ECE/MATH 520, Spring 2008

Homework Problems 3

Solutions (version: March 6, 2008, 17:42)

10.1 Let $\boldsymbol{Q}$ be a real symmetric positive definite $n \times n$ matrix. Given an arbitrary set of linearly independent vectors $\{\boldsymbol{p}^{(0)}, \ldots, \boldsymbol{p}^{(n-1)}\}$ in $\mathbb{R}^n$, the *Gram-Schmidt* procedure generates a set of vectors $\{\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(n-1)}\}$ as follows:

$$\begin{aligned} \boldsymbol{d}^{(0)} &= \boldsymbol{p}^{(0)} \\ \boldsymbol{d}^{(k+1)} &= \boldsymbol{p}^{(k+1)} - \sum_{i=0}^{k} \frac{\boldsymbol{p}^{(k+1)T}\boldsymbol{Q}\boldsymbol{d}^{(i)}}{\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(i)}}\boldsymbol{d}^{(i)} \end{aligned}$$

Show that the vectors $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(n-1)}$ are $\boldsymbol{Q}$-conjugate.

Ans.: We proceed by induction to show that for $k = 0, \ldots, n-1$, the set $\{\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k)}\}$ is $\boldsymbol{Q}$-conjugate. We assume that $\boldsymbol{d}^{(i)} \neq \boldsymbol{0}$, $i = 1, \ldots, k$, so that $\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(i)} \neq 0$ and the algorithm is well defined.

For $k = 0$, the statement trivially holds. So, assume that the statement is true for $k < n-1$, i.e., $\{\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k)}\}$ is $\boldsymbol{Q}$-conjugate. We now show that $\{\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k+1)}\}$ is $\boldsymbol{Q}$-conjugate. For this, we need only to show that for each $j = 0, \ldots, k$, we have $\boldsymbol{d}^{(k+1)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = 0$. To this end,

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

By the induction hypothesis, $\boldsymbol{d}^{(i)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = 0$ for $i \neq j$. Therefore

$$\boldsymbol{d}^{(k+1)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = \boldsymbol{p}^{(k+1)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} - \frac{\boldsymbol{p}^{(k+1)T}\boldsymbol{Q}\boldsymbol{d}^{(j)}}{\boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(j)}}\boldsymbol{d}^{(j)T}\boldsymbol{Q}\boldsymbol{d}^{(j)} = 0.$$

In the above, we have assumed that the vectors $\boldsymbol{d}^{(k)}$ are nonzero (so that $\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)} \neq 0$ and the algorithm is well defined). To prove that this assumption holds, we use induction to show that $\boldsymbol{d}^{(k)}$ is a (nonzero) linear combination of $\boldsymbol{p}^{(0)}, \ldots, \boldsymbol{p}^{(k)}$ (which immediately implies that $\boldsymbol{d}^{(k)}$ is nonzero because of the linear independence of $\boldsymbol{p}^{(0)}, \ldots, \boldsymbol{p}^{(k)}$).

For $k = 0$, we have $\boldsymbol{d}^{(0)} = \boldsymbol{p}^{(0)}$ by definition. Assume that the result holds for $k < n-1$; i.e., $\boldsymbol{d}^{(k)} = \sum_{j=0}^{k} \alpha_j^{(k)} \boldsymbol{p}^{(j)}$, where the coefficients $\alpha_j^{(k)}$ are not all zero. Consider $\boldsymbol{d}^{(k+1)}$:

$$\begin{aligned}
\boldsymbol{d}^{(k+1)} &= \boldsymbol{p}^{(k+1)} - \sum_{i=0}^{k} \beta_i \boldsymbol{d}^{(i)} \\
&= \boldsymbol{p}^{(k+1)} - \sum_{i=0}^{k} \beta_i \sum_{j=0}^{i} \alpha_j^{(k)} \boldsymbol{p}^{(j)} \\
&= \boldsymbol{p}^{(k+1)} - \sum_{j=0}^{k} \sum_{i=j}^{k} \beta_i \alpha_j^{(k)} \boldsymbol{p}^{(j)}.
\end{aligned}$$

So, clearly $\boldsymbol{d}^{(k+1)}$ is a nonzero linear combination of $\boldsymbol{p}^{(0)}, \ldots, \boldsymbol{p}^{(k+1)}$.

