Local Error Bound
2022年3月1日 9:11

Recall our problem:

$$\hat{\theta} \in \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \{ F(\theta) \triangleq \mathcal{L}(\theta) + \underbrace{R(\theta)}_{\text{convex}} \} \quad \text{— } (*)$$

and the residual error

$$E(\theta) = \operatorname{prox}_R(\theta - \nabla \mathcal{L}(\theta)) - \theta$$

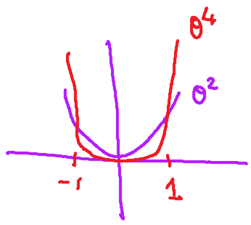
Let $\hat{v}$ be the optimal value and Θ be the optimal solution set of $(*)$, respectively. Note that $E(\theta) = 0 \iff \theta \in \Theta$

Local Error Bound Condition

For any $v \geq \hat{v}$, there exist $\mu, \varepsilon > 0$ s.t.

$$\operatorname{dist}(\theta, \Theta) \leq \mu \cdot \| E(\theta) \|_2 \quad \text{— (EB)}$$

for any $\theta \in \mathbb{R}^d$ satisfying $F(\theta) \leq v$, $\| E(\theta) \|_2 \leq \varepsilon$.



Q: What instances of $(*)$ satisfy (EB)?

Example: (Failure of (EB))

Consider $\min_{\theta \in \mathbb{R}} F(\theta) = \underbrace{\frac{1}{4}\theta^4}_{\mathcal{L}(\theta)}$. (In particular, $R(\theta) = 0$)

Then, $\Theta = \{0\}$, $E(\theta) = \underbrace{\operatorname{prox}_0(\theta - \nabla \mathcal{L}(\theta))}_{\theta - \nabla \mathcal{L}(\theta)} - \theta = \underbrace{-\theta^3}_{-\nabla \mathcal{L}(\theta)}$

Consider $\theta_k \downarrow 0$. Then,

$$\operatorname{dist}(\theta_k, \Theta) = \theta_k, \quad |E(\theta_k)| = \theta_k^3$$

hence $\quad |E(\theta_k)| = o(\operatorname{dist}(\theta_k, \Theta))$

(recall: $f(\theta_k) = o(g(\theta_k))$ if $\lim_{k \to \infty} \frac{f(\theta_k)}{g(\theta_k)} = 0$)

Scenario 1: $\mathcal{L}$ is strongly convex, $\nabla \mathcal{L}$ Lipschitz continuous, R convex

By definition, there exist $\kappa > 0$ and $L > 0$ s.t.

$$\mathcal{L}(\gamma) \geq \mathcal{L}(\theta) + \nabla\mathcal{L}(\theta)^T(\gamma-\theta) + \frac{\kappa}{2}\|\gamma-\theta\|_2^2, \quad \forall \theta,\gamma,$$

and

$$\|\nabla\mathcal{L}(\gamma)-\nabla\mathcal{L}(\theta)\|_2 \le L\|\gamma-\theta\|_2, \quad \forall\theta,\gamma.$$

(Exercise) $$(\nabla\mathcal{L}(\gamma)-\nabla\mathcal{L}(\theta))^T(\gamma-\theta) \ge \kappa\|\gamma-\theta\|_2^2, \quad \forall\theta,\gamma.$$

Let $\hat\theta$ be the unique optimal solution to $(*)$. Then, for any θ,

due to strong convexity of F

$$\kappa\cdot \mathrm{dist}^2(\theta,\Theta) = \kappa\|\theta-\hat\theta\|_2^2 \underset{\substack{\uparrow\\ \text{result}\\ \text{of exercise}}}{\le} (\nabla\mathcal{L}(\theta)-\nabla\mathcal{L}(\hat\theta))^T(\theta-\hat\theta) \quad -(\Delta)$$

$$E(\theta) = \mathrm{prox}_R(\theta-\nabla\mathcal{L}(\theta)) - \theta$$

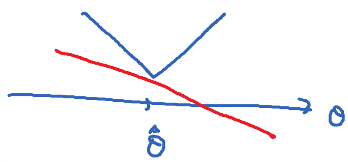
We need to relate $(\nabla\mathcal{L}(\theta)-\nabla\mathcal{L}(\hat\theta))^T(\theta-\hat\theta)$ to $\|E(\theta)\|_2^2$.

