Differentially Private Multi-Party Data Release for Linear Regression (Supplementary Material)

Ruihan Wu[*1] **Xin Yang**[2] **Yuanshun Yao**[2] **Jiankai Sun**[2] **Tianyi Liu**[2] **Kilian Q. Weinberger**[1] **Chong Wang**[2]

[1]Cornell University, USA
[2]ByteDance Inc., USA
[*]Work done as an intern at ByteDance Inc.

A PROOFS OF USEFUL LEMMAS

Lemma 1 (Gaussian mechanism). *For any deterministic real-valued function $f : \mathcal{D} \to \mathbb{R}^m$ with sensitivity S_f, we can define a randomized function by adding Gaussian noise to f:*

$$f^{dp}(D) := f(D) + \mathcal{N}\left(\mathbf{0}, S_f^2\sigma^2 \cdot I\right),$$

where $\mathcal{N}\left(\mathbf{0}, S_f^2\sigma^2 \cdot I\right)$ is a multivariate normal distribution with mean $\mathbf{0}$ and co-variance matrix $S_f^2\sigma^2$ multiplying a $m \times m$ identity matrix I. When $\sigma \geq \frac{\sqrt{2\log(1/(1.25\delta))}}{\varepsilon}$, f^{dp} is (ε, δ)-differentially private.

Lemma 2 (JL Lemma for inner-product preserving (Bernoulli)). *Suppose S be an arbitrary set of l points in $\mathbb{R}^d$ and suppose s is an upper bound for the maximum L2-norm for vectors in S. Let B be a $k \times d$ random matrix, where B_{ij} are independent random variables, which take value 1 and value -1 with probability $1/2$. With the probability at least $1 - (l+1)^2 \exp\left(-k\left(\frac{\beta^2}{4} - \frac{\beta^3}{6}\right)\right)$,*

$$\frac{\mathbf{u}^\top \mathbf{v}}{s^2} - 4\beta \leq \frac{\left(B\mathbf{u}/\sqrt{k}\right)^\top \left(B\mathbf{v}/\sqrt{k}\right)}{s^2} \leq \frac{\mathbf{u}^\top \mathbf{v}}{s^2} + 4\beta.$$

Lemma 3. *1. $\forall x \in [0,1]$, $-\log(1-x) - x \geq \frac{x^2}{2}$.*

2. $\forall x \in [0,1]$, $x - \log(1+x) \geq \frac{x^2}{4}$.

3. $\forall x > 1, x - \log(1+x) \geq \frac{x}{2}$.

Proof. Define $f_1(x) := -\log(1-x) - x - \frac{x^2}{2}$. $f_1'(x) = \frac{x^2}{1-x} \geq 0$. Thus $f_1(x)$ increases on $[0,1]$ and $f_1(x) \geq f_1(0) = 0$.

Define $f_2(x) := x - \log(1+x) - \frac{x^2}{4}$. $f_2'(x) = \frac{x(1-x)}{2(1+x)}$. $f_2(x)$ increases on $[0,1]$ and $f_2(x) \geq f_2(0) = 0$.

Define $f_3(x) := x - \log(1+x) - \frac{x}{4}$. $f_3'(x) = \frac{3x-1}{4(1+x)} > 0$. $f_3(x)$ increases on $[0,1]$ and $f_3(x) \geq f(1) > 0$. □

Lemma 4. *Denote $\hat{H}_n = \frac{1}{n}X^\top X$, $\hat{C}_n = \frac{1}{n}X^\top Y$, $H = \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}}\left[\mathbf{x}\mathbf{x}^\top\right]$ and $C = \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}}\left[\mathbf{x} \cdot y\right]$. Assume $\|\hat{H}_n^{\mathsf{pub}} - \hat{H}_n\| \leq \beta, \|\hat{C}_n^{\mathsf{pub}} - \hat{C}_n\| \leq \beta$ with prob $1 - f(\beta)$. We have that when $\beta \leq \frac{2\|C\|\|H^{-1}\|+5}{8}$,*

$$\mathbb{P}_{X,\mathbf{y}\sim\mathcal{D},R_1,R_2}\left[\|\hat{\mathbf{w}}_n^{\mathsf{pub}} - \mathbf{w}^*\| \leq \beta\right] \geq 1 - h(\beta),$$

where $\hat{\mathbf{w}}_n^{\mathsf{pub}} = \left(\hat{H}_n^{\mathsf{pub}}\right)^{-1}\hat{C}_n^{\mathsf{pub}}$, $c := \|C\|\|H^{-1}\|^2 + 2\|H^{-1}\|$ and $h(\beta) = f(\beta/2c) + d^2\exp\left(-\frac{n\beta^2}{8c^2d^2}\right) + d\exp\left(-\frac{n\beta^2}{8c^2 d}\right)$.

