SUPPLEMENTARY TO "NON-ASYMPTOTIC ANALYSIS OF STOCHASTIC GRADIENT DESCENT UNDER LOCAL DIFFERENTIAL PRIVACY GUARANTEE"

Anonymous authors
Paper under double-blind review

A APPENDIX

The appendix contains all technical proofs and some additional experiment results.

A.1 PROOF OF LEMMAS AND THEOREMS

Proof of Theorem 1. Consider two different data points x and x', at the first iteration, notice that

$$\|\eta_1 \nabla f(\theta_0, x) - \eta_1 \nabla f(\theta_0, x')\|_2 \leq 2\eta_1 C_0.$$

Therefore, the sensitivity of the gradient iterates θ_1 is bounded by $2\eta_1 C_0$. Using Proposition 1, adding a Gaussian noise $2\eta_1 C_0/\mu_i \omega_1$ to a gradient update makes this step μ_1-GDP. Consequently, θ_1 is μ_1-GDP since the initial estimate θ_0 is deterministic. The second iterate θ_2 takes θ_1 as input in addition to the dataset. According to the Proposition 2, we then have the two-fold joint mechanism (θ_1, θ_2) is $\max\{\mu_1, \mu_2\}$-GDP. Using the same argument repeatedly, we have completed the proof of this theorem.

□

Proof of Theorem 2. Following the proof of Theorem 1, the statement in Theorem 2 is also true.

□

Proof of Theorem 3. Recall that

$$\theta_n = \theta_{n-1} - \eta_n \nabla f(\theta_{n-1}, x_n) + \eta_n A_n,$$

where $A_n = 2C_0 \omega_n / \mu_n$. We then have

$$\begin{aligned} \|\theta_n - \theta^*\|^2 = &\|\theta_{n-1} - \theta^*\|^2 - 2\eta_n \langle \theta_{n-1} - \theta^*, \nabla f(\theta_{n-1}, x_n)\rangle + \eta_n^2 \|\nabla f(\theta_{n-1}, x_n)\|^2 + \eta_n^2 \|A_n\|^2 \\ &+ 2\eta_n \langle \theta_{n-1} - \theta^*, A_n\rangle - 2\eta_n^2 \langle \nabla f(\theta_{n-1}, x_n), A_n\rangle. \end{aligned} \tag{1}$$

Notice that $E(A_n) = 0$ and therefore the expectations of the last two terms of (1) are zero. Under Condition 1, we have $E(\|\nabla f(\theta_{i-1}, x_i)\|^2) \leq C_0^2$. For the second term of (1), by Condition **??**, we have

$$\begin{aligned} E(\langle \theta_{n-1} - \theta^*, \nabla f(\theta_{n-1}, x_n)\rangle) =& E(\langle \theta_{n-1} - \theta^*, \nabla f(\theta_{n-1}, x_n) - \nabla f(\theta^*, x_n)\rangle) \\ \geq& mE(\|\theta_{n-1} - \theta^*\|^2), \end{aligned}$$

where the first equality holds due to $E\{\nabla f(\theta^*, x_n)\} = 0$. For the fourth term of (1), we know

$$E(\|A_n\|^2) = \frac{4C_0^2}{\mu_n^2} E(\|\omega_n\|^2) = \frac{64 C_0^2 d}{\mu_n^2} := \tau_n^2.$$

Define $\Delta_n = E(\|\theta_n - \theta^*\|^2)$. Combining the above inequalities, we then have the following recursion rule

$$\Delta_n \leq \Delta_{n-1} - 2m\eta_n \Delta_{n-1} + \eta_n^2 C_0^2 + \gamma_n^2 \tau_n^2 \leq (1 - 2m\eta_n)\Delta_{n-1} + \eta_n^2 (C_0^2 + \tau_n^2). \tag{2}$$

Notice that $(1-2m\eta_n)\Delta_{n-1}$ is always positive with $\eta_n = \eta n^{-\alpha}$ when $2m\eta < 1$. Applying the recursion (2) n times, we obtain

$$\begin{aligned}\Delta_n &\le \prod_{k=1}^{n}(1-2m\eta_k)\Delta_0 + \sum_{k=1}^{n}\prod_{i=k+1}^{n}(1-2m\eta_i)\{\eta_k^2(C_0^2+\tau_k^2)\}\\ &\le A_{1,n}\Delta_0 + A_{2,n}\tilde{\sigma}_\mu^2,\end{aligned}$$

where $\tilde{\sigma}_\mu^2 = \max_k\{C_0^2+\tau_k^2\}$. For $A_{1,n}$, using the fact $e^{-x}\ge 1-x$, we have

$$A_{1,n} \le \exp\left(-2m\sum_{k=1}^{n}\eta_k\right).$$

For $A_{2,n}$, we decompose it as follows

$$\begin{aligned}A_{2,n} =& \sum_{k=1}^{n/2}\prod_{i=k+1}^{n}(1-2m\eta_i)\eta_k^2 + \sum_{k=n/2+1}^{n}\prod_{i=k+1}^{n}(1-2m\eta_i)\eta_k^2\\ \le& \prod_{i=n/2+1}^{n}(1-2m\eta_i)\sum_{k=1}^{n/2}\eta_k^2 + \eta_{n/2}\sum_{k=n/2+1}^{n}\prod_{i=k+1}^{n}(1-2m\eta_i)\,\eta_k\\ \le& \exp\left(-2m\sum_{i=n/2+1}^{n}\eta_i\right)\sum_{k=1}^{n/2}\eta_k^2 + \frac{\eta_{n/2}}{2m}\sum_{k=n/2+1}^{n}\left[\prod_{i=k+1}^{n}(1-2m\eta_i) - \prod_{i=k}^{n}(1-2m\eta_i)\right]\\ \le& \exp\left(-2m\sum_{i=n/2+1}^{n}\eta_i\right)\sum_{k=1}^{n/2}\eta_k^2 + \frac{\eta_{n/2}}{2m}\left[1-\prod_{i=n/2+1}^{n}(1-2m\eta_i)\right]\\ \le& \exp\left(-2m\sum_{i=n/2+1}^{n}\eta_i\right)\sum_{k=1}^{n}\eta_k^2 + \frac{\eta_{n/2}}{2m}.\end{aligned}$$

When $\eta_n = \eta n^{-\alpha}$, we have the two following two inequalities

$$\frac{\eta}{2}\psi_{1-\alpha}(n) \le \sum_{i=1}^{n}\eta_i \le \eta\psi_{1-\alpha}(n), \quad \frac{\eta^2}{2}\psi_{1-2\alpha}(n) \le \sum_{i=1}^{n}\eta_i^2 \le \eta^2\psi_{1-2\alpha}(n),$$

where $\psi_\beta(n) = (n^\beta - 1)/\beta$. Then, we have

$$\begin{aligned}A_{1,n} &\le \exp\{-2m\eta\psi_{1-\alpha}(n)\},\\ A_{2,n} &\le \eta^2\psi_{1-2\alpha}(n)\exp\{-m\eta(\psi_{1-\alpha}(n)-\psi_{1-\alpha}(n/2))\} + \frac{\eta}{mn^\alpha}.\end{aligned}$$

Note that $\psi_{1-\alpha}(n)-\psi_{1-\alpha}(n/2) \ge n^{1-\alpha}/2$. Therefore, using the above inequalities, we have

$$\begin{aligned}\Delta_n \le& \exp\{-2m\eta\psi_{1-\alpha}(n)\}\Delta_0 + \frac{\eta\sigma_M^2}{mn^\alpha} + \eta^2\sigma_M^2\psi_{1-2\alpha}(n)\exp(-m\eta n^{1-\alpha}/2)\\ \le& \exp\{-2m\eta\psi_{1-\alpha}(n)\}\Delta_0 + C_0^2\left(\frac{\eta}{mn^\alpha} + \eta^2\psi_{1-2\alpha}(n)\exp(-m\eta n^{1-\alpha}/2)\right)\left(1+\frac{64d}{\min_k\{\mu_k^2\}}\right).\end{aligned}$$

