ISEN 629: Engineering Optimization

Lecture 14

Sergiy Butenko

Industrial and Systems Engineering
Texas A&M University

Fall 2007

Subgradients

Theorem (3.1.14)

Let f be a closed convex function. Then for any $x_0 \in \operatorname{int}(\operatorname{dom} f)$ and $p \in \mathbb{R}^n$ we have

$$f'(x_0; p) = \max\{g^T p : g \in \partial f(x_0)\}.$$

Proof.

The proof consists of the following steps:

1. Prove that $f'(x_0; p) \geq \max\{g^T p : g \in \partial f(x_0)\}$.
   Indeed, for an arbitrary $g \in \partial f(x_0)$ we have
   $$f'(x_0; p) = \lim_{\alpha \to 0+} \frac{1}{\alpha}[f(x_0 + \alpha p) - f(x_0)] \geq g^T p.$$
2. Find a vector $g_p \in \partial f(x_0)$ such that $f'(x_0; p) \leq g_p^T p$.

Subgradients

$$f(y) \geq f(x) + f'(x; y - x) \qquad (*)$$

To prove the second part, we will consider $f'(x_0; p)$ as a function of p and show the following:

(a) $\partial f(x_0) = \partial_p f'(x_0; 0)$.

(b) For any $g_p \in \partial f'(x_0; p)$ we have $g_p \in \partial_p f'(x_0; 0) \equiv \partial f(x_0)$ and $g_p^T p \geq f'(x_0; p)$.

* * *

(a) Since $f'(x_0; 0) = 0$, the inequality $f'(x_0; p) \geq g^T p$ is equivalent to $f'(x_0; p) - f'(x_0; 0) \geq g^T p$ for any $g \in \partial f(x_0)$. Thus, $\partial f(x_0) \subseteq \partial_p f'(x_0; 0)$. On the other hand, $f'(x_0; p)$ is convex in p and for any $y \in \operatorname{dom} f$ we have [(*)]

$$f(y) \geq f(x_0) + f'(x_0; y - x_0) \geq {}^{1} f(x_0) + g^T (y - x_0)$$

for any $g \in \partial_p f'(x_0; 0)$. Thus, $\partial_p f'(x_0; 0) \subseteq \partial f(x_0)$, so $\partial f(x_0) = \partial_p f'(x_0; 0)$.

[1]Here we used the definition of subdifferential for $\partial_p f'(x_0; 0)$: for any $g \in \partial_p f'(x_0; 0)$ we have $f'(x_0; y - x_0) \geq f'(x_0; 0) + g^T (y - x_0)$

Subgradients

$$f(y) \geq f(x) + f'(x; y - x) \qquad (*)$$

(b) Denote by $\phi(p) = f'(x_0; p)$ and let $g_p \in \partial_p f'(x_0; p) \equiv \partial \phi(p)$. For any $v \in \mathbb{R}^n, \tau > 0$, using $(*)$ with $y = \tau v, x = p$, we have

$$\begin{aligned} \tau f'(x_0; v) = \\ f'(x_0; \tau v) &\equiv \phi(\tau v) \geq \phi(p) + \phi'(p; \tau v - p) \geq \phi(p) + g_p^T(\tau v - p) \\ &\equiv f'(x_0; p) + g_p^T(\tau v - p), \end{aligned}$$

i.e.,

$$\tau f'(x_0; v) = f'(x_0; \tau v) \geq f'(x_0; p) + g_p^T(\tau v - p). \qquad (1)$$

Thus,

$$f'(x_0; v) \geq \frac{1}{\tau}[f'(x_0; p) + g_p^T(\tau v - p)] = g^T v + \frac{1}{\tau}[f'(x_0; p) - g_p^T p],$$

and taking $\tau \to \infty$, we obtain $f'(x_0; v) \geq g_p^T v \Rightarrow g_p \in \partial_p f'(x_0; 0) \equiv \partial f(x_0)$. Finally, taking $\tau \to 0$ in (1), we obtain

$$f'(x_0; p) - g_p^T p \leq 0.$$

□

Optimality conditions

Theorem (3.1.15)

