

**Recap of Conjugate sets, conjugate functions.
Subgradient and subdifferential**

Daniil Merkulov

Optimization methods. MIPT

Conjugate sets

Conjugate set

Let $S \subseteq \mathbb{R}^n$ be an arbitrary non-empty set. Then its conjugate set is defined as:

$$S^* = \{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq -1 \quad \forall x \in S\}$$

A set S^{**} is called double conjugate to a set S if:

$$S^{**} = \{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq -1 \quad \forall x \in S^*\}$$

- The sets S_1 and S_2 are called **inter-conjugate** if $S_1^* = S_2$, $S_2^* = S_1$.



Figure 1: Convex sets may be described in a dual way - through the elements of the set and through the set of hyperplanes supporting it

Conjugate set

Let $S \subseteq \mathbb{R}^n$ be an arbitrary non-empty set. Then its conjugate set is defined as:

$$S^* = \{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq -1 \quad \forall x \in S\}$$

A set S^{**} is called double conjugate to a set S if:

$$S^{**} = \{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq -1 \quad \forall x \in S^*\}$$

- The sets S_1 and S_2 are called **inter-conjugate** if $S_1^* = S_2$, $S_2^* = S_1$.
- A set S is called **self-conjugate** if $S^* = S$.



Figure 1: Convex sets may be described in a dual way - through the elements of the set and through the set of hyperplanes supporting it

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\text{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.
- $\left(\bigcup_{i=1}^m S_i\right)^* = \bigcap_{i=1}^m S_i^*$.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.
- $\left(\bigcup_{i=1}^m S_i\right)^* = \bigcap_{i=1}^m S_i^*$.
- If S is closed, convex, and includes 0, then $S^{**} = S$.

Properties of conjugate sets

- A conjugate set is always closed, convex, and contains zero.
- For an arbitrary set $S \subseteq \mathbb{R}^n$:

$$S^{**} = \overline{\mathbf{conv}(S \cup \{0\})}$$

- If $S_1 \subseteq S_2$, then $S_2^* \subseteq S_1^*$.
- $\left(\bigcup_{i=1}^m S_i\right)^* = \bigcap_{i=1}^m S_i^*$.
- If S is closed, convex, and includes 0, then $S^{**} = S$.
- $S^* = (\overline{S})^*$.

Example 1

Example

Prove that $S^* = (\overline{S})^*$.

Example 1

Example

Prove that $S^* = (\overline{S})^*$.

- $S \subset \overline{S} \rightarrow (\overline{S})^* \subset S^*$.

Example 1

i Example

Prove that $S^* = (\overline{S})^*$.

- $S \subset \overline{S} \rightarrow (\overline{S})^* \subset S^*$.
- Let $p \in S^*$ and $x_0 \in \overline{S}$, $x_0 = \lim_{k \rightarrow \infty} x_k$. Then by virtue of the continuity of the function $f(x) = p^T x$, we have:
 $p^T x_k \geq -1 \rightarrow p^T x_0 \geq -1$. So $p \in (\overline{S})^*$, hence $S^* \subset (\overline{S})^*$.

Example 2

Example

Prove that $(\text{conv}(S))^* = S^*$.

Example 2

Example

Prove that $(\text{conv}(S))^* = S^*$.

- $S \subset \text{conv}(S) \rightarrow (\text{conv}(S))^* \subset S^*$.

Example 2

i Example

Prove that $(\mathbf{conv}(S))^* = S^*$.

- $S \subset \mathbf{conv}(S) \rightarrow (\mathbf{conv}(S))^* \subset S^*$.
- Let $p \in S^*$, $x_0 \in \mathbf{conv}(S)$, i.e., $x_0 = \sum_{i=1}^k \theta_i x_i \mid x_i \in S, \sum_{i=1}^k \theta_i = 1, \theta_i \geq 0$.

So $p^T x_0 = \sum_{i=1}^k \theta_i p^T x_i \geq \sum_{i=1}^k \theta_i (-1) = 1 \cdot (-1) = -1$. So $p \in (\mathbf{conv}(S))^*$, hence $S^* \subset (\mathbf{conv}(S))^*$.

Example 3

Example

Prove that if $B(0, r)$ is a ball of radius r by some norm centered at zero, then $(B(0, r))^* = B(0, 1/r)$.

Example 3

Example

Prove that if $B(0, r)$ is a ball of radius r by some norm centered at zero, then $(B(0, r))^* = B(0, 1/r)$.

- Let $B(0, r) = X$, $B(0, 1/r) = Y$. Take the normal vector $p \in X^*$, then for any $x \in X : p^T x \geq -1$.

Example 3

i Example

Prove that if $B(0, r)$ is a ball of radius r by some norm centered at zero, then $(B(0, r))^* = B(0, 1/r)$.

- Let $B(0, r) = X$, $B(0, 1/r) = Y$. Take the normal vector $p \in X^*$, then for any $x \in X$: $p^T x \geq -1$.
- From all points of the ball X , take such a point $x \in X$ that its scalar product of p : $p^T x$ is minimal, then this is the point $x = -\frac{p}{\|p\|}r$.

$$p^T x = p^T \left(-\frac{p}{\|p\|}r \right) = -\|p\|r \geq -1$$

$$\|p\| \leq \frac{1}{r} \in Y$$

So $X^* \subset Y$.

Example 3

i Example

Prove that if $B(0, r)$ is a ball of radius r by some norm centered at zero, then $(B(0, r))^* = B(0, 1/r)$.

- Let $B(0, r) = X$, $B(0, 1/r) = Y$. Take the normal vector $p \in X^*$, then for any $x \in X$: $p^T x \geq -1$.
- From all points of the ball X , take such a point $x \in X$ that its scalar product of p : $p^T x$ is minimal, then this is the point $x = -\frac{p}{\|p\|}r$.
- Now let $p \in Y$. We need to show that $p \in X^*$, i.e., $\langle p, x \rangle \geq -1$. It's enough to apply the Cauchy-Bunyakovsky inequality:

$$\|\langle p, x \rangle\| \leq \|p\| \|x\| \leq \frac{1}{r} \cdot r = 1$$

The latter comes from the fact that $p \in B(0, 1/r)$ and $x \in B(0, r)$.

