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

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

- 求出局部模型的最优解
- 更新模型信赖域的半径
  - 模型足够好 $arrow$ 增大半径
  - 模型比较差 $arrow$ 缩小半径
  - 否则半径不变
- 对模型进行评价
  - 好 $arrow$ 子问题的解即下一个迭代点
  - 差 $arrow$ 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

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

其中 $t\in(0,1)$ 为和 d 有关的正数

- 和牛顿法相同，利用 $f(x)$ 的二阶近似来刻画 $f(x)$ 在点 x^k 处的性质

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

其中 B^k 是对称矩阵，并且是海瑟矩阵的近似矩阵

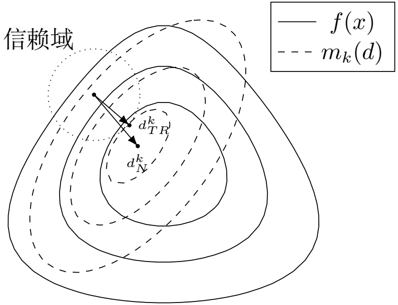
- 由于泰勒展开的局部性, 需对上述模型添加信赖域约束

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

其中 $\varDelta_k>0$ 是信赖域半径

- 因此信赖域算法每一步都需要求解如下子问题

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



模型近似程度好坏的的衡量

- 引入如下定义来衡量 $m_k(d)$ 近似程度的好坏

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

- 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

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**

计算迭代方向 d^k

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

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

$k \leftarrow k + 1$

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

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

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

- 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 $J(x)^\top J(x)$ 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

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

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

- 然后更新 $x_{k+1} = x_k + \alpha_k d_k$

高斯-牛顿方法

```
给定初值 $x_0$，$k \leftarrow 0$
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 $\min_d \quad \frac{1}{2}\|J_k d + r_k\|^2$ 确定下降方向 $d_k$
    使用线搜索准则计算步长 $\alpha_k$
    更新 $x_{k+1} = x_k + \alpha_k d_k$
    $k \leftarrow k + 1$
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

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

- LM 方法将如下近似当作信赖域方法中的 m_k

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

- 同样使用 $(J^k)^\top J^k$ 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

$$\rho_k = \frac{f(x^k) - f(x^k + d^k)}{m_k(0) - m_k(d^k)} \tag{2}$$

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

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

$k \leftarrow k + 1$
end while

子问题求解

- **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

$$\begin{aligned} &(J^\top J + \lambda I)d^* = -J^\top r \\ &\lambda(\Delta - \|d^*\|) = 0 \end{aligned}$$

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

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

LMF

- 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

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

- 调整 λ 的大小等价于调整信赖域半径的大小，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

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

- 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

$$\rho_k = \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)} \tag{3}$$

LMF

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 未达到停机准则 **do**
求解 LM 方程 $((J_k)^\top J_k + \lambda I)d = -(J_k)^\top r_k$ 得到迭代方向 d_k
根据(3)式计算下降率 ρ_k
更新信赖域半径

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

/* 扩大乘子（缩小信赖域半径）*/
/* 缩小乘子（扩大信赖域半径）*/
/* 乘子不变 */

更新自变量

$$x_{k+1} = \begin{cases} x_k + d_k, & \rho_k > \eta \\ x_k, & \text{其他} \end{cases}$$

/* 只有下降比例足够大才更新 */

$k \leftarrow k + 1$
end while

应用实例

- 相位恢复是最小二乘法的重要应用，原始模型为

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

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

- 对于 $f(x)$，定义

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

其中

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

- 在第 k 步，高斯 – 牛顿法求解方程

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

- LM 方法求解正则化方程

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

- λ_k 是与 $f(\mathbf{z}^k)$ 相关的参数，选取

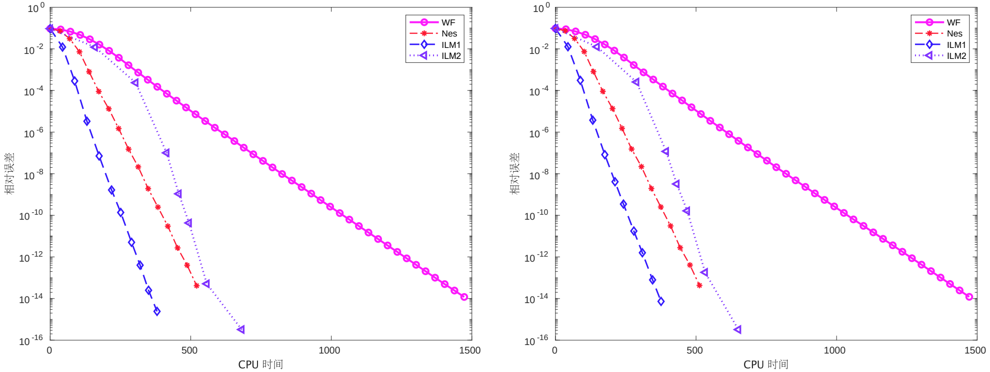
$$\lambda_k = \begin{cases} 70000n\sqrt{nf(z^k)}, & f(z^k) \geq \frac{1}{900n}\|z^k\|_2^2 \\ \sqrt{f(z^k)}, & \text{其他} \end{cases}$$

- 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

- 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