Quasi-Newton method

link of references:

1. 牛顿法和拟牛顿法 basic assumption of Quasi-Newton
2. 牛顿法和拟牛顿法 types of Quasi-Newton method
3. BFGS算法的迭代公式推导 derive BFGS with Sherman-Morrison formula
4. Sherman–Morrison formula
5. 拟牛顿法：SR1，BFGS，DFP，DM条件 math explanation of Quasi-Newton methods

review of Newton method

$$\nabla f\left(x_{k+1}\right) \approx \nabla f\left(x_k\right)+H\left(x_k\right)\left(x_{k+1}-x_k\right)$$

$f(x)$ is local minimal at $x=x_{k+1}$ when

1. $H\left(x_k\right)$ is positive definite
2. $\nabla f\left(x_{k+1}\right)=0$

Thus

$$0=\nabla f\left(x_k\right)+H\left(x_k\right)\left(x_{k+1}-x_k\right) \Longrightarrow x_{k+1}=x_k-H^{-1}\left(x_k\right) \nabla f\left(x_k\right)$$

assumption of Quasi-Newton

H^{-1} is complicated to compute, we find other form G to replace it

define $\delta_k \equiv x_{k+1}-x_k$, and $y_k \equiv g_{k+1}-g_k \equiv \nabla f(x_{k+1})-\nabla f(x_k)$

It still holds

$$\nabla f\left(x_{k+1}\right) \approx \nabla f\left(x_k\right)+H\left(x_k\right)\left(x_{k+1}-x_k\right)$$

A. thus, replace $H^{-1}\left(x_k\right)$ with G_{k+1} to derive the constraint

$$\delta_k \approx G_{k+1} y_k$$

B. modify the Newton method, here G_k is corresponding to the $H^{-1}\left(x_{k-1}\right)$, then replace $H^{-1}\left(x_k\right)$ with $\lambda_k H^{-1}\left(x_{k-1}\right) \approx \lambda_k G_k$

$$x_{k+1}=x_k-\lambda_k G_k g_k$$

moreover, the λ_k is given by

$$\lambda_k \equiv \underset{\lambda_k}{\operatorname{argmin}} f(x_k - \lambda_k G_k g_k), \quad st. \lambda_k \geq 0 \Longrightarrow \frac{df}{d\lambda_k} = (G_k g_k)^T g_{k+1} = g_k^T G_k g_{k+1} = 0$$

recall that:

1. we only care about the direction of x changes δ_k, that is $-G_k g_k$
2. we have to make sure G_k is always positive definite
3. we have to make sure G_{k+1} satisfies $\delta_k \approx G_{k+1} y_k$

type 1: DFP method

Use G_{k+1} to estimate $H^{-1}(x_k)$

1. G_k is positive definite
2. satisfy $\delta_k = G_{k+1} y_k$

Derive the formula of G_{k+1}

a. set up $G_{k+1} = G_k + uu^T - vv^T$

$$\delta_k = G_{k+1} y_k = G_k y_k + u(u^T y_k) + v(v^T y_k)$$

now we could set

$$\begin{aligned} \delta_k &= u(u^T y_k) \\ -G_k g_k &= -v(v^T y_k) \end{aligned}$$

Thus we have $u = k_1 \delta_k, v = k_2 G_k y_k$, solve for k_1, k_2 by comparing coefficients

$$\begin{aligned} 1 &= k_1^2(\delta_k^T y_k) \\ 1 &= k_2^2(y_k^T G_k y_k) \end{aligned}$$

In the end

$$G_{k+1} = G_k + k_1^2 \delta_k \delta_k^T - k_2^2 G_k y_k y_k^T G_k^T = G_k + \frac{1}{\delta_k^T y_k} \delta_k \delta_k^T - \frac{1}{y_k^T G_k y_k} G_k y_k y_k^T G_k^T$$

notice B. $\delta_k = -\lambda_k G_k g_k, g_k^T G_k g_{k+1} = 0$, so that if we want to ensure $G_k \Rightarrow G_{k+1}$ positive definite, have to make sure

$$\delta_k^T y_k = -\lambda_k g_k^T G_k^T (g_{k+1} - g_k) = +\lambda_k g_k^T G_k g_k > 0$$

It can be proved if $G_0 = I$ is positive definite, then G_k is positive definite,

For any X, first part

$$\frac{1}{\delta_k^T y_k} X^T \delta_k \delta_k^T X = \frac{1}{\delta_k^T y_k} (\delta_k^T X)^2 \geq 0$$