So $Y \subset X^*$.

$$p^T x = p^T \left(-\frac{p}{\|p\|} r \right) = -\|p\| r \geq -1$$

$$\|p\| \leq \frac{1}{r} \in Y$$

So $X^* \subset Y$.

Dual cone

A conjugate cone to a cone K is a set K^* such that:

$$K^* = \{y \mid \langle x, y \rangle \geq 0 \quad \forall x \in K\}$$

To show that this definition follows directly from the definitions above, recall what a conjugate set is and what a cone $\forall \lambda > 0$ is.

$$\{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq -1 \quad \forall x \in S\} \rightarrow \{\lambda y \in \mathbb{R}^n \mid \langle y, x \rangle \geq -\frac{1}{\lambda} \quad \forall x \in S\}$$



Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.

Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.
- For an arbitrary set $S \subseteq \mathbb{R}^n$ and a cone $K \subseteq \mathbb{R}^n$:

$$(S + K)^* = S^* \cap K^*$$

Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.
- For an arbitrary set $S \subseteq \mathbb{R}^n$ and a cone $K \subseteq \mathbb{R}^n$:

$$(S + K)^* = S^* \cap K^*$$

- Let K_1, \dots, K_m be cones in \mathbb{R}^n , then:

$$\left(\sum_{i=1}^m K_i \right)^* = \bigcap_{i=1}^m K_i^*$$

Dual cones properties

- Let K be a closed convex cone. Then $K^{**} = K$.
- For an arbitrary set $S \subseteq \mathbb{R}^n$ and a cone $K \subseteq \mathbb{R}^n$:

$$(S + K)^* = S^* \cap K^*$$

- Let K_1, \dots, K_m be cones in \mathbb{R}^n , then:

$$\left(\sum_{i=1}^m K_i \right)^* = \bigcap_{i=1}^m K_i^*$$

- Let K_1, \dots, K_m be cones in \mathbb{R}^n . Let also their intersection have an interior point, then:

$$\left(\bigcap_{i=1}^m K_i \right)^* = \sum_{i=1}^m K_i^*$$

Example

i Example

Find the conjugate cone for a monotone nonnegative cone:

$$K = \{x \in \mathbb{R}^n \mid x_1 \geq x_2 \geq \dots \geq x_n \geq 0\}$$

Example

Example

Find the conjugate cone for a monotone nonnegative cone:

$$K = \{x \in \mathbb{R}^n \mid x_1 \geq x_2 \geq \dots \geq x_n \geq 0\}$$

Note that:

$$\sum_{i=1}^n x_i y_i = y_1(x_1 - x_2) + (y_1 + y_2)(x_2 - x_3) + \dots + (y_1 + y_2 + \dots + y_{n-1})(x_{n-1} - x_n) + (y_1 + \dots + y_n)x_n$$

Since in the presented sum in each summand, the second multiplier in each summand is non-negative, then:

$$y_1 \geq 0, \quad y_1 + y_2 \geq 0, \quad \dots, \quad y_1 + \dots + y_n \geq 0$$

$$\text{So } K^* = \left\{ y \mid \sum_{i=1}^k y_i \geq 0, k = \overline{1, n} \right\}.$$

Polyhedra

The set of solutions to a system of linear inequalities and equalities is a polyhedron:

$$Ax \preceq b, \quad Cx = d$$

Here $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{p \times n}$, and the inequality is a piecewise inequality.

Theorem

Let $x_1, \dots, x_m \in \mathbb{R}^n$. Conjugate to a polyhedral set:

$$S = \mathbf{conv}(x_1, \dots, x_k) + \mathbf{cone}(x_{k+1}, \dots, x_m)$$

is a polyhedron (polyhedron):

$$S^* = \{p \in \mathbb{R}^n \mid \langle p, x_i \rangle \geq -1, i = \overline{1, k}; \langle p, x_i \rangle \geq 0, i = \overline{k+1, m}\}$$

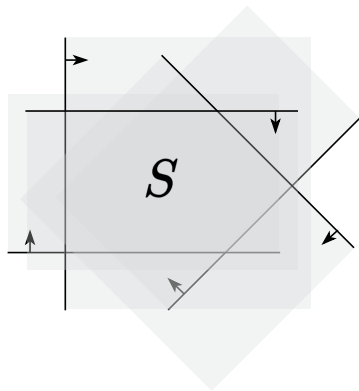


Figure 3: Polyhedra

Proof

- Let $S = X, S^* = Y$. Take some $p \in X^*$, then $\langle p, x_i \rangle \geq -1, i = \overline{1, k}$. At the same time, for any $\theta > 0, i = \overline{k+1, m}$:

$$\langle p, x_i \rangle \geq -1 \rightarrow \langle p, \theta x_i \rangle \geq -1$$

$$\langle p, x_i \rangle \geq -\frac{1}{\theta} \rightarrow \langle p, x_i \rangle \geq 0.$$

So $p \in Y \rightarrow X^* \subset Y$.

Proof

- Let $S = X, S^* = Y$. Take some $p \in X^*$, then $\langle p, x_i \rangle \geq -1, i = \overline{1, k}$. At the same time, for any $\theta > 0, i = \overline{k+1, m}$:

$$\langle p, x_i \rangle \geq -1 \rightarrow \langle p, \theta x_i \rangle \geq -1$$

$$\langle p, x_i \rangle \geq -\frac{1}{\theta} \rightarrow \langle p, x_i \rangle \geq 0.$$

So $p \in Y \rightarrow X^* \subset Y$.

- Suppose, on the other hand, that $p \in Y$. For any point $x \in X$:

$$x = \sum_{i=1}^m \theta_i x_i \quad \sum_{i=1}^k \theta_i = 1, \theta_i \geq 0$$

So:

$$\langle p, x \rangle = \sum_{i=1}^m \theta_i \langle p, x_i \rangle = \sum_{i=1}^k \theta_i \langle p, x_i \rangle + \sum_{i=k+1}^m \theta_i \langle p, x_i \rangle \geq \sum_{i=1}^k \theta_i (-1) + \sum_{i=1}^k \theta_i \cdot 0 = -1.$$

Example

Conjugate functions

Conjugate functions



Recall that given $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the function defined by

$$f^*(y) = \max_x [y^T x - f(x)]$$

is called its conjugate.

