# Several concepts to investigate strongly nonnormal eigenvalue problems

Jos L.M. van Dorsselaer[1]

December 2000

**Abstract.** Eigenvalue analysis plays an important role in understanding physical phenomena. However, if one deals with strongly nonnormal matrices or operators, the eigenvalues alone may not tell the full story. A popular tool which can be useful to get more insight in the reliability or sensitivity of eigenvalues is $\varepsilon$-pseudospectra. Apart from $\varepsilon$-pseudospectra we consider other tools which might help to learn more about the eigenvalue problem, viz. condition numbers of the eigenvalues, condition numbers of sets of eigenvectors and angles between invariant subspaces. All these concepts will be studied and compared for both standard and generalized eigenvalue problems. The tools can be used to analyze large eigenvalue problems. We apply the different concepts to a generalized eigenvalue problem obtained from magnetohydrodynamics. In this problem one is interested in an interior part of the spectrum, called the Alfvén spectrum.

**Key words:** eigenvalues, eigenvalue problem, generalized eigenvalue problem, nonnormality, $\varepsilon$-pseudospectra, condition numbers, invariant subspaces, magnetohydrodynamics, Alfvén spectrum.

**AMS subject classification:** 15A18, 15A22, 15A60, 65F15, 65F35.

## 1 Introduction

Eigenvalues play an important role in many applications. In order to analyze (the behaviour of) a matrix or operator one often computes some eigenvalues and eigenvectors and make predictions based on this information. If the matrix or operator is normal (i.e. the eigenvectors form an orthonormal basis) this approach is reliable and a study of the eigenvalues can be used safely to analyze the problem. However, in many applications the matrices or operators are not normal, and an analysis based on eigenvalues only can be misleading. For example, let us consider the eigenvalue problem

$$Ax = \lambda x \,, \tag{1.1}$$

where $A$ is a square matrix. If the real parts of the eigenvalues of $A$ are negative, the vector $e^{tA}y$ converges to the zero vector as $t \to \infty$, but nevertheless the entries of $e^{tA}y$ can become arbitrarily large for finite $t$, which may lead to instabilities. A similar phenomenon occurs, e.g.,

[1]Mathematical Institute, Utrecht University, P.O. Box 80.010, NL-3508 TA Utrecht, The Netherlands (dorssela@math.uu.nl). Part of this research has been carried out at CWI, Amsterdam, The Netherlands. This research has been supported by the Dutch Organization for Scientific Research (NWO), project 95MPR04.

in problems from fluid dynamics, see, e.g., [27, 29]; this behaviour may show why an analysis based on eigenvalues only cannot explain observations obtained from laboratory experiments.

In this paper we will study and compare several tools which may be helpful for understanding the behaviour of matrices. One of the most popular tools used nowadays is $\varepsilon$-pseudospectra, studied extensively by Trefethen and others since 1990 (see, e.g., [26]). The $\varepsilon$-pseudospectra of $A$ indicate how the eigenvalues of $A$ may change if $A$ is perturbed by a matrix $E$ of which the norm is at most $\varepsilon$. Knowledge about the $\varepsilon$-pseudospectra may help, e.g., to understand the behaviour of $\mathrm{e}^{tA}y$ better, as well as other processes involving $A$. Moreover, $\varepsilon$-pseudospectra may also give some insight in the accuracy of numerically computed eigenvalues. However in applications it may happen that one is only interested in a few eigenvalues and eigenvectors of $A$. In such a situation it may be useful to consider $\varepsilon$-pseudospectra of $A$ restricted to the space spanned by these eigenvectors. For example, if $y$ is in such a subspace, the $\varepsilon$-pseudospectra restricted to that space may give better estimates for $\mathrm{e}^{tA}y$ than the $\varepsilon$-pseudospectra of $A$ (see Section 3.3). Moreover, the computation of $\varepsilon$-pseudospectra may be very time consuming if $A$ is a large matrix and the computation of the $\varepsilon$-pseudospectra of $A$ restricted to a small subspace is much cheaper in that case. In this paper we study both $\varepsilon$-pseudospectra of $A$ as well as $\varepsilon$-pseudospectra restricted to so-called invariant subspaces, and we also compare these sets.

Apart from $\varepsilon$-pseudospectra we also consider various condition numbers related to $A$. The condition number of an eigenvalue, introduced by Wilkinson [30] in 1965, measures how much this eigenvalue may change under small perturbations of $A$. It is obvious that there is a relation between these condition numbers and $\varepsilon$-pseudospectra, and we will discuss this in more detail. If the condition number of an eigenvalue is very large, it may be difficult or even impossible to compute such an eigenvalue accurately by numerical methods. This is another reason to be suspicious about the use of eigenvalues in such situations. Apart from the condition numbers of eigenvalues, we also study the condition number of a set of eigenvectors. This condition number can be used to investigate whether the eigenvectors from such a set make small angles with each other. Furthermore it can be used as well to estimate $\mathrm{e}^{tA}y$ for $y$ in the subspace spanned by those eigenvectors. The condition number of a set of eigenvectors does not give any information about the relation to other eigenvectors. In order to investigate whether two subspaces spanned by eigenvectors are close to each other or not, we also compute the angle between these subspaces.

Apart from the eigenvalue problem (1.1) also generalized eigenvalue problems of the form

$$Ax = \lambda Bx \tag{1.2}$$

occur in applications; here $A$ and $B$ are two square matrices. The matrix $B$ may be singular, and in that case the (mathematical) properties of (1.1) and (1.2) may be different in nature. The concepts mentioned above for the standard eigenvalue problem can be generalized to (1.2), and we will study these generalizations in this paper as well. If $B$ is nonsingular, (1.2) is equivalent to the standard eigenvalue problem

$$B^{-1}Ax = \lambda x \ . \tag{1.3}$$

Although the eigenvalues and eigenvectors of (1.2) and (1.3) are the same, this does *not* hold for the $\varepsilon$-pseudospectra and the condition numbers of the eigenvalues. The difference can be large and we will analyze this also in the paper. In our application from magnetohydrodynamics (see Section 5) we indeed will see that the difference is large. The formulation (1.3) has the advantage that one can use the theory related to (1.1), which is more developed than the theory of matrix pencils (1.2). The condition number of a set of eigenvectors and the angle between two subspaces spanned by eigenvectors are the same for both formulations (1.2) and (1.3), because these numbers depend only on the underlying eigenvectors.

We will also comment on the computational work and implementation aspects of the different tools studied in this paper. The matrices $A$ and $B$ are assumed to be large, so the computation of eigenvalues, eigenvectors etc. is not trivial. The determination of $\varepsilon$-pseudospectra is the most expensive task, and we consider this issue in Section 3.4. The computation of condition numbers and angles between subspaces will be discussed in the sections where they are introduced. If some eigenvalues and the corresponding eigenvectors are known, the condition numbers of a set of eigenvectors and angles between subspaces can be computed relatively cheaply (compared to the computation of eigenvalues and eigenvectors). In order to compute the condition number of an eigenvalue, one also has to determine the corresponding left eigenvector as well (left eigenvectors are defined in Section 2).

All concepts discussed in this paper are applied to a problem in magnetohydrodynamics, taken from [14]; this problem is of the form (1.2) with a nonsingular matrix $B$. As in [14] we consider only a part of the eigenvalues, the so-called Alfvén spectrum. It turns out that both the $\varepsilon$-pseudospectra and the condition numbers of the eigenvalues depend strongly on the choice of the formulation of the eigenvalue problem ((1.2) or (1.3)). In fact the $\varepsilon$-pseudospectra related to (1.2) are so large that even perturbations of order $10^{-12}$ on the matrices $A$ and $B$ may change the spectrum completely. Perturbations of the same order have a less dramatic effect on $B^{-1}A$. A similar behaviour is observed for the condition numbers of the eigenvalues. The $\varepsilon$-pseudospectra restricted to invariant subspaces are much smaller than the $\varepsilon$-pseudospectra for the whole problem in this application, and the same holds for the condition numbers of the eigenvectors from these subspaces.

The paper is organized as follows. In Section 2 we introduce some notation, definitions and present some facts from (numerical) linear algebra. The computation of a few eigenvalues and eigenvectors is also treated in that section. $\varepsilon$-Pseudospectra is the topic of Section 3: the $\varepsilon$-pseudospectrum of a matrix $A$ is defined in Section 3.1, and $\varepsilon$-pseudospectra related to generalized eigenvalue problems (1.2) is treated in Section 3.2. Restrictions of $\varepsilon$-pseudospectra to invariant subspaces is discussed in Section 3.3, and Section 3.4 deals with the computation of $\varepsilon$-pseudospectra. Condition numbers of eigenvalues and sets of eigenvectors are treated in Sections 4.1 and 4.2, respectively. The angles between invariant subspaces are considered in Section 4.3. The application of all these concepts can be found in Section 5: a formulation of the problem is given in Section 5.1, Section 5.2 deals with the $\varepsilon$-pseudospectra of the problem, and the other tools are considered in Section 5.3. The main conclusions of the paper can be found in Section 6.

## 2 Preliminaries and notation

In this section we present some theory from (numerical) linear algebra and introduce some relevant notation.

The matrices $A$ and $B$ are $N \times N$ matrices with complex entries and the pair $(A, B)$ is called a *matrix pencil.* For singular matrices $B$ the (mathematical) properties of the eigenvalue problem (1.2) may differ significantly from (1.1). For example, if there is a vector $x \neq 0$ such that $Ax = Bx = 0$ then (1.2) is satisfied for any $\lambda \in \mathbb{C}$. In that situation it does not make sense to study (the sensitivity of) eigenvalues of the matrix pencil and in this paper we assume that the determinant of $A - zB$ as a function of $z$ is not identically zero, so that (1.2) has at most $N$ eigenvalues. In order to understand (1.2) better for singular $B$ it is useful to write the eigenvalue problem as follows (cf., e.g., [24]):

$$\beta A x \;=\; \alpha B x \;, \tag{2.1}$$

with $\alpha, \beta \in \mathbb{C}$. The pair $\langle \alpha, \beta \rangle$ is called an eigenvalue of this pencil, and $\lambda = \alpha/\beta$ if $\beta \neq 0$. The pair $\alpha$ and $\beta$ in (2.1) is not unique. If $B$ is singular then $\langle \alpha, \beta \rangle = \langle 1, 0 \rangle$ is an eigenvalue of (2.1) which corresponds to the infinite eigenvalue of (1.2) ($\lambda = \infty$). In (2.1) finite and infinite eigenvalues are treated equally, and this formulation has also the advantage that it does not discriminate between $A$ and $B$.

The identity matrix is denoted by $I$. For an arbitrary (possibly rectangular) matrix $C$ we denote its transpose by $C^{\mathrm{T}}$ and its (Hermitian) adjoint by $C^*$; these matrices are the same if all entries of $C$ are real. The eigenvalues of (1.1) or (1.2) are denoted by $\lambda_j$, and the eigenvectors by $x_j$ (it will be clear from the context to which eigenvalue problem we refer); the set of all eigenvalues of (1.1) or (1.2) is called the *spectrum* of $A$ or $(A, B)$. Apart from the eigenvectors $x_j$ we also consider the *left eigenvectors* $y_j \neq 0$ given by

$$y_j^* A \;=\; y_j^* \lambda_j \tag{2.2}$$

corresponding to the standard eigenvalue problem (1.1), or

$$y_j^* A \;=\; y_j^* B \lambda_j \tag{2.3}$$

(related to (1.2)). In connection with left eigenvectors, the vectors $x_j$ are sometimes called right eigenvectors. If we take the transpose of the equations (2.2) and (2.3) we see that $\overline{y_j}$ (the complex conjugate of $y_j$) is the right eigenvector corresponding to $\lambda_j$ of $A^{\mathrm{T}}$ or $(A^{\mathrm{T}}, B^{\mathrm{T}})$, respectively.

