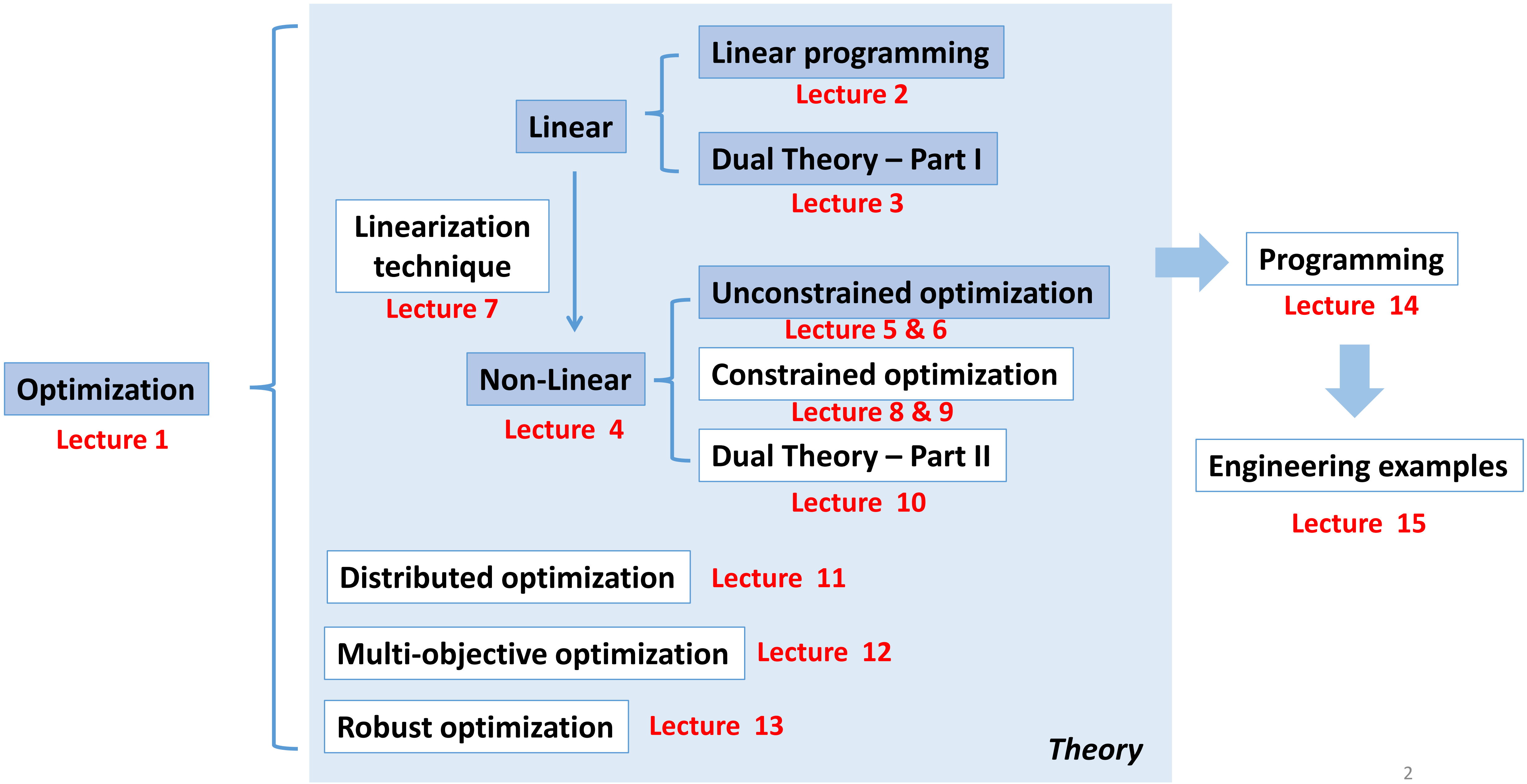
MAEG4070 Engineering Optimization

Lecture 5 Unconstrained Optimization Basics

Yue Chen
MAE, CUHK
email: yuechen@mae.cuhk.edu.hk
Sep 28, 2022

Content of this course (tentative)



Overview

In Lecture 2-3, we introduce the linear programming, which can be solved efficiently by geometrical methods, simplex methods, interior point algorithms, etc; and by mature commercial software, e.g. CPLEX, Gurobi, Lingo, etc.

