CS675: Convex and Combinatorial Optimization
Fall 2019
Convex Functions

Instructor: Shaddin Dughmi
Outline

1. Convex Functions
2. Examples of Convex and Concave Functions
3. Convexity-Preserving Operations
A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if the line segment between any points on the graph of $f$ lies above $f$. i.e. if $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$, then

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$
A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if the line segment between any points on the graph of $f$ lies above $f$. i.e. if $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$, then

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

- Inequality called **Jensen’s inequality** (basic form)
A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if the line segment between any points on the graph of $f$ lies above $f$. i.e. if $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$, then

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

- Inequality called Jensen’s inequality (basic form)
- $f$ is convex iff its restriction to any line $\{x + tv : t \in \mathbb{R}\}$ is convex
Convex Functions

A function \( f : \mathbb{R}^n \to \mathbb{R} \) is **convex** if the line segment between any points on the graph of \( f \) lies above \( f \). i.e. if \( x, y \in \mathbb{R}^n \) and \( \theta \in [0, 1] \), then

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

- Inequality called **Jensen’s inequality** (basic form)
- \( f \) is convex iff its restriction to any line \( \{x + tv : t \in \mathbb{R}\} \) is convex
- \( f \) is **strictly** convex if inequality strict when \( x \neq y \).
A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if the line segment between any points on the graph of $f$ lies above $f$. i.e. if $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$, then

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

- Inequality called Jensen’s inequality (basic form)
- $f$ is convex iff its restriction to any line $\{x + tv : t \in \mathbb{R}\}$ is convex
- $f$ is strictly convex if inequality strict when $x \neq y$.
- Analogous definition when the domain of $f$ is a convex subset $D$ of $\mathbb{R}^n$
A function is $f : \mathbb{R}^n \to \mathbb{R}$ is **concave** if $-f$ is convex. Equivalently:

- Line segment between any points on the graph of $f$ lies below $f$.
- If $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$, then
  
  $$f(\theta x + (1 - \theta)y) \geq \theta f(x) + (1 - \theta)f(y)$$

**Concave and Affine Functions**
A function is $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is **concave** if $-f$ is convex. Equivalently:

- Line segment between any points on the graph of $f$ lies **below** $f$.
- If $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$, then
  $$f(\theta x + (1 - \theta)y) \geq \theta f(x) + (1 - \theta)f(y)$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is **affine** if it is both concave and convex. Equivalently:

- Line segment between any points on the graph of $f$ lies on the graph of $f$.
- $f(x) = a^T x + b$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. 
We will now look at some equivalent definitions of convex functions

**First Order Definition**

A differentiable \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex if and only if the first-order approximation centered at any point \( x \) underestimates \( f \) everywhere.

\[
f(y) \geq f(x) + (\nabla f(x))^T(y - x)
\]
We will now look at some equivalent definitions of convex functions

First Order Definition
A differentiable $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the first-order approximation centered at any point $x$ underestimates $f$ everywhere.

$$f(y) \geq f(x) + (\nabla f(x))^T(y - x)$$

- Local information $\rightarrow$ global information
- If $\nabla f(x) = 0$ then $x$ is a global minimizer of $f$
Second Order Definition

A twice differentiable \( f : \mathbb{R}^n \to \mathbb{R} \) is convex if and only if its Hessian matrix \( \nabla^2 f(x) \) is positive semi-definite for all \( x \). (We write \( \nabla^2 f(x) \succeq 0 \))
**Second Order Definition**

A twice differentiable $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only if its Hessian matrix $\nabla^2 f(x)$ is positive semi-definite for all $x$. (We write $\nabla^2 f(x) \succeq 0$)

**Interpretation**

- Recall definition of PSD: $z^\top \nabla^2 f(x) z \geq 0$ for all $z \in \mathbb{R}^n$.
- When $n = 1$, this is $f''(x) \geq 0$.
- More generally, $\frac{z^\top \nabla^2 f(x) z}{\|z\|^2}$ is the second derivative of $f$ along the line $\{x + tz : t \in \mathbb{R}\}$. So if $\nabla^2 f(x) \succeq 0$ then $f$ curves upwards along any line.
- Moving from $x$ to $x + \delta \bar{z}$, infinitesimal change in gradient is $\delta \nabla^2 f(x) z$. When $\nabla^2 f(x) \succeq 0$, this is in roughly the same direction as $\bar{z}$. 

