COMP7180: Quantitative Methods for Data Analytics and Artificial Intelligence

Lecture 6: Convex Optimization: Theory II

Jun Qi

Research Assistant Professor in Computer Science @ Hong Kong Baptist University

Affiliated Associate Professor in Electronic Engineering @ Fudan University

22/10/2024

CVX Course Scope

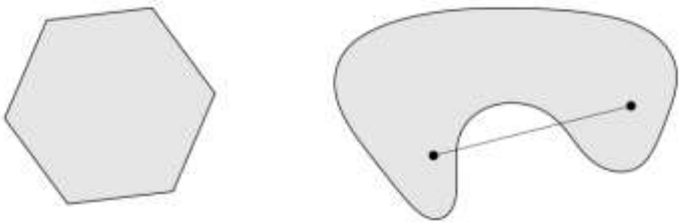
- Convex sets and convex functions
- Convex function properties: 1^{st} and 2^{nd} characterizations
- Conjugate functions
- Duality and optimality (Lagrangian method)

Convex sets

Convex set: $C \subseteq \mathbb{R}^n$ such that

$$x, y \in C \implies tx + (1-t)y \in C \text{ for all } 0 \leq t \leq 1$$

In words, line segment joining any two elements lies entirely in set



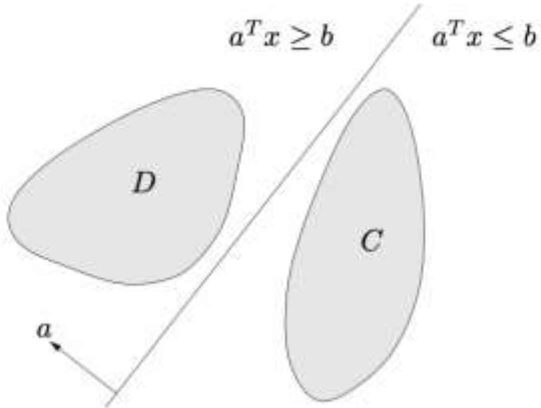
Convex combination of $x_1, \ldots, x_k \in \mathbb{R}^n$: any linear combination

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

with $\theta_i \geq 0$, $i = 1, \ldots, k$, and $\sum_{i=1}^k \theta_i = 1$. Convex hull of a set C, $\mathrm{conv}(C)$, is all convex combinations of elements. Always convex

Key properties of convex sets

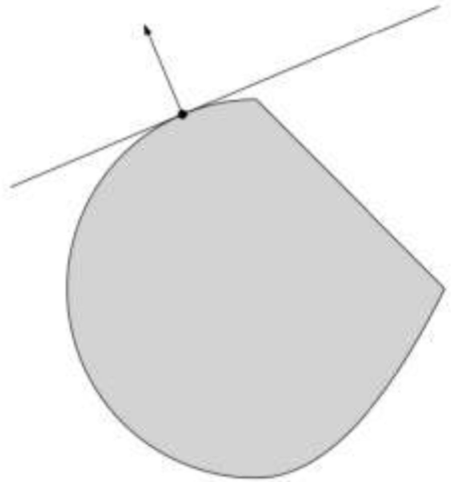
- Separating hyperplane theorem: two disjoint convex sets have a separating between hyperplane them



Formally: if C, D are nonempty convex sets with $C \cap D = \emptyset$, then there exists a, b such that

$$C \subseteq \{x : a^T x \leq b\}$$
$$D \subseteq \{x : a^T x \geq b\}$$

- Supporting hyperplane theorem: a boundary point of a convex set has a supporting hyperplane passing through it



Formally: if C is a nonempty convex set, and $x_0 \in \mathrm{bd}(C)$, then there exists a such that

$$C \subseteq \{x : a^T x \leq a^T x_0\}$$

Operations preserving convexity

- Intersection: the intersection of convex sets is convex
- Scaling and translation: if C is convex, then

$$aC + b = \{ax + b : x \in C\}$$

  is convex for any a, b

- Affine images and preimages: if $f(x) = Ax + b$ and C is convex then

$$f(C) = \{f(x) : x \in C\}$$

  is convex, and if D is convex then

$$f^{-1}(D) = \{x : f(x) \in D\}$$

  is convex

Exercises:

- Let's analyze the convexity for each of the sets:

(a) A slab. A slab is defined as a set of the form:

$$\mathcal{A} = \{ x \in R^n \mid \alpha \le a^T x \le \beta \}.$$

Solution:

To show that this set is convex, consider any two points $x_1, x_2 \in \mathcal{A}$. For any $\lambda \in [0,1]$, we need to show that the linear combination $\lambda x_1 + (1-\lambda)x_2$ is also in $\mathcal{A}$.

$$\alpha \le a^T x_1 \le \beta,$$
$$\alpha \le a^T x_2 \le \beta.$$

Consider the point $\lambda x_1 + (1-\lambda)x_2$:

$$\begin{aligned} a^T(\lambda x_1 + (1-\lambda)x_2) &= \lambda a^T x_1 + (1-\lambda)a^T x_2 \\ &\ge \lambda\alpha + (1-\lambda)\alpha \\ &= \alpha \\ &\le \beta. \end{aligned}$$

Therefore, $\alpha \le a^T(\lambda x_1 + (1-\lambda)x_2) \le \beta$, and $\mathcal{A}$ is convex.

Exercises:

- Let's analyze the convexity for each of the sets:

(b) A rectangle: A rectangle (or hyperrectangle) is defined as:

$$\mathcal{B} = \{ x \in R^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \dots, n \}.$$

Solution:

Again, consider any two points $x_1, x_2 \in \mathcal{B}$ and any $\lambda \in [0, 1]$. We need to show that $\lambda x_1 + (1 - \lambda)x_2 \in \mathcal{B}$.

For each coordinate i:

$$\alpha_i \leq x_{1i} \leq \beta_i,$$
$$\alpha_i \leq x_{2i} \leq \beta_i.$$

Consider the i-th coordinate of $\lambda x_1 + (1 - \lambda)x_2$:

$$\alpha_i \leq \lambda x_{1i} + (1 - \lambda)x_{2i} \leq \beta_i,$$

since x_{1i} and x_{2i} are individually bounded by α_i and β_i. Therefore, $\mathcal{B}$ is convex.

Exercises:

- Let's analyze the convexity for each of the sets:

(c) The set of points closer to a given point than a given set, i.e.,

$$\mathcal{D} = \{\, x \mid |x - x_0|_2 \le |x - y|_2 \text{ for all } y \in S \,\}.$$

Solution:

Consider two points $x_1, x_2 \in \mathcal{D}$ and any $\lambda \in [0,1]$. We need to show that $\lambda x_1 + (1-\lambda)x_2 \in \mathcal{D}$.

Let's examine the distance relations for any point $y \in S$:

$$\|x_1 - x_0\|_2 \le \|x_1 - y\|_2,$$
$$\|x_2 - x_0\|_2 \le \|x_2 - y\|_2.$$

Using the triangle inequality:

$$\begin{aligned}\|\lambda x_1 + (1-\lambda)x_2 - x_0\|_2 &\le \lambda\|x_1 - x_0\|_2 + (1-\lambda)\|x_2 - x_0\|_2 \\ &\le \lambda\|x_1 - y\|_2 + (1-\lambda)\|x_2 - y\|_2 \\ &\le \|\lambda x_1 + (1-\lambda)x_2 - y\|_2.\end{aligned}$$

Therefore, $\lambda x_1 + (1-\lambda)x_2 \in \mathcal{D}$, and $\mathcal{D}$ is convex.

DEPARTMENT OF COMPUTER SCIENCE 計算機科學系

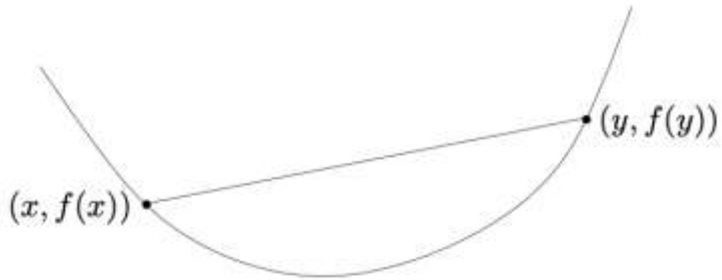
香港浸會大學 HONG KONG BAPTIST UNIVERSITY

Convex functions

Convex function: $f : \mathbb{R}^n \to \mathbb{R}$ such that $\mathrm{dom}(f) \subseteq \mathbb{R}^n$ convex, and

