ORIE 6326: Convex Optimization

Subgradients

Professor Udell

Operations Research and Information Engineering
Cornell

April 25, 2017

Some slides adapted from Stanford EE364b

Outline

Subgradients

Properties

Subgradient calculus

Optimality

Basic inequality

recall basic inequality for convex differentiable f:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

- first-order approximation of f at x is global underestimator
- $(\nabla f(x), -1)$ supports $\mathbf{epi}\, f$ at $(x, f(x))$

what if f is not differentiable?

Non-differentiable functions

are these functions differentiable?

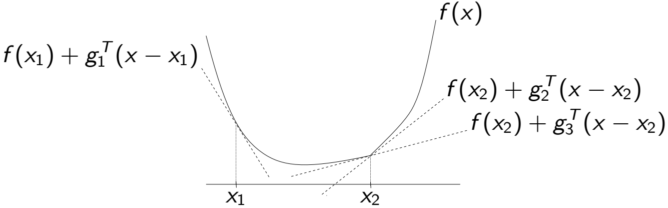
- $|t|$ for $t \in \mathbf{R}$
- $\|x\|_1$ for $x \in \mathbf{R}^n$
- $\|X\|_*$ for $X \in \mathbf{R}^{n \times n}$
- $\max_i a_i^T x + b_i$ for $x \in \mathbf{R}^n$
- $\lambda_{\max}(X)$ for $X \in \mathbf{R}^{n \times n}$
- indicators of convex sets $\mathcal{C}$

if not, where? can we find underestimators for them?

Subgradient of a function

g is a **subgradient** of f (not necessarily convex) at x if

$$f(y) \geq f(x) + g^T(y - x) \quad \text{for all } y$$



g_2, g_3 are subgradients at x_2; g_1 is a subgradient at x_1

Subgradients and convexity

- g is a subgradient of f at x iff $(g, -1)$ supports **epi** f at $(x, f(x))$
- g is a subgradient iff $f(x) + g^T(y - x)$ is a global (affine) underestimator of f
- if f is convex and differentiable, $\nabla f(x)$ is a subgradient of f at x

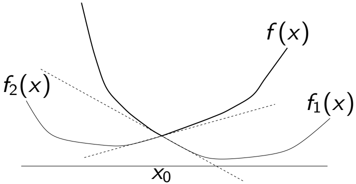
subgradients come up in several contexts:

- algorithms for nondifferentiable convex optimization
- convex analysis, *e.g.*, optimality conditions, duality for nondifferentiable problems

(if $f(y) \leq f(x) + g^T(y - x)$ for all y, then g is a **supergradient**)

Example

$f = \max\{f_1, f_2\}$, with f_1, f_2 convex and differentiable



- $f_1(x_0) > f_2(x_0)$: unique subgradient $g = \nabla f_1(x_0)$
- $f_2(x_0) > f_1(x_0)$: unique subgradient $g = \nabla f_2(x_0)$
- $f_1(x_0) = f_2(x_0)$: subgradients form a line segment $[\nabla f_1(x_0), \nabla f_2(x_0)]$

Subdifferential

set of all subgradients of f at x is called the **subdifferential** of f at x, denoted $\partial f(x)$

$$\partial f(x) = \{g : f(y) \geq f(x) + g^T(y - x) \quad \forall y\}$$

for any f,

- $\partial f(x)$ is a closed convex set (can be empty)
- $\partial f(x) = \emptyset$ if $f(x) = \infty$

proof: use the definition

Subdifferential

set of all subgradients of f at x is called the **subdifferential** of f at x, denoted $\partial f(x)$

$$\partial f(x) = \{g : f(y) \geq f(x) + g^T(y - x) \quad \forall y\}$$

for any f,

- $\partial f(x)$ is a closed convex set (can be empty)
- $\partial f(x) = \emptyset$ if $f(x) = \infty$

proof: use the definition

if f is convex,

- $\partial f(x)$ is nonempty, for $x \in \textbf{relint dom}\, f$
- $\partial f(x) = \{\nabla f(x)\}$, if f is differentiable at x
- if $\partial f(x) = \{g\}$, then f is differentiable at x and $g = \nabla f(x)$

