4.QUASI-NEWTON METHODS

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{S}_k \mathbf{g}_k$$

$$\mathbf{S}_k = \begin{cases} \mathbf{I}_n & \text{for the steepest-descent method} \\ \\ \mathbf{H}_k^{-1} & \text{for the Newton method} \end{cases}$$

Quasi-Newton

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_{k+1}) + \boldsymbol{g}_{\mathrm{k+1}}^{T} * (\boldsymbol{x} - \boldsymbol{x}_{k+1}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_{k+1})^{T} \boldsymbol{G}_{k+1} (\boldsymbol{x} - \boldsymbol{x}_{k+1})$$

$\downarrow$ 求导

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{g}_{\mathrm{k+1}} + \boldsymbol{G}_{k+1}(\boldsymbol{x} - \boldsymbol{x}_{k+1})$$

$\downarrow$ $\boldsymbol{x} = \boldsymbol{x}_k$

$$\boldsymbol{g}_{\mathrm{k}} = \boldsymbol{g}_{\mathrm{k+1}} + \boldsymbol{G}_{k+1}(\boldsymbol{x}_k - \boldsymbol{x}_{k+1})$$

$\downarrow$ Let $\boldsymbol{\delta}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \boldsymbol{\gamma}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$

$$\boldsymbol{G}_{k+1}^{-1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k$$

$\downarrow$ Let $\boldsymbol{S}_{k+1} = \boldsymbol{G}_{k+1}^{-1}$

$$\boldsymbol{S}_{k+1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k$$

quasi-Newton equation $\boldsymbol{\gamma}_k^T \boldsymbol{S}_{k+1}\boldsymbol{\gamma}_k = \boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k \geq 0$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha \boldsymbol{G}_k^{-1}\boldsymbol{g}_k = \boldsymbol{x}_k - \alpha \boldsymbol{S}_k \boldsymbol{g}_k$$

Quasi-newton

$$\boldsymbol{\delta}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$$
$$\boldsymbol{\gamma}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$

Rank-one

$$\mathbf{S}_{k+1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k$$

假定

$$\mathbf{S}_{k+1} = \mathbf{S}_k + \beta_k \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T \longrightarrow \text{半正定}$$

代入上式

$$\boldsymbol{\delta}_k = \mathbf{S}_k\boldsymbol{\gamma}_k + \beta_k \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k$$

$$\downarrow$$

$$
\begin{aligned}
\boldsymbol{\gamma}_k^T(\boldsymbol{\delta}_k - \mathbf{S}_k\boldsymbol{\gamma}_k) &= \beta_k \boldsymbol{\gamma}_k^T \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k \\
&= \beta_k (\boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k)^2
\end{aligned}
$$

$$\downarrow$$

$$(\boldsymbol{\delta}_k - \mathbf{S}_k\boldsymbol{\gamma}_k) = \beta_k \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k = \beta_k (\boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k)\boldsymbol{\xi}_k$$

$$(\boldsymbol{\delta}_k - \mathbf{S}_k\boldsymbol{\gamma}_k)^T = \beta_k \boldsymbol{\gamma}_k^T \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T = \beta_k (\boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k)\boldsymbol{\xi}_k^T$$

$$\downarrow$$

since $\boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k$ is a scalar. Hence

$$(\boldsymbol{\delta}_k - \mathbf{S}_k\boldsymbol{\gamma}_k)(\boldsymbol{\delta}_k - \mathbf{S}_k\boldsymbol{\gamma}_k)^T = \beta_k (\boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k)^2 \beta_k \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T$$

$$\downarrow$$

$$\beta_k \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T = \frac{(\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)(\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)^T}{\beta_k (\boldsymbol{\xi}_k^T \boldsymbol{\gamma}_k)^2}$$

$$= \frac{(\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)(\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)^T}{\boldsymbol{\gamma}_k^T (\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)}$$

$$\downarrow$$

$$\mathbf{S}_{k+1} = \mathbf{S}_k + \frac{(\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)(\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)^T}{\boldsymbol{\gamma}_k^T (\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)}$$

