最优化方法 Optimization

第五讲

最小二乘法

最小二乘法

- 线性最小二乘问题
- 非线性最小二乘问题

最小二乘 (Least Squares)

起源: 19世纪初由德国数学家Gauss在预测行星运行轨道时提出。

发现

预测轨道

Giuseppe Piazzi
(1746—1826)

Ceres（谷神星）
(01/01/1801)

J.C.F.Gauss
(1777—1855)

Ceres Solver: Google,2010,open source C++ library, large-scale nonlinear least squares solver, (http://www.ceres-solver.org/; https://anaconda.org/conda-forge/ceres-solver; Github)

评价: "最小二乘法是十九世纪统计学的主题曲. 从很多方面来看，它之于统计学就相当于十八世纪的微积分之于数学."– Stephen M Stigler 《The History of Statistics 》

最小二乘法

基本思想： 基于观测数据与模型数据之间差的平方和最小来估计模型参数的数据拟合方法。

数据拟合：

输入数据	观测数据	模型数据
t_1	y_1	$\phi(x,t_1)$
t_2	y_2	$\phi(x,t_2)$
$\vdots$	$\vdots$	$\vdots$
t_m	y_m	$\phi(x,t_m)$

- 样本量 $\boldsymbol{m}$;待估参数个数$\boldsymbol{n}$; 大样本情形: $\boldsymbol{m}>\boldsymbol{n}$; 小样本情形: $\boldsymbol{m}\ll\boldsymbol{n}$

最小二乘模型： $$\min_{x\in R^n}\sum_{i=1}^{m}\left(y_i-\phi\left(x,t_i\right)\right)^2$$

北京交通大學

最小二乘

$$\min_{x} \sum_{i=1}^{m} \left(y_i - \phi(x,t_i)\right)^2$$

残量函数

$$r_i(x) = y_i - \phi(x,t_i)$$

$$r(x) = \left(r_1(x)\ \ r_2(x)\ \ \cdots\ \ r_i(x)\right)^T$$

$$\min_{x} \sum_{i=1}^{m} r_i^2(x) = r(x)^T r(x)$$

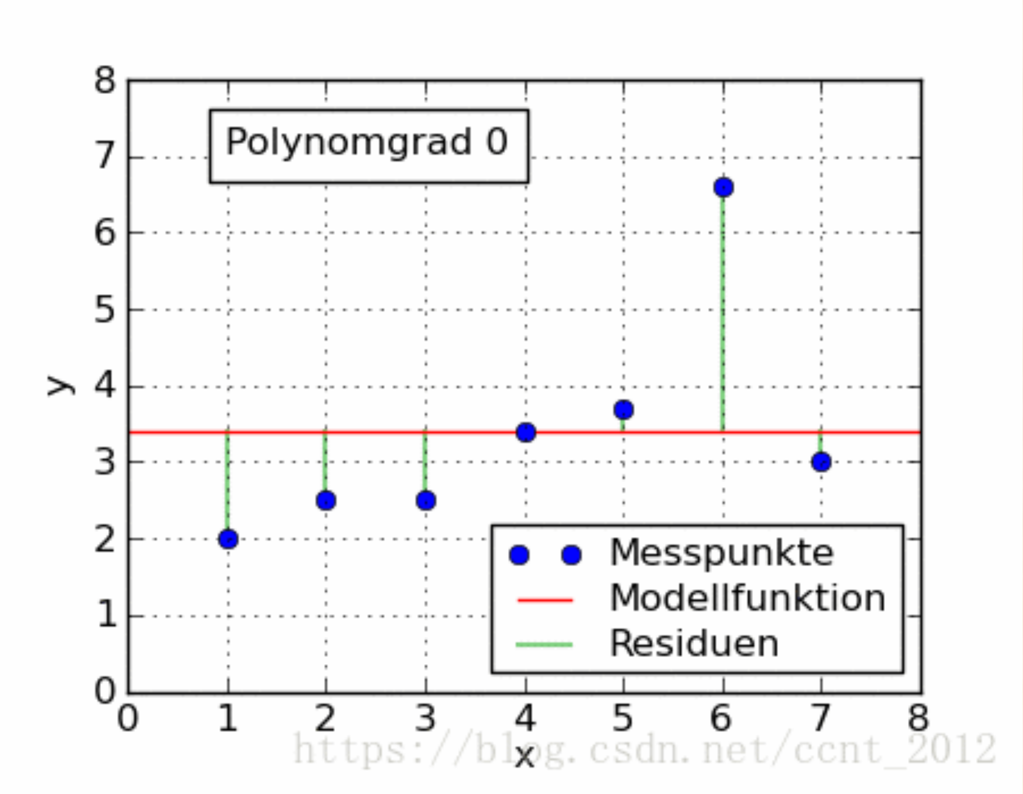
$$\min_{x} \frac{1}{2} r(x)^T r(x) = \frac{1}{2}\|r(x)\|^2$$

