A Note on Perturbation Estimates for Invariant Subspaces of Hessenberg Matrices

Aurél Galántai

Óbuda University
Bécsi út 96/b, 1034 Budapest, Hungary
galantai.aurel@nik.uni-obuda.hu

Abstract. We give survey of polynomial and matrix perturbation results that are necessary to understand and develop the invariant subspace perturbation theorem we investigate in details. The main purpose of this note is to point out special features of that result such as computability and sharpness. We tested our perturbation estimate on several matrices. The numerical results indicate a high precision and also the possibility of further development for theory and applications.

Keywords: perturbation, polynomial, matrix, invariant subspace, angles between subspaces.

1 Introduction

The eigenvalue problem of matrices is very important in theory and applications and raises many questions. The eigenvalue problem and the polynomial equations are intertwined via the characteristic polynomial. The matrix and polynomial perturbations have been studied from many aspects and the subject has quite an enormous literature (see, e.g. [25], [37], [5], [1]).

Here we are seeking for numerically computable perturbation estimates for invariant subspaces. In Sections 2 and 3 we recall those basic results and concepts we need for our investigations and also provide comparisons as well. In Section 4 we present computable estimates for the perturbation of invariant subspaces of unreduced Hessenberg matrices. The last section contains the details of computation and examples of numerical testing with some conclusions.

2 Polynomial Perturbation Results

The first computable estimate for the perturbation of polynomial zeros was given by Ostrowski [28] in 1940. He later extended this result to matrices using the fact that their characteristic polynomials are sufficiently close for perturbations small enough [29],[30].

Theorem 1 (Ostrowski [28]). *Let* $p(z) = z^n + a_1 z^{n-1} + \cdots + a_n$ *and* $q(z) = z^n + b_1 z^{n-1} + \cdots + b_n$. *For any root* x_i *of* $p(z)$, *there exists a root* y_j *of* $q(z)$

J. Fodor and R. Fullér (eds.), *Advances in Soft Computing, Intelligent Robotics and Control,* Topics in Intelligent Engineering and Informatics 8,
DOI: 10.1007/978-3-319-05945-7_16, © Springer International Publishing Switzerland 2014

such that

$$|x_i - y_j| \leq \left\{ \sum_{k=1}^{n} |a_k - b_k| \gamma^{n-k} \right\}^{1/n}, \tag{1}$$

where $\gamma = 2 \max_{1 \leq k \leq n} \{|a_k|^{1/k}, |b_k|^{1/k}\}$. *Furthermore, the roots of* p *and* q *can be enumerated as* $\alpha_1, \ldots, \alpha_n$ *and* $\beta, \ldots, \beta_n$*, respectively, in such a way that*

$$\max_i |\alpha_i - \beta_i| \leq (2n-1) \left\{ \sum_{k=1}^{n} |a_k - b_k| \gamma^{n-k} \right\}^{1/n}. \tag{2}$$

If $|a_j - b_j| \leq \varepsilon$, $j = 1, \ldots, n$, then $\max_i |\alpha_i - \beta_i| = O\left(\varepsilon^{1/n}\right)$. This basic result is widely used in the literature as a final state of the art (see, e.g. [34], [9]). However, Beauzamy significantly improved the estimate in 1999. For polynomial $p(z) = \sum_{j=0}^{n} a_j z^{n-j}$, define the Bombieri-norm as $[p]_B = \left(\sum_{j=0}^{n} |a_j|^2 / \binom{n}{j}\right)^{1/2}$.

Theorem 2 (Beauzamy [3]). *Let* $k \geq 1$ *be an integer,* $p(z)$ *and* $q(z)$ *be two polynomials of degree* n*, with* $[p-q]_B \leq \varepsilon$*. If* x_i *is any zero of* $p(z)$ *with multiplicity* k*, there exists a zero* y_j *of* $q(z)$*, with*

$$|x_i - y_j| \leq \left(\frac{n!}{(n-k)!} \frac{\left(1 + |x_i|^2\right)^{n/2}}{\left|q^{(k)}(x_i)\right|} \right)^{1/k} \varepsilon^{1/k}. \tag{3}$$

If

$$\varepsilon \leq \frac{(n-k)!}{2n!} \frac{\left|p^{(k)}(x_i)\right|}{\left(1 + |x_i|^2\right)^{\frac{n-k}{2}}}, \tag{4}$$

then (3) implies

$$|x_i - y_j| \leq \left(\frac{2n!}{(n-k)!} \frac{\left(1 + |x_i|^2\right)^{n/2}}{\left|p^{(k)}(x_i)\right|} \right)^{1/k} \varepsilon^{1/k}. \tag{5}$$

This result of local character implies that in the neighborhood of a zero x_i of multiplicity $k < n$ the order of perturbation is $O\left(\varepsilon^{1/k}\right)$, which is definitely better than $O\left(\varepsilon^{1/n}\right)$ $(\varepsilon \to 0)$, if $p(z)$ has at least two different zeros.

Inspired by Beauzamy's result we developed the following estimate in a different way [14].

Theorem 3 ([14]). *Assume that* $p(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n$ *has the distinct roots* $z_1, \ldots, z_k$ *with multiplicity* $n_1, \ldots, n_k$*. Let* $\widetilde{p}(z) = z^n + \widetilde{a}_1 z^{n-1} + \cdots + \widetilde{a}_{n-1} z + \widetilde{a}_n$ *be a perturbation of* p *with* $\widetilde{a}_i = a_i + \varepsilon_i$*,* $|\varepsilon_i| \leq \varepsilon$*,* $i = 1, \ldots, n$*.*

For $0 < \varepsilon < \varepsilon'$, there exist constants γ_i $(i = 1, \ldots, k)$ depending only on $p(z)$ such that disk

$$|z - z_i| \leq r_i = \left(\frac{2(n_i)!\gamma_i \varepsilon}{\left|p^{(n_i)}(z_i)\right|} \right)^{1/n_i} \qquad (i = 1, \ldots, k) \tag{6}$$

contains exactly n_i zeros of the perturbed polynomial $\widetilde{p}(z)$ provided that $r_i < \frac{1}{2} \min_{\ell \neq j} |z_\ell - z_j|$.

