Lecture Note 2: Convex function

Xiaoqun Zhang
Shanghai Jiao Tong University

Last updated: March 15, 2017
1.1 Convex functions

- Definition on one variable: $f$ defined on a nonempty interval $I$ is said to be convex on $I$ if
  
  $$f(\alpha x + (1 - \alpha)x') \leq \alpha f(x) + (1 - \alpha)f(x')$$

  for all pairs of points $(x, x')$ in $I$ and all $\alpha \in (0, 1)$.

- Criterion of increasing slopes: A function $f$ is convex on an interval $I$ iff for all $x_0 \in I$, the slope-function $s(x) = \frac{f(x) - f(x_0)}{x - x_0}$ is increasing on $I\setminus\{x_0\}$.

  **Proof.** Let $x_0 = \alpha x_1 + (1 - \alpha)x_2$. By convexity, we have $\alpha f(x_0) + (1 - \alpha)f(x_0) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)$, then $s(x_1) \leq s(x_2)$.

- Let a function $f$ be differentiable with an increasing derivative on an open interval $I$, then $f$ is convex on $I$.

- Let a function $f$ be twice differentiable with a non-negative second derivative on an open interval $I$. Then $f$ is convex on $I$.

**Definition 1** A function $f : C \to \mathbb{R}$, where $C \subset \mathbb{R}^n$ and $C \neq \emptyset$, is convex if

- $C$ is convex;

- For every $x, y \in C$ and every $\lambda \in [0, 1]$ one has
  
  $$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

More definitions:
Lecture note 2  Convex optimization

- **Concave**: if \(-f\) is convex, \(f\) is called concave.

- **Strictly convex**: if the inequality is strict whenever \(x \neq y\), and \(0 < \lambda < 1\), \(f\) is called strictly convex.

- **Strongly convex**: if there exist \(c > 0\) s.t.
  \[
  f(\alpha x + (1 - \alpha)x') \leq \alpha f(x) + (1 - \alpha)f(x') - \frac{1}{2}c\alpha(1 - \alpha)\|x - x'\|^2
  \]
  for all \(x, x' \in C\) and \(\alpha \in (0, 1)\). We say that \(f\) is strongly convex on \(C\) with modulus of strong convexity \(c\).

**Theorem 1** The function \(f\) is strongly convex on \(C\) with modulus \(c\) if and only if the function \(f - \frac{1}{2}c\|\cdot\|^2\) is convex on \(C\).

- **Extended value of \(f\)**. Define
  \[
  \tilde{f}(x) = \begin{cases} f(x), & x \in \text{dom}(f); \\ +\infty, & x \notin \text{dom}(f). \end{cases}
  \]

  1. For all \(x, y \in \mathbb{R}^n\), all \(\alpha \in [0, 1]\), \(\tilde{f}(\alpha x + (1 - \alpha)y) \leq \alpha \tilde{f}(x) + (1 - \alpha)\tilde{f}(y)\).
  2. Arithmetic operations involving \(+\cdot\): \((+\infty) + (+\infty) = +\infty; (-\infty) + (-\infty) = -\infty; (+\infty) + (-\infty) \text{ undefined.} \)
  \(+(\infty) \times (\pm\infty) = \pm\infty\);

- The domain (or effective domain) of \(f\) is the nonempty set such as \(f(x) < \infty\), denoted as \(\text{dom}(f) = \{x \in C | f(x) < \infty\}\) and the epigraph \(f\) is a subset in \(\mathbb{R}^n \times \mathbb{R}\) defined as \(\text{epi}(f)(x) = \{(x, t) \in \mathbb{R}^{n+1} : x \in \text{dom}(f); f(x) \leq t\}\).

**Theorem 2** Let \(f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) be not identically equal to +\(\infty\). The three properties below are equivalent:

1. \(f\) is convex (in the sense of definition).
2. Its epigraph \(\text{epi}(f)\) is a convex set in \(\mathbb{R}^n \times \mathbb{R}\).
3. Its strict epigraph is a convex set in \(\mathbb{R}^n \times \mathbb{R}\).

