Submitted to the Annals of Statistics

MINIMAX AND ADAPTIVE COVARIANCE MATRIX ESTIMATION UNDER DIFFERENTIAL PRIVACY

BY T. TONY CAI[1,a] AND YICHENG LI[2,b]

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

[2]*KLATASDS-MOE, School of Statistics, East China Normal University*
[b]*ycli@sfs.ecnu.edu.cn*

Estimating covariance matrices is fundamental to a wide range of statistical applications. This paper studies minimax and adaptive estimation of high-dimensional covariance matrices under ρ-zero-concentrated differential privacy (ρ-zCDP) over three nested classes: the pointwise-decay class $\mathcal{H}_\alpha$, the row-tail class $\mathcal{G}_\alpha$, and the separated-block class $\mathcal{F}_\alpha$. We consider both squared operator norm loss and normalized squared Frobenius norm loss.

For $\mathcal{H}_\alpha$ and $\mathcal{G}_\alpha$, we develop center–outer dyadic estimators tailored to the refined geometry of the two classes, while for $\mathcal{F}_\alpha$, we develop a blockwise tridiagonal estimator. The resulting minimax-optimal rates reveal a nontrivial interplay among the smoothness α, the loss, the geometry of the covariance class, and the privacy constraint. In contrast to the non-private setting, privacy distinguishes covariance classes that share the same leading non-private rate and induces a polynomial dependence on the ambient dimension.

We further develop procedures that adapt to the unknown decay parameter over all three covariance classes under both losses, at the cost of at most polylogarithmic factors. To establish minimax lower bounds, we develop a novel differentially private van Trees inequality that connects Fisher information with the ρ-zCDP constraint and may be useful for other private estimation problems. We also construct carefully designed prior distributions to obtain matching minimax lower bounds.

1. Introduction. Modern statistical analysis increasingly relies on individual-level data that are both high dimensional and highly sensitive. In applications such as biomedical research, genomics, electronic health records, finance, social science, and federated data analysis, each observation may contain detailed information about a person or institution. Covariance estimation is often a first step in analyzing such data, because covariance structure underlies principal component analysis, discriminant analysis, clustering, regression, graphical modeling, and many other multivariate methods. At the same time, releasing or using covariance estimates without privacy protection can reveal sensitive information about individuals. Improper disclosure can lead to harms such as identity theft, discrimination, loss of confidentiality, or loss of public trust. These risks are particularly acute in healthcare and biomedical studies, where genetic and medical records enable important scientific advances but are highly sensitive [2, 29]. Thus, modern covariance estimation must balance statistical accuracy with rigorous privacy guarantees.

Differential privacy (DP) [26, 28] has emerged as a leading mathematical framework for privacy-preserving data analysis. It provides a guarantee that the inclusion or exclusion of any single individual's data has only a limited effect on the output of an analysis, thereby protecting against inference attacks even when an adversary has auxiliary information. This

MSC2020 subject classifications: Primary 62H12; secondary 62F12.

Keywords and phrases: Adaptive estimation, Covariance matrix, Differential privacy, Minimax rate of convergence.

robustness makes DP especially attractive for statistical procedures intended for sensitive data. However, privacy is not free. A private estimator must inject sufficient randomness, or otherwise limit the information released, in order to protect individuals. This added randomization can substantially degrade statistical accuracy. A central question in private statistics, therefore, is to characterize the optimal privacy–utility trade-off: how much accuracy must be sacrificed to guarantee privacy, and how should estimators be designed to achieve the best possible balance?

This paper studies that question for high-dimensional covariance matrix estimation under differential privacy. Suppose we observe independent and identically distributed random vectors $x_1, \ldots, x_n \in \mathbb{R}^d$ with population covariance matrix Σ. In classical low-dimensional settings, the sample covariance matrix performs well. Modern datasets, however, often operate in the high-dimensional regime where d is comparable to, or much larger than, n. In this regime the sample covariance matrix is unstable or singular, and structural assumptions are needed for accurate estimation. A widely studied structural assumption is that of *bandable* covariance matrices, where off-diagonal entries decay as they move away from the diagonal. Such matrices arise naturally in temporal, spatial, longitudinal, and other ordered data settings, reflecting the weakening of correlations with increasing separation. Optimal rates and adaptive procedures for estimating bandable covariance matrices in the non-private setting have been established under both the operator and Frobenius norms. In particular, Cai, Zhang and Zhou [21] derived the minimax rate of convergence under the operator norm and proposed an optimal-rate tapering estimator. Later, Cai and Yuan [20] introduced a block-thresholding estimator that adaptively attains the optimal rate over a collection of parameter spaces without requiring knowledge of the decay parameter. See Cai, Ren and Zhou [15] for a comprehensive survey of structured covariance estimation.

The interaction between bandable structure and differential privacy is subtle. In non-private covariance estimation, bandability reduces the effective complexity of the parameter space and leads to estimators that outperform the unstructured sample covariance matrix. Under privacy constraints, however, the estimator must also control the sensitivity of the released object. The geometry of the covariance class can therefore affect not only the statistical bias–variance trade-off, but also the amount and placement of privacy noise. Consequently, it is not enough to privatize a non-private estimator in a black-box fashion. One must understand how the structural decay of the covariance matrix, the loss function, and the privacy constraint jointly determine the minimax risk.

We work primarily with the ρ-zero-concentrated differential privacy (ρ-zCDP) framework [10, 27]. While (ϵ, δ)-DP remains the most common privacy formulation, zCDP is based on Rényi divergence and provides a single-parameter privacy budget with tight composition properties. This is particularly useful for the multiscale and blockwise procedures developed in this paper. Recall first that a randomized algorithm M satisfies (ϵ, δ)-DP if, for any adjacent datasets $S, S' \in \mathcal{D}$ differing by one observation and any measurable output set A,

$$\mathbb{P}\{M(S) \in A\} \leq e^{\epsilon}\mathbb{P}\left\{M(S') \in A\right\} + \delta.$$

For $\alpha > 1$, the α-Rényi divergence between random variables X and Y, associated with probability measures P and Q, is defined as

$$D_\alpha(X\|Y) = \frac{1}{\alpha - 1}\log \mathbb{E}\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(X)\right)^{\alpha}.$$

The definition of ρ-zCDP is as follows.

DEFINITION 1.1 (ρ-zCDP). A randomized algorithm $M : \mathcal{D} \to \mathcal{A}$ satisfies ρ-zero-concentrated DP (ρ-zCDP) if for all adjacent $S, S' \in \mathcal{D}$,

$$D_\alpha(M(S)\|M(S')) \leq \rho\alpha, \quad \forall \alpha \in (1, \infty), \tag{1}$$

where $D_\alpha(M(S)\|M(S'))$ is taken only over the randomness of M conditioned on the data.

The connection between ρ-zCDP and (ϵ,δ)-DP is well established [10, 26].

LEMMA 1.2. *If M is $(\epsilon,0)$-DP, then M is $(\epsilon^2/2)$-zCDP. Conversely, if M is ρ-zCDP, then M is $\left(\rho+2\sqrt{\rho\log(1/\delta)},\delta\right)$-DP for every $\delta\in(0,1)$. In particular, for $\epsilon\in(0,1]$ and $\delta\in(0,e^{-1}]$, ρ-zCDP with $\rho\le\epsilon^2(8\log(1/\delta))^{-1}$ implies (ϵ,δ)-DP.*

Although (ϵ,δ)-DP is widely used, the divergence-based formulation of zCDP often simplifies the analysis of complex algorithms and yields sharper composition bounds. For this reason, we adopt ρ-zCDP as the primary privacy notion in this paper; the corresponding (ϵ,δ)-DP guarantees follow from the lemma above.

In this paper, we develop minimax and adaptation theory for estimating bandable covariance matrices under differential privacy. We consider three nested covariance classes: the pointwise-decay class $\mathcal{H}_\alpha$, the row-tail class $\mathcal{G}_\alpha$, and the separated-block class $\mathcal{F}_\alpha$, where α denotes the smoothness parameter. The precise definitions of these classes are provided in Section 2.2. They satisfy the inclusion $\mathcal{H}_\alpha\subset\mathcal{G}_\alpha\subset\mathcal{F}_\alpha$, up to a constant factor in the class radius.

1.1. *Main Contribution.* We first characterize the minimax risks under squared operator loss and normalized squared Frobenius loss over all three classes. The principal finding is a geometric separation created by privacy. Under operator loss, the classes have the same leading non-private rate, but the privacy cost over the separated-block class $\mathcal{F}_\alpha$ is algebraically larger than the costs over the pointwise-decay and row-tail classes. Under Frobenius loss, the non-private rates already separate $\mathcal{F}_\alpha$ from $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, and privacy creates a further separation between the separated-block class and the two smaller classes. Thus privacy depends not only on the decay exponent α, but also on the local geometry through which that decay is imposed. This distinction is invisible in the leading non-private operator rate and is one of the central phenomena identified by the paper.

TABLE 1

Leading algebraic terms under squared operator and normalized squared Frobenius losses in the high-dimensional regime. Logarithmic factors suppressed. Denote $x=d/(\rho n^2)$. The marker "(log)" indicates a fixed polylogarithmic gap between the current upper and lower bounds.

	Non-private rate		Cost of privacy	
Class	Operator	Frobenius	Operator	Frobenius
$\mathcal{F}_\alpha$	$n^{-\frac{2\alpha}{2\alpha+1}}$	$n^{-\frac{2\alpha}{2\alpha+1}}$	$x^{\frac{\alpha}{\alpha+1}}$	$x^{\frac{\alpha}{\alpha+1}}$
$\mathcal{G}_\alpha$	$n^{-\frac{2\alpha}{2\alpha+1}}$	$n^{-\frac{2\alpha+1}{2\alpha+2}}$	$x^{\frac{2\alpha}{2\alpha+1}}\vee x^{\frac{2\alpha+1}{2\alpha+3}}$ (log)	$x^{\frac{2\alpha+1}{2\alpha+3}}$ (log)
$\mathcal{H}_\alpha$	$n^{-\frac{2\alpha}{2\alpha+1}}$	$n^{-\frac{2\alpha+1}{2\alpha+2}}$	$x^{\frac{2\alpha}{2\alpha+1}}\vee x^{\frac{2\alpha+1}{2\alpha+3}}$ (log)	$x^{\frac{2\alpha+1}{2\alpha+3}}$

Moreover, the role of covariance geometry depends on the choice of loss. In the non-private setting, the operator norm does not distinguish among the three classes in terms of rate of convergence, whereas the Frobenius norm separates $\mathcal{F}_\alpha$ from the two smaller classes $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$. Under privacy, both losses distinguish $\mathcal{F}_\alpha$ from $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, but they do so in qualitatively different ways. For $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, the privacy cost under the operator norm exhibits a phase transition: the first operator term in Table 1 dominates when $\alpha<1/2$, the second dominates when $\alpha>1/2$, and the two terms coincide at $\alpha=1/2$. No such transition occurs under the Frobenius loss, where the privacy cost follows a single algebraic regime.

We then construct private estimators tailored to the geometry of each class. For $\mathcal{F}_\alpha$, we introduce a blockwise tridiagonal estimator to balance the statistical cost, privacy cost and truncation bias. For $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, the useful local constraints occur at every distance from the diagonal, and a single block scale does not directly expose this multiscale structure while keeping the privacy noise under control. We therefore organize the off-diagonal entries into dyadic increments and refine them into square blocks that separate the distance scales and support procedures tailored to each class. This construction yields dyadic profile and nonlinear peeling estimators for $\mathcal{G}_\alpha$, and a dyadic greedy procedure for $\mathcal{H}_\alpha$. More specifically, pointwise decay makes each dyadic outer block small in Frobenius norm, which enables a greedy rank-one release under operator loss. The row-tail condition instead provides a local row–column ℓ^1 constraint, leading to profile blocks under operator loss and peeling blocks under Frobenius loss. For $\mathcal{H}_\alpha$ under Frobenius loss, the blockwise tridiagonal construction already exploits the available entrywise decay. These constructions attain the corresponding minimax rates up to logarithmic factors in certain cases.

The methodological novelty is therefore not a single privatized tapering estimator, but a collection of releases tailored to the local geometry of each class: tridiagonal blocks for separated-block control, greedy low-rank releases for pointwise decay, and profile or peeling releases for row-tail decay. A common dyadic decomposition integrates these mechanisms while preserving the correct polynomial dependence on d, n, and ρ.

Another key contribution is the DP van Trees inequality in Theorem 4.1. Under the stated regularity conditions, we establish a contraction of Fisher information implied by zCDP and combine it with the classical Bayesian Cramér–Rao inequality. Let $I_x(\theta)$ denote the Fisher information in one observation, let J_π denote the Fisher information of the prior π, and define $C_\rho = (e^{2\rho} - 1)/\rho$. The denominator in the resulting Bayes risk bound contains

$$\min\left\{ n \int \operatorname{Tr} I_x(\theta)\, \mathrm{d}\pi(\theta),\, C_\rho \rho n^2 \int \|I_x(\theta)\|\, \mathrm{d}\pi(\theta) \right\} + \operatorname{Tr} J_\pi.$$

The first term inside the minimum is the classical contribution from the sample information, governed by the trace of the Fisher information, whereas the second is the information limit imposed by privacy, governed by its operator norm. This distinction between trace and operator norm yields the correct polynomial dependence on the dimension in the private lower bounds without introducing an additional logarithmic loss.

