Optimality Conditions

Cong Ma

University of Chicago, Winter 2026

Outline

- Optimization problems: basic notation and terminology
- Unconstrained optimization
- First- and second-order necessary conditions for optimality
- Second-order sufficient condition for optimality
- Least squares and its solution

An Optimization Problem

We write an optimization problem in the form

$$\min_{x \in \Omega} \ f(x),$$

where

- $x \in \mathbb{R}^n$ are the *decision variables*,
- $f : \mathbb{R}^n \to \mathbb{R}$ is the *objective function*,
- $\Omega \subseteq \mathbb{R}^n$ is the *constraint/feasible set*.

Local and Global Minimizers

Global minimizer:
$x^\star \in \Omega$ is a *global minimizer* of f if

$$f(x) \geq f(x^\star) \quad \forall x \in \Omega.$$

Local minimizer:

$x^\star \in \Omega$ is a *local minimizer* of f if there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that

$$f(x) \geq f(x^\star) \quad \forall x \in \mathcal{N} \cap \Omega.$$

Strict Local Minimizer

Strict local minimizer:

$x^\star \in \Omega$ is a *strict local minimizer* if there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that

$$f(x) > f(x^\star) \quad \forall x \in \mathcal{N} \cap \Omega,\ x \neq x^\star.$$

Unique Minimizer

Unique minimizer:

$x^\star$ is the *unique minimizer* if it is the only global minimizer of f.

$$f(x) \geq f(x^\star) \quad \forall x \in \Omega, \qquad \text{and equality holds only at } x = x^\star.$$

Remark: Uniqueness is a global property, not a local one.

Example: Minimization without a Minimizer

Example L1.1. Consider minimization of the following two functions, both over their domains.

Example 1: Unbounded Objective

$$f(x) = \tan(x), \qquad \Omega = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

- As $x \to -\frac{\pi}{2}^+$, $f(x) \to -\infty$
- Objective is unbounded below
- Values can be made arbitrarily small

$\Rightarrow$ **No minimizer exists.**

Example 2: Lower Bounded but No Minimizer

$$f(x) = -10 + e^{-x}, \qquad \Omega = \mathbb{R}$$

- $f(x) \geq -10$ for all x
- Objective has a lower bound

Why No Minimizer Exists

$$\inf_{x \in \mathbb{R}} f(x) = -10$$

- For any x, and any $x' > x$:

$$f(x') < f(x)$$

- The infimum is approached as $x \to +\infty$
- But no finite x achieves it

$\Rightarrow$ **No minimizer exists.**

Key Takeaways

- Minimization problems may fail to have solutions
- Two common failure modes:
  - Objective is unbounded below
  - Objective is bounded, but the infimum is not attained
- Existence of a minimizer requires more than smoothness

Existence of Minimizers

Theorem 1 (Weierstrass extreme value theorem)

Let $f : \Omega \to \mathbb{R}$ *be continuous and let* $\Omega \subset \mathbb{R}^n$ *be nonempty and compact (closed and bounded). Then there exists* $x^\star \in \Omega$ *such that*

$$f(x^\star) \leq f(x) \quad \forall x \in \Omega.$$

Interpretation: under mild conditions, optimization problems *actually have solutions*.

Weierstrass Theorem (Compact Sublevel Sets)

Theorem 2

Let f be a continuous function defined on a set S. If f has a nonempty and compact sublevel set $C = \{\boldsymbol{x} \in S : f(\boldsymbol{x}) \leq \alpha\}$ for some $\alpha \in \mathbb{R}$, then there exists $\boldsymbol{x}^\star \in S$ such that $f(\boldsymbol{x}^\star) = \min_{\boldsymbol{x} \in S} f(\boldsymbol{x})$.

The compactness of the sublevel set (being nonempty, closed, and bounded) ensures that the minimum value is attained and a global minimizer exists in S.

Unconstrained optimization

Unconstrained optimization refers to problems of the form

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

i.e., the decision variables are not constrained; only the objective matters.
We aim to provide a simple means of determining whether a particular point is a local or global solution

Taylor's Theorem

Taylor's theorem explains how a smooth function can be approximated *locally* by a polynomial.

The approximation depends on:

- the function value
- low-order derivatives of f

This local approximation is fundamental in optimization.

First-order Taylor theorem (integral form)

Theorem 3

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be continuously differentiable. For any* $x, p \in \mathbb{R}^n$*,*

$$f(x+p) = f(x) + \int_0^1 \nabla f(x + \gamma p)^\top p \, d\gamma.$$

Interpretation: the change in f is an accumulated directional derivative along the line segment from x to $x + p$.

