MAT3007 Optimization
Lecture 14 Optimality Conditions
Convexity

Yuang Chen

School of Data Scienc
The Chinese University of Hong Kong, Shenzhen

July 7, 2025

Agenda

- Optimality Conditions: July 7
- Convexity: July 7, 9
- Lagrangian Dual and KKT Conditions: July 10, 14
- Algorithms: July 16, 17
- Final Review: July 21
- Final Exam: July 24 (1:30 - 4:30 pm)
- HW 5 due July 13
- HW 6 due July 20

Outline

Outline

Optimality Conditions for Minimization

We consider the unconstrained minimization problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$$

Theorem (FONC for Minimization)

If $\boldsymbol{x}^*$ is a local minimizer of f for the unconstrained problem, then we must have $\nabla f(\boldsymbol{x}^*) = 0$.

Theorem (SONC for Minimization)

If $\boldsymbol{x}^*$ is a local minimizer of f for the unconstrained problem, then we must have i) $\nabla f(\boldsymbol{x}^*) = 0$; ii) $\nabla^2 f(\boldsymbol{x}^*)$ is positive semi-definite. $\boxed{d^T\nabla^2 f(x)d \geq 0 \ \forall d}$

Theorem (SOSC for Minimization)

Let f be twice continuous differentiable. If $\boldsymbol{x}^*$ satisfies i) $\nabla f(\boldsymbol{x}^*) = 0$; ii) $\nabla^2 f(\boldsymbol{x}^*)$ is positive definite, then $\boldsymbol{x}^*$ is a local minimizer.

$\rightarrow d^T\nabla^2 f(x)d > 0, \ \forall d$

Optimality Conditions for Maximization

OWe consider the unconstrained minimization problem:

$$\max_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$$

Theorem (FONC for Maximization)

If $\boldsymbol{x}^*$ is a local maximizer of f for the unconstrained problem, then we must have $\nabla f(\boldsymbol{x}^*) = 0$.

Theorem (SONC for Maximization)

If $\boldsymbol{x}^*$ is a local maximizer of f for the unconstrained problem, then we must have i) $\nabla f(\boldsymbol{x}^*) = 0$; ii) $\nabla^2 f(\boldsymbol{x}^*)$ is negative semi-definite

Theorem (SOSC for Maximization)

Let f be twoice continuous differentiable. If $\boldsymbol{x}^*$ satisfies i) $\nabla f(\boldsymbol{x}^*) = 0$; ii) $\nabla^2 f(\boldsymbol{x}^*)$ is negative definite, then $\boldsymbol{x}^*$ is a local maximizer.

Critical and Saddle Points

critical Pts $\nabla f(x) = 0$

local min $\nabla^2 f(x) \succeq 0$

local max $\nabla^2 f(x) \preceq 0$

saddle Pt $\nabla^2 f(x)$ indefinite

Definition

- A point $\mathbf{x}$ satisfying $\nabla f(\mathbf{x}) = 0$ is called **critical point** or **stationary point**.
- A stationary point is called **saddle point** if it is neither a local minimizer nor a local maximizer.

Theorem

Suppose that $\mathbf{x}^*$ is a stationary point ($\nabla f(\mathbf{x}^*) = \mathbf{0}$) and that the Hessian $\nabla^2 f(\mathbf{x}^*)$ is indefinite, then $\mathbf{x}^*$ is a **saddle point**.

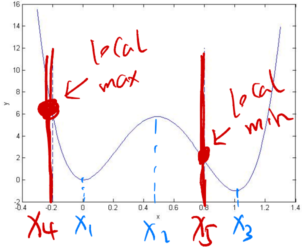
Outline

Constrained Problems

We have derived necessary and sufficient conditions for the local minimum for unconstrained problems.

- What is the difference between constrained and unconstrained problems?

Consider the example $f(x) = 100x^2(1-x)^2 - x$ with constraint $-0.2 \le x \le 0.8$.



In addition to the original local minimizer ($x_1 = 0.013$), there is one more local minimizer on the boundary ($x = 0.8$).

