Appendices

A

The Gaussian Mechanism

Let $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^d$ be an arbitrary d-dimensional function, and define its ℓ_2 sensitivity to be $\Delta_2 f = \max_{\text{adjacent}x,y} \|f(x) - f(y)\|_2$. The *Gaussian Mechanism with parameter* σ adds noise scaled to $\mathcal{N}(0, \sigma^2)$ to each of the d components of the output.

Theorem A.1. Let $\varepsilon \in (0, 1)$ be arbitrary. For $c^2 > 2\ln(1.25/\delta)$, the Gaussian Mechanism with parameter $\sigma \geq c\Delta_2 f/\varepsilon$ is (ε, δ)-differentially private.

Proof. There is a database D and a query f, and the mechanism will return $f(D)+\eta$, where the noise is normally distributed. We are adding noise $\mathcal{N}(0, \sigma^2)$. For now, assume we are talking about real-valued functions, so

$$\Delta f = \Delta_1 f = \Delta_2 f.$$

We are looking at

$$\left| \ln \frac{e^{(-1/2\sigma^2)x^2}}{e^{(-1/2\sigma^2)(x+\Delta f)^2}} \right|. \tag{A.1}$$

We are investigating the probability, given that the database is D, of observing an output that occurs with a very different probability

under D than under an adjacent database D', where the probability space is the noise generation algorithm. The numerator in the ratio above describes the probability of seeing $f(D) + x$ when the database is D, the denominator corresponds the probability of seeing *this same value* when the database is D'. This is a ratio of probabilities, so it is always positive, but the logarithm of the ratio may be negative. Our random variable of interest — the privacy loss — is

$$\ln \frac{e^{(-1/2\sigma^2)x^2}}{e^{(-1/2\sigma^2)(x+\Delta f)^2}}$$

and we are looking at its absolute value.

$$\begin{aligned}\left|\ln \frac{e^{(-1/2\sigma^2)x^2}}{e^{(-1/2\sigma^2)(x+\Delta f)^2}}\right| &= |\ln e^{(-1/2\sigma^2)[x^2-(x+\Delta f)^2]}| \\ &= |-\frac{1}{2\sigma^2}[x^2-(x^2+2x\Delta f+\Delta f^2)]| \\ &= |\frac{1}{2\sigma^2}(2x\Delta f+(\Delta f)^2)|. \end{aligned} \tag{A.2}$$

This quantity is bounded by ε whenever $x < \sigma^2\varepsilon/\Delta f - \Delta f/2$. To ensure privacy loss bounded by ε with probability at least $1-\delta$, we require

$$\Pr[|x| \geq \sigma^2\varepsilon/\Delta f - \Delta f/2] < \delta,$$

and because we are concerned with $|x|$ we will find σ such that

$$\Pr[x \geq \sigma^2\varepsilon/\Delta f - \Delta f/2] < \delta/2.$$

We will assume throughout that $\varepsilon \leq 1 \leq \Delta f$.

We will use the tail bound

$$\Pr[x > t] \leq \frac{\sigma}{\sqrt{2\pi}} e^{-t^2/2\sigma^2}.$$

We require:

$$\begin{aligned} &\frac{\sigma}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/2\sigma^2} < \delta/2 \\ \Leftrightarrow\ &\sigma\frac{1}{t}e^{-t^2/2\sigma^2} < \sqrt{2\pi}\delta/2 \\ \Leftrightarrow\ &\frac{t}{\sigma}e^{t^2/2\sigma^2} > 2/\sqrt{2\pi}\delta \\ \Leftrightarrow\ &\ln(t/\sigma) + t^2/2\sigma^2 > \ln(2/\sqrt{2\pi}\delta). \end{aligned}$$

Taking $t = \sigma^2\varepsilon/\Delta f - \Delta f/2$, we get

$$\ln((\sigma^2\varepsilon/\Delta f - \Delta f/2)/\sigma) + (\sigma^2\varepsilon/\Delta f - \Delta f/2)^2/2\sigma^2 > \ln(2/\sqrt{2\pi}\delta)$$
$$= \ln\left(\sqrt{\frac{2}{\pi}}\frac{1}{\delta}\right).$$