Compute subgradient via definition

$g \in \partial f(x)$ iff

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \textbf{dom}(f)$$

example. let $f(x) = |x|$ for $x \in \mathbf{R}$. suppose $s \in \textbf{sign}(x)$, where

$$\textbf{sign}(x) = \begin{cases} \{1\} & x > 0 \\ [-1, 1] & x = 0 \\ -\{1\} & x < 0. \end{cases}$$

then

$$f(y) = \max(y, -y) \geq sy = s(x + y - x) = |x| + s(y - x)$$

Compute subgradient via definition

$g \in \partial f(x)$ iff

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \mathbf{dom}(f)$$

example. let $f(x) = |x|$ for $x \in \mathbf{R}$. suppose $s \in \mathbf{sign}(x)$, where

$$\mathbf{sign}(x) = \begin{cases} \{1\} & x > 0 \\ [-1, 1] & x = 0 \\ -\{1\} & x < 0. \end{cases}$$

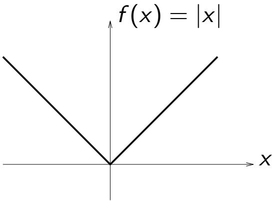
then

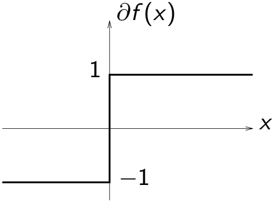
$$f(y) = \max(y, -y) \geq sy = s(x + y - x) = |x| + s(y - x)$$

so $\mathbf{sign}(x) \subseteq \partial f(x)$ (in fact, holds with equality)

Subgradient of $|x|$

$f(x) = |x|$ for $x \in \mathbf{R}$





righthand plot shows $\bigcup \{(x, g) \mid x \in \mathbf{R},\ g \in \partial f(x)\}$

Compute subgradient via definition

$g \in \partial f(x)$ iff

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \mathbf{dom}(f)$$

example. let $f(x) = \max_i a_i^T x + b_i$.

Compute subgradient via definition

$g \in \partial f(x)$ iff

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \mathbf{dom}(f)$$

example. let $f(x) = \max_i a_i^T x + b_i$. then for any i,

$$\begin{aligned} f(y) &= \max_i a_i^T y + b_i \\ &\geq a_i^T y + b_i \\ &= a_i^T(x + y - x) + b_i \\ &= a_i^T x + b_i + a_i^T(y - x) \\ &= f(x) + a_i^T(y - x), \end{aligned}$$

where the last line holds for $i \in \operatorname{argmax}_j a_j^T x + b_j$. so

- $a_i \in \partial f(x)$ for each $i \in \operatorname{argmax}_j a_j^T x + b_j$
- $\partial f(x)$ is convex, so

$$\mathbf{conv}\{a_i : i \in \operatorname*{argmax}_j a_j^T x + b_j\} \subseteq \partial f(x)$$

Compute subgradient via definition

$g \in \partial f(x)$ iff

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \mathbf{dom}(f)$$

example. let $f(X) = \lambda_{\max}(X)$.

Compute subgradient via definition

$g \in \partial f(x)$ iff

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \textbf{dom}(f)$$

example. let $f(X) = \lambda_{\max}(X)$. then

$$\begin{aligned}
f(Y) &= \sup \|v\| \leq 1 v^T Y v \\
&= \sup \|v\| \leq 1 v^T (X + Y - X) v, \quad \|v\| \leq 1 \\
&= \sup \|v\| \leq 1 \left(v^T X v + v^T (Y - X) v \right), \quad \|v\| \leq 1 \\
&= v^T X v + \textbf{tr}(v v^T (Y - X)), \quad v \in \underset{\|v\| \leq 1}{\operatorname{argmax}} v^T X v \\
&= \lambda_{\max}(X) + \textbf{tr}(v v^T (Y - X)), \quad v \in \underset{\|v\| \leq 1}{\operatorname{argmax}} v^T X v
\end{aligned}$$

so

- $vv^T \in \partial f(x)$ for each $v \in \operatorname{argmax}_{\|v\| \leq 1} v^T X v$
- $\partial f(x)$ is convex, so

$$\textbf{conv}\{vv^T : v \in \underset{\|v\| \leq 1}{\operatorname{argmax}} v^T X v\} \subseteq \partial f(x)$$

