高等数值算法与应用(4)

– 最小二乘法 (chap5) –

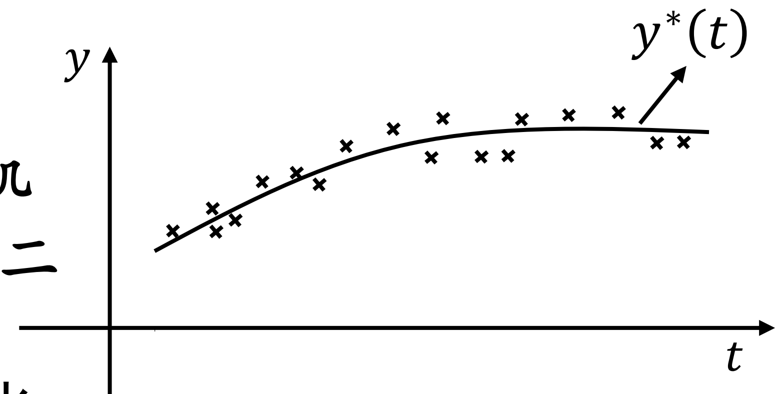
喻文健

Outline

- 曲线拟合与逼近问题
- 线性最小二乘问题的解法
- 伪逆与不满秩情况
- 非线性最小二乘

最小二乘法

- 最小二乘法（least squares）
  - 广泛用于数据拟合: 若误差的随机分布满足适当假设，则通过最小二乘能得到参数的最大似然估计
  - 优化: 使估计误差的平方和最小化
  - “求解”一个超定的方程组 (优化剩余)



$$\min_{S(t)} \sum_{i=1}^{m} [y_i - y(t_i)]^2$$

- 拟合的模型
  - 线性最小二乘: 设待求的$y(t) \approx \beta_1\phi_1(t) + \cdots + \beta_n\phi_n(t)$
  - ➡ 求n个基函数前的系数

超定方程组, 最小化剩余

  - 定义设计矩阵X, $x_{i,j} = \phi_j(t_i)$

$$\begin{bmatrix} \phi_1(t_1) & \phi_2(t_1) & \cdots & \phi_n(t_1) \\ \phi_1(t_2) & \phi_2(t_2) & \cdots & \\ \vdots & \vdots & \vdots & \\ \phi_1(t_m) & \phi_2(t_m) & \cdots & \phi_n(t_m) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \quad \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\boldsymbol{X\beta} \approx \boldsymbol{y}$$

最小二乘法

- 拟合的模型
  - 线性最小二乘的常见拟合模型
  - 直线： $y(t) \approx \beta_1 t + \beta_2$
  - 多项式： $y(t) \approx \beta_1 t^{n-1} + \ldots + \beta_{n-1} t + \beta_n$ Matlab中polyfit命令
  - 对数线性： $y(t) \approx K e^{\lambda t}$ 注意: 拟合目标发生了变化

    $$\log y \approx \beta_1 t + \beta_2, \text{ with } \beta_1 = \lambda, \ \beta_2 = \log K$$

  - 可分离的非线性最小二乘: 基函数中含额外参数
  - 指数函数： $y(t) \approx \beta_1 e^{-\lambda_1 t} + \ldots + \beta_n e^{-\lambda_n t}$

