Stable Perturbations of Nonsymmetric Matrices

James V. Burke
Mathematics Department
University of Washington GN 50
Seattle, WA 98195

Michael L. Overton
Computer Science Department
Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York, NY 10012

Submitted by Hans Schneider

ABSTRACT

A complex matrix is said to be stable if all its eigenvalues have negative real part. Let J be a Jordan block with zeros on the diagonal and ones on the superdiagonal, and consider analytic matrix perturbations of the form $A(\varepsilon) = J + \varepsilon B + O(\varepsilon^2)$, where ε is real and positive. A necessary condition on B for the stability of $A(\varepsilon)$ on an interval $(0, \varepsilon_0)$, and a sufficient condition on B for the existence of such a family $A(\varepsilon)$, is (i) $\operatorname{Re}\operatorname{tr} B \leqslant 0$; (ii) the sum of the elements on the first subdiagonal of B has nonpositive real part and zero imaginary part; (iii) the sum of the elements on each of the other subdiagonals of B is zero. This result is extended to matrices with any number of nonderogatory eigenvalues on the imaginary axis, and to a stability definition based on the spectral radius. A generalized necessary condition, though not a sufficient condition, applies to arbitrary Jordan structures. The proof of our results uses two important techniques: the Puiseux-Newton diagram and the Arnold normal form. In the nonderogatory case our main results were obtained by Levantovskii in 1980 using a different proof. Practical implications are discussed.

655 Avenue of the Americas, New York, NY 10010 1992 0024-3795/92/$5.00

1. INTRODUCTION

A matrix is said to be stable if all its eigenvalues lie in the open left half of the complex plane. A problem of great theoretical interest and practical importance is the following: given a matrix A on the boundary of the set of stable matrices, i.e. with no eigenvalue having positive real part and one or more eigenvalues lying on the imaginary axis, what perturbations to the matrix are associated with stability? To be more specific, suppose that $A^{(0)}$ is an $n \times n$ Jordan block J_0, with zeros on the diagonal and ones on the main superdiagonal, and consider a real perturbation $A(\varepsilon)$, with $A(0) = J_0$. It is well known (e.g. [24, Section 2.2]) that if

$$A(\varepsilon) = \begin{bmatrix} 0 & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ \varepsilon & 0 & & & 0 \end{bmatrix},$$

then the multiple zero eigenvalue splits into n eigenvalues

$$\varepsilon^{1/n}\omega^h, \qquad h = 1, \dots, n, \tag{1}$$

where ω is the nth principal root of unity, $\exp(2\pi i/n)$. Thus if $n > 2$, at least one of these eigenvalues has positive real part for all $\varepsilon \neq 0$.

In this paper we give a much stronger result. We show, for example, that if $A(\varepsilon)$ is a real analytic matrix function of the form

$$A(\varepsilon) = J_0 + \varepsilon B + O(\varepsilon^2),$$

then a necessary condition on B for the stability of $A(\varepsilon)$ on an interval $(0, \varepsilon_0)$, and a sufficient condition on B for the existence of such a family $A(\varepsilon)$, is

(i) $\operatorname{tr} B \leqslant 0$, i.e., the sum of the diagonal elements of B is less than or equal to zero;

(ii) the sum of the elements on the first subdiagonal of B is less than or equal to zero;

(iii) the sum of the elements on each of the other subdiagonals of B is exactly zero.

Note that (iii) excludes (1) for $n > 2$. In fact, we show that the jth diagonal in the lower triangle (counting the main diagonal as the first) is associated with perturbations of the form

$$\varepsilon^{1/j}\omega^h, \qquad h = 1, \dots, j,$$

where ω is the jth principal root of unity. This largely explains (i)–(iii), since eigenvalue splitting separated by angles of less than π in the complex plane cannot be tolerated for stability, while the splitting associated with $j = 2$, namely in opposite directions in the complex plane, is permitted only along the imaginary axis. The case $j = 1$ corresponds to an eigenvalue shift permitted only in the negative direction. Our results are established in quite full generality, allowing $A^{(0)}$ and $A(\varepsilon)$ to be real or complex and requiring knowledge only of the Jordan form of $A^{(0)}$. In the case of a complex perturbation of J_0 (taking ε to be real), we obtain, instead of (i), that $\operatorname{Re}\operatorname{tr} B \leqslant 0$, and, instead of (ii), that the sum of the elements on the first subdiagonal has real part less than or equal to zero and imaginary part equal to zero. This is because an imaginary shift in the eigenvalues is permissible, but even a small rotation of a split along the imaginary axis is not.

In the case that $A^{(0)}$ has one or more derogatory multiple eigenvalues, we obtain a generalized necessary condition but no sufficiency condition. For example, in the nondefective (semisimple) case (e.g., $A^{(0)} = 0$), conditions of the form (ii) and (iii) vanish (since a perturbation of a nondefective eigenvalue is always Lipschitz), and (i) remains an obvious necessary but insufficient condition for stability. In fact, a stronger necessary and sufficient condition is well known for the nondefective case: if $A^{(0)} = 0$, it is trivially that the maximum real part of the eigenvalues of B be less than or equal to zero. For a more complete discussion of the semisimple case, see [13, 21].

To a large extent the present work arose through attempting to answer the questions raised in Section 6 of [21]. The present paper, together with related work [4], not only answers those questions, but establishes much stronger results than anticipated by [21]. For the much more special symmetric or Hermitian case, see [20].

The proof of our results uses two powerful techniques: the Puiseux-Newton diagram for describing roots of polynomials with analytic coefficients, and the Arnold normal form for perturbation of multiple eigenvalues. These are both explained in the text.

In the nonderogatory case our main results have already been obtained by Levantovskii [15–17], who used a different technique to study C^1 (continuously differentiable) perturbations. Levantovskii also made use of the Arnold normal form, but he did not use the Puiseux-Newton diagram, which we consider to be a key point for the understanding of these results. (Analyticity is required for the use of the diagram.) See Levantovskii's papers, as well as [2, pp. 255–257], for some further results, particularly illustrations of the shape of the stability region in parameter space for various cases.

We note that the Puiseux-Newton diagram has been used in a related context, namely the stability of methods for solving ordinary differential equations; see [11]. However, that work has a very different emphasis, not being concerned with multiple eigenvalues but instead with second-order effects of the perturbation of simple eigenvalues.

All the results have a straightforward extension to the case of reducing the spectral radius of a matrix with multiple eigenvalues. This case is relevant, for example, when the definition of matrix stability requires the magnitude of the eigenvalues to be less than one. This situation typically arises in practice when studying the solution of difference equations rather than differential equations.

The paper is organized as follows. Section 2 uses the Puiseux-Newton diagram to study the roots of polynomials with complex coefficients. Section 3 uses the Arnold normal form to extend these results to perturbation of a complex matrix with multiple eigenvalues, both the nonderogatory and derogatory cases. Section 4 briefly discusses the real case. Section 5 is concerned with reducing the spectral radius of a matrix. Section 6 discusses practical implications.

2. ROOTS OF POLYNOMIALS WITH COMPLEX COEFFICIENTS

A complex function $\beta(\varepsilon)$ is said to be analytic (equivalently, holomorphic) near $\varepsilon = 0$ if it may be expanded in a power series in ε, convergent in a neighborhood of $\varepsilon = 0$. Consider

$$P(\lambda, \varepsilon) = (\lambda - \lambda_0)^n + \beta_1(\varepsilon)(\lambda - \lambda_0)^{n-1} + \cdots + \beta_n(\varepsilon) = 0, \qquad (2)$$

a polynomial equation in λ with analytic coefficients $\beta_j(\varepsilon)$, satisfying

$$\beta_j(0) = 0, \qquad j = 1, \dots, n. \qquad (3)$$

We may therefore write

$$\beta_j(\varepsilon) = \beta_j^{(1)}\varepsilon + \beta_j^{(2)}\varepsilon^2 + \cdots . \qquad (4)$$

We shall restrict ε to a nontrivial real interval $[0, \varepsilon_0]$.

