Supplemental Materials for Differential Privacy for Tensor-Valued Queries

Appendix A
Proofs of TVG

A. Proof of Lemma 2 (Tensor norm inequality)

Proof. We prove by induction. First, letting $N=1$, we have $\mathcal{X}=\mathcal{Y}\times_1 U_1$. By Def. 3, we can get:

$$\mathcal{X}_{i_1 i_2\cdots i_N}=(\mathcal{Y}\times_1 U_1)_{i_1 i_2\cdots i_N}=\sum_{j=1}^{I_1} y_{j i_2\cdots i_N} u_{i_1 j}.$$

By Cauchy–Schwarz inequality, we could get the bound for $\mathcal{X}_{i_1 i_2\cdots i_N}$,

$$\left(\sum_{j=1}^{I_1} y_{j i_2\cdots i_N} u_{i_1 j}\right)^2 \le \left(\sum_{j=1}^{I_1} y_{j i_2\cdots i_N}^2\right)\left(\sum_{j=1}^{I_1} u_{i_1 j}^2\right).$$

Therefore, we can calculate that

$$\begin{aligned}\|\mathcal{X}\|^2 &= \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} x_{i_1 i_2\cdots i_N}^{2,}\\ &\le \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N}\left(\sum_{j=1}^{I_1} y_{j i_2\cdots i_N}^2\right)\left(\sum_{j=1}^{I_1} u_{i_1 j}^2\right),\\ &= \|\mathcal{Y}\|^2\|U_1\|_F^2,\end{aligned}$$

which indicates that the statement is true for $N=1$.

Next, we assume that the statement is true for $N=k-1$:

$$\begin{aligned}&\text{if } \mathcal{X}=\mathcal{Y}\times_1 U_1\times_2 U_2\times_3\ldots\times_{k-1} U_{k-1}\\ &\|\mathcal{X}\|\le\|\mathcal{Y}\|\|U_1\|_F\|U_2\|_F\ldots\|U_{k-1}\|_F.\end{aligned}$$

Then for $N=k$, we have $\mathcal{T}=\mathcal{Y}\times_1 U_1\times_2 U_2\times_3\ldots\times_{k-1} U_{k-1}\in\mathbb{R}^{I_1\times\cdots\times I_N}$ satisfies the inequality above. So we have

$$\mathcal{X}=\mathcal{Y}\times_1 U_1\times_2 U_2\times_3\ldots\times_k U_k=\mathcal{T}\times_k U_k.$$

Following the conclusion of the case when $N=1$, we have

$$\|\mathcal{X}\|\le\|\mathcal{T}\|\|U_k\|_F.$$

With the inequality for $N=k-1$, we obtain

$$\|\mathcal{X}\|\le\|\mathcal{Y}\|\|U_1\|_F\|U_2\|_F\ldots\|U_N\|_F,$$

which indicates that the statement is true for $N=k$. By induction, we could claim that statement is true for for every natural number N. □

B. Proof of Thm. 1

Proof. By Def. 1, to guarantee (ϵ,δ)-differential privacy, we should have the following for each pair of datasets $\mathcal{X},\mathcal{X}'$ and any possible output set $\mathcal{O}$ that

$$\Pr(f(\mathcal{X})+\mathcal{Z}\in\mathcal{O})\le e^\epsilon\cdot\Pr(f(\mathcal{X}')+\mathcal{Z}\in\mathcal{O})+\delta, \quad (1)$$

which can be rewritten as

$$\Pr(\mathcal{Z}\in\mathcal{O}-f(\mathcal{X}))\le e^\epsilon\cdot\Pr(\mathcal{Z}\in\mathcal{O}-f(\mathcal{X}'))+\delta.$$

On the other hand, by the definition of TVG, we can rewrite $\mathcal{Z}\sim\mathcal{TVG}\left(0,\Sigma_1,\ldots,\Sigma_N\right)$ as

$$\mathcal{Z}=\mathcal{N}\times_1 U_1\times_2 U_2\times_3\ldots\times_N U_N. \quad (2)$$

Then we define the following events:

$$\mathbf{R}_1=\{\mathcal{N}:\|\mathcal{N}\|^2\le\zeta^2(\delta)\},\mathbf{R}_2=\{\mathcal{N}:\|\mathcal{N}\|^2>\zeta^2(\delta)\}, \quad (3)$$

where $\zeta^2(\delta)$ is defined in Lemma 3, and the $\|\cdot\|$ is defined in Def. 4. Then we observe

$$\begin{aligned}&\Pr[\mathcal{Z}\in\mathcal{O}-f(\mathcal{X})]\\ &\le\Pr[\{\mathcal{Z}\in\mathcal{O}-f(\mathcal{X})\}\cap\mathbf{R}_1]+\Pr[\{\mathcal{Z}\in\mathcal{O}-f(\mathcal{X})\}\cap\mathbf{R}_2].\end{aligned}$$

