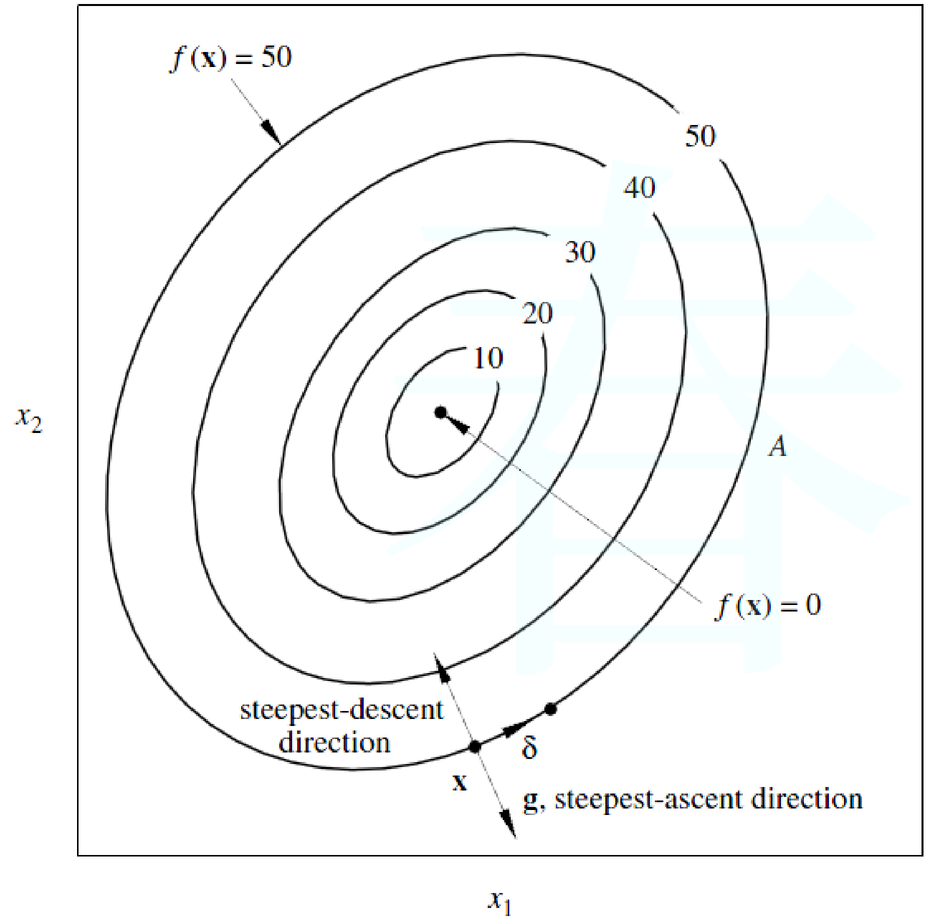
3.Gradient methods

Gradient methods

- 1. Steepest-descent method
- 2. Newton method
- 3. Gauss-Newton Method

Ascent and descent directions

the gradient at point x is orthogonal to contour A



$$F + \Delta F = f(\mathbf{x} + \boldsymbol{\delta}) \approx f(\mathbf{x}) + \mathbf{g}^T\boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}^T\mathbf{H}\boldsymbol{\delta}$$

and as $\|\boldsymbol{\delta}\| \to 0$, the change in F due to change $\boldsymbol{\delta}$ is obtained as