Using the DP van Trees inequality, we derive minimax lower bounds for all three covariance classes. In particular, we construct a blockwise diagonal prior with random blocks whose operator norms are controlled for $\mathcal{F}_\alpha$; a banded entrywise prior for the Frobenius lower bound over $\mathcal{H}_\alpha$; and block priors of intermediate rank that lie simultaneously in $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$ for their common operator norm lower bound. These constructions isolate the different geometric obstructions created by the separated-block, row-tail, and pointwise-decay constraints. Combined with the upper bounds, these results show precisely which distinctions among the three classes are induced by privacy.

The inequality is particularly well suited to structured matrix problems because, unlike approaches that rely on priors with independent coordinates or specially constructed testing packings, it accommodates dependent matrix-valued priors. Because it requires only suitable control of the likelihood Fisher information and the prior Fisher information, the lower-bound method may also be of independent interest for other high-dimensional estimation problems under differential privacy.

Finally, for each specified covariance class and loss, we construct private estimators that adapt to the unknown smoothness parameter α. The unknown α enters the blockwise tridiagonal and center–outer procedures differently, requiring two adaptive constructions. For $\mathcal{F}_\alpha$,

thresholding the dyadic tridiagonal increments yields operator norm adaptation and, consequently, normalized Frobenius adaptation, while a Frobenius norm thresholded version handles $\mathcal{H}_\alpha$ under Frobenius loss. These procedures preserve the known smoothness statistical terms while incurring only logarithmic factors in the privacy terms. For operator loss over $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, and Frobenius loss over $\mathcal{G}_\alpha$, both the local radius and the preferred release branch vary with α and scale. We combine public radius grids, deterministic selection among zero, structured, and dense branches, and Lepski comparisons across scales to match the known smoothness benchmarks uniformly over compact smoothness intervals, up to fixed polylogarithmic factors.

In particular, adaptation introduces no additional algebraic loss: the adaptive procedures retain the powers of n, d, and ρ corresponding to known smoothness, with only fixed polylogarithmic overhead.

1.2. *Related Work.* Existing work on private covariance estimation has largely focused on unstructured matrices. Under appropriate conditions, calibrated Gaussian perturbation of the sample covariance attains the benchmark under squared operator loss

$$\frac{d}{n} + \frac{d^3}{\rho n^2}$$

[4, 23, 32, 34]. Related work studies Gaussian covariance estimation under Mahalanobis loss, including ill-conditioned models[3, 5, 9, 31], as well as private estimation of spiked covariance matrices [18]. These benchmarks do not exploit ordered off-diagonal decay and can therefore be substantially suboptimal for the structured covariance classes considered here. This paper quantifies the resulting minimax improvement and shows that it depends on both the precise form of the decay constraint and the loss function.

In contrast, there is a rich literature on structured covariance estimation in the non-private setting. For bandable covariance matrices, regularization methods including banding [7], tapering [21], thresholding [8], and adaptive block-thresholding [20] have been proposed to exploit the structural decay of correlations. Optimal rates of convergence under both operator and Frobenius norms have been established [21]. Other commonly studied structures include sparse covariance matrices [12, 22], Toeplitz covariance matrices [14], and sparse precision matrices [13]. See Cai, Ren and Zhou [15] for a comprehensive overview of this literature.

Privacy constraints typically introduce a "cost of privacy" in minimax rates. Several techniques have been developed for proving minimax lower bounds under differential privacy. These include private analogues of classical information-theoretic tools such as Fano's inequality and Le Cam's method[1, 24, 25], fingerprinting arguments[6, 30, 33], score attacks[16, 17], and van Trees inequalities[11, 36]. More recently, Narayanan [32], Portella and Harvey [34] used Wishart constructions to establish minimax lower bounds for private estimation of general unstructured covariance matrices.

Within this literature, the differentially private van Trees inequality developed here is especially useful because its trace–operator separation yields the correct dimension dependence for the structured covariance lower bounds; its methodological contribution is described in Section 1.1 and developed in Section 4.

1.3. *Organization.* The rest of the paper is organized as follows. We conclude this section with basic notation. Section 2 introduces the statistical model and the three covariance classes and states the main minimax results with discussions. Section 3 develops the private estimators: a blockwise tridiagonal estimator for the separated-block class, a greedy center–outer procedure for the pointwise-decay class under operator loss, and profile and peeling procedures for the row-tail class under operator and Frobenius losses, respectively.

The pointwise-decay class under Frobenius loss is handled by the blockwise tridiagonal construction. Section 4 presents the DP van Trees inequality and derives minimax lower bounds for bandable covariance matrix estimation under DP. In Section 5, we develop estimators adaptive to smoothness and analyze their performance under both losses. Section 6 considers extensions of our methodology to precision matrix estimation and Section 7 concludes with a discussion of our findings and directions for future research.

1.4. *Notation.* We use C and c to denote generic positive constants that may vary from line to line. Write $a \lesssim b$ if there exists a constant $C > 0$ such that $a \le Cb$, and write $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. Write $a \lesssim_{\log} b$ if $a \le C \log^A(edn/\rho) b$ for fixed $C, A > 0$, and write $a \asymp_{\log} b$ if $a \lesssim_{\log} b$ and $b \lesssim_{\log} a$. We use the notations $a \wedge b$ and $a \vee b$ to denote $\min(a, b)$ and $\max(a, b)$, respectively. The indicator function is denoted by $\mathbf{1}\{\cdot\}$, and $|S|$ denotes the cardinality of a set S. We write $[d] = \{1, 2, \dots, d\}$. For $R > 0$, define the radial clipping map on $\mathbb{R}^m$ by $\chi_R(0) = 0$ and $\chi_R(z) = z \min\{1, R/\|z\|_2\}$ for $z \neq 0$.

For a matrix A, $\|A\|$ denotes the spectral (operator) norm and $\|A\|_F$ the Frobenius norm. We write $\langle A, B\rangle = \mathrm{Tr}(A^\top B)$ for the trace inner product. The matrix ℓ^r norm is defined as $\|A\|_{\ell^r} = \max_{\|x\|_{\ell^r}=1} \|Ax\|_{\ell^r}$. In particular, $\|A\|_{\ell^2} = \|A\|$ is the spectral norm, $\|A\|_{\ell^\infty} = \max_i \sum_j |A_{ij}|$ is the maximum absolute row sum, and $\|A\|_{\ell^1} = \max_j \sum_i |A_{ij}|$ is the maximum absolute column sum. The matrix Schatten-q norm is defined as $\|A\|_{\mathcal{S}_q} = (\sum_i \sigma_i(A)^q)^{1/q}$, where $\sigma_i(A)$ are the singular values of A. In particular, $\|A\|_{\mathcal{S}_2} = \|A\|_F$ is the Frobenius norm and $\|A\|_{\mathcal{S}_\infty} = \|A\|$ is the spectral norm. Write $\Pi_{M_0}(A)$ for the spectral projection of a symmetric matrix to $[0, M_0]$. For a matrix $M \in \mathbb{R}^{d\times d}$ and an index set $B \subseteq [d]^2$, we denote by M_B or $M[B]$ the matrix obtained by setting all entries outside B to zero. A similar notation v_I applies to vectors. We call B a block if $B = I \times J$ for some $I, J \subseteq [d]$, and say that it has size k if $|I| = |J| = k$.

The sub-Gaussian norm [35] for a random variable ξ is defined as $\|\xi\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}(e^{\xi^2/t^2}) \le 2\}$, and for a random vector X as $\|X\|_{\psi_2} := \sup_{v:\|v\|_2=1} \|\langle X, v\rangle\|_{\psi_2}$.

2. Model, Covariance Classes and Results.

2.1. *Statistical model.* Let $X_1, \dots, X_n$ be independent copies of a random vector $X \in \mathbb{R}^d$, with mean $\mu = \mathbb{E}X$ and covariance matrix

$$\Sigma = \mathbb{E}\left[(X - \mu)(X - \mu)^\top\right].$$

For fixed constants $K, M_0 < \infty$, we assume

$$\|X - \mu\|_{\psi_2} \le K, \qquad 0 \preceq \Sigma \preceq M_0 I_d. \tag{2}$$

We consider the pairing symmetrization $W_i = (X_{2i} - X_{2i-1})/\sqrt{2}$, for $1 \le i \le \lfloor n/2 \rfloor$. Then, the paired observations are independent and satisfy $\mathbb{E}W_i = 0$, $\mathrm{Cov}(W_i) = \Sigma$ and $\|W_i\|_{\psi_2} \le CK$. Moreover, replacing one original record changes at most one W_i. Consequently, composing any ρ-zCDP mechanism for the paired data recovers a ρ-zCDP mechanism for the original data. Therefore, since our focus is the optimal rates, we assume that $\mu = 0$ without loss of generality throughout the paper.

In this paper, for an estimator $\hat{\Sigma}$, we consider the squared operator norm loss $\|\hat{\Sigma} - \Sigma\|^2$ and the normalized squared Frobenius norm loss $d^{-1}\|\hat{\Sigma} - \Sigma\|_F^2$. The normalization makes the two losses comparable due to the relation $d^{-1}\|A\|_F^2 \le \|A\|^2$.

Throughout, we consider the multidimensional regime $d \gtrsim \log n$. We focus on the nondegenerate privacy regime $d/(\rho n^2) = o(1)$; otherwise, the minimax lower bounds show that the

rates degenerate. In terms of privacy, we consider the meaningfully private case $0 < \rho \lesssim 1$, where the implicit upper bound is fixed and the constants below may depend on it. Hence, $d = o(n^2)$. All asymptotic statements below are understood along sequences satisfying these conditions.

2.2. *Three covariance classes.* We consider the following classes of bandable covariance matrices. The first two are standard in the non-private covariance estimation literature [7, 8, 20, 21]. The third class is introduced in this paper and turns out to possess especially favorable geometric properties for both non-private and private estimation. The first class imposes a power-law decay condition on the covariance entries as their distance from the main diagonal increases.

DEFINITION 2.1 (Pointwise-decay class). For $\alpha, M_0, M > 0$, let $\mathcal{H}_\alpha(M_0, M)$ be the set of covariance matrices satisfying $0 \preceq \Sigma \preceq M_0 I_d$ and

$$|\Sigma_{ij}| \le M\,|i-j|^{-(\alpha+1)}, \qquad i \ne j. \tag{3}$$

Another closely related class controls the tail sum of each row [20, 21]:

DEFINITION 2.2 (Row-tail class). For $\alpha, M_0, M > 0$, let $\mathcal{G}_\alpha(M_0, M)$ be the set of covariance matrices satisfying $0 \preceq \Sigma \preceq M_0 I_d$ and

$$\max_{j\in[d]} \sum_{i:\,|i-j|>k} |\Sigma_{ij}| \le M k^{-\alpha}, \qquad k \ge 1. \tag{4}$$

These bandable covariance classes encode the decay of covariance away from the main diagonal. Another natural way to express this structure is to control blocks separated from the diagonal. We call a block R k-off-diagonal if it is contained in $\{(i,j) : j \ge i + k\}$ or, symmetrically, in $\{(i,j) : i \ge j + k\}$. For example, $[1, \dots, i_0] \times [i_0 + k, \dots, d]$ is a k-off-diagonal block. This motivates the following definition.

DEFINITION 2.3 (Separated-block class). For $\alpha, M_0, M > 0$, let $\mathcal{F}_\alpha(M_0, M)$ be the set of covariance matrices satisfying $0 \preceq \Sigma \preceq M_0 I_d$ and

$$\|\Sigma_R\| \le M k^{-\alpha}, \qquad \text{for every } k \ge 1 \text{ and every } k\text{-off-diagonal block } R. \tag{5}$$

Alternatively, the class can be described in terms of the operator norm of covariances between two k-separated index sets $I, J \subseteq [d]$, i.e. $\min_{i\in I, j\in J} |i-j| \ge k$:

$$\sup_{k\text{-separated } I,J} \|\mathrm{Cov}(x_I, x_J)\| \le C_1 k^{-\alpha}.$$

From this perspective, the class $\mathcal{F}_\alpha$ captures the decay of long-range dependencies in the random vector x. Thus, it is a natural model for various applications where such decay occurs, e.g., time series analysis and spatial statistics.

Although formulated at different levels, the three classes are closely related. Combining $\|A\|^2 \le \|A\|_{\ell^1}\,\|A\|_{\ell^\infty}$ with tail summation yields the following inclusion relation, up to a constant adjustment of the class radius.

PROPOSITION 2.4 (Class inclusions). *For every fixed $\alpha, M_0, M > 0$, there is a constant $C_\alpha > 0$, depending only on α, such that*

$$\mathcal{H}_\alpha(M_0, M) \subseteq \mathcal{G}_\alpha(M_0, C_\alpha M) \subseteq \mathcal{F}_\alpha\big(M_0, C_\alpha(M + M_0)\big).$$

Moreover, we have reverse inclusions $\mathcal{F}_{\alpha+1/2} \subseteq \mathcal{G}_\alpha$ and $\mathcal{F}_{\alpha+1} \subseteq \mathcal{H}_\alpha$.

