Numerical Optimization

Unconstrained Optimization

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Coordinate Descent Method

Consider the problem,

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x})$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R},\ f \in \mathcal{C}^1$.

Idea :

1. For every coordinate variable $x_i, i = 1, \ldots, n$, minimize $f(\boldsymbol{x})$ w.r.t. x_i, keeping the other coordinate variables $x_j,\ j \neq i$ *constant*.
2. Repeat the procedure in step 1 until some stopping condition is satisfied.

Coordinate Descent Method

(1) Initialize $\boldsymbol{x}^0$, ϵ , set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

for $i = 1, \ldots, n$

$x_i^{new} = \arg\min_{x_i} f(\boldsymbol{x})$

$x_i = x_i^{new}$

endfor

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.

- Globally convergent method if a search along any coordinate direction yields a unique minimum point

Example: Consider the problem,

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \stackrel{\Delta}{=} 4x_1^2 + x_2^2$$

We use coordinate descent method with *exact line search* to solve this problem.

- $\boldsymbol{x}^0 = (-1, -1)^T$
- Let $\boldsymbol{d}^0 = (1, 0)^T$
- $\boldsymbol{x}^1 = \boldsymbol{x}^0 + \alpha^0 \boldsymbol{d}^0$ where

$$\alpha^0 = \arg\min_{\alpha} \phi_0(\alpha) \stackrel{\Delta}{=} f(\boldsymbol{x}^0 + \alpha \boldsymbol{d}^0)$$

- $\phi_0(\alpha) = f\begin{pmatrix} x_1^0 + \alpha d_1^0 \\ x_2^0 + \alpha d_2^0 \end{pmatrix} = 4(\alpha - 1)^2 + 1$
- $\phi_0'(\alpha) = 0 \ \Rightarrow \ \alpha^0 = 1 \ \Rightarrow \ \boldsymbol{x}^1 = ({\color{red}0}, -1)^T$
- $\boldsymbol{d}^1 = (0, 1)^T, \ \boldsymbol{x}^2 = \boldsymbol{x}^1 + \alpha^1 \boldsymbol{d}^1, \ \alpha^1 = \arg\min_\alpha \phi_1(\alpha) \stackrel{\Delta}{=}$ $f\begin{pmatrix} 0 \\ \alpha - 1 \end{pmatrix} = (\alpha - 1)^2 \ \Rightarrow \ \alpha^1 = 1 \ \Rightarrow \ \boldsymbol{x}^2 = (0, {\color{red}0})^T {\color{red}= \boldsymbol{x}^*}$

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \overset{\Delta}{=} 4x_1^2 + x_2^2$$

For the above problem,

- Moving along coordinate directions and using exact lines search gives the solution in at most two steps.
- Same result is obtained even if $\boldsymbol{d}^0$ and $\boldsymbol{d}^1$ are interchanged.

Example: Consider the problem,

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \stackrel{\Delta}{=} 4x_1^2 + x_2^2 - 2x_1x_2$$

We use coordinate descent method with *exact line search* to solve this problem.

- $\boldsymbol{x}^0 = (-1, -1)^T$
- Let $\boldsymbol{d}^0 = (1, 0)^T$
- $\boldsymbol{x}^1 = \boldsymbol{x}^0 + \alpha^0 \boldsymbol{d}^0$ where

$$\alpha^0 = \arg\min_{\alpha} \phi_0(\alpha) \stackrel{\Delta}{=} f(\boldsymbol{x}^0 + \alpha \boldsymbol{d}^0)$$

- $\phi_0(\alpha) = f\begin{pmatrix} x_1^0 + \alpha d_1^0 \\ x_2^0 + \alpha d_2^0 \end{pmatrix} = 4(\alpha - 1)^2 + 1 + 2(\alpha - 1)$
- $\phi_0'(\alpha) = 0 \;\Rightarrow\; \alpha^0 = \frac{3}{4} \;\Rightarrow\; \boldsymbol{x}^1 = (-\frac{1}{4}, -1)^T$
- $\boldsymbol{d}^1 = (0, 1)^T, \; \boldsymbol{x}^2 = \boldsymbol{x}^1 + \alpha^1 \boldsymbol{d}^1, \; \alpha^1 = \arg\min_\alpha \phi_1(\alpha) \stackrel{\Delta}{=}$ $f\begin{pmatrix} -\frac{1}{4} \\ \alpha - 1 \end{pmatrix} = (\alpha - 1)^2 + \frac{\alpha - 1}{2} + \frac{1}{4} \;\Rightarrow\; \alpha^1 = \frac{3}{4} \;\Rightarrow\; \boldsymbol{x}^2 =$ $(-\frac{1}{4}, -\frac{1}{4})^T \neq \boldsymbol{x}^*$

- Example 1:

$$\min_{\boldsymbol{x}} \; f_1(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2$$

  - $\boldsymbol{H} = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}$.
  - $\boldsymbol{x}^*$, attained in *at most two steps* using coordinate descent method

- Example 2:

$$\min_{\boldsymbol{x}} \; f_2(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

  - $\boldsymbol{H} = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}$.
  - $\boldsymbol{x}^*$, could not be attained in two steps using coordinate descent method (if $\boldsymbol{x}^0$ is not on one of the principal axes of the elliptical contours)

Consider the problem:

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \triangleq \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}$$

where $\boldsymbol{H}$ is a symmetric positive definite matrix.

