第三章 无约束非线性规划

自动化学院

§ 3.1 最优性条件

本章主要研究如下优化模型：

$$\min_{x\in R_n} \ f(x)$$

其中，$f: R^n \to R$，且 f 具有二阶连续偏导数。

全局最优解：对于多元函数$f(x)$，若点$x^*\in R^n$，$\forall\ x\in R^n$，均有$f(x^*)\le f(x)$，则x^*为全局最优解。

局部最优解：对于多元函数$f(x)$，若点x^*存在一邻域$\delta(x^*)$，使得$\forall x\in\delta(x^*)$，均有$f(x^*)\le f(x)$，则x^*为$f(x)$的局部极小点。

§ 3.1 最优性条件

多元函数的一阶导数--梯度

对于函数$f(x)$，$x \in R^n$，若在点x_0处自变量$x=(x_1, ..., x_n)^T$的各分量的偏导数$\partial f(x_0)/\partial x_i (i=1,\cdots,n)$都存在，则称$f(x)$在点$x_0$处一阶可导，并称向量

$$\nabla f(x_0) = \left(\frac{\partial f(x_0)}{\partial x_1}, \cdots, \frac{\partial f(x_0)}{\partial x_n} \right)^T$$

是$f(x)$在点x_0处的梯度或一阶导数。

§ 3.1 最优性条件

多元函数的二阶导数

对于函数$f(x)$，$x\in R^n$，若在点x_0处自变量$x=(x_1, ..., x_n)^T$的各分量的二阶偏导数$\partial^2 f(x_0)/\partial x_i \partial x_j (i, j=1,\cdots,n)$ 都存在，则称$f(x)$在点x_0处二阶可导，并称矩阵

$$\nabla^2 f(x_0)=\begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \cdots\cdots\cdots\cdots & & \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

是$f(x)$在点x_0处的二阶导数或黑塞矩阵。

§ 3.1 最优性条件

泰勒展开

对于一元函数$f(x)$，若在点x_0处n阶可导，其泰勒展开为

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \cdots \frac{1}{n!} f^{n}(x_0)(x - x_0)^n$$

对于多元函数$f(x)$，$x \in R^n$，若$f(x)$在点x_0具有二阶连续偏导数，则$f(x)$的二阶泰勒展开为

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

§ 3.1 最优性条件

定理1 (局部极小点的一阶必要条件)

设 $f(x)$在x^*处可微，若x^*为局部极小点，则必有 $\nabla f(x^*)=0$.

定理2 (局部极小点的二阶必要条件)

设 $f(x)$在x^*二阶连续可微，若x^*为局部极小点，则有

$\nabla f(x^*)=0$ 和 $\nabla^2 f(x^*)\geq 0.$

§ 3.1 最优性条件

定理3 (局部极小点的二阶充分条件)

设 $f(x)$在x^*二阶连续可微，若$\nabla f(x^*)=0$且$\nabla^2 f(x^*)>0$，则x^*是局部极小点。

定理4 (全局极小点的充要条件)

设 $f(x)$是定义在R^n上的可微凸函数，$x^*\in R^n$，则x^*为全局极小点的充要条件是$\nabla f(x^*)=0$.

§ 3.1 最优性条件

例1 求解下列问题的局部极小点

$$\min f(x) = (x_1^2 - 1)^2 + x_1^2 + x_2^2 - 2x_1$$

§ 3.1 最优性条件

例2 利用最优性条件求解下列问题：

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

先求 $f(x)$的一阶偏导数：

$$\frac{\partial f}{\partial x_1}=x_1^2-1,\quad \frac{\partial f}{\partial x_2}=x_2^2-2x_2$$

令 $\nabla f(x)=\begin{bmatrix} x_1^2-1 \\ x_2^2-2x_2 \end{bmatrix}=0$

解方程组得，$x^{(1)}=\begin{bmatrix} 1 \\ 0 \end{bmatrix},\quad x^{(2)}=\begin{bmatrix} 1 \\ 2 \end{bmatrix},\quad x^{(3)}=\begin{bmatrix} -1 \\ 0 \end{bmatrix},\quad x^{(4)}=\begin{bmatrix} -1 \\ 2 \end{bmatrix}$

§ 3.1 最优性条件

再求 $f(x)$ 的黑塞矩阵：

$$\nabla^2 f(x)=\begin{bmatrix}2x_1 & 0\\ 0 & 2x_2-2\end{bmatrix}$$

则有

$$\nabla^2 f(x^{(1)})=\begin{bmatrix}2 & 0\\ 0 & -2\end{bmatrix}\qquad \nabla^2 f(x^{(2)})=\begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix}$$

$$\nabla^2 f(x^{(3)})=\begin{bmatrix}-2 & 0\\ 0 & -2\end{bmatrix}\qquad \nabla^2 f(x^{(4)})=\begin{bmatrix}-2 & 0\\ 0 & 2\end{bmatrix}$$

因此，$x^{(1)}$, $x^{(4)}$ 不是极值点，$x^{(2)}$ 是局部极小点，$x^{(3)}$ 是局部极大点。

§ 3.2 一维搜索

一维搜索---解决单变量函数的极小化问题。

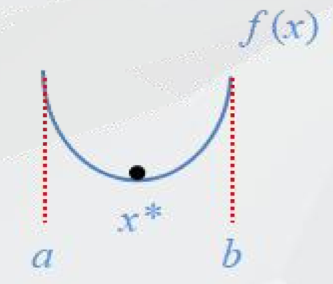
试探法(区间收缩法)：即按某种方式找试探点，通过一系列试探点来确定极小点。

函数逼近法：用某种较简单的曲线逼近/近似代替原来的函数曲线，并利用近似曲线的极小值来估计目标函数的极小值。

这两类方法一般只能求得极小点的近似值。

§ 3.2 一维搜索

(一) 平分法



对于一个区间$[a, b]$，如果$f'(a)<0$，$f'(b)>0$，则在$[a, b]$之间必有极小点x^*，且$f'(x^*)=0$.