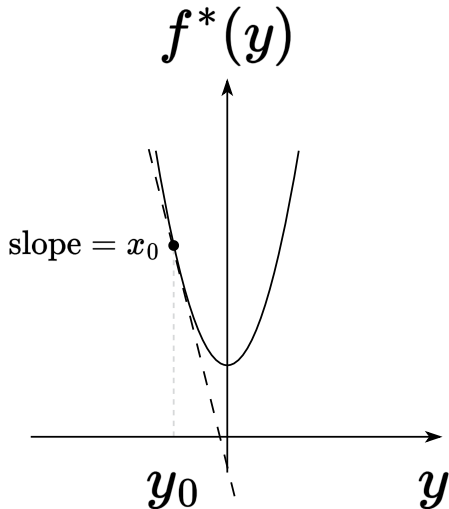
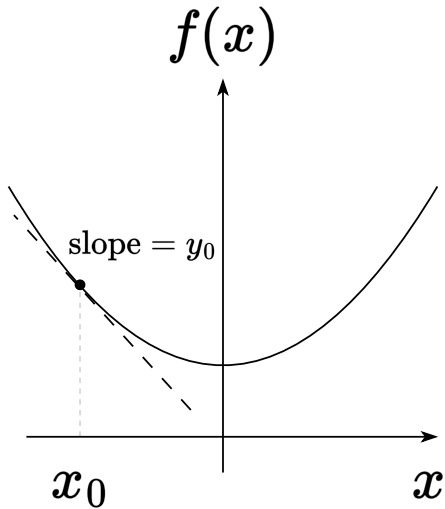
Geometrical intuition



Geometrical intuition



Slopes of f and f^*



Slopes of f and f^*

Assume that f is a closed and convex function. Then f is strongly convex with parameter $\mu \Leftrightarrow \nabla f^*$ is Lipschitz with parameter $1/\mu$.

Slopes of f and f^*

Assume that f is a closed and convex function. Then f is strongly convex with parameter $\mu \Leftrightarrow \nabla f^*$ is Lipschitz with parameter $1/\mu$.

Proof of “ \Rightarrow ”: Recall, if g is strongly convex with minimizer x , then

$$g(y) \geq g(x) + \frac{\mu}{2} \|y - x\|^2, \quad \text{for all } y$$

Slopes of f and f^*

Assume that f is a closed and convex function. Then f is strongly convex with parameter $\mu \Leftrightarrow \nabla f^*$ is Lipschitz with parameter $1/\mu$.

Proof of “ \Rightarrow ”: Recall, if g is strongly convex with minimizer x , then

$$g(y) \geq g(x) + \frac{\mu}{2} \|y - x\|^2, \quad \text{for all } y$$

Hence, defining $x_u = \nabla f^*(u)$ and $x_v = \nabla f^*(v)$,

$$f(x_v) - u^T x_v \geq f(x_u) - u^T x_u + \frac{\mu}{2} \|x_u - x_v\|^2$$

$$f(x_u) - v^T x_u \geq f(x_v) - v^T x_v + \frac{\mu}{2} \|x_u - x_v\|^2$$

Slopes of f and f^*

Assume that f is a closed and convex function. Then f is strongly convex with parameter $\mu \Leftrightarrow \nabla f^*$ is Lipschitz with parameter $1/\mu$.

Proof of “ \Rightarrow ”: Recall, if g is strongly convex with minimizer x , then

$$g(y) \geq g(x) + \frac{\mu}{2} \|y - x\|^2, \quad \text{for all } y$$

Hence, defining $x_u = \nabla f^*(u)$ and $x_v = \nabla f^*(v)$,

$$f(x_v) - u^T x_v \geq f(x_u) - u^T x_u + \frac{\mu}{2} \|x_u - x_v\|^2$$

$$f(x_u) - v^T x_u \geq f(x_v) - v^T x_v + \frac{\mu}{2} \|x_u - x_v\|^2$$

Adding these together, using the Cauchy-Schwarz inequality, and rearranging shows that

$$\|x_u - x_v\|^2 \leq \frac{1}{\mu} \|u - v\|^2$$

Slopes of f and f^*

Proof of “ \Leftarrow ”: for simplicity, call $g = f^*$ and $L = \frac{1}{\mu}$. As ∇g is Lipschitz with constant L , so is $g_x(z) = g(z) - \nabla g(x)^T z$, hence

$$g_x(z) \leq g_x(y) + \nabla g_x(y)^T (z - y) + \frac{L}{2} \|z - y\|_2^2$$

Slopes of f and f^*

Proof of “ \Leftarrow ”: for simplicity, call $g = f^*$ and $L = \frac{1}{\mu}$. As ∇g is Lipschitz with constant L , so is $g_x(z) = g(z) - \nabla g(x)^T z$, hence

$$g_x(z) \leq g_x(y) + \nabla g_x(y)^T (z - y) + \frac{L}{2} \|z - y\|_2^2$$

Minimizing each side over z , and rearranging, gives

$$\frac{1}{2L} \|\nabla g(x) - \nabla g(y)\|^2 \leq g(y) - g(x) + \nabla g(x)^T (x - y)$$

Slopes of f and f^*

Proof of “ \Leftarrow ”: for simplicity, call $g = f^*$ and $L = \frac{1}{\mu}$. As ∇g is Lipschitz with constant L , so is $g_x(z) = g(z) - \nabla g(x)^T z$, hence

$$g_x(z) \leq g_x(y) + \nabla g_x(y)^T (z - y) + \frac{L}{2} \|z - y\|_2^2$$

Minimizing each side over z , and rearranging, gives

$$\frac{1}{2L} \|\nabla g(x) - \nabla g(y)\|^2 \leq g(y) - g(x) + \nabla g(x)^T (x - y)$$

Exchanging roles of x , y , and adding together, gives

$$\frac{1}{L} \|\nabla g(x) - \nabla g(y)\|^2 \leq (\nabla g(x) - \nabla g(y))^T (x - y)$$

Slopes of f and f^*

Proof of “ \Leftarrow ”: for simplicity, call $g = f^*$ and $L = \frac{1}{\mu}$. As ∇g is Lipschitz with constant L , so is $g_x(z) = g(z) - \nabla g(x)^T z$, hence

$$g_x(z) \leq g_x(y) + \nabla g_x(y)^T (z - y) + \frac{L}{2} \|z - y\|_2^2$$

Minimizing each side over z , and rearranging, gives

$$\frac{1}{2L} \|\nabla g(x) - \nabla g(y)\|^2 \leq g(y) - g(x) + \nabla g(x)^T (x - y)$$

Exchanging roles of x , y , and adding together, gives

$$\frac{1}{L} \|\nabla g(x) - \nabla g(y)\|^2 \leq (\nabla g(x) - \nabla g(y))^T (x - y)$$

Let $u = \nabla f(x)$, $v = \nabla g(y)$; then $x \in \partial g^*(u)$, $y \in \partial g^*(v)$, and the above reads $(x - y)^T (u - v) \geq \frac{\|u - v\|^2}{L}$, implying the result.

