南京大學 NANJING UNIVERSITY 工程管理学院 SCHOOL OF MANAGEMENT & ENGINEERING

第五章 非线性规划

- 概述
- 非线性规划问题的解
- 凸函数和凸规划
- 下降迭代算法
- 无约束极值问题
- 约束极值问题

第五章 非线性规划

1. 概述

例5.1 曲线拟合问题

在某化学反应里，已知生成物的浓度φ与时间t之间有如下的经验函数关系：

$$\varphi = c_1 + c_2 t + e^{c_3 t} \qquad (1)$$

其中c_1, c_2, c_3是待定参数。现通过测试获得到了n组浓度φ与时间t之间的实验数据$(t_i, \varphi_i), i = 1,2,\cdots,n$。

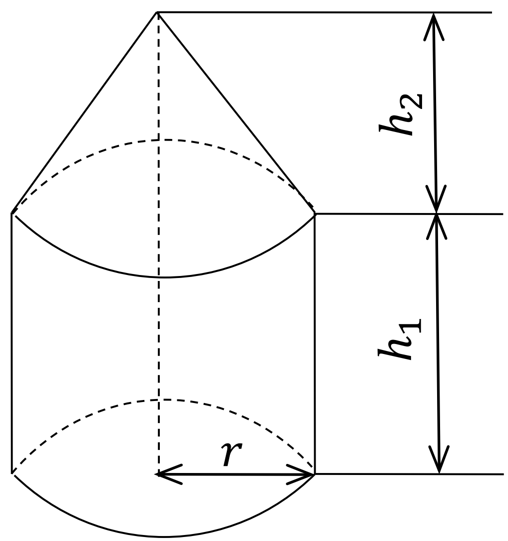
试确定参数c_1, c_2, c_3，使理论曲线（1）尽可能地与n个测试点(t_i, φ_i)吻合。

最小二乘
Least Squares

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

例5.3 构件设计问题

设计一个如图所示的由圆锥和圆柱所组成的构件，要求构件体积为V，圆锥的高h_1和圆柱的高h_2之比为a，确定构件尺寸使其表面积最小。

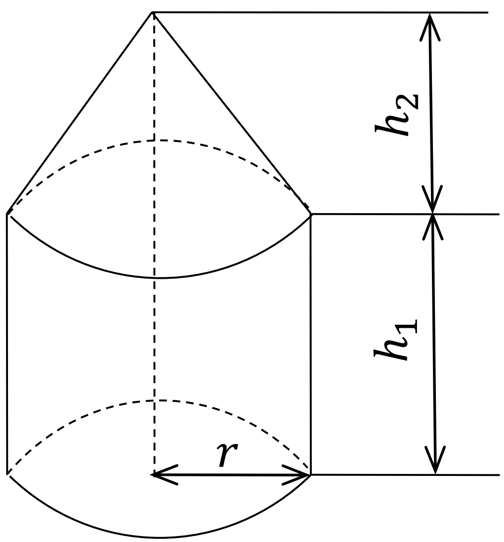


例5.3 构件设计问题

设计一个如图所示的由圆锥和圆柱所组成的构件，要求构件体积为V，圆锥的高h_1和圆柱的高h_2之比为a，确定构件尺寸使其表面积最小。

$$\min \quad 2\pi r h_2 + \pi r^2 + \pi r\sqrt{r^2 + h_1^2}$$

$$\text{s.t.} \quad \begin{cases} \pi r^2 h_2 + \dfrac{1}{3}\pi r^2 h_1 = V \\ \dfrac{h_1}{h_2} = a \\ r, h_1, h_2 \geq 0 \end{cases}$$



非线性规划数学模型的一般形式

$$\min \ f(\mathbf{x})$$

$$\text{s.t.} \ \begin{cases} g_i(\mathbf{x}) \geq 0, i = 1,2,\cdots l \\ h_j(\mathbf{x}) = 0, j = 1,2,\cdots m \end{cases}$$

若令$\Omega = \{\mathbf{x}|g_i(\mathbf{x}) \geq 0, h_j(\mathbf{x}) = 0, i = 1,2,\cdots,l; j = 1,2,\cdots,m\}$，则上述模型可写为：

$$\min_{\mathbf{x}\in\Omega} f(\mathbf{x})$$

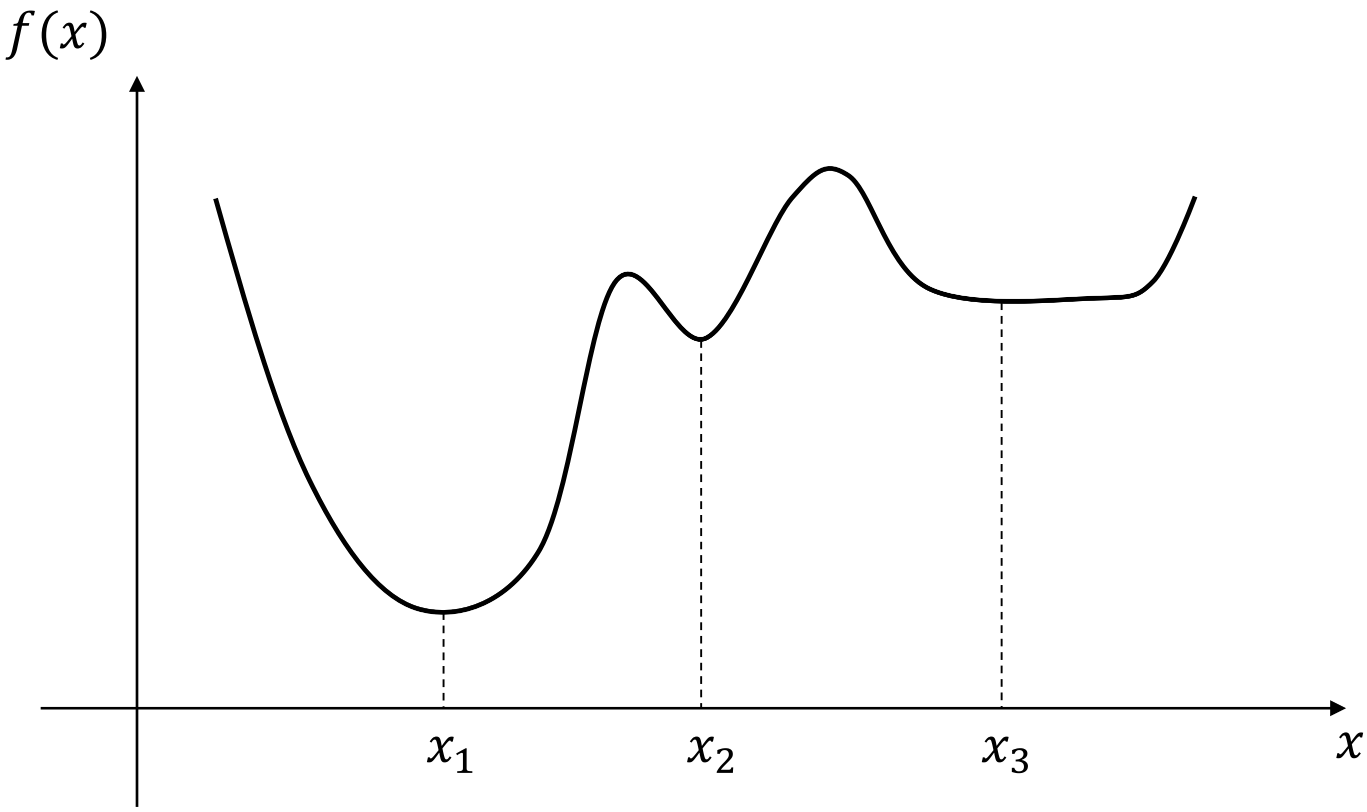
若$\Omega = \Re^n$, 则为无约束优化问题（Unconstrained Optimization Problem）；否则，若$\Omega \subset \Re^n$，则为约束优化问题（Constrained Optimization Problem）；

南京大學 NANJING UNIVERSITY 工程管理学院 SCHOOL OF MANAGEMENT & ENGINEERING

第五章 非线性规划

2. 非线性规划问题的解

2.1 解（极值点）的定义



2.1 解（极值点）的定义

- **局部极小点：**

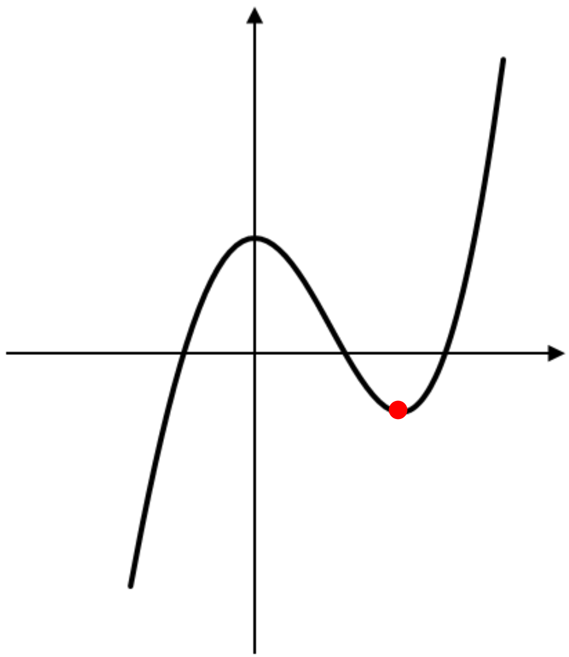
若存在某个$\varepsilon > 0$，使得对所有与$\mathbf{x}^*$的距离小于ε的$\mathbf{x} \in \Omega$，即在$\mathbf{x}^*$的某个邻域$N_\varepsilon(\mathbf{x}^*) = \{\mathbf{x} \in \Omega \mid \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon\}$中，$\forall \mathbf{x} \in N_\varepsilon(\mathbf{x}^*)$都有$f(\mathbf{x}) \geq f(\mathbf{x}^*)$，则称点$\mathbf{x}^*$为非线性规划问题的局部极小点（Local Minimum Point 或 Relative Minimum Point）。

若$\forall \mathbf{x} \in N_\varepsilon(\mathbf{x}^*)$且$\mathbf{x} \neq \mathbf{x}^*$，都有$f(\mathbf{x}) > f(\mathbf{x}^*)$，则称点$\mathbf{x}^*$为严格局部极小点（Strict Local Minimum Point）。

