Journal of Computational Mathematics, Vol.10, No.3, 1992, 193–203.

ON THE SENSITIVITY OF SEMISIMPLE MULTIPLE EIGENVALUES*1)

Sun Ji-guang
(Computing Center, Academia Sinica, Beijing, China)

Abstract

This paper is a supplement to the work in [6] where we investigated the directional derivatives of semisimple multiple eigenvalues of a complex $n \times n$ matrix analytically dependent on several parameters. The result of this paper can be used to define the sensitivity of semisimple multiple eigenvalues in a more resonable way than in [6].

§1. Main Results

This is a supplement to the work in [6]. We shall use the notation described in [5] and [6].

Let $p = (p_1, \cdots, p_N)^T \in \mathbb{C}^N$. Suppose that $A(p) \in \mathbb{C}^{n\times n}$ is an analytic function in a connected open set $\mathcal{B} \in \mathbb{C}^N$. In this paper we consider the eigenproblem

$$A(p)x(p) = \lambda(p)x(p), \quad \lambda(p) \in \mathbb{C}, \quad x(p) \in \mathbb{C}^n, \quad p \in \mathcal{B}. \tag{1.1}$$

Let $\mu(p)$ be a function defined in $\mathcal{B}$. The directional derivative of $\mu(p)$ at $p^* \in \mathcal{B}$ in the direction ν, denoted by $D_\nu \mu(p^*)$, is defined as follows:

$$D_\nu \mu(p^*) \equiv \lim_{\tau \to +0} \frac{\mu(p^* + \tau\nu) - \mu(p^*)}{\tau}, \tag{1.2}$$

where $\nu \in \mathbb{C}^N$ with $\|\nu\|_2 = 1$ and τ is a positive parameter.

Without loss of generality, we may investigate the directional derivatives of the eigenvalues of $A(p)$ at the origin of $\mathbb{C}^N$. The following two theorems are the main results of this paper.

Theorem 1.1. *Let $A(p) \in \mathbb{C}^{n\times n}$ be an analytic function of $p = (p_1, p_2, \cdots, p_N)^T$ in some neighbourhood $\mathcal{B}(0)$ of the origin of $\mathbb{C}^N$, and let $\lambda(A(p)) = \{\lambda_s(p)\}_{s=1}^n$ for $p \in \mathcal{B}(0)$, in which $\lambda_1(0) = \cdots = \lambda_r(0) = \lambda_1$. Suppose that there are matrices $X, Y \in \mathbb{C}^{n\times n}$ satisfying*

$$X = (X_1, X_2), \quad Y = (Y_1, Y_2), \quad X_1, Y_1 \in \mathbb{C}^{n\times r}, \quad X^H Y = I \tag{1.3}$$

* Received September 14, 1988.
1) The project supported by National Natural Science Foundation of China.

and

$$Y^H A(0) X = \begin{pmatrix} \lambda_1 I^{(r)} & 0 \\ 0 & A_2 \end{pmatrix}, \quad \lambda_1 \bar{\in} \lambda(A_2). \tag{1.4}$$

Then, for any fixed direction $\nu \in \mathbb{C}^N$ with $\|\nu\|_2 = 1$, there are a positive scalar β and r single-valued continuous functions $\mu_1(\tau\nu), \cdots, \mu_r(\tau\nu)$ for $\tau \in [-\beta, \beta]$ such that $\{\mu_s(\tau\nu)\}_{s=1}^r$ are r of the eigenvalues of $A(\tau\nu)$, the set $\{\mu_s(\tau\nu)\}_{s=1}^r$ and the set $\{\lambda_s(\tau\nu)\}_{s=1}^r$ are identical, and there is a one-to-one correspondence between the elements of the two sets. Moreover, we have

$$\{D_\nu \mu_s(0)\}_{s=1}^r = \lambda\Big(\sum_{j=1}^N \nu_j Y_1^H \Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0} X_1\Big). \tag{1.5}$$

Theorem 1.2. *Let $A(p), B(0), \lambda_1(p), \cdots, \lambda_r(p), \lambda_1, X, Y$ and ν be described as in Theorem 1.1. Define*

$$s_p^{(\nu)}(\lambda_1) \equiv \lim_{\tau\to 0} \max_{\substack{z\in\mathbb{C} \\ |z|=\tau>0}} \frac{\max_{1\le j\le r} |\lambda_j(z\nu) - \lambda_1|}{|z|}. \tag{1.6}$$

Then

$$s_p^{(\nu)}(\lambda_1) = \rho\Big(\sum_{j=1}^N \nu_j Y_1^H \Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0} X_1\Big), \tag{1.7}$$

where $\rho(\cdot)$ denotes the spectral radius of a matrix.

