Convex Optimization

(EE227A: UC Berkeley)

Lecture 4
(Conjugates, subdifferentials)

31 Jan, 2013

Suvrit Sra

Organizational

- ♡ HW1 due: **14th Feb 2013** in class.
- ♡ Please LaTeX your solutions (contact TA if this is an issue)
- ♡ Discussion with classmates is ok
- ♡ Each person must submit his/her individual solutions
- ♡ Acknowledge any help you receive
- ♡ Do not copy!
- ♡ Make sure you understand the solution you submit
- ♡ Cite any source that you use
- ♡ Have fun solving problems!

Recap

- Eigenvalues, singular values, positive definiteness
- Convex sets, $\theta_1 x + \theta_2 y \in C$, $\theta_1 + \theta_2 = 1$, $\theta_i \geq 0$
- Convex functions, midpoint convex, recognizing convexity
- Norms, mixed-norms, matrix norms, dual norms
- Indicator, distance function, minimum of jointly convex
- Brief mention of other forms of convexity

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}[f(x) + f(y)] + \text{continuity} \implies f \text{ is cvx}$$

$$\nabla^2 f(x) \succeq 0 \text{ implies } f \text{ is cvx.}$$

Fenchel Conjugate

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \leq 1\right\}.$$

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \leq 1\right\}.$$

Exercise: Verify that we may write $\|u\|_* = \sup_{x \neq 0} \frac{u^T x}{\|x\|}$

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \leq 1\right\}.$$

Exercise: Verify that we may write $\|u\|_* = \sup_{x \neq 0} \frac{u^T x}{\|x\|}$

Exercise: Verify that $\|u\|_*$ is a norm.

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual norm** is

$$\|u\|_* := \sup \left\{ u^T x \mid \|x\| \leq 1 \right\}.$$

Exercise: Verify that we may write $\|u\|_* = \sup_{x \neq 0} \frac{u^T x}{\|x\|}$

Exercise: Verify that $\|u\|_*$ is a norm.

- $\|u+v\|_* = \sup \left\{ (u+v)^T x \mid \|x\| \leq 1 \right\}$
- But $\sup\, (A+B) \leq \sup A + \sup B$

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \leq 1\right\}.$$

Exercise: Verify that we may write $\|u\|_* = \sup_{x \neq 0} \frac{u^T x}{\|x\|}$

Exercise: Verify that $\|u\|_*$ is a norm.

- $\|u + v\|_* = \sup\left\{(u+v)^T x \mid \|x\| \leq 1\right\}$
- But $\sup\,(A + B) \leq \sup A + \sup B$

Exercise: Let $1/p + 1/q = 1$, where $p, q \geq 1$. Show that $\|\cdot\|_q$ is dual to $\|\cdot\|_p$. In particular, the ℓ_2-norm is self-dual.

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \leq 1\right\}.$$

Exercise: Verify that we may write $\|u\|_* = \sup_{x \neq 0} \frac{u^T x}{\|x\|}$

Exercise: Verify that $\|u\|_*$ is a norm.

- $\|u+v\|_* = \sup\left\{(u+v)^T x \mid \|x\| \leq 1\right\}$
- But $\sup\,(A+B) \leq \sup A + \sup B$

Exercise: Let $1/p + 1/q = 1$, where $p, q \geq 1$. Show that $\|\cdot\|_q$ is dual to $\|\cdot\|_p$. In particular, the ℓ_2-norm is self-dual.

Hint: Use *Hölder's inequality:* $u^T v \leq \|u\|_p \|v\|_q$

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \operatorname{dom} f} \quad x^T z - f(x).$$

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \operatorname{dom} f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \mathrm{dom}\, f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \leq 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \mathrm{dom}\, f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \leq 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \leq 1$

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \mathrm{dom}\, f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \leq 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \leq 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \leq 1$ and $z^T u > 1$

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \operatorname{dom} f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \leq 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \leq 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \leq 1$ and $z^T u > 1$
- $f^*(z) = \sup_x x^T z - f(x)$. Rewrite $x = \alpha u$, and let $\alpha \to \infty$

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \operatorname{dom} f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \le 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \le 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \le 1$ and $z^T u > 1$
- $f^*(z) = \sup_x x^T z - f(x)$. Rewrite $x = \alpha u$, and let $\alpha \to \infty$
- Then, $z^T x - \|x\| = \alpha z^T u - \|\alpha u\| = \alpha(z^T u - \|u\|)$;

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \text{dom}\, f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \le 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \le 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \le 1$ and $z^T u > 1$
- $f^*(z) = \sup_x x^T z - f(x)$. Rewrite $x = \alpha u$, and let $\alpha \to \infty$
- Then, $z^T x - \|x\| = \alpha z^T u - \|\alpha u\| = \alpha(z^T u - \|u\|)$; $\to \infty$

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \mathrm{dom}\, f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \le 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \le 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \le 1$ and $z^T u > 1$
- $f^*(z) = \sup_x x^T z - f(x)$. Rewrite $x = \alpha u$, and let $\alpha \to \infty$
- Then, $z^T x - \|x\| = \alpha z^T u - \|\alpha u\| = \alpha(z^T u - \|u\|)$; $\to \infty$
- Case (ii): Since $z^T x \le \|x\| \|z\|_*$,

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \operatorname{dom} f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \le 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \le 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \le 1$ and $z^T u > 1$
- $f^*(z) = \sup_x x^T z - f(x)$. Rewrite $x = \alpha u$, and let $\alpha \to \infty$
- Then, $z^T x - \|x\| = \alpha z^T u - \|\alpha u\| = \alpha(z^T u - \|u\|)$; $\to \infty$
- Case (ii): Since $z^T x \le \|x\| \|z\|_*$, $\quad x^T z - \|x\| \le \|x\|(\|z\|_* - 1) \le 0$.