In the following we have $u \in \mathbb{C}^\ell$ and $C$ is a (possibly rectangular) matrix with $\ell$ columns, so that $Cu$ exists. The (Euclidean) norm of $u$ is denoted by $\|u\|$ $(= \sqrt{u^* u})$ and

$$\|C\| \;=\; \max\{\|Cu\| : u \in \mathbb{C}^\ell \text{ and } \|u\| = 1\}$$

is the (*spectral*) norm of $C$. At a few places in the paper we consider other norms as well. Other numbers associated to matrices are the singular values (see, e.g., [13, Section 2]): these numbers are real and nonnegative and we denote the smallest and largest singular value of $C$ by $\sigma_{\min}(C)$ and $\sigma_{\max}(C)$, respectively. One has $\sigma_{\max}(C) = \|C\|$ and for square matrices $\sigma_{\min}(C)$ satisfies the following important equality:

$$\sigma_{\min}(C) \;=\; \min\{\|Cu\| : u \in \mathbb{C}^\ell \text{ and } \|u\| = 1\} \;. \tag{2.4}$$

For a square nonsingular matrix $C$ we define the *condition number* of $C$ as

$$\kappa(C) \;=\; \|C\| \cdot \|C^{-1}\| \quad (= \sigma_{\max}(C)/\sigma_{\min}(C)\ ) \;. \tag{2.5}$$

In the remainder of this paper $Q_k$ and $Z_k$ are $N \times k$ matrices of which the columns are orthonormal (i.e. $Q_k^* Q_k = Z_k^* Z_k = I_k$ (the $k \times k$ identity matrix)), and $R_k$, $S_k$ and $T_k$ are $k \times k$ matrices. If either

$$A Q_k \;=\; Q_k R_k \tag{2.6}$$

or

$$AQ_k = Z_k S_k \quad \text{and} \quad BQ_k = Z_k T_k \tag{2.7}$$

are satisfied, then span$\{Q_k\}$ is called an *invariant subspace* of $A$ (if (2.6) is satisfied) or the matrix pencil $(A, B)$ (if (2.7) is satisfied), respectively. The canonical example of an invariant subspace is a space spanned by some eigenvectors, so invariant subspaces can be seen as a generalization of eigenvectors. In the literature (cf., e.g., [10]) the characterizations (2.6) and (2.7) are sometimes called *partial Schur forms* if the matrices $R_k$, $S_k$ and $T_k$ are upper triangular. One can always determine an orthonormal basis $Q_k$ of an invariant subspace in such a way that the matrices $R_k$ or $S_k$ and $T_k$ are upper triangular.

We conclude this section with mentioning some numerical methods for computing (a few) eigenvalues and eigenvectors of (1.1) or (1.2). If $N$ is small, the QR method (see, e.g., [13]) can be used to compute the eigenvalues of (1.1), and for the generalized eigenvalue problem (1.2) the QZ method (see, e.g., [13]) may be applied to compute the spectrum of (1.2). A nice overview of eigenvalue methods for large matrices, with guidelines how to choose a method which may satisfy the needs of the user and pointers to software, can be found in [1]. In our experiments described in Section 5 we have computed some eigenvalues and eigenvectors of (1.2) with the Jacobi-Davidson QZ (JDQZ) method which has been developed by Fokkema, Sleijpen and Van der Vorst [10]. We briefly discuss some ingredients of this method. The user has to prescribe a target $\tau \in \mathbb{C}$, and the JDQZ method tries to compute a partial generalized Schur form (2.7) with $S_k$ and $T_k$ upper triangular in such a way that the eigenvalues of $(S_k, T_k)$ are the $k$ eigenvalues of $(A, B)$ which are as close to $\tau$ as possible. (The JDQZ method also allows the possibility to compute, e.g., (2.7) so that the eigenvalues of $(S_k, T_k)$ are likely to be those of $(A, B)$ with largest absolute values or largest real parts.) The corresponding eigenvectors of $(A, B)$ can then be computed from the eigenvectors of the small pencil $(S_k, T_k)$. The most expensive part of one step of the JDQZ method is to solve approximately the so-called correction equation, which involves the matrix $\widetilde{\beta}A - \widetilde{\alpha}B$ where $\langle \widetilde{\alpha}, \widetilde{\beta} \rangle$ is an approximation to the eigenvalue $\langle \alpha, \beta \rangle$ from (2.1). This is usually done with an iterative method such as, e.g., GMRES [21], and in order to speed up the convergence of this iterative solver one should use a proper preconditioner: in fact, a preconditioner for the matrix $\widetilde{\beta}A - \widetilde{\alpha}B$ is needed, and we will use a $LU$-factorization of $A - \tau B$ as a preconditioner which seems useful if $\tau \approx \widetilde{\alpha}/\widetilde{\beta}$. The reader may consult `http://www.math.uu.nl/people/sleijpen/` if he or she is interested in obtaining software for the JDQZ method.

# 3 Pseudospectra

## 3.1 An introduction to pseudospectra

The concept of $\varepsilon$-pseudospectra has been studied extensively by L.N. Trefethen and others in the last decade. With the convention $\|(zI - A)^{-1}\| = \infty$ whenever $z$ is an eigenvalue of $A$ the set $\Lambda_\varepsilon(A)$, the $\varepsilon$-*pseudospectrum* of $A$, is defined as follows:

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : \|(zI - A)^{-1}\| \geq \varepsilon^{-1}\} \ . \tag{3.1}$$

It is easy to see that this definition is equivalent to

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : \sigma_{\min}(zI - A) \leq \varepsilon\} \ . \tag{3.2}$$

Another important characterization of $\Lambda_\varepsilon(A)$ is (see, e.g., [17])

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : \exists\ E \text{ with } \|E\| \le \varepsilon \text{ such that } z \text{ is an eigenvalue of } A + E\} \ . \tag{3.3}$$

From (3.3) we see that $\varepsilon$-pseudospectra reveals how sensitive eigenvalues may be under perturbations of the matrix $A$. If the set $\Lambda_\varepsilon(A)$ is large for very small $\varepsilon$, it may be difficult to compute the eigenvalues accurately. $\varepsilon$-Pseudospectra may be used to investigate the accuracy of the approximate eigenvalues: suppose that $\widetilde{\lambda}$ and $\widetilde{x}$ are approximations to $\lambda$ and $x$, respectively, and $\|\widetilde{x}\| = 1$. The norm of the residual $\varepsilon_1 = \|A\widetilde{x} - \widetilde{\lambda}\widetilde{x}\|$ should be small and from (2.4) and (3.2) we get $\widetilde{\lambda} \in \Lambda_{\varepsilon_1}(A)$. Let $\Gamma$ be the closed curve in the complex plane such that $\sigma_{\min}(zI - A) = \varepsilon_1$ for all $z \in \Gamma$ and $\widetilde{\lambda}$ is in the interior of the region surrounded by $\Gamma$. It follows from the maximum principle (see, e.g., [8, Section III.14]) that there must be at least one eigenvalue $\lambda$ of $A$ in that region, and $\max\{|\widetilde{\lambda} - z| \,:\, z \in \Gamma\}$ is an upper bound for $|\widetilde{\lambda} - \lambda|$.

For normal matrices A, i.e. matrices for which the eigenvectors are an orthonormal basis of $\mathbb{C}^N$, the set $\Lambda_\varepsilon(A)$ is the union of the disks with radius $\varepsilon$ centered at the eigenvalues. For nonnormal matrices however, the set $\Lambda_\varepsilon(A)$ can be much larger; see the examples in, e.g., [26, 27] and in Section 5 of the present paper. For such matrices it may be useful to consider $\varepsilon$-pseudospectra instead of the eigenvalues (spectrum). It may not only be difficult to compute the eigenvalues of $A$ numerically, but the *true* eigenvalues may give misleading information about processes involving $A$. Examples of this include the stability of numerical methods for linear ordinary differential equations [7, 17, 18], the convergence of iterative methods for solving linear systems $Ax = b$, applications in fluid dynamics [27, 29] and magnetohydrodynamics [3]; see [28] for more references.

As an illustration we consider the following situation. Assume that the real parts of the eigenvalues of $A$ are negative. Then the matrices $\mathrm{e}^{tA}$ are bounded for $t \ge 0$ and $\mathrm{e}^{tA}$ converges to the zero matrix as $t \to \infty$. However, $\|\mathrm{e}^{tA}\|$ may become large (cf., e.g., [9, Theorem 11$\varepsilon$]):

$$\max_{z \in \Lambda_\varepsilon(A)} \operatorname{Re} z \ > \ C\varepsilon \qquad \Rightarrow \qquad \max_{t \ge 0} \|\mathrm{e}^{tA}\| \ > \ C \ . \tag{3.4}$$

The largest possible constant $C \ge 1$ in the left-hand side of (3.4) cannot be determined from the eigenvalues of $A$. Suppose $f$ is a smooth nonlinear function and that one deals with an ordinary differential equation $y' = f(y)$ which has a stationary solution $y_s$, i.e. $f(y_s) = 0$, and $A$ is the Jacobian matrix of $f$ evaluated at $y_s$. In order to investigate whether $y_s$ is a stable equilibrium one often linearizes the differential equation around $y_s$ and studies the solution of this linearized equation equipped with an initial value $y_s + \Delta y$. The deviation of the solution to that linearized equation from $y_s$ is equal to $\mathrm{e}^{tA}\Delta y$, and $\|\mathrm{e}^{tA}\Delta y\|$ can become very large according to (3.4); hence the equilibrium $y_s$ may actually behave unstable in practice.

One also may define $\varepsilon$-pseudospectra subject to other norms than the spectral norm; this allows the possibility to use, e.g., norms related to an energy norm, or the maximum norm. The equivalence of (3.1) and (3.3), as well as some other properties of $\varepsilon$-pseudospectra, such as, e.g., (3.4), remain valid for norms $\|\cdot\|$ derived from arbitrary vector norms; cf. [7, Section 4].

### 3.2 Pseudospectra for matrix pencils

In applications eigenvalue problems for matrix pencils occur, so it may be useful to study $\varepsilon$-pseudospectra for matrix pencils as well. This has been done by various authors, and several different definitions have appeared in the literature, cf., e.g., [3, 6, 11, 19, 20]. In the first papers on this topic [3, 19] one considers the case where the matrix $B$ is Hermitian and positive definite: with $B = F^*F$ the eigenvalue problem (1.2) is equivalent to $(F^{-1})^*AF^{-1}y = \lambda y$ with $y = Fx$, and the authors define the $\varepsilon$-pseudospectrum related to (1.2) as $\Lambda_\varepsilon((F^{-1})^*AF^{-1})$, which is equal to the $\varepsilon$-pseudospectrum of $B^{-1}A$ with respect to the norm $\|x\|_B := \sqrt{x^*Bx}$. In [20] the author defines the $\varepsilon$-pseudospectrum of the pencil $(A, B)$ as $\Lambda_\varepsilon(B^{-1}A)$. Another generalization has been considered in [6]; in loc. cit. the $\varepsilon$-pseudospectrum of the pencil $(A, B)$ was defined by replacing the identity matrix in (3.1) (or (3.2)) by the matrix $B$. Using this definition, the $\varepsilon$-pseudospectrum is equal to the union of all $z \in \mathbb{C}$ for which there exists a matrix $E$ with $\|E\| \le \varepsilon$ such that $A + E - zB$ is singular (cf. (3.3)). Hence this definition deals only with perturbations of the matrix $A$ (and not of $B$), which may seem unnatural. Perturbations of both $A$ and $B$ were taken into account in [11]; in that paper the $\varepsilon$-pseudospectrum of the pencil $(A, B)$ is defined as the union of the eigenvalues of the pencils $(A + E, B + F)$ where $\|E\| \le c_A\varepsilon$ and $\|F\| \le c_B\varepsilon$ ($c_A$ and $c_B$ are positive constants). The reason for introducing the constants $c_A$ and $c_B$ is to include the possibility to use perturbations relative to the norms of $A$ and $B$: note that for $c \ne 0$ the eigenvalues of $(cA, cB)$ are equal to those of $(A, B)$, but this does not hold in general for the perturbed pencils $(cA + E, cB + F)$ and $(A + E, B + F)$; for $|c| \gg 1$ the effect of perturbations on $(cA, cB)$ with matrices of a fixed norm may be much smaller that for $(A, B)$. The definition in [11] is equivalent to (see [11])

$$\{z \in \mathbb{C} : \sigma_{\min}(zB - A) \le \varepsilon\,(c_A + c_B|z|)\}\ . \tag{3.5}$$

From (3.5) we observe that perturbations of the matrix $B$ do not change the $\varepsilon$-pseudospectra much in the neighbourhood of the origin. In our application from magnetohydrodynamics (see Section 5) we are interested in $\varepsilon$-pseudospectra in the neighbourhood of the origin. For that reason and for the sake of simplicity we take $c_A = 1$ and $c_B = 0$ in (3.5) and define $\Lambda_\varepsilon(A, B)$, the *$\varepsilon$-pseudospectrum* of the pencil $(A, B)$, as

$$\Lambda_\varepsilon(A, B) = \{z \in \mathbb{C} : \sigma_{\min}(zB - A) \le \varepsilon\}\ . \tag{3.6}$$

Once the set $\Lambda_\varepsilon(A, B)$ has been determined numerically it is easy to compute the sets (3.5) for given $c_A$ and $c_B$; moreover, any theoretical result regarding $\Lambda_\varepsilon(A, B)$ can be adapted in such a way that it fits in the form (3.5). The $\varepsilon$-pseudospectra of the pencil $(A, B)$ can also be used to investigate the accuracy of numerically computed eigenvalues: one has to replace the matrix $I$ by $B$ in the discussion on this topic in Section 3.1.

