MAT3007 Optimization
Lecture 13 Optimality Conditions

Yuang Chen

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

July 3, 2025

Outline

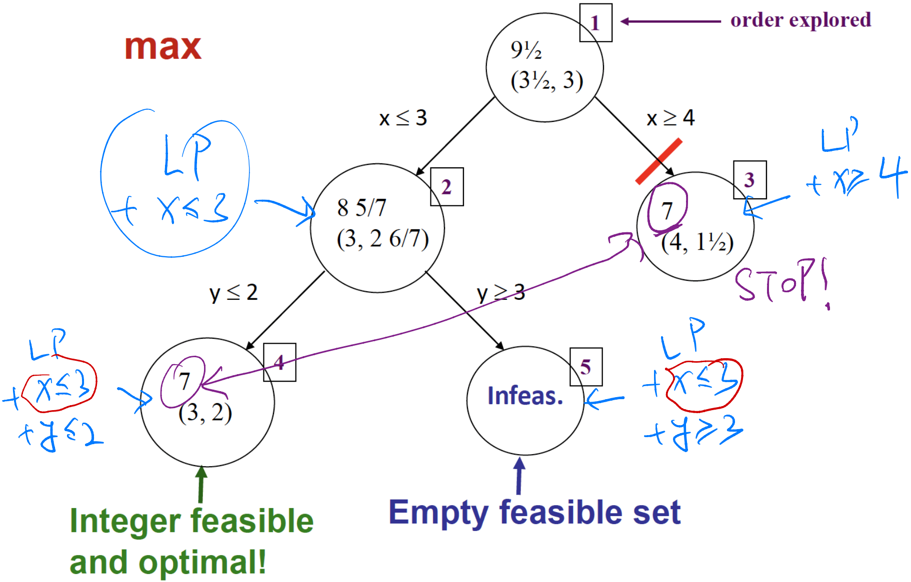
Outline

Example

$$\begin{aligned}
\text{Maximize} \quad & x + 2y \\
\text{subject to} \quad & -2x + 7y \le 14 \\
& 6x + 2y \le 27 \\
& x, y \ge 0 \\
& x, y \in \mathbb{Z}
\end{aligned}$$

LP Relaxation

Example Solution



Outline

Introduction to Nonlinear Optimization

So far we have discussed linear optimization problems. However, in practice, there are many interesting optimization problems that do not take a linear form.

In general, we can write a nonlinear optimization problem as:

$$\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{x} \in \Omega \end{aligned}$$

We call Ω the feasible region and $\boldsymbol{x} \in \Omega$ feasible solutions.

In the following, we study such nonlinear optimization problems.

- Optimality conditions.
- When can we find the optimal solution?
- How to find the optimal solution?
- Without otherwise specified, we always assume we are solving a minimization problem.

Review: Global and Local Minimizers

Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty set and let $f : \Omega \to \mathbb{R}$ be given. We define

$$B_\epsilon(y) := \{x \in \mathbb{R}^n : \|x - y\| < \epsilon\}$$

to be the open ball in $\mathbb{R}^n$ with center y and radius $\epsilon > 0$.

The point $x^* \in \Omega$ is said to be a: For Minimization Problem

- **local minimizer** if there exists $\epsilon > 0$ such that

$$f(x) \geq f(x^*) \quad \text{for all } x \in \Omega \cap B_\epsilon(x^*).$$

- **strict local minimizer** if there is $\epsilon > 0$ with

$$f(x) > f(x^*) \quad \text{for all } x \in (\Omega \cap B_\epsilon(x^*)) \setminus \{x^*\}.$$

