

Nonlinear Optimization Lecture 18

Garrick Aden-Buie

Tuesday, March 29, 2016

Quasi-Newton Method

- Steepest Descent
- Newton's Method

$$x^{k+1} = \lambda_k d^k$$

$$d^k = -D_k \nabla f(x^k)$$

$$D_{k+1} = D_k + C_k$$

Davidon-Fletcher-Powell (DFP) Method

In the DFP method, D_k is updated via

$$C_k^{DFP} = \frac{p^k (p^k)^T}{(p^k)^T q^k} - \frac{D_k q^k (q^k)^T D_k}{(q^k)^T D_k q^k}$$

Where $p^k = x^{k+1} - x^k$ and $q^k = \nabla f(x^{k+1}) - \nabla f(x^k)$.

Broyden Family

In this case, D_k is updated via

$$C_k^B = C_k^{DFP} + \frac{\phi \tau_k v_k (v_k)^T}{(p^k)^T q^k}$$

Where

$$v^k = p^k - \frac{1}{\tau_k} D_k q^k$$

$$\phi: \text{ a constant } \phi \geq 0$$

$$\tau_k: \text{ a constant, with } C_k q^k = p^k - D_k q^k$$

$$\tau_k = \frac{(q^k)^T D_k q^k}{(p^k)^T q^k} > 0$$

Broyden-Fletcher-Goldfarb-Shanno (BFGS)

Basically Broyden family when $\phi = 1$.

$$C_k^{BFGS} = C_k^B \big|_{\phi=1}$$

When $\phi = 0$, $C_k^B = C_k^{DFP}$. Otherwise, C_k^B can be written as a convex combination of DFP and BFGS, when $\phi \in (0, 1)$

$$C_k^B = (1 - \phi)C_k^{DFP} + \phi C_k^{BFGS}$$

Conjugate Directions

- H : $n \times n$ symmetric matrix
- Vectors d^1, \dots, d^n are called H -conjugate if
 - They are linearly independent, and
 - $(d^i)^T H d^j = 0 \ \forall i \neq j$

Consider

$$\min_x f(x) = c^T x + \frac{1}{2} x^T H x$$

for H : $n \times n$ symmetric. Any $x \in \mathbb{R}^n$, x can be represented as

$$x = x^1 + \sum_{j=1}^n \lambda_j d^j$$

$$f(x) = F(\lambda) = c^T (x^1 + \sum \lambda_j d^j) + \frac{1}{2} (x^1 + \sum \lambda_j d^j)^T H (x^1 + \sum \lambda_j d^j)$$

Note that this function $f(x)$ is not really a function of x but rather λ .

$$\begin{aligned} F(\lambda) &= c^T x^1 + \sum_{j=1}^n \lambda_j c^T d^j + \sum_{j=1}^n \left[\frac{1}{2} (x^1 + \lambda_j d^j)^T H (x^1 + \lambda_j d^j) \right] \\ \min F(\lambda) &= \sum_{j=1}^n F_j(\lambda_j) \\ \min F_j(\lambda_j) &= \lambda_j c^T d^j + \frac{1}{2} (x^1 + \lambda_j d^j)^T H (x^1 + \lambda_j d^j) \end{aligned}$$

If H is PD, then we have a strictly convex problem.

$$\begin{aligned}\frac{\partial F_j}{\partial \lambda_j} &= c^T d^j + (d^j)^T H(x^1 \lambda_j d^j) = 0 \\ \Rightarrow \lambda_j &= -\frac{c^T d^j + (d^j)^T H x^1}{(d^j)^T H d^j}\end{aligned}$$

So, rather than solve the larger problem of $F(\lambda)$, we can just solve individual, easier, problems for each $F_j(\lambda_j)$. But, the tricky part of this method is that you need to find each of the d_i . Next class will discuss methods for finding these H -conjugate directions.