By the definition of $\zeta^2(\delta)$ and Lemma 3, we have

$$\Pr(\{\mathcal{Z}\in\mathcal{O}-f(\mathcal{X})\}\cap\mathbf{R}_2)\le\Pr(\mathbf{R}_2)\le\delta.$$

In the rest of the proof, we just need to find sufficient conditions for the following inequality to hold:

$$\Pr(\{\mathcal{Z}\in\mathcal{O}-f(\mathcal{X})\}\cap\mathbf{R}_1)\le e^\epsilon\cdot\Pr(\mathcal{Z}\in\mathcal{O}-f(\mathcal{X}')),$$

for differential privacy (1) to be guaranteed. It is also the sufficient conditions for

$$\Pr(\{\mathcal{Z}\in\mathcal{O}-f(\mathcal{X})\}\cap\mathbf{R}_1)\le e^\epsilon\cdot\Pr(\{\mathcal{Z}\in\mathcal{O}-f(\mathcal{X}')\}\cap\mathbf{R}_1).$$

Letting $\mathcal{O}'=\mathcal{O}-f(\mathcal{X})$ and $\Delta=f(\mathcal{X})-f(\mathcal{X}')$, we have

$$\begin{aligned}&\Pr(\mathcal{Z}\in\mathcal{O}'\cap\mathbf{R}_1)\le e^\epsilon\cdot\Pr(\mathcal{Z}\in(\mathcal{O}'+\Delta)\cap\mathbf{R}_1)\\ \Leftrightarrow&\frac{\Pr(\mathcal{Z}\in\mathcal{O}'\cap\mathbf{R}_1)}{\Pr(\mathcal{Z}\in(\mathcal{O}'+\Delta)\cap\mathbf{R}_1)}\le e^\epsilon\\ \Leftrightarrow&\frac{\int_{\mathcal{O}'\cap\mathbf{R}_1}\exp(-\frac{1}{2}\|\mathcal{Z}\times_1 U_1^{-1}\times_2\ldots\times_N U_N^{-1}\|^2)d\mathcal{Z}}{\int_{(\mathcal{O}'+\Delta)\cap\mathbf{R}_1}\exp(-\frac{1}{2}\|\mathcal{Z}\times_1 U_1^{-1}\times_2\ \ldots\times_N U_N^{-1}\|^2)d\mathcal{Z}}\le e^\epsilon\\ \Leftrightarrow&\frac{\exp(-\frac{1}{2}\|\mathcal{Q}\times_1 U_1^{-1}\times_2 U_2^{-1}\times_3\ldots\times_N U_N^{-1}\|^2)}{\exp(-\frac{1}{2}\|(\mathcal{Q}+\Delta)\times_1 U_1^{-1}\times_2 U_2^{-1}\times_3\ldots\times_N U_N^{-1}\|^2)}\le e^\epsilon\\ \Leftrightarrow&\frac{1}{2}\|\Delta'\|^2+\langle\Delta',\mathcal{Q}'\rangle\le\epsilon\end{aligned}$$

where $\Delta'=\Delta\times_1 U_1^{-1}\times_2\ldots\times_N U_N^{-1}$, and $\mathcal{Q}'=\mathcal{Q}\times_1 U_1^{-1}\times_2\ldots\times_N U_N^{-1},\forall\mathcal{Q}\in\mathcal{O}'\cap\mathbf{R}_1$. Since the above inequality needs to hold for any $\mathcal{O}'$ for the differential privacy mechanism to hold, the last two inequalities have to hold as the sufficient conditions.

