SIAM J. NUMER. ANAL.
Vol. 25, No. 3, June 1988

DERIVATIVES AND PERTURBATIONS OF EIGENVECTORS*

CARL D. MEYER† AND G. W. STEWART‡

Abstract. For a matrix $\mathbf{A}(z)$ whose entries are complex valued functions of a complex variable z, results are presented concerning derivatives of an eigenvector $\mathbf{x}(z)$ of $\mathbf{A}(z)$ associated with a simple eigenvalue $\lambda(z)$ when $\mathbf{x}(z)$ is restricted to satisfy a constraint of the form $\sigma(\mathbf{x}(z))=1$ where σ is a rather arbitrary scaling function. The differentiation formulas lead to a new approach for analyzing the sensitivity of an eigenvector under small perturbations in the underlying matrix. Finally, an application is given which concerns the analysis of a finite Markov chain subject to perturbations in the transition probabilities.

Key words. perturbations, eigenvectors, derivative of eigenvectors, Markov chains

AMS(MOS) subject classifications. 15A13, 65F15, 65F35, 15A12, 15A51, 15A42, 15A09

1. Introduction. For a matrix $\mathbf{A}(z)$ whose entries are complex valued functions of a complex variable z, we present results concerning derivatives of an eigenvector $\mathbf{x}(z)$ of $\mathbf{A}(z)$ associated with a simple eigenvalue $\lambda(z)$ when $\mathbf{x}(z)$ is restricted to satisfy a constraint of the form $\sigma(\mathbf{x}(z))=1$ where σ is a rather arbitrary scaling function. Our differentiation formulas lead to a new approach for analyzing the sensitivity of an eigenvector under small perturbations in the underlying matrix.

The application which motivated this study was the problem of obtaining the derivatives of the stationary probabilities associated with an irreducible finite Markov chain in order to study the effects of small perturbations in such chains. Some of the formulas derived herein are generalizations of results presented by Golub and Meyer [1986], Deutsch and Neumann [1985], Conlisk [1983], and Schweitzer [1968].

2. Background material. We shall be concerned with the perturbation of an eigenvector $\mathbf{x}$ of a matrix $\mathbf{A}_{n\times n}$ associated with a simple eigenvalue λ. Let $\mathbf{y}$ denote the corresponding left-hand eigenvector such that $\mathbf{y}^{\mathbf{H}}\mathbf{x}=1$. If $\mathbf{U}_{n\times n-1}$ is a matrix whose columns form an orthonormal basis for $R(\mathbf{A}-\lambda\mathbf{I})$, ($R(\cdot)$ will denote range and $N(\cdot)$ will denote nullspace) then $\mathbf{P}=(\boldsymbol{x}|\mathbf{U})$ is nonsingular and it is easy to verify that

$$\mathbf{P}^{-1}=\left(\frac{\mathbf{y}^{\mathbf{H}}}{\mathbf{U}^{\mathbf{H}}(\mathbf{I}-\mathbf{x}\mathbf{y}^{\mathbf{H}})}\right).$$

The matrix $\mathbf{P}^{-1}(\mathbf{A}-\lambda\mathbf{I})\mathbf{P}$ has the form

$$\mathbf{P}^{-1}(\mathbf{A}-\lambda\mathbf{I})\mathbf{P}=\begin{pmatrix}0 & \mathbf{0}\\ \mathbf{0} & \mathbf{C}-\lambda\mathbf{I}\end{pmatrix} \tag{2.1}$$

where $\mathbf{C}=\mathbf{U}^{\mathbf{H}}\mathbf{A}\mathbf{U}$. Since λ is simple, $(\mathbf{C}-\lambda\mathbf{I})$ is nonsingular and the matrix

$$(\mathbf{A}-\lambda\mathbf{I})^{\#}=\mathbf{P}\begin{pmatrix}0 & \mathbf{0}\\ \mathbf{0} & (\mathbf{C}-\lambda\mathbf{I})^{-1}\end{pmatrix}\mathbf{P}^{-1} \tag{2.2}$$

is well defined. The matrix $(\mathbf{A}-\lambda\mathbf{I})^{\#}$ is called the *group inverse* of $(\mathbf{A}-\lambda\mathbf{I})$ because it is the inverse of $(\mathbf{A}-\lambda\mathbf{I})$ in the maximal multiplicative subgroup containing $(\mathbf{A}-\lambda\mathbf{I})$.

* Received by the editors June 23, 1986; accepted for publication (in revised form) April 15, 1987.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205.
The work of this author was supported in part by National Science Foundation grant DMS-8521154.
‡ Department of Computer Science, University of Maryland, College Park, Maryland 20742.
The work of this author was supported in part by Air Force Office of Scientific Research grant AFOSR-82-0078.

Since our results will be cast in terms of the group inverse $(\mathbf{A}-\lambda\mathbf{I})^{\#}$, we will collect here its properties which we will use in the sequel. Each of the following may easily be derived from (2.2). Proofs and additional material on group inversion may be found in Campbell and Meyer [1979].

(2.3) A matrix $\mathbf{A}$ belongs to a multiplicative group $\mathcal{G}$ if and only if Rank $(\mathbf{A}^2)=$ Rank $(\mathbf{A})$. With respect to the group $\mathcal{G}$, the matrix $\mathbf{A}$ has a unique inverse $\mathbf{A}^{\#}$ and there is a unique identity element $\mathbf{E}=\mathbf{A}\mathbf{A}^{\#}=\mathbf{A}^{\#}\mathbf{A}$.

(2.4) If $\mathbf{A}$ belongs to a multiplicative group $\mathcal{G}$, then the inverse of $\mathbf{A}$ with respect to $\mathcal{G}$ is the unique matrix $\mathbf{A}^{\#}$ satisfying the three equations $\mathbf{A}\mathbf{A}^{\#}\mathbf{A}=\mathbf{A}$, $\mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#}=\mathbf{A}^{\#}$, and $\mathbf{A}\mathbf{A}^{\#}=\mathbf{A}^{\#}\mathbf{A}$. Pay attention to the fact that $\mathbf{A}^{\#}$ is different from the more familiar pseudo-inverse $\mathbf{A}^{\dagger}$ (known as the Moore–Penrose inverse). Group inversion has the desirable property that

$$(\mathbf{P}^{-1}\mathbf{A}\mathbf{P})^{\#}=\mathbf{P}^{-1}\mathbf{A}^{\#}\mathbf{P}.$$

This is generally not true for the pseudo-inverse $\mathbf{A}^{\dagger}$. It is precisely this property that makes group inversion useful when dealing with questions involving eigensystems. See property (2.8) below.

(2.5) If $\mathbf{A}$ is a group matrix and $\mathbf{b}\in R(\mathbf{A})$, then the set of all solutions for $\mathbf{u}$ in $\mathbf{A}\mathbf{u}=\mathbf{b}$ is given by $\mathbf{u}=\mathbf{A}^{\#}\mathbf{b}+N(\mathbf{A})$.

(2.6) The group identity element $\mathbf{E}=\mathbf{A}\mathbf{A}^{\#}=\mathbf{A}^{\#}\mathbf{A}$ is the projector onto $R(\mathbf{A})=R(\mathbf{A}^{\#})$ along $N(\mathbf{A})=N(\mathbf{A}^{\#})$ and the matrix $\mathbf{I}-\mathbf{E}$ is the spectral projector associated with the zero eigenvalue of $\mathbf{A}$.

