

非线性最小二乘问题

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

<http://bicmr.pku.edu.cn/~wenzw/optbook.html>

致谢：本教案由张轩熙协助准备

- 1 非线性最小二乘问题
- 2 高斯-牛顿方法
- 3 Levenberg-Marquardt 方法
- 4 大残量问题的拟牛顿算法
- 5 应用举例

最小二乘问题

$$\min_x f(x) = \frac{1}{2} \sum_{j=1}^m r_j^2(x) \quad (1)$$

- 其中 $r_j: \mathbb{R}^n \rightarrow \mathbb{R}$ 是光滑函数，并且假设 $m \geq n$. 称 r_j 为残差.
- 记 $r: \mathbb{R}^n \rightarrow \mathbb{R}^m$

$$r(x) = (r_1(x), r_2(x), \dots, r_m(x))^T.$$

问题可以表述为 $\min f(x) = \frac{1}{2} \|r(x)\|_2^2$

- 一般情况下不是凸问题
- 问题1是无约束优化问题，可以直接使用线搜索或拟牛顿法求解

最小二乘问题

- 记 $J(x) \in \mathbb{R}^{m \times n}$ 是向量值函数 $r(x)$ 在点 x 处的雅可比矩阵：

$$J(x) = \begin{bmatrix} \nabla r_1(x)^T \\ \nabla r_2(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{bmatrix}.$$

- $f(x)$ 的梯度和海瑟矩阵：

$$\nabla f(x) = \sum_{j=1}^m r_j(x) \nabla r_j(x) = J(x)^T r(x), \quad (2a)$$

$$\nabla^2 f(x) = \sum_{j=1}^m \nabla r_j(x) \nabla r_j(x) + \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \quad (2b)$$

$$= J(x)^T J(x) + \sum_{i=1}^m r_i(x) \nabla^2 r_i(x), \quad (2c)$$

最小二乘问题

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

- $\nabla^2 f(x)$ 在形式上分为两部分
- 在计算 $f(x)$ 导数时已经求出 $J(x)$ ，第一项可自然得到。第二项需要额外计算。
- 如果在最优解附近，残量值较小或残量函数接近线性函数，第二项可以忽略，可以用 $J(x)^T J(x)$ 近似海瑟矩阵，基于牛顿法，结合线搜索或信赖域方法，可设计出高斯-牛顿方法和Levenberg-Marquardt方法
- 如果第二项不可忽略，则需要引入带结构的拟牛顿方法。

最小二乘问题

- 若 $r(x) = Ax - b$ ，则是线性最小二乘问题，
- $\nabla f(x) = A^T(Ax - b)$. 最优解满足正则化方程：

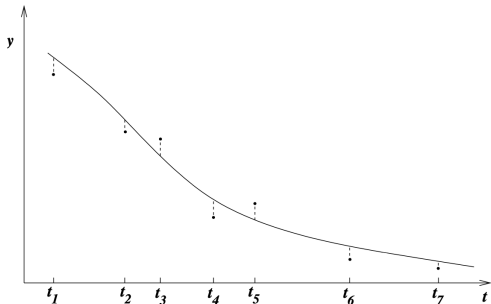
$$A^T Ax = A^T b$$

- 线性问题的求解是非线性问题求解的基础. 求解线性最小二乘问题的方法有：
 - 对 $A^T A$ 直接做 choloskey 分解，简单便捷但受 A 的条件数影响大。
 - 对 A 做 QR 分解，较稳定，相对误差小。
 - 对 A 做 SVD 分解，可以获得更精确的敏感性信息。
 - 当问题规模较大时，迭代法更有效，如共轭梯度法。

实例:模型拟合

$$\min_x \frac{1}{2} \sum_{j=1}^n (\phi(t_j; x) - y_j)^2$$

- 模型 $\phi(t; x) = x_1 + tx_2 + t_3^x + x_4 e^{-x_5 t}$ 依赖于参数向量 x 。
- $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$ 是数据点
- 目标为寻找合适的模型参数 x



- 1 非线性最小二乘问题
- 2 高斯-牛顿方法
- 3 Levenberg-Marquardt 方法
- 4 大残量问题的拟牛顿算法
- 5 应用举例

高斯-牛顿方法

- 可被看作牛顿法+线搜索
- 在迭代点 x_k ，标准牛顿法为，计算更新方向 d_k ：

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

然后做一步更新 $x_{k+1} = x_k + d_k$.

