Lecture-4

Line Search Methods: Search Directions, and step lengths

Line Search Methods

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

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

Steepest descent \( B \) is an identity matrix
Newton \( B \) is a Hessian matrix
Quasi-Newton \( B \) is an approximation to the Hessian matrix
Inverse Hessian

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

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

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

\[ s_k = x_{k+1} - x_k, \quad H_k = B_k^{-1} \]

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

\[ p_k = -H_k \nabla f_k \]

Quasi Newton

Conjugate Gradient

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

\( \beta_k \) 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:

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

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 do not have sufficient reduction.

Sufficient condition

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_{kT} p_k, \ c_1 \in (0, 1)$$

$$c_1 = 10^{-4}$$

$$f(x_k + \alpha p_k) - f(x_k) \leq c_1 \alpha \nabla f_{kT} p_k, \ c_1 \in (0, 1)$$

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

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_{kT} p_k, \ c_1 \in (0, 1)$$

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

St line
Sufficient condition

\[ f(\mathbf{x}_k + \alpha \mathbf{p}_k) \leq l(\alpha) \]

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

Curvature condition

\[ \nabla f(\mathbf{x}_k + \alpha \mathbf{p}_k)^T \mathbf{p}_k \geq c_2 \nabla f_k^T(\mathbf{x}_k) \mathbf{p}_k, \quad c_2 \in (c_1, 1) \]

Derivative \( \phi'(\alpha) \)

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

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

Sufficient decrease

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

Curvature
Strong Wolfe conditions

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

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

This forces step length to lie in at least in a broad neighborhood of a local minimizer or a stationary point of \( \phi \).

\( \phi(\infty) \) should not be too positive, exclude points which are further away from the stationary points of \( \Phi \).

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

Figure 3.6: The Goldstein conditions.
Quadratic Functions

\[ f(x) = \frac{1}{2} x^T Q x - b^T x \]

Q is symmetric, Hessian of \( f \)

\[ \nabla f(x) = Q x - b \]

If \( x^* \) is a unique solution of \( Q x = b \), 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 \( \lambda_i \) be eigenvector and eigenvalue of \( Q \) then

\[ Qu_j = \lambda_j u_j \]
Quadratic Functions

- The change in $f$ when moving away from $x^*$ along the direction $u_j$ depends on the sign of $\lambda_j$
  - If $\lambda_j$ is positive, $f$ will strictly increase as $|\alpha| \text{ increases}$
  - If $\lambda_j$ is negative, $f$ is decreasing as $|\alpha| \text{ increases}$.
  - If $\lambda_j$ 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 eigenvectors 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)**

\[ f(x_1, x_2) = c \]

\[ f(x_1, x_2) = e^c(4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1) \]

\[ c = .2, .4, 1, 1.7, 1.8, 2, 3, 4, 5, 6, 20 \]

---

**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**

\[ Q = \begin{bmatrix} 3 & -1 \\ -1 & -8 \end{bmatrix}, \quad b = \begin{bmatrix} -0.5 \\ 8.5 \end{bmatrix} \]

- Indefinite

---

*Figure 10. Contours of (i) a positive-definite quadratic function; (ii) a positive semi-definite quadratic function; and (iii) an indefinite quadratic function.*