

# 最优化方法 第五讲

Weiwen Wang(王伟文)

暨南大学

2025 年 4 月 18 日

# 目录

## 1 线性共轭梯度法

- 预备知识
- 二次函数最小化
- 共轭方向
- 共轭梯度法
- 线性共轭梯度法收敛性

## 2 非线性共轭梯度法

- Fletcher-Reeves 迭代
- 收敛性分析
- $\beta_{k+1}$  的变体

## 线性共轭梯度法

## 预备知识

二次函数最小化

共轭方向

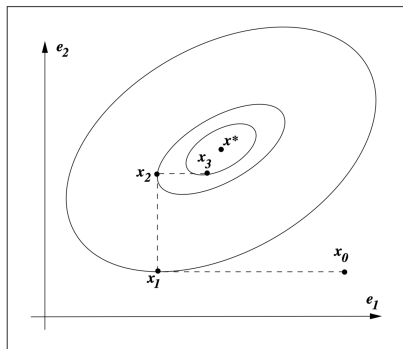
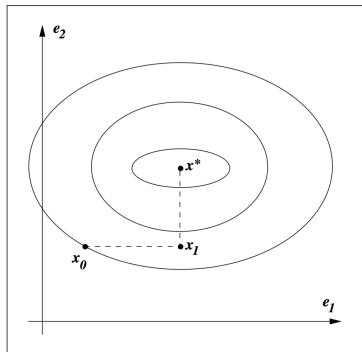
共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 线性共轭梯度法

## 预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 定义 1 (共轭向量)

称一组非零向量  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\} \subset \mathbb{R}^d$  关于对称正定矩阵  $\mathbf{A} \in \mathbb{R}^{d \times d}$  共轭 (conjugate), 若

$$\mathbf{p}_i^T \mathbf{A} \mathbf{p}_j = 0 \quad \forall i \neq j.$$

## 线性共轭梯度法

## 预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 引理 2 (共轭向量线性无关)

若一组非零向量  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\} \subset \mathbb{R}^d$  关于对称正定矩阵  $\mathbf{A} \in \mathbb{R}^{d \times d}$  共轭, 则  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\}$  线性无关.

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 引理 3 (向量投影)

将向量  $\mathbf{v} \in \mathbb{R}^d$  投影到一个非零向量  $\mathbf{u} \in \mathbb{R}^d$  定义为

$$\text{PROJ}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|^2} \mathbf{u}.$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 引理 3 (向量投影)

将向量  $\mathbf{v} \in \mathbb{R}^d$  投影到一个非零向量  $\mathbf{u} \in \mathbb{R}^d$  定义为

$$\text{PROJ}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|^2} \mathbf{u}.$$

## 证明.

向量  $\mathbf{v}$  到向量  $\mathbf{u}$  的投影可以表示为  $\mathbf{z} = \alpha \mathbf{u}$ , 且满足

$$\alpha = \arg \min_{\gamma \in \mathbb{R}} \frac{1}{2} \|\gamma \mathbf{u} - \mathbf{v}\|^2$$

解得  $\alpha = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|^2}$ .



## 线性共轭梯度法

## 预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 定理 4 (Gram-Schmidt 正交化)

设  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  为一组线性无关向量, 令  $\mathbf{u}_1 = \mathbf{v}_1$ ,  $\mathbf{e}_1 = \mathbf{u}_1 / \|\mathbf{u}_1\|$ . 对于  $j = 2, 3, \dots, m$ , 递归定义  $\mathbf{e}_j$

$$\mathbf{u}_j = \mathbf{v}_j - \sum_{i=1}^{j-1} \text{PROJ}_{\mathbf{u}_i}(\mathbf{v}_j), \quad \mathbf{e}_j = \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|}.$$

则  $\mathbf{e}_1, \dots, \mathbf{e}_m$  为一组正交向量且有

$$\text{SPAN}\{\mathbf{v}_1, \dots, \mathbf{v}_j\} = \text{SPAN}\{\mathbf{e}_1, \dots, \mathbf{e}_j\} \quad \forall j \in [m].$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

考虑优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^d} \left\{ f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x} \right\}.$$

其中  $\mathbf{A} \in \mathbb{R}^{d \times d}$  为对称正定矩阵. 记  $\mathbf{g}_k = \nabla f(\mathbf{x}_k) = \mathbf{A} \mathbf{x}_k - \mathbf{b}$ .

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 共轭方向法

- ① 给定初始点  $\mathbf{x}_0$  和一组关于矩阵  $\mathbf{A}$  共轭的向量  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{d-1}\}$ .
- ②  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$
- ③ 精确线搜索

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}} f(\mathbf{x}_k + \alpha \mathbf{p}_k) = -\frac{\mathbf{g}_k^T \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}.$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 定理 5

给定初始点  $\mathbf{x}_0$  和一组关于矩阵  $\mathbf{A}$  共轭的向量  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{d-1}\}$ , 精确线搜索的共轭方法至多需要  $d$  步迭代收敛到  $f(\mathbf{x})$  的全局最小值点  $\mathbf{x}^*$ .

