Differentially private vector aggregation in the case of multivariate gaussian data

Tim Sehested Poulsen

December 22, 2022

1 Preliminaries

1.1 Definitions

The dataset is denoted by X, where $X \in \mathbb{R}^{n \times d}$. I have that n denotes the number of entries in the dataset and d is the number of dimensions of the dataset. I will throughout the report refer to a single entry of the dataset as x_i and a single dimension of the dataset as $X^{(j)}$, and therefore $x_i^{(j)}$ denotes the j'th dimension of the i'th entry.

Differential privacy is the heuristic of releasing a database statistic whilst limiting the impact of any one entry. It builds on the intuition that computing a statistics on a private dataset should not reveal any sensitive information about any one individual as long as that individual has little to no effect on the outcome. Differential privacy has multiple slightly different formal definitions, one such is (ε, δ)-Differential Privacy refered to as (ε, δ)-DP which will be introduced later on. A prerequisite for almost all of the different differential privacy definitions relies on the concept of neighbouring dataset.

Definition 1.1 (Neighbouring dataset [4]). *Two dataset $X, X' \in \mathbb{R}^{n \times d}$ are said to be neigbouring if they differ in at most a single entry. Neighbouring dataset are denoted with the relation $X \sim X'$ and defined as followed*

$$X \sim X' \iff |\{i \in \mathbb{N} \mid i \leq n \wedge x_i \neq x'_i\}| \leq 1$$

Definition 1.2 (Sensitivity [5]). *Let $f(X) : \mathbb{R}^{n \times d} \to \mathbb{R}^k$ be a function. The l_p-sensitivity of f is the maximal possible l_p-norm of the difference between the output of f on two neighbouring dataset. We denote the sensitivity as*

$$\Delta_p(f) = \max_{X \sim X'} \|f(X) - f(X')\|_p$$

and then the total l_2-sensitivity is then

Throughout the report I will only be working with l_2-sensitivity and will just denote this as $\Delta(f)$ for ease of notation.

Definition 1.3 ((ε, δ)-Differential Privacy [4]). *A randomized algorithm $\mathcal{M} : \mathbb{R}^{n\times d} \to \mathcal{R}$ is (ε, δ)-differentially private if for all possible subsets of outputs $S \subseteq \mathcal{R}$ and all pairs of neighbouring dataset $X \sim X'$ we have that*

$$\Pr\left[M(X) \in S\right] \leq e^{\varepsilon} \cdot \Pr\left[M(X') \in S\right] + \delta$$

Theorem 1 ((ε, δ)-DP under post-processing [5]). *Let $\mathcal{M} : \mathbb{R}^{n\times d} \to \mathcal{R}$ be an (ε, δ)-DP algorithm. Let $f : \mathcal{R} \to \mathcal{R}'$ be an arbitrary mapping, then $f \circ \mathcal{M} : \mathbb{R}^{n\times d} \to \mathcal{R}'$ is (ε, δ)-DP* .

Proof

Fix any pair of neighbouring datasets $X \sim X'$ and let $S \subseteq \mathcal{R}'$ be an arbitrary event. We then define $T = \{r \in \mathcal{R} \mid f(r) \in S\}$. We thus have that

$$\begin{aligned}&\Pr\left[f(\mathcal{M}(X)) \in S\right] = \Pr\left[\mathcal{M}(X) \in T\right] \\ &\leq e^{\varepsilon} \cdot \Pr\left[\mathcal{M}(X') \in T\right] + \delta = e^{\varepsilon} \cdot \Pr\left[f(\mathcal{M}(X')) \in S\right] + \delta\end{aligned}$$

■

Error Measure As this report concerns itself exclusively with the sum of entries in a dataset, error will be defined as the expected squared l_2-norm between the true sum and the output of a randomized algorithm. So let $X \in \mathbb{R}^{n\times d}$ be the dataset and $f(X) = \sum_i^n x_i$ be the true sum of all entries. The error of a randomized algorithm $M : \mathbb{R}^{n\times d} \to \mathbb{R}^d$ which estimates $f(X)$ is then

$$\mathrm{Err}(M) := \mathbb{E}\left[\|M(X) - f(X)\|^2\right]$$

extra

If proof of elliptical ... is included then write about edp is preserved during transformation

1.2 Quadratic forms of random variables

Quadratics of random variables have been well studied [1, 9], specially in the case of multivariate gaussian varaibles [6, 9]. Even more research has been done in evaluating the CDF of these quadratic forms for Gaussian random vectors [3, 7].

