最优化方法

哈尔滨工业大学
数学学院

杨　畅

第 五 章

无约束最优化方法

考虑（f） Min $f(x)$

其中 f：$R^n \to R$

5.1 最优性条件

定理：（极小点的一阶必要条件）设 f 连续可微，若 x^*是l.opt.， 则 $\nabla f(x^*)=0$ (驻点)。

定理：当 f 凸时，如果 $x^* \in R^n$是驻点(即$\nabla f(x^*)=0$)，那么 x^*为f的最小点。

注意： $f(x) \geq f(x^*) + \nabla f^T(x^*)(x - x^*), \forall x$

故 $f(x^*) \leq f(x)$， $\forall x$ （由于$\nabla f^T(x^*)=0$）。

定理：（极小点的二阶必要条件）设 f 二阶连续可微，若 x^* 是l.opt.，则 $\nabla f(x^*)=0$，且海塞矩阵 $\nabla^2 f(x^*)$ 半正定。

定理：（极小点的二阶充分条件）设 f 二阶连续可微，$x^*\in R^n$ 使得 $\nabla f(x^*)=0$，$\nabla^2 f(x^*)$ 正定，则 x^* 为f的严格l.opt.。

5.2 最速下降法

方向：在迭代点 $x^{(k)}$ 取方向 $d^{(k)}= -\nabla f(x^{(k)})$

步长：精确一维搜索，即求λ_k，使得

$$f(x^{(k)}+\lambda_k d^{(k)})=\min_{\lambda\in\mathbb{R}} f(x^{(k)}+\lambda d^{(k)})$$

更新：$x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$

最速下降方向：梯度方向函数值变化最快的方向

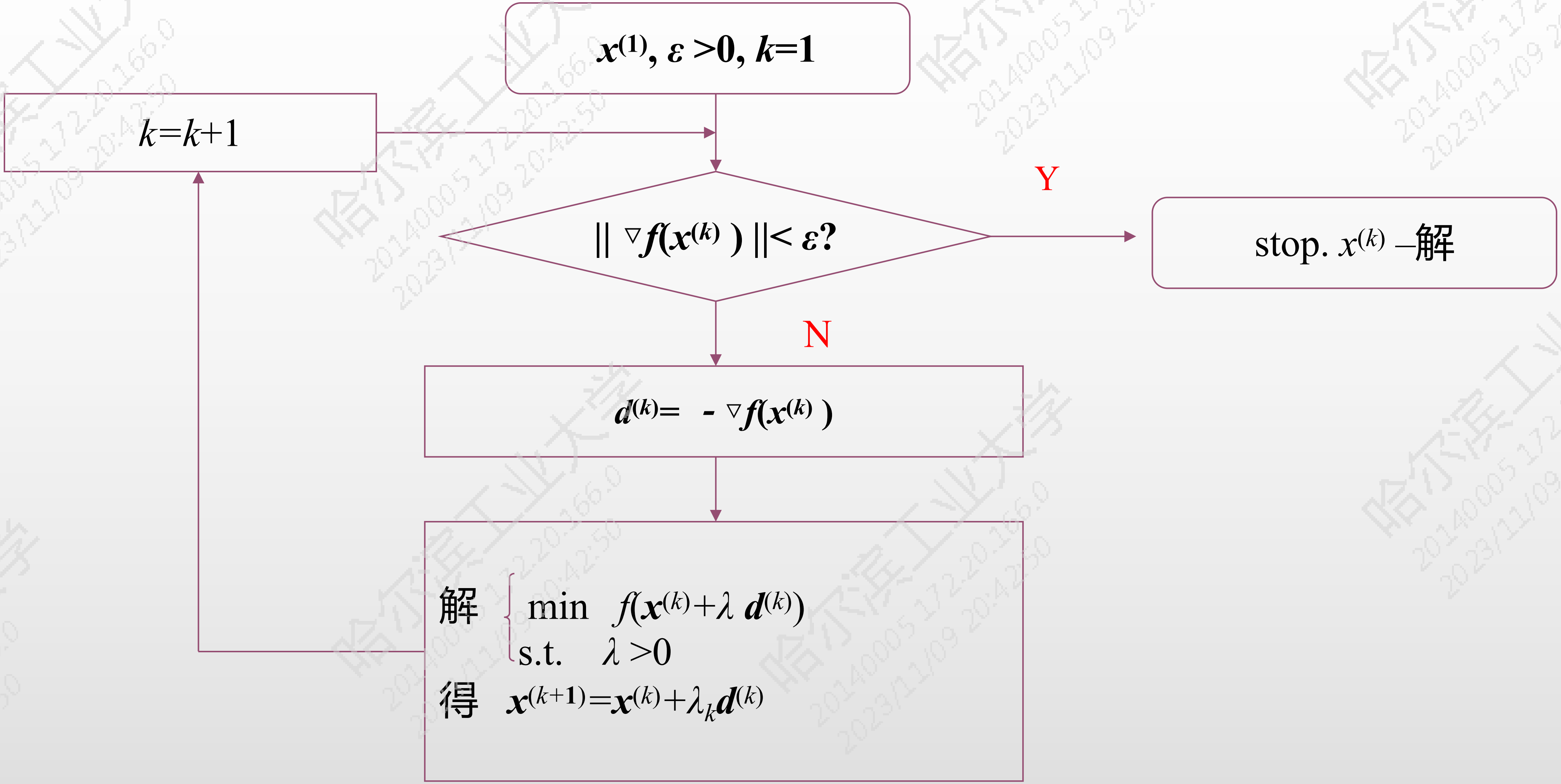
定理：设$f(x):\mathbb{R}^n\to\mathbb{R}$在$x^{(k)}$处可微，$\nabla f(x^{(k)})\neq 0$，则

$$\begin{cases} \min f'(x^{(k)};d) \\ s.t.\ ||d||\leq 1, d\in\mathbb{R}^n \end{cases}$$

的最优解是

$$d^{(k)}=-\frac{\nabla f(x^{(k)})}{\|\nabla f(x^{(k)})\|}$$

最速下降法的算法流程：



$x^{(1)}, \varepsilon>0, k=1$
$k=k+1$
$\|\nabla f(x^{(k)})\|<\varepsilon?$
Y
stop. $x^{(k)}$ –解
N
$d^{(k)}=-\nabla f(x^{(k)})$
解 $\begin{cases} \min & f(x^{(k)}+\lambda d^{(k)}) \\ \text{s.t.} & \lambda>0 \end{cases}$
得 $x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$

特点：全局收敛，线性收敛，易产生扭摆现象（锯齿现象）。

（当$x^{(k)}$距最优点较远时，速度快，而接近最优点时，速度下降）

原因：

用最速下降法极小化目标函数时，相邻两个搜索方向是正交的：

令$\varphi(\lambda)=f(x^{(k)}+\lambda d^{(k)})$，则最小化$\varphi(\lambda)$等价于求

$$\varphi'(\lambda)=\nabla f^T(x^{(k)}+\lambda_k d^{(k)})d^{(k)}=0$$

即（$d^{(k+1)})^T d^{(k)}=0$. 这表明迭代产生的序列$\{x^{(k)}\}$所循路径是“之”字形的。当$x^{(k)}$接近l.opt. x^*时，每次迭代移动的步长很小，这样就呈现出锯齿现象，因此影响了收敛速率。

5.3 共轭梯度法

1.对于正定二次函数

设$f(x)=(1/2)x^{T}Ax-b^{T}x$，其中$A_{n\times n}$对称正定，$b\in R^{n}$

第一步：已知初始点$x^{(1)}$，取最速下降方向

$$d^{(1)}=-\nabla f(x^{(1)})=b-Ax^{(1)}=r^{(1)}\ (\text{剩余向量})$$

接着寻找步长λ_1满足

$$f(x^{(1)}+\lambda_1 d^{(1)})=\lim_{\lambda\in\mathbb{R}} f(x^{(1)}+\lambda d^{(1)})$$

