The Composition Theorem for Differential Privacy

Peter Kairouz, *Member, IEEE*, Sewoong Oh, *Member, IEEE*, and Pramod Viswanath, *Fellow, IEEE*

***Abstract*— Sequential querying of differentially private mechanisms degrades the overall privacy level. In this paper, we answer the fundamental question of characterizing the level of overall privacy degradation as a function of the number of queries and the privacy levels maintained by each privatization mechanism. Our solution is complete: we prove an upper bound on the overall privacy level and construct a sequence of privatization mechanisms that achieves this bound. The key innovation is the introduction of an operational interpretation of differential privacy (involving hypothesis testing) and the use of a data processing inequality along with its converse. Our result improves over the state of the art, and has immediate connections to several problems studied in the literature.**

***Index Terms*— Differential privacy, hypothesis testing.**

I. Introduction

DIFFERENTIAL privacy is a formal framework to quantify to what extent individual privacy in a statistical database is preserved while releasing useful aggregate information about the database. It provides strong privacy guarantees by requiring the indistinguishability of whether or not an individual is in a database based on the released information, regardless of the side information on the other aspects of the database the adversary may possess. Denoting the database when the individual is present as D_1 and as D_0 when the individual is not, a differentially private mechanism provides indistinguishability guarantees with respect to the pair (D_0, D_1). The databases D_0 and D_1 are referred to as "neighboring" databases.

Definition 1 (Differential Privacy [10], [12]): A randomized mechanism M over a set of databases is (ε, δ)-differentially private *if for all pairs of neighboring databases D_0 and D_1, and for all sets S in the output space of the mechanism $\mathcal{X}$,*

$$\mathbb{P}(M(D_0) \in S) \leq e^{\varepsilon}\, \mathbb{P}(M(D_1) \in S) + \delta.$$

Manuscript received January 20, 2014; revised December 3, 2015, May 27, 2016, and January 12, 2017; accepted March 15, 2017. Date of publication March 21, 2017; date of current version May 18, 2017. This work was supported in part by NSF CISE under Award CCF-1422278, Award CCF-1553452, NSF SaTC Award CNS-1527754, NSF CMMI Award MES-1450848, NSF ENG Award ECCS-1232257, and in part by the Google Faculty Research Award. This paper was presented at the 2015 International Conference on Machine Learning in [KOV15].

P. Kairouz is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: kairouz2@illinois.edu).

S. Oh is with the Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana–Champaign, Champaign, IL 61801, USA (e-mail: swoh@illinois.edu).

P. Viswanath is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Champaign, IL 61801, USA (e-mail: pramodv@illinois.edu).

Communicated by A. Smith, Associate Editor for Complexity and Cryptography.

Digital Object Identifier 10.1109/TIT.2017.2685505

A basic problem in differential privacy is how privacy of a fixed pair of neighbors (D_0, D_1) degrades under *composition* of interactive queries when each query, individually, meets certain differential privacy guarantees. A routine argument shows that the composition of k queries, each of which is (ϵ, δ)-differentially private, is at least $(k\epsilon, k\delta)$-differentially private [10]–[12], [16]. A tighter bound of $(\tilde{\varepsilon}_{\tilde{\delta}}, k\delta + \tilde{\delta})$-differential privacy under k-fold adaptive composition is provided, using more sophisticated arguments, in [16] for the case when each of the individual queries is (ϵ, δ)-differentially private. Here $\tilde{\varepsilon}_{\tilde{\delta}} = O\left(k\varepsilon^2 + \varepsilon\sqrt{k\log(1/\tilde{\delta})}\right)$. On the other hand, it was not known if this bound could be improved until this work.

Our main result is the *exact* characterization of the privacy guarantee under k-fold composition. Any k-fold adaptive composition of (ε, δ)-differentially private mechanisms satisfies the privacy guarantee stated in Theorem 9. Further, we demonstrate a specific sequence of (nonadaptive) privacy mechanisms which when composed, degrade the privacy to the level guaranteed in Theorem 9. Our result entails a strict improvement over the state-of-the-art result in [16]. This can be seen immediately in the following approximation – using the same notation as above, the value of $\tilde{\varepsilon}_{\tilde{\delta}}$ is now reduced to $\tilde{\varepsilon}_{\tilde{\delta}} = O\left(k\varepsilon^2 + \varepsilon\sqrt{k\log(e + (\varepsilon\sqrt{k}/\tilde{\delta}))}\right)$. Since a typical choice of $\tilde{\delta}$ is $\tilde{\delta} = \Theta(k\delta)$, in the regime where $\varepsilon = \Theta(\sqrt{k}\delta)$, this improves the existing guarantee by a logarithmic factor. The gain is especially significant when both ε and δ are small.

We view differential privacy as a guarantee on the two types of error (false alarm and missed detection) in a binary hypothesis testing problem involving two neighboring databases. This approach is similar to the previous work of Wasserman and Zhou [33]. Our work leverages two benefits of this *operational interpretation* of differential privacy.

- The first is conceptual. The operational setting directs the logic of the steps of the proof, makes the arguments straightforward, and readily allows for generalizations such as heterogeneous compositions.
- The second is technical. The operational interpretation of hypothesis testing brings both the natural data processing inequality and the strong converse to the data processing inequality. These inequalities, while simple by themselves, lead to surprisingly strong technical results. As an aside, we mention that there is a strong tradition of such derivations in the information theory literature: the Fisher information inequality [5], [34], the entropy power inequality [5], [31], [32], an extremal inequality involving mutual informations [28], matrix determinant

inequalities [7], the Brunn-Minkowski inequality and its functional analytic variants [9] – [8, ch.17] enumerates a detailed list – were all derived using operational interpretations of mutual information and corresponding data processing inequalities.

The optimal composition theorem (Theorem 9) provides a fundamental limit on how much privacy degrades under composition. Such a characterization is a basic result in differential privacy and has been used widely in the literature [2], [16], [19], [22], [24], [29]. In each of these instances, the optimal composition theorem derived here (or the simpler characterization provided in Theorem 10) could be "cut-and-pasted", allowing for a corresponding strengthening of their conclusions. We demonstrate this strengthening for two instances: (a) the variance of noise adding mechanisms in Section V-A, and (b) the utility of graph cut and matrix variance queries in Appendix A. We further show that a variety of existing noise adding mechanisms ensures the same level of privacy with similar variances. This implies that there is nothing special about the popular choice of adding a Gaussian noise when composing multiple queries, and the same utility as measured through the noise variance can be obtained using other known mechanisms. We start our discussions by operationally introducing differential privacy as a guarantee on the error probabilities of a binary hypothesis testing problem.

II. Differential Privacy as Hypothesis Testing

Given a random output Y of a database access mechanism M, consider the following hypothesis testing experiment. We choose a null hypothesis as database D_0 and alternative hypothesis as D_1:

$$H0: Y \text{ came from a database } D_0,$$
$$H1: Y \text{ came from a database } D_1.$$

For a choice of a rejection region S, the probability of false alarm (type I error), when the null hypothesis is true but rejected, is defined as $P_{\text{FA}}(D_0, D_1, M, S) \equiv \mathbb{P}\big(M(D_0) \in S\big)$, and the probability of missed detection (type II error), when the null hypothesis is false but retained, is defined as $P_{\text{MD}}(D_0, D_1, M, S) \equiv \mathbb{P}\big(M(D_1) \in \bar{S}\big)$ where $\bar{S}$ is the complement of S. It turns out that imposing differential privacy conditions on a mechanism M is equivalent to restricting the probability of false alarm and missed detection from being simultaneously small. Wasserman and Zhu proved that $(\varepsilon, 0)$-differential privacy implies the conditions in Equation (1) for the special case when $\delta = 0$ [33, Theorem 2.4]. The same proof technique can be used to prove a similar result for a general $\delta \in [0, 1]$, and to prove that the conditions in Equation (1) imply (ε, δ)-differential privacy as well. We refer the reader to Section VI-B for a proof.

Theorem 2: For any $\varepsilon \geq 0$ and $\delta \in [0, 1]$, a database mechanism M is (ε, δ)-differentially private if and only if the following conditions are satisfied for all pairs of neighboring databases D_0 and D_1, and all rejection region $S \subseteq \mathcal{X}$:

$$P_{\text{FA}}(D_0, D_1, M, S) + e^{\varepsilon} P_{\text{MD}}(D_0, D_1, M, S) \geq 1 - \delta, \text{ and}$$
$$e^{\varepsilon} P_{\text{FA}}(D_0, D_1, M, S) + P_{\text{MD}}(D_0, D_1, M, S) \geq 1 - \delta. \quad (1)$$

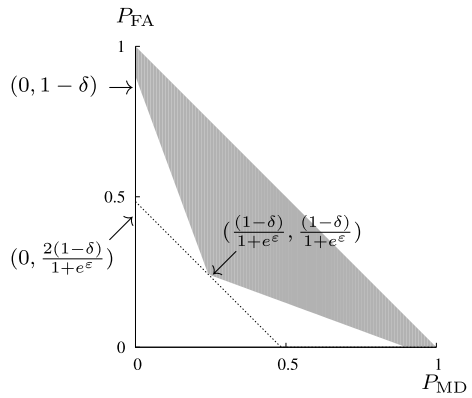

Fig. 1. Privacy region for (ε, δ)-differential privacy. Dotted line represents the solution of a maximization problem (20). For simplicity, we only show the privacy region below the line $P_{\text{FA}} + P_{\text{MD}} \leq 1$, since the whole region is symmetric w.r.t. the line $P_{\text{FA}} + P_{\text{MD}} = 1$.

This operational perspective relates the privacy parameters ε and δ to a set of conditions on probability of false alarm and missed detection. This shows that under differential privacy, it is impossible for both P_{MD} and P_{FA} to be simultaneously small. This operational interpretation of differential privacy suggests a graphical representation of differential privacy as illustrated in Figure 1. We define the *privacy region* for (ε, δ)-differential privacy as

$$\mathcal{R}(\varepsilon, \delta) \equiv \big\{(P_{\text{MD}}, P_{\text{FA}}) \,\big|\, P_{\text{FA}} + e^{\varepsilon} P_{\text{MD}} \geq 1 - \delta, \text{ and } e^{\varepsilon} P_{\text{FA}} + P_{\text{MD}} \geq 1 - \delta\big\}. \quad (2)$$

Similarly, we define the *privacy region* of a database access mechanism M with respect to two neighboring databases D_0 and D_1 as

$$\mathcal{R}(M, D_0, D_1) \equiv \text{conv}\Big(\big\{(P_{\text{MD}}(S), P_{\text{FA}}(S)) \,\big|\, \text{for all } S \subseteq \mathcal{X}\big\}\Big), \quad (3)$$

where $\text{conv}(\cdot)$ is the convex hull of a set and $\mathcal{X}$ is the alphabet of the privatized output, $P_{\text{MD}}(S) = P_{\text{MD}}(D_0, D_1, M, S)$, and $P_{\text{FA}}(S) = P_{\text{FA}}(D_0, D_1, M, S)$). Operationally, by taking the convex hull, the region includes the pairs of false alarm and missed detection probabilities achieved by soft decisions that might use internal randomness in the hypothesis testing rule. Precisely, let $\gamma : \mathcal{X} \to \{H_0, H_1\}$ be any randomized decision. For example, we can accept the null hypothesis with a certain probability p_1 if the output is in a set S_1 and probability p_2 if it is in another set S_2. In full generality, a decision rule γ can be fully described by a partition $\{S_i\}$ of the output space $\mathcal{X}$, and a corresponding accept probabilities $\{p_i\}$. The probabilities of false alarm and missed detection for a decision rule γ is defined as $P_{\text{FA}}(D_0, D_1, M, \gamma) \equiv \mathbb{P}(\gamma(M(D_0)) = H_1)$ and $P_{\text{MD}}(D_0, D_1, M, \gamma) \equiv \mathbb{P}(\gamma(M(D_1)) = H_0)$.

Remark 3: For all neighboring databases D_0 and D_1 and a database access mechanism M, the pair of false alarm and missed detection probabilities achieved by any decision rule γ is included in the privacy region:

$$(P_{\text{MD}}(D_0, D_1, M, \gamma), P_{\text{FA}}(D_0, D_1, M, \gamma)) \in \mathcal{R}(M, D_0, D_1),$$

for all decision rules γ.

Let $D_0 \sim D_1$ denote that the two databases are neighbors. The union over all neighboring databases defines the *privacy region of the mechanism*.

$$\mathcal{R}(M) \equiv \bigcup_{D_0 \sim D_1} \mathcal{R}(M, D_0, D_1).$$

The following corollary, which follows immediately from Theorem 2, gives a necessary and sufficient condition on the privacy region for (ε, δ)-differential privacy.

Corollary 4: A mechanism M is (ε, δ)-differentially private if and only if $\mathcal{R}(M) \subseteq \mathcal{R}(\varepsilon, \delta)$.