Theorem 2 (Expectation of a quadratic random variable [1]). *Let X be a d-dimensional random vector with expected value $\mathbb{E}[X] = \boldsymbol{\mu}_{\boldsymbol{X}}$ and covariance matrix $Var[X] = \boldsymbol{\Sigma}_{\boldsymbol{X}}$. Let also A be a constant $d \times d$ symmetric matrix, then*

$$\mathbb{E}\left[X^T A X\right] = tr\left(A\boldsymbol{\Sigma}_{\boldsymbol{X}}\right) + \boldsymbol{\mu}^T A \boldsymbol{\mu}$$

Proof
Blah blah

$$\begin{aligned}&\mathbb{E}\left[X^T A X\right] = \mathrm{tr}\left(\mathbb{E}\left[X^T A X\right]\right) = \mathbb{E}\left[\mathrm{tr}\left(X^T A X\right)\right] \\ &= \mathbb{E}\left[\mathrm{tr}\left(A X X^T\right)\right] = \mathrm{tr}\left(A\mathbb{E}\left[X X^T\right]\right) = \mathrm{tr}\left(A\left(\mathrm{Var}\left[X\right] + \boldsymbol{\mu}\boldsymbol{\mu}^T\right)\right) \\ &= \mathrm{tr}\left(A\boldsymbol{\Sigma}\right) + \mathrm{tr}\left(A\boldsymbol{\mu}\boldsymbol{\mu}^{\boldsymbol{T}}\right) = \mathrm{tr}\left(A\boldsymbol{\Sigma}\right) + \boldsymbol{\mu}^T A \boldsymbol{\mu}\end{aligned}$$

blah ■

Corollary 1.1. *Let $X \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_X})$ be a d-dimensional gaussian vector with expected value $\mathbf{0}$, and let σ_j^2 denote the variance of the j'th dimension where $1 \leq j \leq d$. By theorem 2 we have that the expected l_2-norm of such a vector is given by*

$$\mathbb{E}\left[\|X\|^2\right] = tr(\mathbf{\Sigma_X}) = \sum_{j=1}^{d} \sigma_j^2$$

2 Algorithms

2.1 The Gaussian Mechanism

One of the most foundational algorithms for achieving (ε, δ)-DP is the Gaussian Mechanism [5]. It computes the real value of a statistic, where the l_2-sensitivity is known. That is it produces a (ε, δ)-DP estimate of a function $g : \mathbb{R}^{n \times d} \to \mathbb{R}^d$ where the l_2-sensitivity $\Delta(g)$ is known. It does so by computing the value of $g(X)$ and then adding noise to each dimension drawn from the normal distribution $\mathcal{N}(0, \sigma_{\varepsilon,\delta}^2)$. This can be seen as adding a noise vector η which is then distributed according to the multivariate normal distribution $\mathcal{N}(\overrightarrow{0}, \sigma_{\varepsilon,\delta}^2 I)$. The algorithm can be seen in Algorithm 1.

Algorithm 1 The Gaussian Mechanism

Input

$\sigma_{\varepsilon,\delta}$	Standard deviation required to achieve (ε, δ)-DP
$X \in \mathbb{R}^{n \times d}$	Dataset

Output

(ε, δ)-DP estimate of $g(X)$

$\eta \leftarrow$ sample from $\mathcal{N}(\overrightarrow{0}, \sigma_{\varepsilon,\delta}^2 I)$

return $g(X) + \eta$

It is quite apparent that the main difficulty of the mechanism lies in determining a $\sigma_{\varepsilon,\delta}$ which achieves (ε, δ)-DP , and preferably the smallest such one.

The following theorem was initially proven

Theorem 3. [5] *Let $g : \mathbb{R}^{n \times d} \to \mathbb{R}^d$ be an arbitrary d-dimensional function with l_2-sensitvity $\Delta(g) = \max_{X \sim X'} \|g(X) - g(X')\|$, and let $\varepsilon \in (0, 1)$. The Gaussian Mechanism with $\sigma_{\varepsilon,\delta} = \Delta(g)\sqrt{2\ln(1.25/\delta)}/\varepsilon$ is (ε, δ)-DP .*

The proof is rather long and is therefore ommitted here.