Remark 1.3. The difficulty in investigating the local behaviours of a semisimple multiple eigenvalue of multiplicity $r > 1$ lies in that the r eigenvalues, as functions of some parameters, may have singularity at the intersection point (ref [2, p.74–76]). Even if in the case of one complex parameter z, the r eigenvalues are in general not continuous in any neighbourhood of the singular point (ref. [2, p.125]). Fortunately, Kato [2, p.125–127] proved that, if the simple parameter z changes over an interval $[\alpha, \beta]$ of the real line, then there exist r single-valued continuous functions, the values of which constitute the set of r eigenvalues for each $z \in [\alpha, \beta]$. Therefore we may take the r single-valued continuous functions as the r eigenvalues for $z \in [\alpha, \beta]$. This fact is just one of the keys to investigating the directional derivatives of semisimple multiple eigenvalues in this paper.

Remark 1.4. M. Overton and R. Womersley [4] discussed directional derivatives of semisimple multiple eigenvalues of the matrix

$$A_0 + \sum_{k=1}^m \xi_k A_k,$$

where $\{A_k\}$ are given real $n \times n$ matrices, and $\{\xi_k\}$ are real parameters.

We shall give the proofs of Theorem 1.1 and Theorem 1.2 in §2 and give some applications in §3.

§2. Proof of Theorems

Proof of Theorem 1.1. The theorem is proved by the following step 2.1 – step 2.4.

2.1. [6] has proved that there exist analytic functions

$$X_1(p) \in \mathbb{C}_r^{n\times r}, \quad Y_1(p)^T \in \mathbb{C}_r^{r\times n}$$

and

$$A_1(p) = (Y_1(p)^T X_1(p))^{-1} Y_1(p)^T A(p) X_1(p),$$
$$\tilde{A}_1(p) = Y_1(p)^T A(p) X_1(p) (Y_1(p)^T X_1(p))^{-1} \tag{2.1}$$

in some neighbourhood $\mathcal{B}_0 \subset \mathcal{B}(0)$ of the origin of $\mathbb{C}^N$, such that

$$A(p)X_1(p) = X_1(p)A_1(p), \quad Y_1(p)^T A(p) = \tilde{A}_1(p) Y_1(p)^T \tag{2.2}$$

and

$$A_1(0) = \tilde{A}_1(0) = \lambda_1 I^{(r)}, \quad X_1(0) = X_1, \quad Y_1(0) = \overline{Y}_1. \tag{2.3}$$

Here $\overline{Y}_1$ denotes the complex conjugate matrix of Y_1. Moreover , by the hypothesis and the relation (2.2),

$$\lambda(A_1(p)) = \{\lambda_s(p)\}_{s=1}^r \subseteq \lambda(A(p)). \tag{2.4}$$

Let $\nu \in \mathbb{C}^N$ be any fixed direction with $\|\nu\|_2 = 1$. Take $p = \tau\nu$, where $\tau \in [-\beta, \beta]$, and β is a small positive scalar so that

$$\tau\nu \in \mathcal{B}_0 \quad \text{for} \quad \tau \in [-\beta, \beta].$$

By [2, p.125–127], we can number the eigenvalues of $A_1(\tau\nu)$ as $\mu_1(\tau\nu), \cdots, \mu_r(\tau\nu)$ such that the functions $\mu_s(\tau\nu)$ are single-valued and continuous on $[-\beta, \beta]$. The set $\{\mu_s(\tau\nu)\}_{s=1}^r$ and the set $\{\lambda_s(\tau\nu)\}_{s=1}^r$ are just the same, and there is one-to-one correspondence between the elements of the two sets. Moreover, from (2.3)

$$\mu_s(0) = \lambda_1, \quad s = 1, \cdots, r.$$

For simplicity, we write

$$\hat{\mu}_s(\tau) = \mu_s(\tau\nu), \quad s = 1, \cdots, r. \tag{2.5}$$

2.2. Let z be a complex variable. Now we consider the function $\hat{A}_1(z) \equiv A_1(z\nu)$ and its eigenvalues, where ν is the fixed direction taken in 2.1, and

$$z \in \mathcal{D}_0 \equiv \{z = \tau + i\sigma \in \mathbb{C} \;:\; \tau \in [-\beta_1, \beta_1], \sigma \in [-\gamma_1, \gamma_1]\},$$

in which $\beta_1 (< \beta)$ and γ_1 are small positive scalars so that

$$z\nu \in \mathcal{B}_0 \quad \text{for} \quad z \in \mathcal{D}_0.$$

Obviously, $\hat{A}_1(z)$ is an analytic function of $z \in \mathcal{D}_0$, and by (2.3) λ_1 is a semisimple multiple eigenvalue of $\hat{A}_1(0)$. Therefore, if we let $\lambda(\hat{A}_1(z)) = \{\hat{\lambda}_t(z)\}_{t=1}^r$, then the eigenvalues $\hat{\lambda}_t(z)$ may be expressed as convergent Piuseux series[3]

$$\hat{\lambda}_t(z) = \lambda_1 + \varphi_t^{(1)} z + \sum_{m>r'} \varphi_t^{(m)} z^{m/r'}, \quad t = 1, \cdots, r,$$

where the natural number $r' \le r$. Specially, take $z = \tau \in (0, \beta_1]$; the eigenvalues $\hat{\lambda}_t(\tau)$ of the matrix $\hat{A}_1(\tau)$ are

$$\hat{\lambda}_t(\tau) = \lambda_1 + \varphi_t^{(1)}\tau + \sum_{m>r'} \varphi_t^{(m)}\tau^{m/r'}, \quad t = 1, \cdots, r. \tag{2.6}$$

