MATRIX ANALYSIS

A Quick Guide

Huan Q. Bui

Colby College

PHYSICS & MATHEMATICS
Statistics

Class of 2021

May 8, 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/Operators & Their Properties

1. **Normal**: $\mathcal{L}$ is normal $\iff \mathcal{L}^*\mathcal{L} = \mathcal{L}\mathcal{L}^*$.

2. **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.*

(d) H is Hermitian $\iff$ H is normal and $\sigma_{\mathbb{C}}(H) \subset \mathbb{R}$.

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 is normal and $\sigma_{\mathbb{C}}(U) \subseteq \Pi = \{e^{i\theta|\theta\in\mathbb{R}}\}$.

(b) U preserves the inner product (is an isometry)

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

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

(d) U is diagonalizable:

$$U = VDV^*,$$

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

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

(f) Its eigenspaces are orthogonal.

(g) U can be written as

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

where H is a Hermitian matrix.

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

3. **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

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

4. **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.

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

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

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

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 $\iff \ker(F) = V \iff \mathrm{Im}(F) = \mathbf{0}_W$

2. $\ker(F)$ is a subspace of V

3. $\mathrm{Im}(F)$ 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

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

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. $\mathbf{e}_j \in \ker(V_1 \dots V_m) \iff V_j = \mathbf{0}_V$, where $\mathbf{e}_j$ denotes a standard basis tuple with a 1 at the j$^{\text{th}}$ 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. $\operatorname{Im}(f) \equiv \operatorname{Im}(V_1 \dots V_m) = \operatorname{span}(V_1 \dots V_m)$

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

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:

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

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

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

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

10. Consider an atrix $A : \mathbb{R}^m \to V$. $A = (V_1 \dots V_m)$. 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: $F \circ E^{[i] \overset{+}{\leftarrow} \alpha \cdot [j]}$.
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. $\text{span}(X_1 \dots X_m), X_j \in V$ is a subspace of V.
2. $\text{span}(X_1 \dots X_j) \subseteq \text{span}(X_1 \dots X_k)$ 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 $\text{Im}(A)$. $i = 1, 2, \dots, m$.
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 $\text{Im}(X_1 \dots X_m) = V$, 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 $\text{Im}(A)$

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.

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

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

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

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

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

Subspace addition is associative.

Subspace addition is commutative.

6.2 Products and Direct Sums

Here is a connection between direct sums and products of vector spaces. Suppose $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are subspaces of $\mathbf{W}$ and $z_1 \in \mathbf{Z}_1$, $z_2 \in \mathbf{Z}_2$. Consider the function $✠ : \mathbf{Z}_1 \times \mathbf{Z}_2 \overset{\text{linear}}{\longrightarrow} \mathbf{Z}_1 \oplus \mathbf{Z}_2 \prec \mathbf{W}$ defined by

$$✠ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \triangleq z_1 + z_2. \tag{1}$$

We can show that

1. ✠ is a linear function.

2. ✠ is an isomorphism, i.e., $\mathbf{Z}_1 \times \mathbf{Z}_2 \sim \mathbf{Z}_1 \oplus \mathbf{Z}_2$.

6.3 Products of Vector Spaces

Definition 6.2. Product of vectors spaces

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

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

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

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

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

Properties 6.2.

1. Product of vectors spaces is a vector space.

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

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:

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

6.4 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 6.3.

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. $\square$

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

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

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. $\square$

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:

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

Suppose $v \in \ker T^n \cap \mathrm{Im}\, T^n$. 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 $v = 0$. $\square$

5. IT IS NOT TRUE THAT $V = \ker(T) \oplus \mathrm{Im}(T)$ 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 6.4.

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 6.5.

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

Definition 7.1. An **Operator** is a linear function from a vector space to itself:

$$\mathcal{E} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}.$$

Definition 7.2. Idempotents: are operators with the property $\mathcal{E}^2 = \mathcal{E}$, i.e.,

$$\mathcal{E} \circ \mathcal{E} = \mathcal{E}.$$

Recall that if $V = \mathbf{W} \oplus \mathbf{Z}$, then there exists an idempotent $\mathcal{E} \in \mathcal{L}(\mathbf{V})$ such that

$$\begin{aligned} \mathbf{W} &= \operatorname{Im}(\mathcal{E}) \\ \mathbf{Z} &= \ker(\mathcal{E}). \end{aligned}$$

In fact, if $\mathbf{V} = \mathbf{W} \oplus \mathbf{Z}$ then there exists at least **two** idempotents $\mathcal{E}$, $\mathbf{W} = \operatorname{Im}(\mathcal{E}), \mathbf{Z} = \ker(\mathcal{E})$ and $\mathcal{F}$, $\mathbf{W} = \ker(\mathcal{F})$, $\mathbf{Z} = \operatorname{Im}(\mathcal{F})$.

A natural question to ask now is whether every idempotent $\mathcal{E} \in \mathcal{L}(\mathbf{V})$ can be generated this way. The answer is **Yes**, but to answer to this, we need to show **existence** and **uniqueness**.

Theorem 7.1. If $\mathcal{E} \circ \mathcal{E} = \mathcal{E}^2 = \mathcal{E}$, then $\mathbf{V} = \operatorname{Im}(\mathcal{E}) \oplus \ker(\mathcal{E})$.

Proof.

1. Show that $\operatorname{Im}(\mathcal{E}) + \ker(\mathcal{E}) = \mathbf{V}$.

   Let $v \in \mathbf{V}$ be given, then $v = \mathcal{E}(v) + v - \mathcal{E}(v)$. Now, observe that

   $$\mathcal{E}(v - \mathcal{E}(v)) = \mathcal{E}(v) - \mathcal{E}^2(v) = \mathcal{E}(v) - \mathcal{E}(v) = 0.$$

   Therefore, $\mathcal{E}(v) \in \operatorname{Im}(\mathcal{E})$ and $(v - \mathcal{E}(v)) \in \ker(\mathcal{E})$. So, $\operatorname{Im}(\mathcal{E}) + \ker(\mathcal{E}) = \mathbf{V}$.

2. Show that $\operatorname{Im}(\mathcal{E}) \oplus \ker(\mathcal{E}) = \mathbf{V}$.

   We want to show that $\operatorname{Im}(\mathcal{E}) \bigcap \ker(\mathcal{E}) = \{\mathbf{0}\}$. So, let $v \in \operatorname{Im}(\mathcal{E}) \bigcap \ker(\mathcal{E})$ be given. Then $v \in \operatorname{Im}(\mathcal{E})$, which implies $\mathcal{E}(\mathcal{E}(x)) = \mathcal{E}(x) = v = \mathbf{0}$ for some x. Hence, the intersection is the trivial subspace.

$\square$

$\square$

Properties 7.1. A by product of this previous item is this fact. For any $x \in \mathbf{V}$

$$\mathcal{E}(\mathcal{E}(x)) = \mathcal{E}(x) \iff \mathcal{E} = \mathcal{E}^2, \tag{2}$$

i.e., $\mathcal{E}$ is an idempotent exactly when it acts as an identity function on its own image.

Theorem 7.2. If $\mathcal{E}^2 = \mathcal{E}$ and $\mathcal{G}^2 = \mathcal{G}$ in $\mathcal{L}(\mathbf{V})$ and

$$\begin{aligned}\mathrm{Im}(\mathcal{G}) &= \mathrm{Im}(\mathcal{E})\\ \ker(\mathcal{G}) &= \ker(\mathcal{E})\end{aligned}$$

then

$$\mathcal{E} = \mathcal{G}.$$

Proof. Let $\mathbf{W} = \mathrm{Im}(\mathcal{E}) = \mathrm{Im}(\mathcal{G})$ and $\mathbf{Z} = \ker(\mathcal{E}) = \ker(\mathcal{G})$. Note that $\mathbf{W} \oplus \mathbf{Z} = \mathbf{V}$. Consider $v \in \mathbf{V}$. Then

$$\mathcal{E}(v) = \mathcal{E}(w+z) = \mathcal{E}(w) + \mathcal{E}(z) = \mathcal{E}(w) + \mathbb{0} = \mathcal{E}(w) = w,$$

where $v = z + w$ is unique by directness. We can do the same thing with G, and get $G = E$. □

□

Theorem 7.3. For each decomposition $V = \mathbf{W} \oplus \mathbf{Z}$, there exists a unique idempotent whose image is $\mathbf{W}$ and kernel is $\mathbf{Z}$. □

Now, notice that for $v = w + z$, we have $v = w + z = \mathcal{E}(w) + \mathcal{F}(z) = \mathcal{E}(v) + \mathcal{F}(v) = Id(v)$. So we have a little theorem:

Theorem 7.4. If $\mathcal{E}$ and $\mathcal{F}$ are idempotents such that $\mathrm{Im}(\mathcal{E}) = \ker(\mathcal{F})$ and $\ker(\mathcal{E}) = \mathrm{Im}(\mathcal{F})$ then

$$\mathcal{E} + \mathcal{F} = Id.$$

□

As a result we get another two results:

1. If $\mathcal{E}^2 = \mathcal{E} \in \mathcal{L}(\mathbf{V})$ then $(Id - \mathcal{E})^2 = (Id - \mathcal{E})$ also an idempotent.

2. $$\begin{aligned}\mathrm{Im}(\mathcal{E}) &= \ker(Id - \mathcal{E})\\ \ker(\mathcal{E}) &= \mathrm{Im}(Id - \mathcal{E})\end{aligned}$$

We say that in this case idempotents come in pairs. So, to informally summarize, if $V = \mathbf{W} \oplus \mathbf{Z}$, then there is a pair of idempotents. More concretely, if $V = \mathbf{W} \oplus \mathbf{Z}$, then there exist unique idempotents $\mathcal{E}$ and $\mathcal{F}$ such that $\mathcal{E} + \mathcal{F} = Id$ and $\mathbf{W} = \mathrm{Im}(\mathcal{E}) = \ker(\mathcal{F})$ and $\mathbf{Z} = \ker(\mathcal{E}) = \mathrm{Im}(\mathcal{F})$. The converse is also true. If there exist idempotents $\mathcal{E}$ and $\mathcal{F}$ such that if $\mathcal{E} + \mathcal{F} = Id$ and $\mathbf{W} = \mathrm{Im}(\mathcal{E}) = \ker(\mathcal{F})$ and $\mathbf{Z} = \ker(\mathcal{E}) = \mathrm{Im}(\mathcal{F})$ then $\mathbf{V} = \mathbf{W} \oplus \mathbf{Z}$.

All is good, but what if the vector space $\mathbf{V}$ is decomposed into three pieces as $\mathbf{V} = \mathbf{W} \oplus \mathbf{Z} \oplus \mathbf{U}$? Well, we simply repeat what we have done before to

$\mathbf{V} = \mathbf{W} \oplus (\mathbf{Z} \oplus \mathbf{U})$. We know that there exist a unique $\mathcal{E}_{\mathbf{W}}$ such that $\mathcal{E}_{\mathbf{W}}^2 = \mathcal{E}_{\mathbf{W}}$ and that $\mathrm{Im}(\mathcal{E}_{\mathbf{W}}) = \mathbf{W}$ and $\ker(\mathcal{E}_{\mathbf{W}}) = \mathbf{U} \oplus \mathbf{Z}$. Similarly, there exist a unique $\mathcal{E}_{\mathbf{U}}$ such that $\mathcal{E}_{\mathbf{U}}^2 = \mathcal{E}_{\mathbf{U}}$ and that $\mathrm{Im}(\mathcal{E}_{\mathbf{U}}) = \mathbf{V}$ and $\ker(\mathcal{E}_{\mathbf{U}}) = \mathbf{W} \oplus \mathbf{Z}$; and there exist a unique $\mathcal{E}_{\mathbf{W}}$ such that $\mathcal{E}_{\mathbf{W}}^2 = \mathcal{E}_{\mathbf{W}}$ and that $\mathrm{Im}(\mathcal{E}_{\mathbf{W}}) = \mathbf{W}$ and $\ker(\mathcal{E}_{\mathbf{W}}) = \mathbf{U} \oplus \mathbf{Z}$.

There are a few observations we can make:

1. Let $v \in \mathbf{V}$ be given, then $(\mathcal{E}_{\mathbf{W}} + \mathcal{E}_{\mathbf{U}} + \mathcal{E}_{\mathbf{Z}})(v) = (\mathcal{E}_{\mathbf{W}} + \mathcal{E}_{\mathbf{U}} + \mathcal{E}_{\mathbf{Z}})(w+u+z) = w + u + z = v$. So $\mathcal{E}_{\mathbf{W}} + \mathcal{E}_{\mathbf{U}} + \mathcal{E}_{\mathbf{Z}} = Id$.

2. $\mathcal{E}_{\mathbf{W}} \circ \mathcal{E}_{\mathbf{Z}}(v) = \mathcal{E}_{\mathbf{W}}(z) = \mathbf{0}_{\mathbf{V}}$. So, in general, $\mathcal{E}_i \circ \mathcal{E}_j = \mathcal{O}$ for $i \neq j$ and $\mathcal{E}_i$ if $i = j$.

What about the converse? Given the linear functions such that $\mathcal{E}_j^2 = \mathcal{E}$ such that $\mathrm{Im}(\mathcal{E}_i) = \ker(\mathcal{E}_j + \mathcal{E}_k)$, then does $V = \mathbf{Z} \oplus \mathbf{W} + \mathbf{U}$ hold? The answer is yes! We can restate this for rigorously:

Theorem 7.5. Suppose $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ are idempotents such that $\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 = Id$ and $\mathcal{E}_i \circ \mathcal{E}_j = \delta_i^j \mathcal{E}_i$ then

$$\mathrm{Im}(\mathcal{E}_1) \oplus \mathrm{Im}(\mathcal{E}_2) \oplus \mathrm{Im}(\mathcal{E}_3) = \mathbf{V}$$

and

$$\ker(\mathcal{E}_1) = \mathrm{Im}(\mathcal{E}_2) \oplus \mathrm{Im}(\mathcal{E}_3),$$

etc. □

The following two statements are equivalent:

1. $\mathbf{W} \oplus \mathbf{Z} \oplus \mathbf{U} = \mathbf{V}$.

