第四章 非线性规划

修贤超

https://xianchaoxiu.github.io

目录

非线性规划问题的数学模型

- 一般形式

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

  - $\mathbf{X} = (x_1, \cdots, x_n)^\top \in \mathbb{R}^n$ 是 n 维欧氏空间中的点 (向量)
  - 目标函数 $f(\mathbf{X})$ 和约束函数 $h_i(\mathbf{X})$、$g_j(\mathbf{X})$ 为 $\mathbf{X}$ 的实函数

- 有时也将非线性规划的数学模型写成

$$
\begin{aligned}
&\min\ f(\mathbf{X}) \\
&\text{s.t.}\ g_j(\mathbf{X}) \geq 0\ (j = 1, \cdots, l)
\end{aligned}
$$

局部极小和全局极小

- 给定 $\mathbf{X}^* \in \mathbb{R}^n$，如果存在某个 $\varepsilon > 0$，使所有与 $\mathbf{X}^*$ 的距离小于 ε 的 $\mathbf{X}$ 都有 $f(\mathbf{X}) \geq f(\mathbf{X}^*)$，则称 $\mathbf{X}^*$ 为局部极小点，$f(\mathbf{X}^*)$ 为局部极小值
- 若对所有 $\mathbf{X} \neq \mathbf{X}^*$ 且与 $\mathbf{X}^*$ 的距离小于 ε 的 $\mathbf{X}$ 都有 $f(\mathbf{X}) > f(\mathbf{X}^*)$，则称 $\mathbf{X}^*$ 为严格局部极小点，$f(\mathbf{X}^*)$ 为严格局部极小值
- 给定 $\mathbf{X}^* \in \mathbb{R}^n$，如果对所有 $\mathbf{X}$ 都有 $f(\mathbf{X}) \geq f(\mathbf{X}^*)$，则称 $\mathbf{X}^*$ 为全局极小点，$f(\mathbf{X}^*)$ 为全局极小值
- 若对所有 $\mathbf{X} \neq \mathbf{X}^*$ 都有 $f(\mathbf{X}) > f(\mathbf{X}^*)$，则称 $\mathbf{X}^*$ 为严格全局极小点，$f(\mathbf{X}^*)$ 为严格全局极小值
- 如将上述不等号取反向，即可得到极大点和极大值的定义

必要条件

- 定理 1 设 $f(\mathbf{X})$ 在 $\mathbb{R}^n$ 上有连续一阶偏导数，且在点 $\mathbf{X}^* \in \mathbb{R}^n$ 取得局部极值，则必有函数 $f(\mathbf{X})$ 在点 $\mathbf{X}^*$ 处的梯度

$$\nabla f(\mathbf{X}^*) = \left(\frac{\partial f(\mathbf{X}^*)}{\partial x_1}, \frac{\partial f(\mathbf{X}^*)}{\partial x_2}, \cdots, \frac{\partial f(\mathbf{X}^*)}{\partial x_n} \right)^\top = 0$$

  称 $\mathbf{X}^*$ 为稳定点或驻点

- 函数 $f(\mathbf{X})$ 在某点 $\mathbf{X}$ 的梯度 $\nabla f(\mathbf{X})$ 必与函数过该点的等值面正交
- 梯度向量的方向是函数值增加最快的方向，而负梯度方向是减少最快的方向

二次型

- 二次型是 $\mathbf{X}=(x_1,\cdots,x_n)^\top\in\mathbb{R}^n$ 的二次齐次函数

$$\begin{aligned} f(\mathbf{X}) &= a_{11}x_1^2+2a_{12}x_1x_2+\cdots+a_{nn}x_n^2 \\ &= \sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j \\ &= \mathbf{X}^\top\mathbf{A}\mathbf{X} \end{aligned}$$

- 若 $\mathbf{A}$ 的所有元素都是实数，则称为实二次型
- 一个二次型唯一对应一个对称矩阵 $\mathbf{A}$, 反之亦成立

二次型

- 正定: $f(\mathbf{X}) = \mathbf{X}^\top \mathbf{A} \mathbf{X} > 0$
- 负定: $f(\mathbf{X}) = \mathbf{X}^\top \mathbf{A} \mathbf{X} < 0$
- 半正定: $f(\mathbf{X}) = \mathbf{X}^\top \mathbf{A} \mathbf{X} \geq 0$
- 半负定: $f(\mathbf{X}) = \mathbf{X}^\top \mathbf{A} \mathbf{X} \leq 0$

