Theory of Advanced

Numerical Optimizations

Duy-Ky Nguyen (1996)

Table of Contents

5. Papers

Fletcher R, 1970, A New Approach to Vraible Metric Algorithm. *Computer Journal,* **13**, 317-322.

Goldfarb D, 1970, A Family of Variable Metric Methods derived by Variational Means. *Mathematics Computation*, **24**, 23-26.

Greenstadt J L, 1970, Variations of Variable Metric Methods. *Mathematics Computation*, **24**, 1-22..

Shanno D F, 1970, Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics Computation*, **24**, 647-656.

Theory of Some Advanced Numerical Methods

Duy-Ky Nguyen **08-05-96**

1. Extrema of Functions with Equality Constraints: Langrange Multiplier

Theorem 1.1:

A necessary condition for

$$\begin{cases} \underset{\mathbf{x}}{\text{Extremize}} \quad f(\mathbf{x}) \\ \text{subject to} \quad \mathbf{g}(\mathbf{x}) = \mathbf{0} \end{cases} \tag{1.1}$$

is

$$\begin{cases} \dfrac{\partial L}{\partial \mathbf{x}} = \mathbf{0} \\ \dfrac{\partial L}{\partial \lambda} = \mathbf{0} \end{cases} \tag{1.2}$$

where

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda' . \mathbf{g}(\mathbf{x}) \tag{1.3}$$

Proof:

Since $f(\mathbf{x})$ has an extremum, the total differential of *f* must be equal to zero, *ie*.

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} . dx_i = \frac{\partial f}{\partial \mathbf{x}}' . \mathbf{x} = 0 \tag{1.4}$$

As $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, we have

$$d\mathbf{g} = \sum_{i=1}^{n} \frac{\partial \mathbf{g}}{\partial x_i} . dx_i = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}' . \mathbf{x} = \mathbf{0} \tag{1.5}$$

By Eq.(1.3), the total differential of L is

$$dL(\mathbf{x}, \lambda) = df(\mathbf{x}) + \lambda' . d\mathbf{g}(\mathbf{x}) \tag{1.6}$$

or

$$dL(\mathbf{x}, \lambda) = \frac{\partial L}{\partial \mathbf{x}}' . d\mathbf{x} + \frac{\partial L}{\partial \lambda}' . d\lambda \tag{1.7}$$

From Eqs.(1.4) to (1.6), we get

$$dL = 0 \tag{1.8}$$

This equality must be hold for any **x** and **λ**, so Eq.(1.2) is achieved.

Q.E.D.

Remark 1.1:

Note that the constraint is

$$\mathbf{g}(\mathbf{x}) = \mathbf{0}$$

if

$$\mathbf{g}(\mathbf{x}) = \mathbf{c}$$

then let

$$\tilde{\mathbf{g}}(\mathbf{x}) = \mathbf{g}(\mathbf{x}) - \mathbf{c}$$

or

$$\tilde{\mathbf{g}}(\mathbf{x}) = \mathbf{c} - \mathbf{g}(\mathbf{x})$$

and we have the required constraint, that is $\tilde{\mathbf{g}}(\mathbf{x}) = \mathbf{0}$.

2. Steepest Descent Method

For a function $f(\mathbf{x}): \Re^n \to \Re$, and a given point $\mathbf{x}_k$, our interest is to find the next point $\mathbf{x}_{k+1}$ to minimize the function *f*. Let

$$\delta = \mathbf{x}_{k+1} - \mathbf{x}_k \tag{2.1}$$

then the Taylor expansion of *f* at $\mathbf{x}_k$ is

$$f(\mathbf{x}_k + \delta) = f_k + \nabla f_k^{/}.\delta + \tfrac{1}{2}\delta^{/}.\nabla^2 f_k.\delta + \cdots = f_k + \mathbf{g}_k^{/}\delta + \tfrac{1}{2}\delta^{/}.\mathbf{Q}_k.\delta + \cdots \tag{2.2}$$

and our concern is to find δ to minimize *f*.

Let α and $\mathbf{u}$ be the magnitude and the unit vector of δ respectively, then Eq.(2.2) can be read as

$$f(\mathbf{x}_k + \delta) = f_k + \alpha.\mathbf{g}_k^{/}\mathbf{u} + \tfrac{1}{2}\alpha^2\mathbf{u}^{/}.\mathbf{Q}_k.\mathbf{u} + \cdots \tag{2.3}$$

or

$$f(\mathbf{x}_k + \alpha.\mathbf{u}) = f_k + \alpha.f_k^{/} + \tfrac{1}{2}\alpha^2.f_k^{//} + \cdots \tag{2.4}$$

so

$$f_k^{/} = \mathbf{g}_k^{/}.\mathbf{u}, \qquad f_k^{//} = \mathbf{u}^{/}.\mathbf{Q}_k.\mathbf{u} \tag{2.5}$$

The use of the negative of the gradient as a direction for minimization was first made by Cauchy in 1847. It is known as the steepest descent method since *for the same magnitude of* δ, this direction produces the largest decrease at that point. We have the following theorem for this method

Theorem 2.1:

The gradient vector represents the direction of the steepest ascent, and its opposite direction is the steepest descent.

Proof:

For a function $f:\Re^n \to \Re$, we have the differentiation

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i = \nabla^{/} f . d\mathbf{x} \tag{2.6}$$

If $\mathbf{u}$ denotes the unit vector in the direction of $d\mathbf{x}$, we have

$$d\mathbf{x} = ds.\mathbf{u} \tag{2.7}$$

where

$$ds = \|d\mathbf{x}\|_2 \quad \Leftrightarrow \quad ds^2 = \sum_{i=1}^{n} dx_i^2 \tag{2.8}$$

and

$$\|\mathbf{u}\|_2 = 1 \quad \Leftrightarrow \quad \mathbf{u}^{/}\mathbf{u} = 1 \tag{2.9}$$

Then Eq.(2.6) can be read as

$$\frac{df}{ds} = \nabla^{/} f . \mathbf{u}$$

or

$$\phi(\mathbf{u}) = \mathbf{g}^{/}\mathbf{u} \tag{2.10}$$

where

$$\phi(\mathbf{u}) = \frac{df}{ds}, \qquad \mathbf{g} = \nabla f \tag{2.11}$$

To find a direction $\mathbf{u}$ to extremize Eq.(2.10), the Lagrange multiplier method is used with

$$L(\mathbf{u}, \lambda) = \phi(\mathbf{u}) + \lambda\left(1 - \mathbf{u}^{/}\mathbf{u}\right) \tag{2.12}$$

where λ is the Lagrange multiplier to be determined. We have the necessary condition as

$$\frac{\partial L}{\partial \mathbf{u}} = \mathbf{g} - 2\lambda\mathbf{u} = \mathbf{0} \quad \Rightarrow \quad 2\lambda\mathbf{u} = \mathbf{g} \quad \Rightarrow \quad \begin{cases} |\lambda| = \dfrac{\|\mathbf{g}\|}{2\|\mathbf{u}\|} \\ \mathbf{u} = \dfrac{\mathbf{g}}{2\lambda} \end{cases} \tag{2.13}$$

$$\frac{\partial L}{\partial \lambda} = 1 - \mathbf{u}^{/}\mathbf{u} = 0 \quad \Rightarrow \quad \mathbf{u}^{/}\mathbf{u} = 1 \quad \Rightarrow \quad \|\mathbf{u}\| = 1 \tag{2.14}$$

so Eq.(2.13) becomes

$$|\lambda| = \tfrac{1}{2}\|\mathbf{g}\| \tag{2.15}$$

(1) If $\lambda > 0$, then Eq.(2.15) becomes

$$\lambda = \tfrac{1}{2}\|\mathbf{g}\| \quad \Rightarrow \quad \mathbf{u} = \frac{\mathbf{g}}{\|\mathbf{g}\|} \tag{2.16}$$

From Eq.(2.10), we have

$$\phi = \|\mathbf{g}\| > 0 \quad \Rightarrow \quad \frac{df}{ds} > 0 \tag{2.17}$$

so it is the steepest ascent direction.