We then divide the left-hand side of the last inequality into two parts and prove the bound for each as follows. The first part is

$$\|\Delta'\|^2 = \|\Delta \times_1 U_1^{-1} \times_2 U_2^{-1} \times_3 \ldots \times_N U_N^{-1}\|^2 \tag{4a}$$

$$\leq \|\Delta\| \|U_1^{-1}\|_F^2 \ldots \|U_N^{-1}\|_F^2 \tag{4b}$$

$$\leq s_2^2(f) \|U_1^{-1}\|_F^2 \ldots \|U_N^{-1}\|_F^2. \tag{4c}$$

The first inequality is derived from Lemma 2, and the second inequality is due to $\|\Delta\|_F \leq s_2(f)$. For conciseness, we define

$$\phi = \|U_1^{-1}\|_F^2 \|U_2^{-1}\|_F^2 \ldots \|U_N^{-1}\|_F^2,$$

so that the bound for the first part is

$$\|\Delta'\|^2 \leq s_2^2(f)\phi^2. \tag{5}$$

The derivation for the second part is similar to the inequality (4a). Observing that $\mathcal{Q}' = \mathcal{Q} \times_1 U_1^{-1} \times_2 U_2^{-1} \times_3 \ldots \times_N U_N^{-1} = \mathcal{N}$, we have

$$\langle \Delta', \mathcal{Q}' \rangle \leq \sqrt{\langle \Delta', \Delta' \rangle \ \langle \mathcal{Q}', \mathcal{Q}' \rangle}.$$

As what we did in the first part, we could get that

$$\langle \mathcal{Q}', \mathcal{Q}' \rangle \leq \zeta(\delta)^2.$$

Therefore the bound for the second part can be written as follow:

$$\langle \Delta', \mathcal{Q}' \rangle \leq s_2(f)\zeta(\delta)\phi. \tag{6}$$

By combining two Eq. (4c)(6), the inequality becomes,

$$s_2(f)^2\phi^2 + 2s_2(f)\zeta(\delta)\phi \leq 2\epsilon. \tag{7}$$

Note that ϕ can only be non-negative and can be obtained by solving inequality Eq. (7). Letting $\alpha = s_2^2(f)$, $\beta = 2s_2(f)\zeta(\delta)$, we have

$$\phi \leq \frac{-\beta + \sqrt{\beta^2 + 8\alpha\epsilon}}{2\alpha},$$

which is the sufficient condition of inequality (14) in Thm. 1. □

C. *Proof of Thm. 2 (UDN)*

Proof. The proof follows the proof of Thm. 1. We follow the proof of Thm. 1 and lead to:

$$\begin{aligned} &\Pr(\mathcal{Z} \in \mathcal{O}' \cap \mathbf{R}_1) \leq e^\epsilon \cdot \Pr(\mathcal{Z} \in (\mathcal{O}' + \Delta) \cap \mathbf{R}_1) \\ \Leftrightarrow &\frac{1}{2}\|\Delta'\|^2 + \langle \Delta', \mathcal{Q}' \rangle \leq \epsilon \end{aligned} \tag{8}$$

for any possible $\mathcal{Q} \in \mathcal{O}' \cap \mathbf{R}_1$, and assuming that $\mathcal{Q}' = \mathcal{Q} \times_1 U_1^{-1}$, $\Delta' = \Delta \times_1 U_1^{-1}$. And thus

$$\begin{aligned} &\|\Delta'\|^2 = \|\Delta \times_1 U_1^{-1}\|^2 \\ \leq &\|\Delta\| \|U_1^{-1}\|_F^2 \leq s_2^2(f)\|U_1^{-1}\|_F^2. \end{aligned} \tag{9}$$

and the other part is:

$$\langle \Delta', \mathcal{Q}' \rangle \leq s_2(f)\zeta(\delta)\|U_1^{-1}\|_F^2 \tag{10}$$

By combining two parts, we could caculate the bound to guarantee (ϵ, δ)-differential privacy:

$$\|U_1{}^{-1}\|_F^2 \leq \frac{\left(-\beta + \sqrt{\beta^2 + 8\alpha\epsilon}\right)^2}{4\alpha^2}. \tag{11}$$

where $\alpha = s_2^2(f)$, and $\beta = 2\zeta(\delta)s_2(f)$.

It is worth noting that our theorem is a sufficient condition for (ϵ, δ)-differential privacy but not a necessary condition. □