$$\Delta F \approx \mathbf{g}^T\boldsymbol{\delta}$$

$$\Delta F \approx \|\mathbf{g}\|\,\|\boldsymbol{\delta}\|\ \cos\theta$$

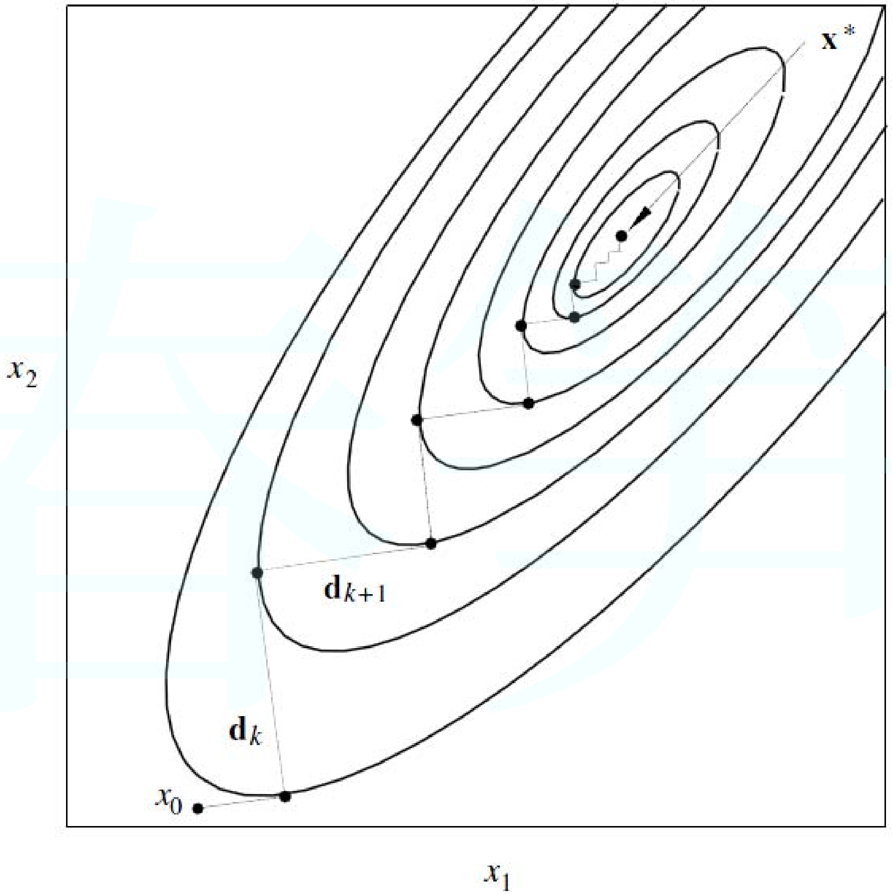
steepest-descent direction

$$\mathbf{d} = -\mathbf{g}$$

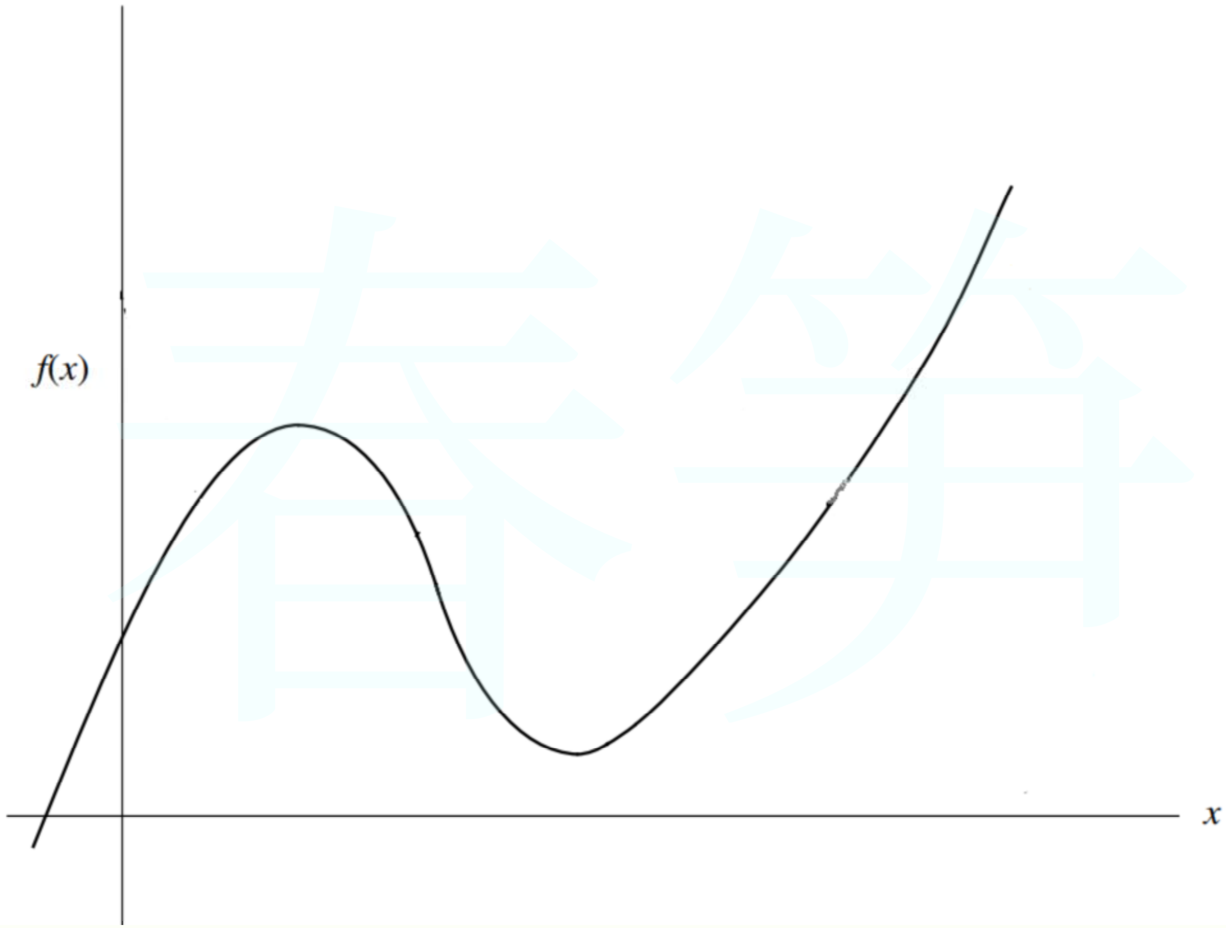
$$\boldsymbol{\delta} = \alpha \mathbf{d}$$

$$\mathbf{d}_{k+1}^T \mathbf{d}_k = 0$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$



steepest-descent direction



Algorithm 5.1 Steepest-descent algorithm

Step 1

Input $\mathbf{x}_0$ and initialize the tolerance ε.

Set $k = 0$.

Step 2

Calculate gradient $\mathbf{g}_k$ and set $\mathbf{d}_k = -\mathbf{g}_k$.

Step 3

Find α_k, the value of α that minimizes $f(\mathbf{x}_k + \alpha\mathbf{d}_k)$, using a line search.

Step 4

Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$ and calculate $f_{k+1} = f(\mathbf{x}_{k+1})$.

