LECTURE 4: CONJUGATE TRANSFORMS

1. A preparation for dual information
2. Subgradients and subdifferentials
3. Conjugate transforms

Motivation

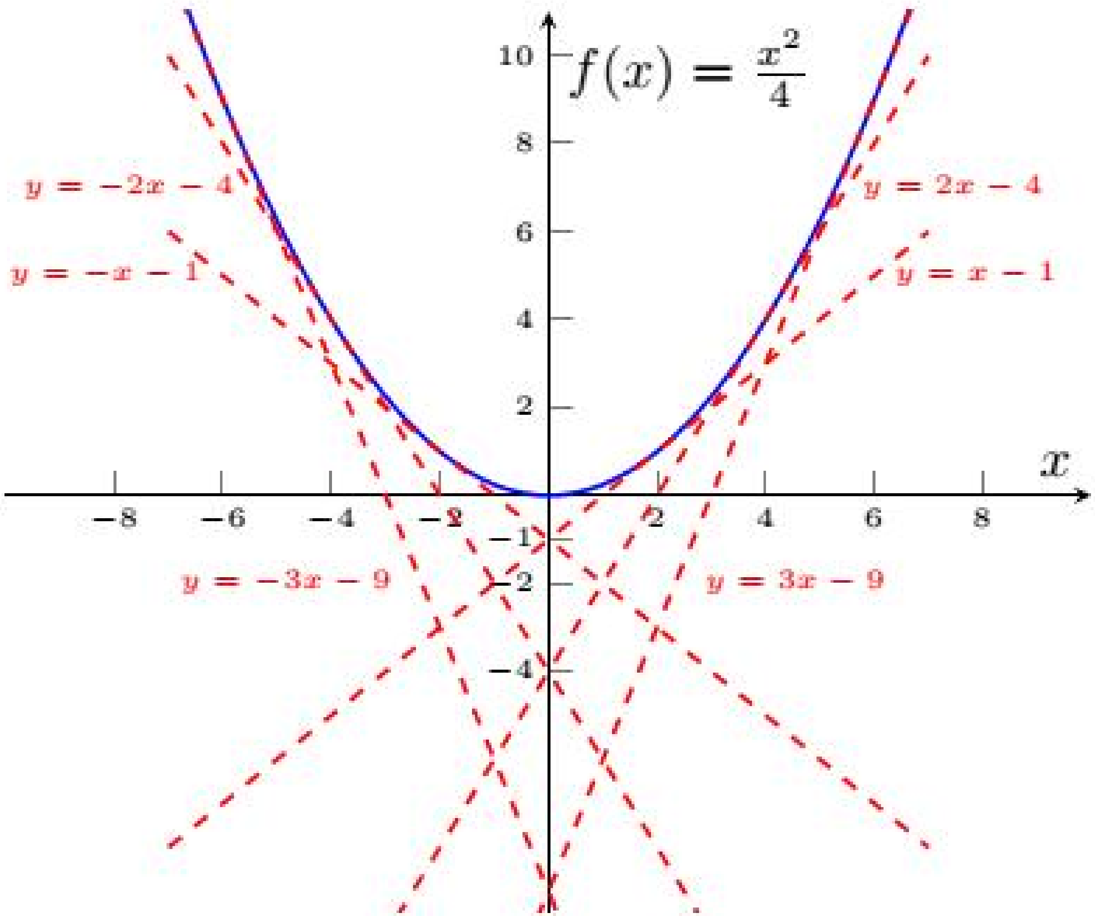


Figure: $(x,\ f(x) = \frac{x^2}{4}) \leftrightarrow (m,\ b(m) = -m^2) : y = mx + b$

Motivation

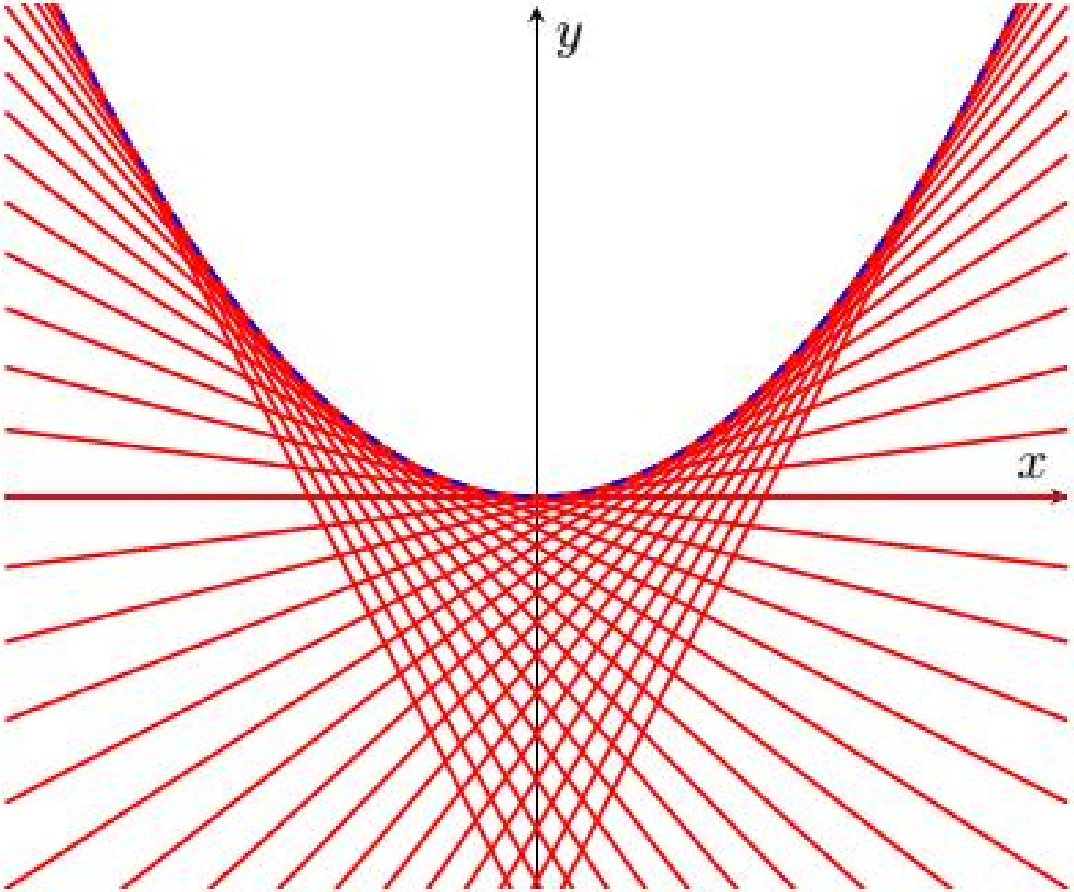


Figure: $(m, b(m) = -m^2) : y = mx + b \leftrightarrow (x,\ f(x) = \frac{x^2}{4})$

