arXiv:2610.33333v1 [cs.LG] 23 Oct 2026

Gaussian Approximation for Composition and Subsampling of Differentially Private Algorithms

Vitaly Feldman
Apple

Moshe Shenfeld*
The Hebrew University of Jerusalem

Abstract

The privacy analysis of differentially private stochastic gradient descent (DP-SGD) relies on bounds for the composition of subsampled mechanisms. The privacy of such a composition is bounded by that of a pair of product distributions, whose privacy loss is a sum of independent terms and has been analyzed via a variety of notions and concentration results. Yet, existing closed-form bounds either lose a logarithmic factor or require restrictive conditions on the noise scale and delicate analysis of subsampling. Further, while it is well known that sums of independent random variables are approximately Gaussian, existing CLT-based techniques give an approximation with a large error in δ.

We show that the privacy loss of this pair is approximately Gaussian as long as each step has a small privacy loss except with small probability. Specifically, its privacy loss distribution is close to a Gaussian with the same mean and variance, after truncating the rare large privacy losses. The closeness in both directions is up to a shift of ε that grows only as $\ln(1/\beta)$ for an additive error β in δ and is a lower-order term in typical applications. In particular, the composition satisfies Gaussian differential privacy up to a lower-order shift of ε and the additive error β. We also give a simple way to analyze the resulting variance-based bounds for subsampled mechanisms which relies on the χ^2 divergence of a single step. For the Gaussian mechanism at sampling rate $1/t$ and $t > \ln^3(1/\delta)$, this gives the predicted rate for $\sigma \gtrsim \max\{1, \sqrt{\ln(1/\delta)}/\ln(t)\}$, whereas the standard Rényi-divergence analyses need $\sigma \gtrsim \sqrt{t \ln(1/\delta)}/\ln(t)$. We extend these approximation and asymptotic results to random allocation (balls-and-bins sampling). As another application, we show that, for bounded privacy losses, a tight Gaussian approximation of the type we give explains the empirically observed high accuracy of the connect-the-dots discretization of privacy loss distributions.

1 Introduction

The privacy analysis of multi-step differentially private (DP) algorithms relies on composition of privacy losses of individual steps. For efficiency purposes each step is often used on a random subsample of the data with the analysis relying on privacy amplification by subsampling. In differentially private stochastic gradient descent (DP-SGD) [ACG+16; BST14; SCS13], for example, each step releases the sum of the clipped gradients of a batch with Gaussian noise of scale σ. It is typically analyzed with *Poisson subsampling* in which each element joins each batch independently with probability λ. The privacy of t steps is then upper bounded by that of a pair of t-fold product distributions [DRS22; ZDW22] and the privacy loss of such a pair is a sum of t independent random variables. Such sums have been analyzed through concentration inequalities [CR21; DRV10] and through several privacy notions, such as Rényi differential privacy [Mir17], concentrated differential privacy [DR16] and its variants [BDRS18; BS16], and Gaussian differential privacy (GDP) [DRS22].

*Work partially done while author was an intern at Apple.

For $\sigma \geq 1$ the variance of this sum predicts (ε, δ)-DP with ε of order $\lambda\sqrt{t\ln(1/\delta)}/\sigma$, which we call the *central limit rate*. Several closed-form bounds at this rate are known and used in theoretical analyses of private learning [BFGT20; BFTT19; Ste25]. However, the known closed-form bounds give the central limit rate only in restricted regimes, and the black-box ones lose additional factors. If each step is shown to be $(\varepsilon_0, \delta_0)$-DP, amplified by subsampling and then combined by advanced composition [DRV10; KLNRS08], the resulting bound exceeds the central limit rate by a factor $\sqrt{\ln(1/\delta)}$ at $\lambda = 1/t$. The moments accountant [ACG+16] and its refinement for the sampled Gaussian mechanism [MTZ19], both based on Rényi differential privacy [Mir17], remove this factor through a per-step bound $O(\lambda^2 q/\sigma^2)$ on the Rényi divergence of order q. This bound holds only for orders up to about $\sigma^2 \ln(1/\lambda)$ [ACG+16, Lemma 3], [MTZ19], whereas the conversion to (ε, δ)-DP at the central limit rate uses orders of about $\ln(1/\delta)/\varepsilon$. For $\lambda = 1/t$ the two requirements are compatible only if $\sigma \gtrsim \sqrt{t\ln(1/\delta)}/\ln(t)$. Practical applications typically rely on σ being a small constant, and thus these closed-form bounds do not give the central limit rate for such σ.

Further, central limit theorems for the privacy loss [DRS22; SMM19] predict that the composition is approximately Gaussian differentially private. Yet, their quantitative (Berry–Esseen) form [DRS22, Theorem 3.4] has an additive error in δ, of order $1/\sqrt{t}$ at $\lambda = 1/t$. Thus it implies the central limit rate only for fixed and relatively large δ.

This leads to the questions of whether the Gaussian approximation predicted by the central limit theorems holds beyond the large error in δ and whether the closed-form bounds at the central limit rate hold under mild and practically realistic conditions on σ and δ.

1.1 Our contributions

We give new and relatively simple bounds on the concentration of the privacy loss which imply that the answer to both questions is yes under mild conditions on the privacy loss of each step: if the privacy loss of each step is small except with small probability then the privacy loss of the product pair (in either direction) is close to a Gaussian random variable with the same mean and variance, after truncating the rare large privacy losses. Specifically, if the privacy loss of every step deviates from its mean by at most Δ, then for every β the distribution function of the privacy loss of the product pair is bounded from above and from below by that of the Gaussian random variable with the same mean μ and variance V, up to a *loss shift* of $O(\Delta \ln(1/\beta))$ in its argument and a *tail shift* β in its value (Theorem 3.3). We say that such a privacy loss is *Gaussian-concentrated*. For a pair with a Gaussian-concentrated privacy loss, the privacy profile is within the same shifts (in ε and in δ) of the privacy profile of $\sqrt{V}$-GDP biased in ε by the excess $\mu - V/2$ of the mean over half the variance (Lemma 2.8). Importantly, the loss shift does not depend on the number of steps t and depends only logarithmically on the tail shift. Therefore, in contrast to the Berry–Esseen bounds, the approximation covers values of δ below every fixed power of $1/t$ (as in the second case of Corollary 4.6). For privacy losses that are bounded only with high probability we truncate the privacy loss of each step. The truncation uses part of the tail shift β for the probability of the truncated events and replaces the mean and the variance by those of the truncated privacy loss.

For upper bounds on the privacy profile of an algorithm we can combine the excess $\mu - V/2$ of the mean with the loss shift to say that an algorithm is (α, β)-*shifted* $\sqrt{V}$-*GDP* if its privacy profiles are at most that of $\sqrt{V}$-GDP shifted by α in ε, plus β (Definition 2.9). Thus an adaptive composition of steps as above is $(\mu - V/2 + O(\Delta \ln(1/\beta)), \beta)$-shifted $\sqrt{V}$-GDP. Both parts of its shift are lower-order terms in typical applications (e.g., Corollary 4.6). Shifted GDP is preserved by post-processing and therefore transfers from the product of the dominating pairs of the steps to their adaptive composition (although the resulting upper bounds need not be tight). In Appendix B we compare shifted GDP with zero-concentrated differential privacy [BS16], truncated

concentrated differential privacy [BDRS18] and concentrated differential privacy [DR16].

From Bernstein's inequality to a Gaussian coupling. As a warm-up, in Section 3.1 we show that the standard Bernstein concentration inequality can be used to derive the upper tail part of our concentration result (up to a second-order factor). This bounds the privacy profile by δ at $\varepsilon = \mu + \sqrt{2V\ln(1/\delta)} + O(\Delta\ln(1/\delta))$, where μ is the mean of the total privacy loss and Δ bounds the deviation of each privacy loss above its mean (Theorem 3.1). After truncation, this is essentially all we need for upper bounds at the central limit rate (Theorem 4.3). A tail inequality cannot, however, bound a privacy profile from below. For the two-sided result we couple the total privacy loss with a Gaussian random variable [Sak96; Zai25]. If Δ also bounds the deviation of each privacy loss below its mean, the error of this coupling is $O(\Delta\ln(1/\beta))$ except with probability β, whatever the number of steps (Lemma 3.2).

Subsampling through the χ^2 divergence. In Section 4 we show a way to simplify the resulting concentration bounds by using the χ^2 divergence $\tilde{V} \coloneqq \chi^2_{P,Q}$ of the pair (P, Q) that dominates one step and a tail condition. This result is based on the observation that the χ^2 divergence bounds the first two moments of the privacy loss (Lemma D.1), and subsampling at rate λ multiplies it by exactly λ^2 (in the add direction, by at most $\lambda^2/(1-\lambda)$). Together with the basic Bernstein concentration inequality, this gives $\varepsilon = O(t\lambda^2\tilde{V} + \lambda\sqrt{t\tilde{V}\ln(1/\delta)})$ for t subsampled steps (Theorem 4.3). Using the Gaussian concentration, we also show that t subsampled steps are shifted $\sqrt{W}$-GDP for $W = t\lambda^2\tilde{V}/(1-\lambda)$ (Corollary 4.4).

The subsampled Gaussian mechanism. In Section 4.2 we apply the bounds for composition of subsampled algorithms to the Gaussian noise addition mechanism on which the analysis of DP-SGD relies. Our bounds give explicit and nearly tight conditions on σ, t and δ, under which t steps at sampling rate λ satisfy $\varepsilon = O(t\lambda^2/\sigma^2 + \lambda\sqrt{t\ln(1/\delta)}/\sigma)$ ((8)). For example, for all sufficiently large t the condition holds when $\sigma \gtrsim \max\{1, \sqrt{\ln(1/\delta)}/\ln(t)\}$ and $t > \ln^3(1/\delta)$, and for a fixed $\sigma \geq 1$ it holds for all $\delta \geq \exp(-\Omega(\sigma^2\ln^2(t)))$. For comparison, we note that at $\lambda = 1/t$ our results hold even for σ smaller by a factor of at least order $\sqrt{t}/\ln(t)$ than that of the closed-form analyses based on Rényi divergence. Indeed, these analyses bound exponential moments of high order of the privacy loss, whereas we need only its first two moments and a tail bound (see Appendix A for a more detailed comparison). Further, the likelihood ratio of the Gaussian mechanism is within Δ of 1 except with probability at most $\beta/(2t)$, for some Δ with $\ln(1+\Delta)$ of order $\sqrt{\ln(t/\beta)}/\sigma$ ((7)). Thus DP-SGD is shifted $\sqrt{W}$-GDP for $W = t\lambda^2(e^{1/\sigma^2}-1)/(1-\lambda)$, with a shift that is of lower order when $\Delta\sqrt{\ln(1/\beta)}$ is small compared with $\sqrt{t}/\sigma$ (Corollary 4.5).

Random allocation. In 1-out-of-t *random allocation* (also known as balls-and-bins sampling), every element is used in exactly one step, chosen independently and uniformly at random [CGH+25; CGHST25]. Its privacy loss is not a sum of independent terms. Feldman and Shenfeld [FS25] analyze it through a reduction to Poisson subsampling. The resulting bound for the Gaussian mechanism therefore inherits the assumption on σ of the analysis of Poisson subsampling, and the reduction needs the additional assumption $\sigma \gtrsim \sqrt{\ln(t/\delta)}$.

Our Gaussian approximation extends to random allocation even more directly than to Poisson subsampling. The likelihood ratio of the dominating pair of random allocation [FS25] is an average of t independent one-step likelihood ratios, each with mean 1 and variance $\tilde{V}$. In Section 5 we couple this average, rather than its logarithm, with a Gaussian random variable with the same mean and variance. We show that if each one-step likelihood ratio is within Δ of 1, then the privacy losses of the pair are Gaussian-concentrated with mean $\tilde{V}/(2t)$ and variance $\tilde{V}/t$, as

for $\sqrt{\tilde{V}/t}$-GDP, with a tail shift β and a loss shift of order $(\Delta/t)\ln(1/\beta)$ whenever the latter is sufficiently small (Theorem 5.2). For the Gaussian mechanism, random allocation is shifted $\sqrt{\tilde{V}/(t-1)}$-GDP, with the same parameter as Poisson subsampling at rate $1/t$. Thus we obtain the central limit rate under a condition of the same form as for Poisson subsampling, with no additional assumption, and the same asymptotic upper bounds (Corollaries 5.3 to 5.5).

Connect-the-dots discretization. As an additional application of our techniques we consider the connect-the-dots (CtD) discretization [DGKKM22], which Google's `dp_accounting` library [Goo26] uses by default for the privacy loss distributions (PLDs) of continuous mechanisms. Its authors found that it "can work with much larger discretization intervals while keeping a similar error bound compared to previous approaches" such as privacy buckets [MM18], and asked "where these improvements come from" [DGKKM22, Section 7] (see Figure 1 for a numerical comparison). A single discretized step can have ε larger by almost the grid spacing h (Claim F.1). A simple analysis thus bounds the increase of ε for the t-step composition only by th. The error analysis in [GLW21, Lemma 5.3], based on a coupling argument, gives an upper bound of $\tilde{O}(\sqrt{t}\,h)$, and it applies to connect-the-dots discretization. Our tight bounds imply that under composition the privacy profile is determined by the mean and the variance of the privacy loss up to the shifts of the Gaussian concentration. We show that the discretization changes the mean and the variance only by $O(h^2)$ per step (Lemma 6.2) and thus, for bounded privacy losses, increases the ε of the t-step composition by $O(th^2)$, up to logarithmic factors and the accuracy of the Gaussian concentration (Corollary 6.4), improving quadratically on the bound in [GLW21].

1.2 Related work

Composition is the most basic and the most extensively studied property of differential privacy [DR14]. Advanced composition [DRV10] shows that the privacy parameter of t steps grows as $\sqrt{t}$ rather than as t, and the optimal composition of $(\varepsilon_0, \delta_0)$-DP steps was characterized in [KOV15]. Concentrated differential privacy [DR16] and its variants [BDRS18; BS16], Rényi differential privacy [Mir17] and Gaussian differential privacy [DRS22] describe privacy by quantities that compose simply. We refer to [Ste25] for a detailed survey.

The analysis of DP-SGD [ACG+16; BST14; SCS13], which combines privacy amplification by subsampling [BBG18; KLNRS08] with composition over many steps, has driven the development of both the theory and the numerical methods. The moments accountant [ACG+16] and the analyses that followed it [MTZ19; ZW19] bound the Rényi divergences of the subsampled Gaussian mechanism. Numerical accountants compute the privacy profile of the composition from the PLDs of the steps [DGKKM22; GLW21; KH21; KJH20; KJPH21; MM18], and dominating pairs [ZDW22] reduce adaptive compositions to products of distributions. Central limit theorems for the privacy loss go back to [SMM19], and Dong et al. [DRS22, Theorem 3.5] prove that the composition of many mechanisms that are each close to perfect privacy is asymptotically GDP (see also [BDLS20]), although Gopi et al. [GLW21] point out that the GDP accountant can significantly underreport the true ε of DP-SGD. Refined approximations use Edgeworth expansions [WGZSSW26; ZDLS20] or the saddle-point method [AGACKS23]. We compare these approaches with ours in Appendix A.

Our approach is closer in spirit to concentration-based analyses of composition, such as the proof of advanced composition by Azuma's inequality [DRV10], concentrated differential privacy [BS16; DR16], and martingale inequalities for privacy losses with subgaussian increments [CR21]. Such subgaussian bounds, however, do not capture amplification by subsampling (Appendix B).

Practical implementations of DP-SGD usually process the data in passes in which every element is used once, whereas the privacy analysis assumes Poisson subsampling, in which batch sizes

vary and an element can be used in any number of steps. In random allocation or *balls-and-bins* sampling every element is used in exactly one of the t steps, chosen uniformly at random. Random allocation was proposed for DP-SGD [CGH+25], for matrix mechanisms [CGHST25] and for private low-communication distributed mean estimation [AFKRT26]. Feldman and Shenfeld [FS25] gave its dominating pair, from which our analysis starts. Several theoretical analyses, including that of its Rényi divergence, were given in [DCÖ25; FS25; SK26]. Feldman and Shenfeld [FS26] showed that its PLD can be computed efficiently and accurately, and Dong and Özgür [DÖ26] compared the sampling schemes asymptotically.

2 Preliminaries

We denote the domain of elements by $\mathcal{X}$ and the set of outputs by $\mathcal{Y}$. An algorithm M maps every dataset $\boldsymbol{s} \in \mathcal{X}^*$ to a distribution over $\mathcal{Y}$. For a distribution P over $\mathcal{Y}$ we write $P(\omega)$ for the probability of ω or, in the continuous case, for its density with respect to a fixed dominating measure. All spaces are standard Borel and all randomized mappings are measurable.

2.1 Privacy losses and dominating pairs

We augment the domain to $\mathcal{X}_\perp \coloneqq \mathcal{X} \cup \{\perp\}$ with a null element $\perp$ such that $M(\boldsymbol{s}, \perp) = M(\boldsymbol{s})$ for all $\boldsymbol{s} \in (\mathcal{X}_\perp)^*$, and call two datasets in $(\mathcal{X}_\perp)^*$ *adjacent* if one is obtained from the other by replacing a single element by $\perp$ (zero-out adjacency), and write their induced output distributions as $M(\boldsymbol{s}, x)$ and $M(\boldsymbol{s}, \perp)$, where $x \in \mathcal{X}$ is the *distinguished* element.

Definition 2.1 (Privacy loss and hockey-stick divergence [BKOZB12; DR16])**.** Let P and Q be distributions over $\mathcal{Y}$. The *privacy loss* of $\omega \in \mathcal{Y}$ is $\ell(\omega; P, Q) \coloneqq \ln(P(\omega)/Q(\omega))$, which is $+\infty$ when $Q(\omega) = 0$. The *privacy loss random variable* $L_{P,Q}$ is $\ell(\omega; P, Q)$ for $\omega \sim P$, and its distribution is the *privacy loss distribution (PLD)*. Denoting $[\,\cdot\,]_+ \coloneqq \max\{\cdot, 0\}$, for any $\kappa \geq 0$, the *κ-hockey-stick divergence* is

$$\boldsymbol{H}_\kappa(P \parallel Q) \coloneqq \int_{\mathcal{Y}} [P(\omega) - \kappa Q(\omega)]_+ d\omega = \mathbb{E}\Big[\Big[1 - \kappa e^{-L_{P,Q}}\Big]_+\Big].$$

The hockey-stick divergence of a pair depends only on the distribution of its privacy loss. Following [FS26, Definition 2.2], for a random variable X with values in $[-\infty, \infty]$ and $\kappa \geq 0$ we write $\boldsymbol{H}_\kappa(X) \coloneqq \mathbb{E}\big[[1 - \kappa e^{-X}]_+\big]$, so that $\boldsymbol{H}_\kappa(P \parallel Q) = \boldsymbol{H}_\kappa(L_{P,Q})$ by Definition 2.1. For every $l \in \mathbb{R}$ we have that $\boldsymbol{H}_{e^\varepsilon}(X + l) = \boldsymbol{H}_{e^{\varepsilon - l}}(X)$.

We denote the mean and the variance of $L_{P,Q}$ by $\mu_{P,Q}$ and $V_{P,Q}$. For a set E, we denote by $\mu'_{P,Q}$ and $V'_{P,Q}$ the mean and the variance of the *truncated privacy loss* $\ell(\omega; P, Q)\,\mathbb{1}\{\omega \in E\}$ for $\omega \sim P$, which is set to zero outside E; the set is specified wherever these moments appear.

We say that distributions P and Q are (ε, δ)-*close* if $\boldsymbol{H}_{e^\varepsilon}(P \parallel Q) \leq \delta$ and $\boldsymbol{H}_{e^\varepsilon}(Q \parallel P) \leq \delta$. An algorithm is (ε, δ)-*differentially private (DP)* [DKMMN06] if for all adjacent datasets their induced output distributions are (ε, δ)-close. We refer to the case where the first dataset has $\perp$ in place of the distinguished element as the *add* direction and to the reverse as the *remove* direction [FS25]. This definition is closely-related to differential privacy with respect to adding or removing an element (with $\perp$ corresponding to removal). The *privacy profile* of a pair (P, Q) is the function $\varepsilon \mapsto \boldsymbol{H}_{e^\varepsilon}(P \parallel Q)$ [BBG18]. We denote by $\varepsilon_{P,Q}(\delta)$ the smallest $\varepsilon \geq 0$ for which P and Q are (ε, δ)-close, and by $\varepsilon_M(\delta)$ the smallest $\varepsilon \geq 0$ for which an algorithm M is (ε, δ)-DP.

For $\kappa = e^\varepsilon$, the random variable $[1 - e^{\varepsilon - L_{P,Q}}]_+$ in Definition 2.1 is at most the indicator of the event $\{L_{P,Q} > \varepsilon\}$. Therefore we have the *tail bound*

$$\boldsymbol{H}_{e^\varepsilon}(P \parallel Q) \leq \mathbb{P}(L_{P,Q} > \varepsilon). \tag{1}$$

For P absolutely continuous with respect to Q we denote the likelihood ratio by $r(\omega; P, Q) \coloneqq e^{\ell(\omega;P,Q)} = P(\omega)/Q(\omega)$, and by $R_{P,Q} \coloneqq e^{L_{P,Q}}$ its value for $\omega \sim P$. For $\omega \sim Q$ the likelihood ratio $r(\omega; P, Q) = e^{-\ell(\omega;Q,P)}$ is distributed as $1/R_{Q,P}$, it has mean 1, and its variance is the χ^2 *divergence*

$$\chi^2_{P,Q} \coloneqq \mathop{\mathbb{E}}_{\omega\sim Q}\Big[(r(\omega; P, Q) - 1)^2\Big] = \mathop{\mathbb{E}}_{\omega\sim Q}\Big[r(\omega; P, Q)^2\Big] - 1.$$

Definition 2.2 (Dominating pair [ZDW22])**.** A pair of distributions (P, Q) *dominates* a pair (P', Q') if $\boldsymbol{H}_\kappa(P' \parallel Q') \le \boldsymbol{H}_\kappa(P \parallel Q)$ for every $\kappa \ge 0$. It *dominates* an algorithm M if it dominates $(M(\boldsymbol{s}, x), M(\boldsymbol{s}, \perp))$ for all adjacent datasets $(\boldsymbol{s}, x)$ and $(\boldsymbol{s}, \perp)$.

We use the following characterization of domination by post-processing.

