Eigenvectors of Matrix Polynomials

Murray Mannos

It is the object of this paper to compare the eigenvectors of an arbitrary \( n \times n \) matrix \( A \) over the complex field with those of the matrix polynomial \( f(A) \). While it is well known that each eigenvector of \( A \) is an eigenvector of \( f(A) \), it is not, in general, true that \( A \) and \( f(A) \) have identical eigenvectors. In this regard a necessary and sufficient condition that \( A \) and \( f(A) \) have identical eigenvectors is given. The condition is that both (1) and (2) hold:

1. \( f'(\lambda) \neq 0 \) for all eigenvalues \( \lambda \) of the matrix \( A \) corresponding to nonlinear elementary divisors.
2. The values of \( f(\mu) \) are distinct for all eigenvalues \( \mu \) of the matrix \( A \) corresponding to linear elementary divisors.

When either (1) or (2) fails to hold, then \( f(A) \) has eigenvectors that are not eigenvectors of \( A \). This situation is also discussed.

The vector space of eigenvectors of the matrix \( A \) corresponding to the eigenvalue \( \lambda \) shall be denoted by \( V_\lambda[A] \). The vector space spanned by the eigenvectors of \( A \) shall be denoted by \( V[A] \). Further, \( d\{V_\lambda[A]\} \) and \( d\{V[A]\} \) shall denote their dimensions. It is clear that each eigenvector of \( A \) is an eigenvector of \( f(A) \). That is, \( V_\lambda[A] \subseteq V_{f(\lambda)}[f(A)] \) for each eigenvalue \( \lambda \) of \( A \). Thus \( V[A] \subseteq V[f(A)] \).

Let \( J \) be the Jordan canonical form of \( A \). Then there exists a nonsingular matrix \( P \) such that \( P^{-1}AP = J \) and so \( P^{-1}f(A)P = f(J) \).

**Lemma 1.** The eigenvectors of \( A \) and \( f(A) \) are identical if, and only if, the eigenvectors of \( J \) and \( f(J) \) are identical.

**Proof.** This follows from the fact that \( P\xi \) is an eigenvector of \( PBP^{-1} \) if \( \xi \) is an eigenvector of the matrix \( B \).

Since \( d\{V_\lambda[J]\} = V_\lambda[A] \), it follows that \( d\{V[J]\} - d\{V[A]\} = d\{V[f(J)]\} - d\{V[J]\} \), where \( P\{V[J]\} \) denotes the space of all vectors \( P\xi \), where \( \xi \) is an eigenvector of \( J \) corresponding to \( \lambda \).

**Lemma 2.** If \( D = \text{diag} \{ \alpha, \beta, \ldots, \beta; \ldots, \pi, \pi, \ldots, \pi \} \), where \( \alpha, \beta, \ldots, \pi \) are distinct, then \( V[D] \equiv V[f(D)] \). Furthermore, the eigenvectors of \( D \) are identical with those of \( f(D) \) if, and only if, \( f(\alpha), f(\beta), \ldots, f(\pi) \) are all distinct.

**Proof.** If \( D \) is of order \( n \), it is clear that both the eigenvectors of \( D \) and \( f(D) \) each generate the whole \( n \)-dimensional vector space. If \( f(\alpha) = f(\beta) \), then it is easily seen that \( f(D) \) has eigenvectors corresponding to the eigenvalue \( f(\alpha) = f(\beta) \), which are not eigenvectors of \( D \).

**Lemma 3.** If \( f'(\lambda) \neq 0 \) for all \( \lambda \) in the upper right-hand corner of the matrix. Then the nullity of \( f(J_m(\lambda)) - f(\lambda)I \) is \( m - (r + 1) \), as may be observed by noting the nonzero minor of order \( m - (r + 1) \) in the upper right-hand corner of the matrix. Thus the nullity of \( f(J_m(\lambda)) - f(\lambda)I \) is \( r + 1 \). If the first \( m \)-derivatives of \( f(\lambda) \) vanish at \( \lambda = \lambda_0 \), then the nullity of \( f(J_m(\lambda)) - f(\lambda)I \) is \( m \). In a similar fashion one finds that the matrix \( J_m(\lambda) - M \) has nullity equal to 1. This establishes the lemma.

