Lemma 1. *For any $\boldsymbol{\mu} \in \mathbb{R}_+^M$, we have*

$$\begin{aligned}&\sum_{t=1}^{T} \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) \leq \\ &\frac{\|\boldsymbol{\mu}\|_2^2}{2}\left(\frac{1}{\beta_1} - \delta_1\right) + MNA^2G^2 J_T(\boldsymbol{\beta}) + \sum_{t=1}^{T}\beta_t\delta_t^2\|\boldsymbol{\mu}_t\|_2^2,\end{aligned} \tag{25}$$

where $J_T(\boldsymbol{\beta}) \triangleq \sum_{t=1}^{T}\beta_t$.

Proof of Lemma 1. Note that $\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu})$ is δ_t-strongly concave in $\boldsymbol{\mu}$. This implies that

$$\begin{aligned}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) \leq &\langle \boldsymbol{\mu} - \boldsymbol{\mu}_t, \nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\rangle \\ &- \frac{\delta_t}{2}\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2,\end{aligned} \tag{26}$$

The updating rule of $\boldsymbol{\mu}_t$ leads to

$$\begin{aligned}\|\boldsymbol{\mu} - \boldsymbol{\mu}_{t+1}\|_2^2 \leq &\left\|\boldsymbol{\mu} - \boldsymbol{\mu}_t - \beta_t\nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\right\|_2^2 \\ =&\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2 + \beta_t^2\|\nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\|_2^2 \\ &- 2\beta_t\langle \boldsymbol{\mu} - \boldsymbol{\mu}_t, \nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\rangle,\end{aligned} \tag{27}$$

which leads to

$$\begin{aligned}&\langle \boldsymbol{\mu} - \boldsymbol{\mu}_t, \nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\rangle \leq \\ &\frac{\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2 - \|\boldsymbol{\mu} - \boldsymbol{\mu}_{t+1}\|_2^2}{2\beta_t} + \frac{\beta_t\|\nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\|_2^2}{2}\end{aligned} \tag{28}$$

Finally, we obtain

$$\begin{aligned}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) \leq &\frac{\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2 - \|\boldsymbol{\mu} - \boldsymbol{\mu}_{t+1}\|_2^2}{2\beta_t} \\ + \frac{\beta_t\|\nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\|_2^2}{2} - &\frac{\delta_t}{2}\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2.\end{aligned} \tag{29}$$

Summing over $t \in \{1, 2, ..., T\}$ leads to

$$\begin{aligned}&\sum_{t=1}^{T} \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) \leq \sum_{t=1}^{T}\frac{\beta_t\|\nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\boldsymbol{x}_t,\boldsymbol{\mu}_t)\|_2^2}{2} \\ &+ \sum_{t=1}^{T}\left(\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_t\|_2^2 - \|\boldsymbol{\mu}-\boldsymbol{\mu}_{t+1}\|_2^2}{2\beta_t} - \frac{\delta_t}{2}\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2\right)\end{aligned} \tag{30}$$

Next we upper bound the above formula in two steps. First, we have

$$\begin{aligned}\sum_{t=1}^{T}\frac{\beta_t\|\nabla_{\boldsymbol{\mu}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\|_2^2}{2} \leq &\sum_{t=1}^{T}\beta_t\left(MNA^2G^2 + \delta_t^2\|\boldsymbol{\mu}_t\|_2^2\right) \\ = &MNA^2G^2J_T(\boldsymbol{\beta}) + \sum_{t=1}^{T}\beta_t\delta_t^2\|\boldsymbol{\mu}_t\|_2^2,\end{aligned} \tag{31}$$

where the inequality follows from $(a_1 + a_2 + ... + a_n)^2 \leq n(a_1^2 + a_2^2 + ... + a_n^2)$.

