

Conjugate sets





Conjugate set

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

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

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

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

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

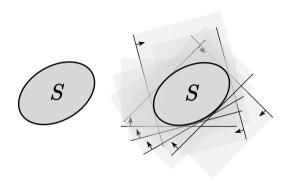


Figure 1: Convex sets may be described in a dual way through the elements of the set and through the set of hyperplanes supporting it

⊕ ∩ **ø**

Conjugate set

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

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

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

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

- 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 $S^* = S$.

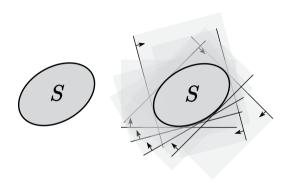


Figure 1: Convex sets may be described in a dual way through the elements of the set and through the set of hyperplanes supporting it

എറ മ

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

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

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

- 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 $S_2^* \subseteq S_1^*$.

େ ପ

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

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

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



i Example

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

i Example

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

• $S \subset \overline{S} \to (\overline{S})^* \subset S^*$.

i Example

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

- $S \subset \overline{S} \to (\overline{S})^* \subset S^*$.
- Let $p \in S^*$ and $x_0 \in \overline{S}, x_0 = \lim_{k \to \infty} x_k$. 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 $S^* \subset (\overline{S})^*$.

i Example

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

i Example

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

• $S \subset \mathbf{conv}(S) \to (\mathbf{conv}(S))^* \subset S^*$.

i Example

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

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

i Example

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

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 $x \in X : p^T x \ge -1$.

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 $x \in X : p^T x \ge -1$.
- 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 $x = -\frac{p}{\|p\|}r$.

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

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

So $X^* \subset Y$.

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 $x \in X : p^T x > -1$.
- From all points of the ball X, take such a point $x \in X$ that its scalar product of p: p^Tx is minimal, then this is the point $x = -\frac{p}{\|x\|}r$.

$$p^{T}x = p^{T} \left(-\frac{p}{\|p\|} r \right) = -\|p\|r \ge -1$$
$$\|p\| \le \frac{1}{r} \in Y$$

So $X^* \subset Y$.

• 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:

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

The latter comes from the fact that $p\in B(0,1/r)$ and $x\in B(0,r).$ So $Y\subset X^*.$

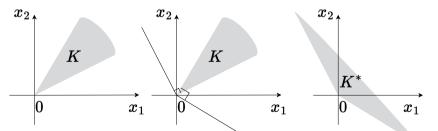
Dual cone

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

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

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.

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



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

Conjugate sets

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

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

େ ଚ

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

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



i Example

Find the conjugate cone for a monotone nonnegative cone:

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



i Example

Find the conjugate cone for a monotone nonnegative cone:

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

Note that:

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

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

$$y_1 > 0$$
, $y_1 + y_2 > 0$, ..., $y_1 + \ldots + y_n > 0$

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

 $f \to \min_{x,y,z}$ Conjugate sets

Polyhedra

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

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

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

i Theorem

Let $x_1, \ldots, x_m \in \mathbb{R}^n$. 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):

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

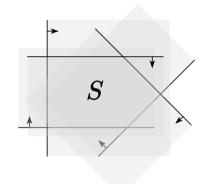


Figure 3: Polyhedra

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 $\theta > 0, i = \overline{k+1, m}$:

$$\langle p, x_i \rangle \ge -1 \to \langle p, \theta x_i \rangle \ge -1$$

$$\langle p, x_i \rangle \ge -\frac{1}{\theta} \to \langle p, x_i \rangle \ge 0.$$

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

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 $\theta>0, i=\overline{k+1,m}$:

$$\langle p, x_i \rangle \ge -1 \to \langle p, \theta x_i \rangle \ge -1$$

$$\langle p, x_i \rangle \ge -\frac{1}{2} \to \langle p, x_i \rangle \ge 0.$$

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

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

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

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

So:

