Optimization Basic concepts

Instructor: Jin Zhang

Department of Mathematics
Southern University of Science and Technology
Spring 2024

Goals of this lecture

The general form of optimization:

$$\begin{aligned} \min \quad & f(x), \\ \text{subject to} \quad & x \in \Omega. \end{aligned}$$

We study the following topics:

- terminology
- types of minimizers
- optimality conditions

Unconstrained vs constrained optimization

$$\begin{aligned} \min \quad & f(x), \\ \text{subject to} \quad & x \in \Omega. \end{aligned}$$

Suppose $x \in \mathbb{R}^n$, Ω is called the feasible set.

- if $\Omega = \mathbb{R}^n$, then the problem is called unconstrained.
- otherwise, the problem is called constrained.

In general, more sophisticated techniques are needed to solve constrained problems.

(off the topic)

Later, we will study some nonsmooth analysis and algorithms that allow f to have the extended value, ∞. Then, we can write any constrained problem in the unconstrained form

$$\min f(x) + \iota_\Omega(x),$$

where the *indicator function*

$$\iota_\Omega(x) = \begin{cases} 0, & x \in \Omega, \\ \infty, & x \notin \Omega. \end{cases}$$

The objective function $f(x) + \iota_\Omega(x)$ is nonsmooth.

Types of solutions

- x^* is a local minimizer if there is $\epsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$ and $\|x - x^*\| < \epsilon$.
- x^* is a global minimizer if $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$
- If "$\geq$" is replaced with "$>$", then they are strict local minimizer and strict global minimizer, respectively.

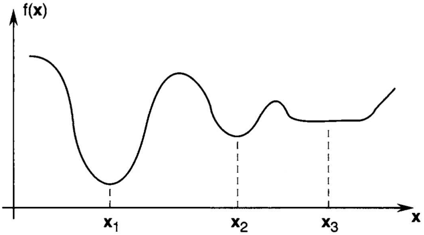


Figure: x_1: strict global minimizer; x_2: strict local minimizer; x_3: local minimizer

Convexity and global minimizers

- A set Ω is convex if $\lambda x+(1-\lambda)y \in \Omega$ for any $x,\ y \in \Omega$ and $\lambda \in [0,1]$.
- A function is convex if

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

  for any $x,\ y \in \Omega$ and $\lambda \in [0,1]$.

  A function is convex if and only if its epigraph is convex.
- An optimization problem is convex if both the objective function and feasible set are convex.
- **Theorem:** Any local minimizer of a convex optimization problem is a global minimizer.

Derivatives

- First-order derivative: row vector

$$Df \triangleq \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right].$$

- **Gradient** of $\nabla f = (Df)^T$, which is a column vector.

- A gradient represents the slope of the tangent of the graph of function. It gives the linear approximation of f at a point. It points toward the greatest rate of increase.

- **Hessian** (i.e., second-derivative) of f:

$$F(x) \triangleq D^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

  which is a symmetric matrix.
- For one-dimensional function $f(x)$ where $x \in \mathbb{R}$, it reduces to $f''(x)$.
- $F(x)$ is the Jacobian of $\nabla f(x)$, that is, $F(x) = J(\nabla f(x))$.
- Alternative notation: $H(x)$ and $\nabla^2 f(x)$ are also used for Hessian.
- A Hessian gives a quadratic approximation of f at a point.
- Gradient and Hessian are local properties that help us recognize local solutions and determine a direction to move at toward the next point.

Example

Consider

$$f(x_1, x_2) = x_1^3 + x_1^2 - x_1 x_2 + x_2^2 + 5x_1 + 8x_2 + 4$$

Then,

$$\nabla f(x) = \begin{bmatrix} 3x_1^2 + 2x_1 - x_2 + 5 \\ -x_1 + 2x_2 + 8 \end{bmatrix} \in \mathbb{R}^2$$

and

$$F(x) = \begin{bmatrix} 6x_1 + 2 & -1 \\ -1 & 2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

Observation: if f is a quadratic function (remove x_1^3 in the above example), $\nabla f(x)$ is a linear vector and $F(x)$ is a symmetric constant matrix for any x.

Taylor expansion

Suppose $\phi \in \mathcal{C}^m$ (m times continuously differentiable). The Taylor expansion of ϕ at a point a is

$$\phi(a+h) = \phi(a) + \phi'(a)h + \frac{\phi''(a)}{2!}h^2 + \cdots + \frac{\phi^m(a)}{m!}h^m + o(h^m).$$

There are other ways to write the last two terms.