To illustrate the strengths of the graphical representation of differential privacy, we provide simpler proofs for some well-known results in differential privacy in Appendix A.

Consider two database access mechanisms $M(\cdot)$ and $M'(\cdot)$. Let X and Y denote the random outputs of mechanisms M and M' respectively. We say that M *dominates* M' if $M'(D)$ is conditionally independent of D given the outcome of $M(D)$. In other words, the database D, $X = M(D)$ and $Y = M'(D)$ form the following Markov chain: D–X–Y. We note that this holds for all distributions on D.

Theorem 5 (Data Processing Inequality for Differential Privacy): If a mechanism M dominates a mechanism M', then for all pairs of neighboring databases D_0 and D_1,

$$\mathcal{R}(M', D_0, D_1) \subseteq \mathcal{R}(M, D_0, D_1).$$

We refer the reader to Section VI-A for a proof. Together with Corollary 4, Theorem 5 recovers a well known result: differential privacy is preserved by postprocessing the output [10], [15], [33]. Perhaps surprisingly, the converse is also true.

Theorem 6 [4, Corollary of Theorem 10]: Fix a pair of neighboring databases D_0 and D_1, and let X and Y denote the random outputs of mechanisms M and M', respectively. If M and M' satisfy

$$\mathcal{R}(M', D_0, D_1) \subseteq \mathcal{R}(M, D_0, D_1),$$

then there exists a coupling of the random outputs X and Y such that they form a Markov chain D–X–Y where $D \in \{D_0, D_1\}$.

In other words, when the privacy region of M' is included in M, there exists a stochastic transformation T that operates on X to produce a random output that has the same marginal distribution as Y conditioned on the database D. We can consider this mechanism T as a privatization mechanism that takes a (privatized) output X and provides even further privatization. The above theorem was proved in [4, Corollary of Theorem 10] in the context of comparing two experiments, where a *statistical experiment* denotes a mechanism in the context of differential privacy.

III. Composition of Differentially Private Mechanisms

In this section, we address how differential privacy guarantees compose: when accessing databases multiple times via differentially private mechanisms, each of which having its own privacy guarantees. Precisely, we address the following fundamental question: how much privacy can be guaranteed after multiple database accesses? To formally define composition, we consider the following scenario known as the 'composition experiment', proposed in [16].

A composition experiment takes as input a parameter $b \in \{0, 1\}$, and an adversary $\mathcal{A}$. From the hypothesis testing perspective proposed in the previous section, b can be interpreted as the hypothesis: null hypothesis for $b = 0$ and alternative hypothesis for $b = 1$. At each time i, a database $D^{i,b}$ is accessed depending on b. For example, one includes a particular individual and another does not. For example, $D^{1,0}$ could be medical records including a particular individual and $D^{1,1}$ does not include the person, and $D^{2,0}$ could be voter registration database with the same person present and $D^{2,1}$ with the person absent. An adversary $\mathcal{A}$ is trying to figure out whether or not a particular individual is in the database by testing the hypotheses on the output of k sequential database accesses via differentially private mechanisms. In full generality, we allow the adversary to have full control over which pair of databases to access, which query to ask, and which mechanism to be used at each repeated access. Further, the adversary is free to make these choices adaptively based on the previous outcomes. The only restrictions are: (a) the differentially private mechanisms belong to a family $\mathcal{M}$ (e.g., the family of all (ε, δ)-differentially private mechanisms), (b) the internal randomness of the mechanisms are independent at each repeated access, and (c) that the hypothesis b is not known to the adversary.

Algorithm 1 COMPOSE$(\mathcal{A}, \mathcal{M}, k, b)$

Input: $\mathcal{A}$, $\mathcal{M}$, k, b
Output: V^b
  Choose internal randomness R for the adversary $\mathcal{A}$
  for $i = 1$ to k **do**
    $\mathcal{A}$ requests $(D^{i,0}, D^{i,1}, q_i, M_i)$ for some $M_i \in \mathcal{M}$;
    $\mathcal{A}$ receives $y_i = M_i(D^{i,b}, q_i)$;
  end for
  Output the view of the adversary $V^b = (R, Y_1^b, \ldots, Y_k^b)$.

The outcome of this k-fold composition experiment is the *view of the adversary* $\mathcal{A}$: $V^b \equiv (R, Y_1^b, \ldots, Y_k^b)$, which is the sequence of random outcomes $Y_1^b, \ldots, Y_k^b$, and the outcome R of any internal randomness of $\mathcal{A}$.

A. Optimal Privacy Region Under Composition

We would like to characterize how much privacy degrades after a k-fold composition experiment. It is known that the privacy degrades under composition by at most the 'sum' of the differential privacy parameters of each access.

Theorem 7 ([10]–[12], [16]): For any $\varepsilon > 0$ and $\delta \in [0, 1]$, the class of (ε, δ)-differentially private mechanisms satisfies $(k\varepsilon, k\delta)$-differential privacy under k-fold adaptive composition.

In general, one can show that if M_i is $(\varepsilon_i, \delta_i)$-differentially private, then the composition satisfies $(\sum_{i \in [k]} \varepsilon_i, \sum_{i \in [k]} \delta_i)$-differential privacy. If we do not allow for any slack in the δ, this bound cannot be tightened. Precisely, there are examples of mechanisms which under k-fold composition violate

$(\varepsilon, \sum_{i\in[k]}\delta_i)$-differential privacy for any $\varepsilon < \sum_{i\in[k]}\varepsilon_i$. We can prove this by providing a set S such that the privacy condition is met with equality: $\mathbb{P}(V^0 \in S) = e^{\sum_{i\in[k]}\varepsilon_i}\mathbb{P}(V^1 \in S) + \sum_{i\in[k]}\delta_i$. However, if we allow for a slightly larger value of δ, then Dwork et al. showed in [16] that one can gain a significantly higher privacy guarantee in terms of ε.

Theorem 8 ([16, Theorem III.3]): For any $\varepsilon > 0$, $\delta \in [0,1]$, and $\tilde{\delta} \in (0,1]$, the class of (ε,δ)-differentially private mechanisms satisfies $(\tilde{\varepsilon}_{\tilde{\delta}}, k\delta+\tilde{\delta})$-differential privacy under k-fold adaptive composition, for

$$\tilde{\varepsilon}_{\tilde{\delta}} = k\varepsilon(e^{\varepsilon}-1) + \varepsilon\sqrt{2k\log(1/\tilde{\delta})}. \tag{4}$$

By allowing a slack of $\tilde{\delta} > 0$, one can get a higher privacy of $\tilde{\varepsilon}_{\tilde{\delta}} = O(k\varepsilon^2 + \sqrt{k\varepsilon^2})$, which is significantly smaller than $k\varepsilon$. This is the best known guarantee so far, and has been used whenever one requires a privacy guarantee under composition (e.g. [2], [16], [24]). However, the important question of optimality has remained open. Namely, is there a composition of mechanisms where the above privacy guarantee is tight? In other words, is it possible to get a tighter bound on differential privacy under composition?

We give a complete answer to this fundamental question in the following theorems. We prove a tighter bound on the privacy guarantee under composition. Further, we also prove the achievability of the privacy guarantee: we provide a set of mechanisms such that the privacy region under k-fold composition is exactly the region defined by the conditions in (5). Hence, this bound on the privacy region is tight and cannot be improved upon.

Theorem 9: For any $\varepsilon \geq 0$ and $\delta \in [0,1]$, the class of (ε,δ)-differentially private mechanisms satisfies

$$\big((k-2i)\varepsilon,\ 1-(1-\delta)^k(1-\delta_i)\big)\text{-differential privacy} \tag{5}$$

under k-fold adaptive composition, for all $i = \{0,1,\ldots,\lfloor k/2\rfloor\}$, where

$$\delta_i = \frac{\sum_{\ell=0}^{i-1}\binom{k}{\ell}\left(e^{(k-\ell)\varepsilon} - e^{(k-2i+\ell)\varepsilon}\right)}{(1+e^{\varepsilon})^k}. \tag{6}$$

Hence, the privacy region of k-fold composition is an intersection of k regions, each of which is $((k-2i)\varepsilon, 1-(1-\delta)^k(1-\delta_i))$-differentially private: $\mathcal{R}(\{(k-2i)\varepsilon, 1-(1-\delta)^k(1-\delta_i)\}_{i\in[k/2]}) \equiv \bigcap_{i=0}^{\lfloor\frac{k}{2}\rfloor}\mathcal{R}((k-2i)\varepsilon, 1-(1-\delta)^k(1-\delta_i))$. We prove this result in Section IV by constructing an explicit mechanism that achieves this region under composition. Hence, this bound on the privacy region is tight, and gives the exact description of how privacy degrades under k-fold adaptive composition. This settles the question that was left open in [10]–[12], [16] by providing, for the first time, the fundamental limit of composition and proving a matching mechanism with the worst-case privacy degradation.

To prove the optimality of our main result in Theorem 9, namely that it is impossible to have a privacy worse than (5), we rely on the operational interpretation of the privacy as hypothesis testing. To this end, we use the new analysis tools (Theorem 5 and Theorem 6) provided in the previous section. Figure 2 illustrates how much the privacy region of Theorem 9 degrades as we increase the number of composition k. Figure 3 provides a comparison of the three privacy guarantees in Theorems 7, 8 and 9 for 30-fold composition of $(0.1, 0.001)$-differentially private mechanisms.

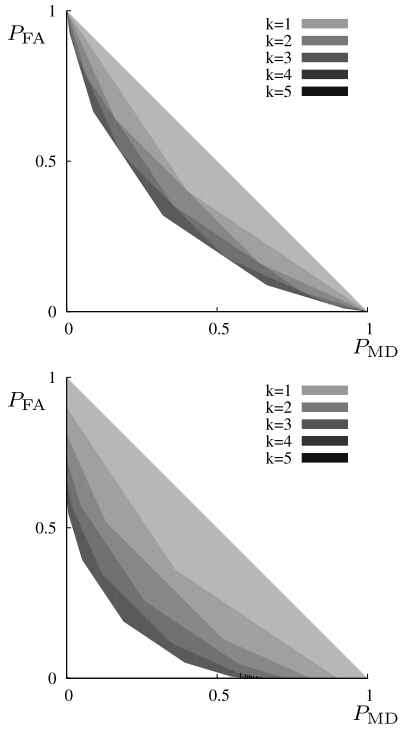

Fig. 2. Privacy region $\mathcal{R}(\{(k-2i)\varepsilon, \delta_i\})$ for the class of $(\varepsilon, 0)$-differentially private mechanisms (top) and (ε,δ)-differentially private mechanisms (bottom) under k-fold adaptive composition.

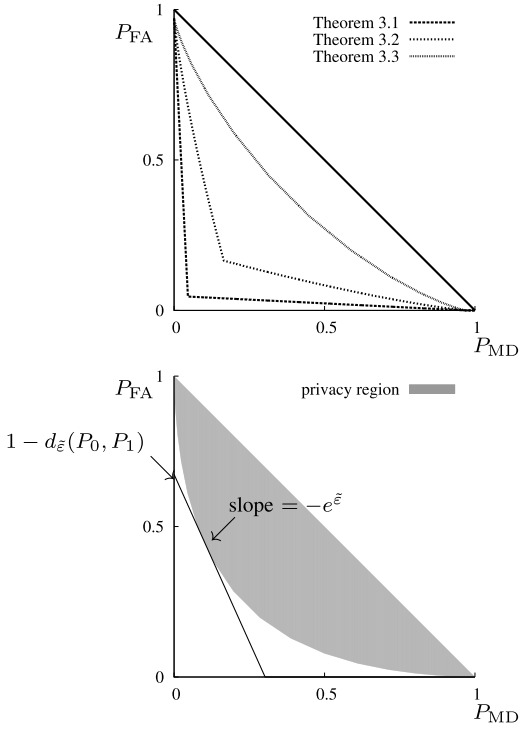

Fig. 3. Theorem 9 provides the tightest bound (top). Given a mechanism M, the privacy region can be completely described by its boundary, which is represented by a set of tangent lines of the form $P_{\rm FA} = -e^{\tilde{\varepsilon}}P_{\rm MD} + 1 - d_{\tilde{\varepsilon}}(P_0, P_1)$ (bottom).

Smaller region gives a tighter bound, since it guarantees higher privacy.

B. Simplified Privacy Region Under Composition

In many applications of the composition theorems, a closed form expression of the composition privacy guarantee is required. The privacy guarantee in (5) is tight, but can be difficult to evaluate. The next theorem provides a simpler expression which is an outer bound on the exact region described in (5). Compared to (4), the privacy guarantee is significantly improved from $\tilde{\varepsilon}_{\tilde{\delta}} = O\Big(k\varepsilon^2 + \sqrt{k\varepsilon^2 \log(1/\tilde{\delta})}\Big)$ to $\tilde{\varepsilon}_{\tilde{\delta}} = O\Big(k\varepsilon^2 + \min\big\{\sqrt{k\varepsilon^2 \log(1/\tilde{\delta})}, \varepsilon \log(\varepsilon/\tilde{\delta})\big\}\Big)$, especially when composing a large number k of interactive queries. Further, the δ-approximate differential privacy degradation of $(1-(1-\delta)^k(1-\tilde{\delta}))$ is also strictly smaller than the previous $(k\delta+\tilde{\delta})$. We discuss the significance of this improvement in the next section using examples from the existing differential privacy literature.