□

Proof of Theorem 4. Notice that

$$\begin{aligned}&E\{\nabla^2 f(\theta^*, x_n)\}(\theta_{n-1}-\theta^*)\\ =&\nabla^2 f(\theta^*, x_n)(\theta_{n-1}-\theta^*) + \left[E\{\nabla^2 f(\theta^*, x_n)\} - \nabla^2 f(\theta^*, x_n)\right](\theta_{n-1}-\theta^*)\\ =&\nabla f(\theta_{n-1}, x_n) - \nabla f(\theta^*, x_n) + \left[\nabla^2 f(\theta^*, x_n)(\theta_{n-1}-\theta^*) - \nabla f(\theta_{n-1}, x_n) + \nabla f(\theta^*, x_n)\right]\\ &+\left[E\left\{\nabla^2 f(\theta^*, x_n)\right\} - \nabla^2 f(\theta^*, x_n)\right](\theta_{n-1}-\theta^*)\\ =&I_1+I_2+I_3+I_4. \end{aligned} \tag{3}$$

For I_1, we have

$$\nabla f\left(\theta_{n-1}, x_n\right)=\frac{1}{\eta_n}\left(\theta_{n-1}-\theta_n\right)+A_n .$$

Then, we get

$$\begin{aligned}
& \frac{1}{n} \sum_{k=1}^n \nabla f\left(\theta_{k-1}, x_k\right)=\frac{1}{n} \sum_{k=1}^n \frac{1}{\eta_k}\left(\theta_{k-1}-\theta_k\right)+\frac{1}{n} \sum_{k=1}^n A_k \\
= & \frac{1}{n} \sum_{k=1}^{n-1}\left(\theta_k-\theta^*\right) \cdot\left(\frac{1}{\eta_{k+1}}-\frac{1}{\eta_k}\right)-\frac{1}{n}\left(\theta_n-\theta^*\right) \cdot \frac{1}{\eta_n}+\frac{1}{n}\left(\theta_0-\theta^*\right) \cdot \frac{1}{\eta_1}+\frac{1}{n} \sum_{k=1}^n A_k .
\end{aligned}$$

For I_2, we have

$$E\left\|E\{\nabla^2 f(\theta^*, x_n)\}^{-1} \frac{1}{n} \sum_{k=1}^n \nabla f\left(\theta^*, x_k\right)\right\|^2 \leq \frac{1}{n} \operatorname{tr}\left[E\{\nabla^2 f\left(\theta^*, x_k\right)\}^{-1} S E\{\nabla^2 f\left(\theta^*, x_k\right)\}^{-1}\right],$$

where $S=\mathbb{E}\{\nabla f(\theta^*, x) \nabla f(\theta^*, x)^{\top}\}$. For I_3, using Condition 3, we obtain

$$\left\|\nabla^2 f\left(\theta^*, x_n\right)\left(\theta_{n-1}-\theta^*\right)-\nabla f\left(\theta_{n-1}, x_n\right)+\nabla f\left(\theta^*, x_n\right)\right\| \leq \frac{C_1}{2}\|\theta_{n-1}-\theta^*\|^2 .$$

Therefore,

$$E\left\|\frac{1}{n} \sum_{k=1}^n\left\{\nabla^2 f\left(\theta^*, x_k\right)\left(\theta_{k-1}-\theta^*\right)-\nabla f\left(\theta_{k-1}, x_k\right)+\nabla f\left(\theta^*, x_k\right)\right\}\right\|^2 \leq \frac{C_1^2}{4 n^2} \sum_{k=0}^{n-1}\left\|\theta_k-\theta^*\right\|^4 .$$

For I_4, under Condition 2, we get

$$E(\left\|\left[E\left\{\nabla^2 f\left(\theta^*, x_n\right)\right\}-\nabla^2 f\left(\theta^*, x_n\right)\right]\left(\theta_{n-1}-\theta^*\right)\right\|^2) \leq 4 M^2\|\theta_{n-1}-\theta^*\|^2 .$$

We then have

$$E\left\|\frac{1}{n} \sum_{k=1}^n\left[E\left\{\nabla^2 f\left(\theta^*, x_k\right)\right\}-\nabla^2 f\left(\theta^*, x_k\right)\right]\left(\theta_{k-1}-\theta^*\right)\right\| \leq \frac{4 M^2}{n^2} \sum_{k=0}^{n-1}\|\theta_k-\theta^*\|^2 .$$

Next, we need to bound $\|\theta_n-\theta^*\|^2$ and $\|\theta_n-\theta^*\|^4$. From the proof of Theorem 3, we know the following recursion formula

$$\begin{aligned}
\|\theta_n-\theta^*\|^2= & \|\theta_{n-1}-\theta^*\|^2-2 \eta_n\langle\theta_{n-1}-\theta^*, \nabla f(\theta_{n-1}, x_n)\rangle+\eta_n^2\|\nabla f(\theta_{n-1}, x_n)\|^2+\eta_n^2\|A_n\|^2 \\
& +2 \eta_n\langle\theta_{n-1}-\theta^*, A_n\rangle-2 \eta_n^2\langle\nabla f(\theta_{n-1}, x_n), A_n\rangle .
\end{aligned}$$

For notation simplicity, denote $\delta_n=\theta_n-\theta^*$ and $\Psi_n=\nabla f(\theta_n, x_{n+1})$. Then, we have the fourth order as follows

$$\begin{aligned}
\|\delta_n\|^4= & \|\delta_{n-1}\|^4-4 \eta_n\langle\delta_{n-1}, \Psi_n\rangle\|\delta_{n-1}\|^2+4 \eta_n\langle\delta_{n-1}, A_n\rangle\|\delta_{n-1}\|^2-4 \eta_n^2\langle\Psi_{n-1}, A_n\rangle\|\delta_{n-1}\|^2 \\
& +2 \eta_n^2\|\Psi_{n-1}\|^2\|\delta_{n-1}\|^2+2 \eta_n^2\|A_n\|^2\|\delta_{n-1}\|^2-4 \eta_n^2\langle\Psi_{n-1}, A_n\rangle\|\delta_{n-1}\|^2 \\
& +4 \eta_n^3\langle\delta_{n-1}, \Psi_{n-1}\rangle\langle\Psi_{n-1}, A_n\rangle-4 \eta_n^3\|\Psi_{n-1}\|^2\langle\delta_{n-1}, \Psi_{n-1}\rangle-4 \eta_n^3\|A_n\|^2\langle\delta_{n-1}, \Psi_{n-1}\rangle \\
& -8 \eta_n^3\langle\delta_{n-1}, A_n\rangle\|A_n\|^2+4 \eta_n^3\langle\delta_{n-1}, A_n\rangle\|\Psi_{n-1}\|^2 \\
& +4 \eta_n^3\langle\delta_{n-1}, A_n\rangle\|A_n\|^2-4 \eta_n^3\langle\Psi_{n-1}, A_n\rangle\|\Psi_{n-1}\|^2-4 \eta_n^3\langle\Psi_{n-1}, A_n\rangle\|A_n\|^2 \\
& +2 \eta_n^4\|\Psi_{n-1}\|^2\|A_n\|^2+4 \eta_n^2\|\Psi_{n-1}\|^2\|\delta_{n-1}\|^2+4 \eta_n^2\|A_n\|^2\|\delta_{n-1}\|^2+4 \eta_n^4\|A_n\|^2\|\Psi_{n-1}\|^2 \\
& +\eta_n^4\|\Psi_{n-1}\|^4+\eta_n^4\|A_n\|^4 .
\end{aligned}$$

Note that all terms containing the first moment of A_n have zero expectations, then we only need to bound the following terms

$$\begin{aligned}
& -4 \eta_n\langle\delta_{n-1}, \Psi_n\rangle\|\delta_{n-1}\|^2 \leq-4 m \eta_n\|\delta_{n-1}\|^4, \\
& 6 \eta_n^2\|\Psi_{n-1}\|^2\|\delta_{n-1}\|^2 \leq m \eta_n\|\delta_{n-1}\|^4+9 \eta_n^3\|\Psi_{n-1}\|^4 / m, \\
& 2 \eta_n^2\|A_n\|^2\|\delta_{n-1}\|^2 \leq m \eta_n\|\delta_{n-1}\|^4 / 3+3 \eta_n^3\|A_n\|^4 / m, \\
& -4 \eta_n^3\|\Psi_{n-1}\|^2\langle\delta_{n-1}, \Psi_{n-1}\rangle \leq \eta_n^3\|\delta_{n-1}\|^4+3 \eta_n^3\|\Psi_{n-1}\|^4, \\
& -4 \eta_n^3\|A_n\|^2\langle\delta_{n-1}, \Psi_{n-1}\rangle \leq \eta_n^3\|\delta_{n-1}\|^4+3 \eta_n^3\|A_n\|^4, \\
& 6 \eta_n^4\|\Psi_{n-1}\|^2\|A_n\|^2 \leq 3 \eta_n^4\|\Psi_{n-1}\|^4+3 \eta_n^4\|A_n\|^4, \\
& 4 \eta_n^2\|A_n\|^2\|\delta_{n-1}\|^2 \leq 2 m \eta_n\|\delta_{n-1}\|^4 / 3+6 \eta_n^3\|A_n\|^4 / m .
\end{aligned}$$