(2.7) If the entries of $\mathbf{A}(z)$ are continuous functions of z on a domain D and if Rank $(\mathbf{A}(z))$ remains constant on D, then the entries of $\mathbf{A}^{\#}(z)$ are also continuous functions of z.

(2.8) For a scalar λ,

$$\lambda^{\#}=\begin{cases}1/\lambda & \text{if } \lambda\neq 0,\\ 0 & \text{if } \lambda=0.\end{cases}$$

A vector $\mathbf{x}$ is an eigenvector for $\mathbf{A}$ corresponding to the eigenvalue λ if and only if $\mathbf{x}$ is an eigenvector for $\mathbf{A}^{\#}$ corresponding to the eigenvalue $\lambda^{\#}$, i.e., $\mathbf{A}\mathbf{x}=\lambda\mathbf{x}$ if and only if $\mathbf{A}^{\#}\mathbf{x}=\lambda^{\#}\mathbf{x}$.

3. Main results. Throughout this section, assume that $\mathbf{A}=\mathbf{A}(z)$ is a matrix whose elements a_{ij} are well-defined complex valued functions of a complex variable $z=\alpha+\mathbf{i}\beta$ on some domain D. Let $\lambda=\lambda(z)$ be an eigenvalue for $\mathbf{A}$ with associated eigenvector

$$\mathbf{x}=\mathbf{x}(z)=\mathbf{u}(\alpha,\beta)+\mathbf{i}\mathbf{v}(\alpha,\beta).$$

Assume that $z_0=\alpha_0+\mathbf{i}\beta_0$ is a point in D such that $\lambda(z_0)$ is a simple eigenvalue for $\mathbf{A}(z_0)$ and that $\mathbf{A}'(z_0)$, $\mathbf{x}'(z_0)$, and $\lambda'(z_0)$ each exist. (We use prime notation to denote differentiation with respect to the complex variable z.) For another vector $\mathbf{y}_{n\times 1}(z)$ such that $\mathbf{y}'(z_0)$ also exists, let $\sigma(\mathbf{x},\mathbf{y})$ be a scalar valued function defined on C^{2n}. The function σ is usually thought of as a scaling function. One standard example of such a function is the inner product $\sigma(\mathbf{x},\mathbf{y})=\mathbf{y}^{\mathrm{H}}\mathbf{x}$.

Write $\mathbf{x}$ and $\mathbf{y}$ as

$$\mathbf{x}=\mathbf{u}(\alpha,\beta)+\mathbf{i}\mathbf{v}(\alpha,\beta),\qquad \mathbf{y}=\mathbf{p}(\alpha,\beta)+\mathbf{i}\mathbf{q}(\alpha,\beta)$$

where $\mathbf{u}$, $\mathbf{v}$, $\mathbf{p}$, $\mathbf{q}$ are real valued functions of two real variables, α and β and consider $\sigma = \sigma(\mathbf{u}, \mathbf{v}, \mathbf{p}, \mathbf{q})$ to be a function of the four real vector variables $\mathbf{u}$, $\mathbf{v}$, $\mathbf{p}$, and $\mathbf{q}$. If $\mathbf{x}'(z_0)$ and $\mathbf{y}'(z_0)$ each exist, then the Cauchy–Riemann equations hold at (α_0, β_0). That is,

$$\begin{aligned} &\mathbf{u}_\alpha(\alpha_0, \beta_0) = \mathbf{v}_\beta(\alpha_0, \beta_0) \quad \text{and} \quad \mathbf{u}_\beta(\alpha_0, \beta_0) = -\mathbf{v}_\alpha(\alpha_0, \beta_0), \\ &\mathbf{p}_\alpha(\alpha_0, \beta_0) = \mathbf{q}_\beta(\alpha_0, \beta_0) \quad \text{and} \quad \mathbf{p}_\beta(\alpha_0, \beta_0) = -\mathbf{q}_\alpha(\alpha_0, \beta_0), \end{aligned} \tag{3.1}$$

where the subscripts denote partial differentiation. Extend the subscript notation by defining $\boldsymbol{\sigma}_\mathbf{x}$ and $\boldsymbol{\sigma}_\mathbf{y}$ to be the vectors

$$\boldsymbol{\sigma}_\mathbf{x} \equiv \boldsymbol{\sigma}_\mathbf{u} + \mathbf{i}\boldsymbol{\sigma}_\mathbf{v} \quad \text{and} \quad \boldsymbol{\sigma}_\mathbf{y} \equiv \boldsymbol{\sigma}_\mathbf{p} + \mathbf{i}\boldsymbol{\sigma}_\mathbf{p}, \tag{3.2}$$

where $\boldsymbol{\sigma}_\mathbf{u}$, $\boldsymbol{\sigma}_\mathbf{v}$, $\boldsymbol{\sigma}_\mathbf{p}$, $\boldsymbol{\sigma}_\mathbf{q}$ are understood to represent the columns

$$\boldsymbol{\sigma}_\mathbf{u} = \begin{pmatrix} \partial\sigma/\partial u_1 \\ \partial\sigma/\partial u_2 \\ \vdots \\ \partial\sigma/\partial u_n \end{pmatrix}, \quad \boldsymbol{\sigma}_\mathbf{v} = \begin{pmatrix} \partial\sigma/\partial v_1 \\ \partial\sigma/\partial v_2 \\ \vdots \\ \partial\sigma/\partial v_n \end{pmatrix}, \quad \boldsymbol{\sigma}_\mathbf{p} = \begin{pmatrix} \partial\sigma/\partial p_1 \\ \partial\sigma/\partial p_2 \\ \vdots \\ \partial\sigma/p_n \end{pmatrix}, \quad \boldsymbol{\sigma}_\mathbf{q} = \begin{pmatrix} \partial\sigma/\partial q_1 \\ \partial\sigma/\partial q_2 \\ \vdots \\ \partial\sigma/\partial q_n \end{pmatrix}.$$

Our primary goal is to examine the components of the derivative $\mathbf{x}'(z_0)$ when $\mathbf{x}$ is constrained to satisfy $\sigma(\mathbf{x}, \mathbf{y}) = \kappa$ on D where κ is a real valued constant.

THEOREM 1. *Let* $\mathbf{A} = \mathbf{A}(z)$, $\lambda = \lambda(z)$, *and* $\mathbf{x} = \mathbf{x}(z)$ *be a matrix, eigenvalue, and associated eigenvector, which are defined on some domain D. Let* $\mathbf{z}_0 = \alpha_0 + \mathbf{i}\beta_0$ *be a point in D such that* $\lambda(z_0)$ *is simple and* $\mathbf{A}'(z_0)$, $\lambda'(z_0)$, *and* $\mathbf{x}'(z_0)$ *each exist. Suppose that* $\mathbf{y} = \mathbf{y}(z)$ *is a vector which is also defined on D such that* $\mathbf{y}'(z_0)$ *exists and let* $\sigma(\mathbf{x}, \mathbf{y})$ *be a function whose value is a real scalar constant for all z in D. If*

$$\boldsymbol{\sigma}_\mathbf{x}^\mathbf{H}\mathbf{x} \neq 0 \quad at\ z = z_0,$$

then the derivative of $\mathbf{x}$ *at* z_0 *is given by*

$$\mathbf{x}' = \left\{ \frac{\boldsymbol{\sigma}_\mathbf{x}^\mathbf{H}(\mathbf{A} - \lambda\mathbf{I})^\#\mathbf{A}'\mathbf{x} - \boldsymbol{\sigma}_\mathbf{y}^\mathbf{H}\mathbf{y}'}{\sigma_\mathbf{x}^\mathbf{H}\mathbf{x}} \right\} \mathbf{x} - (\mathbf{A} - \lambda\mathbf{I})^\#\mathbf{A}'\mathbf{x} \quad at\ z = z_0 \tag{3.3}$$

where $\boldsymbol{\sigma}_\mathbf{x}$ *and* $\boldsymbol{\sigma}_\mathbf{y}$ *are as defined in* (3.2).