**Proof.**

"only if": If \(f\) is convex, \(f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)\). let \((x, t) \in \text{epi}(f)\) and \((y, t_2) \in \text{epi}(f)\). then
  \[
  f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \leq \lambda t_1 + (1 - \lambda)t_2.
  \]

Thus \((\lambda x + (1 - \lambda)y, \lambda t_1 + (1 - \lambda)t_2) \in \text{epi}(f)\). It yields \(\text{epi}(f)\) is convex.

"if part": if \(\text{epi}(f)\) is convex, for \((x, f(x)) \in \text{epi}(f)\) and \((y, f(y)) \in \text{epi}(f)\), \(f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)\). Thus \(f\) is convex.
• The sub level sets \( S_r(f) := \{ x \in \mathbb{R}^n : f(x) \leq r \} \) of a (proper) convex function \( f \) are convex (possibly empty). Conversely it is not necessarily true (such a function is called quasi-convex).

• Closed convex function: If we want to minimize a function \( f \) on some compact set \( K \), we do not need to bother with existence if \( f \) is known to be closed (or l.s.c) and this holds even if \( K \) is not contained in dom(\( f \)).

Lower semi-continuous (ls.c): a function \( f \) is l.s.c if, for each \( x \in \mathbb{R}^n \),
\[
\liminf_{y \to x} f(y) \geq f(x).
\]

**Proposition 1** For \( f : \mathbb{R}^n \times \mathbb{R} \cup \{ +\infty \} \), the following three properties are equivalent:

- \( f \) is l.s.c on \( \mathbb{R}^n \);
- epigraph \( \text{epi}(f) \) is a closed set in \( \mathbb{R}^n \);
- the sublevel-sets \( S_r(f) \) are closed (possibly empty) for all \( r \in \mathbb{R} \).

**Definition 2** The function \( f : \mathbb{R}^n \to \mathbb{R} \cup \{ \infty \} \) is said to be closed if it is l.s.c everywhere, or if its epigraph is closed, or if its sublevel-sets are closed.

**Examples of convex functions:**

1. **Example on \( \mathbb{R} \):**
   - Linear function \( f(x) = ax + b \)
   - \( e^{ax} \) for any \( a \in \mathbb{R} \).
   - \( x^\alpha \) on \( \mathbb{R}_{++} \) for \( \alpha \geq 1 \) or \( \alpha \leq 0 \). (\( f'(x) = \alpha x^{\alpha-1} \) and \( f''(x) = \alpha(\alpha - 1)x^{\alpha-2} \).)
   - \( |x|^p \) on \( \mathbb{R} \) for \( p \geq 1 \).
   - \( x \ln x \) on \( \mathbb{R}_{++} \) (negative entropy).
   - Some concave functions: \( ax + b, x^\alpha \) on \( \mathbb{R}_{++} \) for \( 0 \leq \alpha \leq 1 \), log(\( x \)) on \( \mathbb{R}_{++} \).

2. **Example on \( \mathbb{R}^n \):**
   - All affine function \( f(x) = a^T x + b \).
   - All norms are convex
   \[
   \|x\|_p = \left( \sum_{i=1}^{p} |x_i|^p \right)^{1/p}, \quad \text{for } p \geq 1
   \]
   \[
   \|x\|_\infty = \max_i |x_i|
   \]
   - Max function \( f(x) = \max\{x_1, \ldots, x_n\} \) is convex on \( \mathbb{R}^n \).
Lecture note 2

Convex optimization

- Quadratic-over-linear function: $f(x, y) = \frac{x^2}{y}$, $y > 0$ dom $f = \mathbb{R} \times \mathbb{R}_{++}$

- Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$.
  Note: Log-sum-exp function is a differentiable approximation of max function $f_\beta(x) = \frac{1}{\beta} \log(e^{\beta x_1} + \cdots + e^{\beta x_n})$
  \[
  \max_{x_1, \ldots, x_n} \leq f_\beta(x) \leq \max_i \{x_1, \ldots, x_n\} + \frac{1}{\beta} \log n
  \]

- Geometry mean $f(x) = (\prod_{i=1}^n x_i)^\frac{1}{n}$ is concave on dom($f$) = $\mathbb{R}^n_{++}$.