10.6 Consider the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} - \boldsymbol{x}^T \boldsymbol{b},$$

where $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$. Let $\boldsymbol{D} \in \mathbb{R}^{n \times r}$ be of rank r, and $\boldsymbol{x}_0 \in \mathbb{R}^n$. Define the function $\phi : \mathbb{R}^r \to \mathbb{R}$ by

$$\phi(\boldsymbol{a}) = f(\boldsymbol{x}_0 + \boldsymbol{D}\boldsymbol{a}).$$

Show that ϕ is a quadratic function with a positive definite quadratic term.

Ans.: Expanding $\phi(\boldsymbol{a})$ yields

$$\begin{aligned}
\phi(\boldsymbol{a}) &= \frac{1}{2}(\boldsymbol{x}_0 + \boldsymbol{D}\boldsymbol{a})^T \boldsymbol{Q} (\boldsymbol{x}_0 + \boldsymbol{D}\boldsymbol{a}) - (\boldsymbol{x}_0 + \boldsymbol{D}\boldsymbol{a})^T \boldsymbol{b} \\
&= \frac{1}{2}\boldsymbol{a}^T \left(\boldsymbol{D}^T \boldsymbol{Q} \boldsymbol{D}\right) \boldsymbol{a} + \boldsymbol{a}^T \left(\boldsymbol{D}^T \boldsymbol{Q} \boldsymbol{x}_0 - \boldsymbol{D}^T \boldsymbol{b}\right) + \left(\frac{1}{2}\boldsymbol{x}_0^T \boldsymbol{Q} \boldsymbol{x}_0 - \boldsymbol{x}_0^T \boldsymbol{b}\right).
\end{aligned}$$

Clearly ϕ is a quadratic function on $\mathbb{R}^r$. It remains to show that the matrix in the quadratic term, $\boldsymbol{D}^T \boldsymbol{Q} \boldsymbol{D}$, is positive definite. Since $\boldsymbol{Q} > 0$, for any $\boldsymbol{a} \in \mathbb{R}^r$, we have

$$\boldsymbol{a}^T \left(\boldsymbol{D}^T \boldsymbol{Q} \boldsymbol{D}\right) \boldsymbol{a} = (\boldsymbol{D}\boldsymbol{a})^T \boldsymbol{Q} (\boldsymbol{D}\boldsymbol{a}) \geq 0$$

and

$$\boldsymbol{a}^T \left(\boldsymbol{D}^T \boldsymbol{Q} \boldsymbol{D}\right) \boldsymbol{a} = (\boldsymbol{D}\boldsymbol{a})^T \boldsymbol{Q} (\boldsymbol{D}\boldsymbol{a}) = 0$$

if and only if $\boldsymbol{D}\boldsymbol{a} = \boldsymbol{0}$. Since rank $\boldsymbol{D} = r$, $\boldsymbol{D}\boldsymbol{a} = \boldsymbol{0}$ if and only if $\boldsymbol{a} = \boldsymbol{0}$. Hence, the matrix $\boldsymbol{D}^T \boldsymbol{Q} \boldsymbol{D}$ is positive definite.

10.7 Let $f(\boldsymbol{x})$, $\boldsymbol{x} = [x_1, x_2]^T \in \mathbb{R}^2$, be given by

$$f(\boldsymbol{x}) = \frac{5}{2}x_1^2 + \frac{1}{2}x_2^2 + 2x_1 x_2 - 3x_1 - x_2$$

a. Express $f(\boldsymbol{x})$ in the form of $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$.

b. Find the minimizer of f using the conjugate gradient algorithm. Use a starting point of $\boldsymbol{x}^{(0)} = [0,0]^T$.

c. Calculate the minimizer of f analytically from $\boldsymbol{Q}$ and $\boldsymbol{b}$, and check it with your answer in part b.