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

Proof. Hoeffding inequality and union bound together imply that with prob. $1 - d^2 \exp\left(-\frac{n\beta^2}{2d^2}\right) - d \exp\left(-\frac{n\beta^2}{2d}\right)$,

$$\|\hat{H}_n - H\| \le \beta, \|\hat{C}_n - C\| \le \beta.$$

Thus with prob $1 - g(\delta)$, $\|\hat{H}_n^{\mathsf{pub}} - H\| \le \beta, \|\hat{C}_n^{\mathsf{pub}} - C\| \le \beta$, where $g(\beta) = f(\beta/2) + d^2 \exp\left(-\frac{n\beta^2}{8d^2}\right) + d \exp\left(-\frac{n\beta^2}{8d}\right)$

We further have

- $\|\hat{C}_n^{\mathsf{pub}}\| \le \|C - \hat{C}_n\| + \|C\| \le \|C\| + \beta$
- $\|\left(\hat{H}_n^{\mathsf{pub}}\right)^{-1} - H^{-1}\| \le \|\left(\hat{H}_n^{\mathsf{pub}}\right)^{-1}\| \|H^{-1}\| \cdot \|\hat{H}_n^{\mathsf{pub}} - H\| \le \left(\|H^{-1}\| + \|\left(\hat{H}_n^{\mathsf{pub}}\right)^{-1} - H^{-1}\|\right) \cdot \|H^{-1}\| \cdot \beta$, which implies that when $\beta \le \frac{1}{2\|H^{-1}\|}$, $\|\left(\hat{H}_n^{\mathsf{pub}}\right)^{-1} - H^{-1}\| \le \frac{\|H^{-1}\|^2 \cdot \beta}{1 - \|H^{-1}\| \cdot \beta} \le \frac{\|H^{-1}\|^2 \cdot \beta}{2}$.
- When $\beta \le \frac{1}{2\|H^{-1}\|}$,

$$\begin{aligned}\|\left(\hat{H}_n^{\mathsf{pub}}\right)^{-1} \hat{C}_n^{\mathsf{pub}} - H^{-1}C\| &\le \|\hat{C}_n^{\mathsf{pub}}\| \cdot \|\left(\hat{H}_n^{\mathsf{pub}}\right)^{-1} - H^{-1}\| + \|H^{-1}\| \cdot \|\hat{C}_n^{\mathsf{pub}} - C\| \\ &\le (\|C\| + \beta) \cdot \frac{\|H^{-1}\|^2 \cdot \beta}{2} + \|H^{-1}\| \cdot \beta \\ &\le \frac{(2\|C\|\|H^{-1}\|^2 + 5\|H^{-1}\|)}{4}\beta.\end{aligned}$$

Let $b := \frac{(2\|C\|\|H^{-1}\|^2 + 5\|H^{-1}\|)}{4}$ and replace β by $b^{-1}\beta$, we have that when $\beta \le \frac{2\|C\|\|H^{-1}\| + 5}{8}$

$$\mathbb{P}_{X,\mathbf{y}\sim\mathcal{D},R_1,R_2}\left[\|\hat{\mathbf{w}}_n - \mathbf{w}^*\| \le \beta\right] \ge 1 - h(\beta),$$

where $h(\beta) = g(\beta/b) = f(\beta/2b) + d^2 \exp\left(-\frac{n\beta^2}{8b^2d^2}\right) + d \exp\left(-\frac{n\beta^2}{8b^2d}\right)$. ☐

Lemma 5. *If r is a random variable sampled from standard normal distribution, we have following concentration bound:*

$$\mathbb{P}\left[|r| < \beta\right] \ge 1 - \frac{2}{\sqrt{2\pi}\beta} \exp\left(-\frac{\beta^2}{2}\right)$$

