corollary C.1. □

2) Heterogeneous Linear Regression:

Proof of Proposition IV.4. For client l,

$$I_l(\theta) = \sigma_l^{-2}\Sigma_l, \qquad J_l(\theta) = n_l\sigma_l^{-2}\Sigma_l.$$

Substituting this into Theorem III.4 yields

$$J_l^{1/2} B_l J_l^{1/2} = n_l\sigma_l^{-2}\Sigma_l^{1/2} B_l \Sigma_l^{1/2},$$

where

$$0 \preceq B_l \preceq P_{\Sigma_l}, \qquad \operatorname{Tr} B_l \le (e^{2\rho_l} - 1)n_l.$$

Therefore

$$\mathbb{E}_\pi \mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|_2^2 \ge \inf_{\{B_l\}} \operatorname{Tr}\left[J_\pi + \sum_{l=1}^m n_l\sigma_l^{-2}\Sigma_l^{1/2} B_l \Sigma_l^{1/2}\right]^{-1}.$$

For $\Theta = [-1, 1]^d$, the product cosine prior has $J_\pi = cI_d$ for a universal constant $c > 0$, giving the display in (18) by the minimax reduction. Each summand is supported on $\operatorname{range}(\Sigma_l)$. Consequently the public transcript cannot reduce posterior uncertainty in directions orthogonal to $\operatorname{span}\{\operatorname{range}(\Sigma_l) : 1 \le l \le m\}$. This is the directional obstruction that a scalar trace denominator discards. □

3) Prediction on a Target Subspace:

Proof of Proposition IV.5. Take P_0 to be the projection onto an r-dimensional target subspace, and choose $R_0 \in \mathbb{R}^{d\times r}$ with orthonormal columns such that $P_0 = R_0 R_0^\top$. Restrict the parameter to the submodel $\theta = R_0 u$, where $u \in \mathbb{R}^r$, and consider

$$L_0(\hat\theta, \theta) = (\hat\theta - \theta)^\top P_0 (\hat\theta - \theta) = \left\| R_0^\top \hat\theta - u \right\|_2^2.$$

Only the q clients with $\Sigma_l = P_0$ contribute to L_0; all other clients satisfy $P_0 \Sigma_l P_0 = 0$ and therefore contribute no target information. Indeed, positive semidefiniteness gives $\left\| \Sigma_l^{1/2} P_0 \right\|_{\mathrm{F}}^2 = \mathrm{Tr}(P_0 \Sigma_l P_0) = 0$, so $\Sigma_l P_0 = P_0 \Sigma_l = 0$. In the u-coordinates, each relevant client has raw Fisher information $n\sigma^{-2} I_r$ and a whitened budget κn, where $\kappa = e^{2\rho} - 1$. For the minimax risk over $B_2(0,1)$, choose a product cosine prior for u on a coordinate cube of radius $r^{-1/2}$. Its Fisher matrix is bounded by a constant multiple of rI_r, and the induced prior for $\theta = R_0 u$ is supported in $B_2(0,1)$. Applying the balanced target coordinate calculation with dimension r, coverage q, and target rank r gives

$$\inf_{\hat\theta \in \mathcal{M}_\rho} \sup_{\theta \in B_2(0,1)} \mathbb{E}_\theta L_0(\hat\theta, \theta) \gtrsim 1 \wedge \sigma^2 \left(\frac{r}{qn} \vee \frac{r^2}{q\kappa n^2} \right).$$

For bounded ρ, $\kappa \asymp \rho$, which gives the rate displayed in (22). □

C. Rank-One Binary Regression

Proof of Proposition IV.6. Write $\eta_l = u_l^\top \theta$. For one sign observation from client l, the score and Fisher information are

$$s_l(x;\theta) = \frac{x - \eta_l}{1 - \eta_l^2} u_l, \qquad I_l(\theta) = \frac{1}{1 - \eta_l^2} P_l.$$

The condition $hL_u \le 1/2$ gives

$$P_l \preceq I_l(\theta) \preceq \frac{4}{3} P_l.$$

Let $\kappa_l = e^{2\rho_l} - 1$ and $\widetilde\omega_l = n_l(1 \wedge \kappa_l n_l)$. Because P_l has rank one, every feasible allocation in Theorem III.4 is a scalar multiple of P_l, and the inverse-trace objective decreases as that scalar increases. With the product cosine prior on Θ_h, whose information is $J_\pi = \pi^2 h^{-2} I_d$, the matrix theorem gives

$$\inf_{\widehat\theta \in \mathcal{M}_\rho} \sup_{\theta \in \Theta_h} \mathbb{E}_\theta \left\| \widehat\theta - \theta \right\|_2^2 \ge \mathrm{Tr}\left[\pi^2 h^{-2} I_d + \frac{4}{3} \sum_{l=1}^m \widetilde\omega_l P_l \right]^{-1}.$$

For $\psi_R(\theta) = R\theta$, we have $D_{\pi,\psi_R} = R$. The functional target bound in Theorem B.2 therefore gives, for every $R \in \mathbb{R}^{q\times d}$,

$$\inf_{\widehat\psi_R \in \mathcal{M}_\rho} \sup_{\theta \in \Theta_h} \mathbb{E}_\theta \left\| \widehat\psi_R - R\theta \right\|_2^2 \ge \mathrm{Tr}\left\{ R \left[\pi^2 h^{-2} I_d + \frac{4}{3} \sum_{l=1}^m \widetilde\omega_l P_l \right]^{-1} R^\top \right\}.$$

