

Lecture-16

Lemma 5.6 & Theorem 5.7

Lemma 5.6

Suppose that the Algorithm 5.4 is implemented with a step length such that it satisfies strong wolf conditions with $0 < c_2 < 1/2$. Then the method generates the descent directions p_k that satisfies the following inequalities:

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}, \quad \forall k = 0, 1, \dots$$

$$-1 < \frac{2c_2-1}{1-c_2} < 0, \quad 0 < c_2 < \frac{1}{2} \qquad -2 < \frac{-1}{1-c_2} < -1, \quad 0 < c_2 < \frac{1}{2} \quad (\text{B})$$

Proof

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}, \quad \forall k=0,1,\dots$$

$$-1 < \frac{2c_2-1}{1-c_2} < 0, \quad 0 < c_2 < \frac{1}{2} \quad -2 < \frac{-1}{1-c_2} < -1, \quad 0 < c_2 < \frac{1}{2} \quad (\text{B})$$

$$\text{Induction, } k=0: \quad \frac{\nabla f_0^T p_0}{\|\nabla f_0\|^2} = -\frac{\nabla f_0^T \nabla f_0}{\|\nabla f_0\|^2} = -1$$

So by using (B), it is easy to see the both inequalities are satisfied.

Assume holds for k .

$$p_{k+1} = -\nabla f_{k+1} + \mathbf{b}_{k+1} p_k; \quad \text{Algorithm (5.4)}$$

$$\nabla f_{k+1}^T p_{k+1} = -\nabla f_{k+1}^T \nabla f_{k+1} + \mathbf{b}_{k+1} \nabla f_{k+1}^T p_k;$$

$$\frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \mathbf{b}_{k+1} \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_{k+1}\|^2}$$

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}, \quad \forall k=0,1,\dots$$

$$\frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \mathbf{b}_{k+1} \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_{k+1}\|^2}$$

$$\frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k} \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_{k+1}\|^2}$$

$$\frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2} \quad (\text{C})$$

$$\mathbf{b}_{k+1}^{FR} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k};$$

$$|\nabla f_{k+1}^T p_k| \leq c_2 |\nabla f_k^T p_k| \quad \text{Wolf's strong condition (2)}$$

$$c_2 \nabla f_k^T p_k \leq \nabla f_{k+1}^T p_k \leq -c_2 \nabla f_k^T p_k$$

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}, \quad \forall k = 0, 1, \dots$$

$$\begin{aligned} c_2 \nabla f_k^T p_k &\leq \nabla f_{k+1}^T p_k \leq -c_2 \nabla f_k^T p_k \\ \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} &\leq \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2} \leq -\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \\ -1 + \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} &\leq -1 + \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2} \leq -1 - \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \\ &\qquad \qquad \qquad \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2} \\ -1 + \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} &\leq \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \leq -1 - \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \end{aligned} \quad \text{From (C)}$$

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}, \quad \forall k = 0, 1, \dots$$

$$\begin{aligned} -1 + \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} &\leq \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \leq -1 - \frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \\ -1 - \frac{c_2}{1-c_2} &\leq \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \leq -1 + \frac{c_2}{1-c_2} \\ \frac{-1}{1-c_2} &\leq \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \leq \frac{2c_2-1}{1-c_2} \end{aligned} \quad \text{From induction}$$

QED

FR and PR

- We can use Lemma 5.6 to explain weakness of FR-GC method
 - If the method generates a bad direction and a tiny step, then subsequent directions will also be bad
- On the other hand PR-GC does not have that problem

FR

We know the angle between steepest descent and the direction is given by:

$$\cos \mathbf{q}_k = -\frac{p_k^T \nabla f_k}{\|p_k\| \|\nabla f_k\|}$$

Now from Lemma 5.6

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}, \quad \forall k = 0, 1, \dots$$

$$-\frac{1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \frac{\|\nabla f_k\|}{\|p_k\|} \leq \frac{2c_2-1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|}, \quad \forall k = 0, 1, \dots$$

$$-\frac{1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\| \|p_k\|} \leq \frac{2c_2-1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|}, \quad \forall k = 0, 1, \dots$$

$$c_1 \frac{\|\nabla f_k\|}{\|p_k\|} \leq \cos \mathbf{q}_k \leq c_2 \frac{\|\nabla f_k\|}{\|p_k\|}, \quad \forall k = 0, 1, \dots$$

FR

$$c_1 \frac{\|\nabla f_k\|}{\|p_k\|} \leq \cos \mathbf{q}_k \leq c_2 \frac{\|\nabla f_k\|}{\|p_k\|}, \quad \forall k = 0, 1, \dots \quad (\text{G})$$

$$\begin{aligned} & \cos \mathbf{q}_k \approx 0 \\ \text{iff} & \quad \|\nabla f_k\| \ll \|p_k\| \end{aligned} \quad \cos \mathbf{q}_k = -\frac{p_k^T \nabla f_k}{\|p_k\| \|\nabla f_k\|}$$

Since p_k is almost orthogonal to the gradient, the step from X_k to X_{k+1} is tiny i.e.

$$\begin{aligned} x_k &\approx x_{k+1} & \mathbf{a}_k &= -\frac{\nabla f_k^T p_k}{p_k^T A p_k} \\ \nabla f_{k+1} &\approx \nabla f_k \end{aligned}$$

FR

$$\mathbf{b}_{k+1}^{FR} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k} \approx 1$$

$$p_{k+1} \leftarrow -\nabla f_{k+1} + \mathbf{b}_{k+1}^{FR} p_k; \quad \|\nabla f_k\| \ll \|p_k\|$$

$$p_{k+1} \approx p_k;$$

So the new direction will improve little on the previous one.