Most practical problems involve nonlinearility, e.g.

- The cost function of a thermal generator can be modeled as a quadratic function
- The power output of a hydro unit is the product of water head and flow rate

Two ways to deal with the nonlinearility:

- algorithms to solve nonlinear optimization (Lecture 5-6, 8-9);
- linearization techniques (Lecture 7).

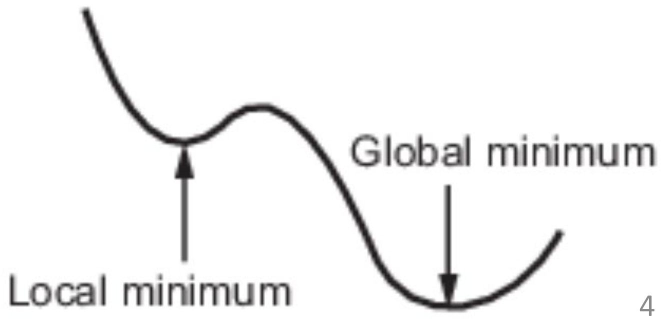
Basic concept

Global optimum. Let $f(x)$ be the objective function, $\mathcal{X}$ be the feasible region, and $x_0 \in \mathcal{X}$. Then x_0 is the global optimum if and only if $f(x) \geq f(x_0), \forall x \in \mathcal{X}$.

Local optimum. Let $f(x)$ be the objective function, $\mathcal{X}$ be the feasible region , and $x_0 \in \mathcal{X}$. If there is a neighborhood of x_0 with radius $\varepsilon > 0$:

$$\mathcal{N}_\epsilon(x_0) = \{x \mid ||x - x_0|| < \epsilon\}$$

Such that $\forall x \in \chi \cap N_\varepsilon(x_0)$, we have $f(x) \geq f(x_0)$. Then x_0 is a local optimum.



Pictures from Google image

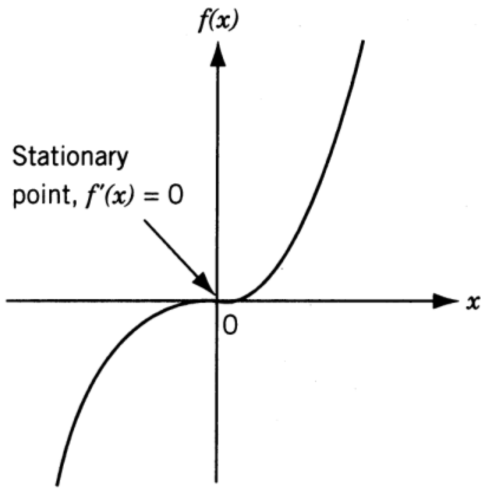
Recall the single variable optimization

Recall what we have learned in Calculus, a ***necessary condition*** for an optimal point is as follows:

Suppose the derivative $df(x)/dx$ exists as a finite number at $x = x^*$. If a function $f(x)$ is defined in the interval $a \leq x \leq b$ and has a local minimum at $x = x^*$, where $a < x^* < b$, we have $df(x^*)/dx = 0$.

Remark-1:
it is a necessary condition, not a sufficient condition.

For example, for $f(x) = x^3$, we have $df(x)/dx = 3x^2$, which equals to 0 when $x = 0$. But…



Pictures from Google image

Recall the single variable optimization

Remark-2:
It is possible that the derivative of $f(x)$ at its maximum or minimum point does not exist.

