150 Years old and still alive: eigenproblems

Henk A. van der Vorst

Utrecht University

Gene H. Golub

Stanford University

Dedicated to J.H. Wilkinson
on the 10^{th} Anniversary of his death

1 Introduction

Linear eigenproblems continue to be an important and highly relevant area of research in numerical linear algebra. Therefore, it should be no surprise that numerical algorithms for eigenproblems are among the oldest known in the modern literature. In 1846, exactly 150 years ago, Jacobi [33] wrote his famous paper on the computation of solutions for the problem $Ax = \lambda x$, with $A^T = A$. Strictly speaking, this is not correct, since matrix notation was unknown at that time. So Jacobi formulated the problem in terms of systems of equations, written out elementwise. His paper contains computational elements that are still in use, like the plane rotations for making a system more diagonally dominant, and the (Gauss)-Jacobi iteration method for diagonally dominant systems. These elements were combined in his approach for the computation of eigenvalues, together with a clever way of setting up a correction equation for eigenvalue and eigenvector approximations.

Later on other methods became popular, such as the Power iteration method, and, when tools became more sophisticated, eigenproblems other than the standard $Ax = \lambda x$ problem required solution: for instance, the generalized eigenproblems $Ax = \lambda Bx$, higher order polynomial eigenproblems

$$(A_0 + \lambda A_1 + ..\lambda^n A_n)\, x = 0,$$

and, of course, the singular value decomposition.

Through pioneering work of such people as Arnoldi, Francis, Givens, Householder, Kublanovskaya, Lanczos, Ostrowski, Rutishauser, Wilkinson, and many others, a sophisticated toolbox of algorithms, together with analysis, became

available. This work was continued by Kahan, Paige, Parlett, Stewart, to mention only a handful of leaders in a lively and productive research area. Ten years ago we were in the situation that many eigenproblems, of order a few hundreds at most, could be solved routinely. This toolbox contains algorithms like, for instance, Householder reduction, the QR-algorithm, and the numerically stable SVD. Quite essential in the computation of eigenvalues and eigenvectors is the concept of inverse iteration, which leads to quadratic convergence or even cubic convergence (in the diagonalizable case) [52, 60].
Most of these algorithms had been described in the famous book by Wilkinson and Reinsch [77], following their earlier publication in Numerische Mathematik. The Algol-60 codes in these documents formed the basis for the standard libraries EISPACK and LAPACK, and modern codes in NAG, and other libraries, still bear the traces of these 'ancient' Algol-60 procedures. Most of this software is available through *netlib*.

With larger problems coming into the picture, it soon became clear that the matrix transforming techniques could not solve these problems with reasonable computing resources, and, as an alternative, iterative methods were investigated. The Power method was not robust enough, although the method is still hidden in state-of-the-art methods, like the powerful (literally) QR method [30, 57]. Lanczos [34] and Arnoldi [2] started the research on modern iteration methods in the early fifties. After a period of little interest in these methods, mainly because of poor understanding of their numerical properties, Paige [53, 54] showed the potential of the Lanczos method. This marked the start of an entire new area of research. About the same time Davidson, a chemist, designed a popular method that became widely known as Davidson's method [18]. Among numerical analysts the method did not receive much attention, partly because it appeared to be applicable only to diagonal dominant problems, and partly because its effects were poorly understood from a numerical point of view. The Lanczos method, for symmetric matrices, was well understood in 1986, see for instance [57], but the methods for unsymmetric systems have begun to blossom by that time.

It should be noted that in the modern iterative *subspace* methods, like Arnoldi's, Lanczos', and Davidson's method, the given large problem is reduced to a much smaller problem. This smaller problem can then be solved by the, by now standard techniques for dense matrices.

This is a rough sketch of the situation halfway through the eighties, say 1986. In the remaining part of this paper we will highlight what we consider to be the main new developments during the last 10 years.

We are dedicating our paper to J.H. Wilkinson, who died on October 5, 1986. He led the way in numerical linear algebra, and we have all benefited by his vision.

2 Highlights of the period 1986-1996

In the past ten years we have seen many new developments in this exciting area. In our opinion the major steps, from an algorithmic point of view, are:

- Attempts to make the two-sided Lanczos process more robust, by introducing a so-called Look-ahead strategy [59, 27, 9, 32].
- The idea of an implicit restart technique for the Arnoldi process, which helps to keep memory requirements reasonable, and which makes the Arnoldi process an attractive algorithm for eigencomputations [68, 38].
- Further improvements on the Davidson method, culminating in the Jacobi-Davidson algorithm [66]. In this algorithm a major deficiency in the original Davidson method has been removed. The Jacobi-Davidson method can be implemented as an accelerated inner-outer iteration scheme.

Of course, much more has happened, and the above reflect only personal impressions. Other important developments and improvements were (the list is still personally colored and incomplete):

- The concept and use of the ε-pseudospectrum [71] as a means to make sensitivities in the spectrum of a nonnormal matrix easily visible. The pseudospectrum does not make classical perturbation theory superfluous, but it helps to detect situations that need further analysis, and it does so in a way that is easily understandable for non-expert users. We will discuss the use of this tool briefly.
- The convergence behavior of Ritz values, for symmetric matrices, has been further analyzed and is now quite well understood. The so-called misconvergence phenomenon, a Ritz value that lingers near some eigenvalue before it converges towards some other eigenvalue [73, 58], and various local effects in the convergence behavior have been analysed and are fairly well understood by now [73]. The notion that rounding errors lead to multiple (spurious) eigenvalues, has been translated into practical strategies also for the unsymmetric Lanczos process [16].
- Strategies have been proposed to improve the efficiency of the various subspace methods. One type of approach amounts to filtering undesired eigenvector components from the starting vector: Chebychev polynomial preconditioning [65]:Chapter VII. The general idea goes back to Flanders and Shortley [25]; Lanczos also suggested polynomial filtering techniques [35]. Other strategies aim for improving the speed of convergence by considering a transformed problem: rational shift-and-invert techniques [62, 63], inexact shift-invert preconditioning for eigenproblems [50, 45], and [65]: Chapter VII; for a general discussion, see also [46].

- Generalizations of the Davidson method, making the method also suitable for unsymmetric matrices. These methods come down to the incorporation of more general preconditioning, instead of the original proposed diagonal preconditioning [49, 48, 14], or to second order corrections for the current eigenvector approximations [51]. The latter variant is strongly related to the Jacobi-Davidson type of algorithms.

- Eigenproblems can be regarded as nonlinear problems, which means that one can employ homotopy methods for the efficient computation of eigenvalues, or tracking these eigenvalues as function of a parameter in the underlying model [43, 44]. We will sketch this approach and point at some relations with other iterative methods

- Parallelism (Cuppen's Divide and Conquer [17, 22], Restructuring of iterative algorithms in an attempt to combine innerproducts [21, 5, 20]).

- Subspace methods for interior eigenvalues [49, 63], rational Lanczos [62, 63], harmonic Ritz values [28, 47, 55]. We will discuss some of these approaches in this paper.

The remainder of this paper has been organized as follows. We start with an introduction to Krylov subspace methods. Then we highlight the implicit restart technique for Arnoldi's method, as well as some shift-and-invert strategies that help to improve the speed of convergence. Special attention will be given to the Jacobi-Davidson method, which works with different subspaces. A novel extension for this method, which makes it possible to compute several eigenvalues in a part of the spectrum efficiently, will be discussed. Then we will pay some attention to the Look-ahead techniques for the two-sided Lanczos method.

Interior eigenvalues are always difficult to compute with subspace methods, if one wants to avoid expensive shift-and-invert operations. Approximate shift-and-invert operations have been suggested for the Jacobi-Davidson method [66], for the Lanczos method [50], and for the Arnoldi method [45]. The notion of harmonic Ritz values offers a helpful tool for restart purposes, since they identify the best approximations with respect to interior eigenvalues.

The concept of homotopy received attention as a means to compute some eigenvalues for 'difficult', or perturbed matrices, starting with available knowledge for a given matrix. We shall briefly examine such techniques. Also, we discuss the concept of ε-pseudospectrum, as a menas to study the sensitivity of a spectrum, for not too large matrices. Finally, we will conclude our paper by an outlook on some problems that are still hard to solve, as a motivation for further research.

