Section 5.2 – Nonlinear Conjugate Gradient Methods

AMS 230 Numerical Optimization

Qi Gong
Dept. of Applied Math & Stats.
Baskin School of Engineering
University of California
Santa Cruz, CA

Nonlinear Conjugate Gradient

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^\intercal Ax - b^\intercal x$$

Algorithm 5.2 (CG).

Given x_0;

Set $r_0 \leftarrow Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$;

while $r_k \neq 0$

$$\alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k};$$

$$x_{k+1} \leftarrow x_k + \alpha_k p_k;$$

$$r_{k+1} \leftarrow r_k + \alpha_k A p_k;$$

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};$$

$$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k;$$

$$k \leftarrow k+1;$$

end (while)

Linear CG

- Replace r by the gradient
- Replace exact line search by inexact line search

$$\min_{x\in\mathbb{R}^n} f(x)$$

Algorithm 5.4 (FR).

Given x_0;

Evaluate $f_0 = f(x_0)$, $\nabla f_0 = \nabla f(x_0)$;

Set $p_0 \leftarrow -\nabla f_0$, $k \leftarrow 0$;

while $\nabla f_k \neq 0$

Compute α_k and set $x_{k+1} = x_k + \alpha_k p_k$;

(based on Wolfe conditions)

Evaluate ∇f_{k+1};

$$\beta_{k+1}^{\mathrm{FR}} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k};$$

$$p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_{k+1}^{\mathrm{FR}} p_k;$$

$$k \leftarrow k+1;$$

end (while)

Nonlinear CG

Fletcher-Reeves Method

Convergence of F-R

Question: is p_k always a descent direction?

$$\beta_{k+1}^{\mathrm{FR}} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}; \qquad p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_{k+1}^{\mathrm{FR}} p_k;$$

$$\Longrightarrow \quad \nabla f_k^T p_k = -\|\nabla f_k\|^2 + \beta_k^{\mathrm{FR}} \nabla f_k^T p_{k-1} \quad < 0\ ?$$

Lemma 5.6.

Suppose that Algorithm 5.4 is implemented with a step length α_k that satisfies the strong Wolfe conditions (5.43) with $\boxed{0 < c_2 < \frac{1}{2}.}$ *Then the method generates descent directions p_k that satisfy the following inequalities:*

$$-\frac{1}{1-c_2} \le \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \boxed{\frac{2c_2 - 1}{1 - c_2},} \qquad \text{for all } k = 0, 1, \ldots. \tag{5.53}$$

Global Convergence of FR

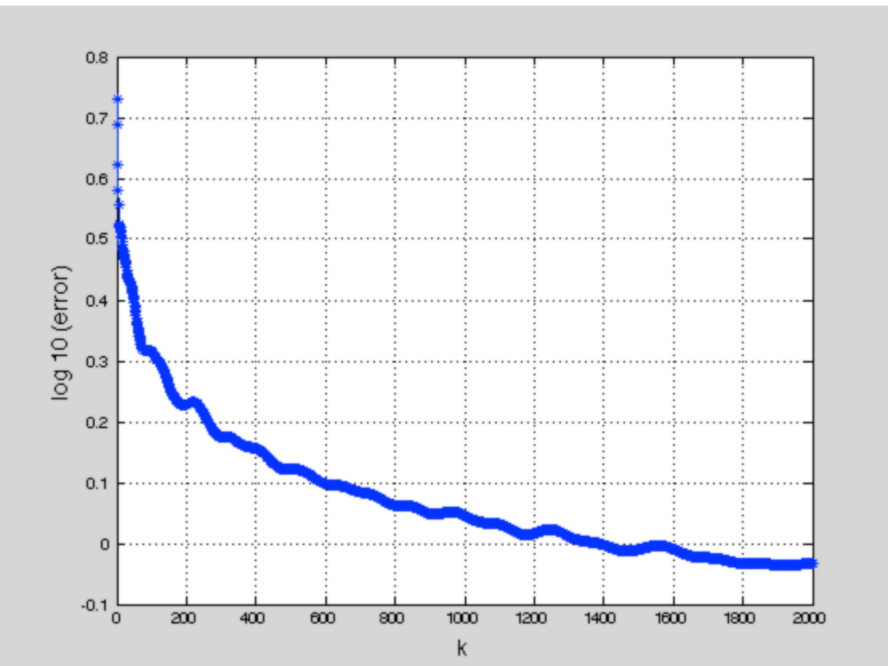
Theorem 5.7 (Al-Baali [3]).

