Math 273a: Optimization
Basic concepts

Instructor: Wotao Yin
Department of Mathematics, UCLA

Spring 2015

slides based on Chong-Zak, 4th Ed.

Goals of this lecture

The general form of optimization:

$$\text{minimize } f(\boldsymbol{x})$$
$$\text{subject to } \boldsymbol{x} \in \Omega$$

We study the following topics:

- terminology
- types of minimizers
- optimality conditions

Unconstrained vs constrained optimization

$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}) \\ &\text{subject to } \boldsymbol{x} \in \Omega \end{aligned}$$

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

$$\text{minimize} \;\; f(\boldsymbol{x}) + \iota_\Omega(\boldsymbol{x})$$

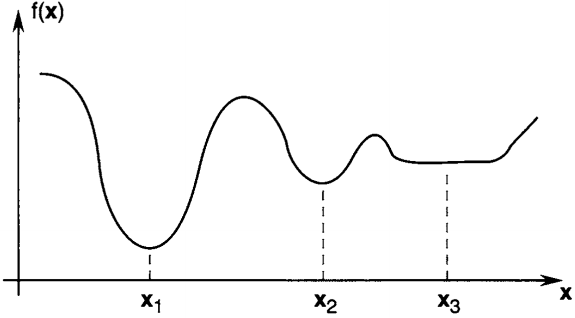
where the *indicator function*

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

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

Types of solutions

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



$\boldsymbol{x}_1$: strict global minimizer; $\boldsymbol{x}_2$: strict local minimizer; $\boldsymbol{x}_3$: local minimizer

Convexity and global minimizers

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

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

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

  A function is convex if and only if its graph 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

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

- **Gradient** of f: $\nabla f = (Df)^T$, 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:

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

which is a symmetric matrix. $(\boldsymbol{F}(\boldsymbol{x}))_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$.

- For one-dimensional function $f(x)$ where $x \in \mathbb{R}$, it reduces to $f''(x)$.

- $\boldsymbol{F}(\boldsymbol{x})$ is the Jacobian of $\nabla f(\boldsymbol{x})$, that is, $\boldsymbol{F}(\boldsymbol{x}) = J(\nabla f(\boldsymbol{x}))$.

- Alternative notation: $\boldsymbol{H}(\boldsymbol{x})$ and $\nabla^2 f(\boldsymbol{x})$ are also used for Hessian.

- A Hessian gives a quadratic approximation of f at a point.

- Gradient and Hessian are <u>local properties</u> 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,

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

and

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

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

Taylor expansion

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

$$\phi(a+h) = \phi(a) + \phi'(a)h + \frac{\phi''(a)}{2!}h^2 + \cdots + \frac{\phi^m(a)}{m!}h^m + o(h^m).$$ [1]

There are other ways to write the last two terms.

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

$$\phi'(\alpha) = \nabla f(\boldsymbol{x} + \alpha \boldsymbol{d})^T \boldsymbol{d}$$
$$\phi''(\alpha) = \boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x} + \alpha \boldsymbol{d})^T \boldsymbol{d}$$

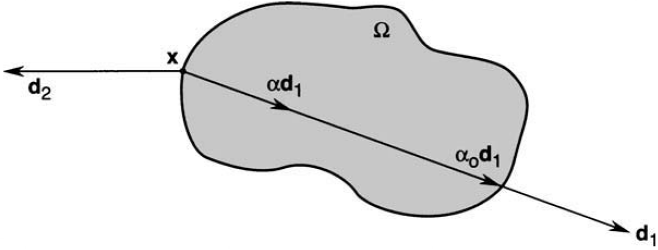
Hence,

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

[1] $o(\alpha)$ collects the term(s) that is "asymptotically smaller than α" near 0, that is, $\frac{o(\alpha)}{\alpha} \to 0$, as $\alpha \downarrow 0$.

Feasible direction

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



$\boldsymbol{d}_1$ is feasible, $\boldsymbol{d}_2$ is infeasible

- If $\Omega = \mathbb{R}^n$ or $\boldsymbol{x}$ lies in the interior of Ω, then any $\boldsymbol{d} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ 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 6.1 ***First-Order Necessary Condition (FONC).*** *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\mathbf{d}$ at $\boldsymbol{x}^*$, we have*

$$\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0.$$

Proof: Let $\boldsymbol{d}$ by any feasible direction. First-order Taylor expansion:

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

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

Corollary 6.1 ***Interior Case.*** *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω and if $\boldsymbol{x}^*$ is an interior point of Ω, then*