Proof. We first must establish that the following equation holds for $z = z_0$:

$$\boldsymbol{\sigma}_\mathbf{x}^\mathbf{H}\mathbf{x}' + \boldsymbol{\sigma}_\mathbf{y}^\mathbf{H}\mathbf{y}' = 0. \tag{3.4}$$

To do this, use the fact that the Cauchy–Riemann conditions (3.1) hold at z_0 and write

$$\begin{aligned} \boldsymbol{\sigma}_\mathbf{x}^\mathbf{H}\mathbf{x}' + \boldsymbol{\sigma}_\mathbf{y}^\mathbf{H}\mathbf{y}' &= (\boldsymbol{\sigma}_\mathbf{u}^\mathbf{T} - \mathbf{i}\boldsymbol{\sigma}_\mathbf{v}^\mathbf{T})(\mathbf{u}_\alpha + \mathbf{i}\mathbf{v}_\alpha) + (\boldsymbol{\sigma}_\mathbf{p}^\mathbf{T} - \mathbf{i}\boldsymbol{\sigma}_\mathbf{q}^\mathbf{T})(\mathbf{p}_\alpha + \mathbf{i}\mathbf{q}_\alpha) \\ &= (\boldsymbol{\sigma}_\mathbf{u}^\mathbf{T}\mathbf{u}_\alpha + \boldsymbol{\sigma}_\mathbf{v}^\mathbf{T}\mathbf{v}_\alpha + \boldsymbol{\sigma}_\mathbf{p}^\mathbf{T}\mathbf{p}_\alpha + \boldsymbol{\sigma}_\mathbf{q}^\mathbf{T}\mathbf{q}_\alpha) + \mathbf{i}(\boldsymbol{\sigma}_\mathbf{u}^\mathbf{T}\mathbf{v}_\alpha - \boldsymbol{\sigma}_\mathbf{v}^\mathbf{T}\mathbf{u}_\alpha + \boldsymbol{\sigma}_\mathbf{p}^\mathbf{T}\mathbf{q}_\alpha - \boldsymbol{\sigma}_\mathbf{q}^\mathbf{T}\mathbf{p}_\alpha) \\ &= (\boldsymbol{\sigma}_\mathbf{u}^\mathbf{T}\mathbf{u}_\alpha + \boldsymbol{\sigma}_\mathbf{v}^\mathbf{T}\mathbf{v}_\alpha + \boldsymbol{\sigma}_\mathbf{p}^\mathbf{T}\mathbf{p}_\alpha + \boldsymbol{\sigma}_\mathbf{q}^\mathbf{T}\mathbf{q}_\alpha) - \mathbf{i}(\boldsymbol{\sigma}_\mathbf{u}^\mathbf{T}\mathbf{u}_\beta + \boldsymbol{\sigma}_\mathbf{v}^\mathbf{T}\mathbf{v}_\beta + \boldsymbol{\sigma}_\mathbf{p}^\mathbf{T}\mathbf{p}_\beta + \boldsymbol{\sigma}_\mathbf{q}^\mathbf{T}\mathbf{q}_\beta). \end{aligned} \tag{3.5}$$

Since $\sigma = \sigma(\mathbf{u}(\alpha, \beta), \mathbf{v}(\alpha, \beta), \mathbf{p}(\alpha, \beta), \mathbf{q}(\alpha, \beta))$ is constant on D, it follows that $\partial\sigma/\partial\alpha = \partial\sigma/\partial\beta = 0$ on D. Assuming all derivatives exist and are continuous in the proper sets, we note the chain rule yields

$$\begin{aligned} \frac{\partial\sigma}{\partial\alpha} &= \sum_i \left(\frac{\partial\sigma}{\partial u_i}\frac{\partial u_i}{\partial\alpha} + \frac{\partial\sigma}{\partial v_i}\frac{\partial v_i}{\partial\alpha} + \frac{\partial\sigma}{\partial p_i}\frac{\partial p_i}{\partial\alpha} + \frac{\partial\sigma}{\partial q_i}\frac{\partial q_i}{\partial\alpha} \right) \\ &= \boldsymbol{\sigma}_\mathbf{u}^\mathbf{T}\mathbf{u}_\alpha + \boldsymbol{\sigma}_\mathbf{v}^\mathbf{T}\mathbf{v}_\alpha + \boldsymbol{\sigma}_\mathbf{p}^\mathbf{T}\mathbf{p}_\alpha + \boldsymbol{\sigma}_\mathbf{q}^\mathbf{T}\mathbf{q}_\alpha = 0 \end{aligned}$$

and

$$\frac{\partial\sigma}{\partial\beta}=\sum_i\left(\frac{\partial\sigma}{\partial u_i}\frac{\partial u_i}{\partial\beta}+\frac{\partial\sigma}{\partial v_i}\frac{\partial v_i}{\partial\beta}+\frac{\partial\sigma}{\partial p_i}\frac{\partial p_i}{\partial\beta}+\frac{\partial\sigma}{\partial q_i}\frac{\partial q_i}{\partial\beta}\right)$$
$$=\boldsymbol{\sigma}_{\mathbf{u}}^{\mathbf{T}}\mathbf{u}_\beta+\boldsymbol{\sigma}_{\mathbf{v}}^{\mathbf{T}}\mathbf{v}_\beta+\boldsymbol{\sigma}_{\mathbf{p}}^{\mathbf{T}}\mathbf{p}_\beta+\boldsymbol{\sigma}_{\mathbf{q}}^{\mathbf{T}}\mathbf{q}_\beta=0$$

at z_0. Using these last two equations in (3.5) produces the desired conclusion that $\boldsymbol{\sigma}_{\mathbf{x}}^{\mathbf{H}}\mathbf{x}'+\boldsymbol{\sigma}_{\mathbf{y}}^{\mathbf{H}}\mathbf{y}'=0$ at z_0. We now proceed with the derivation of the derivative formula (3.3). We start with $(\mathbf{A}-\lambda\mathbf{I})\mathbf{x}=\mathbf{0}$ and apply the elementary product rule to obtain

$$(\mathbf{A}-\lambda\mathbf{I})\mathbf{x}'=-(\mathbf{A}'-\lambda'\mathbf{I})\mathbf{x}\quad\text{at } z=z_0.$$

Because $\mathbf{x}$ is a basis for $N(\mathbf{A}-\lambda\mathbf{I})$, it follows from (2.5) that there must exist a scalar δ such that

$$\mathbf{x}'=\delta\mathbf{x}-(\mathbf{A}-\lambda\mathbf{I})^{\#}(\mathbf{A}'-\lambda'\mathbf{I})\mathbf{x}\quad\text{at } z=z_0.$$

It follows from the properties of group inversion given in § 2 that $N(\mathbf{A}-\lambda\mathbf{I})=N(\mathbf{A}-\lambda\mathbf{I})^{\#}$ so that the above expression for $\mathbf{x}'$ reduces to

$$\mathbf{x}'=\delta\mathbf{x}-(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}\quad\text{at } z=z_0.$$

Now use this expression for $\mathbf{x}'$ in the relationship of (3.4) to produce

$$0=\boldsymbol{\sigma}_{\mathbf{x}}^{\mathbf{H}}\mathbf{x}'+\boldsymbol{\sigma}_{\mathbf{y}}^{\mathbf{H}}\mathbf{y}'=\delta\boldsymbol{\sigma}_{\mathbf{x}}^{\mathbf{H}}\mathbf{x}-\boldsymbol{\sigma}_{\mathbf{x}}^{\mathbf{H}}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}+\boldsymbol{\sigma}_{\mathbf{y}}^{\mathbf{H}}\mathbf{y}'\quad\text{at } z=z_0.$$

Therefore the scalar δ must be given by

$$\delta=\frac{\boldsymbol{\sigma}_{\mathbf{x}}^{\mathbf{H}}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}-\boldsymbol{\sigma}_{\mathbf{y}}^{\mathbf{H}}\mathbf{y}'}{\boldsymbol{\sigma}_{\mathbf{x}}^{\mathbf{H}}\mathbf{x}}\quad\text{at } z=z_0$$

and the desired formula given in (3.3) now follows. □

By making various choices of the scaling function σ, some insight into the problem of eigenvector sensitivity can be obtained.