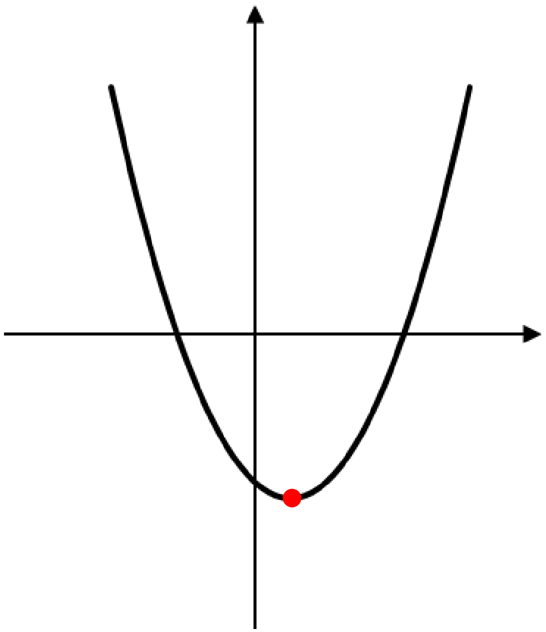
- **全局极小点：**

若存在$\mathbf{x}^* \in \Omega$，$\forall \mathbf{x} \in \Omega$，都有$f(\mathbf{x}) \geq f(\mathbf{x}^*)$，则称点$\mathbf{x}^*$为非线性规划问题的全局极小点（Global Minimum Point）。若$\forall \mathbf{x} \in \Omega$且$\mathbf{x} \neq \mathbf{x}^*$，都有$f(\mathbf{x}) > f(\mathbf{x}^*)$，则称点$\mathbf{x}^*$为严格全局极小点（Strict Global Minimum Point）。

2.1 解（极值点）的定义



(a) 局部极小点



(b) 全局极小点

➢ **定理5.1**

设f是定义在集合$\Omega \subset \Re^n$上的一阶连续可微函数，如果$\mathbf{x}^*$是f在Ω上的局部极小点，那么对$\mathbf{x}^*$的任意可行方向$\mathbf{p} \in \Re^n$，有

$$\nabla f(\mathbf{x}^*)^T \mathbf{p} \geq 0$$

其中

$$\nabla f(\mathbf{x}^*) = \left(\frac{\partial f(\mathbf{x}^*)}{\partial x_1}, \ \frac{\partial f(\mathbf{x}^*)}{\partial x_2}, \ \cdots, \ \frac{\partial f(\mathbf{x}^*)}{\partial x_n}\right)^T$$

是函数f在点$\mathbf{x}^*$处的梯度(Gradient)，即一阶偏导数。

梯度方向是函数f在点$\mathbf{x}^*$的等值线的法线方向，沿这个方向函数值增加最快。

证明：考虑如下的一阶Taylor展式：

$$f(\mathbf{x}^* + \alpha\mathbf{p}) \approx f(\mathbf{x}^*) + \alpha\nabla f(\mathbf{x}^*)^T\mathbf{p}$$

如果$\nabla f(\mathbf{x}^*)^T\mathbf{p} < \mathbf{0}$，那么即便是对充分小的$\alpha > 0$, 都有

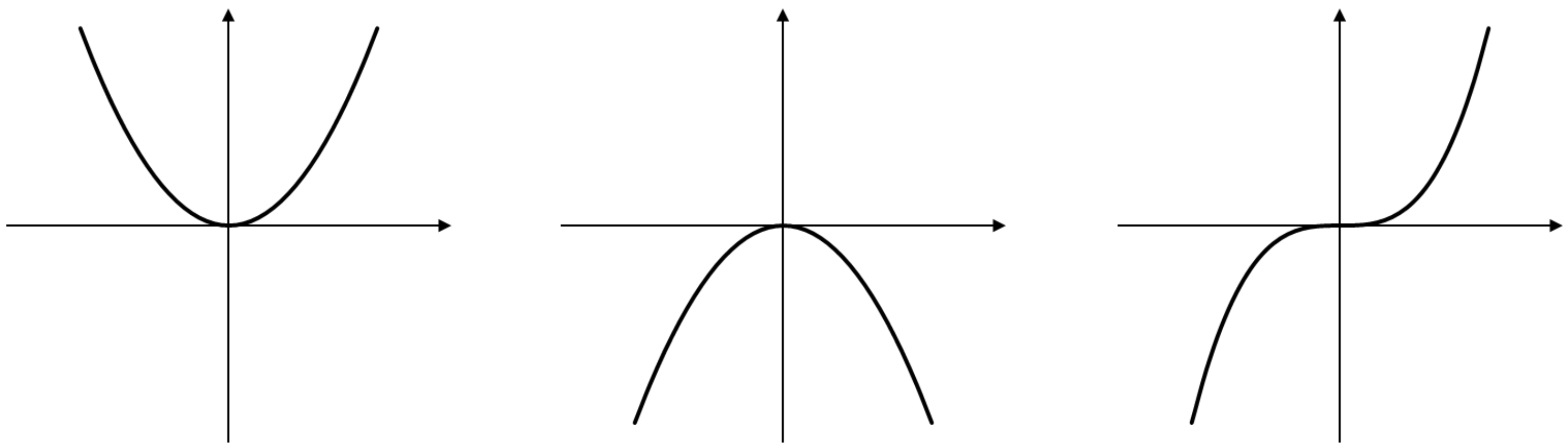
$$\alpha\nabla f(\mathbf{x}^*)^T\mathbf{p} < \mathbf{0}$$

即$f(\mathbf{x}^* + \alpha\mathbf{p}) < f(\mathbf{x}^*)$，这与$\mathbf{x}^*$是局部极小点相矛盾。

若$\Omega = \Re^n$，则所有方向$\mathbf{p}$都是可行方向，那么对任意的$\mathbf{p}$，有$\nabla f(\mathbf{x}^*)^T\mathbf{p} \geq \mathbf{0}$。于是，$\nabla f(\mathbf{x}^*) = 0$。

- **定理5.2（一阶必要条件——无约束情形）**

设f是定义在$\Re^n$上的一阶连续可微函数，如果$\mathbf{x}^*$是f在$\Re^n$上的局部极小点，则必有$\nabla f(\mathbf{x}^*) = 0$。



（a）局部极小点　　（b）局部极大点　　（c）鞍点

$\nabla f(\mathbf{x}^*) = 0 \Longrightarrow$ 稳定点（Stationary Point）

2.2 多元函数极值点的存在条件

考虑下面的二阶Taylor展式：

$$f(\mathbf{x}^* + \alpha\mathbf{p}) \approx f(\mathbf{x}^*) + \alpha\nabla f(\mathbf{x}^*)^T\mathbf{p} + \frac{1}{2}\alpha^2\mathbf{p}^{\mathrm{T}}\nabla^2 f(\mathbf{x}^*)\mathbf{p}$$

$$\nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n^2} \end{pmatrix}$$

是函数f在点$\mathbf{x}^*$处的二阶偏导数，被称为Hessian矩阵。

- Hessian矩阵是对称的。

- 练习

请计算下面函数的梯度和Hessian矩阵:

$$f(x_1, x_2) = 2x_1^4 + 3x_1^2x_2 + 2x_1x_2^3 + 4x_2^2$$

$$\nabla f(x) = \begin{pmatrix} 8x_1^3 + 6x_1x_2 + 2x_2^3 \\ 3x_1^2 + 6x_1x_2^2 + 8x_2 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 24x_1^2 + 6x_2 & 6x_1 + 6x_2^2 \\ 6x_1 + 6x_2^2 & 12x_1x_2 + 8 \end{pmatrix}$$

- **定理5.3（二阶必要条件——无约束情形）**

设函数f在$\Re^n$上具有二阶连续偏导数，如果$\mathbf{x}^*$是f在$\Re^n$上的局部极小点，则必有：

（1）$\nabla f(\mathbf{x}^*) = 0$；

（2）$\forall \mathbf{p} \in \Re^n$，$\mathbf{p}^T \nabla^2 f(\mathbf{x}^*)\mathbf{p} \ge 0$（即Hessian矩阵$\nabla^2 f(\mathbf{x}^*)$半正定）。

- **定理5.4（二阶充分条件——无约束情形）**

设函数f在$\Re^n$上具有二阶连续偏导数，若$\nabla f(\mathbf{x}^*) = 0$，且$\nabla^2 f(\mathbf{x}^*)$正定，则$\mathbf{x}^*$是$f$在$\Re^n$上的严格局部极小点。

$\Downarrow$

$\forall \mathbf{p} \ne \mathbf{0} \in \Re^n$，$\mathbf{p}^T \nabla^2 f(\mathbf{x}^*)\mathbf{p} > 0$

- **例5.4**

求下面函数的局部极小点：

$$f(x_1,x_2)=\frac{1}{3}x_1^3+\frac{1}{2}x_1^2+2x_1x_2+\frac{1}{2}x_2^2-x_2+9$$

解：$\nabla f=\begin{pmatrix}x_1^2+x_1+2x_2\\2x_1+x_2-1\end{pmatrix}=0\ \Longrightarrow\ \mathbf{x}_a=\begin{pmatrix}1\\-1\end{pmatrix}$，$\mathbf{x}_b=\begin{pmatrix}2\\-3\end{pmatrix}$

又 $\nabla^2 f(\mathbf{x}_a)=\begin{pmatrix}3&2\\2&1\end{pmatrix}$，$\nabla^2 f(\mathbf{x}_b)=\begin{pmatrix}5&2\\2&1\end{pmatrix}$

这里只有$\nabla^2 f(\mathbf{x}_b)$正定，所以$\mathbf{x}_b$是严格局部极小点。

第五章 非线性规划

3. 凸函数和凸规划

3.1 凸函数的定义

➢ **凸函数**：设f是定义在凸集Ω上的函数，若对任意实数$0 \le \alpha \le 1$以及任意两点$\mathbf{x}, \mathbf{y} \in \Omega$，有

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \le \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

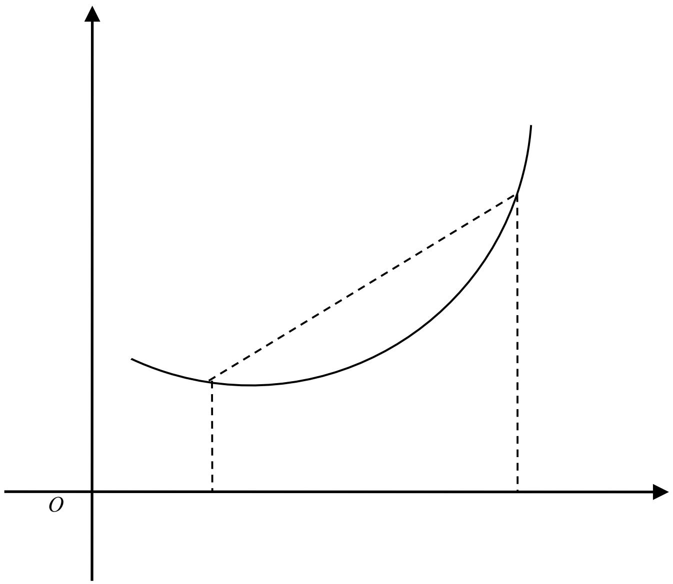
则称f是凸集Ω上的凸函数(Convex Function)。

➢ **严格凸函数**：设f是定义在凸集Ω上的函数，若对任意实数$0 < \alpha < 1$以及任意两点$\mathbf{x}, \mathbf{y} \in \Omega$且$\mathbf{x} \neq \mathbf{y}$，有