Outline

Subgradients

Properties

Subgradient calculus

Optimality

Properties of subgradients

subgradient inequality:

$$g \in \partial f(x) \iff f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \mathbf{dom}(f)$$

for convex f, we'll show

- subgradients are monotone: for any $x, y \in \mathbf{dom}\, f$, $g_y \in \partial f(y)$, and $g_x \in \partial f(x)$,
$$(g_y - g_x)^T(y - x) \geq 0$$
- $\partial f(x)$ is continuous: if f is (lower semi-)continuous, $x^{(k)} \to x$, $g^{(k)} \to g$, and $g^{(k)} \in \partial f(x^{(k)})$ for each k, then $g \in \partial f(x)$
- $\partial f(x) = \mathrm{argmax}\, g^T x - f(x)$

these will help us compute subgradients

Subgradients are monotone

fact. for any $x, y \in \mathbf{dom}\, f$, $g_y \in \partial f(y)$, and $g_x \in \partial f(x)$,

$$(g_y - g_x)^T (y - x) \geq 0$$

proof. same as for differentiable case:

$$f(y) \geq f(x) + g_x^T (y - x) \qquad f(x) \geq f(y) + g_y^T (x - y)$$

add these to get

$$(g_y - g_x)^T (y - x) \geq 0$$

Subgradients are preserved under limits

subgradient inequality:

$$g \in \partial f(x) \iff f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \mathbf{dom}(f)$$

fact. if f is (lower semi-)continuous, $x^{(k)} \to x$, $g^{(k)} \to g$, and $g^{(k)} \in \partial f(x^{(k)})$ for each k, then $g \in \partial f(x)$

proof.

Subgradients are preserved under limits

subgradient inequality:

$$g \in \partial f(x) \iff f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \mathbf{dom}(f)$$

fact. if f is (lower semi-)continuous, $x^{(k)} \to x$, $g^{(k)} \to g$, and $g^{(k)} \in \partial f(x^{(k)})$ for each k, then $g \in \partial f(x)$

proof. For each k and for every y,

$$\begin{aligned} f(y) &\geq f(x^{(k)}) + (g^{(k)})^T(y - x^{(k)}) \\ \lim_{k\to\infty} f(y) &\geq \lim_{k\to\infty} f(x^{(k)}) + (g^{(k)})^T(y - x^{(k)}) \\ f(y) &\geq f(x) + g^T(y - x) \end{aligned}$$

moral. To find a subgradient $g \in \partial f(x)$, find points $x^{(k)} \to x$ where f is differentiable, and let $g = \lim_{k\to\infty} \nabla f(x^{(k)})$.

Subgradients are preserved under limits: example

consider $f(x) = |x|$. we know

$$\partial f(x) = \begin{cases} \{-1\} & x < 0 \\ ? & x = 0 \\ \{1\} & x > 0 \end{cases}$$

so

- $\lim_{x \to 0^+} \nabla(x) = 1$
- $\lim_{x \to 0^-} \nabla(x) = -1$

hence

Subgradients are preserved under limits: example

consider $f(x) = |x|$. we know

$$\partial f(x) = \begin{cases} \{-1\} & x < 0 \\ ? & x = 0 \\ \{1\} & x > 0 \end{cases}$$

so

- $\lim_{x \to 0^+} \nabla(x) = 1$
- $\lim_{x \to 0^-} \nabla(x) = -1$

hence

- $-1 \in \partial f(0)$ and $-1 \in \partial f(0)$
- $\partial f(0)$ is convex, so $[-1, 1] \subseteq \partial f(0)$
- and $\partial f(0)$ is monotone, so $[-1, 1] = \partial f(0)$

Convex functions can't be very non-differentiable

Theorem. (Rockafellar, Convex Analysis, Thm 25.5)
a convex function is differentiable almost everywhere on the interior of its domain.

In other words, if you pick $x \in \textbf{dom}\, f$ uniformly at random, then with probability 1, f is differentiable at x.

Convex functions can't be very non-differentiable

Theorem. (Rockafellar, Convex Analysis, Thm 25.5)
a convex function is differentiable almost everywhere on the interior of its domain.