2.3. Let $\tau = \tau_1^{r'}$. Further, let

$$\tilde{\lambda}_t(\tau_1) \equiv \hat{\lambda}_t(\tau_1^{r'}), \quad t = 1, \cdots, r \tag{2.7}$$

and

$$\tilde{\mu}_s(\tau_1) \equiv \hat{\mu}_s(\tau_1^{r'}), \quad s = 1, \cdots, r. \tag{2.8}$$

Observe the following facts:

(i) By (2.6) and (2.7) the functions $\tilde{\lambda}_t(\tau_1)$ are analytic for $\tau_1 \in (0, \beta_1^{1/r'}]$. Since the zeros of any analytic function of one real variable are isolated[1], we have

$$\tilde{\lambda}_j(\tau_1) \ne \tilde{\lambda}_k(\tau_1), \quad \forall\, \tau_1 \in (0, \beta_1^{1/r'}], \quad j \ne k$$

provided that $\tilde{\lambda}_j(\tau_1) \ne \tilde{\lambda}_k(\tau_1)$ for $\tau_1 \in (0, \beta^{1/r'})$ and the positive scalar β_1 is sufficiently small.

(ii) The functions $\tilde{\mu}_1(\tau_1), \cdots, \tilde{\mu}_r(\tau_1)$ are continuous on $[0, \beta_1^{1/r'}]$.

(iii) The set $\{\tilde{\mu}_s(\tau_1)\}_{s=1}^r$ and $\{\tilde{\lambda}_t(\tau_1)\}_{t=1}^r$ are just the same for each point $\tau_1 \in [0, \beta_1^{1/r'}]$, and there is a one-to-one correspondence between the elements of the two sets.

Therefore there is a permutation π_1 of $\{1, \cdots, r\}$ dependent on ν such that

$$\tilde{\mu}_s(\tau_1) = \tilde{\lambda}_{\pi_1(s)}(\tau_1), \quad s = 1, \cdots, r, \quad \tau_1 \in [0, \beta_1^{1/r'}],$$

and from (2.7), (2.8),

$$\hat{\mu}_s(\tau) = \hat{\lambda}_{\pi_1(s)}(\tau), \quad s = 1, \cdots, r, \quad \tau \in [0, \beta_1]. \tag{2.9}$$

Hence, from (1.2), (2.5), (2.9) and (2.6) we get

$$\begin{aligned} D_\nu \mu_s(0) &= \lim_{\tau \to +0} \frac{\mu_s(\tau\nu) - \mu_s(0)}{\tau} = \lim_{\tau \to +0} \frac{\hat{\mu}_s(\tau) - \hat{\mu}_s(0)}{\tau} \\ &= \lim_{\tau \to +0} \frac{\hat{\lambda}_{\pi_1(s)}(\tau) - \hat{\lambda}_{\pi_1(s)}(0)}{\tau} = \left(\frac{d\hat{\lambda}_{\pi_1(s)}(\tau)}{d\tau}\right)_{\tau=0} = \varphi_{\pi_1(s)}^{(1)}, \quad s = 1, \cdots, r. \end{aligned} \tag{2.10}$$

2.4. Now we seek the explicit expressions of $\varphi_t^{(1)}, t = 1, \cdots, r$.

Let ε be an arbitrary positive number. By the Jordan canonical form theorem there is a nonsingular matrix $Q \in \mathbb{C}^{r\times r}$ such that

$$Q^{-1}\left(\frac{dA_1(\tau\nu)}{d\tau}\right)_{\tau=0} Q = \begin{pmatrix} \delta_1 & \varepsilon_1 & & \\ & \ddots & \ddots & \\ & & \ddots & \varepsilon_{r-1} \\ & & & \delta_r \end{pmatrix}, \tag{2.11}$$

where

$$\left.\begin{aligned} &\delta_1 = \cdots = \delta_{r_1} < \delta_{r_1+1} = \cdots = \delta_{r_1+r_2} \\ &< \cdots < \delta_{r_1+\cdots+r_{q-1}+1} = \cdots = \delta_{r_1+\cdots+r_{q-1}+r_q}, \\ &r_1 + \cdots + r_q = r, \end{aligned}\right\} \tag{2.12}$$