(2) If $\lambda < 0$, then Eq.(2.15) becomes

$$\lambda = -\tfrac{1}{2}\|\mathbf{g}\| \quad \Rightarrow \quad \mathbf{u} = -\frac{\mathbf{g}}{\|\mathbf{g}\|} \tag{2.18}$$

From Eq.(2.10), we have

$$\phi = -\|\mathbf{g}\| < 0 \quad \Rightarrow \quad \frac{df}{ds} < 0 \tag{2.19}$$

so it is the steepest descent direction.

Q.E.D.

3. Newton Method

Theorem 3.1: Newton Method

Consider the second-order approximation to *f* based on the Taylor expansion

$$f(\mathbf{x}_k + \delta_k) \approx f_k + \mathbf{g}_k^{/}\delta_k + \tfrac{1}{2}\delta_k^{/}.\mathbf{Q}_k.\delta_k \tag{3.1}$$

if $\mathbf{Q}_k$ is positive-definite, then

$$\underset{\delta_k}{\text{Minimize}}\ f(\mathbf{x}_k + \boldsymbol{\delta}_k) \quad \Rightarrow \quad \delta_k = -\mathbf{Q}_k^{-1}\mathbf{g}_k \tag{3.2}$$

this is the *exact line-search condition*.

Proof:

A necessary condition for an extremum is

$$\frac{\partial f}{\partial \delta_k} = \mathbf{0} \quad \Rightarrow \quad \mathbf{g}_k + \mathbf{Q}_k\delta_k = \mathbf{0} \quad \Rightarrow \quad \delta_k = -\mathbf{Q}_k^{-1}\mathbf{g}_k$$

since $\mathbf{Q}_k$ is positive-definite and by Eq.(2.5), this extremum is a minimum.

Q.E.D.

Remark 3.1: Newton method for univariable function

The minimum is a solution of the equation

$$\mathbf{g}(\mathbf{x}) = \mathbf{0}$$

and the Newton method gives

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{\delta}_k = \mathbf{x}_k - \mathbf{Q}_k^{-1}\mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - \nabla^{-1}\mathbf{g}(\mathbf{x}_k).\mathbf{g}(\mathbf{x}_k)$$

for a univariable function, we have

$$x_{k+1} = x_k - \frac{g(x_k)}{g^{/}(x_k)}$$

as we have known.

Theorem 3.2: Convergence of Newton Method

If $\boldsymbol{\delta}_k$ forms a basis which spans the vector space of **x**, then the Newton method terminates after *n* iterations .

Proof:

Let $\mathbf{x}^*$ and $\mathbf{x}_0$ be the minimum and a given starting point. Since $\boldsymbol{\delta}_0, \boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_{n-1}$ are *n* independent vectors to form a basis, we have

$$\mathbf{x}^* - \mathbf{x}_0 = \sum_{i=0}^{n-1} \alpha_i \boldsymbol{\delta}_i \quad \Rightarrow \quad \mathbf{x}^* = \mathbf{x}_0 + \sum_{i=0}^{n-1} \alpha_i \boldsymbol{\delta}_i \tag{3.3}$$

If we define

$$\mathbf{x}_k = \mathbf{x}_0 + \sum_{j=0}^{k-1} \alpha_j \boldsymbol{\delta}_j \tag{3.4}$$

then we obtain

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \boldsymbol{\delta}_k \tag{3.5}$$

thus

$$\mathbf{x}_n = \mathbf{x}^* \tag{3.6}$$

Q.E.D.

Remark 3.2:

Newton direction $\boldsymbol{\delta}_k = -\mathbf{Q}_k^{-1}\mathbf{g}_k$ is a descent direction since

$$\frac{df\left(\mathbf{x}_k + \alpha_k \boldsymbol{\delta}_k\right)}{d\alpha_k} = \nabla^{/} f_k . \boldsymbol{\delta}_k = \mathbf{g}_k^{/} \boldsymbol{\delta}_k = -\mathbf{g}_k^{/} \mathbf{Q}_k^{-1} \mathbf{g}_k < 0$$

as $\mathbf{Q}_k$ is positive-definite and so is $\mathbf{Q}_k^{-1}$.

4. Quasi-Newton Methods

The main disadvantage of Newton method is that the Hessian (second derivative matrix) must be supplied. However methods closely related to Newton method can be derived when only the gradient (first derivative vector) is available. The most obvious is a *finite difference Newton method* in which increments h_i in each coordinate direction $\mathbf{e}_i$ are taken sa as to estimate $\mathbf{Q}_k$ by differences in gradient. That is to say, the matrix $\tilde{\mathbf{Q}}$ whose *i*-th column is

$$\tilde{\mathbf{Q}}_{*i} = \frac{\mathbf{g}\left(\mathbf{x}_k + h_i \mathbf{e}_i\right) - \mathbf{g}\left(\mathbf{x}_k\right)}{h_i}$$

Then $\tilde{\mathbf{Q}}$ is made symmetric by taking $\frac{1}{2}\left(\tilde{\mathbf{Q}} + \tilde{\mathbf{Q}}^T\right)$ and the resulting matrix is used to replace $\mathbf{Q}_k$ in Newton method. Disadvantage of this method is the resulting matrix may not be positive-definite (requiring modification technique), *n* gradient evaluations are required to compute $\tilde{\mathbf{Q}}$. The conjugate-gradient method (Hestenes and Stiefel 1952, Fletcher and Reeves 1964) is in this direction where the computation of **Q** is avoided by using a line-search, but the positive-definiteness of **Q** is not guaranteed, so the necessary condition for the existence of a minimum is not satisfied.

The above disadvantages are all avoided in the much more important class of *quasi-Newton methods* where an estimate of the Hessian $\mathbf{Q}_k$ is maintained to be symmetric and positive-definite. The Hessian is estimated using the first-order approximation of the gradient

$$\mathbf{g}(\mathbf{x}_{k+1}) \approx \mathbf{g}(\mathbf{x}_k) + \mathbf{Q}_k \boldsymbol{\delta}_k \tag{4.1}$$

To ensure the minimum due to truncation of the Taylor series in the function and its gradient, the quasi-Newton method minimizes the function in the direction $\mathbf{s}_k$ by introducing a step length α and a search direction $\mathbf{s}_k$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\, \mathbf{s}_k$$

where α is determined as

$$\underset{\alpha}{\text{Minimize}}\ f(\mathbf{x}_k + \alpha \mathbf{s}_k) \tag{4.2}$$

Remark 4.1:

Therefore, *in the rest of this note*, we use

$$\boldsymbol{\delta}_k = \alpha_k \mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k, \qquad \boldsymbol{\gamma}_k = \mathbf{g}_{k+1} - \mathbf{g}_k \tag{4.3}$$

Quasi-Newton methods uses directly the inverse of $\mathbf{Q}_k$ to simplify the process, so the exact line-search condition Eq.(3.2) becomes

$$\mathbf{s}_k = -\mathbf{H}_k \mathbf{g}_k, \qquad \mathbf{H}_k = \mathbf{Q}_k^{-1} \tag{4.4}$$

From Eq.(4.1), the *quasi-Newton condition* below must be satisfied

$$\boldsymbol{\delta}_k = \mathbf{H}_{k+1} \boldsymbol{\gamma}_k \tag{4.5}$$

Since Eq.(4.5) can be determined after the line search, $\mathbf{H}_{k+1}$ is used instead of $\mathbf{H}_k$ in order to estimate $\mathbf{Q}_k^{-1}$.