Lemma 2.3 ([KOV15, Theorem 2.5])**.** *The pair (P, Q) dominates (P', Q') if and only if there is a randomized mapping φ with $\varphi(P) = P'$ and $\varphi(Q) = Q'$.*

In the *t-step adaptive composition* of algorithms $M_1, \ldots, M_t$, the i-th algorithm also receives the outputs $y_1, \ldots, y_{i-1}$ of the previous steps, and the composition outputs $(y_1, \ldots, y_t)$, where y_i is drawn from the output distribution of $M_i(\cdot\,; y_1, \ldots, y_{i-1})$ on the dataset. For $\lambda \in [0, 1]$, the *λ-subsampled* algorithm M_λ runs M on the dataset in which every element is independently replaced by $\perp$ with probability $1 - \lambda$ (Poisson subsampling); in a composition of subsampled algorithms, the subsampling is independent at every step. For distributions P and Q we similarly write $P_\lambda \coloneqq (1 - \lambda)Q + \lambda P$, and we denote by Q^t the product of t copies of Q (and use the same notation for other distributions). We use the following standard forms of composition and subsampling for dominating pairs.

Lemma 2.4 (Composition and subsampling [ZDW22, Theorems 10 and 11(1)])**.** *Let $t \in \mathbb{N}$, and let $M_1, \ldots, M_t$ be algorithms, where M_i also receives the outputs of the previous steps.*

1. *(Composition) If for every $i \in [t]$ and all outputs $y_1, \ldots, y_{i-1}$ the pair (P_i, Q_i) dominates $M_i(\cdot\,; y_1, \ldots, y_{i-1})$, then the pair $(P_1 \times \cdots \times P_t,\ Q_1 \times \cdots \times Q_t)$ dominates the t-step adaptive composition of $M_1, \ldots, M_t$.*
2. *(Subsampling) If (P, Q) dominates an algorithm M, then for every $\lambda \in [0, 1]$ the pair (P_λ, Q) dominates M_λ.*

In particular, Lemma 2.4 implies that if (P, Q) dominates $M_i(\cdot\,; y_1, \ldots, y_{i-1})$ for every $i \in [t]$ and all outputs $y_1, \ldots, y_{i-1}$, then (P_λ^t, Q^t) dominates the t-step adaptive composition of the λ-subsampled algorithms. The pair (P, Q) dominates (P', Q') if and only if (Q, P) dominates (Q', P') [ZDW22, Lemma 28], and thus (Q^t, P_λ^t) dominates the pairs of output distributions in the add direction.

The *Gaussian mechanism* with noise scale σ and unit norm bound, for a function $f\colon \mathcal{X}_\perp \to \mathbb{R}^d$, is the algorithm

$$M(\boldsymbol{s}) = \mathcal{N}\Big(\sum_{z\in\boldsymbol{s}} f(z),\, \sigma^2 I_d\Big), \qquad \|f(z)\|_2 \le 1, \qquad f(\perp) = 0, \tag{2}$$

where in an adaptive composition the function f may depend on the outputs of the previous steps (for example, $f(z)$ is the clipped gradient of the loss on z at the current model). We refer to $(\mathcal{N}(1, \sigma^2), \mathcal{N}(0, \sigma^2))$ as the *Gaussian pair* and use the following standard fact.

Claim 2.5 ([BW18, Lemma 3])**.** *The Gaussian pair dominates the Gaussian mechanism.*

2.2 Concentration and shifted GDP

We characterize the closeness of the composed mechanisms to a Gaussian in two ways, by comparing their privacy profiles to that of the Gaussian mechanism, and by comparing their PLDs to a Gaussian distribution. To do so, we denote by Φ the distribution function of the standard normal distribution, and by $\overline{\Phi} \coloneqq 1 - \Phi$ its tail. For every $w \geq 0$ we have that $\overline{\Phi}(w) \leq e^{-w^2/2}/2$. We also use the following one-sided form of Bernstein's inequality.

Lemma 2.6 (Bernstein's inequality [BLM13, Theorem 2.10])**.** *Let $X_1, \ldots, X_t$ be independent mean-zero random variables such that $X_i \leq \Delta$ almost surely and $\sum_i \mathbb{E}[X_i^2] \leq V$. Then for every $\delta \in (0, 1]$,*

$$\mathbb{P}\left(\sum_{i=1}^{t} X_i > \sqrt{2V \ln(1/\delta)} + \frac{\Delta \ln(1/\delta)}{3}\right) \leq \delta.$$

For $V > 0$ we call $\mathcal{G}_V \coloneqq (\mathcal{N}(1, 1/V), \mathcal{N}(0, 1/V))$ the *Gaussian reference* with parameter V and denote its privacy profile by g_V. Its privacy loss is distributed as $\mathcal{N}(V/2, V)$. By [BW18, Theorem 8],

$$g_V(\varepsilon) = \Phi\left(\frac{\sqrt{V}}{2} - \frac{\varepsilon}{\sqrt{V}}\right) - e^{\varepsilon}\,\Phi\left(-\frac{\sqrt{V}}{2} - \frac{\varepsilon}{\sqrt{V}}\right), \tag{3}$$

and we let $\mathcal{G}_0$ be a pair of identical distributions, whose privacy profile is $g_0(\varepsilon) = [1 - e^{\varepsilon}]_+$. The reflection $\omega \mapsto 1 - \omega$ swaps the two distributions of $\mathcal{G}_V$ and therefore g_V is also the privacy profile of the reversed pair. An algorithm is *$\sqrt{V}$-Gaussian differentially private ($\sqrt{V}$-GDP)* if and only if it is $(\varepsilon, g_V(\varepsilon))$-DP for every $\varepsilon \geq 0$ [DRS22, Corollary 2.13]. For $0 < V' < V$, adding independent noise $G \sim \mathcal{N}(0, 1/V' - 1/V)$ maps $\mathcal{G}_V$ to $\mathcal{G}_{V'}$, and a constant map sends $\mathcal{G}_V$ to a pair of identical distributions. Therefore $\mathcal{G}_V$ dominates $\mathcal{G}_{V'}$ for $0 \leq V' < V$ by Lemma 2.3, that is, $g_V(\varepsilon)$ is nondecreasing in V. We denote by $\varepsilon_{\mathrm{G}}(V, \delta) \coloneqq \inf\{\varepsilon \geq 0 : g_V(\varepsilon) \leq \delta\}$ the smallest ε for which $\sqrt{V}$-GDP implies (ε, δ)-DP. By the tail bound (1) and the bound on $\overline{\Phi}$, we have that $\varepsilon_{\mathrm{G}}(V, \delta) \leq V/2 + \sqrt{2V \ln(1/\delta)}$ for $\delta \in (0, 1)$.

For $G \sim \mathcal{N}(\mu, V)$ the random variable $G - \mu + V/2$ is distributed as the privacy loss of $\mathcal{G}_V$, and therefore

$$\boldsymbol{H}_{e^{\varepsilon}}(G) = g_V\left(\varepsilon - \mu + \frac{V}{2}\right) \qquad \text{for every } \varepsilon \in \mathbb{R}. \tag{4}$$

Thus the excess $\mu - V/2$ of the mean of a Gaussian random variable over half its variance acts as a bias of ε in the privacy profile of $\sqrt{V}$-GDP.

We compare a privacy loss with a Gaussian random variable through their tails, using approximate stochastic domination [FS26, Definition 4.1].

Definition 2.7 (Gaussian concentration and approximation)**.** Let $\alpha, \beta \geq 0$. A random variable Y *(α, β)-approximately stochastically dominates* a random variable X, denoted by $X \preceq_{(\alpha,\beta)} Y$, if

$$\mathbb{P}(X > \ell) \leq \mathbb{P}(Y > \ell - \alpha) + \beta \qquad \text{for every } \ell \in [-\infty, \infty].$$

For $\mu \in \mathbb{R}$ and $V \geq 0$, a random variable X is *(α, β)-approximately (μ, V)-Gaussian-concentrated* if $G \preceq_{(\alpha,\beta)} X \preceq_{(\alpha,\beta)} G$ for $G \sim \mathcal{N}(\mu, V)$. We refer to α as the *loss shift* and to β as the *tail shift*.

Note that a coupling of X and G for which $|X - G| \leq \alpha$ except with probability β (a coupling approximation in the terminology of [GLW21, Definition 5.1]) gives both relations, since $\{X > \ell\} \subseteq \{G > \ell - \alpha\} \cup \{|X - G| > \alpha\}$. Also, directly from the definition, if $X \preceq_{(\alpha_1,\beta_1)} Y$ and $Y \preceq_{(\alpha_2,\beta_2)} Y'$, then $X \preceq_{(\alpha_1+\alpha_2,\beta_1+\beta_2)} Y'$. We use the following bound on the hockey-stick divergence.

If $X \preceq_{(\alpha,\beta)} Y$, then $\boldsymbol{H}_{e^\varepsilon}(X) \le \boldsymbol{H}_{e^{\varepsilon-\alpha}}(Y) + \beta$ for every $\varepsilon \in \mathbb{R}$ [FS26, Claim 4.2]. Applying this fact in both directions to a privacy loss and a Gaussian random variable and combining it with (4), we obtain the following two-sided bound.

Lemma 2.8 (Privacy profiles of Gaussian-concentrated privacy losses)**.** *Let (P, Q) be a pair of distributions whose privacy loss $L_{P,Q}$ is (α, β)-approximately (μ, V)-Gaussian-concentrated. Then for every $\varepsilon \in \mathbb{R}$,*

$$g_V\left(\varepsilon - \mu + \frac{V}{2} + \alpha\right) - \beta \le \boldsymbol{H}_{e^\varepsilon}(P \parallel Q) \le g_V\left(\varepsilon - \mu + \frac{V}{2} - \alpha\right) + \beta.$$

The upper bound of Lemma 2.8 is the privacy profile of $\sqrt{V}$-GDP shifted in ε and in δ, forming a "shifted" GDP profile.

Definition 2.9 (Shifted GDP)**.** Let $V, \beta \ge 0$ and $\alpha \in \mathbb{R}$. An algorithm M is (α, β)-*shifted* $\sqrt{V}$-*GDP* if for any $\boldsymbol{s} \in \mathcal{X}^*, x \in \mathcal{X}$ and every $\varepsilon \in \mathbb{R}$,

$$\max\{\boldsymbol{H}_{e^\varepsilon}(M(\boldsymbol{s}, x) \parallel M(\boldsymbol{s}, \perp)), \boldsymbol{H}_{e^\varepsilon}(M(\boldsymbol{s}, \perp) \parallel M(\boldsymbol{s}, x))\} \le g_V(\varepsilon - \alpha) + \beta.$$

For convenience, we also say that a pair (P, Q) is (α, β)-shifted $\sqrt{V}$-GDP if the same holds for $\boldsymbol{H}_{e^\varepsilon}(P \parallel Q)$ and $\boldsymbol{H}_{e^\varepsilon}(Q \parallel P)$. Note that an algorithm is $(0, 0)$-shifted $\sqrt{V}$-GDP if and only if it is $\sqrt{V}$-GDP, and that the shift α can be negative when $\beta > 0$. Since g_V is continuous and nonincreasing, if (P, Q) is (α, β)-shifted $\sqrt{V}$-GDP, then for every $\beta < \delta \le 1$ the distributions P and Q are (ε, δ)-close for

$$\varepsilon = \left[\varepsilon_{\mathrm{G}}(V, \delta - \beta) + \alpha\right]_+. \tag{5}$$

Note that Gaussian concentration and GDP differ in three aspects. Gaussian concentration is a property of the PLD, requires both upper and lower bounds, and uses an arbitrary Gaussian distribution as reference. On the other hand, GDP characterizes the privacy profile, uses only an upper bound, and uses the PLD of the Gaussian mechanism as reference, for which the mean of the PLD is half of its variance. Nevertheless, the two are closely related. By Lemma 2.8, if the privacy loss of (P, Q) is (α, β)-approximately (μ, V)-Gaussian-concentrated, then $\boldsymbol{H}_{e^\varepsilon}(P \parallel Q) \le g_V(\varepsilon - \mu + V/2 - \alpha) + \beta$ for every ε, and if the same holds for the privacy loss of (Q, P), then (P, Q) is $(\mu - V/2 + \alpha, \beta)$-shifted $\sqrt{V}$-GDP.[1]

Shifted GDP is preserved by post-processing, that is, by domination (Lemma 2.3). Namely, if (P, Q) is (α, β)-shifted $\sqrt{V}$-GDP and dominates (P', Q'), then (P', Q') is also (α, β)-shifted $\sqrt{V}$-GDP. In particular, by Lemma 2.4, the t-step adaptive composition of algorithms whose steps are dominated by (P, Q) for all outputs of the previous steps is (α, β)-shifted $\sqrt{V}$-GDP whenever the pair (P^t, Q^t) is. These upper bounds need not be tight. Gaussian concentration of the privacy loss of the product pair bounds its privacy profiles from both sides, whereas those of the algorithm are bounded only from above.

3 Gaussian approximation of compositions

In this section we show that the privacy loss of a composition of steps with small privacy losses is Gaussian-concentrated, with the mean and the variance of its truncated privacy loss. We first derive an upper bound from Bernstein's inequality and then prove the two-sided concentration for privacy losses that are bounded except with small probability. We defer the proofs to Appendix C.

Throughout this section, (P, Q) is a pair of mutually absolutely continuous distributions. By Lemma 2.4, an upper bound on the privacy profile of (P^t, Q^t) also holds for the t-step adaptive

[1] In fact, this holds even if only the upper bound part of the approximation holds.

composition of algorithms whose steps are dominated by (P, Q) for all outputs of the previous steps. The privacy loss of a product pair is the sum of the privacy losses of its coordinates: for $\omega = (\omega_1, \ldots, \omega_t)$,

$$\ell\Big(\omega; P^t, Q^t\Big) = \sum_{i=1}^{t} \ell(\omega_i; P, Q). \tag{6}$$

Thus L_{P^t,Q^t} is a sum of t independent copies of $L_{P,Q}$, and therefore $\mu_{P^t,Q^t} = t \cdot \mu_{P,Q}$ and $V_{P^t,Q^t} = t \cdot V_{P,Q}$. The proofs of this section do not use that the steps are identical, and by the linearity of the expectation and of the variance of independent random variables, the results extend to the products of different pairs $(P_1, Q_1), \ldots, (P_t, Q_t)$, with the sums of the means and of the variances of their privacy losses in place of μ_{P^t,Q^t} and V_{P^t,Q^t}. We state the results of this section for the pair (P^t, Q^t). Applied to the reversed pair (Q, P), whose privacy loss is $-\ell(\omega; P, Q)$ for $\omega \sim Q$, they give bounds of the same form in the add direction, with the assumptions and the moments of this privacy loss.

3.1 Warm-up: Bernstein's inequality

We first apply Bernstein's inequality (Lemma 2.6) to t independent copies of the centered privacy loss $L_{P,Q} - \mu_{P,Q}$, and the tail bound (1) then gives the following upper bound.

Theorem 3.1 (Composition of bounded privacy losses)**.** *Let $\Delta \geq 0$ and assume that $L_{P,Q} - \mu_{P,Q} \leq \Delta$ almost surely. Then for every $\delta \in (0, 1)$,*

$$\boldsymbol{H}_{e^\varepsilon}\Big(P^t \;\Big\|\; Q^t\Big) \leq \delta \qquad \textit{for} \qquad \varepsilon = \mu_{P^t,Q^t} + \sqrt{2V_{P^t,Q^t} \ln(1/\delta)} + \frac{\Delta}{3} \ln(1/\delta).$$

To compare Theorem 3.1 with the Gaussian approximation, we write its bound as

$$\varepsilon = \left(\frac{V_{P^t,Q^t}}{2} + \sqrt{2V_{P^t,Q^t} \ln(1/\delta)}\right) + \left(\mu_{P^t,Q^t} - \frac{V_{P^t,Q^t}}{2}\right) + \frac{\Delta}{3} \ln(1/\delta).$$

The first term is the upper bound on $\varepsilon_{\mathrm{G}}(V_{P^t,Q^t}, \delta)$ given by (3) and the tail bound of the Gaussian. The second term is the excess of the mean over half the variance, which is a bias of the privacy profile itself rather than an error of the bound. By (4), it is the bias in ε between the privacy profile of $\sqrt{V_{P^t,Q^t}}$-GDP and that of a Gaussian random variable with the same mean and variance as the privacy loss. The third term is a loss shift error and depends neither on the variance nor on the number of steps. It is of lower order than the first term when Δ is well below $\sqrt{tV_{P,Q}/\ln(1/\delta)}$, and of the same order when Δ is of order $\sqrt{tV_{P,Q}/\ln(1/\delta)}$, which we call the *balanced cutoff*. Put simply, up to the unavoidable bias in the expected loss, the bound exceeds the smallest ε at which the privacy profile of $\sqrt{V_{P^t,Q^t}}$-GDP is at most δ only by two errors: the third term and the error of the tail bound. Thus Bernstein's inequality essentially suffices for the upper bounds at the central limit rate in our applications. If the privacy losses are bounded only with high probability, we truncate them from above at the balanced cutoff, which increases neither their means nor their second moments, before applying Bernstein's inequality. This keeps the third term of the same order as the first, and changes the sum of the privacy losses only if the privacy loss of some step exceeds the cutoff, and we add the probability of this event to δ. In Section 4 we do so for subsampled mechanisms, whose privacy losses may be unbounded (Theorem 4.3).

3.2 Gaussian concentration

To bound the privacy profiles also from below, we couple the privacy loss with a Gaussian random variable. As noted above, such a coupling gives Gaussian concentration, with the bound on its error as the loss shift and the probability that this bound fails as the tail shift.

The key concentration result we use is the following coupling of the sum of bounded random variables with a Gaussian random variable of the same mean and variance. We derive it in Appendix C from the bounds on the moments of the error of a Gaussian coupling in [Sak96, Corollary 4 and inequality (1.6)] (Theorem C.2), and it is also a special case of a more general coupling [Zai25, Theorems 1 and 2].

Lemma 3.2 (Gaussian coupling of bounded summands)**.** *There is a universal constant $C > 0$ such that the following holds. Let $X_1, \dots, X_t$ be independent mean-zero random variables with $|X_i| \le \Delta$ almost surely, and let $S \coloneqq \sum_{i=1}^{t} X_i$. Then for every $0 < \beta \le 1/2$, S can be coupled with $G \sim \mathcal{N}(0, \mathrm{Var}(S))$ so that*

$$\mathbb{P}(|S - G| > C\Delta \ln(2/\beta)) \le \beta.$$

We apply Lemma 3.2 to $L_{P,Q}$ by first truncating the $\beta/(2t)$ tail and then coupling the sum of the truncated privacy losses with a Gaussian random variable.

Theorem 3.3 (Gaussian concentration for compositions)**.** *There is a universal constant $C > 0$ such that for any $\Delta > 0$, $0 < \beta \le 1$ and a set E such that $P(E^c) \le \beta/(2t)$ and the truncated privacy loss $\ell(\omega; P, Q)\, \mathbb{1}\{\omega \in E\}$ for $\omega \sim P$ deviates from its mean $\mu'_{P,Q}$ by at most Δ almost surely, the privacy loss L_{P^t,Q^t} is $(C\Delta \ln(2/\beta), \beta)$-approximately $(t \cdot \mu'_{P,Q}, t \cdot V'_{P,Q})$-Gaussian-concentrated, where $V'_{P,Q}$ is the variance of this truncated privacy loss.*

Note that the loss shift $C\Delta \ln(2/\beta)$ depends neither on the variance nor on the number of steps. If $|L_{P,Q} - \mu_{P,Q}| \le \Delta$ almost surely, then Theorem 3.3 applies for every $0 < \beta \le 1$ with E the whole space, and thus with $\mu'_{P,Q} = \mu_{P,Q}$ and $V'_{P,Q} = V_{P,Q}$.

By Theorem 3.3 and Lemma 2.8, the privacy profile of (P^t, Q^t) is bounded from both sides by that of $\sqrt{t \cdot V'_{P,Q}}$-GDP biased in ε by the excess $t(\mu'_{P,Q} - V'_{P,Q}/2)$, up to the loss shift $C\Delta \ln(2/\beta)$ in ε and the tail shift β in δ. As in the bound of Theorem 3.1, the excess is a bias of the privacy profile itself and moves both bounds by the same amount, whereas the loss shift is the error of the approximation and moves them apart. In the upper bound the loss shift adds to the excess: $\boldsymbol{H}_{e^\varepsilon}(P^t \parallel Q^t) \le g_{t \cdot V'_{P,Q}}(\varepsilon - \alpha) + \beta$ for every ε, where $\alpha \coloneqq t(\mu'_{P,Q} - V'_{P,Q}/2) + C\Delta \ln(2/\beta)$. For β of order δ, the loss shift is of lower order than $\sqrt{t \cdot V'_{P,Q} \ln(1/\delta)}$, the order of ε at the central limit rate, when Δ is well below the balanced cutoff $\sqrt{t \cdot V'_{P,Q} / \ln(1/\delta)}$, that is, when $t \gg \Delta^2 \ln(1/\delta)/V'_{P,Q}$. In Section 4 we bound the means and the variances of the truncated privacy losses of subsampled steps, and thus the excess, by the χ^2 divergence of the original pair. In Appendix G we show that $\mu'_{P,Q}$ and $V'_{P,Q}$ can be replaced by $\mu_{P,Q}$ and $V_{P,Q}$ if the deviations of the privacy losses from their means have exponentially decaying tails.

4 Subsampling via the χ^2 divergence

In this section we bound the parameters of Theorems 3.1 and 3.3 for t identical λ-subsampled steps by the χ^2 divergence $\tilde{V}$ of the original pair (P, Q), under tail conditions on $L_{P,Q}$ and $L_{Q,P}$ that do not depend on λ. We state our bounds as upper bounds on the privacy profiles of the pairs (P_λ^t, Q^t) and (Q^t, P_λ^t). They also apply to every algorithm dominated by (P_λ^t, Q^t), such as the t-step adaptive composition of λ-subsampled algorithms whose steps are dominated by (P, Q) (Lemma 2.4). In Section 4.2 we apply these results to the subsampled Gaussian mechanism. We defer all proofs to Appendix D.

4.1 Subsampled compositions

The χ^2 divergence of a pair bounds both the mean and the second moment of its privacy loss (Lemma D.1). Therefore it also bounds the variances of the truncated privacy losses, which are the variance parameters of Theorem 3.3. We now show that subsampling at rate λ scales the χ^2 divergence by λ^2 (up to a factor $1/(1-\lambda)$ in the add direction), and that a large privacy loss of the subsampled pair requires an even larger privacy loss of the original pair.

