Convex Functions (II)

Lecture 4, Nonlinear Programming

National Taiwan University

October 11, 2016

Table of contents

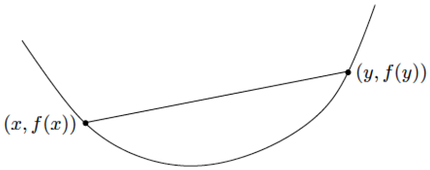
Definitions of Convex Functions

Convex functions

A function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\ f$ is a convex set and if for all $x, y \in \mathbf{dom}\ f$ and for all $\theta \in [0, 1]$,

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

- The line segment between $(x, f(x))$ and $(y, f(y))$, which is the chord from x to y, lies above the graph of f.



Definitions of Convex Functions

Convex functions

A function $f : \mathbf{R}^n \to \mathbf{R}$ is **convex** if **dom** f is a convex set and if for all $x, y \in$ **dom** f and for all $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y). \tag{1}$$

- A function f is **strictly convex** if strict inequality holds in (1) whenever $x \neq y$ and $0 < \theta < 1$:

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y).$$

- We say f is **concave** if $-f$ is convex, and **strictly concave** if $-f$ is strictly convex.

Affine Functions

Affine Functions

For an affine function we always have equality in (1), i.e.,

$$f(\theta x + (1-\theta)y) = \theta f(x) + (1-\theta)f(y),$$

so all affine functions are both convex and concave.

- Conversely, any function that is convex and concave is affine.

Convexity

- A function is convex if and only if it is convex when restricted to any line that intersects its domain:
- That is, f is convex if and only if $\forall x \in \textbf{dom } f, v \in \mathbf{R}^n$, the function $g(t) = f(x + tv)$ is convex on $\{t \mid x + tv \in \textbf{dom } f\}$.
- A convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

Extended-Value Extensions

Extended-Value Extensions

If f is convex we define its **extended-value extension** $\tilde{f} : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ by

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathbf{dom}\ f \\ \infty & x \notin \mathbf{dom}\ f \end{cases}$$

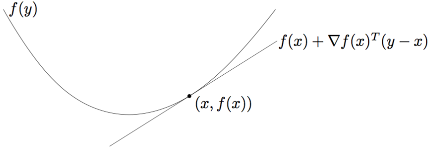
First-Order Conditions

First-Order Conditions

Suppose f is differentiable (implying that **dom** f is open). Then f is convex if and only if **dom** f is convex and

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

- Observation: the first-order Taylor approximation is a global underestimator of the function.
- Conversely, if the first-order Taylor approximation of a function is always a global underestimator of the function, then the function is convex.



First-Order Conditions

- A convex function f satisfies

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

  for all $x, y \in \textbf{dom } f$.
- This shows that from local information about a convex function (i.e., $f(x), \nabla f(x)$), we can derive global information (i.e., a global underestimator).
- Example: if $\nabla f(x) = 0$, then for all $y \in \textbf{dom } f$, $f(y) \geq f(x)$. (x is the global minimizer of f.)

First-Order Conditions – Strict Convexity, Concavity

First-Order Conditions for strict convexity

f is strictly convex if and only if **dom** f is convex and for $x, y \in$ **dom** $f, x \neq y$, we have

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

First-Order Conditions for (strict) concavity

f is concave if and only if **dom** f is convex and for $x, y \in$ **dom** f, we have

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

f is strictly concave if and only if **dom** f is convex and for $x, y \in$ **dom** $f, x \neq y$, we have

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

Proof of First-Order Conditions

Proof ideas:

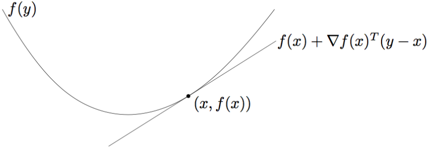
- Consider the special case $n = 1$ first.
  - Then we only need to prove that f is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x).$$

- For the general case $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\ f$ convex, consider the line passing by any two points $x, y \in \mathbf{dom}\ f, x \neq y$, and define a function $g : \mathbf{R} \to \mathbf{R}$ with $g(t) = f(ty + (1 - t)x)$.