## 证明.

要证明定理成立, 即证明  $\mathbf{x}_d = \mathbf{x}^*$

因为共轭向量线性无关, 故存在非零向量  $\beta = (\beta_k)_{k \in [d]}$  使得

$$\mathbf{x}^* - \mathbf{x}_0 = \sum_{k=0}^{d-1} \beta_k \mathbf{p}_k.$$

由共轭性解得

$$\beta_k = \frac{\mathbf{p}_k^T \mathbf{A}(\mathbf{x}^* - \mathbf{x}_0)}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = -\frac{\mathbf{p}_k^T \mathbf{g}_0}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} \quad (\mathbf{A} \mathbf{x}^* = \mathbf{b})$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 证明.

要证明定理成立, 即证明  $\mathbf{x}_d = \mathbf{x}^*$

因为共轭向量线性无关, 故存在非零向量  $\beta = (\beta_k)_{k \in [d]}$  使得

$$\mathbf{x}^* - \mathbf{x}_0 = \sum_{k=0}^{d-1} \beta_k \mathbf{p}_k.$$

由共轭性解得

$$\beta_k = \frac{\mathbf{p}_k^T \mathbf{A}(\mathbf{x}^* - \mathbf{x}_0)}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = -\frac{\mathbf{p}_k^T \mathbf{g}_0}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} \quad (\mathbf{A} \mathbf{x}^* = \mathbf{b})$$

由迭代公式知

$$\mathbf{x}_d - \mathbf{x}_0 = \sum_{k=0}^{d-1} \alpha_k \mathbf{p}_k.$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

$$\mathbf{x}^* - \mathbf{x}_d = \sum_{k=0}^{d-1} \frac{\mathbf{p}_k^T (\mathbf{g}_k - \mathbf{g}_0)}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} \mathbf{p}_k$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

$$\mathbf{x}^* - \mathbf{x}_d = \sum_{k=0}^{d-1} \frac{\mathbf{p}_k^T (\mathbf{g}_k - \mathbf{g}_0)}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} \mathbf{p}_k$$

又因为

$$\begin{aligned} \mathbf{p}_k^T (\mathbf{g}_k - \mathbf{g}_0) &= \mathbf{p}_k^T \mathbf{A} (\mathbf{x}_k - \mathbf{x}_0) \\ &= \mathbf{p}_k^T \mathbf{A} \left( \sum_{j=1}^{k-1} \alpha_j \mathbf{p}_j \right) = 0 \quad (\text{共轭性}) \end{aligned}$$

故

$$\mathbf{x}_d = \mathbf{x}^*.$$



# 精确线搜索的共轭方法与最速坐标下降法

$$\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}.$$

若  $\mathbf{A}$  为非对角矩阵.

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 精确线搜索的共轭方法与最速坐标下降法

$$\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}.$$

若  $\mathbf{A}$  为非对角矩阵.

记  $\hat{\mathbf{x}} = \mathbf{S}^{-1} \mathbf{x}$ , 其中  $\mathbf{S}$  为  $d$  个关于  $\mathbf{A}$  共轭的列向量组成列的矩阵. 上述优化问题可以转换为

$$\min_{\hat{\mathbf{x}} \in \mathbb{R}^d} \frac{1}{2} \hat{\mathbf{x}}^T (\mathbf{S}^T \mathbf{A} \mathbf{S}) \hat{\mathbf{x}} - \mathbf{b}^T \mathbf{S} \hat{\mathbf{x}}$$

由共轭关系,  $\mathbf{S}^T \mathbf{A} \mathbf{S}$  为共轭矩阵.

此时坐标下降法

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{S} \mathbf{e}_k = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 定理 6 (扩张性子空间最小化)

设  $\mathbf{x}_0$  为初始点,  $\{\mathbf{x}_k\}$  为精确线搜索共轭方向法产生的迭代序列, 则有

$$\mathbf{g}_k^T \mathbf{p}_i = 0 \quad \forall i \in \{0, \dots, k-1\}.$$

且  $\mathbf{x}_k$  为目标函数  $f$  在集合

$$\mathcal{A}_{k-1} = \{\mathbf{x} | \mathbf{x}_0 + \text{SPAN}\{\mathbf{p}_0, \dots, \mathbf{p}_{k-1}\}\}$$

上的最小值点.

## 证明.

## 1. 首先证明第二个结论. 优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) \quad \text{s.t. } \mathbf{x} \in \mathcal{A}_{k-1}$$

等价于

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{k-1}} h(\boldsymbol{\beta}) = f(\mathbf{x}_0 + \sum_{i=0}^{k-1} \beta_i \mathbf{p}_i)$$

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 证明.