3. Examples in $\mathbb{R}^{m \times n}$

- Affine function $X \in \mathbb{R}^{m \times n}$:
  \[
  f(X) = \sum_i \sum_j A_{ij} X_{ij} + b \\
  = \sum_i (A^T X)_{ii} + b \\
  = \text{tr}(A^T X) + b
  \]

- Spectral norm
  \[
  f(x) = \|x\|_2 = (\lambda_1(x^T x))^{1/2}
  \]

- Nuclear norm
  \[
  f(x) = \|x\|_*
  \]

- Sum of largest eigenvalues of a symmetric matrix $A$. $f_m(A) := \sum_{j=1}^m \lambda_j(A)$ has also the representation $f_m(A) = \sup\{Q^T Q = I_m\} \text{ trace}(QAQ^T) = \text{ trace}(QQ^T A)$.

- Volume of ellipsoids: still in symmetric matrices $S_n(\mathbb{R})$, define the function $f(A) := \log(\det(A^{-1}))$ for positive definite $A$, otherwise $+\infty$. is convex.

4. Indicator and support functions: given a nonempty subset $S \subset \mathbb{R}^n$, the function $\chi_S : \mathbb{R}^n \to \mathbb{R} \times \{\infty\}$ defined by
  \[
  \chi_S := \begin{cases} 
  0; & \text{if } x \in S, \\
  \infty & \text{if not}
  \end{cases}
  \]

is called the indicator function of $S$. $\chi_S$ is (closed and) convex if and only if $S$ is (closed and) convex. Indeed, $\text{epi} \chi_S = S \times \mathbb{R}_+$ by definition.

The support function of a nonempty subset $S$ is defined as

\[
\sigma_S(x) := \sup\{\langle s, x \rangle : s \in S\}
\]

is convex. In fact its epigraph $\text{epi}(\sigma_S)$ is convex cone in $\mathbb{R} \times \mathbb{R}$.
1.2 Elemental properties of convex function

- Jensen’s inequality

**Proposition 2** Let \( f \) be convex and \( \text{dom}(f) = Q \), then for every convex combination \( \sum_i \lambda_i x_i \) of points from \( Q \), one has

\[
f(\sum_i \lambda_i x_i) \leq \sum_i \lambda_i f(x_i)
\]

**Proof.** Using the equivalence definition \( \text{epi}(f) \) is convex.

- Convexity of sub-level sets.

Let \( f \) be a convex function with the domain \( Q \), then any real \( \alpha \), the set

\[
\text{Lev}_\alpha(f) := \{x \in Q, f(x) \leq \alpha\}
\]

is convex.

(\text{convexity} \Rightarrow \text{convex sub-level sets}, but converse isn’t true).

(any sub-level set is convex \Rightarrow \text{quasiconvex}).

1.3 Operations that preserving convexity of functions

- Non-negative weighted sum: if \( f \) and \( g \) are convex, then \( \lambda f + \mu g \) is convex for any \( \lambda, \mu \geq 0 \).

- Affine substitutions is convex: If \( f \) is convex, then \( f(Ax + b) \) is convex.

- Pointwise max/sup: let \( \{f_j\}_{j \in J} \) for \( J \) be an arbitrary family of convex functions. Then \( f : \sup\{f_j : j \in J\} \) is convex (assume it is not identically \( \infty \)).

- Convex monotone superposition: Let \( f : \mathbb{R}^n \to \mathbb{R} \) be convex, and \( g : \mathbb{R} \to \mathbb{R} \) be convex and increasing. Then the composite \( g \circ f : x \to g(f(x)) \) is convex (set \( g(\infty) = \infty \) and \( g \circ f \) is not identically \( \infty \)).