Theorem 10: For any $\varepsilon > 0$, $\delta \in [0,1]$, and $\tilde{\delta} \in [0,1]$, the class of (ε,δ)-differentially private mechanisms satisfies $\big(\tilde{\varepsilon}_{\tilde{\delta}}, 1-(1-\delta)^k(1-\tilde{\delta})\big)$-differential privacy under k-fold adaptive composition, for

$$\tilde{\varepsilon}_{\tilde{\delta}} = \min\Big\{ k\varepsilon \ , \ \frac{(e^\varepsilon-1)\varepsilon k}{e^\varepsilon+1} + \varepsilon\sqrt{2k\,\log\Big(e + \frac{\sqrt{k\varepsilon^2}}{\tilde{\delta}}\Big)} \ , \ \frac{(e^\varepsilon-1)\varepsilon k}{e^\varepsilon+1} + \varepsilon\sqrt{2k\,\log\Big(\frac{1}{\tilde{\delta}}\Big)}\Big\}. \tag{7}$$

This bound can be further simplified as

$$\tilde{\varepsilon}_{\tilde{\delta}} \le \min\Big\{ k\varepsilon, k\varepsilon^2 + \varepsilon\sqrt{2k\log\big(e + (\sqrt{k\varepsilon^2}/\tilde{\delta})\big)}, \ k\varepsilon^2 + \varepsilon\sqrt{2k\log(1/\tilde{\delta})}\Big\}.$$

A proof is provided in Section VI-D. This privacy guarantee improves over the existing result of Theorem 8 when $\tilde{\delta} = \Theta(\sqrt{k\varepsilon^2})$. Typical regime of interest is the high-privacy regime for composition privacy guarantee, i.e. when $\sqrt{k\varepsilon^2} \ll 1$. The above theorem suggests that we only need the extra slack of approximate privacy $\tilde{\delta}$ of order $\sqrt{k\varepsilon^2}$.

C. Composition Theorem for Heterogeneous Mechanisms

So far, we considered homogeneous mechanisms, where all mechanisms are (ε,δ)-differentially private. Our analysis readily extends to heterogeneous mechanisms, where the ℓ-th query satisfies $(\varepsilon_\ell,\delta_\ell)$-differential privacy (we refer to such mechanisms as $(\varepsilon_\ell,\delta_\ell)$-differentially private mechanisms).

Theorem 11: For any $\varepsilon_\ell > 0$, $\delta_\ell \in [0,1]$ for $\ell \in \{1,\dots,k\}$, and $\tilde{\delta} \in [0,1]$, the class of $(\varepsilon_\ell,\delta_\ell)$-differentially private mechanisms satisfies $\big(\tilde{\varepsilon}_{\tilde{\delta}}, 1-(1-\tilde{\delta})\prod_{\ell=1}^k(1-\delta_\ell)\big)$-differential privacy under k-fold adaptive composition, for

$$\tilde{\varepsilon}_{\tilde{\delta}} = \min\left\{ \sum_{\ell=1}^k \varepsilon_\ell \ , \ \sum_{\ell=1}^k \frac{(e^{\varepsilon_\ell}-1)\varepsilon_\ell}{e^{\varepsilon_\ell}+1} + \sqrt{\sum_{\ell=1}^k 2\,\varepsilon_\ell^2\,\log\Big(\frac{1}{\tilde{\delta}}\Big)} \ , \ \sum_{\ell=1}^k \frac{(e^{\varepsilon_\ell}-1)\varepsilon_\ell}{e^{\varepsilon_\ell}+1} + \sqrt{\sum_{\ell=1}^k 2\,\varepsilon_\ell^2\,\log\Big(e + \frac{\sqrt{\sum_{\ell=1}^k \varepsilon_\ell^2}}{\tilde{\delta}}\Big)}\right\}. \tag{8}$$

This tells us that the ε_ℓ's *sum up under composition*: whenever we have $k\varepsilon$ or $k\varepsilon^2$ in (7) we can replace it by the summation to get the general result for heterogeneous case.

IV. Proof of Theorem 9

We first propose a simple mechanism and prove that it dominates over all (ε,δ)-differentially private mechanisms. Analyzing the privacy region achieved by the k-fold composition of the proposed mechanism, we get a bound on the privacy region under the adaptive composition. This gives an exact characterization of privacy under composition, since we show both converse and achievability. We prove that no other family of mechanisms can achieve 'more degraded' privacy (converse), and that there is a mechanism that we propose which achieves the privacy region (achievability).

A. Achievability

We propose the following simple mechanism $\tilde{M}$. Under the null hypothesis ($b=0$), the outputs $\{X^{i,0} = \tilde{M}(D^{i,0}, q_i)\}_{i\in[k]}$ are independent and identically distributed to a discrete random variable $\tilde{X}_0 \sim \tilde{P}_0(\cdot)$, where

$$\tilde{P}_0(x) \ = \ \mathbb{P}(\tilde{X}_0 = x) \ \equiv \ \begin{cases} \delta & \text{for } x=0, \\ \frac{(1-\delta)\,e^\varepsilon}{1+e^\varepsilon} & \text{for } x=1, \\ \frac{1-\delta}{1+e^\varepsilon} & \text{for } x=2, \\ 0 & \text{for } x=3. \end{cases} \tag{9}$$

Under the alternative hypothesis ($b=1$), the outputs $\{X^{i,1} = \tilde{M}(D^{i,1}, q_i)\}_{i\in[k]}$ are independent and identically distributed to a discrete random variable $\tilde{X}_1 \sim \tilde{P}_1(\cdot)$, where

$$\tilde{P}_1(x) \ = \ \mathbb{P}(\tilde{X}_1 = x) \ \equiv \ \begin{cases} 0 & \text{for } x=0, \\ \frac{1-\delta}{1+e^\varepsilon} & \text{for } x=1, \\ \frac{(1-\delta)\,e^\varepsilon}{1+e^\varepsilon} & \text{for } x=2, \\ \delta & \text{for } x=3. \end{cases} \tag{10}$$

In particular, the output of this mechanism does not depend on the database $D^{i,b}$ or the query q_i, and only depends on the hypothesis b. The privacy region of a single access to this mechanism is $\mathcal{R}(\varepsilon,\delta)$ in Figure 1. Hence, by Theorem 6, all (ε,δ)-differentially private mechanisms are dominated by this mechanism.

In general, the privacy region $\mathcal{R}(M, D_0, D_1)$ of any mechanism can be represented by an intersection of multiple $\{(\tilde{\varepsilon}_j, \tilde{\delta}_j)\}$ privacy regions. For a mechanism M, we can compute the $(\tilde{\varepsilon}_j, \tilde{\delta}_j)$ pairs representing the privacy region as follows. Given a null hypothesis database D_0, an alternative hypothesis database D_1, and a mechanism M whose output

space is $\mathcal{X}$, let P_0 and P_1 denote the probability density function of the outputs $M(D_0)$ and $M(D_1)$, respectively. To simplify notations we assume that P_0 and P_1 are symmetric, i.e. there exists a permutation π over $\mathcal{X}$ such that $P_0(x) = P_1(\pi(x))$ and $P_1(x) = P_0(\pi(x))$. This ensures that we get a symmetric privacy region.

The privacy region $\mathcal{R}(M, D_0, D_1)$ can be described by its boundaries. Since it is a convex set, a tangent line on the boundary with slope $-e^{\tilde{\varepsilon}_j}$ can be represented by the smallest $\tilde{\delta}_j$ such that

$$P_{\text{FA}} \geq -e^{\tilde{\varepsilon}_j} P_{\text{MD}} + 1 - \tilde{\delta}_j, \tag{11}$$

for all rejection sets (cf. Figure 3). Letting S denote the complement of a rejection set, such that $P_{\text{FA}} = 1 - P_0(S)$ and $P_{\text{MD}} = P_1(S)$, the minimum shift $\tilde{\delta}_j$ that still ensures that the privacy region is above the line (11) is defined as $\tilde{\delta}_j = d_{\tilde{\varepsilon}_j}(P_0, P_1)$ where

$$d_{\tilde{\varepsilon}}\ (P_0, P_1) \equiv \max_{S \subseteq \mathcal{X}} \Big\{ P_0(S) - e^{\tilde{\varepsilon}}\ P_1(S) \Big\}.$$

The privacy region of a mechanism is completely described by the set of slopes and shifts, $\{(\tilde{\varepsilon}_j, \tilde{\delta}_j) : \tilde{\varepsilon}_j \in E \text{ and } \tilde{\delta}_j = d_{\tilde{\varepsilon}_j}(P_0, P_1)\}$, where

$$E \equiv \{\, 0 \leq \tilde{\varepsilon} < \infty :\ P_0(x) = e^{\tilde{\varepsilon}}\ P_1(x) \text{ for some } x \in \mathcal{X}\}.$$

Any $\tilde{\varepsilon} \notin E$ does not contribute to the boundary of the privacy region. For the above example distributions $\tilde{P}_0$ and $\tilde{P}_1$, $E = \{\varepsilon\}$ and $d_\varepsilon(\tilde{P}_0, \tilde{P}_1) = \delta$.

Remark 12: For a database access mechanism M over a output space $\mathcal{X}$ and a pair of neighboring databases D_0 and D_1, let P_0 and P_1 denote the probability density function for random variables $M(D_0)$ and $M(D_1)$ respectively. Assume there exists a permutation π over $\mathcal{X}$ such that $P_0(x) = P_1(\pi(x))$. Then, the privacy region is

$$\mathcal{R}(\,M, D_0, D_1\,) = \bigcap_{\tilde{\varepsilon} \in E} \mathcal{R}\big(\,\tilde{\varepsilon}, d_{\tilde{\varepsilon}}(P_0, P_1)\,\big),$$

where $\mathcal{R}(M, D, D')$ and $\mathcal{R}(\tilde{\varepsilon}, \tilde{\delta})$ are defined as in (3) *and* (2)*.*

The symmetry assumption is to simplify notations, and the analysis can be easily generalized to deal with non-symmetric distributions.

Now consider a k-fold composition experiment, where at each sequential access $\tilde{M}_i$, we receive a random output $X^{i,b}$ independent and identically distributed as $\tilde{X}_b$. We can explicitly characterize the distribution of k-fold composition of the outcomes: $\mathbb{P}(X^{1,b} = x_1, \ldots, X^{k,b} = x_k) = \prod_{x=1}^{k} \tilde{P}_b(x_i)$. It follows form the structure of these two discrete distributions that, $E = \{e^{(k-2\lfloor k/2 \rfloor)\varepsilon}, e^{(k+2-2\lfloor k/2 \rfloor)\varepsilon}, \ldots, e^{(k-2)\varepsilon}, e^{k\varepsilon}\}$. After some algebra, it also follows that

$$\begin{aligned} &d_{(k-2i)\varepsilon}\big(\,(\tilde{P}_0)^k, (\tilde{P}_1)^k\,\big) \\ &\quad = 1 - (1-\delta)^k \\ &\qquad + (1-\delta)^k \frac{\sum_{\ell=0}^{i-1} \binom{k}{\ell} \big(e^{\varepsilon(k-\ell)} - e^{\varepsilon(k-2i+\ell)}\big)}{(1+e^{\varepsilon})^k}. \end{aligned}$$

for $i \in \{0, \ldots, \lfloor k/2 \rfloor\}$. From Remark 12, it follows that the privacy region is $\mathcal{R}(\{\varepsilon_i, \delta_i\}) = \bigcap_{i=0}^{\lfloor k/2 \rfloor} \mathcal{R}\big(\varepsilon_i, \delta_i\big)$, where $\varepsilon_i = (k-2i)\varepsilon$ and δ_i's are defined as in (6). Figure 2 shows this privacy region for $k = 1, \ldots, 5$, $\varepsilon = 0.4$, and $\delta = 0$ and $\delta = 0.1$.

B. Converse

We will now prove that this region is the largest region achievable under k-fold adaptive composition of any (ε, δ)-differentially private mechanisms.

From Corollary 4, any mechanism whose privacy region is included in $\mathcal{R}(\{\varepsilon_i, \delta_i\})$ satisfies $(\tilde{\varepsilon}, \tilde{\delta})$-differential privacy. We are left to prove that for the family of all (ε, δ)-differentially private mechanisms, the privacy region of the k-fold composition experiment is included inside $\mathcal{R}(\{\varepsilon_i, \delta_i\})$. To this end, consider the following composition experiment, which reproduces the *view of the adversary* from the original composition experiment.

