# MATH 204 C03 – MULTIPICITY OF EIGENVALUES COMPARED TO THE DIMENSION OF ITS EIGENSPACE

JON FICKENSCHER

## Outline

In section 5.1 of our text, we are given (without proof) the following theorem (it is Theorem 2):

**Theorem.** *Let $p(\lambda)$ be the characteristic polynomial for an $n \times n$ matrix $\mathbf{A}$ and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the roots of $p(\lambda)$. Then the dimension $d_i$ of the $\lambda_i$-eigenspace of $\mathbf{A}$ is at most the multiplicity $m_i$ of $\lambda_i$ as a root of $p(\lambda)$.*

The book will address this theorem in Chapter 7. However, we aim to prove this comment given our information in Chapter 5. Portions of this proof also may serve as a way to understand the concepts in Chapter 7 as well. Even though this theorem was given early in the chapter, we will prove it in as general a context as possible. This means that we will allow $\mathbf{A}$ and the $\lambda_i$'s to be complex. We will prove this by reducing to two lemmas. We will discuss the idea behind each lemma before each proof.

## Terms

We use the following conventions (which may differ slightly from the book):

- $\mathcal{M}_C(m,n)$ is the complex vector space of $m \times n$ matrices with complex coefficients. $\mathcal{M}(m,n)$ is the subset of matrices with real coefficients. In any context $\mathbf{I}$ is the identity matrix with the size omitted.
- For $\mathbf{A} \in \mathcal{M}_C(n,n)$,
  - $\det(\mathbf{A})$ is the determinant of $\mathbf{A}$ (this may be a complex number).
  - $p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$ is the characteristic polynomial of $\mathbf{A}$.
  - $\mathrm{COL}(\mathbf{A})$ is the columnspace of $\mathbf{A}$.
  - $\mathrm{NULL}(\mathbf{A})$ is the nullspace of $\mathbf{A}$.

**Definition.** Given a polynomial $p(x)$ with complex coefficients, we say that a complex value $x_0$ is a **root of multiplicity** $m > 0$ of $p$ if $p(x_0) = p'(x_0) = \cdots = p^{(m-1)}(x_0) = 0$ and $p^{(m)}(x_0) \neq 0$, where $p^{(k)}$ is the $k^{\text{th}}$ derivative of $p$. This is equivalent to the statement $p(x) = (x - x_0)^m q(x)$ for some polynomial $q(x)$ where $q(x_0) \neq 0$. We can also say that $x_0$ is a root of multiplicity 0 if it is not a root.

**Example.** $p(x) = x^3 - 2x^2 + x$ is a polynomial of degree 3 that may be factored as $p(x) = x(x-1)^2$. So 0 is a root of multiplicity 1, and 1 is a root of multiplicity 2. We may check our equivalent definition by noting that $p(0) = 0$ while $p'(0) = 1 \neq 0$ and $p(1) = p'(1) = 0$ while $p''(1) = 2 \neq 0$.

*Date*: April 22, 2012.

## Results from class

We discussed the following results in class (and some results were related to homework problems). The reader is encouraged to prove these statements.

**Exercise.** If $\mathbf{A}, \mathbf{B} \in \mathcal{M}_C(n,n)$ are similar, meaning there exists an invertible $\mathbf{S} \in \mathcal{M}_C(n,n)$ such that $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$, then $p_{\mathbf{A}} = p_{\mathbf{B}}$.

**Exercise.** Let $\mathbf{A} \in \mathcal{M}_C(n,n)$, $\mathbf{B} \in \mathcal{M}_C(n,m)$, $\mathbf{C} \in \mathcal{M}_C(m,n)$ and finally $\mathbf{D} \in \mathcal{M}_C(m,m)$. If $\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$, then $\det(\mathbf{X}) = \det(\mathbf{Y}) = \det(\mathbf{A})\det(\mathbf{D})$.

**Exercise.** Let $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{X}, \mathbf{Y}$ be as in the previous exercise. Then

$$p_{\mathbf{X}}(\lambda) = p_{\mathbf{Y}}(\lambda) = p_{\mathbf{A}}(\lambda)p_{\mathbf{D}}(\lambda).$$

**Exercise.** Let $\mathbf{A} \in \mathcal{M}_C(n,n)$. The matrix $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$.

