A APPENDIX

The appendix is divided into three parts. We first introduce some preliminaries before giving detailed proof of the main results. We finally discuss more details for truncated mechanisms.

A.1 PRELIMINARIES

We give some specific forms of operator T in KM iteration in Table 1. The SGD, SPGD and ADMM algorithms are the special cases of KM iteration.

Table 1: Overview of several first-order algorithms

Algorithm	Operator identity (T)	Subgradient identity
SGD	$I-\gamma\nabla f$	$x^{k+1}=\mathbf{W}x^k-\gamma_k\nabla f(x^k)$
SPGD	$(I+\gamma\partial g)^{-1}(I-\gamma\nabla f)$	$x^{k+1}=\mathbf{W}x^k+(\mathrm{prox}_{\gamma_k g}(I-\gamma_k\nabla f(x^k))-x^k)$
ADMM	$(I+\gamma\partial f)^{-1}\left[(I+\gamma\partial g)^{-1}(I-\gamma\partial f)+\gamma\partial f\right]$	$x^{k+1}=\mathbf{W}x^k+\frac{1}{2}\,\mathrm{refl}_{\gamma_k\partial f}\circ\mathrm{refl}_{\gamma_k\partial g}\left(x^k\right)$

Table 2 concludes the sensitivity for different typologies in decentralized learning when the step size is chosen as $\alpha_k=\alpha$ and $\alpha_k=1/(k+1)$.

Table 2: Δ_K-Sensitivity under Different Graph Typologies

Graph typology	Spectral gap (1-λ)	Sensitivity ($\alpha_k=\alpha$)	Sensitivity ($\alpha_k=(1/(k+1))$)
Ring	$\mathcal{O}(\frac{1}{M^2})$	$\mathcal{O}(\frac{\alpha K}{MN}+M^2\alpha K)$	$\mathcal{O}(\frac{\ln(K)}{MN}+M^2\ln(K))$
Grid	$\mathcal{O}(\frac{1}{M\log(M)})$	$\mathcal{O}(\frac{\alpha K}{MN}+M\log(M)\alpha K)$	$\mathcal{O}(\frac{\ln(K)}{MN}+M\log(M)\ln(K))$
Star	$\mathcal{O}(\frac{1}{M})$	$\mathcal{O}(\frac{\alpha K}{MN}+M\alpha K)$	$\mathcal{O}(\frac{\ln(K)}{MN}+M\ln(K))$
Exponential	$\mathcal{O}(\frac{1}{\log(M)})$	$\mathcal{O}(\frac{\alpha K}{MN}+\log(M)\alpha K)$	$\mathcal{O}(\frac{\ln(K)}{MN}+\log(M)\ln(K))$
Full connected	1	$\mathcal{O}(\frac{\alpha K}{MN}+\alpha K)$	$\mathcal{O}(\frac{\ln(K)}{MN}+\ln(K))$

We know that the connection between local servers is described by a mixing matrix $\mathbf{W}$ for different topologies. The mixing matrix has a special property given in Lemma 1.

Lemma 1 (Convergence of Mixing Matrix Zhu et al. (2022)) *Let $\mathbf{P}\in\mathbb{R}^{M\times M}$ be a matrix whose elements are all $1/M$. Given any $t\in\mathbb{Z}^+$, the mixing matrix $\mathbf{W}\in\mathbb{R}^{M\times M}$ satisfies,*

$$\left\|\mathbf{W}^t-\mathbf{P}\right\|_{op}\leq\lambda^t,$$

where $\|\cdot\|_{op}$ denotes the spectral norm. Note that $\mathbf{W}$ corresponds to some Markov chain's transition matrix, and the parameter $0\leq\lambda<1$ characterizes the speed of convergence to the stationary state.

In this paper, we consider three topologies: ring, star, and full connected graph in Figure 4.

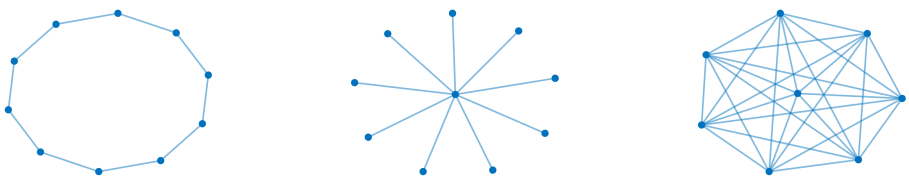

Figure 4: Structures of the connected graph from left to right: ring, star, full connected.

Definition 7 (Nonexpansive Operator) *A mapping $T:\mathbb{X}\to\mathbb{X}$ is nonexpansive if $\|Tx-Ty\|\leq\|x-y\|$ holds for all $x,y\in\mathbb{X}$.*

We next introduce differential privacy based on KL divergence, which measures the distance between two distributions. The bound given in Definition 8 is useful for establishing the composition theorem.

Definition 8 (**(ε, α)-Rényi Differential Private Duchi & Rogers (2019)**) *For distributions P and Q, $D_\alpha(P\|Q) := \frac{1}{\alpha-1}\log\int(dP/dQ)^\alpha dQ$ is the Rényi α-divergence. For $\alpha \geq 1$, $\mathcal{A}$ is (ε, α)-Rényi differential private if for all Ξ, Ξ', $D_\alpha\left(\mathcal{A}(\Xi)\|\mathcal{A}\left(\Xi'\right)\right) \leq \varepsilon$. By taking $\alpha = 1$, we obtain ε-KL-privacy. If $\mathcal{A}$ is ε-differential private, then for any $\alpha \geq 1$, it also satisfies,*

$$D_\alpha\left(\mathcal{A}(\Xi)\|\mathcal{A}\left(\Xi'\right)\right) \leq \min\left\{2(\alpha-1)\varepsilon^2 + \min\left\{2, (e^\varepsilon - 1)\right\}\varepsilon, \varepsilon\right\}.$$

