

共轭梯度法

邓建松

2022 年 10 月 24 日

松弛迭代法的局限性

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- SOR迭代法中如能取得最佳松弛因子，算法的效率会得到数量级上的提高

松弛迭代法的局限性

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- SOR迭代法中如能取得最佳松弛因子，算法的效率会得到数量级上的提高
- 而最佳松弛因子只在系数矩阵具有较好性质时才有可能找到

松弛迭代法的局限性

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- SOR迭代法中如能取得最佳松弛因子，算法的效率会得到数量级上的提高
- 而最佳松弛因子只在系数矩阵具有较好性质时才有可能找到
- 而且上节在计算最佳松弛因子时，还用到了对应的Jacobi迭代矩阵的谱半径

共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 这是一种不需要确定任何参数的求解对称正定线性方程组的方法

共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 这是一种不需要确定任何参数的求解对称正定线性方程组的方法
- 它是上世纪50年代初期由M.R. Hestenes和E. Stiefel首先提出的

共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 这是一种不需要确定任何参数的求解对称正定线性方程组的方法
- 它是上世纪50年代初期由M.R. Hestenes和E. Stiefel首先提出的
- 自后得到了长足的发展，成为求解大型稀疏线性方程组最受欢迎的一类方法

共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 这是一种不需要确定任何参数的求解对称正定线性方程组的方法
- 它是上世纪50年代初期由M.R. Hestenes和E. Stiefel首先提出的
- 自后得到了长足的发展，成为求解大型稀疏线性方程组最受欢迎的一类方法
- 它也是求解大型非线性优化问题的主要方法之一

线性方程组与对应的二次泛函

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 设 A 为对称正定矩阵

线性方程组与对应的二次泛函

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 设 A 为对称正定矩阵
- 考虑线性方程组 $Ax = b$ 的求解

线性方程组与对应的二次泛函

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 设 A 为对称正定矩阵
- 考虑线性方程组 $Ax = b$ 的求解
- 为此我们定义二次函数

$$\varphi(x) = x^T Ax - 2b^T x$$

定理

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

定理

设 A 对称正定，求方程组 $Ax = b$ 的解等价于求二次函数 $\varphi(x)$ 的极小值点

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 直接计算可得

$$\frac{\partial \varphi}{\partial x_i} = 2(a_{i1}x_1 + \cdots + a_{in}x_n) - 2b_i$$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 直接计算可得

$$\frac{\partial \varphi}{\partial x_i} = 2(a_{i1}x_1 + \cdots + a_{in}x_n) - 2b_i$$

- 所以

$$\nabla \varphi(x) = 2(Ax - b)$$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 直接计算可得

$$\frac{\partial \varphi}{\partial x_i} = 2(a_{i1}x_1 + \cdots + a_{in}x_n) - 2b_i$$

- 所以

$$\nabla \varphi(x) = 2(Ax - b)$$

- 若 $\varphi(x)$ 在某点 x_* 达到极小, 则必有 $\nabla \varphi(x_*) = 0$, 从而 $Ax_* = b$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 直接计算可得

$$\frac{\partial \varphi}{\partial x_i} = 2(a_{i1}x_1 + \cdots + a_{in}x_n) - 2b_i$$

- 所以

$$\nabla \varphi(x) = 2(Ax - b)$$

- 若 $\varphi(x)$ 在某点 x_* 达到极小, 则必有 $\nabla \varphi(x_*) = 0$, 从而 $Ax_* = b$
- 若 x_* 是 $Ax = b$ 的解, 而 $\varphi(x)$ 的Hessian阵是 A , 对称正定, 从而 x_* 是 $\varphi(x)$ 的极小值点

求解方法：盲人下山

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 为了求解线性方程组，我们可以计算二次函数 $\varphi(x)$ 的极小值

求解方法：盲人下山

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 为了求解线性方程组，我们可以计算二次函数 $\varphi(x)$ 的极小值
- 为了求二次函数的极小值，我们可以模拟盲人下山：先任意给定一个初始点 x_0 ，确定一个下山的方向 p_0 ，沿着经过点 x_0 而方向为 p_0 的直线 $x = x_0 + \alpha p_0$ 上找一点 x_1 使 $\varphi(x)$ 达到极小

重复：迈更多步

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 第一步走到 x_1

重复：迈更多步

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 第一步走到 x_1
- 然后在 x_1 点，再找一个下山的方向 p_1 , 沿直线 $x = x_1 + \alpha p_1$ 再跨出一步

重复：迈更多步

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 第一步走到 x_1
- 然后在 x_1 点，再找一个下山的方向 p_1 ，沿直线 $x = x_1 + \alpha p_1$ 再跨出一步
-

重复：迈更多步

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 第一步走到 x_1
- 然后在 x_1 点，再找一个下山的方向 p_1 ，沿直线 $x = x_1 + \alpha p_1$ 再跨出一步
-
- 这样就得到一串参数 α_i 和方向 p_i .

重复：迈更多步

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 第一步走到 x_1
- 然后在 x_1 点，再找一个下山的方向 p_1 ，沿直线 $x = x_1 + \alpha p_1$ 再跨出一步
-
- 这样就得到一串参数 α_i 和方向 p_i .
- p_i 称为搜索方向， α_k 为步长

重复：迈更多步

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 第一步走到 x_1
- 然后在 x_1 点，再找一个下山的方向 p_1 ，沿直线 $x = x_1 + \alpha p_1$ 再跨出一步
-
- 这样就得到一串参数 α_i 和方向 p_i .
- p_i 称为搜索方向， α_k 为步长
- 不同的确定搜索方向和步长的方法，就得出不同的算法

步长的确定

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 设 x_k 已确定，下山方向 p_k 也确定

步长的确定

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 设 x_k 已确定，下山方向 p_k 也确定
- 任务：在直线 $x = x_k + \alpha p_k$ 上确定 α_k 使得 $\varphi(x)$ 在 $x_{k+1} = x_k + \alpha_k p_k$ 处达到极小