- 高斯-牛顿法的迭代方向 d_k^{GN} 满足：

$$J_k^T J_k d_k^{GN} = -J_k^T r_k. \quad (3)$$

其中 J_k 、 r_k 分别是 $J(x_k)$ 、 $r(x_k)$ 的简写

- 使用了近似

$$\nabla^2 f_k \approx J_k^T J_k$$

省略了对 $\nabla^2 r_j$ 的计算，极大的减少了计算量。

高斯-牛顿方法

- 方程(3)与线性最小二乘问题的正则化方程类似，迭代方向是如下问题的解

$$\min_d \frac{1}{2} \|J_k d + r_k\|^2$$

- 求解该问题时，可以直接对 J_k 做QR分解或SVD分解，无需计算出 $J_k^T J_k$ 。
- 若使用共轭梯度法求解，需要计算向量和矩阵 $J_k^T J_k$ 的乘法，可以依次乘 J_k 和 J_k^T ，无需计算 $J_k^T J_k$ 。
- 另一种理解高斯-牛顿方法的方式为，在点 x_k 处，考虑下一步更新 $x_k + d$ ，做近似 $r(x_k + d) \approx r_k + J_k d$ ，原问题近似为：

$$\min_d f(x_k + d) = \frac{1}{2} \|r(x_k + d)\|^2 \approx \frac{1}{2} \|J_k d + r_k\|^2$$

高斯-牛顿方法

算法可总结如下

Algorithm 1 高斯-牛顿法

- 1: 给定初始值 x_0 , $k \leftarrow 0$.
 - 2: **while** 未达到收敛准则 **do**
 - 3: 计算残差向量 r_k , 雅可比矩阵 J_k .
 - 4: 求解线性最小二乘问题 $\min_d \frac{1}{2} \|J_k d + r_k\|^2$ 确定下降方向 d_k .
 - 5: 使用线搜索准则计算步长 α_k .
 - 6: 更新: $x_{k+1} = x_k + \alpha_k d_k$.
 - 7: $k \leftarrow k + 1$.
 - 8: **end while**
-

全局收敛性分析

- 线搜索条件可选择Armijo或Wolfe.
- 若 J_k 满秩且 ∇f_k 非零, 则 d_k^{GN} 是一个下降方向:

$$\begin{aligned}(d_k)^T \nabla f(x_k) &= d_k^T J_k^T r_k = -d_k^T J_k^T J_k d_k \\ &= -\|J_k d_k\|^2 \leq 0.\end{aligned}$$

- 那么 d^k 是一个合适的线搜索方向, 全局收敛性的证明可以套用线搜索的证明。

全局收敛性分析

- 注意到，雅可比矩阵 J_k 的非奇异性很关键，在这个条件下建立收敛性.
- 具体为：假设雅可比矩阵 $J(x)$ 的奇异值一致地大于0，即存在 $\gamma > 0$ 使得

$$\|J(x)z\| \geq \gamma\|z\|, \quad \forall x \in \mathcal{N}, \quad (4)$$

其中 \mathcal{N} 是下水平集

$$\mathcal{L} = \{x \mid f(x) \leq f(x_0)\} \quad (5)$$

的一个邻域， x_0 是算法的初始点，且假设 \mathcal{L} 是有界的.