Example: Consider $x, d \in \mathbb{R}^n$ and $f \in \mathcal{C}^2$. Define $\phi(\alpha) = f(x+\alpha d)$. Then,

$$\begin{aligned} \phi'(\alpha) &= \nabla f(x+\alpha d)^T d \\ \phi''(\alpha) &= dF(x+\alpha d)^T d \end{aligned}$$

Hence,

$$\begin{aligned} f(x+\alpha d) &= f(x) + (\nabla f(x)^T d)\alpha + o(\alpha) \\ &= f(x) + (\nabla f(x)^T d)\alpha + \frac{dF(x)^T d}{2}\alpha^2 + o(\alpha^2). \end{aligned}$$

Feasible direction

- A vector $d \in \mathbb{R}^n$ is a feasible direction at $x \in \Omega$ if $d \neq 0$ and $x + \alpha d \in \Omega$ for some small $\alpha > 0$. (It is possible that d is an infeasible step, that is, $x + d \notin \Omega$. But if there is some room in Ω to move from x toward d, then d is a feasible direction.)

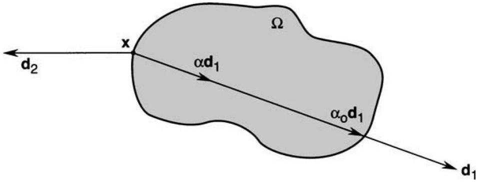


Figure: d_1 is feasible, d_2 is infeasible

- If $\Omega = \mathbb{R}^n$ or x lies in the interior of Ω, then any $d \in \mathbb{R}^n \setminus \{0\}$ is a feasible direction
- Feasible directions are introduced to establish optimality conditions, especially for points on the boundary of a constrained problem

First-order necessary condition

Let $\mathcal{C}^1$ be the set of continuously differentiable functions.

Theorem

First-Order Necessary Condition (FONC). *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-value function on Ω. If x^* is a local minimizer of f over Ω, then for any feasible direction d at x^*, we have*

$$d^T \nabla f(x^*) \geq 0.$$

Proof: Let d by any feasible direction. First-order Taylor expansion:

$$f(x^* + \alpha d) = f(x^*) + \alpha d^T \nabla f(x^*) + o(\alpha).$$

If $d^T \nabla f(x^*) < 0$, which does not depend on α, then $f(x^* + \alpha d) < f(x^*)$ for all sufficiently small $\alpha > 0$ (that is, all $\alpha \in (0, \bar{\alpha})$ for some $\bar{\alpha} > 0$). This is a contradiction since x^* is a local minimizer. ■

Corollary

Interior Case. *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-value function on Ω. If x^* is a local minimizer of f over Ω and if x^* is an interior point, then*

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

Proof: Since any $d \in \mathbb{R}^n \setminus \{0\}$ is a feasible direction, we can set $d = -\nabla f(x^*)$. We have $d^T \nabla f(x^*) = -\|\nabla f(x^*)\| 2 \geq 0$. Since $\|\nabla f(x^*)\|^2 \geq 0$, we have $\|\nabla f(x^*)\|^2 = 0$ and thus $\nabla f(x^*) = 0$. ■

Comment: This condition also reduces the problem

$$\min \ f(x)$$

to solving the equation

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

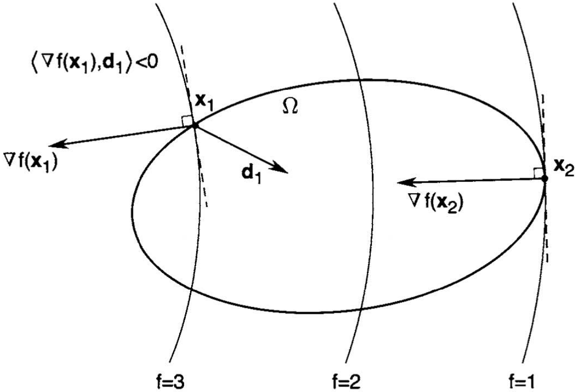


Figure: x_1 fails to satisfy the FONC; x_2 satisfies the FONC

Second-order necessary condition

In FONC, there are two possibilities

- $d^T \nabla f(x^*) > 0$;
- $d^T \nabla f(x^*) = 0$.

In the first case, $f(x^* + \alpha d) > f(x^*)$ for all sufficiently small $\alpha > 0$.
In the second case, the vanishing $d^T \nabla f(x^*)$ allows us to check higher-order derivatives.

Let $\mathcal{C}^2$ be the set of twice continuously differentiable functions.

