《数值计算方法》课程

最优化

(基于梯度的优化)

胡建芳

（研究方向：计算机视觉）

http://sdcs.sysu.edu.cn/content/5143

计算机学院

课程回顾

- **数值微分，数值积分：**

即
$$I=\int_a^b f(x)\mathrm{d}x \approx \sum_{k=0}^{n} A_k f(x_k)=I_n$$

$$(A_k=\int_a^b l_k(x)\mathrm{d}x \quad k=0,1,\cdots,n)$$

由上式确定系数的公式称为插值型求积公式。

节点等距的时候

则 $A_k=(b-a)C_k^{(n)}$，于是得求积公式

$$I_n=(b-a)\sum_{k=0}^{n} C_k^{(n)} f(x_k)$$

称为n 阶牛顿-柯特斯 (Newton-Cotes)公式，$C_k^{(n)}$ 称为柯特斯系数。

课程回顾

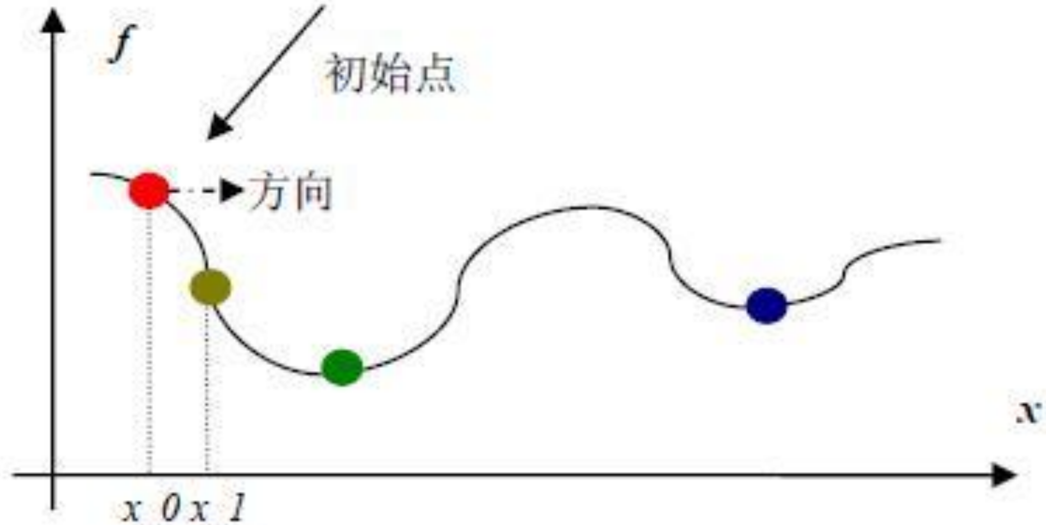
常用的柯特斯系数表

n	$C_k^{(n)}$						
1	1/2	1/2					
2	1/6	4/6	1/6				
3	1/8	3/8	3/8	1/8			
4	7/90	32/90	12/90	32/90	7/90		
5	19/288	75/288	50/288	50/288	75/288	19/288	
6	41/840	216/840	27/840	272/840	27/840	216/840	41/840

还有其它内容，参见详细课程PPT......

