最优化理论与方法

第三章　无约束规划方法

上海财经大学应用数学系

December 26, 2008

3.1最速下降法

最速下降法是求解无约束最优化问题最早使用的方法之一.它是现代方法的基础,许多求解无约束问题的现代方法都是在这种方法基础上或在其启发下建立起来的.

最速下降法的思想

- 我们还是按照逐次求点,得一个点列不断趋近最优点的想法.设在第k步得到一个点x^k.设目标函数$f(x)$在x^k附近连续可微,且$\nabla f(x^k) \neq 0$. 将$f(x)$在x^k处泰勒展开

$$f(x) = f(x^k) + \nabla f(x^k)^T(x - x^k) + o(\|x - x^k\|),$$

记$x - x^k = ad^k$, 其中d^k是一个确定的方向向量.上式可以写为

$$f(x^k + ad^k) = f(x^k) + a\nabla f(x^k)^T d^k + o(\|ad^k\|).$$

- 我们知道,若向量d^k满足$\nabla f(x^k)d^k < 0$,则d^k是下降方向.由上式知,若$\nabla f(x^k)d^k < 0$ 越小, 即$-\nabla f(x^k)d^k < 0$越大,则$f(x)$下降的越大.

最速下降法的思想

- 我们还是按照逐次求点,得一个点列不断趋近最优点的想法.设在第k步得到一个点x^k.设目标函数$f(x)$在x^k附近连续可微,且$\nabla f(x^k) \neq 0$. 将$f(x)$在x^k处泰勒展开

$$f(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + o(\|x - x^k\|),$$

记$x - x^k = ad^k$, 其中d^k是一个确定的方向向量.上式可以写为

$$f(x^k + ad^k) = f(x^k) + a\nabla f(x^k)^T d^k + o(\|ad^k\|).$$

- 我们知道,若向量d^k满足$\nabla f(x^k)d^k < 0$,则d^k是下降方向.由上式知,若$\nabla f(x^k)d^k < 0$ 越小, 即$-\nabla f(x^k)d^k < 0$越大,则$f(x)$下降的越大.

最速下降法的思想

- 由$-\nabla f(x^k)^T d^k = \|\nabla f(x^k)\| \|d^k\| cos\theta_k$,其中$\theta_k$是向量$-\nabla f(x^k)$与向量$d^k$之间的夹角,知,当$a$ 和$\|d^k\|$固定时,取$\theta_k = 0$,也即取$d^k = -\nabla f(x^k)$时,$-\nabla f(x^k)^T d^k$最大,因而,$f(x)$在x^k处下降量最大. 故取搜索方向$d^k = -\nabla f(x^k)$,相应的方法称为最速下降法.
- 其迭代格式为:

$$x^{k+1} = x^k + a_k d^k,$$

a^k由线性搜索确定.
求解问题

$$\min_{x \in R^n} f(x), f \in C^1$$

最速下降法的思想

- 由$-\nabla f(x^k)^T d^k = \|\nabla f(x^k)\|\|d^k\| cos\theta_k$,其中$\theta_k$是向量$-\nabla f(x^k)$与向量$d^k$之间的夹角,知,当$a$ 和$\|d^k\|$固定时,取$\theta_k = 0$,也即取$d^k = -\nabla f(x^k)$时,$-\nabla f(x^k)^T d^k$最大,因而,$f(x)$在x^k处下降量最大. 故取搜索方向$d^k = -\nabla f(x^k)$,相应的方法称为最速下降法.
- 其迭代格式为:

$$x^{k+1} = x^k + a_k d^k,$$

a^k由线性搜索确定.
求解问题

$$\min_{x \in R^n} f(x), f \in C^1$$

3.1.1最速下降法的具体步骤

最速下降法的具体步骤

1、 选定初试点x^1和给定精度要求$\epsilon > 0$, 令$k = 1$;

2、 若$\|\nabla f(x^k)\| < \epsilon$,则停,$x^* = x^k$,否则令$d^k = -\nabla f(x^k)$;

3、 在x^k处沿方向d^k作线性搜索得$x^{k+1} = x^k + a_k d^k$, $k = k + 1$,转步2.

最速下降法的具体步骤

1、 选定初试点x^1和给定精度要求$\epsilon > 0$, 令$k = 1$;

2、 若$\|\nabla f(x^k)\| < \epsilon$,则停,$x^* = x^k$,否则令$d^k = -\nabla f(x^k)$;

3、 在x^k处沿方向d^k作线性搜索得$x^{k+1} = x^k + a_k d^k$, $k = k + 1$,转步2.

最速下降法的具体步骤

1、 选定初试点x^1和给定精度要求$\epsilon > 0$, 令$k = 1$;

2、 若$\|\nabla f(x^k)\| < \epsilon$,则停,$x^* = x^k$,否则令$d^k = -\nabla f(x^k)$;

3、 在x^k处沿方向d^k作线性搜索得$x^{k+1} = x^k + a_k d^k, k = k+1$,转步2.

- 若在第三步中,采用精确线性搜索,即

$$a_k = argminf(x^k + ad^k)$$

- 于是就有

$$\frac{df(x^k + ad^k)}{da}\Big|_{a=a_k} = (d^k)^T \nabla f(x^{k+1}) = 0 .$$

此式表明d^k与d^{k+1}是正交的.

- 若在第三步中,采用精确线性搜索,即

$$a_k = argmin f(x^k + ad^k)$$

- 于是就有

$$\frac{df(x^k + ad^k)}{da}\Big|_{a=a_k} = (d^k)^T \nabla f(x^{k+1}) = 0 \ .$$

此式表明d^k与d^{k+1}是正交的.

例3.1.1用最速下降法求解无约束问题

用最速下降法求解无约束问题

$$\min f(x) = \frac{x_1^2}{a} + \frac{x_2^2}{b}$$

取初始点$x^1 = (a, b)^T$.

解

- 由于

$$\nabla f(x) = \left(\frac{2}{a}x_1, \frac{2}{b}x_2\right)^T,$$

因此,$\nabla f(x^1) = (2,2)^T$, 取$d^1 = -\nabla f(x^1) = (-2,-2)^T$,作一维搜索：构造一元函数

$$\phi(\alpha) = f(x^1 + \alpha d^1) = \frac{(a-2\alpha)^2}{a} + \frac{(b-2\alpha)^2}{b},$$

- 求导得

$$\phi'(\alpha) = -4\left(1 - \frac{2}{a}\alpha\right) - 4\left(1 - \frac{2}{b}\alpha\right),$$

解方程$\phi'(\alpha) = 0$,得最优步长

$$\alpha_1 = \frac{ab}{}$$

解

- 由于

$$\nabla f(x) = \left(\frac{2}{a}x_1, \frac{2}{b}x_2\right)^T,$$

因此,$\nabla f(x^1) = (2,2)^T$, 取$d^1 = -\nabla f(x^1) = (-2,-2)^T$,作一维搜索：构造一元函数

$$\phi(\alpha) = f(x^1 + \alpha d^1) = \frac{(a-2\alpha)^2}{a} + \frac{(b-2\alpha)^2}{b},$$

- 求导得

$$\phi'(\alpha) = -4\left(1 - \frac{2}{a}\alpha\right) - 4\left(1 - \frac{2}{b}\alpha\right),$$

解方程$\phi'(\alpha) = 0$,得最优步长

$$\alpha_1 = \frac{ab}{\ldots}$$

解

- 从而

$$\begin{aligned} x^2 &= x^1 + \alpha_1 d^1 \\ &= (a,b)^T + \tfrac{ab}{a+b}(-2,-2)^T \\ &= \left(-\tfrac{a(a-b)}{a+b}, \tfrac{b(b-a)}{a+b}\right)^T \end{aligned}$$

- 若$a + b$,则$x^2 = (0,0)^T, \nabla f(x^2) = (0,0)^T$, 停在计算,$x^2$ 为无约束问题的最优解.
  若$a \neq b$, 则再进行下一次迭代,得到

$$x^3 = \left(\frac{a(a-b)^2}{(a+b)^2}, \frac{b(b-a)^2}{(a+b)^2}\right)^T,$$

解

- 从而
$$\begin{aligned} x^2 &= x^1 + \alpha_1 d^1 \\ &= (a,b)^T + \tfrac{ab}{a+b}(-2,-2)^T \\ &= \left(-\tfrac{a(a-b)}{a+b}, \tfrac{b(b-a)}{a+b}\right)^T \end{aligned}$$
- 若$a+b$,则$x^2=(0,0)^T, \nabla f(x^2)=(0,0)^T$, 停在计算,$x^2$ 为无约束问题的最优解.
若$a \neq b$, 则再进行下一次迭代,得到
$$x^3 = \left(\frac{a(a-b)^2}{(a+b)^2}, \frac{b(b-a)^2}{(a+b)^2}\right)^T,$$

解

- 如此作下去,可得

$$x^{k+1} = \left(\frac{a(a-b)^k}{(a+b)^k}, \frac{b(b-a)^k}{(a+b)^k} \right)^T,$$

- 当$k \to \infty$时, $x^k \to (0,0)^T$. 得到无约束问题的最优解.

解

- 如此作下去,可得

$$x^{k+1} = \left(\frac{a(a-b)^k}{(a+b)^k}, \frac{b(b-a)^k}{(a+b)^k} \right)^T,$$

- 当$k \to \infty$时, $x^k \to (0,0)^T$. 得到无约束问题的最优解.

3.2牛顿法

牛顿法的基本思想是利用二次近似多项式的极值点求法给出原函数的极值点的求法.

3.2.1牛顿方法的思想

- 设x^*是无约束问题min$f(x), x \in R^n$, 的局部解,其中$f(x)$是二次可微的,则x^*满足

$$\nabla f(x) = 0 \tag{1}$$

- 解方程组可得优化问题的解.但是该方程组一般是非线性方程组,不易求解.

选取初始点x^1(作为x^*的第一次近似),在x^1处Taylor展开,取二次近似多项式

$$f(x) \approx f(x^1) + \nabla f(x^1)^T (x - x^1) + (x - x^1)^T \nabla^2 f(x^1)(x - x^1) \tag{2}$$

3.2.1牛顿方法的思想

- 设x^*是无约束问题$\min f(x), x \in R^n$, 的局部解,其中$f(x)$是二次可微的,则x^*满足

$$\nabla f(x) = 0 \tag{1}$$

- 解方程组可得优化问题的解.但是该方程组一般是非线性方程组,不易求解.

选取初始点x^1(作为x^*的第一次近似),在x^1处Taylor展开,取二次近似多项式

$$f(x) \approx f(x^1)+\nabla f(x^1)^T(x-x^1)+(x-x^1)^T\nabla^2 f(x^1)(x-x^1) \tag{2}$$

3.2.1牛顿方法的思想

- 令近似二次函数的导数为零,得

$$\nabla f(x^1) + \nabla^2 f(x^1)(x - x^1) = 0 \tag{3}$$

  求解线性方程组(3)得到

$$x^2 = x^1 - [\nabla^2 f(x^1)]^{-1} \nabla f(x^1),$$

  作为x^*的第二次近似.
  如果x^1的精度不够,可以在x^1处将$f(x)$展开,求出近似二次函数的极值点x^3,如此下去,可以得到序列x^k 并且满足如下迭代公式

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k), \tag{4}$$

- 称(4)为Newtondi迭代共识. 为了计算方便,(4)改写成

$$x^{k+1} = x^k + d^k, \tag{5}$$

  其中d^k是线性方程组

$$\nabla^2 f(x^k) d = -\nabla f(x^k) \tag{6}$$

3.2.1牛顿方法的思想

- 令近似二次函数的导数为零,得
$$\nabla f(x^1) + \nabla^2 f(x^1)(x - x^1) = 0 \tag{3}$$
求解线性方程组(3)得到
$$x^2 = x^1 - [\nabla^2 f(x^1)]^{-1}\nabla f(x^1),$$
作为x^*的第二次近似.
如果x^1的精度不够,可以在x^1处将$f(x)$展开,求出近似二次函数的极值点x^3,如此下去,可以得到序列x^k 并且满足如下迭代公式
$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1}\nabla f(x^k), \tag{4}$$
- 称(4)为Newtondi迭代共识. 为了计算方便,(4)改写成
$$x^{k+1} = x^k + d^k, \tag{5}$$
其中d^k是线性方程组
$$\nabla^2 f(x^k)d = -\nabla f(x^k) \tag{6}$$

3.2.2牛顿算法的步骤

由上述分析,得到如下算法.

