Supplementary Material for 'Efficient Private Empirical Risk Minimization for High-dimensional Learning'

Shiva Prasad Kasiviswanathan KASIVISW@GMAIL.COM
Samsung Research America, Mountain View, CA 94043

Hongxia Jin HONGXIA.JIN@SAMSUNG.COM
Samsung Research America, Mountain View, CA 94043

A. Proof of Claim 2.1

Proof of Claim 2.1. Consider $\theta_a, \theta_b \in \mathcal{C}$, $\mathbf{x} \in \mathbb{R}^d$ with $\|\mathbf{x}\| \leq 1$, $y \in \mathbb{R}$ with $|y| \leq 1$,

$$\begin{aligned}&|\ell(\langle \mathbf{x}, \theta_a \rangle; y) - \ell(\langle \mathbf{x}, \theta_b \rangle; y)| \\ &= |\ell(\langle \mathbf{x}, \theta_a \rangle; y) - \ell(\langle \mathbf{x}, \theta_a \rangle + \langle \mathbf{x}, \theta_b - \theta_a \rangle; y)| \\ &\leq |\lambda_\ell \langle \mathbf{x}, \theta_b - \theta_a \rangle| \leq \lambda_\ell \|\theta_a - \theta_b\|.\end{aligned}$$

Since this holds for every $\mathbf{x}$ and y in the chosen domain, this completes the proof. □

B. Missing Proofs from Section 3

Proof of Lemma 3.3. We first investigate the function $\ell(\langle \Phi \mathbf{x}_i, \vartheta \rangle; y_i)$. Consider $\vartheta_a, \vartheta_b \in \Phi\mathcal{C}$,

$$\begin{aligned}&|\ell(\langle \Phi\mathbf{x}_i, \vartheta_a \rangle; y_i) - \ell(\langle \Phi\mathbf{x}_i, \vartheta_b \rangle; y_i)| \\ &= |\ell(\langle \Phi\mathbf{x}_i, \vartheta_a \rangle; y_i) - \ell(\langle \Phi\mathbf{x}_i, \vartheta_a \rangle + \langle \Phi\mathbf{x}_i, \vartheta_b - \vartheta_a \rangle; y_i)| \\ &\leq |\lambda_\ell \langle \Phi\mathbf{x}_i, \vartheta_b - \vartheta_a \rangle|.\end{aligned}$$

Using Theorem 3.1, if

$$m = \Theta((\psi^4/\gamma^2) \max\{\log n, \log(1/\beta)\}),$$

then with probability at least $1-\beta$, $\|\Phi\mathbf{x}_i\| \leq (1+\gamma)\|\mathbf{x}_i\| \leq 2\|\mathbf{x}_i\| \leq 2$. Therefore, with probability at least $1-\beta$,

$$|\ell(\langle \Phi\mathbf{x}_i, \vartheta_a \rangle; y_i) - \ell(\langle \Phi\mathbf{x}_i, \vartheta_b \rangle; y_i)| \leq 2\lambda_\ell \|\vartheta_b - \vartheta_a\|.$$

Taking a union bound over all i's, with probability at least $1-\beta n$, for all $i \in [n]$, $|\lambda_\ell \langle \Phi\mathbf{x}_i, \vartheta_b - \vartheta_a \rangle| \leq 2\lambda_\ell \|\vartheta_b - \vartheta_a\|$.

Replacing β by β/n gives the proof. □

Proof of Lemma 3.4. If S is the set of points $\mathbf{x}_1, \dots, \mathbf{x}_n, \hat{\theta}$, then as mentioned above $w(S) = O(\sqrt{\log n})$. For any $i \in [n]$, using Corollary 3.2, with probability at least $1-\beta$,

$$\ell(\langle \Phi\mathbf{x}_i, \Phi\hat{\theta} \rangle; y_i) \leq \ell(\langle \mathbf{x}_i, \hat{\theta} \rangle; y_i) + \lambda_\ell \gamma \|\mathcal{C}\|_2. \quad (5)$$

