The Complexity of the Algebraic Eigenproblem *

Victor Y. Pan[1], Zhao Q. Chen[2], Ailong Zheng[2]

[1] Department of Mathematics & Computer Science, Lehman College, CUNY, Bronx, NY 10468. Internet address: vpan@alpha.lehman.cuny.edu

[2] Ph.D. Program in Math., Graduate Center, CUNY, New York, NY 10036

December 14, 1998

Abstract

The eigenproblem for an n-by-n matrix A is the problem of the approximation (within a relative error bound 2^{-b}) of all the eigenvalues of the matrix A and computing the associated eigenspaces of all these eigenvalues. We show that the arithmetic complexity of this problem is bounded by $O(n^3 + (n\log^2 n)\log b)$. If the characteristic and minimum polynomials of the matrix A coincide with each other (which is the case for generic matrices of all classes of general and special matrices that we consider), then the latter deterministic cost bound can be replaced by the randomized bound $O(K_A(2n) + n^2 + (n\log^2 n)\log b)$ where $K_A(2n)$ denotes the cost of the computation of the $2n-1$ vectors $A^i\mathbf{v}$, $i = 1,\ldots,2n-1$, maximized over all n-dimensional vectors $\mathbf{v}$; $K_A(2n) = O(M(n)\log n)$, for $M(n) = o(n^{2.376})$ denoting the arithmetic complexity of $n\times n$ matrix multiplication. This bound on the complexity of the eigenproblem is optimal up to a logarithmic factor and implies much faster solution of the eigenproblem for the important special classes of matrices. In particular, we prove the bound $O(n^2\log n + (n\log^2 n)\log b)$ on the randomized arithmetic complexity of the eigenproblem for generic matrices of the classes of $n\times n$ Toeplitz, Hankel, Toeplitz-like, Hankel-like and Toeplitz-like-plus-Hankel-like matrices. Then again, this bound is optimal (up to a logarithmic factor) for each of the latter classes of input matrices. We also prove similar nearly optimal upper bounds for the generic Cauchy-like, Vandermonde-like and sparse matrices. All our complexity estimates improve the known ones by order of magnitude.

*Supported by NSF Grants CCR 9625344 and CCR 9732206 and by PSC CUNY Awards 668365 and 669363

Key words: matrix eigenvalues, eigenvectors, characteristic polynomial, polynomial roots, algorithms, computational complexity, structured matrices, sparse matrices

1991 Math. Subject Classification: 65F15, 68Q25, 68Q40, 15A18

1 Introduction

1.1 Matrix eigenproblem and some open issues

The solution of the matrix eigenproblem (that is, the problem of the computation of the matrix eigenvalues and the eigenspaces of the associated eigenvectors) is a central classical topic of applied linear algebra. The history of this problem can be traced back at least to the first half of the 19th century, when, motivated by applications to celestial mechanics, Cauchy discovered his celebrated Interlacing Theorem for the eigenvalues. Recently the eigenproblem turned out to be highly important also for the solution of polynomial systems of equations [AuSt88], [Ste96], [MP98], [BMP98].

The subject has enormous bibliography (see e.g. [GL96] and references therein) and causes never stopping stream of research articles in the leading journals on numerical analysis and applied linear algebra. The importance of this topic is highly recognized by numerical analysts (e.g. works on it have been awarded by the SIAM Prize in Applied Linear Algebra) but much less so by the computer science theoreticians, even though some major practical and theoretical algorithmic problems in this area were widely open.

In particular, the customary iterative algorithms for the eigenproblem are relatively slow for the worst case (unsymmetric) input matrix. We will state this quantitatively, measuring the arithmetic computational cost by the number of ops involved, where "ops" will be our abbreviation for both arithmetic operations (over real or complex numbers) and comparisons (of the absolute values or the moduli of pairs of real or complex numbers). We observe that for a general $n \times n$ input matrix A, the most popular QR algorithm requires order of n^3 ops to reduce the matrix to Hessenberg form and then order of Mn^3 ops to approximate all the real and complex eigenvalues of A, where M, the average number of iterations per eigenvalue, is unbounded both theoretically and practically. Other known practical iterative algorithms have no better computational cost estimates and no better performance [GL96]. All of them routinely fail in practice already for n of order 50 in the highly important case where some eigenvalues are multiple or form clusters. This observation is stated explicitly e.g. in [Go94], pp. 1062-1063, whereas the solution in the case of a much higher order of n and in the presence of clustered eigenvalues is usually required, say, in the cited applications to polynomial systems of equations.

The computational complexity estimates based on the customary algorithms are not satisfactory even for the much simpler *eigenspace problem*, which is a subproblem of the eigenproblem. Namely, in the eigenspace problem, one assumes all the eigenvalues of an $n \times n$ matrix available and seeks the eigenspaces of the eigenvectors associated with these eigenvalues. The solution method of practical choice is the Inverse Power Iteration, but it supports no finite complexity estimates for the worst case input (see Appendix B).

Similar deficiency is observed for the eigenproblem for the various highly important classes of special (sparse and dense structured, e.g. Toeplitz-like and Cauchy-like) matrices, for which one should expect substantial acceleration of the solution by exploiting the sparsity and/or structure. Here again, the full complexity analysis of the customary solution algorithms is missing.

The above background should have made the eigenproblem a major computer science subject in the areas of computational complexity and the design and analysis of algebraic algorithms. So far the situation is quite opposite, however, apart from the estimates for the complexity of some selected algorithms [KW92], the solution of the (simpler) symmetric eigenvalue problem in [BP91], [BP98] and some treatment of the sparse eigenvalue problem in [R95] (cf. Remark A.1 in our Appendix A).

In our present paper, we will study the arithmetic complexity issues for the eigenproblem for both general and special $n \times n$ input matrices, yielding the tight estimates (up to polylogarithmic factors).

1.2 Our results for general matrices

Let us next specify our results for general $n \times n$ matrices and compare them with the known ones. To solve the eigenproblem with controlled computational cost for an input matrix A possibly having clustered and/or multiple eigenvalues, we rely on the well-known observation that all the eigenvalues of an $n \times n$ matrix A are the zeros of its characteristic polynomial $c_A(x)$. This observation immediately enables us to reduce the solution of the eigenproblem to the following stages:

a) the reduction of the eigenproblem for a given matrix to the one for a matrix in a canonical form (that is, for a Frobenius matrix, for a block triangular or block diagonal matrix with Frobenius diagonal blocks or for a tridiagonal matrix); as a by-product of this stage, the coefficients of the characteristic polynomial $c_A(x)$ of the original input matrix A are output or become available readily, in $O(n)$ additional ops,

b) the approximation of the eigenvalues of A as the zeros of $c_A(x)$, and

c) the computation of the eigenspaces of the approximated eigenvalues, by exploiting the available reduction of A to the canonical form.

Stage a) has been thoroughly investigated already several decades ago [FF63]. Several well-known algorithms reduce a generic $n \times n$ matrix A to a Frobenius matrix, and we also prove in section 9 that an unsymmetric modification of Lanczos algorithm reduces it to the tridiagonal form. (Generic $n \times n$ matrices cover all $n \times n$ matrices but ones forming an algebraic variety of a lower dimension.) Furthermore, by combining and extending some known techniques, Keller-Gehrig in [K-G85] devised a solution algorithm that reduced any $n \times n$ input matrix A to the triangular

Frobenius form. All these algorithms are performed at nearly optimal arithmetic cost of $O(M(n)\log n)$ ops, whereas $\Omega(M(n))$ is a lower bound on the arithmetic complexity of the computation of $c_A(x)$ for a general input matrix (see [BP94], exercise 2.6 on p.213). Here and hereafter, $M(n)$ denotes the number of ops required for $n \times n$ matrix multiplication: $M(n) = O(n^\omega)$ where $\omega < 2.376$ in theory, and ω ranges from $\log_2 7 < 2.808$ to 3 in practical algorithms (cf. [BP94], p. 94; [GL96], [Hig96]).

At stage b) we fix some matrix norm $||\cdot||$ and absolute output error bound $\epsilon = 2^{-b}||A||$ and apply the recent nearly optimal polynomial rootfinders (see section 10) to approximate all the eigenvalues of A within ϵ in arithmetic time

$$t(n,b) = O((n\log^2 n)(\log b + \log^2 n)), \tag{1.1}$$

whereas n and $\Omega(\log b)$ are the known lower bounds on the complexity of polynomial rootfinding (cf. [R87] and [P97]).

For an $n \times n$ matrix A given with its eigenvalue λ, stage c) amounts to the computation of the null-space of the matrix $\lambda I_n - A$, I_n being the $n\times n$ identity matrix. The known algorithms compute the null-space in $O(M(n))$ ops (see [BP94], pp. 109-110), which gives us the bound of $O(nM(n)) = O(n^{\omega+1})$ ops for the entire eigenspace problem. Such a known record bound for the eigenspace problem also dominates the overall complexity of the eigenproblem. Our study in section 7, however, shows that $O(n^3)$ ops are sufficient to solve the eigenspace problem for an $n \times n$ matrix A given in the triangular Frobenius form whereas the cost decreases to $O(n^2)$ ops if A is a tridiagonal or Frobenius matrix or is represented in the diagonal Frobenius form. Summarizing, we obtain the following results:

Theorem 1.1 *The deterministic arithmetic complexity of the eigenproblem for any $n \times n$ matrix A is bounded by $O(n^3) + t(n,b)$ ops for $t(n,b)$ of (1.1) and for $2^{-b}||A||$ denoting the required upper bound on the absolute output error of the approximation of the eigenvalues of A where $||\cdot||$ denotes any fixed matrix norm. For generic $n \times n$ matrix A, the complexity is bounded by $O(M(n)\log n) + t(n,b)$ ops, where $M(n)$ denotes the complexity of $n \times n$ matrix multiplication, $M(n) = o(n^{2.376})$.*

Remark 1.1 *Clearly, the term $t(n,b)$ can be deleted from the former of the above estimates if $\log b = O(n^2/\log^2 n)$, that is, if $b = 2^{O(n^2/\log^2 n)}$, and from the latter of the estimates if $\log b = O(M(n)/(n\log n))$, that is , if $b = 2^{O(M(n)/(n\log n))}$.*

The complexity estimates of Theorem 1.1 substantially improve the known estimates due to the improvement from $O(nM(n))$ to $O(n^3)$ for any input matrix A or $O(n^2)$ for generic input matrix A at the bottleneck stage c) of the solution of the eigenspace problem.

1.3 Our results for sparse and dense structured matrices

As we will see in section 11, stage c) of the solution of the eigenproblem is much simplified for any $n \times n$ matrix A that satisfies the equation

$$m_A(x) = c_A(x), \tag{1.2}$$

where $m_A(x)$ denotes the minimum polynomial of A. Equation (1.2) holds generically for an $n \times n$ input matrix A, and if (1.2) holds, the stage a) becomes the bottleneck, that is, the cost of the computations at stage a) dominates the overall computational cost of the solution of the eigenproblem. In section 4, we will prove that the cost can be decreased dramatically in the case where the input matrix A is special, that is, sparse or dense structured.

Sparse matrices A are characterized by the patterns of their nonzero entries, whose total number will be denoted f_A. It is realistic to assume that

$$f_A = O(n^\beta), \tag{1.3}$$

with $\beta < 2$ or even $\beta = 1$ for an $n \times n$ sparse matrix A.

There are several important families of dense structured matrices, the most celebrated being the classes of *Toeplitz, Hankel, Vandermonde and Cauchy matrices,* which we display in the next table (cf. [OP98]).

Table 1. Dense Structured matrices.

Toeplitz matrices: $T = (t_{i-j})_{i,j=0}^{n-1}$	Hankel matrices: $H = (h_{i+j})_{i,j=0}^{n-1}$.
Vandermonde matrices: $V = (x_i^j)_{i,j=0}^{n-1}$	Cauchy matrices: $C = (\frac{1}{x_i - y_j})_{i,j=0}^{n-1}$

For the latter classical matrices, important extensions were more recently developed called *Toeplitz-like, Hankel-like, Vandermonde-like and Cauchy-like matrices* in [BP94]. In some applications, one deals with *Toeplitz-like + Hankel-like matrices* (obtained as the sum of pairs of Toeplitz-like and Hankel-like matrices and thus extending both of these two classes).