Lemma 2 (Mohri et al. (2018)) *Suppose $\{Y_i\}_{i=1}^N$ is a sequence of random variables, $Y_i \in [-c_i, c_i]$. Let $\{X_i\}_{i=1}^N$ be a sequence of random variables such that $\mathbb{E}[Y_i \mid X_{i-1}, \cdots, X_i] \leq C_i$, where $\{C_i\}_{i=1}^N$ is sequence of constant real numbers. Then,*

$$\mathbb{P}\left(\sum_{i=1}^N Y_i \geq \sum_{i=1}^N C_i + t\sqrt{\sum_{i=1}^N c_i^2}\right) \leq e^{-\frac{t^2}{2}}.$$

Lemma 3 (Dwork & Rothblum (2016)) *For any random variables Y, Z, we have that $D_\infty^\delta(Y\|Z) \leq \varepsilon$, $D_\infty^\delta(Z\|Y) \leq \varepsilon$ if and only if there exist r.v.s, Y', Z' such that $\Delta\left(Y\|Y'\right) \leq \frac{\delta}{e^{\varepsilon+1}}$, $\Delta\left(Z\|Z'\right) \leq \frac{\delta}{e^{\varepsilon+1}}$, $D_\infty\left(Y'\|Z'\right) \leq \varepsilon, D_\infty\left(Z'\|Y'\right) \leq \varepsilon$, where $D_\infty(X\|Y) = \max_{\nu\in\text{supp}(X)}\left[\log\frac{P(X\in U)}{P(Y\in U)}\right]$, $D_\infty^\delta(X\|Y) = \max_{U\in\text{supp}(i):P(Y\in U)\geq\delta}\left[\log\frac{P(X\in U)-\delta}{P(Y\in U)}\right]$, and $\Delta(X\|Y) = \max_U|P(X\in U) - P(Y\in U)|$.*

A.2 Proof of main results

A.2.1 Proof of Lemma 4

Lemma 4 *$x^k(m)$ is generated by (1) for a given relaxed parameter $\alpha_k \in (0, 1]$. Define $x^k = \frac{1}{M}\sum_{m=1}^M x^k(m)$ is an average iterate of all local agent, $m \in \{1, \cdots, M\}$. Under Assumption 3, we have*

$$\left[\sum_{m=1}^M \|x^k(m) - x^k\|^2\right]^{\frac{1}{2}} \leq 2\sqrt{M}B\sum_{j=0}^{k-1}\alpha_j\lambda^{k-1-j}.$$

Proof. Let

$$\mathbf{X}^k := [x^k(1), \cdots, x^k(M)]^\top \in \mathbb{R}^{M\times p},$$
$$\Delta^k := \alpha_k[T(x^k(1); \xi_{i_k(1)}) - x^k(1), \cdots, T(x^k(M); \xi_{i_k(M)}) - x^k(M)]^\top.$$

The matrix form of (1) is as follows,

$$\mathbf{X}^{k+1} = \mathcal{A}(\mathbf{X}^k) = \mathbf{W}\mathbf{X}^k + \Delta^k,$$

where $\mathbf{W}$ satisfies Definition 1 and Lemma 1. We then have,

$$(\mathbb{I} - \mathbf{P})\mathbf{X}^{k+1} = (\mathbb{I} - \mathbf{P})\left[\mathbf{W}\mathbf{X}^k + \Delta^k\right].$$

Therefore,

$$\begin{aligned}
\|(\mathbb{I} - \mathbf{P})\mathbf{X}^{k+1}\| &\leq \|\left(\mathbf{W}\mathbf{X}^k + \Delta^k\right) - \left(\mathbf{P}\mathbf{W}\mathbf{X}^k\right)\| + \|\Delta^k\| \\
&\leq \|\mathbf{W}\mathbf{X}^k + \Delta^k - \mathbf{P}\mathbf{W}\mathbf{X}^k\| + \alpha_k\sqrt{M}B \\
&\leq \|\mathbf{W}\mathbf{X}^k - \mathbf{P}\mathbf{W}\mathbf{X}^k\| + 2\alpha_k\sqrt{M}B \\
&= \left\|(\mathbf{W} - \mathbf{P})(\mathbb{I} - \mathbf{P})\mathbf{X}^k\right\| + 2\alpha_k\sqrt{M}B \\
&\leq \lambda\left\|(\mathbb{I} - \mathbf{P})\mathbf{X}^k\right\| + 2\alpha_k\sqrt{M}B \\
&\leq \lambda^{k+1}\|(\mathbb{I} - \mathbf{P})\mathbf{X}^0\| + 2\sqrt{M}B\sum_{l=0}^k\alpha_l\lambda^{k-l}.
\end{aligned}$$

The result of Lemma 4 follows.

A.2.2 PROOF OF THEOREM 1

Without loss of generality, we assume that two random sets Ξ' and Ξ'' differ at one point in the first N samples of agent 1, and the iteration starts with samples selected from this agent. Then with probability, $1-\frac{1}{N}$, the sample selected by D-KM is the same in both Ξ' and Ξ''. Note that,

$$
\begin{aligned}
&x^{K+1}-y^{K+1}\\
&=\frac{1}{M}\sum_{m=1}^{M}\left\{\left[\sum_{l=1}^{M}\omega_{ml}x^K(l)-\alpha_K\left(T(x^K(m);\xi_{i_K(m)})-x^K(m)\right)\right]\right.\\
&\quad\left.-\left[\sum_{l=1}^{M}\omega_{ml}y^K(l)-\alpha_K\left(T(y^K(m);\xi_{i_K(m)})-y^K(m)\right)\right]\right\}\\
&=\frac{1}{M}\sum_{m=1}^{M}\left\{\left[x^K-\alpha_K\left(T(x^K;\xi_{i_K(m)})-x^K\right)\right]-\left[y^K-\alpha_K\left(T(y^K;\xi_{i_K(m)})-y^K\right)\right]\right\}+\varepsilon_K,
\end{aligned}
$$

