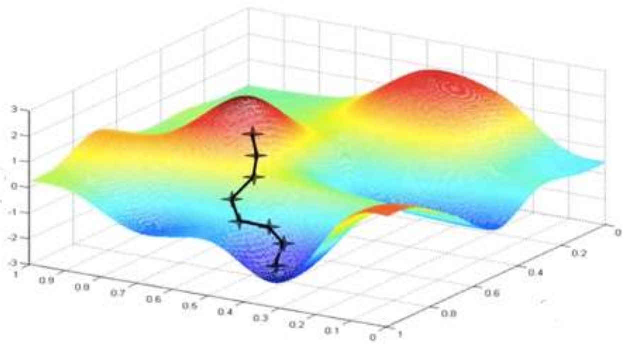
第六章 非线性规划

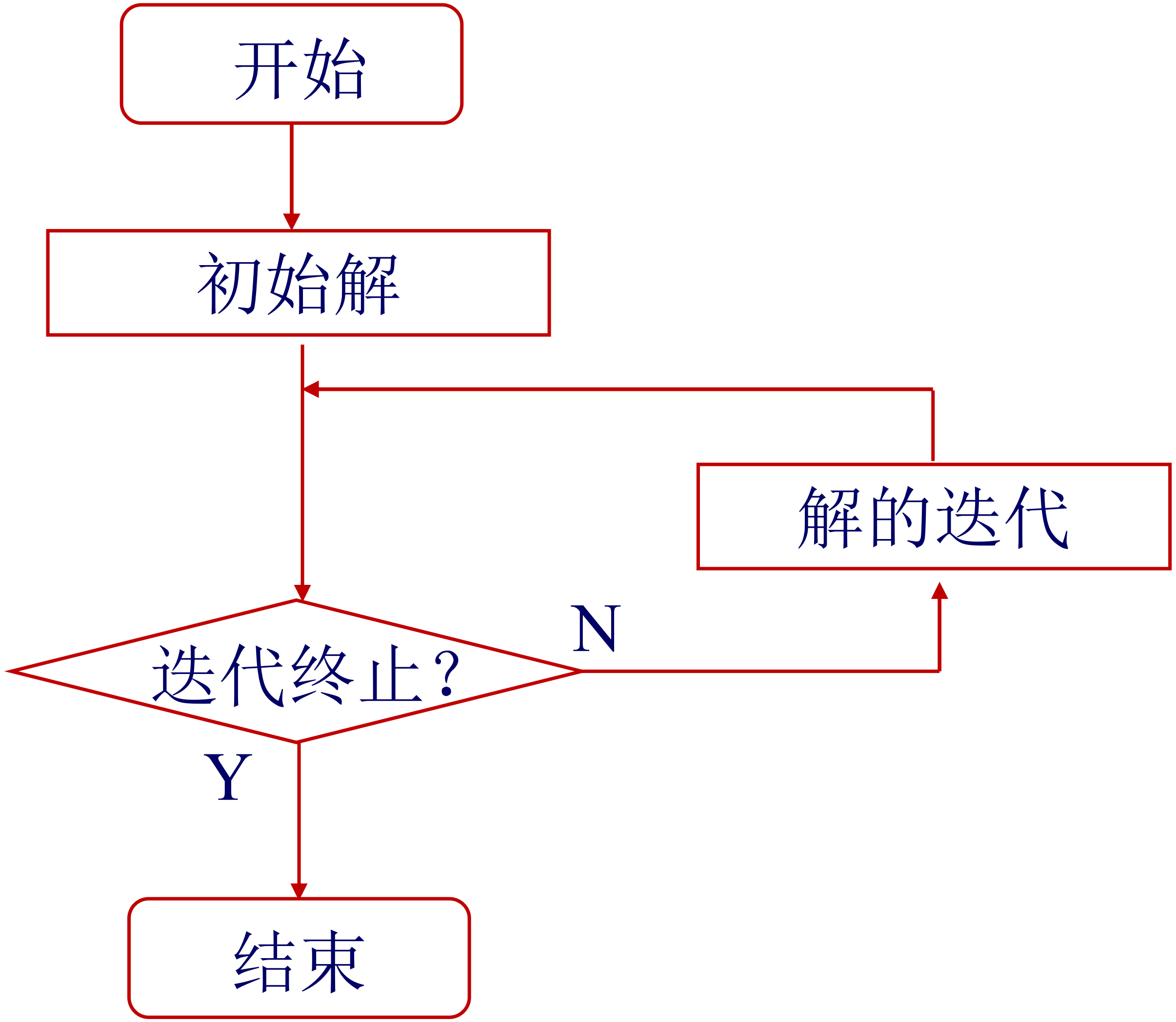
- 非线性规划数值解法
  - 无约束极值问题
    - 下降迭代法
  - 有约束极值问题
    - 可行方向法
    - 制约函数法
    - 逐次逼近法



数值求解的一般思路

- 思想
  - 迭代法
- 方法类型
  - 基于梯度的方法
  - 非梯度的方法（直接法、启发式方法）
- 问题种类
  - 无约束极值问题
  - 有约束极值问题

数值解法的流程图



下降迭代法

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{p}^{(k)} \quad \Rightarrow \quad f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$$

- 迭代方向
  - 基于梯度方法
    - 最速下降法
    - 牛顿法
    - 拟牛顿法
    - 共轭梯度法
  - 直接法
- 步长选择（一维搜索）
  - 基于梯度法
  - 直接法
- 迭代终止准则

最速下降法

- 设计思想：使目标函数值下降

目标函数值：$f(\boldsymbol{x}^{(k)}+\lambda_k \boldsymbol{p}^{(k)})=f(\boldsymbol{x}^{(k)})+\lambda_k \nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{p}^{(k)}+O(\lambda_k)$

可选方向：$\nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{p}^{(k)}=\|\nabla f(\boldsymbol{x}^{(k)})\|\cdot\|\boldsymbol{p}^{(k)}\|\cos\theta<0$ （方向导数）

下降最快方向：$\boldsymbol{p}^{(k)}=-\nabla f(\boldsymbol{x}^{(k)})$ 负梯度方向

性质：最优步长时 $\nabla f(\boldsymbol{x}^{(k+1)})\perp\nabla f(\boldsymbol{x}^{(k)})$

优点：计算量小。
缺点：“之”字迭代路径，接近极值点时尤为严重。
应用场合：迭代前期。

极小值点附近的等值面

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

极值点： $\nabla f(\boldsymbol{x}^*) = 0$ ⟹ $f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^T \nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*) + O(\|\boldsymbol{x}-\boldsymbol{x}^*\|^2)$

极值点附近等值面

$\boldsymbol{x}^{(k)} \in N(\boldsymbol{x}^*, \varepsilon)$ $f(\boldsymbol{x}) = c$ ⟹ $f(\boldsymbol{x}) = c \approx f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^T \nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)$

⟹ $(\boldsymbol{x}-\boldsymbol{x}^*)^T \nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*) \approx c' = 2[c - f(\boldsymbol{x}^*)]$

严格极小点：$\nabla^2 f(\boldsymbol{x}^*) > 0$ ⟹ $\boldsymbol{x}^*$为中心的椭球面

$\nabla^2 f(\boldsymbol{x}^*) \geq 0$ ⟹ 椭球柱面（降维椭球面）/抛物面/平行超平面

$$\frac{d(\boldsymbol{k}^T\boldsymbol{x})}{d\boldsymbol{x}}=\boldsymbol{k}$$

牛顿法

$$\frac{d\left(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\right)}{d\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}$$

- 设计思想：直接指向极值点。

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

两边对$\boldsymbol{x}$求导 ⟹ $\nabla f(\boldsymbol{x})\approx\nabla f(\boldsymbol{x}^{(k)})+\nabla^2 f(\boldsymbol{x}^{(k)})(\boldsymbol{x}-\boldsymbol{x}^{(k)})$

