第8章　数值优化

目　录

8.1 单变量函数的极小值

- 定义8.1 如果存在包含p的开区间I，

使得对所有$x \in I$，有$f(p) \leq f(x)$，则称函数f在$x=p$处有局部极小值。 类似地，

如果对所有$x \in I$，有$f(p) \geq f(x)$，则称函数f在$x=p$处有局部极大值。

如果f在点$x=p$处有局部极大值或极小值，则称f在点$x=p$处有局部极值。

单调性的定义和判定

- 定义8.2 设$f(x)$定义在区间I上。
  - 若对所有$x_1<x_2$，当$x_1, x_2\in I$时有$f(x_1)<f(x_2)$，则称f在区间I上递增。
  - 若对所有$x_1<x_2$，当$x_1, x_2\in I$时有$f(x_1)>f(x_2)$，则称f在区间I上递减。

单调性的定义和判定

- **定理8.1 设$f(x)$ 在区间$I=[a,b]$上连续，并在(a,b)上可微。**
  - **若对所有$x\in(a,b)$有$f'(x)>0$，则$f(x)$在 I上递增。**
  - **若对所有$x\in(a,b)$有$f'(x)<0$，则$f(x)$在 I上递减。**

驻点和一阶导数测试

- 定理8.2 设$f(x)$定义在区间$I=[a, b]$上，并在内点$p\in(a, b)$处有局部极值。若$f(x)$在$x=p$处可微，

  则 $f'(p)=0$。

- 定理8.3 设$f(x)$ 在$I=[a, b]$上连续，并设除$x=p$处外，$f'(x)$对所有$x\in(a, b)$都有定义。
  - 若在(a, p)上$f'(x)<0$，而在(p, b)上$f'(x)>0$，则$f(p)$是局部极小值。
  - 若在(a, p)上$f'(x)>0$，而在(p, b)上$f'(x)<0$，则$f(p)$是局部极大值。

二阶导数测试

- 定理8.4 设f在区间$[a, b]$上连续，并且f''和f''在区间(a, b)上有定义。又设$p \in (a, b)$是关键点，即$f'(p)=0$。
  - 若$f''(p)>0$，则$f(p)$是f的一个局部极小值。
  - 若$f''(p)<0$，则$f(p)$是f的一个局部极大值。
  - 若$f''(p)=0$，则结果不确定。

8.1.1 分类搜索方法

- 直接法是一种数值方法

  基本思想: 迭代，通过迭代产生一个点序列{ $X^{(k)}$ }，使之逐步接近最优点

- 只用到目标函数，通过对函数多次求值来求函数$f(x)$在给定区间上的一个局部极小值

- 要尽量减少函数求值的次数，确定在哪里求$f(x)$值的好策略非常重要

- 如黄金分割搜索法、Fibonacci搜索法、随机搜索法

搜索法必须满足的条件

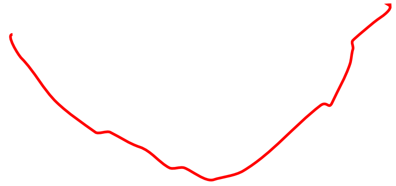
- 使用这些方法来求$f(x)$的极小值必须满足特定的条件，以保证在给定的区间内有合适的极小值
- 这个特定条件就是：函数$f(x)$在给定区间中是单峰的
- 定义8.3 如果存在唯一的$p \in I$，使得

  （1） $f(x)$在$[a, p]$上递减，

  （2） $f(x)$在$[p, b]$上递增，

则函数$f(x)$在$I=[a, b]$上是（下）单峰的。

黄金分割搜索法（0.618法）



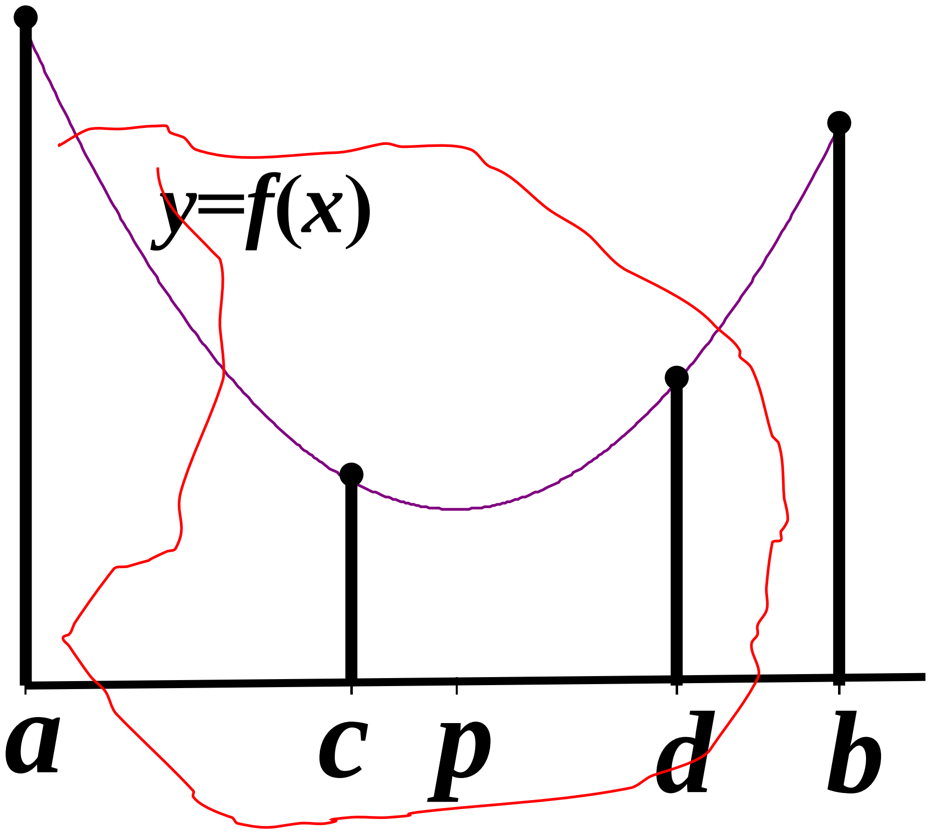
- 如果已知$f(x)$在$[a,b]$上是（下）单峰的，则有可能找到该区间的一个子区间，$f(x)$在该子区间上取得极小值
- 选择两个内点$c<d$，这样就有$a<c<d<b$。

$f(x)$的单峰特性保证了:

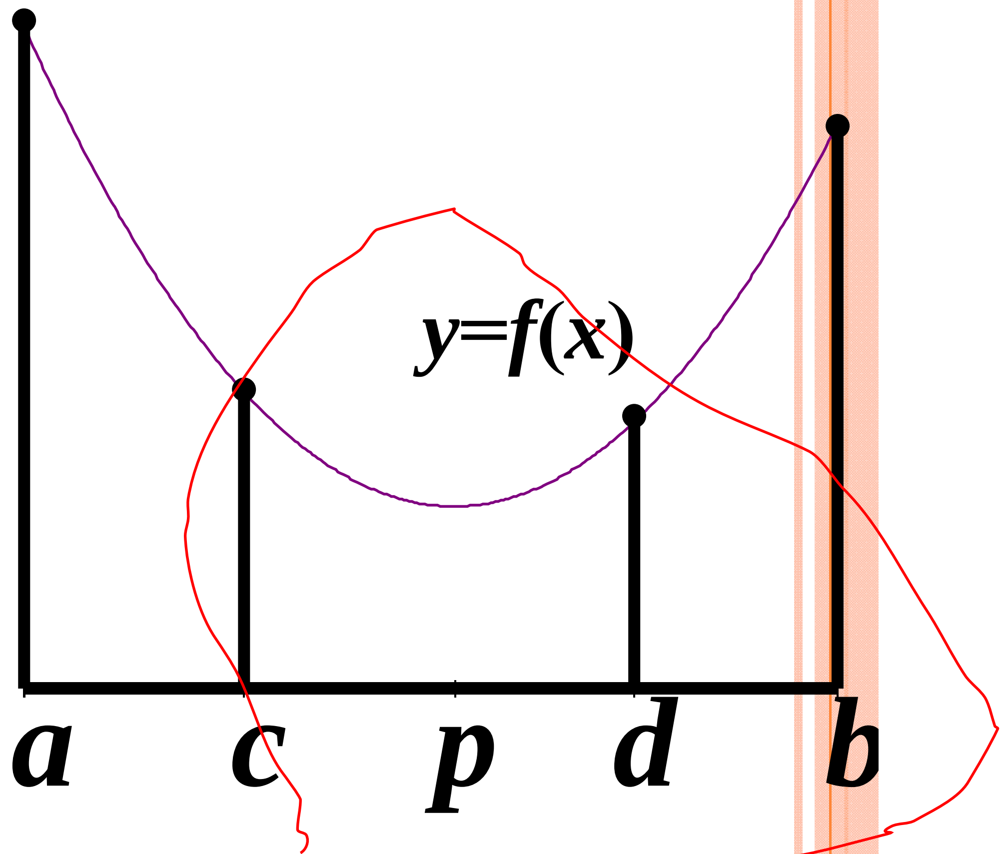
$$\text{函数值} f(c) \text{和} f(d) \text{小于} \max\{f(a), f(b)\}$$

- 出现两种情况:

黄金分割搜索方法的决策过程



如果$f(c)\le f(d)$，则从右侧压缩，使用$[a,d]$



如果$f(c)>f(d)$，则从左侧压缩，使用$[c,b]$

黄金分割法原理

设函数 $f(x)$ 在闭区间 $[a, b]$ 上是（下）单峰函数，

即在 (a, b) 内 $f(x)$ 有唯一的极小点 p.