where,

$$
\begin{aligned}
\|\varepsilon_K\|\leq&\frac{1}{M}\sum_{m=1}^{M}\{\sum_{l=1}^{M}\omega_{ml}\left\|x^K(l)-x^K\right\|\\
&+\alpha_K\|\left(T(x^K(m);\xi_{i_K(m)})-x^K(m)\right)-\left(T(x^K;\xi_{i_K(m)})-x^K\right)\|\}\\
&+\frac{1}{M}\sum_{m=1}^{M}\{\sum_{l=1}^{M}\omega_{ml}\left\|y^K(l)-y^K\right\|\\
&+\alpha_K\|\left(T(y^K(m);\xi_{i_K(m)})-y^K(m)\right)-\left(T(y^K;\xi_{i_K(m)})-y^K\right)\|\}\\
\leq&\frac{1}{M}\sum_{m=1}^{M}\left\{\sum_{l=1}^{M}\omega_{ml}\left\|x^K(l)-x^K\right\|+2\alpha_K\|x^K(m)-x^K\|\right\}\\
&+\frac{1}{M}\sum_{m=1}^{M}\left\{\sum_{l=1}^{M}\omega_{ml}\left\|y^K(l)-y^K\right\|+2\alpha_K\|y^K(m)-y^K\|\right\}\\
\leq&\frac{1}{M}\left\{\sum_{l=1}^{M}\left\|x^K(l)-x^K\right\|+2\alpha_K\sum_{m=1}^{M}\|x^K(m)-x^K\|\right\}\\
&+\frac{1}{M}\left\{\sum_{l=1}^{M}\left\|y^K(l)-y^K\right\|+2\alpha_K\sum_{m=1}^{M}\|y^K(m)-y^K\|\right\}\\
\leq&\frac{1}{M}\left\{(1+2\alpha_K)\sqrt{M}\left[\sum_{m=1}^{M}\|x^K(m)-x^K\|^2\right]^{\frac{1}{2}}\right\}\\
&+\frac{1}{M}\left\{(1+2\alpha_K)\sqrt{M}\left[\sum_{m=1}^{M}\|y^K(m)-y^K\|^2\right]^{\frac{1}{2}}\right\}\\
\leq&4(1+2\alpha_K)B\sum_{j=0}^{K-1}\alpha_j\lambda^{K-1-j}+2\lambda^{K+1}(1+2\alpha_K)\|(\mathbb{I}-\mathbf{P})\mathbf{X}^0\|/\sqrt{M}\\
=&4(1+2\alpha_K)B\sum_{j=0}^{K-1}\alpha_j\lambda^{K-1-j}.
\end{aligned}
\tag{A1}
$$

Due to the nonexpansiveness of the operator T, we have that,

$$
\begin{aligned}
&\left\|\frac{1}{M}\sum_{m=1}^{M}\left\{\left[x^K-\alpha_K\left(T(x^K;\xi_{i_K(m)})-x^K\right)\right]-\left[y^K-\alpha_t\left(T(y^K;\xi_{i_K})-y^K\right)\right]\right\}\right\|\\
&\leq\|x^K-y^K\|=\delta_K.
\end{aligned}
$$

That is,

$$\delta_{K+1} \leq \delta_K + 4(1+2\alpha_K)B\sum_{j=0}^{K-1}\alpha_j\lambda^{K-1-j}. \tag{A2}$$

On the other hand, with probability $\frac{1}{N}$, dKM selects the one sample to update in which Ξ' and Ξ'' differ.

$$\begin{aligned}
&x^{K+1}-y^{K+1}\\
=&\frac{1}{M}\sum_{m=2}^{M}\left\{\left[\sum_{l=1}^{M}\omega_{ml}x^K(l)-\alpha_K\left(T(x^K;\xi_{i_K(m)})-x^K\right)\right]\right.\\
&-\left[\sum_{l=1}^{M}\omega_{ml}y^K(l)-\alpha_K\left(T(y^K;\xi_{i_K(m)}-y^K\right)\right]\\
&+\frac{1}{M}\left\{\left[\sum_{l=1}^{M}\omega_{1l}x^K(l)-\alpha_K\left(T(x^K;\xi_{i_K(1)})-x^K\right)\right]\right.\\
&-\left.\left[\sum_{l=1}^{M}\omega_{1l}y^K(l)-\alpha_K\left(T(y^K;\xi'_{i_K(1)})-y^K\right)\right]+\Im_K\right.\\
=&\frac{1}{M}\sum_{m=2}^{M}\left\{\left[x^K-\alpha_K\left(T(x^K;\xi_{i_K(m)})-x^K\right)\right]-\left[y^K-\alpha_K\left(T(y^K;\xi_{i_K(m)})-y^K\right)\right]\right\}\\
&+\frac{1}{M}\left\{\left[x^K-\alpha_K\left(T(x^K;\xi_{i_K(1)})-x^K\right)\right]-\left[y^K-\alpha_K\left(T(y^K;\xi'_{i_K(1)})-y^K\right)\right]\right\}+\Im_K.
\end{aligned}$$

From the inequality A1, we have $\|\Im_K\| \leq 4\left(1+2\alpha_K\right)B\sum_{j=0}^{K-1}\alpha_j\lambda^{K-1-j}$. We conclude that

$$\|x^{K+1}-y^{K+1}\| \leq \frac{M-1}{M}\delta_K + \frac{\delta_K+2\alpha_K B}{M} + 4\left(1+2\alpha_K\right)B\sum_{j=0}^{K-1}\alpha_j\lambda^{K-1-j}. \tag{A3}$$

Inequalities A2, A3 imply that

$$\mathbb{E}\delta_{K+1} \leq \mathbb{E}\delta_K + \frac{2\alpha_t B}{MN} + 4\left(1+2\alpha_K\right)B\sum_{j=0}^{K-1}\alpha_j\lambda^{K-1-j},$$

which indicates that,

$$\mathbb{E}\delta_{K+1} \leq \frac{2B\sum_{k=0}^{K}\alpha_k}{MN} + 4B\sum_{k=0}^{K}(1+2\alpha_k)\sum_{j=0}^{k-1}\alpha_j\lambda^{k-1-j}.$$

A.2.3 PROOF OF THEOREM 2

Let Ξ^{ξ_0} be the sample set with size MN which differs Ξ at only point ξ_0. Note that

