M05M11084 最优化理论、算法与应用

5 无约束优化方法 III

无约束优化方法 III

参考：

1. Numerical optimization, Chapter 8, Jorge Nocedal Stephen J. Wright
2. 最优化导论，第11章， Edwin K.P. Chong, Stanislaw H. Żak 著, 孙志强等译
3. Practical Optimization Algorithms and Engineering Applications, Chapter 7, A. Antoniou, W. LU

- 直接法
  - 线性最小二乘法
  - 坐标轮换法
  - 共轭方向法
    - GS
    - Powell方法
    - Powell改进
  - Downhill Simplex
  - 模式搜索
- 一阶导数
  - 梯度下降法
    - 固定步长梯度下降法
    - 最速梯度下降法
  - 共轭梯度法
    - FR
    - PRP
- 二阶导数
  - 牛顿法
  - 阻尼牛顿法
  - 拟牛顿法
    - DFP
    - BFGS
    - LBFGS
  - 高斯牛顿法
  - LM
- 信赖域
- 全局优化法
  - 遗传算法、模拟退火、粒子群

1. 拟牛顿法的概念与基本算法
2. 秩1修正法
3. DFP方法
4. BFGS方法
5. Broyden族

1. 拟牛顿法的概念与基本算法
   ① 拟牛顿方向的下降性
   ② 逼近Hesse阵逆矩阵的条件
   ③ 拟牛顿法与共轭方向法的关系
   ④ 构造逼近矩阵的基本方法
2. 秩1修正法
3. DFP方法
4. BFGS方法
5. Broyden族　DFP与BFGS的线性组合

REVIEW

牛顿法

$d_k = -H_k^{-1} g_k$，$H_k \succ 0$　牛顿方向

$x^{k+1} = x^k + \alpha_k d_k$

$\alpha_k = \arg\min_{\alpha \ge 0} f\left(x^k - \alpha H_k^{-1} g_k\right)$

$H_k^{-1} = H^{-1}\left(x^k\right)$

$g_k = \nabla f\left(x^k\right)$

$q(\delta) = f\left(x^k\right) + g_k^T \delta + \frac{1}{2} \delta^T H_k \delta$

$0 = Dq(\delta) = g_k^T + \delta^T H_k$

$0 = \nabla q(\delta) = g_k + H_k \delta$

$\delta = -H_k^{-1} g_k$

牛顿法的问题:

- ✓ $f \in C^2$
- ✓ 计算Hesse阵及其逆矩阵
- ✓ 当Hesse阵非正定时，牛顿方向不是下降方向，需要修正

思考：可否不计算Hesse矩阵，直接构造一个类似于 Hesse逆矩阵 的矩阵?

$$Q \xrightarrow{\text{逼近}} H^{-1}$$

拟牛顿法的基本原理

$$d_k = -Q_k g_k, Q_k \succ 0 \qquad \text{拟牛顿方向}$$
$$x^{k+1} = x^k + \alpha_k d_k$$
$$\alpha_k = \arg\min_{\alpha \ge 0} f\left(x^k + \alpha d_k\right)$$

特殊地，

$$d_k = \begin{cases} -g_k & Q_k = I \quad \text{最速下降方向} \\ -H^{-1}\left(x^k\right) g_k & Q_k = H^{-1}\left(x^k\right) \quad \text{牛顿方向} \end{cases}$$

① 拟牛顿方向是下降的?
② 正定阵有很多，如何选择Q？构造逼近Hesse阵逆矩阵的条件?
③ 拟牛顿法与共轭方向法的关系?
④ 具体构造方法?

①正定矩阵与梯度可以构造一个下降方向

设 $f \in C^1, x^k \in \mathcal{R}^n, g_k = \nabla f\left(x^k\right) \neq 0,\ Q_k \succ 0, Q_k \in \mathcal{R}^{n \times n}$
若取 $x^{k+1} = x^k - \alpha_k Q_k g_k$，其中 $\alpha_k = \arg\min_{\alpha \ge 0} f\left(x^k - \alpha Q_k g_k\right)$，
那么， $\alpha_k > 0$ 且 $f\left(x^{k+1}\right) < f\left(x^k\right)$.

证明
$$x^{k+1} = x^k - \alpha Q_k g_k,\ \alpha > 0$$

$f\left(x^{k+1}\right)$在x^k处作泰勒展开

$$f\left(x^{k+1}\right) = f\left(x^k\right) - \alpha g_k^T Q_k g_k + o(\|Q_k g_k\| \alpha)$$

当$\alpha \to 0_+$， $Q_k \succ 0$， $g_k^T Q_k g_k > 0$

$$f\left(x^{k+1}\right) < f\left(x^k\right) \quad \Longrightarrow \text{拟牛顿方向 } d_k = -Q_k g_k \text{ 下降方向}$$

正定阵很多，选择逼近“Hesse阵逆矩阵”的矩阵

②选择逼近逆Hesse阵的矩阵

逼近阵要满足的割线方程

考虑二次型 $f(x)=\frac{1}{2}x^THx+x^Tb+c,\quad H\succ 0$

$g_{k+1}=Hx^{k+1}+b,\quad g_k=Hx^k+b \qquad g_k=\nabla f(x^k)$

割线方程 $g_{k+1}-g_k=H(x^{k+1}-x^k)$

$y_k=Hs_k$

$y_k=g_{k+1}-g_k$

$s_k=x^{k+1}-x^k$

割线方程 $y_k=Hs_k,\qquad k=0,1,\dots,n-1$

写成矩阵形式 $[y_0\quad y_1\quad\cdots\quad y_{n-1}]=H[s_0\quad s_1\quad\cdots\quad s_{n-1}]$

如果由x^0开始迭代得到的向量组$s_0,s_1,\dots,s_{n-1}$线性无关，则

$$H=[y_0\quad y_1\quad\cdots\quad y_{n-1}][s_0\quad s_1\quad\cdots\quad s_{n-1}]^{-1}$$

构造逼近矩阵Q_k 满足割线方程

$$Q_{k+1}y_i=s_i,\qquad i=0,1,\dots,k$$

观察$k=n-1$时，方程组

$Q_ny_i=s_i,\qquad i=0,1,\dots,n-1$ ✓ 逼近阵需要满足的条件

$Q_n[y_0\quad y_1\quad\cdots\quad y_{n-1}]=[s_0\quad s_1\quad\cdots\quad s_{n-1}]$ 写成矩阵形式

精确相等， $Q_n=H^{-1}$

$$[Q_n]^{-1}=[y_0\quad y_1\quad\cdots\quad y_{n-1}][s_0\quad s_1\quad\cdots\quad s_{n-1}]^{-1}=H$$