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \mathrm{dom}\, f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \leq 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \leq 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \leq 1$ and $z^T u > 1$
- $f^*(z) = \sup_x x^T z - f(x)$. Rewrite $x = \alpha u$, and let $\alpha \to \infty$
- Then, $z^T x - \|x\| = \alpha z^T u - \|\alpha u\| = \alpha(z^T u - \|u\|)$; $\to \infty$
- Case (ii): Since $z^T x \leq \|x\| \|z\|_*$, $\quad x^T z - \|x\| \leq \|x\|(\|z\|_* - 1) \leq 0$.
- $x = 0$ maximizes $\|x\|(\|z\|_* - 1)$, hence $f(z) = 0$.

Fenchel conjugate

Def. The **Fenchel conjugate** of a function f is

$$f^*(z) := \sup_{x \in \mathrm{dom}\, f} \quad x^T z - f(x).$$

Note: f^* pointwise (over x) sup of linear functions of z, so cvx!

Example Let $f(x) = \|x\|$. We have $f^*(z) = \mathbb{I}_{\|\cdot\|_* \leq 1}(z)$. That is, conjugate of norm is the indicator function of dual norm ball.

- Consider two cases: (i) $\|z\|_* > 1$; (ii) $\|z\|_* \leq 1$
- Case (i), by definition of dual norm (sup over $z^T u$) there is a u s.t. $\|u\| \leq 1$ and $z^T u > 1$
- $f^*(z) = \sup_x x^T z - f(x)$. Rewrite $x = \alpha u$, and let $\alpha \to \infty$
- Then, $z^T x - \|x\| = \alpha z^T u - \|\alpha u\| = \alpha(z^T u - \|u\|)$; $\to \infty$
- Case (ii): Since $z^T x \leq \|x\| \|z\|_*$, $\quad x^T z - \|x\| \leq \|x\|(\|z\|_* - 1) \leq 0$.
- $x = 0$ maximizes $\|x\|(\|z\|_* - 1)$, hence $f(z) = 0$.
- Thus, $f(z) = +\infty$ if (i), and 0 if (ii), as desired.

Fenchel conjugate

Example $f(x) = ax + b$; then,

$$f^*(z) = \sup_x zx - (ax + b)$$

Fenchel conjugate

Example $f(x) = ax + b$; then,

$$\begin{aligned} f^*(z) &= \sup_x zx - (ax + b) \\ &= \infty, \quad \text{if } (z - a) \neq 0. \end{aligned}$$

Fenchel conjugate

Example $f(x) = ax + b$; then,

$$\begin{aligned} f^*(z) &= \sup_x zx - (ax + b) \\ &= \infty, \quad \text{if } (z - a) \neq 0. \end{aligned}$$

Thus, $\operatorname{dom} f^* = \{a\}$, and $f^*(a) = -b$.

Fenchel conjugate

Example $f(x) = ax + b$; then,

$$\begin{aligned} f^*(z) &= \sup_x zx - (ax+b) \\ &= \infty, \quad \text{if } (z-a) \neq 0. \end{aligned}$$

Thus, $\operatorname{dom} f^* = \{a\}$, and $f^*(a) = -b$.

Example Let $a \geq 0$, and set $f(x) = -\sqrt{a^2 - x^2}$ if $|x| \leq a$, and $+\infty$ otherwise. Then, $f^*(z) = a\sqrt{1+z^2}$.

Fenchel conjugate

Example $f(x) = ax + b$; then,

$$\begin{aligned} f^*(z) &= \sup_x zx - (ax + b) \\ &= \infty, \quad \text{if } (z - a) \neq 0. \end{aligned}$$

Thus, $\operatorname{dom} f^* = \{a\}$, and $f^*(a) = -b$.

Example Let $a \geq 0$, and set $f(x) = -\sqrt{a^2 - x^2}$ if $|x| \leq a$, and $+\infty$ otherwise. Then, $f^*(z) = a\sqrt{1 + z^2}$.

Example $f(x) = \frac{1}{2}x^T Ax$, where $A \succ 0$. Then, $f^*(z) = \frac{1}{2}z^T A^{-1} z$.

Fenchel conjugate

Example $f(x) = ax + b$; then,

$$\begin{aligned} f^*(z) &= \sup_x zx - (ax+b) \\ &= \infty, \quad \text{if } (z-a) \neq 0. \end{aligned}$$

Thus, $\operatorname{dom} f^* = \{a\}$, and $f^*(a) = -b$.

Example Let $a \geq 0$, and set $f(x) = -\sqrt{a^2 - x^2}$ if $|x| \leq a$, and $+\infty$ otherwise. Then, $f^*(z) = a\sqrt{1+z^2}$.