Second, we have

$$\begin{aligned}&\sum_{t=1}^{T}\left(\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_t\|_2^2 - \|\boldsymbol{\mu}-\boldsymbol{\mu}_{t+1}\|_2^2}{2\beta_t} - \frac{\delta_t}{2}\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2\right) \\ =&\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_1\|_2^2}{2\beta_1} - \frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_{T+1}\|_2^2}{2\beta_T} + \sum_{t=2}^{T}\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_t\|_2^2}{2}\left(\frac{1}{\beta_t} - \frac{1}{\beta_{t-1}}\right) \\ &- \sum_{t=1}^{T}\frac{\delta_t}{2}\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_2^2 \\ \overset{(a)}{\leq}&\frac{\|\boldsymbol{\mu}\|_2^2}{2}\left(\frac{1}{\beta_1} - \delta_1\right) + \sum_{t=2}^{T}\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_t\|_2^2}{2}\left(\frac{1}{\beta_t} - \frac{1}{\beta_{t-1}} - \delta_t\right) \\ \overset{(b)}{\leq}&\frac{\|\boldsymbol{\mu}\|_2^2}{2}\left(\frac{1}{\beta_1} - \delta_1\right)\end{aligned} \tag{32}$$

where the inequality (a) follows from $\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}_{T+1}\|_2^2}{2\beta_T} \geq 0$ and $\boldsymbol{\mu}_1 = \boldsymbol{0}_M$. The inequality (b) follows from (16b).

This completes the proof. □

Lemma 2. *For any feasible allocation $\boldsymbol{x} \in \mathcal{X}$, we have*

$$\begin{aligned}&\sum_{t=1}^{T}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) - \mathcal{L}_t(\boldsymbol{x};\boldsymbol{\mu}_t) \leq \\ &A\left[\alpha_T\ln(N) + \frac{(1+M)G^2}{2}\left(J_T(\boldsymbol{\alpha}) + \sum_{t=1}^{T}\frac{\|\boldsymbol{\mu}_t\|_2^2}{\alpha_t}\right)\right]\end{aligned} \tag{33}$$

where $J_T(\boldsymbol{\alpha}) \triangleq \sum_{t=1}^{T}\frac{1}{\alpha_t}$.

Proof of Lemma 2. Note that $\mathcal{L}_t(\cdot;\boldsymbol{\mu}_t)$ is convex in $\boldsymbol{x}$. This implies that

$$\sum_{t=1}^{T}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) - \mathcal{L}_t(\boldsymbol{x};\boldsymbol{\mu}_t) \leq \sum_{t=1}^{T}\langle\hat{\boldsymbol{x}}_t - \boldsymbol{x}, \nabla_{\boldsymbol{x}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\rangle. \tag{34}$$

Next we will derive an upper bound for the RHS in (34) based on the potential function $\Phi_t(\alpha)$, i.e.,

$$\Phi_t(\alpha) \triangleq \alpha\ln\left(\frac{1}{N}\sum_{n=1}^{N}\exp\left(-\frac{1}{\alpha}\sum_{s=1}^{t}\frac{\partial\mathcal{L}_s(\hat{\boldsymbol{x}}_s;\boldsymbol{\mu}_s)}{\partial x_n}\right)\right). \tag{35}$$

For any $t > 1$, we have

$$\begin{aligned}&\Phi_t(\alpha_t) - \Phi_{t-1}(\alpha_t) \\ =&\alpha_t\ln\left(\frac{\sum_{n=1}^{N}\exp\left(-\frac{1}{\alpha_t}\sum_{s=1}^{t}\frac{\partial\mathcal{L}_s(\hat{\boldsymbol{x}}_s;\boldsymbol{\mu}_s)}{\partial x_n}\right)}{\sum_{n=1}^{N}\exp\left(-\frac{1}{\alpha_t}\sum_{s=1}^{t-1}\frac{\partial\mathcal{L}_s(\hat{\boldsymbol{x}}_s;\boldsymbol{\mu}_s)}{\partial x_n}\right)}\right) \\ \overset{(a)}{=}&\alpha_t\ln\left(\sum_{n=1}^{N}\frac{\hat{x}_{t,n}}{A}\exp\left(-\frac{1}{\alpha_t}\frac{\partial\mathcal{L}(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)}{\partial x_n}\right)\right) \\ \overset{(b)}{\leq}&-\frac{\langle\hat{\boldsymbol{x}}_t, \nabla_{\boldsymbol{x}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t,\boldsymbol{\mu}_t)\rangle}{A} + \frac{(1 + \|\boldsymbol{\mu}_t\|_1)^2G^2}{2\alpha_t} \\ \overset{(c)}{\leq}&-\frac{\langle\hat{\boldsymbol{x}}_t, \nabla_{\boldsymbol{x}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t,\boldsymbol{\mu}_t)\rangle}{A} + \frac{(1 + M)(1 + \|\boldsymbol{\mu}_t\|_2^2)G^2}{2\alpha_t},\end{aligned} \tag{36}$$

where the equality (a) follows from (11) and the weight vector's updating rule (14). The inequality (b) follows from Hoeffding's inequality and $\|[\nabla_{\boldsymbol{x}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)]_{(n)}\| \leq (1+\|\boldsymbol{\mu}_t\|_1)G$. The inequality (c) follows from $(a_1 + a_2 + ... + a_k)^2 \leq k(a_1^2 + a_2^2 + ... + a_k^2)$.