It is a basic property of sparse and dense structured matrices that they can be multiplied by vectors fast. Formally, let v_A denote the number of ops required to multiply a given $n \times n$ matrix A by a vector $\mathbf{v}$ and let $K_A(h)$ denote the number of ops required to compute the $n \times h$ Krylov matrix, $K(A, \mathbf{v}, h) = (A^i \mathbf{v})_{i=0,\dots,h-1}$, where $\mathbf{v}$ is a generic n-dimensional vector. We have

$$K_A(h) \le (h-1)v_A, \tag{1.4}$$

$$v_A = 2n^2 - n, \;\; K_A(h) \le (2n-1)(h-1)n, \tag{1.5}$$

$$v_A \le 2f_A, \;\; K_A(h) \le 2(h-1)f_A \tag{1.6}$$

for any $n \times n$ matrix A (where f_A is the number of nonzero entries of A),

$$v_A = O(n \log n), \;\; K_A(h) = O(hn \log n) \tag{1.7}$$

for any $n \times n$ Toeplitz, Hankel, Toeplitz-like, Hankel-like, or Toeplitz-like+Hankel-like matrix A, and

$$v_A = O(n \log^2 n), \;\; K_A(h) = O(hn \log^2 n), \tag{1.8}$$

for any $n \times n$ Vandermonde, Vandermonde-like, Cauchy, or Cauchy-like matrix A (cf. [BP94]).

We will prove that for every considered class of special matrices (which is either a linear subspace or an algebraic variety in the linear space of dimension n^2 of the entries of all $n \times n$ matrices) equation (1.2) holds for generic matrices of this class (that is, for all matrices of this class except for ones that form an algebraic variety of a lower dimension).

We will next state (in terms of n, b and $K_A(n)$) our estimates for the complexity of the eigenproblems of $n \times n$ sparse and dense structured matrices A satisfying (1.2).

Theorem 1.2 *If an $n \times n$ matrix A satisfies (1.2), then its eigenproblem can be solved by means of generating $4n-2$ random parameters and then performing $t(n,b)+O(n^2+K_A(n))$ ops for $t(n,b)$ of (1.1) and for $2^{-b}||A||$ denoting the required upper bound on the output errors for the eigenvalues, where $||\cdot||$ denotes any fixed matrix norm. The cost bound does not include the cost of the generation of random parameters. Assuming that these parameters are sampled from a fixed finite set S of cardinality $|S|$ independently of each other under the uniform probability distribution on S, the algorithm supporting the above arithmetic complexity estimate either outputs FAILURE or otherwise, with a probability at least $(1-(n+1)n/(2|S|))(1-2n/|S|)$, produces correct output for a matrix A satisfying (1.2). The algorithm can be applied to any $n \times n$ matrix A and outputs FAILURE unless (1.2) holds.*

Remark 1.2 *The term $t(n,b)$ can be deleted from the above complexity estimate if $\log b = O(n/\log^2 n)$, that is, if $b = 2^{O(n/\log^2 n)}$. The term n^2 can be deleted unless $K_A(n) = o(n^2)$.*

If we rely on bounds (1.4) and (1.5), Theorem 1.2 gives no improvement of the complexity estimates of Theorem 1.1. For sparse and dense structured matrices, we immediately obtain a major improvement based on (1.6) –(1.8).

Corollary 1.1 *The complexity estimate of Theorem 1.2 turns into $t(n,b)+O(n^2+nf_A)$ for $n \times n$ sparse matrices A having f_A nonzero entries (which gives us the bound $t(n,b)+O(n^2)$ if $f_A = O(n)$); the estimate turns into $t(n,b)+O(n^2\log n)$ for $n \times n$ Toeplitz-like+Hankel-like matrices A, and into $t(n,b)+O(n^2\log^2 n)$ for Cauchy-like and Vandermonde-like matrices A.*

Remark 1.3 *The latter bounds are optimal (up to polylogarithmic factors) because n^2 ops are required already to output the n^2 coordinates of the n eigenvectors of an $n \times n$ matrix.*

For comparison, the knows estimates for the complexity of the eigenproblem of the above highly important classes of special matrices do not improve the bound $O(nM(n))$=$O(n^{\omega+1})$ known for general matrices.

1.4 Organization of our paper

We organize our paper as follows. In the next section we introduce the matrix eigenproblem. In sections 3 and 4 we study its reduction to canonical forms together with the computation of the characteristic polynomial, $c_A(x)$. In sections 5–7, we define various classes of dense structured matrices and prove that equation (1.2) holds generically for the matrices of these classes as well as for sparse matrices. In section 8, we cover fast randomized solution of singular but consistent Toeplitz or Toeplitz-like linear systems of equations (over any field of constants), which is required in section 4. In section 9, we study the reduction of a matrix to the tridiagonal form by means of an unsymmetric variant of Lanczos algorithm and prove that this works for generic input. In section 10, we cover approximation of the eigenvalues as the zeros of $c_A(x)$ and the computation of their algebraic multiplicities. In section 11, we compute the eigenspaces of all the eigenvalues. In section 12, we briefly discuss some open issues, in particular the solution of the eigenproblem for the special matrices associated with polynomial systems and the extension of our arithmetic complexity results to the bit-complexity estimates. In Appendix A we recall some fast deterministic algorithms for the computation of the characteristic polynomial where A is sparse or is dense and structured. In Appendix B we comment on the Inverse Power Iteration for the computation of the eigenvectors.

The appendix was written by the first two authors, section 8 by the first and the third authors, whereas all other sections are due to the first author.

Acknowledgement Victor Pan thanks Luca Gemignani for pointing out reference [GIKT95].

2 Matrix Eigenproblem

In this section we recall some definitions and basic results (cf. [FF63], [GL96], [Wil65]).

Definition 2.1 *For an $n \times n$ matrix A, let*

$$AV = V\Lambda, \tag{2.1}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_q)$ is a $q \times q$ diagonal matrix, $V = (\mathbf{v}_1, \dots, \mathbf{v}_q)$ is an $n \times q$ matrix of full rank with columns $\mathbf{v}_1, \dots, \mathbf{v}_q$, and q is maximum, $q \leq n$. Then every pair $(\lambda_j, \mathbf{v}_j)$ is the pair of the eigenvalue *λ_j and the* eigenvector *$\mathbf{v}_j$ of A associated to each other, and the computation of such pairs for $j = 1, \dots, q$ is the* eigenproblem *for A. $\lambda_1, \dots, \lambda_q$ must not be all distinct. The number μ_j of times the eigenvalue λ_j is repeated in Λ is called the* geometric multiplicity *of λ_j. λ_j is* geometrically simple *if $\mu_j = 1$. μ_j is the dimension of the* eigenspace of the eigenvalue *λ_j, defined as the linear space of the eigenvectors associated to this eigenvalue.*

Remark 2.1 *One may scale the eigenvalues of A by scaling A because (2.1) implies that $hAV = V(h\Lambda)$ for any scalar h.*

Definition 2.2 *Hereafter,* I_k *denotes the* $k \times k$ identity matrix, 0 *denotes a* null matrix *of appropriate size,* $\det W$ *denotes the* determinant *of a matrix* W*;* W^T *and* W^H *denote the* transpose *and the* Hermitian transpose *of a matrix or a vector* W*, respectively.*

Definition 2.3 $c_A(x)$=$\det(xI_n - A)$ *is* the characteristic polynomial *of* A*.* $m_A(x)$*, the* minimum polynomial *of* A*, is the monic polynomial of the minimum degree such that* $m_A(A) = 0$*.*

Theorem 2.1 $c_A(A) = 0$*. Furthermore,* $m_A(x)$ *divides* $c_A(x)$*;* $m_A(\lambda_j) = c_A(\lambda_j) = 0$ *if and only if* λ_j *is an eigenvalue of* A*.*

Definition 2.4 μ_j^+*, the* algebraic multiplicity *of an eigenvalue* λ_j *of* A*, is the multiplicity of the zero* λ_j *of* $c_A(x)$*, that is,* $\mu_j^+ = \max\{d : (x - \lambda_j)^d$ *divides* $c_A(x)\}$*.*

We have

$$\mu_j^+ \geq \mu_j \text{ for all } j. \tag{2.2}$$

Theorem 2.2 *Let (2.1) hold, let* T *be an* $n \times n$ *nonsingular matrix, and let*

$$\hat{A} = TAT^{-1}. \tag{2.3}$$

Then $\hat{A}(TV) = (TV)\Lambda$*.*

Definition 2.5 *The matrix transformation (2.2) of* A *into* $\hat{A}$ *is called* similarity transformation*, which reduces the eigenproblem for* A *to one for* $\hat{A}$ *and is said to* reduce A *to* $\hat{A}$*.*

The initial customary step of the solution of the eigenproblem for a given matrix A is the reduction of A (by a similarity transformation) to some canonical form for which the solution of the eigenproblem is simpler. The next two definitions describe some of these canonical forms of matrices.

Definition 2.6 $A = (a_{i,j})$ *is a* tridiagonal matrix *if* $a_{i,j} = 0$ *for* $|i - j| > 1$*. More generally,* A *is a* Hessenberg matrix *if* $a_{i,j} = 0$ *for* $i - j > 1$*. A Hessenberg or tridiagonal matrix* $(a_{i.j})$ *is* unreduced *if* $a_{i+1,i} \neq 0$ *for all* i*.*

Remark 2.2 *Actually, we narrowed the class of Hessenberg matrices to the upper Hessenberg matrices and similarly for the matrices in the triangular Frobenius form defined below (in Definition 2.7) because this is sufficient for our eigenstudy.*

The following simple observations specify the structure of Hessenberg and tridiagonal matrices.

Theorem 2.3 *Every Hessenberg matrix A can be represented as a block triangular matrix with unreduced Hessenberg diagonal blocks. If A is tridiagonal, then so are these blocks too, and each superdiagonal block either vanishes or has a single nonzero entry, located in its southwestern corner.*

The next special case of Hessenberg matrices will be important for our study.

Definition 2.7 *A special $n \times n$ Hessenberg matrix of the form*

$$F = F(p) = \begin{pmatrix} 0 & \dots & & 0 & -p_0 \\ 1 & \ddots & & & \vdots \\ & \ddots & & & \vdots \\ & & \ddots & 0 & \vdots \\ 0 & & & 1 & -p_{n-1} \end{pmatrix}$$

is called a Frobenius (companion) matrix. *A Hessenberg matrix A is* in the triangular Frobenius form *if this is a block upper triangular matrix,*

$$A = \begin{pmatrix} F_1 & & & * \\ & \ddots & & \\ & & \ddots & \\ 0 & & & F_k \end{pmatrix}, \tag{2.4}$$

*whose diagonal blocks $F_1, \dots, F_k$ are Frobenius matrices. Here, all the subdiagonal blocks are null matrices, and * stays for the superdiagonal blocks. If all of them are null matrices, then the matrix is* in the diagonal Frobenius form.

Remark 2.3 *A Frobenius matrix F is available as soon as its characteristic polynomial $c_F(x) = x^n + \sum_{n=0}^{n-1} p_i x^i$ is available and vice versa.*

Remark 2.4 *For a matrix A of (2.4), we have*

$$c_A(x) = \prod_{i=1}^{k} c_{F_i}(x). \tag{2.5}$$

Hereafter, we will assume computations with real or complex numbers, will measure the computational cost by the number of arithmetic operations and comparisons involved, and will refer to all such operations as *ops*.

By applying Gaussian elimination with row interchange, we obtain the following results.

Theorem 2.4 *$O(n^2)$ ops suffice to factorize an $n \times n$ nonsingular Hessenberg matrix A as a product PLU where P is a permutation matrix, L and U are lower and upper triangular matrices, respectively. The cost bound decreases to $O(n)$ either if A is tridiagonal or if $A = \lambda I_n - F$ for a scalar λ and a matrix F given in the diagonal Frobenius form. If the PLU factorization of A has been computed, then $O(n)$ additional ops are sufficient to compute $\det A$ and to decide whether A is singular or not.*

Corollary 2.1 *$O(n^2)$ ops suffice to solve a nonsingular linear system of equations with a Hessenberg coefficient matrix A. The cost bound decreases to $O(n)$ if A is a tridiagonal matrix or equals $\lambda I_n - F$ where F is a matrix in the diagonal Frobenius form and λ is a scalar.*

3 Reduction of a Matrix to the Triangular and Diagonal Frobenius Forms

We will start this section with some definitions partly repeating ones of the introduction.

Definition 3.1 *Hereafter, $M(n)$ and $I(n)$ denote the arithmetic cost of $n \times n$ matrix multiplication and inversion, respectively.*

It is well-known (see e.g. [BP94], p. 213) that

$$M(n) = O(n^{\omega}),\ 2 \leq \omega < 2.376, \tag{3.1}$$

$$M(n) = O(I(n)),\ I(n) = O(M(n)).$$

Remark 3.1 *(cf. [BP94], p. 94). The asymptotic upper bound of (3.1) hides a huge overhead constant in the "O" notation. Practical customary algorithms support the bounds $M(n) = O(n^{2.808})$ and $I(n) = O(n^3)$ where the overhead constants are quite small. Estimating the computational cost in terms of $M(n)$, however, has some practical justification because matrix multiplications and the computations intensively involving them are performed very effectively on modern supercomputers [GL96], [Q94].*

Theorem 3.1 *Given an $n \times n$ matrix A, it suffices to use $O(M(n) \log n)$ ops in order to compute a nonsingular matrix W, its inverse W^{-1}, and the matrix WAW^{-1} in the triangular Frobenius form.*

Proof. See [K-G85]. □

Keller-Gehrig's algorithm of [K-G85], supporting Theorem 3.1, involves several steps of modified block Gaussian elimination, which makes it generally ineffective for sparse and/or structured matrices A.