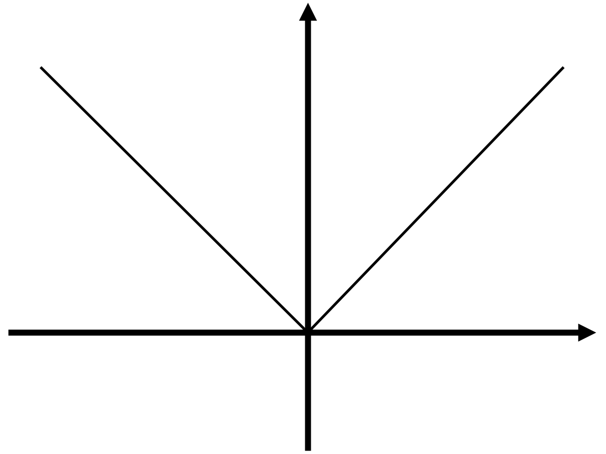
For example, for $f(x) = |x|$ obviously the minimum point is $x^* = 0$. However, its derivative is defined as:

$\lim_{\Delta x \to 0} \frac{|0+\Delta x|-|0|}{\Delta x}$ =1 (positive), -1 (negative)

Therefore, $df(x)/dx$ does not exist.

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$f(x) = |x|$



Recall the single variable optimization

Taylor Expansion

$$f(x^* + \Delta x) = f(x^*) + f'(x^*)\Delta x + \frac{1}{2} f''(x^*)(\Delta x)^2 + \cdots$$

A ***sufficient condition*** for an optimal point is as follows:

Let $f'(x^*) = f''(x^*) = \cdots = f^{(n-1)}(x^*) = 0$, but $f^n(x^*) \neq 0$. Then $x = x^*$ is

- a **minimum point** of $f(x)$ if $f^n(x^*) > 0$ and n is **even**
- a **maximum point** of $f(x)$ if $f^n(x^*) < 0$ and n is **even**
- Neither a minimum nor a maximum point if n is **odd**

For the previous example, $f(x) = x^3$, we have $f'(x) = 3x^2, f''(x) = 6x, f'''(x) = 6$. For $x^* = 0$, due to the above condition, we have $n = 3$, satisfies the 3rd condition, so is neither a minimum nor a maximum.

Example

Determine the maximum and minimum values of the function

$$y = x^5 - 5x^4 + 5x^3 + 1$$

Solution: First, we have the following

$$f'(x) = 5x^4 - 20x^3 + 15x^2$$
$$f''(x) = 20x^3 - 60x^2 + 30x$$
$$f'''(x) = 60x^2 - 120x + 30$$
$$f^{(4)}(x) = 120x - 120$$
$$f^{(5)}(x) = 120$$

Let $f'(x) = 0$, we have $x = 0,1,3$. We check each of them as follows

Example

Solution:

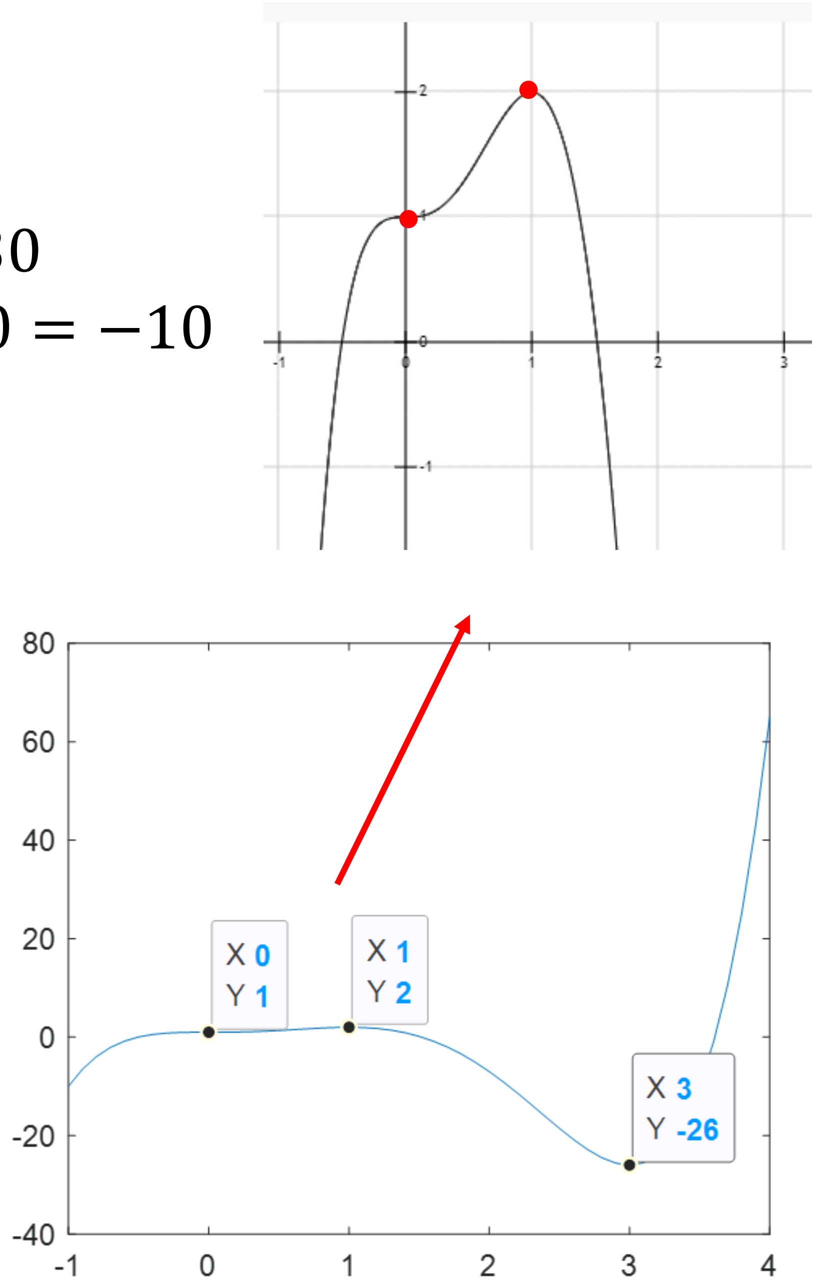
For $x = 0$, we have $f'(0) = 0, f''(0) = 0, f'''(0) = 30$