平分法，即按照某一准则，逐步缩小区间长度，并使其始终含有极小点，直到区间足够小，则得到满足误差要求的解。

§ 3.2 一维搜索

取$x^0 = (a+b)/2$，若$f'(x^0)>0$，则在$[a, x^0]$中有极小点，此时，以$[a, x^0]$作为新的区间；

若$f'(x^0)<0$，则在$[x^0, b]$中有极小点，因此，以$[x^0, b]$作为新的区间。

继续这个过程，直至区间充分小或$f'(x^0)=0$.

对于初始区间的确定，可以先取一个初始点S_0，如果$f'(S_0)<0$，则令$S_1=S_0+\triangle S$，此时，若$f'(S_1)>0$，则区间$[a, b]=[S_0, S_1]$；否则，令$S_2=S_1+\triangle S$，直至$f'(S_i)>0$出现。

§ 3.2 一维搜索

例：利用平分法求解下列问题：

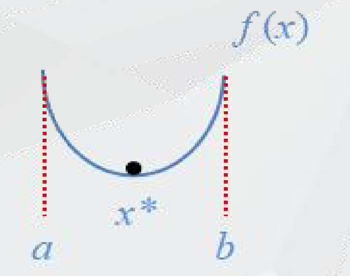
$$\min\ x^2+2x$$
$$-3\le x\le 5$$

令$f(x)=x^2+2x$，则$f'(x)=2x+2$. 对于区间$[a, b]=[-3, 5]$，取$x^0=(-3+5)/2=1$，有$f'(x^0)=4>0$，则令$a_1=-3$，$b_1=1$，取$x^1=(-3+1)/2=-1$，有$f'(x^1)=0$.

故$x^*=x^1=-1$为极小点。

§ 3.2 一维搜索

(二) 0.618法/黄金分割法



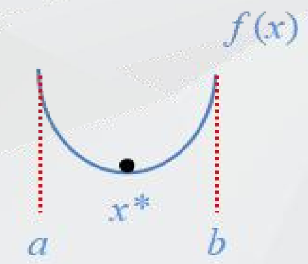
基本思想：通过选择两个试探点，使包含极小点的区间不断缩短，当区间长度小到一定程度时，区间上各点的函数值均接近极小值，因此，任一点都可作为极小点的近似。

对于函数$f(x)$，若存在$x^* \in [a, b]$，使$f(x)$在$[a, x^*]$上严格递减，在$[x^*, \mathrm{b}]$上严格递增，则称$f(x)$为$[a, b]$上的单峰函数。

在区间内选择两个点，令$x_1, x_2 \in [a, b]$，且$x_1 < x_2$，使x^*包含在区间$[x_1, x_2]$内，则有：若$f(x_1) > f(x_2)$，则$x^* \in [x_1, b]$;

若$f(x_1) \le f(x_2)$，则$x^* \in [a, x_2]$.

§ 3.2 一维搜索



在k次迭代后，有$x^* \in [a_k, b_k]$，取$\lambda_k, \mu_k \in [a_k, b_k]$，且$\lambda_k < \mu_k$

若$f(\lambda_k) \le f(\mu_k)$，则令$a_{k+1} = a_k, b_{k+1} = \mu_k$；若$f(\lambda_k) > f(\mu_k)$，则令$a_{k+1} = \lambda_k, b_{k+1} = b_k$，如此下去。

如何选取λ_k和μ_k呢？

λ_k和μ_k取在$[a_k, b_k]$中的对称位置；

$$b_k - \lambda_k = \mu_k - a_k$$

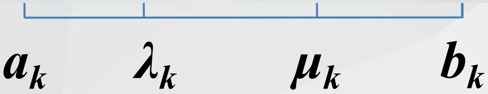
每次迭代区间长度缩短比例相同。

$$b_{k+1} - a_{k+1} = \alpha(b_k - a_k)$$

§ 3.2 一维搜索

$b_k - \lambda_k = \mu_k - a_k$

$b_{k+1} - a_{k+1} = \alpha(b_k - a_k)$



不妨设$f(\lambda_k) \le f(\mu_k)$，$a_{k+1} = a_k, b_{k+1} = \mu_k$，

整理可得，$\mu_k = a_k + \alpha(b_k - a_k)$。

若$f(\lambda_k) > f(\mu_k)$，$a_{k+1} = \lambda_k, b_{k+1} = b_k$，则$\lambda_k = a_k + (1-\alpha)(b_k - a_k)$。

假设在第k次迭代，$f(\lambda_k) \le f(\mu_k)$，则$[a_{k+1}, b_{k+1}]=[a_k, \mu_k]$

在第k+1次迭代，$\mu_{k+1} = a_{k+1} + \alpha(b_{k+1} - a_{k+1}) = a_k + \alpha(\mu_k - a_k)$

$$= a_k + \alpha^2(b_k - a_k)$$

若令$\alpha^2 = 1-\alpha$，则 $\mu_{k+1} = \lambda_k$，因此μ_{k+1}不必重新计算，只需求λ_{k+1}即可。

解得，$\alpha = \dfrac{-1+\sqrt{5}}{2} \approx 0.618$

§ 3.2 一维搜索

0.618法在区间内取点的规则

$$\begin{cases} \mu_k = a_k + 0.618(b_k - a_k) \\ \lambda_k = a_k + 0.382(b_k - a_k) \end{cases}$$

对比平分法与0.618法：

1. 平分法需要求导，因此要求函数可微。

2. 0.618法不需要函数可微，但只适用于单峰函数。

§ 3.2 一维搜索

计算步骤：

1. 设置初始区间$[a_1, b_1]$，精度要求$\varepsilon > 0$，计算

$$\lambda_1 = a_1 + 0.382(b_1 - a_1),\ \mu_1 = a_1 + 0.618(b_1 - a_1)$$

$f(\lambda_1)$和 $f(\mu_1)$，令$k = 1$.