2.3. *Results overview.* For $\mathcal{J} \in \{\mathcal{F}, \mathcal{G}, \mathcal{H}\}$, we write the class of distributions

$$\mathcal{P}_\alpha^{\mathcal{J}}(K, M_0, M) = \{P : P \text{ satisfies (2) and } \Sigma(P) \in \mathcal{J}_\alpha(M_0, M)\}.$$

When K, M_0, M are fixed, we suppress them and write $\mathcal{P}_\alpha^{\mathcal{J}}$ for readability. Let $\mathcal{M}_\rho$ be the collection of all ρ-zCDP estimators based on n observations. To uniformly characterize the optimal rates for the classes $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, define

$$\mathcal{Q}_{\alpha,d}(u) = \sup_{\substack{1 \le k \le \lfloor d/4 \rfloor \\ 1 \le r \le k}} \min\left(\frac{k^{-2\alpha}}{r}, ukr\right), \tag{6}$$

where $u > 0$ and k, r are integers. Under $d/(\rho n^2) = o(1)$, direct optimization in (6) gives the following explicit form. If $\alpha \ge 1/2$, then

$$\mathcal{Q}_{\alpha,d}\left(\frac{d}{\rho n^2}\right) \asymp_\alpha \begin{cases} \dfrac{d^3}{\rho n^2}, & d \lesssim (\rho n^2)^{1/(2\alpha+4)}, \\ \left(\dfrac{d}{\rho n^2}\right)^{\frac{2\alpha+1}{2\alpha+3}}, & d \gtrsim (\rho n^2)^{1/(2\alpha+4)}. \end{cases} \tag{7}$$

If $0 < \alpha < 1/2$, then

$$\mathcal{Q}_{\alpha,d}\left(\frac{d}{\rho n^2}\right) \asymp_\alpha \begin{cases} \dfrac{d^3}{\rho n^2}, & d \lesssim (\rho n^2)^{1/(2\alpha+4)}, \\ \dfrac{d^{1-\alpha}}{\sqrt{\rho}\, n}, & (\rho n^2)^{1/(2\alpha+4)} \lesssim d \lesssim (\rho n^2)^{1/(2\alpha+2)}, \\ \left(\dfrac{d}{\rho n^2}\right)^{\frac{2\alpha}{2\alpha+1}}, & d \gtrsim (\rho n^2)^{1/(2\alpha+2)}. \end{cases} \tag{8}$$

The first result gives matching minimax rates over the separated-block class under both losses.

THEOREM 2.5 (Optimal rates under $\mathcal{F}_\alpha$). *Fix $\alpha, K, M_0, M > 0$. If $\rho n^2/d \gtrsim (\log d)^{2(\alpha+1)}$, then*

$$\begin{aligned} \inf_{\hat{\Sigma} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha^{\mathcal{F}}} \mathbb{E}_P \left\| \hat{\Sigma} - \Sigma(P) \right\|^2 &\asymp \inf_{\hat{\Sigma} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha^{\mathcal{F}}} \frac{1}{d} \mathbb{E}_P \left\| \hat{\Sigma} - \Sigma(P) \right\|_F^2 \\ &\asymp n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \left(\frac{d}{\rho n^2}\right)^{\frac{\alpha}{\alpha+1}} \wedge \frac{d^3}{\rho n^2}. \end{aligned} \tag{9}$$

The stronger local structure of the row-tail and pointwise-decay classes leads to a common and generally smaller operator-norm rate.

THEOREM 2.6 (Operator norm; $\mathcal{G}_\alpha$, $\mathcal{H}_\alpha$). *Fix $\alpha, K, M_0, M > 0$. For each $\mathcal{J} \in \{\mathcal{G}, \mathcal{H}\}$,*

$$\inf_{\hat{\Sigma} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha^{\mathcal{J}}} \mathbb{E}_P \left\| \hat{\Sigma} - \Sigma(P) \right\|^2 \asymp_{\log} n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \mathcal{Q}_{\alpha,d}\left(\frac{d}{\rho n^2}\right). \tag{10}$$

Under squared Frobenius loss, the two smaller classes share the same algebraic rate, with a fixed polylogarithmic gap only for the row-tail class.

THEOREM 2.7 (Frobenius norm; $\mathcal{G}_\alpha$, $\mathcal{H}_\alpha$). *Fix $\alpha, K, M_0, M > 0$. For each $\mathcal{J} \in \{\mathcal{G}, \mathcal{H}\}$,*

$$\inf_{\hat{\Sigma} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha^{\mathcal{J}}} \frac{1}{d} \mathbb{E}_P \left\| \hat{\Sigma} - \Sigma(P) \right\|_F^2 \asymp_{\log} n^{-\frac{2\alpha+1}{2\alpha+2}} \wedge \frac{d}{n} + \left(\frac{d}{\rho n^2}\right)^{\frac{2\alpha+1}{2\alpha+3}} \wedge \frac{d^3}{\rho n^2}. \tag{11}$$

For $\mathcal{J} = \mathcal{H}$, the relation can be strengthened by replacing $\asymp_{\log}$ with $\asymp$.

In the following sections, Section 3 constructs DP estimators attaining the upper bounds when the smoothness α is known, Section 4 gives the minimax lower bounds, and Section 5 establishes procedures adapting to unknown smoothness.

2.4. *Discussions.* The three preceding theorems show how the covariance class, loss function, privacy budget, and ambient dimension jointly determine the minimax risk. We discuss how local geometry changes the optimal procedure and rate, how dimension determines the active regime, and where logarithmic gaps remain. The structured rates below should be read together with their finite-dimensional minima against the unstructured terms.

2.4.1. *Geometry, loss, and estimator design.* Under squared operator loss, the three classes share the same non-private term $n^{-2\alpha/(2\alpha+1)}$, but their privacy costs differ. For the separated-block class $\mathcal{F}_\alpha$, the privacy term is $\big(d/(\rho n^2)\big)^{\alpha/(\alpha+1)}$, whereas the row-tail and pointwise-decay classes have the smaller benchmark $\mathcal{Q}_{\alpha,d}\big(d/(\rho n^2)\big)$. Privacy therefore distinguishes structures whose leading non-private operator rates coincide. The gain for $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$ comes from stronger local constraints that permit less costly distance-specific releases.

Under normalized squared Frobenius loss, both the non-private and private terms separate $\mathcal{F}_\alpha$ from the two smaller classes. The former retains the operator-loss exponents, while the latter have the minimax rate

$$n^{-\frac{2\alpha+1}{2\alpha+2}} + \left(\frac{d}{\rho n^2}\right)^{\frac{2\alpha+1}{2\alpha+3}},$$

before the finite-dimensional minima. Frobenius loss averages entrywise errors, whereas operator loss is governed by the worst spectral direction. Consequently, for $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, the Frobenius privacy term is smaller in the structured high-dimensional regime when $0 < \alpha < 1/2$ and has the same algebraic order when $\alpha \geq 1/2$.

The estimator constructions mirror these geometric distinctions. For $\mathcal{F}_\alpha$, blockwise tridiagonal estimation balances truncation bias, sampling fluctuation, and Gaussian privacy noise under both losses. For $\mathcal{H}_\alpha$ and $\mathcal{G}_\alpha$, a center–outer dyadic decomposition supports class-specific releases. A small blockwise Frobenius radius enables greedy rank-one releases for pointwise decay under operator loss; local row–column ℓ^1 control leads to profile releases under operator loss and peeling releases under Frobenius loss for row-tail decay. For pointwise decay under Frobenius loss, the blockwise tridiagonal estimator already uses the entrywise structure.

Practical implication. The center–outer procedures are primarily theoretical constructions: they require searches over finite nets and reconciliation of feasible sets to attain the sharper rates for individual classes, whereas the blockwise tridiagonal estimator is the computationally practical procedure used in our experiments. For routine implementation, the blockwise tridiagonal estimator remains the natural default because it uses fixed local releases and avoids the candidate searches required by those procedures. The center–outer procedures are chiefly useful for establishing the sharper rates for individual classes.

2.4.2. *Finite-dimensional regimes and the cost of privacy.* Each minimax risk is the sum of a non-private statistical contribution and a privacy contribution. For example, over $\mathcal{F}_\alpha$ under operator loss, privacy dominates when $\rho \lesssim dn^{-2\alpha/(2\alpha+1)}$; sampling error dominates on the other side. The results are stated under ρ-zCDP. Standard conversion to (ϵ, δ)-DP introduces the corresponding $\log(1/\delta)$ dependence in the upper bounds.

Dimension also determines whether the structured rate or the unstructured endpoint is active. The unstructured statistical and privacy terms are d/n and $d^3/(\rho n^2)$. For $\mathcal{F}_\alpha$ under either loss, the statistical and privacy transitions occur at $d \asymp n^{1/(2\alpha+1)}$ and $d \asymp (\rho n^2)^{1/(2\alpha+3)}$.

The operator statistical threshold for $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$ is the same. Under Frobenius loss, their thresholds are $d \asymp n^{1/(2\alpha+2)}$ and $d \asymp (\rho n^2)^{1/(2\alpha+4)}$.

The operator privacy term over $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$ has a finer structure. With $u = d/(\rho n^2)$, the exact benchmark is defined in (6), and its explicit regimes are given in (7) and (8). The two terms are, respectively, the squared operator signal permitted at distance scale k and effective rank r, and the corresponding private information capacity. Balancing them over r gives $u^{1/2}k^{(1-2\alpha)/2}$, explaining the change at $\alpha = 1/2$. For $\alpha \geq 1/2$, the two regimes are those in (7). Here the optimizer is first dimension-limited and then satisfies $r \asymp k \asymp (\rho n^2/d)^{1/(2\alpha+3)}$. For $0 < \alpha < 1/2$, the three regimes are those in (8). In the middle branch, $k \asymp d$ while the effective rank decreases from order d to order one. In the last branch, $r \asymp 1$ and the optimal scale becomes interior. Thus the three branches correspond to both variables being dimension-limited, only the scale being dimension-limited, and the rank reaching its lower endpoint.

Unlike the residual non-private term $(\log d)/n$, every privacy term depends polynomially on d. In particular, when $\rho \asymp 1$, the lower bounds imply that consistency requires $d = o(n^2)$.

2.4.3. *Sharpness, adaptation, and extensions.* For known smoothness, the upper and lower bounds match up to constants over $\mathcal{F}_\alpha$ under both losses and over $\mathcal{H}_\alpha$ under Frobenius loss. Over $\mathcal{G}_\alpha$ under both losses and over $\mathcal{H}_\alpha$ under operator loss, the bounds match up to fixed polylogarithmic factors. The condition $\rho n^2/d \gtrsim (\log d)^{2(\alpha+1)}$ for the constant-factor $\mathcal{F}_\alpha$ comparison excludes a near-degenerate boundary at which the lower bound decreases only logarithmically.

The adaptive procedures incur no algebraic loss relative to their known-smoothness benchmarks. For $\mathcal{F}_\alpha$ under both losses and $\mathcal{H}_\alpha$ under Frobenius loss, they preserve the non-private terms and add logarithmic factors only to the privacy terms. For $\mathcal{G}_\alpha$ under both losses and $\mathcal{H}_\alpha$ under operator loss, they attain fixed-polylogarithmic oracle ratios over compact smoothness intervals. Whether these adaptation costs are intrinsic remains open.

The blockwise analysis also extends to covariance classes with exponential off-diagonal decay. If every block R_k separated by distance k satisfies $\|\Sigma_{R_k}\| \leq C_1 e^{-\gamma k}$, then choosing $k \asymp (\log n) \wedge \log(\rho n^2/d)$ gives squared operator risk:

$$\frac{\log n}{n} + \frac{d}{\rho n^2}\left(\log\left(\frac{\rho n^2}{d}\right) + \log d\right)^2.$$

The corresponding row-tail and pointwise-decay results follow from the class inclusions.

3. DP Estimators and Upper Bounds. In this section, we construct private estimators and establish their upper bounds with the smoothness parameter α known. The separated-block class $\mathcal{F}_\alpha$ uses a single blockwise-tridiagonal core, whereas $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$ use its dyadic square-block refinement with class-specific profile and greedy local procedures. Under Frobenius loss, $\mathcal{G}_\alpha$ uses a nonlinear peeling estimator, $\mathcal{H}_\alpha$ uses the blockwise tridiagonal estimator, and $\mathcal{F}_\alpha$ follows from the Schatten consequence.

3.1. *Basic Privacy Mechanisms.* We use the Gaussian mechanism and the composition and post-processing properties of zCDP summarized below [10, 26].

LEMMA 3.1 (Gaussian Mechanism). *Let $f : \mathcal{D} \to \mathbb{R}^p$ be an algorithm such that*

$$\sup_{D,D' \text{ adjacent}} \left\|f(D) - f(D')\right\|_2 \leq \Delta.$$

Then, the algorithm $M_f(D) = f(D) + \sigma w$ is ρ-zCDP, where $w \sim N(0, I_p)$ and $\sigma^2 = \frac{\Delta^2}{2\rho}$.

