COMP6704 Lecture 3

Advanced Topics in Optimization

Convex Functions

Fall, 2022

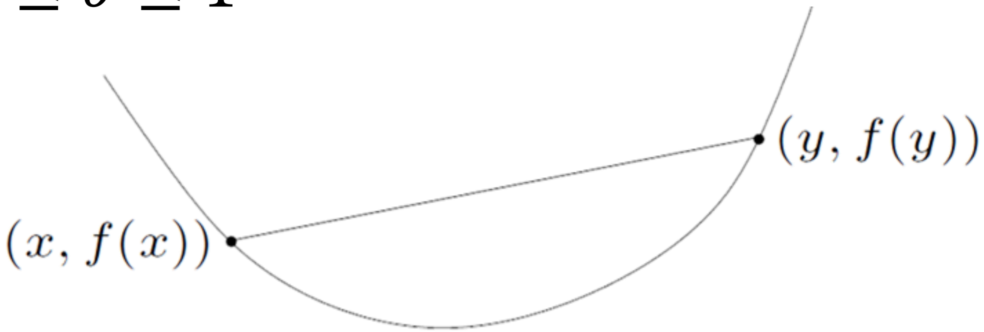
Instructor: WU，Xiao-Ming

Definition

$f: \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

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

for all $x, y \in \mathbf{dom}\, f, 0 \le \theta \le 1$

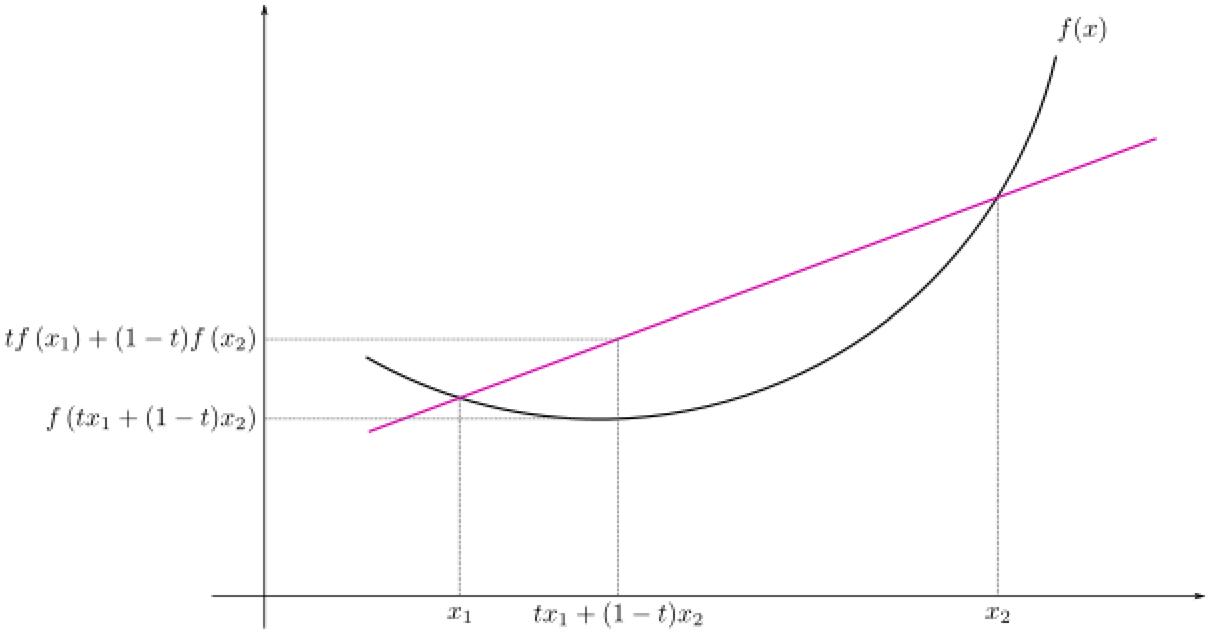


- $-f$ is concave if f is convex
- f is strictly convex if $\mathbf{dom}\, f$ is convex and