At each time step i, we generate a random variable $X^{i,b}$ distributed as $\tilde{X}_b$ independent of any other random events, and call this the output of a database access mechanism $\tilde{M}_i$ such that $\tilde{M}_i(D^{i,b}, q_i) = X^{i,b}$. Since, $X^{i,b}$ only depends on b, and is independent of the actual database or the query, we use $\tilde{M}_i(b)$ to denote this outcome.

We know that $\tilde{M}_i(b)$ has privacy region $\mathcal{R}(\varepsilon, \delta)$ for any choices of $D^{i,0}$, $D^{i,1}$ and q_i. Now consider the mechanism M_i from the original experiment. Since it is (ε, δ)-differentially private, we know from Theorem 2 that $\mathcal{R}(M_i, D^{i,0}, D^{i,1}) \subseteq \mathcal{R}(\varepsilon, \delta)$ for any choice of neighboring databases $D^{i,0}$, $D^{i,1}$. Hence, from the converse of data processing inequality (Theorem 6), we know that there exists a mechanism T_i that takes as input $X^{i,b}$ and produces an output $Y^{i,b}$ which is distributed as $M_i(D^{i,b}, q_i)$ for all $b \in \{0, 1\}$. Hence, $Y^{i,b}$ is independent of the past conditioned on $X^{i,b}, D^{i,0}, D^{i,1}, q_i, M_i$. Precisely we have the following Markov chain:

$$\begin{aligned} (b, R, \{X^{\ell,b}, D^{\ell,0}, D^{\ell,1}, q_\ell, M_\ell\}_{\ell \in [i-1]}) \\ -(X^{i,b}, D^{i,0}, D^{i,1}, q_i, M_i)-Y^{i,b}, \end{aligned}$$

where R is any internal randomness of the adversary $\mathcal{A}$. Since, (X, Y)–Z–W implies X–(Y, Z)–W, we have

$$b\text{–}(R, \{X^{\ell,b}, D^{\ell,0}, D^{\ell,1}, q_\ell, M_\ell\}_{\ell \in [i]})\text{–}Y^{i,b}.$$

Notice that if we know R and the outcomes $\{Y^{\ell,b}\}_{\ell \in [i]}$, then we can reproduce the original experiment until time i. This is because the choices of $D^{i,0}, D^{i,1}, q_i, M_i$ are exactly specified by R and $\{Y^{\ell,b}\}_{\ell \in [i]}$. Hence, we can simplify the Markov chain as

$$b\text{–}(R, X^{i,b}, \{X^{\ell,b}, Y^{\ell,b}\}_{\ell \in [i-1]})\text{–}Y^{i,b}. \tag{12}$$

Further, since $X^{i,b}$ is independent of the past conditioned on b, we have

$$X^{i,b}\text{–}b\text{–}(R, \{X^{\ell,b}, Y^{\ell,b}\}_{\ell \in [i-1]}). \tag{13}$$

It follows that

$$\begin{aligned} &\mathbb{P}(b, r, x_1 \ldots, x_k, y_1, \ldots, y_k) \\ &\quad = \mathbb{P}(b, r, x_1, \ldots, x_k, y_1, \ldots, y_{k-1}) \\ &\qquad \times \mathbb{P}(y_k | r, x_1, \ldots, x_k, y_1, \ldots, y_{k-1}) \\ &\quad = \mathbb{P}(b, r, x_1, \ldots, x_{k-1}, y_1, \ldots, y_{k-1}) \mathbb{P}(x_k | b) \\ &\qquad \times \mathbb{P}(y_k | r, x_1, \ldots, x_k, y_1, \ldots, y_{k-1}), \end{aligned}$$

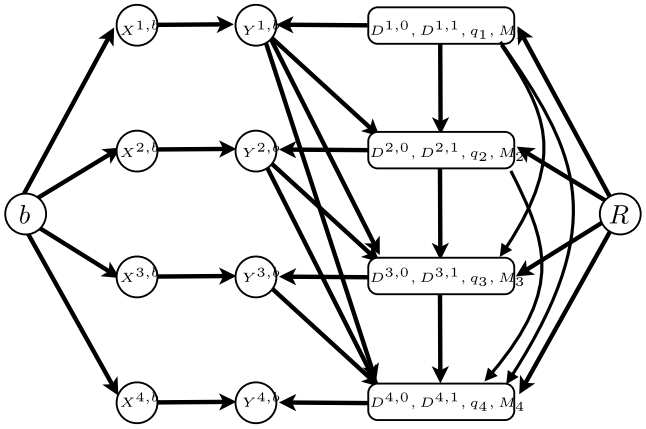

Fig. 4. Bayesian network representation of the composition experiment. The subset of nodes $(X^{1,b}, X^{2,b}, X^{3,b}, X^{4,b})$ d-separates node b from the rest of the network.

where we used (12) in the first equality and (13) in the second. By induction, we get a decomposition

$$\begin{aligned}
&\mathbb{P}(b, r, x_1, \ldots, x_k, y_1, \ldots, y_k) \\
&\quad = \mathbb{P}(b, r) \prod_{i=1}^{k} \mathbb{P}(x_i|b) \prod_{i=1}^{k} \mathbb{P}(y_i|r, x_1, \ldots, x_i, y_1, \ldots, y_{i-1}) \\
&\quad = \mathbb{P}(b, r, x_1, \ldots, x_k)\mathbb{P}(y_1, \ldots, y_k|r, x_1, \ldots, x_k) \\
&\quad = \mathbb{P}(b|r, x_1, \ldots, x_k)\, \mathbb{P}(y_1, \ldots, y_k, r, x_1, \ldots, x_k).
\end{aligned}$$

From the construction of the experiment, it also follows that the internal randomness R is independent of the hypothesis b and the outcomes $X^{i,b}$'s: $\mathbb{P}(b|r, x_1, \ldots, x_k) = \mathbb{P}(b|x_1, \ldots, x_k)$. Then, marginalizing over R, we get $\mathbb{P}(b, x_1, \ldots, x_k, y_1, \ldots, y_k) = \mathbb{P}(b|x_1, \ldots, x_k)\, \mathbb{P}(y_1, \ldots, y_k, x_1, \ldots, x_k)$. This implies the following Markov chain:

$$b\text{–}(\{X^{i,b}\}_{i\in[k]})\text{–}(\{Y^{i,b}\}_{i\in[k]}), \tag{14}$$

and it follows that a set of mechanisms $(M_1, \ldots, M_k)$ dominates $(\tilde{M}_1, \ldots, \tilde{M}_k)$ for two databases $\{D^{i,0}\}_{i\in[k]}$ and $\{D^{i,1}\}_{i\in[k]}$. By the data processing inequality for differential privacy (Theorem 5), this implies that

$$\begin{aligned}
&\mathcal{R}\big(\{M_i\}_{i\in[k]}, \{D^{i,0}\}_{i\in[k]}, \{D^{i,1}\}_{i\in[k]}\big) \\
&\quad \subseteq \mathcal{R}\big(\{\tilde{M}_i\}_{i\in[k]}, \{D^{i,0}\}_{i\in[k]}, \{D^{i,1}\}_{i\in[k]}\big) \;=\; \mathcal{R}\big(\{\varepsilon_i, \delta_i\}\big).
\end{aligned}$$

This finishes the proof of the desired claim.

Alternatively, one can prove (14), using the probabilistic graphical model shown in Figure 4. Precisely, the following Bayesian network describes the dependencies among various random quantities of the experiment described above. Since the set of nodes $(X^{1,b}, X^{2,b}, X^{3,b}, X^{4,b})$ d-separates node b from the rest of the bayesian network, it follows immediately from the Markov property of this Bayesian network that (14) is true (cf. [27]).

V. Applications of the Optimal Composition Theorem

We now apply the results of the previous section to analyze the utility of a complex privacy mechanism that is composed of k sub-mechanisms, each with an $(\varepsilon_0, \delta_0)$-differential privacy guarantee. To ensure an overall of (ε, δ)-differential privacy, we choose $\varepsilon_0 = \varepsilon/(2\sqrt{k\log(e + \varepsilon/\delta)})$ and $\delta_0 = \delta/2k$. The composition theorem presented in the previous section guarantees the desired overall privacy. For each application we study, we first fix k differentially private sub-mechanisms, and then calculate the *utility* of the complex mechanism. Following this recipe, we provide a sufficient condition on the variance of noise adding mechanisms. Our analysis shows that one requires smaller variance than what was previously believed, especially in the regime where $\varepsilon = \Theta(\delta)$. Further, we show that a variety of known mechanisms achieve the desired privacy under composition with the *same* level of variance. Applying this analysis to known mechanisms for cut queries of a graph, we show that in the regime where $\varepsilon = \Theta(\delta)$, one can achieve the desired privacy under composition with improved utility.

For count queries with sensitivity one, the geometric noise adding mechanism is known to be universally optimal in a general cost minimization framework (Bayesian setting in [18] and worst-case setting in [20]). Here we provide a new interpretation of the geometric noise adding mechanism as an optimal mechanism under *composition* for counting queries. Indeed, in the course of proving Theorem 9, we show that a family of mechanisms are optimal under composition, in the sense that they achieve the largest privacy region among k-fold compositions of any $(\varepsilon_i, \delta_i)$-differentially private mechanisms. Larger region under composition implies that one can achieve smaller error rates, while ensuring the same level of privacy at each step of the composition. In section V-B, we show that the geometric mechanism is one such mechanism, thus providing a new interpretation to the optimality of the geometric mechanisms.

A. Variance of Noise Adding Mechanisms Under Composition

Consider a real-valued database query $q : \mathcal{D} \to \mathbb{R}$. The *sensitivity* of q is defined as the maximum absolute difference of the output between any two neighboring databases:

$$\Delta \equiv \max_{D\sim D'} |q(D) - q(D')|,$$

where $\sim$ indicates that the pair of databases are neighbors. The output of q is usually privatized via the addition of random noise the variance of which grows with sensitivity of the query and the desired level of privacy. One of the most popular noise adding mechanisms is the *Laplacian mechanism*, which adds Laplacian noise to real-valued query outputs. When the sensitivity is Δ, one can achieve $(\varepsilon_0, 0)$-differential privacy with the choice of the distribution $\mathrm{Lap}(\varepsilon_0/\Delta) = (\varepsilon_0/2\Delta)e^{-\varepsilon_0|x|/\Delta}$. The resulting variance of the noise is $2\Delta^2/\varepsilon_0^2$. However, the *Laplacian mechanism* has been largely ignored in the context of query compositions. When composing real-valued queries, the *Gaussian mechanism* is a popular choice [2], [3], [13], [14], [24]. To ensure (ε, δ)-differential privacy under k-fold composition, it is sufficient to add Gaussian noise with variance $O(k\Delta^2 \log(1/\delta)/\varepsilon^2)$ to each query output.

In this section, we show that there is nothing special about the Gaussian mechanism. Indeed, we prove that the *Laplacian*

mechanism or the *staircase mechanism* (introduced in [20]) can achieve the same level of privacy under composition with the same variance.

We can use Theorem 10 to find how much noise we need to add to each query output in order to ensure (ε, δ)-differential privacy under k-fold composition. We know that if each query output is $(\varepsilon_0, \delta_0)$-differentially private, then the composed outputs satisfy $(k\varepsilon_0^2+\sqrt{2k\varepsilon_0^2\log(e+\sqrt{k\varepsilon_0^2}/\tilde{\delta})}, k\delta_0+\tilde{\delta})$-differential privacy. With the choice of $\delta_0 = \delta/2k$, $\tilde{\delta} = \delta/\sqrt{2}$, and $\varepsilon_0^2 = \varepsilon^2/4k\log(e+(\varepsilon/\delta))$, this ensures that the target privacy of (ε, δ) is satisfied under k-fold composition as described in the following corollary.

Corollary 13: For any ε and $\delta \in (0, 1]$, if the database access mechanism satisfies $(\sqrt{\varepsilon^2/4k\log(e+(\varepsilon/\delta))}, \delta/2k)$-differential privacy on each query output, then it satisfies (ε, δ)-differential privacy under k-fold composition.

The above corollary implies a sufficient condition on the variance of the Laplacian mechanism to ensure privacy under composition.

Corollary 14: For real-valued queries with sensitivity $\Delta > 0$, the mechanism that adds Laplacian noise with variance $\big(8k\Delta^2\log\big(e+(\varepsilon/\delta)\big)/\varepsilon^2\big)$ satisfies (ε, δ)-differential privacy under k-fold adaptive composition for any ε and $\delta \in (0, 1]$.

In terms of variance-privacy trade-off for real-valued queries, the optimal noise-adding mechanism known as the *staircase mechanism* was introduced in [20]. The probability density function of this noise is piecewise constant, and the probability density on the pieces decays geometrically. It is shown in [21] that with variance of $O(\min\{1/\varepsilon^2, 1/\delta^2\})$, the staircase mechanism achieves (ε, δ)-differential privacy. Corollary 13 implies that with variance $O\big(k\Delta^2\log(e+\varepsilon/\delta)/\varepsilon^2\big)$, the staircase mechanism satisfies (ε, δ)-differential privacy under k-fold composition.