2. There are idempotents $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3 \in \mathcal{L}(\mathbf{V})$ such that

   $$\mathrm{Im}(\mathcal{E}_1) = \mathbf{W}, \mathrm{Im}(\mathcal{E}_2) = \mathbf{Z}, \mathrm{Im}(\mathcal{E}_3) = \mathbf{U}$$

   and

   $$\begin{aligned} &\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 = Id \\ &\mathcal{E}_j \circ \mathcal{E}_i = \delta_i^j \mathcal{E}_i \end{aligned}$$

Proof.
- The forward direction has been shown.
- We will show how (2) implies (1). We first show that $\mathrm{Im}(\mathcal{E}_1) + \mathrm{Im}(\mathcal{E}_2) + \mathrm{Im}(\mathcal{E}_3) = \mathbf{V}$, then

  $$\mathbf{V} \ni v = Id(v) = (\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3)(v) = \mathcal{E}_1(v) + \mathcal{E}_2(v) + \mathcal{E}_3(v) \in \mathrm{Im}(\mathcal{E}_1) + \mathrm{Im}(\mathcal{E}_2) + \mathrm{Im}(\mathcal{E}_3).$$

  It is straightforward to show that $V = \mathrm{Im}(\mathcal{E}_1) + \mathrm{Im}(\mathcal{E}_2) + \mathrm{Im}(\mathcal{E}_3)$. To show directness, suppose that $x_1 + x_2 + x_3 = \mathbf{0}$, $x_i \in \mathrm{Im}(\mathcal{E}_i)$. We want to show that $x_i = \mathbf{0}$ for any $i \in \{1,2,3\}$. By construction, we have that

$$\mathcal{E}_1(x_1) + \mathcal{E}_2(x_2) + \mathcal{E}_3(x_3) = \mathbf{0},$$

so

$$\mathcal{E}_1\left(\mathcal{E}_1(x_1) + \mathcal{E}_2(x_2) + \mathcal{E}_3(x_3)\right) = \mathcal{E}_1(\mathbf{0}) = \mathbf{0}.$$

But this implies that

$$\mathcal{E}_1(\mathcal{E}_1(x_1)) = \mathcal{E}_1(x_1) = x_1 = \mathbf{0}.$$

Similarly, we get $x_j = \mathbf{0}$ for all j. By one of the equivalent statements about directness, we get that $\mathrm{Im}(\mathcal{E}_1) \oplus \mathrm{Im}(\mathcal{E}_2) \oplus \mathrm{Im}(\mathcal{E}_3) = \mathbf{V}$.

□

We can also (of course) look at the kernel. Observe that $\mathcal{E}_1(\mathcal{E}_2(v)) = \mathbf{0}$. So $\mathrm{Im}(\mathcal{E}_2) \subseteq \ker(\mathcal{E}_1)$. Similarly, $\mathrm{Im}(\mathcal{E}_3) \subseteq \ker(\mathcal{E}_1)$. So, $\mathrm{Im}(\mathcal{E}_3) + \mathrm{Im}(\mathcal{E}_2) \subseteq \ker(\mathcal{E}_1)$. But we also know that the images of $\mathcal{E}_1$ and E_2 intersect trivially, so this sum is direct. So we have $\mathrm{Im}(\mathcal{E}_3) \oplus \mathrm{Im}(\mathcal{E}_2) \subseteq \ker(\mathcal{E}_1)$. But we also know that $V = \mathrm{Im}(\mathcal{E}_1) \oplus \ker(\mathcal{E}_1) = \mathrm{Im}(\mathcal{E}_1) \oplus \mathrm{Im}(\mathcal{E}_2) \oplus \mathrm{Im}(\mathcal{E}_3)$, so

$$\mathrm{Im}(\mathcal{E}_2) \oplus \mathrm{Im}(\mathcal{E}_3) = \ker(\mathcal{E}_1)$$

as desired.

So, just a recap of what we have done so far, the following statements are equivalent for $\mathcal{A} \in \mathcal{L}(\mathbf{V})$:

1. $\mathcal{A}^2 = \mathcal{A}$.
2. $\mathcal{A}(x) = x$ for $x \in \mathrm{Im}(\mathcal{A})$.
3. $(Id - \mathcal{A})^2 = (Id - \mathcal{A})$.
4. $\mathrm{Im}(\mathcal{A}) = \ker(Id - \mathcal{A})$.
5. $\mathrm{Im}(Id - \mathcal{A}) = \ker(\mathcal{A})$.

8 Block-representations of operators

8.1 Coordinatization

Consider a finite-dimensional linear space V with basis $\{V_i\}$, $i = 1, 2, \dots, m$. 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

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

8.2 Matricial representation of linear functions

8.2.1 In Cartesian products

Let us start with a Cartesian product of two vector spaces and a linear function $\mathcal{L}$ mapping this Cartesian product to itself $\mathcal{L} : \mathbf{V} \times \mathbf{W} \overset{\text{linear}}{\longrightarrow} \mathbf{V} \times \mathbf{W}$, where a typical element in $\mathbf{V} \times \mathbf{W}$ is given by

$$\begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} v \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ w \end{pmatrix}.$$

Formally, we can represent the linear function $\mathcal{L}$ as a block-matrix function

$$\left(\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right) \begin{pmatrix} v \\ w \end{pmatrix} \triangleq \begin{pmatrix} F_{11}(v) + F_{12}(w) \\ F_{21}(v) + F_{22}(w) \end{pmatrix}.$$

So we have

$$F_{11} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$$
$$F_{12} : \mathbf{W} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$$
$$F_{21} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{W}$$
$$F_{22} : \mathbf{W} \overset{\text{linear}}{\longrightarrow} \mathbf{W}.$$

It can be illuminating if we show what the linear functions F_{ij} does in a block-matrix form:

$$\begin{array}{c|c|c|} & \mathbf{V} & \mathbf{W} \\ \hline \mathbf{V} & \text{↲} & \text{↲} \\ \mathbf{W} & \text{↲} & \text{↲} \end{array}.$$

Now, since F_{ij} are all linear, the block-matrix function

$$\left(\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right) : \mathbf{V} \times \mathbf{W} \to \mathbf{V} \times \mathbf{W}$$

is linear as well.

A natural question to ask would be: "Can every linear function be represented this way?" The answer is Yes, and we shall see how this is done.

Notice that

$$\mathcal{L}\begin{pmatrix} v \\ \mathbf{0} \end{pmatrix} = \left(\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right)\begin{pmatrix} v \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} F_{11}(v) \\ F_{21}(v) \end{pmatrix}. \tag{3}$$

In particular,

$$\Pi_{\mathbf{V}} \circ \mathcal{L}\begin{pmatrix} v \\ \mathbf{0} \end{pmatrix} = F_{11}(v), \tag{4}$$

where $\Pi_{\mathbf{V}} : \mathbf{V} \times \mathbf{W} \to \mathbf{V}$ is a coordinate projection from $\mathbf{V} \times \mathbf{W}$ onto $\mathbf{V}$. In fact, we can also defined the linear function $\gamma_{\mathbf{V}}$ that is a coordinate injection from $\mathbf{V}$ into $\mathbf{V} \times \mathbf{W}$. So, given $\mathcal{L} : \mathbf{V} \times \mathbf{W} \overset{\text{linear}}{\longrightarrow} \mathbf{V} \times \mathbf{W}$, we let:

$$F_{11} := \Pi_{\mathbf{V}} \circ \mathcal{L} \circ \gamma_{\mathbf{V}} \tag{5}$$
$$F_{12} := \Pi_{\mathbf{V}} \circ \mathcal{L} \circ \gamma_{\mathbf{W}} \tag{6}$$
$$F_{21} := \Pi_{\mathbf{W}} \circ \mathcal{L} \circ \gamma_{\mathbf{V}} \tag{7}$$
$$F_{22} := \Pi_{\mathbf{W}} \circ \mathcal{L} \circ \gamma_{\mathbf{W}}. \tag{8}$$

Now, we can easily check that $\mathcal{L}$ is indeed

$$\left(\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right). \tag{9}$$

We do this by considering $\begin{pmatrix} v & w \end{pmatrix}^{\top} \in \mathbf{V} \times \mathbf{W}$. We know that

$$F_{11}(v) + F_{12}(w) = \Pi_1 \circ \mathcal{L}\begin{pmatrix} v \\ \mathbf{0} \end{pmatrix} + \Pi_1 \circ \mathcal{L}\begin{pmatrix} \mathbf{0} \\ w \end{pmatrix} = \Pi_1 \circ \mathcal{L}\begin{pmatrix} v \\ w \end{pmatrix} \tag{10}$$
$$F_{21}(v) + F_{22}(w) = \Pi_2 \circ \mathcal{L}\begin{pmatrix} v \\ \mathbf{0} \end{pmatrix} + \Pi_2 \circ \mathcal{L}\begin{pmatrix} \mathbf{0} \\ w \end{pmatrix} = \Pi_2 \circ \mathcal{L}\begin{pmatrix} v \\ w \end{pmatrix}. \tag{11}$$

So,

$$\begin{pmatrix} F_{11}(v) + F_{12}(w) \\ F_{21}(v) + F_{22}(w) \end{pmatrix} = \begin{pmatrix} \Pi_1 \circ \mathcal{L}\begin{pmatrix} v \\ w \end{pmatrix} \\ \Pi_2 \circ \mathcal{L}\begin{pmatrix} v \\ w \end{pmatrix} \end{pmatrix} = \mathcal{L}\begin{pmatrix} v \\ w \end{pmatrix}.$$

So, if we start with any $\mathcal{L}$, we can represent $\mathcal{L}$ as a matrix of the linear functions F_{ij}.

8.2.2 In direct sums

We don't often work with Cartesian products. So we want to use the idea developed above to break a vector space into direct sums. Suppose that we are given $V = \mathbf{Z} \oplus \mathbf{W}$. Let us use the bad notation

$$\begin{pmatrix} w \\ z \end{pmatrix}_+ = w + z,$$

so that we can mimic our previous idea. Recall the idea

$$\begin{array}{c|c|c|} & \mathbf{V} & \mathbf{W} \\ \hline \mathbf{V} & \text{↲} & \text{↲} \\ \mathbf{W} & \text{↲} & \text{↲} \end{array},$$

given

$$\begin{aligned} F_{11} &: \mathbf{W} \to \mathbf{W} \\ F_{12} &: \mathbf{Z} \to \mathbf{W} \\ F_{21} &: \mathbf{W} \to \mathbf{Z} \\ F_{22} &: \mathbf{Z} \to \mathbf{Z}, \end{aligned}$$

we can define a linear function

$$\left(\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right) : \mathbf{V} \to \mathbf{V}$$

by

$$\left(\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right) \begin{pmatrix} w \\ z \end{pmatrix}_+ = \begin{pmatrix} F_{11}(w) + F_{12}(z) \\ F_{21}(w) + F_{22}(z) \end{pmatrix}_+ .$$

Now, we want to find something that is similar to Π_i defined before. Since we are dealing with direct sums rather than Cartesian products, **idempotents** are perfect candidates. In fact, if

$$\mathcal{L} = \left(\begin{array}{c|c} F_{11} & F_{12} \\ \hline F_{21} & F_{22} \end{array}\right),$$

then

$$\mathcal{L} \begin{pmatrix} w \\ \mathbf{0} \end{pmatrix}_+ = \begin{pmatrix} F_{11}(w) \\ F_{21}(w) \end{pmatrix}_+$$

So it is easy to see that

$$\begin{aligned} \mathcal{E} \circ \mathcal{L} \begin{pmatrix} w & \mathbf{0} \end{pmatrix}^\top &= F_{11}(w) \in \mathbf{W} \\ (Id - \mathcal{E}) \circ \mathcal{L} \begin{pmatrix} w & \mathbf{0} \end{pmatrix}^\top &= F_{21}(w) \in \mathbf{Z}. \end{aligned}$$

The same argument applies to find F_{12} and F_{22}. So we can once again check that

$$\mathcal{L} = \left(\begin{array}{c|c} \mathcal{E} \circ \mathcal{L}\Big|_{\mathbf{W}} & \mathcal{E} \circ \mathcal{L}\Big|_{\mathbf{Z}} \\ \hline (Id - \mathcal{E}) \circ \mathcal{L}\Big|_{\mathbf{W}} & (Id - \mathcal{E}) \circ \mathcal{L}\Big|_{\mathbf{Z}} \end{array}\right).$$

So, if we have a linear function $\mathcal{L} : \mathbf{V} \to \mathbf{V}$ and with respect to the decomposition $|V = \mathbf{W} \oplus \mathbf{Z}$ and that $\mathcal{L}$ has the block form $\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$ then it means

$$A := \mathcal{E} \circ \mathcal{L}\Big|_{\mathbf{W}} : \mathbf{W} \overset{\text{linear}}{\longrightarrow} \mathbf{W},$$

and so on, where $\mathcal{E}$ is an idempotent with $\mathrm{Im}(\mathcal{E}) = \mathbf{W}$ and $\ker(\mathcal{E}) = \mathbf{Z}$.

There is a little caveat: direct sum is commutative, BUT here the order of $\mathbf{W}$ and $\mathbf{Z}$ matters in the definition of the block-matrix representation of $\mathcal{L}$.

