Theorem 9 (Proximal RRM Convergence) *Suppose the loss $\ell(z;\theta)$ is β-jointly smooth and γ-strongly convex. If the distribution map $D(\cdot)$ is ϵ-sensitive, then for*

$$G(\theta) \;=\; \arg\min_{\phi}\; \mathbb{E}_{z\sim D(\theta)}\big[\ell(z;\phi)\big] \;+\; \frac{\lambda}{2}\,\|\theta-\phi\|^2, \tag{54}$$

we have, for all $\theta,\theta'\in\Theta$,

$$\|G(\theta)-G(\theta')\| \;\le\; \frac{\epsilon\beta+\lambda}{\gamma+\lambda}\,\|\theta-\theta'\|.$$

Furthermore, if $\frac{\epsilon\beta+\lambda}{\gamma+\lambda} < 1$, then G is a contraction, possesses a unique fixed point θ_{PS}, and the proximal RRM iterates $\theta_{t+1}=G(\theta_t)$ converge linearly:

$$\|\theta_t-\theta_{PS}\| \;\le\; \left(\frac{\epsilon\beta+\lambda}{\gamma+\lambda}\right)^{t}\,\|\theta_0-\theta_{PS}\|.$$

Proof. Fix θ and θ'. Let

$$f(\phi)=\mathbb{E}_{z\sim D(\theta)}\ell(z;\phi)+\frac{\lambda}{2}\,\|\theta-\phi\|^2,\quad f'(\phi)=\mathbb{E}_{z\sim D(\theta')}\ell(z;\phi)+\frac{\lambda}{2}\,\|\theta'-\phi\|^2.$$

Because $\ell(z;\phi)$ is γ-strongly convex, both f and f' are $(\gamma+\lambda)$-strongly convex. Hence

$$f\big(G(\theta)\big)-f\big(G(\theta')\big)\ge \big(G(\theta)-G(\theta')\big)^{\top}\nabla_{\phi} f\big(G(\theta')\big)+\frac{\gamma+\lambda}{2}\,\|G(\theta)-G(\theta')\|^2, \tag{55}$$

$$f\big(G(\theta')\big)-f\big(G(\theta)\big)\ge \big(G(\theta')-G(\theta)\big)^{\top}\nabla_{\phi} f\big(G(\theta)\big)+\frac{\gamma+\lambda}{2}\,\|G(\theta')-G(\theta)\|^2. \tag{56}$$

Since f is minimized at $G(\theta)$, the inner product in 56 is non-negative. Combining 55 and 56 yields

$$\big(G(\theta)-G(\theta')\big)^{\top}\nabla_{\phi} f\big(G(\theta')\big) \;\le\; -(\gamma+\lambda)\,\|G(\theta)-G(\theta')\|^2. \tag{57}$$

Define the regularized loss

$$\ell_\theta(z;\phi)=\ell(z;\phi)+\frac{\lambda}{2}\,\|\theta-\phi\|^2.$$

The map

$$z \;\longmapsto\; \frac{\big(G(\theta)-G(\theta')\big)^{\top}\nabla_{\phi}\ell_\theta\big(z;G(\theta')\big)}{\beta\,\|G(\theta)-G(\theta')\|}$$

is 1-Lipschitz in z because of the β-joint smoothness of ℓ. The ϵ-sensitivity of $D(\cdot)$ then implies

$$\sup_{g\text{ is 1-Lip}}\big|\mathbb{E}_{z\sim D(\theta)}g(z)-\mathbb{E}_{z\sim D(\theta')}g(z)\big|\le \epsilon\,\|\theta-\theta'\|. \tag{58}$$

Using the 1-Lipschitz function above in 58 gives

$$\left|\tfrac{(G(\theta)-G(\theta'))^{\top}}{\beta\|G(\theta)-G(\theta')\|}\big(\mathbb{E}_{z\sim D(\theta)}\nabla_{\phi}\ell_\theta(z;G(\theta'))-\mathbb{E}_{z\sim D(\theta')}\nabla_{\phi}\ell_\theta(z;G(\theta'))\big)\right| \;\le\; \epsilon\|\theta-\theta'\|.$$

