Optimal Theory and Method

最优化理论与方法

翟春杰

杭州电子科技大学 自动化学院

科技馆512

Email: cjzhai@hdu.edu.cn

目录

- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 最小二乘法
- 信赖域法

目录

- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 最小二乘法
- 信赖域法

■ 牛顿法

➢牛顿法的特点

- 当初始点远离极小点时，牛顿法可能不收敛 **Why?**

原因之一：牛顿方向不一定是下降方向

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

$$\nabla f(x^{(k)})^T d^{(k)} = -\nabla f(x^{(k)})^T \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

原因之二：即使目标函数值下降，得到的值也未必是沿牛顿方向的极小点。步长为1

- **改进：阻尼牛顿法—— 沿牛顿方向一维搜索**

拟牛顿法

■ 算法产生

◆ 牛顿法

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \boxed{\left(\nabla^2 f(\mathbf{x}^{(k)})\right)^{-1}} \nabla f(\mathbf{x}^{(k)})$$

◆ 思路：用不包含二阶导数的矩阵近似海塞矩阵

◆ 关键：近似矩阵的构造

◆ 构造近似矩阵的方法不同，产生不同的拟牛顿法

◆ 拟牛顿法是一类算法，而不是一个算法：秩1校正法、DFP算法、BFGS法

拟牛顿法

■ 拟牛顿条件

目标函数在点 $\mathbf{x}^{(k+1)}$ 展成泰勒级数，并取二阶近似

$$f(\mathbf{x}) \approx f(\mathbf{x}^{(k+1)}) + \nabla f(\mathbf{x}^{(k+1)})^{\mathrm{T}}(x - \mathbf{x}^{(k+1)}) + \frac{1}{2}(x - \mathbf{x}^{(k+1)})^{\mathrm{T}} \nabla^2 f(\mathbf{x}^{(k+1)})(x - \mathbf{x}^{(k+1)})$$

对上式≈两边同时求关于 x 的梯度，得

$$\nabla f(\mathbf{x}) \approx \nabla f(\mathbf{x}^{(k+1)}) + \nabla^2 f(\mathbf{x}^{(k+1)})(x - \mathbf{x}^{(k+1)})$$

令 $\mathbf{x} = \mathbf{x}^{(k)}$，则 $\nabla f(\mathbf{x}^{(k)}) \approx \nabla f(\mathbf{x}^{(k+1)}) + \nabla^2 f(\mathbf{x}^{(k+1)})(\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)})$

记 $\boxed{p^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}$ $\boxed{q^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})}$

则有 $q^{(k)} \approx \nabla^2 f(\mathbf{x}^{(k+1)}) p^{(k)}$

设海塞矩阵可逆 $p^{(k)} \approx \nabla^2 f(\mathbf{x}^{(k+1)})^{-1} q^{(k)}$

令 H_{k+1} 满足 $\boxed{p^{(k)} = H_{k+1} q^{(k)}}$ 拟牛顿条件：pH值+1

拟牛顿法

■ 近似矩阵的构造

- 当$\nabla^2 f(\mathbf{x}^{(k+1)})^{-1}$是n阶对称正定矩阵时，满足拟牛顿条件的矩阵$H_{k+1}$也是n阶对称正定矩阵

- 构造矩阵H_k的策略

• 初始选择任一n阶对称正定矩阵，一般$H_1 = I$

• 对近似矩阵迭代计算

$$H_{k+1} = H_k + \Delta H_k$$

校正矩阵 ΔH_k

拟牛顿法

■ DFP（变尺度法）

- **DFP（变尺度法）计算校正矩阵**

$$\nabla \boldsymbol{H}_k = \frac{\boldsymbol{p}^{(k)}\left(\boldsymbol{p}^{(k)}\right)^{\mathrm{T}}}{(\boldsymbol{p}^{(k)})^{\mathrm{T}}\boldsymbol{q}^{(k)})} - \frac{\boldsymbol{H}_k\boldsymbol{q}^{(k)}\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k}{(\boldsymbol{q}^{(k)})^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)})}$$

- **DFP公式**

$$\boxed{\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{p}^{(k)}\left(\boldsymbol{p}^{(k)}\right)^{\mathrm{T}}}{(\boldsymbol{p}^{(k)})^{\mathrm{T}}\boldsymbol{q}^{(k)})} - \frac{\boldsymbol{H}_k\boldsymbol{q}^{(k)}\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k}{(\boldsymbol{q}^{(k)})^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)})}}$$

$$\boxed{\boldsymbol{p}^{(k)} = \boldsymbol{H}_{k+1}\boldsymbol{q}^{(k)}}$$

拟牛顿条件成立吗?