For a closed convex function f, we have

$$x^* = \arg\min_{x \in \mathrm{dom}\, f} f(x) \quad \Leftrightarrow \quad 0 \in \partial f(x^*).$$

Proof.

$$f(x) \geq f(x^*) \text{ for all } x \in \mathrm{dom}\, f$$
$$\Updownarrow$$
$$f(x) \geq f(x^*) + 0^T(x - x^*) \text{ for all } x \in \mathrm{dom}\, f$$
$$\Updownarrow$$
$$0 \in \partial f(x_0)$$

□

Subgradients and optimization

Theorem (3.1.16)

For any $x_0 \in \mathrm{dom}\, f$ all vectors $g \in \partial f(x_0)$ are supporting to the level set $\mathcal{L}_f(f(x_0))$:

$$g^T(x_0 - x) \geq 0 \text{ for all } x \in \mathcal{L}_f(f(x_0)) \equiv \{x \in \mathrm{dom}\, f : f(x) \leq f(x_0)\}.$$

Proof.

If $f(x) \leq f(x_0)$ and $g \in \partial f(x_0)$, then

$$f(x_0) \geq f(x) \geq f(x_0) + g^T(x - x_0).$$

□

Corollary

Let $Q \subseteq \mathrm{dom}\, f$ be a closed convex set, $x_0 \in Q$ and $x^* = \arg\min_{x \in Q} f(x)$. Then for any $g \in \partial f(x_0)$ we have $g^T(x_0 - x^*) \geq 0$.

Computing subgradients

Lemma (3.1.7)

Let f be closed, convex, and differentiable on the interior of its domain. Then $\partial f(x) = \{f'(x)\}$ for any $x \in \mathrm{int}(\mathrm{dom}\, f)$.

Proof.

Consider an arbitrary $x \in \mathrm{int}(\mathrm{dom}\, f)$. Then for any direction $p \in \mathbb{R}^n$ and any $g \in \partial f(x)$ we have

$$f'(x)^T p = f'(x; p) = \max\{g^T p : g \in \partial f(x)\} \geq g^T p.$$

Using $-p$ instead of p, we have

$$f'(x)^T p \leq g^T p.$$

Thus, $f'(x)^T p = g^T p$ for all $g \in \partial f(x)$.
Using $p = e_k,\ k = 1, \ldots, n$ we conclude that each component of g is equal to the corresponding component of $f'(x)$, therefore $g = f'(x)$.

□

Computing subgradients

Lemma (3.1.8)

Let $f(y)$ be closed and convex with $\mathrm{dom}\, f \subseteq \mathbb{R}^m$. For a linear operator $\mathcal{A}(x) = Ax + b : \mathbb{R}^n \to \mathbb{R}^m$, consider the function $\phi(x) = f(\mathcal{A}(x))$. Then $\phi(x)$ is a closed convex function with $\mathrm{dom}\, \phi = \{x : \mathcal{A}(x) \in \mathrm{dom}\, f\}$, and for any $x \in \mathrm{int}(\mathrm{dom}\, \phi)$ we have

$$\partial \phi(x) = A^T \partial f(\mathcal{A}(x)).$$

Proof.

For $x \in \mathrm{int}(\mathrm{dom}\, \phi)$, let $y = \mathcal{A}(x)$. Then for all $p \in \mathbb{R}^n$ we have

$$\begin{aligned}
\phi'(x; p) &= \lim_{\alpha \to 0+} \tfrac{\phi(x+\alpha p) - \phi(x)}{\alpha} = \lim_{\alpha \to 0+} \tfrac{f(y + \alpha A p) - f(y)}{\alpha} = f'(y; Ap); \\
\phi'(x; p) &= \max\{g^T p : g \in \partial \phi(x)\}; \\
f'(y; Ap) &= \max\{g^T A p : g \in \partial f(y)\} = \max\{\bar{g}^T p : \bar{g} \in A^T \partial f(y)\}.
\end{aligned}$$

