Local Differential Privacy with Correlated Noise Achieves Central-DP Optimal Cost

Srikanth Avasarala, Viveck R. Cadambe, Madhura Pathegama, Juba Ziani
Georgia Institute of Technology, Atlanta, GA, USA
Email: {savasarala, viveck, macharige3, jziani3}@gatech.edu

***Abstract*—We study privately estimating the sum of n user-held values in the presence of an honest-but-curious server. This motivates requiring privacy not only at data release but also throughout server-side computation. We therefore adopt the local (pure) differential privacy model, in which each user transmits a noise-perturbed value. It is well known that independent local noise typically incurs a substantial utility loss compared to the centralized model, where noise is added only after aggregation.**

We show that this gap is not fundamental. By carefully designing correlations among the locally added noise variables, we construct ε-DP mechanisms whose estimation cost matches the optimal cost achievable in the centralized setting, up to an arbitrarily small error.

I. Introduction

Consider n users indexed by $i \in \{1, \dots, n\}$, where user i holds a private datum $x_i \in \mathbb{R}$.[1] A server aims to compute and release the aggregate sum $\sum_{i=1}^n x_i$ (or the average $\frac{1}{n}\sum_{i=1}^n x_i$) while preserving the privacy of each individual datum. This aggregation task arises in applications such as distributed and federated learning [1], [2], consensus algorithms [3], [4], and sensor network computations [5], [6], where users contribute local updates or statistics that must be combined while preserving individual privacy.

In the *central* differential privacy (central-DP) model [7], users transmit their raw data to a trusted server, which then adds noise to the aggregate before releasing it. When the server is not trusted (e.g., it may be honest-but-curious), stronger safeguards are required.

As an alternative, *local* differential privacy (local-DP) was proposed [8], [9]. In this model, privacy is enforced before data leaves the user. We consider an additive local-noise mechanism in which each user transmits $x_i + Z_i$, where Z_i is a random noise variable. The server's estimate of the sum becomes

$$\sum_{i=1}^n x_i + \sum_{i=1}^n Z_i.$$

Because noise is added locally, one expects that the server cannot reliably infer any individual x_i.

[1]Our analysis extends naturally to vector-valued x_i under mild conditions; see the conclusion.

In this work, we quantify privacy using pure ε-differential privacy (ε-DP) and focus exclusively on additive noise mechanisms. Let $x, \widetilde{x} \in \mathbb{R}^n$ be *neighboring* data vectors, meaning they differ in at most one coordinate and the change is bounded by a sensitivity parameter $\Delta > 0$; i.e., there exists $i \in \{1, \dots, n\}$ such that $x_j = \widetilde{x}_j$ for all $j \neq i$ and $|x_i - \widetilde{x}_i| \leq \Delta$. The ε-DP is then defined as follows.

Definition 1. The additive mechanism $x \mapsto x + Z$ satisfies ε-DP for sensitivity Δ if, for all neighboring $x, \widetilde{x} \in \mathbb{R}^n$ and all measurable sets $A \subseteq \mathbb{R}^n$,

$$\Pr\big(x + Z \in A\big) \;\leq\; e^{\varepsilon} \Pr\big(\widetilde{x} + Z \in A\big). \tag{1}$$

We seek to design the joint noise vector $Z^n = (Z_1, \dots, Z_n)$ to satisfy (1) while minimizing the aggregate cost, which in our setting depends only on the noise:

$$\mathbb{E}\left[L\left(\sum_{i=1}^n Z_i \right) \right]. \tag{2}$$

Here $L(\cdot)$ is even and nondecreasing on $\mathbb{R}_+$. A common choice is the L_p loss $L(x) = |x|^p$ for $p \geq 1$.

The prevailing approach in the local differential privacy literature is to add independent noise at each user, often i.i.d. from a common distribution [8], [9]. Independence greatly simplifies the privacy constraints: when the Z_i are independent, satisfying (1) jointly is equivalent to each Z_i satisfying the corresponding one-dimensional version of (1).

However, independent local noise typically incurs a strict utility gap relative to the centralized model [9]. In particular, for quadratic loss $L(x) = x^2$, the optimal independent mechanism is obtained by choosing each Z_i according to the one-dimensional optimal distribution, and the resulting aggregate cost is n times the cost of the corresponding centrally added-noise mechanism.

Our goal is to mitigate this utility gap by allowing correlated noise across users. Note that Definition 1 does not require independence and continues to guarantee privacy under correlation. In particular, letting $Z_{\sim i} := (Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n)$, for any measurable $A_i \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}^{n-1}$, conditioning (1) on the

event $\{Z_{\sim i} \in B\}$ yields

$$\Pr\big(x_i+Z_i \in A_i \mid Z_{\sim i} \in B\big) \le e^{\varepsilon} \Pr\big(\widetilde{x}_i + Z_i \in A_i \mid Z_{\sim i} \in B\big). \quad (3)$$

Thus, even if the server knows all other users' data and noise values (e.g., via collusion), it still cannot reliably distinguish whether user i's datum was x_i or $\widetilde{x}_i$.