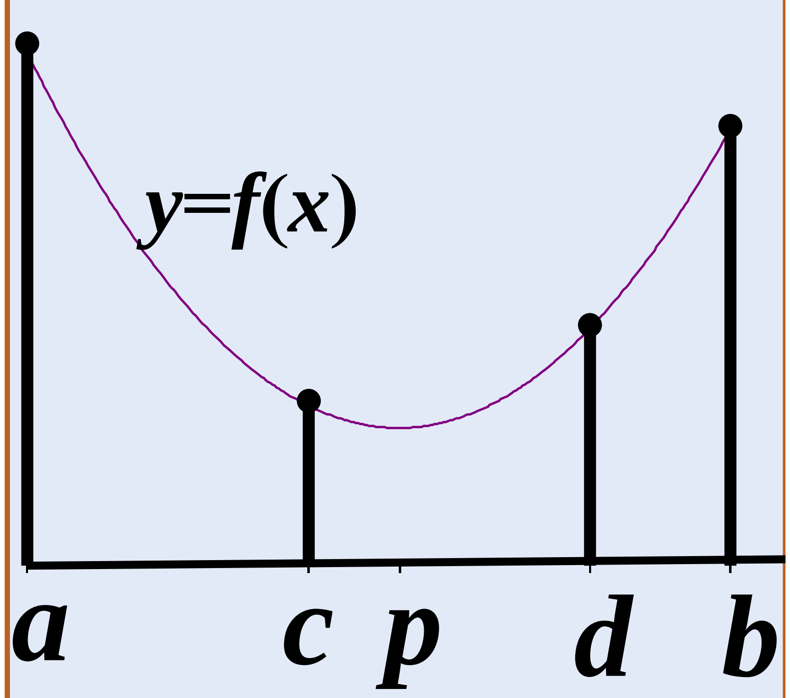
在 p 的左边 $f(x)$ 严格单调下降

在 p 的右边 $f(x)$ 严格单调上升

那么对于 (a, b) 内任意两点 $c < d$，

如果 $f(c) < f(d)$，则 $p \in [a, d]$；

否则 $p \in [c, b]$



如果 $f(c) \le f(d)$，则从右侧压缩，使用 $[a,d]$

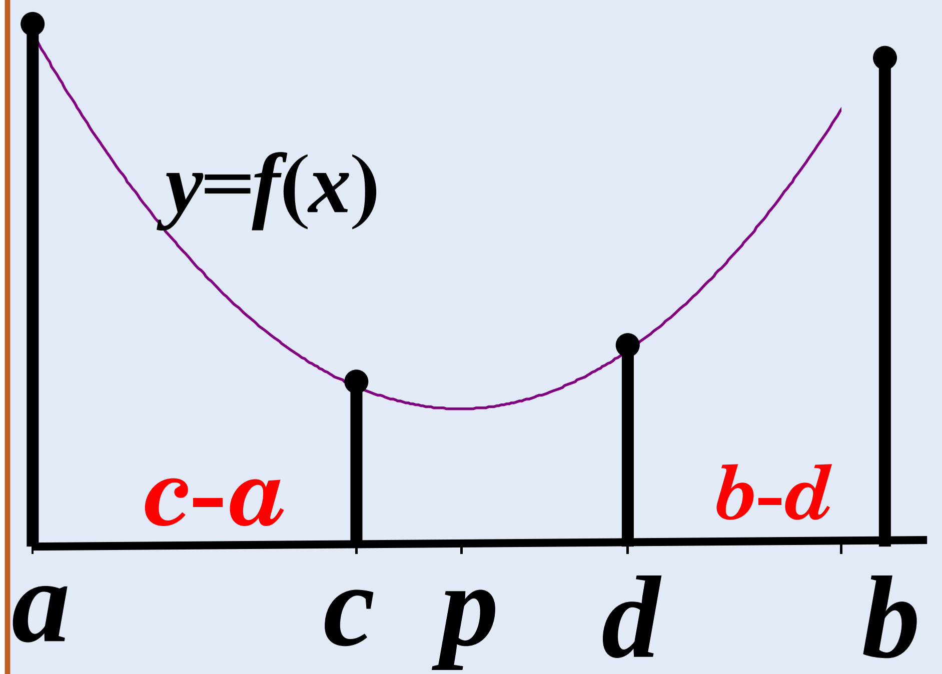
黄金分割法

- 选择内点c和d，使得区间$[a,c]$与$[d,b]$对称，

即$b-d=c-a$，其中　$c=a+(1-r)(b-a)=ra+(1-r)b$

$d=b-(1-r)(b-a)=(1-r)a+rb$

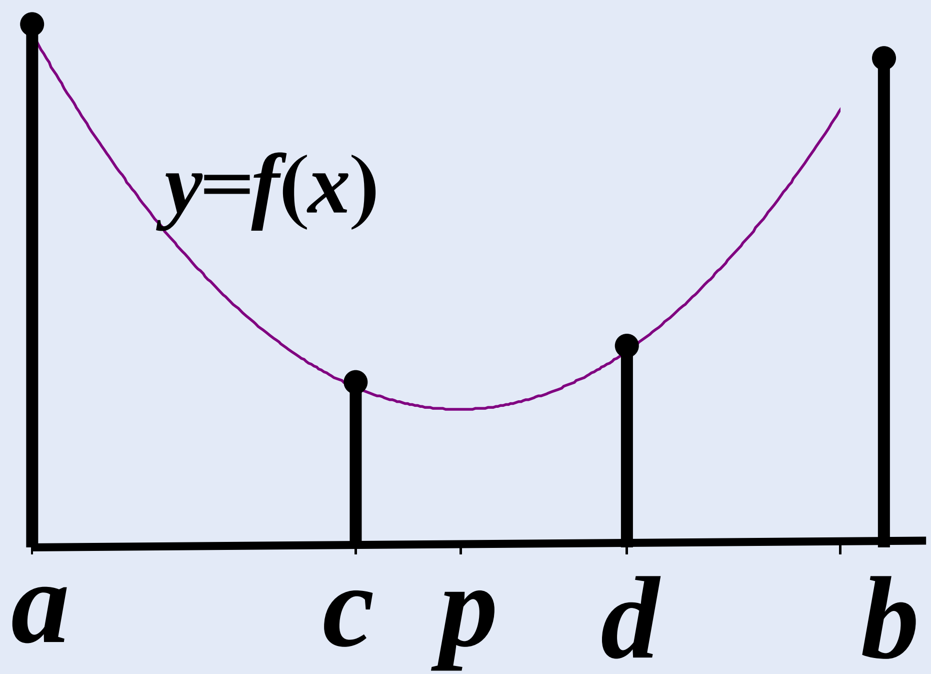
并且$1/2<r<1$（保证$c<d$）



如果$f(c)\le f(d)$，则从右侧压缩，使用$[a,d]$

黄金分割法

- 希望r在每个子区间上保持为常数， 且旧的内点中有一个成为新子区间的一个内点， 而另一个则成为新子区间的一个端点



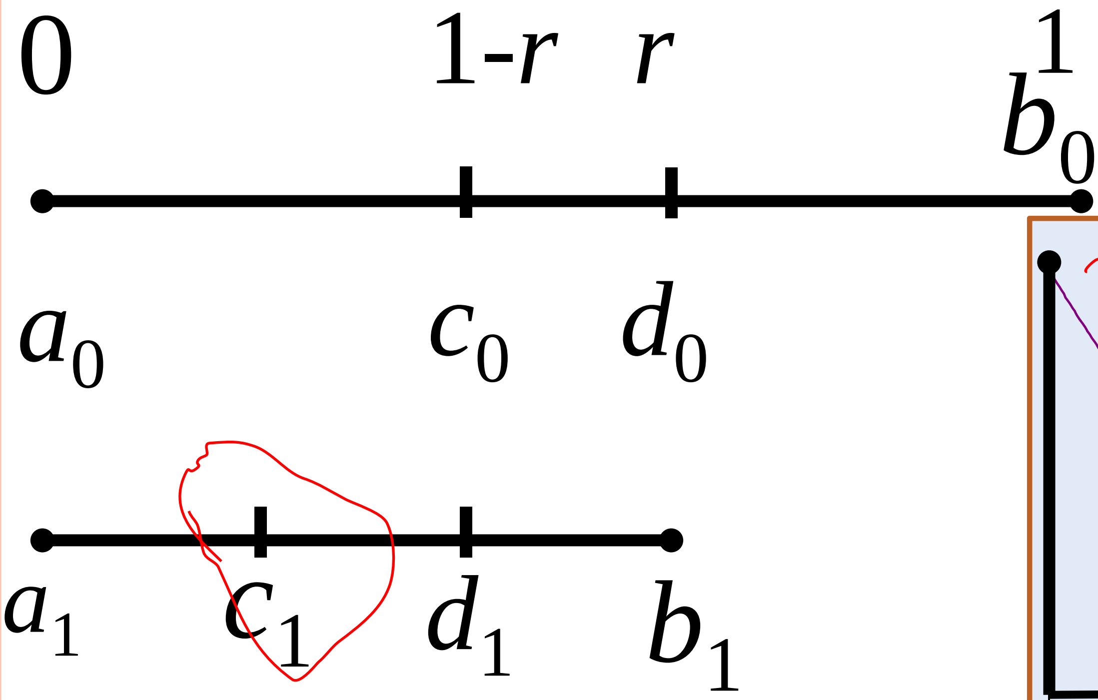
如果$f(c)\le f(d)$，则从右侧压缩，使用$[a,d]$

- 在每次迭代中只需要找一个新的点，则只需要一次新的函数求值计算

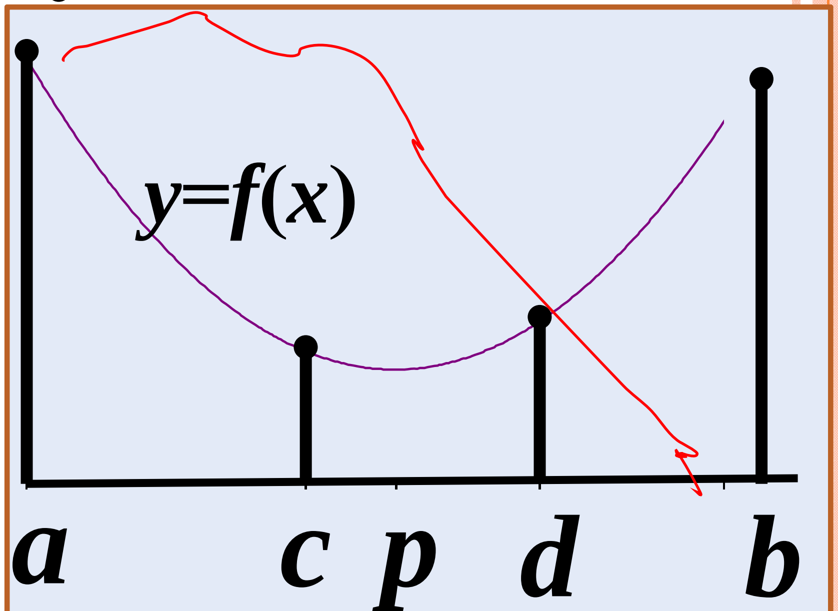
比例因子的选择

设$f(c_0)\le f(d_0)$，则从右侧压缩，使用$[a_0,d_0]$，

即取$a_1=a_0$，$b_1=d_0$和$d_1=c_0$，则需要再求一个新的点c_1



因为$1/2<r<1$（保证$c<d$）

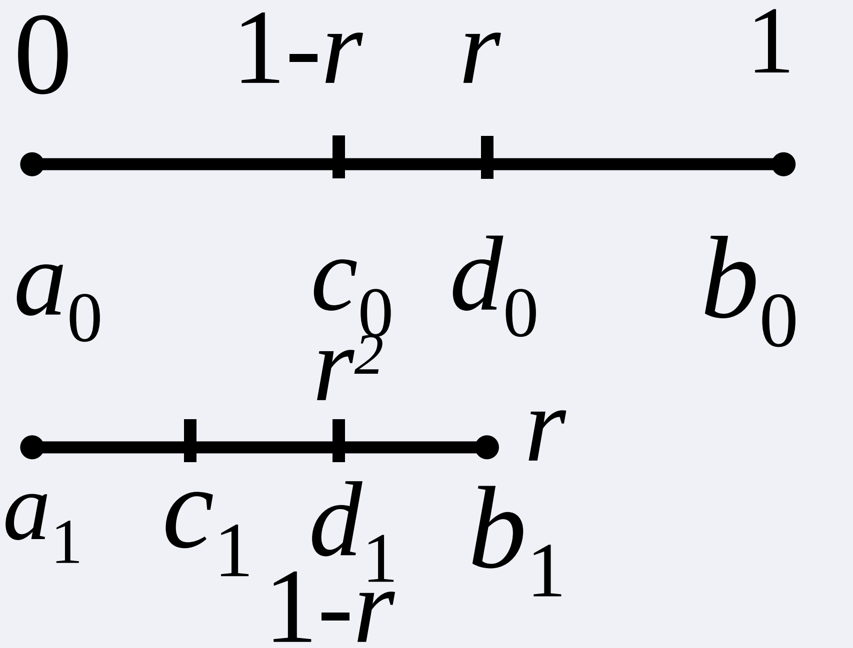


如果$f(c)\le f(d)$，则从右侧压缩，使用$[a,d]$

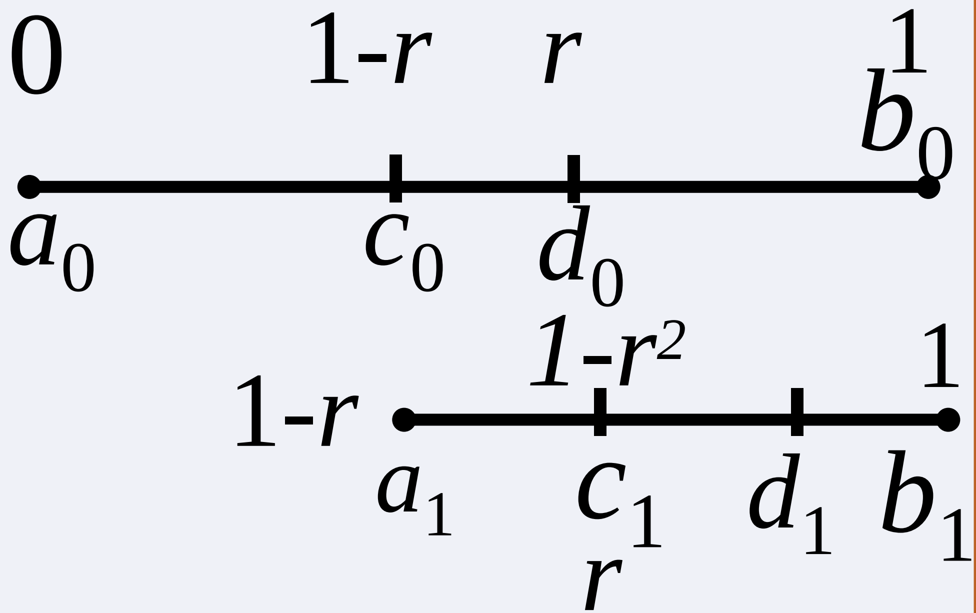
希望：**1**， ***r***在每个区间上为常数

2，旧的内点中的一个为新的子区间的内点

另一个为新的子区间的一个端点



从右侧压缩，新区间***[0,r]***



从左侧压缩，新区间***[1-r,1]***

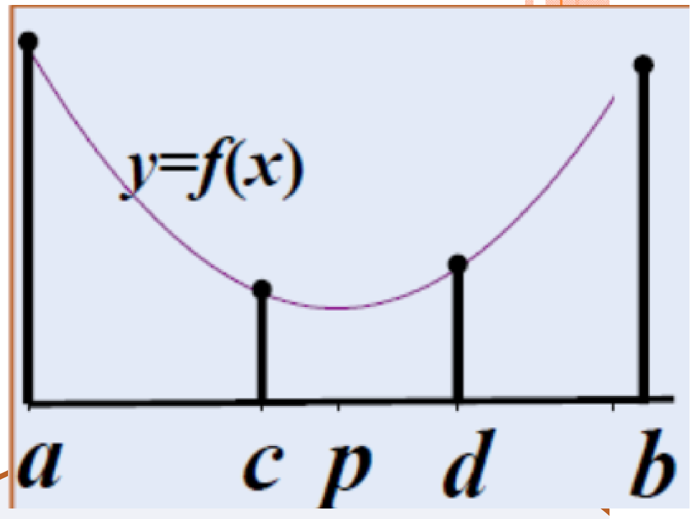
因此：1， 每次迭代中只需要找一个新的点

2，只需要一次新的函数求值计算

假设：*f(c)* ≤*f(d)*,且只进行一次新的函数求值，则有

$$\frac{d-a}{b-a}=\frac{c-a}{d-a}$$

$$\frac{r(b-a)}{b-a}=\frac{(1-r)(b-a)}{r(b-a)}$$



黄金分割比

$$\frac{r}{1}=\frac{1-r}{r}$$

$$r^2+r-1=0$$

$$r=\frac{-1\pm\sqrt{5}}{2}\approx 0.618$$

例:利用黄金分割法求单峰函数

f(x)=x^2-sin(x) 在**[0,1]**上的极小值

解：令：$a_0=0, b_0=1$

$$c_0=0+(1-\frac{-1+\sqrt{5}}{2})(1-0)=\frac{3-\sqrt{5}}{2}\approx 0.38919660$$

$$d_0=1-(1-\frac{-1+\sqrt{5}}{2})(1-0)=\frac{-1+\sqrt{5}}{2}\approx 0.6180340$$

$f(c_0)=-0.22684748,\qquad f(d_0)=-0.19746793$

所以有 $f(c_0)<f(d_0)$

所以新的子区间为：

$[a_0,d_0]=[0.0000000,0.6180340]$

令： $a_1 = a_0, b_0 = d_0, d_1 = c_0$

$$c_1 = a_1 + (1-r)(b_1 - a_1)$$

$$= 0 + (1 - \frac{-1+\sqrt{5}}{2})(0.6180340 - 0) \approx 0.2360680$$

计算$f(c_1)$、$f(d_1)$并比较，确定新的子空间，继续迭代，结果如表8.2

结果。

表 8.2　求 $f(x)=x^2-\sin(x)$ 的极小值的黄金分割搜索法

k	a_k	c_k	d_k	b_k	$f(c_k)$	$f(d_k)$
0	0.0000000	0.3819660	0.6180340	1	−0.22684748	−0.19746793
1	0.0000000	0.2360680	0.3819660	0.6180340	−0.17815339	−0.22684748
2	0.2360680	0.3819660	0.4721360	0.6180340	−0.22684748	−0.23187724
3	0.3819660	0.4721360	0.5278640	0.6180340	−0.23187724	−0.22504882
4	0.3819660	0.4376941	0.4721360	0.5278640	−0.23227594	−0.23187724
5	0.3819660	0.4164079	0.4376941	0.4721360	−0.23108238	−0.23227594