Example 8.1. Consider the linear function $G : \mathbf{V} \to \mathbf{V}$ an idempotent, such that $\mathbf{V} = \mathrm{Im}(G) \oplus \ker(G)$. So,

$$G = \left(\begin{array}{c|c} Id\Big|_{\mathrm{Im}(G)} & \mathcal{O} \\ \hline \mathcal{O} & \mathcal{O} \end{array}\right).$$

□

8.3 Properties of Block-matrix representation

We will show in the homework the following properties:

1. **Addition:** If

$$L = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

and

$$S = \left(\begin{array}{c|c} P & Q \\ \hline R & T \end{array}\right)$$

then

$$L + S = \left(\begin{array}{c|c} A + P & B + Q \\ \hline C + R & D + T \end{array}\right).$$

It is not too hard to show why this is true:

$$[L + S]_{11}(w) = \mathcal{E}[L + S](w) = \mathcal{E} \circ L(w) + \mathcal{E} \circ S(w) = A(w) + P(w) = [A + P](w).$$

2. **Scaling:**

$$\alpha \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) = \left(\begin{array}{c|c} \alpha A & \alpha B \\ \hline \alpha C & \alpha D \end{array}\right).$$

3. **Composition** also works nicely:

$$\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \circ \left(\begin{array}{c|c} P & Q \\ \hline R & T \end{array}\right) = \left(\begin{array}{c|c} A \circ P + B \circ R & A \circ Q + B \circ T \\ \hline C \circ P + D \circ R & C \circ Q + D \circ T \end{array}\right). \tag{12}$$

It is also quite straightforward to show why this is true. Consider $[L \circ S]_{11}$:

$$\begin{aligned} [L \circ S]_{11}(w) &= \mathcal{E}[L \circ S](w) \\ &= \mathcal{E} \circ L[\mathcal{E} + (Id. - \mathcal{E})] \circ S(w) \\ &= \mathcal{E} \circ L \circ \mathcal{E} \circ S(w) + \mathcal{E} \circ L(Id. - \mathcal{E}) \circ S(w) \\ &= \mathcal{E} \circ L \circ P(w) + \mathcal{E} \circ L \circ R(w) \\ &= A \circ P(w) + B \circ R(w) \\ &= [A \circ P + B \circ R](w). \end{aligned}$$

4. Suppose

$$L = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right).$$

If we pick a basis β for $\mathbf{W}$ and γ for $\mathbf{Z}$ then we can write the atrix

$$[L]_{\beta||\gamma \leftarrow \beta||\gamma} = [c_i \ldots c_k].$$

where $\beta||\gamma$ is a concatenation of the bases and the columns:

$$\begin{aligned} c_i &= [L(b_i)]_{\beta||\gamma} \\ &= \left[\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \begin{pmatrix} b_i \\ \mathbf{0} \end{pmatrix}_+ \right]_{\beta||\gamma} \\ &= \left[\begin{pmatrix} A(b_i) \\ C(b_i) \end{pmatrix}_+ \right]_{\beta||\gamma} \\ &= \begin{bmatrix} [A(b_i)]_\beta \\ [C(b_i)]_\gamma \end{bmatrix} \end{aligned}$$

where b_i is the i^{th} element of the basis set β. If we repeat this procedure for all b_i's and g_i's, we will get:

$$[L]_{\beta||\gamma \leftarrow \beta||\gamma} = \left(\begin{array}{c|c} [A]_{\beta \leftarrow \beta} & [B]_{\gamma \leftarrow \beta} \\ \hline [C]_{\beta \leftarrow \gamma} & [D]_{\gamma \leftarrow \gamma} \end{array}\right)$$

where a block, say $[C]_{\beta \leftarrow \gamma}$ is generated by

$$[C]_{\gamma \leftarrow \beta} = \begin{bmatrix} [C(b_1)]_\gamma & [C(b_1)]_\gamma & \ldots [C(b_n)]_\gamma \end{bmatrix}.$$

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

Consider an illuminating example with an idempotent matrix: $\mathcal{E} = \mathcal{E}^2 \in \mathbb{M}_{n\times n}$. We know that $\mathrm{Im}(\mathcal{E}) \oplus \ker(\mathcal{E}) = \mathbb{C}^n$. So, $\mathcal{E} : \mathrm{Im}(\mathcal{E}) \oplus \ker(\mathcal{E}) \overset{\text{linear}}{\longrightarrow} \mathrm{Im}(\mathcal{E}) \oplus \ker(\mathcal{E})$ can be represented as

$$\mathcal{E} = \left(\begin{array}{c|c} \mathcal{E}_{11} & \mathcal{E}_{12} \\ \hline \mathcal{E}_{21} & \mathcal{E}_{22} \end{array}\right)$$

where

$$\begin{aligned}
\mathcal{E}_{11} &= \mathcal{E} \circ \mathcal{E}\Big|_{\mathrm{Im}(\mathcal{E})} = Id. \\
\mathcal{E}_{12} &= \mathcal{E} \circ \mathcal{E}\Big|_{\ker(\mathcal{E})} = \mathcal{O} \\
\mathcal{E}_{21} &= (Id. - \mathcal{E}) \circ \mathcal{E}\Big|_{\mathrm{Im}(\mathcal{E})} = \mathcal{O} \\
\mathcal{E}_{22} &= (Id. - \mathcal{E}) \circ \mathcal{E}\Big|_{\ker(\mathcal{E})} = \mathcal{O},
\end{aligned}$$

i.e.,

$$\mathcal{E} \sim \left(\begin{array}{c|c} Id.\Big|_{\mathrm{Im}(\mathcal{E})} & \mathcal{O} \\ \hline \mathcal{O} & \mathcal{O} \end{array}\right).$$

The idea we have just explored in the beginning is that we can pick a basis for $\mathrm{Im}(\mathcal{E})$ and $\ker(\mathcal{E})$ and concatenate to get a basis for the whole space so that we can write

$$[\mathcal{E}]_{\gamma_1||\gamma_2 \leftarrow \gamma_1||\gamma_2} = \left[\begin{array}{c|c} [Id.]_{\gamma_1\leftarrow\gamma_1} & [\mathcal{O}]_{\gamma_2\leftarrow\gamma_1} \\ \hline [\mathcal{O}]_{\gamma_1\leftarrow\gamma_2} & [\mathcal{O}]_{\gamma_2\leftarrow\gamma_2} \end{array}\right].$$

Now since

$$\begin{aligned}
[Id.]_{\gamma_1\leftarrow\gamma_1} &= I \\
[\mathcal{O}]_{\gamma_i\leftarrow\gamma_j} &= 0,
\end{aligned}$$

we can write

$$\mathcal{E} = \left(\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right).$$

We have a theorem:

Theorem 8.1. Every idempotent $\mathcal{E} \in \mathbb{M}_n$ is similar to matrix of the form

$$\begin{bmatrix} 1 & & \cdots & & 0 \\ & 1 & & & \\ \vdots & & \ddots & & \vdots \\ & & & 0 & \\ 0 & & & & 0 \end{bmatrix}$$

where the k 1's along the diagonal is $\operatorname{rank}(\mathcal{E})$, which also happens to be the trace of $\mathcal{E}$, i.e.,

$$\dim(\operatorname{Im}(\mathcal{E})) = \operatorname{rank}(\mathcal{E}) = \operatorname{Tr}(\mathcal{E}).$$

□

Consequently, we have:

Properties 8.2. 1. Since $\operatorname{rank}(\mathcal{E}) = \operatorname{Tr}(\mathcal{E})$, if $\operatorname{Tr}(M)$ is not an integer or negative, then M is not an idempotent.

2. Suppose we have a resolution of identity of m idempotents:

$$\mathcal{E}_1 + \mathcal{E}_2 + \cdots + \mathcal{E}_m = I_{\mathbb{C}^n}$$

then

$$\operatorname{Tr}(\mathcal{E}_1 + \mathcal{E}_2 + \cdots + \mathcal{E}_m) = \operatorname{Tr}(I_{\mathbb{C}^n}) = n,$$

i.e.,

$$\sum_{i=1}^{m} \operatorname{Tr}(\mathcal{E}_i) = n,$$

i.e.,

$$\sum_{i=1}^{m} \operatorname{rank}(\mathcal{E}_i) = n.$$

3. Consider the same resolution of identity. Let us consider

$$\operatorname{Im}(\mathcal{E}_1 + \mathcal{E}_2 + \cdots + \mathcal{E}_m)$$

and

$$\operatorname{Im}(\mathcal{E}_1) + \operatorname{Im}(\mathcal{E}_2) + \cdots + \operatorname{Im}(\mathcal{E}_m).$$

Next, consider

$$(\mathcal{E}_1 + \mathcal{E}_2 + \cdots + \mathcal{E}_m)(x) = \mathcal{E}_1(x) + \mathcal{E}_2(x) + \cdots + \mathcal{E}_m(x) \in \sum_{i=1}^{m} \operatorname{Im}(\mathcal{E}_i)$$

Therefore,

$$\mathbb{C}^n = \mathrm{Im}(I_{\mathbb{C}^n}) = \mathrm{Im}\left(\sum_{i=1}^{m} \mathcal{E}_m\right) \subseteq \sum_{i=1}^{m} \mathrm{Im}(\mathcal{E}_i) \subseteq \mathbb{C}^n.$$

So,

$$\sum_{i=1}^{m} \mathrm{Im}(\mathcal{E}_i) = \mathbb{C}^n.$$

But we also know that

$$\sum_{i=1}^{m} \dim(\mathrm{Im}(\mathcal{E}_i)) = \sum_{i=1}^{m} \mathrm{rank}(\mathcal{E}_i) = n.$$

So, we have a direct sum:

$$\boxed{\bigoplus_{i=1}^{m} \mathrm{Im}(\mathcal{E}_i) = \mathbb{C}^n}$$

In particular,

$$\dim\left(\bigoplus_{i=2}^{m} \mathrm{Im}(\mathcal{E}_i)\right) = n - \dim(\mathrm{Im}(\mathcal{E}_1)) = \dim(\ker(\mathcal{E}_1)).$$

But we also know that

$$\bigoplus_{i=2}^{m} \mathrm{Im}(\mathcal{E}_i) \supseteq \mathrm{Im}\left(\sum_{i=2}^{m} \mathcal{E}_i\right) = \mathrm{Im}(I - \mathcal{E}_1) = \ker(\mathcal{E}_1).$$

But since

$$\dim(\ker(\mathcal{E}_1)) = \dim\left(\bigoplus_{i=2}^{m} \mathrm{Im}(\mathcal{E}_i)\right),$$

it must be true that

$$\boxed{\ker(\mathcal{E}_1) = \bigoplus_{i=2}^{m} \mathrm{Im}(\mathcal{E}_i)}$$

This implies

$$\mathrm{Im}(\mathcal{E}_2) \subseteq \ker(\mathcal{E}_1),$$

which means

$$\mathcal{E}_1 \circ \mathcal{E}_2 = \mathcal{O}.$$

In general, for $i \neq j$

$$\boxed{\mathcal{E}_i \circ \mathcal{E}_j = \mathcal{O}_{n\times n}}$$

This leads us to the next item:

4. If $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_m$ are idempotents and $\sum_{i=1}^m \mathcal{E}_i = I$ then

$$\begin{cases} \mathcal{E}_i \circ \mathcal{E}_j = \mathcal{O} \\ \bigoplus_{i=1}^m \operatorname{Im}(\mathcal{E}_i) = \mathbb{C}^n. \end{cases}$$

8.5 Direct sums of operators

In the last subsection, we have worked with idempotents and resolution of identity. Now, suppose we have any linear function $\mathcal{L} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$, where $\mathbf{V}$ is finite dimensional. Suppose we have a non-trivial $\mathcal{L}$-invariant proper subspace $\mathbf{W}$ of $\mathbf{V}$. Then we can write

$$\mathbf{V} = \mathbf{W} \oplus \square.$$

Let $v_1, v_2, \ldots, v_m \in \mathbf{W}$ a basis of $\mathbf{W}$, and $v_{m+1}, v_{m+2}, \ldots, v_n$ be a basis of $\square$. We immediately know that

$$\operatorname{span}(v_1, v_2, \ldots, v_m) \oplus \operatorname{span}(v_{m+1}, v_{m+2}, \ldots, v_n) = \mathbf{V}.$$

Now, with respect to the decomposition $\mathbf{V} = \mathbf{W} \oplus \square$, we can represent $\mathcal{L}$ as

	$\mathbf{W}$	$\square$
$\mathbf{W}$	$\mathcal{L}_{11}$	$\mathcal{L}_{12}$
$\square$	$\mathcal{L}_{21}$	$\mathcal{L}_{22}$

where, like what we have done before

$$\mathcal{L}_{11} = \mathcal{E} \circ \mathcal{L}\Big|_{\mathbf{W}}$$
$$\mathcal{L}_{12} = \mathcal{E} \circ \mathcal{L}\Big|_{\square}$$
$$\mathcal{L}_{21} = (Id. - \mathcal{E}) \circ \mathcal{L}\Big|_{\mathbf{W}}$$
$$\mathcal{L}_{22} = (Id. - \mathcal{E}) \circ \mathcal{L}\Big|_{\square}.$$

But notice that since $\mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$

$$\mathcal{L}(w) \in \mathbf{W} = \operatorname{Im}(\mathcal{E}) = \ker(Id. - \mathcal{E}),$$

i.e.,

$$(Id. - \mathcal{E}) \circ \mathcal{L}\Big|_{\mathbf{W}} = \mathcal{O}.$$

As a little aside, the converse is also true:

$$\begin{aligned}
&(Id. - \mathcal{E}) \circ \mathcal{L}\Big|_{\mathbf{W}} = \mathcal{O} \\
\Longrightarrow &\operatorname{Im}\left(\mathcal{L}\Big|_{\mathbf{W}}\right) \subseteq \ker(Id. - \mathcal{E}) = \operatorname{Im}(\mathcal{E}) = \mathbf{W} \\
\Longrightarrow &\mathbf{W} \in \mathfrak{Lat}(\mathcal{L}).
\end{aligned}$$

So, we have established that

$$\mathcal{L}_{12} = \mathcal{O} \iff \begin{array}{c|c|c|} & \mathbf{W} & \square \\ \hline \mathbf{W} & & \\ \square & \mathcal{O} & \end{array} \iff \mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$$

$$\mathcal{L}_{21} = \mathcal{O} \iff \begin{array}{c|c|c|} & \mathbf{W} & \square \\ \hline \mathbf{W} & & \mathcal{O} \\ \square & & \end{array} \iff \square \in \mathfrak{Lat}(\mathcal{L}).$$

So, if $\mathbf{W}, \square \in \mathfrak{Lat}(\mathcal{L})$ and $\mathbf{W} \oplus \square = \mathbf{V}$ then $\mathcal{L}$ can be represented by

$$\begin{array}{c|c|c|} & \mathbf{W} & \square \\ \hline \mathbf{W} & \mathcal{A} & \mathcal{O} \\ \square & \mathcal{O} & \mathcal{B} \end{array}.$$

But of course the question is whether we know there is an invariant subspace of $\mathcal{L}$ to start. While we don't really have a clue as to how to *find* one for any linear function $\mathcal{L}$, we can surely *construct* one. For $v_0 \neq \mathbf{0} \in \mathbf{V}$. We want to construct $\mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$ such with $v_0 \in \mathbf{W}$, i.e.,

$$v_0 \in \mathbf{W}, \mathcal{L}(v_0) \in \mathbf{W}, \ldots, \mathcal{L}^n(v_0) \in \mathbf{W}.$$

Say $\mathcal{L}^k(v_0)$ is that first one that is a linear combination of previous ones, i.e.,

$$\mathcal{L}^k(v_0) = \sum_0^{k-1} a_i \mathcal{L}^i(v_0)$$

then, say for $k = 3$, we have

$$\left(\mathcal{L}^3 - a_0\mathcal{L}^0 - a_1\mathcal{L}^1 - a_2\mathcal{L}^2\right)(v_0) = \mathbf{0}.$$

By the fundamental theorem of algebra, we can factor out the "polynomial" above to get

$$(\mathcal{L} - r_1 I)(\mathcal{L} - r_2 I)(\mathcal{L} - r_3 I)(v_0) = \mathbf{0}.$$

But since $v_0 \neq \mathbf{0}$, at least one of $\mathcal{L} - r_i I$ is NOT injective, i.e., for some $\hat{v} \neq \mathbf{0}$,

$$\mathcal{L}(\hat{v}) = r_i \hat{v}.$$

We have just shown that r_i is an eigenvalue for $\mathcal{L}$ and $\hat{v}$ is an eigenvector for $\mathcal{L}$.