拟牛顿法

■ DFP（变尺度法）拟牛顿条件验证

- **DFP公式**

$$\boxed{\boldsymbol{H}_{k+1}=\boldsymbol{H}_k+\frac{\boldsymbol{p}^{(k)}\left(\boldsymbol{p}^{(k)}\right)^{\mathrm{T}}}{(\boldsymbol{p}^{(k)})^{\mathrm{T}}\boldsymbol{q}^{(k)})}-\frac{\boldsymbol{H}_k\boldsymbol{q}^{(k)}\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k}{(\boldsymbol{q}^{(k)})^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)})}}$$

$$\boldsymbol{H}_{k+1}\boldsymbol{q}^{(k)}=[\boldsymbol{H}_k+\frac{\boldsymbol{p}^{(k)}\left(\boldsymbol{p}^{(k)}\right)^{\mathrm{T}}}{(\boldsymbol{p}^{(k)})^{\mathrm{T}}\boldsymbol{q}^{(k)})}-\frac{\boldsymbol{H}_k\boldsymbol{q}^{(k)}\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k}{(\boldsymbol{q}^{(k)})^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)})}]\boldsymbol{q}^{(k)}$$

$$=\boldsymbol{H}_k\boldsymbol{q}^{(k)}+\frac{\boldsymbol{p}^{(k)}\left(\boldsymbol{p}^{(k)}\right)^{T}\boldsymbol{q}^{(k)}}{(\boldsymbol{p}^{(k)})^{T}\boldsymbol{q}^{(k)})}-\frac{\boldsymbol{H}_k\boldsymbol{q}^{(k)}\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)}}{(\boldsymbol{q}^{(k)})^{T}\boldsymbol{H}_k\boldsymbol{q}^{(k)})}$$

$$=\boldsymbol{H}_k\boldsymbol{q}^{(k)}+\boldsymbol{p}^{(k)}-\boldsymbol{H}_k\boldsymbol{q}^{(k)}=\boldsymbol{p}^{(k)}$$

拟牛顿法

■ DFP算法步骤

步骤1：给定初始点 $\boldsymbol{x}^{(1)} \in \mathbf{E}^{\mathrm{n}}$,允许误差 $\varepsilon < 0$

步骤2：置 $\mathbf{H_1} = \mathbf{I_n}$ （单位矩阵），计算 $\boldsymbol{x}^{(1)}$ 处的梯度 $\boldsymbol{g_1} = \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)})$,置k=1

步骤3：令 $\boldsymbol{d}^{(k)} = -\mathbf{H}_k \boldsymbol{g}_k$

步骤4：从 $\boldsymbol{x}^{(k)}$ 出发，沿方向 $\boldsymbol{d}^{(k)}$ 搜索，求步长$\boldsymbol{\lambda_k}$使其满足 $\boldsymbol{f}(\boldsymbol{x}^{(k)} + \boldsymbol{\lambda}_k \boldsymbol{d}^{(k)}) = \mathbf{min} \boldsymbol{f}(\boldsymbol{x}^{(k)} + \boldsymbol{\lambda} \boldsymbol{d}^{(k)})$

令 $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{\lambda}_k \boldsymbol{d}^{(k)}$

拟牛顿法

■ DFP算法步骤

步骤5：检验是否满足收敛准则，若$\left\|\nabla f\left(x^{(1)}\right)\right\| \le \varepsilon$，则停止迭代，得到点 $x^* = x^{(k+1)}$；否则，转到步骤6

步骤6：若k=n，则令$x^1 = x^{(k+1)}$，返回步骤2；否则，进行步骤7

步骤7：令$g_{k+1} = \nabla f(x^{(k+1)}), p^{(k)} = x^{(k+1)} - x^{(k)}$，$q^{(k)} = g_{k+1} - g_k$。利用公式（10.4.17）计算$H_{k+1}$,置k:=k+1, 转到步骤3

