第五章 无约束优化算法

修贤超

https://xianchaoxiu.github.io

目录

信赖域算法框架

- 在当前迭代点 x^k 建立局部模型，求出最优解

$$d^k = \arg\min_{d} \quad (g^k)^\top d + d^\top B d \quad \text{s.t.} \quad \|d\|_2 \le \Delta_k$$

- 更新模型信赖域的半径
  - 模型足够好 $\Rightarrow$ 增大半径
  - 模型比较差 $\Rightarrow$ 缩小半径
  - 否则半径不变
- 对模型进行评价
  - 好 $\Rightarrow$ 子问题的解即下一个迭代点
  - 差 $\Rightarrow$ 迭代点不改变

信赖域子问题

- 根据带拉格朗日余项的泰勒展开

$$f(x^k + d) = f(x^k) + \nabla f(x^k)^\top d + \frac{1}{2} d^\top \nabla^2 f(x^k + td)d$$

- 利用 $f(x)$ 的二阶近似来刻画 $f(x)$ 在点 x^k 处的性质

$$m_k(d) = f(x^k) + \nabla f(x^k)^\top d + \frac{1}{2} d^\top B^k d$$

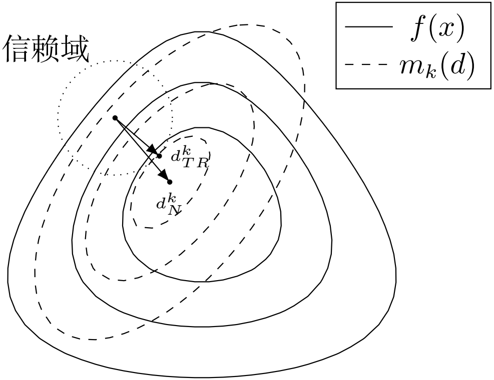
- 由于泰勒展开的局部性，需对上述模型添加信赖域约束

$$\Omega_k = \{x^k + d \mid \|d\| \leqslant \Delta_k\}$$

信赖域子问题

- 信赖域算法每一步都需要求解如下子问题

$$\min_{d\in\mathbb{R}^n} \quad m_k(d) \quad \text{s.t.} \quad \|d\| \le \Delta_k \tag{1}$$



模型近似程度好坏的的衡量

- 引入
$$\rho_k = \frac{f(x^k) - f(x^k + d^k)}{m_k(0) - m_k(d^k)} \tag{2}$$
- 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，则扩大 Δ_k
- 如果 ρ_k 非常小甚至为负，说明过分地相信了二阶模型 $m_k(d)$，则缩小 Δ_k

算法 5.4　信赖域算法

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 未达到停机准则 **do**

计算子问题 (1) 得到迭代方向 d^k

根据 (2) 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

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

$k \leftarrow k + 1$

end while

目录

非线性最小二乘问题

- 考虑最小二乘问题

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

- 记 $r(x) = (r_1(x), r_2(x), ..., r_m(x))^\top$，问题可以表述为

$$\min \quad f(x) = \frac{1}{2}\|r(x)\|_2^2$$

- 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

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

最小二乘问题

- $f(x)$ 的梯度和海瑟矩阵

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x)\nabla r_j(x) = J(x)^\top r(x)$$

$$\begin{aligned}\nabla^2 f(x) &= \sum_{j=1}^{m} \nabla r_j(x)\nabla r_j(x) + \sum_{i=1}^{m} r_i(x)\nabla^2 r_i(x)\\ &= J(x)^\top J(x) + \sum_{i=1}^{m} r_i(x)\nabla^2 r_i(x)\end{aligned}$$

- 小残差 $\Rightarrow$ 高斯-牛顿方法和 Levenberg-Marquardt 方法
- 大残差 $\Rightarrow$ 引入带结构的拟牛顿方法

高斯-牛顿方法

- 使用近似 $\nabla^2 f_k \approx J_k^\top J_k$，省略 $\nabla^2 r_j$ 的计算，减少了计算量
- 高斯-牛顿法的迭代方向 d_k^{GN} 满足

$$J_k^\top J_k d_k^{GN} = -J_k^\top r_k$$

- 另一种理解：在点 x_k 处，考虑近似 $r(x_k + d) \approx r_k + J_k d$ 得到

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

- 然后更新 $x_{k+1} = x_k + \alpha_k d_k$

算法 5.9　高斯-牛顿方法

```
给定初值 x_0，k ← 0
while 未达到停机准则 do
计算残差向量 r_k，雅可比矩阵 J_k
求解线性最小二乘问题 min_d 1/2||r_k + J_k d||^2 确定下降方向 d_k
使用线搜索准则计算步长 α_k
更新 x_{k+1} = x_k + α_k d_k
k ← k + 1
end while
```

Levenberg-Marquardt (LM) 方法

- LM 方法本质为信赖域方法，更新方向为如下问题的解

$$\min_{d} \quad \frac{1}{2}\|J^k d + r^k\|^2 \quad \text{s.t.} \quad \|d\| \le \Delta_k \tag{3}$$