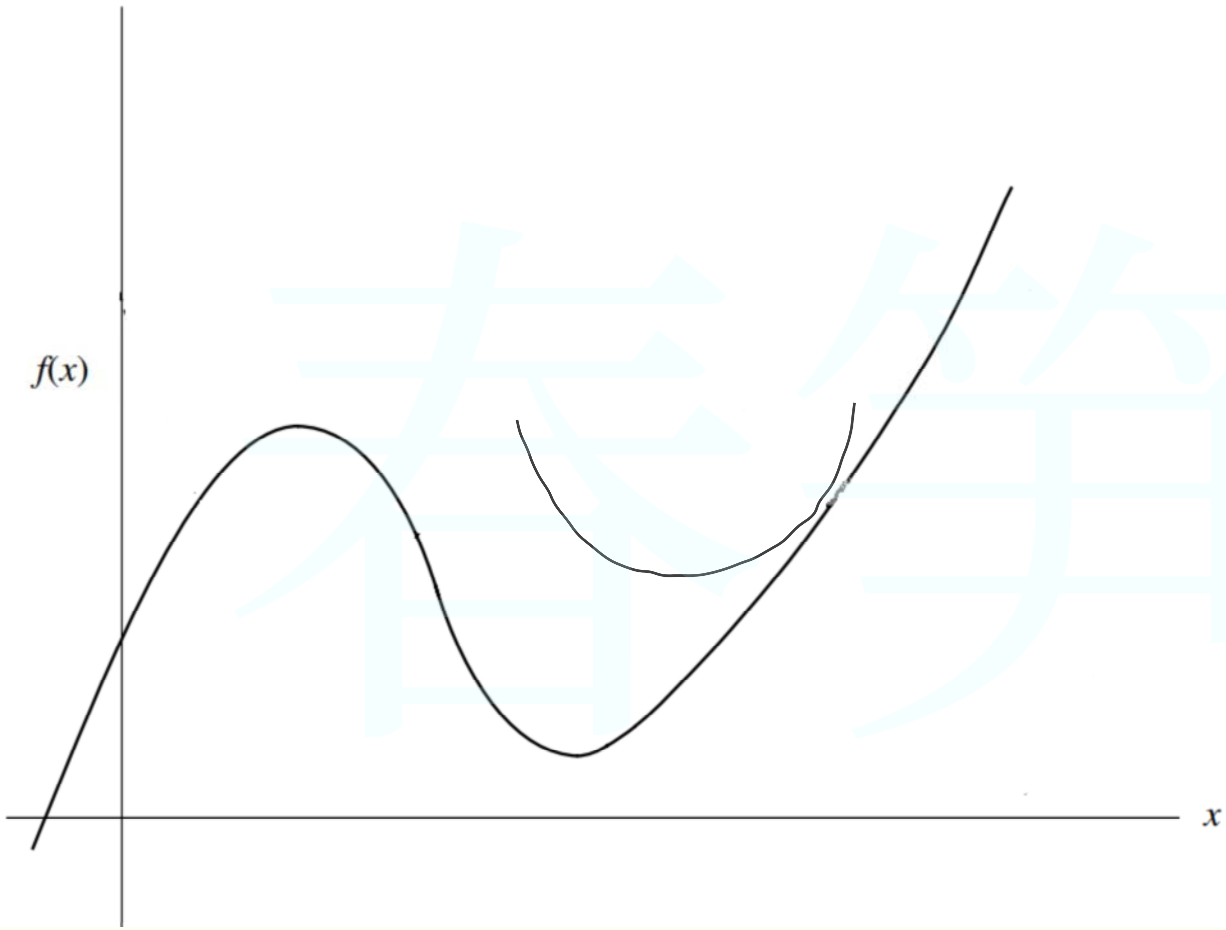
Step 5

If $\|\alpha_k\mathbf{d}_k\| < \varepsilon$, then do:

  Output $\mathbf{x}^* = \mathbf{x}_{k+1}$ and $f(\mathbf{x}^*) = f_{k+1}$, and stop.

Otherwise, set $k = k + 1$ and repeat from Step 2.

Newton Method



Newton Method

$$f(\mathbf{x}+\boldsymbol{\delta}) \approx f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^T\mathbf{H}(\mathbf{x})\boldsymbol{\delta}$$

$$\frac{\partial f}{\partial \boldsymbol{\delta}} = 0$$

$$\boldsymbol{\delta} = -\mathbf{H}^{-1}\mathbf{g}$$

Newton direction

$$\Delta F = \boldsymbol{g}^T\boldsymbol{\delta} = -\boldsymbol{g}^T\boldsymbol{H}\boldsymbol{g} \le 0 \quad \text{iif } \boldsymbol{H} \succeq \boldsymbol{0}$$

Algorithm 5.3 Basic Newton algorithm

Step 1

Input $\mathbf{x}_0$ and initialize the tolerance ε.