We also recall the following results.

Theorem 3.2 *(see [GL96], p.314). For a pair of $n \times n$ matrices $A = \begin{pmatrix} B & C \\ 0 & G \end{pmatrix}$ and $W = \begin{pmatrix} I_\rho & X \\ 0 & I_{n-\rho} \end{pmatrix}$, where B is a $\rho \times \rho$ matrix, we have $W^{-1}AW = \begin{pmatrix} B & C - XG + BX \\ 0 & G \end{pmatrix}$.*

Theorem 3.3 *[GIKT95], pp. 383-384.) Under the assumptions of Theorem 3.2, let B be a Frobenius matrix. Then there exists a matrix X satisfying $XG - BX = C$.*

Based on Theorems 3.2 and 3.3, we may reduce an $n \times n$ matrix A from the triangular Frobenius to the diagonal Frobenius form. The estimated cost of such a reduction is dominated by the cost estimates for the stage of the computation of the matrices X of these theorems. The latter cost estimates generally exceed $M(n)$ by order of magnitude, even where B and G are Frobenius matrices.

4 Randomized Reduction to a Frobenius Matrix

We will again start this section with definitions, partly repeating ones of the introduction.

Definition 4.1 *For an $n \times n$ matrix A and an n-dimensional vector $\mathbf{v}$, the $n \times n$ matrices $K(A, \mathbf{v}, h) = (A^i \mathbf{v})_{i=0,\ldots,h-1}$ for $h = 1, 2, \ldots,$ are called* Krylov matrices *induced by A and $\mathbf{v}$.*

Definition 4.2 *For an $n \times n$ matrix A and a positive integer h, let $K_A(h)$ denote the arithmetic complexity of the computation of the Krylov matrix $K(A, \mathbf{v}, h)$ maximized over all n-dimensional vectors $\mathbf{v}$.*

Definition 4.3 *A subset W of a linear space V or, more generally, of an algebraic variety V of a dimension k is algebraically open or* generic *if $V - W$ is an algebraic variety of dimension less than k. A property holds* generically *for a fixed class of matrices (or, equivalently, for a* generic *matrix of a fixed class) if it holds generically for the matrices of this class.*

Definition 4.4 Uniform sampling *of h elements from a finite set S is their selection from S independently of each other under the uniform probability distribution on S. $|S|$ is the cardinality of a set S.*

We have the following well-known results (cf. e.g [Wi86], [KP91], [BP94] on Theorem 4.1 and [GL96], pp. 348-349, on Theorem 4.2).

Theorem 4.1 *$m_A(x) = c_A(x)$ (that is, equation (1.2) holds) for an $n \times n$ matrix A if and only if*

$$\text{rank } (K(A, \mathbf{v}, n)) = n \tag{4.1}$$

or, equivalently, if and only if the $n \times n$ matrix $H_n = (\mathbf{u}^T A^{i+j} \mathbf{v})_{i,j=0}^{n-1}$ is nonsingular where the vectors $\mathbf{u}$ and $\mathbf{v}$ (both of dimension n) have $2n$ indeterminates as their coordinates.

Corollary 4.1 *Equation (4.1) (for a fixed vector $\mathbf{v}$) holds generically for $n \times n$ matrices. Equation (1.2) holds generically for $n \times n$ matrices A.*

Remark 4.1 *Corollary 4.1 can be deduced alternatively from the well-known fact that equation (1.2) holds if an $n \times n$ matrix A has n distinct eigenvalues and from Theorem 2.1.*

Theorem 4.2 *Let A be an $n \times n$ matrix. $\mathbf{v}$. Let equations (1.2), (4.1) hold. Then*

$$F = K^{-1}AK, \ \ for \ K = K(A, \mathbf{v}, n), \tag{4.2}$$

is the $n \times n$ Frobenius matrix satisfying the equation

$$c_F(x) = c_A(x). \tag{4.3}$$

Theorem 4.2 for $\mathbf{v} = (1, 0, \dots, 0)^T$ leads to the Danilevski algorithm which under (4.1) computes a similarity transformation of a matrix A into the Frobenius matrix F and outputs $c_F(x) = c_A(x)$ and the matrix K of (4.2). The theorem has the following corollary, involving $M(n), I(n)$ and $K_A(n)$ of Definitions 3.1 and 4.2.

Corollary 4.2 *Let an $n \times n$ matrix A and a vector $\mathbf{v}$ satisfy equations (1.2), (4.1). Then the Frobenius matrix F and the Krylov matrix $K = K(A, \mathbf{v}, n)$ satisfying equation (4.2) can be computed at the cost $K_A(n) + 2M(n) + I(n)$.*

Now we observe that for the solution of the eigenproblem we need only the matrices F and K of (4.2) but not K^{-1}. This motivates our next goal to remove the terms $2M(n)$ and $I(n)$ from the cost bound of Corollary 4.1. Such a removal will dramatically accelerate the computations in the case where A is a sparse and/or structured matrix (cf. (1.6)-(1.8)). Towards this goal, we first recall the next theorem, fundamental for the study of randomized algebraic algorithms. The theorem is due to [DL78] but is more widely known from the papers [Z79] and [S80].

Theorem 4.3 *Let $p(x) = p(x_1, \dots, x_m)$ be a nonzero m-variate polynomial over and field or ring R. Let $p(x)$ have total degree d. Let $x_1^* \dots, x_m^*$ be uniformly sampled from a fixed finite subset S of R. Then $p(x_1^* \dots, x_m^*) \neq 0$ with a probability at least $1 - d/|S|$.*

The algorithm supporting the next theorem probabilistically computes the minimum polynomial $m_A(x)$ of a given matrix A.

Theorem 4.4 *Let A be a fixed $n \times n$ matrix, let S be a fixed finite set and let $\mathbf{v}$ be a random vector of dimension n with its entries uniformly sampled from the set S. Then it suffices to sample uniformly $3n - 2$ additional random values from S and to perform $K_A(2n) + O(n^2)$ ops in order to compute the matrix $K(A, \mathbf{v}, n)$ and a monic divisor $d(x)$ of the polynomial $m_A(x)$ such that $d(x) = m_A(x)$ with a probability at least $(1 - \frac{(n+1)n}{2|S|})(1 - \frac{2n}{|S|})$.*

Proof. The modification by [KP91] of the algorithm of [Wi86] uniformly samples $2n$ random parameters from a finite set S, to define the $2n$ components of two vectors $\mathbf{u}$ and $\mathbf{v}$, each of dimension n. Then (cf. [KP91]), at the cost of performing $K_A(2n)+O(n^2)$ ops, the $n \times (2n)$ matrix $K(A, \mathbf{v}, 2n)$ and the vector $\mathbf{h} = \mathbf{u}^T K(A, \mathbf{v}, n) = (\mathbf{u}^T A^i \mathbf{v})_{i=0}^{2n-1}$ are computed, which defines the $k \times k$ Hankel matrices,

$$H_k = (\mathbf{u}^T A^{i+j} \mathbf{v})_{i,j=0}^{k-1}, \quad k = 1, \ldots, n, \tag{4.4}$$

satisfying

$$\det H_\rho \neq 0, \text{ for } \rho = \operatorname{rank}(H_n). \tag{4.5}$$

The coefficient vector $\mathbf{d} = (d_i)_{i=0}^{r-1}$ of the polynomial $d(x) = \sum_{i=0}^{r-1} d_i x^i$ is computed as the solution of the following nonsingular linear system of ρ equations:

$$H_\rho \mathbf{d} = (\mathbf{u}^T A^i \mathbf{v})_{i=\rho}^{2\rho-1}. \tag{4.6}$$

It is immediately observed that the polynomial $d(x)$ divides $m_A(x)$; furthermore, it is proved in [KP91] (based on Theorem 4.3) that $d(x)$ equals the minimum polynomial $m_A(x)$ with a probability at least $1 - \frac{2n}{|S|}$. It remains to compute ρ and to solve the linear system (4.6). This task will be solved in section 8, where we will recall and slightly elaborate further the algorithm of [Kal95]. The solution requires to sample $2n-2$ additional random values uniformly from the set S and in addition to perform $O(n \log^2 n)$ ops. This will complete the proof of Theorem 4.4. □

Now, for an $n \times n$ matrix A we immediately devise the following randomized algorithm, which tests if (4.1) holds and, if so, computes two matrices, F (Frobenius) and K (Krylov) of (4.2), (4.3), and also outputs $c_A(x)$ (cf. Remark 2.3).

Algorithm 4.1, *reduction of a generic matrix to the Frobenius matrix.*

Input: an $n \times n$ matrix A and a finite set S.

Output: FAILURE or an $n \times n$ nonsingular matrix K and the Frobenius matrix F satisfying equations (4.2), (4.3), as well as the characteristic polynomial $c_A(x) = c_F(x)$.

Computations:

1. Apply the randomized algorithm supporting Theorem 4.4 to compute a matrix $K = K(A, \mathbf{v}, n)$ (for a random vector $\mathbf{v}$) and a divisor $d(x)$ of $m_A(x)$.

2. If the degree of $d(x)$ is n, output $d(x)$, the matrix K and the Frobenius matrix F satisfying $c_F(x) = d(x)$. Otherwise output FAILURE.

Correctness of the algorithm follows from Theorems 4.2 and 4.4.

By summarizing the estimates for the computational cost and the failure probability of the algorithm, we obtain the following result.

Theorem 4.5 *If Algorithm 4.1 is applied to an $n \times n$ matrix A not satisfying (1.2), it always outputs FAILURE. If the algorithm is applied to an $n \times n$ matrix A satisfying (1.2), then it either outputs FAILURE or otherwise, with a probability at least $(1 - \frac{(n+1)n}{2|S|})(1 - \frac{2n}{|S|})$, correctly computes a pair of $n \times n$ matrices K and F (K a nonsingular*

Krylov matrix, F a Frobenius matrix) satisfying (4.2), (4.3), so that F defines the characteristic polynomial $c_A(x) = c_F(x)$. The algorithm samples uniformly $4n-2$ random values from S and then performs $K_A(2n) + O(n^2)$ ops (not including the cost of the random sampling).

Remark 4.2 *As soon as the matrices K and F of (4.2) are available, Theorem 2.2 reduces the eigenproblem of A to the one of F.*

Remark 4.3 *The complexity bound of Theorem 4.5 is smaller than one of Theorem 3.1 by order of magnitude if the input matrix A is sparse and/or structured (cf. (1.6)–(1.8)).*

5 Equation (1.2) for General, Vandermonde-like, Cauchy-like, and Sparse Matrices

By Corollary 4.1, equation (1.2) holds generically for $n \times n$ matrices and also for the matrices of any subspace or algebraic variety in the space of $n \times n$ matrices provided that it holds for at least one matrix A in this subspace or variety. To prove the existence of such a matrix A, we will use, in particular, the following definition and result.

Definition 5.1 *An $l \times l$* leading principal *or* northwestern *submatrix of an $n \times n$ matrix W is the submatrix formed by the entries lying in the first l columns and first l rows of W, for $l \leq n$. If the $k \times k$ leading principal (northwestern) submatrices of a matrix W of rank r are nonsingular for $k = 1, \ldots, r$, then W is said to have* generic rank profile. *A matrix is* strongly nonsingular *if it is nonsingular and has generic rank profile.*

Theorem 5.1 *[Wi86]. Equation (1.2) holds for an $n \times n$ matrix $A = WY$, provided that Y is the $n \times n$ diagonal matrix with indeterminates filling its diagonal and that the matrix W has generic rank profile.*

Let us next show that equation (1.2) holds generically for Vandermonde-like matrices and Cauchy-like matrices. We will first recall their definitions (cf. [BP94], [GO94]).