In the case where $g(X) = f(X)$ i.e. it is estimating the sum of entries, we have quite intuitively that the error is given by the norm of the noise introduced as

$$\mathbb{E}\left[\|(f(X) + \eta) - f(X)\|^2\right] = \mathbb{E}\left[\|\eta\|^2\right]$$

Which by corollary 1.1 is

$$\mathbb{E}\left[\|\eta\|^2\right] = \sum_{i}^{d} \sigma_{\varepsilon,\delta}^2 = d \cdot \sigma_{\varepsilon,\delta}^2$$

-Does introduce the same variance for all dimensions regardless of individual variance

-Talk about finding the minimal σ s.t. privacy is held, and the error in that case

3 Problem setup

The problem consists of realeasing the sum of vectors in a dataset under differential privacy. More formally we whish to release the value of $f : \mathbb{R}^{n\times d} \to \mathbb{R}^d$ given by

$$f(X) = \sum_{i=1}^{n} x_i$$

under (ε, δ)-DP .

The common factor for achieving (ε, δ)-DP in both the Gaussian Mechanism and the Elliptical Gaussian Mechanism is the requirement that data lie within some hyperrectangle. It is formally described as each dimension of the data must lie within some range $x_i^{(j)} \in [-\Delta_j/2, \Delta_j/2]$. This requirement is needed to know the l_2-sensitivity $\|\Delta\|$ as defined in equation **??** of the data. In this project I will change this assumption and instead look at the case where each dimension is normally distributed. This means that for each $j \in [d]$ we have that $X^{(j)} \sim \mathcal{N}(\mu_j, \sigma_j^2)$. An equivalent formulation is that a the data is multivariately distributed but with no correlation between dimensions. This means that $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where Σ is a diagonal matrix with the variance of each dimension along its diagonal. It is quite apparent that determining a $\|\Delta\|$ is impossible in this setting as the Gaussian distribution is continously defined on the range $(-\infty, \infty)$. What has been done in several recent papers is that data is *clipped* by some threshhold C [2,8]. Clipping is the process of limiting the norm of any one entry to be at most C. This means that every vector is transformed as such

$$\hat{x_i} := \min\left\{\frac{C}{\|x_i\|}, 1\right\} \cdot x_i$$

Clipping entries by a factor C thus means that $\|\Delta\| = C$ as any one entry cannot have more impact on the summation than C. It can then be seen that if the summation $f(X)$ is instead performed on a clipped dataset $\hat{X}$ then this is equivalent to defining the summation function $\hat{f} : \mathbb{R}^{n\times d} \to \mathbb{R}^d$ as

$$\hat{f}(X) = \sum_{i}^{n} \min\left\{\frac{C}{\|x_i\|}, 1\right\} \cdot x_i$$

Then by theorem 3 the gaussian mechanism with the function $\hat{f}$ is (ε, δ)-DP with $\sigma_{\varepsilon,\delta} = C\sqrt{2\ln(1.25/\delta)}/\varepsilon$. Though the mechanism is still (ε, δ)-DP it will now have a larger error when regarding the true sum $f(X) = \sum_i^n x_i$ as the actual answer.

3.1 Gaussian data

Let $X^{(j)} \sim \mathcal{N}(0, \sigma_j^2)$ As the expected l_2-norm of x_i is given by

$$\mathbb{E}\left[\|x_i\|^2\right] = \sum_{j=1}^{d} \sigma_j^2$$

To achieve an expected norm of 1 I will scale each dimension by a factor $\frac{1}{b_j}$ which achieves this. If

$$\hat{x}_i = \left(\frac{x_i^{(0)}}{b_0}, \frac{x_i^{(1)}}{b_1}, \ldots, \frac{x_i^{(d)}}{b_d} \right)$$