- 实二次型 $\mathbf{X}^\top \mathbf{A} \mathbf{X}$ 为正定的充要条件是 $\mathbf{A}$ 的左上角各阶主子式都大于零
- 实二次型 $\mathbf{X}^\top \mathbf{A} \mathbf{X}$ 为负定的充要条件是 $\mathbf{A}$ 的左上角顺序各阶主子式负正相间

多元函数的泰勒 (Taylor) 公式

- **设 $f(\mathbf{X})$ 在 $\mathbf{X}^0$ 的某一邻域内有连续二阶偏导数，则在 $\mathbf{X}^0$ 的泰勒展开式为**

$$f(\mathbf{X}) = f(\mathbf{X}^0) + \nabla f(\mathbf{X}^0)^\top(\mathbf{X} - \mathbf{X}^0) + \frac{1}{2}(\mathbf{X} - \mathbf{X}^0)^\top \nabla^2 f(\bar{\mathbf{X}})(\mathbf{X} - \mathbf{X}^0)$$

$$\Updownarrow$$

$$f(\mathbf{X}^0 + \mathbf{P}) = f(\mathbf{X}^0) + \nabla f(\mathbf{X}^0)^\top \mathbf{P} + \frac{1}{2}\mathbf{P}^\top \nabla^2 f(\bar{\mathbf{X}})\mathbf{P}$$

$$\Updownarrow$$

$$\begin{aligned} f(\mathbf{X}) =& f(\mathbf{X}^0) + \nabla f(\mathbf{X}^0)^\top(\mathbf{X} - \mathbf{X}^0) + \frac{1}{2}(\mathbf{X} - \mathbf{X}^0)^\top \nabla^2 f(\mathbf{X}^0)(\mathbf{X} - \mathbf{X}^0) \\ &+ o(\|\mathbf{X} - \mathbf{X}^0\|^2) \end{aligned}$$

充分条件

- 定理 2 设 $f(\mathbf{X})$ 在 $\mathbb{R}^n$ 具有连续二阶偏导数，若 $\nabla f(\mathbf{X}^*) = 0$ 且黑塞 (Hesse) 矩阵为

$$\nabla^2 f(\mathbf{X}^*) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_2^2} & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_n \partial x_2} & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_n^2} \end{bmatrix}$$

  正定，则 $\mathbf{X}^*$ 为 $f(\mathbf{X})$ 严格局部极小点

- 若将 $\nabla^2 f(\mathbf{X}^*)$ 正定改为负定，则 $\mathbf{X}^*$ 为 $f(\mathbf{X})$ 的严格局部极大点

- 研究函数 $f(\mathbf{X}) = x_1^2 - x_2^2$ 是否存在极值点

分析 由极值点存在的必要条件求出稳定点

$$\frac{\partial f(\mathbf{X})}{\partial x_1} = 2x_1, \quad \frac{\partial f(\mathbf{X})}{\partial x_2} = -2x_2$$

令 $f(\mathbf{X}) = 0$，即 $2x_1 = 0, -2x_2 = 0$, 得稳定点 $\mathbf{X} = (x_1, x_2)^\top = (0, 0)^\top$

再用充分条件进行检验

$$\nabla^2 f(\mathbf{X}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{X})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{X})}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(\mathbf{X})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{X})}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

由于黑塞矩阵 $\nabla^2 f(\mathbf{X})$ 不定，故 $\mathbf{X} = (0, 0)^\top$ 不是极值点，而是鞍点

凸函数和凹函数

- 设 $f(\mathbf{X})$ 为定义在凸集 Ω 上的函数，若对任何实数 α $(0 < \alpha < 1)$ 以及 Ω 中的任意两点 $\mathbf{X}^{(1)}$ 和 $\mathbf{X}^{(2)}$，恒有

$$f(\alpha\mathbf{X}^{(1)} + (1-\alpha)\mathbf{X}^{(2)}) \leq \alpha f(\mathbf{X}^{(1)}) + (1-\alpha)\mathbf{X}^{(2)}$$