③二次型拟牛顿方向是共轭的 定理

对Hessian阵为$H=H^T$的二次型应用拟牛顿法，对于$k=0,1,\dots,n-1$，有，

$$Q_{k+1}y_i=s_i,\ i=0,1,\dots,k$$

其中 $Q_{k+1}=Q_{k+1}^T$. 如果$\alpha_i\neq 0$，$i=0,1,\dots,k$, 那么$d_0,\cdots,d_{k+1}$是H-共轭的

$$d_k=-Q_kg_k, Q_k\succ 0$$
$$x^{k+1}=x^k+\alpha_kd_k$$
$$\alpha_k=\arg\min_{\alpha\ge 0} f\left(x^k+\alpha d_k\right)$$

$$g_k=\nabla f\left(x^k\right)$$
$$y_k=g_{k+1}-g_k$$
$$s_k=x^{k+1}-x^k=\alpha_kd_k$$
$$y_k=Hs_k$$

证明 归纳法

$$x^1=x^0+\alpha_0d_0$$
$$\Rightarrow d_0=\frac{s_0}{\alpha_0},\alpha_0\neq 0$$

$$d_1^THd_0=-g_1^TQ_1Hd_0$$
$$=-g_1^TQ_1\frac{Hs_0}{\alpha_0}$$
$$=-g_1^TQ_1\frac{y_0}{\alpha_0}$$
$$=-g_1^T\frac{Q_1y_0}{\alpha_0}\quad=-g_1^T\frac{s_0}{\alpha_0}\quad=-g_1^Td_0\overset{?}{=}0$$

REVIEW

根据上章二次型共轭方向法中，搜索方向与梯度的正交关系

在二次型共轭方向法中，对于所有k，$0\le k\le n-1$，$0\le i\le k$，都有
$$d_i^Tg_{k+1}=0$$

所以，$d_1^THd_0=-g_1^Td_0=0$

③二次型拟牛顿方向是共轭的 定理

对Hessian阵为$H=H^T$的二次型应用拟牛顿法，对于$k=0,1,\dots,n-1$，有，

$$Q_{k+1}y_i=s_i,\ i=0,1,\dots,k$$

其中 $Q_{k+1}=Q_{k+1}^T$. 如果 $\alpha_i\neq 0$，$i=0,1,\dots,k$, 那么$d_0,\cdots,d_{k+1}$是 H-共轭的

$$d_k=-Q_kg_k, Q_k\succ 0$$
$$x^{k+1}=x^k+\alpha_kd_k$$
$$\alpha_k=\arg\min_{\alpha\geq 0} f\left(x^k+\alpha d_k\right)$$

$$g_k=\nabla f\left(x^k\right)$$
$$y_k=g_{k+1}-g_k$$
$$s_k=x^{k+1}-x^k$$
$$y_k=Hs_k$$

证明　归纳法

$$d_1^THd_0=-g_1^Td_0=0$$

设定理在$k-1$（$k<n-1$）时成立，　证明在k时也成立

即，若 $Q_{k+1}y_i=s_i$，$0\leq i\leq k-1$ 则 $d_0,\cdots,d_k$是 H-共轭的 $d_j^THd_i=0,\quad 0\leq i\neq j\leq k$ $\Longrightarrow$ 若 $Q_{k+1}y_i=s_i$，$0\leq i\leq k$，则 $d_0,\cdots,d_{k+1}$是 H-共轭的 $d_j^THd_i=0,\quad 0\leq i\neq j\leq k+1$

只需证明 $d_{k+1}^THd_i=0$，$0\leq i\leq k+1$

只需证明$d_{k+1}^THd_i=0, i=0,\dots,k$

$$\begin{aligned}
d_{k+1}^THd_i&=-g_{k+1}^TQ_{k+1}Hd_i\\
&=-g_{k+1}^TQ_{k+1}\frac{Hs_i}{\alpha_i}\\
&=-g_{k+1}^TQ_{k+1}\frac{y_i}{\alpha_i}\\
&=-g_{k+1}^T\frac{Q_{k+1}y_i}{\alpha_i}\\
&=-g_{k+1}^T\frac{s_i}{\alpha_i}\\
&=-g_{k+1}^Td_i=0
\end{aligned}$$

搜索方向与梯度的正交关系

$$d_k=-Q_kg_k$$
$$x^{k+1}=x^k+\alpha_kd_k$$
$$s_k=x^{k+1}-x^k=\alpha_kd_k$$
$$d_k=\frac{s_k}{\alpha_k}$$

$$g_k=\nabla f\left(x^k\right)$$
$$y_k=g_{k+1}-g_k$$
$$y_k=x^{k+1}-x^k$$
$$y_k=Hs_k$$

$$Q_{k+1}y_i=s_i$$

所以，$d_0,\cdots,d_{k+1}$是 H-共轭的

据此定理, 对n维二次型应用拟牛顿法最多只需迭代n次

如何构造矩阵Q？
$Q\succ 0\ \to H^{-1}$

④校正矩阵的构造方法 逼近 Hesse阵的逆矩阵

- ✓ 满足逼近阵条件：割线方程
- ✓ 每次迭代产生
- ✓ 新逼近阵Q_k = 当前逼近阵Q_{k-1} + 修正矩阵C_k

设想 $Q_k = Q_{k-1} + C_k$　　$Q_k \triangleq Q(x^k)$

$Q_{k-1} \succ 0 \ \rightarrow H_{k-1}^{-1}$　　$C_k \triangleq C(x^k)$

C_k点x^k处的校正矩阵 由函数f在x^k处的低于二阶的信息构成，且，信息尽可能少

使用尽可能少的信息构造矩阵：$zz^T, z \in \mathcal{R}^n$

$\operatorname{rank} zz^T = 1$　秩1矩阵　对称矩阵

秩1修正 $C_k = zz^T$

秩2修正 $C_k = zz^T + ww^T$

1. 拟牛顿法的概念与基本算法
2. 秩1修正法
3. DFP方法
4. BFGS法
5. Broyden族

2. 秩1修正公式　Rank One Correction Formula

迭代公式　$Q_{k+1} = Q_k + \underbrace{\alpha_k z_k z_k^T}_{= C_k}$　设 $\alpha_k \in \mathcal{R}$，$z_k \in \mathcal{R}^n$

