运筹学与优化方法

晁国清
计算机科学与技术学院

课程回顾

- 凸优化问题

凸优化问题、凸优化问题的最优性条件$x^* \in S$ 是最优解$\Leftrightarrow$ $\nabla f(x^*)^T(x-x^*) \geq 0, \forall x \in S$、无约束、等式约束、非负约束的凸优化问题最优性条件

- 常见凸优化问题的类型
  - 线性规划(Linear programming, LP)
  - 二次规划(Quadratic programming, QP)
  - 带二次约束的二次规划(Quadratically constrained quadratic program, QCQP)
  - 二阶锥规划(Second-order cone program, SOCP)
  - 半定规划(Semidefinite program, SDP)
  - 几何规划(Geometric programming, GP)

无约束优化问题

- 无约束优化问题的最优性条件
- 迭代算法的基本思想
- 一维（线性）搜索

无约束优化问题的最优性条件

对于无约束优化问题$\min f(x)$

- 若$f(x)$为凸函数且$f(x)$在x^*处一阶可微，则

$$x^*\text{是最优解} \Leftrightarrow \nabla f(x^*) = 0$$

- 若$f(x)$为凸函数且$f(x)$在x^*处二阶可微，则

$$x^*\text{是最优解} \Leftrightarrow \nabla^2 f(x^*) \succcurlyeq 0$$

- 若$f(x)$为一般函数且$f(x)$在x^*处一阶可微，则

$$x^*\text{是最优解} \Rightarrow \nabla f(x^*) = 0$$

- 若$f(x)$为一般函数且$f(x)$在x^*处二阶可微，则

$$x^*\text{是最优解} \Rightarrow \nabla^2 f(x^*) \succcurlyeq 0$$

- 若$f(x)$为一般函数且$f(x)$在x^*处二阶可微

$$\nabla f(x^*) = 0\text{且}\nabla^2 f(x^*) \succ 0 \Rightarrow x^*\text{是最优解}$$

迭代算法的基本思想

- 利用无约束优化问题的最优性条件求解无约束优化问题困难多多：

1. 有些问题导数不存在
2. 有些问题即便导数存在，计算麻烦
3. 多数情况下由$\nabla f(x)=0$得到一个非线性方程组，求解困难，甚至无法得到解析解

- 无约束优化问题一般采用数值计算的迭代方法
- 迭代方法的基本思想：给定初始点x^0, 产生点列$\{x^k\}_{k=1}^{\infty}$, 并且点列满足条件$f\left(x^{k+1}\right)<f(x^k)$
- 关键问题：如何从当前点x^k迭代到下一个点x^{k+1}?
- 两种策略：线搜索方法和信赖域方法

迭代算法一般步骤

- Step 1：给定初始点x_0，k=0；
- Step 2:判断x^k是否满足终止条件；是，结束；不是，进行Step 3；
- Step 3：寻找x^k处的下降方向d^k；
- Step 4:选择合适的步长$\alpha_k > 0$，使$f\left(x^k + \alpha_k d^k\right) < f(x^k)$成立；
- Step 5:令$x^{k+1} = x^k + \alpha_k d^k$ 且k=k+1；转Step 2
- 接下来重点介绍选择步长问题，下一次课介绍寻找下降方向问题

哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

线搜索方法

- d^k方向上的任何一点可表示为： $x^{k+1} = x^k + \alpha_k d^k$，其中$\alpha_k$是步长因子，是实数， d^k方向上的的任何一点的函数值可表示为$f(x^k + \alpha_k d^k)$，是参数α_k的一元函数。沿d^k方向求$f(x)$的极小点就转化为：

$$\alpha_k^*: \min f(x^k + \alpha_k d^k)$$

求解最优步长α_k的过程就是线搜索过程，也叫一维搜索过程

- 求解一元函数$\phi(\alpha) = f(x + \alpha d)$的极小点$\alpha^*$, 采用解析法

即利用一元函数的最优条件$\phi'(\alpha) = 0$, 求α^*

注意：在用函数$\phi(\alpha)$的导数求α^*时，所用的函数$\phi(\alpha)$是以步长α为变量的一元函数，而不是以向量x为变量的多元函数$f(x)$

