RAL-TR-2000-010

The Lanczos method with semi-inner product

Karl Meerbergen

ABSTRACT
The spectral transformation Lanczos method is very popular for solving large scale Hermitian generalized eigenvalue problems. The method uses a special inner product so that the symmetric Lanczos method can be used. Sometimes, a semi-inner product must be used. This may lead to instabilities and break-down. In this paper, we suggest a cure for breakdown by use of an implicit restart in the Lanczos method.

Keywords: Lanczos method, semi-inner product, implicit restart

AMS (MOS) subject classifications: 65F15,65F50

Computational Science and Engineering Department
Atlas Centre
Rutherford Appleton Laboratory
Oxon OX11 0QX

March 31, 2000

Contents

1 Introduction

In structural analysis (Grimes, Lewis and Simon 1986, Grimes, Lewis and Simon 1994) and the determination of the stability of the Stokes problem (Malkus 1981), we want to compute a number of selected eigenvalues λ and corresponding eigenvectors x of

$$Ax = \lambda Bx \tag{1}$$

with A and B Hermitian and B positive semi-definite or ill-conditioned. Very often, B has an explicit zero block

$$B = \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} , \tag{2}$$

with M Hermitian positive-definite. A popular method for solving this problem is the spectral transformation Lanczos method (Ericsson and Ruhe 1980, Ericsson 1986, Nour-Omid, Parlett, Ericsson and Jensen 1987). This is the Lanczos method (Lanczos 1950) applied to the spectral transformation $(A - \mu B)^{-1}B$. Since this transformation is self-adjoint with respect to the B inner-product, the Lanczos method for Hermitian matrices can be used when the B inner product is used.

The B inner-product is a semi inner-product but it is mathematically correct to use it within the Lanczos method (Nour-Omid et al. 1987). The major problem is that this inner-product does not see components in the nullspace of B. These components may grow in an uncontrolled way. When B has an explicit zero block, these components do not disturb the Lanczos method. However, when B does not have an explicit zero block, large components in the nullspace of B may dominate the Lanczos vectors, which may have few significant digits. It is even possible that inner products that should be positive become negative, which leads to breakdown. This paper deals with curing breakdown. The idea is to use an implicit restart (Sorensen 1992) to limit the growth of components in the nullspace of B. It was first proposed in (Meerbergen and Spence 1997) for the determination of the stability of steady-states of the (Navier) Stokes equations, and used in (Lehoucq and Scott 1997) for the solution of the Navier-Stokes problem where B has an explicit zero block, and later studied for the rational Krylov method (De Samblanx, Meerbergen and Bultheel 1997, Meerbergen and Scott 2000). In structural applications, B may not have an explicit zero block, but a large number of very small eigenvalues. This paper analyses the application of an implicit restart when B does not have an explicit zero block.

The paper is organized as follows. In §2, we introduce the eigenvalue problem we want to solve and give the properties that are of most importance to this paper. In §3, we explain the Lanczos method and give an example that leads to breakdown. We also introduce implicit restarting and suggest a cure for breakdown. In §4, we analyse the Lanczos method, the growth of the two-norm of the Lanczos vectors, and the reliability of an implicit restart in this context. Section 5 presents a practical algorithm and illustrates it using a small example. Final conclusions are given in §6. Throughout the paper, we use $\mathcal{R}(A)$ and $\mathcal{N}(A)$ to denote the range and nullspace of A respectively. By $\mathbf{u}$, we denote the precision for finite precision arithmetic.

2 Analysis of the eigenvalue problem

The Lanczos method is not able to solve (1) directly, but instead solves the problem $Sx = \theta x$ with

$$S = (A - \mu B)^{-1} B \ . \tag{3}$$

We assume that $A - \mu B$ is invertible. In order to understand the behaviour of the Lanczos method, we must understand the spectral transformation. The matrix S has the same eigenvectors as (1) and the eigenvalues are given by $\theta = (\lambda - \mu)^{-1}$ which explains the term spectral transformation.