This means that $X^{(j)} \sim \mathcal{N}(0, \frac{\sigma_j^2}{b_j^2})$ and the expected norm is given by

$$\mathbb{E}\left[\|x_i\|^2\right] = \sum_{j=1}^{d} \frac{\sigma_j^2}{b_j^2}$$

and I can introduce that constraint that the expected norm after the transformation should be 1. In such a case when noise is added after the transformation $\hat{X} + \eta$ where $\eta \sim N(0, t^2)$ and achieves (ε, δ)-DP in this space then then due to linearity of transformation the noise introduced in the original space is then given by

$$\hat{\eta} = (b_0\eta_0, b_1\eta_1, \ldots, b_d\eta_d)$$

then the error is

$$\mathbb{E}\left[\|\hat{\eta}\|^2\right] = \sum_{j=1}^{d} b_j^2 \cdot t^2 = t^2 \sum_{j=1}^{d} b_j^2$$

I desire a transformation of $x_i^{(j)}$ such that the expected norm is 1. Thus I must scale each dimension by $\frac{1}{b_j}$, and have that

$$\mathbb{E}\left[\|x_i\|^2\right] = 1$$

Minimize $\|\hat{\eta}\|$ under the constraint that $\mathbb{E}\left[\|x_i\|^2\right] = 1$

Lemma 3.1. *Let $X \sim \mathcal{N}(0, \sigma^2)$, and Φ denote the cumulative density function of $\mathcal{N}(0, 1)$, then the cumulative density function of X^2 is given by*

$$F_{X^2}(x) = \Pr\left[X^2 \le x\right] = 2\Phi\left(\frac{\sqrt{x}}{\sigma}\right) - 1$$

Proof

$$\begin{aligned}
&\Pr\left[X^2 \le x\right] = \Pr\left[|X| \le \sqrt{x}\right] = 2\Pr\left[0 \le X \le \sqrt{x}\right] \\
&= 2\left(\Pr\left[X \le \sqrt{x}\right] - \Pr\left[X \le 0\right]\right) = 2\left(\Pr\left[X \le \sqrt{x}\right] - \frac{1}{2}\right) \\
&= 2\Phi\left(\frac{\sqrt{x}}{\sigma}\right) - 1
\end{aligned}$$

■

Corollary 3.1. *From lemma 3.1 we can give following bound for $X \sim \mathcal{N}(0, \sigma^2)$.*

$$\Pr\left[X^2 > (4\sigma)^2\right] < 10^{-4}$$

Theorem 4. *Running algo 3, with parameter* $\mathbb{E}\left[\|x_i\|^2\right] = 1$ *the error of is minimized when*

$$b_i = \frac{1}{\sqrt{\sigma_i}\sqrt{\sum_{j=1}^{d}\sigma_j}}$$

Proof

Using lagrangian multipliers we find the local maxima or minia of the function subject to equality constraints. To do so we construct the lagrangian function $\mathcal{L} : \mathbb{R}^{d+1} \to \mathbb{R}$, and find the stationary points of it, by setting the derivative of it to $\mathbf{0}$.

$$\mathcal{L}(\boldsymbol{b}, \lambda) = \sum_{j=1}^{d}\left(\frac{t}{b_j}\right)^2 + \lambda\left(\sum_{j=1}^{d}(\sigma_j b_j)^2 - 1\right)$$

The derivative with respect to b_i is

$$\frac{\partial \mathcal{L}}{\partial b_i} = \frac{\partial}{\partial b_i}\left(\frac{t}{b_i}\right)^2 + \lambda\left(\sigma_i b_i\right)^2 = -2\frac{t^2}{b_i^3} + 2\lambda\sigma_i^2 b_i$$

I then solve $\frac{\partial \mathcal{L}}{\partial b_i} = 0$ for b_i

$$-2\frac{t^2}{b_i^3} + 2\lambda\sigma_i^2 b_i = 0 \iff \lambda\sigma_i^2 b_i = \frac{t^2}{b_i^3} \iff b_i^4 = \frac{t^2}{\lambda\sigma_i^2} \iff b_i = \frac{\sqrt{t}}{\lambda^{\frac{1}{4}}\sqrt{\sigma_i}} \tag{1}$$

I now have the last partial derivative $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ which I solve for λ using the previous expression for b_i.

