

## PARTIAL DIFFERENTIAL EQUATIONS FOR EIGENVALUES: SENSITIVITY AND PERTURBATION ANALYSIS

ROY B. LEIPNIK<sup>1</sup>

(Received 30 March 1989; revised 3 February 1992)

### Abstract

The well-known Wilkinson expressions for the first derivatives of (ordinary) eigenvalues and eigenvectors of simple matrices, in terms of the set of eigenvalues and eigenvectors, are redifferentiated and combined to obtain partial differential equations for the eigenvalues. Analogous expressions are obtained for the first derivatives of generalised eigenvalues and eigenvectors of simple pairs of matrices  $(A, B)$ , defined by  $(A - \lambda_j B)x_j = 0$ ,  $y_j^T(A - \lambda_j B) = 0$ . Again, redifferentiation and combination yields slightly more complicated partial differential equations for the generalised eigenvalues. When the matrices depend on a few parameters  $\theta_1, \theta_2, \dots$ , the resulting differential equations for the eigenvalues, with those parameters as independent variables, can easily be derived. These parametric equations are explicit representations of analytic perturbation results of Kato, expressed by him as rather abstract complex matrix integrals. Connections with bounds for eigenvalues derived by Stewart and Sun can also be made. Two applications are exhibited, the first being to a broken symmetry problem, the second being to working out the second-order perturbations for a classical problem in the theory of waves in cold plasmas.

### Introduction

The works of Fourier and Rayleigh underlined the central role of eigenvalues (frequencies) and eigenvectors (modes) in classical physics and engineering. The dilemma of whether to calculate eigenvalues through characteristic equations (highly non-linear) or eigenvectors (often tedious) remains. The mode structure is sometimes of great physical importance (at least for the first few modes), but often it is not. In quantum mechanics, the eigenvalues provide energy levels, measurable by spectroscopy, etc., and the eigenvectors

---

<sup>1</sup>University of California at Santa Barbara.

© Copyright Australian Mathematical Society 1993, Serial-fee code 0334-2700/93

provide electronic densities or (after integration) probabilities, usually not measurable. The characteristic equation involves infinite determinants, so both approaches have drawbacks.

Some relief is offered by perturbation theory. The eigenvectors of the perturbed operator are expanded in series of the eigenvectors of the unperturbed operator. The eigenvalues  $\lambda'_j$  of the perturbed operator are approximated by series whose terms have denominators containing  $\lambda_j - \lambda_l$  and numerators involving dot products of eigenvectors, all summed over  $l \neq j$ . This of course is for operators with discrete spectra, especially simple operators (without multiple eigenvalues). The continuous part of the spectrum, if present, requires more sophisticated expressions. The point is that a few sets of eigenvectors can be used for many eigenvalue problems via perturbations. The situation for finite matrices is simpler than for Schrödinger equations, but paradoxically was not exploited until the 1950's, even though Schrödinger and others had used perturbations in the 1920's.

The question as to whether the changes in eigenvalues can be completely expressed in terms of eigenvalues of the unperturbed operator, without using the eigenvectors (or without seeming to use them) is natural. This depends somewhat on the number of parameters being simultaneously perturbed, from one to  $n^2$ , for an  $n \times n$  matrix. In Section 1, basic perturbation theory is summarised (after Kato [10]), including the relation between parametric perturbations and non-parametric (total) perturbations as in Wilkinson [20].

In Section 2, the well-known sensitivity formulas of Wilkinson are derived, which relate eigenvalue derivatives and eigenvector derivatives to eigenvalues and eigenvectors. One may hope that by continuing this process, the eigenvectors can somehow be eliminated. By suitable manipulations of sextuple products, this is accomplished for ordinary eigenvalues of simple matrices, yielding bilinear partial differential equations. In Section 3, these equations are completely verified for  $2 \times 2$  matrices. Redeveloping the equations for the case of parametric dependence of perturbations is the subject of Section 4.

Section 5 is devoted to analogs of the Wilkinson formulas for generalised eigenproblems  $(A - \lambda_j B)x_j = 0$ ,  $y_j^T(A - \lambda_j B) = 0$ . The results are just a little more complicated than those of Wilkinson. A useful preliminary is the theory of matrix pencils as presented in Stewart and Sun [16], especially for simple matrix pencils. In Section 6, the methods of Section 2 are reapplied to the results of Section 5, obtaining partial differential equations rather like those of Section 2, but for generalised eigenvalues.

Section 7 contains two applications. The first is a symmetry-breaking perturbation of a  $3 \times 3$  matrix, in which everything can be calculated explicitly, but which is too simple to be more than an instructive exercise. The second

is to perturbations of the wave theory of cold plasmas, a substantive question when it is second-order perturbations. Here the results are applicable, and quite likely new, showing the real usefulness of the results. Section 8 shows how the older formula of Rogers for sensitivities relates to that of Wilkinson, which would permit another path to our formulae.

The most likely special applications of these formulae are in quantum physics (when extended to infinite matrices), order-disorder calculations, potential-theory expansions, linear system synthesis, mathematical statistics, and statistical physics. These formulas permit eigenvalue perturbations of all orders to be calculated directly from a knowledge of *all* the unperturbed eigenvalues; in practice, some can be discarded.

Specific problems which look attractive are the Stark effect (done in the 1930's to third order, by great efforts), the energy levels of positronium, the Ising model in an external magnetic field, potential theory of ellipsoidal geometries, amorphous alloys, and nuclear energy levels based on eigenvalues of random matrices. Non-physical applications include the effect of round-off on eigenvalues, and the P. D. Lax theory of invariant spectra (important in soliton theory), which should be reachable via convergent sequences of circulant/continuant matrices of increasing size. In electrical engineering, finding matrices with specified eigenvalue properties should be simplified. Finally, matrix-associated distributions such as the Wishart and A. T. James invariant distributions can be investigated in this way. Since the equations of Sections 2 and 6 give eigenvalues a new twist, other applications should appear.

### **1. The basic perturbation results of Kato for simple and semi-simple matrices**

The theory of matrix perturbations was developed especially by Kato [10] and Householder [9] during 1949–1960 and by Wilkinson [20] during 1950–1965, following early work of Rellich, Friedrichs, A. J. Hoffman/Wielandt, Ostrowski, Taussky, etc. Kato's results are rather abstract, but contain bounds needed for convergence proofs, and use very efficient definitions based on complex-variable theory. Wilkinson's theory is completely nonparametric, depends on the whole matrix, is robust and rather explicit, and is designed for numerical analysis. His so-called sensitivity formulas are shown below to be derivable by Kato's methods, although Wilkinson's technique is more expedient for simple matrices (those which possess distinct eigenvalues). Extensions of our results to infinite-dimensional linear operators would require Kato's methods, it would seem. But even a proper understanding of our

finite dimensional results benefits from Kato’s theory, especially why we restrict ourselves to simple matrices here.

The important book of Stewart and Sun [16], which appeared after the first version of this paper was submitted, contains perturbation bounds for both ordinary and generalised eigenvalues, but does not attempt to improve the results of Wilkinson. The results derived here, though originally obtained from Wilkinson’s formulas by extensive manipulation, can be extracted from some operator power series of Kato by analogs of the Cauchy integral formula for derivatives of analytic functions.

The objective is to calculate 2nd, 3rd, 4th order perturbations of the eigenvalues from the 1st order perturbations and the unperturbed eigenvalues. Rather surprisingly, the second-order derivatives of eigenvalues with respect to matrix elements (or with respect to parameters within the elements) are bilinear functionals of the first-order derivatives, with coefficients dependent on the unperturbed eigenvalues. The system is autonomous. By repeated differentiation, all existing higher-order derivatives can be expressed multilinearly in terms of the first-order derivatives. In most applications, second-order corrections are sufficient, fortunately.

Similar results are derived for generalised eigenproblems, after analogues of Wilkinson’s formulas are derived. Also, eigenvalue perturbations can be partially eliminated from eigenvector perturbations. That effort is not carried out here.

The 1982 book of Kato [10] is an effective introduction, answering Wilkinson’s plea [20, p. 109] for a brief, rigorous treatment. (Rellich’s counterexamples show that *simplicity* is absent.) The account below is drawn from Kato [10], with some help from the recent books of Bhatia [4], and Stewart and Sun [16].

Let  $T$  map a vector space  $X$  into a finite-dimensional vector space  $Y \subset X$ . If  $\lambda$  is a scalar of  $X$ , let  $G_\lambda$  be the subspace of eigenvectors of  $T$  associated with  $\lambda$  (if there are any) and let  $G_\lambda = 0$  otherwise. If  $\zeta \neq \lambda_1, \lambda_2, \dots, \lambda_n$  the eigenvalues of  $T$ , then  $R(\zeta) = (T - \zeta I)^{-1}$  is the resolvent, an operator-valued function of  $\zeta$  and  $T$ .

If  $\lambda_h$  is an eigenvalue of  $T$ , a Laurent-type expansion holds, of the form:

$$R(\zeta) = -(\zeta - \lambda_h)^{-1} P_h - \sum_{n=0}^{\infty} (\zeta - \lambda_h)^n S_h^{n+1} - \sum_{n=1}^{\infty} (\zeta - \lambda_h)^{-n-1} D_h^n. \tag{1.1}$$

(This is the key to Kato’s presentation.) Here  $P_h$  is the eigenprojection,  $S_h$  is the reduced resolvent, and  $D_h$  is the eigen-nilpotent of  $T$  for  $\lambda_h$ , and the

series converge in any matrix norm for  $\zeta$  in a ring  $0 < |\zeta - \lambda_h| < \delta_h$  about  $\lambda_h$ . If  $\Gamma_h$  is a piecewise smooth closed curve about  $\lambda_h$  of small enough diameter, then

$$P_h = (i/2\pi) \int_{\Gamma_h} R(\zeta) d\zeta. \quad (1.2)$$

The usual consequences are

$$P_h P_k = \delta_{hk} P_k, \quad D_h D_k = 0, \quad (1.3)$$

$$D_h = P_h(T - \lambda_h I) = (T - \lambda_h I)P_h, \quad T P_h = P_h T = P_h T P_h.$$

Collectively, if  $E = \sum_h \lambda_h P_h$  and  $D = \sum_h D_h$ , then

$$T = E + D, \quad D^2 = 0, \quad ED = DE, \quad \text{and} \quad \sum_h P_h = I. \quad (1.4)$$

If  $D_h = 0$ , then  $\lambda_h$  is called semi-simple. If all  $\lambda_h$  are semi-simple, then  $T$  is called semi-simple or diagonalizable. The space  $M_h = \text{range } P_h = P_h X$  is called the algebraic eigenspace of  $\lambda_h$ , and the integer  $m_h = \dim M_h$  is the algebraic multiplicity of  $\lambda_h$ . If  $m_h = 1$ ,  $\lambda_h$  is called simple, and if all  $\lambda_h$  are simple,  $T$  is also called simple.

If  $T_0, T_1$  are finite-dimensional linear operators on  $X$ , then  $T(\chi) = T_0 + \chi T_1$  is called a pencil or first-order family of perturbations of  $T_0$ , for  $\chi$  in a complex domain  $D$  or real interval  $J$ . More generally, if  $T_0, T_1, T_2, \dots$  is a sequence of finite-dimensional operators, and

$$T(\chi) = T_0 + \sum_{n=1}^{\infty} \chi^n T_n \quad (1.5)$$

converges in  $D$  (or  $J$ ), then  $T(\chi)$  defines an analytic family of perturbations. The "family" eigenvalues  $\zeta(\chi)$  satisfy

$$\det[T(\chi) - \zeta(\chi)I] = 0. \quad (1.6)$$

If each  $T(\chi)$  is  $N \times N$ , (1.5) is algebraic of degree  $N$ , with coefficients analytic in  $\chi$ . Hence the  $\zeta(\chi)$  are branches of functions analytic in  $\chi$ , except for algebraic singularities. The number of such branches  $\zeta(\chi)$  is a constant  $s \leq N$  for all  $\chi$ , except for a finite number of "exceptional" points in each compact subset of  $D$  (or  $J$ ). If all branches are distinct, then  $s = N$ , and  $T(\chi)$  is simple for all "non-exceptional"  $\chi$ . For  $N = 2$ , six possibilities for exceptional  $\chi$  are shown on page 74 of Kato.

If  $K$  is a simply-connected subdomain of  $D$ , not containing an exceptional point, the eigenvalues can be written as  $\lambda_j(\chi)$ ,  $1 \leq j \leq s$ , where all the  $\lambda_j(\chi)$  are analytic in  $K$ . If however  $\chi_0$  is an exceptional point, then the  $\lambda_j(\chi)$  are still analytic in any small disk outside of  $\chi_0$ . Kato shows how Puiseux (fractional) expansions of  $\lambda_j(\chi)$  are obtained about  $\chi = \chi_0$ .