To proceed, by the first-order optimality conditions,

$$-\nabla\mathcal{L}(\hat\theta)\in\partial R(\hat\theta) \qquad \text{since } \hat\theta = \underset{\theta}{\arg\min}\{\mathcal{L}(\theta)+R(\theta)\},$$

$$\underbrace{-\nabla\mathcal{L}(\theta) - \mathrm{prox}_R(\theta-\nabla\mathcal{L}(\theta)) + \theta \in \partial R(\mathrm{prox}_R(\theta-\nabla\mathcal{L}(\theta)))}_{-(\nabla\mathcal{L}(\theta)+E(\theta))\,\in\,\partial R(\theta+E(\theta))} \quad \text{since } \mathrm{prox}_R(\theta-\nabla\mathcal{L}(\theta)) = \underset{\gamma}{\arg\min}\left\{\frac{1}{2}\|\theta-\nabla\mathcal{L}(\theta)-\gamma\|_2^2 + R(\gamma)\right\}$$

By definition of the subdifferential,



$$R(\theta+E(\theta)) \ge R(\hat\theta) + (\overset{\partial R(\hat\theta)}{-\nabla\mathcal{L}(\hat\theta)})^T(\theta+E(\theta)-\hat\theta),$$

$$R(\hat\theta) \ge R(\theta+E(\theta)) + (\underbrace{-(\nabla\mathcal{L}(\theta)+E(\theta))}_{\in\,\partial R(\theta+E(\theta))})^T(\hat\theta-\theta-E(\theta))$$

Add them up:

$$0 \ge (\nabla\mathcal{L}(\theta)+E(\theta)-\nabla\mathcal{L}(\hat\theta))^T(\theta+E(\theta)-\hat\theta).$$

$$\Rightarrow \quad (\nabla\mathcal{L}(\theta) - \nabla\mathcal{L}(\hat{\theta}))^T(\theta - \hat{\theta}) + \|E(\theta)\|_2^2$$

$$\leq \quad E(\theta)^T\left(\hat{\theta} - \theta + \nabla\mathcal{L}(\hat{\theta}) - \nabla\mathcal{L}(\theta)\right)$$

$$\underset{\substack{\text{triangle}\\ \text{ineq.}}}{\overset{\text{C-S}}{\leq}} \quad \|E(\theta)\|_2 \cdot \Big[\underbrace{\|\hat{\theta} - \theta\|_2}_{= \operatorname{dist}(\theta, \Theta)} + \underbrace{\|\nabla\mathcal{L}(\hat{\theta}) - \nabla\mathcal{L}(\theta)\|_2}_{\leq L \cdot \|\hat{\theta} - \theta\|_2 = L \cdot \operatorname{dist}(\theta, \Theta)} \Big]$$

$$= (L+1) \cdot \|E(\theta)\|_2 \operatorname{dist}(\theta, \Theta). \qquad \text{———} \ (\Delta\Delta)$$

Put (Δ) and $(\Delta\Delta)$ together,

$$\kappa \operatorname{dist}^2(\theta, \Theta) \leq (L+1) \cdot \|E(\theta)\|_2 \cdot \operatorname{dist}(\theta, \Theta)$$

$$\Rightarrow \quad \operatorname{dist}(\theta, \Theta) \leq \underbrace{\frac{L+1}{\kappa}}_{\mu} \|E(\theta)\|_2.$$

Note that this bound is global; i.e., it holds for any θ.

Note that not all loss functions of interest fall into Scenario 1. For instance, consider

$$\mathcal{L}(\theta) = \frac{1}{2}\|A\theta - b\|_2^2$$

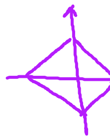
Here, $A \in \mathbb{R}^{n \times d}$ and consider $d \gg n$. Note that for any $\bar{\theta} \in \operatorname{null}(A) = \{\theta : A\theta = 0\} \neq \{0\}$, $\mathcal{L}(\bar{\theta})$ is constant. This suggests that $\mathcal{L}$ is not strongly convex.

Scenario 2: $\mathcal{L}$ takes the form $\mathcal{L}(\theta) = h(A\theta)$ for some $A \in \mathbb{R}^{n \times d}$, h is strongly convex on compact sets, continuously differentiable with ∇h Lipschitz continuous; R has polyhedral epigraph; Θ is compact.

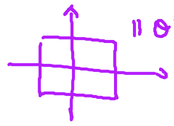
Remarks

① Recall that $\mathrm{epi}(R) = \{ (\theta, t) \in \mathbb{R}^d \times \mathbb{R} : R(\theta) \le t \}$. We say that R has polyhedral epigraph if $\mathrm{epi}(R)$ is a polyhedron. Two examples include

$R(\theta) = \|\theta\|_1$,



$R(\theta) = \|\theta\|_\infty$



② If $\mathcal{L}$ is strongly convex and has Lipschitz $\nabla \mathcal{L}$, then it satisfies the requirements in Scenario 2, since we can take $\mathcal{L} = h$ and $A = I$

③ Some applications that fit into Scenario 2:

- ℓ_1-regularized least-squares regression

$$\mathcal{L}(\theta) = \frac{1}{2} \| A\theta - b \|_2^2 , \quad h(y) = \frac{1}{2} \| y - b \|_2^2$$

$$R(\theta) = \|\theta\|_1$$

- ℓ_1-regularized logistic regression

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \ln(1 + \exp(-b_i a_i^T \theta)),$$

$$h(y) = \sum_{i=1}^{n} \ln(1 + \exp(-b_i y_i)),$$

$$R(\theta) = \|\theta\|_1$$

Exercise: Show that h is strongly convex on compact sets