COROLLARY 1. *If, in addition to the hypothesis of Theorem* 1, $\mathbf{y}=\mathbf{y}(z)$ *is a column vector such that* $\mathbf{y}^{\mathbf{H}}\mathbf{x}$ *is real valued and if* $\mathbf{y}^{\mathbf{H}}\mathbf{x}=1$ *on D, then the derivative of* $\mathbf{x}$ *at* z_0 *is given by the following expression*:

$$\mathbf{x}'=\{\mathbf{y}^{\mathbf{H}}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}-\mathbf{x}^{\mathbf{H}}\mathbf{y}'\}\mathbf{x}-(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}. \tag{3.6}$$

Proof. Take the scaling function σ of Theorem 1 to be $\sigma(\mathbf{x},\mathbf{y})=\mathbf{y}^{\mathbf{H}}\mathbf{x}$. If $\mathbf{x}=\mathbf{u}+i\mathbf{v}$ and $\mathbf{y}=\mathbf{p}+i\mathbf{q}$, where $\mathbf{u}$, $\mathbf{v}$, $\mathbf{p}$, $\mathbf{q}$, are real valued functions of α and β, then

$$\sigma(\mathbf{x},\mathbf{y})=\mathbf{y}^{\mathbf{H}}\mathbf{x}=(\mathbf{p}^{\mathbf{T}}-i\mathbf{q}^{\mathbf{T}})(\mathbf{u}+i\mathbf{v})=(\mathbf{p}^{\mathbf{T}}\mathbf{u}+\mathbf{q}^{\mathbf{T}}\mathbf{v})+i(\mathbf{p}^{\mathbf{T}}\mathbf{v}-\mathbf{q}^{\mathbf{T}}\mathbf{u})=\mathbf{p}^{\mathbf{T}}\mathbf{u}+\mathbf{q}^{\mathbf{T}}\mathbf{v}$$

because $\mathbf{y}^{\mathbf{H}}\mathbf{x}$ is assumed to be real valued. According to (3.2),

$$\boldsymbol{\sigma}_{\mathbf{x}}=\boldsymbol{\sigma}_{\mathbf{u}}+i\boldsymbol{\sigma}_{\mathbf{v}}=\mathbf{p}+i\mathbf{q}=\mathbf{y}\quad\text{and}\quad\boldsymbol{\sigma}_{\mathbf{y}}=\boldsymbol{\sigma}_{\mathbf{p}}+i\boldsymbol{\sigma}_{\mathbf{q}}=\mathbf{u}+i\mathbf{v}=\mathbf{x}$$

so that $\boldsymbol{\sigma}_{\mathbf{x}}^{\mathbf{H}}\mathbf{x}=\mathbf{y}^{\mathbf{H}}\mathbf{x}=1$ at z_0 and thus the desired result (3.6) is produced. □

If, in addition to the hypothesis of Corollary 1, $\mathbf{y}$ is assumed to be a left-hand eigenvector for $\mathbf{A}$ associated with λ (i.e., $\mathbf{y}^{\mathbf{H}}(\mathbf{A}-\lambda\mathbf{I})=\mathbf{0}$), then the following corollary is produced.

COROLLARY 2. *If* $\mathbf{y}$ *is a left-hand eigenvector for* $\mathbf{A}$ *associated with* λ *such that* $\mathbf{y}^{\mathbf{H}}\mathbf{x}=1$, *then the derivative of* $\mathbf{x}$ *at* z_0 *is given by the following expression.*

$$\mathbf{x}'=-\{(\mathbf{x}^{\mathbf{H}}\mathbf{y}')\mathbf{x}+(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}\}. \tag{3.7}$$

Perhaps the most common normalization technique is to require that $\mathbf{x}^{\mathbf{H}}\mathbf{x}=1$. By imposing this constraint, a useful formula for the derivative of $\mathbf{x}$ can be derived from which the effects of perturbations can easily be uncovered. These results are presented in the next theorem.

THEOREM 2. *If, under the hypothesis of Theorem* 1, $\mathbf{x}$ *is constrained to satisfy* $\mathbf{x}^{\mathbf{H}}\mathbf{x}=1$ *everywhere on D, then*

$$\mathbf{x}'=\{\mathbf{x}^{\mathbf{H}}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}\}\mathbf{x}-(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x} \quad \text{at } z=z_0 \tag{3.8}$$

and

$$\|\mathbf{x}'\|=|\sin\theta|\,\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}\| \quad \text{at } z=z_0 \tag{3.9}$$

where θ *is the angle between* $\mathbf{x}$ *and* $(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}$. *Moreover, if* $\mathbf{w}=\mathbf{x}(z)$ *is a left-hand eigenvector for* $\mathbf{A}$ *associated with an eigenvalue* $\mu=\mu(z)\neq\lambda(z)$ *and if* $\|\mathbf{w}\|=1$ *everywhere on D, then*

$$\frac{|\mathbf{w}^{\mathbf{H}}\mathbf{A}'\mathbf{x}|}{|\lambda-\mu|}\leqq\|\mathbf{x}'\|\leqq\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|\,\|\mathbf{A}'\| \quad \text{at } z=z_0 \tag{3.10}$$

where the vector norm is assumed to be the euclidean norm and the matrix norm can be taken to be any matrix norm which is compatible with the euclidean vector norm.

Proof. To prove (3.8), let $\mathbf{y}=\mathbf{x}$ in Corollary 1 so that $\sigma(\mathbf{x},\mathbf{y})=\sigma(\mathbf{x},\mathbf{x})=\mathbf{x}^{\mathbf{H}}\mathbf{x}$. Use the fact that (3.4) holds at z_0 to conclude that $\mathbf{x}^{\mathbf{H}}\mathbf{x}'=0$ at z_0. Thus (3.8) follows from (3.6). To derive (3.9), let $\mathbf{B}$ denote the matrix $\mathbf{B}=(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'$ and let δ denote the scalar

$$\delta=\mathbf{x}^{\mathbf{H}}\mathbf{B}\mathbf{x}=(\cos\theta)\|\mathbf{B}\mathbf{x}\|$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{B}\mathbf{x}$. It follows that

$$\begin{aligned}\|\mathbf{x}'\|^2=\mathbf{x}'^{\mathbf{H}}\mathbf{x}'&=\|\mathbf{B}\mathbf{x}\|^2-\delta^2\\&=\|\mathbf{B}\mathbf{x}\|^2-(\cos^2\theta)\|\mathbf{B}\mathbf{x}\|^2\\&=(\sin^2\theta)\|\mathbf{B}\mathbf{x}\|^2.\end{aligned}$$

