

A PROOF OF PROPOSITIONS

A.1 PROOF OF PROPOSITION 1

Definition 1 A function $f : D \subset \mathbb{R}^p \mapsto \mathbb{R}$ is called L -smooth if its gradient \dot{f} satisfies the Lipschitz continuous condition with the Lipschitz constant L .

Lemma 1 (Boyd et al., 2006) If function $f : D \subset \mathbb{R}^p \mapsto \mathbb{R}$ is L -smooth, then for $x, y \in D$,

$$|f(x) - f(y) - \langle \dot{f}(y), x - y \rangle| \leq \frac{L}{2} \|x - y\|^2.$$

Here, $\langle x, y \rangle$ denotes the inner product of two arbitrary vectors x, y with the same dimension.

Lemma 2 Assume that the global loss function $\mathcal{L}_N(\theta) : D \subset \mathbb{R}^p \mapsto \mathbb{R}$ is L -smooth, convex and has a unique minimum at $\hat{\theta}$. Then, for any $\hat{\theta}^{(t)} \in D, t = 0, 1, \dots, T$, we have

$$\mathcal{L}_N(\hat{\theta}^{(t)}) - \mathcal{L}_N(\hat{\theta}) \leq \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta}^{(t)} - \hat{\theta} \rangle - \frac{1}{2L} \|\dot{\mathcal{L}}_N(\hat{\theta}^{(t)})\|^2.$$

Proof: Consider a point $z = \hat{\theta} + \frac{1}{L} \dot{\mathcal{L}}_N(\hat{\theta}^{(t)})$. On the one hand, since $\mathcal{L}_N(\theta)$ is convex,

$$\mathcal{L}_N(\hat{\theta}^{(t)}) - \mathcal{L}_N(z) \leq \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta}^{(t)} - z \rangle = \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)})\hat{\theta}^{(t)} - \hat{\theta} \rangle + \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta} - z \rangle.$$

However, since $\mathcal{L}_N(\theta)$ is L -smooth and according to Lemma 1, we have

$$\mathcal{L}_N(z) - \mathcal{L}_N(\hat{\theta}) \leq \langle \dot{\mathcal{L}}_N(\hat{\theta}), z - \hat{\theta} \rangle + \frac{L}{2} \|z - \hat{\theta}\|^2 = \frac{L}{2} \|z - \hat{\theta}\|^2.$$

Then,

$$\begin{aligned} \mathcal{L}_N(\hat{\theta}^{(t)}) - \mathcal{L}_N(\hat{\theta}) &= \mathcal{L}_N(\hat{\theta}^{(t)}) - \mathcal{L}_N(z) + \mathcal{L}_N(z) - \mathcal{L}_N(\hat{\theta}) \\ &\leq \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta}^{(t)} - \hat{\theta} \rangle + \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta} - z \rangle + \frac{L}{2} \|z - \hat{\theta}\|^2 \\ &= \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta}^{(t)} - \hat{\theta} \rangle - \frac{1}{2L} \|\dot{\mathcal{L}}_N(\hat{\theta}^{(t)})\|^2. \end{aligned}$$

Q.E.D

Assume that $\hat{\theta}^{(t+1)}$ and $\hat{\theta}^{(t)}$ are solutions after the $(t+1)$ -th and t -th iterations, respectively. Then,

$$\begin{aligned} \|\hat{\theta}^{(t+1)} - \hat{\theta}\|^2 &= \|\hat{\theta}^{(t)} - \alpha \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}) - \hat{\theta}\|^2 \\ &= \|\hat{\theta}^{(t)} - \hat{\theta}\|^2 - 2\alpha \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta}^{(t)} - \hat{\theta} \rangle + \alpha^2 \|\dot{\mathcal{L}}_N(\hat{\theta}^{(t)})\|^2. \end{aligned} \quad (1)$$

According to Lemma 2 and note that $\mathcal{L}_N(\hat{\theta}^{(t)}) \geq \mathcal{L}_N(\hat{\theta})$, we have

$$0 \leq \mathcal{L}_N(\hat{\theta}^{(t)}) - \mathcal{L}_N(\hat{\theta}) \leq \langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta}^{(t)} - \hat{\theta} \rangle - \frac{1}{2L} \|\dot{\mathcal{L}}_N(\hat{\theta}^{(t)})\|^2.$$

which implies

$$-\langle \dot{\mathcal{L}}_N(\hat{\theta}^{(t)}), \hat{\theta}^{(t)} - \hat{\theta} \rangle \leq -\frac{1}{2L} \|\dot{\mathcal{L}}_N(\hat{\theta}^{(t)})\|^2.$$

Substituting this inequality into equation (1), we have

$$\|\hat{\theta}^{(t+1)} - \hat{\theta}\|^2 \leq \|\hat{\theta}^{(t)} - \hat{\theta}\|^2 - \alpha \left(\frac{1}{L} - \alpha \right) \|\dot{\mathcal{L}}_N(\hat{\theta}^{(t)})\|^2. \quad (2)$$

Equation (2) gives the convergence condition of the gradient decent algorithm. As long as $\alpha \leq 1/L$, we have $\|\hat{\theta}^{(t+1)} - \hat{\theta}\|^2 \leq \|\hat{\theta}^{(t)} - \hat{\theta}\|^2$. Note that we only consider the first -order expansion of $\mathcal{L}_N(\theta)$ at point $\hat{\theta}$ in the proof of Lemma 2. Then, the Lipschitz constant L can be substituted by the largest eigenvalue of $\ddot{\mathcal{L}}_N(\hat{\theta})$, that is, λ_1 . As a result, the convergence condition can be written as $\alpha \leq 1/\lambda_1$.

A.2 PROOF OF PROPOSITION 2

On the one hand, we can verify that

$$\mathrm{tr}(\mathrm{cov}(Q_1)) = \mathrm{tr}\left(\frac{B\Sigma_z B^\top}{N}\right) \geq \tau_{\min} \mathrm{tr}\left(\frac{BB^\top}{N}\right) = p\tau_{\min} \mathrm{tr}\left(\frac{BB^\top}{pN}\right) \geq \frac{\tau_{\min}^2 p}{N}.$$

On the other hand, we have

$$\mathrm{tr}(\mathrm{Cov}(Q_2)) = N^{-1} \mathrm{tr}(\Sigma_\epsilon) \leq \frac{p}{N} \tau_{\max}.$$

As a result, we have $\mathrm{tr}\{\mathrm{cov}(Q_1)\} / \mathrm{tr}\{\mathrm{cov}(Q_2)\} \geq \tau_{\min}^2 / \tau_{\max}$.

REFERENCES

Boyd, Vandenberghe, and Foybusovich. Convex optimization. *IEEE Transactions on Automatic Control*, 51(11):1859–1859, 2006.