Another popular mechanism known as the *Gaussian mechanism* privatizes each query output by adding a Gaussian noise with variance σ^2. It is not difficult to show that when the sensitivity of the query is Δ, with a choice of $\sigma^2 \geq 2\Delta^2\log(2/\delta_0)/\varepsilon_0^2$, the Gaussian mechanism satisfies $(\varepsilon_0, \delta_0)$-differential privacy (e.g. [10]). The above corollary implies that the Gaussian mechanism with variance $O(k\Delta^2\log(1/\delta)\log(e+(\varepsilon/\delta))/\varepsilon^2)$ ensures (ε, δ)-differential privacy under k-fold composition. However, we can get a tighter sufficient condition by directly analyzing how Gaussian mechanisms compose, and the proof is provided in Appendix A.

Theorem 15: For real-valued queries with sensitivity $\Delta > 0$, the mechanism that adds Gaussian noise with variance $\big(8k\Delta^2\log\big(e+(\varepsilon/\delta)\big)/\varepsilon^2\big)$ satisfies (ε, δ)-differential privacy under k-fold adaptive composition for any $\varepsilon > 0$ and $\delta \in (0, 1]$.

It is known that it is sufficient to add i.i.d. Gaussian noise with variance $O(k\Delta^2\log(1/\delta)/\varepsilon^2)$ to ensure (ε, δ)-differential privacy under k-fold composition (e.g. [25, Theorem 2.7]). The above theorem shows that when $\delta = \Theta(\varepsilon)$, one can achieve the same privacy with smaller variance by a factor of $\log(1/\delta)$.

B. Geometric Noise Adding Mechanism Under Composition

In this section, we consider integer valued queries $q : \mathcal{D} \to \mathbb{Z}$ with sensitivity one, also called *count queries*. Such queries are common in practice, e.g. "How many individuals have income less than $100,000?". The presence or absence of an individual record changes the output by at most one. Hence, the sensitivity of such queries is equal to one. Count queries are well studied in differential privacy [3], [6], [13], [14] and they provide a primitive for constructing more complex queries [3].

The *geometric mechanism* is a discrete variant of the popular *Laplacian mechanism*. For integer-valued queries with sensitivity one, the mechanism adds a noise distributed according to a double-sided geometric distribution: $p(k) = \big((e^\varepsilon-1)/(e^\varepsilon+1)\big)e^{-\varepsilon|k|}$. This mechanism is known to be universally optimal in a general cost minimization framework (Bayesian setting in [18] and worst-case setting in [20]). In this section, we show that the geometric noise adding mechanism achieves the fundamental limit on the privacy region under composition.

Consider the composition experiment for counting queries. For a pair of neighboring databases D_0 and D_1, some of the query outputs differ by one, since sensitivity is one, and for other queries the output might be the same. Let k denote the number of queries whose output differs with respect to D_0 and D_1. We show, in Appendix A, that the privacy region achieved by the geometric mechanism, is exactly described by the optimal composition theorem of (5). Further, since this is the largest privacy region under composition for the pair of database D_0 and D_1 that differ in k queries, no other mechanism can achieve a larger privacy region. Since the geometric mechanism does not depend on the particular choice of pairs of databases D_0 and D_1, nor does it depend on the specific query being asked, the mechanism achieves the exact composed privacy region universally for every pair of neighboring databases simultaneously.

Among the mechanisms guaranteeing the same level of privacy, the one with larger privacy region under composition is considered to be better in terms of allowing for smaller false alarm and missed detection rate in hypothesis testing whether or not the database contains a particular entry. In this sense, larger privacy degradation under composition has more utility. The geometric mechanism has the largest possible privacy region (or smallest possible privacy degradation) under composition, stated formally below; the proof is deferred to Appendix A.

Theorem 16: Under the k-fold composition experiment of counting queries, the geometric mechanism achieves the largest privacy region among all $(\varepsilon, 0)$-differentially private mechanisms, universally for every pair of neighboring databases simultaneously.

VI. Proofs

A. Proof of Theorem 5

Consider hypothesis testing between D_1 and D_2. If there is a point $(P_{\text{MD}}, P_{\text{FA}})$ achieved by M' but not by M, then we claim that this is a contradiction to the assumption that

D–X–Y forms a Markov chain. Consider a decision maker who have only access to the output of M. Under the Markov chain assumption, he can simulate the output of M' by generating a random variable Y conditioned on $M(D)$ and achieve every point in the privacy region of M' (cf. Theorem 3). Hence, the privacy region of M' must be included in the privacy region of M.

B. Proof of Theorem 2

First we prove that (ε, δ)-differential privacy implies (1). From the definition of differential privacy, we know that for all rejection set $S \subseteq \mathcal{X}$, $\mathbb{P}(M(D_0) \in \bar{S}) \leq e^{\varepsilon}\mathbb{P}(M(D_1) \in \bar{S}) + \delta$. This implies $1 - P_{\mathrm{FA}}(D_0, D_1, M, S) \leq e^{\varepsilon} P_{\mathrm{MD}}(D_0, D_1, M, S) + \delta$. This implies the first inequality of (1), and the second one follows similarly.

The converse follows analogously. For any set S, we assume $1 - P_{\mathrm{FA}}(D_0, D_1, M, S) \leq e^{\varepsilon} P_{\mathrm{MD}}(D_0, D_1, M, S) + \delta$. Then, it follows that $\mathbb{P}(M(D_0) \in \bar{S}) \leq e^{\varepsilon}\mathbb{P}(M(D_1) \in \bar{S}) + \delta$ for all choices of $S \subseteq \mathcal{X}$. Together with the symmetric condition $\mathbb{P}(M(D_1) \in \bar{S}) \leq e^{\varepsilon}\mathbb{P}(M(D_0) \in \bar{S}) + \delta$, this implies (ε, δ)-differential privacy.

C. Proof of Remark 3

We have a decision rule γ represented by a partition $\{S_i\}_{i\in\{1,\ldots,N\}}$ and corresponding accept probabilities $\{p_i\}_{i\in\{1,\ldots,N\}}$, such that if the output is in a set S_i, we accept with probability p_i. We assume the subsets are sorted such that $1 \geq p_1 \geq \ldots \geq p_N \geq 0$. Then, the probability of false alarm is

$$\begin{aligned} &P_{\mathrm{FA}}(D_0, D_1, M, \gamma) \\ &\quad = \sum_{i=1}^{N} p_i\, \mathbb{P}(M(D_0) \in S_i) \\ &\quad = p_N + \sum_{i=2}^{N} (p_{i-1} - p_i)\, \mathbb{P}(M(D_0) \in \cup_{j<i} S_j). \end{aligned}$$

and similarly, $P_{\mathrm{MD}}(D_0, D_1, M, \gamma) = (1 - p_1) + \sum_{i=2}^{N}(p_{i-1} - p_i)\, \mathbb{P}(M(D_1) \notin \cup_{j<i} S_j)$. Recall that $P_{\mathrm{FA}}(D_0, D_1, M, S) = \mathbb{P}(M(D_0) \in S)$ and $P_{\mathrm{MD}}(D_0, D_1, M, S) = \mathbb{P}(M(D_1) \in \bar{S})$. So for any decision rule γ, we can represent the pair $(P_{\mathrm{MD}}, P_{\mathrm{FA}})$ as a convex combination:

$$\begin{aligned} &\big(P_{\mathrm{MD}}(D_0, D_1, M, \gamma), P_{\mathrm{FA}}(D_0, D_1, M, \gamma)\big) \\ &\quad = \sum_{i=1}^{N+1} (p_{i-1} - p_i)\big(P_{\mathrm{MD}}(\cup_{j<i} S_j), P_{\mathrm{FA}}(\cup_{j<i} S_j)\big), \end{aligned}$$

where $P_{\mathrm{MD}}(\cup_{j<i} S_j) = P_{\mathrm{MD}}(D_0, D_1, M, \cup_{j<i} S_j)$, $P_{\mathrm{FA}}(\cup_{j<i} S_j) = P_{\mathrm{FA}}(D_0, D_1, M, \cup_{j<i} S_j)$, and we used $p_0 = 1$ and $p_{N+1} = 0$, and hence it is included in the convex hull of the privacy region achieved by decision rules with hard thresholding.

D. Proof of Theorem 10

We need to provide an outer bound on the privacy region achieved by $\tilde{X}_0$ and $\tilde{X}_1$ defined in (9) and (10) under k-fold composition. Let P_0 denote the probability mass function of $\tilde{X}_0$ and P_1 denote the PMF of $\tilde{X}_1$. Also, let P_0^k and P_1^k denote the joint PMF of k i.i.d. copies of $\tilde{X}_0$ and $\tilde{X}_1$ respectively. Also, for a set $S \subseteq \mathcal{X}^k$, we let $P_0^k(S) = \sum_{x\in S} P_0^k(x)$. In our example, $\mathcal{X} = \{1, 2, 3, 4\}$, and

$$\begin{aligned} P_0 &= \left[\delta \ \tfrac{(1-\delta)e^{\varepsilon}}{1+e^{\varepsilon}} \ \tfrac{1-\delta}{1+e^{\varepsilon}} \ 0\right], \\ P_1 &= \left[0 \ \tfrac{1-\delta}{1+e^{\varepsilon}} \ \tfrac{(1-\delta)e^{\varepsilon}}{1+e^{\varepsilon}} \ \delta\right], \\ P_0^2 &= \begin{bmatrix} \delta^2 & \delta\frac{(1-\delta)e^{\varepsilon}}{1+e^{\varepsilon}} & \delta\frac{(1-\delta)}{1+e^{\varepsilon}} & 0 \\ \delta\frac{(1-\delta)e^{\varepsilon}}{1+e^{\varepsilon}} & \left(\frac{(1-\delta)e^{\varepsilon}}{1+e^{\varepsilon}}\right)^2 & \left(\frac{1-\delta}{1+e^{\varepsilon}}\right)^2 e^{\varepsilon} & 0 \\ \delta\frac{1-\delta}{1+e^{\varepsilon}} & \left(\frac{1-\delta}{1+e^{\varepsilon}}\right)^2 e^{\varepsilon} & \left(\frac{1-\delta}{1+e^{\varepsilon}}\right)^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{etc.} \end{aligned}$$

We can compute the privacy region from P_0^k and P_1^k directly, by computing the line tangent to the boundary. A tangent line with slope $-e^{\tilde{\varepsilon}}$ can be represented as

$$P_{\mathrm{FA}} = -e^{\tilde{\varepsilon}} P_{\mathrm{MD}} + 1 - d_{\tilde{\varepsilon}}(P_0^k, P_1^k). \tag{15}$$

To find the tangent line, we need to maximize the shift, which is equivalent to moving the line downward until it is tangent to the boundary of the privacy region (cf. Figure 3).

$$d_{\tilde{\varepsilon}}(P_0^k, P_1^k) \equiv \max_{S\subseteq\mathcal{X}^k} P_0^k(S) - e^{\tilde{\varepsilon}} P_1^k(S).$$

Notice that the maximum is achieved by a set $B \equiv \{x \in \mathcal{X}^k \mid P_0^k(x) \geq e^{\tilde{\varepsilon}} P_1^k(x)\}$. Then,

$$d_{\tilde{\varepsilon}}(P_0^k, P_1^k) = P_0^k(B) - e^{\tilde{\varepsilon}} P_1^k(B).$$

For the purpose of proving the bound of the form (7), we separate the analysis of the above formula into two parts: one where either $P_0^k(x)$ or $P_1^k(x)$ is zero and the other when both are positive. Effectively, this separation allows us to treat the effects of $(\varepsilon, 0)$-differential privacy and $(0, \delta)$-differential privacy separately. In previous work [16], they separated the analysis in a similar way. Here we provide a simpler proof technique. Further, all the proof techniques we use naturally generalize to compositions of general (ε, δ)-differentially private mechanisms other than the specific example of $\tilde{X}_0$ and $\tilde{X}_1$ we consider in this section.