Constrained Problems

At the boundary ($x^* = 0.8$), the FONC is not satisfied

$$f'(0.8) < 0$$

However, at this point, in order to stay feasible, we can only go leftward. That is, in the Taylor expansion

$$\overset{<0}{d}$$

$$f(x^* + d) = f(x^*) + df'(x^*) + o(d)$$

we can only take d to be negative (otherwise it won't be feasible).

Thus $f(x^* + d) > f(x^*)$ in a small neighborhood of x^* in the feasible region. Thus x^* is a local minimizer.

It means that the developed FONC ($\nabla f(x) = 0$) for unconstrained problem is not enough for constrained problems.

Feasible Directions

Now we formalize the above arguments.

Definition (Feasible Direction)

Given $\boldsymbol{x} \in F$, we call $\boldsymbol{d}$ to be a *feasible direction* at $\boldsymbol{x}$ if there exists $\bar{\alpha} > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in F$ for all $0 \leq \alpha \leq \bar{\alpha}$.

some $\alpha \geq 0$

For example,

1 - If $F = \{\boldsymbol{x} | A\boldsymbol{x} = \boldsymbol{b}\}$, then the feasible directions at $\boldsymbol{x}$ is $\{\boldsymbol{d} | A\boldsymbol{d} = 0\}$
2 - If $F = \{\boldsymbol{x} | A\boldsymbol{x} \geq \boldsymbol{b}\}$, then the feasible directions at $\boldsymbol{x}$ is $\{\boldsymbol{d} | \boldsymbol{a}_i^T \boldsymbol{d} \geq 0 \text{ if } \boldsymbol{a}_i^T \boldsymbol{x} = b_i\}$

$$x \in F \Rightarrow Ax \geq b \Rightarrow a_i^T x \geq b_i \begin{cases} a_i^T x > b_i \\ a_i^T x = b_i \end{cases}$$

$$x+\alpha d \in F$$

1. $\Rightarrow A(x+\alpha d) = b$

$$\Rightarrow \underbrace{Ax}_{b} + \underbrace{\alpha}_{\geq 0} Ad = b$$

$$\Rightarrow Ad = 0$$

2. $x+\alpha d \in F$ $\qquad \begin{bmatrix} 0 \\ \\ \end{bmatrix} \geq \begin{bmatrix} 0 \\ \\ \end{bmatrix}$

$$\Rightarrow A(x+\alpha d) \geq b$$

$$\Rightarrow Ax + \alpha Ad \geq b$$

$$\Rightarrow (Ax+\alpha Ad)_i \geq b_i \quad \forall i$$

$$\Rightarrow a_i^T x + \alpha a_i^T d \geq b_i$$

Case 1: $a_i^T x = b_i$

$$\underbrace{\alpha}_{\geq 0} \underbrace{a_i^T d}_{v} \geq \underbrace{b_i - a_i^T x}_{0}$$

$$\Rightarrow \alpha v \geq 0 \Rightarrow v \geq 0$$

$$\Rightarrow a_i^T d \geq 0$$

$\rightarrow \{d : a_i^T d \geq 0 \text{ if } a_i^T x = b_i\}$

Case 2: $a_i^T x > b_i$

$$\alpha a_i^T d \geq \underbrace{b_i - a_i^T x}_{<0}$$

$\alpha v \geq -\delta$

$\alpha \geq \frac{-\delta}{v}$

We can always satisfy the above inequality by choosing α sufficiently small, no restriction on d.

FONC for Constrained Problems

Theorem (FONC for Constrained Problems)

If $\boldsymbol{x}^*$ is a local minimum of $\min_{x \in F} f(x)$, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, we must have $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0$

Remark

In unconstrained problems, all directions are feasible, thus we must have $\nabla f(\boldsymbol{x}^*) = 0$.

large small

$f(x+\alpha d) = f(x) + \alpha \nabla f(x)^T d + \cdots$

If x^* local min, then $f(x^*) \leq f(x^* + \alpha d)$

So $\underset{\geq 0}{\alpha} \nabla f(x^*)^T d \geq 0 \Rightarrow \nabla f(x^*)^T d \geq 0, \ \forall \text{feasible } d$

An Alternative View

Definition (Descent Direction)

Let f be continuously differentiable. Then $\boldsymbol{d}$ is called a *descent direction* at $\boldsymbol{x}$ if and only if $\nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0$.