## 1. 首先证明第二个结论. 优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) \quad \text{s.t. } \mathbf{x} \in \mathcal{A}_{k-1}$$

等价于

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{k-1}} h(\boldsymbol{\beta}) = f(\mathbf{x}_0 + \sum_{i=0}^{k-1} \beta_i \mathbf{p}_i)$$

其中  $h(\boldsymbol{\beta})$  为关于  $\boldsymbol{\beta}$  的强凸函数,  $\tilde{\boldsymbol{\beta}}$  为唯一最小值点当且仅当

$$\frac{\partial h(\tilde{\boldsymbol{\beta}})}{\partial \beta_i} = 0 \quad \forall i \in \{0, 1, \dots, k-1\}$$

由链式法则, 上式改写为

$$\mathbf{g}(\mathbf{x}_0 + \sum_{i=0}^{k-1} \tilde{\beta}_i \mathbf{p}_i)^T \mathbf{p}_i = 0 \quad \forall i \in \{0, 1, \dots, k-1\}$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

解得  $\tilde{\beta}_i = -\frac{\mathbf{g}_0^T \mathbf{p}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i}$  又因为

$$\alpha_i = -\frac{\mathbf{g}_i^T \mathbf{p}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i} = -\frac{(\mathbf{g}_0 + \sum_{j=0}^{i-1} \alpha_j \mathbf{A} \mathbf{p}_j)^T \mathbf{p}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i} = -\frac{\mathbf{g}_0^T \mathbf{p}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i} \quad \forall i \in \{1, \dots, k-1\}.$$

且  $\alpha_i = -\frac{\mathbf{g}_0^T \mathbf{p}_0}{\mathbf{p}_0^T \mathbf{A} \mathbf{p}_0}$  故  $\tilde{\beta}_i = \alpha_i, \forall i \in \{0, 1, \dots, k-1\}$ .

2. 由迭代公式知, 每一个迭代的梯度满足如下递推关系

$$\mathbf{g}_{j+1} = \mathbf{g}_j + \alpha_j \mathbf{A} \mathbf{p}_j$$

故有

$$\mathbf{g}_j = \mathbf{g}_0 + \sum_{i=0}^{j-1} \alpha_i \mathbf{A} \mathbf{p}_i \quad \forall j \in \{1, 2, \dots, k\}$$

□

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

$$\begin{aligned}
 \mathbf{g}_k^T \mathbf{p}_i &= \mathbf{p}_i^T \mathbf{g}_0 + \alpha_i \mathbf{p}_i^T \mathbf{A} \mathbf{p}_i \quad \forall i \in \{1, 2, \dots, k-1\} \\
 &= \mathbf{p}_i^T (\mathbf{g}_0 - \mathbf{g}_i) \\
 &= \mathbf{p}_i^T \left( - \sum_{l=0}^{i-1} \alpha_l \mathbf{A} \mathbf{p}_l \right) = 0
 \end{aligned}$$

类似地,

$$\mathbf{g}_k^T \mathbf{p}_0 = \mathbf{g}_0^T \mathbf{p}_0 + \alpha_0 \mathbf{p}_0^T \mathbf{A} \mathbf{p}_0 = 0.$$

综上,

$$\mathbf{g}_k^T \mathbf{p}_i = 0 \quad \forall i \in \{0, \dots, k-1\}.$$



- 共轭方向可选矩阵  $\mathbf{A}$  的特征值, 但要较高的计算成本.

- $\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1}\mathbf{p}_k$
- 选择  $\beta_{k+1}$  使得  $\mathbf{p}_{k+1}$  与  $\mathbf{p}_k$  关于  $\mathbf{A}$  共轭
  - $\beta_{k+1} = \frac{\mathbf{g}_{k+1}^T \mathbf{A} \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

- $\mathbf{p}_0 = \mathbf{g}_0 = \mathbf{A}\mathbf{x}_0 - \mathbf{b}$
- - ①  $\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}$
  - ②  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$
  - ③  $\beta_{k+1} = \frac{\mathbf{g}_{k+1}^T \mathbf{A} \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}$
  - ④  $\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1} \mathbf{p}_k$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 定理 7

若共轭梯度下降法的第  $k$  次迭代  $\mathbf{x}_k$  不是  $f(\mathbf{x})$  的最小值点, 有下列性质成立

$$\mathbf{g}_k^T \mathbf{g}_i = 0 \quad \forall i \in \{0, 1, \dots, k-1\} \quad (\circ)$$

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k\} = \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\} \quad (\Delta)$$

$$\text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\} = \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\} \quad (\square)$$

$$\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i = 0 \quad \forall i \in \{0, 1, \dots, k-1\} \quad (\star)$$

从而  $\{\mathbf{x}_k\}$  收敛到  $\mathbf{x}^*$  至多需要  $d$  次迭代.

## 证明.

1. 首先利用归纳法证明  $(\Delta)$  和  $(\square)$ .

当  $k = 0$  时,  $(\Delta)$  和  $(\square)$  均成立, 假设有

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\}$$

$$\text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\}$$

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 证明.

1. 首先利用归纳法证明  $(\Delta)$  和  $(\square)$ .当  $k = 0$  时,  $(\Delta)$  和  $(\square)$  均成立, 假设有

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\}$$

$$\text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\}$$

此时

$$\mathbf{g}_k \in \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\}$$

$$\mathbf{p}_k \in \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\}$$

故

$$\mathbf{A}\mathbf{p}_k \in \text{SPAN}\{\mathbf{A}\mathbf{g}_0, \mathbf{A}^2 \mathbf{g}_0, \dots, \mathbf{A}^{k+1} \mathbf{g}_0\}$$



续.

由递推关系

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \alpha_k \mathbf{A} \mathbf{p}_k$$

得到

$$\mathbf{g}_{k+1} = \mathbf{g}_0 + \sum_{i=0}^k \alpha_i \mathbf{A} \mathbf{p}_i$$

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

续.