结束

验证quasi-newton为下降方向

$$\begin{aligned}
\boldsymbol{\gamma}_i^T \mathbf{S}_{k+1} \boldsymbol{\gamma}_i &= \boldsymbol{\gamma}_i^T \mathbf{S}_k \boldsymbol{\gamma}_i + \frac{\boldsymbol{\gamma}_i^T (\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)(\boldsymbol{\delta}_k^T - \boldsymbol{\gamma}_k^T \mathbf{S}_k)\boldsymbol{\gamma}_i}{\boldsymbol{\gamma}_k^T (\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)} \\
&= \boldsymbol{\gamma}_i^T \mathbf{S}_k \boldsymbol{\gamma}_i + \frac{(\boldsymbol{\gamma}_i^T \boldsymbol{\delta}_k - \boldsymbol{\gamma}_i^T \mathbf{S}_k \boldsymbol{\gamma}_k)(\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_k^T \mathbf{S}_k \boldsymbol{\gamma}_i)}{\boldsymbol{\gamma}_k^T (\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)} \\
&= \boldsymbol{\gamma}_i^T \mathbf{S}_k \boldsymbol{\gamma}_i + \frac{(\boldsymbol{\gamma}_i^T \boldsymbol{\delta}_k - \boldsymbol{\gamma}_i^T \mathbf{S}_k \boldsymbol{\gamma}_k)^2}{\boldsymbol{\gamma}_k^T (\boldsymbol{\delta}_k - \mathbf{S}_k \boldsymbol{\gamma}_k)}
\end{aligned}$$

$\downarrow$

$\mathbf{S}_{k+1}$ 正定

Algorithm 7.2 Basic quasi-Newton algorithm

Step 1

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

Set $k = 0$ and $\mathbf{S}_0 = \mathbf{I}_n$.

Compute $\mathbf{g}_0$.

Step 2

Set $\mathbf{d}_k = -\mathbf{S}_k \mathbf{g}_k$.

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

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

Step 3

If $\|\boldsymbol{\delta}_k\| < \varepsilon$, output $\mathbf{x}^* = \mathbf{x}_{k+1}$ and $f(\mathbf{x}^*) = f(\mathbf{x}_{k+1})$, and stop.

Step 4

Compute $\mathbf{g}_{k+1}$ and set

$$\boldsymbol{\gamma}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$

Compute $\mathbf{S}_{k+1}$ using Eq. (7.20).

Set $k = k + 1$ and repeat from Step 2.

Davidon-Fletcher-Powell (DFP)

$$\mathbf{S}_{k+1} = \mathbf{S}_k + \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} - \frac{\mathbf{S}_k \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T \mathbf{S}_k}{\boldsymbol{\gamma}_k^T \mathbf{S}_k \boldsymbol{\gamma}_k}$$

$$\downarrow$$

$$\mathbf{S}_{k+1}\boldsymbol{\gamma}_k = \mathbf{S}_k\boldsymbol{\gamma}_k + \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} - \frac{\mathbf{S}_k \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T \mathbf{S}_k \boldsymbol{\gamma}_k}{\boldsymbol{\gamma}_k^T \mathbf{S}_k \boldsymbol{\gamma}_k}$$

$$\downarrow$$

$$\mathbf{S}_{k+1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k$$

Broyden-Fletcher-Goldfarb-Shanno Method----BFGS

$$\mathbf{S}_{k+1} = \mathbf{S}_k + \left(1 + \frac{\boldsymbol{\gamma}_k^T \mathbf{S}_k \boldsymbol{\gamma}_k}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k}\right) \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k} - \frac{(\boldsymbol{\delta}_k \boldsymbol{\gamma}_k^T \mathbf{S}_k + \mathbf{S}_k \boldsymbol{\gamma}_k \boldsymbol{\delta}_k^T)}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k}$$

The Broyden Family

$$\mathbf{S}_{k+1} = (1 - \phi_k)\mathbf{S}_{k+1}^{DFP} + \phi_k \mathbf{S}_{k+1}^{BFGS}$$

Comparison of quasi-Newton method vs Newton's method

quasi-Newton method	Newton's method
Only need the function values and gradients	Need the function values, gradients and Hessians
$\{H_k\}$ maintains positive definite for several updates	$\{G_k\}$ is not sure to be positive definite
Need $O(n^2)$ multiplications in each iteration	Need $O(n^3)$ multiplications in each iteration