When B is singular, (1) has an infinite eigenvalue with the nullspace of B as eigenspace. The associated eigenvalues of S are zero. When the zero eigenvalues of S are non-defective, the total space $\mathbf{C}^n$ is the direct sum of $\mathcal{N}(S)$ and $\mathcal{R}(S)$. The range is an invariant subspace of S and $S\mathbf{C}^n = \mathcal{R}(S)$. The Lanczos method, described in the following section, is designed for computing the nonzero eigenvalues and the corresponding eigenvectors of S, i.e. eigenvectors in $\mathcal{R}(S)$. The method is not able to compute the eigenvectors in the nullspace of S. As is shown in Ericsson (1986) and Malkus (1981), the zero eigenvalue may be defective. i.e. may have geometric multiplicity p and algebraic multiplicity $2p$ where p is the nullity of B. We give a small example as illustration.

Example 2.1 Let

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & \beta - 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \ ,$$

then the spectral transformation

$$S = A^{-1}B = \begin{bmatrix} \beta^{-1} & -\beta^{-1} & 0 \\ -\beta^{-1} & \beta^{-1} & 0 \\ (\beta - 1)/\beta & \beta^{-1} & 0 \end{bmatrix}$$

has eigenvalues $\theta_1 = 2\beta^{-1}$ and an eigenvalue 0 with algebraic multiplicity two and geometric multiplicity one. The eigenvalues of (1) are $\beta/2$ and ∞ and the corresponding eigenvectors are

$$x_1 = \begin{pmatrix} 1 \\ -1 \\ 1 - \theta_1^{-1} \end{pmatrix} \quad , \quad x_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \ .$$

The vector $x_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ is a generalized eigenvector or principal vector, since $S^2 x_3 = 0$, but $Sx_3 \neq 0$. We now decompose $\mathbf{C}^n$ into $\mathcal{N}(S^2)$ and $\mathcal{R}(S^2)$. In this example,

$$S^2 = \begin{bmatrix} 2\beta^{-2} & -2\beta^{-2} & 0 \\ -2\beta^{-2} & 2\beta^{-2} & 0 \\ \beta^{-1} - 2\beta^{-2} & -\beta^{-1} + 2\beta^{-2} & 0 \end{bmatrix}$$

which has nullspace $\mathcal{N}(S^2) = \text{span}(x_2, x_3)$ and range $\mathcal{R}(S^2) = \text{span}(x_1)$. Note that $\mathcal{R}(S^2) \not\perp \mathcal{N}(B)$, so vectors in $\mathcal{R}(S^2)$ may still have components in $\mathcal{N}(B)$. For x_1 in this example, this component is $1 - \theta_1^{-1}$ and is large when θ_1 is small.

The nullspace of B is $\text{span}(x_2)$ and is contained by $\mathcal{N}(S^\nu)$. In general, we use the spaces $\mathcal{N}(S^\nu)$ and $\mathcal{R}(S^\nu)$ where ν is the index of the zero eigenvalue of S. Note that in all situations $\mathcal{N}(B) \subset \mathcal{N}(S^\nu)$.

3 The spectral transformation Lanczos method

The spectral transformation Lanczos method is the Lanczos method applied to the spectral transformation $S = (A - \mu B)^{-1}B$ where μ is called the pole. We denote by x^*By the B inner-product of x and y and by $\|x\|_B$ the B norm of x, defined by $\sqrt{x^*Bx}$. The method can also be used when B is singular (Ericsson and Ruhe 1980, Ericsson 1986, Nour-Omid et al. 1987).

Algorithm 3.1 (Lanczos method)

1. Given an initial vector v_1 with $\|v_1\|_B = 1$.
Let $\beta_0 = 0$ and $v_0 = 0$.
2. For $j = 1$ to k do
2.1. Form $w_j = Sv_j$.
2.2. Form $w'_j = w_j - v_{j-1}\beta_{j-1}$.
2.3. Form $\alpha_j = v_j^* w'_j$.
2.4. Form $w''_j = w'_j - v_j\alpha_j$.
2.5. Let $\beta_j = \|w''_j\|_B$ and normalize $v_{j+1} = w''_j/\beta_j$.

