Line Search Globalization for Quasi-Newton Methods

Our goal is to find a local minimizer for $f(\mathbf{x})$, where $f : R^n \to R$. Consider the quasi-Newton iteration

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_{\mathrm{QN}}, \quad k = 0, 1, \ldots, \tag{1}$$

with

$$\mathbf{s}_{\mathrm{QN}} = -A_k^{-1}\mathbf{g}_k,$$

where $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$ is the $n \times 1$ gradient vector and A_k is a symmetric positive definite (SPD) approximation to the $n \times n$ Hessian matrix Hess $f(\mathbf{x}_k)$.

If $\mathbf{x}^*$ is a local minimizer and Hess $f(\mathbf{x}^*)$ is SPD and A_k is a good enough approximation to Hess $f(\mathbf{x}_k)$, there is some theory that guarantees convergence of the quasi-Newton iteration (1) provided we are close enough to $\mathbf{x}^*$. The purpose of globalization is to get close enough from almost *any* initial guess $\mathbf{x}_0$.

The **Line Search Subproblem** is to compute an approximation $\tilde{\tau}$ to

$$\tau^* = \arg\min_{\tau > 0} f(\mathbf{x}_k + \tau \mathbf{s}_{\mathrm{QN}}). \tag{2}$$

If $\tilde{\tau}$ a is "good enough" approximatin, then we replace the quasi-Newton iteration (1) by

$$\mathbf{s}_{\mathrm{QN}} = -A_k^{-1}\mathbf{g}_k \tag{3}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tilde{\tau}\mathbf{s}_{\mathrm{QN}} \tag{4}$$

Note that since A_k is SPD, then $\mathbf{s}_{\mathrm{QN}}$ is a *descent direction* for f at $\mathbf{x}_k$. This means that $f(\mathbf{x}_k + \tau\mathbf{s}_{\mathrm{QN}}) < f(\mathbf{x}_k)$ for all sufficiently small positive values of τ.

Quadratic Backtracking. Here is the basic idea: Suppose we take a full quasi-Newton step $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_{\mathrm{QN}}$ and we find that it is not suitable because $f(\mathbf{x}_k + \mathbf{s}_{\mathrm{QN}}) > f(\mathbf{x}_k)$. The idea is to fit a quadratic polynomial

$$q(\tau) = a + b\tau + c\tau^2$$

using the following 3 pieces of information:

$$\begin{aligned} q(0) &= f(\mathbf{x}_k) \\ q(1) &= f(\mathbf{x}_k + \mathbf{s}_{\mathrm{QN}}) \\ q'(0) &= \nabla f(\mathbf{x}_k)^T \mathbf{s}_{\mathrm{QN}} = \mathbf{g}_k^T \mathbf{s}_{\mathrm{QN}} \end{aligned}$$

The third equation is the directional derivative of f at $\mathbf{x}_k$ in the direction $\mathbf{s}_{\mathrm{QN}}$.

Possible Final Exam Question. First compute the coefficients a, b, c of $q(\tau)$. Then compute

$$\tau_1 = \arg\min_{\tau} q(\tau).$$

One can show under assumptions given above that $0 < \tau_1 < 1$. We then compute $f(\mathbf{x}_k + \tau_1\mathbf{s}_{\mathrm{QN}})$. If this is acceptable (i.e., it is smaller than $f(\mathbf{x}_k)$), then we take τ_1 to be the "approximate solution" $\tilde{\tau}$ to the above line search subproblem. If not, apply the same idea recursively.