- **global minimizer** if

$$f(x) \geq f(x^*) \quad \text{for all } x \in \Omega.$$

- **strict global minimizer** of (1), if $x^* \in \Omega$ and it holds that

$$f(x) > f(x^*) \quad \text{for all } x \in \Omega \setminus \{x^*\}.$$

Remark

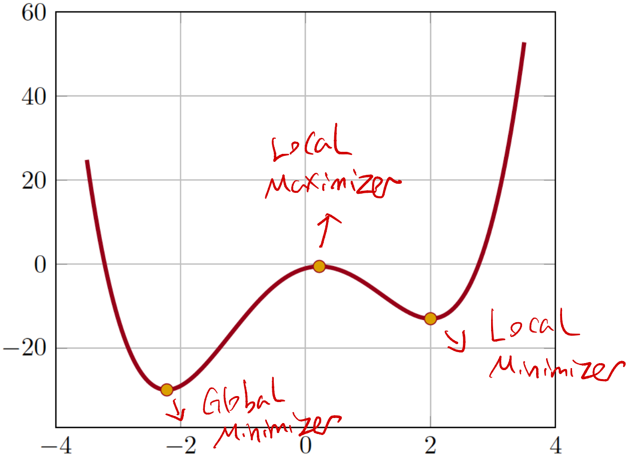
=global optimal solution

- **Remark**: global minimizer $\equiv$ global solution $\equiv$ optimal solution.
- The definitions for **maximizer** is identical, changing: $\geq / > \to \leq / <$.

Example: Minimizer

We consider the unconstrained problem

$$\min_{x\in\mathbb{R}} \quad f(x) := x^4 - 9x^2 + 4x - 1.$$



First-Order Taylor Expansion

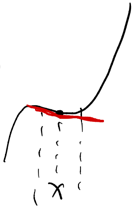
Let $f : \mathbb{R}^n \to \mathbb{R}$. Assume $f(\boldsymbol{x}) = f(x_1, x_2, ..., x_n)$ is continuously differentiable. Then we denote the gradient of f by (an $n \times 1$ vector)

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \quad \in R^n$$

By Taylor expansion, we have

$f(x+t\Delta x)$

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) = f(\boldsymbol{x}) + \alpha \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \overset{0}{\overset{\shortparallel}{o(\alpha)}}$$



Second-Order Taylor Expansion

Let $f : \mathbb{R}^n \to \mathbb{R}$. Assume $f(\boldsymbol{x}) = f(x_1, x_2, ..., x_n)$ is continuously differentiable and second-order differentiable. Then we denote the Hessian matrix of f by (an $n \times n$ matrix)

$$\nabla^2 f(\mathbf{x}) = \begin{array}{cc} & \begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array} & \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix} \end{array}$$

Or

$$[\nabla^2 f(\boldsymbol{x})]_{i,j} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j}$$

By Taylor expansion, we have

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) = f(\boldsymbol{x}) + \alpha \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2} \alpha^2 \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{d} + o(\alpha^2)$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

$\nabla^2 f(x)$ is symmetric

$$\Rightarrow [\nabla^2 f(x)]^T = \nabla^2 f(x)$$

Example

Suppose

$$f(x_1, x_2, x_3) = x_1^2 + x_1 x_2 + x_1 e^{x_3} + x_2 \overset{\ln}{\cancel{\log}}\, x_3$$

Then

$$\nabla f(\boldsymbol{x}) = \left[2x_1 + x_2 + e^{x_3},\ x_1 + \log x_3,\ x_1 e^{x_3} + \frac{x_2}{x_3} \right]^\top$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \!\!\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 2 & 1 & e^{x_3} \\ 1 & 0 & \frac{1}{x_3} \\ e^{x_3} & \frac{1}{x_3} & x_1 e^{x_3} - \frac{x_2}{x_3^2} \end{bmatrix} \end{array}$$

Optimality Conditions

In the following, we first study what conditions an optimal solution has to satisfy for nonlinear optimization problems

- Optimality conditions
- We will start with local optimal solutions

Necessary Optimality Conditions : (a)

If x^* is local optimal, then (a)

Note (a) $\not\Rightarrow$ x^* is local optimal

Sufficient Optimality Conditions : (b)

If x^* satisfies (b), then x^* is local optimal

Note x^* is local optimal $\not\Rightarrow$ (b)

Not necessary

Outline

Optimality Conditions for Unconstrained Problems

We consider the unconstrained problem:

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

What are the optimality conditions for local minimizers for unconstrained problems?