It is well known (e.g. [3, 13]) that the roots of (2) are described by series in fractional powers of ε. These series are commonly called Puiseux series, since it was Puiseux [22] who established their convergence; however, they were derived formally by Newton two centuries earlier. Newton, of course, was concerned only with real coefficients and real roots, but we shall consider the general complex case first for simplicity. We shall obtain the results we need by making use of a diagram devised by Newton for the purpose of calculating coefficients of Puiseux series [19, 23]. Since so many calculation techniques go by the name of Newton, we shall follow [3] in calling this diagram the Puiseux-Newton diagram.

Let $\hat{\beta}_j = \beta_j^{(l_j)}$ be the first nonzero value in the sequence $\{\beta_j^{(1)}, \beta_j^{(2)}, \ldots\}$. By definition, $l_j \geqslant 1$, $j = 1, \ldots, n$. If $\beta_j(\varepsilon)$ is identically zero, take $l_j = \infty$; also, since the coefficient of λ^n in $P(\lambda, \varepsilon)$ is one, take $l_0 = 0$, $\hat{\beta}_0 = 1$. Now plot the values l_j versus j, and consider the lower boundary of the convex hull of the points plotted. Let s_j be the slope of the line on $[j, j+1]$ forming part of this boundary, $j = 0, \ldots, n-1$. Clearly $1/n \leqslant s_0 \leqslant s_1 \leqslant \cdots \leqslant s_{n-1}$. Figure 1 shows the diagram for the following example (taken from [3]):

$$n = 3, \quad \lambda_0 = 0, \quad \beta_1(\varepsilon) = \varepsilon, \quad \beta_2(\varepsilon) = -\varepsilon - \varepsilon^2, \quad \beta_3(\varepsilon) = \varepsilon^2 + 2\varepsilon^3. \tag{5}$$

We have $l_0 = 0$, $l_1 = 1$, $l_2 = 1$, $l_3 = 2$, and so $s_0 = s_1 = \frac{1}{2}$, $s_2 = 1$.

Now consider the following *Ansatz* argument. Suppose a root of (2) is to have the form

$$\lambda(\varepsilon) - \lambda_0 = \alpha\varepsilon^p + \cdots, \tag{6}$$

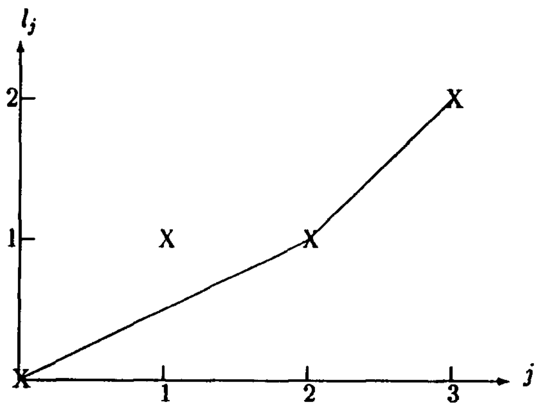


FIG. 1.

where α is nonzero and p is the smallest power of ε in the expansion for this root. Substituting (6) into (2), we need

$$\left(\alpha^n\varepsilon^{np} + \cdots\right) + \left(\hat{\beta}_1\varepsilon^{l_1} + \cdots\right)\left(\alpha^{n-1}\varepsilon^{(n-1)p} + \cdots\right) + \cdots$$
$$+ \left(\hat{\beta}_{n-1}\varepsilon^{l_{n-1}} + \cdots\right)\left(\alpha\varepsilon^{p} + \cdots\right) + \left(\hat{\beta}_n\varepsilon^{l_n} + \cdots\right) = 0.$$

The terms involving the smallest powers of ε are among the terms

$$\alpha^n\varepsilon^{np},\ \hat{\beta}_1\alpha^{n-1}\varepsilon^{l_1+(n-1)p}, \ldots, \hat{\beta}_{n-1}\alpha\varepsilon^{l_{n-1}+p},\ \hat{\beta}_n\varepsilon^{l_n}. \tag{7}$$

For cancellation to take place, at least two terms with the same smallest power of ε must appear. Equivalently, p must equal one or more of the slopes $s_0, \ldots, s_{n-1}$ defined by the Puiseux-Newton diagram. The following discussion will apply to a particular choice of such p. Define f and d by $p = s_f = \cdots = s_{f+d-1}$, so that the line in the diagram with slope p passes from the point (f, l_f) to the point $(f+d, l_{f+d})$. Cancellation of the coefficients of the terms with the smallest powers of ε in (7) requires α to be the root of a polynomial equation with degree d, with leading term $\hat{\beta}_f\alpha^d$ and constant coefficient $\hat{\beta}_{f+d}$, and with an additional intermediate nonzero term for each point (j, l_j) lying on the line in the diagram with slope p, where $f < j < f+d$. Now let $p = q/r$, where q, r are relatively prime integers. By definition, p is an integral multiple of $1/d$, so d is an integral multiple of r, say $d = mr$. It is then clear from the diagram that of the $d-1$ abscissa values j between f and $f+d$, only every rth value is a candidate for the intersection of the line with a point with integer coordinates. Consequently the polynomial of degree d in α reduces to a polynomial of degree m in α^r, which we may denote by $Q(\gamma)$. The conclusion is that the given value of p is associated with d roots with an expansion of the form (6), with α taking the values

$$\gamma_h^{1/r}\omega^j, \qquad h = 1, \ldots, m, \quad j = 1, \ldots, r \tag{8}$$

where the γ_h are the m roots of $Q(\gamma) = 0$, $\gamma_h^{1/r}$ is the principal rth root of γ_h, and ω is the principal rth root of unity.

Completing the example given above, we see that the two values for p are $s_0 = s_1 = \frac{1}{2}$ and $s_2 = 1$. In the case $p = \frac{1}{2}$ we have $f = 0$, $d = 2$, $r = 2$, $m = 1$, with $Q(\gamma) = \gamma - 1$, so the possible values for α are ± 1, giving the Puiseux series

$$\lambda(\varepsilon) - \lambda_0 = \pm\varepsilon^{1/2} + \cdots.$$

In the case $p = 1$ we have $f = 2$, $d = 1$, $r = 1$, $m = 1$, with $Q(\gamma) = \gamma - 1$, so the only possible value for α is 1, giving the Puiseux series

$$\lambda(\varepsilon) - \lambda_0 = \varepsilon + \cdots .$$

The subsequent terms in the series may also be calculated by repeating the process.

The Puiseux-Newton diagram may be used to establish many results. For example, it is a trivial consequence of the diagram that there is a Puiseux series (6) with $p = 1/n$ if and only if $\beta_n^{(1)}$ is nonzero.

We may now state the main result of this section.

