MAEG4070 Engineering Optimization

Summary of Lecture 5-7

Yue Chen
MAE, CUHK
email: yuechen@mae.cuhk.edu.hk
Oct 10, 2022

What have we learned?

Lecture 5:

- Single-variable optimization (necessary condition & sufficiency condition)
- Multivariable optimization (necessary condition & sufficiency condition)

Lecture 6:

- Gradient descent method
- Newton method

Lecture 7:

- Linearization techniques
  - minimizing a convex piecewise linear function
  - A piecewise linear function in constraints
  - the product of a binary and a continuous variable
  - complementary and slackness condition in KKT condition
  - minimum values/maximum values

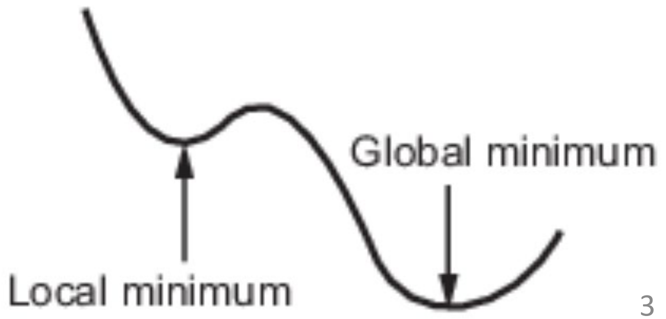
Basic concept

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.



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 $a \le x \le b$ and has a local minimum at $x = x^*$, where $a < x^* < b$, we have $df(x)/dx = 0$.

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**

Example

Determine the optimal the maximum and minimum values of the function:

$$f(x) = 12x^5 - 45x^4 + 40x^3 + 5$$

Solution: Since $f'(x) = 60\left(x^4 - 3x^3 + 2x^2\right) = 60x^2(x-1)(x-2)$
Let $f'(x) = 0$, we have $x = 0, x = 1$, and $x = 2$.

The second derivative is

$$f''(x) = 60(4x^3 - 9x^2 + 4x)$$

- $f''(1) = -60$ and hence $x = 1$ is a relative maximum and $f_{max} = 12$.
- $f''(2) = 240$ and hence $x = 2$ is a relative minimum and $f_{min} = -11$.
- $f''(0) = 0$, so we must investigate the next derivative

$$f'''(x) = 60(12x^2 - 18x + 4) = 240 \text{ at } x = 0$$

Therefore, $x = 0$ is neither a maximum nor a minimum.

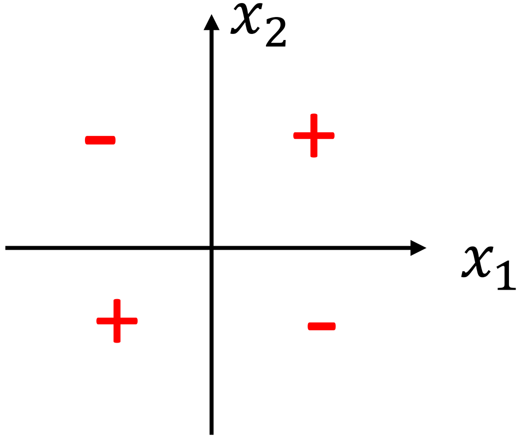
Multivariable optimization

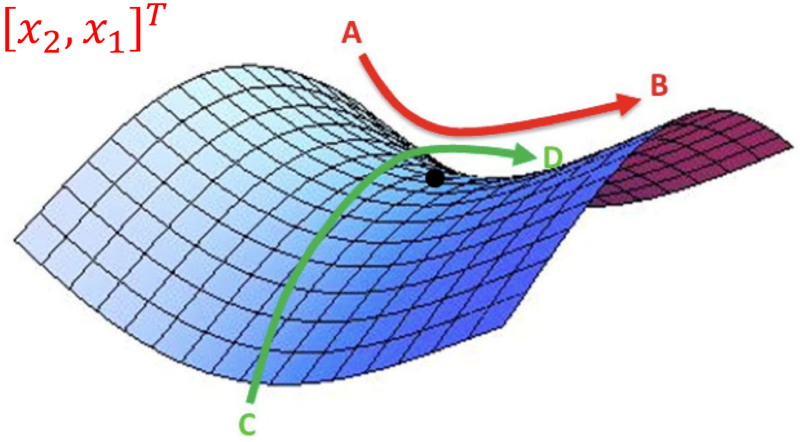
First-order necessary condition: If $f(x)$ has an extreme point at $x = x^*$, and its gradient exists at point x^*, then $\nabla f(x^*) = \mathbf{0}^{\mathbf{T}}$.
vector