Proof. It's shown in page 2 in Pollard [2015]. ☐

Lemma 6. *If r_1, r_2 are two independent random variables sampled from standard normal distribution, $r_1 r_2$ can be written as $\frac{c_1 - c_2}{2}$, where c_1, c_2 are independent two random variables sampled from chi-squared with degree 1. Moreover, $\sum_{i=1}^n r_{1,n} r_{2,n}$ can be written as $\frac{c_{1,1:n} - c_{2,1:n}}{2}$, where $c_{1,1:n}, c_{2,1:n}$ are independent two random variables sampled from chi-squared with degree n.*

Proof. $r_1 r_2 = \frac{\left(\frac{r_1 + r_2}{\sqrt{2}}\right)^2 - \left(\frac{r_1 - r_2}{\sqrt{2}}\right)^2}{2}$. Because r_1, r_2 are two independent standard normal random variables, $\frac{r_1 + r_2}{\sqrt{2}}$, $\frac{r_1 - r_2}{\sqrt{2}}$ are two independent standard normal random variables as well. $c_1 := \frac{r_1 + r_2}{\sqrt{2}}$ and $c_2 := \frac{r_1 - r_2}{\sqrt{2}}$ complete the proof for the first part.

$\sum_{i=1}^n r_{1,n} r_{2,n} = \frac{1}{2} \sum_{i=1}^n (c_{1,i} - c_{2,i}) = \frac{1}{2}(\sum_{i=1}^n c_{1,i} - \sum_{i=1}^n c_{2,i})$. $c_{1,1:n} := \sum_{i=1}^n c_{1,i}$ and $c_{2,1:n} := \sum_{i=1}^n c_{2,i}$ finish the proof. ☐

B PROOFS IN SECTION 4

We restate the assumptions and theorems for the completeness.

Assumption 1. *D_i, $i = 1, \cdots, n$, are i.i.d sampled from an underlying distribution $\mathcal{P}$ over $\mathbb{R}^{d+1}$.*

Assumption 2. *The absolute values of all attributes are bounded by 1.*

Assumption 3. *$\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}}\left[\mathbf{x}\mathbf{x}^\top\right]$ is positive definite.*

Theorem 1. *When $\beta \le c$ for some variable c that depends on $\sigma_{\varepsilon,\delta}$, d and $\mathcal{P}$, but independent of n,*

$$\mathbb{P}\left[\|\hat{\mathbf{w}}_n^{\mathsf{dgm}} - \mathbf{w}^*\| > \beta\right] < 1 - \exp\left(-O\left(\beta^2\frac{n}{\sigma_{\varepsilon,\delta}^4 d^4}\right) + \tilde{O}(1)\right).$$

Proof of Theorem 1. Denote $\left(\max_{j\in[m]} d_j\right)$ by $d_{\max}$. Denote $R \in \mathbb{R}^{n\times d}$ is a random matrix s.t. $R_{i,j} \sim \mathcal{N}\left(0, 4d_{\max}\sigma_{\varepsilon,\delta}^2\right)$. We split R into R_X and R_Y representing the addictive noise to X and Y.

$$\hat{\mathbf{w}}_n^{\mathsf{dgm}} = \left(\frac{1}{n}(X+R_X)^\top(X+R_X) + (\lambda - 4d_{\max}\sigma_{\varepsilon,\delta}^2)I\right)^{-1}\frac{(X+R_X)^\top(Y+R_Y)}{n}.$$

1. For any $i \in [d]$, $\frac{1}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\left[R_X^\top R_X\right]_{i,i}$ is sampled from chi-square distribution with degree n. From the cdf of chi-square distribution, we have following concentration:

$$\mathbb{P}\left[\left|\left[\frac{1}{n}R_X^\top R_X\right]_{i,i} - 4d_{\max}\sigma_{\varepsilon,\delta}^2\right| < \beta\right] \ge 1 - \exp\left(-n\cdot\left(\frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2} - \log\left(1+\frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right)\right) - \exp\left(-n\cdot\left(-\log\left(1-\frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right) - \frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right).$$

Moreover, for $i \ne j$, Lemma 6 implies that $\frac{1}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\left[R_X^\top R_X\right]_{i,j}$ can be written as $\frac{c_{1,1:n} - c_{2,1:n}}{2}$, where $c_{1,1:n}, c_{2,1:n}$ are independent two random variables sampled from chi-squared with degree n. Thus

$$\begin{aligned}
\mathbb{P}\left[\left|\left[\frac{1}{n}R_X^\top R_X\right]_{i,j}\right| < \beta\right] &= \mathbb{P}\left[\left|\frac{c_{1,1:n} - c_{2,1:n}}{2n}\right| < \frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right] \\
&\ge \mathbb{P}\left[|c_{1,1:n} - n| < \frac{n\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}, |c_{1,1:n} - n| < \frac{n\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right] \\
&\ge 1 - 2\mathbb{P}\left[|c_{1,1:n} - n| \ge \frac{n\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right] \\
&\ge 1 - 2\exp\left(-n\cdot\left(\frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2} - \log\left(1+\frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right)\right) \\
&\quad - 2\exp\left(-n\cdot\left(-\log\left(1-\frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right) - \frac{\beta}{4d_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right)
\end{aligned}$$

