智能优化方法及其应用

第二讲 经典优化算法2

课程纲要

- 最优化理论数学基础复习
- 线搜索方法
- 梯度法和牛顿法
- 共轭梯度法
- 拟牛顿法
- 最小二乘法
- 最优性条件(*)
- 线性规划
- 二次规划

课程纲要

- 最优化理论数学基础复习
- 线搜索方法
- 梯度法和牛顿法
- 共轭梯度法
- 拟牛顿法
- 最小二乘法
- 最优性条件(*)
- 线性规划
- 二次规划

共轭梯度法

- **梯度法基本思想与缺陷**: 搜索方向为负梯度方向(也即最速下降方向), 若用精确线搜索方法, 新点梯度与旧点梯度正交, 迭代点列所走路线是锯齿型, 收敛速度比较缓慢.
- **牛顿法基本思想与缺陷**: 用迭代点处的一阶导数和二阶导数对目标函数进行二次函数近似, 然后把二次函数的极小点作为新的迭代点, 重复该过程, 直至求得满足解. 需要求解目标函数Hesse矩阵, 且其需在每个迭代点处正定. 牛顿方向 $\boldsymbol{d}_k = -\boldsymbol{G}_k^{-1}\boldsymbol{g}_k$
- **共轭梯度法**: 是介于梯度法与牛顿法之间的一种无约束优化算法, 具有Q-超线性收敛速度. 只用到目标函数及其梯度值, 避免计算二阶导数(Hesse阵), 降低了计算量和存储量.

线性共轭方向法

- **线性共轭方向法基本思想**: 在求解n维正定二次目标函数极小点时产生一组共轭方向作为搜索方向, 在精确线搜索下算法至多迭代n步即能求得极小点(**二次终止性**). 经过修正后, 线性共轭方向法可推广求解一般非二次目标函数.

- **定义4.1** 设$\boldsymbol{A} \in \boldsymbol{R}^{n\times n}$是**对称正定**矩阵, 若$n$维向量组$\boldsymbol{d}_1, \boldsymbol{d}_2 \cdots, \boldsymbol{d}_m (m \le n)$满足

$$\boldsymbol{d}_i^T \boldsymbol{A} \boldsymbol{d}_j = 0, \qquad i \neq j,$$

则称$\boldsymbol{d}_1, \boldsymbol{d}_2 \cdots, \boldsymbol{d}_m$是$\boldsymbol{A}$**共轭**的.

- 显然, 向量组的共轭是正交的推广, 即当$\boldsymbol{A} = \boldsymbol{I}$(单位阵)时, 共轭变为正交.

线性共轭方向法

- $\boldsymbol{A}$的共轭向量组的性质.
- **性质4.1** 设$\boldsymbol{A} \in \boldsymbol{R}^{n \times n}$是对称正定矩阵, 若非零向量组$\boldsymbol{d}_1, \boldsymbol{d}_2 \cdots, \boldsymbol{d}_m$是$\boldsymbol{A}$共轭的, 则它们是线性无关的.

- 求解正定二次目标函数极小点的线性共轭方向法.

$$\min_{\boldsymbol{x} \in \boldsymbol{R}^n} f(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x}$$

$\boldsymbol{A}$为n阶对称正定阵; $\boldsymbol{b}$为n维常向量.

线性共轭方向法

算法 4.1 (线性共轭方向法)

步骤 0, 给定迭代精度 $0 \leqslant \varepsilon \ll 1$ 和初始点 $\boldsymbol{x}_0 \in \mathbf{R}^n$. 计算 $\boxed{\boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0) = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}.}$ 选取初始方向 $\boldsymbol{d}_0$ 使得 $\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{g}_0 < 0$. 令 $k := 0$.

步骤 1, 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$, 停算, 输出 $\boldsymbol{x}^* \approx \boldsymbol{x}_k$.

步骤 2, 利用精确线搜索方法确定步长因子 α_k.

步骤 3, 令 $\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$, 并计算 $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1}) = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}$.

步骤 4, 选取 $\boldsymbol{d}_{k+1}$ 满足下降性和共轭性条件:

$$\boxed{\boldsymbol{d}_{k+1}^{\mathrm{T}}\boldsymbol{g}_{k+1} < 0,} \quad \boxed{\boldsymbol{d}_{k+1}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i = 0,} \quad i = 0, 1, \cdots, k.$$

步骤 5, 令 $k := k + 1$, 转步骤 1.

线性共轭方向法

- 线性共轭方向法的收敛性

$$\min_{x\in R^n} f(x) = \frac{1}{2}x^T A x - b^T x$$

定理 4.1 设目标函数 f 由式 (4.1) 定义. $\{\boldsymbol{x}_k\}$ 是算法 4.1 产生的迭代序列, 则每一步迭代点 $\boldsymbol{x}_{k+1}$ 都是 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}_0$ 和方向 $\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k$ 所张成的线性流形

$$S_k = \left\{ \boldsymbol{x} \,\middle|\, \boldsymbol{x} = \boldsymbol{x}_0 + \sum_{i=0}^{k} \alpha_i \boldsymbol{d}_i,\ \forall \alpha_i \right\}$$

中的极小点. 特别地, $\boldsymbol{x}_n = \boldsymbol{x}^* = -\boldsymbol{A}^{-1}\boldsymbol{b}$ 是问题 (4.1) 的唯一极小点.

对任意 $\boldsymbol{x} \in S_k$,

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{x}_{k+1}) + \boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{h}_{k+1} + \frac{1}{2}\boldsymbol{h}_{k+1}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{h}_{k+1} \\ &\geqslant f(\boldsymbol{x}_{k+1}) + \boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{h}_{k+1} \\ &= f(\boldsymbol{x}_{k+1}) + \sum_{i=0}^{k} (\beta_i - \alpha_i) \boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{d}_i. \end{aligned}$$

$$\boldsymbol{h}_{k+1} = \boldsymbol{x} - \boldsymbol{x}_{k+1} = \sum_{i=0}^{k} (\beta_i - \alpha_i) \boldsymbol{d}_i$$

$$\boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{d}_i = 0, \quad i = 0, 1, \cdots, k$$

线性共轭梯度法

- **线性共轭梯度法****基本原理**: 线性共轭方向法中取

  $\boldsymbol{d}_0 = -\boldsymbol{g}_0$

  $$\boxed{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_i = 0, \quad i = 0, 1, \cdots, k}$$

- 此处, 以严格凸二次函数$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$为例, 来说明线性共轭梯度法中共轭搜索方向$\boldsymbol{d}_k$的构建.

共轭性

修正负梯度搜索方向

共轭条件

$\boldsymbol{d}_0 = -\boldsymbol{g}_0$, $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0\boldsymbol{d}_0$,

α_0 为最优步长.

$\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{d}_0 = 0.$

令

$$\boxed{\boldsymbol{d}_1 = -\boldsymbol{g}_1 + \beta_0\boldsymbol{d}_0,}$$

其中 β_0 的选取须满足

$$\boxed{\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_1 = 0.}$$

$\boldsymbol{d}_1 = -\boldsymbol{g}_1 + \beta_0\boldsymbol{d}_0$ 两边同乘以 $\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}$, 得

$$\boxed{0 = \boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_1} = -\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_1 + \beta_0\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_0.$$

$$\beta_0 = \frac{\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_1}{\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_0} = \frac{\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_0}{\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_0}$$

$$= \frac{\boldsymbol{g}_1^{\mathrm{T}}(\boldsymbol{g}_1 - \boldsymbol{g}_0)}{\boldsymbol{d}_0^{\mathrm{T}}(\boldsymbol{g}_1 - \boldsymbol{g}_0)}$$

$$\boxed{= \frac{\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{g}_1}{\boldsymbol{g}_0^{\mathrm{T}}\boldsymbol{g}_0}.}$$

推广:$\beta_{k-1} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$

线性共轭梯度法

$$\min_{x\in R^n} f(x) = \frac{1}{2}x^T Ax - b^T x$$

- 求解严格凸二次函数极小点的线性共轭梯度法:

算法 4.1 (线性共轭梯度法)

步骤 0, 给定迭代精度 $0 \leqslant \varepsilon \ll 1$ 和初始点 $\boldsymbol{x}_0 \in \mathbf{R}^n$. 计算

$$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b},\ \boldsymbol{d}_0 = -\boldsymbol{g}_0. \text{ 令 } k := 0.$$

步骤 1, 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$, 停算, 输出 $\boldsymbol{x}^* \approx \boldsymbol{x}_k$.

经$m \le n$步迭代后终止(二次终止性)

步骤 2, 计算

$$\boxed{\alpha_k = -\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_k},}$$

令$\dfrac{\partial f(x_{k+1})}{\partial \alpha_k} = 0$

$$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k,$$

$$\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b},$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_k}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_k} = \boxed{\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}},$$

$$\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k.$$

步骤 3, 令 $k := k+1$, 转步骤 1.

线性共轭梯度法

- 程序详解:

求解

$$\min_{x\in R^n} f(x)=\frac{1}{2}x^TAx-b^Tx$$

$$A=\begin{pmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{pmatrix}\in \mathbf{R}^{n\times n} \qquad b=\begin{pmatrix} 3 \\ 2 \\ \vdots \\ 2 \\ 3 \end{pmatrix}\in \mathbf{R}^n.$$

取 $n=100$, 初始向量为零向量

终止准则 $\|\nabla f(x_k)\|\leqslant 10^{-5}$

迭代 10 次满足终止条件

得近似极小值为 $f(x^*)\approx -101$

```
function [k,x,val]=linecg(A,b,x0,epsilon,N)
if nargin<5, N=1000; end
if nargin<4, epsilon=1.e-5; end
if nargin<3, x0=zeros(length(b),1); end
k=0;
gk=A*x0-b;
dk=-gk;
while(k<N)
    temp=A*dk;
    alpha=-gk'*dk/(dk'*temp);
    x=x0+alpha*dk;
    gk=A*x-b;
    betak=gk'*temp/(dk'*temp);
    dk=-gk+betak*dk;
    if(norm(gk)<epsilon), break; end
    x0=x;
    k=k+1;
end
val=0.5*x'*A*x-b'*x;
```

非线性共轭梯度法

- **非线性共轭梯度法基本原理**: 将线性共轭梯度法用于求解无约束非线性极小化问题 $\min_{\boldsymbol{x}\in \boldsymbol{R}^n} f(\boldsymbol{x})$, 在每一迭代步利用当前点处的最速下降方向与算法的前一个方向的线性组合作为当前步的搜索方向, 且取 $\boldsymbol{d}_0 = -\boldsymbol{g}_0$.
- 搜索方向 $\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k$, 参数 β_k 有多种形式.