- Let $\{\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}\}$ be a set of linearly independent directions and $\boldsymbol{x}^0 \in \mathbb{R}^n$
- Any $\boldsymbol{x} \in \mathbb{R}^n$ can be represented as

$$\boldsymbol{x} = \boldsymbol{x}^0 + \sum_{i=0}^{n-1} \alpha^i \boldsymbol{d}^i$$

- Given $\{\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}\}$ and $\boldsymbol{x}^0 \in \mathbb{R}^n$, the given problem is to minimize $\Psi(\boldsymbol{\alpha})$ defined as,

$$\frac{1}{2}\left(\boldsymbol{x}^0 + \sum_{i=0}^{n-1} \alpha^i \boldsymbol{d}^i\right)^T \boldsymbol{H} \left(\boldsymbol{x}^0 + \sum_{i=0}^{n-1} \alpha^i \boldsymbol{d}^i\right) + \boldsymbol{c}^T \left(\boldsymbol{x}^0 + \sum_{i=0}^{n-1} \alpha^i \boldsymbol{d}^i\right)$$

Define $\boldsymbol{D} = (\boldsymbol{d}^0 | \boldsymbol{d}^1 | \ldots | \boldsymbol{d}^{n-1})$ and $\boldsymbol{\alpha} = (\alpha^0, \alpha^1, \ldots, \alpha^{n-1})$.

$$\Psi(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^T \underbrace{\boldsymbol{D}^T \boldsymbol{H} \boldsymbol{D}}_{\boldsymbol{Q}} \boldsymbol{\alpha} + (\boldsymbol{H}\boldsymbol{x}^0 + \boldsymbol{c})^T \boldsymbol{D}\boldsymbol{\alpha} + \underbrace{\frac{1}{2}{\boldsymbol{x}^0}^T \boldsymbol{H}\boldsymbol{x}^0 + \boldsymbol{c}^T \boldsymbol{x}^0}_{\textit{constant}}$$

$$\boldsymbol{Q} = \boldsymbol{D}^T \boldsymbol{H} \boldsymbol{D} = \begin{pmatrix} {\boldsymbol{d}^0}^T \boldsymbol{H} \boldsymbol{d}^0 & {\boldsymbol{d}^0}^T \boldsymbol{H} \boldsymbol{d}^1 & \ldots & {\boldsymbol{d}^0}^T \boldsymbol{H} \boldsymbol{d}^{n-1} \\ {\boldsymbol{d}^1}^T \boldsymbol{H} \boldsymbol{d}^0 & {\boldsymbol{d}^1}^T \boldsymbol{H} \boldsymbol{d}^1 & \ldots & {\boldsymbol{d}^1}^T \boldsymbol{H} \boldsymbol{d}^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ {\boldsymbol{d}^{n-1}}^T \boldsymbol{H} \boldsymbol{d}^0 & {\boldsymbol{d}^{n-1}}^T \boldsymbol{H} \boldsymbol{d}^1 & \ldots & {\boldsymbol{d}^{n-1}}^T \boldsymbol{H} \boldsymbol{d}^{n-1} \end{pmatrix}$$

$\boldsymbol{Q}$ will be diagonal matrix if ${\boldsymbol{d}^i}^T \boldsymbol{H} \boldsymbol{d}^j = 0, \ \forall \ i \neq j$.

Let ${\boldsymbol{d}^i}^T \boldsymbol{H} \boldsymbol{d}^j = 0, \ \forall \ i \neq j.$

$$\boldsymbol{Q} = \boldsymbol{D}^T \boldsymbol{H} \boldsymbol{D} = \begin{pmatrix} {\boldsymbol{d}^0}^T \boldsymbol{H} \boldsymbol{d}^0 & 0 & \dots & 0 \\ 0 & {\boldsymbol{d}^1}^T \boldsymbol{H} \boldsymbol{d}^1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & {\boldsymbol{d}^{n-1}}^T \boldsymbol{H} \boldsymbol{d}^{n-1} \end{pmatrix}$$

Therefore,

$$\boldsymbol{Q}^{-1}_{ij} = \begin{cases} \frac{1}{{\boldsymbol{d}^i}^T \boldsymbol{H} \boldsymbol{d}^i} & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

$$\Psi(\boldsymbol{\alpha}) = \frac{1}{2}(\boldsymbol{x}^0 + \sum_i \alpha^i \boldsymbol{d}^i)^T \boldsymbol{H}(\boldsymbol{x}^0 + \sum_i \alpha^i \boldsymbol{d}^i) + \boldsymbol{c}^T(\boldsymbol{x}^0 + \sum_i \alpha^i \boldsymbol{d}^i)$$

$$= \frac{1}{2} \sum_i \left[(\boldsymbol{x}^0 + \alpha^i \boldsymbol{d}^i)^T \boldsymbol{H}(\boldsymbol{x}^0 + \alpha^i \boldsymbol{d}^i) + 2\boldsymbol{c}^T(\boldsymbol{x}^0 + \alpha^i \boldsymbol{d}^i) \right] + \text{constant}$$

- $\Psi(\boldsymbol{\alpha})$ is separable in terms of $\alpha^0, \alpha^1, \dots, \alpha^{n-1}$

$$\Psi(\boldsymbol{\alpha}) = \frac{1}{2}\sum_i \left[(\boldsymbol{x}^0 + \alpha^i \boldsymbol{d}^i)^T \boldsymbol{H}(\boldsymbol{x}^0 + \alpha^i \boldsymbol{d}^i) + 2\boldsymbol{c}^T(\boldsymbol{x}^0 + \alpha^i \boldsymbol{d}^i) \right]$$

$$\frac{\partial \Psi}{\partial \alpha^i} = 0 \Rightarrow \alpha^{i*} = -\frac{\boldsymbol{d}^{i^T}(\boldsymbol{H}\boldsymbol{x}^0 + \boldsymbol{c})}{\boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^i}$$

Therefore,

$$\boldsymbol{x}^* = \boldsymbol{x}^0 + \sum_{i=0}^{n-1} \alpha^{i*} \boldsymbol{d}^i$$

Definition

Let $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The vectors $\{\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}\}$ are said to be $\boldsymbol{H}$-*conjugate* if they are linearly independent and $\boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^j = 0 \ \forall\ i \neq j$.