For $x = 1$, we have $f'(1) = 0, f''(1) = 20 - 60 + 30 = -10$

For $x = 3$, we have $f'(3) = 0, f''(3) = 90$

Therefore

- $x = 0$ is neither a maximum nor a minimum point
- $x = 1$ is a maximum point with $f(1) = 2$
- $x = 3$ is a minimum point with $f(3) = -26$



Multivariable optimization – Review of mathematics

Gradient

The *gradient* of a scalar-valued differentiable function f of several variables is the vector field (or vector-valued function) ∇f, whose value at a point x is the vector, whose components are the partial derivatives of f at x. i.e.

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_N}\right]^T$$

For example, $f(x) = 3x_1^2 + 5x_2 + x_1 x_2$, then

$$\frac{\partial f}{\partial x_1} = 6x_1 + x_2, \frac{\partial f}{\partial x_2} = 5 + x_1$$

Therefore

$$\nabla f(x) = [6x_1 + x_2, 5 + x_1]^T$$

Multivariable optimization – Review of mathematics

Hessian

the Hessian matrix or Hessian is a square matrix of second-order partial derivatives of a scalar-valued function, or scalar field

$$\mathbf{H}(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

$$\frac{\partial f}{\partial x_1} = 6x_1 + x_2, \frac{\partial f}{\partial x_2} = 5 + x_1$$

For example, $f(x) = 3x_1^2 + 5x_2 + x_1 x_2$, we have

$$\frac{\partial^2 f}{\partial x_1^2} = 6, \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 1, \frac{\partial^2 f}{\partial x_2^2} = 0$$

Therefore

$$H(x) = \begin{bmatrix} 6 & 1 \\ 1 & 0 \end{bmatrix}$$

For twice continuously differentiable functions, Hessian is always symmetric. (which can be used to double check the calculation of cross term)

Multivariable optimization – Review of mathematics

Taylor Expansion

A function may be approximated locally by its Taylor series expansion about a point x^*

$$f(x^* + \Delta x) = f(x^*) + f'(x^*)\Delta x + \frac{1}{2} f''(x^*)(\Delta x)^2 + \cdots$$

$$f(x^* + \Delta x) \approx f(x^*) + \nabla f^T(x^*)\Delta x + \frac{1}{2}(\Delta x)^T H(x^*)\Delta x$$

where

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_N}\right]^T$$

$$\mathbf{H}(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

Multivariable optimization – Review of mathematics

Positive/negative-definite matrix

M is symmetric

$$M \text{ positive-definite} \iff \mathbf{x}^\top M\mathbf{x} > 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

$$M \text{ positive semi-definite} \iff \mathbf{x}^\top M\mathbf{x} \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

$$M \text{ negative-definite} \iff \mathbf{x}^\top M\mathbf{x} < 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

$$M \text{ negative semi-definite} \iff \mathbf{x}^\top M\mathbf{x} \leq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

Otherwise, ***M*** is indefinite.