解析解法

- 为了直接利用$f(\boldsymbol{x})$的函数式求解最佳步长因子α^*，把$f(\boldsymbol{x}+\alpha\boldsymbol{d})$进行泰勒展开，取到二阶项，即

$$f(\boldsymbol{x}+\alpha\boldsymbol{d}) \approx f(\boldsymbol{x})+\alpha\mathbf{d}^{\mathrm{T}}\nabla f(\boldsymbol{x})+\frac{1}{2}(\alpha\boldsymbol{d})^T\boldsymbol{G}(\alpha\boldsymbol{d})$$

$$= f(\boldsymbol{x})+\alpha\boldsymbol{d}^T\nabla f(\boldsymbol{x})+\frac{1}{2}\alpha^2\boldsymbol{d}^T\boldsymbol{G}\boldsymbol{d}$$

将上式对α进行微分并令其等于零，得到$\boldsymbol{d}^T\nabla \mathrm{f}(\mathbf{x})+\alpha^*\boldsymbol{d}^T\boldsymbol{G}\boldsymbol{d}=\mathbf{0}$

从而求得$\alpha^* = -\dfrac{\boldsymbol{d}^T\nabla \mathrm{f}(\mathbf{x})}{\boldsymbol{d}^T\boldsymbol{G}\boldsymbol{d}}$

注意：这里直接利用函数$f(x)$而不需要将它转化为步长因子α的函数$\phi(\alpha)$。但是，它需要计算$\boldsymbol{x}=\boldsymbol{x}^k$点处梯度$\nabla f(\boldsymbol{x})$和海塞矩阵$\boldsymbol{G}$

解析解法缺点：需要求导，对于函数关系复杂、求导困难或无法求导的情况，使用解析法不合适

因此，在优化设计中常采用数值解法

线搜索方法

- 线搜索方法的数值解法，利用计算机反复迭代求最佳步长因子α^*的近似值

线搜索数值解法
- 精确搜索
  - 区间试探法
    - 进退探索法
    - 黄金分割法
  - 函数逼近法
    - 牛顿法
    - 抛物线法
- 不精确搜索
  - Armijo-Goldstein法则
  - Wolf-Powell法则

精确搜索就是求$\phi(\alpha)$的极小值，使得到的α_k为最优步长因子。不精确搜索就是选取不太小的α_k使得$\phi(\alpha_k) < \phi(x_k + \alpha_k d_k)$

进退探索法

- 区间试探法一般分为两步：

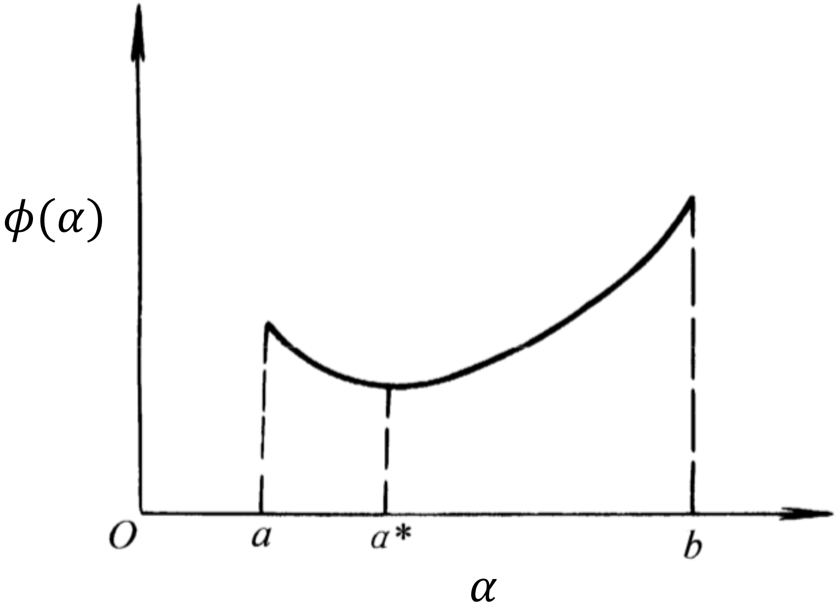
1. 确定包括一维函数极小点在内的单谷区间作为初始搜索区间$[a, b]$ 的方法，如进退探索法
2. 在单谷区间$[a, b]$, 内通过缩小区间寻找极小点的方法，如黄金分割法