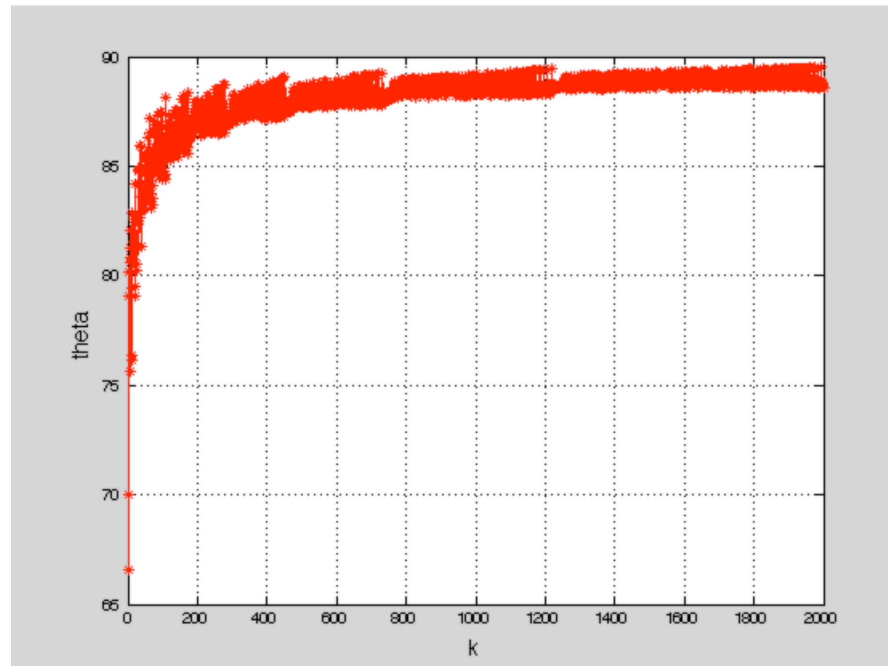
Suppose that Assumptions 5.1 hold, and that Algorithm 5.4 is implemented with a line search that satisfies the strong Wolfe conditions (5.43), with $0 < c_1 < c_2 < \frac{1}{2}$. *Then*

$$\liminf_{k \to \infty} \|\nabla f_k\| = 0. \tag{5.63}$$

Performance of F-R

$$\min_{x \in R^n} f(x) = \ln(1 + x^T Q x), \text{ where } Q = Q^T \text{ is p.d.}$$





- F-R is globally convergent, but the performance is not satisfactory.
- $cos(\theta_k) \approx 0 \Longrightarrow cos(\theta_{k+1}) \approx 0 \Longrightarrow$ very small reduction in the cost function.

Nonlinear CG – Restart

F-R with restart:

If $\boxed{\dfrac{|\nabla f_k^T \nabla f_{k-1}|}{\|\nabla f_k\|^2} \geq \nu \approx 0.1 \qquad \beta_{k+1}^{FR} = 0}$

When F-R encounters a bad step, use steepest descent direction as the search direction.

else

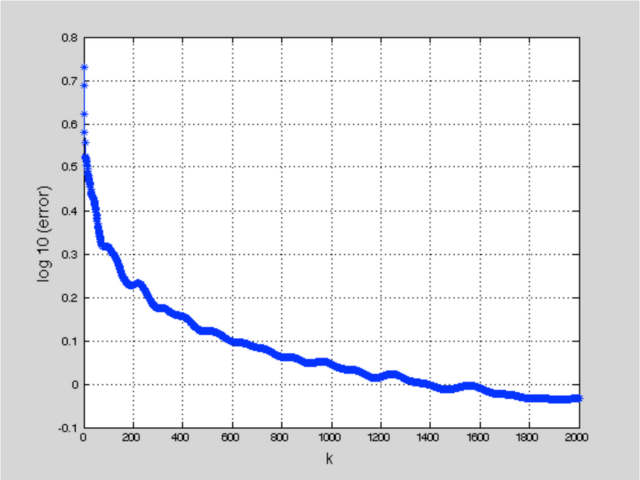
$$\beta_{k+1}^{\text{FR}} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k};$$