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k} \quad \text{(FR 公式)}$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)}{\boldsymbol{d}_k^{\mathrm{T}}(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)} \quad \text{(Crowder-Wolfe 公式)}$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{g}_{k+1}}{-\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}_k} \quad \text{(Dixon 公式)}$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)}{\boldsymbol{d}_k^{\mathrm{T}}(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)} \quad \text{(HS 公式)}$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{g}_{k+1}}{\boldsymbol{d}_k^{\mathrm{T}}(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)} \quad \text{(Dai-Yuan 公式)}$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k)}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k} \quad \text{(PRP 公式)}$$

非线性共轭梯度法

算法 4.2 (非线性共轭梯度法)

步骤 0, 给定迭代精度 $0 \leqslant \varepsilon \ll 1$ 和初始点 $\boldsymbol{x}_0 \in \mathbf{R}^n$. 计算 $\boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0)$, $\boldsymbol{d}_0 = -\boldsymbol{g}_0$. 令 $k := 0$.

步骤 1, 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$, 停算, 输出 $\boldsymbol{x}^* \approx \boldsymbol{x}_k$.

步骤 2, 利用某种线搜索方法确定搜索步长 α_k, 计算

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \qquad \boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1}).$$

步骤 3, 更新搜索方向 $\boldsymbol{d}_{k+1} := -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k$, 其中 β_k 由上页公式中的某一公式 (如 FR 公式) 确定.

$$\beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$$

步骤 4, 令 $k := k+1$, 转步骤 1.

非线性共轭梯度法

- **非线性共轭梯度法收敛性**(了解)

定理 4.3 设 $\{\boldsymbol{x}_k\}$ 是由算法 4.2 产生的序列, 假定函数 $f(\boldsymbol{x})$ 一阶连续可微且水平集 $\mathcal{L}(\boldsymbol{x}_0)=\{\boldsymbol{x}\,|\,f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_0)\}$ 是有界的, 那么算法 4.2 或者有限步终止, 或者 $\lim\limits_{k\to\infty}\boldsymbol{g}(\boldsymbol{x}_k)=\boldsymbol{0}$.

定理 4.4 设 $\{\boldsymbol{x}_k\}$ 是由算法 4.2 利用 Wolfe 准则 (2.11) 和 (2.12) 产生的序列, 假定函数 $f(\boldsymbol{x})$ 一阶连续可微且有下界, 其梯度函数 $\boldsymbol{g}(\boldsymbol{x})$ 在 $\mathbf{R}^n$ 上 Lipschitz 连续, 即存在常数 $L>0$, 使得

$$\|\boldsymbol{g}(\boldsymbol{u})-\boldsymbol{g}(\boldsymbol{v})\|\leqslant L\|\boldsymbol{u}-\boldsymbol{v}\|,\quad \forall\,\boldsymbol{u},\boldsymbol{v}\in\mathbf{R}^n.$$

若选取的搜索方向 $\boldsymbol{d}_k$ 与 $-\boldsymbol{g}_k$ 的夹角 θ_k 满足条件

$$0\leqslant\theta_k\leqslant\frac{\pi}{2}-\mu,\quad \mu\in\left(0,\frac{\pi}{2}\right).$$

那么算法 4.2 或者有限步终止, 或者 $\lim\limits_{k\to\infty}g(\boldsymbol{x}_k)=\boldsymbol{0}$.

拟牛顿法

- **牛顿法的优点**: 不低于二阶的收敛速度
- **牛顿法的缺点**: 要求目标函数的Hesse矩阵$\boldsymbol{G}(\boldsymbol{x})$在每个迭代点$\boldsymbol{x}_k$处是**正定**的, 否则难以保证牛顿方向$\boldsymbol{d}_k = -\boldsymbol{G}_k^{-1}\boldsymbol{g}_k$是$f$在$\boldsymbol{x}_k$处的下降方向. 若$\boldsymbol{G}_k$奇异, 算法无法继续进行.
- **修正牛顿法**: 引入阻尼因子$\mu_k \geq 0$, 即在每一迭代步适当选取参数μ_k使得矩阵$\boldsymbol{A}_k = \boldsymbol{G}(\boldsymbol{x}_k) + \mu_k \boldsymbol{I}$正定. 但修正系数$\mu_k$很难选取, 过大过小都会影响收敛速度. 牛顿法的每一迭代步都需要计算Hesse矩阵, 对于大规模问题计算量惊人!
- **拟牛顿法的优点**: 不需要计算目标函数的Hesse矩阵, 却在某种意义下具有使用Hesse阵时的功效. (1950s, Davidon提出)

拟牛顿法

- **拟牛顿法的基本思想:** 在基本牛顿法的步骤2中用Hesse矩阵的某个近似矩阵取代. 应具有下面三特点:
  - ✓ 在某种意义下$\boldsymbol{B}_k \approx \boldsymbol{G}_k$, 使得算法产生的搜索方向近似于牛顿方向, 保证较快收敛速度.
  - ✓ 对所有k, $\boldsymbol{B}_k$是对称正定的, 使得算法产生的方向是函数f在$\boldsymbol{x}_k$处的下降方向.
  - ✓ 矩阵$\boldsymbol{B}_k$更新规则相对简单, 通常采用一个秩1或秩2矩阵进行校正.
    - **秩:** 矩阵的列秩和行秩总是相等, 都可称作矩阵A的秩, 矩阵A的列秩是A的线性独立的纵列的极大数目.计算矩阵A的秩的最容易的方式是高斯消去法, 其他方法包括: 基本高斯消去(LU分解), 奇异值分解法(SVD), QR分解法

拟牛顿法

位移 $\boldsymbol{s}_k=\boldsymbol{x}_{k+1}-\boldsymbol{x}_k$

梯度差 $\boldsymbol{y}_k=\boldsymbol{g}_{k+1}-\boldsymbol{g}_k$

- 拟牛顿法的对称秩1算法

算法 5.1 (对称秩 1 算法)

步骤 0，选取初始点 $\boldsymbol{x}_0\in\mathbf{R}^n$，终止误差 $0\leqslant\varepsilon\ll 1$，初始对称正定阵 $\boldsymbol{H}_0$(通常取单位阵 $\boldsymbol{I}_n$)。令 $k:=0$。

步骤 1，若 $\|\boldsymbol{g}_k\|\leqslant\varepsilon$，停算，输出 $\boldsymbol{x}_k$ 作为近似极小点。

步骤 2，计算搜索方向 $\boldsymbol{d}_k=-\boldsymbol{H}_k\boldsymbol{g}_k$.

步骤 3，用线搜索技术求步长因子 α_k.

步骤 4，令 $\boldsymbol{x}_{k+1}:=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k$，由对称秩 1 校正公式 (5.6) 确定 $\boldsymbol{H}_{k+1}$.

$$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k+\frac{(\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k)(\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}}{(\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}\boldsymbol{y}_k}.$$

步骤 5，令 $k:=k+1$，转步骤 1.

拟牛顿法

$$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k+\frac{(\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k)(\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}}{(\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}\boldsymbol{y}_k}.$$

位移 $\boldsymbol{s}_k=\boldsymbol{x}_{k+1}-\boldsymbol{x}_k$

梯度差 $\boldsymbol{y}_k=\boldsymbol{g}_{k+1}-\boldsymbol{g}_k$

- 拟牛顿法的对称秩1算法的推导

f 在 $\boldsymbol{x}_{k+1}$ 处的二次近似模型为

$$f(\boldsymbol{x})\approx f(\boldsymbol{x}_{k+1})+\boldsymbol{g}_{k+1}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}_{k+1})+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_{k+1})^{\mathrm{T}}\boldsymbol{G}_{k+1}(\boldsymbol{x}-\boldsymbol{x}_{k+1}).$$

求导数, 得

$$\boldsymbol{g}(\boldsymbol{x})\approx\boldsymbol{g}_{k+1}+\boldsymbol{G}_{k+1}(\boldsymbol{x}-\boldsymbol{x}_{k+1}).$$

$$\boxed{\boldsymbol{G}_{k+1}\boldsymbol{s}_k\approx\boldsymbol{y}_k.}$$

构造出 Hesse 阵的近似矩阵 $\boldsymbol{B}_k$ 满足这种关系式, 即

$$\boxed{\boldsymbol{B}_{k+1}\boldsymbol{s}_k=\boldsymbol{y}_k.}$$

通常称为拟牛顿方程或拟牛顿条件. 令 $\boldsymbol{H}_{k+1}=\boldsymbol{B}_{k+1}^{-1}$, 则得到

拟牛顿方程的另一个形式:

$$\boxed{\boldsymbol{H}_{k+1}\boldsymbol{y}_k=\boldsymbol{s}_k,}$$

拟牛顿法

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{(\boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k)(\boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}}{(\boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}\boldsymbol{y}_k}.$$

位移 $\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$

梯度差 $\boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$

拟牛顿法的对称秩1算法的推导

搜索方向由 $\boldsymbol{d}_k = -\boldsymbol{H}_k\boldsymbol{g}_k$ 或 $\boldsymbol{B}_k\boldsymbol{d}_k = -\boldsymbol{g}_k$ 确定. 可令

$$\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \boldsymbol{E}_k, \quad \boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \boldsymbol{D}_k,$$

式中: $\boldsymbol{E}_k$, $\boldsymbol{D}_k$ 为秩 1 或秩 2 矩阵.

