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:

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

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

Proof

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

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

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

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

Assume holds for k.

$$p_{k+1} = -\nabla f_{k+1} + \beta_{k+1} p_k; \qquad \text{Algorithm (5.4)}$$

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

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

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

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

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

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

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

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

$$c_2 \nabla f_k^T pk \le \nabla f_{k+1}^T pk \le -c_2 \nabla f_k^T pk$$

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

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

$$\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2} \le -\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2}$$

$$-1+\frac{c_2 \nabla f_k^T pk}{\|\nabla f_k\|^2} \le -1+\frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2} \le -1-\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2}$$

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

From (C)

$$-1+\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \le -1-\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2}$$

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

$$-1+\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \le -1-\frac{c_2 \nabla f_k^T p_k}{\|\nabla f_k\|^2}$$

From induction

$$-1-\frac{c_2}{1-c_2} \le \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \le -1+\frac{c_2}{1-c_2}$$

$$\frac{-1}{1-c_2} \le \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \le \frac{2c_2-1}{1-c_2}$$

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:

$$\cos\theta_k = -\frac{p_k^T \nabla f_k}{\| p_k \| \| \nabla f_k \|}$$

Now from Lemma 5.6

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

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

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

$$\chi_1 \frac{\| \nabla f_k \|}{\| p_k \|} \le \quad \cos\theta_k \quad \le \chi_2 \frac{\| \nabla f_k \|}{\| p_k \|}, \ \forall k = 0,1\ldots$$

FR

$$\chi_1 \frac{\|\nabla f_k\|}{\|p_k\|} \leq \quad \cos\theta_k \quad \leq \chi_2 \frac{\|\nabla f_k\|}{\|p_k\|}, \ \forall k = 0,1\ldots \qquad \text{(G)}$$

$$\cos\theta_k \approx 0$$

iff

$$\|\nabla f_k\| << \|p_k\|$$

$$\cos\theta_k = -\frac{p_k{}^T \nabla f_k}{\|p_k\| \|\nabla f_k\|}$$

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

$$x_k \approx x_{k+1} \qquad \alpha_k = -\frac{\nabla f_k^T p_k}{p_k^T A p_k}$$

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

FR

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

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

$$\|\nabla f_k\| << \|p_k\|$$

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

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

PR

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

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

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

$$\beta_{k+1}^{PR} = 0$$

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

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

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

Example

Problem, n=100:

$\cos\theta_k$ 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:

$$\sum_{k\geq 0} \cos^2 \theta_k \, \| \nabla f_k \|^2 < \infty$$

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

$$\cos \theta_k = \frac{-\nabla f_k^T p_k}{\| \nabla f_k \| \| p_k \|}$$

Convergence of Line Search Methods

$$x_{k+1} = x_k + \alpha_k p_k$$ Iteration scheme

$$\nabla f (x_k + \alpha p_k)^T p_k \geq c_2 \nabla f_k^T (x_k) p_k, \quad c_2 \in (c_1, 1)$$ Curvature condition

Therefore

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

$$\| \nabla f (x) - \nabla f (\widetilde{x}) \| \leq L \| x - \widetilde{x} \|$$ Lipschitz continuous

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

Convergence of Line Search Methods

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \le \alpha_k L \| p_k \|^2$$

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

$$\alpha_k \ge \frac{(\nabla f_{k+1} - \nabla f_k)^T p_k}{L \| p_k \|^2} \qquad (3)$$

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

Combining (1) and (3)

$$\alpha_k \ge \frac{c_2 - 1}{L} \frac{\nabla f_k^T p_k}{\| p_k \|^2}$$

Convergence of Line Search Methods

$$\alpha_k \ge \frac{c_2 - 1}{L} \frac{\nabla f_k^T p_k}{\| p_k \|^2}$$

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k \qquad \text{Sufficient decrease}$$

$$f(x_k + \alpha p_k) \le f(x_k) - c_1 (-\alpha_k) \nabla f_k^T p_k$$

Therefore

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

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

$$f_{k+1} \le f_0 - c \sum_{j=0}^{k} \cos^2 \theta_j \| \nabla f_j \|^2$$

Convergence of Line Search Methods

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

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

Taking the limits:

$$\sum_{k \geq 0} \cos^2 \theta_k \, \| \nabla f_k \|^2 < \infty$$

Convergence of Line Search Methods

$$\sum_{k \geq 0} \cos^2 \theta_k \, \| \nabla f_k \|^2 < \infty \qquad \cos^2 \theta_k \, \| \nabla f_k \|^2 \to 0$$

$$\cos \theta_k \geq \delta > 0$$

If angle is bounded away From 90^0

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

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.