

非线性最小二乘问题算法

周旸

武汉大学数学与统计学院

2025 年 11 月 16 日



目录

- ① 非线性最小二乘问题
- ② Gauss-Newton 方法
- ③ LM 方法
- ④ 大残量问题的拟 Newton 法
- ⑤ 应用举例：相位恢复

① 非线性最小二乘问题

② Gauss-Newton 方法

③ LM 方法

④ 大残量问题的拟 Newton 法

⑤ 应用举例：相位恢复

例子：数据拟合

给定观测数据 $\{(t_i, y_i)\}_{i=1}^m$ 和模型 $y(t) = \phi(x, t)$, 拟合参数 $x \in \mathbb{R}^n$,

$$y_i \approx \phi(x; t_i), \quad i = 1, 2, \dots, m.$$

定义残差为 $r_i(x) = \phi(x; t_i) - y_i$, $i = 1, 2, \dots, m$, 构造非线性最小二乘问题

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \sum_{i=1}^m r_i(x)^2.$$

若函数 $y = y(t)$ 形式未知, $\phi(x; t)$ 取多项式, 则这个问题是多项式的离散最佳平方逼近问题。

例子：回归模型

假设观测模型为：

$$y_i = \phi(x; t_i) + \epsilon_i, \quad i = 1, 2, \dots, m,$$

其中 $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, i.i.d. 是独立同分布的正态随机误差。
对未知参数 x 做最大似然估计，

$$\begin{aligned} \arg \max_x L(x) &= \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(y_i - \phi(x; t_i))^2}{2\sigma^2} \right) \\ \iff \arg \min_x \sum_{i=1}^m (y_i - \phi(x; t_i))^2 &= \sum_{i=1}^m r_i(x)^2. \end{aligned}$$

即当误差服从正态分布时，最小二乘估计等价最大似然估计。

例子：非线性方程组求解

考虑求解非线性方程组

$$f_i(x) = 0, \quad i = 1, 2, \dots, m, \quad x \in \mathbb{R}^n, \quad f_i \in C^2.$$

定义残差 $r_i(x) = f_i(x)$, $i = 1, 2, \dots, m$, 构造最小二乘问题

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \sum_{i=1}^m r_i(x)^2 = \frac{1}{2} \|f(x)\|_2^2.$$

思考：与数值分析中学的**非线性方程组求解算法**有什么联系？

非线性最小二乘问题

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \|r(x)\|_2^2 = \frac{1}{2} r(x)^\top r(x), \quad r(x) = \begin{bmatrix} r_1(x) \\ \vdots \\ r_m(x) \end{bmatrix} \in \mathbb{R}^m$$

是一个无约束优化问题。一般有 $m > n$ 。

若 $r_i(x) \in C^2$, $i = 1, 2, \dots, m$, 并记 $J(x)$ 是 $r(x)$ 的 Jacobi 矩阵, 则

$$\nabla f(x) = J(x)^\top r(x),$$

$$\nabla^2 f(x) = J(x)^\top J(x) + \sum_{i=1}^m r_i(x) \nabla^2 r_i(x),$$

可以利用一般 (基于导数的) 无约束优化算法求解这个模型问题。

非线性最小二乘问题的结构

$$\nabla f(x) = J(x)^{\top} r(x),$$
$$\nabla^2 f(x) = \boxed{J(x)^{\top} J(x)} + \sum_{i=1}^m r_i(x) \nabla^2 r_i(x).$$

最小二乘问题好的结构在于其二阶导数包含完全由一阶导数构成的项。可能带来的好处有，

- 一阶导数的信息比二阶导数更容易获得；
- 算法的计算量可能更小或者更稳定；
- $r(x)$ 性质比较好时，最优解附近 $J^{\top} J$ 才是主要的项。