Example: Consider the problem,

$$\min_{\boldsymbol{x}} \ f(\boldsymbol{x}) \stackrel{\Delta}{=} 4x_1^2 + x_2^2 - 2x_1x_2$$

- $\boldsymbol{H} = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}$
- $\boldsymbol{x}^0 = (-1, -1)^T$
- Let $\boldsymbol{d}^0 = (1, 0)^T$
- $\boldsymbol{x}^1 = \boldsymbol{x}^0 + \alpha^0 \boldsymbol{d}^0$ where

$$\alpha^0 = \arg\min_{\alpha} \phi_0(\alpha) \stackrel{\Delta}{=} f(\boldsymbol{x}^0 + \alpha \boldsymbol{d}^0)$$

- $\phi_0(\alpha) = f\begin{pmatrix} x_1^0 + \alpha d_1^0 \\ x_2^0 + \alpha d_2^0 \end{pmatrix} = 4(\alpha - 1)^2 + 1 + 2(\alpha - 1)$
- $\phi_0'(\alpha) = 0 \ \Rightarrow \ \alpha^0 = \frac{3}{4} \ \Rightarrow \ \boldsymbol{x}^1 = (-\frac{1}{4}, -1)^T$
- Choose a non-zero direction $\boldsymbol{d}^1$ such that ${\boldsymbol{d}^1}^T \boldsymbol{H} \boldsymbol{d}^0 = 0$
- Let $\boldsymbol{d}^1 = (a, b)^T$. Therefore,
  $(a \ \ b)\begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0 \ \Rightarrow \ 8a - 2b = 0$

- Let $\boldsymbol{d}^1 = (1, 4)^T$,
- $\boldsymbol{x}^2 = \boldsymbol{x}^1 + \alpha^1 \boldsymbol{d}^1$ where

$$\alpha^1 = \arg\min_{\alpha} \phi_1(\alpha) \triangleq f\begin{pmatrix} \alpha - \frac{1}{4} \\ 4\alpha - 1 \end{pmatrix} = \frac{3}{4}(4\alpha - 1)^2$$

- $\phi_1'(\alpha) = 0 \;\Rightarrow\; \alpha^1 = \frac{1}{4}$
- $\boldsymbol{x}^2 = \boldsymbol{x}^1 + \alpha^1 \boldsymbol{d}^1 = (0, 0)^T = \boldsymbol{x}^*$

A convex quadratic function can be minimized in, *at most*, *n* steps, provided we search along conjugate directions of the Hessian matrix.

Given $\boldsymbol{H}$, does a set of $\boldsymbol{H}$-conjugate vectors exist? If yes, how to get a set of such vectors?

Conjugate Directions

Let $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ be a symmetric matrix.

- Do there exist n conjugate directions w.r.t $\boldsymbol{H}$?
  $\boldsymbol{H}$ is symmetric $\Rightarrow$ $\boldsymbol{H}$ has n mutually orthogonal eigenvectors.
  Let $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ be two orthogonal eigenvectors of $\boldsymbol{H}$.
  $\therefore \boldsymbol{v}_1^T \boldsymbol{v}_2 = 0.$

$$
\begin{aligned}
\boldsymbol{H}\boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1 \quad &\Rightarrow \quad \boldsymbol{v}_2^T \boldsymbol{H} \boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_2^T \boldsymbol{v}_1 \\
&\Rightarrow \quad \boldsymbol{v}_2^T \boldsymbol{H} \boldsymbol{v}_1 = 0 \\
&\Rightarrow \quad \boldsymbol{v}_1 \text{ and } \boldsymbol{v}_2 \text{ are } \boldsymbol{H}\text{-conjugate}
\end{aligned}
$$

$\therefore n$ orthogonal eigenvectors of $\boldsymbol{H}$ are $\boldsymbol{H}$-conjugate.

Conjugate Directions

- Let $\boldsymbol{H}$ be a symmetric positive definite matrix and $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$ be nonzero directions such that

$$\boldsymbol{d}^{i^T} \boldsymbol{H} \boldsymbol{d}^j = 0, \ i \neq j.$$

Are $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$ linearly independent?

$$\begin{aligned}
\sum_{i=0}^{n-1} \mu^i \boldsymbol{d}^i = 0 \quad &\Rightarrow \quad \sum_{i=0}^{n-1} \mu^i \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^i = 0 \text{ for every } j = 0, \ldots, n-1 \\
&\Rightarrow \quad \mu^j \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j = 0 \\
&\Rightarrow \quad \mu^j = 0 \ \text{ for every } j = 0, \ldots, n-1 \\
&\Rightarrow \quad \boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1} \ \text{ are linearly independent}
\end{aligned}$$

Conjugate Directions

Geometric Interpretation:
Consider the problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} \ \frac{1}{2}\boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x}, \quad \boldsymbol{H} \text{ symmetric positive definite matrix.}$$

Let $\boldsymbol{x}^*$ be the solution. $\therefore \boldsymbol{H}\boldsymbol{x}^* = -\boldsymbol{c}$.
Let $\boldsymbol{x}^0$ be any initial point. $\boldsymbol{g}^0 = \boldsymbol{H}\boldsymbol{x}^0 + \boldsymbol{c}$
Let $\boldsymbol{d}^0$ be some direction ($\boldsymbol{d}^0 \neq \boldsymbol{0}$).
$\boldsymbol{x}^1$ is found by doing exact line search along $\boldsymbol{d}^0$. $\therefore {\boldsymbol{g}^1}^T \boldsymbol{d}^0 = 0$.
$\boldsymbol{g}^1 = \boldsymbol{H}\boldsymbol{x}^1 + \boldsymbol{c}$.