拟牛顿法

■ 例子—— 用DFP求解以下问题

$$\min 2\boldsymbol{x}_1^2 + \boldsymbol{x}_2^2 - 4\boldsymbol{x}_1 + 2$$

初始点及初始矩阵分别取 梯度

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \mathbf{H}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{g} = \begin{bmatrix} 4(x_1 - 1) \\ 2x_2 \end{bmatrix}$$

第1次迭代：

在$\boldsymbol{x}^{(1)}$处的梯度：$\boldsymbol{g}_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ $\|\boldsymbol{g}_1\| > 0$

在$\boldsymbol{x}^{(1)}$处的搜索方向： $\boldsymbol{d}^{(1)} = -\mathbf{H}_1\boldsymbol{g}_1 = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$

$$\boldsymbol{d}^{(1)} = -\boldsymbol{g}_1 = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$$

拟牛顿法

■ 例子—— 用DFP求解以下问题

从$\boldsymbol{x}^{(1)}$出发沿$\boldsymbol{d}^{(1)}$作一维搜索，求步长λ_1:

$$\min_{\lambda\geq 0} f\left(\boldsymbol{x}^{(1)}+\lambda \boldsymbol{d}^{(1)}\right),\ \text{得 } \lambda_1=5/18$$

$$\longrightarrow \boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\lambda_1 \boldsymbol{d}^{(1)}=\begin{bmatrix}2\\1\end{bmatrix}+\frac{5}{18}\begin{bmatrix}-4\\-2\end{bmatrix}=\begin{bmatrix}\frac{8}{9}\\\frac{4}{9}\end{bmatrix}$$

在$\boldsymbol{x}^{(2)}$处的梯度：

$$\boldsymbol{g}_2=\begin{bmatrix}4(\frac{8}{9}-1)\\2\cdot\frac{4}{9}\end{bmatrix}=\begin{bmatrix}-\frac{4}{9}\\\frac{8}{9}\end{bmatrix}\quad \|\boldsymbol{g}_2\|>0\longrightarrow$$

进行第2次迭代

拟牛顿法

■ 例子—— 用DFP求解以下问题

第2次迭代：

$$\boldsymbol{p}^{(1)} = \boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)} = \lambda_1 \boldsymbol{d}^{(1)} = \begin{bmatrix} -\frac{10}{9} \\ -\frac{5}{9} \end{bmatrix}$$

$$\boldsymbol{q}^{(1)} = \boldsymbol{g}_2 - \boldsymbol{g}_1 = \begin{bmatrix} -\frac{4}{9} \\ \frac{8}{9} \end{bmatrix} - \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{40}{9} \\ -\frac{10}{9} \end{bmatrix}$$

拟牛顿法

■ 例子—— 用DFP求解以下问题

计算$\boldsymbol{H}_2$：

$$\boldsymbol{H}_2 = \boldsymbol{H}_1 + \frac{\boldsymbol{p}^{(1)}\boldsymbol{p}^{(1)T}}{\boldsymbol{p}^{(1)T}\boldsymbol{q}^{(1)}} - \frac{\boldsymbol{H}_1\boldsymbol{q}^{(1)}\boldsymbol{q}^{(1)T}\boldsymbol{H}_1}{\boldsymbol{q}^{(1)T}\boldsymbol{H}_1\boldsymbol{q}^{(1)}} = \frac{1}{306}\begin{bmatrix} 86 & -38 \\ -38 & 305 \end{bmatrix}$$

在$\boldsymbol{x}^{(2)}$处的搜索方向： $\boldsymbol{d}^{(2)} = -\mathrm{H}_2\boldsymbol{g}_2 = \dfrac{12}{51}\begin{bmatrix} 1 \\ -4 \end{bmatrix}$

从$\boldsymbol{x}^{(2)}$出发沿$\boldsymbol{d}^{(2)}$作一维搜索，求步长λ_2：

$$\min_{\lambda \geq 0} f\left(\boldsymbol{x}^{(2)} + \lambda \boldsymbol{d}^{(2)}\right)，\text{得 } \lambda_2 = 17/36$$

$$\longrightarrow \quad \boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \lambda_2 \boldsymbol{d}^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \boldsymbol{g}_3 = \begin{bmatrix} 4(1-1) \\ 2 \cdot 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\|\boldsymbol{g}_3\| = 0$ 迭代终止 最优点$\boldsymbol{x}^* = \boldsymbol{x}^{(3)}$