并不期待最小二乘问题上算法有更好的收敛性和收敛速度。

无约束优化算法的选取

- 最速下降法：完全没有利用到最小二乘问题的结构 ×
- Newton 法：希望 $\nabla^2 f(x) \approx J(x)^\top J(x)$ ✓
- 拟 Newton 法：可以利用拟 Newton 法的方式更新 $\nabla^2 f(x)$ 的残项 ✓

- ① 非线性最小二乘问题
- ② Gauss-Newton 方法
- ③ LM 方法
- ④ 大残量问题的拟 Newton 法
- ⑤ 应用举例：相位恢复

Gauss-Newton 方法的想法

由于 $f(x)$ 有二阶导数，故可用 Newton 法求解。Newton 方程为

$$\nabla^2 f(x_k) d_k = -\nabla f(x_k).$$

若有 $\nabla^2 f(x_k) \approx J(x_k)^\top J(x_k)$ ，则近似的 Newton 方程为

$$\boxed{J_k^\top J_k d_k = -J_k^\top r_k.} \quad (1)$$

其中 $J_k := J(x_k)$, $r_k := r(x_k)$ 。形如方程 (1) 称为正规方程。
这样做的好处有：

- 计算简单，只需计算 Jacobi 矩阵，无需二阶导数信息；
- 正规方程 (1) 有很好的结构，可用 QR 方法求解；
- d_k 一定是下降方向。

这样的方法就叫做 **Gauss-Newton 方法**。

回顾: 如何求解正规方程?

考虑正规方程 $A^T Ax = A^T b$ 。假设 $A \in \mathbb{R}^{m \times n}$ 是列满秩的, 即 $m \geq n$, 且 $\text{rank}(A) = n \leq m$ 。

Algorithm 1: QR 分解求解正规方程

Input: $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

Output: x 使得 $A^T Ax = A^T b$

- 1 计算 A 的 QR 分解 $A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$ 。
 - 2 计算 $c_1 = Q_1^T b$. // $Q = \begin{bmatrix} (Q_1)_{m \times n}, (Q_2)_{m \times (m-n)} \end{bmatrix}$
 - 3 求解上三角方程组 $Rx = c_1$ 。
 - 4 **return** x
-

Gauss-Newton 方法

Algorithm 2: Gauss-Newton 方法

Input: $x_0 \in \mathbb{R}^n$, $k \leftarrow 0$

Output: $x^* = \arg \min_{x \in \mathbb{R}^n} \frac{1}{2} \|r(x)\|_2^2$

1 **while** 未达到停机准则 **do**

2 计算残差向量 r_k , Jacobi 矩阵 J_k 。

3 利用 QR 分解, 求解 $J_k^\top J_k d_k = -J_k^\top r_k$ 得到 d_k 。

4 使用线搜索准则计算步长 α_k 。

5 更新: $x_{k+1} = x_k + \alpha_k d_k$ 。

6 $k \leftarrow k + 1$ 。

7 **return** x^*

Gauss-Newton 方法的性质

性质 (Gauss-Newton 保下降方向)

若 J_k 是 (列) 满秩矩阵, 则 Gauss-Newton 法得到的方向 d_k 总是一个下降方向

$$d_k^\top \nabla f(x_k) = d_k^\top J_k^\top r_k = -\|J_k d_k\|_2^2 < 0.$$

对于一般的 Newton 方法, 当 $\nabla^2 f(x_k)$ 正定时, Newton 方向

$$\nabla f(x_k)^\top d_k = -\nabla f(x_k)^\top (\nabla^2 f(x_k))^{-1} \nabla f(x_k) < 0$$

是一个下降方向。由于 $J_k^\top J_k$ 的特殊结构 (一定是半正定阵), 要求 J_k (列) 满秩即可保证 $J_k^\top J_k$ 是一个正定阵。

Gauss-Newton 方法的全局收敛性

定理 (Gauss-Newton 的全局收敛性)