最优化（无约束条件）：

- **基于梯度优化方法：**
  牛顿法
  梯度下降法
  随机梯度下降法



- **梯度无关的优化方法**
  黄金分割搜索
  持续抛物插值

不考试，研究生学习的基础知识

最优化（无约束条件）：

求解无约束最优化问题的基本思想

标准形式：

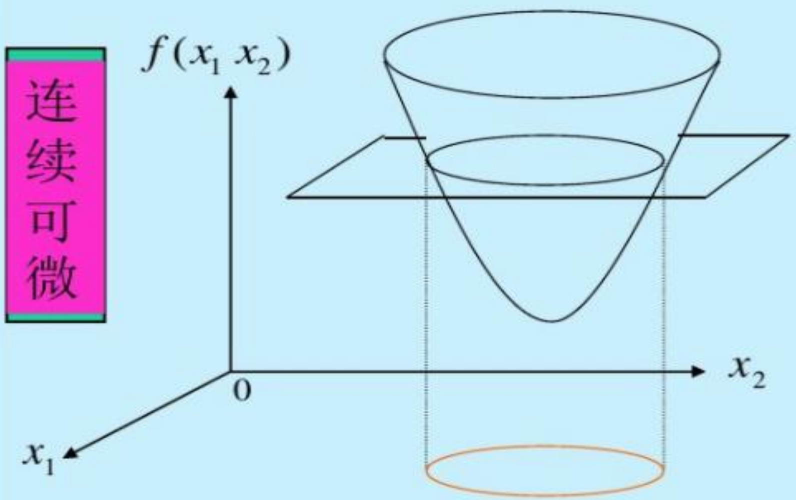
$$\min_{X \in E^n} f(X)$$

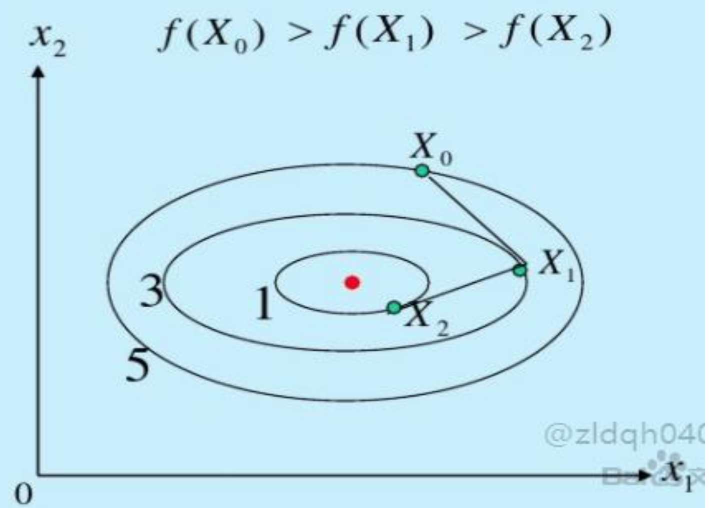
其中 $f: E^n \to E^1$

$$\max_{X \in E^n} f(X) = \min_{X \in E^n} [-f(X)]$$

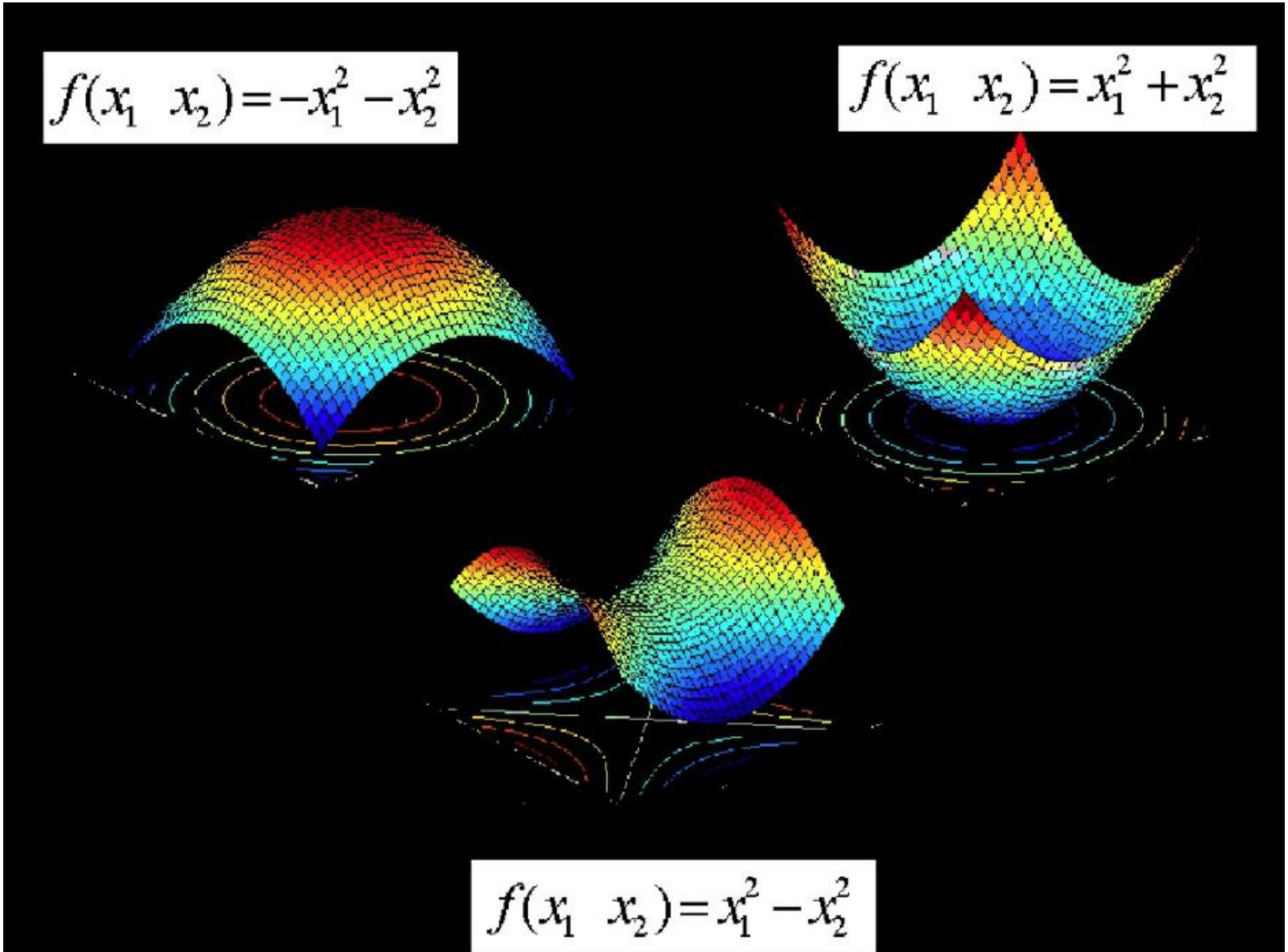
求解的基本思想（以二元函数为例）

连续可微

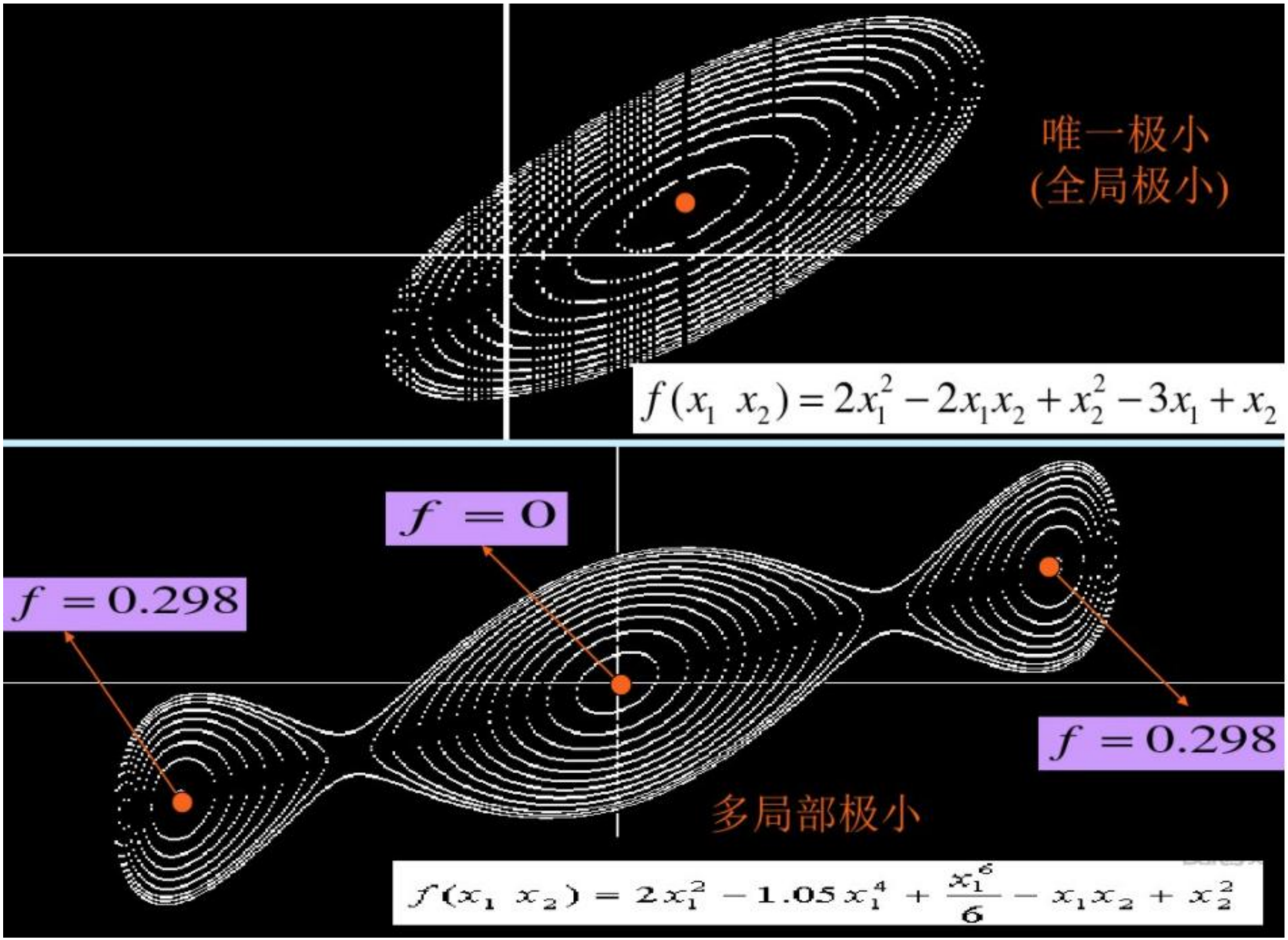




最优化（无约束条件）：



最优化（无约束条件）：



唯一极小
(全局极小)
$f(x_1\ x_2)=2x_1^2-2x_1x_2+x_2^2-3x_1+x_2$
$f=0$
$f=0.298$
$f=0.298$
多局部极小
$f(x_1\ x_2)=2x_1^2-1.05x_1^4+\frac{x_1^6}{6}-x_1x_2+x_2^2$