For any X, the second part

$$X^T \left[G_k - \frac{1}{y_k^T G_k y_k} G_k y_k y_k^T G_k^T \right] X = \frac{1}{y_k^T G_k y_k} [(X^T G_k X)(y_k^T G_k y_k) - (X^T G_k y_k)^2] = \frac{1}{y_k^T G_k y_k} [(X'^T \Lambda_k X')(y_k'^T \Lambda_k y_k') - (X'^T \Lambda_k y_k')^2]$$

here the diagonal matrix $\Lambda_k = \text{diag}(\lambda_{k1}, \cdots, \lambda_{kN})$ shape (N, N), because G_k positive definite, we have all $\lambda_{kn} \geq 0$

$$= \frac{1}{y_k^T G_k y_k}\left[\sum_{n=1}^{N}(\lambda_{kn}x_n'^2)\sum_{n=1}^{N}(\lambda_{kn}y_{kn}'^2) - \sum_{n=1}^{N}(\lambda_{kn}x_n'y_{kn}')^2\right] = \frac{1}{y_k^T G_k y_k}\sum_{n=1}^{N}\sum_{n'=1}^{N}\lambda_{kn}\lambda_{kn'}(x_n'x_{n'}' - y_{kn}'y_{kn'}')^2 \geq 0$$

To sum up, G_{k+1} is positive definite when G_k is positive definite

type2: BFGS method

That is better than DFP

step 1. Use B_{k+1} to estimate $H(x_k)$

1. B_k is positive definite
2. satisfy $y_k = B_{k+1}\delta_k$

step 2. Use $G_{k+1} = B_{k+1}^{-1}$ to get $H^{-1}(x_k)$

Derive the formula of G_{k+1}

For step 1. similarly from $y_k = B_{k+1}\delta_k$

set $B_{k+1} = B_k + uu^T - vv^T$

$$B_{k+1} = B_k + k_1^2 y_k y_k^T - k_2^2 B_k \delta_k \delta_k^T B_k^T = B_k + \frac{1}{y_k^T\delta_k}y_k y_k^T - \frac{1}{\delta_k^T B_k \delta_k}B_k\delta_k\delta_k^T B_k^T$$

notice B. $\delta_k = -\lambda_k G_k g_k, g_k^T G_k g_{k+1} = 0$, so we have $g_k = -\frac{1}{\lambda_k}B_k\delta_k, \delta_k^T B_k \delta_{k+1} = 0$, similarly when $B_0 = I$ is positive definite, we can prove B_k is positive definite too, thus $G_k = B_k^{-1}$ is positive definite

For step 2. with the Sherman–Morrison formula

$$\left(A + \frac{uv^T}{t}\right)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{t + v^T A^{-1}u}$$

Apply it for two times, notice $B^T = B$

$$G_{k+1} \equiv B_{k+1}^{-1} = \left(B_k + \frac{y_k y_k^T}{y_k^T\delta_k} - \frac{B_k\delta_k\delta_k^T B_k^T}{\delta_k^T B_k \delta_k}\right)^{-1}$$

1st time

$$= \left(B_k + \frac{y_k y_k^T}{y_k^T\delta_k}\right)^{-1} + \left(B_k + \frac{y_k y_k^T}{y_k^T\delta_k}\right)^{-1}\frac{B_k\delta_k\delta_k^T B_k}{\delta_k^T B_k\delta_k - \delta_k^T B_k\left(B_k + \frac{y_k y_k^T}{y_k^T\delta_k}\right)^{-1}B_k\delta_k}\left(B_k + \frac{y_k y_k^T}{y_k^T\delta_k}\right)^{-1}$$