- 在给定区间内仅有一个谷值即唯一极小值点的函数称为单谷函数，其区间称为单谷区间

函数值“大--小--大”

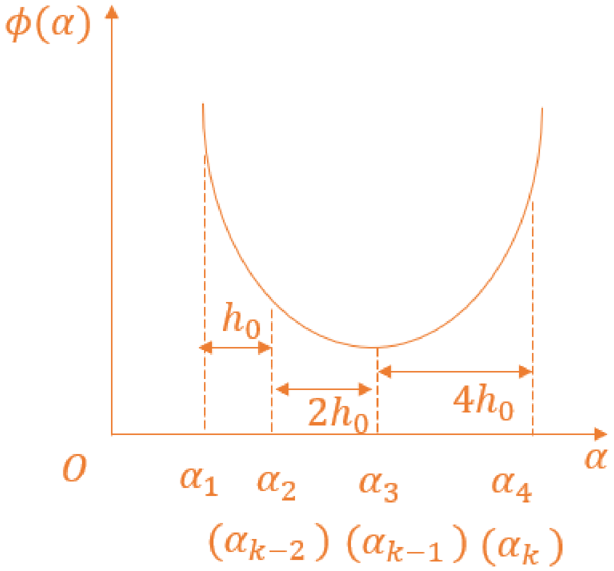
图形上 “高—低—高”

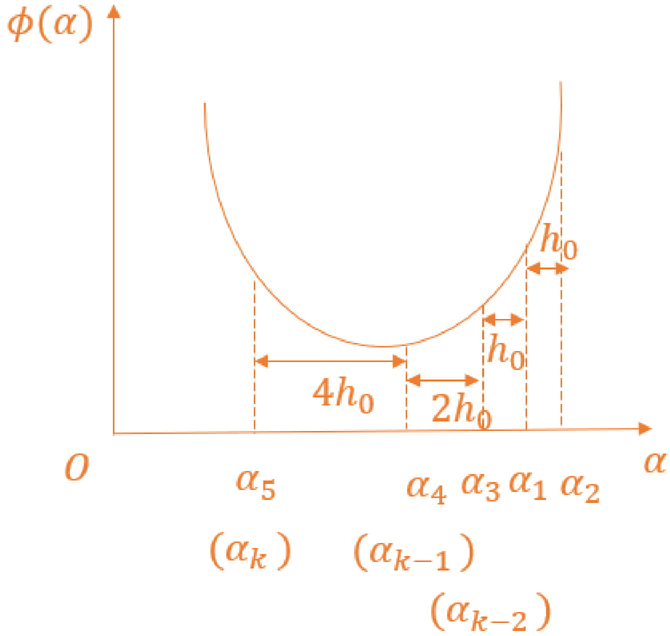
单谷区间中一定能求得一个极小点



进退探索法

- 目的:确定$\phi(\alpha)=f(x+\alpha d)$极小值点$\alpha^*$所在的区间$[a,b]$
- 思路：从一点出发，按一定的步长，试图确定函数值呈现“高-低-高” 的3点。沿着搜索方向若不成功就退回来，再沿相反方向搜索，若搜索正确，则加大步长进行探索，最终找到$\alpha_1,\alpha_2,\alpha_3$3点，满足$\alpha_1<\alpha_2<\alpha_3,\phi(\alpha_1)>\phi(\alpha_2)<\phi(\alpha_3)$为止.