The Lanczos method builds a B orthogonal basis $\{v_1, \ldots, v_{k+1}\}$ of the Krylov space spanned by $v_1, Sv_1, S^2v_1, \ldots, S^kv_1$. A practical implementation requires reorthogonalization in order to keep the Lanczos vectors B orthogonal (Grimes et al. 1994). Eliminating of w_j, w'_j and w''_j from Steps 2.1–2.5, we obtain the relationship

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

Collecting the equations for $j = 1, \ldots, k$ leads to

$$SV_k = V_{k+1}\underline{T}_k$$

with $\underline{T}_k$ the $k+1 \times k$ matrix

$$\underline{T}_k = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \alpha_{k-1} & \beta_{k-1} \\ & & & \beta_{k-1} & \alpha_k \\ & & & & \beta_k \end{bmatrix},$$

and $V_{k+1}^* B V_{k+1} = I$. The matrix $\underline{T}_k$ is called the tridiagonal Lanczos matrix and the columns of V_{k+1} are the Lanczos vectors.

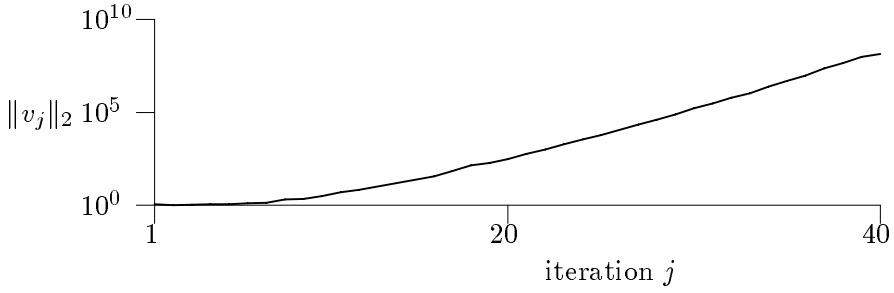

Figure 1: Growth of the two-norm of the Lanczos vectors.

The B orthogonal projection of the eigenvalue problem $Sx = \theta x$ on the range of V_k gives the solution (θ, x) with $x = V_k z$ and $T_k z = \theta z$, where T_k is the $k \times k$ upper submatrix of $\underline{T}_k$. The residual is

$$\begin{aligned} r &= Sx - \theta x \\ &= SV_k z - \theta V_k z \\ &= V_{k+1}\underline{T}_k z - V_k T_k z \\ &= v_{k+1}\beta_k e_k^* z \end{aligned}$$

and $\|r\|_B = \beta_k |e_k^* z|$ is cheaply computed.

Consider the following example.

Example 3.1 Let A and B be generated by the following Matlab code :

```
rand('seed',0)
A = diag(-[1:200]);
B = diag([ones(150,1);zeros(50,1)]);
Z = orth(randn(200));
A = Z'*A*Z;
B = Z'*B*Z;
```

For this example, $\mathcal{R}(S) \perp \mathcal{N}(S)$. Also, $\|x\|_2 = \|x\|_B = 1$ implies that $x \in \mathcal{R}(S)$ and $\|x\|_2 \neq 0$, but $\|x\|_B = 0$ implies $x \in \mathcal{N}(B)$. We ran 50 iterations of the Lanczos method with the initial vector $v_1 = [1, \ldots, 1]/\beta$ where β is so that $\|v_1\|_B = 1$ and pole $\mu = 0$. We used reorthogonalization on each iteration to ensure B orthogonality (Daniel, Gragg, Kaufman and Stewart 1976). Figure 1 shows $\|v_j\|_2$ as a function of the iteration number j. At iteration $j = 39$, the inner product $\|w_j\|_B^2 = w_j^* B w_j$ becomes negative, so β_j is imaginary. At this point, we must stop the calculations, since an imaginary B norm does not make any sense. The Lanczos vectors have large components in $\mathcal{N}(S)$, so the B inner-product may be small but should be positive in theory. Vectors in $\mathcal{N}(S)$ are B orthogonal to themselves, so cannot serve in a B orthogonal basis. The two-norms of v_j grow but the B norms remain equal to one, so v_j has fewer significant digits in $\mathcal{R}(S)$.

In order to get rid of large components in the nullspace of B, we suggest regularly performing an implicit restart (Sorensen 1992, Meerbergen and Spence 1997). The following theorem defines the implicit restart and shows important properties.