LEMMA 3.2 (Composition). *If $M : \mathcal{D} \to \mathcal{Y}$ is ρ-zCDP and $f : \mathcal{Y} \to \mathcal{Z}$ is an arbitrary algorithm, then $f \circ M$ is also ρ-zCDP. If $M_1, \ldots, M_T$ are $\rho_1, \ldots, \rho_T$-zCDP and M is a function of $M_1, \ldots, M_T$, then M is ρ-zCDP for $\rho = \sum_t \rho_t$.*

3.2. *Dyadic block structure.* We first introduce the block notation used by the estimators below and their adaptive counterparts. Without loss of generality, for each construction we append deterministic zero coordinates until $d = 2^J k_0$, where k_0 is its initial block size (with $k_0 = k$ for a fixed-bandwidth estimator), since this preserves the model and privacy and changes the dimension by less than a factor of two. For an integer block size $k \geq 1$, let $N_k = d/k$ and define $I_{k,\ell} = [1 + (\ell - 1)k, \ell k]$, $\ell \in [N_k]$. For $r \geq 0$ and $\ell + r \leq N_k$, put $B_{k;\ell,r} = I_{k,\ell} \times I_{k,\ell+r}$. Only $r = 0, 1, 2, 3$ will be used, and lower-triangular blocks are represented by transposes. Blocks with invalid boundary indices are always omitted. The blockwise tridiagonal region is

$$\mathcal{T}_k = \bigsqcup_{\ell=1}^{N_k} B_{k;\ell,0} \sqcup \bigsqcup_{\ell=1}^{N_k-1} \left(B_{k;\ell,1} \sqcup B_{k;\ell,1}^\top \right). \tag{12}$$

The regions $\mathcal{T}_k$ are nested along a doubling sequence. For $k < d$ and $1 \leq \ell < N_{2k}$, define the upper L-shaped increment $\Gamma_{k,\ell} = B_{2k;\ell,1} \setminus B_{k;2\ell,1}$. The corresponding lower increment is its transpose, and hence

$$\mathcal{T}_{2k} = \mathcal{T}_k \sqcup \bigsqcup_{\ell=1}^{N_{2k}-1} \left(\Gamma_{k,\ell} \sqcup \Gamma_{k,\ell}^\top \right). \tag{13}$$

For the dyadic construction initialized at k_0, write $k_m = 2^m k_0$, $N_m = N_{k_m}$, and abbreviate $\Gamma_{m,\ell} = \Gamma_{k_{m-1},\ell}, m \geq 1$. Once $k_m \geq d/2$, $\mathcal{T}_{k_m} = [d]^2$ and all subsequent increments are empty. The upper-triangular part of the initial region $\mathcal{T}_{k_0}$ consists of $B_{k_0;l,r}$ for $l \in [N_{k_0}]$ and $r \in \{0, 1\}$ satisfying $l + r \leq N_{k_0}$. Each increment then decomposes as $\Gamma_{k,\ell} = B_{k;2\ell-1,2} \sqcup B_{k;2\ell-1,3} \sqcup B_{k;2\ell,2}$. For each $k \in \{k_0, 2k_0, \ldots, d/4\}$, let $\mathcal{B}_k$ be the collection of the three blocks on the right-hand side as $1 \leq \ell < N_{2k}$, which partitions the upper-triangular part of $\mathcal{T}_{2k} \setminus \mathcal{T}_k$. Iterating (13) therefore makes the initial blocks $B_{k_0;l,r}$ together with the families $\mathcal{B}_k$ a disjoint cover of the upper-triangular entries. For budget allocation, write $J(k_0) = 1 \vee (\log_2(d/k_0) - 1)$ for the number of active bands.

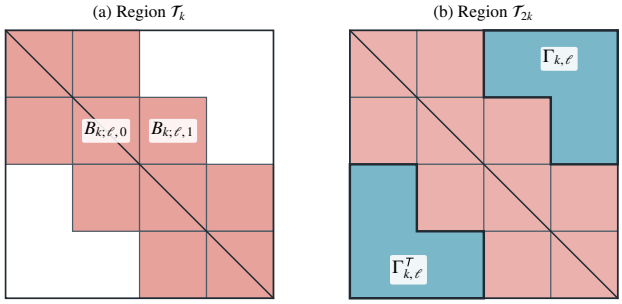

FIG 1. *Dyadic block structure at bandwidths k and $2k$.*

3.3. *Blockwise Tridiagonal Estimator.* We first introduce the simplest blockwise tridiagonal estimator. Algorithm 1 gives a private estimate of a single covariance block. Algorithm 2 applies this primitive to form a blockwise tridiagonal structure. The block size parameter k will be chosen later to balance the bias–variance–privacy trade-off. For the results below, we use $R = CK\sqrt{k}$, where C is a sufficiently large numerical constant. By the Gaussian mechanism and composition, Algorithm 1 is ρ_B-zCDP and Algorithm 2 is ρ-zCDP.

Algorithm 1 DP Covariance Block: $\mathsf{DPCovBlock}(X, B; \rho_B, R)$

Input: X, square block $B = I \times J$, privacy budget ρ_B, clipping radius R.
Let $\widetilde{X}_{i,I} = \chi_R(X_{i,I})$ and $\widetilde{X}_{i,J} = \chi_R(X_{i,J})$.
Let $\tilde{\Sigma}_B = \frac{1}{n}\sum_{i\in[n]} \widetilde{X}_{i,I}\widetilde{X}_{i,J}^\top$,
return $\hat{\Sigma}_B^{\mathsf{DP}} = \tilde{\Sigma}_B + \sigma_B M_B$, where $M_B = \mathrm{GUE}(d)_B$ and $\sigma_B^2 = 18R^4/(\rho_B n^2)$.

Algorithm 2 DP Blockwise Tridiagonal Estimator: $\mathsf{DPCovTri}(X; k, \rho, R)$

Input: X, block size k, privacy parameter ρ, truncation radius R.
Output: ρ-zCDP estimator $\hat{\Sigma}^{\mathsf{DP}}$.
Take $\rho_0 = \rho/(2N_k)$.
Compute $\hat{\Sigma}^{\mathsf{DP}}_{B_{k;l,r}} = \mathsf{DPCovBlock}\left(X, B_{k;l,r}; \rho_0, R\right)$ for all $r \in \{0,1\}$ and $l + r \le N_k$.
Fill the lower triangular part by symmetry, leaving other entries to zero.
return $\hat{\Sigma}^{\mathsf{DP}}$.

Compared with the tapering estimator of Cai, Zhang and Zhou [21], the proposed blockwise tridiagonal construction is simpler and better suited for the private setting, as it streamlines both the calibration of privacy noise and the subsequent statistical analysis. Moreover, because the estimator focuses only on a limited subset of covariance entries, the amount of noise required to guarantee privacy is substantially reduced, enabling optimal estimation accuracy under privacy constraints. We next derive upper bounds for the blockwise tridiagonal estimator under the operator and Frobenius norms.

THEOREM 3.3. *Let $\Sigma \in \mathcal{F}_\alpha$ for some $\alpha > 0$. Consider the blockwise tridiagonal estimator $\hat{\Sigma}^{\mathsf{DP}}$ given in Algorithm 2 with block size $k \le cn$ for some small constant $c > 0$. Then, we have*

$$\mathbb{E}\left\|\hat{\Sigma}^{\mathsf{DP}} - \Sigma\right\|^2 \lesssim \frac{k + \log d}{n} + \frac{dk(k + \log d)}{\rho n^2} + k^{-2\alpha}. \tag{14}$$

Theorem 3.3 characterizes the estimation error under operator norm with arbitrary block size k. Optimizing over k yields the following corollary.

COROLLARY 3.4. *Under the same setting as in Theorem 3.3, suppose that $\rho n^2/d \gtrsim (\log d)^{2(\alpha+1)}$ and $d \gtrsim n^{1/(2\alpha+1)} \wedge (\rho n^2)^{1/(2\alpha+3)}$. Take*

$$k \asymp \min\left(n^{1/(2\alpha+1)}, (\rho n^2/d)^{1/(2\alpha+2)}\right) \vee \log d$$

and let $\hat{\Sigma}_{\mathcal{F},\mathrm{op}}$ be the resulting blockwise tridiagonal estimator. Then

$$\mathbb{E}\left\|\hat{\Sigma}_{\mathcal{F},\mathrm{op}} - \Sigma\right\|^2 \lesssim n^{-\frac{2\alpha}{2\alpha+1}} + \left(\frac{d}{\rho n^2}\right)^{\frac{\alpha}{\alpha+1}}. \tag{15}$$

Under Frobenius loss, the pointwise-decay condition defining $\mathcal{H}_\alpha$ gives the sharper entry-wise truncation bias required by the same estimator.

THEOREM 3.5. *Let $\Sigma \in \mathcal{H}_\alpha$ for some $\alpha > 0$. Consider the blockwise tridiagonal estimator $\hat{\Sigma}^{\mathsf{DP}}$ given in Algorithm 2 with block size $k \le cn$ for some small constant $c > 0$. We have*

$$\frac{1}{d}\mathbb{E}\left\|\hat{\Sigma}^{\mathsf{DP}} - \Sigma\right\|_F^2 \lesssim \frac{k}{n} + \frac{dk^2}{\rho n^2} + k^{-(2\alpha+1)}. \tag{16}$$

COROLLARY 3.6. *Under the same setting as in Theorem 3.5, suppose that $d \gtrsim n^{1/(2\alpha+2)} \wedge (\rho n^2)^{1/(2\alpha+4)}$. Take*

$$k = \min\left(n^{1/(2\alpha+2)}, (\rho n^2/d)^{1/(2\alpha+3)}\right)$$

and let $\hat{\Sigma}_{\mathcal{H},\mathrm{F}}$ be the resulting blockwise tridiagonal estimator. Then

$$\frac{1}{d}\mathbb{E}\left\|\hat{\Sigma}_{\mathcal{H},\mathrm{F}} - \Sigma\right\|_F^2 \lesssim n^{-\frac{2\alpha+1}{2\alpha+2}} + \left(\frac{d}{\rho n^2}\right)^{\frac{2\alpha+1}{2\alpha+3}}. \tag{17}$$

The block size k balances truncation bias, sampling variance, and privacy noise. Under operator loss, balancing the bias with the latter two terms gives $k \asymp n^{1/(2\alpha+1)}$ and $k \asymp (\rho n^2/d)^{1/(2\alpha+2)}$, respectively, yielding Corollary 3.4. The Frobenius choice follows analogously from (16) with loss-specific exponents.

3.4. *Center–Outer Dyadic Estimator.* Now we turn to the two smaller pointwise-decay class $\mathcal{H}_\alpha$ and row-tail class $\mathcal{G}_\alpha$. If there were no privacy requirement, the blockwise tridiagonal estimator, with a properly tuned choice of k, can achieve the minimax optimal rate thanks to the class inclusions and the identical non-private minimax rate, see Table 1 on page 3. However, under privacy, a direct application of the same construction cannot attain the sharper privacy rates for $\mathcal{H}_\alpha$ and $\mathcal{G}_\alpha$ under operator loss, because releasing each dense block incurs too much privacy noise to exploit their local geometry. Likewise, naively privatizing classical estimators such as truncation or tapering does not resolve this issue.

The key difference lies in the geometry of the two smaller bandable covariance classes. There are intermediate regions where we can exploit class geometry to pay less privacy cost than using the dense block estimator in Algorithm 1. The insight is that, if the class geometry admits low-complexity approximation, then we can often construct private procedures whose cost of privacy scales with the intrinsic complexity instead of parameter dimension.

To this end, we split the matrix into three parts: a dense center $\mathcal{T}_{k_0}$, dyadic outer shells between k_0 and a terminal scale k_+, and a tail beyond k_+. The center is estimated by the blockwise-tridiagonal procedure, each outer shell is divided into the square blocks in $\mathcal{B}_k$, and the tail is set to zero. The procedure is given in Algorithm 3. In the following subsections, we will give the different choices of BlockRelease for both classes $\mathcal{H}, \mathcal{G}$ under different losses, and provide corresponding tuning parameters k_0, k_+. The resulting architecture and the three local release modules are illustrated in Figure 2 on page 14. The prescribed budget allocation, zCDP composition, and post-processing make every estimator constructed from this template ρ-zCDP. The resulting estimators are denoted by $\widehat{\Sigma}_{\mathcal{H}}$, $\widehat{\Sigma}_{\mathcal{G}}$ and $\widehat{\Sigma}_{\mathcal{G},\mathrm{F}}$ respectively.