Theorem 8.2. Every linear function on $\mathbb{C}$ on finite dimensional vector space has an eigenvector. $\square$

It follows immediately that

Theorem 8.3. Every linear function on $\mathbb{C}$ on finite dimensional vector space has a one-dimensional invariant subspace. $\square$

9 Invariant subspaces

9.1 Triangularization

Consider $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{Z}$, with $\mathbf{W} \in \mathfrak{Lat}(\mathcal{L})$ and is one-dimensional. Then

$$\mathcal{L} = \begin{array}{c|c|c|} & \mathbf{W}_1 & \mathbf{Z} \\ \hline \mathbf{W}_1 & \mathcal{A} & \mathcal{B} \\ \mathbf{Z} & \mathcal{O} & \mathcal{D} \end{array} = \left(\begin{array}{c|c} \text{“}\lambda_1'' & \mathcal{B} \\ \hline \mathcal{O} & \mathcal{D} \end{array}\right).$$

Now, for $\mathcal{D} : \mathbf{Z} \xrightarrow{\text{linear}} \mathbf{Z}$ we can do the same thing we did for $\mathcal{L}$ if $\mathbf{Z}$ is not one-dimensional. Let's say that $\mathbf{Z} = \mathbf{W}_2 \oplus \mathbf{U}$, with $\mathbf{W}_2 \in \mathfrak{Lat}(\mathcal{L})$, is one-dimensional, then

$$\mathcal{D} = \left(\begin{array}{c|c} \text{“}\lambda_2'' & \mathcal{B}' \\ \hline \mathcal{O} & \mathcal{D}' \end{array}\right)$$

and so on. So,

$$\mathcal{L} = \left(\begin{array}{c|c|c} \text{”}\lambda_2\text{”} & \mathcal{B}_1 & \mathcal{B}_2 \\ \hline \mathcal{O} & \text{”}\lambda_2\text{”} & \square \\ \hline \mathcal{O} & \mathcal{O} & \hat{\mathcal{D}} \end{array}\right),$$

and so on until we get an upper triangular block-matrix. We have a theorem:

Theorem 9.1. Schur's Theorem: Every $n \times n$ (complex-valued) matrix is similar to an upper triangular matrix.

Proof. We prove by induction. The base case of $n = 1$ is trivial. Suppose that n_0 is the smallest size for which there is a counter example $\mathcal{A}$ ($n_0 \geq 2$). Now, $\mathcal{A}$ is a linear function, and hence has an eigenvector called w_0. We know that $\text{span}(w_0) \in \mathfrak{Lat}(\mathcal{A})$. It follows that we can write

$$\mathbb{C}^n = \text{span}(w_0) \oplus \mathbf{Z}.$$

With respect to this decomposition, $\mathcal{A}$ has the form

$$\mathcal{A} = \left(\begin{array}{c|c} \lambda & \mathcal{B} \\ \hline \mathcal{O} & \mathcal{D} \end{array}\right).$$

Now, pick a basis of $\mathbf{Z}$, called $\Gamma_2 : (w_1, \ldots, w_{n_0-1})$. Let the basis of $\text{span}(w_0)$ be Γ_1. We have $\Gamma = \Gamma_1 || \Gamma_2$. We know that

$$[\mathcal{A}]_\Gamma = \big[[\mathcal{A}(w_0)]_\Gamma \quad [\mathcal{A}(w_1)]_\Gamma \quad \ldots \quad [\mathcal{A}(w_{n_0-1})]_\Gamma \big] = \left(\begin{array}{c|c} \lambda & \square \\ \hline \mathcal{O} & [T]_{\Gamma_2 \leftarrow \Gamma_2} \end{array}\right).$$

We know that $\mathcal{A} \sim [\mathcal{A}]_\Gamma$. We also know (or assume) that $\mathcal{A}$ is the smallest counterexample. Then $\mathcal{D}$ cannot be a counter example since it is smaller than $\mathcal{A}$, i.e., $\mathcal{A} = S^{-1} \circ T \circ S$ for some upper triangular T. Hence

$$\mathcal{A} = \left(\begin{array}{c|c} \lambda & \square \\ \hline \mathcal{O} & S^{-1} \circ T \circ S \end{array}\right) = \left(\begin{array}{c|c} 1 & \mathcal{O} \\ \hline \mathcal{O} & S^{-1} \end{array}\right) \left(\begin{array}{c|c} \lambda & \triangle \\ \hline \mathcal{O} & T \end{array}\right) \left(\begin{array}{c|c} 1 & \mathcal{O} \\ \hline \mathcal{O} & S \end{array}\right)$$

But notice that the left and right block-matrices are inverses of each other, so

$$\mathcal{A} \sim \left(\begin{array}{c|c} \lambda & \triangle \\ \hline \mathcal{O} & T \end{array}\right),$$

which is an upper triangular matrix, since T is upper triangular. Hence, $\mathcal{A}$ is similar to an upper triangular matrix. But this is a contradiction, i.e., there is no such $\mathcal{A}$. $\square$

9.2 Reducing subspaces

10 Polynomials applied to operators

In this section we will explore a class of functions $\rho_{\mathcal{L}} : \mathbb{P} \overset{\text{linear}}{\longrightarrow} \mathfrak{L}(\mathbf{V})$. We first know from the problems that $\rho_{\mathcal{L}}$ is

Properties 10.1.

1. Linear
2. Multiplicative.

10.1 Minimal polynomials at a vector

We have shown that for each $\mathbf{0} \neq v_0 \in \mathbf{V}$ there is some $\mathbb{0} \neq P \in \mathbb{P}$ such that $(P(\mathcal{L}))(v_0) = \mathbf{0}$. Now, let us consider

$$\left\{ P \in \mathbb{P} \middle| (P(\mathcal{L}))(v_0) = \mathbf{0} \right\} \subseteq \mathbb{P}.$$

It is easy to show that this is a subspace. But also notice that it also has an absorption property, i.e., if $(P(\mathcal{L}))(v_0) = \mathbf{0}$ then $(q(\mathcal{L}) \cdot P(\mathcal{L}))(v_0) = (q \cdot P)(\mathcal{L})(v_0) = \mathbf{0}$ for any $q \in \mathbb{P}$. We call subspaces like these (with an additional absorption property) an **ideal**.

Properties 10.2.

1. There is the smallest unique monic polynomial.
2. There exists a unique and smallest $P(\mathcal{L})$ such that $P(\mathcal{L})(v_0) = \mathbf{0}$.
3. There exists a $P(\mathcal{L})$ such that $P(\mathcal{L})(v) = \mathbf{0}$ for all $v \in \mathbf{V}$.
4. There exists a unique and smallest $P(\mathcal{A})$ such that $P(\mathcal{A}) = \mathcal{O}$ for any $\mathcal{A} \in \mathbb{M}_{n\times n}$

10.2 Existence of eigenvalues

10.3 Global minimal polynomials

If we think about it a little, it is true that

$$[P(\mathcal{L})]_{\Gamma\leftarrow\Gamma} = P\left([\mathcal{L}]_{\Gamma\leftarrow\Gamma}\right).$$

Let us call

$$\mathcal{A} = [\mathcal{L}]_{\Gamma\leftarrow\Gamma}.$$

It is true that $\mathcal{A}$ is nothing but an element in the space of matrices, and so for some k

$$P(\mathcal{A}) = \mathcal{A}^k - a_0\mathcal{A}^0 - \cdots - a_{k-1}\mathcal{A}^{k-1} = \mathcal{O}.$$

So, we look at the ideal:

$$\left\{ P \in \mathbb{P} \middle| P(\mathcal{A}) = \mathcal{O} \right\}.$$

There exists a singe minimal monic element that sends $\mathcal{A}$ to $\mathcal{O}$ (from the problems). But if $P(\mathcal{A}) = \mathcal{O}$, then $(P(\mathcal{A}))(v_0) = \mathbf{0}$ for any $v \in \mathbf{V}$. Therefore,

$$P(\mathcal{A}) \in \left\{ P \in \mathbb{P} \middle| (P(\mathcal{L}))(v_0) = \mathbf{0} \right\}.$$

In particular, all elements in this ideal are divisors of P where $P(\mathcal{A}) = \mathcal{O}$. (We also showed why this is true in the homework - has to do with the **generator** of the ideal.) It turns out that $P(\mathcal{A})$ is also the smallest degree monic polynomial to annihilates $\mathcal{A}$. We call $P(\mathcal{A})$ the **global minimum polynomial of** $\mathcal{A}$, denoted $\mu_{\mathcal{A}}$.

What about **local minimum polynomials**? These are the minimal polynomials at the vectors $v \in \mathbf{V}$ constructed from looking at the first $\mathcal{A}^k(v_0)$ that can be expressed as a linear combination of the previous ones. We denote the local minimal polynomial of $\mathcal{A}$ at v_0 $\mu_{\mathcal{A},v_0}$. As we have shown before,

$$\boxed{\mu_{\mathcal{A},v_0} \text{ divides } \mu_{\mathcal{A}}}$$

since $\mu_{\mathcal{A}}$ is necessarily a multiple of $\mu_{\mathcal{A},v_0}$, as shown in the problems.

Another point to notice is that a polynomial has at most finitely man monic divisors. So, suppose that

$$\mu_{\mathcal{A}}(z) = (z - r_1)^{\gamma_1} \dots (z - r_k)^{\gamma_k}.$$

Consider a bunch of local minimal polynomials $\mu_{\mathcal{A},v_1}(z), \mu_{\mathcal{A},v_2}(z), \dots \mu_{\mathcal{A},v_1 35}(z)$. Then consider a subspace

$$\left\{ v \middle| (\mu_{\mathcal{A},v_7})(v) = \mathbf{0} \right\} = \ker\left(\mu_{\mathcal{A},v_7}\right).$$

The union of these 135 subspaces is the whole space, $\mathbf{V}$. And so, by problem set 2, one of these subspaces must equal to whole union, i.e., equals the whole space. So, $\mu_{\mathcal{A},v_i} = \mathbf{0}$ for all $v \in \mathbf{V}$ for at least one i, i.e., $\mu_{\mathcal{A},v_i}$ annihilates $\mathcal{A}$. But because $\mu_{\mathcal{A},v_i}$ divides $\mu_{\mathcal{A}} = q \cdot \mu_{\mathcal{A},v_i}$, we have

$$\mu_{\mathcal{A},v_i} = \mu_{\mathcal{A}}$$

for at least one i.

10.4 Existence of eigenvalues

In this section we will see why we are interested in these minimal polynomials. Suppose we have $\mu_{\mathcal{A}}(z) = (z - r_1)^{\gamma_1} \dots (z - r_k)^{\gamma_k}$. Then

$$\mu_{\mathcal{A}}(\mathcal{A}) = (\mathcal{A} - r_1 I)^{\gamma_1} \circ \dots \circ (\mathcal{A} - r_k I)^{\gamma_k} = \mathcal{O}.$$

Now, we can show (by induction) that none of these factors can be invertible (multiplying an inverse of one factor to both sides and show that we get a smaller annihilating polynomial). Therefore, we are guaranteed that $(\mathcal{A} - r_i I)(v) = \mathbf{0}$ for some non-zero vector v for any r_i roots of $\mu_{\mathcal{A}}$. So, we have that **all roots of $\mu_{\mathcal{A}}$ are eigenvalues of** $\mathcal{A}$.

