where the upper bounds are attainable. From Ostrowski's second result we have

$$|\lambda'-\lambda| < 2\times 21^2\times(400\times\tfrac{1}{2}10^{-10})^{\frac{1}{20}}. \tag{2.6}$$

This bound is far poorer than could be obtained by simple norm theory. For we have

$$\left.\begin{aligned}|\lambda| &\leqslant \|A\|_\infty < 20\\ |\lambda'| &\leqslant \|A+E\|_\infty < 20+10\times 10^{-10}\end{aligned}\right\}, \tag{2.7}$$

and hence

$$|\lambda-\lambda'| \leqslant |\lambda|+|\lambda'| < 40+10^{-9}. \tag{2.8}$$

Ostrowski's result is an improvement on that obtained by simple norm theory only when

$$2(n+1)^2(n^2\epsilon)^{\frac{1}{n}} < 2n,$$

giving

$$\epsilon < \frac{n^{n-2}}{(n+1)^{2n}} \doteq \frac{1}{e^2 n^{n+2}}. \tag{2.9}$$

For $n = 20$ this means that ϵ must be less than 10^{-27}. To obtain a *useful* bound from Ostrowski's result ϵ must be far smaller even than this.

It is evident that the factor $\epsilon^{\frac{1}{n}}$ is the main weakness of Ostrowski's result. Unfortunately, as an example due to Forsythe shows, its presence is inevitable in any general result. Consider for example $C_n(a)$, the Jordan submatrix of order n, which has its n eigenvalues equal to a. If we change the $(1, n)$ element from 0 to ϵ, then the characteristic equation becomes

$$(a-\lambda)^n+(-1)^{n-1}\epsilon = 0, \tag{2.10}$$

giving

$$\lambda = a+\omega^r\epsilon^{\frac{1}{n}} \quad (r = 0,\ldots, n-1) \tag{2.11}$$

for the n eigenvalues, where ω is any primitive nth root of unity.

Algebraic functions

3. In the analysis which follows we shall need two results from the theory of algebraic functions (see for example Goursat, 1933). These we now state without proof.

Let $f(x, y)$ be defined by

$$f(x, y) = y^n+p_{n-1}(x)y^{n-1}+p_{n-2}(x)y^{n-2}+\ldots+p_1(x)y+p_0(x), \tag{3.1}$$

where the $p_i(x)$ are polynomials in x. Corresponding to any value of x the equation $f(x, y) = 0$ has n roots $y_1(x), y_2(x),\ldots, y_n(x)$, each root

being given its appropriate multiplicity. The roots of $f(0, y) = 0$ are therefore denoted by $y_1(0), y_2(0),\ldots, y_n(0)$. The two theorems are as follows.

THEOREM 1. *Suppose $y_i(0)$ is a simple root of $f(0, y) = 0$. Then there exists a positive δ_i such that there is a simple root $y_i(x)$ of $f(x, y) = 0$ defined by*

$$y_i(x) = y_i(0)+p_{i1}x+p_{i2}x^2+\ldots, \tag{3.2}$$

where the series on the right of (3.2) is convergent for $|x| < \delta_i$.

Note: (i) We require only that the root $y_i(0)$ under consideration be simple. No assumption is made about the nature of the other roots of $f(0, y) = 0$. (ii) The series on the right of (3.2) may terminate. (iii) $y_i(x) \to y_i(0)$ as $x \to 0$.

THEOREM 2. *If $y_1(0) = y_2(0) = \ldots = y_m(0)$ is a root of multiplicity m of $f(0, y) = 0$ then there exists a positive δ such that there are exactly m zeros of $f(x, y) = 0$ when $|x| < \delta$ having the following properties.*

The m roots fall into r groups of $m_1, m_2,\ldots, m_r$ roots respectively where

$$\sum m_i = m, \tag{3.3}$$

and those roots in the group of m_i are the m_i values of a series

$$y_1(0)+p_{i1}z+p_{i2}z^2+\ldots \tag{3.4}$$

corresponding to the m_i different values of z defined by

$$z = x^{(m_i^{-1})} \qquad z^{\frac{1}{m_i}} \tag{3.5}$$

Note: (i) Again any of the series (3.4) may terminate. (ii) We may have $r = 1$, in which case $m_1 = m$ and all m roots are given by the same fractional power series. (iii) We may have $m_i = 1$ for some i, in which case the corresponding power series (3.4) does not involve fractional powers. (iv) All m of the roots tend to $y_1(0)$ as x tends to zero.