Convex Functions
The epigraph of $f$ is the set of points above the graph of $f$. Formally,

$$\text{epi}(f) = \{(x, t) : t \geq f(x)\}$$
The epigraph of $f$ is the set of points above the graph of $f$. Formally,

$$\text{epi}(f) = \{(x, t) : t \geq f(x)\}$$

**Epigraph Definition**

$f$ is a convex function if and only if its epigraph is a convex set.
**Jensen’s Inequality (General Form)**

\( f : \mathbb{R}^n \to \mathbb{R} \) is convex if and only if

- For every \( x_1, \ldots, x_k \) in the domain of \( f \), and \( \theta_1, \ldots, \theta_k \geq 0 \) such that \( \sum_i \theta_i = 1 \), we have
  \[
  f\left(\sum_i \theta_i x_i\right) \leq \sum_i \theta_i f(x_i)
  \]

- Given a probability measure \( \mathcal{D} \) on the domain of \( f \), and \( x \sim \mathcal{D} \),
  \[
  f(\mathbb{E}[x]) \leq \mathbb{E}[f(x)]
  \]
Jensen’s Inequality (General Form)

\( f : \mathbb{R}^n \to \mathbb{R} \) is convex if and only if

- For every \( x_1, \ldots, x_k \) in the domain of \( f \), and \( \theta_1, \ldots, \theta_k \geq 0 \) such that \( \sum_i \theta_i = 1 \), we have
  \[
  f \left( \sum_i \theta_i x_i \right) \leq \sum_i \theta_i f(x_i)
  \]

- Given a probability measure \( \mathcal{D} \) on the domain of \( f \), and \( x \sim \mathcal{D} \),
  \[
  f \left( \mathbb{E}[x] \right) \leq \mathbb{E}[f(x)]
  \]

Adding noise to \( x \) can only increase \( f(x) \) in expectation.
Local and Global Optimality

**Local minimum**

\( x \) is a **local minimum** of \( f \) if there is a an open ball \( B \) containing \( x \) where \( f(y) \geq f(x) \) for all \( y \in B \).

**Local and Global Optimality**

When \( f \) is convex, \( x \) is a local minimum of \( f \) if and only if it is a global minimum.
Local and Global Optimality

**Local minimum**

$x$ is a **local minimum** of $f$ if there is a an open ball $B$ containing $x$ where $f(y) \geq f(x)$ for all $y \in B$.

**Local and Global Optimality**

When $f$ is convex, $x$ is a local minimum of $f$ if and only if it is a global minimum.

- This fact underlies much of the tractability of convex optimization.
Sub-level sets

Level sets of \( f(x, y) = \sqrt{x^2 + y^2} \)

Sublevel set

The \( \alpha \)-sublevel set of \( f \) is \( \{ x \in \text{domain}(f) : f(x) \leq \alpha \} \).
Sub-level sets

Level sets of $f(x, y) = \sqrt{x^2 + y^2}$

Sublevel set

The $\alpha$-sublevel set of $f$ is $\{x \in \text{domain}(f) : f(x) \leq \alpha\}$.

Fact

Every sub-level set of a convex function is a convex set.

- This fact also underlies tractability of convex optimization.
Sub-level sets

Level sets of \( f(x, y) = \sqrt{x^2 + y^2} \)

Sublevel set

The \( \alpha \)-sublevel set of \( f \) is \( \{ x \in \text{domain}(f) : f(x) \leq \alpha \} \).

Fact

Every sub-level set of a convex function is a convex set.