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right] = \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\mathop{\mathbb{E}}_{\xi\sim\Xi_m}\left[\ell\left(x_{\mathcal{B}(\Xi)},\xi\right)\right]\right]\right] \\
&= \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\Xi}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\ell\left(x_{\mathcal{B}(\Xi)},\xi_{m_{\mathcal{B}(\Xi)}}\right)\right]\right]\right] \\
&= \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\Xi}\left[\int_0^R \mathbb{P}\left(\ell\left(x_{\mathcal{B}(\Xi)},\xi_{m_{\mathcal{B}(\Xi)}}\right)>t\right)dt\right]\right] \\
&= \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\Xi}\left[\sum_{m=1}^{M}\int_0^R \mathbb{P}\left(\ell\left(x_{\mathcal{B}(\Xi)},\xi_m\right)>t, m_{\mathcal{B}(\Xi)}=m\right)dt\right]\right] \\
&\le \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\Xi,\xi^0\sim\mathbb{P}}\left[\sum_{m=1}^{M}\int_0^R \left(e^{\varepsilon}\mathbb{P}\left(\ell(x_{\mathcal{B}(\Xi^{\xi_0})},\xi_m)>t, m_{\mathcal{B}(\Xi)}=m\right)+\delta\right)\mathrm{d}t\right]\right] \\
&\le \sum_{m=1}^{M} e^{\varepsilon}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\Xi,\xi^0\sim\mathbb{P}}\left[\int_0^R \left(\mathbb{P}\left(\ell(x_{\mathcal{B}(\Xi^{\xi_0})},\xi_m)>t, m_{\mathcal{B}(\Xi)}=m\right)\right)\mathrm{d}t\right]\right]+M\delta \\
&\le \sum_{m=1}^{M} e^{\varepsilon}\mathop{\mathbb{E}}_{\Xi'\sim\mathbb{P}^{MN-1}}\left[\mathop{\mathbb{E}}_{\xi_m\sim\mathbb{P},\xi^0\sim\mathbb{P}}\left[\int_0^R \left(\mathbb{P}\left(\ell\left(x_{\mathcal{B}(\Xi'\cup\{\xi_0\})},\xi_m\right)>t, m_{\mathcal{B}(\Xi)}=m\right)\right)\mathrm{d}t\right]\right]+M\delta,
\end{aligned}
$$

where the first inequality holds because the algorithm $\mathcal{B}(\Xi)$ is (ε,δ)-differentially private. Let $\Xi = \Xi'\cup\{\xi^0\}$, and ξ_m,ξ_0 are i.i.d following the distribution $\mathbb{P}$. We have

$$
\begin{aligned}
&\sum_{m=1}^{M} e^{\varepsilon}\mathop{\mathbb{E}}_{\Xi'\sim\mathbb{P}^{MN-1}}\left[\mathop{\mathbb{E}}_{\xi_m\sim\mathbb{P},\xi^0\sim\mathbb{P}}\left[\int_0^R \left(\mathbb{P}\left(\ell\left(x_{\mathcal{B}(\Xi'\cup\{\xi_0\})},\xi_m\right)>t, m_{\mathcal{B}(\Xi)}=m\right)\right)\mathrm{d}t\right]\right]+M\delta \\
&= \sum_{m=1}^{M} e^{\varepsilon}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\mathbb{P}}\left[\int_0^R \left(\mathbb{P}\left(\ell\left(x_{\mathcal{B}(\Xi)},\xi\right)>t, m_{\mathcal{B}(\Xi)}=m\right)\right)\mathrm{d}t\right]\right]+M\delta \\
&= e^{\varepsilon}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\mathbb{P}}\left[\int_0^R \left(\mathbb{P}\left(\ell\left(x_{\mathcal{B}(\Xi)},\xi\right)>t\right)\right)\mathrm{d}t\right]\right]+M\delta \\
&= e^{\varepsilon}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\xi\sim\mathbb{P}}\left[\mathbb{E}^{\mathcal{B}(\Xi)}\left[\ell\left(x_{\mathcal{B}(S)},\xi\right)\right]\right]\right]+M\delta.
\end{aligned}
$$

Thus we obtain,

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right] &\le e^{\varepsilon}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[L(\mathcal{B}(\Xi))\right]\right]+M\delta, \\
-\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[L(\mathcal{B}(\Xi))\right]\right] &\le -e^{-\varepsilon}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right]+e^{-\varepsilon}M\delta.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right]-\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[L(\mathcal{B}(\Xi))\right]\right] \\
&\le e^{-\varepsilon}M\delta - e^{-\varepsilon}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right]+\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right] \\
&= (1-e^{-\varepsilon})\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right]+e^{-\varepsilon}M\delta \\
&\le (1-e^{-\varepsilon})R+e^{-\varepsilon}M\delta.
\end{aligned}
$$

The other side can be similarly obtained. Now we complete the proof.

A.2.4 PROOF OF THEOREM 3

Due to $\hat{L}(\mathcal{B}(\Xi)) \geq 0$, for any $\epsilon > 0$, we have,

$$\begin{aligned}\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[L(\mathcal{B}(\Xi))\right]\right] &\geq \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[L(\mathcal{B}(\Xi))\mathbb{I}_{\{L(\mathcal{B}(\Xi))\geq\hat{L}(\mathcal{B}(\Xi))+\epsilon\}}\right]\right]\\ &\geq \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[(\hat{L}(\mathcal{B}(\Xi))+\epsilon)\mathbb{I}_{\{L(\mathcal{B}(\Xi))\geq\hat{L}(\mathcal{B}(\Xi))+\epsilon\}}\right]\right].\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[L(\mathcal{B}(\Xi))\right]\right] - \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right]\\ &\geq \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[(\hat{L}(\mathcal{B}(\Xi))+\epsilon)\mathbb{I}_{\{L(\mathcal{B}(\Xi))\geq\hat{L}(\mathcal{B}(\Xi))+\epsilon\}}\right]\right]\\ &\quad - \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\mathbb{I}_{\{L(\mathcal{B}(\Xi))\geq\hat{L}(\mathcal{B}(\Xi))+\epsilon\}}\right]\right]\\ &\quad - \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\mathbb{I}_{\{L(\mathcal{B}(\Xi))<\hat{L}(\mathcal{B}(\Xi))+\epsilon\}}\right]\right]\\ &= \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\epsilon\mathbb{I}_{\{L(\mathcal{B}(\Xi))\geq\hat{L}(\mathcal{B}(\Xi))+\epsilon\}}\right]\right]\\ &\quad - \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\mathbb{I}_{\{L(\mathcal{B}(\Xi))<\hat{L}(\mathcal{B}(\Xi))+\epsilon\}}\right]\right]\\ &\geq \epsilon\cdot\mathbb{P}(L(\mathcal{B}(\Xi))\geq\hat{L}(\mathcal{B}(\Xi))+\epsilon) - R\cdot\mathbb{P}(L(\mathcal{B}(\Xi))\leq\hat{L}(\mathcal{B}(\Xi))+\epsilon)\\ &= \epsilon - (\epsilon+R)\cdot\mathbb{P}(L(\mathcal{B}(\Xi))\leq\hat{L}(\mathcal{B}(\Xi))+\epsilon).\end{aligned} \tag{A4}$$