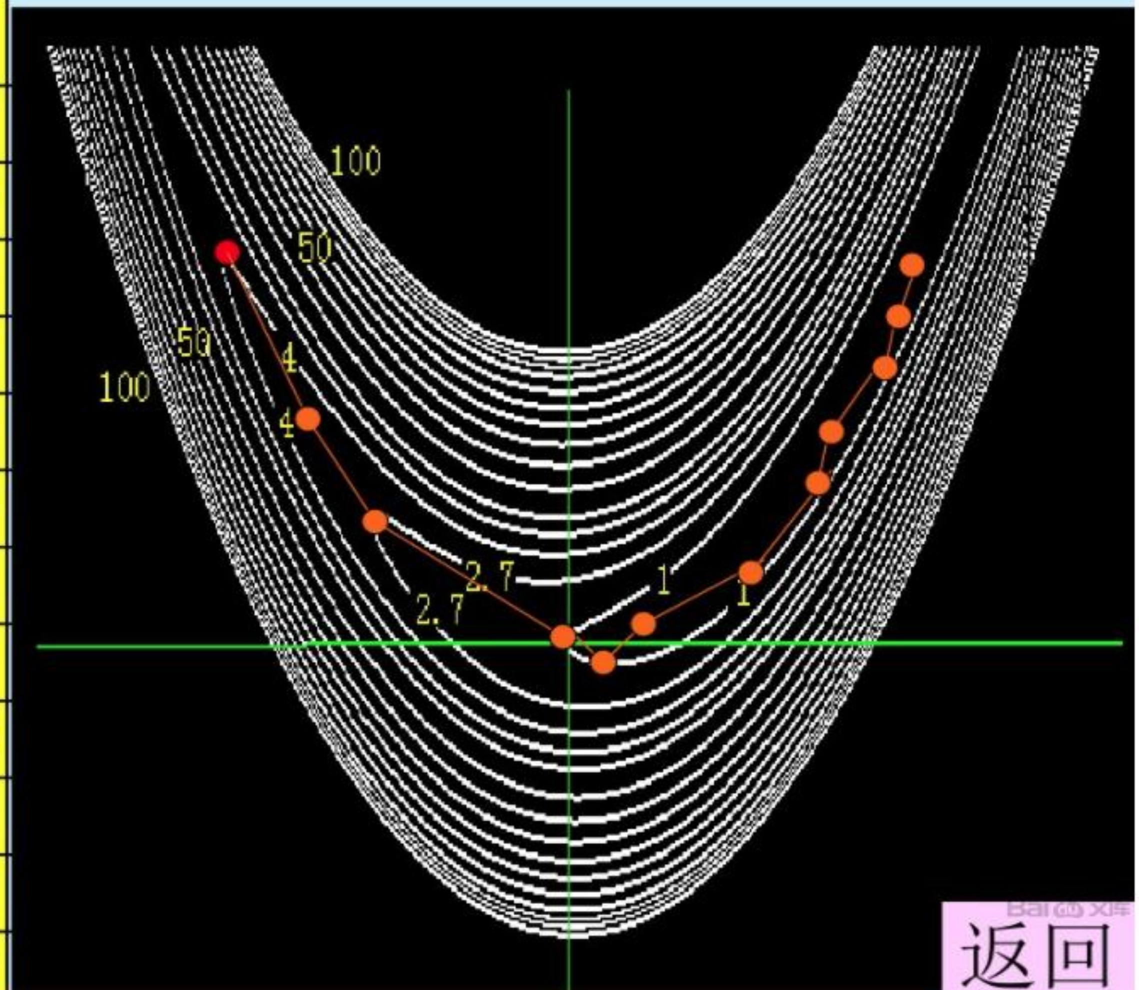
最优化（无约束条件）：

搜索过程

$$\min \; f(x_1 \;\; x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

最优点 (1　1)
初始点 (-1　1)

x_1	x_2	f
-1	1	4.00
-0.79	0.58	3.39
-0.53	0.23	2.60
-0.18	0.00	1.50
0.09	-0.03	0.98
0.37	0.11	0.47
0.59	0.33	0.20
0.80	0.63	0.05
0.95	0.90	0.003
0.99	0.99	1E-4
0.999	0.998	1E-5
0.9997	0.9998	1E-8



返回

基于梯度优化方法：

- **牛顿法：** **求方程：一阶导数=0**

(1) 选定初始点 $X^0 \in E^n$，给定允许误差 $\varepsilon > 0$，令 k=0;

(2) 求 $\nabla f\left(X^k\right), \left(\nabla^2 f\left(X^k\right)\right)^{-1}$，检验：若 $\left\|\nabla f\left(X^k\right)\right\| < \varepsilon$，则

停止迭代，$X^* \approx X^k$．否则，转向(3)；

(3) 令 $S^k = -[\nabla^2 f\left(X^k\right)]^{-1} \nabla f\left(X^k\right)$（牛顿方向）；

(4) $X^{k+1} = X^k + S^k$，$k = k + 1$，转回(2)．

如果f是对称正定矩阵A的二次函数，则用牛顿法经过**一次迭代**就可达到最优点，如不是二次函数，则牛顿法不能一步达到极值点，但由于这种函数在极值点附近和二次函数很近似，因此牛顿法的收敛速度还是很快的．

牛顿法的收敛速度虽然较快，但要求Hessian矩阵要可逆，要计算二阶导数和逆矩阵，就加大了计算机计算量和存储量．

基于梯度优化方法：

- **拟牛顿法：**

为克服牛顿法的缺点，同时保持较快收敛速度的优点，利用第 k 步和第 k+1 步得到的X^k，X^{k+1}，$\nabla f(X^k)$，$\nabla f(X^{k+1})$，构造一个正定矩阵G^{k+1}近似代替$\nabla^2 f(X^k)$，或用H^{k+1}近似代替$(\nabla^2 f(X^k))^{-1}$，将牛顿方向改为：

$$G^{k+1} S^{k+1} = -\nabla f(X^{k+1}),\quad S^{k+1} = -H^{k+1} \nabla f(X^{k+1})$$

从而得到下降方向.

https://blog.csdn.net/songbinxu/article/details/79677948

基于梯度优化方法：

梯度下降法：

通常采用迭代法计算G^{k+1}，H^{k+1}，迭代公式为：

BFGS（Boryden-Fletcher-Goldfarb-Shanno）公式

$$G^{k+1}=G^{k}+\frac{\Delta f^{k}(\Delta f^{k})^{T}}{(\Delta f^{k})^{T}\Delta x^{k}}-\frac{G^{k}\Delta x^{k}(\Delta x^{k})^{T}G^{k}}{(\Delta x^{k})^{T}G^{k}\Delta x^{k}}$$

$$H^{k+1}=H^{k}+\left(1+\frac{(\Delta f^{k})^{T}H^{k}\Delta f^{k}}{(\Delta f^{k})^{T}\Delta x^{k}}\right)\frac{\Delta x^{k}(\Delta x^{k})^{T}}{(\Delta f^{k})^{T}\Delta x^{k}}-\frac{\Delta x^{k}(\Delta f^{k})^{T}H^{k}-H^{k}\Delta f^{k}(\Delta x^{k})^{T}}{(\Delta f^{k})^{T}\Delta x^{k}}$$

基于梯度优化方法：

- **梯度下降法：**

DFP（Davidon-Fletcher-Powell）公式：

$$G^{k+1} = G^k + \left(1 + \frac{(\Delta X^k)^T G^k \Delta X^k}{(\Delta X^k)^T \Delta f^k}\right)\frac{\Delta f^k (\Delta f^k)^T}{(\Delta f^k)^T \Delta X^k} - \frac{\Delta f^k (\Delta X^k)^T G^k - G^k \Delta X^k (\Delta f^k)^T}{(\Delta X^k)^T \Delta f^k}$$

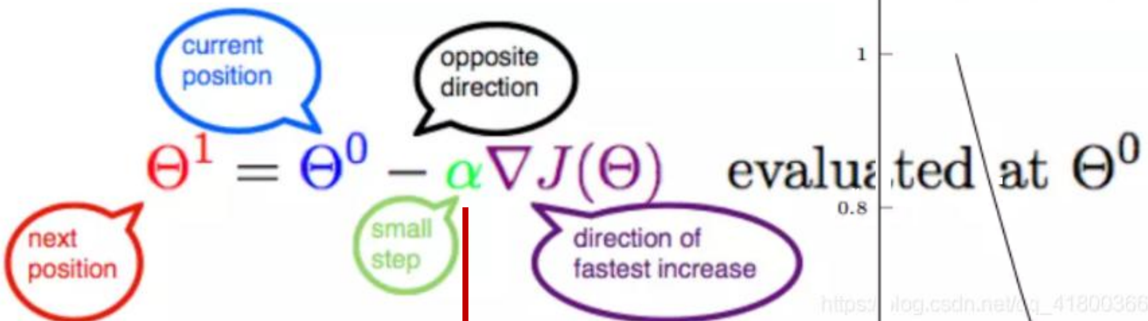
$$H^{k+1} = H^k + \frac{\Delta X^k (\Delta X^k)^T}{(\Delta f^k)^T \Delta X^k} - \frac{H^k \Delta f^k (\Delta f^k)^T H^k}{(\Delta f^k)^T H^k \Delta f^k}$$

