Some Proofs About Eigenvalues and Eigenvectors

Consider a linear operator $\mathcal{L} : \mathcal{V} \to \mathcal{V}$ where $\mathcal{V}$ is a finite dimensional complex vector space. An eigenvector vector for $\mathcal{L}$ is a vector $\mathbf{r} \in \mathcal{V}$ satisfying

$$\mathbf{r} \neq \mathbf{0} \ \text{ such that } \ \mathcal{L}(\mathbf{r}) = \lambda \mathbf{r} \ \text{ for some scalar } \lambda \in \mathbb{C}.$$

I'll first show every such linear operator has at least one eigenvalue/eigenvector pair, $\lambda/\mathbf{r}$.

Given $\dim(\mathcal{V}) = n$, let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ denote a basis set for $\mathcal{V}$. As done on an earlier homework, write a vector $\mathbf{x} \in \mathcal{V}$ in terms of this basis to see

$$\mathbf{x} = x_1 \mathbf{b}_1 + \cdots + x_n \mathbf{b}_n \quad \Longrightarrow \quad \mathcal{L}(\mathbf{x}) = x_1 \mathcal{L}(\mathbf{b}_1) + \cdots + x_n \mathcal{L}(\mathbf{b}_n)$$

where $x_1, \ldots, x_n$ are n complex valued scalars. Also, because $\mathcal{L}(\mathbf{b}_k) \in \mathcal{V}$ for every index $1 \leq k \leq n$, there are n complex scalars $l_{1,k}, \ldots, l_{n,k}$ such that

$$\mathcal{L}(\mathbf{b}_k) = l_{1,k} \mathbf{b}_1 + \cdots + l_{n,k} \mathbf{b}_n \quad \Longrightarrow \quad \mathcal{L}(\mathbf{x}) = \mathbf{y} \equiv y_1 \mathbf{b}_1 + \cdots + y_n \mathbf{b}_n.$$

where the scalars $y_1, \ldots, y_n$ are determined via matrix multiplication

$$\vec{y} = L\vec{x} \ \text{ with } \ \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{C}^n, \ L = \begin{pmatrix} l_{1,1} & \cdots & l_{1,n} \\ \vdots & \ddots & \vdots \\ l_{n,1} & \cdots & l_{n,n} \end{pmatrix} \in \mathbb{C}^{n \times n}, \ \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{C}^n.$$

That is, once a basis for the vector space $\mathcal{V}$ is fixed, the linear operator $\mathcal{L}$ is uniquely identified by a matrix $L \in \mathbb{C}^{n \times n}$.

For the $n \times n$ matrix L above, let's see if we can show there is a nonzero vector $\vec{r} \in \mathbb{C}^n$ with an associated scalar $\lambda \in \mathbb{C}$ such that

$$L\vec{r} = \lambda \vec{r} \quad \Longleftrightarrow \quad (L - \lambda I)\vec{r} = \vec{0}.$$

Call $L_\lambda \equiv L - \lambda I$ and recall from page 7 on Homework 6 we've shown

$$L_\lambda \text{ has independent columns} \Longleftrightarrow \det(L_\lambda) \neq 0,$$

which is equivalent to saying

$$L_\lambda \text{ has dependent columns} \Longleftrightarrow \det(L_\lambda) = 0.$$

On the top of page 6 on Homework 7 we've also shown

$$\det(L_\lambda) = p(\lambda),$$

where $p(\lambda)$ is a degree n "monic" polynomial, the so-called characteristic polynomial. The *fundamental theorem of algebra* says this has $l \geq 1$ distinct (possibly complex) roots and can be factored as follows

$$p(\lambda) = (-1)^n (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_l)^{m_l} \ \text{ where } \ m_1 + \cdots + m_l = n,$$

and where the m's are all positive integers called the multiplicity of the associated characteristic root. For example, suppose

$$p(\lambda) = \lambda^4 - 9\lambda^2 + 4\lambda + 12 \quad \Longrightarrow \quad p(\lambda) = (\lambda+1)^1(\lambda-2)^2(\lambda+3)^1.$$

The root $\lambda = -1$ has multiplicity 1, $\lambda = 2$ has multiplicity 2 and $\lambda = -3$ has multiplicity 1. Note that the sum of the multiplicities here equals 4 which must be equal to the degree of $p(\lambda)$.