D. *Proof of Thm. 3 (IDN)*

Proof. The proof follows the proof of Thm. 1. We define the set of events $\mathbf{R}_1$ and $\mathbf{R}_2$ as in Eq. (3). And we will focus on the sufficient condition of

$$\Pr(\mathcal{Z} \in \mathcal{O}' \cap \mathbf{R}_1) \leq e^\epsilon \cdot \Pr(\mathcal{Z} \in (\mathcal{O}' + \Delta) \cap \mathbf{R}_1), \tag{12}$$

given any $\mathcal{O}' = \mathcal{O} - f(\mathcal{X})$ and $\Delta = f(\mathcal{X}) - f(\mathcal{X}')$. Since U_1 is a diagonal matrix, the tensor-valued random variable $\mathcal{Z} \sim \mathcal{TVG}(0, U_1, \mathbf{E}_2, \ldots, \mathbf{E}_N)$ and can be expressed as

$$\mathcal{Z} = \mathcal{N} \times_1 U_1.$$

And the pdf of $\mathcal{Z}$ is Eq. (20) in the paper. By substituting the pdf of $\mathcal{Z}$ into the inequality (12), we obtain

$$\begin{aligned} &\Pr(\mathcal{Z} \in \mathcal{O}' \cap \mathbf{R}_1) \leq e^\epsilon \cdot \Pr(\mathcal{Z} \in (\mathcal{O}' + \Delta) \cap \mathbf{R}_1) \\ \Leftrightarrow &\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} \frac{\Delta_{i_1 i_2 \cdots i_N}^2 + 2\Delta_{i_1 i_2 \cdots i_N} z_{i_1 i_2 \cdots i_N}}{2\sigma_{i_1}^2} \leq \epsilon. \end{aligned}$$

We bound the two parts respectively in the last equation. Consider that we are normalizing each feature to the same range, i.e., each element of $F(\mathcal{X})$ is in range $[a, b]$. Then we have $0 \leq \Delta_{i_1 i_2 \cdots i_N}^2 \leq (b-a)^2 = \frac{\hat{s}_2^2(f)}{I_1 I_2 \cdots I_N}$ for every $i_n \in [I_n], \forall 1 \leq n \leq N$. Hence we have

$$\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} \frac{\Delta_{i_1 k_2 \cdots k_N}^2}{2\sigma_{i_1}^2} \leq \sum_{i_1=1}^{I_1} \frac{\hat{s}_2^2(f)}{2I_1\sigma_{i_1}^2} = \frac{\hat{s}_2^2(f)}{2I_1}\|U_1^{-1}\|_F^2. \tag{13}$$

In the second part, we rewrite $\mathcal{Z}$ as $\mathcal{N} \times_1 U_1$. By the condition of $U_1 = diag[\sigma_1, ..., \sigma_{I_1}] \in \mathbb{R}^{I_1 \times I_1}$, we represent each entry of $\mathcal{Z}$ as

$$\begin{aligned} &z_{j i_2 \cdots i_N} = x_{j i_2 \cdots i_N}\sigma_j, \\ &x_{j i_2 \cdots i_N} \sim N(0, 1), \ \forall j \in [I_1], i_2 \in [I_2] \ldots, i_N, \in [I_N]. \end{aligned}$$

Then, the second part could be written as

$$\begin{aligned} &\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} \frac{\Delta_{i_1 i_2 \cdots i_N} z_{i_1 i_2 \cdots i_N}}{\sigma_{i_1}^2} \\ = &\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} \frac{\Delta_{i_1 i_2 \cdots i_N} x_{i_1 i_2 \cdots i_N}}{\sigma_{i_1}} \\ \leq &\sqrt{\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} \frac{\Delta_{i_1 i_2 \cdots i_N}^2}{\sigma_{i_1}^2}} \sqrt{\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2}. \end{aligned} \tag{14}$$

The inequality is by Cauchy inequality to single out $x_{i_1 i_2 \cdots i_N}$. According to the definition of $\mathbf{R}_1$, we know that if $\mathcal{N} \in \mathbf{R}_1$, $\|\mathcal{N}\|_F^2 \leq \zeta^2(\delta)$. Hence,

$$\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2 \leq \|\mathcal{N}\|_F^2 \leq \zeta^2(\delta).$$

Thus we have the following inequality holds:

$$\begin{aligned} &\sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} \frac{\Delta_{i_1 i_2 \cdots i_N}^2}{\sigma_{i_1}^2}} \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2} \\ \leq &\sqrt{\frac{\hat{s}_2^2(f)}{I_1}} \|U_1^{-1}\|_F \zeta(\delta), \end{aligned} \tag{15}$$

by the inequality (13). Finally, we combine the Eq.(13)(15) to obtain

$$\frac{\hat{s}_2^2(f)}{2I_1} \|U_1^{-1}\|_F^2 + \frac{\hat{s}_2(f)}{\sqrt{I_1}} \zeta(\delta) \|U_1^{-1}\|_F \leq \epsilon.$$