Definition 5.2 *Let $x_0, \ldots, x_{n-1}$ denote n distinct values and let $s_0, t_0, \ldots, s_{n-1}, t_{n-1}$ denote $2n$ distinct values. Let $\mathbf{x} = (x_i)_{i=0}^{n-1}$, $\mathbf{s} = (s_i)_{i=0}^{n-1}$, $\mathbf{t} = (t_j)_{j=0}^{n-1}$. Then the $n \times n$ matrices $V = V(\mathbf{x}) = (x_i^j)_{i,j=0}^{n-1}$ and $C = C(\mathbf{s}, \mathbf{t}) = (\frac{1}{s_i - t_j})_{i,j=0}^{n-1}$ are called* Vandermonde *and* Cauchy matrices, *respectively. Furthermore, the $n \times n$ matrices $V_f(\mathbf{x}, G, H)$ and $C(\mathbf{s}, \mathbf{t}, G, H)$ are said to be given with their $(f, \mathbf{x})$*-scaling/displacement generator *(G, H) and $(\mathbf{s}, \mathbf{t})$*-scaling generator *(G, H), both* of length *r, respectively, if*

$$V_f(\mathbf{x}, G, H) = \sum_{k=1}^{r} \text{ diag } (f g_i^{(k)}/(f - x_i^n))_{i=0}^{n-1} V(\mathbf{x}) Z_{1/f}^T(\mathbf{h}^{(k)}),$$

$$C(\mathbf{s}, \mathbf{t}, G, H) = \sum_{k=1}^{r} \text{ diag } (s_i g_i^{(k)})_{i=0}^{n-1} C(\mathbf{s}, \mathbf{t}) \text{ diag } (h_j^{(k)})_{j=0}^{n-1}$$

where

$$\mathbf{g}^{(k)} = (g_i^{(k)})_{i=0}^{n-1}, \mathbf{h}^{(k)} = (h_j^{(k)})_{j=0}^{n-1}, k = 0, \ldots, n-1, G = (\mathbf{g}^{(k)})_{k=0}^{n-1}, H = (\mathbf{h}^{(k)})_{k=0}^{n-1},$$

$\text{diag}(\mathbf{w}_i)_{i=0}^{n-1}$ *denotes the diagonal matrix with the diagonal entries* $w_0, \ldots, w_{n-1}$*;* $Z_q(\mathbf{y}) = (y_{i,j})_{i,j=0}^{n-1}$ *is the* q-circulant matrix, $y_{i,j} = y_{i-j \bmod n}$ *for* $i \geq j$*,* $y_{i,j} = q y_{i-j \bmod n}$ *for* $i < j$*,* $\mathbf{y} = (y_k)_{k=0}^{n-1}$*,* $q = 1/f$*, and* f *is a scalar,* $f \neq 0$*,* $f \neq x_i^n$ *for* $i = 1, \ldots, n-1$*. 1-circulant matrices are called* circulant*. The minimum length* r *in the above representations of the matrices* $V_f(\mathbf{x}, G, H)$ *and* $C(\mathbf{s}, \mathbf{t}, G, H)$ *(for all pairs* $(f, \mathbf{x})$ *and* $(\mathbf{s}, \mathbf{t})$*, respectively) is called the* scaling/displacement rank *and the* scaling rank *of these matrices, respectively. If* $r = O(1)$ *as* $n \to \infty$*, then the above matrices are called* Vandermonde-like *and* Cauchy-like matrices*, respectively.*

Note that Vandermonde and Cauchy matrices have scaling/displacement and scaling ranks 1, respectively.

We immediately observe the following results.

Theorem 5.2 *In the* n^2*-dimensional linear space of the entries of* $n \times n$ *matrices, the entries of the* $n \times n$ *matrices having scaling/displacement or scaling ranks at most* r *form two algebraic varieties of dimensions at most* $(2r+1)n+1$ *and* $(2r+2)n$*, respectively.*

It is well known that all $k \times k$ Vandermonde and Cauchy matrices $V = (x_i^j)_{i,j=0}^{k-1}$ and $C = (\frac{1}{s_i - t_j})_{i,j=0}^{k-1}$ have generic rank profile (provided that all the values $x_0, \ldots, x_{k-1}$ are distinct and all the values $s_0, t_0, \ldots, s_{k-1}, t_{k-1}$ are distinct). By Theorem 8.1, equation (1.2) holds for the matrices $A = VY$ and $A = CY$. By Definition 5.1, these are Vandermonde-like and Cauchy-like matrices, respectively, defined with their displacement/scaling generators of length 1 and scaling generators of length 1, respectively. By combining these observations, we obtain the next result.

Theorem 5.3 *Equation (1.2) holds generically for (Vandermonde-like matrices) having displacement/scaling rank at most* r *and for (Cauchy-like) matrices having scaling rank at most* r*, for any* $r \geq 1$*.*

We also have the next trivial corollary of Theorem 5.1.

Corollary 5.1 *Equation (1.2) holds generically for* $n \times n$ *sparse matrices with a fixed pattern of their nonzero entries as long as this pattern allows a matrix to have generic rank profile.*

6 Toeplitz and Toeplitz-like Matrices: Definitions, a Basic Property and Equation (1.2)

Our study of Toeplitz, Toeplitz-like, Hankel, and Hankel-like matrices throughout section 8 will be valid over any field $\mathbf{F}$, though for our study of the eigenproblem we only need the special case where $\mathbf{F} = \mathbf{C}$ and $\mathbf{C}$ denotes the field of complex numbers. Frobenius matrices are not involved in this and the next sections, and we will use capital F to denote the linear operators of shift (displacement).

Definition 6.1 *A matrix $T = (t_{i,j})$ is called a* Toeplitz matrix *if*

$$t_{i+1,j+1} = t_{i,j} \tag{6.1}$$

for all pairs i, j for which $t_{i,j}$ and $t_{i+1,j+1}$ are defined.

Toeplitz matrices are a special class (where $r \leq 2$) of the following important and well-studied class of Toeplitz-like matrices having a small displacement rank r (cf. [BP94], pp. 174-211).

Definition 6.2 *(cf. e.g. [BP94], Definition 11.1). For an $n \times n$ matrix T, define the two displacement generators,*

$$F_-(T) = T - Z^T T Z, \quad F_+(T) = T - Z T Z^T, \tag{6.2}$$

where

$$Z = \begin{pmatrix} 0 & & & & 0 \\ 1 & 0 & & & \\ & 1 & \ddots & & \\ & & \ddots & & \\ 0 & & & 1 & 0 \end{pmatrix}$$

is a down-shift $n \times n$ matrix. If for $F = F_+$ or $F = F_-$ we have

$$F(T) = GH^T, \tag{6.3}$$

*where G and H are $n \times r$ matrices, then the pair of matrices (G, H) is called an F-*generator *or a* displacement generator *of T of length r and will be denoted $d.g._r(T)$. The minimum r allowing the above representation (6.3) is called the F-*rank *or* displacement rank *of T. T is called a* Toeplitz-like matrix *if $r = O(1)$ as $n \to \infty$.*

Next, we will recall a basic property of Toeplitz-like matrices, which will enable us to manipulate with them by means of operating with a few entries of their short displacement generators, rather than with their own more numerous entries.

Theorem 6.1 *[KKM79]. Let F_-, F_+, T, G, H, and r be as in (6.2) and (6.3). Then $F(T) = GH^T = \sum_{k=1}^r \mathbf{g}^{(k)}\mathbf{h}^{(k)T}$ if and only if*

$$T = \sum_{k=1}^{r} L^T(\mathbf{g}^{(k)})L(\mathbf{h}^{(k)}) \ for\ F = F_-,\ T = \sum_{k=1}^{r} L(\mathbf{g}^{(k)})L^T(\mathbf{h}^{(k)}) \ for\ F = F_+, \quad (6.4)$$

where $G = (\mathbf{g}^{(1)}, \ldots, \mathbf{g}^{(r)})$, $H = (\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(r)})$, and $L(\mathbf{v})$ is a lower triangular Toeplitz matrix with the first column $\mathbf{v}$.

Based on equations (6.1)-(6.4), we immediately obtain the following results.

Theorem 6.2 *The class of all $n \times n$ Toeplitz matrices forms a linear subspace T of dimension $2n - 1$ in the space of $n \times n$ matrices.*

Theorem 6.3 *The class of all $n \times n$ (Toeplitz-like) matrices having displacement rank at most r forms an algebraic variety T_r of dimension at most $2rn$ in the space of $n \times n$ matrices. Furthermore, $T \subset T_r$ for $r \geq 2$.*

Now we immediately verify that (1.2) holds for the Toeplitz matrix $A = Z$. Combining the latter observation with Theorems 6.2, 6.3 and Corollary 4.1 gives us the next result.

Corollary 6.1 *Equation (1.2) holds generically for $n \times n$ lower (upper) triangular Toeplitz matrices and $n \times n$ Toeplitz matrices as well as for $n \times n$ (Toeplitz-like) matrices of displacement rank at most r for any $r \geq 1$.*

Theorem 6.4 *Equation (1.2) holds generically for $n \times n$ circulant matrices A (which are a special case of Toeplitz matrices, cf. Definition 5.2).*

Proof. Recall a simple explicit expression of the eigenvalues of a circulant matrix via the entries (cf. e.g. [BP94], Theorem 5.1 on p.134). Based on this expression, define a circulant matrix having n distinct eigenvalues and recall Remark 4.1. $\square$

7 Hankel, Hankel-like, and Hankel-like+Toeplitz-like Matrices: Definitions and Equation (1.2)

Hankel matrices and *Hankel-like matrices* of displacement rank r are obtained from Toeplitz matrices and Toeplitz-like matrices of displacement rank r, respectively, by their pre-multiplication (or post-multiplication) by the *reflection matrix* J having ones on its antidiagonal and zero entries elsewhere. (Note that J^2 is the identity matrix.)

By Theorems 6.2 and 6.3, *the entries of $n \times n$ Hankel matrices $H = (h_{i+j})_{i,j=0}^{k-1}$ form a linear space of dimension $2n-1$, whereas Hankel-like matrices of displacement rank at most r form an algebraic variety of dimension at most $2rn$ in $\mathbf{F}^{n \times n}$.*

Now we recall the following result from [KS91].

Theorem 7.1 *Let an $n \times n$ nonsingular matrix W have generic rank profile. Let U^T and L be generic lower triangular $n \times n$ Toeplitz matrices, and let U have its diagonal entries equal to 1. Then the matrix $A = UWL$ satisfies equation (1.2).*

By applying Theorem 7.1 to the (triangular) Hankel matrix $W = J(I + Z)$, we obtain that (1.2) holds for the triangular Hankel matrix $A = UJ(I + Z)L$. We note that A has displacement rank 1 because JA is a lower triangular Toeplitz matrix.

Corollary 7.1 *Equation (1.2) holds generically for $n \times n$ lower (upper) triangular Hankel matrices and for $n \times n$ Hankel matrices as well as for $n \times n$ (Hankel-like) matrices of displacement rank at most r for any $r \geq 1$.*

Clearly, Corollary 7.1 can be immediately extended to the class of Toeplitz-like +Hankel-like matrices.

8 Fast Randomized Solution of a Consistent Toeplitz-like Linear System of Equations over an Arbitrary Field

In this section we will follow the appendix of the paper [Kal95] and slightly strengthen its results (cf. Remark 8.3). We will first recall the algorithm that supports the next theorem and then will extend this algorithm to complete the proof of Theorem 4.4.

Theorem 8.1 *Let A be an $n \times n$ matrix having generic rank profile (cf. Definition 5.1) and represented as*

$$A = \sum_{k=1}^{r} L^T(\mathbf{g}^{(k)})L(\mathbf{h}^{(k)}). \tag{8.1}$$

Then $O(n \log^2 n)$ ops suffice to compute $\rho = \operatorname{rank}(A)$ and $2r$ vectors $\mathbf{u}^{(1)}, \mathbf{v}^{(1)}, \ldots, \mathbf{u}^{(r)}, \mathbf{v}^{(r)}$ of dimension n defining an F_+-generator of A_ρ^{-1}, that is, satisfying the matrix equation $A_\rho^{-1} = \sum_{k=1}^{r} L(\mathbf{u}^{(k)})L^T(\mathbf{v}^{(k)})$, where A_ρ is the $\rho \times \rho$ northwestern (leading principal) submatrix of A, $A_\rho = A^{-1}$ if A is (strongly) nonsingular.

Remark 8.1 *For the cost estimates of Theorem 8.1, we assumed computations over the complex field, but the algorithms supporting the theorem can be applied over any field of constants with the standard minor adjustment of the cost bounds.*

We will start with some definitions and auxiliary results.