The two-parameter family resolvent

$$R_2(\zeta, \chi) = (T(\chi) - \zeta I)^{-1} \tag{1.7}$$

is analytic in the two-parameter set:  $\{(\zeta, \chi) : \zeta \neq \text{any eigenvalue of any } T(\chi)\}$ . If  $R_1(\zeta) = R_2(\zeta, \chi_0)$ , then

$$\begin{aligned} R_2(\zeta, \chi) &= R_1(\zeta)[I + (T(\chi) - T(\chi_0))R_1(\zeta)]^{-1} \\ &= R_1(\zeta) \sum_{p=0}^{\infty} [(T(\chi_0) - T(\chi))R_1(\zeta)]^p \end{aligned} \tag{1.8}$$

for  $\|T(\chi_0) - T(\chi)\| \|R_1(\zeta)\| < 1$ . If  $T(\chi_0) = T_0$ , and  $T(\chi)$  is analytic around  $\chi = \chi_0$ , then the series (1.8) in powers of  $T(\chi_0) - T(\chi)$  can, by a matrix version of Lagrange’s composite expansion rule, be rewritten as a power series in  $\chi$ . Let

$$R_1^{(n)}(\zeta) = \sum_p \sum_{\nu_1, \dots, \nu_p} (-1)^p R_1(\zeta) [T^{\nu_1} R_1(\zeta)] \cdots [T^{\nu_p} R_1(\zeta)] \tag{1.9}$$

summed over all integers  $p$ , and integers  $\nu_1, \dots, \nu_p$  such that  $1 \leq p \leq n$ ,  $\nu_j \geq 1$ ,  $\nu_1 + \dots + \nu_p = n$ . Then

$$R_2(\zeta, \chi) = R_1(\zeta) + \sum_{n=1}^{\infty} \chi^n R_1^{(n)}(\zeta) \tag{1.10}$$

on re-expanding (1.6) in power series in  $\chi$ . The pattern of two types of terms in (1.9), the “opening term”  $R_1(\zeta)$  and the various “closing terms”  $T^\nu R_1(\zeta)$ , can be discerned in a coordinatised analysis, if higher-order derivatives are calculated. The point is that the eigenprojections can be calculated from integrating the resolvent, and the first derivative of an eigenvalue with respect to its matrix is (the transpose of) its corresponding eigenprojection. Thus the structure of the resolvent expansion is reflected in the structure of the eigenvalue expansion. Some of these results extend to infinite-dimensional operators, especially when the operators are Hermitian, with discrete spectrum, as in many quantum mechanical applications.

Similar formulas for the eigenvectors, the Jordan form (in non-diagonable cases) and for group eigenprojections, etc., are found in [10, pages 77–96].

An important application is to the Motzkin-Taussky theorems:

- (i) if  $T(\chi) = T_0 + \chi T_1$  is diagonable for all complex  $\chi$ , then all eigenvalues of  $T(\chi)$  have the form  $\lambda_0 + \chi \lambda_1$ ,
- (ii) the eigenprojections are entire functions of  $\chi$ ,
- (iii) if  $T_1$  is also diagonable, then  $T_0 T_1 = T_1 T_0$ .

Eigenprojections are more stable numerically than eigenvectors. This justifies the seemingly inverted procedure of determining eigenprojections  $P_1(\chi)$ ,

$P_2(\chi), \dots$  and taking a basis  $\{\phi_1, \dots, \phi_N\}$  of  $X$  to produce eigenvectors  $\phi_{hk}(\chi) = P_h(\chi)\phi_k$ . This double sequence yields (sometimes) zero vectors and (sometimes) duplicates or linear combinations of previous eigenvectors. Such anomalies can be removed automatically by Kato's elegant transformation procedure, which expresses  $P_h(\chi)$  as  $UP_h(\chi_0)U^{-1}$ , where  $U$  is a function of the  $P_h(\chi)$  and  $(d/d\chi)P_h(\chi)$  for  $h = 1, 2, \dots$ . This is numerically impractical, but is formally useful in verifying the correctness of the basic results of this paper when the perturbations depend holomorphically on a single parameter. For more parameters similar formulas may be found but are even more impractical.

Leaving the holomorphic case, one first encounters non-convergent perturbation series, then bounded-order derivatives, and finally non-parametric (Wilkinson-type) perturbations, of numerical analysis interest. When  $T(\chi)$  is merely continuous in  $\chi$ , then  $R_2(\zeta, \chi)$  is jointly continuous in the set where all points such that  $\zeta$  is an eigenvalue of  $T(\chi)$  are *deleted*. It follows that eigenvalues and "joint" eigenprojections  $P_\Gamma(\chi)$  are continuous, where  $\Gamma$  is a closed curve enclosing several eigenvalues of  $T(\chi)$ . The number of distinct eigenvalues of  $T(\chi)$  may vary with  $\chi$ ; therefore the eigenvalues are often listed with their multiplicities in a set  $\mathcal{S}(\chi)$  of  $N$  elements.

The "distance" between  $\mathcal{S}(\chi) = \{\mu_j\}$  and  $\mathcal{S}(\chi') = \{\mu'_j\}$  is defined as  $\min_\Pi \max_n |\mu_{\Pi(n)} - \mu'_n|$ , where  $\Pi$  runs over all permutations of  $\{1, 2, \dots, N\}$ . This distance, though not a metric, is continuous. A stronger result is "separate continuity": the existence of  $N$  eigenvalue functions,  $\mu_j(\chi)$ , which are all continuous. If  $\chi$  is confined to the reals, or if all  $\mu_j(\chi)$  are real, then separate continuity is valid. Continuity of eigenprojections holds except at cross-overs between the separate eigenvalues. Even in the holomorphic case, eigenprojection poles can appear at exceptional points. But when  $T(\chi)$  is infinitely differentiable, the eigenprojections and eigenvectors may still have discontinuities of the second ( $\sin[1/(\chi - \chi_0)]$ ) kind, if  $\chi_0$  is a point of non-analyticity.

On the optimistic side, if  $T(\chi)$  is differentiable, then so are group projections and eigenvalues. If also  $T(\chi_0)$  is diagonalizable, then the separable eigenvalues  $\mu_j(\chi)$  are differentiable at  $\chi = \chi_0$ , thus defining  $\mathcal{S}'(\chi_0)$ .

New pathologies may arise with two parameters,  $(\chi_1, \chi_2)$ , in that totally differentiable, diagonalizable  $T(\chi_1, \chi_2)$  does *not* imply *totally* differentiable eigenvalues  $\lambda_j(\chi_1, \chi_2)$ , only *partially* differentiable eigenvalues. If the number of distinct eigenvalues is fixed (as when  $T(\chi_1, \chi_2)$  is *simple*), total differentiability is restored.

When all the entries can be varied (as in Wilkinson [20]), the eigenvalues become functions of  $T$ , in effect. Then  $\mathcal{S} = \mathcal{S}(T)$  for the repeated eigenvalues, with four important conclusions:

- (i)  $\mathcal{S}(T)$  is a continuous function of  $T$ ;
- (ii)  $\mathcal{S}(T)$  is partially differentiable at  $T = T_0$  if and only if  $T_0$  is diagonalable;
- (iii) if  $T_0$  is simple, then  $T = T_0 + T'$  is simple for  $\|T'\| < \delta$ ;
- (iv) if  $T_0$  is simple, then the  $\lambda_j(T)$  are locally holomorphic in  $T$  (a technical operator-based concept dependent on multilinear expansions), and thus separately continuous and totally differentiable for  $\|T - T_0\| < \delta$ .

It is (iii) and (iv), together with various counterexamples, which motivate the restriction to simple matrices in the sequel. Kato suggests that (iv) may be extended to diagonalable  $T_0$  by some kind of fractional Puiseux concepts, and references the work of Baumgärtel [3].

Eigenvalue derivatives of the second order, as considered in Section 2 et seq. are related to eigenvalue perturbations through the Taylor expansion theorem.

$$\begin{aligned} \lambda_j(T_0 + T') &= \lambda_j(T_0) + \sum_{mn} (\partial\lambda_j/\partial a_{mn})_0 a'_{mn} \\ &\quad + \frac{1}{2} \sum_{mn} \sum_{pq} [\partial^2\lambda_j/(\partial a_{mn}\partial a_{pq})] T_1 a'_{mn} a'_{pq} \end{aligned} \tag{1.11}$$

where  $T_1 = (a_{mn})_1$ ,  $T_0 = (a_{mn})_0$ ,  $T' = (a_{mn})'$  and  $(a_{mn})_0 \leq (a_{mn})_1 \leq (a_{mn})_0 + (a_{mn})'$  for all  $m, n$ . Kato infers that

$$\begin{aligned} &\left| \lambda_j(T_0 + T') - \lambda_j(T_0) - \sum_{mn} (\partial\lambda_j/\partial a_{mn})_0 a'_{mn} \right. \\ &\quad \left. - \frac{1}{2} \sum_{mn} \sum_{pq} [\partial^2\lambda_j/(\partial a_{mn}\partial a_{pq})]_0 a'_{mn} a'_{pq} \right| = o\left(\sum_{mn} |a'_{mn}|^2\right). \end{aligned} \tag{1.12}$$

Higher-order perturbations behave similarly. Clearly  $\lambda_j(T_0)$ ,  $(\partial\lambda_j/\partial T)_0$ , and the formula (2.38) for the second derivatives, derived below, produces the second-order perturbations.

## 2. Wilkinson sensitivity formulas and partial differential equations for eigenvalues

Suppose that  $A$  is a simple  $N \times N$  matrix with eigenvalues  $\lambda_1, \dots, \lambda_N$ , and complete sets of right (column) eigenvectors  $\mathbf{x}_1, \dots, \mathbf{x}_N$  and left (row) eigenvectors  $\mathbf{y}_1^T, \dots, \mathbf{y}_N^T$ . Without loss of generality, the bi-orthogonal conditions



$$\mathbf{y}_k^\top \mathbf{x}_j = \sum_t y_{kt} x_{jt} = s_k \delta_{jk} \tag{2.1}$$

can be assumed. Obviously,

$$s_k = \sum_m y_{km} x_{km} \tag{2.2}$$

where  $y_{km} = (\mathbf{y}_k)_m$  and  $x_{km} = (\mathbf{x}_k)_m$  are the components of  $\mathbf{y}_k$  and  $\mathbf{x}_k$ .

Suppose now that  $A' = A + \varepsilon \Delta_1 A$  is a perturbed matrix to first order and formally write

$$\begin{aligned} \lambda'_j &= \lambda_j + \varepsilon \Delta_1 \lambda + O(\varepsilon^2), \\ \mathbf{x}'_k &= \mathbf{x}_k + \varepsilon \Delta_1 \mathbf{x}_k + O(\varepsilon^2), \\ \mathbf{y}'_j &= \mathbf{y}_j + \varepsilon \Delta_1 \mathbf{y}_j + O(\varepsilon^2). \end{aligned} \tag{2.3}$$

as the eigenvalues and eigenvectors of  $A'$ . Hence formally

$$\begin{aligned} (A + \varepsilon \Delta_1 A)(\mathbf{x}_j + \varepsilon \Delta_1 \mathbf{x}_j + O(\varepsilon^2)) \\ = (\lambda_j + \varepsilon \Delta_1 \lambda_j + O(\varepsilon^2))(\mathbf{x}_j + \varepsilon \Delta_1 \mathbf{x}_j + O(\varepsilon^2)). \end{aligned} \tag{2.4}$$

The first-order corrections satisfy

$$(\Delta_1 A)\mathbf{x}_j + A(\Delta_1 \mathbf{x}_j) = \lambda_j(\Delta_1 \mathbf{x}_j) + (\Delta_1 \lambda_j)\mathbf{x}_k + O(\varepsilon). \tag{2.5}$$

Left multiplication of (1.5) by  $\mathbf{y}_k^\top$  yields

$$\begin{aligned} \mathbf{y}_k^\top (\Delta_1 A)\mathbf{x}_j + \mathbf{y}_k^\top A(\Delta_1 \mathbf{x}_j) &= \mathbf{y}_k^\top (\Delta_1 A)\mathbf{x}_j + \lambda_k \mathbf{y}_k^\top (\Delta_1 \mathbf{x}_j) \\ &= \lambda_h \mathbf{y}_k^\top (\Delta_1 \mathbf{x}_j) + (\Delta_1 \lambda_j) \mathbf{y}_k^\top \mathbf{x}_j + O(\varepsilon) \\ &= \lambda_j \mathbf{y}_k^\top (\Delta_1 \mathbf{x}_j) + (\Delta_1 \lambda_j) \cdot (s_k \delta_{jk}) + O(\varepsilon). \end{aligned} \tag{2.6}$$

If  $j = k$ , then (2.6) becomes

$$\mathbf{y}_j^\top (\Delta_1 A)\mathbf{x}_j + \lambda_j \mathbf{y}_j^\top \Delta_1 \mathbf{x}_j = \lambda_j \mathbf{y}_j^\top \Delta_1 \mathbf{x}_j + s_j \Delta_1 \lambda_j + O(\varepsilon). \tag{2.7}$$

Two terms cancel from (2.7), leaving

$$\begin{aligned} \mathbf{y}_j^\top (\Delta_1 A)\mathbf{x}_j &= s_j \Delta_1 \lambda_j + O(\varepsilon) \\ &= s_j \left[ \sum_{mn} (\partial \lambda_j / \partial a_{mn}) (\Delta_1 A)_{mn} + O(\|\Delta_1 A\|) \right] + O(\varepsilon) \end{aligned} \tag{2.8}$$

since by Kato [10] a total differential of  $\lambda_j$  exists in the simple case. Expansion of the left side of (2.8) as  $\sum_{mn} y_{jm} (\Delta_1 A)_{mn} x_{jn}$  permits the conclusion

that

$$s_j \partial \lambda_j / \partial a_{mn} = y_{jm} x_{jn}, \tag{2.9}$$

$$\partial \lambda_j / \partial a_{mn} = y_{jm} x_{jn} / s_j = y_{jm} x_{jn} / \left( \sum_p y_{jp} x_{jp} \right). \tag{2.10}$$

Thus (2.10) gives the first eigenvalue derivatives in terms of eigenvector components. More geometrically, the matrix  $M_j$  defined by  $(M_j)_{mn} = x_{jm} y_{jn}$  has the property that  $\sum_n (M_j)_{mn} (M_k)_{nq} = x_{jm} (\sum_n y_{jn} x_{kn}) y_{kq} = x_{jm} \delta_{jk} s_j y_{kq} = s_j \delta_{jk} (M_j)_{mq}$ , so that if  $P_j = M_j / s_j$ , then  $P_k P_k = \delta_{jk} P_j$ , and the set of  $P_j$ ,  $j = 1, \dots, N$ , is a set of projections. Read matrixially, (2.10) says that  $\partial \lambda_j / \partial A = P_j^T$ .

