Lecture-7

Step Length Selection

Homework (Due 2/13/03)

- 3.1
- 3.2
- 3.5
- 3.6
- 3.7
- 3.9
- 3.10
- Show equation 3.44
- The last step in the proof of Theorem 3.6. (see slides)
- Show that if c1>.5, the line search would exclude the minimizer of a quadratic, and unit step length may not be admissible. (Theorem 3.5)

Sufficient condition

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha \nabla f_k^T p_k, \quad c_1 \in (0,1) \qquad c_1 = 10^{-4}$$

$$f(x_k + \alpha p_k) - f(x_k) \le c_1 \alpha \nabla f_k^T p_k, \quad c_1 \in (0,1)$$

The reduction should be proportional to both the step length, and directional derivative.

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha \nabla f_k^T p_k, \quad c_1 \in (0,1)$$

$$f(x_k + \alpha p_k) \le l(\alpha)$$

St line

Sufficient condition

$$f(x_k + \alpha p_k) \le l(\alpha)$$

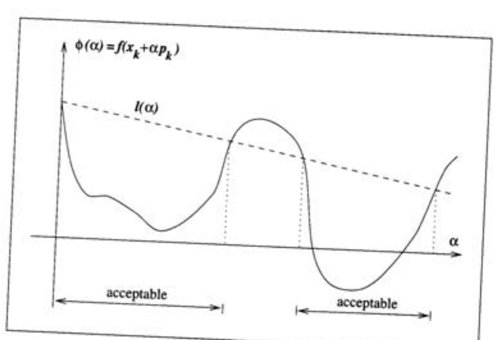


Figure 3.3 Sufficient decrease condition.

Problem:
The sufficient decrease condition is satisfied for all small values of step length

Curvature condition

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

$c_2 = .9$ for Newton and Quasi - Newton

$c_2 = .1$ for conjugate gradient

Derivative

The slope of $\phi(\alpha_k)$ is greater than c_2 times the gradient $\phi'(0)$.

Curvature condition

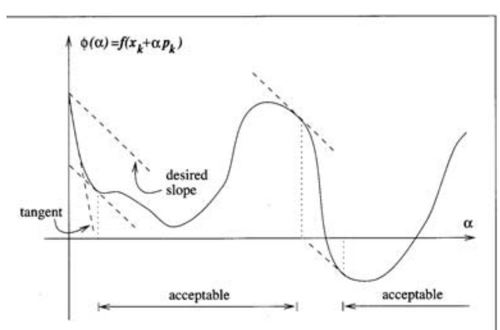


Figure 3.4 The curvature condition.

If the slope is strongly negative, that means we can reduce f further along the chosen direction

If the slope is positive, it indicates we can not decrease f further in this direction.

Wolfe conditions

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha \nabla f_k^T p_k, \quad c_1 \in (0,1)$$ Sufficient decrease

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

Backtracking Line Search

If line search method chooses its step length appropriately, we can dispense with the second condition

Choose $\overline{\alpha} > 0$, $\rho, c \in (0,1)$; set $\alpha \leftarrow \overline{\alpha}$;

repeat until $f(x_k + \alpha p_k) \le f(x_k) + c\alpha \nabla f_k^T p_k$

$\alpha \leftarrow \rho\alpha$;

end(*repeat*)

Terminate with $\alpha_k = \alpha$

$\overline{\alpha} = 1$, for Newton and quasi-Newton

This ensures that the step length is short enough to satisfy the sufficient decrease condition, but not too short.

Searching Step Length Using Interpolation

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha \nabla f_k^T p_k, \quad c_1 \in (0,1)$$ Sufficient decrease

$$\phi(\alpha_k) \le \phi(0) + c_1 \alpha_k \phi'(0) \qquad \phi(\alpha) = f(x_k + \alpha p_k)$$

1. Assume α_0 is the initial guess. Then if we have:

   $$\phi(\alpha_0) \le \phi(0) + c_1 \alpha_0 \phi'(0)$$

   Then this step length satisfies the condition, we terminate the search.

2. Otherwise, we know $[0, \alpha_0]$ contains the acceptable step lengths. We fit quadratic polynomial to three pieces of information:

   $$\phi_q(0) = \phi(0), \phi_q'(0) = \phi'(0), \phi_q(\alpha_0) = \phi(\alpha_0)$$

Searching Step Length Using Interpolation

and find step length α_1 by analytically minimizing this polynomial

If the sufficient decrease condition is satisfied for this α_1 then we terminate the search.

3. If not we fit cubic polynomial to interpolate four pieces of information,and analytically minimize this polynomial to find

$$\phi_c(0) = \phi(0), \phi_c'(0) = \phi'(0), \phi_c(\alpha_0) = \phi(\alpha_0), \phi_c(\alpha_1) = \phi(\alpha_1)$$

If necessary we can repeat this process with $\phi(0), \phi'(0)$ and two most recent values of ϕ .