全局收敛性

Theorem

全局收敛性 如果每个残差函数 r_j 在有界下水平集(5)的一个邻域 \mathcal{N} 内是利普希茨连续可微的, 并且雅可比矩阵 $J(x)$ 在 \mathcal{N} 内满足一致满秩条件(4), 而步长满足 *Wolfe* 准则, 则对高斯-牛顿法得到的序列 $\{x_k\}$ 有

$$\lim_{k \rightarrow \infty} (J_k)^T r_k = 0.$$

证明:

- 首先, 选择有界下水平集 \mathcal{L} 的邻域 \mathcal{N} 足够小, 从而使得存在 $L > 0, \beta > 0$, 对于任何 $x, \tilde{x} \in \mathcal{N}$ 以及任意的 $j = 1, 2, \dots, m$, 以下条件被满足:

$$\begin{aligned} \|r_j(x)\| &\leq \beta, \|\nabla r_j(x)\| \leq \beta, \\ |r_j(x) - r_j(\tilde{x})| &\leq L\|x - \tilde{x}\|, \|\nabla r_j(x) - \nabla r_j(\tilde{x})\| \leq L\|x - \tilde{x}\|. \end{aligned}$$

全局收敛性证明

- 那么对任意的 $x \in \mathcal{L}$ 存在 $\tilde{\beta}$ 使得

$$\|J(x)\| = \|J(x)^T\| \leq \tilde{\beta}$$

并且 $\nabla f(x) = J(x)^T r(x)$ 是利普希茨连续函数.

- 那么Zoutendijk条件满足，成立：

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\nabla f(x_k)\|^2 < +\infty \quad (6)$$

其中 θ_k 是高斯-牛顿方向 d^k 与负梯度方向的夹角

全局收敛性证明

- 则

$$\cos \theta_k = -\frac{\nabla f(x^k)^T d^k}{\|d^k\| \|\nabla f(x^k)\|} = \frac{\|J^k d^k\|^2}{\|d^k\| \|(J^k)^T J^k d^k\|} \geq \frac{\gamma^2 \|d^k\|^2}{\tilde{\beta}^2 \|d^k\|} = \frac{\gamma^2}{\tilde{\beta}^2} > 0$$

根据6 即可得 $\nabla f(x^k) \rightarrow 0$. □

- 关键假设是一致满秩条件. 实际上, 若 J_k 不满秩, 则更新方向 d_k 有无穷多个解. 如果对解的性质不提额外要求, 则无法推出 $\cos \theta_k$ 一致地大于零. 此时收敛性可能不成立.

局部收敛性分析

- 在最优值点 x^* 附近, 当 $\sum_{i=1}^m r_i(x)\nabla^2 r_i(x)$ 较小时 ($r(x^*)$ 很小或在 x^* 附近 r 接近仿射函数), $J_k^T J_k$ 占主导位置, 高斯-牛顿方法有类似牛顿法的收敛速度。

Theorem (局部收敛性)

设 $r_i(x)$ 二阶连续可微, x^* 是最小二乘问题(1)的最优解, 海瑟矩阵 $\nabla^2 f(x)$ 和其近似矩阵 $J(x)^T J(x)$ 均在点 x^* 的一个邻域内利普希茨连续, 则当高斯-牛顿算法步长 α_k 恒为1时,

$$\|x_{k+1} - x^*\| \leq C \|((J^*)^T J^*)^{-1} H^*\| \|x_k - x^*\| + \mathcal{O}(\|x_k - x^*\|^2), \quad (7)$$

其中 $H^* = \sum_{i=1}^m r(x^*)\nabla^2 r(x^*)$ 为海瑟矩阵 $\nabla^2 f(x^*)$ 去掉 $J(x^*)^T J(x^*)$ 的部分, $C > 0$ 为常数。

局部收敛性证明

证明：

- 类似牛顿法二次收敛性的证明。根据迭代公式，

$$\begin{aligned}x^{k+1} - x^* &= x^k + d^k - x^* \\ &= \left((J^k)^T J^k \right)^{-1} \left((J^k)^T J^k (x^k - x^*) + \nabla f(x^*) - \nabla f(x^k) \right)\end{aligned}\tag{8}$$