In other words, if you pick $x \in \textbf{dom}\, f$ uniformly at random, then with probability 1, f is differentiable at x. **intuition.** (in $\mathbf{R}$.) Subgradients are closed convex sets, so in $\mathbf{R}$ subgradients are closed intervals. Subgradients are monotone, so the interiors of the intervals do not intersect. (Use monotone (sub)gradient inequality

$$\tilde{\nabla} f(y)^T(y-x) \geq \tilde{\nabla} f(x)^T(y-x);$$

notice $(y-x)$ is scalar to see $\tilde{\nabla} f(y) \geq \tilde{\nabla} f(x)$ if $y \geq x$.) At each nondifferentiable point x, $\tilde{\nabla} f(y)$ jumps up by some finite amount! It can't do that too often.

More formally, $|\partial f(x)|$ is strictly positive for each x where f is nondifferentiable; and the sum of uncountably many positive numbers is infinite. So the number of x's where f is not differentiable must be countable over the interior of the domain of f; and hence, f is a.e. differentiable on the interior of its domain.

Convex functions can't be very non-differentiable

Theorem. (Rockafellar, Convex Analysis, Thm 25.5)
a convex function is differentiable almost everywhere on the interior of its domain.

In other words, if you pick $x \in \textbf{dom}\, f$ uniformly at random, then with probability 1, f is differentiable at x. **intuition.** (in $\mathbf{R}$.) Subgradients are closed convex sets, so in $\mathbf{R}$ subgradients are closed intervals. Subgradients are monotone, so the interiors of the intervals do not intersect. (Use monotone (sub)gradient inequality

$$\tilde{\nabla} f(y)^T(y-x) \geq \tilde{\nabla} f(x)^T(y-x);$$

notice $(y-x)$ is scalar to see $\tilde{\nabla} f(y) \geq \tilde{\nabla} f(x)$ if $y \geq x$.) At each nondifferentiable point x, $\tilde{\nabla} f(y)$ jumps up by some finite amount! It can't do that too often.

More formally, $|\partial f(x)|$ is strictly positive for each x where f is nondifferentiable; and the sum of uncountably many positive numbers is infinite. So the number of x's where f is not differentiable must be countable over the interior of the domain of f; and hence, f is a.e. differentiable on the interior of its domain. **Moral.** For any x, you can always find a sequence of points $x^{(k)} \to x$ where f is differentiable.

Subgradients and fenchel conjugates

fact. $g \in \partial f(x) \iff f^*(g) + f(x) = g^T x$

(recall the conjugate function $f^*(g) = \sup_x g^T x - f(x)$.)

Subgradients and fenchel conjugates

proof. if $f^*(g) + f(x) = g^T x$,

$$\begin{array}{rcl} f^*(g) & = & \sup_y g^T y - f(y) \\ & \geq & g^T y - f(y) \quad \forall y \\ f(y) & \geq & g^T y - f^*(g) \quad \forall y \\ & = & g^T y - g^T x + f(x) \quad \forall y \\ & = & g^T (y - x) + f(x) \quad \forall y \end{array}$$

so $g \in \partial f(x)$. conversely, if $g \in \partial f(x)$,

Subgradients and fenchel conjugates

proof. if $f^*(g) + f(x) = g^T x$,

$$\begin{array}{rcl} f^*(g) &=& \sup_y g^T y - f(y) \\ &\geq& g^T y - f(y) \quad \forall y \\ f(y) &\geq& g^T y - f^*(g) \quad \forall y \\ &=& g^T y - g^T x + f(x) \quad \forall y \\ &=& g^T (y - x) + f(x) \quad \forall y \end{array}$$

so $g \in \partial f(x)$. conversely, if $g \in \partial f(x)$,

$$\begin{array}{rcl} f(y) &\geq& g^T(y-x) + f(x) \\ g^T x - f(x) &\geq& g^T y - f(y) \\ \sup_y g^T x - f(x) &\geq& \sup_y g^T y - f(y) \\ g^T x - f(x) &\geq& f^*(g) \end{array}$$

so $f^*(g) + f(x) = g^T x$.

Subgradients and fenchel conjugates

Conclusion.

$$\begin{aligned} g \in \partial f(x) \quad &\iff \quad f^*(g) + f(x) = g^T x \\ &\iff \quad x \in \operatorname*{argmax}_x g^T x - f(x) \end{aligned}$$

consider the same implications for the function f^*:

$$\begin{aligned} x \in \partial f^*(g) \quad &\iff \quad f(x) + f^*(g) = x^T g \\ &\iff \quad g \in \operatorname*{argmax}_g g^T x - f^*(g) \end{aligned}$$

so all these conditions are equivalent, and
$g \in \partial f(x) \iff x \in \partial f^*(g)$!