2. 若$b_k - a_k < \varepsilon$，则停止计算；否则，当 $f(\lambda_k) > f(\mu_k)$时，转3；当 $f(\lambda_k) \le f(\mu_k)$时，转4.

3. 令$a_{k+1} = \lambda_k$, $b_{k+1} = b_k$, $\lambda_{k+1} = \mu_k$, $\mu_{k+1} = a_{k+1} + 0.618(b_{k+1} - a_{k+1})$，计算$f(\mu_{k+1})$，转5.

4. 令$a_{k+1} = a_k$, $b_{k+1} = \mu_k$, $\mu_{k+1} = \lambda_k$, $\lambda_{k+1} = a_{k+1} + 0.382(b_{k+1} - a_{k+1})$，计算$f(\lambda_{k+1})$，转5.

5. 令$k = k+1$，返回2.

§ 3.2 一维搜索

$$\begin{cases} \mu_k = a_k + 0.618(b_k - a_k) \\ \lambda_k = a_k + 0.382(b_k - a_k) \end{cases}$$

例：利用0.618法求解下列问题：

$$\min e^x - 5x$$

$$1 \le x \le 2,\ \ \varepsilon = 0.04.$$

k	a_k	b_k	λ_k	μ_k	$f(\lambda_k)$	$f(\mu_k)$	$b_k - a_k$
1	1	2	1.382	1.618	−2.928	−3.048	1
2	1.382	2	1.618	1.764	−3.048	−2.985	0.618
3	1.382	1.764	1.528	1.618	−3.032	−3.048	0.382
4	1.528	1.764	1.618	1.674	−3.048	−3.037	0.236
5	1.528	1.674	1.584	1.618	−3.046	−3.048	0.146
6	1.584	1.674	1.618	1.640	−3.048	−3.046	0.09
7	1.584	1.640	1.605	1.618	−3.048	−3.048	0.056
8	1.584	1.618					0.036

$$x^* = \frac{1.584 + 1.618}{2} = 1.601$$

§ 3.2 一维搜索

(三) 函数逼近法

基本思想：在迭代点附近用二阶泰勒展开多项式近似目标函数$f(x)$，进而求得极小点的估计值。

设$f(x)$二次可微，在$f(x)$在x^k处的二阶泰勒展开式为：

$$f(x) \approx f(x^k) + f'(x^k)(x - x^k) + \frac{1}{2} f''(x^k)(x - x^k)^2$$

令

$$f'(x) = f'(x^k) + f''(x^k)(x - x^k) = 0$$

可得，

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)}$$

若$|f'(x^k)| < \varepsilon$, 则停止迭代，$x^* = x^k$为极小点。

§ 3.2 一维搜索

计算步骤：

1. 给定初始点x^1，精度要求$\varepsilon > 0$，令$k=1$.

2. 若$|f'(x^k)| < \varepsilon$，则停止计算，得到极小点x^k；否则，转3.

3. 计算点x^{k+1}，

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)}$$

令$k = k+1$，返回2.

局限性：对初始点的选择比较敏感。如果初始点靠近极小点，则可能很快收敛；如果初始点远离极小点，迭代产生的点列可能不收敛于极小点。

§ 3.2 一维搜索

$$x^{k+1}=x^k-\frac{f'(x^k)}{f''(x^k)}$$

例：用函数逼近法求解下列问题：

$$\min\ f(x)=e^x-5x$$

取$x^1=2$，$\varepsilon=0.01$

$f'(x)=e^x-5$

$f''(x)=e^x$

k	$x^{(k)}$	$f'\left(x^{(k)}\right)$	$f''\left(x^{(k)}\right)$
1	2	2.388	7.388
2	1.677	0.349	5.349
3	1.612	0.012	5.012
4	1.6096	$\boxed{-0.00002}$	

则$x^*=x^4=1.6096$为极小点。

课堂练习

1. 利用0.618法和平分法求解下列问题：

$$\min\ f(x) = 2x^2 - x - 1$$
$$-1 \le x \le 1,\ \ \varepsilon = 0.16.$$

k	a_k	b_k	λ_k	μ_k	$f(\lambda_k)$	$f(\mu_k)$	$b_k - a_k$
1	-1	1	-0.236	0.236	-0.653	-1.125	2
2	-0.236	1	0.236	0.528	-1.125	-0.970	1.236
3	-0.236	0.528	0.056	0.236	-1.050	-1.125	0.764
4	0.056	0.528	0.236	0.348	-1.125	-1.106	0.472
5	0.056	0.348	0.168	0.236	-1.112	-1.125	0.292
6	0.168	0.348	0.236	0.279	-1.125	-1.123	0.18
7	0.168	0.279					$\boxed{0.111}$

$$x^* = \frac{0.168 + 0.279}{2} = 0.2235$$

§ 3.3 最速下降法

迭代法：

　　给定极小值点的初始估计x^0，利用某种方法找出一系列的点列$\{x^k\}$，使得$\lim\limits_{k\to\infty}\left\|x^k-x^*\right\|=0.$