This shows that using correlated noise additions does not compromise the privacy during the computation phase as long as the mechanism satisfies ε-DP condition in (1). Against this backdrop, we ask the following question.

Can we design ε-DP additive local mechanisms with correlated noise that significantly outperform the standard independent-noise mechanisms? We show that the answer is affirmative. We prove that appropriately designed correlated local mechanisms can achieve costs arbitrarily close to the optimal cost attainable under centralized ε-DP.

Our approach leverages the characterization of the optimal one-dimensional ε-DP mechanism in [10]: we reduce the n-user problem to a one-dimensional optimization along the aggregate noise direction, obtain a one-dimensional lower bound, and show it is essentially tight via a correlated construction.

Note that our results have direct implications for distributed summation tasks, most notably distributed mean estimation [9], [11], which underlies aggregation in gradient-based optimization and learning [12], [13] as well as consensus algorithms [3]. In particular, under quadratic loss, correlated noise can achieve an $O(1)$ error for mean estimation, whereas independent local noise leads to an $O(n)$ scaling of the loss.

A. Related work

It is worth noting that under *approximate* differential privacy[2], gains from correlated noise are known in the *Gaussian* setting [14]–[18]. In particular, [16] and [18] show that correlated Gaussian noise can match the performance of centrally added Gaussian noise under quadratic loss. However, these results rely on structural properties specific to Gaussian mechanisms and are tailored to quadratic loss. They do not extend to pure ε-DP or to general loss functions.

Our focus in this work is to show that carefully constructed correlated local noise can achieve centralized-DP utility. We do not address the efficient or secure generation of such correlations; these may be developed based on secure aggregation techniques [17], [19]–[21].

It is also worth noting that temporal correlations in added noise have been used to improve distributed gradient methods [22], [23], but such correlations are not relevant to our setting.

[2]Approximate differential privacy, also known as (ε, δ)-DP, relaxes (1) to $\Pr(x + Z \in A) \le e^{\varepsilon} \Pr(\widetilde{x} + Z \in A) + \delta$.

II. Preliminaries

A. Notation

We use the following notation throughout: for $A \subseteq \mathbb{R}^n$ and $v \in \mathbb{R}^n$, $A + v := \{x + v : x \in A\}$, and for $\alpha \in \mathbb{R}$, $\alpha A := \{\alpha x : x \in A\}$.

For a function $f : \mathbb{R}^n \to \mathbb{R}$, we write $f(\alpha(\cdot))$ to denote the function $x \mapsto f(\alpha x)$. Similarly for a probability measure P on $\mathbb{R}^n$, we write $P(\alpha(\cdot))$ for $P(\alpha A)$ as a function of A. All probability measures considered in this work are Borel measurable.

B. Differential privacy

In this work, we consider additive mechanisms in which the noise variables Z_i are chosen independently of the corresponding data x_i. Under this assumption, the privacy requirement (1) depends only on the joint distribution P of the noise vector Z, and the mechanism is fully specified by P.

To make this observation explicit, define the set of allowable coordinate-wise perturbations

$$T_n^{\Delta} := \bigcup_{i=1}^{n} \{0\}^{i-1} \times [-\Delta, \Delta] \times \{0\}^{n-i},$$

i.e., the zero-centered union of length-2Δ line segments along each coordinate axis. Then $x, \widetilde{x} \in \mathbb{R}^n$ are neighboring (for sensitivity Δ) if and only if $\widetilde{x} = x + t$ for a unique $t \in T_n^{\Delta}$. Hence (1) is equivalent to requiring that, for all $x \in \mathbb{R}^n$, all $t \in T_n^{\Delta}$, and all measurable $A \subseteq \mathbb{R}^n$,

$$\Pr\big(x + Z \in A\big) \le e^{\varepsilon} \Pr\big(x + t + Z \in A\big).$$

Equivalently

$$P(A) \le e^{\varepsilon} P(A + t). \quad (4)$$

Accordingly, when the mechanism $x \mapsto x + Z$ is ε-DP for sensitivity Δ, we will also say (by abuse of notation) that the noise law P itself is ε-DP for sensitivity Δ.

Definition 2. Given $\varepsilon > 0$ and $\Delta > 0$, a probability measure P on $\mathbb{R}^n$ satisfies ε-DP for sensitivity Δ if, for all measurable $A \subseteq \mathbb{R}^n$ and all $t \in T_n^{\Delta}$, (4) is satisfied.

