MAT 3007 – Optimization

Optimality Conditions

Lecture 12

July 6th

Andre Milzarek

SDS / CUHK-SZ

Repetition

Terminologies:

- Global vs local optimizer (minimizer).
- Gradient, Hessian, Taylor expansions, ...

We started to study optimality conditions for unconstrained optimization problems.

Recap: Optimality for Unconstrained Problems

First-Order Necessary Conditions

If x^* is a local minimizer of the unconstr. problem $\min_{x \in \mathbb{R}^n} f(x)$, then we must have $\nabla f(x^*) = 0$.

Theorem: Second-Order Necessary Conditions

If x^* is a local minimizer of f, then it holds that:

1. $\nabla f(x^*) = 0$;
2. For all $d \in \mathbb{R}^n$: $d^\top \nabla^2 f(x^*) d \geq 0$.

- These necessary conditions can be used to find candidates for local minimizers.

Recap: Definiteness

Definiteness: Let A be a real $n \times n$ matrix.

- A is called positive (negative) semidefinite if $x^\top Ax \geq 0$ (≤ 0) for all $x \in \mathbb{R}^n$.
- The matrix A is positive (negative) definite if $x^\top Ax > 0$ (< 0) for all $x \in \mathbb{R}^n \backslash \{0\}$.
- A is said to be indefinite if A is neither positive semidefinite nor negative semidefinite.

Theorem: Eigenvalues and Definiteness

Let A be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_i \in \mathbb{R}$, $i = 1, ..., n$. It follows:

- A is pos. (neg.) semidefinite iff $\lambda_i \geq 0$ (≤ 0) for all i.
- A is pos. (neg.) definite iff $\lambda_i > 0$ (< 0) for all i.
- A is indefinite iff there are $i, j \in \{1, ..., n\}$ with $\lambda_i > 0$ and $\lambda_j < 0$.

Logistics & Agenda

Logistics:

- The detailed description of the midterm project is online.
- The project is generally designed for groups of three students.
- Please send names and student IDs of the students in your group (and the group name) to `huangzhipeng@cuhk.edu.cn` until Wed., July 8th, 11:00 am.
- Deadline for the report submission is Sat., July 18th, 11:00 pm.
- Solutions for the second exercise sheet are now available.
- The fourth (smaller) exercise sheet is also online. It is due on Sun., July 12th, 11:00 am.

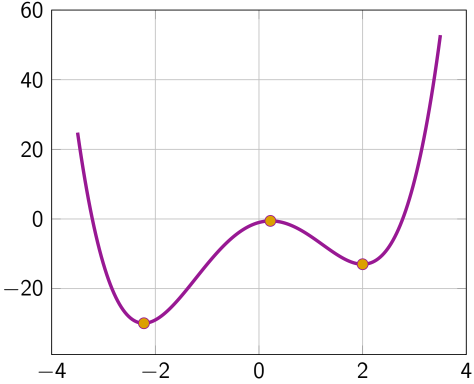
Agenda:

$\rightsquigarrow$ Continue with optimality conditions.

Second-Order Necessary Conditions

We consider the unconstrained problem

$$\text{minimize}_{x \in \mathbb{R}} \; f(x) := x^4 - 9x^2 + 4x - 1.$$



Example: Continued

For $f(x) := x^4 - 9x^2 + 4x - 1$, the second-order condition is:

$$f''(x) = 12x^2 - 18 \geq 0$$

Only $x_1 = -1 - \sqrt{6}/2$ and $x_3 = 2$ satisfy the condition. But for the point $x_2 = -1 + \sqrt{6}/2$, we obtain $f''(x_2) = 12(1 - \sqrt{6}) < 0$ (thus, x_2 is not a local minimizer).

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

- If $f(x) = x^\top Mx$ (M is symmetric), then $\nabla^2 f(x) = 2M$.

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

SONC is Not Sufficient & Saddle Points

However, even if both the first- and second-order necessary cond. hold, we still can not guarantee that the candidate is a local min.!

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

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

- The SONC can used to verify that a candidates are **not** local minimizer.