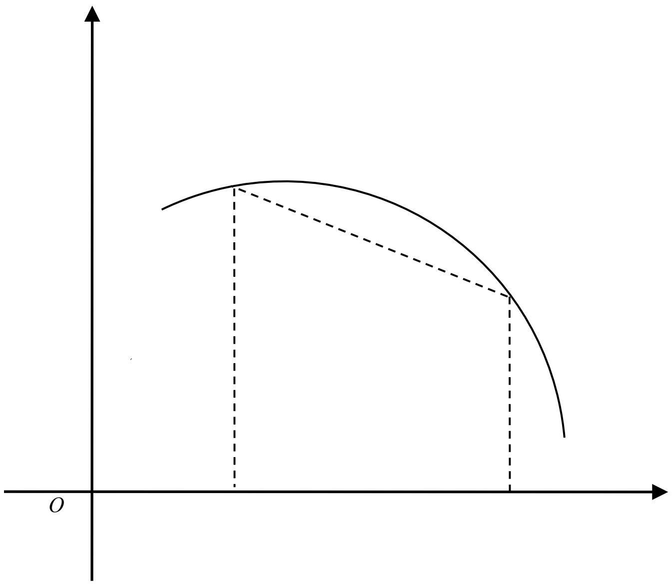
$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

则称f是凸集Ω上的严格凸函数(Strictly Convex Function)。

3.1 凸函数的定义



(a) 凸函数



(b) 凹函数

3.2 凸函数的性质

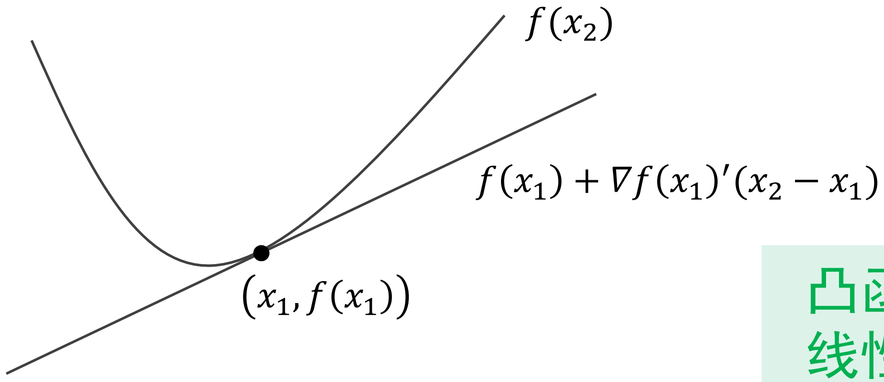
- **性质1** 设$f(\mathbf{x})$是定义在凸集Ω上的凸函数，$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$是$\Omega$中的$m$个点，数$\alpha_1, \alpha_2, \cdots, \alpha_m \geq 0$且$\sum_{i=1}^{m} \alpha_i = 1$，则
$$f(\alpha_1\mathbf{x}_1+\alpha_2\mathbf{x}_2+\cdots+\alpha_m\mathbf{x}_m) \leq \alpha_1 f(\mathbf{x}_1)+\alpha_2 f(\mathbf{x}_2)+\cdots+\alpha_m f(\mathbf{x}_m)$$
- **性质2** 设$f(\mathbf{x})$是定义在凸集Ω上的凸函数，那么对任意实数$\beta \geq 0$，函数$\beta f(\mathbf{x})$也是定义在Ω上的凸函数。
- **性质3** 设f_1, f_2是定义在凸集Ω上的两个凸函数，那么$f_1 + f_2$仍然是定义在Ω上的凸函数。
- **性质4** 设$f(\mathbf{x})$是定义在凸集Ω上的凸函数，那么对任意实数c，集合（称为水平集）$\Gamma_c = \{\mathbf{x} | f(\mathbf{x}) \leq c, \ \mathbf{x} \in \Omega\}$是凸集。

- **性质5.5（一阶充分必要条件）**

设f在开凸集Ω上是可微函数，则f在Ω上是**凸函数**的充分必要条件是：$\forall \mathbf{x}_1, \mathbf{x}_2 \in \Omega$有

$$f(\mathbf{x}_2) \geq f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^T (\mathbf{x}_2 - \mathbf{x}_1)$$

若上式为严格不等式，它就是**严格凸函数**的充要条件。



凸函数中任意点的局部线性近似总低于该函数

- **性质5.6（二阶充分必要条件）**

设f在开凸集Ω上是二阶可微函数，则f在Ω上是**凸函数**的充分必要条件是：$\forall \mathbf{x} \in \Omega$，Hessian矩阵$\mathbf{H}(\mathbf{x}) = \nabla^2 f(\mathbf{x})$是**半正定**的。

若$\forall \mathbf{x} \in \Omega$，$\mathbf{H}(\mathbf{x})$是**正定**的，则$f$在$\Omega$上是**严格凸函数**。

例5.5 证明：$f(\mathbf{x}) = 4x_1^2 + 6x_1x_2 + 9x_2^2$是凸函数。

解：由二阶导数$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 8 & 6 \\ 6 & 18 \end{pmatrix}$处处正定，可得：$f(\mathbf{x})$是严格凸函数。

3.4 凸规划的定义和性质

考虑下面的数学规划：

$$\min_{\mathbf{x}\in\Omega} \ f(\mathbf{x})$$

若其中$f(\mathbf{x})$为凸函数，Ω为凸集，则称该式为**凸规划**（Convex Program）。

例： 如果$f(x)$是凸函数，$\{g_i(x), i=1,\cdots,m\}$是凹函数，那么下面的非线性规划问题是否为凸规划？

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } \ g_i(x) \geq 0, i=1,\cdots,m \end{aligned}$$

3.4 凸规划的定义和性质

- **性质5.7** 凸规划的最优解具有下述特殊性质：

（1）如果最优解存在，那么最优解集为凸集；

（2）任何局部最优解也就是全局最优解；

（3）如果目标函数为严格凸函数且最优解存在，那么最优解唯一。

第五章 非线性规划

4. 下降迭代算法

4.1 下降迭代算法的基本思想

从一个初始估计点$\mathbf{x}_0$出发，按照一定的规则，找到一个比$\mathbf{x}_0$更好的点$\mathbf{x}_1$（即对极小化问题来说，就是要求$f(\mathbf{x}_1)$比$f(\mathbf{x}_0)$更小），再从点$\mathbf{x}_1$出发，找到一个比$\mathbf{x}_1$更好的点$\mathbf{x}_2$，……，如此继续，得到一个解点序列$\{\mathbf{x}_k\}$。

✓ 若该点列$\{\mathbf{x}_k\}$有一极限点$\mathbf{x}^*$，即

$$\lim_{k\to\infty}\|\mathbf{x}_k-\mathbf{x}^*\|=0$$

就称该点列收敛于$\mathbf{x}^*$ 。

✓ 若$\mathbf{x}^*$是问题$\min\limits_{\mathbf{x}\in\Omega} f(\mathbf{x})$的最优解，则称该算法是有效的下降迭代算法。

4.2 下降迭代算法的基本步骤

第1步：**选取某一初始点$\mathbf{x}_0$。**令$k := 0$；

第2步：**确定搜索方向。**若已得出某一迭代点$\mathbf{x}_k$，且$\mathbf{x}_k$不是极小点，那么就要根据一定的规则，从$\mathbf{x}_k$点出发确定一个搜索方向$\mathbf{d}_k$，*沿该方向应能找到使目标函数值下降的可行点。*

第3步：**确定步长。**沿方向$\mathbf{d}_k$前进一个步长，得到新点$\mathbf{x}_{k+1}$，通过选择合适的步长α_k，使得下一个迭代点

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

满足

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k)$$

第4步：**最优性检验。**检验新得到的点是否是极小点或达到近似极小点的要求。如满足，则迭代停止；否则，令$k := k + 1$，返回上面的“第2步”继续迭代。

4.2 下降迭代算法的基本步骤

- 在上述步骤中，搜索方向（也称寻优方向）的确定对算法起着关键性作用。各种算法的区分，事实上主要在于确定搜索方向的方法不同。
- 对有约束优化问题而言，搜索方向不仅要求是目标函数的下降方向，同时还要求是可行方向，即沿此方向搜索的解仍然在可行域内。
- 对于极小化问题，在多数算法中，步长的选定是以使目标函数值沿搜索方向下降最多为依据的。

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) = \min_{\alpha \geq 0} f(\mathbf{x}_k + \alpha \mathbf{d}_k)$$

——“一维搜索(One Dimensional Search)”或“线搜索(Line Search)”。由此确定的步长称为“最佳步长(Optimal Step Size)”。

4.3 迭代终止准则

- 由于真正的极值点$\mathbf{x}^*$事先并不知道，因此在下降迭代算法中只能根据相邻两次迭代得到的计算结果来判断，当前迭代点是否已经达到要求。
- 常用的迭代终止准则有：
  - 按绝对误差：

$$\|\mathbf{x}_{k+1}-\mathbf{x}_k\|<\varepsilon \text{ 和 } |f(\mathbf{x}_{k+1})-f(\mathbf{x}_k)|<\varepsilon$$

  - 按相对误差：

$$\frac{\|\mathbf{x}_{k+1}-\mathbf{x}_k\|}{\|\mathbf{x}_k\|}<\varepsilon \text{ 和 } \frac{|f(\mathbf{x}_{k+1})-f(\mathbf{x}_k)|}{|f(\mathbf{x}_k)|}<\varepsilon$$

  - 按目标函数梯度的模：

$$\|\nabla f(\mathbf{x}_k)\|<\varepsilon$$

第五章 非线性规划

5. 一维搜索

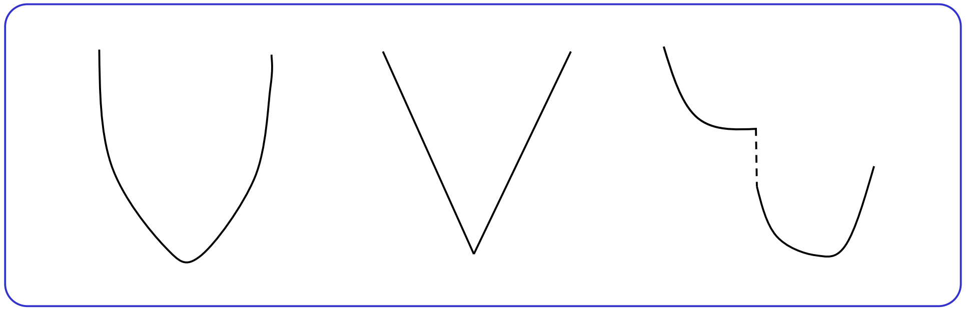
5.1 简介

- 一维搜索就是单变量函数在某个区间上求极值点的问题。
- 一维搜索的方法很多，常用的有：
  - 试探法（如黄金分割法，斐波那契法等）；
  - 微积分中的求根法（如切线法，二分法等）；
  - 函数逼近法，也称插值法（如抛物线插值法，三次插值法等）。