Conjugate function properties

Recall that given $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the function defined by

$$f^*(y) = \max_x [y^T x - f(x)]$$

is called its conjugate.

- Conjugates appear frequently in dual programs, since

$$-f^*(y) = \min_x [f(x) - y^T x]$$

Conjugate function properties

Recall that given $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the function defined by

$$f^*(y) = \max_x [y^T x - f(x)]$$

is called its conjugate.

- Conjugates appear frequently in dual programs, since

$$-f^*(y) = \min_x [f(x) - y^T x]$$

- If f is closed and convex, then $f^{**} = f$. Also,

$$x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \arg \min_z [f(z) - y^T z]$$

Conjugate function properties

Recall that given $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the function defined by

$$f^*(y) = \max_x [y^T x - f(x)]$$

is called its conjugate.

- Conjugates appear frequently in dual programs, since

$$-f^*(y) = \min_x [f(x) - y^T x]$$

- If f is closed and convex, then $f^{**} = f$. Also,

$$x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \arg \min_z [f(z) - y^T z]$$

- If f is strictly convex, then

$$\nabla f^*(y) = \arg \min_z [f(z) - y^T z]$$

Conjugate function properties (proofs)

We will show that $x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x)$, assuming that f is convex and closed.

- **Proof of \Leftarrow :** Suppose $y \in \partial f(x)$. Then $x \in M_y$, the set of maximizers of $y^T z - f(z)$ over z . But

$$f^*(y) = \max_z \{y^T z - f(z)\} \quad \text{and} \quad \partial f^*(y) = \text{cl}(\text{conv}(\bigcup_{z \in M_y} \{z\})).$$

Thus $x \in \partial f^*(y)$.

Conjugate function properties (proofs)

We will show that $x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x)$, assuming that f is convex and closed.

- **Proof of \Leftarrow :** Suppose $y \in \partial f(x)$. Then $x \in M_y$, the set of maximizers of $y^T z - f(z)$ over z . But

$$f^*(y) = \max_z \{y^T z - f(z)\} \quad \text{and} \quad \partial f^*(y) = \text{cl}(\text{conv}(\bigcup_{z \in M_y} \{z\})).$$

Thus $x \in \partial f^*(y)$.

- **Proof of \Rightarrow :** From what we showed above, if $x \in \partial f^*(y)$, then $y \in \partial f^*(x)$, but $f^{**} = f$.

Conjugate function properties (proofs)

We will show that $x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x)$, assuming that f is convex and closed.

- **Proof of \Leftarrow :** Suppose $y \in \partial f(x)$. Then $x \in M_y$, the set of maximizers of $y^T z - f(z)$ over z . But

$$f^*(y) = \max_z \{y^T z - f(z)\} \quad \text{and} \quad \partial f^*(y) = \text{cl}(\text{conv}(\bigcup_{z \in M_y} \{z\})).$$

Thus $x \in \partial f^*(y)$.

- **Proof of \Rightarrow :** From what we showed above, if $x \in \partial f^*(y)$, then $y \in \partial f^*(x)$, but $f^{**} = f$.

Conjugate function properties (proofs)

We will show that $x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x)$, assuming that f is convex and closed.

- **Proof of \Leftarrow :** Suppose $y \in \partial f(x)$. Then $x \in M_y$, the set of maximizers of $y^T z - f(z)$ over z . But

$$f^*(y) = \max_z \{y^T z - f(z)\} \quad \text{and} \quad \partial f^*(y) = \text{cl}(\text{conv}(\bigcup_{z \in M_y} \{z\})).$$

Thus $x \in \partial f^*(y)$.

- **Proof of \Rightarrow :** From what we showed above, if $x \in \partial f^*(y)$, then $y \in \partial f^*(x)$, but $f^{**} = f$.

Clearly $y \in \partial f(x) \Leftrightarrow x \in \arg \min_z \{f(z) - y^T z\}$

Lastly, if f is strictly convex, then we know that $f(z) - y^T z$ has a unique minimizer over z , and this must be $\nabla f^*(y)$.

Subgradient and Subdifferential

ℓ_1 -regularized linear least squares

ℓ_1 induces sparsity

ℓ_2 regularization. $\|Xw - y\|_2^2 \rightarrow \min_{\|w\|_2 \leq 1}$



ℓ_1 regularization. $\|Xw - y\|_2^2 \rightarrow \min_{\|w\|_1 \leq 1}$



@fminxyz

Norms are not smooth

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

A classical convex optimization problem is considered. We assume that $f(x)$ is a convex function, but now we do not require smoothness.