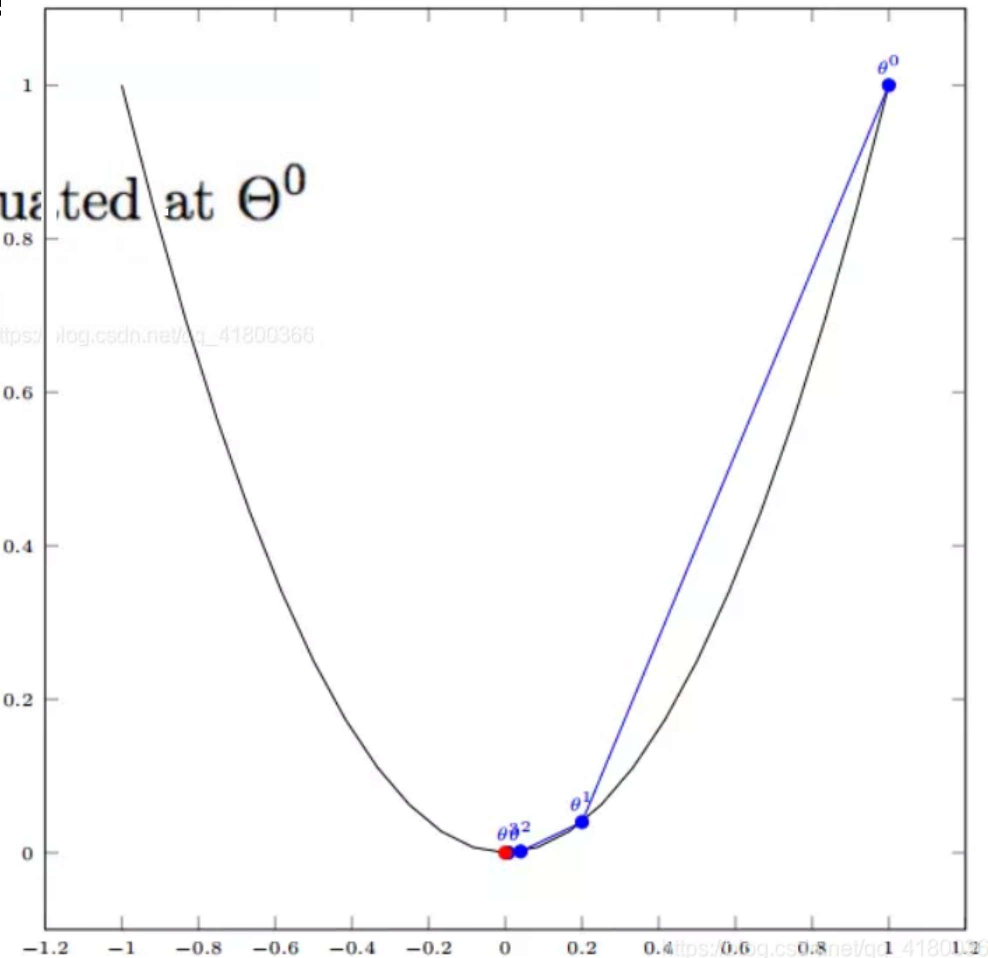
计算时可置$H^1 = I$（单位阵），对于给出的X^1利用上面的公式进行递推. 这种方法称为**拟牛顿法**.

基于梯度优化方法：

- **梯度下降法：**

  沿着负梯度方向，更新参数





步长，调参
核心之一
可以搜索

形式简单，计算快
收敛速度较慢

基于梯度优化方法：

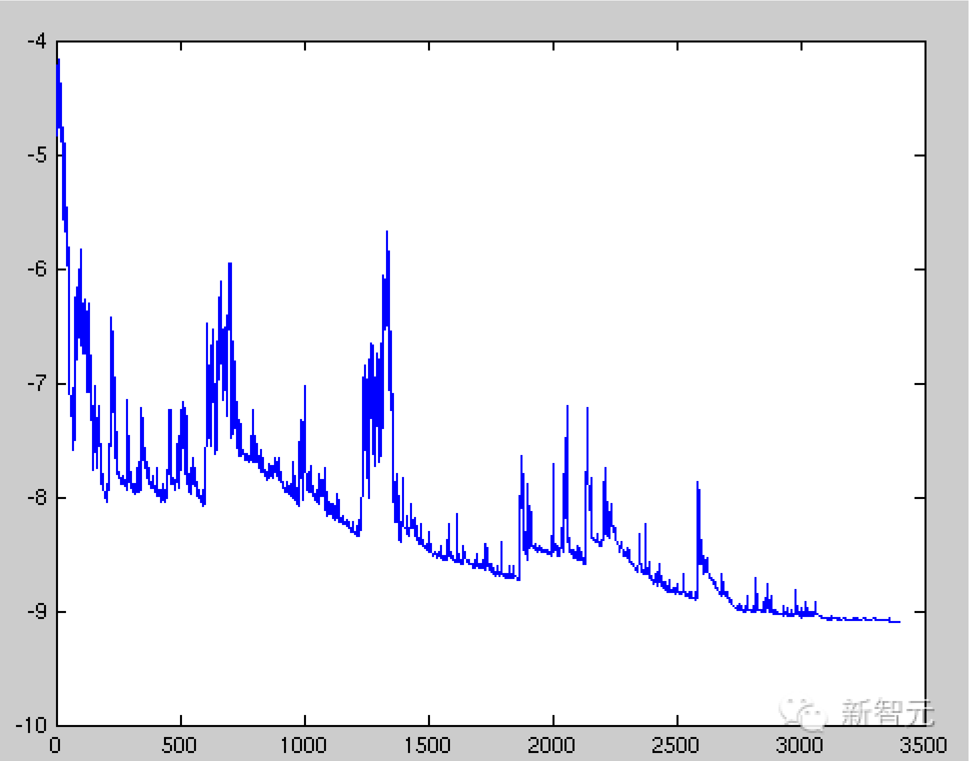
- **随机梯度下降法：**

  随机梯度下降算法每次从训练集中随机选择一部分样本来进行优化，即： $\theta=\theta-\eta\cdot\nabla\theta\ J(\theta;xi;yi)$

波动较大：容易跳出比较差的局部最小值

形式简单，计算快

会收敛吗?

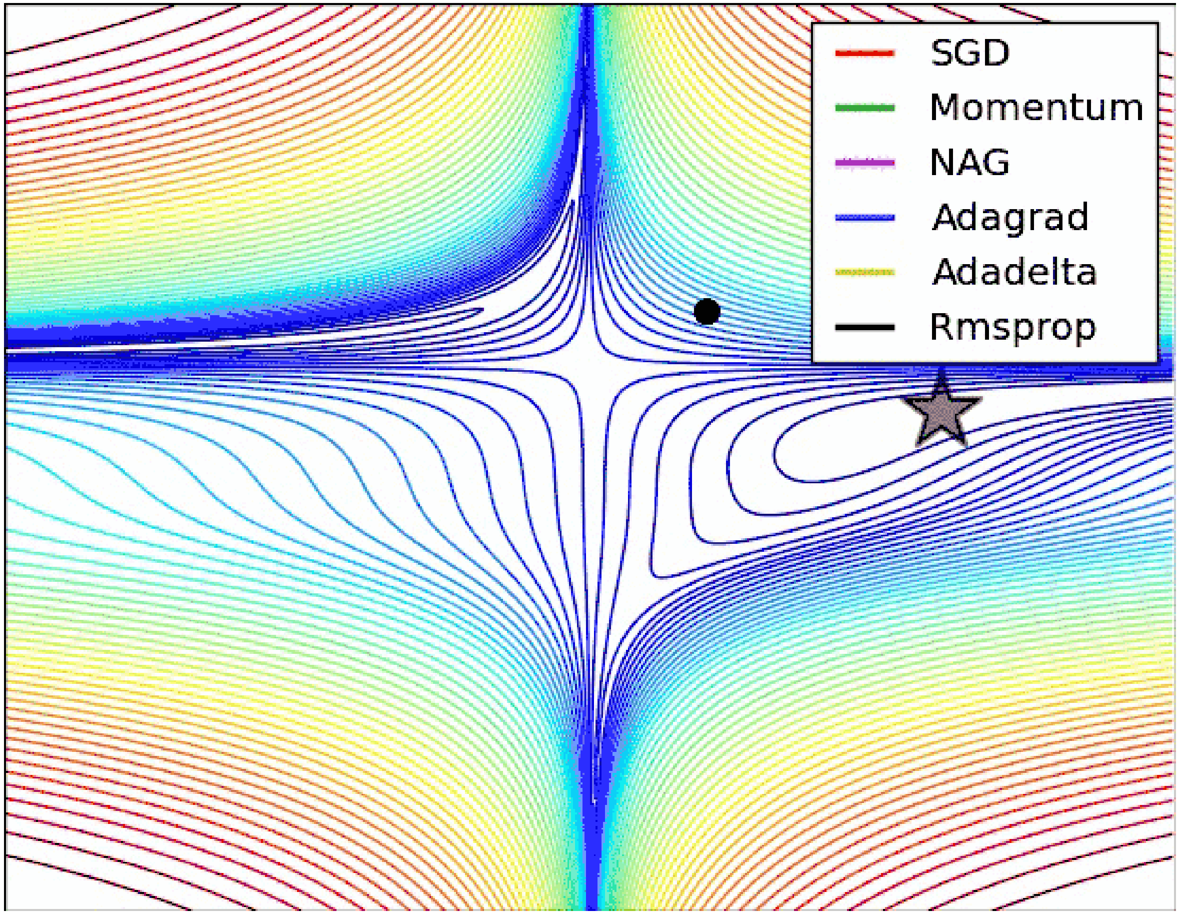


基于梯度优化方法：

- **动量（momentum）：**

保持之前运动的势能

$$\theta_{n+1} = \theta_n + v_{n+1}$$
$$v_{n+1} = \beta v_n - \alpha \nabla_\theta L(\theta_n)$$