    β_i是线性组合系数，λ_i是非线性参数

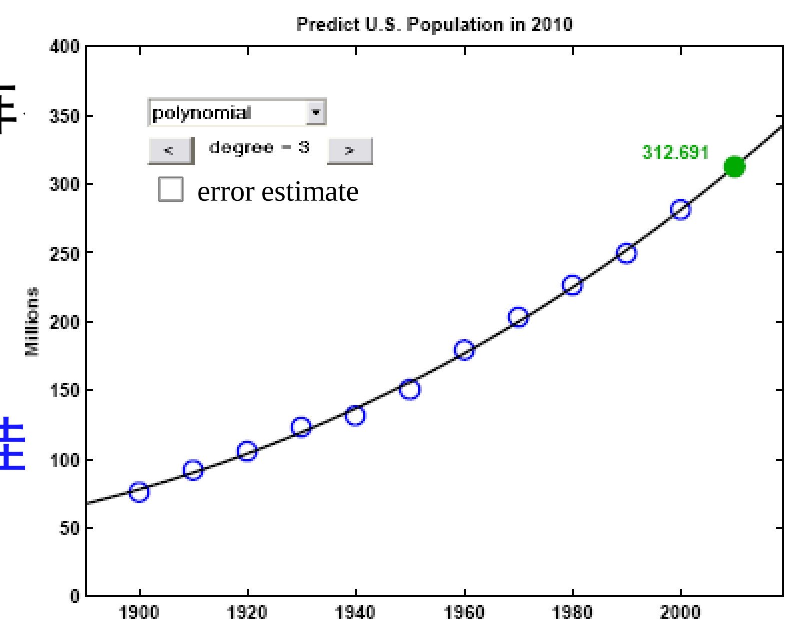
  - 高斯函数： $y(t) \approx \beta_1 e^{-\left(\frac{t-\mu_1}{\sigma_1}\right)^2} + \ldots \beta_n e^{-\left(\frac{t-\mu_n}{\sigma_n}\right)^2}$
  - 有理函数： $y(t) \approx \dfrac{\beta_1 t^{n-1} + \ldots + \beta_{n-1} t + \beta_n}{\alpha_1 t^{n-1} + \ldots + \alpha_{n-1} t + \alpha_n}$

最小二乘法

- 曲线拟合的例子
  - 根据1900~2000年美国人口(每10年采样), 预测2010年人口
  - 一种模型： $y \approx \beta_1 t^3 + \beta_2 t^2 + \beta_3 t + \beta_4$
  - 直接构造的设计矩阵***X***比较病态, (各列近似线性相关), 求解时会导致较大误差

$$\begin{bmatrix} t_1^3 & t_1^2 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ t_n^3 & t_n^2 & \cdots & 1 \end{bmatrix}$$

  - 改进策略：令 $s = (t - 1950)/50$
  - 得到s_i, y_i对应关系，列设计矩阵
  - 演示程序censusgui
  - 其中还包括插值模型
  - 及对数线性 $y(t) \approx K e^{\lambda t}$
  - 程序无法回答“哪个模型是最佳的？”，只能由使用者来决策



其他逼近问题

- 从范数的角度思考
  - 剩余(residual)是观测值和模型值之差 $r_i = y_i - \sum_{1}^{n} \beta_j \phi_j(t_i, \alpha)$ 基函数带的参数
  - 做拟合就是希望剩余尽可能小 即 $r = y - X(\alpha)\beta$
  - “小”的标准不同对应不同的逼近问题
  - 插值：在所有观测点上剩余为0，且m=n
  - 最小二乘：最小化剩余的2-范数(平方和) $\|r\|^2 = \sum_{1}^{m} r_i^2$
  - 加权最小二乘：某些观测值有更高权重(例如误差小的)
  - 若要最小化的是$\|r\|_w^2 = \sum_{i=1}^{m} w_i^2 r_i^2$
  - 只需将观测值y_i和设计矩阵$\boldsymbol{X}$的第i行都乘以w_i
  - 1-范数：最小化剩余的绝对值之和 $\|r\|_1 = \sum_{1}^{m} |r_i|$
  - ∞-范数：最小化绝对值最大分量 $\|r\|_\infty = \max_i |r_i|$

转化为线性规划问题

压缩稀疏信号的恢复(Compressed sensing): min解的1范数

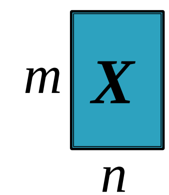
线性最小二乘问题的解法

- 概述
- Householder反射
- QR分解
- 利用QR分解解线性最小二乘

求解线性最小二乘问题

$$X\beta \approx y$$

- 概述
  - 非线性优化
  - 比非线性最小二乘、1-范数、∞-范数逼近容易求解
  - 若设计矩阵为X, 观测向量y，在Matlab中 >> beta = X\y
  - “\”处理“长条形”矩阵的算法基于对矩阵X作QR分解