由于$f(x^{(1)}+\lambda d^{(1)})=\frac{1}{2}\langle A(x^{(1)}+\lambda d^{(1)}),x^{(1)}+\lambda d^{(1)}\rangle-\langle b,x^{(1)}+\lambda d^{(1)}\rangle$

$$=\frac{1}{2}\langle Ax^{(1)},x^{(1)}\rangle+\lambda\langle Ax^{(1)},d^{(1)}\rangle+\frac{1}{2}\lambda^2\langle Ad^{(1)},d^{(1)}\rangle-\langle b,x^{(1)}\rangle-\langle b,\lambda d^{(1)}\rangle$$

$$=f(x^{(1)})+\frac{1}{2}\lambda^2\langle Ad^{(1)},d^{(1)}\rangle-\lambda\langle r^{(1)},d^{(1)}\rangle$$

在上式中，对λ求导数得 $\lambda_1=\dfrac{\langle r^{(1)},d^{(1)}\rangle}{\langle Ad^{(1)},d^{(1)}\rangle}$

于是得到$x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}$

第二步：已知$x^{(2)}$，$r^{(2)}$，$d^{(1)}$，从$x^{(2)}$出发，取$d^{(1)}$的共轭方向$d^{(2)}$为下一个搜索方向，即$\langle Ad^{(1)}, d^{(2)}\rangle = 0.$

令$d^{(2)}$=$r^{(2)} + \beta d^{(1)}$,则$\langle Ad^{(1)}, d^{(2)}\rangle = \langle Ad^{(1)}, r^{(2)} + \beta d^{(1)}\rangle = 0.$

于是有$\beta = -\dfrac{\langle Ad^{(1)}, r^{(2)}\rangle}{\langle Ad^{(1)}, d^{(1)}\rangle}.$

再求λ_2使得

$$f(x^{(2)} + \lambda_2 d^{(2)}) = \lim_{\lambda\in\mathbb{R}} f(x^{(2)} + \lambda d^{(2)})$$

于是有$\lambda_2 = \dfrac{\langle r^{(2)}, d^{(2)}\rangle}{\langle Ad^{(2)}, d^{(2)}\rangle}$，从而$x^{(3)} = x^{(2)} + \lambda_2 d^{(2)}$

第三步：重复第二步

例. 用共轭梯度法求解 $10x_1^2+x_2^2$的最小值

解：取初始点$x^{(1)}=(1/10,1)^T$

得到如下结果：

k	$x^{(k)}$	$\nabla f(x^{(k)})$	β_{k-1}	$d^{(k)}$	λ_k
1	$\left(\frac{1}{10},1\right)^T$	$(2,2)^T$	0	$(-2,-2)^T$	$\frac{1}{11}$
2	$\left(-\frac{9}{110},\frac{9}{11}\right)^T$	$\left(-\frac{18}{11},\frac{18}{11}\right)^T$	$\frac{81}{121}$	$\left(\frac{36}{121},-\frac{360}{121}\right)^T$	$\frac{11}{40}$
3	$(0,0)^T$	$(0,0)^T$			

定理：

（1）剩余向量组构成一个正交向量组，即

$$\langle r^{(i)}, r^{(j)}\rangle = 0, \text{当} i \neq j$$

（2）$\{d^{(k)}\}$构成一个A共轭向量组，即

$$\langle d^{(i)}, Ad^{(j)}\rangle = 0, \text{当} i \neq j$$

（3）$\langle r^{(i)}, r^{(i)}\rangle = \langle r^{(i)}, d^{(i)}\rangle$

共轭方向的例子：二维

（1）任选初始点$x^{(0)}$

（2）最速下降方向$d^{(0)}$

（3）得到$x^{(1)} = x^{(0)} + \lambda_0 d^{(0)}$

（4）共轭方向$d^{(1)}$

（5）得到$x^* = x^{(2)} = x^{(1)} + \lambda_1 d^{(1)}$为最优解

原因：

$$\begin{aligned} &Ax^* - b \\ &= A(x^{(1)} + \lambda_1 d^{(1)}) - b \\ &= Ax^{(1)} - b + \lambda_1 Ad^{(1)} \\ &= \nabla f(x^{(1)}) + \lambda_1 Ad^{(1)} \end{aligned}$$

计算$\langle Ax^* - b, d^{(0)}\rangle = 0$，

即$Ax^* - b$与$d^{(0)}$垂直。

由$x^{(0)}$的任意性，故$d^{(0)}$也是任意的，进而$Ax^* - b = 0$

2.对于一般的函数的共轭梯度法

共轭方向为：

$$\boldsymbol{d}^{(k+1)}=-\nabla f(\boldsymbol{x}^{(k+1)})+\beta_k\boldsymbol{d}^{(k)}$$

其中β_k有以下的选择方式：

$$\beta_k=\frac{\nabla f(x^{(k+1)})^T\nabla f(x^{(k+1)})}{\nabla f(x^{(k)})^T\nabla f(x^{(k)})} \qquad FR\text{法}$$

$$\beta_k=\frac{\nabla f(x^{(k+1)})^T[\nabla f(x^{(k+1)})-\nabla f(x^{(k)})]}{\nabla f(x^{(k)})^T\nabla f(x^{(k)})} \qquad PRP\text{法}$$

$$\beta_k=\frac{\nabla f(x^{(k+1)})^T\nabla f(x^{(k+1)})}{-\nabla f(x^{(k)})^T d^{(k)}} \qquad \text{共轭下降法}$$