Motivation

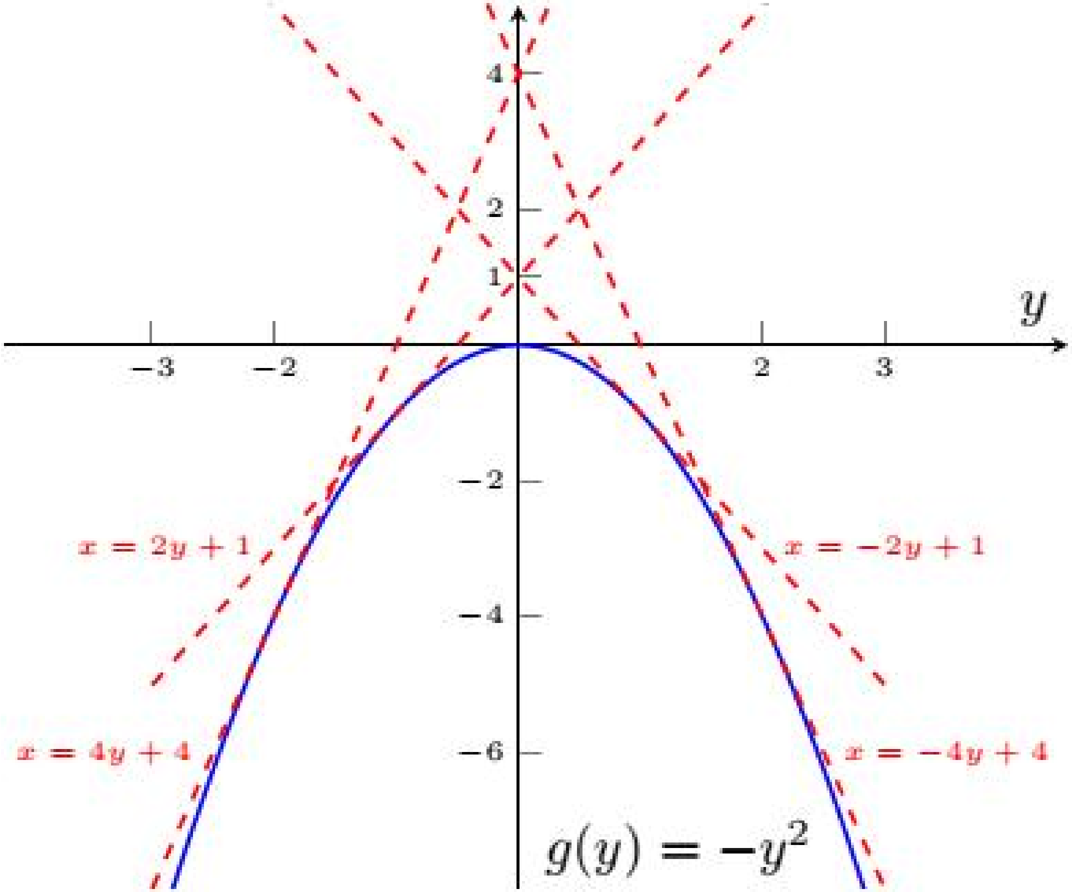


Figure: $(y,\ g(y) = -y^2) \rightarrow (m,\ b(m) = \frac{m^2}{4}) : x = my + b$

Motivation

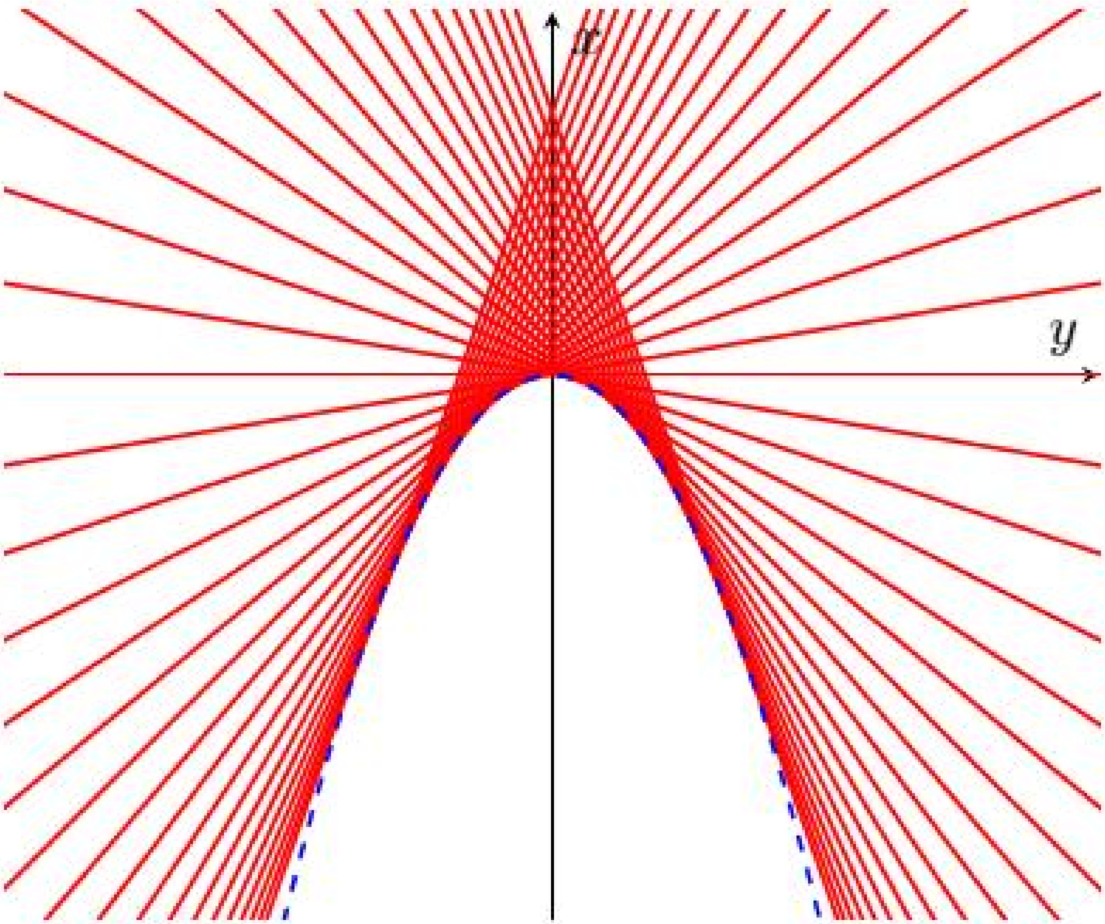
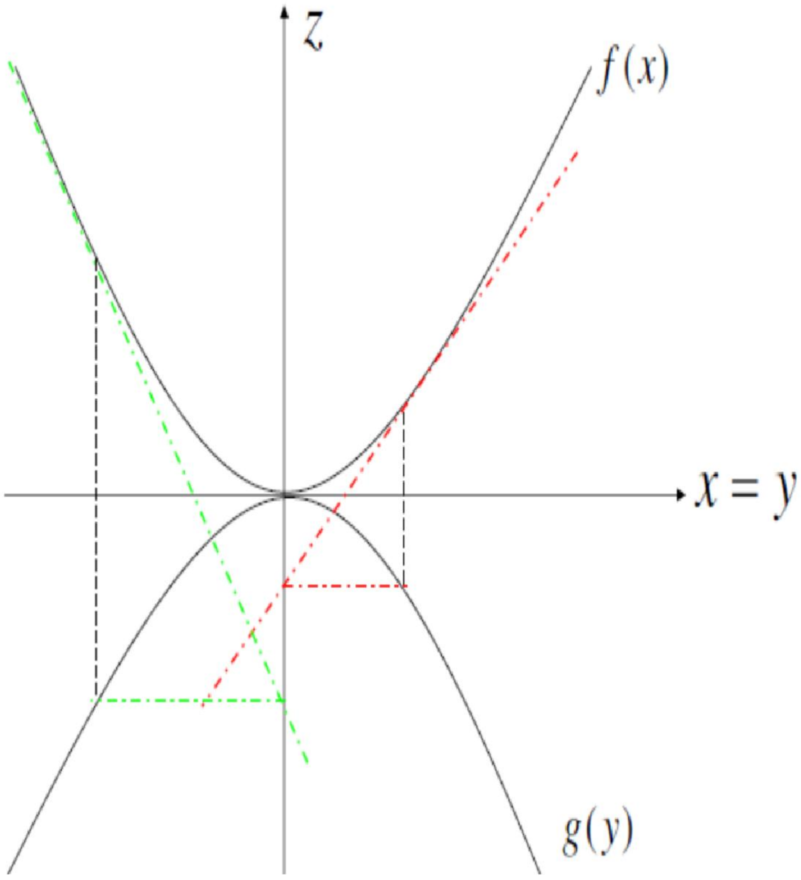


Figure: $(m,\ b(m) = \frac{m^2}{4}) : x = my + b \leftrightarrow (y,\ g(y) = -y^2)$

Where is the hidden dual information?

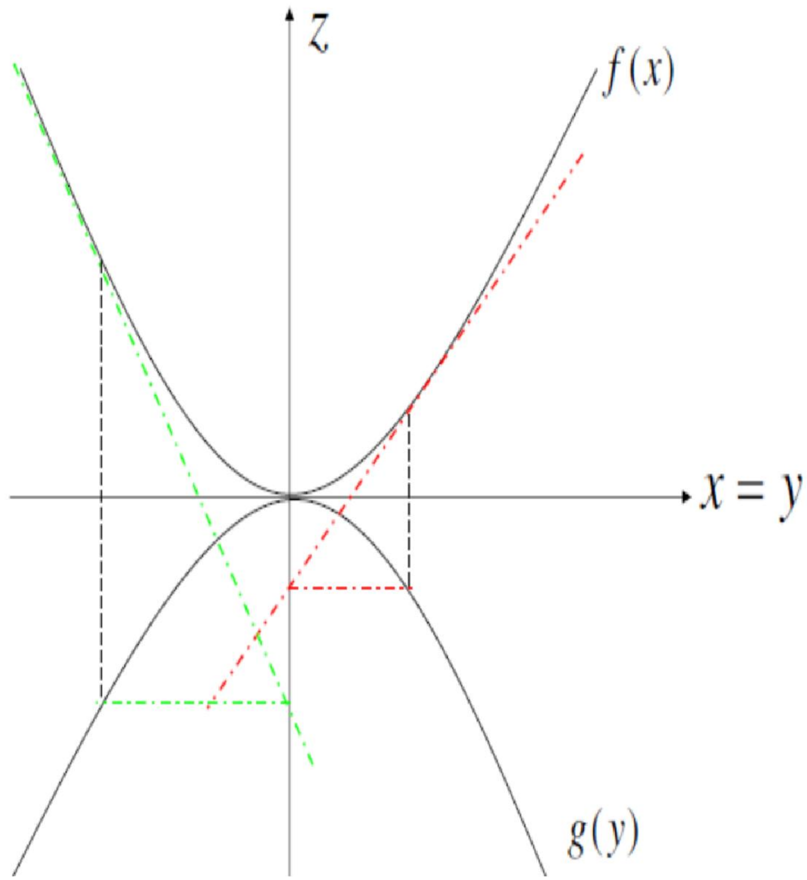
- Consider $f(x) = \frac{1}{2}x^2, \quad x \in R.$



Is there a function g(y) that tells another side of story about f(x) ? How?