According to the Lemma 1.19 of Rigollet & Hütter (2015) and Condition 1, we know $E(\|A_n\|^4) \leq \tau_n^4$ and $E(\|\Psi_{n-1}\|^4) \leq C_0^4$. Combining the above inequalities, the expectation of fourth order has the following recursion

$$E(\|\delta_n\|^4) \leq (1 - 2m\eta_n + 2\eta_n^3)E(\|\delta_{n-1}\|^4) + 4(C_0^4 + \tau_n^4)\eta_n^3 \left(\eta_n + \frac{3m+9}{4m}\right).$$

Denote

$$C_{3,0} = (C_0^4 + \tau_n^4)(3 + 9/m), \quad C_{4,0} = 4(C_0^4 + \tau_n^4).$$

We then have

$$\begin{aligned} E(\|\delta_n\|^4) &\leq (1 - 2m\eta_n + 2\eta_n^3)E(\|\delta_{n-1}\|^4) + C_{3,0}\gamma_n^3 + C_{4,0}\gamma_n^4 \\ &\leq \prod_{k=1}^{n}(1 - 2m\eta_k + 2\eta_k^3)E(\|\delta_0\|^4) + \sum_{k=1}^{n}\prod_{i=k+1}^{n}(1 - 2m\eta_i + 2\eta_i^3)(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ &:= B_{1,n}E(\|\delta_0\|^4) + B_{2,n}. \end{aligned}$$

It is easy to show that

$$B_{1,n} \leq \exp\left(-2m\sum_{k=1}^{n}\eta_k + 2\sum_{k=1}^{n}\eta_k^3\right).$$

For $B_{2,n}$, denote $n_0 = \inf\{n \in \mathbb{N} : 2\eta_n^3 \geq m\eta_n\}$, we obtain

$$\begin{aligned} B_{2,n} \leq &\sum_{k=n_0+1}^{n}\prod_{i=k+1}^{n}\exp(-m\eta_i)(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ &+ \prod_{k=n_0+1}^{n}\exp(-m\eta_k)\sum_{k=1}^{n_0}\prod_{i=k+1}^{n_0}\exp(2\eta_i^3)(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ \leq &\sum_{k=1}^{n}\prod_{i=k+1}^{n}\exp(-m\eta_i)(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ &+ \prod_{k=1}^{n}\exp(-m\eta_k)\prod_{k=1}^{n_0}\exp(m\eta_k)\prod_{i=1}^{n_0}\exp(2\eta_i^3)\sum_{k=1}^{n_0}(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ \leq &\sum_{k=1}^{n}\prod_{i=k+1}^{n}\exp(-m\eta_i)(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ &+ \prod_{k=1}^{n}\exp(-m\eta_k)\prod_{i=1}^{n_0}\exp(4\eta_i^3)\sum_{k=1}^{n_0}(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ := &B_{3,n} + B_{4,n}. \end{aligned}$$

For $B_{3,n}$, we have

$$\begin{aligned} B_{3,n} = &\sum_{k=1}^{n/2}\prod_{i=k+1}^{n}\exp(-m\eta_i)(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) + \sum_{k=n/2+1}^{n}\prod_{i=k+1}^{n}\exp(-m\eta_i)(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) \\ \leq &\prod_{i=n/2+1}^{n}\exp(-m\eta_i)\sum_{k=1}^{n/2}(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) + \sum_{k=n/2+1}^{n}\prod_{i=k+1}^{n}\exp(-m\eta_i)(C_{3,0}\eta_{n/2}^2 + C_{4,0}\eta_{n/2}^3)\eta_k \\ \leq &\prod_{i=n/2+1}^{n}\exp(-m\eta_i)\sum_{k=1}^{n/2}(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) + (C_{3,0}\eta_{n/2}^2 + C_{4,0}\eta_{n/2}^3)\sum_{k=1}^{n}\exp\left(-m\sum_{i=k+1}^{n}\eta_i\right)\eta_k \\ \leq &\prod_{i=n/2+1}^{n}\exp(-m\eta_i)\sum_{k=1}^{n/2}(C_{3,0}\eta_k^3 + C_{4,0}\eta_k^4) + \frac{1}{m}(C_{3,0}\eta_{\frac{n}{2}}^2 + C_{4,0}\eta_{n/2}^3). \end{aligned}$$

When $\eta_n = \eta n^{-\alpha}$ and using the bounds of summations, we obtain

$$
\begin{aligned}
B_{1,n} &\leq \exp\{-m\eta\psi_{1-\alpha}(n) + 2\eta^3\psi_{1-3\alpha}(n)\}, \\
B_{3,n} &\leq \exp\left[-\frac{m\eta}{2}\left\{\psi_{1-\alpha}(n) - \psi_{1-\alpha}(n/2)\right\}\right]\{C_{3,0}\eta^3\psi_{1-3\alpha}(n) + C_{4,0}\eta^4\psi_{1-4\alpha}(n)\} \\
&\quad + \frac{1}{m}\left(C_{3,0}\frac{\eta^2 2^{2\alpha}}{n^{2\alpha}} + C_{4,0}\frac{\eta^3 2^{3\alpha}}{n^{3\alpha}}\right), \\
B_{4,n} &\leq \exp\left\{-\frac{m\eta}{2}\psi_{1-\alpha}(n) + 4\eta^3\psi_{1-3\alpha}(n))(C_{3,0}\eta^3\psi_{1-3\alpha}(n) + C_{4,0}\eta^4\psi_{1-4\alpha}(n)\right\}.
\end{aligned}
$$

Combining the above inequalities, we have

$$
\begin{aligned}
E(\|\delta_n\|^4) \leq{}& \exp\{-m\eta\psi_{1-\alpha}(n) + 2\eta^3\psi_{1-3\alpha}(n)\}E(\|\delta_0\|^4) + \frac{1}{m}\left(C_{3,0}\frac{\eta^2 2^{2\alpha}}{n^{2\alpha}} + C_{4,0}\frac{\eta^3 2^{3\alpha}}{n^{3\alpha}}\right) \\
&+ \exp\left\{-\frac{m\eta}{4}n^{1-\alpha} + 4\eta^3\psi_{1-3\alpha}(n)\right\}\{C_{3,0}\eta^3\psi_{1-3\alpha}(n) + C_{4,0}\eta^4\psi_{1-4\alpha}(n)\}.
\end{aligned}
$$

Therefore, using (3), we get

$$
\begin{aligned}
E(\|\bar{\theta}_n - \theta^*\|^2) \leq{}& \frac{1}{n}\operatorname{tr}\left[E\{\nabla^2 f(\theta^*, x_k)\}^{-1} S E\{\nabla^2 f(\theta^*, x_k)\}^{-1}\right] \\
&+ \frac{1}{mn^2}\sum_{k=1}^{n-1} E(\|\theta_k - \theta^*\|^2)(\eta_{k+1}^{-1} - \eta_k^{-1})^2 + \frac{1}{mn^2\eta_n^2}E(\|\theta_n - \theta^*\|^2) \\
&+ \left(\frac{1}{mn^2\eta_1^2} + \frac{4M^2}{mn^2}\right)E(\|\theta_0 - \theta^*\|^2) + \frac{1}{mn^2}\sum_{k=1}^{n} E(\|A_k\|^2) \\
&+ \frac{C_1}{4n^2 m}\sum_{k=1}^{n-1} E(\|\theta_k - \theta^*\|^4) + \frac{C_1}{4n^2 m}E(\|\theta_0 - \theta^*\|^4) + \frac{4M^2}{mn^2}\sum_{k=1}^{n-1} E\|\theta_k - \theta^*\|^2 \\
&:= J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + J_7 + J_8.
\end{aligned}
$$

