第六章 非线性规划

Non-linear Programming

非线性规划

- ❖基本概念
- ❖凸函数和凸规划
- ❖一维搜索方法
- ❖无约束最优化方法
- ❖约束最优化方法

第一节 基本概念

1、非线性规划模型：

➢数学规划模型的一般形式：

$$\begin{cases} \min \quad f(x) \\ s.t. \quad g_i(x) \le 0, \quad i=1,\cdots,p \\ \qquad\quad h_i(x)=0, \quad j=1,\cdots,q \end{cases}$$

其中,$x=(x_1,x_2,\ldots x_n)^T$，$f(x),g_i(x),h_j(x)$为x的实值函数

简记为MP(Mathematical Programming)

无约束问题：

$$\min\ f(x), \quad x \in R^n$$

约束问题：

$$\min\ f(x) \quad s.t.\ \begin{aligned} & h_i(x) = 0, \quad i = 1, \cdots, m \\ & g_j(x) \geq 0, \quad j = 1, \cdots, l \end{aligned}$$

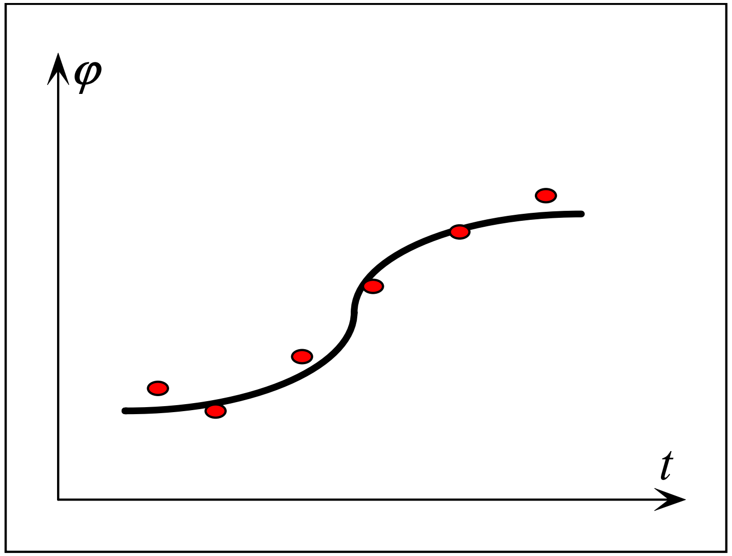
$$\Leftrightarrow\ \min\ f(x) \quad s.t.\ h(x) = 0,\ g(x) \geq 0$$

$$\Leftrightarrow\ \min\ f(x) \quad s.t.\ x \in D = \{x \mid h(x) = 0, g(x) \geq 0\}$$

D 称为可行域。

非线性规划问题举例

例1 曲线的最优拟合问题



已知某物体的温度φ与时间 t 之间有如下形式的经验函数关系：

$$\varphi = c_1 + c_2 t + e^{c_3 t} \qquad (*)$$

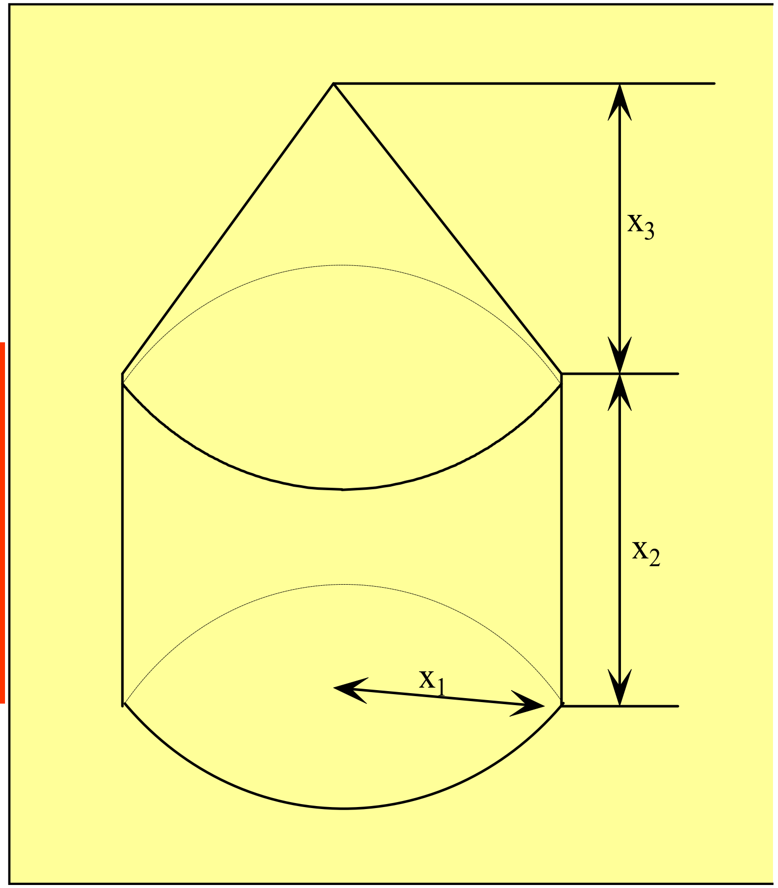
其中c_1，c_2，c_3是待定参数。现通过测试获得 n 组φ与 t 之间的实验数据(t_i, φ_i)，i=1，2，…,n。试确定参数c_1，c_2，c_3，使理论曲线(*)尽可能地与 n 个测试点(t_i, φ_i)拟合。

$$\min \sum_{i=1}^{n} [\varphi_i - (c_1 + c_2 t_i + e^{c_3 t_i})]^2$$

例2 构件容积问题

设计一个右图所示的由圆锥和圆柱面围成的构件，要求构件的表面积为S，圆锥部分的高h和圆柱部分的高x_2之比为a。确定构件尺寸，使其容积最大。

$$
\begin{cases}
\max\ V = (1 + a/3)\pi x_1^2 x_2 \\
s.t.\ \ \pi x_1 \sqrt{x_1^2 + a^2 x_2^2} + 2\pi x_1 x_2 + \pi x_1^2 = S \\
\qquad x_1 \ge 0, x_2 \ge 0
\end{cases}
$$



2、非线性规划方法概述

- 微分学方法的局限性：

实际的问题中，函数可能是不连续或者不可微的。

需要解复杂的方程组，而方程组到目前仍没有有效的算法。

实际的问题可能含有不等式约束，微分学方法不易处理。

基本概念

整体最优解 x^*： $f(x^*) \le f(x)$ ， $\forall\, x \in D$

邻域 $N(\hat{x},\delta) = \{\, x \mid \|x - \hat{x}\| < \delta \,\}$

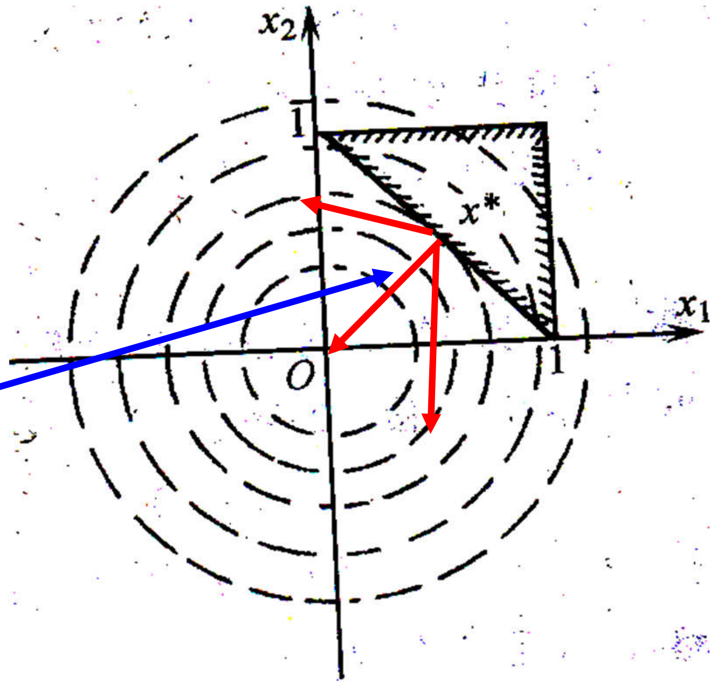
局部最优解 x^*： $f(x^*) \le f(x)$， $\forall\, x \in N(x^*,\delta) \cap D$

定义：下降方向

设$f:R^n \to R^1, \bar{x}\in R^n, p\in R^n, p\neq 0$,若存在$\delta>0$，使

$$f(\bar{x}+tp)<f(\bar{x}),\ \ \forall t\in(0,\delta)$$

则称向量 p是函数 $f(x)$ 在点 $\bar{x}$处的下降方向。

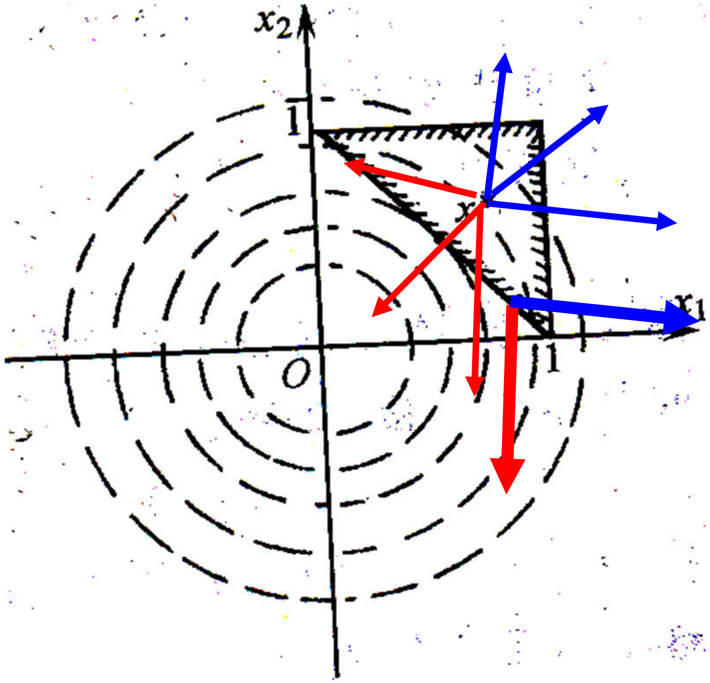


若$f(x)$在$\bar{x}$可导，则$-\nabla f(\bar{x})$就是$f(x)$在$\bar{x}$处下降最快的方向。

定义 设$X \subset R^n, \bar{x} \in X, p \in R^n, p \neq 0$,若存在t>0，使得

$$\bar{x} + tp \in X$$

则称向量 p是点$\bar{x}$处关于X的可行方向。



解非线性规划问题，关键在于找到某个方向，使得在此方向上，目标函数得到下降，同时还是可行方向。

这样的方向称为可行下降方向。

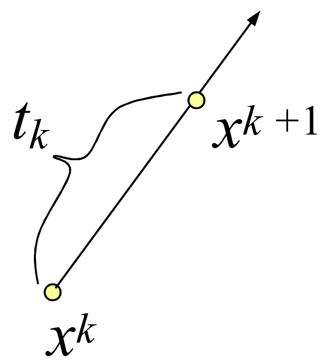
非线性规划基本下降迭代格式

第一步: 给定初始点 x^0 ,设 $k=0$;

