Outline

Question

Can you find any affine function that underestimates $f(x)$ and is tight at $x = 0$? What about when $x \neq 0$?

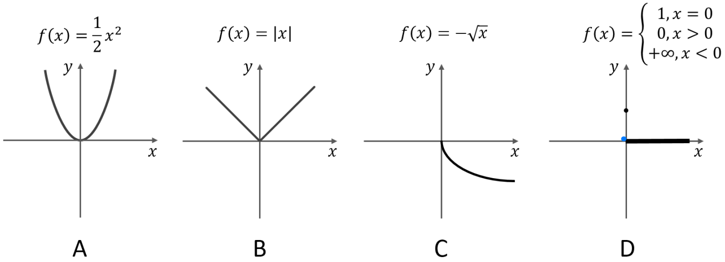


Figure: Convex Functions

Subgradient

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex.

Definition. A vector $g \in \mathbb{R}^n$ is a *subgradient* of f at a point $x_0 \in dom(f)$ if

$$f(x) \geq f(x_0) + g^T(x - x_0), \forall x.$$

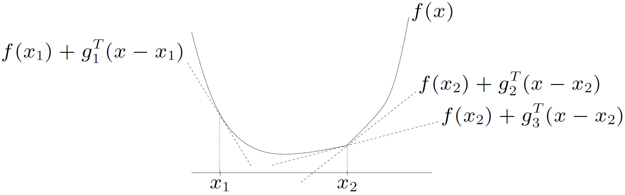


Figure: Subgradients

Definition. The set of all subgradient at x_0 is called the *subdifferential* of f at x_0 denoted as $\partial f(x_0)$.

Subgradient and Epigraph

Subgradients form supporting hyperplanes for the epigraph.

$$
\begin{aligned}
& g \in \partial f(x_0) \\
& \Leftrightarrow f(x) - g^T x \geq f(x_0) - g^T x_0, \forall x \\
& \Leftrightarrow t - g^T x \geq f(x_0) - g^T x_0, \forall (x,t) \in \mathrm{epi}(f) \\
& \Leftrightarrow \begin{bmatrix} -g \\ 1 \end{bmatrix}^T \begin{bmatrix} x \\ t \end{bmatrix} \geq \begin{bmatrix} -g \\ 1 \end{bmatrix}^T \begin{bmatrix} x_0 \\ f(x_0) \end{bmatrix}, \forall (x,t) \in \mathrm{epi}(f) \\
& \Leftrightarrow H := \left\{ (x,t) : (-g,1)^T (x,t) = (-g,1)^T (x_0, f(x_0)) \right\} \\
& \quad \text{is a supporting hyperplane of } \mathrm{epi}(f) \text{ at } (x_0, f(x_0))
\end{aligned}
$$

Examples: Differentiable Functions

Example 1. If f is differentiable at $x \in dom(f)$, then

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

Proof. Let $y = x + \epsilon d$, $g \in \partial f(x)$, then

$$\begin{aligned} & f(x+\epsilon d) \geq f(x) + \epsilon g^T d \\ \Rightarrow & \frac{f(x+\epsilon d) - f(x)}{\epsilon} \geq g^T d, \forall d, \forall \epsilon \\ \Rightarrow & \nabla f(x)^T d \geq g^T d, \forall d, \text{ as } \epsilon \to 0 \\ \Rightarrow & g = \nabla f(x). \end{aligned}$$

Examples: Simple Functions

Example 2.

(a) $f(x) = \frac{1}{2}x^2$, $\partial f(x) = x$

(b) $f(x) = |x|$, $\partial f(x) = \begin{cases} \mathrm{sgn}(x), x \neq 0 \\ [-1, 1], x = 0 \end{cases}$.

(c) $f(x) = \begin{cases} -\sqrt{x}, x \geq 0 \\ +\infty, o.w. \end{cases}$, $\partial f(x) = \begin{cases} -\frac{1}{2\sqrt{x}}, x > 0 \\ \emptyset, x = 0 \end{cases}$.

(d) $f(x) = \begin{cases} 1, x = 0 \\ 0, x > 0 \\ +\infty, o.w. \end{cases}$, $\partial f(x) = \begin{cases} 0, x > 0 \\ \emptyset, x = 0 \end{cases}$.

