Matrix Secant Methods

Matrix Secant Methods

We now consider Newton-Like methods of a special type. In a Newton-Like method the iteration scheme takes the form

$$x_{k+1} := x_k - M_k^{-1} g(x_k),$$

where M_k is meant to approximate $g'(x_k)$. In the one dimensional case, a choice of particular note is the secant approximation

$$M_k = \frac{g(x_{k-1}) - g(x_k)}{x_{k-1} - x_k}.$$

With this approximation one has

$$g'(x_k)^{-1} - M_k^{-1} = \frac{g(x_{k-1}) - [g(x_k) + g'(x_k)(x_{k-1} - x_k)]}{g'(x_k)[g(x_{k-1}) - g(x_k)]}.$$

Also, near a point x^* at which g' is non–singular there exists an $\alpha > 0$ such that $\alpha \|x - y\| \le \|g(x) - g(y)\|$, so

$$\left\|g'(x_k)^{-1} - M_k^{-1}\right\| \le \frac{\frac{L}{2} \|x_{k-1} - x_k\|^2}{\alpha \|g'(x_k)\| \|x_{k-1} - x_k\|} \le K \|x_{k-1} - x_k\|$$

for some constant $K > 0$ whenever x_k and x_{k-1} are sufficiently close to x^*. Therefore, the secant method is locally two step quadratically convergent to a non–singular solution of the equation $g(x) = 0$. An additional advantage of this approach is that no extra function evaluations are required to obtain the approximation M_k.

Matrix Secant Methods

Unfortunately, the secant approximation

$$(\star) \qquad M_k = \frac{g(x_{k-1}) - g(x_k)}{x_{k-1} - x_k}$$

is meaningless in the $n > 1$ dimensional case since division by vectors is undefined. However, this can be rectified by simply writing

$$M_k(x_{k-1} - x_k) = g(x_{k-1}) - g(x_k).$$

This equation is called the Matrix Secant Equation (MSE) or the Quasi-Newton Equation (QNE) at x_k and it determines M_k along an n dimensional manifold in $\mathbf{R}^{n\times n}$. Thus equation $(\star)$ is not enough to uniquely determine M_k since $(\star)$ is n linear equations in n^2 unknowns.

Matrix Secant Methods

Consequently, we may place further conditions on the update M_k if we wish to do so. In order to see what further properties one would like the update to possess, let us consider an overall iteration scheme based on

$$(\star\star) \qquad x_{k+1} := x_k - M_k^{-1} g(x_k).$$

At every iteration we have (x_k, M_k) and compute x_{k+1} by $(\star)$. Then M_{k+1} is constructed to satisfy (MSE).
If M_k is close to $g'(x_k)$ and x_{k+1} is close to x_k, then M_{k+1} should be chosen not only to satisfy $(\star)$ but also to be as "close" to M_k as possible. In what sense should we mean "close" here?
In order to facilitate the computations it is reasonable to mean "algebraically" close in the sense that M_{k+1} is only a rank 1 modification of M_k, i.e. there are vectors $u, v \in \mathbf{R}^n$ such that

$$M_{k+1} = M_k + uv^{\mathsf{T}}.$$

Broyden's update

$$M_{k+1} = M_k + uv^\mathsf{T}$$

Define

$$s_k := x_{k+1} - x_k \quad \text{and} \quad y_k := g(x_{k+1}) - g(x_k).$$

Multiply the matrix update by s_k and use the MSE $M_{k+1}s_k = y_k$ to obtain

$$y_k = M_{k+1}s_k = M_k s_k + uv^\mathsf{T} s_k.$$

Hence, if $v^\mathsf{T} s_k \neq 0$, we obtain

$$u = \frac{y_k - M_k s_k}{v^\mathsf{T} s_k} \quad \text{and} \quad M_{k+1} = M_k + \frac{(y_k - M_k s_k)v^\mathsf{T}}{v^\mathsf{T} s_k}.$$

This equation determines a class of rank one updates that satisfy the MSE by choosing $v \in \mathbf{R}^n$ so that $v^\mathsf{T} s_k \neq 0$. An obvious choice for v is $s_k \neq 0$ yielding the *Broyden update*

$$M_{k+1} = M_k = \frac{(y_k - M_k s_k)s_k^\mathsf{T}}{s_k^\mathsf{T} s_k}.$$

Optimality of Broyden's Update

Theorem: Let $A \in \mathbf{R}^{n \times n}$, $s,\ y \in \mathbf{R}^n$, $s \neq 0$. The Broyden update

$$A_+ = A + \frac{(y - As)s^\mathsf{T}}{s^\mathsf{T} s}$$

is the unique solution to the problem

$$\min\{\|B - A\| : Bs = y\}.$$

Proof:

$$\begin{aligned} \|A_+ - A\| &= \|\frac{(y - As)s^\mathsf{T}}{s^\mathsf{T} s}\| = \|(B - A)\frac{ss^\mathsf{T}}{s^\mathsf{T} s}\| \\ &\leq \|B - A\|\,\|\frac{ss^\mathsf{T}}{s^\mathsf{T} s}\| \leq \|B - A\|. \end{aligned}$$

Broyden's Method

Algorithm:

Initialization: $x_0 \in \mathbf{R}^n$, $M_0 \in \mathbf{R}^{n\times n}$
Having (x_k, M_k) compute (x_{k+1}, M_{x+1}) as follows:
Solve $M_k s_k = -g(x_k)$ for s_k and set

$$
\begin{aligned}
x_{k+1} : &= x_k + s_k \\
y_k : &= g(x_k) - g(x_{k+1}) \\
M_{k+1} : &= M_k + \frac{(y_k - M_k s_k)s_k^{\mathsf{T}}}{s_k^{\mathsf{T}} s_k}.
\end{aligned}
$$