拟牛顿法

由Broyden，Fletcher，Goldfarb和Shanno于1970年提出

■ BFGS

- 另一种形式拟牛顿条件

$$q^{(k)} = B_{k+1} p^{(k)}$$

- 递推公式

$$B_{k+1} = B_k + \frac{(q^{(k)})^{\mathrm{T}} q^{(k)}}{(q^{(k)})^{\mathrm{T}} p^{(k)}} - \frac{(B_k p^{(k)})^{\mathrm{T}} p^{(k)} B_k}{(p^{(k)})^{\mathrm{T}} B_k p^{(k)}}$$

- 令 $H_{k+1} = B_{k+1}{}^{-1}$

$$\boxed{H_{k+1} = H_k + \left(1 + \frac{(q^{(k)})^{\mathrm{T}} H_k q^{(k)}}{(p^{(k)})^{\mathrm{T}} q^{(k)}}\right) \frac{p^{(k)} (p^{(k)})^{\mathrm{T}}}{(p^{(k)})^{\mathrm{T}} q^{(k)}}}$$

- **数值计算经验表明**

比DFP公式还好，因此得到广泛应用

拟牛顿法

■ 讨论

•优点

➢ 无约束最优化方法中最有效的一类算法

➢ 迭代仅需一阶导数，不必计算海塞矩阵

$$g_{k+1} = \nabla f(x^{(k+1)})$$

➢ 当矩阵 H_k 正定，算法产生的方向为下降方向，具有二次终止性

•缺点

➢ 存储量大，不宜适用于大型优化问题。

目录

- 最速下降法
- 牛顿法
- 共轭梯度法
- 拟牛顿法
- 最小二乘法
- 信赖域法

最小二乘法

■ 最小二乘问题

设目标函数由若干个函数的平方和构成，一般写成

$$F(\boldsymbol{x}) = \sum_{i=1}^{m} f_i^2(\boldsymbol{x}), \quad \boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$$

则极小化$F(x)$这类函数的问题

$$\min_{x} F(\boldsymbol{x}) = \sum_{i=1}^{m} f_i^2(\boldsymbol{x})$$

称为最小二乘问题

- ◆ $f_i(\boldsymbol{x})$为$\boldsymbol{x}$的线性函数时，则为线性最小二乘问题
- ◆ $f_i(\boldsymbol{x})$为$\boldsymbol{x}$的非线性函数时，则为非线性最小二乘问题

最小二乘法

■ 线性最小二乘问题

假设$f_i(x) = \boldsymbol{p}_i^T x - b_i$, p_i为n维列向量， b_i为实数

令 $$A = \begin{bmatrix} \boldsymbol{p}_1^T \\ \vdots \\ \boldsymbol{p}_m^T \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

则有 $$F(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x}))\begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$$

$$\begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1^T\boldsymbol{x} - b_1 \\ \boldsymbol{p}_2^T\boldsymbol{x} - b_2 \\ \vdots \\ \boldsymbol{p}_m^T\boldsymbol{x} - b_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1^T\boldsymbol{x} \\ \boldsymbol{p}_2^T\boldsymbol{x} \\ \vdots \\ \boldsymbol{p}_m^T x \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = A\boldsymbol{x} - \boldsymbol{b}$$

最小二乘法

■ 线性最小二乘问题

所以 $F(\boldsymbol{x}) = (A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b})$

$$= (\boldsymbol{x}^T A^T - \boldsymbol{b}^T)(A\boldsymbol{x} - \boldsymbol{b})$$

$\boldsymbol{b}^T A\boldsymbol{x} = (\boldsymbol{x}^T A^T \boldsymbol{b})^T$

一个实数

$$= \boldsymbol{x}^T A^T A\boldsymbol{x} - \boldsymbol{x}^T A^T \boldsymbol{b} - \boldsymbol{b}^T A\boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{b}$$

$$= \boldsymbol{x}^T A^T A\boldsymbol{x} - 2\boldsymbol{b}^T A\boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{b}$$