Closedness of Subdifferential

Proposition. Let f be convex and $x_0 \in \text{dom}(f)$. Then $\partial f(x_0)$ is convex and closed.

Proof. This is because

$$\begin{aligned}\partial f(x_0) &= \left\{g \in \mathbb{R}^n : f(x) \geq f(x_0) + g^T(x - x_0), \forall x\right\} \\ &= \cap_x \left\{g \in \mathbb{R}^n : f(x) \geq f(x_0) + g^T(x - x_0)\right\}\end{aligned}$$

is the solution to an infinite system of linear inequalities.

$\lambda \times f(x) \geq f(x_0) + g^T(x - x_0)$

$(1-\lambda) \times f(x) \geq f(x_0) + h^T(x - x_0)$

Existence of Subgradient

Theorem. Let f be convex and $x_0 \in \mathrm{rint}(\mathrm{dom}(f))$. Then $\partial f(x_0)$ is nonempty and bounded.

Remark. The reverse is also true. If $\forall x_0 \in \mathrm{dom}(f), \partial f(x_0)$ is non-empty, and $\mathrm{dom}(f)$ is convex, then f is convex.

Proof. Let $g \in \partial f(x_0)$ and $x_0 = \lambda x + (1-\lambda)y$, we have

$$\begin{array}{l} \lambda \\ (1-\lambda) \end{array} \begin{cases} f(x) \geq f(x_0) + g^T(x - x_0) & +(1-\lambda)(x-y) \\ f(y) \geq f(x_0) + g^T(y - x_0) & \lambda(y-x) \end{cases}$$

$$\Rightarrow \lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y)$$

Proof of Existence and Boundedness

- **(Nonempty)** By separation theorem, $\exists \alpha = (s, \beta) \neq 0$,

$$\underset{\beta}{s}^T x + \cancel{\beta} \overset{f(x)}{t} \geq \underset{\beta}{s}^T x_0 + \cancel{\beta} f(x_0), \forall (x, t) \in \text{epi}(f)$$

  We must have $\beta > 0$ (why?). Setting $g = -\beta^{-1} s$,

$$f(x) \geq f(x_0) + g^T (x - x_0), \forall x$$

- **(Bounded)** Suppose $\partial f(x_0)$ is unbounded, i.e. $\exists g_k \in \partial f(x_0)$, s.t. $\| g_k \|_2 \to \infty$, as $k \to \infty$. Let $x_k = x_0 + \delta \frac{g_k}{\|g_k\|_2} \in \text{dom}(f)$. By convexity,

$$f(x_k) \geq f(x_0) + g_k^T (x_k - x_0) = f(x_0) + \delta \| g_k \|_2 \to \infty.$$

  Contradicts with the continuity of f over $\text{int}(\text{dom}(f))$.

Monotonicity

Proposition. The subdifferential of a convex function f is a *monotone operator*, i.e.,

$$(u-v)^T(x-y) \geq 0, \forall x, y, u \in \partial f(x), v \in \partial f(y).$$

Proof.

By definition, we have

$$\begin{cases} f(y) \geq f(x) + u^T(y-x) \\ f(x) \geq f(y) + v^T(x-y) \end{cases}$$

Combining the two inequalities leads to the monotonicity.

Directional Derivative

Definition. The ***directional derivative*** of a function f at x along direction d is

$$f'(x;d) = \lim_{\delta \to 0^+} \frac{f(x+\delta d) - f(x)}{\delta}.$$

$= \inf_{\delta > 0}$ ''

Remark.

- If f is differentiable, then $f'(x;d) = \nabla f(x)^T d$.
- $f'(x;d) = \phi'(0^+)$, where $\phi(\alpha) = f(x+\alpha d)$.
- $f'(x;d) = \inf_{t>0}(tf(x+d/t) - tf(x))$ is convex in d (why?).
- $f'(x;d)$ defines a lower bound on f on direction d: $f(x+\alpha d) \geq f(x) + \alpha f'(x;d), \forall \alpha \geq 0$.

Descent Direction

Definition. The direction d is called a *descent direction* if

$$f'(x; d) < 0.$$

- If f is differentiable, then $d = -\nabla f(x)$ is a descent direction, except when it is zero.

Q. Is negative subgradient always a descent direction?

Descent Direction

- Negative subgradient may not be a descent direction.