C. Cost function

The utility in our setting is quantified by the cost in (2). If the noise vector $Z = (Z_1, \ldots, Z_n)$ is distributed according to P then the cost can be written as

$$\int_{\mathbb{R}^n} L\left(\sum_{i=1}^{n} z_i\right) P(dz_1 \cdots dz_n).$$

We impose the following mild and natural assumptions on the loss function $L : \mathbb{R} \to \mathbb{R}$: **(A1)** L is even; **(A2)** L is nondecreasing on $\mathbb{R}_+$; and **(A3)** L has at most subexponential growth, i.e.,

$$L(x) = \exp(o(|x|)) \qquad \text{as } |x| \to \infty.$$

Since both privacy and utility are determined by the noise law P of $Z = (Z_1, \dots, Z_n)$, our focus is to solve the following optimization problem.

$$\inf_P \int_{\mathbb{R}^n} L\left(\sum_{i=1}^n z_i\right) P(dz), \tag{5}$$

where optimization is over all P satisfying ε-DP for sensitivity Δ. Our main result (Theorem 1) resolves (5) by matching its optimum to the centralized benchmark and showing it is approachable under local DP using correlated noise.

Although the cost in (5) is defined in terms of an n-dimensional noise measure P, it is useful to also consider costs associated with one-dimensional probability measures, as much of the analysis reduces to a scalar problem. Accordingly, for a cost function $L : \mathbb{R} \to \mathbb{R}$ and a probability measure Q on $\mathbb{R}$, we define the associated cost

$$C(L, Q) := \int_{\mathbb{R}} L(x)\, Q(dx).$$

Let $\mathcal{P}_n(\varepsilon, \Delta)$ denote the class of Borel probability measures on $\mathbb{R}^n$ that satisfy ε-differential privacy for sensitivity Δ. We define the optimal cost (over one-dimensional noise measures) as

$$C^*_{\varepsilon,\Delta}(L) \triangleq \inf_{Q \in \mathcal{P}_1(\varepsilon,\Delta)} C(L, Q). \tag{6}$$

It was shown in [10] that for any cost function L satisfying Assumptions **(A1)–(A3)**, an optimal noise distribution exists and admits a density, which we denote by $f^*_{L,\varepsilon,\Delta}$. The optimal value of the objective can therefore be written as

$$C^*_{\varepsilon,\Delta}(L) \;=\; \int_{\mathbb{R}} L(x)\, f^*_{L,\varepsilon,\Delta}(x)\, dx.$$

More specifically, [10] shows that the optimal density $f^*_{L,\varepsilon,\Delta}$ belongs to the parametric family of *staircase mechanisms* $\{f^{(\varepsilon,\Delta)}_\gamma : \gamma \in [0,1]\}$. We use these functions to prove the achievability part of our main theorem; see Appendix B for details.

Remark 1. The quantity $C^*_{\varepsilon,\Delta}(L)$ has a direct operational meaning in the centralized model of differential privacy. In the *central* DP setting, users send their raw data to a trusted server, which releases a randomized aggregate. For the additive mechanism $\sum_{i=1}^n x_i + Z_0$, where $Z_0 \sim Q$, the ε-DP requirement for neighboring $x, \widetilde{x}$ is

$$\Pr\left(\sum_{i=1}^n x_i + Z_0 \in A\right) \le e^\varepsilon \Pr\left(\sum_{i=1}^n \widetilde{x}_i + Z_0 \in A\right).$$

Equivalently, the one-dimensional noise law Q must satisfy ε-DP with sensitivity Δ in the sense of Definition 2. Hence $C^*_{\varepsilon,\Delta}(L)$ is exactly the minimum achievable centralized cost.

We state a simple scaling property of the optimal cost under dilation of the loss function, which will be used in the proof of the main theorem. The proof is deferred to Appendix A.

Lemma 1 (Scaling invariance of the optimal cost)**.** *Let $\varepsilon, \Delta, \alpha > 0$ and let $L : \mathbb{R} \to \mathbb{R}$ satisfy **(A1)–(A3)**. Then*

$$C^*_{\varepsilon,\Delta/\alpha}(L(\alpha(\cdot))) = C^*_{\varepsilon,\Delta}(L). \tag{7}$$

III. The main result

Our main result characterizes the optimal cost achievable under ε-DP for sensitivity Δ. The theorem is stated in two parts: a universal lower bound and an achievability statement.