Union bound implies that

$$\mathbb{P}\left[\left\|\frac{1}{n}R_X^\top R_X - 4d_{\max}\sigma_{\varepsilon,\delta}^2\cdot I\right\| \le \beta_1\right] \ge 1 - d^2\cdot\exp\left(-n\cdot\left(\frac{\beta_1}{4dd_{\max}\sigma_{\varepsilon,\delta}^2} - \log\left(1+\frac{\beta_1}{4dd_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right)\right) - d^2\cdot\exp\left(-n\cdot\left(-\log\left(1-\frac{\beta_1}{4dd_{\max}\sigma_{\varepsilon,\delta}^2}\right) - \frac{\beta_1}{4dd_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right)$$

2. $\mathbb{P}\left[\left\|\frac{X^\top R_X}{n}\right\| \le \beta_2\right] \ge 1 - \frac{4\sigma_{\varepsilon,\delta}d^3 d_{\max}^{1/2}}{\sqrt{2\pi n}\beta_2}\exp\left(-\frac{n\beta_2^2}{8d^2 d_{\max}\sigma_{\varepsilon,\delta}^2}\right)$, implied by Lemma 5.
3. $\mathbb{P}\left[\left\|\frac{X^\top R_Y}{n}\right\| \le \beta_3\right] \ge 1 - \frac{4\sigma_{\varepsilon,\delta}d^{3/2} d_{\max}^{1/2}}{\sqrt{2\pi n}\beta_3}\exp\left(-\frac{n\beta_3^2}{8dd_{\max}\sigma_{\varepsilon,\delta}^2}\right)$, implied by Lemma 5.
4. $\mathbb{P}\left[\left\|\frac{R_X^\top Y}{n}\right\| \le \beta_4\right] \ge 1 - \frac{4\sigma_{\varepsilon,\delta}d^{3/2} d_{\max}^{1/2}}{\sqrt{2\pi n}\beta_4}\exp\left(-\frac{n\beta_4^2}{8dd_{\max}\sigma_{\varepsilon,\delta}^2}\right)$, implied by Lemma 5.

5. Similar to 1,

$$\mathbb{P}\left[\left\|\frac{R_X^\top R_Y}{n}\right\| \le \beta_5\right] \ge 1 - 2d\exp\left(-n\cdot\left(\frac{\beta_5}{4d^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2} - \log\left(1+\frac{\beta_5}{4d^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right)\right)$$
$$- 2d\exp\left(-n\cdot\left(-\log\left(1-\frac{\beta_5}{4d^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}\right) - \frac{\beta_5}{4d^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}\right)\right)$$

One can simplify $-\log(1-x) - x$ and $x - \log(1+x)$ by Lemma 3. Set $\beta_1 = \frac{1}{2}\beta,\ \beta_2 = \frac{1}{4}\beta,\ \beta_3 = \beta_4 = \frac{1}{4}\beta,\ \beta_5 = \frac{1}{2}\beta$. The above concentrations together imply that when $\beta < 8dd_{\max}\sigma_{\varepsilon,\delta}^2$, $\|\hat{H}_n^{\mathsf{pub}} - \hat{H}_n\| \le \beta$, $\|\hat{C}_n^{\mathsf{pub}} - \hat{C}_n\| \le \beta$ with prob at least $1 - f(\beta)$, where $f(\beta) = \exp\left(-\min\left\{O\left(n\cdot\frac{\beta^2}{d^2 d_{\max}^2\sigma_{\varepsilon,\delta}^4}\right) + \tilde{O}(1)\right\}\right)$.