- Partial (marginal) minimization: if \( f(x, y) \) is jointly convex, and let \( C \) be nonempty closed convex set and let \( g(x) = \inf_{y \in C} f(x, y) \) is proper and if \( f \) is bounded below on the set \( \{x\} \times C \) for any \( x \in \mathbb{R}^n \), then \( g \) is convex.

**Proof.**

By the boundedness of \( \{x\} \times C \) for any \( x \in \mathbb{R}^n \). The inf on \( y \) is attained on any \( x \). Thus \( \text{epi}(g) = \{(x, t) | \text{for some} y \in C, (x, y, t) \in \text{epi}(f)\} \) and it is the projection of \( \text{epi}(f) \) onto \( \mathbb{R}^n \times \mathbb{R} \). Therefore \( \text{epi}(g) \) is the image of a convex under a linear mapping and it is convex.

Example:
Lecture note 2 Convex optimization

– consider $f(x, y) = x^T A x + 2x^T B y + y^T C y$ with $\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ positive semi-definite, $C$ is positive definite minimizing over $y$ gives $g(x) = \inf_y f(x, y) = x^T (A - BC^{-1}B^T) x$, $g$ is convex, hence Schur complement $A - BC^{-1}B^T$ is positive semi-definite.
– Distance to a set: $\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if $S$ is convex.

• Dilation and perspectives of a function: Dilation: $f_u u = uf(x/u)$ is still convex ($\text{epi}(f_u) = u\text{epi}(f)$, $\text{sr}(f_u) = u\text{sr}(f)$.)
Perspectives: $\tilde{f}(u, x) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\tilde{f}(u, x) := \begin{cases} uf(x/u) & u > 0 \\ +\infty & \text{if not} \end{cases}$$

is convex in $\mathbb{R}^{n+1}$.

Proof.

It is better to look at $\tilde{f}$ with "epigraph".

$$\text{epi}(\tilde{f}) = \{(u, x, r) \in \mathbb{R}^*_+ \times \mathbb{R}^n \times \mathbb{R} : f(x/u) \leq r/u\}$$
$$= \{u(1, x', r') : u > 0, (x', r') \in \text{epi}(f)\}$$
$$= \cup_{u > 0} \{u(\{1\} \times \text{epi}(f))\}$$

and it is therefore a convex cone.

Example:

– $f(x) = x^T x$ convex, $f(x, t) = x^T x/t = tf(x/t)$ is convex.
– $f(x) = -\log(x)$ is convex, hence relative entropy $g(x, t) = t \log t - t \log x$ is convex on $\mathbb{R}^+_+$.  
– KullbackLeibler (KL) divergence divergence between $u, v \in \mathbb{R}^+_+$:

$$D_{kl}(u, v) = \sum_{i=1}^{n} (u_i \log(u_i/v_i) - u_i + v_i)$$

is convex since it is negative entropy plus linear function of $u$ and $v$.
– If $f$ is convex, then $g(x) = (c^T x + d)f((Ax + b)/(c^T x + d))$ is convex on $\text{dom}(g) = \{c^T x + d > 0(Ax + b)/(c^T x + d) \in \text{dom}(f)\}$.

• Restriction of a convex function to a line: If $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \to \mathbb{R}$,

$$g(t) = f(x + tv), \quad \text{dom}(g) = \{t | x + tv \in \text{dom}(f)\}$$
is convex for any \( x \in \text{dom}(f) \), \( v \in \mathbb{R}^n \). This means that one can check convexity of \( f \) by checking convexity of functions of one variable.

*Proof.*

When \( f(x) \) is convex, derive \( g(t) \) is convex by checking the definition. Conversely, for any \( x_0, x_1 \), consider \( g(t) = f(x_0 + t(x_1 - x_0)) \) and let \( t = 0 \) and \( t = 1 \).