- This fact also underlies tractability of convex optimization

Note: converse false, but nevertheless useful check.
Continuity

Real-valued convex functions are continuous on the interior of their domain.
Other Basic Properties

Continuity

Real-valued convex functions are continuous on the interior of their domain.

Extended-value extension

If a function $f : D \to \mathbb{R}$ is convex on its domain, and $D$ is convex, then it can be extended to a convex function on $\mathbb{R}^n$ by setting $f(x) = \infty$ whenever $x \notin D$.

This simplifies notation. Resulting function $\tilde{f} : D \to \mathbb{R} \cup \infty$ is “convex” with respect to the ordering on $\mathbb{R} \cup \infty$. 
Outline

1. Convex Functions

2. Examples of Convex and Concave Functions

3. Convexity-Preserving Operations
Functions on the reals

- **Affine**: $ax + b$
- **Exponential**: $e^{ax}$ convex for any $a \in \mathbb{R}$
- **Powers**: $x^a$ convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$
- **Logarithm**: $\log x$ concave on $\mathbb{R}_{++}$. 
Norms are convex.

\[ \| \theta x + (1 - \theta)y \| \leq \| \theta x \| + \| (1 - \theta)y \| = \theta \| x \| + (1 - \theta) \| y \| \]

- Uses both norm axioms: triangle inequality, and homogeneity.
- Applies to matrix norms, such as the spectral norm (radius of induced ellipsoid)
Norms

Norms are convex.

\[ \|\theta x + (1 - \theta)y\| \leq \|\theta x\| + \|(1 - \theta)y\| = \theta\|x\| + (1 - \theta)\|y\| \]

- Uses both norm axioms: triangle inequality, and homogeneity.
- Applies to matrix norms, such as the spectral norm (radius of induced ellipsoid)

Max

\[ \max_i x_i \text{ is convex} \]

\[ \max_i (\theta x + (1 - \theta)y)_i = \max_i (\theta x_i + (1 - \theta)y_i) \]
\[ \leq \max_i \theta x_i + \max_i (1 - \theta)y_i \]
\[ = \theta \max_i x_i + (1 - \theta) \max_i y_i \]

If i’m allowed to pick the maximum entry of \( \theta x \) and \( \theta y \) independently, I can do only better.
Log-sum-exp: \( \log(e^{x_1} + e^{x_2} + \ldots + e^{x_n}) \) is convex

Geometric mean: \( (\prod_{i=1}^{n} x_i)^{1/n} \) is concave

Log-determinant: \( \log \det X \) is concave

Quadratic form: \( x^\top A x \) is convex iff \( A \succeq 0 \)

Other examples in book

\[ f(x, y) = \log(e^x + e^y) \]
- Log-sum-exp: \( \log(e^{x_1} + e^{x_2} + \ldots + e^{x_n}) \) is convex
- Geometric mean: \( (\prod_{i=1}^{n} x_i)^{\frac{1}{n}} \) is concave
- Log-determinant: \( \log \det X \) is concave
- Quadratic form: \( x^\top A x \) is convex iff \( A \succeq 0 \)
- Other examples in book

Proving convexity often comes down to case-by-case reasoning, involving:

- Definition: restrict to line and check Jensen’s inequality
- Write down the Hessian and prove PSD
- Express as a combination of other convex functions through convexity-preserving operations (Next)
Outline

1. Convex Functions

2. Examples of Convex and Concave Functions

3. Convexity-Preserving Operations
Nonnegative Weighted Combinations

If $f_1, f_2, \ldots, f_k$ are convex, and $w_1, w_2, \ldots, w_k \geq 0$, then
\[ g = w_1 f_1 + w_2 f_2 \ldots + w_k f_k \] is convex.
Nonnegative Weighted Combinations

If $f_1, f_2, \ldots, f_k$ are convex, and $w_1, w_2, \ldots, w_k \geq 0$, then $g = w_1f_1 + w_2f_2 \ldots + w_k f_k$ is convex.

