$$\boxed{\textbf{ELA}}$$

ON THE SENSITIVITY ANALYSIS OF EIGENVALUES[*]

RAFIKUL ALAM[†]

Abstract. Let λ be a simple eigenvalue of an n-by-n matrix A. Let y and x be left and right eigenvectors of A corresponding to λ, respectively. Then, for the spectral norm, the condition number $\text{cond}(\lambda, A) := \|x\|_2 \|y\|_2/|y^*x|$ measures the sensitivity of λ to small perturbations in A and plays an important role in the accuracy assessment of computed eigenvalues. R. A. Smith [Numer. Math., 10(1967), pp.232-240] proved that $\text{cond}(\lambda, A) = \|x\|_2\|y\|_2/|y^*x| = \|\text{adj}(\lambda I - A)\|_2/|p'(\lambda)|$, where $\text{adj}(A)$ is the "adjugate" of A and $p'(\lambda)$ is the derivative of $p(z) := \det(zI - A)$ at λ. We extend Smith's condition number to any matrix norm $\|\cdot\|$ and show that

$$\text{cond}(\lambda, A) = \frac{\|yx^*\|_*}{|y^*x|} = \frac{\|\text{adj}(\lambda I - A)^*\|_*}{|p'(\lambda)|}$$

measures the sensitivity of λ to small perturbations in A, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. The MATLAB command `roots` computes roots of a polynomial $p(x)$ by computing the eigenvalues of a companion matrix C_p associated with p. We analyze the sensitivity of λ as a root of $p(x)$ as well as the sensitivity of λ as an eigenvalue of C_p and compare their condition numbers.

Key words. Eigenvalue, eigenvector, condition number, sensitivity analysis, perturbation.

AMS subject classifications. 65F15, 15A18, 65F35, 15A60

Dedicated to Professor Ravindra B. Bapat on the occasion of his 60th birthday

1. Introduction. Computation of eigenvalues and eigenvectors of matrices is a major task in Numerical Linear Alegbra and the sensitivity analysis of eigenvalues plays an important role in the accuracy assessment of computed eigenvalues [14, 17, 6]. For example, the MATLAB command `[U, D] = eig(A)` provides a diagonal matrix $D \in \mathbb{C}^{n \times n}$ whose diagonal entries are computed eigenvalues and a matrix $U \in \mathbb{C}^{n \times n}$ whose columns are computed eigenvectors satisfying $(A + \Delta A)U = UD$ for some ΔA such that $\|\Delta A\|$ is bounded by a constant multiple of the unit roundoff. Thus the accuracy of the computed eigenvalues obtained by `eig` is strongly influenced by the sensitivity of the eigenvalues of A to small perturbations in the matrix A.

Let $\mathbb{C}^{n \times n}$ denote the set of n-by-n matrices with entries in $\mathbb{C}$. Let $A \in \mathbb{C}^{n \times n}$ and λ be an eigenvalue of A, that is, $\text{rank}(A - \lambda I) < n$. Then there exist nonzero vectors $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^n$ such that

$$Ax = \lambda x \text{ and } y^*A = \lambda y^*,$$

where y^* denotes the conjugate transpose of y. The vectors y and x are called left and right eigenvectors of A corresponding to λ, respectively. We refer to (λ, y, x) as an eigentriple of A. An eigenvalue λ is simple if it is a simple root of the characteristic polynomial $p(z) := \det(zI - A)$. We refer to (λ, y, x) as a simple eigentriple of A when λ is a simple eigenvalue of A. We denote the spectrum of A by $\text{eig}(A)$.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 29, pp. 223-236, March 2016

ELA

224

Let $\lambda \in \text{eig}(A)$. The condition number of λ, denoted by $\text{cond}(\lambda, A)$, is given by

$$\text{cond}(\lambda, A) := \limsup_{\|\Delta A\| \to 0} \frac{\text{dist}(\lambda, \ \text{eig}(A + \Delta A))}{\|\Delta A\|},$$

where $\text{dist}(\lambda, \text{eig}(A + \Delta A)) := \min\{|\lambda - \mu| : \mu \in \text{eig}(A + \Delta A)\}$ and $\|\cdot\|$ is a matrix norm. For the spectral norm, it was shown by Wilkinson [14, 15, 16, 17] that if (λ, y, x) is a simple eigentriple then

$$\text{cond}(\lambda, A) = \frac{\|x\|_2 \|y\|_2}{|y^* x|} \tag{1.1}$$

and that, for a sufficiently small $\|\Delta A\|_2$, there is a simple eigenvalue $\hat{\lambda} \in \text{eig}(A + \Delta A)$ such that $|\hat{\lambda} - \lambda| \leq \text{cond}(\lambda, A)\|\Delta A\|_2 + \mathcal{O}(\|\Delta A\|_2^2)$. The condition number $\text{cond}(\lambda, A)$ was introduced by Wilkinson for accuracy assessment of approximate eigenvalues computed by backward stable algorithms [14, 15]. See [12, 13] for the sensitivity analysis of multiple eigenvalues.

Starting from the condition number (1.1), Smith [10] derived an eigenvector free expression for $\text{cond}(\lambda, A)$ which is given by

$$\text{cond}(\lambda, A) = \frac{\|\text{adj}(A - \lambda I)\|_2}{|p'(\lambda)|} = \frac{\|\text{adj}(A - \lambda I)\|_2}{\prod_{\mu \neq \lambda} |\lambda - \mu|}, \tag{1.2}$$

where $\text{adj}(A - \lambda I)$ is the "adjugate" of $A - \lambda I$, $p'(\lambda)$ is the derivative of the characteristic polynomial $p(z) := \det(zI - A)$ at λ, and the product $\prod_{\mu \neq \lambda} |\lambda - \mu| = p'(\lambda)$ is taken over all the eigenvalues of A except for λ.

It is well known [17] that an eigenvalue μ of A is multiple if and only if there exist left and right eigenvectors u and v of A corresponding to μ such that $u^* v = 0$. Hence it follows from (1.1) that a highly ill-conditioned eigenvalue of A, that is, an eigenvalue with a large condition number, is expected to behave like a multiple eigenvalue when A undergoes a small perturbation. On the other hand, it follows from (1.2) that a simple eigenvalue belonging to a cluster of eigenvalues of A is expected to be highly ill-conditioned and hence behave like a multiple eigenvalue when A undergoes a small perturbation.

The main aim of this paper is to revisit sensitivity analysis of eigenvalues of matrices and roots of polynomials. More specifically, our main contributions are as follows.

• We extend (1.2) to any matrix norm. Smith derived the condition number (1.2) from (1.1) for the spectral norm but the derivation is not amenable to generalization to other matrix norms, see [10]. We take the reverse approach. First, we show that

$$\text{cond}(\lambda, A) = \frac{\|\text{adj}(A - \lambda I)^*\|_*}{|p'(\lambda)|},$$

where $\text{adj}(A - \lambda I)^*$ is the conjugate transpose of $\text{adj}(A - \lambda I)$ and $\|\cdot\|_*$ is the dual norm of a matrix norm $\|\cdot\|$. Our derivation is concise and is independent of (1.1)

ELA

and does not depend on the choice of a particular matrix norm. In particular, for the spectral norm, we deduce yet another version of $\mathrm{cond}(\lambda, A)$ in terms of the singular values of $A - \lambda I$. Indeed, we show that

$$\mathrm{cond}(\lambda, A) = \frac{\prod_{j=1}^{n-1} \sigma_j(A - \lambda I)}{|p'(\lambda)|},$$

where $\sigma_j(A - \lambda I), j = 1, \ldots, n-1$, are the nonzero singular values of $A - \lambda I$. Second, for a simple eigentriple (λ, y, x) of A, we show that $\mathrm{adj}(\lambda I - A) = p'(\lambda) x y^* / y^* x$. Hence we show that

$$\mathrm{cond}(\lambda, A) = \frac{\|\mathrm{adj}(A - \lambda I)^*\|_*}{|p'(\lambda)|} = \frac{\|y x^*\|_*}{|y^* x|}.$$

Thus we show that the eigenvector free version of $\mathrm{cond}(\lambda, A)$ and the eigenvector dependent version of $\mathrm{cond}(\lambda, A)$ are easy consequences of each other and that the derivation is independent of the choice of a particular matrix norm.