## The Proof

**First Lemma.**

**Lemma 1.** *Let $\mathbf{A} \in M_C(n,n)$. If $0$ is a root of $p_{\mathbf{A}}(\lambda)$ of multiplicity $m$, then*

$$\dim(\mathrm{NULL}(\mathbf{A})) \leq m.$$

**Remark.** This lemma is saying that we may look only at the 0-eigenspace (the nullspace) and the multiplicity of 0 as a root for a given matrix. To make the same claim for any other eigenvalue, we just shift our matrix by $\mathbf{I}$ times that eigenvalue.

*Proof that Lemma 1 proves the Theorem.* Let $\mathbf{A} \in M_C(n,n)$ and $\mu$ be a root of $p_{\mathbf{A}}$ of multiplicity $m$. We define

$$\mathbf{B} = \mathbf{A} - \mu\mathbf{I}.$$

By direct calculation,

$$\begin{aligned} p_{\mathbf{B}}(\lambda) &= \det(\mathbf{B} - \lambda\mathbf{I}) \\ &= \det((\mathbf{A} - \mu\mathbf{I}) - \lambda\mathbf{I}) \\ &= \det(\mathbf{A} - (\mu + \lambda)\mathbf{I}) \\ &= p_{\mathbf{A}}(\lambda + \mu), \end{aligned}$$

or equivalently, $p_{\mathbf{B}}(\lambda - \mu) = p_{\mathbf{A}}(\lambda)$. Either way, for every $k \geq 0$, the $k^{\text{th}}$ derivative of $p_{\mathbf{A}}$ evaluated at $\mu$ and the $k^{\text{th}}$ derivative of $p_{\mathbf{B}}$ evaulated at 0 are equal, or

$$\forall k \geq 0,\ p_{\mathbf{A}}^{(k)}(\mu) = p_{\mathbf{B}}^{(k)}(0),$$

by the chain rule of derivatives and induction. So we see that $\mu$ is a root of multiplicity $m$ of $p_{\mathbf{A}}$ if and only if 0 is a root of $p_{\mathbf{B}}$ of multiplicity $m$.

**Exercise.** Show that the nullspace of $\mathbf{B}$ is equal to the $\mu$-eigenspace of $\mathbf{A}$.

Lemma 1 states that the nullity of $\mathbf{B}$ is less than or equal to $m$, which implies that the $\mu$-eigenspace of $\mathbf{A}$ has dimension less than or equal to $m$. This is the conclusion needed for the Theorem. $\square$

**Second Lemma.**

**Lemma 2.** *Let* $\mathbf{A} \in M_C(n,n)$*. Then the following are equivalent:*

1. $\mathbf{A}^k = \mathbf{0}$ *for some* $k > 0$*.*
2. $\mathbf{A}^k = \mathbf{0}$ *for some* $1 \le k \le n$*.*
3. $\mathbf{A}^n = \mathbf{0}$*.*
4. $p_{\mathbf{A}}$ *has* $0$ *as a root of multiplicity* $n$*. In other words,* $p_{\mathbf{A}}(\lambda) = (-1)^n \lambda^n$*.*

**Remark.** In the proof that follows, we remind ourselves of some properties of columnspaces and nullspaces of the powers of a square matrix $\mathbf{A}$. We then find a way to divide our space $\mathbb{C}^n$ into two independent subspaces, one on which $\mathbf{A}$ is invertible and the other on which multiplying by $\mathbf{A}$ repeatedly results in $\{\mathbf{0}\}$. We then use Lemma 2 to see that the multiplicity of 0 as a root of $p_{\mathbf{A}}$ is equal to the dimension of the second space we find. This second space contains the nullspace of $\mathbf{A}$, implying Lemma 1.

*Proof that Lemma 2 proves Lemma 1.* Let $\mathbf{A} \in M_C(n,n)$. We have the following relationships:

$$\begin{array}{ccccccccc} \mathrm{COL}(\mathbf{I} = \mathbf{A}^0) & \supseteq & \mathrm{COL}(\mathbf{A}) & \supseteq & \mathrm{COL}(\mathbf{A}^2) & \supseteq & \mathrm{COL}(\mathbf{A}^3) & \supseteq & \ldots \\ \mathrm{NULL}(\mathbf{I} = \mathbf{A}^0) & \subseteq & \mathrm{NULL}(\mathbf{A}) & \subseteq & \mathrm{NULL}(\mathbf{A}^2) & \subseteq & \mathrm{NULL}(\mathbf{A}^3) & \subseteq & \ldots \end{array}$$