But are there other eigenvalues that are NOT roots of $\mu_{\mathcal{A}}$? The answer, fortunately, is no. Consider a vector $(z - \lambda)$, where λ is an eigenvalue of $\mathcal{A}$ but is not a root of $\mu_{\mathcal{A}}$. It follows that $\mu_{\mathcal{A}}(z)$ and $(z-\lambda)$ are relatively prime (since they share no common roots). Hence, by the problems, there exist $q, p \in \mathbb{P}$ such that

$$p(z)\mu_{\mathcal{A}}(z) + q(z)(z - \lambda) = \mathbb{1},$$

i.e.,

$$p(\mathcal{A})\mu_{\mathcal{A}}(\mathcal{A}) + q(\mathcal{A})(\mathcal{A} - \lambda I) = I.$$

But since $\mu_{\mathcal{A}}(\mathcal{A}) = \mathcal{O}$, we have

$$q(\mathcal{A})(\mathcal{A} - \lambda I) = I.$$

This implies $(\mathcal{A} - \lambda I)$ is invertible, hence $\ker(\mathcal{A} - \lambda I) = \{\mathbf{0}\}$, thus λ is not an eigenvalue of $\mathcal{A}$, a contradiction. So no such λ exists. So, we have a theorem:

Theorem 10.1. All roots of $\mu_{\mathcal{A}}$ are **exactly** the eigenvalues of $\mathcal{A}$. □

We notice that this is a much easier way to find eigenvalues than through using characteristic polynomials since the degrees of minimal polynomials are often much smaller. This leads us to an interesting point. Notice that $\mu_{\mathcal{A}}(z)$ has degree at most n^2 for an $n \times n$ matrix $\mathcal{A}$ (recall how we constructed $\mu_{\mathcal{A}}$ by going through $I, \mathcal{A}, \mathcal{A}^2, \ldots$ and that $\dim(\mathbb{M})_{n\times n} = n^2$). However, the characteristic polynomial of $\mathcal{A}$ has degree at most n. By Cayley-Hamilton theorem, $\deg(\mu_{\mathcal{A}}) \leq \deg(\text{char}(\mathcal{A}))$, so

Theorem 10.2. For $\mathcal{A} \in \mathcal{M}_{n\times n}$, $\deg(\mu_{\mathcal{A}}) \leq n$. □

Theorem 10.3. Similar matrices have the same minimal polynomial.

Proof. Recall that

$$[P(\mathcal{L})]_{\Gamma\leftarrow\Gamma} = P\left([\mathcal{L}]_{\Gamma\leftarrow\Gamma}\right).$$

This simple says if $P(z)$ happens to be annihilating and is smallest, then $P(\mathcal{L}) = \mathcal{O}$. It is helpful to think of polynomials as belong to the linear function, not to matrices, as they are nothing but representations of the linear function, not that linear function itself. Now, of course the converse is not true (consider two identity matrices of different sizes). □

Theorem 10.4.

$$\boxed{\mu_{\mathcal{A}^\top} = \mu_{\mathcal{A}}}$$

Proof. If $\mathcal{A} \in \mathcal{M}_{n\times n}$ and $a_0 I + a_1\mathcal{A} + \cdots + a_k\mathcal{A}^k = \mathcal{O}$. Apply the transformation to this,

$$\sum_j^k a_j(\mathcal{A}^j)^\top = \sum_j^k a_j(\mathcal{A}^\top)^j = \mathcal{O}.$$

So, if $P(\mathcal{A}) = \mathcal{O}$, then $P(\mathcal{A}^\top) = \mathcal{O}$. But if $P(\mathcal{A}^\top) = \mathcal{O}$, then $P(\mathcal{A}) = \mathcal{O}$. Therefore, $\mu_{\mathcal{A}} = \mu_{\mathcal{A}^\top}$. $\square$

10.5 Minimal polynomials of block-diagonal operators

We observe that if

$$T = \begin{pmatrix} \mathcal{A} & \mathcal{O} \\ \mathcal{O} & \mathcal{M} \end{pmatrix}$$

then

$$T^0 = \begin{pmatrix} \mathcal{A}^0 & \mathcal{O} \\ \mathcal{O} & \mathcal{M}^0 \end{pmatrix}, T^2 = \begin{pmatrix} \mathcal{A}^2 & \mathcal{O} \\ \mathcal{O} & \mathcal{M}^2 \end{pmatrix}, \ldots, T^n = \begin{pmatrix} \mathcal{A}^n & \mathcal{O} \\ \mathcal{O} & \mathcal{M}^n \end{pmatrix}.$$

Thus we can convince ourselves that

$$P(T) = \begin{pmatrix} P(\mathcal{A}) & \mathbb{0} \\ \mathbb{0} & P(\mathcal{M}) \end{pmatrix}.$$

We can in fact think about of μ_T relates to $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{M}}$. A polynomial, say P, annihilates $T \iff P$ annihilates both $\mathcal{A}$ and $\mathcal{M} \iff \mu_{\mathcal{A}}$ and $\mu_{\mathcal{M}}$ both divide $P \iff P$ is a common multiple of $\mu_{\mathcal{A}}, \mu_{\mathcal{M}}$. Now, since μ_T is the minimal polynomial, whose multiples are such P's, we get

$$\boxed{\mu_T = \text{LCM}(\mu_{\mathcal{A}}, \mu_{\mathcal{M}})}$$

In particular, for a diagonal matrix, say

$$T = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_k \end{pmatrix},$$

the minimal polynomial

$$\mu_T = \text{LCM}(z-\alpha_1, z-\alpha_2, \ldots, z-\alpha_k) = (z-\alpha_1)\ldots(z-\alpha_j),$$

for distinct α_i, $i \neq j \leq k$.

Example 10.1. Let

$$T = \begin{pmatrix} 3 & & & \\ & 4 & & \\ & & 1 & \\ & & & 3 \end{pmatrix}.$$

Then it is clear that

$$\mu_T(z) = (z-3)(z-4)(z-1).$$

And thus the "spectrum" of T is

$$\sigma_{\mathbb{C}}(T) = \{1, 3, 4\}.$$

$\square$

10.6 (Minimal) polynomials of block-Δ^r operators

Consider

$$T = \begin{pmatrix} \mathcal{A} & \mathcal{K} \\ \mathcal{O} & \mathcal{M} \end{pmatrix}.$$

Then, like before

$$T^0 = \begin{pmatrix} I & \mathcal{O} \\ \mathcal{O} & I \end{pmatrix}, T^2 = \begin{pmatrix} \mathcal{A}^2 & \triangle \\ \mathcal{O} & \mathcal{M}^2 \end{pmatrix}, \dots, T^n = \begin{pmatrix} \mathcal{A}^n & \triangle' \\ \mathcal{O} & \mathcal{M}^n \end{pmatrix}.$$

So,

$$P(T) = \begin{pmatrix} P(\mathcal{A}) & \square \\ \mathcal{O} & P(\mathcal{M}) \end{pmatrix}.$$

If $P(T) = \mathcal{O}$, then $P(\mathcal{A}) = \mathcal{O}$ and $P(\mathcal{M}) = \mathcal{O}$. Thus, P is a common multiple (not necessarily least) of $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{M}}$. (Notice that we no longer have $\iff$ like with block-diagonal matrices.)

Thus we have

$$\mathrm{LCM}(\mu_{\mathcal{A}}, \mu_{\mathcal{M}}) \big| P(T).$$

But we also know that

$$\begin{aligned}(\mu_{\mathcal{A}} \cdot \mu_{\mathcal{M}})(T) &= \mu_{\mathcal{A}}(T) \cdot \mu_{\mathcal{M}}(T) \\ &= \begin{pmatrix} \mathcal{O} & \mathcal{O} \\ \mathcal{O} & \mu_{\mathcal{A}}(\mathcal{M}) \end{pmatrix} \begin{pmatrix} \mu_{\mathcal{M}}(\mathcal{A}) & \mathcal{O} \\ \mathcal{O} & \mathcal{O} \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{O} & \mathcal{O} \\ \mathcal{O} & \mathcal{O} \end{pmatrix} \\ &= [0].\end{aligned}$$

Therefore,

$$\boxed{\mathrm{LCM}(\mu_{\mathcal{A}}, \mu_{\mathcal{M}}) \big| \mu_T \big| \mu_{\mathcal{A}} \mu_{\mathcal{M}}}$$

and hence

$$\boxed{\sigma_{\mathbb{C}}(T) = \sigma_{\mathbb{C}}(\mathcal{A}) \cup \sigma_{\mathbb{C}}(\mathcal{M})}$$

i.e., linear factors of T are exactly those that appear in $\mu_{\mathcal{A}}$ or $\mu_{\mathcal{M}}$ (or both).

Of course "equality" occurs when $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{M}}$ are relatively prime.

Example 10.2. Consider

$$T = \left(\begin{array}{c|c|c|c} 3 & 1 & 2 & 8 \\ \cline{1-2} & 4 & 5 & 1 \\ \hline & & 2 & 7 \\ \cline{3-4} & & & 3 \end{array}\right).$$

Then

$$\begin{aligned} \sigma_{\mathbb{C}}(T) &= \sigma_{\mathbb{C}} \left(\begin{array}{c|c} 3 & 1 \\ \hline & 4 \end{array}\right) \cup \sigma_{\mathbb{C}} \left(\begin{array}{c|c} 2 & 7 \\ \hline & 3 \end{array}\right) \\ &= \sigma_{\mathbb{C}}[3] \cup \sigma_{\mathbb{C}}[4] \cup \sigma_{\mathbb{C}}[2] \cup \sigma_{\mathbb{C}}[3] \\ &= \{2, 3, 4\}. \end{aligned}$$

We could have guessed this result - these are just the values on the diagonal of T. What, then, is $\mu_T(z)$? Unfortunately, we don't know:

$$\mu_T(z) = (z-2)(z-4)(z-3)^{1 \text{ or } 2}.$$

$\square$

Example 10.3. If $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $p(z) = (z-a)(z-d) - bc$ annihilates T.

Proof.

$$\begin{aligned} p(T) &= (T - aI)(T_d I) - bcI \\ &= \begin{pmatrix} 0 & b \\ c & d-a \end{pmatrix} \begin{pmatrix} a-d & b \\ c & 0 \end{pmatrix} - \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix} \\ &= \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix} - \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix} \\ &= [0]. \end{aligned}$$

$\square$

We have a theorem:

Theorem 10.5. If $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is NOT a multiple of identity, then $\mu_T(z) = (z-a)(z-d) - bc$. $\square$

11 Eigentheory

11.1 Minimal polynomials and the spectrum

11.2 Spectral Mapping Theorem

12 Diagonalization

12.1 Spectral resolutions

12.2 Compressions to invariant subspaces

13 Simultaneous triangularization and diagonalization for commuting families

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

Before getting into primary decomposition, we first establish a few stepping stones to get to and understand the theorems to come. Suppose we have relatively prime polynomials p_1 and p_2. Then

$$\ker(p_1(\mathcal{L})) \cap \ker(p_2(\mathcal{L})) = \{\mathcal{O}\}.$$

Sketch. We know (from the problem sets) that given relatively prime p_1 and p_2, there exist q_1, q_2 such that

$$q_1(z)p_1(z) + q_2(z)p_2(z) = \mathbb{1}.$$

Thus, for $\mathcal{L}$ as argument,

$$q_1(\mathcal{L})p_1(\mathcal{L}) + q_2(\mathcal{L})p_2(\mathcal{L}) = I,$$

which we can write as

$$\left(q_1(\mathcal{L})p_1(\mathcal{L}) + q_2(\mathcal{L})p_2(\mathcal{L})\right)(v) = v,$$

i.e.,

$$q_1(\mathcal{L})[(p_1(\mathcal{L})(v))] + q_2(\mathcal{L})[(p_2(\mathcal{L})(v))] = v.$$

Now, if $v \in \ker(p_1(\mathcal{L})) \cap \ker(p_2(\mathcal{L}))$, then

$$q_1(\mathcal{L})(\mathbf{0}) + q_2(\mathcal{L})(\mathbf{0}) = v \implies v = \mathbf{0}.$$

This shows that

$$\ker(p_1(\mathcal{L})) \cap \ker(p_2(\mathcal{L})) = \{\mathbf{0}\}.$$

□

Now, let's say we're given

$$\mu_{\mathcal{A}}(z) = (z-\lambda_1)^{k_1} \dots (z-\lambda_n)^{k_n}.$$

Let us call

$$\begin{aligned} p_1 &= (z-\lambda_1)^{k_1} \dots (z-\lambda_{n-1})^{k_{n-1}} \\ p_2 &= (z-\lambda_n)^{k_n}. \end{aligned}$$

It is clear that p_1, p_2 are relatively prime. So, $\ker(p_1(\mathcal{L})) \cap \ker(p_2(\mathcal{L})) = \{\mathbf{0}\}$. Also,

$$p_1(\mathcal{A})p_2(\mathcal{A}) = \mu_{\mathcal{A}}(\mathcal{A}) = \mathcal{O} = p_2(\mathcal{A})p_1(\mathcal{A}),$$

so,

$$\begin{cases} \operatorname{Im}(p_1(\mathcal{A})) \subseteq \ker(p_2(\mathcal{A})) \\ \operatorname{Im}(p_2(\mathcal{A})) \subseteq \ker(p_1(\mathcal{A})) \end{cases}.$$

Yet also, for some q_1, q_2

$$q_1(z)p_1(z) + q_2(z)p_2(z) = 1.$$

Thus,

$$p_1(\mathcal{A})q_1(\mathcal{A}) + p_2(\mathcal{A})q_2(\mathcal{A}) = I.$$

Hence, for all $y \in \mathbf{V}$

$$p_1(\mathcal{A})[q_1(\mathcal{A})(y)] + p_2[q_2(\mathcal{A})(y)] = y.$$

But from the subset conditions, it is necessary that

$$\ker(p_2(\mathcal{A})) + \ker(p_1(\mathcal{A})) = \mathbf{V}.$$

But since these kernels intersect trivially,

$$\boxed{\mathbf{V} = \ker(p_1(\mathcal{A})) \oplus \ker(p_2(\mathcal{A}))}$$

Now, we observe that $p(\mathcal{A})$ and $\mathcal{A}$ commute, so from the problems, we get two invariant subspaces

$$\begin{cases} \ker(p_1(\mathcal{A})) \in \mathfrak{Lat}(\mathcal{A}) \\ \ker(p_2(\mathcal{A})) \in \mathfrak{Lat}(\mathcal{A}) \end{cases}.$$

With respect to this decomposition, $\mathcal{A}$ can be expressed as

$$\mathcal{A} = \begin{array}{c|c|c|} & \ker(p_1(\mathcal{A})) & \ker(p_2(\mathcal{A})) \\ \hline \ker(p_1(\mathcal{A})) & \mathcal{C} & \mathcal{O} \\ \ker(p_2(\mathcal{A})) & \mathcal{O} & \mathcal{D} \end{array}.$$

What are $\mu_{\mathcal{C}}$ and $\mu_{\mathcal{D}}$? It is clear that $\mu_{\mathcal{C}} = p_1$ and $\mathcal{D} = p_2$. Consider

$$q_1(\mathcal{A}) \begin{pmatrix} x \\ y \end{pmatrix}_+ = q_1(\mathcal{A})(x) + q_1(\mathcal{A})(y) = q_1(\mathcal{C})(x) + q_1(\mathcal{D})(y).$$

We know that $x \in \ker(p_1(\mathcal{A}))$, so this reduces to

$$q_1(\mathcal{A})(y) = q_1(\mathcal{C})(x) + q_1(\mathcal{D})(y),$$

which holds for all x. So, for $y = \mathbf{0}$, we must have $q_1(\mathcal{C}) = \mathcal{O}$, i.e, p_1 annihilates $\mathcal{C}$. So we can keep going and break the blocks $\mathcal{C}$ and $\mathcal{D}$ down into smaller block-diagonal pieces.