5.2 斐波那契法和黄金分割法

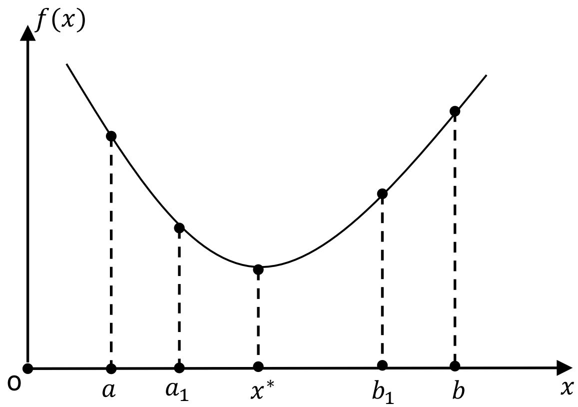
- 斐波那契法和黄金分割法有较广泛的实用性，它们不需要函数具有连续性和可微性，只要求目标函数是单峰函数（Unimodal Function）。
- （下）单峰函数：如果单变量函数$f(x)$在闭区间$[a,b]$上有唯一的极小点x^*，并且函数在x^*的左边严格下降，在x^*的右边严格上升，即$\forall x_1, x_2 \in [a,b]$，$x_1 < x_2$，有

$$\begin{cases} x_2 \le x^* & 则 f(x_1) > f(x_2) \\ x^* \le x_1 & 则 f(x_2) > f(x_1) \end{cases}$$

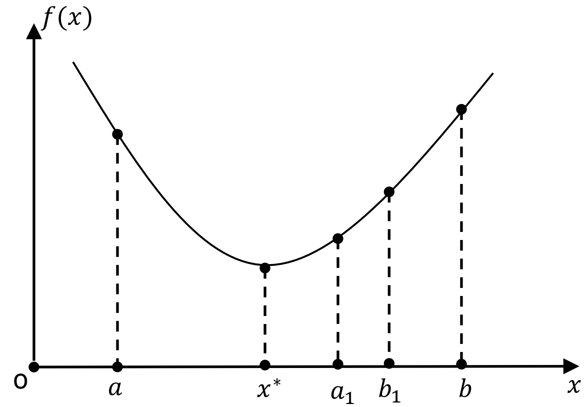


5.2 斐波那契法和黄金分割法

- 对于单峰函数，可以通过计算闭区间$[a,b]$内相异两点的函数值来确定其极小点的位置。若任取a_1,b_1，且$a_1<b_1$，那么有

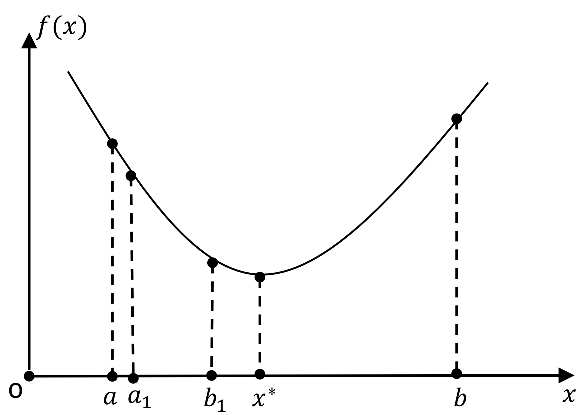


(a)

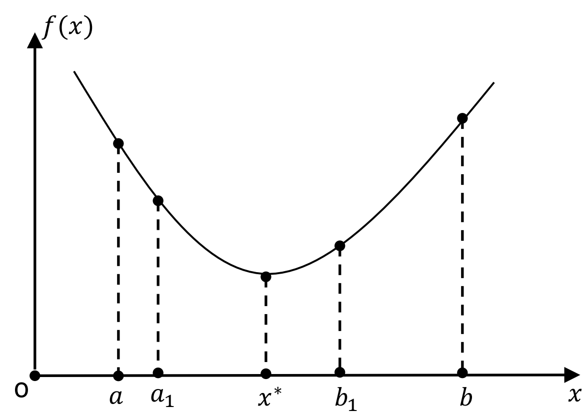


(b)

若 $f(a_1)\le f(b_1)$，那么$x^*\in[a,b_1]$



(a)



(b)

若$f(a_1)\ge f(b_1)$，那么$x^*\in[a_1,b]$

5.2 斐波那契法和黄金分割法

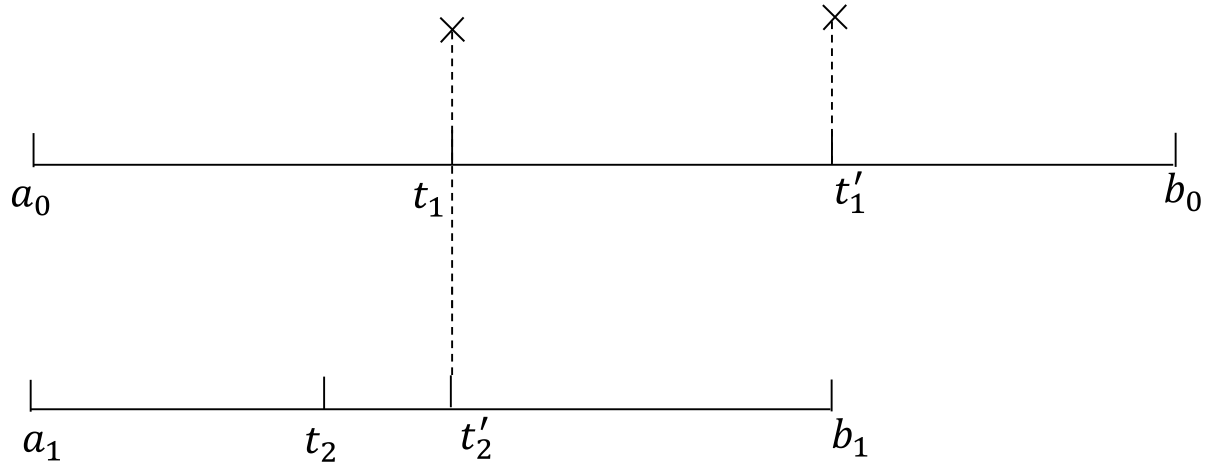
- 通过选择试验点计算函数值，比较函数值的大小，即可缩短包含极小点的区间。
- 区间缩得越短，就越接近函数的极小点，当单峰区间缩短到充分小时，可以将最后的试验点视为极小点的近似值。
- 此类搜索算法也称为序贯试验法（Sequential Experimental Method）。

✓ 基本思想：以尽可能少地选取试验点，获得尽可能大的区间缩短率。

5.2（1）黄金分割法

- 在搜索区间中“对称”取点，等比例得缩小区间
- 假设区间缩短率为α，在初始单峰区间$[a_0, b_0]$中，对称取两点分别为：

$$t_1 = a_0 + (1-\alpha)(b_0 - a_0), \qquad t_1' = a_0 + \alpha(b_0 - a_0)$$



$f(t_1) < f(t_1')$的情形

$$t_2 = a_1 + (1-\alpha)(b_1 - a_1), \qquad t_2' = a_1 + \alpha(b_1 - a_1)$$

5.2（1）黄金分割法

要使$t_1 = t_2'$，则有：

$$a_0 + (1-\alpha)(b_0 - a_0) = a_1 + \alpha(b_1 - a_1)$$

代入$a_1 = a_0, b_1 - a_1 = \alpha(b_0 - a_0)$，得：

$$a_0 + (1-\alpha)(b_0 - a_0) = a_0 + \alpha^2(b_0 - a_0)$$

经过整理，有：

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

求解得：$\alpha = \frac{\sqrt{5}-1}{2} \approx 0.618$（负值舍去）。

5.2（2）斐波那契法

- 如果只计算n次函数值，能将一个给定的包含极小点的区间缩短到多少呢？
- 或者，计算n次函数值，能将一个多大的区间缩短到长度为1的单位区间呢？
- 假设用F_n表示计算n次函数值能将其缩短到1单位长度的最大区间长度。
- 显然，$F_0 = F_1 = 1$
- 那么，F_2？F_3？F_4？……