第二步: 按照某种规则确定搜索方向 p^k ;

第三步: 按照某种规则确定搜索步长 t_k ;

第四步: 计算 $\boldsymbol{x}^{k+1}=\boldsymbol{x}^{k}+\boldsymbol{t}_{k}\boldsymbol{p}^{k}$;



检查得到的 x^{k+1} 是否满足终止准则：是，$x^*=x^{k+1}$ ，终止； 否，$k=k+1$，转步骤 第二步。

迭代算法中如何确定搜索步长 t_k？

(1) 定步长，即取 t_k 等于某一常数。

(2) 可接受点法，即任意选取能使目标函数值下降的 t_k。

(3)（精确）一维搜索，即求解一元函数极小化问题：

$$f(x^k + t_k p^k) = \min_{t \geq 0} f(x^k + t p^k)$$

性质：若 $f(x)$ 连续可微，t_k 由精确一维搜索确定，则

$$\nabla f(x^{k+1})^T p^k = 0$$

证明：考虑函数 $\phi(t) = f(x^k + t p^k)$，则 t_k 是 $\phi(t)$ 的极小点，所以 $\phi'(t_k) = \nabla f(x^k + t_k p^k)^T p^k = 0$，

即 $\nabla f(x^{k+1})^T p^k = 0$。

若迭代算法产生的点列 $\{x^k\}$ 满足

$$f(x^0) > f(x^1) > \cdots > f(x^k) > f(x^{k+1}) > \cdots$$

称为下降算法。

若迭代点列 $\{x^k\}$有子序列收敛于问题的解 x^*，则称该迭代算法收敛。

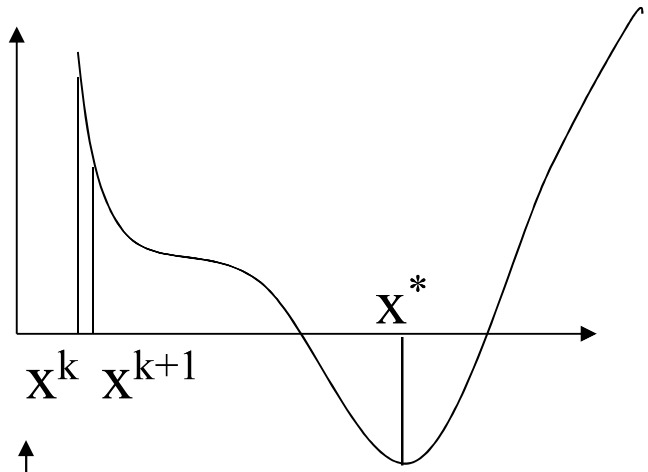
若对于任意初始点 x^0，迭代算法收敛，则称为全局收敛；

若仅当 x^0 属于 x^* 的某个邻域时，迭代算法收敛，则称为局部收敛。

终止准则：

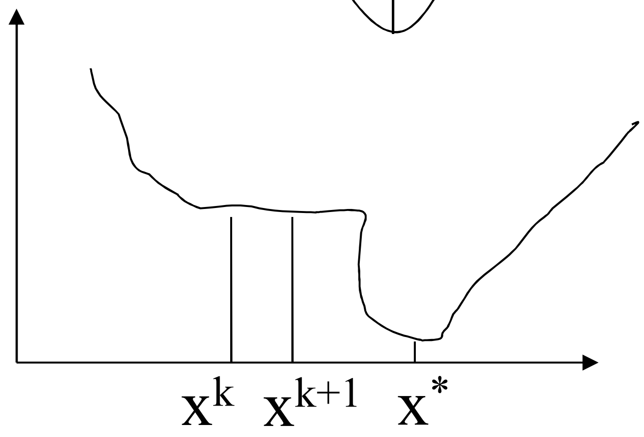
(1) $\| x^{k+1} - x^{k} \| < \varepsilon$

(2) $| f(x^{k+1}) - f(x^{k}) | < \varepsilon$



(3) $\dfrac{\| x^{k+1} - x^{k} \|}{\| x^{k} \|} < \varepsilon \quad (\| x^{k} \| > \varepsilon_0)$

(4) $\dfrac{| f(x^{k+1}) - f(x^{k}) |}{| f(x^{k}) |} < \varepsilon \quad (| f(x^{k}) | > \varepsilon_0)$



(5) $\| \nabla f(x^{k}) \| < \varepsilon$

注： $(3) \Rightarrow (1)$ $\quad \left\| x^{k+1} - x^{k} \right\| < \varepsilon \ <\neq> \ \left\| x^{k+1} - x^{*} \right\| < \varepsilon$

$(4) \Rightarrow (2)$

第二节 凸函数和凸规划

1、凸函数及其性质：

定义 凸集：设 $S \subseteq R^n$，若任给 $x^1, x^2 \in S$ 和 $0 \le \lambda \le 1$，

$\lambda x^1 + (1-\lambda)x^2 \in S$，则称 S 是凸集。

定义：凸函数

设$S \subset R^n$是非空凸集，$f: S \to R^1$,若对$\forall \alpha \in (0,1)$ 有

$$f(\alpha x^1 + (1-\alpha)x^2) \le \alpha f(x^1) + (1-\alpha) f(x^2), \forall x^1, x^2 \in S$$

则称f是S上的凸函数，或f在S上是凸的。

例 $f(x) = \alpha^T x + \beta$,其中$\alpha, x \in R^n, \beta \in R^1$即是凸的也是凹的。

例 $f(x) = \|x\|$其中$x \in R^n$是凸函数

严格凸函数

若$f(\alpha x^1+(1-\alpha)x^2)<\alpha f(x^1)+(1-\alpha)f(x^2), \forall x^1, x^2 \in S$

则称f是S上的严格凸函数，或f在S上是严格凸的。

若$-f$是S上的(严格)凸函数，称f是S上的(严格)凹函数,或f在S上是(严格)凹的。

性质 $f(x)$为凸函数

$$\Leftrightarrow f(x^2) \geq f(x^1)+\nabla f(x^1)^T(x^2-x^1) \quad \forall x^1, x^2$$

$\Leftrightarrow \forall x$，Hesse阵$\nabla^2 f(x)$半正定。

定理　设$S\subset R^n$是非空开凸集，$f:S\to R^1$二阶连续可导，则f是 S上的凸函数的充要条件是

$$\nabla^2 f(x)\text{在S上是半正定的。}$$

当$\nabla^2 f(x)$在S上是正定矩阵时，f是S上的严格凸函数（此时,逆命题不成立）

$\nabla^2 f(x)$称为$Hesse$矩阵，

$$\nabla^2 f(x)=\begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \cdots & & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}$$

例 验证下列MP是凸规划

$$\begin{cases} \min\ f(x_1,x_2,x_3)=2x_1^2+x_2^2+2x_3^2+x_1x_3-x_1x_2+x_1+2x_2 \\ s.t.\ \ g_1(x)=x_1^2+x_2^2-x_3\le 0 \\ \qquad\ \ g_2(x)=x_1+x_2+2x_3\le 16 \\ \qquad\ \ g_3(x)=-x_1-x_2+x_3\le 0 \end{cases}$$

解： （1）目标函数是不是凸函数？

（2）$g_i(x)$是不是凸函数？

➢关于凸函数的一些结论

定理： 设$S\subset R^n$是非空凸集

(1)若f是S上的凸函数， $\alpha\geq 0$， 则αf是S上的凸函数；

(2)若f_1, f_2是S上的凸函数 ，f_1+f_2是S上的凸函数。

定理：设$S\subset R^n$是非空凸集,f是凸函数,$c\in R^1$,则集合

$$H_S(f,c)=\{x\in S \mid f(x)\leq c\}$$ 是凸集。

函数f在集合S上关于c的水平集

凸规划

(1) $\min f(x) \quad s.t. \ x \in D$

其中 $f(x)$ 为凸函数， D 为非空凸集。

(2) 对于非线性规划(MP),

$$\begin{cases} \min \ f(x) \\ s.t. \ g_i(x) \le 0 \quad i=1,\cdots,p \\ \quad h_j(x)=0 \quad j=1,\cdots,q \end{cases}$$

凸函数

线性函数

若$g_i(x)$皆为R^n上的凸函数,$h_j(x)$皆为线性函数，并且f是X上的凸函数，则（MP）是凸规划。

定理 凸规划的任一局部最优解都是它的整体最优解。

二次规划

$$\min \ f(x)=\frac{1}{2}x^TQx+q^Tx+c$$

$$s.t. \quad Ax=b, \ x\geq 0$$

其中 $x\in R^n, \ Q\in R^{n\times n}, q\in R^n, \ c\in R^1,$

$A\in R^{m\times n}, b\in R^m$

$$\nabla f(x)=Qx+q\,, \quad \nabla^2 f(x)=Q$$

极小点的判定条件

必要条件：若 $f(x)$ 在开区域 D 内连续可微，且 $x^* \in D$ 为局部极值点，则 $\nabla f(x^*) = 0$。

充分条件：若 $f(x)$ 在开区域 D 内二阶连续可微，$x^* \in D$，$\nabla f(x^*) = 0$，且 $\nabla^2 f(x^*)$ 正定，则 x^* 为严格局部极小点。

$$f(x)=f(x^*)+\nabla f(x^*)^T(x-x^*)+\frac{1}{2}(x-x^*)^T\nabla^2 f(x^*)(x-x^*)+o(\|x-x^*\|^2)$$

$\nabla^2 f(x^*)$ 正定 $\Rightarrow$ 函数在极小点附近的等值线为近似的同心椭圆。

正定二次函数 $f(x) = \frac{1}{2} x^T Qx + q^T x + c$ 有唯一极小点 $x^* = -Q^{-1} q$.

