MATRIX ANALYSIS

A Quick Guide

Huan Bui

Colby College
Physics & Statistics
Class of 2021

February 28, 2019

Preface

Greetings,

Matrix Analysis: A Quick Guide to is compiled based on my MA353: Matrix Analysis notes with professor Leo Livshits. The sections are based on a number of resources: *Linear Algebra Done Right* by Axler, *A Second Course in Linear Algebra* by Horn and Garcia, *Matrices and Linear Transformations* by Cullen, *Matrices: Methods and Applications* by Barnett, *Problems and Theorems in Linear Algebra* by Prasolov, *Matrix Operations* by Richard Bronson, and professor Leo Livshits' own textbook (in the making). Prerequisites: some prior exposure to a first course in linear algebra.

The development of this text will come in layers. The first layer, one that I am working on during the course of S'19 MA353, will be an overview of the key topics listed in the table of contents. As the semester progresses, I will be constantly updating the existing notes, as well as adding prof. Livshits' problems and my solutions to the problems. The second layer will come after the course is over, when concepts will have hopefully "come together."

I will decide how much narrative I should put into the text as text is developed over the semester. I'm thinking that I will only add detailed explanations wherever I find fit or necessary for my own studies. I will most likely keep the text as condensed as I can.

Enjoy!

Contents

1 List of Special Matrices & Their Properties

1. **Hermitian/Self-adjoint**: $H = H^\dagger$. A Hermitian matrix is matrix that is equal to its own conjugate transpose:

$$H \text{ is Hermitian} \iff H_{ij} = \bar{H}_{ji}$$

Properties 1.1.

(a) H is Hermitian $\iff \langle w, Hv\rangle = \langle Hw, v\rangle$, where $\langle , \rangle$ denotes the inner product.

(b) H is Hermitian $\iff \langle v, Hv\rangle \in \mathbb{R}$.

(c) H is Hermitian $\iff$ it is *unitarily diagonalizable* with *real eigenvalues.*

Unitary: $U^*U = UU^* = I = U^\dagger U = UU^\dagger$. The real analogue of a unitary matrix is an orthogonal matrix. The following list contain the properties of U:

(a) U preserves the inner product:

$$\langle Ux, Uy\rangle = \langle x, y\rangle.$$

(b) U is normal: it commutes with $U^* = U^\dagger$.

(c) U is diagonalizable:

$$U = VDV^*,$$

where D is diagonal and unitary, and V is unitary.

(d) $|\det(U)| = 1$ (hence the real analogue to U is an orthogonal matrix)

(e) Its eigenspaces are orthogonal.

(f) U can be written as

$$U = e^{iH},$$

where H is a Hermitian matrix.

(g) Any square matrix with unit Euclidean norm is the average of two unitary matrices.

2. **Idempotent:** M idempotent $\iff M^2 = M$.

(a) Singularity: its number of independent rows (and columns) is less than its number of rows (and columns).

(b) When an idempotent matrix is subtracted from the identity matrix, the result is also idempotent.

"Proof".

$$[I - M][I - M] = I - M - M + M^2 = I - M - M + M = I - M.$$

□

(c) M is idempotent $\iff \forall n \in \mathbb{N}$, $A^n = A$.

(d) Eigenvalues: an idempotent matrix is always diagonalizable and its eigenvalues are either 0 or 1. (think "projection")

(e) Trace: the trace of an idempotent matrix equals the rank of the matrix and thus is always an integer. So

$$\operatorname{tr}(A) = \dim(\operatorname{Im} A).$$

3. **Nilpotent:** a nilpotent matrix is a square matrix N such that

$$N^k = 0$$

for some positive integer k. The smallest such k is sometimes called the **index** of N.

The following statements are equivalent:

(a) N is nilpotent.

(b) The minimal polynomial for N is x^k for some positive integer $k \leq n$.

(c) The characteristic polynomial for N is x^n.

(d) The only complex eigenvalue for N is 0.

(e) $\operatorname{tr} N^k = 0$ for all $k > 0$.

Properties 1.2.

(a) The degree of an $n \times n$ nilpotent matrix is always less than or equal to n.

(b) $\det N = \operatorname{tr}(N) = 0$.

(c) Nilpotent matrices are not invertible.

(d) The only nilpotent diagonalizable matrix is the zero matrix.

2 List of Operations

1. **Conjugate transpose** is what its name suggests.

2. **Classical adjoint/Adjugate/adjunct** of a square matrix is the transpose of its cofactor matrix.

3 List of Algorithms

4 Complex Numbers

4.1 A different point of view

We often think of complex numbers as

$$a + ib$$

where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$. While there is nothing "bad" about this way of thinking - in fact thinking of complex numbers as $a+ib$ allows us to very quickly and intuitively do arithmetics operations on them - a "matrix representation" of complex numbers can give us some insights on "what we actually do" when we perform complex arithmetics.

Let us think of

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

as a different representation of the same object - the same complex number "$a + ib$." Note that it does not make sense to say the matrix representation **equals** the complex number itself. But we shall see that a lot of the properties of complex numbers are carried into this matrix representation under interesting matricial properties.

$$\boxed{\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \sim a + ib}$$

First, let us break the matrix down:

$$a + ib = a \times 1 + i \times b \sim \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = aI + b\mathcal{I}.$$

Right away, we can make some "mental connections" between the representations:

$$I \sim 1$$
$$\mathcal{I} \sim i.$$

Now, we know that complex number multiplications commute:

$$(a + ib)(c + id) = (c + id)(a + ib).$$

Matrix multiplications are not commutative. So, we might wonder whether commutativity holds under the this new representation of complex numbers. Well, the answer is yes. We can readily verify that

$$(aI + b\mathcal{I})(cI + b\mathcal{I}) = (cI + b\mathcal{I})(aI + b\mathcal{I}).$$

How about additions? Let's check:

$$(a+ib)+(c+id)=(a+c)+i(b+d)\sim\begin{pmatrix}a+c & -(b+d)\\(b+d) & a+c\end{pmatrix}=\begin{pmatrix}a & -b\\b & a\end{pmatrix}+\begin{pmatrix}c & -d\\d & c\end{pmatrix}.$$

Ah! Additions work. So, the new representation of complex numbers seems to be working flawlessly. However, we have yet to gain any interesting insights into the connections between the representations. To do that, we have to look into changing the form of the matrix. First, let's see what conjugation does:

$$(a+ib)^*=a-ib\sim\begin{pmatrix}a & b\\-b & a\end{pmatrix}=\begin{pmatrix}a & -b\\b & a\end{pmatrix}^\top$$

Ah, so conjugation to a complex number in the traditional representation is the same as transposition in the matrix representations. What about the amplitude square? Let us call

$$M=\begin{pmatrix}a & -b\\b & a\end{pmatrix}.$$

We have

$$(a+ib)(a-ib)\sim\begin{pmatrix}a & -b\\b & a\end{pmatrix}\begin{pmatrix}a & b\\-b & a\end{pmatrix}=MM^\top=(a^2+b^2)I=\det(M)I$$

Interesting. But observe that if $\det(M)\neq 0$

$$\frac{1}{\det(M)}MM^\top=I.$$

This tells us that

$$M^\top=M^{-1},$$

where M^{-1} is the inverse of M, and, not surprisingly, it corresponds to the reciprocal to the complex number $a+ib$. We can readily show that

$$M^{-1}\sim(a+ib)^{-1}=\frac{1}{a^2+b^2}(a-ib).$$

Remember that we can also think of a complex number as a column vector:

$$c+id\sim\begin{pmatrix}c\\d\end{pmatrix}.$$

Let us look back at complex number multiplication under matrix representation:

$$(a+ib)(c+id)=(ac-bd)+i(bc+ad)\sim\begin{pmatrix}a & -b\\b & a\end{pmatrix}\begin{pmatrix}c\\d\end{pmatrix}=\begin{pmatrix}ac-bd\\bc+ad\end{pmatrix}.$$

Multiplication actually works in this "mixed" way of representing complex numbers as well. Now, observe that what we just did was performing a linear transformation on a vector in $\mathbb{R}^2$. It is always interesting to look at the geometrical interpretation of this transformation. To do this, let us call N the "normalized" version of M:

$$N = \frac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

We immediately recognize that N is an orthogonal matrix. This means N is an orthogonal transformation (length preserving). Now, it is reasonable to define

$$\cos\theta = \frac{a}{\sqrt{a^2+b^2}}$$
$$\sin\theta = \frac{b}{\sqrt{a^2+b^2}}.$$

We can write N as

$$N = \begin{pmatrix} \cos\theta & -\sin\theta \sin\theta & \cos\theta \end{pmatrix},$$

which is a physicists' favorite matrix: the rotation by θ. So, let us write M in terms of N:

$$M = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \sqrt{a^2+b^2}N = \sqrt{a^2+b^2} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

We can interpret M as a rotation by θ, followed by a scaling by $\sqrt{(}a^2+b^2)$. But what $\sqrt{a^2+b^2}$ exactly is just the "length" or the "amplitude" of the complex number $a+ib$, if we think of it as an arrow in a plane.

4.2 Relevant properties and definitions

1. The *modulus* of $z = a+ib$ is the "amplitude" of z, denoted by $|z| = \sqrt{a^2+b^2} = z\bar{z}$.

2. The modulus is *multiplicative*, i.e.

$$|wz| = |w||z|.$$

3. Triangle inequality:

$$|z+w| \leq |z| + |w|.$$

   We can readily show this geometrically, or algebraically.

4. The *argument* of $z = a+ib$ is θ, where

$$\theta = \begin{cases} \tan^{-1}\left(\frac{b}{a}\right), \text{ if } a > 0 \\ \frac{\pi}{2} + k2\pi, k \in \mathbb{R} \text{ if } a = 0, b > 0 \\ -\frac{\pi}{2} + k2\pi, k \in \mathbb{R} \text{ if } a = 0, b < 0 \\ \text{Undefined if } a = b = 0. \end{cases}$$

5. The *conjugate* of $a+ib$ is $a-ib$. Conjugation is *additive* and *multiplicative*, i.e.

$$\begin{aligned}\overline{z+w} &= \bar{z}+\bar{w}\\ \overline{wz} &= \bar{w}\bar{z}.\end{aligned}$$

Note that we can also show the multiplicative property with the matrix representation as well:

$$\overline{wz} \sim (WZ)^\top = Z^\top W^\top \sim \bar{z}\bar{w} = \bar{w}\bar{z}.$$

6. Euler's identity, generalized to de Moivre's formula:

$$z^n = r^n e^{in\theta}.$$

5 Vector Spaces & Linear Functions

5.1 Review of Linear Spaces and Subspaces

Properties 5.1. of linear spaces:

1. Commutativity and associativity of addition
2. Existence of an additively neutral element (null element). Zero multiples of elements give the null element: $0 \cdot V = \mathbf{0}$
3. Every element has an (unique) additively antipodal element
4. Scalar multiplication distributes over addition
5. Multiplicative identity:
6. $ab \cdot V = a \cdot (bV)$
7. $(a + b)V = aV + bV$

W is a subspace of V if

1. $S \subseteq V$
2. S is non-empty
3. S is closed under addition and scalar multiplication

Properties 5.2. that are interesting/important/maybe-not-so-obvious:

1. If S is a subspace of V and $S \neq V$ then S is a proper subspace of V.
2. If X in a subspace of Y and Y is a subspace of Z, then X is a subspace of W.
3. Non-trivial linear (non-singleton) spaces are infinite.