Note that the order of perturbation bound is given by the multiplicity n_i of the nearest root z_i. It also follows that the perturbation of simple roots is of order $O(\varepsilon)$. The estimates of Theorems 2 and 3 are compared in [14].

3 Eigenvalue and Subspace Perturbations of Matrices

Ostrowski [29], [30] proved the first computable bound for the perturbations of matrix eigenvalues as well using Theorem 1.

Theorem 4 (Ostrowski [29], [30]). *Let $A = [a_{ij}]_{i,j=1}^n$, $B = [b_{ij}]_{i,j=1}^n$ be two matrices and*

$$\varphi(\lambda) \equiv |A - \lambda I| = 0, \quad \psi(\lambda) \equiv |B - \lambda I| = 0 \tag{7}$$

the corresponding characteristic polynomials and equations. Denote the zeros of $\varphi(\lambda)$ by λ_i and those of $\psi(\lambda)$ by μ_i. Put

$$M = \max(|a_{ij}|, |b_{ij}|) \quad (i, j = 1, \ldots, n), \tag{8}$$

$$\frac{1}{nM} \sum_{i,j} |a_{ij} - b_{ij}| = \delta. \tag{9}$$

Then to every root μ_i of $\psi(\lambda)$ belongs to a certain root λ_i of $\varphi(\lambda)$ such that we have

$$|\mu_i - \lambda_i| \leq (n + 2) M\delta^{1/n}. \tag{10}$$

Furthermore, for a suitable ordering of λ_i and μ_i we have

$$|\mu_i - \lambda_i| \leq 2(n + 1)^2 M\delta^{1/n}. \tag{11}$$

We need the concept of eigenvalue variation.

Definition 1. *Let $A, B \in \mathbb{C}^{n \times n}$. Assume that $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$ and $\sigma(B) = \{\mu_1, \ldots, \mu_n\}$. Let S_n be the set of all permutations of $\{1, 2, \ldots, n\}$. The eigenvalue variation of A and B is defined by*

$$v(A, B) = \min_{\pi \in S_n} \left\{ \max_i \left|\mu_{\pi(i)} - \lambda_i\right| \right\}. \tag{12}$$

$v(A,B)$ is also called the (optimal) matching distance between the eigenvalues of A and B (see, e.g. [37] or [5]). The next result is a reformulation and improvement of Ostrowski's matrix perturbation theorem although the order of estimate is the same.

Theorem 5 (Bhatia, Elsner, Krause [4]). *Let $A, E \in \mathbb{C}^{n\times n}$. Then*

$$v(A, A+E) \leq 4 \times 2^{-1/n} \left(\|A\| + \|A+E\|\right)^{1-1/n} \|E\|^{1/n}. \tag{13}$$

The above results suggest an $O\left(\varepsilon^{1/n}\right)$ size perturbation of the eigenvalues ($\varepsilon = \|E\|$). However, Bauer and Fike proved the following result in 1960.

Theorem 6 (Bauer, Fike, [2]). *If A is diagonalizable, i.e., $A = X\Lambda X^{-1}$ with $\Lambda =$diag$(\lambda_1(A), \ldots, \lambda_n(A))$, then to each $\lambda_i(A+E)$ there is a $\lambda_j(A)$ such that*

$$|\lambda_i(A+E) - \lambda_j(A)| \leq \|X\| \left\|X^{-1}\right\| \|E\|$$

using any norm for which $\|\Lambda\| = \max_i |\lambda_i(A)|$.

Hence for diagonalizable matrices the perturbation order of eigenvalues is $O(\varepsilon)$ ($\varepsilon = \|E\|$), which is much better than $O\left(\varepsilon^{1/n}\right)$. For normal matrices this bound is even better since X can be unitary matrix with a norm 1.

The Bauer-Fike theorem indicates a significant difference between the polynomials and matrices. While Theorems 2 and 3 are sharp and in generally cannot be improved, the Bauer-Fike theorems guarantees that the eigenvalue perturbations of diagonalizable matrices are of order $O(\varepsilon)$ independently of the multiplicities of eigenvalues.

For non-normal matrices, Henrici [21] was the first to extend the Bauer-Fike result. His result was improved by Chu [8].

Let $J_k(\lambda) \in \mathbb{C}^{k\times k}$ be an upper Jordan block. For any $A \in \mathbb{C}^{n\times n}$, there exists a nonsingular matrix X such that

$$X^{-1}AX = \operatorname{diag}\left(J_{n_1}(\lambda_1), J_{n_2}(\lambda_2), \ldots, J_{n_k}(\lambda_k)\right) \tag{14}$$

and $\sum_{j=1}^{k} n_j = n$. The eigenvalues λ_i, $i = 1, \ldots, k$ are not necessarily distinct.

Denote by $g_m(c)$ the unique nonnegative real zero of equation

$$\phi_m(x) = \sum_{\ell=1}^{m} x^{\ell} = c \quad (c \geq 0). \tag{15}$$

Function $g_m(c)$ is strictly monotone increasing and

$$\min\left\{c/m, \sqrt[m]{c/m}\right\} \leq g_m(c) \leq \sqrt[m]{c} \tag{16}$$

(for proof, see Henrici [21]).

Theorem 7 (Chu [8]). *If $A \in \mathbb{C}^{n\times n}$ has the Jordan canonical form (14), then for any $\mu \in \sigma(A+E)$ there exists $\lambda_j \in \sigma(A)$ such that*

$$|\mu - \lambda_j| \leq 1/g_{n_j}(1/\theta) \leq \max\left\{n_j\theta, (n_j\theta)^{1/n_j}\right\} \tag{17}$$

holds with $\theta = \left\|X^{-1}EX\right\|_2$.

The result indicates an $O\left(\varepsilon^{1/n_i}\right)$ perturbation of the eigenvalue λ_i having multiplicity n_i in the Jordan form (14) like in the polynomial case. One can replace n_j by $m = \max_i n_i$, the maximum size of Jordan blocks, which might be less than the algebraic multiplicity of the eigenvalue λ_j. For $k \geq 2$, this estimate corresponds to those of Theorems 2 and 3 and is asymptotically better than those of Ostrowski-Elsner type. The result can be rephrased as

$$v\left(A, A+E\right) \leq \left(2n-1\right)/g_m\left(1/\theta\right) \leq \left(2n-1\right)\max\left\{m\theta, \left(m\theta\right)^{1/m}\right\}, \tag{18}$$

where $m = \max_i n_i$ and $\theta = \left\|X^{-1}EX\right\|_2$. The result also follows from a Hoffman-Wielandt type theorem of Song [35] (see also [14]).