For λ equal to any one of the characteristic roots, say λ_*, we have $\det(L_{\lambda_*}) = 0$ which as stated above implies the columns of L_{λ_*} must be linearly dependent. This says there are n (also possibly complex) scalars $r_1, \ldots, r_n$ which are not all zero such that $L_{\lambda_*}\vec{r}_* = \vec{0}$ where $\vec{r}_* = (r_1 \cdots r_n)^T \neq \vec{0}$. That is, for any $n \times n$ matrix L, there is a vector $\vec{r}_* \in \mathbb{C}^n$ and a scalar $\lambda_* \in \mathbb{C}$ such that

$$\vec{r}_* \neq \vec{0} \ \text{ and satisfying } \ L\vec{r}_* = \lambda_*\vec{r}_*.$$

Returning to the linear operator introduced earlier, i.e. $\mathcal{L} : \mathcal{V} \to \mathcal{V}$, set

$$\mathbf{r}_* \equiv r_1\mathbf{b}_1 + \cdots r_n\mathbf{b}_n \neq \mathbf{0}$$

and note that

$$\begin{aligned} \mathcal{L}(\mathbf{r}_*) &= (L\vec{r}_*)_1\mathbf{b}_1 + \cdots + (L\vec{r}_*)_n\mathbf{b}_n \\ &= (\lambda_* r_1)\,\mathbf{b}_1 + \cdots + (\lambda_* r_n)\,\mathbf{b}_n = \lambda_*\mathbf{r}_*, \end{aligned}$$

which shows $\mathbf{r}_* \in \mathcal{V}$ is an eigenvector for $\mathcal{L}$.

Here's an illustrative example. Consider the vector space $\mathcal{V} \equiv \text{span}\{1, \cos(x), \sin(x)\}$ and a linear operator $\mathcal{L}(f) \equiv d^2f/dx^2 + df/dx$. Check that $\mathcal{L} : \mathcal{V} \to \mathcal{V}$ and that $\{1, \cos(x), \sin(x)\}$ is in fact a basis for $\mathcal{V}$. Let's determine $\mathcal{L}$'s matrix L with respect to this basis.

$$f = \alpha_0\,1 + \alpha_1\cos(x) + \alpha_2\sin(x) \quad \Longrightarrow \quad \mathcal{L}(f) = (-\alpha_1 + \alpha_2)\cos(x) + (-\alpha_1 - \alpha_2)\sin(x)$$

and from this see

$$\mathcal{L}(f) = \beta_0 1 + \beta_1\cos(x) + \beta_2\sin(x) \ \text{ where } \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix} = \begin{pmatrix}0 & 0 & 0\\ 0 & -1 & 1\\ 0 & -1 & -1\end{pmatrix}\begin{pmatrix}\alpha_0\\ \alpha_1\\ \alpha_2\end{pmatrix}.$$

L is the 3×3 matrix on the right above. Compute the determinant of $L - \lambda I$ and factor

$$\det(L - \lambda I) = -\lambda\left((\lambda+1)^2 + 1\right) = -\lambda(\lambda+1+i)(\lambda+1-i).$$