$\Updownarrow$

Remark

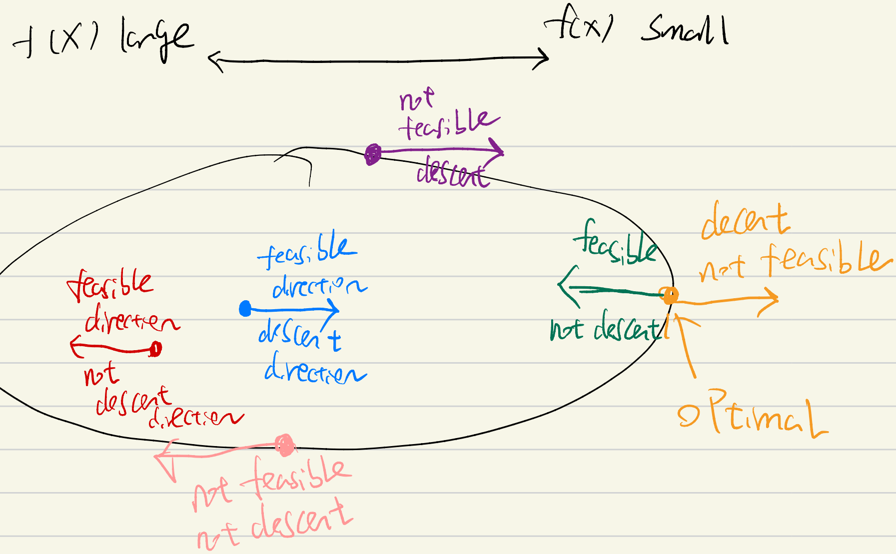
If $\boldsymbol{d}$ is a descent direction at $\boldsymbol{x}$, then there exists $\bar{\gamma} > 0$ such that $f(\boldsymbol{x} + \gamma \boldsymbol{d}) < f(\boldsymbol{x})$ for all $\underbrace{0 < \gamma \le \bar{\gamma}}_{\text{some } \gamma > 0}$.

If we denote the set of feasible directions at $\boldsymbol{x}$ by $S_F(\boldsymbol{x})$ and the set of descent directions at $\boldsymbol{x}$ by $S_D(\boldsymbol{x})$. Then the first order necessary condition can be written as:

feasible direction   descent direction

$$\rightarrow \quad \underline{S_F(\boldsymbol{x}^*) \cap S_D(\boldsymbol{x}^*) = \emptyset}$$

Or in other words, there cannot be any feasible descent directions.



At optimal point, no direction is both descent and feasible.

For algorithm design, I want to find a direction which is feasible and descent.

Nonlinear Optimization with Equality Constraints

linear

Consider

$$\text{minimize}_{\boldsymbol{x}} \quad f(\boldsymbol{x})$$
$$\text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}$$

- The feasible direction set is $\{\boldsymbol{d} | A\boldsymbol{d} = 0\}$.
- The descent direction set is $\{\boldsymbol{d} | \nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0\}$.

The FONC says that at local minimum, there cannot be a solution to both systems (feasible and descent direction)

Theorem (Alternative System)

The system $A\boldsymbol{d} = 0$ and $\nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0$ does not have a solution if and only if there exists $\boldsymbol{y}$ such that

$$A^T \boldsymbol{y} = \nabla f(\boldsymbol{x})$$

Therefore, the first-order necessary condition for

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & A\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{1}$$

is that there exists $\boldsymbol{y}$ such that

$$A^T \boldsymbol{y} = \nabla f(\boldsymbol{x})$$

then the system $\left\{ Ad = 0 \text{ and } \nabla f(x)^T d < 0 \right\}$ does not have solution

Theorem

If $\boldsymbol{x}^*$ is a local minimum for (1), then there must exist $\boldsymbol{y}$ such that

$$A^T \boldsymbol{y} = \nabla f(\boldsymbol{x}^*)$$

Proof: If x^* is local min for (1),

no direction is both feasible and descent.

So the system $\underbrace{Ad=0}_{\text{feasible}}$ and $\nabla f(x)^T d<0$

has no solution.