- 法方程法
  - 最小化的目标 $\|r\|^2$是关于参数β_i的二次函数
  - 利用多元函数求极值方法可得取极小值的必要条件

$$X^T X\beta = X^T y$$ 若X列向量线性无关，则为n阶非奇异矩阵

  - 缺点是数值误差大，$X^T X$可能(近)奇异　　$X^T X$往往病态
  - 例: 若$|\delta| < 10^{-8}$, 则　　$\kappa(X^T X) = \underline{\kappa(X)^2}$

$$X = \begin{pmatrix} 1 & 1 \\ \delta & 0 \\ 0 & \delta \end{pmatrix} \qquad X^T X = \begin{pmatrix} 1+\delta^2 & 1 \\ 1 & 1+\delta^2 \end{pmatrix}$$ 奇异

基于矩阵QR分解的解法

- Householder反射
  - 一种重要的矩阵变换. 矩阵 $H = I - \rho u u^T$
  - u为任意非零向量, 而 $\rho = 2/\|u\|^2$
  - 或, $\boldsymbol{w}=\boldsymbol{u}/\|\boldsymbol{u}\|$, 则$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{w}\boldsymbol{w}^T$

$$\boldsymbol{H}=\begin{bmatrix} 1-2w_1^2 & -2w_1w_2 & \cdots & -2w_1w_n \\ -2w_2w_1 & 1-2w_2^2 & \cdots & -2w_2w_n \\ \vdots & \vdots & \cdots & \vdots \\ -2w_nw_1 & -2w_nw_2 & \cdots & 1-2w_n^2 \end{bmatrix}$$

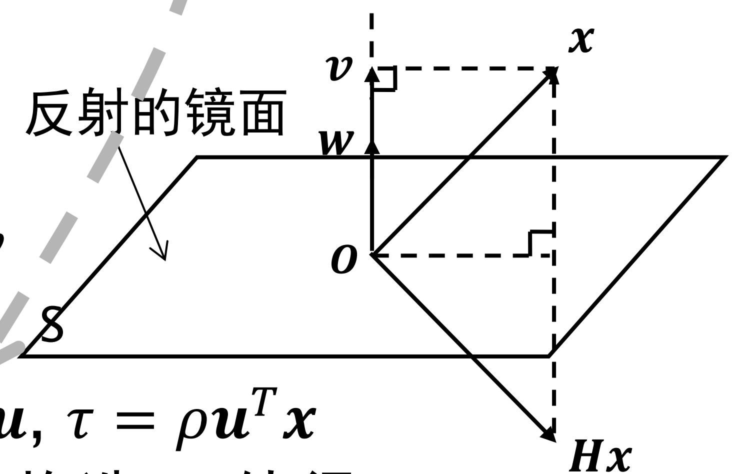
  - 性质: 对称阵、正交阵

$$H^T H = H^2 = I$$

  - $\boldsymbol{Hx}$对向量$\boldsymbol{x}$做变换

$$\boldsymbol{Hx} = (\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T)\boldsymbol{x} = \boldsymbol{x} - 2\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{x}$$

$$\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{x} = (\boldsymbol{w}^T\boldsymbol{x})\boldsymbol{w} = \boldsymbol{v} \Rightarrow \boldsymbol{Hx} = \boldsymbol{x} - 2\boldsymbol{v}$$



  - $\boldsymbol{Hx}$与$\boldsymbol{x}$的2-范数相等, 镜面反射
  - 提示1: 不显式构造$\boldsymbol{H}$, $\boldsymbol{Hx} = \boldsymbol{x} - \tau\boldsymbol{u}$, $\tau = \rho\boldsymbol{u}^T\boldsymbol{x}$
  - 提示2: 对任意$\boldsymbol{y}$, $\|\boldsymbol{x}\| = \|\boldsymbol{y}\|$, 都可构造$\boldsymbol{H}$, 使得$\boldsymbol{Hx} = \boldsymbol{y}$