The same technique, carried out more elaborately when  $j \neq k$ , yields similar formulas for the eigenvectors. From (2.6),

$$y_k^T (\Delta_1 A) x_j + \lambda_k y_k^T \Delta_1 x_j = \lambda_j y_k^T \Delta_1 x_j + O(\epsilon) \text{ or} \tag{2.11}$$

$$y_k^T (\Delta_1 A) x_j = (\lambda_j - \lambda_k) y_k^T \Delta_1 x_j + O(\epsilon) \text{ or}$$

$$y_k^T \Delta_1 x_j = y_k^T (\Delta_1 A) x_j / (\lambda_j - \lambda_k) + O(\epsilon). \tag{2.12}$$

For a more informative result, note that  $\Delta_1 x_j$  is a specific vector in the column space and so has the decomposition, for some matrix  $F$ ,

$$\Delta_1 x_j = \sum_l f_{jl} x_l, \quad j = 1, 2, \dots, N, \tag{2.13}$$

because of completeness of the  $x_l$ . Multiplication of (2.13) by  $y_k^T$  leads to

$$y_k^T \Delta_1 x_j = \sum_l f_{jl} y_k^T x_l = \sum_l f_{jl} s_k \delta_{kl} = f_{jk} s_k \tag{2.14}$$

$$= y_k^T (\Delta_1 A) x_j / (\lambda_j - \lambda_k) + O(\epsilon).$$

Thus

$$f_{jk} = y_k^T (\Delta_1 A) x_j / [s_k (\lambda_j - \lambda_k)] + O(\epsilon) \tag{2.15}$$

and insertion into (2.13) yields

$$\Delta_1 x_j = \sum_{l \neq j} y_l^T (\Delta_1 A) x_j / [s_l (\lambda_j - \lambda_l)] x_l + O(\epsilon). \tag{2.16}$$

The total differentiability of the eigenvectors in the simple case (Kato [10]), following the previous reasoning for the eigenvalues, allows the conclusions

$$\partial x_j / \partial a_{mn} = \sum_{l \neq j} (y_{lm} x_{jn}) x_l / [s_l (\lambda_j - \lambda_l)], \tag{2.17}$$

$$\partial x_{jr} / \partial a_{mn} = \sum_{l \neq j} y_{jm} x_{jn} x_{lr} / [s_l (\lambda_j - \lambda_l)].$$

Such triple products as on the right of (2.17) become sextuples in the second derivatives.

Similar formulas for the left eigenvalues begin with

$$\begin{aligned}
 & (\mathbf{y}_k^\top + \varepsilon \Delta_1 \mathbf{y}_k^\top + O(\varepsilon^2))(A + \varepsilon \Delta_1 A) \\
 & = (\mathbf{y}_k^\top + \varepsilon \Delta_1 \mathbf{y}_k^\top + O(\varepsilon^2))(\lambda_k + \varepsilon \Delta_1 \lambda_k + O(\varepsilon^2)).
 \end{aligned}
 \tag{2.18}$$

Thus the analog of (2.5) is

$$\mathbf{y}_k^\top (\Delta_1 A) + (\Delta_1 \mathbf{y}_k^\top) A = \mathbf{y}_k^\top (\Delta_1 \lambda_k) + \lambda_k \Delta_1 \mathbf{y}_k^\top + O(\varepsilon).
 \tag{2.19}$$

Right multiplication by  $\mathbf{x}_j$  yields equations which for  $j = k$  are the same as (2.8). For  $j \neq k$ , the result is for right eigenvectors, in the simple matrix case

$$(\Delta_1 \mathbf{y}_k^\top) \mathbf{x}_j = (\mathbf{y}_k^\top \Delta_1 A) \mathbf{x}_j / (\lambda_k - \lambda_j) + O(\varepsilon) = -\mathbf{y}_k^\top \Delta_1 \mathbf{x}_j + O(\varepsilon).
 \tag{2.20}$$

Using the completeness of the  $\mathbf{y}_k^\top$  as before and the total differentiability for simple matrices, the equations

$$\begin{aligned}
 \partial \mathbf{y}_k^\top / \partial a_{mn} & = \sum_{l \neq k} y_{km} x_{ln} \mathbf{y}_l^\top / [s_l (\lambda_k - \lambda_l)], \\
 \partial y_{kr} / \partial a_{mn} & = \sum_{l \neq k} y_{km} x_{ln} y_{lr} / [s_l (\lambda_k - \lambda_l)]
 \end{aligned}
 \tag{2.21}$$

finally result. Obviously, they are again sums of triple products of vector components.

The next step is to redifferentiate and substitute in order to express

$$R_{j m n p q} = (\partial / \partial a_{pq})(\partial \lambda_j / \partial a_{mn}),
 \tag{2.22}$$

the second matrix derivatives of  $\lambda_j$ , in terms of the  $\lambda$ 's and their first derivatives. Now from (2.10), (2.17), (2.21),

$$\begin{aligned}
 R_{j m n p q} & = (\partial / \partial a_{pq})(y_{jm} x_{jn} / s_j) \\
 & = (1/s_j^2)[-y_{jm} x_{jn} (\partial / \partial a_{pq}) s_j + s_j x_{jn} (\partial / \partial a_{pq}) y_{jm} \\
 & \quad + s_j y_{jm} (\partial / \partial a_{pq}) x_{jn}]
 \end{aligned}
 \tag{2.23}$$

$$\begin{aligned}
 & = (1/s_j^2) \left\{ -y_{jm} x_{jn} (\partial / \partial a_{pq}) \sum_i y_{ji} x_{ji} \right. \\
 & \quad + \sum_i y_{ji} x_{ji} \sum_{l \neq j} y_{jp} x_{lq} y_{lm} x_{jn} / [s_l (\lambda_j - \lambda_l)] \\
 & \quad \left. + \sum_i y_{ji} x_{ji} \sum_{l \neq j} y_{lp} x_{jq} x_{ln} y_{jm} / [s_l (\lambda_j - \lambda_l)] \right\}.
 \end{aligned}
 \tag{2.24}$$

This does not apparently simplify until  $(\partial/\partial a_{pq}) \sum_t y_{jt} x_{jt}$  is obtained from (2.17), (2.21). Clearly

$$\begin{aligned}
 (\partial/\partial a_{pq}) \left( \sum_t y_{jt} x_{jt} \right) &= \sum_t \{ (\partial y_{jt} / \partial a_{pq}) x_{jt} + (\partial x_{jt} / \partial a_{pq}) y_{jt} \} \\
 &= \sum_t \sum_{l \neq j} \{ [y_{jp} x_{lq} y_{lt} x_{jt} + y_{lp} x_{iq} x_{lt} y_{jt}] / [s_l (\lambda_j - \lambda_l)] \} \tag{2.25}
 \end{aligned}$$

so that

$$\begin{aligned}
 R_{j m n p q} &= (1/s_j^2) \sum_t \sum_{l \neq j} \{ -y_{jp} x_{lq} y_{lt} x_{jt} y_{jm} x_{jn} - y_{lp} x_{jq} x_{lt} y_{jt} y_{jm} x_{jn} \\
 &\quad + [y_{jp} x_{lq} y_{jt} x_{jt} y_{lm} x_{jn} + y_{lp} x_{jq} y_{jt} x_{jt} x_{ln} y_{jm}] / [s_l (\lambda_j - \lambda_l)] \} . \tag{2.26}
 \end{aligned}$$

But  $\sum_t y_{lt} x_{jt} = \sum_t x_{lt} y_{jt} = 0$  for  $l \neq j$ , so from (2.26) the first two terms can be dropped after summation. Thus

$$\begin{aligned}
 R_{j m n p q} &= (1/s_j) \sum_{l \neq j} [y_{jp} x_{lq} y_{lm} x_{jn} + y_{lp} x_{jq} x_{ln} y_{jm}] / [s_l (\lambda_j - \lambda_l)] \\
 &= \sum_{l \neq j} [(y_{jm} x_{jn} / s_j) (y_{lm} x_{lq} / s_l) + (y_{lp} x_{ln} / s_l) (y_{jm} x_{jq} / s_j)] / (\lambda_j - \lambda_l) \\
 &= \sum_{i \neq j} [(\partial \lambda_j / \partial a_{pn}) (\partial \lambda_l / \partial a_{mq}) + (\partial \lambda_l / \partial a_{pn}) (\partial \lambda_j / \partial a_{mq})] / (\lambda_j - \lambda_l) \tag{2.27}
 \end{aligned}$$

on using (2.10) four times. This is the desired set of partial differential equations,  $N^5$  in all. The third-order partial derivatives of  $\lambda_j$  can similarly be expressed in terms of trilinear expressions in the first-order partial derivatives of triples of eigenvalues, with coefficients expressible in terms of differences of eigenvalues. The eigenvectors, modes or principal axes have been eliminated. The rest of the paper consists of verification, extensions, and applications of this startling formula, which for the case when only one parameter is allowed to vary, is a concrete version of a complex Cauchy-like matrix integral formula of Kato [10]. This connects the Wilkinson theory with the Kato theory, numerical matrix analysis with abstract functional analysis at a rather high level, via partial differential equations, which become ordinary differential equations in the case of one parameter. Because all the eigenvalues are coupled in these equations, they are useful when *all the unperturbed eigenvalues* are known as an initial condition. There are a few special but important

cases where this knowledge is in fact available. Also, the eigenvalues should be well separated for numerical stability, but this is not unexpected.

### 3. The case $N = 2$

In the case  $N = 2$ , the equations (2.10) are easy to work out and provide a better understanding of the way things fit together. There are 5 essentially different sub-cases, represented by

$$\partial^2 \lambda_1 / \partial \alpha_{11}^2 = 2(\partial \lambda_2 / \partial \alpha_{11})(\partial \lambda_1 / \partial \alpha_{11}) / (\lambda_1 - \lambda_2), \tag{3.1a}$$

$$\begin{aligned} \partial^2 \lambda_1 / (\partial \alpha_{11} \partial \alpha_{12}) &= [(\partial \lambda_2 / \partial \alpha_{11})(\partial \lambda_1 / \partial \alpha_{12}) \\ &\quad + (\partial \lambda_1 / \partial \alpha_{11})(\partial \lambda_2 / \partial \alpha_{12})] / (\lambda_1 - \lambda_2), \end{aligned} \tag{3.1b}$$

$$\begin{aligned} \partial^2 \lambda_1 / (\partial \alpha_{11} \partial \alpha_{22}) &= [(\partial \lambda_2 / \partial \alpha_{21})(\partial \lambda_1 / \partial \alpha_{12}) \\ &\quad + (\partial \lambda_1 / \partial \alpha_{21})(\partial \lambda_2 / \partial \alpha_{12})] / (\lambda_1 - \lambda_2), \end{aligned} \tag{3.1c}$$

$$\partial^2 \lambda_1 / \partial \alpha_{12}^2 = 2(\partial \lambda_2 / \partial \alpha_{12})(\partial \lambda_1 / \partial \alpha_{12}) / (\lambda_1 - \lambda_2), \tag{3.1d}$$

$$\begin{aligned} \partial^2 \lambda_1 / (\partial \alpha_{12} \partial \alpha_{21}) &= [(\partial \lambda_2 / \partial \alpha_{22})(\partial \lambda_1 / \partial \alpha_{11}) \\ &\quad + (\partial \lambda_1 / \partial \alpha_{22})(\partial \lambda_2 / \partial \alpha_{11})] / (\lambda_1 - \lambda_2) \end{aligned} \tag{3.1e}$$

where

$$\lambda_{1,2} = (\alpha_{11} + \alpha_{22} \pm H^{1/2}) / 2, \quad H = (\alpha_{11} - \alpha_{22})^2 + 4\alpha_{12}\alpha_{21}. \tag{3.2}$$

For the right-hand sides of (3.1 a-e), the results

$$\lambda_1 - \lambda_2 = H^{1/2}, \tag{3.3a}$$

$$\partial \lambda_1 / \partial \alpha_{11} = (1/2)(1 + H^{-1/2}(\alpha_{11} - \alpha_{22})), \tag{3.3b}$$

$$\partial \lambda_2 / \partial \alpha_{11} = (1/2)(1 - H^{-1/2}(\alpha_{11} - \alpha_{22})), \tag{3.3c}$$

$$\partial \lambda_1 / \partial \alpha_{22} = (1/2)(1 + H^{-1/2}(\alpha_{22} - \alpha_{11})), \tag{3.3d}$$

$$\partial \lambda_2 / \partial \alpha_{22} = (1/2)(1 - H^{-1/2}(\alpha_{22} - \alpha_{11})), \tag{3.3e}$$

$$\partial \lambda_1 / \partial \alpha_{12} = H^{-1/2} \alpha_{21}, \tag{3.3f}$$

$$\partial \lambda_2 / \partial \alpha_{12} = -H^{-1/2} \alpha_{21}, \tag{3.3g}$$

$$\partial \lambda_1 / \partial \alpha_{21} = H^{-1/2} \alpha_{12}, \tag{3.3h}$$

$$\partial \alpha_2 / \partial \alpha_{21} = -H^{-1/2} \alpha_{12} \tag{3.3i}$$

are needed.

