ISEN 629: Engineering Optimization

Lecture 15

Sergiy Butenko

Industrial and Systems Engineering
Texas A&M University

Fall 2007

Lower complexity bounds

Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x).$$

Model:	1. Unconstrained minimization 2. f is convex on $\mathbb{R}^n$ and Lipschitz continuous on a bounded set.
Oracle:	First-order black box: at each point $\hat{x}$ we can compute $f(\hat{x}),\ g(\hat{x}) \in \partial f(\hat{x})$.
Solution:	Find $\bar{x} \in \mathbb{R}^n : f(\bar{x}) - f^* \leq \epsilon$
Methods:	Generate a sequence $\{x_k\}$: $x_k \in x_0 + \mathrm{Lin}\{g(x_0), \dots, g(x_{k-1})\}$.

Lower complexity bounds

In addition, we assume that f belongs to the class $\mathcal{P}(x_0, R, M)$ defined as follows:

- There exists a solution x^* of our problem and $x^* \in B_R(x_0)$;
- f is Lipschitz continuous on $B_R(x_0)$ with constant $M > 0$.

Theorem (3.2.1)

For any class $\mathcal{P}(x_0, R, M)$ and any $k, 0 \leq k \leq n-1$, there exists a function $f \in \mathcal{P}(x_0, R, M)$ such that

$$f(x_k) - f^* \geq \frac{MR}{2(1+\sqrt{k+1})}$$

for any scheme generating a sequence $x_k \in x_0 + \mathrm{Lin}\{g(x_0), \dots, g(x_{k-1})\}$.

Nonsmooth optimization methods

Consider the problem

$$\min\{f(x) : x \in Q\},$$

where $Q \subseteq \mathbb{R}^n$ is a closed convex set and $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

Can we use a subgradient $g(x)$ in place of the gradient?

- The objective function may not decrease in the direction $-g(x)$.
- We cannot expect that $g(x) \to 0,\ x \to x^*$.

However, we can use the inequality

$$g(x)^T(x - x^*) \geq 0$$

that holds for any $x \in Q$.

Nonsmooth optimization methods

For an interior point x of Q, consider $y = x - \alpha g(x) \in Q$, where α is a sufficiently small positive scalar. Since $g(x)^T(x - x^*) \geq 0$ for any $x \in Q$, we have

$$\begin{aligned} \|y - x^*\|^2 &= (x - \alpha g(x) - x^*)^T(x - \alpha g(x) - x^*) \\ &= \|x - x^*\|^2 - 2\alpha g(x)^T(x - \tfrac{1}{2}\alpha g(x) - x^*) \\ &\leq \|x - x^*\|^2. \end{aligned}$$

Thus,

> *The distance between x and x^* is decreasing in the direction $-g(x)$.*

Nonsmooth optimization methods

If we fix $x = \bar{x} \in Q$ and consider the hyperplane

$$g(\bar{x})^T(\bar{x} - y) = 0,$$

then for any optimal solution $y = x^*$ we have

$$g(\bar{x})^T(\bar{x} - x^*) \geq 0.$$

Thus,

> *Only one of the two half-spaces defined by the above hyperplane can contain an optimal solution x^* as its interior point.*

Nonsmooth optimization methods

Next, we develop a technique for estimating the quality of an approximate solution to our optimization problem. We define

$$v_f(\bar{x}, x) = \begin{cases} \frac{1}{\|g(x)\|} g(x)^T(x - \bar{x}), & g(x) \neq 0; \\ 0, & g(x) = 0. \end{cases}$$

Then, clearly, $v_f(\bar{x}, x) \leq \|x - \bar{x}\|$.