We have discussed a number of concepts that lead us to the conclusion:

**Exercise.** There exists a unique minimum $j \ge 0$ such that for all $k > j$,

$$\mathrm{COL}(\mathbf{A}^j) = \mathrm{COL}(\mathbf{A}^k) \text{ and } \mathrm{NULL}(\mathbf{A}^j) = \mathrm{NULL}(\mathbf{A}^k).$$

Let $j$ be this unique value. If $j = 0$, then $\mathbf{A}$ is invertible, does not have 0 as an eigenvalue and satisifes Lemma 1 trivially. So we may assume that $j > 0$. Because $\mathrm{COL}(\mathbf{A}^{j+1}) = \mathrm{COL}(\mathbf{A}^j)$, $\mathbf{A}$ is invertible. We mean that for every $\mathbf{X}$ in $\mathrm{COL}(\mathbf{A}^j)$, there exists a unique $\mathbf{Y}$ in $\mathrm{COL}(\mathbf{A}^j)$ such that $\mathbf{AX} = \mathbf{Y}$. Likewise, if $\mathbf{X}$ belongs to $\mathrm{NULL}(A^j)$, then $\mathbf{AX} \in \mathrm{NULL}(\mathbf{A}^{j-1}) \subsetneq \mathrm{NULL}(\mathbf{A}^j)$. Let $M$ be the dimension of $\mathrm{NULL}(\mathbf{A}^j)$ and $N$ be the dimension of $\mathrm{COL}(\mathbf{A}^j)$. We may therefore choose bases $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ of $\mathrm{COL}(\mathbf{A}^j)$ and $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_M\}$ of $\mathrm{NULL}(\mathbf{A}^j)$ in $\mathbb{C}^n$.

**Exercise.** Explain why $M + N = n$, and show that $\{\mathbf{X}_1, \ldots, \mathbf{X}_N, \mathbf{Y}_1, \ldots, \mathbf{Y}_M\}$ is a basis for $\mathbb{C}^n$.

If we define $\mathbf{S} = [\mathbf{X}_1, \ldots, \mathbf{X}_N, \mathbf{Y}_1, \ldots, \mathbf{Y}_M]$ then

$$\mathbf{A} = \mathbf{S} \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix} \mathbf{S}^{-1}$$

where $\mathbf{B} \in \mathcal{M}_C(N,N)$ is invertible and $\mathbf{C} \in \mathcal{M}_C(M,M)$.

**Exercise.** $\mathbf{C}^j = \mathbf{0}$.

So we conclude that $p_{\mathbf{A}} = p_{\mathbf{B}} p_{\mathbf{C}}$ where $p_{\mathbf{B}}$ is degree $N$ that doesn't have 0 as a root, and $p_{\mathbf{C}}$ is an $M$ degree polynomial with 0 as a root of multiplicity $M$ (by Lemma 2). We see that by our assumptions on $\mathbf{A}$, $M$ must be $m$. Also because

$$\mathrm{NULL}(\mathbf{A}) \subseteq \mathrm{NULL}(\mathbf{A}^j),$$

we conclude that $\mathrm{null}(\mathbf{A}) \le \mathrm{null}(\mathbf{A}^j) = m$. $\square$

**Proof of Second Lemma.**

**Remark.** A matrix that satisfies one of the first three condition is called **nilpotent**. This lemma proves that any matrix has zero as its only eigenvalue if and only if it is nilpotent. We begin the proof by showing the equivalence of the first three statements. All we show is that a matrix must become zero by the time we take it to the $n^{th}$ power or it is not nilpotent. We reserve the right to one more remark to follow the proof.