算法3.2.1(Newton法)

1、 取初始点x^1, 置精度要求ϵ, 置k=1.

2、 如果$\|\nabla f(x^k)\| \le \epsilon$, 则停止计算($x^k$作为无约束问题的解);否则求解线性方程组

$$\nabla^2 f(x^k)d = -\nabla f(x^k)$$

得到d^k.

3、 置

$$x^{k+1} = x^k + d^k, k = k+1,$$

转(2).

3.2.2牛顿算法的步骤

由上述分析,得到如下算法.

算法3.2.1(Newton法)

1、 取初始点x^1, 置精度要求ϵ, 置k=1.

2、 如果$||\nabla f(x^k)|| \le \epsilon$, 则停止计算($x^k$作为无约束问题的解);否则求解线性方程组

$$\nabla^2 f(x^k)d = -\nabla f(x^k)$$

得到d^k.

3、 置

$$x^{k+1} = x^k + d^k, k = k + 1,$$

转(2).

3.2.2牛顿算法的步骤

由上述分析,得到如下算法.

算法3.2.1(Newton法)

1、 取初始点x^1, 置精度要求ϵ, 置k=1.

2、 如果$||\nabla f(x^k)|| \leq \epsilon$, 则停止计算($x^k$作为无约束问题的解);否则求解线性方程组

$$\nabla^2 f(x^k)d = -\nabla f(x^k)$$

得到d^k.

3、 置

$$x^{k+1} = x^k + d^k, k = k + 1,$$

转(2).

例3.2.1用Newton法求解无约束问题

用Newton法求解无约束问题

$$\min f(x) = 4x_1^2 + x_2^2 - x_1^2 x_2$$

分别取初始点$x_A = (1,1)^T, x_B = (3,4)^T, x_C = (2,0)^T$,精度要求$\epsilon = 10^{-3}$.

解

- $$f(x) = 4x_1^2 + x_2^2 - x_1^2 x_2$$
  $$\nabla f(x) = (8x_1 - 2x_1x_2, 2x_2 - x_1^2)^T,$$
  $$\nabla^2 f(x) = \begin{pmatrix} 8 - 2x_2 & -2x_1 \\ -2x_1 & 2 \end{pmatrix}$$

- (1) 取$x^1 = x_A = (1,1)^T$, 得到$\nabla f(x^1) = (6,1)^T, \nabla^2 f(x^1) = \begin{pmatrix} 6 & -2 \\ -2 & 2 \end{pmatrix}$,解线性方程组
  $$\nabla^2 f(x^1) d = -\nabla f(x^1)$$
  即
  $$\begin{pmatrix} 6 & -2 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = -\begin{pmatrix} 6 \\ 1 \end{pmatrix}$$
  得到$d_1 = -1.75, d_2 = -2.25$,即$d^1 = (-1.75, -2.25)^T$. 所以
  $$x^2 = x^1 + d^1 = (1,1)^T + (-1.75, -2.25)^T = (-0.75, -1.25)^T$$

解

- $$f(x) = 4x_1^2 + x_2^2 - x_1^2 x_2$$
$$\nabla f(x) = (8x_1 - 2x_1x_2, 2x_2 - x_1^2)^T,$$
$$\nabla^2 f(x) = \begin{pmatrix} 8-2x_2 & -2x_1 \\ -2x_1 & 2 \end{pmatrix}$$

- (1) 取$x^1 = x_A = (1,1)^T$, 得到$\nabla f(x^1) = (6,1)^T, \nabla^2 f(x^1) = \begin{pmatrix} 6 & -2 \\ -2 & 2 \end{pmatrix}$,解线性方程组
$$\nabla^2 f(x^1) d = -\nabla f(x^1)$$
即
$$\begin{pmatrix} 6 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = -\begin{pmatrix} 6 \\ 1 \end{pmatrix}$$
得到$d_1 = -1.75, d_2 = -2.25$,即$d^1 = (-1.75, -2.25)^T$. 所以
$$x^2 = x^1 + d^1 = (1,1)^T + (-1.75, -2.25)^T = (-0.75, -1.25)^T$$

解

- 再进行第二轮计算, 经过5次迭代,x^k收敛到问题的极小点$x=(0,0)^T$.

表3.2.1 $x^1=(1,1)^T$的计算结果

k	x^k	$f(x^k)$	$\nabla f(x^k)$	$\|\nabla f(x^k)\|$	
1	$\begin{pmatrix}1.0000\\1.0000\end{pmatrix}$	4.0000	$\begin{pmatrix}6.0000\\1.0000\end{pmatrix}$	6.0928	6.0 / −2.
2	$\begin{pmatrix}-0.7500\\-1.2500\end{pmatrix}$	4.5156	$\begin{pmatrix}-7.8750\\-3.6350\end{pmatrix}$	8.4495	10 / 1.
3	$\begin{pmatrix}-0.1550\\-0.1650\end{pmatrix}$	0.1273	$\begin{pmatrix}-1.2911\\-0.3540\end{pmatrix}$	1.3388	8. / 0.
4	$\begin{pmatrix}-0.0057\\-0.0111\end{pmatrix}$	0.0003	$\begin{pmatrix}-0.0459\\-0.0223\end{pmatrix}$	0.0511	8. / 0.
5	$\begin{pmatrix}0.0000\\0.0000\end{pmatrix}$	0.0000	$\begin{pmatrix}-0.0001\\-0.0000\end{pmatrix}$	0.0001	8. / 0

- 再进行第二轮计算，经过5次迭代，x^k收敛到问题的极小点$x = (0, 0)^T$.

表3.2.1 $x^1 = (1, 1)^T$的计算结果

-

k	x^k	$f(x^k)$	$\nabla f(x^k)$	$\|\|\nabla f(x^k)\|\|$	
1	$\begin{pmatrix} 1.0000 \\ 1.0000 \end{pmatrix}$	4.0000	$\begin{pmatrix} 6.0000 \\ 1.0000 \end{pmatrix}$	6.0928	6.0 / −2.
2	$\begin{pmatrix} -0.7500 \\ -1.2500 \end{pmatrix}$	4.5156	$\begin{pmatrix} -7.8750 \\ -3.6350 \end{pmatrix}$	8.4495	10 / 1.
3	$\begin{pmatrix} -0.1550 \\ -0.1650 \end{pmatrix}$	0.1273	$\begin{pmatrix} -1.2911 \\ -0.3540 \end{pmatrix}$	1.3388	8. / 0.
4	$\begin{pmatrix} -0.0057 \\ -0.0111 \end{pmatrix}$	0.0003	$\begin{pmatrix} -0.0459 \\ -0.0223 \end{pmatrix}$	0.0511	8. / 0.
5	$\begin{pmatrix} 0.0000 \\ 0.0000 \end{pmatrix}$	0.0000	$\begin{pmatrix} -0.0001 \\ -0.0000 \end{pmatrix}$	0.0001	8. / 0

解

- (2)取$x^1 = x_B = (3,4)^T$,计算步骤同(1),最后x^k收敛到$(2\sqrt{2},4)^T$目标函数的鞍点,计算结果见下表.

表3.2.2 $x_1 = (3,4)^T$的计算结果

-

k	x^k	$f(x^k)$	$\nabla f(x^k)$	$\|\nabla f(x^k)\|$	
1	$\begin{pmatrix} 3.0000 \\ 4.0000 \end{pmatrix}$	16.0000	$\begin{pmatrix} 0.0000 \\ -1.0000 \end{pmatrix}$	1.0000	$\begin{pmatrix} 0.0 \\ -6. \end{pmatrix}$
2	$\begin{pmatrix} 2.8333 \\ 4.0000 \end{pmatrix}$	16.0000	$\begin{pmatrix} 0.0000 \\ -0.2078 \end{pmatrix}$	0.0278	$\begin{pmatrix} 0.0 \\ -5. \end{pmatrix}$
3	$\begin{pmatrix} 2.8284 \\ 4.0000 \end{pmatrix}$	16.0000	$\begin{pmatrix} 0.0000 \\ 0.0000 \end{pmatrix}$	0.0000	$\begin{pmatrix} 0.0 \\ -5. \end{pmatrix}$

解

- (2)取$x^1 = x_B = (3,4)^T$,计算步骤同(1),最后x^k收敛到$(2\sqrt{2},4)^T$目标函数的鞍点,计算结果见下表.

表3.2.2 $x_1 = (3,4)^T$的计算结果

-

k	x^k	$f(x^k)$	$\nabla f(x^k)$	$\|\|\nabla f(x^k)\|\|$	
1	$\begin{pmatrix} 3.0000 \\ 4.0000 \end{pmatrix}$	16.0000	$\begin{pmatrix} 0.0000 \\ -1.0000 \end{pmatrix}$	1.0000	$\begin{pmatrix} 0.0 \\ -6.0 \end{pmatrix}$
2	$\begin{pmatrix} 2.8333 \\ 4.0000 \end{pmatrix}$	16.0000	$\begin{pmatrix} 0.0000 \\ -0.2078 \end{pmatrix}$	0.0278	$\begin{pmatrix} 0.0 \\ -5.6 \end{pmatrix}$
3	$\begin{pmatrix} 2.8284 \\ 4.0000 \end{pmatrix}$	16.0000	$\begin{pmatrix} 0.0000 \\ 0.0000 \end{pmatrix}$	0.0000	$\begin{pmatrix} 0.0 \\ -5.6 \end{pmatrix}$

解

- (3) 取$x^1 = x_C = (2,0)^T$, 得到$\nabla^2 f(x^1) = \begin{pmatrix} 8 & -4 \\ -4 & 2 \end{pmatrix}$, Hessian矩阵奇异,无法进行下一步计算.

例3.2.1表明用Newton法求解无约束问题会出现以下情况:

- (1) 收敛到极小点;
- (2) 收敛到鞍点;
- (3) Hessian矩阵奇异,无法继续计算.

利用牛顿法求解最优化问题,其优点是收敛速度快,但是,它的缺点是每一步不能保证目标函数值总是下降,且当Hesse矩阵奇异时无法计算,所以还有其他相关的修正算法.

例3.2.1表明用Newton法求解无约束问题会出现以下情况:

- (1) 收敛到极小点;
- (2) 收敛到鞍点;
- (3) Hessian矩阵奇异,无法继续计算.

利用牛顿法求解最优化问题,其优点是收敛速度快,但是,它的缺点是每一步不能保证目标函数值总是下降,且当Hesse矩阵奇异时无法计算,所以还有其他相关的修正算法.