Multivariable optimization – Review of mathematics

Positive-definite and positive semidefinite matrices can be characterized in many ways:

Criterion 1: Denote λ as the eigenvalue, then it satisfies the determinental equation

$$|M - \lambda I| = 0$$

Then

- A matrix M is positive definite if all its eigenvalues are positive.
- A matrix M is negative definite if its eigenvalues are negative.
- A matrix M is positive semi-definite if all its eigenvalues are positive or zero.
- A matrix M is negative semi-definite if its eigenvalues are negative or zero.

Multivariable optimization – Review of mathematics

The eigenvalue of $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$.

$$|A - \lambda I| = \begin{vmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 = 0$$

So we have $\lambda_1 = \lambda_2 = 3$.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1\times5\times9 + 4\times8\times3 + 7\times2\times6 \\ -3\times5\times7 - 1\times6\times8 - 2\times4\times9$$

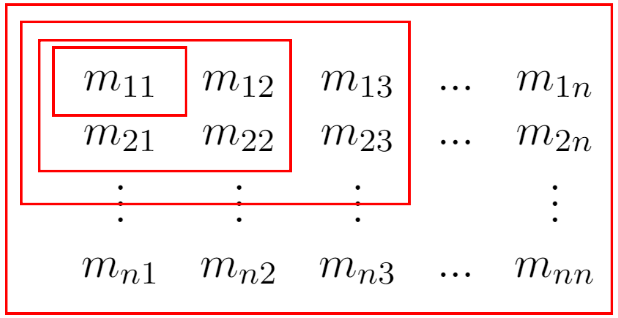
The eigenvalue of $A = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}$.

$$|A - \lambda I| = \begin{vmatrix} 2-\lambda & 2 & 2 \\ 2 & 2-\lambda & 2 \\ 2 & 2 & 2-\lambda \end{vmatrix} = (2-\lambda)^3 + 8 + 8 - 3\times 4(2-\lambda)$$

$$= 8 - 12\lambda + 6\lambda^2 - \lambda^3 + 16 - 24 + 12\lambda = 6\lambda^2 - \lambda^3 = 0$$

So we have $\lambda_1 = \lambda_2 = 0, \lambda_3 = 6$.

Multivariable optimization – Review of mathematics



Criterion 2: Let

$$M_1 = |m_{11}|$$

$$M_2 = \begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix}$$

$$M_3 = \begin{vmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{vmatrix}$$

$$M_n = \begin{vmatrix} m_{11} & m_{12} & m_{13} & \dots & m_{1n} \\ m_{21} & m_{22} & m_{23} & \dots & m_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ m_{n1} & m_{n2} & m_{n3} & \dots & m_{nn} \end{vmatrix}$$

- The matrix M will be positive definite if and only if all the values $M_1, M_2, \dots, M_n$ are positive
- The matrix M will be negative definite if and only if the sign of M_j is $(-1)^j$ for $j = 1,2, \dots, n$
- If some of the M_j are positive and the remaining M_j are zero, the matrix M will be positive semidefinite

Multivariable optimization – Review of mathematics

Consider matrix $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$.

$$M_1 = 3, \qquad M_2 = \begin{vmatrix} 3 & 1 \\ 0 & 3 \end{vmatrix} = 3 \times 3 - 0 \times 1 = 9 > 0$$

So A is a positive definite matrix.

Consider matrix $A = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}$.

$$M_1 = 2, \qquad M_2 = \begin{vmatrix} 2 & 2 \\ 2 & 2 \end{vmatrix} = 0, \qquad M_3 = \begin{vmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{vmatrix} = 0$$

So A is a positive semidefinite matrix.

Example

Consider matrix $M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, we try to prove that it is positive definite in three ways.

