Matrix Secant Methods

Optimization

Outline

Broyden Updates

Given $g : \mathbb{R}^n \to \mathbb{R}^n$ solve $g(x) = 0$.

Algorithm: Broyden's Method

Initialization: $x^0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n \times n}$

Having (x^k, B_k) compute (x^{k+1}, B_{x+1}) as follows:

Solve $B_k s^k = -g(x^k)$ for s^k and set

$$\begin{aligned} x^{k+1} : &= x^k + s^k \\ y^k : &= g(x^k) - g(x^{k+1}) \\ B_{k+1} : &= B_k + \frac{(y^k - B_k s^k) {s^k}^T}{{s^k}^T s^k}. \end{aligned}$$

Broyden Updates

Algorithm: Broyden's Method (Inverse Updating)

Initialization: $x^0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n \times n}$

Having (x^k, B_k) compute (x^{k+1}, B_{x+1}) as follows:

Solve $s^k = -J_k g(x^k)$ for s^k and set

$$\begin{array}{rcl} x^{k+1} : & = & x^k + s^k \\ y^k : & = & g(x^k) - g(x^{k+1}) \\ J_{k+1} i & = & J_k + \dfrac{(s^k - J_k y^k) {s^k}^T J_k}{{s^k}^T J_k y}. \end{array}$$

MSE for Optimization

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

MSE for Optimization

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.
Solve $\nabla f(x) = 0$ using the Newton-Like iteration

$$x^{k+1} = x^k - M_k^{-1} \nabla f(x^k) \;.$$

MSE for Optimization

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.
Solve $\nabla f(x) = 0$ using the Newton-Like iteration

$$x^{k+1} = x^k - M_k^{-1} \nabla f(x^k) \; .$$

The MSE becomes

$$M_{k+1} s^k = y^k \; ,$$

where

$$s^k := x^{k+1} - x^k \text{ and } y^k := \nabla f(x^{k+1}) - \nabla f(x^k).$$

MSE for Optimization

The Broyden update gives

$$M_{k+1} = M_k + \frac{(y^k - M_k s^k){s^k}^T}{{s^k}^T s^k}.$$

MSE for Optimization

The Broyden update gives

$$M_{k+1} = M_k + \frac{(y^k - M_k s^k){s^k}^T}{{s^k}^T s^k}.$$

This is unsatisfactory for two reasons:

1. Since M_k approximates $\nabla^2 f(x^k)$ it must be symmetric.
2. Since we are minimizing, then M_k must be positive definite to insure that $s^k = -M_k^{-1}\nabla f(x^k)$ is a direction of descent for f at x^k.

SR1 Update

The Broyden class of updates is

$$M_{k+1} = M_k + \frac{(y^k - M_k s^k)v^T}{v^T s^k}.$$

SR1 Update

The Broyden class of updates is

$$M_{k+1} = M_k + \frac{(y^k - M_k s^k)v^T}{v^T s^k}.$$

The only symmetric member of this class is obtained by taking

$$v = (y^k - M_k s^k)$$

giving

$$M_{k+1} = M_k + \frac{(y^k - M_k s^k)(y^k - M_k s^k)^T}{(y^k - M_k s^k)^T s^k}.$$

The corresponding inverse update (by SMW) is

$$H_{k+1} = H_k + \frac{(s^k - H_k y^k)(s^k - H_k y^k)^T}{(s^k - H_k y^k)^T y^k} \ .$$

SR1 Update

The Broyden class of updates is

$$M_{k+1} = M_k + \frac{(y^k - M_k s^k)v^T}{v^T s^k}.$$

The only symmetric member of this class is obtained by taking

$$v = (y^k - M_k s^k)$$

giving

$$M_{k+1} = M_k + \frac{(y^k - M_k s^k)(y^k - M_k s^k)^T}{(y^k - M_k s^k)^T s^k}.$$

The corresponding inverse update (by SMW) is

$$H_{k+1} = H_k + \frac{(s^k - H_k y^k)(s^k - H_k y^k)^T}{(s^k - H_k y^k)^T y^k} \ .$$

But this update may not be positive definite.