Second-Order Conditions

- Assume that $f : \mathbf{R} \to \mathbf{R}$ is twice differentiable with $\mathbf{dom}\ f = \mathbf{R}$, then it is convex if and only if its second derivative is nonnegative.



Second-Order Conditions

- Assume that f is twice differentiable, that is, its **Hessian** or second derivative $\nabla^2 f$ exists at each point in **dom** f (open).

Second-Order Conditions

Then, f is convex if and only if **dom** f is convex and its Hessian is positive semidefinite:

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbf{dom}\ f.$$

- For a function on $\mathbf{R}$, this means $f''(x) \geq 0$, and **dom** f is convex.

Second-Order Conditions – Strict Convexity, Concavity

Second-Order Conditions for Concavity

A function f is concave if and only if **dom** f is convex and $\nabla^2 f(x) \preceq 0$ for all $x \in \textbf{dom } f$.

Second-Order Conditions for Strict Convexity

If $\nabla^2 f(x) \succ 0$ for all $x \in \textbf{dom } f$, then f is strictly convex.

- If f is strictly convex, do we have $\nabla^2 f(x) \succ 0$? (e.g., think $f(x) = x^4$)
- Is $f : \mathbf{R} \to \mathbf{R}, f(x) = 1/x^2$ a convex function? Why?

Example – Quadratic Functions

- Consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\ f = \mathbf{R}^n$, given by

$$f(x) = (1/2)x^T P x + q^T x + r,$$

with $P \in \mathbf{S}^n, q \in \mathbf{R}^n$, and $r \in \mathbf{R}$.
  - Note that $\nabla^2 f(x) = P$.
- The function f is convex if and only if $P \succeq 0$.
- The function f is concave if and only if $P \preceq 0$.
- The function f is strictly convex if and only if $P \succ 0$.
- The function f is strictly concave if and only if $P \prec 0$.

Example Convex functions on $\mathbf{R}$

- Exponential: e^{ax} is convex on $\mathbf{R}$, for any $a \in \mathbf{R}$.
- Powers: x^a is convex on $\mathbf{R}_{++}$ when $a \geq 1$ or $a \leq 0$; it is concave when $0 \leq a \leq 1$.
- Powers of absolute value: $|x|^p$ with $p \geq 1$ is convex on $\mathbf{R}$.
- Logarithm: $\log x$ is concave on $\mathbf{R}_{++}$.
- Negative entropy: $x \log x$ is convex on $\mathbf{R}_{++}$ (and also on $\mathbf{R}_{+}$ if defined as 0 for $x = 0$).

Example Convex Functions on $\mathbf{R}^n$

- Norms. Every norm on $\mathbf{R}^n$ is convex.

Norms

A function $f : \mathbf{R}^n \to \mathbf{R}$ (with $\mathbf{dom}\ f = \mathbf{R}^n$) is called a **norm** if for any $x, y \in \mathbf{R}^n, t \in \mathbf{R}$, we have

- $f(x) \geq 0$ (f is nonnegative).
- $f(x) = 0$ only if $x = 0$ (f is definite).
- $f(tx) = |t| f(x)$ (f is homogeneous).
- $f(x + y) \leq f(x) + f(y)$ (f satisfies the triangle inequality).

Example Convex Functions on $\mathbf{R}^n$

- Norms. Every norm on $\mathbf{R}^n$ is convex.
- Max function. $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbf{R}^n$.
- Quadratic-over-linear function. The function $f(x, y) = x^2/y$, with $\mathbf{dom}\ f = \mathbf{R} \times \mathbf{R}_{++} = \left\{(x, y) \in \mathbf{R}^2 \mid y > 0\right\}$, is convex.
- Log-sum-exp. The function $f(x) = \log(e^{x_1} + \ldots + e^{x_n})$ is convex on $\mathbf{R}^n$.
  - Note that $\max\{x_1, \ldots, x_n\} \leq f(x) \leq \max\{x_1, \ldots, x_n\} + \log n$.
- Geometric mean. The geometric mean $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\ f = \mathbf{R}^n_{++}$.
- Log-determinant. The function $f(X) = \log\det X$ is concave on $\mathbf{dom}\ f = \mathbf{S}^n_{++}$.