Theorem

Second-Order Necessary Condition (SONC). *Let $\Omega \subset \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, x^* a local minimizer of f over Ω, and d a feasible direction at x^*. If $d^T \nabla f(x^*) = 0$, then*

$$d^T F(x^*) d \geq 0,$$

where F is the Hessian of f.

Proof: Assume that $\exists$ a feasible direction d with $d^T \nabla f(x^*) = 0$ and $d^T F(x^*) d < 0$. By 2nd-order Taylor expansion (with a vanishing 1st order term), we have

$$f(x^* + \alpha d) = f(x^*) + \frac{d^T F(x^*) d}{2} \alpha^2 + o(\alpha^2),$$

where by our assumption $d^T F(x^*) d < 0$. Hence, for all sufficiently small $\alpha > 0$, we have $f(x^* + \alpha d) < f(x^*)$, which contradicts that x^* is a local minimizer. ■

Corollary

Interior Case *Let x^* be a interior point of $\Omega \subset \mathbb{R}^n$. If x^* is a local minimizer of $f : \Omega \to \mathbb{R}^n$, $f \in \mathcal{C}^2$, then*

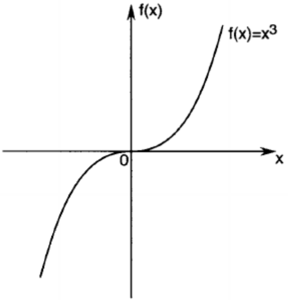
$$\nabla f(x^*) d = 0,$$

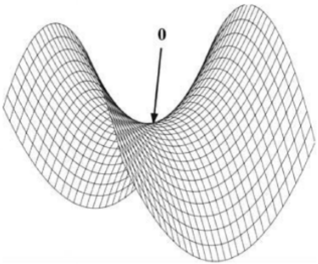
and $F(x^*)$ is positive semidefinite ($F(x^*) \geq 0$); that is, for all $d \in \mathbb{R}^n$,

$$d^T F(x^*) d \geq 0.$$

The necessary conditions are not sufficient

Counter examples





$f(x) = x^3$, $f'(x) = 3x^2$, $f''(x) = 6x$

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

$\mathbf{0}$ is a saddle point: $\nabla f(\mathbf{0}) = \mathbf{0}$ but neither a local minimizer nor maximizer

By SONC, $\mathbf{0}$ is not a local minimizer!

Second-order sufficient condition

Let $\mathcal{C}^2$ be the set of twice continuously differentiable functions.

Theorem

Second-Order Sufficient Condition (SOSC), Interior point. *Let $f \in \mathcal{C}^2$ be defined on a region in which x^* is an interior point. Suppose that*

1. $\nabla f(x^*) = 0$;
2. $F(x^*) > 0$.

Then, x^* is a strict local minimizer of f.

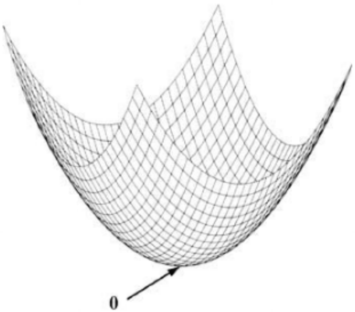
Comments:

- part 2 states $F(x^*)$ is positive definite: $x^T F(x^*) x > 0$ for $x \neq 0$.
- the condition is not necessary for strict local minimizer.

Proof: For any $d \neq 0$ and $\|d\| = 1$, we have $d^T F(x^*) d \geq \lambda_{\min}(F(x^*)) > 0$. Use the 2nd order Taylor expansion

$$f(x^* + \alpha d) = f(x^*) + \frac{\alpha^2}{2} d^T F(x^*) d + o(\alpha^2) \geq f(x^*) + \frac{\alpha^2}{2} \lambda_{\min}(F(x^*)) + o(\alpha^2).$$

Then, $\exists \bar{\alpha} > 0$, regardless of d, such that $f(x^* + \alpha d) > f(x^*)$, $\alpha \in (0, \bar{\alpha})$. ■



Graph of $f(x) = x_1^2 + x_2^2$
The point 0 satisfies the SOSC.

Roles of optimality conditions

- **Recognize a solution:** given a candidate solution, check optimality conditions to verify it is a solution.

- **Measure the quality** of an approximate solution: measure how ¡°close¡± a point is to being a solution

- **Develop algorithms:** reduce an optimization problem to solving a (nonlinear) equation (finding a root of the gradient).

Later, we will see other forms of optimality conditions and how they lead to equivalent subproblems, as well as algorithms