Ans.: a. We have $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$ where

$$\boldsymbol{Q} = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

b. Since f is a quadratic function on $\mathbb{R}^2$, we need to perform only two iterations. For the first iteration we compute

$$\begin{aligned} \boldsymbol{d}^{(0)} &= -\boldsymbol{g}^{(0)} = [3,1]^T \\ \alpha_0 &= \frac{5}{29} \\ \boldsymbol{x}^{(1)} &= [0.51724, 0.17241]^T \\ \boldsymbol{g}^{(1)} &= [-0.06897, 0.20690]^T. \end{aligned}$$

For the second iteration we compute

$$\begin{aligned} \beta_0 &= 0.0047534 \\ \boldsymbol{d}^{(1)} &= [0.08324, -0.20214]^T \\ \alpha_1 &= 5.7952 \\ \boldsymbol{x}^{(2)} &= [1.000, -1.000]^T. \end{aligned}$$

c. The minimizer is given by $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{b} = [1,-1]^T$, which agrees with part b.

11.1 Given $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, consider the algorithm

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

where $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \ldots$ are vectors in $\mathbb{R}^n$, and $\alpha_k \geq 0$ is chosen to minimize $f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)})$; that is,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)}).$$

Note that the above general algorithm encompasses almost all algorithms that we discussed in this part, including the steepest descent, Newton, conjugate gradient, and quasi-Newton algorithms.

Let $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$, and assume that $\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k)} < 0$.

a. Show that $\boldsymbol{d}^{(k)}$ is a descent direction for f, in the sense that there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$,

$$f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) < f(\boldsymbol{x}^{(k)}).$$

b. Show that $\alpha_k > 0$.

c. Show that $\boldsymbol{d}^{(k)T} \boldsymbol{g}^{(k+1)} = 0$.

d. Show that the following algorithms all satisfy the condition $\boldsymbol{d}^{(k)T} \boldsymbol{g}^{(k)} < 0$, if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$:

1. Steepest descent algorithm
2. Newton's method, assuming the Hessian is positive definite
3. Conjugate gradient algorithm
4. Quasi-Newton algorithm, assuming $\boldsymbol{H}_k > 0$

e. For the case where $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} - \boldsymbol{x}^T \boldsymbol{b}$, with $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$, derive an expression for α_k in terms of $\boldsymbol{Q}$, $\boldsymbol{d}^{(k)}$, and $\boldsymbol{g}^{(k)}$.

Ans.: a. Let

$$\phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}).$$

Then, using the chain rule, we obtain

$$\phi'(\alpha) = \boldsymbol{d}^{(k)T} \nabla f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}).$$

Hence

$$\phi'(0) = \boldsymbol{d}^{(k)T} \boldsymbol{g}^{(k)}.$$

Since ϕ' is continuous, then, if $\boldsymbol{d}^{(k)T} \boldsymbol{g}^{(k)} < 0$, there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$, $\phi(\alpha) < \phi(0)$, i.e., $f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) < f(\boldsymbol{x}^{(k)})$.

b. By part a, $\phi(\alpha) < \phi(0)$ for all $\alpha \in (0, \bar{\alpha}]$. Hence,

$$\alpha_k = \arg\min_{\alpha \geq 0} \phi(\alpha) \neq 0$$

which implies that $\alpha_k > 0$.

c. Now,

$$\boldsymbol{d}^{(k)T} \boldsymbol{g}^{(k+1)} = \boldsymbol{d}^{(k)T} \nabla f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}) = \phi_k'(\alpha_k).$$

Since $\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) > 0$, we have $\phi_k'(\alpha_k) = 0$. Hence, $\boldsymbol{g}^{(k+1)T} \boldsymbol{d}^{(k)} = 0$.