To obtain the leftmost inequality of (3.10), multiply (3.8) on the left by $\mathbf{w}^{\mathbf{H}}$ and use the fact that $\mathbf{w}^{\mathbf{H}}\mathbf{x}=0$ in order to produce

$$\mathbf{w}^{\mathbf{H}}\mathbf{x}'=-\mathbf{w}^{\mathbf{H}}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}.$$

Since $\mathbf{w}$ is a left-hand eigenvector for $(\mathbf{A}-\lambda\mathbf{I})$ with associated eigenvalue $(\mu-\lambda)$, property (2.8) guarantees that $\mathbf{w}$ is a left-hand eigenvector for $(\mathbf{A}-\lambda\mathbf{I})^{\#}$ corresponding to the eigenvalue $(\mu-\lambda)^{-1}$. Thus

$$\mathbf{w}^{\mathbf{H}}\mathbf{x}'=\frac{\mathbf{w}^{\mathbf{H}}\mathbf{A}'\mathbf{x}}{\lambda-\mu}.$$

The Cauchy–Schwarz inequality now produces

$$\frac{|\mathbf{w}^{\mathbf{H}}\mathbf{A}'\mathbf{x}|}{|\lambda-\mu|}\leqq\|\mathbf{x}'\|.$$

The rightmost inequality in (3.10) is a direct consequence of (3.9). □

In passing, we remark that any component (say the kth component) of $\mathbf{x}'(z_0)$ is easily isolated. Under the hypothesis of Theorem 2, we have

$$\mathbf{x}_k' = \{\mathbf{x}^{\mathbf{H}}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}\}\mathbf{x}_k - (\mathbf{A}-\lambda\mathbf{I})^{\#}_{k\cdot}\mathbf{A}'\mathbf{x} \quad \text{at } z=z_0 \tag{3.8'}$$

where $(\mathbf{A}-\lambda\mathbf{I})^{\#}_{k\cdot}$ denotes the kth row of $(\mathbf{A}-\lambda\mathbf{I})^{\#}$.

For the case of a constant matrix, the results of Theorem 2 present a complete statement concerning the sensitivity of an eigenvector (associated with a simple eigenvalue) to pertubations in the underlying matrix. Formulas (3.8) and (3.8′) show precisely how the entries of $\mathbf{x}$ change as entries of $\mathbf{A}$ change. The leftmost inequality in (3.10) is a reaffirmation of the well-known fact that an eigenvector will exhibit sensitivities to some perturbation of the underlying matrix when the associated eigenvalue lies near another eigenvalue. The rightmost inequality in (3.10), along with the expressions (3.8) and (3.8′), show that the magnitude of the matrix $(\mathbf{A}-\lambda\mathbf{I})^{\#}$ is the measure of maximum sensitivity. Moreover, it is apparent from Theorem 2 that $(\mathbf{A}-\lambda\mathbf{I})^{\#}$ is always a multiplier on $\mathbf{A}'$ and hence $\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|$ may be interpreted as a condition number that gauges the sensitivity of the associated eigenvector.

Another standard normalization technique is to use a left-hand eigenvector $\mathbf{y}$ for $\mathbf{A}$ associated with λ such that $\mathbf{y}^{\mathbf{H}}\mathbf{y}=1$ and constrain the corresponding right-hand eigenvector $\mathbf{x}$ to satisfy $\mathbf{y}^{\mathbf{H}}\mathbf{x}=1$. For this normalization procedure, the next corollary gives the expression of the derivative of $\mathbf{x}$ at z_0.

COROLLARY 3. *If, in addition to the hypothesis of Theorem* 1, $\mathbf{y}$ *is a left-hand eigenvector associated with* λ *such that* $\mathbf{y}^{\mathbf{H}}\mathbf{y}=1$ *and if* $\mathbf{x}$ *is the corresponding right-hand eigenvector such that* $\mathbf{y}^{\mathbf{H}}\mathbf{x}=1$ *everywhere on D, then the derivative of* $\mathbf{x}$ *at* z_0 *is given by*

$$\mathbf{x}' = -\{\mathbf{y}^{\mathbf{H}}\mathbf{A}'(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{y}\}\mathbf{x} - (\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}.$$

Proof. From (3.7) in Corollary 2, we know that

$$\mathbf{x}' = -\{(\mathbf{x}^{\mathbf{H}}\mathbf{y}')\mathbf{x} + (\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x}\} \quad \text{at } z=z_0.$$

The left-hand analogue of Theorem 2 guarantees that at $z=z_0$, $\mathbf{y}'^{\mathbf{H}}$ must be given by

$$\mathbf{y}'^{\mathbf{H}} = \{\mathbf{y}^{\mathbf{H}}\mathbf{A}'(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{y}\}\mathbf{y}^{\mathbf{H}} - \mathbf{y}^{\mathbf{H}}\mathbf{A}'(\mathbf{A}-\lambda\mathbf{I})^{\#}.$$

Substituting this last expression into the preceding expression for $\mathbf{x}'(z_0)$ and using the fact that $(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{x}=\mathbf{0}$ produces the desired conclusion. □

It is interesting to observe what happens in Theorem 2 when $\mathbf{A}$ is a symmetric matrix.

COROLLARY 4. *If, in addition to the hypothesis of Theorem* 2, *the matrix* $\mathbf{A}$ *is real and symmetric, then at* $z=z_0$

$$\mathbf{x}' = -(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{A}'\mathbf{x} \tag{3.11}$$

and

$$\frac{|\mathbf{w}^{\mathbf{T}}\mathbf{A}'\mathbf{x}|}{|\lambda-\mu|} \leq \|\mathbf{x}'\| \leq \frac{\|\mathbf{A}'\|}{|\lambda-\mu|} \tag{3.12}$$

where μ $(\neq\lambda)$ *is the eigenvalue of* $\mathbf{A}$ *which is closest to* λ *and where the matrix norm is the spectral norm.*

Proof. If $\mathbf{A}$ is real and symmetric, then $\mathbf{x}^{\mathbf{T}}$ is a left-hand eigenvector for $\mathbf{A}$ associated with λ. Hence $\mathbf{x}^{\mathbf{T}}(\mathbf{A}-\lambda\mathbf{I})^{\#}=\mathbf{0}$ so that (3.8) reduces to (3.11). To obtain (3.12), note that if $\mathbf{P}$ is an orthogonal matrix such that

$$\mathbf{P}^T(\mathbf{A}-\lambda\mathbf{I})\mathbf{P}=\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2-\lambda & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3-\lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n-\lambda \end{pmatrix},$$

then

$$(\mathbf{A}-\lambda\mathbf{I})^{\#}=\mathbf{P}\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & (\lambda_2-\lambda)^{-1} & 0 & \cdots & 0 \\ 0 & 0 & (\lambda_3-\lambda)^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & (\lambda_n-\lambda)^{-1} \end{pmatrix}\mathbf{P}^T$$

so that

$$\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|_2=\frac{1}{\min_{\lambda_i\neq\lambda}|\lambda-\lambda_i|}. \qquad \square$$

Observe that (3.11) holds even in the more general case of EP matrices. That is, for matrices such that $N(\mathbf{A})=N(\mathbf{A}^H)$.