如果每个残差函数 r_i 在有界下水平集 $\mathcal{L} = \{x \mid f(x) \leq f(x_0)\}$ 的一个邻域 \mathcal{N} 内是 Lipschitz 连续可微的, 并且 Jacobi 矩阵 $J(x)$ 在 \mathcal{N} 内满足一致满秩条件 $\|Jz\| > \gamma\|z\|, \forall z \in \mathcal{N}$, 而步长满足 Wolfe 准则, 则对 Gauss-Newton 法得到的序列 $\{x_k\}$ 有

$$\lim_{k \rightarrow \infty} \nabla f(x_k) = \lim_{k \rightarrow \infty} (J_k)^\top r_k = 0.$$

只需验证 Zoutendijk 条件,

$$\cos \theta_k = -\frac{\nabla f(x_k)^\top d_k}{\|d_k\| \|\nabla f(x_k)\|} = \frac{\|J_k d_k\|^2}{\|d_k\| \|J_k^\top J_k d_k\|} \geq \frac{\gamma^2 \|d_k\|^2}{\tilde{\beta}^2 \|d_k\|^2} = \frac{\gamma^2}{\tilde{\beta}^2} > 0,$$

即可得 $\nabla f(x_k) \rightarrow 0$.

Gauss-Newton 法的局部收敛性

定理 (Gauss-Newton 的局部收敛性)

设 $r_i(x)$ 二阶连续可微, x^* 是最小二乘问题的最优解, Hesse 矩阵 $\nabla^2 f(x)$ 和其近似矩阵 $J(x)^\top J(x)$ 均在点 x^* 的一个邻域内 Lipschitz 连续, 则当 Gauss-Newton 算法步长 a_k 恒为 1 时,

$$\|x^{k+1} - x^*\| \leq C \|((J^*)^\top J^*)^{-1} H^*\| \|x^k - x^*\| + \mathcal{O}(\|x^k - x^*\|^2),$$

其中 $H^* = \sum_{i=1}^m r_i(x^*) \nabla^2 r_i(x^*)$ 为 Hesse 矩阵 $\nabla^2 f(x^*)$ 去掉 $J(x^*)^\top J(x^*)$ 的部分, $C > 0$ 为常数。

- 当 $\|H(x^*)\|$ 充分小, Gauss-Newton 法的渐进收敛速度是 Q -线性的;
- 特别地, 当 $\|H(x^*)\| = 0$ 时, 收敛速度是 Q -二次的,
- 一般无约束问题的 Newton 法收敛速度是 Q -二次的。

- ① 非线性最小二乘问题
- ② Gauss-Newton 方法
- ③ LM 方法
- ④ 大残量问题的拟 Newton 法
- ⑤ 应用举例：相位恢复

LM 方法的想法

在 Gauss-Newton 法中，有正规方程

$$J_k^T J_k d_k = -J_k^T r_k.$$

若 $J_k^T J_k$ 奇异，则 Gauss-Newton 迭代无法进行下去。
为了能让 GN 迭代下去，取 $\lambda_k > 0$ ，求解

$$(J_k^T J_k + \lambda_k I) d_k = -J_k^T r_k. \quad (2)$$

这样的方法就是 **LM (Levenberg-Marquardt) 方法**。同时，

$$d_k^T \nabla f(x_k) = -r_k^T J_k (J_k^T J_k + \lambda_k I)^{-1} J_k^T r_k < 0.$$

由 (2) 得到的 d_k 一定是一个下降方向。

LM 方法的细节

① 如何选取 λ_k ?