$$\min_{x} \frac{1}{2}\|b - Ax\|^2$$

线性最小二乘

$$\min_{x} \frac{1}{2}\|r(x)\|^2$$

非线性最小二乘

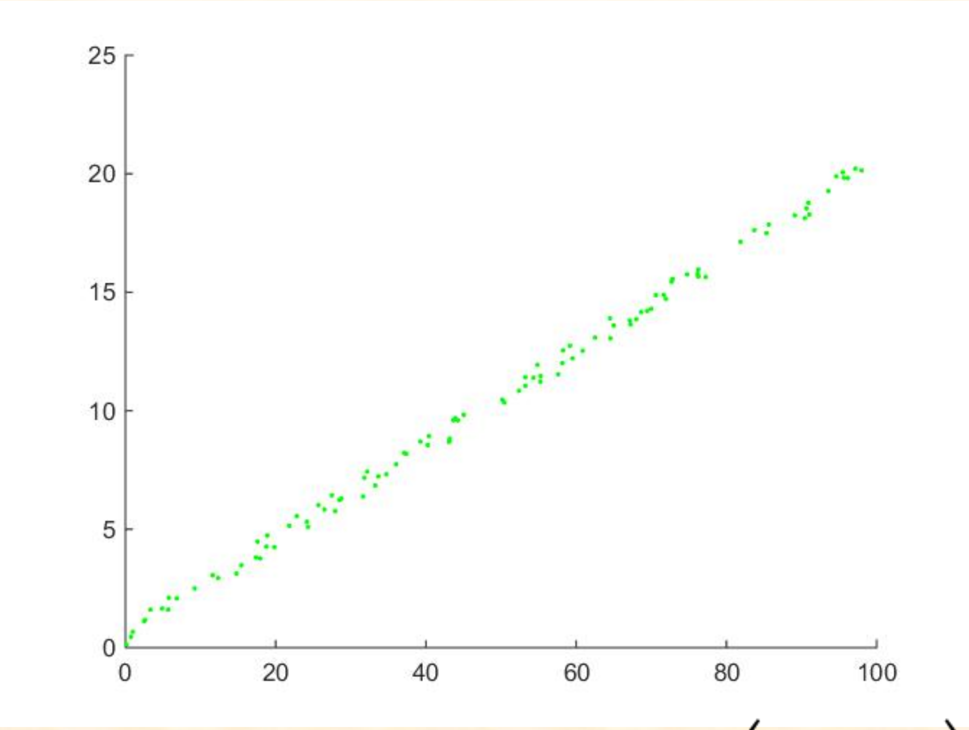


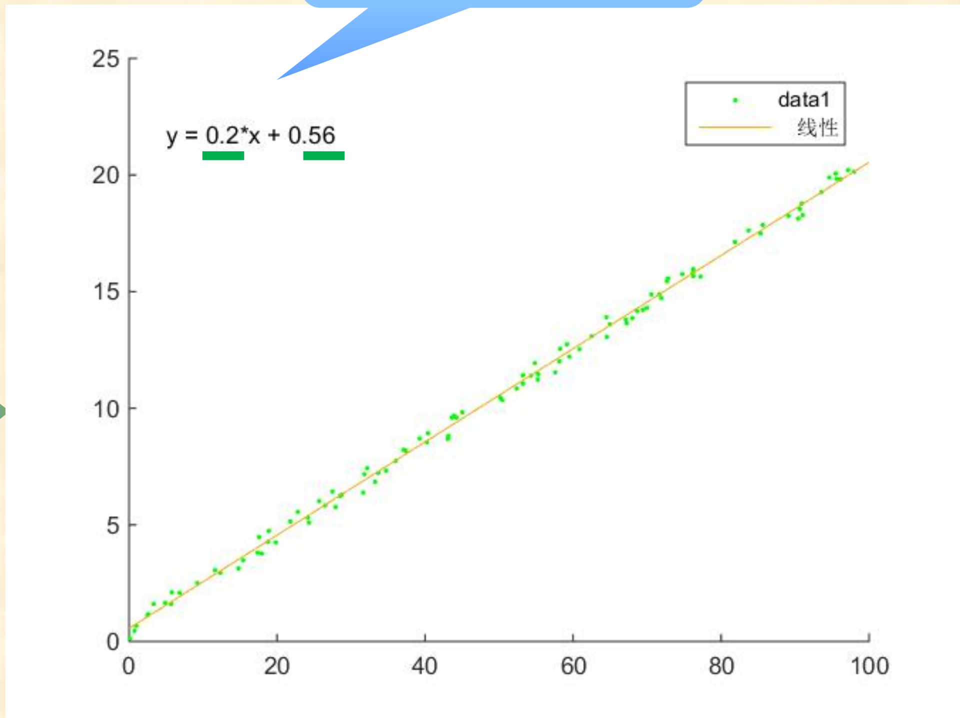
同一组数据(蓝色点),使用不同类型的拟合函数$\boldsymbol{\phi}$, 通过最小二乘法（使绿色线段长度的平方和最小）可以得到不同的拟合曲线（红色曲线）