注：这些表达式在二次函数情况下是等价的。

算法特点

（1）全局收敛（下降算法），线性收敛。

共轭梯度方向为下降方向：

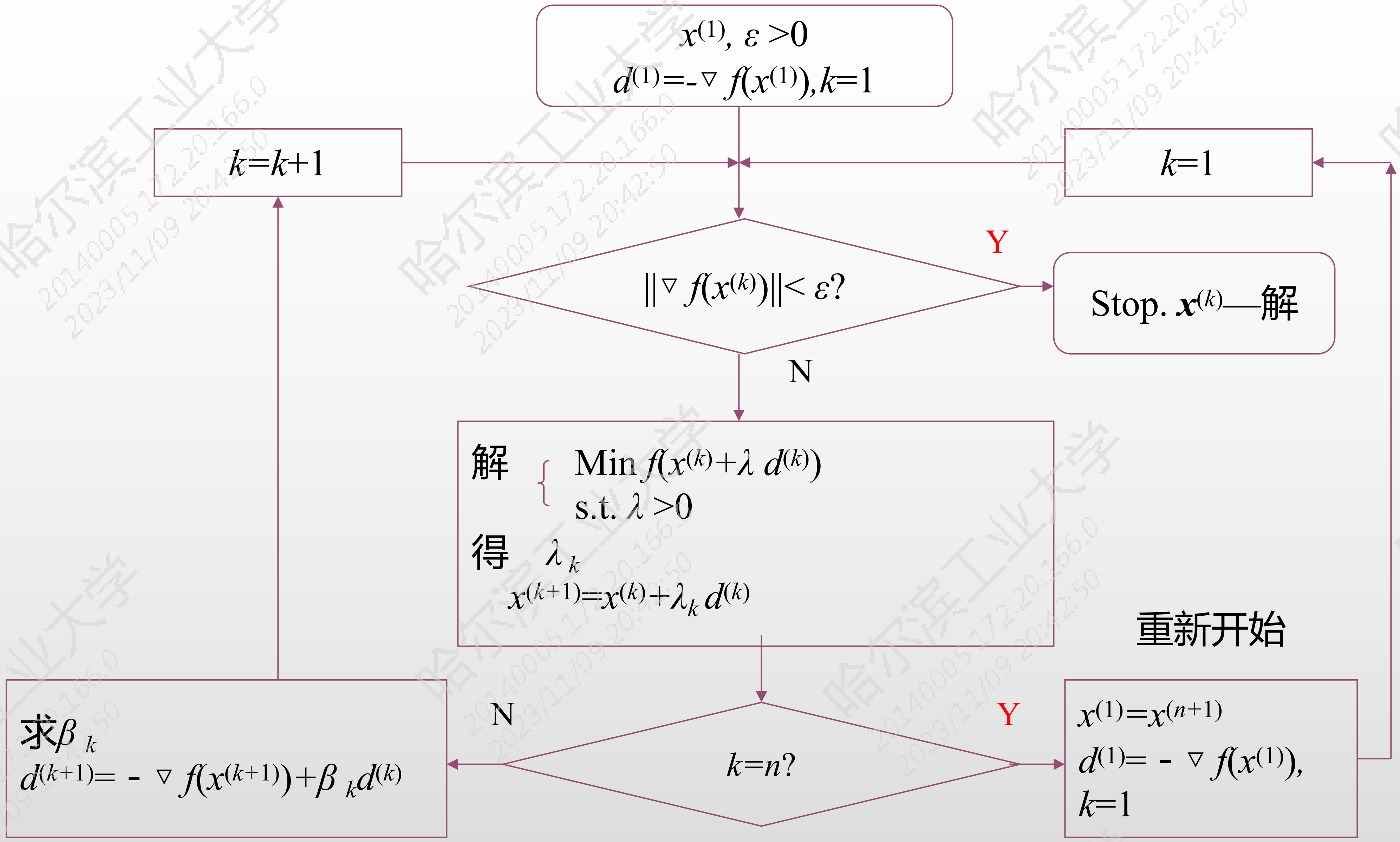
$$\nabla f^T(x^{(k)})d^{(k)}$$
$$= \nabla f^T(x^{(k)})(-\nabla f(x^{(k)}) + \beta_{k-1}d^{(k-1)})$$
$$=- \nabla f^T(x^{(k)})\nabla f(x^{(k)}) < 0$$

（2）每步迭代只需存储若干向量（适用于大规模问题）。

（3）有二次终结性（对于正定二次函数，至多n次迭代可达opt.）。

注：对不同的β_k公式，对于正定二次函数是相等的，对非正定二次函数，有不同的效果，经验上PRP效果较好。

共轭梯度法算法流程图



$x^{(1)}, \varepsilon>0$
$d^{(1)}=-\nabla f(x^{(1)}), k=1$
$k=k+1$
$k=1$
$\|\nabla f(x^{(k)})\|<\varepsilon$?
Y
Stop. $\boldsymbol{x}^{(k)}$—解
N
解 $\begin{cases} \text{Min } f(x^{(k)}+\lambda d^{(k)}) \\ \text{s.t. } \lambda>0 \end{cases}$
得 λ_k
$x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$
重新开始
N
$k=n$?
Y
求β_k
$d^{(k+1)}=-\nabla f(x^{(k+1)})+\beta_k d^{(k)}$
$x^{(1)}=x^{(n+1)}$
$d^{(1)}=-\nabla f(x^{(1)})$,
$k=1$

5.4 牛顿法及其修正

1. 牛顿法

设$f \in C^2(\mathbb{R}^n)$，又设已得到迭代点$x^{(k)}$，将$f(x^{(k)}+s)$在$x^{(k)}$处作Taylor展开得到

$$f(x^{(k)}+s)=f(x^{(k)})+\nabla^T f(x^{(k)})s+1/2s^T\nabla^2 f(x^{(k)})s+o(\|s\|^2)$$

令$s=x-x^{(k)}$，并取$q_k(x)$为

$$q_k(x)=f(x^{(k)})+\nabla^T f(x^{(k)})s+1/2s^T\nabla^2 f(x^{(k)})s$$

则$q_k(x)$是$f(x)$在$x^{(k)}$附近的一个二阶近似。

设$\nabla^2 f(x^{(k)})$正定，则$q_k(x)$有唯一的驻点，即$\nabla q_k(x)=0.$

由此得到(牛顿方向)

$$s^{(k)}=-\left[\nabla^2 f(x^{(k)})\right]^{-1}\nabla f(x^{(k)})$$

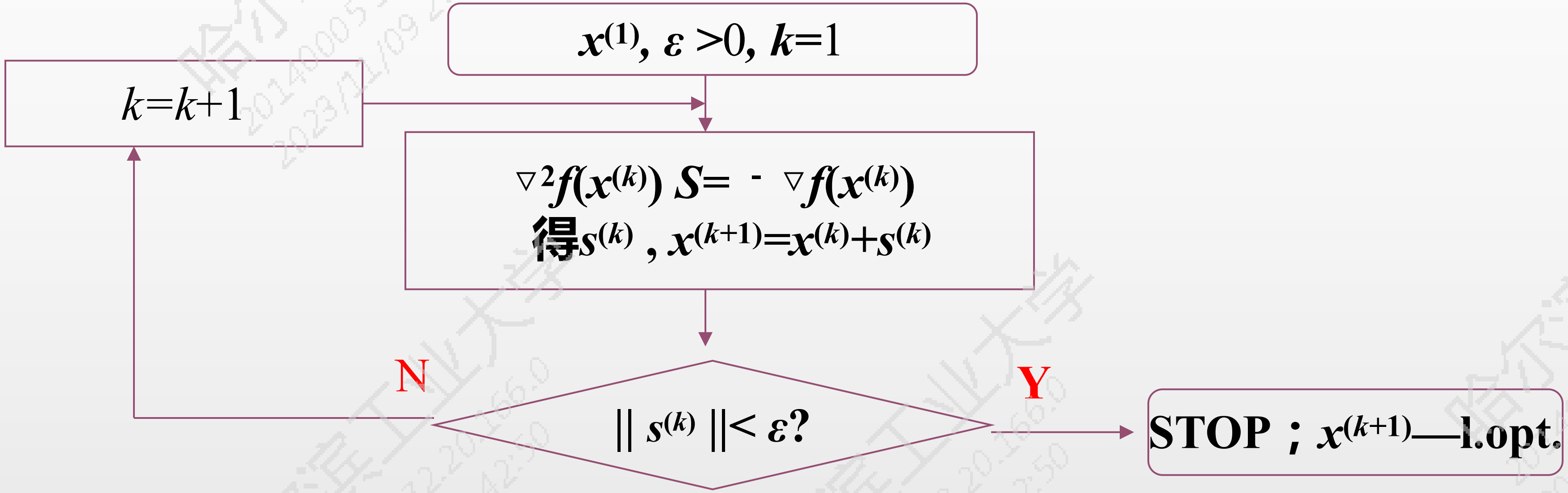
因此

$$x^{(k+1)}=x^{(k)}+s^{(k)}$$

通常$s^{(k)}$是通过求解线性方程组（牛顿方程）得到的

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

牛顿法的算法流程图：



$x^{(1)}, \varepsilon>0, k=1$
$k=k+1$
$\nabla^2 f(x^{(k)})\, S=-\nabla f(x^{(k)})$
得$s^{(k)}$, $x^{(k+1)}=x^{(k)}+s^{(k)}$
N
$\| s^{(k)} \|<\varepsilon$?
Y
STOP ; $x^{(k+1)}$—l.opt.

优点：

(1) 二阶收敛

（当$x^{(k)}$充分接近x^*时,局部函数可用正定二次函数很好地近似，故收敛很快）

(2) 二次终结性：当$f(x)$为正定二次函数时，从任意初始点可一步迭代达到最优解

设$f(x)=(1/2)x^TQx+P^Tx+r$，$Q_{n\times n}$对称正定，$P\in R^n$, $r\in R$。$\forall x^{(1)}$,

$$\nabla f(x^{(1)})=Q x^{(1)}+P \quad , \quad \nabla^2 f(x^{(1)})=Q$$

迭代：$x^{(2)} = x^{(1)} - Q^{-1}(Qx^{(1)}+P) = -Q^{-1}P$ （驻点即opt.)