  则称 $f(\mathbf{X})$ 为定义在 Ω 上的凸函数

- 若对每一个 α $(0 < \alpha < 1)$ 和任意两点 $\mathbf{X}^{(1)} \neq \mathbf{X}^{(2)} \in \Omega$，恒有

$$f(\alpha\mathbf{X}^{(1)} + (1-\alpha)\mathbf{X}^{(2)}) < \alpha f(\mathbf{X}^{(1)}) + (1-\alpha)\mathbf{X}^{(2)}$$

  则称 $f(\mathbf{X})$ 为定义在 Ω 的严格凸函数

- 反之可得到凹函数和严格凹函数的定义

凸函数的性质

- **性质 1** 设 $f(\mathbf{X})$ 为定义在凸集 Ω 上的函数, 则对任意实数 $\beta \geq 0$，函数 $\beta f(\mathbf{X})$ 也是定义在 Ω 上的凸函数
- **性质 2** 设 $f_1(\mathbf{X})$ 和 $f_2(\mathbf{X})$ 为定义在凸集 Ω 上的函数, 则这两个凸函数的和 $f(\mathbf{X}) = f_1(\mathbf{X}) + f_2(\mathbf{X})$ 仍为定义在 Ω 上的凸函数
- **性质 3** 有限个凸函数的非负线性组合 $\beta_1 f_1(\mathbf{X}) + \beta_2 f_2(\mathbf{X}) + \cdots + \beta_m f_m(\mathbf{X})$ 仍为凸函数
- **性质 4** 设 $f(\mathbf{X})$ 为定义在凸集 Ω 上的函数, 则对实数 β，集合 (称为水平集) $S_\beta = \{\mathbf{X} \mid \mathbf{X} \in \Omega, f(\mathbf{X}) \leq \beta\}$ 是凸集

凸函数的判定

- **一阶条件** 设 $f(\mathbf{X})$ 为定义在凸集 Ω 上的可微函数，则 $f(\mathbf{X})$ 是凸函数的充要条件是: 对任意两点 $\mathbf{X}^{(1)}, \mathbf{X}^{(2)} \in \Omega$，恒有

$$f(\mathbf{X}^{(2)}) \geq f(\mathbf{X}^{(1)}) + \nabla f(\mathbf{X}^{(1)})^\top (\mathbf{X}^{(2)} - \mathbf{X}^{(1)})$$

- **二阶条件** 设 $f(\mathbf{X})$ 为定义在凸集 Ω 上的二阶可微函数，则 $f(\mathbf{X})$ 是凸函数的充要条件是: 对所有 $\mathbf{X} \in \Omega$ 有

$$\nabla^2 f(\mathbf{X}) \geq 0$$

进一步，若 $\nabla^2 f(\mathbf{X}) > 0$, 则 $f(\mathbf{X})$ 是 Ω 上的严格凸函数

例 2

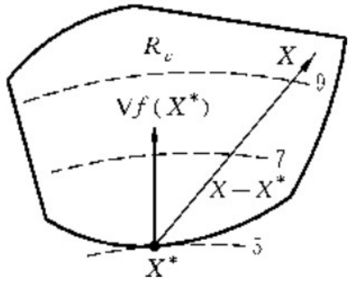
- **证明函数 $f(\mathbf{X}) = x_1^2 + x_2^2$ 为严格凸函数**

  分析　由二阶条件得到

$$\nabla^2 f(\mathbf{X}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{X})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{X})}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(\mathbf{X})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{X})}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

  因 $\nabla^2 f(\mathbf{X})$ 正定，故 $f(\mathbf{X})$ 为严格凸函数

- 设 $f(\mathbf{X})$ 为定义在凸集 Ω 上的可微凸函数，如果存在点 $\mathbf{X}^* \in \Omega$，使得对于所有的 $\mathbf{X} \in \Omega$，都有
$$\nabla f(\mathbf{X}^*)^\top (\mathbf{X} - \mathbf{X}^*) \geq 0$$
则 $\mathbf{X}^*$ 是 $f(\mathbf{X})$ 在 Ω 上的最小点 (全局极小点)



