最优化方法

哈尔滨工业大学
数学学院

杨　畅

第 四 章

最优化搜索算法的结构

与

一维搜索

4.1常用的搜索算法结构

1.收敛性概念： 考虑（*fs*）

设迭代算法产生点列$\{x^{(k)}\} \subseteq S$。

（1） 理想的收敛性： 设$x^* \in S$是g.opt.

1° 当$x^* \in \{x^{(k)}\}$

2° $x^{(k)} \neq x^*$，$\forall k$，但满足：$\lim\limits_{k\to\infty} x^{(k)} = x^*$

时，称算法收敛到最优解 x^*。

由于非线性规划问题的复杂性，实用中建立下列收敛性概念：

（2）实用收敛性：定义解集

$S^* = \{ \boldsymbol{x} \mid \boldsymbol{x}$ 具有某种性质 $\}$

例如：$S^*=\{\boldsymbol{x}|\boldsymbol{x}$—g.opt.$\}$

$S^*=\{\boldsymbol{x}|\boldsymbol{x}$—l.opt.$\}$

$S^*=\{\boldsymbol{x}|\ \nabla f(\boldsymbol{x})=0\}$

$S^*=\{\boldsymbol{x}|f(\boldsymbol{x})\leq\beta\ \}$

(β为给定的实数，称为阈值)

▲**收敛性：** 设解集$S^*\neq\emptyset$，$\{\boldsymbol{x}^{(k)}\}$为算法产生的点列。下列情况之一成立时，称算法收敛：

1° $\exists \boldsymbol{x}^{(k)} \in S^*$;

2° $\boldsymbol{x}^{(k)} \notin S^*$，$\forall k$，但$\{\boldsymbol{x}^{(k)}\}$任意极限点 $\in S^*$ 。

▲**全局收敛：** 对任意初始点$\boldsymbol{x}^{(1)}$,算法均收敛。

局部收敛： 当$\boldsymbol{x}^{(1)}$充分接近解$\boldsymbol{x}^*$时，算法才收敛。

2.收敛速度

设算法产生点列$\{x^{(k)}\}$,收敛到解x^*,且$x^{(k)}\neq x^*$，$\forall k$，

（1）线性收敛：$\dfrac{\| x^{(k+1)} - x^* \|}{\| x^{(k)} - x^* \|} < 1$，当$k$充分大时成立。

（2）超线性收敛：$\lim\limits_{k\to\infty} \dfrac{\| x^{(k+1)} - x^* \|}{\| x^{(k)} - x^* \|} = 0$

（3）二阶收敛：∃ $\alpha > 0$，使得当k充分大时，有

$$\frac{\| x^{(k+1)} - x^* \|}{\| x^{(k)} - x^* \|^2} \le \alpha$$

定理 设算法点列$\{x^{(k)}\}$超线性收敛于x^*,且$x^{(k)}\neq x^*$，$\forall k$，那么

$$\lim_{k\to\infty}\frac{\|x^{(k+1)}-x^{(k)}\|}{\|x^{(k)}-x^*\|}=1$$

证明只需注意

$$\left|\,\|x^{(k+1)}-x^*\| - \|x^{(k)}-x^*\|\,\right| \le \|x^{(k+1)}-x^{(k)}\|$$

$$\le \|x^{(k+1)}-x^*\| + \|x^{(k)}-x^*\|$$

同除以$\|x^{(k)}-x^*\|$，并令$k\to\infty$，利用超线性收敛定义可得结果。

3.线性搜索算法

设已得到迭代点$\boldsymbol{x}^{(k)}$：

（1）确定搜索方向$\boldsymbol{d}^{(k)}$

（2）求λ_k，使$f(\boldsymbol{x}^{(k)}+\lambda_k d^{(k)})=\min(f(\boldsymbol{x}^{(k)}+\lambda d^{(k)}) \mid \lambda \in R_k)$。其中$R_k$是针对问题得到的限制集合。

（3）令$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k d^{(k)}$

- 如果R_k=R，那么在一定条件下线性搜索的结果就当有

$$\frac{df(x^{(k)}+\lambda d^{(k)})}{d\lambda}\Big|_{\lambda=\lambda_k}=0$$

即$\nabla f^T(x^{(k)}+\lambda_k d^{(k)})d^{(k)}=0$

4.二次终结性

- 一个算法用于解正定二次函数（$f(x)=1/2x^TAx-c^Tx$,A为对称正定）的无约束极小时，有限步迭代可达最优解，则称该算法具有二次终结性。
- 二次终结性=共轭方向+精确一维搜索。
- 共轭方向

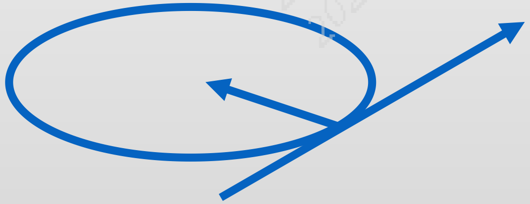
· 定义 设 $A_{n\times n}$ 对称正定，$d^{(1)},d^{(2)}\in R^n$，

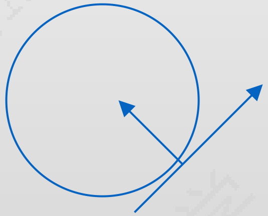
$d^{(1)}\neq 0$，$d^{(2)}\neq 0$，满足$d^{(1)T}Ad^{(2)}=0$，称$d^{(1)},d^{(2)}$关于矩阵A共轭。