5.2（2）斐波那契法

斐波那契（Fibonacci）数列：

（1）$F_0 = F_1 = 1$

（2）$F_n = F_{n-1} + F_{n-2}$

$F_2 = 2$

$F_3 = 3$

$F_4 = 5$

$F_5 = 8$

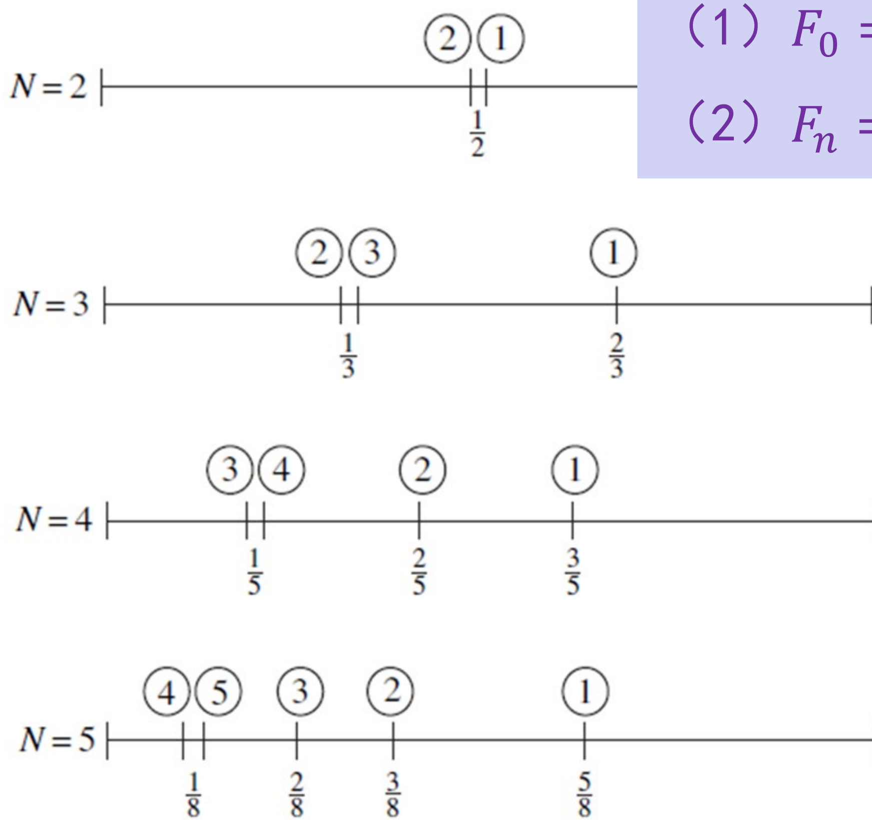


图 5.11 斐波那契法中试验点的选择

5.2（2）斐波那契法

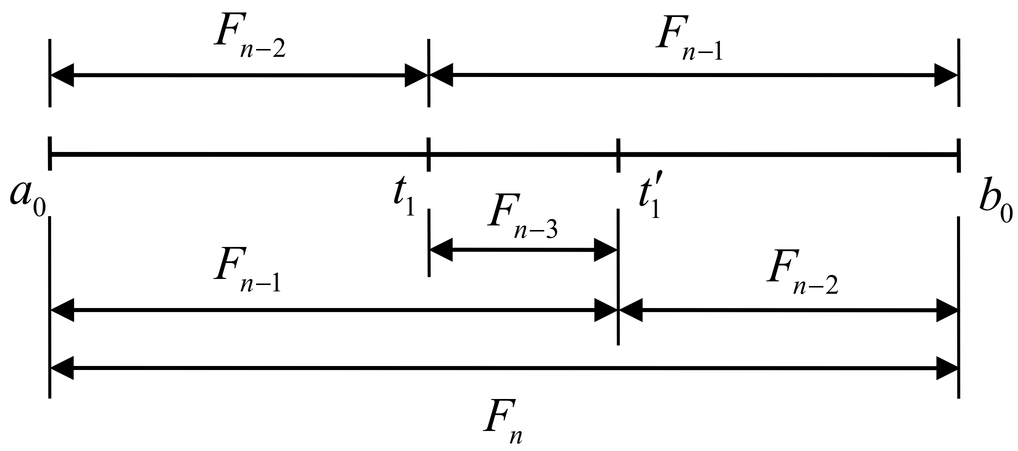


图 5.12 斐波那契法示意图

- 应用斐波那契法时，首先根据精度要求来确定计算函数值的次数n。
- 在其后的迭代计算中，其区间长度的缩短率依次为：

$$0.618 \leftarrow \quad \frac{F_{n-1}}{F_n}, \frac{F_{n-2}}{F_{n-1}}, \cdots, \frac{F_1}{F_2}$$

第五章 非线性规划

6. 无约束极值问题的求解算法

6.1 概述

$$\min_{\mathbf{x}\in\Re^n} f(\mathbf{x})$$

- 对于一个复杂的非线性函数，用求稳定点即求解方程

$$\nabla f(\mathbf{x}) = 0$$

  的方法得到其极值点往往是很困难的，特别是对于多变量问题。

- 在实际计算中通常还是要使用迭代的方法。迭代法一般分为两类：
  - 一类需要用到函数的导数值，称为解析法；
  - 另一类则仅用到各点的函数值，称为直接法。

6.2 梯度法

梯度法（Gradient Method）是一种十分简单，但却又极为重要的解析方法。它使用方便，同时是构建其它更高级优化算法的基础。

基本思想：以负梯度方向作为搜索方向。

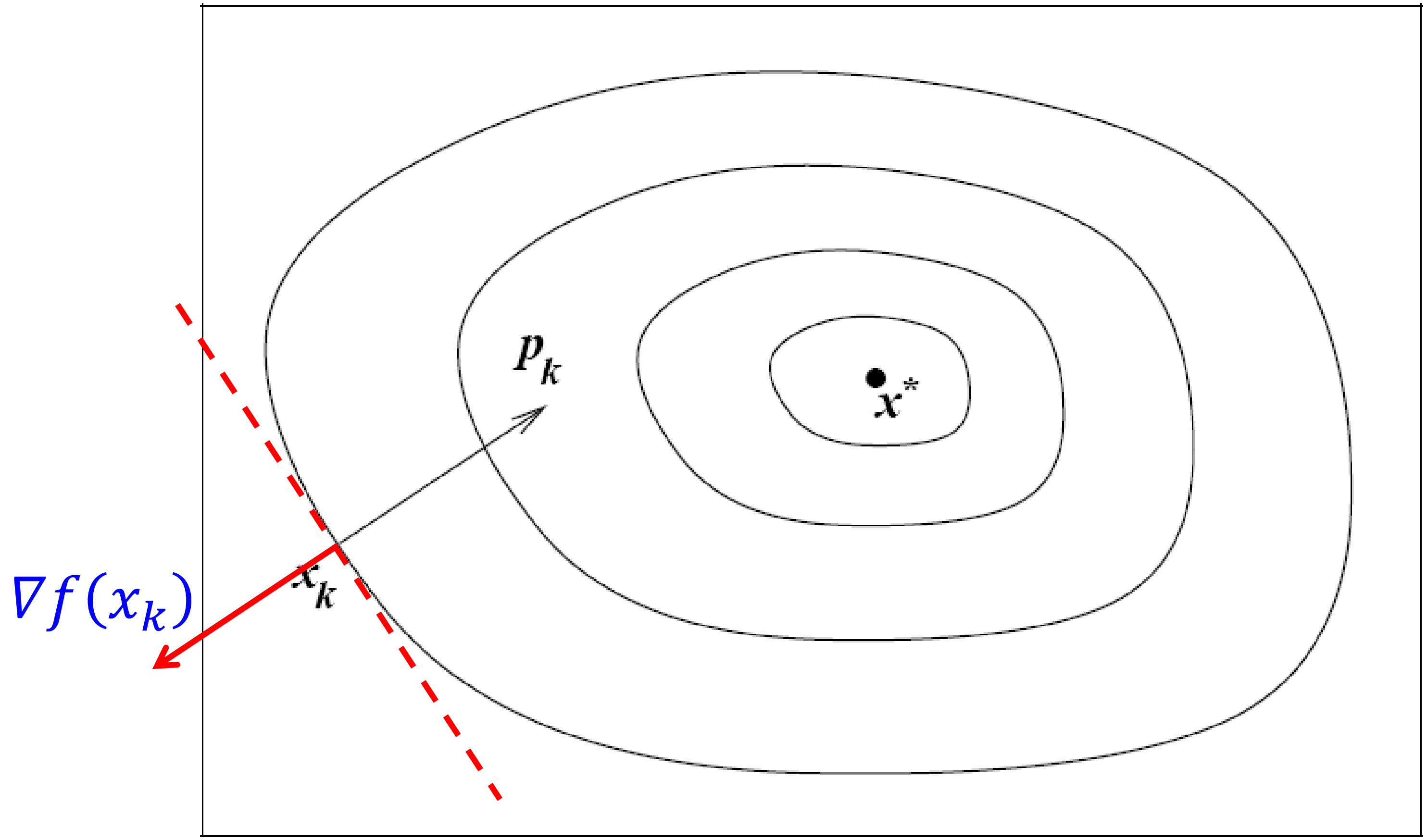
- 满足$\nabla f(\mathbf{x}_k)^T \mathbf{d}_k < 0$的方向$\mathbf{d}_k$被称为下降方向。由于

$$\nabla f(\mathbf{x}_k)^T \mathbf{d}_k = \|\nabla f(\mathbf{x}_k)\| \cdot \|\mathbf{d}_k\| \cdot \cos\theta$$

只要$\theta > \pi/2$时，$\mathbf{d}_k$就是下降方向。

- 特别地，若规格化搜索方向$\mathbf{d}_k$的模为1，那么当$\mathbf{d}_k$和$\nabla f(\mathbf{x}_k)$反向，即$\theta = \pi$时，$\nabla f(\mathbf{x}_k)^T \mathbf{d}_k$的值最小。$\mathbf{d}_k$即负梯度方向。

6.2 梯度法



6.2 梯度法

Step 1. 给定初始点$\mathbf{x}_0$和误差容限$\varepsilon > 0$，令$k := 0$。

Step 2. 计算负梯度方向$\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$。

Step 3. 检验是否满足收敛性准则：

$$\|\nabla f(\mathbf{x}_k)\| < \varepsilon$$

若满足则停止迭代，得到$\mathbf{x}^* := \mathbf{x}_k$；否则，进入Step 4。

Step 4. 进行一维搜索，即求解单变量极值问题

$$\min_{\alpha \geq 0} \ f(\mathbf{x}_k + \alpha \mathbf{d}_k)$$

得到步长α_k。

Step 5. 令$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$，且$k := k + 1$，转到Step 2。

- **例5.8** 用梯度法求解下面的二次函数极值问题：

$$\min f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \boldsymbol{Q}\mathbf{x} - \mathbf{c}^T\mathbf{x}$$

其中$\boldsymbol{Q} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 25 \end{pmatrix}$，$\mathbf{c} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$。

解：这里目标函数的负梯度方向是

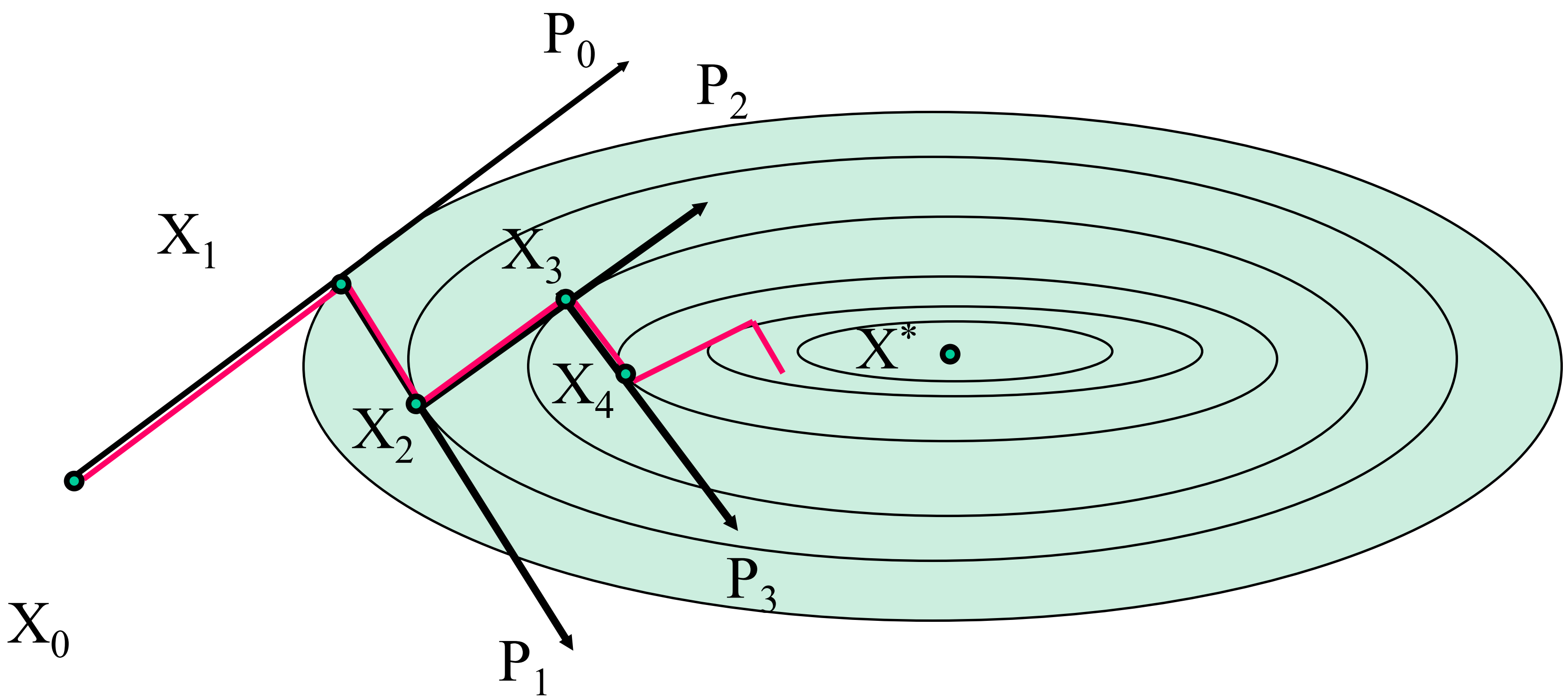
$$\mathbf{d}_k = -\nabla f(\mathbf{x}_k) = -(\boldsymbol{Q}\mathbf{x}_k - \mathbf{c})$$