d.

i. We have $\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)}$. Hence, $\boldsymbol{d}^{(k)T} \boldsymbol{g}^{(k)} = -\|\boldsymbol{g}^{(k)}\|^2$. If $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$, then $\|\boldsymbol{g}^{(k)}\|^2 > 0$, and hence $\boldsymbol{d}^{(k)T} \boldsymbol{g}^{(k)} < 0$.

ii. We have $\boldsymbol{d}^{(k)} = -\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)}$. Since $\boldsymbol{F}(\boldsymbol{x}^{(k)}) > 0$, we also have $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1} > 0$. Therefore, $\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k)} = -\boldsymbol{g}^{(k)T}\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)} < 0$ if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$.

iii. We have

$$\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)} + \beta_{k-1}\boldsymbol{d}^{(k-1)}.$$

Hence,

$$\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k)} = -\|\boldsymbol{g}^{(k)}\|^2 + \beta_{k-1}\boldsymbol{d}^{(k-1)T}\boldsymbol{g}^{(k)}.$$

By part c, $\boldsymbol{d}^{(k-1)T}\boldsymbol{g}^{(k)} = 0$. Hence, if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$, then $\|\boldsymbol{g}^{(k)}\|^2 > 0$, and

$$\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k)} = -\|\boldsymbol{g}^{(k)}\|^2 < 0.$$

iv. We have $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k\boldsymbol{g}^{(k)}$. Therefore, if $\boldsymbol{H}_k > 0$ and $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$, then $\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k)} = -\boldsymbol{g}^{(k)T}\boldsymbol{H}_k\boldsymbol{g}^{(k)} < 0$.

e. Using the equation $\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}$, we get

$$\begin{aligned}
\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k+1)} &= \boldsymbol{d}^{(k)T}(\boldsymbol{Q}\boldsymbol{x}^{(k+1)} - \boldsymbol{b}) \\
&= \boldsymbol{d}^{(k)T}(\boldsymbol{Q}(\boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}) - \boldsymbol{b}) \\
&= \alpha_k\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)} + \boldsymbol{d}^{(k)T}(\boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{b}) \\
&= \alpha_k\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)} + \boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k)}.
\end{aligned}$$

By part c, $\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k+1)} = 0$, which implies

$$\alpha_k = -\frac{\boldsymbol{d}^{(k)T}\boldsymbol{g}^{(k)}}{\boldsymbol{d}^{(k)T}\boldsymbol{Q}\boldsymbol{d}^{(k)}}.$$

11.6 Assuming exact line search, show that if $\boldsymbol{H}_0 = \boldsymbol{I}_n$ ($n \times n$ identity matrix), then the first two steps of the BFGS algorithm yield the same points $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ as conjugate gradient algorithms with the Hestenes-Stiefel, the Polak-Ribiere, as well as the Fletcher-Reeves formulas.

Ans.: The first step for both algorithms is clearly the same, since in either case we have

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \alpha_0\boldsymbol{g}^{(0)}.$$

For the second step,

$$\begin{aligned}
\boldsymbol{d}^{(1)} &= -\boldsymbol{H}_1\boldsymbol{g}^{(1)} \\
&= -\left(\boldsymbol{I}_n + \left(1 + \frac{\Delta\boldsymbol{g}^{(0)T}\Delta\boldsymbol{g}^{(0)}}{\Delta\boldsymbol{g}^{(0)T}\Delta\boldsymbol{x}^{(0)}}\right)\frac{\Delta\boldsymbol{x}^{(0)}\Delta\boldsymbol{x}^{(0)T}}{\Delta\boldsymbol{x}^{(0)T}\Delta\boldsymbol{g}^{(0)}}\right. \\
&\qquad \left. - \frac{\Delta\boldsymbol{g}^{(0)}\Delta\boldsymbol{x}^{(0)T} + (\Delta\boldsymbol{g}^{(0)}\Delta\boldsymbol{x}^{(0)T})^T}{\Delta\boldsymbol{g}^{(0)T}\Delta\boldsymbol{x}^{(0)}}\right)\boldsymbol{g}^{(1)}
\end{aligned}$$