**Corollary.** The matrices \( J_m(\lambda) \) and \( f(J_m(\lambda)) \) have identical eigenvectors if, and only if, \( f'(\lambda) \neq 0 \).

It is to be noted that in this case \( V_{f(\lambda)}[f(J_m(\lambda))] = \) 

\[
\begin{bmatrix}
  f(\lambda) & f'(\lambda) & f''(\lambda) & \cdots & f^{(m-1)}(\lambda) \\
  0 & f(\lambda) & f'(\lambda) & \cdots & f^{(m-2)}(\lambda) \\
  0 & 0 & f(\lambda) & \cdots & f^{(m-3)}(\lambda) \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & 0 & \cdots & f(\lambda) \\
  0 & 0 & 0 & \cdots & 0
\end{bmatrix}
\]

Proofs of the lemma may be found in Wedderburn and MacDuffee.3

**Lemma 4.** If \( f'(\lambda) = f''(\lambda) = \cdots = f^{(r)}(\lambda) = 0 \) and \( f^{(r+1)}(\lambda) \neq 0 \) \( r=1,2, \ldots, m-2 \), then
\[
d\{V_{f(\lambda)}[f(J_m(\lambda))]\} - d\{V_\lambda[J_m(\lambda)]\} = r,
\]
and conversely. Also,
\[
d\{V_{f(\lambda)}[f(J_m(\lambda))]\} - d\{V_\lambda[J_m(\lambda)]\} = m-1
\]
if, and only if, \( f' \neq f'' \neq \cdots \neq f^{(r)} \neq 0 \), but \( f^{(r+1)} \neq 0 \), it follows that the rank of the matrix \( f(J_m(\lambda)) - f(\lambda)I \) is \( m - (r + 1) \), as may be observed by noting the nonzero minor of order \( m - (r + 1) \) in the upper right-hand corner of the matrix. Thus the nullity of \( f(J_m(\lambda)) - f(\lambda)I \) is \( r + 1 \). If the first \( m \)-derivatives of \( f(x) \) vanish at \( x = \lambda \), then the nullity of \( f(J_m(\lambda)) - f(\lambda)I \) is \( m \). In a similar fashion one finds that the matrix \( J_m(\lambda) - M \) has nullity equal to 1. This establishes the lemma.

1. Preparation of this paper was sponsored (in part) by the Office of Scientific Research, US.A.F.
and $V_α[J_α(λ)]$ are each spanned by the single column vector $[1, 0, \ldots, 0]^T$.

At this point it will be convenient to introduce the language of elementary divisors. Each block in the Jordan canonical form of a matrix $A$ corresponds to an elementary divisor $(x-λ)^m$, and conversely. Each such block is called the hypercompanion matrix $J_α(λ)$ of the polynomial $(x-λ)^m$.

Suppose the nonlinear elementary divisors of the matrix $A$ corresponding to the eigenvalue $λ$ are of the form $(x-λ)^{n_1}, (x-λ)^{n_2}, \ldots, (x-λ)^{n_k}$, where $n_1, n_2, \ldots, n_k$ are integers such that $n_1 > n_2 > \ldots > n_k > 1$. Furthermore, suppose that $(x-λ)^{n_1}$ appears as an elementary divisor $n_1$ times. Let $p_λ = \frac{n_1}{\prod n_k}$ denote the total number of nonlinear elementary divisors of $A$ corresponding to the eigenvalue $λ$. Denote by $n_λ$ the number of linear elementary divisors of $A$ corresponding to the eigenvalue $λ$. Further, denote by $K_λ$ the direct sum of the hypercompanion matrices of the elementary divisors corresponding to $λ$. Set

