Recall. Under the assumptions on the non-convex regularizer $R_\lambda$,

Proposition 1:

(i) $R_\lambda$ is $\lambda L$-Lipschitz; (ii) $\forall\, \theta \in \mathbb{R}^d$, $\lambda L \|\theta\|_1 \leq R_\lambda(\theta) + \frac{\mu}{2}\|\theta\|_2^2$

Proposition 2: Let $\theta \in \mathbb{R}^d$ and $S$ be the index set of the $k$ largest elements of $\theta$ in magnitude. Then, for any $\xi > 0$,

$$(1) \text{ — } \quad \xi R_\lambda(\theta_S) - R_\lambda(\theta_{S^c}) \leq \lambda L\left(\xi \|\theta_S\|_1 - \|\theta_{S^c}\|_1\right)$$

Moreover, if $\theta^* \in \mathbb{R}^d$ is $k$-sparse, then for any $\theta \in \mathbb{R}^d$,

$$(2) \text{ — } \quad R_\lambda(\theta^*) - R_\lambda(\theta) \leq \lambda L\left(\|\Delta_T\|_1 - \|\Delta_{T^c}\|_1\right),$$

where $\Delta = \theta - \theta^*$ and $T$ is the index set of the $k$ largest elements in $\Delta$ in magnitude.

Consider the problem

$$\min_{\|\theta\|_1 \leq R} \{ \underbrace{\mathcal{L}(\theta) + R_\lambda(\theta)}_{=F(\theta)} \}, \qquad \text{— } (*)$$

where $R_\lambda$ satisfies the assumptions. Since $(*)$ is non-convex in general, we aim to bound the estimation error $\|\tilde{\theta} - \theta^*\|_2$ for any $\tilde{\theta} \in \mathbb{R}^d$ satisfying the first-order condition

$$\underbrace{\left(\nabla \mathcal{L}(\tilde{\theta}) + \nabla R_\lambda(\tilde{\theta})\right)^T (\theta - \tilde{\theta})}_{\text{directional derivative of } F \text{ in direction } \theta - \tilde{\theta}} \geq 0, \qquad \forall\, \|\theta\|_1 \leq R$$

$$\left( f'(\theta, d) = \lim_{t \downarrow 0} \frac{f(\theta + td) - f(\theta)}{t} \right)$$

Remark: For simplicity, we assume that $R_\lambda$ is differentiable everywhere, though I suspect the argument can still go through for $R_\lambda$ non-smooth at 0.

To bound $\|\tilde{\theta} - \theta^*\|_2$, we need the following notion.

Definition: We say that $\mathcal{L}$ satisfies the RSC condition with parameters $\alpha_1, \alpha_2 > 0$; $\tau_1, \tau_2 \geq 0$ if

$$(\nabla\mathcal{L}(\theta^* + \Delta) - \nabla\mathcal{L}(\theta^*))^T \Delta \geq \begin{cases} \alpha_1 \cdot \|\Delta\|_2^2 - \tau_1 \frac{\log d}{n} \|\Delta\|_1^2 & \text{if } \|\Delta\|_2 \leq 1, \\ \alpha_2 \cdot \|\Delta\|_2^2 - \underbrace{\tau_2 \frac{\log d}{n} \|\Delta\|_1}_{} & \text{if } \|\Delta\|_2 \geq 1. \end{cases}$$

adjust for the effects from probabilistic analysis of generative model

Theorem: Suppose that $\mathcal{L}$ is RSC as above and $R_\lambda$ satisfies the given assumptions with $\alpha_1 > \mu/2$. Suppose that $\theta^*$ is feasible for $(*)$. Provided that

(i) $$\frac{4}{L} \max\left\{ \|\nabla\mathcal{L}(\theta^*)\|_\infty, \alpha_2 \sqrt{\frac{\log d}{n}} \right\} \leq \lambda \leq \frac{\alpha_2}{6LR};$$

(effect of regularization)

(ii) $$n \geq \frac{16 R^2 \max\{\tau_1^2, \tau_2^2\}}{\alpha_2^2} \log d,$$ (sample complexity)

every first-order point $\tilde{\theta}$ of $(*)$ satisfies

$$\|\tilde{\theta} - \theta^*\|_2 \leq \frac{3\lambda L\sqrt{k}}{2\alpha_1 - \mu}, \quad k = \|\theta^*\|_0$$

Proof: Let $\tilde{\Delta} = \tilde{\theta} - \theta^*$.