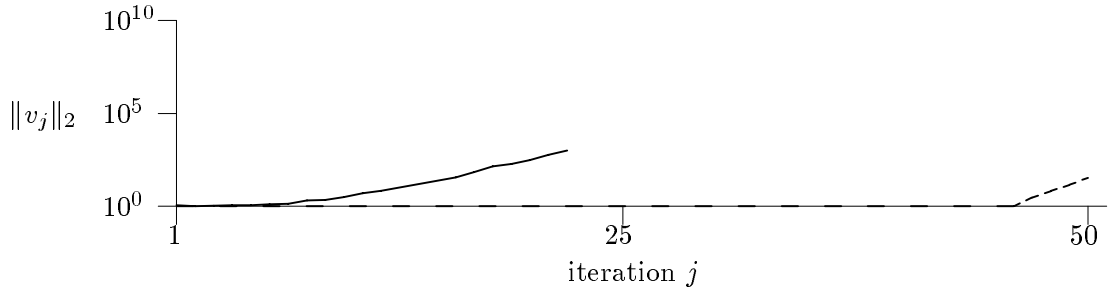

Figure 2: Growth of the two-norm of the Lanczos vectors The dashed line shows the two-norm after implicit restart.

Theorem 3.1 *Let V_{k+1} and $\underline{T}_k$ be computed by the Lanczos method. Consider the QR factorization $\underline{T}_k = \underline{Q}_k R_k$ with $\underline{Q}_k \in \mathbf{C}^{k+1\times k}$ unitary and $R_k \in \mathbf{C}^{k\times k}$ upper triangular. Let $\underline{Q}_{k-1}$ be the $k \times k-1$ upper left part of $\underline{Q}_k$ and define*

$$V_k^+ = V_{k+1}\underline{Q}_k \quad \textit{and} \quad \underline{T}_{k-1}^+ = R_k \underline{Q}_{k-1} \ .$$

Then V_k^+ and $\underline{T}_{k-1}^+$ have the following properties :

1. $V_k^{+*} B V_k^+ = I$
2. *If $\beta_k \neq 0$, $\mathcal{R}(V_k^+) = \mathcal{R}(SV_k)$*
3. *$\underline{T}_{k-1}^+$ is Hermitian tridiagonal and for $V_k^+ = [V_{k-1}^+ \ v_k^+]$,*

$$SV_{k-1}^+ = V_k^+ \underline{T}_{k-1}^+$$

with $v_1^+ = Sv_1/\|Sv_1\|_B$.

The proof is a compilation of results from (Sorensen 1992) and is given in (Meerbergen and Spence 1997). A similar proof is given in (Lehoucq 1996). The implicit restart is defined as the transition of the pair V_{k+1}, $\underline{T}_k$ to the pair V_k^+, $\underline{T}_{k-1}^+$. The theorem tells us that an implicit restart applies S implicitly on the Krylov subspace. This operation also reduces the dimension of the subspace by one.

We performed an implicit restart in the example when $\|v_j\|_2 \geq 10^4$, i.e. after iteration 22. The Lanczos vectors V_{22}^+ have two-norms equal to one, see Figure 2. The new Lanczos vectors added to the subspace retain small two-norm, until the norm starts growing again from the 45th iteration.

4 Analysis

Roughly speaking, the reason why the implicit restart filters away $\mathcal{N}(B)$ is that the nullspace of S is filtered away. The components of v_j in $\mathcal{N}(B)$ are responsible for large $\|v_j\|_2$ and potential breakdown, so their removal reduces $\|v_j\|_2$ and makes the method more reliable with respect to breakdown.

4.1 A bound for the two-norm of the Lanczos vectors

First we prove that limiting v_j to the $\mathcal{R}(B)$ limits the two-norm of v_j.