3 Krylov subspaces

Krylov subspaces play a central role also in iterative methods for eigenvalue computations. To illustrate this, we consider the very well-known Power Method. Assume A is real symmetric, then it has real eigenvalues and a complete set of orthonormal eigenvectors

$$Au_k = \lambda_k u_k \quad , \quad \|u_k\|_2 = 1 \quad (k = 1, 2, \cdots, n).$$

We further assume that the largest eigenvalue in modulus is single and that

$$|\lambda_1| > |\lambda_2| \geq \cdots .$$

Now suppose we are given a vector v_1, which can be expressed in terms of the eigenvectors as $v_1 = \sum_i \gamma_i u_i$, and we assume that $\gamma_1 \neq 0$ (that means that v_1 has a nonzero component in the direction of the largest eigenvector).

Given this v_1, we compute $Av_1, A(Av_1), \ldots$, and it follows that

$$\lim_{j \to \infty} \frac{\|A^j v_1\|}{\|A^{j-1} v_1\|} = |\lambda_1|$$

It is not hard to see why the ratios of these norms approximate the dominant eigenvalue, since

$$\begin{aligned} A^j v_1 &= \sum \gamma_i \lambda_i^j u_i \\ &= \lambda_1^j \left\{ \gamma_1 u_1 + \sum_{i \geq 2} \gamma_i \left(\frac{\lambda_i}{\lambda_1} \right)^j u_i \right\} \quad . \end{aligned}$$

Hence

$$\begin{aligned} \frac{\|A^j v_1\|}{\|A^{j-1} v_1\|} &= |\lambda_1| \frac{\|\gamma_1 u_1 + \sum_{i \geq 2} \gamma_i \left(\frac{\lambda_i}{\lambda_1} \right)^j u_i\|}{\|\gamma_1 u_1 + \sum_{i \geq 2} \gamma_i \left(\frac{\lambda_i}{\lambda_1} \right)^{j-1} u_i\|} \\ &= |\lambda_1| \frac{\| \sum_{i \geq 1} \gamma_i P_j(\lambda_i) u_i\|}{\| \sum_{i \geq 1} \gamma_i P_{j-1}(\lambda_i) u_i\|} \quad , \end{aligned}$$

with $P_j(t) \equiv (t/\lambda_1)^j$.

With the Power method we have built a *Krylov Subspace*

$$K_m(A; v_1) \equiv \text{span}\{v_1, Av_1, \ldots, A^{m-1} v_1\},$$

but note that the method exploits only the last two vectors. The result is that the speed of convergence depends on how fast the polynomial values of $P_j(t)$ decrease for $t = \lambda_k$, for increasing j. The methods of Lanczos and Arnoldi exploit the whole Krylov subspace, and they implicitly construct polynomials P_j that may decrease much faster depending on the eigenvalue distribution.

3.1 Orthogonal basis (Arnoldi, Lanczos)

In order to identify better solutions in the Krylov subspace we need a suitable basis for this subspace, one that can be extended in a meaningful way for subspaces of increasing dimension. The obvious basis r_0, Ar_0, …, $A^{i-1}r_0$, for $K^i(A; r_0)$, is not very attractive from a numerical point of view, since the vectors $A^j r_0$ point more and more in the direction of the dominant eigenvector for increasing j (the power method !), and hence the basis vectors become dependent in finite precision arithmetic.

Lanczos [34] proposed to use an orthogonal basis for the Krylov subspace, and Arnoldi [2] suggested to compute this basis for unsymmetric matrices as follows. Start with $v_1 \equiv r_0/\|r_0\|_2$. Assume that we have already an orthonormal basis v_1, …, v_j for $K^j(A; r_0)$, then this basis is expanded by computing $\widetilde{v} = Av_j$, and by orthonormalizing this vector $\widetilde{v}$ with respect to v_1, …, v_j. In principle the orthonormalization process can be carried out in different ways, but the most commonly used approach is to do this by a modified Gram-Schmidt procedure [30]. This leads to the following algorithm for the creation of an orthonormal basis for $K^m(A; r_0)$:

$$
\begin{array}{l}
v_1 = r_0/\|r_0\|_2; \\
\text{for } j = 1, .., m-1 \\
\quad \widetilde{v} = Av_j; \\
\quad \text{for } i = 1, ..., j \\
\quad\quad h_{i,j} = v_i^* \widetilde{v}; \\
\quad\quad \widetilde{v} = \widetilde{v} - h_{i,j} v_i; \\
\quad \text{end}; \\
\quad h_{j+1,j} = \|\widetilde{v}\|_2; \\
\quad v_{j+1} = \widetilde{v}/h_{j+1,j}; \\
\text{end}
\end{array}
$$

A more stable implementation, useful for ill-conditioned matrices A, was suggested by Walker [75]; his approach was to use Householder transformations rather than modified Gram-Schmidt.

It is easily verified that v_1, …, v_m form an orthonormal basis for $K^m(A; r_0)$ (that is, if the construction does not terminate at a value $h_{j+1,j} = 0$). The orthogonalization leads to relations between the v_j, that can be formulated in a compact algebraic form. Let V_j denote the matrix with columns v_1 up to v_j, $V_j \equiv [v_1 \mid v_2 \mid \cdots \mid v_j]$, then it follows that

$$AV_{m-1} = V_m H_{m,m-1}. \tag{3.1}$$

The m by $m-1$ matrix $H_{m,m-1}$ is upper Hessenberg, and its elements $h_{i,j}$ are defined by the Arnoldi orthogonalization algorithm. From a computational point of view, this construction is composed from three basic elements: a matrix vector product with A, innerproducts, and updates. We see that this orthogonalization becomes increasingly expensive for increasing dimension of the subspace, since the computation of each $h_{i,j}$ requires an inner product and a vector update.

Note that if A is symmetric, then so is $H_{m-1,m-1} = V_{m-1}^* A V_{m-1}$, so that in this situation $H_{m-1,m-1}$ is tridiagonal. This means that in the orthogonalization process, each new vector has to be orthogonalized with respect to the previous two vectors only, since all other innerproducts vanish. The resulting three term recurrence relation for the basis vectors of $K_m(A; r_0)$ is known as the *Lanczos method* and some very elegant methods are derived from it. In the symmetric case the orthogonalization process involves constant arithmetical costs per iteration step: one matrix vector product, two innerproducts, and two vector updates. In [13, 74] it has been shown that the Lanczos algorithm can also be applied to matrices for which $A^* = -A$, with a skew symmetric tridiagonal matrix as a result.

3.2 Subspace Iteration Methods

The standard iteration method for eigenproblems is the power method. From it we can derive many well-known iterative methods, almost in a similar way as the iterative solvers for linear systems could be derived from the standard Richardson iteration method.

The main idea is to create some subspace and to compute approximations for the eigenvalues and eigenvectors with respect to this subspace. Starting from the Power method, the most natural choice for this subspace is the Krylov subspace, for which the standard basis vectors are generated by the Power method. A numerically more stable basis is created by generating an orthogonal basis straight away. This leads to the *Lanczos Method* for symmetric matrices, and to the *Arnoldi Method* for unsymmetric matrices.

The orthogonality relation for the basis vectors of $K^m(A; r_0)$ is:

$$AV_m = V_{m+1} H_{m+1,m}. \tag{3.2}$$

and we observe that

$$V_m^* A V_m = H_{m,m}. \tag{3.3}$$

The matrix $H_{m,m}$ is the projection of A onto $K^m(A; r_0)$. If (θ, y) is an eigenpair of $H_{m,m}$ then

$$H_{m,m} y = \theta y$$

$$V_m^* A V_m y - \theta V_m^* V_m y = 0$$

$$V_m^* (A V_m y - \theta V_m y) = 0$$

$$V_m^* (As - \theta s) = 0,$$

where $s = V_m y$.
Hence, the residual for the approximate eigenpair (θ, s) is orthogonal to the current Krylov subspace. The value θ is called a *Ritz value* of A with respect to $K^m(A; r_0)$, and s is the corresponding *Ritz vector*.

If A is unsymmetric then $H_{m,m}$ is an m by m upper Hessenberg matrix and the method is then known as the *Arnoldi Method*; when A is symmetric (or antisymmetric) then $H_{m,m}$ is tridiagonal and symmetric (or antisymmetric), and the method is then known as the *Lanczos Method*. In this case the orthogonality of a new basis vector is obtained with only two innerproducts.