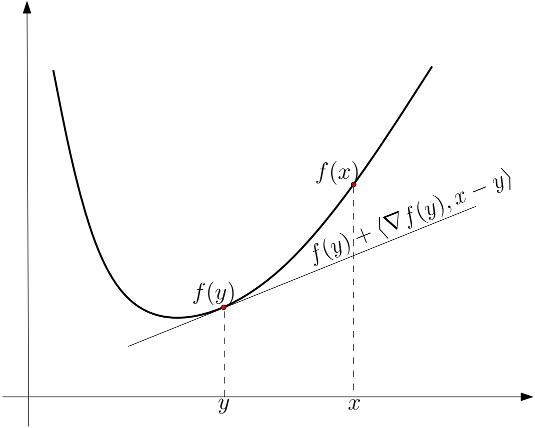
Example $f(x) = \frac{1}{2}x^T Ax$, where $A \succ 0$. Then, $f^*(z) = \frac{1}{2}z^T A^{-1} z$.

Exercise: If $f(x) = \max(0, 1-x)$, then $\operatorname{dom} f^*$ is $[-1, 0]$, and within this domain, $f^*(z) = z$.

Hint: Analyze cases: $\max(0, 1-x) = 0$; and $\max(0, 1-x) = 1-x$

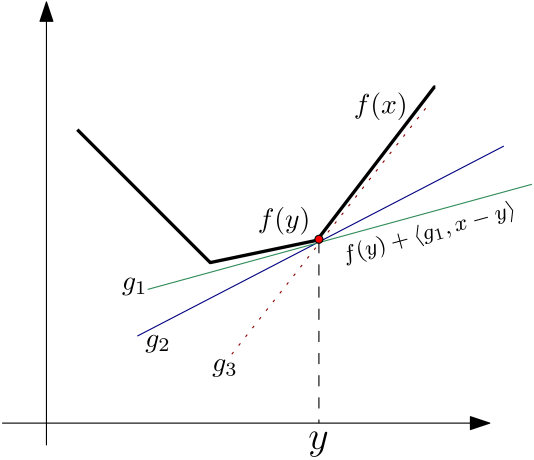
Subdifferentials

First order global underestimator



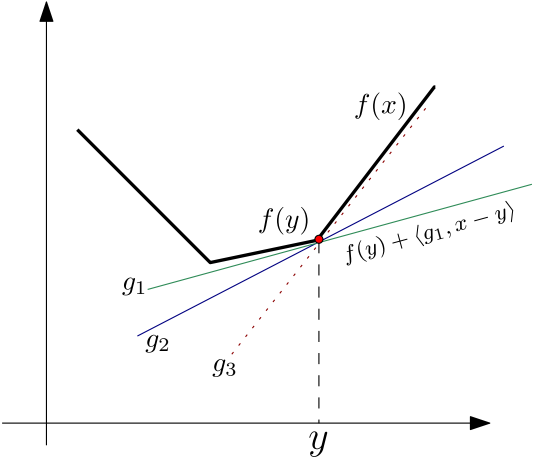
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$$

First order global underestimator



$$f(x) \geq f(y) + \langle g, x - y \rangle$$

Subgradients



g_1, g_2, g_3 are subgradients at y

Subgradients – basic facts

- f is convex, differentiable: $\nabla f(y)$ the **unique** subgradient at y
- A vector g is a subgradient at a point y if and only if $f(y) + \langle g, x - y \rangle$ is **globally** smaller than $f(x)$.
- Usually, **one** subgradient costs approx. as much as $f(x)$

Subgradients – basic facts

- f is convex, differentiable: $\nabla f(y)$ the **unique** subgradient at y
- A vector g is a subgradient at a point y if and only if $f(y) + \langle g, x - y \rangle$ is **globally** smaller than $f(x)$.
- Usually, **one** subgradient costs approx. as much as $f(x)$
- Determining all subgradients at a given point — difficult.
- Subgradient calculus—a great achievement in convex analysis

Subgradients – basic facts

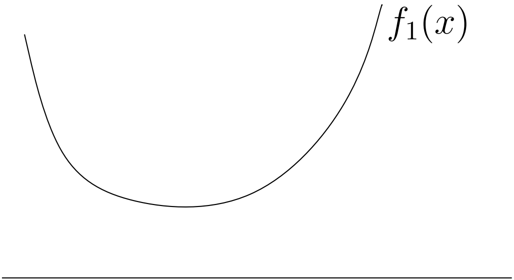
- f is convex, differentiable: $\nabla f(y)$ the **unique** subgradient at y
- A vector g is a subgradient at a point y if and only if $f(y) + \langle g, x - y \rangle$ is **globally** smaller than $f(x)$.
- Usually, **one** subgradient costs approx. as much as $f(x)$
- Determining all subgradients at a given point — difficult.
- Subgradient calculus—a great achievement in convex analysis
- Without convexity, things become wild! — advanced course