This leads to

$$\begin{aligned}\frac{\langle\hat{\boldsymbol{x}}_t, \nabla_{\boldsymbol{x}}\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t)\rangle}{A} \leq &\frac{(1 + M)(1 + \|\boldsymbol{\mu}_t\|_2^2)G^2}{2\alpha_t} \\ &+ \Phi_{t-1}(\alpha_t) - \Phi_t(\alpha_t),\end{aligned} \tag{37}$$

Summing over $t \in \{1, 2, ..., T\}$ leads to

$$\begin{aligned}
&\sum_{t=1}^{T} \frac{\langle \hat{\boldsymbol{x}}_t, \nabla_{\boldsymbol{x}} \mathcal{L}_t(\hat{\boldsymbol{x}}_t; \boldsymbol{\mu}_t) \rangle}{A} \\
&\leq \sum_{t=1}^{T} [\Phi_{t-1}(\alpha_t) - \Phi_t(\alpha_t)] + \sum_{t=1}^{T} \frac{(1+M)(1+\|\boldsymbol{\mu}_t\|_2^2)G^2}{2\alpha_t} \\
&\overset{(a)}{=} \Phi_0(\alpha_1) - \Phi_T(\alpha_T) + \sum_{t=1}^{T-1} [\Phi_t(\alpha_{t+1}) - \Phi_t(\alpha_t)] \\
&\quad + \sum_{t=1}^{T} \frac{(1+M)(1+\|\boldsymbol{\mu}_t\|_2^2)G^2}{2\alpha_t} \\
&\overset{(b)}{\leq} \Phi_0(\alpha_1) - \Phi_T(\alpha_T) + \sum_{t=1}^{T} \frac{(1+M)(1+\|\boldsymbol{\mu}_t\|_2^2)G^2}{2\alpha_t} \\
&\overset{(c)}{=} -\Phi_T(\alpha_T) + \sum_{t=1}^{T} \frac{(1+M)(1+\|\boldsymbol{\mu}_t\|_2^2)G^2}{2\alpha_t},
\end{aligned} \tag{38}$$

where the equality (a) follows from Abel transformation. The inequality (b) follows from that $\Phi_t(\alpha)$ is decreasing in α and the sequence $\{\alpha_t : \forall t \in \mathcal{T}\}$ is increasing in t according to (13). The equality (c) follows from $\Phi_0(\alpha_1) = 0$.

Next we derive a lower bound for $\Phi_T(\alpha_T)$.

$$\begin{aligned}
\Phi_T(\alpha_T) &= \alpha_T \ln\left(\frac{1}{N}\sum_{n=1}^{N} \exp\left(-\frac{1}{\alpha_T}\sum_{t=1}^{T} [\nabla_{\boldsymbol{x}} \mathcal{L}_t(\boldsymbol{x}_t; \boldsymbol{\mu}_t)]_{(n)}\right)\right) \\
&\overset{(a)}{\geq} \alpha_T \ln\left(\sum_{n=1}^{N} \frac{x_n}{A} \exp\left(-\frac{1}{\alpha_T}\sum_{t=1}^{T} [\nabla_{\boldsymbol{x}} \mathcal{L}_t(\boldsymbol{x}_t; \boldsymbol{\mu}_t)]_{(n)}\right)\right) - \alpha_T \ln(N) \\
&\overset{(b)}{\geq} -\sum_{n=1}^{N} \frac{x_n}{A} \sum_{t=1}^{T} [\nabla_{\boldsymbol{x}} \mathcal{L}_t(\boldsymbol{x}_t; \boldsymbol{\mu}_t)]_{(n)} - \alpha_T \ln(N) \\
&= -\sum_{t=1}^{T} \frac{\langle \boldsymbol{x}, \nabla_{\boldsymbol{x}} \mathcal{L}_t(\boldsymbol{x}_t; \boldsymbol{\mu}_t) \rangle}{A} - \alpha_T \ln(N)
\end{aligned} \tag{39}$$

where the inequality (a) follows from $\|\boldsymbol{x}\|_1 = A$ and $\boldsymbol{x} \geq 0$. The inequality (b) follows from Jensen's inequality.