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

for $x, y \in \mathbf{dom}\, f, x \neq y, 0 < \theta < 1$

Convex Functions



Convex Function

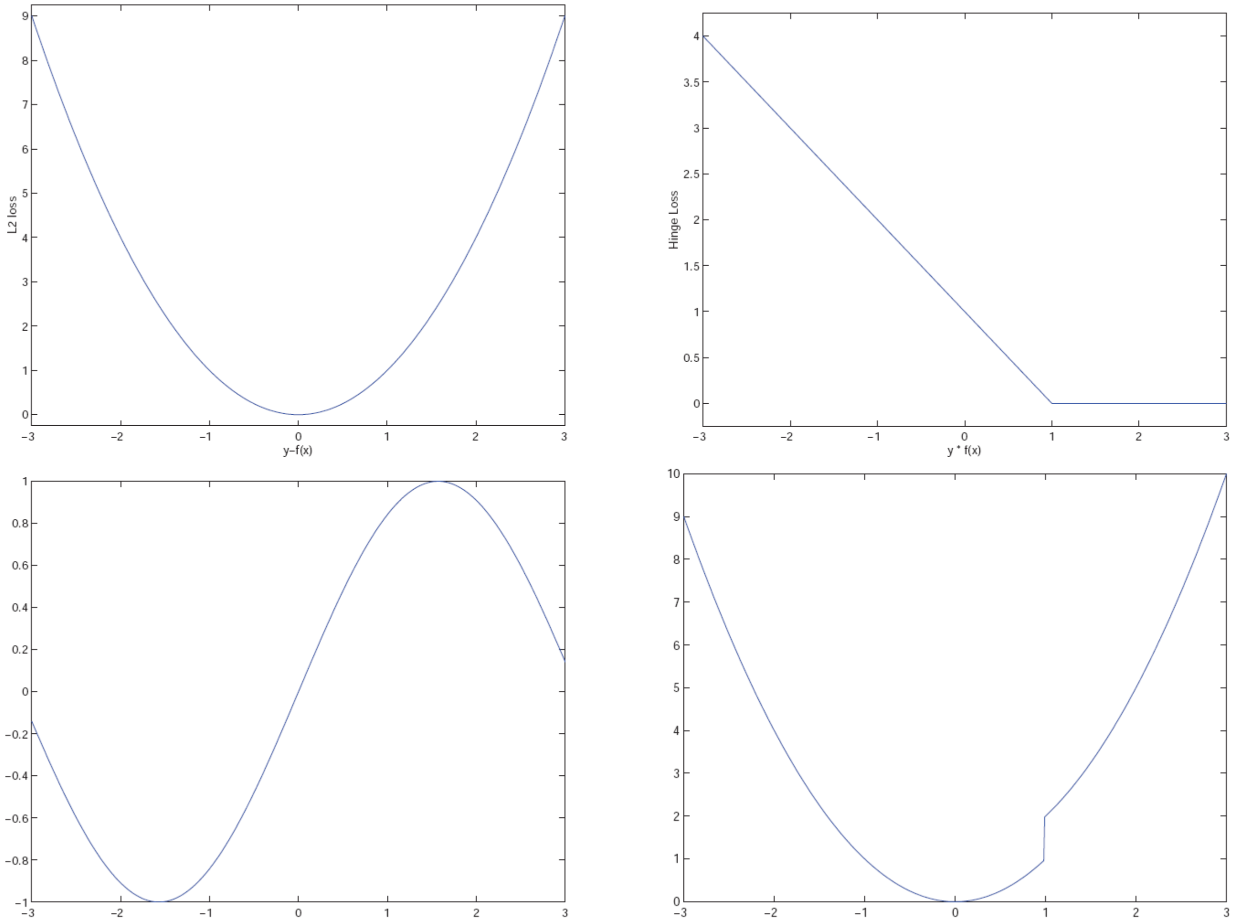
- Definition: A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if:

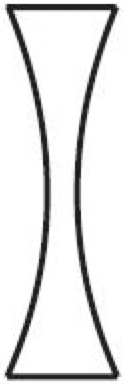
1. For any x_1 and x_2 in the domain of f, for any $\lambda \in [0,1]$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

2. The line segment connecting two points $f(x_1)$ and $f(x_2)$ lies entirely on or above the function f.
3. The set of points lying on or above the function f is convex.

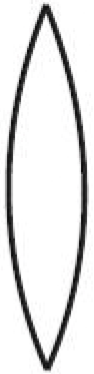
Which are Convex Functions?





Concave

- **affine**: $ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$
- **powers** x^{α} on $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$
- **logarithm**: $\log x$ on $\mathbb{R}_{++}$



Convex

- **affine**:$ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$
- **exponential**: e^{ax}, for any $a \in \mathbb{R}$
- **powers**: x^{α} on $\mathbb{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- **powers of absolute value**: $|x|^{p}$ on $\mathbb{R}$, for $p \geq 1$
- **negative entropy**: $x \log x$ on $\mathbb{R}_{++}$

First-order Condition

f is **differentiable** if **dom** f is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)$$

exists at $x \in$ **dom** f

1st order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \text{ for all } x, y \in \textbf{dom } f$$