- 基于信赖域方法 \rightarrow 信赖域型 LM 方法;
- 对于特定问题设计的选取方式, 如相位恢复问题中,

$$\lambda_k = \begin{cases} 70000n\sqrt{nf(x_k)}, & f(x_k) \geq \frac{1}{900n} \|x_k\|_2^2, \\ \sqrt{f(x_k)}, & \text{otherwise.} \end{cases}$$

② 如何隐式地 (不显式计算 $J_k^\top J_k$) 求解 LM 方程

$$(J_k^\top J_k + \lambda_k I) d_k = -J_k^\top r_k?$$

LM 方程与信赖域问题的关系

定理 (LM 方程与信赖域问题的关系)

向量 d^* 是信赖域子问题

$$\min_d \quad \frac{1}{2} \|Jd + r\|^2, \quad \text{s.t. } \|d\| \leq \Delta \quad (3)$$

的解, 当且仅当 d^* 是可行解, 并且存在数 $\lambda \geq 0$ 使得

$$(J^\top J + \lambda I)d^* = -J^\top r,$$

$$\lambda(\Delta - \|d^*\|) = 0,$$

$$(J^\top J + \lambda I) \succ 0.$$

- 当 Gauss-Newton 法的 d^{GN} 落在信赖域中, 则 $d^* = d^{\text{GN}}$ 且 $\lambda = 0$;
- 我们可以用信赖域方法代替线搜索, 同时获得正定的 Hesse 近似矩阵 $J^\top J + \lambda I$ 。

LM 方程求解

将 LM 方程 $(J^T J + \lambda I)d = -J^T r$ 写成

$$\begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix}^T \begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix} d = - \begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix}^T \begin{bmatrix} r \\ 0 \end{bmatrix}.$$

此问题的系数矩阵带有一定结构，每次改变 λ 进行试探时，有关 J 的块是不变的，因此无需重复计算 J 的 QR 分解。具体来说，设 $J = QR$ 为 J 的 QR 分解，其中 $Q \in \mathbb{R}^{m \times m}$ ， $R \in \mathbb{R}^{m \times n}$ 。我们有

$$\begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix} = \begin{bmatrix} QR \\ \sqrt{\lambda} I \end{bmatrix} = \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix}.$$

矩阵 $\begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix}$ 含有较多零元素，利用这个特点我们可以使用 Householder 变换或 Givens 变换来完成此矩阵的 QR 分解。

信赖域型 LM 方法

Algorithm 3: 信赖域型 LM 方法

Input: $x_0 \in \mathbb{R}^n$, $k \leftarrow 0$

Output: $x^* = \arg \min_{x \in \mathbb{R}^n} \frac{1}{2} \|r(x)\|_2^2$

- 1 **while** 未达到停机准则 **do**
 - 2 用迭代法求解信赖域子问题 (3), 得到 λ_k 。
 - 3 求解 $(J_k^\top J_k + \lambda_k I) d_k = -J_k^\top r_k$, 得到 d_k 。
 - 4 通过线搜索找到合适的步长 $\alpha_k > 0$, 令
 $x_{k+1} = x_k + \alpha_k d_k$ 。
 - 5 $k \leftarrow k + 1$ 。
 - 6 **return** x^*
-

LM 方法的收敛性

定理 (LM 方法的全局收敛性)

假设常数 $\eta \in (0, \frac{1}{4})$, 下水平集 \mathcal{L} 是有界的, 且每个 $r_i(x)$ 在下水平集 \mathcal{L} 的一个邻域 \mathcal{N} 内是 *Lipchitz* 连续可微的。假设对任意 k , 子问题 (3) 的近似解 d_k 满足

$$m_k(0) - m_k(d_k) \geq c_1 \|J_k^\top r_k\| \min \left\{ \Delta_k, \frac{\|J_k^\top r_k\|}{\|J_k^\top J_k\|} \right\},$$

其中 $c_1 > 0$ 且 $\|d_k\| \leq \gamma \Delta_k$, $\gamma \geq 1$, 则

$$\lim_{k \rightarrow \infty} \nabla f(x_k) = \lim_{k \rightarrow \infty} J_k^\top r_k = 0.$$