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



Conjugate sets

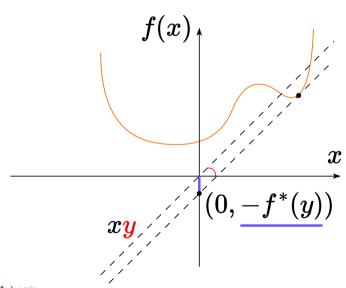


Conjugate functions





Conjugate functions

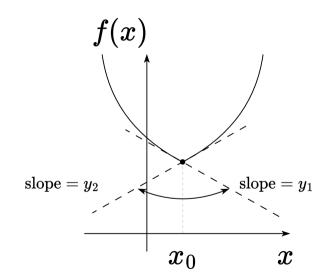


Recall that given $f:\mathbb{R}^n \to \mathbb{R}$, the function defined by

$$f^*(y) = \max_{x} \left[y^T x - f(x) \right]$$

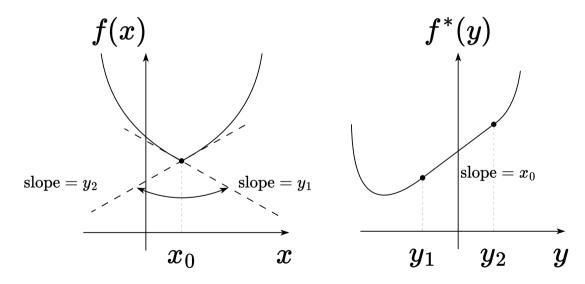
is called its conjugate.

Geometrical intution



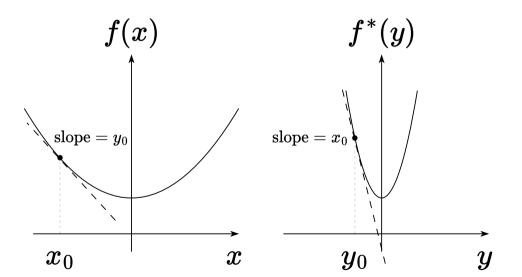


Geometrical intution



Conjugate functions

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 $1/\mu$.



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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

 $f \to \min_{x,y,z}$ Conjugate functions

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

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



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

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

Minimizing each side over z, and rearranging, gives

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



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

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

Minimizing each side over z, and rearranging, gives

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

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

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

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

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

Minimizing each side over z, and rearranging, gives

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

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

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

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 $f: \mathbb{R}^n \to \mathbb{R}$, the function defined by

$$f^*(y) = \max_{x} \left[y^T x - f(x) \right]$$

is called its conjugate.

Conjugates appear frequently in dual programs, since

$$-f^*(y) = \min_{x} \left[f(x) - y^T x \right]$$



Conjugate function properties

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

$$f^*(y) = \max_{x} \left[y^T x - f(x) \right]$$

is called its conjugate.

Conjugates appear frequently in dual programs, since

$$-f^*(y) = \min_{x} \left[f(x) - y^T x \right]$$

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

$$x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \arg\min_{z} \left[f(z) - y^T z \right]$$



Conjugate function properties

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

$$f^*(y) = \max_{x} \left[y^T x - f(x) \right]$$

is called its conjugate.

Conjugates appear frequently in dual programs, since

$$-f^*(y) = \min_{x} \left[f(x) - y^T x \right]$$

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

$$x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \arg\min_{z} \left[f(z) - y^T z \right]$$

• If f is strictly convex, then

$$\nabla f^*(y) = \arg\min_{z} \left[f(z) - y^T z \right]$$



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 $y^Tz - f(z)$ over z. But

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

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



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 $y^Tz - f(z)$ over z. But

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

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 $f^{**} = f$.





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 $y^Tz - f(z)$ over z. But

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

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 $f^{**} = f$.





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 $y^Tz - f(z)$ over z. But

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

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

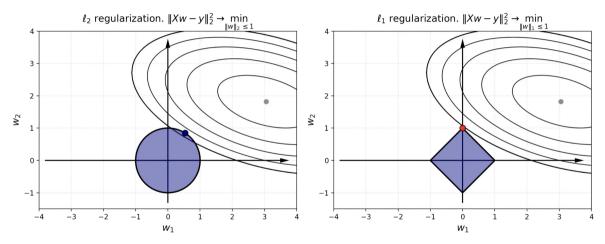
Lastly, if f is strictly convex, then we know that $f(z) - y^T z$ has a unique minimizer over z, and this must be $\nabla f^*(y)$.





ℓ_1 -regularized linear least squares

ℓ_1 induces sparsity



@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.