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 <u>positive semi-definite</u>.

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

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

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

Review of mathematics

Consider matrix $M = \begin{bmatrix} 3 & 3 \\ 3 & 4 \end{bmatrix}$, 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

$$z^T M z = [x, y] \begin{bmatrix} 3 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [3x + 3y \;\; 3x + 4y] \begin{bmatrix} x \\ y \end{bmatrix}$$
$$= 3x^2 + 6xy + 4y^2 = 3(x + y)^2 + y^2 > 0$$

2. Calculate the eigenvalue

$$|M - \lambda I| = \begin{vmatrix} 3 - \lambda & 3 \\ 3 & 4 - \lambda \end{vmatrix} = (3 - \lambda)(4 - \lambda) - 9 = \lambda^2 - 7\lambda + 3 = 0$$

The eigenvalues are $\lambda_1, \lambda_2 = \frac{7 \pm \sqrt{49 - 12}}{2}$.

3. $M_1 = 3, M_2 = \begin{vmatrix} 3 & 3 \\ 3 & 4 \end{vmatrix} = 12 - 9 = 3.$

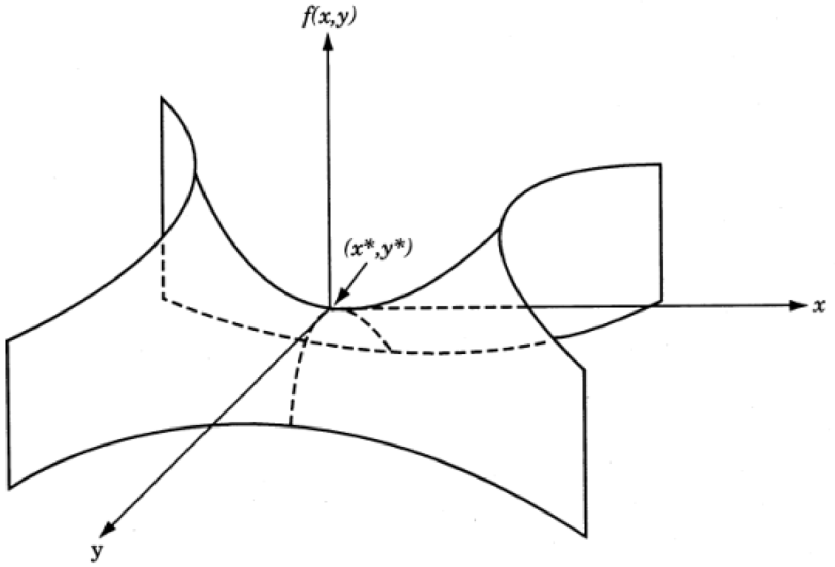
Example-1

Consider the function $f(x, y) = x^2 - y^2$
We have

$$\frac{\partial f}{\partial x} = 2x, \frac{\partial f}{\partial y} = -2y$$

These first derivatives are zero at $x^* = 0, y^* = 0$
The Hessian matrix of f at (x^*, y^*) is given by

$$H = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$



Since this matrix is neither positive definite nor negative definite, the point (x^*, y^*) is a saddle point.

It can be seen that $f(x, y^*) = f(x, 0)$ has a relative minimum and $f(x^*, y) = f(0, y)$ has a relative maximum at the saddle point (x^*, y^*).