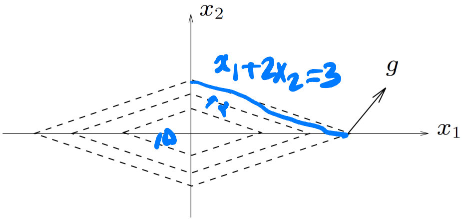


Figure: Contours of function $f(x_1, x_2) = |x_1| + 2|x_2|$

  - At $x = (1, 0)$, $\partial f(x) = \{(1, a) : a \in [-2, 2]\}$.
  - Consider $g = (1, 0)$, $d = -g$ is a descent direction.
  - Consider $g = (1, 2)$, $d = -g$ is not a descent direction.

- Note: let $g_* = \mathrm{argmin}_{g \in \partial f(x)} \{\|g\|_2^2\}$, then $d = -g_*$ is a descent direction if $g_* \neq 0$.

Directional Derivative and Subdifferential

Theorem. Let f be convex and $x \in \text{int}(\text{dom}(f))$, then

$$f'(x; d) = \max_{g \in \partial f(x)} g^T d$$

Proof

- Easy to show $f'(x;d) \geq \max_{g\in\partial f(x)} g^T d$.
- Suffice to show that $\exists \tilde{g} \in \partial f(x)$, s.t. $f'(x;d) \leq \tilde{g}^T d$.
  - Let $\tilde{g}$ be a subgradient of $f'(x;d)$ at d.
  - For any $v, \lambda \geq 0$:

$$\begin{aligned} f(x+\alpha v) - f(x) &\geq \alpha f'(x;v) \\ &= f'(x;\alpha v) \\ &\geq f'(x;d) + \tilde{g}^T(\alpha v - d). \end{aligned}$$

  - Setting $\alpha = \infty$ implies $f(x+v) - f(x) \geq f'(x;v) \geq \tilde{g}^T v$; thus $\tilde{g} \in \partial f(x)$.
  - Setting $\alpha = 0$ implies $f'(x;d) \leq \tilde{g}^T d$.

Calculus of Subgradients

Assume $x \in \text{int}(\text{dom}(h))$.

- Conic combination: Let $h(x) = \beta_1 f_1(x) + \beta_2 f_2(x)$ with $\beta_1, \beta_2 \geq 0$,

$$\partial h(x) = \beta_1 \partial f_1(x) + \beta_2 \partial f_2(x).$$

- Affine transformation: Let $h(x) = f(Ax + b)$,

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

- Pointwise maximum: Let $h(x) = \max_{i=1,\ldots,m} f_i(x)$,

$$\partial h(x) = \text{Conv}\left\{\partial f_i(x) | f_i(x) = h(x)\right\}.$$

- Pointwise supreme: Let $h(x) = \max_{\alpha \in \mathcal{A}} f_\alpha(x)$,

$$\partial h(x) = \text{cl}\left(\text{Conv}\left\{\partial f_\alpha(x) | f_\alpha(x) = h(x)\right\}\right).$$

Weak Calculus

- Maximization: $f(x) = \max_{y \in Y} \phi(x, y)$, where $\phi(x, y)$ is convex in x for any $y \in Y$.
  - Find $\hat{y} \in \operatorname{argmax}_{y \in Y} \phi(x, y)$.
  - $g \in \partial\phi(x, \hat{y})$ is a subgradient of $f(x)$.

- Minimization: $f(x) = \min_{y \in Y} \phi(x, y)$, where $\phi(x, y)$ is convex in (x, y) and Y is convex.
  - Find $\hat{y} \in \operatorname{argmin}_{y \in Y} \phi(x, y)$.
  - $g \in \partial\phi(x, \hat{y})$ is a subgradient of $f(x)$.

- Composition: $f(x) = F(f_1(x), \ldots, f_m(x))$, where $F(y_1, \ldots, y_m)$ is non-decreasing and convex.
  - Find $(d_1, \ldots, d_m) \in \partial F(y_1, \ldots, y_m)|_{y_i = f_i(x), i=1,\ldots,m}$.
  - Find $g_i \in \partial f_i(x), i = 1, \ldots, m$
  - $g = \sum_{i=1}^m d_i g_i$ is a subgradient of $f(x)$.