$$d_λ = d\{V(f(λ)[J])\} - d\{V_α[K_λ]\}.$$

**Lemma 5.** If the first $m^{(i)} - 1$ derivatives of $f(x)$ vanish at $x = λ$, then $d_λ = \frac{n_k}{\prod n_k}(m^{(i)} - 1)$; whereas, if the first $r < m^{(i)} - 1$ derivatives of $f(x)$ vanish at $x = λ$, but $f^{r+1}(λ) \neq 0$, then

$$d_λ = \frac{n_k}{\prod n_k} r + \frac{n_k}{\prod n_k} (m^{(i)} - 1),$$

where $m^{(i)}$ is the largest of the integers $m^{(i)}_i (i = 1, 2, \ldots, k)$ such that $m^{(i)}_i \leq r$.

**Proof.** This follows from lemma 4, the fact that the nullity of $K_λ - λI$ is the sum of the nullities of the characteristic matrices of the hypercompanion matrices of the individual elementary divisors corresponding to the eigenvalue $λ$ and a similar statement about the nullity of $f(λ) - f(λ)I$.

**Corollary.** The matrices $K_λ$ and $f(λ)$ have identical eigenvectors if, and only if, $f'(λ) \neq 0$.

**Proof.** This follows from the fact that $d_λ = 0$.

The Jordan canonical form $J = \text{diag} [K₀, K₁, \ldots, K_s]$ of $A$ is a direct sum of matrices $K_λ$, where $λ = α, β, \ldots$ runs through the distinct eigenvalues of $A$. The subsequent theorems of this paper involve the following main conditions:

**Condition 1.** $f'(λ) \neq 0$ for all eigenvalues $λ$ of the matrix $A$ corresponding to nonlinear elementary divisors.

**Condition 2.** The values $f(µ)$ are distinct for all eigenvalues $µ$ of the matrix $A$ corresponding to linear elementary divisors.

The first theorem concerns the case when condition 1 does not hold, whether condition 2 holds or not.

**Theorem 1.** If condition 1 does not hold, then

$$d\{V_α[A]\} - d\{V[A]\} = \sum_λ d_λ,$$

where $λ$ varies through all distinct eigenvalues of $A$ corresponding to nonlinear elementary divisors, and $d_λ$ is computed as in lemma 5.

**Proof.** One notes first that

$$V[A] = V_α[A] + V_β[A] + \ldots + V_ρ[A],$$

and

$$V[f(λ)] = V_α[f(λ)] + V_β[f(λ)] + \ldots + V_ρ[f(λ)].$$

If $f(α), f(β), \ldots, f(ρ)$ are distinct, for a fixed eigenvalue $λ$, the nullity of $J - λI$ is the same as the nullity of $K_λ - λI$, and the nullity of $f(λ) - f(λ)I$ is the same as the nullity of $f(λ) - f(λ)I$. It follows that

$$d\{V_α[f(λ)]\} - d\{V_β[A]\} = d_λ.$$

Summing over distinct eigenvalues, one obtains

$$d\{V[f(λ)]\} - d\{V[A]\} = \sum_λ d_λ.$$

By the statement following the proof of lemma 1 it follows that

$$d\{V[f(λ)]\} - d\{V[A]\} = \sum_λ d_λ.$$

If $f(α) = f(β) = \ldots = f(ρ)$, then the nullity of $f(λ) - f(α)I$ minus the sum of the nullities of $J - αI$, $J - βI$, $\ldots$, and $J - ρI$ is $d_α + d_β + \ldots + d_ρ + d_λ$ and

$$d\{V_α[f(λ)]\} - d\{V_α[A]\} - d\{V_β[A]\} - \ldots - d\{V_ρ[A]\} = d_α + d_β + \ldots + d_ρ.$$
where \( p_\lambda + n_\lambda \) is the total number of elementary divisors (nonlinear and linear) corresponding to \( \lambda \).

Finally, the next result covers the case in which conditions 1 and 2 both hold.

**Theorem 3.** The matrices \( A \) and \( f(A) \) have identical eigenvectors if, and only if (1) \( f'(\lambda) \neq 0 \) for all eigenvalues \( \lambda \) of the matrix \( A \) corresponding to nonlinear elementary divisors, and (2) the values \( f(\mu) \) are distinct for all eigenvalues \( \mu \) of the matrix \( A \) corresponding to linear elementary divisors.