- 由泰勒展开

$$\begin{aligned}\nabla f(x^k) - \nabla f(x^*) &= \int_0^1 J^T J(x^* + t(x^k - x^*)) (x^k - x^*) dt + \\ &\quad \int_0^1 H(x^* + t(x^k - x^*)) (x^k - x^*) dt\end{aligned}$$

其中 $J^T J(x)$ 是 $J^T(x)J(x)$ 的简写, $H(x) = \nabla^2 f(x) - J^T J(x)$ 为海瑟矩阵剩余部分

局部收敛性的证明

- 将泰勒展开式代入8式右边，取范数进行估计，有

$$\begin{aligned} & \left\| (J^k)^T J^k (x^k - x^*) + \nabla f(x^*) - \nabla f(x^k) \right\| \\ & \leq \int_0^1 \left\| (J^T J(x^k) - J^T J(x^* + t(x^k - x^*))) (x^k - x^*) \right\| dt + \\ & \quad \int_0^1 \left\| H(x^* + t(x^k - x^*)) (x^k - x^*) \right\| dt \\ & \leq \frac{L}{2} \|x^k - x^*\|^2 + C \|H^*\| \|x^k - x^*\| \end{aligned}$$

其中 L 是 $J^T J(x)$ 的利普希茨常数.

- 最后一个不等式是因为我们使用 H^* 来近似 $H(x^* + t(x^k - x^*))$ ，由连续性，存在 $C > 0$ 以及点 x^* 的一个邻域 \mathcal{N} ，对任意的 $x \in \mathcal{N}$ 有 $\|H(x)\| \leq C \|H(x^*)\|$. 将上述估计带入8式即可 □

- 1 非线性最小二乘问题
- 2 高斯-牛顿方法
- 3 Levenberg-Marquardt 方法**
- 4 大残量问题的拟牛顿算法
- 5 应用举例

Levenberg-Marquardt 方法

- 当 J_k 不满秩时, (3)有很多个解, 应该怎么更新?
- LM方法本质为信赖域方法, 更新方向为如下问题的解

$$\min_d \frac{1}{2} \|J^k d + r^k\|^2, \quad \text{s.t.} \quad \|d\| \leq \Delta_k. \quad (9)$$

- LM方法将如下近似当作信赖域方法中的 m_k :

$$m_k(d) = \frac{1}{2} \|r^k\|^2 + d^T (J^k)^T r^k + \frac{1}{2} d^T (J^k)^T J^k d. \quad (10)$$

- 同样使用 $(J^k)^T J^k$ 来近似海瑟矩阵.

Levenberg-Marquardt 方法

- 类似信赖域方法，引入如下定义来衡量 $m_k(d)$ 近似程度的好坏：

$$\rho_k = \frac{f(x^k) - f(x^k + d^k)}{m_k(0) - m_k(d^k)} \quad (11)$$

为函数值实际下降量与预估下降量（即二阶近似模型下降量）的比值。

- 如果 ρ_k 接近1，说明 $m_k(d)$ 来近似 $f(x)$ 是比较成功的，则应该扩大 Δ_k ；如果 ρ_k 非常小甚至为负，就说明我们过分地相信了二阶模型 $m_k(d)$ ，此时应该缩小 Δ_k 。
- 只有当 ρ_k 足够大，也就是对模型拟合较好时，才进行一步更新，否则不更新。

Levenberg-Marquardt 方法

Algorithm 2 Levenberg-Marquardt 方法

- 1: 给定最大半径 Δ_{\max} , 初始半径 Δ_0 , 初始点 x^0 , $k \leftarrow 0$.
- 2: 给定参数 $0 \leq \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1$, $\gamma_1 < 1 < \gamma_2$.
- 3: **while** 未达到收敛准则 **do**
- 4: 计算子问题(9)得到迭代方向 d^k .
- 5: 根据(11) 计算下降率 ρ_k .
- 6: 更新信赖域半径:

$$\Delta_{k+1} = \begin{cases} \gamma_1 \Delta_k, & \rho_k < \bar{\rho}_1, \\ \min\{\gamma_2 \Delta_k, \Delta_{\max}\}, & \rho_k > \bar{\rho}_2 \text{ 以及 } \|d^k\| = \Delta_k, \\ \Delta_k, & \text{其他.} \end{cases}$$