With the application of Lemma 4: when $\beta \le \frac{2\|C\|\|H^{-1}\|+5}{8}$

$$\mathbb{P}_{X,\mathbf{y}\sim\mathcal{D},R_1,R_2}\left[\|\hat{\mathbf{w}}_n - \mathbf{w}^*\| \le \beta\right] \ge 1 - h(\beta),$$

where $h(\beta)$ is:

1. when $\beta < 16bdd_{\max}\sigma_{\varepsilon,\delta}^2$, $h(\beta) = \exp\left(-O\left(n\cdot\frac{\beta^2}{d^2 d_{\max}^2\sigma_{\varepsilon,\delta}^4}\right) + \tilde{O}(1)\right)$;
2. when $\beta \ge 16bdd_{\max}\sigma_{\varepsilon,\delta}^2$, $h(\beta) = \exp\left(-O\left(n\cdot\frac{\beta^2}{d^2 d_{\max}\sigma_{\varepsilon,\delta}^2}\right) + \tilde{O}(1)\right)$;

where $b = \frac{(2\|C\|\|H^{-1}\|^2 + 5\|H^{-1}\|)}{4}$ is a distribution dependent constant. In the other word, when $\beta \le \min\left\{16bdd_{\max}\sigma_{\varepsilon,\delta}^2, \frac{2\|C\|\|H^{-1}\|+5}{8}\right\}$,

$$\mathbb{P}_{X,\mathbf{y}\sim\mathcal{D},R_1,R_2}\left[\|\hat{\mathbf{w}}_n - \mathbf{w}^*\| \le \beta\right] \ge 1 - \exp\left(-O\left(n\cdot\frac{\beta^2}{d^2 d_{\max}^2\sigma_{\varepsilon,\delta}^4}\right) + \tilde{O}(1)\right),$$

□

Theorem 2. *When $\beta \le c$ for some variable c that depends on d and $\mathcal{P}$, but independent of n and $\sigma_{\varepsilon,\delta}$, $\mathbb{P}\left[\|\mathbf{w}_n^{\mathsf{rmgm}} - \mathbf{w}^*\| > \beta\right] < \exp\left(-O\left(\min\left\{\frac{k\beta^2}{d^2}, \frac{n\beta}{kd^2\sigma_{\varepsilon,\delta}^2}, \frac{n^{1/2}\beta}{d^{3/2}\sigma_{\varepsilon,\delta}}\right\}\right) + \tilde{O}(1)\right)$. If we take $k = O\left(\frac{(nd)^{1/2}}{d_{\max}^{1/2}\sigma_{\varepsilon,\delta}}\right)$, $\mathbb{P}\left[\|\hat{\mathbf{w}}_n^{\mathsf{rmgm}} - \mathbf{w}^*\| > \beta\right] < \exp\left(-\frac{n^{1/2}\beta}{d^{3/2}d_{\max}^{1/2}\sigma_{\varepsilon,\delta}}\cdot O(\min\{1,\beta\}) + \tilde{O}(1)\right)$.*

Proof of Theorem 2.

$$\hat{\mathbf{w}}_n^{\mathsf{rmgm}} = \left(\frac{1}{n}\left(X^\top\frac{B^\top B}{k}X + X^\top\frac{B^\top}{\sqrt{k}}R_X + R_X^\top\frac{B}{\sqrt{k}}X + R_X^\top R_X\right)\right)^{-1}\left(\frac{1}{n}\left(X^\top\frac{B^\top B}{k}Y + R_X^\top\frac{B}{\sqrt{k}}Y + X^\top\frac{B^\top}{\sqrt{k}}R_Y + R_X^\top R_Y\right)\right)$$

Define $M := \frac{1}{\sqrt{k}}B$. Then we can make the analysis one by one.

1. JL-lemma applied by Bernoulli random variables implies that with probability $1-(d+2)^2\exp\left(-k\left(\frac{\beta_1^2}{64d} - \frac{\beta_1^3}{96d\sqrt{d}}\right)\right)$,

$$\left\|\frac{1}{n}X^\top M^\top MY - \frac{1}{n}X^\top Y\right\| \le \beta_1.$$

Proof. Lemma 2 implies that with prob $1-(d+2)^2\exp\left(k\left(\frac{\beta^2}{4} - \frac{\beta^3}{6}\right)\right)$, for any $u, v \in \{X_i^\top | i \in [d]\} \cup \{Y\}$,

$$\frac{(Mu)^\top Mv}{n} - 4\beta \le \frac{u^\top v}{n} \le \frac{(Mu)^\top Mv}{n} + 4\beta.$$

This further implies that

$$\left\|\frac{1}{n}X^\top M^\top MY - \frac{1}{n}X^\top Y\right\| \le 4\sqrt{d}\beta.$$