凸规划的主要性质

- 可行解集为凸集
- 最优解集为凸集(如果存在)
- 任何局部最优解也是全局最优解
- 若目标函数为严格凸函数，且最优解存在，则其最优解必唯一

第三节 一维搜索方法

在迭代法中确定搜索步长 t_k 需要一维搜索：

$$f(x^k + t_k p^k) = \min_{t} f(x^k + t p^k)$$

考虑一元函数极小化问题： $\min\limits_{t \in R} f(t)$

方法： 首先定出一个包含 $f(t)$ 的极小点的区间，然后不断缩小区间的长度，当区间充分小时，取其中的一点作为近似极小点。

如何确定一个包含极小点的区间？

进退法：从一点出发，按一定的步长寻找函数值呈现“高 - 低 - 高”的三点，若一个方向不成功，就退回来，然后沿相反方向寻找。

黄金分割法（0.618 法）

二次逼近法（抛物线插值）

4、Newton法

进退法：从一点出发，按一定的步长寻找函数值呈现“高 - 低 - 高”的三点，若一个方向不成功，就退回来，然后沿相反方向寻找。

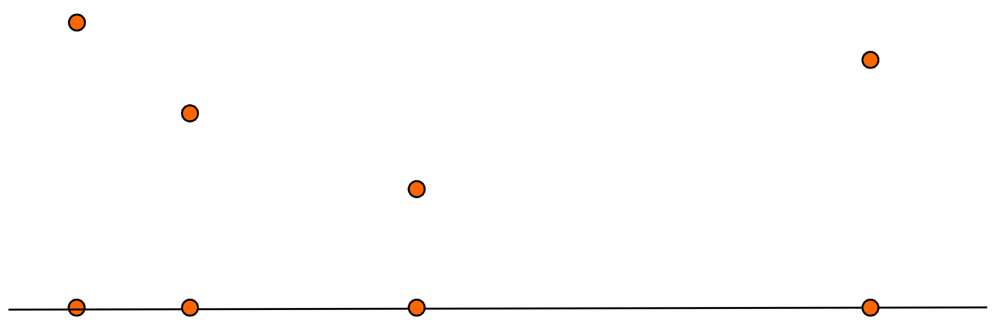
任取初始点 t_0，计算 $f(t_0)$，取步长 $h>0$，计算 $f(t_0+h)$，

若 $f(t_0)>f(t_0+h)$，令 $t_1=t_0+h$，$h=2h$，计算 $f(t_1+h)$，

若 $f(t_1)>f(t_1+h)$，令 $t_2=t_1+h$，$h=2h$，计算 $f(t_2+h)$，

如此继续下去，直到对某个 k 得到 $f(t_k)<f(t_k+h)$，

此时得到一个包含极小点的区间 $[t_{k-1}, t_k+h]$；



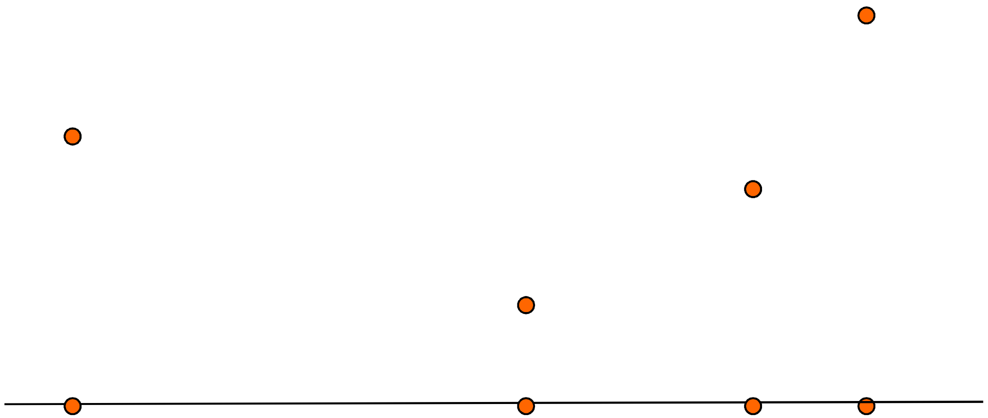
若一开始 $f(t_0) < f(t_0 + h)$,

令 $t_1 = t_0$, $t_0 = t_0 + h$, $h = 2h$, 计算 $f(t_1 - h)$,

若 $f(t_1 - h) < f(t_1)$, 则令 $t_2 = t_1 - h$, $h = 2h$,计算 $f(t_2 - h)$,

如此继续下去, 直到对某个 k 得到 $f(t_k - h) > f(t_k)$,

此时得到一个包含极小点的区间 $[t_k - h , t_{k-1}]$ 。



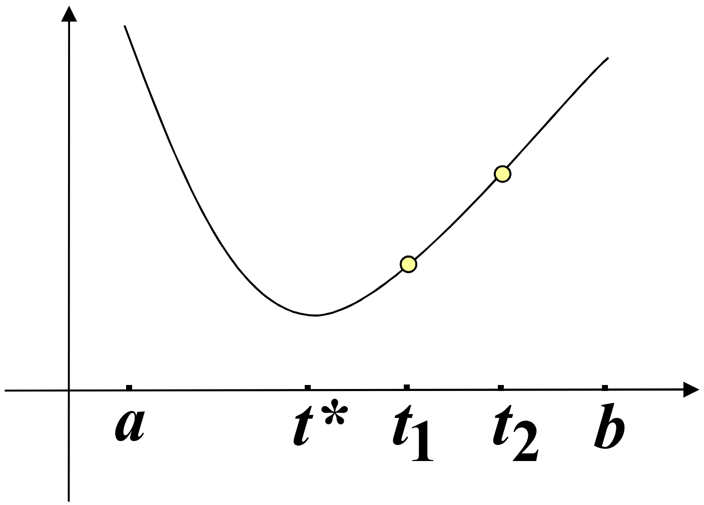
黄金分割法（0.618 法）

解一元函数极小化问题 $\min\limits_{a \le t \le b} f(t)$

假设：$f(t)$ 在$[a,b]$ 上是单谷函数，即 $f(t)$ 在 $[a,b]$ 上有唯一极小点（设为 t^*）且

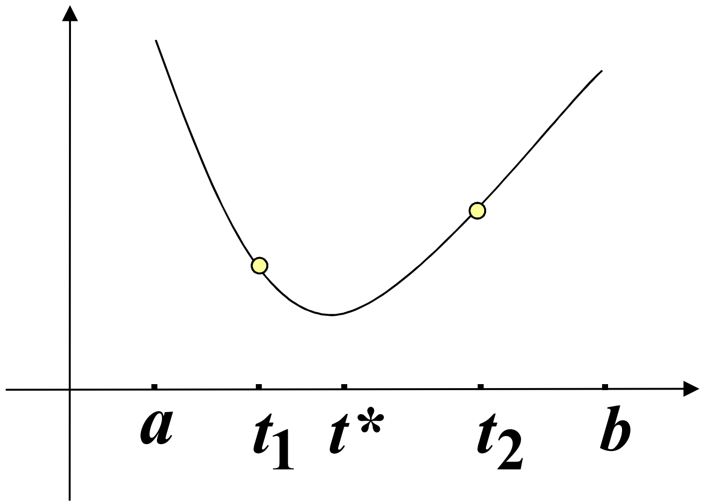
$a \le t_1 < t_2 \le t^*$ 时，$f(t_1) > f(t_2)$；

$t^* \le t_1 < t_2 \le b$ 时，$f(t_1) < f(t_2)$。



在$[a,b]$ 上任取两点 $t_1<t_2$，
计算 $f(t_1)$ 、 $f(t_2)$ ，则

若 $f(t_1)<f(t_2)$，则 $t^*\in[a,t_2]$;
若 $f(t_1)\geq f(t_2)$，则 $t^*\in[t_1,b]$，



即通过比较 $f(t_1)$、$f(t_2)$，可将 t^* 的搜索区间从$[a,b]$缩短为$[a,t_2]$或$[t_1,b]$。为了进一步缩短搜索区 间，需比较 $[a,t_2]$或$[t_1,b]$中两点的函数值。

由于 $t_1\in[a,t_2]$，$t_2\in[t_1,b]$，即保留下来的区间中总 含有一个已计算过函数值 的点，因此要进一步缩 短搜索区间，只需再取一点计 算函数值即可。

n 次函数值计算可使搜索区间缩短 $n-1$ 次。

下面推导黄金分割法如何选取计算点 t_1、t_2。

两个条件:

(1) 由于在比较 $f(t_1)$ 和 $f(t_2)$ 之前不知道 $[a, t_2]$ 和 $[t_1, b]$ 中哪一个将保留下来，因此限定 t_1, t_2 为 $[a, b]$ 中对称的计算点，即满足条件 $t_2 - a = b - t_1$.