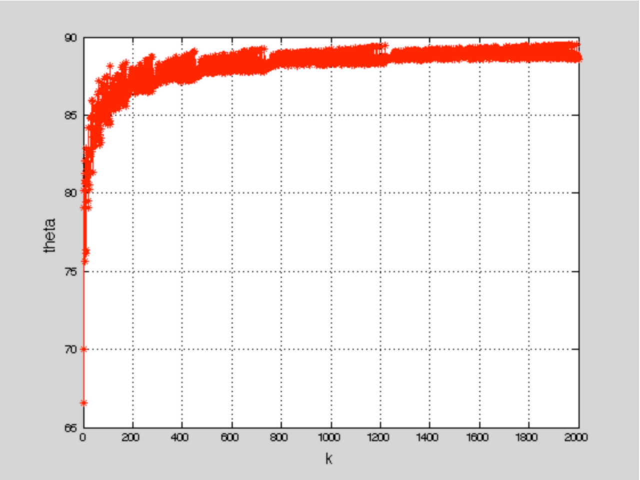
end

$$p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_{k+1}^{\text{FR}} p_k;$$

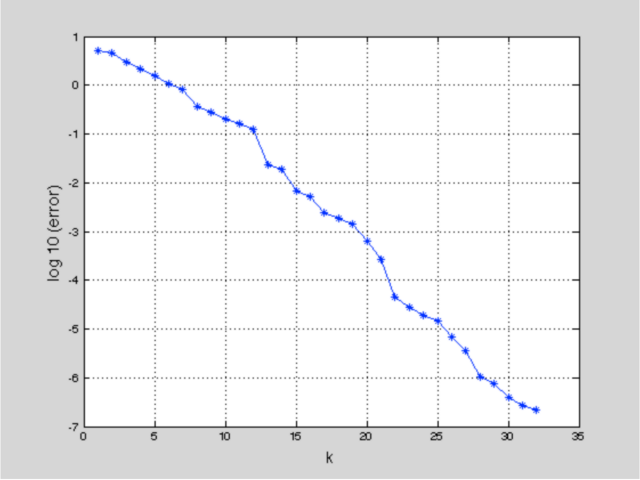
Performance of F-R with Restart

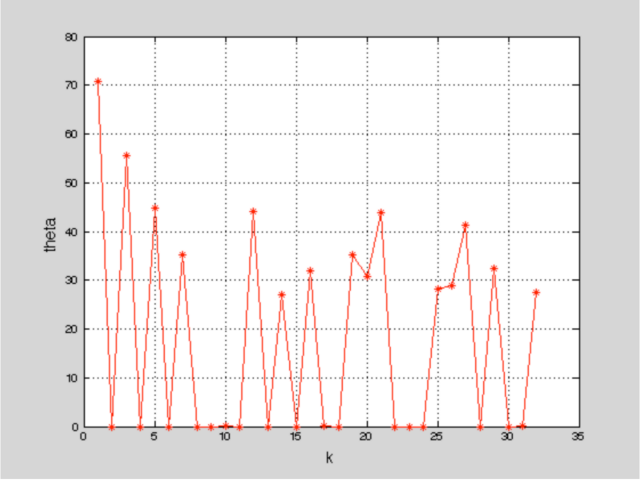
$$\min_{x \in R^n} f(x) = \ln(1 + x^T Q x), \text{ where } Q = Q^T \text{ is p.d.}$$





F-R





F-R w restart

Nonlinear CG Methods

- **Polak-Ribiere:** $$\beta_{k+1}^{\mathrm{PR}} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\|\nabla f_k\|^2}$$

Theorem 5.8.

Consider the Polak–Ribière method method (5.44) with an ideal line search. There exists a twice continuously differentiable objective function $f : \mathbb{R}^3 \to \mathbb{R}$ and a starting point $x_0 \in \mathbb{R}^3$ such that the sequence of gradients $\{\|\nabla f_k\|\}$ is bounded away from zero.

- **PR+:** $$\beta_{k+1}^{+} = \max\{\beta_{k+1}^{\mathrm{PR}}, 0\},$$

- **Hestenes-Stiefel:** $$\beta_{k+1}^{\mathrm{HS}} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{(\nabla f_{k+1} - \nabla f_k)^T p_k}$$