Combining the above two aspects, we have

$$\begin{aligned}
&\sum_{t=1}^{T} \frac{\langle \hat{\boldsymbol{x}}_t - \boldsymbol{x}, \nabla_{\boldsymbol{x}} \mathcal{L}_t(\hat{\boldsymbol{x}}_t; \boldsymbol{\mu}_t) \rangle}{A} \leq \\
&\alpha_T \ln(N) + \sum_{t=1}^{T} \frac{(1+M)(1+\|\boldsymbol{\mu}_t\|_2^2)G^2}{2\alpha_t}
\end{aligned} \tag{40}$$

This completes the proof. □

Proof of Theorem 1.

$$
\begin{aligned}
&\sum_{t=1}^{T} \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\boldsymbol{x};\boldsymbol{\mu}_t) \\
=&\sum_{t=1}^{T} \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) + \sum_{t=1}^{T} \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}_t) - \mathcal{L}_t(\boldsymbol{x};\boldsymbol{\mu}_t) \\
\leq&\frac{\|\boldsymbol{\mu}\|_2^2}{2}\left(\frac{1}{\beta_1}-\delta_1\right) + MNA^2G^2J_T(\boldsymbol{\beta}) + \sum_{t=1}^{T}\beta_t\delta_t^2\|\boldsymbol{\mu}_t\|_2^2 + A\left[\alpha_T\ln(N) + \frac{(1+M)G^2}{2}\left(J_T(\boldsymbol{\alpha}) + \sum_{t=1}^{T}\frac{\|\boldsymbol{\mu}_t\|_2^2}{\alpha_t}\right)\right]
\end{aligned}
\tag{41}
$$

By expanding $\sum_{t=1}^{T} \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\boldsymbol{x};\boldsymbol{\mu}_t)$, we obtain

$$
\begin{aligned}
&\sum_{t=1}^{T} \mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\boldsymbol{x};\boldsymbol{\mu}_t) \\
=&\sum_{t=1}^{T} f_t(\hat{\boldsymbol{x}}_t) - f_t(\boldsymbol{x}) + \sum_{m=1}^{M}\left[\mu_m\sum_{t=1}^{T} g_{t,m}(\hat{\boldsymbol{x}}_t) - \sum_{t=1}^{T}\mu_{t,m}g_{t,m}(\boldsymbol{x})\right] - \sum_{t=1}^{T}\frac{\delta_t\|\boldsymbol{\mu}\|_2^2}{2} + \sum_{t=1}^{T}\frac{\delta_t\|\boldsymbol{\mu}_t\|_2^2}{2}.
\end{aligned}
\tag{42}
$$

Taking the expectation on both sides over $\{\theta_t : \forall t \in \mathcal{T}\}$ leads to

$$
\begin{aligned}
&\sum_{t=1}^{T}\mathbb{E}\left[\mathcal{L}_t(\hat{\boldsymbol{x}}_t;\boldsymbol{\mu}) - \mathcal{L}_t(\boldsymbol{x};\boldsymbol{\mu}_t)\right] \\
=&Reg_T(\boldsymbol{\mathcal{U}}) + \sum_{m=1}^{M}\left[\mu_m\sum_{t=1}^{T}\mathbb{E}[g_{t,m}(\hat{\boldsymbol{x}}_t)] - \sum_{t=1}^{T}\mu_{t,m}\mathbb{E}[g_{t,m}(\boldsymbol{x})]\right] - \sum_{t=1}^{T}\frac{\delta_t\|\boldsymbol{\mu}\|_2^2}{2} + \sum_{t=1}^{T}\frac{\delta_t\mathbb{E}\|\boldsymbol{\mu}_t\|_2^2}{2} \\
\geq&Reg_T(\boldsymbol{\mathcal{U}}) + \sum_{m=1}^{M}\mu_m\sum_{t=1}^{T}\mathbb{E}[g_m(\boldsymbol{x}_t)] - \sum_{t=1}^{T}\frac{\delta_t\|\boldsymbol{\mu}\|_2^2}{2} + \sum_{t=1}^{T}\frac{\delta_t\mathbb{E}\|\boldsymbol{\mu}_t\|_2^2}{2}
\end{aligned}
\tag{43}
$$