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) \quad \text{for } 0 \le t \le 1$$

and all $x, y \in \mathrm{dom}(f)$

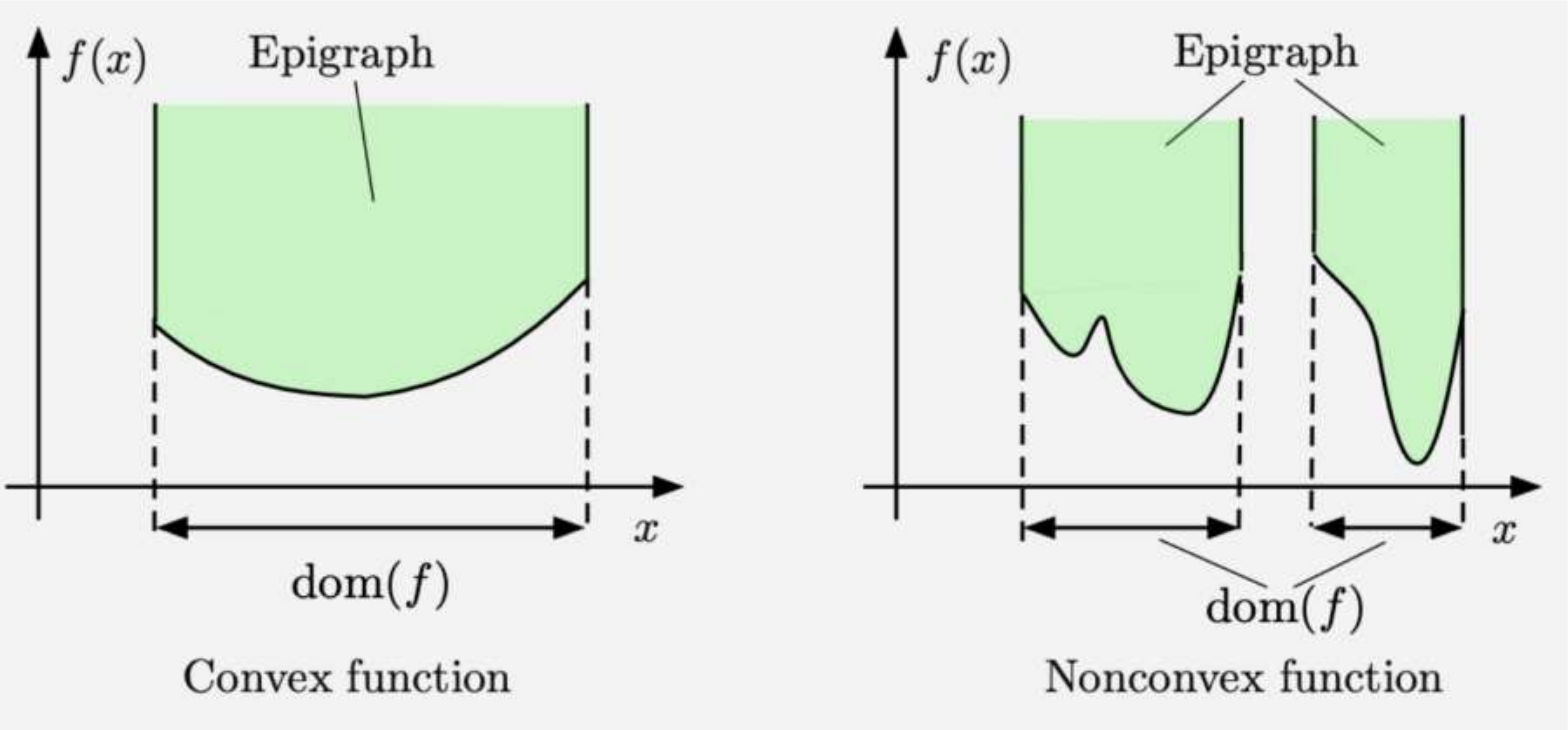


In words, function lies below the line segment joining $f(x), f(y)$

Concave function: opposite inequality above, so that

$$f \text{ concave} \iff -f \text{ convex}$$

Examples:



$f(x)$
Epigraph
x
dom(f)
Convex function
$f(x)$
Epigraph
x
dom(f)
Nonconvex function

Important modifiers:

- Strictly convex: $f\big(tx + (1-t)y\big) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$. In words, f is convex and has greater curvature than a linear function
- Strongly convex with parameter $m > 0$: $f - \frac{m}{2}\|x\|_2^2$ is convex. In words, f is at least as convex as a quadratic function

Note: strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex

(Analogously for concave functions)

Examples of convex functions

- Univariate functions:
  - Exponential function: e^{ax} is convex for any a over $\mathbb{R}$
  - Power function: x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ (nonnegative reals)
  - Power function: x^a is concave for $0 \leq a \leq 1$ over $\mathbb{R}_+$
  - Logarithmic function: $\log x$ is concave over $\mathbb{R}_{++}$
- Affine function: $a^T x + b$ is both convex and concave
- Quadratic function: $\frac{1}{2}x^T Q x + b^T x + c$ is convex provided that $Q \succeq 0$ (positive semidefinite)
- Least squares loss: $\|y - Ax\|_2^2$ is always convex (since $A^T A$ is always positive semidefinite)

Exercises:

Suppose $f: R \to R$ is convex, and $a, b \in dom\ f$ with $a < x < b$.