THEOREM 3.7 (Operator-norm upper bounds). *Fix $\alpha, K, M_0, M > 0$. For each $\mathcal{J} \in \{\mathcal{G}, \mathcal{H}\}$ and all sufficiently large n, the corresponding $\widehat{\Sigma}_{\mathcal{J}}$ defined below is ρ-zCDP and*

$$\sup_{P \in \mathcal{P}_\alpha^{\mathcal{J}}} \mathbb{E}_P \left\|\widehat{\Sigma}_{\mathcal{J}} - \Sigma(P)\right\|^2 \lesssim_{\log} n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \mathcal{Q}_{\alpha,d}\left(\frac{d}{\rho n^2}\right). \tag{18}$$

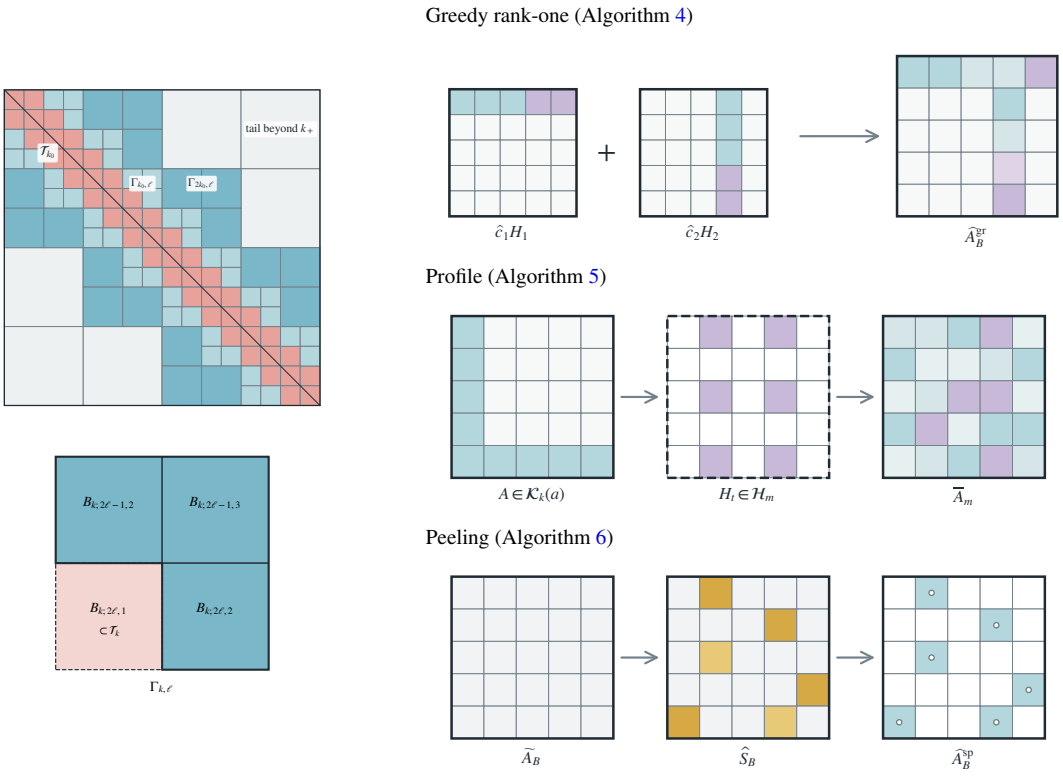

FIG 2. *Center–outer dyadic architecture and local release modules.*

Algorithm 3 DP Center–Outer Dyadic Estimator

Input: X, privacy budget ρ, center size k_0, terminal size $k_+ \in \{k_0, 2k_0, \dots, d/2\}$, and local procedure BlockRelease.
Initialize the center by $\widehat{\Sigma} = \mathsf{DPCovTri}\big(X; k_0, \rho/4, K\sqrt{C_1(k_0 + \log n)}\big)$.
for each $k \in \{k_0, 2k_0, \dots\}$ satisfying $k < k_+$ **do**
  Let $\rho_k = \rho/(4J(k_0)N_{2k})$.
  Set $\widehat{\Sigma}[B] = \mathsf{BlockRelease}(X, B; k, \rho_k)$ for each $B \in \mathcal{B}_k$.
end for
Fill lower-triangular of $\widehat{\Sigma}$ by symmetry.
return $\Pi_{M_0}(\widehat{\Sigma})$.

THEOREM 3.8 (Row-tail Frobenius upper bound). *Fix $\alpha, K, M_0, M > 0$. For all sufficiently large n, $\widehat{\Sigma}_{\mathcal{G},\mathrm{F}}$ defined below is ρ-zCDP and*

$$\sup_{P \in \mathcal{P}_\alpha^{\mathcal{G}}} \frac{1}{d}\mathbb{E}_P \left\| \widehat{\Sigma}_{\mathcal{G},\mathrm{F}} - \Sigma(P) \right\|_F^2 \lesssim_{\log} n^{-\frac{2\alpha+1}{2\alpha+2}} \wedge \frac{d}{n} + \left(\frac{d}{\rho n^2}\right)^{\frac{2\alpha+1}{2\alpha+3}} \wedge \frac{d^3}{\rho n^2}. \tag{19}$$

3.4.1. *Pointwise-decay class under operator loss: greedy blocks.* Pointwise decay makes every outer block small in Frobenius norm. The greedy procedure exploits this by building a rank-one approximation: at each round it selects a direction aligned with the current residual and releases only the corresponding noisy coefficient. Using r rounds produces an approximation term of order a^2/r, but the privacy noise only grows linearly with r, so if r is small, the privacy noise is smaller than that of the dense block estimator.

For a $k \times k$ block A, define $\mathcal{E}_k(a) = \{A : \|A\|_F \le a,\ \|A\| \le M_0\}$. For $\Sigma \in \mathcal{H}_\alpha$, every scale-$k$ square of Σ belongs to $\mathcal{E}_k(a_{\mathcal{H},k})$, where $a_{\mathcal{H},k} = C_{\mathcal{H}} k^{-\alpha}$ and $C_{\mathcal{H}}$ depends only on

(M, M_0). Thus $\mathcal{E}_k(a)$ records the two facts used by the local analysis: the Frobenius radius a controls low-rank approximation, and M_0 supplies a uniform operator bound. Let $\mathcal{W}_k = \{\xi uv^\top : u, v \text{ lie in a fixed } 1/4\text{-net of the unit sphere in } \mathbb{R}^k,\ \xi \in \{-1, 1\}\}$ be the set of rank-one candidates.

Algorithm 4 Greedy Rank-One Block Estimator

Input: X, square block $B = I \times J$ of side length k, radius a, and budget ρ_B.
Choose $r = \arg\min_{1 \le r \le k}(a^2/r + k^2 r/(\rho_B n^2))$.
Set $\varepsilon = \sqrt{\rho_B / r}$ and $L = CK^2 \log n$.
Set $B_0 = 0$.
for $t = 1, \ldots, r$ **do**
For $H = \xi uv^\top \in \mathcal{W}_k$, set

$$s_t(X, H) = \frac{1}{n} \sum_{i=1}^{n} \chi_L \left(\xi (u^\top X_{i,I})(v^\top X_{i,J}) \right) - \langle B_{t-1}, H \rangle_F.$$

Draw H_t by the exponential mechanism with privacy parameter ε and score $s_t(X, H)$.
Draw $G_t \sim N(0, 4L^2 r/(\rho_B n^2))$, set $\widehat{c}_t = s_t(X, H_t) + G_t$, and update $B_t = B_{t-1} + \widehat{c}_t H_t$.
end for
Set $\overline{B}_r = r^{-1} \sum_{t=1}^{r} B_{t-1}$, and project $\overline{B}_r$ onto the operator-norm ball of radius $2M_0$ to obtain $\widehat{A}_B^{\mathrm{gr}}$.
return $\widehat{A}_B^{\mathrm{gr}}$.

Up to logarithmic factors, the blockwise squared error is $k/n + a_{\mathcal{H},k}^2/r + k^2 r/(\rho_k n^2)$ for a fixed rank r, so we minimize the last two terms for an optimal r_k. Balancing the errors yield the choice of k_0 and dyadic scale k_+ satisfying

$$k_0 \asymp \frac{d}{2} \wedge n^{1/(2\alpha+1)} \wedge \left(\frac{d}{\rho n^2} \right)^{-1/(2\alpha+3)}, \qquad k_+ \asymp \frac{d}{2} \wedge n^{1/(2\alpha+1)} \wedge \left(\frac{d}{\rho n^2} \right)^{-1/(2\alpha+1)}. \tag{20}$$

3.4.2. *Row-tail class under operator loss: profile blocks.* The row–column ℓ^1 constraint controls the total mass in every row and column but allows that mass to be distributed across many support sizes. Thus a low rank approximation is no longer suitable. Instead, we use a dual perspective to find the approximation with low complexity. Define the local ℓ^1 feasible set fo $k \times k$ block $A = (A_{uv})$

$$\mathcal{K}_k(a) = \left\{ A : \max_u \sum_v |A_{uv}| \vee \max_v \sum_u |A_{uv}| \le a \right\}.$$

For $\Sigma \in \mathcal{G}_\alpha$, every scale-$k$ square of Σ belongs to $\mathcal{K}_k(a_{\mathcal{G},k})$, where $a_{\mathcal{G},k} = C_{\mathcal{G}} k^{-\alpha}$ and $C_{\mathcal{G}}$ depends only on (M, M_0). Recall that $\|A\| = \sup_{H = uv^\top} \langle A, H \rangle_F$, where u, v range over all unit vectors. The geometry of $\mathcal{K}_k(a)$ allows us to work with the following low-complexity subset instead of all such H's. Define $\mathcal{V}_s = \left\{ u \in \mathbb{R}^k : \|u\|_0 \le s, \|u\|_2 \le 1, \|u\|_\infty \le \sqrt{2/s} \right\}$, and consider

$$\mathcal{H}_m = \left\{ H = \xi uv^\top : u \in \mathcal{V}_s, \quad v \in \mathcal{V}_r, \quad \max(s, r) = m, \quad \xi \in \{-1, 1\} \right\}.$$

Define $d_m(A, B) = \sup_{H \in \mathcal{H}_m} |\langle A - B, H \rangle_F|$ as the approximate error measure over $\mathcal{H}_m$. Then, it suffices to find targets A_m that are close to the non-private empirical covariance

block, and aggregate those targets for the final estimate. We summarize the main procedure as follows, but defer technical details to the supplementary material.

Algorithm 5 Profile Block Estimator (informal)

Input: X, square block $B = I \times J$ of side length k, radius a, and budget ρ_B.
for each dyadic profile m **do**
  for $t = 1, \dots, T_m$ **do**
    Compute the candidate block A_t using $(H_s)_{s<t}$.
    Privately select probe $H_t \in \mathcal{H}_m$ with the largest remaining discrepancy.
  end for
  Compute final candidate block $\overline{A}_m$.
end for
Reconcile $(\overline{A}_m)_m$ into a point estimate $\widehat{A}_B^{\mathrm{prof}}$.
return $\widehat{A}_B^{\mathrm{prof}}$.

We can show that the resulting blockwise squared error is $k/n + ak/(n\sqrt{\rho_B})$ up to logarithmic factors. With $a = a_{\mathcal{G},k}$ and $\rho_B = \rho_k$, balancing the center, outer-block, and tail errors gives the same scales k_0 and k_+ in (20). In the supplement, the profile centers are reconciled through a common feasible set; the true block belongs to it with high probability, and its small diameter yields the pointwise deviation bound. We also remark that this estimator for $\mathcal{G}_\alpha$ also applies to $\mathcal{H}_\alpha \subset \mathcal{G}_\alpha$, giving the same upper rates, but we keep the simpler construction under $\mathcal{H}_\alpha$.

3.4.3. *Row-tail class under Frobenius loss: peeling blocks.* Under Frobenius loss, the row–column ℓ^1 constraint yields a direct sparse approximation. Indeed, $A \in \mathcal{K}_k(a)$ implies $\sum_{u,v} |A_{uv}| \le ka$, so retaining the kr largest entries gives a support $S \subseteq [k]^2$ with $|S| \le kr$ and $\|A_{S^c}\|_F^2 \le ka^2/r$. Because the locations of these entries are unknown, the peeling procedure selects them privately and releases only their noisy values. When r is small, this avoids the privacy cost of releasing all k^2 entries.

Algorithm 6 Peeling Frobenius Block Estimator

Input: X, square block $B = I \times J$ of side length k, radius a, and budget ρ_B.
Choose $r = \arg\min_{1 \le r \le k} (a^2/r + r/n + kr^2/(\rho_B n^2))$.
Set $\widetilde{A}_{B,uv} = n^{-1} \sum_{i=1}^n \chi_L(X_{i,I(u)} X_{i,J(v)})$ for $L = CK^2 \log n$.
Starting with all k^2 coordinates, perform kr successive exponential-mechanism selections with score $\left|\widetilde{A}_{B,uv}\right|$ and privacy parameter $\sqrt{\rho_B/(kr)}$; denote the selected support by $\widehat{S}_B$.
For $(u,v) \in \widehat{S}_B$, set $\widehat{A}^{\mathrm{sp}}_{B,r}(u,v) = \widetilde{A}_{B,uv} + G_{B,uv}$, where $G_{B,uv} \overset{\mathrm{iid}}{\sim} N(0, 4L^2 kr/(\rho_B n^2))$, and set the remaining coordinates to zero.
return $\widehat{A}^{\mathrm{sp}}_{B,r}$.