Definition 8.1 *(cf. [Str69], [BP94], p. 99). Let*

$$A = \begin{pmatrix} B & C \\ E & K \end{pmatrix}$$

be a nonsingular $n \times n$ matrix, with the nonsingular $k \times k$ submatrix B, $k = \lceil \frac{n}{2} \rceil$. Then the balanced triangular factorization (BTF) *of A and A^{-1} is given by the next matrix equations:*

$$A = \begin{pmatrix} I_k & O \\ EB^{-1} & I_{n-k} \end{pmatrix} \begin{pmatrix} B & O \\ O & S \end{pmatrix} \begin{pmatrix} I_k & B^{-1}C \\ O & I_{n-k} \end{pmatrix}, \tag{8.2}$$

$$A^{-1} = \begin{pmatrix} I_k & -B^{-1}C \\ O & I_{n-k} \end{pmatrix} \begin{pmatrix} B^{-1} & O \\ O & S^{-1} \end{pmatrix} \begin{pmatrix} I_k & O \\ -EB^{-1} & I_{n-k} \end{pmatrix}, \tag{8.3}$$

where

$$S = K - EB^{-1}C \tag{8.4}$$

is the Schur complement of B *in* A. *The BTF of A and A^{-1} can be extended to B, S, B^{-1} and S^{-1} and recursively to their submatrices and the Schur complements as long as we deal with nonempty nonsingular matrices. If nonsingularity holds throughout all recursive steps, then we arrive at the* BRTF *(*balanced recursive triangular factorization*) of A and A^{-1}.*

The BRTF of a matrix can be obtained by the Gauss-Jordan elimination algorithm, and we have the following simple results.

Theorem 8.2 *If A and B are nonsingular, then S^{-1} is the southeastern (trailing principal) submatrix of A^{-1}.*

Proof. Theorem immediately follows from (8.3). □

Theorem 8.3 *(cf. [BP94]), exercise 4 on p.212). A matrix A has BRTF if it is strongly nonsingular (cf. Definition 5.1).*

For a matrix A having generic rank profile, the BRTF has a natural extension, which can be efficiently computed by a divide-and-conquer algorithm We are going to describe this algorithm assuming that A is a Toeplitz-like matrix. We will start with some auxiliary results.

Theorem 8.4 *[BA80]. For any $n \times n$ matrix A,*

$$\text{rank}(F_-(A)) - 2 \leq \text{rank}(F_+(A)) \leq \text{rank} F_-(A) + 2.$$

Furthermore, given a $d.g._r(T)$ under $F = F_+$ (resp. $F = F_-$), it suffices to use $O(r)$ ops in order to compute a $d.g._{r+2}(T)$ under $F = F_-$ (resp. $F = F_+$).

Theorem 8.5 (cf. e.g. [BP94], Corollary 12.1 *). Let T_1 and T_2 be two Toeplitz-like matrices, given with their F-generators of lengths r_1 and r_2, respectively, for $F = F_+$ or $F = F_-$. Then an F-generator of length at most $r_1 + r_2 + 1$ for the matrix T_1T_2 can be computed by using $O(r_1 r_2)$ multiplications of polynomials of degree $O(n)$ and $O(r_1 + r_2)$ summations of $O(r_1 + r_2)$ vectors of dimension n, at the overall cost of performing $O((r_1 + r_2)^2 n \log n)$ ops. Furthermore, a d.g.$_r(UAL)$ for a given d.g.$_r(A)$ and a given pair of lower triangular Toeplitz matrices L and U^T can be computed at the cost $O(r^2 n \log n)$, provided that $F = F_-$.*

Theorem 8.6 *(cf. Proposition A.6 of [P92], [P93], or [BP94], Problem 2.2.11b). Given an d.g.$_{\hat{r}}(A) = (G, H)$ and the displacement rank r of A, $r < \hat{r} \leq n$, one can compute d.g.$_r(A)$ by using $O(\hat{r}^2 n)$ ops.*

Theorem 8.7 *[KKM79]. Let A be a nonsingular matrix. Then we have* rank $(F_+(A^{-1})) =$ rank $(F_-(A))$.

Theorem 8.8 *(cf. [M80], [BA80], [BP94]). Let A be an $n \times n$ strongly nonsingular Toeplitz-like matrix such that*

$$A = \begin{pmatrix} B & C \\ E & K \end{pmatrix}, \qquad S = K - EB^{-1}C,$$

B is a $k \times k$ matrix, and S is the $(n-k) \times (n-k)$ Schur complement of B in A. Let $r = \text{rank}(F_+(A))$. Then

$$\text{rank } (F_-(S^{-1})) = \text{rank } (F_+(S)) \leq r,$$

$$\text{rank } (F_-(B^{-1})) = \text{rank } (F_+(B)) \leq r,$$

$$\text{rank } (F_+(S^{-1})) = \text{rank } (F_-(S)) \leq r + 2,$$

$$\text{rank } (F_+(B^{-1})) = \text{rank } (F_-(B)) \leq r + 2.$$

Proof. Theorem 8.6 follows from Theorems 8.2, 8.4 and 8.7. □

(We slightly abuse the notation when we write S, but the Schur complement S and a finite set S will never appear together in this paper and thus will not interfere with each other.)

Given $d.g_r.(A)$ and two lower triangular Toeplitz matrices U^T and L, we may compute $d.g._r(\tilde{A})$ for $\tilde{A} = UAL$ and $F = F_-$ at the cost of performing at most $O(r^2 n \log n)$ ops due to Theorem 8.5.

Now, we will describe an algorithm supporting Theorem 8.1.

Algorithm 8.1, *computation of the largest northwestern (leading principal) inverse.*

Input: n-dimensional vectors $\mathbf{g}^{(1)}, \mathbf{h}^{(1)}, \ldots, \mathbf{g}^{(r)}, \mathbf{h}^{(r)}$ such that a Toeplitz-like matrix

$$A = \sum_{i=1}^{r} L^T(\mathbf{g}^{(i)}) L(\mathbf{h}^{(i)})$$

has generic rank profile.
Output: An integer $\rho \leq n$ and n-dimensional vectors $\mathbf{u}^{(1)},\ \mathbf{v}^{(1)},\ldots,\mathbf{u}^{(r)},\ \mathbf{v}^{(r)}$, such that $\rho = \text{rank}(A)$ and

$$A_\rho^{-1} = \sum_{m=1}^{r} L(\mathbf{u}^{(m)})L^T(\mathbf{v}^{(m)}),$$

where A_ρ denote the $\rho\times\rho$ northwestern submatrix of A.

1. Represent A as a 2×2 block matrix $\begin{pmatrix} B & C \\ E & K \end{pmatrix}$, for $k = \lceil \frac{n}{2} \rceil$, where the $k\times k$ submatrix B of A is singular if and only if $k > \rho$ (since A has generic rank profile). Apply Algorithm 8.1 recursively to the input matrix B replacing A. (Note that the first k components of all the given vectors $\mathbf{g}^{(i)}$ and $\mathbf{h}^{(i)}$ (for all i) define a $d.g._r(B)$.) If $\rho \geq k$, the output of this stage is the desired output of the algorithm. Otherwise, the matrix B is nonsingular, and then we obtain a $d.g._{r+2}(B^{-1})$ for $F = F_-$ and a $d.g._r(B^{-1})$ for $F = F_+$.

2. Apply Theorem 8.5 to compute a $d.g._r(S)$ for the matrix $S = K - E\ B^{-1}C$ and for $F = F_+$.

3. Apply the algorithm recursively to the Toeplitz-like input matrix S, replacing A. Output $\rho = \text{rank}(A) = k + \text{rank}(S)$.

4. By using Theorems 6.1, 8.4-8.8, compute $s.g._r(A_\rho^{-1})$ for $F = F_+$ (see some further comments below).

Let us specify stage 4. Consider A_ρ, the $\rho\times\rho$ northwestern (leading principal) submatrix of A, $A_\rho = \begin{pmatrix} B & V \\ D & R \end{pmatrix}$. Write $\check{S} = R - D\ B^{-1}V$. Note that at the preceding stages we have computed $d.g._r(V)$ and $d.g._r(D)$ for $F = F_-$, $d.g._r(B^{-1})$, $d.g._{2r+1}(-B^{-1}V)$, $d.g._{2r+1}(-DB^{-1})$, and $d.g._r(\check{S}^{-1})$ for $F = F_+$. We obtain the following block representation:

$$A_\rho^{-1} = \begin{pmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & \check{S}^{-1} \end{pmatrix},$$

where $M_{1,2} = -B^{-1}V\ \check{S}^{-1}$, $M_{2,1} = -\check{S}^{-1}D\ B^{-1}$, $M_{1,1} = B^{-1} - M_{1,2}DB^{-1}$. By applying Theorems 6.1, 8.4-8.8 we compute $d.g._r(A_\rho^{-1})$ for $F = F_+$. □

Combining (1.7) with Algorithm 8.1 and Theorems 6.1, 8.2-8.8, we immediately obtain Theorem 8.1. □

Furthermore, Algorithm 8.1 is immediately extended to the solution of a consistent linear system $A\mathbf{x} = \mathbf{b}$, whereas the assumption that the matrix A has generic rank profile is relaxed due to the following result of [KS91].

Theorem 8.9 *Let S be a fixed finite set. Let $L(\mathbf{g})$ and $L(\mathbf{h})$ be $n\times n$ unit lower triangular Toeplitz matrices, each defined by the $n-1$ random entries of its first column, which are uniformly sampled from the set S. Let A be an $n\times n$ matrix of rank ρ. Then the matrix $L^T(\mathbf{g})AL(\mathbf{h})$ has generic rank profile with a probability at least $1-(\rho+1)\rho/(2|S|)$.*

Due to the latter theorem, we may replace the matrix A by the matrix $L^T(\mathbf{g})AL(\mathbf{h})$ and the linear system $A\mathbf{x} = \mathbf{b}$ by the linear system $L^T(\mathbf{g})AL(\mathbf{h})\mathbf{y} = L^T(\mathbf{g})\mathbf{b}$ where $\mathbf{x} = L(\mathbf{h})\mathbf{y}$. Furthermore, (8.1) turns into the equation $L^T(\mathbf{g})AL(\mathbf{h}) = \sum_{k=1}^{r} L^T(\tilde{\mathbf{g}}^{(k)})L^T(\tilde{\mathbf{h}}^{(k)})$, where $\tilde{\mathbf{g}}^{(k)}$ and $\tilde{\mathbf{h}}^{(k)}$ are the convolutions of the pairs of vectors $(\mathbf{g}, \mathbf{g}^{(k)})$ and $(\mathbf{h}, \mathbf{h}^{(k)})$, respectively.

The extension from Toeplitz-like representation (8.1) to Hankel-like with the same generator length is immediate (by means of the pre- or post-multiplication of equation (8.1) by the reflection matrix J).

These observations imply the following result.

Theorem 8.10 *To compute the rank ρ of an $n \times n$ matrix A represented in the form (8.1) and to solve a consistent linear system of equations $A\mathbf{x} = \mathbf{b}$ with such a matrix of coefficients, it suffices to sample uniformly $2n-2$ random values from a fixed finite subset S and to perform $O(r^2 n \log^2 n)$ ops. The resulting algorithm may fail with a probability at most $(\rho+1)\rho/(2|S|) \le (n+1)n/(2|S|)$ but otherwise produces correct output. The same results apply to the computations with the matrices AJ and JA replacing A, where J is the reflection matrix, having ones on its antidiagonal and zero entries elsewhere.*

Now the proof of Theorem 4.4 can be immediately completed due to Theorem 8.1 applied to AJ for a Toeplitz matrix A. $\square$

Remark 8.2 *The proof of Theorem 8.8 applies to Toeplitz-like and to Hankel-like matrices A but can be easily extended to Toeplitz-like + Hankel-like matrices and some other classes of structured matrices by using their known representations via the associated linear operators (cf. [BP94]).*

Remark 8.3 *Algorithm 8.1 and the proof of Theorem 8.1 essentially follow the line of the appendix of [Kal95], which extends the MBA algorithm of [M80], [BA80], to the case of singular input. Due to Theorem 8.6, we derandomize Theorem 8.1, thus decreasing the number of random parameters and the failure probability estimate in Theorem8.10. The computation at stages 1 and 2 of Algorithm 8.1 can be a little simplified further (cf. Lemma 3.1 of [OP98]).*

9 Randomized Reduction of a Matrix to the Tridiagonal Form

A popular alternative to the Frobenius reduction is the reduction to the tridiagonal form, to which a generic $n \times n$ matrix A can be reduced by means of an unsymmetric variation of Lanczos randomized algorithm (cf. [Par92], [BP94], pp. 122-123, 172, 325). To facilitate the analysis of this algorithm, we will next present it in the form based on factorization of the inverse of a Hankel matrix $H^{(0)}$ (which is a *Bezout* *matrix,* cf. [BP94], p. 160) rather than on the more customary computation of three term recurrence relations (cf. [GL96], pp. 503-504).

Algorithm 9.1, *randomized tridiagonalization of a generic matrix by a modified Lanczos algorithm.*

Input: an $n \times n$ matrix A.

Output: FAILURE or a triple of $n \times n$ matrices P, Q and $\tilde{T}$ such that

$$P^H Q = D, \ \tilde{T} = P^H A Q, \tag{9.1}$$

D is a diagonal matrix and T is a complex symmetric tridiagonal matrix.