由递推关系

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \alpha_k \mathbf{A} \mathbf{p}_k$$

得到

$$\mathbf{g}_{k+1} = \mathbf{g}_0 + \sum_{i=0}^k \alpha_i \mathbf{A} \mathbf{p}_i$$

从而

$$\mathbf{g}_{k+1} \in \text{SPAN}\{\mathbf{g}_0, \mathbf{A} \mathbf{g}_0, \dots, \mathbf{A}^{k+1} \mathbf{g}_0\}$$

利用递推关系

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1} \mathbf{p}_k$$

即有

$$\mathbf{p}_{k+1} \in \text{SPAN}\{\mathbf{g}_0, \mathbf{A} \mathbf{g}_0, \dots, \mathbf{A}^{k+1} \mathbf{g}_0\}$$



线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

续.

因此

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k+1}\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\}$$

$$\text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k+1}\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\}$$

接下来证明另外一个方向的包含关系, 假设有

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

续.

因此

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k+1}\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\}$$

$$\text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k+1}\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\}$$

接下来证明另外一个方向的包含关系, 假设有

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k\mathbf{g}_0\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k\}$$

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k\mathbf{g}_0\} \subset \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\}$$

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

因此

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k+1}\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\}$$

$$\text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k+1}\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\}$$

接下来证明另外一个方向的包含关系, 假设有

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k\mathbf{g}_0\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k\}$$

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k\mathbf{g}_0\} \subset \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\}$$

此时

$$\mathbf{A}^k\mathbf{g}_0 \in \text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k\}$$

$$\mathbf{A}^k\mathbf{g}_0 \in \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\}$$





## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

$$\mathbf{A}^{k+1}\mathbf{g}_0 \in \text{SPAN}\{\mathbf{A}\mathbf{g}_0, \mathbf{A}\mathbf{g}_1, \dots, \mathbf{A}\mathbf{g}_k\}$$

$$\mathbf{A}^{k+1}\mathbf{g}_0 \in \text{SPAN}\{\mathbf{A}\mathbf{p}_0, \mathbf{A}\mathbf{p}_1, \dots, \mathbf{A}\mathbf{p}_k\}$$

由递推关系

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \alpha_k \mathbf{A}\mathbf{p}_k$$

$$\mathbf{A}\mathbf{p}_k = \frac{1}{\alpha_k} (\mathbf{g}_k - \mathbf{g}_{k+1}) \quad \alpha_k \neq 0(\text{why?})$$

知

$$\mathbf{A}^{k+1}\mathbf{g}_0 \in \text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k, \mathbf{g}_{k+1}\}$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

同样利用递推关系

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1}\mathbf{p}_k$$

$$\mathbf{g}_{k+1} = \beta_{k+1}\mathbf{p}_k - \mathbf{p}_{k+1}$$

故

$$\mathbf{A}^{k+1}\mathbf{g}_0 \in \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k+1}\}$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

同样利用递推关系

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1}\mathbf{p}_k$$

$$\mathbf{g}_{k+1} = \beta_{k+1}\mathbf{p}_k - \mathbf{p}_{k+1}$$

故

$$\mathbf{A}^{k+1}\mathbf{g}_0 \in \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k+1}\}$$

因此

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\} \subset \text{SPAN}\{\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{k+1}\}$$

$$\text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^{k+1}\mathbf{g}_0\} \subset \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k+1}\}$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

2. 这里先利用性质 (★) 证明 (○)

性质 (★) 表明  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\}$  关于  $\mathbf{A}$  共轭, 因此有

$$\mathbf{g}_k^T \mathbf{p}_i = 0 \quad \forall i \in \{0, 1, \dots, k-1\} \quad (\circ)$$

故

$$\mathbf{g}_k^T \mathbf{g}_i = \mathbf{g}_k^T (\beta_i \mathbf{p}_{i-1} - \mathbf{p}_i) \quad \forall i \in \{1, 2, \dots, k-1\}$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

2. 这里先利用性质 (★) 证明 (○)

性质 (★) 表明  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k\}$  关于  $\mathbf{A}$  共轭, 因此有

$$\mathbf{g}_k^T \mathbf{p}_i = 0 \quad \forall i \in \{0, 1, \dots, k-1\} \quad (\circ)$$

故

$$\mathbf{g}_k^T \mathbf{g}_i = \mathbf{g}_k^T (\beta_i \mathbf{p}_{i-1} - \mathbf{p}_i) \quad \forall i \in \{1, 2, \dots, k-1\}$$

又因为  $\mathbf{g}_0 = \mathbf{p}_0$ , 所以  $\mathbf{g}_k^T \mathbf{g}_0 = 0$ .

综上,

$$\mathbf{g}_k^T \mathbf{g}_i = 0 \quad \forall i \in \{0, 1, \dots, k-1\} \quad (\circ)$$



续.

3. 最后使用归纳法证明性质 (★). 由  $\beta_1$  及递推关系

$$\mathbf{p}_1 = -\mathbf{g}_1 + \beta_1 \mathbf{p}_0$$

可得  $\mathbf{p}_1^T \mathbf{A} \mathbf{p}_0 = 0$ .

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

续.