For $0 < \rho_l \le C$,

$$\widetilde\omega_l \asymp n_l(1 \wedge \rho_l n_l) \asymp \left(\frac{1}{n_l} + \frac{2}{\rho_l n_l^2} \right)^{-1} = \omega_l.$$

When $\rho_l = 0$, both $\widetilde\omega_l$ and ω_l are zero. Hence $\sum_l \widetilde\omega_l P_l \asymp H$ in Loewner order. The assumed condition $H \succeq h^{-2} I_d$ also absorbs the prior term. The two displays above are therefore bounded below by constant multiples of $\mathrm{Tr}(H^{-1})$ and $\mathrm{Tr}(RH^{-1}R^\top)$, respectively.

For the matching upper bound, client l releases

$$Y_l = \overline{X}_l + Z_l, \qquad Z_l \sim \mathcal{N}\left(0, \frac{2}{\rho_l n_l^2}\right)$$

when $\rho_l > 0$, and sends a deterministic null message otherwise. Replacing one record changes $\overline{X}_l$ by at most $2/n_l$, so the Gaussian mechanism with variance $\Delta_2^2/(2\rho_l)$ gives exact ρ_l-zCDP. Define

$$\widehat\theta = H^{-1} \sum_{l=1}^m \omega_l u_l Y_l.$$

Since $\mathbb{E}_\theta Y_l = u_l^\top \theta$, this estimator is unbiased. For every l with $\rho_l > 0$,

$$\mathrm{Var}_\theta(Y_l) \le \frac{1}{n_l} + \frac{2}{\rho_l n_l^2} = \omega_l^{-1}.$$

Independence across clients yields

$$\mathbb{E}_\theta \left\| \widehat{\theta} - \theta \right\|_2^2 \le \mathrm{Tr}\left[H^{-1} \left(\sum_{l=1}^{m} \omega_l P_l \right) H^{-1} \right] = \mathrm{Tr}(H^{-1}).$$

For every $R \in \mathbb{R}^{q \times d}$,

$$\mathbb{E}_\theta \left\| R\widehat{\theta} - R\theta \right\|_2^2 \le \mathrm{Tr}\left[RH^{-1} \left(\sum_{l=1}^{m} \omega_l P_l \right) H^{-1} R^\top \right] = \mathrm{Tr}(RH^{-1}R^\top).$$

□

D. Bradley–Terry Learning on a Comparison Graph

We first justify the condition that the score range is bounded. Fix a finite comparison design and a centered vector $v \in \mathbf{1}_K^\perp$ whose coordinates are all distinct. Along the ray $\tau_t = tv$, let E be the event that every observed comparison favors the higher-scored item. Then $\mathbb{P}_{\tau_t}(E) \to 1$ as $t \to \infty$. Conditional on E, the distribution of any estimator is fixed and does not depend on t. If it has finite second moment, its squared error from tv diverges; otherwise, its squared error is already infinite. Thus no finite comparison experiment has a uniform squared-error bound over the unbounded score space, even without privacy.

Proof of Proposition IV.7. Let $d = K - 1$, and let $Q_0 \in \mathbb{R}^{K \times d}$ have orthonormal columns spanning $\mathbf{1}_K^\perp$. Write $\tau = Q_0\theta$ and, for an oriented edge $e = (a, b)$, put

$$x_e = Q_0^\top (e_a - e_b).$$

Thus $\|x_e\|_2^2 = 2$, and the model is

$$\mathbb{P}_\theta(Y_i^{(e,r)} = 1) = \sigma(x_e^\top \theta).$$

Put

$$H_{\boldsymbol{\rho}} = Q_0^\top L_{\mathcal{G},\boldsymbol{\rho}} Q_0.$$

This matrix is positive definite because $L_{\mathcal{G},\boldsymbol{\rho}}$ is connected.

We first prove the lower bound. One comparison along edge e has Fisher information

$$\sigma(x_e^\top\theta)\{1 - \sigma(x_e^\top\theta)\} x_e x_e^\top \preceq \frac{1}{4} x_e x_e^\top.$$

The rank-one form of the matrix contraction in Theorem III.4 therefore gives, for every client,

$$A_{e,r}(\theta) \preceq C_C \omega_{e,r} x_e x_e^\top.$$

Summing over clients gives

$$\int_\Theta I_G(\theta)\, \mathrm{d}\pi(\theta) \preceq C_C H_{\boldsymbol{\rho}}.$$

Write

$$r_{\boldsymbol{\rho}} = \max_{e \in E} x_e^\top H_{\boldsymbol{\rho}}^{-1} x_e = r_{\mathcal{G},\boldsymbol{\rho}}, \qquad a = \frac{B}{\sqrt{d r_{\boldsymbol{\rho}}}}.$$