Lemma 4.1 *When* $\|v_j\|_B = 1$ *then*

$$(\lambda_{\max}(B))^{-1/2} \leq \|v_j\|_2$$

When in addition $v_j \in \mathcal{R}(B)$, *then*

$$(\lambda_{\max}(B))^{-1/2} \leq \|v_j\|_2 \leq (\lambda_{\min,\lambda\neq 0}(B))^{-1/2} \ .$$

Proof Decompose $B = L^*L$. Since $\|v_j\|_B = 1$, we have that $\|Lv_j\|_2 = 1$. The first statement and the first inequality of the second statement follow from $\|Lv_j\|_2 \leq \|L\|_2\|v_j\|_2$. If, in addition, B is nonsingular, $\|v_j\|_2 = \|L^{-1}Lv_j\|_2 \leq \|L^{-1}\|_2\|Lv_j\|_2 = \|L^{-1}\|_2$ which proves $\|v_j\|_2 \leq (\lambda_{\min}(B))^{-1/2}$. When $v_j \in \mathcal{R}(B)$, then $\|v_j\|_B = \|v_j\|_{B_{|\mathcal{R}(B)}}$, where $B_{|\mathcal{R}(B)}$ is the restriction of B onto $\mathcal{R}(B)$. This proves the lemma. □

The idea is to reduce the two-norm of the Lanczos vectors by the application of implicit restarts, i.e. by removing the nullspace of S^ν from the basis vectors. The following theorem shows an upper bound to $\|v\|_2$ when x is an eigenvector of S lying in $\mathcal{R}(S^\nu)$. In general, vectors in $\mathcal{R}(S^\nu)$ still have components in the nullspace of B so the above bound is not valid.

Theorem 4.2 *Let* $Sx = \theta x$, $\theta \neq 0$ *and* $\|x\|_B = 1$, *then*

$$\|x\|_2 \leq |\theta^{-1}|\|S\|_2(\lambda_{\min,\lambda\neq 0}(B))^{-1/2} \ .$$

Proof Decompose $x = x_1 + x_2$ with $x_1 \in \mathcal{R}(B)$ and $x_2 \perp \mathcal{R}(B)$. Then we have $\|x\|_B = \|x_1\|_B = 1$ and $\|x_1\|_2 \leq (\lambda_{\min,\lambda\neq 0}(B))^{-1/2}$. Since $Sx_2 = 0$, we have $x = \theta^{-1}Sx = \theta^{-1}Sx_1$, so $\|x\|_2 = |\theta^{-1}|\|Sx_1\|_2$, which proves the theorem. □

An illustration was given in Example 2.1. When β increases, θ_1 becomes smaller and the component of x_1 in the nullspace of B becomes larger. This theorem shows that if the Krylov subspace has good approximations to eigenvectors corresponding to eigenvalues of S with small absolute value, reducing $\|v_j\|_2$ by removing the nullspace of S may be insufficient to reduce components in the nullspace of B. We could remove all small eigenvalues of S from the Krylov subspace. This can be achieved by applying S^ν to the Lanczos basis with ν large enough. This filtering can be done by an implicit restart and was illustrated successfully in Example 3.1.

In some practical situations, B does not have an explicit zero block, but has a number of eigenvalues with very small absolute values. In this situation, we should again try to remove the components corresponding to the eigenvalues of S with small modulus.

When B is positive semi-definite and the initial vector v_1 lies in $\mathcal{R}$, all Lanczos vectors lie in $\mathcal{R}$, since $\mathcal{R}$ is an invariant subspace of S. Breakdown cannot occur. In finite precision arithmetic, components from $\mathcal{N}(S)$ may be introduced in the Lanczos vectors so that $\|v_j\|_2$ may become large and breakdown is possible. It is good practice to start the Lanczos method with $v_1 = S^\nu \cdot v$ with v randomly chosen so that $v_1 \in \mathcal{R}(S^\nu)$. This reduces the chance that v_j is quickly corrupted by components in $\mathcal{N}(S)$. When B has eigenvalues with very small absolute values, the conclusions are similar. In this case $\mathcal{N}(S)$ is the invariant subspace corresponding to these eigenvalues.