Example-2

Find the extreme points of the function

$$f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$$

Solution: The necessary conditions for the existence of an extreme point are:

$$\frac{\partial f}{\partial x_1} = 3x_1^2 + 4x_1 = x_1(3x_1 + 4) = 0$$

$$\frac{\partial f}{\partial x_2} = 3x_2^2 + 8x_2 = x_2(3x_2 + 8) = 0$$

These equations are satisfied at the points: $(0,0)$, $\left(0, -\frac{8}{3}\right)$, $\left(-\frac{4}{3}, 0\right)$, $(-\frac{4}{3}, -\frac{8}{3})$.

Example-2

The second-order partial derivatives of f are:

$$\frac{\partial^2 f}{\partial {x_1}^2} = 6x_1 + 4 \qquad \frac{\partial^2 f}{\partial {x_2}^2} = 6x_2 + 8 \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$$

The Hessian matrix of f is given by:

$$H(x) = \begin{bmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{bmatrix}$$

To check whether the nature of $H(x)$, we calculate

$$H_1 = 6x_1 + 4$$

$$H_2 = \begin{vmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{vmatrix}$$

Example-2

Point $\boldsymbol{x}$	Value of $\boldsymbol{H_1}$	Value of $\boldsymbol{H_2}$	Nature of $\boldsymbol{H}$	Nature of $\boldsymbol{x}$	$f(\boldsymbol{x})$
(0,0)	+4	+32	Positive definite	local Strict minimum	6
$(0, -\frac{8}{3})$	+4	-32	Indefinite	Saddle point	$\frac{418}{27}$
$(-\frac{4}{3}, 0)$	-4	-32	Indefinite	Saddle point	$\frac{194}{27}$
$(-\frac{4}{3}, -\frac{8}{3})$	-4	+32	Negative definite	local Strict maximum	$\frac{50}{3}$

Example-2

For a 2×2 matrix $A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, if:

(i) $a > 0, ad - bc > 0$, it is positive definite

(ii) $a < 0, ad - bc > 0$, it is negative definite

This is because $-A = \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$

and $-a > 0$, $(-a)(-d) - (-b)(-c) = ad - bc > 0$

Hence $-A$ is positive definite, and A is negative definite.

(iii) others, it is indefinite

Suppose a point x^* satisfies $\nabla f(x^*) = 0$ and $H(x^*)$ is negative definite.
Then, we have $-\nabla f(x^*) = 0$ and $-H(x^*)$ is positive definite.
Hence, x^* is the strict optimum of $\min_x -f(x)$,
which is equivalent to $\max_x f(x)$.
Hence, x^* is a strict maximum of $f(x)$.

Example-3

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

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

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

The Hessian matrix is

$$H(x) = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$$

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

Example-4

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

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

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 12x_1 - 2x_2 \\ -2x_1 + 2x_2 \end{bmatrix}$$

The Hessian matrix is

$$H(x) = \begin{bmatrix} 12 & -2 \\ -2 & 2 \end{bmatrix}$$

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

Gradient-based algorithms

Algorithm: Choose initial point $x_0 \in \mathbb{R}^n$, repeat:

Gradient Descent: $x_k = x_{k-1} - \alpha \nabla f(x_{k-1})$

Or Newton: $x_k = x_{k-1} - \left[\nabla^2 f(x_{k-1})\right]^{-1} \nabla f(x_{k-1})$

Stop until convergence, e.g. $\|x_k - x_{k-1}\| \leq \varepsilon$

Gradient Descent

Interpretation:

If we approximate the Hessian $\nabla^2 f$ by $\frac{1}{\alpha} I$, then

$$f(y) \approx f(x) + \nabla f(x)^T (y - x) + \boxed{\frac{1}{2\alpha} \|y - x\|_2^2}$$