6	0.4164079	0.4376941	0.4508497	0.4721360	−0.23227594	−0.23246503
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	0.4501574	0.4501730	0.4501827	0.4501983	−0.23246558	−0.23246558
22	0.4501730	0.4501827	0.4501886	0.4501983	−0.23246558	−0.23246558
23	0.4501827	0.4501886	0.4501923	0.4501983	−0.23246558	−0.23246558

在第 23 次迭代时，区间收缩为 $[a_{23},b_{23}]=[0.4501827,0.4501983]$。该区间的宽度为 0.0000156。而在该区间的两个端点处求得的函数值在小数点后有 8 位相同，即 $f(a_{23})\approx -0.23246558\approx f(b_{23})$，因此算法结束。搜索法的一个问题是，函数在极小值附近可能比较平缓，从而限制了精度。割线方法能够求得更精确的解 $p_5=0.4501836$。

尽管本例中黄金分割搜索法的速度较慢，但它的优点是可用于 $f(x)$ 不可微的情况。■

目　录

- ■单变量函数的极小值
- ■内德-米德方法（单纯形法）
- ■最速下降法（梯度方法）
- ■牛顿方法

多元函数求极值的问题

设函数$f(x_1,x_2,\ldots,x_N)$定义在区域

$$R=\left\{(x_1,x_2,\ldots,x_N):\sum_{k=1}^{N}(x_k-p_k)^2<r^2\right\}$$

如果$f(p_1,p_2,\ldots,p_N)\le f(x_1,x_2,\ldots,x_N)$对所有的点$(x_1,x_2,\ldots,x_N)\in R$都成立，则函数$f(x_1,x_2,\ldots,x_N)$在点$(p_1,p_2,\ldots,p_N)$处有局部极小值；

如果$f(p_1,p_2,\ldots,p_N)\ge f(x_1,x_2,\ldots,x_N)$对所有的点$(x_1,x_2,\ldots,x_N)\in R$都成立，则函数$f(x_1,x_2,\ldots,x_N)$在点$(p_1,p_2,\ldots,p_N)$处有局部极大值。

多元函数的极小值问题

- 二元函数的图形是一个几何表面

- 定理8.5（二阶偏导数测试）

  设$f(x,y)$及其一阶和二阶偏导数在区域R上连续。

  设点$(p,q)\in R$是一个临界点，即:

  $$f_x(p,q)=0 \text{且} f_y(p,q)=0。$$

  可用高阶偏导数来确定临界点的属性。

1，若 $f_{xx}(p,q)f_{yy}(p,q)-f_{xy}^2(p,q)>0$ 且 $f_{xx}(p,q)>0$，

则 $f(p,q)$是 f 的局部极小值。

2，若 $f_{xx}(p,q)f_{yy}(p,q)-f_{xy}^2(p,q)>0$ 且 $f_{xx}(p,q)<0$，

则 $f(p,q)$是 f 的局部极大值。

3，若 $f_{xx}(p,q)f_{yy}(p,q)-f_{xy}^2(p,q)<0$

则$f(x,y)$在(p,q)没有局部极 值。

4，若$f_{xx}(p,q)f_{yy}(p,q)-f_{xy}^2(p,q)=0$，则结果不确定。

例8.5

例8.5 求函数**f(x,y)=x^2-4x+y^2-y-xy**的极小值

解： $f_x(x,y)=2x-4-y,\ f_y(x,y)=2y-1-x$

令：$\begin{cases} f_x(x,y)=0 \\ f_y(x,y)=0 \end{cases} \Rightarrow \begin{cases} 2x-y=4 \\ -x+2y=1 \end{cases}$

$\Rightarrow (x,y)=(3,2),\ f(x,y)$的二阶偏导：

$f_{xx}(x,y)=2,\ f_{yy}(x,y)=2,\ f_{xy}(x,y)=-1$

可得为定理8.5中的第一种情况，即

$f_{xx}(3,2)f_{yy}(3,2)-f_{xy}^2(3,2)=3>0$，$f_{xx}(3,2)=2>0$

所以$f(x,y)$在(3, 2)有局部极小值$f(3,2)$=-7

多元函数的直接搜索法

- 多变量目标函数$f(x_1,x_2,\ldots,x_N)$的极值直接搜索法:

  对函数的可微性不作显性或隐性的假设

- 对非光滑（不可微）目标函数而言，直接方法特别有用内德一米德方法(单纯形方法)和鲍威尔方法

单纯形方法的基本思想

- 内德和米德提出了单纯形法，可用于求解多变量函数的局部极小值
- 从可行域中的一个基本可行解出发，判断它是否已是最优解，若不是，寻找下一个基本可行解，并使目标函数得到改进，如此迭代下去，直到找出最优解或判定问题无解为止。
- 从另一个角度说，就是从可行域的某一个极点出发，迭代到另一个极点，并使目标函数的值有所改善，直到找出有无最优解时为止。

单纯形的概念

- 单纯形是指0维中的点，一维中的线段，二维中的三角形，三维中的四面体，n维空间中的有$n+1$个顶点的多面体。

例如在三维空间中的四面体，其顶点分别为:

(0,0,0)，(1,0,0)，(0,1,0)，(0,0,1)。

具有单位截距的单纯形的方程是$\sum x_i \leq 1$，

并且$x_i \geq 0$，$i=1,2,\ldots,n$

二元函数的单纯形方法

- 在二维平面空间中，单纯形就是三角形
- 搜索过程：

　　比较三角形3个顶点处的函数值，$f(x,y)$值最大的顶点为最差顶点(W)，用一个新的顶点代替最差顶点，形成新的三角形式

- 继续这一过程，生成一系列三角形（它们可能具有不同的形状），函数在其顶点处的值越来越小
- 随着三角形的减小就可以找到极小值点的坐标

单纯形的寻找过程

① 初始三角形BGW

② 良边的中点

③ 反射点R

④ 开拓点E

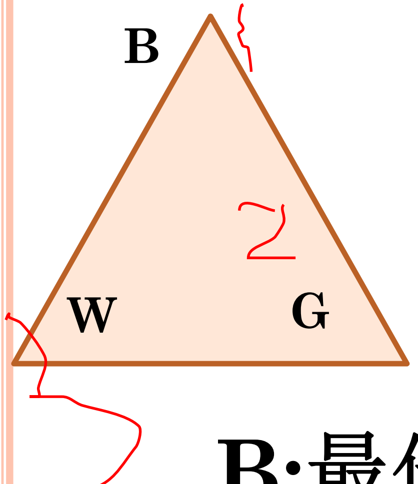
⑤ 收缩点C

⑥ 向B方向收缩

单纯形的寻找过程　　例：求*f*(*x*，*y*)极小值

① 初始三角形BGW

- 给定三角形三个顶点 $V_k=(x_k,y_k),k=1,2,3$
- 求值：$z_k=f(x_k,y_k),k=1,2,3$



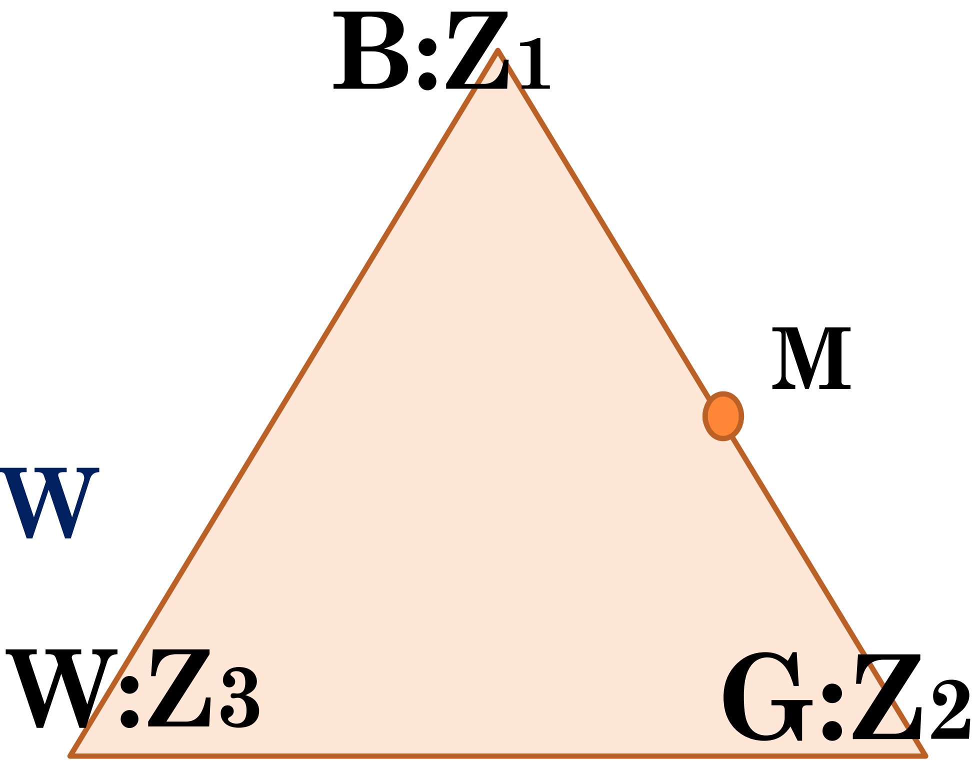
下标满足 $z_1\le z_2\le z_3$

$B(x_1,y_1)$，　$G(x_2,y_2)$，　$W(x_3,y_3)$

B:最佳顶点　　**G:**次最佳顶点　　**W:**最差顶点

单纯形的寻找过程

例：求*f(x，y)*极小值



① 初始三角形BGW

② 良边的中点

M: BG的中点

$$z_1 \le z_2 \le z_3$$

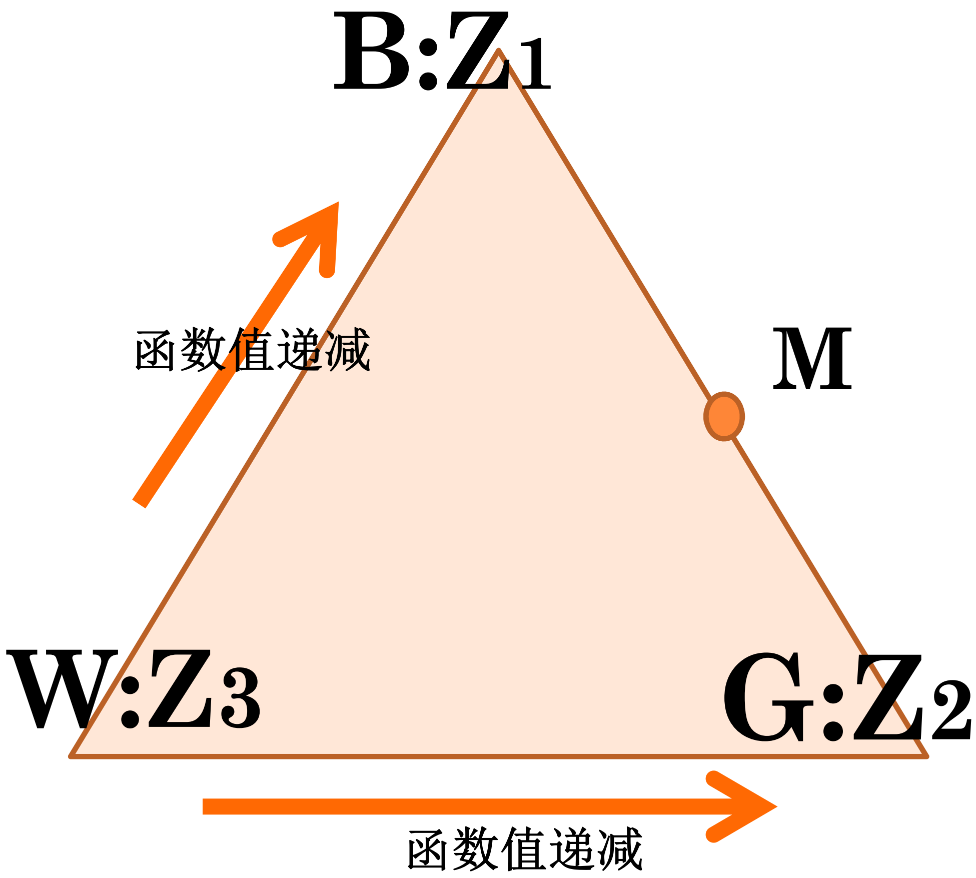
$$M = \frac{B+G}{2} = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$$

单纯形的寻找过程

① 初始三角形BGW

② 良边的中点

③ 反射点R



W→B方向：值递减

W→G方向：值递减

所以：

以BG连线为分界线，与W相对的点函数值*f(x,y)*较小

选择：

测试点R：关于边$\overline{BG}$对三角形进行“反射”

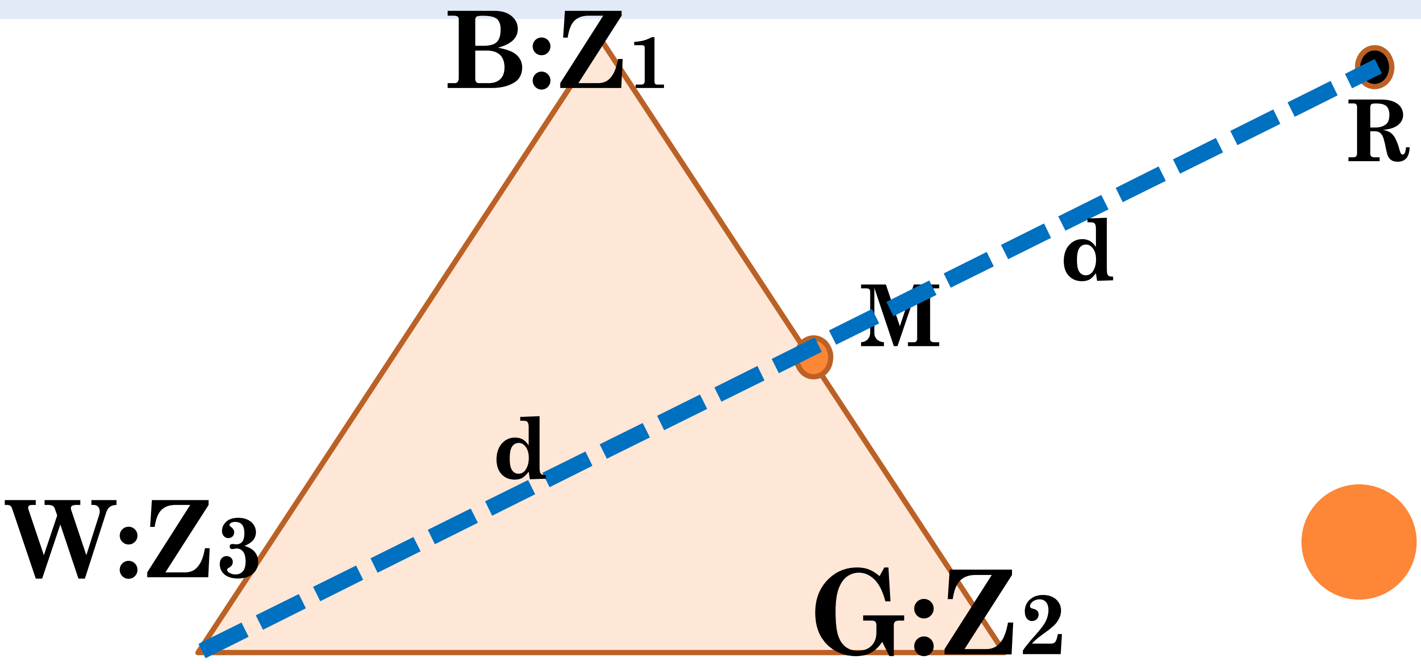
如何确定反射点R?

1, 找到中点M　2,从W到M画线段，　长度为d

3,从M做线段延长线，　长度为d,得到R

R的向量公式为:

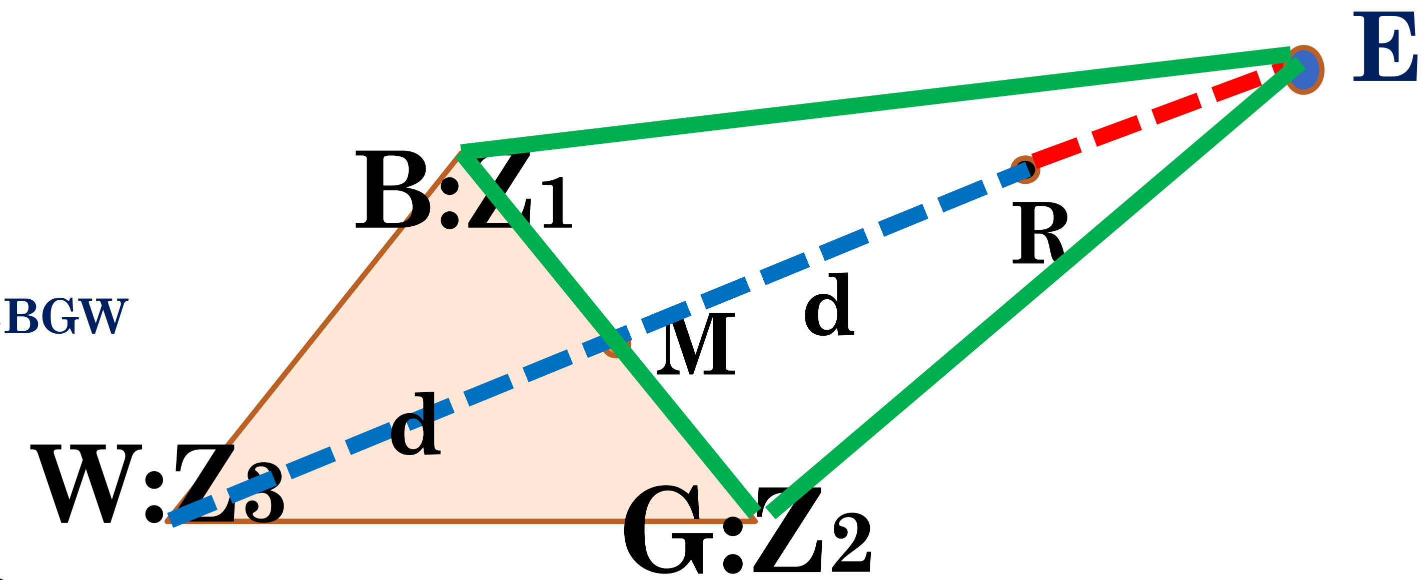
$$R=M+(M-W)$$



① 初始三角形BGW

② 良边的中点

③ 反射点R



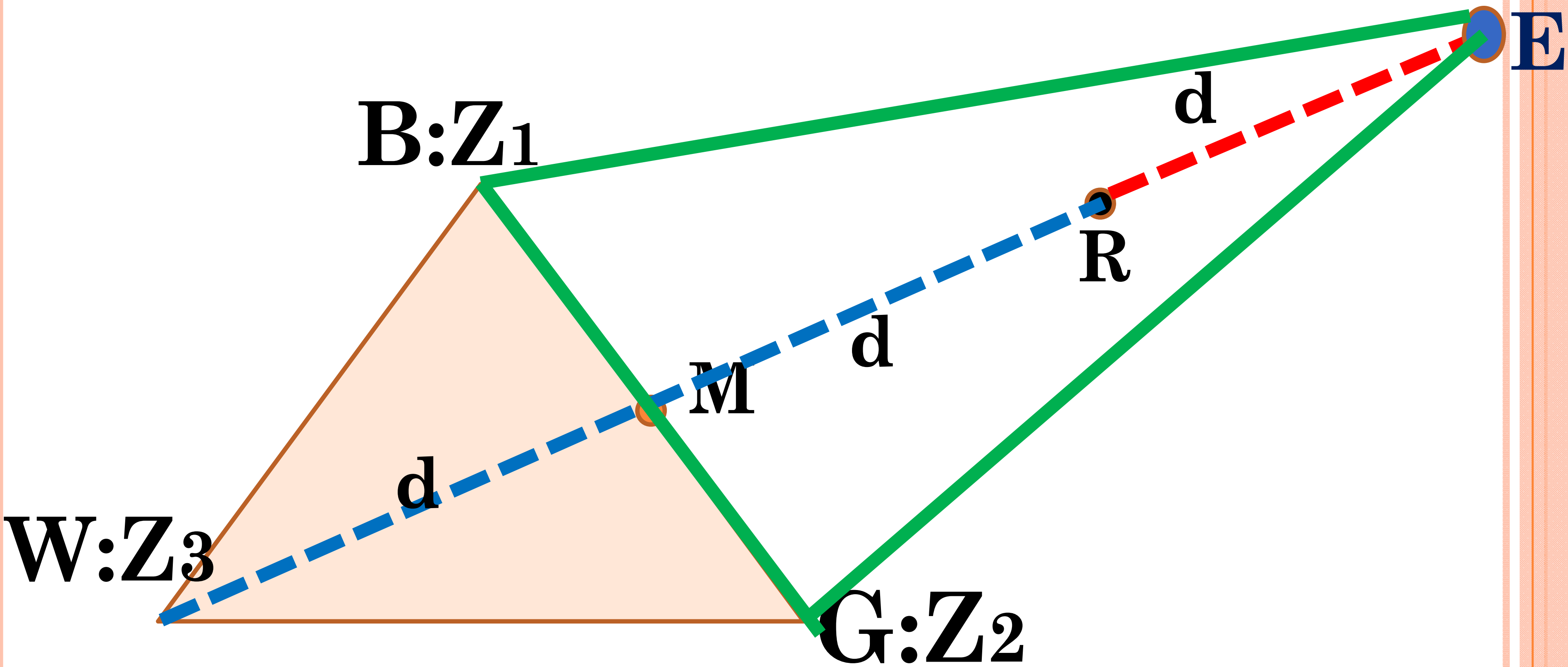
④ 开拓点E

如果：$f_R < f_W$,则求解方向正确

可能的极小点比R远一点

所以：过M和R延长到E

构成新的开拓三角形BGE



点E 沿着连结M和R的线段方向延长距离d

如果$f_E<f_R$,则改点比R好。 点E的向量公式为:

$$E=R+(R-M)=2R-M$$

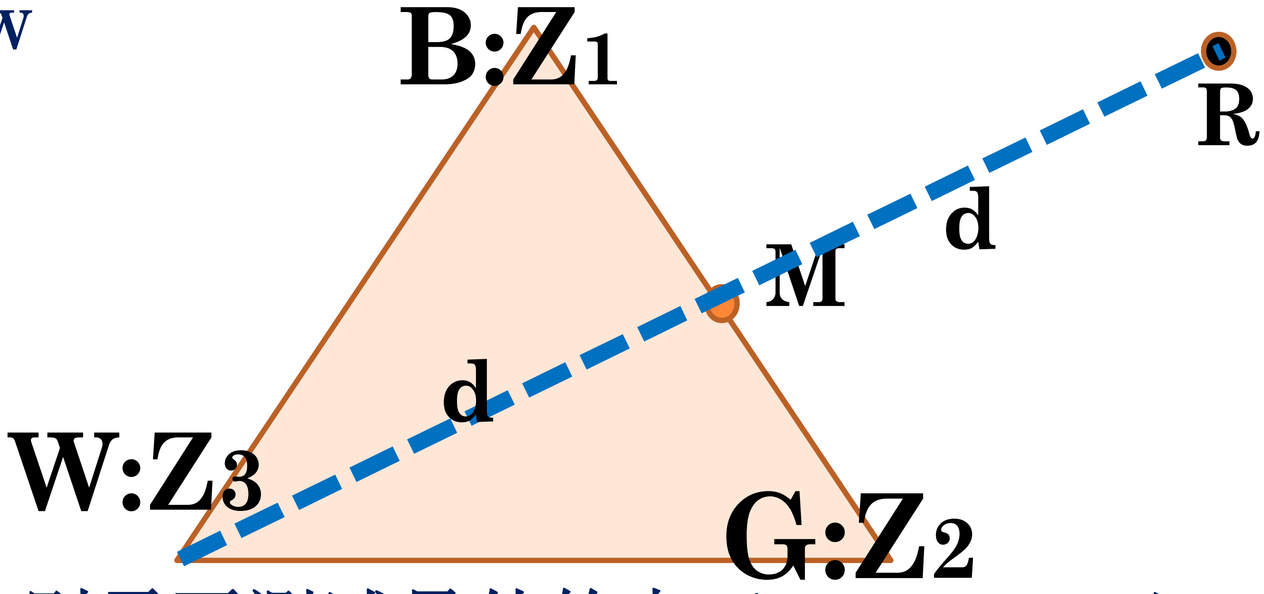
① 初始三角形BGW

② 良边的中点

③ 反射点R

④ 开拓点E

⑤ 收缩点C



如果$f_R=f_W$，则需要测试另外的点（也许M函数值小）

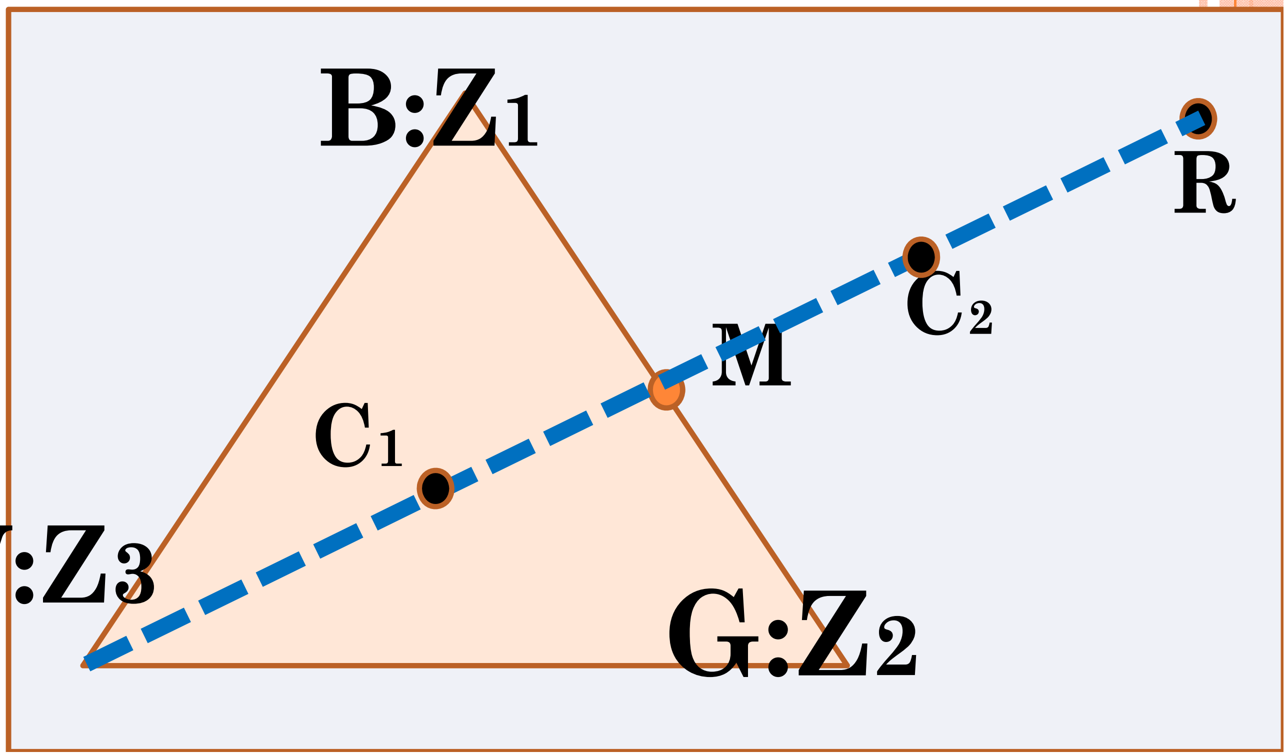
为了保证构成三角形，

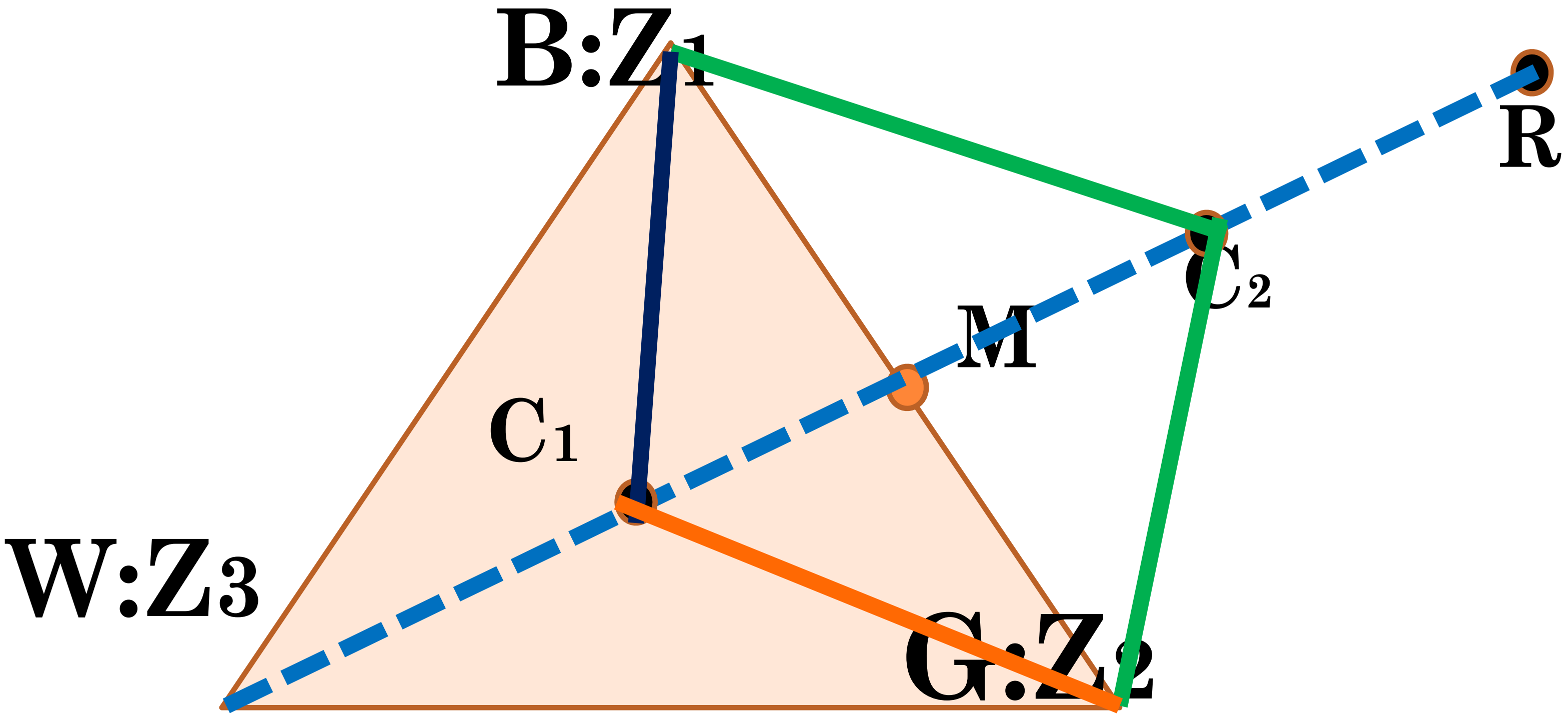
不能取M代替W，

分别考虑

$\overline{WM}$:C1

$\overline{MR}$:C2





$$f_C=\min(f_{C1},f_{C2})$$

具有较小函数值的为C,并由此构成三角形BGC

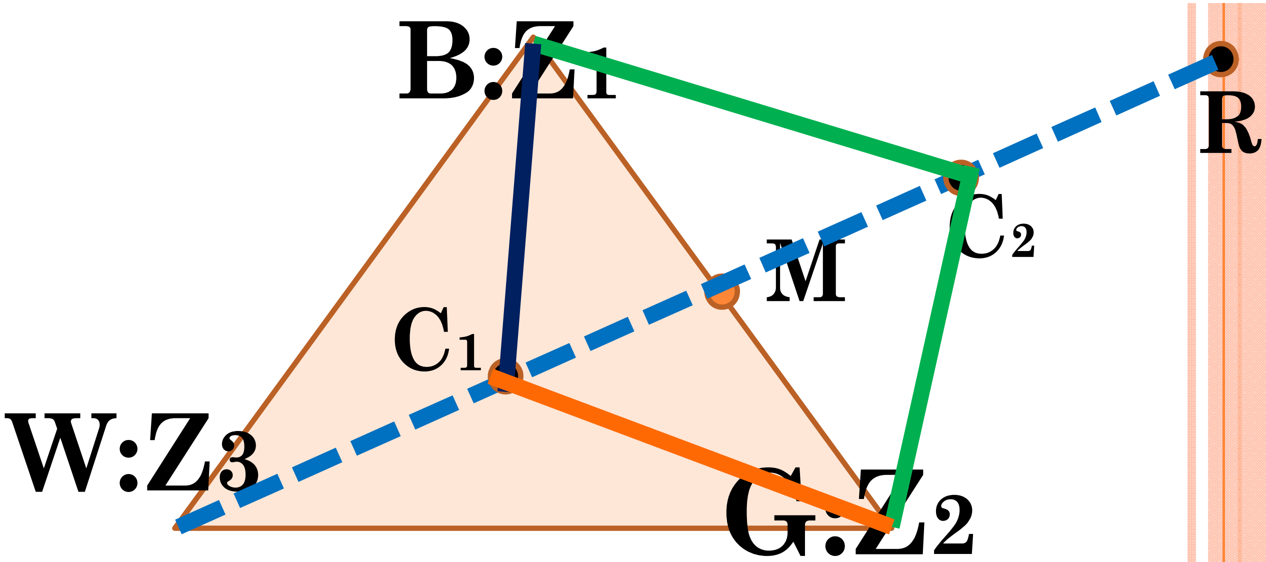
① 初始三角形BGW

② 良边的中点

③ 反射点R

④ 开拓点E

⑤ 收缩点C



⑥ 向B方向收缩

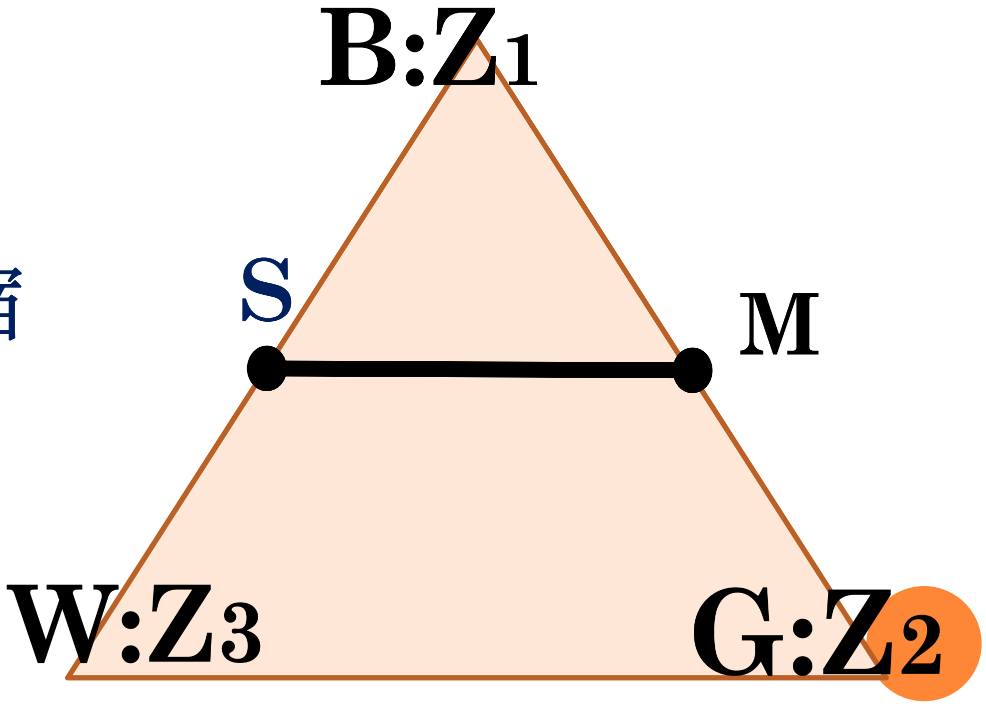
如果： $f_c \geq f_w$

则：G、W必向B收缩

G置换为: M

W置换为: S

S为连接B和W的线段中点



每一步的逻辑判断

- 若$f(R)<f(G)$，则转①（反射点或开拓点）

  否则转②（压缩点或收缩点）

①若$f(B)<f(R)$，则以R代替W

否则计算E和$f(E)$

若$f(E)<f(B)$，则以E代替W

否则以R代替W

例8.6

②若$f(R)<f(W)$，则以R代替W

否则计算$C_1=(M+R)/2$和$f(C_1)$

计算$C_2=(W+M)/2$和$f(C_2)$

取两者中函数值较小者为C

若$f(C)<f(W)$，则以C代替W

否则计算$S=(W+B)/2$和$f(S)$

以S代替W，以M代替G

例如：**8.6** 利用单纯形法求

$$f(x,y)=x^2-4x+y^2-y-xy\text{的极小值}$$

从以下三个顶点开始：

$$V_1=(0,0),V_2=(1.2,0.0),V_3=(0.0,0.8)$$

解：$f(x,y)$在顶点处的值为：

$f(0,0)=0.0,\ f(1.2,0.0)=-3.36,\ f(0.0,0.8)=-0.16$

比较三个顶点的值，确定B, G, W:

$B=(1.2,0.0),G=(0.0,0.8),W=(0,0)$

所以顶点$W=(0,0)$将被替代

点M和R为：

M=$\dfrac{B+G}{2}=(0.6,0.4),R=2M-W=(1.2,0.8)$

$f(R)=f(1.2,0.8)=-4.48<f(G)$

属于第一种情况

因为：$f(R)\le f(B)$,所以移动方向正确

所以：

$E=2R-M=2(1.2,0.8)-(0.6,0.4)=(1.8,1.2)$

$f(E)=f(1.8,1.2)=-5.88<f(B)$

所以新三角形顶点为：

$V_1=(1.8,1.2),V_2=(1.2,0.0),V_3=(0.0,0.8)$

构造。函数值 $f(E)=f(\ldots$

$$V_1=(1.8,1.2),\qquad V_2=(1.2,0.0),$$

继续该过程，得到朝着解点(3,2)收敛的一系列三角形(见图 8.10)。表 8.6 给出了迭代中各步的三角形顶点处的函数值。计算机实现的算法执行到第 33 步，得到的最佳顶点为 $B=(2.99996456,1.99983839)$，函数值为 $f(B)=-6.99999998$。它是例 8.5 中求得的 $f(3,2)=-7$ 的近似值。迭代在到达(3,2)之前停止的原因是，函数在极小值点附近平缓。经过检查(见表 8.6)，函数值 $f(B)$，$f(G)$和 $f(W)$(见表 8.6)相同(这是舍入误差的一个例子)，算法终止。■

表 8.6　例 8.6 中各三角形顶点处的函数值

k	最佳点	较好点	最差点
1	$f(1.2, 0.0)=-3.36$	$f(0.0, 0.8)=-0.16$	$f(0.0, 0.0)=0.00$
2	$f(1.8, 1.2)=-5.88$	$f(1.2, 0.0)=-3.36$	$f(0.0, 0.8)=-0.16$
3	$f(1.8, 1.2)=-5.88$	$f(3.0, 0.4)=-4.44$	$f(1.2, 0.0)=-3.36$
4	$f(3.6, 1.6)=-6.24$	$f(1.8, 1.2)=-5.88$	$f(3.0, 0.4)=-4.44$
5	$f(3.6, 1.6)=-6.24$	$f(2.4, 2.4)=-6.24$	$f(1.8, 1.2)=-5.88$
6	$f(2.4, 1.6)=-6.72$	$f(3.6, 1.6)=-6.24$	$f(2.4, 2.4)=-6.24$