Thus the left-hand side of (3.1a) is

$$\begin{aligned} \partial^2 \lambda_1 / \partial \alpha_{11}^2 &= (1/2)[H^{-1/2} + (\partial H^{-1/2} / \partial \alpha_{11})(\alpha_{11} - \alpha_{22})] \\ &= (1/2)H^{-3/2}[H - (\alpha_{11} - \alpha_{22})^2] \end{aligned} \tag{3.4a}$$

while the right-hand side of (3.1a) is

$$2(1/4 - 1/4H^{-1}(\alpha_{11} - \alpha_{22})^2)/H^{1/2} = (1/2)H^{-3/2}[H - (\alpha_{11} - \alpha_{22})^2]. \tag{3.4b}$$

Similarly,

$$\partial^2 \lambda_1 / (\partial \alpha_{11} \partial \alpha_{12}) = -(1/4)(\alpha_{11} - \alpha_{22})H^{-3/2}(4\alpha_{21}) \tag{3.5a}$$

and the right-hand side of (3.1b) is

$$\begin{aligned} & [(1/2)(1 - H^{-1/2}(\alpha_{11} - \alpha_{22}))(H^{-1/2}\alpha_{21}) \\ & + (-H^{-1/2}\alpha_{21})(1/2)(1 + H^{-1/2}(\alpha_{11} - \alpha_{22}))]/H^{1/2} \\ & = (1/2)[(H^{-1}\alpha_{21} - H^{-1}\alpha_{21}) - H^{-3/2}2\alpha_{21}(\alpha_{11} - \alpha_{22})]. \end{aligned} \tag{3.5b}$$

Also

$$\begin{aligned} \partial^2 \lambda_1 / (\partial \alpha_{11} \partial \alpha_{22}) & = (\partial / \partial \alpha_{22})(1/2)(1 + H^{-1/2}(\alpha_{11} - \alpha_{22})) \\ & = (1/2)[-H^{-1/2} - (1/2)H^{-3/2}2(\alpha_{11} - \alpha_{22})(\alpha_{22} - \alpha_{11})] \\ & = (1/2)H^{-3/2}[-H + (\alpha_{11} - \alpha_{22})^2] \\ & = (1/2)H^{-3/2}[-4\alpha_{12}\alpha_{21}] \end{aligned} \tag{3.6a}$$

while the right-hand side of (3.1c) is

$$\begin{aligned} & \frac{(\partial \lambda_2 / \partial \alpha_{21})(\partial \lambda_1 / \partial \alpha_{12}) + (\partial \lambda_1 / \alpha_{21})(\partial \lambda_2 / \partial \alpha_{12})}{H^{1/2}} \\ & = [(-H^{-1/2}\alpha_{12})(H^{-1/2}\alpha_{21}) + (H^{-1/2}\alpha_{12})(-H^{-1/2}\alpha_{21})]/H^{1/2} \\ & = -2H^{-3/2}\alpha_{12}\alpha_{21}. \end{aligned} \tag{3.6b}$$

Continuing,

$$\partial^2 \lambda_1 / \partial \alpha_{12}^2 = (\partial / \partial \alpha_{12})(H^{-1/2}\alpha_{21}) = \alpha_{21}(-1/2H^{-3/2})(4\alpha_{21}) \tag{3.7a}$$

and

$$\begin{aligned} & (2\partial \lambda_2 / \partial \alpha_{12})(\partial \lambda_1 / \partial \alpha_{12}) / (\lambda_1 - \lambda_2) \\ & = 2(-H^{-1/2}\alpha_{21})(H^{-1/2}\alpha_{21}) / H^{1/2} = -2H^{-3/2}\alpha_{21}^2. \end{aligned} \tag{3.7b}$$

Finally,

$$\begin{aligned} \partial^2 \lambda_1 / (\partial \alpha_{12} \partial \alpha_{21}) & = (\partial / \partial \alpha_{21})(H^{-1/2}\alpha_{21}) \\ & = H^{-1/2} + \alpha_{21}(-1/2H^{-3/2})4\alpha_{12} \\ & = H^{-3/2}(H - 2\alpha_{21}\alpha_{12}) \\ & = H^{-3/2}((\alpha_{11} - \alpha_{22})^2 + 2\alpha_{21}\alpha_{12}) \end{aligned} \tag{3.8a}$$

and

$$\begin{aligned}
 & [(\partial\lambda_2/\partial\alpha_{22})(\partial\lambda_1/\partial\alpha_{11}) + (\partial\lambda_1/\partial\alpha_{22})(\partial\lambda_2/\partial\alpha_{11})]/H^{1/2} \\
 & = \{1/4(1 + H^{-1/2}(\alpha_{11} - \alpha_{22}))(1 + H^{-1/2}(\alpha_{11} - \alpha_{22})) \\
 & \quad + 1/4(1 - H^{-1/2}(\alpha_{11} - \alpha_{22}))(1 - H^{-1/2}(\alpha_{11} - \alpha_{22}))\}/H^{1/2} \quad (3.8b) \\
 & = 1/2(1 + H^{-1}(\alpha_{11} - \alpha_{22})^2)/H^{1/2} \\
 & = H^{-3/2}[(\alpha_{11} - \alpha_{22})^2 + 2\alpha_{12}\alpha_{21}].
 \end{aligned}$$

The other 15 unordered identities follow by interchanges of the 5 subscripts in  $\partial^2\lambda_k/\partial\alpha_{mn}\partial\alpha_{pq}$ .

#### 4. Parametric dependence

In the applications of these formulae, the elements  $\alpha_j$  of  $A$  may be functions of a small number of parameters  $\theta_1, \theta_2, \dots, \theta_r$ . Then

$$\partial\lambda_j/\partial\theta_s = \sum_{mn} (\partial\lambda_j/\partial\alpha_{mn})(\partial\alpha_{mn}/\partial\theta_s), \quad (4.1)$$

$$\begin{aligned}
 \partial^2\lambda_j/(\partial\theta_s\partial\theta_t) & = (\partial/\partial\theta_t) \sum_{mn} (\partial\lambda_j/\partial\alpha_{mn})(\partial\alpha_{mn}/\partial\theta_s) \\
 & = \sum_{mnpq} \left[ (\partial^2\lambda_j/(\partial\alpha_{mn}\partial\alpha_{pq}))(\partial\alpha_{pq}/\partial\theta_t)(\partial\alpha_{mn}/\partial\theta_s) \right. \\
 & \quad \left. + \sum_{mn} (\partial\lambda_j/\partial\alpha_{mn})\partial^2\alpha_{mn}/(\partial\theta_s\partial\theta_t) \right]. \quad (4.2)
 \end{aligned}$$

Insertion of (2.10) into (4.2) yields

$$\begin{aligned}
 & \sum_{mnpq} \sum_{l \neq j} [(\partial\lambda_l/\partial\alpha_{pn})(\partial\lambda_j/\partial\alpha_{mq}) \\
 & \quad + (\partial\lambda_j/\partial\alpha_{pn})(\partial\lambda_l/\partial\alpha_{mq})](\partial\alpha_{pq}/\partial\theta_t)(\partial\alpha_{mn}/\partial\theta_s)/(\lambda_j - \lambda_l) \\
 & \quad + \sum_{mn} (\partial\lambda_j/\partial\alpha_{mn})\partial^2\alpha_{mn}/(\partial\theta_s\partial\theta_t) \\
 & = \sum_{l \neq j} \sum_{pnq} (\partial\lambda_l/\partial\alpha_{pn})(\partial\alpha_{pq}/\partial\theta_t) \sum_m (\partial\lambda_j/\partial\alpha_{mq})(\partial\alpha_{mn}/\partial\theta_s)/(\lambda_j - \lambda_l) \\
 & \quad + \sum_{l \neq j} \sum_{pnq} (\partial\lambda_j/\partial\alpha_{pn})(\partial\alpha_{pq}/\partial\theta_t) \sum_m (\partial\lambda_l/\partial\alpha_{mq})(\partial\alpha_{mn}/\partial\theta_s)/(\lambda_j - \lambda_l) \\
 & \quad + \sum_{mn} (\partial\lambda_j/\partial\alpha_{mn})\partial^2\alpha_{mn}/(\partial\theta_s\partial\theta_t). \quad (4.3)
 \end{aligned}$$

Let

$$L_{nq}^{st} = \sum_p (\partial \lambda_s / \partial \alpha_{pn}) \partial \alpha_{pq} / \partial \theta_t \quad \text{for } s, t, n, q = 1, 2, \dots, N. \quad (4.4)$$

Then

$$\begin{aligned} \partial^2 \lambda_j / \partial \theta_s \partial \theta_t &= \sum_{l \neq j} \sum_n \sum_q L_{nq}^{lt} L_{nq}^{js} / (\lambda_j - \lambda_l) \\ &+ \sum_{l \neq j} \sum_n \sum_q L_{nq}^{jt} L_{qn}^{ls} / (\lambda_j - \lambda_l) \\ &+ \sum_{mn} (\partial \lambda_j / \partial \alpha_{mn}) \partial^2 \alpha_{mn} / (\partial \theta_s \partial \theta_t). \end{aligned} \quad (4.5)$$

The combination  $\sum_{nq} L_{nq}^{lt} L_{qn}^{js}$  is a matrix trace, namely  $\text{tr}(L^{lt}(L^{js}))$ , with respect to the lower indices. Likewise,

$$\sum_{nq} L_{nq}^{jt} L_{qn}^{ls} = \text{tr}(L^{jt}(L^{ls})) \quad (4.6)$$

so that

$$\begin{aligned} \partial^2 \lambda_j / (\partial \theta_s \partial \theta_t) &= \sum_{l \neq j} [\text{tr}(L^{lt}(L^{js})) + \text{tr}(L^{jt}(L^{ls}))] / (\lambda_j - \lambda_l) \\ &+ \sum_{mn} (\partial \lambda_j / \partial \alpha_{mn}) \partial^2 \alpha_{mn} / (\partial \theta_s \partial \theta_t). \end{aligned} \quad (4.7)$$

Thus the second parametric derivatives of  $\lambda_j$  are expressible in terms of first matrix derivatives, which can be reduced to products of first derivatives of eigenvectors, and eigenprojections of  $A$ .

### 5. Generalised eigenproblems and sensitivity formulas

The generalised eigenproblem discussed below is the solution of

$$(A - \lambda_j B)x_j = 0, \quad y_k^T (A - \lambda_j B) = 0. \quad (5.1)$$

This problem is complicated by the fact that  $B$  may be singular, in which case an eigenvalue can be essentially infinite. Projective methods help to unravel the problem. Corresponding to the pair  $(A, B)$  of  $N \times N$  matrices, introduce pair eigenvalues  $(\alpha, \beta)$  for the problems  $(\alpha A - \beta B)x = 0$ ,  $y^T(\alpha A - \beta B) = 0$ . If  $Bx \neq 0$ , it is natural to choose  $\alpha \neq 0$  and write  $\lambda = \beta/\alpha$  as the eigenvalue. If  $Bx = 0$ , it is natural to set  $\alpha = 0$ . The account below is based on Stewart and Sun [16], which has a very full account of the generalised eigenproblem.

If  $(\alpha, \beta)$  is a pair eigenvalue associated with a right eigenvector  $x$  or a left eigenvector  $y$  for  $(A, B)$ , then  $(c\alpha, c\beta)$  is also a pair eigenvalue with  $c \neq 0$ .