凸规划

- 考虑非线性规划

$$
\begin{aligned}
&\min\ f(\mathbf{X}) \\
&\text{s.t. } g_j(\mathbf{X}) \geq 0\ (j = 1, \cdots, l)
\end{aligned}
$$

若其中的 $f(\mathbf{X})$ 为凸函数，$g_j(\mathbf{X})$ 全是凹函数，则称为**凸规划**

- 可行解集为凸集
- 最优解集为凸集 (假定最优解存在)
- 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，则最优解必唯一

例 3

- **验证下式为凸规划**

$$\min\ f(\mathbf{X}) = x_1^2 + x_2^2 - 4x_1 + 4$$

$$\text{s.t. } \begin{cases} g_1(\mathbf{X}) = x_1 - x_2 + 2 \geq 0 \\ g_2(\mathbf{X}) = -x_1^2 + x_2 - 1 \geq 0 \\ g_3(\mathbf{X}) = x_1 \geq 0 \\ g_4(\mathbf{X}) = x_2 \geq 0 \end{cases}$$

下降迭代算法

- 从初始点 $\mathbf{X}^{(0)}$ 出发，按照一定的规则，先找一个比 $\mathbf{X}^{(0)}$ 更好的点 $\mathbf{X}^{(1)}$，再找比 $\mathbf{X}^{(1)}$ 更好的点 $\mathbf{X}^{(2)}$ ... 如此继续就产生了一个序列 $\{\mathbf{X}^{(k)}\}$。若序列有一极限点 $\mathbf{X}^*$，即
$$\lim_{k\to\infty}\|\mathbf{X}^{(k)}-\mathbf{X}^*\|=0$$
则称序列 $\{\mathbf{X}^{(k)}\}$ 收敛于 $\mathbf{X}^*$

- 对于极小化问题，序列 $\{\mathbf{X}^{(k)}\}$ 对应的目标函数值 $\{f(\mathbf{X}^{(k)})\}$ 应是逐步减小的，即
$$f(\mathbf{X}^{(0)})>f(\mathbf{X}^{(1)})>\cdots>f(\mathbf{X}^{(k)})>\cdots$$
具有这种性质的算法称为下降迭代算法

下降迭代算法的一般迭代格式

- 选取初始点: $\mathbf{X}^{(0)}$，令 $k := 0$
- 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。从 $\mathbf{X}^{(k)}$ 出发确定一搜索方向 $\mathbf{P}^{(k)}$，沿这个方向能找到使目标函数值下降的点
- 确定步长: 沿 $\mathbf{P}^{(k)}$ 方向前进得新点 $\mathbf{X}^{(k+1)}$，即在由 $\mathbf{X}^{(k)}$ 出发的射线

  $$\mathbf{X} = \mathbf{X}^{(k)} + \lambda \mathbf{P}^{(k)}, \quad \lambda \geq 0$$

  上，通过选定步长 $\lambda = \lambda_k$，得下一个迭代点

  $$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \lambda_k \mathbf{P}^{(k)}$$

  使得

  $$f(\mathbf{X}^{(k+1)}) = f(\mathbf{X}^{(k)} + \lambda_k \mathbf{P}^{(k)}) < f(\mathbf{X}^{(k)})$$

- 检验是否最优: 如满足则停止，否则令 $k := k + 1$，返回第 (2) 步继续迭代

下降迭代算法的一般迭代格式

- 步长应使目标函数值沿搜索方向下降最多，即

$$f(\mathbf{X}^{(k)} + \lambda_k \mathbf{P}^{(k)}) = \min_{\lambda} f(\mathbf{X}^{(k)} + \lambda \mathbf{P}^{(k)})$$

  这一过程称为(最优) 一维搜索或线搜索，由此确定的步长称为最佳步长

- 定理 3 设目标函数 $f(\mathbf{X})$ 具有连续一阶偏导数，$\mathbf{X}^{(k+1)}$ 按下述规则产生

$$\begin{cases} f(\mathbf{X}^{(k)} + \lambda_k \mathbf{P}^{(k)}) = \min_{\lambda} f(\mathbf{X}^{(k)} + \lambda \mathbf{P}^{(k)}) \\ \mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \lambda_k \mathbf{P}^{(k)} \end{cases}$$