线性最小二乘问题

应 用

线性拟合





输入数据：测量矩阵$\boldsymbol{A}:=\begin{pmatrix} x_1 & 1 \\ \dots & \\ x_{100} & 1 \end{pmatrix}$；观测向量 $\boldsymbol{b}:=(y_1,\dots,y_{100})^T$

假设模型函数: $y=kx+b$, 待估系数向量: $\boldsymbol{x}:=(k,b)^T$

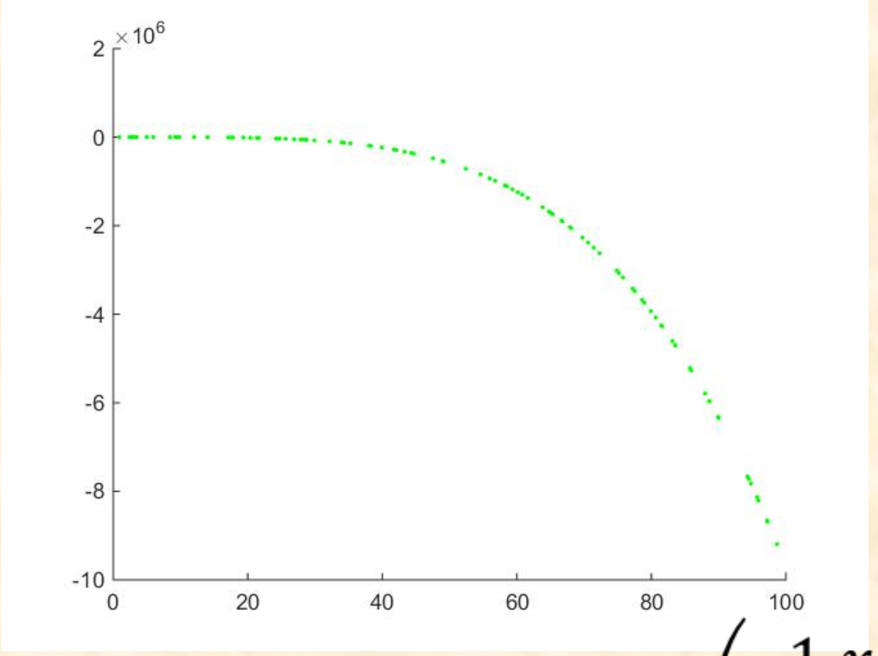
$$\begin{pmatrix} 0.2 \\ 0.56 \end{pmatrix}=\arg\min_{k,b}\frac{1}{2}\sum_{i=1}^{100}\left(y_i-(kx_i+b)\right)^2=:\arg\min_{x\in\boldsymbol{R}^2}\frac{1}{2}\|\boldsymbol{b}-A\boldsymbol{x}\|^2$$

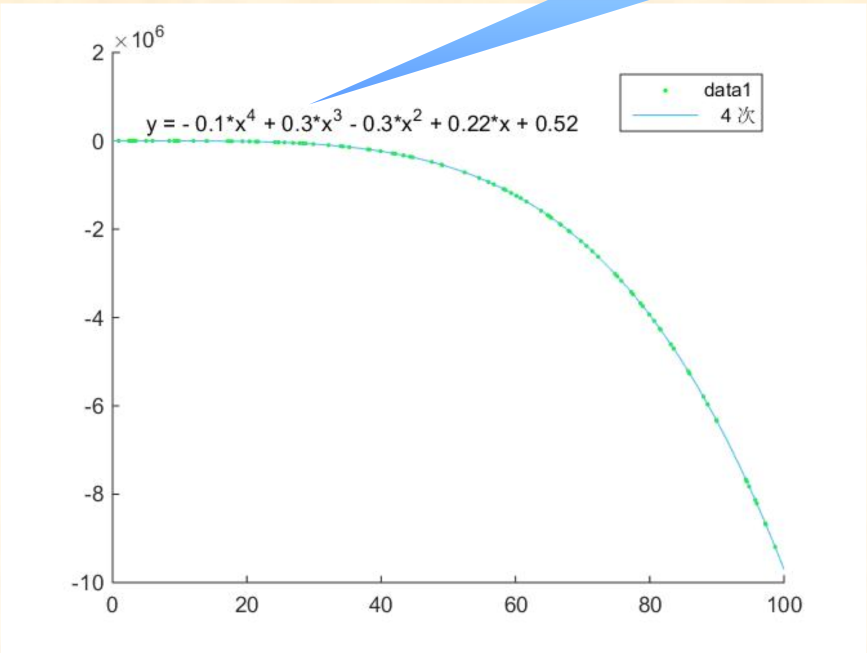
线性最小二乘问题

应 用

多项式拟合

多项式回归函数





输入数据：测量矩阵$\boldsymbol{A}:=\begin{pmatrix} 1\ x_1\ \dots\ x_1{}^4 \\ \cdots \\ 1\ x_{100}\ \dots\ x_{100}^4 \end{pmatrix}$；观测向量 $\boldsymbol{b}:=(y_1,\dots,y_{100})^{\boldsymbol{T}}$