Numerical examples

4. (i) $f(x, y) = y^2(1+x)-3y(1+x^2)+(2+x)$. We have

$$f(0, y) = y^2-3y+2.$$

Hence, for example, $y_1(0) = 1$ is a simple root. The corresponding root $y_1(x)$ is given by

$$y_1(x) = [3(1+x^2)-(1-12x+14x^2+9x^4)^{\frac{1}{2}}]/2(1+x).$$

This may obviously be expanded as a convergent power series in x provided $12\,|x|+14\,|x|^2+9\,|x|^4 < 1$.

(ii) $f(x, y) = y^3 - y^2 - x(1+x)^2$. We have $f(0, y) = y^3 - y^2$, so that $y_1(0) = 1$ is a simple root and $y_2(0) = 0$ is a double root. Clearly we have

$$y_1(x) = 1+x,$$

so that the convergent series in x terminates. Note that the presence of the double root has not led to fractional powers in $y_1(x)$.

The roots corresponding to $y_2(0) = 0$ are the solutions of

$$y^2 + yx + x(1+x) = 0,$$

and are therefore given by the two values of

$$y = \tfrac{1}{2}[-x+(-4x-3x^2)^{\frac{1}{2}}]$$

$$= \tfrac{1}{2}[-x+2ix^{\frac{1}{2}}(1+\tfrac{3}{4}x)^{\frac{1}{2}}].$$

The power series therefore contains only the odd powers of $x^{\frac{1}{2}}$; it is convergent if $|x| < \frac{4}{3}$.

(iii) $f(x, y) = y^4 - y^2 x(1+x)^2 + x^3(1+x)^2$. The equation $f(0, y) = 0$ has the quadruple root $y = 0$. The roots of $f(x, y) = 0$ are given by

$$y = x^{\frac{1}{2}}(1+x)^{\frac{1}{2}}, \quad y = x(1+x)^{\frac{1}{2}}, \quad y = -x(1+x)^{\frac{1}{2}}.$$

In the notation of Theorem 2 we have $m_1 = 2$, $m_2 = 1$, $m_3 = 1$.

Note that although the third polynomial has a quadruple zero, the perturbations in the coefficients of the powers of y are so related that none of the perturbations in the roots is of order $x^{\frac{1}{4}}$. Related perturbations of this kind are common in the eigenvalue theory.

Perturbation theory for simple eigenvalues

5. Consider now two matrices A and B satisfying relations (2.1), and let λ_1 be a simple eigenvalue of A. We wish to examine the corresponding eigenvalue of $(A+\epsilon B)$. Let us denote the characteristic equation of A by

$$\det(\lambda I - A) \equiv \lambda^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \ldots + c_0 = 0. \tag{5.1}$$

Then the characteristic equation of $(A+\epsilon B)$ is given by

$$\det(\lambda I - A - \epsilon B) \equiv \lambda^n + c_{n-1}(\epsilon)\lambda^{n-1} + c_{n-2}(\epsilon)\lambda^{n-2} + \ldots + c_0(\epsilon) = 0, \tag{5.2}$$

where $c_r(\epsilon)$ is a polynomial of degree $(n-r)$ in ϵ such that

$$c_r(0) = c_r. \tag{5.3}$$

This is immediately obvious if we examine the explicit expression for $\det(\lambda I - A - \epsilon B)$. We may write

$$c_r(\epsilon) = c_r + c_{r1}\epsilon + c_{r2}\epsilon^2 + \ldots + c_{r,n-r}\epsilon^{n-r}. \tag{5.4}$$

Now since λ_1 is a simple root of (5.1) we know from Theorem 1 of § 3 that for sufficiently small ϵ there is a simple root $\lambda_1(\epsilon)$ of (5.2), given by a convergent power series

$$\lambda_1(\epsilon) = \lambda_1 + k_1\epsilon + k_2\epsilon^2 + \ldots. \tag{5.5}$$

Clearly $\lambda_1(\epsilon) \to \lambda_1$ as $\epsilon \to 0$. Note that

$$|\lambda_1(\epsilon) - \lambda_1| = 0(\epsilon) \tag{5.6}$$

independent of the multiplicities of other eigenvalues.

