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 $\text{epi } f$ at $(x, f(x))$

what if $f$ is not differentiable?
Non-differentiable functions

are these functions differentiable?

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

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 $\text{epi} f$ 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

\[ f = \max\{f_1, f_2\}, \text{ with } f_1, f_2 \text{ convex and differentiable} \]

![Diagram showing the function f(x) and its subgradients at x_0]

- \( f_1(x_0) > f_2(x_0) \): 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
set of all subgradients of \( f \) at \( x \) is called the \textbf{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 \( x \in \text{relint \ dom \ } f \)
- \( \partial f(x) = \{ \nabla f(x) \} \), 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

\[ g \in \partial f(x) \text{ iff } \]

\[ f(y) \geq f(x) + g^T(y - x) \quad \forall y \in \text{dom}(f) \]

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

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

then

\[ f(y) = \max(y, -y) \geq sy = s(x + y - x) = |x| + s(y - x) \]
Compute subgradient via definition

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

**example.** Let \( f(x) = |x| \) for \( x \in \mathbb{R} \). Suppose \( s \in \text{sign}(x) \), where

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

Then

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

So \( \text{sign}(x) \subseteq \partial f(x) \) (in fact, holds with equality)
Subgradient of $|x|$

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

The righthand plot shows $\bigcup \{(x, g) \mid x \in \mathbb{R}, g \in \partial f(x)\}$
Compute subgradient via definition

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

example. let \( f(x) = \max_i a_i^T x + b_i \).
Compute subgradient via definition

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

**example.** Let \( f(x) = \max_i a_i^T x + b_i \). Then for any \( i \),

\[
\begin{align*}
f(y) &= \max_i a_i^T y + b_i \\
&\geq a_i^T y + b_i \\
&= a_i^T (x + y - x) + b_i \\
&= a_i^T x + b_i + a_i^T (y - x) \\
&= f(x) + a_i^T (y - x),
\end{align*}
\]

where the last line holds for \( i \in \arg\max_j a_j^T x + b_j \). So

- \( a_i \in \partial f(x) \) for each \( i \in \arg\max_j a_j^T x + b_j \)
- \( \partial f(x) \) is convex, so

\[
\text{conv}\{a_i : i \in \arg\max_j a_j^T x + b_j\} \subseteq \partial f(x)
\]
Compute subgradient via definition

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

**example.** let \( f(X) = \lambda_{\text{max}}(X) \).
Compute subgradient via definition

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

**Example.** Let \( f(X) = \lambda_{\text{max}}(X) \). Then

\[
 f(Y) = \sup \|v\| \leq 1 v^T Y v \\
 = \sup \|v\| \leq 1 v^T (X + Y - X) v, \quad \|v\| \leq 1 \\
= \sup \|v\| \leq 1 (v^T X v + v^T (Y - X) v), \quad \|v\| \leq 1 \\
= v^T X v + \text{tr}(vv^T (Y - X)), \quad v \in \arg\max_{\|v\| \leq 1} v^T X v \\
= \lambda_{\text{max}}(X) + \text{tr}(vv^T (Y - X)), \quad v \in \arg\max_{\|v\| \leq 1} v^T X v \\
\]

So

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

\[
 \text{conv}\{vv^T : v \in \arg\max_{\|v\| \leq 1} v^T X v\} \subseteq \partial f(x) 
\]
Outline

Subgradients

Properties

Subgradient calculus

Optimality
Properties of subgradients

subgradient inequality:

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

for convex \( f \), we’ll show

- subgradients are monotone: for any \( x, y \in \text{dom} f \), \( g_y \in \partial f(y) \), 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 \( g \in \partial f(x) \)

- \( \partial f(x) = \arg\max g^T x - f(x) \)

these will help us compute subgradients
Subgradients are monotone

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

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

**proof.** same as for differentiable case:

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

add these to get

$$(g_y - g_x)^T(y - x) \geq 0$$
Subgradients are preserved under limits

Subgradient inequality:

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

**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:

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

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 \( y \),

\[
\begin{align*}
  f(y) & \geq f(x^{(k)}) + (g^{(k)})^T(y - x^{(k)}) \\
  \lim_{k \to \infty} f(y) & \geq \lim_{k \to \infty} f(x^{(k)}) + (g^{(k)})^T(y - x^{(k)}) \\
  f(y) & \geq f(x) + g^T(y - x)
\end{align*}
\]