关键是让u,或w与x-y同方向

基于矩阵QR分解的解法

- 用Householder变换进行消元
  - 对向量x, 可构造正交阵H使得向量Hx只有一个分量非零

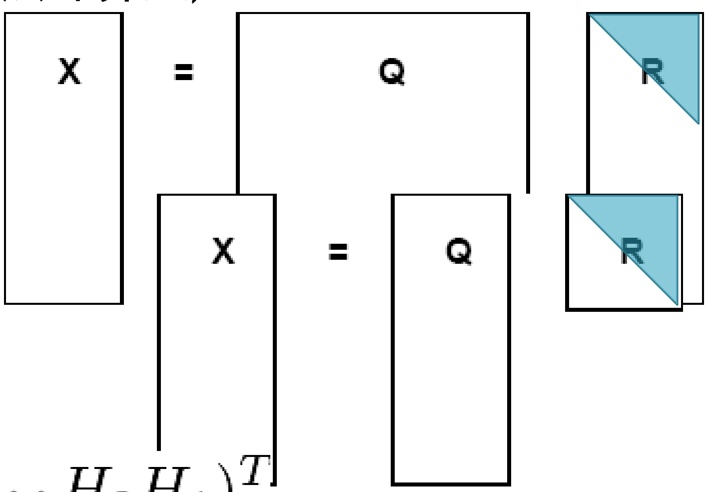
$$\begin{cases} \sigma = -\mathrm{sign}(x_k)\|\boldsymbol{x}\| \\ \boldsymbol{u} = \boldsymbol{x} - \sigma \boldsymbol{e}_k \\ \boldsymbol{H} = \boldsymbol{I} - \rho \boldsymbol{u}\boldsymbol{u}^T \ , \rho = 2/\|\boldsymbol{u}\|^2 \ = -1/(\sigma u_k) \end{cases}$$

希望$\boldsymbol{H}\boldsymbol{x} = \sigma \boldsymbol{e}_k$

可能同符号数相减! 修改后无抵消现象

  - 与向量相乘$\boldsymbol{H}\boldsymbol{y} = \boldsymbol{y} + \dfrac{\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{y}}{\sigma u_k}$ (2m次乘法计算量)
- 矩阵的QR分解
  - “六大矩阵分解”之一, (两种形式)
  - Q为正交阵, R为上三角阵
  - 用Householder变换实现: 对X逐次消元得上三角阵R

$$H_n \cdots H_2 H_1 X = R \Longrightarrow Q = (H_n \cdots H_2 H_1)^T$$



Householder变换实现QR分解

NCM演示 >> qrsteps(randi(5,5,4))

- 用Householder变换实现: $\boldsymbol{H}_n \cdots \boldsymbol{H}_2 \boldsymbol{H}_1 \boldsymbol{A} = \boldsymbol{R}$
  - 设$\boldsymbol{A} \in \mathbb{R}^{m\times n}, m \geq n$, $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$, 其中$\boldsymbol{a}_j$为$m$维向量

构造m阶反射阵$\boldsymbol{H}_1$消$\boldsymbol{A}$的第1列:

$$\boldsymbol{A}^{(2)} = \boldsymbol{H}_1 \boldsymbol{A} = \begin{bmatrix} -\sigma_1 & a_{12}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{m2}^{(2)} & \cdots & a_{mn}^{(2)} \end{bmatrix} = \begin{bmatrix} -\sigma_1 & | & & | \\ 0 & \boldsymbol{H}_1\boldsymbol{a}_2 & \cdots & \boldsymbol{H}_1\boldsymbol{a}_n \\ \vdots & | & & | \\ 0 & | & & | \end{bmatrix}$$

用$\boldsymbol{H}_2$实现$\boldsymbol{A}^{(2)}$第2列的消元, 同时不影响第1列

令$\boldsymbol{H}_2 = \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{H}_2' \end{bmatrix}$

$$\boldsymbol{H}_2 \boldsymbol{A}^{(2)} = \begin{bmatrix} -\sigma_1 & \boldsymbol{r}_1^T \\ \boldsymbol{0} & \boldsymbol{H}_2' \boldsymbol{A}'^{(2)} \end{bmatrix} = \left[\begin{array}{c:cccc} -\sigma_1 & & \boldsymbol{r}_1^T & & \\ \hdashline 0 & -\sigma_2 & | & & | \\ 0 & 0 & \boldsymbol{H}_2'\boldsymbol{a}_2'^{(2)} & \cdots & \boldsymbol{H}_2'\boldsymbol{a}_{n-1}'^{(2)} \\ \vdots & \vdots & | & & | \\ 0 & 0 & | & & | \end{array}\right]$$

记$\boldsymbol{A}^{(2)} = \begin{bmatrix} -\sigma_1 & \boldsymbol{r}_1^T \\ \boldsymbol{0} & \boldsymbol{A}'^{(2)} \end{bmatrix}$　　构造$\boldsymbol{H}_2'$

$\boldsymbol{H}_2$也是Householder阵(其向量$\boldsymbol{u}$的第一个分量为0). 后续$\boldsymbol{H}_j$类似构造