(2) 希望每次迭代保持相同的区间缩短率，

当前迭代的区间缩短率 $\dfrac{t_2 - a}{b - a} = \dfrac{b - t_1}{b - a} = \omega$

即 $t_1 = b - \omega(b - a), \quad t_2 = a + \omega(b - a)$. (1)

不妨设当前迭代后保留下来的区间为 $[a, t_2]$,

下一步迭代的计算点为 $t_1' < t_2'$ (其中一个为 t_1),

则下一步迭代的区间缩短率为

$$\frac{t_2' - a}{t_2 - a} = \frac{t_2 - t_1'}{t_2 - a} = \omega$$

所以 $t_2' - a = t_2 - t_1' = \omega(t_2 - a) = \omega^2(b - a)$.

若 $t_1' = t_1$，则 $t_2 - t_1 = \omega^2(b - a)$,

而由 (1)，$t_2 - t_1 = (2\omega - 1)(b - a)$,

因此 $\omega^2 = 2\omega - 1$， $\omega = 1$，但这不可能 .

设 $t_2' = t_1$，则 $t_1 - a = \omega^2(b - a)$,

而由 (1)，$t_1 - a = (1 - \omega)(b - a)$,

所以 $\omega^2 = 1 - \omega$， $\omega = \frac{\sqrt{5} - 1}{2} \approx 0.618$.

对保留下来的区间为 $[t_1, b]$ 的情形可类似考虑，但此时应 $t_1' = t_2$.

黄金分割法的计算步骤：

(1) 给定初始区间 $[a,b]$，精度要求 $\varepsilon>0$.

(2) 取初始计算点

$$t_1=a+0.382(b-a)=b-0.618(b-a),$$

$$t_2=a+0.618(b-a)=b-0.382(b-a),$$

计算 $f(t_1)$ 、 $f(t_2)$ 。

(3) 若 $f(t_1)<f(t_2)$，转 (4)，否则转 (5) .

(4) 若 $t_2-a<\varepsilon$，停止，输出 t_1；否则，令

$b=t_2$，$t_2=t_1$，$t_1=a+0.382(b-a)=b-0.618(b-a)$，

计算 $f(t_1)$，转(3)。

(5) 若 $b-t_1<\varepsilon$，停止，输出 t_2；否则，令

$a=t_1$，$t_1=t_2$，$t_2=a+0.618(b-a)=b-0.382(b-a)$，

计算 $f(t_2)$，转(3)。

黄金分割法的迭代效果：

每次迭代的区间缩短率 为 $\frac{\sqrt{5}-1}{2}$，

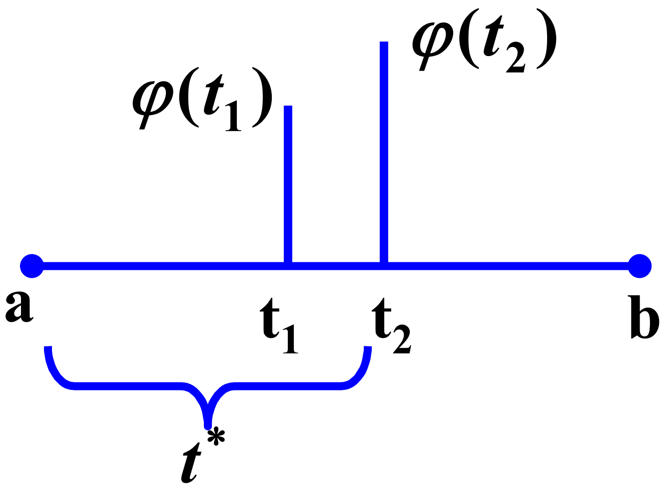
n 次迭代（计算 $n+1$ 个点的函数值）后区间 缩短为

初始区间长度的 $\left(\frac{\sqrt{5}-1}{2}\right)^n$ 倍。

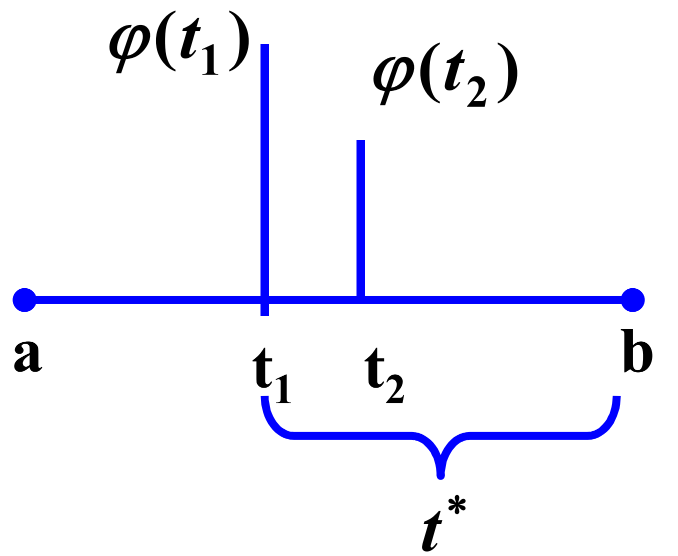
其它试探点算法：Fibonacci 算法

➢假定：已经确定了单谷区间[a,b]

$$\left.\begin{array}{l} \min\limits_{t\geq 0} \varphi(t) \\ \min\limits_{0\leq t\leq t_{\max}} \varphi(t) \end{array}\right\} \Longrightarrow \min\limits_{a\leq t\leq b} \varphi(t)$$

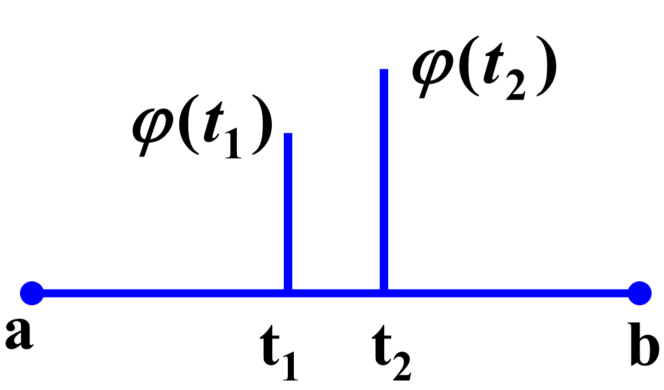


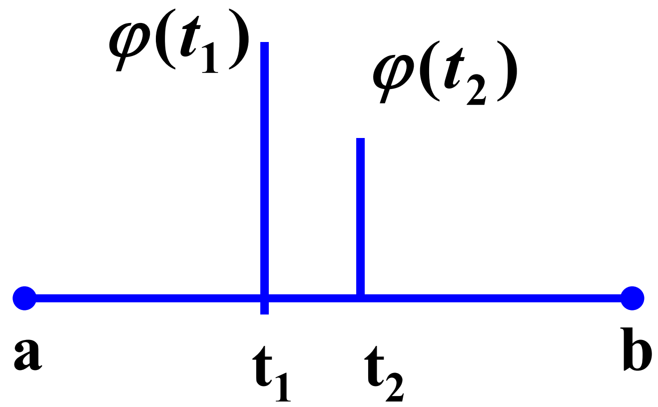
新搜索区间为$[a, t_2]$



新搜索区间为$[t_1, b]$

➢区间缩小比例的确定：





区间缩短比例为$(t_2-a)/(b-a)$　　缩短比例为$(b-t_1)/(b-a)$

缩短比例满足：

每次插入搜索点使得两个区间$[a, t_2]$和$[t_1, b]$相等；

每次迭代都以相等的比例缩小区间。

缩短比例$\omega=\dfrac{\sqrt{5}-1}{2}\approx 0.618$　0.618法

第 1 步　确定单谷区间[a,b]，给定最后区间精度 $\varepsilon > 0$；

第 2 步　计算最初两个探索点

$$t_1 = a + 0.382(b-a) = b - 0.618(b-a)$$

$$t_2 = a + 0.618(b-a)$$

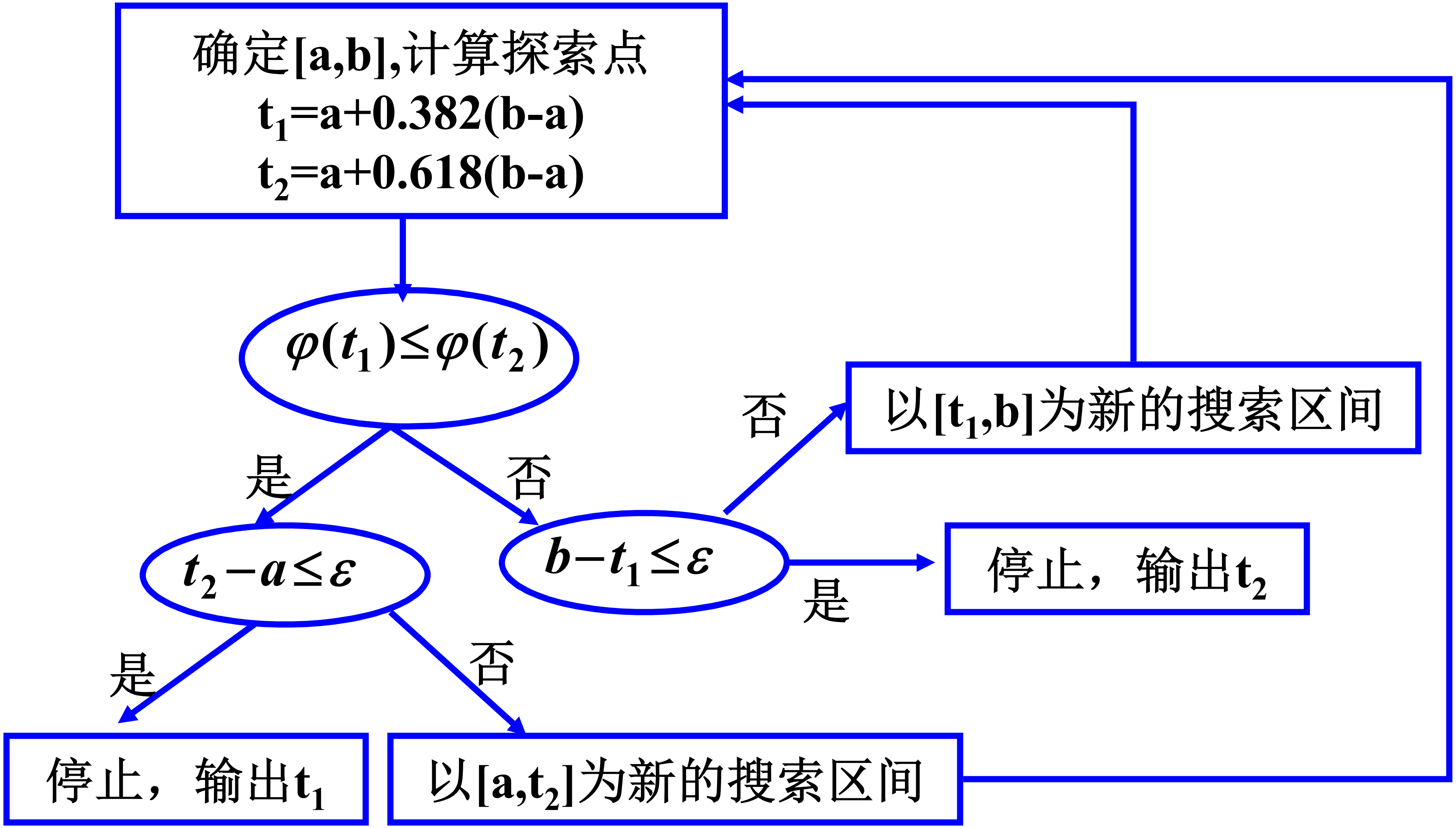
并计算 $\varphi_1 = \varphi(t_1)$，$\varphi_2 = \varphi(t_2)$；