3. 最后使用归纳法证明性质 (★). 由  $\beta_1$  及递推关系

$$\mathbf{p}_1 = -\mathbf{g}_1 + \beta_1 \mathbf{p}_0$$

可得  $\mathbf{p}_1^T \mathbf{A} \mathbf{p}_0 = 0$ . 设当  $k \geq 2$  时, 有  $\mathbf{p}_{k-1}^T \mathbf{A} \mathbf{p}_i = 0, \forall i \in \{0, 1, \dots, k-2\}$ . 此时  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k-2}\}$  关于矩阵  $A$  共轭. 由定理 6 知

$$\mathbf{g}_{k-1}^T \mathbf{p}_i = 0 \quad \forall i \in \{0, 1, \dots, k-2\}.$$

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

3. 最后使用归纳法证明性质 (★). 由  $\beta_1$  及递推关系

$$\mathbf{p}_1 = -\mathbf{g}_1 + \beta_1 \mathbf{p}_0$$

可得  $\mathbf{p}_1^T \mathbf{A} \mathbf{p}_0 = 0$ . 设当  $k \geq 2$  时, 有  $\mathbf{p}_{k-1}^T \mathbf{A} \mathbf{p}_i = 0, \forall i \in \{0, 1, \dots, k-2\}$ . 此时  $\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k-2}\}$  关于矩阵  $A$  共轭. 由定理 6 知

$$\mathbf{g}_{k-1}^T \mathbf{p}_i = 0 \quad \forall i \in \{0, 1, \dots, k-2\}.$$

利用  $\alpha_{k-1}$  及递推关系  $\mathbf{g}_k = \mathbf{g}_{k-1} + \alpha_{k-1} \mathbf{A} \mathbf{p}_{k-1}$  知  $\mathbf{g}_k^T \mathbf{p}_{k-1} = 0$ , 且有归纳假设可得到  $\mathbf{g}_k^T \mathbf{p}_i = 0, \forall i \in \{0, 1, 2, \dots, k-2\}$ , 故

$$\mathbf{g}_k^T \mathbf{p}_i = 0 \quad \forall i \in \{0, 1, \dots, k-1\}.$$

□



续.

$$\begin{aligned}\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i &= -\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i + \beta_{k-1} \mathbf{p}_{k-1}^T \mathbf{A} \mathbf{p}_i \quad \forall i \in \{0, 1, 2, k-2\} \\ &= -\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i\end{aligned}$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

$$\begin{aligned}\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i &= -\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i + \beta_{k-1} \mathbf{p}_{k-1}^T \mathbf{A} \mathbf{p}_i \quad \forall i \in \{0, 1, 2, k-2\} \\ &= -\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i\end{aligned}$$

由性质 (□)

$$\begin{aligned}\mathbf{A} \mathbf{p}_i &\in \text{SPAN}\{\mathbf{A} \mathbf{g}_0, \mathbf{A} \mathbf{g}_0, \dots, \mathbf{A}^{i+1} \mathbf{g}_0\} \\ &\subset \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{i+1}\} \quad \forall i \in \{0, 1, \dots, k-2\}\end{aligned}$$

故  $\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i = 0, \forall i \in \{0, 1, \dots, k-2\}$ , 即

$$\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i = 0, \forall i \in \{0, 1, \dots, k-2\}.$$

利用  $\beta_k$  及递推关系  $\mathbf{p}_k = -\mathbf{g}_k + \beta_k \mathbf{p}_{k-1}$  知  $\mathbf{p}_k^T \mathbf{A} \mathbf{p}_{k-1} = 0$ .

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

$$\begin{aligned}\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i &= -\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i + \beta_{k-1} \mathbf{p}_{k-1}^T \mathbf{A} \mathbf{p}_i \quad \forall i \in \{0, 1, 2, k-2\} \\ &= -\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i\end{aligned}$$

由性质 (□)

$$\begin{aligned}\mathbf{A} \mathbf{p}_i &\in \text{SPAN}\{\mathbf{A} \mathbf{g}_0, \mathbf{A} \mathbf{g}_0, \dots, \mathbf{A}^{i+1} \mathbf{g}_0\} \\ &\subset \text{SPAN}\{\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{i+1}\} \quad \forall i \in \{0, 1, \dots, k-2\}\end{aligned}$$

故  $\mathbf{g}_k^T \mathbf{A} \mathbf{p}_i = 0, \forall i \in \{0, 1, \dots, k-2\}$ , 即

$$\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i = 0, \forall i \in \{0, 1, \dots, k-2\}.$$

利用  $\beta_k$  及递推关系  $\mathbf{p}_k = -\mathbf{g}_k + \beta_k \mathbf{p}_{k-1}$  知  $\mathbf{p}_k^T \mathbf{A} \mathbf{p}_{k-1} = 0$ .综上,  $\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i = 0, \forall i \in \{0, 1, 2, \dots, k-1\}$ . □

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体


$$\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = -\frac{\mathbf{g}_k^T (-\mathbf{g}_k + \beta_k \mathbf{p}_{k-1})}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

$$\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = -\frac{\mathbf{g}_k^T (-\mathbf{g}_k + \beta_k \mathbf{p}_{k-1})}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}$$

$$\beta_{k+1} = \frac{\mathbf{g}_{k+1}^T \mathbf{A} \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = \frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \alpha_k \mathbf{p}_k)}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = \frac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}.$$