**Remark 1.** It may readily be seen that \( V[J] \) and \( V[f(J)] \) each can be generated by a set of linearly independent vectors, each of which has a 1 in a single component and 0 in all the remaining components. Thus if \( J = P^{-1}AP \) is the Jordan canonical form of \( A \), then \( V[f(A)] \) can be generated by a subset of the column vectors of \( P \).

**Remark 2.** A simple application of the foregoing theory shows that if \( A \) is a 2\times2 matrix, then \( A \) and \( f(A) \) have the same eigenvectors unless either (1) \( A \) is diagonalizable and has distinct eigenvalues \( \alpha, \beta \) for which \( f(\alpha) = f(\beta) \), or (2) \( A \) is nondiagonalizable and \( f(x) = k(x)(x-\alpha)^2 + c \), where \( \alpha \) is the eigenvalue of \( A \), \( k(\lambda) \) is an arbitrary polynomial, and \( c \) is an arbitrary constant.

The following example is given as an illustration:

\[
A = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & -1 & -1 & -1 & -1 & -2 & -1 \\
1 & 0 & 1 & 1 & 0 & -1 & 0
\end{bmatrix}
\]

and \( f(x) = x^6 + x^5 - x^4 - x^3 + 2 \).

For \( P = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix} \)

it may be verified that

\[
P^{-1} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1
\end{bmatrix}
\]

and that the Jordan canonical form of \( A \) is

\[
J = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & -1
\end{bmatrix}
\]

where

\[
\begin{align*}
K_1 &= J_1(1) = [1], \\
K_2 &= \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}, \\
K_3 &= \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}, \\
K_4 &= \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}
\end{align*}
\]

it follows that

\[
\begin{align*}
f(0) &= f(1) = f(-1) = 2, \\
f'(0) &= f'(-1) = 0, \\
f''(0) &= f''(-1) = 0
\end{align*}
\]

It is readily seen that \( V[J] \) is generated by \([1,0,0,0,0,0,0]^T \), \( V_0[J] \) is generated by \([1,0,0,0,0,0,0]^T \) and \([0,0,0,0,0,0,0]^T \), and \( V_{-1}[J] \) is generated by \([0,0,0,0,0,1,0]^T \). Since \( V[J] = V[J] + V_0[J] + V_{-1}[J] \), it follows that \( d(V[J]) = 4 \). Since \( f[J] \) is a scalar matrix, \( d(V[f(J)]) = 7 \). Hence \( d(V[A]) = 3 \). The same result is arrived at by the use of theorem 1, where \( \sum d_\lambda \) is calculated as in lemma 5. Since \( f'(0) = 0, f''(0) = 0, \) and \( f'(-1) = 0, \) where 0, -1 are the eigenvalues corresponding to nonlinear elementary divisors, it follows that

\[
d_0 = n_0^{(2)} + n_0^{(3)} = 1.2 = 2
\]

and

\[
d_{-1} = n_{-1}^{(2)} + n_{-1}^{(3)} = 1.1 = 1
\]

35
and
\[ \sum_{\lambda} d_{\lambda} = d_0 + d_{-1} = 2 + 1 = 3, \]
as before. By the corollary following theorem 2,
\[ d(V_2[f(A)]) - d(V_1[A]) = d_0 + d_{-1} + p_0 + p_{-1} + n_0 + n_{-1} \]
\[ = 2 + 1 + 1 + 1 + 0 = 6, \]
which checks with the observed results above.

From a glance at the above eigenvectors generating $V[J]$, it is clear that $V[A]$ is generated by the first, second, fifth, and sixth-column vectors of $P$. Since $f(A)$ is a scalar, $V[f(A)]$ is generated by all seven column vectors of $P$.

The author thanks A. H. Clifford, O. Taussky, and N. Wiegmann for their helpful suggestions.

WASHINGTON, December 1, 1952.