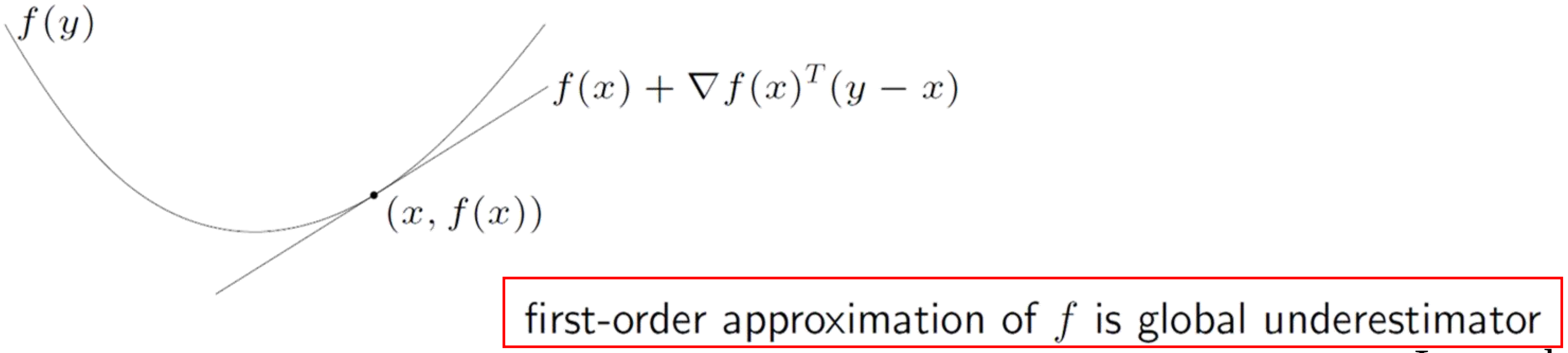


Image by A. d'Aspremont

The Hessian

Suppose that $f: \mathbb{R}^n \to \mathbb{R}$ is a function that takes a vector in $\mathbb{R}^n$ and returns a real number. Then the **Hessian matrix** with respect to x, written as $\nabla_x^2 f(x)$ or simply as H is the $n \times n$ matrix of partial derivatives,

$$\nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

In other words, $\nabla_x^2 f(x) \in \mathbb{R}^{n\times n}$, with

$$\left(\nabla_x^2 f(x)\right)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}.$$

Note that the Hessian is always symmetric, since

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.$$

Similar to the gradient, the Hessian is defined only when $f(x)$ is real-valued.

Second-order Conditions

f is **twice differentiable** if **dom** f is open and the Hessian $\nabla^2 f(x) \in \mathbb{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, i, j = 1, \ldots, n,$$

exists at each $x \in \mathbf{dom}\, f$.

2nd order conditions: for twice differentiable f with convex domain.

- f is convex if and only if

$$\nabla^2 f(x) \succeq 0 \text{ for all } x \in \mathbf{dom}\, f,$$

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

Exercise

- **Exponential.** Let $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = e^{ax}$ for any $a \in \mathbb{R}$. To show f is convex, we can simply take the second derivative $f''(x) = a^2 e^{ax}$, which is positive for all x.

- **Negative logarithm.** Let $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = -\log x$ with domain $\mathcal{D}(f) = \mathbb{R}_{++}$ (here, $\mathbb{R}_{++}$ denotes the set of strictly positive real numbers, $\{x : x > 0\}$). Then $f''(x) = 1/x^2 > 0$ for all x.

- **Affine functions.** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = b^T x + c$ for some $b \in \mathbb{R}^n$, $c \in \mathbb{R}$. In this case the Hessian, $\nabla_x^2 f(x) = 0$ for all x. Because the zero matrix is both positive semidefinite and negative semidefinite, f is both convex and concave. In fact, affine functions of this form are the *only* functions that are both convex and concave.

Are Norms Convex?

- **norm**: a function $\|\cdot\|$ that satisfies
  - $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$
  - $\|tx\| = |t|\|x\|$ for $t \in \mathbb{R}$
  - $\|x + y\| \leq \|x\| + \|y\|$
- **norms**: Let $f: \mathbb{R}^n \to \mathbb{R}$ be some norm on $\mathbb{R}^n$. Then by the triangle inequality and positive homogeneity of norms, for $x, y \in \mathbb{R}^n, 0 \leq \theta \leq 1$.

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

- This is an example of a convex function where it is not possible to prove convexity based on the second-order or first-order conditions because norms are not generally differentiable everywhere (e.g., the 1-nrom, $\|x\|_1 = \sum_{i=1}^{n} |x_i|$, is non-differentiable at all points where any x_i is equal to zero).