Read off L's eigenvalues: $\lambda = 0$, $\lambda = -1-i$ and $\lambda = -1+i$. Next, compute that the eigenvalue $\lambda = 0$ has eigenvector $\vec{r} = (1\ 0\ 0)^T$, the eigenvalue $\lambda = -1-i$ has eigenvector $\vec{r} = (0\ +i\ 1)^T$ and the eigenvalue $\lambda = -1+i$ has eigenvector $\vec{r} = (0\ -i\ 1)^T$. Use the eigenvectors for the matrix L just computed to find the eigenvectors for the original linear operator $\mathcal{L}$,

$$\begin{aligned} \lambda = 0 \quad &\Longrightarrow \quad f(x) = 1, \\ \lambda = -1 - i \quad &\Longrightarrow \quad f(x) = +i\cos(x) + \sin(x), \\ \lambda = -1 + i \quad &\Longrightarrow \quad f(x) = -i\cos(x) + \sin(x). \end{aligned}$$

An important question I want to answer now is the following. Suppose $\mathcal{L} : \mathcal{V} \to \mathcal{V}$ is a given linear operator. If $\mathcal{L}$ has a matrix representation

$$L \in \mathbb{C}^{n\times n} \text{ with respect to a basis with } \mathcal{V} = \operatorname{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_n\},$$

and another matrix representation

$$L' \in \mathbb{C}^{n\times n} \text{ with respect to a different basis with } \mathcal{V} = \operatorname{span}\{\mathbf{b}'_1, \ldots, \mathbf{b}'_n\},$$

how are L and L' related? The answer is the well known *change of basis formula*

$$L' = B^{-1}LB \text{ for some invertible } B \in \mathbb{C}^{n\times n}.$$

A consequence of this fact is

$$\begin{aligned} p'(\lambda) &= \det(L' - \lambda I) = \det(B^{-1}LB - \lambda I) \\ &= \det(B^{-1}(L - \lambda I)\, B) = \det(B^{-1})\det(L - \lambda I)\det(B) \\ &= \frac{1}{\det(B)}\det(L - \lambda I)\det(B) = \det(L - \lambda I) = p(\lambda). \end{aligned}$$

So we see that L and L' share the same characteristic polynomial, and therefore they share the same characteristic roots. This tells us the eigenvalues of the linear operator $\mathcal{L} : \mathcal{V} \to \mathcal{V}$ are independent of any particular basis used to define its matrix representation. Moreover, if we have

$$L'\vec{r}' = \lambda\vec{r}' \text{ we see that } B^{-1}LB\,\vec{r}' = \lambda\vec{r}' \quad \Longrightarrow \quad L(B\,\vec{r}') = \lambda(B\,\vec{r}').$$

That is, if $\vec{r}'$ is an eigenvector for L' then $B\vec{r}'$ is an eigenvector for L.

Now I'm going to derive the change of basis formula stated above. Let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\{\mathbf{b}'_1, \ldots, \mathbf{b}'_n\}$ denote two bases of $\mathcal{V}$. An arbitrary $\mathbf{x} \in \mathcal{V}$ can be decomposed in each

$$\begin{aligned} \mathbf{x} &= x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n, \\ \mathbf{x} &= x'_1\mathbf{b}'_1 + \cdots + x'_n\mathbf{b}'_n. \end{aligned}$$

Since for each $1 \le k \le n$ we have $\mathbf{b}'_k \in \mathcal{V}$, there are scalars $b_{1,k}, \ldots, b_{n,k}$ such that

$$\mathbf{b}'_k = b_{1,k}\mathbf{b}_1 + \cdots + b_{n,k}\mathbf{b}_n.$$

Insert this into the second decomposition

$$\mathbf{x} = \sum_{k=1}^{n} x'_k \left(\sum_{i=1}^{n} b_{i,k}\mathbf{b}_i \right) = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} b_{i,k}\, x'_k \right) \mathbf{b}_i,$$

and equate to the first to find

$$\vec{x} = B\vec{x}', \quad \text{where} \quad \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \; \vec{x}' = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} \quad \text{and } B \in \mathbb{C}^{n\times n} \text{ with } B_{i,j} = b_{i,j}\,.$$