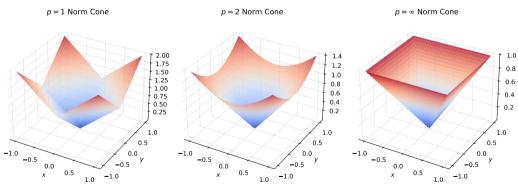
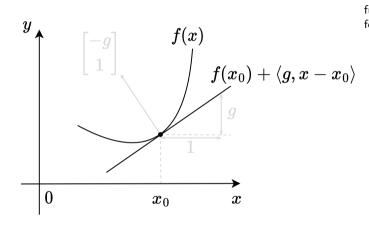


Figure 5: Norm cones for different p - norms are non-smooth



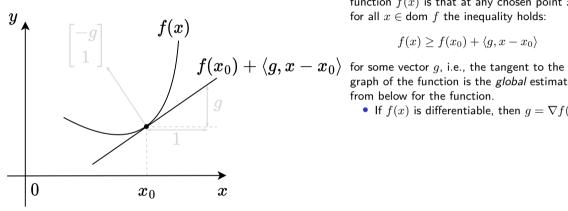




An important property of a continuous convex function f(x) is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

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

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



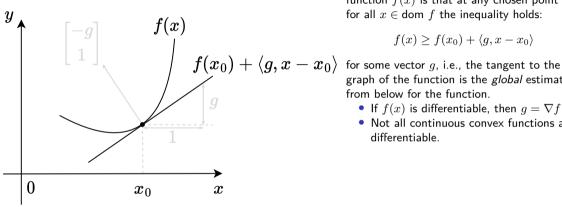
An important property of a continuous convex function f(x) is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

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

graph of the function is the global estimate • 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

Subgradient and Subdifferential



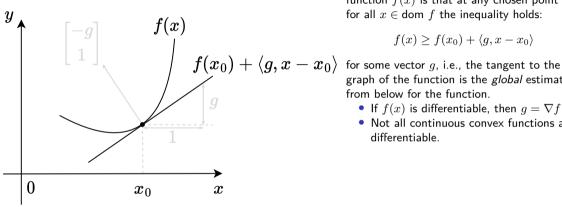
An important property of a continuous convex function f(x) is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

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

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



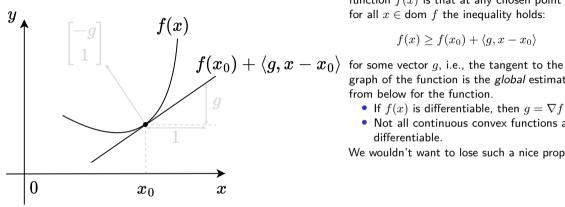
An important property of a continuous convex function f(x) is that at any chosen point x_0 for all $x \in \text{dom } f$ the inequality holds:

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

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



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:

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

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

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

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

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

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

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 $\partial f(x_0)$.

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

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

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 $\partial f(x_0)$.

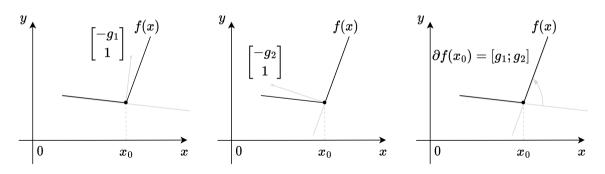
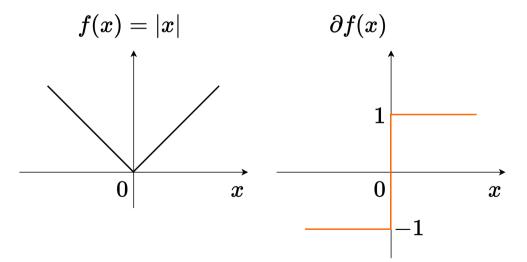


Figure 7: Subdifferential is a set of all possible subgradients

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



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



Subdifferential properties
• If $x_0 \in \mathbf{ri}(S)$, then $\partial f(x_0)$ is a convex compact set.



- 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 $x_0 \in \mathbf{ri}(S)$, 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.

₩ €

- 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.

₩ €

- 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.
- i Subdifferential of a differentiable function

Let $f: S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \mathbf{ri}(S)$ 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.



- 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.
- i Subdifferential of a differentiable function

Let $f: S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \mathbf{ri}(S)$ 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

- 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.
- i Subdifferential of a differentiable function

Let $f: S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \mathbf{ri}(S)$ 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

- If $x_0 \in \mathbf{ri}(S)$, 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 $f: S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \mathbf{ri}(S)$ 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

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

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

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

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 $s = \nabla f(x_0)$.

$$f o \min_{x \in \mathbb{R}^d} \quad f(x_0 + tv) \geq f(x_0) + t\langle s, v
angle$$
 Subgradient and Subdifferential

of the subgradient, we have

Subdifferential properties

- If $x_0 \in \mathbf{ri}(S)$, 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 $\partial f(x_0) \neq \emptyset$ $\forall x_0 \in S$, then f(x) is convex on S.
- i Subdifferential of a differentiable function
- Let $f: S \to \mathbb{R}$ be a function defined on the set S in a Euclidean space \mathbb{R}^n . If $x_0 \in \mathbf{ri}(S)$ 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