4.2 Analysis of the Lanczos recurrence relation

The Lanczos method orthogonalizes Sv_j against $v_1, \ldots, v_j$. It subtracts a linear combination of $v_1, \ldots, v_j$ from Sv_j. In finite precision arithmetic, the error on the recurrence relation is thus

$$Sv_j - V_{j+1}t_j = f_j$$

with

$$\|f_j\|_2 \ll \||V_{j+1}| \cdot |t_j|\|_2 \leq \||V_{j+1}|\|_2 \|t_j\|_2 \ .$$

On the global level, we can conclude that

$$SV_k - V_{k+1}\underline{T}_k = F_k$$

and

$$\|F_k\|_2 \ll \||V_{k+1}| \cdot |\underline{T}_k|\|_2 \leq \||V_{k+1}|\|_2 \|\underline{T}_k\|_2 \ .$$

In practice we want that for a Ritz pair (θ, x) with $x = V_k z$, the residual

$$r = SV_k z - \theta V_k z = v_{k+1}\beta_k e_k^T z + F_k z \approx v_{k+1}\beta_k e_k^T z \ .$$

Often, we want $\|r\|_B \leq \sqrt{\mathbf{u}}\|\underline{T}_k\|_2$, so we also want the term $\|F_k z\|_B$ much smaller than $\|\underline{T}_k\|_2$. This is satisfied if $\||V_{k+1}|\|_2 \leq 1/\sqrt{\|B\|_2}$. When $\||V_{k+1}|\|_2$ is large, the major problem is that the product $V_{k+1}\underline{T}_k$ has modest norm, but the terms in the summation may have large modulus. This is a typical situation where rounding errors may be large due to cancellation. In Example 3.1, we have that $\|F_1\|_2 \simeq 9\,10^{-17}$ and $\|F_{38}\|_2 \simeq 4\,10^{-9}$. This matches the observations that $\|\underline{T}_{38}\|_2 \simeq 1$ and $\|v_1\|_2 \simeq 1.12$ and $\|v_{39}\|_2 \simeq 9.9\,10^7$. In the last Lanczos iteration, we lose 7 digits through cancellation.

4.3 Analysis for the implicit restart

The idea is to apply S^ν to the Lanczos vectors so that they are free of components in $\mathcal{N}(B)$. The expensive way is to explicitly multiply V_{k+1} ν times by S. A cheaper but equivalent way is an implicit restart.

Assume that

$$\begin{aligned} SV_k - V_{k+1}\underline{T}_k &= F_k & \text{(4a)} \\ V_{k+1}^* B V_{k+1} &= I + \Gamma_{k+1} & \text{(4b)} \end{aligned}$$

where the terms F_k and Γ_{k+1} are the errors from computations in finite precision. We assume that $\|F_k\|_2 \ll \||V_{k+1}|\|_2 \|\underline{T}_k\|_2$ and $\|\Gamma_{k+1}\|_2 \ll 1$.

An implicit restart produces $\underline{Q}_k$ and R_k so that $\underline{T}_k = \underline{Q}_k R_k + \Delta$ with $\|\Delta\|_2 \ll \|\underline{T}_k\|_2$. From (4a), we then derive that

$$SV_k - V_{k+1}\underline{Q}_k R_k - V_{k+1}\Delta = F_k$$

and

$$V_{k+1}\underline{Q}_k = (SV_k)R_k^{-1} - (V_{k+1}\Delta + F_k)R_k^{-1} \ .$$

The left-hand side forms the new basis vectors after the implicit restart. They clearly span the columns of SV_k, when the second term in the right-hand side is small. This is the case when $\|R_k^{-1}\|_2$ is not very large. In (Meerbergen and Spence 1997), we find that

$$\|R_k^{-1}\|_2 \geq \left(\min(|\theta|^2 + \rho^2)^{1/2}\right)^{-1}$$

where θ is a Ritz value and $\rho = \beta_k |e_k^T z|$ the corresponding residual norm. So, when T_k has a small eigenvalue with a small corresponding residual norm the implicit filtering may lose accuracy. Since, usually, $\|\Delta\|_2$ is of the order of $\|\underline{T}_k\|_2 \mathbf{u}$ and $\|F_k\|_2$ of the order of $\||V_{k+1}|\|_2 \|\underline{T}_k\|_2 \mathbf{u}$, we may conclude that, roughly speaking,

$$\|V_{k+1}\underline{Q}_k - SV_kR_k^{-1}\|_2 \simeq \mathbf{u}\||V_{k+1}|\|_2 \kappa_2(\underline{T}_k)$$

where κ_2 denotes the two-norm condition number. For Example 3.1, $\kappa_2(\underline{T}_k) \simeq 148$, which is small. So, the filtering is almost perfect.