4. Relation between $\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|$ and the SEP function. Stewart [1971] defined the *separation* between two matrices $\mathbf{B}$ and $\mathbf{C}$ to be

$$\operatorname{Sep}(\mathbf{B},\mathbf{C})=\begin{cases}\dfrac{1}{\|\mathbf{T}^{-1}\|} & \text{if } \mathbf{T}^{-1} \text{ exists,}\\ 0 & \text{otherwise}\end{cases}$$

where $\mathbf{T}$ is the linear operator defined by $\mathbf{T}(\mathbf{X})=\mathbf{XB}-\mathbf{CX}$. This function was used by Stewart to bound perturbations in invariant subspaces. In the context of this paper, it is only natural to inquire about the relation between $\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|$ and $\operatorname{Sep}^{-1}(\lambda,\mathbf{C})=\|(\mathbf{C}-\lambda\mathbf{I})^{-1}\|$ where $\mathbf{C}$ is the matrix defined in (2.1).

THEOREM 3. *For $\mathbf{A}_{n\times n}$, let $\mathbf{x}$ be an eigenvector of unit 2-norm associated with the simple eigenvalue λ and let $\mathbf{U}_{n\times n-1}$ be a matrix whose columns form an orthonormal basis for $R(\mathbf{A}-\lambda\mathbf{I})$. Let $\mathbf{C}=\mathbf{U}^H\mathbf{A}\mathbf{U}$ be the matrix in (2.1). If $\mathbf{A}$ is normal, then*

(4.1) $$\operatorname{Sep}^{-1}(\lambda,\mathbf{C})=\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|.$$

In general,

(4.2) $$\operatorname{Sep}^{-1}(\lambda,\mathbf{C})=\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\|$$

and

(4.3) $$\frac{1}{\|\mathbf{A}-\lambda\mathbf{I}\|}\leqq\operatorname{Sep}^{-1}(\lambda,\mathbf{C})\leqq\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|.$$

The matrix norm is the spectral norm.

Proof. If $\mathbf{A}$ is normal, then the matrix $\mathbf{P}$ in (2.2) can be taken to be unitary and equation (4.1) clearly follows. In general, (2.2) yields

(4.4) $$(\mathbf{C}-\lambda\mathbf{I})^{-1}=\mathbf{U}^H(\mathbf{I}-\mathbf{x}\mathbf{y}^H)(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}=\mathbf{U}^H(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}.$$

Since $\mathbf{UU^H}$ is the orthogonal projector onto $R(\mathbf{A}-\lambda\mathbf{I})=R(\mathbf{A}-\lambda\mathbf{I})^{\#}$, it follows that

$$\mathbf{UU^H}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}=(\mathbf{A}+\lambda\mathbf{I})^{\#}\mathbf{U}. \tag{4.5}$$

Hence

$$\{\mathbf{U^H}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\}^{\mathbf{H}}\{\mathbf{U^H}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\}=\mathbf{U^H}(\mathbf{A}-\lambda\mathbf{I})^{\#^{\mathbf{H}}}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}$$

so that

$$\|\mathbf{U^H}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\|=\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\|\leqq\|(\mathbf{A}-\lambda\mathbf{I})^{\#}\|.$$

This, together with (4.4), produces the right-hand inequality in (4.3). The left-hand inequality in (4.3) is produced by using (4.4) and (4.5) to observe that

$$\begin{aligned}\mathbf{I}&=(\mathbf{C}-\lambda\mathbf{I})(\mathbf{C}-\lambda\mathbf{I})^{-1}\\&=\mathbf{U^H}(\mathbf{A}-\lambda\mathbf{I})\mathbf{UU^H}(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\\&=\mathbf{U^H}(\mathbf{A}-\lambda\mathbf{I})(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}.\end{aligned}$$

Use (4.2) to conclude that

$$\begin{aligned}1&=\|\mathbf{U^H}(\mathbf{A}-\lambda\mathbf{I})(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\|\\&=\|(\mathbf{A}-\lambda\mathbf{I})(\mathbf{A}-\lambda\mathbf{I})^{\#}\mathbf{U}\|\\&\leqq\|\mathbf{A}-\lambda\mathbf{I}\|\mathrm{Sep}^{-1}(\lambda,\mathbf{C}).\end{aligned}$$

□

The inequalities of Theorem 3 may be strict. For example, if

$$\mathbf{A}=\begin{pmatrix}0&0\\-1&0\end{pmatrix},$$

then for $\lambda=0$, $\mathbf{C}=[1]$ and

$$\mathrm{Sep}^{-1}(0,\mathbf{C})=1<\|\mathbf{A}^{\#}\|=\sqrt{2}.$$

5. Linear perturbations. Of particular interest is the situation where $\mathbf{A}(z)$ is linear in z. That is, let $\mathbf{A}_0$ and $\mathbf{E}$ be constant matrices and let $\mathbf{A}(z)=\mathbf{A}_0+z\mathbf{E}$ have a simple eigenvalue $\lambda(z)$ with corresponding eigenvector $\mathbf{x}(z)$ on some neighborhood about the origin. The strategy of the traditional approach to perturbational analysis given in Wilkinson [1965] is to examine the first order term in a Taylor expansion of $\mathbf{x}(z)$ about $z=0$. The analysis in Wilkinson requires $\mathbf{A}_0$ to be diagonalizable and hinges upon the expansion

$$\mathbf{x}(z)=\mathbf{x_0}+z\sum_{i=1}^{n-1}\frac{\mathbf{y_i^H E x_0}}{(\lambda_0-\lambda_i)s_i}\mathbf{x_i}+O(z^2) \tag{5.1}$$

where $\{\lambda_0,\lambda_1,\cdots,\lambda_{n-1}\}$ are the eigenvalues of $\mathbf{A_0}$ corresponding to a complete set of normalized right-hand and left-hand eigenvectors

$$\{\mathbf{x_0},\mathbf{x_1},\cdots,\mathbf{x_{n-1}}\}\quad\text{and}\quad\{\mathbf{y_0},\mathbf{y_1},\cdots,\mathbf{y_{n-1}}\},$$

respectively, and where $s_i=\mathbf{y_i^H x_i}$. In addition to the separation of λ_0 from the other eigenvalues, (5.1) suggests that the sensitivity of $\mathbf{x_0}$ also depends on the s_i $(i\neq0)$ terms. However, as Wilkinson points out, the existence of small s_i terms does not imply sensitivity in $\mathbf{x_0}$ (see Example 2). For these reasons, the expansion (5.1) can be somewhat intractable for the purpose of analyzing eigenvector sensitivity.

On the other hand, the expressions in Theorem 2 do not involve the s_i's nor are they based on the assumption that $\mathbf{A_0}$ is diagonalizable. Using (3.8) with $z_0=0$, we may write

$$\mathbf{x}(z)=\mathbf{x_0}+z\mathbf{x}(0)+O(z^2)=\mathbf{x_0}+z\{(\mathbf{x_0^H G_0 E x_0})\mathbf{I}-\mathbf{G_0E}\}\mathbf{x_0}+O(z^2) \tag{5.2}$$

where $\lambda_0=\lambda(0)$, $\mathbf{x_0}=\mathbf{x}(0)$, and $\mathbf{G_0}=(\mathbf{A_0}-\lambda_0\mathbf{I})^{\#}$. It is clear from (5.2) that the term $\mathbf{G_0}$ is the predominant factor in eigenvector sensitivity.