  则有

$$\nabla f(\mathbf{X}^{(k+1)})^{\top} \mathbf{P}^{(k)} = 0$$

终止迭代准则

- 根据相继两次迭代结果的绝对误差

$$\|\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)}\| \leq \varepsilon_1$$

$$|f(\mathbf{X}^{(k+1)}) - f(\mathbf{X}^{(k)})| \leq \varepsilon_2$$

- 根据相继两次迭代结果的相对误差

$$\frac{\|\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)}\|}{\|\mathbf{X}^{(k)}\|} \leq \varepsilon_3$$

$$\frac{|f(\mathbf{X}^{(k+1)}) - f(\mathbf{X}^{(k)})|}{|f(\mathbf{X}^{(k)})|} \leq \varepsilon_4$$

- 根据函数梯度的模足够小

$$\|\nabla f(\mathbf{X}^{(k)})\| \leq \varepsilon_5$$

小结

- 非线性规划的数学模型
- 局部极小和全局极小
- 多元函数的极值点存在的条件
- 凸函数和凹函数
- 凸规划
- 下降迭代算法
- 课后作业: P183, 习题 6.10

目录

无约束极值问题

- 无约束极值问题可表述为

$$\min \quad f(\mathbf{X})$$

- 不要求函数的解析性质，仅利用函数值，称为直接法
- 利用函数的解析性质，如一阶导数或二阶导数，称为解析法
  - 梯度法（最速下降法）
  - 牛顿法

梯度法

- 给定初始点 $\mathbf{X}^{(0)}$，令 $k := 0$
- 计算 $f(\mathbf{X}^{(k)})$ 和 $\nabla f(\mathbf{X}^{(k)})$, 若 $\|\nabla f(\mathbf{X}^{(k)}\|^2 \leq \varepsilon$，停止迭代，得近似极小点 $\mathbf{X}^{(k)}$ 和近似极小值 $f(\mathbf{X}^{(k)})$; 否则，转下一步
- 做一维搜索

$$\lambda_k : \min_{\lambda} f(\mathbf{X}^{(k)} - \lambda \nabla f(\mathbf{X}^{(k)}))$$

并计算 $\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \lambda_k \nabla f(\mathbf{X}^{(k)})$, 然后令 $k := k + 1$，转回第 (2) 步

梯度法

- 设 $f(\mathbf{X})$ 具有二阶连续偏导数，将 $f(\mathbf{X}^{(k)} - \lambda\nabla f(\mathbf{X}^{(k)})$ 在 $\mathbf{X}^k$ 作泰勒展开

$$f(\mathbf{X}^{(k)} - \lambda\nabla f(\mathbf{X}^{(k)})) \approx f(\mathbf{X}^{(k)}) - \nabla f(\mathbf{X}^{(k)})^\top \lambda\nabla f(\mathbf{X}^{(k)}) + \frac{1}{2}\lambda\nabla f(\mathbf{X}^{(k)})^\top \nabla^2 f(\mathbf{X}^{(k)})\lambda\nabla f(\mathbf{X}^{(k)})$$

- 对 λ 求导并令其等于零，即可得近似最佳步长的计算公式

$$\lambda_k = \frac{\nabla f(\mathbf{X}^{(k)})^\top \nabla f(\mathbf{X}^{(k)})}{\nabla f(\mathbf{X}^{(k)})^\top \nabla^2 f(\mathbf{X}^{(k)})\nabla f(\mathbf{X}^{(k)})}$$

例 1

- 用梯度法求函数 $f(\mathbf{X}) = x_1^2 + 5x_2^2$ 的极小点, 取允许误差 $\varepsilon = 0.7$

  分析 取 $\mathbf{X}^{(0)} = (2,1)^\top$，计算

$$\nabla f(\mathbf{X}^{(0)}) = (4,10)^\top, \nabla^2 f(\mathbf{X}) = \begin{bmatrix} 2 & 0 \\ 0 & 10 \end{bmatrix}$$

$$\lambda_0 = \frac{\begin{bmatrix} 4 & 10 \end{bmatrix} \begin{bmatrix} 4 \\ 10 \end{bmatrix}}{\begin{bmatrix} 4 & 10 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} 4 \\ 10 \end{bmatrix}} = 0.1124$$