进退探索法

➢算法步骤如下：

(1) 给定初始点α_1, 初始步长$h_0 > 0$

(2) 用加倍步长的外推法寻找初始区间

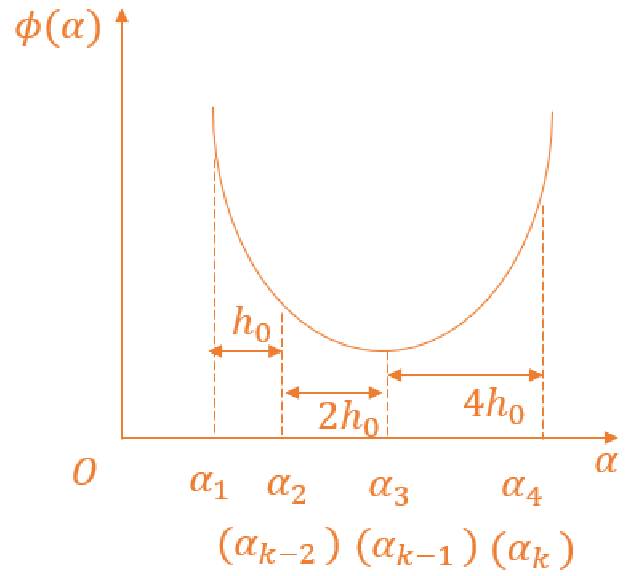
由初始点α_1向某一方向走一步，得到$\alpha_2 = \alpha_1 + h_0$, 比较$\phi(\alpha_2)$和$\phi(\alpha_1)$的大小

1) 若$\phi(\alpha_2) < \phi(\alpha_1)$, 则方向选对，步长加倍，继续得到$\alpha_3 = \alpha_2 + 2h_0$. 若仍有$\phi(\alpha_3) < \phi(\alpha_2)$，则步长再加倍，有$\alpha_4 = \alpha_3 + 4h_0, \cdots$ 直到α_k点函数值刚刚变为增加为止，

也就得到$\alpha_{k-2} < \alpha_{k-1} < \alpha_k$,

$\phi(\alpha_{k-2}) > \phi(\alpha_{k-1}) < \phi(\alpha_k)$,

极小值点必在$[\alpha_{k-2}, \alpha_k]$



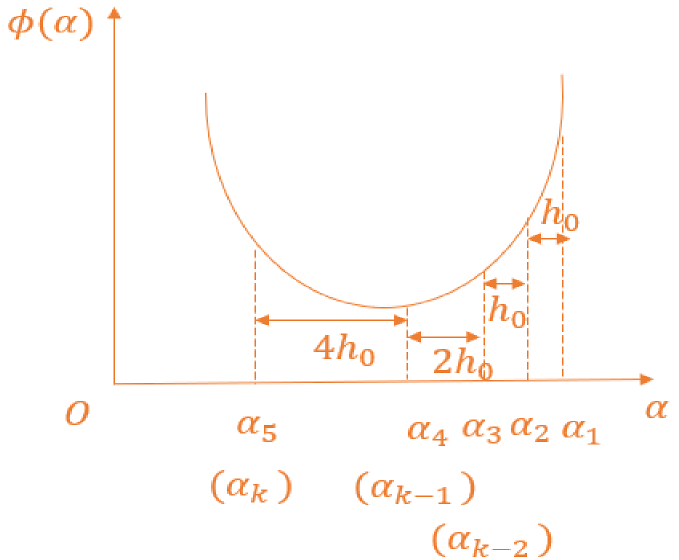
进退探索法

2)若$\phi(\alpha_2)>\phi(\alpha_1)$, 则方向选错，改为向相反方向走一步，得到$\alpha_3=\alpha_1-h_0$，若$\phi(\alpha_3)<\phi(\alpha_1)$. 则方向对，加倍步长走，取$\alpha_4=\alpha_3-2h_0$, 若仍有$\phi(\alpha_4)<\phi(\alpha_3)$, 再加倍步长同方向走，直到函数值刚刚变为增加为止，这样就得到

$\alpha_k<\alpha_{k-1}<\alpha_{k-2}$,

$\phi(\alpha_k)>\phi(\alpha_{k-1})<\phi(\alpha_{k-2})$,

极小值点必在$[\alpha_k,\alpha_{k-2}]$



(3)进一步缩小搜索空间

对于$\alpha_k,\alpha_{k-1},\alpha_{k-2}$三个点，步长是逐次加倍的，故有$\alpha_k-\alpha_{k-1}=2(\alpha_{k-1}-\alpha_{k-2})$