Computations:

1. Fix a pair of n-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$ and compute and output the Krylov matrices $U = K(A^H, \mathbf{u}, n)$ and $V = K(A, \mathbf{v}, n)$.

2. Compute the Hankel matrix $H^{(0)} = (\mathbf{u}^H A^{i+j} \mathbf{v})_{i,j=0}^{n-1}$.

3. Apply the MBA algorithm of [M80], [BA80], that is, Algorithm 8.1 trivially modified so that it either computes the balanced recursive triangular factorization (BRTF) of the matrices $H^{(0)}$ and $(H^{(0)})^{-1}$ whenever the matrix $H^{(0)}$ is nonsingular and has BRTF or outputs FAILURE otherwise.

4. Compress the BRTF of $(H^{(0)})^{-1}$ to compute the unit lower triangular matrix L and the diagonal matrix D satisfying $(H^{(0)})^{-1} = L^T D L$.

5. Compute and output the matrices $P = U L^T$, $Q = V L^T$ and $\tilde{T} = P^H A Q$.

Now, to yield formal tridiagonal reduction of A, it suffices to compute the matrices $W = Q D^{-1}$ and $T = \tilde{T} D^{-1}$, which, by (9.1), satisfy

$$T = W^{-1} A W. \tag{9.2}$$

Correctness of Algorithm 9.1 is proved in [Par92], [BP94]. The algorithm outputs FAILURE if and only if the matrix $H^{(0)}$ is singular or has no BRTF.

How likely is the failure of the algorithm ? By Theorem 8.3, Algorithm 9.1 does not fail if the matrix $H^{(0)}$ is strongly nonsingular, that is, unless at least one of its $k \times k$ northwestern submatrices $H_k^{(0)}$, $k = 1, \ldots, n$, is singular. The singularity of $H_k^{(0)}$ imposes the polynomial equation, det $H_k^{(0)} = 0$, on the entries of $\mathbf{u}, \mathbf{v}$ and A. Such an equation has degree k in the entries of each of the vectors $\mathbf{u}$ and $\mathbf{v}$. For a fixed pair of nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ and for generic A, such an equation holds for no k, and, therefore, the algorithm does not fail. For generic A and random $\mathbf{u}$ and $\mathbf{v}$, we apply Theorem 4.3. Summarizing this analysis, we obtain the next results.

Theorem 9.1 *Algorithm 9.1 does not fail if* $\mathbf{u}$ *and* $\mathbf{v}$ *are two nonzero vectors and if A is a generic $n \times n$ matrix. If, for a fixed $n \times n$ matrix A and for some pair of vectors* $\mathbf{u}$ *and* $\mathbf{v}$*, the matrix $H^{(0)}$ is strongly nonsingular, then for such a matrix A and for the vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *with entries uniformly sampled from a fixed finite set S, Algorithm 9.1 may fail with a probability at most $(n+1)n/(2|S|)$.*

Remark 9.1 *If A is a nonsingular real symmetric or Hermitian matrix, then $P = Q$, $D = I$, and Algorithm 9.1 never fails for generic vectors* $\mathbf{u}$ *and* $\mathbf{v}$. *(To yield nonsingularity, one may shift to the family of matrices $A - bI_n$ for a scalar b and observe that $Q^H AQ = T$ if and only if $Q^H(A - bI_n)Q = T - bI_n$.) In this case, the resulting tridiagonal matrix T is real symmetric and the subsequent approximation of the eigenvalues is considerably simplifed [BP91], [BP98].*

We also have the following result.

Theorem 9.2 *The cost of performing Stages 1-4 of Algorithm 9.1 is bounded by $O(K_A(n) + K_{A^H}(n) + n^2 \log n)$ ops, which turns into $O(n^2 \log n)$ for Toeplitz-like and Hankel-like matrices A and into $O(n^2 \log^2 n)$ for Vandermonde-like and Cauchy-like matrices A.*

Proof: Clearly, the claimed cost bound applies at stages 1 and 2. Stage 3 requires $O(n \log^2 n)$ ops (see section 8), and this bound is immediately extended to the computation of the BRTF of $(H^{(0)})^{-1}$ at the same cost. Since all matrices defining the BRTF are Hankel-like or Toeplitz-like matrices, $O(n^2 \log n)$ ops suffice for the compression of the BRTF at stage 4 (which gives us the matrices L and D). $\square$

Remark 9.2 *The term $K_{A^H}(0)$ can be deleted from the above cost bound due to Tellegen's theorem [PSD70].*

Stage 5 of Algorithm 9.1 requires order of $M(n)$ ops (note that the structure of A is not preserved in the transition to the matrices L^T, U, and V), but next we will show how to stay within the superior cost bound of Theorem 9.2 if we agree to represent the tridiagonalizaiton of A by the matrices V, D, L^T and T such that the matrices T and $W = VL^TD^{-1}$ satisfy (9.2).

Algorithm 9.2, *accelerated randomized tridiagonalization.*

Input: an $n \times n$ matrix A.

Output: FAILURE or four matrices D (diagonal), L (lower triangular), T (complex tridiagonal) and V such that the matrix $W = VL^TD^{-1}$ is nonsingular and satisfies (9.2).

Computations:

1.–4. as in Algorithm 9.1, except that the computation of the Krylov matrix U at stage 1 can be omitted.

5. Compute the matrix $H^{(1)} = (\mathbf{u}^H A^{i+j}\mathbf{v})_{i,j=1}^n$ (most of the entries of $H^{(1)}$ are also the entries of $H^{(0)}$).

6. Compute and output the $n \times n$ tridiagonal matrix T formed by the three main diagonals of the matrix $LH^{(1)}L^TD^{-1}$.

The correctness proof for this algorithm can be found in [Par92].

Since $H^{(1)}$ is a Hankel matrix, the computation of the matrix $B = H^{(1)}L^TD^{-1}$ requires $O(n^2 \log n)$ ops (cf. (1.7)). Clearly, $O(n^2)$ ops suffice to compute the three main diagonals of the matrix LB.

Summarizing our analysis of Algorithms 9.1 and 9.2, we obtain the next result.

Theorem 9.3 *Let Algorithm 9.2 be applied to generic $n \times n$ matrix A and let the $2n$ coordinates of the vectors $\mathbf{u}$ and $\mathbf{v}$ be uniformly sampled from a fixed finite set S. Then the algorithm either outputs FAILURE with a probability of most $(n+1)n/(2|S|)$ or computes an $n \times n$ tridiagonal matrix T and three nonsingular $n \times n$ matrices, V, L (tower triangular) and D(diagonal), such that the matrices T and $W = VL^T D^{-1}$ satisfy (9.2). Not counting the cost of the random sampling, the entire computation by Algorithm 9.2 requires $O(K_A(n) + n^2 \log n)$ ops.*

By Remark 2.4, the characteristic polynomial $c_A(x)$ is immediately available as soon as the matrix A has been reduced to the Frobenius form or to the triangular Frobenius form. For the tridiagonal reduction, the situation is similar:

Theorem 9.4 *(cf. e.g. [BP91], [BP98]). $O(n)$ ops suffice to obtain the characteristic polynomial of an $n \times n$ tridiagonal matrix.*

10 Approximation of the Eigenvalues and the Computation of Their Algebraic Multiplicities

By Theorem 2.1, every eigenvalue of A is a zero of its both minimum and characteristic polynomials. Suppose that we already know such a polynomial (cf. Theorems 3.1, 4.1, 4.2, 4.4 and 4.5). In this case we may first easily estimate from above the absolute values of all the eigenvalues by applying either Gershgorin's theorem to A (cf. [GL96]) or the known root radius estimates for polynomials (cf. [P97]), then scale A to move all the eigenvalues into the unit disc $\{x : |x| \leq 1\}$ (cf. Remark 2.1), and finally approximate them by applying the polynomial rootfinders of [P95], [P96] to the polynomials $c_A(x)$ or $m_A(x)$ (see [P97] on some alternative effective polynomial rootfinders). This gives us the following result.

Theorem 10.1 *All the eigenvalues of an $n \times n$ matrix A given with its minimum or characteristic polynomials can be approximated within the error bound $2^{-b}||A||$, at the cost of performing $O((n \log^2 n)(\log b + \log^2 n))$ ops.*

The algorithms of [P95], [P96] do not compute polynomial zeros exactly, which makes their application to $c_A(x)$ inefficient for the task of computing the algebraic multiplicities μ_j^+ of the eigenvalues λ_j. We may, however, circumvent this obstacle easily by exploiting the following simple result.

Theorem 10.2 *A zero $x = z$ of a polynomial $p(x)$ has multiplicity $m \geq k \geq 0$ if and only if it is a zero of $p^{(k)}(x)$ of multiplicity $m - k$.*

Definition 10.1 *Let gcd $(u_0), \ldots, u_m(x))$ denote the greatest common divisor of polynomials $u_0(x), \ldots, u_m(x)$. For a fixed polynomial $p(x)$ with coefficient vector $\mathbf{p}$, let us write $d_{\mathbf{p},0}(x) = p(x)$, $d_{\mathbf{p},m}(x) = \gcd(p^{(0)}(x), \ldots, p^{(m)}(x))$, $f_{\mathbf{p},m}(x) = d_{\mathbf{p},m-1}(x)/d_{\mathbf{p},m}(x)$, $m = 1, \ldots, n$.*

We have the following corollary of Theorem 6.2.

Corollary 10.1 *$x = z$ is a zero of $p(x)$ having multiplicity $m \geq 0$ if and only if it is a zero of $f_{\mathbf{p},m}(x)$.*

Corollary 6.1 suggests the following algorithm (cf. [Y76]).

Algorithm 10.1, *approximation of the zeros of a polynomial and computation of their multiplicities.*

Input: the coefficient vector $\mathbf{p}$ of a polynomial $p(x)$ of degree n having all its zeros $z_1, \ldots, z_n$ in the unit disc $\{x : |x| \leq 1\}$, and $\epsilon = 2^{-b}, b > 0$.

Output: k pairs (z_j^*, μ_j^+), $j = 1, \ldots, k$, where $|z_j^* - z_j| \leq \epsilon$, z_j is a zero of $p(x)$ of multiplicity μ_j^+, $j = 1, \ldots, k$,

Computations.

1. Compute $d_{\mathbf{p},m}(x)$ for $m = 1, \ldots, n$.
2. Compute $f_{\mathbf{p},m}(x)$ for $m = 1, \ldots, n$.
3. Approximate within ϵ all the roots of $f_{\mathbf{p},m}(x)$ for $m = 1, \ldots, n$. For each m, $m = 1, \ldots, n$, output the computed approximations $z_{j(m)}^*$ to the roots $z_{j(m)}$ of $f_{\mathbf{p},m}(x)$, together with their multiplicity $m = \mu_{j(m)}^+$.

Correctness of the algorithm follows from Corollary 10.1. Stages 1 and 2 can be performed by recursive application of Euclidean algorithm and straightforward polynomial division, respectively, at the overall cost of performing $O(n^2)$ ops. At stage 3 we apply the algorithm of [P95], [P96], which requires $O((n \log n)(\log b + \log^2 n))$ ops.

Summarizing, we perform Algorithm 10.1 by using $O(n^2 + (n \log n) \log b)$ ops.

Due to Theorem 2.1, Algorithm 10.1 applied to the polynomial $p(x) = c_A(x)$ outputs approximations to the eigenvalues λ_j of A and their algebraic multiplicities μ_j^+.

11 Computation of the Eigenspaces of Given Eigenvalues and the Proof of the Main Theorems

By Definition 2.1, the eigenspace of an eigenvalue λ_j of A is the null space of the matrix

$$A_j = \lambda_j I_n - A, \tag{11.1}$$

having dimension

$$\mu_j = n - \operatorname{rank} A_j. \tag{11.2}$$

In this section we will study the following eigenspace problem.

Problem 11.1, *the eigenspace computation. Given an $n \times n$ matrix A and all its eigenvalues (without their algebraic and geometric multiplicities), compute some bases for the eigenspaces of all these eigenvalues.*

The solution of Problem 11.1 depends on the form in which the matrix A is given. We will prove the following results (cf. Definitions 2.6 and 2.7).

Theorem 11.1 *$O(n^3)$ ops suffice to solve Problem 7.1 if its input matrix is in the triangular Frobenius form or (more generally) is a Hessenberg matrix.*

Theorem 11.2 *$O(n^2)$ ops suffice to solve Problem 7.1 if its input matrix is tridiagonal or is in the diagonal Frobenius form.*

The proofs will rely on the next two theorems.