$\beta_1 = 4\sqrt{d}\beta$ helps finish the proof. ☐

2. JL-lemma applied by Bernoulli random variables implies that with probability $1-(d+2)^2\exp\left(k\left(-\frac{\beta_2^2}{64d^2}-\frac{\beta_2^3}{96d^3}\right)\right)$,

$$\left\|\frac{1}{n}X^\top M^\top MX - \frac{1}{n}X^\top X\right\| \le \beta_2.$$

3. With prob. $1-2kd\sqrt{\frac{2kdd_{\max}\sigma_{\varepsilon,\delta}^2}{\pi n\beta_3}}\exp\left(-n\frac{\beta_3}{8kdd_{\max}\sigma_{\varepsilon,\delta}^2}\right)$, $\left\|\frac{R_X^\top R_X}{n}\right\| \le \beta_3$.

   Proof. To simplify the proof, let's assume R_X is a standard gaussian matrix. Because $\mathbb{P}\left(|(R_X)_{ij}|\right) \le \beta) \ge 1 - \frac{2}{\sqrt{2\pi}\beta}\exp\left(-\beta^2/2\right)$ shown in Lemma 5,

$$\mathbb{P}\left(\|R_X^\top R_X\| \le kd\beta\right) \ge \mathbb{P}\left(\|R_X\| \le \sqrt{kd\beta}\right) \ge 1 - \frac{2kd}{\sqrt{2\pi\beta}}\exp\left(-\beta/2\right).$$

   It's equivalent that

$$\mathbb{P}\left(\left\|\frac{R_X^\top R_X}{n}\right\|_2 \le \beta_3\right) \ge 1 - 2kd\sqrt{\frac{kd}{2\pi n\beta_3}}\exp\left(-\frac{n}{2kd}\beta_3\right).$$

   Plug-in the variance of R_X leads to the targeted inequality. ☐

4. With prob. $1-2\left(\frac{1}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta_4\sqrt{nd_{\max}}}}d^{5/4}+2kd\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta_4}{4\sigma_{\varepsilon,\delta}\sqrt{dd_{\max}}}\right)$

$$\left\|\frac{R_X^\top BY}{n\sqrt{k}}\right\| \le \beta_4, \left\|\frac{X^\top B^\top R_Y}{n\sqrt{k}}\right\| \le \beta_4$$

   Proof. Denote $\mathbf{c} := \frac{R_X^\top BY}{n\sqrt{k}}$ and further $\mathbf{c}_i := \frac{\left((R_X)_i\right)^\top \mathbf{b}}{\sqrt{k}}$, where $(R_x)_i$ is the ith column for R_X and $\mathbf{b} = \frac{BY}{n}$.

$$\begin{aligned}
\mathbb{P}\left[|\mathbf{c}_i| \le \beta\right] &= \int_{\mathbf{b}} \mathbb{P}\left[|\mathbf{c}_i| \le \beta|\mathbf{b}\right]\mathbb{P}[\mathbf{b}]d\mathbf{b} \\
&\ge \max_{\alpha>0}\int_{|\mathbf{b}|\le\alpha\cdot\mathbf{1}} \mathbb{P}\left[|\mathbf{c}_i| \le \beta|\mathbf{b}\right]\mathbb{P}[\mathbf{b}]d\mathbf{b} \\
&\ge \max_{\alpha>0}\int_{|\mathbf{b}|\le\alpha\cdot\mathbf{1}} \mathbb{P}\left[|\mathbf{c}_i| \le \beta||\mathbf{b}| = \alpha\cdot\mathbf{1}\right]\mathbb{P}[\mathbf{b}]d\mathbf{b} \\
&\ge \max_{\alpha>0}\mathbb{P}\left[|\mathbf{c}_i| \le \beta|\mathbf{b}| = \alpha\cdot\mathbf{1}\right]\mathbb{P}[|\mathbf{b}| \le \alpha\cdot\mathbf{1}] \\
&\ge \max_{\alpha>0}\left(1 - \frac{4\alpha\sigma_{\varepsilon,\delta}\sqrt{d_{\max}}}{\sqrt{2\pi}\beta}\exp\left(-\frac{\beta^2}{8\alpha^2\sigma_{\varepsilon,\delta}^2 d_{\max}}\right) - 2k\exp\left(-n\cdot\frac{4\alpha^2\sigma_{\varepsilon,\delta}^2 d_{\max}}{2}\right)\right) \\
&\ge 1 - \left(\frac{1}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta\sqrt{d_{\max}n}}} + 2k\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta}{4\sqrt{d_{\max}}\sigma_{\varepsilon,\delta}}\right) \quad //\alpha^2 = \frac{\beta}{2\sqrt{nd_{\max}}\sigma_{\varepsilon,\delta}}.
\end{aligned}$$

