最小二乘问题

教学提纲

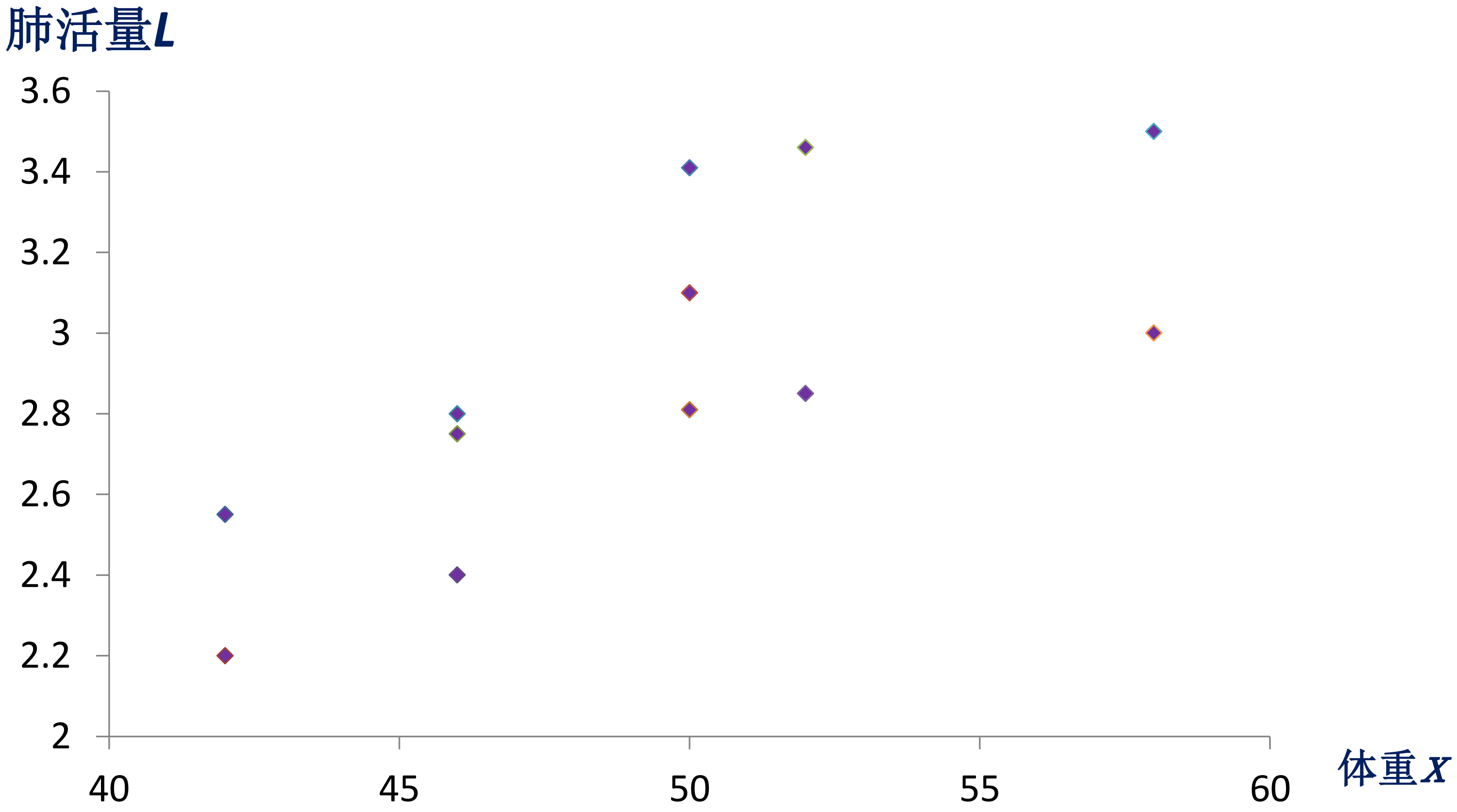
- 最小二乘问题
- 线性最小二乘
- 非线性最小二乘

一、最小二乘问题

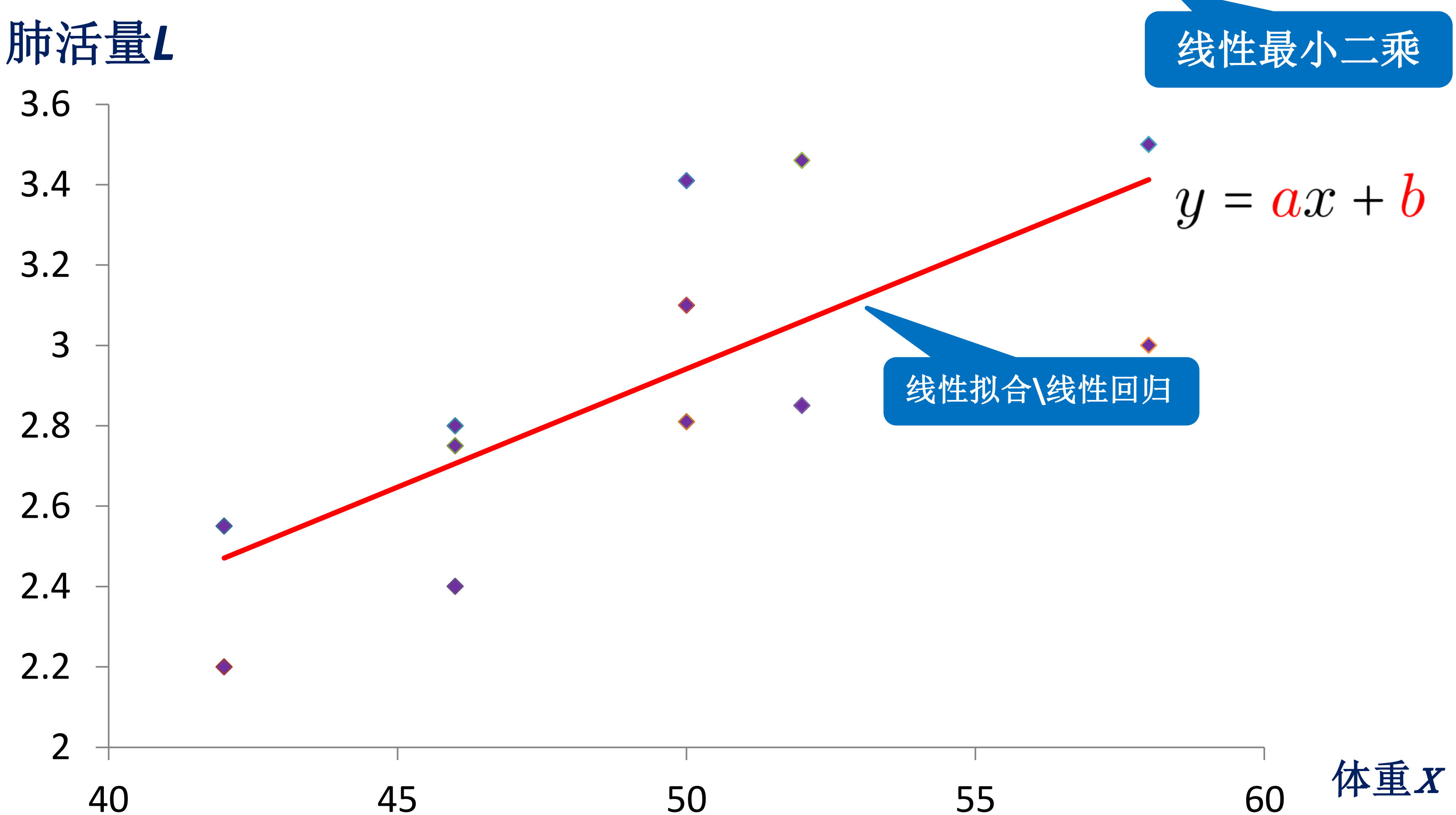
问题引入

例： 考察人的体重和肺活量之间的关系

编号	1	2	3	4	5	6	7	8	9	10	11	12
体重(kg)	42	42	46	46	46	50	50	50	52	52	58	58
肺活量(L)	2.55	2.2	2.75	2.4	2.8	2.81	3.41	3.1	3.46	2.85	3.5	3


肺活量L
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
40
45
50
55
60
体重X

$$\min_{a,b}\sum_{i=1}^{12}\left(y_i - l_i\right)^2 = \sum_{i=1}^{12}\left(\left(ax_i + b\right) - l_i\right)^2$$



最小二乘问题起源:

