提纲

最小二乘问题

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

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

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

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

- 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：

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

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

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

最小二乘问题

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

- $\nabla^2 f(x)$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：

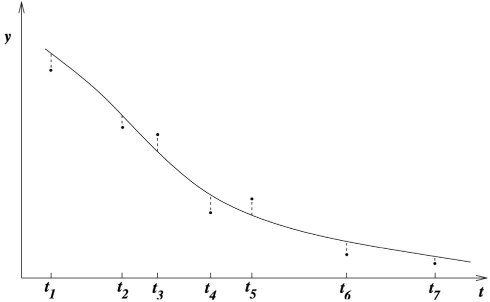
$$A^T A x = A^T b$$

- 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

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

- 模型$\phi(t;x)=x_1+tx_2+t_3^x+x_4e^{-x_5t}$ 依赖于参数向量x。
- $(t_1,y_1),(t_2,y_2),\cdots,(t_n,y_n)$是数据点
- 目标为寻找合适的模型参数x



提纲

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：

$$J_k^T J_k d_k^{GN} = -J_k^T r_k. \tag{3}$$

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：

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

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：

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

- 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

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

其中$\mathcal{N}$是下水平集

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

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

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

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

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$j=1,2,\cdots,m$, 以下条件被满足：

$$\|r_j(x)\|\leqslant\beta, \|\nabla r_j(x)\|\leqslant\beta,$$
$$|r_j(x)-r_j(\tilde{x})|\leqslant L\|x-\tilde{x}\|, \|\nabla r_j(x)-\nabla r_j(\tilde{x})\|\leqslant L\|x-\tilde{x}\|.$$

全局收敛性证明

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

$$\|J(x)\| = \left\|J(x)^{\mathrm{T}}\right\| \leqslant \tilde{\beta}$$

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

- 那么Zoutendijk条件满足，成立：

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

其中θ_k 是高斯– 牛顿方向d^k 与负梯度方向的夹角

- 则

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

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

- 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$J(x)^\mathrm{T}J(x)$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，

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

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

证明：

- 类似牛顿法二次收敛性的证明。根据迭代公式,

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

- 由泰勒展开

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

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

- 将泰勒展开式代入8式右边，取范数进行估计，有

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

  其中L 是$J^{\mathrm{T}}J(x)$ 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$H\left(x^* + t\left(x^k - x^*\right)\right)$, 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$\|H(x)\| \leqslant C \left\| H\left(x^*\right) \right\|$. 将上述估计带入8 式即可 □

提纲

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

$$\min_{d} \quad \frac{1}{2}\|J^k d + r^k\|^2, \quad \text{s.t.} \quad \|d\| \le \Delta_k. \tag{9}$$

- LM 方法将如下近似当作信赖域方法中的m_k:

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

- 同样使用$(J^k)^{\mathrm{T}}J^k$来近似海瑟矩阵.

- 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：

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

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。

Algorithm 2 Levenberg-Marquardt 方法

给定最大半径$\Delta_{\max}$，初始半径Δ_0，初始点x^0，$k \leftarrow 0$.

给定参数$0 \le \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1$，$\gamma_1 < 1 < \gamma_2$.

while 未达到收敛准则 **do**

  计算子问题(9)得到迭代方向d^k.

  根据(11) 计算下降率ρ_k.

  更新信赖域半径：

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

  更新自变量：

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

  $k \leftarrow k + 1$.

end while

Corollary

向量d^*是信赖域子问题

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

的解当且仅当d^*是可行解并且存在数$\lambda \geq 0$使得

$$(J^{\mathrm{T}}J + \lambda I)d^* = -J^{\mathrm{T}}r, \tag{12}$$

$$\lambda(\Delta - \|d^*\|) = 0. \tag{13}$$

- 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

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

子问题求解

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

- 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$\eta \in \left(0, \frac{1}{4}\right)$，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

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

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

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

证明：

- 根据$r_j(x)$的连续性假设，可以推出存在$M>0$，使得$\|J_k^T J^k\| \le M$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。 $\square$

- 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$(J^k)^{\mathrm{T}} J^k$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

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

  该步计算代价较大。

- 注意到在LM方法中，由于$J_k^TJ_k\succ 0$，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。

LMF

- LM的更新基于Δ，LMF的更新直接基于λ，每一步求解子问题：

$$\min_{d} \quad \|Jd+r\|_2^2+\lambda\|d\|_2^2.$$

- 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

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

较大可以减小下一步的λ，较小可以增大下一步的λ。

LMF

Algorithm 3 LMF 方法

给定初始点x_0，初始乘子λ_0，$k \leftarrow 0$.