怎样产生这个点列呢？

在第k次迭代，从x^k出发，寻找x^{k+1}，而一个向量是由其长度和方向所确定的，即总可以写成

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

其中，d^k为方向，λ_k为步长。

各种迭代方法的区别在于确定方向和步长方式的不同，且总是要求所构造出的点列所对应的函数值是单调递减的。

§ 3.3 最速下降法

在x^k和d^k均已知的情况下，$f(x^k+\lambda d^k)$为λ的一元函数。

非精确搜索：

存在λ_k，使得

$$f\left(x^{(k)}+\lambda_k d^{(k)}\right)<f\left(x^{(k)}\right)$$

精确搜索：

求解以λ为变量的一元函数$f(x^k+\lambda d^k)$的极小点，使得

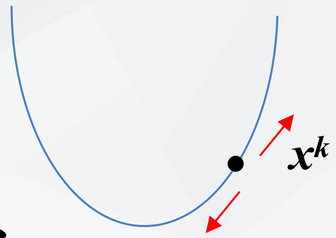
$$f\left(x^{(k)}+\lambda_k d^{(k)}\right)=\min f\left(x^{(k)}+\lambda d^{(k)}\right)$$

此时，λ_k为最优步长。

§ 3.3 最速下降法

搜索方向(下降方向)：

对于充分小的$\lambda > 0$, 有$f(x+\lambda d) < f(x)$, 则d为f在x处的下降方向。



如右图所示，在x^k处有两个运动方向，

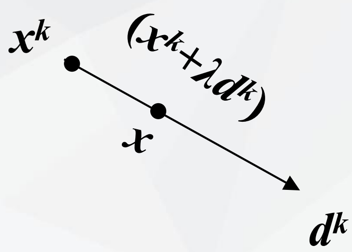
为了求极小值，运动方向应为函数下降的方向。

结论：函数$f(x)$在点x^k的负梯度方向是x^k处函数值下降最快的方向，即 $d^k = -\nabla f(x^k).$

§ 3.3 最速下降法

搜索方向(下降方向):

任取模为1的方向d^k及$\lambda > 0$, 有$f(x+\lambda d) < f(x)$, 则$f(x)$在x^k处的一阶泰勒展开为



$$f(x) = f(x^k + \lambda d^k) = f(x^k) + \lambda \nabla f^T(x^k) d^k + O(\lambda)$$

$$f(x^k + \lambda d^k) - f(x^k) = \lambda \nabla f^T(x^k) d^k + O(\lambda)$$

$$= \lambda\left(\nabla f^T(x^k) d^k + \underbrace{\frac{O(\lambda)}{\lambda}}_{\substack{\downarrow\ \lambda\text{充分小} \\ 0}}\right)$$

$$\nabla f^T(x^k) d^k = \| \nabla f^T(x^k) \| \| d^k \| \cos\theta$$

取$\cos\theta = -1$, 则$f(x^k+\lambda d^k) - f(x^k)$最小, 即$\theta = 180^o$, $d^k = -\nabla f(x^k)$.

§ 3.3 最速下降法

算法步骤：

1. 选定初始点x^0，给定精度要求$\varepsilon > 0$，令$k = 0$.

2. 计算搜索方向$d^k = -\nabla f(x^k)$.

3. 若$\|d^k\| < \varepsilon$，则停止迭代，x^k为极小点的近似值；否则从x^k出发，沿d^k方向，求λ_k，使$f(x^k + \lambda_k d^k) = \min f(x^k + \lambda d^k)$.

4. 令$x^{k+1} = x^k + \lambda_k d^k$, 且$k = k+1$，返回2.

§ 3.3 最速下降法

相邻两个搜索方向是正交的

利用精确搜索求 $\min \varphi(\lambda) = f(x^k+\lambda d^k)$

令$\varphi'(\lambda)=0$，有 $\nabla f^T(x^k+\lambda_k d^k)d^k=0$

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

$$(d^{k+1})^T d^k=0$$

因此，相邻两次迭代搜索方向垂直/正交。

最速下降法可以较快地从初始点到达极小点附近，但在接近极小点时，收敛速度变慢。

§ 3.3 最速下降法

例： 求 $\min f(x)=(x_1-1)^2+(x_2-1)^2$

取 $x^0=(0,0)^T$，$\varepsilon=0.1.$

解： $\nabla f(x)=\big(2(x_1-1),2(x_2-1)\big)^T$，　　$d^k=-\nabla f(x^k)$

$k=0$时，$d^0=-(-2,-2)^T=(2,2)^T$，$\left\|d^0\right\|=2\sqrt{2}>\varepsilon$

从x^0出发，沿d^0方向进行搜索，求步长λ_1

因 $x^0+\lambda d^0=(2\lambda,2\lambda)^T$　　$f(x^0+\lambda d^0)=(2\lambda-1)^2+(2\lambda-1)^2$

$$\frac{d\,f(x^0+\lambda d^0)}{d\lambda}=4(2\lambda-1)+4(2\lambda-1)=0$$

可得$\lambda_0=1/2$

令 $x^1=x^0+\lambda_0 d^0=(1,1)^T.$

$k=1$时，$d^1=(0,0)^T$，$\left\|d^1\right\|=0<\varepsilon$

故 $x^*=x^1=(1,1)^T$为极小点。

§ 3.4 牛顿法和修正牛顿法

牛顿法---一维搜索中函数逼近法的推广

设$f(x)$二次可微，则$f(x)$在点x^k处的二阶泰勒展开为：

$$f(x) \approx \varphi(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k)$$

令 $\nabla \varphi(x) = \nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$

当$\nabla^2 f(x^k)$ **正定，即**$\nabla^2 f(x^k)^{-1}$ **存在，则有**

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k)$$

令搜索方向为 $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$**---牛顿方向，**

此时，步长$\lambda_k = 1$.