In order to avoid the expensive construction of the upper Hessenberg matrix as in the Arnoldi method, we can follow a dual-basis approach for the construction of a basis that satisfies a three term construction. This leads to the *Bi-Lanczos Method*, which is very similar to the Bi-CG method. Both methods lead to the same projected tridiagonal matrix. Freund [27] and Cullum and Willoughby [16] have published codes for the computation of eigenvalues using this approach.

We can also construct other subspaces for the restriction of the matrix A. One such approach has been suggested by Davidson [18].

In actual situations subspace methods are often applied for $(A - \sigma I)^{-1}$, so that convergence towards eigenvalues close to σ is favored. This is called the *shift-and-invert* approach. This technique is attractive if one can compute the vector $(A - \sigma I)^{-1} z$ at relatively low costs. Note that with a direct solver this can be done with an LU decomposition of $A - \sigma I$, and this LU decomposition can be used for subsequent iteration steps.

For a good overview of subspace methods, see [65].

4 Improvements to Arnoldi's method

The main problem with Arnoldi's method is that it becomes increasingly expensive per iteration step, and in order to restrict memory storage, as well as computational work, restarts are necessary. However, at restart we throw away useful information. Instead of restarting with the most current approximation for the approximated eigenvector (the Ritz vector), corresponding to the desired eigenvalue, it has been suggested to restart with a vector that is a mix of Ritz vectors, corresponding to relevant Ritz values. See [65]:Ch.VII and [61] for such strategies. These strategies are related to an earlier approach suggested in [29]. These strategies may have the advantage that the new starting vector also contains information for nearby eigenvalues. The main problem is that in this kind of restart we try to catch the information for an approximate subspace in one single vector, and apart from this, it is not easy to find the optimal mix.

The restart problem has been solved very elegantly by Sorensen [68]. The idea behind his *Implicitly Restarted Arnoldi* (IRA) method is the following. Suppose we are at step $k+m$, and we want to shrink the current subspace to dimension k again. Obviously we want to maintain the subspace with the best k Ritz vectors, that is, the Ritz vectors corresponding to k Ritz values that we have selected, for instance the k rightmost Ritz values if we are after the k rightmost eigenvalues of A. We compute all $k+m$ Ritz values of H_{k+m} and then we carry out m shifted QR iterations with the m undesired Ritz values as shifts. This has the effect that the k wanted Ritz values are contained in the leading k by k part of the transformed matrix. Also the projection matrix V_{m+k} is transformed correspondingly to the transformations on H_{m+k}, and after the operation the first k columns of the transformed V_{m+k} span the subspace of the Ritz vectors for the wanted Ritz values. This subspace should be expanded again in order to improve the current eigenvalue approximations, and the nice observation is that the leading k by k part of the transformed Hessenberg matrix H_{m+k} is still upper Hessenberg. This means that the first k columns of the transformed V_{m+k} and the leading part H_k can be regarded as representing the first k Arnoldi steps on the transformed basis vector v_1. Sorensen [68] then proposes to continue this new Arnoldi process with another m steps, and to repeat the sketched subspace shrinking procedure. The whole procedure can be viewed as a mechanism that rakes repeatedly new supplementary information with m vectors from the entire space, to filter the desired supplements for the k approximated Ritz values, and to keep the gradually improved k-dimensional subspace in stock.

The attractive property of the IRA process is that we explore effectively subspaces of high-dimensional Krylov subspaces, without the high computational costs that would go with the unrestarted process. The process has been further refined by Lehoucq [37], who considers the many practical aspects involved in a careful implementation. In particular, an analysis and comparison of restarting an Arnoldi iteration, its numerical stability and deflation rules which allow computation of clustered and/or multiple eigenvalues is examined. In [39] existing software for this process, as well as for other methods, is evaluated.

In [68] it has been shown that the implicit restart technique effectively is equivalent with applying a polynomial filter on the starting vector, filtering out undesired eigenvetor components. In [23] it is shown how this polynomial filter, or the multishift QR, can be carried out without actually computing the shifts. The implicit-restart idea has been applied also in combination with the two-sided Lanczos method, for model reduction in electronical applications (pole-zero analysis, stability of CD-players) [31]. Similar ideas have recently been exploited to improve the Jacobi-Davidson method (see Section 6) and the rational Krylov method [64]..

The Arnoldi iteration is often carried out with the shift-and-invert approach. For instance, when solving the generalized eigenproblem $Ax = \lambda Bx$, the method is applied to the operator $(A-\sigma B)^{-1}B$. This transformation requires expensive operations with an inverted operator but the advantage is much faster convergence. Meerbergen [45] considers the use of inexact forms of the *Cayley transform* $(A-\sigma B)^{-1}(A-\tau B)$, where the inverse operation is approximated by a few steps of an iterative method, for Arnoldi's method. This technique has a close relation with polynomial preconditioning. Ruhe [63] considers a more general shift-and-invert transform, the so-called *Rational Krylov Sequence (RKS)*:

$$(\delta_j A - \gamma_j B)^{-1}(\sigma_j A - \rho_j B),$$

in which the coefficients may be different for each iteration step j. It has been shown that by generating a subspace with this operator, the given problem can be reduced to a small projected generalized system

$$(\zeta K_{j,j} - \eta L_{j,j})s = 0,$$

where $K_{j,j}$ and $L_{j,j}$ are upper Hessenberg matrices of dimension j. This small system may be solved by the QZ algorithm in order to obtain approximate values for an eigenpair. The parameters in RKS can be chosen to obtain faster convergence to interior eigenvalues. For a comparison of RKS and Arnoldi, see [63, 62].

5 Davidson's method and new variants

The main idea behind Davidson's method is the following one. Suppose we have some subspace K of dimension k, over which the projected matrix A has a Ritz value θ_k (e.g., θ_k is the largest Ritz value) and a corresponding Ritz vector u_k. Let us assume that an orthogonal basis for K is given by the vectors v_1, v_2, ..., v_k.

Now we want to find a successful update for u_k, in order to expand our subspace. To that end we compute the *defect:* $r = Au_k - \theta_k u_k$. Then Davidson, in his original paper [18], suggests to compute $\widetilde{v}$ from $(D_A - \theta_k I)\widetilde{v} = r$, where D_A is the diagonal of the matrix A. The vector $\widetilde{v}$ is made orthogonal to the basis vectors v_1, ..., v_k, and the resulting vector is chosen as the new v_{k+1}, by which K is expanded.

It has been reported that this method can be quite successful in finding dominant eigenvalues of (strongly) diagonally dominant matrices. The matrix $(D_A - \theta_k I)^{-1}$ can be viewed as a preconditioner for the vector r. Davidson [19] suggests that his algorithm (more precisely: the Davidson-Liu variant of it) may be interpreted as a Newton-Raphson scheme, and this has been used as an argument to explain its fast convergence. It is tempting to see the preconditioner also as an approximation for $(A - \theta_k I)^{-1}$, and, indeed, this approach has been followed for the construction of more complicated preconditioners (see, e.g., [14, 47, 50]). However, note that $(A - \theta_k I)^{-1}$ would map r onto u_k, and hence it would not lead to an expansion of our search space. Clearly this is a wrong interpretation for the preconditioner. Originally, the Davidson method had been proposed for symmetric matrices, but in many of the cited publications it has been mentioned that the method can be used successfully for some unsymmetric matrices as well.

5.1 The Jacobi-Davidson iteration method

In this section the matrix A may be unsymmetric and complex, and in order to express this we use the notation v^* for the complex conjugate of a vector (if complex), or the transpose (if real), and likewise for matrices.
For the construction of effective subspaces we observe that for a given approximate *Ritz pair* (θ, s), the residual is given by

$$r = As - \theta s.$$

Following an old and forgotten technique of Jacobi [33] (for strongly diagonally dominant matrices), it was suggested in [66] to compute a correction Δs for s in the subspace orthogonal to s, such that the residual vanishes in that subspace. That is, we want to solve

$$(I - ss^*)(A - \theta I)(I - ss^*)\Delta s = -r, \tag{5.1}$$

for $\Delta s \perp s$.
It can be shown that for $\theta = \lambda$ (an eigenvalue of A), this correction Δs leads immediately to the corresponding eigenvector $y = s + \Delta s$: $Ay = \lambda y$.