· 共轭向量组：$d^{(1)},d^{(2)},\ldots,d^{(m)}\in R^n$均非零，满足

$$d^{(i)T}Ad^{(j)}=0(i\neq j)。$$

· 当$A=E$(单位矩阵)时，$d^{(1)T}Ad^{(2)}=d^{(1)T}d^{(2)}=0$，即正交关系。





· 当$d^{(1)},d^{(2)},\ldots,d^{(m)}$ 关于正定矩阵A两两共轭时， $d^{(1)},d^{(2)},\ldots,d^{(m)}$ 线性无关。

证明. 设$d=\alpha_1 d^{(1)}+\alpha_2 d^{(2)}+\ldots+\alpha_m d^{(m)}=0$，$\forall j=1,2,\ldots,m$，$d^{(j)T}Ad=\alpha_j d^{(j)T}Ad^{(j)}=0$

因为 $d^{(j)T}Ad^{(j)}>0$，故$\alpha_j=0$，即线性无关。

定理：设$d^{(1)},d^{(2)},\ldots,d^{(m)}$，m<n，非零、关于A两两共轭。设$x^{(1)}\in R^n$，

$$x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)},k=1,\ldots,m$$

λ_k为$\lim\limits_{\lambda} f(x^{(k)}+\lambda d^{(k)})$的最优解。

那么$x^{(m+1)}$为$\begin{cases}\min f(x)\\ s.t.\ x\in V\end{cases}$的最优解，其中

$V=\{x|x=x^{(1)}+\sum_{i=1}^{m}\mu_i d^{(i)},\mu_i\in R,i=1,\ldots,m\}$.

注：超线性收敛和二次终结性常用来讨论算法的优点。

5.下降算法模型

考虑（fs） $\begin{cases} \text{Min } f(x) \\ \text{s.t. } x \in S \end{cases}$

常用一种线性搜索的方式来求解：迭代中从一点出发沿下降可行方向找一个新的、性质有改善的点。

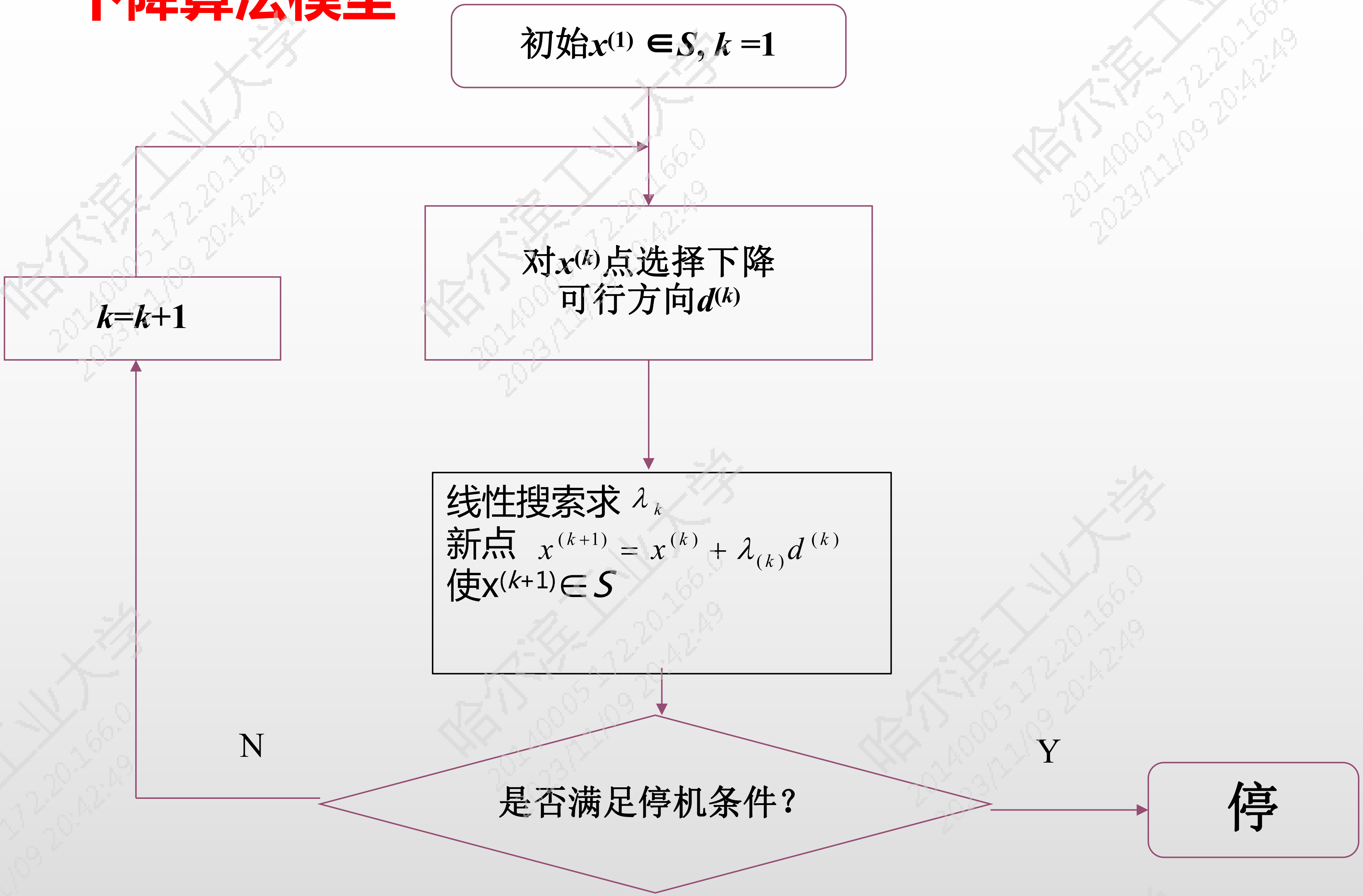
- **下降方向**

设 $\bar{x} \in S$，$d \in R^n$, $d \neq 0$, 若存在 $\delta > 0$，使 $f(\bar{x}+\lambda d) < f(\bar{x}), \forall \lambda \in (0,\delta)$，称 d 为在 $\bar{x}$ 点的下降方向。