A matrix pair  $(A, B)$  is called regular if  $D = \tau A - \sigma B$  is non-singular for some pair  $(\sigma, \tau)$  of scalars. If so, consider the matrix  $C = \sigma A + \tau B$ . Thus, if  $(\alpha, \beta)$  is an eigenvalue for  $(A, B)$ , then  $(-\alpha\sigma + \beta\tau, \beta\sigma + \alpha\tau)$  is a pair eigenvalue for  $(C, D)$  and conversely, where  $(\alpha, \beta) \neq 0$  and  $(\sigma, \tau) \neq 0$ , all for the same right eigenvector  $\mathbf{x}$ , or left eigenvector  $\mathbf{y}^T$ . Now suppose  $B\mathbf{x} = \mathbf{0}$  and  $\alpha = 0$ . Then  $(\beta\tau, \beta\sigma)$  is a pair eigenvalue for  $(C, D) = (\tau A - \sigma B, \sigma A + \tau B)$  and  $\mu = \sigma/\tau$  is the corresponding eigenvalue.

Another approach of physical interest is to replace  $B$  (if singular), by a non-singular approximation  $\tilde{B}$  and consider either the unsymmetric problem  $\tilde{B}^{-1}A\mathbf{x} = \lambda\mathbf{x}$ , or the symmetrical problem  $\tilde{B}^{-1/2}A\tilde{B}^{-1/2}\mathbf{z} = \lambda\mathbf{z}$ , in effect replacing the pair  $(A, \tilde{B})$  by  $(\tilde{B}^{-1/2}A\tilde{B}^{-1/2}, I)$ . Limits as  $\tilde{B} \rightarrow B$  in some topology can then be taken. If  $B$  is finite-dimensional, the topology is not very significant.

One important result to be carried over from the ordinary problem is the completeness of eigenvectors under proper conditions. The ‘‘pencil’’  $A - \lambda B$  is replaced by  $D = \tau A - \sigma B$ . If  $F_k$  is the family of determinantal minors of order  $k$  of  $D$ , and  $G_k(\tau, \sigma)$  is the greatest common denominator of the  $F_k$ , then  $G_k(\tau, \sigma)$  is a homogeneous polynomial, and the invariant polynomials are  $i_m(\tau, \sigma) = G_{N-m+1}(\tau, \sigma)/G_{N-m}(\tau, \sigma)$ . Then the elementary divisors of  $(A, B)$  over a field  $F$  of scalars are the polynomial factors  $e_{m,1}(\tau, \sigma), \dots, e_{m,q_m}(\tau, \sigma)$  of  $i_m(\tau, \sigma)$  over  $\mathcal{F}$ . If  $e_{m,q_m}(\tau, \sigma) = \tau^{\gamma_m}$  for  $\gamma_m \geq 1$ , then  $i_m$  is said to have an infinite root of order  $\gamma_m$ .

The pair  $(A, B)$  is called simple if each elementary divisor is of degree 1 in  $\tau$  (infinite root) or is of degree 1 in  $\sigma$  (zero root) or is linear in  $\tau$  and  $\sigma$  (finite, non-zero root). In that case, Weierstrass’ reduction theorem for regular pairs shows that the right and left eigenvectors for simple regular pairs are complete, which is what is needed to carry out the methods of Section 2 to reach analogous conclusions. The details of the Weierstrass reduction are found in Stewart and Sun, and, most completely, in Gantmacher [7], vol. I, Chapter 6 and vol. II, Chapter 12. Another condition for complete eigenvectors of  $(A, B)$  is that all the finite, non-zero eigenvalues of the pair are different, and that there is at most one zero eigenvalue and one infinite eigenvalue. From now on, we make this assumption, unless otherwise mentioned.

Returning to the generalised problem, define constants  $c_k$  (analogous to the  $s_k$  of Section 2) by  $\mathbf{y}_k^T B \mathbf{x}_j = c_k \delta_{jk}$ ,

$$c_k = \mathbf{y}_k^T B \mathbf{x}_k = \sum_{rs} y_{kr} b_{rs} x_{ks}. \tag{5.2}$$

These follow from (5.1) by left and right multiplication by  $\mathbf{y}_k$  and  $\mathbf{x}_j$ , yielding  $\mathbf{y}_k A \mathbf{x}_j = \lambda_j \mathbf{y}_k^T B \mathbf{x}_j = \lambda_k \mathbf{y}_k^T B \mathbf{x}_j$ , so that  $(\lambda_j - \lambda_k) \mathbf{y}_k^T B \mathbf{x}_j = 0$ . Simplicity implies  $\mathbf{y}_k B \mathbf{x}_j = 0$  for  $j \neq k$ .

Now, perturb both  $A$  and  $B$  to first order so that:

$$\begin{aligned} A' &= A + \varepsilon\Delta_1 A, & B' &= B + \varepsilon\Delta_1 B, \\ \lambda'_j &= \lambda_j + \varepsilon\Delta_1 \lambda_j + O(\varepsilon^2), & \mathbf{x}'_k &= \mathbf{x}_k + \varepsilon\Delta_1 \mathbf{x}_k + O(\varepsilon^2), \\ \mathbf{y}'_j &= \mathbf{y}_j + \varepsilon\Delta_1 \mathbf{y}_j + O(\varepsilon^2). \end{aligned} \tag{5.3}$$

Hence

$$[A + \varepsilon\Delta_1 A - (\lambda_j + \varepsilon\Delta_1 \lambda_j + O(\varepsilon^2))(B + \varepsilon\Delta_1 B)](\mathbf{x}_j + \varepsilon\Delta_1 \mathbf{x}_j + O(\varepsilon^2)) = 0. \tag{5.4}$$

The first-order corrections are, as justified by previous reasoning,

$$(A - \lambda_j B)\Delta_1 \mathbf{x}_j + (\Delta_1 A - (\Delta_1 \lambda_j)B - \lambda_j \Delta_1 B)\mathbf{x}_j = O(\varepsilon) \tag{5.5}$$

relative to the  $\mathbf{x}_j$ , containing 5 terms in place of the 4 terms in (1.5), for all finite eigenvalues. Left multiplication of (5.5) by  $\mathbf{y}_k^T$  yields

$$\begin{aligned} O(\varepsilon) + \mathbf{y}_k^T(\Delta_1 A)\mathbf{x}_j + \mathbf{y}_k^T A(\Delta_1 \mathbf{x}_j) &= \mathbf{y}_k^T(\Delta_1 A)\mathbf{x}_j + \lambda_k \mathbf{y}_k^T B \Delta_1 \mathbf{x}_j \\ &= \lambda_j \mathbf{y}_k^T B \Delta_1 \mathbf{x}_j + (\Delta_1 \lambda_j) \mathbf{y}_k^T B \mathbf{x}_j + \lambda_j \mathbf{y}_k^T \Delta_1 B \mathbf{x}_j. \end{aligned} \tag{5.6}$$

If  $j = k$ , then (5.6) becomes

$$\begin{aligned} \mathbf{y}_j^T(\Delta_1 A)\mathbf{x}_j + \lambda_j \mathbf{y}_j^T B \Delta_1 \mathbf{x}_j \\ = \lambda_j \mathbf{y}_j^T B \Delta_1 \mathbf{x}_j + (\Delta_1 \lambda_j) c_j + \lambda_j \mathbf{y}_j^T(\Delta_1 B)\mathbf{x}_j + O(\varepsilon). \end{aligned} \tag{5.7}$$

Two terms cancel from (5.7), leaving

$$\mathbf{y}_j^T(\Delta_1 A)\mathbf{x}_j - \lambda_j \mathbf{y}_j^T(\Delta_1 B)\mathbf{x}_j = (\Delta_1 \lambda_j) c_j + O(\varepsilon). \tag{5.8}$$

Expansion of the left side of (5.8) leads, using total differentiability, to

$$\begin{cases} \partial \lambda_j / \partial \alpha_{mn} = y_{jm} x_{jn} / c_j = y_{jm} x_{jn} / \sum_{rs} y_{jr} b_{rs} x_{js}, \\ \partial \lambda_j / \partial b_{mn} = -\lambda_j y_{jm} x_{jn} / c_j = -\lambda_j y_{jm} x_{jn} / \sum_{rs} y_{jr} b_{rs} x_{js}. \end{cases} \tag{5.9}$$

Clearly, (5.9) gives eigenvalue sensitivities, in terms of eigenvalues and eigenvectors. As for right eigenvector sensitivities, take  $j \neq k$  in (5.6), so that

$$\begin{aligned} \mathbf{y}_k^T(\Delta_1 A)\mathbf{x}_j &= (\lambda_j - \lambda_k)(\mathbf{y}_k^T B)(\Delta_1 \mathbf{x}_j) + \lambda_j \mathbf{y}_k^T(\Delta_1 B)\mathbf{x}_j + O(\varepsilon), \\ (\mathbf{y}_k^T)(\Delta_1 \mathbf{x}_j) &= \mathbf{y}_k^T(\Delta_1 A - \lambda_j \Delta_1 B)\mathbf{x}_j / (\lambda_j - \lambda_k) + O(\varepsilon). \end{aligned} \tag{5.10}$$

By completeness, there is a matrix  $D$  such that

$$\Delta_1 \mathbf{x}_j = \sum_l d_{jl} \mathbf{x}_l. \tag{5.11}$$

Left multiplication of (5.11) by  $\mathbf{y}_k^\top B$  yields, using (5.10),

$$\begin{aligned} \mathbf{y}_k^\top B \Delta_1 \mathbf{x}_j &= \sum_l d_{jl} \mathbf{y}_k^\top B \mathbf{x}_l \\ &= \sum_l d_{jl} c_k \delta_{kl} = c_k d_{jk} \\ &= \mathbf{y}_k^\top (\Delta_1 A - \lambda_j \Delta_1 B) \mathbf{x}_j / (\lambda_j - \lambda_k) + O(\varepsilon). \end{aligned} \tag{5.12}$$

Thus

$$d_{jk} = \mathbf{y}_k^\top (\Delta_1 A - \lambda_j \Delta_1 B) \mathbf{x}_j / [c_k (\lambda_j - \lambda_k)] + O(\varepsilon) \tag{5.13}$$

and insertion of (5.13) into (5.11) produces

$$\Delta_1 \mathbf{x}_j = \sum_{k \neq j} [\mathbf{y}_k^\top (\Delta_1 A - \lambda_j \Delta_1 B) \mathbf{x}_j] \mathbf{x}_k / [c_k (\lambda_j - \lambda_k)] + O(\varepsilon). \tag{5.14}$$

Hence, from total differentiability

$$\partial \mathbf{x}_j / \partial \alpha_{mn} = \sum_{k \neq j} y_{km} x_{jn} \mathbf{x}_k / [c_k (\lambda_j - \lambda_k)]$$

and

$$\partial x_{jr} / \partial \alpha_{mn} = \sum_{k \neq j} y_{km} x_{jn} x_{kr} / [c_k (\lambda_j - \lambda_k)]. \tag{5.15}$$

Similarly

$$\begin{aligned} \partial \mathbf{x}_j / \partial b_{mn} &= - \sum_{k \neq j} y_{km} \lambda_j x_{jn} \mathbf{x}_k / [c_k (\lambda_j - \lambda_k)], \\ \partial x_{jr} / \partial b_{mn} &= - \lambda_j \sum_{k \neq j} y_{km} x_{jn} x_{kr} / [c_k (\lambda_j - \lambda_k)] \end{aligned} \tag{5.16}$$

provides the right generalised eigenvector sensitivities with respect to  $B$  in terms of eigenvalues and eigenvectors.

As before, the left eigenvector sensitivities are calculated from

$$(\mathbf{y}_k^\top + \varepsilon \Delta \mathbf{y}_k^\top + O(\varepsilon^2)) [A + \varepsilon \Delta A - (\lambda_k + \varepsilon \Delta_1 \lambda_k + O(\varepsilon^2))(B + \varepsilon \Delta B)] = 0. \tag{5.17}$$

The first-order corrections are

$$\Delta_1 \mathbf{y}_k^\top (A - \lambda_j B) + \mathbf{y}_k^\top (\Delta_1 A - (\Delta_1 \lambda_k) B - \lambda_k \Delta_1 B) = O(\varepsilon). \tag{5.18}$$

Again, 5 terms are present instead of the 4 terms in (1.19). Right multiplication by  $\mathbf{x}_j$  yields

$$\begin{aligned} O(\varepsilon) + \mathbf{y}_k^\top (\Delta_1 A) \mathbf{x}_j + \Delta_1 \mathbf{y}_k^\top (A \mathbf{x}_j) &= \mathbf{y}_k^\top (\Delta_1 A) \mathbf{x}_j + \Delta_1 \mathbf{y}_k^\top \lambda_j B \mathbf{x}_j \\ &= \lambda_k \Delta_1 \mathbf{y}_k^\top (B \mathbf{x}_j) + (\Delta_1 \lambda_k) \mathbf{y}_k^\top B \mathbf{x}_j + \lambda_k \mathbf{y}_k^\top (\Delta_1 B) \mathbf{x}_j. \end{aligned} \tag{5.19}$$

If  $j = k$ , then

$$\begin{aligned} & \mathbf{y}_j^\top(\Delta_1 A)\mathbf{x}_j + \lambda_j \Delta_1 \mathbf{y}_j^\top(B\mathbf{x}_j) \\ &= \lambda_j \Delta_1 \mathbf{y}_j^\top(B\mathbf{x}_j) + \Delta_1 \lambda_j (c_j) + \lambda_j \mathbf{y}_j^\top(\Delta_1 B)\mathbf{x}_j + O(\varepsilon). \end{aligned} \tag{5.20}$$