One of the aspects of $\varepsilon$-pseudospectra we will consider in this paper is the relation between $\Lambda_\varepsilon(A, B)$ and $\Lambda_\varepsilon(B^{-1}A)$ (the latter set is considered in [20]). Working with the sets $\Lambda_\varepsilon(B^{-1}A)$ has the advantage that one can use the theory and insight of $\varepsilon$-pseudospectra for matrices; the relevance and application of $\varepsilon$-pseudospectra for matrices is much better understood than for matrix pencils. Computational issues will be considered in Section 3.4.

One can easily show that (see [6])

$$\Lambda_{\varepsilon/\|B\|}(B^{-1}A)\ \subset\ \Lambda_\varepsilon(A, B)\ \subset\ \Lambda_{\varepsilon\|B^{-1}\|}(B^{-1}A)\ , \tag{3.7}$$

which implies that there is not much difference between $\Lambda_\varepsilon(B^{-1}A)$ and $\Lambda_\varepsilon(A, B)$ if both $\|B\|$ and $\|B^{-1}\|$ are close to 1. However, the difference can be large as the following example shows: let $B$ be a multiple of the identity matrix. Then $\Lambda_\varepsilon(A, B) = \Lambda_{\varepsilon/\|B\|}(B^{-1}A)$ (which means that the first inclusion of (3.7) is sharp), so the difference between $\Lambda_\varepsilon(B^{-1}A)$ and $\Lambda_\varepsilon(A, B)$

can be very big if one chooses $\|B\|$ close to zero or very large. Also in our application from magnetohydrodynamics (see Section 5) we observe a big difference between $\Lambda_\varepsilon(B^{-1}A)$ and $\Lambda_\varepsilon(A,B)$.

We now briefly discuss the situation when the matrix $B$ is singular. This will happen in some applications from fluid dynamics, e.g., in problems which are derived from incompressible Navier-Stokes equations. The set $\Lambda_\varepsilon(A,B)$ is always defined and not empty for $\varepsilon > 0$ (unless $B$ is the zero matrix), and $\Lambda_\varepsilon(A,B) = \mathbb{C}$ for $\varepsilon \geq \min\{\,\|Ax\| : x \in \mathbb{C}^N$ with $\|x\| = 1$ and $Bx = 0\}$; see [6] for the proofs of these statements and more illustrations for singular $B$.

### 3.3 Pseudospectra restricted to invariant subspaces

In [25] it was suggested that one might approximate the set $\Lambda_\varepsilon(A)$ for large matrices $A$ by $\Lambda_\varepsilon(R_k)$, with $R_k$ as in (2.6). It is obvious that one may save a lot of computation time doing this. Of course, this approach only makes sense if the $\varepsilon$-pseudospectra of $A$ projected on the invariant subspace do not depend on the choice of $Q_k$: let $\widetilde{Q}_k$ be another $N \times k$ matrix of which the columns form an orthonormal basis of $\operatorname{span}\{Q_k\}$ and define $\widetilde{R}_k = \widetilde{Q}_k^* A \widetilde{Q}_k$. One can easily verify that $\Lambda_\varepsilon(R_k) = \Lambda_\varepsilon(\widetilde{R}_k)$ for all $\varepsilon \geq 0$, so the choice of the basis for $\operatorname{span}\{Q_k\}$ does not influence the $\varepsilon$-pseudospectra of the matrix $A$ projected on $\operatorname{span}\{Q_k\}$. In a similar way one may restrict the $\varepsilon$-pseudospectra of the matrix pencil $(A,B)$ to an invariant subspace, using (2.7): the set $\Lambda_\varepsilon(S_k,T_k)$ does not depend on the choices of the bases of $\operatorname{span}\{Q_k\}$ and $\operatorname{span}\{Z_k\}$. The same holds for the set $\Lambda_\varepsilon(T_k^{-1}S_k)$ which might be used as an approximation for $\Lambda_\varepsilon(B^{-1}A)$: note that (2.7) implies $B^{-1}AQ_k = Q_kT_k^{-1}S_k$. The following theorem, which can be proved by using (2.4), shows that the $\varepsilon$-pseudospectra restricted to invariant subspaces are nested.

**Theorem 3.1** *Let $Q_j$ be a $N \times j$ matrix such that the columns of $Q_j$ form an orthonormal basis of* $\operatorname{span}\{Q_j\}$ $(j = k, k+1)$ *and assume that* $\operatorname{span}\{Q_k\} \subset \operatorname{span}\{Q_{k+1}\} \subset \mathbb{C}^N$. *If both* $\operatorname{span}\{Q_k\}$ *and* $\operatorname{span}\{Q_{k+1}\}$ *are invariant subspaces of $A$, i.e. (2.6) is satisfied for both $k$ and $k+1$, the inclusions*

$$\Lambda_\varepsilon(R_k) \;\subset\; \Lambda_\varepsilon(R_{k+1}) \;\subset\; \Lambda_\varepsilon(A)$$

*hold for all $\varepsilon \geq 0$. If both* $\operatorname{span}\{Q_k\}$ *and* $\operatorname{span}\{Q_{k+1}\}$ *are invariant subspaces of the matrix pencil $(A,B)$, i.e. (2.7) is satisfied for both $k$ and $k+1$, the inclusions*

$$\Lambda_\varepsilon(S_k,T_k) \;\subset\; \Lambda_\varepsilon(S_{k+1},T_{k+1}) \;\subset\; \Lambda_\varepsilon(A,B)$$

*hold for all $\varepsilon \geq 0$.*

Another reason for dealing with $\varepsilon$-pseudospectra restricted to invariant subspaces is to study the effect of perturbations related to that subspace. As an illustration we consider the following result [7, Theorem 4.8], which can be seen as the converse of (3.4).

**Theorem 3.2** *Let $C \geq 1$ be a given constant and assume that $A$ is an $N \times N$ matrix such that $\Lambda_\varepsilon(A) \subset \{z \in \mathbb{C} : \operatorname{Re} z \leq C\varepsilon\}$ holds for all $\varepsilon \geq 0$. Then*

$$\|\mathrm{e}^{tA}\| \;\leq\; C\mathrm{e}N \quad \textit{for all} \quad t \geq 0\ .$$

In actual applications (cf. Section 3.1) one might only be interested in upper bounds $\|e^{tA}\Delta y\|$ with $\Delta y$ in a certain invariant subspace of $A$; the question is whether it is possible to obtain a sharper upper bound for $\|e^{tA}\Delta y\|$ in that case, e.g., an upper bound without the factor $N$ which can be large in practice[2]. This is indeed possible, as we will see below. First we present a lemma involving an arbitrary function $\Psi$ (instead of the exponential function), which may be useful for other applications as well.

**Lemma 3.3** *Let $\Psi : \mathbb{C} \to \mathbb{C}$ be analytic in a neighbourhood of the spectrum of $A$, and let $R_k$ and $Q_k$ be defined by (2.6). Then the matrices $\Psi(A)$ and $\Psi(R_k)$ exist and*

$$\|\Psi(A)y\| \;=\; \|\Psi(R_k)Q_k^* y\| \quad \textit{for} \quad y \in \operatorname{span}\{Q_k\}\ .$$

*Proof.* Let $I_k$ be the $k\times k$ identity matrix. From (2.6) it follows that $(\zeta I - A)Q_k = Q_k(\zeta I_k - R_k)$ holds for all $\zeta \in \mathbb{C}$. Assume that $y \in \operatorname{span}\{Q_k\}$ and $\zeta$ is not an eigenvalue of $A$. Then $y = Q_k Q_k^* y = Q_k(\zeta I_k - R_k)(\zeta I_k - R_k)^{-1}Q_k^* y = (\zeta I - A)Q_k(\zeta I_k - R_k)^{-1}Q_k^* y$ so that

$$(\zeta I - A)^{-1} y \;=\; Q_k(\zeta I_k - R_k)^{-1} Q_k^* y\ .$$

From the matrix version of the Cauchy integral formula (see, e.g., [8, Section VII.3]) one obtains

$$\begin{aligned}
\Psi(A)y &= \frac{1}{2\pi i}\int_\Gamma \Psi(\zeta)(\zeta I - A)^{-1} y \,\mathrm{d}\zeta \\
&= \frac{1}{2\pi i}\int_\Gamma \Psi(\zeta) Q_k(\zeta I_k - R_k)^{-1} Q_k^* y \,\mathrm{d}\zeta \\
&= Q_k \Psi(R_k) Q_k^* y\ ,
\end{aligned}$$

where $\Gamma$ is a closed contour in the complex plane surrounding the eigenvalues of $A$. Hence

$$\|\Psi(A)y\| \;=\; \|Q_k\Psi(R_k)Q_k^* y\| \;=\; \|\Psi(R_k)Q_k^* y\|\ ,$$

which proves the lemma.

□

Lemma 3.3 with $\Psi(\zeta) = e^{t\zeta}$ can be used to estimate $\|e^{tA}y\|$ for $y \in \operatorname{span}\{Q_k\}$. Observe that Lemma 3.3 implies $\|e^{tA}y\| \le \|e^{tR_k}\| \cdot \|y\|$, and Theorem 3.2 provides a bound for $\|e^{tR_k}\|$. This leads to the following result.

**Theorem 3.4** *Let $C_k \ge 1$, and assume that $\Lambda_\varepsilon(R_k) \subset \{z \in \mathbb{C} : \operatorname{Re} z \le C_k\varepsilon\}$ holds for all $\varepsilon \ge 0$ with $R_k$ and $Q_k$ as in (2.6). Then*

$$\|e^{tA}y\| \;\le\; C_k\, e k \|y\| \quad \textit{for} \quad y \in \operatorname{span}\{Q_k\} \quad \textit{and} \quad t \ge 0\ . \tag{3.8}$$

From Theorem 3.1 it follows that the smallest possible $C_k$ in Theorem 3.4 cannot be larger than the smallest possible $C$ in Theorem 3.2. Hence Theorem 3.4 gives a much sharper estimate than Theorem 3.2 for $k \ll N$. This result shows that it may be worthwhile to study $\varepsilon$-pseudospectra restricted to invariant subspaces.

[2] If the assumption in Theorem 3.1 holds with $C = 1$ one has $\|e^{tA}\| \le 1$ for $t \ge 0$ (see, e.g., [7]).

### 3.4 Computation of pseudospectra

Trefethen has written a nice review paper on the computation of $\varepsilon$-pseudospectra [28]; here we will discuss a few aspects of this and refer to [28] and the references cited therein for more information and details.

For small $N$ the standard approach is to compute $\sigma_{\min}(zI-A)$ or $\sigma_{\min}(zB-A)$ for several $z$ in an interesting region in the complex plane, and draw the level curves of this function — usually $z$ is taken from a rectangular grid which contains (part of) the spectrum of the underlying eigenvalue problem. In order to get a picture of reasonable quality, one needs to use about 50 gridpoints or more in each direction. If $B$ is nonsingular the matrix $B^{-1}A$ can be computed explicitly and $\Lambda_\varepsilon(B^{-1}A)$ can be determined in a similar way.