若 Q_k 对称，则 Q_{k+1}也是对称的

给定 Q_k，确定α_k，z_k

$$Q_{k+1} y_i = s_i \xrightarrow{i = k} \boxed{Q_{k+1} y_k = s_k}$$

$$\downarrow Q_{k+1} = Q_k + \alpha_k z_k z_k^T$$

$$(Q_k + \alpha_k z_k z_k^T) y_k = s_k$$

$$\downarrow$$

$$\alpha_k z_k (z_k^T y_k) = s_k - Q_k y_k$$

$$\downarrow$$

$$z_k = \frac{s_k - Q_k y_k}{\alpha_k \, (z_k^T y_k)}$$

$$Q_{k+1} = Q_k + \alpha_k z_k z_k^T$$

$$z_k = \frac{s_k - Q_k y_k}{\alpha_k \, (z_k^T y_k)}$$

$$\alpha_k z_k z_k^T = \alpha_k \frac{s_k - Q_k y_k}{\alpha_k \, (z_k^T y_k)} \left[\frac{s_k - Q_k y_k}{\alpha_k \, (z_k^T y_k)}\right]^T$$

$$= \frac{(s_k - Q_k y_k)(s_k - Q_k y_k)^T}{\alpha_k (z_k^T y_k)^2}$$

$$Q_{k+1} = Q_k + \frac{(s_k - Q_k y_k)(s_k - Q_k y_k)^T}{\alpha_k (z_k^T y_k)^2}$$

?

$$Q_{k+1} = Q_k + \frac{(s_k - Q_k y_k)(s_k - Q_k y_k)^T}{\alpha_k (z_k^T y_k)^2}$$

$$\alpha_k z_k (z_k^T y_k) = s_k - Q_k y_k$$

Pre-multiply y_k^T

$$y_k^T\left(\alpha_k z_k (z_k^T y_k)\right) = y_k^T (s_k - Q_k y_k)$$

$$\alpha_k \left(z_k^T y_k\right)^2 = y_k^T (s_k - Q_k y_k)$$

$$Q_{k+1} = Q_k + \frac{(s_k - Q_k y_k)(s_k - Q_k y_k)^T}{y_k^T (s_k - Q_k y_k)}$$

Rank One Algorithm

Given starting point x^0, convergence tolerance $\varepsilon > 0$,
a real symmetric positive definite $Q_0 = I$;
$k \leftarrow 0$;
while $\|\nabla f(x^k)\| > \varepsilon$;
Compute search direction
$d_k = -Q_k \nabla f(x^k)$
Set $x^{k+1} = x^k + \alpha_k d_k$ where α_k is an Inexact step size following Wolfe criterion or an exact step size ;
Define $s_k = \alpha_k d_k$ and $y_k = \nabla f(x^{k+1}) - \nabla f(x^k)$;
$v_k = s_k - Q_k y_k$
Compute

$$Q_{k+1} = Q_k + \frac{v_k v_k^T}{y_k^T v_k}$$

$k \leftarrow k + 1$;
end (while)

定理 二次型的秩1算法满足割线方程

将秩1算法应用到二次型问题中，其中，Hessian阵 $H = H^T$, 有 $Q_{k+1}y_i = s_i$, $0 \le i \le k$.

证明 $k = 0$时 $Q_1 y_0 = s_0$

$$Q_1 = Q_0 + \frac{(s_0 - Q_0 y_0)(s_0 - Q_0 y_0)^T}{y_0^T(s_0 - Q_0 y_0)}$$

$$Q_1 y_0 = Q_0 y_0 + \frac{(s_0 - Q_0 y_0)(s_0 - Q_0 y_0)^T y_0}{y_0^T(s_0 - Q_0 y_0)}$$

$$= Q_0 y_0 + s_0 - Q_0 y_0$$

$$= s_0$$

$$Q_{k+1} = Q_k + \frac{(s_k - Q_k y_k)(s_k - Q_k y_k)^T}{y_k^T(s_k - Q_k y_k)}$$

$d_0 = -Q_0 g_0$ $\quad$ $d_k = -Q_k g_k$

$g_k = \nabla f(x^k)$

$x^{k+1} = x^k + \alpha_k d_k$ $\quad$ $y_k = g_{k+1} - g_k$

$s_k = x^{k+1} - x^k = \alpha_k d_k$ $\quad$ $s_k = x^{k+1} - x^k$

$y_k = H s_k$

$d_k = \frac{s_k}{\alpha_k}$

$g_k = H x^k + b$

$g_{k+1} = H x^{k+1} + b$

$$Q_{k+1} = Q_k + \frac{(s_k - Q_k y_k)(s_k - Q_k y_k)^T}{y_k^T(s_k - Q_k y_k)}$$

将秩1算法应用到二次型问题中，其中，Hessian阵 $H = H^T$, 有 $Q_{k+1}y_i = s_i$, $0 \le i \le k$.

证明 $k = 0$时 $Q_1 y_0 = s_0$

假设定理在$k-1 \ge 0$时成立，即，$Q_k y_i = s_i$, $0 \le i \le k-1$ $\implies$ 定理在k时也成立

秩1公式推导时使用过 $Q_{k+1}y_k = s_k$ 自然成立

只需证明 $Q_{k+1}y_i = s_i$, $0 \le i < k$

$$Q_{k+1}y_i = \underbrace{Q_k y_i}_{= s_i \text{ 由假设}} + \frac{(s_k - Q_k y_k)(s_k - Q_k y_k)^T y_i}{y_k^T(s_k - Q_k y_k)} \quad = 0$$

$$y_k^T Q_k y_i = y_k^T s_i = s_k^T H s_i = s_k^T y_i \qquad y_k = H s_k$$

$$(s_k - Q_k y_k)^T y_i = s_k^T y_i - y_k^T Q_k y_i = s_k^T y_i - s_k^T y_i = 0$$

Example 11.1

Let $f(\boldsymbol{x}) = x_1^2 + \frac{1}{2}x_2^2 + 3$

Apply the rank one correction algorithm to minimize f. Use $x^0 = [1 \quad 2]^T$ and $Q_0 = I_2$ (identity matrix).

Solution $f(x) = \frac{1}{2}x^T Hx + 3, \qquad H = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \qquad b = 0$

$$g_k = Hx^k + b \qquad \rightarrow g_0 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} - 0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$d_0 = -Q_0 g_0 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$s_0 = \alpha_0 d_0 = \frac{4}{3}\begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$\alpha_0 = -\frac{d_0^T g_0}{d_0^T H d_0} = \frac{2}{3}$$

$$x^1 = x^0 + s_0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \frac{4}{3}\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1/3 \\ 2/3 \end{bmatrix}$$

$$x^1 = \begin{bmatrix} -1/3 \\ 2/3 \end{bmatrix}, \qquad s_0 = \begin{bmatrix} -4/3 \\ -4/3 \end{bmatrix}, \qquad g_0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$g_1 = Hx^1 + b = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1/3 \\ 2/3 \end{bmatrix} - 0 = \begin{bmatrix} -2/3 \\ 2/3 \end{bmatrix}$$

$$y_0 = g_1 - g_0 = \begin{bmatrix} -8/3 \\ -4/3 \end{bmatrix}$$

$$v_0 = s_0 - Q_0 y_0 = \begin{bmatrix} 4/3 \\ 0 \end{bmatrix}$$

$$v_0 v_0^T = \begin{bmatrix} 16/9 & 0 \\ 0 & 0 \end{bmatrix}$$

$$y_0^T v_0 = -\frac{32}{9}$$

$$Q_1 = Q_0 + \frac{v_0 v_0^T}{y_0^T v_0} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q_{k+1} = Q_k + \frac{v_k {v_k}^T}{y_k^T v_k}$$
$$v_k = s_k - Q_k y_k$$