7	$f(3.0, 1.8)=-6.96$	$f(2.4, 1.6)=-6.72$	$f(2.4, 2.4)=-6.24$
8	$f(3.0, 1.8)=-6.96$	$f(2.55, 2.05)=-6.7725$	$f(2.4, 1.6)=-6.72$
9	$f(3.0, 1.8)=-6.96$	$f(3.15, 2.25)=-6.9525$	$f(2.55, 2.05)=-6.7725$
10	$f(3.0, 1.8)=-6.96$	$f(2.8125, 2.0375)=-6.95640625$	$f(3.15, 2.25)=-6.9525$

目　录

梯度方法和牛顿方法

- 内德－米德方法（单纯形法）：

  针对多元函数偏导数不可求的方法，是直接搜索法

- 梯度方法和牛顿方法：

  针对函数$\boldsymbol{f(X)}$的偏导数可得的求极小值的方法，

  其中，$\boldsymbol{X=(x_1, x_2, \ldots, x_N)^T}$

梯度的定义

定义8.4

设$z=f(X)$是X的函数，对$k=1,2,\ldots,N$，

$\partial f(\boldsymbol{X})/\partial x_k$存在。$f$的梯度，记为$\nabla f(X)$，是向量

$$\nabla f(\boldsymbol{X})=\left(\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{x}_1},\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{x}_2},\cdots,\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{x}_N}\right)^T$$

- 梯度向量在局部指向$\boldsymbol{f(X)}$增加得最快的方向。

因此$-\nabla \boldsymbol{f(X)}$在局部指向下降得最快的方向。

关于函数梯度的说明

- 梯度是一个向量。N元函数$f(x_1, x_2, \ldots x_N)$在某点x处的梯度为：

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_N}\right)^T$$

- 梯度的方向与函数f的等值线的一个法线方向相同，从较低的等值线指向较高的等值线。
- 梯度的方向：函数f的值增加最快的方向，

  其相反方向：函数f的值降低最快的方向。

梯度法的定义

- 梯度法又称最速下降法（**steepest descent method**）
- 由法国数学家**Cauchy**于**1847**年首先提出。

在每次迭代中，

沿最速下降方向（负梯度方向）进行搜索，

每步沿负梯度方向取最优步长，

因此这种方法称为最优梯度法。

梯度方法描述

- 从迭代初始点$\boldsymbol{P_0}$开始，沿着过$\boldsymbol{P_0}$， 方向为:

$$S_0 = \frac{-\nabla f(P_0)}{\| -\nabla f(P_0) \|}$$

的直线搜索，到达点$\boldsymbol{P_1}$.

当点$\boldsymbol{X}$满足约束$\boldsymbol{X=P_0+\gamma S_0}$时，在该点取得局部极小值

- 由于偏导数可得，因此极小化过程可以使用二次或三次近似方法

梯度方法描述

- 然后计算-$\nabla f(P_1)$，并沿方向：$S_1 = \dfrac{-\nabla f(P_1)}{\|-\nabla f(P_1)\|}$

  搜索，到达点P_2 .

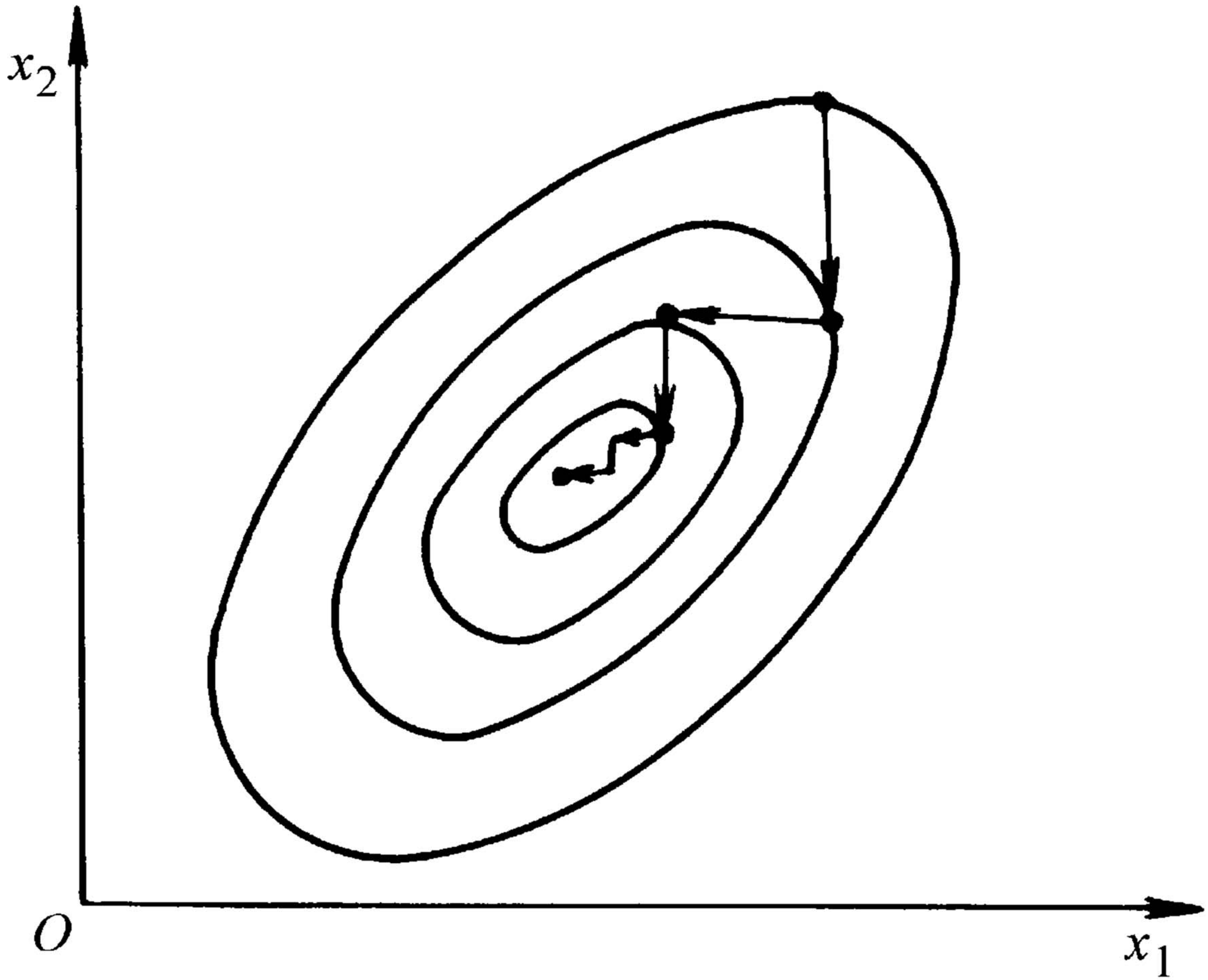
  当X满足约束$X=P_1+\gamma S_1$时，在该点取得局部极小值

- 迭代该计算过程，得到点序列$\{P_k\}_{k=0}^{\infty}$ ，

  满足$f(P_0)>f(P_1)>\ldots> f(P_k)>\ldots.$

  如果$\lim_{k\to\infty}P_k=P$，则$f(P)$是$f(X)$的一个局部极小值

梯度法搜索过程示意图



梯度方法概要

- 设$\boldsymbol{P_k}$已知
  - 求梯度向量$\nabla \boldsymbol{f(P_k)}$
  - 计算搜索方向 $$S_k = \frac{-\nabla f(P_k)}{\| -\nabla f(P_k) \|}$$
  - 在区间$[0,b]$上对$\Phi(\gamma)=f(P_k+\gamma S_k)$进行单参数极小化，$b$为一个较大值。这一过程将产生值$\gamma=h_{\min}$，它是$\Phi(\gamma)$的一个局部极小值点。