在α_{k-1},α_k中间插一点α_{k+1}，令$\alpha_{k+1}=\frac{1}{2}(\alpha_{k-1}+\alpha_k)$得到四个点$\alpha_k,\alpha_{k+1},\alpha_{k-1},\alpha_{k-2}$, 比较它们的函数值，令其中最小的点为$\alpha_2$, 左右邻点分别为$\alpha_1,\alpha_3$, 则得到更小的搜索区间$[\alpha_1,\alpha_3]$, 这三点同样具有“高-低-高”的形式

黄金分割法(0.618法)

- 黄金分割法适用于$[a,b]$(可由进退搜索法得到)区间上的任何单谷函数求极小值问题。
- 单谷函数与单谷区间的性质：若$\phi(\lambda)$是单谷区间$[a,b]$上的单谷函数，极小点为λ^*, 在$[a,b]$中任取两点$[a_1,b_1]$, 且$a_1<b_1$，则(1)当$\phi(\alpha_1)<\phi(b_1)$时，$\lambda^*\in[a,b_1]$;(2)当$\phi(\alpha_1)>\phi(b_1)$时，$\lambda^*\in[a_1,b]$
- 黄金分割法对插入点的要求：

1. 要求插入点a_1,b_1的位置相对于区间$[a,b]$两端点具有对称性，即

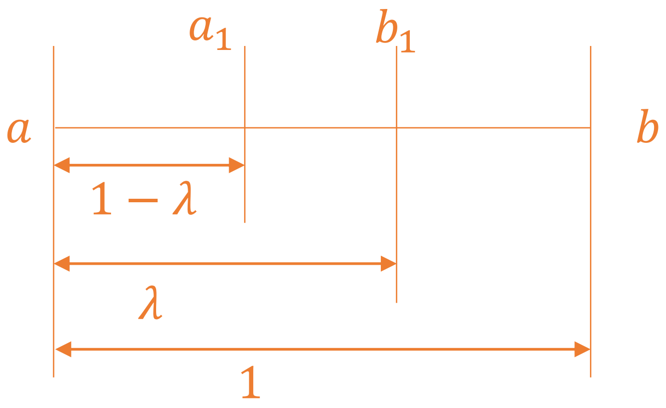
$$a_1=b-\lambda(b-a)$$
$$b_1=a+\lambda(b-a)$$

其中λ为待定常数

2. 要求在保留下来的区间内再插入一点所形成的新的三段区间与原区间的三段具有相同的比例分布。

黄金分割法

- 如下图所示，区间$[a,b]$的大小设为1，设插入点b_1长度为λ，则a_1长度为$1-\lambda$, 新插入点应在$[a,a_1]$之间即$\lambda(1-\lambda)$位置，则此时a_1在原区间的位置$1-\lambda$，相当于保留区间$[a,b_1]$的λ处，即相当于原区间的λ^2处, 故有



$1-\lambda=\lambda^2$，求正解得到

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

总结：黄金分割是指将一线段分成两段，使整段长与较长段的长度比值等于较长段与较短段长度的比值即

$$1:\lambda=\lambda:(1-\lambda)$$

黄金分割法

- 黄金分割法的搜索过程

(1) 给出初始搜索区间$[a,b]$及收敛精度ϵ，设步长$\lambda = 0.618$;

(2) 按坐标点计算公式计算a_1, b_1并计算其对应的函数值$\phi(a_1), \phi(b_1)$

(3) 根据单谷函数的性质缩短搜索区间

(4) 检查区间是否足够小以及函数值收敛到足够精度，如果收敛条件满足，则取最后两插入点的平均值作为极小点的数值近似解。如果条件不满足则转步骤(3)

牛顿法

- 函数逼近法的主要思想是在极小点附近用近似表达式取代目标函数$\phi(\alpha)$，从而求出$\phi(\alpha)$的极小点的估计值
  - 牛顿法是采用二阶泰勒多项式近似取代目标函数$\phi(\alpha)$
  - 抛物线法是采用二次三项式近似取代目标函数$\phi(\alpha)$
- 假定极小点的近似点为α_0，在α_0点将$\phi(\alpha)$作二阶泰勒展开：

$$\phi(\alpha) = \phi(\alpha_0) + \phi'(\alpha_0)(\alpha - \alpha_0) + \frac{1}{2}\phi''(\alpha_0)(\alpha - \alpha_0)^2 + O(|\alpha - \alpha_0|^2)$$