• We analyze sensitivity of roots of a scalar polynomial $p(x)$. The MATLAB command `roots` computes roots of a polynomial $p(x)$ by computing the eigenvalues of a companion matrix C_p associated with p. We therefore compare the sensitivity of λ as a root of $p(x)$ with the sensitivity of λ as an eigenvalue of C_p.

The rest of the paper is organized as follows. Section 2 presents sensitivity analysis in an abstract setting and analyzes sensitivity of roots of scalar polynomials. Section 3 presents sensitivity analysis of eigenvalues of matrices. Section 4 provides a short exposition of holomorphic perturbation of eigenvalues. Section 5 compares sensitivity of λ as a root of a polynomial $p(x)$ with the sensitivity of λ as an eigenvalue of a companion matrix C_p associated with p.

2. Sensitivity analysis. Let V be a finite dimensional Hilbert space equipped with an innerproduct $\langle, \rangle$. If $\| \cdot \| : V \to \mathbb{R}$ is a norm then $\| \cdot \|_* : V \to \mathbb{R}$ defined by $\|x\|_* := \sup_{\|y\|=1}\{|\langle x, y \rangle| : y \in V\}$ is a norm on V and is called the dual norm of $\| \cdot \|$. It follows that $|\langle x, y \rangle| \leq \|x\| \|y\|_*$ for $x \in V$ and $y \in V$.

For the special case when $V = \mathbb{C}^m$, we consider the standard innerproduct $\langle x, y \rangle := y^* x$, where y^* is the conjugate transpose of y. Similarly, when $V = \mathbb{C}^{m \times n}$, we consider the standard innerproduct $\langle X, Y \rangle := \mathrm{Tr}(Y^* X)$. Then $\|X\|_F := \sqrt{\langle X, X \rangle}$ is the Frobenius norm on $\mathbb{C}^{m \times n}$. The spectral norm on $\mathbb{C}^{n \times n}$ is given by $\|A\|_2 := \sup\{\|Ax\|_2 : \|x\|_2 = 1\}$, where $\|x\|_2 := \sqrt{\langle x, x \rangle}$.

Let $U \subset V$ be open and $\lambda : U \to \mathbb{C}$. The directional derivative (also called the Gateaux derivative) of λ at $A \in U$ in the direction of $H \in V$ is given by

$$\delta\lambda(A; H) := \lim_{t \to 0} \frac{\lambda(A + tH) - \lambda(A)}{t}$$

when the limit exists. The derivative (also called the Frechet derivative) $\mathrm{D}\lambda(A)$ of λ at $A \in U$ is a linear map $\mathrm{D}\lambda(A) : V \to \mathbb{C}$ such that

$$\lim_{\|H\| \to 0} \frac{|\lambda(A + H) - \lambda(A) - \mathrm{D}\lambda(A)H|}{\|H\|} = 0.$$

226

If $D\lambda(A)$ exists then $D\lambda(A)H = \delta\lambda(A; H)$ for $H \in V$. Further, there is a unique vector $\nabla\lambda(A) \in V$, called the gradient of λ at A, such that $D\lambda(A)H = \langle H, \nabla\lambda(A)\rangle$ for all $H \in V$. Consequently, we have

$$\|D\lambda(A)\| = \sup_{\|H\|=1} |D\lambda(A)H| = \|\nabla\lambda(A)\|_* = \sup_{\|H\|=1} |\delta\lambda(A; H)|. \qquad (2.1)$$

Suppose that λ is differentiable at $A \in U$. Then the sensitivity of $\lambda(A)$ to small perturbations in A is measured by the norm of the derivative $D\lambda(A)$. Thus the condition number of λ at A is defined by

$$\mathrm{cond}(\lambda, A) := \|D\lambda(A)\| = \|\nabla\lambda(A)\|_*. \qquad (2.2)$$

Note that the first order bound $|\lambda(A + H) - \lambda(A)| \lesssim \mathrm{cond}(\lambda, A)\|H\|$ holds for sufficiently small $\|H\|$. If $H \in V$ is such that $\langle H, \nabla\lambda(A)\rangle = \|\nabla\lambda(A)\|_*\|H\|$ then the first order bound is attained, that is, $|\lambda(A + H) - \lambda(A)| = \mathrm{cond}(\lambda, A)\|H\|$. In such a case, H is called a **fast perturbation** for $\lambda(A)$.

Often the function λ is defined implicitly and the task is to compute $\lambda(A)$ and analyze the sensitivity of λ at A. This is specially the case for eigenvalue problems. For implicitly defined functions, the implicit function theorem plays an important role.

The Implicit Function Theorem: Let $f : \mathbb{C}^n \times \mathbb{C}^m \to \mathbb{C}^m$ and $\mathbb{V}(f) := \{(x, y) \in \mathbb{C}^n \times \mathbb{C}^m : f(x, y) = 0\}$. Let $(a, b) \in \mathbb{V}(f)$. Suppose that f is holomorphic in a neighbourhood of (a, b) and that

$$\partial_y f(a, b) := \begin{bmatrix} \frac{\partial f_1}{\partial y_1}(a, b) & \cdots & \frac{\partial f_1}{\partial y_m}(a, b) \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial y_1}(a, b) & \cdots & \frac{\partial f_m}{\partial y_m}(a, b) \end{bmatrix}$$

is nonsingular. Then there is an open set $U \subset \mathbb{C}^n$ containing a and an open set $W \subset \mathbb{C}^m$ containing b and a holomorphic function $g : U \to W$ such that

$$\{(x, g(x)) : x \in U\} = \mathbb{V}(f) \cap (U \times W).$$

This shows that $g(a) = b$ and the graph of g is precisely the set of all $(x, y) \in U \times W$ such that $f(x, y) = 0$.

Let $F : V \times \mathbb{C} \to \mathbb{C}$ be a smooth function. We denote the derivatives of $F(X, z)$ at (A, λ) with respect to X and z by $\partial_X F(A, \lambda)$ and $\partial_z F(A, \lambda)$, respectively. Then for an implicitly defined function, we have the following result.

THEOREM 2.1. *Let $F : V \times \mathbb{C} \to \mathbb{C}$ be a smooth function and $(A, \lambda_A) \in V \times \mathbb{C}$ be such that $F(A, \lambda_A) = 0$. Suppose that $\partial_z F(A, \lambda_A) \neq 0$. Then there is an open set $U \subset V$ containing A and a smooth function $\lambda : U \to \mathbb{C}$ such that $\lambda(A) = \lambda_A$ and $F(X, \lambda(X)) = 0$ for all $X \in U$. Moreover, for $X \in U$, we have*

$$D\lambda(X)H = -\partial_X F(X, \lambda(X))H/\partial_z F(X, \lambda(X)), \quad \text{for } H \in V, \qquad (2.3)$$

ELA

where $\partial_X F(X, \lambda(X))$ and $\partial_z F(X, \lambda(X))$ are derivatives of $F(X, z)$ with respect to X and z, respectively, evaluated at $(X, \lambda(X))$. Thus the sensitivity of λ at A is measured by the condition number

$$\text{cond}(\lambda_A, A) := \frac{\|\partial_X F(A, \lambda_A)\|}{|\partial_z F(A, \lambda_A)|}. \tag{2.4}$$

Proof. Since $\partial_z F(A, \lambda_A) \neq 0$, by the Implicit Function Theorem, the smooth function $\lambda : U \to \mathbb{C}$ exists satisfying the stated conditions. Again since $F(X, \lambda(X)) = 0$ for all $X \in U$, differentiating with respect to X, we have $0 = \partial_X F(X, \lambda(X))H + \partial_z F(X, \lambda(X))\text{D}\lambda(X)H$ for $H \in V$. Hence evaluating the derivative at A and using the fact that $\lambda(A) = \lambda_A$, the desired results follow. $\square$

Let $\mathcal{P}_m \subset \mathbb{C}[x]$ be the subspace of polynomials of degree at most m. For $p \in \mathcal{P}_m$, let $[p] \in \mathbb{C}^{m+1}$ denote the coordinate of p with respect to the ordered basis $(1, x, \ldots, x^m)$ of $\mathcal{P}_m$. Then for $p(x) = \sum_{j=0}^m a_j x^j$ we have $[p] = [a_0, a_1, \ldots, a_m]^\top$. Obviously the map $\mathcal{P}_m \to \mathbb{C}^{m+1}, p \mapsto [p]$ is an isomorphism and $\langle p, q \rangle_m := [q]^*[p]$ defines an innerproduct on $\mathcal{P}_m$. If $\|\cdot\|$ is a norm on $\mathbb{C}^{m+1}$ then $\|p\| := \|[p]\|$ defines a norm on $\mathcal{P}_m$ and $\|p\|_* := \sup\{|\langle q, p \rangle_m| : \|q\| = 1\}$ is the dual norm on $\mathcal{P}_m$. For $p(x) := \sum_{j=0}^m a_j x^j$, the conjugate polynomial $\bar{p}$ is given by $\bar{p}(x) := \sum_{j=0}^m \bar{a}_j x^j$.