§ 3.4 牛顿法和修正牛顿法

$$\varphi(x)=f(x^k)+\nabla f(x^k)^T(x-x^k)+\frac{1}{2}(x-x^k)^T\nabla^2 f(x^k)(x-x^k)$$

$$\downarrow$$

$$\Gamma(X)=c+b^T X+\frac{1}{2}X^T AX$$

有 $\nabla\Gamma(X)=AX+b, \quad \nabla^2\Gamma(X)=A$

根据牛顿法的迭代方程，有

$$\begin{aligned}X^1&=X^0-\nabla^2\Gamma(X^0)^{-1}\nabla\Gamma(X^0)\\&=X^0-A^{-1}(AX^0+b)\\&=-A^{-1}b \quad \longleftarrow \quad \text{二次型的最优解}\end{aligned}$$

说明，牛顿法具有二次终止性。

§ 3.4 牛顿法和修正牛顿法

牛顿法的计算步骤：

1. 给定初始点x^0，精度要求$\varepsilon > 0$，令$k=0$.

2. 计算$\nabla f(x^k)$**和** $\nabla^2 f(x^k)^{-1}$

3. 若$\|\nabla f(x^k)\| < \varepsilon$**，则停止迭代；否则令** $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$

4. 计算$x^{k+1} = x^k + d^k$, 且$k = k+1$，返回2.

对初始点的选择比较敏感，当初始点远离最优解时，算法可能不收敛。

§ 3.4 牛顿法和修正牛顿法

修正牛顿法：

迭代方向--- $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$

迭代步长--- 一维搜索求λ_k

算法步骤：

1. 给定初始点x^0，精度要求$\varepsilon > 0$，令$k = 0$.

2. 计算$\nabla f(x^k)$和 $\nabla^2 f(x^k)^{-1}$

3. 若$\|\nabla f(x^k)\| < \varepsilon$，则停止迭代；否则令 $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$

4. 从x^k出发，沿方向d^k作一维搜索，计算λ_k，使

$$\lambda_k = \arg\min f(x^k + \lambda d^k)$$

5. 令$x^{k+1} = x^k + \lambda_k d^k$，且$k = k+1$，返回2.

§ 3.4 牛顿法和修正牛顿法

修正牛顿法：

迭代方向--- $d^k = -\nabla^2 f(x^k)^{-1}\nabla f(x^k)$

迭代步长--- 一维搜索求λ_k

优点：二阶收敛，收敛速度快；

迭代点越靠近最优解，收敛速度越快。

局限性：若迭代中目标函数的黑塞矩阵不可逆，则算法无法执行；

黑塞矩阵及其逆的计算量较大。

§ 3.4 牛顿法和修正牛顿法

例1 用牛顿法求解 $\min f(x) = x_1^2 + 25x_2^2$

取 $x^0 = (2,2)^T$，$\varepsilon = 10^{-6}$.

例2 用修正牛顿法求解 $\min f(x) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$

取 $x^0 = (0,0)^T$，$\varepsilon = 10^{-6}$.

§ 3.5 拟牛顿法

在牛顿法中，迭代过程在确定d^k时，需要计算$\nabla^2 f(x^k)^{-1}$，计算量较大，且$\nabla^2 f(x^k)$有时也未必可逆，因此，希望用矩阵H_k来替代$\nabla^2 f(x^k)^{-1}$，则在迭代过程中搜索方向变为

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

根据Taylor展开，有

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

求$f(x)$的梯度，有 $\nabla f(x) \approx \nabla f(x^k) + \nabla^2 f(x^k)(x - x^k)$

$$\Big\downarrow \quad \text{令} x = x^{k-1}$$

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

整理得，$$x^k - x^{k-1} \approx \nabla^2 f(x^k)^{-1}(\nabla f(x^k) - \nabla f(x^{k-1}))$$

§ 3.5 拟牛顿法

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

令 $\gamma_k = \nabla f(x^k) - \nabla f(x^{k-1}),\ \ \delta_k = x^k - x^{k-1}$

则有 $H_k \gamma_k = \delta_k$ $\longrightarrow$ 拟牛顿方程

因 $\nabla^2 f(x^k)$ 对称，所以$\boldsymbol{H_k}$也应是对称的。

如何确定$\boldsymbol{H_{k+1}}$？

令$\boldsymbol{H_{k+1} = H_k + \boxed{E_{k+1}}}$ $\longleftarrow$ 对称

满足上式中的$\boldsymbol{E_{k+1}}$有很多，因此拟牛顿算法是一族算法。

变尺度(DFP)算法最为经典/常用。

§ 3.5 拟牛顿法

$H_k\gamma_k=\delta_k$

变尺度(DFP)算法

令 $E_{k+1}=\alpha_{k+1}U_{k+1}U_{k+1}^T+\beta_{k+1}V_{k+1}V_{k+1}^T$， 显然$\boldsymbol{E_{k+1}}$是对称的。

则有 $H_{k+1}=H_k+E_{k+1}=H_k+\alpha_{k+1}U_{k+1}U_{k+1}^T+\beta_{k+1}V_{k+1}V_{k+1}^T$

$\downarrow$ 两边乘γ_{k+1}

$$H_{k+1}\gamma_{k+1}=H_k\gamma_{k+1}+\alpha_{k+1}U_{k+1}\boxed{U_{k+1}^T\gamma_{k+1}}+\beta_{k+1}V_{k+1}\boxed{V_{k+1}^T\gamma_{k+1}}$$

$$\underbrace{H_{k+1}\gamma_{k+1}}_{\delta_{k+1}}=H_k\gamma_{k+1}+\alpha_{k+1}(U_{k+1}^T\gamma_{k+1})U_{k+1}+\beta_{k+1}(V_{k+1}^T\gamma_{k+1})V_{k+1}$$