Now, we calculate the bound for each term J_j, $j = 1, \ldots, 8$. Let

$$
B = \sum_{k=1}^{n} \exp\left\{-\frac{m\eta}{k}^{1-\alpha}/4 + \psi_{1-2\alpha}(k) + 4\eta^3\psi_{1-3\alpha}(k)\right\}.
$$

We then have

$$
\begin{aligned}
J_2 &\leq \frac{4\alpha^2}{mn^2\eta^2}\sum_{k=1}^{n-1} k^{2(\alpha-1)}E(\|\theta_k - \theta^*\|^2) \\
&\leq \frac{4\alpha^2}{mn^2\eta^2}\sum_{k=1}^{n-1}\frac{\eta\tilde{\sigma}_\mu^2}{m}k^{\alpha-2} + \frac{4\alpha^2}{mn^2\eta^2}\sum_{k=1}^{n-1}\exp\{\psi_{1-2\alpha}(k) - m\eta k^{1-\alpha}/2\}(E\|\delta_0\|^2 + \eta^2\tilde{\sigma}_\mu^2)k^{2(\alpha-1)} \\
&\leq \frac{4\alpha^2\tilde{\sigma}_\mu^2}{m^2 n^2\eta}\psi_{\alpha-1}(n) + \frac{4\alpha^2}{mn^2\eta^2}(E\|\delta_0\|^2 + \eta^2\tilde{\sigma}_\mu^2)B, \\
J_3 &\leq \frac{1}{mn^{2(1-\alpha)}\eta^2}\left[\exp\{-2m\eta\psi_{1-\alpha}(n)\}\Delta_0 + \frac{\eta\tilde{\sigma}_\mu^2}{mn^\alpha} + \eta^2\tilde{\sigma}_\mu^2\psi_{1-2\alpha}(n)\exp(-m\eta n^{1-\alpha}/2)\right] \\
&\leq \frac{\tilde{\sigma}_\mu^2}{m^2 n^{2-\alpha}\eta} + \frac{1}{mn^{2(1-\alpha)}\eta^2}E\|\delta_0\|^2 + \frac{\tilde{\sigma}_\mu^2}{mn^{2(1-\alpha)}}\psi_{1-2\alpha}(n), \\
J_5 &\leq \frac{1}{mn^2}\sum_{k=1}^{n}\tau_k^2 \leq \frac{\tilde{\sigma}_\mu^2}{mn},
\end{aligned}
$$

$$
\begin{aligned}
J_6 \leq & \frac{C_1}{4n^2m^2}\sum_{k=1}^{n-1}\left(C_{3,0}\frac{\eta^2 2^{2\alpha}}{k^{2\alpha}} + C_{4,0}\frac{\eta^3 2^{3\alpha}}{k^{3\alpha}}\right) \\
& + \frac{C_1}{4n^2m}\left\{E\|\delta_0\|^4 + C_{3,0}\eta^3\psi_{1-3\alpha}(n) + C_{4,0}\eta^4\psi_{1-4\alpha}(n)\right\}B \\
\leq & \frac{C_1C_{3,0}\eta^2}{n^2m^2}\psi_{1-2\alpha}(n) + \frac{2C_1C_{4,0}\eta^3}{n^2m^2}\psi_{1-3\alpha}(n) \\
& + \frac{C_1}{4n^2m}\left\{E\|\delta_0\|^4 + C_{3,0}\eta^3\psi_{1-3\alpha}(n) + C_{4,0}\eta^4\psi_{1-4\alpha}(n)\right\}B, \\
J_8 \leq & \frac{4M^2}{mn^2}\sum_{k=1}^{n-1}\frac{\eta\tilde{\sigma}_\mu^2}{m}k^{-\alpha} + \frac{4M^2}{mn^2}(E\|\delta_0\|^2 + \eta^2\tilde{\sigma}_\mu^2)B \\
\leq & \frac{4M^2\eta\tilde{\sigma}_\mu^2}{m^2n^2}\psi_{1-\alpha}(n) + \frac{4M^2}{mn^2}(E\|\delta_0\|^2 + \eta^2\tilde{\sigma}_\mu^2)B.
\end{aligned}
$$

Finally, we obtain the overall bound as

$$
\begin{aligned}
E(\|\bar{\theta}_n - \theta^*\|^2) \leq & \frac{1}{n}\operatorname{tr}\left[E\{\nabla^2 f(\theta^*, x_k)\}^{-1} S E\{\nabla^2 f(\theta^*, x_k)\}^{-1}\right] + \frac{4\alpha^2\tilde{\sigma}_\mu^2}{m^2n^2\eta}\psi_{\alpha-1}(n) + \frac{\tilde{\sigma}_\mu^2}{m^2n^{2-\alpha}\eta} \\
& + \left(\frac{1}{mn^2\eta_1^2} + \frac{1}{mn^{2(1-\alpha)}\eta^2} + \frac{4M^2}{mn^2}\right)E\|\delta_0\|^2 + \frac{\tilde{\sigma}_\mu^2}{mn^{2(1-\alpha)}}\psi_{1-2\alpha}(n) \\
& + \frac{C_1C_{3,0}\eta^2}{n^2m^2}\psi_{1-2\alpha}(n) + \frac{2C_1C_{4,0}\eta^3}{n^2m^2}\psi_{1-3\alpha}(n) + \frac{4M^2\eta\tilde{\sigma}_\mu^2}{m^2n^2}\psi_{1-\alpha}(n) + \frac{\tilde{\sigma}_\mu^2}{mn} \\
& + \left(\frac{4\alpha^2}{mn^2\eta^2} + \frac{4M^2}{mn^2}\right)(E\|\delta_0\|^2 + \eta^2\tilde{\sigma}_\mu^2)B + \frac{C_1}{4mn^2}E\|\delta_0\|^4 \\
& + \frac{4M^2\eta\tilde{\sigma}_\mu^2}{m^2n^2}\left\{E\|\delta_0\|^4 + C_{3,0}\eta^3\psi_{1-3\alpha}(n) + C_{4,0}\eta^4\psi_{1-4\alpha}(n)\right\}B.
\end{aligned}
$$

The proof of this theorem is completed. □

Proof of Theorem 5. Using the same arguments in Theorem 3, we have the following recursion

$$\Delta_n \leq \Delta_{n-1} + C_0^2\eta_n^2 + \tau_n^2\eta_n^2.$$

After applying the previous recursion n times, we have

$$\Delta_n \leq \Delta_0 + \tilde{\sigma}_\mu^2\sum_{k=1}^{n}\eta_k^2 \equiv K_n,$$

where $\tilde{\sigma}_\mu^2 = \max_k\{C_0^2 + \tau_k^2\}$. Using the Lipschitz continuity of $F(\theta)$ and taking conditional expectations, we have

$$
\begin{aligned}
F(\theta_n) &\leq F(\theta_{n-1}) - \eta_n\langle\nabla F(\theta_{n-1}), \nabla F(\theta_{n-1}, x_n) + A_n\rangle + \frac{\eta_n^2 M}{2}\|\nabla F(\theta_{n-1}, x_n) + A_n\|^2, \\
E\{F(\theta_n) \mid \mathcal{F}_{n-1}\} &\leq F(\theta_{n-1}) - \eta_n\|\nabla F(\theta_{n-1})\|^2 + \frac{\eta_n^2 M}{2}E\{\|\nabla F(\theta_{n-1}, x_n)\|^2\} + \frac{\eta_n^2\tau_n^2 M}{2},
\end{aligned}
$$

where $(\mathcal{F}_n)_{n\geqslant 0}$ are an increasing family of σ-fields and θ_0 is $\mathcal{F}_0$-measurable. Using $F(\theta)-F(\theta^*) \leq \langle\nabla F(\theta), \theta - \theta^*\rangle$, denote $D_n = E\{F(\theta_n) - F(\theta^*)\}$, we have

$$D_{n-1}^2 = [E\{F(\theta_{n-1}) - F(\theta^*)\}]^2 \leq E\{\|\nabla F(\theta_{n-1})\|^2\}E(\|\theta_{n-1} - \theta^*\|^2).$$