proof ($k = 2$)

$$g \left( \frac{x + y}{2} \right) = w_1 f_1 \left( \frac{x + y}{2} \right) + w_2 f_2 \left( \frac{x + y}{2} \right)$$

$$\leq w_1 \frac{f_1(x) + f_1(y)}{2} + w_2 \frac{f_2(x) + f_2(y)}{2}$$

$$= \frac{g(x) + g(y)}{2}$$
If \( f_1, f_2, \ldots, f_k \) are convex, and \( w_1, w_2, \ldots, w_k \geq 0 \), then 
\[
g = w_1 f_1 + w_2 f_2 \ldots + w_k f_k \text{ is convex.}
\]

Extends to integrals 
\[
g(x) = \int y \, w(y) f_y(x) \text{ with } w(y) \geq 0,
\]
and therefore expectations 
\[
\mathbb{E}_y f_y(x).
\]
Nonnegative Weighted Combinations

If \( f_1, f_2, \ldots, f_k \) are convex, and \( w_1, w_2, \ldots, w_k \geq 0 \), then
\[
g = w_1 f_1 + w_2 f_2 \ldots + w_k f_k
\]
is convex.

Extends to integrals \( g(x) = \int y w(y) f_y(x) \) with \( w(y) \geq 0 \), and therefore expectations \( \mathbb{E}_y f_y(x) \).

Worth Noting

Minimizing the expectation of a random convex cost function is also a convex optimization problem!

- A **stochastic** convex optimization problem is a convex optimization problem.
Example: Stochastic Facility Location

Average Distance

- \( k \) customers located at \( y_1, y_2, \ldots, y_k \in \mathbb{R}^n \)
- If I place a facility at \( x \in \mathbb{R}^n \), average distance to a customer is

\[
g(x) = \sum_i \frac{1}{k} ||x - y_i||
\]
Example: Stochastic Facility Location

Average Distance

- $k$ customers located at $y_1, y_2, \ldots, y_k \in \mathbb{R}^n$
- If I place a facility at $x \in \mathbb{R}^n$, average distance to a customer is $g(x) = \sum_i \frac{1}{k} ||x - y_i||$

- Since distance to any one customer is convex in $x$, so is the average distance.
- Extends to probability measure over customers
Implication

Convex functions are a convex cone in the vector space of functions from $\mathbb{R}^n$ to $\mathbb{R}$.

The set of convex functions is the intersection of an infinite set of homogeneous linear inequalities indexed by $x, y, \theta$

$$f(\theta x + (1 - \theta)y) - \theta f(x) - (1 - \theta)f(y) \leq 0$$
If \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex, and \( A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n \), then
\[
g(x) = f(Ax + b)
\]
is a convex function from \( \mathbb{R}^m \) to \( \mathbb{R} \).
Composition with Affine Function

If \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex, and \( A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n \), then

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

is a convex function from \( \mathbb{R}^m \) to \( \mathbb{R} \).

Proof

\[(x, t) \in \text{graph}(g) \iff t = g(x) = f(Ax + b) \iff (Ax + b, t) \in \text{graph}(f)\]
Composition with Affine Function

If \( f : \mathbb{R}^n \to \mathbb{R} \) is convex, and \( A \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n \), then

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

is a convex function from \( \mathbb{R}^m \) to \( \mathbb{R} \).

Proof

\[
(x, t) \in \text{graph}(g) \iff t = g(x) = f(Ax + b) \iff (Ax + b, t) \in \text{graph}(f)
\]

\[
(x, t) \in \text{epi}(g) \iff t \geq g(x) = f(Ax + b) \iff (Ax + b, t) \in \text{epi}(f)
\]
Composition with Affine Function

If \( f : \mathbb{R}^n \to \mathbb{R} \) is convex, and \( A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n \), then
\[
g(x) = f(Ax + b)
\]
is a convex function from \( \mathbb{R}^m \) to \( \mathbb{R} \).