For the expansion of the subspace we solve equation (5.1), for a given Ritz value θ, and we expand the subspace with Δs. We compute a new Ritz pair with respect to the expanded subspace and we repeat the above procedure. This is the basis for the *Jacobi-Davidson Method.* The sketched procedure leads to quadratic convergence of the Ritz value to an eigenvalue if A is unsymmetric, and to cubic convergence if A is symmetric.

Of course, solving (5.1) may be an expensive affair, and for that reason we discuss what happens if this is equation is only solved approximately. This question has been addressed in [66], and has led to the following observations:

1. If we take the very crude approximation $\Delta s = -r$, then this method becomes equivalent with the Arnoldi method (and with Lanczos if A is symmetric).
2. If we approximate the projected operator by $D_A - \theta I$, and skip the condition that $\Delta s \perp s$ then we obtain Davidson's method. More recent suggestions made in [14, 47, 48, 50, 19] come down to better approximations for the inverse of $A - \theta_k I$, e.g., incomplete decompositions for this operator. However, as is well-known, this is a risky approach (see [65, 14]), since the exact inverse of this operator leads to failure of the method[1], and therefore the approximation should not be too accurate [65].

[1] Any progress in this case may be attributed to the effects of rounding errors

3. If we take full account of the restriction to the subspace orthogonal to s the we obtain the Jacobi - Davidson methods.

The algorithm for the improved Davidson method then becomes as follows (in the style of [65], in particular we have skipped indices for variables that overwrite old values in an iteration step, e.g., u instead of u_k).

1. **Start:**
   - Compute $v_1 = v/\|v\|$, $w_1 = Av_1$,
     set $V_1 = [\,v_1]$, $W_1 = [\,w_1]$, $H_1 = [\,h_{11}]$,
     $u = v_1$, $\theta = h_{11}$, compute $r = w_1 - \theta u$.
2. **Iterate:** Until convergence do:
3. **Inner Loop:** For $k = 1, ..., m$ do:
   - Solve (approximately)
     $(I - u\,u^*)\,(A - \theta I)\,(I - u\,u^*)\,\Delta s = -r$.
   - Orthogonalize Δs against V_k via Modified Gram-Schmidt, and expand V_k with this vector to V_{k+1}.
   - Compute $w_{k+1} := Av_{k+1}$ and expand W_k with this vector to W_{k+1}.
   - Compute $V_{k+1}^* w_{k+1}$, the last column of $H_{k+1} := V_{k+1}^* A V_{k+1}$, and $v_{k+1}^* W_k$, the last row of H_{k+1} (only if $A \neq A^*$).
   - Compute the largest eigenpair (θ, s) of H_{k+1} (with $\|s\| = 1$).
   - Compute the Ritz vector $u := V_{k+1} s$,
     compute $\widehat{u} := Au$ $(= W_{k+1} s)$, and
     the associated residual vector
   - Test for convergence. Stop if satisfied.
4. **Restart:** Set $v_1 := u$ and goto 3.

Now we will discuss convenient ways for the approximate solution of

$$(I - u_k u_k^*)(A - \theta_k I)(I - u_k u_k^*)\Delta s = -r \quad \text{and} \quad \Delta s \perp u_k. \tag{5.2}$$

Since $\Delta s \perp u_k$, it follows from (5.2) that

$$(A - \theta_k I)\Delta s - \varepsilon u_k = -r \tag{5.3}$$

or

$$(A - \theta_k I)\Delta s = \varepsilon u_k - r.$$

When we have a suitable preconditioner M, for which $M^{-1} \approx (A - \theta_k I)^{-1}$, then we can compute an approximation $\widetilde{\Delta s}$ for Δs:

$$\widetilde{\Delta s} = \varepsilon M^{-1} u_k - M^{-1} r. \tag{5.4}$$

The value of ε is determined by the requirement that $\widetilde{\Delta s}$ should be orthogonal with respect to u_k:

$$\varepsilon = \frac{u_k^* M^{-1} r}{u_k^* M^{-1} u_k}. \tag{5.5}$$

Equation (5.4) leads to several interesting observations:

1. If we choose $\varepsilon = 0$ then we obtain the Davidson method (with preconditioner M). In this case $\widetilde{\Delta s}$ will not be orthogonal to u_k.

2. If we choose ε as in (5.5) then we have a Jacobi-Davidson method. Note that this method requires two operations with the preconditioning matrix per iteration.

3. If $M = A - \theta_k I$, then (5.4) reduces to

$$\Delta s = \varepsilon (A - \theta_k I)^{-1} u_k - u_k.$$

   Since Δs is made orthogonal to u_k afterwards, this choice is equivalent with $\Delta s = (A - \theta_k I)^{-1} u_k$. In this case the method is mathematically equivalent with (accelerated) shift and invert iteration (with optimal shift).

If we solve (5.2) approximately with a preconditioned iterative method, like Bi-CGSTAB or GMRES, then we do not need two preconditioning operations per iteration step (as is necessary if we do only unaccelerated preconditioning), for details see [66].

Successful implementations largely depend on how well an effective preconditioner can be identified. Note that the operator $A - \theta I$ will be indefinite in general, so that one has to be careful with incomplete decomposition techniques. The operator restricted to the subspace orthogonal to the Ritz vector corresponding to θ, however, is not indefinite, and in [67] it is shown how available preconditioners for $A - \theta I$ can be restricted to that subspace.

6 A novel extension for the Jacobi-Davidson method

In some circumstances the Jacobi-Davidson method has apparent disadvantages with respect to Arnoldi's method. For instance, in many cases we see rapid convergence to one single eigenvalue, and what to do if we want more eigenvalues? For Arnoldi this is not a big problem, since the usually slower convergence towards a particular eigenvalue goes hand in hand with simultaneous convergence towards other eigenvalues. So after a number of steps Arnoldi produces approximations for several eigenvalues.
For Jacobi-Davidson the obvious approach would be to restart with a differently selected Ritz pair, with no guarantee that this leads to a new eigenpair. Also the detection of multiple eigenvalues is a problem, but this problem is shared with the other subspace methods.

A well-known way out of this problem is to use a technique, known as *deflation*. If an eigenvector has converged, then we continue in a subspace spanned by the remaining eigenvectors. A problem is then how to re-use information obtained in a previous Jacobi-Davidson cycle. In [26] an algorithm is proposed by which several eigenpairs can be computed. The algorithm is based on the computation of a partial Schur form of A:

$$AQ_k = Q_k R_k,$$

where Q_k is an $n \times k$ orthonormal matrix, and R_k is a $k \times k$ upper triangular matrix, with $k \ll n$. Note that if (x, λ) is an eigenpair of R_k, then $(Q_k x, \lambda)$ is an eigenpair of A.

We now proceed in the following way in order to obtain this partial Schur form for eigenvalues close to a target value τ.

Step I: Given an orthonormal subspace basis $v_1, \ldots, v_i$, with matrix V_i, compute the projected matrix $M = V_i^* A V_i$. For the $i \times i$ matrix M we compute the complete Schur form $MU = US$, with $U^*U = I$, and S upper triangular. This can be done with the standard QR algorithm [30].
Then we order S such that the $|s_{i,i} - \tau|$ form a nondecreasing row for increasing i. The first few diagonal elements of S then represent the eigenapproximations closest to τ, and the first few of the correspondingly reordered columns of V_i represent the subspace of best eigenvector approximations. If memory is limited then this subset can be used for restart, that is the other columns are simply discarded. The remaining subspace is expanded according to the Jacobi-Davidson method. After convergence of this procedure we have arrived at an eigenpair (q, λ) of A: $Aq = \lambda q$. The question is how to expand this partial Schur form of dimension 1. This will be shown in step II.

Step II: Suppose we have already a partial Schur form of dimension k, and we want to expand this by a convenient new column q:

$$A\,[Q_k, q] = [Q_k, q] \begin{bmatrix} R_k & s \\ & \lambda \end{bmatrix}$$

with $Q^* q = 0$.
After some standard linear algebra manipulations it follows that

$$(I - Q_k Q_k^*)(A - \lambda I)(I - Q_k Q_k^*)q = 0,$$

which expresses that the new pair (q, λ) is an eigenpair of

$$\widetilde{A} = (I - Q_k Q_k^*)A(I - Q_k Q_k^*).$$

This pair can be computed by applying the Jacobi-Davidson algorithm (with Schur form reduction, as in step I) for $\widetilde{A}$.