So, just a recap, if q_1, q_2 are relatively prime and $q_1 \cdot q_2$ annihilates $\mathcal{A} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$, then $\mathbf{V} = \ker(p_1(\mathcal{A})) \oplus \ker(p_2(\mathcal{A}))$ and with respect to this decomposition $\mathcal{A}$ has the form

$$A = \left(\begin{array}{c|c} \mathcal{C} & \mathcal{O} \\ \hline \mathcal{O} & \mathcal{D} \end{array}\right)$$

where q_1 annihilates $\mathcal{C}$ and q_2 annihilates $\mathcal{D}$. So, we have a little theorem:

Theorem 14.1. Under the hypotheses of the summary, if q_1, q_2 are monic and $\mu_{\mathcal{A}} = q_1 \cdot q_2$, then

$$\begin{cases} q_1 = \mu_{\mathcal{C}} \\ q_2 = \mu_{\mathcal{D}} \end{cases}.$$

Sketch. We know that

$$\mu_{\mathcal{A}} = \mu_{\begin{pmatrix} \mathcal{C} & \mathcal{O} \\ \mathcal{O} & \mathcal{D} \end{pmatrix}} = \mathrm{LCM}(\mu_{\mathcal{C}}, \mu_{\mathcal{D}}).$$

Now, $\mu_{\mathcal{C}}$ divides q_1 because q_1 annihilates $\mathcal{C}$. Similarly, $\mu_{\mathcal{D}}$ divides q_2 because q_2 annihilates $\mathcal{D}$. But since q_1, q_2 are relatively prime, $\mu_{\mathcal{C}}, \mu_{\mathcal{D}}$ are also relatively prime. Hence,

$$\mathrm{LCM}(\mu_{\mathcal{C}}, \mu_{\mathcal{D}}) = \mu_{\mathcal{C}} \cdot \mu_{\mathcal{D}}.$$

Therefore,

$$\mu_{\mathcal{A}} = \mu_{\mathcal{C}} \cdot \mu_{\mathcal{D}}.$$

But we also know that $\mu_{\mathcal{A}} = q_1 \cdot q_2 = \mu_{\mathcal{C}} \cdot \mu_{\mathcal{D}}$ and $\mu_{\mathcal{C}} \big| p_1$ and $\mu_{\mathcal{D}} \big| p_2$, we must have

$$\begin{cases} q_1 = \mu_{\mathcal{C}} \\ q_2 = \mu_{\mathcal{D}} \end{cases}.$$

□

Theorem 14.2. Primary Decomposition Theorem:
Suppose $\mathcal{A} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ and $\mu_{\mathcal{A}}(z) = (z - \lambda_1)^{p_1} \ldots (z - \lambda_k)^{p_k}$, then

$$\mathbf{V} = \ker(\mathcal{A} - \lambda_1 I)^{p_1} \oplus \cdots \oplus \ker(\mathcal{A} - \lambda_k I)^{p_k}$$

and with respect to this decomposition, $\mathcal{A}$ has the form

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_1 & & \\ & \ddots & \\ & & \mathcal{A}_k \end{pmatrix}$$

with $\mu_{\mathcal{A}_i}(z) = (z - \lambda_i)^{p_i}$.

Proof. We induct on k. The base case where $k = 1$ is trivially true. Consider the inductive hypothesis. Suppose that $k = 1, 2, \ldots, m$ hold. We want to show that the case where $k = m + 1$ also holds, i.e.,

$$\mu_{\mathcal{A}} = (z - \lambda_1)^{q_1} \ldots (z - \lambda_m)^{q_m} (z - \lambda_{m+1})^{q_{m+1}}.$$

Let $p_1 = (z - \lambda_1)^{q_1} \ldots (z - \lambda_m)^{q_m}$ and $p_2 = (z - \lambda_{m+1})^{q_{m+1}}$, i.e., by the lemma we just proved, $\mathcal{A} = \begin{pmatrix} \mathcal{B} & \mathcal{O} \\ \mathcal{O} & \mathcal{C} \end{pmatrix}$ with respect to the decomposition $\mathbf{V} = \ker(p_1(\mathcal{A})) \oplus \ker(p_2(\mathcal{A}))$ and $\mu_{\mathcal{B}} = p_1$ and $\mathcal{C} = p_2$.

By the inductive hypothesis, $\ker(p_1(\mathcal{A})) = \ker(\mathcal{B} - \lambda_1 I)q_1 \oplus \cdots \oplus \ker(\mathcal{B} - \lambda_m)^{q_m}$, and with respect to this decomposition, $\mathcal{B} = \begin{pmatrix} \mathcal{B}_1 & & \\ & \ddots & \\ & & \mathcal{B}_{\Updownarrow} \end{pmatrix}$ and $\mu_{\mathcal{B}_i}(z) = (z - \lambda_i)^{q_i}$. Hence, by the problem sets, $\mathcal{A}$ can be represented as

$$\mathcal{A} = \left(\begin{array}{ccc|c} \mathcal{B}_1 & & & \\ & \ddots & & \\ & & \mathcal{B}_m & \\ \hline & & & \mathcal{C} \end{array}\right)$$

with respect to the decomposition $\mathbf{V} = \ker(\mathcal{B} - \lambda_1 I)q_1 \oplus \cdots \oplus \ker(\mathcal{B} - \lambda_m)^{q_m} \oplus \ker(\mathcal{A} - \lambda_{m+1} I)^{q_{m+1}}$.

So, the next step is to show $\ker(\mathcal{B} - \lambda_i I)^{q_i} = \ker(\mathcal{A} - \lambda_i I)^{q_i}$ for all i. First, we know that

$$\ker(\mathcal{A} - \lambda_i I)^{q_i} \subseteq \ker\left((\mathcal{A} - \lambda_1)^{q_1} \ldots (\mathcal{A} - \lambda_m I)^{q_m}\right)$$

, since the factors can commute among themselves. Hence, we just rewrite this as

$$\ker(\mathcal{A} - \lambda_i I)^{q_i} \subseteq \ker(q_1(\mathcal{A})).$$

Similarly, we can show

$$\ker(\mathcal{B} - \lambda_i I)^{q_i} \subseteq \ker(q_1(\mathcal{A})).$$

Now, recall that

$$\mathcal{A} = \begin{pmatrix} \mathcal{B} & \mathcal{O} \\ \mathcal{O} & \mathcal{C} \end{pmatrix},$$

thus,

$$(\mathcal{A} - \lambda_i I)^{q_i} = \begin{pmatrix} (\mathcal{B} - \lambda_i)^{q_i} & \mathcal{O} \\ \mathcal{O} & \square \end{pmatrix}.$$

Consider $x \in \ker\left((\mathcal{A} - \lambda_i I)^{q_i}\right)$, i.e.,

$$\begin{aligned}
&(\mathcal{A} - \lambda_i I)^{q_i}(x) = \mathbf{0} \\
\iff &\begin{pmatrix} \mathcal{B} & \mathcal{O} \\ \mathcal{O} & \mathcal{C} \end{pmatrix} \begin{pmatrix} x \\ \mathbf{0} \end{pmatrix}_+ \\
\iff &(\mathcal{B} - \lambda_i I)^{q_i}(x) = \mathbf{0} \\
\iff &x \in \ker\left((\mathcal{B} - \lambda_i I)^{q_i}\right).
\end{aligned}$$

So we have just shown that

$$\ker\left((\mathcal{A} - \lambda_i I)^{q_i}\right) = \ker\left((\mathcal{B} - \lambda_i I)^{q_i}\right)$$

for all i. This completes our proof. $\square$

15 Cyclic decomposition and Jordan form

We can immediately follow the primary decomposition theorem with another theorem:

Theorem 15.1. Ever $\mathcal{L} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$, where $\mathbf{V}$ is finite dimensional, can be expressed as a direct sum of operators:

$$\mathcal{L} = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \cdots \oplus \mathcal{L}_m,$$

where

$$\mathcal{L}_i = \lambda_i I + \mathcal{N}_i,$$

where $\mathcal{N}_i$ is nilpotent.

It is worthwhile to study some properties of nilpotents.

15.1 Nilpotents

First, we want to look at the structure of nilpotents. Given a nilpotent $\mathcal{N} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$, such that, say $\mathcal{N}^8 = \mathcal{O}$ (and hence $\mu_{\mathcal{N}}(z) = z^8$) and $v_0 \in \mathbf{V}$, we write

$$\langle v_0 \rangle = \operatorname{span}\left\{v_0, \mathcal{N}(v_0), \ldots, \mathcal{N}^7(v_0)\right\} = \left\{P(\mathcal{N})(v_0) \middle| \deg(P) \leq 7\right\} \in \mathfrak{Lat}(\mathcal{N}),$$

a cyclic invariant subspace.

Theorem 15.2. For any nilpotent $\mathcal{N}$ as above, there exist some $v_1, v_2, \ldots v_k \in \mathbf{V}$ such that

$$\mathbf{V} = \langle v_0 \rangle \oplus \langle v_1 \rangle \oplus \cdots \oplus \langle v_k \rangle$$

□

We will look at the proof later, but with respect to this decomposition, $\mathcal{N}$ is block-diagonal:

$$\mathcal{N} = \begin{pmatrix} \mathcal{N}_1 & \mathcal{O} & \mathcal{O} \\ \mathcal{O} & \ddots & \mathcal{O} \\ \mathcal{O} & \mathcal{O} & \mathcal{N}_k \end{pmatrix}.$$

Let us suppose that the v_0 up to $\mathcal{N}^5(v_0)$ are linear independent, then

$$\mathcal{N}^6(v_0) = \sum_{i=0}^{5} a_i \mathcal{N}^i(v_0),$$

i.e.,

$$(\mathcal{N}^6 - a_5\mathcal{N}^5 - \ldots a_0 I)(v_0) = q(\mathcal{N})(v_0) = \mathbf{0}.$$

But recall that the global minimal polynomial for $\mathcal{N}$ is $\mu_{\mathcal{N}}(z) = z^8$, and that the local minimal polynomial divides th global. This means that $q(\mathcal{N}) = \mathcal{N}^k$ for some $k \leq 8$. So, the local minimal polynomial has to have degree 6. This implies $\mathcal{N}^6(v_0) = \mathbf{0}$. Therefore, $v_0, \mathcal{N}(v_0), \ldots, \mathcal{N}^5(v_0)$ is a basis of $\langle v_0 \rangle$. So, the first block along the diagonal of $\mathcal{N}$ can be written as

$$\mathcal{N}_1 = \begin{array}{c|c|c|c|c|}
 & \operatorname{span}(v_0) & \operatorname{span}(\mathcal{N}(v_0)) & \ldots & \operatorname{span}(\mathcal{N}^5(v_0)) \\
\hline
\operatorname{span}(v_0) & 0 & 0 & \ldots & 0 \\
\operatorname{span}(\mathcal{N}(v_0)) & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & 1 & & \vdots \\
\vdots & \vdots & \vdots & & \vdots \\
\operatorname{span}(\mathcal{N}^5(v_0)) & 0 & 0 & \ldots & 0
\end{array}$$

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

Of course, conventionally, we would prefer upper triangular matrices, so we simply reverse the order of the basis elements to get

$$\mathcal{N}_1 = \begin{pmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{pmatrix}$$

Theorem 15.3. Cyclic Decomposition for Nilpotents:
If $\mathcal{N} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$, $\mathbf{V}$ is finite dimensional, is nilpotent then $\mathbf{V} = \bigoplus_1^k \langle v_i \rangle_{\mathcal{N}}$ for some non-zero v_i.