信赖域型 LM 方法的缺点

- Δ_k 要如何选取？为什么不直接选 λ_k ？
- 计算量过大，一个迭代步需要求解一个信赖域子问题和一个 LM 方程。

如何改进信赖域型 LM 方法？

LMF 方法的想法

根据信赖域算法的思想，我们需要计算目标函数的预测下降量和实际下降量的比值 ρ_k ，来确定下一步信赖域半径的大小，

$$\rho_k = \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)},$$

其中

$$f(x) = \frac{1}{2} \sum_{j=1}^m r_j^2(x), \quad m_k(d) = \|J_k d + r_k\|_2^2.$$

计算 ρ_k 后，我们根据 ρ_k 的大小来更新 λ_k 。当乘子 λ_k 增大时，信赖域半径会变小，反之亦然。因此 λ_k 的变化策略应该和信赖域算法中 Δ_k 的恰好相反。

LMF 方法

Algorithm 4: LMF 方法

Input: $x_0 \in \mathbb{R}^n$, 初始乘子 λ_0 , $k \leftarrow 0$

Input: 给定参数 $0 \leq \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1$, $\gamma_1 < 1 < \gamma_2$

Output: $x^* = \arg \min_{x \in \mathbb{R}^n} f(x)$

1 **while** 未达到停机准则 **do**

2 求解 LM 方程 $(J_k^\top J_k + \lambda_k I) d_k = -J_k^\top r_k$ 得到迭代方向 d_k 。

3 计算下降率 ρ_k 。

4 更新乘子:

$$\lambda_{k+1} = \begin{cases} \gamma_2 \lambda_k, & \rho_k < \bar{\rho}_1, \\ \gamma_1 \lambda_k, & \rho_k > \bar{\rho}_2, \\ \lambda_k, & \text{otherwise.} \end{cases}$$

5 更新自变量: $x_{k+1} = x_k + d_k$, 若 $\rho_k > \eta$; 否则 $x_{k+1} = x_k$ 。

6 **return** x

LMF 方法的好处

和 LM 方法相比, LMF 方法每一次迭代只需要求解一次 LM 方程, 从而极大地减少了计算量, 在编程方面也更容易实现。所以 LMF 方法在求解最小二乘问题中是很常见的做法。

- ① 非线性最小二乘问题
- ② Gauss-Newton 方法
- ③ LM 方法
- ④ 大残量问题的拟 Newton 法
- ⑤ 应用举例：相位恢复

大残量问题拟 Newton 法的想法

- 不能期待所有最小二乘问题都满足

$$\nabla^2 f(x_k) = J(x_k)^\top J(x_k) + \sum_{i=1}^m r_i(x_k) \nabla^2 r_i(x_k)$$

$$\approx J(x_k)^\top J(x_k).$$

尤其在数据不好或建模不好的时候；

- 对比一般无约束问题的 Newton 法，Gauss-Newton 法的渐进收敛速度从二次收敛退化到线性收敛，这是忽略 $\nabla^2 f(x)$ 第二项的结果；
- 拟 Newton 法也可以做到超线性收敛。

大残量问题拟 Newton 法的想法

对于非线性最小二乘问题，我们既希望不直接计算残差函数的二阶导数，又希望考虑到残项。拟 Newton 法是一个好的方案，因为拟 Newton 法不依赖二阶导数。

回顾：拟 Newton 法的割线方程（拟 Newton 条件）

设 $f(x) \in C^2$ 。向量值函数 $\nabla f(x)$ 在点 x_{k+1} 处的近似为

$$\nabla f(x) = \nabla f(x_{k+1}) + \nabla^2 f(x_{k+1})(x - x_{k+1}) + \mathcal{O}(\|x - x_{k+1}\|^2).$$