Note that perturbation bounds of Theorems 2, 3 and 7 are sharp and cannot be improved generally.

The use of Theorem 7 and the companion matrix of polynomials yield the following polynomial perturbation theorem of global character [14].

Assume that $p(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$ has the distinct zeros $z_1, \ldots, z_k$ with multiplicity $n_1, \ldots, n_k$. Then the Jordan form of its companion matrix

$$C = C\left(p\right) = \begin{bmatrix} -a_1 & -a_2 & & & -a_n \\ 1 & 0 & & & 0 \\ & 1 & \ddots & & \\ & & \ddots & \ddots & \\ \mathbf{0} & & & 1 & 0 \end{bmatrix} \tag{19}$$

is given by

$$C = \Pi V J V^{-1} \Pi^T, \tag{20}$$

where $\Pi = [e_n, e_{n-1}, \ldots, e_2, e_1]$, $J = \mathrm{diag}(J_{n_1}\left(z_1\right), \ldots, J_{n_k}\left(z_k\right))$ and

$$V = [V_1, \ldots, V_k] \quad (V_i \in \mathbb{C}^{n\times n_i}), \quad \left(V_i\right)_{pq} = \begin{cases} 0, & \text{if } p < q \\ \binom{p-1}{q-1} z_i^{p-q}, & \text{if } p \geq q \end{cases} \tag{21}$$

(see, e.g. [39], [40], [31] or [17]). The matrix V is called the confluent Vandermonde matrix. The companion matrix is diagonalizable by similarity if and only if all its zeros are distinct, i.e., $k = n$ and $n_i = 1$ $(i = 1, \ldots, n)$, when V is the common Vandermonde matrix. Kittaneh [26] proved that C is normal (unitary) if and only if $p\left(z\right) = z^n + a_n$ with $|a_n| = 1$

Theorem 8. *([14]). Assume that* $p\left(z\right) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$ *has the distinct roots* $z_1, \ldots, z_k$ *with multiplicity* $n_1, \ldots, n_k$*. Let* $\widetilde{p}\left(z\right) = z^n + \widetilde{a}_1 z^{n-1} + \cdots + \widetilde{a}_{n-1}z + \widetilde{a}_n$ *be a perturbation of* p *with* $\widetilde{a}_i = a_i + \varepsilon_i$, $|\varepsilon_i| \leq \varepsilon$, $i = 1, \ldots, n$. *For any root* $\widetilde{z}_i$ *of* $\widetilde{p}\left(z\right)$*, there exists a root* z_j *of* $p\left(z\right)$ *such that*

$$|\widetilde{z}_i - z_j| \leq \max\left\{n_j\theta, \left(n_j\theta\right)^{1/n_j}\right\} \tag{22}$$

with $\theta = \left\|V^{-1}\Delta V\right\|_2$ *and* $\Delta = -e_n w^T$ $(w^T = [\varepsilon_n, \varepsilon_{n-1}, \dots, \varepsilon_1])$*. There also exists a permutation* $\pi \in S_n$ *such that for* $i = 1, \dots, n$*,*

$$\left|\widetilde{z}_{\pi(i)} - z_i\right| \leq (2n-1)\max\left\{m\theta, (m\theta)^{1/m}\right\}, \tag{23}$$

where $m = \max_i n_i$*.*

Since $\theta = O(\varepsilon)$, perturbation bounds (22) and (23) are of order $O\left(\varepsilon^{1/n_j}\right)$ and $O\left(\varepsilon^{1/m}\right)$, respectively.

The perturbation of invariant subspaces is a much more complicated matter than the perturbation of eigenvalues (see, e.g. Davis, Kahan [10] or [37], [5]). A subspace $\mathcal{M} \subset \mathbb{C}^n$ is an invariant subspace of A if $Ax \in \mathcal{M}$ for every $x \in \mathcal{M}$. Particularly, each eigenvector x spans a one dimensional invariant subspace $\mathcal{V} = \{\alpha x \mid \alpha \in \mathbb{C}\}$. For the theory of invariant subspaces we refer to Gohberg, Lancaster and Rodman [18].

For Hermitan matrices there are other type of eigenvalue perturbation estimates that are related to subspace perturbations (see, e.g. [5], [32]). An example of such estimates is the following.

Assume that $A \in C^{n\times n}$ is Hermitian with eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$. If X has orthonormal columns that span an invariant subspace S of A and $M = X^H AX$, then $AX - XM = 0$. Assume that the columns of X span an approximate invariant subspace $\widehat{S}$ of A. Then the residual matrix $R = AX - XM$ is expected to be small. Assume that the eigenvalues of M are $\mu_1 \geq \dots \geq \mu_k$ and $n-k$ eigenvalues are well separated from the eigenvalues of M, that is a number $\delta > 0$ exists such that exactly $n-k$ eigenvalues of A lie outside the interval $[\mu_k - \delta, \mu_1 + \delta]$. Then the following result holds.

Theorem 9 (Stewart [36]). *If* $\rho = \|R\|/\delta < 1$*, then there is an index* j *such that* $\lambda_j, \dots, \lambda_{j+k-1} \in (\mu_k - \delta, \mu_1 + \delta)$ *and*

$$|\mu_i - \lambda_{j+i-1}| \leq \frac{1}{1-\rho^2}\frac{\|R\|^2}{\delta} \quad (i = 1, \dots, k). \tag{24}$$

For nonnormal matrices Kahan, Parlett and Jiang [24] pointed out that "*the norms of residuals of the approximate eigenvectors are not themselves sufficient information to bound an approximate eigenvalue*".

The perturbation of invariant subspaces is measured by the Jordan or canonical angles between two subspaces. For definition and computation of canonical angles, we refer to [12], [13] or [19]. The kth subspace angle between the subspaces $\mathcal{M}$ and $\mathcal{N}$ will be denoted by $\theta_k(\mathcal{M}, \mathcal{N})$, where $k = 1, \dots, j$ and $j = \min\{\dim(\mathcal{M}), \dim(\mathcal{N})\}$.

For Hermitan or normal matrices there are several estimates for the canonical angles given in various forms including quantities such as residual and/or separation, which are difficult to compute in general (see, e.g. [10], [5], [32], [27]). In order to give some insight we recall a result of Ipsen [23] for general matrices, which is close to the results of the subsequent sections at least in character.

Let the perturbed matrix $A+E$ have an invariant subspace $\widehat{\mathcal{M}}$, whose dimension is not necessarily the same as that of $\mathcal{M}$. Let P and $\widehat{P}$ denote the orthogonal projectors onto $\mathcal{M}$ and $\widehat{\mathcal{M}}$, respectively. The absolute separation between A and $A+E$ is defined by

$$\text{abssep}=\text{abssep}_{\{A,A+E\}} = \min_{\|Z\|=1,\ PZ\widehat{P}=Z} \|PAZ - Z(A+E)P\|. \tag{25}$$

Theorem 10 (Ipsen [23]). *If abssep>0 then*

$$\max_i \sin\theta_i \left(\mathcal{M}, \widehat{\mathcal{M}}\right) \leq \|E\| / \textit{abssep}. \tag{26}$$

Next we give a result on the perturbation of the invariant subspaces of unreduced Hessenberg matrices that provides a bound for subspace angles without using any concept of separation.

4 Perturbation Results for Hessenberg Matrices

A matrix is called nonderogatory if exactly one Jordan block may belong to each eigenvalue. A matrix is nonderogatory if and only if it is similar to an unreduced upper Hessenberg matrix. The upper Hessenberg matrix $H \in \mathbb{C}^{n\times n}$ is said to be unreduced, if all $h_{i+1,i}$ elements are nonzero. If H is unreduced, then the last and first entries of the right and left eigenvectors, respectively are nonzero.

Define vectors $x, y \in \mathbb{C}^n$ such that $y^H e_1 = e_n^T x = 1$ and

$$(H - \lambda I)x = p(\lambda)e_1, \tag{27}$$

$$y^H(H - \lambda I) = p(\lambda)e_n^T \tag{28}$$

hold, where λ is real or complex scalar. Here $p(\lambda)$ is the characteristic polynomial of H, which can be easily evaluated at any λ in a numerically stable way from any of the above equations by the Hyman's method (see [40], [41], [22] or [15]).

The following properties hold (see [15]).

Lemma 1. *The components of x and y are polynomials in λ : x_{n-j} and y_{1+j} have degree j $(j = 0, 1, \ldots, n-1)$. The polynomial $p(\lambda)$ is of order n.*

Lemma 2. *The k-th derivative of y, x and $p(\lambda)$ with respect to λ satisfy the relations*

$$(H - \lambda I)\, x^{(k)} = kx^{(k-1)} + p^{(k)}(\lambda)\, e_1, \quad k = 0, 1, \ldots \tag{29}$$

and

$$y^{(k)H}(H - \lambda I) = ky^{(k-1)H} + p^{(k)}(\lambda)\, e_n^T, \quad k = 0, 1, \ldots, \tag{30}$$

where $y^{(k)H}$ denotes the conjugate transpose of $y^{(k)}$ and differentiation is done componentwise.

Lemma 3.

$$p^{(k)}(\lambda) = -ky^H x^{(k-1)} = -ky^{(k-1)H}x, \; k > 0 \tag{31}$$

Define

$$X(r,\lambda) = \left[x(\lambda), x'(\lambda), \frac{1}{2!}x''(\lambda), \ldots, \frac{1}{(r-1)!}x^{(r-1)}(\lambda)\right]. \tag{32}$$

It was shown in [15], that if λ_i is an eigenvalue of H with multiplicity n_i, then the columns of matrix $X(n_i, \lambda_i)$ are the right generalized eigenvectors belonging to λ_i and they span the corresponding invariant subspace. Gohberg, Lancaster and Rodman [18] showed that such an invariant subspace of dimension n_i can define additionally $n_i - 1$ different invariant subspaces of smaller dimension. However, with respect to an eigenvalue, we shall think on the invariant subspace of maximal dimension in the following. Observe that for $s < r$,

$$X(r,\lambda) = \left[X(s,\lambda), \frac{1}{s!}x^{(s)}(\lambda), \ldots, \frac{1}{(r-1)!}x^{(r-1)}(\lambda)\right] \tag{33}$$

and $\mathcal{R}(X(j,\lambda)) \subset \mathcal{R}(X(\ell,\lambda))$ for $j < \ell$. We proved the following results in [16].

Theorem 11 ([16]). *Assume that both $H \in \mathbb{C}^{n\times n}$ and its perturbation $\widehat{H} = H+E$ are unreduced upper Hessenberg matrices for $\|E\|$ ($\|E\| \le \varepsilon$) small enough. Assume that λ_i is an eigenvalue of H with multiplicity n_i and μ_i is a nearby eigenvalue of $\widehat{H}$ with multiplicity m_i ($1 \le m_i \le n_i$). Let P be the orthogonal projection on $\mathcal{R}(X(n_i,\lambda_i))$, $X_1 = X(m_i,\lambda_i)$, $\widehat{X}_1 = \widehat{X}(m_i,\mu_i)$ and $\Delta X_1 = \widehat{X}_1 - X_1$. If θ_k denotes the kth subspace angle between the corresponding invariant subspaces $\mathcal{R}(X(n_i,\lambda_i))$ and $\mathcal{R}\left(\widehat{X}(m_i,\mu_i)\right)$, then for $k = 1,\ldots,m_i$,*

$$0 \le \sin\theta_k \le \left(2\left\|\left(X_1^H X_1\right)^{-1}\right\|\right)^{1/2} \|(I-P)\,\Delta X_1\|. \tag{34}$$

Corollary 1. *There exists a constant $C > 0$ such that*

$$0 \le \sin\theta_k \le \left(2\left\|\left(X_1^H X_1\right)^{-1}\right\|\right)^{1/2} \|\Delta X_1\| \le C\varepsilon^{1/n_i} \quad (k = 1,\ldots,m_i). \tag{35}$$

Corollary 2. *Under the conditions of Theorem 11*

$$\sin\theta_k = O\left((\mu_i - \lambda_i)^{n_i - m_i + 1}\right) = O\left(\varepsilon^{\frac{n_i - m_i + 1}{n_i}}\right) \tag{36}$$

for $k = 1,\ldots,m_i$.