Taking a union bound over all i's and replacing β by β/n completes the proof. □

Proof of Lemma 3.5. Since $\hat{\theta} \in \mathcal{C}$, by definition,

$$\begin{aligned}&\min_{\theta \in \mathcal{C}} \mathcal{L}_{\text{comp}}(\theta; (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n); \Phi) \\ &\qquad \leq \mathcal{L}_{\text{comp}}(\hat{\theta}; (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n); \Phi). \quad (6)\end{aligned}$$

From Lemma 3.4, with probability at least $1-\beta$,

$$\begin{aligned}&\mathcal{L}_{\text{comp}}(\hat{\theta}; (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n); \Phi) \\ &\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \ell(\langle \Phi\mathbf{x}_i, \Phi\hat{\theta} \rangle; y_i) \\ &\leq \frac{1}{n}\sum_{i=1}^n \ell(\langle \mathbf{x}_i, \hat{\theta} \rangle; y_i) + \lambda_\ell \gamma \|\mathcal{C}\|_2,\end{aligned}$$

where we used the fact that $\|\hat{\theta}\| \leq \|\mathcal{C}\|_2$. Using the above inequality with (6) completes the proof. □

Proof of Proposition 3.7. We discuss the proof for the case of (ϵ, δ)-differential privacy (the proof for the case of ϵ-differential privacy proceeds similarly by using Theorem 3.6, Part 2).

Since the inputs $\Phi\mathbf{x}_i$'s are m-dimensional, from guarantees of Theorem 3.6 (Part 1), we know that with probability at least $1-\beta$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \ell(\langle \Phi\mathbf{x}_i, \vartheta^{\text{priv}} \rangle; y_i) - \\ &\qquad \min_{\vartheta \in \Phi\mathcal{C}} \frac{1}{n}\sum_{i=1}^n \ell(\langle \Phi\mathbf{x}_i, \vartheta \rangle; y_i) \\ &= O\left(\frac{\lambda_{\mathcal{L}_{\text{comp}}} \sqrt{m} \|\Phi\mathcal{C}\|_2 \log^{3/2}(n/\delta) \sqrt{\log(\frac{1}{\delta})}\, \text{polylog}(\frac{1}{\beta})}{n\epsilon}\right).\end{aligned} \quad (7)$$

Notice that by definition,

$$\min_{\vartheta\in\Phi\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n}\ell(\langle\Phi\mathbf{x}_i,\vartheta\rangle;y_i) \equiv \min_{\theta\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n}\ell(\langle\Phi\mathbf{x}_i,\Phi\theta\rangle;y_i).$$

Also by construction in Step 2 of Mechanism ProjERM, $\vartheta^{\text{priv}} = \Phi\theta^{\text{priv}}$. Substituting these two identities in (7) provides, that with probability at least $1-\beta$,

$$\frac{1}{n}\sum_{i=1}^{n}\ell(\langle\Phi\mathbf{x}_i,\Phi\theta^{\text{priv}}\rangle;y_i) - \min_{\theta\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n}\ell(\langle\Phi\mathbf{x}_i,\Phi\theta\rangle;y_i) = O\Big(\frac{\lambda_{\mathcal{L}_{\text{comp}}}\sqrt{m}\|\Phi\mathcal{C}\|_2\log^{3/2}(n/\delta)\sqrt{\log(1/\delta)}\,\text{polylog}(1/\beta)}{n\epsilon}\Big).$$

Using the bound from Lemma 3.5, gives that with probability at least $1-2\beta$,

$$\frac{1}{n}\sum_{i=1}^{n}\ell(\langle\Phi\mathbf{x}_i,\Phi\theta^{\text{priv}}\rangle;y_i) - \min_{\theta\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n}\ell(\langle\Phi\mathbf{x}_i,\Phi\theta\rangle;y_i) = O\left(\frac{\lambda_{\mathcal{L}_{\text{comp}}}\sqrt{m}\|\Phi\mathcal{C}\|_2\log^{3/2}(n/\delta)\sqrt{\log(1/\delta)}\,\text{polylog}(1/\beta)}{n\epsilon}\right) + O(\lambda_\ell\gamma\|\mathcal{C}\|_2).$$