1. By definition

For any non-zero vector $z = [x, y]^T$, we have

$$z^T M z = [x, y] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x \;\; y] \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + y^2 > 0$$

2. Calculate the eigenvalue

$$|M - \lambda I| = \begin{vmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 = \lambda^2 - 2\lambda + 1 = 0$$

The eigenvalues are $\lambda_1 = \lambda_2 = 1$.

3. $M_1 = 1, M_2 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$

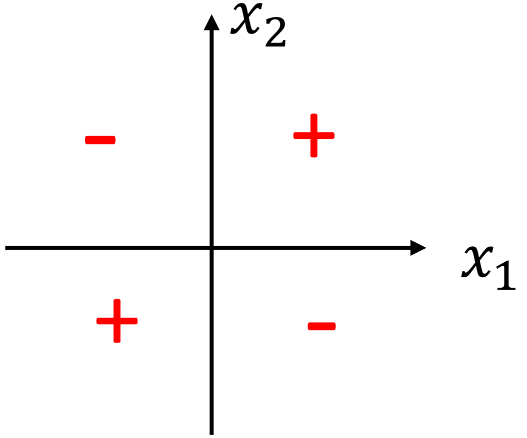
Multivariable optimization

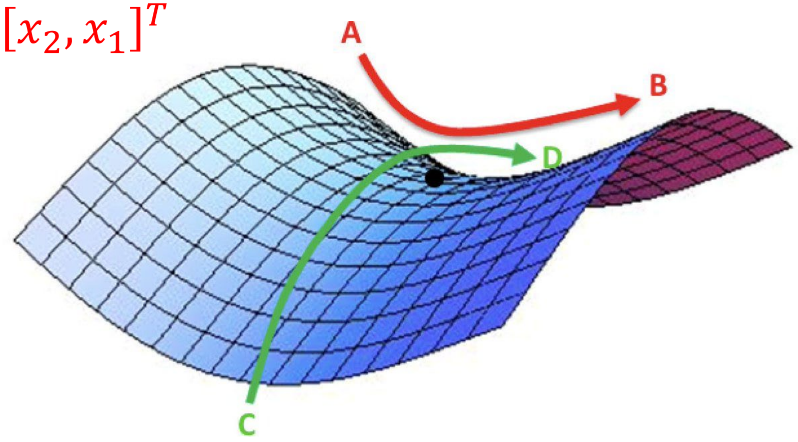
First-order necessary condition: If $f(x)$ has an extreme point at $x = x^*$, and its gradient exists at point x^*, then $\nabla f(x^*) = \mathbf{0}^{\mathbf{T}}$.
vector

Remark: if the gradient of $f(x)$ exists at point x^* and $\nabla f(x^*) = \mathbf{0}^{\mathbf{T}}$, then $x = x^*$ is called a "stationary point"; if a stationary point $x = x^*$ is neither a maximum nor minimum point, then it is called a "saddle point".

For example, for function $f(x) = x_1 x_2$, $x^* = (0{,}0)^T$ is a stationary point and a saddle point. (try to prove it)

$\nabla f = [x_2, x_1]^T$





Pictures from Google image

Multivariable optimization

Second-order necessary condition: If $f(x)$ has a minimum point at $x = x^*$, and it is twice-differentiable at x^*, then $\nabla f(x^*) = 0$ and its Hessian $H(x^*)$ is positive semi-definite.

Proof: Let d be any non-zero vector. λ is a nonzero scalar that can be positive or negative. According to the first-order necessary condition, we have $\nabla f(x) = 0$.

$$f(x^* + \lambda d) = f(x^*) + \lambda \nabla f(x^*)^T d + \frac{1}{2}\lambda^2 d^T \boxed{\nabla^2 f(x^*)} d + \lambda^2 ||d||^2 o(x^*, \lambda d)$$

$\boldsymbol{H(x^*)}$

Then

$$\frac{f(x^* + \lambda d) - f(x^*)}{\lambda^2} = \frac{1}{2} d^T \nabla^2 f(x^*) d + ||d||^2 o(x^*, \lambda d)$$

When $\lambda \to 0$, we have $o(x^*, \lambda d) \to 0$.