*Example:*

Consider \( f : S^n \to \mathbb{R} \) with \( f(X) = \log \det(X) \), \( \text{dom}(f) = S^n_{++} \).

\[
g(t) = \log \det(X + tV) = \log \det(X) + \log \det(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}) = \log \det(X) + \sum_{i=1}^n \log(1 + t\lambda_i) \text{ where } \lambda_i \text{ are the eigenvalues of } X^{-\frac{1}{2}}VX^{-\frac{1}{2}}. \]

\( g \) is concave in \( t \) (for any choice of \( X \in S^n_{++}, V \)); hence \( f \) is concave.

### 1.4 Differential criteria of convexity

Consider \( C \subset \mathbb{R}^n \) be nonempty and convex and a function \( f \) is defined on \( C \) (\( f(x) < \infty \) for all \( x \in C \)). We consider \( f \) is convex and differentiable on \( C \).

**Theorem 3 (First order condition)** Let \( f \) be a function differentiable on an open set \( \Omega \subset \mathbb{R}^n \), and let \( C \) be a convex subset of \( \Omega \). Then

- \( f \) is convex on \( C \) if and only if
  \[
f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle, \quad \text{for all } (x_0, x) \in C \times C
  \]

- \( f \) is strictly convex on \( C \) if and only if strict inequality holds in the above inequality whenever \( x \neq x_0 \).

- \( f \) is strongly convex with modulus \( c \) on \( C \) if and only if for all \( (x_0, x) \in C \times C \),
  \[
f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2}c\|x - x_0\|^2
  \]

*Proof.*

1. Let \( f \) be convex on \( C \), for arbitrary \( (x_0, x) \in C \times C \) and \( \alpha \in (0, 1) \), we have
  \[
f(\alpha x + (1 - \alpha)x_0) - f(x_0) \leq \alpha(f(x) - f(x_0))
  \]

Divide by \( \alpha \) and let \( \alpha \to 0 \), the lefthand side reduces to

\[
\frac{f(x_0 + \alpha(x - x_0)) - f(x_0)}{\alpha} \to \langle \nabla f(x_0), x - x_0 \rangle
\]

and the inequality is established.
Conversely, take $x_1$ and $x_2$ in $C$, $\alpha \in (0,1)$ and set $x_0 = x_2 + \alpha(x_1 - x_2)$, by assumption

$$f(x_i) \geq f(x_0) + \langle \nabla f(x_0), x_i - x_0 \rangle$$

we obtain the convex combination

$$\alpha f(x_1) + (1 - \alpha) f(x_2) \geq f(x_0) + \langle \nabla f(x_0), \alpha x_1 + (1 - \alpha)x_2 - x_0 \rangle$$

which is the definition of convex function.

2. Similar for strictly convex for $x_0 \neq x$

3. Apply the proof in 1) on the function $f - \frac{1}{2}c\|\cdot\|^2$.

Example:

Non-negativity of Kullback-Leibler (KL) divergence

$$D_{\text{kl}}(u,v) = \sum_{i=1}^{n} (u_i \log(u_i/v_i) - u_i + v_i)$$

as

$$D_{\text{kl}}(u,v) = \phi(u) - \phi(v) - \langle \phi(v), u - v \rangle$$

where $\phi(x) = \sum_i u_i \log u_i$ is convex.

**Definition 3** Let $C \subset \mathbb{R}^n$ be convex. The mapping $F : C \to \mathbb{R}^n$ is said to be monotone [resp. strictly monotone, resp. strongly monotone with modulus $c > 0$] on $C$ when, for all $x, x' \in C$,

$$\langle F(x) - F(x'), x - x' \rangle \geq 0$$

[resp.$\langle F(x) - F(x'), x - x' \rangle > 0$ whenever $x \neq x'$, resp.$\langle F(x) - F(x'), x - x' \rangle \geq c\|x - x'\|^2$].

**Theorem 4 (monotonicity of gradient)** Let $f$ be a function differentiable on an open set $\Omega \subset \mathbb{R}^n$, and let $C$ be convex subset of $\Omega$. Then $f$ is convex [resp. strictly convex, resp. strongly convex with modulus $c$] on $C$ if and only if its gradient $\nabla f$ is monotone [resp. strictly monotone, strongly monotone with modulus $c$].