In addition, note that $E(\|\nabla f(\theta_{n-1}, x_n)\|^2) \leq C_0^2$, we have

$$D_n \leq D_{n-1} - \frac{\eta_n}{K_n}D_{n-1}^2 + \frac{\eta_n^2 M\tilde{\sigma}_\mu^2}{2}.$$

For $\boldsymbol{\alpha > 1/2}$, $\sum_{k=1}^{\infty} \eta_k^2$ is bounded and we define $\lim_{n\to\infty} K_n = K_\infty$. Consider another recursion by replacing inequality with equality:

$$\widetilde{D}_n = \widetilde{D}_{n-1} - \frac{\eta_n}{K_\infty}\widetilde{D}_{n-1}^2 + \frac{\eta_n^2 M \tilde{\sigma}_\mu^2}{2},$$

with $D_0 = \widetilde{D}_0$. Next, we show that for any $n \in \mathbb{N}$, $D_n \le \widetilde{D}_n$. Define the function

$$t \mapsto g(t) := t - \frac{\eta_n}{K_\infty} t^2,$$

which is increasing on $[0, K_\infty/(2\eta_n)]$. Note that both D_n and $\widetilde{D}_n$ is bounded by $(M/2)K_\infty$. When $\eta < 1/M$, we have $[0, MK_\infty/2] \subset [0, K_\infty/2\eta_n]$. Then $D_0 = \widetilde{D_0}$ implies $D_1 \le \widetilde{D}_1$. Due to both sequences being within the increasing interval of $g(t)$, we have for any $n \in \mathbb{N}$, $D_n \le \widetilde{D}_n$.
Define $\varepsilon_n = (4M^{1/2}\tilde{\sigma}_\mu \eta^{3/2})^{-1} K_\infty^{1/2} \min\{1, n^{3\alpha/2-1}\}$, note that $\eta_n^{1/2} - \eta_{n+1}^{1/2} \ge \eta^{1/2}/(4n^{\alpha/2})$, we then have for all $n > 0$,

$$\begin{aligned}
\eta_n^{1/2}(1+\varepsilon_n)^{1/2} - \eta_{n+1}^{1/2}(1+\varepsilon_{n+1})^{1/2} &\ge \eta_n^{1/2}(1+\varepsilon_{n+1})^{1/2} - \eta_{n+1}^{1/2}(1+\varepsilon_{n+1})^{1/2} \\
&\ge \frac{\eta^{1/2}}{4n^{\alpha/2}}(1+\varepsilon_{n+1})^{1/2} \ge \frac{\eta^{1/2}}{4n^{\alpha/2}}, \\
\varepsilon_n M^{1/2}\tilde{\sigma}_\mu \eta_n^2 K_\infty^{-1/2}/\sqrt{2} = \frac{\eta^{1/2}\min\{1, n^{3\alpha/2-1}\}}{4\sqrt{2}n^{2\alpha}} &\le \frac{\eta^{1/2}}{4\sqrt{2}n^{\alpha/2}} \\
&\le \eta_n^{1/2}(1+\varepsilon_n)^{1/2} - \eta_{n+1}^{1/2}(1+\varepsilon_{n+1})^{1/2}.
\end{aligned}$$

Let n_0 be the smallest n such that $\widetilde{D}_{n-1}^2 \ge (1+\varepsilon_n)M\eta_n\tilde{\sigma}_\mu^2 K_\infty/2$. Next, we show for all $n \ge n_0$, such that $\widetilde{D}_{n-1}^2 \ge (1+\varepsilon_n)M\eta_n\tilde{\sigma}_\mu^2 K_\infty/2$. In fact, if $\widetilde{D}_{n-1} \ge (1+\varepsilon_n)^{1/2}M^{1/2}\eta_n^{1/2}\tilde{\sigma}_\mu K_\infty^{1/2}/\sqrt{2}$, due to the fact that $g(t)$ is increasing in $\widetilde{D}_{n-1}$, we have

$$\begin{aligned}
\widetilde{D}_n &\ge (1+\varepsilon_n)^{1/2}M^{1/2}\eta_n^{1/2}\tilde{\sigma}_\mu K_\infty^{1/2}/\sqrt{2} - \frac{\eta_n}{2K_\infty}\{(1+\varepsilon_n)^{1/2}M^{1/2}\eta_n^{1/2}\tilde{\sigma}_\mu K_\infty^{1/2}\}^2 + \frac{\eta_n^2 M\tilde{\sigma}_\mu^2}{2} \\
&\ge (1+\varepsilon_{n+1})^{1/2}M^{1/2}\eta_{n+1}^{1/2}\tilde{\sigma}_\mu K_\infty^{1/2}/\sqrt{2} - \varepsilon_n\eta_n^2 M\tilde{\sigma}_\mu^2/2 \\
&\quad + M^{1/2}\tilde{\sigma}_\mu K_\infty^{1/2}\{\eta_n^{1/2}(1+\varepsilon_n)^{1/2} - \eta_{n+1}^{1/2}(1+\varepsilon_{n+1})^{1/2}\}/\sqrt{2} \\
&\ge (1+\varepsilon_{n+1})^{1/2}M^{1/2}\eta_{n+1}^{1/2}\tilde{\sigma}_\mu K_\infty^{1/2}/\sqrt{2}.
\end{aligned}$$

Correspondingly, the recursion for $n > n_0$ can be updated by

$$\widetilde{D}_n \le \widetilde{D}_{n-1} - \frac{\eta_n}{K_\infty}\widetilde{D}_{n-1}^2 \frac{\varepsilon_n}{1+\varepsilon_n}.$$

We then obtain

$$\frac{1}{\widetilde{D}_{n-1}} \le \frac{1}{\widetilde{D}_n} - \frac{\eta_n}{K_\infty}\frac{\widetilde{D}_{n-1}}{\widetilde{D}_n}\frac{\varepsilon_n}{1+\varepsilon_n} \le \frac{1}{\widetilde{D}_n} - \frac{\eta_n}{K_\infty}\frac{\varepsilon_n}{1+\varepsilon_n}.$$

By summing $n - n_0$ times, we get

$$\frac{1}{\widetilde{D}_{n_0}} \le \frac{1}{\widetilde{D}_n} - \frac{1}{K_\infty}\sum_{k=n_0+1}^{n}\frac{\varepsilon_k}{1+\varepsilon_k}\eta_k.$$

Therefore,

$$\begin{aligned}
\widetilde{D}_n &\le \frac{1}{\frac{1}{K_\infty}\sum_{k=n_0+1}^{n}\frac{\varepsilon_k}{1+\varepsilon_k}\eta_k + \frac{1}{\widetilde{D}_{n_0}}} \\
&\le \frac{1}{\frac{1}{K_\infty}\sum_{k=n_0+1}^{n}\frac{\varepsilon_k}{1+\varepsilon_k}\eta_k + \sqrt{2}(1+\varepsilon_{n_0})^{-1/2}M^{-1/2}\eta_{n_0}^{-1/2}\tilde{\sigma}_\mu^{-1}K_\infty^{-1/2}}.
\end{aligned}$$

Moreover, it is easy to show that

$$\begin{aligned}
\eta_n^{-1/2}\left(1+\varepsilon_n\right)^{-1/2} &= \sum_{k=1}^{n-1}\left\{\eta_{k+1}^{-1/2}\left(1+\varepsilon_{k+1}\right)^{-1/2}-\eta_k^{-1/2}\left(1+\varepsilon_k\right)^{-1/2}\right\}+\eta_1^{-1/2}\left(1+\varepsilon_1\right)^{-1/2} \\
&= \sum_{k=1}^{n-1}\frac{\eta_k^{1/2}\left(1+\varepsilon_k\right)^{1/2}-\eta_{k+1}^{1/2}\left(1+\varepsilon_{k+1}\right)^{1/2}}{\eta_k^{1/2}\left(1+\varepsilon_k\right)^{1/2}\eta_{k+1}^{1/2}\left(1+\varepsilon_{k+1}\right)^{1/2}}+\eta_1^{-1/2}\left(1+\varepsilon_1\right)^{-1/2} \\
&\geq \sum_{k=1}^{n-1}\left(1+\varepsilon_{k+1}\right)^{-1}\eta_k^{-1/2}\eta_{k+1}^{-1/2}\left(\eta_k^{1/2}-\eta_{k+1}^{1/2}\right)+\eta_1^{-1/2}\left(1+\varepsilon_1\right)^{-1/2} \\
&\geq \frac{1}{2}\sum_{k=1}^{n}\eta_k\frac{\varepsilon_k}{1+\varepsilon_k}M^{1/2}\tilde{\sigma}_\mu K_\infty^{-1/2}.
\end{aligned}$$