- 将如下近似当作信赖域方法中的 m_k

$$m_k(d) = \frac{1}{2}\|r^k\|^2 + d^\top (J^k)^\top r^k + \frac{1}{2} d^\top (J^k)^\top J^k d$$

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 未达到停机准则 **do**

计算子问题 (3) 得到迭代方向 d^k

根据 (2) 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

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

$k \leftarrow k + 1$

end while

- **推论 5.4** 向量 d^* 是信赖域子问题

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

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

$$(J^\top J + \lambda I)d^* = -J^\top r$$
$$\lambda(\Delta - \|d^*\|) = 0$$

- 实际上, $(J^\top J + \lambda I)d^* = -J^\top r$ 是最小二乘问题的最优性条件

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

LMF 方法

- 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

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

- LM 的更新基于 Δ，LMF 的更新直接基于 λ，每一步求解子问题

$$\min_d \quad \frac{1}{2}\|J^k d + r^k\|^2 \quad \text{s.t.} \quad \|d\| \le \Delta$$

$$\Downarrow$$

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

- 调整 λ 的原则可以参考信赖域半径的调整原则

算法 5.10　LMF 方法

```
给定初始点 x_0，初始乘子 λ_0，k ← 0
给定参数 0 ≤ η < ρ̄_1 < ρ̄_2 < 1，γ_1 < 1 < γ_2
while 未达到停机准则 do
    求解 LM 方程 ((J_k)^⊤ J_k + λI)d = −(J_k)^⊤ r_k 得到迭代方向 d_k
    根据 (2) 式计算下降率 ρ_k
    更新信赖域半径
```

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

```
    更新自变量
```

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

```
    k ← k + 1
end while
```

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

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

- 记 $s_k = x_{k+1} - x_k$，T_{k+1} 应保留原海瑟矩阵的性质

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

大残量问题的拟牛顿算法

- 拟牛顿条件为

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

- Dennis, Gay 和 Welsch 给出的一种更新格式

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

其中

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

应用实例: 相位恢复

- 相位恢复是最小二乘法的重要应用, 原始模型为

$$\min_{z\in\mathbb{C}^n} \quad f(z) = \frac{1}{2}\sum_{j=1}^{m}(|\bar{a}_j^\top z|^2 - b_j)^2$$

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

- 根据 Wirtinger 导数知

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

其中

$$\frac{\partial f}{\partial z} = \sum_{j=1}^{m}(|\bar{a}_j^\top x|^2 - b_j)\bar{z}^\top a_j\bar{a}_j^\top, \quad \frac{\partial f}{\partial \bar{z}} = \sum_{j=1}^{m}(|\bar{a}_j^\top x|^2 - b_j)z^\top \bar{a}_j a_j^\top$$

应用实例: 相位恢复

- 雅可比矩阵和高斯 – 牛顿矩阵分别为

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

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

- 在第 k 步，高斯 – 牛顿法求解方程

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

应用实例: 相位恢复

- LM 方法求解正则化方程

$$(\Psi(\mathbf{z}^k)+\lambda_k)d^k = -\nabla f(\mathbf{z}^k) \tag{4}$$

- 选取

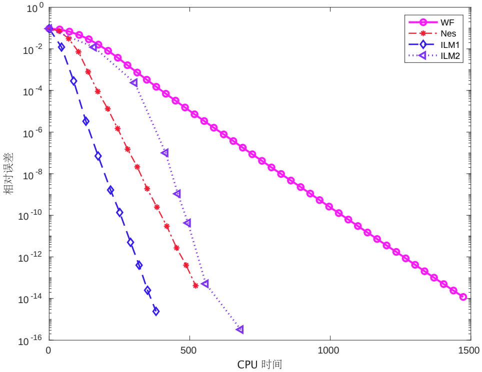
$$\lambda_k = \begin{cases} 70000n\sqrt{nf(z^k)}, & f(z^k) \geq \frac{1}{900n}\|z^k\|_2^2 \\ \sqrt{f(z^k)}, & \text{其他} \end{cases}$$

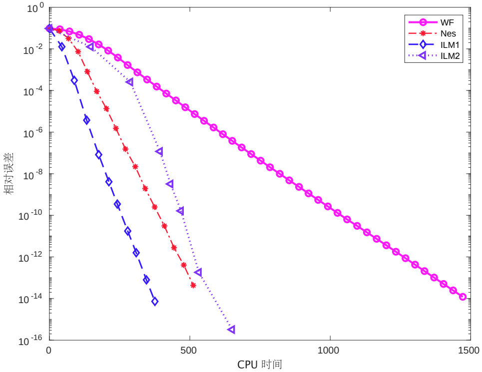
- 利用共轭梯度法求解线性方程 (4)，使得

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

应用实例: 相位恢复

- WF 求解 Wirtinger 梯度下降方法
- LM ILM1 ($\eta_k = 0.1$), ILM2 ($\eta_k = \min\{0.1, \|\nabla f(\mathbf{z}^k)\|\}$), Nes (Nesterov 加速)





Q&A

Thank you!

感谢您的聆听和反馈