Lemma 4.1 (Subsampling and the χ^2 divergence)**.** *Let P and Q be mutually absolutely continuous and let $\lambda \in (0, 1/2)$. Then*

$$\chi^2_{P_\lambda,Q} = \lambda^2 \chi^2_{P,Q} \qquad \text{and} \qquad \chi^2_{Q,P_\lambda} \le \frac{\lambda^2}{1-\lambda} \chi^2_{P,Q}.$$

Lemma 4.2 (Tails of subsampled privacy losses)**.** *Let P and Q be mutually absolutely continuous and $\lambda \in (0, 1/2)$. Then for every $l > 0$,*

$$\mathbb{P}(L_{P_\lambda,Q} > l) \le \mathbb{P}\left(L_{P,Q} > \ln\left(1 + \frac{e^l - 1}{\lambda}\right)\right) \qquad \text{and} \qquad \mathbb{P}(L_{Q,P_\lambda} > l) \le \mathbb{P}\left(L_{Q,P} > \frac{l}{\lambda}\right).$$

Moreover, for every $\Delta > 0$ and every ω with $|r(\omega; P, Q) - 1| \le \Delta$, we have $|\ell(\omega; P_\lambda, Q)| \le 2\lambda\Delta$.

We truncate the privacy losses of (P_λ, Q) and (Q, P_λ) at the balanced cutoff $\sqrt{tV_{P_\lambda,Q}/\ln(1/\delta)}$ (Section 3.1). For the subsampled pair this cutoff is proportional to λ (it is of order $\lambda\Delta$ for the Δ of Theorem 4.3). Thus, by Lemma 4.2, the probability that it is exceeded for (P_λ, Q) (resp. (Q, P_λ)) is at most the probability that the privacy loss of (P, Q) exceeds $\ln(1+\Delta)$ (resp. that of (Q, P) exceeds Δ), and these thresholds do not depend on λ. Applying Bernstein's inequality to the truncated privacy losses, we obtain the following bound.

Theorem 4.3 (Subsampled composition)**.** *Let P and Q be mutually absolutely continuous with $\tilde{V} \coloneqq \chi^2_{P,Q}$, let $\lambda \in (0, 1/2)$, $t \in \mathbb{N}$ and $0 < \delta \le 1/2$ with $\ln(2/\delta) \le t$, and let $\Delta \coloneqq \sqrt{2t\tilde{V}/\ln(2/\delta)}$. If*

$$\mathbb{P}(L_{P,Q} > \ln(1+\Delta)) \le \frac{\delta}{2t} \qquad \text{and} \qquad \mathbb{P}(L_{Q,P} > \Delta) \le \frac{\delta}{2t},$$

then P_λ^t and Q^t are (ε, δ)-close for

$$\varepsilon \coloneqq \frac{W}{2} + \sqrt{2W\ln(2/\delta)} + \frac{2}{3}\lambda\Delta\ln(2/\delta), \qquad \text{where} \qquad W \coloneqq \frac{t\lambda^2\tilde{V}}{1-\lambda}.$$

For an unbounded PLD we instead set the privacy loss to zero wherever the privacy loss of the original pair exceeds $\ln(1+\Delta)$ in absolute value. Elsewhere the likelihood ratio of the original pair is within Δ of 1, and therefore the privacy loss of the subsampled pair is at most $2\lambda\Delta$ in absolute value (Lemma 4.2), and the truncated privacy loss deviates from its mean by at most $4\lambda\Delta$. We bound both the means and the variances of the truncated privacy losses in both directions by the χ^2 divergence, and obtain from Theorem 3.3 the following concentration result and, as its corollary, a bound by a Gaussian privacy profile with the same parameters in both directions.

Corollary 4.4. *There is a universal constant $C > 0$ such that the following holds. Let P and Q be mutually absolutely continuous with $\tilde{V} \coloneqq \chi^2_{P,Q} < \infty$, and let $t \in \mathbb{N}$, $\lambda \in (0, 1/2)$, $\Delta > 0$ and $0 < \beta \le 1$ be such that*

$$\mathbb{P}(L_{P,Q} > \ln(1+\Delta)) \le \frac{\beta}{4t} \qquad \text{and} \qquad \mathbb{P}(L_{Q,P} > \ln(1+\Delta)) \le \frac{\beta}{4t},$$

and let

$$W \coloneqq \frac{t\lambda^2\tilde{V}}{1-\lambda}.$$

Then the privacy losses of (P_λ^t, Q^t) *and* (Q^t, P_λ^t) *are* $(4C\lambda\Delta\ln(2/\beta), \beta)$*-approximately* (μ', V')*-Gaussian-concentrated for some* $\mu' \le W/2 + \lambda\beta/2$ *and* $0 \le V' \le W$ *(which may differ between the two directions). In particular, every algorithm dominated by* (P_λ^t, Q^t) *is* (α, β)*-shifted* $\sqrt{W}$*-GDP for*

$$\alpha \coloneqq 4C\lambda\Delta\ln(2/\beta) + \frac{W}{2} + \frac{\lambda\beta}{2}.$$

If $L_{P,Q}$ *and* $L_{Q,P}$ *have the same distribution, then this holds with* $\mu' \le W/2$ *and without the term* $\lambda\beta/2$ *of* α*.*

By Lemma 2.8, the shift α is the loss shift of the concentration plus an upper bound on $\mu' - V'/2$, for which we use the upper bound on μ' and the nonnegativity of V'. The resulting term $W/2 + \lambda\beta/2$ of α is of lower order than $\sqrt{W\ln(1/\beta)}$ when $\beta \ll W \ll \ln(1/\beta)$. Compared with Theorem 4.3, the loss shift $4C\lambda\Delta\ln(2/\beta)$ is of lower order than $\sqrt{W\ln(1/\beta)}$ only if $\Delta\sqrt{\ln(1/\beta)}$ is small compared with $\sqrt{t\tilde{V}}$, that is, for Δ well below its value in Theorem 4.3, where $\lambda\Delta$ is the balanced cutoff (see the remark after Theorem 3.3). In exchange, we bound the whole privacy profile by that of $\sqrt{W}$-GDP (up to the shift α and the additive term β). This gives the asymptotic upper bounds of Corollary 4.6.

4.2 The subsampled Gaussian mechanism

In the rest of this section (P, Q) is the Gaussian pair, which dominates the Gaussian mechanism (2) (Claim 2.5), and $\tilde{V} = e^{1/\sigma^2} - 1$. Both privacy losses $L_{P,Q}$ and $L_{Q,P}$ are distributed as $\mathcal{N}(1/(2\sigma^2), 1/\sigma^2)$ (Claim D.2).

For the bound by a Gaussian privacy profile we use

$$\Delta \coloneqq \exp\left(\frac{1}{2\sigma^2} + \frac{\sqrt{2\ln(2t/\beta)}}{\sigma}\right) - 1, \tag{7}$$

where $0 < \beta \le 1$. As the privacy losses are Gaussian, the tail conditions of Corollary 4.4 hold for this Δ (see the proof of Corollary 4.5), and thus Corollary 4.4 applies, without the term $\lambda\beta/2$ of α as the two privacy losses have the same distribution. Using Claim 2.5 and Lemma 2.4, we then obtain the following bound for DP-SGD, that is, for the t-step adaptive composition of the λ-subsampled Gaussian mechanism.

Corollary 4.5 (Shifted GDP of the subsampled Gaussian mechanism)**.** *There is a universal constant* $C > 0$ *such that the following holds. Let* $\sigma > 0$*,* $t \in \mathbb{N}$*,* $\lambda \in (0, 1/2)$ *and* $0 < \beta \le 1$*, and let* Δ *be given by (7). Then the* t*-step adaptive composition of the* λ*-subsampled Gaussian mechanism (2) is* (α, β)*-shifted* $\sqrt{W}$*-GDP, where*

$$W \coloneqq \frac{t\lambda^2(e^{1/\sigma^2} - 1)}{1-\lambda} \qquad \text{and} \qquad \alpha \coloneqq 4C\lambda\Delta\ln(2/\beta) + \frac{W}{2}.$$

The ε that Corollary 4.5 gives exceeds that of $\sqrt{W}$-GDP by the shift α, which has two sources. The first is the term $W/2$, which bounds the excess of the mean of the privacy loss over half its variance, a bias of the privacy profile itself. For $\sigma \ge 1$ it is of order $t\lambda^2/\sigma^2$, and thus small compared with the central limit rate $\lambda\sqrt{t\ln(1/\delta)}/\sigma$ when $\lambda\sqrt{t} \ll \sigma\sqrt{\ln(1/\delta)}$, for example for $\lambda = 1/t$. The second is the loss shift of the Gaussian concentration, which is proportional to the

cutoff Δ. The ratio between this term and the central limit rate does not depend on λ, and it is small when Δ is well below $\sqrt{t}/(\sigma\sqrt{\ln(1/\delta)})$ (see also the discussion after Corollary 4.4).

Converting this result into an (ε, δ)-DP guarantee implies that for some universal constant $C > 0$, for every $\lambda \in (0, 1/2)$, $t \geq 2$ and $0 < \delta \leq 1/2$, if $\sigma \geq 1$ and

$$\sigma \ln\left(1 + \frac{\sqrt{t}}{\sigma\sqrt{C\ln(1/\delta)}}\right) \geq C\sqrt{\ln(t/\delta)}, \tag{8}$$

then the t-step adaptive composition of the λ-subsampled Gaussian mechanism is (ε, δ)-DP for $\varepsilon = O(t\lambda^2/\sigma^2 + \lambda\sqrt{t\ln(1/\delta)}/\sigma)$ (Claim D.4). For $\lambda = 1/t$ and sufficiently large t, the bound above is $\varepsilon = O(\sqrt{\ln(1/\delta)}/(\sigma\sqrt{t}))$, the central limit rate. The following corollary gives its leading constant for a fixed σ as $t \to \infty$, by using the sharper bound of Lemma C.1 on ε_{G} instead of the tail bound. Its first claim supports a polynomial decay $\delta = t^{-\theta}$ of δ in t, and its second claim provides guarantees for δ that decays faster than any polynomial in t.

Corollary 4.6 (Asymptotic upper bounds for the subsampled Gaussian mechanism)**.** *Fix $\sigma > 0$, let $\tilde{V} \coloneqq e^{1/\sigma^2} - 1$, and let M be the t-step adaptive composition of the $1/t$-subsampled Gaussian mechanism (2). If $\delta = t^{-\theta}$ for a fixed $\theta > 1/2$, then, as $t \to \infty$,*

$$\varepsilon_M(\delta) \leq (1 + o(1))\sqrt{\frac{(2\theta - 1)\,\tilde{V}\ln(t)}{t}}.$$

If $\delta = \delta(t)$ satisfies $\ln(1/\delta)/\ln(t) \to \infty$ and $\limsup_{t\to\infty} 8\ln(1/\delta)/(\sigma\ln(t))^2 < 1$, then

$$\varepsilon_M(\delta) \leq (1 + o(1))\sqrt{\frac{2\tilde{V}\ln(1/\delta)}{t}}.$$

For $\lambda = 1/t$ both sources of the shift are of lower order: the bias is of order $1/t$, and in both regimes the cutoff Δ grows polynomially slower than $\sqrt{t}$, so the loss shift is of lower order as well. The leading constant therefore comes from $\varepsilon_{\mathrm{G}}(W, \delta/2)$ with $W = \tilde{V}/(t-1)$. The privacy profile of $\sqrt{W}$-GDP is at most a fraction of order $\sqrt{W}$ of the tail of its privacy loss, and accounting for this factor through Lemma C.1, instead of using the tail bound, effectively replaces $\ln(1/\delta)$ by $\ln(1/\delta) - \frac{1}{2}\ln(t)$, which for $\delta = t^{-\theta}$ gives the factor $2\theta - 1$ in place of 2θ.

By Lemma D.3, the condition (8) of Claim D.4 holds in each of the following two cases:

1. $\eta \in (0, 1/2)$, $t \geq t_\eta$, $\ln(1/\delta) \leq t^{1/2-\eta}$ and $\sigma \geq C_\eta \max\{1, \sqrt{\ln(1/\delta)}/\ln(t)\}$, where C_η and t_η depend only on η;
2. $\sigma \geq 1$ and $t \geq \max\{t_\sigma, \exp(3\sqrt{2}\,C\sqrt{\ln(1/\delta)}/\sigma)\}$, where t_σ depends only on σ.

For a fixed $\sigma \geq 1$ and all sufficiently large t, the second case allows $\delta = t^{-\theta}$ for every constant $\theta > 0$.

5 Random allocation

An alternative sampling scheme to Poisson subsampling is 1-*out-of-t random allocation*. This scheme consists of a t-step adaptive composition of algorithms $M_1, \ldots, M_t$, where every element of the dataset is assigned to one of the t steps, chosen independently and uniformly at random, and the i-th step runs M_i on the dataset in which the elements not assigned to it are replaced by $\perp$. We show that the privacy losses of the dominating pair of this scheme are approximately $(\tilde{V}/(2t), \tilde{V}/t)$-Gaussian-concentrated, where $\tilde{V} \coloneqq \chi^2_{P,Q}$ for a pair (P, Q) that dominates every step.

We then apply this result to the Gaussian mechanism to show that it satisfies the same bounds as the t-step composition of Poisson subsampling at rate $1/t$. We defer the proofs to Appendix E.

Feldman and Shenfeld [FS25, Theorem 3.1] show that if (P, Q) dominates $M_i(\cdot\,; y_1, \dots, y_{i-1})$ for every $i \in [t]$ and all outputs $y_1, \dots, y_{i-1}$, then random allocation is dominated by the pair $(\bar{P}_t, Q^t)$ in the remove direction and $(Q^t, \bar{P}_t)$ in the add direction, where $\bar{P}_t \coloneqq \frac{1}{t}\sum_{j=1}^{t} Q^{j-1} \times P \times Q^{t-j}$. Since $r(\omega_{1:t}; Q^{j-1} \times P \times Q^{t-j}, Q^t) = r(\omega_j; P, Q)$ for every $j \in [t]$, the likelihood ratio of $(\bar{P}_t, Q^t)$ is the average $\frac{1}{t}\sum_{j=1}^{t} r(\omega_j; P, Q)$, and we denote by $R_{\bar{P}_t, Q^t}$ and by $1/R_{Q^t, \bar{P}_t}$ its values for $\omega_{1:t} \sim \bar{P}_t$ and for $\omega_{1:t} \sim Q^t$, respectively. We characterize the concentration of these averages, and use it to bound the PLDs dominating the random allocation scheme via the identities $L_{\bar{P}_t, Q^t} = \ln(R_{\bar{P}_t, Q^t})$ and $L_{Q^t, \bar{P}_t} = -\ln(1/R_{Q^t, \bar{P}_t})$ in the remove and add directions, respectively. Their means and variances can be directly calculated.

Lemma 5.1 (Likelihood ratio of random allocation)**.** *Let $t \in \mathbb{N}$ and let $P \ll Q$ with $\tilde{V} \coloneqq \chi^2_{P,Q} < \infty$. Then $\mathbb{E}\Big[1/R_{Q^t, \bar{P}_t}\Big] = 1$ and $\mathrm{Var}\Big(1/R_{Q^t, \bar{P}_t}\Big) = \frac{\tilde{V}}{t}$, and*

$$\mathbb{E}\Big[R_{\bar{P}_t, Q^t}\Big] = \mathbb{E}\Big[1/R_{Q^t, \bar{P}_t}\Big] + \frac{\tilde{V}}{t},$$

$$\mathrm{Var}\Big(R_{\bar{P}_t, Q^t}\Big) = \mathrm{Var}\Big(1/R_{Q^t, \bar{P}_t}\Big) + \frac{\mathop{\mathbb{E}}_{\omega \sim Q}\big[(r(\omega; P, Q) - 1)^3\big] - \tilde{V}^2}{t^2}.$$

The identities for $1/R_{Q^t, \bar{P}_t}$ (sampling from Q^t) follow from direct calculations. The $R_{\bar{P}_t, Q^t}$ part relies on the fact that a sample from $\bar{P}_t$ is a sample from $P \times Q^{t-1}$ with its first coordinate moved to a uniformly random position. Since the average is invariant under permutations of the coordinates, it has the same distribution under $\bar{P}_t$ as under $P \times Q^{t-1}$, which is identical to $1/R_{Q^t, \bar{P}_t}$ up to the change of measure of a single draw from Q to P. In the proof of Theorem 5.2 we account for this increase of the mean in the remove direction and show that the variance of $R_{\bar{P}_t, Q^t}$ is close to that of $1/R_{Q^t, \bar{P}_t}$.

If the likelihood ratio $r \coloneqq r(\cdot; P, Q)$ satisfies $|r - 1| \le \Delta$, then each summand of the two averages deviates from its own mean by at most $2\Delta/t$. We apply Lemma 3.2 separately to $1/R_{Q^t, \bar{P}_t} - 1$ and $R_{\bar{P}_t, Q^t} - \mathbb{E}\Big[R_{\bar{P}_t, Q^t}\Big]$, which couples each of them with a Gaussian random variable up to an error of order $(\Delta/t)\ln(1/\beta)$, except with probability β. Taking logarithms where the averages are close to 1, we couple both privacy losses with a $\mathcal{N}(\tilde{V}/(2t), \tilde{V}/t)$ random variable, which is distributed as the privacy loss of $\mathcal{G}_{\tilde{V}/t}$.

Theorem 5.2 (Gaussian concentration for random allocation)**.** *There are universal constants $c, C > 0$ such that the following holds. Let $P \ll Q$, $t \in \mathbb{N}$ and $\Delta > 0$, assume that $|r(\omega; P, Q) - 1| \le \Delta$ for Q-almost every ω, and let $\tilde{V} \coloneqq \chi^2_{P,Q}$. Then for every $0 < \beta \le 1/2$ with $(\Delta/t)\ln(C/\beta) \le c$, the privacy losses of $(\bar{P}_t, Q^t)$ and of $(Q^t, \bar{P}_t)$ are $((C\Delta/t)\ln(C/\beta), \beta)$-approximately $(\tilde{V}/(2t), \tilde{V}/t)$-Gaussian-concentrated.*

Here the excess of the mean over half the variance is zero, and therefore, by Lemma 2.8, the privacy profiles of both pairs are bounded from both sides by $g_{\tilde{V}/t}$, up to the loss shift $(C\Delta/t)\ln(C/\beta)$ in ε and the tail shift β in δ. In particular, the pair $(\bar{P}_t, Q^t)$ is $((C\Delta/t)\ln(C/\beta), \beta)$-shifted $\sqrt{\tilde{V}/t}$-GDP.

For an unbounded likelihood ratio we condition on an event E with $P(E) = Q(E) \ge 1 - \beta/(2t)$ on which $|r - 1| \le \Delta$. The conditional pair $(P_E, Q_E) \coloneqq (P(\cdot \mid E), Q(\cdot \mid E))$ has likelihood ratio r on E, and conditioning $\bar{P}_t$ and Q^t on E^t simply uses (P_E, Q_E) in place of (P, Q). As a result, if $0 < \beta \le 1$ and $(\Delta/t)\ln(2C/\beta) \le c$ for the constants c and C of Theorem 5.2, then the privacy losses of $(\bar{P}_t, Q^t)$ and of $(Q^t, \bar{P}_t)$ are $((C\Delta/t)\ln(2C/\beta), \beta)$-approximately $(\tilde{V}_E/(2t), \tilde{V}_E/t)$-Gaussian-concentrated for $\tilde{V}_E \coloneqq \chi^2_{P_E, Q_E}$ (Lemma E.1).

The Gaussian mechanism. For the Gaussian pair (P, Q) with $\sigma \geq 1$, we condition on the event $E \coloneqq \{\omega : |\ell(\omega; P, Q)| \leq \ln(1+\Delta)\}$ for Δ of (7), which has the same probability under P and Q, and obtain the following analogue of Corollary 4.5.

Corollary 5.3 (Gaussian random allocation)**.** *There are universal constants $c, C > 0$ such that the following holds. Let (P, Q) be the Gaussian pair with $\sigma \geq 1$ and $\tilde{V} \coloneqq e^{1/\sigma^2} - 1$, and let $t \geq 2$. If $0 < \beta \leq 1$ and $(\Delta/t)\ln(C/\beta) \leq c$ for Δ given by (7), then 1-out-of-t random allocation of the Gaussian mechanism (2) is (α, β)-shifted $\sqrt{\tilde{V}/(t-1)}$-GDP, where $\alpha \coloneqq \frac{C\Delta}{t}\ln(C/\beta)$.*

As for Poisson subsampling, this gives a closed-form condition for the central limit rate.

Corollary 5.4 (Closed-form bound for Gaussian random allocation)**.** *There is a universal constant $C > 0$ such that the following holds. Let $\sigma \geq 1$, $t \geq 2$ and $0 < \delta \leq 1/2$. If*

$$\sigma \ln\left(1 + \frac{\sqrt{t}}{\sigma\sqrt{C\ln(1/\delta)}}\right) \geq C\sqrt{\ln(t/\delta)},$$

then 1-out-of-t random allocation of the Gaussian mechanism (2) is (ε, δ)-DP for

$$\varepsilon = C\left(\frac{1}{\sigma^2 t} + \frac{\sqrt{\ln(1/\delta)}}{\sigma\sqrt{t}}\right).$$

Up to the constants, this is the bound of Claim D.4 for Poisson subsampling at rate $\lambda = 1/t$, under the same condition (8). The two cases after Corollary 4.6 in which (8) holds are proved for every value of its constant (Lemma D.3), and therefore they also apply to random allocation. In contrast, the reduction of random allocation to Poisson subsampling [FS25, Corollary 4.5] needs the assumption on σ of the analysis of Poisson subsampling and, in addition, $\sigma \gtrsim \sqrt{\ln(t/\delta)}$.

Since $\tilde{V}/(t-1)$ is also the parameter of Corollary 4.5 for $\lambda = 1/t$, inverting the Gaussian profile as in Corollary 4.6 gives the same bounds.

