

Conditions of convergence

- Steepest Descent: Wolf's conditions
- Newton and Quasi-Newton: In addition to Wolfe's conditions, PD Hessian, and bounded condition number
- Conjugate Gradient: subsequence of direction cosines cosq_k is bounded away from zero.



- Steepest descent: Linear
- Quasi-Newton: Super-linear
- Newton: Quadratic
- Conjugate Gradient: *n* steps

Convergence of Line Search Methods

- The steepest descent method is globally convergent
- For other algorithms how far *p_k* can deviate from the steepest descent direction and still gives rise to globally convergent iteration.

Convergence of Line Search Methods (Theorem 3.2)

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

$$\cos \boldsymbol{q}_{k} = \frac{-\nabla f_{k}^{T} p_{k}}{\|\nabla f_{k} \| \| p_{k} \|}$$

We will show (Theorem 3.2):

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

Convergence of Line Search Methods

$$x_{k+1} = x_k + \mathbf{a}_k p_k \quad \text{Iteration scheme}$$

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

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

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

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

$$\le \mathbf{a}_k L \| p_k \| \| p_k \|$$

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

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

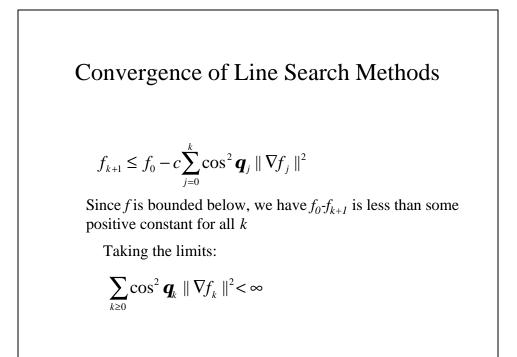
$$a_{k} \geq \frac{c_{2}-1}{L} \frac{\nabla f_{k}^{T} p_{k}}{\|p_{k}\|^{2}}$$

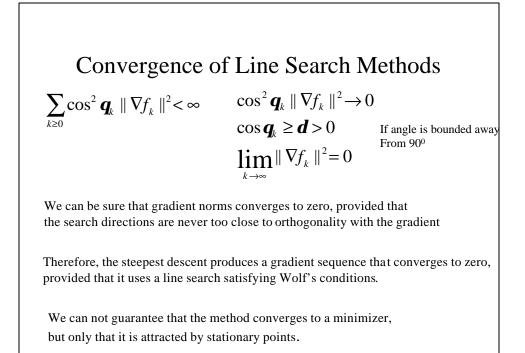
$$f(x_{k} + \mathbf{a}p_{k}) \leq f(x_{k}) + c_{1}\mathbf{a}_{k} \nabla f_{k}^{T} p_{k} \qquad \text{Sufficient decrease}$$
Therefore

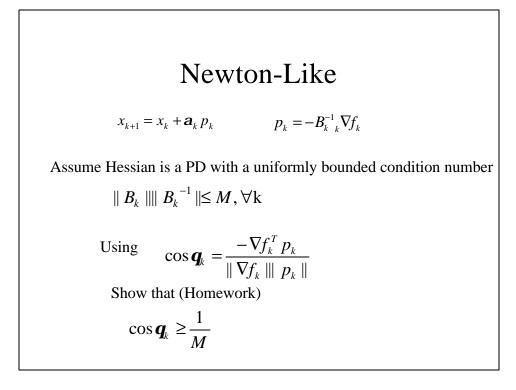
$$f_{k+1} \leq f_{k} - c_{1} \frac{1-c_{2}}{L} \frac{(\nabla f_{k}^{T} p_{k})^{2}}{\|p_{k}\|^{2}}$$

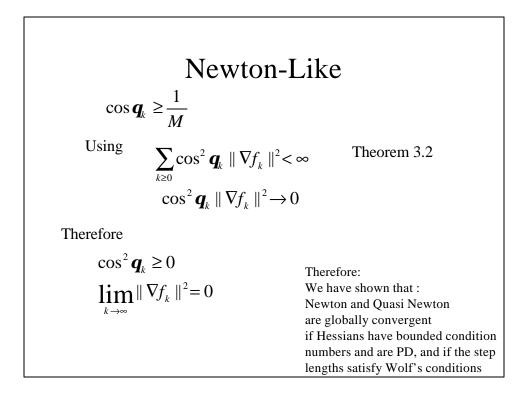
$$f_{k+1} \leq f_{k} - c \cos^{2} \mathbf{q}_{k} \|\nabla f_{k}\|^{2}, \quad c = c_{1}(1-c_{2})/L$$

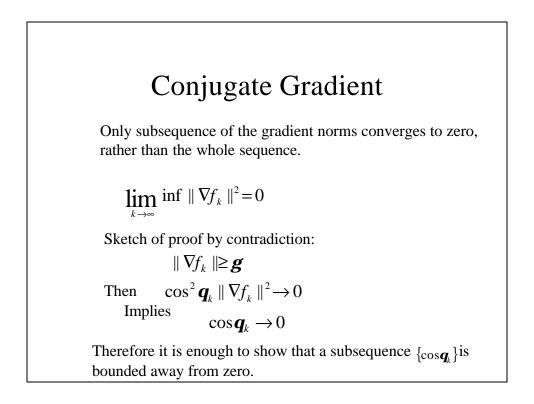
$$f_{k+1} \leq f_{0} - c \sum_{j=0}^{k} \cos^{2} \mathbf{q}_{j} \|\nabla f_{j}\|^{2}$$