$$\begin{aligned} (\boldsymbol{x}^* - \boldsymbol{x}^1)^T \boldsymbol{H}\boldsymbol{d}^0 &= (\boldsymbol{H}\boldsymbol{x}^* - \boldsymbol{H}\boldsymbol{x}^1)^T \boldsymbol{d}^0 \\ &= -{\boldsymbol{g}^1}^T \boldsymbol{d}^0 \\ &= 0 \end{aligned}$$

Therefore, the direction $(\boldsymbol{x}^* - \boldsymbol{x}^1)$ is $\boldsymbol{H}$ conjugate to $\boldsymbol{d}^0$.

Consider the problem:

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \triangleq \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x}+\boldsymbol{c}^T\boldsymbol{x}, \quad \boldsymbol{H} \text{ symmetric positive definite matrix.}$$

Let $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$ be $\boldsymbol{H}$-conjugate. $\therefore \boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$ are linearly independent.

Let $\mathcal{B}^k$ denote the subspace spanned by $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{k-1}$.

Clearly, $\mathcal{B}^k \subset \mathcal{B}^{k+1}$.

Let $\boldsymbol{x}^0 \in \mathbb{R}^n$ be any arbitrary point.

Let $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$ where α^k is obtained by doing exact line search:

$$\alpha^k = \arg\min_{\alpha} \quad f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k)$$

Claim:

$$\begin{aligned} \boldsymbol{x}^k \;&=\; \arg\min\nolimits_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \\ &\quad\;\; \text{s.t.} \quad \boldsymbol{x} \in \boldsymbol{x}^0 + \mathcal{B}^k \end{aligned}$$

Exact line search:

$$\alpha^k = \arg\min_{\alpha \in \mathbb{R}} \quad f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k)$$

Therefore,

$\nabla f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k)^T \boldsymbol{d}^k = 0 \Rightarrow {\boldsymbol{g}^{k+1}}^T \boldsymbol{d}^k = 0 \; \forall \; k = 0, \ldots, n-1$

$\boldsymbol{x}^k = \boldsymbol{x}^{k-1} + \alpha^{k-1}\boldsymbol{d}^{k-1} = \boldsymbol{x}^j + \sum_{i=j}^{k-1} \alpha^i \boldsymbol{d}^i \qquad (j = 0, \ldots, k-1)$

$$\therefore \boldsymbol{H}\boldsymbol{x}^k + \boldsymbol{c} \quad = \quad \boldsymbol{H}\boldsymbol{x}^j + \boldsymbol{c} + \sum_{i=j}^{k-1} \alpha^i \boldsymbol{H}\boldsymbol{d}^i$$

$$\therefore \boldsymbol{g}^k \quad = \quad \boldsymbol{g}^j + \sum_{i=j}^{k-1} \alpha^i \boldsymbol{H}\boldsymbol{d}^i$$

$$\therefore {\boldsymbol{g}^k}^T \boldsymbol{d}^{j-1} \quad = \quad {\boldsymbol{g}^j}^T \boldsymbol{d}^{j-1} + \sum_{i=j}^{k-1} \alpha^i {\boldsymbol{d}^i}^T \boldsymbol{H}\boldsymbol{d}^{j-1} = 0$$

Therefore, ${\boldsymbol{g}^k}^T \boldsymbol{d}^j = 0 \; \forall \; j = 0, \ldots, k-1$ or $\boldsymbol{g}^k \perp \mathcal{B}^k$

Note that for every $j = 0, \ldots, n-1$,

$$\alpha^j = \arg\min_{\alpha} \quad f(\boldsymbol{x}^j + \alpha \boldsymbol{d}^j)$$

$$\therefore f(\boldsymbol{x}^j + \alpha^j \boldsymbol{d}^j) \leq f(\boldsymbol{x}^j + \mu^j \boldsymbol{d}^j), \quad \mu^j \in \mathbb{R}$$

$$\therefore f(\boldsymbol{x}^j) + \alpha^j \boldsymbol{g}^{j^T} \boldsymbol{d}^j + \frac{1}{2} \alpha^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j \leq f(\boldsymbol{x}^j) + \mu^j \boldsymbol{g}^{j^T} \boldsymbol{d}^j + \frac{1}{2} \mu^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j$$

We need to show that $f(\boldsymbol{x}^k) \leq f(\boldsymbol{x}) \; \forall \; \boldsymbol{x} \in \boldsymbol{x}^0 + \mathcal{B}^k$ or

$$f(\boldsymbol{x}^0 + \sum_{j=0}^{k-1} \alpha^j \boldsymbol{d}^j) \leq f(\boldsymbol{x}^0 + \sum_{j=0}^{k-1} \mu^j \boldsymbol{d}^j), \quad \mu^j \in \mathbb{R} \; \forall \; j.$$

That is,

$$f(\boldsymbol{x}^0) + \sum_{j=0}^{k-1} (\alpha^j \boldsymbol{g}^{0^T} \boldsymbol{d}^j + \frac{1}{2} \alpha^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j) \leq f(\boldsymbol{x}^0) + \sum_{j=0}^{k-1} (\mu^j \boldsymbol{g}^{0^T} \boldsymbol{d}^j + \frac{1}{2} \mu^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j)$$

where $\mu^j \in \mathbb{R} \; \forall \; j$.