Let $\tilde{X}_0^k$ denote a k-dimensional random vector whose entries are independent copies of $\tilde{X}_0$. We partition B into two sets: $B = B_0 \bigcup B_1$ and $B_0 \bigcap B_1 = \emptyset$. Let $B_0 \equiv \{x \in \mathcal{X}^k : P_0^k(x) \geq e^{\tilde{\varepsilon}} P_1^k(x)$, and $P_1^k(x) = 0\}$ and $B_1 \equiv \{x \in \mathcal{X}^k : P_0^k(x) \geq e^{\tilde{\varepsilon}} P_1^k(x)$, and $P_1^k(x) > 0\}$. Then, it is not hard to see that $P_0^k(B_0) = 1 - \mathbb{P}(\tilde{X}_0^k \in \{1, 2, 3\}^k) = 1 - (1 - \delta)^k$, $P_1^k(B_0) = 0$, $P_0^k(B_1) = P_0^k(B_1 | \tilde{X}_0^k \in \{1, 2\}^k)\mathbb{P}(\tilde{X}_0^k \in \{1, 2\}^k) = (1 - \delta)^k\, P_0^k(B_1 | \tilde{X}_0^k \in \{1, 2\}^k)$, and $P_1^k(B_1) = (1 - \delta)^k\, P_1^k(B_1 | \tilde{X}_1^k \in \{1, 2\}^k)$. It follows that

$$\begin{aligned} &P_0^k(B_0) - e^{\tilde{\varepsilon}} P_1^k(B_0) \\ &\qquad = 1 - (1 - \delta)^k, \text{ and} \\ &P_0^k(B_1) - e^{\tilde{\varepsilon}} P_1^k(B_1) \\ &= (1 - \delta)^k \big(P_0^k(B_1 | \tilde{X}_0^k \in \{1, 2\}^k) - e^{\tilde{\varepsilon}} P_1^k(B_1 | \tilde{X}_1^k \in \{1, 2\}^k)\big). \end{aligned}$$

Let $\tilde{P}_0^k(x) \equiv P_0^k(x|x \in \{1,2\}^k)$ and $\tilde{P}_1^k(x) \equiv P_1^k(x|x \in \{1,2\}^k)$. Then, we have

$$\begin{aligned} & d_{\tilde{\varepsilon}}(P_0^k, P_1^k) \\ &= P_0^k(B_0) - e^{\tilde{\varepsilon}} P_1^k(B_0) + P_0^k(B_1) - e^{\tilde{\varepsilon}} P_1^k(B_1) \\ &= 1 - (1-\delta)^k + (1-\delta)^k \big(\tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}} \tilde{P}_1^k(B_1)\big). \quad (16) \end{aligned}$$

Now, we focus on upper bounding $\tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}} \tilde{P}_1^k(B_1)$, using a variant of Chernoff's tail bound. Notice that

$$\begin{aligned} & \tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}} \tilde{P}_1^k(B_1) \\ &= \mathbb{E}_{\tilde{P}_0^k}\Big[\mathbb{I}_{\left(\log \frac{\tilde{P}_0^k(\tilde{X}^k)}{\tilde{P}_1^k(\tilde{X}^k)} \geq \tilde{\varepsilon}\right)}\Big] - e^{\tilde{\varepsilon}} \mathbb{E}_{\tilde{P}_0^k}\Big[\mathbb{I}_{\left(\log \frac{\tilde{P}_0^k(\tilde{X}^k)}{\tilde{P}_1^k(\tilde{X}^k)} \geq \tilde{\varepsilon}\right)} \frac{\tilde{P}_1^k(\tilde{X}^k)}{\tilde{P}_0^k(\tilde{X}^k)}\Big] \\ &= \mathbb{E}_{\tilde{P}_0^k}\Big[\mathbb{I}_{\left(\log(\tilde{P}_0^k(\tilde{X}^k)/\tilde{P}_1^k(\tilde{X}^k)) \geq \tilde{\varepsilon}\right)} \Big(1 - e^{\tilde{\varepsilon}} \frac{\tilde{P}_1^k(\tilde{X}^k)}{\tilde{P}_0^k(\tilde{X}^k)}\Big)\Big] \\ &\leq \mathbb{E}[e^{\lambda Z - \lambda\tilde{\varepsilon} + \lambda \log\lambda - (\lambda+1)\log(\lambda+1)}], \quad (17) \end{aligned}$$

where we use a random variable $Z \equiv \log(\tilde{P}_0^k(\tilde{X}_0^k)/\tilde{P}_1^k(\tilde{X}_0^k))$ and the last line follows from $\mathbb{I}_{(x \geq \tilde{\varepsilon})}(1 - e^{\tilde{\varepsilon}-x}) \leq e^{\lambda(x-\tilde{\varepsilon}) + \lambda\log\lambda - (\lambda+1)\log(\lambda+1)}$ for any $\lambda \geq 0$. To show this inequality, notice that the right-hand side is always non-negative. So it is sufficient to show that the inequality holds, without the indicator on the left-hand side. Precisely, let $f(x) = e^{\lambda(x-\tilde{\varepsilon}) + \lambda\log\lambda - (\lambda+1)\log(\lambda+1)} + e^{\tilde{\varepsilon}-x} - 1$. This is a convex function with $f(x^*) = 0$ and $f'(x^*) = 0$ at $x^* = \tilde{\varepsilon} + \log((\lambda+1)/\lambda)$. It follows that this is a non-negative function.

Next, we give an upper bound on the moment generating function of Z.

$$\begin{aligned} \mathbb{E}_{\tilde{P}_0}[e^{\lambda \log(P_0(X)/P_1(X))}] &= \frac{e^{\varepsilon}}{e^{\varepsilon}+1} e^{\lambda\varepsilon} + \frac{1}{e^{\varepsilon}+1} e^{-\lambda\varepsilon} \\ &\leq e^{\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\lambda\varepsilon + \frac{1}{2}\lambda^2\varepsilon^2}, \end{aligned}$$

for any λ, which follows from the fact that $pe^x + (1-p)e^{-x} \leq e^{(2p-1)x + (1/2)x^2}$ for any $x \in \mathbb{R}$ and $p \in [0,1]$ [1, Lemma A.1.5]. Substituting this into (17) with a choice of $\lambda = \frac{\tilde{\varepsilon} - k\varepsilon(e^{\varepsilon}-1)/(e^{\varepsilon}+1)}{k\varepsilon^2}$, we get

$$\begin{aligned} & \tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}} \tilde{P}_1^k(B_1) \\ &\leq e^{\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\lambda\varepsilon k + \frac{1}{2}\lambda^2\varepsilon^2 \;\; k - \lambda\tilde{\varepsilon} + \lambda\log\lambda - (\lambda+1)\log(\lambda+1)} \\ &= e^{-\frac{k\varepsilon^2}{2}\left(\lambda - \frac{1}{k\varepsilon^2}\left(\tilde{\varepsilon} - k\varepsilon\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\right)\right)^2 - \frac{1}{2k\varepsilon^2}\left(\tilde{\varepsilon} - \frac{k\varepsilon(e^{\varepsilon}-1)}{e^{\varepsilon}+1}\right)^2} \\ &\quad \times e^{\lambda\log\frac{\lambda}{\lambda+1} - \log(\lambda+1)} \\ &\leq \exp\Big\{ - \frac{1}{2\;k\varepsilon^2}\Big(\tilde{\varepsilon} - k\varepsilon\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\Big)^2 - \log(\lambda+1)\Big\} \\ &\leq \frac{1}{1 + \frac{\tilde{\varepsilon} - k\varepsilon(e^{\varepsilon}-1)/(e^{\varepsilon}+1)}{k\varepsilon^2}} \exp\Big\{ - \frac{1}{2\;k\varepsilon^2}\Big(\tilde{\varepsilon} - k\varepsilon\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\Big)^2 \Big\} \\ &= \frac{1}{1 + \frac{\sqrt{2k\varepsilon^2\log(e + (\sqrt{k\varepsilon^2}/\tilde{\delta}))}}{k\varepsilon^2}} \frac{1}{e + \frac{\sqrt{k\varepsilon^2}}{\tilde{\delta}}} \\ &\leq \frac{1}{\sqrt{k\varepsilon^2} + \sqrt{2\log(e + (\sqrt{k\varepsilon^2}/\tilde{\delta}))}} \frac{\tilde{\delta}}{\frac{e\tilde{\delta}}{\sqrt{k\varepsilon^2}} + 1}, \end{aligned}$$

for our choice of $\tilde{\varepsilon} = k\varepsilon(e^{\varepsilon}-1)/(e^{\varepsilon}+1) + \varepsilon\sqrt{2k\,\log(e + (\sqrt{k\varepsilon^2}/\tilde{\delta}))}$. The right-hand side is always less than $\tilde{\delta}$.

Similarly, one can show that the right-hand side is less than $\tilde{\delta}$ for the choice of $\tilde{\varepsilon} = k\varepsilon(e^{\varepsilon}-1)/(e^{\varepsilon}+1) + \varepsilon\sqrt{2k\,\log(1/\tilde{\delta})}$. We get that the k-fold composition is $(\tilde{\varepsilon}, 1-(1-\delta)^k\,(1-\tilde{\delta}))$-differentially private.

E. Proof of Theorem 11

In this section, we closely follow the proof of Theorem 10 in Section VI-D carefully keeping the dependence on ℓ, the index of the composition step. For brevity, we omit the details which overlap with the proof of Theorem 10. By the same argument as in the proof of Theorem 9, we only need to provide an outer bound on the privacy region achieved by $\tilde{X}_0^{(\ell)}$ and $\tilde{X}_1^{(\ell)}$ under k-fold composition, defined as

$$\mathbb{P}(\tilde{X}_0^{(\ell)} = x) = \tilde{P}_0^{(\ell)}(x) \equiv \begin{cases} \delta_\ell & \text{for } x = 0, \\ \frac{(1-\delta_\ell)\,e^{\varepsilon_\ell}}{1+e^{\varepsilon_\ell}} & \text{for } x = 1, \\ \frac{1-\delta_\ell}{1+e^{\varepsilon_\ell}} & \text{for } x = 2, \\ 0 & \text{for } x = 3. \end{cases}, \text{ and}$$

$$\mathbb{P}(\tilde{X}_1^{(\ell)} = x) = \tilde{P}_1^{(\ell)}(x) \equiv \begin{cases} 0 & \text{for } x = 0, \\ \frac{1-\delta_\ell}{1+e^{\varepsilon_\ell}} & \text{for } x = 1, \\ \frac{(1-\delta_\ell)\,e^{\varepsilon_\ell}}{1+e^{\varepsilon_\ell}} & \text{for } x = 2, \\ \delta_\ell & \text{for } x = 3. \end{cases}$$

Using the similar notations as Section VI-D, it follows that under k-fold composition,

$$\begin{aligned} & d_{\tilde{\varepsilon}}(P_0^k, P_1^k) \\ &= 1 - \prod_{\ell=1}^{k}(1-\delta_\ell) + \big(\tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}}\tilde{P}_1^k(B_1)\big)\prod_{\ell=1}^{k}(1-\delta_\ell). \end{aligned} \quad (18)$$

Now, we focus on upper bounding $\tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}} \tilde{P}_1^k(B_1)$, using a variant of Chernoff's tail bound. We know that

$$\begin{aligned} & \tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}} \tilde{P}_1^k(B_1) \\ &= \mathbb{E}_{\tilde{P}_0^k}\Big[\mathbb{I}_{\left(\log \frac{\tilde{P}_0^k(\tilde{X}^k)}{\tilde{P}_1^k(\tilde{X}^k)} \geq \tilde{\varepsilon}\right)}\Big] - e^{\tilde{\varepsilon}} \mathbb{E}_{\tilde{P}_0^k}\Big[\mathbb{I}_{\left(\log \frac{\tilde{P}_0^k(\tilde{X}^k)}{\tilde{P}_1^k(\tilde{X}^k)} \geq \tilde{\varepsilon}\right)} \frac{\tilde{P}_1^k(\tilde{X}^k)}{\tilde{P}_0^k(\tilde{X}^k)}\Big] \\ &= \mathbb{E}_{\tilde{P}_0^k}\Big[\mathbb{I}_{\left(\log(\tilde{P}_0^k(\tilde{X}^k)/\tilde{P}_1^k(\tilde{X}^k)) \geq \tilde{\varepsilon}\right)} \Big(1 - e^{\tilde{\varepsilon}} \frac{\tilde{P}_1^k(\tilde{X}^k)}{\tilde{P}_0^k(\tilde{X}^k)}\Big)\Big] \\ &\leq \mathbb{E}[e^{\lambda Z - \lambda\tilde{\varepsilon} + \lambda \log\lambda - (\lambda+1)\log(\lambda+1)}], \quad (19) \end{aligned}$$

where we use a random variable $Z \equiv \log(\tilde{P}_0^k(\tilde{X}_0^k)/\tilde{P}_1^k(\tilde{X}_0^k))$ and the last line follows from the fact that $\mathbb{I}_{(x \geq \tilde{\varepsilon})}(1 - e^{\tilde{\varepsilon}-x}) \leq e^{\lambda(x-\tilde{\varepsilon}) + \lambda\log\lambda - (\lambda+1)\log(\lambda+1)}$ for any $\lambda \geq 0$.

