Lecture-4

Line Search Methods: Search Directions, and step lengths

\[ x_{k+1} \leftarrow x_k + \alpha_k p_k \]

\[ p_k \leftarrow -B_k^{-1} \nabla f_k \]

Steepest descent: is an identity matrix
Newton: is a Hessian matrix
Quasi-Newton: is approximation to the Hessian matrix
Inverse Hessian

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

\[ H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k s_k y_k^T) + \rho_k s_k s_k^T \]

\[ \rho_k = \frac{1}{y_k^T s_k} \]

\[ s_k = x_{k+1} - x_k \]

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

\[ p_k = -H_k \nabla f_k \]

Quasi Newton

Conjugate Gradient

\[ p_k = -\nabla f(x_k) + \beta_k p_{k-1} \]

is scalar such that and are conjugate

Two vectors are conjugate with respect to a PD 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:

\[ \phi(\alpha) = f(x_k + \alpha p_k) \quad \alpha > 0 \]

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
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

Sufficient condition

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

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

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

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

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

St line
Sufficient condition

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

\[ f(x_k + \alpha p_k) \leq l(\alpha) \]

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

Curvature condition

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

Derivative

The slope of is greater than times the gradient.

\[ c_2 = .9 \text{ for Newton and Quasi-Newton} \]

\[ c_2 = .1 \text{ for conjugate gradient} \]
\[ \nabla f(x_k + \alpha p_k)^T p_k \geq c_2 \nabla f_k^T (x_k) p_k, \quad c_2 \in (c_1, 1) \]

**Curvature condition**

If the slope is strongly negative (too steep), that means we can go further along the chosen direction (you should not stop there). If the slope is positive, it indicates we cannot decrease further in this direction.
Wolfe conditions

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

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

Sufficient decrease
Curvature

Strong Wolfe conditions

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

\[ |\nabla f(x_k + \alpha p_k)^T p_k| \leq c_2 |\nabla f(x_k)^T (x_k) p_k| \]

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

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

should not be too positive, exclude points which are
Further away from the stationary points of
Goldstein conditions

\[ f(x_k) + (1-c)\alpha_k \nabla f_k^T p_k \leq f(x_k + \alpha p_k) \leq f(x_k) + c \alpha_k \nabla f_k^T p_k \]

To control step length from the below

\[ 0 < c < \frac{1}{2} \]

Sufficient decrease

Disadvantage:
It may exclude minimizers
Goldstein conditions

\[ f(x_k) + (1-c)\alpha_k \nabla f_k^T p_k \leq f(x_k + \alpha p_k) \leq f(x_k) + \alpha_k \nabla f_k^T p_k \]

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

(finite difference approximation)

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

(Wolf's curvature condition)