Set $k = 0$.

Step 2

Compute $\mathbf{g}_k$ and $\mathbf{H}_k$.

If $\mathbf{H}_k$ is not positive definite, force it to become positive definite.

Step 3

Compute $\mathbf{H}_k^{-1}$ and $\mathbf{d}_k = -\mathbf{H}_k^{-1}\mathbf{g}_k$.

Step 4

Find α_k, the value of α that minimizes $f(\mathbf{x}_k + \alpha\mathbf{d}_k)$, using a line search.

Step 5

Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$.

Compute $f_{k+1} = f(\mathbf{x}_{k+1})$.

Step 6

If $\|\alpha_k\mathbf{d}_k\| < \varepsilon$, then do:

Output $\mathbf{x}^* = \mathbf{x}_{k+1}$ and $f(\mathbf{x}^*) = f_{k+1}$, and stop.

Otherwise, set $k = k + 1$ and repeat from Step 2.

Modification of the Hessian

$$\hat{\mathbf{H}}_k = \frac{\mathbf{H}_k + \beta \mathbf{I}_n}{1 + \beta}$$

where β is set to a large value if $\mathbf{H}_k$ is nonpositive definite, or to a small value if $\mathbf{H}_k$ is positive definite.

If β is large, then

$$\hat{\mathbf{H}}_k \approx \mathbf{I}_n$$

and from Eq. (5.12)

$$\mathbf{d}_k \approx -\mathbf{g}_k$$

Gauss-Newton Method

$$\mathbf{f} = [f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ \cdots \ f_m(\mathbf{x})]^T$$

where $f_p(\mathbf{x})$ for $p = 1, 2, \ldots, m$ are independent functions of $\mathbf{x}$ (see Sec. 1.2). The solution sought is a point $\mathbf{x}$ such that all $f_p(\mathbf{x})$ are reduced to zero simultaneously.

$$F = \sum_{p=1}^{m} f_p(\mathbf{x})^2 = \mathbf{f}^T\mathbf{f}$$

Jacobian

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial F}{\partial x_i} = \sum_{p=1}^{m} 2 f_p(\mathbf{x}) \frac{\partial f_p}{\partial x_i}$$

$$\downarrow$$

$$\begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix} = 2 \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$$

$$\downarrow$$

$$\mathbf{g}_F = 2\mathbf{J}^T\mathbf{f}$$

Assuming that $f_p(\mathbf{x}) \in C^2$, Eq. (5.23) yields

$$\frac{\partial^2 F}{\partial x_i \partial x_j} = 2\sum_{p=1}^{m} \frac{\partial f_p}{\partial x_i}\frac{\partial f_p}{\partial x_j} + 2\sum_{p=1}^{m} f_p(\mathbf{x})\frac{\partial^2 f_p}{\partial x_i \partial x_j}$$

for $i,\ j = 1,\ 2,\ \ldots,\ n$. If the second derivatives of $f_p(\mathbf{x})$ are neglected, we have

$$\frac{\partial^2 F}{\partial x_i \partial x_j} \approx 2\sum_{p=1}^{m} \frac{\partial f_p}{\partial x_i}\frac{\partial f_p}{\partial x_j}$$

Thus the Hessian of F, designated by $\mathbf{H}_F$, can be deduced as

$$\mathbf{H}_F \approx 2\mathbf{J}^T\mathbf{J}$$

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k - \alpha_k(2\mathbf{J}^T\mathbf{J})^{-1}(2\mathbf{J}^T\mathbf{f})\\ &= \mathbf{x}_k - \alpha_k(\mathbf{J}^T\mathbf{J})^{-1}(\mathbf{J}^T\mathbf{f})\end{aligned}$$

思考题：

5-1,5-5,5-7,5-9