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 L1.2.

Let f be a continuous function defined on a set S.

Assumption: Compact Sublevel Set

Assume that f has a nonempty and compact sublevel set.
That is, there exists $\alpha \in \mathbb{R}$ such that

$$\{x \in S : f(x) \leq \alpha\}$$

is:

- nonempty
- bounded
- closed

Conclusion

Under these assumptions,

$$\exists\, x^\star \in S \quad \text{such that} \quad f(x^\star) = \min_{x \in S} f(x).$$

- The minimum value is attained
- A 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 2

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

Examples:

- Gradient descent: first-order approximation
- Newton's method: second-order approximation

First-order necessary condition

Assume f is continuously differentiable.

Theorem 3 (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

Descent Direction

Assume $\nabla f(x^\star) \neq 0$.
Consider the step

$$p = -\alpha \nabla f(x^\star),$$

where $\alpha > 0$ is small.

- This is a step in the direction of steepest descent

First-Order Expansion

By the mean value form of Taylor's theorem,

$$f(x^\star - \alpha \nabla f(x^\star)) = f(x^\star) - \alpha \, \nabla f(x^\star - \gamma\alpha \nabla f(x^\star))^T \nabla f(x^\star),$$

for some $\gamma \in (0, 1)$.

Using Continuity of the Gradient

Since ∇f is continuous,

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

for all sufficiently small $\alpha > 0$ and any $\gamma \in (0, 1)$.

Strict Decrease in Function Value

Substituting into the expansion:

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

for all sufficiently small $\alpha > 0$.

Contradiction with Local Minimality

- For any neighborhood $\mathcal{N}$ of $x^\star$
- Points of the form $x^\star - \alpha \nabla f(x^\star)$ lie in $\mathcal{N}$
- And satisfy $f(x) < f(x^\star)$

$\Rightarrow$ Contradiction.

Conclusion

Therefore,

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

- Every local minimizer must be a stationary point
- This is a necessary, not sufficient, condition

Second-order necessary condition

Assume f is twice continuously differentiable.

Theorem 4 (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 5

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

Proof Strategy (Contradiction)

Suppose, for contradiction, that the Hessian is not positive semidefinite.

- Then $\nabla^2 f(x^\star)$ has a negative eigenvalue
- $\exists v \in \mathbb{R}^n$, $\|v\| = 1$, and $\lambda > 0$ such that

$$v^T \nabla^2 f(x^\star) v \leq -\lambda$$

We will move along direction v.

Second-Order Expansion

Consider $x = x^\star + \alpha v$, with $\alpha > 0$ small.
By the second-order Taylor expansion,

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

for some $\gamma \in (0, 1)$.
Since $\nabla f(x^\star) = 0$, the linear term vanishes.

Strict Decrease and Contradiction

By continuity of $\nabla^2 f$, for α small enough,

$$v^T \nabla^2 f(x^\star + \gamma \alpha v) v \leq -\frac{\lambda}{2}.$$

Substituting,

$$f(x^\star + \alpha v) \leq f(x^\star) - \frac{1}{4}\alpha^2 \lambda < f(x^\star).$$

- Arbitrarily close points have smaller function value
- Contradiction to local minimality

$$\Rightarrow \nabla^2 f(x^\star) \succeq 0.$$

Second-order sufficient condition

Assume f is twice continuously differentiable.

Theorem 6 (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., $f(x) = x^4$

Uniform Positive Curvature Nearby

Since $\nabla^2 f(x^\star) \succ 0$ and $\nabla^2 f$ is continuous:

- $\exists\, \rho > 0$ and $\varepsilon > 0$ such that

$$v^T \nabla^2 f(x^\star + \gamma p)\, v \;\geq\; \varepsilon \|v\|^2$$

  for all $\|p\| \leq \rho$, all $\gamma \in (0,1)$, and all $v \in \mathbb{R}^n$.

(Uses continuity of eigenvalues.)

Taylor Expansion

Apply Taylor's theorem at $x^\star$:

$$f(x^\star + p) = f(x^\star) + \nabla f(x^\star)^T p + \frac{1}{2}\, p^T \nabla^2 f(x^\star + \gamma p)\, p, \quad \gamma \in (0,1).$$

Since $\nabla f(x^\star) = 0$,

$$f(x^\star + p) \;\geq\; f(x^\star) + \frac{1}{2}\, \varepsilon \|p\|^2, \qquad \forall\, \|p\| \leq \rho.$$

Conclusion: Strict Local Minimality

Let

$$\mathcal{N} = \{x^\star + p : \ \|p\| < \rho\}.$$

Then for all $x \in \mathcal{N}$, $x \neq x^\star$,

$$f(x) > f(x^\star).$$

$$\Rightarrow \quad x^\star \text{ is a strict local minimizer.}$$

Least squares: solution via optimality conditions

Consider

$$\min_{x\in\mathbb{R}^n} \ \|Ax-b\|_2^2.$$

Let $f(x) = \|Ax-b\|_2^2 = (Ax-b)^\top(Ax-b)$. Then

$$\nabla f(x) = 2A^\top(Ax-b).$$

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

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

Necessary and sufficient conditions for optimality

Convexity is the key.