Broyden's Method

Algorithm:

Initialization: $x_0 \in \mathbf{R}^n$, $M_0 \in \mathbf{R}^{n \times n}$
Having (x_k, M_k) compute (x_{k+1}, M_{x+1}) as follows:
Solve $M_k s_k = -g(x_k)$ for s_k and set

$$\begin{aligned} x_{k+1} : &= x_k + s_k \\ y_k : &= g(x_k) - g(x_{k+1}) \\ M_{k+1} : &= M_k + \frac{(y_k - M_k s_k)s_k^{\mathsf{T}}}{s_k^{\mathsf{T}} s_k}. \end{aligned}$$

Inverse Updating: $M_k^{-1} = W_k$ where

$$W_{k+1} := W_k + \frac{(s_k - W_k y_k)s_k^{\mathsf{T}} W_k}{s_k^{\mathsf{T}} W_k y_k}$$

Matrix Secant Methods for Optimization

$$\mathcal{P} : \underset{x \in \mathbf{R}^n}{\text{minimize}}\, f(x)$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is C^2.

Goals:

1. Since M_k is intended to approximate $\nabla^2 f(x_k)$ it is desirable that M_k be symmetric.
2. Since we are concerned with minimization, then at least locally one can assume the second-order sufficiency condition holds. Consequently, we would like the M_k's to be positive definite.

Matrix Secant Methods for Optimization

$$\mathcal{P} : \underset{x \in \mathbf{R}^n}{\text{minimize}}\, f(x)$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is C^2.

Goals:

1. Since M_k is intended to approximate $\nabla^2 f(x_k)$ it is desirable that M_k be symmetric.
2. Since we are concerned with minimization, then at least locally one can assume the second-order sufficiency condition holds. Consequently, we would like the M_k's to be positive definite.

The Broyden update fails these conditions.

The BFGS Update

Suppose $M \in \mathcal{S}^n_{++}$ and $s, y \in \mathbf{R}^n \setminus \{0\}$.
Find $\overline{M} \in \mathcal{S}^n_{++}$ so that $\overline{M}s = y$.

The BFGS Update

Suppose $M \in \mathcal{S}^n_{++}$ and $s, y \in \mathbf{R}^n \setminus \{0\}$.
Find $\overline{M} \in \mathcal{S}^n_{++}$ so that $\overline{M}s = y$.

Assume $M = LL^T$ and $\overline{M} = JJ^T$ where both $L, J \in \mathbf{R}^{n\times n}$ are nonsingular.

The MSE implies that if

$$J^{\mathsf{T}}s = v \quad \text{then} \quad Jv = y.$$

Our approach is the apply the Broyden update to J and L giving

$$J = L + \frac{(y - Lv)v^{\mathsf{T}}}{v^{\mathsf{T}}v}.$$

Hence,

$$v = J^{\mathsf{T}}s = L^{\mathsf{T}}s + \frac{v(y - Lv)^{\mathsf{T}}s}{v^{\mathsf{T}}v}.$$

Hence $v = \alpha L^{\mathsf{T}}s$ for some $\alpha \in \mathbf{R}$.

The BFGS Update

Substituting this expression for v back in gives

$$\alpha L^{\mathsf{T}} s = L^{\mathsf{T}} s + \frac{\alpha L^{\mathsf{T}} s (y - \alpha L L^{\mathsf{T}} s)^{\mathsf{T}} s}{\alpha^2 s^{\mathsf{T}} L L^{\mathsf{T}} s}.$$

Hence

$$\alpha^2 = \left[\frac{s^{\mathsf{T}} y}{s^{\mathsf{T}} M s} \right].$$

That is, J exists only if $s^{\mathsf{T}} y > 0$ in which case

$$J = L + \frac{(y - \alpha M s) s^{\mathsf{T}} L}{\alpha s^{\mathsf{T}} M s}, \quad \text{with} \quad \alpha = \left[\frac{s^{\mathsf{T}} y}{s^{\mathsf{T}} M s} \right]^{1/2},$$

yielding

$$\overline{M} = M + \frac{y y^{\mathsf{T}}}{y^{\mathsf{T}} s} - \frac{M s s^{\mathsf{T}} M}{s^{\mathsf{T}} M s}.$$

$s^\top y > 0$

In the iterative context

$$s = s_k = -\lambda_k M_k^{-1} \nabla f(x_k) \quad \text{and} \quad y = y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

So

$$\begin{aligned} y^\top s = y_k^\top s_k &= \nabla f(x_{k+1})^\top s_k - \nabla f(x_k)^\top s_k \\ &= \lambda_k \nabla f(x_k + \lambda_k d_k)^\top d_k - \lambda_k \nabla f(x_k)^\top d_k \\ &= \lambda_k (\nabla f(x_k + \lambda_k d_k)^\top d_k - \nabla f(x_k)^\top d_k), \end{aligned}$$

where $d_k := -M_k^{-1} \nabla f(x_k)$. Since M_k is positive definite the direction d_k is a descent direction for f at x_k and so $\lambda_k > 0$. Thus, we need to show that $\lambda_k > 0$ can be chosen so that

$$\nabla f(x_k + \lambda_k d_k)^\top d_k \geq \beta \nabla f(x_k)^\top d_k$$

for some $\beta \in (0, 1)$.

The Inverse BFGS Update

$$
\begin{aligned}
M_k^{-1} &= W_k \\
&= W + \frac{(s - Wy)s^\mathsf{T} + s(s - Wy)^\mathsf{T}}{y^\mathsf{T} s} - \frac{(s - Wy)^\mathsf{T} y s s^\mathsf{T}}{(y^\mathsf{T} s)^2}.
\end{aligned}
$$