第 3 步　若 $\varphi_1 \le \varphi_2$，转第 4 步。否则转第 5 步；

第 4 步　若 $t_2 - a \le \varepsilon$，停止迭代，输出 t_1。否则令 $b := t_2$，$t_2 := t_1$，$t_1 := b - 0.618(b-a)$，$\varphi_2 := \varphi_1$，计算 $\varphi_1 = \varphi(t_1)$，转第 3 步；

第 5 步　若 $b - t_1 \le \varepsilon$，停止迭代，输出 t_2。否则令 $a := t_1$，$t_1 := t_2$，$t_2 := a + 0.618(b-a)$，$\varphi_1 := \varphi_2$，计算 $\varphi_2 = \varphi(t_2)$，转第 3 步。

➢**0.618法解题步骤：**



确定[a,b],计算探索点
$t_1=a+0.382(b-a)$
$t_2=a+0.618(b-a)$
$\varphi(t_1)\le\varphi(t_2)$
是
否
$t_2-a\le\varepsilon$
$b-t_1\le\varepsilon$
否
以[t_1,b]为新的搜索区间
是
停止，输出t_2
是
否
停止，输出t_1
以[a,t_2]为新的搜索区间

➢例：求解 $\min\limits_{t\geq 0}\varphi(t)=t^3-2t+1$

其中单谷区间[0,3],精度0.5

解：

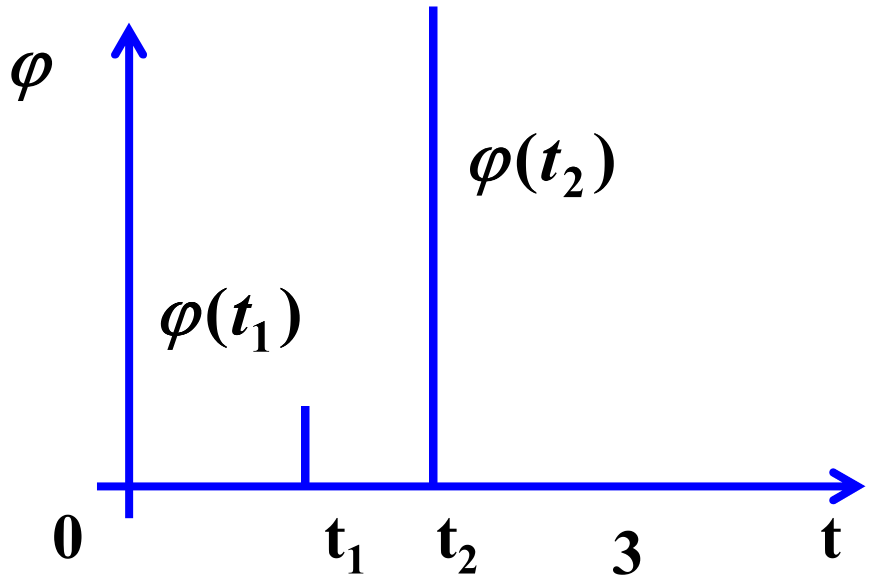
1、第一轮：

$t_1=1.146,\ t_2=1.854$

$\varphi(t_1)=0.2131,$

$\varphi(t_2)=3.6648$

$t_2-0>0.5$



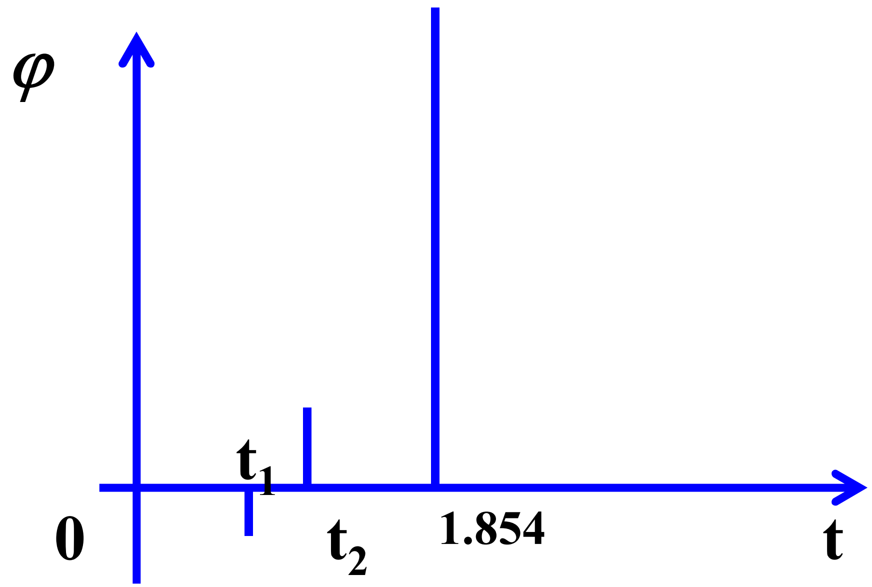
2、第二轮：

$t_2=1.146$, $t_1=0.708$

$\varphi(t_1)=-0.0611$

$\varphi(t_2)=0.2131$

$t_2-0=1.146>0.5$



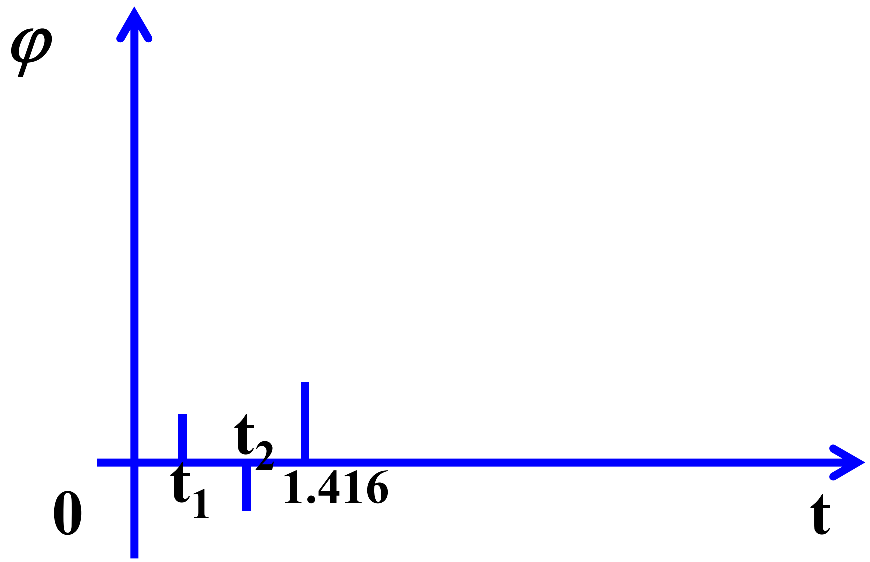
3、第三轮：

$t_1=0.438$, $t_2=0.708$

$\varphi(t_2)=-0.0611$

$\varphi(t_1)=0.2082$

$b-t_1=1.146-0.438>0.5$



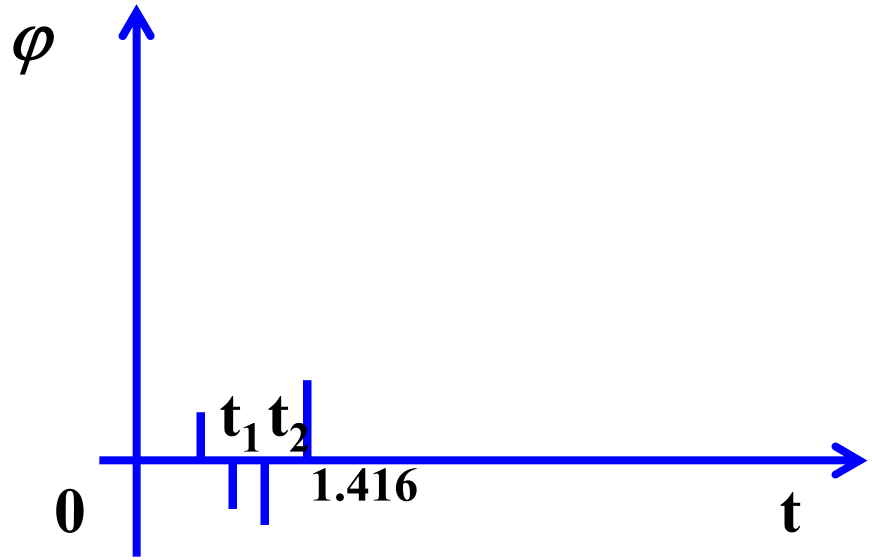
4、第四轮：

t_2=0.876, t_1=0.708

$\varphi(t_1) = -0.0611$

$\varphi(t_2) = -0.0798$

b-t_1=1.146-0.708<0.5



输出：$t^*=t_2$=0.876为最优解，最优值为-0.0798

课下练习：仔细分析上述迭代过程，体会0.618法的实质。

例. 用 0.618 法计算 $f(t)=(t-1)^2$ 在 $[0,3]$ 上的极小点。

解. 计算点 $t_1=0+0.382\times(3-0)=1.146$

$t_2=0+0.618\times(3-0)=1.854$

$f(t_1)=0.146^2=0.0213$， $f(t_2)=0.854^2=0.7293$

$f(t_1)<f(t_2)$， 新的搜索区间 $[0,1.854]$

新的计算点 $t_2=t_1=1.146$， $f(t_2)=0.0213$

$t_1=0+0.382\times(1.854-0)=0.708$