Where is the hidden dual information?

- Secrete is in the conjugate transform



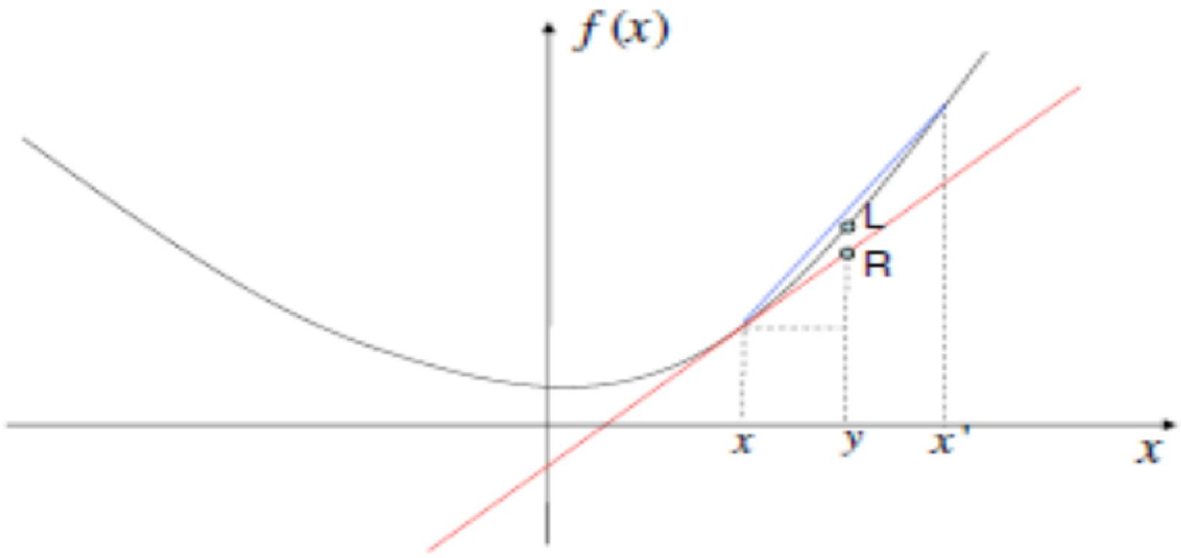
Conjugate:
Joined together, especially in a pair or pairs; coupled

Recall: Basic property - 3

- **Theorem:**

  Let $f \in C^1$. Then f is convex on a convex set $\Omega \subset E^n$ if, and only if,

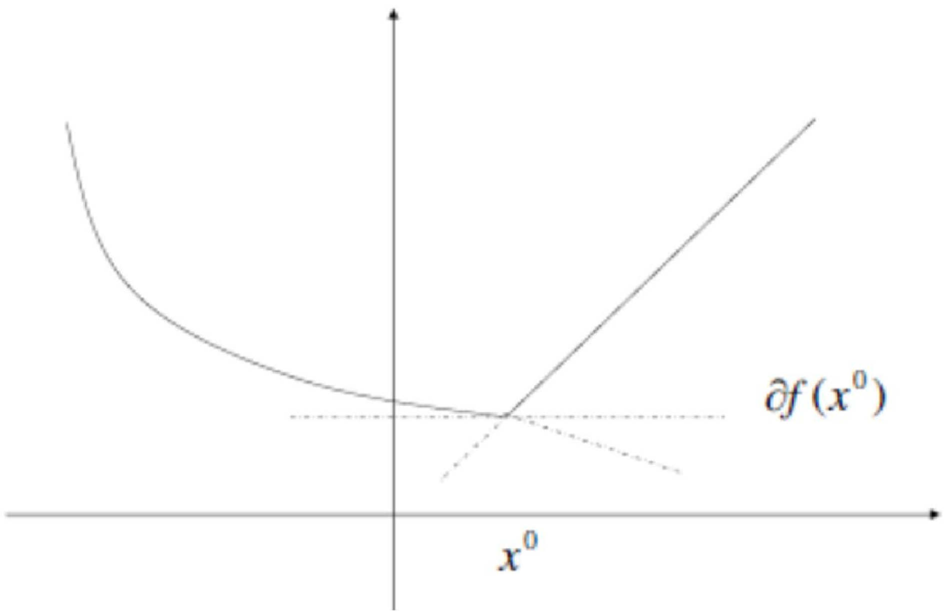
$$f(y) \geq f(x) + \nabla f(x)(y - x), \quad \forall\, x, y \in \Omega$$



(underestimate by one-point information)

Non-differentiable convex functions

- Where is the first order information when $f(x)$ is not differentiable?
  - subgradient and subdifferential



Subgradient and subdifferential

- Definition

  A vector *y* is said to be a subgradient of a convex function *f* (over a set *S*) at a point x^0 if

$$f(x) \geq f(x^0) + \langle y, x - x^0 \rangle, \forall x \in S$$

- Definition

  The set of all subgradients of *f* at x^0 is called the subdifferential of *f* at x^0 and is denoted by

$$\partial f(x^0) = \{ y \in E^n \mid f(x) \geq f(x^0) + \langle y, x - x^0 \rangle, \forall x \in S \}$$

Properties

1. The graph of the affine function

$$h(x) = f(x^0) + \langle y, x - x^0 \rangle$$

is a non-vertical supporting hyperplane to the convex set *epi*(*f*) at the point of (x^0, $f(x^0)$).

2. The subdifferential set $\partial f(x^0)$ is closed and convex.

3. $\partial f(x^0)$ can be empty, singleton, or a set with infinitely many elements. When it is not empty, *f* is said to be subdifferentiable at x^0.

4. $\nabla f(x^0) \in \partial f(x^0)$ if f is differentiable at x^0.

   $\{\nabla f(x^0)\} = \partial f(x^0)$ if f is convex and differentiable at $x^0 \in int(S)$.

Examples

- In R, $f(x) = |x|$ is subdifferentiable at every point and

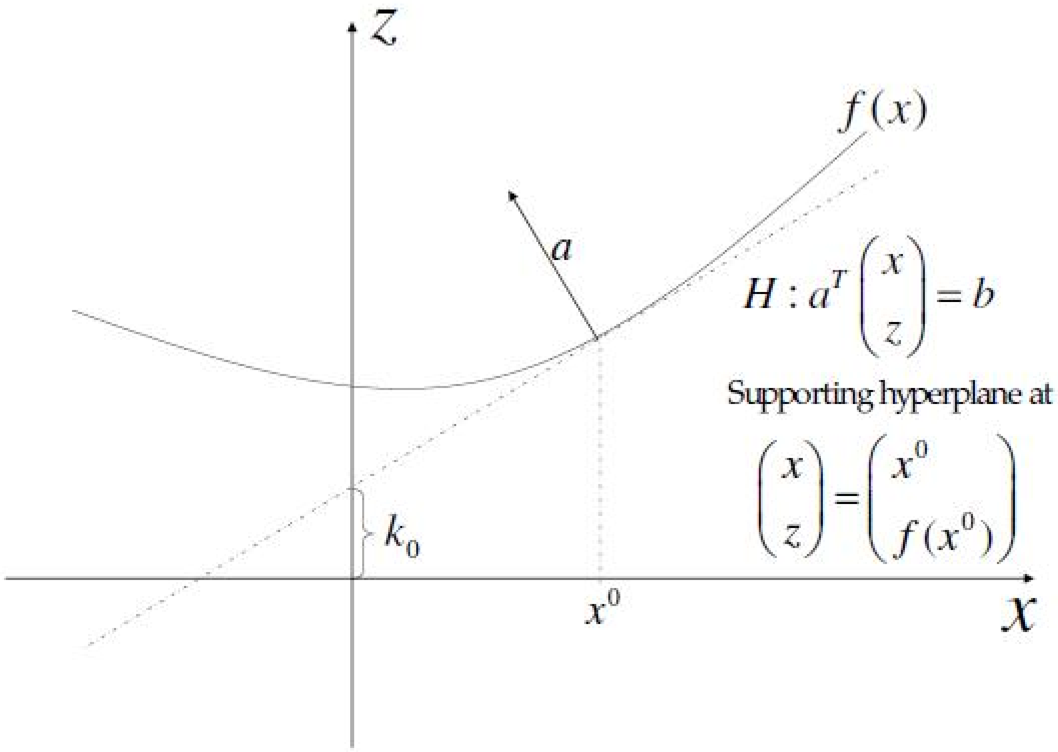
  $$\partial f(0) = [-1, 1].$$

- In E^n the Euclidean norm $f(x) = ||x||$ is subdifferentiable at every point and ∂f (0) consists of all the vectors y such that

  $$||x|| \geq \langle y, x \rangle \quad \text{for all } x.$$

  This means the Euclidean unit ball !

Conjugate Transformation



Let $f: S \subset E^n \rightarrow R$ be convex and $f \in C^1(S)$.
What's the supporting hyperplane at $(x^0, f(x^0))$?