令 $\nabla F(\boldsymbol{x}) = 2A^T A\boldsymbol{x} - \mathbf{2}A^T b = 0$

$F(\boldsymbol{x})$的极小点满足 $A^T A\boldsymbol{x} = A^T b$

设A列满秩， $A^T A$为n阶对称正定矩阵

凸函数$F(\boldsymbol{x})$的全局极小点$\bar{\boldsymbol{x}} = (A^T A)^{-1} A^T b$

最小二乘法

■ 线性最小二乘问题求解

给定方程组 $\begin{cases} x_1 + x_2 = 3 \\ 2x_1 - 3x_2 = 2 \\ -x_1 + 4x_2 = 4 \end{cases}$

记系数矩阵为A，右端向量为b，求函数 $f(\boldsymbol{x}) = (A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b})$ 的极小点。

$$A = \begin{bmatrix} 1 & 1 \\ 2 & -3 \\ -1 & 4 \end{bmatrix}, b = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}$$

进一步得到

$$A^TA = \begin{bmatrix} 6 & -9 \\ -9 & 26 \end{bmatrix},\ A^TA\text{非奇异},\ (A^TA)^{-1} = \frac{1}{75}\begin{bmatrix} 26 & 9 \\ 9 & 6 \end{bmatrix}$$

最小二乘法

■ 线性最小二乘问题求解

根据公式 $\bar{\boldsymbol{x}} = (A^T A)^{-1} A^T b$

$$\bar{\boldsymbol{x}} = (A^T A)^{-1} A^T b = \frac{1}{75}\begin{bmatrix} 26 & 9 \\ 9 & 6 \end{bmatrix}\begin{bmatrix} 1 & 2 & -1 \\ 1 & -3 & 4 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{13}{5} \\ \frac{7}{5} \end{bmatrix}$$

极小点$\bar{\boldsymbol{x}}$也称为最小二乘解。

$$f_{min} = f(\bar{\boldsymbol{x}}) = (A\bar{\boldsymbol{x}} - \boldsymbol{b})^T (A\bar{\boldsymbol{x}} - \boldsymbol{b}) = 3 \neq 0$$

方程组 $\begin{cases} x_1 + x_2 = 3 \\ 2x_1 - 3x_2 = 2 \\ -x_1 + 4x_2 = 4 \end{cases}$ 无解

若方程组有解，最小二乘解，也是f(x)的解

最小二乘法

■ 非线性最小二乘问题

$$\min_{x} F(\boldsymbol{x}) = \sum_{i=1}^{m} f_i^2(\boldsymbol{x})$$

$f_i(\boldsymbol{x})$是非线性函数， $F(\boldsymbol{x})$存在连续偏导数

求解思路：

- 通过解一系列线性最小二乘问题不断逼近非线性最小二乘问题的解
- 具体要用线性函数近似非线性函数

最小二乘法

■ 非线性最小二乘问题

设$\boldsymbol{x}^{(k)}$是解的第k次近似

$f_i(\boldsymbol{x})$在 $\boldsymbol{x}^{(k)}$处展开一阶**Taylor**多项式，得线性近似函数

$$\varphi_i(\boldsymbol{x}) = f_i\left(\boldsymbol{x}^{(k)}\right) + \nabla f_i\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) \approx f_i(\boldsymbol{x})$$

令 $\phi(\boldsymbol{x}) = \sum_{i=1}^{m} \varphi_i^2(\boldsymbol{x})$， 用 $\phi(\boldsymbol{x})$ 近似$F(\boldsymbol{x})$

于是用$\phi(\boldsymbol{x})$的极小点作为目标函数$F(\boldsymbol{x})$的估计

最小二乘法

■ 非线性最小二乘问题

$$\varphi_i(\boldsymbol{x}) = f_i\left(\boldsymbol{x}^{(k)}\right) + \nabla f_i\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right)$$

$$= \nabla f_i\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}}\boldsymbol{x} - \left(\nabla f_i\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}}\boldsymbol{x}^{(k)} - f_i\left(\boldsymbol{x}^{(k)}\right)\right)$$