Thus a quasi-Newton algorithm is as below

(0) $\mathbf{H}_0 = \mathbf{I}$, $\mathbf{x}_0$: *given*

(1) $\mathbf{s}_k = -\mathbf{H}_k \mathbf{g}_k$

(2) $\underset{\alpha}{\text{Minimize}}\ f(\mathbf{x}_k + \alpha \mathbf{s}_k)$, line search

(3) $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{s}_k = \mathbf{x}_k + \boldsymbol{\delta}_k$

(4) $\mathbf{H}_{k+1} = \mathbf{H}_k + \mathbf{E}_k$

where $\mathbf{E}_k$ to be determined such that

- $\mathbf{H}_{k+1}$ is symmetric and positive-definite and converges to $\mathbf{Q}_k^{-1}$ (the Newton method);
- $\boldsymbol{\delta}_k$'s are linear independent (convergence of Newton method)

We will find the smallest correction $\mathbf{E}_k$ in the sense of some norm. To a certain extent, this would tend to keep the elements of $\mathbf{H}$ from growing too large, which might cause an undesirable instability. The simplest type of norm is

$$N_{\mathbf{E}}(\mathbf{E}) = \sum_{i,j} E_{ij}^2 = trace(\mathbf{E}\mathbf{E}^T) \tag{4.6}$$

If we include one degree of freedom for flexibily, **E** is transformed as

$$N_{\mathbf{W}}(\mathbf{E}) = trace\left(\mathbf{WEWE}^{/}\right) \tag{4.7}$$

where

$$\mathbf{W} = \mathbf{W}^{/} \tag{4.8}$$

then we have the following theorem

Theorem 4.1: BFGS Formula (Broyden, Fletcher, Greenstadt/Goldfarb, Shanno)

If

$$\mathbf{H}_{k+1} = \mathbf{H} + \mathbf{E} \tag{4.9}$$

then a solution of the problem

$$\underset{\mathbf{E}}{\text{Minimize}} \|\mathbf{E}\|_{\mathbf{W}} \tag{4.10}$$

subject to the conditions

$$\mathbf{E}^{T} = \mathbf{E} \quad and \quad \boldsymbol{\eta} = \mathbf{E}\gamma \tag{4.11}$$

is

$$\mathbf{H}_{k+1} = \mathbf{H} + \left(1 + \frac{\gamma^{/}\mathbf{H}\gamma}{\gamma^{/}\delta}\right) . \frac{\delta\delta^{/}}{\gamma^{/}\delta} - \left(\frac{\delta\gamma^{/}\mathbf{H} + \mathbf{H}\gamma\delta^{/}}{\gamma^{/}\delta}\right) \tag{4.12.a}$$

or

$$\mathbf{H}_{k+1} = \mathbf{H} + \left(1 + \frac{\gamma^{/}\mathbf{H}\gamma}{\gamma^{/}\mathbf{s}}\right) . \frac{\mathbf{ss}^{/}}{\gamma^{/}\mathbf{s}} - \left(\frac{\mathbf{s}\gamma^{/}\mathbf{H} + \mathbf{H}\gamma\mathbf{s}^{/}}{\gamma^{/}\mathbf{s}}\right) \tag{4.12.b}$$

where all the subscript k in LHS are omitted, and

$$\boldsymbol{\eta} = \delta - \mathbf{H}\gamma, \quad \gamma = \mathbf{g}_{k+1} - \mathbf{g}_{k}, \quad \delta = \alpha\mathbf{s} \tag{4.13}$$

and **W** is chosen such that

$$\mathbf{W}^{/} = \mathbf{W}, \quad \gamma = \mathbf{W}\delta \tag{4.14}$$

Note that conditions in Eq.(4.11) correspond to the symmetry and the quasi-Newton conditions.

Proof:

From Eq.(4.9), we have from the quasi-Newton condition Eq.(4.5)

$$\mathbf{H}_{k+1}\gamma = \mathbf{H}\gamma + \mathbf{E}\gamma \quad \Rightarrow \quad \mathbf{E}\gamma = \mathbf{H}_{k+1}\gamma - \mathbf{H}\gamma = \delta - \mathbf{H}\gamma = \boldsymbol{\eta}$$

and Eq.(4.14) is equivalent to

$$\mathbf{W} = \mathbf{H}_{k+1}^{-1} \tag{4.15}$$

since by the quasi-Newton condition, we have

$$\mathbf{H}_{k+1}\gamma = \delta \quad \Rightarrow \quad \gamma = \mathbf{H}_{k+1}^{-1}\delta = \mathbf{W}\delta$$

After squaring the norm, a suitable Lagrangian function is

$$\mathscr{L} = \tfrac{1}{4} trace\left(\mathbf{WE}^{/}\mathbf{WE}\right) + trace\left[\Lambda^{/}\left(\mathbf{E}^{/} - \mathbf{E}\right)\right] - \lambda^{/}\mathbf{W}(\mathbf{E}\gamma - \boldsymbol{\eta}) = \tfrac{1}{4} trace\left(\mathbf{WE}^{/}\mathbf{WE}\right) + trace\left[\Lambda^{/}\left(\mathbf{E}^{/} - \mathbf{E}\right)\right] - trace\left[\lambda^{/}\mathbf{W}(\mathbf{E}\gamma - \boldsymbol{\eta})\right] \tag{4.16}$$

where Λ is a Lagrange multiplier matrix for the contraint $\mathbf{E}^{/} = \mathbf{E}$ (use of the *trace* is just a convenient way of summing $\Lambda_{ij}\left(E_{ij}^{/} - E_{ij}\right)$ over al *i*, *j*) and $\lambda^{/}\mathbf{W}$ is a vector Lagrange multipliers for the constraint $\boldsymbol{\eta} = \mathbf{E}\gamma$. $\mathscr{L}$ must be stationary with respect to $\mathbf{E}$, Λ and λ . Setting derivatives of with respect to Λ and λ to zero just gives the constraints in Eq.(4.11).

For the derivative with respect to **E**, using Frechet derivative operator with the last term is $d\mathbf{E}'$ and the property

$$trace(\mathbf{AB}) = trace(\mathbf{BA}) = trace(\mathbf{A}'\mathbf{B}') = trace(\mathbf{B}'\mathbf{A}')$$

we have

$$df_1(\mathbf{E}, d\mathbf{E}) = trace\left[\mathbf{W}d\mathbf{E}'\mathbf{WE}\right] + trace\left[\mathbf{WE}'\mathbf{W}d\mathbf{E}\right] = trace\left[\mathbf{WEW}d\mathbf{E}'\right] + trace\left[\mathbf{W}'\mathbf{EW}'d\mathbf{E}'\right] = trace\left[2\mathbf{WEW}d\mathbf{E}'\right]$$

$$df_2(\mathbf{E}, d\mathbf{E}) = trace\left[\Lambda' d\mathbf{E}' - \Lambda' d\mathbf{E}\right] = trace\left[\Lambda' d\mathbf{E}' - \Lambda d\mathbf{E}'\right]$$

$$df_3(\mathbf{E}, d\mathbf{E}) = trace\left[\lambda'\mathbf{W}d\mathbf{E}\gamma\right] = trace\left[\mathbf{W}\lambda\gamma' d\mathbf{E}'\right]$$

so

$$\frac{\partial \mathscr{L}}{\partial \mathbf{E}} = \mathbf{0} \quad \Rightarrow \quad \begin{cases} \frac{1}{2}\mathbf{WEW} + \Lambda' - \Lambda - \mathbf{W}\lambda\gamma' = \mathbf{0} \\ \frac{1}{2}\mathbf{WE}^T\mathbf{W} + \Lambda - \Lambda^T - \gamma\lambda'\mathbf{W} = \mathbf{0} \end{cases}$$

thus, by Eq.(4.14) we can solve for **E**

$$\mathbf{WEW} = \mathbf{W}\lambda\gamma' + \gamma\lambda'\mathbf{W} = \mathbf{W}\delta\lambda'\mathbf{W} + \mathbf{W}\lambda\delta'\mathbf{W} \quad \Rightarrow \quad \mathbf{E} = \lambda\delta' + \delta\lambda' \tag{4.17}$$

substituting into $\eta = \mathbf{E}\gamma$, we solve for λ

$$\eta = \left(\lambda\delta' + \delta\lambda'\right)\gamma \quad \Rightarrow \quad \lambda = \frac{\eta - \delta\left(\lambda'\gamma\right)}{\delta'\gamma} = \frac{\eta - \left(\lambda'\gamma\right)\delta}{\delta'\gamma} \tag{4.18}$$