忽略高阶项，得到

$$\psi(\alpha) = \phi(\alpha_0) + \phi'(\alpha_0)(\alpha - \alpha_0) + \frac{1}{2}\phi''(\alpha_0)(\alpha - \alpha_0)^2$$

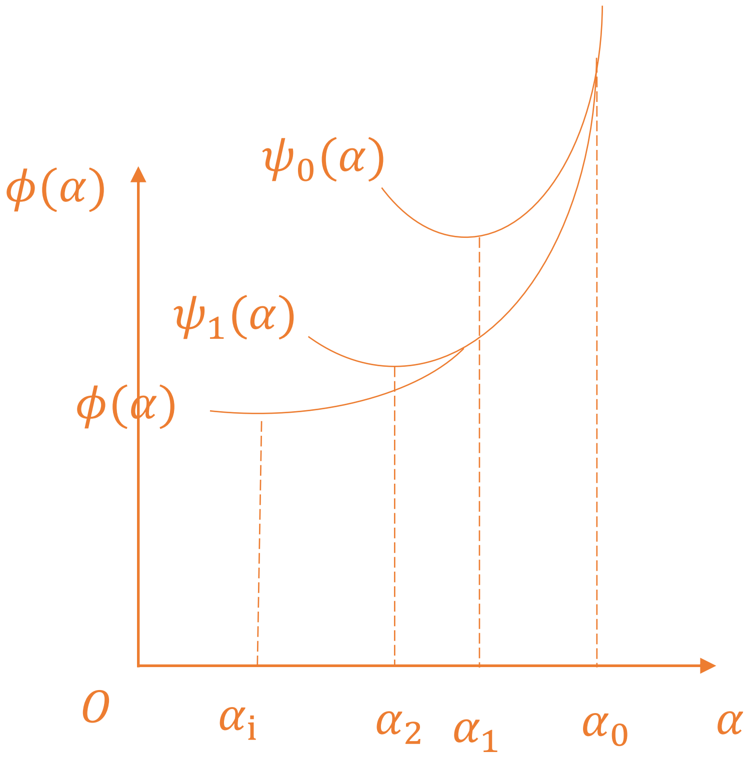
在α_0处可用$\psi(\alpha)$近似$\phi(\alpha)$，然后用$\psi(\alpha)$的极小点作为$\phi(\alpha)$极小点的近似α_1。根据无约束优化问题的极值必要条件得

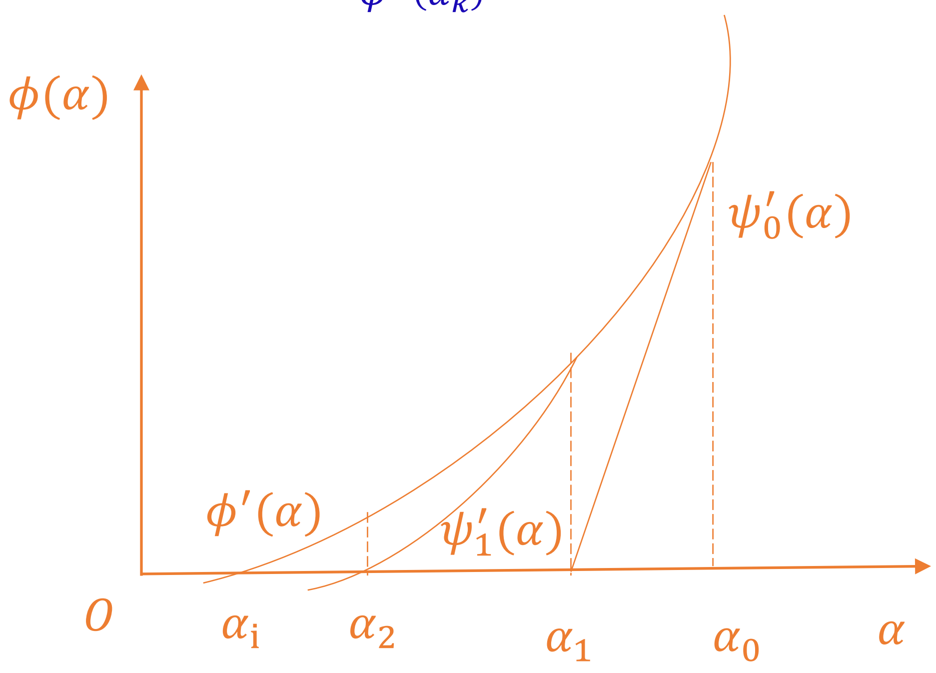
$$\psi'(\alpha) = 0 \Rightarrow \alpha_1 = \alpha_0 - \frac{\phi'(\alpha_0)}{\phi''(\alpha_0)} \Rightarrow \alpha_{k+1} = \alpha_k - \frac{\phi'(\alpha_k)}{\phi''(\alpha_k)}(k = 0, \cdots,)$$