① If $\exists y$ s.t. $A^T y=\nabla f(x)$,

then $\begin{cases} Ad=0 \\ \nabla f(x)^T d<0 \end{cases}$ has no solution.

multiply d on both sides

$$(A^T y)^T d = (\nabla f(x))^T d$$

$$\Rightarrow (Ad)^T y = \nabla f(x)^T d$$

So $\begin{cases} Ad=0 \\ \nabla f(x)^T d<0 \end{cases}$ has no solution

② If $\begin{cases} Ad=0 \\ \nabla f(x)^T d<0 \end{cases}$ has no solution, then $\exists y$: $A^T y=\nabla f(x)$

Consider LP (P)

$$\min. \ \nabla f(x)^T d$$

$$(P) \quad \text{s.t.} \ Ad=0 \quad (y)$$

$$d \text{ free}$$

The dual for the above LP is

$$\max.\ 0^T y = 0$$

$$(D) \quad \text{s.t.}\ A^T y = \nabla f(x)$$

If $\begin{cases} Ad \geq 0 \\ \nabla f(x)^T d < 0 \end{cases}$ has no solution, then (P) is not unbounded.

Also since $d=0$ is a feasible soln for (P), so (P) is feasible.

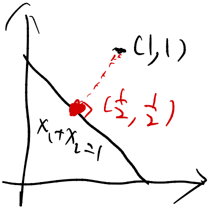
Thus (P) must have optimal soln.

By Strong duality, (D) is feasible and also has optimal soln. (P) and (D) have the same optimal obj value.

So $\exists\, y$ s.t. $A^T y = \nabla f(x)$

Example

$$
\begin{aligned}
\text{minimize} \quad & (x_1 - 1)^2 + (x_2 - 1)^2 \\
\text{s.t.} \quad & x_1 + x_2 = 1
\end{aligned}
$$



This problem finds the nearest point on the line $x_1 + x_2 = 1$ to the point $(1, 1)$

$$f(x) = (x_1-1)^2 + (x_2-1)^2$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 - 2 \\ 2x_2 - 2 \end{bmatrix}$$

$x_1 + x_2 = 1$

$Ax = b$

$A = [1 \quad 1]$ $\quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\quad b = 1$

At local min x^*, I know $\exists y$ s.t.

$$A^T y = \nabla f(x^*)$$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} y = \begin{bmatrix} 2x_1^* - 2 \\ 2x_2^* - 2 \end{bmatrix}$$

$$\begin{cases} 2x_1^* - 2 = y \\ 2x_2^* - 2 = y \end{cases} \Rightarrow 2x_1^* - 2 = 2x_2^* - 2$$

$$\Downarrow$$

$$\begin{cases} x_1^* = x_2^* \\ x_1^* + x_2^* = 1 \end{cases}$$

$$\Downarrow$$

$$\begin{cases} x_1^* = \frac{1}{2} \\ x_2^* = \frac{1}{2} \end{cases}$$ is local min

Another Example

$$\min \|X\beta - y\|_2^2 \Rightarrow \beta^* = (X^T X)^{-1} X^T y$$

Consider a constrained version of the least squares problem:

$$\begin{aligned} \text{minimize}_{\boldsymbol{\beta}} \quad & ||X\boldsymbol{\beta} - \boldsymbol{y}||_2^2 \\ \text{s.t.} \quad & W\boldsymbol{\beta} = \xi \end{aligned}$$

$$f(\beta) = \|X\beta - y\|_2^2$$

$$\nabla f(\beta) = 2X^T(X\beta - y) = 2X^T X\beta - 2X^T y$$

If β is local optimal, then $\exists\, z$ s.t.

$$W^T z = \nabla f(\beta)$$

$$\Rightarrow W^T z = 2X^T X\beta - 2X^T y$$

$$\Rightarrow \underset{n\times m}{X^T}\underset{m\times n}{X}\underset{n\times 1}{\beta} - \frac{1}{2}\underset{n\times d}{W^T}\underset{d\times 1}{z} = \underset{n\times m}{X^T}\underset{m\times 1}{y}$$

n equations of β and z.