In the sequel, note that

$$\gamma'\delta = \delta'\gamma \tag{4.19}$$

Pre-multiplying Eq.(4.18) by γ' gives

$$\gamma'\lambda = \frac{\gamma'\eta - \left(\lambda'\gamma\right)\gamma'\delta}{\delta'\gamma} \quad \Rightarrow \quad \left(\gamma'\lambda\right)\delta'\gamma = \gamma'\eta - \left(\lambda'\gamma\right)\gamma'\delta \quad \Rightarrow \quad \lambda'\gamma = \frac{\gamma'\eta}{2\delta'\gamma}$$

so Eq.(4.18) becomes

$$\lambda = \frac{\eta - \delta\dfrac{\gamma'\eta}{2\delta'\gamma}}{\delta'\gamma} = \frac{\delta - \mathbf{H}\gamma - \delta\dfrac{\gamma'\left(\delta - \mathbf{H}\gamma\right)}{2\delta'\gamma}}{\delta'\gamma} = \frac{\delta - \mathbf{H}\gamma - \left(1 - \dfrac{\gamma'\mathbf{H}\gamma}{\delta'\gamma}\right)\dfrac{\delta}{2}}{\delta'\gamma} = \frac{-\mathbf{H}\gamma + \left(1 + \dfrac{\gamma'\mathbf{H}\gamma}{\delta'\gamma}\right)\dfrac{\delta}{2}}{\delta'\gamma} \tag{4.20}$$

Substituting into Eq.(4.17) gives

$$\mathbf{E} = \delta\frac{-\gamma'\mathbf{H} + \left(1 + \dfrac{\gamma'\mathbf{H}\gamma}{\delta'\gamma}\right)\dfrac{\delta'}{2}}{\delta'\gamma} + \frac{-\mathbf{H}\gamma + \left(1 + \dfrac{\gamma'\mathbf{H}\gamma}{\delta'\gamma}\right)\dfrac{\delta}{2}}{\delta'\gamma}\delta' = \frac{-\delta\gamma'\mathbf{H} - \mathbf{H}\gamma\delta' + \left(1 + \dfrac{\gamma'\mathbf{H}\gamma}{\delta'\gamma}\right)\delta\delta'}{\delta'\gamma} = \left(1 + \frac{\gamma'\mathbf{H}\gamma}{\delta'\gamma}\right)\frac{\delta\delta'}{\delta'\gamma} - \frac{\delta\gamma'\mathbf{H} + \mathbf{H}\gamma\delta'}{\delta'\gamma}$$

Q.E.D.

Theorem 4.2: Dual BFGS Formula

If

$$\mathbf{B} = \mathbf{H}^{-1} \tag{4.21}$$

then the BFGS formula has the dual

$$\mathbf{B}_{k+1} = \mathbf{B} + \frac{\gamma\gamma'}{\gamma'\delta} - \frac{\mathbf{B}\delta\delta'\mathbf{B}}{\delta'\mathbf{B}\delta} \tag{4.22}$$

Proof:

Multiply Eqs.(4.12) and (4.22), using Eq.(4.21), we have

$$\mathbf{B}_{k+1}\mathbf{H}_{k+1} = \mathbf{I}$$

Q.E.D.

Theorem 4.3: Positive-Definiteness of BFGS Formula

If $\gamma_k^{/}\delta_k > 0, \forall k$, *ie* $\gamma^{/}\delta > 0$, the BFGS formula preserves positive definite matrices **H.**

Proof:

Since $\mathbf{H} = \mathbf{H}_k$ is positive-definite, so is $\mathbf{B} = \mathbf{B}_k$, it can be decomposed using Choleski factors as $\mathbf{B} = \mathbf{L}^T\mathbf{L}$, then

$$\mathbf{z}^{/}\left(\mathbf{B} + \frac{\gamma\gamma^{/}}{\gamma^{/}\delta} - \frac{\mathbf{B}\delta\delta^{/}\mathbf{B}}{\delta^{/}\mathbf{B}\delta}\right)\mathbf{z} = \mathbf{z}^{/}\left(\frac{\gamma\gamma^{/}}{\gamma^{/}\delta}\right)\mathbf{z} + \mathbf{a}^{/}\mathbf{a} - \frac{\left(\mathbf{a}^{/}\mathbf{b}\right)^2}{\mathbf{b}^{/}\mathbf{b}}$$

where $\mathbf{a} = \mathbf{L}^{/}\mathbf{z}$, $\mathbf{b} = \mathbf{L}^{/}\delta$. The first term is positive by the assumption, the last 2 terms is positive by Cauchy inequality. Thus the proof is completed.

Q.E.D.

For a quadratic function, Theorem 3.2 guarantees the Newton method with the exact line-search will terminate (reach the minimum) after n iterations for n linear independent directions $\delta_0, \delta_1, \cdots, \delta_{n-1}$. The theorem below also guarantees the BFGS method will terminate after n iterations and **H** will converge to $\mathbf{Q}^{-1}$ on a quadratic function.

Theorem 4.4: Convergence of BFGS Method

The BFGS method with the exact line-search will terminate after n iterations on a quadratic function, and the following hold for all $i = 1, 2, \cdots, n$

Conjugacy $\quad \delta_i^{/}\mathbf{Q}\delta_j = 0, \quad j \le i-1 \quad (1, 2, \cdots, i-1)$, for all subset (4.23)

Quasi-Newton condition $\quad \delta_j = \mathbf{H}_{i+1}\gamma_j, \quad j \le i \quad (1, 2, \cdots, i)$, for all subset (4.24)

$\mathbf{Q}^{-1}$ Convergence $\quad \mathbf{H}_{n+1} = \mathbf{Q}^{-1}$ (4.25)

Proof:

Consider a quadratic function

$$f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^{/}\mathbf{Q}\mathbf{x} + \mathbf{b}^{/}\mathbf{x} + c \tag{4.26}$$

so

$$\nabla f(\mathbf{x}) = \mathbf{g}(\mathbf{x}) = \mathbf{Q}\mathbf{x} + \mathbf{b} \tag{4.27}$$

thus

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \delta_i \quad \Rightarrow \quad \mathbf{Q}\delta_i = \mathbf{g}_{i+1} - \mathbf{g}_i = \gamma_i$$

or

$$\gamma_i = \mathbf{g}_{i+1} - \mathbf{g}_i = \mathbf{Q}\delta_i \tag{4.28}$$

where α_i minimizes the function in the direction $\mathbf{s}_i$ (exact line search)

$$\underset{\alpha_i}{\text{Minimize}}\ f\left(\mathbf{x}_i+\alpha_i\mathbf{s}_i\right)\ \Rightarrow\ \frac{df\left(\mathbf{x}_i+\alpha_i\mathbf{s}_i\right)}{d\alpha_i}=\nabla f\left(\mathbf{x}_i+\alpha_i\mathbf{s}_i\right)^{/}.\mathbf{s}_i=\mathbf{g}_{i+1}^{/}\mathbf{s}_i=0$$

so *the exact line-search condition* at point $\mathbf{x}_i$ is

$$\mathbf{g}_{i+1}^{/}\boldsymbol{\delta}_i=0 \tag{4.29}$$

In the sequel, we employ the following property

$$\mathbf{u}^{/}\mathbf{v}=\mathbf{v}^{/}\mathbf{u},\qquad \forall\mathbf{u},\mathbf{v}\in\Re^{\mathbf{n}} \tag{4.30}$$