Let us write $\sigma = c\Delta f/\varepsilon$; we wish to bound c. We begin by finding the conditions under which the first term is non-negative.

$$\begin{aligned}\frac{1}{\sigma}\left(\sigma^2\frac{\varepsilon}{\Delta f} - \frac{\Delta f}{2}\right) &= \frac{1}{\sigma}\left[\left(c^2\frac{(\Delta f)^2}{\varepsilon^2}\right)\frac{\varepsilon}{\Delta f} - \frac{\Delta f}{2}\right] \\ &= \frac{1}{\sigma}\left[c^2\left(\frac{\Delta f}{\varepsilon}\right) - \frac{\Delta f}{2}\right] \\ &= \frac{\varepsilon}{c\Delta f}\left[c^2\left(\frac{\Delta f}{\varepsilon}\right) - \frac{\Delta f}{2}\right] \\ &= c - \frac{\varepsilon}{2c}.\end{aligned}$$

Since $\varepsilon \leq 1$ and $c \geq 1$, we have $c - \varepsilon/(2c) \geq c - 1/2$. So $\ln(\frac{1}{\sigma}(\sigma^2\frac{\varepsilon}{\Delta f} - \frac{\Delta f}{2})) > 0$ provided $c \geq 3/2$. We can therefore focus on the t^2/σ^2 term.

$$\begin{aligned}\left(\frac{1}{2\sigma^2}\frac{\sigma^2\varepsilon}{\Delta f} - \frac{\Delta f}{2}\right)^2 &= \frac{1}{2\sigma^2}\left[\Delta f\left(\frac{c^2}{\varepsilon} - \frac{1}{2}\right)\right]^2 \\ &= \left[(\Delta f)^2\left(\frac{c^2}{\varepsilon} - \frac{1}{2}\right)\right]^2\left[\frac{\varepsilon^2}{c^2(\Delta f)^2}\right]\frac{1}{2} \\ &= \frac{1}{2}\left(\frac{c^2}{\varepsilon} - \frac{1}{2}\right)^2\frac{\varepsilon^2}{c^2} \\ &= \frac{1}{2}(c^2 - \varepsilon + \varepsilon^2/4c^2).\end{aligned}$$

Since $\varepsilon \leq 1$ the derivative of $(c^2 - \varepsilon + \varepsilon^2/4c^2)$ with respect to c is positive in the range we are considering ($c \geq 3/2$), so $c^2 - \varepsilon + \varepsilon^2/4c^2 \geq c^2 - 8/9$ and it suffices to ensure

$$c^2 - 8/9 > 2\ln\left(\sqrt{\frac{2}{\pi}}\frac{1}{\delta}\right).$$

In other words, we need that

$$c^2 > 2\ln(\sqrt{2/\pi}) + 2\ln(1/\delta) + \ln(e^{8/9}) = \ln(2/\pi) + \ln(e^{8/9}) + 2\ln(1/\delta),$$

which, since $(2/\pi)e^{8/9} < 1.55$, is satisfied whenever $c^2 > 2\ln(1.25/\delta)$.

Let us parition $\mathbb{R}$ as $\mathbb{R} = R_1 \cup R_2$, where $R_1 = \{x \in \mathbb{R} : |x| \leq c\Delta f/\varepsilon\}$ and $R_2 = \{x \in \mathbb{R} : |x| > c\Delta f/\varepsilon\}$. Fix any subset $S \subseteq \mathbb{R}$, and define

$$S_1 = \{f(x) + x \,|\, x \in R_1\}$$
$$S_2 = \{f(x) + x \,|\, x \in R_2\}.$$

We have

$$\begin{aligned}
\Pr_{x\sim\mathcal{N}(0,\sigma^2)}[f(x) + x \in S] &= \Pr_{x\sim\mathcal{N}(0,\sigma^2)}[f(x) + x \in S_1] \\
&\quad + \Pr_{x\sim\mathcal{N}(0,\sigma^2)}[f(x) + x \in S_2] \\
&\leq \Pr_{x\sim\mathcal{N}(0,\sigma^2)}[f(x) + x \in S_1] + \delta \\
&\leq e^{\varepsilon}\left(\Pr_{x\sim\mathcal{N}(0,\sigma^2)}[f(y) + x \in S_1]\right) + \delta,
\end{aligned}$$

yielding (ε, δ)-differential privacy for the Gaussian mechanism in one dimension.