and each ε_l is equal to $\frac{\varepsilon}{2}$ or zero. Let

$$Q^{-1}A_1(\tau\nu)Q = (\theta_{kl}(\tau))_{1\le k,\ l\le r}\ . \tag{2.13}$$

Then by the same argument as used in [6, p.136, 3°], from (2.11) and

$$Q^{-1}\Big(\frac{dA_1(\tau\nu)}{d\tau}\Big)_{\tau=0}\quad Q = \Big[\frac{d}{d\tau}(Q^{-1}A_1(\tau\nu)Q)\Big]_{\tau=0}$$

we have

$$\theta_{k1} = \begin{cases} \lambda_1 + \delta_k + \theta_{kk}^{(2)}\tau^2 + \theta_{kk}^{(3)}\tau^3 + \cdots & \text{if } k = l, \\ \frac{\varepsilon}{2} + \theta_{k,k+1}^{(2)}\tau^2 + \theta_{k,k+1}^{(3)}\tau^3 + \cdots & \\ \quad \text{or } \theta_{k,k+1}^{(2)}\tau^2 + \theta_{k,k+1}^{(3)}\tau^3 + \cdots & \text{if } k = l-1, \\ \theta_{kl}^{(2)}\tau^2 + \theta_{kl}^{(3)}\tau^3 + \cdots & \text{otherwise.} \end{cases} \tag{2.14}$$

By the Gerschgorin theorem, from (2.11)–(2.14) we can prove that (ref. [6. p.137])

$$\varphi_t^{(1)} = \delta_{\pi_2(t)}, \quad t = 1, \cdots, r, \tag{2.15}$$

where π_2 is a permutation of the set $\{1, \cdots, r\}$.

From (2.11) it follows that

$$\{\delta_{\pi_2(t)}\}_{t=1}^r = \lambda\Big(Q^{-1}\Big(\frac{dA_1(\tau\nu)}{d\tau}\Big)_{\tau=0}Q\Big) = \lambda\Big(\Big(\frac{dA_1(\tau\nu)}{d\tau}\Big)_{\tau=0}\Big). \tag{2.16}$$

Moreover, from (2.1)–(2.3) we get (ref. [6, p.136])

$$\Big(\frac{dA_1(\tau\nu)}{d\tau}\Big)_{\tau=0} = Y_1^H\Big(\frac{dA(\tau\nu)}{d\tau}\Big)_{\tau=0}X_1 = Y_1^H\Big(\sum_{j=1}^N \nu_j\Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0}\Big)X_1$$

$$= \sum_{j=1}^N \nu_j Y_1^H\Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0}X_1.$$

Substituting it into (2.16), and combining with (2.10) and (2.15), we get (1.5).

This completes the proof.

Proof of Theorem 1.2. By the step 2.1 of the proof of Theorem 1.1, $A_1(p)$ (see (2.1)) is an analytic matrix-valued function of p in some neighbourhood $\mathcal{B}_0$ of the origin of $\mathbb{C}^N$. Therefore, $A_1(p)$ can be expanded as a convergent power series

$$A_1(p) = \lambda_1 I^{(r)} + \sum_{j=1}^N\Big(\frac{\partial A_1(p)}{\partial p_j}\Big)_{p=0}p_j + \frac{1}{2}\sum_{j,k=1}^N\Big(\frac{\partial^2 A_1(p)}{\partial p_j\partial p_k}\Big)_{p=0}p_jp_k + \cdots,\ p\in\mathcal{B}_0. \tag{2.17}$$

Now we take $p = z\nu$ for any fixed $\nu\in\mathbb{C}^N$ with $\|\nu\|_2 = 1$, in which $z\in\mathbb{C}, z\ne 0$, and $|z|$ is so small that $z\nu\in\mathcal{B}_0$. Then from (2.17),

$$A_1(z\nu) = \lambda_1 I^{(r)} + z\sum_{j=1}^N\Big(\frac{\partial A_1(p)}{\partial p_j}\Big)_{p=0}\nu_j + \frac{z^2}{2}\sum_{j,k=1}^N\Big(\frac{\partial^2 A_1(p)}{\partial p_j\partial p_k}\Big)_{p=0}\nu_j\nu_k + \cdots,$$

$$z\nu\in\mathcal{B}_0. \tag{2.18}$$

Let $\varepsilon > 0$. It is well known that there exists a nonsingular matrix $Q(\varepsilon)$ such that

$$Q(\varepsilon)^{-1}\Big(\sum_{j=1}^{N}\Big(\frac{\partial A_1(p)}{\partial p_j}\Big)_{p=0}\nu_j\Big)Q(\varepsilon)=\begin{pmatrix}\delta_1 & \varepsilon_1 & & \\ & \ddots & \ddots & \\ & & \ddots & \varepsilon_{r-1}\\ & & & \delta_r\end{pmatrix},$$

in which each ε_k is equal to ε or zero, and

$$\{\delta_1,\cdots,\delta_r\}=\lambda\Big(\sum_{j=1}^{N}\Big(\frac{\partial A_1(p)}{\partial p_j}\Big)_{p=0}\nu_j\Big)=\lambda\Big(\sum_{j=1}^{N}\nu_j Y_1^H\Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0}X_1\Big).$$