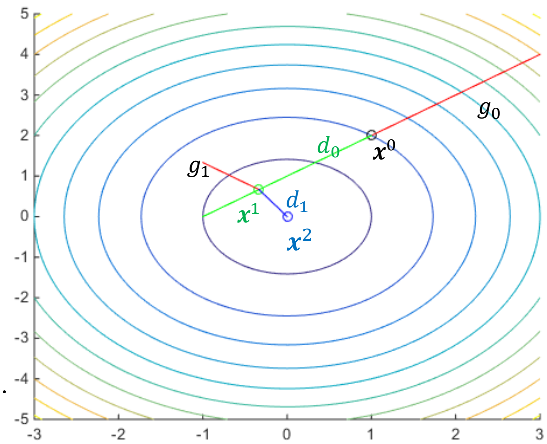
$$Q_1 = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}, \qquad g_1 = \begin{bmatrix} -2/3 \\ 2/3 \end{bmatrix}$$

$$d_1 = -Q_1 g_1 = \begin{bmatrix} 1/3 \\ -2/3 \end{bmatrix}$$

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T H d_1} = 1$$

$$x^2 = x^1 + \alpha_1 d_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$g_2 = 0 \qquad \Longrightarrow x^2 = x^*$$



As expected, the algorithm solves the problem in 2 steps.

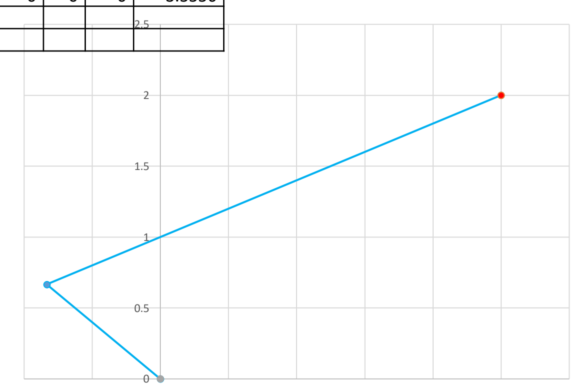
$$d_1^T H d_0 = \left[\frac{1}{3}, -\frac{2}{3}\right] \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix} = 0$$

The directions d_0 and d_1 are H-conjugate

Example 11.1

k	*x*		*Q*				*g*		*d*		*α*
0	1	2	1	0	0	1	2	2	-2	-2	0.6667
1	-0.3333	0.6667	0.5	0	0	1	-0.6667	0.6667	0.3333	-0.6667	1
2	0	0					0	0			

k	Δx		Δg		v=Δx-QΔg		v•v^T				Δg^T•v
0	-1.3333	-1.3333	-2.6667	-1.3333	1.3333	0	1.7778	0	0	0	-3.5556
1											
2											



REVIEW

Outer Product

$$\boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{a}\boldsymbol{b}^T,\ \boldsymbol{a}, \boldsymbol{b} \in \mathcal{R}^n$$

$$\boldsymbol{a} \otimes \boldsymbol{a} = \boldsymbol{a}\boldsymbol{a}^T \succcurlyeq 0$$

$$\boldsymbol{x}^T(\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{x} = (\boldsymbol{a}^T\boldsymbol{x})^T(\boldsymbol{a}^T\boldsymbol{x}) = \|\boldsymbol{a}^T\boldsymbol{x}\|^2 \geq 0$$

Disadvantage

1. If $Q_{k+1} \nsucc 0$, $d_{k+1} = -Q_{k+1}g_{k+1}$ may not be a descent direction, even in the quadratic case.

Supposing $Q_k \succ 0$, if $y_k^T(s_k - Q_k y_k) > 0$, then $Q_{k+1} \succ 0$.

if $y_k^T(s_k - Q_k y_k) < 0$, Q_{k+1} may not be positive definite.

$$Q_{k+1} = Q_k + \frac{\boxed{(s_k - Q_k y_k)(s_k - Q_k y_k)^T}}{y_k^T(s_k - Q_k y_k)} \quad \succcurlyeq 0$$

2. If $y_k^T(s_k - Q_k y_k) \to 0$, there may be numerical problems in evaluating Q_{k+1}.

Denominator $\to 0$

为此，提出了DFP、BFGS为代表的多种拟牛顿方法

1. 拟牛顿法的概念与基本算法
2. 秩1修正法
3. DFP方法
4. BFGS法
5. Broyden族

秩2修正公式

不同的$C_k = C\left(x^k\right)$构造方法形成不同的拟牛顿法，最著名的两个：

① 由Davidon提出，Fletcher和Powell改进的DFP方法

② 由Broyden、Fletcher、Goldfarb、Shannon分别同时提出的BFGS方法 变尺度法

$$Q_k = Q_{k-1} + C_k \qquad C_k = zz^T + ww^T \quad \text{秩2修正}$$

DFP法的校正公式

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k} \qquad k = 0,1,\dots,n-1$$

- 二次型DFP方法满足割线方程 $Q_{k+1} y_i = s_i$，$0 \le i \le k$

Example 11.3

$$v_k = Q_k y_k$$

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{v_k v_k^T}{y_k^T v_k}$$

Locate the minimizer of $f(x) = \frac{1}{2}x^T H x - b^T x$, $H = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

Use the initial point $x^0 = [0,0]^T$ and $Q_0 = I_2$.