Quadratic Functions

A quadratic function is one that can be expressed as

$$q(x) = x \cdot Ax + b \cdot x + \alpha,$$

where $A \in \mathbb{S}^n$, the space of real symmetric $n \times n$ matrices, $b \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$.

Exercise

- Show that the function

$$q(x) = (x_1 - x_2)^2 + (x_1 + 2x_2 + 1)^2 - 8x_1x_2$$

is a quadratic function by finding the appropriate A, b, α.

- **Solution**: If you expand $q(x)$, you can see that it's a quadratic function with the following parameters:

$$A = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \alpha = 1.$$

- Compute the gradient and Hessian of $q(x)$ and express them using the A, b, α computed above in Item 1.

- **Solution**: It is easy to check that $\nabla q(x) = 2Ax + b$ and $\nabla^2 q(x) = 2A$ which is true for any quadratic function.

Exercise

- Show that a quadratic function $q(x)$ is convex if, and only if, A is positive semidefinite and it is strictly convex if A is positive definite.

- Suppose that $q(x)$ is a quadratic function of n variables such that the corresponding matrix A is positive definite. Show that $0 = 2Ax + b$ has a unique solution and that this solution is the strict global minimizer of $q(x)$.

- **Solution**: A is positive definite, so is invertible. Then $2Ax + b = 0$ has the unique solution $x^* = -\frac{1}{2}A^{-1}b$. x^* is a critical point and $q(x)$ is strictly convex, so x^* is the unique global minimizer.

Quadratic Function

Quadratic functions. Let $f: \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{2}x^T A x + b^T x + c$ for a symmetric matrix $A \in \mathbb{S}^n, b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Hessian for this function is given by

$$\nabla_x^2 f(s) = A.$$

Therefore, the convexity or non-convexity of f is determined entirely by whether or not A is positive semidefinite: if A is positive semidefinite then the function is convex (and analogously for strictly convex, concave, strictly concave); if A is indefinite then f is neither convex nor concave.

Note that the squared Euclidean norm $f(x) = \|x\|_2^2 = x^T x$ is a special case of quadratic functions where $A = I, b = 0, c = 0$, so it is therefore a strictly convex function.

Convex or Strictly Convex?

□ $f(x,y) = 5x^2 + 2xy + y^2 - x + 2y + 3$ on $D = \mathbb{R}^2$.

- **Solution**: Hessian of the function is

$$\nabla^2 f(x,y) = \begin{bmatrix} 10 & 2 \\ 2 & 2 \end{bmatrix},$$

which is positive definite, so the function is strictly convex.

□ $f(x_1,x_2) = c_1x_1 + \frac{c_2}{x_1} + c_3x_2 + \frac{c_4}{x_2}$ on $D = \mathbb{R}^2_{++}, c_i > 0$,

- **Solution**: The Hessian of the function is

$$\nabla^2 f(x_1,x_2) = \begin{bmatrix} 2c_2{x_1}^{-3} & 0 \\ 0 & 2c_4{x_2}^{-3} \end{bmatrix},$$

which is positive definite for $(x_1,x_2) \in \mathbb{R}^2_{++}$ and $c_i > 0$, so the function is strictly convex on the domain.

Epigraph and Sublevel Set

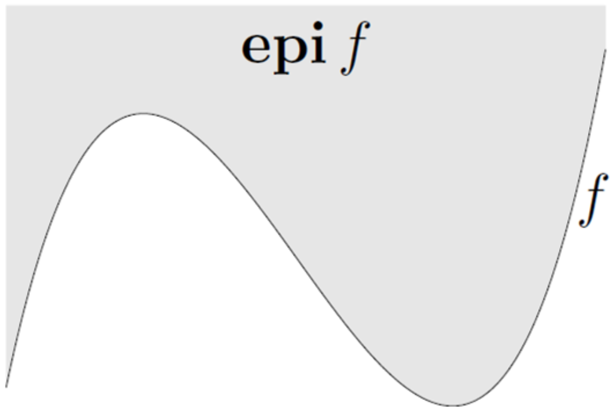
- α-**sublevel set** of $f: \mathbb{R}^n \to \mathbb{R}$:

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

sublevel sets of convex functions are convex (converse is false)

- **epigraph of** $f: \mathbb{R}^n \to \mathbb{R}$:

$$\textbf{epi}\, f = \{(x, t) \in R^{n+1} \mid x \in \textbf{dom}\, f, f(x) \leq t\}$$



Imaged by A. d'Aspremont

f is convex if and only if **epi** f is a convex set

Exercise

- Show that if $f(x)$ is a convex function, then the set

$$L = \{x \in \mathbb{R}^n \mid f(x) \leq b\}$$

is a convex set.