Finding the supporting hyperplane

Intuitively, assume the hyperplane has the form of $z = \nabla f(x^0)x + k.$

Passing $x = 0$ with $z = k_0$ implies that

$$k = k_0.$$

Passing $x = x^0$ with $z = f(x^0)$ implies that

$$k_0 = f(x^0) - \nabla f(x^0)x^0.$$

Therefore, we have

$$\begin{aligned} z &= \nabla f(x^0)x + f(x^0) - \nabla f(x^0)x^0 \\ &= f(x^0) + \nabla f(x^0)(x - x^0) \end{aligned}$$

Finding the supporting hyperplane

or

$$(-\nabla f(x^0), 1) \begin{pmatrix} x \\ z \end{pmatrix} = f(x^0) - \nabla f(x^0) x^0.$$

This means the normal vector

$$a^T = (-\nabla f(x^0), 1),$$

and the right hand side coefficient

$$b = f(x^0) - \nabla f(x^0) x^0$$

Property 1 – upper half space

- If f is convex and $C^1(S)$ then

$$f(x) \;\geq\; f(x^0) + \nabla f(x^0)(x - x^0),\ \forall x \in S;$$

$$f(x) - \nabla f(x^0)x \;\geq\; f(x^0) - \nabla f(x^0)x^0,\ \forall x \in S;$$

$$(-\nabla f(x^0), 1)\begin{pmatrix} x \\ f(x) \end{pmatrix} \;\geq\; k_0.$$

Hence,

$$gra(f) \subset H^U. \quad \text{(upper half space of supporting hyperplane H)}$$

One application

- One product.
- Cost function $f(q)$ with $q \in S$ being the quantit; to produce.
- f is C^1 on S.
- At $q_0 \in S$, the marginal cost is $f'(q_0)$.
- Sell the product with unit price $p_0 = f'(q_0)$.
- If f is convex, then

$$f(q) \;\geq\; f(q_0) + \langle f'(q_0), q - q_0 \rangle, \quad \forall q \in S,$$

i.e., $\langle p_0, q_0 \rangle - f(q_0) \;\geq\; \langle p_0, q \rangle - f(q), \; \forall q \in S.$

- Maximum profit is achieved by selling q_0 products.

Property 2 – conjugate function

- Since

$$f(x) - \nabla f(x^0)x \;\geq\; f(x^0) - \nabla f(x^0)x^0, \;\forall x \in S,$$

$$\inf_{x \in S} \; [f(x) - \nabla f(x^0)x] \;=\; \underbrace{f(x^0) - \nabla f(x^0)x^0}_{k_0}$$

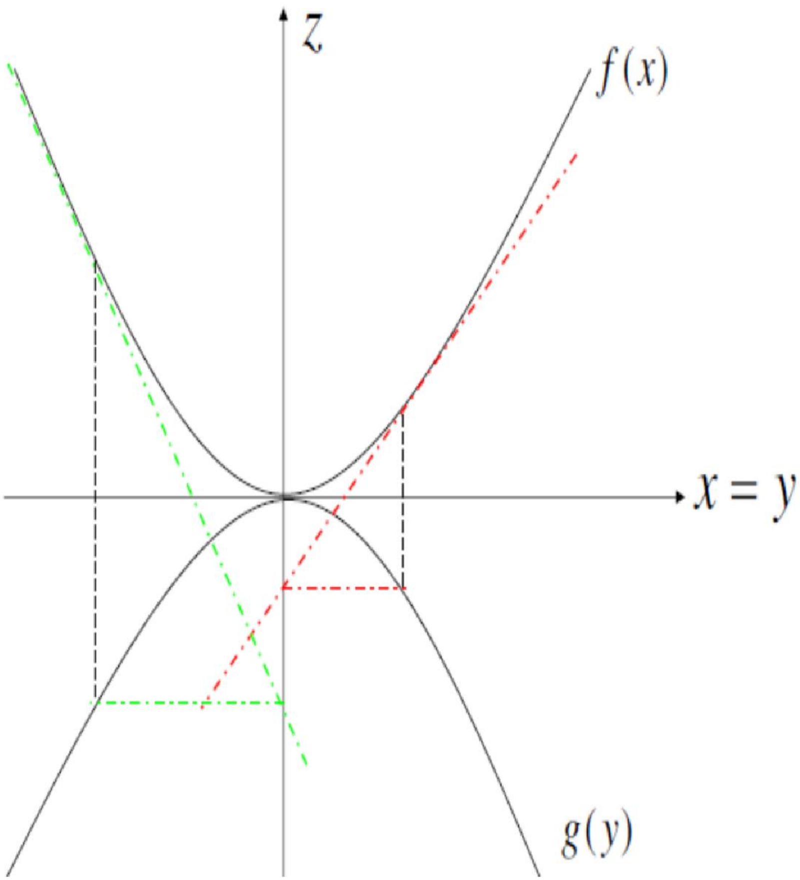
Let $y^0 \triangleq \nabla f(x^0)$. Then the function

$$g(y^0) \triangleq \inf_{x \in S} \; [f(x) - \langle x, y^0 \rangle]$$

takes value k_0.

Example – conjugate function

- Consider $f(x) = \frac{1}{2}x^2, \quad x \in R.$



$$Then \quad f'(x) = x, \ \ \forall x \in R$$

$$k_0 = f(x^0) - f'(x^0)x^0$$

$$= \frac{1}{2}(x^0)^2 - (x^0)^2$$

$$= -\frac{1}{2}(x^0)^2$$

$$\therefore g(y) = g(f'(x)) = g(x) = -\frac{1}{2}x^2 = -\frac{1}{2}y^2$$

Conjugate transform (function)

- **Definition**:

  Let $f : S \subset E^n \to R$. Its conjugate transform is a function h with domain

$$\Omega \triangleq \{y \in E^n \mid \operatorname*{Sup}_{x \in S} [\langle x, y \rangle - f(x)] < +\infty\}$$

and

$$h(y) \triangleq \operatorname*{Sup}_{x \in S} [\langle x, y \rangle - f(x)]$$
$$= -\inf_{x \in S} [f(x) - \langle y, x \rangle], \quad \forall y \in \Omega.$$

Conjugate transform of convex functions

- If $f \in C^1(S)$ is convex, then $\nabla f(x) \in \Omega, \forall x \in S$.

  Take

  $$y^0 = \nabla f(x^0) \quad \text{for some } x^0 \in S.$$

  Then

  $$\begin{aligned} h(y^0) &= -\inf_{x \in S} [f(x) - \nabla f(x^0)x\,] \\ &= -g(y^0) \\ &= -k_0 \\ &= \nabla f(x^0)x^0 - f(x^0) \end{aligned}$$

Special property – knowing each other

- $\langle x, y \rangle \;\le\; f(x) + h(y), \quad \forall x \in S,\; y \in \Omega.$

  Let $f \in C^1(S)$ be convex and $y^0 = \nabla f(x^0)$ for some $x^0 \in S$.

  Then

  $$h(y^0) \text{ is achieved at } x^0$$

  with

  $$\langle x^0, y^0 \rangle \;=\; f(x^0) + h(y^0).$$

Special property – knowing each other

- Similarly,

$$\text{if } \ y^0 \in \partial f(x^0) \ \text{ for some } x^0 \in S,$$

then

$$f(x) \ \geq \ f(x^0) + \langle y^0, x - x^0 \rangle, \ \ \forall x \in S,$$

i.e.,

$$\langle y^0, x^0 \rangle - f(x^0) \ \geq \ \langle y^0, x \rangle - f(x), \ \ \forall x \in S.$$

Hence

$$y^0 \in \Omega \quad \text{and}$$

$$h(y^0) \ = \ \langle y^0, x^0 \rangle - f(x^0),$$

or

$$\langle x^0, y^0 \rangle \ = \ f(x^0) + h(y^0).$$

Property of conjugate transforms

- Theorem:

  Given a function $f : S \subset E^n \to R$, its conjugate transform h, when exists, is a convex function defined on a convex domain Ω.

  Proof: Routine exercise.

Main Theorem

- <u>**Theorem**</u>:

  (Conjugate Inequality / Fenchel's Inequality)

  Let $f : S \subset E^n \rightarrow R$ with its conjugate transform h over Ω.

  Then

  $$\langle x, y \rangle \ \le \ f(x) + h(y), \quad \forall x \in S \text{ and } y \in \Omega.$$

  Moreover,

  $$\langle x, y \rangle \ = \ f(x) + h(y) \quad \text{iff} \quad y \in \partial f(x).$$

Proof

<u>Proof</u>:

We only need to show the necessity part of the equality case.

Suppose that $\langle x^0, y^0 \rangle = f(x^0) + h(y^0)$.

Then

$$\begin{aligned} h(y^0) &= \operatorname*{Sup}_{x \in S} \left[\langle x, y^0 \rangle - f(x) \right] \\ &= \langle x^0, y^0 \rangle - f(x^0). \end{aligned}$$

Hence

$$\langle x, y^0 \rangle - f(x) \leq \langle x^0, y^0 \rangle - f(x^0), \quad \forall x \in S.$$

$$\therefore f(x) \geq f(x^0) + \langle x - x^0, y^0 \rangle, \quad \forall x \in S.$$

This implies $y^0 \in \partial f(x^0)$.

