

求解大型稀疏对称矩阵极端特征值的 子空间加速的截断牛顿法

钱小燕, 刘 浩

(南京工业大学 理学院, 江苏 南京 210009)

摘要:利用扩展子空间的方法,对求解大型稀疏对称矩阵极端特征值的截断牛顿法进行改进,提出了子空间加速的截断牛顿法。理论分析和数值结果均表明,新方法对计算对称矩阵的极端特征值是有效的。

关键词:对称矩阵;特征值;特征向量;截断牛顿法;子空间加速方法

中图分类号: O221.2, O241.6 **文献标识码:** A **文章编号:** 1671-5322(2010)03-0024-04

对于实对称矩阵 A 而言, Rayleigh 商函数的最小值和最大值分别是矩阵 A 的最小特征值和最大特征值,因而求解大型对称矩阵极端特征值的问题,就转化成求解非线性 Rayleigh 商函数的优化问题。目前, Newton 类方法已成为求解对称矩阵极端特征值的最常用优化方法之一。

Absil 等人提出了用 truncated - CG 法^[1,2]求解对称矩阵的极端特征值问题。该方法与 Jacobi - Davidson 法^[3,4]、Tracemin 法^[5]有着密切的联系。它与 DACG 法^[6]不同的是, truncated - CG 是在信赖域内用线性共轭梯度法来近似最小化 Rayleigh 商的两次模型,而 DACG 方是使用非线性共轭梯度法求解 Rayleigh 商函数的最小值。

不带位移的反迭代法一定收敛于最小的特征对,也就是该方法整体收敛,但是局部线性收敛。而 Rayleigh 商迭代,它是局部两次收敛的,甚至对于对称矩阵它是局部 3 次收敛的,但是不一定收敛于最小的特征对。基于上述两种方法的互补特性,本文提出了要求校正向量与当前近似迭代向量正交,保证 Rayleigh 商函数的 Hessian 矩阵在近似特征向量接近最小值点时是正定的,然后在此基础上,将外迭代的一维线搜索推广为多维子空间搜索,从而给出了子空间加速截断 Newton 法。

1 求解特征值问题的截断 Newton 法

设 A 为 n 阶实对称矩阵,对于任一非零向量 x ,称

$$f(x) = \frac{x^T A x}{x^T x} \quad (1)$$

为对应于向量 x 的 Rayleigh 商。设矩阵 A 的 n 个特征值为 $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, 相应的特征向量为 $v_1, v_2, \dots, v_n$ 。若进一步 $\lambda_1 < \lambda_2$ 和 $\lambda_{n-1} < \lambda_n$, 则对于任一既不与 v_1 也不与 v_n 共线的向量有

$$\frac{v_1^T A v_1}{v_1^T v_1} < \frac{x^T A x}{x^T x} < \frac{v_n^T A v_n}{v_n^T v_n} \quad (2)$$

从而求解矩阵 A 的最小特征值问题,就转化为求解最优化问题

$$\min_{x \neq 0} f(x) \quad (3)$$

设 p_k 为当前迭代向量 x_k 的校正向量,考虑函数在当前迭代点的校正

$$f(x_k + p_k) = \frac{(x_k + p_k)^T A (x_k + p_k)}{(x_k + p_k)^T (x_k + p_k)} \quad (4)$$

由于 Rayleigh 商函数 $f(x)$ 的梯度为

$$\nabla f(x) = \frac{2}{x^T x} [A x - f(x) x] \quad (5)$$

Rayleigh 商函数 $f(x)$ 的 Hessian 矩阵为

$$\nabla^2 f(x) = \frac{2}{x^T x} [A - f(x)I] - \frac{4}{(x^T x)^2} [Axx^T + xx^T A - 2f(x)xx^T] \quad (6)$$