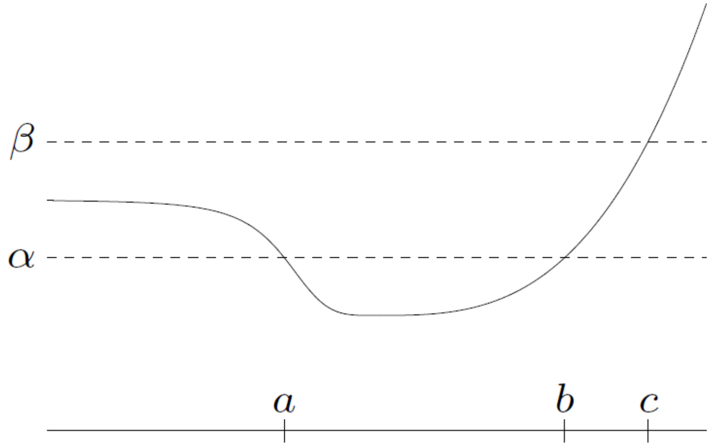
- **Solution**: We need to prove that for every $x_1, x_2 \in L$ the point $\alpha x_1 + (1-\alpha)x_2$ is also in L.

Quasiconvex Functions

$f: \mathbb{R}^n \to \mathbb{R}$ is quasiconvex if $\mathbf{dom}\, f$ is convex and the sublevel sets

$$S_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

are convex for all α



- f is quasiconcave if - f is quasiconvex
- f is quasilinear if it is quasiconvex and quasiconcave

Jensen's Inequality

basic inequality: if f is convex, then for $0 \leq \theta \leq 1$.

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

extension: if f is convex, then

$$f(\boldsymbol{E}z) \leq \boldsymbol{E}f(z)$$

for any random variable z.

Basic inequality is a special case with discrete distribution:

$$\boldsymbol{Prob}(z = x) = \theta, \qquad \boldsymbol{Prob}(z = y) = 1 - \theta$$

Many famous inequalities can be derived by applying Jensen's inequality to some appropriate convex function. (Indeed, convexity and Jensen's inequality can be made the foundation of a theory of inequalities.) As a simple example, consider the arithmetic-geometric mean inequality:

$$\sqrt{ab} \le \frac{a+b}{2} \qquad (2)$$

for $a, b \geq 0$. The function $-\log x$ is convex; Jensen's inequality with $\theta = \frac{1}{2}$ yields

$$-\log\left(\frac{a+b}{2}\right) \le \frac{-\log a - \log b}{2}.$$

Taking the exponential of both sides yields (2).

As a less trivial example we prove Holder's inequality: for $p > 1, \frac{1}{p} + \frac{1}{q} = 1$, and $x, y \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} x_i y_i \le \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^q\right)^{\frac{1}{q}}.$$

Exercise

- Use Jensen's inequality to show that if f is convex and $a_i > 0$ then we have the following.

$$\sum_i a_i f(x_i) \geq \left(\sum_i a_i\right) f\left(\frac{\sum_i a_i x_i}{\sum_i a_i}\right)$$

Solution: Define Z as a categorical random variable (the categorical distribution is a special case of the multinomial distribution) and then apply Jensen's Inequality.

Operations that Preserve Convexity

- Practical methods for establishing convexity of a function

1. Verify definition (often simplified by restricting to a line)
2. For twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
3. Show that f is obtained from simple convex functions by operations that preserve convexity
    - nonnegative weighted sum
    - composition with affine function
    - pointwise maximum and supremum
    - composition
    - minimization
    - perspective

Exercise

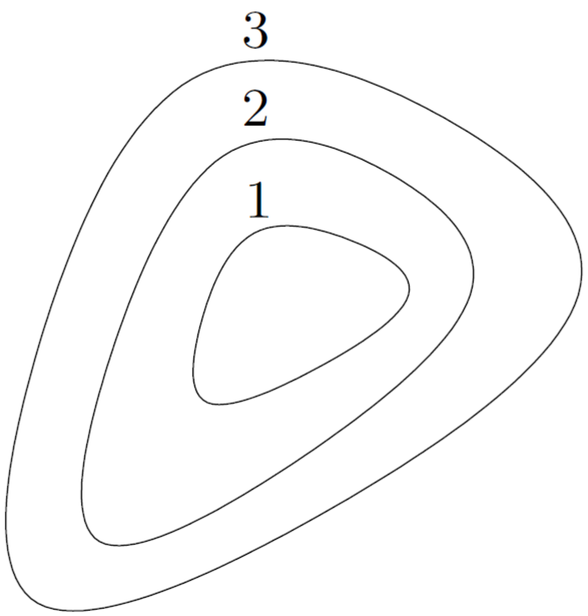
- **Nonnegative weighted sums of convex functions.** Let $f_1, f_2, \ldots, f_k$ be convex functions and $w_1, w_2, \ldots, w_k$ be nonnegative real numbers. Then