The first corollary is a consequence of the perturbation theorems of Sections 2 and 3. It is somewhat crude in view of Corollary 2 but corresponds to the classic eigenvalue perturbation results. It proves that invariant subspace perturbation is continuous in a sense. It also indicates a positive distance from the set of derogatory matrices (for other approach, see Gohberg, Lancaster, Rodman [18], and Gracia, de Hoyos, Velasco [20]).

The second corollary is based upon a refined estimate of $(I - P)\,\Delta X_1$ and it is somewhat surprising. If an eigenvalue λ of multiplicity n_i splits up into n_i simple ones, then $\sin\theta_1 = O(\varepsilon)$ in contrast to the eigenvalue perturbation, which might be of $O\left(\varepsilon^{1/n_i}\right)$. Examples show the possibility of even better perturbation results [16].

The results of this section were extended to dense perturbations of Hessenberg matrices and general nonderogatory matrices as well [16].

The aim of this paper is to show the computational character and goodness of the above results. Details of computations and numerical testing will be presented in the next section.

5 The Computational Algorithm and Testing

We need to compute the matrix $X(r,\lambda)$ for a given H, λ and r, where

$$x(\lambda) = \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ \vdots \\ x_{n-1} \\ 1 \end{bmatrix}, \quad x^{(k)}(\lambda) = \begin{bmatrix} x_1^{(k)} \\ \vdots \\ x_{n-k}^{(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

First we consider equation $(H - \lambda I)\,x = p(\lambda)\,e_1$. The nth row of the system is

$$h_{n,n-1}x_{n-1} + (h_{nn} - \lambda) = 0.$$

For $1 < i < n$, the ith row is given by

$$h_{i,i-1}x_{i-1} + (h_{ii} - \lambda)\,x_i + h_{i,i+1}x_{i+1} + \cdots + h_{i,n-1}x_{n-1} + h_{in} = 0.$$

We obtain the solution by the following backward substitution algorithm

$$x_{n-1} = (\lambda - h_{nn})\,/h_{n,n-1}, \tag{37}$$

$$x_{i-1} = -\left(h_{in} + (h_{ii} - \lambda)\,x_i + \sum_{j=i+1}^{n-1} h_{ij}x_j\right)/h_{i,i-1}, \quad i = n-1,\ldots,2. \tag{38}$$

This gives the vector x and also $p(\lambda) = e_1^T(H - \lambda)\,x(\lambda)$.

We calculate $x^{(k)}(\lambda)$ for $0 < k < n$ using relation

$$(H - \lambda I)\,x^{(k)} = kx^{(k-1)} + p^{(k)}(\lambda)\,e_1, \tag{39}$$

which having the form

$$(H-\lambda I)\begin{bmatrix} x_1^{(k)} \\ \vdots \\ x_{n-k}^{(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = k\begin{bmatrix} x_1^{(k-1)} \\ \vdots \\ x_{n-k}^{(k-1)} \\ x_{n-k+1}^{(k-1)} \\ \vdots \\ 0 \end{bmatrix} + p^{(k)}(\lambda)\, e_1 \tag{40}$$

reduces to a $(n-k+1) \times (n-k)$ problem. The last equation (row $n-k+1$) reads as

$$h_{n-k+1,n-k} x_{n-k}^{(k)} = k x_{n-k+1}^{(k-1)}. \tag{41}$$

Equation i $(1 < i < n-k+1)$ has the form

$$h_{i,i-1} x_{i-1}^{(k)} + (h_{ii} - \lambda)\, x_i^{(k)} + \sum_{j=i+1}^{n-k} h_{ij} x_j^{(k)} = k x_i^{(k-1)}. \tag{42}$$

Hence the algorithm is the following

$$x_{n-k}^{(k)} = k x_{n-k+1}^{(k-1)} / h_{n-k+1,n-k}, \tag{43}$$

$$x_{i-1}^{(k)} = \left(k x_i^{(k-1)} - (h_{ii} - \lambda)\, x_i^{(k)} - \sum_{j=i+1}^{n-k} h_{ij} x_j^{(k)} \right) / h_{i,i-1}, \quad i = n-k, \ldots, 2. \tag{44}$$

Thus we obtain $x^{(k)}$ and also $p^{(k)}(\lambda)$ from the relation $e_1^T (H - \lambda I)\, x^{(k)} = k x^{(k-1)} + p^{(k)}(\lambda)$ (substitution into the first row). Observe that for computing $X(r, \lambda)$ we do not need to compute $p(\lambda)$ or $p^{(k)}(\lambda)$. Since the computations are performed on the same matrix $H - \lambda I$ in a numerically stable way (see Wilkinson [40], [41] or Higham [22]), the whole computation of $X(r, \lambda)$ is numerically stable.

For the numerical testing we wrote a Matlab program to compute $X(r, \lambda)$ for a given unreduced upper Hessenberg matrix H and eigenvalue λ with known multiplicity r.

The other essential elements of computing an estimate are provided in Matlab. However, instead of the original subroutine `subspace.m` for computing the largest subspace angle, we used subroutines `subspace.m` and `subspacea.m` which are due to Andrew Knyazev and can be downloaded from the site MATLAB Central File Exchange.

We made two different types of numerical testing of our estimate.

1. H and $H + E$ are companion matrices with known zeros (H, $H + E$ and the zeros are known exactly).

2. H and $H+E$ are given unreduced Hessenberg matrices with known zeros and the approximate eigenvalue μ_i is computed by Matlab's `eig` routine. This routine is based on the QR-algorithm that is backward stable, which means that it computes the exact eigenvalues of a perturbed matrix $A+E$ with $\|E\| \approx \epsilon_{machine}\|A\|$ (see Golub, van Loan [19], Tisseur [38] or Kressner [27]). However, the algorithm does not recognize the multiple eigenvalues and there are some precision problems as well (see, e.g. [15]).

Next we show some characteristic results of the numerical testing.