Finally, we obtain

$$D_n \leq \widetilde{D}_n \leq \frac{1}{\frac{1}{2K_\infty}\sum_{k=1}^n \frac{\varepsilon_k}{1+\varepsilon_k}\eta_k}.$$

When $1/2 < \alpha \leq 2/3$, we have

$$\begin{aligned}
D_n &\leq \frac{2K_\infty}{\sum_{k=1}^n \frac{\eta k^{-\alpha}}{1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2}K_\infty^{-1/2}k^{1-3\alpha/2}}} \\
&= \frac{2K_\infty^{1/2}(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{\sum_{k=1}^n k^{\alpha/2-1}} \leq \frac{2K_\infty^{1/2}(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{\psi_{\alpha/2}(n)} \\
&\leq \frac{2\{\Delta_0+\tilde{\sigma}_\mu^2\eta^2\psi_{1-2\alpha}(n)\}^{1/2}(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{\psi_{\alpha/2}(n)}.
\end{aligned}$$

When $2/3 < \alpha < 1$, we have

$$\begin{aligned}
D_n &\leq \frac{2K_\infty}{\sum_{k=1}^n \frac{\eta k^{-\alpha}}{1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2}K_\infty^{-1/2}}} \\
&= \frac{2K_\infty^{1/2}(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{\sum_{k=1}^n k^{-\alpha}} \leq \frac{2K_\infty^{1/2}(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{\psi_{1-\alpha}(n)} \\
&\leq \frac{2\{\Delta_0+\tilde{\sigma}_\mu^2\eta^2\psi_{1-2\alpha}(n)\}^{1/2}(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{\psi_{1-\alpha}(n)}.
\end{aligned}$$

For $\mathbf{0 \leq \alpha \leq 1/2}$, $\sum_{k=1}^\infty \eta_k^2$ diverges, then we consider the following equality:

$$\widetilde{D}_n = \widetilde{D}_{n-1} - \frac{\eta_n}{K_n}\widetilde{D}_{n-1}^2 + \frac{\eta_n^2 M\tilde{\sigma}_\mu^2}{2},$$

with $K_n \leq \Delta_0 + \tilde{\sigma}_\mu^2\eta^2\psi_{1-2\alpha}(n)$. We follow the same procedure above and get

$$\begin{aligned}
D_n \leq \widetilde{D}_n &\leq \frac{1}{\frac{1}{2}\sum_{k=1}^n \frac{1}{K_k}\frac{\varepsilon_k}{1+\varepsilon_k}\eta_k} \\
&\leq \frac{2(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{\sum_{k=1}^n K_k^{-1/2}k^{\alpha/2-1}} \\
&\leq \frac{2(\Delta_0+\tilde{\sigma}_\mu^2\eta^2)(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{(1-2\alpha)^{1/2}\sum_{k=1}^n k^{3(\alpha-1)/2}} \\
&\leq \frac{2(\Delta_0+\tilde{\sigma}_\mu^2\eta^2)(1+4M^{1/2}\tilde{\sigma}_\mu\eta^{3/2})/\eta}{(1-2\alpha)^{1/2}\psi_{3\alpha/2-1/2}(n)}.
\end{aligned}$$

We have completed the proof of this theorem. $\square$

Proof of Theorem 6. By convexity of $F(\theta)$, we have

$$F\left(\frac{1}{n}\sum_{k=1}^{n}\theta_k\right) \leq \frac{1}{n}\sum_{k=1}^{n}F(\theta_k).$$

The convexity also implies $F(\theta_k) - F(\theta^*) \leq \langle \nabla F(\theta_k), \theta_k - \theta^* \rangle$. From the previous inequality, we then have

$$E(\|\theta_k - \theta^*\|^2 \mid \mathcal{F}_{k-1}) \leq \|\theta_{k-1} - \theta^*\|^2 - 2\eta_n E\{\langle \theta_{k-1} - \theta^*, \nabla F(\theta_{k-1})\rangle\} + \eta_n^2(C_0^2 + \tau_k^2).$$

Therefore,

$$\begin{aligned}
E\left\{F\left(n^{-1}\sum_{k=1}^{n}\theta_k\right) - F(\theta^*)\right\} &\leq E\left[\frac{1}{n}\sum_{k=1}^{n}\{F(\theta_k) - F(\theta^*)\}\right] \\
&\leq E\left\{\frac{1}{n}\sum_{k=1}^{n}\langle \nabla F(\theta_k), \theta_k - \theta^*\rangle\right\} \\
&\leq \frac{1}{2n}\sum_{k=1}^{n}\left(\frac{\Delta_k - \Delta_{k-1}}{\eta_k} + \tilde{\sigma}_\mu^2\sum_{k=0}^{n}\eta_k\right) \\
&\leq \frac{\Delta_n}{2n\eta_n} + \frac{\tilde{\sigma}_\mu^2\eta\psi_{1-\alpha}(n)}{2n} \\
&= \frac{n^{\alpha-1}}{2\eta}\{\Delta_0 + \tilde{\sigma}_\mu^2\eta^2\psi_{1-2\alpha}(n)\} + \frac{\tilde{\sigma}_\mu^2\eta\psi_{1-\alpha}(n)}{2n}.
\end{aligned}$$

Hence, the proof of this theorem is completed. □

REFERENCES

Phillippe Rigollet and Jan-Christian Hütter. High dimensional statistics. *Lecture notes for course 18S997*, 813(814):46, 2015.

B ADDITIONAL EXPERIMENTS RESULTS

In this section, we showcase the remaining numerical results for both linear and logistic regressions, maintaining consistent configurations with dimensions $d = 10, 20$. These findings align with the trends discussed in the main body of the paper.

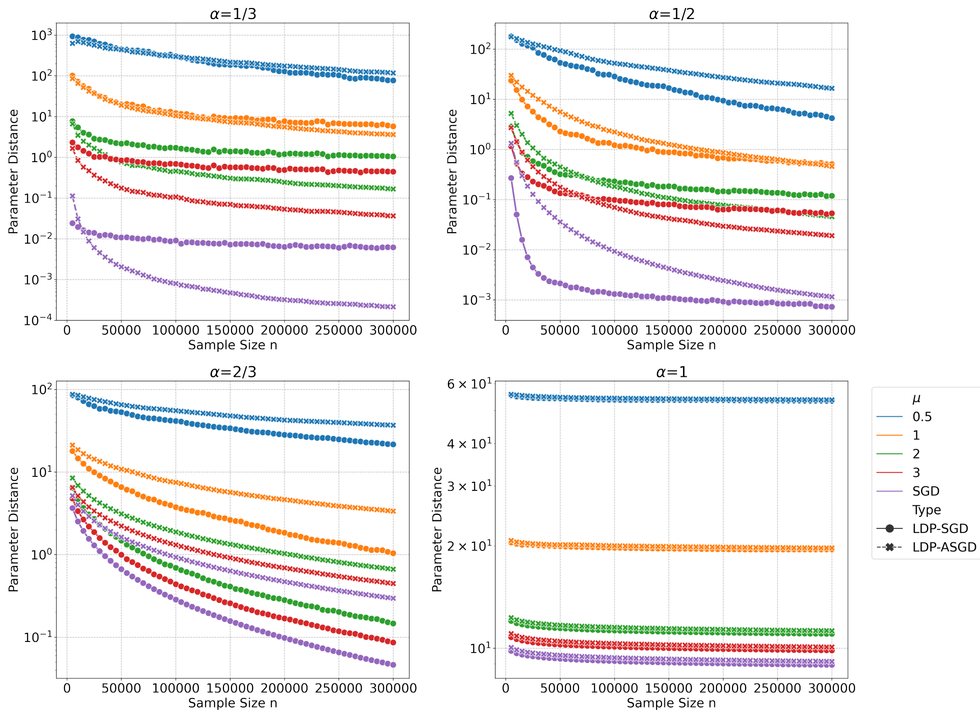

Figure 1: Trajectories of the distance between DP-SGD estimators and the optimal for linear regression with $d = 10$.