Claim: $\|\tilde{\Delta}\|_2 \leq 1$.

Assuming the claim, we get from RSC condition that

$$(\nabla\mathcal{L}(\tilde{\theta}) - \nabla\mathcal{L}(\theta^*))^T \tilde{\Delta} \geq \alpha_1 \|\tilde{\Delta}\|_2^2 - \tau_1 \frac{\log d}{n} \|\tilde{\Delta}\|_1^2$$

Since $R_{\lambda,\mu}(\theta) = R_\lambda(\theta) + \frac{\mu}{2}\|\theta\|_2^2$ is convex,

$$\underset{\substack{\| \\ R_\lambda(\theta^*) + \frac{\mu}{2}\|\theta^*\|_2^2}}{R_{\lambda,\mu}(\theta^*)} - \underset{\substack{\| \\ R_\lambda(\tilde\theta) \\ + \frac{\mu}{2}\|\tilde\theta\|_2^2}}{R_{\lambda,\mu}(\tilde\theta)} \geq \nabla R_{\lambda,\mu}(\tilde\theta)^T(\theta^* - \tilde\theta) = (\nabla R_\lambda(\tilde\theta) + \mu\tilde\theta)^T(\theta^* - \tilde\theta).$$

It follows that

$$\nabla R_\lambda(\tilde\theta)^T \underbrace{(\theta^* - \tilde\theta)}_{-\tilde\Delta} \leq R_\lambda(\theta^*) - R_\lambda(\tilde\theta) + \frac{\mu}{2}\|\tilde\theta - \theta^*\|_2^2$$

By the first-order condition, we have

$$\alpha_1 \|\tilde\Delta\|_2^2 - \tau_1 \frac{\log d}{n}\|\tilde\Delta\|_1^2 \leq (\nabla\mathcal{L}(\tilde\theta) - \nabla\mathcal{L}(\theta^*))^T\tilde\Delta$$

$$\leq -\nabla R_\lambda(\tilde\theta)^T\tilde\Delta - \nabla\mathcal{L}(\theta^*)^T\tilde\Delta$$ (by first-order condition; $\theta = \theta^*$)

$$\leq -\nabla\mathcal{L}(\theta^*)^T\tilde\Delta + R_\lambda(\theta^*) - R_\lambda(\tilde\theta) + \frac{\mu}{2}\|\tilde\Delta\|_2^2$$

$$\leq \underbrace{\|\nabla\mathcal{L}(\theta^*)\|_\infty}_{\leq \lambda L/4 \text{ by assumption}} \cdot \|\tilde\Delta\|_1 + \lambda L(\|\tilde\Delta_S\|_1 - \|\tilde\Delta_{S^c}\|_1) + \frac{\mu}{2}\|\tilde\Delta\|_2^2$$

($S$ is the index set of the $k$ largest elements of $\tilde\Delta$ in magnitude; inequality follows from generalized Cauchy-Schwarz and Proposition 2)

$$\leq \frac{5\lambda L}{4}\|\tilde\Delta_S\|_1 - \frac{3\lambda L}{4}\|\tilde\Delta_{S^c}\|_1 + \frac{\mu}{2}\|\tilde\Delta\|_2^2$$ ($\|\nabla\mathcal{L}(\theta^*)\|_\infty \leq \frac{\lambda L}{4}$, $\|\tilde\Delta\|_1 = \|\tilde\Delta_S\|_1 + \|\tilde\Delta_{S^c}\|_1$)