THEOREM 2.2. *Let* $p(x) := \sum_{j=0}^m a_j x^j$ *and* λ_0 *be a simple root of* $p(x)$. *Then there is an open set* $U \subset \mathcal{P}_m$ *containing* p *and a smooth function* $\lambda : U \to \mathbb{C}$ *such that* $\lambda(p) = \lambda_0$ *and* $\lambda(s)$ *is a simple root of* $s(x)$ *for all* $s \in U$. *Further, for* $s \in U$, *we have*

$$\text{D}\lambda(s)h = -\frac{h(\lambda)}{s'(\lambda)} = \langle h, -\bar{\Lambda}_m(s)/\bar{s}'(\lambda) \rangle_m \text{ and } \nabla\lambda(s) = -\frac{\bar{\Lambda}_m(s)}{\bar{s}'(\lambda)}$$

for all $h \in \mathcal{P}_m$, *where* $s'(\lambda)$ *is the derivative of* $s(x)$ *evaluated at* $\lambda(s)$ *and* $\Lambda_m(s)$ *is a polynomial given by* $\Lambda_m(s)(x) := \sum_{j=0}^m (\lambda(s))^j x^j$. *The condition number* $\text{cond}(\lambda_0, p)$ *of the root* λ_0 *of* $p(x)$ *is given by*

$$\text{cond}(\lambda_0, p) = \|\text{D}\lambda(p)\| = \frac{\|\bar{\Lambda}_m(p)\|_*}{|p'(\lambda_0)|} = \frac{\|[1, \lambda_0, \ldots, \lambda_0^m]^*\|_*}{|p'(\lambda_0)|}.$$

In particular, if $p(x)$ *is monic then the sensitivity of* λ_0 *relative to perturbations in the coefficients* $a_0, \ldots, a_{m-1}$ *is measured by the condition number*

$$\text{cond}^{\mathbf{S}}(\lambda_0, p) = \frac{\|\bar{\Lambda}_{m-1}(p)\|_*}{|p'(\lambda_0)|} = \frac{\|[1, \lambda_0, \ldots, \lambda_0^{m-1}]^*\|_*}{|p'(\lambda_0)|}.$$

Proof. Define $F : \mathcal{P}_m \times \mathbb{C} \to \mathbb{C}$ by $F(s, z) = s(z)$. Then $\partial_z F(p, \lambda_0) = p'(\lambda_0) \neq 0$ and $\partial_s F(s, z)h = h(z)$ for all $h \in \mathcal{P}_m$. Hence $\text{D}\lambda(s), \nabla\lambda(s)$ and $\text{cond}(\lambda_0, p)$ follow from Theorem 2.1. If $p(x)$ is monic and the perturbations are restricted to $a_0, \ldots, a_{m-1}$ then $p(x)$ is perturbed to $p(x) + b(x)$ with $b(x) := \sum_{j=0}^{m-1} b_j x^j$. Hence the sensitivity of λ_0 is measured by the condition number $\text{cond}^{\mathbf{S}}(\lambda_0, p) = \sup\{|\text{D}\lambda(p)h| : h \in \mathcal{P}_{m-1}\} = \sup\{|h(\lambda_0)/p'(\lambda_0)| : h \in \mathcal{P}_{m-1}\} = \frac{\|\bar{\Lambda}_{m-1}(p)\|_*}{|p'(\lambda_0)|} = \frac{\|[1, \lambda_0, \ldots, \lambda_0^{m-1}]^*\|_*}{|p'(\lambda_0)|}$ $\square$

228

Suppose that $p \in \mathcal{P}_m$ is a monic polynomial and λ_0 is a simple root of $p(x)$. Then observe that $|p'(\lambda_0)| = \prod_{\mu \neq \lambda_0} |\lambda_0 - \mu| =: \delta_p(\lambda_0)$, where the product is taken over all the roots μ of $p(x)$ except for λ_0. This shows that if $\delta_p(\lambda_0)$ is small, that is, if λ_0 is not well separated from rest of the roots of p, then λ_0 is expected to be sensitive to small perturbations in $p(x)$. By Theorem 2.2, we have

$$\lambda(p + \Delta p) = \lambda_0 - \langle \Delta p, \, \bar{\Lambda}_m(p) \rangle_m / p'(\lambda_0) + \mathcal{O}(\|\Delta p\|^2)$$

for all $\Delta p \in \mathcal{P}_m$ such that $\|\Delta p\|$ is small. Note that $\lambda(p) = \lambda_0$. Hence the first order bound $|\lambda(p + \Delta p) - \lambda(p)| \lesssim \mathrm{cond}(\lambda_0, p) \|\Delta p\|$ holds. A polynomial Δp_* is said to be a *fast perturbation* for $\lambda(p)$ if

$$\langle \Delta p, \, \bar{\Lambda}_m(p) \rangle_m / p'(\lambda_0) = \|\bar{\Lambda}_m(p)\|_* \|\Delta p_*\| / \delta_p(\lambda_0).$$

In such a case, we have $|\lambda(p + \Delta p_*) - \lambda(p)| = \mathrm{cond}(\lambda_0, p) \|\Delta p_*\|$ showing that the first order bound for $|\lambda(p + \Delta p) - \lambda(p)|$ is attained at Δp_*.

A fast perturbation can be constructed as follows. Let $\| \cdot \|$ be a norm on $\mathbb{C}^m$ and $x_0 \in \mathbb{C}^m$ be nonzero. Then it is well known that [18, 19]

$$\partial \|x_0\| = \{ y \in \mathbb{C}^m : \langle x_0, \, y \rangle = \|x_0\| \text{ and } \|y\|_* = 1 \}$$

is the subdifferential (subgradient) of the map $x \mapsto \|x\|$ at x_0, where $\langle x, \, y \rangle := y^* x$. If $\| \cdot \|$ is differentiable at x_0 then $\partial \|x_0\| = \{ \nabla \|x_0\| \}$. In such a case, we have $\|(\nabla \|x_0\|)\|_* = 1$ and $\langle x_0, \nabla \|x_0\| \rangle = \|x_0\|$. For example, if $\| \cdot \|$ is strictly convex then it is differentiable on $\mathbb{C}^m \setminus \{0\}$. For the special case of the Hölder p-norm on $\mathbb{C}^m$, it is easy to determine the subdifferential $\partial \|x_0\|$, see [1, 18, 19].

Thus a fast perturbation for the root $\lambda(p) = \lambda_0$ of $p(x)$ is a scaled polynomial in $\partial \|\bar{\Lambda}_m(p) / p'(\lambda_0)\|_*$. Indeed, for $p_* \in \partial \|\bar{\Lambda}_m(p) / p'(\lambda_0)\|_*$ setting $\Delta p_*(x) := \epsilon \, p_*(x)$, we have $\langle \Delta p_*, \, \bar{\Lambda}_m(p) \rangle_m / p'(\lambda_0) = \|\bar{\Lambda}_m(p)\|_* \|\Delta p_*\| / \delta_p(\lambda_0)$ and hence the first order bound $|\lambda(p + \Delta p_*) - \lambda(p)| = \mathrm{cond}(\lambda_0, p) \|\Delta p_*\|$ holds for sufficiently small $\epsilon > 0$.