Geometrically, $v_f(\bar{x}, x)$ is a distance from $\bar{x}$ to hyperplane $\{y : g(x)^T(x - y) = 0\}$. Indeed, for $y = \bar{x} + v_f(\bar{x}, x)g(x)/\|g(x)\|$ we have

$$\|y - \bar{x}\| = v_f(\bar{x}, x)$$

and

$$g(x)^T(x - y) = g(x)^T(x - \bar{x}) - v_f(\bar{x}, x)\|g(x)\| = 0.$$

Nonsmooth optimization methods

The following function measures variation of f with respect to $\bar{x}$:

$$w_f(\bar{x}; t) = \begin{cases} \max\{f(x) - f(\bar{x}) : \|x - \bar{x}\| \leq t\}, & \text{if } t \geq 0; \\ 0, & \text{if } t < 0 \end{cases}$$

This function has the following properties:

- $w_f(\bar{x}; t)$ is a nondecreasing function of t, $t \in \mathbb{R}^1$.
- $f(x) - f(\bar{x}) \leq w_f(\bar{x}; \|x - \bar{x}\|)$.

Nonsmooth optimization methods

Lemma (3.2.1)

For any $x \in \mathbb{R}^n$ we have

$$f(x) - f(\bar{x}) \leq w_f(\bar{x}; v_f(\bar{x}; x)). \qquad (1)$$

If $f(x)$ is Lipschitz continuous on $B_R(\bar{x})$ with constant M then

$$f(x) - f(\bar{x}) \leq M(v_f(\bar{x}; x))_+ \qquad (2)$$

for all $x \in \mathbb{R}^n$ with $v_f(\bar{x}; x) \leq R$.

Nonsmooth optimization methods

Consider the problem $\quad \min\{f(x) : x \in Q\} \qquad (*)$

Definition

Let $\{x_i : i \geq 0\}$ be a sequence in Q. Define

$$S_k = \{x \in Q : g(x_i)^T(x_i - x) \geq 0, i = 0, \ldots, k\}.$$

We call S_k a localization set of our problem generated by the sequence $\{x_i : i \geq 0\}$.

Let x^* be a solution to the problem $(*)$. Then for all $k \geq 0$ we have $x^* \in S_k$. Denote by

$$v_i = v_f(x^*, x_i), \quad v_k^* = \min_{0 \leq i \leq k} v_i.$$

Thus,

$$v_k^* = \max\{r : g(x_i)^T(x_i - x) \geq 0, i = 0, \ldots, k, \forall x \in B_r(x^*)\}.$$

Nonsmooth optimization methods

Lemma (3.2.2)

Let $f_k^* = \min_{0 \leq i \leq k} f(x_i)$. Then $f_k^* - f^* \leq w_f(x^*, v_k)$.

Proof.

Using the previous lemma,

$$w_f(x^*, v_k^*) = \min_{0 \leq i \leq k} w_f(x^*, v_i) \geq \min_{0 \leq i \leq k} [f(x_i) - f^*] = f_k^* - f^*.$$

□

Subgradient method

We consider the problem $\quad \min\{f(x) : x \in Q\}, \qquad (*)$
where f is convex on $\mathbb{R}^n$ and Q is a simple closed convex set.

- By "simple" we mean that we can easily solve some minimization problems over Q.
- In particular, we should be able to easily find a projection of any point onto Q.

We assume the first-order oracle, which for any x provides $f(x)$ and one of the subgradients $g(x)$.

Subgradient method

0. Choose $x_0 \in Q$ and a sequence $\{h_k : k \geq 0\}$:

$$h_k > 0,\ h_k \to 0,\ \sum_{k=0}^{\infty} h_k = \infty.$$

1. k-th iteration ($k \geq 0$):
   Compute $f(x_k),\ g(x_k)$ and set

$$x_{k+1} = \pi_Q\left(x_k - h_k \frac{g(x_k)}{\|g(x_k)\|}\right).$$

Subgradient method

Theorem (3.2.2)

Let f be Lipschitz continuous on $B_R(x^*)$ with constant M and $x_0 \in B_R(x^*)$. Then

$$f_k^* - f^* \leq M \frac{R^2 + \sum_{i=0}^{k} h_i^2}{2\sum_{i=0}^{k} h_i}.$$

Subgradient method

Denote by

$$\Delta_k = \frac{R^2 + \sum_{i=0}^{k} h_i^2}{2\sum_{i=0}^{k} h_i}.$$

Then $\Delta_k \to 0$ if $\sum_{i=0}^{\infty} h_i = \infty$.