High Dimension. To extend this to functions in R^m, define $\Delta f = \Delta_2 f$. We can now repeat the argument, using Euclidean norms. Let v be any vector satisfying $\|v\| \leq \Delta f$. For a fixed pair of databases x, y we are interested in $v = f(x) - f(y)$, since this is what our noise must obscure. As in the one dimensional, case we seek conditions on σ under which the privacy loss

$$\left|\ln \frac{e^{(-1/2\sigma^2)\|x-\mu\|^2}}{e^{(-1/2\sigma^2)\|x+v-\mu\|^2}}\right|$$

is bounded by ε; here x is chosen from $\mathcal{N}(0, \Sigma)$, where (Σ) is a diagonal matrix with entries σ^2, whence $\mu = (0, \ldots, 0)$.

$$\left| \ln \frac{e^{(-1/2\sigma^2)\|x-\mu\|^2}}{e^{(-1/2\sigma^2)\|x+v-\mu\|^2}} \right| = |\ln e^{(-1/2\sigma^2)[\|x-\mu\|^2 - \|x+v-\mu\|^2]}|$$
$$= \left| \frac{1}{2\sigma^2}(\|x\|^2 - \|x+v\|^2)) \right|.$$

We will use the fact that the distribution of a spherically symmetric normal is independent of the orthogonal basis from which its constituent normals are drawn, so we may work in a basis that is alligned with v. Fix such a basis $b_1, \ldots, b_m$, and draw x by first drawing signed lengths $\lambda_i \sim \mathcal{N}(0, \sigma^2)$, for $i \in [m]$, then defining $x^{[i]} = \lambda_i b_i$, and finally letting $x = \sum_{i=1}^m x^{[i]}$. Assume without loss of generality that b_1 is parallel to v. We are interested in $|\, \|x\|^2 - \|x+v\|^2 \,|$.

Consider the right triangle with base $v + x^{[1]}$ and edge $\sum_{i=2}^m x^{[i]}$ orthogonal to v. The hypotenuse of this triangle is $x + v$.

$$\|x+v\|^2 = \|v + x^{[1]}\|^2 + \sum_{i=2}^m \|x^{[i]}\|^2$$
$$\|x\|^2 = \sum_{i=1}^m \|x^{[i]}\|^2.$$

Since v is parallel to $x^{[1]}$ we have $\|v + x^{[1]}\|^2 = (\|v\| + \lambda_1)^2$. Thus, $\|x+v\|^2 - \|x\|^2 = \|v\|^2 + 2\lambda_1 \cdot \|v\|$. Recall that $\|v\| \leq \Delta f$, and $\lambda \sim \mathcal{N}(0, \sigma)$, so we are now exactly back in the one-dimensional case, writing λ_1 instead of x in Equation (A.2):

$$\left| \frac{1}{2\sigma^2}(\|x\|^2 - \|x+v\|^2)) \right| \leq \left| \frac{1}{2\sigma^2}(2\lambda_1 \Delta f - (\Delta f)^2) \right|$$

and the rest of the argument proceeds as above. □

The argument for the high dimensional case highlights a weakness of (ε, δ)-differential privacy that does not exist for $(\varepsilon, 0)$-differential privacy. Fix a database x. In the $(\varepsilon, 0)$-case, the guarantee of indistinguishability holds for all adjacent databases *simultaneously*. In the

(ε, δ) case indistinguishability only holds "prospectively," i.e., for any fixed y adjacent to x, the probability that the mechanism will allow the adversary to distinguish x from y is small. In the proof above, this is manifested by the fact that we fixed $v = f(x) - f(y)$; we did not have to argue about all possible directions of v simultaneously, and indeed we cannot, as once we have fixed our noise vector $x \sim \mathcal{N}(0, \Sigma)$, so that the output on x is $o = f(x) + x$, there may exist an adjacent y such that output $o = f(x) + x$ is much more likely when the database is y than it is on x.

A.1 Bibliographic notes

Theorem A.1 is folklore initially observed by the authors of [23]. A generalization to non-spherical gaussian noise appears in [66].