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

for all $0 < t < \delta$. Taking the limit as t approaches 0 and

using the definition of the gradient, we get:
$$\langle \nabla f(x_0), v \rangle = \lim_{t \to 0: 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

2. From this, $\langle s - \nabla f(x_0), v \rangle > 0$. 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 $\nabla f(x_0) \in \partial f(x_0)$.

i Question

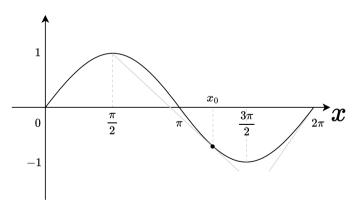
Is it correct, that if the function has a subdifferential at some point, the function is convex?



i Question

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

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



⊕ ი

i Question

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





i 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 $f:[0,\infty)\to\mathbb{R}$ 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 $s\leq -\sqrt{1}$ for all x>0. Taking the limit as x approaches 0 from the right, we get $s\leq -\infty$, which is impossible.



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 $\bigcap_{i=1}^n \mathbf{ri}(S_i) \neq \emptyset$ then the function

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

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

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



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 $\bigcap^n \mathbf{ri}(S_i) \neq \emptyset$ then the function

$$f(x)=\sum\limits_{i=1}^n a_if_i(x),\ a_i>0$$
 has a subdifferential $\partial_S f(x)$ on the set $S=\bigcap\limits_{i=1}^n S_i$ and

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

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 $f(x) = \max f_i(x)$. Then:

$$\partial_S f(x_0) = \mathbf{conv} \left\{ igcup_{i \in I(x_0)} \partial_S f_i(x_0)
ight\}, \quad I(x) = \{i \in [1], i \in [n]\}$$



•
$$\partial(\alpha f)(x) = \alpha \partial f(x)$$
, for $\alpha \ge 0$

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



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



- $\partial(\alpha f)(x) = \alpha \partial f(x)$, for $\alpha > 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, 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 $I_S : \mathbb{R}^n \to \mathbb{R}$,

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

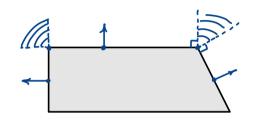
For $x \in S$, $\partial I_S(x) = \mathcal{N}_S(x)$, the **normal cone** of S at x is, recall

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

Why? By definition of subgradient g,

$$I_S(y) \ge I_S(x) + g^T(y - x)$$
 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 $I_S:\mathbb{R}^n\to\mathbb{R}$,

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

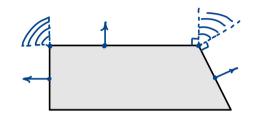
For $x \in S$, $\partial I_S(x) = \mathcal{N}_S(x)$, the **normal cone** of S at x is, recall

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

Why? By definition of subgradient g,

$$I_S(y) \ge I_S(x) + g^T(y-x)$$
 for all y

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





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

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

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

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

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

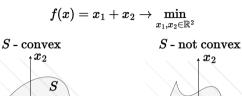
$$\nabla f(x)^T (y - x) \ge 0 \quad \text{for all } y \in S$$

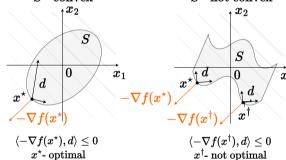
Intuitively: this says that the gradient increases as we move away from x. How to prove it? First, recast the problem as

$$\min_{x} f(x) + I_S(x)$$

Now apply subgradient optimality:

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







Derivation of first-order optimality

Observe

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

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

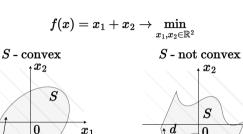
$$\Leftrightarrow -\nabla f(x) \in \mathcal{N}_S(x)$$

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

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

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).





i Example

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



i Example

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

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

So

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

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 $q \ge 1$.



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 $q \ge 1$.

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:

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

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 $\langle s, x_0 \rangle < ||x_0||$ becomes an equality.

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,

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

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.

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

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

or equivalently,

$$\sup_{s \in \mathcal{S}} \{ \langle s, x \rangle - ||x|| \} \le \langle s, x_0 \rangle - ||x_0||.$$

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

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,

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

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.

$$\langle s,x_0
angle \leq \|x_0\|$$
 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,

$$\sup\{\langle s, x \rangle - ||x||\} \le \langle s, x_0 \rangle - ||x_0||.$$

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

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

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.