- **可行方向**：设 $\bar{x} \in S$，$d \in R^n$, $d \neq 0$, 若存在 $\delta > 0$，使 $\bar{x}+\lambda d \in S, \forall \lambda \in (0,\delta)$，称 d 为 $\bar{x}$ 点的可行方向。

同时满足上述两个性质的方向称为下降可行方向。

下降算法模型


初始$x^{(1)} \in S$, k =1
对$x^{(k)}$点选择下降
可行方向$d^{(k)}$
线性搜索求 λ_k
新点 $x^{(k+1)} = x^{(k)} + \lambda_{(k)} d^{(k)}$
使$x^{(k+1)} \in S$
是否满足停机条件？
Y
停
N
k=k+1

4.2 一维搜索

一元函数求极小值及线性搜索均为一维搜索。常用于求：

$$\text{Min } f(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{d}^{(k)})=\varphi(\lambda)$$

$$\text{s.t.} \quad \lambda \in S$$

S有3种情况（-∞，+∞）或（0， +∞ ）或$[a,b]$

1.缩小区间的精确一维搜索：考虑问题(P)

$$\text{Min } \varphi(\lambda)$$

$$\text{s.t. } \lambda \in [\alpha, \beta]$$

其中$\varphi(\lambda)$:$\mathbf{R}\to\mathbf{R}$

（1）不确定区间及单峰函数

➢ **不确定区间：$[\alpha, \beta]$含$\varphi(\lambda)$的最小点，但不知其位置。**

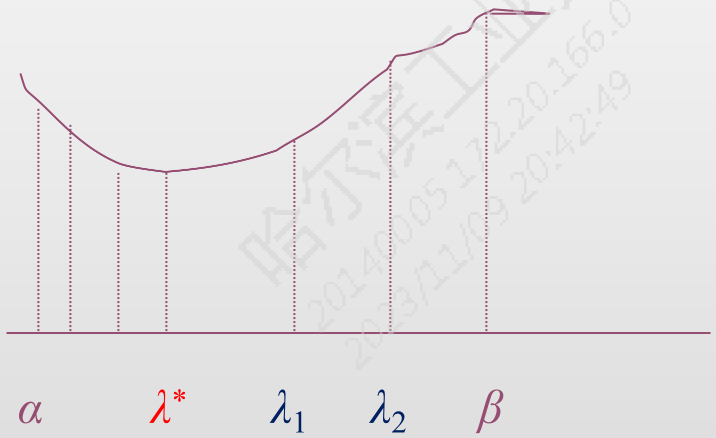
定义 设φ：$[\alpha,\beta]\to\mathbf{R}$，$\exists\lambda^*\in[\alpha,\beta]$ 是φ在$[\alpha,\beta]$上的最小点，若对任意λ_1,λ_2, $\alpha\le\lambda_1<\lambda_2\le\beta$满足：

1° 若$\lambda_2\le\lambda^*$，则$\varphi(\lambda_1)>\varphi(\lambda_2)$;

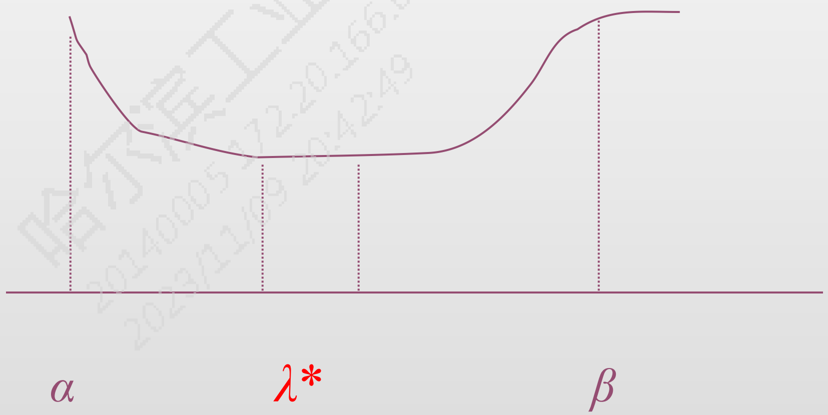
2° 若$\lambda^*\le\lambda_1$，则$\varphi(\lambda_1)<\varphi(\lambda_2)$。

则称$\varphi(\lambda)$在$[\alpha,\beta]$上强单峰。

若只有当$\varphi(\lambda_1)\ne\varphi(\lambda^*)$，$\varphi(\lambda_2)\ne\varphi(\lambda^*)$时，上述1°, 2° 中的式子才成立，则称$\varphi(\lambda)$在$[\alpha,\beta]$上单峰。