令 $x = x_k$, $s_k = x_{k+1} - x_k$ 及 $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, 得到

$$\nabla^2 f(x_{k+1}) s_k + \mathcal{O}(\|s_k\|^2) = y_k.$$

忽略高阶项 $\|s_k\|^2$, 我们希望 Hesse 矩阵的近似矩阵 B_{k+1} 满足方程

$$y_k = B_{k+1} s_k,$$

或者其逆的近似矩阵 H_{k+1} 满足方程

$$s_k = H_{k+1} y_k.$$

并称上式为割线方程。

回顾：拟 Newton 法的矩阵更新

- 秩 1 更新公式

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

- DFP 公式

$$B_{k+1} = B_k + \left(1 + \frac{s_k^\top B_k s_k}{s_k^\top y_k}\right) \frac{y_k y_k^\top}{s_k^\top y_k} - \frac{y_k s_k^\top B_k + B_k s_k y_k^\top}{s_k^\top y_k}.$$

- BFGS 公式

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

回顾：拟 Newton 法的算法框架

Algorithm 5: 拟 Newton 法

Input: $x_0 \in \mathbb{R}^n$, 初始矩阵 $B_0 \in \mathbb{R}^{n \times n}$ (或 H_0), $k \leftarrow 0$

Output: $x^* = \arg \min_{x \in \mathbb{R}^n} f(x)$

- 1 **while** 未达到停机准则 **do**
 - 2 计算方向 $d_k = -(B_k)^{-1} \nabla f(x_k)$ 或 $d_k = -H_k \nabla f(x_k)$ 。
 - 3 通过线搜索找到合适的步长 $\alpha_k > 0$, 令
 $x_{k+1} = x_k + \alpha_k d_k$ 。
 - 4 更新 Hesse 矩阵的近似矩阵 B_{k+1} 或其逆矩阵的近似
 H_{k+1} 。
 - 5 $k \leftarrow k + 1$ 。
 - 6 **return** x
-

大残量问题拟 Newton 法的想法

但是，直接使用拟 Newton 法，就没有充分利用非线性最小二乘问题的导数结构。

我们并不排斥计算 $J^T J$ 项。自然，我们设想：是否可以利用拟 Newton 估计残项，而不是整个二阶导数？

拟 Newton 法的构建

设 B_k 是 $\nabla^2 f(x_k)$ 的近似矩阵, 即

$$B_k = J_k^\top J_k + T_k.$$

我们的目标是让 T_{k+1} 和 Hesse 矩阵的第二部分尽量相似, 即

$$T_{k+1} \approx \sum_{j=1}^m r_j(x_{k+1}) \nabla^2 r_j(x_{k+1}).$$

大残量问题的拟 Newton 法的近似矩阵

由一阶 Taylor 展开得知, T_{k+1} 应该尽量保留原 Hesse 矩阵的性质, 即

$$\begin{aligned} T_{k+1} s_k &\approx \left(\sum_{j=1}^m r_j(x_{k+1}) \nabla^2 r_j(x_{k+1}) \right) s_k = \sum_{j=1}^m r_j(x_{k+1}) (\nabla^2 r_j(x_{k+1})) s_k \\ &\approx \sum_{j=1}^m r_j(x_{k+1}) (\nabla r_j(x_{k+1}) - \nabla r_j(x_k)) \\ &= (J_{k+1})^\top r_{k+1} - (J_k)^\top r_{k+1}. \end{aligned}$$

令 $\hat{y}_k = (J_{k+1})^\top r_{k+1} - (J_k)^\top r_{k+1}$, 则 T 满足的拟 Newton 条件为

$$T_{k+1} s_k = \hat{y}_k.$$

在这里注意 \hat{y}_k 不是原有的 y_k .