做精确一维线搜索：

$$\alpha_k = -\frac{\nabla f(\mathbf{x}_k)^T \mathbf{d}_k}{\mathbf{d}_k^T \boldsymbol{Q}\mathbf{d}_k}$$

梯度法的特点：

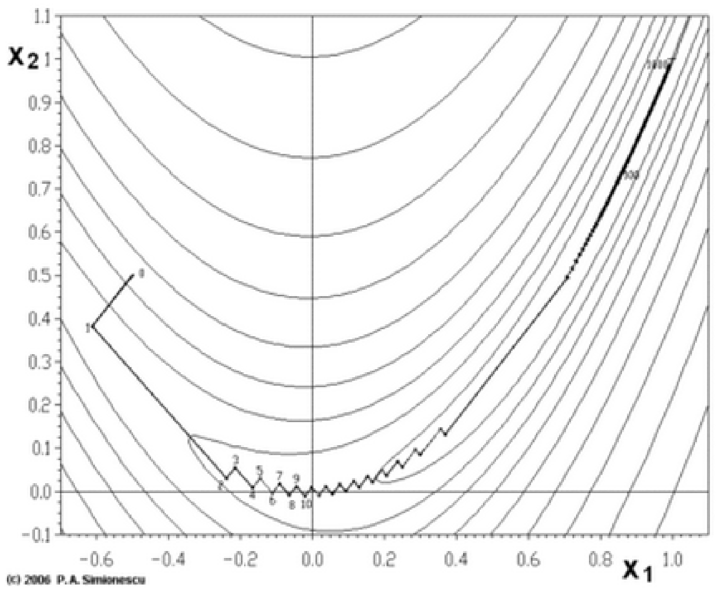
- ✓ 相邻两次寻优方向是垂直的；
- ✓ 寻优路线呈锯齿状（Zig-Zag），在极小点附近收敛缓慢。

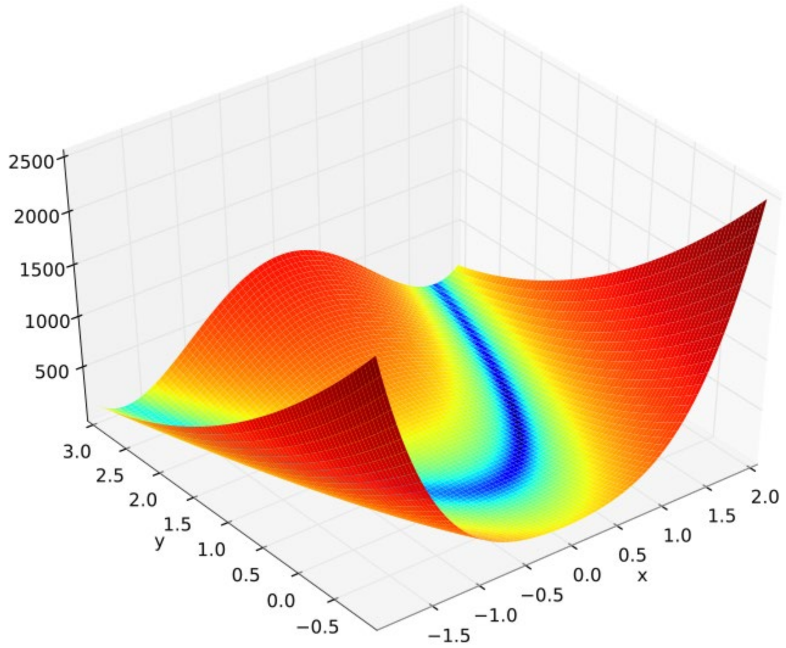


南京大学 NANJING UNIVERSITY 工程管理学院 SCHOOL OF MANAGEMENT & ENGINEERING

Rosenbrock function (香蕉函数):

$$f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$$





梯度下降法的应用

- 以线性回归模型为例，最小二乘为评价准则，介绍：批量梯度下降(Batch Gradient Descent, BGD)，随机梯度下降(Stochastic Gradient Descent, SGD)，小批量梯度下降(Mini-batch Gradient Descent, MBGD)
- 线性回归模型的假设函数(hypothesis function)为：

$$h_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{j=0}^{n} \theta_j x_j = \boldsymbol{\theta}^T \mathbf{x}$$

其中，$\mathbf{x} = (x_0, x_1, \dots, x_n)^T, x_0 = 1, \boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_n)^T$.

- 线性回归模型的**代价函数(cost function)**为：

$$J(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^{m} \left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right)^2$$

- 接下来，就是找到$\boldsymbol{\theta}$，使得$J(\boldsymbol{\theta})$最小（**采用梯度下降法**）。

梯度下降法的应用

- 梯度下降法是指按照梯度的方向改变$\boldsymbol{\theta}$的值：

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta})$$

其中，α为学习率(learning rate)，决定了$\boldsymbol{\theta}$改变的大小，即决定了学习速度的快慢。

$$J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{m}\left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right)^2 \Rightarrow \frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) = \sum_{i=1}^{m}\left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

$$\Rightarrow \theta_j := \theta_j - \alpha \sum_{i=1}^{m}\left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

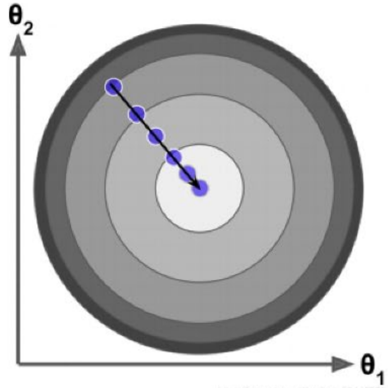
- 可以看到每次更新$\boldsymbol{\theta}$要用到训练集中的所有样本数据$\mathbf{x}^{(1)} \sim \mathbf{x}^{(m)}$，这种梯度下降方法就叫做**批量梯度下降**，每次更新$\boldsymbol{\theta}$要保证**所有样本的代价函数之和变化最快**。

梯度下降法的应用

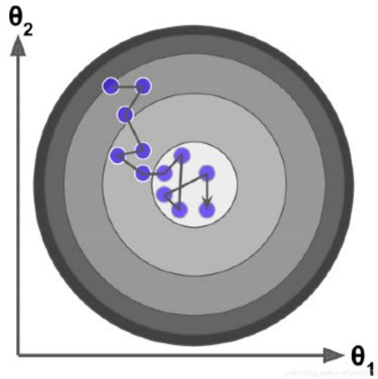
- **批量梯度下降**每次更新**θ**要用到训练集中的所有样本，对于训练集中样本数比较大的时候是不方便的，因此有了**随机梯度下降**。
- 随机梯度下降是指每次更新**θ**的时候只**随机**考虑一个样本的梯度值，不去考虑所有样本的梯度值，每次更新**θ**只保证**<u>某一个样本</u>的代价函数变化最快**，即此时的代价函数变成了

$$J(\boldsymbol{\theta}) = \frac{1}{2}\left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right)^2 \Rightarrow \frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) = \left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

$$\Rightarrow \theta_j \coloneqq \theta_j - \alpha\left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$



(批量梯度下降)



(随机梯度下降)

梯度下降法的应用

- 小批量梯度下降综合了上述两种方法，在每次更新$\boldsymbol{\theta}$的时候选择$\boldsymbol{k}$个样本，每次更新$\boldsymbol{\theta}$保证这k个**样本的代价函数之和变化最快**，即此时的代价函数变成了

$$J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{k=i}^{i+9}\left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(k)}\right) - y^{(k)}\right)^2$$

$$\Rightarrow \quad \theta_j := \theta_j - \alpha\sum_{k=i}^{i+9}\left(h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(k)}\right) - y^{(k)}\right)x_j^{(k)}$$

- 在线性回归模型中，**批量梯度下降**总能够找到全局最优解，它每一步都是考虑全部样本代价函数的梯度，因而迭代次数要少，但样本数比较多时速度慢；**随机梯度下降**不一定能够找到最优解，可能会在最优解附近徘徊，由于它每一步只考虑某个样本代价函数的梯度，所以在迭代过程中可能会发生全部样本的代价函数之和增加的现象，迭代次数比较多，但速度快；**小批量梯度下降**则综合考虑了两者，但对于“小”的程度，则需要具体设计。

6.3 牛顿法

- 在搜索方向上比梯度法有所改进
- 不仅利用了目标函数在搜索点的梯度（一阶导数）
- 而且还利用了目标函数在搜索点的二阶导数

（不但考虑了函数的梯度，还考虑了梯度的变化趋势，因而收敛速度较快）

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

由$\nabla f(\mathbf{x}) = 0$，得：

$$\nabla f(\mathbf{x}) \approx \nabla f(\mathbf{x}_k) + \nabla^2 f(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) = 0$$

$$\Longrightarrow \mathbf{x}_{k+1} = \mathbf{x}_k - [\nabla^2 f(\mathbf{x}_k)]^{-1} \nabla f(\mathbf{x}_k) = \mathbf{d}_k$$

Step 1. 给定初始点$\mathbf{x}_0$和误差容限$\varepsilon > 0$，令$k := 0$。

Step 2. 检验是否满足收敛性准则：

$$\|\nabla f(\mathbf{x}_k)\| < \varepsilon$$

若满足则停止迭代，得到$\mathbf{x}^* := \mathbf{x}_k$；否则，进入Step 3。

Step 3. 计算牛顿方向：$\mathbf{d}_k = -[\nabla^2 f(\mathbf{x}_k)]^{-1}\nabla f(\mathbf{x}_k)$。

Step 4. 令$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$，且$k := k+1$，转到Step 2。