例 1

于是

$$\mathbf{X}^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - 0.1124 \begin{bmatrix} 4 \\ 10 \end{bmatrix} = \begin{bmatrix} 1.5504 \\ -0.1240 \end{bmatrix}$$

此时

$$\nabla f(\mathbf{X}^{(1)}) = \begin{bmatrix} 3.1008 \\ -1.2400 \end{bmatrix}$$

检查误差

$$\|\nabla f(\mathbf{X}^{(1)})\|^2 = 11.1526 > \varepsilon$$

继续迭代直至满足误差 $\varepsilon = 0.7$

牛顿法

- 考虑正定二次函数

$$f(\mathbf{X}) = \frac{1}{2}\mathbf{X}^\top \mathbf{A}\mathbf{X} + \mathbf{B}^\top \mathbf{X} + c$$

其中 $\mathbf{A}$ 为对称正定阵，$\mathbf{B}$ 和 $\mathbf{X}$ 为向量，c 为常数

- 假定极小点是 $\mathbf{X}^*$，则必有

$$\nabla f(\mathbf{X}^*) = \mathbf{A}\mathbf{X}^* + \mathbf{B} = 0$$

对任一点 $\mathbf{X}^{(0)}$，函数在该点的梯度

$$\nabla f(\mathbf{X}^{(0)}) = \mathbf{A}\mathbf{X}^{(0)} + \mathbf{B}$$

消去 $\mathbf{B}$，得到 $\nabla f(\mathbf{X}^{(0)}) = \mathbf{A}\mathbf{X}^{(0)} - \mathbf{A}\mathbf{X}^*$, 由此解出

$$\mathbf{X}^* = \mathbf{X}^{(0)} - \mathbf{A}^{-1}\nabla f(\mathbf{X}^{(0)})$$

例 2

- 用牛顿法求函数 $f(\mathbf{X}) = x_1^2 + 5x_2^2$ 的极小点, 取允许误差 $\varepsilon = 0.7$

  分析　取 $\mathbf{X}^{(0)} = (2, 1)^\top$ 算出

$$\nabla f(\mathbf{X}^{(0)}) = (4, 10)^\top$$

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 10 \end{bmatrix}, \mathbf{A}^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/10 \end{bmatrix}$$

$$\begin{aligned} \mathbf{X}^* &= \mathbf{X}^{(0)} - \mathbf{A}^{-1} \nabla f(\mathbf{X}^{(0)}) \\ &= \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/2 & 0 \\ 0 & 1/10 \end{bmatrix} \begin{bmatrix} 4 \\ 10 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

  可知 $\mathbf{X}^*$ 确实为极小点

牛顿法

- 考虑一般 n 元实函数 $f(\mathbf{X})$，具有连续二阶偏导数，在 $\mathbf{X}^{(k)}$ 附近取二阶泰勒多项式逼近

$$f(\mathbf{X}) \approx f(\mathbf{X}^{(k)}) + \nabla f(\mathbf{X}^{(k)})^\top \Delta\mathbf{X} + \frac{1}{2}\Delta\mathbf{X}^\top \nabla^2 f(\mathbf{X}^{(k)})\Delta\mathbf{X}$$

其中 $\Delta\mathbf{X} = \mathbf{X} - \mathbf{X}^{(k)}$

- 近似函数的极小点应满足一阶必要条件，即

$$\nabla f(\mathbf{X}^{(k)}) + \nabla^2 f(\mathbf{X}^{(k)})\Delta\mathbf{X} = 0$$

- 设 $\nabla^2 f(\mathbf{X}^{(k)})$ 的逆阵存在，可得

$$\mathbf{X} = \mathbf{X}^{(k)} - [\nabla^2 f(\mathbf{X}^{(k)})]^{-1}\nabla f(\mathbf{X}^{(k)})$$

牛顿法

- 为求 $f(\mathbf{X})$ 的极小点，可令

$$-[\nabla^2 f(\mathbf{X}^{(k)})]^{-1}\nabla f(\mathbf{X}^{(k)})$$

为搜索方向 (牛顿方向)，按下述公式进行迭代

$$\begin{cases} \mathbf{P}^{(k)} = -[\nabla^2 f(\mathbf{X}^{(k)})]^{-1}\nabla f(\mathbf{X}^{(k)}) \\ \lambda_k : \ \min_\lambda f(\mathbf{X}^{(k)} + \lambda \mathbf{P}^{(k)}) \\ \mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \lambda_k \mathbf{P}^{(k)} \end{cases}$$