假设模型函数:$y=a_0+a_1x+\cdots+a_4x^4$，待估系数向量：$\boldsymbol{x}:=(a_0,\dots,a_4)^{\boldsymbol{T}}$

$$(0.52,0.22,-0.3,0.3,-0.1)^T=\arg\min_{a_0,\dots,a_4}\frac{1}{2}\sum_{i=1}^{100}\left(y_i-\sum_{j=0}^{4}a_jx^j\right)^2=:\arg\min\frac{1}{2}\|\boldsymbol{b}-A\boldsymbol{x}\|^2$$

线性最小二乘问题

- 残量函数的线性与否，不是由模型函数的线性与否所决定，而是由待估参数与已知数据之间是否为线性关系所决定.

线性最小二乘问题的求解

模型分析

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

无约束凸二次规划

一阶最优性条件：

$$\boxed{A^T(b-Ax)=0}$$

充要条件

线性最小二乘问题的求解

解集分析

$$A^T(b-Ax)=0 \Rightarrow A^TAx=A^Tb$$

线性方程组有解,因为
$R(A^TA)=R(A^T)$

定理6.1.1 对任意$A\in \mathbb{R}^{m\times n}$和$b\in \mathbb{R}^m$,线性最小二乘问题

$$\min_x \frac{1}{2}\|b-Ax\|^2$$ 存在全局最优解，而它有唯一最优解的充分必要条件是矩阵A列满秩。

课后练习1

若$f(x)$是$\mathbb{R}^n$上的二次函数 $f(x)=\frac{1}{2}x^TAx+b^Tx+c$,其中$A\in\mathbb{R}^{n\times n}$对称矩阵，$b\in\mathbb{R}^n$，$c\in\mathbb{R}$常数，此时,优化问题 $\min_{x\in\mathbb{R}^n} f(x)$是否一定有全局最优解?

线性最小二乘问题的求解

A列满秩时MATLAB求解

$$A^T A x = A^T b$$

直接求解

$MATLAB: x = A \backslash b;$

Cholesky分解： $A^T A = LL^T$

$$x = L^{-T} L^{-1} A^T b$$

$MATLAB:$

$n = 10000;$

$m = 3 * n;$

$A = rand(m,n);$

$b = rand(m,1);$

$L = chol(A'*A,'lower');$

$z = A'*b;$

$z = L \backslash z;$

$x = L' \backslash z;$

QR分解： $A^T A = QR$

$$x = R^{-1} Q^T A^T b$$

$MATLAB:$

$n = 10000;$

$m = 3 * n;$

$A = rand(m,n);$

$b = rand(m,1);$

$[Q,R] = qr(A'*A);$

$z = A'*b;$

$z = Q'*z;$

$x = R \backslash z;$

PCG迭代求解

$MATLAB:$

$x = pcg(A'*A, A'*b, 1e-10, 200);$

线性最小二乘问题的求解

A列亏秩时的MATLAB求解

$$A^T A x = A^T b$$

通解：$x = A^+ b + \left(I - A^+ A\right)u, u \in R^n$

特解：$x = A^+ b$

$MATLAB:$

$G = pinv(A);$

$x = G * b$

矩阵的广义逆：设$A \in R^{m\times n}$,若矩阵$G \in R^{n\times m}$满足

$AGA = A, GAG = G,$

$(AG)^T = AG, (GA)^T = GA,$

则称G为A的广义逆，记为A^+

$(Moore, 1920; Penrose, 1955)$

若$rank(A) = n$,则$A^+ = \left(A^T A\right)^{-1} A^T$

- 无穷多个解！对于具体的实际问题，选择哪个解呢？--约束正则化!