First-order Taylor theorem (mean-value form)

Under the same assumptions, there exists some $\gamma \in (0, 1)$ such that

$$f(x+p) = f(x) + \nabla f(x+\gamma p)^\top p.$$

Interpretation: locally, f behaves like a linear function evaluated at an intermediate point.

Second-order Taylor expansion: gradient

If f is twice continuously differentiable, then

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+\gamma p)\, p \, d\gamma.$$

The Hessian controls how the gradient changes locally.

Second-order Taylor theorem

If f is twice continuously differentiable, then there exists some $\gamma \in (0, 1)$ such that

$$f(x+p) = f(x) + \nabla f(x)^\top p + \frac{1}{2}\, p^\top \nabla^2 f(x+\gamma p)\, p.$$

Interpretation:

- first-order term: linear approximation
- second-order term: curvature correction

Terminology

- $$f(x+p) = f(x) + \int_0^1 \nabla f(x+\gamma p)^\top p \, d\gamma$$

  is called the **integral form** of Taylor's theorem.

- $$f(x+p) = f(x) + \nabla f(x+\gamma p)^\top p$$

  is called the **mean-value form**.

Why Taylor's theorem matters in optimization

Taylor expansions allow us to:

- approximate complicated objectives locally
- reason about descent directions
- design efficient algorithms

First-order necessary condition

Assume f is continuously differentiable.

Theorem 4 (First-order necessary condition)

If $x^\star$ is an unconstrained local minimizer of f, then

$$\nabla f(x^\star) = 0.$$

A point satisfying $\nabla f(x) = 0$ is called a **stationary/critical point**.

This condition is necessary but, in general, not sufficient. (With convexity, it becomes sufficient for global optimality.)

Proof Strategy

We prove by contradiction.

- Assume $x^\star$ is a local minimizer
- Suppose $\nabla f(x^\star) \neq 0$
- Show there exists a nearby point with strictly smaller function value

Proof: $\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$ is Necessary

To prove by contradiction, assume $\boldsymbol{x}^\star$ is a local minimizer but $\nabla f(\boldsymbol{x}^\star) \neq \mathbf{0}$. Using a steepest descent step $\boldsymbol{p} = -\alpha \nabla f(\boldsymbol{x}^\star)$ and the mean value form of Taylor's theorem:

$$f(\boldsymbol{x}^\star - \alpha \nabla f(\boldsymbol{x}^\star)) = f(\boldsymbol{x}^\star) - \alpha \nabla f(\boldsymbol{x}^\star - \gamma\alpha \nabla f(\boldsymbol{x}^\star))^\top \nabla f(\boldsymbol{x}^\star)$$

for some $\gamma \in (0, 1)$. By the continuity of the gradient, for sufficiently small $\alpha > 0$:

$$\nabla f(\boldsymbol{x}^\star - \gamma\alpha \nabla f(\boldsymbol{x}^\star))^\top \nabla f(\boldsymbol{x}^\star) \geq \frac{1}{2} \|\nabla f(\boldsymbol{x}^\star)\|^2$$

Substituting this back into the expansion yields:

$$f(\boldsymbol{x}^\star - \alpha \nabla f(\boldsymbol{x}^\star)) \leq f(\boldsymbol{x}^\star) - \frac{1}{2}\alpha \|\nabla f(\boldsymbol{x}^\star)\|^2 < f(\boldsymbol{x}^\star)$$

Second-order necessary condition

Assume f is twice continuously differentiable.

Theorem 5 (Second-order necessary condition)

If $x^\star$ is an unconstrained local minimizer of f, then

$$\nabla f(x^\star) = 0 \quad \textit{and} \quad \nabla^2 f(x^\star) \succeq 0.$$

Interpretation: curvature at a local minimum cannot be "negative" in any direction.

Second-order conditions: linear algebra interlude

For twice differentiable f, the Hessian at x is $\nabla^2 f(x)$ (symmetric).

PSD/PD definitions

A symmetric matrix H is:

- **positive semidefinite (psd)** if $v^\top H v \geq 0$ for all v,
- **positive definite (pd)** if $v^\top H v > 0$ for all $v \neq 0$.

If A is not symmetric, one often considers its symmetric part $\frac{1}{2}(A + A^\top)$.

Eigenvalue characterizations

Theorem 6

A symmetric matrix H is psd iff all eigenvalues of H are ≥ 0. It is pd iff all eigenvalues are > 0.