Positive Definite Updating

Basic Problem:
Let $M \in \mathbb{R}^{n \times n}$ be a given symmetric positive definite (spd) matrix. Given $s, y \in \mathbb{R}^n$ find an spd matrix M_+ such that $M_+ s = y$ and M_+ is "close" to M.

Positive Definite Updating

Basic Problem:
Let $M \in \mathbb{R}^{n\times n}$ be a given symmetric positive definite (spd) matrix. Given $s, y \in \mathbb{R}^n$ find an spd matrix M_+ such that $M_+ s = y$ and M_+ is "close" to M.

We know that "close" cannot mean a rank one modification since the SR1 may not preserve positive definiteness.

Positive Definite Updating

Basic Problem:
Let $M \in \mathbb{R}^{n \times n}$ be a given symmetric positive definite (spd) matrix. Given $s, y \in \mathbb{R}^n$ find an spd matrix M_+ such that $M_+ s = y$ and M_+ is "close" to M.

We know that "close" cannot mean a rank one modification since the SR1 may not preserve positive definiteness.

Let's try another approach that combines the Broyden update with yet another important property of symmetric matrices.

Positive Definite Updating

Every spd matrix M can be written in the form $M = LL^T$ for some non-singular matrix L.

Positive Definite Updating

Every spd matrix M can be written in the form $M = LL^T$ for some non-singular matrix L.

Indeed, the LU factorization can be used to give the so called Cholesky factorization $M = LL^T$ where L is lower triangular.

Positive Definite Updating

Every spd matrix M can be written in the form $M = LL^T$ for some non-singular matrix L.

Indeed, the LU factorization can be used to give the so called Cholesky factorization $M = LL^T$ where L is lower triangular.

Similarly, if M_+ exists, then there must be a nonsingular matrix J such that $M_+ = JJ^T$.

Positive Definite Updating

So $M = LL^T$ and $M_+ = JJ^T$ with $M_+ s = y$.

Therefore, if we set $v = J^T s$, then $Jv = y$.

Positive Definite Updating

So $M = LL^T$ and $M_+ = JJ^T$ with $M_+ s = y$.

Therefore, if we set $v = J^T s$, then $Jv = y$.

We apply the Broyden update to $Jv = y$ and L to get

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

Positive Definite Updating

So $M = LL^T$ and $M_+ = JJ^T$ with $M_+ s = y$.

Therefore, if we set $v = J^T s$, then $Jv = y$.

We apply the Broyden update to $Jv = y$ and L to get

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

Using $v = J^T s$, we get

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

Positive Definite Updating

So $M = LL^T$ and $M_+ = JJ^T$ with $M_+ s = y$.

Therefore, if we set $v = J^T s$, then $Jv = y$.

We apply the Broyden update to $Jv = y$ and L to get

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

Using $v = J^T s$, we get

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

But then $v = \alpha L^T s$ for some $\alpha \in \mathbb{R}$.

Positive Definite Updating

Hence

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

with $v = \alpha L^T s$.

Positive Definite Updating

Hence

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

with $v = \alpha L^T s$.

Plugging in we get

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

Positive Definite Updating

Hence

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

with $v = \alpha L^T s$.

Plugging in we get

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

Hence

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

Broyden-Fletcher-Goldfarb-Shanno (BFGS) Updating

Therefore, this update only exists if $s^T y > 0$ in which case

$$J = L + \frac{(y - \alpha M s)s^T L}{\alpha s^T M s},$$

with

$$\alpha = \left[\frac{s^T y}{s^T M s}\right]^{1/2},$$

yielding

$$M_+ = M + \frac{yy^T}{y^T s} - \frac{Mss^T M}{s^T M s}.$$

This is called the BFGS update. It is currently viewed as the best MSE update available for optimization due to its outstanding performace in practise.

BFGS Updating

In addition, we note that the M_+ can be obtained directly from the matrices J.

If the QR factorization of J^T is $J^T = QR$, we can set $L_+ = R$ yielding

$$M_+ = JJ^T = R^T Q^T QR = L_+ L_+^T.$$

Inverse BFGS Updating

Sherman-Morrison-Woodbury formula again gives the inverse update

$$H_+ = H + \frac{(s + Hy)^T yss^T}{(s^T y)^2} - \frac{Hys^T + sy^T H}{s^T y} ,$$

where $H = M^{-1}$.