Since $x = x^*$ is a relative minimum point, then

$$f(x^* + \lambda d) \geq f(x^*)$$

Therefore, let $\lambda \to 0$, we have

$$d^T \nabla^2 f(x^*) d \geq 0$$

Multivariable optimization

Sufficient condition: If $f(x)$ is twice-differentiable at x^*, $\nabla f(x^*) = 0$ and its Hessian $H(x^*)$ is positive definite, then $x = x^*$ is a strict local minimum point.

Proof: Let d be any non-zero vector.

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

$$f(x^* + \lambda d) = f(x^*) + \lambda \nabla f(x^*)^T d + \frac{1}{2}\lambda^2 d^T \nabla^2 f(x^*) d + \lambda^2 ||d||^2 o(x^*, \lambda d)$$

Since $\nabla^2 f(x)$ is positive definite, $d^T \nabla^2 f(x^*) d > 0$.
There exists δ, given any $\lambda \in (0, \delta]$

$$\frac{1}{2}\lambda^2 d^T \nabla^2 f(x^*) d + \lambda^2 ||d||^2 o(x^*, \lambda d) > 0$$

Therefore, $f(x^* + \lambda d) > f(x^*)$.

this means $f(x^*)$ is strict minimum in any ball centered at x^* with radius $\delta||d|| = \delta$.

Remark: If $H(x^*)$ is positive semi-definite, then $x = x^*$ is a relative minimum point.

$\geq$

Example

Find the minimum point of $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$

Solution:
The gradient of $f(x)$ is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4(x_1 - 2)^3 + 2(x_1 - 2x_2) \\ -4(x_1 - 2x_2) \end{bmatrix}$$

The Hessian matrix is

$$H(x) = \begin{bmatrix} 12(x_1 - 2)^2 + 2 & -4 \\ -4 & 8 \end{bmatrix}$$

Let $\nabla f(x) = 0$, we have $x^* = (2, 1)^T$ and

$$H(x^*) = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix}$$

Since $H(x^*)$ is positive semi-definite, x^* is a *local relative* minimum point.

Example

Find the minimum point of $f(x) = 5x_1^2 - 6x_1x_2 + 5x_2^2$

Solution:

The gradient of $f(x)$ is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 10x_1 - 6x_2 \\ -6x_1 + 10x_2 \end{bmatrix}$$

The Hessian matrix is

$$H(x) = \begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix}$$

Let $\nabla f(x) = 0$, we have $x^* = (0, 0)$.

Since $H(x^*)$ is positive definite, x^* is a strict minimum point.

Hessian is positive semi-definite (definite) everywhere in a domain implies (strict) convex function in that domain

Actually, it is a *global* minimum point since $f(x)$ is a convex function.

Multivariable optimization

Necessary and sufficient condition: If $f(x)$ is twice-differentiable at x^* and is a <u>convex function</u>, then x^* is a global minimum if and only if $\nabla f(x^*) = 0$.

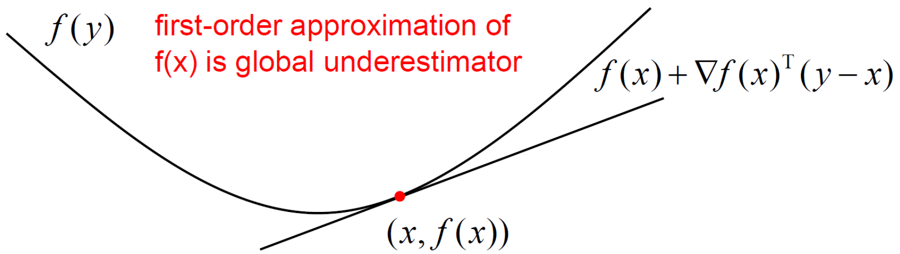
Proof:

$\Rightarrow$ If x^* is a global minimum, then it is a local minimum. According to the first-order necessary condition, we have $\nabla f(x^*) = 0$.

$\Leftarrow$ If $\nabla f(x^*) = 0$, then for any $x \in \mathbb{R}^n$, we have $\nabla f(x^*)^T(x - x^*) = 0$. Since $f(x)$ is a convex function, we have

$$f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*) = f(x^*)$$

Therefore, x^* is a global minimum.