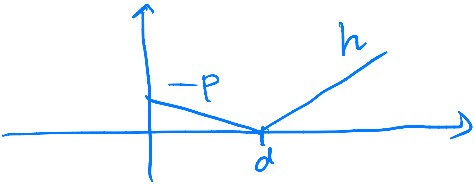
Example: Piecewise Linear Function

Example 3. Consider a single period inventory system. The cost $f(x)$ at inventory level x given demand d is

$$f(x) = h \cdot \max(x - d, 0) + p \cdot \max(d - x, 0).$$

The subgradient of $f(x)$ is

$$\partial f(x) = \begin{cases} h, & x > d \\ [-p, h], & x = d \\ -p, & x < d \end{cases}.$$



Example: ℓ_1-Norm

$= \sum s_i x_i$

Example 5. $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$

$= \sum |x_i|$

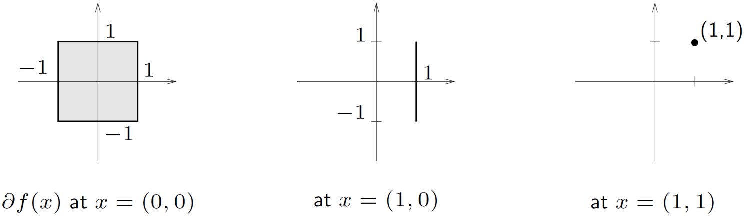


Figure: Subgradient of $f(x) = \|x\|_1$ on $\mathbb{R}^2$

Example: general norm

Example 6. $f(x) = \|x\|$, here $\|\cdot\|$ is an arbitrary norm

$$\partial f(x) = \{g : g^T x = \|x\| \text{ and } \|g\|_* \leq 1\}.$$

- $\|\cdot\|_*$ is the dual norm: $\|y\|_* = \max_{x:\|x\|\leq 1} y^T x$.
- In particular, $\partial f(0) := \{g : \|g\|_* \leq 1\}$.

Outline

Recap: Subgradient

- Subgradient and subdifferential
  - $g \in \partial f(x_0)$ if $f(x) \geq f(x_0) + g^T(x - x_0), \forall x$.
- Properties
  - Subdifferential is closed and convex.
  - Subgradient exists and is bounded at interior.
  - Subdifferential is a monotone operator.
- Directional derivative
  - $f'(x; d) = \max_{g \in \partial f(x)} g^T d$
- Calculus of Subgradients
  - Conic combination
  - Affine transformation
  - Point maximum/supreme
  - Taking minimization
  - Composition

Simple Examples

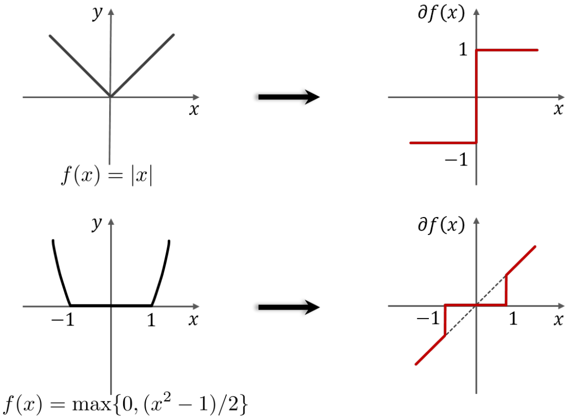


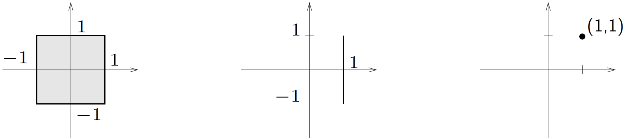
Figure: Examples of subdifferential sets

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$

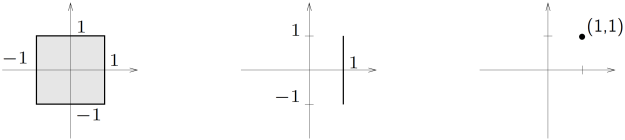


$\partial f(x)$ at $x = (0,0)$ | at $x = (1,0)$ | at $x = (1,1)$

Figure: Subgradient of $f(x) = \|x\|_1$ on $\mathbb{R}^2 (d = 2)$

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$



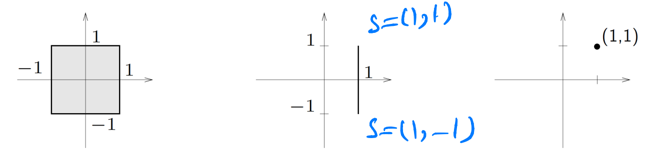
$\partial f(x)$ at $x = (0,0)$ | at $x = (1,0)$ | at $x = (1,1)$