$$
\begin{aligned}
&= -\boldsymbol{g}^{(1)} - \left(1 + \frac{\Delta \boldsymbol{g}^{(0)T} \Delta \boldsymbol{g}^{(0)}}{\Delta \boldsymbol{g}^{(0)T} \Delta \boldsymbol{x}^{(0)}}\right) \frac{\Delta \boldsymbol{x}^{(0)} \Delta \boldsymbol{x}^{(0)T} \boldsymbol{g}^{(1)}}{\Delta \boldsymbol{x}^{(0)T} \Delta \boldsymbol{g}^{(0)}} \\
&\qquad + \frac{\Delta \boldsymbol{g}^{(0)} \Delta \boldsymbol{x}^{(0)T} \boldsymbol{g}^{(1)} + \Delta \boldsymbol{x}^{(0)} \Delta \boldsymbol{g}^{(0)T} \boldsymbol{g}^{(1)}}{\Delta \boldsymbol{g}^{(0)T} \Delta \boldsymbol{x}^{(0)}}.
\end{aligned}
$$

Since the line search is exact, we have

$$\Delta \boldsymbol{x}^{(0)T} \boldsymbol{g}^{(1)} = \alpha_0 \boldsymbol{d}^{(0)T} \boldsymbol{g}^{(1)} = 0.$$

Hence,

$$
\begin{aligned}
\boldsymbol{d}^{(1)} &= -\boldsymbol{g}^{(1)} + \left(\frac{\Delta \boldsymbol{g}^{(0)T} \boldsymbol{g}^{(1)}}{\Delta \boldsymbol{g}^{(0)T} \Delta \boldsymbol{x}^{(0)}}\right) \Delta \boldsymbol{x}^{(0)} \\
&= -\boldsymbol{g}^{(1)} + \left(\frac{\boldsymbol{g}^{(1)T} \Delta \boldsymbol{g}^{(0)}}{\Delta \boldsymbol{g}^{(0)T} \boldsymbol{d}^{(0)}}\right) \boldsymbol{d}^{(0)} \\
&= -\boldsymbol{g}^{(1)} + \beta_0 \boldsymbol{d}^{(0)}
\end{aligned}
$$

where

$$\beta_0 = \frac{\boldsymbol{g}^{(1)T} \Delta \boldsymbol{g}^{(0)}}{\boldsymbol{d}^{(0)T} \Delta \boldsymbol{g}^{(0)}} = \frac{\boldsymbol{g}^{(1)T} (\boldsymbol{g}^{(1)} - \boldsymbol{g}^{(0)})}{\boldsymbol{d}^{(0)T} (\boldsymbol{g}^{(1)} - \boldsymbol{g}^{(0)})}$$

is the Hestenes-Stiefel update formula for β_0. Since $\boldsymbol{d}^{(0)} = -\boldsymbol{g}^{(0)}$, and $\boldsymbol{g}^{(1)T} \boldsymbol{g}^{(0)} = 0$, we have

$$\beta_0 = \frac{\boldsymbol{g}^{(1)T} (\boldsymbol{g}^{(1)} - \boldsymbol{g}^{(0)})}{\boldsymbol{g}^{(0)T} \boldsymbol{g}^{(0)}},$$

which is the Polak-Ribiere formula. Applying $\boldsymbol{g}^{(1)T} \boldsymbol{g}^{(0)} = 0$ again, we get

$$\beta_0 = \frac{\boldsymbol{g}^{(1)T} \boldsymbol{g}^{(1)}}{\boldsymbol{g}^{(0)T} \boldsymbol{g}^{(0)}},$$

which is the Fletcher-Reeves formula.