Some notes are appropriate:

1. Although we see that after each converged eigenpair the explicitly deflated matrix $\widetilde{A}$ leads to more expensive computations, it is shown in [26], by

numerical experiments, that the entire procedure leads to a very efficient computational process. An explanation for this is that after convergence of some eigenvectors, the matrix $\widetilde{A}$ will be better conditioned, so that the correction equation in the Jacobi-Davidson step is more easily solved.

2. The correction equation may be solved by a preconditioned iterative solver, and it is shown in [26] that the same preconditioner can be used with great efficiency for different eigenpairs. Hence, it pays to construct better preconditioners.

3. In [26] a similar algorithm for generalized eigenproblems $Ax = \lambda Bx$ is proposed, based on partial QZ reduction [30].

7 Bi-Lanczos and Look-ahead techniques

In the unsymmmetric Lanczos method [34] dual bases $\{r_j\}$ and $\{s_j\}$ are generated for the Krylov subspace $K^i(A; r_0)$ and its adjoint $K^i(A^*; s_0)$. The r_j are generated with a three term recurrence relation, with A:

$$\gamma_j r_{j+1} = Ar_j - \alpha_j r_j - \beta_j r_{j-1},$$

and the s_j with a similar recurrence for A^*:

$$\gamma_j s_{j+1} = A^* s_j - \alpha_j s_j - \beta_j s_{j-1}.$$

The constants γ_j, α_j, and β_j, are chosen so that $s_k^* r_i = 0$ for $k \neq i$, and $s_j^* r_j = 1$ (this requirement can not always be fulfilled). In algebraic form these recurrencies read as

$$AR_j = R_{j+1} T_{j+1,j},$$

and

$$A^* S_j = S_{j+1} T_{j+1,j}.$$

The matrix S_j is now used for the projection of the first equation:

$$S_j^* A R_j = T_{j,j}.$$

The eigenvalues of $T_{j,j}$ are taken as approximations for those of A, and the eigenvector approximations are taken as $R_j y_k$, where y_k is an eigenvector of $T_{j,j}$.

The breakdown occurs when $s_j^* r_j = 0$, and for numerical stability reasons one also wants to avoid the situation that

$$\frac{s_j^* r_j}{||s_j||_2 ||r_j||_2} \approx 0.$$

See [76]:Ch6.36 for a discussion on the failures in the Lanczos process. Despite the bad reputation of the standard unsymmetric Lanczos process, software produced by Cullum and Willoughby [16] has been quite successful. In their approach the inaccurate eigenapproximations, due to numerical instabilities, were identified from comparison of results for submatrices of $T_{j,j}$. The main idea is that the starting vector has nonzero components in directions of eigenvectors corresponding to eigenvectors that we are interested in. With this starting vector the Krylov subspaces are generated. When we compare the eigenvalues of $T_{j,j}$ with those of the tridiagonal matrix of order $j-1$ that is obtained if we skip the first row and column from $T_{j,j}$, then we compare Ritz values of the current Krylov subspace with Ritz values for a subspace from which the starting vector has been removed. Since this starting vector contains essential information of the desired eigenvector directions, no Ritz value can have converged *unless* all information for the corresponding eigenvector has entered through rounding errors. This kind of heuristics help to identify the so-called *spurious* eigenvalues.

In the early eighties Taylor [70] and Parlett et al [59], suggested to expand the Krylov subspace with sets of basis vectors that were block-dual. The idea is that certain dimensions, namely those for which breakdown occurs, can not be used for projection, and one has to postpone the inspection of the projected system until a block R_i of sufficient large dimension was discovered for which $S_i^* R_i$ is not too small. This process of delaying the actual projection step was called *Look-ahead*. In the period '85–'95 the look-ahead technique was generally accepted as a necessary element for the Lanczos method. Gutknecht gave a detailed theoretic basis for the look-ahead mechanism [32]. Freund and Nachtigal incorporated this technique in their QMR method [27], and it was shown later that their codes could also be used for effective solution of large sparse unsymmetric eigenproblems, see for instance [24] for an example of this.

Brezinski and co-workers, see e.g., [9], considered the breakdown problem from a polynomial point of view. The vectors in the Krylov subspaces can be seen as the results of matrix polynomials acting on the starting vectors, and likewise, the bi-orthogonality relations can be viewed as orthogonality relations for polynomials, with respect to a finite innerproduct (in which the weights are defined by the starting vectors). They showed that for some degrees the polynomials do not exist, and they also showed how higher degree orthogonal polynomials could be defined, by temporarily using polynomials of degree larger than 1 in the recursion formulas. This is equivalent with the block-wise expansion for the dual bases in the look-ahead approach.

8 Related issues

8.1 Approximations for interior eigenvalues

It is well-known that the subspace methods lead to eigenvalue approximations that tend to converge towards exterior eigenvalues. It may happen that an eigenvalue approximation is close to an interior eigenvalue of A, but in the next iteration step this needs not to be the case. In that situation we say that the eigenvalue approximation was on its way towards some exterior eigenvalue.

It is easy to reverse the direction of convergence of the eigenvalue approximations to the interior eigenvalues closest to the origin, by working with A^{-1}, but this is expensive. It is also possible to obtain eigenvalue approximations that converge (slowly) to the eigenvalues of A closest to the origin, from the subspaces generated with A. We will explain this for the Arnoldi process.

The Arnoldi proces leads to

$$AV_m = V_{m+1}H_{m+1,m},$$

with $H_{m+1,m}$ a upper Hessenberg matrix with $m+1$ rows and m columns. The upper m by m part of this matrix will be denoted as H_m.

With $H_{m+1,m}^* V_{m+1}^* = V_m^* A^*$ it follows that:

$$V_m^* A^* A V_m = H_{m+1,m}^* V_{m+1}^* V_{m+1} H_{m+1,m}$$

or

$$V_m^* A^* A V_m = H_{m+1,m}^* H_{m+1,m} \equiv M_m^2 .$$

From the equation

$$M_m^{-1} V_m^* A^* A V_m M_m^{-1} = I$$

it follows that the columns of $AV_m M_m^{-1}$ form a set of m orthonormal vectors.

Note that these vectors form an orthonormal basis for $AK_m(A; v_1)$. The matrix A^{-1} maps $AK_m(A; v_1)$ onto $K_m(A; v_1)$, and we may try to find a suitable form for the projection of the operator A^{-1}, that is the orthogonal restriction of A^{-1} with respect to $AK_m(A; v_1)$. To that end we use the orthonormal basis, and find that the projection can be given as:

$$M_m^{-1}(AV_m)^* A^{-1} AV_m M_m^{-1} = M_m^{-1}(AV_m)^* V_m M_m^{-1} = M_m^{-1} H_m^* M_m^{-1} .$$

So approximations for the eigenvalues of A^{-1} follow from

$$M_m^{-1} H_m^* M_i^{-1} t = \theta t,$$

or

$$M_m^{-2} H_m^* s = \theta s \, .$$

The vector s can be represented in terms of the basisvectors for $AK_m(A; v_1)$, and can then be viewed as an approximate eigenvector of A^{-1}:

$$y = AV_m s.$$

Since we are actually looking for eigenvalue approximations for A, we may wish to solve

$$H_m^{-*} M_m^2 s = \theta^{-1} s.$$

The idea behind this approach is the following. The Arnoldi projection process for A leads to approximations that tend to converge to exterior eigenvalues of A. Likewise, we might hope that the exterior eigenvalues of the projected A^{-1} converge (slowly) to the exterior eigenvalues of A^{-1}. Note that these happen to be the eigenvalues of A closest to the origin. Note that we then can force (slow) convergence to eigenvalues close to any point in the spectrum of A.

In [55] these eigenvalue approximations, in connection with the related Lanczos process, were called Harmonic Ritz values, and some nice relations for Harmonic Ritz values for symmetric indefinite matrices are given in that paper. Harmonic Ritz values had already been studied from a different viewpoint by other authors. Freund [28] has studied them as the zeros of the GMRES and MINRES iteration polynomials. Morgan [47] had observed that the Harmonic Ritz values and vectors are very suitable for restarting purposes if one wants to compute interior eigenvalues with subspaces of restricted dimension. In [66, 26] the Harmonic Ritz values are considered in connection with the Jacobi-Davidson process.