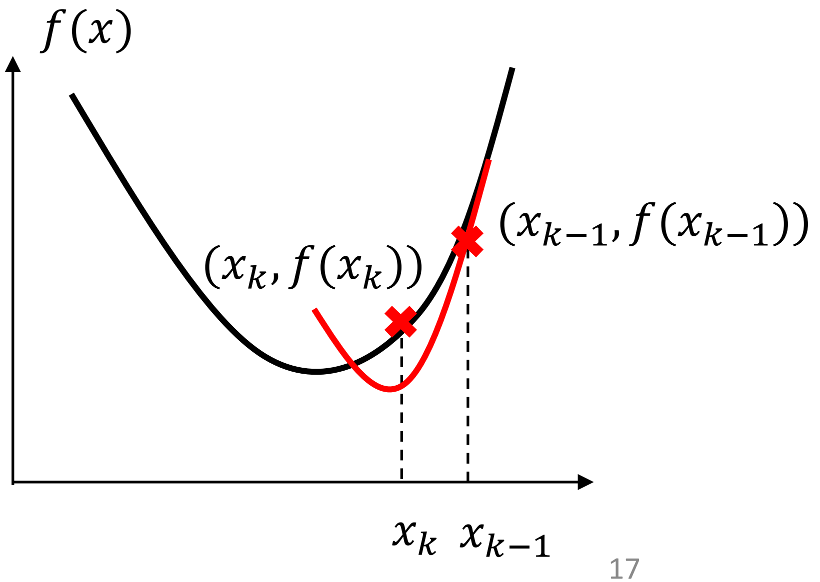
Let $x = x_{k-1}$, we want to choose $x_k = y$ that minimizes $f(y)$

$$\min_y \frac{1}{2\alpha} \|y - x\|_2^2 + \nabla f(x)^T (y - x)$$

$$\frac{1}{\alpha}(y - x) + \nabla f(x_{k-1}) = 0$$

Therefore

$$x_k = x_{k-1} - \alpha \nabla f(x_{k-1})$$



Newton Method

Interpretation:

Consider the second-order Taylor approximation

$$f(y) \approx f(x) + \nabla f(x)^T (y - x) + \boxed{\frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x)}$$

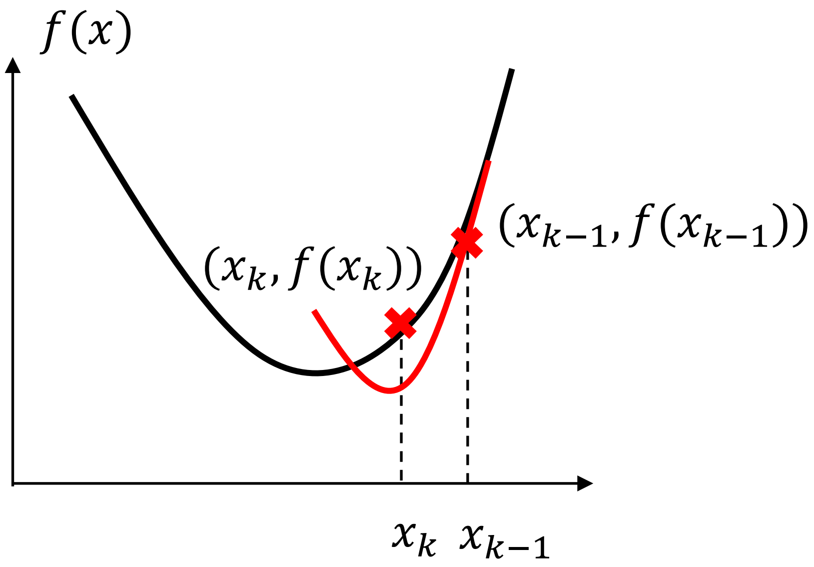
Assume $\nabla^2 f(x)$ is positive definite, so that $f(x)$ has a strict global optimum. Let $x = x_{k-1}$, we want to choose $x_k = y$ that minimizes $f(y)$

$$\min_y \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) + \nabla f(x)^T (y - x)$$