- **例5.9** 用牛顿法求解下面的无约束极值问题：

$$\min\ (x_1-1)^4+(x_2-2)^2$$

解：这里目标函数的梯度和Hessian矩阵分别为

$$\nabla f(\mathbf{x})=\begin{pmatrix}4(x_1-1)^3\\2(x_2-2)\end{pmatrix},\ \nabla^2 f(\mathbf{x})=\begin{pmatrix}12(x_1-1)^2 & 0\\0 & 2\end{pmatrix}$$

第一次迭代：

$$\nabla f(\mathbf{x}_0)=\begin{pmatrix}-4\\-4\end{pmatrix},\ \nabla^2 f(\mathbf{x}_0)=\begin{pmatrix}12 & 0\\0 & 2\end{pmatrix}$$

$$\Rightarrow\ \mathbf{x}_1=\mathbf{x}_0-[\nabla^2 f(\mathbf{x}_0)]^{-1}\nabla f(\mathbf{x}_0)=\begin{pmatrix}1/3\\2\end{pmatrix},\ \nabla f(\mathbf{x}_1)=\begin{pmatrix}-32/27\\0\end{pmatrix}$$

6.3 牛顿法

牛顿法迭代结果

k	x_k	$\nabla f(x_k)$	$\Vert \nabla f(x_k) \Vert$
1	(0.333,2.000)	(-1.185,0.000)	1.19e+00
2	(0.556,2.000)	(-0.351,0.000)	3.51e-01
3	(0.704,2.000)	(-0.104,0.000)	1.04e-01
4	(0.802,2.000)	(-0.031,0.000)	3.08e-02
5	(0.868,2.000)	(-0.009,0.000)	9.13e-03
6	(0.912,2.000)	(-0.003,0.000)	2.71e-03
7	(0.941,2.000)	(-0.001,0.000)	8.02e-04
8	(0.961,2.000)	(-0.000,0.000)	2.38e-04
9	(0.974,2.000)	(-0.000,0.000)	7.04e-05
10	(0.983,2.000)	(-0.000,0.000)	2.09e-05
11	(0.988,2.000)	(-0.000,0.000)	6.18e-06
12	(0.992,2.000)	(-0.000,0.000)	1.83e-06
13	(0.995,2.000)	(-0.000,0.000)	5.43e-07

6.3 牛顿法

- **牛顿法的缺点：**

  计算复杂，每步迭代都需要计算目标函数的二阶偏导数（Hessian矩阵）和矩阵的逆。

- **拟牛顿法的基本思想：**

  用目标函数f以及其一阶导数∇f，构造Hessian矩阵的近似矩阵，由此获得一个搜索方向，生成新的迭代点。

  由于近似矩阵的构造方法不同，因此出现了不同的拟牛顿法。经过理论证明和实践检验，拟牛顿法已被公认为一类收敛速度比较快的无约束优化算法。

第五章 非线性规划

7. 约束极值问题

7.1 起作用约束

- **有效约束**：对于约束优化问题$\min_{\mathbf{x}\in\Omega} f(\mathbf{x})$，设点$\bar{\mathbf{x}}\in\Omega$，若有$g_i(\bar{\mathbf{x}})=0(1\le i\le l)$，则称不等式约束$g_i(\mathbf{x})\ge 0$为点$\bar{\mathbf{x}}$处的有效约束（起作用约束），且将下标集

$$A(\bar{\mathbf{x}})=\{i|g_i(\bar{\mathbf{x}})=0,\ 1\le i\le l\}$$

称为点$\bar{\mathbf{x}}$处的有效约束下标集（Active Set）。

- **正则点**：对于约束优化问题$\min_{\mathbf{x}\in\Omega} f(\mathbf{x})$，如果在可行点$\bar{\mathbf{x}}$处，各个有效约束的梯度向量，即$\{\nabla g_i(\bar{\mathbf{x}}),\ i\in A(\bar{\mathbf{x}})\}$，线性无关，则称点$\bar{\mathbf{x}}$是约束条件的一个正则点（Regular Point）。

7.2 可行方向和可行下降方向

$$\min \ f(\mathbf{x})$$
$$\text{s.t.} \ \ g_i(\mathbf{x}) \geq 0, i = 1,2,\cdots,l$$

- **可行方向**：任取一可行点$\mathbf{x}_0$，若对方向$\mathbf{d}$，存在$\alpha_0 > 0$，当$0 \leq \alpha \leq \alpha_0$时，下式成立

$$g_i(\mathbf{x}_0 + \alpha\mathbf{d}) \geq 0, \ \ i = 1,2,\cdots,l$$

则称$\mathbf{d}$为$\mathbf{x}_0$的可行方向（Feasible Direction）。

$$\begin{cases} \textit{case 1}: g_i(\mathbf{x}_0) = 0: \ g_i(\mathbf{x}_0 + \alpha\mathbf{d}) = g_i(\mathbf{x}_0) + \alpha\nabla g_i(\mathbf{x}_0)^T\mathbf{d} + o(\alpha) \\ \qquad \Rightarrow g_i(\mathbf{x}_0 + \alpha\mathbf{d}) \geq 0 \text{只要} \nabla g_i(\mathbf{x}_0)^T\mathbf{d} > 0 \text{成立} \\ \\ \textit{case 2}: g_i(\mathbf{x}_0) > 0: \text{由} g_i(\mathbf{x}) \text{连续} \Rightarrow g_i(\mathbf{x}_0 + \alpha\mathbf{d}) \geq 0 \text{只要} \alpha \text{足够小} \end{cases}$$

7.2 可行方向和可行下降方向

- **下降方向**：任取一可行点$\mathbf{x}_0$，若方向$\mathbf{d}$，满足$\nabla f(\mathbf{x}_0)^T\mathbf{d} < 0$，则称$\mathbf{d}$为$\mathbf{x}_0$的下降方向（Descent Direction）。

- **可行下降方向**：任取一可行点$\mathbf{x}_0$，若方向$\mathbf{d}$，既是该点的可行方向，又是该点的下降方向，那么就称它是这个点的可行下降方向（Feasible Descent Direction），即有

$$\begin{cases} \nabla f(\mathbf{x}_0)^T\mathbf{d} < 0 \\ \nabla g_i(\mathbf{x}_0)^T\mathbf{d} > 0,\ \ i \in A(\mathbf{x}_0) \end{cases}$$

即方向$\mathbf{d}$与$-\nabla f(\mathbf{x}_0)$和$\nabla g_i(\mathbf{x}_0)$的夹角都为锐角。

7.3 库恩-塔克条件

南京大学 NANJING UNIVERSITY 工程管理学院 SCHOOL OF MANAGEMENT & ENGINEERING

- **库恩-塔克（Kuhn-Tucker）条件**（简称**K-T条件**），是非线性规划领域中最重要的理论成果之一，是由H. W. Kuhn 和 A. W. Tucker在1951年提出的。
- K-T条件是确定某点为局部最优解的一阶必要条件，只要是最优点（同时是正则点）就必须满足这个条件。
- 但一般来说，它不是充分条件，即满足这个条件的点不一定是最优点。
- 然而，对于凸规划来说，K-T条件既是必要条件，也是充分条件。

南京大学 NANJING UNIVERSITY 工程管理学院 SCHOOL OF MANAGEMENT & ENGINEERING

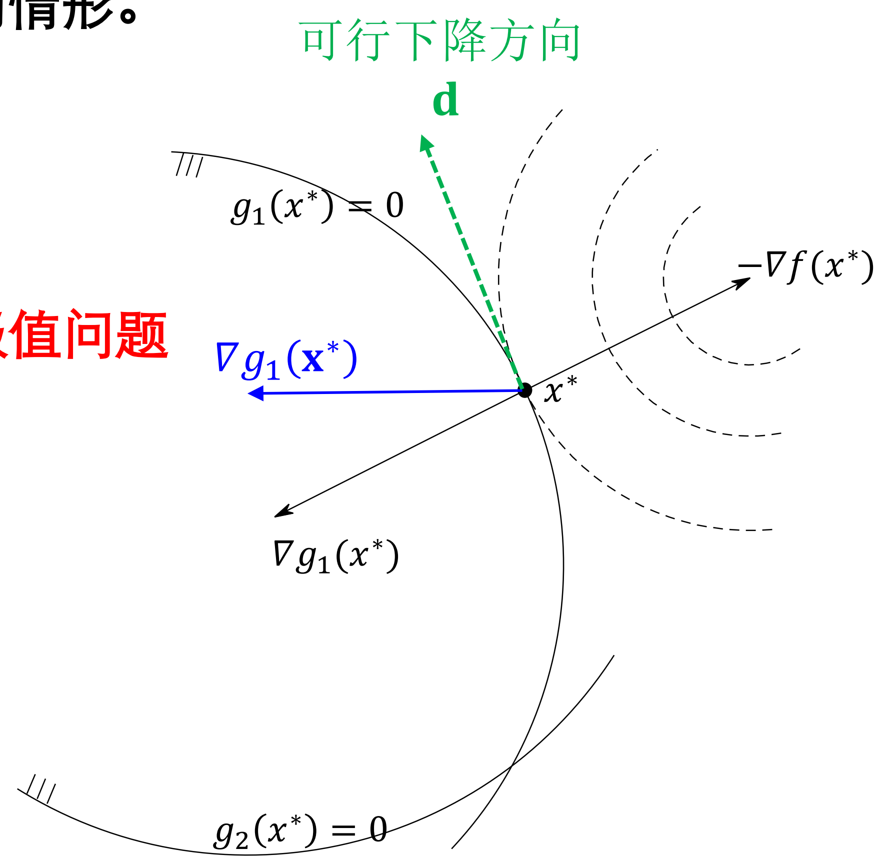
首先，考虑只含有不等式约束的情形。

$$\min\ f(\mathbf{x})$$
$$\text{s.t.}\ \ g_i(\mathbf{x}) \geq 0, i = 1,2,\cdots l$$

$$\begin{cases} \mathbf{x}^* \text{在可行域内部} \Leftrightarrow \text{无约束极值问题} \\ \mathbf{x}^* \text{在可行域边界上} \end{cases}$$

