Note on Optimization Methods

Chunpai Wang

October 2015

1 Steepest Descent Method

1.1 Notations

- $\triangle x$ is a vector in R^n, called the step direction or search direction.
- t is the step size or step length.
- $||\cdot||$ is any norm on $\mathbb{R}^n$
- quadratic norm $||x||_Q = (x^T Qx)^{\frac{1}{2}} = ||Q^{\frac{1}{2}}x||_2$, where $Q \in S^n_{++}$ (symmetric positive definite $n \times n$ matrices).
- $||\cdot||_*$ is dual norm, given by $||u||_* = sup\{u^T x| \ \ ||x|| \leq 1\}$
- In Euclidean space, inner product $< x, y >= x^T y = ||x||_2||y||_2 cos\theta, \theta$ is the angle between x and y

1.2 Intuition and Interpretation

The first-order Taylor approximation of $f(x + v)$ around x is

$$f(x + v) \approx f(x) + \nabla f(x)^T v \tag{1}$$

where $\nabla f(x)^T v$ is the directional derivative of f at x in the direction v. If the directional derivative is negative, the step v is a descent direction.
We now address the question of how to choose v to make the directional derivative as negative as possible. Since the directional derivative $\nabla f(x)^T v$ is linear in v, it can be made as negative as we like by taking v large. To make the question sensible we have to limit the size of v, or normalize by the length of v.
Therefore, we define a normalized steepest descent direction is

$$\triangle x_{nsd} = \arg\min_v \{\nabla f(x)^T v \ | \ \ ||v|| = 1\} \tag{2}$$

Since $\nabla f(x)^T v = ||\nabla f(x)||||v||cos\theta$, where θ is the angle between $\nabla f(x)$ and v, it is easy to see that the minimizer is attained when $cos\theta = -1$ and

$$\triangle x_{nsd} = v = -\nabla f(x)/||\nabla f(x)|| \tag{3}$$

Note: the direction v is a vector, such as $< 2, 1 >$, after normalized, it becomes $< \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} >$ A normalized steepest descent direction $\triangle x_{nsd}$ is a unit norm step that with most negative directional derivative. You can see the figure below, the direction is orthogonal to the contours of the function.
(Unnormalized) steepest descent direction (at x, for norm $||\cdot||$)is

$$\triangle x_{sd} = ||\nabla f(x)||_* \triangle x_{nsd} \tag{4}$$

satisfies $\nabla f(x)^T \triangle x_{sd} = ||\nabla f(x)||_* \nabla f(x)^T \triangle x_{nsd} = -||\nabla f(x)||^2_*$
Note the dual norm here, it may reflect the relation between DFP and BFGS.

1.3 Other Line Search Method

The steepest descent method is a line search method that moves along $\triangle x_k = -\nabla f_k$ at every step. It can choose the step length α_k in a variety of ways.

- advantage: require calculation of the gradient ∇f_k but not of second derivatives.
- disadvantage: can be excruciatingly slow on difficult problems.

Line search methods may use search directions other than the steepest descent direction. In general, any descent direction - one that makes an angle of strictly less than $\pi/2$ radians with ∇f_k is guaranteed to produce a decrease in f, provided that the step length is sufficient small. (In conjugate direction descent, we need to find the step length. Also, nonlinear conjugate gradient directions are much more effective than the steepest descent direction and are almost as simple to compute).

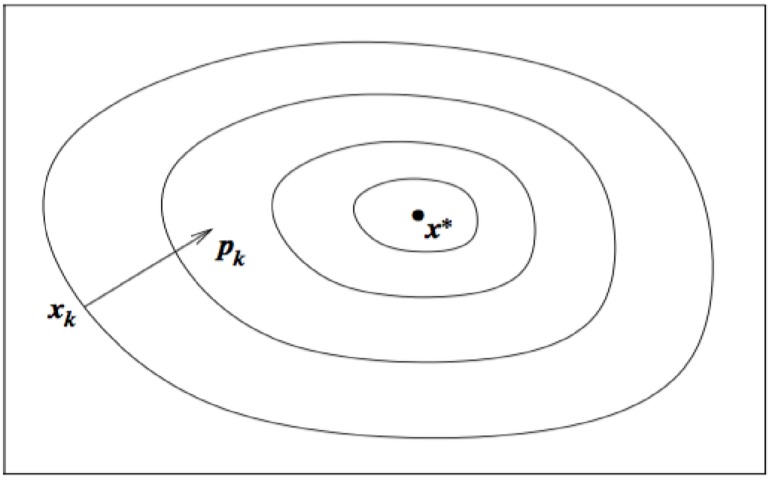

Figure 1: steepest descent direction for a function of two variables