Corollary 5.5 (Asymptotic upper bounds for random allocation)**.** *Fix $\sigma \geq 1$, let (P, Q) be the Gaussian pair and $\tilde{V} \coloneqq e^{1/\sigma^2} - 1$. If $\delta = t^{-\theta}$ for a fixed $\theta > 1/2$, then, as $t \to \infty$,*

$$\varepsilon_{\bar{P}_t, Q^t}(\delta) \leq (1 + o(1))\sqrt{\frac{(2\theta - 1)\,\tilde{V}\ln(t)}{t}}.$$

If $\delta = \delta(t)$ satisfies $\ln(1/\delta)/\ln(t) \to \infty$ and $\limsup_{t\to\infty} 8\ln(1/\delta)/(\sigma\ln(t))^2 < 1$, then

$$\varepsilon_{\bar{P}_t, Q^t}(\delta) \leq (1 + o(1))\sqrt{\frac{2\tilde{V}\ln(1/\delta)}{t}}.$$

By Claim 2.5 and [FS25, Theorem 3.1], the same upper bounds hold for $\varepsilon_M(\delta)$, where M is 1-out-of-t random allocation of the Gaussian mechanism (2).

6 Application: connect-the-dots discretization

In this section we use our approximation analysis for composition to explain the properties of connect-the-dots (CtD) discretization. Specifically, we show that while discretizing a single step can increase ε by almost the grid spacing, after composition the increase per step is only of second order in the spacing (up to the accuracy of the Gaussian concentration).

Numerical PLD accountants discretize the PLD of every step on a grid and convolve the resulting PLDs. The simplest conservative approach is to round the loss up to the nearest grid

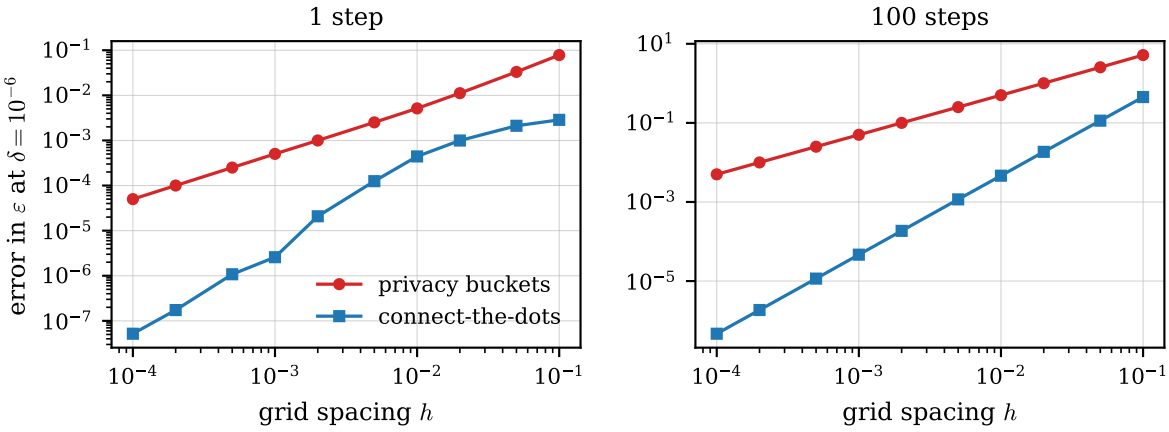


Figure 1: The error of the upper bound on ε at $\delta = 10^{-6}$ for the Gaussian mechanism with $\sigma = 10$, whose privacy loss is distributed as $\mathcal{N}(1/(2\sigma^2), 1/\sigma^2)$, discretized on the grid $h\mathbb{Z}$ by privacy buckets and by CtD, as a function of the grid spacing h, for a single step (left) and for the composition of 100 steps (right). The upper bounds are the pessimistic estimates of `dp_accounting` [Goo26] with its default truncation of the tails, and the error is computed relative to the exact value. For small h, the error of privacy buckets is close to $th/2$, whereas that of CtD is of order h^2.

point, which is the approach used in the privacy buckets method [MM18]. The CtD discretization [DGKKM22] instead splits the probability mass of every privacy loss between the two adjacent grid points, and it is the least, under domination, of the dominating pairs whose PLD is supported on the grid [DGKKM22, Lemma 4.1]. Experimentally, it keeps a similar error bound with a discretization interval 66.66 times larger than that of privacy buckets [DGKKM22, Section 6]. Figure 1 shows the error of both discretizations for the Gaussian mechanism, for a single step and for 100 steps. This plot shows that CtD discretization is better even for a single step but its advantage becomes close to quadratic in h after composition. We demonstrate this effect in Figure 2 by examining the ratio of the two errors for $t = 1, 5, 25, 100$ and 1000 steps. We defer the proofs to Appendix F.

The discretization. Let P and Q be mutually absolutely continuous and $L \coloneqq L_{P,Q}$. We split the probability mass of L at every value between the two knots adjacent to it in a *grid*, that is, in a set $\Lambda \subset \mathbb{R}$ of *knots* with finitely many knots in every bounded interval and $\inf \Lambda \le L \le \sup \Lambda$ almost surely. The resulting distribution is that of the random post-processing of L that moves it to one of the two adjacent knots, and we define the discretization through this post-processing. The grid has *spacing* at most h if adjacent knots differ by at most h, as for $h\mathbb{Z}$. We analyze the effect of discretization separately from the truncation of the tails, which in standard implementations is done together with discretization.

Definition 6.1 (Connect-the-dots discretization)**.** The *CtD discretization* of L on a grid Λ is the random variable $\tilde{L}$ that is equal to L if $L \in \Lambda$, and otherwise is equal to one of the knots $\ell^- < \ell^+$ adjacent to L, with the conditional probabilities given, for $\ell^- < \ell < \ell^+$, by

$$\mathbb{P}\Big(\tilde{L} = \ell^- \mid L = \ell\Big) = \frac{e^{-\ell} - e^{-\ell^+}}{e^{-\ell^-} - e^{-\ell^+}} \qquad \text{and} \qquad \mathbb{P}\Big(\tilde{L} = \ell^+ \mid L = \ell\Big) = \frac{e^{-\ell^-} - e^{-\ell}}{e^{-\ell^-} - e^{-\ell^+}}. \tag{9}$$

The *CtD pair* of (P, Q) on Λ consists of the distributions $\tilde{P}(\ell) \coloneqq \mathbb{P}\Big(\tilde{L} = \ell\Big)$ and $\tilde{Q}(\ell) \coloneqq e^{-\ell}\tilde{P}(\ell)$ over Λ.

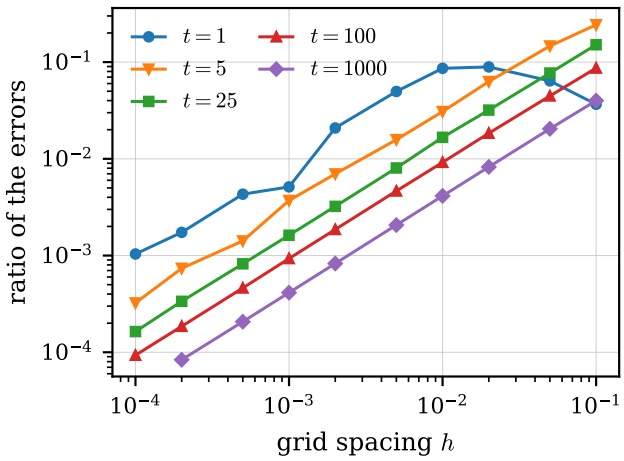


Figure 2: The ratio of the error of CtD to that of privacy buckets, in the setting of Figure 1, as a function of the grid spacing h, for the composition of t steps.

Note that this post-processing preserves the conditional mean of e^{-L}, rather than that of L:

$$\mathbb{E}\left[e^{-\tilde{L}} \mid L\right] = e^{-L}. \tag{10}$$

By this identity and $\mathbb{E}\left[e^{-L}\right] = 1$, $\tilde{Q}$ is a probability distribution, $\tilde{P}$ and $\tilde{Q}$ are mutually absolutely continuous, and the privacy loss of $(\tilde{P}, \tilde{Q})$ is distributed as $\tilde{L}$ (and thus $\tilde{L}$ couples the privacy losses of the two pairs).

The resulting error. In the worst case, the discretization of a single step increases ε by almost the grid spacing. For example, for randomized response with privacy loss $\pm\varepsilon_0$, where $0 < \varepsilon_0 < h$, the two distributions are $(\varepsilon_0, 0)$-close, whereas their CtD discretizations on the grid $h\mathbb{Z}$ are (ε, δ)-close only for ε above a threshold that tends to h as $\delta \to 0$ (Claim F.1). The mean and the variance of the privacy loss, in contrast, change only by $O(h^2)$ (Lemma 6.2), whereas rounding up to the grid, as in privacy buckets, increases every privacy loss, and thus also the mean, by up to h. As the privacy profile of a composition is determined by the mean and the variance of its privacy loss up to the shifts of the Gaussian concentration (Theorem 3.3 and Lemma 2.8), this is why the increase of ε per step tends to be quadratic in h after composition (Corollary 6.4). The bounds on the moments also give an explanation of why CtD is more accurate in practice even for a single step (Figure 1), although by themselves they cannot guarantee a smaller error for small δ and general algorithms, as the example of randomized response shows.

Lemma 6.2 (CtD discretization of one step)**.** *Let $\tilde{L}$ and $(\tilde{P}, \tilde{Q})$ be as in Definition 6.1 for a grid with spacing at most h, and assume that $V_{P,Q} < \infty$. Then*

1. $0 \le \mu_{\tilde{P},\tilde{Q}} - \mu_{P,Q} \le h^2/8$;
2. $-(h^2/8)\sqrt{V_{P,Q}} \le V_{\tilde{P},\tilde{Q}} - V_{P,Q} \le h^2/4 + (h^2/8)\sqrt{V_{P,Q}}$.

We state the results of this section for the privacy loss L of (P, Q). They also apply to the reversed pairs, since the CtD pair of (Q, P) on the grid $-\Lambda$ is $(\tilde{Q}, \tilde{P})$ with every knot ℓ replaced by $-\ell$: by (9) for the privacy loss $-\ell(\omega; P, Q)$, $\omega \sim Q$, of (Q, P), the knot $-\ell$ receives

the mass $e^{-\ell}\tilde{P}(\ell) = \tilde{Q}(\ell)$. In practice, the two directions are usually discretized separately, as in `dp_accounting` [Goo26] and `PLD_accounting` [She26].

We now discretize every step of the composition (P^t, Q^t) of Section 3, whose privacy loss L_{P^t,Q^t} has mean μ_{P^t,Q^t} and variance V_{P^t,Q^t}. Let $(\tilde{P}, \tilde{Q})$ be the CtD pair of (P, Q) on a grid Λ, and for every $i \in [t]$, let L_i be a copy of $L_{P,Q}$ and $\tilde{L}_i$ its CtD discretization on Λ (Definition 6.1), where $L_1, \ldots, L_t$ and the post-processings are independent. We write $\varepsilon(\delta)$ and $\tilde{\varepsilon}(\delta)$ for the smallest $\varepsilon \geq 0$ such that $\boldsymbol{H}_{e^\varepsilon}(P^t \parallel Q^t) \leq \delta$ and $\boldsymbol{H}_{e^\varepsilon}\left(\tilde{P}^t \,\middle\|\, \tilde{Q}^t\right) \leq \delta$, respectively. We apply Lemma 6.2 to (P, Q) and Theorem 3.3 to the discretized composition, and obtain the following comparison.

Theorem 6.3 (Composition of CtD discretizations)**.** *Let C be the constant of Theorem 3.3 and let $\Delta, h > 0$. Assume that the grid Λ has spacing at most h and that $|L_{P,Q} - \mu_{P,Q}| \leq \Delta$ almost surely. Then $(\tilde{P}^t, \tilde{Q}^t)$ dominates (P^t, Q^t), and*

$$t \cdot \mu_{P,Q} \leq \mu_{\tilde{P}^t,\tilde{Q}^t} \leq t \cdot \mu_{P,Q} + \frac{th^2}{8}, \quad t \cdot V_{P,Q} - \frac{th^2}{8}\sqrt{V_{P,Q}} \leq V_{\tilde{P}^t,\tilde{Q}^t} \leq t \cdot V_{P,Q} + \frac{th^2}{4} + \frac{th^2}{8}\sqrt{V_{P,Q}}.$$

Furthermore, for every $0 < \beta \leq 1$, the privacy loss of $(\tilde{P}^t, \tilde{Q}^t)$ is $(C(\Delta + 2h)\ln(2/\beta), \beta)$-approximately $(\mu_{\tilde{P}^t,\tilde{Q}^t}, V_{\tilde{P}^t,\tilde{Q}^t})$-Gaussian-concentrated.

We bound the cost of replacing the parameters $(\mu_{\tilde{P}^t,\tilde{Q}^t}, V_{\tilde{P}^t,\tilde{Q}^t})$ of the Gaussian random variable by $(\mu_{P^t,Q^t}, V_{P^t,Q^t})$ using (11), and obtain the following bound on the error caused by the discretization.

Corollary 6.4 (Error of the discretization)**.** *Assume the conditions of Theorem 6.3 and that $V_{P,Q} > 0$. Then for every $0 < \delta \leq 1$ and $0 < \beta \leq 1$ with $3\beta < \delta$,*

$$\varepsilon(\delta) \leq \tilde{\varepsilon}(\delta) \leq \varepsilon(\delta - 3\beta) + 2C(\Delta + h)\ln(2/\beta) + \frac{th^2}{8} + \frac{h^2\sqrt{t}}{8}\left(\frac{1}{\sqrt{V_{P,Q}}} + 1\right)\sqrt{2\ln(1/\beta)}.$$

The term $2C(\Delta + h)\ln(2/\beta)$ is the error of the two Gaussian concentrations: the comparison uses the lower bound for the original composition and the upper bound for the discretized one, and therefore both loss shifts enter. It depends neither on the variance nor on the number of steps. The term $th^2/8$ bounds the increase of the mean, which biases the privacy profile itself rather than adding an error of the approximation, and the last term bounds the effect of the change of the variance. The last term is $O(h^2\sqrt{t\ln(1/\beta)/V_{P,Q}})$ whenever $V_{P,Q} \leq 1$, which holds for $\Delta \leq 1$, since $V_{P,Q} \leq \Delta^2$. This explains the finding in [DGKKM22] for bounded privacy losses.

For comparison, we note that the error of a mean-preserving discretization can be bounded by coupling the original and the discretized privacy losses and applying Hoeffding's inequality [Hoe63], as in [GLW21, Lemma 5.3]. This argument also applies to CtD, since given $L_1, \ldots, L_t$ the corrections $\tilde{L}_i - L_i$ are independent, with means at most $h^2/8$ and ranges of length at most h (see the proof of Lemma 6.2). Thus the privacy loss of (P^t, Q^t) approximately stochastically dominates that of $(\tilde{P}^t, \tilde{Q}^t)$ with tail shift β and loss shift $th^2/8 + h\sqrt{t\ln(1/\beta)/2}$, which is of order $h\sqrt{t\ln(1/\beta)}$ when $th^2 \ll 1$. For t of order $1/V_{P,Q}$, the last term of Corollary 6.4 is $O(th^2\sqrt{\ln(1/\beta)})$, and therefore our bound on the error caused by the discretization is quadratically better (ignoring logarithmic factors).

Acknowledgments

Shenfeld's work was supported in part by ERC grant 101125913, Simons Foundation Collaboration 733792, Israel Science Foundation (ISF) grant 2861/20, and Apple. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

[ACG+16] Martín Abadi, Andy Chu, Ian J. Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. "Deep Learning with Differential Privacy". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*. Ed. by Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi. Full version: arXiv:1607.00133v2. We refer to the numbering of the statements in this version. ACM, 2016, pp. 308–318. URL: https://arxiv.org/pdf/1607.00133v2.

[AFKRT26] Hilal Asi, Vitaly Feldman, Hannah Keller, Guy Rothblum, and Kunal Talwar. "PREAMBLE: Private and Efficient Aggregation via Block Sparse Vectors". In: *Advances in Neural Information Processing Systems* 38 (2026), pp. 20093–20137.

[AGACKS23] Wael Alghamdi, Juan Felipe Gómez, Shahab Asoodeh, Flávio P. Calmon, Oliver Kosut, and Lalitha Sankar. "The Saddle-Point Method in Differential Privacy". In: *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*. Ed. by Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett. Vol. 202. Proceedings of Machine Learning Research. PMLR, 2023, pp. 508–528. URL: https://proceedings.mlr.press/v202/alghamdi23a.html.

[BBG18] Borja Balle, Gilles Barthe, and Marco Gaboardi. "Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. Ed. by Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett. Curran Associates, Inc., 2018, pp. 6280–6290. URL: https://proceedings.neurips.cc/paper/2018/hash/3b5020bb891119b9f5130f1fea9bd773-Abstract.html.

[BDLS20] Zhiqi Bu, Jinshuo Dong, Qi Long, and Weijie J. Su. "Deep Learning with Gaussian Differential Privacy". In: *Harvard Data Science Review* 2.3 (Sept. 2020). Full version: arXiv:1911.11607v3. We refer to the numbering of the statements in this version.

[BDRS18] Mark Bun, Cynthia Dwork, Guy N. Rothblum, and Thomas Steinke. "Composable and Versatile Privacy via Truncated CDP". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*. Ed. by Ilias Diakonikolas, David Kempe, and Monika Henzinger. We refer to the numbering of the statements in the 36-page full version. ACM, 2018, pp. 74–86. DOI: 10.1145/3188745.3188946. URL: https://www.thomas-steinke.net/tcdp.pdf.

[BFGT20] Raef Bassily, Vitaly Feldman, Cristóbal Guzmán, and Kunal Talwar. "Stability of Stochastic Gradient Descent on Nonsmooth Convex Losses". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. Ed. by Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin. Curran Associates, Inc., 2020, pp. 4381–4391. URL: https://proceedings.neurips.cc/paper/2020/hash/2e2c4bf7ceaa4712a72dd5ee136dc9a8-Abstract.html.

[BFTT19] Raef Bassily, Vitaly Feldman, Kunal Talwar, and Abhradeep Guha Thakurta. "Private Stochastic Convex Optimization with Optimal Rates". In: *Advances in*

Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada. Ed. by Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett. Curran Associates, Inc., 2019, pp. 11279–11288. URL: https://proceedings.neurips.cc/paper/2019/hash/3bd8fdb090f1f5eb66a00c84dbc5ad51-Abstract.html.

[BKOZB12] Gilles Barthe, Boris Köpf, Federico Olmedo, and Santiago Zanella Beguelin. "Probabilistic Relational Reasoning for Differential Privacy". In: *Proceedings of the 39th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2012, Philadelphia, Pennsylvania, USA, January 22-28, 2012*. Ed. by John Field and Michael Hicks. ACM, 2012, pp. 97–110. DOI: 10.1145/2103656.2103670.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. DOI: 10.1093/ACPROF:OSO/9780199535255.001.0001.

[BS16] Mark Bun and Thomas Steinke. "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds". In: *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part I*. Ed. by Martin Hirt and Adam D. Smith. Vol. 9985. Lecture Notes in Computer Science. Full version: arXiv:1605.02065v1. We refer to the numbering of the statements in this version. Springer, 2016, pp. 635–658. URL: https://arxiv.org/pdf/1605.02065v1.

[BST14] Raef Bassily, Adam D. Smith, and Abhradeep Thakurta. "Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds". In: *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*. IEEE Computer Society, 2014, pp. 464–473. DOI: 10.1109/FOCS.2014.56.

[BW18] Borja Balle and Yu-Xiang Wang. "Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Ed. by Jennifer G. Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. Full version: arXiv:1805.06530v2. We refer to the numbering of the statements in this version. PMLR, 2018, pp. 403–412. URL: https://arxiv.org/pdf/1805.06530v2.

[CGH+25] Lynn Chua, Badih Ghazi, Charlie Harrison, Pritish Kamath, Ravi Kumar, Ethan Leeman, Pasin Manurangsi, Amer Sinha, and Chiyuan Zhang. "Balls-and-Bins Sampling for DP-SGD". In: *International Conference on Artificial Intelligence and Statistics, AISTATS 2025, Mai Khao, Thailand, 3-5 May 2025*. Ed. by Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Mohammad Emtiyaz Khan. Vol. 258. Proceedings of Machine Learning Research. PMLR, 2025, pp. 946–954. URL: https://proceedings.mlr.press/v258/chua25a.html.

[CGHST25] Christopher A. Choquette-Choo, Arun Ganesh, Saminul Haque, Thomas Steinke, and Abhradeep Guha Thakurta. "Near-Exact Privacy Amplification for Matrix Mechanisms". In: *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL: https://openreview.net/forum?id=txV4dNeusx.

[CR21] Mark Cesar and Ryan Rogers. "Bounding, Concentrating, and Truncating: Unifying Privacy Loss Composition for Data Analytics". In: *Algorithmic Learning*

Theory, ALT 2021, 16-19 March 2021, Virtual Conference, Worldwide. Ed. by Vitaly Feldman, Katrina Ligett, and Sivan Sabato. Vol. 132. Proceedings of Machine Learning Research. Full version: arXiv:2004.07223v3. We refer to the numbering of the statements in this version. PMLR, 2021, pp. 421–457. URL: arxiv.org/pdf/2004.07223v3.

[DCÖ25] Andy Dong, Wei-Ning Chen, and Ayfer Özgür. "Leveraging Randomness in Model and Data Partitioning for Privacy Amplification". In: *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*. Ed. by Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu. Vol. 267. Proceedings of Machine Learning Research. PMLR, 2025. URL: https://proceedings.mlr.press/v267/dong25a.html.

[DGKKM22] Vadym Doroshenko, Badih Ghazi, Pritish Kamath, Ravi Kumar, and Pasin Manurangsi. "Connect the Dots: Tighter Discrete Approximations of Privacy Loss Distributions". In: *Proceedings on Privacy Enhancing Technologies* 2022.4 (2022). Full version: arXiv:2207.04380v1. We refer to the numbering of the statements in this version, pp. 552–570.

[DKMMN06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. "Our Data, Ourselves: Privacy Via Distributed Noise Generation". In: *Advances in Cryptology - EUROCRYPT 2006, 25th International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*. Ed. by Serge Vaudenay. Vol. 4004. Lecture Notes in Computer Science. Springer, 2006, pp. 486–503. DOI: 10.1007/11761679_29.

[DÖ26] Andy Dong and Ayfer Özgür. "Less Random, More Private: What is the Optimal Subsampling Scheme for DP-SGD?" In: *arXiv preprint arXiv:2605.07072* (2026). DOI: 10.48550/ARXIV.2605.07072. arXiv: 2605.07072.

[DR14] Cynthia Dwork and Aaron Roth. "The Algorithmic Foundations of Differential Privacy". In: *Foundations and Trends in Theoretical Computer Science* 9.3-4 (2014), pp. 211–407. DOI: 10.1561/0400000042.