Two terms cancel, leaving

$$\mathbf{y}_j^\top(\Delta_1 A)\mathbf{x}_j - \lambda_j \mathbf{y}_j^\top(\Delta_1 B)\mathbf{x}_j = (\Delta_1 \lambda_j) c_j + O(\varepsilon)$$

which coincides with (5.8). However, if  $j \neq k$ , then (5.19) yields

$$\begin{aligned} \mathbf{y}_k^\top(\Delta_1 A)\mathbf{x}_j &= (\lambda_k - \lambda_j) \Delta_1 \mathbf{y}_k^\top(B\mathbf{x}_j) + \lambda_k \mathbf{y}_k^\top(\Delta_1 B)\mathbf{x}_j + O(\varepsilon), \\ (\Delta_1 \mathbf{y}_k^\top)(B\mathbf{x}_j) &= \mathbf{y}_k^\top(\Delta_1 A - \lambda_k \Delta_1 B)\mathbf{x}_j / [\lambda_k - \lambda_j] + O(\varepsilon). \end{aligned} \tag{5.21}$$

By completeness of the  $\mathbf{y}_k$ , there is a matrix  $E$  such that

$$\Delta_1 \mathbf{y}_k^\top = \sum_l e_{kl} \mathbf{y}_l^\top \tag{5.22}$$

Hence insertion of (5.22) into (5.21) yields

$$\begin{aligned} \left( \sum_l e_{kl} \mathbf{y}_l^\top \right) (B\mathbf{x}_j) &= \sum_l e_{kl} \delta_{jl} c_l \\ &= e_{kj} c_j = \mathbf{y}_k^\top(\Delta_1 A - \lambda_k \Delta_1 B)\mathbf{x}_j / [\lambda_k - \lambda_j] + O(\varepsilon) \end{aligned} \tag{5.23}$$

and thus

$$e_{kj} = \mathbf{y}_k^\top(\Delta_1 A - \lambda_k \Delta_1 B)\mathbf{x}_j / [c_j(\lambda_k - \lambda_j)] + O(\varepsilon). \tag{5.24}$$

From (5.22)

$$\Delta_1 \mathbf{y}_k^\top = \sum_{l \neq k} \mathbf{y}_k^\top(\Delta_1 A - \lambda_k \Delta_1 B)\mathbf{x}_l \mathbf{y}_l^\top / [c_l(\lambda_k - \lambda_l)] + O(\varepsilon). \tag{5.25}$$

Hence

$$\begin{aligned} \partial \mathbf{y}_k^\top / \partial \alpha_{mn} &= \sum_{l \neq k} y_{km} x_{ln} \mathbf{y}_l^\top / [c_l(\lambda_k - \lambda_l)], \\ \partial y_{kr} / \partial \alpha_{mn} &= \sum_{l \neq k} y_{km} x_{ln} y_{lr} / [c_l(\lambda_k - \lambda_l)], \end{aligned} \tag{5.26}$$

and similarly

$$\begin{aligned} \partial \mathbf{y}_k^\top / \partial b_{mn} &= \sum_{l \neq k} (-\lambda_k) y_{km} x_{ln} \mathbf{y}_l^\top / [c_l(\lambda_k - \lambda_l)], \\ \partial y_{kr} / \partial b_{mn} &= -\lambda_k \sum_{l \neq k} y_{km} x_{ln} y_{lr} / [c_l(\lambda_k - \lambda_l)]. \end{aligned} \tag{5.27}$$

The six sensitivity formulas (5.9), (5.15), (5.16), (5.26), (5.27) are obviously generalisations of the Wilkinson formulas, reducing to them when  $B = I$  and  $c_l = s_l$ . This problem is also treated by Stewart [13].

## 6. Partial differential equations for generalised eigenvalues

For generalised eigenvalues, the replacement of  $s_l = \sum_t y_{lt} x_{lt}$  by  $c_l = \sum_{rs} y_{lr} b_{rs} x_{ls}$  in the sensitivity formulas is the major complication in redifferentiation. Note that

$$\partial c_l / \partial \alpha_{mn} = \sum_{rs} [(\partial y_{lr} / \partial \alpha_{mn}) b_{rs} x_{ls} + y_{lr} (\partial b_{rs} / \partial \alpha_{mn}) x_{ls} + y_{lr} b_{rs} (\partial x_{ls} / \partial \alpha_{mn})] \quad (6.1)$$

and

$$\begin{aligned} \partial c_l / \partial b_{mn} &= \sum_{rs} [(\partial y_{lr} / \partial b_{mn}) b_{rs} x_{ls} + y_{lr} \delta_{rm} \delta_{sn} x_{ls} + y_{lr} b_{rs} (\partial x_{ls} / \partial b_{mn})] \\ &= y_{lm} x_{ln} + \sum_{rs} b_{rs} [(\partial y_{lr} / \partial b_{mn}) x_{ls} + y_{lr} (\partial x_{ls} / \partial b_{mn})]. \end{aligned} \quad (6.2)$$

If the entries  $b_{rs}$  in  $B$  are independent of the entries  $a_{mn}$  in  $A$ , then

$$\partial c_l / \partial \alpha_{mn} = \sum_{rs} b_{rs} [(\partial y_{lr} / \partial \alpha_{mn}) x_{ls} + y_{lr} (\partial x_{ls} / \partial \alpha_{mn})]. \quad (6.3)$$

However, in a variety of applications,  $B$  and  $A$  have entries which depend on the same parameters and so the expression  $\partial b_{rs} / \partial \alpha_{mn}$  must be retained to calculate (6.1). Also, (5.9), (5.5), (5.16), (5.26), (5.27) must be modified in that case.

Three types of second derivatives require notice. Besides

$$R_{jmn pq} = \partial^2 \lambda_j / (\partial \alpha_{mn} \partial \alpha_{pq}) \quad (6.4)$$

there are

$$S_{jmn pq} = \partial^2 \lambda_j / (\partial \alpha_{mn} \partial b_{pq}) \quad (6.5)$$

and

$$T_{jmn pq} = \partial^2 \lambda_j / (\partial b_{mn} \partial b_{pq}) \quad (6.6)$$

to consider. Now

$$\begin{aligned}
 R_{jnmpq} &= (\partial/\partial\alpha_{pq})(\partial\lambda_j/\partial\alpha_{mn}) \\
 &= (\partial/\partial\alpha_{pq}) \left[ (y_{jm}x_{jn}) / \left( \sum_{rs} y_{jr}b_{rs}x_{js} \right) \right] \\
 &= (\partial y_{jm}/\partial\alpha_{pq})(x_{jn}/c_j) + (\partial x_{jn}/\partial\alpha_{pq})(y_{jm}/c_j) \\
 &\quad - (y_{jm}x_{jn}/c_j^2) \left( \sum_{rs} \partial y_{jr}/\partial\alpha_{pq} \right) b_{rs}x_{js} \\
 &\quad - (y_{jm}x_{jn}/c_j^2) \left( \sum_{rs} y_{jr}(\partial b_{rs}/\partial\alpha_{pq}) \right) x_{js} \\
 &\quad - (y_{jm}x_{jn}/c_j^2) \left( \sum_{rs} y_{jr}b_{rs}(\partial x_{js}/\partial\alpha_{pq}) \right)
 \end{aligned} \tag{6.7}$$

which by (5.26) and (5.16) and (6.1) becomes

$$\begin{aligned}
 &\sum_{l \neq j} (y_{jp}x_{lq}y_{lm}/[c_l(\lambda_j - \lambda_l)])(x_{jn}/c_j) \\
 &\quad + \sum_{l \neq j} y_{lp}x_{jq}x_{ln}/[c_l(\lambda_j - \lambda_l)](y_{jm}/c_j) \\
 &\quad - (y_{jm}x_{jn}/c_j^2) \left[ \sum_{rs} \sum_{l \neq j} y_{jp}x_{lq}y_{lr}b_{rs}x_{js}/[c_l(\lambda_j - \lambda_l)] \right. \\
 &\quad\quad\quad + \sum_{rs} y_{jr}(\partial b_{rs}/\partial\alpha_{pq})x_{js} \\
 &\quad\quad\quad \left. + \sum_{rs} y_{jr}b_{rs} \cdot \sum_{l \neq j} y_{lp}x_{jq}x_{ls}/[c_l(\lambda_j - \lambda_l)] \right].
 \end{aligned} \tag{6.8}$$

Now (6.8) can be expanded to

$$\begin{aligned}
 &\sum_{l \neq j} (y_{jp}x_{lq}y_{lm}x_{jn})/[c_jc_l(\lambda_j - \lambda_l)] \\
 &\quad + \sum_{l \neq j} \sum_{rs} y_{lp}x_{jq}x_{ln}y_{jm}/[c_jc_l(\lambda_j - \lambda_l)] \\
 &\quad - \sum_{l \neq j} \sum_{rs} (y_{jm}x_{jn}y_{jp}x_{lq}y_{lr}x_{js}b_{rs}/c_j^2c_l(\lambda_j - \lambda_l)) \\
 &\quad - \sum_{l \neq j} \sum_{rs} (y_{jm}x_{jn}y_{jr}b_{rs}y_{lp}x_{jq}x_{ls}/c_j^2c_l(\lambda_j - \lambda_l)) \\
 &\quad - y_{jm}x_{jn} \sum_{rs} y_{jr}x_{js}(\partial b_{rs}/\partial\alpha_{pq})/c_j^2.
 \end{aligned} \tag{6.9}$$

As in the derivation of (2.10) from (1.10), the first four terms of (6.9) resemble (2.10) as seen from (5.9):

$$\begin{aligned}
 R_{jmn pq} = & \sum_{l \neq j} (\partial \lambda_j / \partial \alpha_{pn}) (\partial \lambda_l / \partial \alpha_{mq}) / (\lambda_j - \lambda_l) \\
 & + \sum_{l \neq j} (\partial \lambda_l / \partial \alpha_{pn}) (\partial \lambda_j / \partial \alpha_{mq}) / (\lambda_j - \lambda_l) \\
 & - \sum_{l \neq j} y_{jm} x_{jn} y_{jp} x_{lq} c_l \delta_{jl} / [c_j^2 c_l (\lambda_j - \lambda_l)] \quad (6.10) \\
 & - \sum_{l \neq j} y_{jm} x_{jn} y_{lp} x_{jq} c_j \delta_{jl} / [c_j^2 c_l (\lambda_j - \lambda_l)] \\
 & - y_{jm} x_{jn} \sum_{rs} y_{jr} x_{js} (\partial b_{rs} / \partial \alpha_{pq}) / c_j^2.
 \end{aligned}$$

From (5.9), the above sum is

$$\begin{aligned}
 R_{jmn pq} = & \sum_{l \neq j} [(\partial \lambda_j / \partial \alpha_{pn}) (\partial \lambda_l / \partial \alpha_{mq}) + (\partial \lambda_l / \partial \alpha_{pn}) (\partial \lambda_j / \partial \alpha_{mq})] / (\lambda_j - \lambda_l) \\
 & - \sum_{rs} (\partial \lambda_j / \partial \alpha_{mn}) (\partial \lambda_j / \partial \alpha_{rs}) (\partial b_{rs} / \partial \alpha_{pq}), \quad (6.11)
 \end{aligned}$$

since two terms drop out because of the appearance of  $\delta_{jl}$ .