   Then

$$\mathbb{P}\left[\|\mathbf{c}\| \le \beta\right] \ge 1 - \sum_{i=1}^{d}\mathbb{P}\left[\|\mathbf{c}_i\| > \frac{\beta}{\sqrt{d}}\right] \ge 1 - \left(\frac{1}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta\sqrt{d_{\max}n}}}d^{5/4} + 2kd\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta}{4\sigma_{\varepsilon,\delta}\sqrt{dd_{\max}}}\right).$$

   Similarly,

$$\mathbb{P}\left[\left\|\frac{X^\top B^\top R_Y}{n\sqrt{k}}\right\| \le \beta\right] \ge 1 - \left(\frac{1}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta\sqrt{nd_{\max}}}}d^{5/4} + 2kd\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta}{4\sigma_{\varepsilon,\delta}\sqrt{dd_{\max}}}\right).$$

☐

Union bound gives the conclusion.

5. With prob. $1-2\left(\frac{1}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta_5\sqrt{d_{\max}n}}}d^{5/2}+2kd^2\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta_5}{4\sigma_{\varepsilon,\delta}dd_{\max}^{1/2}}\right)$

$$\left\|\frac{R_X^\top BX}{n\sqrt{k}}\right\|\leq\beta_5,$$

which is implied similar to 4.

6. With prob. $1-2k(d+1)\sqrt{\frac{2kd^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}{\pi n\beta_6}}\exp\left(-n\frac{\beta_6}{8kd^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}\right)$, $\left\|\frac{R_X^\top R_Y}{n}\right\|\leq\beta_6$.

Proof. To simplify the proof, let's assume R_X and R_Y is a standard gaussian matrix first. Because $\mathbb{P}\left(|(R_X)_{ij}|)\leq\beta\right)\geq 1-\frac{2}{\sqrt{2\pi}\beta}\exp\left(-\beta^2/2\right)$ shown in Lemma 5,

$$\mathbb{P}\left(\|R_X^\top R_Y\|\leq k\sqrt{d}\beta\right)\geq\mathbb{P}\left(\|R_X\|\leq\sqrt{kd\beta},\|R_Y\|\leq\sqrt{k\beta}\right)\geq 1-\frac{2k(d+1)}{\sqrt{2\pi\beta}}\exp\left(-\beta/2\right).$$

It's equivalent that

$$\mathbb{P}\left(\left\|\frac{R_X^\top R_Y}{n}\right\|\leq\beta_6\right)\geq 1-2k(d+1)\sqrt{\frac{k\sqrt{d}}{2\pi n\beta_6}}\exp\left(-\frac{n}{2k\sqrt{d}}\beta_6\right).$$

Plug-in the variance of R_X and R_Y leads to the targeted inequality. ☐

Define $\hat{H}_n^{\mathsf{rmgm}} := \frac{1}{n}\left(X^\top\frac{A^\top A}{k}X+X^\top\frac{A^\top}{\sqrt{k}}R_X+R_X^\top\frac{A}{\sqrt{k}}X+\frac{1}{n}R_X^\top R_X\right)$, $\hat{H}_n := \frac{X^\top X}{n}$, $\hat{C}_n^{\mathsf{rmgm}} = \frac{1}{n}\left(X^\top\frac{A^\top A}{k}Y+R_X^\top\frac{A}{\sqrt{k}}Y+X\frac{A}{\sqrt{k}}R_Y+R_X^\top R_Y\right)$, $\hat{C}_n=\frac{1}{n}X^\top Y$.