假设当前迭代向量 x_k 满足 $\|x_k\|_2 = 1$, 由最优化问题(式3)的一阶必要条件知, 校正向量 p_k 应满足 Newton 方程

$$[A - \theta_k I - 2(x_k r_k^T + r_k x_k^T)] p_k = -r_k \quad (7)$$

其中 $\theta_k = f(x_k)$, $r_k = Ax_k - \theta_k x_k$ 。

于是求解特征值问题的截断 Newton 法的算法(求解矩阵 A 的最小特征值):

算法1 截断 Newton 法

(1) 选取初始单位向量;

(2) 对 $k=0, 1, \dots$ 执行

① 计算 $\nabla f(x_k) = 2[Ax_k - f(x_k)x_k]$, 检验 $\nabla f(x_k)$ 是否满足收敛性条件, 若满足则停止, 否则继续。

② 求解 Newton 方程(式7)

③ 计算 α_k 为 $\left\{1, \frac{1}{2}, \frac{1}{2^2}, \dots\right\}$ 中满足下式的最大者:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \alpha_k c_1 [\nabla f(x_k)]^T p_k \quad (8)$$

④ 计算 $x_{k+1} = \frac{x_k + \alpha_k p_k}{\|x_k + \alpha_k p_k\|_2}$

当迭代向量 x_k 接近 Rayleigh 函数的极小值点, 即矩阵 A 的最小特征值所对应的特征向量时, Newton 方程(式7)接近奇异, 坏条件的, 收敛速度只是线性的。

2 求解特征值问题的子空间加速的截断 Newton 法

对于对称矩阵而言, Rayleigh 商迭代的收敛速度可达到3次, 因此用 Rayleigh 商迭代来加速收敛, 但它不一定收敛于最小的特征对, 为此将 Rayleigh 商迭代方法和最速下降方法结合起来, 具体的过程如下:

设对称矩阵 A 的初始近似特征对 (θ_0, x_0) , 其中 $\theta_0 = \frac{x_0^T A x_0}{x_0^T x_0}$, $\|x_0\|_2 = 1$ 。

一方面, 由于 Rayleigh 商函数 $f(x)$ 在 x_0 处的梯度是 $r_0 = (A - \theta_0 I)x_0$, 最速下降方法是选取新的近似特征向量 x_1 , 使得 $f(x_1)$ 是子空间 $\text{span}\{x_0, r_0\}$ 上的最小值。很明显, 这可以看作是在子空间 $K_1 = \text{span}\{x_0, r_0\}$ 上运用 Rayleigh - Ritz 投影方法。另一方面, 由 Rayleigh 商迭代得到的近似特

征向量是 $x_1 = (A - \theta_0 I)^{-1} x_0$, 如果 Rayleigh 商迭代是通过用迭代方法求解方程组 $(A - \theta_0 I)x = x_0$ 而得到的近似解, 则 x_1 实际上来自于由 $A - \theta_0 I$ 产生的 Krylov 子空间方法。

综合考虑以上两个方面, Gloub^[7]等提出了用 Krylov 子空间投影方法的形式来求解极端特征值问题。其基本思想是将 Rayleigh - Ritz 投影方法应用到带位移的矩阵 $A - \theta_0 I$ 产生的特定的 Krylov 子空间 $K_m = \text{span}\{x_0, (A - \theta_0 I)x_0, \dots, (A - \theta_0 I)^m x_0\}$ (其中 m 是某个固定正整数) 上去, 求出近似特征对。本文将上述基本思想运用到算法1上, 得到下面的子空间加速的截断 Newton 法, 具体算法如下:

算法2 子空间加速的截断 Newton 法

(1) 选取初始单位向量 x_0 和自然数 $m, m \geq 1$, 计算 $\theta_0 = x_0^T A x_0$ 。

(2) 对 $k=0, 1, \dots$ 执行

① 计算 $\theta_k = x_k^T A x_k$ ($\|x_k\|_2 = 1$), $g(x_k) = 2(Ax_k - \theta_k x_k)$, 检验 $g(x_k)$ 是否满足收敛性条件, 若满足则停止, 否则继续。