THEOREM 1. *Consider the polynomial equation* (2), *with roots given by one or more Puiseux series of the form* (6). *Suppose that there exists* $\varepsilon_0 > 0$ *such that all the roots* $\lambda(\varepsilon)$ *of* (2) *satisfy*

$$\operatorname{Re}[\lambda(\varepsilon) - \lambda_0] \leqslant 0 \tag{9}$$

for $0 \leqslant \varepsilon \leqslant \varepsilon_0$. *Then*

$$\operatorname{Re} \beta_1^{(1)} \geqslant 0, \tag{10}$$

$$\operatorname{Re} \beta_2^{(1)} \geqslant 0, \qquad \operatorname{Im} \beta_2^{(1)} = 0, \tag{11}$$

$$\beta_j^{(1)} = 0, \qquad j = 3, \ldots, n. \tag{12}$$

Proof. The coefficient $\beta_1(\varepsilon)$ is the sum of the differences $\lambda_0 - \lambda(\varepsilon)$ over the roots $\lambda(\varepsilon)$ of (2); thus (10) follows from (9), letting $\varepsilon \to 0$. The other results follow from the Puiseux-Newton diagram as follows. Consider the Puiseux series corresponding to $p = s_0$, the smallest possible value. In order for the roots on the right-hand side of (6) to all be in the left half plane, it is necessary, because of (8), that $r = 1$ or $r = 2$, with $\operatorname{Re} \gamma_h \leqslant 0$, $h = 1, \ldots, m$, in the former case, and

$$\operatorname{Re} \gamma_h \leqslant 0, \quad \operatorname{Im} \gamma_h = 0, \qquad h = 1, \ldots, m, \tag{13}$$

in the latter case. In both cases $p \geqslant \frac{1}{2}$. Since the points (j, l_j) lie on or above the line through the origin with slope $p = s_0$, (12) holds. In the case $r = 1$ we have $p \geqslant 1$, so $\beta_2^{(1)}$ is also zero. In the case $r = 2$, we have

$$Q(\gamma) = \gamma^m + \beta_2^{(1)}\gamma^{m-1} + \cdots + \beta_d^{(m)}. \tag{14}$$

Since $-\beta_2^{(1)}$ is the sum of the roots of $Q(\gamma)$, (11) follows from (13). ∎

The result of Theorem 1 was first obtained by Levantovskii, using a completely different technique developed for C^1 polynomial coefficients. We were unable to verify some of the steps in his proof, particularly the justification for the division by the quantity in [15, p. 20, line D]. Although we do not doubt that the result is correct for C^1 perturbations, we think that our approach using the Puiseux-Newton diagram, which requires analyticity, gives far more insight.

The next theorem shows that the conditions (10)–(12), as well as being necessary for (9), are sufficient for the existence of a polynomial whose roots have real part strictly less than Re λ_0.

THEOREM 2. *Suppose that* (10)–(12) *hold. Then there exists a polynomial of the form* (2) *such that*

$$\operatorname{Re}\left[\lambda(\varepsilon) - \lambda_0\right] < 0$$

on an open interval $(0, \varepsilon_0)$.

Proof. The proof is trivial if $n = 1$. Otherwise, let $\beta_k(\varepsilon)$, $k = 1, \dots, n$, be defined by the coefficients of

$$\left(\lambda - \lambda_0 + \frac{\beta_1^{(1)}}{n}\varepsilon + \varepsilon^2\right)^{n-2}\left(\lambda - \lambda_0 + i\left(\beta_2^{(1)}\varepsilon\right)^{1/2} + \frac{\beta_1^{(1)}}{n}\varepsilon + \varepsilon^2\right)$$
$$\times\left(\lambda - \lambda_0 - i\left(\beta_2^{(1)}\varepsilon\right)^{1/2} + \frac{\beta_1^{(1)}}{n}\varepsilon + \varepsilon^2\right).$$

(Here $i = \sqrt{-1}$.) ■

Note that two roots split from the rest, but that all the roots have the same real part. (It is interesting to compare this with the more complicated example in [15, p. 20, line F], as well as with [21, p. 489].)

It is also possible to use the Puiseux-Newton diagram to make statements about higher-order coefficients of $\beta_j(\varepsilon)$. For example, it seems reasonable to suppose that the following theorem was known to Newton, since its proof using the diagram is trivial.

THEOREM 3. *Let g be the minimum, over all the Puiseux series describing the roots, of the associated values p. Then*

$$\beta_j^{(k)} = 0 \qquad \textit{for all} \quad j > \lfloor k/g \rfloor. \tag{15}$$

Here $\lfloor x \rfloor$ is the largest integer less than or equal to x. This theorem may be viewed as a form of converse to the obvious fact that, if $\beta_j(\varepsilon)$ is identically zero, $j = n - k, \dots, n$, then for any of the Puiseux series describing the roots, $p \geqslant 1/k$, since dividing $P(\lambda, \varepsilon)$ by the factor λ^{n-k} leaves a polynomial of degree k.

Finally we note that Theorem 1 can be generalized in a straightforward way to products of polynomials of the form (2). Since our main interest is in matrices, we omit the details (but see [4]).

3. EIGENVALUES OF COMPLEX MATRICES

Let $A^{(0)} \in \mathscr{C}^{n \times n}$ be a matrix with eigenvalues $\lambda_1, \dots, \lambda_\sigma$ having (algebraic) multiplicities $t_1, \dots, t_\sigma$ respectively, where $\operatorname{Re} \lambda_1 = \cdots = \operatorname{Re} \lambda_\rho = 0$ and $\operatorname{Re} \lambda_{\rho+1} < 0, \dots, \operatorname{Re} \lambda_\sigma < 0$. Let the Jordan form of $A^{(0)}$ be given by

$$A^{(0)} = PJP^{-1},$$

where

$$J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_\sigma \end{bmatrix},$$

$$J_k = \begin{bmatrix} J_{k1} & & \\ & \ddots & \\ & & J_{km_k} \end{bmatrix},$$

and the Jordan block

$$J_{kl} = \begin{bmatrix} \lambda_k & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \lambda_k \end{bmatrix}$$

has dimension n_{kl}. We have

$$n_{k1} + \cdots + n_{km_k} = t_k, \qquad k = 1, \dots, \sigma.$$

If $m_k = 1$, then λ_k is said to be a nonderogatory eigenvalue, while if $m_k = t_k$, i.e. $n_{k1} = \cdots = n_{km_k} = 1$, then λ_k is said to be nondefective (semisimple).

We will now consider an analytic perturbation of the matrix $A^{(0)}$, i.e. a matrix $A(\varepsilon) \in \mathscr{C}^{n\times n}$ each of whose elements is an analytic function of ε near $\varepsilon = 0$, with $A(0) = A^{(0)}$. It will be convenient to define

$$J(\varepsilon) = P^{-1}A(\varepsilon)P,$$

so that $J(0) = J$, the Jordan form of $A^{(0)}$.

We will use a powerful result of V. I. Arnold [1, Section 1 together with Theorem 4.4] which states that *any* such analytic perturbation $J(\varepsilon)$ of J has the following local representation:

$$J(\varepsilon) = Y(\varepsilon)C(\varepsilon)Y(\varepsilon)^{-1}$$

where $Y(\varepsilon)$ is analytic, with $Y(0) = I$,

$$C(\varepsilon) = \mathrm{Diag}\big(C_k(\varepsilon)\big),$$

$$C_k(\varepsilon) = J_k + E_k(\varepsilon),$$

and $E_k(\varepsilon)$, which satisfies $E_k(0) = 0$, has a structure best described by an example as follows:

$$\begin{bmatrix} & & & & & \\ & & & & & \\ X & X & X & X & X & X \\ X & & & & & \\ X & & & X & X & X \\ X & & & X & & X \end{bmatrix}.$$

Here we suppose that J_k has three blocks J_{k1}, J_{k2}, J_{k3}, with dimensions $n_{k1} = 3$, $n_{k2} = 2$, $n_{k3} = 1$ respectively, so that $t_k = 6$. Each X indicates a distinct complex analytic function of ε. Denote the diagonal blocks of $C_k(\varepsilon)$ by $C_{kl}(\varepsilon)$; each of these has the form commonly known as a companion matrix. Let us write

$$C_{kl}(\varepsilon) = \begin{bmatrix} \lambda_k & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda_k & 1 \\ -\theta_{kl}^{(m_k)}(\varepsilon) & \cdots & \cdots & -\theta_{kl}^{(2)}(\varepsilon) & \lambda_k - \theta_{kl}^{(1)}(\varepsilon) \end{bmatrix}, \tag{16}$$

so that, for the example given above,

$$C_k(\varepsilon) = \begin{bmatrix} \lambda_k & 1 & & & & \\ & \lambda_k & 1 & & & \\ -\theta_{k1}^{(3)}(\varepsilon) & -\theta_{k1}^{(2)}(\varepsilon) & \lambda_k - \theta_{k1}^{(1)}(\varepsilon) & X & X & X \\ X & & & \lambda_k & 1 & \\ X & & & -\theta_{k2}^{(2)}(\varepsilon) & \lambda_k - \theta_{k2}^{(1)}(\varepsilon) & X \\ X & & & X & & \lambda_k - \theta_{k3}^{(1)}(\varepsilon) \end{bmatrix}.$$

Again each X indicates a distinct unnamed analytic function of ε. Note that $C_k(0) = J_k$.

