Practical Session on Convex Optimization: Convex Analysis

Mark Schmidt
INRIA/ENS

September 2011
Two key properties of convex functions:

- All local minima are global minima.
- Global rate of convergence analysis.
A real-valued function is **convex** if

\[ f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \]

for all \( x, y \in \mathbb{R}^n \) and all \( 0 \leq \theta \leq 1 \).

- Function is *below the chord* from \( x \) to \( y \).
A real-valued function is **convex** if

\[ f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \]

for all \( x, y \in \mathbb{R}^n \) and all \( 0 \leq \theta \leq 1 \).

- Function is *below the chord* from \( x \) to \( y \).
- Show that all local minima are global minima.
A real-valued function $f$ is a **norm** if:

1. $f(x) \geq 0$, $f(0) = 0$.
2. $f(\theta x) = |\theta| f(x)$.
3. $f(x + y) \leq f(x) + f(y)$.

Show that norms are convex.
Exercise: Convexity of Norms

A real-valued function $f$ is a norm if:

1. $f(x) \geq 0$, $f(0) = 0$.
2. $f(\theta x) = |\theta|f(x)$.
3. $f(x + y) \leq f(x) + f(y)$.

Show that norms are convex.

1. Use triangle inequality then homogeneity.
A real-valued function is **strictly convex** if

\[ f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y), \]

for all \( x, y \in \mathbb{R}^n \) and all \( 0 < \theta < 1 \).

- Function is *strictly below the chord* between \( x \) to \( y \).
A real-valued function is **strictly convex** if

\[ f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y), \]

for all \( x, y \in \mathbb{R}^n \) and all \( 0 < \theta < 1 \).

- Function is *strictly below the chord* between \( x \) to \( y \).
- Show that global minimum of strictly convex function is unique.
A real-valued differentiable function is convex iff

\[ f(x) \geq f(y) + \nabla f(y)^T (x - y), \]

for all \( x, y \in \mathbb{R}^n \).

- The function is globally above the tangent at \( y \).
A real-valued \textit{differentiable} function is \textbf{convex} \textit{iff}

\[
f(x) \geq f(y) + \nabla f(y)^T (x - y),
\]

for all \(x, y \in \mathbb{R}^n\).

- The function is globally \textit{above the tangent} at \(y\).
- Show that any stationary point is a global minimum.
Show that zero-order condition,

\[ f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \]

implies first-order condition,

\[ f(x) \geq f(y) + \nabla f(y)^T (x - y). \]
Show that zero-order condition,

\[ f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \]

implies first-order condition,

\[ f(x) \geq f(y) + \nabla f(y)^T (x - y). \]

1. Use: \( \theta x + (1 - \theta)y = y + \theta(x - y). \)

2. Use:

\[
\nabla f(y)^T d = \lim_{\theta \to 0} \frac{f(y + \theta d) - f(y)}{\theta}
\]
A real-valued *twice-differentiable* function is **convex** iff

\[ \nabla^2 f(x) \succeq 0 \]

for all \( x \in \mathbb{R}^n \).

- The function is *flat or curved upwards* in every direction.
A real-valued twice-differentiable function is convex iff

\[ \nabla^2 f(x) \succeq 0 \]

for all \( x \in \mathbb{R}^n \).

The function is flat or curved upwards in every direction.

A real-valued function \( f \) is a quadratic if it can be written in the form:

\[ f(x) = x^T Ax + b^T x + c. \]

Show sufficient conditions for a quadratic function to be convex.
Exercise: Convexity of Basic Functions

Show that the following are convex:

1. \( f(x) = \exp(ax) \)
2. \( f(x) = x \log x \) (for \( x > 0 \))
3. \( f(x) = a^T x \)
4. \( f(x) = \|x\|^2 \)
5. \( f(x) = \max_i \{x_i\} \)
Some other notable convex functions:

1. \( f(x, y) = \log(e^x + e^y) \)
2. \( f(X) = \log \det X \) (for \( X \) positive-definite).
3. \( f(x, Y) = x^T Y^{-1} x \) (for \( Y \) positive-definite)
Operations that Preserve Convexity

1. **Non-negative weighted sum:**
   \[ f(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \cdots + \theta_n f_n(x). \]

2. **Composition with affine mapping:**
   \[ g(x) = f(Ax + b). \]

3. **Pointwise maximum:**
   \[ f(x) = \max\{f_i(x)\}. \]
Operations that Preserve Convexity

1. Non-negative weighted sum:

\[ f(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \cdots + \theta_n f_n(x). \]

2. Composition with affine mapping:

\[ g(x) = f(Ax + b). \]

3. Pointwise maximum:

\[ f(x) = \max\{f_i(x)\}. \]

Show that least-residual problems are convex for any \( \ell_p \)-norm:

\[ f(x) = ||Ax - b||_p \]

Show that SVMs are convex:

\[ f(x) = ||x||^2 + C \sum_{i=1}^{n} \max\{0, 1 - b_i a_i^T x\}. \]
Two key properties of convex functions:

- All local minima are global minima.
- Global rate of convergence analysis.
Assume that $f$ is a twice-differentiable function, where for all $x$ we have

$$\mu I \preceq \nabla f(x) \preceq LI,$$

for some $\mu > 0$ and $L < \infty$. 
Assume that $f$ is a twice-differentiable, where for all $x$ we have
\[ \mu I \preceq \nabla f(x) \preceq LI, \]
for some $\mu > 0$ and $L < \infty$.

By Taylor’s theorem, for any $x$ and $y$ we have
\[ f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2 f(z) (y - x), \]
for some $z$. 
From the previous slide, we get for all $x$ and $y$ that

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

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} ||y - x||^2.$$
From the previous slide, we get for all $x$ and $y$ that
\[
f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} ||y - x||^2,
\]
\[
f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} ||y - x||^2.
\]

Use these to show that the gradient iteration
\[
x_{k+1} = x_k - (1/L)\nabla f(x_k),
\]
has the linear convergence rate
\[
f(x_k) - f(x^*) \leq (1 - \mu/L)^k [f(x_0) - f(x^*)].
\]
Convergence Rate: Strongly-Convex Functions

- From the previous slide, we get for all $x$ and $y$ that

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

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\mu}{2}\|y - x\|^2.$$

- Use these to show that the gradient iteration

$$x_{k+1} = x_k - (1/L)\nabla f(x_k),$$

has the linear convergence rate

$$f(x_k) - f(x^*) \leq (1 - \mu/L)^k[f(x_0) - f(x^*)].$$

- Use this result to get a convergence rate on $\|x_k - x^*\|$.
Convergence Rate: Strongly-Convex Functions

- From the previous slide, we get for all \( x \) and \( y \) that

\[
f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|^2,
\]

\[
f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} \|y - x\|^2.
\]

- Use these to show that the gradient iteration

\[
x_{k+1} = x_k - (1/L) \nabla f(x_k),
\]

has the linear convergence rate

\[
f(x_k) - f(x^*) \leq (1 - \mu/L)^k [f(x_0) - f(x^*)].
\]

- Use this result to get a convergence rate on \( \|x_k - x^*\| \).

- Show that if \( \mu = 0 \) we get the sublinear rate \( O(1/k) \).
Most of this lecture is based on material from Boyd and Vandenberghe’s very good “Convex Optimization” book, as well as their online notes.