1. 加速收敛
2. 提高精度(减少收敛过程中的振荡)

还有很多其它方法……

梯度无关优化方法：

- **黄金分割搜索：**

  逐步缩小搜索区间，直至极小值点存在的区间达到允许的误差范围

设一元函数$\boldsymbol{f}(\boldsymbol{a})$的起始搜索区间为$[\boldsymbol{a},\boldsymbol{b}]$，$\alpha^*$是函数的极小点。

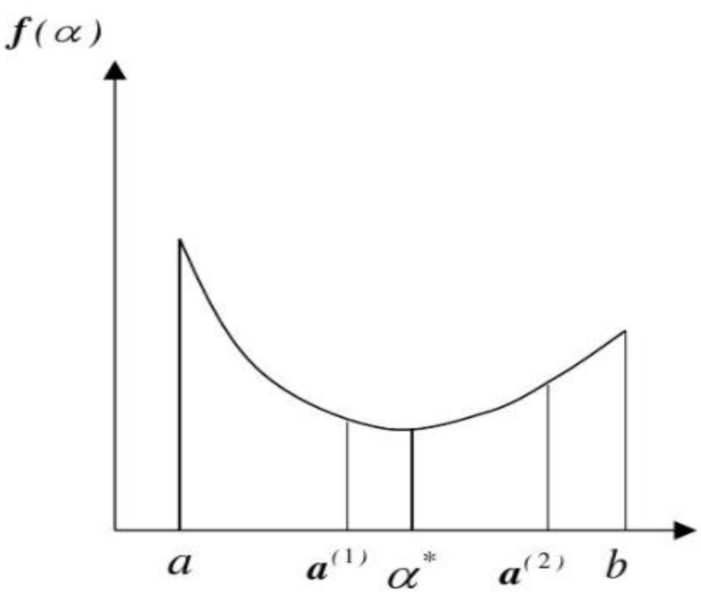
在搜索区间$[\boldsymbol{a},\boldsymbol{b}]$内任取两点$\alpha^{(1)}$、$\alpha^{(2)}$。且$\boldsymbol{a}<\alpha^{(1)}<\alpha^{(2)}<\boldsymbol{b}$，计算$\boldsymbol{f}(\alpha^{(1)})$、$\boldsymbol{f}(\alpha^{(2)})$。将$\boldsymbol{f}(\alpha^{(1)})$与$\boldsymbol{f}(\alpha^{(2)})$进行比较，可能出现三种情况：

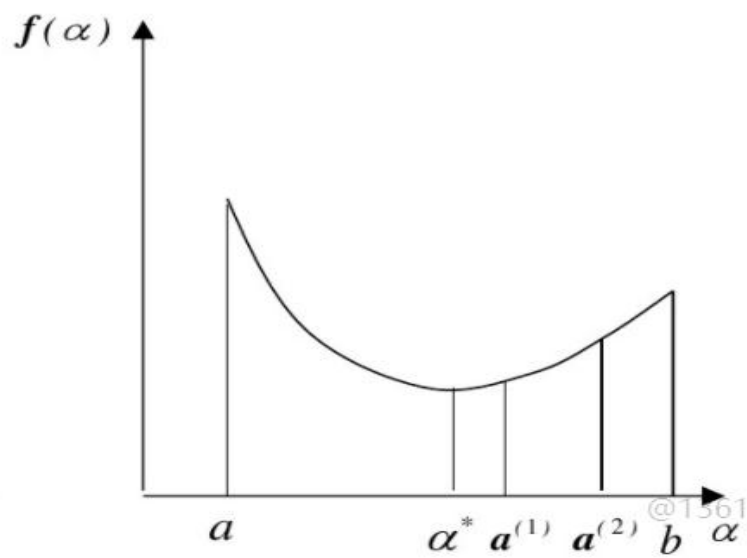
梯度无关优化方法：

- **黄金分割搜索：**

  逐步缩小搜索区间，直至极小值点存在的区间达到允许的误差范围

  (1) $f(\alpha^{(1)}) < f(\alpha^{(2)})$.在这种情况下，可以丢掉$(\alpha^{(2)}, \boldsymbol{b}]$部分，而最小点必定在$[\boldsymbol{a}, \alpha^{(2)}]$内。



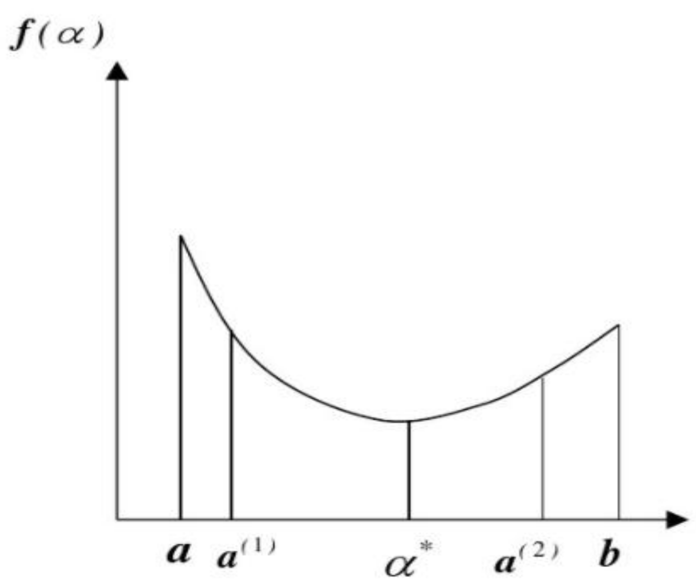


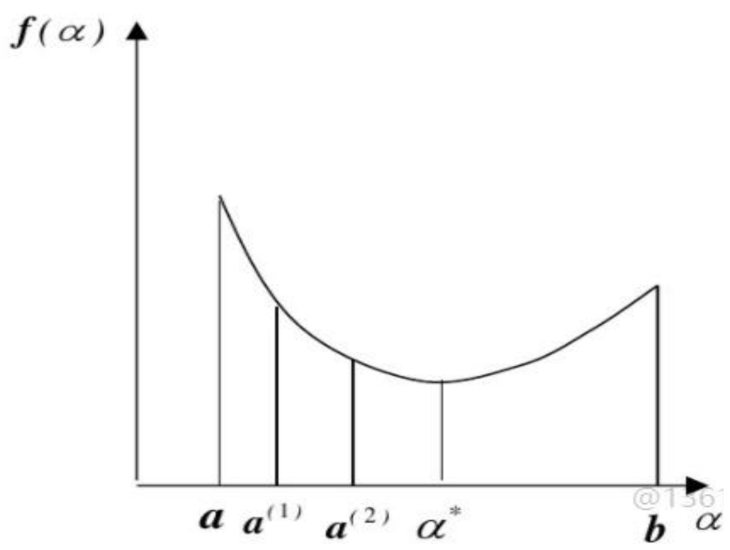
梯度无关优化方法：

- **黄金分割搜索：**

  逐步缩小搜索区间，直至极小值点存在的区间达到允许的误差范围

  (2) $f(\alpha^{(1)}) > f(\alpha^{(2)})$.在这种情况下，可以丢掉$[a,\alpha^{(1)})$部分，而最小点必定在$[\alpha^{(1)},b]$内。



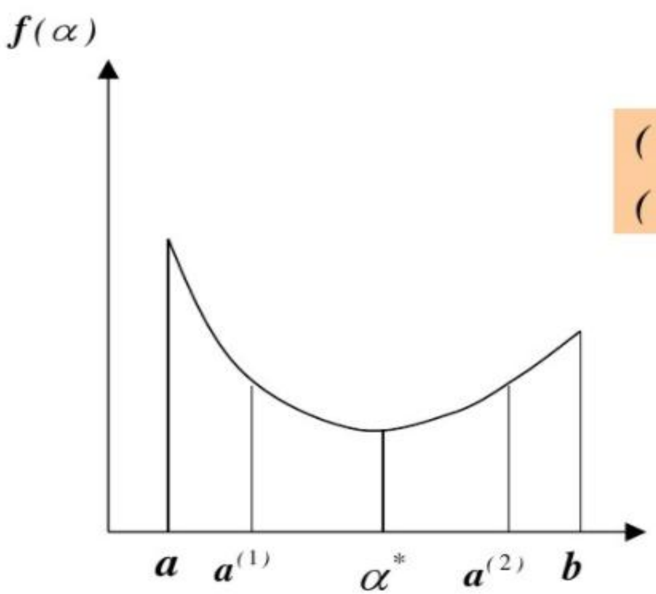


梯度无关优化方法：

- **黄金分割搜索：**

  逐步缩小搜索区间，直至极小值点存在的区间达到允许的误差范围

(3) $f(\alpha^{(1)})=f(\alpha^{(2)})$.在这种情况下，可以丢掉$[a,\alpha^{(1)})$部分，也可以丢掉$(\alpha^{(2)},b]$部分，而最小点必定在$[\alpha^{(1)},\alpha^{(2)}]$内。因此这种情况可以并入上面的任意一种情况。