Combining the above upper bound and lower bound, we obtain

$$
\begin{aligned}
&Reg_T(\boldsymbol{\mathcal{U}}) + \sum_{m=1}^{M}\mu_m\sum_{t=1}^{T}\mathbb{E}[g_m(\boldsymbol{x}_t)] - \sum_{t=1}^{T}\frac{\delta_t\|\boldsymbol{\mu}\|_2^2}{2} + \sum_{t=1}^{T}\frac{\delta_t\mathbb{E}\|\boldsymbol{\mu}_t\|_2^2}{2} \\
\leq&\frac{\|\boldsymbol{\mu}\|_2^2}{2}\left(\frac{1}{\beta_1}-\delta_1\right) + MNA^2G^2J_T(\boldsymbol{\beta}) + \sum_{t=1}^{T}\beta_t\delta_t^2\mathbb{E}\|\boldsymbol{\mu}_t\|_2^2 \\
&+ A\alpha_T\ln(N) + \frac{(1+M)G^2AJ_T(\boldsymbol{\alpha})}{2} + \frac{(1+M)G^2A}{2}\sum_{t=1}^{T}\frac{\mathbb{E}\|\boldsymbol{\mu}_t\|_2^2}{\alpha_t}
\end{aligned}
\tag{44}
$$

which leads to

$$
\begin{aligned}
&Reg_T(\boldsymbol{\mathcal{U}}) + \sum_{m=1}^{M}\left[\mu_m\sum_{t=1}^{T}\mathbb{E}[g_m(\boldsymbol{x}_t)] - \frac{\frac{1}{\beta_1}-\delta_1+J_T(\boldsymbol{\delta})}{2}\cdot(\mu_m)^2\right] \\
\leq&\left[\alpha_T\ln(N) + \frac{(1+M)G^2J_T(\boldsymbol{\alpha})}{2}\right]A + MNA^2G^2J_T(\boldsymbol{\beta}) + \sum_{t=1}^{T}\mathbb{E}\|\boldsymbol{\mu}_t\|_2^2\left(\beta_t\delta_t^2 + \frac{(1+M)G^2A}{2\alpha_t} - \frac{\delta_t}{2}\right) \\
\leq&\left[\alpha_T\ln(N) + \frac{(1+M)G^2J_T(\boldsymbol{\alpha})}{2}\right]A + MNA^2G^2J_T(\boldsymbol{\beta})
\end{aligned}
\tag{45}
$$

where the last inequality follows from condition C2.

Maximizing left-hand-side over $\boldsymbol{\mu}$ leads to

$$
\mathsf{Reg}(T) + \frac{\sum_{m=1}^{M}\mathbb{E}\left(\left[\sum_{t=1}^{T}g_m(\boldsymbol{x}_t)\right]^+\right)^2}{2\left(\frac{1}{\beta_1}-\delta_1+J_T(\boldsymbol{\delta})\right)} \leq \left[\alpha_T\ln(N) + \frac{(1+M)G^2J_T(\boldsymbol{\alpha})}{2}\right]A + MNA^2G^2J_T(\boldsymbol{\beta}).
\tag{46}
$$

Note that on the LHS, we have

$$
\sum_{m=1}^{M}\mathbb{E}\left(\left[\sum_{t=1}^{T}g_m(\boldsymbol{x}_t)\right]^+\right)^2 \geq \frac{[Vio_T(\boldsymbol{\mathcal{U}})]^2}{M},
\tag{47}
$$

which follows from the inequality $z_1^2 + z_2^2 + \ldots + z_M^2 \geq (z_1 + z_2 + \ldots + z_M)^2/M$.

Therefore, we obtain

$$Reg_T(\mathcal{U}) + \frac{[Vio_T(\mathcal{U})]^2}{2\left(\frac{1}{\beta_1} - \delta_1 + J_T(\boldsymbol{\delta})\right)M} \leq \left[\alpha_T \ln(N) + \frac{(1+M)G^2 J_T(\boldsymbol{\alpha})}{2}\right] A + MNA^2G^2J_T(\boldsymbol{\beta}) \tag{48}$$