    关系式$\Phi(h_{\min})=f(P_k+h_{\min}S_k)$表明，

    它是$f(X)$沿搜索线$X=P_k+h_{\min}S_k$的一个极小值

梯度方法概要

- 构造下一个点$P_{k+1}=P_k+h_{\min}S_k$
- 进行极小化过程的终止判断，

若函数值满足$|f(P_{k+1})-f(P_k)|<\varepsilon$

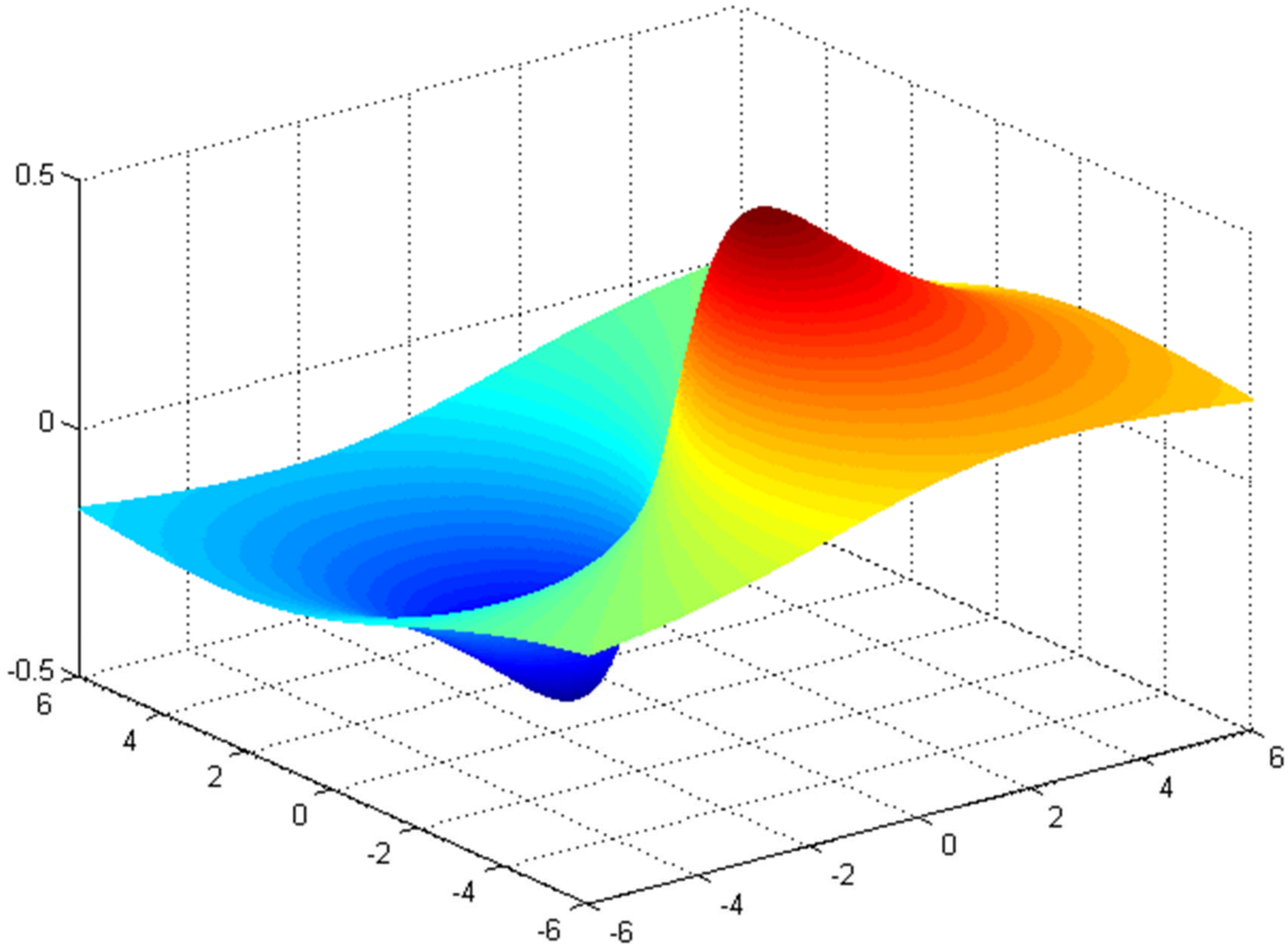
或

两点距离满足$\|P_{k+1}-P_k\|<\varepsilon$，则迭代终止，

否则转第①步

例8.9 用梯度法求函数 $f(x,y)=\dfrac{x-y}{x^2+y^2+2}$ 的$\boldsymbol{P_1}$和$\boldsymbol{P_2}$。

初始点采用$\boldsymbol{P_0}$=(-3, -2)



$\| P_{k+1} - P_k \|$ 是否足够小……

重复该过程。

例 8.9 用梯度法求函数 $f(x,y) = \dfrac{x-y}{x^2+y^2+2}$ 的 P_1 和 P_2。初始点采用 $P_0 = (-3,-2)$。

解：当 $P_0 = (-3,-2)$ 时，

$$
\begin{aligned}
S_0 &= \frac{1}{\| -\nabla f(P_0) \|}(-\nabla f(P_0)) \\
&= \frac{1}{\| -\nabla f(-3,-2) \|}(-\nabla f(-3,-2)) \\
&= (-0.4280863, 0.9037378)
\end{aligned}
$$

函数

$$
\begin{aligned}
f(P_0 + \gamma S_0) &= f((-3,-2) + \gamma(-0.4280863, 0.9037378)) \\
&= f(-3 - 0.4280863\gamma, -2 + 0.9037378\gamma) \\
&= \frac{(-3 - 0.4280863\gamma) - (-2 + 0.9037378\gamma)}{(-3 - 0.4280863\gamma)^2 + (-2 + 0.9037378\gamma)^2 + 2}
\end{aligned}
$$

在 $\gamma = h_{min_0} = 4.8186760$ 处有极小值(见程序 8.3，二次插值)。因此

$$
\begin{aligned}
P_1 &= P_0 + h_{min_0} S_0 \\
&= (-3,-2) + 4.8186760(-0.4280863, 0.9037378) \\
&= (-5.0628094, 2.3548199)
\end{aligned}
$$

函数

$$f(\boldsymbol{P}_0+\gamma\boldsymbol{S}_0)=f((-3,-2)+\gamma(-0.4280863,0.9037378))$$
$$=f(-3-0.4280863\gamma,-2+0.9037378\gamma)$$
$$=\frac{(-3-0.4280863\gamma)-(-2+0.9037378\gamma)}{(-3-0.4280863\gamma)^2+(-2+0.9037378\gamma)^2+2}$$

在 $\gamma=h_{min_0}=4.8186760$ 处有极小值(见程序 8.3,二次插值)。因此

$$\boldsymbol{P}_1=\boldsymbol{P}_0+h_{min_0}\boldsymbol{S}_0$$
$$=(-3,-2)+4.8186760(-0.4280863,0.9037378)$$
$$=(-5.0628094,2.3548199)$$

当 $\boldsymbol{P}_1=(-5.0628094,2.3548199)$ 时,

$$\boldsymbol{S}_1=\frac{1}{\|-\nabla f(\boldsymbol{P}_1)\|}(-\nabla f(\boldsymbol{P}_1))$$
$$=\frac{1}{\|-\nabla f(-5.0628094,2.3548199)\|}(-\nabla f(-5.0628094,2.3548199))$$
$$=(0.9991231,-0.0418690)$$

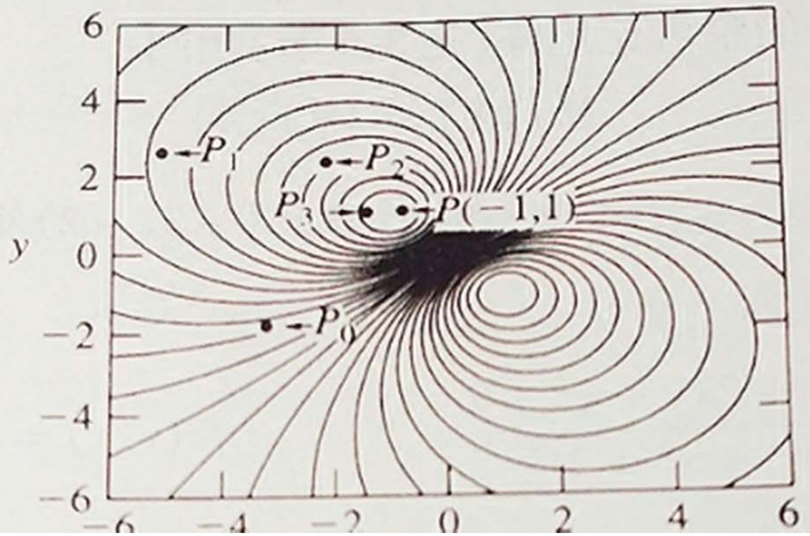
函数

$$f(\boldsymbol{P}_1+\gamma \boldsymbol{S}_1)=f((-5.0628094, 2.3548199)+\gamma(0.9991231, -0.0418690))$$
$$=f(-5.0628094+0.9991231\gamma, 2.3545199-0.0418690\gamma)$$
$$=\frac{(-5.0628094+0.9991231\gamma)-(2.3545199-0.0418690\gamma)}{(-5.0628094+0.9991231\gamma)^2+(2.3545199-0.0418690\gamma)^2+2}$$

在 $\gamma = h_{\min_1} = 2.7708281$ 处有极小值(见程序 8.3,二次插值)。因此

$$\boldsymbol{P}_2=\boldsymbol{P}_1+h_{min_1}\boldsymbol{S}_0$$
$$=(-5.0628094, 2.3548199)+2.7708281(0.9991231, -0.0418690)$$
$$=(-2.2944111, 2.2388080)$$

函数 $f(x,y)=(x-y)/(x^2+y^2+2)$ 在 $\boldsymbol{P}=(-1,1)$ 处有相对极小值。图 8.13 显示了函数 f 的一个等值线图,以及点 $\boldsymbol{P}_0,\boldsymbol{P}_1,\boldsymbol{P}_2$ 和 $\boldsymbol{P}$ 的相对位置。表 8.7 中列出了更多的计算结果。■



初始点:P_0=(-3, -2)

