Lecture-3

Search Directions

Homework Due 1/25/01

- 2.1, 2.2, 2.3, 2.8, 2.13, 2.14

Rate of Convergence

Definition : Suppose $\{p_n\}_{n=0}^{\infty}$ is a sequence that converges to p and that $e_n = p_n - p$

$$\lim_{n\to\infty}\frac{|p_{n+1}-p|}{|p_n - p|^a} = \lim_{n\to\infty}\frac{|e_{n+1}|}{|e_n|^a} = l$$

then the seq is said to converge to p of order a with asymptotic error constant l.

$a = 1$, linear

$a = 2$, quadratic

$a = 1$, and $l = 0$, superlinear

Example

$$\lim_{n\to\infty}\frac{|e_{n+1}|}{|e_n|^a} = .75$$

assume $e_0 = .5$, then error not exceeding 10^{-8} requires 5 iterations for quadratic convergence, and 62 iterations for the linear convergence.

Problem

$$\min_x f(x)$$

Definitions

A point x^* is a stationary point if $f(x^*) = 0$

A point x^* is a global minimizer if $f(x^*) \le f(x) \ \forall x$

A point x^* is a local minimizer if there is a neighborhood N s.t.
$f(x^*) \le f(x) \ \forall x \in \mathrm{N}$

A point x^* is a strict local minimzer if
there is a neighborhood N s.t.
$f(x^*) < f(x) \ \forall x \in \mathrm{N}, x \ne x^*$

if $\nabla f(x^*) = 0$, but x^* is neither a minimum nor
a maxima, it is called a saddle point.

First Order necessary conditions

If x^* is a local minimizer and f is continuously differentiable in an open neighborhood of x^*, then $\nabla f(x^*) = 0$.

Second order necessary conditions

If x^* is a local minimizer and $\nabla^2 f$ is continuous in an open neighborhood of x^*, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite.

Second order sufficient conditions

Suppose that $\nabla^2 f$ is continuous in an open neighborho od of x^* and that $\nabla f(x^*) = 0$ and $\nabla f(x^*)$ is positive definite. Then x^* is a strict local minimizer of f.

Convex Function

is a convex function if for any two points x and y in its domain, the graph of f lies below straight line connecting $(x, f(x))$ to $(y, f(y))$

$$f(ax + (1-a)y) \leq af(x) + (1-a)f(y) \quad \forall a \in [0,1]$$

Convex Function

When f is convex, any local minimizer x^* is a global minimzer of f. If in addition f is differentiable, then any stationary point x^* is a gobal minimzer of f.

Line Search

$$\min_{a>0} f\left(x_k + a p_k\right)$$

x_k current iterate
p_k direction of a search
a distance to move along

Model Algorithm for Smooth Functions

- Let x_k be the current estimate of x^* .
  - [Test for convergence.] If the conditions for convergence are satisfied, the algorithm terminates with x_k as a solution.
  - [Compute a search direction.] Compute a non-zero *n*-vector p_k, the direction of search.
  - [Compute a step length.] Compute a positive scalar, a_k , the step length, for which it holds that

$$f(x_k + a_k p_k) < f(x_k)$$

  - [Update the estimate of the minimum.] Set

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

and go back to the first step.

Steepest Descent

$$f(x_k + ap) = f(x_k) + ap^T \nabla f_k + \frac{1}{2} a^2 p^T \nabla^2 f(x_k + p)p$$

Taylor series

$$\min_p p^T \nabla f_k \qquad \text{subject to } \| p \| = 1$$

$$p^T \nabla f_k = \| p \| \| \nabla f_k \| \cos q \qquad \text{dot product}$$

$$p^T \nabla f_k = \| p \| \| \nabla f_k \| (-1) \qquad \text{minimum value}$$

$$p = -\frac{\nabla f_k}{\| p \| \| \nabla f_k \|}$$

$$p = -\frac{\nabla f_k}{\| \nabla f_k \|}$$

Steepest Descent

$p_k = -\nabla f$ Steepest descent direction

$p_k{}^T \nabla f_k = \| p_k \| \| \nabla f_k \| \cos q_k < 0$ down hill direction

Any descent direction-one that
makes an angle of strictly less than
90 degrees with the gradient vector
produces a decrease in f, provided, that the
step length is sufficiently small.

Newton’s Direction

$$f(x_k + p) = f_k + p^T \nabla f_k + \frac{1}{2} p^T \nabla^2 f_k\, p = m_k$$

Taylor series approximation

Hessian

$$\frac{dm_k}{dp} = \nabla f_k + \nabla^2 f_k\, p = 0$$

$$p = -\nabla^2 f_k{}^{-1} \nabla f$$

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

N superscript for Newton

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

$$-\nabla^2 f_k\, p_k^N = \nabla f_k$$

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

$$\nabla f_k{}^T p_k^N = -p_k^{N^T} \nabla^2 f_k\, p_k^N \le -s_k \| p_k^N \|^2$$

Because $\nabla^2 f_k$ is p.d.

$$\nabla f_k{}^T p_k^N < 0$$

Therefore p_k^N is a descent direction

Newton's Direction

- There is a natural step length, α_k of 1 for Newton's direction.
- If $\nabla^2 f_k$ is not p.d., the Newton's directions may not be defined, because inverse may not exists.
- Even inverse exists, the descent property may not be satisfied.
- In that case, the search direction is modified to be a down hill direction.
- Newton direction gives a quadratic local convergence.
- The main drawback of Newton's method is computation of a Hessian matrix.

Approximation of Hessian

Taylor Series

$$\nabla f(x+p) = \nabla f(x) + \nabla^2 f(x)p$$

Let

$$p = x_{k+1} - x_k,\ x = x_k$$

$$\nabla f_{k+1} = \nabla f_k + \nabla^2 f_{k+1}(x_{k+1} - x_k)$$

$$\nabla^2 f_{k+1}(x_{k+1} - x_k) = \nabla f_{k+1} - \nabla f_k$$

$$B_{k+1} s_k = y_k$$

Approximate Hessian (→ B_{k+1})

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f_{k+1} - \nabla f_k$$

B_{k+1} should be symmetric

The difference between successive approximation B_{k+1} to B_k have a low rank.

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$ SRI (symmetric rank one)

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$ Broyden,Fletcher, Shanno

Quasi-Newton

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

Inverse Hessian

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

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

$$r_k = \frac{1}{y_k^T s_k} \qquad H_k = B_k^{-1}$$

$$p_k = -H_k \nabla f_k$$

Conjugate Gradient

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

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

Two vectors are conjugate with respect to a matrix G if

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