主要缺点：

(1)局部收敛

(2)用到二阶Hesse矩阵，且要求正定

(3)需计算Hesse矩阵逆或解n阶线性方程组，计算量大

(4)需要$f \in C^2$

2. 修正的牛顿算法

(1) 减少计算量的直接求法

为减小工作量，取m（正整数），使每m次迭代使用同一个Hesse矩阵，迭代公式变为

$$x^{(km+j+1)}=x^{(km+j)}-[\nabla^2 f(x^{(km)})]^{-1}\nabla f(x^{(km+j)})$$

$j=0,1,2,\dots,m-1$，　$k=0,1,2,\dots$

特点：收敛速度随m的增大而下降。

此方法的收敛阶数为$p=(m+1)^{1/m}\le 2$

当 $m=1$时即Newton法，收敛阶数$p=2$

当 $m\to\infty$时 即线性收敛。收敛阶数$p=1$（因为$\lim\limits_{m\to\infty} ln(m+1)^{1/m}=0$)

(2)带线性搜索的Newton法(阻尼牛顿法，阻尼因子λ_k)：

在Newton迭代中，取$d^{(k)}=-[\nabla^2 f(x^{(k)})]^{-1}\nabla f(x^{(k)})$,加入线性搜索：

$$\min \quad f(x^{(k)}+\lambda_k d^{(k)})$$

求得 λ_k，$x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$

特点：可改善局部收敛性，当$d^{(k)}$为函数上升方向时，可向负方向搜索，但可能出现$\pm d^{(k)}$均非下降方向的情况。

(3)Goldstein-Price方法(G-P法)

取 $d^{(k)}=\begin{cases} -[\nabla^2 f(x^{(k)})]^{-1}\nabla f(x^{(k)}), & \nabla^2 f(x^{(k)}) \text{ 正定} \\ -\nabla f(x^{(k)}), & \text{否则} \end{cases}$

采用下列不精确一维搜索： 求λ_k，使其中$\delta\in(0,1/2)$

$1°\ f(x^{(k)}+\lambda_k d^{(k)}) \le f(x^{(k)})+\delta\nabla f^T(x^{(k)})d^{(k)}\lambda_k$

$2°\ f(x^{(k)}+\lambda_k d^{(k)}) \ge f(x^{(k)})+(1-\delta)\nabla f^T(x^{(k)})d^{(k)}\lambda_k$

特点：在一定条件下， G-P法全局收敛。

但当$\nabla^2 f(x^{(k)})$非正定情况较多时，收敛速度降为接近线性。

(4)Levenberg-Marguardt法（L-M法）（强迫Hesse矩阵正定）

主要思想：

用$[\nabla^2 f(x^{(k)}) + \mu E]$取代$\nabla^2 f(x^{(k)})$进行迭代，其中$E$为单位矩阵。$\mu>0$ 使$[\nabla^2 f(x^{(k)}) + \mu E]$正定，$\mu$尽量小。

特点：全局二阶收敛。

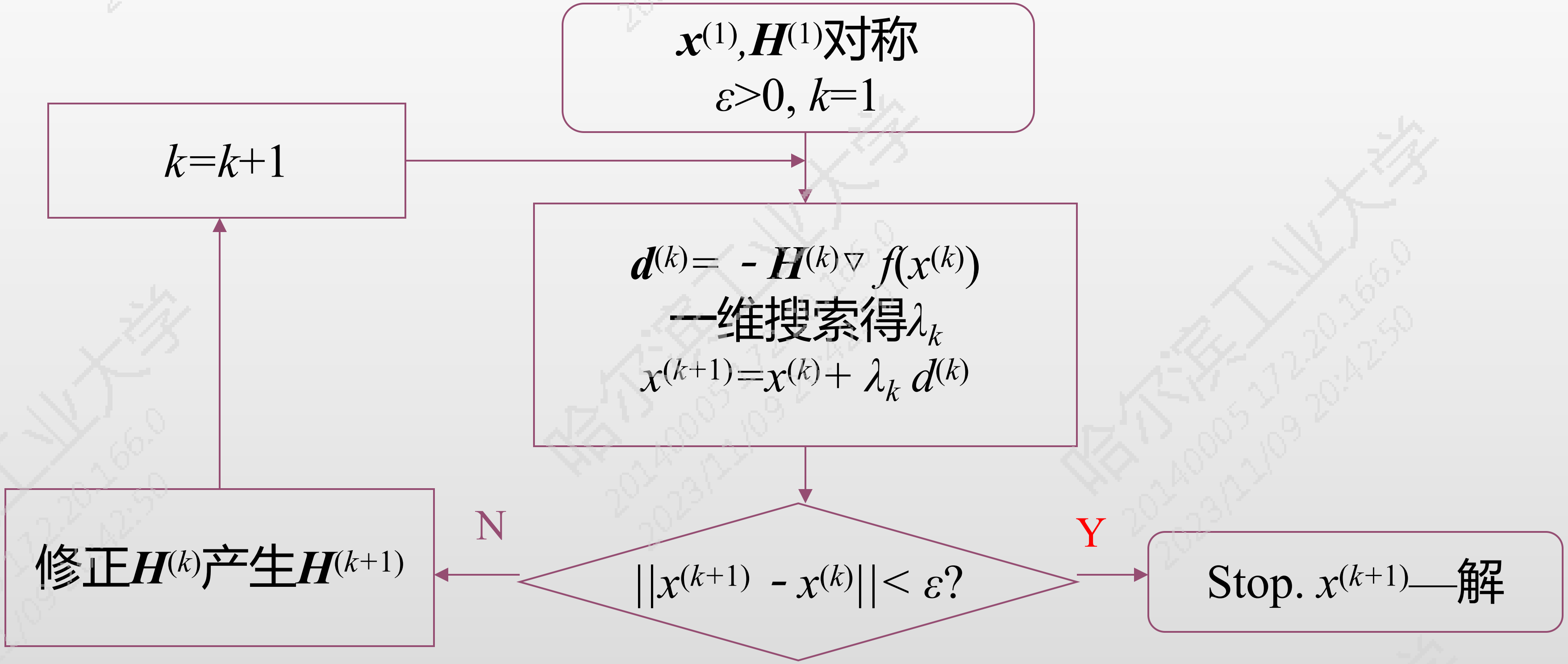
5.5 变尺度法

1.变尺度法的基本思路： $(f)\text{Min}\quad f(x),\quad f:\ R^n \to R$

（1）基本思想

用对称正定矩阵$H^{(k)}$近似$(\nabla^2 f(x^{(k)}))^{-1}$，而$H^{(k)}$的产生是从给定$H^{(1)}$开始逐步修正得到。

（2）算法框图



（3）拟Newton方程

记 $s^{(k)}=x^{(k+1)}-x^{(k)}$，$y^{(k)}=\nabla f(x^{(k+1)})-\nabla f(x^{(k)})$

当 f 为二次函数时：$f(x)=(1/2)x^{T}Bx+c^{T}x+b$

$$\nabla f=Bx+c$$

有 $y^{(k)}=Bs^{(k)}$ 或 $s^{(k)}=B^{-1}y^{(k)}$

称 $Hy=S$ 或 $y=BS$ 为拟Newton方程。

显然，当H 正定时，$B^{-1}=H.$

（4）“近似” $x^{(k+2)}=x^{(k+1)}+\lambda_{k+1}d^{(k+1)}$，其中$d^{(k+1)}=-(\nabla^2 f(x^{(k+1)}))^{-1}\nabla f(x^{(k+1)})$