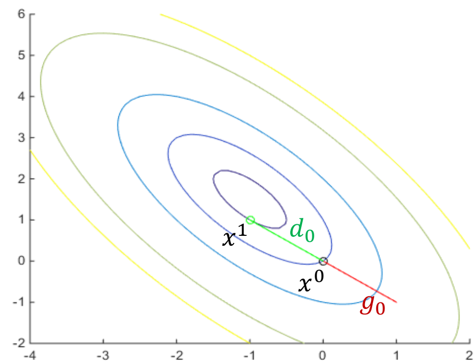
Solution

$$g_k = Hx^k - b$$
$$g_0 = Hx^0 - b = (1,-1)$$
$$d_0 = -Q_0 g_0 = (-1,1)$$

f 是二次型，$\alpha_0 = -\dfrac{g_0^T d_0}{d_0^T H d_0} = 1$

$$s_0 = \alpha_0 d_0 = (-1,1)$$

$$x^1 = x^0 + s_0 = (-1,1)$$
$$g_1 = Hx^1 - b = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$
$$y_0 = g_1 - g_0 = (-2,0)$$



Example_11_03_Chong.m

$$v_k = Q_k y_k$$

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{v_k v_k^T}{y_k^T v_k}$$

$$s_0 s_0^T = \begin{bmatrix} -1 \\ 1 \end{bmatrix}[-1 \quad 1] = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$s_0^T y_0 = [-1 \quad 1]\begin{bmatrix} -2 \\ 0 \end{bmatrix} = 2$$

$$v_0 = Q_0 y_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \rightarrow v_0 v_0^T = \begin{bmatrix} -2 \\ 0 \end{bmatrix}[-2 \quad 0] = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

$$y_0^T v_0 = [-2 \quad 0]\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ 0 \end{bmatrix} = 4$$

$$\frac{s_0 s_0^T}{s_0^T y_0} - \frac{v_0 v_0^T}{y_0^T v_0} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$$

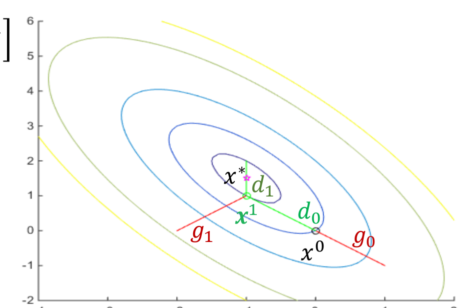
$$Q_1 = Q_0 + \frac{s_0 s_0^T}{s_0^T y_0} - \frac{v_0 v_0^T}{y_0^T v_0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$$

$$d_1 = -Q_1 g_1 = -\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T H d_1} = \frac{1}{2}$$

$$x^2 = x^1 + \alpha_1 d_1 = \begin{bmatrix} -1 \\ 3/2 \end{bmatrix} = x^*$$

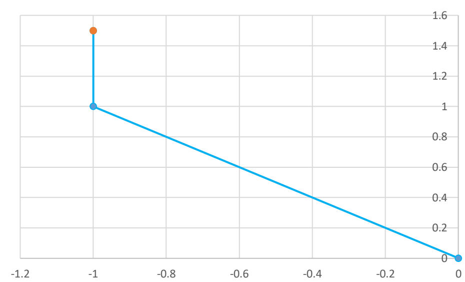
$$g_2 = 0$$



k	x		g		Δg		Q				d		α
0	0	0	1	-1	-2	0	1	0	0	1	-1	1	1
1	-1	1	-1	-1			0.5	-0.5	-0.5	1.5	0	1	0.5
2	-1	1.5	0	0									

k	Δx		Q•Δg		Δx•Δx ^{T}				Δx^T•Δg	Δg^T•Q•Δg	(Q•Δg)•(Q•Δg) ^{T}			
0	-1	1	-2	0	1	-1	-1	1	2	4	4	0	0	0
1														
2														

$\boldsymbol{g}^2 = \mathbf{0}$



OPT_Example_11_3.m

CH11-Example-11.3.xlsx

DFP法的性质

① $Q_k \succ 0 \iff s_k^T y_k > 0$

② 对于凸二次函数，迭代搜索过程中如果都使用一维精确方法获得最优步长，那么向量组$s_0, s_1, \ldots, s_{n-1}$关于$H$共轭 Theorem 7.3 Conjugate directions in DFP method [3]

③ 二次型的DFP方法满足割线方程

④ 当$s_k^T y_k > 0$时，采用Wolfe非精确一维搜索方法的DFP方法是下降算法

⑤ 当$g_k \neq 0$时，如果$Q_k \succ 0$，那么$Q_{k+1} \succ 0$

① $Q_k \succ 0 \iff s_k^T y_k > 0$

观察DFP公式 $$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k}$$

$$\begin{aligned} s_k^T y_k &= s_k^T (g_{k+1} - g_k) \\ &= \alpha_k d_k^T (g_{k+1} - g_k) \\ &= -\alpha_k d_k^T g_k \\ &= \alpha_k g_k^T Q_k g_k > 0 \end{aligned}$$

$$d_k^T g_{k+1} = 0$$

$s_k^T y_k > 0 \quad \iff \quad Q_k \succ 0$，且，$g_k \neq 0$

③ 二次型DFP方法满足割线方程

采用 DFP方法求解Hessian矩阵为 $H = H^T$ 的二次型问题，有

$$Q_{k+1} y_i = s_i,\ 0 \le i \le k.$$

从而，DFP 方法是一种共轭方向法

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k}$$

证明 采用归纳法

$k = 0$时， $$\begin{aligned} Q_1 y_0 &= Q_0 y_0 + \frac{s_0 s_0^T}{s_0^T y_0} y_0 - \frac{(Q_0 y_0)(Q_0 y_0)^T}{y_0^T Q_0 y_0} y_0 \\ &= Q_0 y_0 + \frac{s_0 s_0^T y_0}{s_0^T y_0} - \frac{(Q_0 y_0)(Q_0 y_0)^T y_0}{y_0^T Q_0 y_0} \\ &= Q_0 y_0 + s_0 - Q_0 y_0 \\ &= s_0 \end{aligned}$$

$$(Q_0 y_0)^T y_0 = y_0^T Q_0 y_0$$

$$Q_1 y_0 = s_0$$

假设在$k-1$时定理成立，即，$Q_k y_i = s_i$，$0 \le i \le k-1 \Longrightarrow Q_{k+1} y_i = s_i$，$0 \le i \le k$

首先，证明$i = k$时

$$Q_{k+1} y_k = Q_k y_k + \frac{s_k \boxed{s_k^T}}{\boxed{s_k^T y_k} y_k} - \frac{(Q_k y_k)\boxed{(Q_k y_k)^T}}{\boxed{y_k^T Q_k y_k} y_k}$$

$$= Q_k y_k + s_k \qquad - Q_k y_k$$

$$= s_k$$

当 $i < k$时，

$$Q_{k+1} y_i = Q_k y_i + \frac{s_k \boxed{s_k^T}}{\boxed{s_k^T y_k} y_i} - \frac{(Q_k y_k)\boxed{(Q_k y_k)^T}}{\boxed{y_k^T Q_k y_k} y_i} \qquad Q_k y_i = s_i\text{假设成立}$$

$$= s_i + \frac{s_k s_k^T y_i}{s_k^T y_k} - \frac{(Q_k y_k)(y_k^T Q_k y_i)}{y_k^T Q_k y_k}$$

$$= s_i$$

$s_k^T y_i = s_k^T H s_i = 0 \Longleftarrow$ ② $s_0, s_1, \dots, s_{n-1}$是H共轭　$y_i = H s_i$ 割线方程