The matrix $C(\varepsilon)$ is called a versal deformation of J, and the proof that this versal deformation exists relies on some elementary concepts from differential geometry, in particular the orbit of a matrix and the notion of transversality with respect to the orbit. The orbit of J is simply the set of matrices which are similar to J, i.e. have exactly the same Jordan form. The orbit is a differential manifold in the matrix space $\mathscr{C}^{n\times n}$, and so there exists a linear manifold T which is tangent to the orbit at the point J. The main idea is as follows: since J is similar to all the matrices in its orbit, any perturbation $J(\varepsilon)$ is similar to a representation $C(\varepsilon)$ which lies entirely in a linear manifold of matrices transversal to the orbit at J, e.g. the orthogonal complement of T. For example, consider the "least generic" case, where J is the identity matrix (which has exactly one nondefective eigenvalue); then the orbit consists of the single point I, and the orthogonal complement of T is $\mathscr{C}^{n\times n}$. Thus $C(\varepsilon)$ is a matrix with n^2 distinct analytic component functions. On the other hand, the most generic case is where J is a diagonal matrix with distinct elements $\lambda_1, \ldots, \lambda_n$. In this case it turns out that the orthogonal complement to T is the set of all diagonal matrices, so that it suffices for $C(\varepsilon)$ to contain only n distinct analytic component functions, the off-diagonal elements being zero. In general it turns out that the orthogonal complement to T is the set of matrices which commute with the transpose of J, which leads to the general minimum-parameter form described above. (This particular normal form is not in the orthogonal complement, but in another transversal linear manifold.) For further details, see either Arnold's original paper, or the excellent summary given by Fairgrieve [6].

Before stating the main results we need a definition.

DEFINITION. Define the jth *generalized trace* of a square matrix A, denoted by

$$\operatorname{tr}^{(j)} A,$$

as the sum of the elements on that diagonal of A which is $j-1$ positions below the main diagonal. Thus one obtains the ordinary trace in the case $j = 1$ and the bottom left element of the matrix in the case that j is the dimension of the matrix. If j exceeds the dimension of A, take $\operatorname{tr}^{(j)} A = 0$.

THEOREM 4. *Let $A(\varepsilon)$ be an analytic matrix function with $A(0) = A^{(0)}$, $A'(0) = A^{(1)}$. We shall restrict ε to a nontrivial real interval $[0, \varepsilon_0]$. Partition $P^{-1}A^{(1)}P$ conformally with the partition of J given above, and denote its diagonal block corresponding to J_k by B_k, $k = 1, \ldots, \sigma$, with each B_k having diagonal blocks B_{kl} corresponding to J_{kl}, $l = 1, \ldots, m_k$. Define*

$$\tau_k^{(j)} = \sum_{l=1}^{m_k} \operatorname{tr}^{(j)} B_{kl}, \qquad j = 1, \ldots, \max_l \{n_{kl}\}, \quad k = 1, \ldots, \sigma.$$

(This sum includes zero terms when j exceeds n_{kl}.) Then, for $k = 1, \ldots, \sigma$, the eigenvalues of $A(\varepsilon)$ corresponding to λ_k are the roots of

$$(\lambda - \lambda_k)^{t_k} + \beta_{k1}(\varepsilon)(\lambda - \lambda_k)^{t_k - 1} + \cdots + \beta_{kt_k}(\varepsilon) = 0, \tag{17}$$

where the $\beta_{kj}(\varepsilon)$ are analytic, with

$$\beta_{kj}'(0) = -\tau_k^{(j)}, \qquad j = 1, \ldots, \max_l \{n_{kl}\}.$$

Proof. Define $J(\varepsilon) = P^{-1}A(\varepsilon)P$, and let $C(\varepsilon)$ be the versal deformation of $J(\varepsilon)$ described above. We have

$$\begin{aligned} J'(0) &= C'(0) + Y'(0)J + J(Y^{-1})'(0) \\ &= C'(0) + [Y'(0), J], \end{aligned} \tag{18}$$

where the latter term is the Lie bracket

$$[Y'(0), J] = Y'(0)J - JY'(0).$$

Following Arnold, we note that

$$\langle [Y'(0), J], K \rangle = 0 \tag{19}$$

for all K such that K^* commutes with J, where

$$\langle A, B \rangle = \operatorname{tr} A^* B,$$

the Frobenius matrix inner product, and A^* denotes the complex conjugate transpose of A. Such matrices K^* are block diagonal, $K^* = \mathrm{Diag}(K_k^*)$, where K_k^* commutes with J_k. From [1, 8], the matrices commuting with J_k have a block upper triangular Toeplitz form, specifically, for the example given above,

$$K_k^* = \begin{bmatrix} a & b & c & d & e & f \\ & a & b & & d & \\ & & a & & & \\ & g & h & i & j & k \\ & & g & & i & \\ & & l & & m & n \end{bmatrix}.$$

Consequently, K_k has a block lower triangular Toeplitz structure. Note that if $m_k = 1$ (the nonderogatory case), then K_k is a lower triangular Toeplitz matrix, while if $m_k = t_k$ (the nondefective case), then there is no restriction on the structure of K_k. By (19), then, we have

$$\mathrm{tr}^{(j)} G_{kl} = 0, \quad j = 1, \dots, \max_l \{n_{kl}\}, \qquad k = 1, \dots, \sigma,$$

where the G_{kl} are the diagonal blocks of $[Y'(0), J]$, corresponding in dimension to B_{kl} and J_{kl}. (Indeed, more can be said about the off-diagonal blocks of $[Y'(0), J]$ but we shall not need this.) Now, by definition, the diagonal blocks of $J'(0)$ are B_{kl}, so, by (18),

$$\tau_k^{(j)} = \sum_{l=1}^{m_k} \mathrm{tr}^{(j)} C'_{kl}(0) = -\sum_{l=1}^{m_k} \theta_{kl}^{(j)\prime}(0); \tag{20}$$

here $\theta_{kl}^{(j)}$ is taken to be zero if j exceeds n_{kl}.

Since $A(\varepsilon)$ is similar to $J(\varepsilon)$ and therefore to $C(\varepsilon)$, we need only examine the eigenvalues of $C(\varepsilon)$. We need the following:

LEMMA 1.

$$\det[\lambda I - C_k(\varepsilon)] = (\lambda - \lambda_k)^{t_k} + \beta_{k1}(\varepsilon)(\lambda - \lambda_k)^{t_k - 1} + \cdots + \beta_{kt_k}(\varepsilon), \tag{21}$$

where

$$\beta'_{kj}(0) = \sum_{l=1}^{m_k} \theta_{kl}^{(j)\prime}(0).$$

In the case that λ_k is nonderogatory, the proof of Lemma 1 is trivial. In general, it may be proved by a fairly straightforward but complicated induction argument which gives little insight; we shall not give it here. For a more direct but perhaps less intuitive proof, see [4]. The proof of Theorem 4 is completed by combining (20) and (21). ■

The next theorem continues to use the notation of Theorem 4.