[DR16] Cynthia Dwork and Guy N. Rothblum. "Concentrated Differential Privacy". In: *arXiv preprint arXiv:1603.01887* (2016). Full version: arXiv:1603.01887v2. We refer to the numbering of the statements in this version. DOI: 10.48550/ARXIV.1603.01887. arXiv: 1603.01887.

[DRS22] Jinshuo Dong, Aaron Roth, and Weijie J. Su. "Gaussian Differential Privacy". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.1 (2022). Full version: arXiv:1905.02383v3. We refer to the numbering of the statements in this version, pp. 3–37.

[DRV10] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. "Boosting and Differential Privacy". In: *51st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*. IEEE Computer Society, 2010, pp. 51–60. DOI: 10.1109/FOCS.2010.12.

[FS25] Vitaly Feldman and Moshe Shenfeld. "Privacy Amplification by Random Allocation". In: *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*. Ed. by Danielle Belgrave, Cheng Zhang, Laura N. Montoya, Hsuan-Tien Lin, Razvan Pascanu, Piotr Koniusz, Marzyeh Ghassemi, Nancy Chen, Iván

Vladimir Meza Ruíz, and Arturo Loaiza-Bonilla. Full version: arXiv:2502.08202v4. We refer to the numbering of the statements in this version. Curran Associates, Inc., 2025, pp. 96188–96230. URL: https://arxiv.org/pdf/2502.08202v4.

[FS26] Vitaly Feldman and Moshe Shenfeld. "Efficient Privacy Loss Accounting for Subsampling and Random Allocation". In: *Forty-third International Conference on Machine Learning, ICML 2026, Seoul, South Korea, July 6-11, 2026*. Full version: arXiv:2602.17284v2. We refer to the numbering of the statements in this version. 2026. URL: https://arxiv.org/pdf/2602.17284v2.

[GLW21] Sivakanth Gopi, Yin Tat Lee, and Lukas Wutschitz. "Numerical Composition of Differential Privacy". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. Ed. by Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan. Full version: arXiv:2106.02848v3. We refer to the numbering of the statements in this version. Curran Associates, Inc., 2021, pp. 11631–11642. URL: https://arxiv.org/pdf/2106.02848v3.

[Goo26] Google Differential Privacy Team. *Differential Privacy Accounting (dp_accounting), part of the Google differential privacy library*. https://github.com/google/differential-privacy/tree/main/python/dp_accounting. Accessed October 2026. 2026.

[Hoe63] Wassily Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables". In: *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30. DOI: 10.1080/01621459.1963.10500830.

[KH21] Antti Koskela and Antti Honkela. "Computing Differential Privacy Guarantees for Heterogeneous Compositions Using FFT". In: *arXiv preprint arXiv:2102.12412* (2021). DOI: 10.48550/ARXIV.2102.12412. arXiv: 2102.12412.

[KJH20] Antti Koskela, Joonas Jälkö, and Antti Honkela. "Computing Tight Differential Privacy Guarantees Using FFT". In: *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*. Ed. by Silvia Chiappa and Roberto Calandra. Vol. 108. Proceedings of Machine Learning Research. PMLR, 2020, pp. 2560–2569. URL: https://proceedings.mlr.press/v108/koskela20b.html.

[KJPH21] Antti Koskela, Joonas Jälkö, Lukas Prediger, and Antti Honkela. "Tight Differential Privacy for Discrete-Valued Mechanisms and for the Subsampled Gaussian Mechanism Using FFT". In: *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event*. Ed. by Arindam Banerjee and Kenji Fukumizu. Vol. 130. Proceedings of Machine Learning Research. PMLR, 2021, pp. 3358–3366. URL: https://proceedings.mlr.press/v130/koskela21a.html.

[KLNRS08] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. "What Can We Learn Privately?" In: *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, Philadelphia, PA, USA, October 25-28, 2008*. IEEE Computer Society, 2008, pp. 531–540. DOI: 10.1109/FOCS.2008.27.

[KOV15] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. "The Composition Theorem for Differential Privacy". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*. Ed. by Francis R. Bach and David M. Blei. Vol. 37. Proceedings of Machine Learning Research.

Full version: arXiv:1311.0776v4. We refer to the numbering of the statements in this version. PMLR, 2015, pp. 1376–1385. URL: https://arxiv.org/pdf/1311.0776v4.

[Mir17] Ilya Mironov. "Rényi Differential Privacy". In: *30th IEEE Computer Security Foundations Symposium, CSF 2017, Santa Barbara, CA, USA, August 21-25, 2017*. IEEE Computer Society, 2017, pp. 263–275. DOI: 10.1109/CSF.2017.11.

[MM18] Sebastian Meiser and Esfandiar Mohammadi. "Tight on Budget?: Tight Bounds for r-Fold Approximate Differential Privacy". In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*. Ed. by David Lie, Mohammad Mannan, Michael Backes, and XiaoFeng Wang. ACM, 2018, pp. 247–264. DOI: 10.1145/3243734.3243765.

[MTZ19] Ilya Mironov, Kunal Talwar, and Li Zhang. "Rényi Differential Privacy of the Sampled Gaussian Mechanism". In: *arXiv preprint arXiv:1908.10530* (2019).

[Sak96] A. I. Sakhanenko. "Estimates for the Accuracy of Coupling in the Central Limit Theorem". In: *Siberian Mathematical Journal* 37.4 (1996). Translated from Sibirskiĭ Matematicheskiĭ Zhurnal 37(4):919–931, pp. 811–823. DOI: 10.1007/BF02104672.

[SCS13] Shuang Song, Kamalika Chaudhuri, and Anand D. Sarwate. "Stochastic Gradient Descent with Differentially Private Updates". In: *IEEE Global Conference on Signal and Information Processing, GlobalSIP 2013, Austin, TX, USA, December 3-5, 2013*. IEEE, 2013, pp. 245–248. DOI: 10.1109/GLOBALSIP.2013.6736861.

[She26] Moshe Shenfeld. *PLD_accounting: A PLD accounting service for random allocation and subsampling*. https://github.com/moshenfeld/PLD_accounting. Accessed October 2026. 2026.

[SK26] Jan Schuchardt and Nikita Kalinin. "Sampling-Free Privacy Accounting for Matrix Mechanisms under Random Allocation". In: *arXiv preprint arXiv:2601.21636* (2026). DOI: 10.48550/ARXIV.2601.21636. arXiv: 2601.21636.

[SMM19] David M. Sommer, Sebastian Meiser, and Esfandiar Mohammadi. "Privacy Loss Classes: The Central Limit Theorem in Differential Privacy". In: *Proceedings on Privacy Enhancing Technologies* 2019.2 (2019), pp. 245–269. DOI: 10.2478/POPETS-2019-0029.

[Ste25] Thomas Steinke. "Composition of Differential Privacy & Privacy Amplification by Subsampling". In: *Differential Privacy in Artificial Intelligence: From, Theory to Practice*. Ed. by Ferdinando Fioretto and Pascal Van Hentenryck. Full version: arXiv:2210.00597v4. We refer to the numbering of the statements in this version. Emerald Publishing Limited, 2025.

[WGZSSW26] Hua Wang, Sheng Gao, Huanyu Zhang, Milan Shen, Weijie J. Su, and Jiayuan Wu. "Edgeworth Accountant: An Analytical Approach to Differential Privacy Composition". In: *Journal of the American Statistical Association* (2026). Full version: arXiv:2206.04236v3. We refer to the numbering of the statements in this version, pp. 1–11.

[Zai25] Andrei Yu. Zaitsev. "Estimates of Transport Distance in the Central Limit Theorem". In: *arXiv preprint arXiv:2511.02033* (2025). Full version: arXiv:2511.02033v2. We refer to the numbering of the statements in this version. DOI: 10.48550/ARXIV.2511.02033. arXiv: 2511.02033.

[ZDLS20] Qinqing Zheng, Jinshuo Dong, Qi Long, and Weijie J. Su. "Sharp Composition Bounds for Gaussian Differential Privacy via Edgeworth Expansion". In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. Proceedings of Machine Learning Research. Full version: arXiv:2003.04493v2. We refer to the numbering of the statements in this version. PMLR, 2020, pp. 11420–11435. URL: https://arxiv.org/pdf/2003.04493v2.

[ZDW22] Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. "Optimal Accounting of Differential Privacy via Characteristic Function". In: *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event*. Ed. by Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera. Vol. 151. Proceedings of Machine Learning Research. Full version: arXiv:2106.08567v3. We refer to the numbering of the statements in this version. PMLR, 2022, pp. 4782–4817. URL: https://arxiv.org/pdf/2106.08567v3.

[ZW19] Yuqing Zhu and Yu-Xiang Wang. "Poission Subsampled Rényi Differential Privacy". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, 2019, pp. 7634–7642. URL: https://proceedings.mlr.press/v97/zhu19c.html.

A Comparison with existing approaches

We compare existing analyses of the Gaussian mechanism at sampling rate $\lambda = 1/t$ (which is also the regime of random allocation) by the noise scale they need and by how their leading term relates to the central limit rate $\sqrt{\ln(1/\delta)}/(\sigma\sqrt{t})$. Table 1 summarizes the comparison.

Rényi divergence. For the λ-subsampled Gaussian mechanism, the moments accountant [ACG+16] shows that the Rényi divergence of integer order $q \le \sigma^2 \ln(1/(\lambda\sigma))$ is $O(\lambda^2 q/\sigma^2)$. This quadratic behavior fails for orders much larger than $\sigma^2 \ln(1/\lambda)$. Indeed, by the exact binomial expression [MTZ19], the divergence is at least $q/(2\sigma^2) - \frac{q}{q-1}\ln(1/\lambda)$ (the contribution of the term in which the element is sampled in all q draws; see also [Ste25, Section 6.6]). At the same time, the conversion of a quadratic bound for t steps to (ε, δ)-DP at the central limit rate uses the order $q \asymp \sigma\sqrt{\ln(1/\delta)/t}/\lambda$. The two constraints are compatible only if $\lambda\sigma \ln(1/(\lambda\sigma)) \gtrsim \sqrt{\ln(1/\delta)/t}$, which at $\lambda = 1/t$ reads $\sigma \gtrsim \sqrt{t \ln(1/\delta)}/\ln(t/\ln(1/\delta))$ up to lower-order terms.

Amplification and advanced composition. A black-box analysis starts from the $(\varepsilon_0, \delta_0)$-DP guarantee of the Gaussian mechanism with $\varepsilon_0 = 1/(2\sigma^2) + \sqrt{2\ln(1/\delta_0)}/\sigma$, amplifies it by subsampling to $(\ln(1+\lambda(e^{\varepsilon_0}-1)), \lambda\delta_0)$-DP and composes t steps by advanced composition [DRV10]. At $\lambda = 1/t$ the t steps fail with total probability δ_0, which forces $\delta_0 \lesssim \delta$. The amplified ε is of order ε_0/t only when $\varepsilon_0 = O(1)$, that is, when $\sigma \gtrsim \sqrt{\ln(1/\delta)}$. Advanced composition then gives $\varepsilon \lesssim \ln(1/\delta)/(\sigma\sqrt{t})$. The condition on σ is mild, but the bound is a factor $\sqrt{\ln(1/\delta)}$ above the central limit rate.

Central limit theorems. Bu et al. [BDLS20, Section 3.1 and Theorem 6] show that when $\lambda\sqrt{t}$ converges to a positive constant, t steps of the Poisson-subsampled Gaussian mechanism tend to $\lambda\sqrt{t\tilde{V}}$-GDP, where $\tilde{V} = e^{1/\sigma^2} - 1$ is the χ^2 divergence of the Gaussian pair. Through the Gaussian

Method	Scheme	Sufficient noise scale	Leading term
Quadratic Rényi bounds [ACG+16; MTZ19]	Poisson	$\sigma \gtrsim \sqrt{t \ln(1/\delta)/\ln(t/\ln(1/\delta))}$	$\sqrt{\ln(1/\delta)}/(\sigma\sqrt{t})$
$(\varepsilon_0, \delta_0)$ amplification with advanced composition	Poisson	$\sigma \gtrsim \sqrt{\ln(1/\delta)}$	$\ln(1/\delta)/(\sigma\sqrt{t})$
Reduction to Poisson subsampling [FS25]	allocation	$\sigma \gtrsim \sqrt{\ln(t/\delta)}$ and the condition of the Poisson analysis	$\sqrt{\ln(t/\delta)}/(\sigma\sqrt{t})$
This work (Claim D.4 and Corollary 5.4)	both	$\sigma \ln\left(1 + \frac{\sqrt{t}}{\sigma\sqrt{C \ln(1/\delta)}}\right) \gtrsim \sqrt{\ln(t/\delta)}$	$\sqrt{\ln(1/\delta)}/(\sigma\sqrt{t})$

Table 1: Noise scales at which each analysis gives its leading privacy term for the Gaussian mechanism with Poisson subsampling at rate $1/t$ or with 1-out-of-t random allocation (lower-order terms omitted). For polynomially small δ, the condition of this work holds for every fixed $\sigma \geq 1$ once t is sufficiently large.

tail bound this limit yields $\varepsilon \approx t\lambda^2\tilde{V}/2 + \lambda\sqrt{2t\tilde{V}\ln(1/\delta)}$, the form of the bound of Theorem 4.3. It is a limit statement, however, and its hypotheses exclude $\lambda = 1/t$, for which $\lambda\sqrt{t} \to 0$.

For every finite t, there is a Berry–Esseen bound [DRS22]. For batches of a fixed size λn out of n examples, the trade-off function of t steps of the Gaussian mechanism is at least the GDP trade-off function moved to the left and down by β_{BE} and truncated at 0 [DRS22, Corollary 5.6], where β_{BE} equals $0.56/\sqrt{t}$ times a normalized third moment of the privacy loss of one step [DRS22, Lemma 5.5]. At type I error 0 this lower bound is at most $1 - \beta_{\mathrm{BE}}$, whereas the trade-off function of (ε, δ)-DP equals $1 - \delta$ there. Therefore the guarantee yields (ε, δ)-DP only for $\delta \geq \beta_{\mathrm{BE}}$. For $\lambda = 1/t$ and fixed σ, the error β_{BE} is of order $1/\sqrt{t}$.

Edgeworth expansions [WGZSSW26; ZDLS20] and the saddle-point accountant [AGACKS23, Theorem 5.6] approximate the privacy profile more accurately. The first of these expansions applies only to type I errors bounded away from 0 and 1 [ZDLS20, Section 3.2], while small values of δ depend on type I errors near 0. The finite-sample bound on the error of the second expansion is additive and uniform in the argument of the distribution function, of order $1/\sqrt{t}$ with constants given by normalized moments of the privacy losses [WGZSSW26, Lemma 4.3], and it improves to order $1/t$ under a decay condition on the characteristic function and up to an unspecified lower-order remainder [WGZSSW26, Remark 4.4 and Corollary E.1]. As for the Berry–Esseen bound, such an error gives (ε, δ)-DP only for δ above it, and we have not specialized these bounds to $\lambda = 1/t$. The saddle-point bounds are evaluated numerically, and the accompanying asymptotic composition theorem lets δ depend on the number of steps [AGACKS23, Theorem 4.1], but it assumes that the standard deviation of the total privacy loss is much larger than $\sqrt{\ln(1/\delta)}$. In the regime of the central limit rate $\varepsilon \lesssim \ln(1/\delta)$ this assumption fails for the subsampled Gaussian mechanism, whose total privacy loss has variance at most $t\lambda^2\tilde{V}/(1-\lambda)$ by Lemmas D.1 and 4.1 and therefore standard deviation at most of order $\varepsilon/\sqrt{\ln(1/\delta)}$.

B Comparison with other privacy notions

The definitions below apply to the output distributions P and Q on every pair of adjacent datasets, in both directions. For $q > 1$, we write $D_q(P\|Q) \coloneqq \frac{1}{q-1}\ln\left(\mathbb{E}\left[e^{(q-1)L_{P,Q}}\right]\right)$ for the Rényi divergence of order q.

Zero-concentrated differential privacy. An algorithm is (ξ, ρ)-zCDP if $D_q(P\|Q) \le \xi + \rho q$ for every $q > 1$, and it is ρ-zCDP if it is $(0, \rho)$-zCDP [BS16, Definition 1.1]. For a Gaussian privacy loss $\mathcal{N}(\mu, V)$, equality holds for $(\xi, \rho) = (\mu - V/2, V/2)$. Every $\sqrt{V}$-GDP algorithm is $V/2$-zCDP [BS16, Lemmas 1.8 and 2.4]. Shifted GDP does not imply zCDP when $\beta > 0$, since the Rényi divergences can be infinite. Subsampling does not amplify zCDP for the Gaussian mechanism [Ste25, Section 6.6].

Truncated concentrated differential privacy. For $\rho > 0$ and $\bar{q} > 1$, an algorithm is $(\rho, \bar{q})$-tCDP if $D_q(P\|Q) \le \rho q$ for every $q \in (1, \bar{q})$ [BDRS18, Definition 1]. For a Gaussian privacy loss $\mathcal{N}(\mu, V)$, this holds exactly when $\rho \ge \mu$. Bun et al. [BDRS18, Theorem 13] prove privacy amplification by uniform subsampling without replacement for tCDP. We discuss the limitations of Rényi bounds for Poisson subsampling in Appendix A.

Concentrated differential privacy. An algorithm is (μ, τ)-CDP if its privacy loss $L \coloneqq L_{P,Q}$ has mean at most μ and $L - \mathbb{E}[L]$ is τ-subgaussian [DR16, Definitions 2.5, 3.2 and 3.3]. For a Gaussian privacy loss $\mathcal{N}(\mu, V)$, both bounds hold with equality for $(\mu, \sqrt{V})$. Subsampling does not reduce τ for the Gaussian mechanism [Ste25, Section 6.6], [BS16, Lemma 4.2]. Unlike zCDP and shifted GDP, CDP is not closed under post-processing [BS16, Proposition A.1].

C Proofs for Section 3

We use the notation of Section 3. We first bound the privacy profile of GDP, which we use for the leading constants in Corollary 4.6. We then prove Theorem 3.3 from a Gaussian coupling of sums of bounded independent random variables. Finally, we bound the cost of changing the parameters of a Gaussian random variable, which we also use in later sections.

Lemma C.1 (Privacy profile of GDP)**.** *Let $V > 0$ and $\delta \in (0, 1)$ be such that $k \coloneqq 2\ln(1/\delta) - \ln(1/V) \ge 1$. Then*

$$\varepsilon_{\mathrm{G}}(V, \delta) \le \frac{V}{2} + \sqrt{kV}.$$

Proof. The pair $\mathcal{G}_V$ consists of the output distributions of the Gaussian mechanism with noise variance $1/V$ on two adjacent datasets on which a sensitivity-1 query takes the values 1 and 0, and this mechanism is $V/2$-zCDP [BS16, Proposition 1.6]. Therefore, by [BS16, Lemma 3.6] with $\rho = V/2$, for every $\varepsilon \ge V/2$,

$$g_V(\varepsilon) \le \frac{e^{-(\varepsilon - V/2)^2/(2V)}}{1 + (\varepsilon - V/2)/V}.$$

For $\varepsilon = V/2 + \sqrt{kV}$ the right-hand side equals $e^{-k/2}/(1 + \sqrt{k/V}) = \delta/(\sqrt{V} + \sqrt{k}) \le \delta$, since $\sqrt{V} e^{-k/2} = \delta$ and $k \ge 1$. □

We derive Lemma 3.2 from the following bound on the moments of the error of a Gaussian coupling.

Theorem C.2 ([Sak96, Corollary 4 and inequality (1.6)])**.** *There is a universal constant $C_0 > 0$ such that the following holds. Let $X_1, \dots, X_t$ be independent mean-zero random variables, let $S \coloneqq \sum_{i=1}^t X_i$ with $V_S \coloneqq \mathrm{Var}(S) > 0$, and let $p \ge 2$. Then S can be coupled with a standard normal random variable G_0 so that*

$$\mathbb{E}\left[\left|\frac{S}{\sqrt{V_S}} - G_0\right|^p\right] \le (C_0 p)^p \, V_S^{-(p+2)/2} \sum_{i=1}^t \mathbb{E}\left[|X_i|^{p+2}\right].$$

Proof of Lemma 3.2. We may assume that $V_S \coloneqq \mathrm{Var}(S) > 0$, since otherwise $S = 0$ almost surely and we can take $G = 0$. For every $p \geq 2$, let G_0 be the standard normal random variable of Theorem C.2, and let C_0 be its constant. Since $|X_i|^{p+2} \leq \Delta^p X_i^2$ and $\sum_i \mathbb{E}[X_i^2] = V_S$, the random variable $G \coloneqq \sqrt{V_S}\, G_0 \sim \mathcal{N}(0, V_S)$ satisfies $\mathbb{E}[|S-G|^p] \leq (C_0 p \Delta)^p$. We choose $p \coloneqq \max\{2, \lceil u \rceil\}$ for $u \coloneqq \ln(2/\beta) > 1$, so that $u \leq p \leq 2u$. Then Markov's inequality for $|S-G|^p$ gives, for $C \coloneqq 2eC_0$,

$$\mathbb{P}(|S-G| > C\Delta u) \leq \left(\frac{C_0 p}{Cu}\right)^p \leq e^{-p} \leq e^{-u} = \frac{\beta}{2}.$$

□

Proof of Theorem 3.3. We prove the claim with $C \coloneqq 2C_1$, where C_1 is the constant of Lemma 3.2. Let $\omega \sim P^t$ and $L \coloneqq \ell(\omega; P^t, Q^t) = \sum_i \ell(\omega_i; P, Q)$ (by (6)), and let $L' \coloneqq \sum_i \ell(\omega_i; P, Q)\mathbb{1}\{\omega_i \in E\}$ be the truncated privacy loss. By a union bound over the events $\omega_i \notin E$, we have that $L' = L$ except with probability at most $\beta/2$.