This is a quadratic inequality of $\|U_1^{-1}\|_F$, and with the condition $\|U_1^{-1}\|_F > 0$, we can solve that

$$\|U_1^{-1}\|_F^2 \leq \frac{I_1}{\hat{s}_2^2(f)} \left(-\zeta(\delta) + \sqrt{\zeta^2(\delta) + 2\epsilon} \right)^2,$$

which completes the proof. ☐

APPENDIX B
PROOF FOR THE TVG ERROR

We generate the noise $\mathcal{Z}$ from $\mathcal{N}$ such that

$$\mathcal{Z} = \mathcal{N} \times_1 U_1 \times_2 U_2 \times_3 \ldots \times_N U_N.$$

For the proof of the TVG error, we first need to prove the following lemma:

Lemma B.1. *Suppose that A is a matrix valued variable with the size $m \times m$, then*

$$\mathbb{E}\ (\boldsymbol{tr} A) = \boldsymbol{tr}\ (\mathbb{E} A),$$

where $\boldsymbol{tr} A$ represents the trace of A.

Proof. We find that

$$\mathbb{E}\ (\boldsymbol{tr} A) = \mathbb{E}(\sum_{i=1}^{m} A_{ii}) = \sum_{i=1}^{m} \mathbb{E} A_{ii},$$

$$\boldsymbol{tr}\ (\mathbb{E} A) = \boldsymbol{tr}\ (\mathbb{E} A_{ij})_{m \times m} = \sum_{i=1}^{m} \mathbb{E} A_{ii}.$$

Therefore, $\mathbb{E}\ (\boldsymbol{tr} A) = \boldsymbol{tr}\ (\mathbb{E} A)$. ☐

With the above lemma, we could present the theorem of calculating the expection of $\|\mathcal{Z}\|^2$.

Theorem B.1. *For the given noise*

$$\mathcal{Z} = \mathcal{N} \times_1 U_1 \times_2 U_2 \times_3 \ldots \times_N U_N \in \mathbb{R}^{I_1 \times \cdots \times I_N}, \tag{16}$$

where $\mathcal{N} \in \mathcal{R}^{I_1 \times \cdots \times I_N}$ is a SND noise from Def 6, $U_k \in \mathbb{R}^{I_k \times J_k}, \forall k \in [N]$, we have

$$\mathbb{E}\ \|\mathcal{Z}\|^2 = \|U_N\|_F^2 \|U_{N-1}\|_F^2 \cdots \|U_1\|_F^2. \tag{17}$$

Proof. First, we obtain the matricization of $\mathcal{Z}$ with the lemma 1:

$$\mathcal{Z}_{(1)} = U_1 \mathcal{N}_{(1)} \left(U_N \otimes \cdots \otimes U_2 \right)^\top = U_1 \mathcal{N}_{(1)} V^\top,$$

where $V = U_N \otimes \cdots \otimes U_2$ for presentation conciseness.

$$\mathbb{E}\ \|\mathcal{Z}\|^2 = \mathbb{E}\ \|\mathcal{Z}_{(1)}\|_F^2 = \mathbb{E}\ \boldsymbol{tr}(\mathcal{Z}_{(1)} \mathcal{Z}_{(1)}^\top) \tag{18}$$

Then, with the Lemma. B.1, we could switch the trace and the expectation. Thus

$$\begin{aligned} &\mathbb{E}\ \boldsymbol{tr}(\mathcal{Z}_{(1)} \mathcal{Z}_{(1)}^\top) = \boldsymbol{tr}(\mathbb{E}\ \mathcal{Z}_{(1)} \mathcal{Z}_{(1)}^\top) \\ &= \boldsymbol{tr}(\mathbb{E}\ U_1 \mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top U_1^\top) \\ &= \boldsymbol{tr}(U_1 \mathbb{E}\ [\mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top] U_1^\top). \end{aligned} \tag{19}$$

Hence we focus on the $\mathbb{E}\ [\mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top]$. Assume that $\mathcal{N}^\top = (\boldsymbol{n}_1, \boldsymbol{n}_2, \cdots, \boldsymbol{n}_{I_1})$, $I = I_1 I_2 I_3 \cdots I_N$, and $I' = I / I_1$. We have

$$\left(\mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top \right)_{ij} = \boldsymbol{n}_i^\top V^\top V \boldsymbol{n}_j.$$

Therefore, if $i \neq j$, all the random variables in $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$ are independent. Hence we could get that

$$\mathbb{E}\ [\mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top]_{ij} = 0$$