- Claim: We must have

$$\nabla f(\boldsymbol{x}) = 0$$

Reason: If $\nabla f(\boldsymbol{x}) \neq 0$, then we can find a vector $\boldsymbol{d}$ such that $\nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0$. Therefore by Taylor expansion

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) = f(\boldsymbol{x}) + \alpha \nabla f(\boldsymbol{x})^T \boldsymbol{d} + o(\alpha)$$

By choosing α small enough, we can find a point $\boldsymbol{x}' = \boldsymbol{x} + \alpha \boldsymbol{d}$ in the neighborhood of $\boldsymbol{x}$ such that $f(\boldsymbol{x}') < f(\boldsymbol{x})$.

$$\overset{\text{large}}{f(x+\alpha d)} = \overset{\text{small}}{f(x)} + \underline{\alpha \nabla f(x)^T d} + \overbrace{o(\alpha)}^{0}$$

If x^* is local optimal, then

$$f(x) \le f(x+\alpha d) \quad \forall d$$

$$\Rightarrow \quad \underset{\substack{\vee \\ 0}}{\alpha} \nabla f(x)^T d \ge 0 \quad \forall d$$

$$\Rightarrow \quad \nabla f(x)^T d \ge 0 \quad \forall d$$

$$\Rightarrow \quad \nabla f(x) = 0$$

First-Order Necessary Condition (FONC)

Theorem (First-Order Necessary Condition (FONC))

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

First-order necessary condition provides all the candidates for local minimizers.

Example: $f(\boldsymbol{x}) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$.

$$\nabla f = \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - 3 \end{bmatrix}$$

The FONC is

$$2x_1 - x_2 = 0, \quad -x_1 + 2x_2 = 3$$

There is a unique solution $(x_1 = 1, x_2 = 2)$, which turns out to be the global minimizer for f.

Another Example: Least Squares Problem

Assume a variable y is affected by n factors $x_1, ..., x_n$. We know that they approximately have a linear relationship:

$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$

$$y \approx \beta_1 x_1 + \cdots + \beta_n x_n$$

random noise $\varepsilon \sim N(0, \sigma^2)$

Now we want to find out this relationship (parameters βs).

Goal: estimate β's

- We have m observations ($m > n$):

$$\{\underset{\text{fixed}}{\boldsymbol{x}_i}, \underset{\text{random}}{y_i}\} = \{(x_{i1}, ..., x_{in}), y_i\}, i = 1, ..., m \qquad \hat{\beta}: \text{RV}$$

We want to find $\beta = (\beta_1, ..., \beta_n)$ such that the sum of squared errors is minimized.

$E[\hat{\beta}] = \beta$

unbiased

$$\text{minimize}_{\beta} \quad \sum_{i=1}^{m} \Big(\underbrace{y_i - \sum_{j=1}^{n} \beta_j x_{ij}}_{\text{residual}} \Big)^2$$

Least Squares Problem Continued

Our problem is

$$\text{minimize}_{\boldsymbol{\beta}} \quad \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n} \beta_j x_{ij} \right)^2$$

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

The matrix form of this problem is

$$\text{minimize}_{\boldsymbol{\beta}} \quad \|X\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 = \boldsymbol{\beta}^T X^T X \boldsymbol{\beta} - 2\boldsymbol{\beta}^T X^T \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{y}$$

where $\|\boldsymbol{w}\|_2^2 = \boldsymbol{w}^T \boldsymbol{w} = w_1^2 + \cdots + w_n^2$. $\quad \|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$

Facts:

- If $f(\boldsymbol{x}) = \boldsymbol{x}^T M \boldsymbol{x}$ (M is symmetric), then $\nabla f(\boldsymbol{x}) = 2M\boldsymbol{x}$
- If $f(\boldsymbol{x}) = \boldsymbol{c}^T \boldsymbol{x}$, then $\nabla f(\boldsymbol{x}) = \boldsymbol{c}$

Therefore, the FONC for the least squares problem is

$$X^T X \boldsymbol{\beta} = X^T \boldsymbol{y}$$

Solving this equation gives candidates for local minimizer.