How to insure $s^T y > 0$

The BFGS update preserves both symmetry and positive definiteness when ${s^k}^T y^k > 0$. But is this condition reasonable or even implementable?

How to insure $s^T y > 0$

The BFGS update preserves both symmetry and positive definiteness when ${s^k}^T y^k > 0$. But is this condition reasonable or even implementable?

Recall that

$$y = y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

and

$$s^k = x^{k+1} - x^k = t_k d^k \ ,$$

where

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

is the search direction and $t_k > 0$ is the stepsize.

How to insure $s^T y > 0$

The BFGS update preserves both symmetry and positive definiteness when ${s^k}^T y^k > 0$. But is this condition reasonable or even implementable?

Recall that

$$y = y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

and

$$s^k = x^{k+1} - x^k = t_k d^k \ ,$$

where

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

is the search direction and $t_k > 0$ is the stepsize.

Hence

$$\begin{aligned} {y^k}^T s^k &= \nabla f(x^{k+1})^T s^k - \nabla f(x^k)^T s^k \\ &= t_k(\nabla f(x^k + t_k d_k)^T d^k - \nabla f(x^k)^T d^k) \ , \end{aligned}$$

How to insure $s^T y > 0$

$$
\begin{aligned}
{y^k}^T s^k &= \nabla f(x^{k+1})^T s^k - \nabla f(x^k)^T s^k \\
&= t_k(\nabla f(x^k + t_k d_k)^T d^k - \nabla f(x^k)^T d^k) \ ,
\end{aligned}
$$

How to insure $s^T y > 0$

$$\begin{aligned} {y^k}^T s^k &= \nabla f(x^{k+1})^T s^k - \nabla f(x^k)^T s^k \\ &= t_k(\nabla f(x^k + t_k d_k)^T d^k - \nabla f(x^k)^T d^k) \ , \end{aligned}$$

Since H_k is sdp, $\nabla f(x^k)^T d^k < 0$ so $t_k > 0$.

How to insure $s^T y > 0$

$$
\begin{aligned}
{y^k}^T s^k &= \nabla f(x^{k+1})^T s^k - \nabla f(x^k)^T s^k \\
&= t_k(\nabla f(x^k + t_k d_k)^T d^k - \nabla f(x^k)^T d^k) \ ,
\end{aligned}
$$

Since H_k is sdp, $\nabla f(x^k)^T d^k < 0$ so $t_k > 0$.
Therefore, to get ${s^k}^T y^k > 0$ we must show $t_k > 0$ can be choosen so that

$$
\nabla f(x^k + t_k d^k)^T d^k \geq \beta \nabla f(x^k)^T d^k
$$

for some $\beta \in (0, 1)$ since in this case

$$
\nabla f(x^k + t_k d_k)^T d^k - \nabla f(x^k)^T d^k \geq (\beta - 1)\nabla f(x^k)^T d^k > 0.
$$

How to insure $s^T y > 0$

$$\begin{aligned} {y^k}^T s^k &= \nabla f(x^{k+1})^T s^k - \nabla f(x^k)^T s^k \\ &= t_k(\nabla f(x^k + t_k d_k)^T d^k - \nabla f(x^k)^T d^k) \ , \end{aligned}$$

Since H_k is sdp, $\nabla f(x^k)^T d^k < 0$ so $t_k > 0$.
Therefore, to get ${s^k}^T y^k > 0$ we must show $t_k > 0$ can be choosen so that

$$\nabla f(x^k + t_k d^k)^T d^k \geq \beta \nabla f(x^k)^T d^k$$

for some $\beta \in (0, 1)$ since in this case

$$\nabla f(x^k + t_k d_k)^T d^k - \nabla f(x^k)^T d^k \geq (\beta - 1)\nabla f(x^k)^T d^k > 0.$$

But this precisely the second condition in the weak Wolfe conditions with $\beta = c_2$. Hence a successful BFGS update can always be obtained.