Legendre transform

- **Definition**: The Legendre Transform of $f : S$ is the "restriction"

$$h : \bigcup_{x \in S} \partial f(x)$$

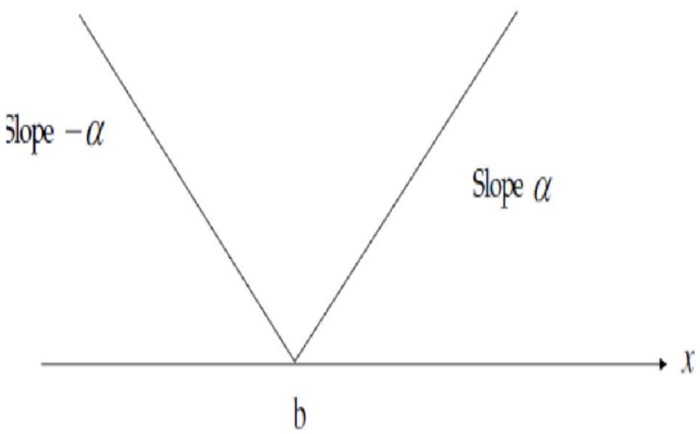
of $h : \Omega$

- In this case, for $y = \nabla f(x)$, we may write

$$\begin{aligned} h(y) &= \langle y, x \rangle - f(x) \\ &= \sum x_i \frac{\partial f}{\partial x_i}(x) - f(x). \end{aligned}$$

Example 1

$$f(x) = \alpha\,|\,x - b\,|,\ \ \alpha > 0,\ x \in S = R$$



$$h(y) \triangleq \sup_{x\in R}\left[\ xy - \alpha|x-b|\ \right]$$

$$= \sup_{z\in R}\left[\ (z+b)y - \alpha|z|\ \right]$$

$$= \sup_{z\in R}\left[\ yz - \alpha|z|\ \right] + by$$

$$\therefore h(y) = \underbrace{by}_{\text{linear function}} \ \ \text{on}\ \ \Omega = [-\alpha, \alpha]$$

Example 2

$$f(x) = \sum_{i=1}^{n} c_i e^{x_i}, \quad c_i > 0, \; S = E^n.$$

$$h(y) \triangleq \sup_{x \in E^n} \left[\langle x, y \rangle - \sum_{i=1}^{n} c_i e^{x_i} \right]$$

$$= \sup_{x \in E^n} \left[\sum_{i=1}^{n} \left(x_i y_i - c_i e^{x_i} \right) \right]$$

$$= \sum_{i=1}^{n} \underbrace{\sup_{x_i \in R}}_{\text{unrestricted}} \underbrace{\left[x_i y_i - c_i e^{x_i} \right]}_{\text{concave function}}$$

Set

$$y_i - c_i e^{x_i} = 0$$

$$\Rightarrow \quad y_i = c_i e^{x_i} > 0$$

$$\Rightarrow \quad x_i = \log_e\left(\frac{y_i}{c_i}\right)$$

Hence,

$$h(y) = \sum_{i=1}^{n} \left[y_i \log\left(\frac{y_i}{c_i}\right) - y_i \right]$$

$$= \sum_{i=1}^{n} y_i \log\left(\frac{y_i}{c_i}\right) - \sum_{i=1}^{n} y_i, \quad \forall y > 0.$$

Example 2 - continue

Also, when $y_i = 0$,

$$\sup_{x \in R} \left[-c_i e^{x_i} \right] = 0$$

i.e.,

$$h(y) = 0 \quad \text{for} \quad y = 0.$$

Take $0 \cdot \log 0 \triangleq 0$. We have

$$\Omega = \{ y \in E^n \mid y_i \geq 0, \ \forall i \} \text{ and}$$

$$h(y) = \underbrace{\sum_{i=1}^{n} y_i \log(\frac{y_i}{c_i})}_{\text{cross-entropy functoin}} - \underbrace{\sum_{i=1}^{n} y_i}_{\text{linear function}} .$$

More examples

For $f(x) \triangleq \sum_{j=1}^{n} x_j \log(\frac{x_j}{c_j})$ with $c_j > 0$

on $S \triangleq \{x \in E^n \mid x_j \geq 0 \text{ and } \sum_{j=1}^{n} x_j = 1\}$,

then

$$\underbrace{h(y) = \log(\sum_{j=1}^{n} c_j e^{y_j})}_{\text{log exponential function}} \quad \text{on} \quad \Omega = E^n.$$

More examples

For $f(x) \triangleq \sum_{j=1}^{n-1} p_j^{-1}|x_j - b_j|^{p_j} + x_n - b_n$

with $p_j > 1$ on $S = E^n$,

then

$$h(y) = \sum_{j=1}^{n-1} (q_j^{-1}|y_j|^{q_j} + b_j y_j) + b_n$$

on $\Omega = \{y \in E^n \mid y_n = 1\}$

where q_i is defined by $\frac{1}{p_i} + \frac{1}{q_i} = 1$.

$\longrightarrow$ norm change.

More examples

- Example:

  For $$\begin{aligned} f(x) &= d \quad \text{on} \quad S = \{c\} \subset E^n, \\ h(y) &= c^T y - d \quad \text{on} \quad \Omega = E^n. \end{aligned}$$

- Example:

  For $$\begin{aligned} f(x) &= c^T x - d \quad \text{on} \quad S = E^n, \\ h(y) &= d \quad \text{on} \quad \Omega = \{c\}. \end{aligned}$$

More examples

- Example:

  For $f(x) = c^T x - d$ on $S = E^n_+$,

  $$h(y) = d \quad \text{on} \quad \Omega = \{y \in E^n \mid y \leq c\}.$$

- Example:

  For $f(x) = cx - d$ on $S = (a, b) \subset R$

  $$h(y) = \begin{cases} a(y-c)+d & \text{for } y \leq c \\ b(y-c)+d & \text{for } y \geq c \end{cases}$$

  on $\Omega = R$.

Constructing conjugate transforms

Let $f : S \subset E^n \to R$ be a function with its conjugate transform $h : \Omega$.

(1) For $\alpha \in R$, the conjugate of $f + \alpha : S$ is $h - \alpha : \Omega$.

(2) For $a \in E^n$, if $\tilde{f}(x) \triangleq f(x) + \langle x, a \rangle : S$, then

$$\tilde{h}(y) = \underbrace{h(y-a)}_{\text{translation of } h(\cdot)}, \quad \forall y \in \Omega + a.$$

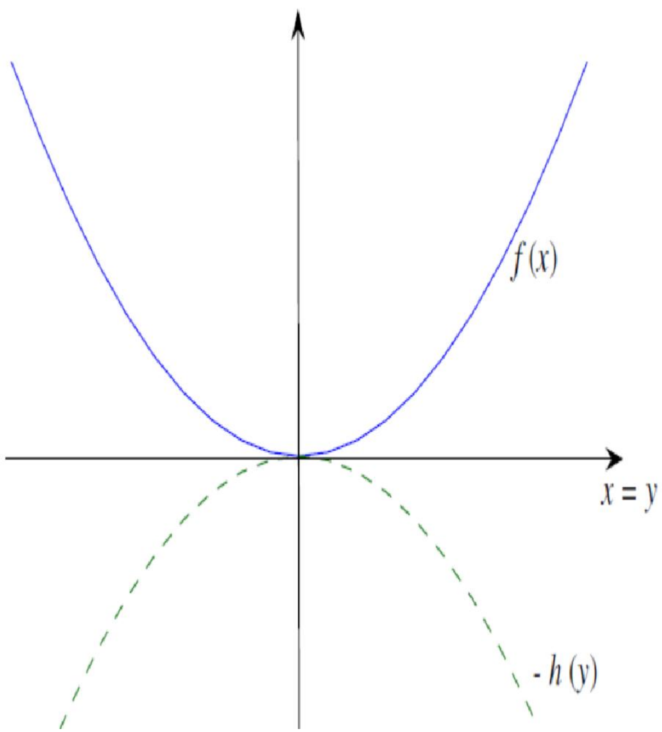
(3) For $a \in E^n$, if $\bar{f}(x) \triangleq f(x-a) : S + a$, then