Note that Theorem 2 indicates

$$\mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[L(\mathcal{B}(\Xi))\right]\right] - \mathop{\mathbb{E}}_{\Xi\sim\mathbb{P}^{MN}}\left[\mathop{\mathbb{E}}_{\mathcal{B}(\Xi)}\left[\hat{L}(\mathcal{B}(\Xi))\right]\right] \leq (1-e^{-\varepsilon})R + e^{-\varepsilon}M\delta.$$

This result combined with equation A4 implies that

$$\mathbb{P}(L(\mathcal{B}(\Xi))\leq\hat{L}(\mathcal{B}(\Xi))+\epsilon) \geq \frac{\epsilon-(1-e^{-\varepsilon})R-e^{-\varepsilon}M\delta}{\epsilon+R}.$$

A.2.5 PROOF OF THEOREM 4

For any subsets $H, H_1, \cdots, H_M$, with Definition 3, we have that,

$$\begin{aligned}&\log\frac{\mathbb{P}_{\Xi}\left(\{\tilde{X}(m)\}_{m=1}^{M}\in H\right)}{\mathbb{P}_{\Xi'}\left(\{\tilde{X}(m)\}_{m=1}^{M}\in H\right)} = \log\prod_{m=1}^{M}\frac{\mathbb{P}_{\Xi}\left(\tilde{X}(m)\in H_m\right)}{\mathbb{P}_{\Xi'}\left(\tilde{X}(m)\in H_m\right)}\\ &= \sum_{m=1}^{M}\log\frac{\mathbb{P}_{\Xi}\left(\tilde{X}(m)\in H_m\right)}{\mathbb{P}_{\Xi'}\left(\tilde{X}(m\in H_m)\right)} = \sum_{m=1}^{M}\log\frac{\mathbb{P}_{\Xi}\left(\tilde{\mathcal{A}}_m(\Xi_m)\in H_m\mid\tilde{X}^0(m)\right)}{\mathbb{P}_{\Xi'}\left(\tilde{\mathcal{A}}_m(\Xi'_m)\in H_m\mid\tilde{X}^0(m)\right)}\end{aligned}$$

And, we have,

$$\begin{aligned}&\mathbb{E}_{\Xi}\left\{\log\left(\frac{\mathbb{P}(\tilde{\mathcal{A}}_m(\Xi)\in H_m)}{\mathbb{P}(\tilde{\mathcal{A}}_m(\Xi')\in H_m)}\right)\mid\tilde{X}^0(m)\right\} \leq D_{KL}\left(\tilde{\mathcal{A}}(\Xi))\|\tilde{\mathcal{A}}(\Xi'))\right)\\ &\leq \min\left\{\min\left\{2, e^{\varepsilon_m}-1\right\}\varepsilon_m, \varepsilon_m\right\} := C_{KL}(m).\end{aligned}$$

Based on Lemma 2, we have that,

$$\mathbb{P}\left(\sum_{m=1}^{M}\log\left(\frac{\mathbb{P}\left(\tilde{\mathcal{A}}_m(\Xi)\in H_m\right)}{\mathbb{P}\left(\tilde{\mathcal{A}}_m(\Xi')\in H_m\right)}\right) \geq \sum_{m=1}^{M}C_{kL}(m) + t\cdot\sqrt{\sum_{m=1}^{M}\varepsilon_m^2}\right) \leq e^{-\frac{t^2}{2}}.$$

Let $\tilde{\delta} = e^{-\frac{t^2}{2}}$, that is, $t = \sqrt{2\log\frac{1}{\delta'}}, \forall \tilde{\delta} > 0$. It follows

$$\mathbb{P}\left(\sum_{m=1}^{M} \log\left(\frac{\mathbb{P}\left(\tilde{\mathcal{A}}_m(\Xi) \in H_m\right)}{\mathbb{P}\left(\tilde{\mathcal{A}}_m(\Xi') \in H_m\right)}\right) \geq \sum_{m=1}^{M} C_{kL}(m) + \sqrt{2\log\frac{1}{\tilde{\delta}}} \cdot \sqrt{\sum_{m=1}^{M} \varepsilon_m^2}\right) \leq \tilde{\delta}.$$

Therefore, $\{\tilde{X}(m)\}_{m=1}^{M}$ is $(\varepsilon', \tilde{\delta})$-pDP with $\varepsilon' = \sum_{m=1}^{M} C_{kL}(m) + \sqrt{\left(2\log\frac{1}{\tilde{\delta}}\right)\left(\sum_{m=1}^{M}\varepsilon_m^2\right)}$.

Since $\left\{\tilde{x}^k(m)\right\}_{k=0}^{K}$ is $(\varepsilon_m, \delta_m)$-DP, for each $m = 1, \cdots, M$. Based on Lemma 3, we have that there exists r.v.s. $\tilde{Y}_\Xi(m)$, $\tilde{Z}_{\Xi'}(m)$, such that,

$$\Delta\left(\tilde{X}_\Xi(m) \| \tilde{Y}_\Xi(m)\right) \leq \frac{\delta_m}{e^{\varepsilon_m}+1}, \quad \Delta\left(\tilde{X}_{\Xi'}(m) \| \tilde{Z}_{\Xi'}(m)\right) \leq \frac{\delta_m}{e^{\varepsilon_m}+1}$$

$$D_\infty\left(\tilde{Y}_\Xi(m) \| \tilde{Z}_{\Xi'}(m)\right) \leq \varepsilon_m, \quad D_\infty\left(\tilde{Z}_{\Xi'}(m) \| \tilde{Y}_\Xi(m)\right) \leq \varepsilon_m.$$

