CS 726 Assignment 6

Ruochen Lin

April 12, 2018

1

Without loss of generality, we write the quadratic function as $f(x) = \frac{1}{2}x^T Ax - b^T x + c$, and its gradient is $\nabla f(x) = Ax - b$. It's obvious that, given the same starting point, Fletcher-Reeves CG reduces to linear CG on $f(x)$, because it simply replaced the r_ks in linear CG with ∇f_ks.

In addition, if starting from the same point, the linear, FR, PR, and HS flavours of CG will all generate the same first step, with exact line search along the negative gradient direction. This is to say that

$$\begin{aligned} x_1^{FR} = x_1^{PR} = x_1^{HS} = x_1, \\ r_1^{FR} = r_1^{PR} = r_1^{HS} = r_1 = \nabla f_1 = Ax_1 - b. \end{aligned}$$

Then, if we apply the properties of linear CG up to the first step, together with the fact that $p_0 = -r_0$we have

$$\begin{aligned} \beta_1^{FR} &= \frac{\|\nabla f_1\|^2}{\|\nabla f_0\|^2} = \frac{\|r_1\|^2}{\|r_0\|^2} = \beta_1, \\ \beta_1^{PR} &= \frac{\nabla f_1^T(\nabla f_1 - \nabla f_0)}{\|\nabla f_0\|^2} = \frac{r_1^T r_1 - r_1^T r_0}{\|r_0\|^2} = \frac{\|r_1\|^2}{\|r_0\|^2} \\ &= \beta_1, \\ \beta_1^{HS} &= \frac{\nabla f_1^T(\nabla f_1 - \nabla f_0)}{(\nabla f_1 - \nabla f_0)^T p_0} = \frac{r_1^T r_1 - r_1^T r_0}{-r_1^T r_0 + r_0^T r_0} = \frac{\|r_1\|^2}{\|r_0\|^2} \\ &= \beta_1, \\ p_1^{FR} &= p_1^{PR} = p_1^{HS} = -\nabla f_1 + \beta_1 p_0 \\ &= p_1, \\ \implies x_2^{FR} &= x_2^{PR} = x_2^{HS} = x_2 = x_1 + \alpha_1 p_1. \end{aligned}$$

The last equation is valid because x_2 is obtained from doing exact line search from x_1 along p_1, and it should lead to the same x_2s if we have the same $x_1$and p_1s in all algorithms. Thus we have shown that all three nonlinear CG algoithms reduces to linear CG on $f(x)$ in the first two steps.

In addition, if all four algorithm lead to the same result after $k-1$ iterations, then if we replace all the 0s with $k-1$s in the previous analysis, and 1s with ks and 2s with $k+1$s, it would still be valid[1]; hence, we have proved that, if linear, FR, PR, and HS CG algorithms produce the same result at the $k-1$th iteration, their outcomes will still be the same at the kth step. Since our base case of $k=1$ is valid, we have proved inductively that FR, PR, and HS CG all reduce to linear CG for a quadratic function.

2

Given $H_k B_k = B_k H_k = I$ and $\rho_k = (y_k^T s_k)^{-1}$, together with the updating formulae

$$\begin{aligned} H_{k+1} &= (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T, \\ B_{k+1} &= b_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \rho_k y_k y_k^T, \end{aligned}$$

[1]When we're calculating the denominator of β_k^{HS}, although $p_k = -r_k$ is no longer valid with $k > 0$, we can use $p_{k-1} = -r_{k-1} + \beta_{k-1} p_{k-2}$ and the fact that $r_j^T p_{j-1},\ j = 1, 2, \cdots, k$ to get the same result.

we have

$$
\begin{aligned}
H_{k+1}B_{k+1} =& H_kB_k - \rho_k s_k y_k^T H_k B_k - \rho_k H_k y_k s_k^T B_k + \rho_k^2 s_k y_k^T H_k y_k s_k^T B_k \\
&+ \rho_k s_k s_k^T B_k - \frac{H_k B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \rho_k \frac{s_k y_k^T H_k B_k s_k s_k^T B_k}{s_k^T B_k s_k} \\
&+ \rho_k \frac{H_k y_k s_k^T B_k s_k s_k^T B_k}{s_k^T B_k s_k} - \rho_k^2 \frac{s_k y_k^T H_k y_k s_k^T B_k s_k s_k^T B_k}{s_k B_k s_k} \\
&- \rho_k \frac{s_k s_k^T B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \rho_k H_k y_k y_k^T - \rho_k^2 s_k y_k^T H_k y_k y_k^T \\
&- \rho_k^2 H_k y_k s_k^T y_k y_k^T + \rho_k^3 s_k y_k^T H_k y_k s_k^T y_k y_k^T + \rho_k^2 s_k s_k^T y_k y_k^T \\
=& I + (\rho_k s_k y_k^T - \rho_k s_k y_k^T) + (\rho_k H_k y_k s_k^T B_k - \rho_k H_k y_k s_k^T B_k) \\
&+ \Big(\rho_k^2 (y_k^T H_k y_k) s_k s_k^T B_k - \rho_k^2 (y_k^T H_k y_k) s_k s_k^T B_k\Big) \\
&+ (\rho_k s_k s_k^T B_k - \rho_k s_k s_k^T B_k) + \Big(\frac{s_k s_k^T B_k}{s_k^T B_k s_k} - \frac{s_k s_k^T B_k}{s_k^T B_k s_k}\Big) \\
&+ (\rho_k H_k y_k y_k^T - \rho_k H_k y_k y_k^T) \\
&+ \Big(\rho_k^2 (y_k^T H_k y_k) s_k y_k^T - \rho_k^2 (y_k^T H_k y_k) s_k y_k^T\Big) \\
=& I.
\end{aligned}
$$

Thus, we have proved that B_{k+1} and H_{k+1} are inverses of each other, given that B_k and H_k are as well.

3

Given that $H_k = B_k^{-1}$ and $H_k = H_k^T$, $B_k = B_k^T$, together with the updating formula

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

and the SM formula

$$(A + ab^T)^{-1} = A^{-1} - \frac{A^{-1} a b^T a^{-1}}{1 + b^T A^{-1} a},$$

by setting $A = B_k$, $a = \frac{y_k - B_k s_k}{(y_k - B_k s_k)^T s_k}$ and $b = y_k - B_k s_k$, we have

$$
\begin{aligned}
H_{k+1} = B_{k+1}^{-1} &= B_k^{-1} - \frac{[(y_k - B_k s_k)^T s_k]^{-1} B_k^{-1} (y_k - B_k s_k)(y_k - B_k s_k)^T B_k^{-1}}{1 + [(y_k - B_k s_k)^T s_k]^{-1} (y_k - B_k s_k)^T B_k^{-1} (y_k - B_k s_k)} \\
&= H_k - \frac{H_k (y_k - B_k s_k)(y_k - B_k s_k)^T H_k}{(y_k - B_k s_k)^T s_k + (y_k - B_k s_k)^T H_k (y_k - B_k s_k)} \\
&= H_k - \frac{(H_k y_k - s_k)(H_k y_k - s_k)^T}{y_k^T s_k - s_k^T B_k s_k + y_k^T H_k y_k - s_k^T y_k - y_k^T s_k + s_k^T B_k s_k} \\
&= H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k}.
\end{aligned}
$$