If $i = j$, we assume that $\boldsymbol{z}_i = V \boldsymbol{n}_i = (z_{1i}, z_{2i}, \ldots, z_{Ii})^\top$, and thus

$$\mathbb{E}\ [\mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top]_{ii} = \mathbb{E}\ \sum_{k=1}^{I'} z_{ki}^2$$

where $z_{ki} \sim \mathcal{N}(0, \sum_{l=1}^{I} V_{lk}^2)$. Therefore,

$$\mathbb{E}\ \sum_{k=1}^{I'} z_{ki}^2 = \sum_{k=1}^{I'} \mathbb{E}\ z_{ki}^2 = \sum_{k=1}^{I'} \sum_{l=1}^{I'} V_{lk}^2 = \|V\|_F^2.$$

Hence,

$$\mathbb{E}\ [\mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top] = \|V\|_F^2 \mathbf{E}_1$$

Finally, we substitute the expectation into (19) to obtain

$$\boldsymbol{tr}(U_1 \mathbb{E}\ [\mathcal{N}_{(1)} V^\top V \mathcal{N}_{(1)}^\top] U_1^\top) = \|U_1\|_F^2 \|V\|_F^2. \tag{20}$$

With the properties of Kronecker product, we have that

$$\|V\|_F^2 = \|U_N\|_F^2 \|U_{N-1}\|_F^2 \cdots \|U_2\|_F^2.$$

The amount of noise is

$$\mathbb{E}\ \|\mathcal{Z}\|^2 = \|U_N\|_F^2 \|U_{N-1}\|_F^2 \cdots \|U_1\|_F^2.$$

The proof completes. ☐

By Thm. B.1, we could formulate the optimization problem as follows:

$$\begin{aligned} &\min_{U_1 \cdots U_N}\ \mathbb{E} \|\mathcal{Z} \times_1 W_1 \times_2 W_2 \cdots \times_N W_N\|^2 \\ \Leftrightarrow &\min_{U_1 \cdots U_N}\ \|W_1 U_1\|_F^2 \|W_2 U_2\|_F^2 \cdots \|W_N U_N\|_F^2 \\ \Leftrightarrow &\min_{U_1 \cdots U_N}\ \|W_1 W_{U_1} S_{U_1}\|_F^2 \|W_2 W_{U_2} S_{U_2}\|_F^2 \cdots \|W_N W_{U_N} S_{U_N}\|_F^2, \end{aligned} \tag{21}$$

where $S_{U_k} = diag(\sigma_{k1}, ..., \sigma_{kI_k})$. If we let $P_{ki} = \sum_{j=1}^{J_k} (W_k W_{U_k})_{ji}^2$, we can write our objective as

$$\min_{U_1 \cdots U_N} \prod_{k=1}^{N} \sum_{i=1}^{I_k} P_{ki} \sigma_{ki}^2. \tag{22}$$

Together with the differential privacy constraint, we have a geometric programming problem:

$$\begin{aligned} &\min_{U_1 \cdots U_N} \prod_{k=1}^{N} \sum_{i=1}^{I_k} P_{ki} \sigma_{ki}^2 \\ &\text{s.t.} \quad \prod_{k=1}^{N} \sum_{i=1}^{I_k} \frac{1}{\sigma_{ki}^2} \leq B, \end{aligned} \tag{23}$$

where $B = \frac{\left(-\beta + \sqrt{\beta^2 + 8\alpha\epsilon}\right)^2}{4\alpha^2}$. The we convert the problem into a convex one by letting $e^{x_{ki_k}} = \sigma_{ki_k}^2$:

$$\begin{aligned} &\underset{x}{\text{minimize}} \log(g(x)), \\ &s.t. \quad \log(g_1(x)) \leq \log(B), \end{aligned}$$

where

$$\begin{aligned} g(x) = &\sum_{n=1}^{N} \sum_{i_n=1}^{I_n} \prod_{k=1}^{N} P_{ki_k} e^{x_{ki_k}}, \\ g_1(x) = &\sum_{n=1}^{N} \sum_{i_n=1}^{I_n} \prod_{k=1}^{N} e^{-x_{ki_k}}. \end{aligned}$$