PR

$$x_k \approx x_{k+1}$$

$$\nabla f_{k+1} \approx \nabla f_k$$

$$\mathbf{b}_{k+1}^{PR} \leftarrow \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\nabla f_k^T \nabla f_k}$$

$$\mathbf{b}_{k+1}^{PR} = 0$$

$$p_{k+1} \leftarrow -\nabla f_{k+1} + \mathbf{b}_{k+1}^{PR} p_k;$$

$$p_{k+1} = -\nabla f_{k+1}$$

The new search direction will be close to the steepest descent;
Restart!

Example

Problem, $n=100$:

$\cos \mathbf{q}$ is order of 10^{-2}

$\|x_k - x_{k-1}\|$ are order of 10^{-2}

FR-CG takes thousands of iterations (improves with restarts)

PR-CG 37 iterations

Convergence of Line Search Methods (Theorem 3.2) (Theorem 5.7)

If step length satisfies the Wolf's conditions, the function is continuously differentiable, the gradient is Lipschitz continuous then:

$$\sum_{k \geq 0} \cos^2 \mathbf{q}_k \|\nabla f_k\|^2 < \infty$$

The angle between p_k and steepest descent direction $-\nabla f_k^T$

$$\cos \mathbf{q}_k = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \|p_k\|}$$

Convergence of Line Search Methods

$$x_{k+1} = x_k + \mathbf{a}_k p_k$$

Iteration scheme

$$\nabla f(x_k + \mathbf{a}_k p_k)^T p_k \geq c_2 \nabla f_k^T(x_k) p_k, \quad c_2 \in (c_1, 1) \quad \text{Curvature condition}$$

Therefore

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \geq (c_2 - 1) \nabla f_k^T p_k \quad (1)$$

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L \|x - \tilde{x}\| \quad \text{Lipschitz continuous}$$

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \leq \mathbf{a}_k L \|p_k\|^2 \quad (2)$$

Convergence of Line Search Methods

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \leq \mathbf{a}_k L \|p_k\|^2 \quad (2)$$

$$\frac{(\nabla f_{k+1} - \nabla f_k)^T p_k}{L \|p_k\|^2} \leq \mathbf{a}_k$$

$$\mathbf{a}_k \geq \frac{(\nabla f_{k+1} - \nabla f_k)^T p_k}{L \|p_k\|^2}$$

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \geq (c_2 - 1) \nabla f_k^T p_k \quad (1)$$

Combining (1) and (2)

$$\mathbf{a}_k \geq \frac{c_2 - 1}{L} \frac{\nabla f_k^T p_k}{\|p_k\|^2}$$

Convergence of Line Search Methods

$$\mathbf{a}_k \geq \frac{c_2 - 1}{L} \frac{\nabla f_k^T p_k}{\|p_k\|^2}$$

$$f(x_k + \mathbf{a} p_k) \leq f(x_k) + c_1 \mathbf{a}_k \nabla f_k^T p_k$$

$$f(x_k + \mathbf{a} p_k) \leq f(x_k) - c_1 (-\mathbf{a}_k) \nabla f_k^T p_k$$

Sufficient decrease

Therefore

$$f_{k+1} \leq f_k - c_1 \frac{1 - c_2}{L} \frac{(\nabla f_k^T p_k)^2}{\|p_k\|^2}$$

$$f_{k+1} \leq f_k - c \cos^2 \mathbf{q}_k \|\nabla f_k\|^2, \quad c = c_1(1 - c_2)/L$$

$$f_{k+1} \leq f_0 - c \sum_{j=0}^k \cos^2 \mathbf{q}_j \|\nabla f_j\|^2$$

Convergence of Line Search Methods

$$f_{k+1} \leq f_0 - c \sum_{j=0}^k \cos^2 \mathbf{q}_j \|\nabla f_j\|^2$$

Since f is bounded below, we have $f_0 - f_{k+1}$ is less than some positive constant for all k

Taking the limits:

$$\sum_{k \geq 0} \cos^2 \mathbf{q}_k \|\nabla f_k\|^2 < \infty$$

Convergence of Line Search Methods

$$\sum_{k \geq 0} \cos^2 \mathbf{q}_k \|\nabla f_k\|^2 < \infty \quad \cos^2 \mathbf{q}_k \|\nabla f_k\|^2 \rightarrow 0$$

$$\cos \mathbf{q}_k \geq \mathbf{d} > 0 \quad \text{If angle is bounded away From } 90^\circ$$

$$\lim_{k \rightarrow \infty} \|\nabla f_k\|^2 = 0$$

We can be sure that gradient norms converges to zero, provided that the search directions are never too close to orthogonality with the gradient

Therefore, the steepest descent produces a gradient sequence that converges to zero, provided that it uses a line search satisfying Wolf's conditions.

We can not guarantee that the method converges to a minimizer, but only that it is attracted by stationary points.