强单峰

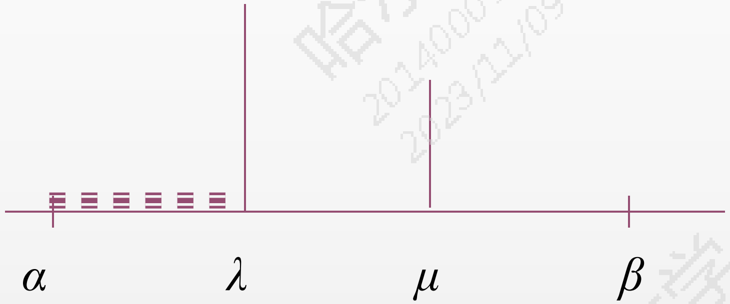


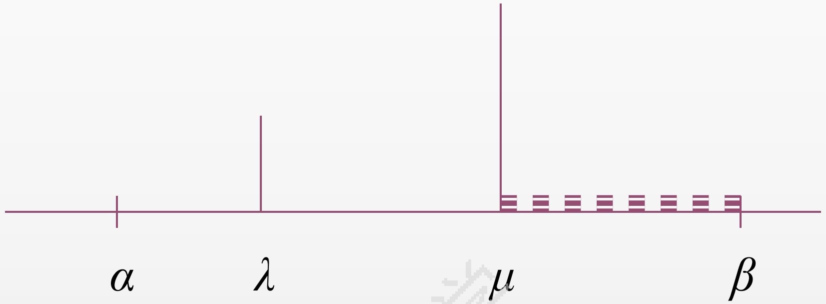
单峰

定理 设Φ：R→R 在$[\alpha，\beta]$上单峰，$\alpha\le\lambda<\mu\le\beta$。那么

1°若$\Phi(\lambda)>\Phi(\mu)$，则$\Phi(\rho)\ge\Phi(\mu)$，$\forall\ \rho\in[\alpha，\lambda]$，如左下图所示；

2°若$\Phi(\lambda)<\Phi(\mu)$，则$\Phi(\rho)\ge\Phi(\lambda)$，$\forall\ \rho\in[\mu，\beta]$，如右下图所示。





证明. 1°反证法：设$\lambda^*\in[\alpha,\beta]$为最小点，假设$\gamma\in[\alpha,\lambda]$及$\gamma<\lambda<\lambda^*$，使$\Phi(\gamma)<\Phi(\mu)<\Phi(\lambda)$。

若$\lambda^*\in[\lambda,\beta]$，由定义$\Phi(\gamma)>\Phi(\lambda)$，矛盾（假设）；

若$\lambda^*\in[\alpha,\lambda)$，由定义及$\mu>\lambda\ge\lambda^*$，$\Phi(\mu)>\Phi(\lambda)$，矛盾（条件）；于是结论成立。

2°的证明类似（略）。

启示：通过上述定理，选两点λ和μ，比较$\Phi(\lambda)$与$\Phi(\mu)$，可去掉$[\alpha,\lambda]$或者$[\mu,\beta]$.

(2)进退法求初始不确定区间

找三点使两端点的函数值大于中间点的函数值。

思路：任取λ_0，步长$\delta>0$，取$\lambda_1=\lambda_0+\delta$，

1°若$\Phi(\lambda_0)<\Phi(\lambda_1)$，令$\delta=2\delta$（步长加倍），$\lambda_2=\lambda_0-\delta$，

若$\Phi(\lambda_2)<\Phi(\lambda_0)$，则令$\lambda_1=\lambda_0$，$\lambda_0=\lambda_2$，重复 1°

若$\Phi(\lambda_2)>\Phi(\lambda_0)$，则停，$\alpha=\lambda_2$，$\beta=\lambda_1$　　(图1)

2°若$\Phi(\lambda_0)>\Phi(\lambda_1)$，令$\delta=2\delta$，$\lambda_2=\lambda_1+\delta$，

若$\Phi(\lambda_2)<\Phi(\lambda_1)$，则令$\lambda_0=\lambda_1$，$\lambda_1=\lambda_2$，重复 2°

若$\Phi(\lambda_2)>\Phi(\lambda_1)$，则停，$\alpha=\lambda_0$，$\beta=\lambda_2$　　(图2)