Substituting it into (2.18) we get

$$Q(\varepsilon)^{-1}A_1(z\nu)Q(\varepsilon)=\lambda_1 I^{(r)}+z\begin{pmatrix}\delta_1 & \varepsilon_1 & & \\ & \ddots & \ddots & \\ & & \ddots & \varepsilon_{r-1}\\ & & & \delta_r\end{pmatrix}+\frac{z^2}{2}Q(\varepsilon)^{-1}R(\nu)Q(\varepsilon),$$

where $R(\nu)=\sum_{j,k=1}^{N}\Big(\frac{\partial^2 A_1(p)}{\partial p_j\partial p_k}\Big)_{p=0}\nu_j\nu_k+O(|z|)(|z|\to 0)$. By the Gerschgorin theorem each eigenvalue $\lambda_j(z\nu)$ of $A_1(z\nu)$ satisfies

$$\frac{|\lambda_j(z\nu)-\lambda_1|}{|z|}\le\max_{1\le j\le r}|\delta_j|+\varepsilon+\frac{|z|}{2}\|Q(\varepsilon)^{-1}R(\nu)Q(\varepsilon)\|_\infty.$$

Let $|z|=\tau\to 0$. Then

$$\lim_{\tau\to 0}\max_{\substack{z\in\mathbb{C}\\ |z|=\tau>0}}\frac{\max_{1\le j\le r}|\lambda_j(z\nu)-\lambda_1|}{|z|}\le\max_{1\le j\le r}|\delta_j|+\varepsilon. \tag{2.19}$$

Thus we have proved that for any $\varepsilon > 0$ the inequality (2.19) holds. Hence by definition (1.6) we have

$$s_p^{(\nu)}(\lambda_1)\le\rho\Big(\sum_{j=1}^{N}\nu_j Y_1^H\Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0}X_1\Big). \tag{2.20}$$

On the other hand, for any fixed $\nu\in\mathbb{C}^N$ with $\|\nu\|_2=1$ and for any $\varepsilon>0$, by Theorem 1.1 there are a single-valued continuous function $\mu_k(\tau\nu)$ and a sufficiently small $\tau>0$ such that

$$\frac{|\mu_k(\tau\nu)-\lambda_1|}{\tau}\ge\rho\Big(\sum_{j=1}^{N}\nu_j Y_1^H\Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0}X_1\Big)-\varepsilon. \tag{2.21}$$

Observe that by Theorem 1.1 $\mu_k(\tau\nu)=\lambda_l(\tau\nu)$ for some $l\in\{1,\cdots,r\}$, hence from (2.21) and (1.6)

$$S_p^{(\nu)}(\lambda_1)\ge\rho\Big(\sum_{j=1}^{N}\nu_j Y_1^H\Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0}X_1\Big). \tag{2.22}$$

Combining (2.22) with (2.20) we get the relation (1.7). The proof is completed.

§3. Applications

First of all we consider a simple example.

Example 3.1.[6] consider the eigenproblem (1.1) with $N = 2, \mathcal{B} = \mathbb{C}^2$, and

$$A(p) = \begin{pmatrix} 1+2p_1+2p_2 & p_2 \\ 2p_1 & 1+4p_2 \end{pmatrix}.$$

The matrix $A(p)$ is an analytic function of $p \in \mathbb{C}^2, A(0)$ has a semisimple multiple eigenvalue $\lambda_1 = 1$, and the eigenvalues of $A(p)$ are

$$\lambda_1(p) = 1 + p_1 + 3p_2 + \sqrt{p_1^2 + p_2^2}, \quad \lambda_2(p) = 1 + p_1 + 3p_2 - \sqrt{p_1^2 + p_2^2}.$$

It is known that the functions $\lambda_1(p)$ and $\lambda_2(p)$ are not differentiable at $p = 0$ (see [6, p.132]).

Let $\nu = (\nu_1, \nu_2)^T \in \mathbb{C}^2$ be any fixed direction with $\|\nu\|_2 = 1$, and let τ be a real variable. From

$$\lambda_1(\tau\nu) = 1 + \tau(\nu_1 + 3\nu_2) + |\tau|\sqrt{\nu_1^2 + \nu_2^2},$$

$$\lambda_2(\tau\nu) = 1 + \tau(\nu_1 + 3\nu_2) - |\tau|\sqrt{\nu_1^2 + \nu_2^2}$$

we see that the functions $\lambda_1(\tau\nu)$ and $\lambda_2(\tau\nu)$ are single-valued and continuous for the real variable τ. Therefore we may take the functions $\mu_s(\tau\nu)$ described in step 2.1 of §2 as $\lambda_s(\tau\nu)$. Straightforward calculations give

$$D_\nu\mu_1(0) = \nu_1 + 3\nu_2 + \sqrt{\nu_1^2 + \nu_2^2}, \quad D_\nu\mu_2(0) = \nu_1 + 3\nu_2 - \sqrt{\nu_1^2 + \nu_2^2}. \tag{3.1}$$

On the other hand, applying Theorem 1.1 we have

$$\{D_\nu\mu_s(0)\}_{s=1}^2 = \left\{\lambda \;:\; \lambda \in \lambda\left(\nu_1\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix} + \nu_2\begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}\right)\right\}$$
$$= \left\{\nu_1 + 3\nu_2 + \sqrt{\nu_1^2 + \nu_2^2}, \nu_1 + 3\nu_2 - \sqrt{\nu_1^2 + \nu_2^2}\right\}.$$

This coincides with (3.1).