For large sparse matrices $A$ and $B$ this approach is not practical and one has to proceed differently. Several authors have studied question; see [28] for an overview. Some papers deal with the reduction of the number of gridpoints $z$ for which $\sigma_{\min}(zI-A)$ needs to be computed. If this is done in a proper way, one may reduce the computation time significantly without destroying the quality of the picture. Now we will discuss the computation of the smallest singular value of $zI-A$, $zB-A$ or $zI-B^{-1}A$ for a given $z$. In the literature two types of techniques have been considered (cf. [28]): the first approach is to restrict the matrices $A$ and $B$ to a low-dimensional subspace and approximate the $\varepsilon$-pseudospectra by the $\varepsilon$-pseudospectra of the projected problem related to this low-dimensional subspace. The second approach is to compute the smallest singular value for each $z$ with an appropriate eigenvalue method.

#### 3.4.1 Restriction to a low-dimensional subspace

We now discuss the first approach, which has already been mentioned in Section 3.3. In that section we explain how invariant subspaces can be used to approximate the different types of $\varepsilon$-pseudospectra. However it is not clear what the size of these subspaces should be to obtain a reasonable approximation at least in the neighbourhood of the eigenvalues of the projected system (one might not expect to get a good approximation far away from these eigenvalues, in particular not in the neighbourhood of the eigenvalues of the original problem which are not eigenvalues of the projected problem). See, e.g., [25] and Chapter 5 for illustrations. It is also possible to approximate $\varepsilon$-pseudospectra by using subspaces that are not invariant. In [25, 31] the authors use variants of Arnoldi's method to obtain the identity $AV_k = V_{k+1}H_{k+1,k}$; here $V_j$ is an $N \times j$ matrix of which the columns are an orthonormal basis of a Krylov subspace with respect to $A$ and $H_{k+1,k}$ is an $(k+1)\times k$ Hessenberg matrix. In these papers $\Lambda_\varepsilon(A)$ is approximated by $\Lambda_\varepsilon(H_{k+1,k})$ (although $H_{k+1,k}$ is not a square matrix one can still define $\Lambda_\varepsilon(H_{k+1,k})$: see [25, 31]) or $\Lambda_\varepsilon(H_{k,k})$, where $H_{k,k}$ is the matrix obtained by removing the last row of $H_{k+1,k}$. These ideas seem to be useful: in [31] the authors have approximated $\Lambda_\varepsilon(A)$ for a matrix of order $N = 200,000$ using the implicitly restarted Arnoldi method. The main computational work for such large matrices lies in the determination of a suitable Hessenberg matrix $H_{k+1,k}$.

For generalized eigenvalue problems one might use, e.g., rational Krylov iterations to obtain $(k+1)\times k$ Hessenberg matrices $H_{k+1,k}$ and $K_{k+1,k}$ satisfying the relation $AV_{k+1}H_{k+1,k} = BV_{k+1}K_{k+1,k}$; the columns of the matrix $V_{k+1}$ are again orthonormal. These Hessenberg matrices may be used to approximate $\Lambda_\varepsilon(A,B)$ and $\Lambda_\varepsilon(B^{-1}A)$: see [20] for details.

The main disadvantage of projecting the original problem onto low-dimensional subspaces it that it is not known how large $k$ should be taken to obtain reasonable approximations of the $\varepsilon$-pseudospectra. In practice one might take, e.g., $k$ between 20 and 50.

### 3.4.2 Computing the singular values with sparse eigenvalue methods

Another approach to approximate $\varepsilon$-pseudospectra is the following: $\lambda_{\min}(C(z))$, the smallest eigenvalue of the Hermitian matrix $C(z) = (zB-A)^*(zB-A)$ is the square of $\sigma_{\min}(zB-A)$ and several authors (see, e.g., [4, 5, 11, 16]) have computed $\lambda_{\min}(C(z))$ with, e.g., the inverse power method, Lanczos method or the (Jacobi-)Davidson method (see, e.g., [1] for a description of these methods). Although most papers deal with the case $B = I$, the generalization to arbitrary $B$ is straightforward. In order to apply these methods one needs a starting vector, and it was suggested by Lui [16] to order the gridpoints $z$ in such a way that the new gridpoint $z_{\text{new}}$ is close to the previous one $z_{\text{old}}$. The eigenvector corresponding to $\lambda_{\min}(C(z_{\text{old}}))$ may be a good starting vector for the computation of $\lambda_{\min}(C(z_{\text{new}}))$. Methods like the Jacobi-Davidson method [22] can be started with a subspace instead of one vector; (unpublished) experiments show that it is more efficient to start the iteration for $\lambda_{\min}(C(z_{\text{new}}))$ with the last subspace used to compute $\lambda_{\min}(C(z_{\text{old}}))$ instead of the (computed) eigenvector corresponding to $\lambda_{\min}(C(z_{\text{old}}))$.

Another observation which may be useful in practice is the following: if $\kappa(C(z))$ is in the order of the reciprocal of the machine precision or larger, it may be impossible to compute $\lambda_{\min}(C(z))$ accurately, due to rounding errors. In that situation the quantity $\|(zB - A)v\|$ with $v$ a normalized eigenvector of $C(z)$ corresponding to the computed eigenvalue $\lambda_{\min}(C(z))$ might be a more accurate approximation to $\sigma_{\min}(zB - A)$ (cf. (2.4)) than the computed quantity $\sqrt{\lambda_{\min}(C(z))}$.

In general it may very time-consuming or even impossible to compute $\Lambda_\varepsilon(B^{-1}A)$ with this technique. In order to apply the methods mentioned above one has to be able to determine $B^{-1}x$ for arbitrary vectors $x$ very accurately. Even if this is possible, it will in general be much more expensive than a matrix vector multiplication with $A$ or $B$. In our experiments (see Section 5) it was possible to determine an $LU$-factorization of $zB - A$ and we used this to apply the inverse power method for computing $\Lambda_\varepsilon(A, B)$. By noting that $(zI - B^{-1}A)^{-1} = (zB - A)^{-1}B$ one can also use the $LU$-factorization of $zB - A$ to compute $\Lambda_\varepsilon(B^{-1}A)$ with the inverse power method: so both $\sigma_{\min}(zB - A)$ and $\sigma_{\min}(zI - B^{-1}A)$ can be computed simultaneously with the inverse power method at the cost of one $LU$-factorization.

# 4 Condition numbers and invariant subspaces

## 4.1 Condition numbers of eigenvalues

The condition number of an eigenvalue of a matrix $A$ was introduced by Wilkinson [30, Section 2.8] and is a well known concept in numerical linear algebra. Assume that the eigenvalue $\lambda_j$ of $A$ is simple, and consider for small $t \geq 0$ the eigenvalue problem

$$(A + tE)\, x_j(t) \;=\; \lambda_j(t)\, x_j(t) \;; \tag{4.1}$$

here $E$ is an $N \times N$ matrix, $\lambda_j(0) = \lambda_j$ and $x_j(0) = x_j$. The function $\lambda_j(t)$ determines how much the eigenvalue $\lambda_j$ may change due to perturbations of $A$ with the matrix $tE$. It is obvious that there is a strong connection with $\varepsilon$-pseudospectra: cf. Definition (3.3). The identity (4.1) is differentiable in $t = 0$ and $\lambda_j(t) \approx \lambda_j + \lambda_j'(0)t$ for small $t$; note that

$\lambda_j'(0) = (y_j^* E x_j)/(y_j^* x_j)$ where $y_j$ is the left eigenvector of $A$ corresponding to $\lambda_j$ (see (2.2)). This implies

$$|\lambda_j'(0)| \leq \frac{\|x_j\|\,\|y_j\|}{|y_j^* x_j|} \cdot \|E\| \;, \tag{4.2}$$

and the bound (4.2) is sharp for $E = y_j x_j^*$. The number

$$\kappa(\lambda_j) \;=\; \frac{\|x_j\|\,\|y_j\|}{|y_j^* x_j|} \tag{4.3}$$

occurring in the right-hand side of (4.2) is called the *condition number* of the eigenvalue $\lambda_j$ and $\kappa(\lambda_j)$ gives an indication of the sensitivity of $\lambda_j$ if $A$ is perturbed slightly: ignoring higher order terms, one obtains from (4.2) that $|\lambda_j(t) - \lambda_j| \lesssim \kappa(\lambda_j)\|tE\|$ for $\|tE\| \ll 1$.

One might ask whether it is useful to compute $\varepsilon$-pseudospectra if the condition numbers of the eigenvalues are known and vice versa. However, both concepts have their own merit. For example, it may happen that a picture of the $\varepsilon$-pseudospectra suggests that all eigenvalues in a certain cluster have a large condition number, but that needs not to be the case: it may happen that one or more eigenvalues from the cluster are well-conditioned. On the other hand, condition numbers only provide information about the $\varepsilon$-pseudospectra for $\varepsilon \downarrow 0$, and the (theoretical) results mentioned in Section 3 cannot be rewritten in terms of condition numbers. One might say that condition numbers provide local information (for each eigenvalue separately) while $\varepsilon$-pseudospectra give more global information about the matrix $A$.

There is an analogue of (4.1) for generalized eigenvalue problems. Following [12], we consider for $t \geq 0$ the eigenvalue problem

$$(A + tE)\,x_j(t) \;=\; \lambda_j(t)\,(B + tF)\,x_j(t) \;, \tag{4.4}$$

with $E$ and $F$ $N \times N$ matrices. If we assume that $\lambda_j$ is simple and finite then we can differentiate (4.4) for $t = 0$ and we arrive at

$$\lambda_j'(0) \;=\; \frac{y_j^*(E - \lambda_j F)x_j}{y_j^* B x_j}$$

with $y_j$ as in (2.3), which implies

$$|\lambda_j'(0)| \leq \frac{\|x_j\|\,\|y_j\|}{|y_j^* B x_j|} \cdot (\|E\| + |\lambda_j|\,\|F\|) \;. \tag{4.5}$$

The right-hand side of (4.5) multiplied by $t$ provides a first-order "bound" for $|\lambda_j(t) - \lambda_j|$.

The expression (4.5) is more complicated than (4.2) (one actually obtains (4.2) for $B = I$ when $F$ is the zero matrix); note, e.g., that the bound for $|\lambda_j'(0)|$ depends on $\lambda_j$. However, for $|\lambda_j| \approx 1$ the quantity

$$\nu(\lambda_j) \;=\; \frac{\|x_j\|\,\|y_j\|}{|y_j^* B x_j|} \tag{4.6}$$

may give an indication how sensitive $\lambda_j$ is under small perturbations, and therefore one might interpret $\nu(\lambda_j)$ as a condition number for $|\lambda_j| \approx 1$. In [12], other definitions of condition numbers, based on upper bounds for $\|E\|$ and $\|F\|$, have been obtained from (4.5).

In case $B$ is invertible one can also determine the condition number of $\lambda_j$ with respect to $B^{-1}A$: if $\widetilde{y_j}$ is the corresponding left eigenvector of $B^{-1}A$ and $y_j$ is as in (2.3), one has $\widetilde{y_j}^* = y_j^* B$ (or $\widetilde{y_j} = B^* y_j$) which leads to (note that the right eigenvector is the same for both eigenvalue problems)

$$\kappa(\lambda_j) = \frac{\|x_j\|\,\|\widetilde{y_j}\|}{|\widetilde{y_j}^* x_j|} = \frac{\|x_j\|\,\|B^* y_j\|}{|y_j^* B x_j|}\,. \tag{4.7}$$

Hence if one intends to compute (4.6) one gets (4.7) almost for free if the denominator in (4.6) is computed as $y_j^* B x_j = (B^* y_j)^* x_j$; the left eigenvector $\widetilde{y_j}$ is not needed to determine $\kappa(\lambda_j)$.

**Remark 4.1** In [24] the generalized eigenvalue problem is written in the form (2.1) and the authors study the effect of small perturbations of $A$ and $B$ on $\langle \alpha, \beta \rangle$. (We omit the subindex $j$ in $\langle \alpha, \beta \rangle$.) In order to do this one has to measure the distance between $\langle \alpha, \beta \rangle$ and $\langle \tilde{\alpha}, \tilde{\beta} \rangle$, with $\langle \tilde{\alpha}, \tilde{\beta} \rangle$ an eigenvalue of the perturbed pencil $(A + E, B + F)$, in some metric and the so-called chordal distance (see, e.g., [24, p. 283]) turns out to be a proper measure. The chordal distance between $\langle \alpha, \beta \rangle$ and $\langle \tilde{\alpha}, \tilde{\beta} \rangle$ can be bounded by the condition number

$$\frac{\|x_j\|\,\|y_j\|}{\sqrt{|y_j^* A x_j|^2 + |y_j^* B x_j|^2}} \tag{4.8}$$

multiplied by the norm of the perturbation $(E, F)$ (see, e.g., [24, Section VI.2]). As in [12] we call (4.8) the chordal condition number. An advantage of this approach is that one can define the condition number of an infinite eigenvalue. A drawback is that the chordal distance might not be the most natural metric to measure differences in finite eigenvalues. Furthermore, the chordal distance behaves counter-intuitively for large numbers. The chordal condition number (4.8) is of the same order of magnitude as (4.6) for $|\lambda_j| \approx 1$, and we will consider only such eigenvalues in the example presented in Section 5. Therefore we consider only (4.6) as a measure for the condition of the eigenvalues of the generalized eigenvalue problem (1.2) in the remainder of this paper, and we refer to [12, 24] for more discussion and comparisons of the different concepts of condition numbers for eigenvalues stemming from generalized eigenvalue problems.

