Local Error Bound (cont'd)

2022年3月3日 15:32

Recall the Setting of Scenario 2:

$$\hat{\theta} \in \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \Big\{ F(\theta) \triangleq \underbrace{h(A\theta)}_{\substack{\text{Strongly convex} \\ \text{on compact sets,} \\ \nabla h \text{ Lipschitz}}} + \underbrace{R(\theta)}_{\substack{\text{norm with} \\ \text{polyhedral} \\ \text{epigraph}}} \Big\} \qquad — (*)$$

$A \in \mathbb{R}^{n \times d}$

As before, $\hat{v}$ is the optimal value and Θ is the optimal solution set.

Local Error Bound Condition

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

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

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

Proposition 1: There exists a $\bar{y} \in \mathbb{R}^n$ such that for all $\hat{\theta} \in \Theta$,

$$A\hat{\theta} = \bar{y}, \qquad \nabla \mathcal{L}(\hat{\theta}) = A^T \nabla h(A\hat{\theta}) = A^T \nabla h(\bar{y}) \triangleq \bar{g}$$

In particular,

$$\Theta = \{ \theta \in \mathbb{R}^d : A\theta = \bar{y}, \ -\bar{g} \in \partial R(\theta) \}$$

Proof: Let $\theta_1, \theta_2 \in \Theta$ and set $\bar{y}_1 = A\theta_1$, $\bar{y}_2 = A\theta_2$.

Since h is strongly convex on compact sets, if $\bar{y}_1 \neq \bar{y}_2$,

then

$$h\Big(\frac{\overset{=A\theta_1}{\bar{y}_1} + \overset{=A\theta_2}{\bar{y}_2}}{2}\Big) < \frac{1}{2} h(\bar{y}_1) + \frac{1}{2} h(\bar{y}_2).$$

$$\Leftrightarrow \quad \mathcal{L}\Big(\frac{\theta_1 + \theta_2}{2}\Big) < \frac{1}{2}\mathcal{L}(\theta_1) + \frac{1}{2}\mathcal{L}(\theta_2) \qquad — (1)$$

By convexity of R,

$$R\Big(\frac{\theta_1 + \theta_2}{2}\Big) \leq \frac{1}{2} R(\theta_1) + \frac{1}{2} R(\theta_2) \qquad — (2)$$

Adding (1) and (2), $F\Big(\frac{\theta_1 + \theta_2}{2}\Big) < \frac{1}{2} \underset{\hat{v}}{\underset{\shortparallel}{F(\theta_1)}} + \frac{1}{2} \underset{\hat{v}}{\underset{\shortparallel}{F(\theta_2)}} = \hat{v}$, which is a contradiction.

Consider the structure of Θ. We can write

$$\Theta = \Theta_L \cap \Theta_R ,$$

where $\Theta_L = \{ \theta \in \mathbb{R}^d : A\theta = \bar{y} \}$, $\Theta_R = \{ \theta \in \mathbb{R}^d : -\bar{g} \in \partial R(\theta) \}$.

Observe: Θ_L is polyhedral

It turns out that Θ_R is also polyhedral. This relies on the following fact:

Fact (Rockafellar: Convex Analysis)

(1) (Theorem 19.2) If R has polyhedral epigraph, so does its conjugate $\tilde{R}$ (recall: $\tilde{R}(\gamma) = \sup_\theta \{ \theta^T\gamma - R(\theta) \}$)

(2) (Corollary 23.5.1) $\partial \tilde{R} = (\partial R)^{-1}$; i.e.,

$$\partial \tilde{R}(\gamma) = (\partial R)^{-1}(\gamma) = \{ \theta \in \mathbb{R}^d : \gamma \in \partial R(\theta) \}$$

Hence, $\Theta_R = (\partial R)^{-1}(-\bar{g}) = \partial \tilde{R}(-\bar{g})$.

(3) (Theorem 23.10) If R has polyhedral epigraph and $R(\theta)$ is finite, then $\partial R(\theta)$ is a polyhedron.

By (2), $\Theta_R = \partial \tilde{R}(-\bar{g})$. By (1), since R has polyhedral epigraph, so does $\tilde{R}$. By (3) (taking $R = \tilde{R}$, $\theta = -\bar{g}$), we know that $\partial \tilde{R}(-\bar{g})$ is polyhedral if $\tilde{R}(-\bar{g})$ is finite.

To prove the finiteness of $\tilde{R}(-\bar{g})$, we use the assumption that R is a norm.