(1) $f(\alpha^{(1)})\leq f(\alpha^{(2)})$.取区间$[a,\alpha^{(2)}]$;

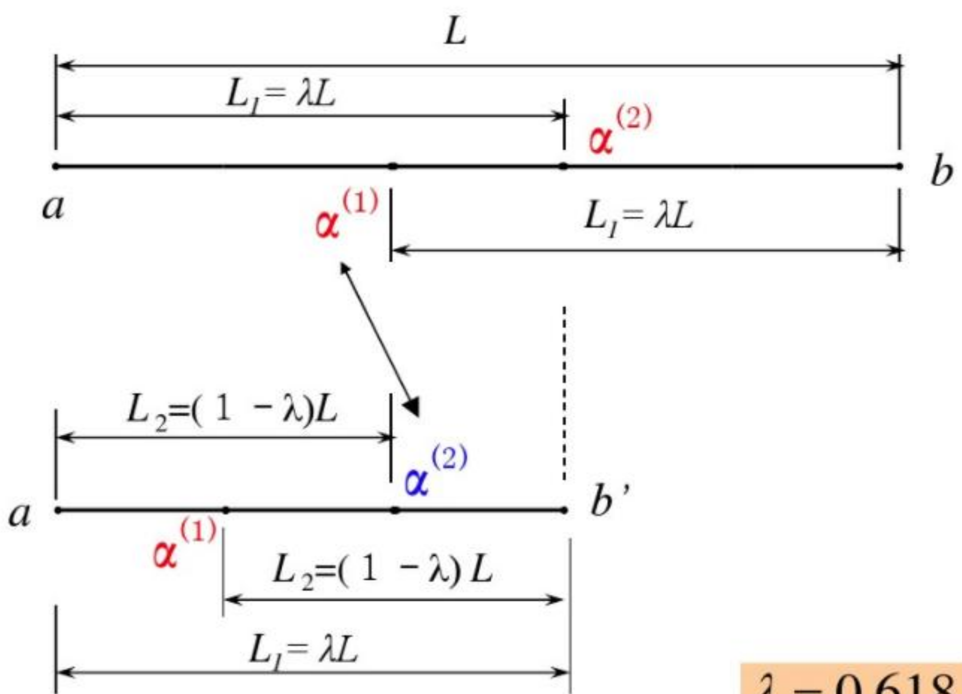
(2) $f(\alpha^{(1)})>f(\alpha^{(2)})$.取区间$[\alpha^{(1)},b]$。