Theorem 11.3 *Let B be a nonsingular $(n-\mu)\times(n-\mu)$ northwestern submatrix of $n\times n$ matrix $W=\begin{pmatrix} B & C \\ E & G \end{pmatrix}$ of rank r, where $n\geq r\geq n-\mu$. Let S denote the Schur complement of B in W, that is, let $S=G-EB^{-1}C$. Let the columns of a $\mu\times(n-r)$ matrix M form a basis of the null-space of S. Then the $n-r$ columns of the $n\times(n-r)$ matrix*

$$N=\begin{pmatrix} I_{n-\mu} & -B^{-1}C \\ 0 & I_{\mu} \end{pmatrix}\begin{pmatrix} 0 \\ M \end{pmatrix} \tag{11.3}$$

form a basis of the null-space of W.

Proof. Observe that

$$W=\begin{pmatrix} I_{n-\mu} & 0 \\ EB^{-1} & I_{\mu} \end{pmatrix}\begin{pmatrix} B & 0 \\ 0 & S \end{pmatrix}\begin{pmatrix} I_{n-\mu} & B^{-1}C \\ 0 & I_{\mu} \end{pmatrix}, \tag{11.4}$$

$$\begin{pmatrix} I_{n-\mu} & -B^{-1}C \\ 0 & I_{\mu} \end{pmatrix}=\begin{pmatrix} I_{n-\mu} & B^{-1}C \\ 0 & I_{\mu} \end{pmatrix}^{-1}.$$

Multiply W by N and substitute (11.3) and (11.4) into the product, to turn it into a null matrix. This shows that all the $n-r$ linearly independent columns of N are in the null-space of W, whereas the dimension of this null-space is exactly $n-r$. $\square$

Theorem 11.4 *Let an $n\times n$ Hessenberg matrix A have exactly k unreduced Hessenberg diagonal blocks H_s of sizes $\rho_s\times\rho_s$, s=1, ..., k. Let λ_j be an eigenvalue of A having algebraic multiplicity μ_j^+ and geometric multiplicity μ_j. Let exactly μ unreduced diagonal blocks $\lambda_j I_{\rho_s}-H_s$ of the matrix A_j of (11.1) be singular. For every such a singular block, let the row and the column of A_j containing the first row and the last column of the block, respectively, be deleted. Let B denote the resulting submatrix of A_j. Then*

a) B is a nonsingular $(n-\mu)\times(n-\mu)$ matrix,

b) $\mu_j\leq\mu\leq\{\min k,\mu_j^+\}$,

c) $n-\mu=\operatorname{rank} A_j$ if A is a tridiagonal matrix or is block diagonal with unreduced Hessenberg diagonal blocks simultaneously,

d) singularity or nonsingularity of $\lambda_j I_{\rho_s}-H_s$, the s-th diagonal block of A_j, can be decided in $O(\rho_s^2)$ ops; this bound turns into $O(\sum_{s=1}^k\rho_s^2)=O(n^2)$ ops for deciding the singularity or nonsingularity of every one of the k blocks simultaneously,

e) the latter bounds turn into $O(\rho_s)$ and $O(\sum_{s=1}^k\rho_s)=O(n)$, respectively, if the matrix H is tridiagonal or is in the triangular Frobenius form.

Proof. a) Observe that B is an $(n-\mu)\times(n-\mu)$ block triangular matrix. Its diagonal blocks are either nonsingular blocks $\lambda_j I_{\rho_s}-H_s$ or their nonsingular triangular submatrices.

b) The inequality $\mu_j \le \mu$ follows from part a) and from (11.2). Clearly, $\mu \le k$. The bound $\mu \le \mu_j^+$ follows because $c_A(x)=\prod_{s=1}^k c_{H_s}(x)$.

c) If A is a tridiagonal matrix or a block diagonal matrix with unreduced $\rho_s \times \rho_s$ Hessenberg blocks H_s, $s=1,\dots,k$, then for each s, the deleted column of A_j containing the last column of H_s is a linear combination of the ρ_s-1 preceding columns of A_j, and a similar property holds for the deleted rows of A_j. Therefore, in this case, B is a nonsingular submatrix of A_j having the maximum size, that is, $n-\mu=\text{rank } A_j$.

Finally, d) and e) follow from Theorem 2.4. □

Proof of Theorem 11.2 By combining Theorems 11.3 and 11.4 a), c) and e), we compute a basis of the eigenspace of λ_j for any fixed j by using $O((n-\mu)\mu)=O(n\mu)$ ops provided that A is a tridiagonal matrix or is in the diagonal Frobenius form. By Theorem 11.4b), this bound implies the bound $O(n\sum_{j=1}^{l}\mu_j^+)$ for all eigenvalues $\lambda_1,\dots,\lambda_l$ of A. Now we recall that

$$\sum_{j=1}^{l}\mu_j^+ = n, \tag{11.5}$$

and Theorem 11.2 follows. □

Proof Theorem 11.1 We first deduce from Theorem 11.4 that the computation of a basis of the eigenspace of an eigenvalue λ_j of A or, equivalently, of the null-space of the matrix A_j can be performed as follows:

Algorithm 11.1, *eigenspace computation.*

Input: Hessenberg matrix A of Theorem 11.4 and its eigenvalue λ_j.

Output: a basis for the eigenspace of λ_j.

Computations:

1) compute a nonsingular northwestern submatrix B of the matrix $W=P_jA_j\overline{P}_j$ where P_j and $\overline{P}_j$ are permutation matrices, and B has a sufficiently large size $(n-\mu)\times(n-\mu)$,

2) compute the matrix $-B^{-1}C$ of (11.3) and the Schur complement S of B in W,

3) compute a matrix M whose columns form a basis of the null-space of S,

4) compute and output the matrix N of (11.3).

Theorem 11.4 fully specifies stage 1) and shows its computational cost bound.

Let us next specify stages 2)–4) and estimate their computational cost.

Stage 2). The computation of the matrix $-B^{-1}C$ amounts to the solution of μ linear systems of equations with the common coefficient matrix B. By Corollary 2.1, this computation uses $O((n-\mu)^2\mu)=O(n^2\mu)$ ops provided that A_j and B are Hessenberg matrices. $O(n^2\mu)$ dominates the overall cost bound at stage 2). Indeed, the multiplication of the $\mu\times(n-\mu)$ matrix E by the $(n-\mu)\times\mu$ matrix $-B^{-1}C$ requires $O((n-\mu)\mu^2)$ ops.

Stage 3). A basis for the null-space of the $\mu \times \mu$ matrix S can be computed in $O(M(\mu)) = O(\mu^3)$ ops (cf. [BP94], pp. 109-110).

Stage 4). The computations amount to multiplication of $-B^{-1}C$ by M, which requires $O((n-\mu)\mu^2)$ ops.

Summarizing, we obtain a basis for the eigenspace of λ_j by using $O(n^2\mu)$ ops. Recall that $\mu \leq \mu_j^+$ (by Theorem 11.4b), sum the bounds $O(n^2\mu_j)$ over all j, recall (11.5) and obtain Theorem 11.1. □

Proof of Theorems 1.1 and 1.2 Theorem 1.1 immediately follows from Theorems 3.1, 10.1 and 11.1. Likewise, Theorem 1.2 immediately follows from Theorems 4.5, 10.1 and 11.2. □

The reader is referred to [GL96] on the estimates for the errors of computing the eigenspace where the input eigenvalue of Algorithm 11.1 is given approximately within a fixed error bound. Note that λ_j are approximated according to Theorem 10.1, whereas the input matrix H of Algorithm 11.1 can be obtained exactly, by application of the algorithms of sections 3 and 4.

12 Discussion

A natural open question is whether the deterministic complexity bound of Theorem 1.1 can be decreased towards the lower bound $\Omega(M(n))$.

Another natural subject of further study is the Boolean complexity of the matrix eigenproblem. Based on the flowchart given in our introduction, one may obtain some reasonably good upper estimates. Indeed, the Boolean cost of stage b) (of polynomial rootfinding) has been estimated in [P95], [P96] (cf. also [Kir,a]), yielding the optimal (up to a polylogarithmic factor) bound $O(n^2b)$. The other stages of the flowchart, a) and c), are rational. One may perform the computations at these stages modulo several selected primes and then recover the output values by means of the Chinese remainder algorithm, thus bounding the precision of the computations in terms of the output precision p. (p is defined as the minimum precision sufficient to ensure the desired bounds on the output approximation errors.) Then we will only need to apply the well known estimates for the Boolean complexity of an arithmetic operation performed with a fixed finite precision. The remaining nontrivial problems are, of course, the estimation of the above output precision p and the Boolean complexity of the alternative approaches. We leave these open problems as a challenge for the reader.

Finally, let us briefly comment on the arithmetic complexity of the eigenproblem for the resultant matrices associated with multivariate polynomial systems of equations. Let D be a finite upper bound on the number of common roots of these equations. The $D \times D$ resultant matrices associated with such systems have certain structure and form an algebraic variety of dimension at most D in $\mathbf{C}^{D\times D}$. Equation (1.2) is extended easily to this class of matrices, whose eigenvalues equal the fixed coordinate of the common roots of the system represented in a fixed polynomial basis. This follows because, clearly, there are polynomial systems where all roots are simple and have distinct values of this coordinate (cf. Remark 4.3). By Corollary 4.1, equation (1.2) holds generically for such matrices. The extension of Corollary 1.1, however, depends on the solution of the important open problem of proving the bound $K_A(D) = O(D^h)$ for $h < 3$, for the matrices A of this class. Solution of this problem would have immediately given us fast solution of the polynomial system because its common roots are exactly the eigenvectors of A [AuSt88], [Ste96], [BMP98].

References

[AuSt88] W. Auzinger, H. Stetter, An Elimination Algorithm for the Computation of All Zeros of a System of Multivariate Polynomial Equations. In *Proc. Intern. Conf. on Numerical Math., Int. Series of Numerical Math*, **86**, 11–30, Birkhäuser, 1988.

[BA80] R. R. Bitmead, B. D. O. Anderson, Asymptotically Fast Solution of Toeplitz and Related Systems of Linear Equations, *Linear Algebra Appl.*, **34**, 103-116, 1980.

[BM75] A. Borodin, I. Munro, *The Computational Complexity of Algebraic and Numeric Problems*, American Elsevier, New York, 1975.

[BMP98] D. Bondyfalat, B. Mourrain, V. Y. Pan, Controlled Iterative Methods for Solving Polynomial Systems, *Proc. of ACM International Symp. on Symbolic and Algebraic Computation*, (ISSAC'98), 252-259, ACM Press, New York, 1998.

[BP91] D. Bini, V. Y. Pan, Parallel Complexity of Tridiagonal Symmetric Eigenvalue Problem, *Proc. 2nd Annual ACM-SIAM Symp. on Discrete Algorithms*, 384-393, ACM Press, New York, and SIAM Publications, Philadelphia, 1991.

[BP94] D. Bini, V. Y. Pan, *Polynomial and Matrix Computations, Volume 1: Fundamental Algorithms*, Birkhäuser, Boston, 1994.

[BP98] D. Bini, V. Y. Pan, Computing Matrix Eigenvalues and Polynomial Zeros Where the Output Is Real, *SIAM J. on Computing*, **27, 4**, 1099-1115, 1998.

[DL78] R. A. Demillo, R. J. Lipton, A Probabilistic Remark on Algebraic Program Testing, *Information Process. Letters*, **7**, **4**, 193–195, 1978.

[FF63] D. K. Faddeev, V. N. Faddeeva, *Computational Methods of Linear Algebra,* W.H. Freeman, San Francisco, 1963.

[GIKT95] A. George, K. Ikramov, A. N. Krivoshapova, W.-P. Tang, A Finite Procedure for the Tridiagonalization of a General Matrix, *SIAM J. on Matrix Analysis & Applics.,* **16, 2**, 377-387, 1995.

[GL96] G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, Maryland, 1996 (third edition).

[Go94] S. Goedecker, Remark on Algorithms to Find Roots of Polynomials, *SIAM J. on Sci. Computing*, **15**, 1094-1063, 1994.

[GO94] I. Gohberg, V. Olshevsky, Complexity of Multiplication with Vectors for Structured Matrices, *Linear Algebra and Its Applications,* **202**, 163–192, 1994.

[Hei95] G. Heinig, Inversion of Generalized Cauchy Matrices and the Other Classes of Structured Matrices, *Linear Algebra for Signal Processing, IMA Volume in Math. and Its Applications*, **69**, 95-114, Springer, 1995.

[Hig96] N. J. Higham, *Accuracy and Stability in Numerical Analysis*, SIAM Publications, Philadelphia, 1996.

[Kal95] E. Kaltofen, Analysis of Coppersmith's Block Wiedemann Algorithm for the Parallel Solution of Sparse Linear Systems, *Math. of Comput.,* **64, 210,** 777-806, 1995.

[K-G85] W. Keller-Gehrig, Fast Algorithms for Characteristic Polynomial, *Theoretical Computer Science*, **36**, 309-317, 1985.