$\boldsymbol{x}=\boldsymbol{x}^*$ ⟹ $\nabla f(\boldsymbol{x}^*)=0\approx\nabla f(\boldsymbol{x}^{(k)})+\nabla^2 f(\boldsymbol{x}^{(k)})(\boldsymbol{x}^*-\boldsymbol{x}^{(k)})$

⟹ $\boldsymbol{x}^*\approx\boldsymbol{x}^{(k)}-[\nabla^2 f(\boldsymbol{x}^{(k)})]^{-1}\nabla f(\boldsymbol{x}^{(k)})$

⟹ $\boldsymbol{p}^{(k)}=-[\nabla^2 f(\boldsymbol{x}^{(k)})]^{-1}\nabla f(\boldsymbol{x}^{(k)})$ 迭代方向

优点：极值点附近收敛速率快。

缺点：计算量大，需要求二阶导数和矩阵逆。

远离极值点时，牛顿方向不一定是下降方向。

应用场合：二次目标函数或极值点附近，通常需采用进一步修正。

Levenberg-Marquardt修正

- 设计思想：使 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 变为正定矩阵，保证牛顿法方向为下降方向

L-M修正方向： $\boldsymbol{p}^{(k)} = -[\nabla^2 f(\boldsymbol{x}^{(k)}) + \mu_k \boldsymbol{I}]^{-1} \nabla f(\boldsymbol{x}^{(k)})$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - [\nabla^2 f(\boldsymbol{x}^{(k)}) + \mu_k \boldsymbol{I}]^{-1} \nabla f(\boldsymbol{x}^{(k)})$$

$\mu_k > |\lambda_{\min}^-|$ 　　 $\lambda^-_{\min}$为 $\nabla^2 f(\boldsymbol{x}^{(k)})$最小负特征值

$\mu \to 0$： 　　牛顿法

$\mu \to \infty$： 　　最速下降法

如果不求特征值，可以从较小的μ值试探

拟牛顿条件

- 设计思想：避免求二阶导数和逆。

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

在$\boldsymbol{x}^{(k+1)}$展开

两边对$\boldsymbol{x}$求导

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

$\boldsymbol{x}=\boldsymbol{x}^{(k)}$

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

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

$$\Longrightarrow \quad \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \approx [\nabla^2 f(\boldsymbol{x}^{(k+1)})]^{-1}[\nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})]$$

$$\boldsymbol{\delta}_k \triangleq \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \qquad \boldsymbol{H}_k \approx \left[\nabla^2 f(\boldsymbol{x}^{(k)})\right]^{-1} \qquad \boldsymbol{\gamma}_k \triangleq \nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})$$

$$\Longrightarrow \quad \boldsymbol{\delta}_k = \boldsymbol{H}_{k+1}\boldsymbol{\gamma}_k$$

拟牛顿条件

逆Hesse矩阵求解（变尺度法）

- 目标：给出一种求取H_k的方法

由**Davidon**提出，**Fletcher**和**Powell**改进，也称**DFP算法**。

迭代方向：
$$\boldsymbol{p}^{(k)} = -\boldsymbol{H}_k \nabla f(\boldsymbol{x}^{(k)})$$

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \Delta \boldsymbol{H}_k \qquad \boldsymbol{H}_1 = \boldsymbol{I}$$

$$\Delta \boldsymbol{H}_k = \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} - \frac{\boldsymbol{H}_k \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T \boldsymbol{H}_k}{\boldsymbol{\gamma}_k^T \boldsymbol{H}_k \boldsymbol{\gamma}_k}$$

$$\boldsymbol{\delta}_k \triangleq \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$$

$$\boldsymbol{\gamma}_k \triangleq \nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})$$

可以证明：**1**、H_k满足拟牛顿条件，为**Hesse**矩阵的逆。

2、当目标函数为正定二次函数时，可经有限步迭代收敛于极值（二次终止性）。为什么不是**1**步？

共轭梯度法

- 共轭方向定义：若A正定，且$\boldsymbol{p}^{(1)}$和$\boldsymbol{p}^{(2)}$满足$\boldsymbol{p}^{(1)T} A\, \boldsymbol{p}^{(2)}=0$，则$\boldsymbol{p}^{(1)}$、$\boldsymbol{p}^{(2)}$关于A共轭。（共轭是正交概念的推广）

$\Rightarrow$若$\boldsymbol{p}^{(1)}$、$\boldsymbol{p}^{(2)}$称关于A共轭，则$\boldsymbol{p}^{(1)}$和A$\boldsymbol{p}^{(2)}$正交。

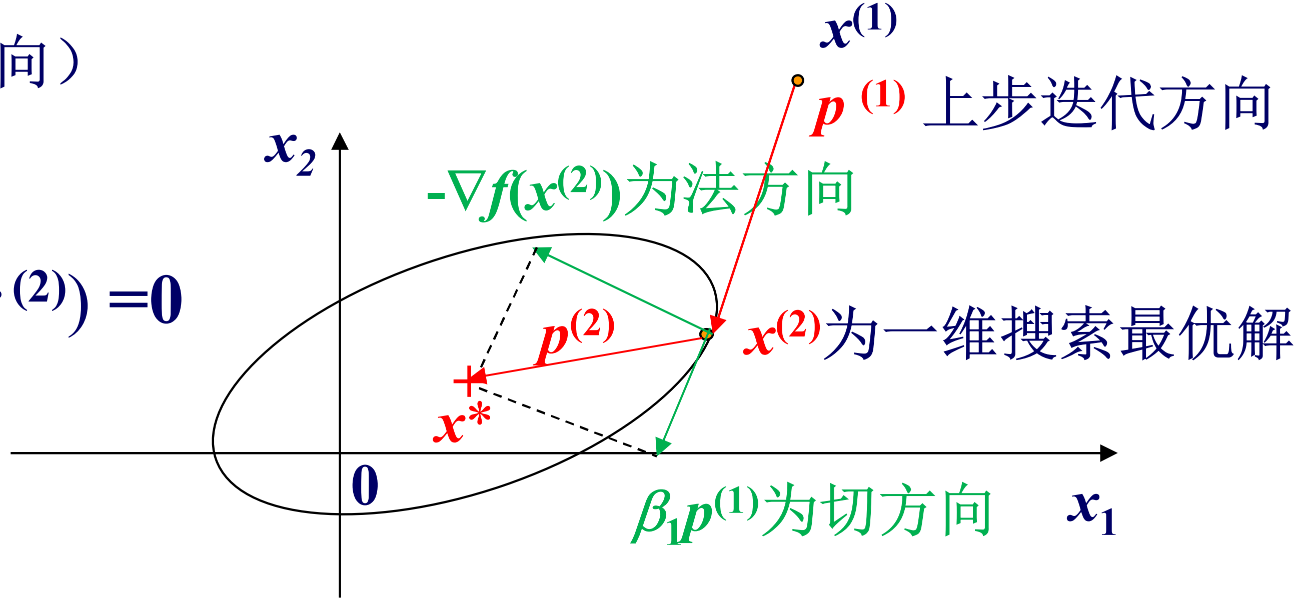
- 二次凸函数：$f(x)=1/2*(x-x^*)^T A\,(x-x^*)$

$$\nabla f(x)=A(x-x^*)$$

$\boldsymbol{p}^{(1)} := \boldsymbol{p}_t$（$x^{(2)}$点切线方向）

$\boldsymbol{p}^{(2)} := -(x^{(2)}-x^*)$

$$\boldsymbol{p}^{(1)T}A\boldsymbol{p}^{(2)} = -\boldsymbol{p}_t^{\,T}\nabla f(x^{(2)}) = 0$$



Fletcher-Reeves法

■ 设计思想：利用梯度构造共轭方向进行迭代

迭代方向： $\boldsymbol{p}^{(k+1)}=-\nabla f(\boldsymbol{x}^{(k+1)})+\beta_k \boldsymbol{p}^{(k)}$ $\quad \boldsymbol{p}^{(1)}=-\nabla f(\boldsymbol{x}^{(1)})$