Subgradients – example

$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable

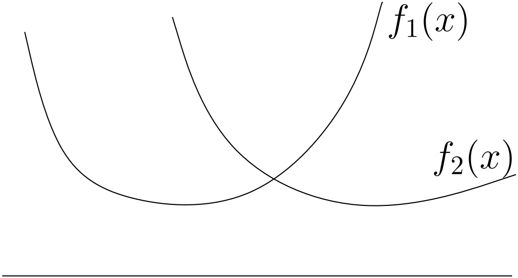
Subgradients – example

$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable



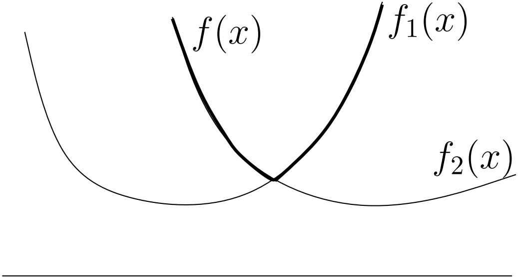
Subgradients – example

$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable



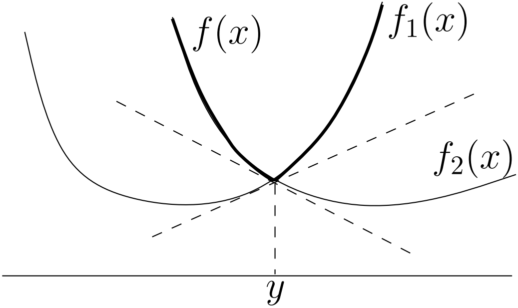
Subgradients – example

$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable



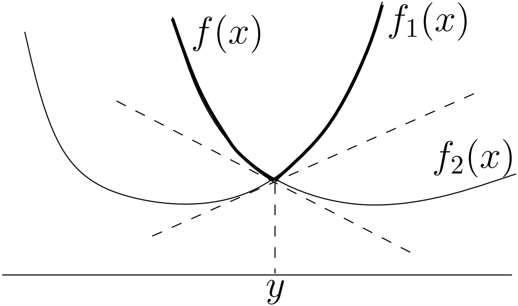
Subgradients – example

$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable



Subgradients – example

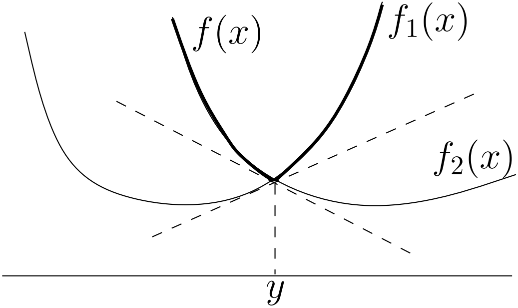
$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable



⋆ $f_1(x) > f_2(x)$: unique subgradient of f is $f_1'(x)$

Subgradients – example

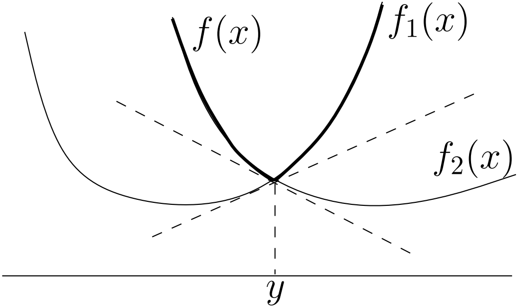
$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable



- ⋆ $f_1(x) > f_2(x)$: unique subgradient of f is $f_1'(x)$
- ⋆ $f_1(x) < f_2(x)$: unique subgradient of f is $f_2'(x)$

Subgradients – example

$f(x) := \max(f_1(x), f_2(x))$; both f_1, f_2 convex, differentiable



- ⋆ $f_1(x) > f_2(x)$: unique subgradient of f is $f_1'(x)$
- ⋆ $f_1(x) < f_2(x)$: unique subgradient of f is $f_2'(x)$
- ⋆ $f_1(y) = f_2(y)$: subgradients, the segment $[f_1'(y), f_2'(y)]$ (imagine all supporting lines turning about point y)

Subgradients

Def. A vector $g \in \mathbb{R}^n$ is called a **subgradient** at a point y, if for all $x \in \operatorname{dom} f$, it holds that

$$f(x) \geq f(y) + \langle g, x - y \rangle$$

Subdifferential

Def. The set of all subgradients at y denoted by $\partial f(y)$. This set is called **subdifferential** of f at y

Subdifferential

Def. The set of all subgradients at y denoted by $\partial f(y)$. This set is called **subdifferential** of f at y

If f is convex, $\partial f(x)$ is nice:

- ♣ If $x \in$ relative interior of $\operatorname{dom} f$, then $\partial f(x)$ nonempty

Subdifferential

Def. The set of all subgradients at y denoted by $\partial f(y)$. This set is called **subdifferential** of f at y

If f is convex, $\partial f(x)$ is nice:

- ♣ If $x \in$ relative interior of $\operatorname{dom} f$, then $\partial f(x)$ nonempty
- ♣ If f differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$

Subdifferential

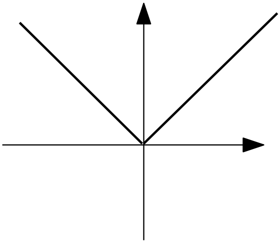
Def. The set of all subgradients at y denoted by $\partial f(y)$. This set is called **subdifferential** of f at y

If f is convex, $\partial f(x)$ is nice:

- ♣ If $x \in$ relative interior of $\operatorname{dom} f$, then $\partial f(x)$ nonempty
- ♣ If f differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$
- ♣ If $\partial f(x) = \{g\}$, then f is differentiable and $g = \nabla f(x)$

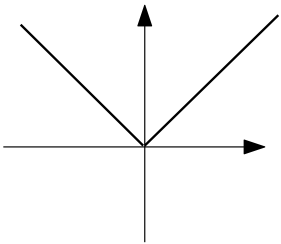
Subdifferential – example

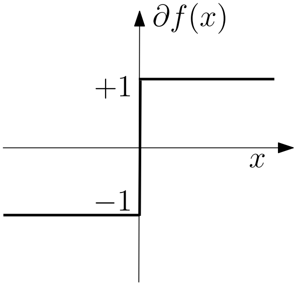
$$f(x) = |x|$$



Subdifferential – example

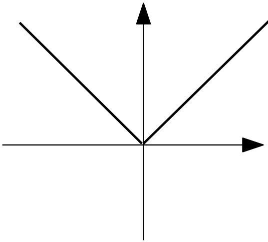
$$f(x) = |x|$$

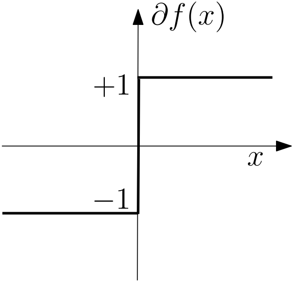




Subdifferential – example

$$f(x) = |x|$$





$$\partial|x| = \begin{cases} -1 & x < 0, \\ +1 & x > 0, \\ [-1, 1] & x = 0. \end{cases}$$