下面介绍一个对称秩 1 校正公式

取 $\boldsymbol{E}_k = \alpha\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}$

$$\boxed{\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k.}$$

$$(\boldsymbol{B}_k + \alpha\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}})\boldsymbol{s}_k = \boldsymbol{y}_k,$$

$$\alpha(\boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{s}_k)\boldsymbol{u}_k = \boldsymbol{y}_k - \boldsymbol{B}_k\boldsymbol{s}_k.$$

表明向量 $\boldsymbol{u}_k$ 平行于向量 $\boldsymbol{y}_k - \boldsymbol{B}_k\boldsymbol{s}_k$, 即存在常数 β 使得

$\boldsymbol{u}_k = \beta(\boldsymbol{y}_k - \boldsymbol{B}_k\boldsymbol{s}_k)$. 因此有

$$\boldsymbol{E}_k = \alpha\beta^2(\boldsymbol{y}_k - \boldsymbol{B}_k\boldsymbol{s}_k)(\boldsymbol{y}_k - \boldsymbol{B}_k\boldsymbol{s}_k)^{\mathrm{T}}.$$

拟牛顿法

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^{\mathrm{T}}}{(s_k - H_k y_k)^{\mathrm{T}} y_k}.$$

位移 $s_k = x_{k+1} - x_k$

梯度差 $y_k = g_{k+1} - g_k$

- 拟牛顿法的对称秩1算法的推导

$$E_k = \alpha\beta^2 (y_k - B_k s_k)(y_k - B_k s_k)^{\mathrm{T}}.$$

$$\boxed{\alpha(u_k^{\mathrm{T}} s_k) u_k = y_k - B_k s_k.}$$

$$\boxed{E_k = \alpha u_k u_k^{\mathrm{T}}}$$

$$\alpha\beta^2 [(y_k - B_k s_k)^{\mathrm{T}} s_k](y_k - B_k s_k) = y_k - B_k s_k.$$

若 $(y_k - B_k s_k)^{\mathrm{T}} s_k \neq 0$, 可取 $\alpha\beta^2 [(y_k - B_k s_k)^{\mathrm{T}} s_k] = 1$, 即

$$\alpha\beta^2 = \frac{1}{(y_k - B_k s_k)^{\mathrm{T}} s_k}, \qquad E_k = \frac{(y_k - B_k s_k)(y_k - B_k s_k)^{\mathrm{T}}}{(y_k - B_k s_k)^{\mathrm{T}} s_k}.$$

故得对称秩 1 校正公式如下:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^{\mathrm{T}}}{(y_k - B_k s_k)^{\mathrm{T}} s_k}.$$

类似地,

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^{\mathrm{T}}}{(s_k - H_k y_k)^{\mathrm{T}} y_k}.$$

拟牛顿法——BFGS算法

- **BFGS校正**是目前最流行也是最有效的拟牛顿校正，是Broyden, Fletcher, Goldfarb和Shanno在1970年各自独立提出的拟牛顿法，故称BFGS算法.
- **BFGS算法基本思想:** 令$\boldsymbol{B}_{k+1}=\boldsymbol{B}_k+\boldsymbol{E}_k$中的修正矩阵$\boldsymbol{E}_k$为秩2矩阵

$$\boldsymbol{E}_k=\alpha\boldsymbol{u}_k\boldsymbol{u}_k^{\mathrm{T}}+\beta\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$$

式中: $\boldsymbol{u}_k,\boldsymbol{v}_k\in\mathbf{R}^n$ 为待定向量; $\alpha,\beta\in\mathbf{R}$ 为待定实数.

$$\boldsymbol{B}_{k+1}=\begin{cases}\boldsymbol{B}_k, & \boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k\leqslant 0,\\ \boldsymbol{B}_k-\dfrac{\boldsymbol{B}_k\boldsymbol{s}_k\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{s}_k}+\dfrac{\boldsymbol{y}_k\boldsymbol{y}_k^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k}, & \boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k>0.\end{cases}$$

拟牛顿法——BFGS算法

算法 5.2 (BFGS 算法)

步骤 0，选取参数 $\beta \in (0,1)$，$\sigma \in (0,0.5)$，初始点 $\boldsymbol{x}_0 \in \mathbf{R}^n$，终止误差 $0 \leqslant \varepsilon \ll 1$，初始对称正定阵 $\boldsymbol{B}_0$ (通常取为 $\boldsymbol{G}(\boldsymbol{x}_0)$ 或单位阵 $\boldsymbol{I}_n$)．令 $k := 0$.

步骤 1，计算 $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$．若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$，停算，输出 $\boldsymbol{x}_k$ 作为近似极小点.

步骤 2，解线性方程组

$$\boxed{\boldsymbol{B}_k \boldsymbol{d} = -\boldsymbol{g}_k,}$$

得解 $\boldsymbol{d}_k$.

步骤 3，设 m_k 是满足下列不等式的最小非负整数 m：

$$\boxed{f(\boldsymbol{x}_k + \beta^m \boldsymbol{d}_k) \leqslant f(\boldsymbol{x}_k) + \sigma \beta^m \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k}$$

Armijo非精确线搜索

令 $\alpha_k := \beta^{m_k}$，$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$.

步骤 4，由 $\boxed{\text{校正公式 (5.11)}}$ 确定 $\boldsymbol{B}_{k+1}$.

$$\boldsymbol{B}_{k+1} = \begin{cases} \boldsymbol{B}_k, & \boldsymbol{y}_k^{\mathrm{T}} \boldsymbol{s}_k \leqslant 0, \\ \boldsymbol{B}_k - \dfrac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{T}} \boldsymbol{B}_k}{\boldsymbol{s}_k^{\mathrm{T}} \boldsymbol{B}_k \boldsymbol{s}_k} + \dfrac{\boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}} \boldsymbol{s}_k}, & \boldsymbol{y}_k^{\mathrm{T}} \boldsymbol{s}_k > 0. \end{cases}$$

步骤 5，令 $k := k+1$，转步骤 1.

拟牛顿法——BFGS算法

- BFGS拟牛顿法的推导(略)

$$B_{k+1} = \begin{cases} B_k, & y_k^{\mathrm{T}} s_k \leqslant 0, \\ B_k - \dfrac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} + \dfrac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}, & y_k^{\mathrm{T}} s_k > 0. \end{cases}$$

位移 $s_k = x_{k+1} - x_k$

梯度差 $y_k = g_{k+1} - g_k$

最小二乘法

- **最小二乘问题:** 是一类特殊的无约束优化问题, 在某种意义下可看作一个超定方程组(方程的个数远多于未知变量的个数)问题.
- **基本思想:** 由于超定方程组一般是无解的, 此时期望得到的是其残量的最小范数解. 前人在牛顿法基础上提出了多种有效的算法.
- 最小二乘问题来源于数据拟合问题, 它是一种基于观测数据与模型数据之间的**残差的平方和最小**来**估计数学模型中的参数**的方法.

最小二乘法

- **最小二乘问题的数学描述**

设某系统中输入数据t与输出数据y服从函数关系

$$y = f(\boldsymbol{x}, t)$$

为估计待定参数向量$\boldsymbol{x} \in \boldsymbol{R}^n$的值, 需经过多次试验取得观测数据$(t_1, y_1), (t_2, y_2), \cdots, (t_m, y_m)$ (通常$m \gg n$)

然后, 基于模型输出值和实际观测值的误差平方和

$$S(\boldsymbol{x}) = \sum_{i=1}^{m} [y_i - f(\boldsymbol{x}, t_i)]^2$$

最小来求参数$\boldsymbol{x}$的值, 上述便是最小二乘问题.

最小二乘法

- **最小二乘问题的数学描述**

  若引入残差函数$r_i(\boldsymbol{x}) = y_i - f(\boldsymbol{x}, t_i)$, $i = 1,2,\cdots,m$, 记
  $$\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), r_2(\boldsymbol{x}), \cdots, r_m(\boldsymbol{x}))^T$$
  最小二乘问题可表示为
  $$\min_{\boldsymbol{x}\in \boldsymbol{R}^n} \|\boldsymbol{r}(\boldsymbol{x})\|_2^2$$
  通常, 将上述模型等价写成
  $$\min_{\boldsymbol{x}\in \boldsymbol{R}^n} \frac{1}{2}\|\boldsymbol{r}(\boldsymbol{x})\|_2^2$$

- 如果残差函数$\boldsymbol{r}(\boldsymbol{x})$关于$\boldsymbol{x}$是线性的, 则称之为**线性最小二乘**问题, 也称为**线性回归**; 否则, 称之为**非线性最小二乘**问题, 也称**非线性回归**.

线性最小二乘问题

- **定义 7.1** 设$A \in R^{m \times n}, b \in R^m$, 线性最小二乘问题为

$$\min_{x \in R^n} \frac{1}{2} \|b - Ax\|_2^2$$

线性最小二乘问题 等价于: 确定 $x_{\mathrm{LS}} \in \mathbf{R}^n$ 使得

$$\|b - Ax_{\mathrm{LS}}\|_2 = \min_{x \in \mathbf{R}^n} \|b - Ax\|_2.$$

称 x_{LS} 为最小二乘解或极小解，解 x_{LS} 又可称为线性方程组

$$\boxed{Ax = b, \quad A \in \mathbf{R}^{m \times n}, \ b \in \mathbf{R}^m}$$

的最小二乘解, 即 x_{LS} 在残量 $r(x) = b - Ax$ 的 2- 范数最小的意义下满足方程组. 当 $\boxed{m > n \text{ 时, 称为超定方程组}}$

$\boxed{m < n \text{ 时, 称为欠定方程组}}$

线性最小二乘问题

- **定理 7.1** $\boldsymbol{x}_{LS}$是最小二乘问题$\min\limits_{\boldsymbol{x}\in R^n}\frac{1}{2}\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_2^2$的极小解的充分必要条件是$\boldsymbol{x}_{LS}$为方程组

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^T\boldsymbol{b}$$

的解,该方程也称为最小二乘问题的法方程.

证明

$$\min f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x}-(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x}+\frac{1}{2}\|\boldsymbol{b}\|_2^2.$$

由于 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\in\mathbf{R}^{n\times n}$ 是半正定矩阵, 因此 n 元实函数 $f(\boldsymbol{x})$ 是凸函数, 故 $\boldsymbol{x}_{\mathrm{LS}}$ 是 极小解等价于

$$\nabla f(\boldsymbol{x}_{\mathrm{LS}})=\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}_{\mathrm{LS}}-\boldsymbol{b})=\mathbf{0}.$$

满秩线性最小二乘问题

- 线性最小二乘问题方程组中的$\boldsymbol{A} \in \boldsymbol{R}^{m \times n}(m \geq n)$为列**满秩矩阵**, 称为满秩线性最小二乘问题, 此时$\boldsymbol{A}^T\boldsymbol{A}$为对称正定矩阵, 故法方程$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$**存在唯一解**. 法方程可用Cholesky方法求解.

算法 7.1 (**法方程 Cholesky 分解法**) $\boldsymbol{A} \in \mathbf{R}^{m \times n}$ $(m \geqslant n)$ 为列满秩矩阵, $\boldsymbol{b} \in \mathbf{R}^m$.

步骤 0, 对 n 阶对称正定矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 作 Cholesky 分解 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boxed{\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}\text{, 其中 } \boldsymbol{L} \text{ 为下三角矩阵.}}$

步骤 1, 依次解 $\boxed{\boldsymbol{L}\boldsymbol{y} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b},\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}}$ 得到最小二乘问题 $\min\limits_{\boldsymbol{x} \in \mathbf{R}^n} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ 的解 $\boldsymbol{x}_{\mathrm{LS}}$.