To choose h_k in a way that would be optimal for a fixed number of iterations,

- assume that we perform N steps of the subgradient method.
- Then Δ_k is a convex function of $\{h_k\}_{k=0}^{N}$.
- Minimizing Δ_k as a function of $\{h_k\}_{k=0}^{N}$, we find that the optimal strategy is given by

$$h_i = \frac{R}{\sqrt{N+1}},\ i = 0, \ldots, N.$$

- Then $\Delta_N = \frac{R}{\sqrt{N+1}}$ and $f_k^* - f^* \leq \frac{MR}{\sqrt{N+1}}$.

Subgradient method

Note that this bound is proportional to the lower bound for first-order methods:

$$f(x_k) - f^* \geq \frac{MR}{2(1+\sqrt{k+1})};$$

$$f_k^* - f^* \leq \frac{MR}{\sqrt{N+1}}.$$

Problems with functional constraints

Consider the problem

$$\min\{f(x) : x \in Q, f_j(x) \leq 0, j = 1, \dots, m\}, \qquad (**)$$

where f, f_j are convex and Q is a simple bounded closed convex set,

$$\|x - y\| \leq R \quad \forall x, y \in Q.$$

Denoting by

$$\bar{f}(x) = \left(\max_{1\leq j\leq m} f_j(x)\right)_+,$$

we can write our problem as

$$\min\{f(x) : x \in Q, \bar{f}(x) \leq 0\}. \qquad (***)$$

Problems with functional constraints

Let x^* be a solution to the problem $(**)$. We have

- $\bar{f}(x^*) = 0$ and
- $v_{\bar{f}}(x^*; x) \geq 0$ for all $x \in \mathbb{R}^n$.
- Thus, $\bar{f}(x) \leq w_{\bar{f}}(x^*, v_{\bar{f}}(x^*, x))$.
- If f_j are Lipschitz continuous on Q with constant M, then for any $x \in \mathbb{R}^n$ we have

$$\bar{f}(x) \leq M v_{\bar{f}}(x^*, x).$$

Subgradient method

0. Choose $x_0 \in Q$ and sequence $\{h_k : k \geq 0\}$:

$$h_k = \frac{R}{\sqrt{k + 0.5}}.$$

1. k-th iteration ($k \geq 0$):
Compute $f(x_k)$, $g(x_k)$, $\bar{f}(x_k)$, $\bar{g}(x_k)$ and set

$$p_k = \begin{cases} g(x_k), & \text{if } \bar{f}(x_k) < \|\bar{g}(x_k)\| h_k, \\ \bar{g}(x_k), & \text{if } \bar{f}(x_k) \geq \|\bar{g}(x_k)\| h_k. \end{cases}$$

$$x_{k+1} = \pi_Q\left(x_k - h_k \frac{p_k}{\|p_k\|}\right).$$

Subgradient method

Theorem (3.2.3)
Let f be Lipschitz continuous on $B_R(x^*)$ with constant M_1 and

$$M_2 = \max_{1\leq j\leq m}\{\|g\| : g \in \partial f_j(x), x \in B_R(x^*)\}.$$

Then for any $k \geq 3$ there exists a number i', $0 \leq i' \leq k$, such that

$$f(x_{i'}) - f^* \leq \frac{\sqrt{3}M_1 R}{\sqrt{k - 1.5}}, \quad \bar{f}(x_{i'}) \leq \frac{\sqrt{3}M_2 R}{\sqrt{k - 1.5}}.$$