## 4.2 Condition numbers of eigenvectors

Let $V$ be an $N \times N$ matrix so that the columns of $V$ are the eigenvectors of $A$. The number $\kappa(V)$ may sometimes be used as a measure of the nonnormality of $A$. This condition number is related to the condition numbers of the eigenvalues and $\varepsilon$-pseudospectra: Kantorovič has showed that (cf., e.g., [2])

$$\kappa(V) \geq \kappa(\lambda_j) + \sqrt{\kappa(\lambda_j)^2 - 1} \tag{4.9}$$

for all simple eigenvalues $\lambda_j$ of $A$, and $\Lambda_\varepsilon(A)$ is contained in the union of the disks with radius $\kappa(V)\varepsilon$ centered at the eigenvalues (see, e.g., [28]). If $\kappa(V)$ is very large, then the matrix of eigenvectors is ill-conditioned. This does not necessarily mean that there are eigenvectors which are nearly dependent: as an illustration we consider the following matrix. Let $0 < \delta \ll 1$ and

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & \delta \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & \delta \end{pmatrix}^{-1} . \tag{4.10}$$

For this matrix one has $\kappa(\lambda_j) = \mathcal{O}(\delta^{-1})$ for each eigenvalue $\lambda_j$, $\kappa(V) = \mathcal{O}(\delta^{-1})$, but the angle between two different eigenvectors is larger than $45^o$. However, each eigenvector makes a small angle with the subspace spanned by the two other eigenvectors, and this causes $\kappa(V)$ to be large. So $\kappa(V)$ gives an indication how small the angles between the invariant subspaces may be.

In applications one might be interested only in a part of the spectrum of a matrix or a matrix pencil, and one might ask whether the corresponding set of eigenvectors is well-conditioned. Let $V_k$ be an $N \times k$ matrix of which the columns are eigenvectors of the matrix $A$ and determine a $QR$-factorization of $V_k$,

$$V_k \;=\; Q_k \mathcal{R}_k \;; \tag{4.11}$$

the columns of $Q_k$ form an orthonormal basis of $\operatorname{span}\{V_k\}$ and $\mathcal{R}_k$ is a $k \times k$ upper triangular matrix. Note that $\mathcal{R}_k$ is different from the matrix $R_k$ in (2.6); in fact the columns of $\mathcal{R}_k$ are the eigenvectors of $R_k$. We define the *condition number* of $V_k$ as follows:

$$\kappa(V_k) \;=\; \kappa(\mathcal{R}_k) \;. \tag{4.12}$$

The number $\kappa(V_k)$ does not depend on the choice of the orthonormal basis $Q_k$ of $\operatorname{span}\{V_k\}$. However one should note that $\kappa(V_k)$ depends on, e.g., the scaling of the eigenvectors in the matrix $V_k$: for example, let $V_2$ consist of the first two eigenvectors of (4.10) written as $(1, 0, 0)^{\mathrm{T}}$ and $(0, \gamma, 0)^{\mathrm{T}}$ with $|\gamma| \geq 1$; then $\kappa(V_2) = |\gamma|$, so one should take $|\gamma| = 1$ in order to minimize $\kappa(V_2)$. In general it is not known how the eigenvectors should be chosen to minimize $\kappa(V_k)$. A permutation of the columns of $V_k$ does not change $\kappa(V_k)$, so the only thing one can try to do is to scale the columns of $V_k$ properly. The length of the $j$th column of $V_k$ is equal to the length of the $j$th column of $\mathcal{R}_k$, and it is shown in [23, Theorem 3.5] that $\kappa(\mathcal{R}_k)$ and hence $\kappa(V_k)$ is minimized up to a factor $\sqrt{k}$ if all columns of $\mathcal{R}_k$ (and $V_k$) have the same length. In practice $k$ will not be very large, so one should use (4.12) with the columns of $V_k$ having equal length.

The following theorem, which might be seen as the counterpart of Theorem 3.1, shows that $\kappa(V_k)$ increases with $k$ if the matrices $V_k$ are chosen properly.

**Theorem 4.2** *Let $V_{k+1}$ be an $N \times (k+1)$ matrix of rank $k+1 \leq N$ obtained by adding a column to $V_k$. Then*

$$\kappa(V_k) \;\leq\; \kappa(V_{k+1}) \;.$$

*Proof.* Let $Q_k$ and $\mathcal{R}_k$ be as in (4.11) and denote the $(k+1)$-th column of $V_{k+1}$ as $v$. Similarly, we define the matrix $Q_{k+1}$ by adding a vector $q$ to $Q_k$ where $q$ is such that the columns of $Q_{k+1}$ form an orthonormal basis of $\operatorname{span}\{V_{k+1}\}$. Define

$$\mathcal{R}_{k+1} \;=\; Q_{k+1}^* V_{k+1} \;=\; \begin{pmatrix} \mathcal{R}_k & Q_k^* v \\ 0 & q^* v \end{pmatrix}$$

(note that $q^*V_k = 0$). From this relation it follows that $\sigma_{\max}(\mathcal{R}_{k+1}) = \|\mathcal{R}_{k+1}\| \geq \|\mathcal{R}_k\| = \sigma_{\max}(\mathcal{R}_k)$ and ((cf. 2.4)) $\sigma_{\min}(\mathcal{R}_{k+1}) \leq \sigma_{\min}(\mathcal{R}_k)$; hence (2.5) implies $\kappa(\mathcal{R}_{k+1}) \geq \kappa(\mathcal{R}_k)$, and from $V_{k+1} = Q_{k+1}\mathcal{R}_{k+1}$ and the definition of $\kappa(V_k)$ one gets $\kappa(V_k) \leq \kappa(V_{k+1})$.

□

The condition number $\kappa(V_k)$ provides *only* information of the eigenspaces in the invariant subspace span$\{V_k\}$ and not about the other eigenspaces. The condition number $\kappa(V_k)$ is small if and only if the eigenvectors are a well-conditioned basis of span$\{V_k\}$. Another application is the following: let $\alpha[C] = \max\{\mathrm{Re}\,\lambda : \lambda \text{ is an eigenvalue of } C\}$. It is easy to see that $\|\mathrm{e}^{tA}\| \leq \mathrm{e}^{t\alpha[A]}\kappa(V)$ if $A$ is diagonalizable, and an application of Lemma 3.3 yields for $y \in \mathrm{span}\{V_k\}$ that, with $Q_k$ and $V_k$ as in (4.11) and $R_k = Q_k^* A Q_k$ as in (2.6) (cf. Theorem 3.4),

$$\|\mathrm{e}^{tA} y\| \;\leq\; \mathrm{e}^{t\alpha[R_k]}\,\kappa(V_k)\,\|y\| \quad \text{for} \quad t \geq 0\;;$$

if $\kappa(V_k)$ is not too large this upper bound may be a useful competitor for (3.8).

The condition number $\kappa(V_k)$ of eigenvectors of a matrix pencil $(A, B)$ can be defined in a similar way. For invertible $B$ the columns of $V_k$ are also eigenvectors of $B^{-1}A$; note that $\kappa(V_k)$ does not depend on the actual formulation of the generalized eigenvalue problem.

As in (4.1) one might study what happens with the eigenvector $x_j$ under perturbations of $A$. It is not obvious how the condition number defined in (4.12) can be used to investigate this: note, e.g., that $\kappa(x_j) = 1$ for any eigenvector $x_j$. One can determine $x_j'(0)$ with $x_j(t)$ as in (4.1) if $\lambda_j$ is a simple eigenvalue and $A$ is diagonalizable [30, Section 2.10]:

$$x_j'(0) \;=\; \sum_{i \neq j} \frac{y_i^* E x_j}{(\lambda_j - \lambda_i) y_i^* x_i}\, x_i\;.$$

In order to compute $x_j'(0)$ one has to determine all eigenvalues and left- and right eigenvectors of $A$, which is impractical for large matrices $A$. Instead of considering only one eigenvector one might also study what happens with an invariant subspace under small perturbations of the matrix $A$. We refer to [24, Chapter 5] for more discussion on this topic.

## 4.3 Angles between invariant subspaces

In the previous sections we observed that large condition numbers of eigenvalues and eigenvectors are related to small angles between subspaces spanned by eigenvectors. For instance it may happen that $\kappa(V)$ is large while the condition numbers $\kappa(V_{k_1})$ and $\kappa(V_{k_2})$ (with $V_{k_i}$ an $N \times k_i$ matrix of eigenvectors) are small. In such cases it might be useful to determine the angle between span$\{V_{k_1}\}$ and span$\{V_{k_2}\}$. We now consider two linear subspaces $\mathcal{U}$ and $\mathcal{V}$ of dimension $\ell$ and $k$, respectively, and define the angle $\varphi \in [0, \frac{\pi}{2}]$ between $\mathcal{U}$ and $\mathcal{V}$ as follows (cf. [13, Section 12.4]):

$$\cos\varphi \;=\; \max\left\{\frac{|u^*v|}{\|u\|\,\|v\|} \;:\; u \in \mathcal{U},\; v \in \mathcal{V}\right\}\;. \tag{4.13}$$

Let $Q_{\mathcal{U}}$ and $Q_{\mathcal{V}}$ be orthonormal bases of $\mathcal{U}$ and $\mathcal{V}$ respectively. It is easy to see that (cf. [13, Section 12.4])

$$\cos\varphi \;=\; \sigma_{\max}(Q_{\mathcal{U}}^* Q_{\mathcal{V}}) \;=\; \sigma_{\max}(Q_{\mathcal{V}}^* Q_{\mathcal{U}})\;; \tag{4.14}$$

the angle $\varphi$ in (4.14) does not depend on the choice of the matrices $Q_{\mathcal{U}}$ and $Q_{\mathcal{V}}$.

For $k = \ell$ the arccosines of the singular values of $Q_{\mathcal{U}}^* Q_{\mathcal{V}}$ are called the *canonical angles* between $\mathcal{U}$ and $\mathcal{V}$.

In our applications $\mathcal{U}$ and $\mathcal{V}$ are invariant subspaces with $\mathcal{U} \cap \mathcal{V} = \{0\}$ of which bases (consisting, e.g., of eigenvectors) are available, and after determination of orthonormal bases of the spaces, the angle $\varphi$ can be computed easily from (4.14). We assume that $k$ and $\ell$ are not too large, so that the singular values of $Q_{\mathcal{U}}^* Q_{\mathcal{V}}$ can be computed. If the JDQR or JDQZ method [10] is used to compute eigenvectors, then an orthonormal basis of a corresponding invariant subspace is generated, so the angle can be determined directly from (4.14).

**Remark 4.3** In the literature other ways of measuring the "distance" between linear subspaces is studied; cf., e.g., [13, 24]. The subspaces $\mathcal{U}$ and $\mathcal{V}$ contain both the zero vector, so the ordinary distance $\min\{\|u - v\| : u \in \mathcal{U},\ v \in \mathcal{V}\}$ between $\mathcal{U}$ and $\mathcal{V}$ is not a useful concept in this case. In [24] the gap between $\mathcal{U}$ and $\mathcal{V}$ has been defined for arbitrary norms. We restrict ourselves to the Euclidean norm and adopt the notation from [24]: the *gap* between $\mathcal{U}$ and $\mathcal{V}$, denoted as $\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V})$, is defined as follows[3]:

$$\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V}) \;=\; \max\left\{\, \max\{\mathrm{d}(u, \mathcal{V}) : u \in \mathcal{U}, \|u\| = 1\}\,,\ \ \max\{\mathrm{d}(v, \mathcal{U}) : v \in \mathcal{V}, \|v\| = 1\}\,\right\},$$

where $d(x, \mathcal{U}) = \min\{\|x - u\| : u \in \mathcal{U}\}$ is the distance from $x$ to $\mathcal{U}$. Note that for $x \in \mathbb{C}^N$, $d(x, \mathcal{U}) = \|(I - Q_{\mathcal{U}} Q_{\mathcal{U}}^*)x\| \le \|x\|$ ($Q_{\mathcal{U}}$ is as above) and similarly $d(x, \mathcal{V}) \le \|x\|$, so that $\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V}) \le 1$. For $k \neq \ell$ one always has[4] $\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V}) = 1$ (cf., e.g., [24, p. 99]) so $\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V})$ does not provide much information about $\mathcal{U}$ and $\mathcal{V}$ in that case. If $k = \ell$ one can show that (see, e.g., [24, Section II.4])

$$\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V}) \;=\; \sqrt{1 - \sigma_{\min}^2(Q_{\mathcal{U}}^* Q_{\mathcal{V}})}\ ;$$

this equality implies that $\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V})$ is the sine of the largest canonical angle between $\mathcal{U}$ and $\mathcal{V}$ if the dimensions of $\mathcal{U}$ and $\mathcal{V}$ are equal.

# 5 An illustration from magnetohydrodynamics

We apply the different concepts discussed in the previous sections to a problem from magnetohydrodynamics (MHD), which has been taken from [14]. This leads to a generalized eigenvalue problem of the form (1.2) with a matrix $B$ which is Hermitian and positive definite. The $\varepsilon$-pseudospectra and condition numbers related to this problem are very large, and depend also on the formulation of the eigenvalue problem.

## 5.1 A description of the model

The model deals with a plasma in a tokamak reactor, which has the geometry of a torus. In [14] the authors consider equations linearized around an equilibrium. In this equilibrium the

[3] For readers familiar with the concept of Hausdorff distance we note that $\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V})$ is equal to the Hausdorff distance between $\mathcal{U} \cap \mathcal{S}$ and $\mathcal{V} \cap \mathcal{S}$, where $\mathcal{S} = \{x \in \mathbb{C}^N : \|x\| = 1\|\}$.

[4] Assume $\ell < k$: then $N - \ell + k > N$ which implies that the intersection of $\mathcal{V}$ and the orthogonal complement of $\mathcal{U}$ have a vector $\widehat{v}$ with $\|\widehat{v}\| = 1$ in common, and $\|(I - Q_{\mathcal{U}} Q_{\mathcal{U}}^*)\widehat{v}\| = 1$ implies $\rho_{\mathrm{g}}(\mathcal{U}, \mathcal{V}) = 1$. If $\ell > k$ one should interchange the roles of $\mathcal{U}$ and $\mathcal{V}$.

velocity of the plasma is equal to zero, and the linearized equations read as follows:

$$
\begin{aligned}
\frac{\partial \rho_1}{\partial t} &= -\nabla \cdot (\rho_0 v_1) \ , \\
\rho_0 \frac{\partial v_1}{\partial t} &= -\nabla p_1 \ + \ (\nabla \times B_0) \times B_1 \ + \ (\nabla \times B_1) \times B_0 \ , \\
\frac{\partial p_1}{\partial t} &= -v_1 \cdot \nabla p_0 \ - \ \gamma p_0 \nabla \cdot v_1 \ , \\
\frac{\partial B_1}{\partial t} &= \nabla \times (v_1 \times B_0 \ - \ \eta \nabla \times B_1) \ , \quad \nabla \cdot B_1 \ = \ 0 \ .
\end{aligned}
\tag{5.1}
$$

Here a subindex 0 corresponds to the equilibrium, and a subindex 1 refers to the perturbation of this equilibrium. The letter $\rho$ stands for the density, $v$ for the velocity (note that $v_0 = 0$), $p$ for the pressure and $B$ for the magnetic field. Furthermore $\gamma$ and $\eta$ are constants of which $\eta$, the resistivity, plays an important role in the analysis in [14]. The boundary conditions are given in [14]. Another important quantity in this problem which does not appear in the differential equations is the aspect ratio, i.e. the ratio of the large and the small circle defining the torus. The dependence of the MHD spectrum on this aspect ratio is studied in [14]; here we take the aspect ratio equal to 5 and the (normalized) resistivity $\eta = 2.5 \cdot 10^{-5}$; this situation has been considered also in [14].

The system (5.1) is discretized using finite elements in the radial direction of the small circle of the torus and Fourier modes in the poloidal direction (which is related to the angle of the small circle). This leads to an eigenvalue problem of the form (1.2) with

$$
N \ = \ 16\, M \cdot N_r \ , \tag{5.2}
$$

where $M$ is the number of Fourier modes and $N_r$ the number of radial gridpoints. The matrices $A$ and $B$ are block tridiagonal: the number of blocks is equal to $N_r$ and the blocksize is $16M$. The matrices $A$ and $B$ are generated by a code called CASTOR [15]. For more details on this problem and the discretization we refer to [14].

We are interested in a part of the MHD spectrum only: the Alfvén spectrum. The Alfvén spectrum has been computed in [14] for $M = 4$ and $N_r = 1000$. A variant of the Jacobi-Davidson method [22] has been applied in that paper to $(A - \tau B)^{-1}B$ where $\tau$ is the target: if $\mu$ is an eigenvalue of $(A - \tau B)^{-1}B$ then $\lambda = \tau + 1/\mu$ is an eigenvalue of (1.2). Due to the fact that $A - \tau B$ is block tridiagonal it is possible to compute an $LU$-factorization of that matrix at a reasonable cost.

For this problem it turns out that taking $N_r = 100$ instead of $N_r = 1000$ is sufficient to compute the Alfvén spectrum (one needs, e.g., a larger resolution for smaller values of $\eta$); we have computed the eigenvalues displayed in Figure 5.1 for both values of $N_r$ and $M = 4$, and the difference between the corresponding eigenvalues was less than 0.5%. Therefore we take $M = 4$ and $N_r = 100$ in our experiments; this reduces the size of the matrices from $N = 64000$ to $N = 6400$ (cf. (5.2)). Working with these smaller matrices is in particular important for the computation of $\varepsilon$-pseudospectra, which can be very time consuming. All our computations have been performed on the CRAY C90 at SARA, Amsterdam (The Netherlands).

We have applied the JDQZ method [10] to compute the Alfvén spectrum, using the LU-factorization of $A - \tau B$ as a preconditioner for the correction equation. Some eigenvalues could not be computed unless the target $\tau$ was chosen very close to these eigenvalues. This might be a consequence of the fact that these eigenvalues are very ill-conditioned: see Sections

5.2 and 5.3. With one target we were able to find at most a few eigenvalues. The Alfvén spectrum is displayed in Figure 5.1.

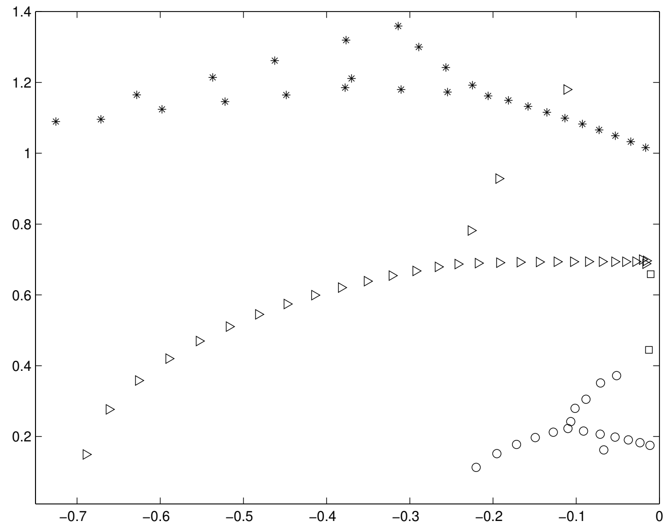


**Figure 5.1.** *The Alfvén spectrum.*

This figure corresponds to [14, Figure 3.b]; in [14, Figure 3.b] the eigenvalues are scaled by the inverse aspect ratio 1/5. In the upper part of Figure 5.1 there are some eigenvalues which are not displayed in [14, Figure 3.b]. In order to be able to distinguish between the different invariant subspaces later on (cf. Sections 3.3, 4.2 and 4.3), we have used different markers to plot the eigenvalues. (For instance, we will speak of the invariant subspace spanned by the eigenvectors corresponding to the eigenvalues marked with a small circle ($\circ$), etc.) The subdivision of the eigenvalues is based on physical properties of the MHD problem; see [14] for more discussion on the model. There are also eigenvalues on the negative real axis, but these are not considered in [14] and the present paper.

## 5.2 Pseudospectra of the MHD problem

We have determined both $\Lambda_\varepsilon(A, B)$ and $\Lambda_\varepsilon(B^{-1}A)$ on the domain in Figure 5.1. The quantities $\sigma_{\min}(zB - A)$ and $\sigma_{\min}(zI - B^{-1}A)$ have been computed with the inverse power method as explained in Section 3.4.2, on a $100 \times 100$ grid. The iterations were stopped if the difference between two consecutive approximations of the smallest singular value is less than 0.1%. The total CPU-time for the computation of $\Lambda_\varepsilon(A, B)$ and $\Lambda_\varepsilon(B^{-1}A)$ is $1.56 \cdot 10^4$ seconds (= 260 minutes). To obtain a picture on a $50 \times 50$ grid would cost about 25% of the computational time for the $100 \times 100$ grid and the picture would be essentially of the same quality. On average we needed 3.72 inverse power iterations per grid point for $\Lambda_\varepsilon(A, B)$ and 6.74 inverse power iterations per grid point for $\Lambda_\varepsilon(B^{-1}A)$.

The $\varepsilon$-pseudospectra are plotted in Figure 5.2. From this figure we see that the $\varepsilon$-pseudospectra of $B^{-1}A$ differ a lot from $\Lambda_\varepsilon(A, B)$ and the latter is very large: for example, the eigenvalues of the pencil $(A + E, B)$ with $\|E\| \leq 1.0 \cdot 10^{-11}$ can be anywhere in the area considered in Figure 5.2. One might try to understand this difference by looking at (3.7). One has $\|B\| = 9.3 \cdot 10^1$, $\|B^{-1}\| = 1.3 \cdot 10^{13}$, so both inclusions in (3.7) are far from sharp. The big

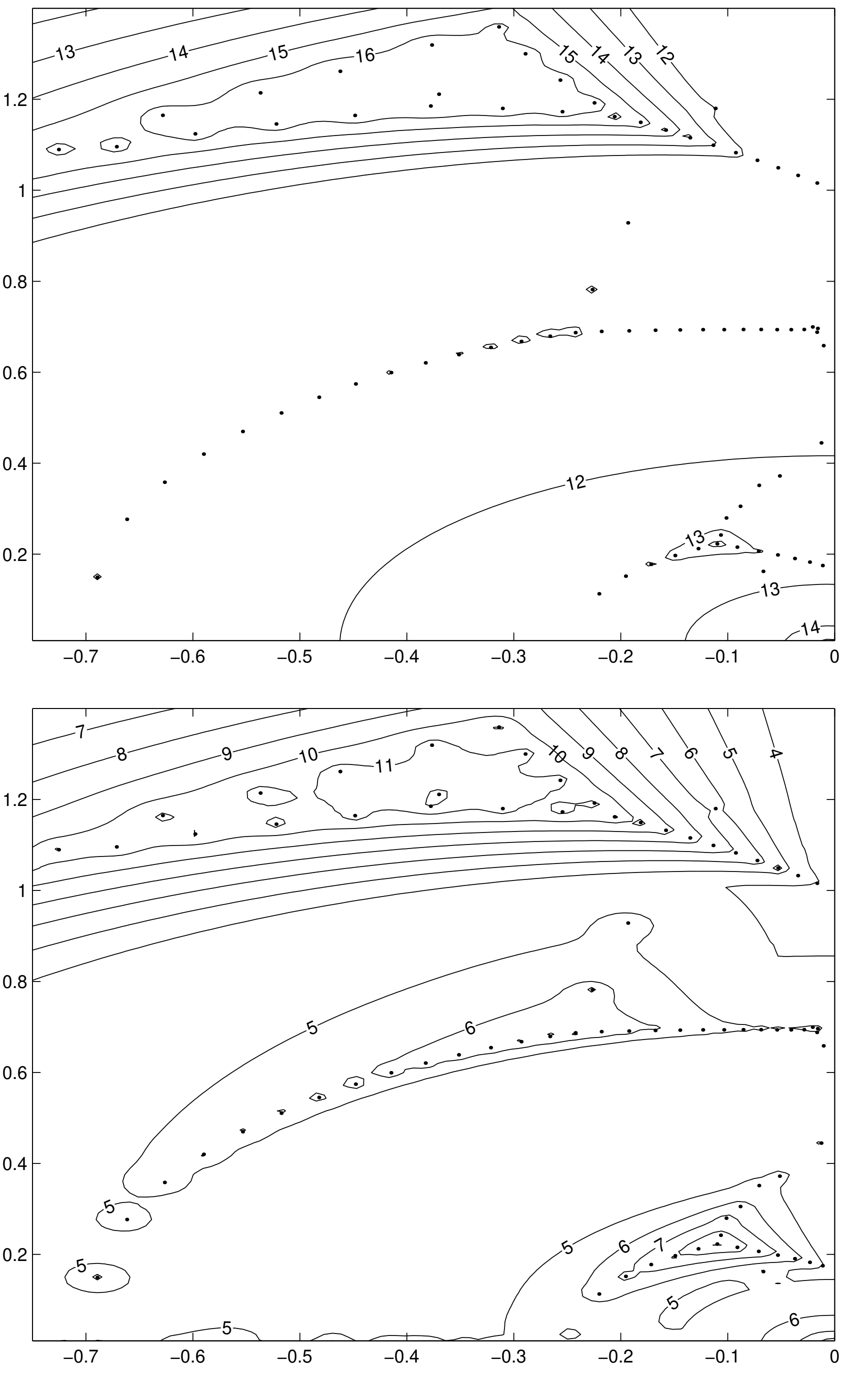


**Figure 5.2.** *The sets* $\Lambda_\varepsilon(A, B)$ *(upper picture) and* $\Lambda_\varepsilon(B^{-1}A)$ *(lower picture). An integer* $j$ *on a curve means that this curve is part of the boundary of the* $\varepsilon$*-pseudospectrum with* $\varepsilon = 10^{-j}$*. The eigenvalues are indicated by dots.*

difference between $\Lambda_\varepsilon(A,B)$ and $\Lambda_\varepsilon(B^{-1}A)$ might be explained as follows: let $B = U_B \Sigma_B V_B^*$ be the singular value decomposition (see, e.g., [13]) of $B$. Then $\Lambda_\varepsilon(A,B) = \Lambda_\varepsilon(U_B^* A V_B, \Sigma_B)$ and $\Lambda_\varepsilon(B^{-1}A) = \Lambda_\varepsilon(\Sigma_B^{-1} U_B^* A V_B)$; the effect of the small singular values of $B$ is that some entries of $\Sigma_B^{-1} U_B^* A V_B$ are much larger than the corresponding entries of $U_B^* A V_B$ and this might be the reason why perturbations may have less effect on the eigenvalues of $\Sigma_B^{-1} U_B^* A V_B$ than on the eigenvalues of $(U_B^* A V_B, \Sigma_B)$ or $(A,B)$. This example nicely illustrates what can happen if there is a large difference between the size of the smallest and largest singular values of $B$ (i.e. $\kappa(B)$ is large).

From Figure 5.2 we see that the $\varepsilon$-pseudospectra are very large in a triangular shaped region in the upper part of the plots, and one might ask the question whether the eigenvalues found in that region are actually correct. (In Section 5.3 we will also see that these eigenvalues are very ill-conditioned.)

The other paper we are aware of that deals with $\varepsilon$-pseudospectra for an MHD problem is [3]; however, the problem considered in that paper differs essentially from ours: the authors consider an incompressible cylindrical plasma while the problem studied in [14] and this section is a compressible plasma in a tokamak. One property of $\varepsilon$-pseudospectra for MHD problems observed both in [3] and Figure 5.2 is that for small $\varepsilon$, the $\varepsilon$-pseudospectra become larger in the neighbourhood of a triple point, i.e., a point where three branches on which eigenvalues are intersect; cf., e.g., Figure 5.2 in the neighbourhood of the triple point near $-0.11 + 0.22i$.

We now consider $\varepsilon$-pseudospectra restricted to invariant subspaces. Let $Q_k$ be a matrix of which the columns form an orthonormal basis of the subspace spanned by the eigenvectors corresponding to the eigenvalues indicated by small circles in Figure 5.1, and let $S_k$ and $T_k$ be as in (2.7); $k = 18$ in this case. The sets $\Lambda_\varepsilon(S_k, T_k)$ and $\Lambda_\varepsilon(T_k^{-1} S_k)$ are visualized in Figure 5.3 (note that the scale in Figure 5.3 is different from the scale in the previous figures).

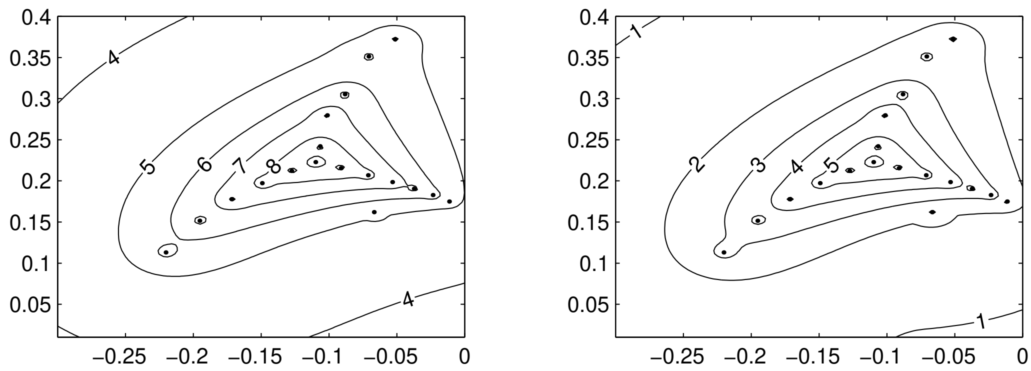


**Figure 5.3.** *The sets $\Lambda_\varepsilon(S_k, T_k)$ (left) and $\Lambda_\varepsilon(T_k^{-1} S_k)$ (right). An integer $j$ on a curve means that this curve is part of the boundary of the $\varepsilon$-pseudospectrum with $\varepsilon = 10^{-j}$. The eigenvalues are indicated by dots.*

From Figure 5.3 we see that $\Lambda_\varepsilon(A,B)$ is much larger than $\Lambda_\varepsilon(S_k, T_k)$, and $\Lambda_\varepsilon(B^{-1}A)$ is much larger than $\Lambda_\varepsilon(T_k^{-1} S_k)$; the approximation of the $\varepsilon$-pseudospectra of the large problems with $N = 6400$ by the $\varepsilon$-pseudospectra of the small problems with $k = 18$ leads to inaccurate results, even in the neighbourhood of the part of the spectrum considered here, despite the fact that the $k$ eigenvalues are well separated from the other part of the spectrum (see Figure

5.1). This example illustrates that one should be careful with using invariant subspaces to approximate $\varepsilon$-pseudospectra.

If we compare Figure 5.2 and 5.3, we see that the shape of the curves in the neighbourhood of the eigenvalues considered is the same; in particular, for small $\varepsilon$, the $\varepsilon$-pseudospectra are large in the neighbourhood of the triple point. Moreover, there is hardly any difference between the shape of the curves in the left and right picture of Figure 5.3; only the corresponding values of $\varepsilon$ differ roughly by a factor $10^3$. This behaviour can be explained as follows: one has $\|S_k\| = 9.1 \cdot 10^{-4}$, $\|T_k\| = 5.3 \cdot 10^{-3}$ and $\kappa(T_k) = 7.2$. The identity $\Lambda_{10^{-3}\varepsilon}(S_k, T_k) = \Lambda_\varepsilon(10^3 S_k, 10^3 T_k)$ combined with (3.7), with $A$ and $B$ replaced by $S_k$ and $T_k$, respectively, explains the resemblance between the left and right picture in Figure 5.3. Note that the scaling of $S_k$ and $T_k$ depends only on the scaling of $A$ and $B$ in the original problem.

We have also computed $\varepsilon$-pseudospectra restricted to the subspaces corresponding to the eigenvalues indicated with triangles ($\triangleright$) and asterisks ($*$) in Figure 5.1, and the behaviour turns out to be similar as for the subspace considered before; therefore we will not discuss these $\varepsilon$-pseudospectra further.

## 5.3 Condition numbers for the MHD problem

For each eigenvalue $\lambda_j$ in Figure 5.1 we have computed the quantity $\nu(\lambda_j)$ from (4.6), which gives a first order indication of the effect of perturbations of $A$ and $B$ on the eigenvalues of (1.2). Also the condition numbers $\kappa(\lambda_j)$ of the eigenvalues of $B^{-1}A$ have been determined (cf. (4.7)). We have applied the JDQZ method [10] to $(A^{\mathrm{T}}, B^{\mathrm{T}})$ to obtain the left eigenvectors of (1.2), and $\kappa(\lambda_j)$ is then computed from the second equality in (4.7). The numbers $\nu(\lambda_j)$ and $\kappa(\lambda_j)$ are visualized in Figure 5.4.

From Figure 5.4 we observe that $\nu(\lambda_j) \gg \kappa(\lambda_j)$, and this is not a surprise in view of Section 5.2. The number $\nu(\lambda_j)$ ranges from $1.0 \cdot 10^7$ to $1.0 \cdot 10^{16}$, so that some eigenvalues of the matrix pencil are better conditioned than others; this could not be predicted from the plot of $\Lambda_\varepsilon(A, B)$ in Figure 5.2, because the $\varepsilon$-pseudospectra are large throughout the area.

The condition numbers and the $\varepsilon$-pseudospectra of $B^{-1}A$ (Figure 5.2) are nicely in agreement in the sense that the magnitude of the condition numbers $\kappa(\lambda_j)$ may be predicted from $\Lambda_\varepsilon(B^{-1}A)$ (and vice versa).

If we compare both pictures in Figure 5.4 we see that the pattern is very similar: this behaviour is not obvious, because $\nu(\lambda_j)$ measures the effect of perturbations of $(A, B)$ while $\kappa(\lambda_j)$ deals with perturbations of the matrix $B^{-1}A$. For example, in both cases the eigenvalues closest to the imaginary axis are the ones with the smallest condition number. Also the eigenvalues near the triple point $-0.11 + 0.22i$ have larger condition numbers than the eigenvalues at the endpoints of those branches which intersect at that triple point.

Now we discuss the condition numbers of the clusters of eigenvectors (cf. Section 4.2) and angles between invariant subspaces (cf. Section 4.3). Note that the dimension of the subspaces is different, so the gap between the subspaces, as defined in Remark 4.3, does not provide any information in this case. We denote the set of eigenvectors corresponding to the eigenvalues indicated by small circles in Figure 5.1 by $V_k(\circ)$, etc. The results are displayed in Table 5.1.

The magnitude of $\kappa(V_k)$ is not surprising in view of the $\varepsilon$-pseudospectra of $B^{-1}A$ and the condition numbers of the eigenvalues. The angle between $\mathrm{span}\{V_k(\circ)\}$ and the other two invariant subspaces is not small, which means that $\mathrm{span}\{V_k(\circ)\}$ is well separated from the other two. On the other hand, the angle between $\mathrm{span}\{V_k(\triangleright)\}$ and $\mathrm{span}\{V_k(*)\}$ is very small.