More on Norms

Norms

A function $f : \mathbf{R}^n \to \mathbf{R}$ (with $\mathbf{dom}\ f = \mathbf{R}^n$) is called a **norm** if for any $x, y \in \mathbf{R}^n$, $t \in \mathbf{R}$, we have

- $f(x) \geq 0$ (f is nonnegative).
- $f(x) = 0$ only if $x = 0$ (f is definite).
- $f(tx) = |t| f(x)$ (f is homogeneous).
- $f(x + y) \leq f(x) + f(y)$ (f satisfies the triangle inequality).

l_p-norm

Let $p \geq 1$. Then the **l_p-norm** is defined as

$$||x||_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

Question: When $p < 1$, is $||x||_p$ still a norm?

Examples of l_p-norm

- When $p = 2$, the l_2-norm is actually the Euclidean norm:

$$||x||_2 = (x_1^2 + \cdots + x_n^2)^{1/2}.$$

- When $p = 1$, the l_1-norm is the sum-absolute-value:

$$||x||_1 = |x_1| + \cdots + |x_n|.$$

- When $p \to \infty$, the l_∞-norm is:

$$||x||_\infty \triangleq \lim_{p \to \infty} ||x||_p = \lim(|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

It can be shown that $||x||_\infty = \max\{|x_1|, \ldots, |x_n|\}$.

Other Examples of Norms

- For $P \in \mathbf{S}^n_{++}$, the ***P*-quadratic norm** is defined as

$$||x||_P = (x^T P x)^{1/2} = ||P^{1/2} x||_2.$$

  - The **unit ball** of a quadratic norm,

$$\{x \in \mathbf{R}^n \mid ||x||_P \leq 1\},$$

    is an ellipsoid.

- The **Frobenius norm**, defined on $\mathbf{R}^{m \times n}$, is

$$||X||_F = (\mathbf{tr}\ X^T X)^{1/2} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^2 \right)^{1/2}.$$

Norms and Max function

- If $f : \mathbf{R}^n \to \mathbf{R}$ is a norm, and $0 \leq \theta \leq 1$, then

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

since f satisfies the triangle inequality and f is homogeneous.

  - Therefore any norm is convex.

- The function $f(x) = \max_i x_i$ is convex since

$$\begin{aligned} \max_i(\theta x_i + (1-\theta)y_i) &\leq \max_i \theta x_i + \max_i (1-\theta)y_i \\ &= \theta \max_i x_i + (1-\theta)\max_i y_i. \end{aligned}$$

  - In addition, $f(|x|) = \max_i |x_i|$ is a norm.

Quadratic-Over-Linear Function

- The quadratic-over-linear function $f : \mathbf{R}^2 \to \mathbf{R}, \mathbf{dom}\ f = \mathbf{R} \times \mathbf{R}_{++}$, $f(x, y) = x^2/y$, is convex since:

$$\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0.$$

Log-Sum-Exp

- The log-sum-exp function $f(x) = \log(e^{x_1} + \ldots + e^{x_n})$ is convex on $\mathbf{R}^n$ since

$$\nabla^2 f(x) = \frac{1}{(\mathbf{1}^T z)^2}\left((\mathbf{1}^T z)\mathbf{diag}\,(z) - zz^T\right),$$

where $z = (e^{x_1}, \ldots, e^{x_n})$, and

- for all v,

$$v^T \nabla^2 f(x) v = \frac{1}{(\mathbf{1}^T z)^2}\left(\left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n v_i^2 z_i\right) - \left(\sum_{i=1}^n v_i z_i\right)^2\right) \geq 0.$$

Geometric mean

- The **geometric mean** function $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\ f = \mathbf{R}^n_{++}$ since its Hessian $\nabla^2 f(x)$ can be shown to be negative semidefinite.
- Note that

$$\frac{\partial f(x)}{\partial x_k} = \frac{1}{n}\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{\prod_{i=1}^n x_i} \cdot \prod_{i=1, i\neq k}^{n} x_i = \frac{1}{n}\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{x_k}$$

and

$$\frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2 x_k^2}, \quad \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2 x_k x_l} (k \neq l)$$