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

$$\nabla f(\beta) = 2\underbrace{X^T}_{m\times n}(\underbrace{\underbrace{X}_{m\times n}\underbrace{\beta}_{n\times 1}}_{m\times 1} - \underbrace{y}_{m\times 1})$$

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

$$f'(\beta) = 2X(X\beta - y)$$

$$2X^T(X\beta - y) = 0$$

Google: Matrix Cookbook

$$\Rightarrow 2X^TX\beta = 2X^Ty$$

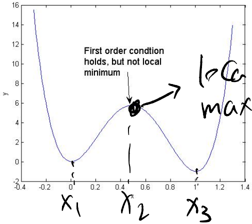
$$\Rightarrow \hat{\beta}^{OLS} = (X^TX)^{-1}X^Ty$$

FONC is Not Sufficient

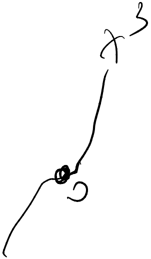
In the example $f(x) = 100x^2(1-x)^2 - x$. The FONC is

$$f'(x) = 400x^3 - 600x^2 + 200x - 1 = 0$$

with solutions $x_1 = 0.01032$, $x_2 = 0.47997$ and $x_3 = 1.00972$.



local maximizer



We see that FONC is not sufficient

- In fact, each local maximum also satisfies the FONC.
- Or it could be neither a local minimum nor maximum (e.g. $f(x) = x^3$).

Outline

Second-Order Necessary Condition

Consider the Taylor expansion again but to the 2nd order (assuming f is second-order differentiable):

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) = f(\boldsymbol{x}) + \alpha \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2}\alpha^2 \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{d} + o(\alpha^2)$$

When the first-order necessary condition holds, we have

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) = f(\boldsymbol{x}) + \frac{1}{2}\alpha^2 \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{d} + o(\alpha^2)$$

In order for $\boldsymbol{x}$ to be a local minimizer, we also need $\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{d}$ to be nonnegative for any $\boldsymbol{d}$.

large small

$$f(x+\alpha d) = f(x) + \cancelto{0}{\alpha \nabla f(x)^T d} + \frac{1}{2}\alpha^2 d^T \nabla^2 f(x) d + \cdots$$

If x is a local minimizer,

FONC $\Rightarrow$ $\nabla f(x) = 0$

$$f(x) \le f(x+\alpha d) \qquad \forall d$$

$$\Rightarrow \underbrace{\frac{1}{2}\alpha^2}_{>0} d^T \nabla^2 f(x) d \ge 0 \qquad \forall d$$

$$\Rightarrow d^T \nabla^2 f(x) d \ge 0 \qquad \forall d$$

Second-Order Necessary Condition (SONC)

Theorem (Second-Order Necessary Condition (SONC))

If $\boldsymbol{x}^*$ is a local minimizer of $f(\cdot)$ for an unconstrained problem, then we must have

1. $\nabla f(\boldsymbol{x}^*) = 0$;
2. *For all $\boldsymbol{d}$, $\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} \geq 0$.* $\Leftrightarrow$ $\nabla^2 f(x) \succeq 0$

$\nabla^2 f(x)$ is PSD

Definition

We call a (symmetric) matrix A positive semi-definite (PSD) if and only if for all $\boldsymbol{x}$, $\boldsymbol{x}^T A \boldsymbol{x} \geq 0$.

Therefore, the second-order necessary condition requires the Hessian matrix at $\boldsymbol{x}^*$ is PSD. In the one-dimensional case, this is equivalent to that the second derivative at $\boldsymbol{x}^*$ is nonnegative (i.e., $f''(x^*) \geq 0$).

Positive Semidefinite Matrices

Here are some useful facts about PSD matrices:

- We usually only talk about PSD for symmetric matrix. If a matrix A is not symmetric, we use $\frac{1}{2}(A + A^T)$ to define the PSD properties (because $\boldsymbol{x}^T A \boldsymbol{x} = \frac{1}{2}\boldsymbol{x}^T(A + A^T)\boldsymbol{x}$)
- A symmetric matrix is PSD if and only if all the eigenvalues are nonnegative.
- A symmetric matrix is PSD if and only if all the principal submatrices have nonnegative determinants
- For any matrix A, $A^T A$ is a (symmetric) PSD matrix

If A is PSD, we call $-A$ a negative semi-definite matrix.

Example Continued

In the example $f(x) = 100x^2(1-x)^2 - x$, the second-order condition is

$$6x^2 - 6x + 1 \geq 0$$

Only $x_1 = 0.01032$ and $x_3 = 1.00972$ satisfy the condition. But $x_2 = 0.47997$ does not (thus x_2 is not a local minimizer)

In the example of least squares problem, we have the following fact:

- If $f(\boldsymbol{x}) = \boldsymbol{x}^T M \boldsymbol{x}$ (M is symmetric), then $\nabla^2 f(\boldsymbol{x}) = 2M$

Therefore, the Hessian matrix in that problem is $2X^TX$, which is always a PSD matrix. Therefore, the SONC always holds.