给定参数$0 \le \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1$，$\gamma_1 < 1 < \gamma_2$.

while 未达到收敛准则 **do**

  求解LM 方程$((J_k)^{\mathrm{T}} J_k + \lambda I)d = -(J_k)^{\mathrm{T}} r_k$得到迭代方向$d_k$.

  根据(14)式计算下降率ρ_k.

  更新乘子：

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

  更新自变量：

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

  $k \leftarrow k + 1$.

end while

提纲

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：

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

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

$$B_k = (J_k)^{\mathrm{T}} J_k + T_k,$$

- 目标为

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

- 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

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

大残量问题的拟牛顿算法

- 拟牛顿条件为：

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

- Dennis, Gay, 和Welsch给出的一种更新格式为：

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

其中

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

提纲

应用实例

- 相位恢复是最小二乘法的重要应用，原始模型为

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

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

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

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：

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

替代$(x,y) \in \mathbb{R}^{2n}$的表示方法。

应用实例

- 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

$$\frac{\partial f}{\partial z} := \left. \frac{\partial f(z,\bar{z})}{\partial z} \right|_{\bar{z}=\text{ constant}} = \left[\frac{\partial f}{\partial z_1}, \frac{\partial f}{\partial z_2}, \ldots, \frac{\partial f}{\partial z_n} \right]_{\bar{z}=\text{ constant}}$$

$$\frac{\partial f}{\partial \bar{z}} := \left. \frac{\partial f(z,\bar{z})}{\partial \bar{z}} \right|_{z=\text{ constant}} = \left[\frac{\partial f}{\partial \bar{z}_1}, \frac{\partial f}{\partial \bar{z}_2}, \ldots, \frac{\partial f}{\partial \bar{z}_n} \right]_{z=\text{ constant}}$$

- 注：这里遵循求导为行向量，梯度为列向量

应用实例

- 在本问题中，有

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

以及

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

最后

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

应用实例

- 雅可比矩阵和高斯– 牛顿矩阵分别为

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

$$\Psi(\mathbf{z}) \stackrel{\text{def}}{=\!=} \overline{J(\mathbf{z})}^{\mathrm{T}} J(\mathbf{z}) = \sum_{j=1}^{m} \begin{bmatrix} |\bar{a}_j^{\mathrm{T}} z|^2 a_j \bar{a}_j^{\mathrm{T}} & (\bar{a}_j^{\mathrm{T}} z)^2 a_j a_j^{\mathrm{T}} \\ \overline{(\bar{a}_j^{\mathrm{T}} z)^2} \bar{a}_j \bar{a}_j^{\mathrm{T}} & |\bar{a}_j^{\mathrm{T}} z|^2 \bar{a}_j a_j^{\mathrm{T}} \end{bmatrix}.$$

- 因此在第k步，高斯– 牛顿法求解方程

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

得到更新方向d^k

应用实例

- LM 方法求解正则化方程

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

- λ_k是与$f(\mathbf{z}^k)$相关的参数，选取

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

- 当$f(z^k)\geq\frac{1}{900n}\|z^k\|_2^2$ 时，参数$\lambda_k=70000n\sqrt{nf(z^k)}$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

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

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

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

- 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$d_l, l = 0, 1, \cdots, L$ 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

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

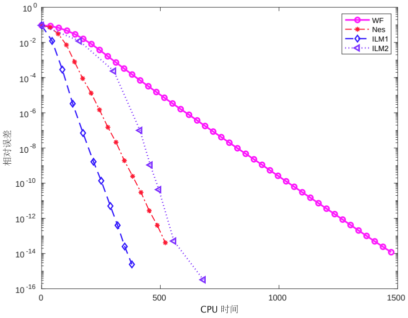
应用实例

- 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．

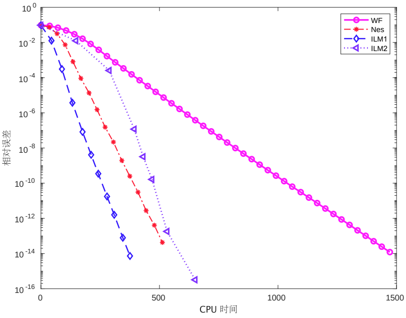
(a) 博雅塔，图片像素为601×541

(b) 华表，图片像素为601×541

- 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z的相对误差，即$\min_{\phi\in[0,2\pi]} \frac{1}{\|z\|}\|\mathbf{z}^k - ze^{i\phi}\|$．



(a) 博雅塔



(b) 华表