Thus, if  $\partial b_{rs} / \partial \alpha_{pq} = 0$  for all  $r, s, p, q$ , then the differential equation (6.11) has the same form as (2.10). If not, then (6.11) possibly has the additional  $N^2$  terms along with the previous  $2(N - 1)$  terms, but retains the bilinear structure in the first derivatives even after taking account of dependences between  $A$  and  $B$ . Next,

$$\begin{aligned}
 S_{jmn pq} = & (\partial / \partial b_{pq}) (\partial \lambda_j / \partial \alpha_{mn}) \\
 = & (\partial / \partial b_{pq}) (y_{jm} x_{jn} / c_j) \\
 = & (\partial y_{jm} / \partial b_{pq}) (x_{jn} / c_j) + (\partial x_{jn} / \partial b_{pq}) (y_{jm} / c_j) \quad (6.12) \\
 & - (1 / c_j^2) y_{jm} x_{jn} (\partial c_j / \partial b_{pq})
 \end{aligned}$$

Expanding with the help of (6.1), (5.16), and (5.27), and exchange of summations, and finally both parts of (5.9),

$$\begin{aligned}
 S_{jmn pq} = & -\lambda_j \sum_{l \neq j} (y_{jp} x_{lq} y_{lm} / [c_l (\lambda_j - \lambda_l)]) (x_{jn} / c_j) \\
 & - \lambda_j \sum_{l \neq j} (y_{lp} x_{jq} x_{ln} / [c_l (\lambda_j - \lambda_l)]) (y_{jm} / c_j)
 \end{aligned}$$

$$\begin{aligned}
 & - (1/c_j^2)y_{jm}x_{jn} \left[ y_{jp}x_{jq} + \sum_{rs} b_{rs}(\partial y_{jr}/\partial b_{pq})x_{js} \right. \\
 & \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \left. + \sum_{rs} b_{rs}y_{jr}(\partial x_{js}/\partial b_{pq}) \right] \\
 = & \sum_{l \neq j} [(\partial \lambda_j/\partial b_{pn})(\partial \lambda_l/\partial \alpha_{mq}) + (\partial \lambda_l/\partial b_{pn})(\partial \lambda_j/\partial \alpha_{mq})]/(\lambda_j - \lambda_l) \\
 & + \left\{ \begin{aligned} & -(1/c_j^2)y_{jm}x_{jn}y_{jp}x_{jq} \\ & -(1/c_j^2)y_{jm}x_{jn} \sum_{rs} b_{rs}x_{js}(-\lambda_j) \sum_{l \neq j} y_{jp}x_{lq}y_{lr}/[c_l(\lambda_j - \lambda_l)] \\ & -(1/c_j^2)y_{jm}x_{jn} \sum_{rs} b_{rs}y_{jr} \sum_{l \neq j} (-\lambda_l)y_{lp}x_{jq}x_{ls}/[c_l(\lambda_j - \lambda_l)] \end{aligned} \right. \\
 = & \sum_{l \neq j} [(\partial \lambda_j/\partial b_{pn})(\partial \lambda_l/\partial \alpha_{mq}) + (\partial \lambda_l/\partial b_{pn})(\partial \lambda_j/\partial \alpha_{mq})]/(\lambda_j - \lambda_l) \\
 & + (\partial \lambda_j/\partial \alpha_{mn})(\partial \lambda_j/\partial b_{pq})/\lambda_j \\
 & + (\lambda_j/c_j^2) \sum_{l \neq j} y_{jm}x_{jn}c_l(\delta_{lj}y_{jp}x_{lq} + \delta_{lj}y_{lp}x_{jq})/[c_l(\lambda_j - \lambda_l)] \\
 = & \sum_{l \neq j} [(\partial \lambda_j/\partial b_{pn})(\partial \lambda_l/\partial \alpha_{mq}) + (\partial \lambda_l/\partial b_{pn})(\partial \lambda_j/\partial \alpha_{mq})]/(\lambda_j - \lambda_l) \\
 & + (\partial \lambda_j/\partial \alpha_{mn})(\partial \lambda_j/\partial b_{pq})/\lambda_j. \tag{6.13}
 \end{aligned}$$

Thus the partial differential equations for the mixed second derivatives have  $2N - 1$  terms, if all eigenvalues are non-zero.

Finally,

$$\begin{aligned}
 T_{jmn pq} - & = (\partial/\partial b_{pq})(\partial \lambda_j/\partial b_{mn}) = (\partial/\partial b_{pq})(-\lambda_j y_{jm} x_{jn}/c_j) \\
 = & - \left\{ (\partial \lambda_j/\partial b_{pq})y_{jm}x_{jn}/c_j \right. \\
 & + \lambda_j(\partial y_{jm}/\partial b_{pq})x_{jn}/c_j + \lambda_j y_{jm}(\partial x_{jn}/\partial b_{pq})/c_j \\
 & - (1/c_j^2)\lambda_j y_{jm}x_{jn} \left[ y_{jp}x_{jq} + \sum_{rs} b_{rs}(\partial y_{jr}/\partial b_{pq})x_{js} \right. \\
 & \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \left. + \sum_{rs} b_{rs}y_{jr}(\partial x_{js}/\partial b_{pq}) \right] \left. \right\}. \tag{6.14}
 \end{aligned}$$



From (6.14), (5.9), (5.16), (5.27)

$$\begin{aligned}
 T_{jmnpq} = - \left\{ \right. & -\lambda_j y_{jp} x_{jq} y_{jm} x_{jn} / c_j^2 \\
 & + \lambda_j (-\lambda_j) \sum_{l \neq j} y_{jp} x_{lq} y_{lm} x_{jn} [c_j c_l (\lambda_j - \lambda_l)] \\
 & + \lambda_j y_{jm} (-\lambda_j) \sum_{l \neq j} y_{lp} x_{jq} x_{ln} / [c_j c_l (\lambda_j - \lambda_l)] - \lambda_j y_{jm} x_{jn} y_{jp} x_{jq} / c_j^2 \\
 & - \sum_{rs} (\lambda_j y_{jm} x_{jn} / c_j^2) (b_{rs} x_{js}) (-\lambda_j) \sum_{l \neq j} y_{jp} x_{lq} y_{lr} / [c_l (\lambda_j - \lambda_l)] \\
 & \left. - \sum_{rs} b_{rs} y_{jr} \lambda_j y_{jm} x_{jn} (-\lambda_j) \sum_{l \neq j} y_{lp} x_{jq} x_{ls} / [c_l (\lambda_j - \lambda_l)] \right\}.
 \end{aligned}
 \tag{6.15}$$

Reapplying (5.9) six times to (6.13) and dropping the last two terms by orthogonality,

$$\begin{aligned}
 T_{jmnpq} = & 2(\partial \lambda_j / \partial b_{mn})(\partial \lambda_j / \partial b_{pq}) / \lambda_j \\
 & + \sum_{l \neq j} [(\partial \lambda_j / \partial b_{pn})(\partial \lambda_l / \partial b_{mq}) + (\partial \lambda_j / \partial b_{mq})(\partial \lambda_l / \partial b_{pn})] / (\lambda_j - \lambda_l).
 \end{aligned}
 \tag{6.16}$$

### 7. Examples and applications

The interesting applications are primarily to  $3 \times 3$ ,  $4 \times 4$ ,  $6 \times 6$  and infinite matrices. For illustrations of the theory, perhaps it will be sufficient to use  $3 \times 3$  matrices in showing how to get the second-order derivatives. One such example, or exercise, is deliberately chosen to require relatively few calculations. The other is quite a realistic problem in cold plasma theory, which has not been carried out to this extent previously, and where the new techniques seem to be quite useful.

#### I. Partially-broken symmetry.

$$\text{Let } A = \begin{bmatrix} a & b & c \\ c & a & -b \\ -b & c & a \end{bmatrix}
 \tag{7.1}$$

so that

$$\det(A - \lambda I) = (a - \lambda)^3 + b^3 + c^3 + bc(a - \lambda).
 \tag{7.2}$$

Taking  $c = b + b'$  to be a perturbation of  $b$ , we see that the unperturbed equation is  $(a - \lambda)^3 + 2b^3 + b^2(a - \lambda) = (a + b - \lambda)((a - \lambda)^2 - b(a - \lambda) + 2b^2) = 0$  whose roots are

$$\lambda_1 = a - bz_1, \quad \lambda_2 = a + b, \quad \lambda_3 = a - bz_3 \tag{7.3}$$

where  $z_{1,3} = (1 \pm \sqrt{-7})/2$ .

The unperturbed right eigenvectors are

$$\begin{bmatrix} 1 \\ -1 \\ 1 - z_1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -1 \\ 1 - z_3 \end{bmatrix} \tag{7.4}$$

and the unperturbed left eigenvectors are

$$[1, -1, z_1 - 1], \quad [1, 1, 0], \quad [1, -1, z_3 - 1] \tag{7.5}$$

The projections are

$$P_1 = \frac{\begin{bmatrix} 1 & -1 & z_1 - 1 \\ -1 & 1 & 1 - z_1 \\ 1 - z_1 & z_1 - 1 & 1 + z_1 \end{bmatrix}}{(z_1 + 3)}, \quad P_2 = \frac{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}{2}, \tag{7.6}$$

$$P_3 = \frac{\begin{bmatrix} 1 & -1 & z_3 - 1 \\ -1 & 1 & 1 - z_3 \\ 1 - z_3 & z_3 - 1 & 1 + z_3 \end{bmatrix}}{z_3 + 3}$$

The second-order approximation to the  $\lambda$ 's requires the first derivatives, which are the transposes of  $P_1, P_2, P_3$ , and the second derivatives, contained in (2.10). Since only  $a_{13}, a_{21}$ , and  $a_{32}$  are perturbed, only six second derivatives need calculation.

Writing out (2.10) in 3 typical cases,

$$R_{1,13,12} = 2(P_1)_{31} \left[ \frac{(P_2)_{31}}{\lambda_1 - \lambda_2} + \frac{(P_3)_{31}}{\lambda_1 - \lambda_3} \right],$$

$$R_{1,13,21} = [(P_2)_{32}(P_1)_{11} + (P_1)_{32}(P_2)_{11}]/(\lambda_1 - \lambda_2) + [(P_3)_{32}(P_1)_{11} + (P_1)_{32}(P_3)_{11}]/(\lambda_1 - \lambda_3),$$

$$R_{1,13,32} = [(P_2)_{33}(P_1)_{21} + (P_1)_{33}(P_2)_{21}]/(\lambda_1 - \lambda_2) + [(P_3)_{33}(P_1)_{21} + (P_1)_{33}(P_3)_{21}]/(\lambda_1 - \lambda_3). \tag{7.7}$$

Note how many different entries from the projections are needed to get just these three second-order terms: 5 of the  $P_1$ , 5 of the  $P_2$  and 5 of the  $P_3$ . The actual results are not very interesting except in the peculiar mixture of symmetry and asymmetry, and in that any six second derivatives of interest involve all of the entries in all three projections.

**II. Waves in cold magnetised plasmas, and other examples.** A non-trivial example of some independent interest, which avoids extensive computing, is provided by the accepted theory of waves in a “cold” magnetised plasma. The account, based on Stix [17], is centered on the “plane wave” matrix

$$C = \begin{bmatrix} S - n^2 \cos^2 \theta & -iD & n^2 \cos \theta \sin \theta \\ iD & S - n^2 & 0 \\ n^2 \cos \theta \sin \theta & 0 & P - n^2 \sin^2 \theta \end{bmatrix} \tag{7.8}$$

which is Hermitian and just a touch complex. This matrix follows from Maxwell’s equations in a cold plasma, Ohm’s law for current density, and a special coordinate system for the wave and the magnetic field. The parameters  $S$ ,  $D$ ,  $P$  depend on the assumed frequency  $\omega$  and the characteristics of the various components of the plasma and on the magnitude  $B$  of the magnetic field. The cold-plasma derivation is based on warm-plasma theory in the low-temperature limit, assuming a Gaussian zero-mean distribution for the transverse particle velocities, and a drifted mean distribution for the parallel (to the magnetic field) velocities [17, Chapters 8, 9].

If  $\mathbf{k}$  is the wave-order of the electric field  $\mathbf{E} = \hat{\mathbf{E}} \exp(i\omega t - i\mathbf{k} \cdot \mathbf{x})$ , and  $\mathbf{n}$  has the direction  $\mathbf{k}$  and the magnitude  $n = kc/\omega$ , then  $\mathbf{n}$  is dimensionless. If  $\mathbf{n}$  is taken in the  $xz$  plane, the magnetic field  $\mathbf{B}$  is along the  $z$ -axis, and  $\theta$  is the angle between  $\mathbf{B}$  and  $n$ , then  $C\hat{\mathbf{E}} = 0$  is the free wave equation. This is a generalised eigenvector problem for a pair  $(A, M)$  of matrices, of the form  $(A - n^2M)\hat{\mathbf{E}} = 0$ , where

$$A = \begin{bmatrix} S & -iD & 0 \\ iD & S & 0 \\ 0 & 0 & P \end{bmatrix}, \quad M = \begin{bmatrix} \cos^2 \theta & 0 & \cos \theta \sin \theta \\ 0 & 1 & 0 \\ \cos \theta \sin \theta & 0 & \sin^2 \theta \end{bmatrix}. \tag{7.9}$$

Obviously  $\det M = 0$ ,  $\det A = (S^2 - D^2)P$  so we will assume  $P \neq 0$ ,  $\omega^2$ . The more basic quantities are  $R = S + D$ ,  $L = S - D$ , resulting in  $RL = S^2 - D^2$ . Clearly

$$\det C = an^4 - bn^2 - c, \tag{7.10}$$

where

$$\begin{aligned} a &= S \sin^2 \theta + P \cos^2 \theta, \\ b &= RL \sin^2 \theta + PS(1 + \cos^2 \theta), \\ c &= \det A = PRL. \end{aligned} \tag{7.11}$$

Since  $n_{1,3}^2 = (b \pm \sqrt{b^2 - 4ac})/2a$  are the two roots, the root  $n_1^2 = (b + \sqrt{b^2 - 4ac})/2a \rightarrow \infty$  as  $a \rightarrow 0$ , while  $n_3^2 \rightarrow c/b$  as  $a \rightarrow 0$ .

Since (7.10) is quadratic in  $n^2$ , one of the eigenvalues is infinite even for  $a \neq 0$  because  $\det M = 0$ . When  $a \rightarrow 0$ ,  $\theta \rightarrow \theta_0$  where  $\tan^2 \theta_0 = -P/S$ .

This  $\theta_0$  is called the resonance angle. Another important angle is  $\theta = \theta_1$ , where  $b = RL + PS + (PS - RL) \cos^2 \theta_1 = 0$ . When  $S \rightarrow 0$ , the resonance angle  $\theta_0 \rightarrow \pi/2$ .