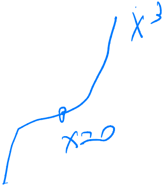
$$f(\beta) = \|X\beta - y\|_2^2$$

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

$$\nabla^2 f(\beta) = 2X^T X \succeq 0$$

SONC is Not Sufficient

However, even both the first- and second-order necessary conditions hold, it still can't guarantee a local minimum.



Consider $f(x) = x^3$ at 0.

- $f'(0) = f''(0) = 0$, thus FONC and SONC hold.
- 0 is not a local minimum

By modifying the SONC, we can get a sufficient condition.

$f(x)=x^3$, $f'(x)=3x^2=0 \Rightarrow x=0$

$f''(x)=6x=0 \Rightarrow x=0$

Outline

Critical and Saddle Points

Critical Pt
local optimal (max or min)
Saddle Point

Definition

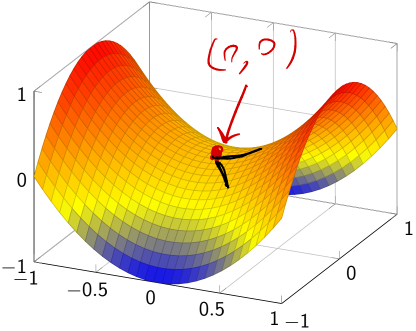
- A point $\mathbf{x}$ satisfying $\nabla f(\mathbf{x}) = 0$ is called **critical point** or **stationary point**.
- A stationary point is called **saddle point** if it is neither a local minimizer nor a local maximizer.

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**.

Illustration of Saddle Points

Consider $f(\boldsymbol{x}) = x_1^2 - x_2^2$



The gradient is $\nabla f(\boldsymbol{x}) = (2x_1, -2x_2)^\top$ and $\boldsymbol{x}^* = (0,0)^\top$ is the single stationary point of f.

- However, $\nabla^2 f(\boldsymbol{x}^*) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$, which is an indefinite matrix. Therefore, $\boldsymbol{x}^*$ is a saddle point.

Second-Order Sufficient Condition (SOSC)

Theorem (Second-Order Sufficient Condition (SOSC))

Let f be twice continuously differentiable. If $\boldsymbol{x}^*$ satisfies:

1. $\nabla f(\boldsymbol{x}^*) = 0$;
2. *For all* $\boldsymbol{d} \neq 0$, $\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} > 0$. $\Rightarrow$ $\nabla^2 f(x) \succ 0$
   $\nabla^2 f(x)$ is PD

Then $\boldsymbol{x}^*$ is a strict local ~~minimum~~ of f for the unconstrained problem.
minimizer

Definition

We call a (symmetric) matrix A positive definite (PD) if and only if for all $\boldsymbol{x} \neq 0$, $\boldsymbol{x}^T A \boldsymbol{x} > 0$.

- PD matrix must be PSD (thus PD is a stronger notion)
- A symmetric matrix is PD if and only if all its eigenvalues are positive
- If A is PD, then we call $-A$ a negative definite matrix.

Proof

The proof is again by Taylor expansion.

Since $\nabla^2 f(\boldsymbol{x}^*)$ is positive definite, for any $\boldsymbol{x}^*$ that satisfies the second-order sufficient condition, for any $\boldsymbol{d}$, and small enough α, we have

$$\Rightarrow \quad f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f(\boldsymbol{x}^*) + \frac{1}{2}\alpha^2 \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} + \overset{\approx 0}{o(\alpha^2)} > f(\boldsymbol{x}^*)$$

which implies that $\boldsymbol{x}^*$ is a local minimizer.

We want $f(x^*) < f(x^* + \alpha d)$

$\Rightarrow \frac{1}{2}\alpha^T d^T \nabla^2 f(x^*) d < 0 \quad \forall d$

$\Rightarrow d^T \nabla^2 f(x^*) d < 0 \quad \forall d$

Example — SOSC is not necessary

Is there a case where local minimizers satisfy SONC but not SOSC?