$f(t_1)=(0.708-1)^2=0.0851>f(t_2)$

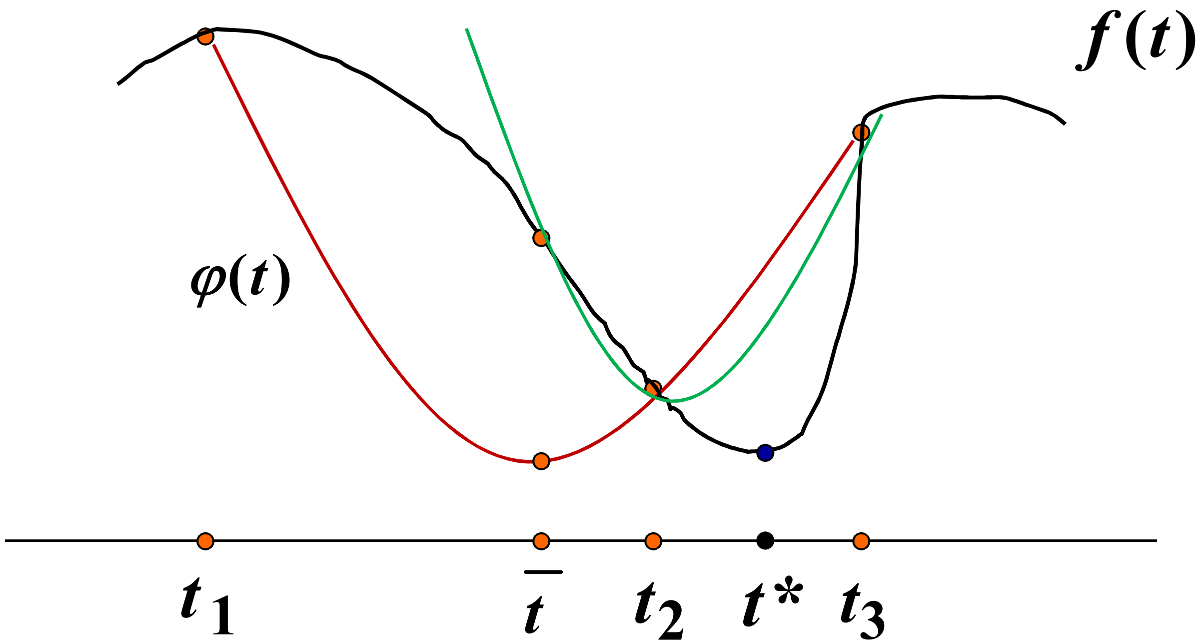
新的搜索区间 $[0.708,1.854]$

新的计算点 $t_1=t_2=1.146$， $f(t_1)=0.0213$

$t_2=0.708+0.618\times(1.854-0.708)=1.416$

二次逼近法（抛物线插值）

思想 在极小点附近，用二次三项式 $\varphi(t)$ 逼近目标函数 $f(t)$，令 $\varphi(t)$ 与 $f(t)$ 在三点 $t_1 < t_2 < t_3$ 处有相同的函数值，并假设 $f(t_1) > f(t_2)$，$f(t_2) < f(t_3)$.



如何计算函数 $\varphi(t)$？

设 $\varphi(t) = a + bt + ct^2$， 则

$$\varphi(t_1) = a + bt_1 + ct_1{}^2 = f(t_1)$$

$$\varphi(t_2) = a + bt_2 + ct_2{}^2 = f(t_2)$$

$$\varphi(t_3) = a + bt_3 + ct_3{}^2 = f(t_3)$$

解上述方程组，可得逼 近函数 $\varphi(t)$ 的系数 b 和 c.

再求函数 $\varphi(t)$ 的极小点，令

$$\varphi'(t) = b + 2ct = 0,$$

解得 $\bar{t} = -\dfrac{b}{2c}$.

以 $\bar{t}$ 作为 $f(t)$ 的极小点的估计值。

抛物线插值算法步骤：

(1) 给定初始区间 $[t_1, t_3]$，设 $f(t_1) > f(t_2), f(t_2) < f(t_3)$.

令 $k := 1$, $\bar{t}^{(0)} = t_2$，给定精度 ε .

(2) 设 $\varphi(t) = a + bt + ct^2$. 令 $\varphi(t_i) = f(t_i)$, $i = 1, 2, 3$, 解出

系数 a, b, c 及 $\varphi(t)$ 的极小点 $\bar{t}^{(k)} = -\dfrac{b}{2c}$.

若 $|f(\bar{t}^{(k)}) - f(\bar{t}^{(k-1)})| < \varepsilon$，则算法停止，否则转（3).

(3) 从 t_1, t_2, t_3 和 $\bar{t}^{(k)}$ 中选择 $f(t)$ 的函数值最小的点 及

其左、右两点，重新标 记为 t_1, t_2, t_3. 令 $k := k + 1$,

转（2).

4、Newton法

考虑$\min\varphi(t)$,其中$\varphi(t)$二次可微， $\varphi''(t)\neq 0$

- **Newton法基本思想：**

  用探索点t_k处的二阶Taylor展开式近似代替目标函数，以展开式的最小点为新的探索点。

展开式：$g(t)=\varphi(t_k)+\varphi'(t_k)(t-t_k)+\frac{1}{2}\varphi''(t_k)(t-t_k)^2$

$g(t)$的最小点即导数为0的点,求导得：

$$t_{k+1}=t_k-\frac{\varphi'(t_k)}{\varphi''(t_k)}$$

Newton法

$$\min \quad \varphi(t)$$

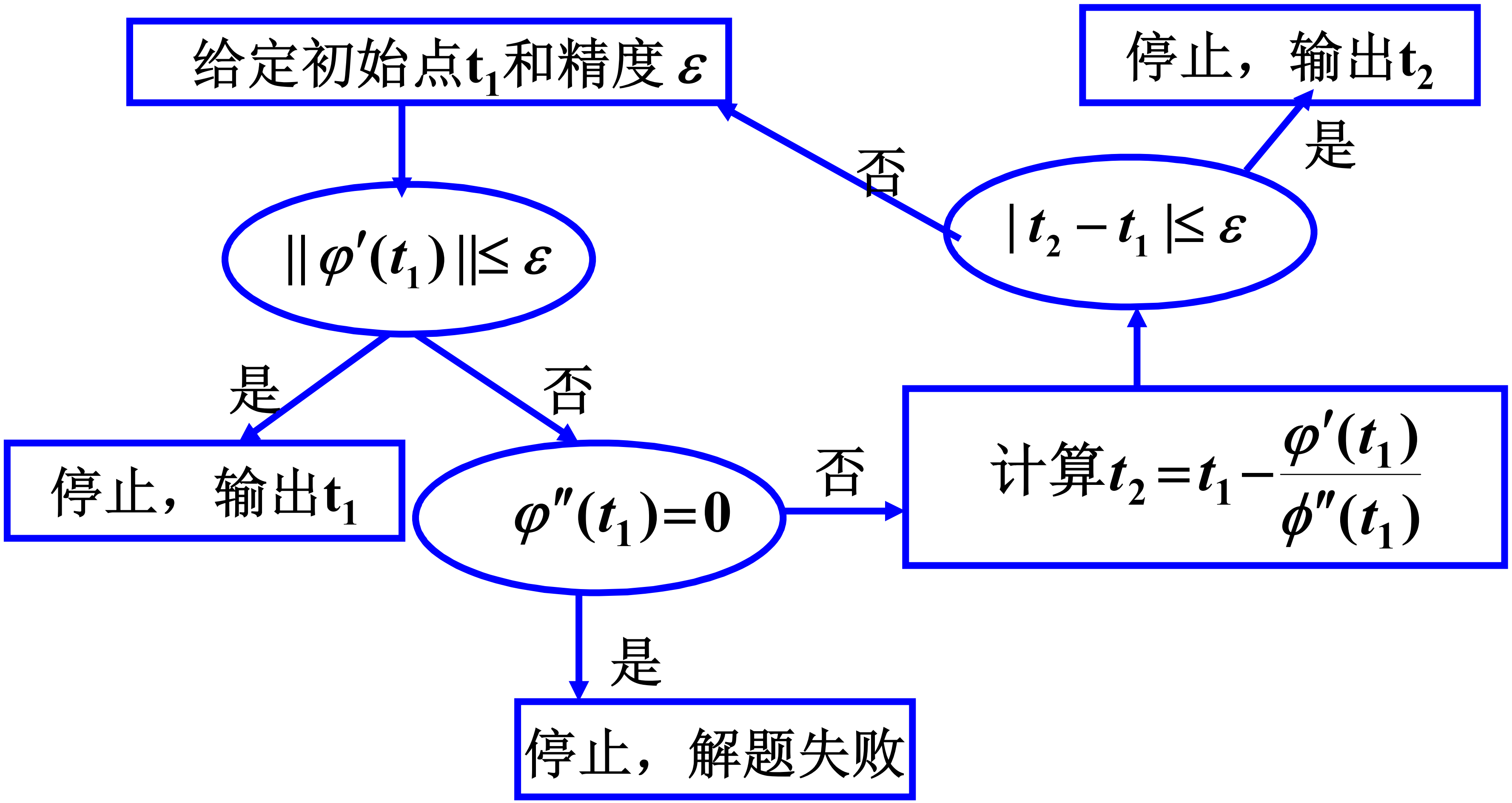
其中 $\varphi(t)$ 是二次可微的，且 $\varphi''(t) \neq 0$。