✓ **"Thus there is a need to constrain, or regularize the estimation process."**

引自Trevor Hastie,Robert Tibshirani,Martin Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press,2015

非线性最小二乘问题

非线性最小二乘应用

指数函数拟合

例 在化学工业的可靠性研究中，对象是某种产品 A. 在制造时单位产品中必须含有 0.50 的有效氯气. 已知产品中的氯气随着时间增加而减少. 在产品到达用户之前的最初 8 周内，氯气含量衰减到 0.49. 但由于随后出现了许多无法控制的因素 (如库房环境、处理设备等), 因而在后 8 周理论的计算对有效氯气的进一步预报是不可靠的. 为了有利于管理需要决定 (1) 库存产品何时应该报废？(2) 何时应该更换存货？在一段时间中观测若干盒产品得到的数据如表 所示. 假定非线性模型

$$\boxed{Y = \alpha + (0.49 - \alpha)\exp(-\beta(X - 8)) + \varepsilon}$$

需要确定的拟合参数

非线性最小二乘

指数函数拟合

X

Y

表　单位产品中有效氯气的百分数

序号	生产后的时间	有效氯气	序号	生产后的时间	有效氯气
1	8	0.49	23	22	0.41
2	8	0.49	24	22	0.40
3	10	0.48	25	24	0.42
4	10	0.47	26	24	0.40
5	10	0.48	27	24	0.40
6	10	0.47	28	26	0.41
7	12	0.46	29	26	0.40
8	12	0.46	30	26	0.41
9	12	0.45	31	28	0.41
10	12	0.43	32	28	0.40
11	14	0.45	33	30	0.40
12	14	0.43	34	30	0.40
13	14	0.43	35	30	0.38
14	16	0.44	36	32	0.41
15	16	0.43	37	32	0.40
16	16	0.43	38	34	0.40
17	18	0.46	39	36	0.41
18	18	0.45	40	36	0.38
19	20	0.42	41	38	0.40
20	20	0.42	42	38	0.40
21	20	0.43	43	40	0.39
22	22	0.41	44	42	0.39

$$(\alpha^*,\beta^*)^T$$

$$=\arg\min_{\alpha,\beta}\frac{1}{2}\sum_{i=1}^{44}\left(Y_i-(\alpha+(0.49-\alpha)\exp(-\beta(X_i-8)))\right)^2$$

$$=:\arg\min_{x\in R^2}\frac{1}{2}\|r(x)\|^2 \qquad \text{其中 } x=(\alpha\ \ \beta)^T$$

非线性最小二乘问题

非线性最小二乘问题的求解

模型分析

$$\min_x \frac{1}{2}\left\|r(x)\right\|^2$$

无约束优化问题

无约束的优化方法均可以用于求解这一模型：

(1) 最速下降

(2) 牛顿法

(3) 共轭梯度法

(4) 拟牛顿法

非线性最小二乘问题的求解

模型分析

$$\min_x f(x) := \frac{1}{2}\|r(x)\|^2$$

- **梯度：**

Jacobi矩阵： $J(x) := \left(\nabla r_1(x)^T, \ldots, \nabla r_m(x)^T\right)^T \in R^{m\times n}$

$$\nabla f(x) = g(x) = \nabla\left(\frac{1}{2}\sum_{i=1}^{m} r_i^2(x)\right) = \sum_{i=1}^{m} r_i(x)\nabla r_i(x) =: J(x)^T r(x)$$

- **Hesse阵：**

$$\nabla^2 f(x) = G(x) = \nabla\left(\sum_{i=1}^{m} r_i(x)\nabla r_i(x)\right) = \sum_{i=1}^{m}\nabla r_i(x)\nabla r_i(x)^T + \sum_{i=1}^{m} r_i(x)\nabla^2 r_i(x)$$

$$=: J(x)^T J(x) + S(x)$$

非线性最小二乘问题的求解

牛顿方法

$$\min_{x} f(x) := \frac{1}{2}\|r(x)\|^2$$

➤ 牛顿方向：

$$d_k^N = -G(x_k)^{-1} g(x_k) = -G_k^{-1} g_k = -\left(J_k^T J_k + \boxed{S_k}\right)^{-1} J_k^T r_k$$

$$S_k = \sum_{i=1}^{m} r_i(x)\nabla^2 r_i(x)$$

- 二阶导的计算量较大
- 残量函数的最优值很小或者线性度较高时，S_k很小

忽略S_k

Gauss–Newton 方法

- 迭代方向：$d_k^{GN} = -\left(J_k^T J_k\right)^{-1} J_k^T r_k$

线性最小二乘问题

$$d_k^{GN} = \arg\min_d \frac{1}{2}\left\|\underline{r_k + J_k d}\right\|^2$$

$$r(x_k + d) \approx r_k + J_k d$$

- 迭代公式：$x_{k+1} = x_k + d_k^{GN}$

➤ 若矩阵J_k列满秩，则为下降方向

$$g_k{}^T d_k^{GN} = -\left(J_k^T r_k\right)^T \left(J_k^T J_k\right)^{-1} J_k^T r_k < 0$$