Summary

First-order necessary condition: If $f(x)$ has an extreme point at $x = x^*$, and its gradient exists at point x^*, then $\nabla f(x^*) = 0$.

$f(x^*)$ **is optimal → ?**

Second-order necessary condition: If $f(x)$ has a minimum point at $x = x^*$, and it is twice-differentiable at x^*, then $\nabla f(x^*) = 0$ and its Hessian $H(x^*)$ is positive semi-definite.

? → $f(x^*)$ **is optimal**

Sufficient condition: If $f(x)$ is twice-differentiable at x^*, $\nabla f(x^*) = 0$ and its Hessian $H(x^*)$ is positive definite, then $x = x^*$ is a strict minimum point.

Remark: If $H(x^*)$ is positive semi-definite, then $x = x^*$ is a relative minimum point.

$f(x^*)$ **is optimal ↔ ?**

Necessary and sufficient condition: If $f(x)$ is twice-differentiable at x^* and is a convex function, then x^* is a global minimum if and only if $\nabla f(x^*) = 0$.

Example

Find the minimum point of $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$

Solution:
The gradient of $f(x)$ is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4(x_1 - 2)^3 + 2(x_1 - 2x_2) \\ -4(x_1 - 2x_2) \end{bmatrix}$$

The Hessian matrix is

$$H(x) = \begin{bmatrix} 12(x_1 - 2)^2 + 2 & -4 \\ -4 & 8 \end{bmatrix}$$

Let $\nabla f(x) = 0$, we have $x^* = (2, 1)^T$ and

$$H(x^*) = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix}$$

Since $H(x^*)$ is positive semi-definite, x^* is a *local relative* minimum point.

Example

Find the minimum point of $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2 + (x_1 - 2)^2$

Solution:
The gradient of $f(x)$ is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4(x_1 - 2)^3 + 2(x_1 - 2x_2) + 2(x_1 - 2) \\ -4(x_1 - 2x_2) \end{bmatrix}$$

The Hessian matrix is

$$H(x) = \begin{bmatrix} 12(x_1 - 2)^2 + 4 & -4 \\ -4 & 8 \end{bmatrix}$$

Let $\nabla f(x^*) = 0$, we have $x^* = (2, 1)^T$ and

$$H(x^*) = \begin{bmatrix} 4 & -4 \\ -4 & 8 \end{bmatrix}$$

Since $H(x^*)$ is positive definite, x^* is a *local strict* minimum point.

Example

Find the minimum point of $f(x) = 7x_1^2 - 3x_1x_2 + 4x_2^2$

Solution:
The gradient of $f(x)$ is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 14x_1 - 3x_2 \\ -3x_1 + 8x_2 \end{bmatrix}$$

The Hessian matrix is

$$H(x) = \begin{bmatrix} 14 & -3 \\ -3 & 8 \end{bmatrix}$$

Let $\nabla f(x) = 0$, we have $x^* = (0, 0)$.
Since $H(x^*)$ is positive definite, x^* is a *global strict* minimum point.

Comparison

Example 1 **local**

The Hessian matrix is

$$H(x) = \begin{bmatrix} 12(x_1-2)^2+2 & -4 \\ -4 & 8 \end{bmatrix}$$

Let $\nabla f(x) = 0$, we have $x^* = (2,1)^T$ and

$$H(x^*) = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix}$$

Relative minimum

Example 2

The Hessian matrix is

$$H(x) = \begin{bmatrix} 12(x_1-2)^2+4 & -4 \\ -4 & 8 \end{bmatrix}$$

Let $\nabla f(x^*) = 0$, we have $x^* = (2,1)^T$ and

$$H(x^*) = \begin{bmatrix} 4 & -4 \\ -4 & 8 \end{bmatrix}$$

Strict minimum

Example 3 **global**

The Hessian matrix is

$$H(x) = \begin{bmatrix} 14 & -3 \\ -3 & 8 \end{bmatrix}$$

Question 1:
If is a convex function?
Hessian matrix positive definite/semi-definite over the whole domain?

- Yes: global
- No: local

Question 2:
the Hessian matrix at a specific point

- Positive definite: strict minimum
- Positive semidefinite: relative minimum

Thanks！