Show that: $$f(x) \le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

Solution:

Since f is convex, by the definition of convexity, we have:

$$f(\lambda a + (1-\lambda)b) \le \lambda f(a) + (1-\lambda) f(b)$$

where $\lambda = \frac{b-x}{b-a}$ and $1 - \lambda = \frac{x-a}{b-a}$. Notice that $\lambda \in [0,1]$ since $x \in [a,b]$.

Now, substituting $\lambda = \frac{b-x}{b-a}$, we get:

$$x = \lambda a + (1-\lambda)b = \frac{b-x}{b-a} a + \frac{x-a}{b-a} b$$

Then, using the convexity condition:

$$f\left(\frac{b-x}{b-a} a + \frac{x-a}{b-a} b\right) \le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

Since x is a convex combination of a and b, we can substitute back x:

$$f(x) \le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

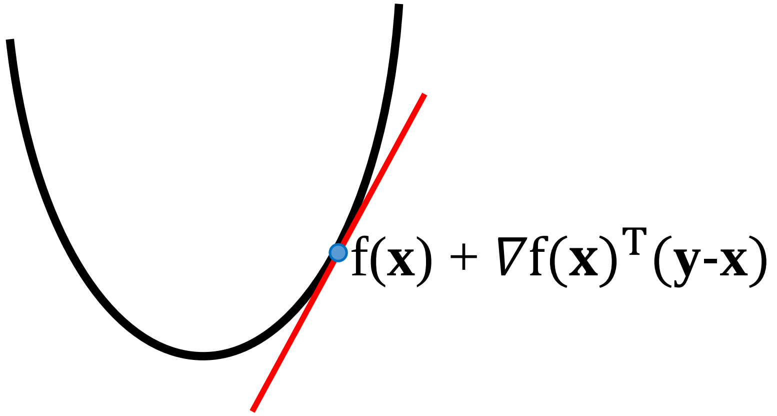
First-order Convexity Condition

When $f(\mathbf{x})$ is differentiable, can we check whether $f(\mathbf{x})$ is convex by the gradient? Following theorem gives the answer:

Assume that $f(\mathbf{x})$ is differential, then $f(\mathbf{x})$ is convex

if and only if

the domain C is convex and $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})$.



First-order Convexity Condition

Theorem 1. Assume that f($\mathbf{x}$) is differentiable, then f($\mathbf{x}$) is convex

if and only if

the domain C is convex and $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})$.

The proof can be found in Proposition 4 in
https://wiki.math.ntnu.no/_media/tma4180/2016v/note2.pdf

We use the linear function to check the result.

$f(\mathbf{x}) = \mathbf{Ax}+\mathbf{b}$, then $\nabla f(\mathbf{x})^T = \mathbf{A}$.

$f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) = \mathbf{Ax}+\mathbf{b}+\mathbf{A}(\mathbf{y}-\mathbf{x}) = \mathbf{Ay}+\mathbf{b}$

So $f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})$

Second-order Convexity Condition

When f(**x**) is twice differentiable, how can we check whether f(**x**) is convex?
Answer: We use Hessian Matrix.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_d)^T$, then the Hessian Matrix of f(**x**) is a dxd matrix, for the ij-th element in the matrix is
the second-order partial derivatives of f: $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_d \partial x_d} \end{bmatrix}$$

Second-order Convexity Condition

- If the second-order partial derivatives are continuous functions, then the Hessian Matrix is symmetric, i.e., $\mathbf{H}(\mathbf{x}) = \mathbf{H}(\mathbf{x})^T$

Example: $f(x_1, x_2) = x_1^2 + x_2^2$

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} = 2, \quad \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_2} = 2, \quad \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} = 0, \quad \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} = 0$$

So

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Exercises

Please Compute:

- the Hessian Matrix of $f(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2$

- the Hessian Matrix of $f(x_1, x_2) = 2x_1^3 + 6x_1x_2 + x_2^2 + x_2^3$

Exercises

Solution:

- the Hessian Matrix of $f(x_1, x_2) = x_1{}^2 + 2x_1x_2 + x_2{}^2$

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} = 2, \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_2} = 2, \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} = 2, \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} = 2$$

So

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

Exercises

Solution:

the Hessian Matrix of $f(x_1, x_2) = 2x_1^3 + 6x_1x_2 + x_2^2 + x_2^3$

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} = 12x_1, \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_2} = 2+6x_2, \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} = 6, \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} = 6$$

So

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} 12x_1 & 6 \\ 6 & 2+6x_2 \end{bmatrix}$$

Second-order Convexity Condition

Theorem 2. Assume that f($\mathbf{x}$) is twice differentiable, then f($\mathbf{x}$) is convex

if and only if

the domain C is convex and the Hessian Matrix $\mathbf{H}(\mathbf{x})$ is positive semi-definite.

The proof can be found in Proposition 7 in *https://wiki.math.ntnu.no/_media/tma4180/2016v/note2.pdf*

- What is positive semi-definite matrix **M**?

Positive semi-definite matrix **M** is a nxn sysmetric matrix $\mathbf{M} = \mathbf{M}^{\mathrm{T}}$ and for any real n-dimensional vector **z,** $\mathbf{z}^{\mathrm{T}}\mathbf{M}\mathbf{z} \geq 0$.

Second-order Convexity Condition

Examples of positive semi-definite matrix:

- **M** = $\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ is a positive semi-definite matrix, if a≥0 and b≥0

Because for any (x,y), (x,y)**M**$(x, y)^T$ = ax^2+by^2 ≥0

- Please check that **M** = $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$ is a positive semi-definite matrix

Second-order Convexity Condition

- Please check that **M** = $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$ is a positive semi-definite matrix

For any (x,y,z), we can check that

$$(x,y,z)\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}(x,y,z)^T = 2x^2\text{-}2xy+2y^2\text{-}2yz+2z^2$$

$$= x^2 + (x-y)^2 + (z-y)^2 + z^2 \geq 0$$

Examples

- Please verify the following univariate functions are convex.

Exponential function: $f(x) = e^{ax}$

Solution:

First derivative:

$$f(x) = e^{ax}$$
$$f'(x) = ae^{ax}$$

Second derivative:

$$f''(x) = a^2e^{ax}$$

Since $e^{ax} > 0$ for all $x \in \mathbb{R}$, the sign of $f''(x) = a^2e^{ax}$ depends on a^2.

For any real a, $a^2 \geq 0$. Therefore, $f''(x) = a^2e^{ax} \geq 0$.

This means e^{ax} is convex for any a over $\mathbb{R}$.

Examples

- Please verify the following univariate functions are convex.

Power function: $f(x) = x^a$ is convex for $a \geq 1$ or $a \leq 0$ over R_+.

Solution:

First derivative:

$$f(x) = x^a$$
$$f'(x) = ax^{a-1}$$

Second derivative:

$$f''(x) = a(a-1)x^{a-2}$$

We analyze the sign of $f''(x)$:

For $x > 0$:

If $a \geq 1$:

$a \geq 1$ and $x^{a-2} > 0$, hence $f''(x) = a(a-1)x^{a-2} \geq 0$.

If $a \leq 0$:

$a - 1 \leq -1$, hence $x^{a-2} \geq 0$. Therefore, $f''(x) = a(a-1)x^{a-2} \geq 0$.

This means x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$.

Examples

- Please verify the following univariate functions are convex.

Logarithmic function: $f(x) = \log x$

Solution:

First derivative:

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

Second derivative:

$$f''(x) = -\frac{1}{x^2}$$

For $x > 0$:

$f''(x) = -\frac{1}{x^2} \leq 0$

This means $\log x$ is concave over $\mathbb{R}_{++}$ (the set of positive real numbers).