Test problem No. 1: $H = C(p(z))$, $H+E = C(\widetilde{p}(z))$, where $p(z) = z^3 - 2z^2 + z$ and $\widetilde{p}(z) = z^3 - (2+\varepsilon)z^2 + (1+\varepsilon)z - \left(\varepsilon - \varepsilon^2\right)$. H has the single eigenvalue $\lambda = 0$ and the double eigenvalue $\lambda = 1$ $(n_2 = 2)$. $H+E$ has the nearby simple eigenvalues $\lambda = \varepsilon$, $\lambda = 1+\sqrt{\varepsilon}$ $(m_2 = 1)$ and $\lambda = 1-\sqrt{\varepsilon}$. Selecting $\lambda_1 = 1$, $n_1 = 2$, $\mu_1 = 1+\sqrt{\varepsilon}$, $m_1 = 1$ and making elementary calculations we have

$$\sin\theta_1\left(\mathcal{R}\left(X\left(2,1\right)\right), \mathcal{R}\left(\widehat{X}\left(1, 1+\sqrt{\varepsilon}\right)\right)\right) = \\ = \frac{\varepsilon}{\left(42\varepsilon + 6\varepsilon^2 + 36\sqrt{\varepsilon} + 24\varepsilon^{\frac{3}{2}} + 18\right)^{1/2}} = O\left(\varepsilon\right),$$

which is exactly the bound of Corollary 2.

The following and the subsequent figures show the following quantities versus $\|E\|$:

- the exact $\max_i \sin(\theta_i)$ values computed with the routine `subspace.m` by Knyazev [red line],

- the ratio $\max_i \sin(\theta_i) / \|E\|^a$ $(a = (n_i - m_i + 1)/n_i)$ to see if estimate (36) can be improved [green line],

- the estimates (35) [*est1* or cyan dashed line] and (34) [*est2* or black dotted line].

Logarithmic scales are used for both axes.

The results of test problem No. 1 are the following.

These results clearly correspond to the theory. Estimate (35) is indeed crude, but it is still acceptable.

Test problem No. 2: $H = C(p)$, $H+E = C(\widetilde{p})$, where $p(z) = z^5 - z^4$ and $\widetilde{p}(z) = (z+\sqrt{\varepsilon})^2 (z-\sqrt{\varepsilon})^2 (z-1)$, respectively. Selecting $\lambda_1 = 0$, $n_1 = 4$, $\mu_1 = \sqrt{\varepsilon}$, $m_1 = 2$ we obtain by a simple calculation that

$$\sin\theta_1\left(\mathcal{R}\left(X\left(4,0\right)\right), \mathcal{R}\left(\widehat{X}\left(2, \sqrt{\varepsilon}\right)\right)\right) = 0$$

and

$$\sin\theta_2\left(\mathcal{R}\left(X\left(4,0\right)\right), \mathcal{R}\left(\widehat{X}\left(2, \sqrt{\varepsilon}\right)\right)\right) = O\left(\varepsilon^{3/2}\right),$$

which is definitely better than $O\left(\varepsilon^{3/4}\right)$ shown by estimate (36). The computational results are shown on the next figure.

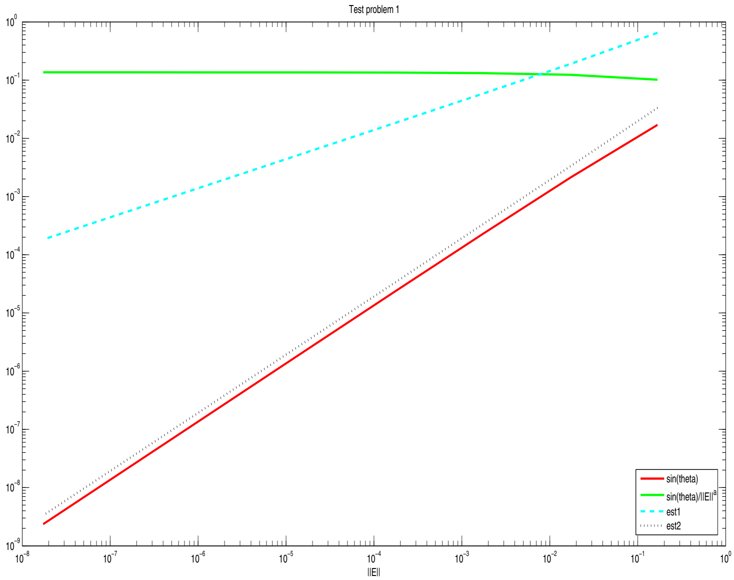

Fig. 1. Test problem 1

Estimate (34) is very sharp. The ratio $\max_i \sin(\theta_i) / \|E\|^a$ indicates that the perturbation order is much better than estimate (36).

Test problem No. 3: $H = C(p)$, $H + E = C(\widetilde{p})$, where

$$p(z) = (z-1)^3 (z+1)(z-2)$$

and

$$\widetilde{p}(z) = (z-1-\varepsilon)^2 (z-1+\varepsilon)(z+1-\varepsilon)(z-2+2\varepsilon).$$

Here $\lambda_1 = 1$, $n_1 = 3$, $m_1 = 1+\varepsilon$, $m_1 = 2$ and the computational results are shown on the next figure.

Here we see again that estimate (34) is very sharp. The $\max_i \sin(\theta_i) / \|E\|^a$ ratio indicates again that the perturbation order is much better than estimate (36). The precision problem shown for the range $\|E\| \lessapprox 10^{-6}$ is due to the fact, that the computed numbers are close to machine epsilon.