THEOREM 5. *Suppose now that all the eigenvalues of* $A(\varepsilon)$ *have nonpositive real part on a nontrivial interval* $[0, \varepsilon_0]$. *Then, for* $k = 1, \dots, \rho$,

$$\operatorname{Re} \tau_k^{(1)} \leqslant 0, \tag{22}$$

$$\operatorname{Re} \tau_k^{(2)} \leqslant 0, \qquad \operatorname{Im} \tau_k^{(2)} = 0, \tag{23}$$

and

$$\tau_k^{(j)} = 0, \qquad j = 3, \dots, \max_l \{n_{kl}\}. \tag{24}$$

Proof. First note that there is no restriction on the eigenvalues of $C_k(\varepsilon)$ for $k > \rho$ as long as ε_0 is small enough, since $\operatorname{Re} \lambda_k < 0$ for $k > \rho$. The proof is completed by applying Theorem 1 to Equation (17), $k = 1, \dots, \rho$. ■

If λ_k is nonderogatory, i.e. $m_k = 1$ and $J_k = J_{k1}$, $B_k = B_{k1}$, then we have $\tau_k^{(j)} = \operatorname{tr}^{(j)} B_k$, $j = 1, \dots, t_k$. The next theorem shows that the necessary condition of Theorem 5 is also a sufficient condition for the existence of a stable family $A(\varepsilon)$ when the eigenvalues with real part equal to zero are all nonderogatory.

THEOREM 6. *Assume that* λ_k *is nonderogatory, i.e.* $m_k = 1$, *for* $k = 1, \dots, \rho$. *Let* B *be a given matrix, and denote its diagonal blocks, partitioned conformally with* J, *by* B_k, $k = 1, \dots, \sigma$. *Suppose that* B *satisfies*

$$\operatorname{Re} \operatorname{tr}^{(1)} B_k \leqslant 0,$$

$$\operatorname{Re} \operatorname{tr}^{(2)} B_k \leqslant 0, \qquad \operatorname{Im} \operatorname{tr}^{(2)} B_k = 0,$$

$$\operatorname{tr}^{(j)} B_k = 0, \qquad j = 3, \dots, t_k,$$

for $k = 1, \dots, \rho$. *Then there exists an analytic family* $A(\varepsilon)$, *with* $A(0) = A^{(0)}$, $A'(0) = PBP^{-1}$, *such that the maximum real part of the eigenvalues of* $A(\varepsilon)$ *is strictly negative on an open interval* $(0, \varepsilon_0)$.

Proof. There exists an Arnold normal form for $J + \varepsilon B$, namely

$$J + \varepsilon B = Y(\varepsilon) C(\varepsilon) Y(\varepsilon)^{-1},$$

where $Y(\varepsilon)$ is analytic and satisfies $Y(0) = I$, and where $C(\varepsilon) = \mathrm{Diag}(C_k(\varepsilon))$. We need consider only $k \leqslant \rho$ as long as ε is small enough. By the nonderogatory assumption, $C_k(\varepsilon)$ consists of a single companion matrix block; see (16). We have

$$B = C'(0) + \left[Y'(0), J\right].$$

Observe that the relevant diagonal blocks of the second term have generalized traces equal to zero, so the first-order terms in the last row of $C_k(\varepsilon)$ are $\mathrm{tr}^{(j)} B_k$, $j = t_k, \dots, 1$. Now define $\tilde{C}(\varepsilon)$ to be another block diagonal matrix with the same structure as $C(\varepsilon)$. Specifically, choose $\tilde{C}(\varepsilon)$ so that $\tilde{C}(0) = C(0) = J$ and $\tilde{C}'(0) = C'(0)$, with the kth block of $\tilde{C}(\varepsilon)$ set equal to the companion matrix for the corresponding stable polynomial (of degree t_k) defined by Theorem 2. Let

$$\begin{aligned} J(\varepsilon) &= J + \varepsilon B + Y(\varepsilon)\left[\tilde{C}(\varepsilon) - C(\varepsilon)\right]Y(\varepsilon)^{-1} \\ &= Y(\varepsilon)\tilde{C}(\varepsilon)Y(\varepsilon)^{-1}, \end{aligned}$$

so that $J'(0) = B$ and, by Theorem 2, $J(\varepsilon)$ is stable on an interval $(0, \varepsilon_0)$. The proof is completed by defining

$$A(\varepsilon) = PJ(\varepsilon)P^{-1}. \qquad \blacksquare$$

We note that, in the nonderogatory case, Theorem 4 is implicit in the work of Levantovskii. However, he apparently did not obtain our necessary condition in the derogatory case. Regarding analyticity, Arnold [1, p. 37] mentions that his normal form may be extended to C^1 perturbations, although the proof is tedious. As in the polynomial case, this extension does not seem to be worth the effort involved.

Having discussed the nonderogatory case, consider now the other extreme case, namely when all the eigenvalues with real part equal to zero are nondefective (semisimple), i.e. $m_k = t_k$, $k = 1, \dots, \rho$. The necessary condition of Theorem 5 reduces to

$$\mathrm{Re}\,\mathrm{tr}\, B_k \leqslant 0, \qquad k = 1, \dots, \rho,$$

since all the B_{kl} are of dimension one. This condition is an obvious necessary condition for stability, and equally clearly not sufficient for the existence of a stable family $A(\varepsilon)$ with $A'(0) = PBP^{-1}$. A necessary and sufficient condition is, however, well known in this case, as the following theorem states.

THEOREM 7. *Assume that* λ_k *is nondefective, i.e.* $m_k = t_k$, *for* $k = 1, \dots, \rho$. *Let* $A(\varepsilon)$ *be any matrix family with*

$$A(0) = A^{(0)}, \qquad A'(0) = PBP^{-1}, \tag{25}$$

and where, as before, B_k *is the diagonal block of* B *corresponding to* λ_k. *Then a necessary condition for all the eigenvalues of* $A(\varepsilon)$ *to have nonpositive real part on a nontrivial real interval* $[0, \varepsilon_0]$ [*respectively, sufficient condition for all the eigenvalues of* $A(\varepsilon)$ *to have negative real part on an open interval* $(0, \varepsilon_0)$] *is that, for* $k = 1, \dots, \rho$, *the maximum real part of the eigenvalues of* B_k *is less than or equal to zero* [*respectively, less than zero*].

The proof uses [13, 21]. Note that in this case the relevant first ρ columns of P are eigenvectors. Note also that, unlike in Theorem 6, the sufficient condition applies to *any* family $A(\varepsilon)$ satisfying (25); this is because the eigenvalues are Lipschitz in ε. Analyticity of $A(\varepsilon)$ is not required; C^1 parameter dependence is sufficient.

When λ_k is simple, i.e. $m_k = t_k = 1$, $k = 1, \dots, \rho$, the necessary conditions of Theorems 5 and 7 and the sufficient condition of Theorem 6 all reduce to the following: the first ρ diagonal elements of B have nonpositive real part.

The most difficult case is therefore the derogatory, defective case, for which we have given only a necessary condition. Perhaps a sufficient condition could be derived by consideration of the off-diagonal blocks of $[Y'(0), J]$.

4. THE REAL CASE

Every eigenvalue of a real matrix is either real or one of a complex conjugate pair. The results given in Section 3 apply, in particular, in the real case. However, they may be simplified. When $A^{(0)}$ is real, its Jordan form may be taken to have the form

$$A^{(0)} = PJQ^T$$

where $P^TQ = I$ and

$$J = \mathrm{Diag}\left(J_0, J_1, \bar{J}_1, \dots, J_s, \bar{J}_s, \dots\right),$$

$$P = \left[P_0, P_1, \bar{P}_1, \dots, P_s, \bar{P}_s, \dots\right],$$

$$Q = \left[Q_0, Q_1, \bar{Q}_1, \dots, Q_s, \bar{Q}_s, \dots\right].$$

Here J_0 corresponds to the eigenvalue $\lambda_0 = 0$, the only possible real eigenvalue on the imaginary axis; this block and the corresponding blocks P_0, Q_0 may be absent. The blocks P_0 and Q_0 can be taken to be real. Blocks $J_1, \ldots, J_s$ correspond to eigenvalues with positive imaginary part and zero real part. The symbol $\bar{z}$, of course, denotes the complex conjugate of z. The number of eigenvalues on the imaginary axis is $\rho = 2s + 1$ if the zero eigenvalue is present and $\rho = 2s$ otherwise. Blocks not explicitly listed are of no interest, since their eigenvalues have negative real part.