将$\boldsymbol{p}^{(k+1)}$代入 共轭方程 $\quad \boldsymbol{p}^{(k)T}\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{p}^{(k+1)}=0$

$$\Rightarrow \quad -\boldsymbol{p}^{(k)T}\nabla^2 f(\boldsymbol{x}^{(k)})\nabla f(\boldsymbol{x}^{(k+1)})+\beta_k \boldsymbol{p}^{(k)T}\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{p}^{(k)}=0$$

$$\Rightarrow \quad \beta_k=\frac{\boldsymbol{p}^{(k)T}\nabla^2 f(\boldsymbol{x}^{(k)})\nabla f(\boldsymbol{x}^{(k+1)})}{\boldsymbol{p}^{(k)T}\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{p}^{(k)}}$$ Fletcher-Reeves法

注意： **n>2**时，$\boldsymbol{p}^{(1)}$ 的共轭方向不止**1**个， $-\nabla f(x^{(2)})+\beta_1 p^{(1)}$ 未必指向x^*

共轭方向组

- 共轭方向组定义：若A是n阶正定矩阵，且$p^{(1)}, p^{(2)} \ldots p^{(k)}$为$k$个关于$A$两两共轭的方向，则称这组方向关于$A$共轭。

- 性质：k≤n，则A的k个非零共轭方向线性无关。（反证法）

$$\sum_{i=1}^{k} \alpha_i \boldsymbol{p}^{(i)} = 0 \Rightarrow \sum_{i=1}^{k} \alpha_i \boldsymbol{p}^{(i)T} A \boldsymbol{p}^{(j)} = \alpha_j \boldsymbol{p}^{(j)T} A \boldsymbol{p}^{(j)} = 0 \Rightarrow \alpha_j = 0$$

- 定理：沿着A的n个非零共轭方向依次做一维最优步长搜索，则最多经过n步可找到二次凸函数的极小值。

数学归纳法 $\Rightarrow$

$$\left.\begin{array}{r} \nabla f(\boldsymbol{x}^{(k+1)})^T \boldsymbol{p}^{(j)} = 0 \\ 1 \le j \le k \le n \\ \\ \nabla f(\boldsymbol{x}^{(n+1)}) = \sum_{i=1}^{n} \alpha_i \boldsymbol{p}^{(i)} \end{array}\right\} \Rightarrow \nabla f(\boldsymbol{x}^{(n+1)})^T \left(\sum_{i=1}^{n} \alpha_i \boldsymbol{p}^{(i)} \right) = 0 \Rightarrow \nabla f(\boldsymbol{x}^{(n+1)}) = \mathbf{0}$$

等价公式

- 目标：避免求取二阶导数

共轭梯度法迭代方向：$$\boldsymbol{p}^{(k+1)} = -\nabla f(\boldsymbol{x}^{(k+1)}) + \beta_k \boldsymbol{p}^{(k)}$$

Crowder-Wolfe公式：
$$\beta_k = \frac{\nabla f(\boldsymbol{x}^{(k+1)})^T[\nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})]}{\boldsymbol{p}^{(k)T}[\nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})]}$$

Polak-Ribiere公式：
$$\beta_k = \frac{\nabla f(\boldsymbol{x}^{(k+1)})^T[\nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)})]}{\nabla f(\boldsymbol{x}^{(k)})^T \nabla f(\boldsymbol{x}^{(k)})}$$

Fletcher-Reeves公式：
$$\beta_k = \frac{\nabla f(\boldsymbol{x}^{(k+1)})^T \nabla f(\boldsymbol{x}^{(k+1)})}{\nabla f(\boldsymbol{x}^{(k)})^T \nabla f(\boldsymbol{x}^{(k)})}$$

下降迭代法

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{p}^{(k)} \quad \Rightarrow \quad f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$$

- 迭代方向
  - 基于梯度方法
    - 最速下降法
    - 牛顿法
    - 拟牛顿法
    - 共轭梯度法
  - 直接法
- 步长选择（一维搜索）
  - 基于梯度法
  - 直接法
- 迭代终止准则

迭代步长

设计思想：沿给定方向搜索目标函数值最小的距离

$$\lambda_k = \arg\min_{\lambda} f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{p}^{(k)})$$

- 基于梯度方法
- 直接法
  - **Fibonacci**斐波那契法
  - **0.618**法

基于梯度方法

目标： $\lambda_k = \arg\min_{\lambda} f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{p}^{(k)})$

$$f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{p}^{(k)}) = f(\boldsymbol{x}^{(k)}) + \lambda \nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{p}^{(k)} + \frac{1}{2}\lambda^2 \boldsymbol{p}^{(k)T} \nabla^2 f(\boldsymbol{x}^{(k)}) \boldsymbol{p}^{(k)} + O(\lambda^2)$$

极小值

$$\Rightarrow \frac{df(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{p}^{(k)})}{d\lambda} \approx \nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{p}^{(k)} + \lambda \boldsymbol{p}^{(k)T} \nabla^2 f(\boldsymbol{x}^{(k)}) \boldsymbol{p}^{(k)} = 0$$

$$\Rightarrow \boxed{\lambda^* = \frac{-\nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{p}^{(k)}}{\boldsymbol{p}^{(k)T} \nabla^2 f(\boldsymbol{x}^{(k)}) \boldsymbol{p}^{(k)}}}$$

特点：1、解析解，不需迭代

2、需求**Hesse**矩阵

3、$\boldsymbol{x}^{(k)}+\lambda \boldsymbol{p}^{(k)}$未必是极小值

最速下降法的性质

- 性质：最优步长时 $\nabla f(\boldsymbol{x}^{(k+1)}) \perp \nabla f(\boldsymbol{x}^{(k)})$
- 证明

$$\frac{df(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{p}^{(k)})}{d\lambda}$$

$$=\left(\frac{df(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{p}^{(k)})}{d(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{p}^{(k)})}\right)^T \frac{d(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{p}^{(k)})}{d\lambda}$$

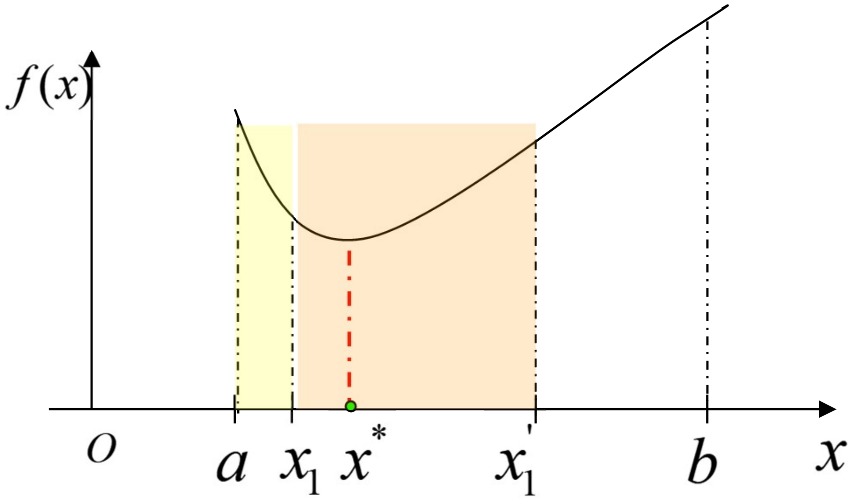
$$=\nabla f(\boldsymbol{x}^{(k+1)})^T \boldsymbol{p}^{(k)}$$