Up to logarithmic factors, the blockwise squared error is $k(a^2/r + r/n + kr^2/(\rho_B n^2))$ for a fixed sparsity index r, so we minimize the three terms for an optimal r_k. The three terms balance sparse approximation, sampling on the selected entries, and privacy noise from selection and release. We set $k_+ = d/2$ and balance these errors to yield the choice

$$k_0 \asymp \frac{d}{2} \wedge n^{1/(2\alpha+2)} \wedge \left(\frac{d}{\rho n^2}\right)^{-1/(2\alpha+3)}.$$

4. Minimax Lower Bounds. In this section, we present our main minimax lower bound results for covariance estimation under DP. As one of our main technical contributions, we first introduce a novel DP van Trees inequality, which is of independent interest for other private estimation problems beyond this paper. Then, we use it to derive minimax lower bounds for covariance estimation under zCDP.

4.1. *A DP van Trees Inequality.* The key technique for proving the lower bounds is Theorem 4.1, a novel van Trees inequality under DP. It extends the classical van Trees inequality to the differentially private setting by incorporating a Fisher information bound under zCDP. The proof of Theorem 4.1 is deferred to the supplementary material.

THEOREM 4.1 (DP van Trees Inequality). *Let $\hat{\theta}$ be a ρ-zCDP estimator computed from n i.i.d. samples from P_θ, $\theta \in \mathbb{R}^p$. Let $I_x(\theta)$ be the Fisher information matrix of a single sample from P_θ. Let π be a prior distribution on $\Theta \subseteq \mathbb{R}^p$ with Fisher information matrix J_π. Under common regularity conditions, we have*

(21)
$$\mathbb{E}_\pi \mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|_2^2 \geq \frac{p^2}{I + \operatorname{Tr} J_\pi}, \quad I = \left\{ C_\rho \rho n^2 \int_\Theta \|I_x(\theta)\| \, \mathrm{d}\pi(\theta) \right\} \wedge \left\{ n \int_\Theta \operatorname{Tr} I_x(\theta) \, \mathrm{d}\pi(\theta) \right\},$$

where $C_\rho = (e^{2\rho} - 1)/\rho$.

Let us make some remarks on Theorem 4.1 from the following perspectives.

4.1.1. *Cost of Privacy.* We note that there are three terms in the denominator of the lower bound in (21): the terms $n \int_\Theta \operatorname{Tr} I_x(\theta) \, \mathrm{d}\pi(\theta)$ and $\operatorname{Tr} J_\pi$ are standard in the classical van Trees inequality, while the additional term $\rho n^2 \int_\Theta \|I_x(\theta)\| \, \mathrm{d}\pi(\theta)$ captures the restriction of the information due to privacy. We highlight that this term involves the operator norm $\|I_x(\theta)\|$ of the Fisher information matrix instead of the trace $\operatorname{Tr} I_x(\theta)$, which is generally smaller than $\operatorname{Tr} I_x(\theta)$ by a factor of the parameter dimension p. Hence, the cost of privacy often has a stronger dependence on the dimension of the parameter compared to the non-private case. This difference is commonly observed in private estimation problems [16, 17].

4.1.2. *Comparison with Other Lower Bound Techniques.* There are several existing techniques for deriving minimax lower bounds under DP. One approach is to use private versions of classical information-theoretic inequalities such as Fano's inequality [1, 16, 24, 25]. However, these techniques often require delicate constructions and fail to yield tight lower bounds in many problems.

Another popular approach is the fingerprinting method [6] or the recently developed score attack [17]. However, these methods typically require that the prior distribution of the parameter has independent coordinates, which limits their applicability in complex problems where such independence may not hold, such as covariance estimation [30, 32, 34]. Notably, the use of van Trees inequality for deriving minimax lower bounds under DP was first explored in Cai, Chakraborty and Vuursteen [11], but it was still interpreted through the lens of score attack. Moreover, these applications typically require cumbersome control of remainder terms and often suffer from logarithmic-factor gaps.

4.1.3. *Applications.* Because it is computationally convenient, Theorem 4.1 can be used to derive minimax lower bounds under DP for various estimation problems, yielding tight lower bounds in many cases. To use Theorem 4.1, it suffices to (i) compute and bound the Fisher information matrix with respect to the parameter of interest, and (ii) construct an

appropriate prior distribution over the parameter space. For more complex problems, the prior distribution may need to be carefully designed to balance I and J_π, as we will see in the proof of our main results in Section 4.

Let us further illustrate the use of Theorem 4.1 with simple examples of mean estimation and linear regression. For mean estimation, we take $x \sim N(\mu, I_d)$ and thus $I_x(\mu) = I_d$. Taking a prior distribution of μ with independent coordinates satisfying regularity conditions, we obtain the lower bound $\frac{d}{n} \vee \frac{d^2}{\rho n^2}$. For linear regression, we take $y = \langle \beta, x \rangle + \varepsilon$ with $x \sim N(0, I_d)$ and $\varepsilon \sim N(0,1)$ independent, and we have $I_{(x,y)}(\beta) = I_d$. Taking a similar prior distribution on β, we obtain the same lower bound as in mean estimation. These examples easily recover the known minimax lower bounds for mean estimation and linear regression [16, 17]. Therefore, we believe that Theorem 4.1 can serve as a general-purpose tool for deriving minimax lower bounds under DP.

4.2. *Class-Specific Minimax Lower Bounds.* Using the DP van Trees inequality, we can provide matching minimax lower bounds for the three classes and two loss norms. Recall that $\mathcal{M}_\rho$ denotes the set of all ρ-zCDP estimators based on n samples. We begin with the separated-block class and then turning to the two smaller classes.

4.2.1. *Separated-block class.* The separated-block lower bound holds simultaneously for all Schatten losses between the Frobenius and operator norms. Taking $q = \infty$ gives the operator-norm lower bound, whereas $q = 2$ gives the normalized Frobenius lower bound.

THEOREM 4.2 (Lower bound in $\mathcal{F}_\alpha$). *For every $q \in [2, \infty]$,*

$$\inf_{\hat{\Sigma} \in \mathcal{M}_\rho} \sup_{\Sigma \in \mathcal{F}_\alpha} d^{-\frac{2}{q}} \mathbb{E}_\Sigma \left\| \hat{\Sigma} - \Sigma \right\|_{\mathcal{S}_q}^2 \gtrsim n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \left(\frac{d}{\rho n^2} \right)^{\frac{\alpha}{\alpha+1}} \wedge \frac{d^3}{\rho n^2}. \tag{22}$$

4.2.2. *Pointwise-decay and row-tail classes.* The stronger local structure of $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$ changes the privacy rate under operator loss. Thanks to Proposition 2.4, it suffices to show lower bounds for the class $\mathcal{H}_\alpha$. Thus the two smaller classes share a common hard submodel for the lower-bound argument, even though their upper bounds require different private local procedures. We have the following two theorems.

THEOREM 4.3 (Operator lower under $\mathcal{G}_\alpha, \mathcal{H}_\alpha$). *For $\mathcal{J} \in \{\mathcal{G}, \mathcal{H}\}$, we have*

$$\inf_{\hat{\Sigma} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha^{\mathcal{J}}} \mathbb{E}_P \left\| \hat{\Sigma} - \Sigma(P) \right\|^2 \gtrsim n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \mathcal{Q}_{\alpha,d} \left(\frac{d}{\rho n^2} \right). \tag{23}$$

THEOREM 4.4 (Frobenius lower bound under $\mathcal{G}_\alpha, \mathcal{H}_\alpha$). *For $\mathcal{J} \in \{\mathcal{G}, \mathcal{H}\}$, we have*

$$\inf_{\hat{\Sigma} \in \mathcal{M}_\rho} \sup_{P \in \mathcal{P}_\alpha^{\mathcal{J}}} \frac{1}{d} \mathbb{E}_P \left\| \hat{\Sigma} - \Sigma(P) \right\|_F^2 \gtrsim n^{-\frac{2\alpha+1}{2(\alpha+1)}} \wedge \frac{d}{n} + \left(\frac{d}{\rho n^2} \right)^{\frac{2\alpha+1}{2\alpha+3}} \wedge \frac{d^3}{\rho n^2}. \tag{24}$$

4.2.3. *Proof Strategy.* All three constructions use Theorem 4.1 with $x \sim N(0, \Sigma)$. We regard the covariance parameter as an element of the Euclidean space of symmetric matrices, equipped with the Frobenius inner product, and represent its Fisher information in any orthonormal basis for this inner product. It can be computed that the trace and operator norm of the Fisher information matrix are given by

$$\operatorname{Tr} I_x(\Sigma) = \frac{1}{4} \left[\operatorname{Tr}(\Sigma^{-2}) + \left(\operatorname{Tr} \Sigma^{-1} \right)^2 \right], \quad \|I_x(\Sigma)\| = \frac{1}{2} \left\| \Sigma^{-1} \right\|^2.$$

For $\mathcal{F}_\alpha$, we need to control the operator norm of off-diagonal blocks. To this end, we need to carefully design a prior distribution over matrices with bounded operator norm and small Fisher information trace. This is accomplished in Lemma 4.5 below.

LEMMA 4.5. *There is a distribution Υ_d over matrices $X \in \mathbb{R}^{d\times d}$ with continuously differentiable density $q(X)$ supported on $\|X\| \le 1$ such that the trace of the Fisher information matrix J_{Υ_d} of the distribution satisfies*

$$\operatorname{Tr} J_{\Upsilon_d} = \sum_{i,j=1}^{d} \mathbb{E}_{\Upsilon_d}\left[\frac{\partial}{\partial x_{ij}} \log q(X)\right]^2 \le Cd^3, \tag{25}$$

where C is an absolute constant.

The proof of Lemma 4.5 is done by constructing an explicit distribution based on truncated Gaussian random matrices. We note that if we only consider distributions with independent entries, the upper bound in (25) is at least of order d^4, which is too large for our purpose.

With Lemma 4.5, we introduce a tridiagonal block structure for the prior distribution of Σ, setting diagonal blocks as I_k and the off-diagonal blocks as $ck^{-\alpha}W_l$ for $1 \le l \le N_k$, where W_l are i.i.d. from the distribution Υ_k. Similar calculations then yield the desired lower bound in Theorem 4.2, where the balancing of the terms also corresponds to the optimal block size in Corollary 3.4.

For the operator loss over $\mathcal{H}_\alpha$, we use rank-r perturbations within blocks of width k, using the same idea of low-rank approximation in Subsection 3.4.1. For every $1 \le r \le k \le \lfloor d/4 \rfloor$, the resulting Gaussian submodel gives the privacy lower bound $\min\left(k^{-2\alpha}/r, dkr/(\rho n^2)\right)$. Optimizing over k and r yields $\mathcal{Q}_{\alpha,d}(d/(\rho n^2))$ in Theorem 4.3. Finally, the Frobenius loss over $\mathcal{H}_\alpha$ is simpler, we take $\Sigma_{ij} = ck^{-(\alpha+1)}u_{ij}$ for $1 \le |i-j| \le k$, where the u_{ij} are independent with density $\cos^2(\pi t/2)$ on $[-1,1]$.

5. Adaptive Estimators. In this section, we construct adaptive estimators for unknown decay parameter α. In parallel to the non-adaptive estimators for different bandable covariance classes, we use two adaptive frameworks. The blockwise tridiagonal estimators that threshold dyadic increments handle the separated-block class and the pointwise-decay class under Frobenius loss. The center–outer estimators combine a public radius grid with the greedy, profile, and peeling block estimators from Section 3 for the pointwise-decay operator mode and the row-tail operator and Frobenius modes.

5.1. *Adaptive Blockwise Tridiagonal Estimators.* We first adapt the blockwise tridiagonal estimator in Algorithm 2. Fix sufficiently large constants C_0, C, L_1 and a sufficiently small constant c_0, and set

$$k_0 = d \wedge \lceil C_0 \log n \rceil, \qquad k_m = 2^m k_0,$$

$$M = 1 + \max\{m \ge 0 : k_m \le \min(c_0 n, d)\}, \quad N_m = \frac{d}{k_m} \quad (0 \le m < M).$$

These public scales do not depend on α. At each noncentral scale, the estimator releases the corresponding dyadic increments and retains only those exceeding a scale-dependent threshold.

Algorithm 7 DP Adaptive Blockwise Tridiagonal Estimator (Operator Norm)

Input: X and privacy parameter ρ.
Initialize $\hat{\Sigma} = \mathbf{0}_{d\times d}$ and set $\rho_0 = \rho/(2MN_0)$.
Compute $\hat{\Sigma}[B_{k_0;l,r}] = \mathsf{DPCovBlock}\left(X, B_{k_0;l,r}; \rho_0, \sqrt{Ck_0}\right)$ for all $l \in [N_0]$ and $r \in \{0,1\}$.
for $m = 1, \dots, M-1$ **do**
  Set $\rho_m = \rho/(MN_m)$ and $\tau_m^2 = L_1\left(\frac{k_m + \log d}{n} + \frac{k_m^2(k_m+\log d)}{\rho_m n^2} + \exp(-2k_m)\right)$.
  for $l = 1, \dots, N_m - 1$ **do**
    Compute $A = \mathsf{DPCovBlock}\left(X, B_{k_m;l,1}; \rho_m, \sqrt{Ckm}\right)[\Gamma_{m,l}]$.
    Set $\hat{\Sigma}[\Gamma_{m,l}] = A \cdot \mathbf{1}\{\|A\| > \tau_m\}$.
  end for
end for
Fill the lower triangular part by symmetry and **return** $\hat{\Sigma}^{\mathsf{Ada}} = \hat{\Sigma}$.