问题：如何找$\nabla^2 f(x^{(k+1)})$或$(\nabla^2 f(x^{(k+1)}))^{-1}$的近似？

对于$\nabla f(x)$作Taylor展式，舍去高阶项有

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

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

用矩阵$B^{(k+1)}$或$H^{(k+1)}$分别取代$\nabla^2 f(x^{(k+1)})$ 或者 $(\nabla^2 f(x^{(k+1)}))^{-1}$ 使拟Newton方程成立。

$B^{(k+1)}$或$H^{(k+1)}$可看做是对$\nabla^2 f(x^{(k+1)})$或$(\nabla^2 f(x^{(k+1)}))^{-1}$的一种近似。

此种近似H 或 B 不唯一。

（5）变尺度法的主要特点

1）只需用到函数的一阶梯度（Newton法用到二阶Hesse矩阵）；
2）下降算法，故全局收敛；
3）不需求矩阵逆（计算量小）；
4）一般可达到超线性收敛（速度快）；
5）有二次终结性。

2.DFP（Davidon(1959),Fletcher and Powell(1963)）法和 BFGS（Broyden(1970), Fletcher (1970),Goldfarb(1970),Schanno(1970)）法

（1）DFP法

以下省去各量上标，x, s, y, H 表示第k 步的量，$\bar{x}, \bar{s}, \bar{y}, \bar{H}$ 等表示第k+1步的量。

令修正公式为$\bar{H} = H + \alpha u u^T + \beta v v^T, u, v \in \mathbb{R}^n$

代入拟Newton条件$s = \bar{H} y$有

$$s = (H + \alpha u u^T + \beta v v^T) y = Hy + \alpha (u^T y) u + \beta (v^T y) v$$

取：$u = s, v = Hy$

则有 $\alpha = \frac{1}{s^T y}, \beta = -\frac{1}{y^T H y}$

DFP公式：$\bar{H} = H + \frac{s s^T}{s^T y} - \frac{H y y^T H}{y^T H y}$

例. 用DFP法求解 $10x_1^2+x_2^2$

解：取初始点 $x^{(1)}=(1/10,1)^T$, $H^{(1)}=E$ (单位矩阵)

得到如下结果：（计算过程见下页）

$$\nabla f(x)=\begin{pmatrix}20x_1\\2x_2\end{pmatrix},\qquad d=-H\nabla f(x)\qquad s=\bar{x}-x=\lambda d$$

k	$x^{(k)}$	$\nabla f(x^{(k)})$	$H^{(k)}$	$d^{(k)}$	λ_k	$S^{(k)}$	$y^{(k)}$
1	$\begin{pmatrix}\frac{1}{10}\\1\end{pmatrix}$	$\begin{pmatrix}2\\2\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-2\\-2\end{pmatrix}$	$\frac{1}{11}$	$\begin{pmatrix}-\frac{2}{11}\\-\frac{2}{11}\end{pmatrix}$	$\begin{pmatrix}-\frac{40}{11}\\-\frac{4}{11}\end{pmatrix}$
2	$\begin{pmatrix}-\frac{9}{110}\\\frac{9}{11}\end{pmatrix}$	$\begin{pmatrix}-\frac{18}{11}\\\frac{18}{11}\end{pmatrix}$	$\frac{1}{2222}\begin{pmatrix}123&-119\\-119&2301\end{pmatrix}$	$\begin{pmatrix}1\\-10\end{pmatrix}$	$\frac{9}{110}$		
3	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$	停				

计算过程

第一步：

$$d^{(1)}=-H^{(1)}\nabla f(x^{(1)})=-\begin{pmatrix}2\\2\end{pmatrix} \qquad x^{(1)}+\lambda d^{(1)}=\begin{pmatrix}\frac{1}{10} & -2\lambda\\ 1 & -2\lambda\end{pmatrix}$$

一维搜索：
$$\begin{cases}\min\ f(x^{(1)}+\lambda d^{(1)})=10\left(\frac{1}{10}-2\lambda\right)^2+(1-2\lambda)^2\\ \qquad s.t.\quad \lambda\ge 0\end{cases}$$

求 $f(x^{(1)}+\lambda d^{(1)})$的驻点：

$$f'=-40\left(\tfrac{1}{10}-2\lambda\right)-4(1-2\lambda)=0 \qquad 得\ \lambda=\tfrac{1}{11}$$

于是 $$x^{(2)}=x^{(1)}+\lambda_1 d^{(1)}=\left(-\tfrac{9}{110},\tfrac{9}{11}\right)^T$$

第二步：

$$\nabla f(x^{(2)})=\left(-\tfrac{18}{11},\tfrac{18}{11}\right)^T \qquad s^{(1)}=x^{(2)}-x^{(1)}=\left(-\tfrac{2}{11},-\tfrac{2}{11}\right)^T$$

$$y^{(1)}=\nabla f(x^{(2)})-\nabla f(x^{(1)})=\left(-\tfrac{40}{11},-\tfrac{4}{11}\right)^T$$

$$H^{(2)}=H^{(1)}+\frac{s^{(1)}s^{(1)^T}}{s^{(1)^T}y^{(1)}}-\frac{H^{(1)}y^{(1)}y^{(1)^T}H^{(1)}}{y^{(1)^T}H^{(1)}y^{(1)}}$$

$$s^{(1)^T}y^{(1)}=\frac{8}{11}\qquad H^{(1)}y^{(1)}=y^{(1)}\qquad y^{(1)^T}H^{(1)}y^{(1)}=\frac{1616}{121}$$

$$s^{(1)}s^{(1)^T}=\begin{pmatrix}\frac{4}{121} & \frac{4}{121}\\ \frac{4}{121} & \frac{4}{121}\end{pmatrix}\qquad H^{(1)}y^{(1)}y^{(1)^T}H^{(1)}=\begin{pmatrix}\frac{1600}{121} & \frac{160}{121}\\ \frac{160}{121} & \frac{16}{121}\end{pmatrix}$$

于是： $H^{(2)}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\frac{1}{22}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}-\frac{1}{101}\begin{pmatrix}100 & 10\\ 10 & 1\end{pmatrix}=\frac{1}{2222}\begin{pmatrix}123 & -119\\ -119 & 2301\end{pmatrix}$

$d^{(2)}=-H^{(2)}\nabla f(x^{(2)})=\frac{4356}{2222}\begin{pmatrix}1\\ -10\end{pmatrix}$ 取 $d^{(2)}=\begin{pmatrix}1\\ -10\end{pmatrix}$ （与长度无关 ）

$\min\ f(x^{(2)}+\lambda d^{(2)})=10(-\frac{9}{110}+\lambda)^2+(\frac{9}{11}-10\lambda)^2$ 求驻点： $\lambda_2=\frac{9}{110}$

$x^{(3)}=x^{(2)}+\lambda_2 d^{(2)}=(0,0)^T$

第三步： $\nabla f(x^{(3)})=(0,0)^T$ 停。 $x^*=(0,0)^T$

定理：设H对称正定，$s^{T}y>0$那么DFP法产生的$\bar{H}$对称正定。

注：下列各情况下有$s^{T}y>0$：

1° $f(x)$为正定二次函数；

2° 精确一维搜索时；

3° 前章介绍的不精确一维搜索时。

可以证明： DFP法在精确一维搜索前提下，超线性收敛。

（2）BFGS法

若把前面的推导平行地用在$y=Bs$公式上，可得到

$$\bar{B} = B + \frac{yy^T}{y^T s} - \frac{Bss^T B}{s^T Bs} \quad \text{——BFGS的B公式}$$