$y_k^T Q_k y_i = y_k^T s_i = s_k^T H s_i = 0$　　　**证毕**

④ 当$s_k^T y_k > 0$时，采用Wolfe非精确线搜法的DFP方法是下降的

或 $Q_k \succ 0$

证：　$\rho \in (0,1), \sigma \in (\rho, 1)$

$f(x^{k+1}) \le f(x^k) + [\rho g_k^T d_k]\alpha_k$

$g_{k+1}^T d_k \ge \sigma g_k^T d_k \quad \longrightarrow \quad g_{k+1}^T d_k - \sigma g_k^T d_k \ge 0$

$s_k = \alpha_k d_k$

$x_{k+1} - x_k = \alpha_k d_k$

$y_k = g_{k+1} - g_k$

$$0 < s_k^T y_k = y_k^T s_k = [g_{k+1} - g_k]^T [\alpha_k d_k]$$
$$= \alpha_k [g_{k+1}^T d_k - g_k^T d_k]$$
$$\ge \alpha_k (\sigma - 1) g_k^T d_k$$

当$s_k^T y_k > 0$时，因为$\alpha_k > 0$，$\sigma - 1 < 0$，则$g_k^T d_k < 0 \Longrightarrow$　d_k是下降方向

DFP方法是下降算法

⑤ 当$g_k \neq 0$时，如果$Q_k \succ 0$，那么$Q_{k+1} \succ 0$

在DFP法中，假定$g_k \neq 0$，如果Q_k是正定的，那么，Q_{k+1}也是正定的.

$$p^T Q_{k+1} p = p^T Q_k p + \frac{p^T s_k s_k^T p}{s_k^T y_k} - \frac{p^T (Q_k y_k)(Q_k y_k)^T p}{y_k^T Q_k y_k}$$

$$= p^T Q_k p + \frac{(p^T s_k)^2}{s_k^T y_k} - \frac{(p^T Q_k y_k)^2}{y_k^T Q_k y_k}$$

$$= a^T a + \frac{(p^T s_k)^2}{s_k^T y_k} - \frac{(a^T b)^2}{b^T b}$$

$$= \underbrace{\frac{\|a\|^2 \|b\|^2 - \|a^T b\|^2}{\|b\|^2}}_{\geq 0} + \underbrace{\frac{(p^T s_k)^2}{s_k^T y_k}}_{> 0} > 0$$

柯西-施瓦茨不等式

$Q_{k+1} \succ 0$

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k}$$

$$Q_k = Q_k^{\frac{1}{2}} Q_k^{\frac{1}{2}}$$

说明见下页

$$a \triangleq Q_k^{\frac{1}{2}} p, \qquad b \triangleq Q_k^{\frac{1}{2}} y_k$$

$$Q_k \succ 0 \quad \Leftrightarrow \quad s_k^T y_k > 0$$

REVIEW

$$Q \succcurlyeq 0 \implies Q = Q^{\frac{1}{2}} Q^{\frac{1}{2}}, \quad Q^{\frac{1}{2}} \succcurlyeq 0$$

$$Q \succcurlyeq 0 \implies Q = D \Lambda D^T, \; D^T D = I, \; \Lambda = \mathrm{diag}(\lambda_i), \lambda_i \geq 0$$

$$Q^{\frac{1}{2}} = D \Lambda^{\frac{1}{2}} D^T, \qquad \Lambda^{\frac{1}{2}} = \mathrm{diag}\left(\lambda_i^{\frac{1}{2}}\right)$$

Advantage and Disadvantage

- The DFP algorithm is superior to the rank one algorithm, for it *preserves the positive definiteness* of Q_k.
- For *larger nonquadratic problems*, the algorithm may get "*stuck*" and Q_k become *nearly singular*.

So, we discuss the following algorithm that *alleviates* this problem.

[əˈliviˌet]减轻，缓和；

1. 拟牛顿法的概念与基本算法
2. 秩1修正法
3. DFP方法
4. BFGS法
5. Broyden族

BFGS校正公式 （秩2修正公式）

In 1970, an alternative formula was suggested independently by Broyden, Fletcher, Goldfarb, and Shanno. The method, now called the *BFGS algorithm*.

$$y_i \rightleftharpoons s_i$$
$$Q_k \rightleftharpoons B_k$$

Dual or *Complementary* 对偶

DFP	BFGS
$Q_{k+1}y_i = s_i,\ 0 \le i \le k$	$y_i = B_{k+1}s_i,\ 0 \le i \le k$
$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k}$	$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k}$
Inverse of Hessian	Hessian

DFP校正公式

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k}$$

Inverse of Hessian

对偶式

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k}$$

Hessian

$$Q_{k+1}^{BFGS} = [B_{k+1}]^{-1}$$

BFGS校正公式

$$Q_{k+1}^{BFGS} = \left[B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k}\right]^{-1}$$

REVIEW

Sherman-Morrison公式

设A为可逆矩阵，$\boldsymbol{u}$、$\boldsymbol{v}$均为列向量，若$1+\boldsymbol{v}^TA^{-1}\boldsymbol{u}\neq 0$且$A+\boldsymbol{u}\boldsymbol{v}^T$可逆，则有

$$(A+\boldsymbol{u}\boldsymbol{v}^T)^{-1}=A^{-1}-\frac{A^{-1}\boldsymbol{u}\boldsymbol{v}^TA^{-1}}{1+\boldsymbol{v}^TA^{-1}\boldsymbol{u}}$$

两次应用Sherman-Morrison公式，得BFGS校正公式

$$Q_{k+1}^{BFGS}=\left[B_k+\frac{y_ky_k^T}{y_k^Ts_k}-\frac{(B_ks_k)(B_ks_k)^T}{s_k^TB_ks_k}\right]^{-1}$$

$$Q_{k+1}^{BFGS}=Q_k^{BFGS}+\left[1+\frac{y_k^TQ_k^{BFGS}y_k}{y_k^Ts_k}\right]\frac{s_ks_k^T}{s_k^Ty_k}-\frac{Q_k^{BFGS}y_ks_k^T+\left[Q_k^{BFGS}y_ks_k^T\right]^T}{y_k^Ts_k}$$

右乘y_k得，

$$Q_{k+1}^{BFGS}y_k=s_k$$

可验证其合理性

BFGS校正公式记号说明

$$Q_{k+1}^{BFGS}=\left[B_k+\frac{y_k y_k^T}{y_k^T s_k}-\frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k}\right]^{-1} \triangleq \bar{Q}_{k+1}$$

$$Q_{k+1}^{BFGS}=Q_k^{BFGS}+\left[1+\frac{y_k^T Q_k^{BFGS} y_k}{y_k^T s_k}\right]\frac{s_k s_k^T}{s_k^T y_k}-\frac{Q_k^{BFGS} y_k s_k^T+\left[Q_k^{BFGS} y_k s_k^T\right]^T}{y_k^T s_k}$$

$$\bar{Q}_{k+1}=\bar{Q}_k+\left[1+\frac{y_k^T \bar{Q}_k y_k}{y_k^T s_k}\right]\frac{s_k s_k^T}{s_k^T y_k}-\frac{\bar{Q}_k y_k s_k^T+\left[\bar{Q}_k y_k s_k^T\right]^T}{y_k^T s_k}$$

BFGS方法的性质

对于二次型问题

① $Q_n^{BFGS}=H^{-1}$

② Directions $s_0, s_1, \ldots, s_{n-1}$ form a conjugate set.

