Differentially private random projections

Stephen Tu

1 Background

We extend the work of releasing differentially private random projections in Kenthapadi et al. [2] to handle row level user privacy instead of attribute level privacy. We use their same notation to keep the derivations consistent.

2 Derivation

For this section, let \(\|\cdot\|\) denote the \(L_2\) norm, let \(\|\cdot\|_F\) denote the matrix Frobenius norm, and let \(\langle \cdot, \cdot \rangle\) denote the usual Euclidean inner product.

Let \(P\) be a random \(d \times k\) Gaussian matrix with each entry drawn independently from \(\mathcal{N}(0, \sigma^2_p)\). Let \(X\) and \(X'\) be two \(n \times d\) matrices of user data, such that \(X\) and \(X'\) only differ in one row \(i\), and \(\|X_i - X'_i\| \leq B\).

2.1 Directly bounding the output

Lemma 1. With probability at least \(1 - \delta\), we have \(\|XP - X'P\|_F \leq B\sigma_p \sqrt{k + 2\sqrt{k \log(1/\delta)} + 2 \log(1/\delta)}\).

Proof. Since \(X\) and \(X'\) only differ in row \(i\), we have \((XP - X'P)_{mn} = 0\) for \(m \neq i\), and that
\[
(XP - X'P)_{ij} = \langle X_i, P_j \rangle - \langle X'_i, P_j \rangle = \langle X_i - X'_i, P_j \rangle
\]
where \(P_j\) is the \(j\)-th column of \(P\). Let \(z = X_i - X'_i\). Then by the scaling properties of Gaussians (e.g. if \(a, b\) are constants, \(X \sim \mathcal{N}(0, \sigma_x^2)\), and \(Y \sim \mathcal{N}(0, \sigma_y^2)\), then \(aX + bY \sim \mathcal{N}(0, a^2\sigma_x^2 + b^2\sigma_y^2)\)), we know \(\langle z, P_j \rangle \sim \mathcal{N}(0, \|z\|^2 \sigma_p^2)\). Let \(Y_j \sim \mathcal{N}(0, 1)\) and \(\chi^2_k\) denote a random variable drawn from a chi-squared distribution with \(k\) degrees of freedom. We now bound the matrix norm as follows
\[
\|XP - X'P\|_F = \sqrt{\sum_{j=1}^{k} \langle z, P_j \rangle^2} = \sqrt{\sum_{j=1}^{k} (\|z\| \sigma_p Y_j)^2} = \|z\| \sigma_p \sqrt{\chi^2_k}
\]
where the second equality follows since if \(X \sim \mathcal{N}(0, \sigma^2)\), then \(X/\sigma \sim \mathcal{N}(0, 1)\). From Laurent and Massart (Lemma 1, [3]), we have the following tail bound on a random variable \(X\) drawn from a \(k\) degrees of freedom chi-squared distribution
\[
\Pr[X \geq k + 2\sqrt{kx} + 2x] \leq \exp(-x)
\]
The claim now follows by setting \(x = \log(1/\delta)\). □

Let \(f : D^n \to \mathbb{R}^d\) be a function with \(L_2\) sensitivity bounded by \(\Delta_2(f)\). Then from Kenthapadi et al. [2], we have the following differentially private mechanism construction using Gaussian noise

Lemma 2. (Lemma 1, [2]) The mechanism given by \(M(D) = f(D) + G\), where \(G\) is a random Gaussian vector with entries drawn from \(\mathcal{N}(0, 2\Delta_2(f)^2(\log(1/2\delta) + \epsilon)/\epsilon^2)\) satisfies \((\epsilon, \delta)\)-differential privacy provided \(\delta < \frac{1}{2}\).

By combining Lemma 1 and Lemma 2, we have the following \((\epsilon, \delta)\)-differentially private algorithm for releasing randomized projections
Theorem 1. Let $\epsilon > 0$ and $0 < \delta < 1/2$. Fix a randomized gaussian projection matrix $P$. Then the mechanism $M_P(X) = XP + G$, where $G$ is an $n \times k$ random gaussian matrix with entries drawn from $\mathcal{N}(0, \sigma^2)$ with

$$
\sigma = B\sigma_p \sqrt{k + 2k \log(2/\delta) + 2 \log(2) \sqrt{2\log(1/\delta) + \epsilon}}/\epsilon
$$

is $(\epsilon, \delta)$-differentially private.

Proof. The claim follows immediately by invoking both Lemma 1 and Lemma 2 with $\delta/2$. □

2.2 Composition approach

We can derive another algorithm by utilizing the following composition theorem from Dwork et al. [1]

Lemma 3. (Theorem S.3, [1]) Suppose we have $k$ mechanisms which are each $(\epsilon, \delta)$-differentially private. Let $\delta' > 0$. Then the composition of the $k$ mechanisms is $(\epsilon', k\delta + \delta')$-differentially private for

$$
e' = \sqrt{2k \log(1/\delta')} + k\epsilon(\exp(\epsilon) - 1)
$$

Lemma 4. Let $H = XP - X' P$, and $H_i$ denote the $i$-th column of $H$. Then we have $\Pr[\|H_i\| > B\sigma_p \sqrt{2 \log(1/\delta)}] \leq \delta$ for all $i$.

Proof. From Lemma 1, we know that $\|H_i\| \sim (z, P_i) \sim \mathcal{N}(0, \|z\|^2 \sigma_i^2)$. Standard Gaussian tail bounds tell us that if $X \sim \mathcal{N}(0, \sigma^2)$, then $\Pr[\|X\| > \sigma \sqrt{2 \log(1/\delta)}] \leq \delta$. Plugging $\|H_i\|$ into the bound yields the claim. □

Theorem 2. Let $0 < \epsilon < 1$ and $0 < \delta < 1/2$. Fix a randomized gaussian projection matrix $P$. Then the mechanism $M_P(X) = XP + G$, where $G$ is an $n \times k$ random gaussian matrix with entries drawn from $\mathcal{N}(0, \sigma^2)$ with

$$
\sigma = \frac{B\sigma_p}{\epsilon_1} \sqrt{4 \log^2(k/\delta) + 2\epsilon_1 \log(2k/\delta)}
$$

where

$$
\epsilon_1 = \frac{\sqrt{2k \log(2/\delta) + 8k\epsilon - \sqrt{2k \log(2/\delta)}}}{4k}
$$

is $(\epsilon, \delta)$-differentially private.

Proof. By invoking Lemma 4 and Lemma 2, the mechanism $M_{P,i}(X) = (XP)_i + G$, where $(XP)_i$ denotes the $i$-th column of $XP$ and $G$ is an $n \times 1$ random gaussian vector with entries drawn from $\mathcal{N}(0, \sigma_i^2)$ with $\sigma_i = \frac{B\sigma_p}{\epsilon_1} \sqrt{4 \log^2(1/2\delta_1) + 2\epsilon_1 \log(1/\delta_1)}$ is $(\epsilon_1, \delta_1)$-differentially private. Also, by setting $\delta_1 = \delta/2k$ and $\delta' = \delta/2$ and invoking Lemma 3, we have that the composition is $(\epsilon', \delta)$-differentially private with $\epsilon' = \sqrt{2k \log(2/\delta)}\epsilon_1 + k\epsilon_1(\exp(\epsilon_1) - 1)$. Noting that $\exp(\epsilon_1) \leq 1 + 2\epsilon_1$ if $0 < \epsilon_1 < 1$, then solving for $\epsilon_1$ using the quadratic formula yields the result. □

References