The centered summands $\ell(\omega_i; P, Q)\mathbb{1}\{\omega_i \in E\} - \mu'_{P,Q}$ of $L' - t \cdot \mu'_{P,Q}$ are independent and at most Δ in absolute value almost surely. By Lemma 3.2 with $\beta/2$ in place of β, we can couple $L' - t \cdot \mu'_{P,Q}$ with $G_0 \sim \mathcal{N}(0, t \cdot V'_{P,Q})$ so that $|L' - t \cdot \mu'_{P,Q} - G_0| \leq C_1 \Delta \ln(4/\beta)$ except with probability $\beta/2$ (given ω, we sample G_0 from the conditional distribution of G_0 given L' in this coupling). Thus $G \coloneqq t \cdot \mu'_{P,Q} + G_0 \sim \mathcal{N}(t \cdot \mu'_{P,Q}, t \cdot V'_{P,Q})$ satisfies $|L' - G| \leq C_1 \Delta \ln(4/\beta)$ except with probability $\beta/2$. Since $\ln(4/\beta) \leq 2\ln(2/\beta)$ for $\beta \leq 1$, we get that $|L - G| \leq C\Delta \ln(2/\beta)$ except with probability β. As noted after Definition 2.7, this coupling gives the claim. □

Changing the parameters of a Gaussian random variable costs a loss shift: for $\mu, \mu_0 \in \mathbb{R}$, $V, V_0 \geq 0$ and $0 < \beta \leq 1$, random variables with distributions $\mathcal{N}(\mu_0, V_0)$ and $\mathcal{N}(\mu, V)$ each (α, β)-approximately stochastically dominate the other for α given by (11) below. Indeed, for $G \sim \mathcal{N}(0,1)$ the random variables $\mu + \sqrt{V}\, G$ and $\mu_0 + \sqrt{V_0}\, G$ are distributed as $\mathcal{N}(\mu, V)$ and $\mathcal{N}(\mu_0, V_0)$. By the Gaussian tail bound, $|G| \leq \sqrt{2\ln(1/\beta)}$ except with probability β, and on this event the two random variables differ by at most

$$|\mu - \mu_0| + \left|\sqrt{V} - \sqrt{V_0}\right| \sqrt{2\ln(1/\beta)}. \tag{11}$$

As noted after Definition 2.7, this coupling gives the claim.

D Proofs for Section 4

In this appendix r is the likelihood ratio of (P, Q), and $r_\lambda \coloneqq 1 - \lambda + \lambda r$ is the likelihood ratio of (P_λ, Q). We first prove the bounds on the moments of the privacy loss that we use in Section 4.

Lemma D.1 (Moments of the privacy loss)**.** *Let $P \ll Q$ with $\chi^2_{P,Q} < \infty$. Then*

$$0 \leq \mu_{P,Q} \leq \chi^2_{P,Q} \qquad \text{and} \qquad \mathbb{E}\left[L^2_{P,Q}\right] \leq \chi^2_{P,Q}.$$

In particular, $V'_{P,Q} \leq \chi^2_{P,Q}$ for every set E.

Proof. By Jensen's inequality,

$$\begin{aligned}\mu_{P,Q} = \mathop{\mathbb{E}}_{\omega\sim P}[\ln(r(\omega; P, Q))] &\leq \ln\left(\mathop{\mathbb{E}}_{\omega\sim P}[r(\omega; P, Q)]\right) \\ &= \ln\left(\mathop{\mathbb{E}}_{\omega\sim Q}\left[r(\omega; P, Q)^2\right]\right) = \ln(1 + \chi^2_{P,Q}) \leq \chi^2_{P,Q},\end{aligned}$$

and similarly $\mathbb{E}[-L_{P,Q}] \le \ln\Big(\mathop{\mathbb{E}}_{\omega\sim P}[1/r(\omega;P,Q)]\Big) \le 0$. For the second moment we use that $s\ln^2(s) \le (s-1)^2$ for $s \ge 0$, which for $s = e^{2a}$ is the inequality $|a| \le |\sinh a|$. Thus

$$\mathbb{E}\Big[L_{P,Q}^2\Big] = \mathop{\mathbb{E}}_{\omega\sim Q}\Big[r(\omega;P,Q)\ln^2(r(\omega;P,Q))\Big] \le \mathop{\mathbb{E}}_{\omega\sim Q}\Big[(r(\omega;P,Q)-1)^2\Big] = \chi^2_{P,Q}.$$

The bound on the variance follows since a variance is at most the second moment. □

Proof of Lemma 4.1. By the definition $P_\lambda = (1-\lambda)Q + \lambda P$, for Q-almost every ω we have

$$r_\lambda(\omega) = \frac{(1-\lambda)Q(\omega) + \lambda P(\omega)}{Q(\omega)} = 1 - \lambda + \lambda r(\omega),$$

so $r_\lambda(\omega) - 1 = \lambda(r(\omega)-1)$. The definition of the χ^2 divergence now gives

$$\chi^2_{P_\lambda,Q} = \mathop{\mathbb{E}}_{\omega\sim Q}\Big[(r_\lambda(\omega)-1)^2\Big] = \lambda^2 \mathop{\mathbb{E}}_{\omega\sim Q}\Big[(r(\omega)-1)^2\Big] = \lambda^2\chi^2_{P,Q}.$$

For the reverse direction, $r_\lambda(\omega) \ge 1-\lambda > 0$, so Q and P_λ are mutually absolutely continuous and the likelihood ratio of (Q, P_λ) is $1/r_\lambda(\omega)$. Changing the sampling distribution from P_λ to Q introduces the factor $r_\lambda(\omega)$, giving

$$\begin{aligned}
\chi^2_{Q,P_\lambda} &= \mathop{\mathbb{E}}_{\omega\sim P_\lambda}\left[\left(\frac{1}{r_\lambda(\omega)} - 1\right)^2\right] \\
&= \mathop{\mathbb{E}}_{\omega\sim Q}\left[r_\lambda(\omega)\left(\frac{1}{r_\lambda(\omega)} - 1\right)^2\right] \\
&= \mathop{\mathbb{E}}_{\omega\sim Q}\left[\frac{(r_\lambda(\omega)-1)^2}{r_\lambda(\omega)}\right] = \lambda^2 \mathop{\mathbb{E}}_{\omega\sim Q}\left[\frac{(r(\omega)-1)^2}{r_\lambda(\omega)}\right].
\end{aligned}$$

Finally, using $r_\lambda(\omega) \ge 1-\lambda$ and the nonnegativity of the numerator, we obtain

$$\chi^2_{Q,P_\lambda} \le \frac{\lambda^2}{1-\lambda}\mathop{\mathbb{E}}_{\omega\sim Q}\Big[(r(\omega)-1)^2\Big] = \frac{\lambda^2}{1-\lambda}\chi^2_{P,Q}.$$

□

Proof of Lemma 4.2. We have that $\ell(\omega;P_\lambda,Q) = \ln(r_\lambda(\omega))$ and $\ell(\omega;Q,P_\lambda) = -\ln(r_\lambda(\omega))$.

For the first bound, let $a \coloneqq 1 + (e^l - 1)/\lambda > 1$ and $A \coloneqq \{r > a\}$. Then

$$\ell(\omega;P_\lambda,Q) > l \iff r_\lambda(\omega) > e^l \iff r(\omega) > a \iff \ell(\omega;P,Q) > \ln(a),$$

and $P > aQ \ge Q$ on A, so

$$\mathbb{P}(L_{P_\lambda,Q} > l) = P_\lambda(A) = (1-\lambda)Q(A) + \lambda P(A) \le P(A) = \mathbb{P}(L_{P,Q} > \ln(a)).$$

For the second bound, concavity of the logarithm gives $\ln(r_\lambda) \ge (1-\lambda)\ln(1) + \lambda\ln(r) = \lambda\ln(r)$, that is, $\ell(\omega;Q,P_\lambda) \le \lambda\ell(\omega;Q,P)$. Hence, for $\omega \sim Q$,

$$\mathbb{P}(L_{Q,P_\lambda} > l) = \mathop{\mathbb{P}}_{\omega\sim Q}(\ell(\omega;Q,P_\lambda) > l) \le \mathop{\mathbb{P}}_{\omega\sim Q}\Big(\ell(\omega;Q,P) > \tfrac{l}{\lambda}\Big) = \mathbb{P}\Big(L_{Q,P} > \tfrac{l}{\lambda}\Big).$$

For the last claim, let $|r(\omega)-1| \le \Delta$. Then $|r_\lambda(\omega)-1| = \lambda|r(\omega)-1| \le \lambda\Delta$ and $r_\lambda(\omega) \ge 1-\lambda$, and since $\ln'(u) = 1/u \le 1/(1-\lambda)$ for $u \ge 1-\lambda$, the mean value theorem gives

$$|\ell(\omega;P_\lambda,Q)| = |\ln(r_\lambda(\omega)) - \ln(1)| \le \frac{|r_\lambda(\omega)-1|}{1-\lambda} \le \frac{\lambda\Delta}{1-\lambda} \le 2\lambda\Delta.$$

□

Proof of Theorem 4.3. We bound $\boldsymbol{H}_{e^\varepsilon}(P_\lambda^t \parallel Q^t)$ and may assume that $\tilde{V} > 0$, since otherwise $P = Q$. The bound on $\boldsymbol{H}_{e^\varepsilon}(Q^t \parallel P_\lambda^t)$ follows in the same way from the pair (Q, P_λ), the second tail bound of Lemma 4.2 and the second assumption.

Let $L \coloneqq L_{P_\lambda,Q}$. By Lemmas D.1 and 4.1, $\mathbb{E}[L^2] \le \chi^2_{P_\lambda,Q} = \lambda^2\tilde{V} \le W/t$, and in the add direction the second moment is at most $\chi^2_{Q,P_\lambda} \le W/t$. For the mean, let $x \coloneqq \lambda(r-1)$, so that $\mathbb{E}[L] = \mathop{\mathbb{E}}_{Q}[(1+x)\ln(1+x)]$. For every $x > -1$,

$$(1+x)\ln(1+x) \le x + \frac{x^2}{2(1+\min\{x,0\})} \qquad \text{and} \qquad -\ln(1+x) \le -x + \frac{x^2}{2(1+\min\{x,0\})}. \tag{12}$$

Indeed, for $x \ge 0$ both differences vanish at $x = 0$ and have the nonnegative derivatives $x - \ln(1+x)$ and $x^2/(1+x)$. For $-1 < x < 0$ and $u \coloneqq 1+x$, the two differences are $(3u^2 - 4u + 1 - 2u^2\ln u)/(2u)$ and $(1 - u^2 + 2u\ln u)/(2u)$; their numerators vanish at $u = 1$ and are nonincreasing on $(0,1]$, since their derivatives $4(u - 1 - u\ln u)$ and $2(1 - u + \ln u)$ are nonpositive. Since $x \ge -\lambda$ and $\mathop{\mathbb{E}}_{Q}[x] = 0$, (12) gives that $\mathbb{E}[L] \le \mathop{\mathbb{E}}_{Q}[x^2]/(2(1-\lambda)) = W/(2t)$, and in the same way the mean $\mathop{\mathbb{E}}_{Q}[-\ln(1+x)]$ in the add direction is at most $W/(2t)$. We truncate L at the balanced cutoff $\lambda\Delta = \sqrt{2t\lambda^2\tilde{V}/\ln(2/\delta)}$. As $(e^{\lambda\Delta} - 1)/\lambda \ge \Delta$, the first tail bound of Lemma 4.2 and the first assumption give that

$$\mathbb{P}(L > \lambda\Delta) \le \mathbb{P}(L_{P,Q} > \ln(1+\Delta)) \le \frac{\delta}{2t}.$$

The truncated privacy loss $L' \coloneqq L\,\mathbb{1}\{L \le \lambda\Delta\}$ satisfies $L' \le L$ and $(L')^2 \le L^2$, since $\lambda\Delta > 0$. Thus $\mathbb{E}[L'] \le W/(2t)$, $\mathbb{E}[(L')^2] \le W/t$ and $|\mathbb{E}[L']| \le \sqrt{W/t} \le \lambda\sqrt{2\tilde{V}} \le \lambda\Delta$, where we used that $\lambda < 1/2$ and that the last inequality is equivalent to the assumption that $\ln(2/\delta) \le t$. In particular, $L' - \mathbb{E}[L'] \le 2\lambda\Delta$.

Let $L_1, \dots, L_t$ be independent copies of L and $L_i' \coloneqq L_i\mathbb{1}\{L_i \le \lambda\Delta\}$. By (6), $\sum_i L_i$ is distributed as $L_{P_\lambda^t,Q^t}$, and by a union bound, $L_i = L_i'$ for all i except with probability $\delta/2$. By Lemma 2.6 for the centered L_i' with $2\lambda\Delta$ in place of Δ, $V = W$ and $\delta/2$ in place of δ, except with probability δ,

$$\sum_{i=1}^t L_i = \sum_{i=1}^t L_i' \le \frac{W}{2} + \sqrt{2W\ln(2/\delta)} + \frac{2}{3}\lambda\Delta\ln(2/\delta) = \varepsilon.$$

By the tail bound (1), this gives the claim. □

Proof of Corollary 4.4. Let C be the constant of Theorem 3.3 and $E \coloneqq \{\omega : |\ell(\omega;P,Q)| \le \ln(1+\Delta)\}$. On E the likelihood ratio $r = e^{\ell(\omega;P,Q)}$ is between $1/(1+\Delta)$ and $1+\Delta$, and therefore $|r-1| \le \Delta$. Since $\ell(\omega;Q,P) = -\ell(\omega;P,Q)$, and since $P \le Q$ where $\ell(\omega;P,Q) < 0$ and $Q \le P$ where $\ell(\omega;P,Q) > 0$, both $P(E^c)$ and $Q(E^c)$ are at most $\mathbb{P}(L_{P,Q} > \ln(1+\Delta)) + \mathbb{P}(L_{Q,P} > \ln(1+\Delta)) \le \beta/(2t)$. Therefore $P_\lambda(E^c) = (1-\lambda)Q(E^c) + \lambda P(E^c) \le \beta/(2t)$, and since $|r-1| \le \Delta$ on E, Lemma 4.2 gives that $|\ell(\omega;Q,P_\lambda)| = |\ell(\omega;P_\lambda,Q)| \le 2\lambda\Delta$ for every $\omega \in E$. Thus in both directions the truncated privacy loss of one step, which is zero outside E, and its mean are at most $2\lambda\Delta$ in absolute value, and the truncated privacy loss deviates from its mean by at most $4\lambda\Delta$. Therefore Theorem 3.3 for the pairs (P_λ, Q) and (Q, P_λ), with the set E and $4\lambda\Delta$ in place of Δ, gives the claimed concentration with μ' and V' the mean and the variance of the truncated privacy loss, that is, of $\sum_{i=1}^t \ell(\omega_i;P_\lambda,Q)\,\mathbb{1}\{\omega_i \in E\}$ for $\omega \sim P_\lambda^t$ in the remove direction and of $\sum_{i=1}^t \ell(\omega_i;Q,P_\lambda)\,\mathbb{1}\{\omega_i \in E\}$ for $\omega \sim Q^t$ in the add direction.

We now bound the means and the variances. By Lemmas D.1 and 4.1, the variance of the truncated privacy loss of one step is at most $\chi^2_{P_\lambda,Q} = \lambda^2\tilde{V}$ in the remove direction and at most

$\chi^2_{Q,P_\lambda} \le \lambda^2\tilde{V}/(1-\lambda)$ in the add direction, so both variances are at most W. For the means, let $x \coloneqq \lambda(r-1) \ge -\lambda$, so that the likelihood ratio of (P_λ, Q) is $1+x$, and recall (12) from the proof of Theorem 4.3. Since $1+\min\{x,0\} \ge 1-\lambda$, the change of measure from P_λ to Q and (12) give that the means of the truncated privacy losses of one step satisfy

$$\mathop{\mathbb{E}}_{\omega\sim P_\lambda}[\ell(\omega;P_\lambda,Q)\mathbb{1}\{\omega\in E\}] \le \mathop{\mathbb{E}}_{Q}[x\,\mathbb{1}\{\omega\in E\}] + \frac{\mathop{\mathbb{E}}_{Q}[x^2]}{2(1-\lambda)},$$
$$\mathop{\mathbb{E}}_{\omega\sim Q}[\ell(\omega;Q,P_\lambda)\mathbb{1}\{\omega\in E\}] \le -\mathop{\mathbb{E}}_{Q}[x\,\mathbb{1}\{\omega\in E\}] + \frac{\mathop{\mathbb{E}}_{Q}[x^2]}{2(1-\lambda)}.$$

Here $\mathop{\mathbb{E}}_{Q}[x^2] = \lambda^2\tilde{V}$, and $\mathop{\mathbb{E}}_{Q}[x\,\mathbb{1}\{\omega\in E\}] = \lambda(P(E)-Q(E)) = \lambda(Q(E^c)-P(E^c))$ is at most $\lambda\beta/(2t)$ in absolute value. Summing over the t steps, both means are at most $W/2+\lambda\beta/2$. If $L_{P,Q}$ and $L_{Q,P}$ have the same distribution, then $P(E^c)=Q(E^c)$, since $L_{Q,P}$ is distributed as $-\ell(\omega;P,Q)$ for $\omega\sim Q$, and therefore both means are at most $W/2$.

For the shifted GDP bound, fix one of the two directions and let μ' and V' be as above. By Lemma 2.8, the hockey-stick divergence $\boldsymbol{H}_{e^\varepsilon}$ of the pair in this direction is at most $g_{V'}(\varepsilon-(\mu'-V'/2)-4C\lambda\Delta\ln(2/\beta))+\beta$ for every $\varepsilon\in\mathbb{R}$. Since $V'\ge 0$ and $\mu'\le W/2+\lambda\beta/2$, and as $g_V(\varepsilon)$ is nonincreasing in ε and nondecreasing in V with $V'\le W$, this is at most $g_W(\varepsilon-\alpha)+\beta$. If $L_{P,Q}$ and $L_{Q,P}$ have the same distribution, then $\mu'\le W/2$, and the same argument gives this bound without the term $\lambda\beta/2$ of α. If (P_λ^t, Q^t) dominates an algorithm, then it dominates the pairs of output distributions of the algorithm in the remove direction, and (Q^t, P_λ^t) dominates those in the add direction, since reversing both pairs preserves domination (Lemma 2.3). Hence the algorithm is (α,β)-shifted $\sqrt{W}$-GDP. □

D.1 The subsampled Gaussian mechanism

In the rest of this appendix (P,Q) is the Gaussian pair $(\mathcal{N}(1,\sigma^2),\mathcal{N}(0,\sigma^2))$ and $\tilde{V}=e^{1/\sigma^2}-1$.

Claim D.2 (The Gaussian pair)**.** *Let $\sigma>0$ and let (P,Q) be the Gaussian pair. Then $L_{P,Q}$ and $L_{Q,P}$ are both distributed as $\mathcal{N}(1/(2\sigma^2),1/\sigma^2)$, and thus for every $\ell\in\mathbb{R}$,*

$$\mathbb{P}(L_{P,Q}>\ell) = \mathbb{P}(L_{Q,P}>\ell) = \overline{\Phi}\Big(\sigma\ell - \frac{1}{2\sigma}\Big). \tag{13}$$

Moreover, $\chi^2_{P,Q} = e^{1/\sigma^2}-1$, and if $\sigma\ge 1$, then $1/\sigma^2 \le \chi^2_{P,Q} \le (e-1)/\sigma^2$.

Proof. The privacy loss $\ell(\omega;P,Q) = (2\omega-1)/(2\sigma^2)$ is distributed as $\mathcal{N}(1/(2\sigma^2),1/\sigma^2)$ for $\omega\sim P$ and as $\mathcal{N}(-1/(2\sigma^2),1/\sigma^2)$ for $\omega\sim Q$, and $L_{Q,P}$ is $-\ell(\omega;P,Q)$ for $\omega\sim Q$. Completing the square gives that $\mathop{\mathbb{E}}_{\omega\sim Q}\Big[r(\omega;P,Q)^2\Big] = e^{1/\sigma^2}$, that is, $\chi^2_{P,Q} = e^{1/\sigma^2}-1$. For $\sigma\ge 1$ we have that $1+1/\sigma^2 \le e^{1/\sigma^2} \le 1+(e-1)/\sigma^2$ by the convexity of the exponential function and $1/\sigma^2\le 1$. □

We prove that (8) holds in the two cases stated after Corollary 4.6 for every value of its constant, since we also use them for random allocation (Section 5).

Lemma D.3 (Sufficient conditions for (8))**.** *Let $C>0$. For every $\sigma\ge 1$, $t\ge 2$ and $0<\delta\le 1/2$, the inequality (8) holds in each of the following two cases:*

1. *$\eta\in(0,1/2)$, $t\ge t_\eta$, $\ln(1/\delta)\le t^{1/2-\eta}$ and $\sigma \ge C_\eta \max\{1, \sqrt{\ln(1/\delta)}/\ln(t)\}$, where C_η and t_η depend only on η and C;*
2. *$t \ge \max\{t_\sigma, \exp(3\sqrt{2}\,C\sqrt{\ln(1/\delta)}/\sigma)\}$, where t_σ depends only on σ and C.*

Proof. Let $H \coloneqq \ln(t/\delta)$. We write (8) as $F(\sigma) \geq C\sqrt{H}$, where $F(s) \coloneqq s\ln(1+a/s)$ and $a \coloneqq \sqrt{t/(C\ln(1/\delta))}$. The function F is increasing, since $F'(s) = \ln(1+b) - b/(1+b) > 0$ for $b = a/s$. We take $t_\eta, t_\sigma \geq 3$, so that $\ln(t) \geq 1$.