3.3共轭梯度法

共轭梯度法是用来求解正定二次规划的一种优化方法,它具有下述性质:

1、 产生的搜索方向是下降方向;

2、 不必计算Hessian矩阵,只计算目标函数值和梯度;

3、 具有二次终止性.

3.3共轭梯度法

共轭梯度法是用来求解正定二次规划的一种优化方法,它具有下述性质:

1、 产生的搜索方向是下降方向;
2、 不必计算Hessian矩阵, 只计算目标函数值和梯度;
3、 具有二次终止性.

3.3共轭梯度法

共轭梯度法是用来求解正定二次规划的一种优化方法,它具有下述性质:

1、 产生的搜索方向是下降方向;

2、 不必计算Hessian矩阵, 只计算目标函数值和梯度;

3、 具有二次终止性.

- 首先考虑一类特殊的正定二次函数

$$f(x) = \frac{1}{2}x^T x + r^T x + \delta \tag{7}$$

- 其中$x = (x_1, x_2, \cdots, x_n)^T, r = (r_1, r_2, \cdots, r_n)^T \in R^n, \delta \in R$. 注意到该特殊的正定二次函数可以看成如下形式的变量分离的函数

$$f(x) = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n)$$

- 其中,$f_i(x_i) = \frac{1}{2}x_i^2 + r_i x_i + \delta_i, i = 1, 2, \cdots, n$.则从任意一点$x^1$出发,依次沿着每个坐标轴方向进行一维搜索,进行一遍（共进行n次搜索）以后,就能得到$min f(x)$的最优解.

3.3.1正交方向和共轭方向

- 首先考虑一类特殊的正定二次函数

$$f(x) = \frac{1}{2}x^T x + r^T x + \delta \tag{7}$$

- 其中$x = (x_1, x_2, \cdots, x_n)^T, r = (r_1, r_2, \cdots, r_n)^T \in R^n, \delta \in R$. 注意到该特殊的正定二次函数可以看成如下形式的变量分离的函数

$$f(x) = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n)$$

- 其中,$f_i(x_i) = \frac{1}{2}x_i^2 + r_i x_i + \delta_i, i = 1, 2, \cdots, n$.则从任意一点$x^1$出发,依次沿着每个坐标轴方向进行一维搜索,进行一遍（共进行n次搜索）以后,就能得到$minf(x)$的最优解.

3.3.1正交方向和共轭方向

- 首先考虑一类特殊的正定二次函数

$$f(x) = \frac{1}{2}x^T x + r^T x + \delta \tag{7}$$

- 其中$x = (x_1, x_2, \cdots, x_n)^T, r = (r_1, r_2, \cdots, r_n)^T \in R^n, \delta \in R$. 注意到该特殊的正定二次函数可以看成如下形式的变量分离的函数

$$f(x) = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n)$$

- 其中,$f_i(x_i) = \frac{1}{2}x_i^2 + r_i x_i + \delta_i, i = 1, 2, \cdots, n$.则从任意一点$x^1$出发,依次沿着每个坐标轴方向进行一维搜索,进行一遍（共进行n次搜索）以后,就能得到$min f(x)$的最优解.

- 这里$\nabla f(x) = x + r$,令$\nabla f(x) = 0$,得$f(x)$的极小点$x^* = -r$,由$f(x)$在x^*的Taylor展开式得到

$$\begin{aligned} f(x) &= f(x^*) + \frac{1}{2}(x - x^*)^T(x - x^*) \\ &= f(x^*) + \frac{1}{2}||x - x^*)||^2 \end{aligned} \tag{8}$$

- (8)式表明$f(x)$的等高面是一族(超)球面.特别地,当$n = 2$时,其等高线是一族圆.任取初始点x^1沿两个相互正交的方向进行精确一维搜索,得到点$x^3 = x^*$, 即为问题的最优解.
- 当$n = 3$时,由类似的结果.设q^1, q^2, q^3是三个相互正交的非零向量,则任取初始点x^1,依次沿它们进行精确一维搜索, 达到极小点$x^4 = x^*$.特别注意:x^2是$f(x)$沿方向q^1生成的直线上的极小点,x^2是$f(x)$过x^1 由$\{q^1, q^2\}$张成的平面π上的极小点,x^4是整个空间上极小点.

3.3.1正交方向和共轭方向

- 这里$\nabla f(x)=x+r$,令$\nabla f(x)=0$,得$f(x)$的极小点$x^*=-r$,由$f(x)$在x^*的Taylor展开式得到

$$\begin{aligned} f(x) &= f(x^*)+\frac{1}{2}(x-x^*)^T(x-x^*) \\ &= f(x^*)+\frac{1}{2}\|x-x^*)\|^2 \end{aligned} \tag{8}$$

- (8)式表明$f(x)$的等高面是一族(超)球面.特别地,当$n=2$时,其等高线是一族圆.任取初始点x^1沿两个相互正交的方向进行精确一维搜索,得到点$x^3=x^*$, 即为问题的最优解.
- 当$n=3$时,由类似的结果.设q^1,q^2,q^3是三个相互正交的非零向量,则任取初始点x^1,依次沿它们进行精确一维搜索, 达到极小点$x^4=x^*$.特别注意:x^2是$f(x)$沿方向q^1生成的直线上的极小点,x^2是$f(x)$过x^1 由$\{q^1,q^2\}$张成的平面π上的极小点,x^4是整个空间上极小点.

- 这里$\nabla f(x)=x+r$,令$\nabla f(x)=0$,得$f(x)$的极小点$x^*=-r$,由$f(x)$在x^*的Taylor展开式得到

$$\begin{aligned} f(x) &= f(x^*)+\frac{1}{2}(x-x^*)^T(x-x^*) \\ &= f(x^*)+\frac{1}{2}||x-x^*)||^2 \end{aligned} \tag{8}$$

- (8)式表明$f(x)$的等高面是一族(超)球面.特别地,当$n=2$时,其等高线是一族圆.任取初始点x^1沿两个相互正交的方向进行精确一维搜索,得到点$x^3=x^*$, 即为问题的最优解.
- 当$n=3$时,由类似的结果.设q^1,q^2,q^3是三个相互正交的非零向量,则任取初始点x^1,依次沿它们进行精确一维搜索, 达到极小点$x^4=x^*$.特别注意:x^2是$f(x)$沿方向q^1生成的直线上的极小点,x^2是$f(x)$过x^1 由$\{q^1,q^2\}$张成的平面π上的极小点,x^4是整个空间上极小点.

定义3.3.1

若R^n中$k(k \leq n)$个向量$q^1, q^2, \cdots, q^k$两两正交,即

$$< q^i, q^j >= 0, i \neq j, 1 \leq i, j \leq k \tag{9}$$

则称它们为k个正交方向,若又满足

$$q^i \neq 0, i = 1, 2, \cdots, k, \tag{10}$$

则称为k个非零正交方向.

定理3.3.1

定理3.3.1

设目标函数为

$$f(x) = \frac{1}{2}x^T x + r^T x + \delta$$

$q^1, q^2, \cdots, q^k$是$k(k \le n)$个两两正交的非零向量.从任意初始点x^1出发,依次沿$q^1, q^2, \cdots, q^k$做精确一维搜索,得到$x^2, x^3, \cdots, x^{k+1}$,则$x^{k+1}$是$f(x)$在线性流形

$$\overline{x}^k = \{x = x^1 + \sum_{i=1}^{k} \widehat{\alpha}_i q^i | -\infty < \alpha_i < +\infty\} \tag{11}$$

上的唯一极小点.特别地,当$k = n$时,则x^{k+1}是$f(x)$在整个空间上的唯一极小点.

定义3.3.2

设G是$n \times n$正定矩阵.若d^1, d^2满足

$$(d^1)^T G d^2 = 0 \tag{12}$$

则称d^1, d^2关于G共轭.若$q^1, q^2, \cdots, q^k, (k \leq n)$两两关于G共轭,即

$$(d^i)^T G d^j = 0, i \neq j, i, j = 1, 2, \cdots, k \tag{13}$$

则称$q^1, q^2, \cdots, q^k, , (k \leq n)$为G的k个共轭方向.若$d^i \neq 0, i = 1, 2, \cdots, k$,则称为G的k个非零共轭方向. 特别,当G=I时,共轭方向就是正交方向.

定理3.3.2扩展子空间定理

扩展子空间定理

设目标函数为

$$f(x) = \frac{1}{2}x^T Gx + r^T x + \delta.$$

$d^1, d^2, \cdots, d^k$是G的$k(k \leq n)$个非零共轭方向,从任意的初始点x^1出发,依次沿$d^i \neq 0, i = 1, 2, \cdots, k,$ 作一维精确搜索,得到$x^2, x^3, \cdots, x^{k+1}$,则$x^{k+1}$是$f(x)$在线性流行

$$x^k = \{x = x^1 + \sum_{i=1}^{k} \alpha_i d^i | -\infty < \alpha_i < +\infty\}$$

上的唯一极小点.特别当$k = n$时,则x^{k+1}是$f(x)$在整个空间上的唯一极小点.

证明

- 因为G是正定矩阵,存在正定矩阵C使得$G = C^TC$.那么

$$f(x) = \frac{1}{2}x^TGx + r^Tx + \delta = \frac{1}{2}(Cx)^TCx + r^TC^{-1}Cx + \delta.$$

令$y = Cx, \bar{r} = r^TC^{-1}$,则

$$f(x) = g(y) = \frac{1}{2}y^Ty + \bar{r}^Ty + \delta.$$

- 由定理3.3.1,存在$k(k \le n)$个两两正交的非零向量$q^1, q^2, \cdots, q^k$使得目标函数$g(y)$依次沿着这些方向进行一维搜索得到y^*使得$g(y^*)$达到最小值. 这时,我们再令$d^i = C^{-1}q^i, i = 1, 2, \cdots, k$,则由$(q^i)^Tq^j = 0$得$(d^i)^TC^TCd^j = 0$,即$(d^i)^TGd^j = 0$,即$d^1, d^2, \cdots, d^k$关于$G$是非零共轭向量.而令$x^* = C^{-1}y^*$,则$f(x)$在$x^*$取到最小值$f(x^*) = g(y^*)$.

证明

- 因为G是正定矩阵,存在正定矩阵C使得$G = C^T C$.那么

$$f(x) = \frac{1}{2}x^T Gx + r^T x + \delta = \frac{1}{2}(Cx)^T Cx + r^T C^{-1} Cx + \delta.$$

令$y = Cx, \bar{r} = r^T C^{-1}$,则

$$f(x) = g(y) = \frac{1}{2}y^T y + \bar{r}^T y + \delta.$$

- 由定理3.3.1,存在$k(k \le n)$个两两正交的非零向量$q^1, q^2, \cdots, q^k$使得目标函数$g(y)$依次沿着这些方向进行一维搜索得到y^*使得$g(y^*)$达到最小值. 这时,我们再令$d^i = C^{-1}q^i, i = 1, 2, \cdots, k$,则由$(q^i)^T q^j = 0$得$(d^i)^T C^T C d^j = 0$,即$(d^i)^T G d^j = 0$,即$d^1, d^2, \cdots, d^k$关于$G$是非零共轭向量.而令$x^* = C^{-1}y^*$,则$f(x)$在$x^*$取到最小值$f(x^*) = g(y^*)$.

推论3.3.3

在定理3.2.2的假设下,有

$$< \nabla f(x^{k+1}), d_i >= 0, i = 1, 2, \cdots, k \tag{14}$$

推论3.3.3的证明

证明

因为x^{k+1}是线性流形x^k上的极小点,所以也是d^i方向上的极小点,因此(18)式成立.