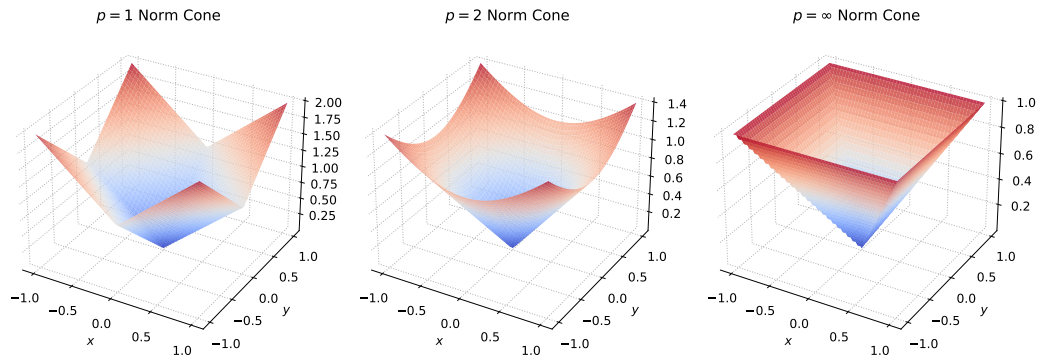


Figure 5: Norm cones for different p - norms are non-smooth

Convex function linear lower bound

An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

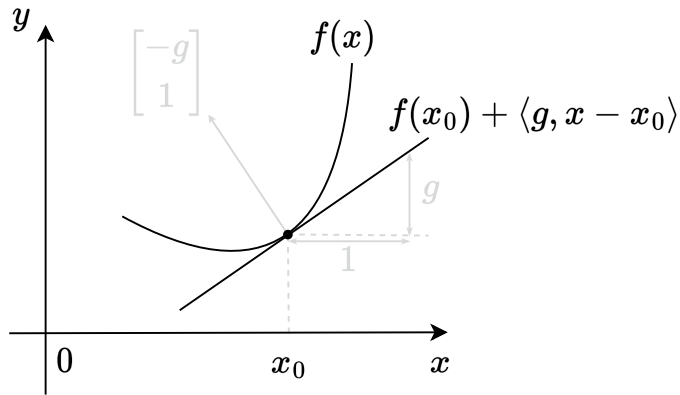
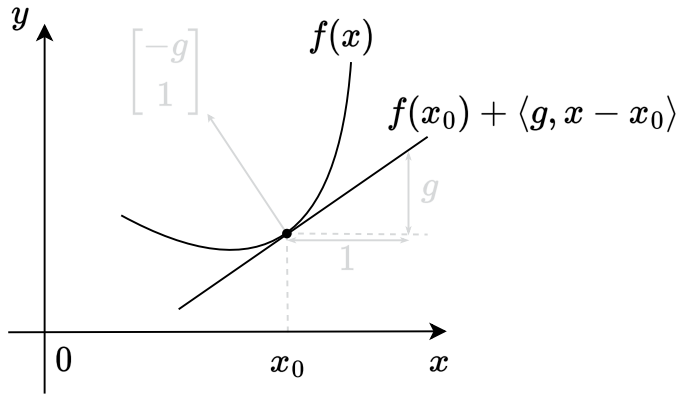


Figure 6: Taylor linear approximation serves as a global lower bound for a convex function

Convex function linear lower bound



An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

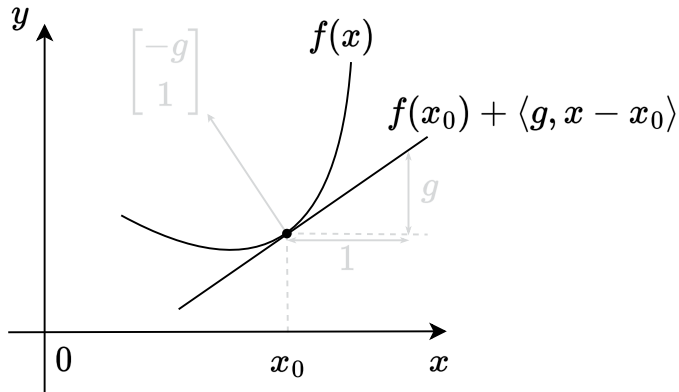
$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g , i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$

Figure 6: Taylor linear approximation serves as a global lower bound for a convex function

Convex function linear lower bound



An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

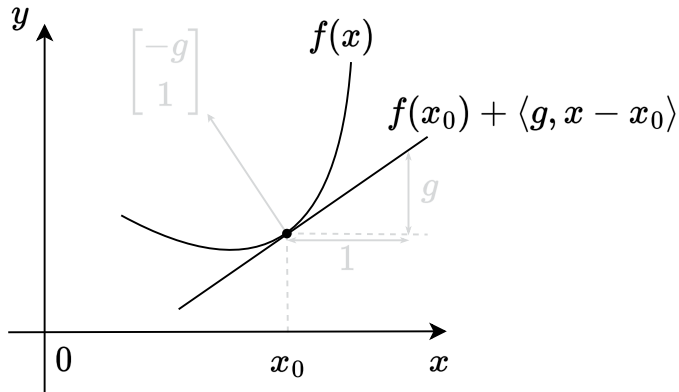
$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g , i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$
- Not all continuous convex functions are differentiable.

Figure 6: Taylor linear approximation serves as a global lower bound for a convex function

Convex function linear lower bound



An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

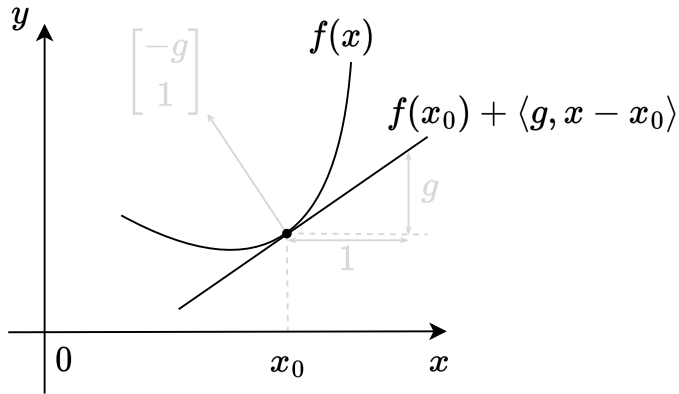
$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g , i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$
- Not all continuous convex functions are differentiable.

Figure 6: Taylor linear approximation serves as a global lower bound for a convex function

Convex function linear lower bound



An important property of a continuous convex function $f(x)$ is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

for some vector g , i.e., the tangent to the graph of the function is the *global* estimate from below for the function.

- If $f(x)$ is differentiable, then $g = \nabla f(x_0)$
- Not all continuous convex functions are differentiable.

We wouldn't want to lose such a nice property.

Figure 6: Taylor linear approximation serves as a global lower bound for a convex function

Subgradient and subdifferential

A vector g is called the **subgradient** of a function $f(x) : S \rightarrow \mathbb{R}$ at a point x_0 if $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

Subgradient and subdifferential

A vector g is called the **subgradient** of a function $f(x) : S \rightarrow \mathbb{R}$ at a point x_0 if $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

The set of all subgradients of a function $f(x)$ at a point x_0 is called the **subdifferential** of f at x_0 and is denoted by $\partial f(x_0)$.