Figure: Subgradient of $f(x) = \|x\|_1$ on $\mathbb{R}^2 (d = 2)$

Example . $f(x) = \|x\|_2 = \max_{s:\|s\|_2\le 1}\{s^T x\}$

Examples: ℓ_1-Norm and ℓ_2-Norm

Example . $f(x) = \|x\|_1 = \max_{s\in\{-1,1\}^d}\{s^T x\}$ $= s_1 x_1 + s_2 x_2$



$\partial f(x)$ at $x = (0,0)$ | at $x = (1,0)$ | at $x = (1,1)$

Figure: Subgradient of $f(x) = \|x\|_1$ on $\mathbb{R}^2 (d = 2)$

Example . $f(x) = \|x\|_2 = \max_{s:\|s\|_2\le 1}\{s^T x\}$

$$\partial f(x) = \begin{cases} \frac{x}{\|x\|_2}, & x \neq 0 \\ \{s : \|s\|_2 \le 1\}, & x = 0 \end{cases}.$$

Question

Which function below is different from others?

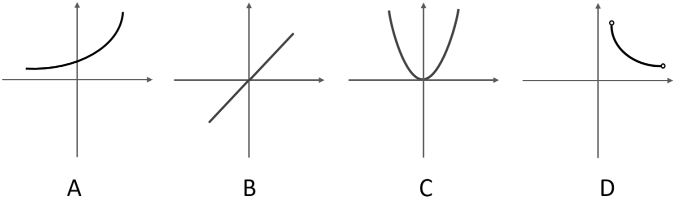


Figure: Convex functions

Existence of Global Minimizer

Definition. x^* is a *global minimizer* of $f(x)$ if

$$f(x^*) \leq f(x), \forall x.$$

Existence of Global Minimizer

Definition. x^* is a *global minimizer* of $f(x)$ if

$$f(x^*) \leq f(x), \forall x.$$

Definition. f is called *coercive* if all level sets are bounded, i.e., $f(x_k) \to \infty$ if $\|x_k\|_2 \to \infty$.

Existence of Global Minimizer

Definition. x^* is a *global minimizer* of $f(x)$ if

$$f(x^*) \leq f(x), \forall x.$$

Definition. f is called *coercive* if all level sets are bounded, i.e., $f(x_k) \to \infty$ if $\|x_k\|_2 \to \infty$.

Theorem. If f is closed (l.s.c.) and coercive, then it has a global minimizer.

Uniqueness of Global Minimizer

Recall f is strictly convex if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \forall \lambda \in (0,1), x \neq y.$$

- (sufficient condition): $\nabla^2 f(x) \succ 0$ (why?)

Uniqueness of Global Minimizer

Recall f is strictly convex if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \forall \lambda \in (0,1), x \neq y.$$

- (sufficient condition): $\nabla^2 f(x) \succ 0$ (why?)

Theorem. If f is strictly convex, then the global minimizer (if exists) must be unique.

Finding Global Minimizer

Theorem. Let f be convex. Then x^* is a global minimizer if and only if

$$0 \in \partial f(x^*).$$

Finding Global Minimizer

Theorem. Let f be convex. Then x^* is a global minimizer if and only if

$$0 \in \partial f(x^*).$$

- If f is convex and differentiable, x^* is a global minimizer iff $\nabla f(x^*) = 0$.

Finding Global Minimizer

Theorem. Let f be convex. Then x^* is a global minimizer if and only if

$$0 \in \partial f(x^*).$$

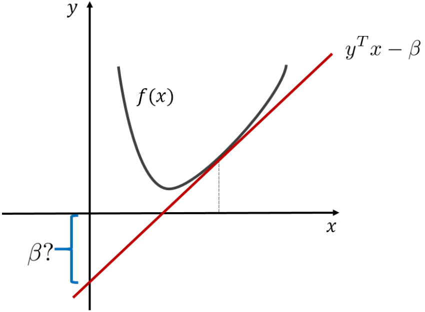
- If f is convex and differentiable, x^* is a global minimizer iff $\nabla f(x^*) = 0$.