Theorem 1. *Let $\varepsilon, \Delta > 0$ and consider an n-user additive-noise mechanism in which the noise vector $Z = (Z_1, \dots, Z_n) \sim P$ and the mechanism satisfies ε-DP for sensitivity Δ. Then*

$$\int_{\mathbb{R}^n} L\left(\sum_{i=1}^n z_i\right) P(dz_1 \cdots dz_n) \;\ge\; C^*_{\varepsilon,\Delta}(L) \tag{8}$$

Moreover, for every $\eta > 0$, there exists a probability measure $\widetilde{P}$ on $\mathbb{R}^n$ such that the additive-noise mechanism with $Z \sim \widetilde{P}$ satisfies ε-DP for sensitivity Δ and

$$\int_{\mathbb{R}^n} L\left(\sum_{i=1}^n z_i\right) \widetilde{P}(dz_1 \cdots dz_n) - C^*_{\varepsilon,\Delta}(L) \;\le\; \eta. \tag{9}$$

This result shows that by using a correlated local differential privacy mechanism, we can arbitrarily approximate the optimal cost of the central DP setting, while still preventing the server from learning individual data.

Before proceeding with the proof, we introduce some notation. Let S be an orthogonal matrix whose first row is $\frac{1}{\sqrt{n}}(1, \dots, 1)$, and define $U = SZ$. Under this transformation,

$$U_1 = \frac{1}{\sqrt{n}}(Z_1 + \cdots + Z_n),$$

while the remaining coordinates span the subspace orthogonal to the all-ones vector.

The distribution of U is denoted by P^S and is obtained from P via the change of coordinates induced by S:

$$P^S(A) := P(S^{-1}A), \qquad A \subseteq \mathbb{R}^n \text{ measurable},$$

where $S^{-1}A = \{S^{-1}u : u \in A\}$. With this notation in place, we prove the lower bound in (8).

Proof of the lower bound. Under the change of variables $U = SZ$, the cost depends only on the first coordinate:

$$\int_{\mathbb{R}^n} L\left(\sum_{i=1}^n z_i\right) P(dz) = \int_{\mathbb{R}^n} L(\sqrt{n}\, u_1)\, P^S(du)$$
$$= \int_{\mathbb{R}} L(\sqrt{n}\, u_1)\, P_1^S(du_1), \quad (10)$$

where P_1^S is the marginal law of U_1 under P^S. This yields a one-dimensional representation of the cost function.

Next, we express the ε-DP constraint (4) in the rotated coordinates. The measure P satisfies (4) if and only if, for all measurable $A \subseteq \mathbb{R}^n$ and all $t \in T_n^\Delta$,

$$P^S(A) \le e^\varepsilon P^S(A + St). \quad (11)$$

Since the first row of S is $\frac{1}{\sqrt{n}}(1, \dots, 1)$, for any $t \in T_n^\Delta$ we have

$$|(St)_1| = \frac{\left|\sum_{i=1}^n t_i\right|}{\sqrt{n}} \le \frac{\Delta}{\sqrt{n}}.$$

Moreover, every $t' \in [-\Delta/\sqrt{n}, \Delta/\sqrt{n}]$ can be realized as $(St)_1$ for some $t \in T_n^\Delta$. Now fix any measurable $A_1 \subseteq \mathbb{R}$ and set $A := A_1 \times \mathbb{R}^{n-1}$. Applying (11) to this set and taking the marginal of the first coordinate yields

$$P_1^S(A_1) \le e^\varepsilon P_1^S(A_1 + t'), \qquad |t'| \le \frac{\Delta}{\sqrt{n}}, \quad (12)$$

so P_1^S satisfies ε-DP for sensitivity $\Delta/\sqrt{n}$. Therefore,

$$\int_{\mathbb{R}} L(\sqrt{n}\, u)\, P_1^S(du) \ge C^*_{\varepsilon,\Delta/\sqrt{n}}\big(L(\sqrt{n}(\cdot))\big) = C^*_{\varepsilon,\Delta}(L),$$

where the last equality follows from Lemma 1. Together with (10), this proves the desired lower bound. □

Proof (sketch) of achievability. We work in the rotated coordinates $U = (U_1, \dots, U_n)$ to construct $\widetilde{P}$. Specifically, we first define the coordinate-transformed law $\widetilde{P}^S$ to be a product measure, so that $U \sim \widetilde{P}^S$ has independent components. We choose U_1 to have a density close to $f^*_{L(\sqrt{n}(\cdot)),\varepsilon,\Delta/\sqrt{n}}$ but satisfying a stricter ε_0-DP constraint for some $\varepsilon_0 < \varepsilon$, and take $U_2, \dots, U_n$ to have sufficiently slowly decaying densities. Intuitively, U_1 controls the aggregate $\sum_i Z_i$ (and hence the cost), while the remaining coordinates provide enough "slack" to absorb the remaining privacy budget $\varepsilon - \varepsilon_0$ under shifts in ST_n^Δ. Consequently, $\widetilde{P}$ satisfies ε-DP, and letting $\varepsilon_0 \uparrow \varepsilon$ yields cost arbitrarily close to the lower bound. A rigorous proof is given in Appendix B. □

We numerically evaluate the two-user case ($n = 2$) with sensitivity $\Delta = 2$. Figure 1 compares the aggregate squared loss under several mechanisms; the dashed black curve shows the lower bound $C^*_{\varepsilon,\Delta}$. All other curves are estimated from 5×10^5 samples of (Z_1, Z_2).