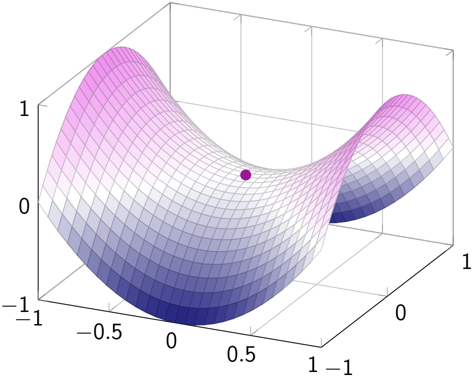
$\rightsquigarrow$ By modifying the SONC, we can get a sufficient condition.

Definition: Stationary Points and Saddle Points

- A point x satisfying $\nabla f(x) = 0$ 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.

Corollary: Saddle Points

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



Plot of the function $f(x) = x_1^2 - x_2^2$.

- The gradient is $\nabla f(x) = (2x_1, -2x_2)^\top$ and $x^* = (0,0)^\top$ is the single stationary point of f. Since $\nabla^2 f(x^*)$ is indefinite, x^* has to be a saddle point.

Second-Order Sufficient Conditions

Second-Order Sufficient Condition (SOSC)

Theorem: Second-Order Sufficient Conditions

Let f be twice continuously differentiable. If x^* satisfies:

1. $\nabla f(x^*) = 0$;
2. For all $d \in \mathbb{R}^n \backslash \{0\}$: $d^\top \nabla^2 f(x^*) d > 0$;

then x^* is a strict local minimum of f.

- A symmetric matrix is PD $\iff$ the determinants of all leading principal submatrices are positive.

The proof uses:

Lemma: Bounds and Eigenvalues

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then

$$\lambda_{\min}(A)\|x\|^2 \leq x^\top A x \leq \lambda_{\max}(A)\|x\|^2 \quad \forall\, x \in \mathbb{R}^n,$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the smallest and largest EV of A.

Our conditions are derived for minimization problems. For maximization problems, we just change the inequalities. Let $f \in C^2$.

Theorem: FONC for Maximization

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

Theorem: SONC for Maximization

If x^* is a local maximizer of f, then we must have 1.) $\nabla f(x^*) = 0$; 2.) $\nabla^2 f(x^*)$ is negative semidefinite.

Theorem: SOSC for Maximization

If x^* satisfies 1.) $\nabla f(x^*) = 0$; 2.) $\nabla^2 f(x^*)$ is negative definite, then x^* is a strict local maximizer.

Optimality Conditions for Unconstrained Problems:

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

In many cases, we can utilize these conditions to identify local and global optimal solutions.

General Strategy:

- Use FONC and SONC to identify all possible candidates. Then, use the sufficient conditions to verify.
- If a problem only has one stationary point and one can reason that the problem must have a finite optimal solution, then this point must be the (global) optimum.

Examples – I

In the example $f(x) = x^4 - 9x^2 + 4x - 1$, the points x_1 and x_3 satisfy the second-order sufficient conditions ($f''(x) > 0$) and are local minimizer.

In the least squares problem, if $X^\top X$ is positive definite (or if it is invertible), then the solution β of the FONC

$$X^\top X\beta = X^\top y$$

is unique and it satisfies the second-order sufficient conditions.

$\rightsquigarrow$ It must be the unique global minimizer of the problem.

Example – II

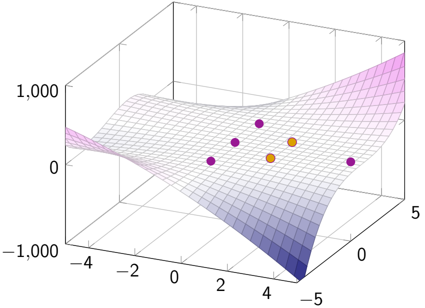
We consider the two-dimensional optimization problem

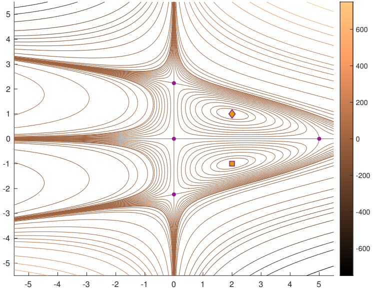
$$\min_{x\in\mathbb{R}^2} \ f(x) = x_1^2 x_2 + x_1 x_2^3 - 5x_1 x_2$$