基于矩阵QR分解的解法

- 利用QR分解求解
  - 求解线性最小二乘问题，不需显式地算出Q
  - Householder反射也作用于$X\beta \approx y$ 的右端项
  - $R\beta \approx z$，其中$H_n \cdots H_2 H_1 y = z$　　　正交变换不改变2-范数！
  - 求$\boldsymbol{\beta}$，使得$\|\boldsymbol{z} - \boldsymbol{R\beta}\|$达到最小值

$$\text{设} \boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 \\ \boldsymbol{O} \end{bmatrix}^{\text{n行}}, \boldsymbol{z} = \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix}^{\text{n行}}, \boldsymbol{z} - \boldsymbol{R\beta} = \begin{bmatrix} \boldsymbol{z}_1 - \boldsymbol{R}_1\boldsymbol{\beta} \\ \boldsymbol{z}_2 \end{bmatrix}$$

$$\|\boldsymbol{z} - \boldsymbol{R\beta}\|_2^2 = \|\boldsymbol{z}_1 - \boldsymbol{R}_1\boldsymbol{\beta}\|_2^2 + \|\boldsymbol{z}_2\|_2^2 \geq \|\boldsymbol{z}_2\|_2^2$$

取等号条件: $\boldsymbol{R}_1\boldsymbol{\beta} = \boldsymbol{z}_1$

- 例: 美国人口预测 (pp.127)
  - $y(s) \approx \beta_1 s^2 + \beta_2 s + \beta_3$
  - 拟合1950开始后6个点

结果与beta =X\y一样

```
>> s=((1950:10:2000)'-1950)/50
>> X=[s.*s s ones(size(s))]
>> [R, z] = qrsteps(X, y)
>> beta = R(1:3, 1:3)\ z(1:3)
>> p2010= polyval(beta, 1.2)
```

伪逆与不满秩情况

- 矩阵的伪逆
  设计矩阵不满秩?
  基本解与最小二乘解
  有关Matlab命令

伪逆与不满秩情况

- 矩阵的伪逆 (pesudo inverse)
  - 矩阵的Frobenius范数

$$\|A\|_F = \left(\sum_i \sum_j a_{i,j}^2\right)^{1/2}$$

```
>> norm(A, 'fro');
```

```
norm(A, 'fro') == norm(A( : ))
```

  - Moore–Penrose伪逆将非奇异阵的逆推广到一般矩阵
    矩阵A的伪逆记为Z= A^+

```
>> Z= pinv(A);
```

  - 若A列满秩，$A^+ = (A^T A)^{-1} A^T \Longrightarrow Ax \approx y$的解为 $x = A^+ y$
  - A^+具有A^{-1}的一些性质，例如:它是"左逆" $A^+ A = I$
  - 但不是"右逆" $A A^+ \neq I$
  - 若列不满秩，有另外的定义（矩阵的奇异值分解）

扩展矩阵的条件数: $\mathrm{cond}(A) = \|A\| \cdot \|A^+\|$

伪逆与不满秩情况

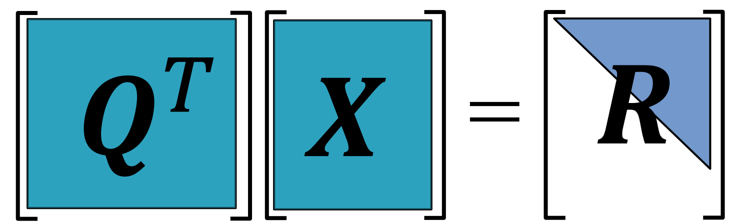
- 矩阵的伪逆 (pesudo inverse)
  - 定义一般矩阵伪逆的方式之一: 从“逼近右逆”的角度
  - 伪逆A^+是最逼近右逆条件的矩阵Z:
  - 且还是满足最小化条件 $\|AA^+ - I\|_F = \min_Z \|AZ - I\|_F$ 的Z矩阵中F–范数最小的那个
  - 例: 标量x的伪逆(近似右逆)
  - 要最小化|xz–1|. 若x≠0, 伪逆是1/x, 是唯一的
  - 若x=0, |xz–1|=1, 伪逆取绝对值最小的那个z, 即0
- 设计矩阵不满秩 $X\beta \approx y$
  - 问题的建模或误差导致
  - 存在非零向量η, 使得$X\eta = \mathbf{0}$ (因此最小二乘解不唯一)

思考：行向量a的伪逆是?