② 求解(式7)。

③ 计算 α_k 为 $\left\{1, \frac{1}{2}, \frac{1}{2^2}, \dots\right\}$ 中满足下式的最大者:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \alpha_k c_1 [\nabla f(x_k)]^T p_k;$$

并且记 $\tilde{x}_{k+1} = \frac{x_k + \alpha_k p_k}{\|x_k + \alpha_k p_k\|_2}$, $\tilde{\theta}_{k+1} = \tilde{x}_{k+1}^T A \tilde{x}_{k+1}$

④ 形成子空间 $K_m = \text{span}\{\tilde{x}_{k+1}, (A - \tilde{\theta}_{k+1} I)\tilde{x}_{k+1}, \dots, (A - \tilde{\theta}_{k+1} I)^m \tilde{x}_{k+1}\}$ 的规范正交基矩阵 $Z_m = (z_0, z_1, \dots, z_m)$

⑤ 计算 $A_m = Z_m^T A Z_m$ 的最小特征对 (θ_{k+1}, v_1) , $x_{k+1} = \frac{Z_m v_1}{\|Z_m v_1\|_2}$ 。

关于子空间加速的截断 Newton 法的收敛性, 有下列结果。

定理1 设 λ_1 是对称矩阵 A 的最小特征值, (θ_k, x_k) 是由算法2 在第 k 步得到的近似特征对, 则有

$$\lambda_1 \leq \theta_{k+1} \leq \theta_k \quad \text{与} \quad x_k^T (A - \theta_k I) x_k = 0 \quad (9)$$

证明: 因算法1 中的第②步的方向 p_k 是下降方向, 从而第④步中的 $\tilde{\theta}_{k+1} = \tilde{x}_{k+1}^T A \tilde{x}_{k+1}$ 满足 $\tilde{\theta}_{k+1} \leq \theta_k$, 再由第⑤步知 $\theta_{k+1} \leq \tilde{\theta}_{k+1}$, 所以 $\lambda_1 \leq \theta_{k+1} \leq \theta_k$

成立。而等式 $x_k^T(A - \theta_k I)x_k = 0$ 是显然成立的。

定理2 假设对称正定矩阵 A 的 n 个特征值为 $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, 相应的规范正交特征向量为 $v_1, v_2, \dots, v_n$, (θ_k, x_k) 是由算法2在第 k 步得到的近似特征对, 则 θ_k 收敛于矩阵 A 的一个特征值 λ , 而且 $\|(A - \lambda I)x_k\| \rightarrow 0$ 。

证明: 由定理1知, $\{\theta_k\}$ 是单调下降且有下界的序列, 则序列 $\{\theta_k\}$ 收敛, 记 $\lim \theta_k = \lambda$ 。因为 $\{x_k\}$ 有界, 所以存在收敛的子序列 $\{x_{n_k}\}$, 记

$$\lim x_{n_k} = x$$

由式(9)可得:

$$x^T(A - \lambda I)x = x^T r = 0 \quad (10)$$

其中 $r = (A - \lambda I)x$

现假设 $r \neq 0$, 记 $\hat{r} = \frac{(A - \lambda I)x}{\|(A - \lambda I)x\|_2}$, $\hat{x} = \frac{x}{\|x\|_2}$ 。考虑矩阵 A 在 $\text{span}\{\hat{x}, \hat{r}\}$ 上得投影, 定义

$$\hat{A} = [\hat{x}, \hat{r}]^T A [\hat{x}, \hat{r}], \quad I_2 = [\hat{x}, \hat{r}]^T [\hat{x}, \hat{r}]$$

由(10)知 $\{x, r\}$ 是正交的, 所以

$$\hat{A} - \lambda I_2 = \begin{pmatrix} 0 & \hat{r}^T \hat{r} \\ \hat{r}^T \hat{r} & \hat{r}^T (A - \lambda I_2) \hat{r} \end{pmatrix}$$