19世纪初由德国数学家Gauss在预测行星运行轨道时提出。

模型输出变量

模型参数

模型输入变量

基本思想: 通过函数模型 $y = \phi(\boldsymbol{x}, t)$，对观测数据进行拟合，探讨输入数据和输出数据之间的数学关系。

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	$\phi(x, t_m)$

最小二乘问题数学模型

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

模型特点：

- $m \gg n$
- 目标函数为平方和形式，函数值非负。
- 拟合函数选取恰当时，目标函数的最优值靠近零。

只讨论拟合函数中参数的取值，不关心拟合函数的选取。

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

$r_i(\boldsymbol{x}) = y_i - \phi(\boldsymbol{x}, t_i),\ i = 1, 2, \cdots, m,$

$\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}); r_2(\boldsymbol{x}); \cdots; r_m(\boldsymbol{x}))$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x})$$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x})$$

$\boldsymbol{r}(\boldsymbol{x})$ 线性 → **线性最小二乘**

$\boldsymbol{r}(\boldsymbol{x})$ 非线性 → **非线性最小二乘**

二、线性最小二乘

数学模型

$$\min_{x\in\mathbb{R}^n} \frac{1}{2} r(x)^{\mathrm{T}} r(x)$$

$$\Big\downarrow \quad r(x) = Ax - b,\quad A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$$

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

定理 线性最小二乘问题存在全局最优解。它有唯一最优解的充分必要条件是矩阵 A 列满秩。

证明：

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

$$\Big\Downarrow \text{凸优化最优性条件}$$

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

由线性代数的知识

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}\in\mathcal{R}(\boldsymbol{A}^{\mathrm{T}})=\mathcal{R}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$$

线性方程组 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 有解，且在矩阵$\boldsymbol{A}$列满秩时，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 非奇异。 ■

二、非线性最小二乘

数学模型

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

无约束优化

最速下降法？ **太慢！**

牛顿算法？**太累！ 不收敛！**

共轭梯度法？**太繁琐！**

问题： 目标函数为平方和形式，且最优值靠近零。

有无针对该问题的有效算法？

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

➢ **梯度：**

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

$$\boldsymbol{J}(\boldsymbol{x}) = D_{\boldsymbol{x}}\boldsymbol{r}(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial r_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial r_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial r_1(\boldsymbol{x})}{\partial x_n} \\ \frac{\partial r_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial r_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial r_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial r_m(\boldsymbol{x})}{\partial x_1} & \frac{\partial r_m(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial r_m(\boldsymbol{x})}{\partial x_n} \end{pmatrix}$$

- **Hesse阵**

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

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

- **牛顿算法**

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \left(\boldsymbol{J}_k^{\mathrm{T}} \boldsymbol{J}_k + \boldsymbol{S}_k\right)^{-1} \boldsymbol{J}_k^{\mathrm{T}} \boldsymbol{r}(\boldsymbol{x}_k)$$

二阶导数项 $\boldsymbol{S}(\boldsymbol{x}) = \sum_{i=1}^{m} r_i(\boldsymbol{x}) \nabla^2 r_i(\boldsymbol{x})$ 的计算量大，且未必收敛！

Newton方法

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \left(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k + \boldsymbol{S}_k\right)^{-1}\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}_k)$$

残量函数$\boldsymbol{r}(\boldsymbol{x})$最优值很小时，

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

Gauss-Newton方向 $\boldsymbol{d}_k^{\mathrm{GN}} = -\left(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k\right)^{-1}\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}_k)$

若$\boldsymbol{J}_k$列满秩，

$\boldsymbol{d}_k^{\mathrm{GN}}$为下降方向

Gauss-Newton方法 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k^{\mathrm{GN}}$

Gauss-Newton算法收敛性

定理 设目标函数水平集有界，残量函数 $\boldsymbol{r}(\boldsymbol{x})$ 及其**Jacob** $\boldsymbol{J}(\boldsymbol{x})$ 在水平集上**Lipschitz**连续, 且在水平集上满足正则性条件, 即存在 $\gamma > 0$，使对任意 $\boldsymbol{x} \in \mathcal{L}(\boldsymbol{x}_0)$，都有

$$\|\boldsymbol{J}(\boldsymbol{x})\boldsymbol{y}\| \geqslant \gamma\|\boldsymbol{y}\|, \quad \forall\ \boldsymbol{y} \in \mathbb{R}^n.$$

则**Wolfe**步长规则下的**G-N**算法产生点列的任一聚点为最小二乘问题的稳定点.