- 7: 更新自变量:

$$x^{k+1} = \begin{cases} x^k + d^k, & \rho_k > \eta, \\ x^k, & \text{其他.} \end{cases} \quad / * \text{ 只有下降比例足够大才更新} */$$

- 8: $k \leftarrow k + 1$.
- 9: **end while**

子问题求解

Corollary

向量 d^* 是信赖域子问题

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

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

$$(J^T J + \lambda I)d^* = -J^T r, \quad (12)$$

$$\lambda(\Delta - \|d^*\|) = 0. \quad (13)$$

- 问题(12)等价于线性最小二乘问题，具体实现时可利用系数矩阵的结构

$$\min_d \frac{1}{2} \left\| \begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix} p + \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|^2.$$

子问题求解

$$\min_d \frac{1}{2} \left\| \begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix} p + \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|^2.$$

- 在试探 λ 的值时, J 的块不变, 设 $J = QR$, 则

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

- 矩阵 $\begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix}$ 有较多的零元素, 可以使用Household变换或Givens变换完成QR分解。
- 如果矩阵 J 没有显式形式, 只能提供矩阵乘法, 则仍然可以用截断共轭梯度法。

Theorem

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} 内是利普希茨连续可微的。假设对于任意的 k ，子问题(9)的近似解 d_k 满足

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

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

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

收敛性分析

证明：

- 根据 $r_j(x)$ 的连续性假设，可以推出存在 $M > 0$ ，使得 $\|J_k^T J^k\| \leq M$ 对任意的 k 成立。注意到 f 是有下界的，可以直接套用信赖域算法全局收敛性的证明。 □
- 事实上，为保证全局收敛性，精确求解子问题是不必要的。
- 关于局部收敛性，同样使用 $(J^k)^T J^k$ 来近似海瑟矩阵，与高斯-牛顿方法有着类似的局部收敛性质

- 信赖域型LM方法本质上是固定信赖域半径 Δ , 通过迭代寻找满足条件的乘子 λ , 每一步迭代需要求解线性方程组

$$(J^T J + \lambda I) d = -J^T r$$

该步计算代价较大。

- 注意到在LM方法中, 由于 $J_k^T J_k \succ 0$, 那么有 $-\lambda_1 < 0$, 此时有 $\lambda > -\lambda_1$, 因此若 λ 越大, d 的模长就越小。
- 调整 λ 的大小等价于调整信赖域半径的大小, 这意味着, Δ 被 λ 隐式决定。

- LM的更新基于 Δ ，LMF的更新直接基于 λ ，每一步求解子问题：

$$\min_d \|Jd + r\|_2^2 + \lambda \|d\|_2^2.$$

- 调整 λ 的原则可以参考信赖域半径的调整原则
- 考虑参数 ρ_k

$$\rho_k = \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)} \quad (14)$$

较大可以减小下一步的 λ ，较小可以增大下一步的 λ 。

Algorithm 3 LMF 方法

- 1: 给定初始点 x_0 , 初始乘子 λ_0 , $k \leftarrow 0$.
- 2: 给定参数 $0 \leq \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1$, $\gamma_1 < 1 < \gamma_2$.
- 3: **while** 未达到收敛准则 **do**
- 4: 求解LM方程 $((J_k)^T J_k + \lambda I)d = -(J_k)^T r_k$ 得到迭代方向 d_k .
- 5: 根据(14)式计算下降率 ρ_k .
- 6: 更新乘子:

$$\lambda_{k+1} = \begin{cases} \gamma_2 \lambda_k, & \rho_k < \bar{\rho}_1, & /* \text{扩大乘子 (缩小信赖域半径)} */ \\ \gamma_1 \lambda_k, & \rho_k > \bar{\rho}_2, & /* \text{缩小乘子 (扩大信赖域半径)} */ \\ \lambda_k, & \text{其他.} & /* \text{乘子不变} */ \end{cases}$$

