

巨型机求解大型稀疏问题的 Krylov 子空间法

杨岳湘 李晓梅

(电子计算机系)

摘要 求解大型稀疏问题最流行的方法是建立在子空间投影技术之上的。这类方法的主要吸引力是仅需要使用矩阵向量乘法。我们给出 PCG、GMRES、Lanczos 和 Davidson 算法及在 YH-I 上的实现, 然后比较每一方法的优缺点, 并尽可能讨论其并行执行。

关键词 矩阵, 大型稀疏问题, Krylov 子空间

分类号 TP301.6

Krylov 子空间法是用子空间 $K_m = \text{span}(v, Av, \dots, A^{m-1}v)$ 的投影法, 其中 A 为 $N \times N$ 阶数矩阵, v 为任一单位向量。它的主要思想是利用 Krylov 子空间的信息构造近似解, 其特点是每次迭代重复使用大型稀疏矩阵向量乘法。巨型机的出现后, 用此种方法能有效地求解大型稀疏线性方程组 $AX=b$ 及大型稀疏特征值问题 $AX=\lambda X$ 。本文将上述两类问题统称为大型稀疏问题。

在工程计算中, 要解二维、三维数理方程等领域大型稀疏问题, 直接法开销太大, 而迭代法存贮省、并行度大、收敛速度快, 从而使这类方法解题效率很高。

虽然目前还没有充分理由说明哪一个算法最好, 但实践表明, 对大型稀疏线性方程组求解问题, 当 A 为对称正定时, PCG 算法最合适, 具体采用哪一个 PCG 算法依赖于巨型机的类型。当 A 非对称时, PGMES 算法最合适; 当 A 为对称不定时, Lanczos 算法最合适。对大型稀疏特征问题, Lanczos 算法和 Davidson 算法最合适。

多色排序、分块、预处理、稀疏矩阵向量乘等是提高并行性的关键技术。

1 求解大型稀疏线性方程组

线性方程组:

$$AX=b \quad (1)$$

给定一个初始向量 x_0 , 从 $x_0 + K_m$ 中寻找一个近似解 X_m , 使 $b - AX_m \perp L_m$, 其中 K_m, L_m 为 m 维子空间。Krylov 子空间方法中 $K_m = K_m(A, r_0) = \text{span}(r_0, Ar_0, \dots, A^{m-1}r_0)$, 其中 $r_0 = b - AX_0$ 。根据 K_m, L_m 的选择可以得到不同 Krylov 子空间方法。

(1) $L_m = K_m = K_m(A, r_0)$ 。当矩阵 A 对称正定时, 其轭梯度法即为特例。

(2) $L_m = AK_m, K_m = K_m(A, r_0)$ 。这种方示主要是非对称情形, X_m 使残差模 $\|b - Ax\|_2$ 在 $x_0 + K_m$ 上最小。

(3) $L_m = K_m(A^T, r_0), K_m = K_m(A, r_0)$ 。对称情况下, 这种方法成为第(1)种情形; 非对称情况下,

计算 $x_m = x_0 + U_m y_m$ 。

4) 重新计算。若满足条件则停止, 否则 $x_0 \leftarrow x_m$, 转到 2。

算法 2 在并行向量机上的执行

- 1) 建立预处理器;
- 2) 矩阵向量乘法;
- 3) 相对于一组正交向量, 正交化一个向量;
- 4) 向量更新;
- 5) 预处理操作。

上述两类算法有很多共同之处, 在巨型机上的主要瓶颈是: (1) 内积计算和矩阵向量乘法; (2) 建立预处理和预处理操作。通常巨型机上均提供内积指令或实用内积程序, 用户直接调用, 在上述算法中的内积还可以提前计算增加并行性, 见文献[7]。矩阵向量乘法文献[6]已作了详细研究, 我们已建立了相应高效软件。