(1) By Eqs.(4.2) and (4.28), we have

$$\boldsymbol{\delta}_{i+1}^{/}\mathbf{Q}\boldsymbol{\delta}_i=\left(\alpha_{i+1}\mathbf{s}_{i+1}\right)^{/}\gamma_i=-\alpha_{i+1}\left(\mathbf{g}_{i+1}^{/}\mathbf{H}_{i+1}\right)\gamma_i \tag{4.31}$$

From the BFGS formula in Eq.(4.12), we have

$$\boldsymbol{\delta}_{i+1}^{/}\mathbf{Q}\boldsymbol{\delta}_i=-\alpha_{i+1}\mathbf{g}_{i+1}^{/}\left[\mathbf{H}_i+\left(1+\frac{\gamma_i^{/}\mathbf{H}_i\gamma_i}{\gamma_i^{/}\boldsymbol{\delta}_i}\right).\frac{\boldsymbol{\delta}_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}-\left(\frac{\boldsymbol{\delta}_i\gamma_i^{/}\mathbf{H}_i+\mathbf{H}_i\gamma_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}\right)\right]\gamma_i$$

$$\boldsymbol{\delta}_{i+1}^{/}\mathbf{Q}\boldsymbol{\delta}_i=-\alpha_{i+1}\mathbf{g}_{i+1}^{/}\left[\mathbf{H}_i\gamma_i+\left(1+\frac{\gamma_i^{/}\mathbf{H}_i\gamma_i}{\gamma_i^{/}\boldsymbol{\delta}_i}\right).\frac{\boldsymbol{\delta}_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}\gamma_i-\left(\frac{\boldsymbol{\delta}_i\gamma_i^{/}\mathbf{H}_i+\mathbf{H}_i\gamma_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}\right)\gamma_i\right]$$

$$\boldsymbol{\delta}_{i+1}^{/}\mathbf{Q}\boldsymbol{\delta}_i=-\alpha_{i+1}\mathbf{g}_{i+1}^{/}\left[\left(1+\frac{\gamma_i^{/}\mathbf{H}_i\gamma_i}{\gamma_i^{/}\boldsymbol{\delta}_i}\right).\boldsymbol{\delta}_i-\boldsymbol{\delta}_i\left(\frac{\gamma_i^{/}\mathbf{H}_i\gamma_i}{\gamma_i\boldsymbol{\delta}_i^{T}}\right)\right]=-\mathbf{g}_{i+1}^{/}\boldsymbol{\delta}_i \tag{4.32}$$

By Eq.(4.29), we have

$$\boldsymbol{\delta}_{i+1}^{/}\mathbf{Q}\boldsymbol{\delta}_i=-\mathbf{g}_{i+1}^{/}\boldsymbol{\delta}_i=0 \tag{4.33}$$

Next, we will prove $\boldsymbol{\delta}_{i+1}^{/}\mathbf{Q}\boldsymbol{\delta}_{i-1}=0$, by Eqs.(4.28), (4.29) and (4.33), we have

$$\mathbf{g}_{i+1}^{/}\boldsymbol{\delta}_{i-1}=\left(\mathbf{g}_i+\mathbf{Q}\boldsymbol{\delta}_i\right)^{/}\boldsymbol{\delta}_{i-1}=\mathbf{g}_i^{/}\boldsymbol{\delta}_{i-1}+\boldsymbol{\delta}_i^{/}\mathbf{Q}\boldsymbol{\delta}_{i-1}=0$$

so from Eq.(4.32), we get

$$\boldsymbol{\delta}_{i+1}^{/}\mathbf{Q}\boldsymbol{\delta}_{i-1}=-\alpha_{i+1}\mathbf{g}_i^{/}\boldsymbol{\delta}_{i-1}=0$$

By induction we obtain Eq.(4.23) and the exact line-search condition holds for all subspace of $\mathbf{x}_i$

$$\mathbf{g}_i^{/}\boldsymbol{\delta}_j=0,\quad j\le i-1 \tag{4.34}$$

(2) By the quasi-Newton condition Eq.(4.5), we have

$$\boldsymbol{\delta}_i=\mathbf{H}_{i+1}\gamma_i \tag{4.35}$$

For $j\le i-1$, we have

$$\mathbf{H}_{i+1}\gamma_{i-1}=\left[\mathbf{H}_i+\left(1+\frac{\gamma_i^{/}\mathbf{H}_i\gamma_i}{\gamma_i^{/}\boldsymbol{\delta}_i}\right).\frac{\boldsymbol{\delta}_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}-\left(\frac{\boldsymbol{\delta}_i\gamma_i^{/}\mathbf{H}_i+\mathbf{H}_i\gamma_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}\right)\right]\gamma_{i-1}=\boldsymbol{\delta}_{i-1}+\left(1+\frac{\gamma_i^{/}\mathbf{H}_i\gamma_i}{\gamma_i^{/}\boldsymbol{\delta}_i}\right).\frac{\boldsymbol{\delta}_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}\gamma_{i-1}-\left(\frac{\boldsymbol{\delta}_i\gamma_i^{/}\mathbf{H}_i+\mathbf{H}_i\gamma_i\boldsymbol{\delta}_i^{/}}{\gamma_i^{/}\boldsymbol{\delta}_i}\right)\gamma_{i-1}$$

By Eqs.(4.28) and (4.23), we have

$$\boldsymbol{\delta}_i^{/}\gamma_{i-1}=\boldsymbol{\delta}_{i-1}^{/}\mathbf{Q}\boldsymbol{\delta}_i=0$$

and

$$\gamma_{i-1}^{/}\mathbf{H}_i\gamma_i=\boldsymbol{\delta}_{i-1}^{/}\gamma_i=\boldsymbol{\delta}_{i-1}^{/}\mathbf{Q}\boldsymbol{\delta}_i=0$$

so

$$\boldsymbol{\delta}_{i-1}=\mathbf{H}_{i+1}\gamma_{i-1}$$

By induction, we obtain Eq.(4.24).

(3) By Eqs.(4.24) and (4.27), we have

$$\delta_j = \mathbf{H}_i \gamma_j = \mathbf{H}_i \mathbf{Q} \delta_j, \quad j \le i-1$$

since $j = 1, 2, \cdots, n$, we must have

$$\delta_j = \mathbf{H}_{n+1} \mathbf{Q} \delta_j$$

thus Eq.(4.25) is obtained.

Q.E.D.