- 7: 更新自变量:

$$x_{k+1} = \begin{cases} x_k + d_k, & \rho_k > \eta, & /* \text{只有下降比例足够大才更新} */ \\ x_k, & \text{其他.} \end{cases}$$

- 8: $k \leftarrow k + 1$.
- 9: **end while**

- 1 非线性最小二乘问题
- 2 高斯-牛顿方法
- 3 Levenberg-Marquardt 方法
- 4 大残量问题的拟牛顿算法
- 5 应用举例

大残量问题的拟牛顿算法

- 大残量问题中，海瑟矩阵的第二部分不可忽视，此时高斯-牛顿法和LM方法可能只有线性的收敛速度。
- 此时如果直接使用牛顿法，则开销太大；直接使用拟牛顿法，又似乎忽略了问题的特殊结构。
- 重新写出海瑟矩阵：

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

第一项是容易求解，可以保留。第二项不易求解但不可忽略，用拟牛顿法进行近似。

大残量问题的拟牛顿算法

- 使用 B_k 来表示 $\nabla^2 f(x_k)$ 的近似矩阵, T_k 表示 $\sum_{j=1}^m r_j(x_k) \nabla^2 r_j(x_k)$ 的近似, 即

$$B_k = (J_k)^T J_k + T_k,$$

- 目标为

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

- 记 $s_k = x_{k+1} - x_k$, $T_k + 1$ 应该尽量保留原海瑟矩阵的性质

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

大残量问题的拟牛顿算法

- 拟牛顿条件为：

$$T_{k+1}s_k = (J_{k+1})^T r_{k+1} - (J_k)^T r_{k+1}$$

- Dennis, Gay, 和Welsch给出的一种更新格式为：

$$T_{k+1} = T_k + \frac{(y^\# - T_k s_k) y^T + y (y^\# - T_k s_k)^T}{y^T s_k} - \frac{(y^\# - T_k s_k)^T s_k}{(y^T s)^2} y y^T$$

其中

$$\begin{aligned} s_k &= x_{k+1} - x_k \\ y &= J_{k+1}^T r_{k+1} - J_k^T r_k \\ y^\# &= J_{k+1}^T r_{k+1} - J_k^T r_{k+1} \end{aligned}$$

- 1 非线性最小二乘问题
- 2 高斯-牛顿方法
- 3 Levenberg-Marquardt 方法
- 4 大残量问题的拟牛顿算法
- 5 应用举例

应用实例

- 相位恢复是最小二乘法的重要应用, 原始模型为

$$\min_{z \in \mathbb{C}^n} f(z) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{j=1}^m \left(|\bar{a}_j^T z|^2 - b_j \right)^2, \quad (15)$$

其中 $a_j \in \mathbb{C}^n$ 是已知的采样向量, $b_j \in \mathbb{R}$ 是观测的模长

- 注意到此时 f 关于 z 并不是全纯函数, 因此我们考虑Wirtinger导数表示梯度和雅可比矩阵。
- 对任意的实值或复值函数 f , $z = x + iy$,

$$f(z) = f(x, y) = u(x, y) + iv(x, y)$$

可以写成 $f(z, \bar{z})$ 的形式。

应用实例

- 当实函数 u, v 关于 x, y 都可导时, 固定 \bar{z} , 复函数 $f(z, \bar{z})$ 关于 z 是全纯的; 固定 z , f 关于 \bar{z} 是全纯的;
- 在本问题中

$$r_j(z) = (|\bar{a}_j^T z|^2 - b_j)^2 = (\bar{z}^T a_j \bar{a}_j^T z - b_j)^2 = r_j(z, \bar{z})$$