$X \in \mathbb{R}^{m\times n}$, $\beta \in \mathbb{R}^{n\times 1}$, $W \in \mathbb{R}^{d\times n}$, $z \in \mathbb{R}^{d\times 1}$, $y \in \mathbb{R}^{m\times 1}$

I want to find β and z : $n+d$ unknowns in total.

$$\begin{cases} X^T X\beta - \frac{1}{2}W^T z = X^T y & (n \text{ equations}) \\ W\beta = \xi & (d \text{ equations}) \end{cases}$$

$$\begin{bmatrix} X^T X & -\frac{1}{2}W^T \\ W & 0 \end{bmatrix}\begin{bmatrix} \beta \\ z \end{bmatrix} = \begin{bmatrix} X^T y \\ \xi \end{bmatrix}$$

This is a linear system with $n+d$ equations and $n+d$ unknowns.

So, to solve the OPT problem, we can just solve the above linear system.

Linearly

Now we consider an inequality constrained problem:

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & A\boldsymbol{x} \geq \boldsymbol{b} \end{aligned} \tag{2}$$

Theorem

If $\boldsymbol{x}^*$ is a local minimum of (2), then there exists some $\boldsymbol{y} \geq 0$ satisfying

$$\left\{ \begin{aligned} \nabla f(\boldsymbol{x}^*) &= A^T \boldsymbol{y} \\ y_i \cdot (\boldsymbol{a}_i^T \boldsymbol{x}^* - b_i) &= 0, \quad \forall i \end{aligned} \right.$$

where $\boldsymbol{a}_i^T$ is the ith row of A.

Prove in Homework 6.

More General Cases

We have discussed unconstrained optimization and constrained optimization with linear equality constraints or linear inequality constraints and derived the (necessary) optimality conditions.

- We want to extend them to more general cases — KKT conditions .
- Before that, we want to learn a special but important class of nonlinear optimization - convex optimization.

Outline

Collection of Vectors

weighted sum of vectors

Definition

Given a collection of vectors $x_1, ...x_k \in \mathbb{R}^n$:

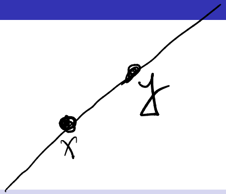
Linear combination: $\sum_{i=1}^k \lambda_i x_i$, $\lambda_i \in \mathbb{R}$, $\forall i$

Affine combination: $\sum_{i=1}^k \lambda_i x_i$, $\sum_{i=1}^k \lambda_i = 1$

Conic combination: $\sum_{i=1}^k \lambda_i x_i$, $\lambda_i \geq 0$, $\forall i$

Convex combination: $\sum_{i=1}^k \lambda_i x_i$, $\sum_{i=1}^k \lambda_i = 1, \lambda_i \geq 0$, $\forall i$

Affine Set

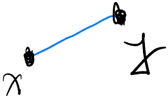


Definition

A set $\mathcal{X}$ is affine if $\forall x, y \in \mathcal{X}$ such that $\lambda x + (1-\lambda)y \in \mathcal{X}, \forall \lambda$.

- Affine set contains the line through any two distinct points in the set.
- Example: solution set of linear equations $\{x | Ax = b\}$. Conversely, every affine set can be expressed as solution set of linear equations

Convex Set

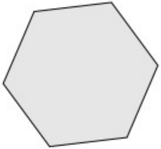


Definition

A set $\mathcal{X}$ is **convex** if $\forall x, y \in \mathcal{X}$ such that $\lambda x + (1 - \lambda) y \in \mathcal{X}, \forall \lambda \in [0, 1]$.

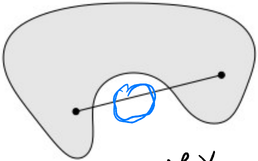
- Convex set contains line segment between any two points in the set
- Example:

1.



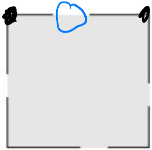
convex

2.



not convex

3.



not convex

convex set   non-convex set

Remark: A set is either convex or not convex.

There's no concave set definition!