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

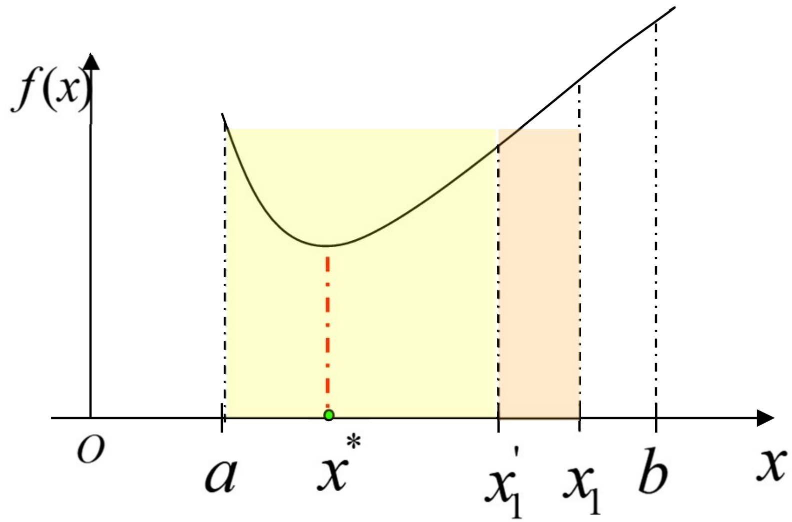
$$=0$$

分数法

- 假设：$f(x)$在$[a,b]$区间内单峰或无峰。假设是否合理？
- 特点：取x_1 , $x'_1 \in (a, b)$，将$[a,b]$分为3个区间。则最小值x^*必定位于x_1 、 x'_1两点中目标函数值较小一点所在的两个相邻区间内。
- 算法思想：将x_1 、 x'_1两点中目标函数值大的点作为新的边界，不断缩小最优值所在区间的范围。