Now we give some applications of Theorem 1.1 and Theorem 1.2.

3.1. Sensitivity of semisimple multiple eigenvalues

Let $C(\nu)$ be a continuous matrix-valued function of $\nu \in \mathbb{C}^N$. Then it is easy to prove that the function $\rho(C(\nu))$ is continuous for $\nu \in \mathbb{C}^N$. Hence according to Theorem 1.1 and Theorem 1.2 we may introduce the following definition:

Definition 3.2. *Let $p, A(p), \mathcal{B}(0), X, Y$ and λ_1 be as in Theorem 1.1. Then the quantity $s_p^{(\nu)}(\lambda_1)$ defined by (1.6) is called the sensitivity of the semisimple multiple eigenvalue λ_1 in the direction $\nu \in \mathbb{C}^N$ with $\|\nu\|_2 = 1$, and the quantity*

$$\hat{s}_p(\lambda_1) \equiv \max_{\substack{\nu \in \mathbb{C}^N \\ \|\nu\|_2 = 1}} s_p^{(\nu)}(\lambda_1) \tag{3.2}$$

is called the sensitivity of the semisimple eigenvalue λ_1 of $A(p)$ at $p = 0$.

By (1.7), we have

$$\hat{s}_p(\lambda_1) = \max_{\sum_{j=1}^{N} |\nu_j|^2=1} \rho\Big(\sum_{j=1}^{n} \nu_j Y_1^H \Big(\frac{\partial A(p)}{\partial p_j}\Big)_{p=0} X_1\Big).$$

Recall that the quantity

$$s_p(\lambda_1) \equiv \Big(\sum_{j=1}^{N} \Big[\rho\Big(Y_1^H \Big(\frac{A(p)}{p_j}\Big)_{p=0} X_1\Big)\Big]^2\Big)^{1/2} \tag{3.3}$$

has been defined as the sensitivity of the semisimple eigenvalue λ_1 (see [6]).

Consider Example 3.1. By (3.2),

$$\begin{aligned}\hat{s}_p(\lambda_1) &= \max_{|\nu_1|^2+|\nu_2|^2=1} \Big\{\Big|\nu_1 + 3\nu_2 + \sqrt{\nu_1^2+\nu_2^2}\Big|, \Big|\nu_1 + 3\nu_2 - \sqrt{\nu_1^2+\nu_2^2}\Big|\Big\}\\ &= \max_{|\nu_1|^2+|\nu_2|^2=1} \{|\nu_1| + 3|\nu_2| + 1\} = \max_{0\le t\le 1} \{t + 3\sqrt{1-t^2} + 1\}\\ &= \sqrt{10} + 1 \approx 4.16228,\end{aligned}$$

and by (3.3)[6],

$$s_p(\lambda_1) = 2\sqrt{5} \approx 4.47214.$$

Since perturbations of the eigenvalues of $A(p)$ at $p = 0$ are not only the perturbations caused by the variances of the parameters along the axes $p_j (j = 1, \cdots, N)$, the quantity $\hat{s}_p(\lambda_1)$ defined by (3.2) is a more reasonable measure of the sensitivity of the semisimple multiple eigenvalue λ_1 of $A(p)$ at $p = 0$. Let us consider the following example:

$$A(p) = \begin{pmatrix} 1 & p_1 \\ p_2 & 1 \end{pmatrix}, \quad p = (p_1, p_2)^T \in \mathbb{C}^2.$$

The matrix $A(p)$ has a semisimple multiple eigenvalue $\lambda_1 = 1$ at $p = 0$, and the eigenvalues of $A(p)$ are $\lambda_1(p) = 1 + \sqrt{p_1 p_2}$, $\lambda_2(p) = 1 - \sqrt{p_1 p_2}$. Let $\nu = (\nu_1, \nu_2)^T \in \mathbb{C}^2$ be any fixed direction with $\|\nu\|_2 = 1$, and let τ be a real variable. Simple calculations show that we may take the functions $\mu_1(\tau\nu), \mu_2(\tau\nu)$ described in step 2.1 of §2 as

$$\mu_1(\tau\nu) = 1 + |\tau|\sqrt{\nu_1\nu_2}, \quad \mu_2(\tau\nu) = 1 - |\tau|\sqrt{\nu_1\nu_2},$$

