Lecture-4

Line Search Methods: Search Directions, and step lengths

Line Search Methods

$$x_{k+1} \leftarrow x_k + a_k p_k$$

$$p_k \leftarrow -B_k^{-1} \nabla f_k$$

Steepest descent B_k is and identity matrix
Newton B_k is a Hessian matrix
Quasi-Newton B_k is approximation to the Hessian matrix

Inverse Hessian

Instead of inverting approximation of Hessian, we can directly compute the approximation of inverse of Hessian:

$$H_{k+1} = (I - r_k s_k y_k^T) H_k (I - r_k s_k y_k^T) + r_k s_k s_k^T,$$

$$r_k = \frac{1}{y_k^T s_k} \qquad s_k = x_{k+1} - x_k, \qquad H_k = B_k^{-1}$$

$$y_k = \nabla f_{k+1} - \nabla f_k$$

$$p_k = -H_k \nabla f_k$$

Quasi Newton

Conjugate Gradient

$$p_k = -\nabla f(x_k) + b_k p_{k-1}$$

b_k is scalar such that p_{k-1} and p_k are conjugate

Two vectors are conjugate with respect to a matrix G if

$$p_k{}^T G p_{k-1} = 0$$

Non-interfering directions, with the special property that minimization along one direction is not spoiled by subsequent minimization along another.

Step Length

(Exact Search) The global minimizer of the univariate function:

$$f(a) = f(x_k + ap_k) \quad a > 0$$

Too many evaluations of a function, and its gradient

(In-exact search): adequate reduction in f at minimal cost.
Two step method:

Bracketing (find the interval containing desirable step lengths)
bisection (compute step length within this interval)

Step Length

Ideal step length is the global minimizer
Step length should achieve sufficient decrease
And it should not be too small

CHAPTER 3. LINE SEARCH METHODS

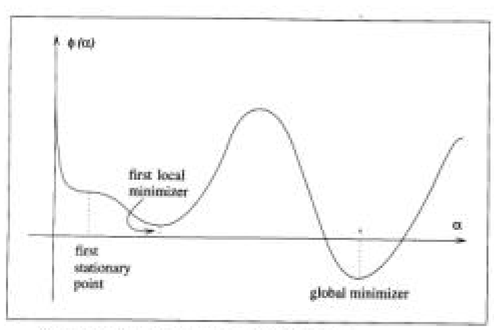


Figure 3.1 The ideal step length is the global minimizer

Simple Condition

Simple condition: reduction in *f*

$$f(x_k + \alpha p_k) < f(x_k)$$

This is not appropriate.

$$\left\{\frac{5}{k}\right\}, \; k = 1,2,3,\ldots$$

We don not have sufficient reduction

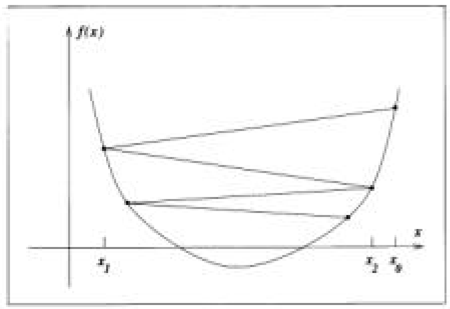


Figure 3.2 Insufficient reduction in *f*.

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

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

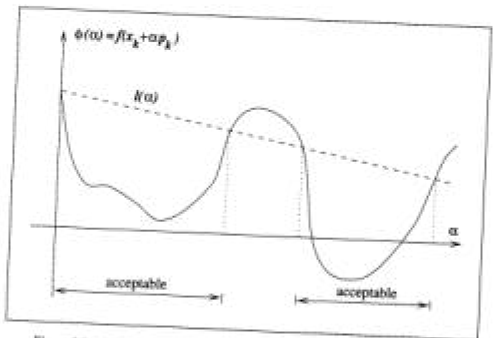


Figure 3.3 Sufficient decrease condition.

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

Curvature condition

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

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

$c_2 = .1$ for conjugate gradient

Derivative $f'(a_k)$

The slope of $f(a_k)$ is greater than c_2 times the gradient $f'(0)$.

Curvature condition

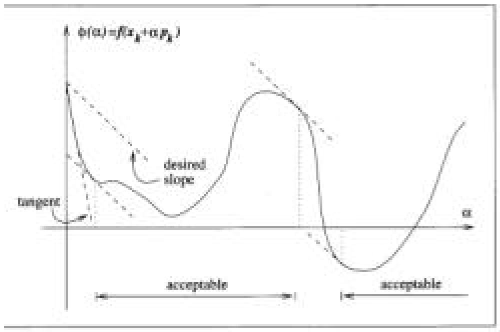


Figure 3.4 The curvature condition.

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

Strong Wolfe conditions

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

$$|\nabla f \geq (x_k + \alpha p_k)^T p_k| \leq c_2 |\nabla f_k^T(x_k) p_k|$$

This forces step length to lie in at least in a broad neighborhood of a local minimizer or a stationary point of f .