Then,

$$(\alpha_1 - \frac{\mu}{2})\|\tilde\Delta\|_2^2 \leq \frac{5\lambda L}{4}\|\tilde\Delta_S\|_1 - \frac{3\lambda L}{4}\|\tilde\Delta_{S^c}\|_1 + \tau_1\frac{\log d}{n}\|\tilde\Delta\|_1^2$$

$$\leq \frac{5\lambda L}{4}\|\tilde\Delta_S\|_1 - \frac{3\lambda L}{4}\|\tilde\Delta_{S^c}\|_1 + 2R\tau_1\frac{\log d}{n}\|\tilde\Delta\|_1$$ ($\because \|\tilde\Delta\|_1 = \|\tilde\theta - \theta^*\|_1 \leq \|\tilde\theta\|_1 + \|\theta^*\|_1 \leq 2R$)

$$\leq \frac{5\lambda L}{4}\|\tilde\Delta_S\|_1 - \frac{3\lambda L}{4}\|\tilde\Delta_{S^c}\|_1 + \alpha_2\sqrt{\frac{\log d}{n}}\|\tilde\Delta\|_1$$

(by assumption on $\alpha_2$ in (ii))

$$\leq \frac{3\lambda L}{2}\|\tilde\Delta_S\|_1 - \frac{\lambda L}{2}\|\tilde\Delta_{S^c}\|_1$$ (by assumption on $\alpha_2$ in (i))

Finally,

$$2(\alpha_1 - \frac{\mu}{2})\|\tilde{\Delta}\|_2^2 \leq 3\lambda L\|\tilde{\Delta}_S\|_1 \leq 3\lambda L\sqrt{k}\|\tilde{\Delta}_S\|_2 \leq 3\lambda L\sqrt{k}\,\|\tilde{\Delta}\|_2$$

which gives the final result.

Proof of Claim:

Suppose that $\|\tilde{\Delta}\|_2 > 1$, where $\tilde{\Delta} = \tilde{\theta} - \theta^*$. By the RSC of $\mathcal{L}$,

$$(\nabla\mathcal{L}(\tilde{\theta}) - \nabla\mathcal{L}(\theta^*))^T\tilde{\Delta} \geq \alpha_2\|\tilde{\Delta}\|_2^2 - \tau_2\sqrt{\frac{\log d}{n}}\|\tilde{\Delta}\|_1$$

By the first-order condition of (*),

$$(\nabla\mathcal{L}(\tilde{\theta}) + \nabla R_\lambda(\tilde{\theta}))^T(-\tilde{\Delta}) \geq 0$$

Hence,

$$-(\nabla\mathcal{L}(\theta^*) + \nabla R_\lambda(\tilde{\theta}))^T\tilde{\Delta} \geq \alpha_2\|\tilde{\Delta}\|_2^2 - \tau_2\sqrt{\frac{\log d}{n}}\|\tilde{\Delta}\|_1$$

On the other hand,

$$-(\nabla\mathcal{L}(\theta^*) + \nabla R_\lambda(\tilde{\theta}))^T\tilde{\Delta} \leq \left(\|\nabla\mathcal{L}(\theta^*)\|_\infty + \|\nabla R_\lambda(\tilde{\theta})\|_\infty\right)\cdot\|\tilde{\Delta}\|_1$$

$$\leq \left(\frac{\lambda L}{4} + \lambda L\right)\|\tilde{\Delta}\|_1$$ (by assumption (i) and Proposition 1)

Hence,

$$\alpha_2\|\tilde{\Delta}\|_2^2 \leq \left(\frac{5\lambda L}{4} + \tau_2\sqrt{\frac{\log d}{n}}\right)\|\tilde{\Delta}\|_1$$

$$\leq 2R\left(\frac{5\lambda L}{4} + \tau_2\sqrt{\frac{\log d}{n}}\right)$$ ($\because \|\tilde{\Delta}\|_1 \leq 2R$)

By assumption (i), $\lambda LR \leq \frac{\alpha_2}{6} \Rightarrow \frac{5\lambda LR}{2} \leq \frac{5\alpha_2}{12}$.

Also, by assumption (ii), $2R\tau_2\sqrt{\frac{\log d}{n}} \leq \frac{1}{2}\alpha_2$.

Putting these together, $\|\tilde{\Delta}\|_2 \leq \sqrt{\frac{5}{12} + \frac{1}{2}} < 1$, a contradiction.