Therefore, the support functions $\Psi_{\partial \phi(x)}(p)$ and $\Psi_{A^T \partial f(y)}(p)$ are equal for any $p \in \mathbb{R}^n$, yielding $\partial \phi(x) \equiv A^T \partial f(\mathcal{A}(x))$. □

Computing subgradients

Lemma (3.1.9)

Let $f_1(x), f_2(x)$ be closed convex functions and $\alpha_1, \alpha_2 \geq 0$. Then $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x)$ is a closed convex function and for any $x \in \mathrm{int}(\mathrm{dom}\, f) = \mathrm{int}(\mathrm{dom}\, f_1) \cap \mathrm{int}(\mathrm{dom}\, f_2)$:

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x).$$

Proof.

For $f(x) = f_1(x) + f_2(x)$, consider a sequence $\{(x_k, t_k)\} \subset \mathrm{epi}(f)$:

$$t_k \geq f(x_k) = f_1(x_k) + f_2(x_k), \ \lim_{k\to\infty} x_k = \bar{x}, \ \lim_{k\to\infty} t_k = \bar{t}; \{x_k\} \subseteq \mathrm{dom}\, f.$$

Note that for a closed function ϕ, for any sequence $\{x_k\} \subseteq \mathrm{dom}(\phi)$ such that $x_k \to \bar{x}$, we have $\liminf_{k\to\infty} \phi(x_k) \geq \phi(\bar{x})$. Indeed, if we assume the opposite, then there exists a subsequence $\{y_k\}$ of $\{x_k\}$ such that $\bar{\tau} = \lim_{k\to\infty} \phi(y_k) < \phi(\bar{x}) \Rightarrow (\bar{x}, \bar{\tau}) \notin \mathrm{epi}(\phi)$. Since $(y_k, \phi(y_k)) \in \mathrm{epi}(\phi)$ and $\lim_{k\to\infty} y_k = \bar{x}$, this contradicts to the assumption that ϕ is closed.

Computing subgradients

Since f_1 and f_2 are closed, we have

$$\liminf_{k\to\infty} f_1(x_k) \geq f_1(\bar{x}), \ \liminf_{k\to\infty} f_2(x_k) \geq f_2(\bar{x}), \text{ and}$$

$$\bar{t} = \lim_{k\to\infty} t_k \geq \liminf_{k\to\infty} f_1(x_k) + \liminf_{k\to\infty} f_2(x_k) \geq f(\bar{x}) \Rightarrow (\bar{x}, \bar{t}) \in \mathrm{epi}(f).$$

To prove the relation for the subdifferentials, consider $x \in \mathrm{int}(\mathrm{dom}\, f_1) \cap \mathrm{int}(\mathrm{dom}\, f_2)$. Then for any $p \in \mathbb{R}^n$:

$$\begin{aligned} f'(x;p) &= \alpha_1 f_1'(x;p) + \alpha_2 f_2'(x;p) \\ &= \max\{g_1^T(\alpha_1 p) : g_1 \in \partial f_1(x)\} + \max\{g_2^T(\alpha_2 p) : g_2 \in \partial f_2(x)\} \\ &= \max\{(\alpha_1 g_1 + \alpha_2 g_2)^T p : g_1 \in \partial f_1(x), g_2 \in \partial f_2(x)\} \\ &= \max\{g^T p : g \in \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)\}. \end{aligned}$$

On the other hand, $f'(x;p) = \max\{g^T p : g \in \partial f(x)\}$. Since $\partial f_1(x)$ and $\partial f_2(x)$ are bounded and for any $p \in \mathbb{R}^n$ $\Psi_{\partial f(x)}(p) = \Psi_{\alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)}(p)$, we conclude that $\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$. □

Computing subgradients

Lemma (3.1.10)

Let functions $f_i(x), i = 1, \ldots, m$, be closed and convex. Then $f(x) = \max_{1\leq i\leq m} f_i(x)$ is also closed and convex, and for any $x \in \mathrm{int}(\mathrm{dom}\, f) = \bigcap_{i=1}^{m} \mathrm{int}(\mathrm{dom}\, f_i)$ we have

$$\partial f(x) = \mathrm{Conv}\{\partial f_i(x) : i \in I(x)\}, \text{ where } I(x) = \{i : f_i(x) = f(x)\}.$$

Proof.

