Line Search Methods

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

January 26, 2024

Line Search Method

- In each iteration of a line search method a search direction p_k is computed, and
- then its decided how far to move along that direction.
- An iteration is given by

$$x_{k+1} = x_k + \alpha_k p_k \tag{1}$$

  where $\alpha_k > 0$ (scalar) called the step length.

The success of a line search method depends on effective choices of both:

1. the direction p_k
2. the step length α_k

The Steepest Descent Direction

- The steepest descent direction $-\nabla f_k$ is the most obvious choice for search direction for a line search method.
- Among all Directions once could from x_k, along $-\nabla f_k$, f decreases most rapidly.

Justification

- Consider any search direction p and step-length α, we have

$$f(x_k+\alpha p) = f(x_k)+\alpha p^T\nabla f_k+\frac{1}{2}\alpha^2 p^T\nabla^2 f(x_k+tp)p \text{ for some } \ t \in (0,$$

- Let $\alpha << 1$ (small) and we consider the first-order approximation of f at $x_k + \alpha p$ around x_k as:

$$f(x_k + \alpha p) \approx f(x_k) + \alpha p^T \nabla f_k$$

- Change in f moving from x_k to $x_k + \alpha p$ is

$$f(x_k + \alpha p) - f(x_k)$$

The Steepest Descent Direction

- As the distance moved in the direction is α, therefore, the rate of change of f along the direction p at x_k is

$$\frac{f(x_k + \alpha p) - f(x_k)}{\alpha}$$

- which is coefficient of α, i.e.

$$p^T \nabla f_k$$

- This implies smaller the above value is, more descent can be achieved.
- Hence, the unit direction p of most rapid decrease is the solution to the problem

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

The Steepest Descent Direction

- $$p^T \nabla f_k = ||p|| \; ||\nabla f_k|| \; \cos\theta = ||\nabla f_k|| \; \cos\theta$$

  where θ is the angle between p and ∇f_k.
- The minimiser is attained when $\cos\theta = -1$ and

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

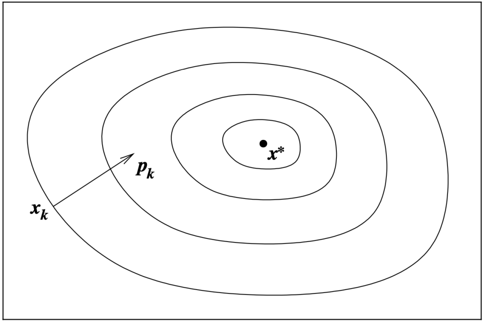

Figure illustrates this direction is orthogonal to the contours of the function

The Steepest Descent Direction

- At every step (iteration) in the steepest descent method the search direction is chosen along

$$p = -\nabla f_k$$

- α_k can be chosen in a variety of ways.
- One advantage of this method is it requires only the calculation of gradient (∇f_k), but not second derivatives.
- Line search methods may use search directions other than the steepest descent direction.

Descent Direction

Descent Direction

Any direction that makes an angle of strictly less than $\frac{\pi}{2}$ radians with ∇f_k is guaranteed to produce a decrease in f, provided the step length is sufficiently small and is called a descent direction.

- Now consider

$$f(x_k + \epsilon p_k) = f(x_k) + \epsilon p_k^T \nabla f_k + \mathcal{O}(\epsilon^2).$$

- When p_k is a downhill (descent) direction, the angle θ_k between p_k and ∇f_k has $\cos\theta_k < 0$, so that

$$p_k^T \nabla f_k = ||p_k|| \; ||\nabla f_k|| \; \cos\theta_k < 0$$
$$\implies f(x_k + \epsilon p_k) < f(x_k)$$

- Most line search algorithms require p_k to be descent direction, because this property guarantees that the function f can be reduced along this direction.

Newton Search Direction

- This direction is derived from the second-order Taylor series approximation to $f(x_k + p)$, which is

$$f(x_{k+p}) \approx f_k + p\nabla f_k + p\nabla f_k p =^{def} m_k(p). \qquad (2)$$

- Assuming for the moment that $\nabla^2 f_k$ is positive definite, we obtain the Newton direction by finding the vector p that minimizes $m_k(p)$.
- By simply setting the derivative of $m_k(p)$ to zero, we obtain the following explicit formula:

$$p_k^N = -(\nabla^2 f_k)^{-1}\nabla f_k. \qquad (3)$$

- The Newton direction is reliable when the difference between the true function $f(x_k + p)$ and its quadratic model $m_k(p)$ is not too large.

Newton Search Direction

- If $\nabla^2 f$ is sufficiently smooth, this difference introduces a perturbation of only $\mathscr{O}(||p||^3)$.
- Therefore, when $||p||$ is small

$$f(x_k + p) \approx m_k(p) \qquad \text{quite accurately}$$

- The Newton direction can be used in a line search method when $\nabla^2 f_k$ is positive definite, as in this case we have

$$\begin{aligned} \nabla f_k^T p_k^N &= -{p_k^N}^T \nabla^2 f_k p_k^N \\ &< -\sigma_k ||p_k^N||^2 \end{aligned}$$

  for some $\sigma_k > 0$ (+ve definiteness of $\nabla^2 f_k$)
- Unless the gradient ∇f_k (and therefore the step p_k^N) is zero, we have

$$\nabla f^T p_k^N < 0$$

- so the Newton direction is a descent direction.

Newton Search Direction

- Unlike the steepest descent direction, there is a "natural" step length of 1 associated with the Newton direction.
- Adjust α only when it does not produce a satisfactory reduction in the value of f .
- Note that when ∇_k^f is not positive definite the Newton direction may not exist, since $(\nabla^2 f_k)^{-1}$ may not exist.
- Even when it is defined, it may not satisfy the descent property, and therefore is unsuitable.
- Methods that use Newton direction have fast rate of local convergence (more on this later).
- After a neighbourhood of the solution is reached, convergence to high accuracy often occurs in just a few iterations.

Newton Search Direction

- The main drawback is the need to calculate the Hessian $\nabla^2 f(x)$.
- Explicit computation of this matrix of second derivatives can sometimes be a cumbersome, error-prone, and expensive process.
- Finite-difference and automatic differentiation techniques come useful in avoiding the need to calculate second derivatives by hand.