共轭梯度法(Conjugate Gradient Mehtod)是共轭方向方法与梯度方法(即最速下降方法)结合起来考虑的一种优化算法.

3.3.2共轭梯度法的推导

设

$$f(x) = \frac{1}{2}x^T Gx + r^T x + \delta$$

其中G是正定对称矩阵,由扩展子空间定理可知,若$d^1, d^2, \cdots, d^n$为n个G共轭方向,那么从任意的初始点x^1出发,至多作n次精确一维搜索就可以得到目标函数唯一的极小点.根据这一思想导出对于正定二次函数的共轭梯度法.

3.3.2共轭梯度法的推导

- 任取初始点x^1.若$\nabla f(x_1)=0$,则停止计算,x^1作为无约束问题的极小点. 当$\nabla f(x_1)\neq 0$,令

$$d^1=-\nabla f(x^1), \tag{15}$$

- 沿d^1方向进行一维搜索得到点x^2, 若$\nabla f(x^2)\neq 0$, 令

$$d^2=-\nabla f(x^2)+\beta_1^{(2)}d^1, \tag{16}$$

并且使d^1,d^2满足

$$(d^1)^T G d^2=0, \tag{17}$$

- 即d^1,d^2关于G共轭.将(20)式带入(21)式,可得到$\beta_1^{(2)}$的表达式

$$\beta_1^{(2)}=\frac{(d^1)^T G\nabla f(x^2)}{(d^1)^T G d^1} \tag{18}$$

3.3.2共轭梯度法的推导

- 任取初始点x^1.若$\nabla f(x_1)=0$,则停止计算,x^1作为无约束问题的极小点. 当$\nabla f(x_1)\neq 0$,令

$$d^1=-\nabla f(x^1), \tag{15}$$

- 沿d^1方向进行一维搜索得到点x^2, 若$\nabla f(x^2)\neq 0$, 令

$$d^2=-\nabla f(x^2)+\beta_1^{(2)}d^1, \tag{16}$$

并且使d^1,d^2满足

$$(d^1)^T G d^2=0, \tag{17}$$

- 即d^1,d^2关于G共轭.将(20)式带入(21)式,可得到$\beta_1^{(2)}$的表达式

$$\beta_1^{(2)}=\frac{(d^1)^T G\nabla f(x^2)}{(d^1)^T G d^1} \tag{18}$$

- 任取初始点x^1.若$\nabla f(x_1)=0$,则停止计算,x^1作为无约束问题的极小点. 当$\nabla f(x_1)\neq 0$,令

$$d^1=-\nabla f(x^1), \tag{15}$$

- 沿d^1方向进行一维搜索得到点x^2, 若$\nabla f(x^2)\neq 0$, 令

$$d^2=-\nabla f(x^2)+\beta_1^{(2)}d^1, \tag{16}$$

并且使d^1, d^2满足

$$(d^1)^T G d^2=0, \tag{17}$$

- 即d^1, d^2关于G共轭.将(20)式带入(21)式,可得到$\beta_1^{(2)}$的表达式

$$\beta_1^{(2)}=\frac{(d^1)^T G\nabla f(x^2)}{(d^1)^T G d^1} \tag{18}$$

- 将此结果带入(20)式, 这样得到d^2是与d^1关于G共轭.再从x^2出发,沿d^2作一维搜索,得到x^3,如此下去,.... 假设在x^k处,$\nabla f(x^k) \neq 0$,构造x^k处的搜索方向d^k如下:
-
$$d^k = -\nabla f(x^k) + \sum_{i=1}^{k-1} \beta_i^{(k)} d^i, \tag{19}$$

- 将此结果带入(20)式, 这样得到d^2是与d^1关于G共轭.再从x^2出发,沿d^2作一维搜索,得到x^3,如此下去,.... 假设在x^k处,$\nabla f(x^k)\neq 0$,构造x^k处的搜索方向d^k如下:
- $$d^k = -\nabla f(x^k) + \sum_{i=1}^{k-1} \beta_i^{(k)} d^i, \tag{19}$$

- 下面来确定系数$\beta_i^{(k)}(i=1,2,\cdots,k-1)$.由于要求构造的搜索方向是关于G的共轭的,即满足

$$\begin{aligned} 0 &= (d^i)^T G d^k \\ &= -(d^i)^T G \nabla f(x^k) + \sum_{j=1}^{k-1} \beta_j^{(k)} (d^i)^T G d^{(j)} \\ &= -(d^i)^T G \nabla f(x^k) + \beta_i^{(k)} (d^i)^T G d^i, i=1,2,\cdots,k-1 \end{aligned}$$

- 所以

$$\beta_i^{(k)} = \frac{(d^i)^T G \nabla f(x^k)}{(d^i)^T G d^i} \tag{20}$$

- 将它带入(23)式,得到的d^k是与$d^1,d^2,\cdots,d^{k-1}$关于G共轭. 又由推导过程可知,前面已得到的搜索方向$d^1,d^2,\cdots,d^{k-1}$已是G的k-1个共轭方向,所以,$d^1,d^2,\cdots,d^k$是G的k个共轭方向.
- 由扩展子空间定理,当k=n时,得到n个非零的G共轭的方向,x^{n+1}为整个空间上的唯一极小点.

3.3.2共轭梯度法的推导

- 下面来确定系数$\beta_i^{(k)}(i=1,2,\cdots,k-1)$.由于要求构造的搜索方向是关于G的共轭的,即满足

$$
\begin{aligned}
0 &= (d^i)^T G d^k \\
&= -(d^i)^T G \nabla f(x^k) + \sum_{j=1}^{k-1} \beta_j^{(k)} (d^i)^T G d^{(j)} \\
&= -(d^i)^T G \nabla f(x^k) + \beta_i^{(k)} (d^i)^T G d^i, i = 1,2,\cdots,k-1
\end{aligned}
$$

- 所以

$$
\beta_i^{(k)} = \frac{(d^i)^T G \nabla f(x^k)}{(d^i)^T G d^i} \tag{20}
$$

- 将它带入(23)式,得到的d^k是与$d^1,d^2,\cdots,d^{k-1}$关于G共轭. 又由推导过程可知,前面已得到的搜索方向$d^1,d^2,\cdots,d^{k-1}$已是G的k-1个共轭方向,所以,$d^1,d^2,\cdots,d^k$是G的k个共轭方向.
- 由扩展子空间定理,当k=n时,得到n个非零的G共轭的方向,x^{n+1}为整个空间上的唯一极小点.

3.3.2共轭梯度法的推导

- 下面来确定系数$\beta_i^{(k)}(i=1,2,\cdots,k-1)$.由于要求构造的搜索方向是关于G的共轭的,即满足

$$\begin{aligned} 0 &= (d^i)^T G d^k \\ &= -(d^i)^T G \nabla f(x^k) + \sum_{j=1}^{k-1} \beta_j^{(k)} (d^i)^T G d^{(j)} \\ &= -(d^i)^T G \nabla f(x^k) + \beta_i^{(k)} (d^i)^T G d^i, i=1,2,\cdots,k-1 \end{aligned}$$

- 所以

$$\beta_i^{(k)} = \frac{(d^i)^T G \nabla f(x^k)}{(d^i)^T G d^i} \tag{20}$$

- 将它带入(23)式,得到的d^k是与$d^1, d^2, \cdots, d^{k-1}$关于G共轭. 又由推导过程可知,前面已得到的搜索方向$d^1, d^2, \cdots, d^{k-1}$已是G的k-1个共轭方向,所以,$d^1, d^2, \cdots, d^k$是G的k个共轭方向.
- 由扩展子空间定理,当k=n时,得到n个非零的G共轭的方向,x^{n+1}为整个空间上的唯一极小点.

- 下面来确定系数$\beta_i^{(k)}(i=1,2,\cdots,k-1)$.由于要求构造的搜索方向是关于G的共轭的,即满足

$$\begin{aligned} 0 &= (d^i)^T G d^k \\ &= -(d^i)^T G \nabla f(x^k) + \textstyle\sum_{j=1}^{k-1} \beta_j^{(k)} (d^i)^T G d^{(j)} \\ &= -(d^i)^T G \nabla f(x^k) + \beta_i^{(k)} (d^i)^T G d^i, i=1,2,\cdots,k-1 \end{aligned}$$

- 所以

$$\beta_i^{(k)} = \frac{(d^i)^T G \nabla f(x^k)}{(d^i)^T G d^i} \tag{20}$$

- 将它带入(23)式,得到的d^k是与$d^1,d^2,\cdots,d^{k-1}$关于G共轭. 又由推导过程可知,前面已得到的搜索方向$d^1,d^2,\cdots,d^{k-1}$已是G的k-1个共轭方向,所以,$d^1,d^2,\cdots,d^k$是G的k个共轭方向.
- 由扩展子空间定理,当k=n时,得到n个非零的G共轭的方向,x^{n+1}为整个空间上的唯一极小点.

3.3.3计算公式的简化

- 由于d^i的表达式中第一项是相应点处的负梯度,它要与后面点处的梯度作内积,因此这个特点可以简化计算公式.
- 由于构造的搜索方向是非零的G共轭方向,由推论3.3.3,得到

$$\nabla f(x^k)^T d^i = 0, i = 1, 2, \cdots, k-1$$

- 结合(23)式,有

$$\begin{aligned} & \nabla f(x^k)^T \nabla f(x^{(i)}) \\ = \; & \nabla f(x^k)^T(-d^i + \beta_1^{(i)} d^1 + \cdots + \beta_{i-1}^{(i)} d^{i-1}) \\ = \; & 0, i = 1, 2, \cdots, k-1 \end{aligned} \tag{21}$$

- 由于d^i的表达式中第一项是相应点处的负梯度,它要与后面点处的梯度作内积,因此这个特点可以简化计算公式.
- 由于构造的搜索方向是非零的G共轭方向,由推论3.3.3,得到

$$\nabla f(x^k)^T d^i = 0, i = 1, 2, \cdots, k-1$$

- 结合(23)式,有

$$\begin{aligned} & \nabla f(x^k)^T \nabla f(x^{(i)}) \\ = & \nabla f(x^k)^T(-d^i + \beta_1^{(i)} d^1 + \cdots + \beta_{i-1}^{(i)} d^{i-1}) \\ = & 0, i = 1, 2, \cdots, k-1 \end{aligned} \tag{21}$$

- 由于d^i的表达式中第一项是相应点处的负梯度,它要与后面点处的梯度作内积,因此这个特点可以简化计算公式.
- 由于构造的搜索方向是非零的G共轭方向,由推论3.3.3,得到

$$\nabla f(x^k)^T d^i = 0, i = 1, 2, \cdots, k-1$$

- 结合(23)式,有

$$\begin{aligned} & \nabla f(x^k)^T \nabla f(x^{(i)}) \\ = & \nabla f(x^k)^T(-d^i + \beta_1^{(i)} d^1 + \cdots + \beta_{i-1}^{(i)} d^{i-1}) \\ = & 0, i = 1, 2, \cdots, k-1 \end{aligned} \tag{21}$$

- 下面计算$\beta_i^{(k)}(i=1,2,\cdots,k-2)$. 为了方便起见,引进记号

$$s^i = x^{i+1} - x^i = \alpha_i d^i, \tag{22}$$

其中α_i是一维搜索的最优步长.