步长的确定

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 设 x_k 已确定，下山方向 p_k 也确定
- 任务：在直线 $x = x_k + \alpha p_k$ 上确定 α_k 使得 $\varphi(x)$ 在 $x_{k+1} = x_k + \alpha_k p_k$ 处达到极小
- 令 $f(\alpha) = \varphi(x_k + \alpha p_k)$ ，则

$$\begin{aligned} f(\alpha) &= (x_k + \alpha p_k)^T A(x_k + \alpha p_k) - 2b^T(x_k + \alpha p_k) \\ &= \alpha^2 p_k^T A p_k - 2\alpha r_k^T p_k + \varphi(x_k) \end{aligned}$$

其中 $r_k = b - Ax_k$ 为 $\varphi(x)$ 在 $x = x_k$ 的负梯度方向

求导确定 α_k

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 计算 $f(\alpha)$ 的导数:

$$f'(\alpha) = 2\alpha p_k^T A p_k - 2r_k^T p_k$$

求导确定 α_k

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 计算 $f(\alpha)$ 的导数:

$$f'(\alpha) = 2\alpha p_k^T A p_k - 2r_k^T p_k$$

- 令 $f'(\alpha) = 0$ 即得 $\alpha_k = \frac{r_k^T p_k}{p_k^T A p_k}$, 由此算出 x_{k+1}

求导确定 α_k

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 计算 $f(\alpha)$ 的导数:

$$f'(\alpha) = 2\alpha p_k^T A p_k - 2r_k^T p_k$$

- 令 $f'(\alpha) = 0$ 即得 $\alpha_k = \frac{r_k^T p_k}{p_k^T A p_k}$, 由此算出 x_{k+1}

- 验证:

$$\begin{aligned}\varphi(x_{k+1}) - \varphi(x_k) &= \alpha_k^2 p_k^T A p_k - 2\alpha_k r_k^T p_k \\ &= -\frac{(r_k^T p_k)^2}{p_k^T A p_k}\end{aligned}$$

因此只要 $r_k^T p_k \neq 0$, 我们就有 $\varphi(x_{k+1}) < \varphi(x_k)$

新到达点原搜索方向与新梯度方向垂直

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在第 k 步确定了搜索方向 p_k 后, 按照前述公式确定步长 α_k , 那么到达了新点 $x_{k+1} = x_k + \alpha_k p_k$

新到达点原搜索方向与新梯度方向垂直

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在第 k 步确定了搜索方向 p_k 后, 按照前述公式确定步长 α_k , 那么到达了新点 $x_{k+1} = x_k + \alpha_k p_k$
- 显然 p_k 与等值线 $\varphi(x) = \varphi(x_{k+1})$ 相切于点 $x = x_{k+1}$

新到达点原搜索方向与新梯度方向垂直

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在第 k 步确定了搜索方向 p_k 后, 按照前述公式确定步长 α_k , 那么到达了新点 $x_{k+1} = x_k + \alpha_k p_k$
- 显然 p_k 与等值线 $\varphi(x) = \varphi(x_{k+1})$ 相切于点 $x = x_{k+1}$
- $r_{k+1} = b - Ax_{k+1}$ 是上述等值线在 $x = x_{k+1}$ 处的法向量

新到达点原搜索方向与新梯度方向垂直

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在第 k 步确定了搜索方向 p_k 后, 按照前述公式确定步长 α_k , 那么到达了新点 $x_{k+1} = x_k + \alpha_k p_k$
- 显然 p_k 与等值线 $\varphi(x) = \varphi(x_{k+1})$ 相切于点 $x = x_{k+1}$
- $r_{k+1} = b - Ax_{k+1}$ 是上述等值线在 $x = x_{k+1}$ 处的法向量
- 所以 $r_{k+1}^T p_k = 0$

新到达点原搜索方向与新梯度方向垂直

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在第 k 步确定了搜索方向 p_k 后, 按照前述公式确定步长 α_k , 那么到达了新点 $x_{k+1} = x_k + \alpha_k p_k$
- 显然 p_k 与等值线 $\varphi(x) = \varphi(x_{k+1})$ 相切于点 $x = x_{k+1}$
- $r_{k+1} = b - Ax_{k+1}$ 是上述等值线在 $x = x_{k+1}$ 处的法向量
- 所以 $r_{k+1}^T p_k = 0$
- 后面我们用代数方法证明更一般性的结论

下山方向的确定：最速下降法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- $\varphi(x)$ 增加最快的方向是梯度方向

下山方向的确定：最速下降法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- $\varphi(x)$ 增加最快的方向是梯度方向
- 因此负梯度方向应该是 $\varphi(x)$ 减小最快的方向

下山方向的确定：最速下降法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- $\varphi(x)$ 增加最快的方向是梯度方向
- 因此负梯度方向应该是 $\varphi(x)$ 减小最快的方向
- 所以我们取 p_k 为负梯度方向

$$r_k = b - Ax_k$$

收敛定理的准备：一个引理

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

引理

设 A 的特征值为 $0 < \lambda_1 \leq \dots \leq \lambda_n$, $P(t)$ 是一个关于 t 的多项式, 则对 $\forall x \in \mathbb{R}^n$

$$\|P(A)x\|_A \leq \max_{1 \leq i \leq n} |P(\lambda_i)| \|x\|_A$$

其中 $\|x\|_A = \sqrt{x^T A x}$

引理的证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 取由 A 对应于 $\lambda_1, \dots, \lambda_n$ 的特征向量 y_1, \dots, y_n , 其构成 \mathbb{R}^n 的一组标准正交基