**Proof.**

Combine the case for "convex $\equiv$ monotone" and "strongly convex $\equiv$ strongly monotone" allowing $c = 0$.

Let $f$ be strongly convex on $C$, and for arbitrary $x, x_0 \in C$, we have

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2}c\|x - x_0\|^2$$

$$f(x_0) \geq f(x) - \langle \nabla f(x), x - x_0 \rangle + \frac{1}{2}c\|x - x_0\|^2$$

and mere addition leads to $\nabla f$ is strongly monotone.
Conversely, let \((x_0, x_1)\) be a pair in \(C\) and consider \(\phi(t) := f(x_t)\) where \(x_t := x_0 + t(x_1 - x_0)\) for \(t\) in an open interval containing \([0, 1]\) and \(\phi\) is well-defined differentiable. Its derivative at \(t\) is \(\phi'(t) = \langle \nabla f(x_t), x_1 - x_0 \rangle\). Thus for \(0 \leq t' < t \leq 1\) we have
\[
\phi'(t) - \phi'(t') = \langle \nabla f(x_t) - \nabla f(x_{t'}), x_1 - x_0 \rangle = \frac{1}{t-t'} \langle \nabla f(x_t) - \nabla f(x_{t'}), x_t - x_{t'} \rangle
\]
and the monotonicity for \(\nabla f\) shows that \(\phi'\) is increasing, \(\phi\) is therefore convex.

For strongly convex, set \(t' = 0\), and use the strongly monotonicity \(\phi'(t') - \phi'(0) \geq \frac{1}{t} c \|x_t - x_0\|^2 = tc \|x_1 - x_0\|^2\). On the other hand, we have
\[
\phi(1) - \phi(0) - \phi'(0) = \int_0^1 [\phi'(t) - \phi'(0)] dt \geq \frac{1}{2} c \|x_1 - x_0\|^2
\]
which by definition of \(\phi\), is just the definition of strongly convex function.

**Theorem 5 (Second order differentiation)** Let \(f\) be twice differentiable on an open convex set \(\Omega \subset \mathbb{R}^n\). Then

- \(f\) is convex on \(\Omega\) iff \(\nabla^2 f(x_0)\) is positive semi-definite for all \(x_0 \in \Omega\).
- if \(\nabla^2 f(x_0)\) is positive definite for all \(x_0 \in \Omega\), then \(f\) is strictly convex on \(\Omega\).
- \(f\) is strongly convex with modulus \(c\) on \(\Omega\) if and only if the smallest eigenvalue of \(\nabla^2 f(\cdot)\) is minorized by \(c\) on \(\Omega\): for all \(x_0 \in \Omega\) and all \(d \in \mathbb{R}^n\),
\[
\langle \nabla^2 f(x_0), d \rangle \geq c \|d\|^2
\]

**Proof.** Suppose \(f\) is convex. As \(f\) is twice differentiable, we have
\[
f(x + \delta x) = f(x) + \nabla f(x)^T \delta x + \frac{1}{2} (\delta x)^T \nabla^2 f(x) \delta x + R(x; \delta x) \|\delta x\|^2
\]
where \(R(x; \delta x) \to 0\) as \(\delta x \to 0\). As \(f\) is convex, by the first-order condition,
\[
f(x + \delta x) \geq f(x) + \nabla f(x)^T \delta x
\]
Hence
\[
\delta x^T \nabla^2 f(x) \delta x + R(x; \delta x) \|\delta x\|^2 \geq 0
\]
for any \(\delta x\). Let \(\delta x = \epsilon d\) and taking \(\epsilon \to 0\) yields \(d^T \nabla^2 f(x) d \geq 0\) for any \(d\), thus \(\nabla^2 f(x) \geq 0\). It is easy to check when \(\nabla^2 f(x) \geq 0\), the first order condition hold, thus \(f\) is convex. Note: similarly \(f\) is concave iff \(\text{dom}(f)\) is convex and \(\nabla^2 f(x)\) is negative semi-definite for \(x \in \text{dom}(f)\).