Geometric mean

- The **geometric mean** function $f(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$ is concave on $\mathbf{dom}\ f = \mathbf{R}^n_{++}$ since its Hessian $\nabla^2 f(x)$ can be shown to be negative semidefinite.
- So,

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2}\left(n\ \mathbf{diag}\ (1/x_1^2, \ldots, 1/x_n^2) - qq^T\right)$$

where $q_i = 1/x_i$
- For any $v \in \mathbf{R}^n$, we have

$$v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2}\left(n \sum_{i=1}^{n} v_i^2/x_i^2 - \left(\sum_{i=1}^{n} v_i/x_i\right)^2\right) \leq 0$$

Log-Determinant

- The function $f : \mathbf{S}^n \to \mathbf{R}, f(X) = \log\det X$, with **dom** $= \mathbf{S}^n_{++}$ is concave.
- Proof idea: consider an arbitrary line in $\mathbf{S}^n$ (that passes through some point in $\mathbf{S}^n_{++}$) given by $X = Z + tV$, where $Z \in \mathbf{S}^n_{++}, V \in \mathbf{S}^n$, and define $g(t) = f(Z + tV)$, **dom** $g = \{t \mid Z + tV \succ 0\}$.
- Then it can be shown that

$$g(t) = \sum_{i=1}^{n} \log(1 + t\lambda_i) + \log\det Z$$

where λ_i are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$.

- So,

$$g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0.$$

Sublevel sets

Sublevel Sets

The **α-sublevel set** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$C_\alpha = \{x \in \textbf{dom } f \mid f(x) \leq \alpha\}.$$

Sublevel sets of a convex function are convex

If f is a convex function, then for any $\alpha \in \mathbf{R}$, the α-sublevel set, C_α, is convex.

- The converse is not true. A function can have all its sublevel sets convex, but not be a convex function. (e.g., $f(x) = -e^x$.)
- If f is concave, then its **α-superlevel** set, given by $\{x \in \textbf{dom } f | f(x) \geq \alpha\}$, is a convex set.

Sublevel sets – Example

Example

The geometric and arithmetic means of $x \in \mathbf{R}_+^n$ are

$$G(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad A(x) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

respectively. Suppose $0 \leq \beta \leq 1$, then the set

$$\left\{x \in \mathbf{R}_+^n \mid G(x) \geq \beta A(x)\right\}$$

is convex since it is the 0-superlevel set of the concave function $G(x) - \beta A(x)$.

- It is also a convex cone.

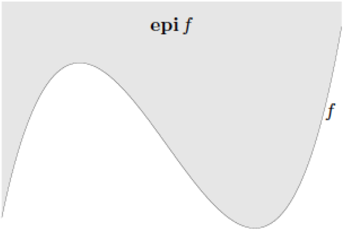
Epigraph

Graph

The **graph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as $\{(x, f(x)) \mid x \in \mathbf{dom}\ f\}$, a subset of $\mathbf{R}^{n+1}$.

Epigraph

The **epigraph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as
$\mathbf{epi}\ f = \{(x, t) | x \in \mathbf{dom}\ f, f(x) \leq t\}$, which is a subset of $\mathbf{R}^{n+1}$.



Epigraph

Graph

The **graph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as $\{(x, f(x)) \mid x \in \mathbf{dom}\ f\}$, a subset of $\mathbf{R}^{n+1}$.

Epigraph

The **epigraph** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as $\mathbf{epi}\ f = \{(x, t) | x \in \mathbf{dom}\ f, f(x) \leq t\}$, which is a subset of $\mathbf{R}^{n+1}$.

The epigraph of convex functions

A function is convex if and only if its epigraph is a convex set.

The epigraph of concave functions

A function is concave if and only if its **hypograph**, defined as $\mathbf{hypo}\ f = \{(x, t) \mid x \in \mathbf{dom}\ f, f(x) \geq t\}$, is a convex set.