引理的证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 取由 A 对应于 $\lambda_1, \dots, \lambda_n$ 的特征向量 y_1, \dots, y_n , 其构成 \mathbb{R}^n 的一组标准正交基
- 对任意 $x \in \mathbb{R}^n$, $x = \sum_{i=1}^n \beta_i y_i$, 从而有

$$\begin{aligned} x^T P(A) A P(A) x &= \left(\sum_{i=1}^n \beta_i P(\lambda_i) y_i \right)^T A \left(\sum_{i=1}^n \beta_i P(\lambda_i) y_i \right) \\ &= \sum_{i=1}^n \lambda_i \beta_i^2 P^2(\lambda_i) \leq \max_{1 \leq i \leq n} P^2(\lambda_i) \sum_{i=1}^n \lambda_i \beta_i^2 \\ &= \max_{1 \leq i \leq n} P^2(\lambda_i) x^T A x \end{aligned}$$

收敛定理

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

定理

设 A 的特征值为 $0 < \lambda_1 \leq \dots \leq \lambda_n$, 则由最速下降法产生的序列 $\{x_k\}$ 满足

$$\|x_k - x_*\|_A \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^k \|x_0 - x_*\|_A$$

其中 $x_* = A^{-1}b$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 展开 $(x - x_*)^T A(x - x_*)$ ，并利用 $Ax_* = b$ 可得

$$\varphi(x) + x_*^T Ax_* = (x - x_*)^T A(x - x_*)$$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 展开 $(x - x_*)^T A(x - x_*)$ ，并利用 $Ax_* = b$ 可得

$$\varphi(x) + x_*^T Ax_* = (x - x_*)^T A(x - x_*)$$

- 根据 x_k 的构造方法，我们有

$$\varphi(x_k) \leq \varphi(x_{k-1} + \alpha r_{k-1}), \quad \forall \alpha \in \mathbb{R}$$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 展开 $(x - x_*)^T A(x - x_*)$ ，并利用 $Ax_* = b$ 可得

$$\varphi(x) + x_*^T Ax_* = (x - x_*)^T A(x - x_*)$$

- 根据 x_k 的构造方法，我们有

$$\varphi(x_k) \leq \varphi(x_{k-1} + \alpha r_{k-1}), \quad \forall \alpha \in \mathbb{R}$$

- 所以

$$\begin{aligned} & (x_k - x_*)^T A(x_k - x_*) \\ & \leq (x_{k-1} + \alpha r_{k-1} - x_*)^T A(x_{k-1} + \alpha r_{k-1} - x_*) \end{aligned}$$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 根据构造方法,
$$r_{k-1} = b - Ax_{k-1} = A(x_* - x_{k-1})$$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 根据构造方法,

$$r_{k-1} = b - Ax_{k-1} = A(x_* - x_{k-1})$$

- 所以我们有

$$\begin{aligned} & (x_k - x_*)^T A(x_k - x_*) \\ & \leq (x_{k-1} + \alpha r_{k-1} - x_*)^T A(x_{k-1} + \alpha r_{k-1} - x_*) \\ & = ((I - \alpha A)(x_{k-1} - x_*))^T A((I - \alpha A)(x_{k-1} - x_*)) \end{aligned}$$

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 取 $P_\alpha(t) = 1 - \alpha t$, 则由引理对 $\forall \alpha \in \mathbb{R}$ 有

$$\begin{aligned}\|x_k - x_*\|_A &\leq \|P_\alpha(A)(x_{k-1} - x_*)\|_A \\ &\leq \max_{1 \leq i \leq n} |P_\alpha(\lambda_i)| \|x_{k-1} - x_*\|_A\end{aligned}$$

- 取 $P_\alpha(t) = 1 - \alpha t$, 则由引理对 $\forall \alpha \in \mathbb{R}$ 有

$$\begin{aligned} \|x_k - x_*\|_A &\leq \|P_\alpha(A)(x_{k-1} - x_*)\|_A \\ &\leq \max_{1 \leq i \leq n} |P_\alpha(\lambda_i)| \|x_{k-1} - x_*\|_A \end{aligned}$$

- 根据Chebyshev多项式的性质,

$$\min_{\alpha} \max_{\lambda_1 \leq t \leq \lambda_n} |1 - \alpha t|$$

应在 $1 - \alpha\lambda_1$ 与 $1 - \alpha\lambda_n$ 互为相反数时达到, 此时 $\alpha = \frac{2}{\lambda_1 + \lambda_n}$, 极值为 $\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}$

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述定理表明，从任一初始向量 x_0 出发，由最速下降法产生的点列总是收敛到方程组的解

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述定理表明，从任一初始向量 x_0 出发，由最速下降法产生的点列总是收敛到方程组的解
- 收敛速度由 $(\lambda_n - \lambda_1)/(\lambda_n + \lambda_1)$ 的大小决定的

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述定理表明，从任一初始向量 x_0 出发，由最速下降法产生的点列总是收敛到方程组的解
- 收敛速度由 $(\lambda_n - \lambda_1)/(\lambda_n + \lambda_1)$ 的大小决定的
- 最速下降法简单易实现，而且可以充分利用 A 的稀疏性，但在 $\lambda_1 \ll \lambda_n$ 时速度变得非常之慢

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述定理表明，从任一初始向量 x_0 出发，由最速下降法产生的点列总是收敛到方程组的解
- 收敛速度由 $(\lambda_n - \lambda_1)/(\lambda_n + \lambda_1)$ 的大小决定的
- 最速下降法简单易实现，而且可以充分利用 A 的稀疏性，但在 $\lambda_1 \ll \lambda_n$ 时速度变得非常之慢
- 在求解线性方程组时很少用它，但它的想法很重要，并且在非线性优化求解中有大量的应用和拓展

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述定理表明，从任一初始向量 x_0 出发，由最速下降法产生的点列总是收敛到方程组的解
- 收敛速度由 $(\lambda_n - \lambda_1)/(\lambda_n + \lambda_1)$ 的大小决定的
- 最速下降法简单易实现，而且可以充分利用 A 的稀疏性，但在 $\lambda_1 \ll \lambda_n$ 时速度变得非常之慢
- 在求解线性方程组时很少用它，但它的想法很重要，并且在非线性优化求解中有大量的应用和拓展