Two unperturbed eigenvectors  $\widehat{E}_1, \widehat{E}_3$  lie in the subspace whose 2nd component is zero. They are related by the equations

$$(\widehat{E}_3)_j = [\lambda_j \cos \theta \sin \theta / (P - \lambda_j \sin^2 \theta)] (\widehat{E}_1)_j \tag{7.12}$$

for  $j = 1, 3$ . If  $b^2 - 4ac > 0$ ,  $\widehat{E}_{1,3}$  can be real. These facts are stated in Stix [17], but the behavior of the remaining eigenvector is not.

The third ‘‘eigenvector’’  $\widehat{E}'$  satisfies the inequation  $\widehat{E}'_2 \neq 0$  and the conditions  $C\widehat{E}' \neq 0$  and  $C^2\widehat{E}' = 0$ , and exists if

$$a'\lambda_j^2 - b'\lambda_j + c' \neq 0 \quad \text{for } j = 1, 3, \tag{7.13}$$

where

$$\begin{aligned} a' &= P^2 \cos^2 \theta + (S^2 + D^2) \sin^2 \theta - (S - P)^2 \cos^2 \theta \sin^2 \theta, \\ b' &= 2P[SP \cos^2 \theta + (S^2 + D^2) \sin^2 \theta], \\ c' &= P^2(S + P)^2 \end{aligned} \tag{7.14}$$

are related to  $C^2$ . The parameters  $S, P, D, R, L$  are not optimal, as (7.11) and (7.14) and Kato’s theory shows, but they are conventional.

A long calculation shows that both conditions of (7.13) are satisfied except on a singular manifold in  $(S, P, D)$  space. If  $\widehat{E}'_2$  is real, then  $\widehat{E}'_1$  and  $\widehat{E}'_3$  are purely imaginary.

The full set of eigenvalues can be restored by regularising  $M$  into

$$\widetilde{M} = M_\alpha = \begin{bmatrix} \cos^2 \theta & 0 & \alpha \cos \theta \sin \theta \\ 0 & 1 & 0 \\ \alpha \cos \theta \sin \theta & 0 & \sin^2 \theta \end{bmatrix} \tag{7.15}$$

and writing  $C_\alpha = A - \lambda M_\alpha$ , for  $|\alpha|^2 < 1$ .

Then  $\det C_\alpha = (\det M_\alpha) \det(N - \lambda I)$ , where

$$N = M_\alpha^{-1/2} A M_\alpha^{-1/2}. \tag{7.16}$$

The eigenvalues  $\mu_1, \mu_2, \mu_3$  and eigenprojections  $R_1, R_2, R_3$  of  $M_\alpha$  are needed. Clearly

$$\mu_2 = 1 \quad \text{and} \quad R_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The other eigenvalues satisfy  $\mu^2 - \mu + (1 - \alpha^2) \sin^2 \theta \cdot \cos^2 \theta = 0$ , and

$$\mu_{1,3} = (1 \pm \sqrt{1 - 4(1 - \alpha^2) \sin^2 \theta \cos^2 \theta})/2 \tag{7.17}$$

which are real if  $(1 - \alpha^2) \sin^2 \theta \cos^2 \theta < 1/4$ . The un-normalised eigenvectors are

$$\mathbf{r}_j^\top = [1, 0, \rho_j], \quad \rho_j = (\mu_j - \cos^2 \theta) / (\alpha \sin \theta \cos \theta) \tag{7.18}$$

for  $j = 1, 3$ . Thus

$$R_j = \begin{bmatrix} 1 & 0 & \rho_j \\ 0 & 0 & 0 \\ \rho_j & 0 & \rho_j^2 \end{bmatrix} / (1 + \rho_j^2) \quad \text{for } \alpha = 1, 3. \tag{7.19}$$

Obviously  $M_\alpha^{-1/2} = \sum_j \mu_j^{-1/2} R_j$  and

$$\begin{aligned} N &= \sum_{jk} \mu_j^{-1/2} R_j A R_k \mu_k^{-1/2} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & 0 \end{bmatrix} + \sum_{j,k \neq 2} \frac{S + \rho_j \rho_k P}{(1 + \rho_j^2)(1 + \rho_k^2)} \begin{bmatrix} 1 & 0 & \rho_k \\ 0 & 0 & 0 \\ \rho_j & 0 & \rho_j \rho_k \end{bmatrix} \mu_j^{-1/2} \mu_k^{-1/2} \\ &\quad + \sum_{j \neq 2} \frac{\mu_j^{-1/2}}{1 + \rho_j^2} \begin{bmatrix} 0 & -iD & 0 \\ +iD & 0 & +iD \rho_j \\ 0 & -iD \rho_j & 0 \end{bmatrix}. \end{aligned} \tag{7.20}$$

Element-wise

$$\begin{aligned} N_{11} &= m_1^2 S + m_2^2 P, & N_{12} &= -iD m_1, & N_{21} &= \bar{N}_{12}, \\ N_{22} &= S, & N_{13} &= N_{31} = m_1 m_2 S + m_2 m_3 P, \\ N_{32} &= -iD m_2, & N_{23} &= \bar{N}_{32}, & N_{33} &= m_2^2 S + m_3^2 P, \end{aligned} \tag{7.21}$$

where

$$m_{i+1} = \frac{\mu_1^{-1/2} \rho_1^i}{1 + \rho_1^2} + \frac{\mu_3^{-1/2} \rho_3^i}{1 + \rho_3^2}, \quad i = 0, 1, 2. \tag{7.22}$$

The eigenvalues and eigenvectors of  $N$  are obtainable with further effort even when  $D \neq 0$ . To apply (2.10), we need the first derivatives of these eigenvalues with respect to  $\theta$ , or alternatively, the projections and a few other facts. All this becomes quite tedious, unless  $D = 0$  in which case

$$N = \begin{bmatrix} m_1^2 S + m_2^2 P & 0 & m_2(m_1 S + m_3 P) \\ 0 & S & 0 \\ m_2(m_1 S + m_3 P) & 0 & m_2^2 S + m_3^2 P \end{bmatrix} \tag{7.23}$$

where  $\rho_2 = 0$ . The eigenvalues of  $N$  are  $\nu_1, \nu_2, \nu_3$ , where  $\nu_2 = S$ , and  $\nu_1, \nu_3$  satisfy

$$\nu^2 - \nu[(m_1^2 S + m_2^2 P) + (m_2^2 S + m_3^2 P)] + SP[m_1 m_3 - m_2^2]^2 = 0. \tag{7.24}$$

Now from (7.17) and (7.18),  $\rho_1, \rho_3$  depend on  $\theta$ , and from (7.22),  $m_1, m_2, m_3$  also depend on  $\theta$ . From here on, the algebraic details become rather

unpleasant and the conclusions are of primary interest to plasma physicists. However, the possibility of obtaining the perturbations around the resonance angle, at which many of the expressions will simplify, may be sufficiently clear. Without the formulas previously obtained, there would be little hope of working the results out, and indeed this has not been done to the writer's knowledge.

Similar calculations appear in the dynamics of streaming filaments, in which the unperturbed state is a moving helix and instability is manifested in various kink and balloon modes. Again, the longitudinal eigenvalues and modes are fairly simple and the transverse modes are quadratics or iterated quadratics, algebraically. A book on this subject is in preparation by Dr. M. Zak of the Jet Propulsion Laboratory, and the present writer. However applied mathematics is full of unworked higher-order perturbations of low-dimensional dynamics, in which parametric and not merely numerical solutions are desirable. The interest in the stability theory of fusion reactions, in which both  $3 \times 3$  and  $6 \times 6$  matrices are known to enter centrally, was another motivating example. However, the immediate impetus was problems in platform stability of thin structures, to which the writer intends to return.

### 8. A formula of Rogers

In the compendium of Rogers [12], many of the formulas derived here are absent. However, one of them is equivalent to (2.10), due to Wilkinson, and it could have been used as the basis for our work. It is less convenient, as seen from its form:

$$\partial \lambda_i / \partial A = \prod_{j \neq i} [(A^T - \lambda_j I) / (\lambda_i - \lambda_j)]. \quad (8.1)$$

Given this form, however, it is easy to see how to derive (8.1) from the spectral theorem (1.4 et seq.) for simple matrices and (2.10).

Let

$$G_i(A) = \prod_{j \neq i} (A - \lambda_j I), \quad (8.2)$$

then for  $r \neq i$ ,

$$G_i(A) = (A - \lambda_r I) \prod_{j \neq i, r} (A - \lambda_j I).$$

If  $A$  is simple,  $A = \sum_k \lambda_k P_k$ . Since  $G_i(A)$  commutes with  $A$ ,  $G_i(A) = \sum_j g_{ij} P_j$  and thus  $P_r G_i(A) = P_r (A - \lambda_r I) \prod_{j \neq i, r} (A - \lambda_j I) = 0$  for  $r \neq i$  and  $P_r G_i(A) = \sum_i g_{ij} P_r P_j = g_{ir} P_r$ . Hence  $g_{ir} = 0$  for  $r \neq i$  and  $G_i(A) = g_{ii} P_i =$

$\prod_{j \neq i} (\sum_k \lambda_k P_k - \lambda_j \sum_k P_k) = \prod_{j \neq i} (\sum_k (\lambda_k - \lambda_j) P_k) = \prod_{j \neq i} (\lambda_i - \lambda_j) P_i + \text{terms orthogonal to } P_i$ .

From the above,  $g_{ii} = \prod_{j \neq i} (\lambda_i - \lambda_j)$  and so  $P_i = \prod_{j \neq i} [(A - \lambda_j I) / (\lambda_i - \lambda_j)]$ . Now  $P_i^T = \partial \lambda_i / \partial A$ , is the coordinate-free form of (2.10), and so  $\partial \lambda_i / \partial A = \prod_{j \neq i} [(A - \lambda_j I) / (\lambda_i - \lambda_j)]^T = \prod_{j \neq i} [(A^T - \lambda_j I) / (\lambda_i - \lambda_j)]$ , which is (8.1).

## 9. Acknowledgments

The support received by the author from Professor Alan Laub of UCSB in the search for eigenvalue differential equations and from the Office of Naval Research through contract N-00-14-85-K-0553 was essential.

The interest of Professors H. S. Green and R. B. Potts and the staff of the Departments of Pure and Applied Mathematics of the University of Adelaide in the completion of this paper have been most helpful, especially the accurate typing of the many tedious formulas by Mary Mattaliano in Adelaide, and by Phyllis Claudio in Santa Barbara.

## References

- [1] R. E. Bargman and D. G. Nel, "On the matrix differentiation of the characteristic roots of matrices", *S. Afr. Statistic J.* **13** (1974) 135–144.
- [2] F. L. Bauer and C. T. Fike, "Norms and exclusion theorems", *Numerische Mathematik* **2** (1960) 42–43.
- [3] H. Baumgärtel, *Endlichdimensionale analytische Störungstheorie* (Akademie-Verlag, Berlin, 1972).
- [4] R. Bhatia, *Perturbation Bounds Research for Matrix Eigenvalues* (Pitman Publishing, Harlow, Essex, U.K., 1987).
- [5] C. R. Crawford, "A stable generalized eigenvalue problem", *SIAM J. Numer. Anal.* **13** (1976) 854–860.
- [6] L. Elsner, "On the variation of the spectra of matrices", *Linear Algebra and its Applications* **47** (1982) 128–138.
- [7] F. R. Gantmacher, *The Theory of Matrices*, (vol. I (transl. 1959), vol. II (transl. 1960), Chelsea Publishing, New York).
- [8] P. Henrici, "Bounds for iterates, inverses, and fields of values of non-normal matrices," *Numerische Mathematik* **4** (1962) 24–39.
- [9] W. T. Householder, *Theory of Matrices in Numerical Analysis* (Ginn-Blaisdell, Boston, 1964); reprinted Dover Publishing, New York, 1975).
- [10] T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1982).
- [11] A. Ostrowski, "Über die Stetigkeit von charakteristischen Wurzeln in Abhängigkeit von den Matrizenelementen", *Jahresbericht der Deutsch Math. Verein* **60** (1957), 40–42.
- [12] G. S. Rogers, *Matrix Derivatives* (M. Dekker, New York, 1980).
- [13] G. W. Stewart, "On the sensitivity of the eigenproblem  $Ax = \lambda Bx$ ", *SIAM J. Numer. Anal.* **4** (1972), 669–686.

- [14] G. W. Stewart, "Gerschgorin theory in the generalised eigenvalue problem  $Ax = \lambda Bx$ ", *Math. of Comp.* **29** (1975) 600–606.
- [15] W. G. Stewart, "A second-order perturbation expansion for small singular values", *Linear Algebra and its Applications* **56** (1984) 23–35.
- [16] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory* (Academic Press, New York, 1990).
- [17] T. H. Stix, *The Theory of Plasma Waves* (McGraw-Hill, New York, 1962).
- [18] J. G. Sun, *Perturbation Analysis for Generalized Eigenproblems and Generalized Singular Value Problems*, Matrix Pencils (Kagström and Ruhe, eds.), (Springer-Verlag, Berlin, 1983).
- [19] J. G. Sun, "Perturbation bounds for eigenspaces of a definite matrix pair", *Numerische Mathematik* **41** (1983) 321–343.
- [20] J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon Press, Oxford, 1965).