More examples

Example $f(x) = \|x\|_2$. Then,

$$\partial f(x) := \begin{cases} \|x\|_2^{-1} x & x \neq 0, \\ \{z \mid \|z\|_2 \leq 1\} & x = 0. \end{cases}$$

More examples

Example $f(x) = \|x\|_2$. Then,

$$\partial f(x) := \begin{cases} \|x\|_2^{-1} x & x \neq 0, \\ \{z \mid \|z\|_2 \leq 1\} & x = 0. \end{cases}$$

Proof.

$$\|z\|_2 \quad \geq \quad \|x\|_2 + \langle g,\, z - x\rangle$$

More examples

Example $f(x) = \|x\|_2$. Then,

$$\partial f(x) := \begin{cases} \|x\|_2^{-1} x & x \neq 0, \\ \{z \mid \|z\|_2 \leq 1\} & x = 0. \end{cases}$$

Proof.

$$\begin{aligned} \|z\|_2 &\geq \|x\|_2 + \langle g,\, z - x \rangle \\ \|z\|_2 &\geq \langle g,\, z \rangle \end{aligned}$$

More examples

Example $f(x) = \|x\|_2$. Then,

$$\partial f(x) := \begin{cases} \|x\|_2^{-1} x & x \neq 0, \\ \{z \mid \|z\|_2 \leq 1\} & x = 0. \end{cases}$$

Proof.

$$\begin{aligned} \|z\|_2 &\geq \|x\|_2 + \langle g,\, z - x \rangle \\ \|z\|_2 &\geq \langle g,\, z \rangle \\ &\implies \|g\|_2 \leq 1. \end{aligned}$$

More examples

Example A convex function need not be subdifferentiable everywhere.

Let

$$f(x) := \begin{cases} -(1 - \|x\|_2^2)^{1/2} & \text{if } \|x\|_2 \le 1, \\ +\infty & \text{otherwise.} \end{cases}$$

f diff. for all x with $\|x\|_2 < 1$, but $\partial f(x) = \emptyset$ whenever $\|x\|_2 \ge 1$.

Calculus

Recall basic calculus

If f and k are differentiable, we know that

- **Addition**: $\nabla(f+k)(x) = \nabla f(x) + \nabla k(x)$
- **Scaling**: $\nabla(\alpha f(x)) = \alpha \nabla f(x)$

Recall basic calculus

If f and k are differentiable, we know that

- **Addition**: $\nabla(f+k)(x) = \nabla f(x) + \nabla k(x)$
- **Scaling**: $\nabla(\alpha f(x)) = \alpha \nabla f(x)$

Chain rule

If $f : \mathbb{R}^n \to \mathbb{R}^m$, and $k : \mathbb{R}^m \to \mathbb{R}^p$. Let $h : \mathbb{R}^n \to \mathbb{R}^p$ be the composition $h(x) = (k \circ f)(x) = k(f(x))$. Then,

$$Dh(x) = Dk(f(x))Df(x).$$

Recall basic calculus

If f and k are differentiable, we know that

- **Addition**: $\nabla(f+k)(x) = \nabla f(x) + \nabla k(x)$
- **Scaling**: $\nabla(\alpha f(x)) = \alpha \nabla f(x)$

Chain rule

If $f : \mathbb{R}^n \to \mathbb{R}^m$, and $k : \mathbb{R}^m \to \mathbb{R}^p$. Let $h : \mathbb{R}^n \to \mathbb{R}^p$ be the composition $h(x) = (k \circ f)(x) = k(f(x))$. Then,

$$Dh(x) = Dk(f(x))Df(x).$$

Example If $f : \mathbb{R}^n \to \mathbb{R}$ and $k : \mathbb{R} \to \mathbb{R}$, then using the fact that $\nabla h(x) = [Dh(x)]^T$, we obtain

$$\nabla h(x) = k'(f(x))\nabla f(x).$$

Subgradient calculus

♠ Finding one subgradient within $\partial f(x)$

Subgradient calculus

- ♠ Finding one subgradient within $\partial f(x)$
- ♠ Determining entire subdifferential $\partial f(x)$ at a point x

Subgradient calculus

- ♠ Finding one subgradient within $\partial f(x)$
- ♠ Determining entire subdifferential $\partial f(x)$ at a point x
- ♠ Do we have the chain rule?

Subgradient calculus

- ♠ Finding one subgradient within $\partial f(x)$
- ♠ Determining entire subdifferential $\partial f(x)$ at a point x
- ♠ Do we have the chain rule?
- ♠ Usually not easy!