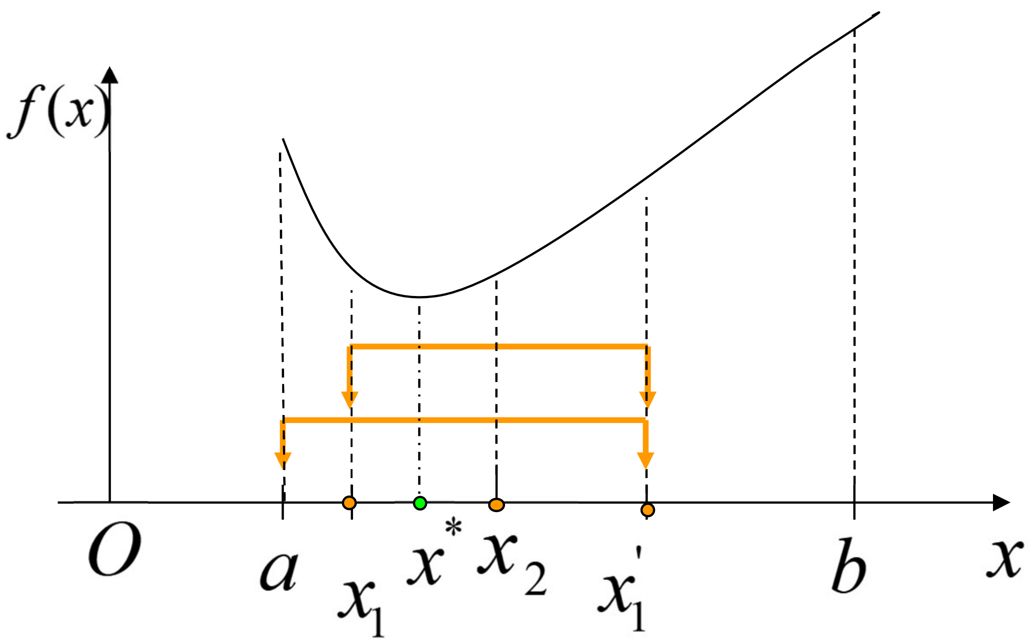
$x^* \in [a, x'_1]$



$x^* \in [a, x_1]$

搜索点选取

- 目标：计算量最小
- 方法：当前目标区的搜索点可直接用于下一次搜索



单峰函数

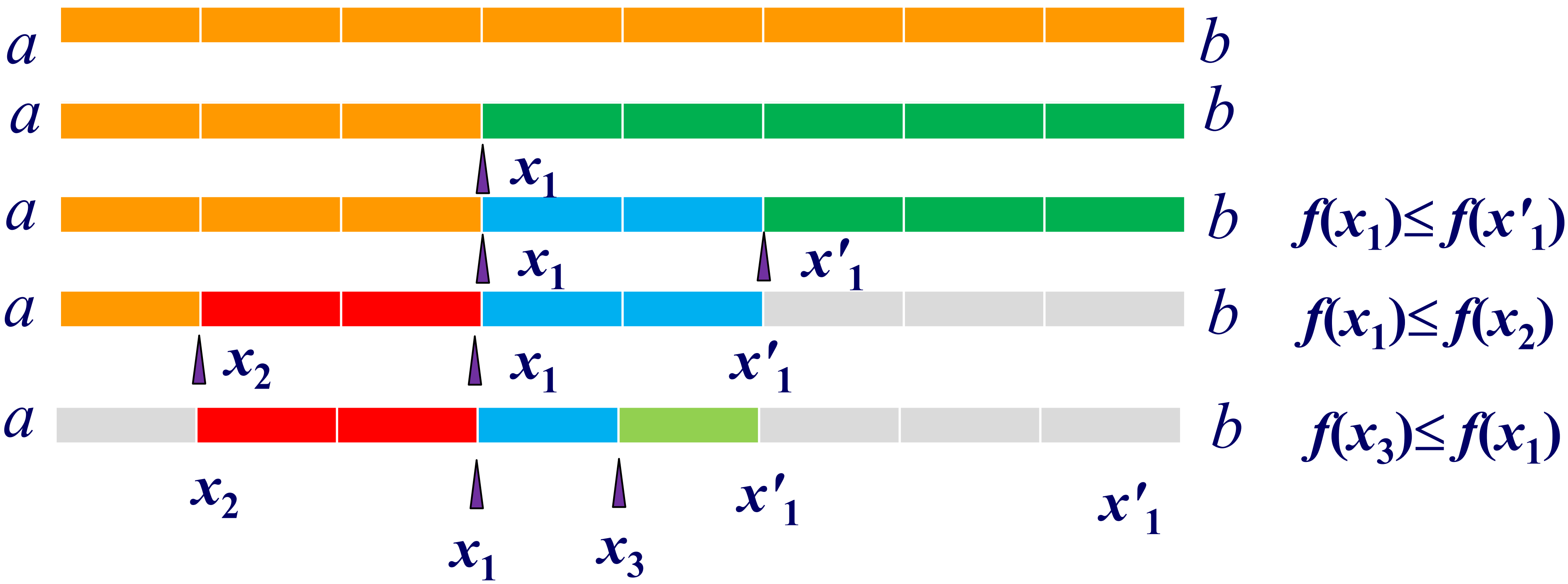
问题：如果单峰点是最大值会怎样？

斐波那契法

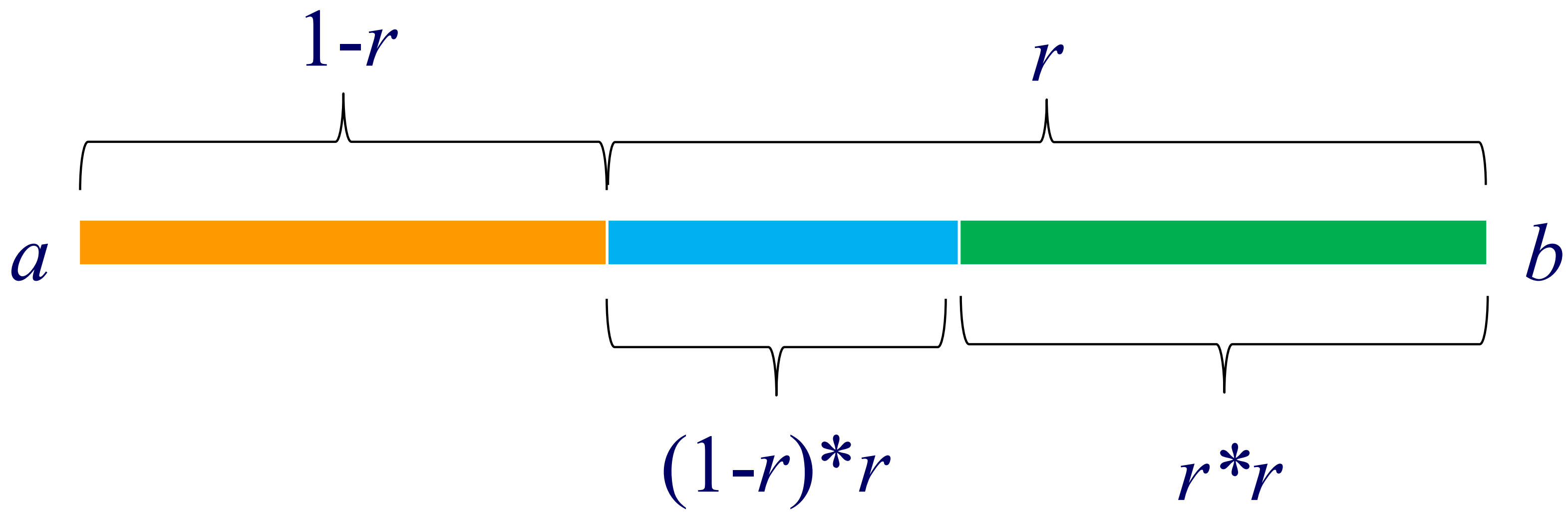
- **Fibonacci数列**

$$F_n=F_{n-1}+F_{n-2}, \quad n\geq 2 \quad F_0=1 \quad F_1=1$$

n	0	1	2	3	4	5
Fn	1	1	2	3	5	8



0.618法（黄金分割点法）



$$r*r=1-r \quad \Rightarrow \quad r^2+r-1=0 \quad \Rightarrow \quad r=0.618$$

可以证明：$\lim\limits_{n\to\infty}\dfrac{F_{n-1}}{F_n}=r=0.618$

下降迭代法

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{p}^{(k)} \quad \Rightarrow \quad f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$$

- 迭代方向
  - 基于梯度方法
    - 最速下降法
    - 牛顿法
    - 拟牛顿法
    - 共轭梯度法
  - 直接法
- 步长选择（一维搜索）
  - 基于梯度法
  - 直接法
- 迭代终止准则

迭代终止准则

- 迭代收敛准则
  - 绝对误差准则

$$\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \| \le \varepsilon_1 \qquad | f(\boldsymbol{x}^{(k+1)}) - f(\boldsymbol{x}^{(k)}) | \le \varepsilon_2$$

  - 相对误差准则

$$\frac{\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \|}{\| \boldsymbol{x}^{(k)} \|} \le \varepsilon_3 \qquad \frac{| f(\boldsymbol{x}^{(k+1)}) - f(\boldsymbol{x}^{(k)}) |}{| f(\boldsymbol{x}^{(k)}) |} \le \varepsilon_4$$

- 梯度模准则**(first-order optimality measure)**

$$\| \nabla f(\boldsymbol{x}^{(k)}) \| \le \varepsilon_5 \qquad \frac{\| \nabla f(\boldsymbol{x}^{(k)}) \|}{\| \nabla f(\boldsymbol{x}^{(0)}) \|} \le \varepsilon_6$$

完整算法举例

1、设置初始点$\boldsymbol{x}^{(1)}$、迭代终止阈值ε，$\boldsymbol{k}$=1；

2、如果 $\|\nabla f(\boldsymbol{x}^{(k)})\| \le \varepsilon$ ，$\boldsymbol{x}^{*}=\boldsymbol{x}^{(k)}$ ，迭代结束。

3、否则继续迭代

$$\boldsymbol{p}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$$

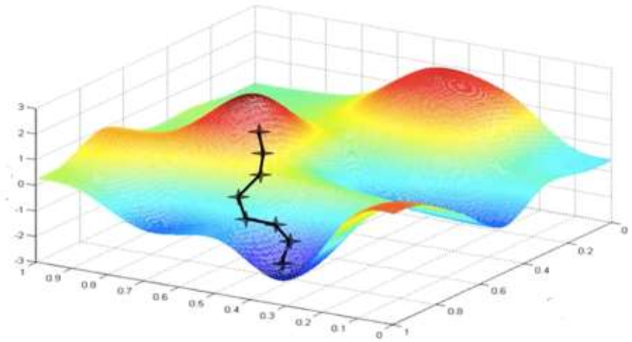
$$\lambda_k = \arg\min_{\lambda} f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{p}^{(k)})$$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{p}^{(k)}$$

$\boldsymbol{k} \leftarrow \boldsymbol{k}+1$，返回第2步。

第六章 非线性规划

- 非线性规划数值解法
  - 无约束极值问题
    - 下降迭代法
  - 有约束极值问题
    - 可行方向法
    - 制约函数法
    - 逐次逼近法



约束极值问题的数值解

- 可行方向法
  - **Zoutendijk**可行方向法
- 制约函数法
  - 外点法
  - 内点法
  - 混合法
- 逐次逼近法（近似规划法）
  - **SLP (Sequential Linear Programming)**
  - **SQP (Sequential Quadratic Programming)**

Zoutendijk可行方向法

- 可行下降方向

$$\begin{cases} \nabla f(\boldsymbol{x}^{k)})^T \boldsymbol{p} < 0 \\ -\nabla g_j(\boldsymbol{x}^{(k)})^T \boldsymbol{p} < 0 \quad j \in J \end{cases}$$

J为起作用约束集合

$$\Leftrightarrow \begin{cases} \nabla f(\boldsymbol{x}^{k)})^T \boldsymbol{p} < \eta \\ -\nabla g_j(\boldsymbol{x}^{(k)})^T \boldsymbol{p} < \eta \quad j \in J \\ \eta < 0 \end{cases}$$

$X^{(k)}$为可行点

$$\begin{aligned} &\min \eta \\ s.t.\quad &\nabla f(\boldsymbol{x}^{k)})^T \boldsymbol{p} \le \eta \\ &-\nabla g_j(\boldsymbol{x}^{(k)})^T \boldsymbol{p} \le \eta \quad j \in J \\ &-1 \le \boldsymbol{p} \le 1 \end{aligned}$$

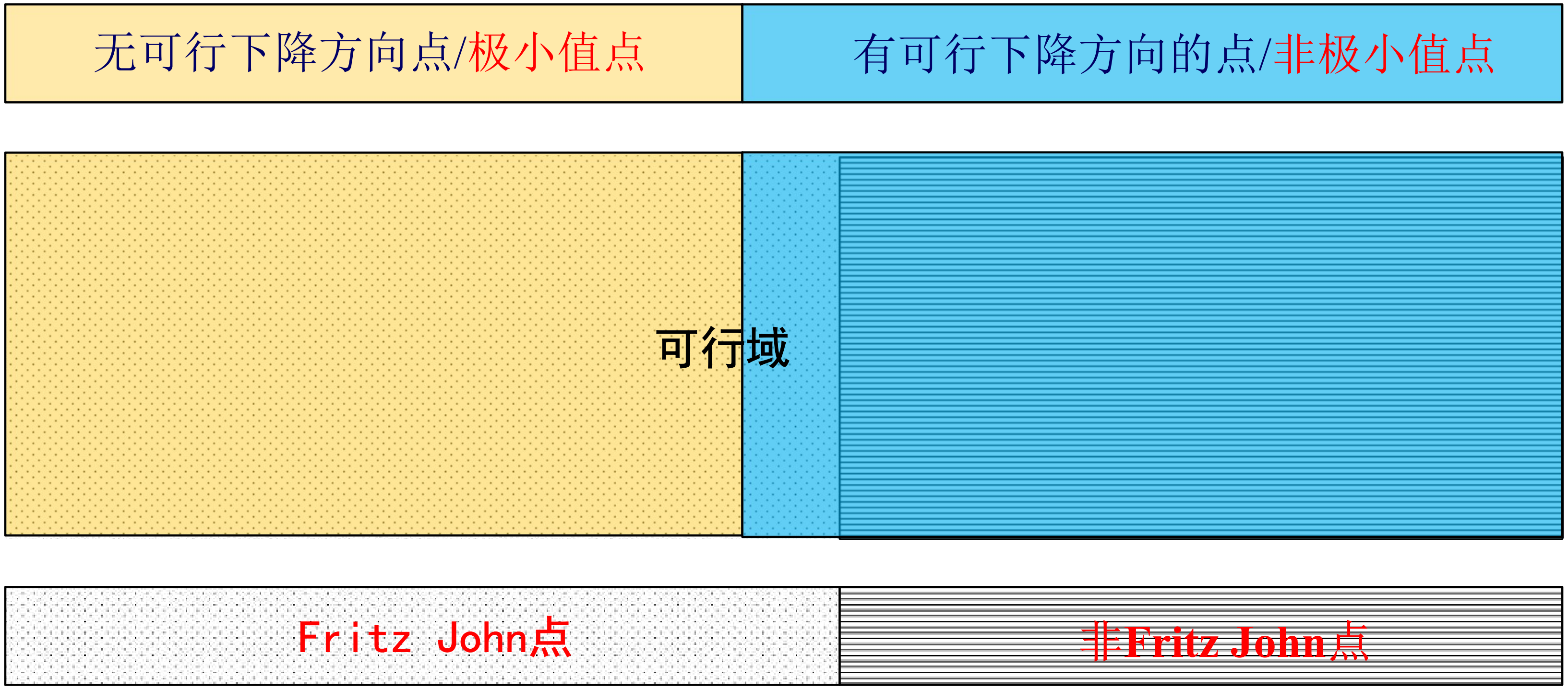
$\eta<0$: $\boldsymbol{p}$为$\boldsymbol{x}^{(k)}$的可行下降方向