③ Q_{k+1} is positive definite if Q_k is positive definite.

④ $s_k^T y_k>0$

BFGS方法的特点

① 与DFP方法相比，校正性更加良好

② Q_k^{BFGS}不易变为奇异矩阵

③ 具有全局超线性收敛

④ 是解多维无约束优化问题最常用的方法

对于非二次型问题，拟牛顿法一般不是n步收敛的

修正:
每迭代n步，搜索方向重新取负梯度方向，直至满足“终止条件”

DFP/BFGS 校正公式记号

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k}$$

$$v_k = Q_k y_k$$

$$Q_{k+1} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{v_k v_k^T}{y_k^T v_k}$$

DFP

$$\bar{Q}_{k+1} = \bar{Q}_k + \left[1 + \frac{y_k^T \bar{Q}_k y_k}{y_k^T s_k}\right] \frac{s_k s_k^T}{s_k^T y_k} - \frac{\bar{Q}_k y_k s_k^T + \left[\bar{Q}_k y_k s_k^T\right]^T}{y_k^T s_k}$$

$$\bar{v}_k = \bar{Q}_k y_k$$

$$\Psi_k = \bar{Q}_k y_k s_k^T$$

$$\bar{Q}_{k+1} = \bar{Q}_k + \left[1 + \frac{y_k^T \bar{v}_k}{y_k^T s_k}\right] \frac{s_k s_k^T}{s_k^T y_k} - \frac{\Psi_k + \Psi_k^T}{y_k^T s_k}$$

BFGS

Example 11.4

Use the BFGS method to minimize $f(x) = \frac{1}{2}x^T H x - x^T b + \ln\pi$, $H = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$, $b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Take $\bar{Q}_0 = I_2$ and $x^0 = 0$. Verify that $\bar{Q}_2 = H^{-1}$.

$$\bar{Q}_{k+1} = \bar{Q}_k + \left[1 + \frac{y_k^T \bar{v}_k}{y_k^T s_k}\right] \frac{s_k s_k^T}{s_k^T y_k} - \frac{\Psi_k + \Psi_k^T}{y_k^T s_k}$$
$$\bar{v}_k = \bar{Q}_k y_k$$
$$\Psi_k = \bar{Q}_k y_k s_k^T$$

Proof $g_k = Hx^k - b$ $\qquad g_0 = (0, -1)$

$$d_0 = -g_0 = (0,1)$$
$$\alpha_0 = -\frac{g_0^T d_0}{d_0^T H d_0} = 1/2$$
$$s_0 = \alpha_0 d_0 = (0, 1/2)$$
$$x^1 = x^0 + s_0 = (0, 1/2)$$

$$g_1 = Hx^1 - b = (-3/2, 0)$$
$$y_0 = g_1 - g_0 = (-3/2, 1)$$
$$\bar{v}_0 = \bar{Q}_0 y_0 = (-3/2, 1)$$

$$\Psi_0 = \bar{Q}_0 y_0 s_0^T = \begin{bmatrix} 0 & -3/4 \\ 0 & 1/2 \end{bmatrix}$$

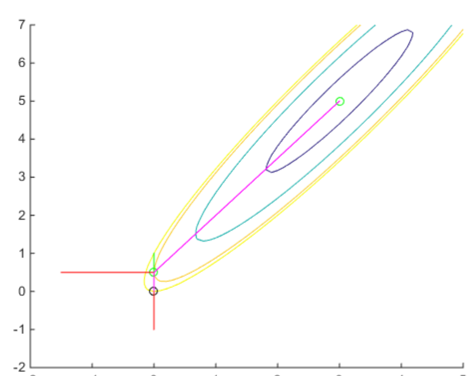
$$\bar{Q}_1 = \bar{Q}_0 + \left[1 + \frac{y_0^T \bar{v}_0}{y_0^T s_0}\right] \frac{s_0 s_0^T}{s_0^T y_0} - \frac{\Psi_0 + \Psi_0^T}{y_0^T s_0} = \begin{bmatrix} 1 & 3/2 \\ 3/2 & 11/4 \end{bmatrix}$$

$$\bar{Q}_{k+1} = \bar{Q}_k + \left[1 + \frac{y_k^T \bar{v}_k}{y_k^T s_k}\right] \frac{s_k s_k^T}{s_k^T y_k} - \frac{\Psi_k + \Psi_k^T}{y_k^T s_k}$$
$$\bar{v}_k = \bar{Q}_k y_k$$
$$\Psi_k = \bar{Q}_k y_k s_k^T$$

$$d_1 = -\bar{Q}_1 g_1 = (3/2, 9/4) \qquad \bar{Q}_1 = \begin{bmatrix} 1 & 3/2 \\ 3/2 & 11/4 \end{bmatrix}$$

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T H d_1} = 2$$
$$s_1 = \alpha_1 d_1 = (3, 9/2)$$
$$x^2 = x^1 + \alpha_1 d_1 = (3,5) = x^*$$
$$g_2 = 0$$

$$y_1 = g_2 - g_1 = (3/2, 0)$$
$$\bar{v}_1 = \bar{Q}_1 y_1 = (3/2, 9/4)$$
$$\Psi_1 = \bar{Q}_1 y_1 s_1^T = \begin{bmatrix} 9/2 & 27/4 \\ 0 & 0 \end{bmatrix}$$



$$\bar{Q}_2 = \bar{Q}_1 + \left[1 + \frac{y_1^T \bar{v}_1}{y_1^T s_1}\right] \frac{s_1 s_1^T}{s_1^T y_1} - \frac{\Psi_1 + \Psi_1^T}{y_1^T s_1} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

$$Q_2 H = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} = I_2 \rightarrow Q_2 = H^{-1}$$

Algorithm DFP/BFGS

Given x^0, $k=0$， $Q_0=I$

Evaluate $g_k=\nabla f(x^k)$; 二次型 一般目标函数

While $\boxed{\|g_k\|>tol\ \ \&\ k<n}$ $\|g_k\|>tol\ \ \&\ \|x^{k+1}-x^k\|>tol\cdot\min\{1,\|x^k\|\}$