这就是阻尼牛顿法 (广义牛顿法)，可用于求解非正定二次函数的极小点

小结

- 梯度法
  - 下降方向
  - 搜索步长
- 牛顿法
  - 正定二次函数
  - 一般多元实函数
- 课后作业: P183, 习题 6.13

目录

约束极值问题

- 考虑约束极值问题

$$\min\ f(\mathbf{X})$$
$$\text{s.t. } g_j(\mathbf{X}) \geq 0\ (j = 1, \cdots, l)$$

- 将约束极值问题转化为无约束极值问题来求解，称为**序列无约束极小化技术**
  - 罚函数法 (外点法)
  - 障碍函数法 (内点法)

罚函数法

- 构造函数

$$\varphi(t)=\begin{cases}0, & \text{当 } t\geq 0\\ \infty, & \text{当 } t<0\end{cases}$$

- 将约束问题转化为无约束形式

$$\min\ f(\mathbf{X}) \quad \text{s.t. } g_j(\mathbf{X})\geq 0\ (j=1,\cdots,l) \tag{1}$$

$$\Downarrow$$

$$\min\ f(\mathbf{X})+\sum_{j=1}^{l}\varphi(g_j(\mathbf{X})) \tag{2}$$

- 若 $\mathbf{X}^*$ 是 (2) 的极小点，同时也是 (1) 的极小点

罚函数法

- 但函数 $\varphi(t)$ 在 $t=0$ 处不连续，更没有导数，于是将该函数做如下修改

$$\varphi(t)=\begin{cases}0, & 当\ t \geq 0 \\ t^2, & 当\ t<0\end{cases}$$

- 通过罚函数法转化为无约束问题

$$\min\ P(\mathbf{X}, M)=f(\mathbf{X})+M \sum_{j=1}^{l} \varphi(g_j(\mathbf{X})) \tag{3}$$

$$\Updownarrow$$

$$\min\ P(\mathbf{X}, M)=f(\mathbf{X})+M \sum_{j=1}^{l}[\min(0, g_j(\mathbf{X}))]^2 \tag{4}$$

- 当罚因子 M 足够大时，(3) 和 (4) 的极小点也是 (1) 的极小点

罚函数法

- 对于等式约束问题

$$\min\ f(\mathbf{X})$$
$$\text{s.t. } h_i(\mathbf{X}) = 0\ (i = 1, \cdots, m)$$

采用以下形式的罚函数

$$\min\ P(\mathbf{X}, M) = f(\mathbf{X}) + M\sum_{j=1}^{m}[h_j(\mathbf{X})]^2$$

- 对于一般非线性规划问题，罚函数为

$$\min\ P(\mathbf{X}, M) = f(\mathbf{X}) + M\sum_{j=1}^{m}[h_j(\mathbf{X})]^2 + M\sum_{j=1}^{l}[\min(0, g_j(\mathbf{X}))]^2$$

罚函数法的迭代步骤

- 取第一个罚因子 $M_1 > 0$，允许误差 $\varepsilon > 0$，并令 $k := 1$
- 求下述无约束极值问题的最优解

$$\mathbf{X}^{(k)} = \arg\min\ P(\mathbf{X}, M)$$

- 若存在某一个 j $(1 \leq j \leq l)$，有

$$-g_j(\mathbf{X}^{(k)}) > \varepsilon$$

或存在某一个 i $(1 \leq i \leq m)$，有

$$|h_i(\mathbf{X}^{(k)})| > \varepsilon$$

则取 $M_{k+1} > M_k$，并令 $k := k + 1$，转回第 (2) 步。否则，停止迭代

例 1

- **用罚函数法求解**

$$\min\ f(x) = (x-1/2)^2$$
$$\text{s.t. } x \le 0$$

分析 **构造罚函数**

$$\min\ P(x,M) = f(x) + M\sum_{j=1}^{l}[\min(0,g(x))]^2$$
$$= (x-1/2)^2 + M[\min(0,-x)]^2$$