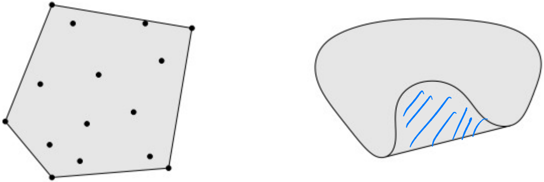
Convex Hull

Definition

Convex hull: set of all convex combinations of elements in the set

$$conv(\mathcal{X}) = \{\sum_i \lambda_i x_i, x_i \in \mathcal{X}, \lambda_i \in [0,1], \sum_i \lambda_i = 1\}$$

conv($\mathcal{X}$) is convex set.



The convex hull of the set X is the smallest convex set containing all elements in X.

Convex Set Examples

1 - **Hyperplane:** set of the form $\{x \mid a^T x = b\}$ where $a \neq 0$
2 - **Halfspace:** set of the form $\{x \mid a^T x \leq b\}$ where $a \neq 0$
3 - **(Euclidean) ball** with center x_c and radius r:

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\}$$

  where $\|\cdot\|_2$ denotes the Euclidean norm
4 - **Polyhedron**: set of the form $\{x \mid a^T x \leq b\}$ where $a \neq 0$

1. hyper plane: $X = \{x : a^T x = b\}$

Proof: $\forall\ x, y \in X,\ \forall \lambda$

$$a^T(\lambda x + (1-\lambda) y) = \lambda a^T x + (1-\lambda)\, a^T y$$

$$= \lambda b + (1-\lambda)\, b$$

$$= b$$

So $\lambda x + (1-\lambda) y \in X,\ \forall \lambda$

Hyperplane is an affine set and a convex set.

2. Half space: $Y = \{x : a^T x \le b\}$

$\forall x, y \in Y,\quad \forall \lambda \in [0, 1]$

$$a^T(\lambda x + (1-\lambda) y) = \lambda a^T x + (1-\lambda)\, a^T y$$

$$\le \lambda b + (1-\lambda)\, b$$

$$= b$$

So $\lambda x + (1-\lambda) y \in Y$ for $\lambda \in [0, 1]$

Half space / polyhedron is a convex set.

$\underset{\swarrow\,\text{center}}{B}(x_c, r)$ $\swarrow$ radius

3. Ball: $B = \{x : \|x - x_c\|_2 \le r\}$

$\|a\|_2 = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$

$= \sqrt{a^T a}$

Triangle Inequality

$\|a+b\|_2$

$\le \|a\|_2 + \|b\|_2$

$\forall x, y \in B$, $\forall \lambda \in [0,1]$

$\|\lambda x + (1-\lambda) y - x_c\|_2$

$= \|\lambda(x - x_c) + (1-\lambda)(y - x_c)\|_2$

$\le \|\lambda(x - x_c)\|_2 + \|(1-\lambda)(y - x_c)\|_2$

$= \lambda \underbrace{\|x - x_c\|_2}_{\le r} + (1-\lambda) \underbrace{\|y - x_c\|_2}_{\le r}$

$\le \lambda r + (1-\lambda) r$

$= r$

So $\lambda x + (1-\lambda) y \in B$

Ball is a convex set.

Operations that Preserve Convexity I

- **Intersection:** Given a family of convex sets $\{X_a\}_{a\in A}$, the set

$$X = \bigcap_{a\in A} X_a$$

is convex.

- **Sum:** Given a family of convex sets $\{X_a\}_{a\in A}$, the set

$$X = \left\{ \sum_{a\in A} x_a : x_a \in X_a \ \forall a \in A \right\}$$

is convex.

- **Product:** If $X_i \subseteq \mathbb{R}^{n_i}$ is a convex set for $i = 1, \ldots, k$, then the set

$$X_1 \times X_2 \times \cdots \times X_k = \left\{ (x_1, \ldots, x_k) \in \mathbb{R}^{n_1+\cdots+n_k} : x_i \in X_i \ \forall i \right\}$$

is convex.

Operations that Preserve Convexity II

- **Affine image:** If $X \subseteq \mathbb{R}^n$ is a convex set, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, then

$$AX + b = \{Ax + b : x \in X\}$$

  is a convex set in $\mathbb{R}^m$.