In the coordinates $z = H_{\boldsymbol{\rho}}^{1/2}\theta$, take a product cosine prior on $[-a, a]^d$. For every edge e, its support satisfies

$$\left| x_e^\top \theta \right| \le \left\| H_{\boldsymbol{\rho}}^{-1/2} x_e \right\|_2 \|z\|_2 \le \sqrt{r_{\boldsymbol{\rho}}} \sqrt{d}\, a = B,$$

so the induced prior is supported on $\Theta_{\mathcal{G}}(B)$. Its Fisher information in the θ coordinates is

$$J_\pi = \frac{\pi^2}{a^2} H_{\boldsymbol{\rho}} = \frac{\pi^2 d r_{\boldsymbol{\rho}}}{B^2} H_{\boldsymbol{\rho}}.$$

The matrix van Trees inequality and the preceding contraction bound therefore yield

$$\inf_{\widehat{\tau} \in \mathcal{M}_{\boldsymbol{\rho}}} \sup_{\tau \in \Theta_{\mathcal{G}}(B)} \mathbb{E}_\tau \left\| \widehat{\tau} - \tau \right\|_2^2 \ge \frac{\mathrm{Tr}(H_{\boldsymbol{\rho}}^{-1})}{C_C + \pi^2 d r_{\boldsymbol{\rho}} / B^2} \gtrsim \frac{\mathrm{Tr}(L_{\mathcal{G},\boldsymbol{\rho}}^\dagger)}{1 + (K-1) r_{\mathcal{G},\boldsymbol{\rho}} / B^2}.$$

For the matching upper bound, each client releases

$$Z_{e,r} = \overline{Y}_{e,r} + W_{e,r}, \qquad W_{e,r} \sim \mathcal{N}\left(0, \frac{1}{2\rho_{e,r} n^2}\right),$$

independently across clients whenever $\rho_{e,r} > 0$; clients with $\rho_{e,r} = 0$ are ignored. Replacing one comparison changes $\overline{Y}_{e,r}$ by at most $1/n$, so this release is $\rho_{e,r}$-zCDP. For $\rho_{e,r} > 0$, write

$$v_{e,r} = \frac{1}{4n} + \frac{1}{2\rho_{e,r} n^2}, \qquad H_v = \sum_{e \in E} \sum_{r:\, \rho_{e,r} > 0} v_{e,r}^{-1} x_e x_e^\top.$$

Then

$$2H_{\boldsymbol{\rho}} \preceq H_v \preceq 4H_{\boldsymbol{\rho}}.$$

Moreover, if $p_e = \sigma(x_e^\top \theta)$, then

$$\mathbb{E}_\theta Z_{e,r} = p_e, \qquad \mathrm{Var}_\theta\, Z_{e,r} \le v_{e,r}.$$

Let $\widehat{\theta}$ minimize, over

$$\widetilde{\Theta}_{\mathcal{G}}(B) = \{\vartheta : Q_0 \vartheta \in \Theta_{\mathcal{G}}(B)\}$$

the convex criterion

$$\mathcal{L}_Z(\vartheta) = \sum_{e \in E} \sum_{r:\, \rho_{e,r} > 0} v_{e,r}^{-1} \left\{ \log(1 + e^{x_e^\top \vartheta}) - Z_{e,r} x_e^\top \vartheta \right\},$$

and set $\widehat{\tau} = Q_0 \widehat{\theta}$. On $\widetilde{\Theta}_{\mathcal{G}}(B)$,

$$\nabla^2 \mathcal{L}_Z(\vartheta) \succeq q_B H_v, \qquad q_B = \min_{|t| \le B} \sigma(t)\{1 - \sigma(t)\} > 0.$$

Since $B \le B_0$, we have $q_B \ge q_{B_0} > 0$.

$$g = \sum_{e \in E} \sum_{r:\, \rho_{e,r} > 0} v_{e,r}^{-1} x_e \{Z_{e,r} - p_e\},$$
$$\mathrm{Cov}_\theta(g) \preceq H_v.$$

The first-order optimality condition and strong convexity therefore imply

$$\sup_{\tau \in \Theta_{\mathcal{G}}(B)} \mathbb{E}_\tau \left\| \widehat{\tau} - \tau \right\|_2^2 \lesssim \frac{d}{q_B^2 \lambda_{\min}(H_v)} \le \frac{d}{q_{B_0}^2 \lambda_{\min}(H_v)} \lesssim \frac{d}{\lambda_2(L_{\mathcal{G},\boldsymbol{\rho}})}.$$

The condition-number bound in (26) bounds this by a constant multiple of

$$\mathrm{Tr}(L_{\mathcal{G},\boldsymbol{\rho}}^\dagger),$$

which proves the upper bound and completes the proof. □

E. Infinite-Dimensional Models via Finite-Dimensional Reductions

1) Nonparametric Regression:

Proof of Proposition IV.8. Let $\phi_1, \dots, \phi_p$ be an orthonormal system in $L^2([0,1]^d)$ adapted to the Sobolev norm, and consider the finite-dimensional submodel

$$f_\theta(u) = \sum_{j=1}^{p} \theta_j \phi_j(u).$$

Choose a product cosine prior on the coefficients with radius

$$r_p \asymp R p^{-1/2 - \alpha/d}.$$

Then the prior is supported in $H^\alpha(R)$, and

$$\mathrm{Tr}\, J_{\pi,p} \asymp p r_p^{-2} \asymp R^{-2} p^{2 + 2\alpha/d}.$$

The design density is bounded above and below, so the Fisher information of one observation in the submodel satisfies

$$\|I_x(\theta)\|_{\mathrm{op}} \lesssim \sigma^{-2}, \qquad \mathrm{Tr}\, I_x(\theta) \lesssim \sigma^{-2} p.$$