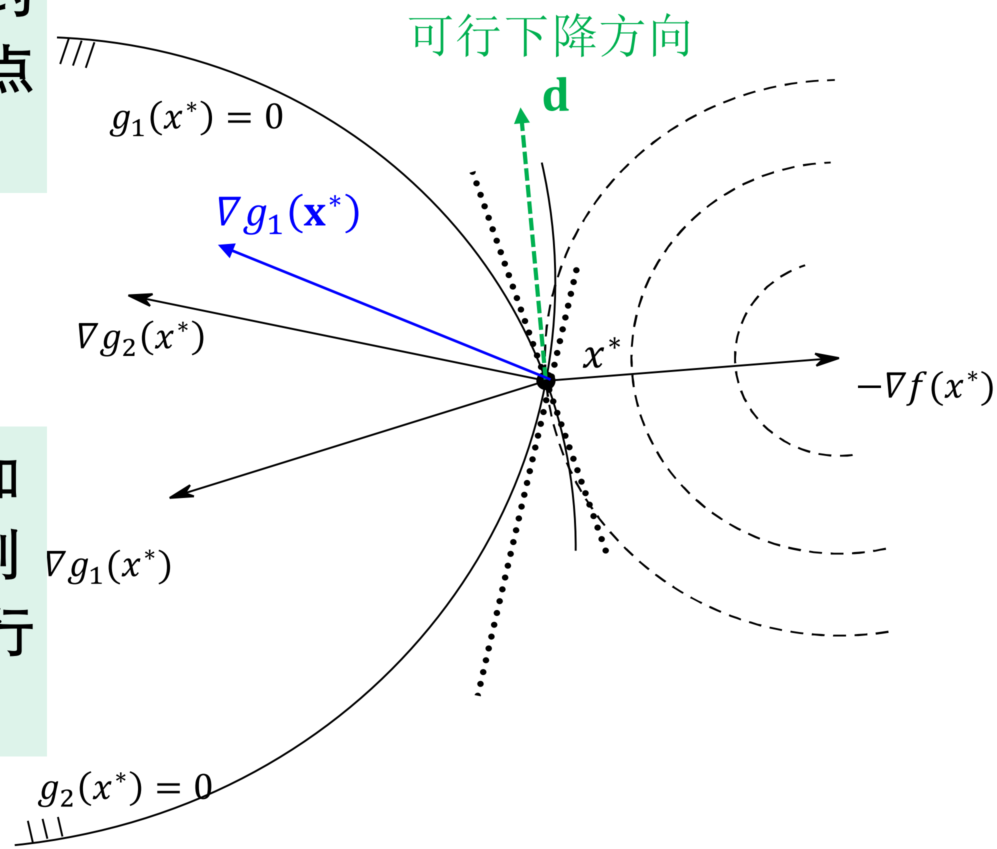
⇓

先假定它位于某一个约束条件形成的边界上，即在点$\mathbf{x}^*$只有一个起作用约束。



$$\nabla f(\mathbf{x}^*) - \mu_1^* \nabla g_1(\mathbf{x}^*) = 0，\ \mu_1^* \geq 0$$

接下来，假定$\mathbf{x}^*$同时位于两个约束条件所形成的边界上，即在点$\mathbf{x}^*$有两个起作用约束。



此时，$\nabla f(\mathbf{x}^*)$必位于$\nabla g_1(\mathbf{x}^*)$和$\nabla g_2(\mathbf{x}^*)$所形成的夹角内；否则也一定可以在点$\mathbf{x}^*$找到一个可行下降方向。

$$\nabla f(\mathbf{x}^*) - \mu_1^* \nabla g_1(\mathbf{x}^*) - \mu_2^* \nabla g_2(\mathbf{x}^*) = 0,\ \mu_1^*, \mu_2^* \geq 0$$

7.3 库恩-塔克条件

- 同理，按上述分析类推：

$$\nabla f(\mathbf{x}^*) - \sum_{i \in A(\mathbf{x}^*)} \mu_i^* \nabla g_i(\mathbf{x}^*) = 0,\ \ \mu_i^* \geq 0, i \in A(\mathbf{x}^*)$$

- 同时，为把不起作用约束也包含在内，可以增加一个互补松弛条件：

$$\mu_i^* g_i(\mathbf{x}^*) = 0,\ \ i = 1,2,\cdots,l$$

- 综合上面两组条件，可得：

$$\boxed{\begin{cases} \nabla f(\mathbf{x}^*) - \sum_{i=1}^{l} \mu_i^* \nabla g_i(\mathbf{x}^*) = 0 \\ \mu_i^* g_i(\mathbf{x}^*) = 0,\ \ i = 1,2,\cdots,l \\ \mu_i^* \geq 0,\ \ i = 1,2,\cdots,l \end{cases}}$$

K-T条件

对于同时含有等式和不等式约束的情形：

$$\min f(\mathbf{x})$$
$$\text{s.t.} \begin{cases} g_i(\mathbf{x}) \geq 0, i = 1,2,\cdots,l \\ h_j(\mathbf{x}) = 0, j = 1,2,\cdots,m \end{cases}$$

$$\begin{cases} h_j(\mathbf{x}) \geq 0 \\ -h_j(\mathbf{x}) \geq 0 \end{cases}$$

$$\begin{cases} \nabla f(\mathbf{x}^*) - \sum_{i=1}^{l} \mu_i^* \nabla g_i(\mathbf{x}^*) - \sum_{j=1}^{m} \lambda_j^* \nabla h_j(\mathbf{x}^*) = 0 \\ \mu_i^* g_i(\mathbf{x}^*) = 0,\ \ i = 1,2,\cdots l \\ \mu_i^* \geq 0,\ \ i = 1,2,\cdots l \end{cases}$$

完整K－T条件还要加上可行条件：

$$g_i(\mathbf{x}^*) \geq 0, i = 1,2,\cdots,l$$
$$h_j(\mathbf{x}^*) = 0, j = 1,2,\cdots,m$$

- K-T条件是确定某点为局部最优解的一阶必要条件，但一般来说，它不是充分条件。
- 只有对于凸规划来说，K-T条件既是必要条件，也是充分条件。

- **例5.12** 求下列非线性规划问题的K-T点：

$$\min \ f(\mathbf{x}) = (x_1 - 1)^2 + x_2$$

$$\text{s.t.} \ \begin{cases} g_1(\mathbf{x}) = x_1 + x_2 - 2 \le 0 \\ g_2(\mathbf{x}) = x_2 \ge 0 \end{cases}$$

解：由K-T条件，得：

$$\begin{cases} 2(x_1^* - 1) + \mu_1^* = 0 \\ 1 + \mu_1^* - \mu_2^* = 0 \\ x_1^* + x_2^* - 2 \le 0 \\ x_2^* \ge 0 \\ \mu_1^*(-x_1^* - x_2^* + 2) = 0 \\ \mu_2^* x_2^* = 0 \\ \mu_1^*, \mu_2^* \ge 0 \end{cases}$$

该问题只有一个K-T点$(1,0)^T$。

例题

- 考虑约束优化问题：

$$\begin{aligned} \min \ & (x_1-1)^2+(x_2-1)^2 \\ \text{s.t.} \ & x_1^2+x_2^2 \le 1 \end{aligned}$$

检验$\boldsymbol{x}^{(1)}=\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}$和$\boldsymbol{x}^{(2)}=\begin{pmatrix}-1/\sqrt{2}\\-1/\sqrt{2}\end{pmatrix}$是否为K-T点？

解： 这里 $\nabla f=\begin{pmatrix}2(x_1-1)\\2(x_2-1)\end{pmatrix}$，$\nabla g_1=\begin{pmatrix}2x_1\\2x_2\end{pmatrix}$，由一阶条件：

$$\nabla f+\mu_1\nabla g_1=0$$

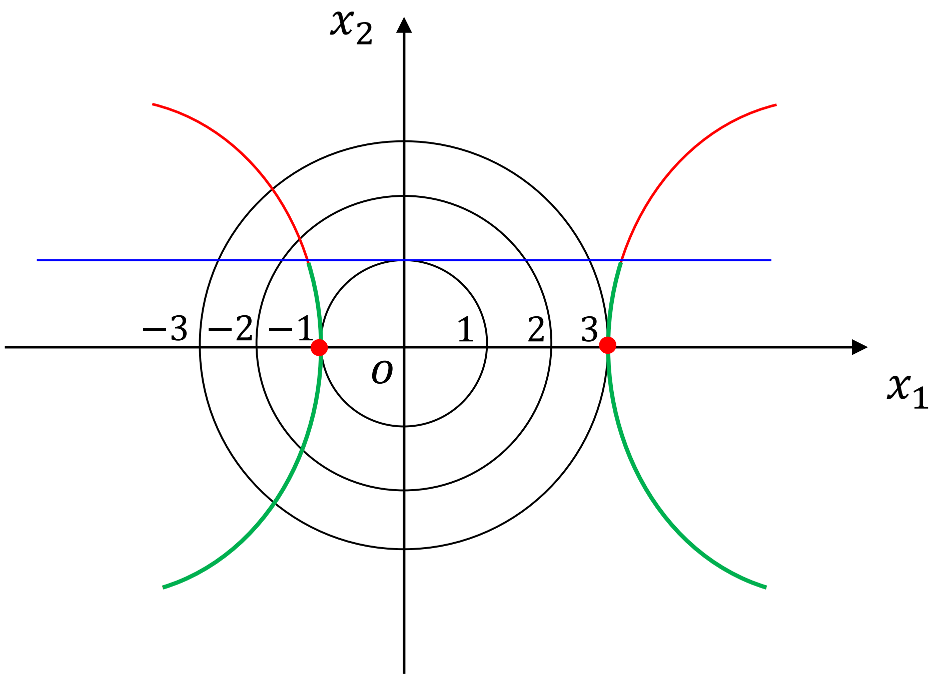
代入 $\boldsymbol{x}^{(1)}$，$\mu_1=\sqrt{2}-1$，因此 $\boldsymbol{x}^{(1)}$ 是K-T点；

代入 $\boldsymbol{x}^{(2)}$，$\mu_1=-\sqrt{2}-1$，因此 $\boldsymbol{x}^{(2)}$ 不是K-T点。

例题（局部极小点、全局极小点）

$$\min\ x_1^2 + x_2^2$$
$$s.t.\ \begin{cases}(x_1-1)^2 - x_2^2 - 4 = 0\\ x_2 - 1 \le 0\end{cases}$$

求出该非线性规划问题的K-T点。



- 局部极小点：

  $x^* = (-1, 0)$

  $\hat{x}^* = (3, 0)$

- 全局极小点：

  $x^* = (-1, 0)$

正则点（约束规范）

- 考虑约束优化问题：

$$\min \quad f(\boldsymbol{x}) = 3x_1 + 4x_2$$

$$s.t. \begin{cases} h_1(\boldsymbol{x}) = (x_1+1)^2 + x_2^2 = 1 \\ h_2(\boldsymbol{x}) = (x_1-1)^2 + x_2^2 = 1 \end{cases}$$

这里，$\boldsymbol{x}^* = (0,0)^T$是局部极小点。

$$\nabla h_1(\boldsymbol{x}) = \begin{pmatrix} 2(x_1+1) \\ 2x_2 \end{pmatrix}, \nabla h_2(\boldsymbol{x}) = \begin{pmatrix} 2(x_1-1) \\ 2x_2 \end{pmatrix}$$

计算可得：$\nabla h_1(\boldsymbol{x}^*) = (2,0)^T$和$\nabla h_2(\boldsymbol{x}^*) = (-2,0)^T$**线性相关**。

在该点，

$$\nabla_x L(\boldsymbol{x}^*, \boldsymbol{\mu}) = \begin{cases} 3 - 2\mu_1 + 2\mu_2 = 0 \\ 4 = 0 \end{cases}$$，显然无法成立！