伪逆与不满秩情况

$$X\beta \approx y$$

- 设计矩阵X不满秩的情况 (rank deficiency)
  - 还能用QR分解的方法求最小二乘解吗?
  - Householder变换过程不会中断
  - 得到**R**的列秩与**X**的一样, 因此**R**的对角元有0
  - $\boldsymbol{R}_1\boldsymbol{\beta}=\boldsymbol{z}_1$有无穷多解　怎么求一个特殊性质的解？
  - 设计矩阵X近似不满秩的例子



$$\boldsymbol{X}=\begin{bmatrix}0.641 & 0.242\\ 0.321 & 0.121\\ 0.962 & 0.363\end{bmatrix} \xrightarrow{\text{QR分解}} \boldsymbol{R}=\begin{bmatrix}-1.1997 & -0.4527\\ 0 & 0.0002\\ 0 & 0\end{bmatrix}$$

近似0
矩阵R_1近奇异

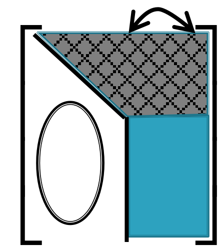
  - “数值秩”为1；求解最小二乘解时很敏感(误差大)

对敏感性问题，怎么求数值秩？

列主元QR分解

- 做正交变换时适当地交换列

$$H_n\cdots H_1X=\begin{bmatrix}R\\O\end{bmatrix} \Longrightarrow H_k\cdots H_1XP_1\cdots P_k=\begin{bmatrix}R & S\\O & O\end{bmatrix}$$



  - 在未消去的子矩阵中找2-范数最大的列，做列交换
  - 若2-范数为0, 则后续列均为0. 这发生在第k步变换后, k=rank($\boldsymbol{X}$) (正交变换不改变列秩)
  - 求最小二乘问题$X\beta\approx y$的基本解 $\quad Q^TX\beta\approx Q^Ty$
  - 列主元QR分解 $Q^TX=\begin{bmatrix}R & S\\O & O\end{bmatrix}P^T$ 变换后的右端项 $Q^Ty=\begin{bmatrix}z_1\\z_2\end{bmatrix}$
  - 等价于求解

$$\begin{bmatrix}R & S\\O & O\end{bmatrix}P^T\beta\approx\begin{bmatrix}z_1\\z_2\end{bmatrix}$$ 设$c=P^T\beta$, 可取 $c=\begin{bmatrix}R^{-1}z_1\\\mathbf{0}\end{bmatrix}$ $\Rightarrow$ $\beta=Pc$

```
>> beta=X\y;
```

基本解：非零元数目不超过X的秩

同时算出X的数值秩

用SVD算最小范数解

$$X\beta \approx y$$

- 矩阵X的奇异值分解(singular value decomposition)
  - 任意m×n矩阵可分解为 $X = U\Sigma V^T$
  - 其中U, V为正交阵，Σ为对角元≥0 (降序排列)的对角阵
  - 线性最小二乘问题：$\min\|y - X\beta\|_2 \Longleftrightarrow \min\|y - U\Sigma V^T\beta\|_2$
  - 正交阵不改变2-范数 $\Longleftrightarrow \min\|U^T y - \Sigma V^T\beta\|_2$
  - 设$c = V^T\beta$, 最小化目标变为$\|U^T y - \Sigma c\|_2$
  - 一个解是$c = \begin{bmatrix} u_1^T y/\sigma_1 \\ \vdots \\ u_r^T y/\sigma_r \\ \mathbf{0} \end{bmatrix}$，r为非零奇异值个数($X$的秩)
    所有可能的解中2–范数最小的
    $\Longrightarrow \beta = Vc$ 称为"最小范数解"
  - 按最一般的伪逆定义，最小范数解就是$\beta = X^+ y$
    求最小范数解比求基本解计算量大些

```
>> beta=pinv(X)*y;
```