- 所以

$$Gs^i = Gx^{i+1} - Gx^i = \nabla f(x^{i+1}) - \nabla f(x^i) \tag{23}$$

- 由(17)式,有

$$\begin{aligned}(d^i)^T G\nabla f(x^k) &= \nabla f(x^k)^T G d^i \\ &= \frac{1}{\alpha_i}\nabla f(x^k)^T G s^i \\ &= \frac{1}{\alpha_i}\nabla f(x^k)^T(\nabla f(x^{i+1}) - \nabla f(x^i)) \\ &= 0, i = 1,2,\cdots,k-2\end{aligned} \tag{24}$$

3.3.3计算公式的简化

- 下面计算$\beta_i^{(k)}(i=1,2,\cdots,k-2)$. 为了方便起见,引进记号

$$s^i = x^{i+1} - x^i = \alpha_i d^i, \tag{22}$$

其中α_i是一维搜索的最优步长.

- 所以

$$Gs^i = Gx^{i+1} - Gx^i = \nabla f(x^{i+1}) - \nabla f(x^i) \tag{23}$$

- 由(17)式,有

$$\begin{aligned} (d^i)^T G \nabla f(x^k) &= \nabla f(x^k)^T G d^i \\ &= \frac{1}{\alpha_i} \nabla f(x^k)^T G s^i \\ &= \frac{1}{\alpha_i} \nabla f(x^k)^T (\nabla f(x^{i+1}) - \nabla f(x^i)) \\ &= 0, i = 1, 2, \cdots, k-2 \end{aligned} \tag{24}$$

3.3.3计算公式的简化

- 下面计算$\beta_i^{(k)}(i=1,2,\cdots,k-2)$. 为了方便起见,引进记号

$$s^i = x^{i+1} - x^i = \alpha_i d^i, \tag{22}$$

其中α_i是一维搜索的最优步长.

- 所以

$$Gs^i = Gx^{i+1} - Gx^i = \nabla f(x^{i+1}) - \nabla f(x^i) \tag{23}$$

- 由(17)式,有

$$\begin{aligned} (d^i)^T G \nabla f(x^k) &= \nabla f(x^k)^T G d^i \\ &= \frac{1}{\alpha_i} \nabla f(x^k)^T G s^i \\ &= \frac{1}{\alpha_i} \nabla f(x^k)^T (\nabla f(x^{i+1}) - \nabla f(x^i)) \\ &= 0, i = 1, 2, \cdots, k-2 \end{aligned} \tag{24}$$

3.3.3计算公式的简化

- 因此,(24)式可以简化为

$$\beta_i^{(k)} = 0, i = 1, 2, \cdots, k-2 \tag{25}$$

- 故(23)式可以写成

$$d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1}, \tag{26}$$

- (此时不要β_{k-1}上标)其中

$$\beta_{k-1} = \frac{(d^{k-1})^T G \nabla f(x^k)}{(d^{k-1})^T G d^{k-1}} \tag{27}$$

3.3.3计算公式的简化

- 因此,(24)式可以简化为

$$\beta_i^{(k)} = 0, i = 1, 2, \cdots, k-2 \tag{25}$$

- 故(23)式可以写成

$$d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1}, \tag{26}$$

- (此时不要β_{k-1}上标)其中

$$\beta_{k-1} = \frac{(d^{k-1})^T G \nabla f(x^k)}{(d^{k-1})^T G d^{k-1}} \tag{27}$$

3.3.3计算公式的简化

- 因此,(24)式可以简化为

$$\beta_i^{(k)} = 0, i = 1, 2, \cdots, k-2 \tag{25}$$

- 故(23)式可以写成

$$d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1}, \tag{26}$$

- (此时不要β_{k-1}上标)其中

$$\beta_{k-1} = \frac{(d^{k-1})^T G \nabla f(x^k)}{(d^{k-1})^T G d^{k-1}} \tag{27}$$

3.3.3计算公式的简化

- 再简化β_{k-1}, 类似于(28)的推导过程,可以得到

$$(d^{k-1})^T G\nabla f(x^k) = \frac{1}{\alpha_{k-1}}\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1})) \quad (28)$$

- 和

$$(d^{k-1})^T G d^{k-1} = \frac{1}{\alpha_{k-1}}(d^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1})) \quad (29)$$

- 再注意到定理2的推论1,

$$\begin{aligned} & (d^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1})) \\ = \; & -(d^{k-1})^T\nabla f(x^{k-1}) \\ = \; & (\nabla f(x^{(k-1)}) - \beta_{k-2}d^{k-2})^T\nabla f(x^{k-1}) \\ = \; & \nabla f(x^{k-1})^T\nabla f(x^{k-1}), \end{aligned} \quad (30)$$

3.3.3计算公式的简化

- 再简化β_{k-1}, 类似于(28)的推导过程,可以得到

$$(d^{k-1})^T G\nabla f(x^k) = \frac{1}{\alpha_{k-1}}\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1})) \quad (28)$$

- 和

$$(d^{k-1})^T G d^{k-1} = \frac{1}{\alpha_{k-1}}(d^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1})) \quad (29)$$

- 再注意到定理2的推论1,

$$\begin{aligned} & (d^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1})) \\ = & -(d^{k-1})^T\nabla f(x^{k-1}) \\ = & (\nabla f(x^{(k-1)}) - \beta_{k-2}d^{k-2})^T\nabla f(x^{k-1}) \\ = & \nabla f(x^{k-1})^T\nabla f(x^{k-1}), \end{aligned} \quad (30)$$

- 再简化β_{k-1}, 类似于(28)的推导过程,可以得到

$$(d^{k-1})^T G\nabla f(x^k) = \frac{1}{\alpha_{k-1}}\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1})) \quad (28)$$

- 和

$$(d^{k-1})^T G d^{k-1} = \frac{1}{\alpha_{k-1}}(d^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1})) \quad (29)$$

- 再注意到定理2的推论1,

$$\begin{aligned} & (d^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1})) \\ = & -(d^{k-1})^T\nabla f(x^{k-1}) \\ = & (\nabla f(x^{(k-1)}) - \beta_{k-2}d^{k-2})^T\nabla f(x^{k-1}) \\ = & \nabla f(x^{k-1})^T\nabla f(x^{k-1}), \end{aligned} \quad (30)$$

- 所以

$$\beta_{k-1} = \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{\nabla f(x^{k-1})^T \nabla f(x^{k-1}))} \tag{31}$$

- 或

$$\beta_{k-1} = \frac{||\nabla f(x^k)||^2}{||\nabla f(x^{k-1})||^2} \tag{32}$$

- 所以

$$\beta_{k-1} = \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{\nabla f(x^{k-1})^T \nabla f(x^{k-1}))} \tag{31}$$

- 或

$$\beta_{k-1} = \frac{||\nabla f(x^k)||^2}{||\nabla f(x^{k-1})||^2} \tag{32}$$

- 这样就得到了用于一般可微函数的共轭梯度法. 其搜索方向构造如下:

$$\begin{cases} d_1 = -\nabla f(x_1), \\ d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1}, \end{cases} \tag{33}$$

- 由(36)和(37)式构造的计算公式称为PRP(Pulak-Ribiere-Polyak)公式,相应的方法称为PRP算法.
- 由(35)和(37)式构造的公式称为FR (Fletcher-Reeves)公式,相应的算法称为FR算法.

- 这样就得到了用于一般可微函数的共轭梯度法. 其搜索方向构造如下:

$$\begin{cases} d_1 = -\nabla f(x_1), \\ d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1}, \end{cases} \tag{33}$$

- 由(36)和(37)式构造的计算公式称为PRP(Pulak-Ribiere-Polyak)公式,相应的方法称为PRP算法.
- 由(35)和(37)式构造的公式称为FR (Fletcher-Reeves)公式,相应的算法称为FR算法.

- 这样就得到了用于一般可微函数的共轭梯度法. 其搜索方向构造如下:

$$\begin{cases} d_1 = -\nabla f(x_1), \\ d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1}, \end{cases} \tag{33}$$

- 由(36)和(37)式构造的计算公式称为PRP(Pulak-Ribiere-Polyak)公式,相应的方法称为PRP算法.
- 由(35)和(37)式构造的公式称为FR (Fletcher-Reeves)公式,相应的算法称为FR算法.

算法6.2.1(FR算法或PRP算法)

FR算法或PRP算法

1、 取初始点x_1,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \leq \epsilon$,停止计算($x^k$作为无约束问题的解);否则置

$$d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1},$$

其中

$$\beta_{k-1} = \begin{cases} 0. & k = 1 \\ \frac{||\nabla f(x^k)||^2}{||\nabla f(x^{k-1})||^2}, & k > 1 \end{cases}$$

或

$$\beta_{k-1} = \begin{cases} 0. & k = 1 \\ \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{\nabla f(x^{k-1})^T \nabla f(x^{k-1})}, & k > 1 \end{cases}$$

FR算法或PRP算法

1、 取初始点x_1,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \leq \epsilon$,停止计算($x^k$作为无约束问题的解);否则置

$$d^k = -\nabla f(x^k) + \beta_{k-1} d^{k-1},$$

其中

$$\beta_{k-1} = \begin{cases} 0. & k = 1 \\ \frac{||\nabla f(x^k)||^2}{||\nabla f(x^{k-1})||^2}, & k > 1 \end{cases}$$

或

$$\beta_{k-1} = \begin{cases} 0. & k = 1 \\ \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{\nabla f(x^{k-1})^T \nabla f(x^{k-1})}, & k > 1 \end{cases}$$

FR算法或PRP算法

3、 一维搜索.求解一维问题

$$\min\phi(\alpha) = f(x^k + \alpha d^k)$$

得α_k,置

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

4、 置k=k+1, 转(2).

FR算法或PRP算法

3、 一维搜索.求解一维问题

$$\min \phi(\alpha) = f(x^k + \alpha d^k)$$

得α_k,置

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

4、 置k=k+1, 转(2).

例3.3.1: 用共轭梯度法(FR算法)求解无约束问题

用共轭梯度法(FR算法)求解无约束问题

$$\min f(x) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1,$$

取$x_1 = (0,0)^T$.