$f'(\alpha)$ should not be too positive, exclude points which are Further away from the stationary points of f

Goldstein conditions

$$f(x_k) + (1\text{-}c)\alpha_k \nabla f_k^T p_k \leq f(x_k + \alpha p_k) \leq f(x_k) + c\ \alpha_k \nabla f_k^T p_k$$

To control step length from the below $0 < c < \frac{1}{2}$

Sufficient decrease

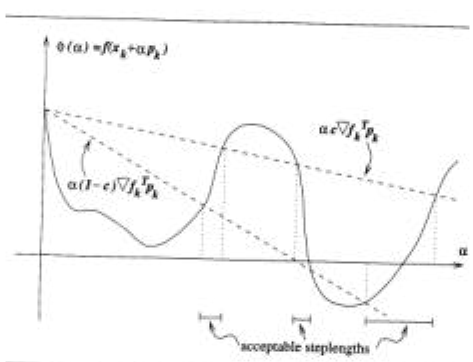


Figure 3.6 The Goldstein conditions.

Disadvantage:
It may exclude minmizers

Quadratic Functions

$$f(x) = \frac{1}{2} x^T Q x - b^T x$$ Q is symmetric, Hessian of f

$$\nabla f(x) = Qx - b$$

if x^* is a unique solution of $Qx = b$, then it is a stationary point of f

$$f(x^* + \alpha p) = f(x^*) + \frac{1}{2}\alpha^2 p^T Q p$$

Let u_i and λ_i be eigenvector and eigenvalue of Q then

$$Qu_j = \lambda_j u_j$$

Quadratic Functions

$$f(x^* + \alpha p) = f(x^*) + \frac{1}{2}\alpha^2 p^T Q p$$

Let p is equal to u_i

$$Qu_j = \lambda_j u_j$$

$$f(x^* + \alpha u_j) = f(x^*) + \frac{1}{2}\alpha^2 u_j^T Q u_j$$

$$f(x^* + \alpha u_j) = f(x^*) + \frac{1}{2}\alpha^2 u_j^T \lambda_j u_j$$

$$f(x^* + \alpha u_j) = f(x^*) + \frac{1}{2}\alpha^2 \lambda_j$$

Q is orthonormal

Quadratic Functions

- The change in f when moving away from x^* along the direction u_j depends on the sign of λ_j
  - If λ_j is positive f will strictly increase as $|\alpha|$ increases
  - If λ_j is negative, f is decreasing as $|\alpha|$ increases.
  - If λ_j is zero, the value of f remains constant when moving along any direction parallel to u_j
  - f reduces to a linear function along any such direction, since quadratic term vanishes.

$$f(x^* + \alpha u_j) = f(x^*) + \frac{1}{2}\alpha^2 \lambda_j$$

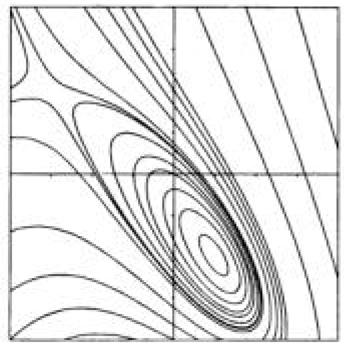
Quadratic Functions

- When all eigenvalues of Q are positive, x^* is the unique global minimum.
  - The contours of f are ellipsoid whose principal axes are in the directions of the eigevectors of Q, with lengths proportional to squareroot of corresponding eigenvalues.
- If Q is positive semi-definite, a stationary point (if it exists) is a week local minimum.
- If Q is indefinite and non-singular, x^* is a saddle point, f is unbounded.

$$f(x^* + \alpha u_j) = f(x^*) + \frac{1}{2}\alpha^2 \lambda_j$$

Iso Contours (Contour Map)

$f(x_1, x_2) = c$



$$f(x_1, x_2) = e^{x_1}(4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1)$$
$$c = .2, .4, 1, 1.7, 1.8, 2, 3, 4, 5, 6, 20$$

Quadratic Functions

Two positive eigenvalues

$$Q = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} -5.5 \\ -3.5 \end{bmatrix}$$

PD

One positive eigenvalue,
one zero eigenvalue

$$Q = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$$

Semi PD

One positive eigenvalue,
one negative eigenvalue

$$Q = \begin{bmatrix} 3 & -1 \\ -1 & -8 \end{bmatrix}, \quad b = \begin{bmatrix} -.5 \\ 8.5 \end{bmatrix}$$

indefinite

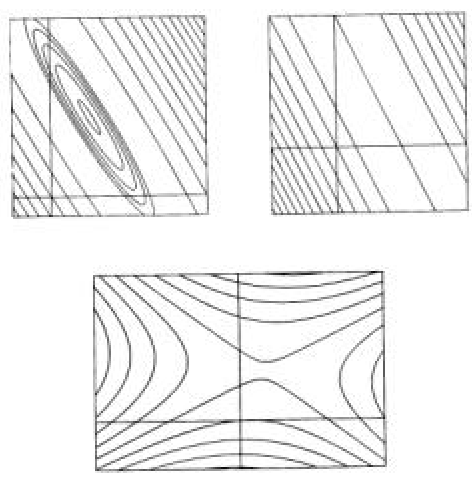


Figure 3f. Contours of: (i) a positive-definite quadratic function; (ii) a positive semi-definite quadratic function; and (iii) an indefinite quadratic function.