为了满足上式，令

$$\alpha_{k+1}(U_{k+1}^T\gamma_{k+1})=1 \qquad U_{k+1}=\delta_{k+1}$$

$$\beta_{k+1}(V_{k+1}^T\gamma_{k+1})=-1 \qquad V_{k+1}=H_k\gamma_{k+1}$$

§ 3.5 拟牛顿法

整理得

$$\alpha_{k+1}=\frac{1}{U_{k+1}^{T}\gamma_{k+1}}=\frac{1}{\delta_{k+1}^{T}\gamma_{k+1}}$$

$$\beta_{k+1}=-\frac{1}{V_{k+1}^{T}\gamma_{k+1}}=-\frac{1}{\gamma_{k+1}^{T}H_{k}\gamma_{k+1}}$$

则有

$$H_{k+1}=H_{k}+E_{k+1}=H_{k}+\frac{\delta_{k+1}\delta_{k+1}^{T}}{\delta_{k+1}^{T}\gamma_{k+1}}-\frac{H_{k}\gamma_{k+1}\gamma_{k+1}^{T}H_{k}}{\gamma_{k+1}^{T}H_{k}\gamma_{k+1}}$$

$$\gamma_{k}=\nabla f(x^{k})-\nabla f(x^{k-1}),\quad \delta_{k}=x^{k}-x^{k-1}$$

§ 3.5 拟牛顿法

DFP算法的计算步骤：

1. 给定初始点x^0.

2. 令H_0=I，k =0.

3. 若 $\nabla f(x^k)=0$ **，则停止迭代；否则令** $d^k=-H_k\nabla f(x^k)$

4. 从x^k出发，沿方向d^k作一维搜索，计算λ_k，令$x^{k+1}=x^k+\lambda_k d^k$

5. 计算 $\delta_{k+1}=x^{k+1}-x^k$，$\gamma_{k+1}=\nabla f(x^{k+1})-\nabla f(x^k)$

$$H_{k+1}=H_k+\frac{\delta_{k+1}\delta_{k+1}^T}{\delta_{k+1}^T\gamma_{k+1}}-\frac{H_k\gamma_{k+1}\gamma_{k+1}^T H_k}{\gamma_{k+1}^T H_k\gamma_{k+1}}$$

且$k=k$+1，返回3.

§ 3.5 拟牛顿法

例：在DFP方法求解问题的过程中，已知

$$H_k = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \qquad \gamma_{k+1} = (1,\ 1)^T, \qquad \delta_{k+1} = (1,\ 2)^T$$

求矩阵$\boldsymbol{H_{k+1}}$.

$$H_{k+1} = H_k + \frac{\delta_{k+1}\delta_{k+1}^T}{\delta_{k+1}^T\gamma_{k+1}} - \frac{H_k\gamma_{k+1}\gamma_{k+1}^T H_k}{\gamma_{k+1}^T H_k \gamma_{k+1}}$$

$$H_{k+1} = \begin{pmatrix} \dfrac{2}{3} & \dfrac{1}{3} \\ \dfrac{1}{3} & \dfrac{5}{3} \end{pmatrix}$$

§ 3.5 拟牛顿法

例：用拟牛顿法求 $\min 2x_1^2 + x_2^2 - 4x_1 + 2$

取 $x^{(0)} = (2,1)^T, H_0 = I$

解：$\nabla f(x) = \left(4x_1 - 4, 2x_2\right)^T$，　$\nabla f(x^0) = \left(4,2\right)^T$

因 $\nabla f(x^0) \neq 0$，令 $d^0 = -H_0 \nabla f(x^0) = \left(-4,-2\right)^T$

有 $x^0 + \lambda d^0 = (2-4\lambda, 1-2\lambda)^T$

$$f(x^0 + \lambda d^0) = 36\lambda^2 - 20\lambda + 3$$

$$\frac{d\, f(x^0 + \lambda d^0)}{d\lambda} = 72\lambda - 20 = 0 \longrightarrow \boldsymbol{\lambda_0 = 5/18}$$

令$x^1 = x^0 + \lambda_0 d^0 = \left(\dfrac{8}{9}, \dfrac{4}{9}\right)^T$，有 $\nabla f(x^1) = \left(-\dfrac{4}{9}, \dfrac{8}{9}\right)^T$

§ 3.5 拟牛顿法

$$\delta_{k+1}=x^{k+1}-x^{k},\quad \gamma_{k+1}=\nabla f(x^{k+1})-\nabla f(x^{k})$$

$$H_{k+1}=H_k+\frac{\delta_{k+1}\delta_{k+1}^T}{\delta_{k+1}^T\gamma_{k+1}}-\frac{H_k\gamma_{k+1}\gamma_{k+1}^T H_k}{\gamma_{k+1}^T H_k\gamma_{k+1}}$$

计算 δ_1，γ_1 **，代入** $\boldsymbol{H_{k+1}}$ **的公式，得有**

$$H_1=\frac{1}{360}\begin{pmatrix}86 & -38\\ -38 & 305\end{pmatrix}$$

此时， $d^1=-H_1\nabla f(x^1)=\frac{12}{51}(1,-4)^T$ **，且** $\nabla f(x^1)\neq 0$

根据 $\frac{d\,f(x^1+\lambda d^1)}{d\,\lambda}=0 \longrightarrow$ $\boldsymbol{\lambda_1=17/36}$

令 $x^2=x^1+\lambda_1 d^1=(1,0)^T$，$\quad \nabla f(x^2)=(0,0)^T$.