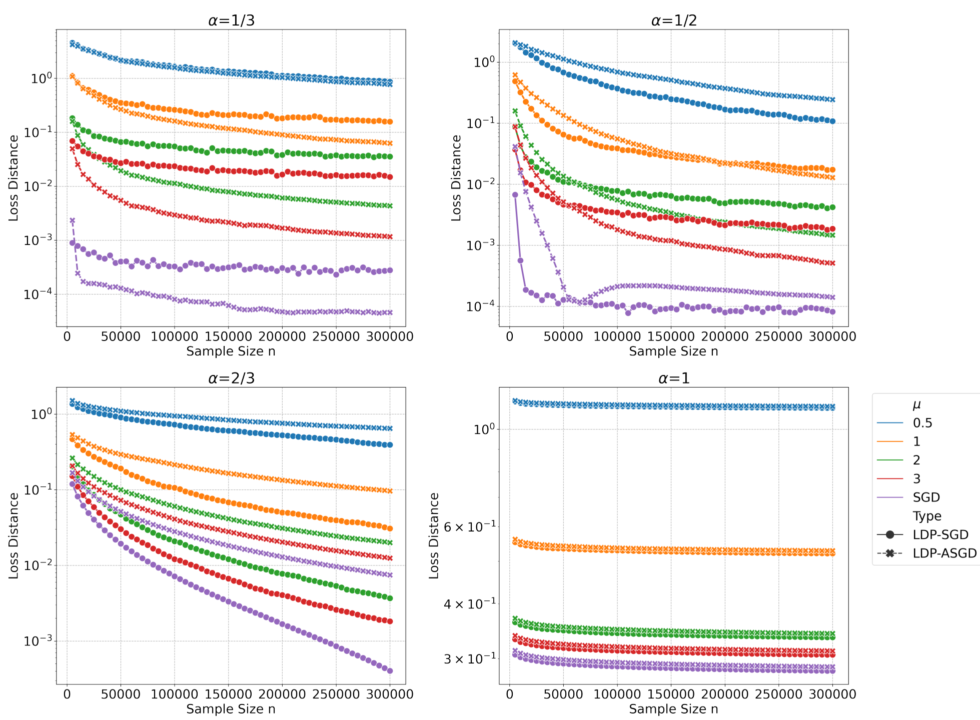

Figure 2: Trajectories of the distance between the loss incurred by the estimators and the optimal loss for linear regression with $d = 10$.

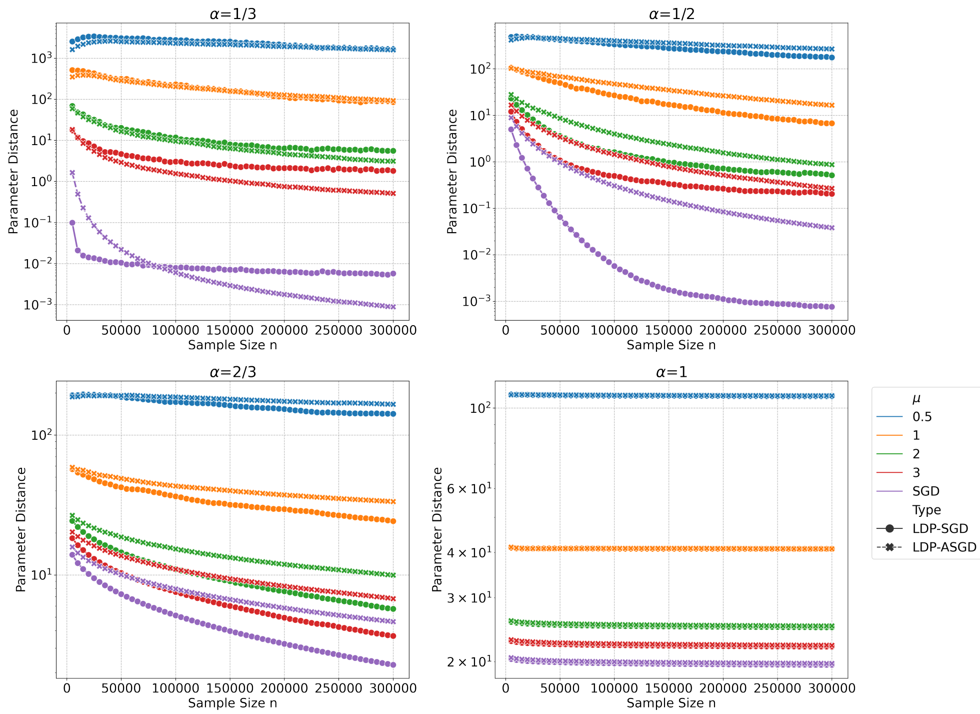

Figure 3: Trajectories of the distance between DP-SGD estimators and the optimal for linear regression with $d = 20$.

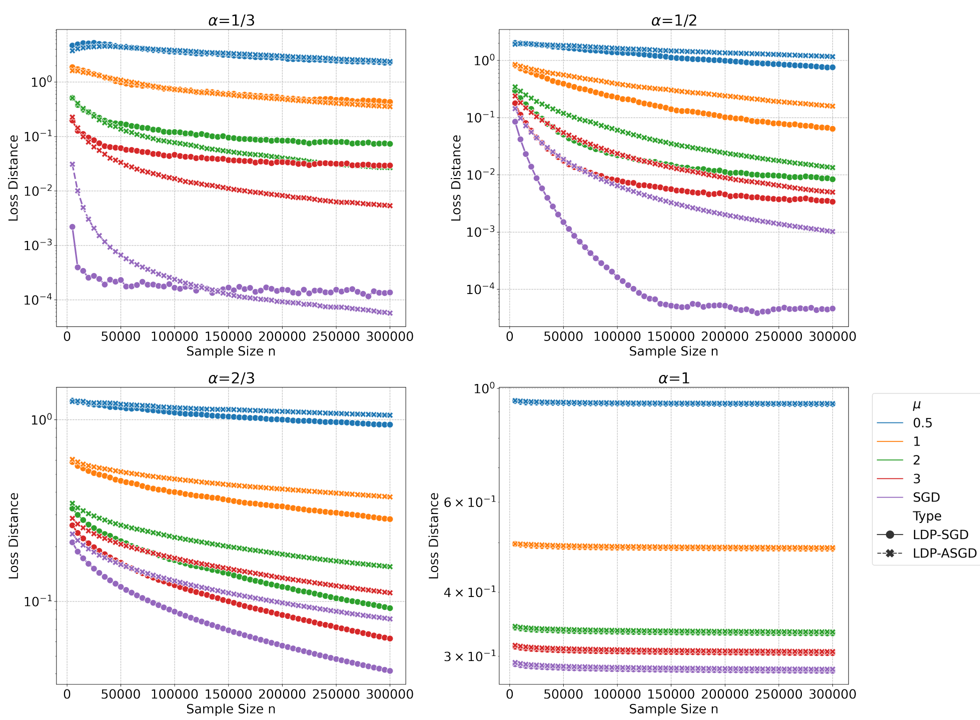

Figure 4: Trajectories of the distance between the loss incurred by the estimators and the optimal loss for linear regression with $d = 20$.

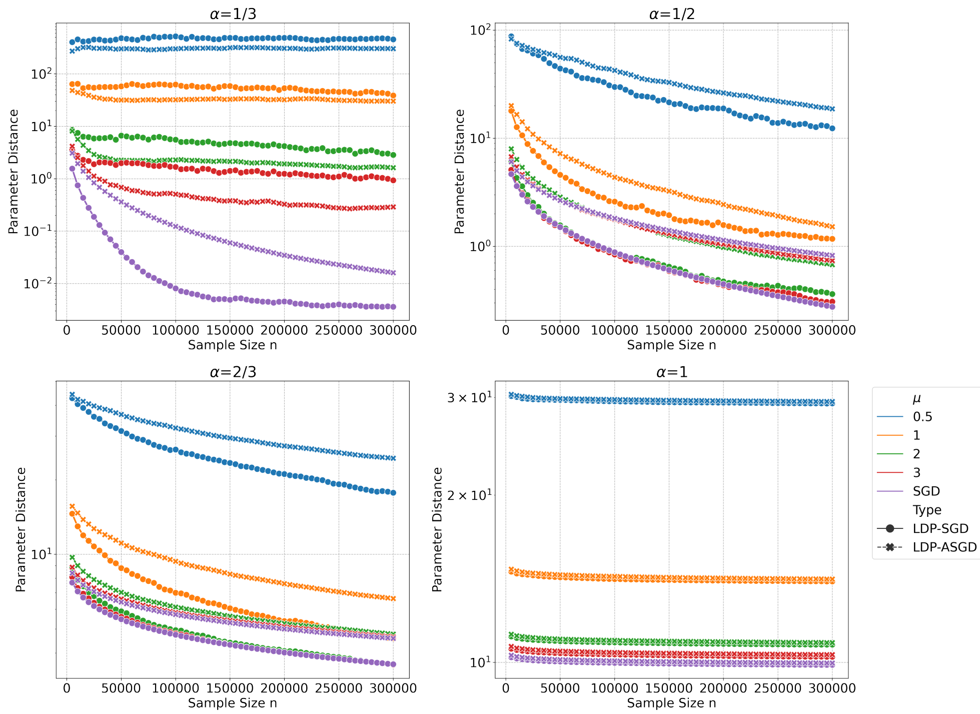

Figure 5: Trajectories of the distance between DP-SGD estimators and the optimal for logistic regression with $d = 10$.