Next, similar to what I just did, write

$$\mathcal{L}(\mathbf{x}) = y_1\mathbf{b}_1 + \cdots + y_n\mathbf{b}_n,$$
$$\mathcal{L}(\mathbf{x}) = y_1'\mathbf{b}_1' + \cdots + y_n'\mathbf{b}_n'.$$

and again insert $\mathbf{b}_k' = b_{1,k}\mathbf{b}_1 + \cdots + b_{n,k}\mathbf{b}_n$ into the second expression for $\mathcal{L}(\mathbf{x})$

$$\mathcal{L}(\mathbf{x}) = \sum_{k=1}^{n} y_k' \left(\sum_{i=1}^{n} b_{i,k}\,\mathbf{b}_i \right) = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} b_{i,k}\,y_k' \right) \mathbf{b}_i.$$

Equate this with first expression for $\mathcal{L}(\mathbf{x})$ to find

$$\vec{y} = B\vec{y}' \quad \text{where} \quad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \; \vec{y}' = \begin{pmatrix} y_1' \\ \vdots \\ y_n' \end{pmatrix} \quad \Longrightarrow \quad \vec{y}' = B^{-1}\vec{y}.$$

(Convince yourself that B must be invertible. This follows from the fact that it represents a change of basis matrix.) But $\vec{y}' = L'\vec{x}'$ and $\vec{y} = L\vec{x}$. These and the fact that $\vec{x} = B\vec{x}'$ established above gives

$$L'\,\vec{x}' = B^{-1}L\,\vec{x} = B^{-1}LB\,\vec{x}' \quad \Longrightarrow \quad L' = B^{-1}LB,$$

where the matrix form of the change of basis formula on the right hand side above follows from the left hand identity because the vector $\vec{x}'$ is arbitrary.

Homework 8 exercise 3b gives you a 4×4 matrix

$$\begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{pmatrix} \quad \begin{aligned} &\text{which has characteristic polynomial } p(\lambda) = (\lambda+1)^2(\lambda-3)^2 \\ &\Longrightarrow \quad \lambda_1 = -1 \text{ with multiplicity } m_1 = 2, \\ &\qquad\;\; \lambda_2 = \;\; 3 \text{ with multiplicity } m_2 = 2. \end{aligned}$$

Here you should have determined that the eigenvalue λ_1 has two independent eigenvectors associated to it, and so does λ_2. Specifically, the two dimensional *eigenspaces* are

$$\mathcal{E}_{\lambda=-1} = \operatorname{span}\{(1\ {-1}\ 0\ 0)^T,\ (0\ 0\ 1\ {-1})^T\},$$
$$\mathcal{E}_{\lambda=3} = \operatorname{span}\{(1\ 1\ 0\ 0)^T,\ (0\ 0\ 1\ 1)^T\}.$$

On the other hand, the 2×2 matrix

$$\begin{pmatrix} -1 & 4 \\ -1 & 3 \end{pmatrix} \quad \begin{aligned} &\text{has characteristic polynomial } p(\lambda) = (\lambda-1)^2 \\ &\Longrightarrow \quad \lambda_1 = 1 \text{ with multiplicity } m_1 = 2. \end{aligned}$$

But for this example we only have a one dimensional associated eigenspace

$$\mathcal{E}_{\lambda=1} = \operatorname{span}\{(2\ 1)^T\}.$$

I'm going to finish these notes by proving the following important result.

If λ_* is an eigenvalue of $\mathcal{L}:\mathcal{V}\to\mathcal{V}$ with multiplicity m_*
then $1\le \dim(\mathcal{E}_{\lambda_*})\le m_*$.

As illustrated above, for eigenvalues having multiplicity greater than one, the dimension of the associated eigenspace may or may not be equal to the eigenvalue's multiplicity. In the latter case conclude that $\mathcal{V}$ will <u>not</u> have a basis of eigenvectors of $\mathcal{L}$.