Subgradient and subdifferential

A vector g is called the **subgradient** of a function $f(x) : S \rightarrow \mathbb{R}$ at a point x_0 if $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

The set of all subgradients of a function $f(x)$ at a point x_0 is called the **subdifferential** of f at x_0 and is denoted by $\partial f(x_0)$.



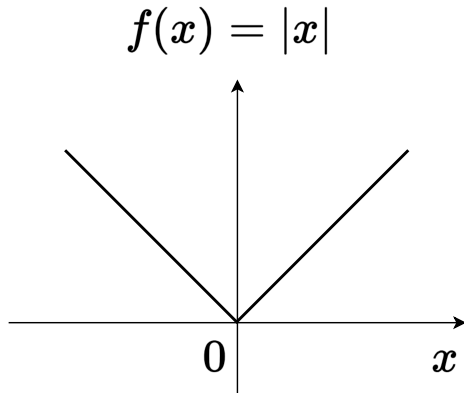
Figure 7: Subdifferential is a set of all possible subgradients

Subgradient and subdifferential

Find $\partial f(x)$, if $f(x) = |x|$

Subgradient and subdifferential

Find $\partial f(x)$, if $f(x) = |x|$



Subdifferential properties

- If $x_0 \in \text{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.

Subdifferential properties

- If $x_0 \in \text{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S .

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S .

Subdifferential properties

- If $x_0 \in \text{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S .

i Subdifferential of a differentiable function

Let $f : S \rightarrow \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \text{ri}(S)$ and f is differentiable at x_0 , then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Subdifferential properties

- If $x_0 \in \text{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S .

i Subdifferential of a differentiable function

Let $f : S \rightarrow \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \text{ri}(S)$ and f is differentiable at x_0 , then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S , there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

Subdifferential properties

- If $x_0 \in \text{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S .

i Subdifferential of a differentiable function

Let $f : S \rightarrow \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \text{ri}(S)$ and f is differentiable at x_0 , then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S , there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

Subdifferential properties

- If $x_0 \in \text{ri}(S)$, then $\partial f(x_0)$ is a convex compact set. which implies:
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S .

i Subdifferential of a differentiable function

Let $f : S \rightarrow \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \text{ri}(S)$ and f is differentiable at x_0 , then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S , there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

for all $0 < t < \delta$. Taking the limit as t approaches 0 and using the definition of the gradient, we get:

$$\langle \nabla f(x_0), v \rangle = \lim_{t \rightarrow 0; 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

2. From this, $\langle s - \nabla f(x_0), v \rangle \geq 0$. Due to the arbitrariness of v , one can set

$$v = -\frac{s - \nabla f(x_0)}{\|s - \nabla f(x_0)\|},$$

leading to $s = \nabla f(x_0)$.

Subdifferential properties

- If $x_0 \in \text{ri}(S)$, then $\partial f(x_0)$ is a convex compact set. which implies:
- The convex function $f(x)$ is differentiable at the point $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- If $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, then $f(x)$ is convex on S .

i Subdifferential of a differentiable function

Let $f : S \rightarrow \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \text{ri}(S)$ and f is differentiable at x_0 , then either $\partial f(x_0) = \emptyset$ or $\partial f(x_0) = \{\nabla f(x_0)\}$. Moreover, if the function f is convex, the first scenario is impossible.

Proof

1. Assume, that $s \in \partial f(x_0)$ for some $s \in \mathbb{R}^n$ distinct from $\nabla f(x_0)$. Let $v \in \mathbb{R}^n$ be a unit vector. Because x_0 is an interior point of S , there exists $\delta > 0$ such that $x_0 + tv \in S$ for all $0 < t < \delta$. By the definition of the subgradient, we have

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

for all $0 < t < \delta$. Taking the limit as t approaches 0 and using the definition of the gradient, we get:

$$\langle \nabla f(x_0), v \rangle = \lim_{t \rightarrow 0; 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

2. From this, $\langle s - \nabla f(x_0), v \rangle \geq 0$. Due to the arbitrariness of v , one can set

$$v = -\frac{s - \nabla f(x_0)}{\|s - \nabla f(x_0)\|},$$

leading to $s = \nabla f(x_0)$.

3. Furthermore, if the function f is convex, then according to the differential condition of convexity $f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$ for all $x \in S$. But by definition, this means $\nabla f(x_0) \in \partial f(x_0)$.

Subdifferentiability and convexity

i Question

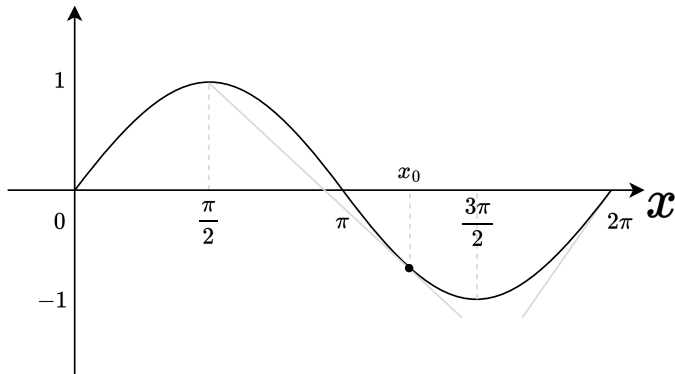
Is it correct, that if the function has a subdifferential at some point, the function is convex?

Subdifferentiability and convexity

i Question

Is it correct, that if the function has a subdifferential at some point, the function is convex?

Find $\partial f(x)$, if $f(x) = \sin x, x \in [\pi/2; 2\pi]$



Subdifferentiability and convexity

Question

Is it correct, that if the function is convex, it has a subgradient at any point?

Subdifferentiability and convexity

Question

Is it correct, that if the function is convex, it has a subgradient at any point?

Convexity follows from subdifferentiability at any point. A natural question to ask is whether the converse is true: is every convex function subdifferentiable? It turns out that, generally speaking, the answer to this question is negative.

Let $f : [0, \infty) \rightarrow \mathbb{R}$ be the function defined by $f(x) := -\sqrt{x}$. Then, $\partial f(0) = \emptyset$.

Assume, that $s \in \partial f(0)$ for some $s \in \mathbb{R}$. Then, by definition, we must have $sx \leq -\sqrt{x}$ for all $x \geq 0$. From this, we can deduce $s \leq -\sqrt{1/x}$ for all $x > 0$. Taking the limit as x approaches 0 from the right, we get $s \leq -\infty$, which is impossible.