Perturbation of corresponding eigenvectors

6. Turning now to the perturbation in the corresponding eigenvector, we first develop explicit expressions for the components of an eigenvector x_1 corresponding to a simple eigenvalue λ_1 of A. Since λ_1 is a simple eigenvalue, $(A-\lambda_1 I)$ has at least one non-vanishing minor of order $(n-1)$. Suppose, without loss of generality, that this lies in the first $(n-1)$ rows of $(A-\lambda_1 I)$. Then from the theory of linear equations the components of x may be taken to be

$$(A_{n1}, A_{n2}, \ldots, A_{nn}), \tag{6.1}$$

where A_{ni} denotes the cofactor of the (n, i) element of $(A-\lambda_1 I)$, and hence is a polynomial in λ_1 of degree not greater than $(n-1)$.

We now apply this result to the simple eigenvalue $\lambda_1(\epsilon)$ of $(A+\epsilon B)$. We denote the eigenvector of A by x_1 and the eigenvector of $(A+\epsilon B)$ by $x_1(\epsilon)$. Clearly the elements of $x_1(\epsilon)$ are polynomials in $\lambda_1(\epsilon)$ and ϵ, and since the power series for $\lambda_1(\epsilon)$ is convergent for all sufficiently small ϵ we see that each element of $x_1(\epsilon)$ is represented by a convergent power series in ϵ, the constant term in which is the corresponding element of x_1. We may write

$$x_1(\epsilon) = x_1 + \epsilon z_1 + \epsilon^2 z_2 + \ldots, \tag{6.2}$$

where each component of the vector series on the right is a convergent power series in ϵ. Corresponding to the result of (5.6) for the eigenvalue, we have for the eigenvector the result

$$|x_1(\epsilon) - x_1| = 0(\epsilon), \tag{6.3}$$

and again there are no fractional powers of ϵ.

Matrix with linear elementary divisors

7. If the matrix A has linear elementary divisors, then we may assume the existence of complete sets of right-hand and left-hand eigenvectors $x_1, x_2,\ldots, x_n$ and $y_1, y_2,\ldots, y_n$ respectively, such that

$$y_i^T x_j = 0 \quad (i \neq j), \tag{7.1}$$

though these vectors are unique only if all the eigenvalues are simple.

Expressing each of the vectors z_i of (6.2) in terms of the x_j in the form

$$z_i = \sum_{j=1}^{n} s_{ji} x_j, \tag{7.2}$$

we have

$$x_1(\epsilon) = x_1 + \epsilon \sum_{j=1}^{n} s_{j1} x_j + \epsilon^2 \sum_{j=1}^{n} s_{j2} x_j + \ldots, \tag{7.3}$$

and collecting together the terms in the vector x_i

$$x_1(\epsilon) = (1+\epsilon s_{11}+\epsilon^2 s_{12}+\ldots)x_1 + (\epsilon s_{21}+\epsilon^2 s_{22}+\ldots)x_2 + \\ +\ldots+(\epsilon s_{n1}+\epsilon^2 s_{n2}+\ldots)x_n. \tag{7.4}$$

The convergence of the n power series in the brackets is a simple consequence of the absolute convergence of the series in (6.2).

We are not interested in a multiplying factor in $x_1(\epsilon)$ and we may divide by $(1+\epsilon s_{11}+\epsilon^2 s_{12}+\ldots)$ since this is non-zero for sufficiently small ϵ. We may then redefine the new vector as $x_1(\epsilon)$. This gives

$$x_1(\epsilon) = x_1 + (\epsilon t_{21}+\epsilon^2 t_{22}+\ldots)x_2 + \ldots + (\epsilon t_{n1}+\epsilon^2 t_{n2}+\ldots)x_n, \tag{7.5}$$

where the expressions in brackets are still convergent power series for sufficiently small ϵ.