*Proof of Lemma 2.* We will first show that the first three statements are equivalent with relative ease (and brevity). We note that (3) implies (2) directly. If we assume (2), then $\mathbf{A}^k = \mathbf{0}$ for some $1 \leq k \leq n$. We see then that

$$\mathbf{A}^n = \mathbf{A}^{n-k}\mathbf{A}^k = \mathbf{A}^{n-k}\mathbf{0} = \mathbf{0},$$

so (3) holds. So (2) and (3) are equivalent. We also see that $k$ and $n$ from statements (2) and (3) are both greater than 0, so either one implies (1). We will prove (1) implies (3) by the contrapositive, namely: if (3) fails then (1) fails. Suppose $\mathbf{A}^n \neq \mathbf{0}$. Then in the sequence of $n+1$ non-increasing values:

$$0 = \mathrm{null}(\mathbf{A}^0 = \mathbf{I}) \leq \mathrm{null}(\mathbf{A}) \leq \mathrm{null}(\mathbf{A}^2) \leq \cdots \leq \mathrm{null}(\mathbf{A}^n) < n$$

there must be a value $i < n$ such that $\mathrm{null}(\mathbf{A}^i) = \mathrm{null}(\mathbf{A}^{i+1})$.

**Exercise.** For all $j > i$, $\mathrm{null}(\mathbf{A}^j) = \mathrm{null}(\mathbf{A}^i) < n$.

This implies that no power of $\mathbf{A}$ may be $\mathbf{0}$, as $\mathbf{A}^j$ for any $j > i$ does not have full nullity. So (1) fails if (3) fails. We have conluded out proof of (1), (2) and (3).

We will show that (4) implies (3) by contrapositive. We therefore assume that $\mathbf{A}^n \neq \mathbf{0}$. From our previous paragraph and our previous proof, there is a unique minimum $j \geq 0$ such that $\mathbf{A}$ is invertible on $\mathrm{COL}(\mathbf{A}^j)$, and we call $r > 0$ the dimension of $\mathrm{COL}(\mathbf{A}^j)$. As before, we may find a basis such that, if $\mathbf{S}$ is the change of basis matrix,

$$\mathbf{A} = \mathbf{S}\begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}\mathbf{S}^{-1}$$

where $\mathbf{B} \in \mathcal{M}_C(r,r)$ is invertible. We conclude that $p_{\mathbf{A}} = p_{\mathbf{B}}p_{\mathbf{C}}$ may only have zero as a root of multiplicity at most $n - r < n$ (zero is not a root of $p_{\mathbf{B}}$). So (4) fails if (3) fails.

Now we will show that (2) implies (4). Let $k \leq n$ be the minimum integer such that $\mathbf{A}^k = \mathbf{0}$ and define $n_i = \mathrm{null}(\mathbf{A}^i)$ for $1 \leq i \leq k$. We must have that

$$0 < n_1 < n_2 < n_3 < \cdots < n_k = n.$$

We are then able to choose a basis of $\mathbb{C}^n$,

$$\{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$$

such that for each $i$, $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_i}\}$ is a basis for $\mathrm{NULL}(\mathbf{A}^i)$.

**Exercise.** Show that if $i$ satisfies $1 \leq i \leq n_1$, then $\mathbf{A}\mathbf{Y}_1 = \mathbf{0}$. If $i$ instead satisfies $n_{j-1} < i \leq n_j$ for some $j \leq k$, then

$$\mathbf{Y}_i \in \mathrm{NULL}(\mathbf{A}^j) \setminus \mathrm{NULL}(\mathbf{A}^{j-1}), \text{ and } \mathbf{A}\mathbf{Y}_i \in \mathrm{NULL}(\mathbf{A}^{j-1}),$$

or in words, $\mathbf{Y}_i$ belongs to $\mathrm{NULL}(\mathbf{A}^j)$ but not $\mathrm{NULL}(\mathbf{A}^{j-1})$ and $\mathbf{A}\mathbf{Y}_i$ belongs to $\mathrm{NULL}(\mathbf{A}^{j-1})$.

So we see that for each $i$, $\mathbf{AY}_i$ is either $\mathbf{0}$ or a linear combination of $\mathbf{Y}_j$'s for $j$'s strictly less than $i$. If we let $\mathbf{S} = \begin{bmatrix}\mathbf{Y}_1, \dots, \mathbf{Y}_n\end{bmatrix}$, then $\mathbf{A} = \mathbf{SBS}^{-1}$ where $\mathbf{B}$ is upper triangular. Moreover the diagonal entries of $\mathbf{B}$ must be zero.

**Exercise.** Show that $p_{\mathbf{B}}(\lambda) = (-1)^n\lambda^n$ if $\mathbf{B} \in \mathcal{M}_C(n,n)$ is upper (or lower) triangular with zeros along the diagonal.

Because $p_{\mathbf{A}} = p_{\mathbf{B}}$, we have finished showing that (4) follows from (2).