固定 \bar{z} , 关于 z 全纯; 固定 z , 关于 \bar{z} 全纯。

- 据此可导出Wirtinger导数。定义共轭表示:

$$\mathbf{z} = \begin{bmatrix} z \\ \bar{z} \end{bmatrix} \in \mathbb{C}^n \times \mathbb{C}^n, \quad z = x + iy \quad \bar{z} = x - iy$$

替代 $(x, y) \in \mathbb{R}^{2n}$ 的表示方法。

- 对于 $f(z)$ ，定义：

$$\nabla f(\mathbf{z}) = \left[\frac{\partial f}{\partial z}, \frac{\partial f}{\partial \bar{z}} \right]^*$$

其中

$$\begin{aligned} \frac{\partial f}{\partial z} &:= \left. \frac{\partial f(z, \bar{z})}{\partial z} \right|_{\bar{z} = \text{constant}} = \left[\frac{\partial f}{\partial z_1}, \frac{\partial f}{\partial z_2}, \dots, \frac{\partial f}{\partial z_n} \right]_{\bar{z} = \text{constant}} \\ \frac{\partial f}{\partial \bar{z}} &:= \left. \frac{\partial f(z, \bar{z})}{\partial \bar{z}} \right|_{z = \text{constant}} = \left[\frac{\partial f}{\partial \bar{z}_1}, \frac{\partial f}{\partial \bar{z}_2}, \dots, \frac{\partial f}{\partial \bar{z}_n} \right]_{z = \text{constant}} \end{aligned}$$

- 注：这里遵循求导为行向量，梯度为列向量

- 在本问题中，有

$$\frac{\partial f}{\partial \mathbf{z}} = \sum_{j=1}^m \left(|\bar{\mathbf{a}}_j^T \mathbf{x}|^2 - b_j \right) \bar{\mathbf{z}}^T \mathbf{a}_j \bar{\mathbf{a}}_j^T$$

以及

$$\frac{\partial f}{\partial \bar{\mathbf{z}}} = \sum_{j=1}^m \left(|\bar{\mathbf{a}}_j^T \mathbf{x}|^2 - b_j \right) \mathbf{z}^T \bar{\mathbf{a}}_j \mathbf{a}_j^T$$

最后

$$\nabla f(\mathbf{z}) = \left[\frac{\partial f}{\partial \mathbf{z}}, \frac{\partial f}{\partial \bar{\mathbf{z}}} \right]^*$$

- 雅可比矩阵和高斯-牛顿矩阵分别为

$$J(\mathbf{z}) = \overline{\begin{bmatrix} a_1(\bar{a}_1^T \mathbf{z}), & a_2(\bar{a}_2^T \mathbf{z}), & \cdots, & a_m(\bar{a}_m^T \mathbf{z}) \\ \bar{a}_1(a_1^T \bar{\mathbf{z}}), & \bar{a}_2(a_2^T \bar{\mathbf{z}}), & \cdots, & \bar{a}_m(a_m^T \bar{\mathbf{z}}) \end{bmatrix}}^T,$$
$$\Psi(\mathbf{z}) \stackrel{\text{def}}{=} \overline{J(\mathbf{z})}^T J(\mathbf{z}) = \sum_{j=1}^m \begin{bmatrix} |\bar{a}_j^T \mathbf{z}|^2 a_j \bar{a}_j^T & (\bar{a}_j^T \mathbf{z})^2 a_j a_j^T \\ (\bar{a}_j^T \mathbf{z})^2 \bar{a}_j \bar{a}_j^T & |\bar{a}_j^T \mathbf{z}|^2 \bar{a}_j \bar{a}_j^T \end{bmatrix}.$$