用此公式求方向时，需用到矩阵求逆或解方程：

$d= -B^{-1}\nabla f(x)$ 或 $Bd= -\nabla f(x)$

由于每次只有秩2的变换，这里的计算量仍可以降下来（但相对于H-公式还是高一些）。

为了得到H-公式，可对上面$\bar{B}$求逆

（利用求逆公式$(M+uv^T)^{-1} = M^{-1} - \frac{M^{-1}uv^T M^{-1}}{1+v^T M^{-1}u}$）：

$$\bar{H} = H + \left(1 + \frac{y^T Hy}{s^T y}\right)\frac{ss^T}{s^T y} - \frac{Hys^T + sy^T H}{s^T y} \quad \text{——BFGS的}H\text{-公式}$$

BFGS法有变尺度法的全部优点，并且在一定条件下可以证明在BFGS法中使用前文中介绍的不精确一维搜索有全局收敛性。

例. 用BFGS法求解 $10x_1^2+x_2^2$

解：取初始点$x^{(1)}=(1/10,1)^T$，$H^{(1)}=E$ (单位矩阵)

得到如下结果：

$$\nabla f(x)=\begin{pmatrix}20x_1\\2x_2\end{pmatrix},\quad d=-H\nabla f(x)\qquad S=\bar{x}-x=\lambda d$$

k	$x^{(k)}$	$\nabla f(x^{(k)})$	$H^{(k)}$	$d^{(k)}$	λ_k	$s^{(k)}$	$y^{(k)}$
1	$\begin{pmatrix}\frac{1}{10}\\1\end{pmatrix}$	$\begin{pmatrix}2\\2\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-2\\-2\end{pmatrix}$	$\frac{1}{11}$	$\begin{pmatrix}-\frac{2}{11}\\-\frac{2}{11}\end{pmatrix}$	$\begin{pmatrix}-\frac{40}{11}\\-\frac{4}{11}\end{pmatrix}$
2	$\begin{pmatrix}-\frac{9}{110}\\\frac{9}{11}\end{pmatrix}$	$\begin{pmatrix}-\frac{18}{11}\\\frac{18}{11}\end{pmatrix}$	$\frac{1}{242}\begin{pmatrix}15&-29\\-29&411\end{pmatrix}$	$\frac{36}{121}\begin{pmatrix}1\\-10\end{pmatrix}$	$\frac{11}{40}$	$\begin{pmatrix}\frac{9}{110}\\-\frac{9}{11}\end{pmatrix}$	$\begin{pmatrix}\frac{18}{11}\\-\frac{18}{11}\end{pmatrix}$
3	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$					

3.Broyden族

当在秩2公式中，α、β 任意选取时可得到不同的公式，经过理论推导，可得到下列结果：

(1)DFP 的 B 公式： $\overline{B} = B + (1 + \frac{s^T Bs}{s^T y})\frac{yy^T}{s^T y} - \frac{Bsy^T + ys^T B}{s^T y}$

DFP 公式与 $BFGS$ 公式，通过用 $s, y, H(B)$分别取代 $y, s, B(H)$ 即可互相得到，称它们 为对偶公式。

(2)令 DFP 法， $BFGS$ 法的 H 公式分别为 $\overline{H}_{DFP}, \overline{H}_{BFGS}$,引入参数 Φ（一般取 $0 \le \Phi \le 1$）有下列公式：

$1^\circ \quad \overline{H}_\Phi = (1 - \Phi)\overline{H}_{DFP} + \Phi \overline{H}_{BFGS}$

$2^\circ \quad \overline{H}_\Phi = \overline{H}_{DFP} + \Phi vv^T$

其中： $v=(y^T Hy)^{1/2}\left(\dfrac{s}{s^T y}-\dfrac{Hy}{y^T Hy}\right)$

$$3^\circ\quad \overline{H}_\Phi=\overline{H}_{DFP}+(s,Hy)A_\Phi\begin{pmatrix} s^T \\ Hy^T \end{pmatrix}$$

其中： $A_\Phi=\begin{pmatrix} (1+\Phi\dfrac{y^T Hy}{s^T y})\dfrac{1}{s^T y} & -\dfrac{\Phi}{s^T y} \\ -\dfrac{\Phi}{s^T y} & -\dfrac{1-\Phi}{y^T Hy} \end{pmatrix}_{2\times 2}$

上述公式均是等价的， 称为*Broyden*族。

显然当$\Phi=0$时即*DFP*公式；

当$\Phi=1$时即*BFGS*公式。

*Broyden*族的任何公式都具备一 般变尺度法的优点。

5.6 直接算法

Min $f(x)$

1.单纯形法及可变多面体算法

（1）单纯形法基本思路

设 $x^{(0)}, x^{(1)}, \dots, x^{(n)}$是$R^n$中$n$+1个点构成的一个当前的单纯形。

比较各点的函数值得到：x_{max}，x_{min}使

$$f(x_{max})=\max\{f(x^{(0)}),f(x^{(1)}),\dots,f(x^{(n)})\}$$

$$f(x_{min})=\min\{f(x^{(0)}),f(x^{(1)}),\dots,f(x^{(n)})\}$$

取单纯形中除去x_{max}点外，其他各点的形心

$$\bar{x}=\frac{1}{n}\left(\sum_{i=0}^{n}x^{(i)}-x_{max}\right)$$

取$x^{(n+1)}$为x_{max}关于$\bar{x}$的反射点：$x^{(n+1)}=\bar{x}+(\bar{x}-x_{max})$

去掉x_{max}，加入$x^{(n+1)}$得到新的单纯形。

重复上述过程。

几点注意：

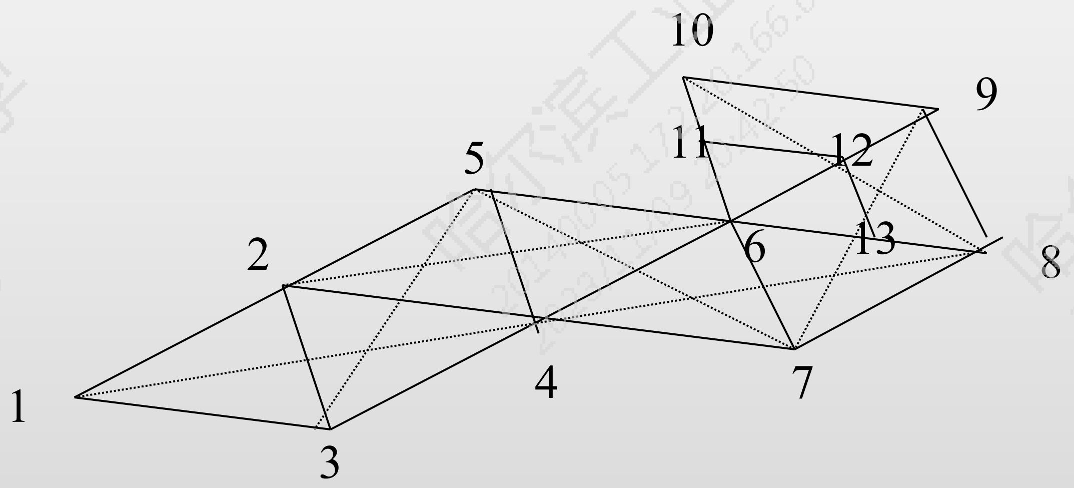
(1)当$x^{(n+1)}$又是新单纯形的最大值点时，取次大值点进行反射；

(2)若某一个点x'出现在连续m个单纯形中的时候，取各点与x'连线的中点（n个）与x'点构成新的单纯形，继续进行。

经验上取 $m \geq 1.65n+0.05n^2$

例如：$n=2$时，可取$m \geq 1.65\times 2+0.05\times 4=3.5$

可取 $m=4$。



优点：不需要求导数，不需要一维搜索。

缺点：无法加速，收敛慢，效果差。

（2）改进单纯形法（可变多面体算法）

设第k步迭代得到$n+1$个点：$x^{(0)},x^{(1)},\ldots,x^{(n)}$，得到$x_{\max}$，$x_{\min}$及

通过下列4步操作选新迭代点：

1° 反射：取反射系数$\alpha>0$(单纯形法中$\alpha=1$)

$$y^{(1)}=\bar{x}+\alpha(\bar{x}-x_{\max})$$

2° 扩展：给定扩展系数$\gamma>1$,计算（加速）

$$y^{(2)}=\bar{x}+\gamma(y^{(1)}-\bar{x})$$

若$f(y^{(1)})<f(x_{\min})$，则若$f(y^{(1)})>f(y^{(2)})$，那么$y^{(2)}$取代$x_{\max}$；否则，$y^{(1)}$取代$x_{\max}$