Proof.

$$0 \in \partial f(x^*) \Leftrightarrow f(x) \geq f(x^*) + \langle 0, x - x^* \rangle = f(x^*), \forall x.$$

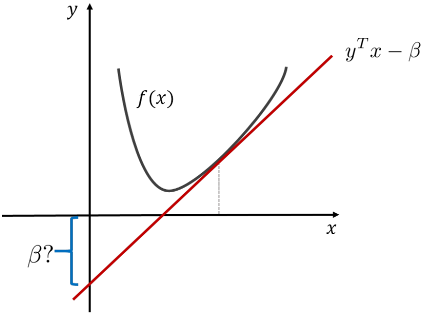
Question

For a given slope y, when is it an affine minorant?



Question

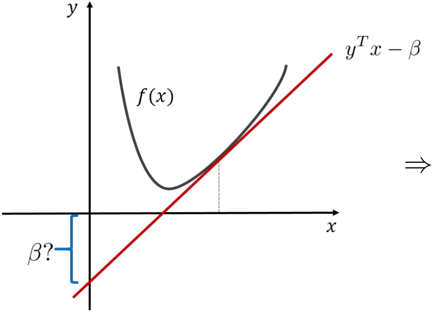
For a given slope y, when is it an affine minorant?



$$f(x) \geq y^T x - \beta, \forall x$$

Question

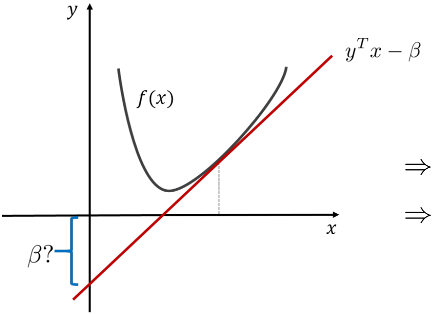
For a given slope y, when is it an affine minorant?



$$f(x) \geq y^T x - \beta, \forall x$$
$$\Rightarrow \beta \geq y^T x - f(x), \forall x$$

Question

For a given slope y, when is it an affine minorant?



$$f(x) \geq y^T x - \beta, \forall x$$
$$\Rightarrow \beta \geq y^T x - f(x), \forall x$$
$$\Rightarrow \beta \geq \sup_x \{y^T x - f(x)\}$$

Conjugate Function

Definition. The conjugate function of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\} = \sup_{x \in dom(f)} \left\{ y^T x - f(x) \right\}$$

Also called Legendre-Fenchel transformation.


A. Legendre (1752-1833)


Werner Fenchel (1905-1988)

Conjugate Function

Definition. The <u>conjugate function</u> of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\} = \sup_{x \in dom(f)} \left\{ y^T x - f(x) \right\}$$

Also called Legendre-Fenchel transformation.

Remark.

- **Fenchel's inequality:**

$$f(x) + f^*(y) \geq x^T y, \forall x, y$$


A. Legendre (1752-1833)


Werner Fenchel (1905-1988)

Conjugate Function

Definition. The conjugate function of $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\} = \sup_{x \in dom(f)} \left\{ y^T x - f(x) \right\}$$

Also called Legendre-Fenchel transformation.

Remark.

- **Fenchel's inequality:**

  $$f(x) + f^*(y) \geq x^T y, \forall x, y$$

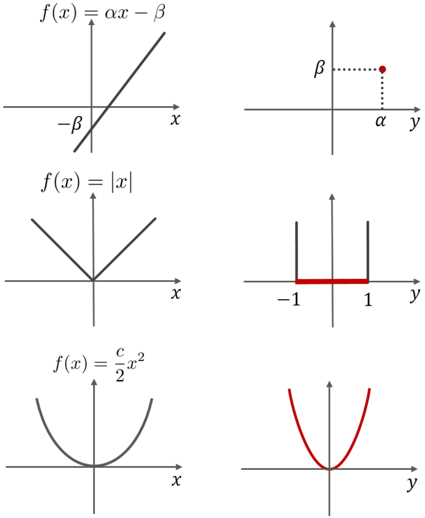
  $\frac{x^2}{2} + y^2/2$

- f^* is convex and closed.