解

- $$\nabla f(x) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$
  $$\nabla^2 f(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = G$$
- 一维搜索步长为
  $$\alpha_k = \frac{(d^k)^T \nabla f(x^k)}{(d^k)^T G d^k} = -\frac{d_1 g_1 + d_2 g_2}{3d_1^2 + d_2^2 - 2d_1 d_2} \tag{34}$$
- 取$x_1 = (0,0)^T$，则$\nabla f(x_1) = (-2,0)^T$, $d_1 = -\nabla f(x_1) = (2,0)^T$. 由(38)得步长$\alpha_1 = \frac{2^2}{3\cdot 2^2} = \frac{1}{3}$,因此
  $$x_2 = x_1 + \alpha_1 d_1 = (0,0)^T + \frac{1}{3}(2,0)^T = (\frac{2}{3},0)^T$$

解

- $$\nabla f(x) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$
  $$\nabla^2 f(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = G$$
- 一维搜索步长为
  $$\alpha_k = \frac{(d^k)^T \nabla f(x^k)}{(d^k)^T G d^k} = -\frac{d_1 g_1 + d_2 g_2}{3d_1^2 + d_2^2 - 2d_1 d_2} \qquad (34)$$
- 取$x_1 = (0,0)^T$, 则$\nabla f(x_1) = (-2,0)^T, d_1 = -\nabla f(x_1) = (2,0)^T$. 由(38)得步长$\alpha_1 = \frac{2^2}{3 \cdot 2^2} = \frac{1}{3}$,因此
  $$x_2 = x_1 + \alpha_1 d_1 = (0,0)^T + \frac{1}{3}(2,0)^T = (\frac{2}{3},0)^T$$

解

-
$$\nabla f(x) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$
$$\nabla^2 f(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = G$$
- 一维搜索步长为
$$\alpha_k = \frac{(d^k)^T \nabla f(x^k)}{(d^k)^T G d^k} = -\frac{d_1 g_1 + d_2 g_2}{3d_1^2 + d_2^2 - 2d_1 d_2} \tag{34}$$
- 取$x_1 = (0,0)^T$, 则$\nabla f(x_1) = (-2,0)^T, d_1 = -\nabla f(x_1) = (2,0)^T$. 由(38)得步长$\alpha_1 = \frac{2^2}{3\cdot 2^2} = \frac{1}{3}$,因此
$$x_2 = x_1 + \alpha_1 d_1 = (0,0)^T + \frac{1}{3}(2,0)^T = (\frac{2}{3},0)^T$$

解

- 再计算第二轮循环:因为$\nabla f(x_2) = (0, -\frac{2}{3})^T$, 有

$$\beta_1 = \frac{||\nabla f(x_2)||^2}{||\nabla f(x_1)||^2} = \frac{\left(\frac{2}{3}\right)^2}{2^2} = \frac{1}{9},$$

- 因此

$$d_2 = -\nabla f(x_2) + \beta_1 d_1 = (0, \frac{2}{3})^T + \frac{1}{9}(2, 0)^T = \left(\frac{2}{9}, \frac{2}{3}\right)^T.$$

- 步长为

$$\alpha_2 = -\frac{\frac{2}{9}\cdot 0 + \frac{2}{3}\cdot\left(-\frac{2}{3}\right)}{3\left(\frac{2}{9}\right)^2 + \left(\frac{2}{3}\right)^2 - 2\cdot\frac{2}{9}\cdot\frac{2}{3}} = \frac{3}{2}$$

解

- 再计算第二轮循环:因为$\nabla f(x_2) = (0, -\frac{2}{3})^T$, 有

$$\beta_1 = \frac{||\nabla f(x_2)||^2}{||\nabla f(x_1)||^2} = \frac{\left(\frac{2}{3}\right)^2}{2^2} = \frac{1}{9},$$

- 因此

$$d_2 = -\nabla f(x_2) + \beta_1 d_1 = (0, \frac{2}{3})^T + \frac{1}{9}(2,0)^T = \left(\frac{2}{9}, \frac{2}{3}\right)^T.$$

- 步长为

$$\alpha_2 = -\frac{\frac{2}{9}\cdot 0 + \frac{2}{3}\cdot\left(-\frac{2}{3}\right)}{3\left(\frac{2}{9}\right)^2 + \left(\frac{2}{3}\right)^2 - 2\cdot\frac{2}{9}\cdot\frac{2}{3}} = \frac{3}{2}$$

解

- 再计算第二轮循环:因为$\nabla f(x_2) = (0, -\frac{2}{3})^T$, 有

$$\beta_1 = \frac{||\nabla f(x_2)||^2}{||\nabla f(x_1)||^2} = \frac{\left(\frac{2}{3}\right)^2}{2^2} = \frac{1}{9},$$

- 因此

$$d_2 = -\nabla f(x_2) + \beta_1 d_1 = (0, \frac{2}{3})^T + \frac{1}{9}(2, 0)^T = \left(\frac{2}{9}, \frac{2}{3}\right)^T.$$

- 步长为

$$\alpha_2 = -\frac{\frac{2}{9}\cdot 0 + \frac{2}{3}\cdot\left(-\frac{2}{3}\right)}{3\left(\frac{2}{9}\right)^2 + \left(\frac{2}{3}\right)^2 - 2\cdot\frac{2}{9}\cdot\frac{2}{3}} = \frac{3}{2}$$

解

- 由此得到

$$x_3 = x_2 + \alpha_2 d_2 = \left(\frac{2}{3}, 0\right)^T + \frac{3}{2}\left(\frac{2}{9}, \frac{2}{3}\right)^T = (1,1)^T.$$

- 计算梯度$\nabla^f(x_3) = (0,0)^T, ||f(x_3)|| = 0.$
- 所以$x_3 = (1,1)^T$为最优解, 最优函数值为$f(x_3) = -1.$

解

- 由此得到

$$x_3 = x_2 + \alpha_2 d_2 = \left(\frac{2}{3}, 0\right)^T + \frac{3}{2}\left(\frac{2}{9}, \frac{2}{3}\right)^T = (1,1)^T.$$

- 计算梯度$\nabla^f(x_3) = (0,0)^T, ||f(x_3)|| = 0.$
- 所以$x_3 = (1,1)^T$为最优解, 最优函数值为$f(x_3) = -1.$

解

- 由此得到

$$x_3 = x_2 + \alpha_2 d_2 = \left(\frac{2}{3}, 0\right)^T + \frac{3}{2}\left(\frac{2}{9}, \frac{2}{3}\right)^T = (1,1)^T.$$

- 计算梯度$\nabla^f(x_3) = (0,0)^T, ||f(x_3)|| = 0.$
- 所以$x_3 = (1,1)^T$为最优解, 最优函数值为$f(x_3) = -1.$

解

- 另外,还有两个计算β_{k-1}的公式,和

$$\beta_{k-1} = \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{(d^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1}))}$$

称为Growder-Wolfe公式,和

$$\beta_{k-1} = \frac{\nabla f(x^k)^T \nabla f(x^k)}{(d^{k-1})^T \nabla f(x^{k-1})}$$

称为Dixon公式.

3.3.4共轭方向的下降性和算法的二次终止性

共轭梯度法虽然是针对正定二次函数导出的, 但仍适用于一般的可微函数. 这里首先需要解决的一个问题:共轭梯度法产生的搜索方向是否是下降方向?

定理3.3.4

定理3.3.4

设$f(x)$具有连续的一阶偏导数,并假设一维搜索是精确的, 考虑用共轭梯度法求解无约束问题$\min f(x), x \in R^n$. 若$\nabla f(x^k) \neq 0$, 则搜索方向d^k是x^k处的下降方向.

证明

- 考虑到$f(x)$在x^k处沿d^k的展开式，只需证明$\nabla f(x^k)^T d^k < 0$. 注意到x^k是d^{k-1}方向上的精确极小点,有$\nabla f(x^k)^T d^{k-1} = 0$,
- 因此

$$\begin{aligned} \nabla f(x^k)^T d^k &= -\nabla f(x^k)^T \nabla f(x^k) + \beta_{k-1} \nabla f(x^k)^T d^{k-1} \\ &= -\|\nabla f(x^k)\|^2 < 0 \end{aligned} \tag{35}$$

定理3.3.4的证明

证明

- 考虑到$f(x)$在x^k处沿d^k的展开式，只需证明$\nabla f(x^k)^T d^k < 0$. 注意到x^k是d^{k-1}方向上的精确极小点,有$\nabla f(x^k)^T d^{k-1} = 0$,
- 因此

$$\begin{aligned}\nabla f(x^k)^T d^k &= -\nabla f(x^k)^T \nabla f(x^k) + \beta_{k-1} \nabla f(x^k)^T d^{k-1} \\ &= -||\nabla f(x^k)||^2 < 0\end{aligned} \tag{35}$$

定理3.3.5精确一维搜索

精确一维搜索

若一维搜索是精确的，则共轭梯度法具有二次终止性.

(1) 考虑用梯度法求解正定二次目标函数,若$\nabla f(x^k)=0,(k\le 0)$,则x^k为无约束问题的最优解;否则由算法得到的搜索方向$d_1,d_2,\cdots,d^n$是关于二次部分正定矩阵共轭的,由扩展子空间定理可知,x^{n+1}是无约束问题的最优解.

(2) 由定理可知,对于正定二次函数,共轭梯度法至多n步终止. 如果算法在n步没有终止,则说明目标函数不是正定二次函数,或者说目标函数没有进入一个正定二次函数的区域,因此这种共轭已经没有意义,此时,搜索方向应重新开始,即令

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

(3) 算法每n步重新开始一次,称为n步重新开始策略.实际计算表明,n步重新开始的FR算法要优越于FR算法. 对于PFP算法,当$\nabla f(x^k)\approx -\nabla f(x^{k-1})$时,有$\beta_{k-1}\approx 0$, 因此$d^k\approx \nabla f(x^k)$,即自动重新开始.试验表明对于大型问题,PRP算法优越于FR算法.

3.4拟牛顿法

从前面几节的内容知,最速下降法具有结构简单,计算量小的优点,但其收敛速度较慢,牛顿法和修正牛顿法, 虽然收敛速度快,但在迭代过程中的每一步构造搜索方向时,首先要计算目标函数的Hesse阵,然后需要解一个线性方程组, 计算工作量很大, 这就抵消了牛顿法收敛速度快的优点.本节介绍的拟牛顿法构造搜索方向只需利用目标函数及其一阶导数的信息, 避免了Hessian阵的计算,减少了计算量,且具有超线性收敛的优点.拟牛顿法也称为变度量法.

在修正Newton法中,需要求解Newton方程

$$\nabla^2 f(x^k)d = -\nabla f(x^k)$$

来确定搜索方向.如果用迭代过程中的某些性质来确定一个近似于Hessian矩阵的矩阵B^k来代替Hessian矩阵.用方程组

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

的解向量d^k来作为搜索方向.就得到一个新的算法.称此类算法为拟牛顿法

3.4.1拟牛顿法的导出

- 由于B^k近似于$\nabla^2 f(x^k)$,可以猜想,该算法应该有较好的收敛速率,并且克服了计算量大的缺点.那么问题的关键是如何构造矩阵B^k? 首先,B^k不可能在数值上近似$\nabla^2 f(x^k)$,因为我们的目标是不计算$\nabla^2 f(x^k)$.那么B^k只能在性质上近似$\nabla^2 f(x^k)$.因此,先研究一下Hessian矩阵$\nabla^2 f(x^k)$具有什么样的性质.
- 考虑$\nabla f(x)$在x^{k+1}处的Taylor展开式

$$\nabla f(x) \approx \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x - x^{k+1}) \tag{36}$$

取$x = x^k$得到

$$\nabla f(x^{k+1}) - \nabla f(x^k) \approx \nabla^2 f(x^{k+1})(x^{k+1} - x^k) \tag{37}$$

- 记

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

- 由于B^k近似于$\nabla^2 f(x^k)$,可以猜想,该算法应该有较好的收敛速率,并且克服了计算量大的缺点.那么问题的关键是如何构造矩阵B^k? 首先,B^k不可能在数值上近似$\nabla^2 f(x^k)$,因为我们的目标是不计算$\nabla^2 f(x^k)$.那么B^k只能在性质上近似$\nabla^2 f(x^k)$.因此,先研究一下Hessian矩阵$\nabla^2 f(x^k)$具有什么样的性质.
- 考虑$\nabla f(x)$在x^{k+1}处的Taylor展开式

$$\nabla f(x) \approx \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x - x^{k+1}) \tag{36}$$

取$x = x^k$得到

$$\nabla f(x^{k+1}) - \nabla f(x^k) \approx \nabla^2 f(x^{k+1})(x^{k+1} - x^k) \tag{37}$$

- 记

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

3.4.1拟牛顿法的导出

- 由于B^k近似于$\nabla^2 f(x^k)$,可以猜想,该算法应该有较好的收敛速率,并且克服了计算量大的缺点.那么问题的关键是如何构造矩阵B^k? 首先,B^k不可能在数值上近似$\nabla^2 f(x^k)$,因为我们的目标是不计算$\nabla^2 f(x^k)$.那么B^k只能在性质上近似$\nabla^2 f(x^k)$.因此,先研究一下Hessian矩阵$\nabla^2 f(x^k)$具有什么样的性质.
- 考虑$\nabla f(x)$在x^{k+1}处的Taylor展开式

$$\nabla f(x) \approx \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x - x^{k+1}) \tag{36}$$

取$x = x^k$得到

$$\nabla f(x^{k+1}) - \nabla f(x^k) \approx \nabla^2 f(x^{k+1})(x^{k+1} - x^k) \tag{37}$$

- 记

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

3.4.1拟牛顿法的导出

- 所以(41)就可以改写成

$$\nabla^2 f(x^{k+1})s^k \approx y^k \tag{38}$$

- 由于要求B^k近似$\nabla^2 f(x^k)$, 即B^{k+1}近似$\nabla^2 f(x^{k+1})$,也就是B^{k+1}满足(42)式,并将≈改写成=,
- 得到

$$B^{k+1}s^k = y^k \tag{39}$$

称(44)式为拟Newton方程,并且假设B^{k+1}是对称矩阵.