Optimization terminology

Reminder: a convex optimization problem (or program) is

$$\begin{aligned} \min_{x \in D} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where f and g_i, $i = 1, \ldots, m$ are all convex, and the optimization domain is $D = \mathrm{dom}(f) \cap \bigcap_{i=1}^m \mathrm{dom}(g_i)$ (often we do not write D)

- f is called criterion or objective function
- g_i is called inequality constraint function
- If $x \in D$, $g_i(x) \leq 0$, $i = 1, \ldots, m$, and $Ax = b$ then x is called a feasible point
- The minimum of $f(x)$ over all feasible points x is called the optimal value, written $f^\star$

- If x is feasible and $f(x) = f^\star$, then x is called optimal; also called a solution, or a minimizer[1]
- If x is feasible and $f(x) \leq f^\star + \epsilon$, then x is called ϵ-suboptimal
- If x is feasible and $g_i(x) = 0$, then we say g_i is active at x
- Convex minimization can be reposed as concave maximization

$$
\begin{array}{ll}
\min\limits_{x} & f(x) \\
\text{subject to} & g_i(x) \leq 0, \\
 & \quad i = 1, \ldots, m \\
 & Ax = b
\end{array}
\quad \Longleftrightarrow \quad
\begin{array}{ll}
\max\limits_{x} & -f(x) \\
\text{subject to} & g_i(x) \leq 0, \\
 & \quad i = 1, \ldots, m \\
 & Ax = b
\end{array}
$$

Both are called convex optimization problems

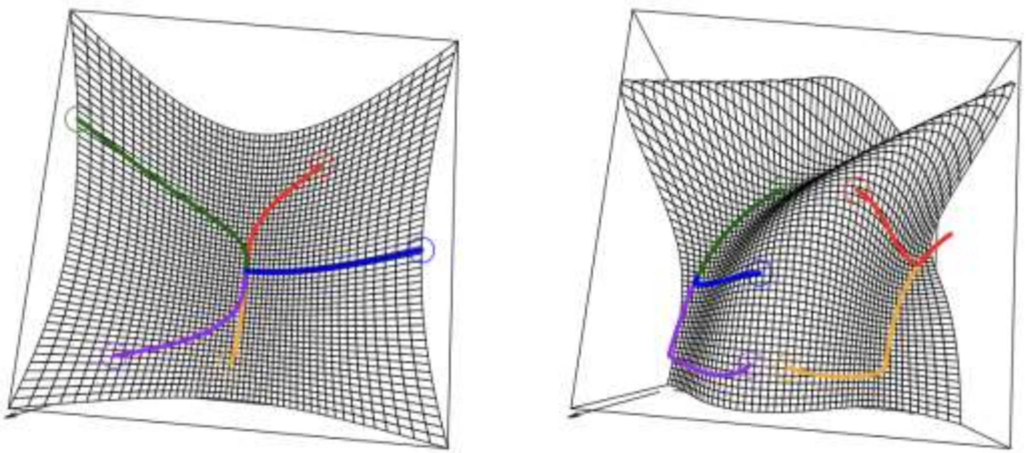
Local minima are global minima

For a convex problem, a feasible point x is called locally optimal is there is some $R > 0$ such that

$$f(x) \le f(y) \quad \text{for all feasible } y \text{ such that } \|x - y\|_2 \le R$$

Reminder: for convex optimization problems, local optima are global optima

Proof simply follows from definitions



Convex   Nonconvex

Example: lasso

Given $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, consider the lasso problem:

$$\begin{aligned} \min_{\beta} \quad & \|y - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \le s \end{aligned}$$

Is this convex? What is the criterion function? The inequality and equality constraints? Feasible set? Is the solution unique, when:

- $n \ge p$ and X has full column rank?
- $p > n$ ("high-dimensional" case)?

How do our answers change if we changed criterion to Huber loss:

$$\sum_{i=1}^{n} \rho(y_i - x_i^T \beta), \quad \rho(z) = \begin{cases} \frac{1}{2} z^2 & |z| \le \delta \\ \delta |z| - \frac{1}{2}\delta^2 & \text{else} \end{cases} \quad ?$$

Example: support vector machines

Given $y \in \{-1, 1\}^n$, $X \in \mathbb{R}^{n \times p}$ with rows $x_1, \ldots x_n$, consider the support vector machine or SVM problem:

$$\begin{array}{ll} \min\limits_{\beta, \beta_0, \xi} & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n} \xi_i \\ \text{subject to} & \xi_i \geq 0, \ i = 1, \ldots, n \\ & y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i, \ i = 1, \ldots, n \end{array}$$

Is this convex? What is the criterion, constraints, feasible set? Is the solution (β, β_0, ξ) unique? What if changed the criterion to

$$\frac{1}{2}\|\beta\|_2^2 + \frac{1}{2}\beta_0^2 + C\sum_{i=1}^{n} \xi_i^{1.01}?$$

For original criterion, what about β component, at the solution?

Continuing CVX Next Week

- Convex sets and convex functions
- Convex function properties: 1st and 2nd characterizations
- Conjugate functions
- Duality and optimality (Lagrangian method)