- 因此在第k步，高斯-牛顿法求解方程

$$\Psi(\mathbf{z}^k) d^k = -\nabla f(\mathbf{z}^k)$$

得到更新方向 d^k

应用实例

- LM 方法求解正则化方程

$$(\Psi(\mathbf{z}^k) + \lambda_k) d^k = -\nabla f(\mathbf{z}^k), \quad (16)$$

- λ_k 是与 $f(\mathbf{z}^k)$ 相关的参数，选取

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

- 当 $f(\mathbf{z}^k) \geq \frac{1}{900n}\|\mathbf{z}^k\|_2^2$ 时，参数 $\lambda_k = 70000n\sqrt{nf(\mathbf{z}^k)}$ 能够保证算法有全局Q-线性收敛速度。
- 利用共轭梯度法求解线性方程(16)，使得

$$\|(\Psi(\mathbf{z}^k) + \lambda_k) d^k + \nabla f(\mathbf{z}^k)\| \leq \eta_k \|\nabla f(\mathbf{z}^k)\|,$$

其中 $\eta_k > 0$ 是人为设置的参数。

应用实例

- 考虑编码衍射模型，其中信号采集的格式为

$$b_j = \left| \sum_{t=0}^{n-1} z_t \bar{d}_l(t) e^{-i2\pi kt/n} \right|^2, \quad j = (l, k), \quad 0 \leq k \leq n-1, \quad 1 \leq l \leq L.$$

- 对给定的 l ，我们采集在波形 d_l 下信号 $\{z_t\}$ 的衍射图的模长。通过改变 l 和相应的波形 d_l ，可以生成一系列编码衍射图。
- 这里假设 $d_l, l = 0, 1, \dots, L$ 是独立同分布的随机向量来模拟实际场景。具体地，令 $d_l(t) = c_1 c_2$ ，其中

$$c_1 = \begin{cases} +1, & \text{依概率 } 0.25, \\ -1, & \text{依概率 } 0.25, \\ +i, & \text{依概率 } 0.25, \\ -i, & \text{依概率 } 0.25, \end{cases} \quad c_2 = \begin{cases} \frac{\sqrt{2}}{2}, & \text{依概率 } 0.8, \\ \sqrt{3}, & \text{依概率 } 0.2. \end{cases}$$

应用实例

- 真实信号 x 取为两张自然图片，分别为博雅塔和华表的图片，如下图所示。这里图片可以看成 $m \times n$ 矩阵，其中行、列指标表示像素点所在位置，取值表示像素点的灰度值。选取 $L = 20$ ，并收集相应的衍射图模长。



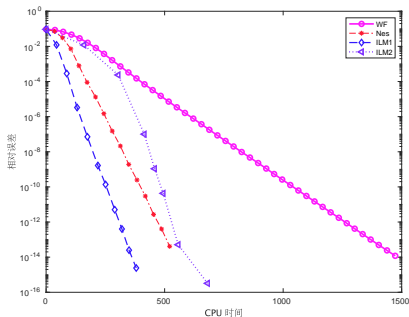
(a) 博雅塔，图片像素为 601×541



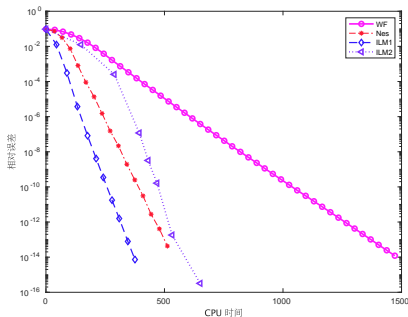
(b) 华表，图片像素为 601×541

应用实例

- 分别测试不精确求解正则化方程(16) ($\eta_k = 0.1$) 的LM方法 (ILM1) 以及更精确 ($\eta_k = \min\{0.1, \|\nabla f(\mathbf{z}^k)\|\}$) 的LM方法 (ILM2), 求解Wirtinger梯度下降方法 (WF) 以及其加速版本Nesterov加速算法 (Nes). 下图给出了不同算法的收敛情况, 其中横坐标为CPU时间, 纵坐标为当前迭代点 \mathbf{z}^k 与真实信号 \mathbf{z} 的相对误差, 即 $\min_{\phi \in [0, 2\pi]} \frac{1}{\|\mathbf{z}\|} \|\mathbf{z}^k - \mathbf{z}e^{i\phi}\|$.



(a) 博雅塔



(b) 华表