When $\alpha_t = t^\epsilon G\sqrt{\frac{1+M}{2\ln(N)}}$, we obtain

$$\begin{aligned} Reg_T(\mathcal{U}) + \frac{[Vio_T(\mathcal{U})]^2}{2\left(\frac{1}{\beta_1} - \delta_1 + J_T(\boldsymbol{\delta})\right)M} \leq & \left[\alpha_T \ln(N) + \frac{(1+M)G^2}{2} J_T(\boldsymbol{\alpha})\right] A + MNA^2G^2J_T(\boldsymbol{\beta}) \\ = & \left(T^\epsilon + \frac{T^{1-\epsilon}}{1-\epsilon}\right) AG\sqrt{\frac{(1+M)\ln(N)}{2}} + MNA^2G^2J_T(\boldsymbol{\beta}) \\ \leq & \left(T^\epsilon + \frac{T^{1-\epsilon}}{1-\epsilon}\right) AG\sqrt{\frac{(1+M)\ln(N)}{2}} + \frac{MNAG}{3\sqrt{2(1+M)\ln(N)}} \cdot \frac{T^\epsilon}{\epsilon} \\ = & AG \cdot \Omega_T(\epsilon) \end{aligned} \tag{49}$$

Note that $\frac{1}{\beta_1} - \delta_1 = 3AG\sqrt{2(1+M)\ln(N)}$. This leads to

$$\begin{aligned} Vio_T(\mathcal{U}) \leq & \sqrt{6AGM\sqrt{2(1+M)\ln(N)}\left(1 + \frac{T^{1-\epsilon}}{1-\epsilon}\right)\left(AG \cdot \Omega_T(N,M) + TAG\sqrt{2N}\right)} \\ \leq & AG\sqrt{\frac{24MC}{1-\epsilon}\left[\frac{\Omega_T(\epsilon)}{T} + \sqrt{2N}\right]} \cdot T^{1-\epsilon/2} \end{aligned} \tag{50}$$

$C = \sqrt{(1+M)\ln(N)/2}$

For notations simplicity, we let $\Psi \triangleq \sqrt{\frac{(1+M)\ln(N)}{2MN}}$ and obtain

$$\begin{aligned} \mathtt{Reg}(T) \leq & AG\sqrt{MN}\left[\left(\Psi + \frac{1}{6\Psi\epsilon}\right)T^\epsilon + \frac{\Psi \cdot T^{1-\epsilon}}{1-\epsilon}\right] \\ \mathtt{Vio}(T) \leq & AG\sqrt{\frac{24\Psi\sqrt{MN}}{1-\epsilon}\left[\frac{\mathtt{Reg}(T)}{AGT} + \sqrt{2N}\right]} \cdot T^{1-\frac{\epsilon}{2}} \end{aligned} \tag{51}$$

□

Proof .

$$\begin{aligned} \mathtt{Reg}(T) \leq & AG\sqrt{MN}\left[\left(\Psi + \frac{1}{6\Psi\epsilon}\right)T^\epsilon + \frac{\Psi \cdot T^{1-\epsilon}}{1-\epsilon}\right] \\ \mathtt{Vio}(T) \leq & AG\sqrt{\frac{68MN\Psi}{1-\epsilon}}T^{1-\frac{\epsilon}{2}} \\ \mathtt{Vio}(T) \leq & AG\sqrt{48\sqrt{2}MN\Psi\frac{T^{1-\epsilon}}{1-\epsilon}T} \end{aligned} \tag{52}$$

□

Lemma 3. *Given the following parameters*

$$\begin{aligned}\alpha_t &= t^{\epsilon} G\sqrt{\frac{1+M}{2\ln(N)}}\\ \delta_t &= \frac{3(1+M)G^2A}{\alpha_t} = \frac{3GA\sqrt{2(1+M)\ln(N)}}{t^{\epsilon}}\\ \beta_t &= \frac{1}{\delta_t(t+1)} = \frac{1}{3GA\sqrt{2(1+M)\ln(N)}}\cdot\frac{t^{\epsilon}}{1+t}\end{aligned} \tag{53}$$

we have

$$\begin{aligned}J_T(\boldsymbol{\alpha}) &\le \frac{1}{G}\sqrt{\frac{2\ln(N)}{1+M}}\frac{T^{1-\epsilon}}{1-\epsilon},\\ J_T(\boldsymbol{\beta}) &\le \frac{1}{3GA\sqrt{2(1+M)\ln(N)}}\cdot\frac{T^{\epsilon}}{\epsilon}\\ J_T(\boldsymbol{\delta}) &\le 3GA\sqrt{2(1+M)\ln(N)}\frac{T^{1-\epsilon}}{1-\epsilon}.\end{aligned} \tag{54}$$