Thus the quasi-Newton BFGS algorithm is

(0) Initialize

$$\mathbf{x}_0, \mathbf{H}_0 = \mathbf{I}$$

(1) Direction

$$\mathbf{s}_k = -\mathbf{H}_k \mathbf{g}_k$$

(2) Line search

$$\underset{\boldsymbol{\alpha}_k}{\text{Minimize}}\, f\left(\mathbf{x}_k + \alpha_k \mathbf{s}_k\right)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{\alpha}_k \mathbf{s}_k$$

(3) Update Hessian (BFGS)

$$\mathbf{H}_{k+1} = \mathbf{H} + \left(1 + \frac{\gamma'\mathbf{H}\gamma}{\gamma'\mathbf{s}}\right) . \frac{\mathbf{s}\mathbf{s}'}{\gamma'\mathbf{s}} - \left(\frac{\mathbf{s}\gamma'\mathbf{H} + \mathbf{H}\gamma\mathbf{s}'}{\gamma'\mathbf{s}}\right), \quad \gamma = \mathbf{g}_{k+1} - \mathbf{g}_k$$

(4) Terminate

If $\left\|\mathbf{g}_{k+1}\right\| < \boldsymbol{\varepsilon}$, then terminate, else goto step (1)

Remark 4.2:

To save memory, let

$$\mathbf{H}_k = \mathbf{I}$$

in the update formula of BFGS, we have

$$\mathbf{s}_{k+1} = -\mathbf{H}_{k+1}\mathbf{g}_{k+1} = -\mathbf{g}_{k+1} - \left(1 + \frac{\gamma'_k \gamma_k}{\mathbf{s}'_k \gamma_k}\right) \frac{\mathbf{s}_k \mathbf{s}'_k}{\mathbf{s}'_k \gamma_k} \mathbf{g}_{k+1} - \left(\frac{\mathbf{s}_k \gamma'_k + \gamma_k \mathbf{s}'_k}{\mathbf{s}'_k \gamma_k}\right) \mathbf{g}_{k+1}$$

but by Eq.(4.29), we have

$$\mathbf{s}'_k \mathbf{g}_{k+1} = 0$$

so

$$\mathbf{s}_{k+1} = -\mathbf{g}_{k+1} - \frac{\mathbf{s}_k \gamma'_k}{\mathbf{s}'_k \gamma_k} \mathbf{g}_{k+1} = -\mathbf{g}_{k+1} - \beta_{k+1} \mathbf{s}_k, \qquad \beta_{k+1} = \frac{\gamma'_k \mathbf{g}_{k+1}}{\mathbf{s}'_k \gamma_k}$$

we will see in the sequel that this is the conjugate-gradient formula.

5. Conjugate-Gradient Method (Hestenes and Stiefel 1952, Fletcher and Reeves 1964)

Suppose we want to find the minimum of a second-order approximation Taylor expansion

$$f(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \tfrac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x} \tag{5.1}$$

so

$$\mathbf{g}(\mathbf{x}) = \mathbf{c} + \mathbf{Q}\mathbf{x} \tag{5.2}$$

and **Q**: symmetric, positive-definite second-order derivative matrix.

Given $k+1$ linearly independent vectors $\mathbf{s}_0, \mathbf{s}_1, \cdots, \mathbf{s}_k \in \Re^n$. Let $\mathbf{S}_k \in \Re^{n\times(k+1)}$ be the matrix with columns $\mathbf{s}_0, \mathbf{s}_1, \cdots, \mathbf{s}_k$. We have

$$\underset{\mathbf{w}\in\Re^{k+1}}{Minimize}\ f\left(\mathbf{x} + \mathbf{S}_k\mathbf{w}\right) = \underset{\mathbf{w}\in\Re^{k+1}}{Minimize}\left\{\mathbf{w}'\mathbf{S}_k'\mathbf{g}_k + \tfrac{1}{2}\mathbf{w}'\mathbf{S}_k'\mathbf{Q}\mathbf{S}_k\mathbf{w} + \mathbf{x}_k'\left(\mathbf{g}_k - \mathbf{Q}\mathbf{x}_k\right)\right\} \tag{5.3}$$

then the minimum is

$$\mathbf{w} = -\left(\mathbf{S}_k'\mathbf{Q}\mathbf{S}_k\right)^{-1}\mathbf{S}_k'\mathbf{g}_k \tag{5.4}$$

so

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{S}_k\left(\mathbf{S}_k'\mathbf{Q}\mathbf{S}_k\right)^{-1}\mathbf{S}_k'\mathbf{g}_k \tag{5.5}$$

Note that

$$\mathbf{S}_k'\mathbf{g}_{k+1} = \mathbf{S}_k'\left(\mathbf{c} + \mathbf{Q}\mathbf{x}_{k+1}\right) = \mathbf{S}_k'\left[\mathbf{c} + \mathbf{Q}\mathbf{x}_k - \mathbf{Q}\mathbf{S}_k\left(\mathbf{S}_k'\mathbf{Q}\mathbf{S}_k\right)^{-1}\mathbf{S}_k'\mathbf{g}_k\right] = \mathbf{S}_k'\mathbf{g}_k - \mathbf{S}_k'\mathbf{g}_k = 0$$

so

$$\mathbf{g}_{k+1}'\mathbf{s}_i = 0; \quad i = 0,\cdots,k$$

by induction

$$\mathbf{g}_j'\mathbf{s}_i = 0; \quad i < j \tag{5.6}$$

Great simplifications occur in Eq.(5.5) when the matrix $\mathbf{S}_k'\mathbf{Q}\mathbf{S}_k$ is diagonal. Suppose that $k+1$ vectors $\left\{\mathbf{s}_j,\ j = 0,\cdots,k\right\}$ are mutually conjugate with respect to the matrix **Q**, *ie.* conjugacy condition is

$$\mathbf{s}_i'\mathbf{Q}\mathbf{s}_j = 0; \quad i \neq j: 0,\cdots,k \tag{5.7}$$

When Eq.(5.7) holds, Eq.(5.5) becomes

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\mathbf{s}_k \tag{5.8}$$

where

$$\alpha_k = -\frac{\mathbf{g}_k'\mathbf{s}_k}{\mathbf{s}_k'\mathbf{Q}\mathbf{s}_k} \tag{5.9}$$

From Eq.(5.2), we have

$$\gamma_i = \mathbf{g}_{i+1} - \mathbf{g}_i = \mathbf{Q}\left(\mathbf{x}_{i+1} - \mathbf{x}_i\right) = \alpha_i\mathbf{Q}\mathbf{s}_i \tag{5.10}$$

By Eq.(5.7), we obtain the orthogonality condition equivalent to the conjugacy condition

$$\gamma_j'\mathbf{s}_i = 0, \quad i < j \tag{5.11}$$

Now we create a set of mutually conjugate directions by taking $\mathbf{s}_0 = -\mathbf{g}_0$ (the steepest descent direction) and computing each subsequent direction as

$$\mathbf{s}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1}\mathbf{s}_k = -\mathbf{g}_{k+1} + \sum_{i=0}^{k}\beta_i\mathbf{s}_i \tag{5.12}$$

By Eq.(5.6), we have

$$\mathbf{g}_j^{/}\mathbf{g}_i = 0, \quad i < j \tag{5.13}$$

For the conjugacy condition, we must have

$$0 = \mathbf{s}_{k+1}^{/}\mathbf{Q}\mathbf{s}_k = -\left(\mathbf{g}_{k+1} + \beta_{k+1}\mathbf{s}_k\right)^{/}\mathbf{Q}\mathbf{s}_k$$

from Eq.(5.1), we obtain

$$\beta_{k+1} = -\frac{\mathbf{g}_{k+1}^{/}\mathbf{Q}\mathbf{s}_k}{\mathbf{s}_k^{/}\mathbf{Q}\mathbf{s}_k} = -\frac{\mathbf{g}_{k+1}^{/}\gamma_k}{\mathbf{s}_k^{/}\gamma_k} \tag{5.14}$$

by Eqs.(5.11) to (5.13), we have

$$\mathbf{s}_k^{/}\gamma_k = \gamma_k^{/}\mathbf{s}_k = \gamma_k^{/}\left(-\mathbf{g}_k + \sum_{i=0}^{k-1}\beta_i\mathbf{s}_i\right) = -\gamma_k^{/}\mathbf{g}_k = -\left(\mathbf{g}_{k+1} - \mathbf{g}_k\right)^{/}\mathbf{g}_k = \mathbf{g}_k^{/}\mathbf{g}_k$$

so

$$\beta_{k+1} = -\frac{\mathbf{g}_{k+1}^{/}\gamma_k}{\mathbf{s}_k^{/}\gamma_k} = -\frac{\mathbf{g}_{k+1}^{/}\mathbf{g}_{k+1}}{\mathbf{g}_k^{/}\mathbf{g}_k}$$