This relation has been obtained under the assumption that the components of the x_i were obtained directly from their determinantal expressions. However, if we replace these x_i by new x_i which are normalized so that

$$\|x_i\|_2 = 1, \tag{7.6}$$

this merely introduces a constant factor into each of the expressions in brackets and this may be absorbed into the t_{ij}. Equation (7.5) then holds for normalized x_i though $x_i(\epsilon)$ will not, of course, be normalized for $\epsilon \neq 0$.

First-order perturbations of eigenvalues

8. We now derive explicit expressions for the first-order perturbations in terms of the x_i and y_i. We first introduce quantities s_i which will be used repeatedly in subsequent analysis; they are defined by the relations

$$s_i = y_i^T x_i \quad (i = 1, 2,\ldots, n), \tag{8.1}$$

where the y_i and x_i are *normalized* left-hand and right-hand vectors. If y_i and x_i are real, then s_i is the cosine of the angle between these vectors.

If there are any multiple eigenvalues the corresponding x_i and y_i will not be uniquely determined; in this case we assume that the s_i we are using correspond to some particular choice of the x_i and y_i. Even when x_i and y_i correspond to a simple λ_i they are arbitrary to the extent of a complex multiplier of modulus unity, but in this case $|s_i|$ *is* fully determined.

We have in any case

$$|s_i| = |y_i^T x_i| \leqslant \|y_i\|_2 \|x_i\|_2 = 1. \tag{8.2}$$

We define also quantities β_{ij} by the relations

$A+\varepsilon B$

$$\beta_{ij} = y_i^T B x_j \tag{8.3}$$

and, since $\|B\|_2 \leqslant n$, we have

$$|\beta_{ij}| = |y_i^T(Bx_j)| \leqslant \|y_i\|_2 \|Bx_j\|_2 \leqslant \|B\|_2 \|y_i\|_2 \|x_j\|_2 \leqslant n. \tag{8.4}$$

9. From their definitions we have

$$(A+\epsilon B)x_1(\epsilon) = \lambda_1(\epsilon) x_1(\epsilon), \tag{9.1}$$

and since $\lambda_1(\epsilon)$ and all components of $x_1(\epsilon)$ are represented by convergent power series we may equate terms in the same power of ϵ on each side of this equation. Equating terms in ϵ gives, from (5.5) and (7.5)

$\lambda_1(\varepsilon) = \lambda_1 + k_1\varepsilon$

$$A\Big(\sum_{i=2}^{n} t_{i1} x_i\Big) + Bx_1 = \lambda_1\Big(\sum_{i=2}^{n} t_{i1} x_i\Big) + k_1 x_1, \tag{9.2}$$

$x_1(\varepsilon) = x_1 + \varepsilon t_{21} x_2 \cdots + \varepsilon t_{n1} x_n$

or

$$\sum_{i=2}^{n} (\lambda_i - \lambda_1) t_{i1} x_i + Bx_1 = k_1 x_1. \tag{9.3}$$

Pre-multiplying this by y_1^T and remembering that $y_1^T x_i = 0$, $(i \neq 1)$, we obtain

$$k_1 = y_1^T B x_1 / y_1^T x_1 = \beta_{11}/s_1, \tag{9.4}$$

and hence

$$|k_1| \leqslant n/|s_1| \quad \text{from (8.4)}. \tag{9.5}$$

For sufficiently small ϵ the main term in the perturbation of λ_1 is $k_1\epsilon$, so we see that the sensitivity of this eigenvalue is primarily dependent on s_1. Unfortunately an s_i may be arbitrarily small.

First-order perturbations of eigenvectors

10. Pre-multiplying (9.3) by y_i^T we have

$$(\lambda_i - \lambda_1) t_{i1} s_i + \beta_{i1} = 0 \quad (i = 2, 3,\ldots, n), \tag{10.1}$$