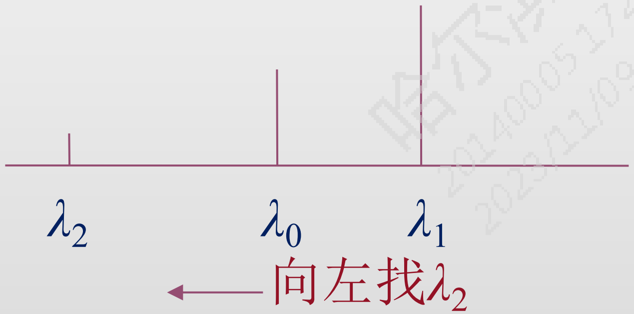


图1

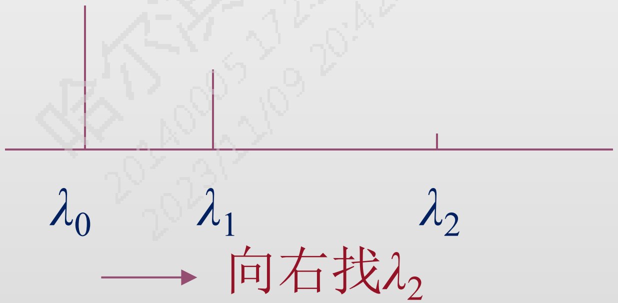


图2

注：1° δ选择要适当（太大含多个单峰区间，太小迭代次数增加）；

2° $\Phi(\lambda)$单调时无结果（加迭代次数限制）；

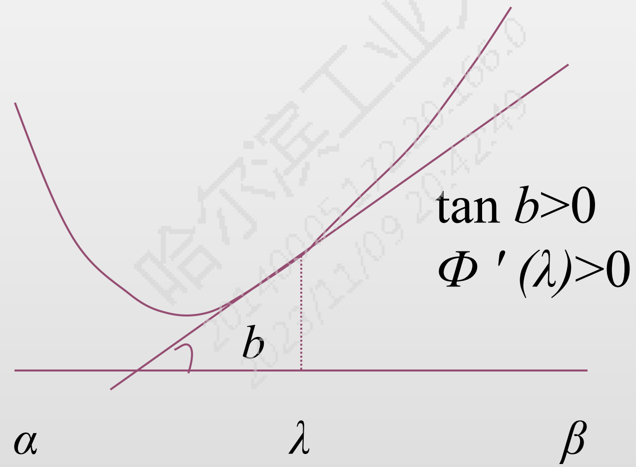
（3）中点法

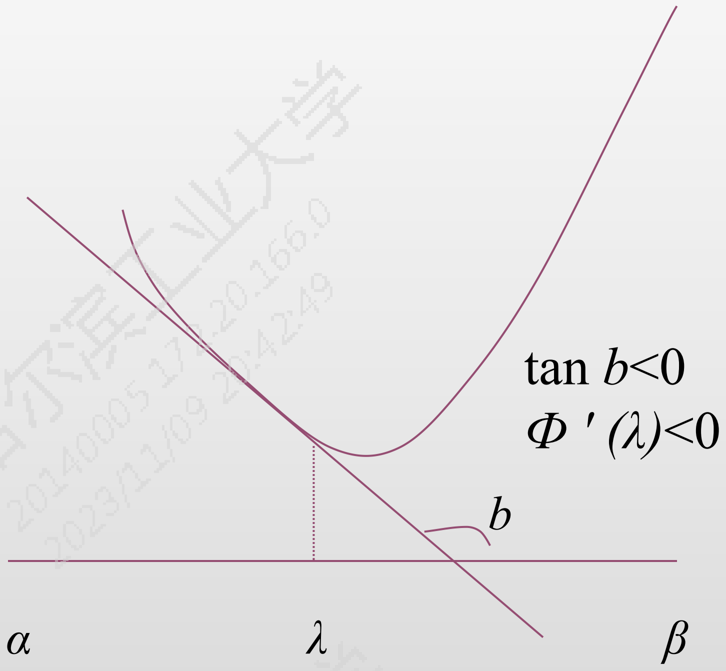
设$\Phi(\lambda)$在$[\alpha, \beta]$上可微，且当导数为零时是解。取$\lambda=(\alpha+\beta)/2$，那么

$\Phi'(\lambda)=0$ 时，λ为最小点，$\lambda=\lambda^*$；

$\Phi'(\lambda)>0$ 时，λ在上升段，$\lambda^*<\lambda$，去掉$[\lambda,\beta]$；(左下图)

$\Phi'(\lambda)<0$ 时，λ在下降段，$\lambda^*>\lambda$，去掉$[\alpha,\lambda]$。(右下图)





（4）黄金分割法（0.618 法）

考虑条件：

1°对称：$\lambda-\alpha=\beta-\mu$ ①

（使“坏”的情况去掉，区间长度不小于“好”的情况）

2°保持缩减比 t=(保留的区间长度 / 原区间长度） 不变。

（使每次保留下来的节点，λ或 μ，在下一次的比较中成为一个相应比例位置的节点 ）。

推导缩减比 t： 如图设第一次保留[α, μ] (去掉[μ, β]),那么第二次保留的长度为[α, λ],则

$$t=\frac{\mu-\alpha}{\beta-\alpha}=\frac{\lambda-\alpha}{\mu-\alpha} \quad ②$$

不妨设$\beta-\alpha=1$，**则**$\mu-\alpha=t$，$\lambda-\alpha=1-t$。**于是由式②得**

$$t/1=(1-t)/t$$

整理得：$t^2+t-1=0$

故 $t\approx0.618$

$$\boxed{t=\frac{-1\pm\sqrt{5}}{2}(\text{舍去负值})}$$

注：**整理式②**：

$$\mu=\alpha+t(\beta-\alpha)$$

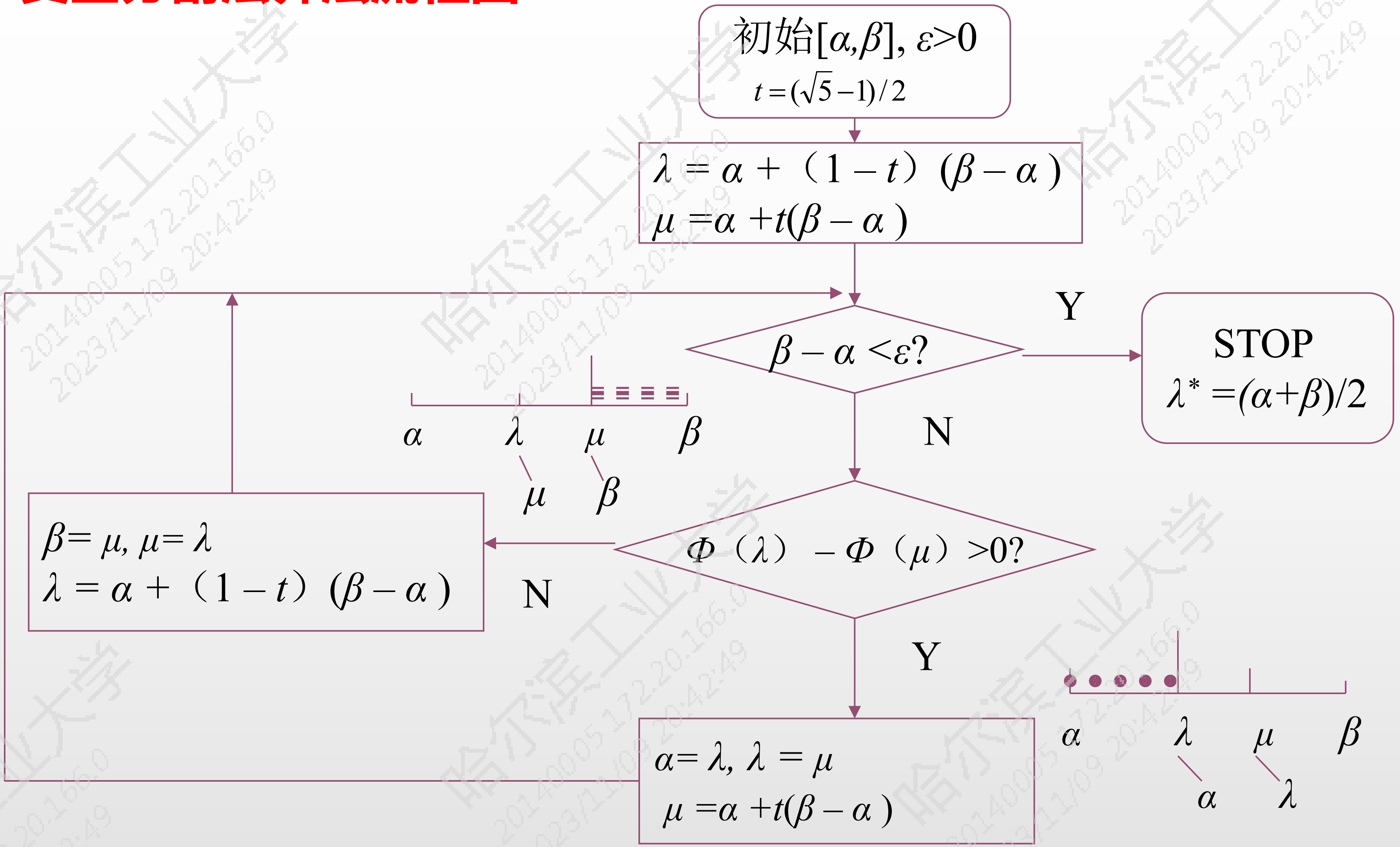
$$\lambda=\alpha+t(\mu-\alpha)=\alpha+t^2(\beta-\alpha)$$