$\eta=0$：$\boldsymbol{x}^{(k)}$为原非线性规划问题的Fritz John点

$\eta>0$:　会不会出现?

$\eta=0$时是否是极小值？

Zoutendijk法集合分析



Fritz John点

不存在与$\nabla f(\boldsymbol{X})$和所有$-\nabla g_{j\in J}(\boldsymbol{X})$夹角大于90方向的点

$\nabla f(\boldsymbol{X})$和$-\nabla g_{j\in J}(\boldsymbol{X})$正线性相关的点

等价问题最优值 $\eta^*=0$的点

非Fritz John点

存在与$\nabla f(\boldsymbol{X})$和所有$-\nabla g_{j\in J}(\boldsymbol{X})$夹角大于90方向的点

$\nabla f(\boldsymbol{X})$和$-\nabla g_{j\in J}(\boldsymbol{X})$不是正线性相关的点

等价问题最优值$\eta^*<0$的点

$\eta^*=0$只能找到Fritz John点，但未必是极值点！

约束极值问题的数值解

- 可行方向法
  - **Zoutendijk**可行方向法
- 制约函数法
  - 外点法
  - 内点法
  - 混合法
- 逐次逼近法（近似规划法）
  - **SLP (Sequential Linear Programming)**
  - **SQP (Sequential Quadratic Programming)**

制约函数法

- 思想：化为无约束极值问题求解，
- 名称：也称为序列无约束极小化技术，**SUMT(Sequential Unconstrained Minimization Technique)**
- 制约函数法的种类
  - 外点法
    - 从可行域外部逼近极值
  - 内点法
    - 从可行域内部逼近极值
  - 混合法
    - 内点法和外点法的结合

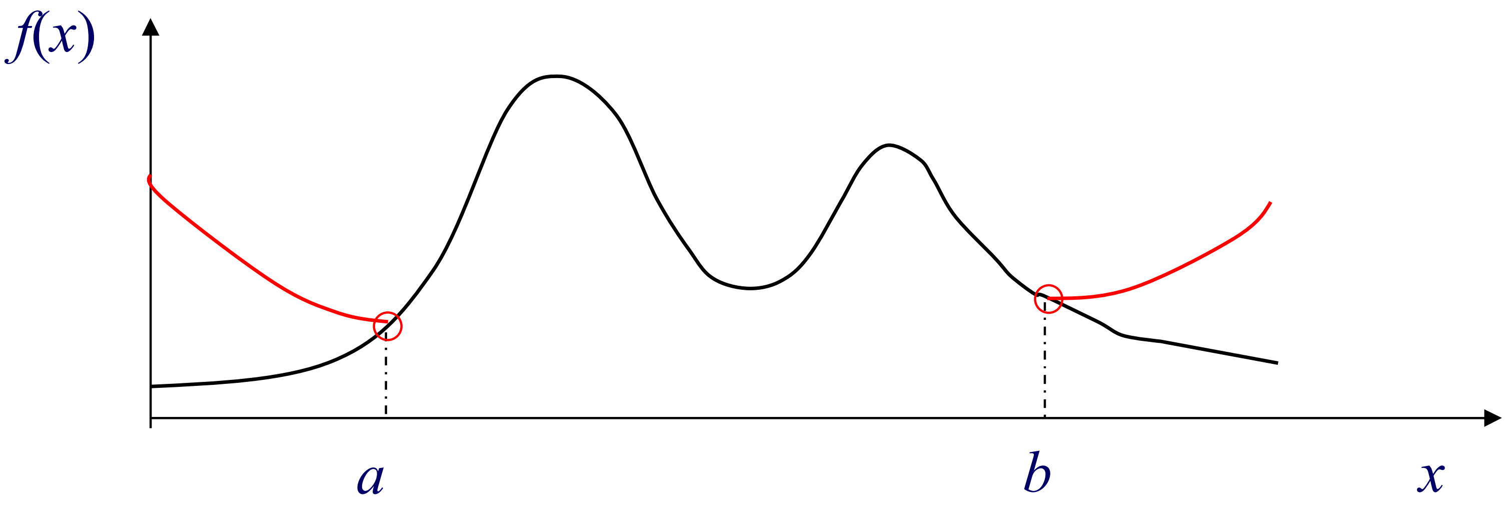
外点法

- 构造罚函数

$$\min P(\boldsymbol{x}, M) = f(\boldsymbol{x}) + M\sum_{i=1}^{m} h_i^2(\boldsymbol{x}) + M\sum_{j=1}^{l}[\min(0, g_j(\boldsymbol{x}))]^2$$