Proof

Suppose $\nabla^2 f(\boldsymbol{x}^\star)$ is not positive semidefinite. Then there exists a unit vector $\boldsymbol{v}$ such that $\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}^\star)\boldsymbol{v} \leq -\lambda$ for some $\lambda > 0$. Consider $\boldsymbol{x} = \boldsymbol{x}^\star + \alpha\boldsymbol{v}$. Since $\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$, the second-order Taylor expansion gives:

$$f(\boldsymbol{x}^\star + \alpha\boldsymbol{v}) = f(\boldsymbol{x}^\star) + \frac{1}{2}\alpha^2 \boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}^\star + \gamma\alpha\boldsymbol{v})\boldsymbol{v}$$

for $\gamma \in (0, 1)$. By continuity of the Hessian, for sufficiently small $\alpha > 0$:

$$\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}^\star + \gamma\alpha\boldsymbol{v})\boldsymbol{v} \leq -\frac{\lambda}{2} \implies f(\boldsymbol{x}^\star + \alpha\boldsymbol{v}) \leq f(\boldsymbol{x}^\star) - \frac{1}{4}\alpha^2\lambda < f(\boldsymbol{x}^\star)$$

This contradicts the local minimality of $\boldsymbol{x}^\star$, hence $\nabla^2 f(\boldsymbol{x}^\star) \succeq \mathbf{0}$.

Second-order sufficient condition

Assume f is twice continuously differentiable.

Theorem 7 (Second-order sufficient condition)

If $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succ 0$, then $x^\star$ is a strict local minimum.

Remarks (important):

- $\nabla^2 f(x^\star) \succeq 0$ is not sufficient for local optimality, e.g., $f(x) = x^3$
- $\nabla^2 f(x^\star) \succ 0$ is not necessary for (even strict global) optimality, e.g., $f(x) = x^4$

Proof: Second-Order Sufficient Condition

Since $\nabla^2 f(\boldsymbol{x}^\star) \succ \mathbf{0}$ and the Hessian is continuous, there exist $\rho > 0$ and $\epsilon > 0$ such that for any direction $\boldsymbol{v}$:

$$\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}^\star + \gamma \boldsymbol{p})\boldsymbol{v} \geq \epsilon \|\boldsymbol{v}\|^2$$

for all steps $\|\boldsymbol{p}\| \leq \rho$ and $\gamma \in (0,1)$.

Applying Taylor's theorem at $\boldsymbol{x}^\star$ for a step $\boldsymbol{p}$ with $\|\boldsymbol{p}\| \leq \rho$:

$$f(\boldsymbol{x}^\star + \boldsymbol{p}) = f(\boldsymbol{x}^\star) + \nabla f(\boldsymbol{x}^\star)^\top \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^\top \nabla^2 f(\boldsymbol{x}^\star + \gamma \boldsymbol{p})\boldsymbol{p}$$

Given the stationary point condition $\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$, we substitute the curvature bound:

$$f(\boldsymbol{x}^\star + \boldsymbol{p}) \geq f(\boldsymbol{x}^\star) + \frac{1}{2}\epsilon \|\boldsymbol{p}\|^2$$

For all $\boldsymbol{x} \in \mathcal{N} = \{\boldsymbol{x} : \|\boldsymbol{x} - \boldsymbol{x}^\star\| < \rho\}$ where $\boldsymbol{x} \neq \boldsymbol{x}^\star$, it follows that $f(\boldsymbol{x}) > f(\boldsymbol{x}^\star)$. This confirms $\boldsymbol{x}^\star$ is a strict local minimizer.

Least Squares: Gradient and Hessian via Taylor Expansion

Consider the loss function $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$. To find the derivatives, we expand $f(\boldsymbol{x}+\boldsymbol{p})$ and identify the linear and quadratic terms in $\boldsymbol{p}$:

$$\begin{aligned} f(\boldsymbol{x}+\boldsymbol{p}) &= \tfrac{1}{2}(\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{p}) - \boldsymbol{b})^\top(\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{p}) - \boldsymbol{b}) \\ &= \tfrac{1}{2}((\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{A}\boldsymbol{p})^\top((\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{A}\boldsymbol{p}) \\ &= \tfrac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 + (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^\top\boldsymbol{A}\boldsymbol{p} + \tfrac{1}{2}\boldsymbol{p}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{p} \end{aligned}$$

Cont'd

By comparing this to the Taylor series
$f(\boldsymbol{x}+\boldsymbol{p}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^\top \nabla^2 f(\boldsymbol{x})\boldsymbol{p}$:

1. Gradient Identification:
The linear term is $(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^\top \boldsymbol{A}\boldsymbol{p} = (\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}))^\top \boldsymbol{p}$. Thus:

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})$$

2. Hessian Identification:
The quadratic term is $\frac{1}{2}\boldsymbol{p}^\top(\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{p}$. Thus:

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}^\top\boldsymbol{A}$$

Least squares: solution via optimality conditions

First-order optimality gives the **normal equations**:

$$A^\top A x^\star = A^\top b.$$

If A has full column rank, then $A^\top A$ is invertible and

$$x^\star = (A^\top A)^{-1} A^\top b.$$