Next, we give an upper bounds on the moment generating function of Z. From the definition of $\tilde{P}_0^{(\ell)}$ and $\tilde{P}_1^{(\ell)}$, $\mathbb{E}[e^{\lambda Z}] = \Big(\mathbb{E}_{\tilde{P}_0^{(\ell)}}[e^{\lambda \log(\tilde{P}_0^{(\ell)}(\tilde{X}_0^{(\ell)})/\tilde{P}_1^{(\ell)}(\tilde{X}_0^{(\ell)}))}]\Big)^k$. Let $\tilde{\varepsilon} = \sum_{\ell=1}^k (e^{\varepsilon_\ell} - 1)\varepsilon_\ell/(e^{\varepsilon_\ell}+1) + \sqrt{2\sum_{\ell=1}^k \varepsilon_\ell^2 \log\big(e + (\sqrt{\sum_{\ell=1}^k \varepsilon_\ell^2}/\tilde{\delta})\big)}$. Next we show that the k-fold composition is $(\tilde{\varepsilon}, 1-(1-\tilde{\delta})\prod_{\ell\in[k]}(1-\delta_\ell)\,)$-differentially private.

$$\mathbb{E}_{\tilde{P}_0^{(\ell)}}[e^{\lambda \log(P_0^{(\ell)}(X)/P_1^{(\ell)}(X))}] \leq e^{\frac{e^{\varepsilon_\ell}-1}{e^{\varepsilon_\ell}+1}\lambda\varepsilon_\ell + \frac{1}{2}\lambda^2\varepsilon_\ell^2},$$

for any λ. Substituting this into (19) with a choice of $\lambda = \frac{\tilde{\varepsilon} - \sum_{\ell\in[k]} \varepsilon_\ell (e^{\varepsilon_\ell}-1)/(e^{\varepsilon_\ell}+1)}{\sum_{\ell\in[k]} \varepsilon_\ell^2}$, we get

$$\begin{aligned}
&\tilde{P}_0^k(B_1) - e^{\tilde{\varepsilon}} \tilde{P}_1^k(B_1) \\
&\quad \le \frac{1}{1 + \frac{\tilde{\varepsilon} - \sum_{\ell\in[k]} \varepsilon_\ell (e^{\varepsilon_\ell}-1)/(e^{\varepsilon_\ell}+1)}{\sum_{\ell\in[k]} \varepsilon_\ell^2}} \\
&\qquad \times \exp\Big\{ - \frac{1}{2\sum_{\ell\in[k]} \varepsilon_\ell^2} \Big(\tilde{\varepsilon} - \sum_{\ell\in[k]} \varepsilon_\ell \frac{e^{\varepsilon_\ell}-1}{e^{\varepsilon_\ell}+1}\Big)^2 \Big\}.
\end{aligned}$$

Substituting $\tilde{\varepsilon}$, we get the desired bound.

Similarly, we can prove that with $\tilde{\varepsilon} = \sum_{\ell=1}^k (e^{\varepsilon_\ell} - 1)\varepsilon_\ell/(e^{\varepsilon_\ell}+1) + \sqrt{2\sum_{\ell=1}^k \varepsilon_\ell^2 \log\big(1/\tilde{\delta}\big)}$, the desired bound also holds.

APPENDIX

Remark 17: The following statements are true.

(a) *If a mechanism is (ε, δ)-differentially private, then it is $(\tilde{\varepsilon}, \tilde{\delta})$-differentially private for all pairs of $\tilde{\varepsilon}$ and $\tilde{\delta} \ge \delta$ satisfying*

$$\frac{1-\delta}{1+e^{\varepsilon}} \ge \frac{1-\tilde{\delta}}{1+e^{\tilde{\varepsilon}}}.$$

(b) *For a pair of neighboring databases D and D', and all (ε, δ)-differentially private mechanisms, the total variation distance defined as $\|M(D) - M(D')\|_{\mathrm{TV}} = \max_{S\subseteq\mathcal{X}} \mathbb{P}(M(D') \in S) - \mathbb{P}(M(D) \in S)$ is bounded by*

$$\sup_{(\varepsilon,\delta)\text{-differentially private } M} \|M(D) - M(D')\|_{\mathrm{TV}} \le 1 - \frac{2(1-\delta)}{1+e^{\varepsilon}}.$$

Proof: **Proof of (*a*).** From Figure 1, it is immediate that $\mathcal{R}(\varepsilon,\delta) \subseteq \mathcal{R}(\tilde{\varepsilon},\tilde{\delta})$ when the conditions are satisfied. Then, for a (ε,δ)-private M, it follows from $\mathcal{R}(M) \subseteq \mathcal{R}(\varepsilon,\delta) \subseteq \mathcal{R}(\tilde{\varepsilon},\tilde{\delta})$ that M is $(\tilde{\varepsilon},\tilde{\delta})$-differentially private.

Proof of (*b*). By definition, $\|M(D) - M(D')\|_{\mathrm{TV}} = \max_{S\subseteq\mathcal{X}} \mathbb{P}(M(D') \in S) - \mathbb{P}(M(D) \in S)$. Letting S be the rejection region in our hypothesis testing setting, the total variation distance is defined by the following optimization problem:

$$\begin{aligned}
&\max_S \ 1 - P_{\mathrm{MD}}(S) - P_{\mathrm{FA}}(S) \\
&\text{subject to} (P_{\mathrm{MD}}(S), P_{\mathrm{FA}}(S)) \in \mathcal{R}(\varepsilon,\delta), \text{for all } S \subseteq \mathcal{X}. \quad (20)
\end{aligned}$$

From Figure 1 it follows immediately that the total variation distance cannot be larger than $\delta + (1-\delta)(e^{\varepsilon}-1)/(e^{\varepsilon}+1)$. □

Following the analysis in Section VI-D, we know that the privacy region of a composition of mechanisms is described by a set of (ε,δ) pairs that satisfy the following:

$$\delta = \mu_0^k(B) - e^{\varepsilon}\mu_1^k(B),$$

where μ_0^k and μ_1^k are probability measures of the mechanism under k-fold composition when the data base is D_0 and D_1 respectively, and the subset $B = \arg\max_{S\subseteq\mathbb{R}^k} \mu_0^k(S) - e^{\varepsilon}\mu_1^k(S)$.

In the case of Gaussian mechanisms, we can assume without loss of generality that D_0 is such that $q_i(D_0) = 0$ and D_1 is such that $q_i(D_1) = \Delta$ for all $i \in \{1,\dots,k\}$. When adding Gaussian noises with variances σ^2, we want to ask how small the variance can be and still ensure (ε,δ)-differential privacy under k-fold composition.

Let $f_0^k(x_1,\dots,x_k) = \prod_{i=1}^k f_0(x_i) = (1/\sqrt{2\pi\sigma^2})^k e^{-\sum_{i=1}^k x_i^2/2\sigma^2}$ and $f_1^k(x_1\dots,x_k) = \prod_{i=1}^k f_1(x_i) = (1/\sqrt{2\pi\sigma^2})^k e^{-\sum_{i=1}^k (x_i-\Delta)^2/2\sigma^2}$ be the probability density functions of Gaussians centered at zero and $\Delta\mathbb{1}_k$ respectively. Using a similar technique as in (17), we know that

$$\begin{aligned}
&\mu_0^k(B) - e^{\varepsilon}\mu_1^k(B) \\
&\quad = \mathbb{E}_{\mu_0^k}\Big[\mathbb{I}_{\left(\log\frac{f_0^k(\tilde{X}^k)}{f_1^k(\tilde{X}^k)} \ge \varepsilon\right)}\Big] - e^{\varepsilon}\mathbb{E}_{\mu_0^k}\Big[\mathbb{I}_{\left(\log\frac{f_0^k(\tilde{X}^k)}{f_1^k(\tilde{X}^k)} \ge \varepsilon\right)} \frac{f_1^k(\tilde{X}^k)}{f_0^k(\tilde{X}^k)}\Big] \\
&\quad = \mathbb{E}_{\mu_0^k}\Big[\mathbb{I}_{\left(\log(f_0^k(\tilde{X}^k)/f_1^k(\tilde{X}^k)) \ge \varepsilon\right)} \Big(1 - e^{\varepsilon}\frac{f_1^k(\tilde{X}^k)}{f_0^k(\tilde{X}^k)}\Big)\Big] \\
&\quad \le \mathbb{E}[e^{\lambda Z - \lambda\varepsilon + \lambda\log\lambda - (\lambda+1)\log(\lambda+1)}], \qquad (21)
\end{aligned}$$

where $\tilde{X}^k$ is a random vector distributed according to μ_0^k, $Z \equiv \log(f_0^k(\tilde{X}^k)/f_1^k(\tilde{X}^k))$, and the last line follows from $\mathbb{I}_{(x\ge\varepsilon)}(1 - e^{\varepsilon - x}) \le e^{\lambda(x-\varepsilon)+\lambda\log\lambda-(\lambda+1)\log(\lambda+1)}$ for any $\lambda \ge 0$.

Next, we give an upper bound on the moment generating function of Z.

$$\begin{aligned}
\mathbb{E}_{\mu_0}[e^{\lambda\log(f_0(X)/f_1(X))}] &= \mathbb{E}[e^{-\lambda\Delta X/\sigma^2}] e^{\lambda\Delta^2/2\sigma^2} \\
&\le e^{(\Delta^2/2\sigma^2)\lambda^2 + (\Delta^2/2\sigma^2)\lambda},
\end{aligned}$$

for any $\lambda \ge 0$. Substituting this into (21) with a choice of $\lambda = \frac{\sigma^2}{k\Delta^2}\big(\varepsilon - \frac{k\Delta^2}{2\sigma^2}\big)$, which is positive for $\varepsilon > k\Delta^2/2\sigma^2$, we get

$$\begin{aligned}
&\mu_0^k(B) - e^{\varepsilon}\mu_1^k(B) \\
&\quad \le e^{(k\Delta^2/2\sigma^2)\lambda^2 + (k\Delta^2/2\sigma^2)\lambda - \varepsilon\lambda + \lambda\log\lambda - (\lambda+1)\log(\lambda+1)} \\
&\quad \le \frac{1}{1 + \frac{\sigma^2}{k\Delta^2}\big(\varepsilon - \frac{k\Delta^2}{2\sigma^2}\big)} \exp\Big\{ - \frac{\sigma^2}{2k\Delta^2}\Big(\varepsilon - \frac{k\Delta^2}{2\sigma^2}\Big)^2 \Big\} \\
&\quad \le \frac{1}{1 + \sqrt{\frac{2\sigma^2}{k\Delta^2}\log\big(e + \frac{1}{\delta}\sqrt{\frac{k\Delta^2}{\sigma^2}}\big)}} \frac{1}{e + \frac{1}{\delta}\sqrt{\frac{k\Delta^2}{\sigma^2}}} \\
&\quad \le \frac{1}{\sqrt{\frac{k\Delta^2}{\sigma^2}} + \sqrt{2\log(e + (1/\delta)\sqrt{k\Delta^2/\sigma^2})}} \frac{\delta}{e\delta\sqrt{\frac{\sigma^2}{k\Delta^2}} + 1},
\end{aligned}$$

for our choice of σ^2 such that $\varepsilon \ge k\Delta^2/(2\sigma^2) + \sqrt{(2k\Delta^2/\sigma^2)\log(e + (1/\delta)\sqrt{k\Delta^2/\sigma^2})}$. The right-hand side is always less than δ.

With $\sigma^2 \ge (4k\Delta^2/\varepsilon^2)\log(e + (\varepsilon/\delta))$ and $\sigma^2 \ge k\Delta^2/(4\varepsilon)$, this ensures that the above condition is satisfied. This implies that we only need $\sigma^2 = O((k\Delta^2/\varepsilon^2)\log(e + (\varepsilon/\delta)))$.

Theorem 16 follows directly from the proof of Theorem 9, once the appropriate associations are made. Consider two databases D_0 and D_1, and a single query q such that $q(D_1) = q(D_0) + 1$. The geometric mechanism produces two random outputs $q(D_0) + Z$ and $q(D_1) + Z$ where Z is distributed accruing to the geometric distribution. Let P_0 and P_1 denote the distributions of the random output respectively. For

$x \le q(D_0)$, $P_0(x) = e^{\varepsilon} P_1(x)$, and for $x > q(D_0)$, $e^{\varepsilon} P_0(x) = P_1(x)$. Then, it is not difficult to see that the privacy region achieved by the geometric mechanism is equal to the privacy region achieved by the canonical binary example of $\tilde{X}_0$ and $\tilde{X}_1$ in (9) and (10) with $\delta = 0$. This follows from the fact there is a stochastic transition from the pair $\tilde{X}_0$ and $\tilde{X}_1$ to $q(D_0) + Z$ and $q(D_1) + Z$; further, the converse is also true. Hence, from the perspective of hypothesis testing, those two (pairs of) outcomes are equivalent.

It now follows from the proof of Theorem 9 that the k-fold composition privacy region is exactly the optimal privacy region described in (5) with $\delta = 0$. We also know that this is the largest possible privacy region achieved by a class of $(\varepsilon, 0)$-differentially private mechanisms.

A. Cut Queries of a Graph and Variance Queries of a Matrix

Blocki et. al. [2] showed that classical Johnson-Lindenstrauss transform can be used to produce a differentially private version of a database. Further, they show that this achieves the best tradeoff between privacy and utility for two applications: cut queries of a graph and variance queries of a matrix. In this section, we show how the best known trade off can be further improved by applying Theorem 10.