In the first case let $C_\eta \coloneqq 4C/\eta$. As F is increasing, it suffices to verify the condition at $\sigma_* \coloneqq C_\eta \max\{1, \sqrt{\ln(1/\delta)}/\ln(t)\}$. Since $\ln(t) \geq 1$, $\eta < 1/2$ and $\ln(1/\delta) \leq t^{1/2-\eta}$, we have that $\sqrt{\ln(1/\delta)}\,\sigma_* \leq C_\eta \max\{\sqrt{\ln(1/\delta)}, \ln(1/\delta)\} \leq C_\eta t^{1/2-\eta}$, that is, $a/\sigma_* \geq t^\eta/(C_\eta\sqrt{C})$. Therefore $\ln(1+a/\sigma_*) \geq \frac{\eta}{2}\ln(t)$ for all t larger than a threshold that depends only on η and C, and for such t,

$$F(\sigma_*) \geq \frac{\eta C_\eta}{2}\max\left\{\ln(t), \sqrt{\ln(1/\delta)}\right\} \geq C\left(\ln(t) + \sqrt{\ln(1/\delta)}\right) \geq C\sqrt{\ln(t) + \ln(1/\delta)} = C\sqrt{H}.$$

In the second case let $c \coloneqq 1/(18C^2)$. The assumption $t \geq \exp(3\sqrt{2}\,C\sqrt{\ln(1/\delta)}/\sigma)$, that is, $\ln(1/\delta) \leq c\sigma^2\ln^2(t)$, gives that $a/\sigma \geq \sqrt{t}/(\sigma^2\sqrt{Cc}\,\ln(t))$. Thus for all t larger than a threshold that depends only on σ and C we have that $F(\sigma) \geq \frac{\sigma}{3}\ln(t)$ and $\ln(t) \leq c\sigma^2\ln^2(t)$, and therefore

$$C\sqrt{H} \leq C\sqrt{\ln(t) + c\sigma^2\ln^2(t)} \leq C\sigma\sqrt{2c}\,\ln(t) = \frac{\sigma}{3}\ln(t) \leq F(\sigma).$$

□

Proof of Corollary 4.5. Let $w \coloneqq \sqrt{2\ln(2t/\beta)}$, so that $\ln(1+\Delta) = 1/(2\sigma^2) + w/\sigma$. By (13),

$$\mathbb{P}(L_{P,Q} > \ln(1+\Delta)) = \mathbb{P}(L_{Q,P} > \ln(1+\Delta)) = \overline{\Phi}(w) \leq \frac{e^{-w^2/2}}{2} = \frac{\beta}{4t}.$$

Thus Corollary 4.4 applies with $\tilde{V} = e^{1/\sigma^2} - 1$, and since $L_{P,Q}$ and $L_{Q,P}$ have the same distribution (Claim D.2), it applies without the term $\lambda\beta/2$ of α. Since the Gaussian pair dominates the Gaussian mechanism (Claim 2.5), the pair (P_λ^t, Q^t) dominates the t-step adaptive composition of the λ-subsampled Gaussian mechanism by Lemma 2.4, which gives the claim. The proof of Corollary 4.4 gives the same bound on the hockey-stick divergences of the pairs (P_λ^t, Q^t) and (Q^t, P_λ^t). □

We now state formally the bound for every sampling rate that is stated informally before Corollary 4.6.

Claim D.4 (Closed-form bound for the subsampled Gaussian mechanism)**.** *There is a universal constant $C > 0$ such that the following holds. Let $\sigma \geq 1$, $t \geq 2$, $\lambda \in (0, 1/2)$ and $0 < \delta \leq 1/2$. If (8) holds, then the t-step adaptive composition of the λ-subsampled Gaussian mechanism (2) is (ε, δ)-DP for*

$$\varepsilon \coloneqq C\left(\frac{t\lambda^2}{\sigma^2} + \frac{\lambda\sqrt{t\ln(1/\delta)}}{\sigma}\right).$$

Proof. Let C_1 be the constant of Corollary 4.5. By Corollary 4.5 with $\beta = \delta/2$, the t-step adaptive composition of the λ-subsampled Gaussian mechanism is $(\alpha, \delta/2)$-shifted $\sqrt{W}$-GDP for

$$W = \frac{t\lambda^2\tilde{V}}{1-\lambda} \qquad \text{and} \qquad \alpha = 4C_1\lambda\Delta\ln\left(\frac{4}{\delta}\right) + \frac{W}{2},$$

where $\ln(1+\Delta) = \frac{1}{2\sigma^2} + \frac{\sqrt{2\ln(4t/\delta)}}{\sigma}$. Since $\alpha \geq 0$, (5) gives that it is $(\varepsilon_{\mathrm{G}}(W, \delta/2) + \alpha, \delta)$-DP.

Let $H \coloneqq \ln(t/\delta)$, so that $H \geq \ln(4)$ for $t \geq 2$ and $\delta \leq 1/2$, and let $C \geq \max\{5/2,\, 2(e-1),\, \sqrt{8(e-1)} + 8C_1\}$. As $\ln(4t/\delta) = \ln(4) + H \leq 2H$ and $1/(2\sigma^2) \leq \sqrt{H}/(2\sigma)$ for $\sigma \geq 1$, we have that $\ln(1+\Delta) \leq \frac{5}{2}\sqrt{H}/\sigma \leq C\sqrt{H}/\sigma$. Therefore (8) gives that

$$\Delta \leq \frac{\sqrt{t}}{\sigma\sqrt{C\ln(1/\delta)}}.$$

As $\ln(4/\delta) \leq 3\ln(1/\delta)$ for $\delta \leq 1/2$ and $C \geq 5/2$, the first term of α satisfies

$$4C_1\lambda\Delta\ln\!\left(\frac{4}{\delta}\right) \leq \frac{12C_1}{\sqrt{C}} \cdot \frac{\lambda\sqrt{t\ln(1/\delta)}}{\sigma} \leq \frac{8C_1\lambda\sqrt{t\ln(1/\delta)}}{\sigma}.$$

By Claim D.2 and $\lambda < 1/2$, we have that $W \leq 2(e-1)t\lambda^2/\sigma^2$, and since $\varepsilon_{\mathrm{G}}(W, \delta/2) \leq W/2 + \sqrt{2W\ln(2/\delta)}$ and $\ln(2/\delta) \leq 2\ln(1/\delta)$,

$$\begin{aligned}\varepsilon_{\mathrm{G}}(W,\delta/2) + \alpha &\leq W + \sqrt{2W\ln(2/\delta)} + \frac{8C_1\lambda\sqrt{t\ln(1/\delta)}}{\sigma} \\ &\leq \frac{2(e-1)t\lambda^2}{\sigma^2} + (\sqrt{8(e-1)} + 8C_1)\frac{\lambda\sqrt{t\ln(1/\delta)}}{\sigma},\end{aligned}$$

which is at most ε by the choice of C. □

Proof of Corollary 4.6. Let C_1 be the constant of Corollary 4.5. By Corollary 4.5 with $\lambda = 1/t$ and $\beta = \delta/2$, the composition M is $(\alpha, \delta/2)$-shifted $\sqrt{W}$-GDP for

$$W = \frac{\tilde{V}}{t-1} \qquad \text{and} \qquad \alpha = \frac{4C_1\Delta}{t}\ln\!\left(\frac{4}{\delta}\right) + \frac{W}{2},$$

where $\ln(1+\Delta) = \frac{1}{2\sigma^2} + \frac{\sqrt{2\ln(4t/\delta)}}{\sigma}$. By (5), $\varepsilon_M(\delta) \leq \varepsilon_{\mathrm{G}}(W, \delta/2) + \alpha$.

The shift. Note that $W \leq 2\tilde{V}/t$ and $\ln(4/\delta) = O(\ln^2(t))$. If $\delta = t^{-\theta}$, then $\ln(1+\Delta) = O(\sqrt{\ln(t)})$, so $\Delta = t^{o(1)}$ and $\alpha \leq t^{-1+o(1)} = o(\sqrt{\ln(t)/t})$. In the second regime, by the assumption on the lim sup there is a fixed $\gamma > 0$ such that $8\ln(1/\delta) \leq (\sigma\ln(t)/(1+\gamma))^2$ for all sufficiently large t. Then $\sqrt{2\ln(1/\delta)} \leq \sigma\ln(t)/(2(1+\gamma))$ and $\ln(t) = o(\ln(1/\delta))$, so

$$\ln(1+\Delta) \leq (1+o(1))\frac{\ln(t)}{2(1+\gamma)} \quad \Longrightarrow \quad \alpha \leq t^{-1+\frac{1}{2(1+\gamma)}+o(1)} = o(1/\sqrt{t}).$$

The Gaussian term. Let $k \coloneqq 2\ln(2/\delta) - \ln(1/W)$ be the quantity of Lemma C.1 for W and $\delta/2$. As $W = (1+o(1))\tilde{V}/t$,

$$k = 2\ln(1/\delta) - \ln(t) + O(1) = \begin{cases}(1+o(1))(2\theta-1)\ln(t) & \delta = t^{-\theta}, \\ (1+o(1))\,2\ln(1/\delta) & \text{second regime.}\end{cases}$$

In both regimes $k \to \infty$ and $W \to 0$, so $W/2 = o(\sqrt{kW})$ and Lemma C.1 gives $\varepsilon_{\mathrm{G}}(W, \delta/2) \leq (1+o(1))\sqrt{kW} = (1+o(1))\sqrt{k\tilde{V}/t}$, which equals the claimed bound in each regime. In the second regime it is at least of order $\sqrt{\ln(t)/t}$, so in both regimes α is of lower order, which completes the proof. □

E Proofs for Section 5

In this appendix r is the likelihood ratio of (P, Q).

Proof of Lemma 5.1. As noted before Lemma 5.1, $r\Big(\omega_{1:t}; \bar{P}_t, Q^t\Big) = \frac{1}{t}\sum_{j=1}^{t} r(\omega_j)$. This average is invariant under permutations of the coordinates, and a sample from $\bar{P}_t$ is a sample from $P \times Q^{t-1}$ with its first coordinate moved to a uniformly random position. Therefore, for mutually independent $\omega_0 \sim P$ and $\omega_1, \dots, \omega_t \sim Q$, the random variables $1/R_{Q^t,\bar{P}_t}$ and $R_{\bar{P}_t,Q^t}$ are distributed as

$$\frac{1}{t}\sum_{i=1}^{t} r(\omega_i) \qquad \text{and} \qquad \frac{1}{t}\left(r(\omega_0) + \sum_{i=2}^{t} r(\omega_i)\right), \tag{14}$$

respectively. Let $X_i \coloneqq r(\omega_i) - 1$ for $i \in [t]$, so that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}\big[X_i^2\big] = \tilde{V}$. Since P has density r with respect to Q, we have that $\mathbb{E}[F(r(\omega_0))] = \mathbb{E}[r(\omega_1)F(r(\omega_1))]$ for every nonnegative function F. In particular, $\mathbb{E}[r(\omega_0)] = \mathbb{E}\big[(1+X_1)^2\big] = 1 + \tilde{V}$ and

$$\mathrm{Var}(r(\omega_0)) = \mathbb{E}\Big[(1+X_1)^3\Big] - (1+\tilde{V})^2 = \tilde{V} + \mathbb{E}\Big[X_1^3\Big] - \tilde{V}^2.$$

The claims follow from (14) by the linearity of the expectation and of the variance of independent random variables. □

Proof of Theorem 5.2. Let C_1 be the constant of Lemma 3.2, let $V \coloneqq \tilde{V}/t$ and $u \coloneqq \ln(4/\beta)$, and let $c \coloneqq 1/(100(1+C_1))$. We take $C \geq 4$, so that $u \leq \ln(C/\beta)$ and the assumption gives $(\Delta/t)u \leq c$. We can assume that $\tilde{V} > 0$ (otherwise $P = Q$ and both privacy losses are 0). Since $r \leq 1 + \Delta$ and $\mathop{\mathbb{E}}_{Q}[r] = 1$, we have that $\tilde{V} = \mathop{\mathbb{E}}_{Q}\big[r^2\big] - 1 \leq \Delta$, and therefore $V \leq \Delta/t$. As $u > 1$, also $V \leq Vu \leq (\Delta/t)u \leq c$ and $\Delta/t \leq c$.

Let $\omega_0 \sim P$ and $\omega_1, \dots, \omega_t \sim Q$ be mutually independent, and let $X_i \coloneqq r(\omega_i) - 1$ for $i \in [t]$, so that $\mathbb{E}[X_i] = 0$, $\mathbb{E}\big[X_i^2\big] = \tilde{V}$ and $|X_i| \leq \Delta$. The privacy losses are $L_{Q^t,\bar{P}_t} = -\ln(1/R_{Q^t,\bar{P}_t})$ and $L_{\bar{P}_t,Q^t} = \ln(R_{\bar{P}_t,Q^t})$, and since the claim depends only on their distributions, by (14) in the proof of Lemma 5.1 we can take

$$1/R_{Q^t,\bar{P}_t} = 1 + \frac{1}{t}\sum_{i=1}^{t} X_i \qquad \text{and} \qquad R_{\bar{P}_t,Q^t} = \frac{1}{t}\left(r(\omega_0) + \sum_{i=2}^{t} r(\omega_i)\right).$$

By Lemma 5.1, $\mathbb{E}\Big[R_{\bar{P}_t,Q^t}\Big] = 1 + V$, and the variance V^* of $R_{\bar{P}_t,Q^t}$ satisfies $|V^* - V| = |\mathbb{E}\big[X_1^3\big] - \tilde{V}^2|/t^2 \leq 2(\Delta/t)V$, where we used that $|\mathbb{E}\big[X_1^3\big]| \leq \Delta\tilde{V}$ and $\tilde{V}^2 \leq \Delta\tilde{V}$. Therefore $V^* \leq 3V$ and $|\sqrt{V^*} - \sqrt{V}| \leq |V^* - V|/\sqrt{V} \leq 2(\Delta/t)\sqrt{V}$.

The centered terms X_i/t of $1/R_{Q^t,\bar{P}_t}$ are bounded by Δ/t in absolute value, and the centered terms of $R_{\bar{P}_t,Q^t}$ are bounded by $2\Delta/t$, since $|r(\omega_0) - 1| \leq \Delta$ almost surely (as $P \ll Q$) and $\tilde{V} \leq \Delta$. By Lemma 3.2 with $2\Delta/t$ in place of Δ and $\beta/2$ in place of β, and since $\mathbb{P}\Big(|G| > \sqrt{2u}\Big) \leq e^{-u} < \beta/2$ for $G \sim \mathcal{N}(0,1)$, in each direction there is a standard normal random variable G such that, except with probability β, we have that $|G| \leq \sqrt{2u}$ and

$$1/R_{Q^t,\bar{P}_t} = 1 + b \quad \text{for} \quad b = \sqrt{V}\,G + \zeta, \quad \text{respectively} \quad R_{\bar{P}_t,Q^t} = 1 + b \quad \text{for} \quad b = V + \sqrt{V^*}\,G + \zeta,$$

where $|\zeta| \leq 2C_1(\Delta/t)u$. By the choice of c, on this event $|\zeta| \leq 1$ and $|b| \leq V + \sqrt{6Vu} + |\zeta| \leq 1/2$. Thus $\ln(1+b) = b - b^2/2 + \psi$ with $|\psi| \leq |b|^3 \leq b^2$, and we get that

$$L_{Q^t,\bar{P}_t} - \left(\frac{V}{2} - \sqrt{V}\,G\right) = -\zeta + \frac{b^2 - V}{2} - \psi,$$

$$L_{\bar{P}_t,Q^t} - \left(\frac{V}{2} + \sqrt{V}\,G\right) = \zeta + \frac{V - b^2}{2} + \left(\sqrt{V^*} - \sqrt{V}\right)G + \psi.$$

In both directions $b^2 \le 3(V^2 + 3VG^2 + \zeta^2) \le 21Vu + 3|\zeta|$, where we used that $V^* \le 3V$, $V^2 \le V \le Vu$ and $\zeta^2 \le |\zeta|$. Further, $|\sqrt{V^*} - \sqrt{V}|\,|G| \le 2(\Delta/t)\sqrt{2Vu} \le 3\Delta/t$. Therefore both differences are at most

$$|\zeta| + \frac{V + b^2}{2} + b^2 + \frac{3\Delta}{t} \le \frac{11}{2}|\zeta| + 32Vu + \frac{3\Delta}{t} \le (11C_1 + 35)\,\frac{\Delta}{t}\,u \le \frac{C\Delta}{t}\ln(C/\beta),$$

where the last inequality holds for $C \ge 11C_1 + 35$, since $u \le \ln(C/\beta)$. Finally, $V/2 \pm \sqrt{V}\,G$ is distributed as $\mathcal{N}(V/2, V)$, and therefore, except with probability β, each of the two privacy losses is within $(C\Delta/t)\ln(C/\beta)$ of a $\mathcal{N}(V/2, V)$ random variable. As noted after Definition 2.7, this coupling gives the claim. □

We now prove the claim for unbounded likelihood ratios stated before Corollary 5.3.

Lemma E.1 (Conditioning for random allocation)**.** *Let c and C be the constants of Theorem 5.2, let $t \in \mathbb{N}$, $\Delta > 0$ and $0 < \beta \le 1$ with $(\Delta/t)\ln(2C/\beta) \le c$, and let E be an event such that $P(E) = Q(E) \ge 1 - \beta/(2t)$ and $|r(\omega; P, Q) - 1| \le \Delta$ for every $\omega \in E$. Then the privacy losses of $(\bar{P}_t, Q^t)$ and of $(Q^t, \bar{P}_t)$ are $((C\Delta/t)\ln(2C/\beta), \beta)$-approximately $(\tilde{V}_E/(2t), \tilde{V}_E/t)$-Gaussian-concentrated, where $P_E \coloneqq P(\cdot \mid E)$, $Q_E \coloneqq Q(\cdot \mid E)$ and $\tilde{V}_E \coloneqq \chi^2_{P_E,Q_E}$.*

Proof. Let $\bar{P}_{t,E} \coloneqq \frac{1}{t}\sum_{j=1}^{t} Q_E^{j-1} \times P_E \times Q_E^{t-j}$ be the distribution $\bar{P}_t$ with (P_E, Q_E) in place of (P, Q). Since $P = P(E)\,P_E$ and $Q = P(E)\,Q_E$ on E, the restrictions of $\bar{P}_t$ and Q^t to E^t are $P(E)^t\,\bar{P}_{t,E}$ and $P(E)^t\,Q_E^t$, and the complement of E^t has probability $1 - P(E)^t \le t(1 - P(E)) \le \beta/2$ under both distributions. Thus on E^t the privacy loss of $(\bar{P}_t, Q^t)$ is equal to that of $(\bar{P}_{t,E}, Q_E^t)$, and given E^t a sample from $\bar{P}_t$ is a sample from $\bar{P}_{t,E}$. Therefore we can couple the two privacy losses so that they are equal except with probability at most $\beta/2$, and the same holds for the privacy losses of $(Q^t, \bar{P}_t)$ and $(Q_E^t, \bar{P}_{t,E})$. By Theorem 5.2 for (P_E, Q_E), whose likelihood ratio is equal to r on E, and $\beta/2$ in place of β, the privacy losses of $(\bar{P}_{t,E}, Q_E^t)$ and $(Q_E^t, \bar{P}_{t,E})$ are $((C\Delta/t)\ln(2C/\beta), \beta/2)$-approximately $(\tilde{V}_E/(2t), \tilde{V}_E/t)$-Gaussian-concentrated. By the coupling and the transitivity noted after Definition 2.7, the same holds for the privacy losses of $(\bar{P}_t, Q^t)$ and $(Q^t, \bar{P}_t)$ with β in place of $\beta/2$, as claimed. □

Proof of Corollary 5.3. Let $E \coloneqq \{\omega : |\ell(\omega; P, Q)| \le \ln(1 + \Delta)\}$ for Δ of (7), and let c_1 and $C_1 \ge 1$ be the constants of Theorem 5.2. Let $w \coloneqq \sqrt{2\ln(2t/\beta)}$, so that $\ln(1 + \Delta) = 1/(2\sigma^2) + w/\sigma$. The privacy loss $\ell(\omega; P, Q)$ is $1/(2\sigma^2) + Z/\sigma$ for $\omega \sim P$ and $-1/(2\sigma^2) + Z/\sigma$ for $\omega \sim Q$, where Z is a standard normal random variable. Thus E is the event $\{-w - 1/\sigma \le Z \le w\}$ under P and the event $\{-w \le Z \le w + 1/\sigma\}$ under Q, and therefore $P(E^c) = Q(E^c) = \overline{\Phi}(w) + \overline{\Phi}(w + 1/\sigma) \le e^{-w^2/2} = \beta/(2t)$. As in the proof of Corollary 4.4, $|r - 1| \le \Delta$ on E. Thus, by Lemmas 2.8 and E.1, for $c = c_1$ and every $C \ge 2C_1$ the pairs $(\bar{P}_t, Q^t)$ and $(Q^t, \bar{P}_t)$ are (α, β)-shifted $\sqrt{\tilde{V}_E/t}$-GDP. Since the likelihood ratio of (P_E, Q_E) is equal to r on E, we have that

$$\tilde{V}_E = \frac{\mathop{\mathbb{E}}_{\omega\sim Q}\left[(r(\omega; P, Q) - 1)^2 \mathbb{1}\{\omega \in E\}\right]}{Q(E)} \le \frac{\tilde{V}}{1 - \beta/(2t)},$$

and thus $\tilde{V}_E/t \le \tilde{V}/(t - \beta/2) \le \tilde{V}/(t - 1)$. As $g_V(\varepsilon)$ is nondecreasing in V, the two pairs are (α, β)-shifted $\sqrt{\tilde{V}/(t-1)}$-GDP. By Claim 2.5 and [FS25, Theorem 3.1], which holds for every $\varepsilon \in \mathbb{R}$, random allocation of the Gaussian mechanism is $(\varepsilon, g_{\tilde{V}/(t-1)}(\varepsilon - \alpha) + \beta)$-DP for every $\varepsilon \in \mathbb{R}$. That is, it is (α, β)-shifted $\sqrt{\tilde{V}/(t-1)}$-GDP, which gives the claim. □

Proof of Corollary 5.4. Let c_1 and $C_1 \geq 1$ be the constants of Theorem 5.2, let $\beta \coloneqq \delta/2$, $H \coloneqq \ln(t/\delta) \geq \ln(4)$ and $K \coloneqq 1 + \log_2(4C_1)$, and let $C \geq \max\{3,\, 2C_1,\, \sqrt{K/c_1},\, \sqrt{8(e-1)} + 2C_1K\}$. As $\ln(1+y) \leq y$, (8) implies that $\sqrt{t/(C\ln(1/\delta))} \geq C\sqrt{H}$, that is, $t \geq C^3 H \ln(1/\delta)$. Further, as in the proof of Claim D.4 (here $\ln(2t/\beta) = \ln(4t/\delta)$ and $C \geq 5/2$), condition (8) gives that

$$\Delta \leq \frac{\sqrt{t}}{\sigma\sqrt{C\ln(1/\delta)}}.$$

For $\delta \leq 1/2$ we have that $\ln(2C_1/\beta) = \ln(4C_1) + \ln(1/\delta) \leq K\ln(1/\delta)$, and since $\ln(1/\delta) \leq t/C^3$ and $\sigma \geq 1$,

$$\frac{\Delta}{t}\ln(2C_1/\beta) \leq \frac{K\sqrt{\ln(1/\delta)}}{\sigma\sqrt{Ct}} \leq \frac{K}{C^2} \leq c_1.$$

Therefore, by Corollary 5.3, which holds with the constants c_1 and $2C_1$ by its proof, 1-out-of-t random allocation of the Gaussian mechanism is $(\alpha_1, \delta/2)$-shifted $\sqrt{\tilde{V}/(t-1)}$-GDP, where

$$\alpha_1 \coloneqq 2C_1\frac{\Delta}{t}\ln(2C_1/\beta) \leq \frac{2C_1K\sqrt{\ln(1/\delta)}}{\sigma\sqrt{t}}.$$