In fact, we can choose any v_1 to start with. And with respect to this decomposition, $\mathcal{N}$ has the form $\mathcal{N}_1 \oplus \cdots \oplus \mathcal{N}_k$ $\square$

Theorem 15.4. If $\mathcal{N} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ nilpotent, $\mathbf{V}$ is finite dimensional, then there is a basis ψ (which can be constructed by concatenation) for $\mathbf{V}$ such that

$$[\mathcal{N}]_{\psi \leftarrow \psi} = \begin{pmatrix} [\mathcal{N}_1] & & \\ & \ddots & \\ & & [\mathcal{N}_k] \end{pmatrix},$$

where each block $[\mathcal{N}_i]$ has the form

$$[\mathcal{N}_i] = \begin{pmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{pmatrix}$$

□

15.2 Jordan canonical form theorem

Theorem 15.5.

$$\boxed{\mathcal{A}, \mathcal{B} \text{ similar} \iff \mathcal{A}, \mathcal{B} \text{ have the same JCF}}$$

□

We can "improve" the previous idea by considering $\mathcal{A} - \lambda I = \mathcal{N}$, i.e., $\mathcal{A} = \lambda I + \mathcal{N}$, and so

$$[\lambda I + \mathcal{N}]_{\psi \leftarrow \psi} = [\lambda I]_{\psi \leftarrow \psi} + [\mathcal{N}]_{\psi \leftarrow \psi} = \begin{pmatrix} \mathcal{A}_1 & & \\ & \ddots & \\ & & \mathcal{A}_k \end{pmatrix}$$

where each $\mathcal{A}_i$ is called a **Jordan block**, denoted $J_{\lambda, size}$. For example,

$$J_{\lambda,2} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

Theorem 15.6. Jordan canonical form theorem:
Every matrix is similar to a *unique (up to changing order of appearance)* direct sum of Jordan blocks. □

Proof. There are two parts to this proof. First, we want to show that this holds for nilpotents $\mathcal{N}$. Then, we extend it and show it also holds for $\mathcal{A} = \lambda I + \mathcal{N}$. □

Step 1: Let us revisit the idea of Weyr characteristic. Consider

$$\mathcal{N} = \begin{pmatrix} J_{0,9} & & \\ & \ddots & \\ & & J_{0,1} \end{pmatrix}.$$

Let's say that from$\mathcal{N}$ we find that

$$\{\mathbf{0}\} \overset{+8}{\subset} \ker(\mathcal{N}) \overset{+7}{\subset} \ker(\mathcal{N}^2) \overset{+7}{\subset} \ker(\mathcal{N}^3) \subset \cdots \subset \ker(\mathcal{N}^9) = \mathbf{V}$$

where the numbers indicate the increase in the dimension. Then, if we were given only the Weyr characteristics, i.e., the inclusion chain above, is it possible for us to generate $\mathcal{N}$? It turns out that the answer is YES. Basically, each number, going from left-to-right, gives us some extra information about $\mathcal{N}$. For instance, the first dimension increase gives us the number of Jordan blocks in $\mathcal{N}$. The difference between the second and the first number gives us the number of $J_{0,1}$ blocks. The different between the third and the second number gives us the number of $J_{0,2}$ blocks, and so on. (The most straightforward way to see how this works is to generate an example and go through the algorithm.)

So, we can see that no matter how the decomposition might differ, this Jordan construction is determined by just $\mathcal{N}$ (since we can reconstruct the Jordan form of $\mathcal{N}$ by just looking at the Weyr characteristics). So, if we have two similar idempotents, then they have the same Weyr characteristics, i.e., they must also have the same Jordan canonical form. But since we also know that if two nilpotents have the same Jordan-block decomposition then they are similar. Therefore, the **Weyr characteristics of $\mathcal{N}$ completely determines the Jordan block sizes and their multiplicities, i.e., completely determines the Jordan form of $\mathcal{N}$ (of course, up to order of appearance)**.

Similar matrices have the same Weyr characteristics, similar nilpotents have the same Jordan form. Conversely, if two nilpotents have the same Jordan form, hen they are similar to it, and so to each other. Therefore, we say that **Jordan form is a "complete invariant" for similarity of nilpotents.**

Step 2: But what about $\mathcal{A} + \lambda I + \mathcal{N}$? We observe that if $A = \lambda I + \mathcal{N}$ then the Weyr characteristics of $\mathcal{A} - \lambda I$ determines the Jordan form of $\mathcal{A}$. But what if we don't know λ? What if all we know is that $\mathcal{A} = \square \cdot I + \mathcal{N}$? It turns out that we can simply use $\mathrm{Tr}(\mathcal{A})$ to figure out what λ is because $\mathcal{A}$ in matrix form has only λ's along the diagonal:

$$\lambda = \frac{\mathrm{Tr}(\mathcal{A})}{\text{Size of } \mathcal{A}}.$$

(recall that $\mathrm{Tr}(\mathcal{A})$ does not change in different representations, i.e., it is "similarity-invariant").

Now, similar matrices have the same Jordan forms.

15.3 Diagonalizability

Recall that Every $n \times n$ matrix is triangularizable; i.e., every square matrix is similar to an upper or lower triangular matrix. Equivalently, any linear operator $\mathcal{L} \in \mathfrak{L}(\mathbf{V})$ can be represented by an upper triangular matrix for some basis.

So which matrices are diagonalizable? To answer this, we can reverse engineer: if $\mathcal{A} \sim \mathcal{D}$, where $\mathcal{D}$ is diagonal, then their minimal polynomials are the same: $\mu_\mathcal{A} = \mu_\mathcal{D}$, and

$$\mu_\mathcal{D}(z) = (z - d_1)\dots(z - d_k)$$

where all d_i's are distinct, i.e., $\mu_\mathcal{D}$ has no repeated roots, i.e., all roots of $\mu_\mathcal{D}$ have multiplicity 1. Next, we observe that for a given Jordan block $\mathcal{J}_{\lambda,\beta}$, its minimal polynomial is $\mu_\mathcal{J}(z) = (z-\lambda)^\beta$. Now, if $\mathcal{A} \sim \mathcal{J}_1 \oplus \mathbf{J}_2 \oplus \cdots \oplus \mathbf{J}_k$, then $\mu_\mathcal{A} = \mathrm{lcm}(\mu_{\mathcal{J}_1}, \dots, \mu_{\mathcal{J}_k})$. So in particular, $\mu_{\mathcal{J}_i}$ is a factor of $\mu_\mathcal{A}$ for each i. So if $\mu_\mathcal{A}$ has no repeated roots, then no $\mu_{\mathcal{J}_i}$ have no repeated roots, i.e., $\mu_{\mathcal{J}_i}(z) = z - \lambda_i$, i.e., $\mathcal{J}$ has size 1. This says $\mathcal{A}$ is similar to a direct sum of 1×1 Jordan blocks, i.e., similar to a diagonal matrix. We the following:

Theorem 15.7. The following statements are equivalent:

1. $\mathcal{A}$ is a diagonalizable.
2. $\mu_\mathcal{A}$ has no repeated roots.
3. There is a basis of $\mathbb{C}^n$ made up entirely of eigenvectors of $\mathcal{A}$, called an $\mathcal{A}$-eigenbasis of $\mathbb{C}^n$.

□

Of course, we can always check for repeated roots by looking at $\mu_\mathcal{A}$ and μ'_A. If $\mu_\mathcal{A}$ has repeated roots, then there exists a root r such that $\mu_A(r) = \mu'_\mathcal{A}(r) = 0$. To check if $\mu_\mathcal{A}$ and $\mu'_\mathcal{A}$ have no common roots, we use the division algorithm and find the greatest common divisor of the polynomials. If the gcd turns out to be the constantly one polynomial, then the polynomial have no common roots.

Theorem 15.8. Spectral Mapping Theorem: Let a matrix $\mathcal{A}$ be given, then

$$\sigma_\mathbb{C}(p(\mathcal{A})) = p\left[\sigma_\mathbb{C}(\mathcal{A})\right].$$

□

For example,

$$\sigma_\mathbb{C}(\alpha(\mathcal{A})) = \alpha\left[\sigma_\mathbb{C}(\mathcal{A})\right]$$
$$\sigma_\mathbb{C}(\mathcal{A} + \beta\mathcal{I}) = \sigma_\mathbb{C}(\mathcal{A}) + \beta$$

Now, suppose that $\mathcal{A}$ is invertible. Then for an eigenvector v,

$$\mathcal{A}^{-1}v = \lambda v \iff \frac{1}{v}v = \mathcal{A}v.$$

Theorem 15.9.

$$\gamma \in \sigma_\mathbb{C}(\mathcal{A}^{-1}) \iff \frac{1}{\gamma} \in \sigma_\mathbb{C}(\mathcal{A}).$$

□

15.4 Simultaneous triangularizability and diagonalizability

We first look at **simultaneous diagonalizability**. Suppose we have $\mathcal{A}, \mathcal{B} \in \mathfrak{L}(\mathbf{V})$, and an invertible $\mathcal{S}$ such that $\mathcal{S}^{-1}\mathcal{A}\mathcal{S}$ and $\mathcal{S}^{-1}\mathcal{B}\mathcal{S}$ are diagonal. Then

$$\mathcal{S}^{-1}\mathcal{A}\mathcal{S}\mathcal{S}^{-1}\mathcal{B}\mathcal{S} = \mathcal{S}^{-1}\mathcal{A}\mathcal{B}\mathcal{S} = \mathcal{S}^{-1}\mathcal{B}\mathcal{A}\mathcal{S} = \mathcal{S}^{-1}\mathcal{B}\mathcal{S}\mathcal{S}^{-1}\mathcal{A}\mathcal{S} \iff \mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}.$$

So we have the necessary conditions for simultaneous diagonalizability:

1. Individual diagonalizability
2. Commutativity.

It turns out that they are also sufficient.

Theorem 15.10. (†) The following are equivalent for a collection of $\mathcal{F} \in \mathbb{M}_n$:

1. $\mathcal{F}$ is a commutative collection and all element of $\mathcal{F}$ are individually diagonalizable.
2. There is an invertible $\mathcal{S}$ such that $\mathcal{S}^{-1}\mathcal{A}\mathcal{S}$ is diagonal for every $\mathcal{A} \in \mathcal{F}$.

□

To prove this, we first prove the following smaller theorem:

Theorem 15.11. $\mathcal{A}$ is diagonalizable $\iff$ $\mathcal{A} = \alpha_1\mathcal{E}_1 + \dots \alpha_k\mathcal{E}_k$ for some non-zero idempotents such that $\mathcal{E}_1 + \dots \mathcal{E}_k = \mathcal{I}$.

Proof. ($\implies$) $\mathcal{A}$ diagonalizable if and only if $\mathcal{S}^{-1}\mathcal{A}\mathcal{S} = \beta_1\mathcal{F}_1 + dots\beta_m\mathcal{F}_m$ diagonal, where $\mathcal{F}_i$'s are non-zero idempotents that resolve identity. Then.

$$\mathcal{A} = \mathcal{S}\left[\beta_1\mathcal{F}_1 + dots\beta_m\mathcal{F}_m\right]\mathcal{S}^{-1} = \sum_{i=1}^{m}\beta_i(\mathcal{S}\mathcal{F}_i\mathcal{S}^{-1}).$$

Now, since each $\mathcal{S}\mathcal{F}_i\mathcal{S}^{-1}$ is also an idempotent, we can write

$$\mathcal{A} = \sum_{i=1}^{m}\beta_i\mathcal{E}_i,$$

where $\sum_{i=1}^{m}\mathcal{E}_m = \mathcal{S}^{-1}\left(\sum_{i=1}^{m}\mathcal{F}_i\right)\mathcal{S} = \mathcal{I}$. □

($\Longleftarrow$) Now, suppose that $\mathcal{A} = \sum_{i=1}^{k}\alpha_i\mathcal{E}_i$ where $\mathcal{E}_i$'s idempotent that together resolve identity. Then we know that $\mathbf{V} = \bigoplus \mathrm{Im}(\mathcal{E}_i) = \mathbb{C}^p$. Now, let Γ_i be a basis for $\mathrm{Im}(\mathcal{E}_i)$, then $\Gamma = \Gamma_1||\Gamma_2||\ dots||\Gamma_k$ is a basis for $\mathbb{C}^p$. With respect to this decomposition, we can express $\mathcal{A}$ as a block-diagonal matrix with block-entries:

$$\mathcal{A}_{ij} = \mathcal{E}_i \circ \mathcal{A} \circ \mathcal{E}_j\Big|_{\mathrm{Im}(\mathcal{E}_j)} = \delta^i_j\alpha_i\mathcal{E}_i\Big|_{\mathrm{Im}(\mathcal{E}_i)} = \alpha\mathcal{I}\Big|_{\mathrm{Im}(\mathcal{E}_i)}.$$

So it is obvious that

$$[\mathcal{A}]_{\Gamma\leftarrow\Gamma} = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_k \end{bmatrix}$$

So, $\mathcal{A}$ is diagonalizable. □

□

Before we prove Theorem (†), we must show that α_i's are exactly the eigenvalues of $\mathcal{A}$, that $\mathrm{Im}(\mathcal{E}_i) = \mathbf{E}_{\mathcal{A}(\alpha_i)}$ for any i, and the representation of $\mathcal{A}$ as $\mathcal{A} = \sum_{i=1}^{k} \alpha_k \mathcal{E}_k$ is unique, where the α_i's are distinct. The proof will not be reproduced here, since it is one of the problems in the back. But the key to this is Lagrange's Interpolation formula:

$$p_i(z) = \prod_{j=1, j\neq i}^{k} \frac{z-\alpha_j}{\alpha_i - \alpha_j}$$

where (you can check) that $p_i(\mathcal{A}) = \mathcal{E}_i$. Another key to the the proof is the fact that any idempotent can be written as a diagonal with zero everywhere and a block-identity along the diagonal. Once this is done, we will have "mined" another theorem

Theorem 15.12. Each **spectral idempotent** of a diagonalizable $\mathcal{A}$ is a polynomial in $\mathcal{A}$, i.e., is of the form $p(\mathcal{A})$ for some polynomial p (given by the Lagrange Interpolation formula). □

Now we are ready for the proof of Theorem (†). We will dod this by induction on n.

Proof. ($\implies$) The base case where $n = 1$ is trivially true. Now, suppose that the theorem holds for all $1 \leq n < n_0$. We want to show that it also holds for n_0. If $\mathcal{F}$ contains only scalar multiples of identity, then we are done So, let us assume that $\mathcal{F}$ contains $\mathcal{A}$'s that are not multiples of identity. Then, by the theorem we just proved

$$\mathcal{A} = \sum_{i=1}^{k} \alpha_i \mathcal{E}_i,$$

for some idempotents $\mathcal{E}_i$ that together resolve identity, and distinct α_i's which are exactly the eigenvalues of $\mathcal{A}$. Now, since the α_i's are distinct, $n_0 \geq 2$. Next, recall that $\mathcal{E}_i = p_i(\mathcal{A})$, for some polynomial p_i and that every $\mathcal{B} \in \mathcal{F}$ commutes with $\mathcal{A}$. This means $\mathcal{B}$ commutes with $p_i(\mathcal{A})$, and in particular, commutes with every $\mathcal{E}_i$. So, the block-matrix representation, the block-entries of $\mathcal{B}$ can be written as

$$\mathcal{B}_{ij} = \mathcal{E}_i B \mathcal{E}_j \Big|_{\mathrm{Im}(\mathcal{E}_j)} = \mathcal{E}_i \mathcal{E}_j B = \delta^i_j \mathcal{B}\Big|_{\text{restricted}}.$$

So, $\mathcal{B}$ is block-diagonal:

$$\mathcal{B} = \begin{bmatrix} \mathcal{B}_{11} & & \\ & \ddots & \\ & & \mathcal{B}_{kk} \end{bmatrix}.$$

Now, since $\mathcal{A}$ and $\mathcal{B}$ commute, $\mathcal{A}_{ij}$ commutes with $\mathcal{B}_{ij}$. But $\mathcal{B}_{ij}$ are individually diagonalizable by hypothesis, so we can go on and find a basis for which each $[\mathcal{B}_{ij}]$ is diagonal, concatenate, and form the basis for the entire underlying space, making $\mathcal{B}$ diagonalizable.