Now assume also that $A(\varepsilon)$ is real. We have, if the zero eigenvalue is present, that $B_0 = Q_0^T A^{(1)} P_0$ is real, and hence that $\tau_0^{(j)}$ is real for all j. Consequently, the conditions (22)–(24) reduce, in the case $k = 0$, to $\tau_0^{(1)} \leqslant 0$, $\tau_0^{(2)} \leqslant 0$, and $\tau_0^{(j)} = 0$ for $j > 2$.

Now consider a complex conjugate pair of eigenvalues. In this case the $\tau_k^{(j)}$ are generally complex, but they occur in complex conjugate pairs. Since each of conditions (22)–(24) applies equally to a complex number or its conjugate, each condition appears exactly twice. Thus it is necessary to apply them only to $\tau_k^{(j)}$ corresponding to eigenvalues with positive imaginary part.

Most likely it is possible to rewrite these conditions using only real quantities by using the real Jordan form [7, 14]. However, our attempts to do this have led only to complicated results. See [7] for the extension of the Arnold normal form to the real field.

Regarding complex conjugate roots of polynomials with real coefficients: instead of (2) we may consider

$$P(\lambda, \varepsilon) = \left[(\lambda - ia)^n + \beta_1(\varepsilon)(\lambda - ia)^{n-1} + \cdots + \beta_n(\varepsilon)\right] \times \left[(\lambda + ia)^n + \bar{\beta}_1(\varepsilon)(\lambda + ia)^{n-1} + \cdots + \bar{\beta}_n(\varepsilon)\right], \quad (26)$$

where a is real and nonzero. Suppose now that $P(\lambda, \varepsilon)$ has no root with positive real part on a nontrivial interval $[0, \varepsilon_0]$. By the generalization of Theorem 1 mentioned at the end of Section 2, it follows that (10)–(12) hold. Let

$$R(\lambda) = \lambda^2 + a^2.$$

Multiplying out the factors in (26), we see that

$$P(\lambda, \varepsilon) = R(\lambda)^{n-2}\left[R(\lambda)^2 + \nu_1 \varepsilon \lambda R(\lambda) + \nu_2 \varepsilon R(\lambda) + \nu_3 \varepsilon\right] + O(\varepsilon^2),$$

where $\nu_1 \geqslant 0$, $\nu_3 \leqslant 0$, and there is no sign restriction on ν_2. The same result (with a different derivation) is given in [15, p. 20, line E].

5. REDUCING THE SPECTRAL RADIUS

The spectral radius of a matrix is the radius of the smallest disk in the complex plane containing all its eigenvalues. For the purposes of this section, we change the definition of matrix stability to require the spectral radius to be less than one. We first modify the polynomial theorems accordingly.

THEOREM 8. *Consider the polynomial equation* (2), *with roots given by one or more Puiseux series of the form* (6). *Suppose that there exists* $\varepsilon_0 > 0$ *such that all the roots* $\lambda(\varepsilon)$ *of* (2) *satisfy*

$$|\lambda(\varepsilon)| \leqslant |\lambda_0| \tag{27}$$

for $0 \leqslant \varepsilon \leqslant \varepsilon_0$. *Then*

$$\operatorname{Re} \bar{\lambda}_0 \beta_1^{(1)} \geqslant 0, \tag{28}$$

$$\operatorname{Re} \bar{\lambda}_0^2 \beta_2^{(1)} \geqslant 0, \qquad \operatorname{Im} \bar{\lambda}_0^2 \beta_2^{(1)} = 0, \tag{29}$$

$$\beta_j^{(1)} = 0, \qquad j = 3, \dots, n. \tag{30}$$

Proof. First note that when λ_0 is real and positive the result is the same as before; this is because the circle of radius $|\lambda_0|$ centered at the origin lies to the left of the vertical axis $\operatorname{Re} z = \operatorname{Re} \lambda_0$. When λ_0 is real and negative we obtain the negative of the original conditions, since the circle lies to the right of the corresponding axis. In general, (27) is equivalent to

$$\operatorname{Re} \bar{\lambda}_0[\lambda(\varepsilon) - \lambda_0] \leqslant -\tfrac{1}{2}|\lambda(\varepsilon) - \lambda_0|^2,$$

so that a necessary condition for (27) to hold is that

$$\operatorname{Re} \bar{\lambda}_0[\lambda(\varepsilon) - \lambda_0] \leqslant 0. \tag{31}$$

The proof is now an easy modification of the proof of Theorem 1 as follows. The coefficient $\beta_1(\varepsilon)$ is the sum of the differences $\lambda_0 - \lambda(\varepsilon)$ over the roots $\lambda(\varepsilon)$ of (2); thus the first condition follows from (27), letting $\varepsilon \to 0$. The other results follow from the Puiseux-Newton diagram. In order for the roots on the right-hand side of (6) to lie in the half plane defined by (31), it is necessary that $r = 1$ or $r = 2$, with $\operatorname{Re} \bar{\lambda}_0 \gamma_h \leqslant 0$ in the former case and

$$\operatorname{Re} \bar{\lambda}_0^2 \gamma_h \leqslant 0, \quad \operatorname{Im} \bar{\lambda}_0^2 \gamma_h = 0, \qquad h = 1, \dots, m,$$

in the latter case. The rest of the proof is exactly as before. ∎

A sufficient condition is somewhat harder to obtain in the spectral-radius case. The difficulty is, by way of example, that an eigenvalue splitting of the form $1 \pm i\varepsilon^{1/2}$, while having the same real part as the unperturbed eigenvalue 1, has spectral radius $\sqrt{1+\varepsilon}$. Consequently, we need to make an assumption that the left-hand side of (28) is sufficiently bigger than the first left-hand side of (29). For example, we have:

THEOREM 9. *Suppose that* $\lambda_0 \neq 0$ *and that* (28)–(30) *hold. Suppose also that*

$$|\beta_2^{(1)}| < 2\,\mathrm{Re}\,\bar{\lambda}_0 \frac{\beta_1^{(1)}}{n}. \tag{32}$$

Then there exists a polynomial of the form (2) *such that*

$$|\lambda(\varepsilon)| < |\lambda_0|$$

on an open interval $(0, \varepsilon_0)$.

Proof. Let $\beta_k(\varepsilon)$, $k = 1, \ldots, n$, be defined by the coefficients of

$$\left(\lambda - \lambda_0 + \frac{\beta_1^{(1)}}{n}\varepsilon\right)^{n-2}\left(\lambda - \lambda_0 + i\left(\beta_2^{(1)}\varepsilon\right)^{1/2} + \frac{\beta_1^{(1)}}{n}\varepsilon\right)$$
$$\times\left(\lambda - \lambda_0 - i\left(\beta_2^{(1)}\varepsilon\right)^{1/2} + \frac{\beta_1^{(1)}}{n}\varepsilon\right). \qquad \blacksquare$$

Note that the larger $|\lambda_0|$ is, the more likely (32) is to be satisfied for given $\beta_1^{(1)}$, $\beta_2^{(1)}$. This reflects the fact that the smaller a circle is, the greater is the distance between it and a tangent line segment of given length. In the limit case $\lambda_0 = 0$, reduction of the spectral radius is, of course, impossible. The factor n in (32) reflects the particular choice of polynomial used in Theorem 2 and can be improved (see [4]).

Now take $A^{(0)}$ to have the same Jordan form as in Section 3, but assuming $|\lambda_1| = \cdots = |\lambda_\rho| = 1$ and $|\lambda_{\rho+1}| < 1, \ldots, |\lambda_\sigma| < 1$. Instead of Theorem 5, we obtain:

THEOREM 10. *Suppose that the spectral radius of* $A(\varepsilon)$ *is less than or equal to one on a nontrivial interval* $[0, \varepsilon_0]$. *Then, for* $k = 1, \ldots, \rho$,

$$\mathrm{Re}\,\bar{\lambda}_k \tau_k^{(1)} \leqslant 0,$$
$$\mathrm{Re}\,\bar{\lambda}_k^2 \tau_k^{(2)} \leqslant 0, \qquad \mathrm{Im}\,\bar{\lambda}_k^2 \tau_k^{(2)} = 0,$$

and

$$\tau_k^{(j)} = 0, \qquad j = 3, \dots, \max_l \{n_{kl}\}.$$

The proof is identical to the proof of Theorem 5, using Theorem 8 instead of Theorem 1.