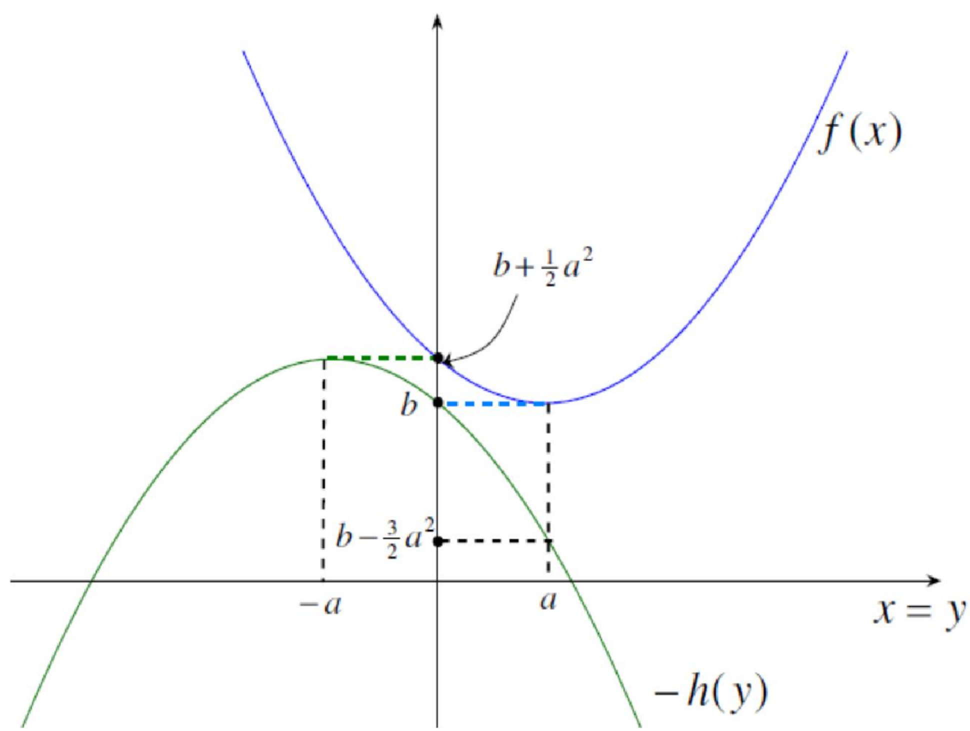
$$\bar{h}(y) = h(y) + \underbrace{\langle a, y \rangle}_{\text{linear function}}, \quad \forall y \in \Omega.$$

Illustration

$$\begin{cases} f(x) = \frac{1}{2}x^2 & \text{for } x \in R \\ h(y) = \frac{1}{2}y^2 & \text{for } y \in R \end{cases}$$

$$\begin{cases} f(x) = \frac{1}{2}(x-a)^2 + b\ , & \text{for } x \in R \\ h(y) = \frac{1}{2}y^2 + ay - b\ , & \text{for } y \in R \end{cases}$$





Constructing conjugate transforms

Let $f : S$ have conjugate transform $h : \Omega$.

(1) For $\lambda > 0$, $f_1(x) \triangleq \lambda f(x)$ on S. Then $h_1 : \lambda\Omega$ with $h_1(y) = \lambda h(\frac{y}{\lambda})$.

(2) For $\lambda > 0$, $f_2(x) \triangleq f(\frac{x}{\lambda})$ on λS. Then $h_2 : \Omega/\lambda$ with $h_2(y) = h(\lambda y)$.

Question

- Assume $f : S$ has conjugate transform $h : \Omega$. When will the conjugate transform of $h : \Omega$ become $f : S$?

A hint

- Recall that

- **Theorem**:

  Given a function $f : S \subset E^n \to R$, its conjugate transform h, when exists, is a convex function defined on a convex domain Ω.

  Proof: Routine exercise.

Proper and closed functions

- **Definition**:

  A *proper* function on E^n is a function obtained by taking a finite convex function f on a non-empty convex set S and then extending it to E^n by setting $f(x) = +\infty$, $\forall\ x \notin S$.

- **Definition**:

  A function $f : S \subset E^n \to R$ is a closed function if $epi(f)$ is a closed subset of E^{n+1}.

Properties of conjugate transforms

- **<u>Observation</u>:**

  If S is closed and $f \in C(S)$, then f is a closed function.

- **<u>Theorem</u>:**

  When it exists, the conjugate transform $h : \Omega$ is closed.

Properties of conjugate transforms

- <u>**Theorem**</u>:

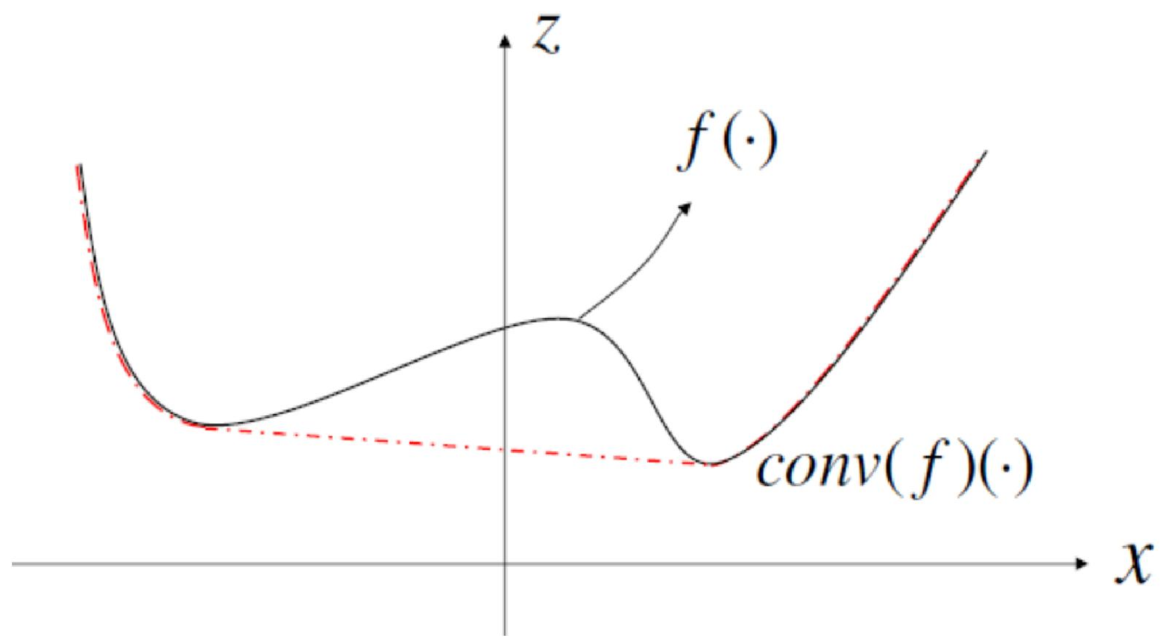
  Let $f : S \subset E^n \to R$ be a proper closed convex function with conjugate transform $h : \Omega$. Then the conjugate transform of $h : \Omega$ is $f : S$. Moreover, $y \in \partial f(x)$ iff $x \in \partial h(y)$. In this case, $\langle x, y \rangle = f(x) + h(y)$, iff

  $$y \in \partial f(x) \text{ or } x \in \partial h(y).$$

Proof: See Fang and Xing's "Linear Conic Optimization". Theorem 2.38.

Convex hull function

Given a function $f : S \subset E^n \to R$, its <u>convex hull function</u> $conv(f)$ is a function on S such that $epi(\ conv(f)\ ) = co(\ epi(f)\ )$.



Properties of conjugate transforms

- **Theorem**:

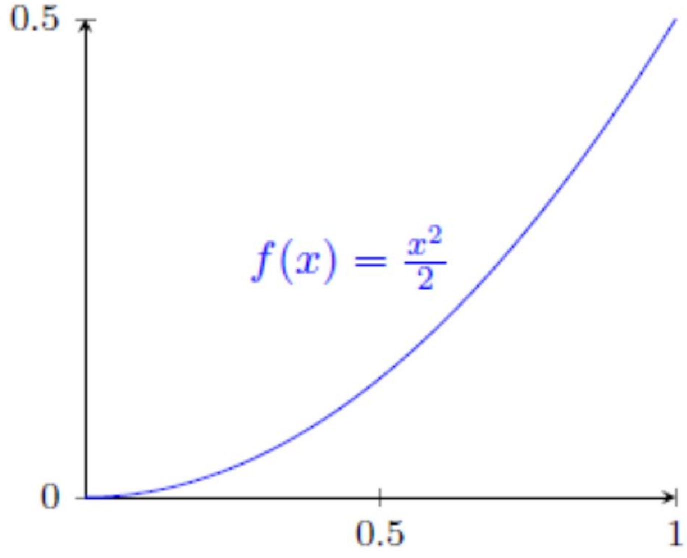
  Assume $f_1 : S$ and $f_2 : S$ have the same convex hull function. Then they have the same conjugate transform $h : \Omega$, when it exists.

  Proof: See Fang and Xing's "Linear Conic Optimization". Theorem 2.39.

Examples

Example 1:

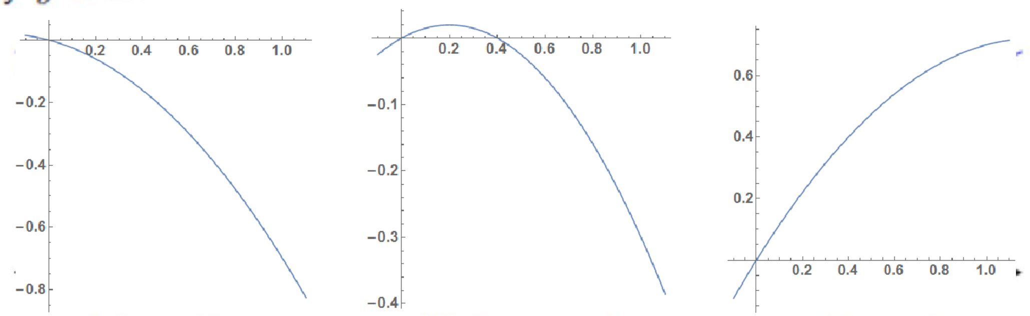
Consider a convex function $f(x) = \frac{1}{2}x^2$ over $S = [0,\ 1]$.