Subgradient calculus

- If f is differentiable, $\partial f(x) = \{\nabla f(x)\}$

Subgradient calculus

- If f is differentiable, $\partial f(x) = \{\nabla f(x)\}$
- **Scaling** $\alpha > 0$, $\partial(\alpha f)(x) = \alpha \partial f(x) = \{\alpha g \mid g \in \partial f(x)\}$

Subgradient calculus

- If f is differentiable, $\partial f(x) = \{\nabla f(x)\}$
- **Scaling** $\alpha > 0$, $\partial(\alpha f)(x) = \alpha \partial f(x) = \{\alpha g \mid g \in \partial f(x)\}$
- **Addition*****:** $\partial(f + k)(x) = \partial f(x) + \partial k(x)$ (set addition)

Subgradient calculus

- If f is differentiable, $\partial f(x) = \{\nabla f(x)\}$
- **Scaling** $\alpha > 0$, $\partial(\alpha f)(x) = \alpha \partial f(x) = \{\alpha g \mid g \in \partial f(x)\}$
- **Addition***: $\partial(f + k)(x) = \partial f(x) + \partial k(x)$ (set addition)
- **Chain rule***: Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$, and $h : \mathbb{R}^n \to \mathbb{R}$ be given by $h(x) = f(Ax + b)$. Then,

$$\partial h(x) = A^T \partial f(Ax + b).$$

Subgradient calculus

- If f is differentiable, $\partial f(x) = \{\nabla f(x)\}$
- **Scaling** $\alpha > 0$, $\partial(\alpha f)(x) = \alpha \partial f(x) = \{\alpha g \mid g \in \partial f(x)\}$
- **Addition*:** $\partial(f + k)(x) = \partial f(x) + \partial k(x)$ (set addition)
- **Chain rule*:** Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$, and $h : \mathbb{R}^n \to \mathbb{R}$ be given by $h(x) = f(Ax + b)$. Then,
$$\partial h(x) = A^T \partial f(Ax + b).$$
- **Chain rule*:** $h(x) = f \circ k$, where $k : X \to Y$ is diff.
$$\partial h(x) = \partial f(k(x)) \circ Dk(x) = [Dk(x)]^T \partial f(k(x))$$

Subgradient calculus

- If f is differentiable, $\partial f(x) = \{\nabla f(x)\}$
- **Scaling** $\alpha > 0$, $\partial(\alpha f)(x) = \alpha \partial f(x) = \{\alpha g \mid g \in \partial f(x)\}$
- **Addition*:** $\partial(f + k)(x) = \partial f(x) + \partial k(x)$ (set addition)
- **Chain rule*:** Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$, and $h : \mathbb{R}^n \to \mathbb{R}$ be given by $h(x) = f(Ax + b)$. Then,

$$\partial h(x) = A^T \partial f(Ax + b).$$

- **Chain rule*:** $h(x) = f \circ k$, where $k : X \to Y$ is diff.

$$\partial h(x) = \partial f(k(x)) \circ Dk(x) = [Dk(x)]^T \partial f(k(x))$$

- **Max function*:** If $f(x) := \max_{1 \le i \le m} f_i(x)$, then

$$\partial f(x) = \text{conv} \bigcup \{\partial f_i(x) \mid f_i(x) = f(x)\},$$

convex hull over subdifferentials of "active" functions at x

Subgradient calculus

- If f is differentiable, $\partial f(x) = \{\nabla f(x)\}$
- **Scaling** $\alpha > 0$, $\partial(\alpha f)(x) = \alpha \partial f(x) = \{\alpha g \mid g \in \partial f(x)\}$
- **Addition***: $\partial(f+k)(x) = \partial f(x) + \partial k(x)$ (set addition)
- **Chain rule***: Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$, and $h : \mathbb{R}^n \to \mathbb{R}$ be given by $h(x) = f(Ax+b)$. Then,
$$\partial h(x) = A^T \partial f(Ax+b).$$
- **Chain rule***: $h(x) = f \circ k$, where $k : X \to Y$ is diff.
$$\partial h(x) = \partial f(k(x)) \circ Dk(x) = [Dk(x)]^T \partial f(k(x))$$
- **Max function***: If $f(x) := \max_{1\le i\le m} f_i(x)$, then
$$\partial f(x) = \text{conv} \bigcup \{\partial f_i(x) \mid f_i(x) = f(x)\},$$
convex hull over subdifferentials of "active" functions at x
- **Conjugation:** $z \in \partial f(x)$ if and only if $x \in \partial f^*(z)$

Examples

It can happen that $\partial(f_1 + f_2) \neq \partial f_1 + \partial f_2$

Examples

It can happen that $\partial(f_1 + f_2) \neq \partial f_1 + \partial f_2$

Example Define f_1 and f_2 by

$$f_1(x) := \begin{cases} -2\sqrt{x} & \text{if } x \geq 0, \\ +\infty & \text{if } x < 0, \end{cases} \quad \text{and} \quad f_2(x) := \begin{cases} +\infty & \text{if } x > 0, \\ -2\sqrt{-x} & \text{if } x \leq 0. \end{cases}$$

Then, $f = \max\{f_1, f_2\} = \mathbb{I}_0$, whereby $\partial f(0) = \mathbb{R}$
But $\partial f_1(0) = \partial f_2(0) = \emptyset$.

Examples

It can happen that $\partial(f_1 + f_2) \neq \partial f_1 + \partial f_2$

Example Define f_1 and f_2 by

$$f_1(x) := \begin{cases} -2\sqrt{x} & \text{if } x \geq 0, \\ +\infty & \text{if } x < 0, \end{cases} \quad \text{and} \quad f_2(x) := \begin{cases} +\infty & \text{if } x > 0, \\ -2\sqrt{-x} & \text{if } x \leq 0. \end{cases}$$

Then, $f = \max\{f_1, f_2\} = \mathbb{I}_0$, whereby $\partial f(0) = \mathbb{R}$
But $\partial f_1(0) = \partial f_2(0) = \emptyset$.