2nd time

$$
\begin{aligned}
&= \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) + \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) \frac{B_k \delta_k \delta_k^T B_k}{\delta_i^T B_k \delta_k - \delta_k^T B_k \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) B_k \delta_k} \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) \\
&= \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) + \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) \frac{B_k \delta_k \delta_k^T B_k}{\frac{\delta_k^T y_k y_k^T \delta_k}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}} \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) \\
&= \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) + \frac{B_k^{-1} B_k \delta_k \delta_k^T B_k B_k^{-1}}{\frac{\delta_k^T y_k y_k^T \delta_k}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}} - \frac{B_k^{-1} B_k \delta_k \delta_k^T B_k}{\frac{\delta_k^T y_k y_k^T \delta_k}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}} \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k} \\
&- \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k} \frac{B_k \delta_k \delta_k^T B_k}{\frac{\delta_k^T y_k y_k^T \delta_k}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}} B_k^{-1} + \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k} \frac{B_k \delta_k \delta_k^T B_k}{\frac{\delta_k^T y_k y_k^T \delta_k}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}} \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k} \\
&= \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) + \frac{\delta_k \delta_k^T \left(y_k^T \delta_k + y_k^T B_k^{-1} y_k\right)}{\delta_k^T y_k y_k^T \delta_k} - \frac{\delta_k \delta_k^T y_k y_k^T B_k^{-1}}{\delta_k^T y_k y_k^T \delta_k} - \frac{B_k^{-1} y_k y_k^T \delta_k \delta_k^T}{\delta_k^T y_k y_k^T \delta_k} \\
&+ \frac{B_k^{-1} y_k \left(y_k^T \delta_k \delta_k^T y_k\right) y_k^T B_k^{-1}}{\left(y_k^T \delta_k + y_k^T B_k^{-1} y_k\right) \left(\delta_k^T y_k y_k^T \delta_k\right)} \\
&= \left(B_k^{-1} - \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{y_k^T \delta_k + y_k^T B_k^{-1} y_k}\right) + \frac{\delta_k \delta_k^T \left(y_k^T \delta_k + y_k^T B_k^{-1} y_k\right)}{\left(\delta_k^T y_k\right)^2} - \frac{\delta_k \left(\delta_k^T y_k\right) y_k^T B_k^{-1}}{\left(\delta_k^T y_k\right)^2} - \frac{B_k^{-1} y_k \left(y_k^T \delta_k\right) \delta_k^T}{\left(\delta_k^T y_k\right)^2} \\
&+ \frac{B_k^{-1} y_k y_k^T B_k^{-1}}{\left(y_k^T \delta_k + y_k^T B_k^{-1} y_k\right)}
\end{aligned}
$$

then

$$
\begin{aligned}
&= B_k^{-1} + \frac{\delta_k \delta_k^T \left(y_k^T \delta_k\right)}{\left(\delta_k^T y_k\right)^2} + \frac{\delta_k \delta_k^T \left(y_k^T B_k^{-1} y_k\right)}{\left(\delta_k^T y_k\right)^2} - \frac{\delta_k y_k^T B_k^{-1}}{\delta_k^T y_k} - \frac{B_k^{-1} y_k \delta_k^T}{\delta_k^T y_k} \\
&= B_k^{-1} - \frac{B_k^{-1} y_k \delta_k^T}{\delta_k^T y_k} - \frac{\delta_k y_k^T B_k^{-1}}{\delta_k^T y_k} + \frac{\delta_k \left(y_k^T B_k^{-1} y_k\right) \delta_k^T}{\left(\delta_k^T y_k\right)^2} + \frac{\delta_k \delta_k^T}{\delta_k^T y_k} \\
&= B_k^{-1} \left(I - \frac{y_k \delta_k^T}{\delta_k^T y_k}\right) - \frac{\delta_k y_k^T B_k^{-1}}{\delta_k^T y_k} \left(I - \frac{y_k \delta_k^T}{\delta_k^T y_k}\right) + \frac{\delta_k \delta_k^T}{\delta_k^T y_k} \\
&= \left(I - \frac{\delta_k y_k^T}{\delta_k^T y_k}\right) B_k^{-1} \left(I - \frac{y_k \delta_k^T}{\delta_k^T y_k}\right) + \frac{\delta_k \delta_k^T}{\delta_k^T y_k} \\
&= \left(I - \frac{\delta_k y_k^T}{\delta_k^T y_k}\right) B_k^{-1} \left(I - \frac{\delta_k y_k^T}{\delta_k^T y_k}\right)^T + \frac{\delta_k \delta_k^T}{\delta_k^T y_k}
\end{aligned}
$$

Eventually

$$
G_{k+1} = \left(I - \frac{\delta_k y_k^T}{\delta_k^T y_k}\right) G_k \left(I - \frac{\delta_k y_k^T}{\delta_k^T y_k}\right)^T + \frac{\delta_k \delta_k^T}{\delta_k^T y_k}
$$

That is the iterative formula for G_k

1. if $G_0 = I$ is positive definite, then G_k is positive definite

2. satisfy $\delta_k = G_{k+1} y_k$