满秩线性最小二乘问题

算法 7.1 (法方程 Cholesky 分解法) $\boldsymbol{A} \in \mathbf{R}^{m \times n}\,(m \geqslant n)$ 为列满秩矩阵, $\boldsymbol{b} \in \mathbf{R}^m$.

步骤 0, 对 n 阶对称正定矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 作 Cholesky 分解 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$, 其中 $\boldsymbol{L}$ 为下三角矩阵.

步骤 1, 依次解 $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$, $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ 得到最小二乘问题 $\min\limits_{\boldsymbol{x}\in\mathbf{R}^n} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ 的解 $\boldsymbol{x}_{\mathrm{LS}}$.

```
function [x,res]=nels(A,b)
B=A'*A;  f=A'*b;
L=chol(B,'lower');
y=L\f;   x=L'\y;
res=norm(b-A*x);
```

Cholesky 分解

求解线性方程组

满秩线性最小二乘问题

- 求解满秩线性最小二乘问题更常用方法是**QR分解法**.

算法 7.2 (QR **分解法**) $\boldsymbol{A} \in \mathbf{R}^{m\times n}\,(m \geqslant n)$ 为列满秩矩阵, $\boldsymbol{b} \in \mathbf{R}^m$.

步骤 0, 计算系数矩阵 $\boldsymbol{A}$ 的 QR 分解 $[Q, R] = \mathrm{qr}(A)$.

步骤 1, 用回代法求解上三角形方程组

$$\begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{pmatrix} \boldsymbol{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b},$$

得最小二乘解 $\boldsymbol{x}_{\mathrm{LS}}$.

求解上三角形方程组

```
function [x,res]=qrls(A,b)
[Q,R]=qr(A);   f=Q'*b;
x=R\f;   res=norm(b-A*x);
```

满秩线性最小二乘问题

- **QR分解法**原理

$\boxed{\boldsymbol{Q}\text{ 为正交矩阵}}$

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \Longleftrightarrow \boldsymbol{Q}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{O}\end{pmatrix}\boldsymbol{x}=\boldsymbol{b} \Longleftrightarrow \begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{O}\end{pmatrix}\boldsymbol{x}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}=(c_1,c_2,\cdots,c_n,c_{n+1},\cdots,c_m)^{\mathrm{T}}$，则 $\boldsymbol{R}\boldsymbol{x}=(c_1,\cdots,c_n)^{\mathrm{T}}$ 有唯一解 $\boldsymbol{x}_{\mathrm{LS}}$.

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 &= \left\|\boldsymbol{Q}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{O}\end{pmatrix}\boldsymbol{x}-\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}\right\|_2^2=\left\|\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{O}\end{pmatrix}\boldsymbol{x}-\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}\right\|_2^2\\ &=\left\|\boldsymbol{R}\boldsymbol{x}-\begin{pmatrix}c_1\\ \vdots\\ c_n\end{pmatrix}\right\|_2^2+\left\|-\begin{pmatrix}c_{n+1}\\ \vdots\\ c_m\end{pmatrix}\right\|_2^2 \quad \boxed{\geqslant c_{n+1}^2+\cdots+c_m^2}\end{aligned}$$

$$\|\boldsymbol{A}\boldsymbol{x}_{\mathrm{LS}}-\boldsymbol{b}\|_2^2=\left\|\boldsymbol{R}\boldsymbol{x}_{\mathrm{LS}}-\begin{pmatrix}c_1\\ \vdots\\ c_n\end{pmatrix}\right\|_2^2+\left\|-\begin{pmatrix}c_{n+1}\\ \vdots\\ c_m\end{pmatrix}\right\|_2^2=\boxed{c_{n+1}^2+\cdots+c_m^2}$$

亏秩线性最小二乘问题

- 广义逆的定义

定义 7.2 设 $\boldsymbol{A} \in \mathrm{R}^{m\times n}$, 若有 $\boldsymbol{X} \in \mathrm{R}^{n\times m}$ 满足

(1) $\boldsymbol{AXA} = \boldsymbol{A}$;

(2) $\boldsymbol{XAX} = \boldsymbol{X}$;

(3) $(\boldsymbol{AX})^{\mathrm{T}} = \boldsymbol{AX}$;

(4) $(\boldsymbol{XA})^{\mathrm{T}} = \boldsymbol{XA}$.

则称 $\boldsymbol{X}$ 为矩阵 $\boldsymbol{A}$ 的广义逆, 记为 $\boldsymbol{A}^{\dagger}$.

亏秩线性最小二乘问题

- 广义逆的计算

定理 7.2 设秩为 $r(r \geqslant 1)$ 的实 $m \times n$ 矩阵 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U} \begin{pmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{V}^{\mathrm{T}},$$

式中: $\boldsymbol{U} \in \mathbf{R}^{m\times m}$ 和 $\boldsymbol{V} \in \mathbf{R}^{n\times n}$ 均为正交阵; $\boldsymbol{\Sigma}_r = \mathrm{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_r\}$, $\sigma_i > 0\,(i = 1, 2, \cdots, r)$ 为矩阵 $\boldsymbol{A}$ 的正奇异值. 则

$$\boldsymbol{A}^{\dagger} = \boldsymbol{V} \begin{pmatrix} \boldsymbol{\Sigma}_r^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{U}^{\mathrm{T}}.$$

亏秩线性最小二乘问题

- 亏秩线性最小二乘问题求解

定理 7.3 如果线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解，则它的极小范数解 $\boldsymbol{x}_{\mathrm{LS}}$ 唯一，并且 $\boldsymbol{x}_{\mathrm{LS}}=\boldsymbol{A}^{\dagger}\boldsymbol{b}$.

定理 7.4 如果线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 无解，则它的极小范数最小二乘解 $\boldsymbol{x}_{\mathrm{LS}}$ 唯一，并且 $\boldsymbol{x}_{\mathrm{LS}}=\boldsymbol{A}^{\dagger}\boldsymbol{b}$.

$$\boldsymbol{x}_{\mathrm{LS}}=\boldsymbol{A}^{\dagger}\boldsymbol{b}=\boldsymbol{V}\begin{pmatrix}\boldsymbol{\Sigma}_r^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{b}=\sum_{i=1}^{r}\frac{\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{b}}{\sigma_i}\boldsymbol{v}_i.$$

$$\boldsymbol{\Sigma}_r=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r),\ \sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0.$$

亏秩线性最小二乘问题

- 亏秩线性最小二乘问题实例

求解超定方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$, 其中

$$\boldsymbol{A}=\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 5 & 6 \\ 1 & 5 & 6 & 7 \\ 1 & 8 & 9 & 10 \\ 1 & 11 & 12 & 13 \end{pmatrix},$$

$$\boldsymbol{x}=\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \quad \boldsymbol{b}=\begin{pmatrix} 11 \\ 13 \\ 15 \\ 18 \\ 20 \end{pmatrix}.$$

$$\boldsymbol{x}_{\mathrm{LS}}=\boldsymbol{A}^{\dagger}\boldsymbol{b}=\sum_{i=1}^{r}\frac{\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{b}}{\sigma_i}\boldsymbol{v}_i.$$

```
A=[1 2 3 4; 1 4 5 6; 1 5 6 7;
   1 8 9 10; 1 11 12 13];
b=[11 13 15 18 20]';
[m,n]=size(A); x=zeros(n,1);
[U,S,V]=svd(A);
r=rank(S);
for i=1:r
    x=x+(U(:,i)'*b/S(i,i))*V(:,i);
end
res=norm(b-A*x)
```

非线性最小二乘问题

- 非线性最小二乘问题是求向量$\boldsymbol{x} \in \boldsymbol{R}^n$, 使得$\|\boldsymbol{F}(\boldsymbol{x})\|^2$最小, 其中$\boldsymbol{F}: \boldsymbol{R}^n \to \boldsymbol{R}^m$是连续可微函数.

记 $\boldsymbol{F}(\boldsymbol{x}) = \left(F_1(\boldsymbol{x}), F_2(\boldsymbol{x}), \cdots, F_m(\boldsymbol{x})\right)^{\mathrm{T}}$, 则非线性最小二乘问题可以表示为

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{F}(\boldsymbol{x})\|^2 = \frac{1}{2}\sum_{i=1}^{m} F_i^2(\boldsymbol{x}).$$

目标函数 f 的梯度和 Hesse 阵分别为

$$\boldsymbol{g}(\boldsymbol{x}) := \nabla f(\boldsymbol{x}) = \nabla\left(\frac{1}{2}\|\boldsymbol{F}(\boldsymbol{x})\|^2\right) = \boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{F}(\boldsymbol{x}) = \sum_{i=1}^{m} F_i(\boldsymbol{x})\nabla F_i(\boldsymbol{x})$$

$$\begin{aligned}\boldsymbol{G}(\boldsymbol{x}) := \nabla^2 f(\boldsymbol{x}) &= \sum_{i=1}^{m} \nabla F_i(\boldsymbol{x})(\nabla F_i(\boldsymbol{x}))^{\mathrm{T}} + \sum_{i=1}^{m} F_i(\boldsymbol{x})\nabla^2 F_i(\boldsymbol{x}) \\ &= \boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \sum_{i=1}^{m} F_i(\boldsymbol{x})\nabla^2 F_i(\boldsymbol{x}) := \boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \boldsymbol{S}(\boldsymbol{x})\end{aligned}$$

非线性最小二乘问题

- 也即:

$$g(x) := J(x)^{\mathrm{T}} F(x)$$

$$G(x) := J(x)^{\mathrm{T}} J(x) + S(x)$$

$$J(x) = F'(x) = (\nabla F_1(x), \nabla F_2(x), \cdots, \nabla F_m(x))^{\mathrm{T}},$$

$$S(x) = \sum_{i=1}^{m} F_i(x) \nabla^2 F_i(x).$$

Gauss-Newton算法缺点: 要求矩阵J_k列满秩

利用牛顿型迭代算法, 可得非线性最小二乘问题的迭代算法:

$$x_{k+1} = x_k - \left(J_k^{\mathrm{T}} J_k + S_k\right)^{-1} J_k^{\mathrm{T}} F_k$$

其中$S(x)$的计算量较大, 忽略该项, 可得Gauss-Newton迭代算法:

$$\boxed{x_{k+1} = x_k + d_k^{\mathrm{GN}},} \qquad \boxed{d_k^{\mathrm{GN}} = -\left[J_k^{\mathrm{T}} J_k\right]^{-1} J_k^{\mathrm{T}} F_k}$$

最优性条件

- 等式约束问题的最优性条件
- 不等式约束问题的最优性条件
- 一般约束问题的最优性条件
- 鞍点和对偶问题

等式约束问题的最优性条件

- 讨论下列等式约束问题的最优性条件

$$\begin{aligned} &\min \ f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbf{R}^n, \\ &\text{s.t.} \ \ h_i(\boldsymbol{x}) = 0, \ \ i = 1, 2, \cdots, l. \end{aligned}$$

求其拉格朗日函数:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{i=1}^{l} \lambda_i h_i(\boldsymbol{x}),$$

$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_l)^{\mathrm{T}}$ 为乘子向量.