Compute subgradient via fenchel conjugate

$$\partial f(x) = \operatorname*{argmax}_g g^T x - f^*(g)$$

example. let $f(x) = \|x\|_1$. compute

$$\begin{aligned} f^*(g) &= \sup_x g^T x - \|x\|_1 \\ \\ &= \end{aligned}$$

Compute subgradient via fenchel conjugate

$$\partial f(x) = \operatorname*{argmax}_g g^T x - f^*(g)$$

example. let $f(x) = \|x\|_1$. compute

$$\begin{aligned} f^*(g) &= \sup_x g^T x - \|x\|_1 \\ &= \begin{cases} 0 & \|g\|_\infty \leq 1 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

Compute subgradient via fenchel conjugate

$$\partial f(x) = \operatorname*{argmax}_g g^T x - f^*(g)$$

example. let $f(x) = \|x\|_1$. compute

$$\begin{aligned} f^*(g) &= \sup_x g^T x - \|x\|_1 \\ &= \begin{cases} 0 & \|g\|_\infty \le 1 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

given x,

$$\begin{aligned} \partial f(x) &= \operatorname*{argmax}_g g^T x - f^*(g) \\ &= \operatorname*{argmax}_{\|g\|_\infty \le 1} g^T x \\ &= \mathbf{sign}(x) \end{aligned}$$

where **sign** is computed elementwise.

Compute subgradient via fenchel conjugate

$$\partial f(x) = \operatorname*{argmax}_{g} g^T x - f^*(g)$$

example. let $f(X) = \|X\|_*$. compute

$$\begin{aligned} f^*(G) &= \sup_X \mathbf{tr}(G, X) - \|X\|_* \\ \\ &= \end{aligned}$$

Compute subgradient via fenchel conjugate

$$\partial f(x) = \operatorname*{argmax}_g g^T x - f^*(g)$$

example. let $f(X) = \|X\|_*$. compute

$$\begin{aligned} f^*(G) &= \sup_X \mathbf{tr}(G, X) - \|X\|_* \\ &= \begin{cases} 0 & \|G\| \leq 1 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

where $\|G\| = \sigma_1(G)$ is the operator norm of G.

Compute subgradient via fenchel conjugate

$$\partial f(x) = \operatorname*{argmax}_g g^T x - f^*(g)$$

example. let $f(X) = \|X\|_*$. compute

$$\begin{aligned} f^*(G) &= \sup_X \mathbf{tr}(G, X) - \|X\|_* \\ &= \begin{cases} 0 & \|G\| \le 1 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

where $\|G\| = \sigma_1(G)$ is the operator norm of G.

given $X = U \,\mathbf{diag}(\sigma) V^T$,

$$\begin{aligned} \partial f(x) &= \operatorname*{argmax}_G \mathbf{tr}(G, X) - f^*(G) \\ &= \operatorname*{argmax}_{\|G\| \le 1} \mathbf{tr}(G, X) \\ &= U \,\mathbf{diag}(\mathbf{sign}(\sigma)) V^T \end{aligned}$$

where **sign** is computed elementwise.

Outline

Subgradients

Properties

Subgradient calculus

Optimality

Subgradient calculus

- **weak subgradient calculus**: formulas for finding **one** subgradient $g \in \partial f(x)$
- **strong subgradient calculus**: formulas for finding the whole subdifferential $\partial f(x)$, *i.e.*, **all** subgradients of f at x
- many algorithms for nondifferentiable convex optimization require only **one** subgradient at each step, so weak calculus suffices
- some algorithms, optimality conditions, etc., need whole subdifferential
- roughly speaking: if you can compute $f(x)$, you can usually compute a $g \in \partial f(x)$
- we'll assume that f is convex, and $x \in \mathbf{relint\,dom}\, f$

Some basic rules

- $\partial f(x) = \{\nabla f(x)\}$ if f is differentiable at x
- **scaling:** $\partial(\alpha f) = \alpha \partial f$ (if $\alpha > 0$)
- **addition:** $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$ (RHS is addition of point-to-set mappings)
- **affine transformation of variables:** if $g(x) = f(Ax + b)$, then $\partial g(x) = A^T \partial f(Ax + b)$
- **finite pointwise maximum:** if $f = \max_{i=1,\ldots,m} f_i$, then