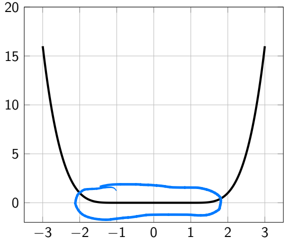
- Yes. Actually, any non-strict local minimizers will be the case.

Consider the following function:

$$f(x) = \begin{cases} (x+1)^4, & x < -1 \\ 0, & -1 \le x \le 1 \\ (x-1)^4, & x \ge 1. \end{cases}$$

$f''(x) = 0$

$-1 \le x \le 1$



Even for some strict local minimizer, SOSC is not satisfied. For example, $f(x) = x^4$ at $x = 0$.

- If x^* does not satisfy

$\cdot\ f'(x) = 0$

$\cdot\ f''(x) > 0$] SOSC

then x^* is not a local minimizer,

False !

$f(x) = x^4$

$x^* = 0$ — $f'(x) = 0$, $f''(x) = 0$

does not satisfy SOSC

global minizer

For Maximization Problems

Our conditions are derived for minimization problems.

- For maximization problems, it is just the opposite direction

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

$f''(x) \leq 0$

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.

$f''(x) < 0$

Optimality Conditions

We have learned for unconstrained problems:

- First-order necessary condition
- Second-order necessary condition
- Second-order sufficient condition

In many cases, we can often use these conditions to identify local optimal solutions.

- Use FONC and SONC to identify candidates, and then use the sufficient condition to verify.
- Or if a problem only has one solution that satisfies FONC, and one can reason that the problem must have a finite optimal solution, then that point must be (global) optimal.

Example

In the example $f(x) = 100x^2(1-x)^2 - x$, x_1 and x_3 all satisfy second-order sufficient conditions ($f''(x) > 0$), thus are local minimum.

In the least squares problem, if $X^T X$ is positive definite, then the solution $\boldsymbol{\beta}$ solved from FONC

$$X^T X \boldsymbol{\beta} = X^T \boldsymbol{y}$$

satisfies the second-order sufficient conditions and must be optimal.

Example

$$f(x, y, z) = 2x^2 + 2y^2 + z^2 - 2xy - 2xz - 6y + 7$$

$$f(x_1, x_2) = x_1^2 x_2 + x_1 x_2^3 - 5x_1 x_2$$

$$f(x,y,z) = 2x^2 + 2y^2 + z^2 - 2xy - 2xz - 6y + 7$$

Step 1: Find Candidates by $\nabla f = 0$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} = \begin{bmatrix} 4x - 2y - 2z \\ 4y - 2x - 6 \\ 2z - 2x \end{bmatrix} = 0$$

$\Rightarrow x = y = z = 3$

$(3, 3, 3)$ is a critical / stationary Pt.

Step 2. Verify $(3,3,3)$ is local minimizer

$$\nabla^2 f = \begin{array}{c c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} 4 & -2 & -2 \\ -2 & 4 & 0 \\ -2 & 0 & 2 \end{bmatrix} \end{array}$$

All eigenvalues of $\nabla^2 f$ are Positive.

$$\nabla^2 f \succ 0 \quad (\text{PD})$$

So $(3,3,3)$ is a local minimizer

$$f(x) = x_1^2 x_2 + x_1 x_2^3 - 5 x_1 x_2$$

Step 1 Find Critical Pts

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 x_2 + x_2^3 - 5x_2 \\ x_1^2 + 3x_1 x_2^2 - 5x_1 \end{bmatrix} = 0$$

$$\begin{cases} x_2(2x_1 + x_2^2 - 5) = 0 \\ x_1(x_1 + 3x_2^2 - 5) = 0 \end{cases}$$

- $x_1 = x_2 = 0 \Rightarrow \underline{x^{1*} = (0,0)}$

- $x_2 = 0,\ x_1 + 3x_2^2 - 5 = 0 \Rightarrow x_1 = 5$

  $\Rightarrow \underline{x^{2*} = (5, 0)}$

- $x_1 = 0,\ 2x_1 + x_2^2 - 5 = 0 \Rightarrow x_2 = \pm\sqrt{5}$

  $\Rightarrow \underline{x^{3*} = (0, \sqrt{5})},\ \underline{x^{4*} = (0, -\sqrt{5})}$