Find all local minimizer, local maximizer, and saddle points of f!

x^*	$f(x^*)$	$\nabla^2 f(x^*)$	$\lambda_1,\ \lambda_2$	Definiteness	Conclusion
$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$		$\begin{pmatrix} 0 & -5 \\ -5 & 0 \end{pmatrix}$	$5,\ -5$		
$\begin{pmatrix} 5 \\ 0 \end{pmatrix}$		$\begin{pmatrix} 0 & 5 \\ 5 & 0 \end{pmatrix}$	$5,\ -5$		
$\begin{pmatrix} 0 \\ \sqrt{5} \end{pmatrix}$		$\begin{pmatrix} \sqrt{5} & 10 \\ 10 & 0 \end{pmatrix}$	$5\sqrt{5},\ -4\sqrt{5}$		
$\begin{pmatrix} 0 \\ \text{-}\sqrt{5} \end{pmatrix}$		$\begin{pmatrix} -\sqrt{5} & 10 \\ 10 & 0 \end{pmatrix}$	$4\sqrt{5},\ -5\sqrt{5}$		
$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$		$\begin{pmatrix} 1 & 2 \\ 2 & 12 \end{pmatrix}$	$12.4,\ 0.7$		
$\begin{pmatrix} 2 \\ \text{-}1 \end{pmatrix}$		$\begin{pmatrix} -1 & 2 \\ 2 & -12 \end{pmatrix}$	$-0.7,\ -12.4$		

Example – II: Continued





Existence of Solutions

Weierstraß Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and let $\Omega \subset \mathbb{R}^n$ be a bounded, closed, and nonempty set. Then, f attains a global maximum and global minimum on the set Ω.

- The set Ω is closed if for every convergent sequence $(x^k)_k$ with $x^k \in \Omega$ for all k and $\lim_{k\to\infty} x^k = x$, it holds that $x \in \Omega$.
- Ω is bounded if there is $B > 0$ with $\|x\| \leq B$ for all $x \in \Omega$.
- A closed and bounded set is also called compact.

Example:

$$\min \; f(x) = x_1^2 - x_2^2 \quad \text{s.t.} \quad h(x) = x_1^2 + x_2^2 - 4 = 0.$$

- The feasible set $\Omega = \{x : h(x) = 0\} = \{x \in \mathbb{R}^2 : \|x\| = 2\}$ is closed and bounded and f is continuous (on $\mathbb{R}^2$).

$\rightsquigarrow$ By Weierstraß: f attains a global max. and min. on Ω.

The Weierstraß Theorem guarantees existence of global minima if we minimize a continuous function on a compact set.

- For unconstr. problems, this result is not directly applicable.

Definition: Coercivity

A continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be coercive if

$$\lim_{\|x\| \to \infty} f(x) = +\infty.$$

- Geometrically, coercivity means that $f(x)$ increases as x moves away from the origin in any possible direction.
- Mathematically, coercivity means:

$$\forall \ B > 0 \quad \exists \ r > 0 \quad \text{such that} \quad \|x\| > r \implies f(x) > B.$$

Examples:

- The mappings $f(x) = x^2$, $f(x) = x^4$, and $f(x) = |x|^3$ are simple coercive functions.

Examples – Continued:

- The functions $f(x) = x$, $f(x) = x^3$, $f(x) = e^x$, and $f(x) = 1$ are not coercive.

Coercivity is often established by estimating the function and by finding a suitable lower bound for sufficiently large x. (What are the dominating terms in f?).

- Coercivity guarantees existence of solutions:

Theorem: Coercivity and Existence of Solutions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous and coercive function. Then, for all $\alpha > 0$, the level set

$$L_{\leq\alpha} := \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$$

is compact and f has at least one global minimizer.

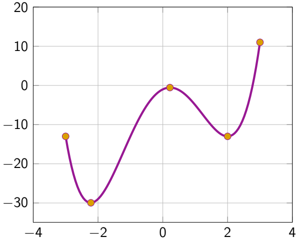
Optimality Conditions for Constrained Problems

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

- What is the difference between constrained and unconstrained problems? How can we characterize optimality?