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

  i.e., convex hull of union of subdifferentials of 'active' functions at x

Minimization

define $g(y)$ as the optimal value of

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq y_i, \quad i = 1, \dots, m \end{array}$$

(f_i convex; variable x)

with $\lambda^\star$ an optimal dual variable, we have

$$g(z) \geq g(y) - \sum_{i=1}^{m} \lambda_i^\star (z_i - y_i)$$

i.e., $-\lambda^\star$ is a subgradient of g at y

Composition

- $f(x) = h(f_1(x), \ldots, f_k(x))$, with h convex nondecreasing, f_i convex
- find $q \in \partial h(f_1(x), \ldots, f_k(x))$, $g_i \in \partial f_i(x)$
- then, $g = q_1 g_1 + \cdots + q_k g_k \in \partial f(x)$
- reduces to standard formula for differentiable h, f_i

proof:

$$\begin{aligned} f(y) &= h(f_1(y), \ldots, f_k(y)) \\ &\geq h(f_1(x) + g_1^T(y-x), \ldots, f_k(x) + g_k^T(y-x)) \\ &\geq h(f_1(x), \ldots, f_k(x)) + q^T(g_1^T(y-x), \ldots, g_k^T(y-x)) \\ &= f(x) + g^T(y-x) \end{aligned}$$

Outline

Subgradients

Properties

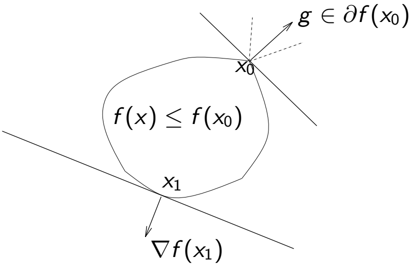
Subgradient calculus

Optimality

Subgradients and sublevel sets

g is a subgradient at x means $f(y) \geq f(x) + g^T(y - x)$

hence $f(y) \leq f(x) \Longrightarrow g^T(y - x) \leq 0$



- f differentiable at x_0: $\nabla f(x_0)$ is normal to the sublevel set $\{x \mid f(x) \leq f(x_0)\}$
- f nondifferentiable at x_0: subgradient defines a supporting hyperplane to sublevel set through x_0

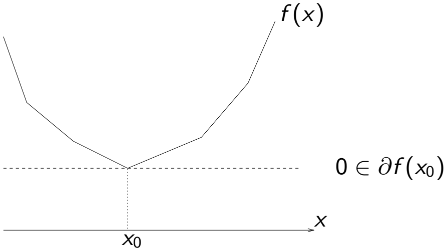
Optimality conditions — unconstrained

recall for f convex, differentiable,

$$f(x^\star) = \inf_x f(x) \iff 0 = \nabla f(x^\star)$$

generalization to nondifferentiable convex f:

$$f(x^\star) = \inf_x f(x) \iff 0 \in \partial f(x^\star)$$



proof.

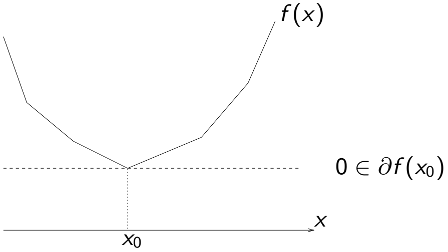
Optimality conditions — unconstrained

recall for f convex, differentiable,

$$f(x^\star) = \inf_x f(x) \iff 0 = \nabla f(x^\star)$$

generalization to nondifferentiable convex f:

$$f(x^\star) = \inf_x f(x) \iff 0 \in \partial f(x^\star)$$



proof. by definition (!)

$$f(y) \geq f(x^\star) + 0^T(y - x^\star) \text{ for all } y \quad \iff \quad 0 \in \partial f(x^\star)$$

...seems trivial but isn't

Example: piecewise linear minimization

$f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$

$x^\star$ minimizes $f \iff 0 \in \partial f(x^\star) = \mathbf{conv}\{a_i \mid a_i^T x^\star + b_i = f(x^\star)\}$