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{j=1}^{d}(\sigma_j b_j)^2 - 1$$

$$\sum_{j=1}^{d}(\sigma_j b_j)^2 - 1 = 0 \iff \sum_{j=1}^{d}\sigma_j^2\left(\frac{t}{\sqrt{\lambda}\sigma_j}\right) = 1 \iff$$

$$\frac{t}{\sqrt{\lambda}}\sum_{j=1}^{d}\frac{\sigma_j^2}{\sigma_j} = 1 \iff t\sum_{j=1}^{d}\sigma_j = \sqrt{\lambda}$$

Inserting back into equation 1

$$b_i = \frac{\sqrt{t}}{\lambda^{\frac{1}{4}}\sqrt{\sigma_i}} = \frac{\sqrt{t}}{\sqrt{t\sum_{j=1}^{d}\sigma_j}\sqrt{\sigma_i}} = \frac{1}{\sqrt{\sum_{j=1}^{d}\sigma_j}\sqrt{\sigma_i}}$$

■

Theorem 5. *Expected Error is given by*

$$\mathbb{E}\left[\|\eta\|^2\right] = t^2\left(\sum_{i=1}^{d}\sigma_i\right)^2 \leq \sigma_{\varepsilon,\delta}\cdot\alpha\cdot\left(\sum_{i=1}^{d}\sigma_i\right)^2$$

Lemma 3.2. *Let $X_1, X_2, \ldots, X_d$ be d independent random gaussian variables where for $1 \leq j \leq d$ we have that $X_j \sim \mathcal{N}(0, \sigma_j)$. Then the probability for the sum of variables squared is bounded by*

$$\Pr\left[\sum_{j\in[d]} X_j^2 \geq \sqrt{32} \cdot t \cdot \frac{\sqrt{\sum_{j\in[d]} \sigma_j^2}}{\sum_{j\in[d]} \sigma_j}\right] < e^{-t^2}$$

for

$$0 \leq t \leq \sqrt{4} \cdot \frac{\sqrt{\sum_{i\in[d]} \sigma_i^2}}{\sum_{i\in[d]} \sigma_i}$$

Proof
Let $Y_j = X_j^2 - \mathbb{E}\left[X_j^2\right]$ be the shifted chi-square with 1 degree of freedom using the j'th gaussian random variable. As $\mathbb{E}\left[Y_j\right] = \mathbb{E}\left[X_j^2 - \mathbb{E}\left[X_j^2\right]\right] = \mathbb{E}\left[X_j^2\right] - \mathbb{E}\left[X_j^2\right] = 0$ we have that Y_j is zero centered. We are thus interested in giving bounds on $\Pr\left[\sum_{j\in[d]} Y_j\right]$. We can use Bernsteins inequality, if the following holds for all $k \in \mathbb{N}$ with $k \geq 2$ and for all $j \in [d]$, and for some $L \in \mathbb{R}$

$$\mathbb{E}\left[|Y_j^k|\right] \leq \frac{1}{2}\mathbb{E}\left[Y_j^2\right] L^{k-2} k!$$

Initially we have that $\mathbb{E}\left[|Y_j^k|\right] = \mathbb{E}\left[|\left(X_j^2 - \mathbb{E}\left[X_j^2\right]\right)^k|\right]$ and since $X_j^2 \geq 0$ and therefore also $\mathbb{E}\left[X_j^2\right] \geq 0$ we can therefore bound it by

$$\mathbb{E}\left[|\left(X_j^2 - \mathbb{E}\left[X_j^2\right]\right)^k|\right] \leq \mathbb{E}\left[|X_j^{2k}|\right] = \mathbb{E}\left[|\left(\sigma_j \hat{X}\right)^{2k}|\right] = \sigma_j^{2k} \cdot \mathbb{E}\left[|\hat{X}^k|\right]$$

Where $\hat{X} \sim \chi_1^2$ is a chi square with 1 degree of freedom. The moment generating function of $\hat{X}$ is $\mathbb{E}\left[X\right](\hat{X}^m) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m-1)$