梯度无关优化方法：

- **黄金分割搜索：**

  逐步缩小搜索区间，直至极小值点存在的区间达到允许的误差范围

  1. $\alpha^{(1)},\alpha^{(2)}$在[a, b]中位置对称
  2. 每次缩短的区间缩短率不变，减少计算量。



$$\frac{L_2}{L_1}=\frac{L_1}{L}$$

$$\Downarrow$$

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

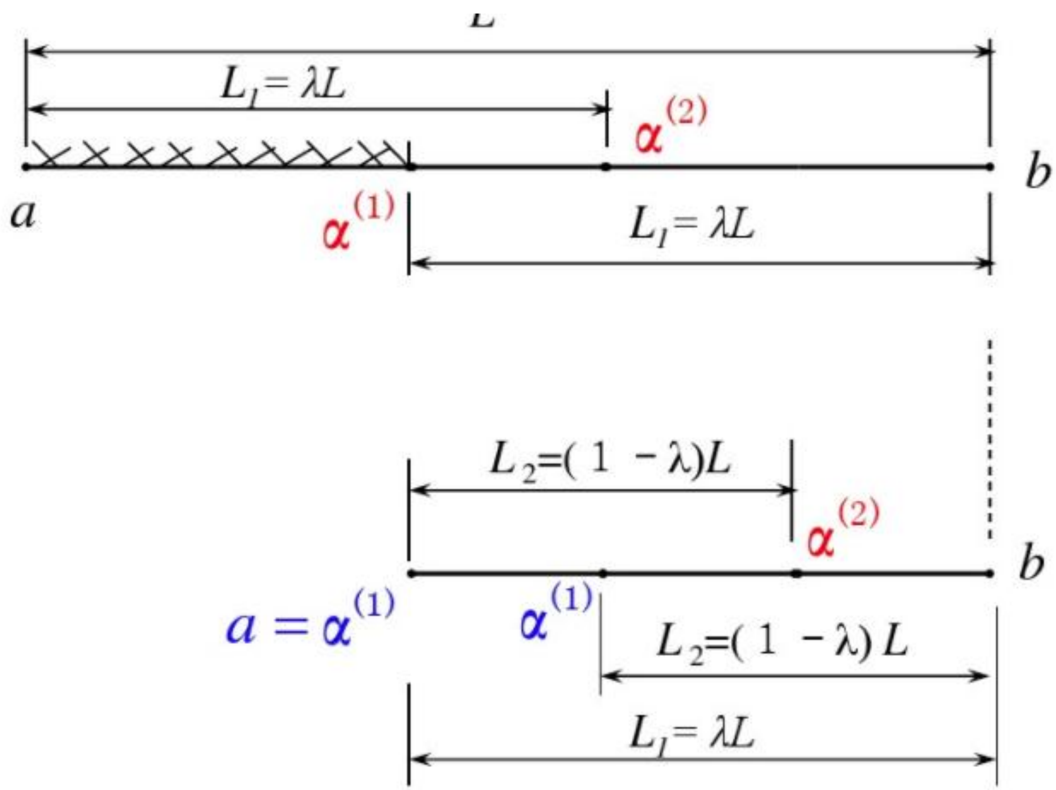
$$\Downarrow$$

$$\lambda^2+\lambda-1=0$$

$$\lambda=0.618$$

梯度无关优化方法：

- **黄金分割搜索：**
  逐步缩小搜索区间，直至极小值点存在的区间达到允许的误差范围



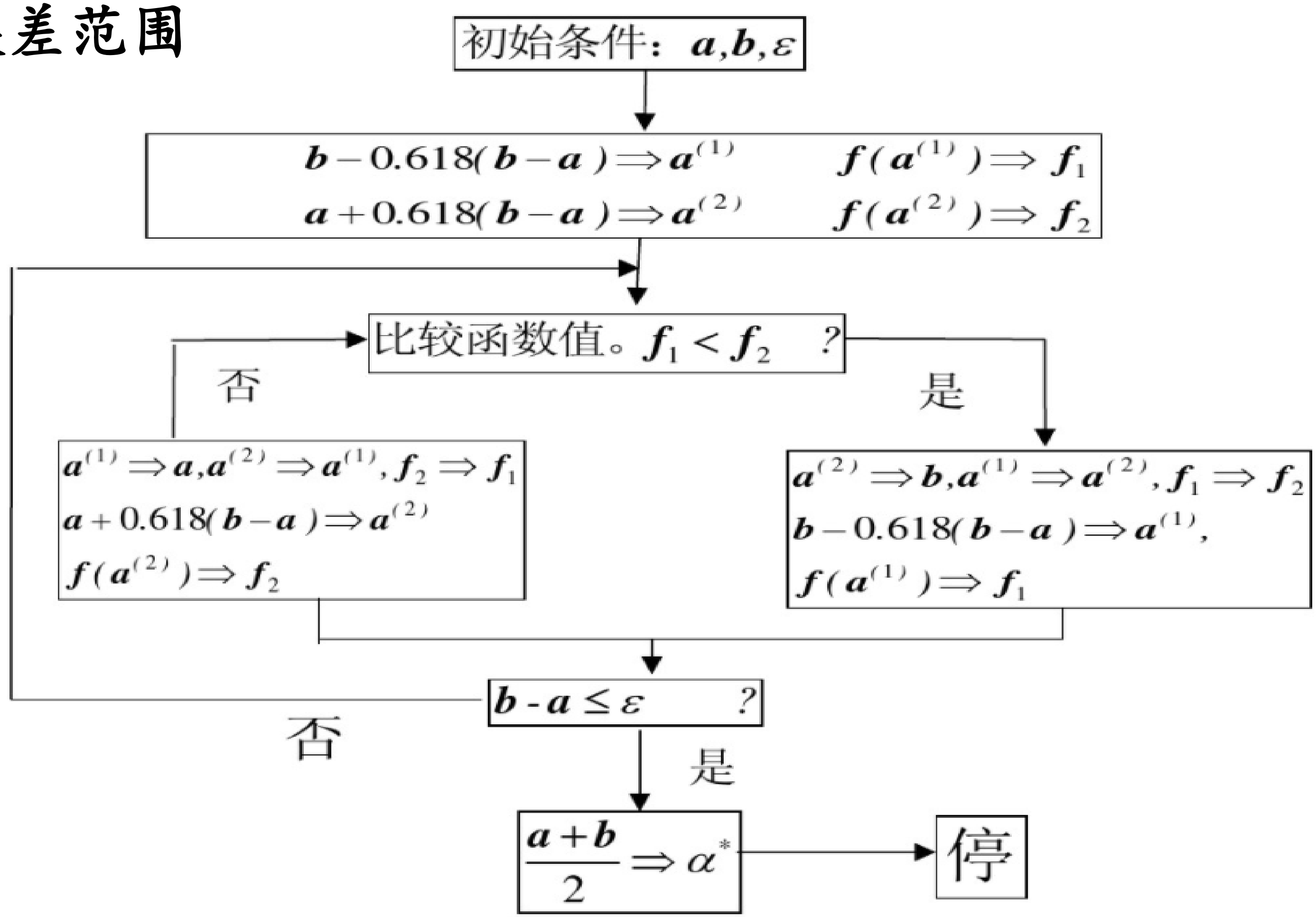
$\lambda = 0.618$

梯度无关优化方法：

- **黄金分割搜索：**

  逐步缩小搜索区间，直至极小值点存在的区间达到允许的误差范围

黄金分割法计算框图



基于梯度优化方法：

- **抛物插值法：**

设函数在$[a,b]$内呈现“大一小一大”单峰变化

在$[a,b]$上以低次（二次或三次）插值多项式P(x)

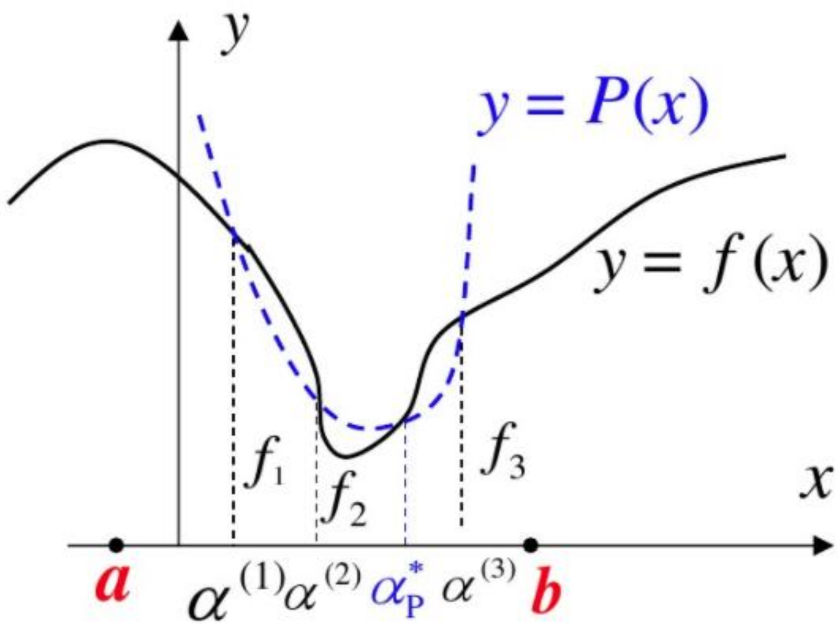
来逼近原目标函数，求得多项式的极值点，

逐步缩短搜速区间。反复计算，直至给定精度。

此法应用较广。

基于梯度优化方法：

- **抛物插值法：**



在[a, b]中设定三点$\alpha^{(1)}$，$\alpha^{(2)}$和$\alpha^{(3)}$

$$\boldsymbol{a} \le \alpha^{(1)} \le \alpha^{(2)} \le \alpha^{(3)} \le \boldsymbol{b}$$

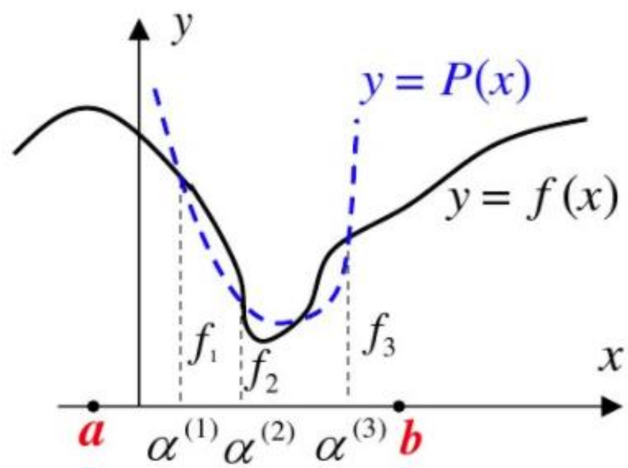
对应有 f_1, f_2, f_3，且$f_1 > f_2 < f_3$

基于梯度优化方法：

抛物插值法：

构造二次多项式

$$P(x)=a_0+a_1x+a_2x^2$$



由插值条件确定待定的a_i(i=0,1,2)

$$\left.\begin{aligned}P(\alpha^{(1)})&=a_0+a_1\alpha^{(1)}+a_2(\alpha^{(1)})^2=f_1\\P(\alpha^{(2)})&=a_0+a_1\alpha^{(2)}+a_2(\alpha^{(2)})^2=f_2\\P(\alpha^{(3)})&=a_0+a_1\alpha^{(3)}+a_2(\alpha^{(3)})^2=f_3\end{aligned}\right\}\cdots(1)$$

基于梯度优化方法：

- **抛物插值法：**

求$\mathrm{P}(x)$极小点， $\mathrm{P}(x)=a_1+2a_2x=0$

$$\Rightarrow \alpha_{\mathrm{P}}^{*}=-\frac{a_1}{2a_2}$$

$$\alpha_{\mathrm{P}}^{*}=\frac{1}{2}\left(\alpha^{(1)}+\alpha^{(3)}-\frac{\mathrm{C}_1}{\mathrm{C}_2}\right)$$

其中， $\mathrm{C}_1=\dfrac{f_3-f_1}{\alpha^{(3)}-\alpha^{(1)}}$

$$\mathrm{C}_2=\frac{(f_2-f_1)/(\alpha^{(2)}-\alpha^{(1)})-\mathrm{C}_1}{\alpha^{(2)}-\alpha^{(3)}}$$