$$f(x) = \sum_{i=1}^{k} w_i f_i(x)$$

is a convex function, since

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \sum_{i=1}^{k} w_i f_i(\theta x + (1-\theta)y) \\ &\leq \sum_{i=1}^{k} w_i(\theta f_i(x) + (1-\theta) f_i(y)) \\ &= \theta \sum_{i=1}^{k} w_i f_i(x) + (1-\theta) \sum_{i=1}^{k} w_i f_i(y) \\ &= \theta f(x) + (1-\theta) f(x). \end{aligned}$$

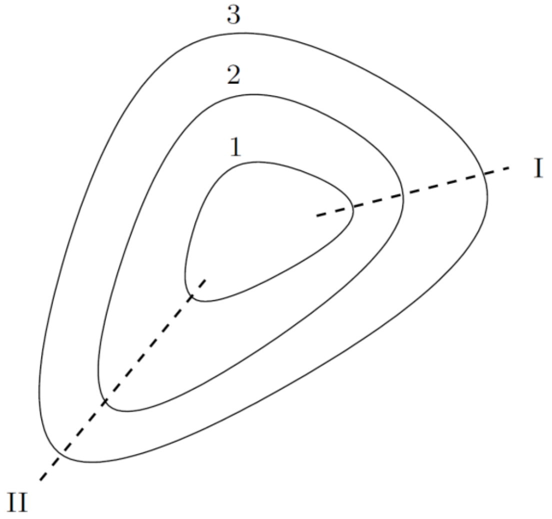
Exercise

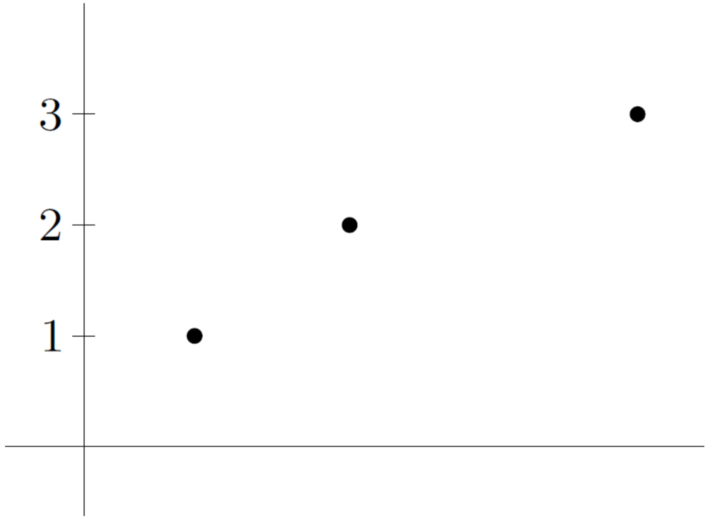


Could f be convex (concave, quasiconvex, quasiconcave)?

Exercise

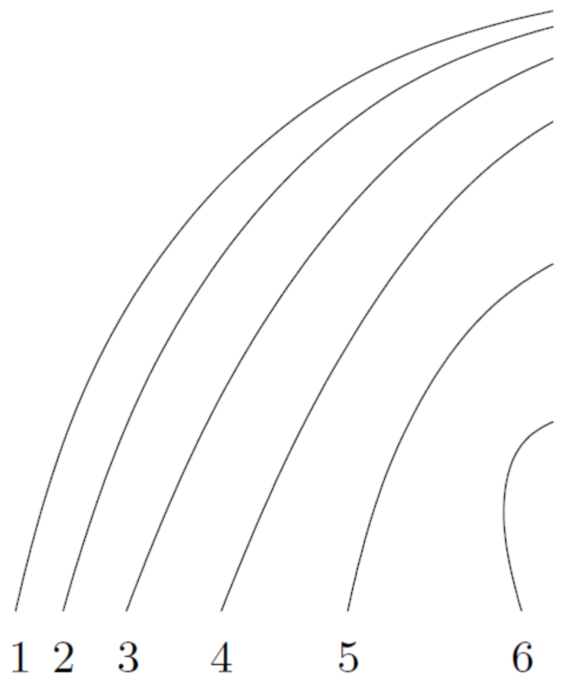
- **Solution**: The first function could be quasiconvex because the sublevel sets appear to be convex. It is definitely not concave or quasiconcave because the superlevel sets are not convex. It is also not convex, for the following reason. We plot the function values along the dashed line labeled I.





- Along this line the function passes through the points marked as black dots in the figure below. Clearly along this line segment, the function is not convex.