运行及收敛性条件：残量函数的Jacob阵 $\boldsymbol{J}(\boldsymbol{x})$ 列满秩

Gauss-Newton算法的改进

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k^{\mathrm{GN}}$$

$$\boldsymbol{d}^{\mathrm{GN}} = -(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}))^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x})$$

$$=\arg\min_{\boldsymbol{d}\in\mathbb{R}^n} \frac{1}{2}\|\boxed{\boldsymbol{r}(\boldsymbol{x}) + \boldsymbol{J}(\boldsymbol{x})\boldsymbol{d}}\|^2$$

运行条件：Jacob阵$J(x)$满足正则性条件（列满秩）

$r(\boldsymbol{x}+\boldsymbol{d})$ 线性近似

添加正则项

$\mu > 0$

$$\min_{\boldsymbol{d}\in\mathbb{R}^n} \|\boldsymbol{r}(\boldsymbol{x}) + \boldsymbol{J}(\boldsymbol{x})\boldsymbol{d}\|^2 + \mu\|\boldsymbol{d}\|^2$$

线性近似度提升

最优性条件

Levenberg-Marquardt方法

$$\boldsymbol{d}^{\mathrm{LM}} = -(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I})^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}).$$

Levenberg-Marquardt方法

$$\boldsymbol{d}^{\mathrm{LM}} = -(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu \boldsymbol{I})^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}).$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k^{\mathrm{LM}}$$

参数 μ 影响 $\boldsymbol{d}^{\mathrm{LM}}$ 的长度和方向

性质1 $\|\boldsymbol{d}(\mu)\|$关于$\mu > 0$ 单调不增，且$\mu \to \infty$ 时，$\|\boldsymbol{d}(\mu)\| \to 0$

性质2 $\boldsymbol{d}(\mu)$ 与 $-\boldsymbol{g}(\boldsymbol{x})$的夹角关于$\mu > 0$ 单调不增。

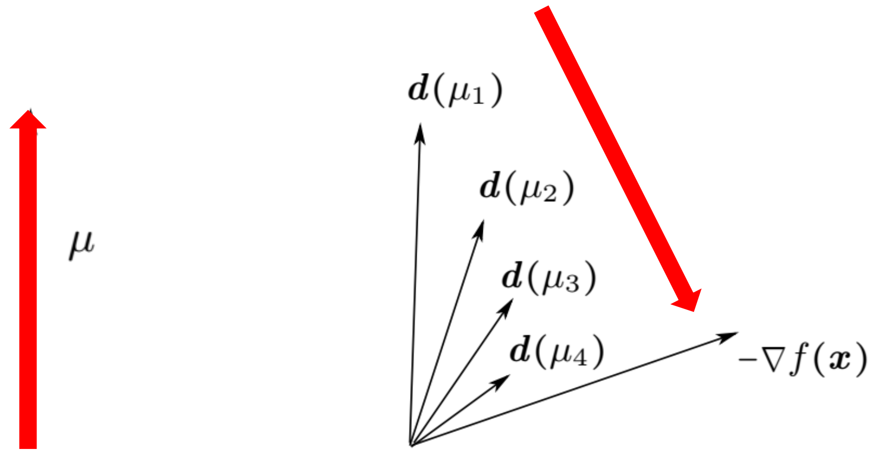
即参数μ 越大，$\boldsymbol{d}(\mu)$越靠近目标函数的负梯度方向，

其下降性越强。

证明要点： 利用导数的符号。

参数 μ 对 $\boldsymbol{d}^{\mathrm{LM}}$ 的影响

$$\boldsymbol{d}^{\mathrm{LM}} = -\left(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I}\right)^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}).$$



Armijo步长规则下的L-M方法

对任意 的μ >0,

$$d^{\mathrm{LM}} = -(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I})^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}).$$

都是下降方向。 对其进行线搜索，即得带线搜索的LM方法.

定理 设Armijo步长规则下的LM方法产生无穷迭代点列$\{\boldsymbol{x}_k\}$.
若$\{\boldsymbol{x}_k, \mu_k\}$的聚点$(\boldsymbol{x}^*, \mu^*)$满足$((\boldsymbol{J}^*)^{\mathrm{T}}\boldsymbol{J}^* + \mu^*\boldsymbol{I})$正定,
则算法全局收敛，即$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

参数 μ 的选取

对LM方法，μ 取值太小，$\boldsymbol{d}(\mu)$ 靠近Gauss-Newton方向，因 $\boldsymbol{J}(\boldsymbol{x})$ 的奇异性，算法执行时会出现困难；μ 取值过大，$\boldsymbol{d}(\mu)$ 靠近负梯度方向，算法效率会受影响。

如果非线性最小二乘问题的最优值近似为零，则 $\mu_k = \|\boldsymbol{r}(\boldsymbol{x}_k)\|^2$ 是一个具有自适应性质的选项.