Quadratic Interpolation

$$\phi_q(\alpha) = a\alpha^2 + b\alpha + c$$

$$\phi_q(0) = \phi(0), \phi_q'(0) = \phi'(0), \phi_q(\alpha_0) = \phi(\alpha_0)$$

$$\phi_q(\alpha) = \left(\frac{\phi(\alpha_0) - \phi(0) - \alpha_0\phi'(0)}{\alpha_0^2}\right)\alpha^2 + \phi'(0)\alpha + \phi(0)$$

$$\frac{d}{d\alpha}\phi_q(\alpha) = 2\left(\frac{\phi(\alpha_0) - \phi(0) - \alpha_0\phi'(0)}{\alpha_0^2}\right)\alpha + \phi'(0) = 0$$

$$\alpha_1 = -\left(\frac{\phi'(0)\alpha_0^2}{2(\phi(\alpha_0) - \phi(0) - \alpha_0\phi'(0))}\right)$$

Cubic Interpolation

$$\phi_c(\alpha) = a\alpha^3 + b\alpha^2 + c\alpha + d$$

$$\phi_c(0) = \phi(0), \phi_c'(0) = \phi'(0), \phi_c(\alpha_0) = \phi(\alpha_0), \phi_c(\alpha_1) = \phi(\alpha_1)$$

$$\phi_c(\alpha) = a\alpha^3 + b\alpha^2 + \phi'(0)\alpha + \phi(0)$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\alpha_0^2\alpha_1^2(\alpha_1 - \alpha_0)}\begin{bmatrix} \alpha_0^2 & -\alpha_1^2 \\ -\alpha_0^3 & \alpha_1^3 \end{bmatrix}\begin{bmatrix} \phi(\alpha_1) - \phi(0) - \phi'(0)\alpha_1 \\ \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0 \end{bmatrix}$$

$$\alpha_2 = -\left(\frac{-b + \sqrt{b^2 - 3a\phi'(0)}}{3a}\right)$$

Algorithm 3.2 (Line Search Algorithm)

Set $\alpha_0 \leftarrow 0$, choose $\alpha_1 > 0$, and $\alpha_{\max}$;

$i \leftarrow 1$

repeat

Evaulate $\phi(\alpha_i)$;

if $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$ or $[\phi(\alpha_i) > \phi(\alpha_{i-1}), i > 1]$ — 1st Wolfe's condition

$\alpha_* \leftarrow zoom(\alpha_{i-1}, \alpha_i)$, and *stop*;

Evaulate $\phi'(\alpha_i)$;

if $|\phi'(\alpha_i)| \le -c_2\phi'(0)$ — 2nd Wolfe's condition

set $\alpha_* \leftarrow \alpha_i$, and *stop*;

if $\phi'(\alpha_i) \ge 0$

set $\alpha_* \leftarrow zoom(\alpha_i, \alpha_{i-1})$, and *stop*;

choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$

$i \leftarrow i + 1$;

end(*repeat*)

Algorithm 3.3 (Zoom)

repeat

Interpolate to find a trial step length

α_j between α_{lo}, α_{hi};

Evaulate $\phi(\alpha_j)$;

if $\phi(\alpha_j) > \phi(0) + c_1\alpha_j\phi'(0)$ or $[\phi(\alpha_j) > \phi(\alpha_{lo})]$ — 1st Wolfe's condition

$\alpha_{\mathrm{hi}} \leftarrow \alpha_{\mathrm{j}}$;

esle

Evaulate $\phi'(\alpha_j)$;

if $|\phi'(\alpha_j)| \le -c_2\phi'(0)$ — 2nd Wolfe's condition

set $\alpha_* \leftarrow \alpha_j$, and *stop*;

if $\phi'(\alpha_j)(\alpha_{hi} - \alpha_{lo}) \ge 0$

set $\alpha_{hi} \leftarrow \alpha_{lo}$

set $\alpha_{lo} \leftarrow \alpha_j$

end(*repeat*)

CHAPTER 3. LINE SEARCH METHODS

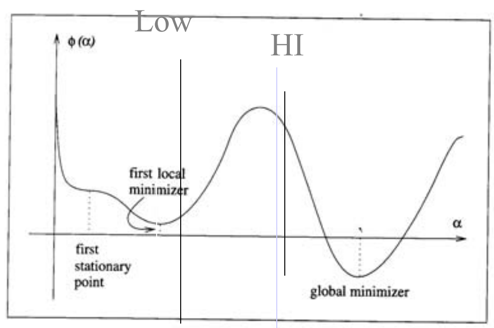


Figure 3.1 The ideal step length is the global minimizer

CHAPTER 3. LINE SEARCH METHODS

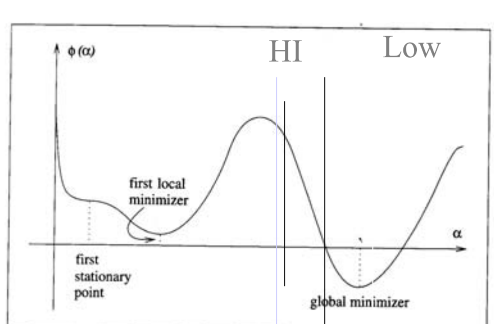


Figure 3.1 The ideal step length is the global minimizer

Theorem 3.5 (Any Descent Direction)