However, $\partial f_1(x) + \partial f_2(x) \subset \partial(f_1 + f_2)(x)$ always holds.

Examples

Example $f(x) = \|x\|_\infty$. Then,

$$\partial f(0) = \mathsf{conv}\left\{\pm e_1, \ldots, \pm e_n\right\},$$

where e_i is i-th canonical basis vector (column of identity matrix).

Examples

Example $f(x) = \|x\|_\infty$. Then,

$$\partial f(0) = \text{conv}\left\{\pm e_1, \ldots, \pm e_n\right\},$$

where e_i is i-th canonical basis vector (column of identity matrix).

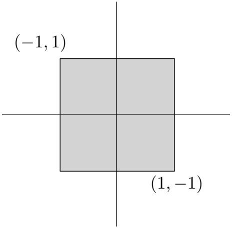
To prove, notice that $f(x) = \max_{1 \le i \le n} \left\{|e_i^T x|\right\}$.

Example (S. Boyd)

Example Let $f(x) = \max\left\{s^T x \mid s_i \in \{-1, 1\}\right\}$ (2^n members)

Example (S. Boyd)

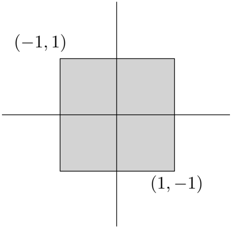
Example Let $f(x) = \max\left\{s^T x \mid s_i \in \{-1, 1\}\right\}$ (2^n members)



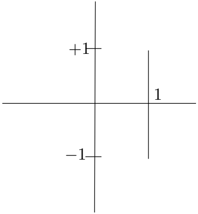
∂f at $x = (0,0)$

Example (S. Boyd)

Example Let $f(x) = \max\left\{s^T x \mid s_i \in \{-1, 1\}\right\}$ (2^n members)



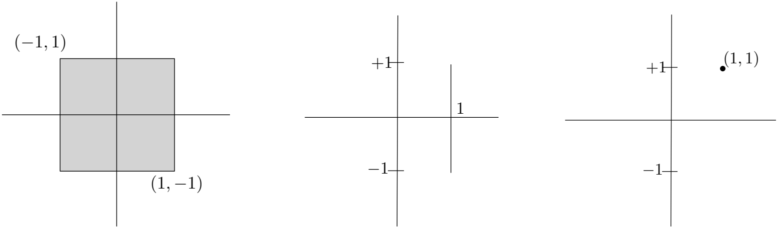
∂f at $x = (0, 0)$



∂f at $x = (1, 0)$

Example (S. Boyd)

Example Let $f(x) = \max\left\{s^T x \mid s_i \in \{-1, 1\}\right\}$ (2^n members)



∂f at $x = (0, 0)$ ∂f at $x = (1, 0)$ ∂f at $x = (1, 1)$

Rules for subgradients

Subgradient for pointwise sup

$$f(x) := \sup_{y \in \mathcal{Y}} \quad h(x, y)$$

Getting $\partial f(x)$ is complicated!

Subgradient for pointwise sup

$$f(x) := \sup_{y \in \mathcal{Y}} \quad h(x, y)$$

Getting $\partial f(x)$ is complicated!

Simple way to obtain some $g \in \partial f(x)$:

Subgradient for pointwise sup

$$f(x) := \sup_{y \in \mathcal{Y}} \quad h(x, y)$$

Getting $\partial f(x)$ is complicated!

Simple way to obtain some $g \in \partial f(x)$:

- Pick any y^* for which $h(x, y^*) = f(x)$

Subgradient for pointwise sup

$$f(x) := \sup_{y \in \mathcal{Y}} \quad h(x, y)$$

Getting $\partial f(x)$ is complicated!

Simple way to obtain some $g \in \partial f(x)$:

- Pick any y^* for which $h(x, y^*) = f(x)$
- Pick any subgradient $g \in \partial h(x, y^*)$

Subgradient for pointwise sup

$$f(x) := \sup_{y \in \mathcal{Y}} \quad h(x, y)$$

Getting $\partial f(x)$ is complicated!

Simple way to obtain some $g \in \partial f(x)$:

- Pick any y^* for which $h(x, y^*) = f(x)$
- Pick any subgradient $g \in \partial h(x, y^*)$
- This $g \in \partial f(x)$

Subgradient for pointwise sup

$$f(x) := \sup_{y \in \mathcal{Y}} \quad h(x, y)$$

Getting $\partial f(x)$ is complicated!

Simple way to obtain some $g \in \partial f(x)$:

- Pick any y^* for which $h(x, y^*) = f(x)$
- Pick any subgradient $g \in \partial h(x, y^*)$
- This $g \in \partial f(x)$

$$\begin{aligned} h(z, y^*) &\geq h(x, y^*) + g^T(z - x) \\ h(z, y^*) &\geq f(x) + g^T(z - x) \end{aligned}$$

Subgradient for pointwise sup

$$f(x) := \sup_{y \in \mathcal{Y}} \quad h(x, y)$$

Getting $\partial f(x)$ is complicated!