预处理与机器类型有关, Dubois 的多项式预处理是基于问题的对角结构, 将原方程两边同乘以一个多项式: $S(A)AX = S(A)b$ 。Meijerink 和 Vander vorst 的不完全分解预处理, 最早是使用不完全循环约化法, 后来 Vander vorst 建议用一个简单的 Neumann 序列代替两对角线的逆。Axesson 和 Concus、Golub 各自独立地给出了块不完全分解。以上方法适于短向量巨型机。E·L·Poole 给出适合于长向量的多色不完全分解预处理。

1.3 Lanczos 算法

在实际问题中经常碰到系数矩阵对称不定的方程组, 即系数矩阵有正的特征值, 也有负的特征值。此时共轭梯度法失效, 因为此时 (Ax, y) 不是 x, y 的一种内积。C·C·Paige 等给出了 Lanczos 过程来求解这种方程组的方法。它的基本思想是: 利用 Lanczos 过程将方程组化成对称三对方程组。当 A 为稀疏矩阵时, 用到的只是 A 与向量 v 的乘法, 易于向量机上实现。详细过程类似于下节。

2 求解特征值问题

巨型机上求解大型稀疏特征值问题的最为有效的算法是 Krylov 子空间法。Lanczos 算法和 Davidson 算法是最著名的代表。Lanczos 算法求解对称矩阵特征值问题是五十年代提出来的。基本思想是取一个单位向量 v , 通过 Lanczos 过程构造一组正交化序列 v_1, v_2, \dots, v_n , $Q = [v_1, v_2, \dots, v_n]$, 则 $Q^*AQ = T$ 是一个对称三角阵。这样 A 的特征值问题化成三对角特征值问题。由于舍入误差影响, 在求 v_i 的过程中, 所求 v_i 很快失去正交性, 因此 Lanczos 算法在 50 年代到 60 年代认为不稳定, 很少被使用。直到 1971 年 Paige 通过对舍入误差分析, 发现失去的正交性恰与近似特征值精度提高相关, 重新肯定 Lanczos 算法是一种求解大型稀疏特征值问题的有效方法。特别是每次迭代用到的只是 A 与向量的乘法, 迭代过程中的计算都是向量间运算; 正适于向量并行处理。

算法 3 Lanczos 算法

开始 选择初始向量 v_1

迭代 $j = 1, 2, \dots, m, D_0$

$$y = Av_j$$

$$\text{如果 } j > 1, \text{ 计算 } y := y - \beta_j v_{j-1}$$

$$\alpha_j = (y, v_j)$$

$$y = y - \alpha_j v_j$$

$$\beta_{j+1} = \|y\|_2$$

$$v_{j+1} = y / \beta_{j+1}$$

式中主要迭代可以描述为三项递归式:

$$\beta_{j+1}v_{j-1} = Av_j - \alpha_j v_j - \beta_j v_{j-1}$$

α_j, β_{j+1} 被选择为使 v_{j+1} 与 v_j, v_{j-1} 正交。

Lanczos 算法在巨型机的执行：无重新正交化带来的好处是适于并行、向量处理。部分正交化和有选择正交化在本质上与无正交化相同。当使用全正交化时，必须执行一个好的 Gram-Schmidt 算法。

MCG (修正 Gram-Schmidt) 过程：

```
do k=1, m
  normalize v_k
  do j=k+1, m
    v_j := v_j - v_k(v_j, v_k)
```

其中 $v_k, k=1, \dots, m$ 已被正交化。内循环能并行化。为了提高效率，使用微任务和宏任务语句达到同步。

三对角阵 T 的特征值的求解，已有许多好的方法，我们使用多段算法，参见文[2]。

类似于 Lanczos 算法，Davidson 算法是建立在矩阵在子空间投影的基础上，其中子空间维数不断增加。在一定程度上，Davidson 方法可以认为是 Lanczos 方法的预处理版本。这种预处理与线性方程组的预处理技术大不一样。如果预处理高效的，则收敛快。主要缺点是每一次迭代计算量要随空间维数增加而增加。