若$\max\{f(x^{(i)})|\ x^{(i)}\neq x_{\max}\}\geq f(y^{(1)})\geq f(x_{\min})$，$y^{(1)}$取代$x_{\max}$

3° 收缩：若$f(x_{max}) > f(y^{(1)}) > f(x^{(i)})$, $x^{(i)} \neq x_{max}$,计算

$$y^{(3)} = \bar{x} + \beta(y^{(1)} - \bar{x}) \qquad \beta \in (0,1)$$

以$y^{(3)}$取代x_{max} 。

4 ° 减半：若$f(y^{(1)}) > f(x_{max})$, 重新取各点，使$x^{(i)} = x_{min} + 1/2(x^{(i)} - x_{min})$ 得到新单纯形。

经验上：$\alpha=1, 0.4 \le \beta \le 0.6, 2.3 \le \gamma \le 3.0$

有人建议:$\alpha=1, \beta=0.5$ ， $\gamma=2$

算法停机准则取：

$$\frac{1}{n+1}\sum_{i=0}^{n}[f(x^{(i)}) - f(x_{min})]^2 < \varepsilon$$

2.模式搜索法： Hooke & Jeeves （1961）

（1）基本思想与主要过程

- 利用两类移动（探测性移动和模式性移动）进行一步迭代：

探测性移动的目的：探求一个沿各坐标方向的新点并得到一 个“有前途”的方向；

模式性移动的目的：沿上述“有前途”方向加速移动。

- 主要过程：第k步迭代，设已得到$x^{(k)}$

1°探测性移动：

给定步长α_k，设通过模式性移动得到$y^{(0)}$,

依次沿各坐标方向$e^{(i)}=(0,\ldots,\underset{i}{1},0,\ldots,0)^T$

移动α_k步长：$i=0,1,\ldots,n-1$， $\bar{y}=y^{(i)}+\alpha_k e^{(i+1)}$

若$f(\bar{y})<f(y^{(i)})$，则 $y^{(i+1)}=\bar{y}$

否则取 $\bar{y}=y^{(i)}-\alpha_k e^{(i+1)}$

若$f(\bar{y})<f(y^{(i)})$, 则 $y^{(i+1)}=\bar{y}$

否则 $y^{(i+1)}=y^{(i)}$

最后得到$y^{(n)}$。

若$f(y^{(n)})<f(x^{(k)})$, 令$x^{(k+1)}=y^{(n)}$。

2°模式性移动

$x^{(k+1)}-x^{(k)}$为一个有前途的方向，取

$y^{(0)}=x^{(k+1)}+(x^{(k+1)}-x^{(k)})=2x^{(k+1)}-x^{(k)}$

[但不一定保证$f(y^{(0)})<f(x^{(k+1)})$]

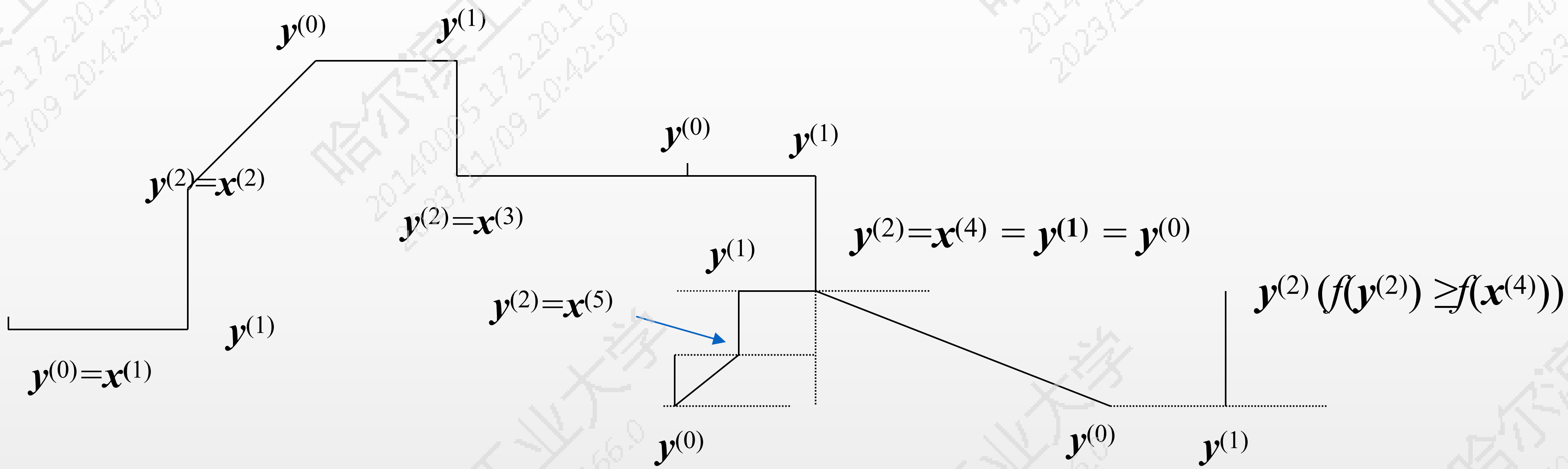
3 °几点措施：

①若探测性移动得到$y^{(n)}$使$f(y^{(n)})\geq f(x^{(k)})$, 则跳过模式性移动而令$y^{(0)}=x^{(k)}$ 重新进行探测性移动