We will use the following fact in the proof: for any set of values $a_1, \ldots, a_k$ we have

$$\max_{1\leq i\leq k} a_i = \max\left\{\sum_{i=1}^{k} \lambda_i a_i : \{\lambda_i\} \subset \Delta_k\right\},$$

where $\Delta_k = \{\lambda_i \geq 0 : \sum_{i=1}^{k} \lambda_i = 1\}$.

Computing subgradients

Consider $x \in \bigcap_{i=1}^{m} \mathrm{int}(\mathrm{dom}\, f_i)$. Assume that $I(x) = \{1, \ldots, k\}$ (i.e. $f(x) = f_1(x) = \ldots = f_k(x)$). Then for any $p \in \mathbb{R}^n$:

$$\begin{aligned} f'(x;p) &= \lim_{\alpha\to 0+} \frac{f(x+\alpha p) - f(x)}{\alpha} = \lim_{\alpha\to 0+} \frac{\max_{1\leq i\leq k}\{f_i(x+\alpha p) - f_i(x)\}}{\alpha} \\ &= \max_{1\leq i\leq k} f_i'(x;p) = \max_{1\leq i\leq k} \max\{g_i^T p : g_i \in \partial f_i(x)\} \\ &= \max_{\{\lambda_i\}\subset\Delta_k} \left\{\sum_{i=1}^{k} \lambda_i \max\{g_i^T p : g_i \in \partial f_i(x)\}\right\} \\ &= \max\left\{\left(\sum_{i=1}^{k} \lambda_i g_i\right)^T p : g_i \in \partial f_i(x), \{\lambda_i\} \subset \Delta_k\right\} \\ &= \max\{g^T p : g = \sum_{i=1}^{k} \lambda_i g_i, g_i \in \partial f_i(x), \{\lambda_i\} \subset \Delta_k\} \\ &= \max\{g^T p : g \in \mathrm{Conv}\{\partial f_i(x), i \in I(x)\}\}. \end{aligned}$$

Hence, $\partial f(x) = \mathrm{Conv}\{\partial f_i(x) : i \in I(x)\}$. □

Computing subgradients

Lemma (3.1.11)

Let Δ be a set such that for any fixed $y \in \Delta$ the function $\phi(y,x)$ is closed and convex in x. Then $f(x) = \sup\{\phi(y,x) : y \in \Delta\}$ is a closed convex function with the domain

$$\mathrm{dom}\, f = \{x \in \bigcap_{y\in\Delta} \mathrm{dom}\, \phi(y,\cdot) | \exists \gamma : \phi(y,x) \leq \gamma \forall y \in \Delta\}$$

and for any $x \in \mathrm{dom}\, f$ we have

$$\partial f(x) \supseteq \mathrm{Conv}\{\partial_x \phi(y,x) : y \in I(x)\},$$

where $I(x) = \{y : \phi(y,x) = f(x)\}$.

Proof.

For any $x_0, x \in \mathrm{dom}\, f, y \in I(x_0), g \in \partial_x \phi(y, x_0)$:
$f(x) \geq \phi(y,x) \geq \phi(y,x_0) + g^T(x - x_0) = f(x_0) + g^T(x - x_0)$.
Thus, $g \in \partial f(x_0)$. □

Computing subgradients

Example (1)

Consider $f(x) = \sum_{i=1}^{m} |a_i^T x - b_i|$.