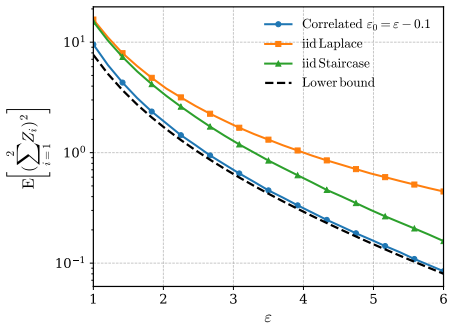

Fig. 1: Quadratic loss vs ε for different DP schemes.

The orange curve corresponds to independent Laplace(Δ/ε) noise[3], the green curve to the product of one-dimensional optimal densities $f^*_{L,\varepsilon,\Delta}$, and the blue curve to our correlated construction $\widetilde{P}$ with $\varepsilon_0 = \varepsilon - 0.1$ (specified in Appendix B). These results empirically validate the optimality of the correlated construction relative to standard local DP schemes.

IV. Conclusion

We showed that the utility gap between local and centralized differential privacy is not inherent to locality, but to the restriction to independent noise. By introducing carefully designed correlations among locally added noise variables, we constructed ε-DP mechanisms whose aggregate cost can approach the optimal centralized cost arbitrarily closely.

Our approach extends beyond pure ε-DP: the lower-bound and scaling arguments carry over to approximate DP and Rényi DP [24], and achievability should persist when the relevant one-dimensional optimizers are stable under small parameter changes.

Our analysis also extends to vector-valued data. Under ℓ_∞ sensitivity, the problem decouples coordinatewise and reduces to the scalar setting studied here. When sensitivity is measured using other norms, similar lower-bound continue to apply. Moreover, recent results on optimal mechanisms for vector-valued sensitivities [25] suggest that our construction can be extended as well.

An important direction for future work is the efficient and secure generation of the required correlations. In the approximate DP setting, correlated Gaussian schemes use independent noise combined with pairwise anticorrelated components that cancel upon aggregation [17], [19], [20]. While similar ideas may be relevant here, our construction is not directly reducible to such a decomposition, as the resulting aggregate noise cannot generally be expressed as a sum of independent random variables.

[3]Laplace(λ) has density $\frac{1}{2\lambda} e^{-|x|/\lambda}$, and Laplace($\Delta/\varepsilon$) is ε-DP for sensitivity Δ [7].

REFERENCES

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[3] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proceedings of the 2012 ACM workshop on Privacy in the electronic society*, 2012, pp. 81–90.

[4] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus with optimal noise selection," *IFAC-PapersOnLine*, vol. 48, no. 22, pp. 203–208, 2015.

[5] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, 2010, pp. 735–746.

[6] J. Won, C. Y. Ma, D. K. Yau, and N. S. Rao, "Proactive fault-tolerant aggregation protocol for privacy-assured smart metering," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 2804–2812.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*. Springer, 2006, pp. 265–284.

[8] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.

[9] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *2013 IEEE 54th annual symposium on foundations of computer science*. IEEE, 2013, pp. 429–438.

[10] Q. Geng and P. Viswanath, "The optimal noise-adding mechanism in differential privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 925–951, 2015.

[11] H. Asi, V. Feldman, and K. Talwar, "Optimal algorithms for mean estimation under local differential privacy," in *International Conference on Machine Learning*. PMLR, 2022, pp. 1046–1056.

[12] A. Rajkumar and S. Agarwal, "A differentially private stochastic gradient descent algorithm for multiparty classification," in *Artificial Intelligence and Statistics*. PMLR, 2012, pp. 933–941.

[13] S. Song, K. Chaudhuri, A. D. Sarwate *et al.*, "Stochastic gradient descent with differentially private updates." in *GlobalSIP*, 2013, pp. 245–248.

[14] H. Imtiaz and A. D. Sarwate, "Distributed differentially private algorithms for matrix and tensor factorization," *IEEE journal of selected topics in signal processing*, vol. 12, no. 6, pp. 1449–1464, 2018.