For every $j = 0, \ldots, n-1$,

$$f(\boldsymbol{x}^j) + \alpha^j \boldsymbol{g}^{j^T} \boldsymbol{d}^j + \frac{1}{2}\alpha^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j \quad \leq \quad f(\boldsymbol{x}^j) + \mu^j \boldsymbol{g}^{j^T} \boldsymbol{d}^j + \frac{1}{2}\mu^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j$$

Suppose $\boldsymbol{g}^{j^T} \boldsymbol{d}^j = \boldsymbol{g}^{0^T} \boldsymbol{d}^j \; \forall \; j$

$$\therefore \alpha^j \boldsymbol{g}^{0^T} \boldsymbol{d}^j + \frac{1}{2}\alpha^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j \leq \mu^j \boldsymbol{g}^{0^T} \boldsymbol{d}^j + \frac{1}{2}\mu^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j \quad \forall \; j$$

Therefore,

$$f(\boldsymbol{x}^0) + \sum_{j=0}^{k-1} (\alpha^j \boldsymbol{g}^{0^T} \boldsymbol{d}^j + \frac{1}{2}\alpha^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j) \leq f(\boldsymbol{x}^0) + \sum_{j=0}^{k-1} (\mu^j \boldsymbol{g}^{0^T} \boldsymbol{d}^j + \frac{1}{2}\mu^{j^2} \boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j)$$

$$\therefore f(\boldsymbol{x}^0 + \sum_{j=0}^{k-1} \alpha^j \boldsymbol{d}^j) \quad \leq \quad f(\boldsymbol{x}^0 + \sum_{j=0}^{k-1} \mu^j \boldsymbol{d}^j), \quad \mu^j \in \mathbb{R} \; \forall \; j$$

$$\therefore f(\boldsymbol{x}^k) \quad \leq \quad f(\boldsymbol{x}), \quad \forall \; \boldsymbol{x} \in \boldsymbol{x}^0 + \mathcal{B}^k$$

We need to show that

$$\boldsymbol{g}^{j^T}\boldsymbol{d}^j = \boldsymbol{g}^{0^T}\boldsymbol{d}^j \quad \forall\, j$$

Consider, $\boldsymbol{x}^j = \boldsymbol{x}^0 + \sum_{i=0}^{j-1} \alpha^i \boldsymbol{d}^i$.

$$\begin{aligned}
\therefore \boldsymbol{H}\boldsymbol{x}^j + \boldsymbol{c} &= \boldsymbol{H}\boldsymbol{x}^0 + \boldsymbol{c} + \sum_{i=0}^{j-1} \alpha^i \boldsymbol{H}\boldsymbol{d}^i \\
\therefore \boldsymbol{g}^j &= \boldsymbol{g}^0 + \sum_{i=0}^{j-1} \alpha^i \boldsymbol{H}\boldsymbol{d}^i \\
\therefore \boldsymbol{g}^{j^T}\boldsymbol{d}^j &= \boldsymbol{g}^{0^T}\boldsymbol{d}^j + \sum_{i=0}^{j-1} \alpha^i \boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^j \\
\therefore \boldsymbol{g}^{j^T}\boldsymbol{d}^j &= \boldsymbol{g}^{0^T}\boldsymbol{d}^j \quad \forall\, j
\end{aligned}$$

Expanding Subspace Theorem

Consider the problem to minimize $f(\boldsymbol{x}) \triangleq \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}$ where $\boldsymbol{H}$ is symmetric positive definite matrix. Let $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$ be $\boldsymbol{H}$-conjugate and let $\boldsymbol{x}^0 \in \mathbb{R}^n$ be any initial point. Let

$$\alpha^k = \arg\min_{\alpha \in \mathbb{R}} f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k), \; \forall\, k = 0, \ldots, n-1$$

and
$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k, \; \forall\, k = 0, \ldots, n-1.$$

Then, for all $k = 0, \ldots, n-1$,

1. ${\boldsymbol{g}^k}^T \boldsymbol{d}^j = 0, \; j = 0, \ldots, k$
2. ${\boldsymbol{g}^k}^T \boldsymbol{d}^k = {\boldsymbol{g}^0}^T \boldsymbol{d}^k$
3.

$$\begin{aligned} \boldsymbol{x}^{k+1} \quad = \quad & \arg\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \\ & \text{s.t.} \quad \boldsymbol{x} \in \boldsymbol{x}^0 + \mathcal{B}^k \end{aligned}$$

Given a set of n directions, $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$ which are $\boldsymbol{H}$-conjugate and $\boldsymbol{x}^0 \in \mathbb{R}^n$, it is easy to determine $\alpha^{i^*},\ \forall\, i = 0, \ldots, n-1,$

$$\alpha^{i^*} = -\frac{\boldsymbol{d}^{i^T}(\boldsymbol{H}\boldsymbol{x}^0 + \boldsymbol{c})}{\boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^i}$$

and get

$$\boldsymbol{x}^* = \boldsymbol{x}^0 + \sum_{i=0}^{n-1} \alpha^{i^*}\boldsymbol{d}^i$$

- How do we construct the $\boldsymbol{H}$-conjugate directions, $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$?
- Given the $\boldsymbol{H}$-conjugate directions, $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{k-1}$, how do we determine α^k where

$$\alpha^k = \arg\min_{\alpha}\ f(\boldsymbol{x}^k + \alpha\boldsymbol{d}^k)?$$

$$\boldsymbol{x}^* - \boldsymbol{x}^0 = \sum_{i=0}^{n-1} \alpha^i \boldsymbol{d}^i$$

$$\therefore \boldsymbol{d}^{k^T} \boldsymbol{H}(\boldsymbol{x}^* - \boldsymbol{x}^0) = \alpha^k \boldsymbol{d}^{k^T} \boldsymbol{H} \boldsymbol{d}^k$$

$$\therefore \alpha^k = \frac{\boldsymbol{d}^{k^T} \boldsymbol{H}(\boldsymbol{x}^* - \boldsymbol{x}^0)}{\boldsymbol{d}^{k^T} \boldsymbol{H} \boldsymbol{d}^k}$$