Applying Theorem III.2 and using $\kappa_l \asymp \rho_l$ for bounded ρ_l gives, for each fixed resolution p,

$$\mathbb{E}_\pi \mathbb{E}_f \left\| \hat{f} - f \right\|_{L^2}^2 \gtrsim \frac{p^2}{R^{-2} p^{2 + 2\alpha/d} + \sigma^{-2} \sum_{l=1}^{m} (\rho_l n_l^2 \wedge p n_l)}.$$

Dividing numerator and denominator by p^2 gives

$$\mathbb{E}_\pi \mathbb{E}_f \left\| \hat{f} - f \right\|_{L^2}^2 \gtrsim \left[\frac{p^{2\alpha/d}}{R^2} + \frac{1}{\sigma^2} \sum_{l=1}^{m} \left(\frac{p}{n_l} \vee \frac{p^2}{\rho_l n_l^2} \right)^{-1} \right]^{-1}.$$

Since the prior is supported in $H^\alpha(R)$, the minimax reduction applies for each p. Taking the supremum over p proves the heterogeneous display.

For homogeneous clients, the nonprivate component is obtained by balancing

$$R^{-2} p^{2\alpha/d} \asymp \frac{mn}{\sigma^2 p},$$

which yields

$$R^{\frac{2d}{2\alpha+d}} \left(\frac{\sigma^2}{mn} \right)^{\frac{2\alpha}{2\alpha+d}}.$$

The privacy component is obtained by balancing

$$R^{-2} p^{2\alpha/d} \asymp \frac{\rho m n^2}{\sigma^2 p^2},$$

which yields

$$R^{\frac{2d}{\alpha+d}} \left(\frac{\sigma^2}{\rho mn^2} \right)^{\frac{\alpha}{\alpha+d}}.$$

Since a lower bound by the larger of finitely many components is equivalent, up to a universal constant, to a lower bound by their sum, the homogeneous rate in (30) follows whenever the optimizing resolutions are at least one. If a formal optimizer is smaller than one, the one-dimensional submodel gives a constant multiple of R^2. This proves the truncated homogeneous rate for all parameter values. □

2) Functional Mean Estimation: The privacy unit is one entire discretely observed curve. The finite-dimensional reductions below use localized wavelet submodels

$$f_\theta = \sum_{r=1}^{p} \theta_r \phi_r$$

with coefficient radius $p^{-\alpha-1/2}$. For these submodels,

$$\operatorname{Tr} J_\pi \asymp p^{2\alpha+2},$$

so the Sobolev truncation contribution after the p^2 van Trees numerator is $p^{2\alpha}$. Let Z have the cosine density p_0 above with $a = 1$. Then $\mathbb{E}Z = 0$, $|Z| \le 1$, and its location Fisher information is π^2. Because both p_0 and p_0' vanish at the endpoints, its translated location family satisfies the regularity conditions used below.

a) Independent design:

Proof of Proposition IV.9. For the independent design lower bound, take the measurement locations $T_{ij}^{(l)}$ to be private observations generated independently from a density on $[0,1]$ bounded above and below. Use the finite-dimensional bounded curve submodel

$$X_i^{(l)}(t) = f_\theta(t) + \sum_{r=1}^{p} \eta_{ir}^{(l)} \phi_r(t), \qquad \eta_{ir}^{(l)} = \tau_p Z_{ir}^{(l)}, \qquad Z_{ir}^{(l)} \overset{\text{i.i.d.}}{\sim} p_0, \qquad \tau_p^2 \asymp p^{-2\alpha-1}.$$

This variation at the curve level is admissible deterministically because the wavelet norm equivalence and $\left| Z_{ir}^{(l)} \right| \le 1$ give

$$\left\| X_i^{(l)} - f_\theta \right\|_{H^\alpha}^2 \lesssim p^{2\alpha} \sum_{r=1}^{p} \left| \eta_{ir}^{(l)} \right|^2 \le p^{2\alpha+1} \tau_p^2 \lesssim 1.$$

For a fixed client l, write

$$\varphi(t) = (\phi_1(t), \dots, \phi_p(t))^\top, \qquad \Phi_T = \begin{pmatrix} \varphi(T_{i1}^{(l)})^\top \\ \vdots \\ \varphi(T_{ik_l}^{(l)})^\top \end{pmatrix}.$$

Conditionally on the location vector $T = (T_{i1}^{(l)}, \dots, T_{ik_l}^{(l)})$, the observed vector $Y_i^{(l)} = (Y_{i1}^{(l)}, \dots, Y_{ik_l}^{(l)})^\top$ satisfies

$$Y_i^{(l)} = \Phi_T \theta + \Phi_T \eta_i^{(l)} + \xi_i^{(l)}, \qquad \xi_i^{(l)} \sim \mathcal{N}(0, I_{k_l}).$$