故 $\boldsymbol{x^*=x^2=(1,0)^T}$ **为极小点。**

§ 3.6 共轭梯度法

定义1 (共轭方向)

设x, y是n维欧氏空间中两个向量，即$x, y \in R^n$, 若有$x^T y = 0$, 则称x与y是两个正交的向量。又设A是一个n阶对称正定矩阵，若有$x^T A y = 0$, 则称向量x与y关于A共轭正交，简称关于A共轭。

定义2

设一组非零向量$p_1, ..., p_n \in R^n$, A为n阶对称正定阵，若有下式成立：

$$p_i^T A p_j = 0 \quad (i \neq j,\ i, j = 1, \cdots, n)$$

则称向量组$p_1, ..., p_n$关于A共轭，也称为A的n个共轭方向。

特别地，当A为单位阵时，$p_i^T A p_j = p_i^T p_j = 0$, p_i与p_j正交。

§ 3.6 共轭梯度法

性质1

设A是n阶对称正定阵，$p_1, \ldots, p_k$是 k 个A共轭的非零 n 维向量，则向量组 $p_1, \ldots, p_k$ 必线性无关($k \le n$)。

性质2

设 $p_1, \ldots, p_k$ 是 k 个相互共轭的非零向量，则有如下结论成立：

$\nabla f(x^{k+1})$与 $p_1, \ldots, p_k$ 的任意线性组合正交，$1 \le k < n$.

§ 3.6 共轭梯度法

证明：向量$(1, 0)^T$和$(3, -2)^T$关于矩阵

$$A=\begin{bmatrix}2 & 3\\3 & 5\end{bmatrix}$$

共轭。

§ 3.6 共轭梯度法

对于二次函数

$$f(x)=\frac{1}{2}x^T Ax+b^T x+c$$

若***f*(*x*)**可写成变量分离的形式，即 $f(x)=f_1(x_1)+f_2(x_2)+\cdots+f_n(x_n)$

则从任意一点出发，只需分别沿每个坐标轴方向进行一维搜索，就可得到***f*(*x*)**的最优解。

因此，对于二次函数***f*(*x*)**，只要我们能够找到一组基$\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\}$，使$\boldsymbol{p}_i$满足条件$p_i^T Ap_j=0\ (i\neq j)$，则在这组基下，从任一点出发进行搜索，可得最优解。

§ 3.6 共轭梯度法

如何构造两两A共轭的方向？

考虑在点x^k处，利用负梯度方向$-\nabla f(x^k)$和前一个搜索方向p_{k-1}的组合构成p_k，使p_k与前k-1个搜索方向 $p_1, \ldots, p_{k-1}$两两共轭。

对于

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

有 $\nabla f(x) = Ax + b$

对于 $\forall x, y \in R^n$，有 $\nabla f(y) - \nabla f(x) = Q(y - x)$

从初始点为x^1和$p_1 = -\nabla f(x^1)$出发，可得(x^2, p_2) …, (x^k, p_k), 且有$p_1, \ldots, p_k$ 相互共轭和 $x^{k+1} = x^k + \lambda_k p_k$.

§ 3.6 共轭梯度法

$$\frac{d\,f(x^k+\lambda p_k)}{d\lambda}=0$$ 求λ_k

即
$$\frac{d\,f(x^k+\lambda p_k)}{d\lambda}=p_k^T\nabla f(x^{k+1})=0$$

$$\Downarrow \qquad \nabla f(x)=Ax+b$$

$$p_k^T[Ax^{k+1}+b]=0$$

$$\Downarrow \qquad \boldsymbol{x^{k+1}=x^k+\lambda_k p_k}$$

$$p_k^T[A(x^k+\lambda_k p_k)+b]=0$$

$$\Downarrow$$

$$p_k^T[\nabla f(x^k)+\lambda_k A p_k]=0$$

$$\Downarrow$$

$$\lambda_k=-\frac{p_k^T\nabla f(x^k)}{p_k^T A p_k}$$

§ 3.6 共轭梯度法

对于$\forall\, j \le k$, 有

$$\frac{d\,f(x^j+\lambda p_j)}{d\lambda} = p_j^T \nabla f(x^{j+1}) = 0$$

和

$$p_j^T \nabla f(x^{k+1}) = 0$$

又对于$\forall\, j < k$，根据 $\nabla f(y)-\nabla f(x) = A(y-x)$ 有

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

$$\Big\downarrow \text{两边乘 } p_j^T$$

$$0 = p_j^T \nabla f(x^{k+1}) = p_j^T \nabla f(x^{j+1}) + p_j^T A[(x^{k+1}-x^k)+(x^k-x^{k-1})+\cdots+(x^{j+2}-x^{j+1})]$$

$$= \quad 0 \quad + \quad p_j^T A\left(\sum_{i=j+1}^{k}\lambda_i p_i\right) = \sum_{i=j+1}^{k}\lambda_i (p_j^T A p_i) = 0$$

§ 3.6 共轭梯度法

另外，令 $p_{k+1} = -\nabla f(x^{k+1}) + \alpha_k p_k$

$\downarrow$ 两边乘 $p_k^T A$

$$p_k^T A p_{k+1} = p_k^T A(-\nabla f(x^{k+1})) + p_k^T A(\alpha_k p_k) = 0$$

$\downarrow$

$$\alpha_k = \frac{p_k^T A \nabla f(x^{k+1})}{p_k^T A p_k}$$

不难证明，对于 $j \le k$-1，也有

$$p_j^T A p_{k+1} = 0$$

§ 3.6 共轭梯度法

由 $p_j^T \nabla f(x^{k+1}) = 0$，有

$$\nabla^T f(x^{k+1}) p_j = 0$$

$$\Downarrow \quad p_j = -\nabla f(x^j) + \alpha_{j-1} p_{j-1}$$

$$\nabla^T f(x^{k+1})[-\nabla f(x^j) + \alpha_{j-1} p_{j-1}] = 0$$

$$\Downarrow$$

$$-\nabla^T f(x^{k+1}) \nabla f(x^j) + \alpha_{j-1} \nabla^T f(x^{k+1}) p_{j-1} = 0$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$0 \qquad\qquad 0$$