Since $\{\tilde{X}(m)\}_{m=1}^{M}$ is ε'-pDP, we need to prove that,

$$\prod_{m=1}^{M} \mathbb{P}\left(\tilde{X}_\Xi(m) \in H_m\right) \leq e^{\varepsilon'} \prod_{m=1}^{M} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right) + \delta'.$$

Note that,

$$\mathbb{P}\left(\tilde{Y}_\Xi(m) \in H_i\right) \leq \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)$$

which indicates that,

$$\prod_{m=1}^{M} \mathbb{P}\left(\tilde{Y}_\Xi(m) \in H_i\right) \leq \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right).$$

Case 1: $\prod_{m=1}^{M}\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m)\in H_m\right)}\right\} \leq e^{\varepsilon'}$

$$\begin{aligned}
&\prod_{m=1}^{M} \mathbb{P}\left(\tilde{Y}_\Xi(m) \in H_m\right) \\
&\leq \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right) \\
&+ \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right) \\
&- \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \left\{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right) - \frac{\delta_m}{1+e^{\varepsilon_m}}\right\}.
\end{aligned}$$

Because

$$\left\{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right) - \frac{\delta_m}{1+e^{\varepsilon_m}}\right\} \leq \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right)$$

$$\prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right)$$
$$- \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \left\{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right) - \frac{\delta_m}{1+e^{\varepsilon_m}}\right\}$$
$$\leq 1 - \prod_{m=1}^{M} \left\{1 - \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \frac{\delta_m}{1+e^{\varepsilon_m}}\right\} := \Delta_1.$$

We have that,

$$\prod_{m=1}^{M} \mathbb{P}\left(\tilde{Y}_{\Xi}(m) \in H_m\right) \leq \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right) + \Delta_1.$$

Combing the inequality $\mathbb{P}\left(\tilde{Y}_{\Xi}(m) \in H_m\right) \geq \mathbb{P}\left(\tilde{X}_{\Xi}(m) \in H_m\right) - \frac{\delta_m}{1+e^{\varepsilon_m}}$. We have,

$$\prod_{m=1}^{M} \mathbb{P}\left(\tilde{X}_{\Xi}(m) \in H_m\right)$$
$$\leq \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right) + \Delta_1 + \prod_{m=1}^{M} \frac{\delta_m}{1+e^{\varepsilon_m}}$$
$$\leq e^{\varepsilon'} \prod_{m=1}^{M} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right) + \Delta_1$$
$$+ \prod_{m=1}^{M} \mathbb{P}\left(\tilde{X}_{\Xi}(m) \in H_m\right) - \prod_{m=1}^{M} \left\{\mathbb{P}\left(\tilde{X}_{\Xi}(m) \in H_m\right) - \frac{\delta_m}{1+e^{\varepsilon_m}}\right\}$$
$$\leq e^{\varepsilon'} \prod_{m=1}^{M} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right) + \Delta_1 + 1 - \prod_{m=1}^{M} \left(1 - \frac{\delta_m}{1+e^{\varepsilon_m}}\right).$$

That is, $\delta' = 1 - \prod_{m=1}^{M} \left(1 - \frac{\delta_m}{1+e^{\varepsilon_m}}\right) + \Delta_1$.

Case 2: $\prod_{m=1}^{M} \left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} > e^{\varepsilon'}$

There exists a sequence of real numbers $\{a_m\}_{m=1}^{M}$, such that,

$$e^{a_m} \leq \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\}, \quad \sum_{m=1}^{M} a_m = \varepsilon'.$$

Note that

$$\prod_{m=1}^{M} \mathbb{P}\left(\tilde{Y}_{\Xi}(m) \in H_m\right)$$
$$\leq \prod_{m=1}^{M} \min\left\{e^{\varepsilon_m}, \frac{1}{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)}\right\} \mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right)$$
$$+ \prod_{m=1}^{M} e^{a_m} \mathbb{P}\left(\tilde{X}_{\Xi'}(m) \in H_m\right) - \prod_{m=1}^{M} e^{a_m} \left\{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m) \in H_m\right) - \frac{\delta_m}{1+e^{\varepsilon_m}}\right\}.$$

On the other hand,

$$\prod_{m=1}^{M} e^{a_m}\mathbb{P}\left(\tilde{Z}_{\Xi'}(m)\in H_M\right)-\prod_{m=1}^{M} e^{a_m}\left\{\mathbb{P}\left(\tilde{Z}_{\Xi'}(m)\in H_m\right)-\frac{\delta_m}{1+e^{\varepsilon_m}}\right\}$$
$$\leq 1-\prod_{m=1}^{M}\left\{1-e^{a_m}\frac{\delta_m}{1+e^{\varepsilon_m}}\right\}:=\Delta_2.$$

which indicates,

$$\prod_{m=1}^{M}\mathbb{P}\left(\tilde{Y}_{\Xi}(m)\in H_m\right)\leq\prod_{m=1}^{M} e^{a_m}\mathbb{P}\left(\tilde{X}_{\Xi'}(m)\in H_m\right)+\Delta_2.$$

In further,

$$\prod_{m=1}^{M}\mathbb{P}\left(\tilde{Y}_{\Xi}(m)\in H_m\right)\geq\prod_{m=1}^{M}\left\{\mathbb{P}\left(\tilde{X}_{\Xi}(m)\in H_m\right)-\frac{\delta_m}{1+e^{\varepsilon_m}}\right\}$$

$$\prod_{m=1}^{M}\left\{\mathbb{P}\left(\tilde{X}_{\Xi}(m)\in H_m\right)-\frac{\delta_m}{1+e^{\varepsilon_m}}\right\}\leq\prod_{m=1}^{M} e^{a_m}\mathbb{P}\left(\tilde{X}_{\Xi'}(m)\in H_m\right)+\Delta_2.$$