By (5), it is $(\varepsilon_{\mathrm{G}}(\tilde{V}/(t-1), \delta/2) + \alpha_1, \delta)$-DP. Finally, $\tilde{V}/(t-1) \leq 2\tilde{V}/t \leq 2(e-1)/(\sigma^2 t)$ by Claim D.2. Since g_V is nondecreasing in V and $\varepsilon_{\mathrm{G}}(V,\delta) \leq V/2 + \sqrt{2V\ln(1/\delta)}$, we have that $\varepsilon_{\mathrm{G}}(\tilde{V}/(t-1), \delta/2) \leq \varepsilon_{\mathrm{G}}(2(e-1)/(\sigma^2 t), \delta/2) \leq (e-1)/(\sigma^2 t) + \sqrt{4(e-1)\ln(2/\delta)/(\sigma^2 t)}$, and since $\ln(2/\delta) \leq 2\ln(1/\delta)$,

$$\varepsilon_{\mathrm{G}}\left(\frac{\tilde{V}}{t-1}, \frac{\delta}{2}\right) \leq \frac{e-1}{\sigma^2 t} + \frac{\sqrt{8(e-1)\ln(1/\delta)}}{\sigma\sqrt{t}}.$$

Thus $\varepsilon_{\mathrm{G}}(\tilde{V}/(t-1), \delta/2) + \alpha_1 \leq \varepsilon$ by the choice of C, which gives the claim. □

Proof of Corollary 5.5. Let $\beta \coloneqq \delta/2$. The proof of Corollary 4.6 uses only that the parameter of the shifted GDP guarantee is $(1 + o(1))\,\tilde{V}/t$ and that the shift is $o(\sqrt{\ln(t)/t})$ in the first regime and $o(1/\sqrt{t})$ in the second. The proof of Corollary 5.3 shows that the pairs $(\bar{P}_t, Q^t)$ and $(Q^t, \bar{P}_t)$ are shifted GDP with the parameter $\tilde{V}/(t-1)$, and both the condition of Corollary 5.3 and the shift are given by $(\Delta/t)\ln(2C/\delta)$, so we also need that this condition holds for all sufficiently large t. By the proof of Corollary 4.6, $\Delta = t^{o(1)}$ and $\ln(2C/\delta) = O(\ln(t))$ in the first regime, and $\Delta \leq t^{1/(2(1+\gamma))+o(1)}$ for the γ of that proof and $\ln(2C/\delta) = O(\ln^2(t))$ in the second. Thus $(\Delta/t)\ln(2C/\delta)$ is $t^{-1+o(1)}$ and at most $t^{-1+1/(2(1+\gamma))+o(1)}$, respectively, which gives the order of the shift and shows that the condition holds for all sufficiently large t. □

F Proofs for Section 6

In this appendix, (P, Q), L, Λ, $\tilde{L}$ and $(\tilde{P}, \tilde{Q})$ are as in Definition 6.1. If $L \notin \Lambda$, we denote by $\ell^- < \ell^+$ the two knots adjacent to L. The privacy losses of (P, Q) and $(\tilde{P}, \tilde{Q})$ are distributed as L and $\tilde{L}$, and therefore, by Definition 2.1, for every $\kappa \geq 0$,

$$\boldsymbol{H}_\kappa(P \parallel Q) = \mathbb{E}\Big[[1 - \kappa e^{-L}]_+\Big] \qquad \text{and} \qquad \boldsymbol{H}_\kappa\Big(\tilde{P} \,\Big\|\, \tilde{Q}\Big) = \mathbb{E}\Big[[1 - \kappa e^{-\tilde{L}}]_+\Big].$$

Claim F.1 (CtD discretization of randomized response)**.** *Let $0 < \varepsilon_0 < h$, let (P, Q) be randomized response over $\{0, 1\}$ with $P(1) = Q(0) = e^{\varepsilon_0}/(1 + e^{\varepsilon_0})$, and let $(\tilde{P}, \tilde{Q})$ be its CtD pair on the grid $h\mathbb{Z}$. Then P and Q are $(\varepsilon_0, 0)$-close, but for every $0 \leq \delta < \tanh(\varepsilon_0/2)$ and $\varepsilon \geq 0$, the distributions $\tilde{P}$ and $\tilde{Q}$ are (ε, δ)-close exactly when $\varepsilon \geq h + \ln(1 - \delta/\tilde{P}(h))$, where $\tilde{P}(h) = \tanh(\varepsilon_0/2)/(1 - e^{-h})$. This threshold tends to h as $\delta \to 0$.*

Proof. The privacy loss of (P, Q) is $\pm\varepsilon_0$, and (9) splits the mass at $\pm\varepsilon_0$ between $\pm h$ and 0, with $\tilde{P}(h) = \tanh(\varepsilon_0/2)/(1-e^{-h})$ and $\tilde{P}(-h) = e^{-h}\tilde{P}(h)$. Both pairs are symmetric ($Q(\omega) = P(1-\omega)$ and $\tilde{Q}(\ell) = \tilde{P}(-\ell)$), and for every $\varepsilon \geq 0$ their privacy profiles in both directions are

$$\boldsymbol{H}_{e^\varepsilon}(P \parallel Q) = \frac{[e^{\varepsilon_0} - e^{\varepsilon}]_+}{1+e^{\varepsilon_0}} \qquad \text{and} \qquad \boldsymbol{H}_{e^\varepsilon}\Big(\tilde{P} \,\Big\|\, \tilde{Q}\Big) = \tilde{P}(h)\,[1 - e^{\varepsilon - h}]_+.$$

The first is zero for $\varepsilon \geq \varepsilon_0$. Since $\delta < \tanh(\varepsilon_0/2) = \tilde{P}(h)(1-e^{-h})$, the second is at most δ exactly when $e^{\varepsilon-h} \geq 1 - \delta/\tilde{P}(h)$. □

Proof of Lemma 6.2. We bound the correction $\zeta \coloneqq \tilde{L} - L$, which is 0 when $L \in \Lambda$. Let $L = \ell \notin \Lambda$, $h_0 \coloneqq \ell^+ - \ell^- \leq h$ and $p \coloneqq \mathbb{P}\Big(\tilde{L} = \ell^+ \mid L = \ell\Big)$. Given $L = \ell$, the correction is $\ell^- - \ell \in (-h_0, 0)$ with probability $1-p$ and $\ell^- - \ell + h_0$ otherwise, and thus $|\zeta| \leq h_0 \leq h$. By (10), $\mathbb{E}\Big[e^{-\zeta} \mid L = \ell\Big] = 1$, which means that $\ell^- - \ell = \ln(1 - p + pe^{-h_0})$. Therefore, given $L = \ell$, the correction is

$$\zeta = \ln(1 - p + pe^{-h_0}) + h_0\, \mathbb{1}\{\tilde{L} = \ell^+\}, \tag{15}$$

where the indicator has mean p given $L = \ell$. Thus $\mathbb{E}[\zeta \mid L = \ell] \geq 0$, since $e^{-\mathbb{E}[\zeta \mid L=\ell]} \leq \mathbb{E}\Big[e^{-\zeta} \mid L = \ell\Big] = 1$ by Jensen's inequality. At the same time, by Hoeffding's lemma [Hoe63] for a random variable in an interval of length h_0,

$$0 = \ln\Big(\mathbb{E}\Big[e^{-\zeta} \mid L = \ell\Big]\Big) \leq -\mathbb{E}[\zeta \mid L = \ell] + \frac{h_0^2}{8},$$

and thus $\mathbb{E}[\zeta \mid L = \ell] \leq h_0^2/8 \leq h^2/8$. Taking expectations, we get the bounds on $\mu_{\tilde{P},\tilde{Q}} - \mu_{P,Q} = \mathbb{E}[\zeta]$.

To bound $\mathrm{Var}(\zeta)$ we let $c_h \coloneqq \ln(\cosh(h/2))$ and show that $\mathbb{E}\big[(\zeta - c_h)^2 \mid L\big] \leq h^2/4$ almost surely, which suffices since $\mathrm{Var}(\zeta) \leq \mathbb{E}\big[(\zeta - c_h)^2\big]$. When $L \in \Lambda$, the left-hand side is $c_h^2 \leq h^2/4$, as $0 \leq \ln(\cosh(h/2)) \leq h/2$. When $L = \ell \notin \Lambda$, by (15) the conditional variance of ζ is $h_0^2 p(1-p)$ and its conditional mean is $m_{h_0}(p) \coloneqq h_0 p + \ln(1 - p + pe^{-h_0})$. The function m_{h_0} is concave on $[0,1]$, with $m_{h_0}(0) = m_{h_0}(1) = 0$, $m_{h_0}(1/2) = c_{h_0} \coloneqq \ln(\cosh(h_0/2)) \in [0, h_0/2]$ and $m'_{h_0}(1/2) = h_0 - 2\tanh(h_0/2) \in [0, h_0]$. By concavity, the chord of m_{h_0} between p and $1/2$ has slope between $m'_{h_0}(1/2)$ and the slope $\pm 2c_{h_0}$ of the chord from $1/2$ to the endpoint of $[0,1]$ on the side of p. Both slopes lie in $[-h_0, h_0]$, and thus $|m_{h_0}(p) - c_{h_0}| \leq h_0|p - 1/2|$. Therefore

$$\mathbb{E}\Big[(\zeta - c_{h_0})^2 \mid L = \ell\Big] = h_0^2 p(1-p) + \big(m_{h_0}(p) - c_{h_0}\big)^2 \leq h_0^2 p(1-p) + h_0^2\Big(p - \frac{1}{2}\Big)^2 = \frac{h_0^2}{4}.$$

Now $c_h \geq c_{h_0}$ and $m_{h_0}(p) \geq 0$, and therefore

$$\mathbb{E}\Big[(\zeta - c_h)^2 \mid L = \ell\Big] = \mathbb{E}\Big[(\zeta - c_{h_0})^2 \mid L = \ell\Big] - 2(c_h - c_{h_0})\, m_{h_0}(p) + c_h^2 - c_{h_0}^2 \leq \frac{h_0^2}{4} + c_h^2 - c_{h_0}^2 \leq \frac{h^2}{4},$$

where the last inequality uses that

$$c_h^2 - c_{h_0}^2 = \int_{h_0}^{h} \ln(\cosh(s/2)) \tanh(s/2)\, ds \leq \int_{h_0}^{h} \frac{s}{2}\, ds = \frac{h^2 - h_0^2}{4}.$$

For the covariance, let $\bar{\zeta} \coloneqq \mathbb{E}[\zeta \mid L]$. As ζ is bounded and $V_{P,Q} < \infty$, we have that

$$\mathrm{Cov}(L, \zeta) = \mathbb{E}[(L - \mu_{P,Q})\,\zeta] = \mathbb{E}\Big[(L - \mu_{P,Q})\,\bar{\zeta}\Big] = \mathrm{Cov}(L, \bar{\zeta}).$$

We showed that $\bar{\zeta} \in [0, h^2/8]$, and therefore $\mathrm{Var}(\bar{\zeta}) \le (h^2/16)^2$ by Popoviciu's inequality. By the Cauchy–Schwarz inequality, $|\operatorname{Cov}(L, \zeta)| \le (h^2/16)\sqrt{V_{P,Q}}$, and the bounds on $V_{\tilde{P},\tilde{Q}} - V_{P,Q}$ follow from $V_{\tilde{P},\tilde{Q}} = \mathrm{Var}(\tilde{L}) = V_{P,Q} + \mathrm{Var}(\zeta) + 2\operatorname{Cov}(L, \zeta)$ and $0 \le \mathrm{Var}(\zeta) \le h^2/4$. □

Proof of Theorem 6.3. By (10) and the conditional Jensen inequality for the convex function $[1 - \kappa \cdot]_+$, we have that $\mathbb{E}\Big[[1 - \kappa e^{-\tilde{L}}]_+ \mid L\Big] \ge [1 - \kappa e^{-L}]_+$. Taking expectations, we get that $(\tilde{P}, \tilde{Q})$ dominates (P, Q), and the domination of the products follows by applying the mappings of Lemma 2.3 coordinatewise. Multiplying the bounds of Lemma 6.2 by t, we obtain the bounds on $\mu_{\tilde{P}^t,\tilde{Q}^t} = t\mu_{\tilde{P},\tilde{Q}}$ and $V_{\tilde{P}^t,\tilde{Q}^t} = tV_{\tilde{P},\tilde{Q}}$.

The pair $(\tilde{P}, \tilde{Q})$ is mutually absolutely continuous, and its privacy loss is distributed as $\tilde{L} = L + (\tilde{L} - L)$. By the proof of Lemma 6.2, $|\tilde{L} - L| \le h$. Thus the correction $\tilde{L} - L$ deviates from its mean by at most $2h$, and therefore $|\tilde{L} - \mu_{\tilde{P},\tilde{Q}}| \le |L - \mu_{P,Q}| + 2h \le \Delta + 2h$. Thus Theorem 3.3 with E the whole space and $\Delta + 2h$ in place of Δ gives the Gaussian concentration of the privacy loss of $(\tilde{P}^t, \tilde{Q}^t)$. □

Proof of Corollary 6.4. The first inequality follows from the domination of Theorem 6.3. We denote the privacy losses of the original and of the discretized composition by L and $\tilde{L}$, their means by μ and $\tilde{\mu}$ and their variances by $V = t \cdot V_{P,Q}$ and $\tilde{V}$. Let $G \sim \mathcal{N}(\mu, V)$ and $\tilde{G} \sim \mathcal{N}(\tilde{\mu}, \tilde{V})$. By Theorem 3.3 with E the whole space and by Theorem 6.3, $G \preceq_{(C\Delta \ln(2/\beta),\beta)} L$ and $\tilde{L} \preceq_{(C(\Delta+2h)\ln(2/\beta),\beta)} \tilde{G}$. By (11), $\tilde{G} \preceq_{(\alpha_1,\beta)} G$ for $\alpha_1 \coloneqq \tilde{\mu} - \mu + |\sqrt{\tilde{V}} - \sqrt{V}|\sqrt{2\ln(1/\beta)}$. By Theorem 6.3, $0 \le \tilde{\mu} - \mu \le th^2/8$. If $\tilde{V} \ge V$, then $\sqrt{\tilde{V}} - \sqrt{V} \le (\tilde{V} - V)/(2\sqrt{V}) \le th^2/(8\sqrt{V}) + h^2\sqrt{t}/16 = h^2\sqrt{t}/(8\sqrt{V_{P,Q}}) + h^2\sqrt{t}/16$, and otherwise $\sqrt{V} - \sqrt{\tilde{V}} \le (V - \tilde{V})/\sqrt{V} \le h^2\sqrt{t}/8$. Chaining the three relations and applying the bound on the hockey-stick divergence [FS26, Claim 4.2], we get that for every $\varepsilon \in \mathbb{R}$,

$$\boldsymbol{H}_{e^\varepsilon}(\tilde{L}) \le \boldsymbol{H}_{e^{\varepsilon-\alpha}}(L) + 3\beta,$$

where

$$\alpha \coloneqq 2C(\Delta + h)\ln(2/\beta) + \frac{th^2}{8} + \frac{h^2\sqrt{t}}{8}\left(\frac{1}{\sqrt{V_{P,Q}}} + 1\right)\sqrt{2\ln(1/\beta)}.$$

At $\varepsilon = \varepsilon(\delta - 3\beta) + \alpha$ the right-hand side is at most δ, and therefore $\tilde{\varepsilon}(\delta) \le \varepsilon(\delta - 3\beta) + \alpha$. □

G Subexponential tails

In this appendix we show that if the deviations of the privacy losses from their means have exponentially decaying tails, then we can replace the mean $\mu'_{P,Q}$ and the variance $V'_{P,Q}$ of the truncated privacy loss in Theorem 3.3 by the mean $\mu_{P,Q}$ and the variance $V_{P,Q}$ of the privacy loss itself, which do not depend on the truncation set E. We use the notation of Section 3.

Theorem G.1 (Gaussian concentration under subexponential tails)**.** *There is a universal constant $C > 0$ such that the following holds. Let $\Delta > 0$, $0 < \beta \le 1$ and $0 \le \beta_0 \le \beta/(4t)$, and assume that $L_{P,Q}$ has a finite mean and that for every $\ell \ge \Delta$,*

$$\mathbb{P}(|L_{P,Q} - \mu_{P,Q}| > \ell) \le \beta_0 e^{1-\ell/\Delta}.$$

Then L_{P^t,Q^t} is $(C\Delta\ln(2/\beta), \beta)$-approximately $(\mu_{P^t,Q^t}, V_{P^t,Q^t})$-Gaussian-concentrated.

If $\beta_0 = 0$, then the hypothesis says that $|L_{P,Q} - \mu_{P,Q}| \le \Delta$ almost surely, and the conclusion also follows from Theorem 3.3 with E the whole space. By Lemma 2.8, we also get that

$\boldsymbol{H}_{e^\varepsilon}(P^t \parallel Q^t) \le g_{V_{P^t,Q^t}}(\varepsilon - \alpha) + \beta$ for every ε, where $\alpha \coloneqq \mu_{P^t,Q^t} - V_{P^t,Q^t}/2 + C\Delta\ln(2/\beta)$. Compared with Theorem 3.3, the theorem assumes a tail bound at every scale beyond Δ rather than at a single cutoff. Such bounds hold for subexponential privacy losses. Specifically, if the random variable $e^{|L_{P,Q}-\mu_{P,Q}|/\Delta}$ has expectation at most 2, then by Markov's inequality the tail assumption holds with $\Delta\ln(8t/\beta)$ in place of Δ and with $\beta_0 \coloneqq \beta/(4t)$. The cutoff grows only logarithmically in t/β.

Proof. Let C_1 be the constant of Lemma 3.2. We prove the claim with $C \coloneqq \frac{9}{2}C_1 + 3$. Let $L_1, \dots, L_t$ be independent copies of $L_{P,Q}$, let $\mu \coloneqq \mu_{P,Q}$, and let $L'_i \coloneqq \mu + (L_i - \mu)\mathbb{1}\{|L_i - \mu| \le \Delta\}$ be the privacy losses truncated at their mean, that is, set to μ where they deviate from it by more than Δ. (The truncation at zero of Theorem 3.3 would add $|\mu|$ to the deviations of the truncated privacy losses from their mean.) By the tail assumption with $\ell = \Delta$ and a union bound, $\sum_i L'_i = \sum_i L_i$ except with probability at most $t\beta_0 \le \beta/4$.

Since $L_i - L'_i = (L_i - \mu)\mathbb{1}\{|L_i - \mu| > \Delta\}$, by integrating the tail assumption we get that

$$\mathbb{E}\big[|L_i - L'_i|\big] = \Delta\,\mathbb{P}(|L_i - \mu| > \Delta) + \int_\Delta^\infty \mathbb{P}(|L_i - \mu| > \ell)\,d\ell \le 2\beta_0\Delta,$$
$$\mathbb{E}\Big[(L_i - L'_i)^2\Big] = \Delta^2\,\mathbb{P}(|L_i - \mu| > \Delta) + \int_\Delta^\infty 2\ell\,\mathbb{P}(|L_i - \mu| > \ell)\,d\ell \le 5\beta_0\Delta^2.$$

In particular, since $\mathbb{E}[L_i] = \mu$, the mean of L'_i is within $2\beta_0\Delta$ of μ, and as $|L'_i - \mu| \le \Delta$, the truncated privacy loss L'_i deviates from its mean by at most $(1 + 2\beta_0)\Delta \le \frac{3}{2}\Delta$. By Lemma 3.2 with $\frac{3}{2}\Delta$ in place of Δ and $\beta/4$ in place of β, we can couple $\sum_i L'_i$ with $G' \sim \mathcal{N}(\mathbb{E}[\sum_i L'_i], \mathrm{Var}(\sum_i L'_i))$ so that $|\sum_i L'_i - G'| \le \frac{3}{2}C_1\Delta\ln(8/\beta)$ except with probability $\beta/4$ (as in the proof of Theorem 3.3, we shift the coupling of the centered sum by its mean). Thus $|\sum_i L_i - G'| \le \alpha_1 \coloneqq \frac{3}{2}C_1\Delta\ln(8/\beta)$ except with probability $\beta/2$, and as noted after Definition 2.7, $L_{P^t,Q^t} \preceq_{(\alpha_1,\beta/2)} G'$ and $G' \preceq_{(\alpha_1,\beta/2)} L_{P^t,Q^t}$.

Further, $|\mu_{P^t,Q^t} - \mathbb{E}[\sum_i L'_i]| \le 2t\beta_0\Delta \le \beta\Delta/2$. By the triangle inequality for the standard deviation and the independence of the differences $L_i - L'_i$, we also get that

$$\left|\sqrt{V_{P^t,Q^t}} - \sqrt{\mathrm{Var}\left(\sum_{i=1}^t L'_i\right)}\right| \le \sqrt{\mathrm{Var}\left(\sum_{i=1}^t (L_i - L'_i)\right)} \le \left(\sum_{i=1}^t \mathbb{E}\Big[(L_i - L'_i)^2\Big]\right)^{1/2} \le \Delta\sqrt{5\beta/4}.$$

Let $G \sim \mathcal{N}(\mu_{P^t,Q^t}, V_{P^t,Q^t})$. By (11) with $\beta/2$ in place of β, we have that $G' \preceq_{(\alpha_2,\beta/2)} G$ and $G \preceq_{(\alpha_2,\beta/2)} G'$ for

$$\alpha_2 \coloneqq \frac{\beta\Delta}{2} + \Delta\sqrt{\frac{5}{2}\beta\ln(2/\beta)} \le 3\Delta\ln(2/\beta),$$

where we used that $\beta \le 2\ln(2/\beta)$ and $\frac{5}{2}\beta \le 4\ln(2/\beta)$ for $\beta \le 1$. Chaining these relations with those for L_{P^t,Q^t} (the loss shifts and the tail shifts add up, as noted after Definition 2.7), we get that $L_{P^t,Q^t} \preceq_{(\alpha_1+\alpha_2,\beta)} G$ and $G \preceq_{(\alpha_1+\alpha_2,\beta)} L_{P^t,Q^t}$. This gives the claim, since $\ln(8/\beta) \le 3\ln(2/\beta)$ and therefore $\alpha_1 + \alpha_2 \le (\frac{9}{2}C_1 + 3)\Delta\ln(2/\beta) = C\Delta\ln(2/\beta)$. □