Exercise



If we repeat the same analysis for the second function, we see that it could be concave (and therefore it could be quasiconcave). It cannot be convex or quasiconvex, because the sublevel sets are not convex.

Exercise

Negative entropy: $x \log x$ (either on $\mathbb{R}_{++}$, or on $\mathbb{R}_{+}$, defined as 0 for $x = 0$) is convex.

Convexity or concavity of these examples can be shown by verifying the basic inequality (1), or by checking that the second derivative is nonnegative or nonpositive. For example, with $f(x) = x \log x$ we have

$$f'(x) = \log x + 1, \qquad f''(x) = \frac{1}{x},$$

So that $f''(x) > 0$ for $x > 0$. This shows that the negative entropy function is (strictly) convex.

Exercise

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

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

The inequality follows from the triangle inequality, and the equality follows from homogeneity of a norm.

Exercise

- **Max function**. The function $f(x) = \max_i x_i$ satisfies, for $0 \le \theta \le 1$,

$$
\begin{aligned}
f(\theta x + (1-\theta)y) &= \max_i(\theta x_i + (1-\theta)y_i) \\
&\le \theta \max_i x_i + (1-\theta) \max_i y_i \\
&= \theta f(x) + (1-\theta) f(y) \, .
\end{aligned}
$$

- Quadratic-over-linear functions. The function $f(x, y) = \frac{x^2}{y}$, with

$$\mathbf{dom}\, f = \mathbb{R} \times \mathbb{R}_{++} = \{(x, y) \in \mathbb{R}^2 | y > 0\},$$

is convex

- To show that the quadratic-over-linear function $f(x, y) = \frac{x^2}{y}$ is convex, we note that (for $y > 0$),

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

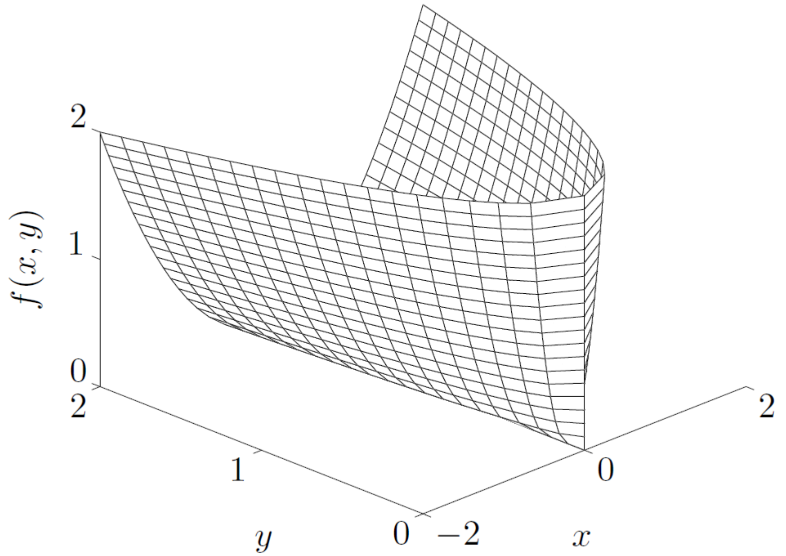


Figure 3.3 Graph of $f(x, y) = x^2/y$.

- **Log-sum-exp**. The function $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$.
- Log-sum-exp. The Hessian of the log-sum-exp function is

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

Where $z = (e^{x_1}, \dots, e^{x_n})$. To verify that $\nabla^2 f(x) \succeq 0$ we must show that for all v, $v^T \nabla^2 f(x) v \geq 0$, i.e.,

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

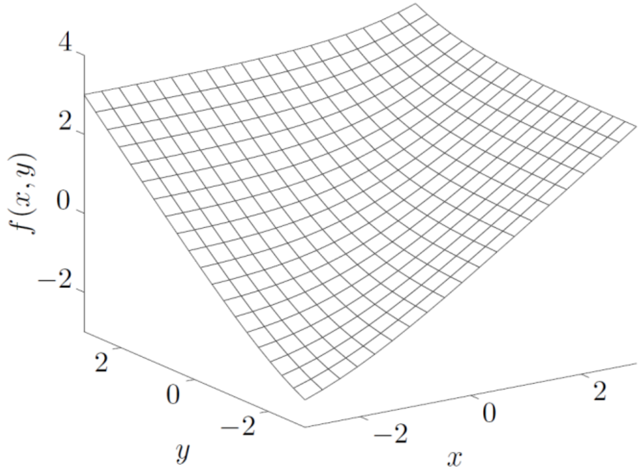


Figure 3.4 Graph of $f(x, y) = \log(e^x + e^y)$.