Therefore

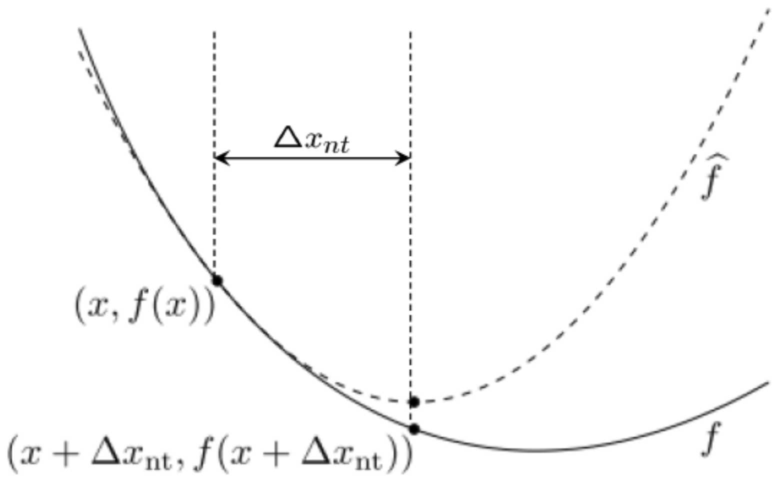
$$x_k = x_{k-1} - \left[\nabla^2 f(x_{k-1})\right]^{-1} \nabla f(x_{k-1})$$

Comparison of Gradient Descent & Newton Method



Gradient Descent

$$x_k = x_{k-1} - \alpha \nabla f(x_{k-1})$$



Newton Method

$$x_k = x_{k-1} - \left[\nabla^2 f(x_{k-1})\right]^{-1} \nabla f(x_{k-1})$$

Tradeoff: Newton method takes fewer steps, but more time for each step

Pictures from Google image

Example

Solve the optimization $\min_{x_1,x_2} f(x) = x_1^2 + 25x_2^2$ for one step, using gradient descent and Newton method, respectively. Choose $\alpha = 0.1$.

Solution: Let $x^{(0)} = (2,2)^T$, then

$$\nabla f\left(x^{(0)}\right) = \begin{pmatrix} 2x_1 \\ 50x_2 \end{pmatrix}\Bigg|_{x^{(0)}} = \begin{pmatrix} 4 \\ 100 \end{pmatrix}$$

$$\nabla^2 f\left(x^{(0)}\right) = \begin{pmatrix} 2 & 0 \\ 0 & 50 \end{pmatrix}, \nabla^2 f\left(x^{(0)}\right)^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{50} \end{pmatrix}$$

Gradient descent: $x^{(1)} = x^{(0)} - \alpha \nabla f\left(x^{(0)}\right) = \begin{pmatrix} 1.6 \\ -8 \end{pmatrix}$

Newton method: $x^{(1)} = x^{(0)} - \nabla^2 f\left(x^{(0)}\right)^{-1} \nabla f\left(x^{(0)}\right)$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{50} \end{pmatrix} \begin{pmatrix} 4 \\ 100 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Example

Solve the optimization $\min_{x_1,x_2} f(x) = 4x_1^2 + x_2^2 - x_1^2 x_2$ using Newton method, with initial points $x_A = (1,1)^T, x_B = (3,4)^T$, and $x_c = (2,0)^T$, respectively.

Solution:

The gradient is

$$\nabla f(x) = \left(8x_1 - 2x_1x_2, 2x_2 - x_1^2\right)^T$$

The Hessian matrix is

$$\nabla^2 f(x) = \begin{pmatrix} 8 - 2x_2 & -2x_1 \\ -2x_1 & 2 \end{pmatrix}$$

Example

$$x^{(0)} = x_A = (1,1)^T$$

k	$x^{(k)}$	$f(x^{(k)})$	$\nabla f(x^{(k)})$	$\|\|\nabla f(x^{(k)})\|\|$	$\nabla^2 f(x^{(k)})$
0	1.0000 1.0000	4.000	6.0000 1.0000	6.0828	6.0000 -2.0000 -2.0000 2.0000
1	-0.7500 -1.2500	4.5156	-7.8750 -3.0625	8.4495	10.500 1.5000 1.5000 2.0000
2	-0.1550 -0.1650	0.1273	-1.2911 -0.3540	1.3388	8.3300 0.3100 0.3100 2.0000
3	-0.0057 -0.0111	0.0003	-0.0459 -0.0223	0.0511	8.0222 0.0115 0.0115 2.0000
4	-0.0000 -0.0000	0.0000	-0.0001 -0.0000	0.0001	8.0000 0.0000 0.0000 2.0000