and

$$\{D_\nu \mu_s(0)\}_{s=1}^{2} = \{\pm\sqrt{\nu_1\nu_2}\}.$$

By (3.2) we have

$$\begin{aligned}\hat{s}_p(\lambda_1) &= \max_{|\nu_1|^2+|\nu_2|^2=1} \rho\left(\nu_1 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \nu_2 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right)\\ &= \max_{|\nu_1|^2+|\nu_2|^2=1} |\sqrt{\nu_1\nu_2}| = \frac{1}{\sqrt{2}},\end{aligned} \tag{3.4}$$

and the maximum can be attained at the directions $\nu = \frac{1}{\sqrt{2}}(e^{i\theta_1}, e^{i\theta_2})^T$, where θ_1, θ_2 are any real scalars. But by (3.3)

$$s_p(\lambda_1) = \sqrt{\Big[\rho\Big(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\Big)\Big]^2 + \Big[\rho\Big(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\Big)\Big]^2} = 0, \tag{3.5}$$

which can not represent the sensitivity of the semisimple multiple eigenvalue λ_1 of $A(p)$ at $p = 0$.

From (3.4) and (3.5) we see that the quantity $s_p(\lambda_1)$ can not give an upper bound of $\hat{s}_p(\lambda_1)$, but (3.2) shows that

$$\rho\left(Y_1^H \left(\frac{\partial A(p)}{\partial p_j}\right)_{p=0} X_1\right) \leq \hat{s}_p(\lambda_1), \quad j = 1, \cdots, N,$$

and so we have

$$s_p(\lambda_1) \leq \sqrt{N}\,\hat{s}_p(\lambda_1).$$

It follows from (3.2) that

$$\hat{s}_p(\lambda_1) \leq \sqrt{\sum_{j=1}^{N} \left\| Y_1^H \left(\frac{\partial A(p)}{\partial p_j}\right)_{p=0} X_1 \right\|^2} \equiv \bar{s}_p(\lambda_1) \tag{3.6}$$

for any consistent norm $\|\cdot\|$ on $\mathbb{C}^{r\times r}$. Hence we may introduce the following definition.

Definition 3.3. *Let $p, A(p), \mathcal{B}(0), X, Y$ and λ_1 be as in Theorem 1.1, and let $\|\cdot\|$ be a consistent norm on $\mathbb{C}^{r\times r}$. Then the quantity $\bar{s}_p(\lambda_1)$ defined by (3.6) is also called the sensitivity of the semisimple multiple eigenvalue λ_1 of $A(p)$ at $p = 0$.*

Consider Example 3.1. Taking $\|\cdot\| = \|\cdot\|_2, \|\cdot\|_\infty$ and $\|\cdot\|_1$, we have

$$\bar{s}_p(\lambda_1) = \sqrt{\left\|\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}\right\|_2^2 + \left\|\begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}\right\|_2^2} \approx 5.02998,$$

$$\bar{s}_p(\lambda_1) = \sqrt{\left\|\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}\right\|_\infty^2 + \left\|\begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}\right\|_\infty^2} \approx 4.47214$$

and

$$\bar{s}_p(\lambda_1) = \sqrt{\left\|\begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}\right\|_1^2 + \left\|\begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}\right\|_1^2} \approx 5.38516,$$

respectively.

Since the right-hand side of (3.2) is, in general, a constrained non-smooth optimization problem, the computation of the quantity $\bar{s}_p(\lambda_1)$ defined by (3.6) is much simpler than that of $\hat{s}_p(\lambda_1)$.

3.2. Condition numbers of semisimple eigenvalues

Let $A = (\alpha_{jk}) \in \mathbb{C}^{n\times n}$. Assume that there are matrices $X, Y \in \mathbb{C}^{n\times n}$ satisfying

$$Y^H X = I, \quad X = (X_1, X_2), \quad Y = (Y_1, Y_2), \quad X_1, Y_1 \in \mathbb{C}^{n\times r}$$

and

$$Y^H A X = \begin{pmatrix} \lambda_1 I^{(r)} & 0 \\ 0 & A_2 \end{pmatrix}, \quad \lambda_1 \overline{\in} \lambda(A_2).$$

Regarding the elements α_{jk} as parameters, by (3.2) we obtain the sensitivity of the semisimple multiple eigenvalue λ_1 with respect to A:

$$\hat{s}_A(\lambda_1) = \max_{\sum_{j,k=1}^{n} |\nu_{jk}|^2 = 1} \rho\left(\sum_{j,k=1}^{n} \nu_{jk} Y_1^H \frac{\partial A}{\partial \alpha_{jk}} X_1\right). \tag{3.7}$$

Write

$$X_1^H = (x_1' \cdots, x_n'), \quad Y_1^H = (y_1', \cdots, y_n'), \quad x_j', y_j' \in \mathbb{C}^r \quad \forall\, j$$

and

$$V = (\nu_{jk}) \in \mathbb{C}^{n\times n}.$$

Then from (3.7) we get

$$\hat{s}_A(\lambda_1) = \max_{\sum_{j,k=1}^{n} |\nu_{jk}|^2=1} \rho\Big(\sum_{j,k=1}^{n} \nu_{jk} Y_j' x_k'^H \Big) = \max_{\|V\|_F=1} \rho(Y_1^H V X_1). \tag{3.8}$$

The quantity $\hat{s}_A(\lambda_1)$ expressed by (3.8) may be obviously regarded as a condition number of the eigenvalue λ_1.