是不定的。若设 $\hat{A}$ 的最小特征值为 $\hat{\lambda}$, 则有

$$\hat{\lambda} < \lambda \quad (11)$$

另一方面, 由于第 k 步得到的近似特征对为 (θ_k, x_k) , 再根据算法2第②和③步的构造, 以及设第

③步得到 $\tilde{x}_{k+1} = \frac{x_k + \alpha_k p_k}{\|x_k + \alpha_k p_k\|_2}$, $\tilde{\theta}_{k+1} = \tilde{x}_{k+1}^T A \tilde{x}_{k+1}$,

我们有 $\tilde{\theta}_{k+1} < \theta_k$ 。

定义 $\tilde{r}_{k+1} = (A - \tilde{\theta}_{k+1} I) \tilde{x}_{k+1}$, 并且记 $\hat{r}_k = \frac{(A - \tilde{\theta}_{k+1} I) \tilde{x}_{k+1}}{\|(A - \tilde{\theta}_{k+1} I) \tilde{x}_{k+1}\|_2}$, $\hat{x}_k = \frac{\tilde{x}_{k+1}}{\|\tilde{x}_{k+1}\|_2}$, 所以 $\hat{A}_k =$

$[\hat{x}_k, \hat{r}_k]^T A [\hat{x}_k, \hat{r}_k]$, $I_2 = [\hat{x}_k, \hat{r}_k]^T [\hat{x}_k, \hat{r}_k]$ 。设 $\hat{A}_k$ 的

最小特征值为 $\hat{\lambda}_{k+1}$ 。显然当 $n_k \rightarrow \infty$ 时, $\hat{A}_k \rightarrow \hat{A}$, 从而 $\hat{\lambda}_{n_k+1} \rightarrow \hat{\lambda}$ 。又因为 θ_{k+1} 为矩阵 A 在子空间 $K_m = \text{span}\{\tilde{x}_{k+1}, (A - \tilde{\theta}_{k+1} I) \tilde{x}_{k+1}, \dots, (A - \tilde{\theta}_{k+1} I)^m \tilde{x}_{k+1}\}$ 上的最小特征值, 所以 $\theta_{k+1} \leq \hat{\lambda}_{k+1}$ 。综上可得

$$\lambda = \lim \theta_{k+1} \leq \lim \hat{\lambda}_{n_k+1} = \hat{\lambda} \quad (12)$$

显然(11)与(12)矛盾。所以 $r = (A - \lambda I)x = 0$ 这说明 λ 是矩阵 A 的特征值。即 θ_k 收敛于矩阵 A

的一个特征值 λ , 而且 $\|(A - \lambda I)x_{n_k}\| \rightarrow 0$ 。

下面采用反证法证明 $\|(A - \lambda I)x_k\| \rightarrow 0$ 。假设存在一个子序列 m_k 使得 $\|(A - \lambda I)x_{m_k}\| \geq c \geq 0$ 。对于子序列 m_k 而言, 它有子序列 t_k 使得 x_{t_k} 收敛。因此, 根据上面的证明过程可得 $\|(A - \lambda I)x_{t_k}\| \rightarrow 0$ 。这样就产生矛盾, 所以 $\|(A - \lambda I)x_k\| \rightarrow 0$ 。

3 数值算例

我们用 Fortra77 采用双精度编写计算程序并在 CPU2.0G, 256M 内存的台式机上计算。算例中的初始向量都是随机产生的。为表述方便, 用 ITN 代表算法的迭代次数, CPU 代表所需要的计算时间, 单位为 s, 用 TRNW 表示截断 Newton 法, 用 SubNW 表示子空间加速的截断 Newton 法, m 代表子空间加速截断 Newton 法中选定的自然数。

例1 矩阵 A 是 1 200 阶的对称矩阵

$$A = \begin{bmatrix} 1 & 0.5 & & 0.5 \\ 0.5 & 2 & 0.5 & \\ \ddots & \ddots & \ddots & \\ & & & 0.5 \\ 0.5 & & & 0.5 & 1000 \end{bmatrix}$$

取精度 $\varepsilon = 1.0e-6$ 。计算 A 的最小特征值, 其结果见表1。

表1 例1的计算结果表

Table 1 Numerical results of Problem 1

程序名	特征值	ITN	CPU/s
TRNW	0.774 392 946 118 47	27	3.82
SubNW($m=1$)	0.774 392 987 217 51	20	3.30
SubNW($m=5$)	0.774 393 053 259 52	20	3.28

例2 矩阵 A 是 1 000 阶的对称矩阵

$$A = \begin{bmatrix} 4 & -1 & & 0 \\ -1 & 4 & 0 & \\ & -1 & \ddots & \ddots \\ & & \ddots & 4 & -1 \\ 0 & & & -1 & 4 \end{bmatrix}$$

取精度 $\varepsilon = 1.0e-6$, 计算 A 的最小特征值, 其结果见表2。

表 2 例 2 的计算结果表
Table 2 Numerical results of Problem 2

程序名	特征值	ITN	CPU/s
TRNW	2.000 010 041 187 83	13	2.71
SubNW($m=1$)	2.000 010 088 026 30	9	2.60
SubNW($m=5$)	2.000 010 092 375 10	8	4.00

例 3 矩阵 A 是 1 000 阶的对角矩阵 $\text{diag}(1.01, \dots, i+0.01, \dots, 1\ 000.01)$ ($i=1, \dots, 1\ 000$), 取精度 $\varepsilon=1.0e-6$ 。计算 A 的最小特征值,其结果见表 3。

表 3 例 3 的计算结果表
Table 3 Numerical results of Problem 3

程序名	特征值	ITN	CPU/s
TRNW	1.010 000 098 806 28	26	2.23
SubNW($m=1$)	1.010 000 113 898 84	20	2.10
SubNW($m=5$)	1.010 000 106 157 66	20	3.08

由以上数值结果可以看出子空间加速的截断 Newton 法比截断 Newon 法的迭代次数少,优于截断 Newon 法,并且是有效的算法。

参考文献:

[1] Absil P A, Baker C G, Gallivan K A. A truncated – CG style method for symmetric generalized eigenvalue problems[J]. Journal of Computational and Applied Mathematics,2006,189:274 – 285.

[2] Absil P A, Baker C G, Gallivan K A, et al. A daptive model trust region methods for generalized eigenvalue eproblems[J]. Lecture Notes in Compute Science,2005,35(14):33 – 41.

[3] Gerard L G, Van Der Vorst H A. A Jacobi – Davidson iteration method for linear eigenvalue problems[J]. SIAM J Matrix Anal Appl,1996,17:401 – 425.

[4] Sleijpen G L G, Van Der Vorst H A. Efficient expansion of subspaces in the Jacobi – Davidson method for standard and generalized eigenproblems[J]. ETNA,1998,7:75 – 89.

[5] Sameh A, Wisniewski J. A trace minimization algorithm for the generalized eigenvalue problem[J]. SIAM J Numer Anal, 1982,19:1 243 – 1 259.

[6] Perdon A M, Gambolati G. Extreme eigenvalues of large sparse matrices by Rayleigh quotient and modified conjugate gradients[J]. Comp. Methods App. Mech. Eng. 1986,56:251 – 264.

[7] Gloub G H, Ye Q. An inverse free preconditioned krylov subspace method for symmetric generalized eigenvalue problems [J]. SIAM J. Sci. Comput. ,2002,1(24):312 – 334.

A Subspace Accelerated Truncated Newton Method for Solving the Extreme Eigenvalue of Large – scale Sparse Symmetric Matrix

QIAN Xiao-yan, LIU Hao
(College of Science, Nanjing University of technology, Jiangsu Nanjing 210009, China)

Abstract: By using the subspace accelerated method, the authors improved the truncated Newton method for solving the eigenvalue of large – scale sparse symmetric matrix, and proposed a subspace accelerated truncated Newton method. Theoretical analysis and numerical results show the efficiency of the new method.

Keywords: symmetric matrix; eigenvalue; eigen – vector; truncated Newton method; subspace accelerated method

(责任编辑:张英健; 校对:沈建新)