5.2 Review of Linear Maps

Consider linear spaces V and W and elements $v \in V$ and $w \in W$ and scalars $\alpha, \beta \in \mathbb{R}$, a function $F : V \to W$ is a linear map if

$$F[\alpha v + \beta w] = \alpha F[v] + \beta F[w].$$

Properties 5.3.

1. $F[\mathbf{0}_V] = \mathbf{0}_W$
2. $G[w] = (\alpha \cdot F)[w] = \alpha \cdot F[w]$
3. Given $F : V \to W$ and $G : V \to W$, $H[v] = F[v] + G[v] = (F + G)[v]$ is call the sum of the functions F and G.

4. Linear combinations of linear maps are linear.
5. Compositions of linear maps are linear.
6. Compositions distributes over linear combinations of linear maps.
7. Inverses of linear functions (if they exist) are linear.
8. Inverse of a bijective linear function is a bijective linear function

5.3 Review of Kernels and Images

Definition 5.1. Let $F : V \to W$ be given. The kernel of F is defined as

$$\ker(F) = \{v \in V | F[v] = \mathbf{0}_W\}.$$

Properties 5.4. Let $F : V \to W$ a linear map be given. Also, consider a linear map G such that $F \circ G$ is defined

1. F is null $\iff \ker(F) = V \iff \operatorname{Im}(F) = \mathbf{0}_W$
2. $\ker(F)$ is a subspace of V
3. $\operatorname{Im}(F)$ is a subspace of W
4. F is injective $\iff \ker(F) = \mathbf{0}_V$
5. $\ker(F) \subseteq \ker(F \circ G)$.
6. F injective $\implies \ker(F) = \ker(F \circ G)$

Definition 5.2. Let $F : V \to W$ be given. The image of F is defined as

$$\operatorname{Im}(F) = \{w \in W | \exists v \in V, F[v] = w\}.$$

5.4 Atrices

Definition 5.3. Atrix functions: Let $V_1, \ldots, V_m \in \mathbf{V}$ be given. Consider $f : \mathbb{R}^m \to \mathbf{V}$ be defined by

$$f\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \sum_{i=1}^{m} a_i V_i.$$

We denote f by

$$\begin{pmatrix} V_1 & V_2 & \ldots & V_m \end{pmatrix}.$$

We refer to the V_i's as the **columns** of f, even though there doesn't have to be any columns. Basically, f is simply a function that takes in an ordered list of coefficients and returns a linear combination of V_i with the respective coefficients. A matrix is a special atrix. Not every atrix is a matrix.

Properties 5.5. of atrices

1. The V_i's - the columns of an atrix - are the images of the standard basis tuples.

2. $\mathbf{e}_j \in \ker(V_1 \ldots V_m) \iff V_j = \mathbf{0}_V$, where $\mathbf{e}_j$ denotes a standard basis tuple with a 1 at the j^{th} position. To put in words, a kernel of an atrix contains a standard basis if and only if one of its columns in a null element.

3. $\mathrm{Im}(f) \equiv \mathrm{Im}(V_1 \ldots V_m) = \mathrm{span}(V_1 \ldots V_m)$

4. B is a null atrix $\iff \ker(B) = \mathbb{R}^m \iff \mathrm{Im}(B) = \mathbf{0}_V \iff V_j = \mathbf{0}_V \forall j = 1, 2, \ldots, m$.

5. f is a linear function $\mathbb{R}^m \to V \iff f$ is an atrix function $\mathbb{R}^m \to V$.

6. An atrix A is bijective/invertible, then its inverse A^{-1} is a linear function, but is an atrix only if A is a matrix.

7. Linear combinations of atrices are atrices:

$$\begin{aligned} \alpha \cdot \begin{pmatrix} V_1 & \cdots & V_m \end{pmatrix} + \beta \cdot \begin{pmatrix} W_1 & \ldots & W_m \end{pmatrix} \\ = \begin{pmatrix} \alpha V_1 + \beta W_1 & \ldots & \alpha V_m + \beta W_m \end{pmatrix} \end{aligned}$$

8. Compositions of two atrices are NOT defined unless the atrix going first is a matrix. Consider $F : \mathbb{R}^m \to W$ and $G : \mathbb{R}^n \to T$. $F \circ G$ is only defined if $T = \mathbb{R}^m$. This make G an $n \times m$ matrix. It follows that the atrix $F \circ G$ has the form

$$\begin{pmatrix} f(g_1) & f(g_2) & \ldots & f(g_m) \end{pmatrix}.$$

9. Consider $F : \mathbb{R}^m \to V$ and $G : \mathbb{R}^k \to V$. $\mathrm{Im}(F) \subseteq \mathrm{Im}(G) \iff F = G \circ C$, with $C \in \mathbb{M}_{k \times m}$, i.e. C is an $k \times m$ matrix.

10. Consider an atrix $A : \mathbb{R}^m \to V$. $A = (V_1 \ldots V_m)$. A is NOT injective.
$\iff \exists$ a non-trivial linear combination of the columns of A that gives $\mathbf{0}_V$
$\iff A = [\mathbf{0}_v]$ or $\exists j | V_j$ is linear combination of other columns of A
$\iff$ The first column of A is $\mathbf{0}_V$ or $\exists j | V_j$ is a linear combination some of $V_i, i < j$.

11. If atrix $A : \mathbb{R} \to V$ has a single column then it is injective if the column is not $\mathbf{0}_V$.

Properties 5.6. of elementary column operations for atrices. Elementary operations on the columns of F can be expressed as a composition of F and an appropriate elementary matrix E, $F \circ E$.

1. Swapping i^{th} and j^{th} columns: $F \circ E^{[i] \leftrightarrow [j]}$.

2. Scaling the j^{th} column by α: $F \circ E^{\alpha \cdot [j]}$.

3. Adjust the j^{th} column by adding to it $\alpha \times i^{th}$ column: $F \circ E^{[i] \xleftarrow{+} \alpha \cdot [j]}$.

4. Elementary column operations do not change the 'jectivity nor image of F.

5. If a column of F is a linear combination of some of the other columns then elementary column operations can turn it into $\mathbf{0}_V$.

6. Removing/Inserting null columns or columns that are linear combinations of other columns does not change the image of F.

7. Given atrix A, it is possible to eliminate (or not) columns of A to end up with an atrix B with $\mathrm{Im}(B) = \mathrm{Im}(A)$.

8. If B is obtained from insertion of columns into atrix A, then $\mathrm{Im}(B) = \mathrm{Im}(A) \iff$ the insert columns $\in \mathrm{Im}(A)$.

9. Inserting columns to a surjective A does not destroy surjectivity of A. (A is already having extra or just enough columns)

10. A surjective $\iff$ A is obtained by inserting columns (or not) into an invertible atrix $\iff$ deleting some columns of A (or not) gives an invertible matrix.

11. If A is injective, then column deletion does not destroy injectivity. (A is already "lacking" or having just enough columns)

12. The new atrix obtained from inserting columns from $\mathrm{Im}(A)$ into A is injective $\iff$ A is injective.

5.5 Linear Independence, Span, and Bases

5.5.1 Linear Independence

$$
\begin{aligned}
& X_1 \dots X_m \text{ are linearly independent} \\
\iff & F = [X_1 \dots X_m] \text{ injective} \\
\iff & \textstyle\sum a_i X_i = 0 \iff a_i = 0 \forall i \\
\iff & \mathbf{0}_V \notin \{X_i\} \text{ and none are linear combinations of some of the others.} \\
\iff & X_1 \neq \mathbf{0}_V \text{ and } X_j \text{ is not a linear combination of any of } X_i\text{'s for } i < j.
\end{aligned}
$$

Properties 5.7.

1. The singleton list is linearly independent if its entry is not the null element

2. Sublists of a linearly independent list are linearly independent

3. List operations cannot create/destroy linearly independence.

4. If a linear map $L : X \to W$ injective, then $X_1 \dots X_m$ linearly independent $\iff$ $L(X_1) \dots L(X_m)$ linearly independent.

5.5.2 Span

Properties 5.8.

1. Spans are subspaces. $\operatorname{span}(X_1 \dots X_m), X_j \in V$ is a subspace of V.
2. $\operatorname{span}(X_1 \dots X_j) \subseteq \operatorname{span}(X_1 \dots X_k)$ if $j \leq k$.
3. Adding elements to a list that spans V produces a list that spans V.
4. The following list operations do not change the span of V: removing the null/linearly dependent element, inserting a linearly dependent element, scaling element(s), adding (multiples) of an element to another element.
5. It is possible to reduce a list that spans to a list that spans AND have linearly independent elements.
6. Consider $A : V \to W$. If X_i span W then $A(X_i)$ span $\operatorname{Im}(A)$. $i = 1, 2, \dots, m$.
7. If $A : V \to W$ invertible, then X_i span $V \iff A(X_i)$ span W, $i = 1, 2, \dots, m$.

5.5.3 Bases

Properties 5.9.

1. A list is a basis of V if the elements are linearly independent and they span V.
2. A singleton linear space has no basis.
3. Re-ordering the elements of a basis gives another basis.
4. $\{X_1 \dots X_m\}$ is a basis of V if $(X_1 \dots X_m)$ is injective AND $\operatorname{Im}(X_1 \dots X_m) = V$, i.e. $(X_1 \dots X_m)$ invertible.
5. If $\{X_i\}$ is a basis of V and $\{Y_i\}$ is a list of elements in W, then there exists a unique linear function $L : V \to W$ satisfying
$$L[X_i] = Y_i$$
6. If $A : V \to W$ bijective, then $\{V_i\}$ forms a basis of V and $\{A(X_i)\}$ forms a basis of W.
7. Elementary operations on bases give bases.

5.6 Linear Bijections and Isomorphisms

Definition 5.4. Let linear spaces V, W be given. V is **isomorphic** to W if $\exists F : V \to W$ bijective. We say $V \sim W$.

Properties 5.10.

1. A non-zero scalar multiple of an isomorphism is an isomorphism.
2. A composition of isomorphism is an isomorphism.
3. "Isomorphism" behaves like an equivalence relation:
   (a) Reflexivity: $V \sim V$.
   (b) Symmetry: if $V \sim W$ then $W \sim V$.
   (c) Transitivity: if $V \sim W$ and $W \sim Z$ then $V \sim Z$.
4. Consider $F : V \to W$ an isomorphism.
   (a) Isomorphisms preserve linear independence. V_i's are linearly independent in $V \iff F(V_i)$'s are linearly independent in W.
   (b) isomorphisms preserve spanning. V_i's span $V \iff F(V_i)$'s span W.
   (c) Isomorphisms preserve bases. $\{V_i\}$ is a basis of $V \iff F\{(V_i)\}$ is a basis of W.
5. If $V \sim W$ then $\dim(V) = \dim(W)$ (finite or infinite).
6. If a linear map $A : V \to W$ is given and $A(X_i)$'s are linearly independent, then A_i's are linearly independent.
7. If a linear map $A : V \to W$ is injective and $\{X_i\}$ is a basis of V then $\{A(X_i)\}$ is a basis of $\mathrm{Im}(A)$

5.7 Finite-Dimensional Linear Spaces

5.7.1 Dimension

Properties 5.11.

1. $\mathbb{R}^m \sim R^n \iff m = n$.
2. Isomorphisms $F : \mathbb{R}^n \to V$ are bijective atrices.
3. Isomorphisms $G : W \to \mathbb{R}^m$ are inverses of bijective atrices.
4. Consider a non-singleton linear space V
   (a) V has a basis with n elements.
   (b) $V \sim \mathbb{R}^n$.
   (c) $V \sim W$, where W is any linear space with a basis of n elements.

5. If V is a linear space with a basis with n elements, then any basis of V has n elements.

6. Linear space $V \sim W$ where W is n-dimensional if V is n-dimensional.

7. For a non-singleton linear space V, $V \sim \mathbb{R}^n \iff \dim(V) = n$.

8. $V \sim W \iff \dim(V) = \dim(W)$.

9. If W is a subspace of V, then $\dim(W) \leq \dim(V)$. Equality holds when $W = V$.

5.7.2 Rank-Nullity Theorem

Let finite-dimensional linear space V and linear map $F : V \to W$ be given. Then $\mathrm{Im}(F)$ is finite-dimensional and

$$\dim(\mathrm{Im}(F)) + \dim(\ker(F)) = \dim(V)$$

A stronger statement: If a linear map $F : V \to W$ has finite rank and finite nullity $\iff$ V is finite-dimensional, then

$$\mathrm{Im}(F) + \ker(F) = \dim(V).$$

5.8 Infinite-Dimensional Spaces

Consider a non-singleton linear space V. The following statements are equivalent:

1. V is infinite-dimensional.

2. Every linearly independent list in V can be enlarged to a strictly longer linearly independent set in V.

3. Every linearly independent list in V can be enlarged to an arbitrarily long (finite) linearly independent set in V.

4. There are arbitrarily long (finite) linearly independent lists in V.

5. There are linearly independent lists in V of any (finite) length.

6. No list of finitely many elements of V spans V.

6 Sums of Subspaces & Products of vector spaces

6.1 Direct Sums

Definition 6.1. Let U_j, $j = 1, 2, \dots m$ are subspaces of V. $\sum_1^m U_j$ is a *direct sum* if each $u \in \sum U_j$ can be written in only one way as $u = \sum_1^m u_j$. The direct sum $\sum_i^m U_j$ is denoted as $U_1 \oplus \cdots \oplus U_m$.

Properties 6.1.

1. Condition for direct sum: If all U_j are subspaces of V, then $\sum_1^m U_j$ is a direct sum $\iff$ the only way to write 0 as $\sum_1^m u_j$, where $u_j \in U_j$ is to take $u_j = 0$ for all j.

2. If U, W are subspaces of V and $U \bigcap W = \{0\}$ then $U + W$ is a direct sum.

6.2 Products of Vector Spaces

Definition 6.2. Product of vectors spaces

$$V_1 \times \cdots \times V_m = \{(v_1, \dots, v_m) : v_j \in V_j, j = 1, 2, \dots, m\}.$$

Definition 6.3. Addition on $V_1 \times \cdots \times V_m$:

$$(u_1, \dots, u_m) + (v_1, \dots, v_m) = (u_1 + v_1, \dots, v_m + u_m).$$

Definition 6.4. Scalar multiplication on $V_1 \times \cdots \times V_m$:

$$\lambda(v_1, \dots, v_m) = (\lambda v_1, \dots, \lambda v_m).$$

Properties 6.2.

1. Product of vectors spaces is a vector space.

   V_j are vectors spaces over $\mathbb{F} \implies V_1 \times \cdots \times V_m$ is a vector space over $\mathbb{F}$.

2. Dimension of a product is the sum of dimensions:

$$\dim(V_1 \times \cdots \times V_m) = \sum_1^m \dim(V_j)$$

3. Vector space products are NOT commutative:

$$W \times V \neq V \times W.$$

   However,

$$V \times W \sim W \times V.$$

4. Vector space products are NOT associative:

$$V \times (W \times Z) \neq (V \times W) \times Z$$

6.3 Products & Direct Sums

Properties 6.3.

1. Let $U_1, \ldots, U_m$ be subspaces of V. Define a linear map $\Gamma : U_1 \times \cdots \times U_m \to U_1 + \cdots + U_m$ by:

$$\Gamma(u_1, \ldots, u_m) = \sum_1^m u_j.$$

$U_1 + \cdots + U_m$ is a direct sum $\iff$ Γ is injective.

2. Let U_j be finite-dimensional and are subspaces of V.

$$U_1 \oplus \cdots \oplus U_m \iff \dim(U_1 + \cdots + U_m) = \sum_1^m \dim(U_j)$$

6.4 Rank-Nullity Theorem

Suppose Z_1 and Z_2 are subspaces of a finite-dimensional vector space W. Consider $z_1 \in Z_1$, $z_2 \in Z_2$, and a function $\phi : Z_1 \times Z_2 \to Z_1 + Z_2 \prec W$ defined by

$$\phi \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1 + z_2.$$

First, ϕ is a linear function, as it satisfies the linearity condition:

$$\phi\left(\alpha \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \beta \begin{pmatrix} z_1' \\ z_2' \end{pmatrix}\right) = \alpha\phi \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \beta\phi\beta \begin{pmatrix} z_1' \\ z_2' \end{pmatrix}.$$

By rank-nullity theorem,

$$\dim(Z_1 \times Z_2) = \dim(Z_1 + Z_2) + \dim(\ker(\phi)).$$

But this is equivalent to

$$\dim(Z_1) + \dim(Z_2) = \dim(Z_1 + Z_2) + \dim(\ker(\phi))$$

The kernel of ϕ is:

$$\ker(\phi) = \left\{ \begin{pmatrix} v \\ -v \end{pmatrix} \middle| v \in z \in Z_1, z \in Z_2 \right\} = \left\{ \begin{pmatrix} v \\ -v \end{pmatrix} \middle| v \in z \in Z_1 \cap Z_2 \right\}$$

We can readily verify that $Z_1 \cap Z_2$ is a subspace of W. With this, $\dim(\ker(\phi)) = \dim(Z_1 \cap Z_2)$. So we end up with

$$\dim(Z_1 + Z_2) = \dim(Z_1) + \dim(Z_2) - \dim(Z_1 \cap Z_2).$$

Properties 6.4.

1. When $Z_1 \cap Z_2$ is trivial, then $Z_1 + Z_2$ is direct.

2. When $\dim(\ker(\phi)) = 0$, ϕ is injective. But ϕ is also surjective by definition, this implies ϕ is a bijection, in which case

$$Z_1 \oplus Z_2 \sim Z_1 + Z_2.$$

6.5 Nullspaces & Ranges of Operator Powers

1. Sequence of increasing null spaces: Suppose $T \in \mathcal{L}(V)$, i.e., T is some linear function mapping $V \to V$, then

   $$\{0\} = \ker(T^0) \subset \ker(T^1) \subset \ker(T^2) \subset \cdots \subset \ker(T^k) \subset \ker(T^{k+1}) \subset \ldots.$$

   Proof Outline. Let k be a nonnegative integer and $v \in \ker(T^k)$. Then $T^k v = 0$, so $T^{k+1}v = T(T^k v) = T(0) = 0$, so $v \in \ker T^{k+1}$. So $\ker(T^k) \subset \ker(T^{k+1})$. □

2. Equality in the sequence of null spaces: Suppose m is a nonnegative integer such that $\ker(T^m) = \ker(T^{m+1})$, then

   $$\ker(T^m) = \ker(T^{m+1}) = \ker(T^{m+2}) = \ldots.$$

   Proof Outline. We want to show

   $$\ker(T^{m+k}) = (T^{m+k+1}).$$

   We know that $\ker T^{m+k} \subset \ker T^{m+k+1}$. Suppose $v \in \ker T^{m+k+1}$, then

   $$T^{m+1}(T^k v) = T^{m+k+1}v = 0.$$

   So

   $$T^k v \in \ker T^{m+1} = \ker T^m.$$

   So

   $$0 = T^m(T^k v) = T^{m+k}v,$$

   i.e., $v \in \ker T^{m+k}$. So $\ker T^{m+k+1} \subset \ker T^{m+k}$. This completes the proof. □

3. Null spaces stop growing: If $n = \dim(V)$, then

   $$\ker(T^n) = \ker(T^{n+1}) = \ker(T^{n+2}) = \ldots.$$

   Proof Outline. To show:

   $$\ker T^n = \ker T^{n+1}.$$

   Suppose this is not true. Then the dimension of the kernel has to increase by at least 1 every step until $n+1$. Thus $\dim\ker T^{n+1} \geq n+1 > n = \dim(V)$. This is a contradiction. □

4. V is the direct sum of $\ker(T^{\dim(V)})$ and $\mathrm{Im}(T^{\dim(V)})$: If $n = \dim(V)$, then

   $$V = \ker(T^n) \oplus \mathrm{Im}(T^n).$$

Proof Outline. To show:

$$\ker T^n \cap \operatorname{Im} T^n = \{0\}.$$

Suppose $v \in \ker T^n \cap \operatorname{Im} T^n$. Then $T^n v = 0$ and $\exists u \in V$ such that $v = T^n u$. So

$$T^n v = T^{2n} u = 0.$$

So

$$T^n u = 0.$$

But this means $v = 0$. □

5. IT IS NOT TRUE THAT $V = \ker(T) \oplus \operatorname{Im}(T)$ in general.

6.6 Generalized Eigenvectors and Eigenspaces

Definition 6.5. Suppose $T \in \mathcal{L}(V)$ and Λ is an eigenvalue of T. A vector $v \in V$ is called a **generalized eigenvector** of T corresponding to λ if $v \neq 0$ and

$$(T - \lambda I)^j v = 0$$

for some positive integer j.

Definition 6.6. Generalized Eigenspace: Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The **generalized eigenspace** of T corresponding to λ, denoted $G(\lambda, T)$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Properties 6.5. 1. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Then

$$G(\lambda, T) = \ker(T - \lambda I)^{\dim(V)}.$$

Proof Outline. Suppose $v \in \ker(T_\lambda I)^{\dim(V)}$. Then $v \in G(\lambda, T)$. So, $\ker(T - \lambda I)^{\dim V} \subset G(\lambda, T)$. Next, suppose $v \in G(\lambda, T)$. Then these is a positive integer j such that

$$v\ in \ker(T - \lambda I)^j.$$

But if this is true, then

$$v\ in \ker(T - \lambda I)^{\dim V},$$

since $\ker(T - \lambda I)^{\dim V}$ is the largest possible kernel, in a sense. □

2. Linearly independent generalized eigenvectors: Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding generalized eigenvectors. Then $v_1, \ldots, v_m$ is linearly independent.

6.7 Nilpotent Operators

Definition 6.7. An operator is called **nilpotent** if some power of it equals 0.

Properties 6.6.

Nilpotent operator raised to dimension of domain is 0: Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then

$$N^{\dim(V)} = 0.$$

Matrix of a nilpotent operator: Suppose N is a nilpotent operator on V. Then there is a basis of V with espect to which the matrix of N has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix};$$

here all entries on and below the diagonal are 0's.

6.8 Weyr Characteristic

7 Idempotents & Resolutions of Identity

8 Block-representations of operators

8.1 Direct sums of operators

8.2 Coordinatization & matricial representation of linear functions

Consider a finite-dimensional linear space V with basis $\{V_i\}$, $i = 1, 2, \ldots, m$. An element $\tilde{V}$ in V can be expressed in exactly one way:

$$\tilde{V} = \sum_{i=1}^{m} a_i V_i,$$

where $\{a_i\}$ is unique. We call $\{a_i\}$ the coordinate tuple of $\tilde{V}$ and a_i's the coordinates.

Properties 8.1.

1. Inverse of a bijective atrix outputs the coordinates. Suppose $A = [V_i]$. Then

$$\tilde{V} = \sum_{i=1}^{m} a_i V_i \iff A^{-1}(Z) = \begin{pmatrix} a_1 & \ldots & a_m \end{pmatrix}^\top$$

8.3 Equality of rank and trace for idempotents; Resolution of Identity Revisited

9 Invariant subspaces

9.1 Reducing subspaces

10 Polynomials applied to operators

10.1 Minimal polynomials of block-Δ^r operators

10.2 Minimal polynomials at a vector

11 Eigentheory

11.1 Spectral Mapping Theorem

12 Triangularization

12.1 Compression to invariant subspaces

12.2 Simultaneously Δ-ity of commuting families

13 Diagonalization

13.1 Spectral resolutions

13.2 Compressions to reducing subspaces

13.3 Simultaneous diagonalizability for commuting families

14 Primary decomposition over $\mathbb{C}$ and generalized eigenspaces

15 Cyclic decomposition and Jordan form

15.1 Square roots of operators

15.2 Similarity of a matrix and its transpose

15.3 Similarity of a matrix and its conjugate

15.4 Jordan forms of $\mathcal{AB}$ and $\mathcal{BA}$

15.5 Power-convergent operators

15.6 Power-bounded operators

15.7 Row-stochastic matrices

16 Determinant & Trace

16.1 Classical adjoints

16.2 Cayley-Hamilton theorem

17 Inner products and norms

17.1 Riesz representation theorem

17.2 Adjoints

17.3 Grammians

17.4 Orthogonal complements and orthogonal decompositions

17.5 Ortho-projections

17.6 Closest point solutions

17.7 Gram-Schmidt and orthonormal bases

18 Isometries and unitary operators

19 Ortho-triangularization

20 Spectral resolutions

21 Ortho-diagonalization; self-adjoint and normal operators; spectral theorems

22 Positive (semi-)definite operators

22.1 Classification of inner products

22.2 Positive square roots

23 Polar decomposition

24 Single value decomposition

24.1 Spectral/operator norm

24.2 Singular values and approximation

24.3 Singular values and eigenvalues

25 Problems and Solutions

25.1 Problem set 1 (under corrections)

25.2 Problem set 2

Problem. 1. Finite unions of subspaces are rarely a subspace
Suppose that $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n$ are subspaces of a vector space $\mathbf{V}$. Prove that the following are equivalent:

1. $\mathbf{W}_1 \cup \mathbf{W}_2 \cup \cdots \cup \mathbf{W}_n$ is a subspace of $\mathbf{V}$.
2. One of the $\mathbf{W}_i$'s contains all the others.

Solution. 1.

- [1. $\implies$ 2.] : Suppose 2. is false and define

$$\mathbf{S} = \bigcup_{i=1}^{n_0} \mathbf{W}_i$$

  a subspace of $\mathbf{V}$, where none of the $\mathbf{W}_i$'s $\prec \mathbf{V}$ contains all the others and n_0 is minimal. If $n_0 = 1$, then the implication 1. $\implies$ 2. is true because $\mathbf{W}_1$ contains itself. Therefore, in order for this implication to fail, $n_0 \geq 2$.

  Because n_0 is minimal, $\mathbf{S}$ cannot be obtained from a union of less than n_0 of the $\mathbf{W}_i$'s. Therefore, for any i, there exists $w_i \in \mathbf{W}_i$ such that $w_i \notin \bigcup_{j \neq i}^{n_0} \mathbf{W}_j$; i.e., each $\mathbf{W}_i$ contains an element w_i that does not belong to any other $\mathbf{W}_j$'s.

  Take $w_1 \in \mathbf{W}_1$ and $w_2 \in \mathbf{W}_2 \setminus \mathbf{W}_1$ such that $w_1 \notin \bigcup_{j \neq 1}^{n_0} \mathbf{W}_j$. It follows that $w_1, w_2 \in \mathbf{S}$ because $\mathbf{W}_1, \mathbf{W}_2 \subset \mathbf{S}$. And since $\mathbf{S}$ is a subspace, $w_1 + w_2 \in \mathbf{S}$ by closure under addition. Let a collection of the n_0 linear combinations of w_1 and w_2, defined as

$$T = \{mw_1 + w_2 \,|\, m = 1, 2, \ldots, n_0\}.$$

  Consider a typical element, $mw_1 + w_2 \in \mathbf{S}$. Suppose $mw_1 + w_2 \in \mathbf{W}_1$. By closure under addition,

$$(mw_1 + w_2) - mw_1 = w_2 \in \mathbf{W}_1$$

  for any $m = 1, 2, \ldots, n_0$. But this contradicts the choice of w_2. Therefore, $mw_1 + w_2 \notin \mathbf{W}_1$, for any m.

  It follows that $T \subset \bigcup_{i=2}^{n_0} \mathbf{W}_i$. By construction, T has n_0 elements, while $\bigcup_{i=2}^{n_0} \mathbf{W}_i$ has $(n_0 - 1)$ $\mathbf{W}_i$'s. By the Pigeonhole Principle, a $\mathbf{W}_k$ of $\mathbf{W}_2, \ldots, \mathbf{W}_{n_0}$ contains $mw_1 + w_2$ for two distinct values of m. Let these values be m_a and m_b. By closure under addition, we have

$$(m_a w_1 + w_2) - (m_b w_1 + w_2) = (m_a - m_b) w_1 \in \mathbf{W}_k.$$

  But since $m_a \neq m_b$, $w_1 \in \mathbf{W}_k$. This contradicts our initial choice of w_1. Therefore, by contraposition, the implication 1. $\implies$ 2. must hold.

- [2. $\implies$ 1.] : Without loss of generality, assume that $\mathbf{W}_1$ contains all the other $\mathbf{W}_j$'s. It follows that

$$\bigcup_{i=1}^{n_0} \mathbf{W}_i = \mathbf{W}_1.$$

  Since $\mathbf{W}_1$ is a subspace of $\mathbf{V}$, $\bigcup_{i=1}^{n_0} \mathbf{W}_i$ is also a subspace of $\mathbf{V}$. Therefore, 2. $\implies$ 1. must hold.

Problem. 2. Images of pre-images and pre-images of images: Is there a 3×3 matrix $\mathcal{A}$ such that

$$\mathbf{W} := \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \middle| x, y \in \mathbb{C} \right\}$$

is an invariant subspace for $\mathcal{A}$, and

$$\mathcal{A}\left[\mathcal{A}^{-1}[\mathbf{W}]\right] \subsetneq \mathbf{W} \subsetneq \mathcal{A}^{-1}\left[\mathcal{A}[\mathbf{W}]\right]?$$

Justify your answer.

Solution. 2.

First, we observe that $\dim(\mathbf{W}) = 2$, because a basis set of $\mathbf{W}$,

$$\left\{ \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} \right\},$$

where $x, y, \in \mathbb{C}$, has two elements. Next, consider $\mathbb{C}^3$. Since $\mathbf{W}$ and $\mathbb{C}^3$ are subspaces and any element of $\mathbf{W}$ is contained in $\mathbb{C}^3$ by construction, $\mathbf{W}$ is a subspace of $\mathbb{C}^3$.

By the condition $\mathcal{A}\left[\mathcal{A}^{-1}[\mathbf{W}]\right] \subsetneq \mathbf{W} \subsetneq \mathcal{A}^{-1}\left[\mathcal{A}[\mathbf{W}]\right]$, it must be true that

$$\dim\left(\mathcal{A}\left[\mathcal{A}^{-1}[\mathbf{W}]\right]\right) < \dim(\mathbf{W}) < \dim\left(\mathcal{A}^{-1}\left[\mathcal{A}[\mathbf{W}]\right]\right).$$

Since $\dim(\mathbf{W}) = 2$ and $\mathcal{A}$ is a 3×3 matrix (i.e., neither the dimension of $\ker(\mathcal{A})$ nor that of $\mathrm{Im}(A)$ can exceed 3), it is required that $\dim\left(\mathcal{A}\left[\mathcal{A}^{-1}[\mathbf{W}]\right]\right) \leq 1$ and $\dim\left(\mathcal{A}^{-1}\left[\mathcal{A}[\mathbf{W}]\right]\right) = 3$.

Because $\mathbf{W}$ is invariant under $\mathcal{A}$, $\mathcal{A}[\mathbf{W}] \subseteq \mathbf{W}$. It follows that $\mathcal{A}^{-1}\left[\mathcal{A}[\mathbf{W}]\right] \subseteq \mathcal{A}^{-1}[\mathbf{W}]$. Therefore,

$$\dim\left(\mathcal{A}^{-1}\left[\mathcal{A}[\mathbf{W}]\right]\right) \leq \dim\left(\mathcal{A}^{-1}[\mathbf{W}]\right).$$

So $3 \leq \dim\left(\mathcal{A}^{-1}[\mathbf{W}]\right)$. But since the dimension of any image of $\mathcal{A}$ cannot exceed 3 (because $\mathcal{A} : \mathbb{C}^3 \to \mathbb{C}^3$),

$$3 \leq \dim\left(\mathcal{A}^{-1}[\mathbf{W}]\right) \leq 3,$$

which implies $\dim\left(\mathcal{A}^{-1}[\mathbf{W}]\right) = 3 = \dim(\mathbf{V})$.

Consider $v \in \mathcal{A}^{-1}[\mathbf{W}]$. v can have the form $(x\ y\ z)^\top$, $x, y, z \in \mathbb{C}$, so $v \in \mathbb{C}^3$. This implies $\mathcal{A}^{-1}[\mathbf{W}] \subseteq \mathbb{C}^3$. But since their dimensions are both 3, $\mathcal{A}^{-1}[\mathbf{W}] = \mathbb{C}^3$. It follows that

$$\mathcal{A}\left[\mathcal{A}^{-1}[\mathbf{W}]\right] = \mathcal{A}[\mathbb{C}^3] \supseteq \mathcal{A}[\mathbf{W}],$$

where the second relation comes from the fact that $\mathbf{W}$ is a subspace of $\mathbb{C}^3$. Hence

$$1 \geq \dim\left(\mathcal{A}\left[\mathcal{A}^{-1}[\mathbf{W}]\right]\right) \geq \dim\left(\mathcal{A}[\mathbf{W}]\right),$$

which implies $\dim(\mathcal{A}[\mathbf{W}]) = 0$ or 1. Such a matrix $\mathcal{A}$ can be

$$\mathcal{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

First, $\mathbf{W}$ is invariant under $\mathcal{A}$ because take any $(x \;\; y \;\; 0)^\top \in \mathbf{W}$, $\mathcal{A}(x \;\; y \;\; 0)^\top = (x \;\; 0 \;\; 0)^\top \in \mathbf{W}$, which implies $\mathcal{A}[\mathbf{W}] \subseteq [\mathbf{W}]$ (a).

Second, it is easy to see that the pre-image of $\mathbf{W}$ under $\mathcal{A}$ is $\mathbb{C}^3$, since take any $(x \;\; y \;\; z)^\top \in \mathbb{C}^3$, we have $\mathcal{A}(x \;\; y \;\; z)^\top = (x \;\; 0 \;\; 0)^\top \in \mathbf{W}$. But observe that $\mathrm{Im}(\mathcal{A})$ is proper subset of $\mathbf{W}$, because any $(x \;\; y \;\; 0)^\top \in \mathbf{W}$ with $y \neq 0$ is not contained in $\mathrm{Im}(\mathcal{A})$. This implies $\mathcal{A}\left[\mathcal{A}^{-1}[\mathbf{W}]\right] \subsetneq \mathbf{W}$ (b).

Lastly, as we have pointed out, $\dim\left(A^{-1}\left[\mathcal{A}[\mathbf{W}]\right]\right) = 3$ and $A^{-1}\left[\mathcal{A}[\mathbf{W}]\right] \prec \mathbb{C}^3$, so $A^{-1}\left[\mathcal{A}[\mathbf{W}]\right] = \mathbb{C}^3 \supsetneq \mathbf{W}$ (c).

From $(a), (b), (c)$, we see that this 3×3 matrix $\mathcal{A}$ works as desired.

Problem. 3. Invariant subspaces form a "lattice":

1. Argue that $\mathfrak{Lat}(\mathcal{L})$ is closed under intersection; i.e., that the intersection of any two invariant subspaces for $\mathcal{L}$ is also an invariant subspace for $\mathcal{L}$.
2. Argue that $\mathfrak{Lat}(\mathcal{L})$ is closed under subspace sums; i.e., that a subspace sum of two invariant subspaces for $\mathcal{L}$ is again an invariant subspace for $\mathcal{L}$.
3. Show that an image under $\mathcal{L}$ of an invariant subspace for $\mathcal{L}$ is again an invariant subspace for $\mathcal{L}$.
4. Show that a pre-image under $\mathcal{L}$ of an invariant subspace for $\mathcal{L}$ is again an invariant subspace for $\mathcal{L}$.

Solution. 3.

1. Let subspaces $\mathbf{V}, \mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$ be given. $\mathfrak{Lat}(\mathcal{L})$ is closed under intersection if $\mathbf{V} \cap \mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$. Consider $u \in \mathbf{V} \cap \mathbf{W}$, then $u \in \mathbf{V}$ and $u \in \mathbf{W}$. It follows that $\mathcal{L}(u) \in \mathcal{L}[\mathbf{V}]$ and $\mathcal{L}(u) \in \mathcal{L}[\mathbf{W}]$.

   Now, since $\mathcal{L}[\mathbf{V}] \subseteq \mathbf{V}$ and $\mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$ because $\mathbf{W}, \mathbf{V} \in \mathfrak{Lat}(\mathcal{L})$, we have $\mathcal{L}(u) \in \mathbf{V}$ and $\mathcal{L}(u) \in \mathbf{W}$, which implies $\mathcal{L}(u) \in \mathbf{V} \cap \mathbf{W}$. Since this implication holds for any $u \in \mathbf{V} \cap \mathbf{W}$, $\mathcal{L}[\mathbf{V} \cap \mathbf{W}] \subseteq \mathbf{V} \cap \mathbf{W}$. Therefore, $\mathbf{V} \cap \mathbf{W}$ is invariant under $\mathcal{L}$; i.e., $\mathbf{W} \cap \mathbf{V} \in \mathfrak{Lat}(\mathcal{L})$. This completes the argument.

2. **(MORE CARE NEEDED)** Let subspaces $\mathbf{V}, \mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$ be given. $\mathfrak{Lat}(\mathcal{L})$ is closed under subspace sums if $\mathbf{V} + \mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$. Consider $\mathcal{L}[\mathbf{V} + \mathbf{W}] = \mathcal{L}[\mathbf{V}] + \mathcal{L}[\mathbf{W}]$. Since $\mathcal{L}[\mathbf{V}] \subseteq \mathbf{V}$ and $\mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$ by the definitions of $\mathbf{V}$ and $\mathbf{W}$, it must be true that $\mathcal{L}[\mathbf{V}] + \mathcal{L}[\mathbf{W}] \subseteq \mathbf{V} + \mathbf{W}$. Therefore, $\mathcal{L}[\mathbf{V} + \mathbf{W}] \subseteq \mathbf{V} + \mathbf{W}$, which implies $\mathbf{V} + \mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$. This completes the argument.

3. Let $\mathbf{V} \in \mathfrak{Lat}(\mathcal{L})$ be given. By definition, $\mathcal{L}[\mathbf{V}] \subseteq \mathbf{V}$. It follows that $\mathcal{L}\left[\mathcal{L}[\mathbf{V}]\right] \subseteq \mathcal{L}[\mathbf{V}]$. So, $\mathcal{L}[\mathbf{V}]$ is invariant under $\mathcal{L}$.

4. Let $\mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$ be given. Claim: $\mathcal{L}\left[\mathcal{L}^{-1}[\mathbf{W}]\right] \subseteq \mathcal{L}^{-1}[\mathbf{W}]$.

   It follows from the definition of the pre-image of $\mathbf{W}$ under $\mathcal{L} : \mathbf{V} \to \mathbf{W}$ that $\mathcal{L}^{-1}[\mathbf{W}] = \{v \in \mathbf{V} | \mathcal{L}(v) \in \mathbf{W}\}$. This implies $\mathcal{L}\left[\mathcal{L}^{-1}[\mathbf{W}]\right] \subseteq \mathbf{W}$ (i).

   Next, consider $w \in \mathbf{W} \subseteq \mathbf{V}$. By definition, $\mathcal{L}^{-1}\left[\mathcal{L}[\mathbf{W}]\right] = \{v \in \mathbf{V} | \mathcal{L}(v) \in \mathcal{L}[\mathbf{W}]\}$, which implies $w \in \mathcal{L}^{-1}[\mathcal{L}[\mathbf{W}]]$, because $\mathcal{L}(w) \in \mathcal{L}[\mathbf{W}]$. Therefore, $\mathbf{W} \subseteq \mathcal{L}^{-1}\left[\mathcal{L}[\mathbf{W}]\right]$ (ii).

   Moreover, because $\mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$, $\mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$ (iii).

From (i), (ii), and (iii) we have

$$\mathcal{L}\left[\mathcal{L}^{-1}[\mathbf{W}]\right] \overset{(1)}{\subseteq} \mathbf{W} \overset{(2)}{\subseteq} \mathcal{L}^{-1}\left[\mathcal{L}[\mathbf{W}]\right] \overset{(3)}{\subseteq} \mathcal{L}^{-1}[\mathbf{W}].$$

Therefore, $\mathcal{L}\left[\mathcal{L}^{-1}[\mathbf{W}]\right] \subseteq \mathcal{L}^{-1}[\mathbf{W}]$, verifying the claim.

Problem. 4. Cyclic invariant subspaces For a given $\mathcal{L} \in \mathfrak{L}(\mathbf{V})$ and a fixed $v_0 \in \mathbf{V}$, define

$$\mathbf{P}(\mathcal{L}, v_0) := \left\{ \left(a_0\mathcal{L}^0 + a_1\mathcal{L}^1 + a_2\mathcal{L}^2 + \cdots + a_k\mathcal{L}^k\right)(v_0) \middle| k \geq 0, a_i \in \mathbb{C} \right\}.$$

1. Argue that $\mathbf{P}(\mathcal{L}, v_0)$ is a subspace of $\mathbf{V}$.
2. Argue that $\mathbf{P}(\mathcal{L}, v_0)$ is invariant under $\mathcal{L}$.
3. Argue that $\mathbf{P}(\mathcal{L}, v_0)$ is the smallest invariant subspace for $\mathcal{L}$ that contains v_0. We refer to $\mathbf{P}(\mathcal{L}, v_0)$ as **the cyclic invariant subspace for $\mathcal{L}$ generated by** v_0.
4. Argue that $\mathbf{P}(\mathcal{L}, v_0)$ is 1-dimensional exactly when v_0 is an eigenvector of $\mathcal{L}$.
5. Argue a subspace $\mathbf{W}$ of $\mathbf{V}$ is invariant under $\mathcal{L}$ exactly when it is a union of cyclic invariant subspaces for $\mathcal{L}$.

Solution. 4.

1. Consider a typical element $P(\mathcal{L}, v_0) \in \mathbf{P}(\mathcal{L}, v_0)$. Since $\mathcal{L} \in \mathfrak{L}(\mathbf{V})$, $\mathcal{L}^i(v_0) \in \mathbf{V}$ for all i and $v_0 \in \mathbf{V}$. By closure under scalar multiplication and addition, for any $a_i \in \mathbb{C}$, $v_0 \in \mathbf{V}$, and $k \geq 0$,

   $$P(\mathcal{L}, v_0) = \sum_{i=0}^{k} a_i\mathcal{L}^i(v_0) \in \mathbf{V}.$$

   Therefore $\mathbf{P}(\mathcal{L}, v_0) \subseteq \mathbf{V}$ (i).

   With $a_i = 0$ for all i, $\mathbf{P}(\mathcal{L}, v_0) \ni P(\mathcal{L}, v_0) = \sum_{i=1}^{k} 0 \cdot \mathcal{L}^i(v_0) = \mathbf{0_P}$. So $\mathbf{P}(\mathcal{L}, v_0)$ contains the null element (ii).

   Consider $P_1(\mathcal{L}, v_0) = \sum_{i=0}^{k} b_i\mathcal{L}^i(v_0)$ and $P_2(\mathcal{L}, v_0) = \sum_{j=0}^{l} c_j\mathcal{L}^j(v_0)$, with $b_i, c_j \in \mathbb{C}$ for all i, j. Without loss of generality, assume that $k \geq l$. It is clear that $\mathbf{P}(\mathcal{L}, v_0)$ is closed under addition because,

   $$\begin{aligned} P_1(\mathcal{L}, v_0) + P_2(\mathcal{L}, v_0) &= \sum_{i=0}^{k} b_i\mathcal{L}^i(v_0) + \sum_{j=0}^{l} c_j\mathcal{L}^j(v_0) \\ &= \sum_{i=0}^{k} b_i\mathcal{L}^i(v_0) + \sum_{i=0}^{k} c_i\mathcal{L}^j(v_0) \quad \text{with } c_i = 0 \text{ for all } i > l \\ &= \sum_{i=0}^{k} (b_i + c_i)\mathcal{L}^i(v_0) \\ &= \sum_{i=0}^{k} d_i\mathcal{L}^i(v_0) \in \mathbf{P}(\mathcal{L}, v_0), \end{aligned}$$

where $d_i = b_i + c_i \in \mathbb{C}$. (iii).

It is also clear that $\mathbf{P}(\mathcal{L}, v_0)$ is closed under scalar multiplication because given $\mu \in \mathbb{C}$,

$$\mu P_1(\mathcal{L}, v_0) = \mu \sum_{i=0}^{k} b_i \mathcal{L}^i(v_0) = \sum_{i=0}^{k} \mu b_i \mathcal{L}^i(v_0) = \sum_{i=0}^{k} e_i \mathcal{L}^i(v_0) \in \mathbf{P}(\mathcal{L}, v_0),$$

where $e_i = \mu b_i \in \mathbb{C}$ (iv).

By (i), (ii), (iii), and (iv), $\mathbf{P}(\mathcal{L}, v_0)$ is a subspace of $\mathbf{V}$.

2. Let $P(\mathcal{L}, v_0)$ be given. It suffices to show that $\mathcal{L}(P(\mathcal{L}, v_0)) \in \mathbf{P}(\mathcal{L}, v_0)$, for all $P(\mathcal{L}, v_0)$.

$$\mathcal{L}(P(\mathcal{L}, v_0)) = \mathcal{L}\left(\sum_{i=0}^{k} a_i \mathcal{L}^i(v_0)\right) = \sum_{i=0}^{k} a_i \mathcal{L}^{i+1}(v_0).$$

Define coefficients $b_j = a_i$ where $j = i + 1$ and $b_0 = 0$, then

$$\mathcal{L}(P(\mathcal{L}, v_0)) = \sum_{j=1}^{k+1} b_j \mathcal{L}^j(v_0) = \sum_{j=0}^{k+1} b_j \mathcal{L}^j(v_0) \in \mathbf{P}(\mathcal{L}, v_0).$$

Therefore, $\mathcal{L}\left[\mathbf{P}(\mathcal{L}, v_0)\right] \subseteq \mathbf{P}(\mathcal{L}, v_0)$; i.e., $\mathbf{P}(\mathcal{L}, v_0)$ is invariant under $\mathcal{L}$.

3. Let $\mathbf{P}(\mathcal{L}, v_0)$ be given. $\mathbf{P}(\mathcal{L}, v_0)$ is an invariant subspace under $\mathcal{L}$ that contains v_0 by definition. Suppose $\mathbf{Q}(\mathcal{L}, v_0)$ is some invariant subspace under $\mathcal{L}$ that also contains v_0. It suffices to show $\mathbf{P}(\mathcal{L}, v_0) \subseteq \mathbf{Q}(\mathcal{L}, v_0)$ for any such $\mathbf{Q}(\mathcal{L}, v_0)$.

   Consider $\mathcal{L}^j(v_0)$ for some $j \geq 0$. $\mathcal{L}^j(v_0) \in \mathbf{P}(\mathcal{L}, v_0)$ by definition. We claim that $\mathcal{L}^j(v_0) = q \in \mathbf{Q}(\mathcal{L}, v_0)$ for any non-negative j. Assume that this is true. The base case where $j = 0$ is true since $v_0 \in \mathbf{Q}(\mathcal{L}, v_0)$. The inductive case is also true because

$$\mathcal{L}^{(j+1)}(v_0) = \mathcal{L}(\mathcal{L}^j(v_0)) = \mathcal{L}(q) \in \mathbf{Q}(\mathcal{L}, v_0)$$

   where the last relation comes from the fact that $\mathbf{Q}(\mathcal{L}, v_0)$ is invariant under $\mathcal{L}$. Therefore, by the principle of induction, $\mathcal{L}^j(v_0) \in \mathbf{Q}(\mathcal{L}, v_0)$ for any non-negative j. This implies $\mathbf{P}(\mathcal{L}, v_0) \subseteq \mathbf{Q}(\mathcal{L}, v_0)$ for any such $\mathbf{Q}(\mathcal{L}, v_0)$. Thus $\mathbf{P}(\mathcal{L}, v_0)$ must be the smallest invariant subspace under $\mathcal{L}$ that contains v_0.

4. (a) ($\Longrightarrow$) : Let a $\mathbf{P}(\mathcal{L}, v_0)$ be given such that $\dim(\mathbf{P}(\mathcal{L}, v_0)) = 1$. Consider the subspace $\operatorname{span}(v_0)$. We know that $\dim(\operatorname{span}(v_0)) = 1$. Consider $v \in \operatorname{span}(v_0)$. Then v can be expressed as $v = av_0$ where $a \in \mathbb{C}$.

It follows immediately that $v \in \mathbf{P}(\mathcal{L}, v_0)$. Therefore, $\text{span}(v_0) \subseteq \mathbf{P}(\mathcal{L}, v_0)$. But because $\dim(\mathbf{P}(\mathcal{L}, v_0)) = 1 = \dim(\text{span}(v_0))$, it follows that $\mathbf{P}(\mathcal{L}, v_0) = \text{span}(v_0)$; i.e., all elements of $\mathbf{P}(\mathcal{L}, v_0)$ are some scalar multiples of v_0.

Now, it suffices to show $\mathcal{L}(v_0) = \lambda v_0$, where $\lambda \in \mathbb{C}$. Consider $P(\mathcal{L}, v_0) = \mathcal{L}^0(v_0) + \mathcal{L}(v_0) \in \mathbf{P}(\mathcal{L}, v_0) = \text{span}(v_0)$, $v_0 \neq \mathbf{0_V}$. By closure under addition, $(\mathcal{L}^0(v_0) + \mathcal{L}(v_0)) - \mathcal{L}^0(v_0) = \mathcal{L}(v_0) \in \text{span}(v_0)$. This implies $\mathcal{L}(v_0) = \lambda v_0$ for some $\lambda \in \mathbb{C}$; i.e., v_0 is an eigenvector of $\mathcal{L}$.

(b) ($\Longleftarrow$) : If v_0 is an eigenvector of $\mathcal{L}$, then $\mathcal{L}^i(v_0) = \lambda^i v_0$, where $\lambda^i \in \mathbb{C}$ is the $\mathcal{L}$'s v_0-eigenvalue raised to the i^{th} power. It follows that a typical element of $\mathbf{P}(\mathcal{L}, v_0)$ is

$$P(\lambda, v_0) = \sum_{i=0}^{k} a_i \mathcal{L}^i v_0 = \left(\sum_{i=0}^{k} a_i \lambda^i \right) v_0 = \mu v_0,$$

where $\mu = \sum_{i=0}^{k} a_i \lambda^i \in \mathbb{C}$. Therefore, $\mathbf{P}(\mathcal{L}, v_0) = \text{span}(v_0)$; i.e., $\dim(\mathbf{P}(\mathcal{L}, v_0)) = 1$. This completes the argument.

5. (a) ($\Longrightarrow$) : Let $\mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$ be given. Then $\mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$. Let $w_1 \in \mathbf{W}$ be given. We first show that $\mathcal{L}^k(w_1) \in \mathbf{W}$ for all non-negative k by induction. Assume that $\mathcal{L}^k(w_1) = v_1 \in \mathbf{W}$ for all non-negative k holds. The base case where $k = 0$ is trivial since $\mathcal{L}^0(w_1) = w_1 \in \mathbf{W}$. The inductive case is also true because

$$\mathcal{L}^{(k+1)}(w_1) = \mathcal{L}(\mathcal{L}^k(w_1)) = \mathcal{L}(v_1) \in \mathbf{W}.$$

By the principle of induction, $\mathcal{L}^k(w_1) \in \mathbf{W}$ for all $w_1 \in \mathbf{W}$ and $k \geq 0$ (†). Now, consider $P \in \mathbf{P}(\mathcal{L}, w_1)$. By the definition of $\mathbf{P}(\mathcal{L}, w_1)$, P is a linear combination of $\mathcal{L}^0(w_1), \ldots, \mathcal{L}^k(w_1)$ for some $k \geq 0$. So, by (†), $P \in \mathbf{W}$. This implies $\mathbf{P}(\mathcal{L}, w_1) \subseteq \mathbf{W}$.

We can repeat the argument above, starting with $w_2, w_3, \dots \in \mathbf{W}$, and conclude that $\mathbf{P}(\mathcal{L}, w_2) \subseteq \mathbf{W}$, $\mathbf{P}(\mathcal{L}, w_3) \subseteq \mathbf{W}$, and so on. This means $\mathbf{W}$ contains a union of $\mathbf{P}(\mathcal{L}, w_1), \mathbf{P}(\mathcal{L}, w_2), \mathbf{P}(\mathcal{L}, w_3), \ldots$. Moreover, observe the fact that any element $w_n \in \mathbf{W}$ can be used to generate a $\mathbf{P}(\mathcal{L}, w_n)$ that is contained in the union of itself and the other $\mathbf{P}(\mathcal{L}, w_m)$'s, which is ultimately contained in $\mathbf{W}$. Therefore, $\mathbf{W}$ *is* itself a union of cyclic invariant subspaces for $\mathcal{L}$.

(b) ($\Longleftarrow$) : Let $\mathbf{W} = \mathbf{P_1} \cup \mathbf{P_2} \cup \cdots \cup \mathbf{P_n}$ be a subspace of $\mathbf{V}$, where $\mathbf{P_i} \equiv \mathbf{P}(\mathcal{L}, v_i)$ and v_i is an arbitrary element of $\mathbf{V}$. By problem 1, because $\mathbf{P_1} \cup \mathbf{P_2} \cup \cdots \cup \mathbf{P_n}$ is a subspace of $\mathbf{V}$, a $\mathbf{P_i}$, $1 \leq i \leq n$, contains all the other $\mathbf{P_j}$'s. Without loss of generality, assume such a $\mathbf{P_i}$ is $\mathbf{P_1}$. It follows that $\mathbf{W} = \mathbf{P_1}$. By part 2. of this problem, we know that $\mathbf{P_1}$ is invariant under $\mathcal{L}$. Therefore, $\mathbf{W}$ is invariant under $\mathcal{L}$.

Problem. 5. Invariant subspaces of commuting operators: Suppose that linear functions $\mathcal{L}, \mathcal{M} \in \mathfrak{L}(\mathbf{V})$ commute; i.e.,

$$\mathcal{L} \circ \mathcal{M} = \mathcal{M} \circ \mathcal{L}.$$

1. Argue that for any non-negative integer k, $\operatorname{Im}(\mathcal{M}^k)$ and $\ker(\mathcal{M}^k)$ are invariant subspace for $\mathcal{L}$.
2. Argue that every eigenspace of $\mathcal{M}$ is an invariant subspace for $\mathcal{L}$.
3. By giving a general example (with justification, of course!) show that for each $n > 1$ there are commuting matrices $\mathcal{A}$ and $\mathcal{B}$ in $\mathbb{M}_n$ such that

$$\mathfrak{Lat}(\mathcal{A}) \neq \mathfrak{Lat}(\mathcal{B}).$$

Solution. 5.

1. Because $\mathcal{L} \circ \mathcal{M} = \mathcal{M} \circ \mathcal{L}$, $\mathcal{L} \circ \mathcal{M}^k = \mathcal{M}^k \circ \mathcal{L}$ for any non-negative k. We can verify this by a short proof by induction. Assume that $\mathcal{L} \circ \mathcal{M}^k = \mathcal{M}^k \circ \mathcal{L}$ holds for any non-negative k. The base case where $k = 0$ is trivially true, for $\mathcal{L} \circ \mathcal{M}^0 = \mathcal{L} \circ \mathcal{I} = \mathcal{I} \circ \mathcal{L} = \mathcal{M}^0 \circ \mathcal{L}$, where $\mathcal{I}$ is the identity operator. The inductive case is also true because

$$\mathcal{L} \circ \mathcal{M}^{k+1} = \mathcal{L} \circ \mathcal{M}^k \circ \mathcal{M} = \mathcal{M}^k \circ \mathcal{L} \circ \mathcal{M} = \mathcal{M}^k \circ \mathcal{M} \circ \mathcal{L} = \mathcal{M}^{k+1} \circ \mathcal{L}.$$

   Therefore, by the principle of induction, $\mathcal{L} \circ \mathcal{M}^k = \mathcal{M}^k \circ \mathcal{L}$ is indeed true for all $k \geq 0$.

   (a) <u>To show</u>: $\mathcal{L}\left[\operatorname{Im}(\mathcal{M}^k)\right] \subseteq \operatorname{Im}(\mathcal{M}^k)$.

   Since $\mathcal{L} \circ \mathcal{M}^k = \mathcal{M}^k \circ \mathcal{L}$, $\mathcal{L} \circ \mathcal{M}^k[\mathbf{V}] = \mathcal{M}^k \circ \mathcal{L}[\mathbf{V}]$; i.e., $\mathcal{L}[\operatorname{Im}(\mathcal{M}^k)] = \mathcal{M}^k[\operatorname{Im}(\mathcal{L})]$. But since $\operatorname{Im}(\mathcal{L}) \subseteq \mathbf{V}$, $\mathcal{M}^k[\operatorname{Im}(\mathcal{L})] \subseteq \mathcal{M}^k[\mathbf{V}] = \operatorname{Im}(\mathcal{M}^k)$. It follows that $\mathcal{L}[\operatorname{Im}(\mathcal{M}^k)] \subseteq \operatorname{Im}(\mathcal{M}^k)$. So, $\operatorname{Im}(\mathcal{M}^k)$ is an invariant subspace for $\mathcal{L}$.

   (b) <u>To show</u>: $\mathcal{L}[\ker(\mathcal{M}^k)] \subseteq \ker(\mathcal{M}^k)$.

   Since $\mathcal{L} \circ \mathcal{M}^k = \mathcal{M}^k \circ \mathcal{L}$, $\mathcal{L} \circ \mathcal{M}^k[\ker(\mathcal{M}^k)] = \{\mathbf{0_V}\} = \mathcal{M}^k \circ \mathcal{L}[\ker(\mathcal{M}^k)]$. It follows that $\mathcal{L}[\ker(\mathcal{M}^k)] \subseteq \ker(\mathcal{M}^k)$. This completes the argument.

2. Let $\mathbf{E}_\lambda$ the eigenspace of $\mathcal{M}$ associated with eigenvalue λ be given. Consider $e \in \mathbf{E}_\lambda$, we have

$$\mathcal{M} \circ \mathcal{L}(e) = \mathcal{L} \circ \mathcal{M}(e) = \mathcal{L}(\lambda e) = \lambda \mathcal{L}(e)$$

   . This implies $\mathcal{L}(e) \in \mathbf{E}_\lambda$. Since this relation holds for any $e \in \mathbf{E}_\lambda$, $\mathcal{L}[\mathbf{E}_\lambda] \subseteq \mathbf{E}_\lambda$. Therefore, $\mathbf{E}_\lambda$ is an invariant subspace for $\mathcal{L}$ for any eigenvalue λ. This proves the claim that *every* eigenspace of $\mathcal{M}$ is an invariant subspace for $\mathcal{L}$.

3. It suffices to find commuting matrices $\mathcal{A}$ and $\mathcal{B}$ such that there exists an invariant subspace under $\mathcal{A}$, $\mathbf{J} \in \mathfrak{Lat}(\mathcal{A})$, but $\mathbf{J} \notin \mathfrak{Lat}(\mathcal{B})$.

   For $n = 2$, consider

$$\mathcal{A} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

   $\mathcal{A}$ and $\mathcal{B}$ commute:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

   Consider $\mathbf{J} = \operatorname{span}\{(1\ \ 1)^\top\}$. $\mathbf{J} \in \mathfrak{Lat}(\mathcal{A})$ since $\mathcal{A}[\mathbf{J}] = \{\mathbf{0_V}\} \subseteq \mathbf{J}$. However, consider $j = (1\ \ 1)^\top \in \mathbf{J}$, we have

$$\mathcal{B}(j) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \notin \operatorname{span}\{(1\ \ 1)^\top\},$$

   which implies $\mathbf{J} \notin \mathfrak{Lat}(\mathcal{B})$. Therefore, $\mathfrak{Lat}(\mathcal{A}) \neq \mathfrak{Lat}(\mathcal{B})$.

   Observe that we can give a more general example for any $n > 1$. Let $\mathcal{A}$ be an $n \times n$ zero matrix, and $\mathcal{B}$ is an $n \times n$ matrix of the form

$$\mathcal{B} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}.$$

   It is clear that $\mathcal{A}$ and $\mathcal{B}$ commute, since $\mathcal{A}$ is the zero matrix: $\mathcal{AB} = [0]_{n\times n} = \mathcal{BA}$.

   Consider $j = (1\ 1\ \dots\ 1)^\top \in \mathbf{J} = \operatorname{span}\{(1\ 1\ \dots\ 1)^\top\} \subseteq \mathbb{C}^n$. It is clear that $\mathcal{B}(j) = (1\ 0\ \dots\ 0)^\top \notin \mathbf{J}$ but $\mathcal{A}(j') = \mathbf{0_V} \in \mathbf{J}$ for any $j' \in \mathbf{J}$. So, $\mathbf{J} \in \mathfrak{Lat}(\mathcal{A})$ but $\mathbf{J} \notin \mathfrak{Lat}(\mathcal{B})$; i.e., $\mathfrak{Lat}(\mathcal{A}) \neq \mathfrak{Lat}(\mathcal{B})$.

Problem. 6. Invariant subspace chains: Suppose that $\mathcal{L} \in \mathfrak{L}(\mathbf{V})$ and $\mathbf{W}$ is an invariant subspace of $\mathcal{L}$.

1. Argue that the following inclusions hold:

   $\dots \subseteq \mathcal{L}^3[\mathbf{W}] \subseteq \mathcal{L}^2[\mathbf{W}] \subseteq \mathcal{L}[\mathbf{W}] \subseteq \mathbf{W} \subseteq \mathcal{L}^{-1}[\mathbf{W}] \subseteq \mathcal{L}^{-2}[\mathbf{W}] \subseteq \mathcal{L}^{-3}[\mathbf{W}] \subseteq \dots$

   Note that by Problem 3, all of the subspaces listed in the chain are invariant under $\mathcal{L}$.

2. Argue that the following implications hold:

   $$\mathcal{L}^{k+1}[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}] \implies \mathcal{L}^m[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}] \text{ for any } m \geq k,$$

   and

   $$\mathcal{L}^{-k}[\mathbf{W}] = \mathcal{L}^{-(k+1)}[\mathbf{W}] \implies \mathcal{L}^{-k}[\mathbf{W}] = \mathcal{L}^{-m}[\mathbf{W}] \text{ for any } m \geq k,$$

   and then explain why the chain in part 1 stabilizes in both directions, and has at most $\dim(\mathbf{V})$ proper inclusions.

3. Argue that any subspace $\mathbf{M}$ that falls between two consecutive subspaces in the chain shown in part 1. is also invariant under $\mathcal{L}$.

Solution. 6.

1. (a) To show: $\dots \subseteq \mathcal{L}^3[\mathbf{W}] \subseteq \mathcal{L}^2[\mathbf{W}] \subseteq \mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$.

   Assume that $\mathcal{L}^{k+1}[\mathbf{W}] \subseteq \mathcal{L}^k[\mathbf{W}]$ for any $k \geq 0$. The base case $k = 0$ is true because $\mathbf{W}$ is an invariant subspace of $\mathcal{L}$. The inductive case is also true because

   $$\mathcal{L}^{k+2}[\mathbf{W}] = \mathcal{L}[\mathcal{L}^{k+1}[\mathbf{W}]] \subseteq \mathcal{L}[\mathcal{L}^k[\mathbf{W}]] = \mathcal{L}^{k+1}[\mathbf{W}],$$

   where the second relation follows from assumption. By principle of induction, $\dots \subseteq \mathcal{L}^3[\mathbf{W}] \subseteq \mathcal{L}^2[\mathbf{W}] \subseteq \mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$.

   (b) To show: $\mathbf{W} \subseteq \mathcal{L}^{-1}[\mathbf{W}] \subseteq \mathcal{L}^{-2}[\mathbf{W}] \subseteq \mathcal{L}^{-3}[\mathbf{W}] \subseteq \dots$.

   Assume that $\mathcal{L}^{-k}[\mathbf{W}] \subseteq \mathcal{L}^{-k-1}[\mathbf{W}]$ for all $k \geq 0$. The base case where $k = 0$ is true: Given $w \in \mathbf{W} \subseteq \mathbf{V}$, $\mathcal{L}(w) \in \mathbf{W}$ because $\mathbf{W}$ is invariant under $\mathcal{L}$. But since $\mathcal{L}^{-1}[\mathbf{W}] = \{v \in \mathbf{V} | \mathcal{L}(v) \in \mathbf{W}\}$, $w \in \mathcal{L}^{-1}[\mathbf{W}]$. Therefore, $\mathbf{W} \in \mathcal{L}^{-1}[\mathbf{W}]$ (i).

   The inductive case is also true:

   $$\mathcal{L}^{-k-1}[\mathbf{W}] = \mathcal{L}^{-1}[\mathcal{L}^{-k}[\mathbf{W}]] \subseteq \mathcal{L}^{-1}[\mathcal{L}^{-k-1}[\mathbf{W}]] \subseteq \mathcal{L}^{-k-2}[\mathbf{W}].$$

   So by the principle of induction, $\mathbf{W} \subseteq \mathcal{L}^{-1}[\mathbf{W}] \subseteq \mathcal{L}^{-2}[\mathbf{W}] \subseteq \mathcal{L}^{-3}[\mathbf{W}] \subseteq \dots$ (ii).

   From (i) and (ii),

   $$\dots \subseteq \mathcal{L}^2[\mathbf{W}] \subseteq \mathcal{L}[\mathbf{W}] \subseteq \mathbf{W} \subseteq \mathcal{L}^{-1}[\mathbf{W}] \subseteq \mathcal{L}^{-2}[\mathbf{W}] \subseteq \dots$$

2. (a) Assume the hypothesis that $\mathcal{L}^{k+1}[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}]$ for all $k \geq 0$. Also assume that $\mathcal{L}^{k+j}[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}]$ for all non-negative j. The base case where $j = 0$ is trivially true. The inductive case is also true since:

$$\mathcal{L}^{k+j+1}[\mathbf{W}] = \mathcal{L}[\mathcal{L}^{k+j}[\mathbf{W}]] = \mathcal{L}[\mathcal{L}^k[\mathbf{W}]] = \mathcal{L}^{k+1}[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}].$$

By the principle of induction, $\mathcal{L}^m[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}]$ must hold for any $m \geq k \geq 0$ if $\mathcal{L}^{k+1}[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}]$.

(b) Assume that hypothesis that $\mathcal{L}^{-k}[\mathbf{W}] = \mathcal{L}^{-(k+1)}[\mathbf{W}]$ for all $k \geq 0$. Also assume that $\mathcal{L}^{-k}[\mathbf{W}] = \mathcal{L}^{-(k+j)}[\mathbf{W}]$ for all non-negative j. The base case where $j = 0$ is trivially true. In inductive case is also true since:

$$\begin{aligned}\mathcal{L}^{-(k+j+1)}[\mathbf{W}] &= \mathcal{L}^{-1}[\mathcal{L}^{-(k+j)}[\mathbf{W}] \\ &= \mathcal{L}^{-1}[\mathcal{L}^{-k}[\mathbf{W}]] \\ &= \mathcal{L}^{-(k+1)}[\mathbf{W}] \\ &= \mathcal{L}^{-k}[\mathbf{W}].\end{aligned}$$

By the principle of induction, $\mathcal{L}^{-k}[\mathbf{W}] = \mathcal{L}^{-m}[\mathbf{W}]$ must hold for $m \geq k$ if $\mathcal{L}^{-k}[\mathbf{W}] = \mathcal{L}^{-(k+1)}[\mathbf{W}]$.

(c) It follows from part 1. that

$$\cdots \leq \dim(\mathcal{L}[\mathbf{W}]) \leq \dim(\mathbf{W}) \leq \dim(\mathcal{L}^{-1}[\mathbf{W}]) \leq \ldots$$

Because $\dim(\mathcal{L}^{-k}[\mathbf{W}]) \leq \dim(\mathbf{V})$ for any $k \geq 0$, there exists an $\mathcal{L}^{-h}[\mathbf{W}]$, $h \geq 0$, that is an invariant subspace in the chain with the highest dimension where h is maximal. It immediately follows that $\dim(\mathcal{L}^{-(h+1)}[\mathbf{W}]) = \dim(\mathcal{L}^{-(h)}[\mathbf{W}])$ is maximal. But because $\mathcal{L}^{-h}[\mathbf{W}] \subseteq \mathcal{L}^{-(h+1)}[\mathbf{W}]$, $\mathcal{L}^{-h}[\mathbf{W}] = \mathcal{L}^{-(h+1)}[\mathbf{W}]$. As a result, $\mathcal{L}^{-h}[\mathbf{W}] = \mathcal{L}^{-j}[\mathbf{W}]$ for any $j \geq h$, which follows from (b). Therefore, the chain in part 1. stabilizes in the "inclusion" direction.

A similar argument can be given to show that the chain in part 1. also stabilizes in the "inclusion by" direction. Since $0 \leq \dim(\mathcal{L}^l[\mathbf{W}])$ for any $l \geq 0$, there exists an $\mathcal{L}^l[\mathbf{W}]$ in the chain with the lowest dimension but l is minimal. It immediately follows that $\dim(\mathcal{L}^{(l+1)}[\mathbf{W}]) = \dim(\mathcal{L}^l[\mathbf{W}])$ is minimal. But because $\mathcal{L}^{(l+1)}[\mathbf{W}] \subseteq \mathcal{L}^l[\mathbf{W}]$, $\mathcal{L}^{(l+1)}[\mathbf{W}] = \mathcal{L}^l[\mathbf{W}]$. As a result, $\mathcal{L}^k[\mathbf{W}] = \mathcal{L}^j[\mathbf{W}]$ for any $k \leq l$, which follows from (a). Therefore, the chain in part 1. stabilizes in the "inclusion by" direction as well.

For every *proper* inclusion, there is a dimension loss. Revisiting $\mathcal{L}^{-h}[\mathbf{W}]$, we know that $\dim(\mathcal{L}^{-h}[\mathbf{W}]) \leq \dim(\mathbf{V})$. Consider the first proper inclusion:

$$\dots \mathcal{L}^{-(h+j)}[\mathbf{W}] \subsetneq \mathcal{L}^{-(h+j-1)}[\mathbf{W}] \subseteq \dots \subseteq \mathcal{L}^{-h}[\mathbf{W}] \subseteq \dots$$

This implies $\dim(\mathcal{L}^{-(h+j)}[\mathbf{W}]) < \dim(\mathcal{L}^{-h}[\mathbf{W}]) \leq \dim(\mathbf{V})$. Suppose $\dim(\mathcal{L}^{-h}[\mathbf{W}]) = \dim(\mathbf{V})$, i.e., maximal, then

$$\dim(\mathcal{L}^{-(h+j)}[\mathbf{W}]) \leq \dim(\mathbf{V}) - 1.$$

By assuming equality is met and consider the next proper inclusion in the chain, a similar argument shows that the invariant subspace of interest will have at most $\dim(\mathbf{V}) - 2$ dimensions. It follows that, at the $\dim(\mathbf{V})^{\text{th}}$ proper inclusion, the considered invariant subspace has at most $\dim(\mathbf{V}) - \dim(\mathbf{V}) = 0$ dimensions; i.e. the subspace is $\{\mathbf{0_V}\}$. At this point, no further proper inclusion is possible. Therefore, the chain in part 1. has at most $\dim(\mathbf{V})$ proper inclusion.

3. Let $\mathcal{L}^{(k+1)}[\mathbf{W}] \subseteq \mathbf{M} \subseteq \mathcal{L}^{k}[\mathbf{W}]$ be given for some integer k. $\mathbf{M}$ is a subspace. Observe that

$$\mathcal{L}[\mathbf{M}] \subseteq \mathcal{L}[\mathcal{L}^{k}[\mathbf{W}]] = \mathcal{L}^{(k+1)}[\mathbf{W}] \subseteq \mathbf{M}.$$

Therefore, $\mathbf{M}$ is invariant under $\mathcal{L}$.

25.3 Problem set 3