Let

$$X_1 Y_1^H = P_1 \Sigma_1 Q_1^H$$

be the singular value decomposition of $X_1Y_1^H$, where $P_1, Q_1 \in \mathbb{C}^{n\times r}$, $P_1^H P_1 = Q_1^H Q_1 = I^{(r)}$, and $\Sigma_1 = \operatorname{diag}(\sigma_1, \cdots, \sigma_r)$ in which $\sigma_1, \cdots, \sigma_r, 0, \cdots, 0$ are the singular values of $X_1Y_1^H$. Take $V = \frac{1}{\sqrt{r}} Q_1 P_1^H$. Then from (3.5),

$$\hat{s}_A(\lambda_1) = \max_{\|V\|_F=1} \rho(X_1 Y_1^H V) \geq \rho\big(X_1Y_1^H \cdot \frac{1}{\sqrt{r}} Q_1 P_1^H\big) = \frac{1}{\sqrt{r}} \rho(\Sigma_1) = \frac{1}{\sqrt{r}} \|X_1 Y_1^H\|_2. \tag{3.9}$$

On the other hand, we have

$$\hat{s}_A(\lambda_1) = \max_{\|V\|_F=1} \rho(X_1 Y_1^H V) \leq \max_{\|V\|_F=1} \|X_1 Y_1^H V\|_2 \leq \|X_1 Y_1^H\|_2. \tag{3.10}$$

Further, observe that

$$\begin{aligned} \|X_1Y_1^H\|_2 &= \|(X_1Y_1^HY_1X_1^H)^{1/2}\|_2 = [\lambda_{\max}(X_1Y_1^HY_1X_1^H)]^{1/2} \\ &= [\lambda_{\max}((X_1^HX_1)^{1/2}Y_1^HY_1(X_1^HX_1)^{1/2})]^{1/2} \\ &= \|[(X_1^HX_1)^{-1/2}X_1^HY_1(Y_1^HY_1)^{-1}Y_1^HX_1(X_1^HX_1)^{-1/2}]^{-1/2}\|_2, \end{aligned} \tag{3.11}$$

where $\lambda_{\max}(\cdot)$ denotes the maximal eigenvalue of a Hermitian matrix. Hence, if we let

$$U_1 = X_1(X_1^HX_1)^{-1/2}, \quad V_1 = Y_1(Y_1^HY_1)^{-1/2}$$

and define

$$\Theta(X_1, Y_1) \equiv \arccos (U_1^H V_1 V_1^H U_1)^{1/2} > 0$$

then from (3.9)–(3.11) we have

$$\|X_1Y_1^H\|_2 = \|[\cos\Theta(X_1,Y_1)]^{-1}\|_2$$

and

$$\frac{1}{\sqrt{r}} \|[\cos\Theta(X_1,Y_1)]^{-1}\|_2 \leq \hat{s}_A(\lambda_1) \leq \|[\cos\Theta(X_1,Y_1)]^{-1}\|_2. \tag{3.12}$$

The relations (3.12) show that we may also use the quantity

$$\frac{1}{S_1^{(2)}} \equiv \|[\cos\Theta(X_1,Y_1)]^{-1}\|_2$$

as a condition number of the eigenvalue λ_1.

Further discussion on condition numbers of semisimple multiple eigenvalues is given in [7].

References

[1] H. Cartan, Elementary Theory of Analytic Functions of One or Several Complex Variables, Hermann, Paris, 1963.

[2] T. Kato, A Short Introduction to Perturbation Theory for Linear Operators, Springer-Verlag, New York, 1982.

[3] P. Lancaster, Theory of Matrices, Academic Press, New York and London, 1969.

[4] M.L. Overton and R.S. Womersley, On minimizing the spectral radius of a non-symmetric matrix function-optimality conditions and duality theory, submitted to *SIAM J. Matrix Anal. and Applic.*

[5] Sun Ji-guang, Sensitivity analysis of multiple eigenvalues (I), *J. Comp. Math.*, **6** : 1 (1988), 28–38.

[6] Sun Ji-guang, Sensitivity analysis of multiple eigenvalues (II), *J. Comp. Math.*, **6** : 2 (1988), 131–141.

[7] Sun Ji-guang, Condition unmbers of semisimple multiple eigenvalues, *Math. Numer. Sinica.*, 13 (1991), 58–66.