Generically, the condition number of a problem is inversely proportional to the distance from the problem to the nearest ill-posed problem [7, 15, 16]. This is easily verified for roots of polynomials. Suppose that $p(x)$ is monic and has m distinct roots $\lambda_1, \ldots, \lambda_m$. Let $\widehat{\mathcal{P}}_m$ denote the set of polynomials in $\mathcal{P}_m$ having a multiple root. Set $\mathrm{dist}(p, \widehat{\mathcal{P}}_m) := \inf \{ \|p - s\| : s \in \widehat{\mathcal{P}}_m \}$. Then we have

$$\mathrm{dist}(p, \widehat{\mathcal{P}}_m) \leq \min_j \frac{\| [\lambda_j, 1]^\top \| \, \| [1, \lambda_j, \ldots, \lambda_j^m]^* \|_*}{\mathrm{cond}(\lambda_j, p)}.$$

Indeed, we have $p(x) = p'(\lambda_j)(x - \lambda_j) + \cdots + p^{(m)}(\lambda_j)(x - \lambda_j)^m / m!$. Now defining $\Delta p(x) := -p'(\lambda_j)(x - \lambda_j)$, it follows that $(p + \Delta p)(x) = q(x)(x - \lambda_j)^2$ has a double root at λ_j. Since $\|\Delta p\| = |p'(\lambda_j)| \, \| [\lambda_j, 1]^\top \| = \| [\lambda_j, 1]^\top \| \, \| [1, \lambda_j, \ldots, \lambda_j^m]^* \|_* / \mathrm{cond}(\lambda_j, p)$, the desired result follows.

3. Sensitivity of eigenvalues. We now generalize Smith's version of the condition number of a simple eigenvalue of a matrix to the case of an arbitrary matrix norm. Our derivation is based on derivatives of eigenvalues and is independent of

Wilkinson's version of the condition number. We proceed as follows. The adjugate of a matrix $A \in \mathbb{C}^{n \times n}$, denoted by $\mathrm{adj}(A)$, is defined by

$$(\mathrm{adj}(A))_{ji} := (-1)^{i+j} \det(A(i,j)),$$

where $A(i,j)$ is the matrix obtained from A by deleting the i-th row and the j-th column of A. Let $\det : \mathbb{C}^{n \times n} \to \mathbb{C}$ denote the determinant map which takes a matrix X to its determinant $\det(X)$. It is well known that $A \, \mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)I$. Also det is a differentiable function and the derivative $\mathrm{D} \det(A)$ at A is given by the Jacobi formula

$$\mathrm{D} \det(A)X = \mathrm{Tr}(\mathrm{adj}(A)X) = \langle X, \mathrm{adj}(A)^* \rangle \ \text{ for } X \in \mathbb{C}^{n \times n}. \qquad (3.1)$$

Note that $\nabla \det(A) := \mathrm{adj}(A)^*$ is the gradient of det at A. Also note that $\mathrm{adj}(A) = 0$ when $\mathrm{rank}(A) < n-1$ and that $\mathrm{adj}(A) \neq 0$ when $\mathrm{rank}(A) \geq n-1$. For the special case when $\mathrm{rank}(A) = n-1$, the following elementary result holds.

THEOREM 3.1. *Let $A \in \mathbb{C}^{n \times n}$. Suppose that $\mathrm{rank}(A) = n-1$. Then $\mathrm{adj}(A) = vu^*$ for some nonzero vectors u and v such that $Av = 0$ and $u^*A = 0$. Consequently, we have $\mathrm{D} \det(A)H = \mathrm{Tr}(\mathrm{adj}(A)H) = u^*Hv$ for all $H \in \mathbb{C}^{n \times n}$ and $\nabla \det(A) = (\mathrm{adj}(A))^* = uv^*$. In particular, if λ is a geometrically simple eigenvalue of A and $p(z) := \det(zI - A)$ then $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(\lambda I - A)) = y^*x$ for some left eigenvector y and right eigenvector x of A corresponding λ, where $p'(\lambda)$ is the derivative of p at λ.*

Proof. Let $\mathrm{R}(A)$ and $\mathrm{N}(A)$ denote the range space and the null space of A, respectively. Set $X := \mathrm{adj}(A)$. Then $XA = AX = \det(A)I = 0$. Now $AX = 0$ implies that $\mathrm{R}(X) \subset \mathrm{N}(A)$. This shows that X is a rank one matrix. Hence X is given by $X = vu^*$ for some nonzero vectors u and v. Since $\mathrm{R}(X) = \mathrm{span}(v) \subset \mathrm{N}(A)$, we have $Av = 0$. Similarly, $XA = 0 \Rightarrow A^*X^* = 0 \Rightarrow \mathrm{R}(X^*) = \mathrm{span}(u) \subset \mathrm{N}(A^*)$, that is, $u^*A = 0$. By the Jacobi's formula (3.1), $\mathrm{D} \det(A)H = \mathrm{Tr}(\mathrm{adj}(A)H) = u^*Hv$ for all $H \in \mathbb{C}^{n \times n}$ and hence $\nabla \det(A) = (\mathrm{adj}(A))^* = uv^*$. Finally, by the chain rule, we have $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(\lambda I - A)) = y^*x$. $\square$

A well known result due to Wilkinson [17] states that λ is a multiple eigenvalue of A if and only if there exist left and right eigenvectors u and v, respectively, of A corresponding to λ such that $u^*v = 0$. This is immediate from Theorem 3.1 when λ is geometrically simple, that is, $\mathrm{rank}(A - \lambda I) = n-1$. Indeed, in such a case, $p'(\lambda) = y^*x$ which shows that $p'(\lambda) = 0 \Leftrightarrow y^*x = 0$. In fact, a more general result holds.

Let $A : \mathbb{C} \to \mathbb{C}^{n \times n}$ be meromorphic and **regular**, that is, $\det(A(z)) \neq 0$ for some $z \in \mathbb{C}$. Let $\lambda \in \mathbb{C}$. If $\mathrm{rank}(A(\lambda)) < n$ then λ is said to be an **eigenvalue** of $A(z)$. If λ is an eigenvalue of $A(z)$ then there exist nonzero vectors x and y such that $A(\lambda)x = 0$ and $y^*A(\lambda) = 0$. The vectors x and y are called right and left eigenvectors of $A(z)$ corresponding to λ, respectively. If λ is a multiple root of $p(z) := \det(A(z))$ then λ is said to be a multiple eigenvalue of $A(z)$.

THEOREM 3.2. *Let $A : \mathbb{C} \to \mathbb{C}^{n \times n}$ be meromorphic and regular. Set $p(z) := \det(A(z))$. Then we have $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(A(\lambda))A'(\lambda))$, where $p'(\lambda)$ and $A'(\lambda)$ are the derivatives of $p(z)$ and $A(z)$ at λ, respectively. Further, λ is a multiple eigenvalue of $A(z)$ if and only if there exist left and right eigenvectors u and v, respectively, of $A(z)$ corresponding to λ such that $u^*A'(\lambda)v = 0$.*

ELA

230

Proof. By Jacobi formula and the chain rule, we have $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(A(\lambda))A'(\lambda))$. Further, if $\mathrm{rank}(A(\lambda)) = n - 1$ then by Theorem 3.1, we have $\mathrm{adj}(A(\lambda)) = vu^*$ for some left eigenvector u and right eigenvector v of $A(z)$ corresponding to λ.

Suppose that λ is a multiple eigenvalue. Then $p'(\lambda) = 0$. If $\mathrm{rank}(A(\lambda)) = n - 1$ then $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(A(\lambda))A'(\lambda)) = u^*A'(\lambda)v = 0$. On the other hand, if $\mathrm{rank}(A(\lambda)) < n - 1$ then, for any nonzero u such that $u^*A(\lambda) = 0$, the map

$$\phi : \mathrm{N}(A(\lambda)) \to \mathbb{C}, \, x \mapsto u^*A'(\lambda)x$$

is a linear functional and hence $\phi(v) = 0$ for some nonzero $v \in \mathrm{N}(A(\lambda))$. Hence the desired result follows.