Example

$$x^{(0)} = x_B = (3,4)^T$$

k	$x^{(k)}$	$f(x^{(k)})$	$\nabla f(x^{(k)})$	$\|\|\nabla f(x^{(k)})\|\|$	$\nabla^2 f(x^{(k)})$
0	3.0000 4.0000	16.000	0.0000 -1.0000	1.0000	0.0000 -6.0000 -6.0000 2.0000
1	2.8333 4.0000	16.000	0.0000 -0.2078	0.0278	0.0000 -5.6667 -5.6667 2.0000
2	2.8284 4.0000	16.000	0.0000 0.0000	0.0000	0.0000 -5.6569 -5.6569 2.0000

indefinite

Example

$$x^{(0)} = x_c = (2,0)^T$$

$\nabla^2 f\left(x^{(0)}\right) = \begin{pmatrix} 8 & -4 \\ -4 & 2 \end{pmatrix}$ which is irreversible, cannot calculate $x^{(1)}$.

Applying Newton method may:

- Converges to the minimum point
- Converges to the saddle point
- Hessian matrix is irreversible, cannot proceed

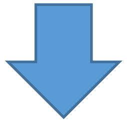
Linearization techniques

Minimizing a convex piecewise linear function (univariate)

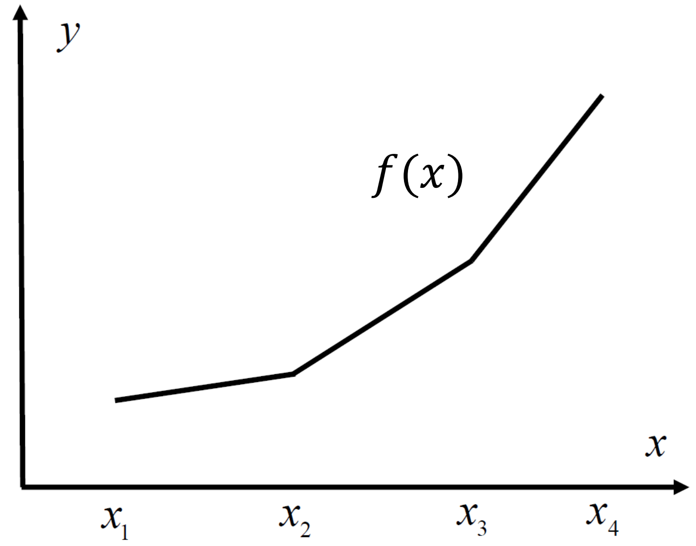
$$\min_{x} \; f(x)$$
$$\text{s.t. } x_1 \le x \le x_4$$