Let us consider the constrained problem

$$\min_{x \in \mathbb{R}} \; f(x) := x^4 - 9x^2 + 4x - 1 \quad \text{subject to} \quad x \in [-3, 3].$$



In addition to the original local maximizer ($x_2 = \sqrt{6}/2 - 1$), there are two more local maximizer at the boundary ($x = \pm 3$).

At the boundary ($x = 3$), the FONC is not satisfied:

$$f'(3) = 58 > 0.$$

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

$$f(x + d) = f(x) + f'(x) \cdot d + o(d)$$

we can only take d to be negative (otherwise $x + d$ is not feasible).

Thus, $f(x + d) < f(x)$ in a small neighborhood of x in the feasible region. Hence, $x = 3$ is a local (even global) maximizer.

We now formalize the above arguments.

Definition: Feasible Direction

Given $x \in \Omega$, we call d a feasible direction at x if there exists $\bar{t} > 0$ such that $x + td \in \Omega$ for all $0 \le t \le \bar{t}$.

Example:

- If $\Omega = \{x : Ax = b\}$, then all feasible directions at x are given by $\{d : Ad = 0\}$.
- If $\Omega = \{x : Ax \ge b\}$, then the feasible directions at x are given by $\{d : a_i^\top d \ge 0 \text{ if } a_i^\top x = b_i\}$.

Theorem: FONC for Constrained Problems

Let x^* be a local minimum of $\min_{x \in \Omega} f(x)$. Then for any feasible direction d at x^*, we must have $\nabla f(x^*)^\top d \geq 0$.

Proof:

- $\rightsquigarrow$ As before by Taylor expansion and using $x + td \in \Omega$ for all t sufficiently small.

Remark:

- In the unconstrained case, all directions are feasible. Thus, we must have $\nabla f(x^*) = 0$.
- $\rightsquigarrow$ This is also true, when x^* lies in the interior of Ω.

An Alternative View: Descent Directions

Definition: Descent Direction

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

Remark:

- If d is a descent direction at x, then there exists $\bar{\gamma} > 0$ such that $f(x + \gamma d) < f(x)$ for all $0 < \gamma \leq \bar{\gamma}$.

If we denote the set of feasible directions at x by $S_\Omega(x)$ and the set of descent directions at x by $S_D(x)$, then the first order necessary condition can be written as:

$$S_\Omega(x^*) \cap S_D(x^*) = \emptyset$$

$\rightsquigarrow$ There are no feasible descent directions.

General Optimality Conditions

We want to use the notion of feasible and descent directions to obtain optimality conditions for nonlinear programs of the form:

General Nonlinear Optimization Problem:

$$\begin{aligned} \text{minimize}_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad \forall\ i = 1, ..., m, \\ & h_j(x) = 0, \quad \forall\ j = 1, ..., p. \end{aligned}$$

- The feasible set is $\Omega = \{x \in \mathbb{R}^n : g(x) \leq 0,\ h(x) = 0\}$.

Definition: Active and Inactive Set

At a point $x \in \Omega$, the set $\mathcal{A}(x) := \{i : g_i(x) = 0\}$ denotes the set of active constraints. The set of inactive constraints is given by $\mathcal{I}(x) := \{i : g_i(x) < 0\}$.

$\rightsquigarrow$ We first consider linear constraints as special case.

Optimality Conditions: Linear Constraints

We now first consider an inequality constrained problem:

$$\text{minimize}_x \; f(x) \quad \text{s.t.} \quad Ax \geq b. \tag{1}$$

How can we express the necessary optimality conditions?

Theorem: FONC for Linearly Constrained Problems

If x^* is a local minimum of (1), then there exists some $y \geq 0$ with

$$\nabla f(x^*) - A^\top y = 0$$
$$y_i \cdot (a_i^\top x^* - b_i) = 0 \quad \forall\ i,$$

where $a_i^\top$ is the ith row of A.

As a consequence, the first-order conditions for the problem

$$\text{minimize}_x \ f(x) \quad \text{s.t.} \quad Ax = b \tag{2}$$

are given by:

Theorem: Linear Equality Constraints

If x^* is a local minimum of (2), then there is some $y \in \mathbb{R}^m$ with

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

Questions?