**Examples:**

- Quadratic function \(f(x) = \frac{1}{2} x^T P x + q^T x + r\) with \(P \in \mathbb{S}^n\) is convex if \(P \in \mathbb{S}^n_+\), \((\nabla f(x)) = P x + q, \nabla^2 f(x) = P\). Special case: \(f(x) = \|Ax - b\|^2\) is convex as \(\nabla^2 f(x) = 2A^T A\) is positive semi-definite for any \(A\).
• Quadratic over linear \( f(x, y) = \frac{x^2}{y} \) for \( y > 0 \). Show the Hessian matrix is
\[
\nabla^2 f(x, y) = \frac{2}{y^2} \begin{pmatrix}
y^2 & -yx \\
-yx & x^2
\end{pmatrix} = \frac{2}{y^2} \begin{pmatrix} y & x \\
x & -x
\end{pmatrix}.
\]

• Log-sum-exp: \( f(x) = \log(\sum_{k=1}^{n} \exp(x_k)) \) is convex. Show the Hessian matrix
\[
\nabla^2 f(x) = \frac{1}{1^T z} \text{diag}(z) - \frac{1}{(1^T z)^2} z z^T
\]
is positive semi-definite for \( z_k = \exp(x_k) \). In fact
\[
\max\{x_1, x_2, \cdots, x_n\} \leq f(x) \leq \max\{x_1, \cdots, x_n\} + \log n,
\]
so \( f \) can be viewed as a differentiable approximation of the max function.

• Geometric mean: \( f(x) = (\prod_{k=1}^{n} x_k)^{1/n} \) on \( \mathbb{R}^n_+ \) is concave.

1.5 Conjugacy

The conjugate of \( f \):
\[
f^*(y) := \sup_{x \in \text{dom}(f)} \langle y, x \rangle - f(x)
\]

Figure 1.1: A function \( f : \mathbb{R} \to \mathbb{R} \), and a value \( y \in \mathbb{R} \). The conjugate function \( f^*(y) \) is the maximum gap between the linear function \( yx \) and \( f(x) \). If \( f \) is differentiable, this occurs at a point \( x \) where \( f'(x) = y \).

• The domain of \( f^*(y) \) consists of \( y \in \mathbb{R}^n \) for which the supremum is finite.

• \( f^* \) is convex even \( f \) is not convex since \( f^* \) is a set supremum of a convex (affine function of \( y \)).

• For differentiable function, the mapping \( f \to f^* \) is called Legendre-Fenchel transform.

Examples:
Lecture note 2 Convex optimization

- $f(x) = a^T x + b$, $f^*(y) = \sup_x y^T x - ax - b$ is bounded iff $y = a$. $\text{dom}(f^*) = \{a\}$ and $f^*(a) = b$.

- Negative logarithm $f(x) = -\log(x)$, $\text{dom}(f) = \mathbb{R}_{++}$.

\[
f^*(y) = \sup_{x \in \text{dom}(f)} y^T x + \log(x) = \begin{cases} \text{unbounded} & y \geq 0 \\ -\log(-y) - 1 & y < 0 \end{cases}
\]

(for $x = -\frac{1}{y}$) when $\text{dom}(f^*) = \{y < 0\}$.

- $f(x) = e^x$. For $y < 0$, $xy = e^x$ is unbounded (let $x \to -\infty$); For $y > 0$, when $x = \log y$, $xy - e^x$ reaches its maximum and $f^*(y) = y \log y - y$. For $y = 0$, $f^*(y) = \sup -e^x = 0$, thus $f^*(y) = y \log y - y$ (with $0 \log 0 = 0$).