- 所以(41)就可以改写成

$$\nabla^2 f(x^{k+1})s^k \approx y^k \tag{38}$$

- 由于要求B^k近似$\nabla^2 f(x^k)$, 即B^{k+1}近似$\nabla^2 f(x^{k+1})$,也就是B^{k+1}满足(42)式,并将≈改写成=,
- 得到

$$B^{k+1}s^k = y^k \tag{39}$$

称(44)式为拟Newton方程,并且假设B^{k+1}是对称矩阵.

- 所以(41)就可以改写成

$$\nabla^2 f(x^{k+1})s^k \approx y^k \tag{38}$$

- 由于要求B^k近似$\nabla^2 f(x^k)$，即B^{k+1}近似$\nabla^2 f(x^{k+1})$,也就是B^{k+1}满足(42)式,并将≈改写成=,
- 得到

$$B^{k+1}s^k = y^k \tag{39}$$

称(44)式为拟Newton方程,并且假设B^{k+1}是对称矩阵.

3.4.1拟牛顿法的导出

- 假设B^{k+1}是由B^k结果修正得到的.

$$B^{k+1)} = B^k + \triangle B^k \tag{40}$$

- 设$\triangle B^k = uv^T$, 即

$$B^{k+1} = B^k + uv^T \tag{41}$$

$$y^k = B^{k+1}s^k = B^k s^k + (v^T s^k)u \tag{42}$$

- 所以

$$y^k - B^k s^{(k)} = (v^T s^k)u \tag{43}$$

- 假设B^{k+1}是由B^k结果修正得到的.

$$B^{k+1}) = B^k + \triangle B^k \tag{40}$$

- 设$\triangle B^k = uv^T$, 即

$$B^{k+1} = B^k + uv^T \tag{41}$$

$$y^k = B^{k+1}s^k = B^k s^k + (v^T s^k)u \tag{42}$$

- 所以

$$y^k - B^k s^{(k)} = (v^T s^k)u \tag{43}$$

- 假设B^{k+1}是由B^k结果修正得到的.

$$B^{k+1}) = B^k + \triangle B^k \tag{40}$$

- 设$\triangle B^k = uv^T$, 即

$$B^{k+1} = B^k + uv^T \tag{41}$$

$$y^k = B^{k+1}s^k = B^k s^k + (v^T s^k)u \tag{42}$$

- 所以

$$y^k - B^k s^{(k)} = (v^T s^k)u \tag{43}$$

3.4.1拟牛顿法的导出

- 只需取$u = y^k - B^k s^k$, v满足

$$v^T s^k = 1 \tag{44}$$

则(53)式成立.

- 由B^k对称,并希望B^{k+1}对称,因此只需v与u共线,即

$$v = \lambda u = \lambda(y^k - B^k s^k) \tag{45}$$

- 将(54)式代入(53)式,得到

$$1 = \lambda(y^k - B^k s^k) s^k$$

3.4.1拟牛顿法的导出

- 只需取$u = y^k - B^k s^k$, v满足

$$v^T s^k = 1 \tag{44}$$

则(53)式成立.

- 由B^k对称,并希望B^{k+1}对称,因此只需v与u共线,即

$$v = \lambda u = \lambda(y^k - B^k s^k) \tag{45}$$

- 将(54)式代入(53)式,得到

$$1 = \lambda(y^k - B^k s^k)s^k$$

- 只需取$u = y^k - B^k s^k$, v满足

$$v^T s^k = 1 \tag{44}$$

则(53)式成立.

- 由B^k对称,并希望B^{k+1}对称,因此只需v与u共线,即

$$v = \lambda u = \lambda(y^k - B^k s^k) \tag{45}$$

- 将(54)式代入(53)式,得到

$$1 = \lambda(y^k - B^k s^k)s^k$$

- 所以

$$\lambda = \frac{1}{(y^k - B^k s^k) s^k} \tag{46}$$

- 结合(55)和(51)得到

$$B^{k+1} = B^k + \frac{(y^k - B^k s^k)(y^k - B^k s^k)^T}{(y^k - B^k s^k)^T s^k} \tag{47}$$

- (57)式称为对B^k的对称秩1修正公式.由此可推导出相应的拟牛顿法.

- 所以

$$\lambda = \frac{1}{(y^k - B^k s^k) s^k} \tag{46}$$

- 结合(55)和(51)得到

$$B^{k+1} = B^k + \frac{(y^k - B^k s^k)(y^k - B^k s^k)^T}{(y^k - B^k s^k)^T s^k} \tag{47}$$

- (57)式称为对B^k的对称秩1修正公式.由此可推导出相应的拟牛顿法.

- 所以

$$\lambda = \frac{1}{(y^k - B^k s^k) s^k} \tag{46}$$

- 结合(55)和(51)得到

$$B^{k+1} = B^k + \frac{(y^k - B^k s^k)(y^k - B^k s^k)^T}{(y^k - B^k s^k)^T s^k} \tag{47}$$

- (57)式称为对B^k的对称秩1修正公式.由此可推导出相应的拟牛顿法.

算法3.3.2 (对称秩1修正公式的拟牛顿法)

1、 取初始点x^1,置初始矩阵$B^1(=I)$,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \le \epsilon$, 则停止计算($x^k$作为无约束问题的最优解);否则求解线性方程组

$$B^k d = -\nabla f(x^k),$$

得到d^k.

3、 一维搜索.求解一维问题

$$\mathrm{min}\phi(\alpha) = f(x^k + \alpha d^k),$$

得α_k.置

$$x^{k+1} = x^k + \alpha d^k$$

算法3.3.2 (对称秩1修正公式的拟牛顿法)

1、 取初始点x^1,置初始矩阵$B^1(=I)$,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \leq \epsilon$, 则停止计算($x^k$作为无约束问题的最优解);否则求解线性方程组

$$B^k d = -\nabla f(x^k),$$

得到d^k.

3、 一维搜索.求解一维问题

$$\mathrm{min}\phi(\alpha) = f(x^k + \alpha d^k),$$

得α_k.置

$$x^{k+1} = x^k + \alpha d^k$$

算法3.3.2 (对称秩1修正公式的拟牛顿法)

1、 取初始点x^1,置初始矩阵$B^1(=I)$,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \leq \epsilon$, 则停止计算($x^k$作为无约束问题的最优解);否则求解线性方程组

$$B^k d = -\nabla f(x^k),$$

得到d^k.

3、 一维搜索.求解一维问题

$$\mathrm{min}\,\phi(\alpha) = f(x^k + \alpha d^k),$$

得α_k.置

$$x^{k+1} = x^k + \alpha d^k$$

对称秩1修正公式的拟牛顿法

4、 修正B^k,记

$$s^k = x^{k+1} - x^k, y^k = -\alpha f(x^{k+1}) - \alpha f(x^k)$$

置

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

5、 置$k = k+1$,转(2).

算法3.3.2 (对称秩1修正公式的拟牛顿法)

对称秩1修正公式的拟牛顿法

4、 修正B^k,记

$$s^k = x^{k+1} - x^k, y^k = -\alpha f(x^{k+1}) - \alpha f(x^k)$$

置

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

5、 置$k = k + 1$,转(2).

前面介绍了对称秩1修正公式的拟牛顿法. 但秩1公式有两个主要缺点.

1、 要求$(y^k - B^k s^k)^T s^k \neq 0$,若出现$(y^k - B^k s^k)^T s^k = 0$,则无法进行进一步计算.

2、 不能保证正定性的传递,即B^k正定,由秩1公式得到的B^{k+1}不一定正定.只有$(y^k - B^k s^k)^T s^k > 0$ 才能保证B^{k+1}正定,而这个条件在算法中很难实现,即使$(y^k - B^k s^k)^T s^k > 0$,它也可能很小,会产生较大的舍入误差,数值计算效果不好.这样使得秩1修正公式在应用中受到限制.

前面介绍了对称秩1修正公式的拟牛顿法. 但秩1公式有两个主要缺点.

1、 要求$(y^k - B^k s^k)^T s^k \neq 0$,若出现$(y^k - B^k s^k)^T s^k = 0$,则无法进行进一步计算.

2、 不能保证正定性的传递,即B^k正定,由秩1公式得到的B^{k+1}不一定正定.只有$(y^k - B^k s^k)^T s^k > 0$ 才能保证B^{k+1}正定,而这个条件在算法中很难实现,即使$(y^k - B^k s^k)^T s^k > 0$,它也可能很小,会产生较大的舍入误差,数值计算效果不好.这样使得秩1修正公式在应用中受到限制.

- 为了克服秩1 公式的缺点木块率对称秩2修正公式. 设

$$\triangle B^k = u^1(v^1)^T + u^2(v^2)^T \tag{48}$$

即

$$B^{k+1} = B^k + u^1(v^1)^T + u^2(v^2)^T \tag{49}$$

- 将(59)代入拟Newton方程得到

$$B^k s^k + ((v^1)^T s^k)u^1 + ((v^2)^T s^k)u^2 = y^k \tag{50}$$

- 为了克服秩1 公式的缺点木块率对称秩2修正公式. 设

$$\triangle B^k = u^1(v^1)^T + u^2(v^2)^T \tag{48}$$

即

$$B^{k+1} = B^k + u^1(v^1)^T + u^2(v^2)^T \tag{49}$$

- 将(59)代入拟Newton方程得到

$$B^k s^k + ((v^1)^T s^k)u^1 + ((v^2)^T s^k)u^2 = y^k \tag{50}$$

- 与秩1公式类似推导,只需取$u^i, v^i (i = 1, 2)$满足

$$\begin{cases} u^1 = y^k, \\ (v^1)^T s^k = 1 \end{cases} \begin{cases} u^2 = -B^k s^k \\ (v^2)^T s^k = 1 \end{cases} \tag{51}$$

- 则(60)式成立. 再由B^k的对称性和类似于秩1 公式的推导过程,得到

$$B^{k+1} = B^k - \frac{B^k s^k (s^k)^T B^k}{(s^k)^T B^k s^k} + \frac{y^k (y^k)^T}{(y^k)^T s^k} \tag{52}$$

- (62)式称为对B^k的对称秩2修正公式,也称为BFGS公式相应的算法称为BFGS算法.