Let $I_{l,p}^{\mathrm{ind}}(\theta)$ be the Fisher information matrix of $(T, Y_i^{(l)})$ for θ. Revealing $\eta_i^{(l)}$ can only increase Fisher information, and conditionally on $(T, \eta_i^{(l)})$ the observation is a Gaussian location experiment. Therefore

$$I_{l,p}^{\mathrm{ind}}(\theta) \preceq \mathbb{E}_T\left[\Phi_T^\top \Phi_T\right]. \tag{46}$$

By the bounded design density and the L^2-normalization of the localized wavelets,

$$\mathbb{E}_T\left[\Phi_T^\top \Phi_T\right] = k_l \mathbb{E}_T[\varphi(T)\varphi(T)^\top] \preceq C k_l I_p,$$

which gives the first information bound. For the second bound, let $W = \theta + \eta_i^{(l)}$. The Markov chain $\theta \to W \to (T, Y_i^{(l)})$ and Fisher information processing give

$$I_{l,p}^{\mathrm{ind}}(\theta) \preceq I_W(\theta) = \pi^2 \tau_p^{-2} I_p \preceq C p^{2\alpha+1} I_p.$$

Combining the two bounds gives the explicit bounds

$$\left\| I_{l,p}^{\mathrm{ind}}(\theta) \right\|_{\mathrm{op}} \lesssim k_l \wedge p^{2\alpha+1}, \qquad \operatorname{Tr} I_{l,p}^{\mathrm{ind}}(\theta) \lesssim k_l p \wedge p^{2\alpha+2}. \tag{47}$$

Applying the trace contraction to both upper bounds and retaining the smaller allowed transcript information gives the following client contribution after division by the p^2 numerator:

$$\frac{k_l n_l}{p} \wedge \frac{\rho_l k_l n_l^2}{p^2} \wedge p^{2\alpha} n_l \wedge p^{2\alpha-1} \rho_l n_l^2.$$

Combining this contribution with the truncation term $p^{2\alpha}$ proves

$$\begin{aligned} &\inf_{\hat{f} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha(R, C_X)} \mathbb{E}_P \left\| \hat{f} - f_P \right\|_{L^2}^2 \\ &\qquad \gtrsim \sup_{p \geq 1} \left(p^{2\alpha} + \sum_{l=1}^{m} \left(\frac{k_l n_l}{p} \wedge \frac{\rho_l k_l n_l^2}{p^2} \wedge p^{2\alpha} n_l \wedge p^{2\alpha-1} \rho_l n_l^2 \right) \right)^{-1}. \end{aligned}$$

In the homogeneous case, the four displayed rate components are obtained as follows. The sample size component at the curve level is obtained by taking $p = 1$, giving $1/(mn)$. Balancing $p^{2\alpha}$ with mkn/p gives

$$p^{2\alpha+1} \asymp mkn, \qquad \text{risk} \asymp (mkn)^{-2\alpha/(2\alpha+1)}.$$

Balancing $p^{2\alpha}$ with $\rho mkn^2/p^2$ gives

$$p^{2\alpha+2} \asymp \rho mkn^2, \qquad \text{risk} \asymp (\rho mkn^2)^{-\alpha/(\alpha+1)}.$$

Finally, balancing the privacy bottleneck at low frequencies, $\rho mn^2 p^{2\alpha-1}$, at $p = 1$ gives $1/(\rho mn^2)$. Together these yield the homogeneous lower bound displayed in (34). The restriction $p \geq 1$ gives the outer truncation by the fixed Sobolev-radius scale in that display. □

b) Common design:

Proof of Proposition IV.10. We give the common design construction explicitly. Use the same sufficiently small constant c_{grid} as in Proposition IV.10, and suppose $k \geq c_{\mathrm{grid}}^{-1}$. For clarity, take a regular public grid $t_j = j/k$, $j = 1, \ldots, k$; the same argument requires only quasi uniform grid spacing. Fix $1 \leq p \leq c_{\mathrm{grid}} k$ and partition the grid into p consecutive blocks

$$B_r = \{j : \ t_j \in I_r\}, \qquad I_r = \left[\frac{r-1}{p}, \frac{r}{p} \right], \qquad r = 1, \ldots, p.$$

For a sufficiently small fixed c_{grid}, each block contains uniformly many grid points. Choose a fixed compactly supported smooth wavelet ψ with $\|\psi\|_{L^2} = 1$, and define

$$\phi_r(t) = p^{1/2} \psi(pt - r + 1), \qquad r = 1, \ldots, p,$$

after translating the supports inside I_r if necessary. The functions ϕ_r have disjoint supports, are orthonormal in L^2, and satisfy

$$\|\phi_r\|_{H^\alpha} \lesssim p^\alpha.$$

Hence the submodel

$$f_\theta(t) = \sum_{r=1}^{p} \theta_r \phi_r(t), \qquad |\theta_r| \lesssim p^{-\alpha-1/2},$$

is contained in the Sobolev ball. As in the independent design calculation, the product prior on these coefficients has

$$\operatorname{Tr} J_\pi \asymp p^{2\alpha+2}.$$

Let $v_r = (\phi_r(t_1), \dots, \phi_r(t_k))^\top \in \mathbb{R}^k$, and let Φ be the $p \times k$ matrix with rows $v_r^\top$. By disjoint support and the grid Riemann sum estimate,