[15] H. Imtiaz, J. Mohammadi, R. Silva, B. Baker, S. M. Plis, A. D. Sarwate, and V. D. Calhoun, "A correlated noise-assisted decentralized differentially private estimation protocol, and its application to fMRI source separation," *IEEE Transactions on Signal Processing*, vol. 69, pp. 6355–6370, 2021.

[16] Y. Allouah, A. Koloskova, A. El Firdoussi, M. Jaggi, and R. Guerraoui, "The privacy power of correlated noise in decentralized learning," in *Proceedings of the 41st International Conference on Machine Learning*, 2024, pp. 1115–1143.

[17] S. Vithana, V. R. Cadambe, F. P. Calmon, and H. Jeong, "Correlated privacy mechanisms for differentially private distributed mean estimation," in *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE, 2025, pp. 590–614.

[18] ——, "Differentially private distributed mean estimation with constrained user correlations," in *2025 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2025, pp. 1–6.

[19] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1175–1191.

[20] W.-N. Chen, C. A. C. Choo, P. Kairouz, and A. T. Suresh, "The fundamental price of secure aggregation in differentially private federated learning," in *International Conference on Machine Learning*. PMLR, 2022, pp. 3056–3089.

[21] T. Stevens, C. Skalka, C. Vincent, J. Ring, S. Clark, and J. Near, "Efficient differentially private secure aggregation for federated learning via hardness of learning with errors," in *31st USENIX security symposium (USENIX Security 22)*, 2022, pp. 1379–1395.

[22] P. Kairouz, B. McMahan, S. Song, O. Thakkar, A. Thakurta, and Z. Xu, "Practical and private (deep) learning without sampling or shuffling," in *International Conference on Machine Learning*. PMLR, 2021, pp. 5213–5225.

[23] K. Pillutla, J. Upadhyay, C. A. Choquette-Choo, K. Dvijotham, A. Ganesh, M. Henzinger, J. Katz, R. McKenna, H. B. McMahan, K. Rush *et al.*, "Correlated noise mechanisms for differentially private learning," *arXiv preprint arXiv:2506.08201*, 2025.

[24] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 2017, pp. 263–275.

[25] A. Kulesza, A. T. Suresh, and Y. Wang, "General staircase mechanisms for optimal differential privacy," in *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, vol. 258, 2025, pp. 4564–4572.

APPENDIX A
PROOF OF LEMMA 1

For convenience, we restate Lemma 1.

Lemma (Scaling invariance of the optimal cost)**.** *Let $\varepsilon, \Delta, \alpha > 0$ and let $L : \mathbb{R} \to \mathbb{R}$ satisfy **(A1)–(A3)**. Then*

$$C^*_{\varepsilon,\Delta/\alpha}(L(\alpha(\cdot))) = C^*_{\varepsilon,\Delta}(L).$$

Proof. We need to prove the following equality:

$$\int_{\mathbb{R}} L(\alpha x)\, f^*_{L(\alpha(\cdot)),\varepsilon,\Delta/\alpha}(x)dx = \int_{\mathbb{R}} L(x) f^*_{L,\varepsilon,\Delta}(x)dx.$$

Observe that if a density g satisfies ε-DP for sensitivity Δ, then the scaled density $x \mapsto \alpha g(\alpha x)$ satisfies ε-DP for sensitivity Δ/α. In particular, $\alpha f^*_{L,\varepsilon,\Delta}(\alpha x)$ is feasible for the optimization defining $f^*_{L(\alpha(\cdot)),\varepsilon,\Delta/\alpha}$.

By optimality,

$$\begin{aligned}\int_{\mathbb{R}} L(\alpha x) f^*_{L(\alpha(\cdot)),\varepsilon,\Delta/\alpha}(x)dx &\le \int_{\mathbb{R}} L(\alpha x)\alpha f^*_{L,\varepsilon,\Delta}(\alpha x)dx \\ &= \int_{\mathbb{R}} L(u) f^*_{L,\varepsilon,\Delta}(u)du,\end{aligned}$$

where the last equality follows from the change of variables $u = \alpha x$.

Applying the same argument with α replaced by $1/\alpha$ yields the reverse inequality, which proves (7). □

APPENDIX B
PROOF OF ACHIEVABILITY OF THEOREM 1

For the reader's convenience, we restate the achievability claim in Theorem 1:

For every $\eta > 0$, there exists a probability measure $\widetilde{P}$ on $\mathbb{R}^n$ such that the additive-noise mechanism with $Z \sim \widetilde{P}$ satisfies ε-DP for sensitivity Δ and

$$\int_{\mathbb{R}^n} L\left(\sum_{i=1}^n z_i\right) \widetilde{P}(dz_1 \cdots dz_n) - C^*_{\varepsilon,\Delta}(L) \;\le\; \eta. \quad (13)$$

To prove achievability, we construct a probability measure $\widetilde{P}$ on $\mathbb{R}^n$ that satisfies ε-DP for sensitivity Δ and whose cost approaches the lower bound in (8).