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

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

so

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

ehence
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 $x \in \text{dom } f$ 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 \( x \in \text{dom} f \) uniformly at random, then with probability 1, \( f \) is differentiable at \( x \). **intuition.** (in \( \mathbb{R} \)) Subgradients are closed convex sets, so in \( \mathbb{R} \) 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 \( x \in \text{dom} \, f \) uniformly at random, then with probability 1, \( f \) is differentiable at \( x \). **Intuition.** (in \( \mathbb{R} \).) Subgradients are closed convex sets, so in \( \mathbb{R} \) 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 \( x^{(k)} \rightarrow x \) 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 \( f^*(g) + f(x) = g^T x, \)

\[
f^*(g) = \sup_y g^T y - f(y)
\]

\[
\geq g^T y - f(y) \quad \forall y
\]

\[
f(y) \geq g^T y - f^*(g) \quad \forall y
\]

\[
= g^T y - g^T x + f(x) \quad \forall y
\]

\[
= g^T (y - x) + f(x) \quad \forall y
\]

so \( g \in \partial f(x). \) Conversely, if \( g \in \partial f(x), \)
Subgradients and fenchel conjugates

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

\[
f^*(g) = \sup_y g^T y - f(y)
\]

\[
\geq g^T y - f(y) \quad \forall y
\]

\[
f(y) \geq g^T y - f^*(g) \quad \forall y
\]

\[
= g^T y - g^T x + f(x) \quad \forall y
\]

\[
= g^T (y - x) + f(x) \quad \forall y
\]

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

\[
f(y) \geq g^T (y - x) + f(x)
\]

\[
g^T x - f(x) \geq g^T y - f(y)
\]

\[
\sup_y g^T x - f(x) \geq \sup_y g^T y - f(y)
\]

\[
g^T x - f(x) \geq f^*(g)
\]

so \( f^*(g) + f(x) = g^T x \).
Subgradients and fenchel conjugates

Conclusion.

\[ g \in \partial f(x) \iff f^*(g) + f(x) = g^T x \]
\[ \iff x \in \arg\max_x g^T x - f(x) \]

consider the same implications for the function \( f^* \):

\[ x \in \partial f^*(g) \iff f(x) + f^*(g) = x^T g \]
\[ \iff g \in \arg\max_g g^T x - f^*(g) \]

so all these conditions are equivalent, and

\[ g \in \partial f(x) \iff x \in \partial f^*(g)! \]
Compute subgradient via fenchel conjugate

\[ \partial f(x) = \arg\max_g g^T x - f^*(g) \]

**Example.** Let \( f(x) = \|x\|_1 \). Compute

\[ f^*(g) = \sup_x g^T x - \|x\|_1 \]

\[ = \]
Compute subgradient via fenchel conjugate

\[ \partial f(x) = \arg \max_g g^T x - f^*(g) \]

**example.** Let \( f(x) = \|x\|_1 \). Compute

\[ f^*(g) = \sup_x g^T x - \|x\|_1 \]

\[ = \begin{cases} 
0 & \|g\|_\infty \leq 1 \\
\infty & \text{otherwise} 
\end{cases} \]
Compute subgradient via fenchel conjugate

\[ \partial f(x) = \arg\max_{g} g^T x - f^*(g) \]

**example.** Let \( f(x) = \|x\|_1 \). Compute

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

Given \( x \),

\[
\partial f(x) = \arg\max_{g} g^T x - f^*(g)
= \arg\max_{\|g\|_\infty \leq 1} g^T x
= \text{sign}(x)
\]

where \( \text{sign} \) is computed elementwise.
Compute subgradient via fenchel conjugate

\[ \partial f(x) = \arg\max_g g^T x - f^*(g) \]

**example.** Let \( f(X) = \|X\|_* \). Compute

\[ f^*(G) = \sup_X \text{tr}(G, X) - \|X\|_* \]

\[ = \]

\[ = \]
Compute subgradient via fenchel conjugate

\[ \partial f(x) = \arg\max_g g^T x - f^*(g) \]

**example.** Let \( f(X) = \|X\|_* \). Compute

\[ f^*(G) = \sup_X \text{tr}(G, X) - \|X\|_* \]

\[ = \begin{cases} 
0 & \|G\| \leq 1 \\
\infty & \text{otherwise}
\end{cases} \]

where \( \|G\| = \sigma_1(G) \) is the operator norm of \( G \).
Compute subgradient via fenchel conjugate

\[ \partial f(x) = \arg\max_g g^T x - f^*(g) \]

**Example.** Let \( f(X) = \|X\|_* \). Compute

\[ f^*(G) = \sup_X \text{tr}(G, X) - \|X\|_* \]

\[ = \begin{cases} 
0 & \|G\| \leq 1 \\
\infty & \text{otherwise}
\end{cases} \]