- $2x_1 + x_2^2 - 5 = 0$ and $x_1 + 3x_2^2 - 5 = 0$

  $\Rightarrow x_2^2 = 5 - 2x_1$ plug in

$\Rightarrow x_1 + 3(5 - 2x_1) - 5 = 0$

$\Rightarrow x_1 + 15 - 6x_1 - 5 = 0$

$\Rightarrow -5x_1 + 10 = 0 \Rightarrow x_1 = 2$

$x_2^2 = 5-4 = 1 \Rightarrow x_2 = 1$ or -1

$\Rightarrow \underline{x^{5*} = (2,1)}$, $\underline{x^{6*} = (2,-1)}$

Step 2 Verify critical pts

$$\nabla^2 f(x) = \begin{bmatrix} 2x_2 & 2x_1+3x_2^2-5 \\ 2x_1+3x_2^2-5 & 6x_1x_2 \end{bmatrix}$$

① $x^{1*} = (0,0)$

$$\nabla^2 f(x^{1*}) = \begin{bmatrix} 0 & -5 \\ -5 & 0 \end{bmatrix}$$ indefinite

eigenvalues: $\lambda_1 = 5$, $\lambda_2 = -5$

$x^{1*} = (0,0)$ is a saddle Point

② $x^{2*} = (5,0)$

$$\nabla^2 f(x^{2*}) = \begin{bmatrix} 0 & 5 \\ 5 & 0 \end{bmatrix}$$ indefinite

$\lambda_1 = 5$, $\lambda_2 = -5$

x^{2*} is a saddle Pt

③ $x^{3*} = (0, \sqrt{5})$

$$\nabla^2 f(x^{3*}) = \begin{bmatrix} 2\sqrt{5} & 10 \\ 10 & 0 \end{bmatrix} \to$$ indefinite

$\lambda_1 = 12.5 \quad \lambda_2 = -8$

x^{3*} is a Saddle Point

④ $x^{4*} = (0, -\sqrt{5})$

$\nabla^2 f(x^{4*}) = \begin{bmatrix} -2\sqrt{5} & 10 \\ 10 & 0 \end{bmatrix} \rightarrow$ indefinite

$\lambda_1 = 8 \quad \lambda_2 = -12.5$

x^{4*} is a saddle Pt

⑤ $x^{5*} = (2, 1)$

$\nabla^2 f(x^{5*}) = \begin{bmatrix} 2 & 2 \\ 2 & 12 \end{bmatrix}$ Positive definite

$\det = 2 \times 12 - 4 = 20 = \lambda_1 \lambda_2$

$\text{trace} = 2 + 12 = 14 = \lambda_1 + \lambda_2$

$\lambda_1 > 0, \quad \lambda_2 > 0$

x^{5*} is a local minimizer

⑥ $x^{6*} = (2, -1)$

$\nabla^2 f(x^{6*}) = \begin{bmatrix} -2 & 2 \\ 2 & -12 \end{bmatrix}$ negative definite

$\det = 24 - 4 = 20 = \lambda_1 \lambda_2$

$\text{trace} = -14 = \lambda_1 + \lambda_2$

$\lambda_1 < 0 \qquad \lambda_2 < 0$

γ^{6*} is a local maximizer

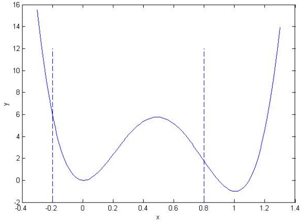
Outline

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

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

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 $0 \leq \alpha \leq \bar{\alpha}$.

For example,

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

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 $\nabla f(\boldsymbol{x}^*) = 0$.

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 $\nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0$.

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 $0 < \gamma \leq \bar{\gamma}$.

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:

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

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

Nonlinear Optimization with Equality Constraints

Consider

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

- 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

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

Nonlinear Optimization with Equality Constraints

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

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

Theorem

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

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

Example

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

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

Another Example

Consider a constrained version of the least squares problem:

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

Inequality Constraints

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

$$\nabla f(\boldsymbol{x}^*) = A^T \boldsymbol{y}$$
$$y_i \cdot (\boldsymbol{a}_i^T \boldsymbol{x}^* - b_i) = 0, \quad \forall i$$

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

Proof in Homework.

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