Suppose that after k iterative steps and obtaining k $\boldsymbol{H}$-conjugate directions,

$$\boldsymbol{x}^k - \boldsymbol{x}^0 = \sum_{i=0}^{k-1} \alpha^i \boldsymbol{d}^i$$

$$\therefore \boldsymbol{d}^{k^T} \boldsymbol{H}(\boldsymbol{x}^k - \boldsymbol{x}^0) = 0$$

Given, $\boldsymbol{d}^{k^T}\boldsymbol{H}(\boldsymbol{x}^k - \boldsymbol{x}^0) = 0$,

$$
\begin{aligned}
\therefore \alpha^k &= \frac{\boldsymbol{d}^{k^T}\boldsymbol{H}(\boldsymbol{x}^* - \boldsymbol{x}^k + \boldsymbol{x}^k - \boldsymbol{x}^0)}{\boldsymbol{d}^{k^T}\boldsymbol{H}\boldsymbol{d}^k} \\
&= \frac{\boldsymbol{d}^{k^T}(\boldsymbol{H}\boldsymbol{x}^* - \boldsymbol{H}\boldsymbol{x}^k)}{\boldsymbol{d}^{k^T}\boldsymbol{H}\boldsymbol{d}^k} \\
&= \frac{\boldsymbol{d}^{k^T}(-c - \boldsymbol{H}\boldsymbol{x}^k)}{\boldsymbol{d}^{k^T}\boldsymbol{H}\boldsymbol{d}^k} \\
&= -\frac{\boldsymbol{g}^{k^T}\boldsymbol{d}^k}{\boldsymbol{d}^{k^T}\boldsymbol{H}\boldsymbol{d}^k}
\end{aligned}
$$

Therefore,

$$
\alpha^k = -\frac{\boldsymbol{g}^{k^T}\boldsymbol{d}^k}{\boldsymbol{d}^{k^T}\boldsymbol{H}\boldsymbol{d}^k}
$$

Suppose $\{-\boldsymbol{g}^0, -\boldsymbol{g}^1, \ldots, -\boldsymbol{g}^{n-1}\}$ is a *linearly independent* set of vectors.

Use Gram-Schmidt procedure to determine the $\boldsymbol{H}$-conjugate vectors, $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$.

- Let $\boldsymbol{d}^0 = -\boldsymbol{g}^0$
- In general,

$$\boldsymbol{d}^k = -\boldsymbol{g}^k + \sum_{j=0}^{k-1} \beta^j \boldsymbol{d}^j, \quad k = 1, \ldots, n-1$$

But we want $\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{n-1}$ to be $\boldsymbol{H}$-conjugate vectors.

$$\begin{aligned}
\boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^k &= -\boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{g}^k + \sum_{j=0}^{k-1} \beta^j \boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^j, \qquad i = 0, \ldots, k-1 \\
\therefore 0 &= -\boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{g}^k + \beta^i \boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^i, \qquad i = 0, \ldots, k-1 \\
\therefore \beta^i &= \frac{\boldsymbol{g}^{k^T}\boldsymbol{H}\boldsymbol{d}^i}{\boldsymbol{d}^{i^T}\boldsymbol{H}\boldsymbol{d}^i}
\end{aligned}$$

$$\therefore \boldsymbol{d}^k = -\boldsymbol{g}^k + \sum_{j=0}^{k-1} \left(\frac{\boldsymbol{g}^{k^T} \boldsymbol{H} \boldsymbol{d}^j}{\boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j} \right) \boldsymbol{d}^j$$

We now need to show that $\{-\boldsymbol{g}^0, -\boldsymbol{g}^1, \ldots, -\boldsymbol{g}^{n-1}\}$ is a *linearly independent* set of vectors.
Note that
$\text{span}\{\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{k-1}\} = \text{span}\{-\boldsymbol{g}^0, -\boldsymbol{g}^1, \ldots, -\boldsymbol{g}^{k-1}\}$
We have already shown that

$$\{\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{k-1}\} \text{ are } \boldsymbol{H}\text{-conjugate} \Rightarrow \boldsymbol{g}^k \perp \mathcal{B}^k$$
$$\therefore -\boldsymbol{g}^k \perp \text{span}\{\boldsymbol{d}^0, \boldsymbol{d}^1, \ldots, \boldsymbol{d}^{k-1}\}$$
$$\therefore -\boldsymbol{g}^k \perp \text{span}\{-\boldsymbol{g}^0, -\boldsymbol{g}^1, \ldots, -\boldsymbol{g}^{k-1}\}$$

Therefore, $\{-\boldsymbol{g}^0, -\boldsymbol{g}^1, \ldots, -\boldsymbol{g}^{n-1}\}$ is a *linearly independent* set of vectors.

Now, consider

$$\begin{aligned} \boldsymbol{d}^0 &= -\boldsymbol{g}^0 \\ \boldsymbol{d}^k &= -\boldsymbol{g}^k + \sum_{j=0}^{k-1} \underbrace{\left(\frac{\boldsymbol{g}^{k^T} \boldsymbol{H} \boldsymbol{d}^j}{\boldsymbol{d}^{j^T} \boldsymbol{H} \boldsymbol{d}^j} \right)}_{\beta^j} \boldsymbol{d}^j \ \forall\, k = 1, \ldots, n-1 \end{aligned}$$

Note that $\boldsymbol{x}^{j+1} = \boldsymbol{x}^j + \alpha^j \boldsymbol{d}^j$ and $\boldsymbol{g}^{j+1} = \boldsymbol{g}^j + \alpha^j \boldsymbol{H}\boldsymbol{d}^j$.