Fact: Given a norm R, its conjugate $\tilde{R}$ is

$$\tilde{R}(\gamma) = \begin{cases} 0 & \text{if } R^*(\gamma) \le 1, \end{cases}$$

$$\tilde{R}(\gamma) = \begin{cases} 0 & \text{if } \underbrace{R^*(\gamma)}_{\text{dual norm}} \le 1, \\ +\infty & \text{o/w}; \end{cases}$$

i.e., $\tilde{R}$ is the indicator of the unit ball of R^*.

Since $\Theta \neq \phi$, there exists an $\hat{\theta} \in \Theta_R = \partial \tilde{R}(-\bar{g})$. Hence,

$$\tilde{R}(g) \geq \tilde{R}(-\bar{g}) + \hat{\theta}^T(g+\bar{g}) \quad \forall g \in \mathbb{R}^d$$

Taking $g=0$, we get $0 \geq \tilde{R}(-\bar{g}) + \hat{\theta}^T\bar{g}$, which implies

$\tilde{R}(-\bar{g}) \leq -\hat{\theta}^T\bar{g} < +\infty$. $\tilde{R}(0)=0$ since $R^*(0)=0 \leq 1$.

Summary: $\Theta = \Theta_L \cap \Theta_R$,

$$\underbrace{\Theta_L = \{\theta \in \mathbb{R}^d : A\theta = \bar{y}\}}_{\text{polyhedral}}, \quad \underbrace{\Theta_R = \{\theta \in \mathbb{R}^d : -\bar{g} \in \partial R(\theta)\}}_{\text{polyhedral}}$$

Since we are interested in bounding $\text{dist}(\theta, \Theta)$, the above motivates us to consider point-polyhedron distance.

Theorem 1 (Hoffman error bound)

Let $P = \{z \in \mathbb{R}^n : Az \leq b\}$ be a non-empty polyhedron. Then, there exists a $c>0$, which depends only on A, s.t.

$$\text{dist}(x, P) \leq c \cdot \|(Ax-b)_+\|_2 \quad \forall x \in \mathbb{R}^n$$

where $(v)_+ = \max\{v, 0\}$.

Corollary 1: Let $\{P_1, P_2, \cdots, P_M\}$ be a finite collection of polyhedra. Suppose that $P = \bigcap_{i=1}^{M} P_i \neq \phi$. Then, there exists an $\alpha > 0$ s.t.

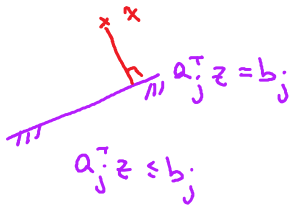
$$\text{dist}^2(x, P) \leq \alpha \cdot \sum_{i=1}^{M} \text{dist}^2(x, P_i) \quad \forall x \in \mathbb{R}^n.$$

(Thus, $\{P_1, \cdots, P_M\}$ is linearly regular.)

<u>Proof of Corollary 1:</u> Let $H_j = \{ z \in \mathbb{R}^n : a_j^T z \le b_j \}$ for $j = 1, \cdots, L$

and $\{K_1, \cdots, K_M\}$ be a partition of $\{1, \cdots, L\}$ s.t.

$P_i = \bigcap_{j \in K_i} H_j$. Observe

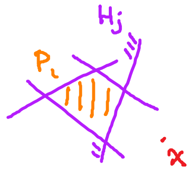


(Exercise) (Δ) $$\text{dist}(x, H_j) = \frac{(a_j^T x - b_j)_+}{\|a_j\|_2}$$

Hence,

$$\text{dist}^2(x, P) \le c' \cdot \sum_{j=1}^{L} \text{dist}^2(x, H_j)$$ (by Theorem 1) and (Δ)

$$= c' \sum_{i=1}^{M} \sum_{j \in K_i} \text{dist}^2(x, H_j)$$



$$\le c'' \sum_{i=1}^{M} \text{dist}^2(x, P_i)$$ ($\because \text{dist}(x, H_j) \le \text{dist}(x, P_i)$ for $j \in K_i$)

By the corollary,

$$\text{dist}(\theta, \Theta) \le c \cdot [\text{dist}(\theta, \Theta_L) + \text{dist}(\theta, \Theta_R)]$$

for some constant $c > 0$. By Theorem 1,

$$\text{dist}(\theta, \Theta_L) \le c' \cdot \|A\theta - \tilde{y}\|_2 .$$

<u>Next tasks:</u>

1. The RHS of the error bound is in terms of $\|E(\theta)\|_2$. Hence, we need to relate $\|A\theta - \tilde{y}\|_2$ to $\|E(\theta)\|_2$.

2. We cannot apply the same trick to bound $\text{dist}(\theta, \Theta_R)$, since we do not have an explicit description of Θ_R.

Thus, some new tools are needed.