对于固定的 M**，令**

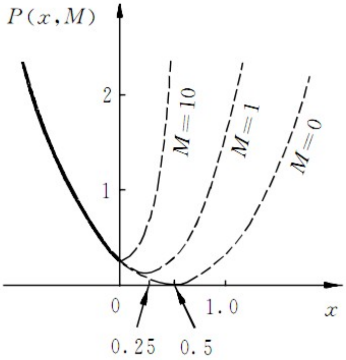
$$\frac{dP(x,M)}{dx} = 2(x-1/2) - 2M[\min(0,-x)] = 0$$
$$\Downarrow$$
$$2(x-1/2) - 2Mx = 0$$

例 1

从而求得其极小点

$$x(M) = \frac{1}{2(1+M)}$$

- 当 $M = 0$ 时，$x(M) = 1/2$
- 当 $M = 1$ 时，$x(M) = 1/4$
- 当 $M = 10$ 时，$x(M) = 1/22$
- 当 $M \to \infty$ 时，$x(M) \to 0$



障碍函数法

- 考虑非线性规划

$$\begin{aligned} &\min\ f(\mathbf{X}) \\ &\text{s.t. } g_j(\mathbf{X}) \geq 0\ (j=1,\cdots,l) \end{aligned}$$

- 取实数 $r_k > 0$，通过障碍函数法转化为无约束问题

$$\min\ \bar{P}(\mathbf{X}, r_k) = f(\mathbf{X}) + r_k \sum_{j=1}^{l} \frac{1}{g_j(\mathbf{X})}$$

$$\Updownarrow$$

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

- 上式要求 $g_j(\mathbf{X}) > 0$

障碍函数法的迭代步骤

- 取第一个障碍因子 $r_1 > 0$，允许误差 $\varepsilon > 0$，并令 $k := 1$
- 构造障碍函数，障碍项可采用倒数函数，也可采用对数函数
- 求下述无约束极值问题的最优解 (满足内点要求)

$$\mathbf{X}^{(k)} = \arg\min \ \bar{P}(\mathbf{X}, r_k)$$

- 检查是否满足收敛准则

$$r_k \sum_{j=1}^{l} \frac{1}{g_j(\mathbf{X})} \leq \varepsilon, \quad \left| r_k \sum_{j=1}^{l} log(g_j(\mathbf{X})) \right| \leq \varepsilon$$

如果满足，则以 $\mathbf{X}^{(k)}$ 为原约束问题的近似极小解，停止迭代; 否则取 $r_{k+1} < r_k$，并令 $k := k + 1$，转回第 (3) 步

例 2

- **用障碍函数法求解**

$$\min\ f(x) = x - 2$$
$$\text{s.t. } x \geq 0$$

分析 **构造障碍函数**

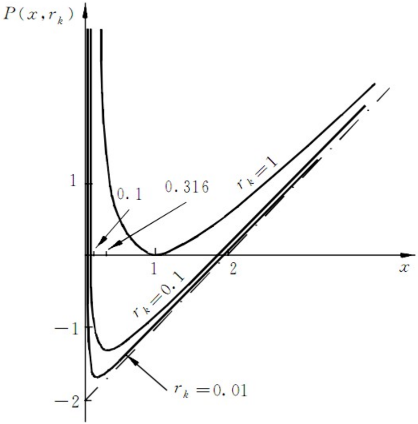
$$\begin{aligned}\bar{P}(\mathbf{X}, r_k) &= f(\mathbf{X}) + r_k \sum_{j=1}^{l} \frac{1}{g_j(\mathbf{X})} \\ &= x - 2 + \frac{r_k}{x}\end{aligned}$$

对于固定的 r_k**，令**

$$\frac{d\bar{P}(\mathbf{X}, r_k)}{dx} = 1 - \frac{r_k}{x^2} = 0 \quad \Rightarrow \quad \mathbf{X} = \pm\sqrt{r_k}$$

例 2

令 $r_k \to 0$, 并考虑到约束条件，即可得该问题的极小点 $\mathbf{X}^* = 0$



小结

- 罚函数法
  - 罚函数
  - 罚因子
- 障碍函数法
  - 障碍函数
  - 障碍因子
- 混合法
- 课后作业: P184, 习题 6.19, 6.22(2)

Q&A

Thank you!

感谢您的聆听和反馈