牛顿法的几何解释

- 在下图中，在α_0处用一抛物线$\psi(\alpha)$代替曲线$\phi(\alpha)$, 相当于用一斜线$\psi'(\alpha)$代替曲线$\phi'(\alpha)$。这样各个近似点是通过对$\psi'(\alpha)$作切线求得与α轴的交点找到的，所以牛顿法也被称为切线法

$$\alpha_{k+1} = \alpha_k - \frac{\phi'(\alpha_k)}{\phi''(\alpha_k)} (k = 0, \cdots,)$$





牛顿法的计算步骤

1. 给定初始点α_0, 控制误差ϵ, 并令$k = 0$;
2. 计算$\phi'(\alpha_k), \phi''(\alpha_k)$;
3. 求$\alpha_{k+1} = \alpha_k - \dfrac{\phi'(\alpha_k)}{\phi''(\alpha_k)}$;
4. 若$|\alpha_{k+1} - \alpha_k| \le \epsilon$, 则求得近似解$\alpha^* = \alpha_{k+1}$, 结束；否则转步骤5
5. 令$k = k + 1$, 转步骤2

牛顿法的特点

优点：收敛速度快

缺点：每一点都要求二阶导数，工作量大，要求初始点离极小点不太远，否则可能使极小化发散或收敛到非极小点

抛物线法(二次插值法)

- 基本思想：二次插值的基本思想是极小点近似点附近3个不同点的函数值构成一个与原函数$\phi(\alpha)$近似的二次多项式$\psi(\alpha)$，以函数$\psi(\alpha)$的极小值点作为目标函数$\phi(\alpha)$的近似极值点
- 函数$\phi(\alpha)$有$\alpha_1 < \alpha_2 < \alpha_3$ 3个点，且满足$\phi(\alpha_1) > \phi(\alpha_2) < \phi(\alpha_3)$，令二次三项式$\psi(\alpha) = \lambda_0 + \lambda_1\alpha + \lambda_2\alpha^2$，并设

$$\begin{aligned}\psi(\alpha_1) &= \lambda_0 + \lambda_1\alpha_1 + \lambda_2\alpha_1^2 \\ \psi(\alpha_2) &= \lambda_0 + \lambda_1\alpha_2 + \lambda_2\alpha_2^2 \qquad (*) \\ \psi(\alpha_3) &= \lambda_0 + \lambda_1\alpha_3 + \lambda_2\alpha_3^2\end{aligned}$$

由$\psi(\alpha)$取极值点的必要条件

$$\psi'(\alpha_i) = \lambda_1 + 2\lambda_2\alpha_i = 0, i = 1,2,3$$

得到$\alpha_i = -\frac{\lambda_1}{2\lambda_2}$

通过求解方程组(*)且消去λ_0可得

$$\lambda_1 = \frac{(\alpha_2^2 - \alpha_3^2)\psi(\alpha_1) + (\alpha_3^2 - \alpha_1^2)\psi(\alpha_2) + (\alpha_1^2 - \alpha_2^2)\psi(\alpha_3)}{(\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_3 - \alpha_1)}$$