The above analysis implies that, with prob.

$$1-(d+2)^2\exp\left(-k\left(\frac{\beta_1^2}{64d}-\frac{\beta_1^3}{96d\sqrt{d}}\right)\right)-(d+2)^2\exp\left(-k\left(\frac{\beta_2^2}{64d^2}-\frac{\beta_2^3}{96d^3}\right)\right)$$

$$-2kd\sqrt{\frac{2kdd_{\max}\sigma_{\varepsilon,\delta}^2}{\pi n\beta_3}}\exp\left(-n\frac{\beta_3}{8kdd_{\max}\sigma_{\varepsilon,\delta}^2}\right)-\left(\frac{1}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta_4\sqrt{nd_{\max}}}}d^{5/4}+2kd\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta_4}{4\sigma_{\varepsilon,\delta}\sqrt{dd_{\max}}}\right)$$

$$-\left(\frac{1}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta_5\sqrt{nd_{\max}}}}d^{5/2}+2kd^2\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta_5}{4\sigma_{\varepsilon,\delta}dd_{\max}^{1/2}}\right)-2k(d+1)\sqrt{\frac{2kd^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}{\pi n\beta_6}}\exp\left(-n\frac{\beta_6}{8kd^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}\right)$$

we have

$$\|\hat{H}_n^{\mathsf{rmgm}}-\hat{H}_n\|\leq\beta_2+\beta_3+2\beta_5,\|\hat{B}_n^{\mathsf{rmgm}}-\hat{B}_n\|\leq\beta_1+2\beta_4+\beta_6.$$

Let $\beta_2=\frac{2\beta}{3}$, $\beta_3=\frac{1}{6}$, $\beta_5=\frac{\beta}{12}$ and $\beta_1=\beta_4=\beta_6=\frac{\beta}{4}$. We will have $\|\hat{H}_n^{\mathsf{pub}}-\hat{H}_n\|\leq\beta$, $\|\hat{C}_n^{\mathsf{pub}}-\hat{C}_n\|\leq\beta$, with prob. $1-f(\beta)$, $\forall\beta\leq 4\sqrt{d}$ (implies $\beta_1\leq\sqrt{d}$ and $\beta_2\leq d$), where

$$f(\beta)=(d+2)^2\exp\left(-\frac{k\beta^2}{768d}\right)+(d+2)^2\exp\left(-\frac{k\beta^2}{432d^2}\right)$$

$$+2kd\sqrt{\frac{12kdd_{\max}\sigma_{\varepsilon,\delta}^2}{\pi n\beta}}\exp\left(-n\frac{\beta}{48kdd_{\max}\sigma_{\varepsilon,\delta}^2}\right)+\left(\frac{2}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta\sqrt{nd_{\max}}}}d^{5/4}+2kd\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta}{16\sigma_{\varepsilon,\delta}\sqrt{dd_{\max}}}\right)$$

$$+\left(\frac{2\sqrt{3}}{\sqrt{\pi\sigma_{\varepsilon,\delta}\beta\sqrt{nd_{\max}}}}d^{5/2}+2kd^2\right)\cdot\exp\left(-\sqrt{n}\cdot\frac{\beta}{48\sigma_{\varepsilon,\delta}dd_{\max}^{1/2}}\right)+2k(d+1)\sqrt{\frac{8kd^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}{\pi n\beta}}\exp\left(-n\frac{\beta}{48kd^{1/2}d_{\max}\sigma_{\varepsilon,\delta}^2}\right)$$

Lemma 4 implies that for $\beta \leq \min\{8b\sqrt{d}, \frac{2\|C\|\|H^{-1}\|+5}{8}\}$, we have

$$\mathbb{P}\left[\|\hat{\mathbf{w}}_n - \mathbf{w}^*\| \leq \beta\right] \geq 1 - h(\beta),$$

where $h(\beta)$ is:

$$h(\beta) = \exp\left(-\min\left\{O\left(\frac{k\beta^2}{d^2}\right), O\left(n\frac{\beta}{kdd_{\max}\sigma_{\varepsilon,\delta}^2}\right), O\left(n^{1/2}\frac{\beta}{dd_{\max}^{1/2}\sigma_{\varepsilon,\delta}}\right)\right\} + \tilde{O}(1)\right),$$

where $\tilde{O}(1)$ includes log terms of $n, d, d_{\max}, k, \beta$. If we take $k = O\left(\frac{(nd)^{1/2}}{d_{\max}^{1/2}\sigma_{\varepsilon,\delta}}\right)$,

$$h(\beta) = \exp\left(-\frac{n^{1/2}\beta}{d^{3/2}d_{\max}^{1/2}\sigma_{\varepsilon,\delta}} \cdot O\left(\min\left\{1, \beta\right\}\right) + \tilde{O}(1)\right).$$

□

REFERENCES

David Pollard. A few good inequalities, November 2015.