Gauss–Newton 方法

Wolfe步长规则下的全局收敛性

定理6.2.1 设最小二乘问题的水平集$L(x_0)$有界，$r(x)$以及$J(x)$在水平集上Lipschitz连续，且$J(x)$在$L(x_0)$上满足正则性条件，即存在$\gamma > 0$,使得对于任意的$x \in L(x_0)$，有

$$\boxed{\|J(x)y\| \geq \gamma \|y\|, \quad \forall y \in R^n}$$

$J(x)$列满秩从而为下降方向

则Wolfe步长规则下的Gauss-Newton算法产生的点列$\{x_k\}$满足$\lim_{k\to\infty} J_k^T r_k = 0$,从而算法产生的点列$\{x_k\}$的任一聚点是问题的稳定点。

Gauss–Newton 方法

单位步长规则下的收敛速度

定理6.2.2 设单位步长规则下的Gauss-Newton算法产生的点列 $\{x_k\}$ 收敛到问题的局部最优解 x^*，且 $J(x^*)^T J(x^*)$ 正定，则当 $J(x)^T J(x), S(x), \left(J(x)^T J(x)\right)^{-1}$ 在 x^* 的邻域内Lipschitz连续时, 对充分大的 k 有

$$\|x_{k+1} - x^*\| \le \left\|\left(J(x^*)^T J(x^*)\right)^{-1}\right\| \boxed{\|S(x^*)\|} \; \|x_k - x^*\| + O\left(\|x_k - x^*\|^2 \right)$$

若 $r(x)$ 的线性度较高或者最优值接近于0，则 $S(x)$ 约等于0，算法收敛速度快。

Gauss-Newton 方法

- 条件：$J(x)$列满秩

若列满秩条件不成立呢？

$J(x)^T J(x)$可逆！

$d^{GN} = -\left(J(x)^T J(x)\right)^{-1} J(x)^T r(x)$有意义

Levenberg–Marquardt 方法

起源:

为克服Gauss-Newton算法过程中要求$J(x)$列满秩这一局限性，由Levenberg（1944）提出，后来Marquardt (1963)重新提出，并给出理论探讨，得到Levenberg-Marquardt（LM）方法。

基本思想: 拟牛顿+ 信赖域策略

Levenberg–Marquardt 方法

- 迭代方向：$d_k^{LM} = -\left(J_k^T J_k + \mu_k I\right)^{-1} J_k^T r_k$

$\Updownarrow$

2-范数正则的线性最小二乘

$$d_k^{LM} = \arg\min_d \frac{1}{2}\left\|r_k + J_k d\right\|^2 + \underline{\frac{1}{2}\mu_k \left\|d\right\|^2}$$

正则项

$\Updownarrow$

$\Delta = \left\|d_k^{LM}\right\|$

信赖域子问题

$$d_k^{LM} = \arg\min_d \left\{\left\|r_k + J_k d\right\|^2 \,\middle|\, d \in R^n, \left\|d\right\| \le \Delta\right\}$$

1. 折线法
2. 二维子空间法
3. 精确法

信赖域策略：其中 μ_k 为约束优化的最优性条件中的Lagrange乘子

Levenberg–Marquardt 方法

- LM方向：

$$d_k^{LM} = -\left(J_k^T J_k + \mu_k I\right)^{-1} J_k^T r_k$$

$$\Downarrow$$

$$d(\mu) = -\left(J^T J + \mu I\right)^{-1} J^T r$$

$d(\mu)$的性质

- **性质6.2.1:** 信赖域半径单调不增

$$d(\mu) = -\left(J^T J + \mu I\right)^{-1} J^T r$$

$\|d(\mu)\|$关于$\mu > 0$单调不增， 且当$\mu \to \infty$时,$\|d(\mu)\| \to 0$.

证明：

$$(J(x)^{\mathrm{T}} J(x) + \mu I) d(\mu) = -J(x)^{\mathrm{T}} r(x).$$

$$d(\mu) + (J(x)^{\mathrm{T}} J(x) + \mu I) \frac{\partial d(\mu)}{\partial \mu} = 0.$$

$$\frac{\partial d(\mu)}{\partial \mu} = -(J(x)^{\mathrm{T}} J(x) + \mu I)^{-1} d(\mu).$$

$$\frac{\partial \| d(\mu) \|^2}{\partial \mu} = 2 d(\mu)^{\mathrm{T}} \frac{\partial d(\mu)}{\partial \mu}.$$

$$\frac{\partial \| d(\mu) \|^2}{\partial \mu} = -2 d(\mu)^{\mathrm{T}} (J(x)^{\mathrm{T}} J(x) + \mu I)^{-1} d(\mu).$$

$$\frac{\partial \| d(\mu) \|}{\partial \mu} = -\frac{1}{\| d(\mu) \|} d(\mu)^{\mathrm{T}} (J(x)^{\mathrm{T}} J(x) + \mu I)^{-1} d(\mu) < 0.$$

$d(\mu)$的性质

性质6.2.2： $d(\mu)$与$-g(x)$的夹角θ关于$\mu>0$单调不增.