下面的拉格朗日定理给出了该等式约束问题取极小值的一阶必要条件, 也即KKT条件

等式约束问题的最优性条件

- 等式约束问题的一阶必要条件

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{i=1}^{l} \lambda_i h_i(\boldsymbol{x}),$$

定理 8.1 （**拉格朗日定理**） 假设 $\boldsymbol{x}^*$ 是问题 (8.1) 的局部极小点，$f(\boldsymbol{x})$ 和 $h_i(\boldsymbol{x})\,(i = 1, 2, \cdots, l)$ 在 $\boldsymbol{x}^*$ 的某邻域内连续可微．若向量组 $\nabla h_i(\boldsymbol{x}^*)\,(i = 1, 2, \cdots, l)$ 线性无关，则存在乘子向量 $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_l^*)^{\mathrm{T}}$ 使得

$$\nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0},$$

即

$$\nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{l} \lambda_i^* \nabla h_i(\boldsymbol{x}^*) = \mathbf{0}.$$

等式约束问题的最优性条件

- **等式约束问题的二阶充分条件**(目标函数和约束函数都是二阶连续可微)

定理 8.2 对于等式约束问题 (8.1)，假设 $f(\boldsymbol{x})$ 和 $h_i(\boldsymbol{x})(i=1,2,\cdots,l)$ 都是二阶连续可微的，并且存在 $(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\in\mathbf{R}^n\times\mathbf{R}^l$ 使得 $\boxed{\nabla L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)=\mathbf{0}.}$ 若对任意的 $\mathbf{0}\neq\boldsymbol{d}\in\mathbf{R}^n$，$\boxed{\nabla h_i(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d}=0}(i=1,2,\cdots,l)$，均有 $\boxed{\boldsymbol{d}^{\mathrm{T}}\nabla^2_{\boldsymbol{xx}}L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{d}>0,}$ 则 $\boldsymbol{x}^*$ 是问题 (8.1) 的一个严格局部极小点.

类似二阶导数大于0/凸函数定义的原理

不等式约束问题的最优性条件

- 讨论下列不等式约束问题的最优性条件

$$\min \ f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbf{R}^n,$$

$$\text{s.t.} \ \ g_i(\boldsymbol{x}) \geqslant 0, \ \ i = 1, 2, \cdots, m.$$

可行域为 $\mathcal{D} = \{\boldsymbol{x} \in \mathbf{R}^n | \, g_i(\boldsymbol{x}) \geqslant 0, i = 1, 2, \cdots, m\}$, 指标集 $\mathcal{I} = \{1, 2, \cdots, m\}$.

定义 8.1 若问题 (8.5) 的一个可行点 $\bar{\boldsymbol{x}} \in \mathcal{D}$ 使得 $g_i(\bar{\boldsymbol{x}}) = 0$, 则称不等式约束 $g_i(\boldsymbol{x}) \geqslant 0$ 为 $\bar{\boldsymbol{x}}$ 的有效约束. 反之, 若有 $g_i(\bar{\boldsymbol{x}}) > 0$, 则称不等式约束 $g_i(\boldsymbol{x}) \geqslant 0$ 为 $\bar{\boldsymbol{x}}$ 的非有效约束. 称集合

$$\mathcal{I}(\bar{\boldsymbol{x}}) = \{i | g_i(\bar{\boldsymbol{x}}) = 0\}$$

为 $\bar{\boldsymbol{x}}$ 处的有效约束指标集, 简称为 $\boldsymbol{x}$ 处的有效集 (或积极集).

不等式约束问题的最优性条件

- 不等式约束问题的一阶必要条件

定理 8.3（**KKT 条件**） 设 $\boldsymbol{x}^*$ 是不等式约束问题 (8.5) 的局部极小点，有效约束集 $\mathcal{I}(\boldsymbol{x}^*) = \{i \mid g_i(\boldsymbol{x}^*) = 0,\ i = 1, 2, \cdots, m\}$. 并设 $f(\boldsymbol{x})$ 和 $g_i(\boldsymbol{x})\ (i = 1, 2, \cdots, m)$ 在 $\boldsymbol{x}^*$ 处可微. 若向量组 $\nabla g_i(\boldsymbol{x}^*)\ (i \in \mathcal{I}(\boldsymbol{x}^*))$ 线性无关，则存在向量 $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^{\mathrm{T}}$ 使得

$$
\begin{cases}
\boxed{\nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) = \mathbf{0},} \\
g_i(\boldsymbol{x}^*) \geqslant 0, \quad \lambda_i^* \geqslant 0, \quad \boxed{\lambda_i^* g_i(\boldsymbol{x}^*) = 0,} \quad i = 1, 2, \cdots, m.
\end{cases}
$$

拉格朗日函数的一阶导数为0

$$\boxed{\lambda_i^* = 0, \quad i \in \mathcal{I} \backslash \mathcal{I}(\boldsymbol{x}^*)}$$

一般约束问题的最优性条件

- 讨论下列一般约束问题的最优性条件

$$
\begin{aligned}
\min\ & f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbf{R}^n, \\
\text{s.t.}\ & h_i(\boldsymbol{x}) = 0,\ i = 1, 2, \cdots, l, \\
& g_i(\boldsymbol{x}) \geqslant 0,\ i = 1, 2, \cdots, m.
\end{aligned}
$$

记可行域为 $\mathcal{D} = \{\boldsymbol{x} \in \mathbf{R}^n | \, h_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E},\ g_i(\boldsymbol{x}) \geqslant 0,\ i \in \mathcal{I}\}$, 指标集 $\mathcal{E} = \{1, 2, \cdots, l\}$, $\mathcal{I} = \{1, 2, \cdots, m\}$.

一般约束问题的最优性条件

- 一般约束问题的KKT一阶必要条件

定理 8.4（KKT 一阶必要条件） 设 $\boldsymbol{x}^*$ 是一般约束问题的局部极小点，在 $\boldsymbol{x}^*$ 处的有效约束集为

$$S(\boldsymbol{x}^*) = \mathcal{E} \cup \mathcal{I}(\boldsymbol{x}^*) = \mathcal{E} \cup \{i \mid g_i(\boldsymbol{x}^*) = 0, i \in \mathcal{I}\}.$$

并设 $f(\boldsymbol{x})$, $h_i(\boldsymbol{x})\,(i \in \mathcal{E})$ 和 $g_i(\boldsymbol{x})\,(i \in \mathcal{I})$ 在 $\boldsymbol{x}^*$ 处可微. 若向量组 $\boxed{\nabla h_i(\boldsymbol{x}^*)\,(i \in \mathcal{E}), \nabla g_i(\boldsymbol{x}^*)\,(i \in \mathcal{I}(\boldsymbol{x}^*))\text{ 线性无关}}$, 则存在向量 $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \in \mathbf{R}^l \times \mathbf{R}^m$, 其中 $\boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*, \cdots, \mu_l^*)^{\mathrm{T}}$, $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^{\mathrm{T}}$, 使得

$$\begin{cases} \boxed{\nabla f(\boldsymbol{x}^*) - \sum\limits_{i=1}^{l} \mu_i^* \nabla h_i(\boldsymbol{x}^*) - \sum\limits_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) = \boldsymbol{0},} \\ h_i(\boldsymbol{x}^*) = 0, \ i \in \mathcal{E}, \\ g_i(\boldsymbol{x}^*) \geqslant 0, \quad \lambda_i^* \geqslant 0, \quad \lambda_i^* g_i(\boldsymbol{x}^*) = 0, \ i \in \mathcal{I}. \end{cases}$$

一般约束问题的最优性条件

- 一般约束问题的二阶充分条件

$$S(\boldsymbol{x}^*) = \mathcal{E} \cup \mathcal{I}(\boldsymbol{x}^*) = \mathcal{E} \cup \{i \,|\, g_i(\boldsymbol{x}^*) = 0, i \in \mathcal{I}\}$$

定理 8.5 对于一般约束优化问题，假设 $f(\boldsymbol{x})$, $h_i(\boldsymbol{x})\,(i \in \mathcal{E})$ 和 $g_i(\boldsymbol{x})\,(i \in \mathcal{I})$ 都是二阶连续可微的，有效约束集 $S(\boldsymbol{x}^*)$ 由式 (8.12) 所定义. 且 $(\boldsymbol{x}^*, (\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*))$ 是问题 (8.11) 的 KKT 点. 若对任意的 $\mathbf{0} \neq \boldsymbol{d} \in \mathbf{R}^n$, $\nabla h_i(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d} = 0\,(i \in \mathcal{E})$, $\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d} = 0\,(i \in \mathcal{I}(\boldsymbol{x}^*))$, 均有 $\boldsymbol{d}^{\mathrm{T}}\nabla^2_{\boldsymbol{xx}}L(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)\boldsymbol{d} > 0$, 则 $\boldsymbol{x}^*$ 是问题 (8.11) 的一个严格局部极小点.

$$L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{i=1}^{l} \mu_i h_i(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x}).$$

约束优化问题的鞍点和对偶

- 一般约束优化问题

$$\begin{aligned} \min \ & f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbf{R}^n, \\ \text{s.t.} \ & h_i(\boldsymbol{x}) = 0, \ i = 1, 2, \cdots, l, \\ & g_i(\boldsymbol{x}) \geqslant 0, \ i = 1, 2, \cdots, m. \end{aligned}$$

定义 8.3 对约束优化问题 (8.11), 若存在 $\boldsymbol{x}^*$ 和 $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$, 其中 $\boldsymbol{\lambda}^* \geqslant \mathbf{0}$, 满足

$$L(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leqslant L(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \leqslant L(\boldsymbol{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*), \quad \forall (\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m_+,$$

则称 $(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ 为约束优化问题 (8.11) 的拉格朗日函数的鞍点, 通常简称 $\boldsymbol{x}^*$ 为问题 (8.11) 的鞍点.