$d_k=-Q_k g_k$

α_k by line search to satisfy Wolfe rule

$s_k=\alpha_k d_k$， $x^{k+1}=x^k+s_k, y_k=g_{k+1}-g_k$

If $s_k^T y_k>0$, Compute Q_{k+1} by DFP / BFGS

else $Q_{k+1}=I$

end (if)

$k\leftarrow k+1$

end(while)

$\|x^{k+1}-x^k\|<tol\cdot\min\{1,\|x^k\|\}$
$\|g_{k+1}\|<tol$

$$Q_{k+1}=Q_k+\frac{s_k s_k^T}{s_k^T y_k}-\frac{v_k v_k^T}{y_k^T v_k}\qquad v_k=Q_k y_k$$

$$\bar{Q}_{k+1}=\bar{Q}_k+\left[1+\frac{y_k^T\bar{v}_k}{y_k^T s_k}\right]\frac{s_k s_k^T}{s_k^T y_k}-\frac{\Psi_k+\Psi_k^T}{y_k^T s_k}\qquad \bar{v}_k=\bar{Q}_k y_k,\quad \Psi_k=\bar{Q}_k y_k s_k^T$$

Algorithm 7.3 Practical quasi-Newton algorithm [3]

DFP/BFGS拟牛顿法的MATLAB程序

`DFP_Wolfe.m`　　`BFGS_Wolfe.m`

`Wolfe_Search.m`

注意:
DFP_Wolfe.m中迭代点和函数值写入数据文件testdata.txt只适合于二维变量
如果维数不是2，需要改写“写入格式”
如果“写入格式”可以表述为可变维数的，就理想了！！！

梯度范数小于1e-2* tol时，终止

例

增加了梯度范数小于1e-2* tol时，终止

例6.4 用DFP法求解多维无约束最优化问题

（取初始点$\boldsymbol{x}^0 = (1,1), tol = 1\times 10^{-6}$）

$$\min f(\boldsymbol{x}) = -0.8e^{-x_1^2-4x_2^2}$$

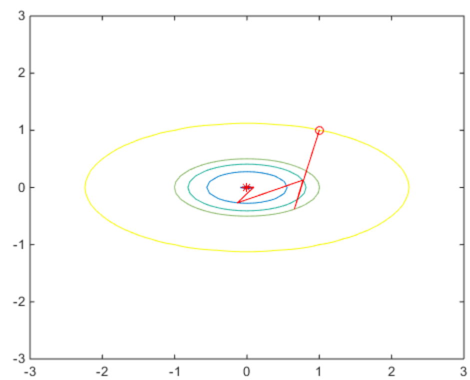
example_6_4_CH06.m

DFP_Wolfe.m

Wolfe__Search.m

testdata.txt

x_optimal = 1.0e-11 *[-0.0687 0.4504]
f_optimal =-0.8000
k = 10



例6.5 用DFP法求解多维无约束最优化问题

（取初始点$\boldsymbol{x}^0 = (1,-4), tol = 1\times 10^{-6}$）

$$\min f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_1x_2 + 2$$

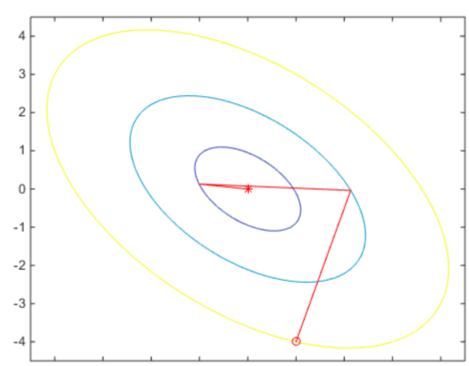
example_6_5_CH06.m

Conjugate_gradient_DY.m

Wolfe__Search.m

testdata.txt

x_optimal = 1.0e-17 *[0.6614 -0.1301]
f_optimal = 2
k = 5



例6.6 用BFGS法求解多维无约束最优化问题

（取初始点$\boldsymbol{x}^0 = (1,-4), tol = 1 \times 10^{-6}$）

$$\min f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_1 x_2 + 2$$

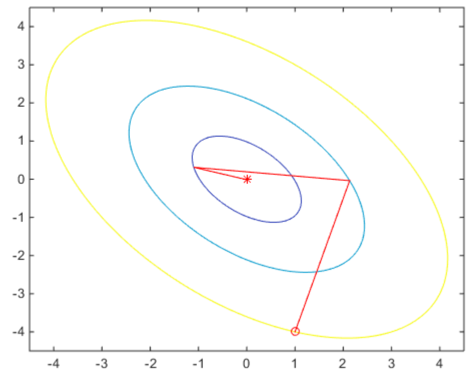
example_6_6_CH06.m

Conjugate_gradient_DY.m

Wolfe__Search.m

testdata.txt

x_optimal = 1.0e-16 *[-0.1128   0]
f_optimal = 2
k = 5



例6.7 用BFGS法求解多维无约束最优化问题

（取初始点$\boldsymbol{x}^0 = (-4,0,-4,-1,1,1), tol = 1 \times 10^{-6}$）

$$\min f(\boldsymbol{x}) \qquad f(\boldsymbol{x}) = 1 + x_1 + x_2 + x_3 + x_4$$
$$+x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4$$
$$+x_1^2 + x_2^2 + +x_3^2 + x_4^2 - 0.4e^{-x_5^2-6x_6^2}$$

x_optimal =[-0.2000 -0.2000
-0.2000 -0.2000
0.0000   0.0000]

f_optimal = 0.2000

k =11

example_6_7_CH06.m

Conjugate_gradient_DY.m

Wolfe__Search.m

1. 拟牛顿法的概念与基本算法
2. 秩1修正法
3. DFP方法
4. BFGS法
5. Broyden族　DFP与BFGS的线性组合

Broyden族

$$Q_{k+1}^{\phi} = (1-\phi_k)Q_{k+1}^{DFP} + \phi_k Q_{k+1}^{BFGS}$$

$$\phi_k = \frac{s_k^T y_k}{s_k^T y_k \pm y_k^T Q_k y_k}$$

$$Q_{k+1}^{DFP} = Q_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{(Q_k y_k)(Q_k y_k)^T}{y_k^T Q_k y_k}$$

$$Q_{k+1}^{BFGS} = Q_k + \left[1 + \frac{y_k^T Q_k y_k}{y_k^T s_k}\right]\frac{s_k s_k^T}{s_k^T y_k} - \frac{Q_k y_k s_k^T + \left[Q_k y_k s_k^T\right]^T}{y_k^T s_k}$$

Practical Optimization Algorithms and Engineering Applications, §7. 8, A. Antoniou, W. LU

作业

用牛顿法、DFP拟牛顿法和BFGS拟牛顿法完成以下各题

6-1

6-3

6-5