Examples

Example 1 - Cont'd

The conjugate function $h(y) = \sup_{x \in [0,\ 1]} \left\{ xy - \frac{x^2}{2} \right\} < +\infty$ for any $y \in E$.



(a) $y < 0$ (b) $0 \le y \le 1$ (c) $y > 1$

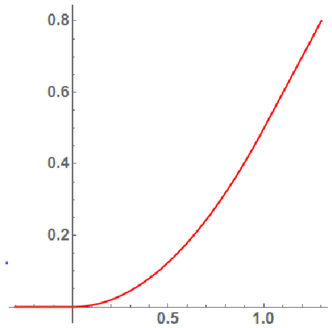
Figure : Plot of $xy - \frac{x^2}{2}$ v.s. x for different cases of y

Hence, $\Omega = E$ and

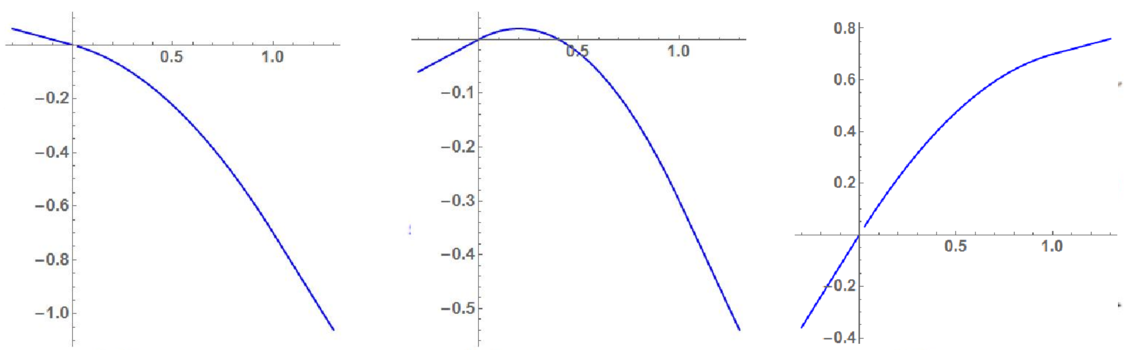
$$h(y) = \begin{cases} 0, & y < 0, \\ y^2/2, & 0 \le y \le 1, \\ y - 1/2, & y > 1. \end{cases}$$

Examples

Example 1 - Cont'd



The conjugate of $h(y)$ is $\tilde{f}(x) = \sup_{y \in E} \{xy - h(y)\}$.



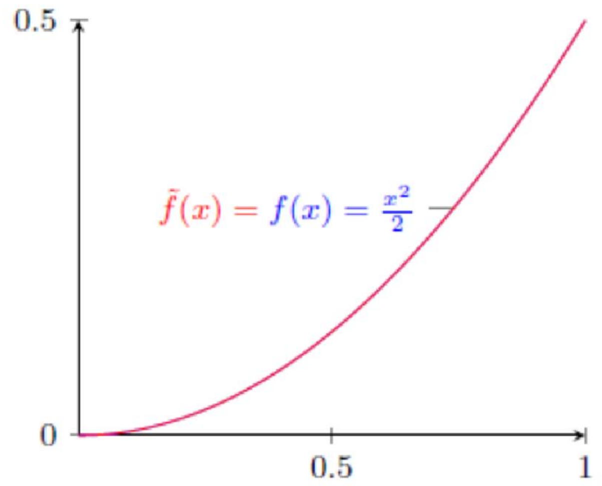
(a) $x < 0$ (b) $0 \leq x \leq 1$ (c) $x > 1$

Figure : Plot of $xy - h(y)$ v.s. y for different cases of x

Examples

Example 1 - Cont'd

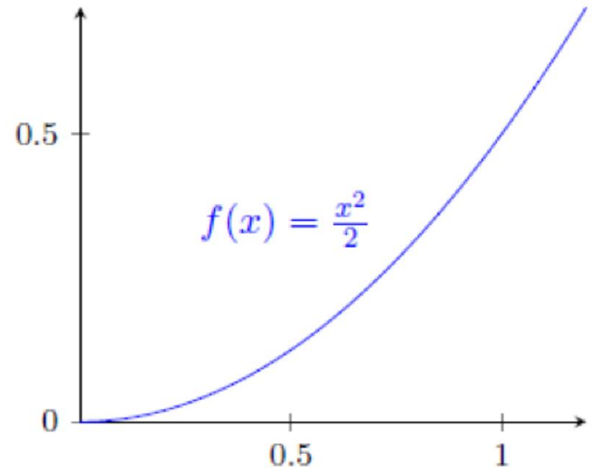
$\tilde{f}(x) < +\infty$ only when $0 \leq x \leq 1$. Hence, $\tilde{\Omega} = [0,\ 1]$, and the supreme is achieved at $y = x$. Thus, $\tilde{f}(x) = \frac{1}{2}x^2$.



Examples

Example 2

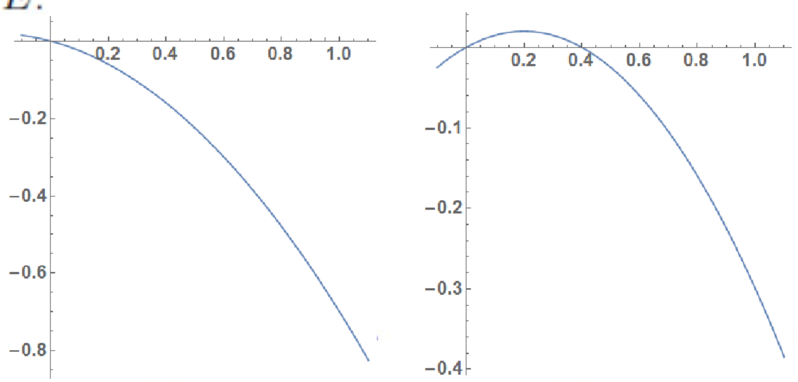
Consider a convex function $f(x) = \frac{1}{2}x^2$ over $S = [0, \ +\infty)$.



Examples

Example 2 - Cont'd

The conjugate function $h(y) = \sup_{x \in [0,\ +\infty)} \left\{ xy - \frac{x^2}{2} \right\} < +\infty$ for any $y \in E$.



(a) $y < 0$ (b) $y \geq 0$

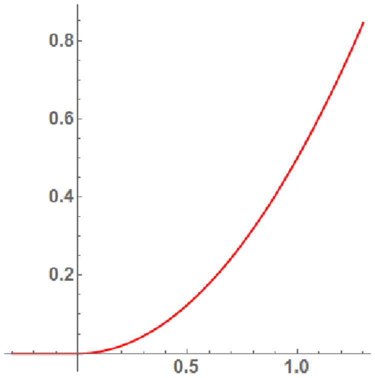
Figure : Plot of $xy - \frac{x^2}{2}$ v.s. x for different cases of y

Hence, $\Omega = E$ and

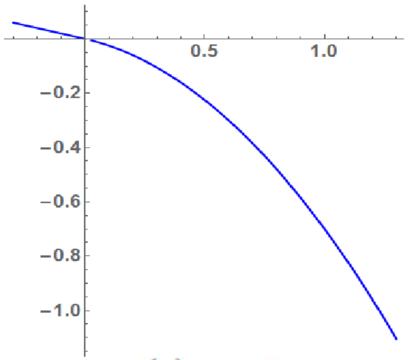
$$h(y) = \begin{cases} 0, & y < 0, \\ y^2/2, & y \geq 0. \end{cases}$$

Examples

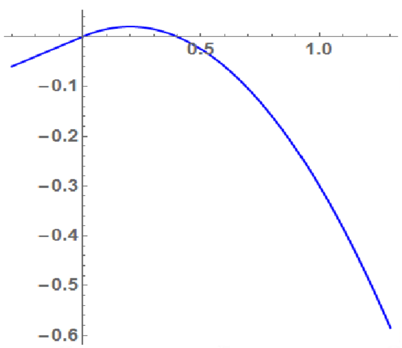
Example 2 - Cont'd



The conjugate of $h(y)$ is $\tilde{f}(x) = \sup_{y \in E} \{xy - h(y)\}$.