- **Inverse affine image:** If $Y \subseteq \mathbb{R}^m$ is a convex set, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, then

$$S = \{x \in \mathbb{R}^n : Ax + b \in Y\}$$

  is a convex set.

Outline

Convex Function

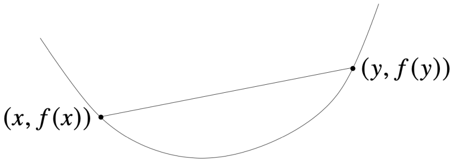
Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is a **convex** function iff the $\mathrm{dom}(f)$ is a convex set and

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

for all $x, y \in \mathrm{dom}(f)$ and $\lambda \in [0,1]$.
f is **concave** if $-f$ is convex.
f is **strictly** convex iff the $\mathrm{dom}(f)$ is a convex set and

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



Convex Function Examples

- 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 $x > 0$, 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 $x > 0$

Concave

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

Examples: Epigraph and Sublevel Set

Definition

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

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

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

$$C_\alpha(f) = \{x \in \mathrm{dom}(f) \mid f(x) \le \alpha\}$$

- The sublevel sets of convex functions are convex (converse is false).
- f is convex iff $epi(f)$ is a convex set.

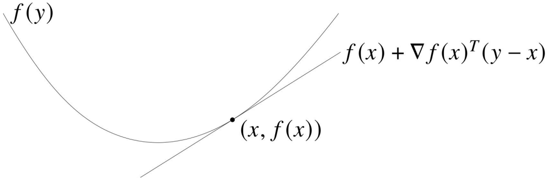
First-order Condition of Convex Functions

First-order condition of convex functions

A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ with convex domain is convex iff

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

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



Second-order Condition of Convex Functions

Second-order condition of convex functions

A twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ with convex domain is convex iff

$$\nabla^2 f(x) \succcurlyeq 0 \quad \text{(the Hessian is positive semi-definite)}$$

for all $x \in \mathrm{dom}(f)$.

Positive Semi-definite

A $n \times n$ matrix A is positive semi-definite if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$. Equivalently, A is positive semi-definite if all its eigenvalues are nonnegative.

Remark: Concave function has a negative semi-definite Hessian.

Examples of Using First and Second-order Conditions

- Quadratic function: $f(x) = (1/2)x^\intercal Px + q^\intercal x + r$
- Least squares objective: $f(x) = ||Ax - b||_2^2$

Convex Function Calculus I

Lemma: Sum Rule

If $a_1, \ldots, a_m \geq 0$, and $f_1, \ldots, f_m$ are convex (concave) functions, then $f = a_1 f_1 + \cdots + a_m f_m$ is a convex (concave) function.

- Nonnegative linear combination preserves convexity (concavity).
- Examples: $x_1^2 + x_2^2$, $e^x + |x|$.

Lemma: Composition with Linear Functions

If $g : \mathbb{R}^m \to \mathbb{R}$ is convex (concave) and $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ are given, then $f(x) := g(Ax + b)$ is convex (concave).

- Examples: e^{2x+3} (convex), $(x_1 - x_2)^2 + (x_2 + x_3)^2$ (convex), $\|Ax - b\|$ (convex), $\log(-2x_1 + 3x_2 + 5)$ (concave).

Convex Function Calculus II

Lemma: Pointwise Maximum

If $f_1, \dots, f_m$ are convex functions, then

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

is a convex function (this can be extended to uncountably many).

- Examples: $|x| = \max\{-x, x\}$, $\max_i\{a_i^\top \mathbf{x} + b_i\}$, $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$.

Lemma: Pointwise Minimum

If $f_1, \dots, f_m$ are concave functions, then

$$f(x) = \min\{f_1(x), \dots, f_m(x)\}$$

is a concave function (this can be extended to uncountably many).

- Examples: $-|x| = \min\{-x, x\}$, $\min_i\{a_i^\top \mathbf{x} + b_i\}$.

Lemma: Composition with Convex and Nondecreasing Function
If $g : \mathbb{R}^m \to \mathbb{R}$ is convex and componentwise nondecreasing, and each $f_i : \mathbb{R}^n \to \mathbb{R}$ is convex, then

$$f(x) = g(f_1(x), \ldots, f_m(x))$$

is a convex function.