②若$y^{(n)}=y^{(0)}$（即每一个坐标方向的移动都失败），减小α_k，重复上述过程。

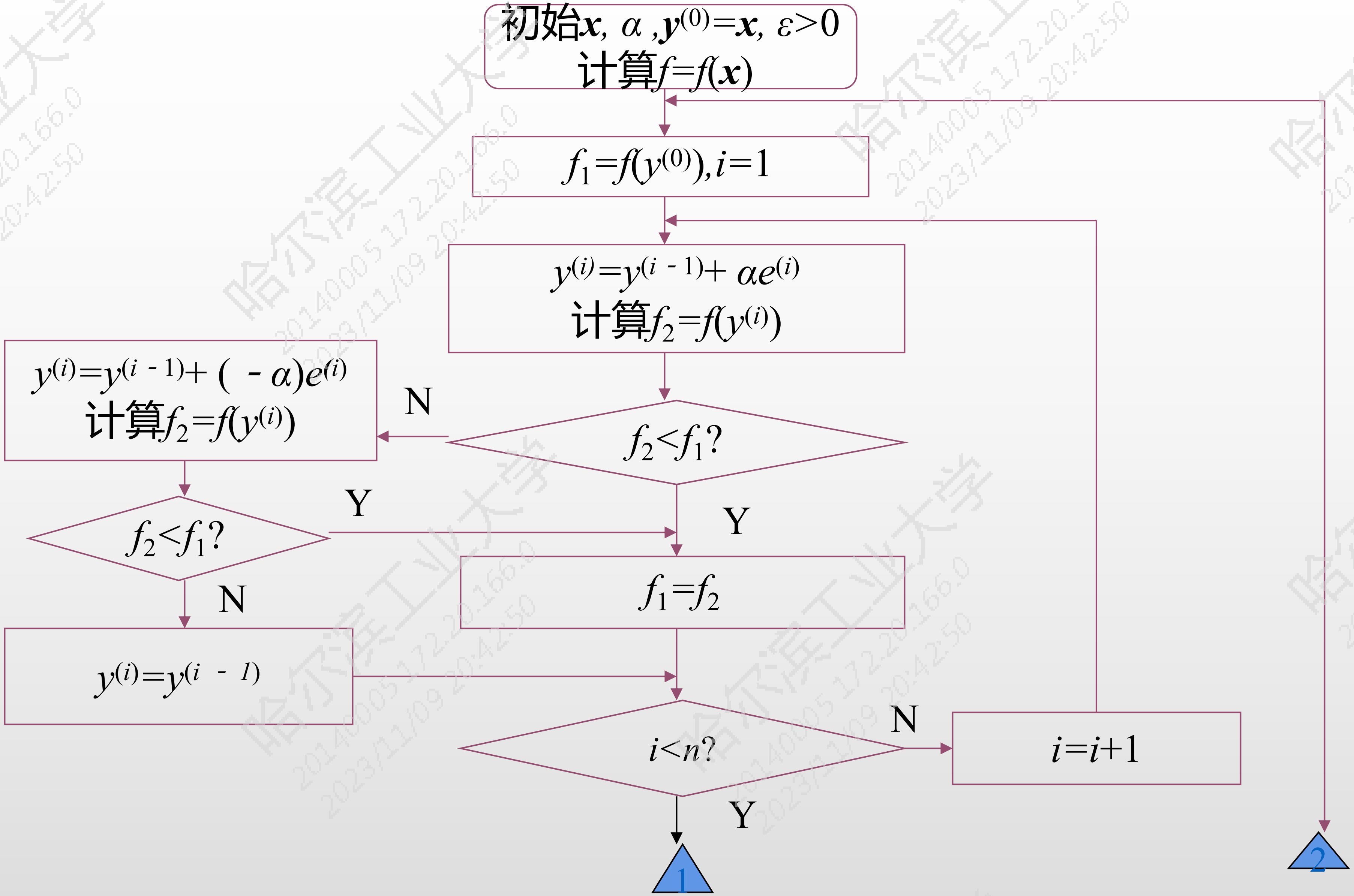
③当进行到α_k充分小（$\alpha_k<\varepsilon$）时，终止计算。最新的迭代点$x^{(k)}$为解。

例：

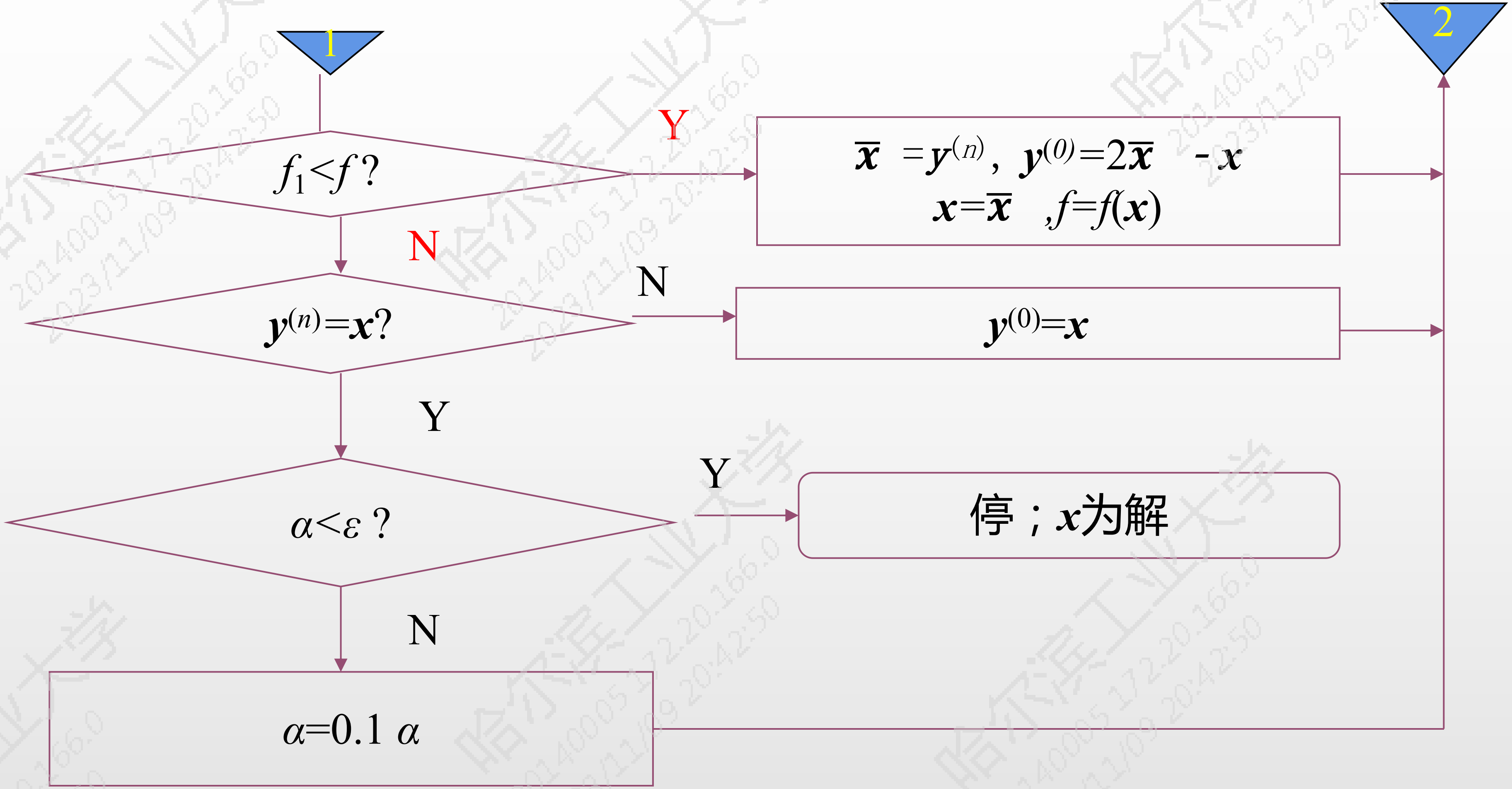


$y^{(0)}$
$y^{(1)}$
$y^{(2)}=x^{(2)}$
$y^{(0)}$
$y^{(1)}$
$y^{(2)}=x^{(3)}$
$y^{(2)}=x^{(4)} = y^{(1)} = y^{(0)}$
$y^{(1)}$
$y^{(2)}=x^{(5)}$
$y^{(2)}\ (f(y^{(2)}) \geq f(x^{(4)}))$
$y^{(1)}$
$y^{(0)}=x^{(1)}$
$y^{(0)}$
$y^{(0)}$
$y^{(1)}$

（2）算法：



初始x, α ,$y^{(0)}=x$, $\varepsilon>0$
计算$f=f(x)$
$f_1=f(y^{(0)}),i=1$
$y^{(i)}=y^{(i-1)}+\alpha e^{(i)}$
计算$f_2=f(y^{(i)})$
$f_2<f_1$?
N
$y^{(i)}=y^{(i-1)}+(-\alpha)e^{(i)}$
计算$f_2=f(y^{(i)})$
Y
$f_2<f_1$?
Y
N
$f_1=f_2$
$y^{(i)}=y^{(i-1)}$
$i<n$?
N
$i=i+1$
Y
1
2


1
2
$f_1<f$?
Y
$\bar{x}=y^{(n)}$, $y^{(0)}=2\bar{x}-x$
$x=\bar{x}$, $f=f(x)$
N
N
$y^{(n)}=x$?
$y^{(0)}=x$
Y
Y
$\alpha<\varepsilon$?
停；x为解
N
$\alpha=0.1\,\alpha$