Conversely, suppose that there exist left and right eigenvectors y and x such that $y^*A'(\lambda)x = 0$. If $\mathrm{rank}(A(\lambda)) < n - 1$ then $\mathrm{adj}(A(\lambda)) = 0$. Consequently, we have $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(A(\lambda))A'(\lambda)) = 0$ showing that λ is a multiple eigenvalue. On the other hand, if $\mathrm{rank}(A(\lambda)) = n - 1$ then $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(A(\lambda))A'(\lambda)) = u^*A'(\lambda)v = \alpha\bar{\beta}\, y^*A'(\lambda)x = 0$, where $v = \alpha x$ and $u = \beta y$. This shows that λ is a multiple eigenvalue. $\square$

Now consider the map $F : \mathbb{C}^{n \times n} \times \mathbb{C} \to \mathbb{C}$ given by $F(X, s) := \det(sI - X)$. If λ_A is an eigenvalue of A then obviously (A, λ_A) is a solution of $F(X, s) = 0$. So, we consider the algebraic variety $\mathbb{V}(F) := \{(X, s) \in \mathbb{C}^{n \times n} \times \mathbb{C} : F(X, s) = 0\}$. We say that (A, λ_A) is a **simple point** of $\mathbb{V}(F)$ if λ_A is a simple eigenvalue of A. We now show that if (A, λ_A) is a simple point of $\mathbb{V}(F)$ then there is an open set Ω containing A such that $\mathbb{V}(F) \cap (\Omega \times \mathbb{C})$ is the graph of a smooth function $\Omega \to \mathbb{C}, X \mapsto \lambda(X)$ such that $\lambda(A) = \lambda_A$.

Recall that $\langle X, Y \rangle := \mathrm{Tr}(Y^*X)$ for $X, Y \in \mathbb{C}^{n \times n}$ is the standard innerproduct on $\mathbb{C}^{n \times n}$ and $\|Y\|_* = \sup_{\|X\|=1} |\langle X, Y \rangle|$ is the dual norm of a norm $\| \cdot \|$ on $\mathbb{C}^{n \times n}$.

THEOREM 3.3. *Let $A \in \mathbb{C}^{n \times n}$ and λ_A be a simple eigenvalue of A. Then there is an open set $\Omega \subset \mathbb{C}^{n \times n}$ containing A and a smooth function $\lambda : \Omega \to \mathbb{C}$ such that $\lambda(A) = \lambda_A$ and $\lambda(X)$ is a simple eigenvalue of X for all $X \in \Omega$. Moreover, for $X \in \Omega$, we have*

$$\mathrm{D}\lambda(X)H = \frac{\mathrm{Tr}(\mathrm{adj}(\lambda(X)I - X)H)}{\partial_s \det((\lambda(X)I - X)} \; \text{and} \; \nabla\lambda(X) = \left(\frac{\mathrm{adj}(\lambda(X)I - X)}{\partial_s \det(\lambda(X)I - X)} \right)^*,$$

for all $H \in \mathbb{C}^{n \times n}$, where $\partial_s \det(\lambda(X)I - X)$ is the partial derivative of $\det(sI - X)$ with respect to s evaluated at $(\lambda(X), X)$. Thus for any matrix norm, the sensitivity of the eigenvalue $\lambda(X)$ is measured by the condition number

$$\mathrm{cond}(\lambda, X) := \frac{\|(\mathrm{adj}(\lambda(X)I - X))^*\|_*}{|\partial_s \det(\lambda(X)I - X)|} = \frac{\|(\mathrm{adj}(\lambda(X)I - X))^*\|_*}{|\mathrm{Tr}(\mathrm{adj}(\lambda(X)I - X))|}.$$

Proof. Consider the map $F : \mathbb{C}^{n \times n} \times \mathbb{C} \to \mathbb{C}, (X, s) \longmapsto \det(sI - X)$ and the variety $\mathbb{V}(F) := \{(X, s) \in \mathbb{C}^{n \times n} \times \mathbb{C} : F(X, s) = 0\}$. Obviously $(A, \lambda_A) \in \mathbb{V}(F)$. The derivative of $F(X, s)$ with respect to s evaluated at (A, λ_A) is given by $\partial_s F(A, \lambda_A) = p'(\lambda_A) \neq 0$. Hence by the Implicit function theorem there is an open set Ω containing A and a smooth function $\lambda : \Omega \to \mathbb{C}$ such that $\lambda(A) = \lambda_A$ and $\mathbb{V}(F) \cap (\Omega \times \mathbb{C}) =$

ELA

$\{(X, \lambda(X)) : X \in \Omega\}$. In other words, $X \mapsto \lambda(X)$ is a smooth function on Ω such that $\lambda(A) = \lambda_A$ and $F(X, \lambda(X)) = 0$ for $X \in \Omega$. Since $\mathbb{V}(F) \cap (\Omega \times \mathbb{C}) = \{(X, \lambda(X)) : X \in \Omega\}$ is the graph of $X \mapsto \lambda(X)$ for $X \in \Omega$, it follows that $\lambda(X)$ is a simple eigenvalue of X for all $X \in \Omega$. Indeed, if $\lambda(X_0)$ is a multiple eigenvalue of X_0 for some $X_0 \in \Omega$ then the eigenvalue λ_A and an eigenvalue $\mu_A \neq \lambda_A$ of A must move and coalesce at $\lambda(X_0)$ when X varies from A to X_0. However, the intersection of two eigenvalue paths $(X, \lambda(X))$ and $(X, \mu(X))$ with $\mu(A) = \mu_A$ at $(X_0, \lambda(X_0))$ would contradict the fact that $\mathbb{V}(F) \cap (\Omega \times \mathbb{C}) = \{(X, \lambda(X)) : X \in \Omega\}$ is the graph of the map $X \mapsto \lambda(X)$ for $X \in \Omega$.

Since $F(X, \lambda(X)) = 0$ for $X \in \Omega$, differentiating with respect to X, we have $\mathrm{D}\lambda(X)H = -\partial_X F(X, \lambda(X))H/\partial_s F(X, \lambda(X))$ for $H \in \mathbb{C}^{n \times n}$, where $\partial_X F(X, \lambda(X))$ and $\partial_s F(X, \lambda(X))$ are derivatives of $F(X, s)$ with respect to X and s, respectively, evaluated at $(X, \lambda(X))$. By Jacobi formula, $\partial_X F(X, \lambda(X))H = -\mathrm{Tr}(\mathrm{adj}(\lambda(X)I - X)H)$ and $\partial_s F(X, \lambda(X)) = \mathrm{Tr}(\mathrm{adj}(\lambda(X)I - X))$. Hence the desired results follow from Theorem 2.1. □

Observe that the condition number in Theorem 3.3 generalizes Smith's condition number (1.2) to the case of an arbitrary norm on $\mathbb{C}^{n \times n}$ when $X = A$. We now show that Smith's condition number of a simple eigenvalue is an immediate consequence of Wilkinson's condition number (1.1) and vice-versa. This follows from a representation of $\mathrm{adj}(A)$ in terms of eigenvalues and eigenvectors of A. Note that if $\mathrm{rank}(A) = n - 1$ then by Theorem 3.1, we have $\mathrm{adj}(A) = vu^*$ for some left and right eigenvectors u and v of A corresponding to 0. On the other hand, if 0 is a simple eigenvalue of A then for any left and right eigenvectors y and x of A corresponding to 0, it is easy to see that $\mathrm{adj}(A) = -\mathrm{Tr}(\mathrm{adj}(A))xy^*/y^*x$. More generally, we have the following result.

THEOREM 3.4. *Let* $A : \mathbb{C} \to \mathbb{C}^{n \times n}$ *be meromorphic and regular and let* (λ, y, x) *be a simple eigentriple of* $A(z)$. *Set* $p(z) := \det(A(z))$. *Then we have*

$$\mathrm{adj}(A(\lambda)) = \frac{p'(\lambda)xy^*}{y^*A'(\lambda)x} = \frac{\mathrm{Tr}(\mathrm{adj}(A(\lambda))A'(\lambda))xy^*}{y^*A'(\lambda)x},$$

where $p'(\lambda)$ *and* $A'(\lambda)$ *are the derivatives of* $p(z)$ *and* $A(z)$ *at* λ, *respectively.*

Proof. By Theorem 3.1, we have $\mathrm{adj}(A(\lambda)) = vu^*$ for some nonzero vectors u and v such that $A(\lambda)v = 0$ and $u^*A(\lambda) = 0$. Since λ is a simple eigenvalue, we have $v = \alpha x$ and $u = \beta y$ for some scalars α and β. Thus $\mathrm{adj}(A(\lambda)) = \alpha\bar{\beta}\, xy^*$. By Jacobi formula, we have $p'(\lambda) = \mathrm{Tr}(\mathrm{adj}(A(\lambda))A'(\lambda)) = \alpha\bar{\beta}\,\mathrm{Tr}(xy^*A'(\lambda)) = \alpha\bar{\beta}\, y^*A'(\lambda)x$ which gives $\alpha\bar{\beta} = p'(\lambda)/y^*A'(\lambda)x$. Hence the desired result follows. □

Now, by considering $A(z) := zI - A$ and a simple eigentriple (λ, x, y) of A, it follows from Theorem 3.4 that

$$\frac{\mathrm{adj}(\lambda I - A)}{p'(\lambda)} = \frac{xy^*}{y^*x}. \tag{3.2}$$

Hence the Smith condition number $\mathrm{cond}(\lambda, A) = \|\mathrm{adj}(A - \lambda I)^*\|_*/|p'(\lambda)|$ is an immediate consequence of the Wilkinson condition number $\mathrm{cond}(\lambda, A) = \|xy^*\|_*/|y^*x|$ and vice-versa. Indeed, we have the following result.

THEOREM 3.5. *Let* $A \in \mathbb{C}^{n \times n}$. *Let* (λ_A, y, x) *be a simple eigentriple of* A. *Then there is an open set* $\Omega \subset \mathbb{C}^{n \times n}$ *containing* A *and a smooth function* $\lambda : \Omega \to \mathbb{C}$ *such*

ELA

that $\lambda(A) = \lambda_A$ and $\lambda(X)$ is a simple eigenvalue of X for all $X \in \Omega$. Moreover,

$$\mathrm{D}\lambda(A)H = \frac{\mathrm{Tr}(\mathrm{adj}(\lambda_A I - A)H)}{p'(\lambda_A)} = \frac{\mathrm{Tr}(\mathrm{adj}(\lambda_A I - A)H)}{\mathrm{Tr}(\mathrm{adj}\,(\lambda_A I - A))} = \frac{y^* H x}{y^* x},$$

$$\nabla\lambda(A) = \left(\frac{\mathrm{adj}(\lambda_A I - A)}{p'(\lambda_A)}\right)^* = \left(\frac{\mathrm{adj}(\lambda_A I - A)}{\mathrm{Tr}(\mathrm{adj}\,(\lambda_A I - A))}\right)^* = \frac{y x^*}{x^* y},$$

for all $H \in \mathbb{C}^{n \times n}$, where $p(z) := \det(zI - A)$. For any matrix norm, the condition number $\mathrm{cond}(\lambda_A, A)$ is given by

$$\mathrm{cond}(\lambda_A, A) = \frac{\|(\mathrm{adj}(A - \lambda_A I))^*\|_*}{|p'(\lambda_A)|} = \frac{\|(\mathrm{adj}(A - \lambda_A I))^*\|_*}{|\mathrm{Tr}(\mathrm{adj}\,(\lambda_A I - A))|} = \frac{\|y x^*\|_*}{|y^* x|}.$$

Further, for a subordinate matrix norm, we have

$$\mathrm{cond}(\lambda_A, A) = \frac{\|(\mathrm{adj}(A - \lambda_A I))^*\|_*}{|p'(\lambda_A)|} = \frac{\|y x^*\|_*}{|y^* x|} = \frac{\|y\|_* \|x\|}{|y^* x|}.$$

Furthermore, for the spectral norm (as well as the Frobenius norm), we have

$$\mathrm{cond}(\lambda_A, A) = \frac{\|(\mathrm{adj}(A - \lambda_A I))\|_2}{|p'(\lambda_A)|} = \frac{\prod_{j=1}^{n-1} \sigma_j(A - \lambda_A I)}{\prod_{\mu \neq \lambda_A} |\mu - \lambda_A|} = \frac{\|x\|_2 \|y\|_2}{|y^* x|},$$

where $\sigma_j(A - \lambda_A I), j = 1 : n - 1$, are nonzero singular values of $A - \lambda_A I$ and μ varies over all eigenvalues of A.

Proof. The derivative $\mathrm{D}\lambda(A)$ and the gradient $\nabla\lambda(A)$ follow from Theorem 3.3 and the equality (3.2). As for the condition number, we only need to prove the results for a subordinate matrix norm and the spectral norm. For a subordinate matrix norm, we have $\|A\| = \max\{\|Ax\| : \|x\| = 1\}$. Hence $\|y x^*\|_* = \|x\| \|y\|_*$, where $\|\cdot\|_*$ is the dual norm of the norm $\|\cdot\|$ on $\mathbb{C}^n$. Consequently, for a subordinate matrix norm, the desired results follow from Theorem 3.3.

For the spectral norm, we have $\|y x^*\|_2 = \|x\|_2 \|y\|_2$. Further, by ([10], Theorem 4) we have $\|\mathrm{adj}(A - \lambda_A I)\|_2 = \prod_{j=1}^{n-1} \sigma_j(A - \lambda_A I)$. Also we have $p'(\lambda_A) = \prod_{\mu \neq \lambda_A}(\lambda_A - \mu)$ for all eigenvalues μ of A. Finally, note that $\|\mathrm{adj}(A - \lambda_A I)\|_2 = \|\mathrm{adj}(A - \lambda_A I)\|_F$. Hence the desired results follow from Theorem 3.3. ☐

Theorem 3.5 provides a first order perturbation bound for a simple eigenvalue of A. Indeed, a first order bound as well as a fast perturbation for a simple eigenvalue are as follows.

COROLLARY 3.6. *Let (λ_A, y, x) be a simple eigentriple of A. Then there is an open set $\Omega \subset \mathbb{C}^{n \times n}$ containing A and a smooth function $\lambda : \Omega \to \mathbb{C}$ such that $\lambda(A) = \lambda_A$ and $\lambda(X)$ is a simple eigenvalue of X for all $X \in \Omega$. Further, we have*

$$\lambda(A + \Delta A) = \lambda(A) + \langle \Delta A, \nabla\lambda(A)\rangle + \mathcal{O}(\|\Delta A\|^2)$$
$$= \lambda(A) + \langle \Delta A, y x^*/x^* y\rangle + \mathcal{O}(\|\Delta A\|^2)$$

and the first order bound $|\lambda(A + \Delta A) - \lambda(A)| \lesssim \mathrm{cond}(\lambda, A)\|\Delta A\|$ for small $\|\Delta A\|$.

If $Z \in \partial\|yx^*\|_*$ then $X := Z/\mathrm{sign}(y^*x)$ is a fast perturbation for λ_A, that is, $\langle X, \nabla\lambda(A)\rangle = \|\nabla\lambda(A)\|_*$ and the first order bound

$$|\lambda(A+tX) - \lambda(A)| = \mathrm{cond}(\lambda, A)\|tX\|$$

holds for sufficiently small $|t|$, where $\mathrm{sign}(z) := \bar{z}/|z|$ if $z \neq 0$ and $\mathrm{sign}(0) = 1$. In particular, if $\|yx^*\|_* = \|x\|\|y\|_*$ and, $u \in \partial\|y\|_*$ and $v \in \partial\|x\|$ then $uv^* \in \partial\|yx^*\|_*$ and hence $X := uv^*/\mathrm{sign}(y^*x)$ is a first perturbation for λ_A.

Let $\mathbb{V}_0(F)$ denote the set of simple points of $\mathbb{V}(F)$, that is, $(X, s) \in \mathbb{V}_0(F)$ if s is a simple eigenvalue of X. Then the map $\mathcal{C} : \mathbb{V}_0(F) \to (\mathbb{C}^{n\times n}, \|\cdot\|)^*, (A, \lambda) \mapsto \mathrm{D}\lambda(A)$ defines the condition map for eigenvalue problems such that $\mathrm{cond}(\lambda, A) = \|\mathcal{C}(A, \lambda)\|$, where $(\mathbb{C}^{n\times n}, \|\cdot\|)^*$ is the dual space of $(\mathbb{C}^{n\times n}, \|\cdot\|)$. On the other hand, the map $\mathcal{K} : \mathbb{V}_0(F) \to (\mathbb{C}^{n\times n}, \|\cdot\|_*), (A, \lambda) \mapsto \nabla\lambda(A)$ is the matrix representation of the condition map $\mathcal{C}(\lambda, A)$ such that $\mathrm{cond}(\lambda, A) = \|\mathcal{K}(A, \lambda)\|_*$.

4. Holomorphic perturbation. Let $A : \mathbb{C}^m \to \mathbb{C}^{n\times n}$ be holomorphic. The analysis of the eigenvalues of $A(t)$ when t varies in $\mathbb{C}^m$ is a classical subject and has been studied extensively, see [4, 8, 9, 5, 11] and the references therein, also see [3]. We mention that the holomorphic evolution of a simple eigenvalue $\lambda(t)$ of $A(t)$ also follows from the results in Section 3. For completeness, we provide a short exposition of the holomorphic evolutions of simple eigenvalues of $A(t)$. Our derivation is slightly different from those in [11, 5].

THEOREM 4.1. *Let* $A : \mathbb{C}^m \to \mathbb{C}^{n\times n}$ *be holomorphic. Let* λ_0 *be a simple eigenvalue of* $A(t_0)$ *and,* y_0 *and* x_0 *be left and right eigenvectors of* $A(t_0)$ *corresponding to* λ_0*. Then there is an open set* $\Omega \subset \mathbb{C}^m$ *containing* t_0 *and a holomorphic function* $\lambda : \Omega \to \mathbb{C}$ *such that* $\lambda(t_0) = \lambda_0$ *and* $\lambda(t)$ *is a simple eigenvalue of* $A(t)$ *for all* $t \in \Omega$*. Further, for* $t \in \Omega$ *we have*

$$\lambda(t+h) = \lambda(t) + \mathrm{D}\lambda(t)h + \mathcal{O}(\|h\|_2^2) = \lambda(t) + \langle h, \nabla\lambda(t)\rangle + \mathcal{O}(\|h\|_2^2)$$

for sufficiently small $\|h\|_2$*. The derivative* $\mathrm{D}\lambda(t)$ *and the gradient* $\nabla\lambda(t)$ *are given by*

$$\mathrm{D}\lambda(t)h = \frac{\sum_{j=1}^m \mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t))\partial_{t_j}A(t))h_j}{\partial_z p(t, \lambda(t))} = \frac{\sum_{j=1}^m \mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t))\partial_{t_j}A(t))h_j}{\mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t)))}$$

$$\nabla\lambda(t) = \left(\frac{[\mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t))\partial_{t_1}A(t)), \ldots, \mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t))\partial_{t_m}A(t))]}{\mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t)))}\right)^*$$

for all $h \in \mathbb{C}^m$*, where* $\partial_{t_j}A(t)$ *is the partial derivative of* $A(t)$ *with respect to* t_j *and* $\partial_z p(t, \lambda(t))$ *is the partial derivative of* $p(t, z)$ *with respect to* z *evaluated at* $(t, \lambda(t))$*. In particular, we have*

$$\frac{\partial\lambda(t_0)}{\partial t_j} = \frac{\mathrm{Tr}\left(\mathrm{adj}(\lambda_0 I - A(t_0))\frac{\partial A(t_0)}{\partial t_j}\right)}{\partial_z p(t_0, \lambda_0)} = \frac{\mathrm{Tr}\left(\mathrm{adj}(\lambda_0 I - A(t_0))\frac{\partial A(t_0)}{\partial t_j}\right)}{\mathrm{Tr}(\mathrm{adj}(\lambda_0 I - A(t_0)))} = \frac{y_0^*\frac{\partial A(t_0)}{\partial t_j}x_0}{y_0^*x_0}.$$

Proof. Consider $p(t, z) := \det(zI - A(t))$ for $(t, z) \in \mathbb{C}^m \times \mathbb{C}$ and the analytic variety $\mathbb{V}(p) := \{(t, z) \in \mathbb{C}^m \times \mathbb{C} : p(t, z) = 0\}$. Then $(t_0, \lambda_0) \in \mathbb{V}(p)$ and $\partial_z p(t_0, \lambda_0) \neq$

0. Hence by the Implicit function theorem and following similar arguments as those in the proof of Theorem 3.3, we obtain a holomorphic function $\lambda : \Omega \to \mathbb{C}$ such that $\lambda(t_0) = \lambda_0$ and $\lambda(t)$ is a simple eigenvalue of $A(t)$ for $t \in \Omega$. Again by Theorem 3.3 and the chain rule, we have $\mathrm{D}\lambda(t)h = \dfrac{\mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t))\mathrm{D}A(t)h)}{\partial_z p(t, \lambda(t))}$ for all $h \in \mathbb{C}^m$, where $\mathrm{D}A(t)$ is the derivative of $A(t)$. The derivative $\mathrm{D}A(t) : \mathbb{C}^m \to \mathbb{C}^{n \times n}$ is a linear map and is given by $\mathrm{D}A(t)h = \sum_{j=1}^m \partial_{t_j} A(t)h_j$ for all $h \in \mathbb{C}^m$. Also $\partial_z p(t, \lambda(t)) = \mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t)))$. Hence for $h \in \mathbb{C}^m$, we have

$$\mathrm{D}\lambda(t)h = \frac{\sum_{j=1}^m \mathrm{Tr}\left(\mathrm{adj}(\lambda(t)I - A(t))\partial_{t_j} A(t)\right) h_j}{\mathrm{Tr}(\mathrm{adj}(\lambda(t)I - A(t)))} = \langle h, \nabla\lambda(t)\rangle$$

which yields the desired results $\square$

Notice that the results in Theorem 4.1 can be easily extended to parameter dependent nonlinear eigenvalue problems [3]. Let $A : \mathbb{C}^m \times \mathbb{C} \longrightarrow \mathbb{C}^{n \times n}, (t, z) \longmapsto A(t, z)$ be holomorphic and regular, that is, for each $t \in \mathbb{C}^m$ there is a $\lambda \in \mathbb{C}$ such that $\det(A(t, \lambda)) \neq 0$. If $\lambda(t)$ is a simple eigenvalue of $A(t, z)$, that is, $\lambda(t)$ is a simple zero of $p(t, z) := \det(A(t, z)) = 0$, then the analyticity of $\lambda(t)$ in a neighbourhood of t and the derivative $\mathrm{D}\lambda(t)$ can be deduced by applying the implicit function theorem to $p(t, z) = 0$ at $(t, \lambda(t))$ followed by the Jacobi formula for the derivative of determinant.

5. Sensitivity and linearization. Consider a monic scalar polynomial $p(z) := z^n + a_{n-1}z^{n-1} + \cdots + a_0$. Then it is well known that the roots of $p(z)$ are the eigenvalues of the companion matrix

$$C_p := \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Indeed, $\det(zI - C_p) = p(z)$ and hence $p(z)$ is the characteristic polynomial of C_p. The matrices of the form C_p are also known as Frobenius companion matrices of polynomials. The Frobenius companion matrices are the basic blocks in the rational canonical forms of matrices. The MATLAB command `roots` uses Frobenius companion matrices for computing roots of scalar polynomials. So, suppose that the roots of $p(z)$ are computed by solving the eigenvalue problem $C_p v = \lambda v$. This raises a natural question: *How is the sensitivity of λ as a root of $p(z)$ related to the sensitivity of λ as an eigenvalue of C_p?* This is a pertinent question because the sensitivity of eigenvalues influence the accuracy of the computed eigenvalues. Thus, with a view to providing an answer to the question, we compare the condition numbers $\mathrm{cond}(\lambda, p)$ and $\mathrm{cond}(\lambda, C_p)$.

PROPOSITION 5.1. *Let λ be a simple root of $p(z)$. Then λ is a simple eigenvalue of C_p and $v := [\lambda^{n-1}, \ldots, \lambda, 1]^\top$ is an eigenvector of C_p corresponding to λ. Further,*

$$u := [1, \lambda + a_{n-1}, \ldots, \lambda^{n-1} + a_{n-1}\lambda^{n-2} + \cdots + a_2\lambda + a_1]^*$$

ELA

is a left eigenvector of C_p corresponding to λ. Furthermore, $u^*v = p'(\lambda)$ and that $\mathrm{adj}(\lambda I - C_p) = vu^*$, *where $p'(\lambda)$ is the derivative of $p(z)$ at λ. Thus we have*

$$\mathrm{cond}(\lambda, C_p) = \frac{\|(\mathrm{adj}(\lambda I - C_p))^*\|_*}{|p'(\lambda)|} = \frac{\|uv^*\|_*}{|p'(\lambda)|}.$$

Proof. It is easy to see that $C_p v = \lambda v$ and $u^* C_p = \lambda u^*$. A simple calculation shows that $u^*v = p'(\lambda)$. Hence by Theorem 3.4 we have $\mathrm{adj}(\lambda I - C_p) = vu^*$ and the desired expression for $\mathrm{cond}(\lambda, C_p)$ follows from Theorem 3.5. This completes the proof. $\square$

Set $\Lambda_n := [1, \lambda, \ldots, \lambda^n]^\top$. Then recall from Theorem 2.2 that the condition number of λ as a root of $p(z)$ is given by $\mathrm{cond}(\lambda, p) = \|\bar{\Lambda}_n\|_* / |p'(\lambda)|$. We now compare the sensitivity of λ as a root of $p(x)$ with the sensitivity of λ as an eigenvalue of C_p.