Thus the conjugate-gradient algorithm is

(0) Initialize $\mathbf{x}_0, \mathbf{H}_0 = \mathbf{I}$

(1) Direction

$$\mathbf{s}_k = -\mathbf{H}_k\mathbf{g}_k$$

(2) Line search

$$\underset{\alpha_k}{\text{Minimize}}\ f\left(\mathbf{x}_k + \alpha_k\mathbf{s}_k\right) \quad \Rightarrow \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{s}_k$$

(3) Update direction

$$\mathbf{s}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1}\mathbf{s}_k$$

where

$$\beta_{k+1} = -\frac{\mathbf{g}_{k+1}^{/}\gamma_k}{\mathbf{s}_k^{/}\gamma_k} - \frac{\mathbf{g}_{k+1}^{/}\mathbf{g}_{k+1}}{\mathbf{g}_k^{/}\mathbf{g}_k}, \quad \gamma_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$

(4) Terminate

If $\left\|\mathbf{g}_{k+1}\right\| < \varepsilon$, then terminate, else goto step (1)

Remark 5.1:

$$\beta_{k+1} = \frac{\mathbf{g}_{k+1}^{/}\mathbf{g}_{k+1}}{\mathbf{g}_k^{/}\mathbf{g}_k} : \text{Fletcher-Reeves} \tag{5.15.a}$$

$$\beta_{k+1} = \frac{\left(\mathbf{g}_{k+1} - \mathbf{g}_k\right)^{/}\mathbf{g}_{k+1}}{\mathbf{g}_k^{/}\mathbf{g}_k} : \text{Polak-Ribiere} \tag{5.15.b}$$

$$\beta_{k+1} = \frac{\mathbf{g}_{k+1}^{/}\left(\mathbf{g}_{k+1} - \mathbf{g}_k\right)}{\mathbf{d}_k^{/}\left(\mathbf{g}_{k+1} - \mathbf{g}_k\right)} : \text{Hestenes-Stiefel} \tag{5.15.c}$$

6. Sum of Squares

Consider the function

$$f(\mathbf{x}) = \tfrac{1}{2}\sum_{i=1}^{m} F_i^2(\mathbf{x}) = \tfrac{1}{2}\left\|\mathbf{F}(\mathbf{x})\right\|^2 \tag{6.1}$$

where

$$\mathbf{F}(\mathbf{x}) = \left[F_1(\mathbf{x}), \cdots, F_m(\mathbf{x})\right]^{/} \tag{6.2}$$

then

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^{m} F_i \frac{\partial F_i}{\partial x_j} = \frac{\partial \mathbf{F}}{\partial x_j}^{/} \mathbf{F}(\mathbf{x}) \tag{6.3}$$

and

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \frac{\partial \mathbf{F}^{/}}{\partial \mathbf{x}} \mathbf{F}(\mathbf{x}) = \mathbf{J}^{/}(\mathbf{x})\mathbf{F}(\mathbf{x}) \tag{6.4}$$

where

$$\mathbf{J}(\mathbf{x}) = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} = \begin{bmatrix} \partial F_1/\partial \mathbf{x} \\ \vdots \\ \partial F_m/\partial \mathbf{x} \end{bmatrix} = \begin{bmatrix} \partial F_1/\partial x_1 & \cdots & \partial F_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial F_m/\partial x_1 & \cdots & \partial F_m/\partial x_n \end{bmatrix} \tag{6.5}$$

For the second-order derivative matrix, we have

$$\frac{\partial^2 f}{\partial x_k \partial x_j} = \sum_{i=1}^{m}\left(\frac{\partial F_i}{\partial x_k}\frac{\partial F_i}{\partial x_j} + F_i \frac{\partial^2 F_i}{\partial x_k \partial x_j}\right) = \frac{\partial \mathbf{F}^{/}}{\partial x_k}\frac{\partial \mathbf{F}}{\partial x_j} + \mathbf{F}^{/}\frac{\partial \mathbf{F}}{\partial x_k \partial x_j} \tag{6.6}$$

and

$$\mathbf{Q}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \mathbf{J}^{/}(\mathbf{x})\mathbf{J}(\mathbf{x}) + \mathbf{F}^{/}(\mathbf{x})\nabla^2\mathbf{F}(\mathbf{x}) = \mathbf{J}^{/}(\mathbf{x})\mathbf{J}(\mathbf{x}) + \mathbf{G}(\mathbf{x}) \tag{6.7}$$

6.1. Gauss-Newton Method

From the Newton condition

$$\mathbf{Q}_k \delta_k = -\mathbf{g}_k \tag{6.8}$$

we have from Eqs.(6.4) and (6.7)

$$\left(\mathbf{J}_k^{/}\mathbf{J}_k + \mathbf{G}_k\right)\delta_k = -\mathbf{J}_k^{/}\mathbf{F}_k \tag{6.9}$$

If $\mathbf{F}_k \to \mathbf{0}$, then $\mathbf{G}_k \to \mathbf{0}$, by Eq.(6.9) we obtain

$$\mathbf{J}_k^{/}\mathbf{J}_k\delta_k = -\mathbf{J}_k^{/}\mathbf{F}_k$$

or

$$\mathbf{J}_k^{/}\left(\mathbf{J}_k\delta_k + \mathbf{F}_k\right) = \mathbf{0} \tag{6.10}$$

hence an algorithm

$$\boxed{\mathbf{x}_{k+1} = \mathbf{x}_k + \delta_k, \qquad \delta_k = -\left(\mathbf{J}_k^{/}\mathbf{J}_k\right)^{-1}\mathbf{J}_k^{/}\mathbf{F}_k}$$

6.2. Levenberg-Marquardt Method (Levenberg 1944, Marquardt 1963)

We modify Eq.(6.9) to have

$$\left(\mathbf{J}_k^{/}\mathbf{J}_k + \lambda_k \mathbf{I}_k\right)\boldsymbol{\delta}_k = -\mathbf{g}_k \tag{6.11}$$

thus

$$\lambda_k \to 0 \;\Rightarrow\; L-M \to G-N$$

$$\lambda_k \to \infty \;\Rightarrow\; L-M \to S-D \left(\text{steepest descent}\right)$$

hence an algorithm

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{\delta}_k, \qquad \boldsymbol{\delta}_k = -\left(\mathbf{J}_k^{/}\mathbf{J}_k + \lambda_k \mathbf{I}_k\right)^{-1}\mathbf{g}_k$$

7. Cubic Line Search

The most efficient line search is the cubic interpolation proposed by Davidon in 1959 for the following problem

$$\underset{h}{\text{Minimize}}\; f\left(\mathbf{x}_k + h\mathbf{d}_k\right) = \underset{h}{\text{Minimize}}\; \tilde{f}(h)$$