- Denoting by $f_i(x) = |a_i^T x - b_i|$, we have $f(x) = \sum_{i=1}^{m} f_i(x)$.
- If $\phi(y) = |y|$, then $f_i(x) = \phi(\mathcal{A}_i(x))$, where $\mathcal{A}_i(x) = a_i^T x - b_i$. Since

$$\partial \phi(y) = \begin{cases} \{1\}, & y > 0; \\ \{-1\}, & y < 0; \\ [-1, 1], & y = 0; \end{cases}$$

we have

$$\partial f_i(x) = \begin{cases} \{a_i\}, & a_i^T x - b_i > 0; \\ \{-a_i\}, & a_i^T x - b_i < 0; \\ [-a_i, a_i], & a_i^T x - b_i = 0. \end{cases}$$

Computing subgradients

- Denote by

$$\begin{aligned} I_-(x) &= \{i : a_i^T x - b_i < 0\}, \\ I_+(x) &= \{i : a_i^T x - b_i > 0\}, \\ I_0(x) &= \{i : a_i^T x - b_i = 0\}. \end{aligned}$$

Then

$$\partial f_i(x) = \begin{cases} \{a_i\}, & i \in I_+(x); \\ \{-a_i\}, & i \in I_-(x); \\ [-a_i, a_i], & i \in I_0(x). \end{cases} .$$

- Since $\partial f(x) = \sum_{i=1}^{m} f_i(x)$, we obtain

$$\partial f(x) = \sum_{i \in I_+(x)} a_i - \sum_{i \in I_-(x)} a_i + \sum_{i \in I_0(x)} [-a_i, a_i].$$

Computing subgradients

Example (2)

For l_1-norm $f(x) = \|x\|_1 = \sum_{i=1}^{n} |x^{(i)}|$, we have:

$$f(x) = \sum_{i=1}^{n} |a_i^T x - b_i|,$$

where $a_i = e_i$ and $b_i = 0,\ i = 1, \ldots, n$. Therefore,

$$\partial f(x) = \sum_{i \in I_+(x)} e_i - \sum_{i \in I_-(x)} e_i + \sum_{i \in I_0(x)} [-e_i, e_i],$$

where

$$\begin{aligned} I_-(x) &= \{i : x^{(i)} < 0\}, \\ I_+(x) &= \{i : x^{(i)} > 0\}, \\ I_0(x) &= \{i : x^{(i)} = 0\}. \end{aligned}$$

Computing subgradients

Example (3)

Consider the function $f(x) = \max_{1 \le i \le n} x^{(i)}$.

- Denote by $f_i(x) = x^{(i)} = e_i^T x$.
- Then $\partial f_i(x) = \{e_i\}$.
- Thus,

$$\partial f(x) = \mathrm{Conv}\{\partial f_i(x) : i \in I(x)\} = \mathrm{Conv}\{e_i : i \in I(x)\},$$

  where $I(x) = \{i : x^{(i)} = f(x)\}$.
  (Note that this implies that $\partial f(0) = \{x \in \mathbb{R}^n : \|x\|_\infty \le 1\}$)

Computing subgradients

Example (4)

For the Euclidean norm $f(x) = \|x\|$ we have:

- f is differentiable in all points except for $x = 0$ and

$$\partial f(0) = \{g \in \mathbb{R}^n : \|x\| \ge g^T x \;\; \forall x \in \mathbb{R}^n\}.$$

- Since $g^T x \le \|g\| \cdot \|x\|$, we have $\{g \in \mathbb{R}^n : \|g\| \le 1\} \subseteq \partial f(0)$.
- On the other hand, if $\|g\| > 1$, then selecting $x = g$, we obtain $g^T x = \|x\|^2 > \|x\|$, since $\|x\| > 1$. Thus, $\partial f(0) = \{g \in \mathbb{R}^n : \|g\| \le 1\}$.
- Since $f'(x) = x/\|x\|$ for $x \neq 0$, we get

$$\partial f(x) = \begin{cases} \{x/\|x\|\}, & x \neq 0; \\ \{x \in \mathbb{R}^n : \|x\| \le 1\}, & x = 0. \end{cases}$$