But this follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$ applied to the vectors with components $a_i = v_i\sqrt{z_i}, b_i = \sqrt{z_i}$.

Composition with an Affine Mapping

- Suppose $f: \mathbb{R}^n \to \mathbb{R}, A \in \mathbb{R}^{n \times m}$, and $b \in \mathbb{R}^n$. Define $g: \mathbb{R}^m \to \mathbb{R}$ by
$$g(x) = f(Ax + b),$$
- With **dom** $g = \{x |\ Ax + b \in \textbf{dom}\ f\}$. Then if f is convex, so is g; if f is concave, so is g.

Pointwise Maximum and Supremum

If f_1 and f_2 are convex functions then their pointwise maximum f, defined by

$$f(x) = \max\{f_1(x), f_2(x)\},$$

With **dom** f=**dom** $f_1 \cap$ **dom** f_2, is also convex. This property is easily verified: if $0 \leq \theta \leq 1$ and $x, y \in$ **dom** f, then

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max_i\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\ &\leq \max\{\theta f_1(x) + (1-\theta)f_1(y), \theta f_2(x) + (1-\theta)f_2(y)\} \\ &\leq \theta \max\{f_1(x), f_2(x)\} + (1-\theta)\max\{f_1(y), f_2(y)\} \\ &= \theta f(x) + (1-\theta)f(y), \end{aligned}$$

Example 3.5 *Piecewise-linear functions.* The function

$$f(x) = \max\{a_1^T x + b_1, \ldots, a_L^T x + b_L\}$$

defines a piecewise-linear (or really, affine) function (with L or fewer regions). It is convex since it is the pointwise maximum of affine functions.

Composition

- In this section we examine conditions on $h: \mathbb{R}^k \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}^k$ that guarantee convexity or concavity of their composition $f = h \circ g: \mathbb{R}^n \to \mathbb{R}$, defined by

$$f(x) = h\big(g(x)\big), \qquad \textbf{dom}\, f = \{x \in \textbf{dom}\, g \mid g(x) \in \textbf{dom}\, h\}.$$

- We first consider the case $k = 1$, so $h: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}$. We can restrict ourselves to the case $n = 1$ (since convexity is determined by the behavior of a function on arbitrary lines that intersect its domain).
- To discover the composition rules, we start by assuming that h and g are twice differentiable, with $\textbf{dom}\, g = \textbf{dom}\, h = \mathbb{R}$. In this case, convexity of f reduces to $f'' \geq 0$ (meaning, $f''(x) \geq 0$ for all $x \in \mathbb{R}$)

- The second derivative of the composition function $f = h \circ g$ is given by

$$f''(x) = h''\big(g(x)\big)g'(x)^2 + h'\big(g(x)\big)g''(x). \tag{3}$$

- Now suppose, for example, that g is convex (so $g'' \geq 0$) and h is convex and nondecreasing (so $h'' \geq 0$ and $h' \geq 0$). It follows from (3) that $f'' \geq 0$, i.e., f is convex. In a similar way, the expression (4) gives the results:
  - f is convex if h is convex and nondecreasing, and g is convex,
  - f is convex if h is convex and nonincreasing, and g is concave, (4)
  - f is concave if h is concave and nondecreasing, and g is concave,
  - f is concave if h is concave and nonincreasing, and g is convex.

Exercise

(a) $f(x) = \|Ax - b\|$, where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $\|\cdot\|$ is a norm on $\mathbf{R}^m$.

Solution. f is the composition of a norm, which is convex, and an affine function.

(a) $f(x) = \max_{i=1,\ldots,k} \|A^{(i)}x - b^{(i)}\|$, where $A^{(i)} \in \mathbf{R}^{m \times n}$, $b^{(i)} \in \mathbf{R}^m$ and $\|\cdot\|$ is a norm on $\mathbf{R}^m$.

Solution. f is the pointwise maximum of k functions $\|A^{(i)}x - b^{(i)}\|$. Each of those functions is convex because it is the composition of an affine transformation and a norm.

Exercise

(a) $f(x) = -\log(-\log(\sum_{i=1}^m e^{a_i^T x + b_i}))$ on $\mathbf{dom}\, f = \{x \mid \sum_{i=1}^m e^{a_i^T x + b_i} < 1\}$. You can use the fact that $\log(\sum_{i=1}^n e^{y_i})$ is convex.

Solution. $g(x) = \log(\sum_{i=1}^m e^{a_i^T x + b_i})$ is convex (composition of the log-sum-exp function and an affine mapping), so $-g$ is concave. The function $h(y) = -\log y$ is convex and decreasing. Therefore $f(x) = h(-g(x))$ is convex.