Matrix fractional function

- The function $f : \mathbf{R}^n \times \mathbf{S}^n \to \mathbf{R}$, defined as

$$f(x, Y) = x^T Y^{-1} x,$$

  is called a **matrix fractional function**, and is convex on $\mathbf{dom}\ f = \mathbf{R}^n \times \mathbf{S}^n_{++}$.
- Proof:

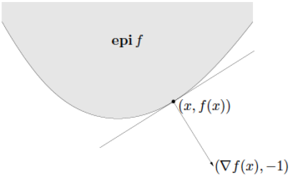
$$\begin{aligned} \mathbf{epi}\ f &= \left\{(x, Y, t) \mid Y \succ 0, x^T Y^{-1} x \leq t\right\} \\ &= \left\{(x, Y, t) \mid Y \succ 0, \begin{bmatrix} Y & x \\ x^T & t \end{bmatrix} \succeq 0\right\} \end{aligned}$$

  is a convex set.

Epigraph and first-order condition for convexity

- If $(y, t) \in \textbf{epi } f$, then
$$t \geq f(y) \geq f(x) + \nabla f(x)^T (y - x),$$
implying
$$\begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0.$$

- This means that the hyperplane defined by $(\nabla f(x), -1)$ supports **epi** f at the boundary point $(x, f(x))$;



Jensen's Inequality

- The basic inequality for convex functions

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

is called **Jensen's inequality**.

- Jensen's inequality can be extended to more than two points: If f is convex, $x_1, \ldots, x_k \in \textbf{dom } f$, and $\theta_1, \ldots, \theta_k \geq 0$ with $\theta_1 + \ldots + \theta_k = 1$, then

$$f(\theta_1 x_1 + \ldots + \theta_k x_k) \leq \theta_1 f(x_1) + \ldots + \theta_k f(x_k).$$

Jensen's Inequality

- Extension to infinite sum:

$$f\left(\int_S p(x)x dx\right) \leq \int_S f(x)p(x)dx,$$

with $p(x) \geq 0$ on S, $\int_S p(x)dx = 1, S \subseteq \textbf{dom } f$.

- If x is a random variable such that Prob $(x \in \textbf{dom } f) = 1$, then

$$f(\mathbf{E}x) \leq \mathbf{E}f(x).$$

- Suppose $x \in \textbf{dom } f \subset \mathbf{R}^n$ and $z \in \mathbf{R}^n, \mathbf{E}(z) = 0$. Then we have

$$\mathbf{E}f(x+z) \geq f(x).$$

.

Inequalities

- Many famous inequalities can be derived by applying Jensen's inequality on some convex functions.
- The arithmetic-geometric mean inequality: $(a+b)/2 \geq \sqrt{ab}$.
- Noting that $-\log x$ is convex, and letting $\theta = 1/2$, we obtain

$$-\log \frac{a+b}{2} \leq \frac{-\log a - \log b}{2},$$

  implying the AM-GM inequality: $\sqrt{ab} \leq \dfrac{a+b}{2}$.
- Further, by taking

$$a = \frac{x_i^2}{\sum_{j=1}^n x_j^2}, b = \frac{y_i^2}{\sum_{j=1}^n y_j^2},$$

  and summing over i, we get the Cauchy's inequality

$$\left(\sum_{j=1}^n x_j y_j\right)^2 \leq \left(\sum_{j=1}^n x_j^2\right)\left(\sum_{j=1}^n y_j^2\right).$$

Inequalities

- Apply the Jensen's inequality on the function $-\log x$ again, with an arbitrary θ, $0 < \theta < 1$, we get an inequality more general than the AM-GM inequality:
$$a^{\theta} b^{1-\theta} \leq \theta a + (1-\theta) b.$$
- If we take $\theta = 1/p$, where $p > 1$. Let $q = 1/(1-\theta)$, then $q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.
- By taking
$$a = \frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}, b = \frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q},$$
and summing over i, we obtain the Hölder's inequality
$$\sum_{j=1}^n x_j y_j \leq \left(\sum_{j=1}^n |x_j|^p \right)^{1/p} \left(\sum_{j=1}^n |y_j|^q \right)^{1/q}.$$