Based on $h, \mathbf{x}_k$ *and* $\mathbf{d}_k$, this technique finds 2 points $\mathbf{x}_p$ *and* $\mathbf{x}_q$ that bracket the minimum then uses $h, \tilde{f}_p, \tilde{f}_p^{/}, \tilde{f}_q, \tilde{f}_q^{/}$ to fit the following cubic polynomial in order to compuite the minimal step h^*

$$\psi(h) = a + bh + ch^2 + dh^3 \tag{7.1}$$

$$\psi^{/}(h) = b + 2ch + 3dh^2 \tag{7.2}$$

$$\psi^{//}(h) = 2c + 6dh \tag{7.3}$$

then the minimum is the solution of $\psi^{/}\left(h^*\right) = 0$ and $\psi^{//}\left(h^*\right) > 0$, *ie.*

$$h^* = \frac{-c + \sqrt{c^2 - 3bd}}{3d} \tag{7.4}$$

since

$$\psi^{//}\left(h^*\right) = 2\sqrt{c^2 - 3bd}$$

We have

$$\tilde{f}^{/}(h) = \frac{d\tilde{f}}{dh} = \nabla f\left(\mathbf{x}_p + h\mathbf{d}_k\right).\mathbf{d}_k^T = \mathbf{g}\left(\mathbf{x}_p + h\mathbf{d}_k\right).\mathbf{d}_k^T \tag{7.5}$$

thus

$$\tilde{f}_p^{/} = \tilde{f}^{/}(0) = \mathbf{g}\left(\mathbf{x}_p\right).\mathbf{d}_k^T = \mathbf{g}_p\mathbf{d}_k^T \tag{7.6}$$

$$\tilde{f}_q^{/} = \tilde{f}^{/}(h) = \mathbf{g}\left(\mathbf{x}_p + h\mathbf{d}\right).\mathbf{d}_k^T = \mathbf{g}_q\mathbf{d}_k^T \tag{7.7}$$

Then a, b, c, d are solutions of the equation set below

$$\left.\begin{aligned} \tilde{f}(0) &= \tilde{f}_p = a \\ \tilde{f}^{/}(0) &= \tilde{f}_p^{/} = b \\ \tilde{f}(h) &= \tilde{f}_q = a + bh + ch^2 + dh^3 \\ \tilde{f}^{/}(h) &= \tilde{f}_q^{/} = b + 2ch + 3dh^2 \end{aligned}\right\} \tag{7.8}$$

so

$$a = \tilde{f}_p, \quad b = \tilde{f}_p', \quad c = -\frac{z + \tilde{f}_p'}{h}, \quad d = \frac{\tilde{f}_p' + \tilde{f}_q' + 2z}{3h^2} \tag{7.9}$$

where

$$z = \tilde{f}_p' + \tilde{f}_q' + \frac{3\left(\tilde{f}_p - \tilde{f}_q\right)}{h} \tag{7.10}$$

The minimum Eq.(7.4) becomes

$$h^* = h\frac{w + z + \tilde{f}_p'}{2z + \tilde{f}_p' + \tilde{f}_q'} = h\frac{w + z - \tilde{f}_p'}{2z + \tilde{f}_q' - \tilde{f}_p'} \tag{5.16}$$

where

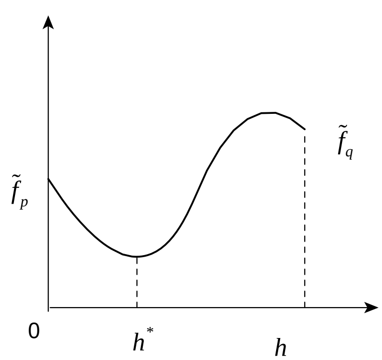
$$w = \sqrt{z^2 - \tilde{f}_p' \tilde{f}_q'} \tag{7.11}$$

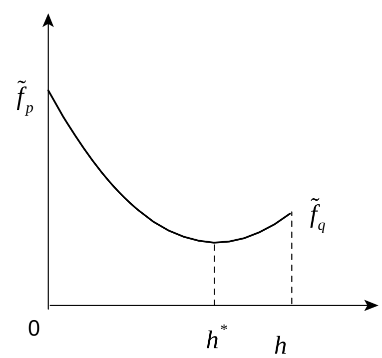
The choice of *h* is at our discretion. If $\tilde{f}_p' < 0$ we would take $h > 0$, *ie.* take a step in the direction of decreasing $\tilde{f}(h)$; otherwise we would take $h < 0$. The magnitude of *h* is such that the interval $[0, h]$ includes the minimum. This will be so if $\tilde{f}_q > \tilde{f}_p$ or if $\tilde{f}_q' > 0$. When neither of these conditions is satisfied, we double the value of *h*, repeatedly if necessary, until our interval does bracket a minimum.

The problem of finding an initial value for *h* remains. There are real difficulties in finding a value that will be satisfactory for *all* problems. Davidon and Fletcher and Powell suggest

$$h_0 = \min\left\{1, -\frac{2\left(\tilde{f}_p - \tilde{f}_m\right)}{\tilde{f}_p'}\right\} \tag{7.12}$$

where $\tilde{f}_m$ is an estimate of the minimum of $\tilde{f}(h)$.





The cubic algorithm is below

(0) Guess $\tilde{f}_m$

Let $\mathbf{x}_p = \mathbf{x}_k$

(1) Find the first point **P** such that $\tilde{f}_p^{/} < 0$

Repeat

$$\tilde{f}_p = \tilde{f}(\mathbf{x}_p), \quad \tilde{f}_p^{/} = \mathbf{g}_p \mathbf{d}_k^T, \quad h = \min\left\{1, \left|\frac{2\left(\tilde{f}_p - \tilde{f}_m\right)}{\tilde{f}_p^{/}}\right|\right\}$$

$\mathbf{x}_p = \mathbf{x}_p - h\mathbf{d}_k$ // Reverse serach direction to get **P**

Until $f_p^{/} < 0$

(2) Find the second point **Q** such that $\tilde{f}_q^{/} > 0$ or $\tilde{f}_q > \tilde{f}_p$

Repeat

$$\tilde{h} = h$$

$$\mathbf{x}_q = \mathbf{x}_p + \tilde{h}\mathbf{d}_k$$

$$\tilde{f}_q = \tilde{f}(\mathbf{x}_q), \quad \tilde{f}_q^{/} = \mathbf{g}_q \mathbf{d}_k^T$$

$h = 2h$ // Double step to get **Q**

Until $\tilde{f}_q^{/} > 0$ or $\tilde{f}_q > \tilde{f}_p$

(3) Find the minimum after having 2 points

$$z = \tilde{f}_p^{/} + \tilde{f}_q^{/} + \frac{3\left(\tilde{f}_p - \tilde{f}_q\right)}{h}$$

$$w = \sqrt{\max\left(0, z^2 - \tilde{f}_p^{/}\tilde{f}_q^{/}\right)}$$

$$h^* = h\frac{w + z - \tilde{f}_p^{/}}{2z + \tilde{f}_q^{/} - \tilde{f}_p^{/}}$$

8. Conclusion

The steepest descent method minimizes a function on its first-order approximation in Taylor series, while the Newton and quasi-Newton methods use the second-order approximation. The main advantage of the Newton method is it reaches the minimum point after *n* iterations from any given point (convergence). The main disadvantages of the Newton method are requirement of an inverse of the positive-definite Hessian (second derivative matrix) which is mathematically expensive and a mechanism to guarantee the positive-definiteness.

The quasi-Newton method does have the advantage (convergence) and does not have the disadvantages (Hessian) of the Newton method. The main features of quasi-Newton method are

- $\mathbf{H}_i$ is maintained to be symmetric and positive-definite;
- $\mathbf{H}_i \rightarrow \mathbf{Q}_i^{-1}$ after *n* iterations from any given symmetric and positive-definite $\mathbf{H}_0$ for an exact line-search
- $\mathbf{x}_k \rightarrow \mathbf{x}^*$ after *n* iterations from any given point $\mathbf{x}_0$ for an exact line-search

For an inexact line-search, the number of iterations increases accordingly.

When there is a memory limitation we can use the conjugate-gradient method, but it is slower than the quasi-Newton method since it estimates $\mathbf{H}_k$ as $\mathbf{I}$ in the BFGS update formula.