Example 1. Consider the matrix

$$\mathbf{A_0}=\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 2 & -1 & \cdots & -1 \\ 0 & 0 & 2 & \cdots & -1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix}_{n\times n}$$

and analyze the condition of the eigenvector

$$\mathbf{x_0}=\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

associated with the eigenvalue $\lambda_0=1$ for large values of n. The traditional approach using (5.1) is not applicable because $\mathbf{A_0}$ is deficient in eigenvectors. However, the results of this paper make it absolutely clear that $\mathbf{x_0}$ is terribly ill conditioned, in spite of the fact that $\lambda_0=1$ is well separated from the other eigenvalues. To see this, simply observe that $\mathbf{G_0}=(\mathbf{A_0}-\lambda_0\mathbf{I})^{\#}$ is given by

$$\mathbf{G_0}=\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{pmatrix}=\begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 2 & \cdots & 2^{n-3} \\ 0 & 0 & 1 & 1 & \cdots & 2^{n-4} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

where $\mathbf{T}$ is the matrix

$$\mathbf{T}=\begin{pmatrix} 1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

It is evident that $\|\mathbf{G_0}\|$ becomes huge an n grows and therefore $\mathbf{x_0}$ becomes violently ill conditioned as n grows. To corroborate this fact and to appreciate just how sensitive $\mathbf{x_0}$ is, consider the matrix $\mathbf{A}(z)=\mathbf{A_0}+z\mathbf{E_{n1}}$ where

$$\mathbf{E_{n1}}=\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}.$$

For all z, the matrix $\mathbf{A}(z)$ has exactly the same eigenvalues as $\mathbf{A_0}$ but the normalized eigenvector $\mathbf{x}(z)$ of $\mathbf{A}(z)$ associated with $\lambda(z)=1$ is extremely sensitive near $z=0$. To see this, simply verify that $\mathbf{x}(z)$ is given exactly by

$$\mathbf{x}(z)=\frac{1}{[1+z^2(4^{n-2})/3]^{1/2}}\begin{pmatrix}1\\ -z2^{n-3}\\ -z2^{n-4}\\ \vdots\\ \vdots\\ -z2^1\\ -z2^0\\ -z\end{pmatrix}.$$

For example, if $n=50$ and $z=10^{-8}$, then the first component in $\mathbf{x}(10^{-8})$ is

$$\mathbf{x}_1(10^{-8})=6.15\times 10^{-7}.$$

This means that $\mathbf{x}(10^{-8})$ is nearly orthogonal to $\mathbf{x}(0)$. If direction is neglected, then two eigenvectors of unit norm have maximal separation when they are orthogonal. Therefore, for sufficiently large n and for z near 0, the eigenvector $\mathbf{x}(z)$ given above is about as sensitive as any eigenvector can be.

Example 2. The following matrix is essentially that given on page 85 in Wilkinson [1965]:

$$\mathbf{A}=\mathbf{A}(z)=\begin{pmatrix}2 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1+z\end{pmatrix}.$$

For $z\neq 0$,

$$|s_1|=|s_2|=z/(1+z^2)^{1/2}$$

are both small when z is small. However, the eigenvector $\mathbf{x}(z)$ associated with $\lambda(z)=2$ cannot be sensitive near $z=0$. This is clear from the point of view of this paper because

$$(\mathbf{A}-2\mathbf{I})^{\#}=\begin{pmatrix}0 & 0 & 0\\ 0 & -1 & -1/(1-z)\\ 0 & 0 & -1/(1-z)\end{pmatrix}$$

has no large entries near $z=0$.

6. The effect of perturbing isolated entries in A. An important special case which is particularly revealing is the situation in which only a single entry of $\mathbf{A}$ is perturbed. Depending on the entry chosen, the effect of a small perturbation on an eigenvector can be negligible or it can be tremendous (e.g., see Example 3). It is therefore desirable to be able to predict which positions in $\mathbf{A}$ can be slightly perturbed without greatly affecting an associated eigenvector and which positions in $\mathbf{A}$, when perturbed, significantly alter the eigenvector. The analysis is easily accomplished by means of a simple linear perturbation.

As before, let $\mathbf{A}_0$ be a constant matrix with a simple eigenvalue λ_0 and a corresponding eigenvector $\mathbf{x}_0$ which has unit length. Let

$$\mathbf{A}(z)=\mathbf{A}_0+z\mathbf{E}_{ij}=\mathbf{A}_0+z\mathbf{e}_i\mathbf{e}_j^T.$$

This represents a perturbation to only the (i, j)-entry of $\mathbf{A}_0$. Let $\mathbf{x}(z)$ be a unit eigenvector of $\mathbf{A}(z)$ associated with $\lambda(z)$ such that $\mathbf{x}(0)=\mathbf{x}_0$ and $\lambda(0)=\lambda_0$. Since $\mathbf{A}'(0)=\mathbf{e}_i\mathbf{e}_j^T$, it follows that $\mathbf{A}'(0)\mathbf{x}_0=\mathbf{e}_i\mathbf{x}_{0_j}$ where $\mathbf{x}_{0_j}$ is the jth component of $\mathbf{x}_0$. Statement (3.8) of Theorem 2 reduces to

$$\mathbf{x}'(0)=\mathbf{x}_{0_j}\{\mathbf{x}_0^H(\mathbf{A}_0-\lambda_0\mathbf{I})^{\#}_{*i}\mathbf{x}_0-(\mathbf{A}_0-\lambda_0\mathbf{I})^{\#}_{*i}\} \tag{6.1}$$

where $(\mathbf{A_0}-\lambda_0 I)^{\#}_{\cdot i}$ is the ith column of $(\mathbf{A_0}-\lambda_0\mathbf{I})^{\#}$. Furthermore, it follows from the rest of Theorem 2 that

$$(6.2) \qquad \|\mathbf{x}'(0)\| = |\mathbf{x}_{\mathbf{0}_j} \sin \theta_i| \, \|(\mathbf{A_0}-\lambda_0\mathbf{I})^{\#}_{\cdot \mathbf{i}}\|$$

where θ_i is the angle between $\mathbf{x_0}$ and $(\mathbf{A_0}-\lambda_0\mathbf{I})^{\#}_{\cdot \mathbf{i}}$. Moreover, it is now easy to show that

$$(6.3) \qquad \frac{|\mathbf{w}_{\mathbf{0}_i}|}{|\lambda_0-\mu_0|} \leqq \|\mathbf{x}'(0)\| \leqq \|(\mathbf{A_0}-\lambda\mathbf{I})^{\#}_{\cdot \mathbf{i}}\|$$

where $\mathbf{w}_{\mathbf{0}_i}$ is the ith component of a unit length left-hand eigenvector, $\mathbf{w_0}$, of $\mathbf{A_0}$ with associated eigenvalue $\mu_0 \neq \lambda_0$. The observations (6.1)-(6.3) justify the following statement.

THEOREM 4. *For a constant matrix* $\mathbf{A}$, *let* $\mathbf{x}$ *be an eigenvector of unit length associated with a simple eigenvalue* λ. *The sensitivity of* $\mathbf{x}$ *to perturbations in the ith row of* $\mathbf{A}$ *is governed only by the entries of the ith column of* $(\mathbf{A}-\lambda\mathbf{I})^{\#}$ *in conjunction with the entries of* $\mathbf{x}$ *itself.*

The following example illustrates the utility of the preceding results.