□

**Remark.** The basis $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ may be chosen in a particular fashion (we leave the details as an exercise for the reader) to resemble the results in Chapter 7. Consider $\mathbf{Y}_n \in \mathrm{NULL}(\mathbf{A}^j) \setminus \mathrm{NULL}(\mathbf{A}^{j-1})$. Then $\mathbf{A}^i\mathbf{Y}_n \in \mathrm{NULL}(\mathbf{A}^{j-i}) \setminus \mathrm{NULL}(\mathbf{A}^{j-i-1})$ for each $1 \le i \le j-1$ and $\mathbf{A}^j\mathbf{Y}_n = \mathbf{0}$. So $\mathbf{Y}_n$,$\mathbf{AY}_n, \dots, \mathbf{A}^{j-1}\mathbf{Y}_n$ is an independent set, which is called $\mathcal{B}$. If we may perform this same process on each $\mathbf{Y}_i$ for all $i > n_{j-1}$ and add them to our set $\mathcal{B}$. This set will remain independent. We continue in this fashion by finding the largest index $i$ such that $\mathbf{Y}_i$ is independent of $\mathcal{B}$ and add all powers of $\mathbf{A}$ times $\mathbf{Y}_i$ that are non-zero. When we finish, we may reorder our set $\mathcal{B}$ as $\{\mathbf{Y}'_1, \dots, \mathbf{Y}'_n\}$ such that for each $i$, $\mathbf{AY}'_i = \mathbf{Y}'_{i-1}$ or $\mathbf{AY}'_i = \mathbf{0}$.

**Example.** Let

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 & -2 \\ 2 & 0 & 0 & -2 \\ 0 & 2 & -2 & 0 \\ 0 & 2 & -2 & 0 \end{bmatrix}.$$

We may verify the following:

- $p_{\mathbf{A}}(\lambda) = \lambda^4$.
- $null(\mathbf{A}) = 2$.
- $null(\mathbf{A}^2) = 3$.
- $null(\mathbf{A}^3) = 4$ or equivalently, $\mathbf{A}^3 = \mathbf{0}$.

So from our terminology in the proof, $n_1 = 2$, $n_2 = 3$ and $n_3 = 4$. The vector $\mathbf{Y}'_3 = [1,0,0,0]^t$ is not in $\mathrm{NULL}(\mathbf{A}^2)$. Let $\mathbf{Y}'_2 = \mathbf{AY}'_3 = [2,2,0,0]^t$ and $\mathbf{Y}'_1 = \mathbf{AY}'_2 = [4,4,4,4]^t$. The set $\{\mathbf{Y}'_1 . \mathbf{Y}'_2, \mathbf{Y}'_3\}$ is independent. Because $null(\mathbf{A}) = 2$, we must find another vector in $\mathrm{NULL}(\mathbf{A})$ that is independent of our three vectors. However, we note that we only need to find an element that is independent of $\mathbf{Y}'_1$ (it will automatically be independent of $\mathbf{Y}'_2$ and $\mathbf{Y}'_3$). The vector $\mathbf{Y}'_4 = [1,0,0,1]^t$ satisfies our conditions. If we let $\mathbf{S} = [\mathbf{Y}'_1, \mathbf{Y}'_2, \mathbf{Y}'_3, \mathbf{Y}'_4]$, then $\mathbf{A} = \mathbf{SBS}^{-1}$ where

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

**Example.** If $\mathbf{A}$ is from the previous example, let $\mathbf{A}' = \mathbf{A} + \mu\mathbf{I}$. Then we see that (to reconcile with Lemma 1),

- $p_{\mathbf{A}'}(\lambda) = (\lambda - \mu)^4$.
- $null(\mathbf{A}' - \mu\mathbf{I}) = 2$.
- $null[(\mathbf{A}' - \mu\mathbf{I})^2] = 3$.
- $null[(\mathbf{A}' - \mu\mathbf{I})^3] = 4$.

By choosing the same vectors and $\mathbf{S}$ from the previous example, then $\mathbf{A}' = \mathbf{S}\mathbf{B}'\mathbf{S}^{-1}$ where

$$\mathbf{B}' = \begin{bmatrix} \mu & 1 & 0 & 0 \\ 0 & \mu & 1 & 0 \\ 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & \mu \end{bmatrix}.$$

We will discuss this form of matrix in Chapter 7.