Davidson 算法用于计算最大特征值，其描述如下：

算法 4 Davidson 算法

开始：选择初始单位向量 v_1 ；

迭代： $i=1, 2, \dots, D_0$

 迭代 $j=1, 2, \dots, k D_0$

$\omega_j = Av_j$ (仅计算最后一列)

$H_j = v_j^T \omega_j$ (仅计算最后一列)

 计算 H_j 最大特征对 (λ, y)

$x = v_j y$

$r = \omega_j y - \lambda x$ (残差)

 收敛测试

 if $i < k$;

$t = (M - \lambda_i)^{-1} r$ (M 为预处理阵)

$v_{i+1} = \text{MGS}([v_i, t])$ (修正 Gram-Schmidt 过程)

$v_1 = x$

其中预处理最简单形式是 Jacobi 预处理。若无预处理，即 $M=I$ ，则子空间序列同 Krylov 子空间等价，Lanczos 算法与 Davidson 算法理论上等价，但由于正交基不同，计算过程不同。

若需同时找几个特征值，则可以使用 Davidson 块方法，同时计算 H_i 的几个特征值对，并且几个向量被加到基向量 v_i 中。

Davidson 算法执行：

初始：给出第一块的初始猜测

外循环：

 内循环

 CALL MATMULI (A 乘最后的块)

 CALL MVTW (H 的更新)

 由 EISPACK 中的 TREDI--TRIDZB--TINVIT--TRBAK1) 计算 NB 个最大特征值和特征向

量;

CALL RITZ (计算 Ritz 向量);

收敛测试;

测试内循环结束否 (其中是否包含太多向量?)

CALL CORRECT (预处理新块的每一个向量);

CALL COMPL (新块在子空间正交部分上的投影);

CALL GRAMS (修正 Gram-Schmidt 正交化过程);

CALL GRAMS (在一个即将作为新迭代的 Ritz 向量上)。

Davidson 算法主要用于量子化学, 而 Lanczos 算法应用更广泛。

3 结 论

我们给出了巨型机上 Krylov 子空间法的主要技术和算法。它们的主要共同点有:

- (1) 每次迭代都要多次使用矩阵向量乘法;
- (2) 每次迭代中要进行的计算均是向量计算;
- (3) 采用稀疏存储结构, 节省空间。

线性方程组的预处理技术和内积计算, 矩阵向量乘法是共轭梯法的关键。我们给出了并行向量化措施。特征值问题 Lanczos 过程及 Davidson 计算方法的关键是矩阵向量乘法和正交化过程。

参 考 文 献

- 1 Parlet B N. The Symmetric Eigenvalue Problem. Prentice Hall, Englewood Cliffs, 1980
- 2 Lo S S, Philippe B and Sameh A. A Multiprocessor Algorithm for Symmetric Tridiagonal Eigenvalue Problem. SLAMJ Stat Sci Comp 8, 1987, (2)
- 3 Umar U M. Multitasking the Davidson Algorithm for the Large. Sparse Eigenvalue Problem, Supercomputer Applications, 1989, 3(2)
- 4 Youcef Saad B P. Solving Large Sparse Eigenvalue Problems on Supercomputer. N89-26423, 1989
- 5 蒋尔雄. 对称矩阵计算. 上海科技出版社, 1984
- 6 杨岳湘, 李晓梅. 大型稀疏矩阵向量乘法的并行计算. 计算机工程与科学. 1992, (3)
- 7 Rosendale J V. Minimizing Inner Product Data Dependences in Conjugate Gradient Iteration. IEEE 1983

The Krylov Subspace Methods to Solve Large Sparse Problems on Supercomputers

Yang Yuexiang Li Xiaomei

(Department of Computer Science)

Abstract

The most popular methods to solve large sparse problems are based on projection techniques on appropriate subspaces. The main attraction of these methods is that they only require using the matrix by vector multiplications. We give the implementations of PCG, GMRES, Lanczos's method and Davidson's method on YH-1. We then compare their advantages and disadvantages. Finally we discuss the parallel implementations as possible as we can.

Key words matrix, large sparse problems, Krylov subspace