[Kir,a] P. Kirrinnis, Partial Fraction Decomposition in $C(z)$ and Simultaneous Newton's Iteration for Factorization in $C[Z]$, *J. of Complexity*, to appear.

[KKM79] T. Kailath, S.Y. Kung, M. Morf, Displacement Ranks of Matrices and Linear Equations, *J. Math. Anal. Appl.*, **68, 2**, 395-407, 1979.

[KP91] E. Kaltofen, V. Y. Pan, Processor Efficient Parallel Solution of Linear Systems over an Abstract Field, *Proc. 3rd Ann. ACM Symp. on Paralled Algorithms and Architectures,* 180-191, ACM Press, New York, 1991.

[KS91] E. Kaltofen, B. D. Saunders, On Wiedemann's Method for Solving Sparse Linear Systems, *Proc. AAECC-5, Lecture Notes in Computer Science*, **536**, 29–38, Springer, Berlin, 1991.

[KW92] J. Kuczyński, H. Woźniakovski, Estimating the Largest Eigenvalue by the Power and Lanczos Algorithms with Random Start, *SIAM J. on Matrix Analysis and Applications,* **13**, 1094-1122, 1992.

[LT79] R. J. Lipton, R. E. Tarjan, A Separator Theorem for Planar Graphs, *SIAM J. on Applied Math.,* **36, 2**, 177-189, 1979.

[M80] M. Morf, Doubling Algorithms for Toeplitz and Related Equations, *Proc. IEEE Internet. Conf. on ASSP*, 954-959, IEEE Computer Society Press, 1980.

[MP96] B. Mourrain, V. Y. Pan, Multidimensional Structured Matrices and Polynomial Systems, *Calcolo*, **33**, 389-401, 1996.

[MP97] B. Mourrain, V. Y. Pan, Solving Special Polynomial Systems by Using Structured Matrices and Algebraic Residues, *Proc. of the AMS-SIAM Conference on Foundation of Computational Math.* (F. Cucker and M. Shub editors), Rio-de-Janeiro, pp. 287-304, Springer, January 1997,

[MP98] B. Mourrain, V. Y. Pan, Asymptotic Acceleration of Solving Multivariate Polynomial Systems of Equations, *Proc. 30th Ann. ACM Symp. on Theory of Computing,* 488-496, ACM Press, New York, 1998.

[OP98] V. Olshevsky, V. Y. Pan, A Unified Superfast Algorithm for Boundary Rational Tangential Interpolation Problem and for Inversion and Factorization of Dense Structured Matrices, *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science,* IEEE Computer Society Press, 1998.

[P92] V. Y. Pan, Parametrization of Newton's Iteration for Computations with Structured Matrices and Applications, *Computers & Math. (with Applics.)*, **24, 3**, 61–75, 1992.

[P93] V. Y. Pan, Parallel Solution of Sparse Linear and Path Systems, in *Synthesis of Parallel Algorithms* (J. H. Reif editor), chapter 14, pp. 621-678. Morgan Kaufmann publishers, San Mateo, California, 1993.

[P95] V. Y. Pan, Optimal (up to Polylog Factors) Sequential and Parallel Algorithms for Approximating Complex Polynomial Zeros, *Proc. 27th Ann. ACM Symp. on Theory of Computing*, 741-750, ACM Press, New York, 1995.

[P96] V. Y. Pan, Optimal and Nearly Optimal Algorithms for Approximating Polynomial Zeros, *Computers & Math. (with Applications),* **31**, **12**, 97–138, 1996.

[P97] V. Y. Pan, Solving a Polynomial Equation: Some History and Recent Progress, *SIAM Review*, **39**, **2**, 187-220, 1997.

[P98] V. Y. Pan, Approximate Polynomial Gcds, Padé Approximation, Polynomial Zeros, and Bipartite Graphs, *Proc. 9th Ann. ACM-SIAM Symp. on Discrete Algorithms,* 68-77, ACM Press, New York, and SIAM Publications, Philadelphia, 1998.

[Par92] B. N. Parlett, *The Symmetric Eigenvalue Problem,* Prentice-Hall, 1980.

[PR93] V. Y. Pan, J. Reif, Fast and Efficient Parallel Solution of Sparse Linear Systems, *SIAM J. on Computing* , **22, 6**, 1227-1250, 1993.

[PSD70] P. Penfield Jr., R. Spencer, S. Duinker, *Tellegen's Theorem and Electrical Networks*, MIT Press, Cambridge, Massachusetts, 1970.

[Q94] M. J. Quinn, *Parallel Computing: Theory and Practice*, McGraw-Hill, New York, 1994.

[R87] J. Renegar, On the Worst-case Arithmetic Complexity of Approximating the Zeros of Polynomials, *J. of Complexity,* **3**, 90-113, 1987.

[R95] J. H. Reif, Efficient Parallel Solution of Sparse Eigenvalue and Eigenvector Problems, *Proc. of Ann. IEEE Symp. on Foundations of Computer Science,* 123-132, IEEE Computer Soc. Press, 1995.

[S80] J. T. Schwartz, Fast Probabilistic Algorithms for Verification of Polynomial Identities, *J. of ACM*, **27**, **4**, 701–717, 1980.

[Ste96] H. J. Stetter, Eigenproblems Are at the Heart of Polynomial System Solving, *ACM Bulletin,* **30, 4**, ACM Press, New York, 22-25, 1996.

[Str69] V. Strassen, Gaussian Elimination Is Not Optimal, *Numer. Math.,* **13**, 354–356, 1969.

[T96] E. E. Tyrtyshnikov, A Unified Approach to Old and New Theorems on Distribution and Clustering, *Linear Algebra and Its Applications*, **232**, 1-43, 1996.

[Wi86] D. H. Wiedemann, Solving Sparse Linear Equations Over Finite Fields, *IEEE Trans. on Inf. Theory,* **IT-32, 1**, 54-62, 1986.

[Wil65] J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[Y76] D. Y. Yun, On Square-free Decompositon Algorithm, *Proc. ACM Symp. on Symbolic and Algebraic Computation*, 26-35, ACM Press, New York, 1976.

[Z79] R. E. Zippel, Probabilistic Algorithms for Sparse Polynomials, *Proc. EUROSAM 79, Lecture Notes in Computer Science*, **72**, 216–226, Springer, Berlin, 1979.

Appendix A. Fast Deterministic Computation of the Characteristic Polynomials of Special Matrices

For some important special classes of matrices A, there are fast deterministic algorithms for computing the characteristic polynomial of A. In particular, this is the case for Toeplitz and Toeplitz-like matrices and sparse matrices.

Theorem A.1 *[P92], [BP94]. The coefficients of the characteristic polynomial of an $n \times n$ Toeplitz, Toeplitz-like, Hankel, Hankel-like or Toeplitz-like+Hankel-like matrix can be computed in $O(n^2 \log n)$ ops.*

For some $n \times n$ matrices A, it is easy to compute $\det(b_j I_n - A)$ for fixed distinct scalars b_j, $j = 0, 1, \ldots, n$. In particular det A is immediately obtained in $n - 1$ ops if one has computed either *triangular factorization (TF)*, $A = PLU$ (where P is a permutation matrix, L and U^T are lower triangular matrices and L has its main diagonal filled with ones), or the balanced recursive triangular factorization (BRTF) of A (see Definition 8.1). Similarly, $O(n)$ ops suffice to compute $\det(b_j I_n - A)$ if one has TF or the BRTF of $b_j I_n - A$. (Note that, by Theorem 8.3, there exists the BRTF of $b_j I_n - A$ if $|b_j|$ is sufficiently large.)

Theorem A.2 *Let $b_0, b_1, \ldots, b_n$ be $n+1$ distinct scalars, let A be a fixed $n \times n$ matrix, and let $F(A)$ ops suffice to compute TFs or BRTF of the matrices $b_j I_n - A$ for each j, $j = 0, \ldots, n$. Then $O(n^2) + F(A)(n+1)$ ops suffice to obtain $c_A(x) = \det(xI_n - A)$, the characteristic polynomial of A.*

Proof. $n^2 - 1$ ops suffice to compute $\det(b_j I_n - A)$ for all j if we are given TFs or BRTFs of $b_j I_n - A$, for all j. Then the characteristic polynomial $c_A(x)$ =det $(xI_n - A)$ is obtained by interpolation in $O(n \log^2 n)$ ops ([BP94], pp. 25-26). (For special choices of b_j, e.g. scaled roots of 1 or Chebyshev points, $O(n \log n)$ ops suffice at the interpolation stage.) $\square$

Remark A.1 *Theorem A.2 is the main result of [R95]. We give a much shorter proof.*

Remark A. 2 *$F(A)$ depends on sparsity and structure of A. For an $n \times n$ matrix A defined with its $s(n)$-separator family, we have $F(A) = O(nM(s(n)))$ [P93], [PR93]. For some important classes of sparse matrices A, we have $s(n) = \sqrt{n}$ [LT79], and then $F(A) = O(n^{3/2})$, even where we rely on the rough bound $M(n) \leq 2n^3 - n^2$, supported by the straightforward algorithm.*

If we only seek the eigenvalues of a special matrix A of one of the above classes, then Theorems 10.1, A.1 and A.2 enable fast deterministic solution.

By applying Theorem 11.2 to the Frobenius matrix F satisfying (4.3), one may solve the eigenspace problem for F and A. If for a random vector $\mathbf{v}$, the Krylov matrix $K = K(A, \mathbf{v}, n)$ is nonsingular, Theorem 4.2 immediately gives us an extension to the solution of the entire eigenproblem for the matrix A. It is not clear, however, if one may verify nonsingularity of K at arithmetic cost which is lower than the cost estimated in Theorem 4.5.

B. Approximation of the Eigenvectors by the Inverse Power Iteration

The customary numerical computation of an eigenvector $\mathbf{v}_j$ associated to a given eigenvalue λ_j goes by the Inverse Power Iteration,

$$\tilde{\mathbf{w}}^{(h)} = (\lambda_j^* I_n - A)^{-1} \mathbf{w}^{(h-1)}, \ \mathbf{w}^{(h)} = \tilde{\mathbf{w}}^{(h)} / ||\tilde{\mathbf{w}}^{(h)}||, \ h = 1, 2, \ldots. \qquad (B.1)$$

Here λ_j^* is a fixed approximation to λ_j, and $\mathbf{w}^{(0)}$ is a fixed normalized initial vector, $||\mathbf{w}^{(0)}|| = 1$. Iteration (B.1) stops where the vector norm $||\mathbf{w}^{(h)} - \mathbf{w}^{(h-1)}||$ becomes small enough. The vectors $\mathbf{w}^{(h)}$ converge to an eigenvector $\mathbf{v}_j$ associated with λ_j if

$$\mathbf{w}^{(0)} = \sum_{g=1}^{l} \gamma_g \mathbf{v}_g, \ \gamma_j \neq 0, j \leq l, \qquad (B.2)$$

where $\mathbf{v}_g$ are some eigenvectors of A associated with eigenvalues λ_g, $g = 1, \ldots, l$, and if

$$\rho_j = |\lambda_j^* - \lambda_j| / \min_{k \neq j} |\lambda_j^* - \lambda_k| < 1; \qquad (B.3)$$

furthermore, the convergence is rapid if $|\gamma_j|$ is not small and if ρ_j is small. Namely, we have the following result (cf. [GL96]):

Theorem B.1 $||\mathbf{w}^{(h)} - \mathbf{v}_j|| = O(\rho_j^h / \gamma_j)$ *as* $h \to \infty$, *under (B.1)-(B.3).*

Theorem B.2 *The computational cost of step (B.1) is* $O(n^2)$ *for a general matrix* A *reduced to Hessenberg form,* $O(n \log^2 n)$ *for a Toeplitz or Toeplitz-like matrix* A *and* $O(n)$ *for a tridiagonal matrix* A.

By Theorem B.1, the Inverse Power Iteration (B.1) converges quite fast if (B.2) and (B.3) hold, $|\gamma_j|$ is not small, and ρ_j is small. For ρ_j close or equal to 1, however, the iteration may converge very slowly or diverge. If A has n linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, then (B.2) holds for $l = n$ unless the vector $\mathbf{w}^{(0)}$ lies in the $(n-1)$-dimensional linear space generated by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n$. On the other hand, no general recipes are known for the choice of $\mathbf{w}^{(0)}$ that would ensure that ρ_j is small (or would just ensure this requirement even probabilistically). Moreover, ρ_j is necessarily close to 1 in the important case where several eigenvalues are clustered about λ_j . Thus, the Inverse Power Iteration should be classified as a heuristic method: it works for some (and in a certain sense for "most") of the input pairs $(A, \mathbf{w}^{(0)})$ but certainly not for all pairs.