Since $T_k = V_k^* B S V_k$, $V_{k+1}^* B V_{k+1} = I$ and B filters away the zero eigenvalues of S it is very unlikely that T_k has an eigenvalue very close to zero. There may be a problem when 0 is a defective eigenvalue of S. In this case, T_k may have a zero eigenvalue (Ericsson 1986, Meerbergen and Spence 1997). One way to prevent this is regularly performing implicit restarts.

5 Numerical example

Algorithm A practical algorithm using the ideas discussed in this paper is the following.

Algorithm 5.1

1. *Choose σ, an initial vector v and the power ν.*
2. *Prefilter $v_1 = S^\nu \cdot v$ and normalize $v_1 \leftarrow v_1/\|v_1\|_B$.*
3. *Perform k steps of the Lanczos method. If $\|v_j\|_2 > \mathbf{u}^{-1/2}\|v_1\|_2$, go to Step 4. If $v_j^* B v_j < 0$, let $j = j-1$, go to Step 4.*
4. *Perform ν implicit restarts.*
5. *Compute Ritz values and Ritz vectors.*
6. *Check convergence.*
7. *Continue the Lanczos method from iteration $j+1$.*

Since $v_j^* B v_j < 0$ leads to breakdown, we hope that it is curable by applying an implicit restart on the first $j-1$ Lanczos iterations. When $\|v_j\|_2$ becomes too large, we also perform ν implicit restarts in order to reduce this norm. Since an implicit restart for $j < 2$ does not make any sense, a breakdown for $j = 2$ is incurable. This algorithm is implemented in the code `EA16` (Meerbergen and Scott 2000, HSL 2000).

An example Consider the 200×200 matrices $A = L^T D_A L$ and $B = L^T D_B L$ where

$$D_A = \begin{bmatrix} \mathrm{diag}(1, \ldots, 150) & I_{150,50} \\ I_{50,150} & 0_{50} \end{bmatrix} \qquad \text{and} \qquad D_B = \begin{bmatrix} I_{150} & 0 \\ 0 & M_{50} \end{bmatrix},$$

M_{50} a diagonal matrix with uniformly distributed elements between -10^{-10} and 10^{10}, and L a lower triangular matrix with ones on the main diagonal and with the off-diagonal elements uniformly distributed between -0.1 and 0.1. The condition number of B is of the order of $2.5\,10^{18}$. If M_{50} was the zero matrix, the index of the zero eigenvalue of S would be $\nu = 2$. We used `EA16` to compute the leftmost eigenvalue using shift-invert mode (`MODE=4`), blocksize `BLK=1`, pole `SIGMA=0`, and using `NV=31` Lanczos vectors. The experiments were performed on a SUN SPARC Ultra 1 with the EPC Fortran 90 compiler. The code automatically performs an implicit restart when the inner product becomes negative or the ratio of $\|v_1\|_2/\|v_j\|_2$ is smaller than $\sqrt{\mathbf{u}}$ as suggested in Algorithm 5.1.

The following table shows at which iteration breakdown ($v_j^* B v_j < 0$) occurs for different values of ν.

ν	0	1	2	3	4	5	6
j	19	18	29	28	28	36	40

The iteration number j at which breakdown occurs increases as ν increases, since the initial vector has smaller components in the eigenvectors corresponding to the small eigenvalues of S.

When we use Algorithm 5.1, an implicit restart is performed when the inner-product is negative or $\|v_j\|_2$ is too large, which improves the reliability of `EA16`. For all values of $\nu > 0$ listed in the table above, the code completes the computations and returns the wanted eigenvalues with the desired accuracy. For $\nu = 0$, the algorithm breaks down.

We have computed the eigenvalues nearest zero using the code `ARPACK` (Lehoucq, Sorensen and Yang 1998) in shift-invert mode with pole $\sigma = 0$. The eigenvalues returned are -1.365, 1.34, and 1.476, which are completely wrong. `ARPACK` is using a pre-filtered initial vector $v_1 = Sv$ but not an implicit restart (as in the spirit of this paper). However, when M_{50} has positive diagonal elements, `ARPACK` returns the right answers 51, 52 and 53.