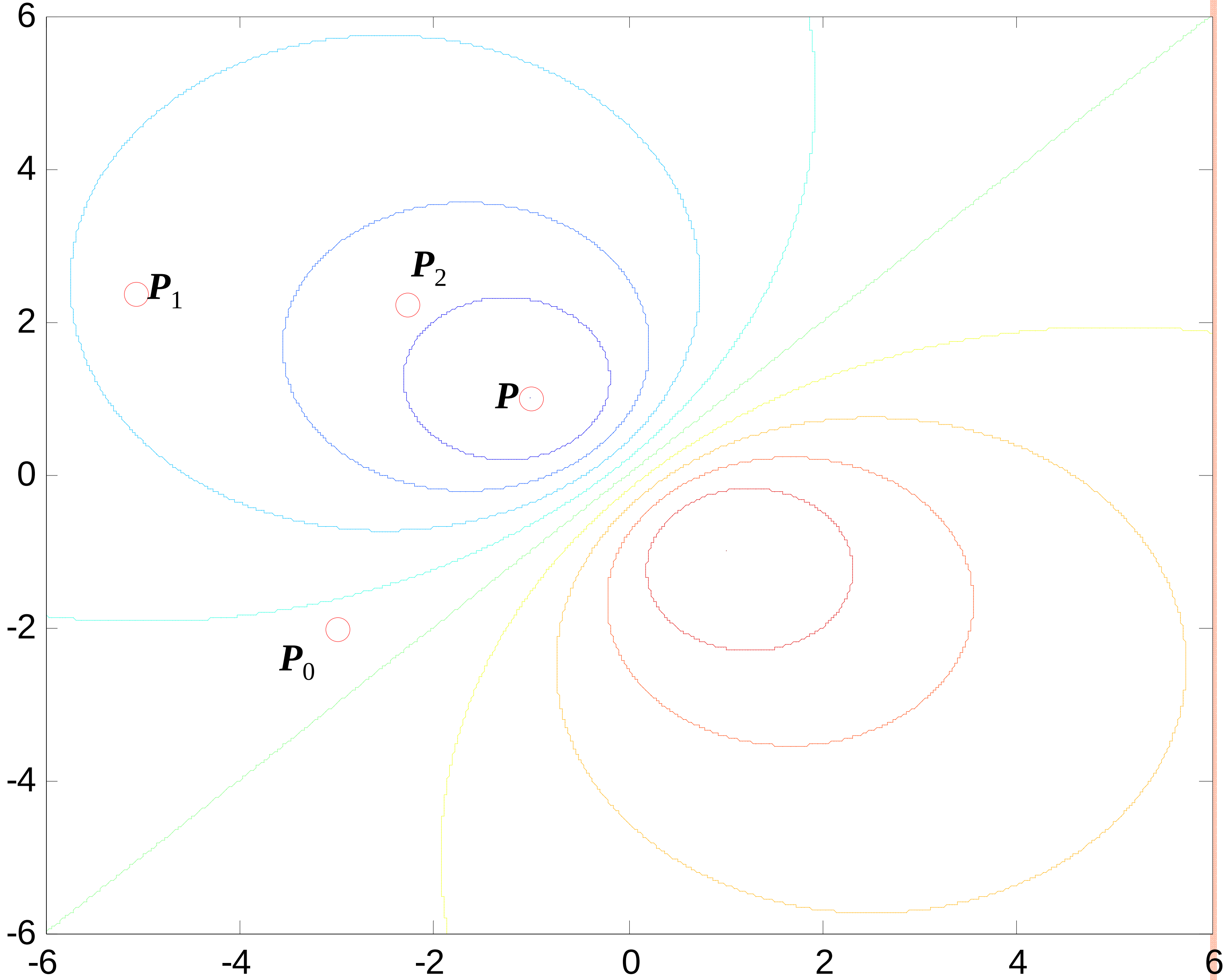


表 8.7 函数 $f(x, y) = (x - y)/(x^2 + y^2 + 2)$ 的梯度方法

k	x_k	y_k	$f(x_k, y_k)$
0	−3.0000000	−2.0000000	−0.0666667
1	−5.0628094	2.3548199	−0.2235760
2	−2.2944111	2.2388080	−0.3692574
3	−1.3879337	1.3859313	−0.4743948
4	−1.0726050	1.0724933	−0.4987762
5	−1.0035351	1.0035334	−0.4999969
6	−1.0000091	1.0000091	−0.5000000
7	−1.0000000	1.0000000	−0.5000000

梯度法小结

- 梯度法是从梯度的几何含义自然延伸得到的

  所以几何上比较直观

- 梯度法的基本思想:

  从当前点$\boldsymbol{x}_k$出发寻找使得目标函数下降最快的方向，

  即负梯度方向。

- 优点：迭代点列总是收敛的，而且计算过程简单

梯度法的缺点

- 梯度法相邻的两个搜索方向是相互垂直的，迭代点越靠近极小值点则函数下降的速度越慢
- 对于N变量函数$f(x_1,x_2,...,x_N)$而言，收敛到一个极小值的速度可能很慢
- 一般地，函数f的极小值在几何上使得值$h_{\min}$很小，这导致需要大量的极小化过程
- 梯度法一般用于最优化过程开始的几步搜索

目　录

牛顿方法原理

- 由单变量函数的二次逼近求极小值的方法推广而得，以有N个独立变量的二次多项式的极小值来近似代替目标函数f的极小值
- 对有N个独立变量的函数$z=f(x_1,x_2,\ldots,x_N)$而言，从初始点P_0开始，递归构造出一个含N个变量的二次多项式序列
- 如果目标函数是良态的，并且初始点在实际极小值附近，则该二次多项式的极小值序列将收敛到目标函数f的极小值

黑森（HESSIAN）矩阵

- 定义8.5 设$z=f(X)$是X的函数，对于$i, j=1, 2, \ldots, N$,

$$\frac{\partial^2 f(\boldsymbol{X})}{\partial x_i \partial x_j}$$

存在。f在X处的黑森矩阵是一个$N\times N$矩阵

$$\boldsymbol{H}f(\boldsymbol{X})=\left[\frac{\partial^2 f(\boldsymbol{X})}{\partial x_i \partial x_j}\right]_{N\times N}$$

其中，$i, j=1, 2, \ldots, N$

例8.10

求函数$f(x,y)=\dfrac{x-y}{x^2+y^2+2}$在点$(-3,-2)$处的黑森矩阵

解：$f_x(x,y)=\dfrac{-x^2+2xy+y^2+2}{(x^2+y^2+2)^2}$

$$f_y(x,y)=\frac{-x^2-2xy+y^2-2}{(x^2+y^2+2)^2}$$

二阶偏导数为

$$f_{xx}(x,y)=\frac{2(x^3-3x^2y-3x(y^2+2)+y(y^2+2))}{(x^2+y^2+2)^3}$$

$$f_{xy}(x,y)=\frac{2(x^3+3x^2y+x(2-3y^2)-y(y^2+2))}{(x^2+y^2+2)^3}$$

$$f_{yx}(x,y)=\frac{2(x^3+3x^2y+x(2-3y^2)-y(y^2+2))}{(x^2+y^2+2)^3}$$

$$f_{yy}(x,y)=-\frac{2(2x+x^3-6y-3x^2y-3xy^2+y^3)}{(x^2+y^2+2)^3}$$

在$(x,y)=(-3,-2)$处求黑森矩阵

$$\boldsymbol{H}f(x,y)=\begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix}$$

得

$$\boldsymbol{H}f(-3,-2)=\frac{1}{3375}\begin{bmatrix} 138 & -78 \\ -78 & -122 \end{bmatrix}$$

二阶泰勒多项式

- 定义8.6 $f(X)$在中心A处的二阶泰勒多项式为

$$Q(X)=f(\boldsymbol{A})+\nabla f(\boldsymbol{A})\bullet(\boldsymbol{X}-\boldsymbol{A})^{T}+\frac{1}{2}(\boldsymbol{X}-\boldsymbol{A})\boldsymbol{H}f(\boldsymbol{A})(\boldsymbol{X}-\boldsymbol{A})^{T}$$

- 设$z=f(x_1,x_2,\ldots,x_N)$的一阶和二阶偏导数存在，并在包含P_0的一个区间内连续，f在点P处有极小值。

则f在P_0的泰勒展开为

$$Q(X)=f(\boldsymbol{P}_0)+\nabla f(\boldsymbol{P}_0)\bullet(\boldsymbol{X}-\boldsymbol{P}_0)^{T}+\frac{1}{2}(\boldsymbol{X}-\boldsymbol{P}_0)\boldsymbol{H}f(\boldsymbol{P}_0)(\boldsymbol{X}-\boldsymbol{P}_0)^{T}$$

二阶泰勒多项式

- $Q(X)$是一个N变量的二阶多项式，它可能在 $\nabla Q(X)=0$有极小值
- 由$\nabla Q(X)=0$得

$$Q(\boldsymbol{X}) = f(\boldsymbol{P}_0) + \nabla f(\boldsymbol{P}_0) \bullet (\boldsymbol{X} - \boldsymbol{P}_0)^T + \frac{1}{2}(\boldsymbol{X} - \boldsymbol{P}_0)\boldsymbol{H}f(\boldsymbol{P}_0)(\boldsymbol{X} - \boldsymbol{P}_0)^T$$

$$\nabla Q(\boldsymbol{X}) = \nabla f(\boldsymbol{P}_0) + (\boldsymbol{X} - \boldsymbol{P}_0)(\boldsymbol{H}f(\boldsymbol{P}_0))^T = 0$$

- 若P_0在点P（f的极小值点）附近，则$H\,f(P_0)$可逆，则：可求得$X=P_0-\nabla f(P_0)((H\,f(P_0))^{-1})^{T}$

$$\nabla Q(\boldsymbol{X}) = \nabla f(\boldsymbol{P}_0) + (\boldsymbol{X} - \boldsymbol{P}_0)(\boldsymbol{H}f(\boldsymbol{P}_0))^T = 0$$

$$(\boldsymbol{X} - \boldsymbol{P}_0)(\boldsymbol{H}f(\boldsymbol{P}_0))^T = -\nabla f(\boldsymbol{P}_0)$$

$$(\boldsymbol{X} - \boldsymbol{P}_0) = -\nabla f(\boldsymbol{P}_0)((\boldsymbol{H}f(\boldsymbol{P}_0))^{-1})^T$$

可得：$\boldsymbol{X} = \boldsymbol{P}_0 - \nabla f(\boldsymbol{P}_0)((\boldsymbol{H}f(\boldsymbol{P}_0))^{-1})^T$

- 用$\boldsymbol{P}_1$代替上式中的$\boldsymbol{X}$,

$$\boldsymbol{P}_1 = \boldsymbol{P}_0 - \nabla f(\boldsymbol{P}_0)((\boldsymbol{H}f(\boldsymbol{P}_0))^{-1})^T$$

- 用$\boldsymbol{P}_{k\text{-}1}$代替$\boldsymbol{P}_0$，$\boldsymbol{P}_k$代替$\boldsymbol{P}_1$，得

$$\boldsymbol{P}_{\mathrm{k}} = \boldsymbol{P}_{\mathrm{k\text{-}1}} - \nabla f(\boldsymbol{P}_{\mathrm{k\text{-}1}})((\boldsymbol{H}f(\boldsymbol{P}_{\mathrm{k\text{-}1}}))^{-1})^T$$

$$\boldsymbol{P}_{\mathrm{k}} = \boldsymbol{P}_{\mathrm{k-1}} - \nabla f(\boldsymbol{P}_{\mathrm{k-1}})((\boldsymbol{H}f(\boldsymbol{P}_{\mathrm{k-1}}))^{-1})^{T}$$

- 上式建立了一个迭代关系式，由此迭代关系式可生成点序列$\{\boldsymbol{P}_k\}(k=0,1,2,\ldots)$
- 但算法中涉及到矩阵求逆的运算，理论上虽然可行，实际上用它进行计算则效率很低

改进的牛顿方法概要

- 设$\boldsymbol{P}_k$已知

  - 计算搜索方向

$$S_k = -\nabla f(\boldsymbol{P}_{\text{k-1}})((\boldsymbol{H}f(\boldsymbol{P}_{\text{k-1}}))^{-1})^T$$

  - 在区间[0, b]上对$\Phi(\gamma)=f(P_k+\gamma S_k)$进行单变量极小化，$b$为较大值。 得到值$\gamma=h_{min}$

它是$\Phi(\gamma)$的极小值点。

关系式$\Phi(h_{min})=f(P_k+h_{min}S_k)$表明，

它是$f(X)$沿搜索线$X=P_k+h_{min}S_k$的一个极小值

改进的牛顿方法概要

- 构造下一点，$P_{k+1}=P_k+h_{min}S_k$
- 进行终止条件测试，$\| f(P_{k+1})-f(P_k)\| < \varepsilon$

  $\| P_{k+1}-P_k \| < \varepsilon$

  即函数值$f(P_k)$和$f(P_{k+1})$是否足够相近，

  距离$||P_{k+1}-P_k||$是否足够小？

- 重复上述过程

无约束优化方法——直接法总结

Nelder-Mead法（单纯形法）

思路清楚，收敛慢

无约束优化方法——间接法总结

1 梯度法

- 方向： 负梯度，用到一阶导数
- 适合于精度不高或用于复杂函数寻找一个好的初始点

2 牛顿法

- 用到一阶导数和海色矩阵，具有二次收敛性
- 要求海色矩阵非奇异，且维数不宜太高

作业

P342　　8.2.4　2

P352　　4,5,6

报告：单纯形法程序，梯度法程序，牛顿法程序

要求：每个程序分为测试主程序**main**和具体函数程序，**main**函数里面包含测试算例所有内容。

比如牛顿法程序：**main.m**，**Newton.m**