$$v_r^\top v_s = 0 \quad (r \neq s), \qquad \|v_r\|_2^2 \asymp k.$$

Use the bounded submodel for the mean curve

$$X_i(t) = f_\theta(t) + \sum_{r=1}^{p} \eta_{ir}\phi_r(t), \qquad \eta_{ir} = \tau_p Z_{ir}, \qquad Z_{ir} \overset{\text{i.i.d.}}{\sim} p_0, \qquad \tau_p^2 = c/p,$$

with $c > 0$ fixed. The random curve variation has deterministically bounded L^2 energy since $\|X_i - f_\theta\|_{L^2}^2 \le p\tau_p^2 = c$. On the common grid, the observed vector $Y_i = (Y_{i1}, \dots, Y_{ik})^\top$ satisfies

$$Y_i = \Phi^\top \theta + \Phi^\top \eta_i + \xi_i, \qquad \xi_i \sim \mathcal{N}(0, I_k).$$

Let $W = \theta + \eta_i$. Since $\theta \to W \to Y_i$ is a Markov chain, Fisher information processing gives

$$I_p^{\text{com}}(\theta) \preceq I_W(\theta) = \pi^2 \tau_p^{-2} I_p \preceq CpI_p. \tag{48}$$

Consequently,

$$\left\| I_p^{\text{com}}(\theta) \right\|_{\text{op}} \lesssim p, \qquad \operatorname{Tr} I_p^{\text{com}}(\theta) \lesssim p^2. \tag{49}$$

After dividing the transcript information terms by p^2, client l's contribution is

$$n_l \wedge \frac{\rho_l n_l^2}{p}.$$

The public grid also creates the usual discretization obstruction. To make this explicit, choose a smooth bump h supported in $(0,1)$, and place a rescaled copy strictly between two adjacent grid points. If its width is $w \asymp k^{-1}$ and its amplitude is $cRw^{\alpha-1/2}$, then Sobolev scaling gives

$$\|h_k\|_{H^\alpha} \lesssim R, \qquad \|h_k\|_{L^2}^2 \asymp R^2 k^{-2\alpha}, \qquad h_k(t_j) = 0 \quad (j = 1, \dots, k).$$

Take the scalar subexperiment $X_i(t) = \beta h_k(t)$, with a smooth one-dimensional prior for $\beta \in [-1, 1]$. All public grid observations, and hence every public transcript, have distributions independent of β. The scalar van Trees inequality therefore gives a constant lower bound for the mean-squared error of estimating β. Since

$$\left\| \widehat{f} - \beta h_k \right\|_{L^2}^2 \ge \|h_k\|_{L^2}^2 \left(\frac{\langle \widehat{f}, h_k \rangle}{\|h_k\|_{L^2}^2} - \beta \right)^2,$$

this yields the required $k^{-2\alpha}$ term. Therefore

$$\inf_{\hat{f} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha(R, C_X)} \mathbb{E}_P \left\| \hat{f} - f_P \right\|_{L^2}^2 \gtrsim k^{-2\alpha} + \sup_{1 \le p \le c_{\text{grid}} k} \left(p^{2\alpha} + \sum_{l=1}^{m} \left(n_l \wedge \frac{\rho_l n_l^2}{p} \right) \right)^{-1}.$$

For homogeneous clients, taking $p = 1$ gives the $1/(mn)$ component when the raw sample size term is active. The privacy component is obtained by balancing

$$p^{2\alpha} \asymp \frac{\rho m n^2}{p}, \qquad p^{2\alpha+1} \asymp \rho m n^2,$$

which yields

$$(\rho m n^2)^{-2\alpha/(2\alpha+1)}.$$

If the optimizing p exceeds $c_{\text{grid}} k$, the discretization term $k^{-2\alpha}$ is active instead. This proves the rate displayed in (37). The restriction $p \ge 1$ gives the outer truncation by the fixed Sobolev-radius scale in that display. □

REFERENCES

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "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*. New York, NY, USA: Springer, 2006, pp. 265–284.

[2] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014. [Online]. Available: https://www.nowpublishers.com/article/Details/TCS-042

[3] T. T. Cai, Y. Wang, and L. Zhang, "The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy," *The Annals of Statistics*, vol. 49, no. 5, Oct. 2021. [Online]. Available: https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-5/The-cost-of-privacy--Optimal-rates-of-convergence-for/10.1214/21-AOS2058.full

[4] R. F. Barber and J. C. Duchi, "Privacy and Statistical Risk: Formalisms and Minimax Bounds," Dec. 2014. [Online]. Available: http://arxiv.org/abs/1412.4451

[5] J. Acharya, Z. Sun, and H. Zhang, "Differentially Private Assouad, Fano, and Le Cam," in *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 132. PMLR, mar 2021, pp. 48–78. [Online]. Available: https://proceedings.mlr.press/v132/acharya21a.html

[6] M. Bun, J. Ullman, and S. Vadhan, "Fingerprinting Codes and the Price of Approximate Differential Privacy," in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery, May 2014, pp. 1–10.