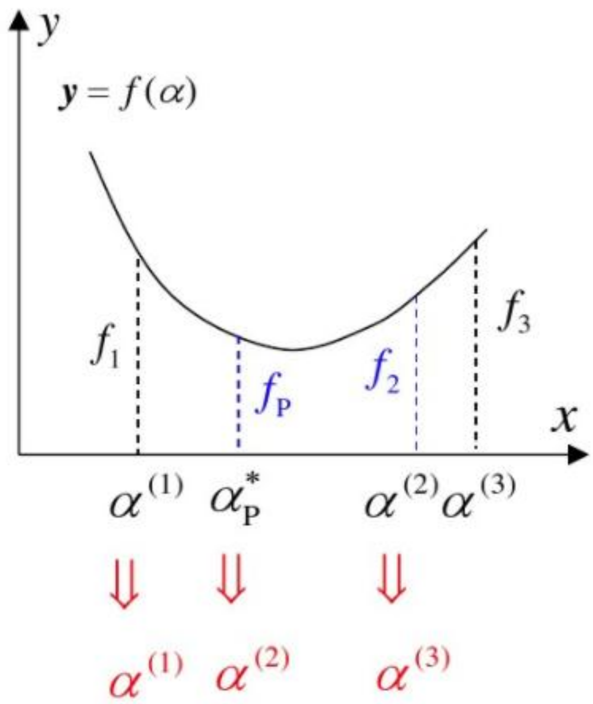
基于梯度优化方法：

- **抛物插值法：**

（1） $\alpha_P^* < \alpha^{(2)}$， $f_P < f_2$

$$\alpha^{(1)} = \alpha^{(1)}, \alpha^{(3)} = \alpha^{(2)}, \alpha^{(2)} = \alpha_P^*$$

$$f_1 = f_1\ ,\quad f_3 = f_2\ ,\quad f_2 = f_P$$



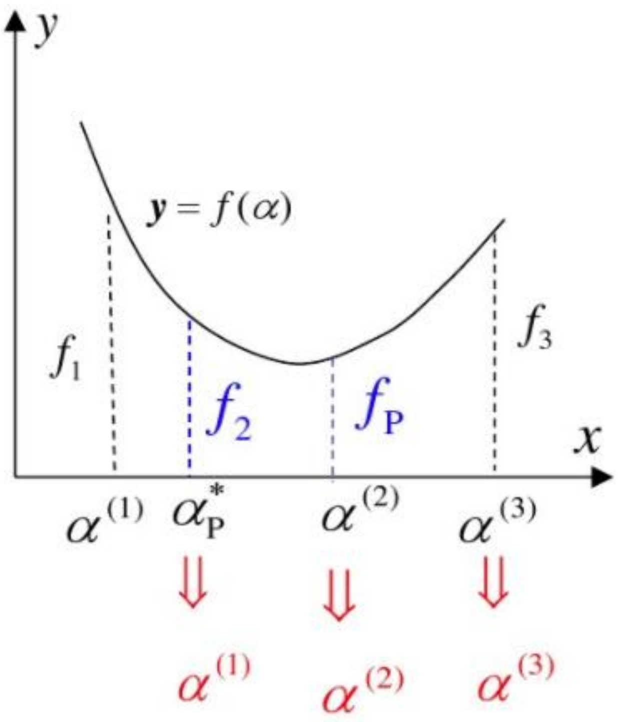
基于梯度优化方法：

- **抛物插值法：**

（2） $\alpha_P^* > \alpha^{(2)}$， $f_P < f_2$

$$\alpha^{(1)} = \alpha_P^*, \alpha^{(2)} = \alpha^{(2)}, \alpha^{(2)} = \alpha^{(3)}$$

$$f_1 = f_P \ , \quad f_2 = f_2 \ , \quad f_3 = f_3$$



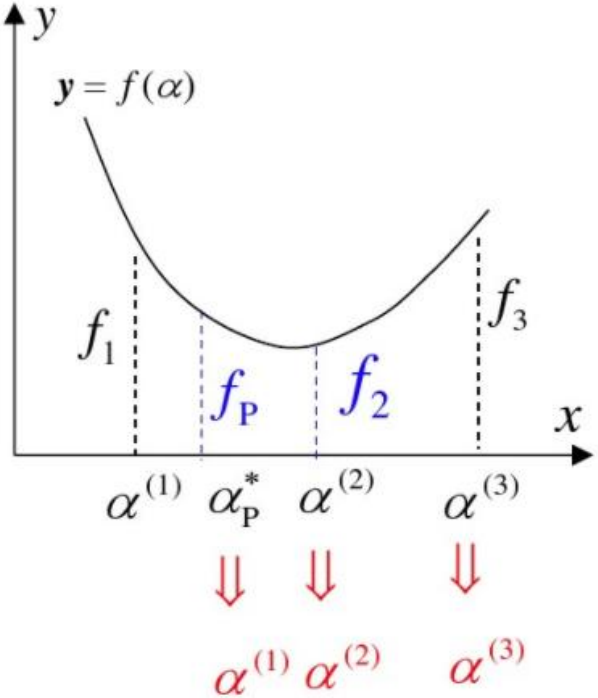
基于梯度优化方法：

- **抛物插值法：**

(3) $\alpha_P^* < \alpha^{(2)}$，$f_P > f_2$

$$\alpha^{(1)} = \alpha_P^*, \alpha^{(2)} = \alpha^{(2)}, \alpha^{(3)} = \alpha^{(3)}$$

$$f_1 = f_P\ ,\quad f_3 = f_2\ ,\quad f_3 = f_3$$



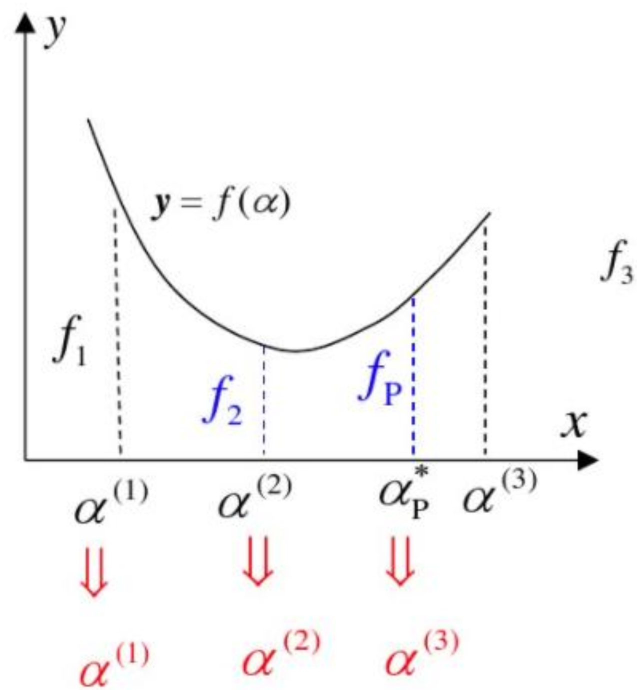
基于梯度优化方法：

- **抛物插值法：**

（4）$\alpha_P^* > \alpha^{(2)}$，$f_P > f_2$

$$\alpha^{(1)} = \alpha^{(1)}, \alpha^{(2)} = \alpha^{(2)}, \alpha^{(3)} = \alpha_P^*$$

$$f_1 = f_1\ ,\quad f_2 = f_2\ ,\quad f_3 = f_P$$



基于梯度优化方法：

- **抛物插值法：**

比较α_P^*与$\alpha^{(2)}$两点，$f(x)$的大小；缩短搜索区间。

（1） $\alpha_P^* < \alpha^{(2)}$， $f_P < f_2$

$$\alpha^{(1)} \Rightarrow \alpha^{(1)}, \alpha^{(2)} \Rightarrow \alpha^{(3)}, \alpha_P^* \Rightarrow \alpha^{(2)}$$

（2） $\alpha_P^* > \alpha^{(2)}$， $f_P < f_2$

$$\alpha^{(2)} \Rightarrow \alpha^{(1)}, \alpha_P^* \Rightarrow \alpha^{(2)}, \alpha^{(3)} \Rightarrow \alpha^{(3)}$$

（3） $\alpha_P^* < \alpha^{(2)}$， $f_P > f_2$

$$\alpha_P^* \Rightarrow \alpha^{(1)}, \alpha^{(2)} \Rightarrow \alpha^{(2)},, \alpha^{(3)} \Rightarrow \alpha^{(3)}$$

（4） $\alpha_P^* > \alpha^{(2)}$， $f_P > f_2$

$$\alpha^{(1)} \Rightarrow \alpha^{(1)},\ \alpha^{(2)} \Rightarrow \alpha^{(2)},\ \alpha_P^* \Rightarrow \alpha^{(3)}$$

基于梯度优化方法：

- **抛物插值法：**

注意点

1) 三个试探点$\alpha^{(1)} < \alpha^{(2)} < \alpha^{(3)}$

且 $f_1 > f_2, f_2 < f_3$； "大-小-大"

2) 迭代中止条件

记$\alpha^{(2)}$与f_2是前一次迭代的插值函数极小点与极小值；

记$\alpha^{(4)}$与f_4是本次迭代的插值函数极小点与极小值。

ε_1、ε_2是给定的精度

若 $\left|\alpha^{(4)}-\alpha^{(2)}\right|<\varepsilon_1$ 或 $\left|f_4-f_2\right|<\varepsilon_2$

THE END