大残量问题拟 Newton 法的矩阵更新

- 秩 1 更新公式

$$T_{k+1} = T_k + \frac{(\hat{y}_k - T_k s_k)(\hat{y}_k - T_k s_k)^\top}{(\hat{y}_k - T_k s_k)^\top s_k}.$$

- DFP 公式

$$T_{k+1} = T_k + \left(1 + \frac{s_k^\top T_k s_k}{s_k^\top \hat{y}_k}\right) \frac{\hat{y}_k \hat{y}_k^\top}{s_k^\top \hat{y}_k} - \frac{\hat{y}_k s_k^\top T_k + T_k s_k \hat{y}_k^\top}{s_k^\top \hat{y}_k}.$$

- BFGS 公式

$$T_{k+1} = T_k + \frac{\hat{y}_k \hat{y}_k^\top}{s_k^\top \hat{y}_k} - \frac{T_k s_k (T_k s_k)^\top}{s_k^\top T_k s_k}.$$

大残量问题的拟 Newton 法

Algorithm 6: 大残量问题的拟 Newton 法

Input: $x_0 \in \mathbb{R}^n$, 初始矩阵 $T_0 \in \mathbb{R}^{n \times n}$, $k \leftarrow 0$

Output: $x^* = \arg \min_{x \in \mathbb{R}^n} f(x)$

1 **while** 未达到停机准则 **do**

2 计算方向 $d_k = -(J_k^\top J_k + T_k)^{-1} J_k^\top r_k$ 。

3 通过线搜索找到合适的步长 $\alpha_k > 0$, 令

$$x_{k+1} = x_k + \alpha_k d_k。$$

4 更新 Hesse 矩阵第二项的近似矩阵 T_{k+1} 。

5 $k \leftarrow k + 1$ 。

6 **return** x

一般无约束问题拟 Newton 法的性质

定理

- 设近似 Hesse 矩阵 B_k 对称正定, 且 $s_k^\top y_k > 0$, 则 DFP 公式或 BFGS 公式下的 B_{k+1} 对称正定;
- 对于使用精确线搜索或 (强) Wolfe 准则的 DFP 方法或 BFGS 方法, 有 $s_k^\top y_k > 0$ 。

对于一般无约束问题的拟 Newton 法, 给定正定的初始近似 Hesse 阵 $B_0 \succ 0$ 并使用精确线搜索或 (强) Wolfe 准则的 DFP 方法或 BFGS 方法, 则由定理保证每一步 $B_k \succ 0$, 从而

$$\nabla f(x_k)^\top d_k = -\nabla f(x_k)^\top B_k^{-1} \nabla f(x_k) < 0$$

保证了每一步方向都是下降方向。

大残量问题的拟 Newton 法有没有这样的性质?

大残量问题的拟 Newton 法不保下降方向

下面的算例否定了大残量问题的拟 Newton 法存在如此良好的性质。取

$$r(x) = x_1^3 + x_2 - 10, \quad x_0 = [-0.29322872, -1.51547262]^\top, \quad T_0 = I_2,$$

并取强 Wolfe 搜索参数 $\alpha = 0.4386725591$, $c_1 = 10^{-4}$, $c_2 = 0.9$, 使用 BFGS 更新公式, 得到

$$B_1 = J_1^\top J_1 + T_1 = \begin{bmatrix} -0.17513883 & 0.10228130 \\ 0.10228130 & 1.06238674 \end{bmatrix},$$

是一个不定的矩阵。从而,

$$\nabla f(x_k)^\top d_k = -\nabla f(x_k)^\top B_k^{-1} \nabla f(x_k)$$

d_k 不一定是下降方向。

- ① 非线性最小二乘问题
- ② Gauss-Newton 方法
- ③ LM 方法
- ④ 大残量问题的拟 Newton 法
- ⑤ 应用举例：相位恢复

相位恢复问题的背景

将光信号看作一个复向量。则

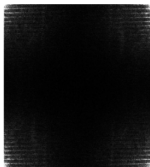
$$x = |x|e^{i \arg x}$$

工程上，测量强度是容易的，直接测量相位是困难的。
然而，下面我们将看到，相位信息比强度信息更重要。

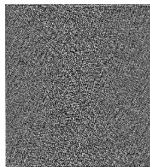
相位恢复问题的背景



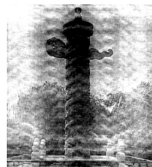
(a) Y



(b) $|\mathcal{F}(Y)|$



(c) $\text{phase}(\mathcal{F}(Y))$



(d) \hat{S}



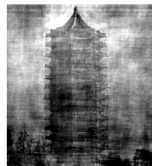
(e) S



(f) $|\mathcal{F}(S)|$



(g) $\text{phase}(\mathcal{F}(S))$



(h) \hat{Y}

相位信息比强度信息更重要!