令

$$\boldsymbol{A}_k = \begin{bmatrix} \nabla f_1\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}} \\ \vdots \\ \nabla f_m\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1\left(\boldsymbol{x}^{(k)}\right)}{\partial x_1} & \cdots & \frac{\partial f_1\left(\boldsymbol{x}^{(k)}\right)}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m\left(\boldsymbol{x}^{(k)}\right)}{\partial x_1} & \cdots & \frac{\partial f_m\left(\boldsymbol{x}^{(k)}\right)}{\partial x_n} \end{bmatrix} \quad f^{(k)} = \begin{bmatrix} f_1\left(\boldsymbol{x}^{(k)}\right) \\ \vdots \\ f_m\left(\boldsymbol{x}^{(k)}\right) \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \nabla f_1\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}}\boldsymbol{x}^{(k)} - f_1\left(\boldsymbol{x}^{(k)}\right) \\ \vdots \\ \nabla f_m\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}}\boldsymbol{x}^{(k)} - f_m\left(\boldsymbol{x}^{(k)}\right) \end{bmatrix} = A_k\boldsymbol{x}^{(k)} - f^{(k)},$$

最小二乘法

■ 非线性最小二乘问题

所以 $\phi(\boldsymbol{x}) = (A_k\boldsymbol{x} - \boldsymbol{b})^T(A_k\boldsymbol{x} - \boldsymbol{b})$ ，$\boldsymbol{b} = A_k\boldsymbol{x}^{(k)} - f^{(k)}$

$$= \left(A_k\boldsymbol{x} - A_k\boldsymbol{x}^{(k)} + f^{(k)}\right)^T\left(A_k\boldsymbol{x} - A_k\boldsymbol{x}^{(k)} + f^{(k)}\right)$$

$$= \left(A_k\left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) + f^{(k)}\right)^T\left(A_k\left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) + f^{(k)}\right)$$

令 $\nabla\phi(\boldsymbol{x}) = 2A_k^TA_k(\boldsymbol{x} - \boldsymbol{x}^{(k)}) - \mathbf{2}A_k^T(A_k\boldsymbol{x}^{(k)} - f^{(k)}) = 0$

则 $A_k^TA_k\left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) = -A_k^Tf^{(k)}$

设A_k列满秩， $A_k^TA_k$为n阶对称正定矩阵

$\phi(\boldsymbol{x})$的极小点 $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \left(A_k^TA_k\right)^{-1}A_k^Tf^{(k)}$

$\boldsymbol{x}^{(k+1)}$为$F(\boldsymbol{x})$的极小点的第$\boldsymbol{k+1}$次近似。

最小二乘法

■ 非线性最小二乘问题

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \left(A_k^T A_k\right)^{-1} A_k^T f^{(k)}$$ Causs-Newton公式

$$\boldsymbol{d}^{(k)} = -\left(A_k^T A_k\right)^{-1} A_k^T f^{(k)}$$ Causs-Newton方向

在求出$\boldsymbol{d}^{(k)}$后，为保证目标函数下降（至少不上升）不直接用$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)}$作为第k+1次近似

沿着$d^{(k)}$方向一维搜索获得λ_k

$$\min_{\lambda} F(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{d}^{(k)})$$

$$\longrightarrow \boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}$$

$\boldsymbol{x}^{(k+1)}$为$F(\boldsymbol{x})$的极小点的第$\boldsymbol{k+1}$次近似。

最小二乘法

■ 非线性最小二乘算法步骤

步骤1：给定初始点 $\boldsymbol{x}^{(1)} \in \mathbf{E}^{\mathrm{n}}$,允许误差 $\varepsilon > 0$

步骤2：计算函数值$f_i\left(\boldsymbol{x}^{(k)}\right)(\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{m})$,求

$$\boldsymbol{f}^{(k)} = \begin{bmatrix} f_1\left(\boldsymbol{x}^{(k)}\right) \\ f_2\left(\boldsymbol{x}^{(k)}\right) \\ \vdots \\ f_m\left(\boldsymbol{x}^{(k)}\right) \end{bmatrix}, \quad \boldsymbol{a}_{ij} = \frac{\partial f_i(\boldsymbol{x}^{(k)})}{\partial x_j}, \quad \boldsymbol{A}_k = \left(a_{ij}\right)_{m\times n}$$