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

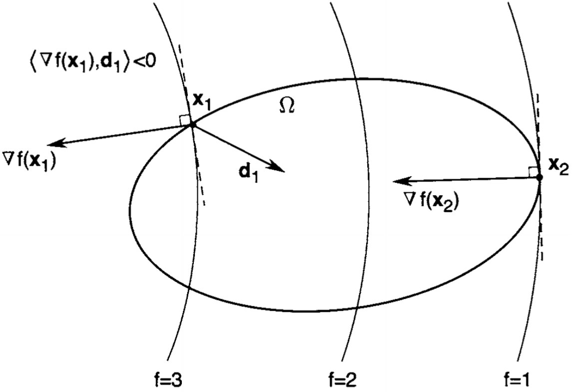
Proof: Since any $\boldsymbol{d} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ is a feasible direction, we can set $\boldsymbol{d} = -\nabla \boldsymbol{f}(\boldsymbol{x}^*)$. From Theorem 6.1, we have $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) = -\|\nabla f(\boldsymbol{x}^*)\|^2 \geq 0$. Since $\|\nabla f(\boldsymbol{x}^*)\|^2 \geq 0$, we have $\|\nabla f(\boldsymbol{x}^*)\|^2 = 0$ and thus $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$. □

Comment: This condition also reduces the problem

$$\text{minimize} \;\; f(\boldsymbol{x})$$

to solving the equation

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



x_1 fails to satisfy the FONC; x_2 satisfies the FONC

Second-order necessary condition

In FONC, there are two possibilities

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

In the first case, $f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) > f(\boldsymbol{x}^*)$ for all sufficiently small $\alpha > 0$.

In the second case, the *vanishing* $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*)$ allows us to check higher-order derivatives.

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

Theorem 6.2 ***Second-Order Necessary Condition (SONC).*** *Let $\Omega \subset \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, $\boldsymbol{x}^*$ a local minimizer of f over Ω, and $\boldsymbol{d}$ a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then*

$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0,$$

where $\boldsymbol{F}$ is the Hessian of f. □

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

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

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

Corollary 6.2 ***Interior Case.*** *Let $\boldsymbol{x}^*$ be an interior point of $\Omega \subset \mathbb{R}^n$. If $\boldsymbol{x}^*$ is a local minimizer of $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, then*

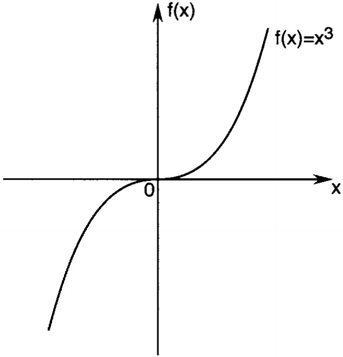
$$\nabla f(\boldsymbol{x}^*) = \mathbf{0},$$

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

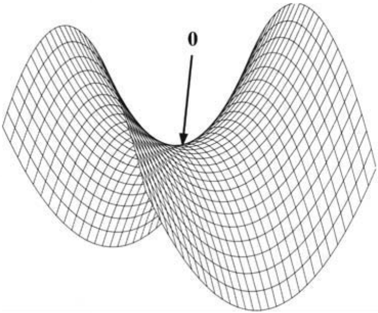
$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0.$$

The necessary conditions are not sufficient

Counter examples



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



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

$\mathbf{0}$ 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

Theorem 6.3 ***Second-Order Sufficient Condition (SOSC), Interior Case.*** *Let $f \in \mathcal{C}^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that*

1. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.
2. $\boldsymbol{F}(\boldsymbol{x}^*) > 0$.

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f. □

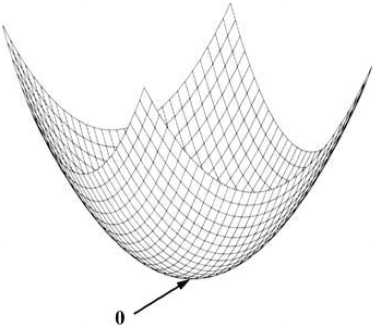
Comments:

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

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

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

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



Graph of $f(\boldsymbol{x}) = x_1^2 + x_2^2$
The point $\mathbf{0}$ 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

Quiz questions

1. Show that for $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$, $\boldsymbol{d} \neq \boldsymbol{0}$ is a feasible direction at $\boldsymbol{x} \in \Omega$ if and only if $\boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$.

2. Show that for any unconstrained quadratic program, which has the form

$$\text{minimize}\, f(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x},$$

   if $\boldsymbol{x}^*$ satisfies the second-order necessary condition, then $\boldsymbol{x}^*$ is a global minimizer.

3. Show that for any unconstrained quadratic program with $\boldsymbol{Q} \geq 0$ ($\boldsymbol{Q}$ is symmetric and positive semi-definite), $\boldsymbol{x}^*$ is a global minimizer if and only if $\boldsymbol{x}^*$ satisfies the first-order necessary condition. That is, the problem is equivalent to solving $\mathbf{Q}\boldsymbol{x} = \mathbf{b}$.

4. Consider $\text{minimize}\, \boldsymbol{c}^T\boldsymbol{x}, \ \text{subject to}\ \boldsymbol{x} \in \Omega$. Suppose that $\boldsymbol{c} \neq \boldsymbol{0}$ and the problem has a global minimizer. Can the minimizer lie in the interior of Ω?