Subdifferential calculus

i Moreau - Rockafellar theorem (subdifferential of a linear combination)

Let $f_i(x)$ be convex functions on convex sets S_i , $i = \overline{1, n}$. Then if $\bigcap_{i=1}^n \text{ri}(S_i) \neq \emptyset$ then the function

$f(x) = \sum_{i=1}^n a_i f_i(x)$, $a_i > 0$ has a subdifferential

$\partial_S f(x)$ on the set $S = \bigcap_{i=1}^n S_i$ and

$$\partial_S f(x) = \sum_{i=1}^n a_i \partial_{S_i} f_i(x)$$

Subdifferential calculus

i Moreau - Rockafellar theorem (subdifferential of a linear combination)

Let $f_i(x)$ be convex functions on convex sets S_i , $i = \overline{1, n}$. Then if $\bigcap_{i=1}^n \text{ri}(S_i) \neq \emptyset$ then the function

$f(x) = \sum_{i=1}^n a_i f_i(x)$, $a_i > 0$ has a subdifferential

$\partial_S f(x)$ on the set $S = \bigcap_{i=1}^n S_i$ and

$$\partial_S f(x) = \sum_{i=1}^n a_i \partial_{S_i} f_i(x)$$

i Dubovitsky - Milutin theorem (subdifferential of a point-wise maximum)

Let $f_i(x)$ be convex functions on the open convex set $S \subseteq \mathbb{R}^n$, $x_0 \in S$, and the pointwise maximum is defined as $f(x) = \max_i f_i(x)$. Then:

$$\partial_S f(x_0) = \text{conv} \left\{ \bigcup_{i \in I(x_0)} \partial_S f_i(x_0) \right\}, \quad I(x) = \{i \in [1, n] \mid f_i(x) = f(x)\}$$

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - convex functions

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - convex functions
- $\partial(f(Ax + b))(x) = A^T \partial f(Ax + b)$, f - convex function

Subdifferential calculus

- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, f_i - convex functions
- $\partial(f(Ax + b))(x) = A^T \partial f(Ax + b)$, f - convex function
- $z \in \partial f(x)$ if and only if $x \in \partial f^*(z)$.

Connection to convex geometry

Convex set $S \subseteq \mathbb{R}^n$, consider indicator function $I_S : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

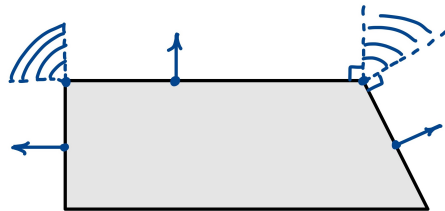
For $x \in S$, $\partial I_S(x) = \mathcal{N}_S(x)$, the **normal cone** of S at x is, recall

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in S\}$$

Why? By definition of subgradient g ,

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

- For $y \notin S$, $I_S(y) = \infty$



Connection to convex geometry

Convex set $S \subseteq \mathbb{R}^n$, consider indicator function $I_S : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$I_S(x) = I\{x \in S\} = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{if } x \notin S \end{cases}$$

For $x \in S$, $\partial I_S(x) = \mathcal{N}_S(x)$, the **normal cone** of S at x is, recall

$$\mathcal{N}_S(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in S\}$$

Why? By definition of subgradient g ,

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

- For $y \notin S$, $I_S(y) = \infty$
- For $y \in S$, this means $0 \geq g^T(y - x)$



Optimality Condition

For any f (convex or not),

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

That is, x^* is a minimizer if and only if 0 is a subgradient of f at x^* . This is called the **subgradient optimality condition**.

Why? Easy: $g = 0$ being a subgradient means that for all y

$$f(y) \geq f(x^*) + 0^T(y - x^*) = f(x^*)$$

Note the implication for a convex and differentiable function f , with

$$\partial f(x) = \{\nabla f(x)\}$$

Derivation of first-order optimality

Example of the power of subgradients: we can use what we have learned so far to derive the **first-order optimality condition**. Recall

$$\min_x f(x) \text{ subject to } x \in S$$

is solved at x , for f convex and differentiable, if and only if

$$\nabla f(x)^T(y - x) \geq 0 \quad \text{for all } y \in S$$

Intuitively: this says that the gradient increases as we move away from x . How to prove it? First, recast the problem as

$$\min_x f(x) + I_S(x)$$

Now apply subgradient optimality:

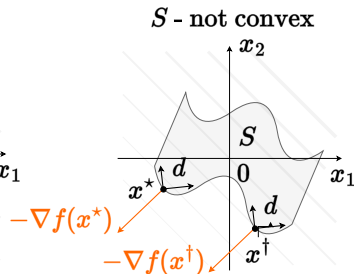
$$0 \in \partial(f(x) + I_S(x))$$

$$f(x) = x_1 + x_2 \rightarrow \min_{x_1, x_2 \in \mathbb{R}^2}$$



$$\langle -\nabla f(x^*), d \rangle \leq 0$$

x^* - optimal



$$\langle -\nabla f(x^\dagger), d \rangle \leq 0$$

x^\dagger - not optimal

Derivation of first-order optimality

Observe

$$0 \in \partial(f(x) + I_S(x))$$

$$\Leftrightarrow 0 \in \{\nabla f(x)\} + \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x) \in \mathcal{N}_S(x)$$

$$\Leftrightarrow -\nabla f(x)^T x \geq -\nabla f(x)^T y \text{ for all } y \in S$$

$$\Leftrightarrow \nabla f(x)^T (y - x) \geq 0 \text{ for all } y \in S$$

as desired.