Matlab有关命令的说明

- qr命令
  - 能处理稠密或“稀疏”矩阵, 允许m<n (列不满秩)
  - [Q,R] = qr(A); 完整的QR分解, Householder, Givens
  - [Q,R] = qr(A,0); 简化形式的QR分解，若m<n同上
  - [Q,R,E] = qr(A); E为排列阵，AE=QR，R对角元递减
  - [Q,R,E] = qr(A,0); E为排列向量，A(:, E)=QR
  - R = qr(A); A稠密时，返回矩阵上三角部分为R; A稀疏时，R就是上三角阵，且$R^TR=A^TA$
- \命令
  - 求解线性最小二乘问题，列重排的Householder变换
  - 列不满秩会报warning，得到一个基本解
- polyfit命令
  - 多项式最佳平方逼近：p = polyfit(x,y,n)

非线性最小二乘

- 可分离最小二乘问题

可分离最小二乘问题

例如, $y \approx \beta_1 e^{-\lambda_1 t} + \beta_2 e^{-\lambda_2 t}$

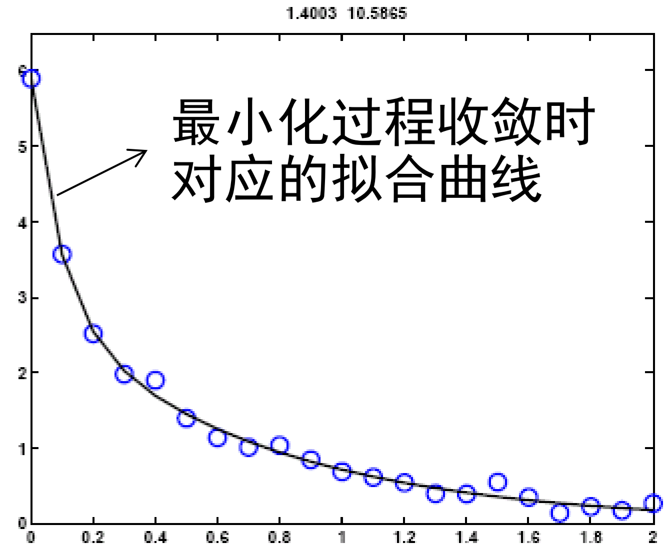
- 问题定义与求解思路
  - 拟合公式中含非线性参数与线性参数
  - 若已知非线性参数($\lambda_{1/2}$)的值, 则用\可求线性参数值($\beta_{1/2}$)
  - 需最小化的是剩余的2-范数 $\sum_{i=1}^{m}\left[y_i - (\beta_1 e^{-\lambda_1 t_i} + \beta_2 e^{-\lambda_2 t_i})\right]^2$
  - ~ 非线性参数($\lambda_{1/2}$)的函数 ⇒ 求函数最小值问题
  - 优化求解器+线性最小二乘求解器

比全看成非线性参数做优化的方法更有效/鲁棒

- 求多元实函数的最小值
  - fminserach, (无约束非线性优化, 直接搜索法)
  - [x,fval] = fminsearch(fun,x0,options) 后面加给fun的参数
  - 对于例子, 定义fun的输入为λ_1, λ_2, 返回值为最小剩余

可分离最小二乘问题

- 例子 $y \approx \beta_1 e^{-\lambda_1 t} + \beta_2 e^{-\lambda_2 t}$ (放射性物质衰变的观测)
  - 数据: t = (0:.1:2) '
  - y = [5.8955 3.5639 2.5173 1.9790
  - ..., 0.1370 0.2211 0.1704 0.2636]';
  - λ_1, λ_2的初值为: [3 6]'



- 求解过程与演示
  - 线性最小二乘的解定义为函数:
    res = expfitfun(lambda, t, y)　将观测数据设为函数参数
  - 非线性最小化: fminsearch(@expfitfun,lambda0,[],t,y)
  - 程序细节与动态演示程序
    expfitdemo

```
>> edit expfitfun.m;
>> edit expfitdemo.m;
```

非线性最小二乘及最优化问题

- fminsearch
  - 适合仅含几个变量的小规模优化问题
- Matlab优化工具箱(Toolbox)
  - 带约束极小化fmincon
  - 无约束极小化fminunc
  - 最小二乘问题也属于最优化问题
  - 特别针对最小二乘问题的函数: lsqnonlin, lsqcurvefit
- 曲线拟合工具箱
  - 方便解决曲线、曲面拟合问题的图形交互界面