- $\mathbf{p}_0 = \mathbf{g}_0 = \mathbf{A}\mathbf{x}_0 - \mathbf{b}$
- - ①  $\alpha_k = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}$
  - ②  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$
  - ③  $\beta_{k+1} = \frac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}$
  - ④  $\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1} \mathbf{p}_k$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 线性共轭梯度法收敛性

## 引理 8

设  $\mathbf{x}_0$  为初始点,  $\mathbf{x}^*$  为  $f$  的全局最小值点, 线性共轭梯度法有

$$\|\mathbf{x}_{k+1} - \mathbf{x}_0\|_{\mathbf{A}}^2 \leq \min_{P_k} \max_{1 \leq i \leq n} [1 + \lambda_i P_k(\lambda_i)]^2 \|\mathbf{x}_0 - \mathbf{x}^*\|_{\mathbf{A}}^2$$

成立, 其中  $P_k$  表示  $k$  次多项式,  $\lambda_i$  为矩阵  $\mathbf{A}$  从小到大排序的特征值,  $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$ .

## 证明思路

- 首先利用  $\mathbf{x}_0 + \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\}$  表示  $\mathbf{x}_{k+1}$ .
- $\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{x}^*\|_{\mathbf{A}}^2$  与  $\min_{\mathbf{x}} f(\mathbf{x})$  等价.
- 使用矩阵  $\mathbf{A}$  的特征向量表示  $\mathbf{x}_0 - \mathbf{x}^*$

## 证明.

由扩张性子空间最小化定理和定理 7 性质 (□),

$$\mathbf{x}_{k+1} \in \mathbf{x}_0 + \text{SPAN}\{\mathbf{g}_0, \mathbf{A}\mathbf{g}_0, \dots, \mathbf{A}^k \mathbf{g}_0\} \triangleq \mathcal{G}_k$$

故存在  $\gamma^* = (\gamma_i^*)_{i \in \{0, 1, \dots, k\}}$  使得

$$\mathbf{x}_{k+1} = \mathbf{x}_0 + \left( \gamma_0^* \mathbf{I} + \sum_{i=1}^k \gamma_i^* \mathbf{A}^i \right) \mathbf{g}_0$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

即

$$\mathbf{x}_{k+1} \in \mathbf{x}_0 + \left\{ P_k^\gamma(\mathbf{A}) \triangleq \gamma_0 + \sum_{i=1}^k \gamma_i \mathbf{A}^i : \gamma = (\gamma_0, \gamma_1, \dots, \gamma_k) \in \mathbb{R}^{k+1} \right\} \mathbf{g}_0$$

另一方面

$$\frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|_{\mathbf{A}}^2 = \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \mathbf{A} (\mathbf{x} - \mathbf{x}^*) = f(\mathbf{x}) + \frac{1}{2} (\mathbf{x}^*)^T \mathbf{A} \mathbf{x}^*$$

因此

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{G}_k} f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{G}_k} \|\mathbf{x} - \mathbf{x}^*\|_{\mathbf{A}}^2 = \operatorname{argmin}_{\mathbf{x}_0 + P_k^\gamma(\mathbf{A}) \mathbf{g}_0} \|\mathbf{x}_0 + P_k^\gamma(\mathbf{A}) \mathbf{g}_0 - \mathbf{x}^*\|_{\mathbf{A}}^2$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

进而

$$\begin{aligned}
 \|\mathbf{x}_{k+1} - \mathbf{x}^*\|_A^2 &= \min_{P_k^\gamma} \|\mathbf{x}_0 + P_k^\gamma(\mathbf{A})\mathbf{g}_0 - \mathbf{x}^*\|_A^2 \\
 &= \min_{P_k^\gamma} \|\mathbf{x}_0 + P_k^\gamma(\mathbf{A})(\mathbf{A}\mathbf{x}_0 - \mathbf{b}) - \mathbf{x}^*\|_A^2 \\
 &= \min_{P_k^\gamma} \|\mathbf{x}_0 + P_k^\gamma(\mathbf{A})(\mathbf{A}\mathbf{x}_0 - \mathbf{A}\mathbf{x}^*) - \mathbf{x}^*\|_A^2 \\
 &= \min_{P_k^\gamma} \|\left(\mathbf{I} + P_k^\gamma(\mathbf{A})\mathbf{A}\right)(\mathbf{x}_0 - \mathbf{x}^*)\|_A^2
 \end{aligned}$$

设矩阵  $\mathbf{A}$  的谱分解为  $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ , 其中  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_d]$ ,  
 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d)$ , 存在  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ , 使得

$$\mathbf{x}_0 - \mathbf{x}^* = \sum_{i=1}^d \beta_i \mathbf{v}_i = \mathbf{V}\boldsymbol{\beta}.$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