$\iff$ there is a λ with

$$\lambda \succeq 0, \qquad \mathbf{1}^T\lambda = 1, \qquad \sum_{i=1}^m \lambda_i a_i = 0$$

where $\lambda_i = 0$ if $a_i^T x^\star + b_i < f(x^\star)$

. . . but these are the KKT conditions for the epigraph form

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{array}$$

with dual

$$\begin{array}{ll} \text{maximize} & b^T \lambda \\ \text{subject to} & \lambda \succeq 0, \qquad A^T \lambda = 0, \qquad \mathbf{1}^T \lambda = 1 \end{array}$$

Optimality conditions — constrained

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \;\; i = 1, \ldots, m \end{array}$$

we assume

- f_i convex, defined on $\mathbf{R}^n$ (hence subdifferentiable)
- strict feasibility (Slater's condition)

$x^\star$ is primal optimal ($\lambda^\star$ is dual optimal) iff

$$\begin{array}{l} f_i(x^\star) \leq 0, \;\; \lambda_i^\star \geq 0 \\ 0 \in \partial f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \partial f_i(x^\star) \\ \lambda_i^\star f_i(x^\star) = 0 \end{array}$$

... generalizes KKT for nondifferentiable f_i

Directional derivative

directional derivative of f at x in the direction δx is

$$f'(x; \delta x) \stackrel{\Delta}{=} \lim_{h \searrow 0} \frac{f(x + h\delta x) - f(x)}{h}$$

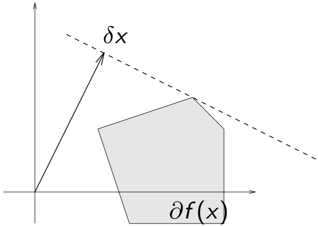
can be $+\infty$ or $-\infty$

- f convex, finite near $x \Longrightarrow f'(x; \delta x)$ exists

- f differentiable at x if and only if, for some g $(= \nabla f(x))$ and all δx, $f'(x; \delta x) = g^T \delta x$ (*i.e.*, $f'(x; \delta x)$ is a linear function of δx)

Directional derivative and subdifferential

general formula for convex f: $f'(x;\delta x) = \sup_{g \in \partial f(x)} g^T \delta x$



Descent directions

δx is a **descent direction** for f at x if $f'(x; \delta x) < 0$

for differentiable f, $\delta x = -\nabla f(x)$ is always a descent direction (except when it is zero)

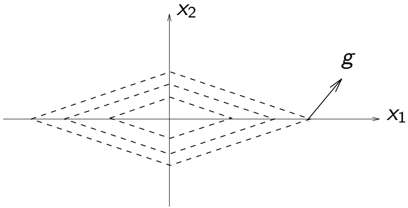
Descent directions

δx is a **descent direction** for f at x if $f'(x; \delta x) < 0$

for differentiable f, $\delta x = -\nabla f(x)$ is always a descent direction (except when it is zero)

warning: for nondifferentiable (convex) functions, $\delta x = -g$, with $g \in \partial f(x)$, need not be descent direction

example: $f(x) = |x_1| + 2|x_2|$



Subgradients and distance to sublevel sets

if f is convex, $f(z) < f(x)$, $g \in \partial f(x)$, then for small $t > 0$,

$$\|x - tg - z\|_2 < \|x - z\|_2$$

thus $-g$ is descent direction for $\|x - z\|_2$, for **any** z with $f(z) < f(x)$ (*e.g.*, $x^\star$)

negative subgradient is descent direction for distance to optimal point

$$\text{proof:} \quad \begin{aligned} \|x - tg - z\|_2^2 &= \|x - z\|_2^2 - 2tg^T(x - z) + t^2\|g\|_2^2 \\ &\leq \|x - z\|_2^2 - 2t(f(x) - f(z)) + t^2\|g\|_2^2 \end{aligned}$$

Descent directions and optimality

fact: for f convex, finite near x, either

- $0 \in \partial f(x)$ (in which case x minimizes f), or
- there is a descent direction for f at x

i.e., x is optimal (minimizes f) iff there is no descent direction for f at x

proof: define $\delta x_{\mathsf{sd}} = -\operatorname*{argmin}_{z \in \partial f(x)} \|z\|_2$

if $\delta x_{\mathrm{sd}} = 0$, then $0 \in \partial f(x)$, so x is optimal; otherwise $f'(x; \delta x_{\mathsf{sd}}) = -\left(\inf_{z \in \partial f(x)} \|z\|_2\right)^2 < 0$, so δx_{sd} is a descent direction