[7] C. Dwork, A. Smith, T. Steinke, J. Ullman, and S. Vadhan, "Robust Traceability from Trace Amounts," in *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*. Berkeley, CA, USA: IEEE Computer Society, Oct. 2015, pp. 650–669. [Online]. Available: https://doi.org/10.1109/FOCS.2015.46

[8] G. Kamath, A. Mouzakis, and V. Singhal, "New lower bounds for private estimation and a generalized fingerprinting lemma," *Advances in neural information processing systems*, vol. 35, pp. 24 405–24 418, 2022. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/hash/9a6b27821896649919449 1f55ccf8b75-Abstract-Conference.html

[9] N. Peter, E. Tsfadia, and J. Ullman, "Smooth Lower Bounds for Differentially Private Algorithms via Padding-and-Permuting Fingerprinting Codes," in *Proceedings of Thirty Seventh Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 247. PMLR, jun 2024, pp. 4207–4239. [Online]. Available: https://proceedings.mlr.press/v247/peter24a.html

[10] T. T. Cai, Y. Wang, and L. Zhang, "Score attack: A lower bound technique for optimal differentially private learning," Jul. 2025. [Online]. Available: http://arxiv.org/abs/2303.07152

[11] T. T. Cai and Y. Li, "Minimax and adaptive covariance matrix estimation under differential privacy," Mar. 2026. [Online]. Available: http://arxiv.org/abs/2603.19703

[12] T. T. Cai, A. Chakraborty, and L. Vuursteen, "Federated nonparametric hypothesis testing with differential privacy constraints: Optimal rates and adaptive tests," Jun. 2024. [Online]. Available: http://arxiv.org/abs/2406.06749

[13] M. Li, Y. Tian, Y. Feng, and Y. Yu, "Federated Transfer Learning with Differential Privacy," Apr. 2024. [Online]. Available: http://arxiv.org/abs/2403.11343

[14] T. T. Cai, A. Chakraborty, and L. Vuursteen, "Optimal federated learning for nonparametric regression with heterogeneous distributed differential privacy constraints," Jun. 2024. [Online]. Available: http://arxiv.org/abs/2406.06755

[15] A. Auddy, T. T. Cai, and A. Chakraborty, "Minimax and adaptive transfer learning for nonparametric classification under distributed differential privacy constraints," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, jul 2026. [Online]. Available: https://doi.org/10.1093/jrsssb/qkaf070

[16] G. Xue, Z. Lin, and Y. Yu, "Optimal estimation in private distributed functional data analysis," Dec. 2024. [Online]. Available: http://arxiv.org/abs/2412.06582

[17] E. K. H. Hung and Y. Yu, "Optimal Cox Regression under federated differential privacy: Coefficients and cumulative hazards," Aug. 2025. [Online]. Available: http://arxiv.org/abs/2508.19640

[18] T. T. Cai, A. Chakraborty, and L. Vuursteen, "The cost of adaptation under differential privacy: Optimal adaptive federated density estimation," Dec. 2025. [Online]. Available: http://arxiv.org/abs/2512.14337

[19] J. Acharya, C. L. Canonne, Z. Sun, and H. Tyagi, "Unified lower bounds for interactive high-dimensional estimation under information constraints," *Advances in Neural Information Processing Systems*, vol. 36, pp. 51 133–51 165, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/hash/a07e87ecfa8a651d62257571669b0150-Abstract-Conference.html

[20] L. P. Barnes, Y. Han, and A. Özgür, "Fisher information for distributed estimation under a blackboard communication protocol," in *2019 IEEE International Symposium on Information Theory (ISIT)*. Paris, France: IEEE, 2019, pp. 2704–2708. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8849821

[21] L. P. Barnes, W.-N. Chen, and A. Özgür, "Fisher Information Under Local Differential Privacy," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 645–659, Nov. 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9264231

[22] T. T. Cai, A. Chakraborty, and L. Vuursteen, "Optimal federated learning for functional mean estimation under heterogeneous privacy constraints," Dec. 2024. [Online]. Available: http://arxiv.org/abs/2412.18992

[23] R. Bassily, A. Smith, and A. Thakurta, "Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds," in *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 2014, pp. 464–473. [Online]. Available: https://doi.org/10.1109/focs.2014.56

[24] J. Duchi, M. J. Wainwright, and M. I. Jordan, "Local privacy and minimax bounds: Sharp rates for probability estimation," *Advances in Neural Information Processing Systems*, vol. 26, 2013. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2013/hash/5807a685d1a9ab3b599035bc566ce2b9-Abstract.html

[25] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local Privacy and Statistical Minimax Rates," in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 2013, pp. 429–438. [Online]. Available: https://doi.org/10.1109/FOCS.2013.53

[26] ——, "Minimax Optimal Procedures for Locally Private Estimation," *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 182–201, Jan. 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1389735

[27] W.-N. Chen and A. Özgür, "L_q lower bounds on distributed estimation via Fisher information," in *2024 IEEE International Symposium on Information Theory (ISIT)*. IEEE, apr 2024. [Online]. Available: https://doi.org/10.1109/isit57864.2024.10619530

[28] J. Acharya, Y. Liu, and Z. Sun, "Discrete Distribution Estimation under User-Level Local Differential Privacy," in *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, F. Ruiz, J. Dy, and J.-W. van de Meent, Eds., vol. 206. PMLR, Apr. 2023, pp. 8561–8585. [Online]. Available: https://proceedings.mlr.press/v206/acharya23a.html