It
Bernsteins inequality then states that

$$\Pr\left[\sum_{j\in[d]} X_j \geq 2t\sqrt{\sum_{j\in[d]} \mathbb{E}\left[X_i^2\right]}\right] < e^{-t^2}$$

for all

$$0 \leq t \leq \frac{1}{2L}\sqrt{\sum}$$

$\blacksquare$

I have that $\bar{X}$ is a standard gaussian variable.
Let $Y_j = X_j^2$ I am interested in $\sum_{j\in[d]} Y_j$, then it has

$$\mathbb{E}\left[|Y_j^k|\right] = \mathbb{E}\left[|X_j^{2k}|\right] = \mathbb{E}\left[|X_j|^{2k}\right] = \mathbb{E}\left[|\sigma_j \bar{X}|^{2k}\right] =$$
$$|\sigma_j|^{2k}\mathbb{E}\left[|\bar{X}|^{2k}\right] = \sigma_j^{2k}\prod_{c=1}^{k} 2c = \sigma_j^{2k} \cdot 2^k \cdot k!$$

Which means $\mathbb{E}\left[Y_j^2\right] = 8\sigma_j^4$. To check the constraint (find value for L)

$$\begin{aligned}
&\sigma_j^{2k} \cdot 2^k \cdot k! \leq \frac{1}{2} \cdot 8\sigma_j^4 \cdot L^{k-2}k! \iff \\
&\sigma_j^{2k} \cdot 2^k \leq 4\sigma_j^4 \cdot L^{k-2} \iff \\
&\sigma_j^{2(k-2)} \cdot 2^k \leq 4Ł^{k-2}
\end{aligned}$$

We have that $\sigma_j = \sigma_j b_j = \frac{\sqrt{\sigma_j}}{\sqrt{\sum_{i\in[d]} \sigma_i}} = \sqrt{\frac{\sigma_j}{\sum_{i\in[d]} \sigma_i}}$. Therefore

$$\sigma_j^{2(k-2)} \cdot 2^k = \left(\sqrt{\frac{\sigma_j}{\sum_{i\in[d]} \sigma_i}}\right)^{2(k-2)} \cdot 2^k =$$
$$\left(\frac{\sigma_j}{\sum_{i\in[d]} \sigma_i}\right)^{k-2} \cdot 2^k \leq 2^k$$

This therefore implies

$$2^k \leq 4L^{k-2} \iff 2^{k-2} \leq L^{k-2} \iff 2 \leq L$$

If it is desired that less than 10^{-p} points are removed then $t = \sqrt{p\ln(10)}$ as long as

$$\sqrt{p\ln(10)} \leq \frac{1}{2L}\sqrt{\sum_{j\in[d]} \mathbb{E}\left[X_j^2\right]} = \sqrt{4} \cdot \frac{\sqrt{\sum_{i\in[d]} \sigma_i^2}}{\sum_{i\in[d]} \sigma_i}$$

An alternative Again minimize $\|\hat{\eta}\|$, but instead the constraint comes from Chebyshev's inequality, I always have that

$$\Pr\left[|\|x_i\|^2 - \mathbb{E}\left[\|x_i\|^2\right]| \geq k \cdot \sqrt{\mathrm{Var}\left[\|x_i\|^2\right]}\right] \leq \frac{1}{k^2}$$

Therefore I can set $\frac{1}{k^2} = 0.05$, and find a transformation where I decide how many standard deviations I must be away from the mean to have norm greater than 1, i.e.

$$\mathbb{E}\left[\|x_i\|^2\right] + k \cdot \sqrt{\mathrm{Var}\left[\|x_i\|^2\right]} = 1 \implies k = \frac{1 - \mathbb{E}\left[\|x_i\|^2\right]}{\sqrt{\mathrm{Var}\left[\|x_i\|^2\right]}}$$

I then have that

$$\frac{1}{k^2} = \frac{1}{\left(\frac{1-\mathbb{E}[\|x_i\|^2]}{\sqrt{\mathrm{Var}[\|x_i\|^2]}}\right)^2} = \left(\frac{\sqrt{\mathrm{Var}\left[\|x_i\|^2\right]}}{1 - \mathbb{E}\left[\|x_i\|^2\right]}\right)^2 = \frac{\mathrm{Var}\left[\|x_i\|^2\right]}{\left(1 - \mathbb{E}\left[\|x_i\|^2\right]\right)^2}$$