步骤3：求causs-newton方向

$$\boldsymbol{d}^{(k)} = -\left(A_k^T A_k\right)^{-1} A_k^T f^{(k)}$$

最小二乘法

■ 非线性最小二乘算法步骤

步骤4：从 $\boldsymbol{x}^{(k)}$ 出发，沿方向 $\boldsymbol{d}^{(k)}$ 搜索，求步长$\boldsymbol{\lambda_k}$使其满足

$$\boldsymbol{F}(\boldsymbol{x}^{(k)}+\boldsymbol{\lambda_k}\boldsymbol{d}^{(k)})=\mathbf{min}\boldsymbol{F}(\boldsymbol{x}^{(k)}+\boldsymbol{\lambda}\boldsymbol{d}^{(k)})$$

$$\boldsymbol{x}^{(\boldsymbol{k}+\mathbf{1})}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}$$

步骤5：若$\left\|\boldsymbol{x}^{(\boldsymbol{k}+\mathbf{1})}-\boldsymbol{x}^{(k)}\right\|\leq\varepsilon$，则停止迭代，得到点$\overline{\boldsymbol{x}}=\boldsymbol{x}^{(k+1)}$；否则，转到步骤2

最小二乘法

■ 讨论

•线性最小二乘问题

➢ **直接用下面的公式求最小二乘解**

$$\bar{\boldsymbol{x}} = (A^T A)^{-1} A^T b$$

•非线性最小二乘问题

➢ **需要迭代，使用线性近似函数确定搜索方向**

$$\boldsymbol{d}^{(k)} = -\left(A_k^T A_k\right)^{-1} A_k^T f^{(k)}$$

➢ **需要通过多次迭代逐步逼近最小二乘解。**

$$\boldsymbol{x}^{(\boldsymbol{k}+\mathbf{1})} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}$$

目录

信赖域法

- 迭代方法一般策略是，给定点$x^{(k)}$后，确定搜索方向$d^{(k)}$，再从点$x^{(k)}$出发沿着$d^{(k)}$进行一维搜索，找到最优步长λ_k，继而$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$。

- 信赖域法则不同，它以 $x^{(k)}$ 为中心确定一个球域，称为信赖域，在此域内优化目标函数的二次逼近式，按照一定模式求出$x^{(k+1)}$。

信赖域法

■ 基本形式

◆ 模型选择为二次近似模型，采用函数二阶泰勒展开，即

$$f(x_k+\boldsymbol{p})\approx\varphi(\boldsymbol{p})=f(x_k)+\nabla f(x_k)^{\mathrm{T}}\boldsymbol{p}+\frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\nabla^2 f(x_k)\boldsymbol{p}$$

◆ 优化求解如下模型

$$\min\varphi_k(\boldsymbol{p})=f(x_k)+\nabla f(x_k)^{\mathrm{T}}\boldsymbol{p}+\frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\nabla^2 f(x_k)\boldsymbol{p}$$

$$s.t.\quad \|\boldsymbol{p}\|\le r_k$$ **信赖域**

信赖域法

■ 基本形式

$$\min \varphi_k(\boldsymbol{p}_k) = f(x_k) + \nabla f(x_k)^{\mathrm{T}} \boldsymbol{p}_k + \frac{1}{2} \boldsymbol{p}_k^{\mathrm{T}} \nabla^2 f(x_k) \boldsymbol{p}_k$$

$$s.t. \quad \|\boldsymbol{p}_k\| \le r_k$$

最优性条件

K-T条件

$$\begin{cases} \nabla^2 f(x_k)^{\mathrm{T}} \boldsymbol{p}_k + w\boldsymbol{p}_k = -\nabla f(x_k) \\ w(\|\boldsymbol{p}_k\| - r_k) = 0 \\ w \ge 0 \\ \|\boldsymbol{p}_k\| \le r_k \end{cases}$$

$\boldsymbol{r}_k$ **充分大** $\quad \boldsymbol{p}_k \approx -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$ 接近牛顿方向

$\boldsymbol{r}_k \to 0 \quad \boldsymbol{p}_k \approx -\dfrac{1}{w} \nabla f(x_k)$ 接近最速下降方向

作业

P330:

2、14(偶数)

P331:

19