Simple way to obtain some $g \in \partial f(x)$:

- Pick any y^* for which $h(x, y^*) = f(x)$
- Pick any subgradient $g \in \partial h(x, y^*)$
- This $g \in \partial f(x)$

$$\begin{aligned}
h(z, y^*) &\geq h(x, y^*) + g^T(z - x) \\
h(z, y^*) &\geq f(x) + g^T(z - x) \\
f(z) &\geq h(z, y) \qquad \text{(because of sup)} \\
f(z) &\geq f(x) + g^T(z - x).
\end{aligned}$$

Example

Suppose $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. And

$$f(x) := \max_{1 \leq i \leq n} (a_i^T x + b_i).$$

This f a max (in fact, over a finite number of terms)

Example

Suppose $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. And

$$f(x) := \max_{1 \leq i \leq n} (a_i^T x + b_i).$$

This f a max (in fact, over a finite number of terms)

- Suppose $f(x) = a_k^T x + b_k$ for some index k

Example

Suppose $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. And

$$f(x) := \max_{1 \leq i \leq n} (a_i^T x + b_i).$$

This f a max (in fact, over a finite number of terms)

- Suppose $f(x) = a_k^T x + b_k$ for some index k
- Here $f(x; y) = f_k(x) = a_k^T x + b_k$, and $\partial f_k(x) = \{\nabla f_k(x)\}$

Example

Suppose $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. And

$$f(x) := \max_{1 \le i \le n} (a_i^T x + b_i).$$

This f a max (in fact, over a finite number of terms)

- Suppose $f(x) = a_k^T x + b_k$ for some index k
- Here $f(x; y) = f_k(x) = a_k^T x + b_k$, and $\partial f_k(x) = \{\nabla f_k(x)\}$
- Hence, $a_k \in \partial f(x)$ works!

Subgradient of expectation

Suppose $f = \mathbf{E} f(x,u)$, where f is convex in x for each u (an r.v.)

$$f(x) := \int f(x,u)p(u)du$$

Subgradient of expectation

Suppose $f = \mathbf{E} f(x, u)$, where f is convex in x for each u (an r.v.)

$$f(x) := \int f(x, u) p(u) du$$

- For each u choose any $g(x, u) \in \partial_x f(x, u)$

Subgradient of expectation

Suppose $f = \mathbf{E} f(x, u)$, where f is convex in x for each u (an r.v.)

$$f(x) := \int f(x, u) p(u) du$$

- For each u choose any $g(x, u) \in \partial_x f(x, u)$
- Then, $g = \int g(x, u) p(u) du = \mathbf{E} g(x, u) \in \partial f(x)$

Subgradient of composition

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ cvx and nondecreasing; each f_i cvx

$$f(x) := h(f_1(x), f_2(x), \ldots, f_n(x)).$$

Subgradient of composition

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ cvx and nondecreasing; each f_i cvx

$$f(x) := h(f_1(x), f_2(x), \ldots, f_n(x)).$$

To find a vector $g \in \partial f(x)$, we may:

Subgradient of composition

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ cvx and nondecreasing; each f_i cvx

$$f(x) := h(f_1(x), f_2(x), \ldots, f_n(x)).$$

To find a vector $g \in \partial f(x)$, we may:

- For $i = 1$ to n, compute $g_i \in \partial f_i(x)$

Subgradient of composition

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ cvx and nondecreasing; each f_i cvx

$$f(x) := h(f_1(x), f_2(x), \ldots, f_n(x)).$$

To find a vector $g \in \partial f(x)$, we may:

- For $i = 1$ to n, compute $g_i \in \partial f_i(x)$
- Compute $u \in \partial h(f_1(x), \ldots, f_n(x))$

Subgradient of composition

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ cvx and nondecreasing; each f_i cvx

$$f(x) := h(f_1(x), f_2(x), \ldots, f_n(x)).$$

To find a vector $g \in \partial f(x)$, we may:

- For $i = 1$ to n, compute $g_i \in \partial f_i(x)$
- Compute $u \in \partial h(f_1(x), \ldots, f_n(x))$
- Set $g = u_1 g_1 + u_2 g_2 + \cdots + u_n g_n$; this $g \in \partial f(x)$

Subgradient of composition

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ cvx and nondecreasing; each f_i cvx

$$f(x) := h(f_1(x), f_2(x), \ldots, f_n(x)).$$

To find a vector $g \in \partial f(x)$, we may:

- For $i = 1$ to n, compute $g_i \in \partial f_i(x)$
- Compute $u \in \partial h(f_1(x), \ldots, f_n(x))$
- Set $g = u_1 g_1 + u_2 g_2 + \cdots + u_n g_n$; this $g \in \partial f(x)$
- Compare with $\nabla f(x) = J \nabla h(x)$, where J matrix of $\nabla f_i(x)$

Subgradient of composition

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ cvx and nondecreasing; each f_i cvx

$$f(x) := h(f_1(x), f_2(x), \ldots, f_n(x)).$$

To find a vector $g \in \partial f(x)$, we may:

- For $i = 1$ to n, compute $g_i \in \partial f_i(x)$
- Compute $u \in \partial h(f_1(x), \ldots, f_n(x))$
- Set $g = u_1 g_1 + u_2 g_2 + \cdots + u_n g_n$; this $g \in \partial f(x)$
- Compare with $\nabla f(x) = J \nabla h(x)$, where J matrix of $\nabla f_i(x)$

Exercise: Verify $g \in \partial f(x)$ by showing $f(z) \geq f(x) + g^T(z - x)$

References

1. R. T. Rockafellar. *Convex Analysis*
2. S. Boyd (Stanford); EE364b Lecture Notes.