[29] A. Kent, T. B. Berrett, and Y. Yu, "Rate optimality and phase transition for user-level local differential privacy," *Journal of the American Statistical Association*, jul 2026. [Online]. Available: https://doi.org/10.1080/01621459.2026.2677739

[30] A. Pensia, A. R. Asadi, V. Jog, and P.-L. Loh, "Simple Binary Hypothesis Testing Under Local Differential Privacy and Communication Constraints," *IEEE Transactions on Information Theory*, vol. 71, no. 1, pp. 592–617, Jan. 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10680565

[31] C. Butucea, A. Dubois, M. Kroll, and A. Saumard, "Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids," *Bernoulli*, vol. 26, no. 3, pp. 1727–1764, Aug. 2020. [Online]. Available: https://projecteuclid.org/journals/bernoulli/volume-26/issue-3/Local-differential-privacy--Elbow-effect-in-optimal-density-estimation/10.3150/19-BEJ1165.full

[32] C. Butucea, A. Rohde, and L. Steinberger, "Interactive versus non-interactive locally differentially private estimation: Two elbows for the quadratic functional," *The Annals of Statistics*, apr 2023. [Online]. Available: https://doi.org/10.1214/22-aos2254

[33] C. Butucea, K. Klockmann, and T. Krivobokova, "Nonparametric spectral density estimation using interactive mechanisms under local differential privacy," Apr. 2025. [Online]. Available: http://arxiv.org/abs/2504.00919

[34] M. Kroll, "Nonparametric spectral density estimation under local differential privacy," *Statistical Inference for Stochastic Processes*, vol. 27, no. 3, pp. 725–759, Oct. 2024. [Online]. Available: https://doi.org/10.1007/s11203-024-09308-3

[35] L. Zhu, M. Ding, V. Aggarwal, J. Xu, and D. Wang, "Improved Analysis of Sparse Linear Regression in Local Differential Privacy Model," Oct. 2023. [Online]. Available: http://arxiv.org/abs/2310.07367

[36] A. Lowy and M. Razaviyayn, "Private Federated Learning Without a Trusted Server: Optimal Algorithms for Convex Losses," Nov. 2024. [Online]. Available: http://arxiv.org/abs/2106.09779

[37] Z. Zhang, R. Nakada, and L. Zhang, "Differentially Private Federated Learning: Servers Trustworthiness, Estimation, and Statistical Inference," Apr. 2024. [Online]. Available: http://arxiv.org/abs/2404.16287

[38] Y. Allouah, R. Guerraoui, N. Gupta, R. Pinot, and J. Stephan, "On the privacy-robustness-utility trilemma in distributed learning," in *International Conference on Machine Learning*. Honolulu, Hawaii, USA: PMLR, 2023, pp. 569–626. [Online]. Available: https://proceedings.mlr.press/v202/allouah23a.html

[39] J. Li, T. T. Cai, D. Xia, and A. R. Zhang, "Federated PCA and Estimation for Spiked Covariance Matrices: Optimal Rates and Efficient Algorithm," Nov. 2024. [Online]. Available: http://arxiv.org/abs/2411.15660

[40] C. Dwork and G. N. Rothblum, "Concentrated Differential Privacy," Mar. 2016. [Online]. Available: http://arxiv.org/abs/1603.01887

[41] M. Bun and T. Steinke, "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds," in *Lecture Notes in Computer Science*, ser. Lecture Notes in Computer Science. Springer, 2016, vol. 9985, pp. 635–658. [Online]. Available: https://doi.org/10.1007/978-3-662-53641-4_24

[42] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, Dec. 1952. [Online]. Available: https://doi.org/10.1093/biomet/39.3-4.324

[43] S. Negahban, S. Oh, and D. Shah, "Rank centrality: Ranking from pairwise comparisons," *Operations Research*, vol. 65, no. 1, pp. 266–287, 2017. [Online]. Available: https://doi.org/10.1287/opre.2016.1534

[44] R. D. Gill and B. Y. Levit, "Applications of the van Trees inequality: A Bayesian Cramér-Rao bound," *Bernoulli*,

vol. 1, no. 1-2, pp. 59–79, Mar. 1995. [Online]. Available: https://projecteuclid.org/journals/bernoulli/volume-1/issue-1-2/Applications-of-the-van-Trees-inequality--a-Bayesian-Cram%c3%a9r/bj/1186078362.full

[45] E. Gassiat and G. Stoltz, "The van Trees inequality in the spirit of Hájek and Le Cam," *Statistical Science*, vol. 39, no. 4, pp. 644–653, 2024, arXiv: 2402.06431. [Online]. Available: https://projecteuclid.org/journals/statistical-science/volume-39/issue-4/The-van-Trees-Inequality-in-the-Spirit-of-H%C3%A1jek-and/10.1214/24-STS941.short

[46] T. van Erven and P. Harremoës, "Rényi Divergence and Kullback-Leibler Divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014. [Online]. Available: http://arxiv.org/abs/1206.2459

[47] M. Sion, "On general minimax theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, pp. 171–176, 1958.