因为  $P_k^\gamma(\mathbf{A}) = \mathbf{V}P_k^\gamma(\mathbf{\Lambda})\mathbf{V}^T$ , 故

$$\begin{aligned}
 \|(\mathbf{I} + P_k^\gamma(\mathbf{A})\mathbf{A})(\mathbf{x}_0 - \mathbf{x}^*)\|_{\mathbf{A}}^2 &= \|(\mathbf{V}\mathbf{V}^T + \mathbf{V}P_k^\gamma(\mathbf{\Lambda})\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)\mathbf{V}\beta\|_{\mathbf{A}}^2 \\
 &= \|\mathbf{V}(\mathbf{I} + P_k^\gamma(\mathbf{\Lambda})\mathbf{\Lambda})\beta\|_{\mathbf{A}}^2 \\
 &= \left(\mathbf{V}(\mathbf{I} + P_k^\gamma(\mathbf{\Lambda})\mathbf{\Lambda})\beta\right)^T \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \left(\mathbf{V}(\mathbf{I} + P_k^\gamma(\mathbf{\Lambda})\mathbf{\Lambda})\beta\right) \\
 &= \beta^T \left(\mathbf{I} + P_k^\gamma(\mathbf{\Lambda})\mathbf{\Lambda}\right) \mathbf{\Lambda} \left(\mathbf{I} + P_k^\gamma(\mathbf{\Lambda})\mathbf{\Lambda}\right) \beta \\
 &\leq \max_{i \in [d]} [1 + P_k^\gamma(\lambda_i)\lambda_i]^2 \beta^T \mathbf{\Lambda} \beta \\
 &= \max_{i \in [d]} [1 + P_k^\gamma(\lambda_i)\lambda_i]^2 \|\mathbf{x}_0 - \mathbf{x}^*\|_{\mathbf{A}}^2
 \end{aligned}$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

综合上述结论

$$\begin{aligned}\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_{\mathbf{A}}^2 &= \min_{P_k^\gamma} \left\| \left( \mathbf{I} + P_k^\gamma(\mathbf{A})\mathbf{A} \right) (\mathbf{x}_0 - \mathbf{x}^*) \right\|_{\mathbf{A}}^2 \\ &\leq \min_{P_k^\gamma} \max_{i \in [d]} [1 + P_k^\gamma(\lambda_i)\lambda_i]^2 \|\mathbf{x}_0 - \mathbf{x}^*\|_{\mathbf{A}}^2\end{aligned}$$



# 线性共轭梯度法收敛性

## 定理 9

若矩阵  $\mathbf{A}$  有  $n$  个不同特征值, 线性共轭梯度法至多需要  $n$  次迭代收敛到  $f$  的全局最小值点.

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 证明.

记矩阵  $\mathbf{A}$  的  $n$  个不同特征值为  $0 < \xi_1 < \xi_2 < \cdots < \xi_n$ , 对应  $d$  个特征值  $(\lambda_i)_{i \in [d]}$  的取值.

设迭代次数  $k+1 = n$ , 定义多项式

$$Q^n(\lambda) = \frac{(-1)^n}{\prod_{i=1}^n \xi_i} (\lambda - \xi_1) \cdots (\lambda - \xi_n)$$

此时  $Q^n(\lambda) - 1$  为不带常数项的  $k+1$  次多项式, 因此

$$\bar{P}_k = \frac{Q^n(\lambda) - 1}{\lambda}$$

为  $k$  次多项式.

$$\min_{P_k^\gamma} \max_{i \in [d]} [1 + P_k^\gamma(\lambda_i) \lambda_i]^2 \leq \max_{i \in [d]} [1 + \bar{P}_k(\lambda_i) \lambda_i]^2 = \max_{i \in [d]} [Q^n(\lambda_i)]^2 = 0$$

考虑更一步的目标函数

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}),$$

其中  $f$  可微

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

# Fletcher-Reeves 迭代

- $\nabla f_0 = \nabla f(\mathbf{x}_0), \mathbf{p} = -\nabla f(\mathbf{x}_0)$
- 选择步长  $\alpha_k$ , 更新  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$
- 计算梯度  $\nabla f_{k+1} = \nabla f(\mathbf{x}_{k+1})$
- 更新共轭方向

$$\beta_{k+1} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$$

$$\mathbf{p}_{k+1} = -\nabla f_{k+1} + \beta_{k+1} \mathbf{p}_k$$

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体



## 确定迭代步长 $\alpha_k$

强 Wolfe 条件

$$f(\mathbf{x}_k + \alpha \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f_k^T \mathbf{p}_k \quad (\text{Wolfe-a})$$

$$|\nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)^T \mathbf{p}_k| \leq -c_2 \nabla f_k^T \mathbf{p}_k \quad (\text{Wolfe-b})$$

上式中  $0 < c_1 < c_2 < \frac{1}{2}$  为给定参数.

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

$\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 引理 10

在 *Fletcher-Reeves* 迭代中, 若步长  $\alpha_k$  满足强 *Wolfe-a&b* 条件, 其中  $0 < c_1 < c_2 < \frac{1}{2}$ , 则关于更新方向  $\mathbf{p}_k$ , 有

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2} \quad \forall k \in \{0, 1, \dots\}.$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 证明.

此处我们使用归纳法证明.

定义  $h(c_2) = \frac{2c_2-1}{1-c_2}$ , 因为  $h(c_2)$  在  $(0, \frac{1}{2})$  单调递增, 故

$$h(0) = -1 < \frac{2c_2-1}{1-c_2} < h\left(\frac{1}{2}\right) = 0 \quad \forall c_2 \in \left(0, \frac{1}{2}\right).$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

## 证明.

此处我们使用归纳法证明.

定义  $h(c_2) = \frac{2c_2-1}{1-c_2}$ , 因为  $h(c_2)$  在  $(0, \frac{1}{2})$  单调递增, 故

$$h(0) = -1 < \frac{2c_2-1}{1-c_2} < h\left(\frac{1}{2}\right) = 0 \quad \forall c_2 \in \left(0, \frac{1}{2}\right).$$

当  $k=0$  时,

$$\frac{\nabla f_0^T \mathbf{p}_0}{\nabla f_0^T \nabla f_0} = -1,$$

故

$$-\frac{1}{1-c_2} < \frac{\nabla f_0^T \mathbf{p}_0}{\nabla f_0^T \nabla f_0} < \frac{2c_2-1}{1-c_2}$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

假设当  $k \geq 1$ , 有

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}.$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

假设当  $k \geq 1$ , 有

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}.$$

利用  $\beta_{k+1}$  及递推关系  $\mathbf{p}_{k+1} = -\nabla f_{k+1} + \beta_{k+1}\mathbf{p}_k$ ,

$$\begin{aligned} \frac{\nabla f_{k+1}^T \nabla \mathbf{p}_{k+1}}{\nabla f_{k+1}^T \nabla f_{k+1}} &= \frac{\nabla f_{k+1}^T \nabla (-\nabla f_{k+1} + \beta_{k+1}\mathbf{p}_k)}{\nabla f_{k+1}^T \nabla f_{k+1}} \\ &= -1 + \frac{\nabla f_{k+1}^T \mathbf{p}_k}{\nabla f_k^T \nabla f_k} \end{aligned}$$



## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

由 Wolfe-b,

$$c_2 \nabla f_k^T \mathbf{p}_k \leq \nabla f_{k+1}^T \mathbf{p}_k \leq -c_2 \nabla f_k^T \nabla \mathbf{p}_k$$

故

$$-1 + \frac{c_2 \nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2} \leq -1 + \frac{\nabla f_{k+1}^T \mathbf{p}_k}{\nabla f_k^T \nabla f_k} \leq -1 - \frac{c_2 \nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2}$$

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

由 Wolfe-b,

$$c_2 \nabla f_k^T \mathbf{p}_k \leq \nabla f_{k+1}^T \mathbf{p}_k \leq -c_2 \nabla f_k^T \nabla \mathbf{p}_k$$

故

$$-1 + \frac{c_2 \nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2} \leq -1 + \frac{\nabla f_{k+1}^T \mathbf{p}_k}{\nabla f_k^T \nabla f_k} \leq -1 - \frac{c_2 \nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2}$$

结合归纳假设

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T \mathbf{p}_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}.$$

得到

$$-1 - \frac{c_2}{1-c_2} \leq -1 + \frac{\nabla f_{k+1}^T \mathbf{p}_k}{\nabla f_k^T \nabla f_k} \leq -1 + \frac{c_2}{1-c_2}$$





## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

续.

因此

$$-\frac{1}{1-c_2} \leq \frac{\nabla f_{k+1}^T \mathbf{p}_{k+1}}{\|\nabla f_{k+1}\|^2} \leq \frac{2c_2 - 1}{1 - c_2}.$$



- 引理 9 说明在强 Wolfe 条件下, Fletcher-Reeves 迭代的更新方向  $\mathbf{p}_k$  为下降方向, 即  $\nabla f_k^T \mathbf{p}_k < 0$ .

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体

# 收敛性分析

## 假设 11

- 存在  $\mathbf{x}_0 \in \mathbb{R}^d$ , 水平集  $\mathcal{L} = \{\mathbf{x} | f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$  有界
- 存在  $\hat{\gamma}$  使得

$$\|\nabla f(\mathbf{x})\| \leq \hat{\gamma} \quad \forall \mathbf{x} \in \mathcal{L}.$$

- $f$  为  $L$ -smooth

# 收敛性分析

## 定理 12

若  $f$  满足假设 10, 在 *Fletcher-Reeves* 迭代步长  $\alpha_k$  满足 *Wolfe-a&b*, 则

$$\liminf_{k \rightarrow \infty} \|\nabla f_k\| = 0.$$

线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

非线性共轭梯度法

*Fletcher-Reeves* 迭代

收敛性分析

$\beta_{k+1}$  的变体

## 线性共轭梯度法

预备知识

二次函数最小化

共轭方向

共轭梯度法

线性共轭梯度法收敛性

## 非线性共轭梯度法

Fletcher-Reeves 迭代

收敛性分析

 $\beta_{k+1}$  的变体 $\beta_{k+1}$  的变体

- Polak-Ribière:

$$\beta_{k+1}^{PR} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\nabla f_k^T \nabla f_k}$$

PR 迭代即使在强 Wolfe 条件下也无法保证  $\mathbf{p}_k$  为下降方向.

- PR+:

$$\beta_{k+1}^{PR+} = \max\{\beta_{k+1}^{PR}, 0\}$$

- Hestenes-Stiefel:

$$\beta_{k+1}^{HS} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{(\nabla f_{k+1} - \nabla f_k)^T \mathbf{p}_k}$$

- Dai-Yuan:

$$\beta_{k+1}^{DY} = \frac{\nabla f_k^T \nabla f_k}{(\nabla f_{k+1} - \nabla f_k)^T \mathbf{p}_k}$$