共轭梯度法的动机

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 从局部来看，负梯度方向确实是最佳的下山方向

共轭梯度法的动机

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 从局部来看，负梯度方向确实是最佳的下山方向
- 但从整体看看，它并非最佳：迭代得到的各点连线具有明显的锯齿形状

共轭梯度法的动机

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 从局部来看，负梯度方向确实是最佳的下山方向
- 但从整体看看，它并非最佳：迭代得到的各点连线具有明显的锯齿形状
- 我们要寻找更好的下山方向，而且在方向寻找上付出的代价不要太大

共轭梯度法的计算过程

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 给定初始点 x_0 , 第一步仍然选负梯度方向为下山方向, 即 $p_0 = r_0$, 于是有

$$\alpha_0 = \frac{r_0^T r_0}{r_0^T A r_0}, x_1 = x_0 + \alpha_0 p_0, r_1 = b - A x_1$$

共轭梯度法的计算过程

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 给定初始点 x_0 , 第一步仍然选负梯度方向为下山方向, 即 $p_0 = r_0$, 于是有

$$\alpha_0 = \frac{r_0^T r_0}{r_0^T A r_0}, x_1 = x_0 + \alpha_0 p_0, r_1 = b - A x_1$$

- 在第 $k+1$ ($k \geq 1$)步, 下山方向不再简单地取 r_k , 而是在过点 x_k 由向量 r_k, p_{k-1} 所张成的二维平面 π_2 内找出使函数 φ 下降最快的方向作为 p_k

共轭梯度法的计算过程

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 给定初始点 x_0 , 第一步仍然选负梯度方向为下山方向, 即 $p_0 = r_0$, 于是有

$$\alpha_0 = \frac{r_0^T r_0}{r_0^T A r_0}, x_1 = x_0 + \alpha_0 p_0, r_1 = b - A x_1$$

- 在第 $k+1$ ($k \geq 1$)步, 下山方向不再简单地取 r_k , 而是在过点 x_k 由向量 r_k, p_{k-1} 所张成的二维平面 π_2 内找出使函数 φ 下降最快的方向作为 p_k
- 注意: $r_k \perp p_{k-1}$

新下山方向的计算

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 考虑 φ 在 π_2 上的限制:

$$\begin{aligned}\psi(\xi, \eta) &= \varphi(x_k + \xi r_k + \eta p_{k-1}) \\ &= (x_k + \xi r_k + \eta p_{k-1})^T A (x_k + \xi r_k + \eta p_{k-1}) \\ &\quad - 2b^T (x_k + \xi r_k + \eta p_{k-1})\end{aligned}$$

新下山方向的计算

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 考虑 φ 在 π_2 上的限制:

$$\begin{aligned}\psi(\xi, \eta) &= \varphi(x_k + \xi r_k + \eta p_{k-1}) \\ &= (x_k + \xi r_k + \eta p_{k-1})^T A (x_k + \xi r_k + \eta p_{k-1}) \\ &\quad - 2b^T (x_k + \xi r_k + \eta p_{k-1})\end{aligned}$$

- 分别对 ξ, η 求偏导, 得到局部下降最快的方向

新下山方向的计算

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 考虑 φ 在 π_2 上的限制:

$$\begin{aligned}\psi(\xi, \eta) &= \varphi(x_k + \xi r_k + \eta p_{k-1}) \\ &= (x_k + \xi r_k + \eta p_{k-1})^T A(x_k + \xi r_k + \eta p_{k-1}) \\ &\quad - 2b^T(x_k + \xi r_k + \eta p_{k-1})\end{aligned}$$

- 分别对 ξ, η 求偏导, 得到局部下降最快的方向
 - 实际上直接求出的是在 π_2 中达到最小值的点

新下山方向的计算

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

● 求偏导：

$$\begin{aligned}\frac{\partial \psi}{\partial \xi} &= 2r_k^T A(x_k + \xi r_k + \eta p_{k-1}) - 2b^T r_k \\ &= 2(\xi r_k^T A r_k + \eta r_k^T A p_{k-1} - r_k^T r_k) \\ \frac{\partial \psi}{\partial \eta} &= 2p_{k-1}^T A(x_k + \xi r_k + \eta p_{k-1}) - 2b^T p_{k-1} \\ &= 2(\xi r_k^T A p_{k-1} + \eta p_{k-1}^T A p_{k-1})\end{aligned}$$

这里利用了 $r_k^T p_{k-1} = 0$

- 由此得唯一极值点 $\tilde{x} = x_k + \xi_0 r_k + \eta_0 p_{k-1}$, 其中 ξ_0 和 η_0 满足

$$\begin{cases} \xi_0 r_k^T A r_k + \eta_0 r_k^T A p_{k-1} = r_k^T r_k \\ \xi_0 r_k^T A p_{k-1} + \eta_0 p_{k-1}^T A p_{k-1} = 0 \end{cases}$$

- 由此得唯一极值点 $\tilde{x} = x_k + \xi_0 r_k + \eta_0 p_{k-1}$, 其中 ξ_0 和 η_0 满足

$$\begin{cases} \xi_0 r_k^T A r_k + \eta_0 r_k^T A p_{k-1} = r_k^T r_k \\ \xi_0 r_k^T A p_{k-1} + \eta_0 p_{k-1}^T A p_{k-1} = 0 \end{cases}$$

- 由上式可知若 $r_k \neq 0$, 则必有 $\xi_0 \neq 0$ (为什么?) 因此可取新的下山方向为

$$p_k = \frac{1}{\xi_0}(\tilde{x} - x_k) = r_k + \frac{\eta_0}{\xi_0} p_{k-1}$$

- 由此得唯一极值点 $\tilde{x} = x_k + \xi_0 r_k + \eta_0 p_{k-1}$, 其中 ξ_0 和 η_0 满足

$$\begin{cases} \xi_0 r_k^T A r_k + \eta_0 r_k^T A p_{k-1} = r_k^T r_k \\ \xi_0 r_k^T A p_{k-1} + \eta_0 p_{k-1}^T A p_{k-1} = 0 \end{cases}$$

- 由上式可知若 $r_k \neq 0$, 则必有 $\xi_0 \neq 0$ (为什么?) 因此可取新的下山方向为

$$p_k = \frac{1}{\xi_0}(\tilde{x} - x_k) = r_k + \frac{\eta_0}{\xi_0} p_{k-1}$$

- η_0 是否可以等于0呢?