约束优化问题的鞍点和对偶

$$\begin{aligned}\min\ & f(\boldsymbol{x}), && \boldsymbol{x}\in\mathbf{R}^n,\\ \text{s.t.}\ & h_i(\boldsymbol{x})=0, && i=1,2,\cdots,l,\\ & g_i(\boldsymbol{x})\geqslant 0, && i=1,2,\cdots,m.\end{aligned}$$

- 约束优化问题的对偶问题.

$$\boldsymbol{H}(\boldsymbol{x})=\big(h_1(\boldsymbol{x}),\cdots,h_l(\boldsymbol{x})\big)^{\mathrm{T}},\quad \boldsymbol{G}(\boldsymbol{x})=\big(g_1(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x})\big)^{\mathrm{T}}.$$

令 $\boldsymbol{y}\in\mathbf{R}^l$, $\boldsymbol{z}\in\mathbf{R}^m$, 定义函数

$$L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=f(\boldsymbol{x})-\boldsymbol{H}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{G}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{z}$$

$$\psi(\boldsymbol{y},\boldsymbol{z})=\inf_{\boldsymbol{x}\in\mathbf{R}^n} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}).$$

易知 $\psi(\boldsymbol{y},\boldsymbol{z})$ 关于 $(\boldsymbol{y},\boldsymbol{z})$ 是凹函数.

拉格朗日对偶定义为

$$\begin{aligned}\max\ & \psi(\boldsymbol{y},\boldsymbol{z}),\\ \text{s.t.}\ & \boldsymbol{y}\in\mathbf{R}^l,\ \boldsymbol{z}\in\mathbf{R}^m_+.\end{aligned}$$

Wolfe 对偶定义为

$$\begin{aligned}\max\ & L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}),\\ \text{s.t.}\ & \nabla_{\boldsymbol{x}}L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=\mathbf{0},\\ & \boldsymbol{y}\in\mathbf{R}^l,\ \boldsymbol{z}\in\mathbf{R}^m_+.\end{aligned}$$

线性规划问题

- 线性规划问题的标准形式:

$$\begin{aligned} \min \ & \sum_{i=1}^{n} c_i x_i, \\ \text{s.t.} \ & \boldsymbol{a}_i^{\mathrm{T}} \boldsymbol{x} = b_i, \ i = 1, 2, \cdots, m, \\ & x_i \geqslant 0, \ i = 1, 2, \cdots, n. \end{aligned}$$

令:

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{R}^n, \quad \boldsymbol{c} = (c_1, c_2, \cdots, c_n)^{\mathrm{T}} \in \mathbf{R}^n,$$

$$\boldsymbol{A} = (\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_m})^{\mathrm{T}} \in \mathbf{R}^{m \times n}, \quad \boldsymbol{b} = (b_1, b_2, \cdots, b_m)^{\mathrm{T}} \in \mathbf{R}^m,$$

线性规划问题

- 线性规划问题的标准型:

$$\begin{aligned} \min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}, \end{aligned}$$

约定$\boldsymbol{A}$是行满秩
约定$\boldsymbol{b}$是非负向量

其中向量不等式 $\boldsymbol{x} \geqslant 0$ 按分量取.

- 任何线性规划问题均可化为标准型

$$\max \ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \Longrightarrow -\min\left(-\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}\right)$$

$$\sum_{i=1}^{n}\alpha_i x_i \geqslant \beta \Longrightarrow \sum_{i=1}^{n}\alpha_i x_i - x_{n+1} = \beta$$

松弛变量 x_{n+1}

同时, 增加非负约束 $x_{n+1} \geqslant 0$.

对于自由变量 x_i, 可引入两个非负变量 u_{i1} 和 u_{i2}, 并令 $x_i = u_{i1} - u_{i2}$.

线性规划问题

- 凸集的顶点(极点)和极方向

顶点:具备相似几何意义

定义 9.1 设 $\mathcal{C} \subset \mathbf{R}^n$ 是闭凸集, $\boldsymbol{x} \in \mathcal{C}$. 若不存在两个不同的点 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in \mathcal{C}$ 以及数 $\alpha \in (0,1)$, 使得 $\boldsymbol{x} = \alpha \boldsymbol{x}^{(1)} + (1-\alpha)\boldsymbol{x}^{(2)}$, 则称 $\boldsymbol{x}$ 是凸集 $\mathcal{C}$ 的一个顶点或极点, 即 $\boldsymbol{x} \in \mathcal{C}$ 是顶点的充分必要条件是 $\boldsymbol{x}$ 不能表示为 $\mathcal{C}$ 中两个不同点的凸组合.

类似区域边界方向

定义 9.2 设 $\mathcal{C} \subset \mathbf{R}^n$ 是闭凸集, $\boldsymbol{d} \in \mathbf{R}^n$ 为非零向量. 若对任意的 $\boldsymbol{x} \in \mathcal{C}$, 均有

$$\{\boldsymbol{x} + \alpha \boldsymbol{d} | \alpha \geqslant 0\} \subset \mathcal{C},$$

则称 $\boldsymbol{d}$ 是 $\mathcal{C}$ 的一个方向. 若 $\mathcal{C}$ 的方向 $\boldsymbol{d}$ 不能表示为 $\mathcal{C}$ 的其他两个不同方向的正线性组合, 则称它为 $\mathcal{C}$ 的一个极方向.

线性规划问题

- 线性规划问题的可行域和方向

$$\mathcal{D} = \{x | Ax = b,\ x \geqslant 0\}$$

为线性规划问题的可行域. 显然 $\mathcal{D}$ 是一个凸集. 事实上, $\mathcal{D}$ 是一个多面体区域. 由凸集的性质知 $\mathcal{D}$ 无界的充要条件是它有方向.

定理 9.1 $\boldsymbol{d} \in \mathbf{R}^n$ 是线性规划问题 (9.2) 的可行域 $\mathcal{D}$ 的一个方向的充要条件是它满足 $\boldsymbol{d} \geqslant 0$ 且 $\boldsymbol{A}\boldsymbol{d} = \mathbf{0}$.

线性规划问题

- 线性规划可行域的表示定理

定理 9.2 (表示定理) 设线性规划问题 (9.2) 的可行域 $\mathcal{D}$ 非空. 则

(1) $\mathcal{D}$ 有有限个顶点 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$;

(2) $\mathcal{D}$ 有极方向的充要条件是 $\mathcal{D}$ 无界. 而且, 若 $\mathcal{D}$ 无界, 则存在有限个极方向 $\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_t$;

(3) $\boldsymbol{x} \in \mathcal{D}$ 的充要条件是存在非负数 $\alpha_i \in \mathbf{R}$ $(i = 1, 2, \cdots, r)$ 和非负数 $\beta_i \in \mathbf{R}$ $(i = 1, 2, \cdots, t)$ 使得

$$\boldsymbol{x} = \sum_{i=1}^{r} \alpha_i \boldsymbol{x}_i + \sum_{i=1}^{t} \beta_i \boldsymbol{d}_i,$$

式中: $\sum\limits_{i=1}^{r} \alpha_i = 1$.

线性规划问题

- 线性规划问题的基

$$\min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

$$\text{s.t.}\ \ \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},$$

$$\boldsymbol{x}\geqslant\boldsymbol{0},$$

最多可能有C_n^m组基

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbf{R}^n,\quad \boldsymbol{c}=(c_1,c_2,\cdots,c_n)^{\mathrm{T}}\in\mathbf{R}^n,$$

$$\boldsymbol{A}=(\boldsymbol{a_1},\boldsymbol{a_2},\cdots,\boldsymbol{a_m})^{\mathrm{T}}\in\mathbf{R}^{m\times n},\quad \boldsymbol{b}=(b_1,b_2,\cdots,b_m)^{\mathrm{T}}\in\boxed{\mathbf{R}^m},$$

定义 9.3 线性规划问题 (9.2) 的系数矩阵 $\boldsymbol{A}$ 的 $m\times m$ 阶非奇异子矩阵称为线性规划问题的一组基. 换言之, 线性规划问题的基是由矩阵 $\boldsymbol{A}$ 的 m 个线性无关列组成的子矩阵. 构成基的列向量称为基向量. 相应的变量称为基变量. 其他变量称为非基变量.

线性规划问题

- 线性规划问题的基

例 9.1 线性规划问题

$$
\begin{aligned}
\min \quad & 2x_1 + x_2 \\
\text{s.t.} \quad & 2x_1 + x_2 + x_3 = 2, \\
& -3x_1 + 2x_2 + x_4 = 3, \\
& x_i \geqslant 0, \ i = 1, 2, 3, 4
\end{aligned}
$$

中有基

$$
\boldsymbol{B}_1 = \boxed{\begin{pmatrix} 2 & 1 \\ -3 & 2 \end{pmatrix}}, \quad \boldsymbol{B}_2 = \begin{pmatrix} 2 & 1 \\ -3 & 0 \end{pmatrix}, \quad \boldsymbol{B}_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
$$

等. 相应的基变量分别为: x_1, x_2; x_1, x_3; x_3, x_4.

线性规划问题

线性规划问题的基可行解

定义 9.4 线性规划问题 (9.2) 的可行点称为可行解. 令非基变量为 0 所得到的可行解称为线性规划问题的基可行解.

线性规划可以有多个基可行解, 线性规划问题 (9.2) 最多可以有 C_n^m 个基可行解. 例如, 在例 9.1 **中, 有** $(1/7, 12/7, 0, 0)^{\mathrm{T}}$ **和** $(0, 0, 2, 3)^{\mathrm{T}}$ **等基可行解. 相应的基分别为** $\boldsymbol{B}_1$ **和** $\boldsymbol{B}_3$ **等.**

定理 9.3 线性规划问题 (9.2) 的可行解是基可行解当且仅当它的正分量所对应的系数构成的列向量组线性无关.

定理 9.4 线性规划问题 (9.2) 的基可行解对应于可行域的顶点.

线性规划问题

线性规划的**最优解**必可在**可行域**的**顶点**达到

- 线性规划理论的基本定理

> **定理 9.5** (线性规划基本定理)
>
> (1) 若线性规划问题有可行解, 则必有基可行解.
>
> (2) 若线性规划问题有最优解, 则必有最优基可行解.
>
> (3) 若线性规划问题的可行域有界, 则必有最优解.

由定理 9.4 和定理 9.5 易知, 线性规划问题若有最优解, 则必有可行域的顶点为最优解. 由于线性规划是凸规划, 其最优解集合是凸集. 因此, 若目标函数在多个顶点达到最优, 则在这些点的凸组合处也达到最优, 此时问题有无穷多个解.