Therefore,

$$\boldsymbol{H}\boldsymbol{d}^j = \frac{1}{\alpha^j}(\boldsymbol{g}^{j+1} - \boldsymbol{g}^j)$$

Thus,

$$\begin{aligned} \boldsymbol{d}^k &= -\boldsymbol{g}^k + \sum_{j=0}^{k-1} \left(\frac{\boldsymbol{g}^{k^T} (\boldsymbol{g}^{j+1} - \boldsymbol{g}^j)}{\boldsymbol{d}^{j^T} (\boldsymbol{g}^{j+1} - \boldsymbol{g}^j)} \right) \boldsymbol{d}^j \\ &= -\boldsymbol{g}^k + \left(\frac{\boldsymbol{g}^{k^T} \boldsymbol{g}^k}{\boldsymbol{d}^{k-1^T} (\boldsymbol{g}^k - \boldsymbol{g}^{k-1})} \right) \boldsymbol{d}^{k-1} \end{aligned}$$

$$\boldsymbol{d}^k = -\boldsymbol{g}^k + \left(\frac{\boldsymbol{g}^{k^T}\boldsymbol{g}^k}{\boldsymbol{d}^{k-1^T}(\boldsymbol{g}^k - \boldsymbol{g}^{k-1})} \right) \boldsymbol{d}^{k-1}$$

Due to exact line search, $\boldsymbol{g}^{k^T}\boldsymbol{d}^{k-1} = 0$.

$$\begin{aligned} \boldsymbol{d}^{k-1} &= -\boldsymbol{g}^{k-1} + \beta^{k-2}\boldsymbol{d}^{k-2} \\ -\boldsymbol{d}^{k-1^T}\boldsymbol{g}^{k-1} &= \boldsymbol{g}^{k-1^T}\boldsymbol{g}^{k-1} + \beta^{k-2}\boldsymbol{g}^{k-1^T}\boldsymbol{d}^{k-2} \end{aligned}$$

Therefore,

$$\boldsymbol{d}^k = -\boldsymbol{g}^k + \frac{\boldsymbol{g}^{k^T}\boldsymbol{g}^k}{\boldsymbol{g}^{k-1^T}\boldsymbol{g}^{k-1}}\boldsymbol{d}^{k-1}, \quad k = 1, \ldots, n-1$$

Fletcher-Reeves method

Conjugate Gradient Algorithm (Fletcher-Reeves)

For Quadratic function, $\frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}$, $\boldsymbol{H}$ symmetric positive definite

(1) Initialize $\boldsymbol{x}^0$, ϵ, $\boldsymbol{d}^0 = -\boldsymbol{g}^0$, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

(a) $\alpha^k = -\frac{{\boldsymbol{g}^k}^T \boldsymbol{d}^k}{{\boldsymbol{d}^k}^T \boldsymbol{H}\boldsymbol{d}^k}$

(b) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$

(c) $\boldsymbol{g}^{k+1} = \boldsymbol{H}\boldsymbol{x}^{k+1} + \boldsymbol{c}$

(d) $\beta^k = \frac{{\boldsymbol{g}^{k+1}}^T \boldsymbol{g}^{k+1}}{{\boldsymbol{g}^k}^T \boldsymbol{g}^k}$

(e) $\boldsymbol{d}^{k+1} = -\boldsymbol{g}^{k+1} + \beta^k \boldsymbol{d}^k$

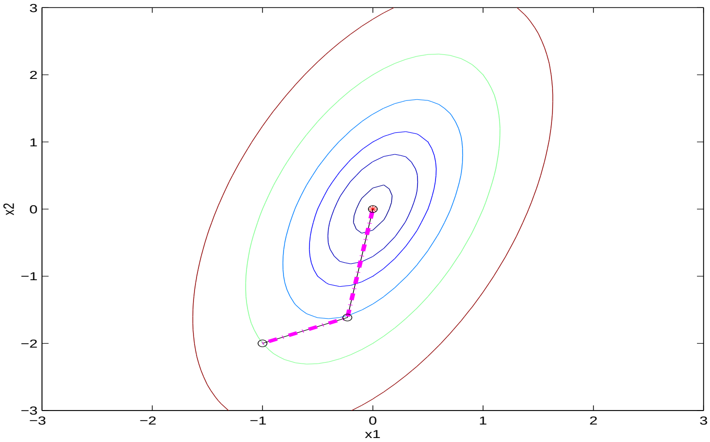
(f) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, global minimum of $f(\boldsymbol{x})$.