KKT conditions [1] can be applied and we obtain the optimal solution:

$$\prod_{k=1}^{N} \sigma_{ki_k}^2 = \frac{\prod_{k=1}^{N} \sum_{i=1}^{I_k} \sqrt{P_{ki}}}{\prod_{k=1}^{N} \sqrt{P_{ki_k}} B}, \quad \forall i_k \in [I_k], \ k \in [N], \tag{24}$$

as well as the minimum value of the objective:

$$\text{Error}_{\text{TVG}}(\mathcal{Y}, \epsilon, \delta) = \frac{\left(\prod_{k=1}^{N} \sum_{i=1}^{I_k} \sqrt{P_{ki}}\right)^2}{B}. \tag{25}$$

Appendix C
Proofs for Corollaries

A. Proof of Corollary 1

Proof. Let $I = I_1 \cdots I_N = J_1 \cdots J_M$. Suppose that $U_k U_k^\top = \Sigma_k, \ \forall k \in [N]$ and $V_m V_m^\top = \Gamma_m, \ \forall m \in [M]$. We reshape $\mathcal{Z}_1$ to a vector $vec(\mathcal{Z}_1) \in \mathbb{R}^I$ and $\mathcal{Z}_2$ to a vector $vec(\mathcal{Z}_2) \in \mathbb{R}^I$. Then we have that

$$\begin{aligned} vec(\mathcal{Z}_1) &\sim \mathcal{N}(0, (\Sigma_1 \otimes \cdots \otimes \Sigma_N)), \\ vec(\mathcal{Z}_2) &\sim \mathcal{N}(0, (\Gamma_1 \otimes \cdots \otimes \Gamma_M)). \end{aligned}$$

Let $\Sigma = \Sigma_1 \otimes \cdots \otimes \Sigma_N, \ \Gamma = \Gamma_1 \otimes \cdots \otimes \Gamma_M$ and $UU^\top = \Sigma, VV^\top = \Gamma$. Obviously, $U = U_1 \otimes \cdots \otimes U_N, V = V_1 \otimes \cdots \otimes V_M$ and

$$vec(\mathcal{Z}_1) = UN, \ vec(\mathcal{Z}_2) = VN$$

where $N \sim \mathcal{N}(0, \mathbf{E}_I)$ is a standard normal random vector.

Due to $\mathcal{Z}_2$ is a reshaped tensor from $\mathcal{Z}_1$, $vec(\mathcal{Z}_1)$ and $vec(\mathcal{Z}_2)$ have the same elements, which only differ by different permutation. Therefore, there exist a series of row-switching transformations matrices $T_1, T_2, \ldots, T_k$, such that

$$\begin{aligned} &T_k \cdots T_1 vec(\mathcal{Z}_1) = vec(\mathcal{Z}_2) \\ \Leftrightarrow &T_k \cdots T_1 U = V. \end{aligned}$$

With the properties of row-switching transformations matrices $T_1^{-1} = T_1 = T_1^\top$, we could get that

$$\|V\|_F = \|U\|_F.$$

Therefore, with the mechanism $\mathbf{TVG}_1$ satisfies (ϵ, δ)-differential privacy, we could give mechanism $\mathbf{TVG}_2$ a explicit bound

$$\begin{aligned} &\|V_1^{-1}\|_F^2 \cdots \|V_M^{-1}\|_F^2 = \|V\|_F^2 = \|U\|_F^2 \\ =&\|U_1^{-1}\|_F^2 \cdots \|U_N^{-1}\|_F^2 \leq B, \end{aligned}$$

where $B = \frac{1}{s_2^2(f)} \left(-\zeta(\delta) + \sqrt{\zeta^2(\delta) + 2\epsilon}\right)^2$ and $\zeta(\delta) = -2\ln\delta + 2\sqrt{-I\ln\delta} + I$. Therefore, the mechanism $\mathbf{TVG}_2$ satisfies the same (ϵ, δ)-differential privacy as $\mathbf{TVG}_1$ does. □

B. Proof of Corollary 2

Proof. We borrow the notations and settings from Sec. C-A. Assume $\mathbf{TVG}_1$ and $\mathbf{TVG}_2$ satisfy the same (ϵ, δ)-differential privacy, and $U_k U_k^\top = \Sigma_k, \ \forall k \in [N]$ and $V_m V_m^\top = \Gamma_m, \ \forall m \in [M]$. The differential privacy constraints are

$$\begin{aligned} \|U_1^{-1}\|_F^2 \ldots \|U_N^{-1}\|_F^2 \leq B_1, \\ \|V_1^{-1}\|_F^2 \ldots \|V_M^{-1}\|_F^2 \leq B_2, \end{aligned}$$