THEOREM 5.1 (Operator-norm adaptation over $\mathcal{F}_\alpha$). *Let $\Sigma \in \mathcal{F}_\alpha$ for some $\alpha > 0$. Suppose that $\rho n^2/d \gtrsim (\log n)^{2\alpha+3}$. The estimator $\hat{\Sigma}^{\mathsf{Ada}}$ in Algorithm 7 is ρ-zCDP and satisfies*

$$\mathbb{E}\left\|\hat{\Sigma}^{\mathsf{Ada}} - \Sigma\right\|^2 \lesssim n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \left(\frac{d\log n}{\rho n^2}\right)^{\frac{\alpha}{\alpha+1}} \wedge \frac{d^3 \log n}{\rho n^2}. \tag{26}$$

The same upper bound applies to the normalized Frobenius loss because

$$\frac{1}{d}\mathbb{E}\left\|\hat{\Sigma}^{\mathsf{Ada}} - \Sigma\right\|_F^2 \le \mathbb{E}\left\|\hat{\Sigma}^{\mathsf{Ada}} - \Sigma\right\|^2.$$

For Frobenius adaptation over $\mathcal{H}_\alpha$, replace the operator threshold in Algorithm 7 by

$$\hat{\Sigma}[\Gamma_{m,l}] = A \cdot \mathbf{1}\left\{\|A\|_F^2 > k_m \tau_m^2\right\}. \tag{27}$$

THEOREM 5.2 (Frobenius adaptation over $\mathcal{H}_\alpha$). *Let $\Sigma \in \mathcal{H}_\alpha$ for some $\alpha > 0$. Suppose that $\rho n^2/d \gtrsim (\log n)^{2\alpha+4}$. The Frobenius-thresholded estimator is ρ-zCDP and satisfies*

$$\frac{1}{d}\mathbb{E}\left\|\hat{\Sigma}^{\mathsf{Ada}} - \Sigma\right\|_F^2 \lesssim n^{-\frac{2\alpha+1}{2\alpha+2}} \wedge \frac{d}{n} + \left(\frac{d\log n}{\rho n^2}\right)^{\frac{2\alpha+1}{2\alpha+3}} \wedge \frac{d^3 \log n}{\rho n^2}. \tag{28}$$

For both thresholding procedures, the statistical term is attained without an adaptation loss, while the privacy term incurs a logarithmic factor.

5.2. *Adaptive Center–Outer Dyadic Estimators.* The center–outer estimators in Subsection 3.4 are not adaptive: their center size, terminal scale in the operator modes, class-specific local radius of order $k^{-\alpha}$, and the rank or sparsity indices in the greedy and peeling procedures are calibrated to the unknown smoothness α. Adaptation is therefore not a single bandwidth choice, because the preferred local release can switch among zero, structured, and dense branches across scales, while all radius candidates and outer blocks must be controlled under a common privacy budget. We fix the smallest center and traverse public dyadic scales; at each scale, we form a geometric radius grid, choose the branch by public criteria, and then select one radius jointly for all outer blocks by Lepski comparisons before assembling them with the dense center. For the selected mode, write $\mathcal{J} = \mathcal{H}$ in mode $\mathcal{H}_{\mathrm{op}}$ and $\mathcal{J} = \mathcal{G}$ in the two $\mathcal{G}$-modes.

Algorithm 8 DP Adaptive Center–Outer Dyadic Estimator

Input: X, privacy parameter ρ, interval $[\underline{\alpha}, \overline{\alpha}]$, and mode $\mathcal{G}_{\mathrm{op}}$, $\mathcal{H}_{\mathrm{op}}$, or $\mathcal{G}_{\mathrm{F}}$.
Set $J_{\mathrm{sq}} = \log_2 d - 1$ and initialize $\widehat{\Sigma}$ on $\mathcal{T}_1$ as in Algorithm 3, using budget $\rho/4$.
for each $k \in \{1, 2, 4, \ldots, d/4\}$ **do**
  if the mode is operator norm and $k > n/(\log n)^C$ **then** Break.
  end if
  Form the public radius grid $(a_{\mathcal{J},k,j})_{j=0}^{J_k}$ specified below.
  Set $\rho_{k,j} = \rho/\{4J_{\mathrm{sq}} N_{2k}(J_k + 1)\}$ for $0 \le j \le J_k$.
  for $j = 0, \ldots, J_k$ **do**
    Select a branch $t_{k,j}$ and deviation radius $D_{k,j}$ by the public tuning rules below.
    for $B \in \mathcal{B}_k$ **do**
      Run branch $t_{k,j}$ with radius $a_{\mathcal{J},k,j}$ and budget $\rho_{k,j}$ to obtain a point candidate.
      Radially clip the candidate to the loss-norm ball of radius $a_{\mathcal{J},k,0}$ in the operator modes or $\sqrt{k}a_{\mathcal{J},k,0}$ in the Frobenius mode; denote the result by $Z_{B,k,j}$.
    end for
  end for
  Set

$$\widehat{j}_k = \max\left\{ j : \max_{B \in \mathcal{B}_k} \left\| Z_{B,k,j} - Z_{B,k,i} \right\|_\star \le D_{k,j} + D_{k,i} \text{ for every } i < j \right\}.$$

  Set $\widehat{\Sigma}[B] = Z_{B,k,\widehat{j}_k}$ for each $B \in \mathcal{B}_k$.
end for
Fill the lower triangular part of $\widehat{\Sigma}$ by symmetry.
return $\Pi_{M_0}(\widehat{\Sigma})$.

At a dyadic scale k, set

$$J_k = \lceil (\overline{\alpha} - \underline{\alpha}) \log_2 k \rceil, \qquad a_{\mathcal{J},k,j} = C_{\mathcal{J}} k^{-\underline{\alpha}} 2^{-j}, \qquad 0 \le j \le J_k,$$

and abbreviate $a_j = a_{\mathcal{J},k,j}$. For every $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, the grid contains a radius between $C_{\mathcal{J}} k^{-\alpha}$ and $2C_{\mathcal{J}} k^{-\alpha}$, where $C_{\mathcal{J}}$ depends only on the class and model bounds. At each radius, choose publicly among zero, structured, and dense candidates. Their deterministic scores are

mode	$W_{k,j,0}$	$W_{k,j,\mathrm{str}}$	$W_{k,j,\mathrm{dense}}$
$\mathcal{H}_{\mathrm{op}}$	a_j^2	$\frac{k}{n} + \min_{1 \le r \le k} \left\{ \frac{a_j^2}{r} + \frac{k^2 r}{\rho_{k,j} n^2} \right\}$	$\frac{k}{n} + \frac{k^3}{\rho_{k,j} n^2}$
$\mathcal{G}_{\mathrm{op}}$	a_j^2	$\frac{k}{n} + \frac{a_j k}{n\sqrt{\rho_{k,j}}}$	$\frac{k}{n} + \frac{k^3}{\rho_{k,j} n^2}$
$\mathcal{G}_{\mathrm{F}}$	ka_j^2	$k \min_{1 \le r \le k} \left\{ \frac{a_j^2}{r} + \frac{r}{n} + \frac{kr^2}{\rho_{k,j} n^2} \right\}$	$\frac{k^2}{n} + \frac{k^4}{\rho_{k,j} n^2}$

Set

$$t_{k,j} \in \operatorname*{arg\,min}_{t \in \{0,\mathrm{str},\mathrm{dense}\}} W_{k,j,t}, \qquad D_{k,j}^2 = C(\log n)^C W_{k,j,t_{k,j}},$$

using a fixed tie-breaking rule. Thus $D_{k,j}$ is a public deviation radius for the point candidate $Z_{B,k,j}$. In modes $\mathcal{H}_{\mathrm{op}}$, $\mathcal{G}_{\mathrm{op}}$, and $\mathcal{G}_{\mathrm{F}}$, the structured branch is Algorithm 4, Algorithm 5, and Algorithm 6, respectively. In every mode, the dense branch is $\mathsf{DPCovBlock}(X, B; \rho_{k,j}, R_k)$ with $R_k^2 = CK^2(k + \log n)$.

Denote the outputs in modes $\mathcal{G}_{\mathrm{op}}$, $\mathcal{H}_{\mathrm{op}}$, and $\mathcal{G}_{\mathrm{F}}$ by $\widehat{\Sigma}_{\mathcal{G}}^{\mathsf{Ada}}$, $\widehat{\Sigma}_{\mathcal{H}}^{\mathsf{Ada}}$, and $\widehat{\Sigma}_{\mathcal{G},\mathrm{F}}^{\mathsf{Ada}}$, respectively. These adaptive estimators attain the minimax rates up to polylogarithmic factors.

THEOREM 5.3 (Operator-norm adaptation over $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$). *Fix $0 < \underline{\alpha} \le \overline{\alpha} < \infty$ and $K, M_0, M > 0$. For each $\mathcal{J} \in \{\mathcal{G}, \mathcal{H}\}$ and all sufficiently large n, the corresponding estimator $\hat{\Sigma}^{\mathsf{Ada}}_{\mathcal{J}}$ in Algorithm 8 is ρ-zCDP and, uniformly over $\alpha \in [\underline{\alpha}, \overline{\alpha}]$,*

$$\sup_{P \in \mathcal{P}^{\mathcal{J}}_\alpha} \mathbb{E}_P \left\| \hat{\Sigma}^{\mathsf{Ada}}_{\mathcal{J}} - \Sigma(P) \right\|^2 \lesssim_{\log} n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \mathcal{Q}_{\alpha,d}\left(\frac{d}{\rho n^2}\right). \tag{29}$$

THEOREM 5.4 (Frobenius adaptation over $\mathcal{G}_\alpha$). *Under the conditions of Theorem 5.3, the estimator $\hat{\Sigma}^{\mathsf{Ada}}_{\mathcal{G},\mathrm{F}}$ in Algorithm 8 is ρ-zCDP and, uniformly over $\alpha \in [\underline{\alpha}, \overline{\alpha}]$,*

$$\sup_{P \in \mathcal{P}^{\mathcal{G}}_\alpha} \frac{1}{d}\mathbb{E}_P \left\| \hat{\Sigma}^{\mathsf{Ada}}_{\mathcal{G},\mathrm{F}} - \Sigma(P) \right\|_F^2 \lesssim_{\log} n^{-\frac{2\alpha+1}{2\alpha+2}} \wedge \frac{d}{n} + \left(\frac{d}{\rho n^2}\right)^{\frac{2\alpha+1}{2\alpha+3}} \wedge \frac{d^3}{\rho n^2}. \tag{30}$$

6. Estimating the Precision Matrix under DP. Estimating the precision matrix, i.e., the inverse of the covariance matrix, is also a fundamental problem in statistics and machine learning, with important applications in graphical models, portfolio optimization, and other areas. As in the non-private setting [21], an estimator of the precision matrix can be obtained by applying matrix inversion to the DP covariance estimator developed in this paper. Following the classical framework, we assume that the precision matrix is bounded in operator norm, which is equivalent to requiring the smallest eigenvalue of the covariance matrix to be bounded away from zero. Fix $0 < c_0 < M_0$. For $\mathcal{J} \in \{\mathcal{F}, \mathcal{G}, \mathcal{H}\}$, define the uniformly conditioned covariance and distribution classes

$$\tilde{\mathcal{J}}_\alpha = \mathcal{J}_\alpha \cap \{\Sigma : \lambda_{\min}(\Sigma) \ge c_0\}, \qquad \tilde{\mathcal{P}}^{\mathcal{J}}_\alpha = \left\{P \in \mathcal{P}^{\mathcal{J}}_\alpha : \Sigma(P) \in \tilde{\mathcal{J}}_\alpha\right\},$$

and write $\Omega(P) = \Sigma(P)^{-1}$.

Since a covariance matrix estimator $\hat{\Sigma}$ may not be invertible, we regularize it via eigenvalue truncation. Let the eigen-decomposition of $\hat{\Sigma}$ be $\hat{\Sigma} = \hat{U}\hat{\Lambda}\hat{U}^\top$, where $\hat{\Lambda} = \mathrm{Diag}(\hat{\lambda}_i)_{i\in[d]}$ is the diagonal matrix of eigenvalues. Then, we define the precision matrix estimator as

$$\hat{\Omega} = \hat{U}\mathrm{Diag}\left(\max(\hat{\lambda}_i, L_2^{-1})^{-1}\right)_{i\in[d]}\hat{U}^\top, \tag{31}$$

for any fixed constant $L_2 \ge c_0^{-1}$.