Courant罚函数

M*>0**为罚因子，当M***趋向无穷时，***x****为原问题约束极值解。

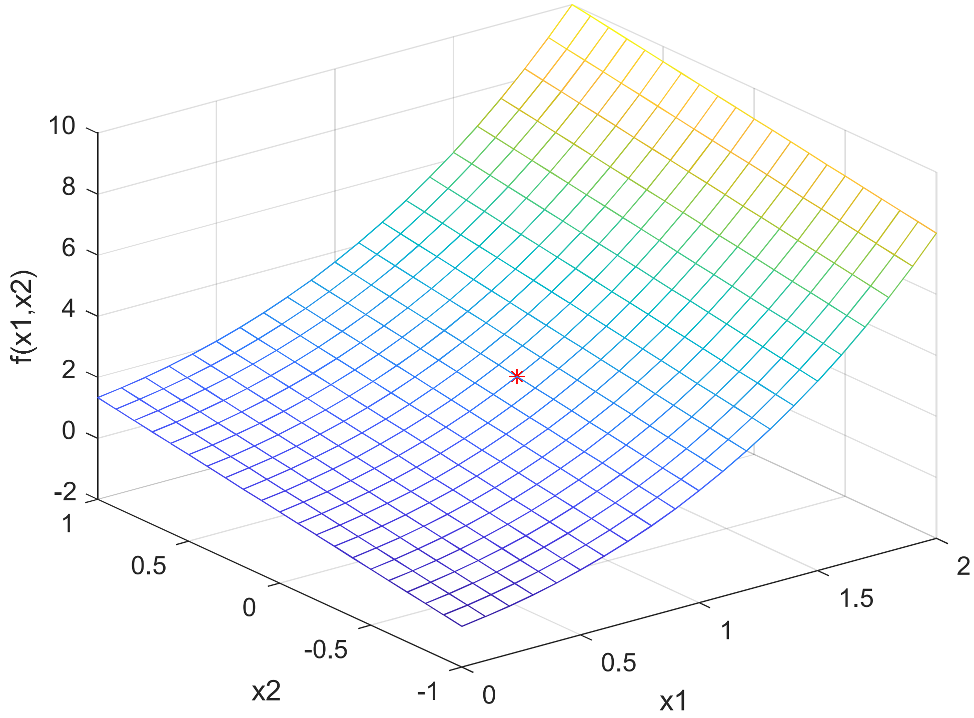


解析分析举例

$$\min f(\boldsymbol{x})=\frac{1}{3}(x_1+1)^3+x_2$$

s.t. $x_1-1\geq 0$

$x_2\geq 0$



$$\nabla f(\boldsymbol{x})=\begin{bmatrix}(x_1+1)^2\\ 1\end{bmatrix}\neq 0$$

没有内点极值

求解

构造罚函数：

$$P(\boldsymbol{x}, M) = \frac{1}{3}(x_1 + 1)^3 + x_2 + M[\min(0, x_1 - 1)]^2 + M[\min(0, x_2)]^2$$

根据一阶驻点条件，有

$$\frac{\partial P}{\partial x_1} = (x_1 + 1)^2 + 2M[\min(0, x_1 - 1)] = 0$$

$$\frac{\partial P}{\partial x_2} = 1 + 2M[\min(0, x_2)] = 0$$

如果$x_1 \geq 1 \Rightarrow x_1 = -1$，矛盾

如果$x_2 \geq 0 \Rightarrow 1 = 0$，不成立

所以考虑$x_1 < 1, x_2 < 0$区域的驻点。

$$\frac{\partial P}{\partial x_1} = (x_1 + 1)^2 + 2M(x_1 - 1) = 0$$

$$\frac{\partial P}{\partial x_2} = 1 + 2Mx_2 = 0$$

可得：

$$\begin{cases} x_1 = -1 - M \pm \sqrt{M^2 + 4M} \\ x_2 = -\dfrac{1}{2M} \end{cases} \xRightarrow{M \to +\infty} \begin{cases} x_1^* = 1 \\ x_2^* = 0 \end{cases}$$

$$\nabla^2 P(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 + 1) + 2M & 0 \\ 0 & 2M \end{bmatrix} > 0$$

故为极小值

外点法的数值解法

- 构造罚函数

$$\min P(\boldsymbol{x}, M_k) = f(\boldsymbol{x}) + M_k \sum_{i=1}^{m} h_i^2(\boldsymbol{x}) + M_k \sum_{j=1}^{l} [\min(0, g_j(\boldsymbol{x}))]^2$$

1、 取$M_1>0$（通常$M_1=1$），允许误差$\varepsilon>0$，$k:=1$

2、 求$\min P(x,M_k)$，得$x^{*(k)}$。

3、 若存在$-g_j(x^{*(k)})>\varepsilon$ 或 $|h_i(x^{*(k)})|>\varepsilon$，取$M_{k+1}=cM_k$ (c>1,通常取5或10），$k:=k+1$转第2步重新求解。否则，停止迭代，$x^*=x^{*(k)}$。

定理：若$\varepsilon>0$，上述算法必在有限步内终止。

外点法聚点分析

- 罚函数

$$P(\boldsymbol{x}, M_k) = f(\boldsymbol{x}) + M_k \sum_{i=1}^{m} h_i^2(\boldsymbol{x}) + M_k \sum_{j=1}^{l} [\min(0, g_j(\boldsymbol{x}))]^2$$

- 第**k**步罚函数的局部极小值满足

$$\nabla P(\boldsymbol{x}^{*(k)}, M_k) = \nabla f(\boldsymbol{x}^{*(k)}) + 2M_k \sum_{i=1}^{m} h_i(\boldsymbol{x}^{*(k)}) \nabla h_i(\boldsymbol{x}^{*(k)}) + 2M_k \sum_{g_j(\boldsymbol{x}_k^*) \le 0} g_j(\boldsymbol{x}^{*(k)}) \nabla g_j(\boldsymbol{x}^{*(k)}) = 0$$

$$\nabla P(\boldsymbol{x}^{*(k)}, M_k) = \nabla f(\boldsymbol{x}^{*(k)}) - \sum_{i=1}^{m} \lambda_i^*(k) \nabla h_i(\boldsymbol{x}^{*(k)}) - \sum_{g_j(\boldsymbol{x}_k^*) \le 0} \mu_j^*(\boldsymbol{x}_k^*) \nabla g_j(\boldsymbol{x}^{*(k)}) = 0$$

$$\lambda_i^*(k) = -2M_k h_i(\boldsymbol{x}^{*(k)}) \qquad \mu_i^*(k) = -2M_k g_j(\boldsymbol{x}^{*(k)}) \ge 0$$

- $\boldsymbol{x}_k$*迭代收敛时

$$\nabla P(\boldsymbol{x}^*, M) = \nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\boldsymbol{x}^*) - \sum_{g_j(\boldsymbol{x}^*) \to 0} \mu_j^* \nabla g_j(\boldsymbol{x}^*) = 0$$

外点法特点

1、只有**M**$\to\infty$，才保证$\lambda^*(k)\to\lambda^*$、$\mu^*(k)\to\mu^*$， $x_k^*\to x^*$。为减小收敛点的误差，需$M_k\to\infty$ 。

2、M_k过大，有可能导致**Hesse**矩阵病态（条件数很大），极小值将位于狭长的深谷，导致目标函数值对搜索方向敏感。

$$cond(\boldsymbol{H})=\|\boldsymbol{H}\|\cdot\|\boldsymbol{H}^{-1}\|\geq \lambda_{\max}/\lambda_{\min}$$

3、改进策略

1）设计精确罚函数（例如l_1罚函数），令有限M_k的$P(x,M_k)$驻点巧好为x^*。

2）构造增广**Lagrange**函数对罚函数进行修正，在有限M_k时也能使x_k^*逼近最优解x^* 。

内点法

- 构造障碍函数

$$\min_{x\in R_0} \bar{P}(\boldsymbol{x},r)=f(\boldsymbol{x})+r\sum_{j=1}^{l}\frac{1}{g_j(\boldsymbol{x})}$$

或

$$\min_{x\in R_0} \bar{P}(\boldsymbol{x},r)=f(\boldsymbol{x})-r\sum_{j=1}^{l}\log g_j(\boldsymbol{x})$$

其中 $R_0=\{\boldsymbol{x}\mid g_j(\boldsymbol{x})>0, j=1,2,\cdots,l\}$ 严格内点

r>0为障碍因子，其在迭代中的取值会不断减小，趋向于**0**，使x可趋向于边界。

解析分析举例

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

s.t. $-x_1^2 + x_2 \geq 0$

$$x_1 \geq 0$$

求解

构造障碍函数：

$$\overline{P}(\boldsymbol{x},r)=x_1+x_2-r\cdot[\log(-x_1^2+x_2)+\log(x_1)]$$

根据驻点一阶条件，有

$$\frac{\partial\overline{P}}{\partial x_1}=1-r\cdot\frac{-2x_1}{-x_1^2+x_2}-r\cdot\frac{1}{x_1}=0$$