In general,

$$\begin{aligned}
&\prod_{m=1}^{M}\mathbb{P}\left(\tilde{X}_{\Xi}(m)\in H_i\right)\\
\leq&\prod_{m=1}^{M} e^{a_m}\mathbb{P}\left(\tilde{X}_{\Xi'}(m)\in H_m\right)+\Delta_2\\
&-\left\{\prod_{m=1}^{M}\left[\mathbb{P}\left(\tilde{X}_{\Xi}(m)\in H_m\right)-\frac{\delta_m}{1+e^{\varepsilon_m}}\right]\right\}+\prod_{m=1}^{M}\mathbb{P}\left(\tilde{X}_{\Xi}(m)\in H_m\right)\\
\leq& e^{\varepsilon'}\prod_{m=1}^{M}\mathbb{P}\left(\tilde{X}_{\Xi'}(m)\in H_m\right)+\Delta_2+\left\{1-\prod_{m=1}^{M}(1-\frac{\delta_m}{1+e^{\varepsilon_i}})\right\}
\end{aligned}$$

Let $\delta'=1-\left\{\prod_{m=1}^{M}(1-e^{a_m}\frac{\delta_m}{1+e^{\varepsilon_m}})\right\}+\left\{1-\prod_{m=1}^{M}(1-\frac{\delta_m}{1+e^{\varepsilon_m}})\right\}$.
Thus, we have $\varepsilon'=\min\{\varepsilon_1,\varepsilon_2,\varepsilon_3\}$, where,

$$\varepsilon_1=\sum_{m=1}^{M}\varepsilon_m,\quad \varepsilon_2=\sum_{m=1}^{M}C_{KL}(m)+\sqrt{2\log(\frac{1}{\delta'})(\sum_{m=1}^{M}\varepsilon_m^2)},$$
$$\varepsilon_3=\sum_{m=1}^{M}\frac{(e^{\varepsilon_m}-1)\varepsilon_m}{e^{\varepsilon_m}+1}+\sqrt{\sum_{m=1}^{M}2\varepsilon_m^2\log\left(e+\frac{\sqrt{\sum_{m=1}^{M}\varepsilon_m^2}}{\tilde{\delta}}\right)}.$$

A.2.6 PROOF OF THEOREM 5

Denote $x_m^k := x^k(m)$, $x^{v,k}$ as the v-th element of $x^k \in \mathbb{R}^p$. We know that

$$\begin{aligned}
&\left\|\mathbf{W}\left(\mathbf{X}^k - \widetilde{\mathbf{X}}^k\right)\right\|_F^2 = \sum_{m=1}^{M}\left\|\sum_{l=1}^{M} w_{ml}\left(x_l^k - \widetilde{x}_l^k\right)\right\|^2 = \sum_{v=1}^{p}\sum_{m=1}^{M}\left\|\sum_{l=1}^{M} w_{ml}\left(x_l^{v,k} - \widetilde{x}_l^{v,k}\right)\right\|^2 \\
&= \sum_{v=1}^{p}\sum_{m=1}^{M}\sigma_{k,m}^2\left(\sum_{l=1}^{M} w_{ml}^2\right)\left\{\frac{\sum_{l=1}^{M} w_{ml}\left[\left(x_l^{v,k} - \tilde{x}_l^{v,k}\right) - \mu_{k,m}^v\right] + \sum_{l=1}^{M} w_{ml}\mu_{k,l}^v}{\sigma_{k,m}\sqrt{\sum_{l=1}^{M} w_{ml}^2}}\right\}^2 \\
&\leq \sum_{v=1}^{p}\sum_{m=1}^{M}\sigma_{k,m}^2\left(\sum_{l=1}^{M} w_{ml}^2\right)\left\{\frac{\sum_{l=1}^{M} w_{ml}\left[\left(x_l^{v,k} - \widetilde{x}_l^{v,k}\right) - \mu_{km}^v\right]}{\sigma_{k,m}\sqrt{\sum_{l=1}^{M} w_{ml}^2}}\right\}^2 + \mu^2 pM \cdot \sum_{m=1}^{M}\sum_{l=1}^{M} w_{ml}^2 \\
&+ \sum_{v=1}^{p}\sum_{m=1}^{M}\left(\sum_{l=1}^{M} w_{ml}^2\right)\frac{\sum_{l_1=1}^{M}\sum_{l_2=1}^{M} w_{ml_1}w_{ml_2}\mu_{k,l_2}^v\left[\left(x_{l_1}^{v,k} - \tilde{x}_{l_1}^{v,k}\right) - \mu_{k,l_1}^v\right]}{\sqrt{\sum_{l_1=1}^{M} w_{ml_1}^2}\sqrt{\sum_{l_2=1}^{M} w_{ml_2}^2}}.
\end{aligned}$$

The last inequality holds since $\sum_{v=1}^{p}\left(\mu_{k,l}^v\right)^2 \leq p\mu^2$. Considering that the weight difference is normally distributed,

$$\left(x_m^k - \tilde{x}_m^k\right) \sim \mathcal{N}\left(\mu_{k,m}, \sigma_{k,m}^2 I_p\right), \quad m = 1, \ldots, M,$$

with $\mu_{k,m}$ satisfying $\|\mu_{k,m}\|_2^2 \leq p\mu^2$ and $\sigma_{k,m}^2 \in \mathbb{R}$ being bounded by σ^2, we obtain,

$$\frac{\sum_{l=1}^{M} w_{ml}\left[\left(x_l^{v,k} - \tilde{x}_l^{v,k}\right) - \mu_{k,m}^v\right]}{\sigma_{k,m}\sqrt{\sum_{l=1}^{M} w_{ml}^2}} \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$$

It follows that

$$\left\{\frac{\sum_{l=1}^{M} w_{m,l}\left[\left(x_l^{v,k} - \widetilde{x}_l^{v,k}\right) - \mu_{k,m}^v\right]}{\sigma_{k,m}\sqrt{\sum_{l=1}^{M} w_{ml}^2}}\right\}^2 \sim \mathcal{X}^2(1), \quad l = 1, \ldots, M.$$