THEOREM 6.1. *Suppose that $\rho n^2/d \gtrsim (\log d)^{2(\alpha+1)}$. Then we have*

$$\inf_{\hat{\Omega} \in \mathcal{M}_\rho} \sup_{P \in \tilde{\mathcal{P}}^{\mathcal{F}}_\alpha} \mathbb{E}_P \left\| \hat{\Omega} - \Omega(P) \right\|^2 \asymp n^{-\frac{2\alpha}{2\alpha+1}} \wedge \frac{d}{n} + \left(\frac{d}{\rho n^2}\right)^{\frac{\alpha}{\alpha+1}} \wedge \frac{d^3}{\rho n^2}. \tag{32}$$

In particular, the upper bound can be achieved by the estimators in (31) based on the DP covariance matrix estimators. For each $\mathcal{J} \in \{\mathcal{G}, \mathcal{H}\}$, under the conditions of Theorem 2.6, the same conclusion holds with $\tilde{\mathcal{P}}^{\mathcal{F}}_\alpha$ replaced by $\tilde{\mathcal{P}}^{\mathcal{J}}_\alpha$, the right-hand side replaced by that of (10), and $\asymp$ replaced by $\asymp_{\log}$.

Theorem 6.1 shows that the minimax rates for estimating the precision matrix under DP constraints are the same as those for estimating the covariance matrix itself, similar to the non-private setting [21]. We remark that the adaptive estimator in Section 5 can also be used in (31) for adaptive estimation of the precision matrix.

The proof of Theorem 6.1 is contained in the supplementary material, and we give a brief sketch here. For the upper bound, eigenvalue truncation moves the covariance estimator by at most its distance from Σ, and the matrix inversion formula then gives $\left\|\hat{\Omega} - \Omega\right\| \le$

$2L_2c_0^{-1}\left\|\hat{\Sigma}-\Sigma\right\|$. Thus, the covariance upper bounds transfer directly by post-processing. For the lower bound, any DP precision matrix estimator can be symmetrized, spectrally truncated to the uniformly conditioned parameter space, and inverted to obtain a DP covariance matrix estimator. The inversion map is bi-Lipschitz on this parameter space, so the covariance lower bounds in Theorem 4.2 and Theorem 4.3 transfer directly.

7. Discussion. This paper develops minimax and adaptive theory for differentially private estimation over the nested covariance classes $\mathcal{H}_\alpha \subset \mathcal{G}_\alpha \subset \mathcal{F}_\alpha$ under operator and Frobenius losses. The central statistical message is that privacy fundamentally changes how these classes compare. Under operator loss, all three classes share the same leading non-private algebraic rate, yet their privacy costs separate the broad separated-block class from the row-tail and pointwise-decay classes. For the latter two classes, the operator privacy cost exhibits an additional phase transition at $\alpha = 1/2$, whereas no analogous smoothness transition occurs under Frobenius loss. Thus, the decay exponent alone does not determine the difficulty of private estimation: the local covariance geometry and the loss function also play essential roles.

The estimator constructions reflect these geometric distinctions. Over $\mathcal{F}_\alpha$, blockwise tridiagonal estimation balances truncation bias, sampling error, and privacy noise. Over $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$, a center–outer dyadic decomposition separates distance scales and supports profile, peeling, and greedy procedures tailored to each class. The lower bounds are obtained by combining a zCDP van Trees inequality with dependent matrix-valued priors tailored to the geometry of each class. These constructions identify the same scale and rank obstructions that govern the upper bounds. Under uniform spectral conditioning, the operator norm guarantees further extend to precision matrix estimation through spectral truncation and inversion.

Several important questions remain open. First, the minimax rates exhibit an unavoidable polynomial dependence on the ambient dimension d, even in the presence of ordered off-diagonal decay, which may limit their applicability in extremely high-dimensional regimes. It would be valuable to determine whether alternative structural assumptions, including sparsity, low rank, or Toeplitz structure [15], can reduce this dimensional dependence. Differentially private estimation of covariance functions for functional data [19] is another promising direction.

Second, the private covariance estimators developed here may serve as building blocks for downstream procedures, including principal component analysis, uncertainty quantification, discriminant analysis, and graphical model estimation under privacy constraints. Establishing optimality guarantees for these downstream procedures would clarify how error in covariance estimation propagates through private multivariate analysis.

Third, extending the present central DP framework to distributed or federated settings is of substantial practical interest. In such settings, observations are partitioned across multiple users or institutions that may have heterogeneous sample sizes, covariance structures, communication constraints, and privacy requirements. These features introduce new statistical and algorithmic trade-offs that are absent from the centralized model.

Finally, it remains open whether the polylogarithmic overheads in the adaptive guarantees are intrinsic. In particular, it is unknown whether the oracle ratios bounded by fixed polylogarithmic factors for the row-tail procedures and the pointwise-decay operator procedure can be sharpened, or whether the logarithmic privacy overheads for the separated-block procedures and the pointwise-decay Frobenius procedure are unavoidable.

We hope that the results and techniques developed here will stimulate further research on differentially private multivariate analysis and high-dimensional statistics.

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

SUPPLEMENTARY MATERIAL

Supplementary Material for "Minimax and Adaptive Covariance Matrix Estimation under Differential Privacy"

The supplementary material contains numerical experiments and all the proofs omitted in the main text.

REFERENCES

[1] ACHARYA, J., SUN, Z. and ZHANG, H. (2021). Differentially Private Assouad, Fano, and Le Cam. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory* (V. FELDMAN, K. LIGETT and S. SABATO, eds.). *Proceedings of Machine Learning Research* **132** 48–78. PMLR.

[2] ADNAN, M., KALRA, S., CRESSWELL, J. C., TAYLOR, G. W. and TIZHOOSH, H. R. (2022). Federated Learning and Differential Privacy for Medical Image Analysis. *Scientific Reports* **12** 1953. https://doi.org/10.1038/s41598-022-05539-7

[3] ALABI, D., KOTHARI, P. K., TANKALA, P., VENKAT, P. and ZHANG, F. (2023). Privately Estimating a Gaussian: Efficient, Robust, and Optimal. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. *STOC 2023* 483–496. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3564246.3585194

[4] AMIN, K., DICK, T., KULESZA, A., MUÑOZ MEDINA, A. and VASSILVITSKII, S. (2019). Differentially Private Covariance Estimation. In *Advances in Neural Information Processing Systems* **32** 14190–14199. Curran Associates, Inc.

[5] ASHTIANI, H. and LIAW, C. (2022). Private and Polynomial Time Algorithms for Learning Gaussians and Beyond. In *Proceedings of Thirty Fifth Conference on Learning Theory* (P.-L. LOH and M. RAGINSKY, eds.). *Proceedings of Machine Learning Research* **178** 1075–1076. PMLR.

[6] BASSILY, R., SMITH, A. and THAKURTA, A. (2014). Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science* 464–473. IEEE. https://doi.org/10.1109/FOCS.2014.56

[7] BICKEL, P. J. and LEVINA, E. (2008a). Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics* **36** 199–227. https://doi.org/10.1214/009053607000000758

[8] BICKEL, P. J. and LEVINA, E. (2008b). Covariance Regularization by Thresholding. *The Annals of Statistics* **36** 2577–2604. https://doi.org/10.1214/08-AOS600

[9] BISWAS, S., DONG, Y., KAMATH, G. and ULLMAN, J. (2020). Coinpress: Practical Private Mean and Covariance Estimation. *Advances in Neural Information Processing Systems* **33** 14475–14485.

[10] BUN, M. and STEINKE, T. (2016). Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. In *Theory of Cryptography*. *Lecture Notes in Computer Science* **9985** 635–658. Springer. https://doi.org/10.1007/978-3-662-53641-4_24

[11] CAI, T. T., CHAKRABORTY, A. and VUURSTEEN, L. (2024). Optimal Federated Learning for Nonparametric Regression with Heterogeneous Distributed Differential Privacy Constraints.

[12] CAI, T. and LIU, W. (2011). Adaptive Thresholding for Sparse Covariance Matrix Estimation. *Journal of the American Statistical Association* **106** 672–684. https://doi.org/10.1198/jasa.2011.tm10560

[13] CAI, T., LIU, W. and LUO, X. (2011). A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association* **106** 594–607. https://doi.org/10.1198/jasa.2011.tm10155

[14] CAI, T. T., REN, Z. and ZHOU, H. H. (2013). Optimal Rates of Convergence for Estimating Toeplitz Covariance Matrices. *Probability Theory and Related Fields* **156** 101–143. https://doi.org/10.1007/s00440-012-0422-7

[15] CAI, T. T., REN, Z. and ZHOU, H. H. (2016). Estimating Structured High-Dimensional Covariance and Precision Matrices: Optimal Rates and Adaptive Estimation. *Electronic Journal of Statistics* **10** 1–59. https://doi.org/10.1214/15-EJS1081

[16] CAI, T. T., WANG, Y. and ZHANG, L. (2021). The Cost of Privacy: Optimal Rates of Convergence for Parameter Estimation with Differential Privacy. *The Annals of Statistics* **49** 2825–2850. https://doi.org/10.1214/21-AOS2058

[17] CAI, T. T., WANG, Y. and ZHANG, L. (2025). Score Attack: A Lower Bound Technique for Optimal Differentially Private Learning. https://doi.org/10.48550/arXiv.2303.07152

[18] CAI, T. T., XIA, D. and ZHA, M. (2024). Optimal Differentially Private PCA and Estimation for Spiked Covariance Matrices. https://doi.org/10.48550/arXiv.2401.03820

[19] CAI, T. T. and YUAN, M. (2010). Nonparametric Covariance Function Estimation for Functional and Longitudinal Data Technical Report, University of Pennsylvania and Georgia Institute of Technology.

[20] CAI, T. T. and YUAN, M. (2012). Adaptive Covariance Matrix Estimation through Block Thresholding. *The Annals of Statistics* **40** 2014–2042. https://doi.org/10.1214/12-AOS999

[21] CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal Rates of Convergence for Covariance Matrix Estimation. *The Annals of Statistics* **38** 2118–2144. https://doi.org/10.1214/09-AOS752

[22] CAI, T. T. and ZHOU, H. H. (2012). Optimal Rates of Convergence for Sparse Covariance Matrix Estimation. *The Annals of Statistics* **40** 2389–2420. https://doi.org/10.1214/12-AOS998

[23] DONG, W., LIANG, Y. and YI, K. (2022). Differentially Private Covariance Revisited. In *Advances in Neural Information Processing Systems* **35**. Neural Information Processing Systems Foundation.

[24] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local Privacy and Statistical Minimax Rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* 429–438. IEEE. https://doi.org/10.1109/FOCS.2013.53

[25] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax Optimal Procedures for Locally Private Estimation. *Journal of the American Statistical Association* **113** 182–201. https://doi.org/10.1080/01621459.2017.1389735

[26] DWORK, C. and ROTH, A. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science* **9** 211–407. https://doi.org/10.1561/0400000042

[27] DWORK, C. and ROTHBLUM, G. N. (2016). Concentrated Differential Privacy. https://doi.org/10.48550/arXiv.1603.01887

[28] DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography*. *Lecture Notes in Computer Science* **3876** 265–284. Springer. https://doi.org/10.1007/11681878_14

[29] HAWES, M. B. (2020). Implementing Differential Privacy: Seven Lessons from the 2020 United States Census. *Harvard Data Science Review* **2**. https://doi.org/10.1162/99608f92.353c6f99

[30] KAMATH, G., MOUZAKIS, A. and SINGHAL, V. (2022). New Lower Bounds for Private Estimation and a Generalized Fingerprinting Lemma. *Advances in neural information processing systems* **35** 24405–24418.

[31] KAMATH, G., MOUZAKIS, A., SINGHAL, V., STEINKE, T. and ULLMAN, J. (2022). A Private and Computationally-Efficient Estimator for Unbounded Gaussians. In *Proceedings of Thirty Fifth Conference on Learning Theory* (P.-L. LOH and M. RAGINSKY, eds.). *Proceedings of Machine Learning Research* **178** 544–572. PMLR.

[32] NARAYANAN, S. (2025). Better and Simpler Lower Bounds for Differentially Private Statistical Estimation. *IEEE Transactions on Information Theory* **71** 1376–1388. https://doi.org/10.1109/TIT.2024.3511624

[33] PETER, N., TSFADIA, E. and ULLMAN, J. (2024). Smooth Lower Bounds for Differentially Private Algorithms via Padding-and-Permuting Fingerprinting Codes. In *Proceedings of Thirty Seventh Conference on Learning Theory* (S. AGRAWAL and A. ROTH, eds.). *Proceedings of Machine Learning Research* **247** 4207–4239. PMLR.

[34] PORTELLA, V. S. and HARVEY, N. J. A. (2025). Lower Bounds for Private Estimation of Gaussian Covariance Matrices under All Reasonable Parameter Regimes. In *Proceedings of Thirty Eighth Conference on Learning Theory* (N. HAGHTALAB and A. MOITRA, eds.). *Proceedings of Machine Learning Research* **291** 4640–4667. PMLR.

[35] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge University Press, Cambridge.

[36] XUE, G., LIN, Z. and YU, Y. (2024). Optimal Estimation in Private Distributed Functional Data Analysis. https://doi.org/10.48550/arXiv.2412.06582