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

Given \( X = U \text{diag}(\sigma)V^T \),

\[ \partial f(x) = \arg\max_G \text{tr}(G, X) - f^*(G) \]

\[ = \arg\max_G \text{tr}(G, X) \]

\[ = U \text{diag}(\text{sign}(\sigma))V^T \]

where \text{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 \( x \in \text{relint dom } f \)
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 \( f = \max_{i=1,...,m} f_i \), then
  \[
  \partial f(x) = \text{conv} \bigcup \{ \partial f_i(x) \mid f_i(x) = f(x) \},
  \]
i.e., convex hull of union of subdifferentials of ‘active’ functions at \( x \)
Minimization

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

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

($f_i$ convex; variable $x$)

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

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

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

\[ f(x) = h(f_1(x), \ldots, f_k(x)), \text{ with } h \text{ convex nondecreasing, } f_i \text{ convex} \]

\[ \text{find } q \in \partial h(f_1(x), \ldots, f_k(x)), \quad g_i \in \partial f_i(x) \]

\[ \text{then, } g = q_1 g_1 + \cdots + q_k g_k \in \partial f(x) \]

\[ \text{reduces to standard formula for differentiable } h, f_i \]

proof:

\[
\begin{align*}
  f(y) & = h(f_1(y), \ldots, f_k(y)) \\
  \geq & \quad h(f_1(x) + g_1^T (y - x), \ldots, f_k(x) + g_k^T (y - x)) \\
  \geq & \quad h(f_1(x), \ldots, f_k(x)) + q^T (g_1^T (y - x), \ldots, g_k^T (y - x)) \\
  = & \quad f(x) + g^T (y - x)
\end{align*}
\]
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 \( f(y) \leq f(x) \iff g^T(y - x) \leq 0 \)

\( g \in \partial f(x_0) \)

\( f(x) \leq f(x_0) \)

\( \nabla f(x_1) \)

- \( f \) 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,

\[ f(x^*) = \inf_x f(x) \iff 0 = \nabla f(x^*) \]

generalization to nondifferentiable convex $f$:

\[ f(x^*) = \inf_x f(x) \iff 0 \in \partial f(x^*) \]

proof.
Optimality conditions — unconstrained

recall for $f$ convex, differentiable,

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

generalization to nondifferentiable convex $f$:

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

proof. by definition (!)

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

... seems trivial but isn’t
Example: piecewise linear minimization

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

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

\( \iff \) there is a \( \lambda \) with

\[ \lambda \succeq 0, \quad 1^T \lambda = 1, \quad \sum_{i=1}^{m} \lambda_i a_i = 0 \]

where \( \lambda_i = 0 \) if \( a_i^T x^* + b_i < f(x^*) \)
...but these are the KKT conditions for the epigraph form

\[
\begin{align*}
& \text{minimize} \quad t \\
& \text{subject to} \quad a_i^T x + b_i \leq t, \quad i = 1, \ldots, m
\end{align*}
\]

with dual

\[
\begin{align*}
& \text{maximize} \quad b^T \lambda \\
& \text{subject to} \quad \lambda \succeq 0, \quad A^T \lambda = 0, \quad 1^T \lambda = 1
\end{align*}
\]
minimize \( f_0(x) \)
subject to \( f_i(x) \leq 0, \ i = 1, \ldots, m \)

we assume

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

\( x^* \) is primal optimal (\( \lambda^* \) is dual optimal) iff

\[
\begin{align*}
    f_i(x^*) & \leq 0, \ \lambda_i^* \geq 0 \\
    0 & \in \partial f_0(x^*) + \sum_{i=1}^m \lambda_i^* \partial f_i(x^*) \\
    \lambda_i^* f_i(x^*) & = 0
\end{align*}
\]

...generalizes KKT for nondifferentiable \( f_i \)
**Directional derivative**

**directional derivative** of $f$ at $x$ in the direction $\delta x$ is

$$f'(x; \delta x) \triangleq \lim_{h \to 0} \frac{f(x + h\delta x) - f(x)}{h}$$

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

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

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

\( \delta x \) 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

$\delta x$ 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: $f(x) = |x_1| + 2|x_2|$
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., $x^*$)

negative subgradient is descent direction for distance to optimal point

proof:  

$$
\|x - tg - z\|_2^2 = \|x - z\|_2^2 - 2tg^T(x - z) + t^2\|g\|_2^2 \\
\leq \|x - z\|_2^2 - 2t(f(x) - f(z)) + t^2\|g\|_2^2
$$
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 $\delta x_{sd} = - \arg\min_{z \in \partial f(x)} \|z\|_2$

if $\delta x_{sd} = 0$, then $0 \in \partial f(x)$, so $x$ is optimal; otherwise

$f'(x; \delta x_{sd}) = - (\inf_{z \in \partial f(x)} \|z\|_2)^2 < 0$, so $\delta x_{sd}$ is a descent direction