线性规划问题——单纯形算法

- **单纯形法基本原理**: 线性规划问题若有最优解, 则必有最优基可行解. 基可行解只有有限个, 故求解线性规划问题只需要求出基可行解中使目标函数数值最小者.
- **单纯形法基本思想**: 从一个基可行解出发, 若该基可行解不是问题的最优解, 则按某种法则寻找另一个基可行解, 如此下去, 直到求得问题的一个最优基可行解.

线性规划问题——单纯形算法

- 单纯形法实例讲解:

例 9.2 求解下列线性规划问题

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -2x_1 - 3x_2, \\
\text{s.t.} \quad & -x_1 + x_2 \leqslant 3, \\
& -2x_1 + x_2 \leqslant 2, \\
& 4x_1 + 2x_2 \leqslant 16, \\
& x_1,\ x_2 \geqslant 0.
\end{aligned}
$$

线性规划问题——单纯形算法

- 单纯形法实例讲解:

引入三个松弛变量 x_3, x_4, x_5 将其转换成标准形

$$
\begin{aligned}
\min\ & f(\boldsymbol{x}) = -2x_1 - 3x_2, \\
\text{s.t.}\ & -x_1 + x_2 \leqslant 3, \\
& -2x_1 + x_2 \leqslant 2, \\
& 4x_1 + 2x_2 \leqslant 16, \\
& x_1,\ x_2 \geqslant 0.
\end{aligned}
$$

$$
\begin{aligned}
\min\ & f(\boldsymbol{x}) = -2x_1 - 3x_2, \\
\text{s.t.}\ & -x_1 + x_2 + x_3 = 3, \\
& -2x_1 + x_2 + x_4 = 2, \\
& 4x_1 + x_2 + x_5 = 16, \\
& x_i \geqslant 0,\quad i = 1, 2, \cdots, 5.
\end{aligned}
$$

线性规划问题——单纯形算法

- 单纯形法实例讲解:

对问题的可行域变形, 使基变量位于方程组的左边, 非基变量位于方程组的右边

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -2x_1 - 3x_2, \\
\text{s.t.} \quad & -x_1 + x_2 + x_3 = 3, \\
& -2x_1 + x_2 + x_4 = 2, \\
& 4x_1 + x_2 + x_5 = 16, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -2x_1 - 3x_2 \\
\text{s.t.} \quad & x_3 = 3 + x_1 - x_2, \\
& x_4 = 2 + 2x_1 - x_2, \\
& x_5 = 16 - 4x_1 - x_2, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.
\end{aligned}
$$

令非基变量 $x_1 = 0$, $x_2 = 0$ 得基可行解 $\boldsymbol{x}_0 = (0, 0, 3, 2, 16)^{\mathrm{T}}$. 相应的目标函数值为 $f(\boldsymbol{x}_0) = 0$. $\boldsymbol{x}_0$ 显然不是问题的最优解, 因为当 x_1 或 x_2 取正值时, 目标函数的值可以减小.

线性规划问题——单纯形算法

- 单纯形法实例讲解:

寻找一个新的基可行解，使x_1或x_2取正值. 将上一步中的非基变量x_1或x_2(换入变量)取代原来的基变量x_3, x_4或x_5(换出变量)

$$\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -2x_1 - 3x_2 \\
\text{s.t.} \quad & x_3 = 3 + x_1 - x_2, \\
& x_4 = 2 + 2x_1 - x_2, \\
& x_5 = 16 - 4x_1 - x_2, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.
\end{aligned}$$

$$\begin{cases}
x_3 = 3 - x_2 \geqslant 0, & \Longleftrightarrow \quad x_2 \leqslant 3, \\
x_4 = 2 - x_2 \geqslant 0, & \Longleftrightarrow \quad \boxed{x_2 \leqslant 2,} \\
x_5 = 16 - x_2 \geqslant 0, & \Longleftrightarrow \quad x_2 \leqslant 16.
\end{cases}$$

取x_4换出，可以保证其他都为非负性

若取x_2作为换入变量，x_1仍然是非基变量，取值为0

确定换出变量的原则是使得新的基解可行，即满足非负性条件.

线性规划问题——单纯形算法

- 单纯形法实例讲解:

新的基变量x_3, x_2, x_5

$$\begin{aligned}\min\quad & f(\boldsymbol{x}) = -2x_1 - 3x_2 \\ \text{s.t.}\quad & x_3 = 3 + x_1 - x_2, \\ & x_4 = 2 + 2x_1 - x_2, \\ & x_5 = 16 - 4x_1 - x_2, \\ & x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}\min\quad & f(\boldsymbol{x}) = -2x_1 - 3x_2, \\ \text{s.t.}\quad & x_2 + x_3 = 3 + x_1, \\ & x_2 = 2 + 2x_1 - x_4, \\ & x_2 + x_5 = 16 - 4x_1, \\ & x_i \geqslant 0, \quad i = 1, 2, \cdots, 5,\end{aligned}$$

线性规划问题——单纯形算法

- 单纯形法实例讲解:

代入消减, 调整形式, 让基变量位于方程组左边

用非基变量来描述

$$
\begin{aligned}
\min\ \ & f(\boldsymbol{x}) = -2x_1 - 3x_2, \\
\text{s.t.}\ \ & x_2 + x_3 = 3 + x_1, \\
& x_2 = 2 + 2x_1 - x_4, \\
& x_2 + x_5 = 16 - 4x_1, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5,
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
\min\ \ & f(\boldsymbol{x}) = \boxed{-6 - 8x_1 + 3x_4,} \\
\text{s.t.}\ \ & x_3 = 1 - x_1 + x_4, \\
& x_2 = 2 + 2x_1 - x_4, \\
& x_5 = 14 - 6x_1 + x_4, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.
\end{aligned}
$$

令非基变量 $x_1 = 0, x_4 = 0$ 得基可行解 $\boldsymbol{x}_1 = (0, 2, 1, 0, 14)^{\mathrm{T}}$. 相应的目标函数为 $f(\boldsymbol{x}_1) = -6 < f(\boldsymbol{x}_0)$. 由于 x_1 的系数 -8 为负数, 因此, 当它取正值时, 目标函数值还会减小, 即 $\boldsymbol{x}_1$ 不是问题的最优解.

线性规划问题——单纯形算法

- 单纯形法实例讲解:

令x_1成为换入变量取代原来的基变量x_3, x_2或x_5

$$\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -6 - 8x_1 + 3x_4, \\
\text{s.t.} \quad & x_3 = 1 - x_1 + x_4, \\
& x_2 = 2 + 2x_1 - x_4, \\
& x_5 = 14 - 6x_1 + x_4, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.
\end{aligned}$$

$$\begin{cases}
x_3 = 1 - x_1 \geqslant 0, & \Longleftrightarrow x_1 \leqslant 1, \\
x_2 = 2 + 2x_1 \geqslant 0, & \Longleftrightarrow x_1 \geqslant -1, \\
x_5 = 14 - 6x_1 \geqslant 0, & \Longleftrightarrow x_1 \leqslant \frac{7}{3}
\end{cases}$$

故换出变量为x_3

线性规划问题——单纯形算法

- 单纯形法实例讲解:

换出变量为x_3, 调整形式

$$\begin{aligned}\min\quad & f(\boldsymbol{x}) = -6 - 8x_1 + 3x_4,\\ \text{s.t.}\quad & x_3 = 1 - x_1 + x_4,\\ & x_2 = 2 + 2x_1 - x_4,\\ & x_5 = 14 - 6x_1 + x_4,\\ & x_i \geqslant 0,\quad i = 1, 2, \cdots, 5.\end{aligned}$$

$$\Rightarrow$$

$$\begin{aligned}\min\quad & f(\boldsymbol{x}) = -14 + 8x_3 - 5x_4,\\ \text{s.t.}\quad & x_1 = 1 - x_3 + x_4,\\ & x_2 = 4 - 2x_3 + x_4,\\ & x_5 = 8 + 6x_3 - 5x_4,\\ & x_i \geqslant 0,\quad i = 1, 2, \cdots, 5.\end{aligned}$$

令非基变量 $x_3 = 0, x_4 = 0$ 得基可行解 $\boldsymbol{x}_2 = (1, 4, 0, 0, 8)^{\mathrm{T}}$. 相应的目标函数为 $f(\boldsymbol{x}_2) = -14 < f(\boldsymbol{x}_1)$. 由于 x_4 的系数 -5 为负数, 因此, 当它取正值时, 目标函数值还会减小, 即 $\boldsymbol{x}_2$ 不是问题的最优解.

线性规划问题——单纯形算法

- 单纯形法实例讲解:

同理,换出变量为x_5,调整形式

系数均为正

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -14 + 8x_3 - 5x_4, \\
\text{s.t.} \quad & x_1 = 1 - x_3 + x_4, \\
& x_2 = 4 - 2x_3 + x_4, \\
& x_5 = 8 + 6x_3 - 5x_4, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -22 + \boxed{2x_3 + x_5,} \\
\text{s.t.} \quad & x_1 = 2.6 + 0.2x_3 - 0.2x_5, \\
& x_2 = 5.6 - 0.8x_3 - 0.2x_5, \\
& x_4 = 1.6 + 1.2x_3 - 0.2x_5, \\
& x_i \geqslant 0, \quad i = 1, 2, \cdots, 5.
\end{aligned}
$$

令非基变量 $x_3 = 0, x_5 = 0$ 得基可行解 $\boldsymbol{x}_3 = (2.6, 5.6, 0, 1.6, 0)^{\mathrm{T}}$. 相应的目标函数为 $f(\boldsymbol{x}_3) = -22 < f(\boldsymbol{x}_2)$. 由于, 目标函数中非基变量 x_3, x_5 的系数均为正数, 因此, 当 x_3 或 x_5 的值变大时, 目标函数值不会再减小, 即 $\boldsymbol{x}_3$ 是问题的最优解.

线性规划问题——单纯形算法

$\boldsymbol{A} = (\boldsymbol{I} \ \ \boldsymbol{N})$　　$\boldsymbol{c}^{\mathrm{T}} = (\boldsymbol{c}_B^{\mathrm{T}}, \boldsymbol{c}_N^{\mathrm{T}})$　　每个元素都大于0

算法 9.1 (单纯形法)

步骤 0，取初始基可行解 $\boldsymbol{x} = (\boldsymbol{x}_B^{\mathrm{T}}, \boldsymbol{x}_N^{\mathrm{T}})^{\mathrm{T}} = (\boldsymbol{b}^{\mathrm{T}}, \boldsymbol{0}^{\mathrm{T}})^{\mathrm{T}}$.

