Remarks on the Davidon-Fletcher-Powell Method

This is one of the earliest and most effective methods quasi-Newton methods. Its effectiveness stems from the fact that, for quadratic problems, the method simultaneouly generates conjugate directions and constructs an approximation of the inverse of the Hessian matrix.

In each step, the inverse of the Hessian is updated by the sum of two symmetric rank 1 matrices. For this reason, it is referred to as a "rank 2" correction procedure. It is also called a "variable metric method".

The Davidon-Fletcher-Powell Method can be described as follows:

(0) Initial Step:

Select a positive definite matrix S_o and an initial estimate $\boldsymbol{x}_o$ of $\boldsymbol{x}^*$. Set $k = 0$

(1) Iterative Step:

(a) Compute $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$.

(b) Compute $\boldsymbol{d}_k = -S_k\, \boldsymbol{g}_k$.

(c) Compute $\alpha_k = \text{ argmin } f(\boldsymbol{x}_k + \alpha\, \boldsymbol{d}_k)$.

(d) Set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$.

(2) Update Step:

(a) $p_k = \alpha_k\, \boldsymbol{d}_k; \boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1})$.

(b) $\boldsymbol{q}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$.

(c) $S_{k+1} = S_k + \frac{\boldsymbol{p}_k\, \boldsymbol{p}_k^\top}{\langle \boldsymbol{p}_k, \boldsymbol{q}_k \rangle} - \frac{S_k \boldsymbol{q}_k\, \boldsymbol{q}^\top\, S_k}{\langle \boldsymbol{q}_k, S_k \boldsymbol{q}_k \rangle}$

(3) Replace k by $k+1$; go to 1(a).

Let us check that if the matrix S_k is positive definite, then so is S_{k+1}. From the definition

$$\begin{aligned}
\langle \boldsymbol{x}, S_{k+1}\boldsymbol{x} \rangle &= \langle \boldsymbol{x}, S_k \boldsymbol{x} \rangle + \frac{\langle \boldsymbol{x}, \boldsymbol{p}_k \boldsymbol{p}_k^\top \boldsymbol{x} \rangle}{\langle \boldsymbol{p}_k, \boldsymbol{q}_k \rangle} - \frac{\langle \boldsymbol{x}, S_k \boldsymbol{q}_k\, \boldsymbol{q}^\top\, S_k\, \boldsymbol{x} \rangle}{\langle \boldsymbol{q}_k, S_k \boldsymbol{q}_k \rangle} \\
&= \langle \boldsymbol{x}, S_k \boldsymbol{x} \rangle + \frac{\langle \boldsymbol{p}_k, \boldsymbol{x} \rangle^2}{\langle \boldsymbol{p}_k, \boldsymbol{q}_k \rangle} - \frac{\langle \boldsymbol{x}, S_k \boldsymbol{q}_k \rangle^2}{\langle \boldsymbol{q}_k, S_k \boldsymbol{q}_k \rangle}
\end{aligned}$$

Now, set $\boldsymbol{a} = S_k^{1/2} \boldsymbol{x}$ and $\mathbf{b} = S_k^{1/2} \boldsymbol{q}_k$, and rewrite:

$$
\begin{aligned}
\langle \boldsymbol{x}, S_{k+1}\boldsymbol{x}\rangle &= \langle \boldsymbol{a}, \boldsymbol{a}\rangle - \frac{\langle \boldsymbol{a}, \mathbf{b}\rangle^2}{\langle \mathbf{b}, \mathbf{b}\rangle} + \frac{\langle \boldsymbol{p}_k, \boldsymbol{x}\rangle^2}{\langle \boldsymbol{p}_k, \boldsymbol{q}_k\rangle} \\
&= \frac{\langle \boldsymbol{a}, \boldsymbol{a}\rangle \langle \mathbf{b}, \mathbf{b}\rangle - \langle \boldsymbol{a}, \mathbf{b}\rangle^2}{\langle \mathbf{b}, \mathbf{b}\rangle} + \frac{\langle \boldsymbol{p}_k, \boldsymbol{x}\rangle^2}{\langle \boldsymbol{p}_k, \boldsymbol{q}_k\rangle} \,.
\end{aligned}
$$

Now

$$
\langle \boldsymbol{p}_k, \boldsymbol{q}_k\rangle \;=\; \langle \boldsymbol{p}_k, \boldsymbol{g}_{k+1} - \boldsymbol{g}_k\rangle \;=\; \langle \boldsymbol{p}_k, \boldsymbol{g}_{k+1}\rangle - \langle \boldsymbol{p}_k, \boldsymbol{g}_k\rangle \,.
$$

Now $0 = \partial f/\partial\alpha|_{\alpha_k} = \langle \boldsymbol{d}_k, \nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k\rangle = \langle \boldsymbol{d}_k, \boldsymbol{g}_{k+1}\rangle$. Hence $\langle \boldsymbol{p}_k, \boldsymbol{g}_{k+1}\rangle = 0$ and so

$$
\begin{aligned}
\langle \boldsymbol{p}_{,}\boldsymbol{q}_k\rangle &= -\langle \boldsymbol{p}_k, \boldsymbol{g}_k\rangle \;=\; -\langle \alpha_k \boldsymbol{g}_k, \boldsymbol{g}_k\rangle \\
&= -\langle -\alpha_k S_k \boldsymbol{g}_k, \boldsymbol{g}_k\rangle \;=\; \alpha_k \langle \boldsymbol{g}_k, S_k \boldsymbol{g}_k\rangle \,.
\end{aligned}
$$

Putting these results together, we have

$$
\langle \boldsymbol{x}, S_{k+1}\boldsymbol{x}\rangle \;=\; \frac{\langle \boldsymbol{a}, \boldsymbol{a}\rangle \langle \mathbf{b}, \mathbf{b}\rangle - \langle \boldsymbol{a}, \mathbf{b}\rangle^2}{\langle \mathbf{b}, \mathbf{b}\rangle} + \frac{\langle \boldsymbol{x}, \boldsymbol{p}_k\rangle^2}{\alpha_k \langle \boldsymbol{g}_k, S\boldsymbol{g}_k\rangle}
$$

Now, from the Cauchy-Schwarz inequality $\langle \boldsymbol{a}, \mathbf{b}\rangle^2 \le \|\boldsymbol{a}\|^2 \|\mathbf{b}\|^2$. Thus the first term ont the right is non-negative and the second is as well. Can they be zero simultaneously? We get equality in the Cauchy-Schwarz inequality if and only if $\boldsymbol{a} = \beta\,\mathbf{b}$. If $S_k^{1/2}\boldsymbol{x} = \beta\, S_k^{1/2}\boldsymbol{q}_k$ then $\boldsymbol{x} = \beta\,\boldsymbol{q}_k$ since $S_k^{1/2}$ is non-singular. Then

$$
\langle \boldsymbol{p}_k, \boldsymbol{x}\rangle \;=\; \beta \langle \boldsymbol{p}_k, \boldsymbol{q}_k\rangle \;=\; \beta\,\alpha_k \langle \boldsymbol{g}_k, S_k \boldsymbol{g}_k\rangle > 0 \,.
$$

Hence, for all $\boldsymbol{x}$, $\langle \boldsymbol{x}, S_{k+1}\boldsymbol{x}\rangle > 0$ which was to be proved. $\square$