An more difficult example We now solve the same problem with the same parameters for `EA16` but with

$$D_A = \begin{bmatrix} 10^5\,\mathrm{diag}(1,\ldots,150) & I_{150,50} \\ I_{50,150} & 0_{50} \end{bmatrix}$$

which only differs in the factor 10^5. The λ's are large, so the θ's are small. Following Theorem 4.2, this may lead to larger components in $\mathcal{N}(B)$. For this example the Lanczos method breaks down after the first or second iteration for $\nu = 0,\ldots,20$. We did not try higher values of of ν.

6 Conclusions

In this paper, we suggested the use of implicit restarts to avoid breakdown of the Lanczos method with semi inner-product. We conclude that this method works when the eigenvectors of S corresponding to the nonzero eigenvalues of S have small components in the nullspace of B.

The numerical results are promising, but the solution of eigenvalue problems with ill-conditioned mass matrix remains a difficult problem. Usually problems of this form arise

from DAE's (differential algebraic equations) which may be transformed into eigenvalue problems that have more favourable structure for the Lanczos method, e.g. with a well conditioned positive definite mass matrix.

Acknowledgement

The author thanks Jennifer Scott from the Rutherford Appleton Laboratory for reading an earlier draft of this manuscript.

References

Daniel, J., Gragg, W., Kaufman, L. and Stewart, G. (1976), 'Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization', *Math. Comp.* **30**, 772–795.

De Samblanx, G., Meerbergen, K. and Bultheel, A. (1997), 'The implicit application of a rational filter in the RKS method', *BIT* **37**, 925–947.

Ericsson, T. (1986), A generalised eigenvalue problem and the Lanczos algorithm, *in* J. Cullum and R. Willoughby, eds, 'Large Scale Eigenvalue Problems', Elsevier Science Publishers BV, pp. 95–119.

Ericsson, T. and Ruhe, A. (1980), 'The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems', *Math. Comp.* **35**, 1251–1268.

Grimes, R., Lewis, J. and Simon, H. (1986), Eigenvalue problems and algorithms in structural engineering, *in* J. Cullum and R. Willoughby, eds, 'Large Scale Eigenvalue Problems', Elsevier Science Publishers B.V., pp. 81–93.

Grimes, R., Lewis, J. and Simon, H. (1994), 'A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems', *SIAM J. Matrix Anal. Applic.* **15**, 228–272.

HSL (2000), 'A collection of Fortran codes for large scale scientific computation'.

Lanczos, C. (1950), 'An iteration method for the solution of the eigenvalue problem of linear differential and integral operators', *J. Res. Nat. Bur. Stand.* **45**, 255–282.

Lehoucq, R. (1996), Restarting an Arnoldi reduction, Preprint MCS-P591-0496, Argonne National Laboratory, Argonne, IL 60439.

Lehoucq, R. B. and Scott, J. A. (1997), Implicitly restarted Arnoldi methods and eigenvalues of the discretized Navier-Stokes equations, Technical Report RAL-TR-97-058, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK. Available through http://www.numerical.rl.ac.uk/reports/reports.html.

Lehoucq, R. B., Sorensen, D. C. and Yang, C. (1998), *ARPACK USERS GUIDE: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, PA.

Malkus, D. (1981), 'Eigenproblems associated with the discrete LBB condition for incompressible finite elements', *Internat. J. for Engrg. Sci.* **19**, 1299–1310.

Meerbergen, K. and Scott, J. (2000), Design and development of a block rational Lanczos method with partial reorthogonalization and implicit restarting, Technical Report RAL-TR-2000-011, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK. Available through `http://www.numerical.rl.ac.uk/reports/reports.html`. Information on HSL2000 available on `http://www.numerical.rl.ac.uk/hsl/`.

Meerbergen, K. and Spence, A. (1997), 'Implicitly restarted Arnoldi and purification for the shift-invert transformation', *Math. Comp.* **66**, 667–689.

Nour-Omid, B., Parlett, B., Ericsson, T. and Jensen, P. (1987), 'How to implement the spectral transformation', *Math. Comp.* **48**, 663–673.

Sorensen, D. (1992), 'Implicit application of polynomial filters in a k-step Arnoldi method', *SIAM J. Matrix Anal. Applic.* **13**, 357–385.