抛物线法

$$\lambda_2 = -\frac{(\alpha_2-\alpha_3)\psi(\alpha_1)+(\alpha_3-\alpha_1)\psi(\alpha_2)+(\alpha_1-\alpha_2)\psi(\alpha_3)}{(\alpha_1-\alpha_2)(\alpha_2-\alpha_3)(\alpha_3-\alpha_1)}$$

故

$$\alpha_i = -\frac{\lambda_1}{2\lambda_2}$$
$$= \frac{1}{2}\frac{(\alpha_2^2-\alpha_3^2)\psi(\alpha_1)+(\alpha_3^2-\alpha_1^2)\psi(\alpha_2)+(\alpha_1^2-\alpha_2^2)\psi(\alpha_3)}{(\alpha_2-\alpha_3)\psi(\alpha_1)+(\alpha_3-\alpha_1)\psi(\alpha_2)+(\alpha_1-\alpha_2)\psi(\alpha_3)}$$

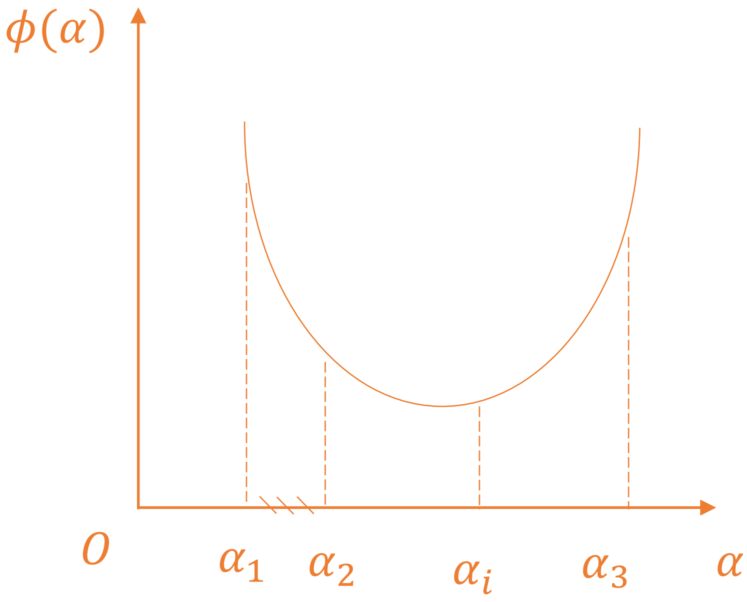
如此便可求出$\phi(\alpha)$极小点α^*的近似解α_i,

然后

1. 如果区间长度$|\alpha_3-\alpha_1|$足够小，则由$|\alpha_i-\alpha^*|<|\alpha_3-\alpha_1|$, 得出近似极小点$\alpha^*\approx\alpha_i$;
2. 如果区间长度$|\alpha_3-\alpha_1|$不足够小，必须根据单谷函数的性质收缩区间$[\alpha_1,\alpha_3]$.

抛物线法

3. 依据单谷函数性质，需要已知区间内两点的函数值，我们有$\phi(\alpha_2)$和$\phi(\alpha_i)$，当$\phi(\alpha_2) < \phi(\alpha_i)$时$[\alpha_1, \alpha_i]$为缩小后的区间；当$\phi(\alpha_2) \geq \phi(\alpha_i)$时取$[\alpha_2, \alpha_3]$为缩小后的区间



区间试探法和函数逼近法的比较

- 区间试探法中试探点位置是由某种给定的规律确定的，它不考虑函数值的分布。比如，黄金分割法是按等比例0.618缩短率确定的。函数逼近法中，试探点位置是按照函数值近似分布的极小点确定的
- 区间试探法只对试探点函数值的大小进行比较，但函数值本身的特性没有得到充分利用，这样即使对一些简单的函数如二次函数，也必须像一般函数那样进行同样多的函数值计算。函数逼近法是利用函数在已知试探点的值(或导数值)来确定新试探点的位置
- 当函数具有比较好的解析性质比如连续可微性，函数逼近法要优于区间试探法

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

参考文献

- 运筹学与最优化方法 第4章 Page 97-105
- 最优化方法，第3章
- 最优化基础理论与方法，第二章
- Numerical optimization, Chapter 2-3