- 与秩1公式类似推导,只需取$u^i, v^i(i=1,2)$满足

$$\begin{cases} u^1 = y^k, \\ (v^1)^T s^k = 1 \end{cases} \begin{cases} u^2 = -B^k s^k \\ (v^2)^T s^k = 1 \end{cases} \tag{51}$$

- 则(60)式成立. 再由B^k的对称性和类似于秩1 公式的推导过程,得到

$$B^{k+1} = B^k - \frac{B^k s^k (s^k)^T B^k}{(s^k)^T B^k s^k} + \frac{y^k (y^k)^T}{(y^k)^T s^k} \tag{52}$$

- (62)式称为对B^k的对称秩2修正公式,也称为BFGS公式相应的算法称为BFGS算法.

- 与秩1公式类似推导,只需取$u^i, v^i (i=1,2)$满足

$$\begin{cases} u^1 = y^k, \\ (v^1)^T s^k = 1 \end{cases} \begin{cases} u^2 = -B^k s^k \\ (v^2)^T s^k = 1 \end{cases} \tag{51}$$

- 则(60)式成立. 再由B^k的对称性和类似于秩1 公式的推导过程,得到

$$B^{k+1} = B^k - \frac{B^k s^k (s^k)^T B^k}{(s^k)^T B^k s^k} + \frac{y^k (y^k)^T}{(y^k)^T s^k} \tag{52}$$

- (62)式称为对B^k的对称秩2修正公式,也称为BFGS公式相应的算法称为BFGS算法.

定理3.4.1

设B^k是对称正定矩阵,若$(y^k)^T s^k > 0$, 通过秩2公式得到的B^{k+1}保持正定.

BFGS算法

1、 取初始点x_1,置初始矩阵$B^1(=I)$,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \leq \epsilon$, 则停止计算($x^k$作为无约束问题的最优解);否则求解线性方程组

$$B^k d = -\nabla f(x^k),$$

得到d^k.

3、 一维搜索.求解一维问题

$$\mathrm{m}in\phi(\alpha) = f(x^k + \alpha d^k),$$

得α_k.置

$$x^{k+1} = x^k + \alpha d^k$$

算法3.3.3(BFGS算法)

BFGS算法

1、 取初始点x_1,置初始矩阵$B^1(=I)$,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \le \epsilon$, 则停止计算($x^k$作为无约束问题的最优解);否则求解线性方程组

$$B^k d = -\nabla f(x^k),$$

得到d^k.

3、 一维搜索.求解一维问题

$$\mathrm{m}in\phi(\alpha) = f(x^k + \alpha d^k),$$

得α_k.置

$$x^{k+1} = x^k + \alpha d^k$$

BFGS算法

1、 取初始点x_1,置初始矩阵$B^1(=I)$,置精度要求ϵ,置k=1.

2、 如果$||\nabla f(x^k)|| \le \epsilon$, 则停止计算($x^k$作为无约束问题的最优解);否则求解线性方程组

$$B^k d = -\nabla f(x^k),$$

得到d^k.

3、 一维搜索.求解一维问题

$$\mathrm{min}\phi(\alpha) = f(x^k + \alpha d^k),$$

得α_k.置

$$x^{k+1} = x^k + \alpha d^k$$

BFGS算法

4、 修正B^k,记

$$s^k = x^{k+1} - x^k, y^k = -\alpha f(x^{k+1}) - \alpha f(x^k)$$

置

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

5、 置$k = k + 1$,转(2).

BFGS算法

4、 修正B^k,记

$$s^k = x^{k+1} - x^k, y^k = -\alpha f(x^{k+1}) - \alpha f(x^k)$$

置

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

5、 置$k = k + 1$,转(2).

例3.3.2: 用变度量法(BFGS算法)求解无约束问题

用变度量法(BFGS算法)求解无约束问题

$$\min f(x) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1,$$

取$x^1 = (0,0)^T$.

解

- $$\nabla f(x) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$
  $$\nabla^2 f(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = G$$

- 一维搜索步长为
  $$\alpha_k = \frac{(d^k)^T \nabla f(x^k)}{(d^k)^T G d^k} = -\frac{d_1 g_1 + d_2 g_2}{3d_1^2 + d_2^2 - 2d_1 d_2} \tag{53}$$

- 取$x^1 = (0,0)^T$, 则$\nabla f(x^1) = (-2,0)^T, B^1 = I, d^1 = -(B^1)^{-1}\nabla f(x^1) = -\nabla f(x^1) = (2,0)^T$. 由(44)得步长$\alpha_1 = \frac{2^2}{3 \cdot 2^2} = \frac{1}{3}$,

解

- $$\nabla f(x) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$
$$\nabla^2 f(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = G$$
- 一维搜索步长为
$$\alpha_k = \frac{(d^k)^T \nabla f(x^k)}{(d^k)^T G d^k} = -\frac{d_1 g_1 + d_2 g_2}{3d_1^2 + d_2^2 - 2d_1 d_2} \tag{53}$$
- 取$x^1 = (0,0)^T$, 则$\nabla f(x^1) = (-2,0)^T, B^1 = I, d^1 = -(B^1)^{-1}\nabla f(x^1) = -\nabla f(x^1) = (2,0)^T$. 由(44)得步长$\alpha_1 = \frac{2^2}{3\cdot 2^2} = \frac{1}{3}$,

解

- $$\nabla f(x) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$
$$\nabla^2 f(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = G$$
- 一维搜索步长为
$$\alpha_k = \frac{(d^k)^T \nabla f(x^k)}{(d^k)^T G d^k} = -\frac{d_1 g_1 + d_2 g_2}{3d_1^2 + d_2^2 - 2d_1 d_2} \tag{53}$$
- 取$x^1 = (0,0)^T$, 则$\nabla f(x^1) = (-2,0)^T, B^1 = I, d^1 = -(B^1)^{-1}\nabla f(x^1) = -\nabla f(x^1) = (2,0)^T$. 由(44)得步长$\alpha_1 = \frac{2^2}{3\cdot 2^2} = \frac{1}{3}$,

解

- 因此

$$x^2 = x^1 + \alpha_1 d^1 = (0,0)^T + \frac{1}{3}(2,0)^T = (\frac{2}{3},0)^T$$

- 再计算第二轮循环:因为$\nabla f(x^2) = (0, -\frac{2}{3})^T$, 有

$$\begin{aligned}
&s^1 = x^2 - x^1 = \left(\tfrac{2}{3}, 0\right)^T - (0,0)^T = \left(\tfrac{2}{3}, 0\right)^T, \\
&y^1 = \nabla f(x^2) - \nabla f(x^1) = \left(0, -\tfrac{2}{3}\right)^T - (-2,0)^T = \left(2, -\tfrac{2}{3}\right)^T, \\
&(s^1)^T B^1 s^1 = \left(\tfrac{2}{3}, 0\right) \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \tfrac{4}{9}, \\
&(y^1)^T s^1 = \left(2, -\tfrac{2}{3}\right) \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \tfrac{3}{4}, \\
&B^1 s^1 (s^1)^T B^1 = \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} \left(\tfrac{2}{3}, 0\right) = \begin{pmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{pmatrix},
\end{aligned}$$

解

- 因此

$$x^2 = x^1 + \alpha_1 d^1 = (0,0)^T + \frac{1}{3}(2,0)^T = (\frac{2}{3},0)^T$$

- 再计算第二轮循环:因为$\nabla f(x^2) = (0, -\frac{2}{3})^T$, 有

$$\begin{aligned}
&s^1 = x^2 - x^1 = \left(\tfrac{2}{3},0\right)^T - (0,0)^T = \left(\tfrac{2}{3},0\right)^T,\\
&y^1 = \nabla f(x^2) - \nabla f(x^1) = \left(0,-\tfrac{2}{3}\right)^T - (-2,0)^T = \left(2,-\tfrac{2}{3}\right)^T,\\
&(s^1)^T B^1 s^1 = \left(\tfrac{2}{3},0\right)\begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \tfrac{4}{9},\\
&(y^1)^T s^1 = \left(2,-\tfrac{2}{3}\right)\begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \tfrac{3}{4},\\
&B^1 s^1 (s^1)^T B^1 = \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix}\left(\tfrac{2}{3},0\right) = \begin{pmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{pmatrix},
\end{aligned}$$

解

- 因此

$$\begin{aligned} B^2 &= B^1 - \frac{B^1 s^1 (s^1)^T B^1}{(s^1)^T B^1 s^1} + \frac{y^{(1)} (y^{(1)})^T}{(y^{(1)})^T s^1} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{9}{4} \begin{pmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{pmatrix} + \frac{3}{4} \begin{pmatrix} 4 & -\frac{4}{3} \\ -\frac{4}{3} & \frac{4}{9} \end{pmatrix} \\ &= \begin{pmatrix} 3 & -1 \\ -1 & \frac{4}{3} \end{pmatrix} \end{aligned}$$

- 解方程组

$$B^2 d = -\nabla f(x_2)$$

解

- 因此

$$
\begin{aligned}
B^2 &= B^1 - \frac{B^1 s^1 (s^1)^T B^1}{(s^1)^T B^1 s^1} + \frac{y^{(1)} (y^{(1)})^T}{(y^{(1)})^T s^1} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{9}{4} \begin{pmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{pmatrix} + \frac{3}{4} \begin{pmatrix} 4 & -\frac{4}{3} \\ -\frac{4}{3} & \frac{4}{9} \end{pmatrix} \\
&= \begin{pmatrix} 3 & -1 \\ -1 & \frac{4}{3} \end{pmatrix}
\end{aligned}
$$

- 解方程组

$$B^2 d = -\nabla f(x_2)$$

解

- 即

$$\begin{pmatrix} 3 & -1 \\ -1 & \frac{4}{3} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{2}{3} \end{pmatrix}$$

得$d_1 = \frac{2}{9}, d_2 = \frac{2}{3}$, 即$d_2 = \left(\frac{2}{9}, \frac{2}{3}\right)$. 步长为

$$\alpha_2 = -\frac{\frac{2}{9} \cdot 0 + \frac{2}{3} \cdot (-\frac{2}{3})}{3\left(\frac{2}{9}\right)^2 + \left(\frac{2}{3}\right)^2 - 2 \cdot \frac{2}{9} \cdot \frac{2}{3}} = \frac{3}{2}$$

- 由此得到

$$x^3 = x^2 + \alpha_2 d^2 = \left(\frac{2}{3}, 0\right)^T + \frac{3}{2}\left(\frac{2}{9}, \frac{2}{3}\right)^T = (1, 1)^T.$$

计算梯度$\nabla f(x^3) = (0, 0)^T, ||f(x^3)|| = 0$. 所以$x^3 = (1, 1)^T$为最优解, 最优函数值为$f(x^3) = -1$.

解

- 即

$$\begin{pmatrix} 3 & -1 \\ -1 & \frac{4}{3} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{2}{3} \end{pmatrix}$$

得$d_1 = \frac{2}{9}, d_2 = \frac{2}{3}$, 即$d_2 = \left(\frac{2}{9}, \frac{2}{3}\right)$. 步长为

$$\alpha_2 = -\frac{\frac{2}{9} \cdot 0 + \frac{2}{3} \cdot (-\frac{2}{3})}{3\left(\frac{2}{9}\right)^2 + \left(\frac{2}{3}\right)^2 - 2 \cdot \frac{2}{9} \cdot \frac{2}{3}} = \frac{3}{2}$$

- 由此得到

$$x^3 = x^2 + \alpha_2 d^2 = \left(\frac{2}{3}, 0\right)^T + \frac{3}{2}\left(\frac{2}{9}, \frac{2}{3}\right)^T = (1,1)^T.$$

计算梯度$\nabla f(x^3) = (0,0)^T, ||f(x^3)|| = 0$. 所以$x^3 = (1,1)^T$为最优解, 最优函数值为$f(x^3) = -1$.