第 1 步　给定初始点 t_1，$\varepsilon > 0$，$k := 1$；

第 2 步　如果 $|\varphi'(t_k)| < \varepsilon$，停止迭代，输出 t_k。否则，当 $\varphi''(t_k) = 0$ 时，停止，解题失败；当 $\varphi''(t_k) \neq 0$ 时，转下一步；

第 3 步　计算 $t_{k+1} = t_k - \dfrac{\varphi'(t_k)}{\varphi''(t_k)}$，如果 $|t_{k+1} - t_k| < \varepsilon$，停止迭代，输出 t_{k+1}。否则 $k := k+1$，转第 2 步。

➢解题步骤：


给定初始点t_1和精度ε
$\|\varphi'(t_1)\| \le \varepsilon$
是
停止，输出t_1
否
$\varphi''(t_1)=0$
是
停止，解题失败
否
计算$t_2 = t_1 - \dfrac{\varphi'(t_1)}{\phi''(t_1)}$
$|t_2 - t_1| \le \varepsilon$
是
停止，输出t_2
否

➢例：　求解　$\min\varphi(t)=\int_0^t \arctan x dx$

解：　取$t_1=1$,计算：

$$\varphi'(t)=\arctan t,\varphi''(t)=\frac{1}{1+t^2}$$

迭代过程如下表：

k	t_k	$\varphi'(t_k)$	$\varphi''(t_k)$
1	1	0.7854	2
2	-0.5708	-0.5178	1.3258
3	0.1169	0.1163	1.137
4	-0.001061		

$$t_{k+1}=t_k-\frac{\varphi'(t_k)}{\varphi''(t_k)}$$

第四节　无约束最优化方法

➢本节课讨论n元函数的无约束非线性规划问题：

$$\min f(x),\text{其中}x=(x_1,x_2\ldots x_n)^T\in R^n$$

➢求解此类模型(UMP)的方法称为无约束最优化方法。

➢无约束最优化方法通常有两类：

解析法：要使用导数的方法；

直接法：无须考虑函数是否可导，直接使用函数值。

无约束问题的最优性条件

最速下降法（一种解析法）

1、无约束问题的最优性条件

定理1 设$f: R^n \to R$在点$\bar{x} \in R^n$处可微，若存在$p \in R^n$,使

$$\nabla f(\bar{x})^T p < 0$$

则向量 p是f在点 $\bar{x}$处的下降方向

定理2 设f在点$x^* \in R^n$可微，若x^*是$\min f(x)$的局部最优解

$$\text{则} \nabla f(x^*) = 0$$

注：梯度为0的点称为函数的驻点。

驻点可能是极小点，也可能是极大点，也可能即不是极大也不是极小，这时称为函数的鞍点。

定理2说明：UMP问题的局部最优解必是目标函数的驻点。

定理3 设f在点$x^* \in R^n$处的$Hesse$矩阵$\nabla^2 f(x^*)$存在，

若$\nabla f(x^*)=0$,并且$\nabla^2 f(x^*)$正定

则x^*是$\min f(x)$的严格局部最优解

定理4 设$f:R^n \to R^1, x^* \in R^n, f$是$R^n$上的可微凸函数

若$\nabla f(x^*)=0$，则

则x^*是$\min f(x)$的整体最优解

例　　求无约束非线性规划问题

$$\min f(x_1,x_2,x_3)=x_1^2+4x_2^2+x_3^2-2x_1$$

解：

1、先求出目标函数的全部驻点；

2、利用充分条件判断驻点是不是最优点。

2. 最速下降法（梯度法）

问题： $\min\limits_{x\in R^n} f(x)$，其中 $f(x)$ 一阶连续可微 .

思想：取目标函数值在当前点 x^k 处下降最快的方向（即负梯度方向） $-\nabla f(x^k)$ 为搜索方向；

迭代公式： $x^{k+1}=x^k-t_k\nabla f(x^k)$，其中步长 t_k 由精确一维搜索 $\min\limits_{t\ge 0} f(x^k-t\nabla f(x^k))$ 确定。

- 关于梯度的复习：

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。

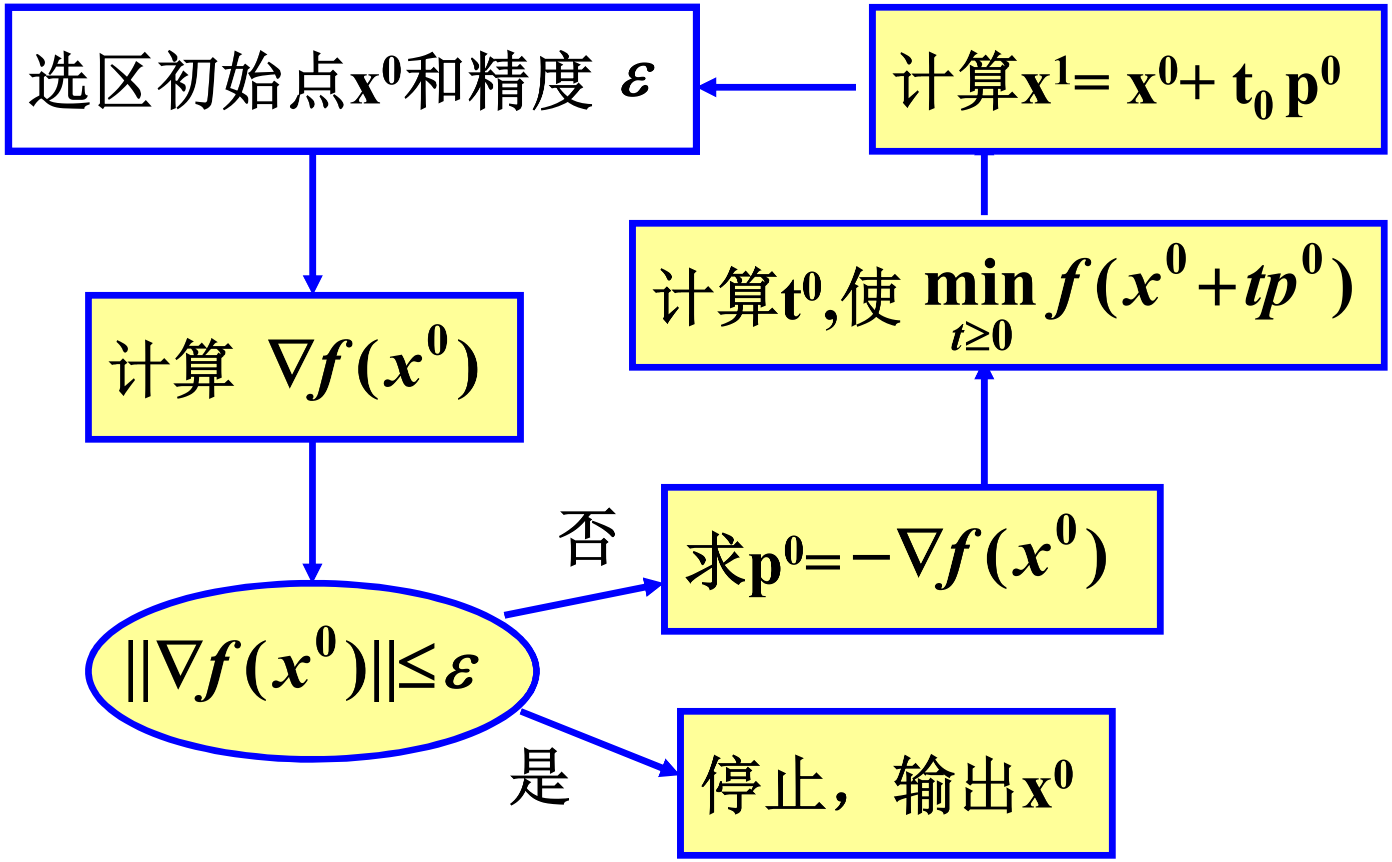
梯度的方向就是函数f的值增加最快的方向，其相反方向就是函数值降低最快的方向。

- 最速下降法又称为梯度法，由**Cauchy**于**1847**年给出。

- 最速下降法解决的是具有连续可微的目标函数的**UMP**问题。

- 最速下降法的基本思想：从当前点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。

➢最速下降法计算步骤：


选区初始点x^0和精度 ε
计算$x^1= x^0+ t_0 p^0$
计算 $\nabla f(x^0)$
计算t^0,使 $\min\limits_{t\geq 0} f(x^0+tp^0)$
否
求$p^0=-\nabla f(x^0)$
$\|\nabla f(x^0)\|\leq\varepsilon$
是
停止，输出x^0

例. $\min f(x)=(x_1)^2+2(x_2)^2$，初始点 $x^0=(1,1)^T$.