where $B_1 = \frac{1}{s_2^2(f)} \left(-\zeta(\delta)_1 + \sqrt{\zeta^2(\delta)_1 + 2\epsilon}\right)^2$, and $B_2 = \frac{1}{s_2^2(f)} \left(-\zeta(\delta)_2 + \sqrt{\zeta^2(\delta)_2 + 2\epsilon}\right)^2$ by definition. With lemma 3, we could calculate that

$$\begin{aligned} \zeta(\delta)_1 &= -2\ln\delta + 2\sqrt{-I_1 I_2 \cdots I_N \ln\delta} + I_1 I_2 \cdots I_N, \\ \zeta(\delta)_2 &= -2\ln\delta + 2\sqrt{-J_1 J_2 \cdots J_N \ln\delta} + J_1 J_2 \cdots J_M, \end{aligned}$$

Because $I_1 \cdots I_N = J_1 \cdots J_M$, we have $B_1 = B_2$. Since ${W_k}', k \in [N]$ are reshaped from $W_k, k \in [N]$, ensuring each element in $f(\mathcal{X})$ is multiplied by the same coefficient in $f'(\mathcal{X})$, we obtain that

$$\begin{aligned} &\mathbb{E}\|\mathcal{Z} \times_1 W_1 \times_2 W_2 \cdots \times_N W_N\|^2 \\ =&\mathbb{E}\|\mathcal{Z}' \times_1 {W_1}' \times_2 {W_2}' \cdots \times_N W_N'\|^2. \end{aligned}$$

And with the same constraints, the optimal solution should be the same, i.e.,

$$\text{Error}_{\text{TVG}}(\mathcal{Y}, \epsilon, \delta) = \text{Error}_{\text{TVG}}(\mathcal{Y}', \epsilon, \delta). \tag{26}$$

□

Appendix D
Composition and Sampling

Here we also introduce the composition theorem used in the paper, which follows [2](Theorem 3.16).

Theorem D.1. *Let $\mathcal{M}_i : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_i$ be an (ϵ_i, δ_i)-differentially private algorithm for $i \in [k]$. Then if $\mathcal{M}_{[k]} : \mathbb{N}^{|\mathcal{X}|} \to \prod_{i=1}^{k} \mathcal{R}_i$ is defined* *to be* $\mathcal{M}_{[k]}(x) = (\mathcal{M}_1(x), \ldots, \mathcal{M}_k(x))$*, then* $\mathcal{M}_{[k]}$ *is* $\left(\sum_{i=1}^{k} \varepsilon_i, \sum_{i=1}^{k} \delta_i\right)$*-differentially private.*

Assume that we add noise to a linear query $f(\mathcal{X}) = \mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. If we apply Gaussian mechanism $\mathcal{M}$ to each element of $\mathcal{X}$, each mechanism satisfies (ϵ, δ)-differential privacy. Therefore, the entire linear query answer satisfies $(I_1 I_2 \cdots I_N \epsilon, I_1 I_2 \cdots I_N \delta)$-differential privacy.

For compositions in differentially-private SGD, we adopt the Theorem 3.4 from [3] to ensure the overall (ϵ, δ)-differential privacy. We also adopt the privacy amplification via sampling such that:

Theorem D.2 (lemma 2 in [4])**.** *Let* $\mathcal{A}$ *be an* ϵ^**-differentially private algorithm. Construct an algorithm* $\mathcal{B}$ *that on input a database* $D = (d_1, \ldots, d_n)$*, constructs a new database* D_s *whose i-th entry is* d_i *with probability* $f(\epsilon, \epsilon^*) = (\exp(\epsilon) - 1)/(\exp(\epsilon^*) + \exp(\epsilon) - \exp(\epsilon - \epsilon^*) - 1)$*, and* $\perp$ *otherwise, and then runs* $\mathcal{A}$ *on* D_s*. Then,* $\mathcal{B}$ *is* ϵ*-differentially private.*

For example, during training process, we take a random sample from the training set with sampling probability q. Then we have $f(\epsilon, \epsilon^*) = q$, and ϵ^* can be calculated as the private budget of the mechanism after sampling.

References

[1] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[2] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[3] P. Kairouz, S. Oh, and P. Viswanath, "The Composition Theorem for Differential Privacy," *IEEE Transactions on Information Theory (TIT)*, vol. 63, no. 6, pp. 4037–4049, 2017.

[4] A. Beimel, S. P. Kasiviswanathan, and K. Nissim, "Bounds on the sample complexity for private learning and private data release," in *Theory of Cryptography Conference*. Springer, 2010, pp. 437–454.