Proof of Lemma 3. The proof of this lemma relies on the fact

$$\sum_{t=1}^{T}\frac{1}{t^{\epsilon}} \le \frac{T^{1-\epsilon}}{1-\epsilon}. \tag{55}$$

We upper bound $J_T(\boldsymbol{\alpha})$ as follows:

$$J_T(\boldsymbol{\alpha}) = \sum_{t=1}^{T}\frac{1}{\alpha_t} = \frac{1}{G}\sqrt{\frac{2\ln(N)}{1+M}}\sum_{t=1}^{T}\frac{1}{t^{\epsilon}} \le \frac{1}{G}\sqrt{\frac{2\ln(N)}{1+M}}\frac{T^{1-\epsilon}}{1-\epsilon}. \tag{56}$$

We upper bound $J_T(\boldsymbol{\beta})$ as follows:

$$\begin{aligned}J_T(\boldsymbol{\beta}) = \frac{1}{3GA\sqrt{2(1+M)\ln(N)}}\sum_{t=1}^{T}\frac{t^{\epsilon}}{1+t} &\le \frac{1}{3GA\sqrt{2(1+M)\ln(N)}}\sum_{t=1}^{T}\frac{t^{\epsilon}}{t}\\ &= \frac{1}{3GA\sqrt{2(1+M)\ln(N)}}\cdot\frac{T^{\epsilon}}{\epsilon}\end{aligned} \tag{57}$$

We upper bound $J_T(\boldsymbol{\delta})$ as follows:

$$J_T(\boldsymbol{\delta}) = 3GA\sqrt{2(1+M)\ln(N)}\sum_{t=1}^{T}\frac{1}{t^{\epsilon}} \le 3GA\sqrt{2(1+M)\ln(N)}\frac{T^{1-\epsilon}}{1-\epsilon}. \tag{58}$$

□

Lemma 4. *Conditions C1 and C2 hold if the sequence $\{\alpha_t\}_{t\ge1}$ is increasing, the sequences $\{\delta_t\}_{t\ge1}$ and $\{\beta_t\}_{t\ge1}$ satisfy*

$$\begin{aligned}\delta_t &= \frac{3(1+M)G^2A}{\alpha_t}\\ \beta_t &= \frac{1}{\delta_t(t+1)}\end{aligned} \tag{59}$$

$$\begin{aligned}&\text{C1: } \frac{1}{\beta_t} - \frac{1}{\beta_{t-1}} - \delta_t \le 0, \quad \forall t \ge 2\\ &\text{C2: } \beta_t\delta_t^2 + \frac{(1+M)G^2A}{2\alpha_t} - \frac{\delta_t}{2} \le 0, \quad \forall t \ge 2\end{aligned} \tag{60}$$

Proof of Lemma 4. We prove C1 as follows:

$$\frac{1}{\beta_t} - \frac{1}{\beta_{t-1}} = \delta_t(t+1) - \delta_{t-1}t = (\delta_t - \delta_{t-1})t + \delta_t \le \delta_t, \tag{61}$$

where the last inequality follows from that the sequence $\{\delta_t\}_{t>1}$ is decreasing (i.e., $\delta_t < \delta_{t-1}$).

We prove C2 as follows:

$$
\begin{aligned}
\beta_t \delta_t^2 + \frac{(1+M)G^2 A}{2\alpha_t} - \frac{\delta_t}{2} &\overset{(a)}{=} \frac{\delta_t}{t+1} + \frac{(1+M)G^2 A}{2\alpha_t} - \frac{\delta_t}{2} \\
&= \delta_t \left(\frac{1}{t+1} - \frac{1}{2} \right) + \frac{(1+M)G^2 A}{2\alpha_t} \\
&\overset{(b)}{\leq} -\frac{\delta_t}{6} + \frac{(1+M)G^2 A}{2\alpha_t} \\
&\overset{(c)}{=} 0
\end{aligned}
\tag{62}
$$

where the equality (a) follows from $\beta_t = \frac{1}{\delta_t (t+1)}$, the inequality (b) follows $t \geq 2$ in C2, the equality (c) follows from $\delta_t = \frac{3(1+M)G^2 A}{\alpha_t}$. This completes the proof. □