解. $\nabla f(x)=\begin{pmatrix}2x_1\\4x_2\end{pmatrix}$　　$\nabla f(x^0)=\begin{pmatrix}2\\4\end{pmatrix}$

$$x^1=x^0-t\nabla f(x^0)=\begin{pmatrix}1-2t\\1-4t\end{pmatrix}$$

$$f(x^1)=(1-2t)^2+2(1-4t)^2=36t^2-20t+3$$

$t=\dfrac{5}{18}$时，$f(x^1)$ 最小，$x^1=\begin{pmatrix}\frac{4}{9}\\-\frac{1}{9}\end{pmatrix}$

$\nabla f(x^1)=\begin{pmatrix}\frac{8}{9}\\-\frac{4}{9}\end{pmatrix}$　　$\cdots$

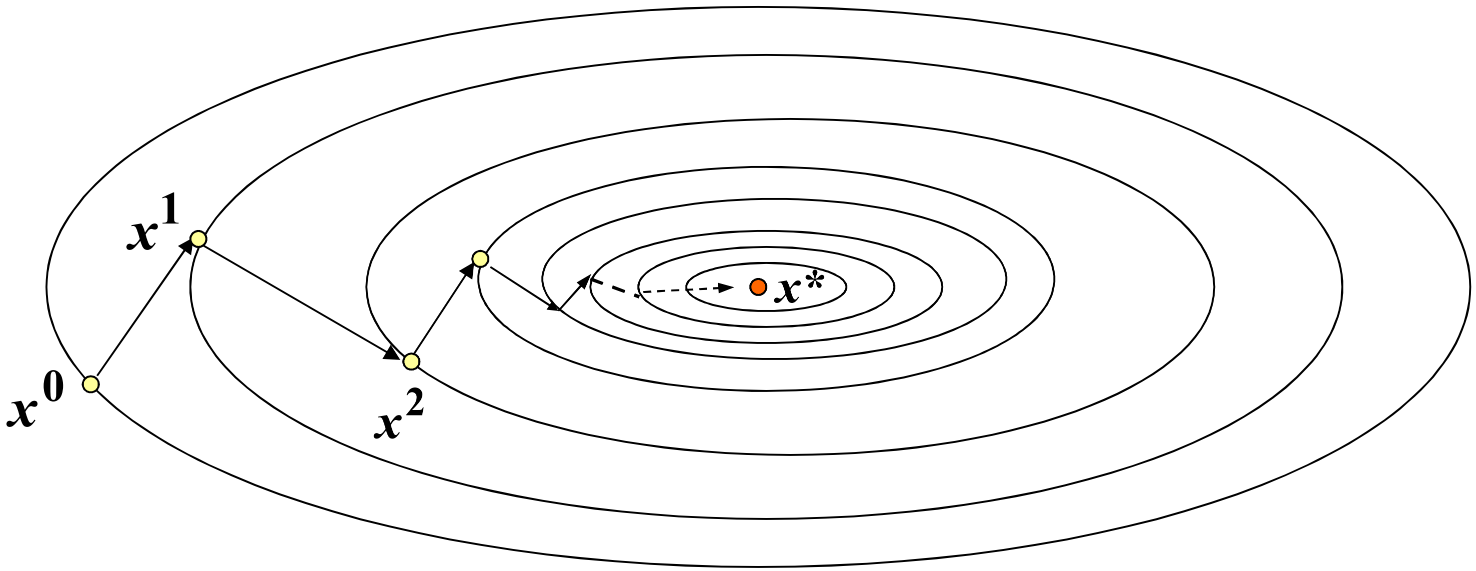
性质： 在最速下降算法中，若 步长由精确一维搜索确定，则 $\nabla f(x^{k+1})^T \nabla f(x^k) = 0$。

证明：设 $\varphi(t) = f(x^k - t\nabla f(x^k))$. 由于搜索步长 t_k 使得 $\varphi(t)$ 达到最小，所以

$$\varphi'(t_k) = -\nabla f(x^k - t_k \nabla f(x^k))^T \nabla f(x^k) = 0$$

即 $\nabla f(x^{k+1})^T \nabla f(x^k) = 0$。

注：在极小点附近，目标函数可用二次函数近似，因此等值面近似为椭球面，上述性质说明最速下降法的搜索路径呈直角锯齿状：



最速下降方向只是目标函数的局部下降最快方向，最速下降算法收敛速度慢，但具有整体收敛性。

近似最佳步长：

若 $f(x)$ 二阶连续可微，则

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

右端为二次函数，

$$t = \frac{\nabla f(x^k)^T \nabla f(x^k)}{\nabla f(x^k)^T \nabla^2 f(x^k) \nabla f(x^k)}$$ 时达到最小。

梯度法$\left(\text{最速下降法}\right)$步骤：

Step1：选取初始点x^0，给定终止误差$\varepsilon > 0$,令$k = 0$;

Step2：计算梯度向量$\nabla f\left(x^k\right)$.

若$\left\|\nabla f\left(x^k\right)\right\| < \varepsilon$,终止迭代计算，输出最 优解$x^* = x^k$;

否则，转***Step3***。

Step3：构造最速下降方向。令$d^k = -\nabla f\left(x^k\right)$

Step4：进行一维搜索。求t_k，使得

$$f\left(x^k + t_k d^k\right) = \min_{t>0} \quad f\left(x^k + t d^k\right)$$

令 $x^{k+1} = x^k + t_k d^k, k = k + 1$,转***Step2***。

3. 共轭方向法

共轭方向：设 A 是 n 阶对称正定阵，对于 R^n 中的两个非零向量 p^1 和 p^2，若 $(p^1)^T Ap^2 = 0$，则称 p^1 和 p^2 关于 A 共轭。

若非零向量组 $p^1, p^2, \cdots, p^k$ 中的向量两两关于 A 共轭，则称该向量组关于 A 共轭。

注：共轭是正交的推广。

性质：共轭向量组一定线性无关。

正定二次函数的共轭方向法：

$$\min \ f(x) = \frac{1}{2} x^T Ax + b^T x + c,$$ 其中 A 正定。

取 n 个 A 共轭方向 $p^0, p^1, \cdots, p^{n-1}$，依次以它们为搜索方向进行迭代，即

$$x^{k+1} = x^k + t_k p^k, \qquad k = 0, 1, \cdots, n-1$$

$$t_k: \quad \min_{t \geq 0} f(x^k + t p^k)$$

性质：共轭方向法用于正定二次函数，则至多经过 n 次迭代即得极小点。

问题：如何选择一组共轭方向？

共轭梯度法：

以已知迭代点处的梯度方向为基础产生共轭方向。

在初始点 x^0，取 $p^0 = -\nabla f(x^0)$

在 x^1，设 $p^1 = -\nabla f(x^1) + \alpha_0 p^0$

p^1 与 p^0 关于 A 共轭，即 $(-\nabla f(x^1) + \alpha_0 p^0)^T A p^0 = 0$

$$\therefore \quad \alpha_0 = \frac{\nabla f(x^1)^T A p^0}{(p^0)^T A p^0}$$

一般地，若已得 A 共轭方向 $p^0, p^1, \cdots, p^k$，可设

$$p^{k+1} = -\nabla f(x^{k+1}) + \sum_{i=0}^{k} \alpha_i p^i$$

可证：欲使 $(p^{k+1})^T A p^j = 0$ $(j = 0, 1, \cdots, k)$，必有

$$\alpha_0 = \alpha_1 = \cdots = \alpha_{k-1} = 0.$$

因此

$$p^{k+1} = -\nabla f(x^{k+1}) + \alpha_k p^k, \quad \alpha_k = \frac{\nabla f(x^{k+1})^T A p^k}{(p^k)^T A p^k}$$

共轭梯度法的搜索方向：

$$p^0 = -\nabla f(x^0)$$

$$p^{k+1} = -\nabla f(x^{k+1}) + \alpha_k p^k$$

$$\alpha_k = \frac{\nabla f(x^{k+1})^T A p^k}{(p^k)^T A p^k}, \quad k = 0, 1, \cdots, n-1$$

共轭梯度法用于正定二次函数时具有性质：

(1) 若 $i \neq j$，则 $\nabla f(x^i)^T \nabla f(x^j) = 0$；

(2) $t_k = -\dfrac{\nabla f(x^k)^T p^k}{(p^k)^T A p^k}$.

为了将共轭梯度法用于非二次函数，需消去公式中的 A。

$\because \ \nabla f(x) = Ax + b$,

$\therefore \ \nabla f(x^{k+1}) - \nabla f(x^k) = A(x^{k+1} - x^k) = t_k A p^k$

在 α_k 的表达式的分子、分母 同乘以 t_k ， 则

$$\alpha_k = \frac{\nabla f(x^{k+1})^T (t_k A p^k)}{(p^k)^T (t_k A p^k)} = \frac{\nabla f(x^{k+1})^T (\nabla f(x^{k+1}) - \nabla f(x^k))}{(p^k)^T (\nabla f(x^{k+1}) - \nabla f(x^k))}$$

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

$$= \frac{\nabla f(x^{k+1})^T \nabla f(x^{k+1})}{-(-\nabla f(x^k) + \alpha_{k-1} p^{k-1})^T \nabla f(x^k)}$$

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

5. 牛顿法

问题：$\min\limits_{x\in R^n} f(x)$，其中 $f(x)$ 二阶连续可微．

在 x^k 附近，$f(x)$ 有近似式：

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

若 $(\nabla^2 f(x^k))^{-1}$ 存在，取 $g(x)$ 的极小点为 x^{k+1}，则

$$\nabla g(x^{k+1})=\nabla f(x^k)+\nabla^2 f(x^k)(x^{k+1}-x^k)=0$$

$$\therefore\quad x^{k+1}=x^k-(\nabla^2 f(x^k))^{-1}\nabla f(x^k)$$ 牛顿法的迭代公式

$$p^k=-(\nabla^2 f(x^k))^{-1}\nabla f(x^k)$$ 牛顿方向

几点说明：

(1) 若 $f(x)$ 为正定二次函数，则牛顿法只需一步迭代即得极小点，因为此时有 $f(x)=g(x)$，所以

$$\nabla f(x^1)=\nabla g(x^1)=0.$$

(2) 牛顿法具有二阶收敛速率，但要求初始点选在极小点附近。

(3) $(\nabla^2 f(x^k))^{-1}$ 的存在性不易保证，即 使存在，也不易计算。

(4) 由于 $f(x)\approx g(x)$ 仅在 x^k 附近成立，因此 $f(x^{k+1})\le f(x^k)$ 不一定成立，改进的一 个途径是沿 $-(\nabla^2 f(x^k))^{-1}\nabla f(x^k)$ 线性搜索确定 x^{k+1}.