由于 $t^2=1-t$，**故有**

$$\mu=\alpha+t(\beta-\alpha)$$

$$\lambda=\alpha+(1-t)(\beta-\alpha)$$

黄金分割法算法流程图：



初始[α,β], ε>0
t=(√5−1)/2
λ = α +（1 − t）(β − α)
μ =α +t(β − α)
β − α <ε?
Y
STOP
λ* =(α+β)/2
N
Φ（λ） − Φ（μ）>0?
N
β= μ, μ= λ
λ = α +（1 − t）(β − α)
α λ μ β
μ β
Y
α= λ, λ = μ
μ =α +t(β − α)
α λ μ β
α λ

2.牛顿法（Newton）和插值法

（1）牛顿法

对Φ 在λ_k 点展开：

$$\Phi(\lambda) = \Phi(\lambda_k) + \Phi'(\lambda_k)(\lambda - \lambda_k) + (1/2)\Phi''(\lambda_k)(\lambda - \lambda_k)^2 + o(\lambda - \lambda_k)^2$$

取二次式（略去高阶项）：

$$q_k(\lambda) = \Phi(\lambda_k) + \Phi'(\lambda_k)(\lambda - \lambda_k) + (1/2)\Phi''(\lambda_k)(\lambda - \lambda_k)^2$$

用$q_k(\lambda)$作为$\Phi(\lambda)$的近似，当$\Phi''(\lambda_k) > 0$时$q_k(\lambda)$为凸函数，其驻点为极小点：

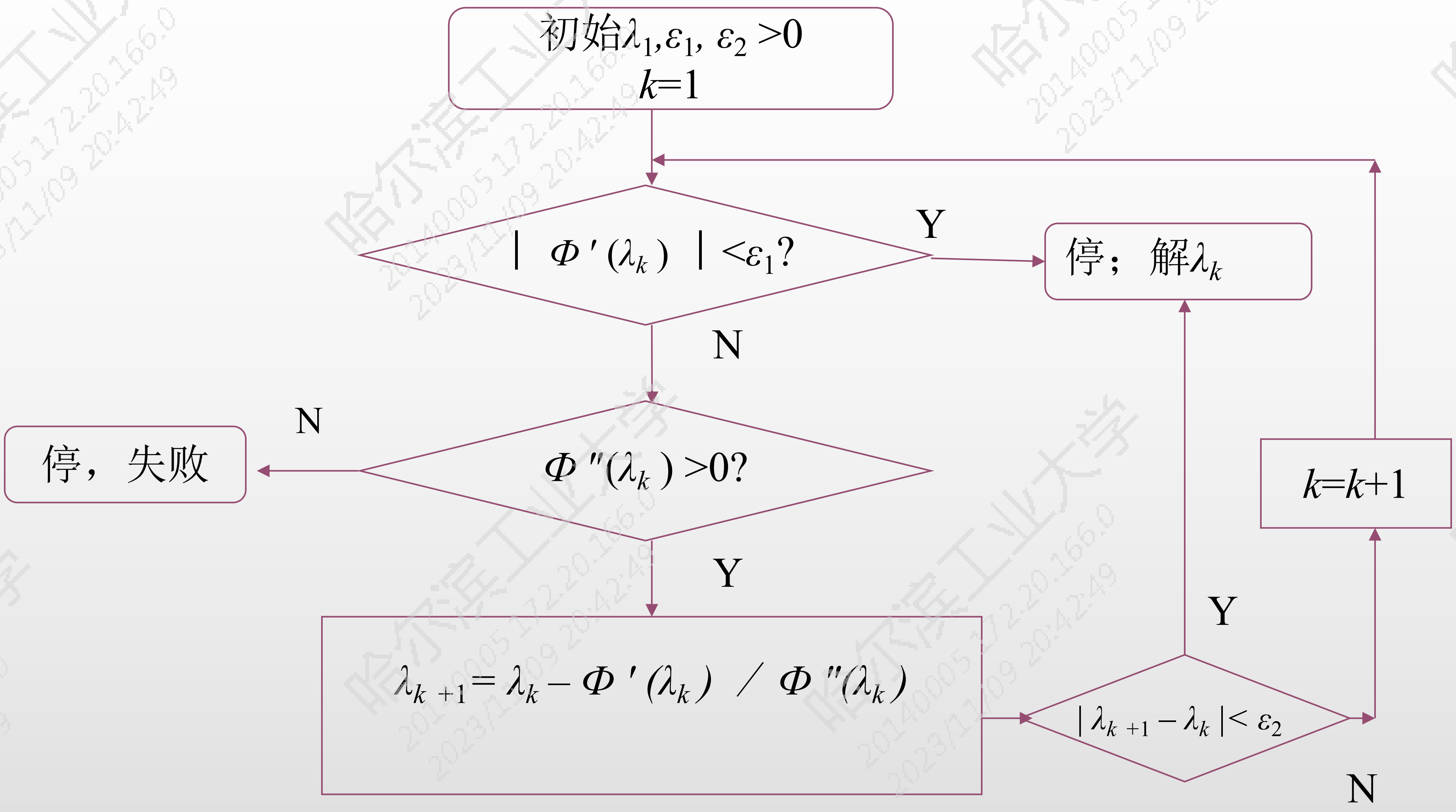
$$q'_k(\lambda) = \Phi'(\lambda_k) + \Phi''(\lambda_k)(\lambda - \lambda_k) = 0$$

令 $\lambda_{k+1} = \lambda = \lambda_k - \Phi'(\lambda_k) / \Phi''(\lambda_k)$为新的迭代点。

以上过程即牛顿法。

特点：二阶、局部收敛。

牛顿法算法框图：



初始$\lambda_1, \varepsilon_1, \varepsilon_2 > 0$
$k=1$
$|\Phi'(\lambda_k)| < \varepsilon_1$?
Y
停；解λ_k
N
$\Phi''(\lambda_k) > 0$?
N
停，失败
Y
$\lambda_{k+1} = \lambda_k - \Phi'(\lambda_k) / \Phi''(\lambda_k)$
$|\lambda_{k+1} - \lambda_k| < \varepsilon_2$
Y
N
$k=k+1$

例. 求 Min $\Phi(\lambda)=\int_0^{\lambda}\arctan(t)\,dt$

解 $\Phi'(\lambda)=\arctan\lambda$ ，$\Phi''(\lambda)=1/(1+\lambda^2)$

迭代公式：$\lambda_{k+1}=\lambda_k-(1+\lambda_k^2)\arctan\lambda_k$

取$\lambda_1=1$，计算结果：

k	λ_k	$\Phi'(\lambda_k)$	$1/\Phi''(\lambda_k)$
1	1	0.7854	2
2	– 0.5708	– 0.5187	1.3258
3	0.1169	– 0.1164	1.0137
4	– 0.001095	– 0.001095	

$\lambda_4\approx\lambda^*=0$

取$\lambda_1=2$，计算结果如下：

k	λ_k	$\Phi'(\lambda_k)$	$1/\Phi''(\lambda_k)$
1	2	1.1071	5
2	– 3.5357	– 1.2952	13.50
3	13.95	不收敛	

（2）插值法

用$\Phi(\lambda)$在2 或3 个点的函数值或导数值，构造2 次或3次多项式作为$\Phi(\lambda)$的近似值，以这个多项式的极小点为新的迭代点。

3点2次，2点2次，4点3次，3点3次，2点3次等

下面以3点2次为例：

取λ_1，λ_2，λ_3，求出$\Phi(\lambda_1)$， $\Phi(\lambda_2)$， $\Phi(\lambda_3)$

设二次插值多项式：$a\lambda^2+b\lambda+c=\Phi(\lambda)$

$$a\lambda_1^2+b\lambda_1+c=\Phi(\lambda_1)$$
$$a\lambda_2^2+b\lambda_2+c=\Phi(\lambda_2)$$
$$a\lambda_3^2+b\lambda_3+c=\Phi(\lambda_3)$$

解得

$$a=-\frac{(\lambda_1-\lambda_2)\Phi(\lambda_3)+(\lambda_2-\lambda_3)\Phi(\lambda_1)+(\lambda_3-\lambda_1)\Phi(\lambda_2)}{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)(\lambda_3-\lambda_1)}$$

$$b=\frac{(\lambda_1^2-\lambda_2^2)\Phi(\lambda_3)+(\lambda_2^2-\lambda_3^2)\Phi(\lambda_1)+(\lambda_3^2-\lambda_1^2)\Phi(\lambda_2)}{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)(\lambda_3-\lambda_1)}$$

$$\bar{\lambda}=-\frac{b}{2a}$$

3.不精确一维搜索： Min $f(x)$

考虑从 $x^{(k)}$ 点出发，沿方向 $d^{(k)}$ 寻找新迭代点： $x^{(k)}+\lambda_k d^{(k)}$

要求: 1° $f(x^{(k)}+\lambda_k d^{(k)})<f(x^{(k)})$;

2° $\lambda_k>0$ 不能太小。

每一步不要求达到精确最小，因此一维搜索速度更快。虽然步数增加，则整体收敛速度更快。

一个实用方法：为了方便，省去上标：

设 $f:R^n \to R$。在 x 取方向 d ，有 $\nabla f^T(x)d<0$ （即 d 为下降方向）

求 λ 使

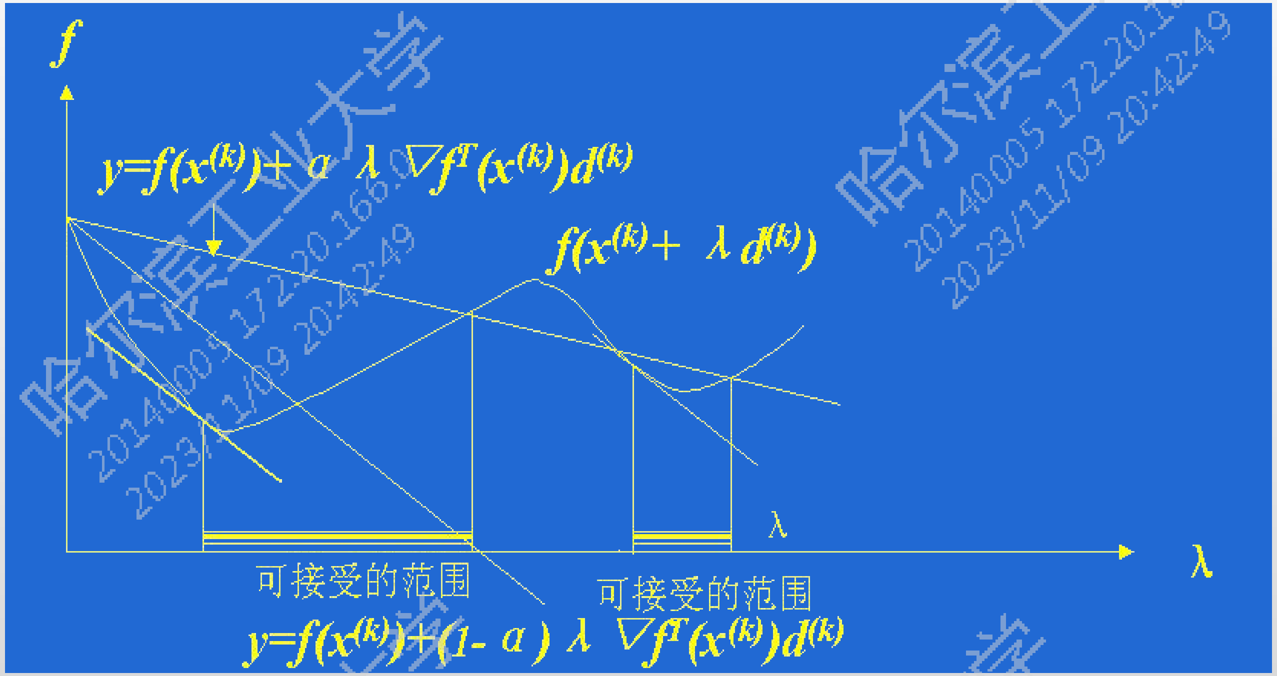
1° $f(x+\lambda d)\le f(x)+\alpha\lambda\nabla f^T(x)d$

2° $\nabla f^T(x+\lambda d)d \ge \beta\nabla f^T(x)d$

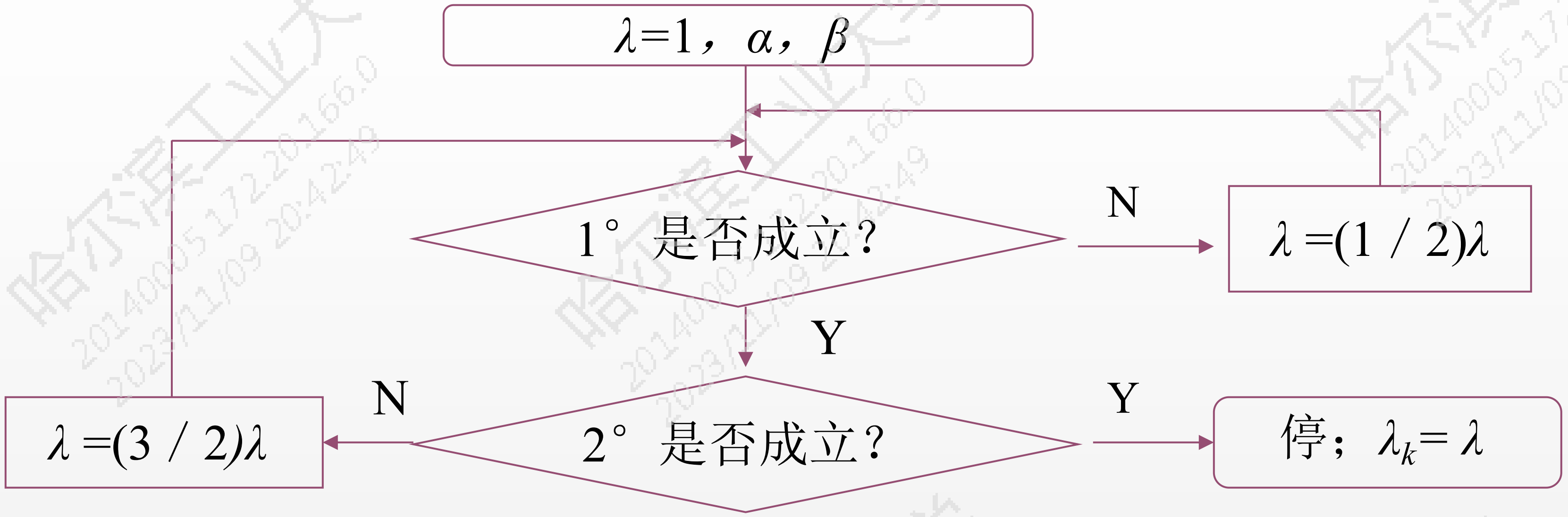
其中 $\alpha\in(0,1/2)$ ，

$\beta\in(\alpha,1)$ 为取定参数。

实际中常取 $\alpha=0.1$ ，$\beta=0.7$ 附近。



不精确一维搜索流程图



要提高精确度可把2° 改为：

$2°\,'\quad |\nabla f^{\mathrm{T}}(x+\lambda d)\, d| \le -\gamma\, \nabla f^{\mathrm{T}}(x)\, d$

当γ =0 时变成精确一维搜索。

此方法一般经几次迭代即可得到满意的λ_k 。