8.2 Sensitivity of Eigenproblems

Eigenvalues are used as a source of information on stability or convergence problems, and the question arises how valid the information of the mere values is. Many authors have studied the problem of sensitivity of the eigenvalues with respect to perturbations, see for instance [76, 69, 12]. These studies are usually related to perturbations caused by rounding errors, and not so much by the relevance of the eigenvalues due to the particular representation of a given problem, for instance the choice of basis.

Around 1987 Trefethen [72] started to emphasize this aspect of eigencomputations, and he propagated the idea of inspecting the pseudospectrum of a matrix as a relatively simple means for getting an idea of the significance of a particular part of the spectrum, without getting involved in complicated matters such as angles between eigenvectors or eigenspaces.
The definition of pseudospectrum $\Lambda_\varepsilon(A)$ for a matrix A is directly related to perturbations:

$$\Lambda_\varepsilon(A) \equiv \{z \in \mathbb{C} : z \in \Lambda(A+E) \text{ for some } E \text{ with } \|E\| \leq \varepsilon\}.$$

The pseudospectrum is usually graphically shown as a set of level curves for various values of ε. The level curves, or contour integrals, are more apparent from the original definition for the ε-pseudospectrum, in terms of the norm of the resolvent $(zI - A)^{-1}$:

$$\Lambda_\varepsilon(A) \equiv \{z \in \mathbb{C} : \|(zI - A)^{-1}\| \geq \varepsilon^{-1}\},$$

with the convention $\|(zI - A)^{-1}\| = \infty$ for $z \in \Lambda(A)$.

For symmetric matrices, the pseudospectrum of A is a collection of discs around the eigenvalues of A (note that the perturbation E needs not be symmetric). For unsymmetric matrices the pseudospectrum can be any collection of closed curves, containing the set of eigenvalues of A. These level curves may give information that is hidden by the information provided by the eigenvalues themselves. For instance, when studying stability of integration methods for systems of ODE's, or in bifurcation problems, the eigenvalues may be in a proper region, for instance, in the left-half plane, while the level curves even for rather small values of ε may intersect with the right-half plane. In such cases it may be time to ask further questions about the problem. On the other hand the pseudospectrum may not tell the full story. For instance, the sensitivity problems may be due to a single pair of ill-conditioned eigenvectors for which the more global level curves are too pessimistic. It may be the case that it is not realistic to assume equal perturbations for all matrix entries, but nevertheless the pseudospectrum points the attention to critical places in the spectrum. A nice introduction to the relevance of pseudospectra is given in [71], where for a number of matrices pseudospectra are actually computed and discussed.

Due to the nature of pseudospectra, this useful tool is often restricted to matrices of moderate size. This poses another problem: if we study the pseudospectrum for a discretized PDE with rather course meshsize (since we have to compute the smallest singular value of $zI - A$, for various values of z), what does this tell us on the effects to be expected for finer meshsizes?

We have carried out some experiments, not reported here, with a simple discretized convection-diffusion equation:

$$-u_{xx} - \beta u_x = f.$$

If we discretize this equation over the interval $[0, 1]$, with central difference approximation, then for large values of β and relatively large meshsize, we see that the spectrum is on a line parallel to the imaginary axis. If we decrease the meshsize, then the spectrum gradually shrinks to one single point on the real axis (a highly defective case), and then spreads along the real axis. It turns out that amazingly soon, after the spectrum starts to spread along the real axis, the pseudospectrum gives quite accurate information on the sensitivity of the given problem, information that is not essentially different from the information that we would have obtained for very fine meshsizes. On the other hand, the information for rather crude meshsizes can be quite misleading. This illustrates that pseudospectra do not necessarily represent more reliable information than the standard spectrum, when the information is obtained for lower-dimensional problems than the problems that are used in the actual large-scale scientific computations.

More recently, tools have become available for computing the pseudospectrum of large sparse matrices. Carpraux et al [10] propose an algorithm for computing the smallest singular value of $zI - A$, that is based on Davidson's method with ILU preconditioning. Lui [41] (see also [8]) suggests to use Lanczos' method in combination with continuation techniques. This is a plausible approach, since we need to do the computation for many values of z, well-distributed over the region of interest, in order to obtain a complete picture of the pseudospectrum. We know that currently such tools are being used for the analysis of instability problems of large sets of ODE's, related to climate modelling, but results have not yet been published.

8.3 Homotopy methods

The subspace methods that we have discussed before, are often applied in combination with shift and invert operations. That means that if one wants to have eigenvalues close to a value σ, then the methods are applied to the inverse $(A - \sigma I)^{-1}$ of the shifted matrix. As we have seen, the Jacobi-Davidson method can be interpreted as an inexact shift-invert method, since the invert step is usually approximated by a few steps of some convenient preconditioned inner iteration method.

Related to these inexact shift-invert approaches is the homotopy approach, that has received attention in the past five years. The idea is to compute some of the eigenvalues of a perturbed matrix $A + E$, when eigenvalues of A are known, or can be relatively easily computed. In order to this we use the homotopy $H(t) = A + tE$, $0 \le t \le 1$. If eigenpairs of $H(t_0)$ are known, then they are used as approximations for those of $H(t_0 + \Delta t)$. These approximations are improved by a convenient subspace iteration, for instance in [43] Rayleigh quotient iterations are used for symmetric A and E (see references in [43] for earlier work on homotopy for eigenproblems). For the Rayleigh quotient iteration one needs to solve systems like $(H(t_0 + \Delta t) - \lambda I)y = x$, where (λ, x) represents the current approximation for an eigenpair of $H(t_0 + \Delta t)$. In the context of large sparse matrices, it may be undesirable to do this with a direct solver, and in [43] the system is solved with SYMMLQ [56]. Of course, one could restrict oneself to only a few steps with SYMMLQ, and then try to accelerate the inexact Rayleigh quotient steps, as is done in the Jacobi-Davidson method. This indicates relations between these different approaches, but as far as we know, these relations have not yet been explored. There seem to be more relations. In [43] it is observed that SYMMLQ may find difficulty in converging for the nearly singular system $H(t_0 + \Delta t) - \lambda I)y = x$, and it is suggested to improve the situation by applying the Rayleigh quotient iteration to the approximately deflated matrix $H(t_0 + \Delta t) + xx^T$, where the term approximately deflated is used to indicate that x is only an approximation to the desired eigenvector. Note that similar deflation procedures are incorporated in the Jacobi-Davidson process.

The whole procedure is repeated for successive increments Δt, until the final value $t = 1$ is reached. In [43] an elegant approach is followed for the selection of the step size Δt.

The homotopy approach lends itself quite naturally for situations where the matrix A varies in time, or where it varies as a linearization of a nonlinear operator, as in bifurcation problems. Other situations that are obvious are for instance the Schrödinger eigenvalue problem [43]:

$$-\Delta u + fu = \lambda u$$

in the unit square in two space dimensions with homogeneous Dirichlet boundary condition. With the usual finite difference approximations on a uniform grid, this leads to the discrete Laplacian for $-\Delta u$, of which we know the eigensystem.

In [78] the homotopy approach is used for symmetric generalized eigenproblems, very much along the same lines as sketched above. The application for real unsymmetric eigenproblems is considered in [44].

8.4 Implementation aspects

In the past ten years sophisticated software has been produced for the computation of eigenvalues and eigenvectors. We have mentioned already the Implicitly Restarted Arnoldi method, for which Lehoucq et al have written the package ARPACK [40]. We also mention the code SRRIT [6] for the identification of a dominant invariant subspace of a nonsymmetric matrix. We have discussed the approaches by Cullum and Willougby, and by Freund and Nachtigal, but we also want to mention a recent approach to design robust software for the unsymmetric Lanczos process, the so-called ABLE package [4].

In [39] an overview of relevant software is given as well as numerical results for representative test problems. In [3] a project is announced for the production of software for relevant algorithm for large eigenproblems.

Most of the iterative methods reduce the given large problem to a much smaller problem. The smaller problem can then be handled by standard software for dense or banded matrices. Excellent software is available in LAPACK [1], the modern successor of the famous packages EISPACK and LINPACK.

Parallelism in the subspace methods is, except for the (approximate) shift-and-invert steps, usually no big problem. For some modern computers the required innerproducts may form a bottleneck with respect to scalable performance. For an overview of techniques to improve parallel behavior of algorithms for eigenproblems, see [21].

9 Some open problems

We have highlighted some of the progress that has been made in the past period. A novice in this area might easily have got the impression that most of the relevant problems, associated with the computation of eigenvalues and eigenvectors have been solved. Fortunately, this is not true. Like in most other areas of science there are many challenging problems that wait for solution and for further analysis.

We mention only a few of these open problems. Despite all our knowledge and experience with the Lanczos process, its simple basic three-term recurrence still contains some mysteries, when computing in finite precision arithmetic. We know that duplicate eigenvalues enter the process, due to rounding errors, and we see that they enter at more or less regular intervals during the iteration process, but this phenomenon is still not well understood.
For unsymmetric problems the picture is still rather obscure. Although the two-sided Lanczos process often leads to good approximations for some eigenvalues, its convergence behavior is not well understood. Even the convergence behavior of the Arnoldi method is not well understood, notwithstanding the fact that it can be viewed as an accelerated power iteration. In a recent paper, Cullum [15] shows that the convergence behavior of the two-sided Lanczos process, for eigenvalues, can be mimicked by Arnoldi's process and vice versa. For the unsymmetric eigenproblem we also face the problem that there are no efficient and reliable algorithms for computing the eigenvalues and eigenvectors of unsymmetric tridiagonal matrices, in contrast with the symmetric situation.

Although serious attempts have been undertaken for the computation of the Kronecker canonical form, by for instance Kågström and Van Dooren, this still needs much further research. Also the computation of invariant subspaces of highly nonnormal matrices is still in its infancy, notwithstanding useful contributions by, for instance, Chaitin-Chatelin et al [11, 7] and Lee [36]. It is also necessary that we get effcient tools for checking the condition of (partial) eigensystems, angles between invariant subspaces, etcetera. There is a need for efficient algorithms for the computation of a (partial) SVD for very large sparse matrices.

Several special types of eigenproblems require special algorithms, for instance large sparse polynomial eigenproblems, of the form as has been mentioned in our Introduction. Such problems occur in acoustics and in structural engineering.

For very large sparse matrices we need efficient approximate inverses to replace shift-and-invert operations. This leads to the problem of approximating the inverse of (highly) indefinite matrices, and these inverses should be effective in the eigendirections close to the one associated with the shift. Currently available approximation techniques are not very good in this respect, most often they do not work at all for interior eigenvalues.
Recently, it has been shown by Lui [42] that domain decomposition techniques can be used for the computation of eigenvalues of partial differential operators. This is certainly a promising direction in view of parallel computation.

Some problems lead to eigenproblems with constraints, for instance Stokes problems, but also semi-indefinite optimization problems. For general cases, the required tools are missing at the moment.

Finally we mention the well-known QR iteration for eigenproblems, which still needs attention when it comes to parallel implementation.

We hope that the reader is convinced that there are still many open problems in this exciting field and these problems involve interesting work to be done by theorists as well as more practical oriented computational scientists.

Acknowledgements

We would like to thank Rich Lehoucq, Karl Meerbergen, Nick Trefethen and Gerard Sleijpen, for specific comments that helped to improve our presentation.
The work of Gene H. Golub was in part supported by the National Science Foundation, Grant Number DMS-9403899.
The work of Henk van der Vorst was in part supported by the Dutch Research Organization NWO, MPR cluster project 95MPR04.

Bibliography

1. E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. DuCroz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. So ren sen. *LAPACK User's Guide.* SIAM, Philadelphia, PA, 1992.

2. W. E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenproblem. *Quart. Appl. Math.*, 9:17–29, 1951.

3. Z. Bai, D. Day, J. Demmel, J. Dongarra, M. Gu, A. Ruhe, and H. van der Vorst. Templates for linear algebra problems. In J. van Leeuwen, editor, *Computer Science Today.* Springer Verlag, Berlin, 1995.

4. Z. Bai, D. Day, and Q. Ye. ABLE: An adaptive block Lanczos method for non-Hermitian eigenvalue problems. Technical Report Research Report 95-04, University of Kentucky, Lexington, KY, 1995.

5. Z. Bai, D. Hu, and L. Reichel. A Newton basis GMRES implementation. *IMA J. Numer. Anal.*, 14:563–581, 1991.

6. Z. Bai and G.W. Stewart. SRRIT - A FORTRAN subroutine to calculate the dominant invariant subspace of a nonsymmetric matrix. Technical Report 2908, Department of Computer Science, University of Maryland, 1992.

7. T. Braconnier, F. Chatelin, and V. Frayssé. The influence of large nonnormality on the quality of convergence of iterative methods in linear algebra. Technical Report TR-PA-94/07, CERFACS, Toulouse, 1994.

8. T. Braconnier and N. J. Higham. Computing the field of values and pseudspectra using the Lanczos method with continuation. Technical Report 279, Manchester centre for computational mathematics, Manchester, UK, 1995.

9. C. Brezinski and M. Redivo-Zaglia. Treatment of near breakdown in the CGS algorithm. *Numerical Algorithms*, 1994.

10. J. F. Carpraux, J. Erhel, and M. Sadkane. Spectral portrait for non-Hermitian large sparse matrices. *Computing*, 53:301–310, 1994.

11. F. Chaitin-Chatelin. Is nonnormality a serious difficulty? Technical Report TR-PA-94/18, CERFACS, Toulouse, 1994.

12. F. Chatelin. *Valeurs propres de matrices.* Collection mathématiques appliquées pour la maîtrise. Masson, 1988.

13. P. Concus and G. H. Golub. A generalized Conjugate Gradient method for nonsymmetric systems of linear equations. Technical Report STAN-CS-76-535, Stanford University, Stanford, CA, 1976.

14. M. Crouzeix, B. Philippe, and M. Sadkane. The Davidson method. *SIAM J. Sci. Comp.*, 15:62–76, 1994.

15. J. Cullum. Arnoldi versus nonsymmetric Lanczos algorithms for solving matrix eigenvalue problems. *BIT*, 1996, to appear.

16. J. Cullum and R.A. Willoughby. A practical procedure for computing eigenvalues of large sparse nonsymmetric matrices. In J. Cullum and R.A. Willoughby, editors, *Large Scale Eigenvalue Problems*, pages 193–240, Amsterdam, 1986. North-Holland.

17. J. J. M. Cuppen. A divide and conquer method for the symmetric eigenproblem. *Numerische Mathematik*, 36:177–195, 1981.

18. E.R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real symmetric matrices. *J. Comp. Phys.*, 17:87–94, 1975.

19. E.R. Davidson. Monster matrices: their eigenvalues and eigenvectors. *Computers in Physics*, 7:519–522, 1993.

20. E. De Sturler and H.A. van der Vorst. Reducing the effect of global communication in GMRES(m) and CG on parallel distributed memory computers. *J. Appl. Num. Math.*, 18:441–459, 1995.

21. J. Demmel, M. Heath, and H. van der Vorst. Parallel numerical linear algebra. In *Acta Numerica 1993.* Cambridge University Press, Cambridge, 1993.

22. J. J. Dongarra and D. C. Sorensen. A fully parallel algorithm for the symmetric eigenvalue problem. *SIAM J. Scientific and Statistical Computing*, 8:139–154, 1987.

23. A. A. Dubrulle and G. H. Golub. A Multishift QR iteration without computation of the shifts. *Numerical Algorithms*, 7:173–181, 1994.

24. P. Feldmann and R.W. Freund. A linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Trans. Computer-Aided Design*, 14, 1995.

25. D. A. Flanders and G. Shortley. Numerical determination of fundamental modes. *J. Appl. Phys.*, 21:1322–1328, 1950.

26. D.R. Fokkema, G.L.G. Sleijpen, and H.A. van der Vorst. Jacobi-Davidson style QR and QZ algorithms for the partial reduction of matrix pencils. Technical Report Preprint 941, Mathematical Institute, Utrecht University, 1996.

27. R. W. Freund and N. M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Num. Math.*, 60:315–339, 1991.

28. R.W. Freund. Quasi-kernel polynomials and their use in non-Hermitian matrix iterations. *J. Comp. Appl. Math.*, 43:135–158, 1992.

29. G. H. Golub and R. Underwood. The block Lanczos method for computing eigenvalues. In J.R. Rice, editor, *Mathematical Software III*, pages 361–377. Academic Press, New York, 1977.

30. G. H. Golub and C. F. van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, 1989.

31. E. J. Grimme, D. C. Sorensen, and P. Van Dooren. Model reduction of state space systems via an Implicitly Restarted Lanczos method. Technical Report Report TR94-21, Dept. of Comp. and Appl. Math., Rice University, Houston, TX, 1994.

32. M. H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related algorithms, Part I. *SIAM J. Matrix Anal. Appl.*, 13:594–639, 1992.

33. C.G.J. Jacobi. Ueber ein leichtes Verfahren, die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen. *Journal für die reine und angewandte Mathematik*, pages 51–94, 1846.

34. C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl. Bur. Stand*, 45:225–280, 1950.

35. C. Lanczos. *Applied Analysis.* Prentice Hall, Englewood Cliffs, NJ, 1956.

36. S. L. Lee. A practical upper bound for departure from normality. *SIAM J. Matrix Anal. Appl.*, 16:462–468, 1995.

37. R. B. Lehoucq. *Analysis and Implementation of an Implicitly Restarted Iteration.* PhD thesis, Rice University, Houston, TX, 1995.

38. R. B. Lehoucq. Restarting an Arnoldi reduction. Technical Report Preprint MCS-P591-0496, Math. and Comp. Science Div., Argonne National Laboratory, Argonne, IL 60439, 1996.

39. R. B. Lehoucq and J. A. Scott. An evaluation of software for computing eigenvalues of sparse nonsymmetric matrices. Technical Report Preprint MCS-P547-1195, Math. and Comp. Science Div., Argonne National Laboratory, Argonne, IL 60439, 1996.

40. R.B. Lehoucq, D.C. Sorensen, P. Vu,a nd C. Yang. ARPACK: An implementation of the Imlicitly Re-started Arnoldi Iteration that computes some of the eigenvalues and eigenvectors of a large sparse matrix. Available from `netlib@ornl.gov` under the directory scalapack, 1996.

41. S. H. Lui. Computation of pseudospectra by continuation. *SIAM J. Scient. Comput.*, 1996.

42. S. H. Lui. Domain decomposition methods for eigenvalue problems. Technical report, Hong Kong University of Science and Technology, Hong Kong, 1996.

43. S. H. Lui and G. H. Golub. Homotopy method for the numerical solution of the eigenvalue problem of self-adjoint partial differential operators. *Numerical Algorithms*, 10:363–378, 1995.

44. S. H. Lui, H. B. Keller, and W. C. Kwok. Homotopy method for the large sparse real nonsymmetric eigenvalue problem. Technical Report, Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong, 1996.

45. K. Meerbergen. *Robust methods for the calculation of rightmost eigenvalues of nonsymmetric eigenvalue problems.* PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, 1996.

46. K. Meerbergen and D. Roose. Matrix transformations for computing rightmost eigenvalues of real nonsymmetric matrices. *IMA J. Numer. Anal.*, 16:297–346, 1996.

47. R.B. Morgan. Computing interior eigenvalues of large matrices. *Lin. Alg. and its Appl.*, 154–156:289–309, 1991.

48. R.B. Morgan. Generalizations of Davidson's method for computing eigenvalues of large nonsymmetric matrices. *J. Comp. Phys.*, 101:287–291, 1992.

49. R.B. Morgan and D.S. Scott. Generalizations of Davidson's method for computing eigenvalues of sparse symmetric matrices. *SIAM J. Sci. Stat. Comput.*, 7(3):817–825, 1986.

50. R.B. Morgan and D.S. Scott. Preconditioning the Lanczos algorithm for sparse symmetric eigenvalue problems. *SIAM J. Sci. Comput.*, 14:585–593, 1993.

51. J. Olsen, P. Jørgensen, and J. Simons. Passing the one-billion limit in full configuration-interaction (FCI) calculations. *Chemical Physics Letters*, 169:463–472, 1990.

52. A.M. Ostrowski. On the convergence of the Rayleigh quotient iteration for the computation of characteristic roots and vectors. V. *Arch. Rational. Mesh. Anal.*, 3:472–481, 1959.

53. C. C. Paige. Computational variants of the Lanczos method for the eigenproblem. *J. Inst. Math. Appl.*, 10:373–381, 1972.

54. C. C. Paige. Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix. *J. Inst. Math. Appl.*, 18:341–349, 1976.

55. C. C. Paige, B. N. Parlett, and H. A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Num. Lin. Alg with Appl.*, 2(2):115–134, 1995.

56. C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12:617–629, 1975.

57. B. N. Parlett. *The Symmetric Eigenvalue Problem.* Prentice-Hall, Englewood Cliffs, N.J., 1980.

58. B. N. Parlett. Misconvergence in the Lanczos algorithm. Technical Report PAM-404, University of California, Berkeley, CA, 1987.

59. B. N. Parlett, D. R. Taylor, and Z. A. Liu. A look-ahead Lanczos algorithm for unsymmetric matrices. *Math. Comp.*, 44:105–124, 1985.

60. B.N. Parlett. The Rayleigh Quotient iteration and some generalizations. *Math. Comp.*, 28:679–693, 1974.

61. S. G. Petiton. Parallel subspace method for non-Hermitian eigenproblems on the Connection Machine (CM2). *Appl. Num. Math.*, 10:19–36, 1992.

62. A. Ruhe. The rational Krylov algorithm for nonsymmetric eigenvalue problems. iii: Complex shifts for real matrices. *BIT*, 34:165–176, 1994.

63. A. Ruhe. Rational Krylov algorithms for nonsymmetric eigenvalue problems. ii. matrix pairs. *Lin. Alg. and its Appl.*, 197, 198:283–295, 1994.

64. A. Ruhe. Rational Krylov, a practical algorithm for large sparse nonsymmetric matrix pencils. Technical Report UCB/CSD-95-871, Computer Science Division, University of California, Berkeley, CA, 1995.

65. Y. Saad. *Numerical methods for large eigenvalue problems.* Manchester University Press, Manchester, UK, 1992.

66. G. L. G. Sleijpen and H.A. Van der Vorst. A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 17:401–425, 1996.

67. G.L.G. Sleijpen, J.G.L. Booten, D.R. Fokkema, and H.A. van der Vorst. Jacobi-Davidson type methods for generalized eigenproblems and polynomial eigenproblems: Part I. Technical Report 923, University Utrecht, Department of Mathematics, 1995.

68. D. C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Mat. Anal. Appl.*, 13(1):357–385, 1992.

69. G. W. Stewart and Ji-Guang Sun. *Matrix Perturbation Theory.* Academic Press, San Diego, CA, 1990.

70. D.R. Taylor. *Analysis of the look ahead Lanczos algorithm.* PhD thesis, University of California, Berkeley, CA, 1982.

71. L. N. Trefethen. Pseudospectra of matrices. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis 1991*, pages 234–266. Longman, 1992.

72. L. N. Trefethen and M. R. Trummer. An instability phenomenon in spectral methods. *SIAM J. Numer. Anal.*, 24:1008–1023, 1987.

73. A. van der Sluis and H. A. van der Vorst. The convergence behavior of Ritz values in the presence of close eigenvalues. *Lin. Alg. and its Appl.*, 88/89:651–694, 1987.

74. H. A. van der Vorst. A generalized Lanczos scheme. *Math. Comp.*, 39(160):559–561, 1982.

75. H. F. Walker. Implementation of the GMRES method using Householder transformations. *SIAM J. Sci. Stat. Comp.*, 9:152–163, 1988.

76. J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.

77. J.H. Wilkinson and C. Reinsch, editors. *Handbook for Automatic Computation, Vol 2, Linear Algebra.* Springer Verlag, Heidelberg – Berlin – New York, 1971.

78. T. Zhang, K. H. Law, and G. H. Golub. On the homotopy method for symmetric modified generalized eigenvalue problems. Technical Report, Stanford University, Stanford, CA, 1996.