证明：

$$\frac{\partial\cos\theta}{\partial\mu}=\frac{\partial}{\partial u}\left(\frac{-g(x)^{\mathrm{T}}d(\mu)}{\|d(\mu)\|\|g(x)\|}\right)$$

$$=\frac{-g(x)^{\mathrm{T}}\frac{\partial d(\mu)}{\partial\mu}\|d(\mu)\|\|g(x)\|+g(x)^{\mathrm{T}}d(\mu)\|g(x)\|\frac{\partial\|d(\mu)\|}{\partial\mu}}{\|d(\mu)\|^2\|g(x)\|^2}.$$

分子可转化为：

$$\begin{aligned}
&\|d(\mu)\|\|g(x)\|g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-1}d(\mu)\\
&-g(x)^{\mathrm{T}}d(\mu)\|g(x)\|d(\mu)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-1}d(\mu)/\|d(\mu)\|\\
=&-\|g(x)\|\|d(\mu)\|g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-2}g(x)\\
&+\|g(x)\|g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-1}g(x)g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-3}g(x)/\|d(\mu)\|\\
=&\frac{\|g(x)\|}{\|d(\mu)\|}\Big(-g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-2}g(x)g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-2}g(x)\\
&+g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-1}g(x)g(x)^{\mathrm{T}}\left(J(x)^{\mathrm{T}}J(x)+\mu I\right)^{-3}g(x)\Big).
\end{aligned}$$

$$Q^{\mathrm{T}} J(x)^{\mathrm{T}} J(x) Q = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

记$v_i = (Q^{\mathrm{T}} g(x))_i,$

$$g(x)^{\mathrm{T}}(J(x)^{\mathrm{T}}J(x)+\mu I)^{-1}g(x)=\sum_{i=1}^{n}\frac{1}{\lambda_i+\mu}v_i^2,$$
$$g(x)^{\mathrm{T}}(J(x)^{\mathrm{T}}J(x)+\mu I)^{-2}g(x)=\sum_{i=1}^{n}\frac{1}{(\lambda_i+\mu)^2}v_i^2,$$
$$g(x)^{\mathrm{T}}(J(x)^{\mathrm{T}}J(x)+\mu I)^{-3}g(x)=\sum_{i=1}^{n}\frac{1}{(\lambda_i+\mu)^3}v_i^2.$$

$$\frac{\|g(x)\|}{\|d(\mu)\|}\Bigg(-\Big(\sum_{i=1}^{n}\frac{v_i^2}{(\lambda_i+\mu)^2}\Big)^2+\Big(\sum_{i=1}^{n}\frac{v_i^2}{\lambda_i+\mu}\Big)\Big(\sum_{i=1}^{n}\frac{v_i^2}{(\lambda_i+\mu)^3}\Big)\Bigg)$$
$$=\frac{\|g(x)\|}{\|d(\mu)\|}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big(\frac{-v_i^2v_j^2}{(\lambda_i+\mu)^2(\lambda_j+\mu)^2}+\frac{v_i^2v_j^2}{(\lambda_i+\mu)(\lambda_j+\mu)^3}\Big)$$
$$=\frac{\|g(x)\|}{2\|d(\mu)\|}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{v_i^2v_j^2}{(\lambda_i+\mu)^3(\lambda_j+\mu)^3}\Big(-2(\lambda_i+\mu)(\lambda_j+\mu)$$
$$+(\lambda_i+\mu)^2+(\lambda_j+\mu)^2\Big)\geqslant 0.$$

$\mu\uparrow, \cos\theta\uparrow, \theta\downarrow$

$d(\mu)$的性质

性质6.2.3：

设$x \in R^n$为非线性最小二乘问题的非稳定点，则对任意的$\rho \in (0,1)$，存在$\bar{\mu} > 0$，使对于任意的$\mu \geq \bar{\mu}$，$\left\langle \frac{-g(x)}{\|g(x)\|}, \frac{d(\mu)}{\|d(\mu)\|} \right\rangle \geq \rho$

$\mu \to \infty$，$d(\mu) \to -g$

性质6.2.3的证明

证明 对任意发散到∞的正数列$\{\mu_k\}$, 数列$\left\{\dfrac{d(\mu_k)}{\|d(\mu_k)\|}\right\}$有界, 故有收敛子列, 不妨设为其本身, 即

$$\lim_{k\to\infty}\frac{d(\mu_k)}{\|d(\mu_k)\|}=\bar{d}\neq 0. \tag{6.2.12}$$

由$d(\mu_k)$的定义,

$$-\frac{g(x)}{\|g(x)\|}=\frac{(J(x)^{\mathrm{T}}J(x)+\mu_k I)d(\mu_k)}{\|(J(x)^{\mathrm{T}}J(x)+\mu_k I)d(\mu_k)\|}$$
$$=\frac{J(x)^{\mathrm{T}}J(x)\dfrac{d(\mu_k)}{\mu_k\|d(\mu_k)\|}+\dfrac{d(\mu_k)}{\|d(\mu_k)\|}}{\left\|J(x)^{\mathrm{T}}J(x)\dfrac{d(\mu_k)}{\mu_k\|d(\mu_k)\|}+\dfrac{d(\mu_k)}{\|d(\mu_k)\|}\right\|}.$$