- Strictly convex quadratic: $f(x) = \frac{1}{2}x^T Q x$, $Q \in S^n_{++}$. $y^T x - \frac{1}{2}x^T Q x$ is bounded above for all $x$, maximum at $Q^{-1}y = x$, thus $f^*(y) = \frac{1}{2}y^T Q^{-1} y$.

- Indicator function: let $I_S(x) = \begin{cases} 0 & x \in S \\ +\infty & \text{if not} \end{cases}$, thus $I_S^*(y) = \sup_{x \in S} y^T x$ (support function of $S$).

- Log-sum-sup: $f(x) = \log(\sum_{i=1}^n e^{x_i})$, $f^*(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i & y \geq 0, 1^T y = 1 \\ +\infty & \text{if not} \end{cases}$

- Norm: let $\| \cdot \|$ be a norm in $\mathbb{R}^n$, with dual norm $\| \cdot \|^*$. We will show that the conjugate of $f(x) = \| x \|$ is

\[
f^*(y) = \begin{cases} 0 & \|y\|^* \leq 1 \\ +\infty & \text{if not} \end{cases}
\]

Recall that $\|y\|^* = \sup\{\|y^T x\|, \|y\| \leq 1\}$. For $\|y\|^* > 1$, by definition, there exists $z$ such that $\|z\| \leq 1$ and $y^T z > 1$. Let $x = tz$ and $t \to \infty$, then

$y^T x - \|x\| = t(y^T z - \|z\|) \to \infty$, thus $f^*(y)$ is unbounded. Conversely, if $\|y\|^* \leq 1$, then $y^T x \leq \|y\|^* \|x\|$ and $y^T x - \|x\| \leq 0$ and the maximum is attained at $x = 0$. Thus

\[
f^*(y) = \begin{cases} 0 & \|y\|^* \leq 1 \\ +\infty & \text{if not} \end{cases}
\]

- Norm square: let $f(x) = \frac{1}{2}\| x \|^2$ for some norm. Then $f^*(y) = \frac{1}{2}\|y\|^2$.

Properties of conjugate functions

- Fenchel’s inequality: $f^*(y) + f(x) \geq x^T y$ (Young’s inequality for differentiable $f$). Example: $x^T y \leq 1/2 x^T Q x + 1/2 y^T Q^{-1} y$.

- Conjugate of conjugate: if $f$ is convex and $f$ is closed, then $f^{**} = f$. 
Suppose $f$ is convex and differentiable with $\text{dom}(f) = \mathbb{R}^n$. Any maximizer $x^*$ s.t. $\nabla f(x^*) = y$ and conversely if $x^*$ s.t $y = \nabla f(x^*)$, then $x^*$ maximize $y^T x - f(x)$ and $f^*(y) = \nabla f(x^*)^T x^* - f(x^*) = y^T (\nabla f)^{-1}(y) - f((\nabla f)^{-1}(y))$. Thus $df^*(y) = \langle y, dx \rangle + \langle x^*, y \rangle - \langle \nabla f(x^*), dx^* \rangle = \langle dy, x^* \rangle$. Thus $\nabla f^*(y) = x^*$.

Scaling and composition with affine transformation: for $a > 0, b \in \mathbb{R}$ $g(x) = ax + b$, thus $g^*(y) = a f^*(y/a) - b$. Suppose that $A \in \mathbb{R}^{n \times n}$ is nonsingular and $b \in \mathbb{R}^n$, then the conjugate of $f(Ax + b)$ is $g^*(y) = f^*(A^{-T} y) - b^T A^{-T} y$ with $\text{dom} g^* = A^T \text{dom} f^*$.

Sum of independent function $f(u, v) = f_1(u) + f_2(v)$ where $f_1, f_2$ are both convex with conjugate $f^*(w, z) = f_1^*(w) + f_2^*(z)$

Convexity of the conjugation: if $\text{dom}(f_1) \cap \text{dom}(f_2) \neq \emptyset$, and $\alpha \in (0, 1)$, then

$$[\alpha f_1 + (1 - \alpha)f_2]^* \leq \alpha f_1^* + (1 - \alpha)f_2^*$$