Suppose f is three times continuously differentiable. Consider iteration $x_{k+1} = x_k + \alpha_k p_k$, where p_k is a descent direction, α_k Satisfies Wolfe's conditions, with $c_1 \le \frac{1}{2}$. If the $\{x_k\}$ converges to a point x^* such that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is pd, and if the search direction satisfies

$$\lim_{k \to 0} \frac{\| \nabla f_k + \nabla^2 f_k p_k \|}{\| p_k \|} = 0$$

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

Then

(i) $\alpha_k = 1$ is admissible for all $k > k_0$ and

(ii) if $\alpha_k = 1$ for all $k > k_0$, then $\{x_k\}$ converges to x^* superlinearly.

Theorem 3.6 (Quasi-Newton)

Suppose f is three times continuously differentiable. Consider iteration $x_{k+1} = x_k + p_k$, where p_k is given by Quasi-Newton direction. Assume the sequence $\{x_k\}$ converges to a x^* point such that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is pd, the $\{x_k\}$ converges superlinearly 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

$$\eta_k = O(\nu_k)$$
$$if \quad |\eta_k| \le C|\nu_k|, \text{for } C > 0, \forall k$$

$$\eta_k = o(\nu_k)$$
$$if \ \lim_{k\to\infty} \frac{\eta_k}{\nu_k} = 0$$

Sketch of a Proof

$$\eta_k = O(\nu_k)$$
$$if \quad |\eta_k| \le C|\nu_k|, \text{for } C > 0$$
$$\eta_k = o(\nu_k)$$
$$if \ \lim_{k\to\infty} \frac{\eta_k}{\nu_k} = 0$$
$$\lim_{k\to 0} \frac{\|(B_k - \nabla^2 f(x^*))p_k\|}{\|p_k\|} = 0$$
$$\lim_{k\to 0} \frac{\|\nabla f_k + \nabla^2 f_k p_k\|}{\|p_k\|} = 0$$

$$\begin{aligned} p_k - p_k^N &= \nabla^2 f_k^{-1}(\nabla^2 f_k p_k + \nabla f_k) \\ &= \nabla^2 f_k^{-1}(\nabla^2 f_k - B_k)p_k \\ &= O(\|(\nabla^2 f_k - B_k)p_k\|) \\ &= o(\|p_k\|) \end{aligned}$$

Norm of Hessian is bounded.

Sketch of a Proof

$$\| x_k + p_k - x^* \| = \| x_k + p_k^N - p_k^N + p_k - x^* \| \le \| x_k + p_k^N - x^* \| + \| p_k - p_k^N \|$$

$$= O(\| x_k - x^* \|^2) + o(\| p_k \|)$$

$$\| x_k + p_k - x^* \| \le o(\| x_k - x^* \|)$$

$\eta_k = o(\nu_k)$

if $\lim_{k \to \infty} \frac{\eta_k}{\nu_k} = 0$

Theorem 3.7

Super-linear

Show this in Homework

Theorem 3.7 (Newton)

Suppose that f is twice differentiable and that Hessian is Lipschitze continuous. Consider the iteration $x_{k+1} = x_k + p_k$ where p_k is given by

$$p_k^N = -\nabla^2 f_k^{-1} \nabla f_k$$

Then:

1. If the starting point x_0 is sufficiently close to x^*, the sequence converges to x^*.

2. The rate of convergence is quadratic
3. The sequence of gradient norms $\| \nabla f(x_k) \|$ converges quadratically to zero.

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 $\cos\theta$ approach to zero, but not the gradient.

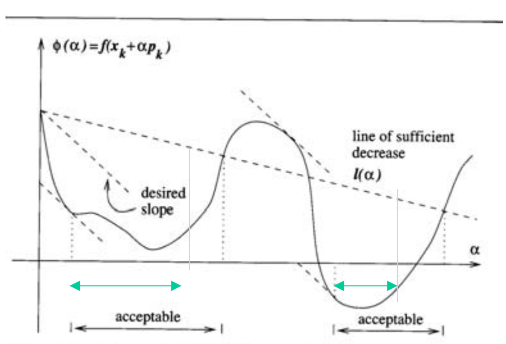
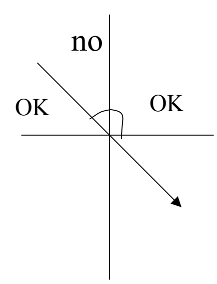
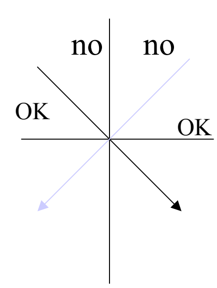


Figure 3.5 Step lengths satisfying the Wolfe conditions.





angle	0	30	45	60	90	120	135	150	180
tan	0	.5774	1	1.732	∞	-1.732	-1	$-.5774$	0
$\vert \tan \vert$	0	.577	1	1.732	∞	1.732	1	.5774	0