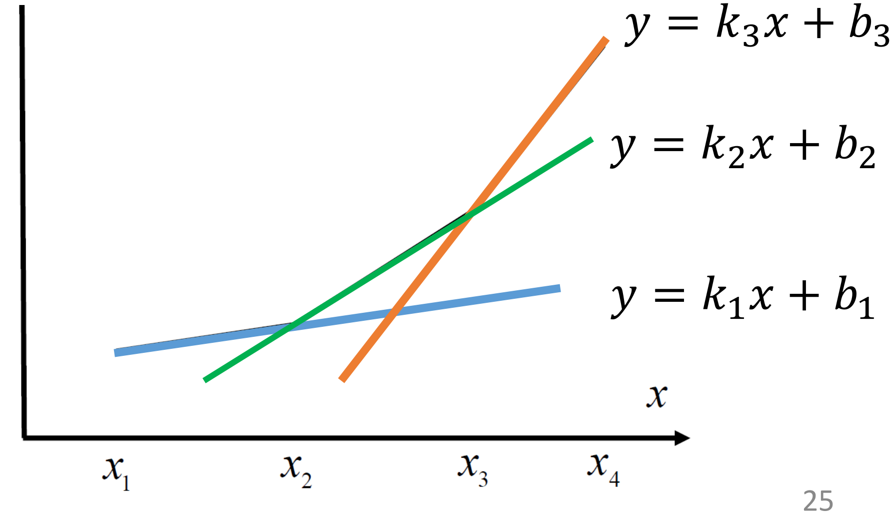
where

$$f(x) = \begin{cases} k_1 x + b_1, & x \in [x_1, x_2] \\ k_2 x + b_2, & x \in [x_2, x_3] \\ k_3 x + b_3, & x \in [x_3, x_4] \end{cases}$$



$$\min_{x,\sigma} \; \sigma$$
$$\text{s.t. } \sigma \ge k_1 x + b_1$$
$$\sigma \ge k_2 x + b_2$$
$$\sigma \ge k_3 x + b_3$$
$$x_1 \le x \le x_4$$





Linearization techniques

Minimizing a convex piecewise linear function (univariate)

Another equivalent form

$$\min_{x,y,\lambda} \ y$$

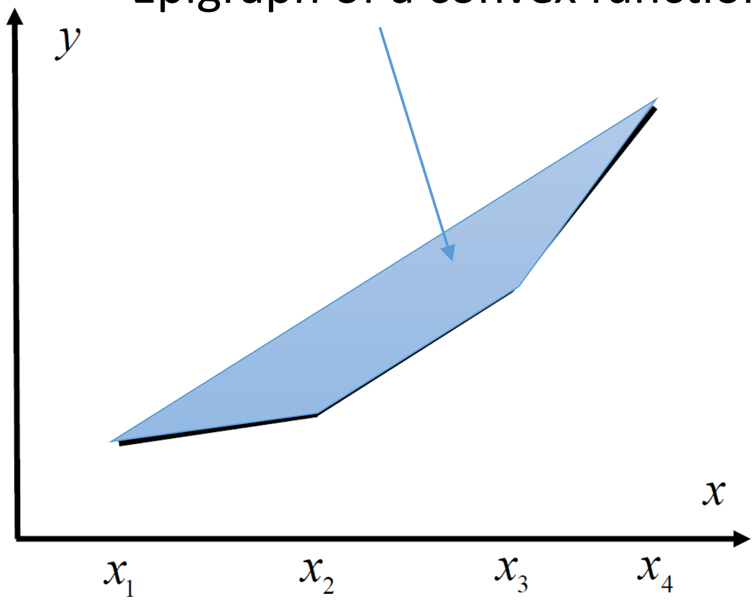
$$\text{s.t. } x = \sum_{n=1}^{N} \lambda_n x_n$$

$$y = \sum_{n=1}^{N} \lambda_n f(x_n)$$

$$0 \le \lambda_n \le 1, \forall n = 1, ..., N$$

$$\sum_{n=1}^{N} \lambda_n = 1$$

Epigraph of a convex function is a convex set



Example

$$\min_{x} \ f(x) \quad \text{s.t. } 1 \le x \le 4$$

where

$$f(x) = \begin{cases} 2x + 1, & 1 \le x \le 2 \\ 3x - 1, & 2 \le x \le 4 \end{cases}$$

Method 1:

$$\begin{aligned} \min_{x,\sigma} \ & \sigma \\ \text{s.t. } & \sigma \ge 2x + 1 \\ & \sigma \ge 3x - 1 \\ & 1 \le x \le 4 \end{aligned}$$

Method 2:

$$\begin{aligned} \min_{x,y,\lambda} \ & y \\ \text{s.t. } & x = \lambda_1 + 2\lambda_2 + 4\lambda_3 \\ & y = 3\lambda_1 + 5\lambda_2 + 11\lambda_3 \\ & 0 \le \lambda_1, \lambda_2, \lambda_3 \le 1 \\ & \lambda_1 + \lambda_2 + \lambda_3 = 1 \end{aligned}$$