A. Legendre (1752-1833)

Werner Fenchel (1905-1988)

Examples



$$f^*(y) = \begin{cases} \beta, & y = \alpha \\ +\infty, & y \neq \alpha \end{cases}$$

$$f^*(y) = \begin{cases} 0, & |y| < 1 \\ +\infty, & |y| > 1 \end{cases}$$

$$f^*(y) = \frac{1}{2c}y^2$$

Figure: Examples of conjugate functions

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^TQx +$ [illegible] $(Q \succ 0)$

$$\max_{x} \; y^Tx - \frac{1}{2}x^TQx$$

$$\nabla(\cdots) = y - Qx = 0 \rightarrow x = Q^{-1}y$$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^TQx \cancel{+ b^T x + c}$ $\quad (Q \succ 0)$

$$f^*(y) = \frac{1}{2}(x \cancel{- b})^T Q^{-1}(x \cancel{- b}) \cancel{- c}$$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

$\nabla = 0 \longrightarrow \tilde{y} = \left[1 + \log x_i \right] = y_i \longrightarrow$

$i \longrightarrow$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^TQx + b^Tx + c \quad (Q \succ 0)$

$$f^*(y) = \frac{1}{2}(x-b)^TQ^{-1}(x-b) - c$$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

$$f^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

More Examples

Example 1. Quadratic: $f(x) = \frac{1}{2}x^TQx + b^Tx + c \quad (Q \succ 0)$

$$f^*(y) = \frac{1}{2}(x-b)^TQ^{-1}(x-b) - c$$

Example 2. Negative entropy: $f(x) = \sum_{i=1}^n x_i \log(x_i)$

$$f^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

Example 3. Negative logarithm: $f(x) = -\sum_{i=1}^n \log(x_i)$

$$f^*(y) = -\sum_{i=1}^n \log(-y_i) - n$$

More Examples

Example 4. Indicator function: $I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$

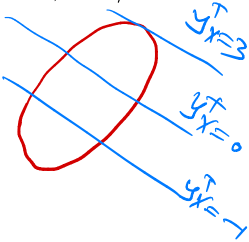
$\max_x \langle y, x \rangle - I_C(x) = \max_{x \in C} \langle y, x \rangle$

Example 5. Norm: $f(x) = \|x\|$

More Examples

Example 4. Indicator function: $I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$

$$I_C^*(y) = \sup_{x \in C} y^T x$$



Example 5. Norm: $f(x) = \|x\|$

$$\sup_x \langle x, y \rangle - \|x\|$$

$$= \sup_{x = z \cdot \|x\|} \|x\| \Big(\underbrace{\langle \frac{x}{\|x\|}, y \rangle - 1}_{\max = \|y\|_* - 1} \Big)$$

$z : \|z\| = 1$

More Examples

Example 4. Indicator function: $I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$

$$I_C^*(y) = \sup_{x \in C} y^T x$$

Example 5. Norm: $f(x) = \|x\|$

$$f^*(y) = \begin{cases} 0, & \|y\|_* \leq 1 \\ +\infty, & \|y\|_* > 1 \end{cases}$$

Calculus of Conjugate Functions

- **Separable sum**: If $g(x_1, x_2) = f_1(x_1) + f_2(x_2)$, then

$$g^*(y_1, y_2) = f_1^*(y_1) + f_2^*(y_2).$$

- **Scaling**: If $g(x) = \alpha f(x)$ with $\alpha > 0$, then

$$g^*(y) = \alpha f^*(y/\alpha).$$

$$g^*(y) = \sup_x \underbrace{y^T x - \alpha f(x)}_{\alpha\left(\left(\frac{y}{\alpha}\right)^T x - f(x)\right)}$$

- **Summation**: If $g(x) = f_1(x) + f_2(x)$, then

$$g^*(y) = \inf_z \{f_1^*(z) + f_2^*(y - z)\}$$

Biconjugate Function

- The conjugate of f is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\}$$

- The conjugate of the conjugate function $f^*(y)$,

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \left\{ x^T y - f^*(y) \right\}$$

Biconjugate Function

- The conjugate of f is

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - f(x) \right\}$$

- The conjugate of the conjugate function $f^*(y)$,

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \left\{ x^T y - f^*(y) \right\}$$

Q: is it true that $f^{**} = f$?