Replacing β by $\beta/2$ completes the proof. $\square$

Proof of Lemma 3.10. Consider the set S as $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}\cup\mathcal{C}$, then $w(S)\leq w(\mathcal{C})+\sqrt{\log n}$. Using an argument similar to Lemma 3.4 (based on Theorem 3.1), for $\theta^{\text{priv}}\in\mathcal{C}$, with probability at least $1-\beta$,

$$\begin{aligned}&\mathcal{L}_{\text{comp}}(\theta^{\text{priv}};(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n);\Phi)\\ &\stackrel{\text{def}}{=}\frac{1}{n}\sum_{i=1}^{n}\ell(\langle\Phi\mathbf{x}_i,\Phi\theta^{\text{priv}}\rangle;y_i)\\ &\geq \frac{1}{n}\sum_{i=1}^{n}\ell(\langle\mathbf{x}_i,\theta^{\text{priv}}\rangle;y_i)-\lambda_\ell\gamma\|\mathcal{C}\|_2.\end{aligned}$$

This implies that with probability at least $1-\beta$,

$$\mathcal{L}_{\text{comp}}(\theta^{\text{priv}};(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n);\Phi) \geq \frac{1}{n}\sum_{i=1}^{n}\ell(\langle\mathbf{x}_i,\theta^{\text{priv}}\rangle;y_i)-\lambda_\ell\gamma\|\mathcal{C}\|_2.$$

Using the definition of $\mathcal{L}(\theta^{\text{priv}};(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n))$ completes the proof. $\square$

Proof of Theorem 3.11. We first discuss the proof for the case of (ϵ,δ)-differential privacy (Part 1). Here, we set

$$\gamma = \frac{\psi\sqrt{w(\mathcal{C})}}{\sqrt{\epsilon n}},$$

and correspondingly set m as,

$$m=\Theta\left(\frac{\psi^2\epsilon n(w(\mathcal{C})+\sqrt{\log n})^2\log(n/\beta)}{w(\mathcal{C})}\right).$$

With the choice of m, with probability $1-\beta$, the diameter of $\Phi\mathcal{C}$ ($\|\Phi\mathcal{C}\|_2$) is at most $(1+\gamma)\|\mathcal{C}\|_2\leq 2\|\mathcal{C}\|_2$. Using this, along with Lemma 3.10 and Proposition 3.7 (Part 1) gives that with probability at least $1-3\beta$,

$$\mathcal{L}(\theta^{\text{priv}};(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n))-\mathcal{L}(\hat{\theta};(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n)) = O\left(\frac{\lambda_{\mathcal{L}_{\text{comp}}}\sqrt{m}\|\mathcal{C}\|_2\log^{3/2}(n/\delta)\sqrt{\log(\frac{1}{\delta})}\,\text{polylog}(\frac{1}{\beta})}{n\epsilon}\right) + O\left(\lambda_\ell\gamma\|\mathcal{C}\|_2\right).$$

Using the bound on $\lambda_{\mathcal{L}_{\text{comp}}}$ from Lemma 3.3 gives that with probability at least $1-4\beta$,

$$\mathcal{L}(\theta^{\text{priv}};(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n))-\mathcal{L}(\hat{\theta};(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n)) = O\left(\frac{\lambda_\ell\sqrt{m}\|\mathcal{C}\|_2\log^{3/2}(n/\delta)\sqrt{\log(\frac{1}{\delta})}\,\text{polylog}(\frac{1}{\beta})}{n\epsilon}\right) + O\left(\lambda_\ell\gamma\|\mathcal{C}\|_2\right).$$

Replacing β by $\beta/5$, and simplifying the resulting expression completes the Part 1 of the proof.

For Part 2, we set $\gamma=\frac{\psi^{4/3}w(\mathcal{C})^{2/3}}{(\epsilon n)^{1/3}}$ and correspondingly set m as,

$$m=\Theta\left(\frac{\psi^{4/3}(n\epsilon)^{2/3}(w(\mathcal{C})+\sqrt{\log n})^2\log(n/\beta)}{w(\mathcal{C})^{4/3}}\right).$$

The proof follows along the same lines as the case of (ϵ,δ)-differential privacy. Here we use Proposition 3.7 (Part 2). $\square$