Test problem No. 4: $H = H_n^T(\alpha)$, $H + E = H_n^T(\alpha+\varepsilon)$, where $H_n(\alpha)$ is the Chow matrix [7], [11] defined by

$$H_n(\alpha) = [h_{ij}]_{i,j=1}^n, \quad h_{ij} = \begin{cases} \alpha^{i-j+1}, & i \ge j \\ 1, & i = j-1 \\ 0, & i < j-1 \end{cases} \tag{45}$$

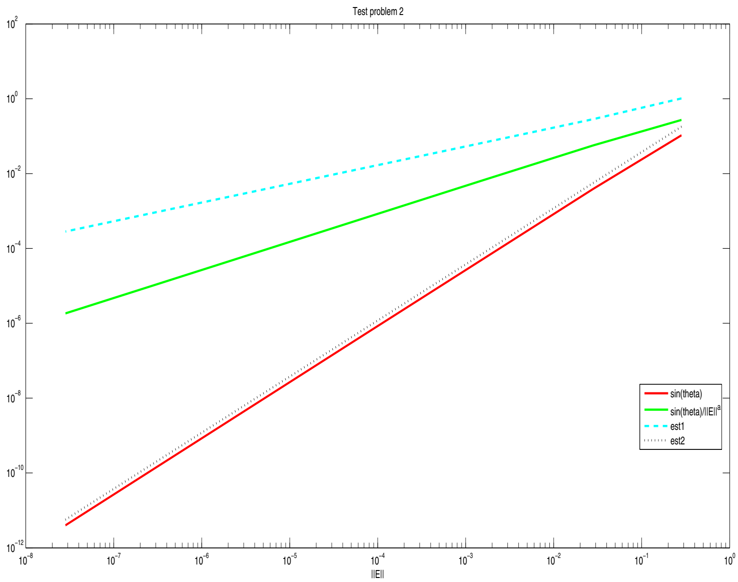

Fig. 2. Test problem 2

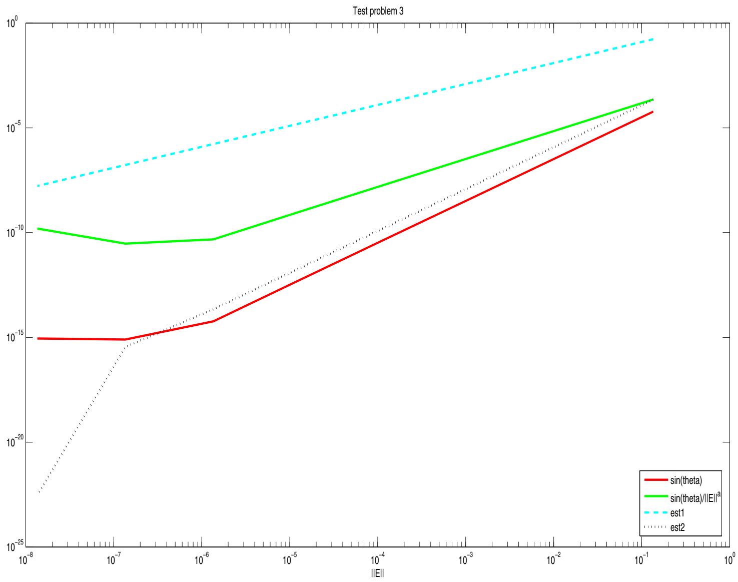

Fig. 3. Test problem 3

The Chow matrix can be found in Matlab gallery library. Chow [7] proved that H_n has $m = \lfloor n/2 \rfloor$ zero eigenvalues and $n - m$ eigenvalues of the form

$$4\alpha \cos^2 \frac{k\pi}{n+2} \quad (k = 1, \ldots, n - m),$$

where $\lfloor n/2 \rfloor$ stands for the lower integer part of $n/2$.

The computational results are shown in the next figure for the parameters $n = 8$, $H = H_n^T(1)$, $H + E = H_n^T(1 + \varepsilon)$, $\lambda_1 = 0$, $n_i = 4$,. Observe that $H + E$ is also Chow matrix and it also has m zero eigenvalues. We present four cases (Version 1-Version 4):

1. μ_1 is the nearest to 0 eigenvalue of $H + E$ provided by Matlab's `eig` routine, $m_1 = 1$.
2. μ_1 is the nearest to 0 eigenvalue of $H + E$ provided by Matlab's `eig` routine, $m_1 = 4$.
3. μ_1 is the average of the eigenvalues (of `eig`) in the zero cluster near to 0, $m_1 = 0$.
4. $\mu_1 = 0$, $m_1 = 4$ (The exact values).

The use of average for clustered (and suspected multiple) eigenvalues was suggested by Saad [33] (Theorem 3.5). Numerical testing also indicates that the multiple eigenvalues when perturbed, show a symmetric pattern around the nonperturbed eigenvalue in the complex plane such that the mean of the errors is fairly zero [6].

Version 1 corresponds the 1-dimensional invariant subspace $\widehat{X}(1, \mu_1)$. The estimate seems to be sharp and it corresponds to theory even if μ_1 is only an approximate eigenvalue of $H + E$. Verson 2 simply shows that we can not take

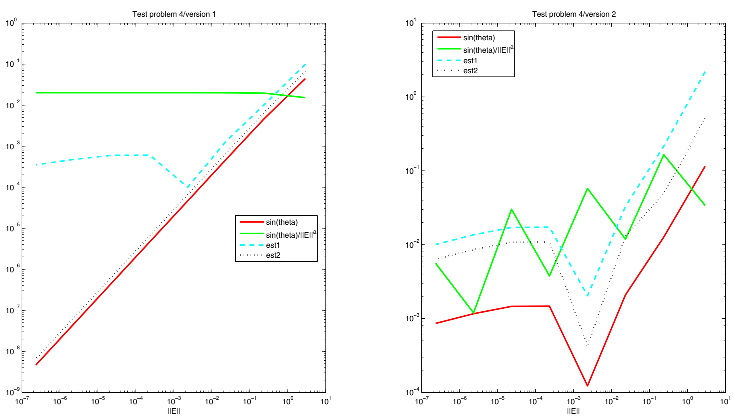

Fig. 4. Test problem 4/Versions 1-2

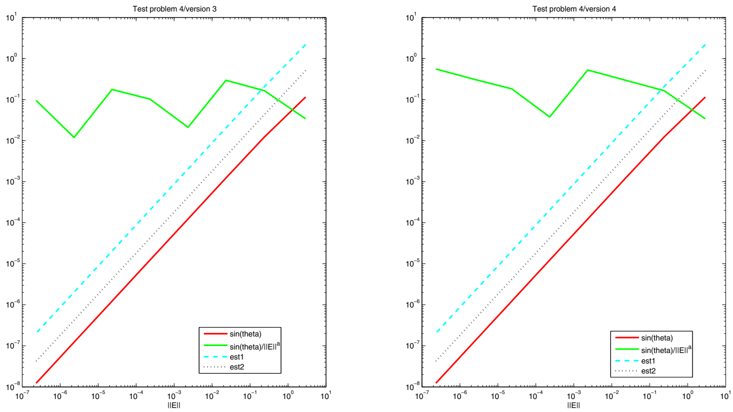

Fig. 5. Test problem 4/Versions 3-4

the approximate μ_1 as a multiple eigenvalue (for problems with Matlab's `eig` routine, see, e.g. [15]).