THEOREM 5.2. *Let $p(x) := \det(xI - C_p)$ and λ be a simple eigenvalue of C_p. Let u and v be as in Proposition 5.1 and $\Lambda_n := [1, \lambda, \ldots, \lambda_n]^\top$. Then, for the spectral and Frobenius norms, we have*

$$\frac{\|u\|_2}{\sqrt{1 + |\lambda|^2}} \leq \frac{\mathrm{cond}(\lambda, C_p)}{\mathrm{cond}(\lambda, p)} \leq \frac{\sqrt{2}\,\|u\|_2}{\sqrt{1 + |\lambda|^2}}.$$

Proof. By Theorem 2.2 and Proposition 5.1, for the spectral and Frobenius norms, we have $\mathrm{cond}(\lambda, C_p)/\mathrm{cond}(\lambda, p) = \|v\|_2 \|u\|_2 / \|\Lambda_n\|_2$. Now a straightforward calculation shows that

$$1 \leq \frac{\sqrt{1 + |\lambda|^2}\,\|v\|_2}{\|\Lambda_n\|_2} \leq \sqrt{2}$$

which yields the desired bounds. $\square$

Note that $\|u\|_2 \geq 1$ and $\|u\|_2 \to \infty$ as $|\lambda| \to \infty$. Thus, a root λ of $p(x)$ with a large absolute value is more ill-conditioned as an eigenvalue of C_p than as a root of $p(x)$. This is to be expected as $\mathrm{cond}(\lambda, C_p)$ measures the sensitivity of λ to an arbitrary small perturbation ΔA to C_p and the perturbed matrix $C_p + \Delta A$ almost always will not be a companion matrix. Since the perturbed matrix $C_p + \Delta A$ cannot be expected to be a companion matrix, restricting the perturbation ΔA such that $C_p + \Delta A$ is a companion matrix we obtain structured sensitivity analysis of λ. Observe that if $b = [b_{n-1}, \ldots, b_0]^\top \in \mathbb{C}^n$ then $\Delta A := -e_1 b^\top$ is a structure preserving perturbation of C_p, that is, $C_p + \Delta A$ is a companion matrix of the polynomial $(p + \Delta p)(x) = x^n + (a_{n-1} + b_{n-1})x^{n-1} + \cdots + (a_0 + b_0)$, where $e_1 = [1, 0, \ldots, 0]^\top \in \mathbb{C}^n$. The sensitivity of an eigenvalue λ of C_p relative to perturbations of the form $e_1 b^\top$ is measured by the *structured condition number* of λ which we denote by $\mathrm{cond}^{\mathbf{S}}(\lambda, C_p)$. Then by Theorem 3.5 we have

$$\mathrm{cond}^{\mathbf{S}}(\lambda, C_p) = \sup\{|\langle e_1 b^\top, uv^*/v^*u\rangle| : b \in \mathbb{C}^n \text{ and } \|e_1 b^*\| = 1\}. \qquad (5.1)$$

Also recall from Theorem 2.2 that $\mathrm{cond}^{\mathbf{S}}(\lambda, p)$ measures the sensitivity of λ relative to small perturbations $\Delta p(x) \in \mathcal{P}_{n-1}$ to $p(x)$. In such a case, the perturbed polynomial $p(x) + \Delta p(x)$ is again a monic polynomial.

PROPOSITION 5.3. *Let $p(x) := \det(xI - C_p)$ and λ be a simple eigenvalue of C_p. Let u and v be as in Proposition 5.1. Then, for the spectral and Frobenius norms, we*

$$\boxed{\textbf{ELA}}$$

236

have

$$\text{cond}^{\mathbf{S}}(\lambda, C_p) = \text{cond}^{\mathbf{S}}(\lambda, p) = \frac{\|v\|_2}{|p'(\lambda)|} = \frac{\|v\|_2}{|u^*v|}.$$

Proof. By Theorem 2.2 we have $\text{cond}^{\mathbf{S}}(\lambda, p) = \|v\|_2/|p'(\lambda)|$. Since

$$|\langle e_1 b^*,\, uv^*/v^*u\rangle| = |b^*v|/|u^*v|$$

and $p'(\lambda) = u^*v$, taking supremum over $\|b\|_2 = 1$ for $b \in \mathbb{C}^n$, by (5.1) we have $\text{cond}^{\mathbf{S}}(\lambda, C_p) = \frac{\|v\|_2}{|p'(\lambda)|} = \frac{\|v\|_2}{|u^*v|} = \text{cond}^{\mathbf{S}}(\lambda, p)$. $\square$

We mention that the results presented in this paper can be generalized, with appropriate modifications, to nonlinear eigenvalue problems, see [2].

REFERENCES

[1] S. AHMAD, R. ALAM AND R. BYERS, *Pseudospectra, critical points and multiple eigenvalues of matrix pencils*, SIAM J. Matrix Anal Appl., 31(2010), pp.1915-1933.

[2] S. AHMAD AND R. ALAM, *Sensivity analysis of nonlinear eigenproblems*, preprint.

[3] A. L. ANDREW, K.-W. E. CHU AND P. LANCASTER,*Derivatives of eigenvalues and eigenvectors of matrix functions,* SIAM J. Matrix Anal Appl., 14(1993), pp.903-926.

[4] H. BAUMGARTEL, *Analytic Perturbation Theory for Matrices and Operators*, Birkhauser Verlag, 1985.

[5] K.-W. E. CHU, *On multiple eigenvalues of matrices depending on several parameters,* SIAM J. Numer. Anal., 27(1990), pp. 1368-1385.

[6] J. W. DEMMEL,*Computing stable eigendecompositions of matrices,* Linear Algebra & Appl., 79(1986), pp.163-193.

[7] J. W. DEMMEL,*On condition numbers and the distance to the nearest ill-posed problem,* Numer. Math., 51(1987), pp.251-289.

[8] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlag, 1980.

[9] K. R. PARTHASARATHY,*Eigenvalues of matrix-valued analytic maps,*J. Austral. Math. Soc. (Series A), 26 (1978), pp.179-197.

[10] R. A. SMITH, *The condition numbers of the matrix eigenvalue problem,* Numer. Math., 10(1967), pp.232-240.

[11] J.G. SUN, *Eigenvalues and eigenvectors of a matrix dependent on several parameters,* J. Comput. Math., 3(1985), pp. 351-364.

[12] J.G. SUN, *Multiple eigenvalue sensitivity anaylysis,* Linear Algebra Appl., 37/138 (1990), pp. 183-211.

[13] J.G. SUN, *On condition numbers of a nondefective multiple eigenvalue,* Numer. Math., 61(1992), pp.265-275.

[14] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Oxford University press, 1965.

[15] J. H. WILKINSON, *Notes on matrices with a very ill-conditioned eigenproblem,* Numer. Math., 19(1972), pp.175-178.

[16] J. H. WILKINSON, *On neighbouring matrices with quadratic elementary divisors,* Numer. Math., 44(1984), pp.1-21.

[17] J. H. WILKINSON, *Sensitivity of eigenvalues,* Utilitas Mathematica, 25(1984), pp.5-76.

[18] G. A. WATSON, *Characterization of the Subdifferential of Some Matrix Norms,* Linear Algebra Appl., 170(1992), pp.33-44.

[19] K. ZIETAK, *Subdifferentials, Faces, and Dual Matrices,* Linear Algebra Appl., 185(1993), pp.125-141.