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 令 $\beta_{k-1} = \eta_0/\xi_0$, 则由 ξ_0 和 η_0 满足的第二个方程可知

$$\beta_{k-1} = -\frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}$$

- 令 $\beta_{k-1} = \eta_0/\xi_0$, 则由 ξ_0 和 η_0 满足的第二个方程可知

$$\beta_{k-1} = -\frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}$$

- 如此确定的 p_k 满足

$$p_k^T A p_{k-1} = \left(r_k - \frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}} p_{k-1} \right)^T A p_{k-1} = 0$$

即 p_k 与 p_{k-1} 是关于 A 相互共轭的

公式初步梳理

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- p_k 确定以后, 可以采用前面的方法确定 α_k

公式初步梳理

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- p_k 确定以后, 可以采用前面的方法确定 α_k
- 总结公式为

$$\alpha_k = \frac{r_k^T p_k}{p_k^T A p_k}, \quad x_{k+1} = x_k + \alpha_k p_k$$

$$r_{k+1} = b - A x_{k+1}$$

$$\beta_k = -\frac{r_{k+1}^T A p_k}{p_k^T A p_k}, \quad p_{k+1} = r_{k+1} + \beta_k p_k$$

r_{k+1} 的简化

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 根据 r_{k+1} 的定义,

$$\begin{aligned}r_{k+1} &= b - Ax_{k+1} = b - A(x_k + \alpha_k p_k) \\ &= r_k - \alpha_k A p_k\end{aligned}$$

r_{k+1} 的简化

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 根据 r_{k+1} 的定义,

$$\begin{aligned}r_{k+1} &= b - Ax_{k+1} = b - A(x_k + \alpha_k p_k) \\&= r_k - \alpha_k A p_k\end{aligned}$$

- $A p_k$ 在计算 α_k 时已求出, 所以计算 r_{k+1} 时就可以用上述递推公式得到

r_{k+1} 的简化

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 根据 r_{k+1} 的定义,

$$\begin{aligned}r_{k+1} &= b - Ax_{k+1} = b - A(x_k + \alpha_k p_k) \\&= r_k - \alpha_k A p_k\end{aligned}$$

- $A p_k$ 在计算 α_k 时已求出, 所以计算 r_{k+1} 时就可以用上述递推公式得到
- 由上式可得

$$A p_k = \frac{1}{\alpha_k} (r_k - r_{k+1})$$

α_k 和 β_k 的简化

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 注意等式（证明后面给出）

$$r_k^T r_{k+1} = r_k^T p_{k-1} = r_{k+1}^T p_k = 0, k = 1, 2, \dots$$

α_k 和 β_k 的简化

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 注意等式（证明后面给出）

$$r_k^T r_{k+1} = r_k^T p_{k-1} = r_{k+1}^T p_k = 0, k = 1, 2, \dots$$

- 从而我们有

$$\begin{aligned} r_{k+1}^T A p_k &= \frac{1}{\alpha_k} r_{k+1}^T (r_k - r_{k+1}) = -\frac{1}{\alpha_k} r_{k+1}^T r_{k+1} \\ p_k^T A p_k &= \frac{1}{\alpha_k} p_k^T (r_k - r_{k+1}) = \frac{1}{\alpha_k} p_k^T r_k \\ &= \frac{1}{\alpha_k} r_k^T (r_k + \beta_{k-1} p_{k-1}) = \frac{1}{\alpha_k} r_k^T r_k \end{aligned}$$

α_k 和 β_k 的简化

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

● 回忆:

$$\alpha_k = \frac{r_k^T p_k}{p_k^T A p_k} \quad \beta_k = -\frac{r_{k+1}^T A p_k}{p_k^T A p_k}$$

α_k 和 β_k 的简化

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 回忆:

$$\alpha_k = \frac{r_k^T p_k}{p_k^T A p_k} \quad \beta_k = -\frac{r_{k+1}^T A p_k}{p_k^T A p_k}$$

- 对 α_k 的分式进行简化; 并且把前页两式相除, 可对 β_k 进行简化:

$$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k} \quad \beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$$

基本性质

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

由共轭梯度法得到的向量组 $\{r_i\}$ 和 $\{p_i\}$ 具有下面的性质：

$$\textcircled{1} \quad p_i^T r_j = 0, \quad 0 \leq i < j \leq k$$

基本性质

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

由共轭梯度法得到的向量组 $\{r_i\}$ 和 $\{p_i\}$ 具有下面的性质：

$$\textcircled{1} \quad p_i^T r_j = 0, \quad 0 \leq i < j \leq k$$

$$\textcircled{2} \quad r_i^T r_j = 0, \quad i \neq j, \quad 0 \leq i, j \leq k$$

基本性质

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

由共轭梯度法得到的向量组 $\{r_i\}$ 和 $\{p_i\}$ 具有下面的性质：

$$\textcircled{1} \quad p_i^T r_j = 0, \quad 0 \leq i < j \leq k$$

$$\textcircled{2} \quad r_i^T r_j = 0, \quad i \neq j, \quad 0 \leq i, j \leq k$$

$$\textcircled{3} \quad p_i^T A p_j = 0, \quad i \neq j, \quad 0 \leq i, j \leq k$$

基本性质

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

由共轭梯度法得到的向量组 $\{r_i\}$ 和 $\{p_i\}$ 具有下面的性质:

- ① $p_i^T r_j = 0, 0 \leq i < j \leq k$
- ② $r_i^T r_j = 0, i \neq j, 0 \leq i, j \leq k$
- ③ $p_i^T A p_j = 0, i \neq j, 0 \leq i, j \leq k$
- ④ 定义 $\mathcal{K}(A, r_0, k+1) = \text{span}\{r_0, A r_0, \dots, A^k r_0\}$, 称为Krylov子空间, 则 $\text{span}\{r_0, \dots, r_k\} = \text{span}\{p_0, \dots, p_k\} = \mathcal{K}(A, r_0, k+1)$

$k = 1$ 时性质的证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本
性质

共轭梯度法

基本性质

实用共轭梯度法及其
收敛性

实用共轭梯度法

收敛性分析

- 对 k 进行归纳证明

$k = 1$ 时性质的证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 对 k 进行归纳证明
- 当 $k = 1$ 时, 因为

$$p_0 = r_0, r_1 = r_0 - \alpha_0 A p_0, p_1 = r_1 + \beta_0 p_0,$$

$$\begin{aligned} p_0^T r_1 &= r_0^T r_1 = r_0^T (r_0 - \alpha_0 A r_0) \\ &= r_0^T r_0 - \alpha_0 r_0^T A r_0 = 0, \end{aligned}$$

$$\begin{aligned} p_1^T A p_0 &= (r_1 + \beta_0 r_0)^T A r_0 \\ &= r_1^T A r_0 - \frac{r_1^T A r_0}{r_0^T A r_0} r_0^T A r_0 = 0 \end{aligned}$$

所以性质成立

性质(1)

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

假设性质在 k 时成立，我们证明在 $k + 1$ 时也成立

① 由于 $r_{k+1} = r_k - \alpha_k A p_k$ 以及归纳假设，我们有

$$p_i^T r_{k+1} = p_i^T r_k - \alpha_k p_i^T A p_k = 0, 0 \leq i \leq k - 1$$

又由于

$$p_k^T r_{k+1} = p_k^T r_k - \frac{p_k^T r_k}{p_k^T A p_k} p_k^T A p_k = 0$$

所以性质(1)在 $k + 1$ 时成立

性质(2)

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

② 由归纳假设

$$\text{span}\{r_0, \dots, r_k\} = \text{span}\{p_0, \dots, p_k\}$$

则由性质(1)可知 r_{k+1} 与上述空间正交，从而性质(2)在 $k+1$ 时成立

性质(3)

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

③ 根据归纳假设, 当 $i = 0, 1, \dots, k-1$ 时

$$\begin{aligned} p_i^T A p_{k+1} &= p_i^T A(r_{k+1} + \beta_k p_k) = r_{k+1}^T A p_i \\ &= \frac{1}{\alpha_i} r_{k+1}^T (r_i - r_{i+1}) = 0 \\ p_{k+1}^T A p_k &= (r_{k+1} + \beta_k p_k)^T A p_k \\ &= r_{k+1}^T A p_k - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k^T A p_k = 0 \end{aligned}$$

所以性质(3)成立

性质(4)

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

④ 由归纳假设可知

$$r_k, p_k \in \mathcal{K}(A, r_0, k+1) = \text{span}\{r_0, Ar_0, \dots, A^k r_0\}$$

于是

$$r_{k+1} = r_k - \alpha_k A p_k \in \mathcal{K}(A, r_0, k+2),$$

$$p_{k+1} = r_k + \beta_k A p_k \in \mathcal{K}(A, r_0, k+2),$$

而根据性质(2),(3), r_0, \dots, r_{k+1} 和 p_0, \dots, p_{k+1} 都是线性无关的, 所以性质(4)成立

Krylov子空间

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 前述性质表明，向量组 r_0, \dots, r_k 和 p_0, \dots, p_k 分别是 Krylov 子空间 $\mathcal{K}(A, r_0, k+1)$ 的正交基和共轭正交基

Krylov子空间

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 前述性质表明，向量组 r_0, \dots, r_k 和 p_0, \dots, p_k 分别是 Krylov 子空间 $\mathcal{K}(A, r_0, k+1)$ 的正交基和共轭正交基
- 所以采用共轭梯度法至多 n 步就得到方程组的解 x_*

Krylov子空间

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 前述性质表明，向量组 r_0, \dots, r_k 和 p_0, \dots, p_k 分别是 Krylov 子空间 $\mathcal{K}(A, r_0, k+1)$ 的正交基和共轭正交基
- 所以采用共轭梯度法至多 n 步就得到方程组的解 x_*
- 理论上讲，共轭梯度法是直接法

精度估计

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

定理

用共轭梯度法计算得到的近似解 x_k 满足

$$\varphi(x_k) = \min\{\varphi(x) : x \in x_0 + \mathcal{K}(A, r_0, k)\}$$

或者等价地表示为

$$\|x_k - x_*\|_A = \min\{\|x - x_*\|_A : x \in x_0 + \mathcal{K}(A, r_0, k)\}$$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 由 $\varphi(x) + x_*^T A x_* = (x - x_*)^T A (x - x_*)$ 可知要证的两式是等价的。下面只证第二式成立

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 由 $\varphi(x) + x_*^T A x_* = (x - x_*)^T A (x - x_*)$ 可知要证的两式是等价的。下面只证第二式成立
- 假设共轭梯度法计算到 ℓ 步出现 $r_\ell = 0$, 那么有

$$\begin{aligned}x_* &= x_\ell = x_{\ell-1} + \alpha_{\ell-1} p_{\ell-1} \\&= x_{\ell-2} + \alpha_{\ell-2} p_{\ell-2} + \alpha_{\ell-1} p_{\ell-1} \\&= \dots\dots\dots \\&= x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \dots + \alpha_{\ell-1} p_{\ell-1}\end{aligned}$$

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本
性质

共轭梯度法

基本性质

实用共轭梯度法及其
收敛性

实用共轭梯度法

收敛性分析

- 而对计算过程中任一步 $k < \ell$, 我们有

$$x_k = x_0 + \sum_{j=0}^{k-1} \alpha_j p_j \in x_0 + \mathcal{K}(A, r_0, k)$$

- 而对计算过程中任一步 $k < \ell$, 我们有

$$x_k = x_0 + \sum_{j=0}^{k-1} \alpha_j p_j \in x_0 + \mathcal{K}(A, r_0, k)$$

- 设 $x \in x_0 + \mathcal{K}(A, r_0, k)$ 为任一向量, 则 x 有表示

$$x = x_0 + \sum_{j=0}^{k-1} \gamma_j p_j$$

- 而对计算过程中任一步 $k < \ell$, 我们有

$$x_k = x_0 + \sum_{j=0}^{k-1} \alpha_j p_j \in x_0 + \mathcal{K}(A, r_0, k)$$

- 设 $x \in x_0 + \mathcal{K}(A, r_0, k)$ 为任一向量, 则 x 有表示

$$x = x_0 + \sum_{j=0}^{k-1} \gamma_j p_j$$

- 于是

$$x_* - x = \sum_{j=0}^{k-1} (\alpha_j - \gamma_j) p_j + \sum_{j=k}^{\ell-1} \alpha_j p_j$$

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 而 $\mathbf{x}_* - \mathbf{x}_k = \sum_{j=k}^{\ell} \alpha_j \mathbf{p}_j$, 根据共轭梯度法的性质(3)可得

$$\begin{aligned}\|\mathbf{x}_* - \mathbf{x}\|_A^2 &= \left\| \sum_{j=0}^{k-1} (\alpha_j - \gamma_j) \mathbf{p}_j \right\|_A^2 + \left\| \sum_{j=k}^{\ell-1} \alpha_j \mathbf{p}_j \right\|_A^2 \\ &\geq \left\| \sum_{j=k}^{\ell-1} \alpha_j \mathbf{p}_j \right\|_A^2 = \|\mathbf{x}_* - \mathbf{x}_k\|_A^2\end{aligned}$$

实用共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 共轭梯度法在理论上确保至多 n 步得到方程组的精确解

实用共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 共轭梯度法在理论上确保至多 n 步得到方程组的精确解
- 在实际使用时由于误差的存在, 使得 r_k 之间的正交性很快损失, 所以有限步终止性不再成立

实用共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 共轭梯度法在理论上确保至多 n 步得到方程组的精确解
- 在实际使用时由于误差的存在, 使得 r_k 之间的正交性很快损失, 所以有限步终止性不再成立
- 而且在实际应用中, 由于 n 一般很大, 迭代 n 次迭代所耗费的计算时间令人无法接受

实用共轭梯度法

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 共轭梯度法在理论上确保至多 n 步得到方程组的精确解
- 在实际使用时由于误差的存在, 使得 r_k 之间的正交性很快损失, 所以有限步终止性不再成立
- 而且在实际应用中, 由于 n 一般很大, 迭代 n 次迭代所耗费的计算时间令人无法接受
- 所以通常仍把共轭梯度法作为一种迭代法使用, 当 $\|r_k\|$ 足够小或者达到指定迭代次数时终止

共轭梯度法的优点

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在算法中，系数矩阵 A 仅仅用来由已知向量 p 产生向量 Ap ，因此可以充分利用 A 的稀疏性，而且对某些提供矩阵 A 困难，而可以方便由 p 产生向量 Ap 的应用问题，这种方法十分有用

共轭梯度法的优点

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在算法中，系数矩阵 A 仅仅用来由已知向量 p 产生向量 Ap ，因此可以充分利用 A 的稀疏性，而且对某些提供矩阵 A 困难，而可以方便由 p 产生向量 Ap 的应用问题，这种方法十分有用
- 不需要预先估计任何参数就可以计算

共轭梯度法的优点

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 在算法中，系数矩阵 A 仅仅用来由已知向量 p 产生向量 Ap ，因此可以充分利用 A 的稀疏性，而且对某些提供矩阵 A 困难，而可以方便由 p 产生向量 Ap 的应用问题，这种方法十分有用
- 不需要预先估计任何参数就可以计算
- 每次迭代的主要计算就是向量之间的运算，因此便于并行化

作为迭代法的收敛性估计

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

若系数矩阵与单位矩阵的差是一个秩为 r 的矩阵, 而且 r 又很小的话, 那么共轭梯度法收敛得很快

定理

如果 $A = I + B$, $\text{rank } B = r$, 那么共轭梯度法至多迭代 $r + 1$ 步即可得到方程组 $Ax = b$ 的精确解

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 注意到 $\text{rank } B = r$ 蕴涵着

$$\text{span}\{r_0, Ar_0, \dots, A^k r_0\} = \text{span}\{r_0, Br_0, \dots, B^k r_0\}$$

的维数不会超过 $r + 1$, 因此定理成立。

误差估计

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

定理

用共轭梯度法求得的 r_k 有如下的误差估计：

$$\|x_k - x_*\|_A \leq 2 \left(\frac{\sqrt{\kappa_2} - 1}{\sqrt{\kappa_2} + 1} \right)^k \|x_0 - x_*\|_A$$

其中 $\kappa_2 = \kappa_2(A) = \|A\|_2 \|A^{-1}\|_2$

定理证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 由共轭梯度法的性质可知, 对任意的 $x \in x_0 + \mathcal{K}(A, r_0, k)$ 有

$$\begin{aligned} x_* - x &= x_* - x_0 + \sum_{j=0}^{k-1} a_{k,j+1} A^j r_0 \\ &= \left(I + \sum_{j=1}^k a_{kj} A^j \right) A^{-1} r_0 = P_k(A) A^{-1} r_0 \end{aligned}$$

$$\text{其中 } P_k(\lambda) = 1 + \sum_{j=1}^k a_{kj} \lambda^j$$

- 令 \mathcal{P}_k 为所有满足 $P_k(0) = 1$, 且次数不超过 k 的实系数多项式全体, 则 (第一个不等号应用了“最速下降法”一节中已证的引理)

$$\begin{aligned}\|x_* - x_k\|_A &= \min\{\|x - x_*\|_A : x \in x_0 + \mathcal{K}(A, r_0, k)\} \\ &= \min_{P_k \in \mathcal{P}_k} \|P_k(A)A^{-1}r_0\|_A \\ &\leq \min_{P_k \in \mathcal{P}_k} \max_{1 \leq i \leq n} |P_k(\lambda_i)| \|A^{-1}r_0\|_A \\ &\leq \min_{P_k \in \mathcal{P}_k} \max_{a \leq \lambda \leq b} |P_k(\lambda)| \|x_* - x_0\|_A\end{aligned}$$

其中 $0 < a = \lambda_0 \leq \dots \leq \lambda_n = b$ 是 A 的特征值

- 根据Chebyshev多项式的性质，最优化问题 $\min_{P_k \in \mathcal{P}_k} \max_{a \leq \lambda \leq b} |P_k(\lambda)|$ 有唯一解

$$\tilde{P}_k(\lambda) = \frac{T_k\left(\frac{b+a-2\lambda}{b-a}\right)}{T_k\left(\frac{b+a}{b-a}\right)}$$

其中 $T_k(x)$ 是 k 次Chebyshev多项式

暂停：Chebyshev多项式

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- Chebyshev多项式: $T_n(x) = \cos(n \arccos x)$, 它是定义在 $[-1, 1]$ 上, 在所有同首项系数的 n 次多项式中, 它在 $[-1, 1]$ 上的绝对最大值最小

暂停：Chebyshev多项式

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- Chebyshev多项式: $T_n(x) = \cos(n \arccos x)$, 它是定义在 $[-1, 1]$ 上, 在所有同首项系数的 n 次多项式中, 它在 $[-1, 1]$ 上的绝对最大值最小
- 在 $[-1, 1]$ 外用多项式形式直接延拓

暂停：Chebyshev多项式

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- Chebyshev多项式: $T_n(x) = \cos(n \arccos x)$, 它是定义在 $[-1, 1]$ 上, 在所有同首项系数的 n 次多项式中, 它在 $[-1, 1]$ 上的绝对最大值最小
- 在 $[-1, 1]$ 外用多项式形式直接延拓
- 递推公式: $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$,
 $T_0(x) = 1, T_1(x) = x$

暂停：Chebyshev多项式

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- Chebyshev多项式: $T_n(x) = \cos(n \arccos x)$, 它是定义在 $[-1, 1]$ 上, 在所有同首项系数的 n 次多项式中, 它在 $[-1, 1]$ 上的绝对最大值最小

- 在 $[-1, 1]$ 外用多项式形式直接延拓

- 递推公式: $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$,
 $T_0(x) = 1, T_1(x) = x$

- 基于上述公式, 可以证明当 $|x| \geq 1$ 时有

$$T_n(x) = \frac{1}{2} \left(\left(x - \sqrt{x^2 - 1} \right)^n + \left(x + \sqrt{x^2 - 1} \right)^n \right)$$

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 当 $\gamma > 1$, $x = (\gamma + 1)/(\gamma - 1)$ 时

$$x \pm \sqrt{x^2 - 1} = \frac{(\sqrt{\gamma} \pm 1)^2}{\gamma - 1},$$

- 当 $\gamma > 1$, $x = (\gamma + 1)/(\gamma - 1)$ 时

$$x \pm \sqrt{x^2 - 1} = \frac{(\sqrt{\gamma} \pm 1)^2}{\gamma - 1},$$

- 从而有

$$\begin{aligned} T_n \left(\frac{\gamma + 1}{\gamma - 1} \right) &= \frac{(\sqrt{\gamma} + 1)^{2n} + (\sqrt{\gamma} - 1)^{2n}}{2(\gamma - 1)^n} \\ &\geq \frac{(\sqrt{\gamma} + 1)^{2n}}{2(\gamma - 1)^n} \end{aligned}$$

重回证明

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 根据Chebyshev多项式的性质，我们有

$$\begin{aligned}\max_{a \leq x \leq b} |\tilde{P}_k(\lambda)| &= \frac{1}{T_k\left(\frac{b+a}{b-a}\right)} \\ &\leq \frac{2(b-a)^k}{(\sqrt{b} + \sqrt{a})^{2k}} \\ &= 2 \left(\frac{\sqrt{\kappa_2} - 1}{\sqrt{\kappa_2} + 1} \right)^k\end{aligned}$$

这就完成了证明($\kappa_2 = b/a$)

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本
性质

共轭梯度法

基本性质

实用共轭梯度法及其
收敛性

实用共轭梯度法

收敛性分析

- 上述估计是十分粗糙的

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述估计是十分粗糙的
- 实际收敛速度往往比这个估计快得多

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述估计是十分粗糙的
- 实际收敛速度往往比这个估计快得多
- 但这个结果告诉我们，只要系数矩阵是十分良态的（即 $\kappa_2 \approx 1$ ），那么共轭梯度法就会收敛得很快

注解

共轭梯度法

邓建松

基本框架

步长的确定

最速下降法

共轭梯度法及其基本性质

共轭梯度法

基本性质

实用共轭梯度法及其收敛性

实用共轭梯度法

收敛性分析

- 上述估计是十分粗糙的
- 实际收敛速度往往比这个估计快得多
- 但这个结果告诉我们，只要系数矩阵是十分良态的（即 $\kappa_2 \approx 1$ ），那么共轭梯度法就会收敛得很快
- 对比于最速下降法收敛估计中的因子，我们有

$$\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \geq \frac{\sqrt{\lambda_n} - \sqrt{\lambda_1}}{\sqrt{\lambda_n} + \sqrt{\lambda_1}}$$