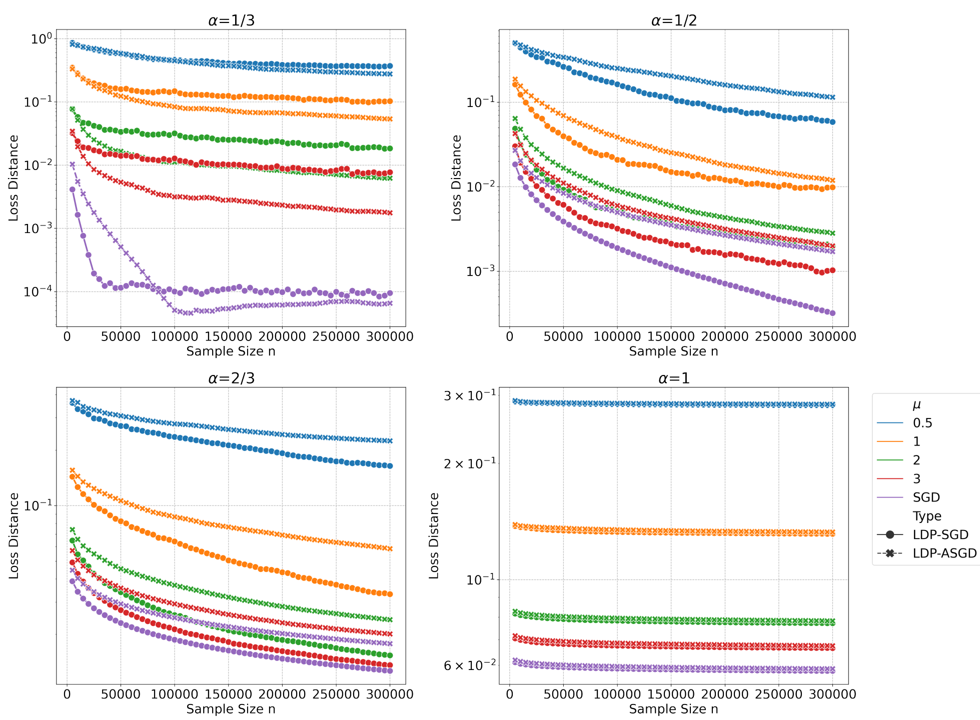

Figure 6: Trajectories of the distance between the loss incurred by the estimators and the optimal loss for logistic regression with $d = 10$.

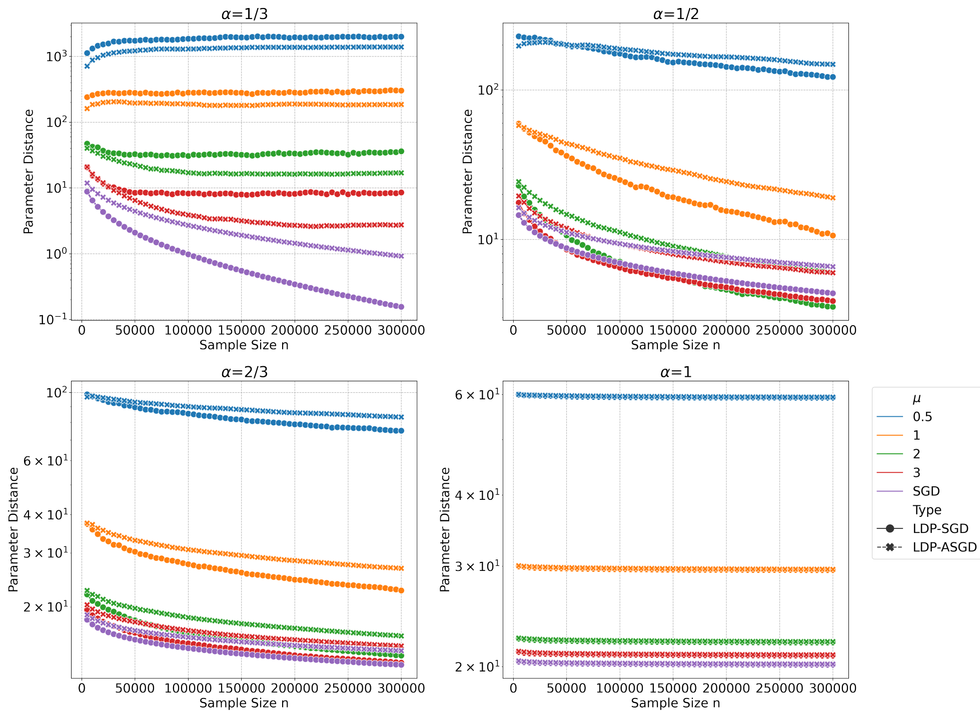

Figure 7: Trajectories of the distance between DP-SGD estimators and the optimal for logistic regression with $d = 20$.

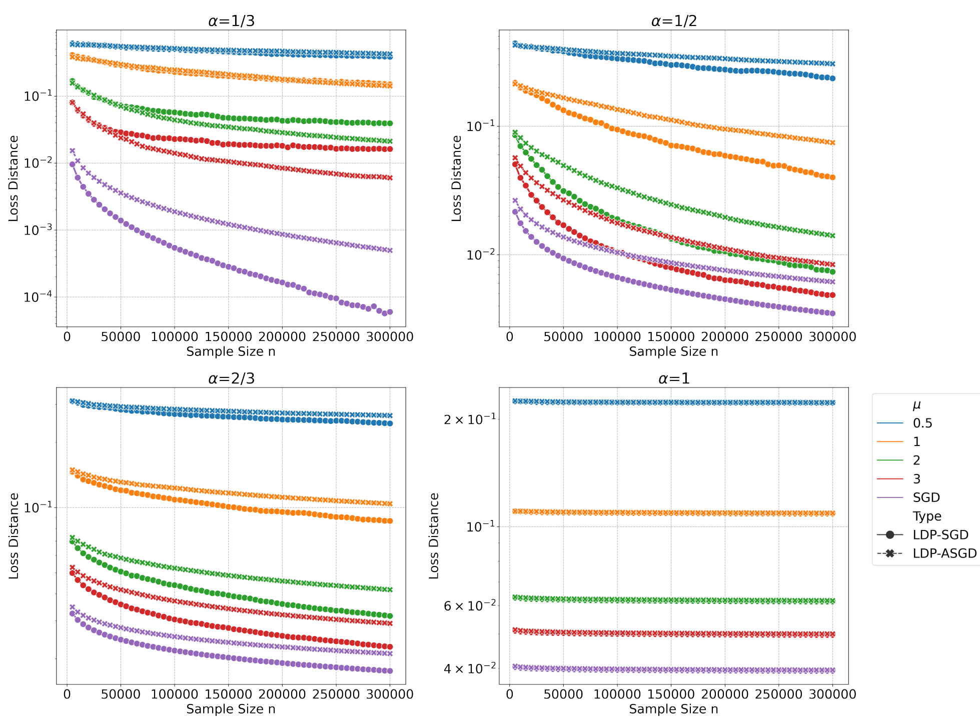

Figure 8: Trajectories of the distance between the loss incurred by the estimators and the optimal loss for logistic regression with $d = 20$.