First, Blocki et. al. provide a differentially private mechanism for cut queries $q(G, S)$: the number of edges crossing a $(S, \bar{S})$-cut in a weighted undirected graph G. This mechanism produces a sanitized graph satisfying (ε, δ)-differential privacy, where two graphs are neighbors if they only differ on a single edge. The *utility* of the mechanism is measured via the additive error τ incurred by the privatization. Precisely, a mechanism M is said to give a (η, τ, ν)-approximation for a *single* cut query $q(\cdot, \cdot)$, if for every graph G and every nonempty S it holds that

$$\mathbb{P}\Big((1-\eta)\, q(G,S) - \tau \;\le\; M(G,S) \;\le\; (1+\eta)\, q(G,S) + \tau \Big) \ge 1 - \nu. \tag{22}$$

For the proposed Johnson-Lindenstrauss mechanism satisfying (ε, δ)-differential privacy, it is shown that the additive error τ_0 incurred by querying the database k times is bounded by [2, Theorem 3.2][1]

$$\tau_0 = O\Big(|S| \frac{\sqrt{\log(1/\delta)\log(k/\nu)}}{\varepsilon} \log\Big(\frac{\log(k/\nu)}{\eta^2 \delta}\Big)\Big). \tag{23}$$

Compared to other state-of-the-art privacy mechanisms such as the Laplace noise adding mechanism [17], Exponential mechanism [30], Multiplicative weights [23], and Iterative Database Construction [19], it is shown in [2] that the Johnson-Lindenstrauss mechanism achieves the best tradeoff between the additive error τ_0 and the privacy ε. This tradeoff in (23) is proved using the existing Theorem 8. We can improve this analysis using the optimal composition theorem of Theorem 10, which gives

$$\tau = O\Big(|S| \frac{\sqrt{\log(e + \varepsilon/\delta)\log(k/\nu)}}{\varepsilon} \log\Big(\frac{\log(k/\nu)}{\eta^2 \delta}\Big)\Big). \tag{24}$$

[1]The original theorem is stated for a single query with $k = 1$. Here we state it more generally with arbitrary k. This requires scaling ν by $1/k$ to take into account the union bound over k query outputs in the *utility* guarantee in (22). This is smaller than (23) by (a square root of) a logarithmic factor when $\varepsilon = \Theta(\delta)$. The proof of the analysis in (24) is provided below.

A similar technique has been used in [2] to provide a differentially private mechanism for variance queries $v(A, x) = x^T A^T A x$: the variance of a given matrix in a direction x. The proposed mechanism produces a sanitized covariance matrix that satisfy (ε, δ)-differential privacy, where two matrices are neighbors if they differ only in a single row and the difference is by Euclidean distance at most one. With the previous composition theorem in Theorem 8, the authors of [2] get an error bound $\tau_1 = O\Big(\frac{\log(1/\delta)\log(k/\nu)}{\varepsilon^2 \eta} \log^2\Big(\frac{\log(k/\nu)}{\eta^2\delta}\Big)\Big)$. Using our tight composition theorem, this can be improved as $\tau = O\Big(\frac{\log(e+\varepsilon/\delta)\log(k/\nu)}{\varepsilon^2 \eta} \log^2\Big(\frac{\log(k/\nu)}{\eta^2\delta}\Big)\Big)$. Again, for $\varepsilon = \Theta(\delta)$, this is an improvement of a logarithmic factor.

For cut queries, Johnson-Lindenstrauss mechanism proceeds as follows:

Algorithm 2 JL Mechanism for Cut Queries [2]

Input: A n-node graph G, parameters $\varepsilon, \delta, \eta, \nu > 0$
Output: An approximate Laplacian of G: $\tilde{L}$
1: Set $r \leftarrow 8\log(2/\nu)/\nu^2$
2: Set $w \leftarrow \sqrt{32 r \log(2/\delta)}\log(4r/\delta)/\varepsilon$
3: For every pair of nodes $I \neq j$,
  Set new weights $w_{i,j} = w/n + (1 - w/n) w_{i,j}$
4: Randomly draw a matrix N of size $r \times \binom{n}{2}$, whose entries are i.i.d. samples of $\mathcal{N}(0, 1)$
5: Output $\tilde{L} = (1/r) E_G^T N^T N E_G$, where E_G is $\binom{n}{2} \times n$ matrix whose (i, j)-th row is $\sqrt{w_{i,j}}(e_i - e_j)$

Here e_i is the standard basis vector with one in the i-th entry. Given this synopsis of the sanitized graph Laplacian, a cut query $q(G, S)$ returns $1/(1-w/n)(\mathbb{1}_S^T \tilde{L} \mathbb{1}_S - w|S|(n-|S|)/n)$, where $\mathbb{1}_S \in \{0,1\}^n$ is the indicator vector for the set S. If the matrix N is an identity matrix, this returns the correct cut value of G.

We have the choice of $w \in \mathbb{R}$ and $r \in \mathbb{Z}$ to ensure that the resulting mechanism is (ε, δ)-differentially private, and satisfy (η, τ, ν)-approximation guarantees of (22). We utilize the following lemma from [2].

Lemma 18: With the choice of

$$w = \frac{4}{\varepsilon_0}\log(2/\delta_0) \text{ and } r = \frac{8\log(2/\nu)}{\eta^2},$$

each row of NE_G satisfy $(\varepsilon_0, \delta_0)$-differential privacy, and the resulting Johnson-Lindenstrauss mechanism satisfy (η, τ, ν)-approximation guarantee with

$$\tau = 2|S|\,\eta\, w,$$

where $|S|$ is the size of the smaller partition S of the cut $(S, \bar{S})$. The error bound in (23) follows from choosing

$$\varepsilon_0 = \frac{\varepsilon}{\sqrt{4r\log(2/\delta)}} \text{ textand } \delta_0 = \frac{\delta}{2\,r},$$

and applying Theorem 8 to ensure that the resulting mechanism with r-composition of the r rows of ME_G is (ε, δ)-differentially private. Here it is assumed that $\varepsilon < 1$.

Now, with Theorem 9, we do not require ε_0 to be as small, which in turn allows us to add smaller noise w, giving us an improved error bound on τ. Precisely, using Theorem 10 it follows that a choice of

$$\varepsilon_0 = \frac{\varepsilon}{\sqrt{4r\log(e+2\varepsilon/\delta)}} \text{ and } \delta_0 = \frac{\delta}{2r},$$

suffices to ensure that after r-composition we get (ε, δ)-differential privacy. Resulting noise is bounded by $w \leq 4\sqrt{4r\log(e+2\varepsilon/\delta)}\log(4r/\delta)/\varepsilon$, which gives the error bound in (24). The proof follows analogously for the matrix variance queries.

ACKNOWLEDGMENT

The authors thank Maxim Raginsky for helpful discussions and for pointing out [4], and Moritz Hardt for pointing out an error in an earlier version of this paper. The authors thank anonymous reviewers for their careful reading and suggestions in improving the readability of the paper.

REFERENCES

[1] N. Alon and J. H. Spencer, *The Probabilistic Method*. Hoboken, NJ, USA: Wiley, 2004.
[2] J. Blocki, A. Blum, A. Datta, and O. Sheffet, "The Johnson-lindenstrauss transform itself preserves differential privacy," in *Proc. IEEE 53rd Annu. Symp. Found. Comput. Sci. (FOCS)*, Oct. 2012, pp. 410–419.
[3] A. Blum, C. Dwork, F. McSherry, and K. Nissim, "Practical privacy: The SuLQ framework," in *Proc. 24th ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 2005, pp. 128–138.
[4] D. Blackwell, "Equivalent comparisons of experiments," *Ann. Math. Statist*, vol. 24, no. 2, pp. 265–272, 1953.
[5] N. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, vol. 11, no. 2, pp. 267–271, Apr. 1965.
[6] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to noninteractive database privacy," *J. ACM*, vol. 60, no. 2, p. 12, 2013.
[7] T. M. Cover and A. Thomas, "Determinant inequalities via information theory," *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 3, pp. 384–392, 1988.
[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[9] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.
[10] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology—EUROCRYPT*, Springer, 2006, pp. 486–503.
[11] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proc. 41st Annu. ACM Symp. Theory Comput.*, 2009, pp. 371–380.
[12] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*. Berlin, Germany: Springer, 2006, pp. 265–284.
[13] I. Dinur and K. Nissim, "Revealing information while preserving privacy," in *Proc. 22nd ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 2003, pp. 202–210.
[14] C. Dwork and K. Nissim, "Privacy-preserving datamining on vertically partitioned databases," in *Advances in Cryptology–CRYPTO*. Berlin, Germany: Springer, 2004, pp. 528–544.
[15] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.
[16] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *Proc. 51st Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2010, pp. 51–60.
[17] C. Dwork, "Differential privacy, automata, languages and programming," in *Proc. 33rd Int. Colloq. Autom., Lang. Programm.*, 2006, pp. 1–12.
[18] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," *SIAM J. Comput.*, vol. 41, no. 6, pp. 1673–1693, 2012.
[19] A. Gupta, A. Roth, and J. Ullman, "Iterative constructions and private data release," in *Theory of Cryptography*. Berlin, Germany: Springer, 2012, pp. 339–356.
[20] Q. Geng and P. Viswanath. (May 2013). "Optimal noise adding mechanisms for approximate differential privacy." [Online]. Available: https://arxiv.org/abs/1305.1330
[21] Q. Geng and P. Viswanath. (May 2013). "The optimal mechanism in (ϵ, δ)-differential privacy." [Online]. Available: https://arxiv.org/abs/1305.1330
[22] M. Hardt, K. Ligett, and F. McSherry. (Dec. 2010). "A simple and practical algorithm for differentially private data release." [Online]. Available: https://arxiv.org/abs/1012.4763
[23] M. Hardt and G. N. Rothblum, "A multiplicative weights mechanism for privacy-preserving data analysis," in *Proc. 51st Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2010, pp. 61–70.
[24] M. Hardt and A. Roth, "Beyond worst-case analysis in private singular vector computation," in *Proc. 45th Annu. ACM Symp. Symp. Theory Comput.*, 2013, pp. 331–340.
[25] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Proc. 42nd ACM Symp. Theory Comput*, 2010, pp. 705–714.
[26] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *Proc. Int. Conf. Mach. Learn.*, 2015.
[27] S. L. Lauritzen, *Graphical Models*. London, U.K.: Oxford Univ. Press, 1996.
[28] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.
[29] S. Muthukrishnan and A. Nikolov, "Optimal private halfspace counting via discrepancy," in *Proc. 44th Symp. Theory Comput.*, 2012, pp. 1285–1292.
[30] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2007, pp. 94–103.
[31] A. Stam, "Some inequalities satisfied by the quantities of information of fisher and Shannon," *Inf. Control*, vol. 2, no. 2, pp. 101–112, 1959.
[32] S. Verdú and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2165–2166, May 2006.
[33] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *J. Amer. Statist. Assoc.*, vol. 105, no. 489, pp. 375–389, 2010.
[34] R. Zamir, "A proof of the fisher information inequality via a data processing argument," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1246–1250, May 1998.

Peter Kairouz received his Ph.D and M.S. in ECE from the University of Illinois at Urbana Champaign (UIUC) in 2016 and 2012, and his B.E. in ECE from the American University of Beirut in 2010. He is currently a Postdoctoral Researcher at Stanford Unviersity. He was an R&D intern at Qualcomm Research in 2012 and 2013, and a research intern at Google in 2015. He has received numerous scholarships and awards including the 2016 Harold L. Olesen Award for Excellence in Undergraduate Teaching, the SIGMETRICS 2015 Best Paper Award, the 2012 Roberto Padovani Scholarship from Qualcomm's Research Center in 2012, and the 2012 Benjamin Franklin Scholarship. His research interests include statistical data privacy and security, machine learning, and big data.

Sewoong Oh received his Ph.D. from the department of Electrical Engineering at Stanford University in 2011. He is an Assistant Professor of Industrial and Enterprise Systems Engineering at the University of Illinois at Urbana Champaign. He was a Postdoctoral Researcher at the Laboratory for Information and Decision Systems (LIDS) at MIT. His research interests are in statistical inference and privacy. He is a recipient of the NSF CAREER award (2016) and GOOGLE Faculty Research Award (2016).

Pramod Viswanath received his Ph.D. degree in EECS from the University of California at Berkeley, Berkeley, in 2000. He was a Member of Technical Staff at Flarion Technologies until August 2001 before joining the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign (UIUC), Urbana. Dr. Viswanath is a recipient of the Xerox Award for Faculty Research from the College of Engineering at UIUC (2010), the Eliahu Jury Award from the Electrical Engineering and Computer Science Department of the University of California at Berkeley (2000), the Bernard Friedman Award from the Mathematics Department of the University of California at Berkeley (2000), and the National Science Foundation (NSF) CAREER Award (2003). He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY for the period 2006-2008.