Our construction relies on the *staircase* densities of [10], named for their piecewise-constant form. They are parametrized by (ε, Δ) and a shape parameter $\gamma \in [0, 1]$. For notational simplicity, we suppress (ε, Δ) and write f_γ for $f^{(\varepsilon,\Delta)}_\gamma$:

$$f_\gamma(x) = \begin{cases} a_\gamma, & 0 \le x < \gamma\Delta, \\ e^{-\varepsilon} a_\gamma, & \gamma\Delta \le x < \Delta, \\ e^{-k\varepsilon} f_\gamma(x - k\Delta), & k\Delta \le x < (k+1)\Delta, k \in \mathbb{N}, \\ f_\gamma(-x), & x < 0, \end{cases} \quad (14)$$

where $a_\gamma := a(\gamma, \varepsilon, \Delta)$ is the normalizing constant. Under Assumptions **(A1)–(A3)**, [10] shows that the optimizer $f^*_{L,\varepsilon,\Delta}$ belongs to this family.

Armed with this characterization, we now prove achievability.

Proof. Let γ^* be such that

$$f^*_{L(\sqrt{n}(\cdot)),\varepsilon,\Delta/\sqrt{n}} = f^{(\varepsilon,\Delta/\sqrt{n})}_{\gamma^*}. \quad (15)$$

That is, γ^* indexes the staircase density that is optimal for privacy level ε and sensitivity $\Delta/\sqrt{n}$.

Fix $\varepsilon_0 \in (0, \varepsilon]$. Define $\widetilde{P}$ via its pushforward under the orthogonal transform S as

$$\widetilde{P}^S(d(u_1, \ldots, u_n)) = f^{(\varepsilon_0,\Delta/\sqrt{n})}_{\gamma^*}(u_1)\, g(u_2) \cdots g(u_n),$$

where g is a density on $\mathbb{R}$ satisfying $\frac{\varepsilon-\varepsilon_0}{n-1}$-DP for sensitivity Δ. For example, one may take g to be $\mathrm{Laplace}(\lambda)$ with $\lambda = \frac{\Delta(n-1)}{\varepsilon-\varepsilon_0}$.[4]

First, we need to verify that this choice satisfies the ε-DP constraint in (11). Observe that for any shift

$$\widetilde{t} \in [-\Delta/\sqrt{n}, \Delta/\sqrt{n}] \times [-\Delta, \Delta]^{n-1}, \quad (16)$$

we have

$$\begin{aligned} &f^{(\varepsilon_0,\Delta/\sqrt{n})}_{\gamma^*}(u_1)\, g(u_2) \cdots g(u_n) \\ &\quad\le e^{\varepsilon_0}\, f^{(\varepsilon_0,\Delta/\sqrt{n})}_{\gamma^*}(u_1 + \widetilde{t}_1) \prod_{i=2}^n e^{\frac{\varepsilon-\varepsilon_0}{n-1}}\, g(u_i + \widetilde{t}_i) \\ &\quad= e^{\varepsilon}\, f^{(\varepsilon_0,\Delta/\sqrt{n})}_{\gamma^*}(u_1 + \widetilde{t}_1)\, g(u_2 + \widetilde{t}_2) \cdots g(u_n + \widetilde{t}_n). \end{aligned}$$

Let ST^Δ_n denote the image of T^Δ_n under S, i.e., $ST^\Delta_n := \{Sx : x \in T^\Delta_n\}$. If we show that ST^Δ_n is contained in the cuboid in (16), then the ε-DP claim follows from the characterization in (11).

Indeed, by (12) we have $(St)_1 \in [-\Delta/\sqrt{n}, \Delta/\sqrt{n}]$ for all $t \in T^\Delta_n$. Moreover, since $\|t\|_2 \le \Delta$ and S is orthogonal, $\|St\|_2 = \|t\|_2 \le \Delta$, which implies $|(St)_i| \le \Delta$ for every $i \ge 2$. Hence $ST^\Delta_n \subset [-\Delta/\sqrt{n}, \Delta/\sqrt{n}] \times [-\Delta, \Delta]^{n-1}$ and therefore $\widetilde{P}$ satisfies ε-DP for sensitivity $\Delta/\sqrt{n}$.