令$k\to\infty$得

$$-\frac{g(x)}{\|g(x)\|}=\bar{d}.$$

结合(6.2.12)得,

$$\lim_{k\to\infty}\left\langle\frac{-g(x)}{\|g(x)\|},\frac{d(\mu_k)}{\|d(\mu_k)\|}\right\rangle=1.$$

由极限的保号性知命题结论成立. 证毕

μ对算法稳定性的影响

性质6.2.4：

$\left(J(x)^T J(x) + \mu I\right)$的条件数关于$\mu$单调不增.

证明 由于$\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x})$对称半正定, 故存在正交阵$\boldsymbol{Q}$使

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x})\boldsymbol{Q} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

其中, $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$. 根据定义, 矩阵$(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I})$的条件数为$\dfrac{\lambda_1+\mu}{\lambda_n+\mu}$. 由

$$\frac{\partial}{\partial\mu}\left(\frac{\lambda_1+\mu}{\lambda_n+\mu}\right) = \frac{\lambda_n-\lambda_1}{(\lambda_n+\mu)^2} \leqslant 0$$

知$(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I})$的条件数关于$\mu$单调不增. 证毕

LM方法实下降量的估计

定理6.2.3 设x为非线性最小二乘问题的非稳定点，则对任意$\sigma \in (0,1)$,存在$\bar{\mu} > 0$,使得对任意的$\mu \geq \bar{\mu}$,

$$f\left(x + d\left(\mu\right)\right) \leq f\left(x\right) + \sigma \left\langle g(x), d\left(\mu\right)\right\rangle.$$

- **μ取值太大，相应的$\|d(\mu)\|$ 较小，新的迭代点离当前迭代点太近**
- **μ取值太小，相应的下降量得不到满足**

如何调整参数μ？

参数μ调整策略：

$$m_k(d)=\frac{1}{2}\|r_k+J_k d\|^2+\frac{1}{2}\mu_k\|d\|^2$$

实下降量与预下降量

预下降量： $\mathrm{Pred}_k := m_k(0)-m_k(d_k)$

实下降量： $\mathrm{Ared}_k := f(x_k)-f(x_k+d_k)$

比值： $\gamma_k=\dfrac{\mathrm{Ared}_k}{\mathrm{Pred}_k}$

调整策略

若$\gamma_k<\dfrac{1}{4}$， 令$\mu_{k+1}=\dfrac{1}{\alpha}\mu_k\ (\alpha\in(0,1))$

若$\gamma_k>\dfrac{3}{4}$， 令$\mu_{k+1}=\beta\mu_k\ (\beta\in(0,1))$

若$\gamma_k\in\left[\dfrac{1}{4},\dfrac{3}{4}\right]$， 令$\mu_{k+1}=\mu_k$

→ 每次调整μ_k后，需要重新计算线性方程组

$$(J_k^T J_k+\mu_k I)d=-J_k^T r_k$$

→ 计算量较大

带线搜索的Levenberg–Marquardt方法

- **基本思想:** LM方法 线搜索
- **Armijo步长规则下的收敛性分析:**

定理 6.2.4 设Armijo步长规则下的LM方法产生无穷迭代点列$\{x_k\}$. 若$\{x_k, \mu_k\}$的聚点(x^*, μ^*)满足$((J^*)^T J^* + \mu^* I)$正定, 则$\nabla f(x^*) = 0$.

定理 6.2.5 对最小二乘问题(6.2.1), 设$r(x)$二阶连续可微, 并设Armijo步长规则下的LM方法产生的点列$\{x_k\}$收敛到(6.2.1)的局部最优解x^*且$\mu_k \to 0$. 若$[(J^*)^T J^*]$非奇异, 矩阵$\left(\frac{1}{2} - \sigma\right) J^{*T} J^* - \frac{1}{2} S^*$正定, 则当$k$充分大时$\alpha_k = 1$,且

$$\limsup_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leqslant \|((J^*)^T J^*)^{-1}\| \|S(x^*)\|.$$

最小二乘问题

- 基本模型: $\min\limits_{x} \frac{1}{2}\|r(x)\|^2$
- 线性最小二乘的求解：线性方程组求解
- 非线性最小二乘的求解:
  - Gauss–Newton 方法
  - Levenberg–Marquardt方法
    - Levenberg–Marquardt–Fletcher方法
    - 带线搜索的Levenberg–Marquardt方法