步骤 1，计算非基变量的检验数 $\boldsymbol{\sigma}_N = \boldsymbol{c}_N^{\mathrm{T}} - \boldsymbol{c}_B^{\mathrm{T}}\boldsymbol{N}$. 若 $\boldsymbol{\sigma}_N \boxed{\geqslant \boldsymbol{0}}$, 则停止计算，得解 $\boldsymbol{x}$; 否则，在非基变量中确定换入变量 x_j 使 $\sigma_j = \min\{\sigma_k, k \in N\}$.

步骤 2，令 $\boldsymbol{\alpha}_j = (\alpha_{ij})_{i=1}^m$ 表示矩阵 $\boldsymbol{N}$ 的第 j 列. 若 $\boldsymbol{\alpha}_j \leqslant \boldsymbol{0}$, 则停止计算. 此时，问题的解不存在; 否则，转步骤 3.

步骤 3，确定下标 i 使

$$\frac{b_i}{\alpha_{ij}} = \min\left\{\frac{b_k}{\alpha_{kj}} \;\middle|\; \alpha_{kj} > 0\right\}.$$

步骤 4，以 α_{ij} 为元，对方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 实施行初等变换，将第 j 列变成 $\boldsymbol{e}_i$, 其中 e_i 为单位矩阵 $\boldsymbol{I}_m$ 的第 i 列，转步骤 1.

线性规划问题

- **线性规划实战:**

用Matlab优化工具箱的函数linprog求解下列线性规划模型

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\
\text{s.t.} \quad & \boldsymbol{A}_1\boldsymbol{x} \leqslant \boldsymbol{b}_1, \\
& \boldsymbol{A}_2\boldsymbol{x} = \boldsymbol{b}_2, \\
& \boldsymbol{l} \leqslant \boldsymbol{x} \leqslant \boldsymbol{u},
\end{aligned}
$$

$\boldsymbol{l}, \boldsymbol{u}$ 分别为决策变量的下界和上界

线性规划问题

- **线性规划实战:**

例 9.5 (下料问题) 某单位需要加工制作 100 套工架, 每套工架需用长为 2.9m、2.1m 和 1.5m 的圆钢各一根. 已知原材料长 7.4m, 现在的问题是如何下料使得所用的原材料最省?

套截更省材料, 几种可能的方案如下:

长度 \ 方案	A	B	C	D	E
2.9	1	2	0	1	0
2.1	0	0	2	2	1
1.5	3	1	2	0	3
合计	7.4	7.3	7.2	7.1	6.6
料头	0	0.1	0.2	0.3	0.8

设按方案ABCDE下料的原材料数分别为 x_1, x_2, x_3, x_4, x_5, 可得下列线性规划模型:

$$
\begin{aligned}
\min\ z &= 0x_1 + 0.1x_2 + 0.2x_3 + 0.3x_4 + 0.8x_5,\\
\text{s.t.}\ & x_1 + 2x_2 + x_4 = 100,\\
& 2x_3 + 2x_4 + x_5 = 100,\\
& 3x_1 + x_2 + 2x_3 + 3x_5 = 100,\\
& x_1,\ x_2,\ x_3,\ x_4,\ x_5 \geqslant 0.
\end{aligned}
$$

线性规划问题

线性规划实战:

例 9.5 (下料问题) 某单位需要加工制作 100 套工架，每套工架需用长为 2.9m、2.1m 和 1.5m 的圆钢各一根. 已知原材料长 7.4m，现在的问题是如何下料使得所用的原材料最省?

$$
\begin{aligned}
\min \quad & z = 0x_1 + 0.1x_2 + 0.2x_3 + 0.3x_4 + 0.8x_5, \\
\text{s.t.} \quad & x_1 + 2x_2 + x_4 = 100, \\
& 2x_3 + 2x_4 + x_5 = 100, \\
& 3x_1 + x_2 + 2x_3 + 3x_5 = 100, \\
& x_1,\ x_2,\ x_3,\ x_4,\ x_5 \geqslant 0.
\end{aligned}
$$

Matlab求解代码:

```
c=[0 0.1 0.2 0.3  0.8]';
A1=[-1,0,0,0,0;0,-1,0,0,0; ...
    0,0,-1,0,0;0,0,0,-1,0;0,0,0,0,-1];
b1=[0,0,0,0,0]';
A2=[1,2,0,1,0;0,0,2,2,1;3,1,2,0,3];
b2=[100,100,100]';
[x,fv]=linprog(c,A1,b1,A2,b2)
```

运行该程序之后, 立即可以得到最优解为

$$\boldsymbol{x} = (12.8243, 27.1757, 17.1757, 32.8243, 0)^{\mathrm{T}},$$

按四舍五入的方法取整得 $\boldsymbol{x} = (13, 27, 17, 33, 0)^{\mathrm{T}}$, 最优值为 $z = 16$.

二次规划问题

- 二次规划的目标函数是二次实函数，约束函数都是线性函数
- 介绍求解等式约束凸二次规划的零空间方法和拉格朗日乘子法(略)

$$
\begin{aligned}
\min \quad & \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},
\end{aligned}
$$

$\boldsymbol{H} \in \mathrm{R}^{n\times n}$ 对称正定; $\boldsymbol{A} \in \mathrm{R}^{m\times n}$ 行满秩; $\boldsymbol{c}, \boldsymbol{x} \in \mathrm{R}^{n}$; $\boldsymbol{b} \in \mathrm{R}^{m}$.

二次规划问题——零空间方法

- 零空间方法

设 $\boldsymbol{x}_0$ 满足 $\boldsymbol{A}\boldsymbol{x}_0=\boldsymbol{b}$. 记 $\boldsymbol{A}$ 的零空间为

$$\mathcal{N}(\boldsymbol{A})=\left\{\boldsymbol{z}\in\mathbf{R}^n \,|\, \boldsymbol{A}\boldsymbol{z}=\mathbf{0}\right\},$$

任一可行点 $\boldsymbol{x}$ 可表示成 $\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{z}\,(\boldsymbol{z}\in\mathcal{N}(\boldsymbol{A}))$.

$$\begin{aligned}\min\ & \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x}+\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},\\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\end{aligned} \quad\Longrightarrow\quad \begin{aligned}\min\ & \frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{z}+\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{c}+\boldsymbol{H}\boldsymbol{x}_0),\\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{z}=\mathbf{0}.\end{aligned}$$

二次规划问题——零空间方法

- 零空间方法

令 $\boldsymbol{Z} \in \mathrm{R}^{n\times(n-m)}$ 是 $\mathcal{N}(\boldsymbol{A})$ 的一组基组成的矩阵, 则对任意的 $\boldsymbol{d} \in \mathrm{R}^{n-m}$ 有 $\boldsymbol{z} = \boldsymbol{Z}\boldsymbol{d} \in \mathcal{N}(\boldsymbol{A})$. 于是问题 (10.2) 变为无约束优化问题

$$\boxed{\begin{aligned} \min\ & \frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{z} + \boldsymbol{z}^{\mathrm{T}}(\boldsymbol{c} + \boldsymbol{H}\boldsymbol{x}_0), \\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{0}. \end{aligned}} \Rightarrow \min\ \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Z})\boldsymbol{d} + \boldsymbol{d}^{\mathrm{T}}[\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{c} + \boldsymbol{H}\boldsymbol{x}_0)].$$

当 $\boldsymbol{H}$ 是半正定时, $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Z}$ 也是半正定的.

$\boldsymbol{d}^*$ 是稳定点, 则 $\boldsymbol{d}^*$ 也是全局极小点

同时 $\boldsymbol{x}^* = \boldsymbol{x}_0 + \boldsymbol{Z}\boldsymbol{d}^*$ 是 全局极小点

二次规划问题—零空间方法

- 零空间方法

$$\min \ \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Z})\boldsymbol{d}+\boldsymbol{d}^{\mathrm{T}}[\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{c}+\boldsymbol{H}\boldsymbol{x}_0)].$$

确定可行点 $\boldsymbol{x}_0$ 和零空间 $\mathcal{N}(\boldsymbol{A})$ 的基矩阵 $\boldsymbol{Z}$.

先对 $\boldsymbol{A}^{\mathrm{T}}$ 作 QR 分解

$$\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{Q}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}=\begin{pmatrix}\boldsymbol{Q}_1 & \boldsymbol{Q}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}$$

那么确立 $\boldsymbol{x}_0$ 和 $\boldsymbol{Z}$ 为

$$\boldsymbol{x}_0=\boldsymbol{Q}_1\boldsymbol{R}^{-\mathrm{T}}\boldsymbol{b},\quad \boldsymbol{Z}=\boldsymbol{Q}_2,$$

同时有

$$\boldsymbol{A}^{\dagger}=\boldsymbol{Q}_1\boldsymbol{R}^{-\mathrm{T}}.$$

二次规划问题—零空间方法

$$A^{\mathrm{T}} = Q\begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}\begin{pmatrix} R \\ 0 \end{pmatrix}$$

算法 10.1 (零空间方法)

步骤 0, 输入矩阵 $\boldsymbol{H}$, $\boldsymbol{A}$ 和向量 $\boldsymbol{c}$, $\boldsymbol{b}$.

步骤 1, 由式 (10.4) 对 $\boldsymbol{A}^{\mathrm{T}}$ 进行 QR 分解得矩阵 $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$ 和 $\boldsymbol{R}$.

步骤 2, 按式 (10.5) 计算可行点 $\boldsymbol{x}_0$ 和零空间 $\mathcal{N}(\boldsymbol{A})$ 的基矩阵 $\boldsymbol{Z}$.

$$\boldsymbol{x}_0 = \boldsymbol{Q}_1\boldsymbol{R}^{-\mathrm{T}}\boldsymbol{b}, \quad \boldsymbol{Z} = \boldsymbol{Q}_2,$$

步骤 3, 求解无约束优化子问题 (10.3) 得解 $\boldsymbol{d}^*$.

步骤 4, 计算全局极小点 $\boldsymbol{x}^* = \boldsymbol{x}_0 + \boldsymbol{Z}\boldsymbol{d}^*$ 和相应的拉格朗日乘子 $\boldsymbol{\lambda}^* = \boldsymbol{A}^{\dagger}(\boldsymbol{H}\boldsymbol{x}^* + \boldsymbol{c})$, 其中 $\boldsymbol{A}^{\dagger}$ 由式 (10.6) 确定.

$$\min\ \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Z})\boldsymbol{d} + \boldsymbol{d}^{\mathrm{T}}[\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{c} + \boldsymbol{H}\boldsymbol{x}_0)].$$

$$\boldsymbol{A}^{\dagger} = \boldsymbol{Q}_1\boldsymbol{R}^{-\mathrm{T}}.$$