相位恢复问题的建模

定义测量为

$$b_k = |\langle a_k, x \rangle|^2,$$

其中 a_k 可以理解为测量的方式或测量的基, b_k 为测得的强度。

测量的过程可以理解为原始信号向某个方向上的投影。

于是, 相位恢复问题可以建模为一个非线性最小二乘问题

$$\min_{x \in \mathbb{C}^n} f(x) := \frac{1}{2} \sum_{j=1}^m \left(\left| \bar{a}_j^\top x \right|^2 - b_j \right)^2.$$

求解算法相关的矩阵

$$\min_{x \in \mathbb{C}^n} f(x) := \frac{1}{2} \sum_{j=1}^m \left(\left| \bar{a}_j^\top x \right|^2 - b_j \right)^2.$$

计算得到

$$J(x) = \overline{\begin{bmatrix} a_1(\bar{a}_1^\top x) & a_2(\bar{a}_2^\top x) & \cdots & a_m(\bar{a}_m^\top x) \\ \bar{a}_1(a_1^\top \bar{x}) & \bar{a}_2(a_2^\top \bar{x}) & \cdots & \bar{a}_m(a_m^\top \bar{x}) \end{bmatrix}}^\top,$$

$$\Psi(x) := \overline{J(x)}^\top J(x) = \sum_{j=1}^m \begin{bmatrix} |\bar{a}_j^\top x|^2 a_j \bar{a}_j^\top & (\bar{a}_j^\top x)^2 a_j a_j^\top \\ (\bar{a}_j^\top x)^2 \bar{a}_j \bar{a}_j^\top & |\bar{a}_j^\top x|^2 \bar{a}_j a_j^\top \end{bmatrix}.$$

相位恢复问题的 LM 方法

LM 方法求解正则化方程

$$(\Psi(x_k) + \lambda_k I) d_k = -\nabla f(x_k), \quad (4)$$

其中

$$\lambda_k = \begin{cases} 70000n\sqrt{nf(x_k)}, & f(x_k) \geq \frac{1}{900n} \|x_k\|_2^2, \\ \sqrt{f(x_k)}, & \text{otherwise,} \end{cases}$$

并利用共轭梯度法求解 (4), 使得

$$\|(\Psi(x^k) + \lambda_k) d^k + \nabla f(x^k)\| \leq \eta_k \|\nabla f(x^k)\|,$$

其中 $\eta > 0$ 是人为设置的参数。

Take-home Page

① 什么是最小二乘问题？ $\min_x f(x) = \frac{1}{2} \|r(x)\|_2^2$.

② 为什么要研究非线性最小二乘问题？

- 最小化 $\|r(x)\|$ 的问题建模是广泛应用的；
- 最小二乘问题有很好的结构，从而有很好的算法；
- 最小二乘问题在一些问题上有更深的背景，如统计上。

③ 最小二乘问题有什么结构？

$$\nabla f = J^\top r, \quad \nabla^2 f = \boxed{J^\top J} + \sum_{i=1}^m r_i \nabla^2 r_i,$$

二阶导数可以被一阶导数的信息很好地近似。

④ 最小二乘问题有什么算法？

- $\nabla^2 f \approx J^\top J$: Gauss-Newton 方法、LM 方法 ($J^\top J$ 不正定)；
- $\nabla^2 f \approx J^\top J + T$: 大残量问题的拟 Newton 法。