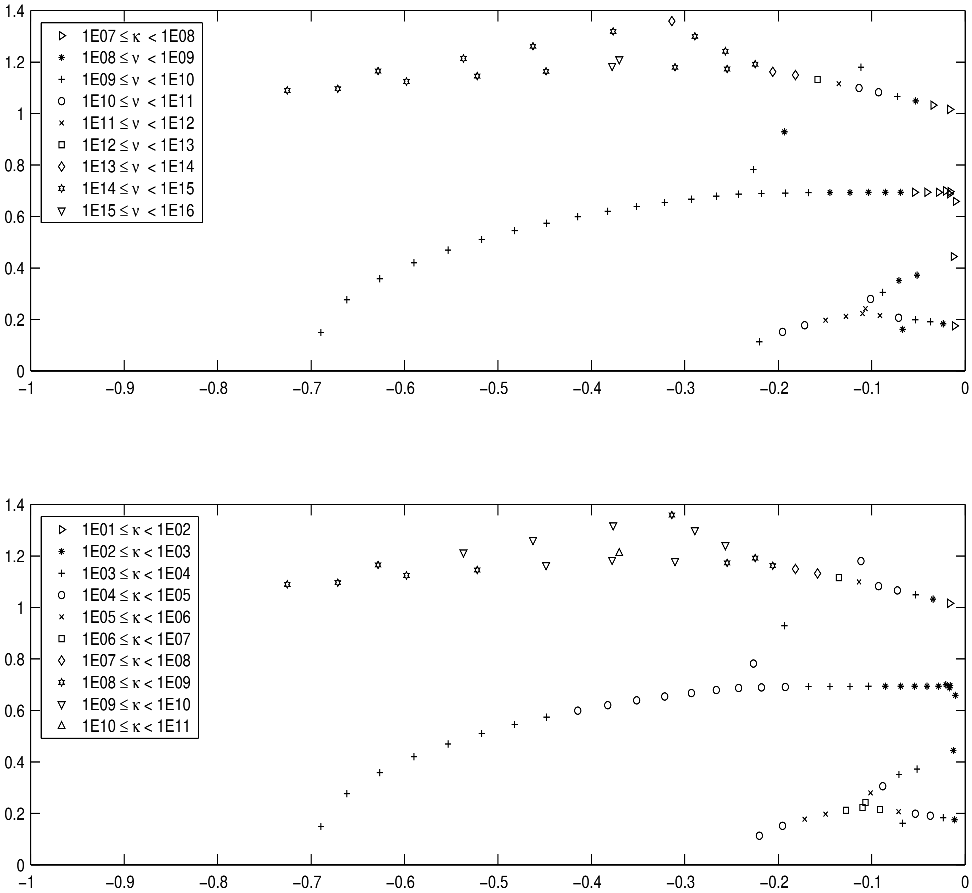


**Figure 5.4.** *The numbers $\nu$ from (4.6) (upper picture) and the condition numbers $\kappa$ from (4.7) (lower picture). Different markers are used to indicate the magnitude of the numbers $\nu$ and $\kappa$.*

| $V_k$ | $k$ | $\kappa(V_k)$ | $\angle(V_k, V_k(\circ))$ | $\angle(V_k, V_k(\triangleright))$ | $\angle(V_k, V_k(*))$ |
|---|---|---|---|---|---|
| $V_k(\circ)$ | 18 | $2.2 \cdot 10^5$ | 0 | 39.5 | 45.5 |
| $V_k(\triangleright)$ | 32 | $4.4 \cdot 10^2$ | 39.5 | 0 | 0.22 |
| $V_k(*)$ | 27 | $4.4 \cdot 10^7$ | 45.5 | 0.22 | 0 |

**Table 5.1.** *In this table the condition numbers of the different sets of eigenvectors are displayed, as well as the angles (in degrees) between the corresponding invariant subspaces. The number of eigenvectors in each set is given in the second column of the table.*

The behaviour of the angles cannot be explained by inspecting the $\varepsilon$-pseudospectra. Because the angle between span$\{V_k(\triangleright)\}$ and span$\{V_k(*)\}$ is so small, we also consider the set of eigenvectors which consists of the union of $V_k(\triangleright)$ and $V_k(*)$, and we call this set $V_k(\diamond)$. We have computed $\kappa(V_k(\diamond))$ and the angle between span$\{V_k(\circ)\}$ and span$\{V_k(\diamond)\}$. One has $\kappa(V_k(\diamond)) = 6.3 \cdot 10^7$ which is the same order of magnitude as $\kappa(V_k(*))$; Theorem 4.2

implies that $\kappa(V_k(\diamond)) \geq \kappa(V_k(*))$. Note that the condition numbers of all sets of eigenvectors considered here is much smaller than $\kappa(V)$ if $V$, the set of all eigenvectors, exists: if we apply (4.9) to all $\lambda_j$ displayed in the lower picture of Figure 5.4 we find $\kappa(V) > 2.1 \cdot 10^{10}$. The angle between $V_k(\diamond)$ and $V_k(\circ)$ is $35.3^o$, which is slightly smaller than the angle between $\text{span}\{V_k(\circ)\}$ and $\text{span}\{V_k(\triangleright)\}$, so $\text{span}\{V_k(\circ)\}$ and $\text{span}\{V_k(\diamond)\}$ are well separated.

# 6 Conclusions

We deal with several concepts which can be used to understand the behaviour of standard and generalized eigenvalue problems that are strongly nonnormal, i.e. problems for which the eigenvalues or the matrix of eigenvectors (if it exists) are ill-conditioned. For those problems it may be difficult to compute the eigenvalues accurately and moreover, an analysis based on the exact eigenvalues only may not reveal some important properties of the matrices under consideration. The concepts we consider are $\varepsilon$-pseudospectra, condition numbers of eigenvalues and a set of eigenvectors, and angles between invariant subspaces. Apart from the $\varepsilon$-pseudospectra related to the whole eigenvalue problem, we consider also $\varepsilon$-pseudospectra restricted to invariant subspaces. This might be useful in some applications, and these $\varepsilon$-pseudospectra are much cheaper to compute. We compare all these tools and discuss their advantages and disadvantages.

In practice these tools may be applied to large matrix problems originating from physics, chemistry or other applications, and therefore we also discuss the computation of these tools. In particular the computation of $\varepsilon$-pseudospectra may be very expensive and this is considered in more detail.

We apply the concepts studied in this paper to a problem from magnetohydrodynamics. This leads to a large generalized eigenvalue problem of the form (1.2) with a nonsingular matrix $B$. It turns out that this problem is very ill-conditioned and that the $\varepsilon$-pseudospectra and the condition number of the eigenvalues depend strongly on whether the problem is considered in the form (1.2) or (2.1). Each tool reveals some information about the eigenvalue problem that the others do not, so this illustrates that all concepts considered in this paper are worthwhile to investigate for strongly nonnormal eigenvalue problems arising in practice.

**Acknowledgements.** The author wishes to thank Bart van der Holst and Margreet Nool for providing the CASTOR code which generates the matrices $A$ and $B$ from Section 5, and Diederik Fokkema and Martin van Gijzen for providing the JDQZ code [10]. He further acknowledges Sander Beliën, Bart van der Holst, Margreet Nool and Henk van der Vorst for stimulating discussions on the topic of this paper. Further he is indebted to Henk van der Vorst for suggesting improvements to a preliminary version of this paper, and to Michiel Hochstenbach for pointing out Formula (4.9). Our numerical computations were performed on the CRAY C90 at SARA (Amsterdam, The Netherlands) and the Dutch National Computing Facilities Foundation (NCF) is acknowledged for providing the possibility to use this computer.

# References

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe & H. van der Vorst, eds: *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide.* Philadelphia: SIAM 2000.

[2] F.L. Bauer & C.T. Fike: Norms and exclusion theorems. *Numer. Math.* **2**, 137-141 (1960).

[3] D. Borba, K.S. Riedel, W. Kerner, G.T.A. Huysmans, M. Ottaviani & P.J. Schmid: The pseudospectrum of the resistive magnetohydrodynamics operator: Resolving the resistive Alfvén paradox. *Phys. Plasmas* **1**, 3151-3160 (1994).

[4] T. Braconnier & N.J. Higham: Computing the field of values and pseudospectra using the Lanczos method with continuation. *BIT* **36**, 422-440 (1996).

[5] J.F. Carpraux, J. Erhel & M. Sadkane: Spectral portrait for non-Hermitian large sparse matrices. *Computing* **53**, 301-310 (1994).

[6] J.L.M. van Dorsselaer: Pseudospectra for matrix pencils and stability of equilibria. *BIT* **37**, 833-845 (1997).

[7] J.L.M. van Dorsselaer, J.F.B.M. Kraaijevanger & M.N. Spijker: Linear stability analysis in the numerical solution of initial value problems. *Acta Numerica 1993*, 199-237 (1993).

[8] N. Dunford & J.T. Schwartz: *Linear Operators Part I: General Theory.* New York: Interscience Publishers Inc. 1957.

[9] M. Embree & L.N. Trefethen: Generalizing eigenvalue theorems to pseudospectra theorems. Manuscript. Oxford University Computing Laboratory, Oxford. Submitted to *SIAM J. Sci. Comput.*

[10] D.R. Fokkema, G.L.G. Sleijpen & H.A. van der Vorst: Jacobi-Davidson style QR and QZ algorithms for the reduction of matrix pencils. *SIAM J. Sci. Comput.* **20**, 94-125 (1998).

[11] V. Frayssé, M. Gueury, F. Nicoud & V. Toumazou: Spectral portraits for matrix pencils. Tech. Report TR/PA/96/19, CERFACS, Toulouse, 1996.

[12] V. Frayssé & V. Toumazou: A note on the normwise perturbation theory for the regular generalized eigenproblem. *Numer. Linear Algebra Appl.* **5**, 1-10 (1998).

[13] G.H. Golub & C.F. van Loan: *Matrix Computations*, 3rd Ed. Baltimore: The John Hopkins University Press 1996.

[14] B. van der Holst, A.J.C. Beliën, J.P. Goedbloed, M. Nool & A. van der Ploeg: Calculation of resistive magnetohydrodynamic spectra in tokamaks. *Phys. Plasmas* **6**, 1554-1561 (1999).

[15] W. Kerner, J.P. Goedbloed, G.T.A. Huysmans, S. Poedts & E. Schwartz: CASTOR: Normal-mode analysis of resistive MHD plasmas. *J. Comp. Phys.* **142**, 271-303 (1998).

[16] S.H. Lui: Computing of pseudospectra by continuation. *SIAM J. Sci. Comput.* **18**, 565-573 (1997).

[17] S.C. Reddy & L.N. Trefethen: Lax-stability of fully discrete spectral methods via stability regions and pseudo-eigenvalues. *Comp. Meth. Appl. Mech. Eng.*, **80**, 147-164 (1990).

[18] S.C. Reddy & L.N. Trefethen: Stability of the method of lines. *Numer. Math.* **62**, 235-267 (1992).

[19] K.S. Riedel: Generalized epsilon-pseudospectra. *SIAM J. Numer. Anal.* **31** (1994), 1219-1225 (1994).

[20] A. Ruhe: The rational Krylov algorithm for large nonsymmetric eigenvalues – mapping the resolvent norms (pseudospectrum). Unpublished note (1994).

[21] Y. Saad & M.H. Schultz: GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.* **7**, 856-869 (1986).

[22] G.L.G. Sleijpen & H.A. van der Vorst: A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.* **17**, 401-425 (1996).

[23] A. van der Sluis: Condition numbers and equilibration of matrices. *Numer. Math.* **14**, 14-23 (1969).

[24] G.W. Stewart & J.-G. Sun: *Matrix Perturbation Theory.* San Diego: Academic Press Inc. 1990.

[25] K.-C. Toh and L.N. Trefethen: Calculation of pseudospectra by the Arnoldi iteration. *SIAM J. Sci. Comput.* **17**, 1-15 (1996).

[26] L.N. Trefethen: Pseudospectra of matrices. In *Numerical Analysis 1991* (D.F. Griffiths & G.A. Watson eds), 234 - 266. Harlow: Longman 1992.

[27] L.N. Trefethen: Pseudospectra of linear operators. *SIAM Review* **39**, 383-406 (1997).

[28] L.N. Trefethen: Computation of pseudospectra. *Acta Numerica 1999*, 247-295 (1999).

[29] L.N. Trefethen, A.E. Trefethen, S.C. Reddy, & T.A. Driscoll: Hydrodynamic stability without eigenvalues. *Science* **261**, 578-584 (1993).

[30] J.H. Wilkinson: *The Algebraic Eigenvalue Problem.* Oxford: Oxford University Press 1965.

[31] Th.G. Wright & L.N. Trefethen: Large-scale computation of pseudospectra using ARPACK and eigs. Report 00/11 Oxford University Computing Laboratory, Numerical Analysis group, 2000. Submitted to *SIAM J. Sci. Comput.*