In the nonderogatory case, we obtain the following sufficient condition.

THEOREM 11. *Assume that λ_k is nonderogatory, i.e. $m_k = 1$, for $k = 1, \dots, \rho$. Let B be a given matrix and denote its diagonal blocks, partitioned conformally with J, by B_k, $k = 1, \dots, \sigma$. Suppose that B satisfies*

$$\operatorname{Re} \bar{\lambda}_k \operatorname{tr}^{(1)} B_k \leqslant 0,$$

$$\operatorname{Re} \bar{\lambda}_k^2 \operatorname{tr}^{(2)} B_k \leqslant 0, \qquad \operatorname{Im} \bar{\lambda}_k^2 \operatorname{tr}^{(2)} B_k = 0,$$

$$\operatorname{tr}^{(j)} B_k = 0, \qquad j = 3, \dots, t_k,$$

for $k = 1, \dots, \rho$. Suppose also that

$$\left|\operatorname{tr}^{(2)} B_k\right| < -\frac{2}{n} \operatorname{Re} \bar{\lambda}_k \operatorname{tr}^{(1)} B_k. \tag{33}$$

Then there exists an analytic family $A(\varepsilon)$, with $A(0) = A^{(0)}$, $A'(0) = PBP^{-1}$, such that the spectral radius of $A(\varepsilon)$ is strictly less than one on an open interval $(0, \varepsilon_0)$.

The proof combines those of Theorem 6 and 9. As before, (33) can be weakened.

Finally, in the nondefective case we have the following modification to Theorem 7.

THEOREM 12. *A necessary condition for the spectral radius of $A(\varepsilon)$ to be less than or equal to one on a nontrivial real interval $[0, \varepsilon_0]$ [respectively,* *sufficient condition for the spectral radius of* $A(\varepsilon)$ *to be less than one on an open interval* $(0, \varepsilon_0)]$ *is that, for* $k = 1, \dots, \rho$, *the maximum real part of the eigenvalues of* $\bar{\lambda}_k B_k$ *is less than or equal to zero [respectively, less than zero].*

The proof uses [21].

6. PRACTICAL IMPLICATIONS

How useful are our results in practice? The necessary condition is practical in the sense that, if the Jordan form of $A^{(0)}$ is known, a candidate first-order perturbation $A^{(1)}$ can be rejected immediately if it does not satisfy the given necessary condition. However, the sufficient condition is of little practical use, since nothing is said about how to compute the actual matrix family $A(\varepsilon)$. Indeed, it is well known that the nonsymmetric eigenvalue problem is very ill conditioned in the case of multiple eigenvalues. The reason for the ill-conditioning is immediately apparent from the Puiseux series (6); the eigenvalues are not Lipschitz when the Puiseux exponent is less than one. Computing the Jordan form of $A^{(0)}$ is itself a very ill-conditioned numerical problem, and, in fact, it is the instability of this process which originally led to Arnold's normal form for perturbed matrices.

Instead of the Jordan form, one may consider reduction of $A^{(0)}$ to some other canonical form. The most stable of these is the triangular or Schur form, since only unitary similarity transformations are required; but the Schur form gives little information beyond the eigenvalues themselves. Nonunitary transformations are required to further reduce a matrix to block diagonal form, with one triangular block associated with each eigenvalue [10, Section 7.1.3]. In fact, our necessary and sufficient conditions have recently been extended to apply to the block diagonal canonical form [4].

The idea of the Schur or block diagonal forms, as opposed to the Jordan form, is to avoid the reduction to too small a number of parameters. On the other hand the Arnold theory uses minimum-parameter versal deformations; hence the use of the Jordan form and the block triangular Toeplitz matrices which commute with the Jordan form. Several proposals have been made to stabilize the computation of the Jordan form of a matrix $A^{(0)}$; see [5, 9, 12]. One approach is to compute the "least generic" of the Jordan forms of those matrices which are within some tolerance of $A^{(0)}$. Perhaps the most promising approach along these lines is that of Fairgrieve [6], who has introduced a path-following method for this purpose. Fairgrieve's method is based on Arnold's normal form and in particular uses Arnold's classification of the degeneracy hierarchy [1], in which, as is well known, a nonderogatory eigenvalue is the most generic and a nondefective one the least generic, but for which the ordering of the "middle" cases is by no means obvious.

We conclude with the description of two numerical experiments, conducted in Matlab [18], which give substantial insight. In the first experiment, we set J to the Jordan block of order $n = 10$ with zeros on the diagonal and ones on the superdiagonal. We then computed the maximum real part of the eigenvalues of $J + \varepsilon B^{(k)}$, say r_k, where $\varepsilon = 10^{-10}$ and $B^{(k)}$ is a random real matrix of order n with the property that

$$\mathrm{tr}^{(j)} B^{(k)} = 0, \qquad j = n - k + 1, \ldots, n. \tag{34}$$

Thus, for $k = 1$, the bottom left element of $B^{(k)}$ is zero; in addition, for $k = 2$, the two elements on the $(n - 1)$th subdiagonal sum to zero, etc. This matrix was computed using the Matlab pseudo-random-number generator, first producing a matrix with elements independently uniformly distributed in the interval $[-1, 1]$, then adding the necessary k constant diagonals to enforce (34). The results were averaged over 100 random matrices and the resulting averaged values of r_k plotted, as a function of k, as *'s in Figure 2. The + symbols in Figure 2 show the quantities $\varepsilon^{1/(n-k)}$. Inspecting Figure 2, one sees a steady reduction in the averaged value of r_k as k increases from 0 to $n/2$, with close agreement between the two plotted values. At $k = n/2$, however, a "floor" is reached. Variation of ε and the machine precision showed that this floor is not a consequence of roundoff error, as might be suspected. Instead, it is easily explained by the Puiseux-Newton diagram. By imposing (34), we are setting the corresponding polynomial coefficients $\beta_j^{(1)}$ to zero. Thus for each k between 1 and $n/2$, slopes with value $1/n, 1/(n-1), \ldots, 1/(n-k+1)$ are not possible. Consequently, roots of the form (6) with corresponding powers are not possible, and the perturbed eigenvalues are instead dominated by terms of order $\varepsilon^{1/(n-k)}$. Hence the close association of the "*" and "+" plots. However, when k reaches $n/2$, the second-order coefficient $\beta_n^{(2)}$ comes into play. Since the second-order terms arising from the Arnold normal form are not zero, the slope with value $2/n$ cannot be eliminated from the Puiseux-Newton diagram. Hence the "floor." We have succeeded only in reducing the order of the eigenvalue perturbations from $\varepsilon^{1/n}$ to $\varepsilon^{2/n}$.