Let's prove the result. Suppose $\mathcal{L}:\mathcal{V}\to\mathcal{V}$ has an eigenvalue λ_* with multiplicity m_*. Let $\{\mathbf{r}_1,\ldots,\mathbf{r}_{n_*}\}$ denote a basis for the associated eigenspace $\mathcal{E}_{\lambda_*}$. We've shown earlier that $n_*=\dim(\mathcal{E}_{\lambda_*})\ge 1$. In order to show $\dim(\mathcal{E}_{\lambda_*})\le m_*$, determine vectors $\mathbf{s}_{n_*+1},\ldots,\mathbf{s}_n$ so that

$$\{\mathbf{r}_1,\ldots,\mathbf{r}_{n_*},\mathbf{s}_{n_*+1},\ldots,\mathbf{s}_n\}\ \text{ is a basis for } \mathcal{V}.$$

Now let's figure out $\mathcal{L}$'s matrix with respect to this basis. For $1\le k\le n_*$ we have

$$\mathcal{L}(\mathbf{r}_k)=\lambda_*\mathbf{r}_k,$$

and for $n_*+1\le k\le n$, since $\mathcal{L}(\mathbf{s}_k)\in\mathcal{V}$, there are scalars $l_{1,k},\ldots,l_{n,k}$ such that

$$\mathcal{L}(\mathbf{s}_k)=l_{1,k}\,\mathbf{r}_1+\cdots+l_{n,k}\,\mathbf{s}_n.$$

Let

$$\mathbf{x}=x_1\,\mathbf{r}_1+\cdots+x_n\,\mathbf{s}_n \quad\Longrightarrow\quad \mathcal{L}(\mathbf{x})=x_1\mathcal{L}(\mathbf{r}_1)+\cdots+x_n\mathcal{L}(\mathbf{s}_n),$$

and use $\mathcal{L}(\mathbf{r}_k)$ and $\mathcal{L}(\mathbf{s}_k)$ given above, together with Σ-notation, to write

$$\begin{aligned}\mathcal{L}(\mathbf{x})&=\sum_{k=1}^{n_*}x_k\lambda_*\mathbf{r}_k+\sum_{k=n_*+1}^{n}x_k\left(\sum_{i=1}^{n_*}l_{i,k}\,\mathbf{r}_i+\sum_{i=n_*+1}^{n}l_{i,k}\,\mathbf{s}_i\right)\\&=\sum_{i=1}^{n_*}\left(\lambda_*x_i+\sum_{k=n_*+1}^{n}l_{i,k}\,x_k\right)\mathbf{r}_i+\sum_{i=n_*+1}^{n}\left(\sum_{k=n_*+1}^{n}l_{i,k}\,x_k\right)\mathbf{s}_i.\end{aligned}$$

Equate this to $\mathcal{L}(\mathbf{x})=y_1\mathbf{r}_1+\cdots+y_n\mathbf{s}_n$ to find

$$\vec{y}=L\,\vec{x}\ \text{ where } L \text{ can be written in block form } L=\begin{pmatrix}\lambda_*I & L_{top}\\ 0 & L_{bot}\end{pmatrix},$$

and where here I've used notation I for the $n_*\times n_*$ identity, 0 for the $(n-n_*)\times n_*$ block of zeros, L_{top} for a $n_*\times(n-n_*)$ block and L_{bot} for a $n_*\times n_*$ block, both filled with values $l_{i,j}$. Successively cofactor $L-\lambda I$ along columns 1 through n_* to find

$$\det(L-\lambda I)=(\lambda_*-\lambda)^{n_*}\det(L_{bot}-\lambda I)=p(\lambda),$$

where the I's here are appropriately sized. However, as pointed out earlier, the FTA says the characteristic polynomial $p(\lambda)$ can be factored by its $l\ge 1$ distinct roots

$$p(\lambda)=(-1)^n(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_l)^{m_l}.$$

Therefore, $(\lambda_* - \lambda)^{n_*}$ must divide $(\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_l)^{m_l}$ which is only possible when

$$\dim(\mathcal{E}_{\lambda_*}) = n_* \leq m_*.$$

That's it.