B

Composition Theorems for (ε, δ)-DP

B.1 Extension of Theorem 3.16

Theorem B.1. Let $T_1(D) : D \mapsto T_1(D) \in \mathcal{C}_1$ be an (ϵ, δ)-d.p. function, and for any $s_1 \in \mathcal{C}_1$, $T_2(D, s_1) : (D, s_1) \mapsto T_2(D, s_1) \in \mathcal{C}_2$ be an (ϵ, δ)-d.p. function given the second input s_1. Then we show that for any neighboring D, D', for any $S \subseteq \mathcal{C}_2 \times \mathcal{C}_1$, we have, using the notation in our paper

$$P((T_2, T_1) \in S) \leq e^{2\epsilon} P'((T_2, T_1) \in S) + 2\delta. \tag{B.1}$$

Proof. For any $C_1 \subseteq \mathcal{C}_1$, define

$$\mu(C_1) = \left(P(T_1 \in C_1) - e^{\epsilon} P'(T_1 \in C_1)\right)_+,$$

then μ is a measure on $\mathcal{C}_1$ and $\mu(\mathcal{C}_1) \leq \delta$ since T_1 is (ϵ, δ)-d.p. As a result, we have for all $s_1 \in \mathcal{C}_1$,

$$P(T_1 \in ds_1) \leq e^{\epsilon} P'(T_1 \in ds_1) + \mu(ds_1). \tag{B.2}$$

Also note that by the definition of (ϵ, δ)-d.p., for any $s_1 \in \mathcal{C}_1$,

$$\begin{aligned} P((T_2, s_1) \in S) &\leq \left(e^{\epsilon} P'((T_2, s_1) \in S) + \delta\right) \wedge 1 \\ &\leq \left(e^{\epsilon} P'((T_2, s_1) \in S)\right) \wedge 1 + \delta. \end{aligned} \tag{B.3}$$

Then (B.2) and (B.3) give (B.1):

$$\begin{aligned}
P((T_2, T_1) \in S) &\leq \int_{S_1} P((T_2, s_1) \in S) P(T_1 \in ds_1) \\
&\leq \int_{S_1} \left((e^{\epsilon} P'((T_2, s_1) \in S)) \wedge 1 + \delta \right) P(T_1 \in ds_1) \\
&\leq \int_{S_1} \left((e^{\epsilon} P'((T_2, s_1) \in S)) \wedge 1 \right) P(T_1 \in ds_1) + \delta \\
&\leq \int_{S_1} \left((e^{\epsilon} P'((T_2, s_1) \in S)) \wedge 1 \right) \\
&\quad \times \left(e^{\epsilon} P'(T_1 \in ds_1) + \mu(ds_1) \right) + \delta \\
&\leq e^{2\epsilon} \int_{S_1} P'((T_2, s_1) \in S) P'(T_1 \in ds_1) + \mu(S_1) + \delta \\
&\leq e^{2\epsilon} P'((T_2, T_1) \in S) + 2\delta. \qquad \text{(B.4)}
\end{aligned}$$

In the equations above, S_1 denotes the projection of S onto $\mathcal{C}_1$. The event $\{(T_2, s_1) \in S\}$ refers to $\{(T_2(D, s_1), s_1) \in S\}$ (or $\{(T_2(D', s_1), s_1) \in S\}$). □

Using induction, we have:

Corollary B.2 (general composition theorem for (ϵ, δ)-d.p. algorithms)**.** Let $T_1 : D \mapsto T_1(D)$ be (ϵ, δ)-d.p., and for $k \geq 2$, $T_k : (D, s_1, \ldots, s_{k-1}) \mapsto T_k(D, s_1, \ldots, s_{k-1}) \in \mathcal{C}_k$ be (ϵ, δ)-d.p., for all given $(s_{k-1}, \ldots, s_1) \in \bigotimes_{j=1}^{k-1} \mathcal{C}_j$. Then for all neighboring D, D' and all $S \subseteq \bigotimes_{j=1}^{k} \mathcal{C}_j$

$$P((T_1, \ldots, T_k) \in S) \leq e^{k\epsilon} P'((T_1, \ldots, T_k) \in S) + k\delta.$$