Example 3. Consider the matrix of Example 1.

$$\mathbf{A_0} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 2 & -1 & \cdots & -1 \\ 0 & 0 & 2 & \cdots & -1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix}_{n\times n}.$$

The vector

$$\mathbf{x_0} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is an eigenvector for $\mathbf{A_0}$ corresponding to the simple eigenvalue $\lambda_0 = 1$. By examining the matrix

$$(\mathbf{A_0}-\mathbf{I})^{\#} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 2 & \cdots & 2^{n-3} \\ 0 & 0 & 1 & 1 & \cdots & 2^{n-4} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

it is clear that the nth column of $(\mathbf{A_0}-\mathbf{I})^{\#}$ has the greatest magnitude. Hence our results predict that the eigenvector $\mathbf{x_0}$ should exhibit maximum sensitivity when the entries of the nth row of $\mathbf{A_0}$ are perturbed. In fact, more can be said. Suppose that the (i,j)th entry of $\mathbf{A_0}$ is perturbed and let

$$\mathbf{A}(z) = \mathbf{A_0} + z\mathbf{E_{ij}}$$

be the resulting perturbed matrix with associated unit eigenvector $\mathbf{x}(z)$. Since $\mathbf{x_0^H}(\mathbf{A_0}-\mathbf{I})^{\#} = \mathbf{0}$, (6.1) reduces to

$$(6.4) \qquad \mathbf{x}'(0) = -\mathbf{x}_{\mathbf{0}_j}(\mathbf{A_0}-\mathbf{I})^{\#}_{\cdot \mathbf{i}}.$$

Since $\|(\mathbf{A_0}-\mathbf{I})^{\#}_{\cdot \mathbf{i}}\|$ becomes progressively smaller as i decreases from n to 1, (6.4) shows that $\|\mathbf{x}'(0)\|$ is maximal for $i=n, j=1$, and becomes progressively smaller as i decreases from n to 1 with j fixed at $j=1$. Furthermore, $\mathbf{x}'(0)=0$ for $j>1$. Therefore, we may conclude that $\mathbf{x}_0$ is most sensitive to a perturbation in the $(n, 1)$ entry of $\mathbf{A}_0$ and progressively less sensitive in positions $(n-1, 1)$, $(n-2, 1); \cdots, (1, 1)$.

Because $\mathbf{x}'(0)=0$ whenever $j>1$, $\mathbf{x}_0$ should be unaffected by perturbations to the (i, j)-entries for $j>1$. Indeed, this is easily verified to be true by noting that $\mathbf{A}_0$ is triangular. In Example 1 it was demonstrated just how terribly sensitive $\mathbf{x}_0$ is to a perturbation of the $(n, 1)$-entry of $\mathbf{A}_0$.

7. An application to Markov chains. An immediate application of our results concerns the problem of computing the derivatives of the stationary probabilities of an ergodic Markov chain. For this application, z is considered to be a real variable and $\mathbf{P}(z)$ is an irreducible row stochastic matrix for each z in some open interval. It follows that for each z, $\lambda(z)=1$ is a simple eigenvalue for $\mathbf{P}(z)$ and that

$$\mathbf{e}=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

is always a corresponding right-hand eigenvector. For each z, the stationary distribution $\boldsymbol{\pi}_{1\times n}(z)$ associated with $\mathbf{P}(z)$ is the left-hand eigenvector corresponding to $\lambda(z)=1$ which satisfies the condition $\boldsymbol{\pi}(z)\mathbf{e}=1$. To compute the derivatives of the stationary probabilities, use (3.7) of Corollary 2 with $\mathbf{A}$ replaced by $\mathbf{P}^T$, $\mathbf{y}$ replaced by $\mathbf{e}$, and $\mathbf{x}$ replaced by $\boldsymbol{\pi}^T$. This yields

$$\boldsymbol{\pi}'=-\boldsymbol{\pi}\mathbf{P}'(\mathbf{P}-\mathbf{I})^{\#}=\boldsymbol{\pi}\mathbf{P}'(\mathbf{I}-\mathbf{P})^{\#} \tag{7.1}$$

and

$$\boldsymbol{\pi}'_i=\boldsymbol{\pi}\mathbf{P}'(\mathbf{I}-\mathbf{P})^{\#}_{\bullet\mathbf{i}} \tag{7.1'}$$

for all z in the interval under question. These are the results of Golub and Meyer [1986]. Equation (7.1′) shows that the sensitivity of the ith stationary probability is dependent only on the magnitude of the entries in the ith column of $(\mathbf{I}-\mathbf{P})^{\#}$ in conjunction with the components of $\boldsymbol{\pi}$ itself. In fact, it was (7.1) that motivated the results of this paper.

In the analysis of a Markov chain, it is particularly important to predict the effect of a perturbation to a single pair of entries in a certain row of the transition matrix. That is, suppose that the (i, j)-entry increases by ε and the (i, r)-entry decreases by ε while all other transition probabilities remain fixed. How is $\boldsymbol{\pi}$ affected? To analyze the situation, let $\mathbf{P}_0$ be a constant transition matrix with stationary distribution $\boldsymbol{\pi}_0$ and consider the linear perturbation

$$\mathbf{P}=\mathbf{P}(z)=\mathbf{P}_0+z\mathbf{e}_\mathbf{i}(\mathbf{e}_\mathbf{j}-\mathbf{e}_\mathbf{r})^T$$

for $z\in(-\varepsilon, \varepsilon)$. Thus $\mathbf{P}'=\mathbf{e}_\mathbf{i}(\mathbf{e}_\mathbf{j}-\mathbf{e}_\mathbf{r})^T$ and hence (7.1) yields

$$\boldsymbol{\pi}'(z)=\boldsymbol{\pi}_i(z)\{(\mathbf{I}-\mathbf{P})^{\#}_{\mathbf{j}\bullet}-(\mathbf{I}-\mathbf{P})^{\#}_{\mathbf{r}\bullet}\}. \tag{7.2}$$

If $\boldsymbol{\pi}_{0_i}$ denotes the ith component of $\boldsymbol{\pi}_0$, then (7.2) implies that

$$\boldsymbol{\pi}'(0)=\boldsymbol{\pi}_{0_i}\{(\mathbf{I}-\mathbf{P}_0)^{\#}_{\mathbf{j}\bullet}-(\mathbf{I}-\mathbf{P}_0)^{\#}_{\mathbf{r}\bullet}\}$$

and

$$\boldsymbol{\pi}'_k(0)=\boldsymbol{\pi}_{0_i}\{(\mathbf{I}-\mathbf{P}_0)^{\#}_{\mathbf{jk}}-(\mathbf{I}-\mathbf{P}_0)^{\#}_{\mathbf{rk}}\}.$$

These observations justify the following statement.

THEOREM 5. *The effect on the kth stationary probability of slightly increasing p_{ij} by the same amount that p_{ir} is decreased is governed strictly by the difference of the (j, k)- and (r, k)- entries of $(\mathbf{I}-\mathbf{P})^{\#}$ in conjunction with the ith stationary probability.*

REFERENCES

S. L. CAMPBELL AND C. D. MEYER [1979], *Generalized Inverses of Linear Transformations*, Pitman, London.

J. CONLISK [1983], *Comparative statistics for Markov chains*, Univ. of California at San Diego, La Jolla, CA.

E. DEUTSCH AND M. NEUMANN [1985], *On the first and second order derivatives of the Perron vector*, Linear Algebra Appl., 71, pp. 57–76.

G. H. GOLUB AND C. D. MEYER [1986], *Using the QR factorization and group inversion to compute, differentiate, and estimate the sensitivity of stationary probabilities for Markov chains*, SIAM J. Algebraic Discrete Methods, 7, pp. 273–281.

P. SCHWEITZER [1968], *Perturbation theory and finite Markov chains*, J. Appl. Probab., 5, pp. 401–413.

G. W. STEWART [1971], *Error bounds for approximate invariant subspaces for closed linear operators*, SIAM J. Numer. Anal., 8, pp. 796–808.

J. WILKINSON [1965], *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford.