Chapter 2

Invariant Subspaces

Reminder: Unless explicitly stated we are talking about finite dimensional vector spaces, and linear transformations and operators between finite dimensional vector spaces. However, we will from time to time explicitly discuss the infinite dimensional case. Similarly, although much of the theory holds for vector spaces over some field, we focus (for practical purposes) on vector spaces of the real of complex field.

Definition 2.1. *Let V be a vector space (real or complex) and $\mathcal{L} : V \to V$ be a linear* **operator** *over V. We say a vector space $W \subseteq V$ is an invariant subspace of $\mathcal{L}$ if for every $\mathbf{w} \in W$, $\mathcal{L}\mathbf{w} \in W$ (we also write $\mathcal{L}W \subseteq W$).*

Note that V, $\{0\}$ (the set containing only the zero vector in V), Null($\mathcal{L}$), and Range($\mathcal{L}$) are all invariant subspaces of $\mathcal{L}$.

Exercise 2.1. *Prove the statement above.*

Theorem 2.2. *Let V and $\mathcal{L}$ be as before, and let W_1, W_2, W_3 be invariant subspaces of $\mathcal{L}$. Then (1) $W_1 + W_2$ is an invariant subspace of $\mathcal{L}$, (2) $(W_1 + W_2) + W_3 = W_1 + (W_2 + W_3)$, (3) $W_1 + \{0\} = \{0\} + W_1$.*

Exercise 2.2. *Prove theorem 2.2. (The set of all invariant subspaces of a linear operator with the binary operation of the sum of two subspaces is a semigroup and a monoid).*

Exercise 2.3. *Prove that the sum of invariant subspaces is commutative.*

If an invariant subspace of a linear operator, $\mathcal{L}$, is one-dimensional, we can say a bit more about it, and hence we have a special name for non-zero vectors in such a space.

Definition 2.3. *We call a nonzero vector* $\mathbf{v} \in V$ *an eigenvector if* $\mathcal{L}\mathbf{v} = \lambda\mathbf{v}$ *for some scalar* λ. *The scalar* λ *is called an eigenvalue. The (ordered) pair,* $(\mathbf{v}, \lambda)$ *is called an eigenpair. (Note: this is the generalization from Chapter 1 to cover any linear operator.)*

Exercise 2.4. *Given a one-dimensional invariant subspace, prove that any nonzero vector in that space is an eigenvector and all such eigenvectors have the same eigenvalue.*

Vice versa the span of an eigenvector is an invariant subspace. From Theorem 2.2 then follows that the span of a set of eigenvectors, which is the sum of the invariant subspaces associated with each eigenvalue, is an invariant subspace.

Example 2.1. *As mentioned before, the matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *defines a linear operator over* $\mathbb{R}^n$. *Consider the real matrix* $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ *and vector* $\mathbf{v} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$. *Then* $\mathbf{A}\mathbf{v} = \mathbf{0} = 0 \cdot \mathbf{v}$. *Hence* $\mathbf{v}$ *is an eigenvector of* $\mathbf{A}$ *and* 0 *is an eigenvalue of* $\mathbf{A}$. *The pair* $(\mathbf{v}, 0)$ *is an eigenpair. Note that a matrix with an eigenvalue* 0 *is singular.*

Example 2.2. *The following is an example for an infinite dimensional subspace. Let* $C^\infty[a,b]$ *be the set of all infinitely differentiable real functions on the (closed) interval* $[a,b]$. *We define the addition of two functions* $f, g \in C^\infty[a,b]$ *by* $h = f + g$ *is the function* $h(x) \equiv (f+g)(x) = f(x) + g(x)$ *(for* $x \in [a,b]$*), and for all* $\alpha \in \mathbb{R}$ *and* $f \in C^\infty[a,b]$ *we define* $h = \alpha f$ *by* $h(x) \equiv (\alpha f)(x) = \alpha f(x)$. *Then* $C^\infty[a,b]$ *with this definition of scalar multiplication and vector addition is a vector space (show this).*

Let $\mathcal{L}$ *be defined by* $\mathcal{L}u = u_{xx}$ *for* $u \in C^\infty[a,b]$. *Then* $\mathcal{L}$ *is a linear operator over* $C^\infty[a,b]$ *and* $(\sin \omega x, -\omega^2)$ *is an eigenpair for any* $\omega \in \mathbb{R}$.

We have $\mathcal{L}v = \lambda v \Leftrightarrow \mathcal{L}v - \lambda v = 0$. We can rewrite the right-hand side of the last expression as $(\mathcal{L} - \lambda\mathcal{I})v = 0$. Since $v \neq 0$, the operator $(\mathcal{L} - \lambda\mathcal{I})$ is singular (i.e. not invertible).

Although in some cases the eigenvalues and eigenvectors of a linear operator are clear by inspection, in general we need some procedure to find them. *As all linear operators over finite dimensional subspaces can be represented as matrices*, all we need is a systematic procedure to find the eigenvalues and eigenvectors of matrices (we get our answer for the original operator by the corresponding linear combinations of basis vectors). Remember that the function that maps linear transformations between finite dimensional vector spaces given bases for the spaces to matrices is an isomorphism (i.e. an invertible linear transformation).

Given a basis $\mathcal{B}$, let $\mathbf{A} = [\mathcal{L}]_{\mathcal{B}}$. Using the linearity of the standard transformation from operator to matrix (given a basis), we also have $\mathbf{A} - \lambda\mathbf{I} = [\mathcal{L} - \mathcal{I}]_{\mathcal{B}}$

and the matrix $\mathbf{A} - \lambda\mathbf{I}$ must be singular. Hence, $\mathbf{det}(\mathbf{A} - \lambda\mathbf{I}) = 0$

Definition 2.4. *The polynomial* $\mathbf{det}(\mathbf{A} - \lambda\mathbf{I})$ *is called the characteristic polynomial of* $\mathbf{A}$*. The (polynomial) equation* $\mathbf{det}(\mathbf{A} - \lambda\mathbf{I}) = 0$ *in* λ *is called the characteristic equation for* $\mathbf{A}$ *(and for* $\mathcal{L}$*!). The eigenvalues of* $\mathbf{A}$ *(and of* $\mathcal{L}$*) are the roots of this characteristic equation. The multiplicity of an eigenvalue as a root of this equation is called the algebraic multiplicity of that eigenvalue.*

IMPORTANT!! It may seem from the above that eigenvalues depend on the choice of basis, but this is not the case! To show that this is the case, we need only pick two different bases for V, and show that the corresponding matrix of the transformation for one basis is *similar to* the matrix of the transformation for the other basis. Let $\mathbf{A} = [\mathcal{L}]_{\mathcal{B}}$. Let $\mathcal{C}$ also be a basis for V and define $\mathbf{B} = [\mathcal{L}]_{\mathcal{C}}$. Furthermore, let $\mathbf{X} = [\mathcal{I}]_{\mathcal{C} \leftarrow \mathcal{B}}$. Then we have $\mathbf{A} = \mathbf{X}^{-1}\mathbf{B}\mathbf{X}$ (recall that this is called a similarity transformation between $\mathbf{A}$ and $\mathbf{B}$).

Theorem 2.5. $\mathbf{A}$ *and* $\mathbf{B}$ *as defined above have the same eigenvalues.*

Proof. From $\mathbf{A} = \mathbf{X}^{-1}\mathbf{B}\mathbf{X}$ we see that $\mathbf{A} - \lambda\mathbf{I} = \mathbf{X}^{-1}\mathbf{B}\mathbf{X} - \lambda\mathbf{I} = \mathbf{X}^{-1}(\mathbf{B} - \lambda\mathbf{I})\mathbf{X}$. Hence $\mathbf{det}(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{det}(\mathbf{X}^{-1})\mathbf{det}(\mathbf{B} - \lambda\mathbf{I})\mathbf{det}(\mathbf{X})$ and $\mathbf{det}(\mathbf{A} - \lambda\mathbf{I}) = 0 \Leftrightarrow \mathbf{det}(\mathbf{B} - \lambda\mathbf{I}) = 0$. □

So, it is indeed fine to define the eigenvalues of an operator over a vector space *using any basis for the vector space and its corresponding matrix*. In fact, as the characteristic polynomial of an $n \times n$ matrix has leading term of $(-1)^n\lambda^n$ (check this), the characteristic polynomials of $\mathbf{A}$ and $\mathbf{B}$ are equal. So, we can call this polynomial the characteristic polynomial of $\mathcal{L}$ without confusion. Note that the proof above does not rely on the fact that $\mathbf{A}$ and $\mathbf{B}$ are representations of the (same) linear operator $\mathcal{L}$, only that $\mathbf{A}$ is obtained from a similarity transformation of $\mathbf{B}$. So, this is a general result for similar matrices.

We say that *similarity transformations preserve eigenvalues*. The standard methods for computing eigenvalues of any but the smallest matrices are in fact based on sequences of cleverly chosen similarity transformations the limit of which is an uppertriangular (general case) or diagonal matrix (Hermitian or real symmetric case). Take a course in Numerical Linear Algebra for more details!

Eigenvectors of matrices are not preserved under similarity transformations, but they change in a straightforward way. Let $\mathbf{A}$ and $\mathbf{B}$ be as above and let $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Then, $\mathbf{X}^{-1}\mathbf{B}\mathbf{X}\mathbf{v} = \lambda\mathbf{v} \Leftrightarrow \mathbf{B}(\mathbf{X}\mathbf{v}) = \lambda(\mathbf{X}\mathbf{v})$, so $\mathbf{X}\mathbf{v}$ is an eigenvector of $\mathbf{B}$ corresponding to λ.

If we are interested in computing eigenvectors of an operator $\mathcal{L}$, then again the choice of basis is irrelevant (at least in theory; in practice, it can matter a lot). Let $\mathbf{A}$ and $\mathbf{B}$ be representations of $\mathcal{L}$ with respect to the bases $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_n\}$ as above, with the change of coordinate matrix $\mathbf{X} = [\mathcal{I}]_{\mathcal{C} \leftarrow \mathcal{B}}$. By some procedure we obtain $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$, which corresponds to $\mathbf{B}(\mathbf{X}\mathbf{u}) = \lambda(\mathbf{X}\mathbf{u})$. Define $\mathbf{y} = \sum_{i=1}^{n} u_i\mathbf{v}_i$ (so that $\mathbf{u} = [\mathbf{y}]_{\mathcal{B}}$), then we have $[\mathcal{L}\mathbf{y}]_{\mathcal{B}} = [\lambda\mathbf{y}]_{\mathcal{B}}$, which implies (by the standard isomorphism) that $\mathcal{L}\mathbf{y} = \lambda\mathbf{y}$. However, we also have

$[\mathbf{y}]_{\mathcal{C}} = [\mathcal{I}]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{y}]_{\mathcal{B}} = \mathbf{Xu}$. This gives $[\mathcal{L}\mathbf{y}]_{\mathcal{C}} = [\lambda\mathbf{y}]_{\mathcal{C}} \Leftrightarrow \mathbf{B}(\mathbf{Xu}) = \lambda(\mathbf{Xu})$. So, computing eigenvectors of $\mathbf{B}$ leads to the same eigenvectors for $\mathcal{L}$ as using $\mathbf{A}$.

Theorem 2.6. *A linear operator, $\mathcal{L}$, is diagonalizable if and only if there is a basis for the vector space with each basis vector an eigenvector of $\mathcal{L}$. A matrix $\mathbf{A}^{n\times n}$ is diagonalizable if and only if there is a basis for $\mathbb{R}^n$ (respectively $\mathbb{C}^n$) that consists of eigenvectors[5] of $\mathbf{A}$.*

We often say $\mathbf{A}$ diagonalizable if there exists an invertible matrix $\mathbf{U}$ such that $\mathbf{U}^{-1}\mathbf{AU}$ is diagonal. But clearly, if we set $\mathbf{D} = \mathbf{U}^{-1}\mathbf{AU}$, we rearrage, we get

$$\begin{aligned}\mathbf{AU} = \mathbf{UD} &\Rightarrow [\mathbf{Au}_1, \ldots, \mathbf{Au}_n] = [d_{11}\mathbf{u}_1, d_{22}\mathbf{u}_2, \ldots, d_{nn}\mathbf{u}_n] \\ &\Rightarrow \mathbf{Au}_i = d_{ii}\mathbf{u}_i, i = 1, \ldots, n\end{aligned}$$

and since the $\mathbf{u}_i$ cannot be zero (since $\mathbf{U}$ was assumed invertible), the columns of $\mathbf{U}$ must be eigenvectors and elements of $\mathbf{D}$ must be eigenvalues.

Furthermore, you should make sure you are able to show that if $\mathcal{L}(\mathbf{x}) = \mathbf{Ax}$ and $\mathbf{A}$ is diagonalizable, then if you use the eigenvectors for the (only) basis on the right and left of the linear transformation/matrix transformation picture, you find that the matrix of the transformation is precisely the diagonal matrix containing the eigenvalues.

As mentioned in the previous chapter, a matrix may not be diagonalizable. We now consider similarity transformations to block diagonal as an alternative. (Refer also to definition of block diagonal in Chapter 1)

Definition 2.7. *Let $\mathbf{A}$ be a complex or real $n \times n$ matrix and let the numbers m_1, m_2, …, m_s be given such that $1 \le m_i \le n$ for $i = 1, \ldots, s$ and $\sum_{i=1}^{s} m_i = n$. Furthermore, for $i = 1, \ldots, s$ let $f_i = 1 + \sum_{j=1}^{i-1} m_j$ (where $f_1 = 1$), $\ell_i = \sum_{j=1}^{i} m_j$, and $P_i = \{f_i, \ldots, \ell_i\}$. We say that $\mathbf{A}$ is a (complex or real) block diagonal matrix with s diagonal blocks of sizes m_1, m_2, …, m_s, if its coefficients satisfy $a_{r,c} = 0$ if r and c are not elements of the same index set P_i. (See note below).*

Note that $\mathbf{A}$ is a block diagonal matrix if the coefficients outside the diagonal blocks are all zero. The first diagonal block is $m_1 \times m_1$, the second block is $m_2 \times m_2$, and so on. The first coefficient of block i has index f_i; the last coefficient of block i has index ℓ_i.

Theorem 2.8. *Let $\mathcal{L}$ be a linear operator over V, with $\dim(V) = n$, and with invariant subspaces V_1, …, V_s, such that $V_1 \bigoplus V_2 \bigoplus \cdots \bigoplus V_s = V$. Further, let there be bases $\mathbf{v}_1^{(i)}$, …, $\mathbf{v}_{m_i}^{(i)}$ for each V_i (for $i = 1, \ldots, s$), and define the ordered set $\mathcal{B} = \{\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_{m_1}^{(1)}, \mathbf{v}_1^{(2)}, \ldots, \mathbf{v}_{m_2}^{(2)}, \ldots, \mathbf{v}_{m_s}^{(s)}\}$ (i.e., $\mathcal{B}$ is basis for V). Then*

[5] Here again, if the matrix is real, we must be careful to specify whether or not we are considering the matrix transformation as a map on $\mathbb{R}^n$ or on $\mathbb{C}^n$. If the former and the eigenpairs are not all real, then we are forced to conclude it is not diagonalizable with respect to $\mathbb{R}^n$, even though it may be if we take it with respect to $\mathbb{C}^n$.

$\mathbf{A} = [\mathcal{L}]_{\mathcal{B}}$ is block-diagonal.

Proof. (This is a sketch of the proof.) As each V_i is an invariant subspace, $\mathcal{L}\mathbf{v}_j^{(i)} \in V_i$. Hence, $\mathcal{L}\mathbf{v}_j^{(i)} = \sum_{k=1}^{m_i} \alpha_k \mathbf{v}_k^{(i)}$. These coefficients correspond to columns $f_i, f_i + 1, \ldots, \ell_i$ and the same rows. So, only the coefficients in the diagonal blocks of $\mathbf{A}$, $\mathbf{A}_{1,1}^{m_1 \times m_1}, \ldots, \mathbf{A}_{s,s}^{m_s \times m_s}$ can have nonzero coefficients. ▯

Exercise 2.5. *Write the proof in detail.*

Note that a block-diagonal matrix (as a linear operator over $\mathbb{R}^n$ or $\mathbb{C}^n$) reveals invariant subspace information quite trivially. Vectors with zero coefficients except corresponding to one diagonal block obviously lie within an invariant subspace. Moreover, bases for the invariant subspaces can be trivially found. Finally, finding eigenvalues and eigenvectors (generalized eigenvectors) for block diagonal matrices is greatly simplified. For this reason we proceed by working with matrices. However, the standard isomorphism between the n-dimensional vector space V and $\mathbb{C}^n$) (or $\mathbb{R}^n$) given a basis B for V guarantees that the invariant subspaces we find for the matrix $[\mathcal{L}]_B$ correspond to invariant subspaces for $\mathcal{L}$, as we observe in the following theorem.

Theorem 2.9. *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a block diagonal matrix with s diagonal blocks of sizes $m_1, m_2, \ldots, m_s$. Define the integers $f_i = 1 + \sum_{j=1}^{i-1} m_j$ (where $f_1 = 1$) and $\ell_i = \sum_{j=1}^{i} m_j$ for $i = 1, \ldots, s$. Then the subspaces (of $\mathbb{C}^n$)*

$$V_i = \{x \in \mathbb{C}^n | x_j = 0 \text{ for all } j < f_i \text{ and } j > \ell_i\} = \operatorname{Span}(\mathbf{e}_{f_i}, \mathbf{e}_{f_i+1}, \ldots, \mathbf{e}_{\ell_i})$$

for $i = 1, \ldots, s$ are invariant subspaces of A.

Proof. The proof is left as an exercise. ▯

Exercise 2.6. *Prove Theorem 2.9.*

Using the previous theorem we can also make a statement about invariant subspaces of $\mathcal{L}$ if the matrix representing $\mathcal{L}$ with respect to a particular basis is block diagonal.

Theorem 2.10. *Let $\mathcal{L}$ be a linear operator over V, with $\dim(V) = n$ and let the ordered set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V. Furthermore, let $\mathbf{A} = [\mathcal{L}]_{\mathcal{B}}$ be block-diagonal with block sizes $m_1, m_2, \ldots, m_s$ (in that order). Let f_i and ℓ_i for $i = 1, \ldots, s$ be as defined in Definition 2.7.*

Then the subspaces (of V) $V_1, \ldots, V_s$, defined by $V_i = \operatorname{Span}(\mathbf{v}_{f_i}, \mathbf{v}_{f_i+1}, \ldots, \mathbf{v}_{m_i})$ are invariant subspaces of $\mathcal{L}$ and $V_1 \bigoplus V_2 \bigoplus \cdots \bigoplus V_s = V$.

Proof. The proof is left as an exercise. ▯

Exercise 2.7. *Prove Theorem 2.10.*

Procedure to find invariant SS for $\mathcal{L}$

1. Get the matrix representation of $\mathcal{L}$ first. That is, pick a basis $\mathcal{B}$ for V and let $\mathbf{A} = [\mathcal{L}]_{\mathcal{B}}$.
2. Find an invertible matrix $\mathbf{S}$ such that $\mathbf{F} := \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ is block diagonal with blocks satisfying certain nice properties (nice is TBD).
3. The columns of $\mathbf{S}$ will span the invariant subspaces of $\mathbf{A}$ (group the columns according to the block sizes of $\mathbf{F}$, as we've been doing in the preceding discussion).
4. Use $\mathbf{S}$ and $\mathcal{B}$ to compose the invariant subspaces (in V) for $\mathcal{L}$.

Now, Step 2 is non-trivial, but we'll put this off and just assume it can be done. What remains is HOW do we finish Step 4? The following two theorem addresss this issue.

Theorem 2.11. *Let $\mathcal{L}$ be a linear operator over a complex n-dimensional vector space V, and let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a (arbitrary) basis for V. Let $\mathbf{A} = [\mathcal{L}]_{\mathcal{B}}$ and let $\mathbf{F} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$, for any invertible matrix $\mathbf{S} \in \mathbb{C}^{n\times n}$, be block diagonal with block sizes $m_1, m_2, \ldots, m_s$. Furthermore, let $f_i = 1 + \sum_{j=1}^{i-1} m_j$ ($f_1 = 1$) and $\ell_i = \sum_{j=1}^{i} m_j$, and let the ordered basis $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ be defined by $\mathbf{c}_i = \sum_{j=1}^{n} b_j s_{j,i}$ for $i = 1, \ldots n$. Then the spaces $V_i = \mathrm{Span}(\mathbf{c}_{f_i}, \ldots, \mathbf{c}_{\ell_i})$ are invariant subspaces of $\mathcal{L}$ and $V = V_1 \bigoplus V_2 \bigoplus \cdots \bigoplus V_s$.*

Proof. The proof is left as an exercise. Hint: consider $\mathcal{I}\circ\mathcal{L}\circ\mathcal{I}$, and note that with this definition of the $\mathcal{C}$ basis, $\mathbf{S} = [\mathcal{I}]_{\mathcal{C}\leftarrow\mathcal{B}}$. □

We will now consider in some more detail a set of particularly useful and revealing invariant subspaces that span the vector space. Hence, we consider block-diagonal matrices of a fundamental type. Most of the following results hold only for complex vector spaces, which are, from a practical point of view, the most important ones.

Next we provide some links between the number of distinct eigenvalues, the number of eigenvectors, and the number of invariant subspaces of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ (working over the complex field).

Theorem 2.12. *Let λ be an eigenvalue of $\mathbf{A}$ (of arbitrary algebraic multiplicity). There is at least one eigenvector $\mathbf{v}$ of $\mathbf{A}$ corresponding to λ.*

Proof. Since $\mathbf{A} - \lambda\mathbf{I}$ is singular, there is at least one nonzero vector $\mathbf{v}$ such that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$. □

Theorem 2.13. *Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues and let $\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{m_i}^{(i)}$ be* *independent eigenvectors associated with λ_i, for $i = 1, \ldots, k$. Then*

$$\{\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_{m_1}^{(1)}, \mathbf{v}_1^{(2)}, \ldots, \mathbf{v}_{m_2}^{(2)}, \ldots, \mathbf{v}_1^{(k)}, \ldots, \mathbf{v}_{m_k}^{(k)}\}$$

is an independent set.

Proof. Left as an exercise for the reader (it's in most linear algebra textbooks.) □

2.1 Toward a Direct Sum Decomposition

In general the above set of eigenvectors does not always give a direct sum decomposition for the vector space V. That is, it is not uncommon that we will not have a complete set of n linearly independent eigenvectors. So we need to think of another way to get what we're after. A complete set of independent eigenvectors for an n-dimensional vector space (n independent eigenvectors for an n-dimensional vector space) would give n 1-dimensional invariant subspaces, each the span of a single eigenvector. These 1-dimensional subspaces form a direct sum decomposition of the vector space V. Hence the representation of the linear operator in this basis of eigenvectors is a block diagonal matrix with each block size equal to one, that is, a diagonal matrix. In the following we try to get as close as possible to such a block diagonal matrix and hence to such a direct sum decomposition. We will aim for the following two properties. First, we want the diagonal blocks to be as small as possible and we want each diagonal block to correspond to a single eigenvector. The latter means that the invariant subspace corresponding to a diagonal block contains a single eigenvector. Second, we want to make the blocks as close to diagonal as possible. It turns out that bidiagonal, with a single nonzero diagonal right above (or below) the main diagonal (picture?) is the best we can do.

In the following discussion, **polynomials of matrices** or linear operators play an important role. Note that for a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ the matrices $\mathbf{A}^2 \in \mathbb{C}^{n\times n}$, $\mathbf{A}^3 \in \mathbb{C}^{n\times n}$, etc. are well-defined and that $\mathbb{C}^{n\times n}$ (over the complex field) is itself a vector space.

Hence, polynomials (of finite degree) in $\mathbf{A}$, expressed as $\alpha_0\mathbf{I}+\alpha_1\mathbf{A}+\cdots+\alpha_n\mathbf{A}^n$ are well-defined as well and, for fixed $\mathbf{A}$, *are elements of* $\mathbb{C}^{n\times n}$. Note there is a difference between a polynomial as a function of a free variable of a certain type and the evaluation of a polynomial for a particular choice of that variable.

Similarly, for linear operators over a vector space, $\mathcal{L} : V \to V$, composition of (or product of) the operator with itself, one or more (but finite) times, results in another linear operator over the same space, $(\mathcal{L})^m \equiv \mathcal{L}\circ(\mathcal{L}^{m-1})$ and $(\mathcal{L})^m : V \to V$. Indeed, the **set of all linear operators over a vector space** V, (often expressed as $L(V)$) is itself a vector space (over the same field). (Exercise: Prove this!)

A nice property of polynomials in a fixed matrix (or linear operator) is that they commute (in contrast to two general matrices).

Exercise 2.8. *Prove this for linear matrix polynomials* $\mathbf{A} - \lambda\mathbf{I}$ *and* $\mathbf{A} - \mu\mathbf{I}$ *and*