I already know that

$$\begin{aligned} \mathbb{E}\left[\|x_i\|^2\right] &= \sum_{j=1}^{d} \frac{\sigma_j^2}{b_j^2} \\ \mathrm{Var}\left[\|x_i\|^2\right] &= 2\sum_{j=1}^{d} \frac{\sigma_j^4}{b_j^4} \end{aligned}$$

I therefore have the constraint

$$\frac{\mathrm{Var}\left[\|x_i\|^2\right]}{\left(1 - \mathbb{E}\left[\|x_i\|^2\right]\right)^2} = \frac{2\sum_{j=1}^{d} \frac{\sigma_j^4}{b_j^4}}{\left(1 - \sum_{j=1}^{d} \frac{\sigma_j^2}{b_j^2}\right)^2} = 0.05$$

Be aware that this could find cases where expected norm is greater than 1 and the constraint then says that they are never less than 1 in norm. Another restriction could be that expected norm is ¡ 1.

Chebyshev's inequality could also be used when expected norm should be 1 and then put a bound on number of std away one must be to have less than 0.05 fraction of data removed.

Extra

Determining α using Bernsteins inequality

$$\Pr\left[\|x_i\|^2 - \mathbb{E}\left[\|x_i\|^2\right] > t\right] < 2 \cdot \exp\left(-\frac{t^2/2}{\mathrm{Var}\left[\|x_i\|^2\right] + C \cdot t/3}\right)$$

In the case where $\mathbb{E}\left[\|x_i\|^2\right] = 1$ and b_i is optimized in this case, we have that

$$P(\|x_i\|^2 > 1 + t) < 2 \exp\left(-\frac{t^2/2}{\mathrm{Var}\left[\|x_i\|^2\right] + C \cdot t/3}\right)$$

$$\mathrm{Var}\left[\|x_i\|^2\right] = 2\frac{\sum_i^d \sigma_i^2}{\left(\sum_i^d \sigma_i\right)^2}$$

$$C = \max_{i \in [d]}(\sigma_i) \cdot \frac{16}{\sum_i^d \sigma_i}$$

C is decided such that less than 0.0001 fraction of the data is outside this bound in each dimension. i.e.

$$\forall j \in [d] : \Pr\left[X^{(j)} > C\right] < 0.0001$$

Solving for t

$$2 \cdot \exp\left(-\frac{t^2/2}{\mathrm{Var}\left[\|x_i\|^2\right] + C \cdot t/3}\right) = 0.0001 \iff \ln\left(\frac{0.0001}{2}\right) = -\frac{t^2/2}{\mathrm{Var}\left[\|x_i\|^2\right] + C \cdot t/3} \implies$$

$$t = -\frac{C \ln(\frac{0.0001}{2}) - \sqrt{\ln(\frac{0.0001}{2})(-18 \cdot \mathrm{Var}\left[\|x_i\|^2\right] + C^2 \ln(\frac{0.0001}{2}))}}{3}$$

Then we have that $\alpha = 1 + t$.

Determine α for the bound using `https://en.wikipedia.org/wiki/Concentration_inequality`, or `https://web.stanford.edu/class/cs229t/2017/Lectures/concentration-slides.pdf`

References

[1] Bates, D. *Quadratic Forms of Random Variables.* University of Wisconsin-Madison: STAT 849 lectures, 2010.

[2] Biswas, S., Dong, Y., Kamath, G., and Ullman, J. Coinpress: Practical private mean and covariance estimation. *Advances in Neural Information Processing Systems 33* (2020), 14475–14485.

[3] Chen, T., and Lumley, T. Numerical evaluation of methods approximating the distribution of a large quadratic form in normal variables. *Computational Statistics & Data Analysis 139* (2019), 75–81.

[4] Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality 7*, 3 (2016), 17–51.

[5] Dwork, C., Roth, A., et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science 9*, 3–4 (2014), 211–407.

[6] Hallam, A. *Some Theorems on Quadratic Forms and Normal Variables.* Iowa State University: Econ 671 lectures, 2004.

[7] Hanson, D. L., and Wright, F. T. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics 42*, 3 (1971), 1079–1083.

[8] Huang, Z., Liang, Y., and Yi, K. Instance-optimal mean estimation under differential privacy.

[9] Mathai, A., and Provost, S. Quadratic forms in random variables.