Further, $\forall l, \mathbb{E}\left[\left(x_l^{v,k} - \widetilde{x}_l^{v,k}\right) - \mu_{k,l}^v\right] = 0$, therefore, we have,

$$\mathbb{E}\left[\sum_{v=1}^{p}\sum_{m=1}^{M}\left(\sum_{l=1}^{M} w_{ml}^2\right)\frac{\sum_{l_1=1}^{M}\sum_{l_2=1}^{M} w_{ml_1}w_{ml_2}\mu_{k,l_2}^v\left[\left(x_{l_1}^{v,k} - \widetilde{x}_{l_1}^{v,k}\right) - \mu_{k,l_1}^v\right]}{\sqrt{\sum_{l_1=1}^{M} w_{ml_1}^2}\sqrt{\sum_{l_2=1}^{M} w_{ml_2}^2}}\right] = 0,$$

As a consequence,

$$\mathbb{E}\left[\left\|\mathbf{W}\left(\mathbf{X}^k - \widetilde{\mathbf{X}}^k\right)\right\|_F^2 \leq p\sigma^2\sum_{m=1}^{M}\lambda_m^2 + p\mu^2\sum_{m=1}^{M}\lambda_m^2 \leq p\left(\sigma^2 + \mu^2\right)\left[1 + (M-1)\lambda^2\right].\right.$$

A.2.7 PROOF OF THEOREM 6

We discuss the cases by the value of z. Note that, for any $S \in \mathbb{R}$, $z \in \left(-\infty, -\frac{\Delta}{2}\right), \mathbb{P}_z(z) \leq \mathbb{P}_z(z+v)$. Therefore, $\mathbb{P}_z(S) - e^{\varepsilon}\mathbb{P}_z(S+v) \leq 0 \leq \delta$.

If $z \in \left[-\frac{\Delta}{2}, 0\right), z + v \in \left[-\frac{\Delta}{2}, 0\right)$, then $0 \leq v < \frac{\Delta}{2}$,

$$\frac{p_z(z)}{p_z(z+v)} = e^{\frac{v^2+2vz}{\sigma^2}} \leq 1 \leq e^{\varepsilon}.$$

Therefore, $\mathbb{P}_z(S) - e^{\varepsilon}\mathbb{P}_z(S+v) = \int_S \left(p_z(z) - e^{\varepsilon}p_z(z+v)\right)dz \leq 0 \leq \delta$.

If $z \in \left[-\frac{\Delta}{2}, 0\right), z+v \in [0, \Delta)$, $e^{\frac{v^2+2vz}{\sigma^2}} \leq e^{\frac{\Delta^2}{\sigma^2}} \leq e^{\varepsilon}$. Therefore, $\mathbb{P}_z(S) - e^{\varepsilon}\mathbb{P}_z(S+v) \leq \delta$.

For any $z \in [0, A-\Delta]$, we still have $\frac{p_z(z)}{p_z(z+v)} \leq \frac{p_z(z)}{p_z(z+\Delta)} \leq e^{\varepsilon}$.

For any $z \in [A-\Delta, +\infty)$,

$$\begin{aligned}
\mathbb{P}_z(S) - e^{\varepsilon}\mathbb{P}_z(S+v) &\leq \mathbb{P}_z(A-\Delta) - e^{\varepsilon}\mathbb{P}_z(A-\Delta+v) \\
&= \int_{A-\Delta}^{A-\Delta+v} p_z(z)dz + \int_{A-\Delta+v}^{+\infty} p_z(z)dz - e^{\varepsilon}\int_{A-\Delta+v}^{+\infty} p_z(z)dz \\
&= \int_{A-\Delta}^{A-\Delta+v} p_z(z)dz + \int_{A-\Delta+v}^{+\infty} (1-e^{\varepsilon})\, p_z(z)dz \\
&\leq \int_{A-\Delta}^{A-\Delta+v} p_z(z)dz \leq \int_{A-\Delta}^{A} p_z(z)dz = \delta.
\end{aligned}$$

Based on the above inequality, for any $S \in \mathbb{R}, |d| \leq \Delta, \mathbb{P}_z(S) - e^{\varepsilon}\mathbb{P}_z(S+d) \leq \delta$, we have the conclusion of the theorem.

A.3 Truncated Generalized Gaussian Mechanisms

In this section, we give more details to determine an appropriate bounding parameter in noise addition. Given the privacy budget ε and δ, we are interested in deriving the minimum amount of noise added to shared parameters to achieve the highest utility while preserving differential privacy. Motivated by the observation that under (ε, δ)-differential privacy, the decay rate defined as $p(z)/p(z+\Delta)$ shall be as high as possible without exceeding e^{ε}, except for a set of points with a probability mass δ [Mironov (2017)]. We then truncate a probability density function (pdf) with an appropriate normalization factor to restore the integral over the truncated span to be 1 and calculate an appropriate scaling parameter. The required scaling parameter depends upon the privacy parameter (ε, δ). Specifically,

- The probability mass δ is equal to the area under the pdf in the last interval with length Δ over the support of pdf, i.e., the interval $[A-\Delta, A]$.
- The decay rate $\frac{p(z)}{p(z+\Delta)}$ is exactly e^{ε} for $z \in [0, A-\Delta)$.

The parameter A is then derived by solving the equations $\int_{-A}^{A} p(z)dz = 1$ and $\int_{A-\Delta}^{A} p(z)dz = \delta$.

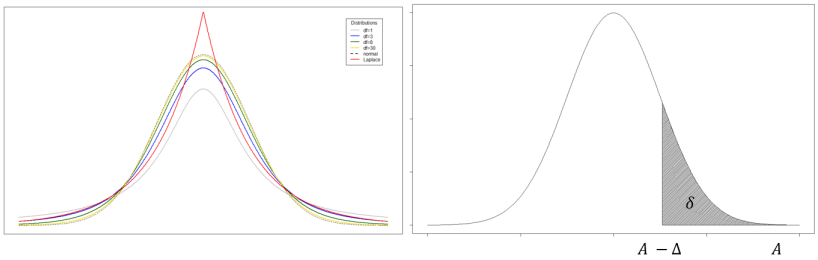

Figure 5: Noise probability density function.

We address that the truncated GG mechanisms can be extended to the case $b \geq 3$ in practice to preserve a differential privacy guarantee. The major challenge is to calculate the boundary parameter, as indicated by A, given the pre-specific parameters, such as Δ, σ, and ε. As it requires solving multiple equations or using the approximation methods thereof that could limit its utility in the real application. This paper thus limits the discussion to the case $b = 1, 2$ which is relatively more applicable in practice.