Proof

\((x, t) \in \text{graph}(g) \iff t = g(x) = f(Ax + b) \iff (Ax + b, t) \in \text{graph}(f)\)

\((x, t) \in \text{epi}(g) \iff t \geq g(x) = f(Ax + b) \iff (Ax + b, t) \in \text{epi}(f)\)

\(\text{epi}(g)\) is the inverse image of \(\text{epi}(f)\) under the affine mapping
\((x, t) \to (Ax + b, t)\)
Examples

- $||Ax + b||$ is convex
- $\max(Ax + b)$ is convex
- $\log(e^{a_1^T x + b_1} + e^{a_2^T x + b_2} + \ldots + e^{a_n^T x + b_n})$ is convex
Maximum

If $f_1, f_2$ are convex, then $g(x) = \max \{ f_1(x), f_2(x) \}$ is also convex.

Generalizes to the maximum of any number of functions, $\max_{i=1}^{k} f_i(x)$, and also to the supremum of an infinite set of functions $\sup_y f_y(x)$. 
If $f_1, f_2$ are convex, then $g(x) = \max \{f_1(x), f_2(x)\}$ is also convex.

Generalizes to the maximum of any number of functions, $\max_{i=1}^{k} f_i(x)$, and also to the supremum of an infinite set of functions $\sup_y f_y(x)$.

$$\text{epi } g = \text{epi } f_1 \cap \text{epi } f_2$$
Example: Robust Facility Location

Maximum Distance

- *k* customers located at \( y_1, y_2, \ldots, y_k \in \mathbb{R}^n \)
- If I place a facility at \( x \in \mathbb{R}^n \), maximum distance to a customer is
  
  \[
g(x) = \max_i \| x - y_i \|
  \]
Example: Robust Facility Location

Maximum Distance

- $k$ customers located at $y_1, y_2, \ldots, y_k \in \mathbb{R}^n$
- If I place a facility at $x \in \mathbb{R}^n$, maximum distance to a customer is 
  $$g(x) = \max_i ||x - y_i||$$

Since distance to any one customer is convex in $x$, so is the worst-case distance.
Example: Robust Facility Location

Maximum Distance

- $k$ customers located at $y_1, y_2, \ldots, y_k \in \mathbb{R}^n$
- If I place a facility at $x \in \mathbb{R}^n$, maximum distance to a customer is $g(x) = \max_i \|x - y_i\|$

Worth Noting

When a convex cost function is uncertain, minimizing the worst-case cost is also a convex optimization problem!

- A robust (in the worst-case sense) convex optimization problem is a convex optimization problem.
Other Examples

- Maximum eigenvalue of a symmetric matrix $A$ is convex in $A$

\[
\max \{ v^\top A v : \|v\| = 1 \}
\]

- Sum of k largest components of a vector $x$ is convex in $x$

\[
\max \left\{ \vec{1}_S \cdot x : |S| = k \right\}
\]
Minimization

If $f(x, y)$ is convex and $C$ is convex and nonempty, then $g(x) = \inf_{y \in C} f(x, y)$ is convex.
Minimization

If $f(x, y)$ is convex and $C$ is convex and nonempty, then $g(x) = \inf_{y \in C} f(x, y)$ is convex.

Proof (for $C = \mathbb{R}^k$)

$\text{epi } g$ is the projection of $\text{epi } f$ onto hyperplane $y = 0$. 

\[ f(x, y) = x^2 + y^2 \]

\[ g(x) = x^2 \]
Example

Distance from a convex set $C$

\[ f(x) = \inf_{y \in C} ||x - y|| \]
Composition Rules

If \( g : \mathbb{R}^n \rightarrow \mathbb{R}^k \) and \( h : \mathbb{R}^k \rightarrow \mathbb{R} \), then \( f = h \circ g \) is convex if

- \( g_i \) are convex, and \( h \) is convex and nondecreasing in each argument.
- \( g_i \) are concave, and \( h \) is convex and nonincreasing in each argument.

Proof \((n = k = 1)\)

\[
f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)
\]
Perspective

If $f$ is convex then $g(x, t) = tf(x/t)$ is also convex.

Proof

$\text{epi } g$ is inverse image of $\text{epi } f$ under the perspective function.