Figure 5 indicates that Version 3 gives a definitely much better result in agreement with Version 4 that uses exact values.

The presented numerical results show the high precision of estimate (34) and the limits of estimate (35). They also show that the obtained theoretical estimates can be further improved in many cases the reason of which is yet to be understood.

References

1. Baumgartel, H.: Analytic Perturbation Theory for Matrices and Operators. Birkhauser Verlag (1985)
2. Bauer, F.L., Fike, C.T.: Norms and exclusion theorems. Numerische Mathematik 2, 137–144 (1960)
3. Beauzamy, B.: How the roots of a polynomial vary with its coefficients: a local quantitative result. Canad. Math. Bull. 42, 3–12 (1999)
4. Bhatia, R., Elsner, L., Krause, G.: Bounds for the variation of the roots of a polynomial and the eigenvalues of a matrix. Linear Algebra and its Applications 142, 195–209 (1990)
5. Bhatia, R.: Matrix Analysis. Springer (1997)
6. Chatelin, F., Braconnier, T.: About the qualitative computation of Jordan forms. ZAMM 74, 105–113 (1994)
7. Chow, T.S.: A class of Hessenberg matrices with known eigenvalues and inverses. SIAM Review 11, 391–395 (1969)
8. Chu, E.K.: Generalization of the Bauer-Fike Theorem. Numerische Mathematik 49, 685–691 (1986)
9. Ćurgus, B., Mascioni, V.: Roots and polynomials as homeomorphic spaces. Expositiones Mathematicae 24, 81–95 (2006)

10. Davis, C., Kahan, W.M.: The rotation of eigenvectors by a perturbation III. SIAM J. Numer. Anal. 7, 1–46 (1970)
11. Fairweather, G.: On the eigenvalues and eigenvectors of a class of Hessenberg matrices. SIAM Review 13, 220–221 (1971)
12. Galántai, A.: Projectors and Projection Methods. Kluwer (2004)
13. Galántai, A., Hegedűs, C.J.: Jordan s principal angles in complex vector spaces. Numerical Linear Algebra with Applications 13, 589–598 (2006)
14. Galántai, A., Hegedűs, C.J.: Perturbation bounds for polynomials. Numerische Mathematik 109, 77–100 (2008)
15. Galántai, A., Hegedűs, C.J.: Hyman's method revisited. J. Comput. Appl. Math. 226, 246–258 (2009)
16. Galántai, A., Hegedűs, C.J.: Perturbations of invariant subspaces of unreduced Hessenberg matrices. Computers and Mathematics with Applications 65, 423–434 (2013)
17. Gilbert, R.C.: Companion matrices with integer entries and integer eigenvalues and eigenvectors. Amer. Math. Monthly 95, 947–950 (1988)
18. Gohberg, I., Lancaster, P., Rodman, L.: Invariant Subspaces of Matrices with Applications. SIAM (2006)
19. Golub, G.H., van Loan, C.F.: Matrix Computations. The John Hopkins Univ. Press (1996)
20. Gracia, J.M., de Hoyos, I., Velasco, F.E.: Safety neighborhoods for the invariants of the matrix similarity. Linear Multilinear A 46, 25–49 (1999)
21. Henrici, P.: Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices. Numerische Mathematik 4, 24–40 (1962)
22. Higham, N.J.: Accuracy and Stability of Numerical Algorithms. SIAM (1996)
23. Ipsen, I.C.F.: Absolute and relative perturbation bounds for invariant subspaces of matrices. Linear Algebra and its Applications 309, 45–56 (2000)
24. Kahan, W., Parlett, B.N., Jiang, E.: Residual bounds on approximate eigensystems of nonnormal matrices. SIAM Journal on Numerical Analysis 19, 470–484 (1982)
25. Kato, T.: Perturbation Theory for Linear Operators. Springer (1966)
26. Kittaneh, F.: Singular values of companion matrices and bounds on zeros of polynomials. SIAM J. Matrix Anal. Appl. 16, 333–340 (1995)
27. Kressner, D.: Numerical Methods for General and Structures Eigenvalue Problems. Springer (2005)
28. Ostrowski, A.: Recherches sur la méthode de Gräffe et les zeros des polynômes et des series de Laurent. Acta Math. 72, 99–257 (1940)
29. Ostrowski, A.: Über die Stetigkeit von charakteristischen Wurzeln in Abhängigkeit von den Matrizenelementen. Jahresber. Deut. Mat.-Ver. 60, 40–42 (1957)
30. Ostrowski, A.: Solution of Equations and Systems of Equations. Academic Press (1960)
31. Parlett, B.: Canonical decomposition of Hessenberg matrices. Mathematics of Computation 21, 223–227 (1967)
32. Parlett, B.: The Symmetric Eigenvalue Problem. SIAM (1998)
33. Saad, Y.: Numerical Methods for Large Eigenvalue Problems. Manchester University Press (1992)
34. Schönhage, A.: The fundamental theorem of algebra in terms of computational complexity. Technical report, Mathematisches Institut der Universität Tübingen (August 1982) (revised in July 2004)
35. Song, Y.: A note on the variation of the spectrum of an arbitrary matrix. Linear Algebra and its Applications 342, 41–46 (2002)

36. Stewart, G.W.: Two simple residual bounds for the eigenvalues of Hermitian matrices. Technical report, CS-TR 2364, University of Maryland (December 1989)
37. Stewart, G., Sun, J.: Matrix Perturbation Theory. Academic Press (1990)
38. Tisseur, F.: Backward stability of the QR algorithm. Technical report No. 239, UMR 5585 Lyon Saint-Etienne (October 1996)
39. Turnbull, H.W., Aitken, A.C.: An Introduction to the Theory of Canonical Matrices. Blackie & Son (1952)
40. Wilkinson, J.H.: The Algebraic Eigenvalue Problem. Oxford University Press (1965)
41. Wilkinson, J.H.: Rounding Errors in Algebraic Processes. Dover (1994)