则有 $\nabla f^T (x^{k+1}) \nabla f(x^j) = 0, \quad j \le k$

§ 3.6 共轭梯度法

$$f(x)=\frac{1}{2}x^T Ax+b^T x+c$$

求解步骤：

1. 任选初始点x^1, $p_1=-\nabla f(x^1)$ **，k =1.**

2. 若 $\nabla f(x^k)=0$ **, 则停止迭代；否则**

$$x^{k+1}=x^k+\lambda_k p^k \qquad \lambda_k=-\frac{p_k^T\nabla f(x^k)}{p_k^T Ap_k}$$

$$p_{k+1}=-\nabla f(x^{k+1})+\alpha_k p_k \qquad \alpha_k=\frac{p_k^T A\nabla f(x^{k+1})}{p_k^T Ap_k}$$

3. k = k+1，返回2.

§ 3.6 共轭梯度法

对于一般二次可微函数，在每一点的局部，可进行二阶泰勒展开，因此，也可利用共轭梯度法进行求解。

这时，需要修改公式，使之不含A，则有

$$\alpha_k = \frac{p_k^T A \nabla f(x^{k+1})}{p_k^T A p_k} = \frac{(A\lambda_k p_k)^T \nabla f(x^{k+1})}{(A\lambda_k p_k)^T p_k}$$

$\cdots\ \boldsymbol{x^{k+1} = x^k + \lambda_k p_k}$

$$= \frac{[A(x^{k+1} - x^k)]^T \nabla f(x^{k+1})}{[A(x^{k+1} - x^k)]^T p_k}$$

$\nabla f(y) - \nabla f(x) = A(y - x)$

$$= \frac{[\nabla f(x^{k+1}) - \nabla f(x^k)]^T \nabla f(x^{k+1})}{[\nabla f(x^{k+1}) - \nabla f(x^k)]^T p_k}$$

$\nabla f^T(x^{k+1}) \nabla f(x^k) = 0$

$p_{k+1} = -\nabla f(x^{k+1}) + \alpha_k p_k$

$$= \frac{\left\|\nabla f(x^{k+1})\right\|^2}{\left\|\nabla f(x^k)\right\|^2}$$

$p_j^T \nabla f(x^{k+1}) = 0$

§ 3.6 共轭梯度法

求解步骤：

1. 任选初始点x^1, $d^1=-\nabla f(x^1)$ **，$k=1$.**

2. 若 $\nabla f(x^k)=0$ **, 则停止迭代；否则**

$$\lambda_k=\arg\min f(x^k+\lambda d^k)$$

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

$$d^{k+1}=-\nabla f(x^{k+1})+\frac{\left\|\nabla f(x^{k+1})\right\|^2}{\left\|\nabla f(x^k)\right\|^2}d^k$$

3. $k=k+1$，返回2.

§ 3.6 共轭梯度法

无约束最优化方法	搜索方向	步长	停止准则	特性
最速下降法	$d^k = -\nabla f(x^k)$	精确搜索求λ_k	$\|\nabla f(x^k)\| < \varepsilon$	相邻两次迭代搜索方向正交
牛顿法	$d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$	λ_k=1	$\|\nabla f(x^k)\| < \varepsilon$	黑塞矩阵可逆，计算量大
修正牛顿法		精确搜索求λ_k		
拟牛顿法 (变尺度法)	$d^k = -H_k \nabla f(x^k)$ $H_{k+1} = H_k + \frac{\delta_{k+1}\delta_{k+1}^T}{\delta_{k+1}^T \gamma_{k+1}} - \frac{H_k \gamma_{k+1} \gamma_{k+1}^T H_k}{\gamma_{k+1}^T H_k \gamma_{k+1}}$	精确搜索求λ_k	$\nabla f(x^k) = 0$	只需一阶导数，不需求逆
共轭梯度法	$d^{k+1} = -\nabla f(x^{k+1}) + \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2} d^k$	精确搜索求λ_k	$\nabla f(x^k) = 0$	只需一阶导数

§ 3.6 共轭梯度法

例： 求 $\min x_1^2 + 2x_2^2 - 2x_1x_2 + 2x_2 + 2$

取 $x^1=(0, 0)^T$

解： $\nabla f(x) = \left(2x_1 - 2x_2, -2x_1 + 4x_2 + 2\right)^T$ ，

k =1时， $\nabla f(x^1) = \left(0, 2\right)^T \neq 0$ **，$d^1=(0, -2)^T$，** $f(x^1 + \lambda d^1) = 8\lambda^2 - 4\lambda + 2$

可得$\lambda_1 = 1/4$，

$$x^2 = x^1 + \lambda_1 d^1 = \left(0, -\frac{1}{2}\right)^T, \quad \nabla f(x^2) = \left(1, 0\right)^T$$

$d^2=(-1, -1/2)^T$

k =2时， $f(x^1 + \lambda d^1) = \lambda^2 + \frac{1}{2}(1+\lambda)^2 - \lambda(1+\lambda) - (1+\lambda) + 2$

可得$\lambda_2 = 1$，

$$x^3 = \left(-1, -1\right)^T, \quad \nabla f(x^3) = \left(0, 0\right)^T$$

故 $x^* = x^3 = (-1, -1)^T$为最优解。