Linearization techniques

Linearize the product of a binary and a continuous variable

Consider $z = xy, x \in [x_l, x_u], y \in \{0,1\}$
It can be linearized by

$$x_l y \leq z \leq x_u y$$
$$x_l(1 - y) \leq x - z \leq x_u(1 - y)$$

Proof of equivalence:

1. If $y = 0$, then the first inequality becomes $z = 0$ and the second $x_l \leq x \leq x_u$. Meanwhile, we have $z = xy = 0$.
2. If $y = 1$, then the second inequality becomes $x = z$ and the first $x_l \leq x = z \leq x_u$. Meanwhile, we have $z = xy = x$.

Linearization techniques

Complementary condition in KKT condition

Consider condition $0 \leq x \perp y \geq 0$
It is equivalent to $x, y \geq 0, xy = 0$
And can be linearized by

$$0 \leq x \leq Mz$$
$$0 \leq y \leq M(1 - z)$$
$$z \in \{0,1\}^n$$

Proof of equivalence:

1. If $x = 0, y > 0$, then let $z = 0$
2. If $x > 0, y = 0$, then let $z = 1$
3. If $x = 0, y = 0$, then let $z = 0$ or $z = 1$

Remark: *M* can be chosen as the upper bound of the values of x, y; called Big-M method in literature.

Linearization techniques

Minimum values

Consider $y = \min\{x_1, \dots, x_n\}, x_i \in [x_i^l, x_i^u]$
Let $L = \min\{x_1^l, \dots, x_n^l\}$. It can be represented as

$$x_i^l \le x_i \le x_i^u, \forall i$$
$$y \le x_i, \forall i$$
$$x_i - (x_i^u - L)(1 - z_i) \le y, \forall i$$
$$z_i \in \{0,1\}, \sum_{i=1}^{n} z_i = 1$$

Proof of equivalence:

- Only one $z_i = 1$ and others =0.
- If $z_i = 1$, we have $x_i^l \le x_i \le x_i^u, y \le x_i, x_i \le y$
- If $z_i = 0$, we have $x_i^l \le x_i \le x_i^u, y \le x_i, x_i - y \le x_i^u - L$

Linearization techniques

Maximum values

Consider $y = \max\{x_1, \dots, x_n\}, x_i \in [x_i^l, x_i^u]$
Let $U = \max\{x_1^u, \dots, x_n^u\}$. It can be represented as

$$x_i^l \le x_i \le x_i^u, \forall i$$
$$y \ge x_i, \forall i$$
$$x_i + (U - x_i^l)(1 - z_i) \ge y, \forall i$$
$$z_i \in \{0,1\}, \sum_{i=1}^{n} z_i = 1$$

Proof of equivalence:

- Only one $z_i = 1$ and others =0.
- If $z_i = 1$, we have $x_i^l \le x_i \le x_i^u, y \ge x_i, x_i \ge y$
- If $z_i = 0$, we have $x_i^l \le x_i \le x_i^u, y \ge x_i, y - x_i \le U - x_i^l$

Example

Consider $z = 5xy, x \in [4,8], y \in \{0,1\}$
It can be linearized by

$$20y \leq z \leq 40y$$

$$20(1-y) \leq 5x - z \leq 40\ (1-y)$$

Consider $y = \min\{x_1, x_2, x_3\}$, $x_1 \in [1,10], x_2 \in [0,8], x_3 \in [3,12]$
Let $L = \min\{1,0,3\} = 0$. It can be represented as

$$1 \leq x_1 \leq 10, 0 \leq x_2 \leq 8, 3 \leq x_3 \leq 12$$

$$y \leq x_1, y \leq x_2, y \leq x_3$$

$$x_1 - 10(1 - z_1) \leq y$$

$$x_2 - 8(1 - z_2) \leq y$$

$$x_3 - 12(1 - z_3) \leq y$$

$$z_i \in \{0,1\}, \sum_{i=1}^{3} z_i = 1$$

Thanks！