Unfolding ℓ_θ and rearranging, we obtain

$$\begin{aligned}-\epsilon\beta\|\theta-\theta'\|\,\|G(\theta)-G(\theta')\| \le& \big(G(\theta)-G(\theta')\big)^{\top}\big[\nabla_{\phi} f\big(G(\theta')\big)-\nabla_{\phi} f'\big(G(\theta')\big)\big]\\ &+\lambda\big(G(\theta)-G(\theta')\big)^{\top}(\theta'-\theta).\end{aligned} \tag{59}$$

Since $G(\theta')$ minimizes f', we have $\big(G(\theta)-G(\theta')\big)^{\top}\nabla_{\phi} f'\big(G(\theta')\big)\ge 0$. Thus

$$\begin{aligned}-\epsilon\beta\|\theta-\theta'\|\,\|G(\theta)-G(\theta')\| \le& \big(G(\theta)-G(\theta')\big)^{\top}\nabla_{\phi} f(G(\theta'))\\ &+\lambda\|G(\theta)-G(\theta')\|\,\|\theta-\theta'\|\end{aligned} \tag{60}$$

by the Cauchy–Schwarz inequality. Combining 59 and 60 gives

$$(-\epsilon\beta-\lambda)\,\|\theta-\theta'\|\,\|G(\theta)-G(\theta')\| \;\le\; \big(G(\theta)-G(\theta')\big)^{\top}\nabla_{\phi} f(G(\theta')), \tag{61}$$

and substituting the upper bound 57 for the right-hand side yields

$$(-\epsilon\beta-\lambda)\,\|\theta-\theta'\|\,\|G(\theta)-G(\theta')\| \;\le\; -(\gamma+\lambda)\,\|G(\theta)-G(\theta')\|^2. \tag{62}$$

Since $\|G(\theta)-G(\theta')\|\ge 0$, dividing both sides of 62 by $(\gamma+\lambda)\|G(\theta)-G(\theta')\|$ gives

$$\|G(\theta)-G(\theta')\| \;\le\; \frac{\epsilon\beta+\lambda}{\gamma+\lambda}\,\|\theta-\theta'\|. \tag{63}$$

Stability of the Proximal RRM Solution in Standard RRM

We examine whether the fixed point of the Proximal RRM introduced in Theorem 9 coincides with the performatively stable point of RRM.

Assume that $\frac{\epsilon\beta}{\gamma} < 1$, Under this condition, a performatively stable point of RRM exists and it's unique. Because the proximal term vanishes at $\phi = \theta_{PS}$, optimality of θ_{PS} yields

$$\mathbb{E}_{z\sim\mathcal{D}(\theta_{PS})}\,\ell(z;\theta_{PS}) \;\leq\; \mathbb{E}_{z\sim\mathcal{D}(\theta_{PS})}\,\ell(z;\phi) \;\leq\; \mathbb{E}_{z\sim\mathcal{D}(\theta_{PS})}\,\ell(z;\phi) + \frac{\lambda}{2}\|\theta_{PS} - \phi\|^2 \quad \forall\phi.$$

Hence

$$\theta_{PS} \;=\; \arg\min_{\phi}\; \mathbb{E}_{z\sim\mathcal{D}(\theta_{PS})}\,\ell(z;\phi) + \frac{\lambda}{2}\|\theta_{PS} - \phi\|^2, \tag{64}$$

so θ_{PS} is also a fixed point of the Proximal RRM.

Note that, for any $\lambda > 0$,

$$\frac{\epsilon\beta}{\gamma} \;\leq\; 1 \;\Longrightarrow\; \frac{\epsilon\beta + \lambda}{\gamma + \lambda} \;\leq\; 1. \tag{65}$$

By Theorem 9, inequality 65 guarantees that the Proximal RRM admits a *unique* fixed point, denoted by θ_{PS}^{λ}. Since θ_{PS} satisfies 64 and the minimiser of 64 is unique, we conclude

$$\boxed{\theta_{PS}^{\lambda} \;=\; \theta_{PS}.}$$

Thus, under $\frac{\epsilon\beta}{\gamma} < 1$, the performatively stable solution of RRM is identical to the unique fixed point of the Proximal RRM.