- Examples: $f(x) = e^{\sum_{i=1}^m |a_i x - b_i|}$.

Convex Function in ML Examples

- Linear Regression

$$\min_{\beta \in \mathbb{R}^n} \quad \sum_{i=1}^{m} \left(x_i^\top \beta - y_i \right)^2 = \| X\beta - \mathbf{y} \|_2^2$$

- Robust Regression

$$\min_{\beta \in \mathbb{R}^n} \quad \sum_{i=1}^{m} \left| x_i^\top \beta - y_i \right| = \| X\beta - \mathbf{y} \|_1$$

- Logistic Regression

$$\min_{w \in \mathbb{R}^n} \quad \frac{1}{m} \sum_{i=1}^{m} \log \left(1 + \exp(-y_i \cdot \mathbf{w}^\top x_i) \right)$$

Outline

Convex optimization

$$\begin{aligned} \min. \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

If f is a convex function and X is a convex set, then the problem is a convex optimization problem. (Minimize a convex function is equivalent to maximize a concave function.) Otherwise, we call it non-convex optimization problem
Examples:

- o $\min\{x^2 : -1 \le x \le 1\}$ is a convex opt. problem.
- o $\min\{-x^2 : -1 \le x \le 1\}$ is not a convex opt. problem.
- o $\min\{x^2 : -5 \le x \le 5, x \in \mathbb{Z}\}$ is not a convex opt. problem.

Constraint Types

What constraints would make the feasible region convex?

Convex Constraint

Let f be a convex (concave) function. Then for any number c, the set $\Gamma_c = \{x : f(x) \leq (\geq) c\}$ is a convex set.

Therefore,

- If we have constraint $g(x) \leq 0$, and $g(x)$ is convex, then this is a convex constraint.
- If we have constraint $g(x) \geq 0$, and $g(x)$ is concave, then this is a convex constraint.
- Linear constraints are always convex constraints.
- Sometimes, even if a constraint doesn't appear to be in the above form, it still could be a convex constraint.

Being able to identify convex problem is an important skill.

Recognizing Convex Optimization Problem

$$\begin{aligned} \min. \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

Suppose X is in the following format

$$X = \{g_i(x) \le b_i \ \ \forall i \in I, \ \ h_j(x) = d_j \ \ \forall j \in J\},$$

if f is convex, g_i is convex for all $i \in I$ and h_j is affine for all $j \in J$, then the problem is a convex optimization problem.

Checking Convexity

- Check that all variables are continuous
- Check that the objective function is convex
- Check each equality constraint to see if it is linear
- Write each constraint as an inequality constraint in $\leq$ form with a constant on the right-hand-side, and check the convexity of the function on the left-hand-side

If it passes all the checks then you have a convex optimization problem. Otherwise, it may or may not be convex (the conditions are sufficient, not necessary).

Convex Optimization Example

$$\begin{aligned} \min. \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & \frac{x_1}{1 + x_2^2} \leq 0 \\ & (x_1 + x_2)^2 = 0 \end{aligned}$$

the problem is equivalent to

$$\begin{aligned} \min. \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 \leq 0 \\ & x_1 + x_2 = 0 \end{aligned}$$

Example

$$
\begin{aligned}
\max. \quad & xyz \\
\text{s.t.} \quad & z^2 - xy \leq 0 \\
& x, y, z \geq 0
\end{aligned}
$$

Convex Optimization: Local is Global

$$\text{(P):} \quad \min . \{ f(x) : \ x \in X \}$$

Theorem

If (P) is a convex optimization problem then a solution $x^* \in X$ is a local optimal solution of (P) iff it is a global optimal solution.

The ‘Easy’ and ‘Difficult’ Optimization Problems

- Linear v.s. nonlinear?
- Differentiable v.s. nondifferentiable?

Classify whether a problem is hard or easy: **Convex (easy)** v.s. **nonconvex (hard)**.

- **Convex optimization:** Benign global geometry and hence reasonable algorithms can almost always find the global minimum.
- **Nonconvex optimization:** Can be anything (e.g., the *saddle point*). In general, finding a stationary point is NP-hard.