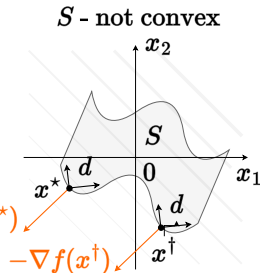
Note: the condition $0 \in \partial f(x) + \mathcal{N}_S(x)$ is a **fully general condition** for optimality in convex problems. But it's not always easy to work with (KKT conditions, later, are easier).

$$f(x) = x_1 + x_2 \rightarrow \min_{x_1, x_2 \in \mathbb{R}^2}$$



$$\langle -\nabla f(x^*), d \rangle \leq 0$$

x^* - optimal



$$\langle -\nabla f(x^\dagger), d^\dagger \rangle \leq 0$$

x^\dagger - not optimal

Example 1

Example

Find $\partial f(x)$, if $f(x) = |x - 1| + |x + 1|$

Example 1

i Example

Find $\partial f(x)$, if $f(x) = |x - 1| + |x + 1|$

$$\partial f_1(x) = \begin{cases} -1, & x < 1 \\ [-1; 1], & x = 1 \\ 1, & x > 1 \end{cases} \quad \partial f_2(x) = \begin{cases} -1, & x < -1 \\ [-1; 1], & x = -1 \\ 1, & x > -1 \end{cases}$$

So

$$\partial f(x) = \begin{cases} -2, & x < -1 \\ [-2; 0], & x = -1 \\ 0, & -1 < x < 1 \\ [0; 2], & x = 1 \\ 2, & x > 1 \end{cases}$$

Example 2

Find $\partial f(x)$ if $f(x) = [\max(0, f_0(x))]^q$. Here, $f_0(x)$ is a convex function on an open convex set S , and $q \geq 1$.

Example 2

Find $\partial f(x)$ if $f(x) = [\max(0, f_0(x))]^q$. Here, $f_0(x)$ is a convex function on an open convex set S , and $q \geq 1$.

According to the composition theorem (the function $\varphi(x) = x^q$ is differentiable) and $g(x) = \max(0, f_0(x))$, we have:

$$\partial f(x) = q(g(x))^{q-1} \partial g(x)$$

By the theorem on the pointwise maximum:

$$\partial g(x) = \begin{cases} \partial f_0(x), & f_0(x) > 0, \\ \{0\}, & f_0(x) < 0, \\ \{a \mid a = \lambda a', 0 \leq \lambda \leq 1, a' \in \partial f_0(x)\}, & f_0(x) = 0 \end{cases}$$

Example 3. Subdifferential of the Norm

Let V be a finite-dimensional Euclidean space, and $x_0 \in V$. Let $\|\cdot\|$ be an arbitrary norm in V (not necessarily induced by the scalar product), and let $\|\cdot\|_*$ be the corresponding conjugate norm. Then,

$$\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0, 1), & \text{if } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{otherwise.} \end{cases}$$

Where $B_{\|\cdot\|_*}(0, 1)$ is the closed unit ball centered at zero with respect to the conjugate norm. In other words, a vector $s \in V$ with $\|s\|_* = 1$ is a subgradient of the norm $\|\cdot\|$ at point $x_0 \neq 0$ if and only if the Hölder's inequality $\langle s, x_0 \rangle \leq \|x_0\|$ becomes an equality.

Example 3. Subdifferential of the Norm

Let V be a finite-dimensional Euclidean space, and $x_0 \in V$. Let $\|\cdot\|$ be an arbitrary norm in V (not necessarily induced by the scalar product), and let $\|\cdot\|_*$ be the corresponding conjugate norm. Then,

$$\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0, 1), & \text{if } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{otherwise.} \end{cases}$$

Where $B_{\|\cdot\|_*}(0, 1)$ is the closed unit ball centered at zero with respect to the conjugate norm. In other words, a vector $s \in V$ with $\|s\|_* = 1$ is a subgradient of the norm $\|\cdot\|$ at point $x_0 \neq 0$ if and only if the Hölder's inequality $\langle s, x_0 \rangle \leq \|x_0\|$ becomes an equality.

Let $s \in V$. By definition, $s \in \partial\|\cdot\|(x_0)$ if and only if

$$\langle s, x \rangle - \|x\| \leq \langle s, x_0 \rangle - \|x_0\|, \text{ for all } x \in V,$$

or equivalently,

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} \leq \langle s, x_0 \rangle - \|x_0\|.$$

By the definition of the supremum, the latter is equivalent to

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} = \langle s, x_0 \rangle - \|x_0\|.$$

Subgradient and Subdifferential

Example 3. Subdifferential of the Norm

Let V be a finite-dimensional Euclidean space, and $x_0 \in V$. Let $\|\cdot\|$ be an arbitrary norm in V (not necessarily induced by the scalar product), and let $\|\cdot\|_*$ be the corresponding conjugate norm. Then,

$$\partial\|\cdot\|(x_0) = \begin{cases} B_{\|\cdot\|_*}(0, 1), & \text{if } x_0 = 0, \\ \{s \in V : \|s\|_* \leq 1; \langle s, x_0 \rangle = \|x_0\|\} = \{s \in V : \|s\|_* = 1; \langle s, x_0 \rangle = \|x_0\|\}, & \text{otherwise.} \end{cases}$$

Where $B_{\|\cdot\|_*}(0, 1)$ is the closed unit ball centered at zero with respect to the conjugate norm. In other words, a vector $s \in V$ with $\|s\|_* = 1$ is a subgradient of the norm $\|\cdot\|$ at point $x_0 \neq 0$ if and only if the Hölder's inequality $\langle s, x_0 \rangle \leq \|x_0\|$ becomes an equality.

Let $s \in V$. By definition, $s \in \partial\|\cdot\|(x_0)$ if and only if

$$\langle s, x \rangle - \|x\| \leq \langle s, x_0 \rangle - \|x_0\|, \text{ for all } x \in V,$$

or equivalently,

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} \leq \langle s, x_0 \rangle - \|x_0\|.$$

By the definition of the supremum, the latter is equivalent to

It is important to note that the expression on the left side is the supremum from the definition of the Fenchel conjugate function for the norm, which is known to be

$$\sup_{x \in V} \{\langle s, x \rangle - \|x\|\} = \begin{cases} 0, & \text{if } \|s\|_* \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

Thus, equation is equivalent to $\|s\|_* \leq 1$ and $\langle s, x_0 \rangle = \|x_0\|$.

Example 3. Subdifferential of the Norm

Consequently, it remains to note that for $x_0 \neq 0$, the inequality $\|s\|_* \leq 1$ must become an equality since, when $\|s\|_* < 1$, Hölder's inequality implies $\langle s, x_0 \rangle \leq \|s\|_* \|x_0\| < \|x_0\|$.

The conjugate norm in Example above does not appear by chance. It turns out that, in a completely similar manner for an arbitrary function f (not just for the norm), its subdifferential can be described in terms of the dual object — the Fenchel conjugate function.