(a) $x < 0$



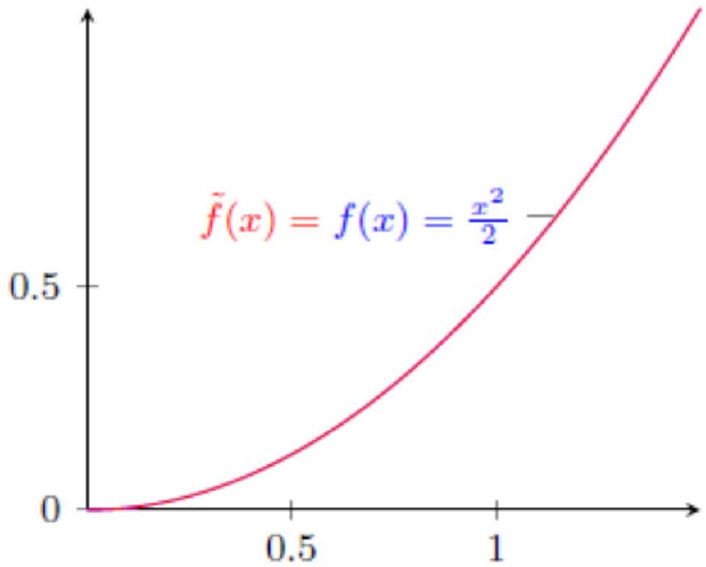
(b) $x \geq 0$

Figure : Plot of $xy - h(y)$ v.s. y for different cases of x

Examples

Example 2 - Cont'd

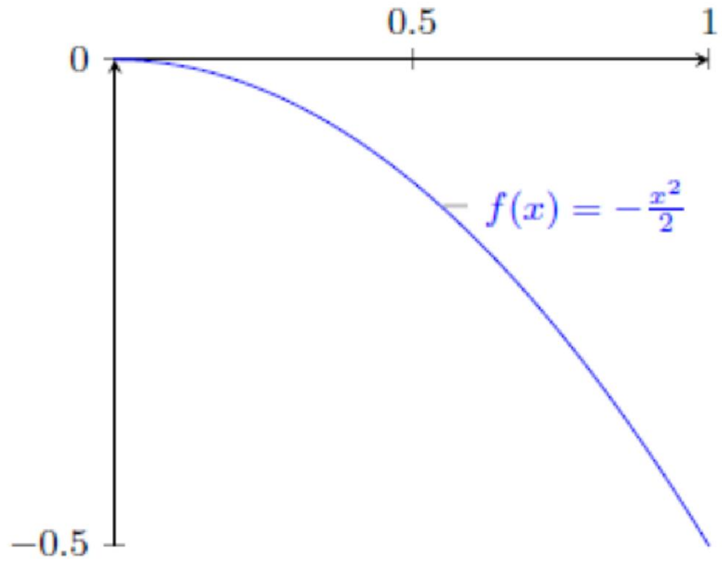
$\tilde{f}(x) < +\infty$ only when $x \geq 0$. Hence, $\tilde{\Omega} = [0, \ +\infty)$, and the supreme is achieved at $y = x$. Thus, $\tilde{f}(x) = \frac{1}{2}x^2$.



Examples

Example 3

Consider a concave function $f(x) = -\frac{1}{2}x^2$ over $S = [0, 1]$.

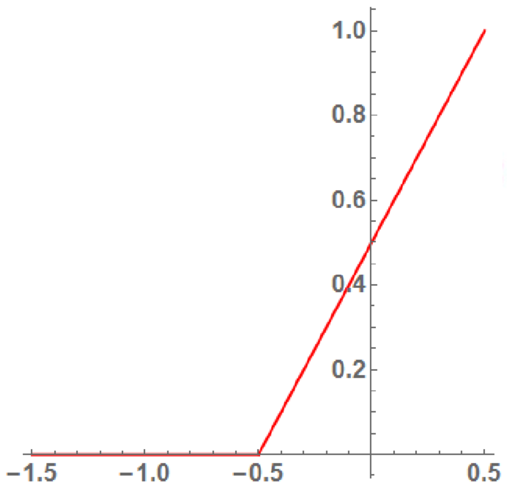


Examples

Example 3 - Cont'd

The conjugate function $h(y) = \sup_{x\in[0,\ 1]} \left\{xy + \frac{x^2}{2}\right\}$ involves a maximization problme of a convex function over a compact set, whose supreme is achieved at either $x = 0$ or $x = 1$ for any $y \in E$. Hence, $\Omega = E$ and

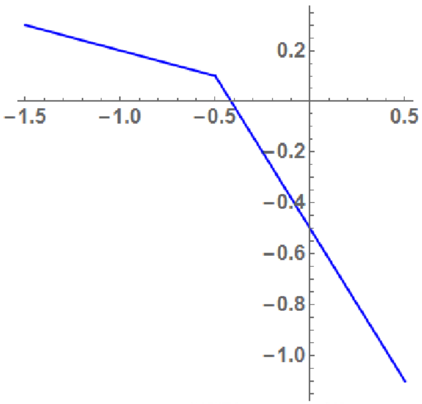
$$h(y) = \begin{cases} 0, & y < -1/2, \\ y + 1/2, & y \geq -1/2. \end{cases}$$



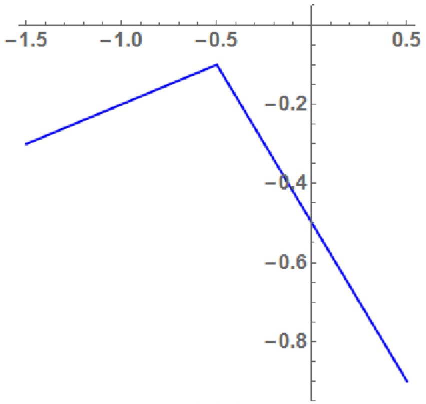
Examples

Example 3 - Cont'd

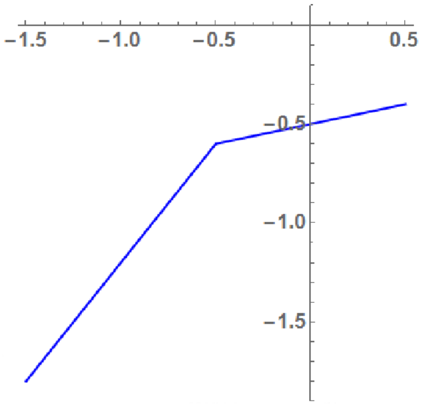
The conjugate of $h(y)$ is $\tilde{f}(x) = \sup_{y \in E} \{xy - h(y)\}$.



(a) $x < 0$



(b) $0 \leq x \leq 1$



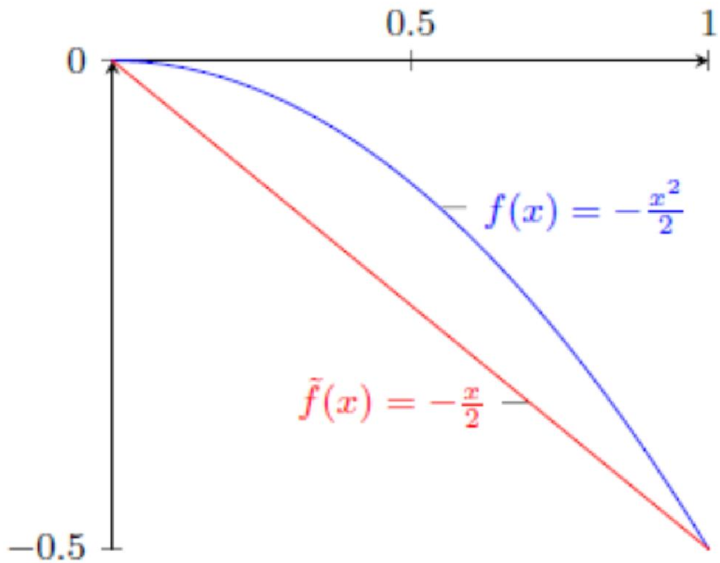
(c) $x > 1$

Figure : Plot of $xy - h(y)$ v.s. y for different cases of x

Examples

Example 3 - Cont'd

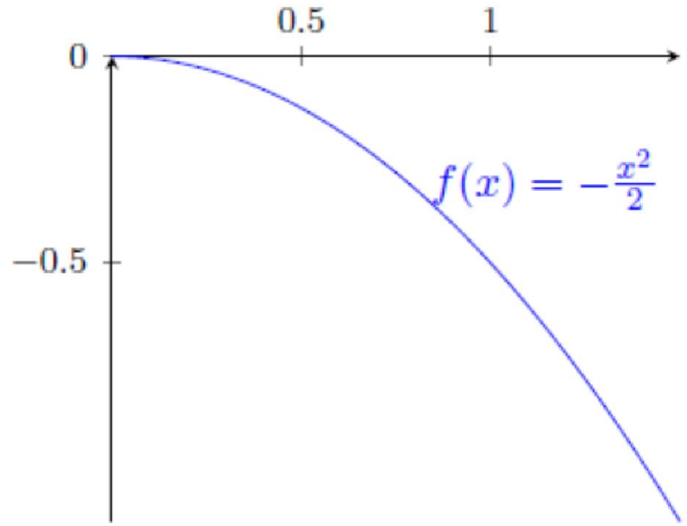
$\tilde{f}(x) < +\infty$ only when $0 \leq x \leq 1$. Hence, $\tilde{\Omega} = [0,\ 1]$, and the supreme is achieved at $y = -0.5$. Thus, $\tilde{f}(x) = -\frac{1}{2}x$.



Examples

Example 4

Consider a concave function $f(x) = -\frac{1}{2}x^2$ over $S = [0, +\infty)$.



Since $\sup_{x \in [0,\ +\infty)} \left\{ xy + \frac{x^2}{2} \right\} = +\infty$ for every $y \in E$, $\Omega = \emptyset$ and the conjugate function $h(y)$ does not exists.