Example:

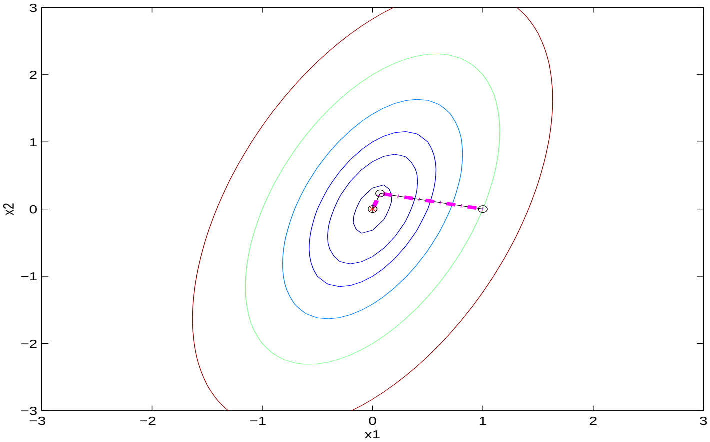
$$\min \quad f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$



Conjugate Gradient algorithm (Fletcher-Reeves) with exact line search applied to $f(\boldsymbol{x})$

Example:

$$\min \ f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$



Conjugate Gradient algorithm (Fletcher-Reeves) with exact line search applied to $f(\boldsymbol{x})$

Extension to Nonquadratic function, $f(\boldsymbol{x})$:

Conjugate Gradient Algorithm (Fletcher-Reeves)

(1) Initialize $\boldsymbol{x}^0$, ϵ, $\boldsymbol{d}^0 = -\boldsymbol{g}^0$, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

  (a) $\alpha^k = \arg\min_{\alpha>0} \; f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k)$

  (b) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$

  (c) Compute $\boldsymbol{g}^{k+1}$

  (d) **if** $k < n - 1$

   - $\beta^k = \frac{\boldsymbol{g}^{k+1^T}\boldsymbol{g}^{k+1}}{\boldsymbol{g}^{k^T}\boldsymbol{g}^k}$
   - $\boldsymbol{d}^{k+1} = -\boldsymbol{g}^{k+1} + \beta^k \boldsymbol{d}^k$
   - $k := k + 1$

  else

   - $\boldsymbol{x}^0 = \boldsymbol{x}^{k+1}$
   - $\boldsymbol{d}^0 = -\boldsymbol{g}^{k+1}$
   - $k := 0$

  endif

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.

β^k Determination

- Fletcher-Reeves method

$$\beta^k_{FR} = \frac{\boldsymbol{g}^{k^T}\boldsymbol{g}^k}{\boldsymbol{g}^{k-1^T}\boldsymbol{g}^{k-1}}$$

- Polak-Ribiere method

$$\beta^k_{PR} = \frac{\boldsymbol{g}^{k^T}(\boldsymbol{g}^k - \boldsymbol{g}^{k-1})}{\boldsymbol{g}^{k-1^T}\boldsymbol{g}^{k-1}}$$

- Hestenes-Steifel method

$$\beta^k_{HS} = \frac{\boldsymbol{g}^{k^T}(\boldsymbol{g}^k - \boldsymbol{g}^{k-1})}{(\boldsymbol{g}^k - \boldsymbol{g}^{k-1})^T\boldsymbol{d}^{k-1}}$$

$$\boldsymbol{B}^k_{BFGS} = \boldsymbol{B} + \left(1 + \frac{\gamma^T \boldsymbol{B}\gamma}{\boldsymbol{\delta}^T\gamma}\right)\frac{\boldsymbol{\delta}\boldsymbol{\delta}^T}{\boldsymbol{\delta}^T\gamma} - \left(\frac{\boldsymbol{\delta}\gamma^T\boldsymbol{B} + \boldsymbol{B}\gamma\boldsymbol{\delta}^T}{\boldsymbol{\delta}^T\gamma}\right)$$

Memoryless BFGS iteration

$$\boldsymbol{B}^k_{BFGS} = \boldsymbol{I} + \left(1 + \frac{\gamma^T\gamma}{\boldsymbol{\delta}^T\gamma}\right)\frac{\boldsymbol{\delta}\boldsymbol{\delta}^T}{\boldsymbol{\delta}^T\gamma} - \left(\frac{\boldsymbol{\delta}\gamma^T + \gamma\boldsymbol{\delta}^T}{\boldsymbol{\delta}^T\gamma}\right)$$

With exact line search, ${\boldsymbol{\delta}^{k-1}}^T\boldsymbol{g}^k = \alpha^{k-1}{\boldsymbol{d}^{k-1}}^T\boldsymbol{g}^k = 0$. Therefore,

$$\boldsymbol{d}^k_{BFGS} = -\boldsymbol{B}^k_{BFGS}\boldsymbol{g}^k = -\boldsymbol{g}^k + \frac{\boldsymbol{\delta}\gamma^T\boldsymbol{g}^k}{\boldsymbol{\delta}^T\gamma} = -\boldsymbol{g}^k + \underbrace{\frac{{\boldsymbol{g}^k}^T(\boldsymbol{g}^k - \boldsymbol{g}^{k-1})}{(\boldsymbol{g}^k - \boldsymbol{g}^{k-1})^T\boldsymbol{d}^{k-1}}}_{\beta^k_{HS}}\boldsymbol{d}^{k-1}$$

For nonquadratic function, $f(\boldsymbol{x})$:

Conjugate Gradient Algorithm (Fletcher-Reeves)

(1) Initialize $\boldsymbol{x}^0$, ϵ, $\boldsymbol{d}^0 = -\boldsymbol{g}^0$, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

(a) $\alpha^k = \arg\min_{\alpha>0} \; f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k)$

(b) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$

(c) Compute $\boldsymbol{g}^{k+1}$

(d) **if** $k < n - 1$

- $\beta^k = \frac{\boldsymbol{g}^{k+1^T} \boldsymbol{g}^{k+1}}{\boldsymbol{g}^{k^T} \boldsymbol{g}^k}$
- $\boldsymbol{d}^{k+1} = -\boldsymbol{g}^{k+1} + \beta^k \boldsymbol{d}^k$
- $k := k + 1$

else

- $\boldsymbol{x}^0 = \boldsymbol{x}^{k+1}$
- $\boldsymbol{d}^0 = -\boldsymbol{g}^{k+1}$
- $k := 0$

endif

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.