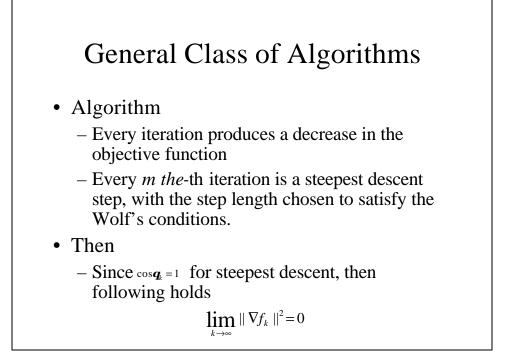












Convergence Rate of Steepest Descent: Quadratic Function

$$\|x_{k+1} - x^*\|_{Q}^{2} \le \left(\frac{I_n - I_1}{I_n + I_1}\right)^{2} \|x_k - x^*\|_{Q}^{2}$$
 Theorem 3.3

As the condition number increases the contours of the quadratic become more elongated, the zigzags of line search becomes more

pronounced.

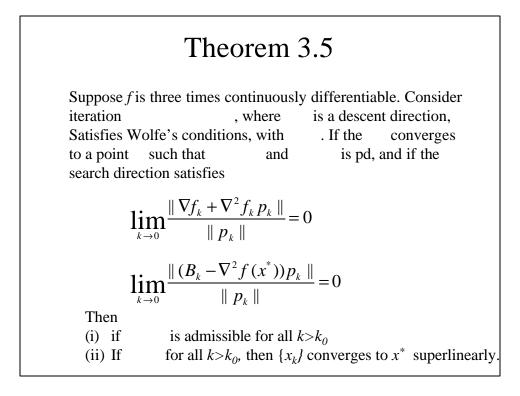
9

Theorem 3.4: Steepest Descent

$$f(x_{k+1}) - f(x^*) \le \left(\frac{I_n - I_1}{I_n + I_1}\right)^2 (f(x_k) - f(x^*))$$

where $0 \le l_1 \le l_2 \le \dots l_n$ are eigenvalues of Hessian

If the condition number is 800, and $f(x_I)=1$ and $f(x^*)=0$, After 1000 iterations the value of function will be .08.



Theorem 3.6

Suppose f is three times continuously differentiable. Consideriteration, whereis given by Quasi-Newtondirection. Assume the sequenceconverges to a pointsuch thatandis pd, theconvergessuperlinearly ifif the following condition holds.

$$\lim_{k \to 0} \frac{\|(B_k - \nabla^2 f(x^*))p_k\|}{\|p_k\|} = 0$$

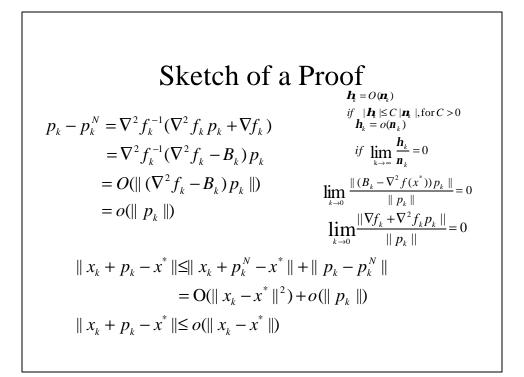
Order Notations

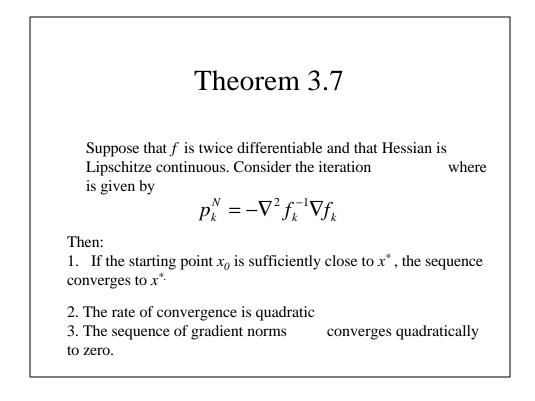
Given two non-negative infinite sequences

.

$$\boldsymbol{h}_{k} = O(\boldsymbol{n}_{k})$$
if $|\boldsymbol{h}_{k}| \leq C |\boldsymbol{n}_{k}|$, for $C > 0, \forall k$

$$\boldsymbol{h}_{k} = o(\boldsymbol{n}_{k})$$
if $\lim_{k \to \infty} \frac{\boldsymbol{h}_{k}}{\boldsymbol{n}_{k}} = 0$





Coordinate Descent Method

Cycle through *n* coordinate directions $e_1, e_2, \dots e_n$ using each in turn as a search direction.

Fix all other variables except one, and minimize the function.

It is an inefficient method, it can iterate infinitely without Ever approaching a point, where the gradient vanishes.

The gradient may become more and more perpendicular to search Directions, making $\cos q$ approach to zero, but not the gradient.