In the second experiment, we set J to the Jordan block of order n and defined B to be a real random matrix (in the same sense as above) with the property that

$$\mathrm{tr}^{(j)} B = 0, \qquad j = 1, \quad j = 3, \ldots, n.$$

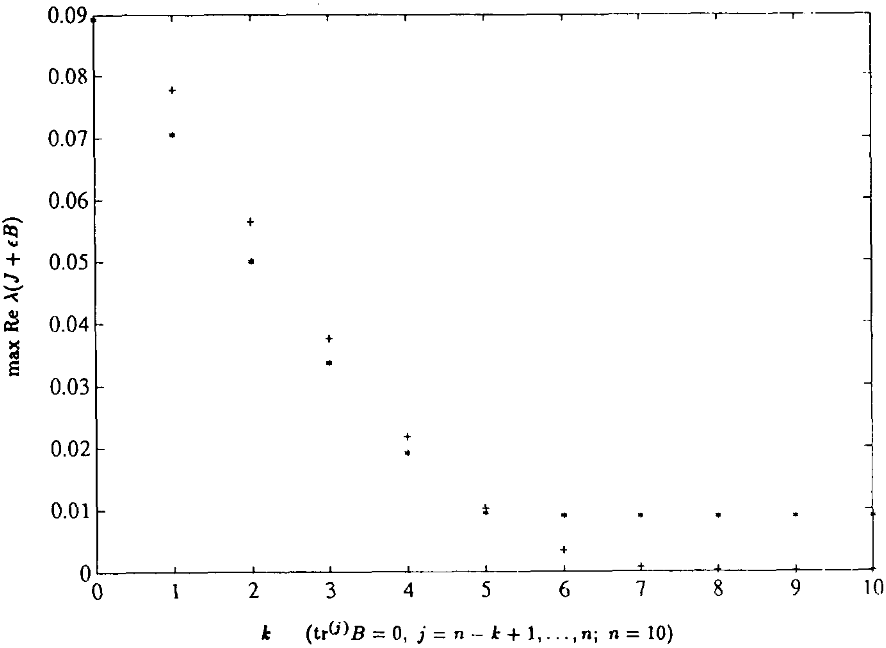
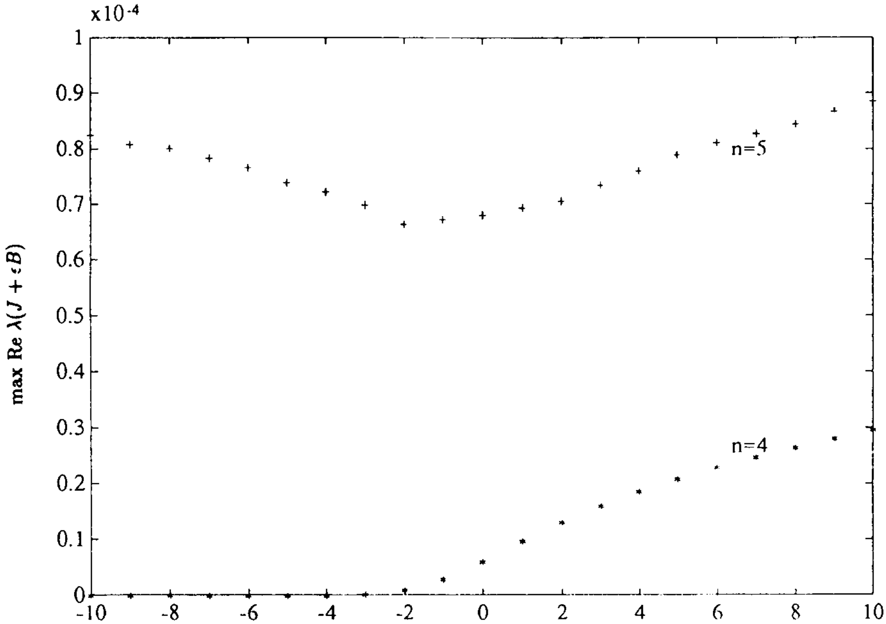


FIG. 2.

This time we plotted the maximum real part of the eigenvalues of $J + \varepsilon B$ against the value of $\mathrm{tr}^{(2)} B$. Again, the results were averaged over 100 random cases. Plots are shown in Figure 3 for $n = 4$ and $n = 5$. The reason for the two very different graphs is similar to the explanation for the result of the first experiment. In the case $n = 5$, the nonzero second-order coefficient $\beta_5^{(2)}$ yields a slope with value $\frac{2}{5}$ in the Puiseux-Newton diagram; this prevents a reduction in the maximum real part of the perturbed eigenvalues. However, in the case $n = 4$, the slopes in the diagram must be at least $\frac{1}{2}$, so that stability is possible when $\mathrm{tr}^{(2)} B \leqslant 0$. Theorem 5 suggests that the larger the quantity $-\mathrm{tr}^{(2)} B$ is, the "more stable" the perturbed matrix should be, and indeed this is supported by the graph for $n = 4$. Stability is not actually achieved, however, because of the effect of $\beta_4^{(2)}$, which is in general nonzero.

We thank V. I. Arnold for bringing the work of Levantovskii to our attention. We also thank R. S. Womersley for some helpful discussions during the early stages of this work. This research was supported in part by National Science Foundation grants DMS-9102059 and CCR-9101640.



$\mathrm{tr}^{(2)}B \quad (\mathrm{tr}^{(j)}B = 0,\ j \neq 2)$

FIG. 3.

REFERENCES

1 V. I. Arnold, On matrices depending on parameters, *Russian Math. Surveys* 26:2:29–43 (1971).

2 V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer-Verlag, New York, 1983.

3 H. Baumgärtel, *Analytic Perturbation Theory for Matrices and Operators*, Birkhäuser, Boston, 1985.

4 J. V. Burke and M. L. Overton, On the subdifferentiability of functions of a matrix spectrum. I: Mathematical Foundations. II: Subdifferential formulas. In: Nonsmooth Optimization, Methods and Applications (F. Giannessi, ed.), Proceedings of a Conference held at Erice, Italy, June 1991, to appear.

5 J. W. Demmel, Computing stable eigendecompositions of matrices, *Linear Algebra Appl.* 79:163–193 (1986).

6 T. F. Fairgrieve, The Application of Singularity Theory to the Computation of the Jordan Canonical Form, M.Sc. Thesis, Dept. of Computer Science, Univ. of Toronto, 1986.

7 D. M. Galin, On real matrices depending on parameters (in Russian), *Uspekhi Mat. Nauk* 27:241-242 (1974).
8 F. R. Gantmacher, *The Theory of Matrices*, Vols. 1, 2, Chelsea, New York, 1959.
9 G. H. Golub and J. H. Wilkinson, Ill-conditioned eigensystems and the computation of the Jordan canonical form, *SIAM Rev.* 18:578-619 (1976).
10 G. H. Golub and C. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins U.P., 1989.
11 R. Jeltsch, Multistep methods using higher derivatives and damping at infinity, *Math. Comp.* 31:124-138 (1977).
12 B. Kågström and A. Ruhe, An algorithm for numerical computation of the Jordan normal form of a complex matrix, *ACM Trans. Math. Software* 6:398-419 (1980).
13 T. Kato, *Perturbation Theory for Linear Operators*, Grundlehren Math. Wiss. 132, Springer-Verlag, New York, 1980.
14 P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Academic, New York, 1985.
15 L. V. Levantovskii, On singularities of the boundary of a stability region, *Moscow Univ. Math. Bull.* 35:6:19-22 (1980).
16 L. V. Levantovskii, The boundary of a set of stable matrices, *Russian Math. Surveys* 35:2:249-250 (1980).
17 L. V. Levantovskii, Singularities of the boundary of the stability domain, *Functional Anal. Appl.* 16:34-37 (1982).
18 *Matlab User's Guide*, The Mathworks, Sherborn, Mass., 1989.
19 I. Newton, *Enumeration of Lines of Third Order, Generation of Curves by Shadows, Organic Description of Curves, and Construction of Equations*, transl. from Latin by C. R. M. Talbot, printed by H. G. Bohn, London, 1860.
20 M. L. Overton, Large-scale optimization of eigenvalues, *SIAM J. Optim.* 2 (1992).
21 M. L. Overton and R. S. Womersley, On minimizing the spectral radius of a nonsymmetric matrix function: Optimality conditions and duality theory, *SIAM J. Matrix Anal. Appl.* 9:473-498 (1988).
22 V. Puiseux, Recherches sur les fonctions algébriques, *J. Math. Pures Appl.* 51:465-581 (1850).
23 M. M. Vainberg and V. A. Trenogen, *Theory of Branching of Solutions of Nonlinear Equations*, Noordhoff International, Leyden, 1974.
24 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford, 1965.

Received 16 October 1990; final manuscript accepted 18 June 1991