We now compute the cost under $\widetilde{P}$. As in (10),

$$\int_{\mathbb{R}^n} L\Big(\sum_{i=1}^n z_i\Big)\, \widetilde{P}(dz) = \int_{\mathbb{R}} L(\sqrt{n}\, u)\, f^{(\varepsilon_0,\Delta/\sqrt{n})}_{\gamma^*}(u)\, du.$$

It remains to show that, for a suitable choice of ε_0, the right-hand-side integral approximates $C^*_{\varepsilon,\Delta}(L)$. Recall that

$$\begin{aligned} C^*_{\varepsilon,\Delta}(L) &= C^*_{\varepsilon,\Delta/\sqrt{n}}(L(\sqrt{n}(\cdot))) \\ &= \int_{\mathbb{R}} L(\sqrt{n}\, u)\, f^{(\varepsilon,\Delta/\sqrt{n})}_{\gamma^*}(u)\, du, \end{aligned}$$

where the first equality follows from Lemma 1, and the second one follows from (15).

We prove the approximation via the dominated convergence theorem. By (14), $f^{(\varepsilon_0,\Delta/\sqrt{n})}_{\gamma^*}$ converges

[4] In our implementation of $\widetilde{P}$ in Figure 1, we take g to be Laplace with this choice of λ.

to $f_{\gamma^*}^{(\varepsilon,\Delta/\sqrt{n})}$ pointwise as $\varepsilon_0 \to \varepsilon$, and hence $L(\sqrt{n}\,(\cdot))\, f_{\gamma^*}^{(\varepsilon_0,\Delta/\sqrt{n})}$ converges to $L(\sqrt{n}(\cdot))\, f_{\gamma^*}^{(\varepsilon,\Delta/\sqrt{n})}$ pointwise.

Moreover, the staircase characterization (14) implies a uniform envelope: for $\varepsilon_0 \in (\varepsilon/2, \varepsilon]$ the densities $f_{\gamma^*}^{(\varepsilon_0,\Delta/\sqrt{n})}$ are dominated by an even function h with geometric tail decay. Since L has at most subexponential growth, we have $\int_{\mathbb{R}} L(\sqrt{n}\,u)\, h(u)\, du < \infty$. Therefore, dominated convergence yields

$$\int_{\mathbb{R}} L(\sqrt{n}\,u) f_{\gamma^*}^{(\varepsilon_0,\Delta/\sqrt{n})}(u)du \longrightarrow \int_{\mathbb{R}} L(\sqrt{n}\,u)\, f_{\gamma^*}^{(\varepsilon,\Delta/\sqrt{n})}(u)du \quad \text{as } \varepsilon_0 \to \varepsilon, \tag{17}$$

and hence we can choose ε_0 sufficiently close to ε so that the absolute difference is at most η. □

We visualize the asymptotic convergence of $\widetilde{P}$ by numerically simulating two users ($n = 2$) with $\Delta = 2$. The samples of Z are generated where the choice for g is a Laplace($\frac{\Delta}{\varepsilon-\varepsilon_0}$). We plot the quadratic loss versus ε for various choices of ε_0 and observe that as ε_0 approaches ε, the quadratic loss decreases and is closer to the theoretical lower bound $C^*_{\varepsilon,\Delta}(L)$.

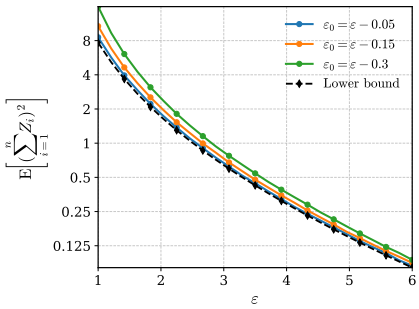

Fig. 2: quadratic loss vs ε for different ε_0.

Figure 3 illustrates the density of $\widetilde{P}$ for $n = 2$.

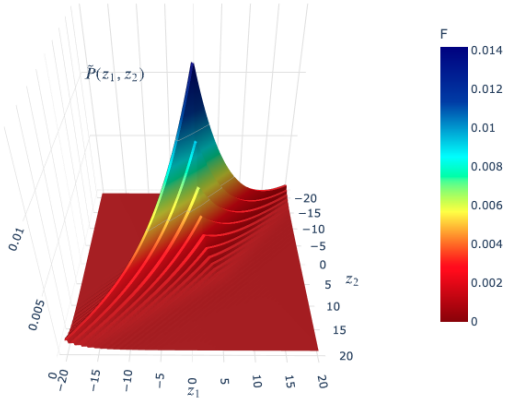

Fig. 3: Density of $\tilde{P}$

The plot corresponds to the two-user case with $\varepsilon = 0.5$, $\Delta = 1$, and $\varepsilon_0 = 0.4$, where g is chosen as $\mathrm{Laplace}\big(\Delta/(\varepsilon-\varepsilon_0)\big)$. As expected, the mass concentrates near the line $z_1 + z_2 = 0$ (i.e., $u_1 = 0$) and exhibits staircase decay along $z_1 = z_2$, equivalently along the u_1 direction.