$$\frac{\partial\overline{P}}{\partial x_2}=1-r\cdot\frac{1}{-x_1^2+x_2}=0$$

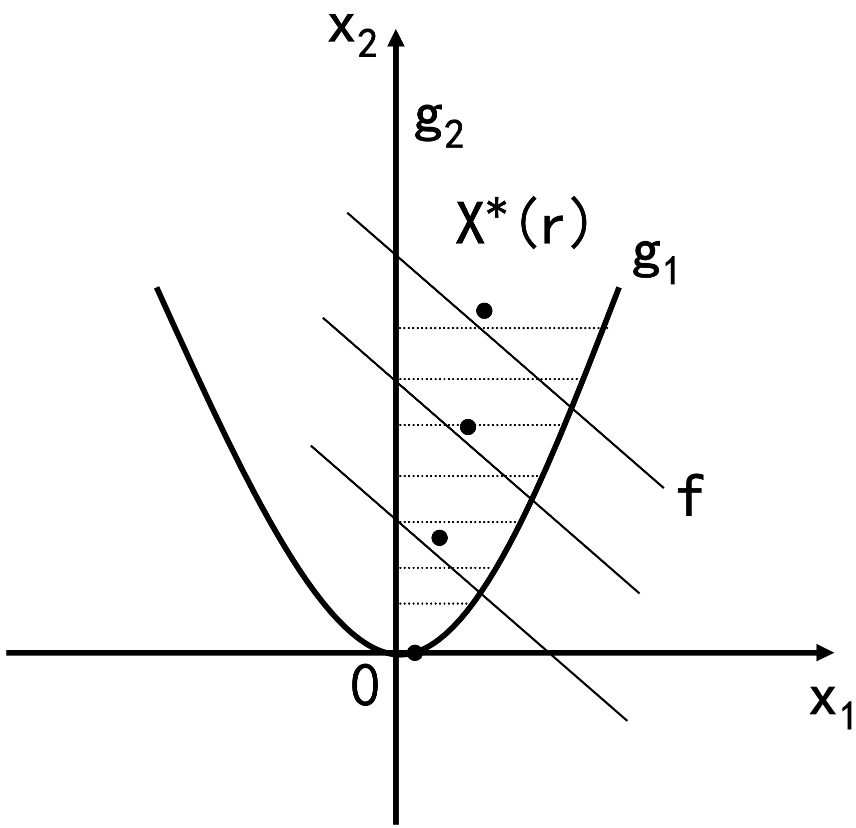
求解得到：

$$x_1 = \frac{\sqrt{1+8r}-1}{4}$$

$$x_2 = \frac{3}{2}r - \frac{\sqrt{1+8r}-1}{8}$$

$r \to 0$时，有

$$\begin{cases} x_1^* = 0 \\ x_2^* = 0 \end{cases}$$



数值解法

- 构造障碍函数

$$\min_{X\in R_0}\bar{P}(\boldsymbol{x},r_k)=f(\boldsymbol{x})-r_k\sum_{j=1}^{l}\log g_j(\boldsymbol{x})$$

1、 取$r_1>0$（通常$r_1=1$），允许误差ε>0，$k:=1$

2、 求min$P(x,r_k)$，得$x^{*(k)}\in \mathrm{R}_0$（须保证是内点）

3、 若存在 $\left|r_k\sum_{j=1}^{l}\log g_j(\boldsymbol{x}^{*(k)})\right|>\varepsilon$

取$r_{k+1}=r_k/c$，（c>1,通常取5或10），$k:=k+1$转第2步重新求解。否则，停止迭代，$x^*=x^{*(k)}$。

定理：若ε>0，上述算法必在有限步内终止。

说明： 对于复杂问题，初始内点可用算法获取。

内点法聚点分析

- 障碍函数

$$\bar{P}(\boldsymbol{x}, r_k) = f(\boldsymbol{x}) - r_k \sum_{j=1}^{l} \log g_j(\boldsymbol{x})$$

- 第k步障碍函数局部极小值满足

$$\nabla \bar{P}(\boldsymbol{x}_k^*, r_k) = \nabla f(\boldsymbol{x}_k^*) - r_k \sum_{j=1}^{l} \frac{1}{g_j(\boldsymbol{x}_k^*)} \nabla g_j(\boldsymbol{x}_k^*) = 0$$

$$\nabla \bar{P}(\boldsymbol{x}_k^*, r_k) = \nabla f(\boldsymbol{x}_k^*) - \sum_{j=1}^{l} \mu_j^*(k) \nabla g_j(\boldsymbol{x}_k^*) = 0$$

$$\mu_j^*(k) = \frac{r_k}{g_j(\boldsymbol{x}_k^*)} \geq 0$$

- $\boldsymbol{x}_k$*迭代收敛时

$$\nabla \bar{P}(\boldsymbol{x}^*, r) = \nabla f(\boldsymbol{x}^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(\boldsymbol{x}^*) = 0$$

$$\mu_j^* = \frac{r}{g_j(\boldsymbol{x}^*)} \geq 0$$

$$\mu_j^* g_j(\boldsymbol{x}^*) = r = 0$$

说明：　对于复杂问题，初始内点可用算法获取。

混合法

- 内点法不能处理等式约束问题
- 外点法不能处理目标函数在可行域外不存在的问题
- 对等式约束和当前不被满足的不等式约束，使用罚函数法，对满足的不等式约束，使用障碍函数法。

约束极值问题的数值解

- 可行方向法
  - **Zoutendijk**可行方向法
- 制约函数法
  - 外点法
  - 内点法
  - 混合法
- 逐次逼近法（近似规划法）
  - **SLP (Sequential Linear Programming)**
  - **SQP (Sequential Quadratic Programming)**

逐次逼近法

- 思想：**Taylor**展开近似为简单规划问题
- 序贯线性规划法**SLP (Sequential Linear Programming)**
- 序贯二次规划法**SQP (Sequential Quadratic Programming)**

一般约束极值问题

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

s.t. $$h_i(\boldsymbol{x}) = 0 \qquad i = 1, 2, \ldots, m$$

$$g_j(\boldsymbol{x}) \geq 0 \qquad j = 1, 2, \ldots, l$$

$$\boldsymbol{x} \in R^n$$

求解思路：通过低阶近似，化为容易求解的规划问题

序贯线性规划法SLP

$$\min f(\boldsymbol{x}) = f(\boldsymbol{x}^{(k)}) + \nabla f(\boldsymbol{x}^{(k)})^T (\boldsymbol{x} - \boldsymbol{x}^{(k)})$$

s.t. $h_i(\boldsymbol{x}^{(k)}) + \nabla h_i(\boldsymbol{x}^{(k)})^T (\boldsymbol{x} - \boldsymbol{x}^{(k)}) = 0 \qquad i = 1, 2, \ldots, m$

$g_j(\boldsymbol{x}^{(k)}) + \nabla g_j(\boldsymbol{x}^{(k)})^T (\boldsymbol{x} - \boldsymbol{x}^{(k)}) \geq 0 \qquad j = 1, 2, \ldots, l$

$|x_s - x_s^{(k)}| \leq \delta_s^{(k)} \qquad s = 1, 2, \ldots, n \qquad$ $\delta_s^{(k)}$是步长限制量

设第k步线性规划的最优解为$\boldsymbol{x}^{*(k)}$

当$\boldsymbol{x}^{*(k)}$是可行解时，取 $\delta_s^{(k+1)} = \delta_s^{(k)}$ ，$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{*(k)}$ 继续迭代；

当$\boldsymbol{x}^{*(k)}$不是可行解时，取 $\delta_s^{(k)} = \beta \delta_s^{(k)}$ ，$\beta < 1$ ，重新寻优；

当$|\delta_s^{(k)}| < \varepsilon, \forall s = 1, 2, \cdots, n$ 时，或 $\|\nabla f(\boldsymbol{x}^{(k)})\| < \varepsilon$ 迭代结束。

序贯二次规划法SQP

起作用约束集法序贯求解如下近似问题

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

s.t. $$h_i(\boldsymbol{x}^{(k)}) + \nabla h_i(\boldsymbol{x}^{(k)})^T \Delta \boldsymbol{x} = 0 \qquad i = 1, 2, \ldots, m$$

$$g_j(\boldsymbol{x}^{(k)}) + \nabla g_j(\boldsymbol{x}^{(k)})^T \Delta \boldsymbol{x} \geq 0 \qquad j = 1, 2, \ldots, l$$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \Delta \boldsymbol{x}^*$$