The proof in the other direction is easy. This will be reproduced in the problems. □ □

15.5 Square roots of operators

15.6 Similarity of a matrix and its transpose

One of the problems in the back covers this. The key here is Jordan forms.

15.7 Similarity of a matrix and its conjugate

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

15.9 Power-convergent operators

15.10 Power-bounded operators

15.11 Row-stochastic matrices

16 Determinant & Trace

16.1 Classical adjoints

16.2 Cayley-Hamilton theorem

16.3 General Laplace expansions

16.4 Cauchy-Binet formula

17 Inner products and norms

Definition 17.1. Inner product An inner product $\phi : \mathbf{V} \times \mathbf{V} \to \mathbb{C}$ on a vector space $\mathbf{V}$ is a function satisfying the following conditions:

1. ϕ is partially linear in the first slot, partially conjugate linear in the second, i.e.,

$$\begin{aligned}\phi(\alpha v_1 + \beta v_2, w) &= \alpha\phi(v_1, w) + \beta\phi(v_2, w) \\ \phi(w, \alpha v_1 + \beta v_2) &= \bar{\alpha}\phi(w, v_1) + \bar{\beta}\phi(w, v_2).\end{aligned}$$

2. Conjugate symmetry: $\phi(\bar{v}, w) = \phi(w, v)$.
3. Definiteness: $\phi(v, v) \geq 0$, equality holds if and only if $v = \mathbf{0}$

□

Definition 17.2. Standard inner product on $\mathbb{C}^n$

$$\langle \begin{pmatrix} a \\ b \\ c \end{pmatrix}, (\alpha, \beta, \gamma) \rangle = a\bar{\alpha} + b\bar{b}eta + c\bar{\gamma},$$

so that

$$\langle \begin{pmatrix} a \\ b \\ c \end{pmatrix}, (a, b, c) \rangle = |a|^2 + |b|^2 + |c|^2 = \left\| \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right\|^2.$$

□

Definition 17.3. Inner product space An inner product space is a vector space with an inner product. □

We also observe that if (v, ϕ) is an inner product space, then $\|v\| := \sqrt{\phi(v, v)}$ defines a "norm" such that

1. $\| \ \| : \mathbf{V} \to [0, \infty)$
2. $\|v\| = 0 \iff v = \mathbf{0}$
3. $\|\alpha v\| = |\alpha| \cdot \|v\|$
4. $\|u + v\| \leq \|u\| + \|v\|$ (triangle inequality)
5. $\|\langle u, v\rangle\| \leq \|u\| \cdot \|v\|$ (Cauchy-Schwarz inequality).

17.1 Orthogonality

Let's have a preliminary "definition:" u, v orthogonal $\iff \langle u, v\rangle = 0$.

But there's a little more to the story. For any $S \subseteq \mathbf{V}$, we let

$$S^{\perp} = \left\{ v \in \mathbf{V} \middle| \langle v, s\rangle = 0 \text{ for any } s \in S \right\}.$$

We observe three things:

1. $\mathbf{0_V} \in \S^{\perp}$.
2. $\mathbf{0_V} \in \mathbf{V}^{\perp} = \left\{ v \in \mathbf{V} \middle| \langle v, w\rangle = 0 \text{ for any } w \in \mathbf{V} \right\}$.
3. Suppose $v_0 \in \mathbf{V}^{\perp}$, then $\langle v_0, v_0\rangle = 0$, then $v_0 = \mathbf{0}$.

Theorem 17.1. If $\mathbf{W} \prec \mathbf{V}$, where $\mathbf{V}$ is finite-dimensional, then $\mathbf{W} \oplus \mathbf{W}^{\perp} = \mathbf{V}$.

□

Proof. There are two things we need to do to prove this:

1. Showing that $\mathbf{W} \cap \mathbf{W}^{\perp} = \{\mathbf{0}\}$
2. $\mathbf{W} + \mathbf{W}^{\perp} = \mathbf{V}$.

The first item is already true, so we need to show the second item. Let $w_1, \ldots, w_k$ be a basis for $\mathbf{W}$ and let $v \in \mathbf{V}$ be given. We want to show that there exist $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$ such that $\sum_{i=1}^{k} \alpha_i w_i + y = v$ for some $y \in \mathbf{W}^{\perp}$, i.e., we want to show that there exist the α_i's such that

$$v - \sum_{i=1}^{k} \alpha_i w_i \in \mathbf{W}^{\perp},$$

i.e., there exist the α_i's such that for $j = 1, 2, 3, \ldots, k$

$$\langle v - \sum_{i=1}^{k} \alpha_i w_i, w_j\rangle = 0.$$

Thus we have a system of equations, which can be put into matrix form

$$\begin{bmatrix} \langle w_1, w_1\rangle & \ldots & \langle w_k, w_1\rangle \\ \vdots & \ddots & \vdots \\ \langle w_1, w_k\rangle & \ldots & \langle w_k, w_k\rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} \langle v, w_1\rangle \\ \vdots \\ \langle v, w_k\rangle \end{bmatrix}.$$

We shall show that the above matrix, called the **Gramian** matrix, is invertible. But since it is a square matrix, we can just show it is injective. Suppose that

$$G(\gamma) = \mathbf{0},$$

where $\gamma = \begin{bmatrix}\alpha_1 & \dots & \alpha_k\end{bmatrix}^\top$, then for any $j = 1, 2, \dots, k$

$$0 = \sum_{i=1}^{k} \gamma_i \langle w_i, w_j \rangle.$$

So, $\sum_{i=1}^k \gamma_i w_i \perp w_j$ for any $j = 1, 2, \dots, k$. In particular,

$$\sum_{i=1}^{k} \gamma_i w_i \perp \sum_{i=1}^{k} \gamma_i w_i \iff \sum_{i=1}^{k} \gamma_i w_i = \mathbf{0}.$$

Thus $\gamma_i 0$ for any $i = 1, 2, \dots, k$. Hence, G is injective, i.e., it is invertible. □

□

Note that this theorem **does not** hold for infinite dimensional cases. We will not go into the details of why this is the case, but it is an important thing to keep in mind.

Now, we make the following observation that an orthonormal list in an inner product space is linearly independent. This is quite easy to show. Suppose $\sum_{i=1}^k \alpha_i v_i = \mathbf{0}$ where $v_1, \dots, v_k$ are mutually orthogonal unit vectors. Then we have

$$0 = \langle \mathbf{0}, v_i \rangle = \alpha_i \langle v_i, v_i \rangle + 0 = \alpha_i.$$

This means all v_i's are linearly independent. We have a theorem.

Theorem 17.2. Every finite dimensional inner product space (that is non-trivial) has an orthonormal basis, and in fact any orthonormal list can be enlarged to one. □

17.2 Riesz representation theorem

Theorem 17.3. Riesz Representation Theorem Suppose $\mathbf{V}$ is an finite-dimensional inner product space (fdips) and $\rho : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbb{C}$. Then there exists exactly one $w_0 \in \mathbf{V}$ such that

$$\rho(v) = \langle v, w_0 \rangle$$

for any $v \in \mathbf{V}$. □

Proof. By rank-nullity theorem, we first have that

$$\text{rank}(\rho) + \text{nullity}(\rho) = \dim(\mathbf{V}).$$

Now, ρ maps to $\mathbb{C}$, so its rank is either 0 or 1. If its rank is zero,then $\rho \equiv 0$, so $w_0 = \mathbf{0}$ unique. If its rank if 1, then we look at

$$\ker(\rho) \oplus \ker \rho^\perp = \mathbf{V},$$

which tells us that $\ker(\rho)^\top$ has dimension one. Thus, $\ker(\rho)^\top = \text{span}(z_0)$, where z_0 is some unit vector. Now,

$$\rho(z_0) \neq 0 \implies \frac{1}{\rho(z_0)} z_0 \in \text{span}(z_0),$$

and

$$\rho\left(\frac{1}{\rho(z_0)}\right) = \rho(z_1) = \frac{1}{\rho(z_0)}\rho(z_0) = 1.$$

And so, $\text{span}(z_0) = \text{span}(z_1)$. Thus for any $v \in \mathbf{V}$, $v = w + \alpha z_1$ where $w \in \ker(\rho)$ and $z_1 \in \ker(\rho)^\perp$ uniquely. So,

$$\rho(v) = \rho(w + \alpha z_1) = 0 + \alpha\rho(z_1) = \alpha.$$

So the next question is for which vector h such that $\rho(v) = \langle v, h\rangle$ for all v? Well, we already know that $\rho(v) = \alpha$. After a bit of thinking, if we set $h = z_1/\|z_1\|$, then

$$\rho(v) = \rho(w + \alpha z_1) = \alpha\langle z_1, \frac{z_1}{\|z_1\|}\rangle = \alpha\frac{\|z_1\|}{\|z_1\|} = \alpha.$$

And so, let us pick $w_0 = \frac{z_1}{\|z_1\|}$, then we have

$$\rho(v) = \langle v, w_0\rangle$$

for any $v \in \mathbf{V}$. Showing uniqueness of w_0 is not too hard: Assume that $\rho(v) = \langle v, w_0\rangle = \langle v, w_1\rangle$. Then

$$\langle v, w_0\rangle - \langle v, w_1\rangle = \langle v, w_0 - w_1\rangle = 0 \iff w_0 = w_1.$$

This completes the proof. $\square$ $\square$

What are the consequences of this theorem? Well, suppose that we have $\mathcal{L} : \mathbf{V} \xrightarrow{\text{linear}} \mathbf{V}$, where $\mathbf{V}$ is an fdips. Then consider

$$f() := \langle \mathcal{L}(), y_0\rangle,$$

where $f : \mathbf{V} \xrightarrow{\text{linear}} \mathbb{C}$. So, f must be like

$$f() = \langle (), w_0\rangle$$

for some unique $w_0 \in \mathbf{V}$ and w_0 depends on y_0 and $\mathcal{L}$. So let us write $w_0 = \mathcal{L}(y_0)$. Now, consider $w_{\mathcal{L}} : \mathbf{V} \to \mathbf{V}$. Now,

$$f() = \langle (), w_0\rangle = \langle \mathcal{L}(), y_0\rangle \implies \langle x, w_{\mathcal{L}}(\alpha y_0)\rangle = \langle \mathcal{L}x, \alpha y_0\rangle = \cdots = \langle x, \alpha w_{\mathcal{L}}(y_0)\rangle.$$

So we have

$$w_{\mathcal{L}}(\alpha y_0) = \alpha w_{\mathcal{L}}(y_0).$$

We can indeed show that additivity also works here as well. And so $w_{\mathcal{L}} : \mathbf{V} lin \mathbf{V}$ has the property that

$$\langle \mathcal{L}(), y \rangle = \langle (), w_{\mathcal{L}}(y) \rangle.$$

We call $w_{\mathcal{L}}$ the **adjoint** of $\mathcal{L}$, denoted $\mathcal{L}^*$. Note that in this text I will use both $\mathcal{L}^*$ and $\mathcal{L}^\dagger$ notations. Another (mathematical) thing to notice is that $\mathcal{L}^*$ depends on the inner product, but not the basis.

17.3 Adjoints

Now, let's say we have $\mathcal{L} : \mathbf{V} \xrightarrow{\text{linear}} \mathbf{U}$ where both $\mathbf{V}, \mathbf{U}$ are fdips. Let Γ be an orthonormal basis for $\mathbf{V}$ and Ω be an orthonormal basis for $\mathbf{U}$. Then we can show that

$$\boxed{[\mathcal{L}^*]_{\gamma \leftarrow \Omega} = \left(\bar{[\mathcal{L}]}_{\Omega \leftarrow \Gamma} \right)^\top}$$

This is an exercise at the back. There's no good way to do this problem except entry-by-entry. Of course, "think indices."

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

The OPTIONAL problems covers this in detail. I will add the theorems to this section later. (source: Leo's Optional Practice Problem Set).

19 Ortho-triangularization

Theorem 19.1. Schur's Theorem: For each operator $\mathcal{L} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ where $\mathbf{V}$is an fdips there is an orthonormal basis Γ of $\mathbf{V}$ such that $[\mathcal{L}]_{\Gamma\leftarrow\Gamma}$ is upper triangular. $\square$

Proof. Here is a sketch of the proof. The full proof (I believe) is in the back. The theorem is trivially true for the one-dimensional case. So let us look at the two-dimensional case. Now, there is a unit eigenpair (λ, v_0) for $\mathcal{L}$ (existence of eigenvalues). Now, let $\mathbf{W} = \text{span}(v_0)$. This subspace is one-dimensional. Then, $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^{\perp}$. Let v_1 be a unit vector in $\mathbf{W}^{\perp}$. Then $\Gamma(v_0, v_1)$ is an orthonormal basis of $\mathbf{V}$. Then

$$[\mathcal{L}]_{\Gamma\leftarrow\Gamma} = \begin{bmatrix} \lambda & \square \\ & \triangle \end{bmatrix},$$

which is upper triangular. Next, we look at the three-dimensional case. Once again,

$$[\mathcal{L}]_{\Gamma\leftarrow\Gamma} = \begin{bmatrix} \lambda & \square \\ & \mathcal{M} \end{bmatrix},$$

where $\mathcal{M}$ is a 2×2 block. But we have already proven the two-dimensional case (we can find a basis such that $\mathcal{M}$ is upper triangular...) And so by induction we get $[\mathcal{L}]_{\Gamma\leftarrow\Gamma}$ upper-triangular. $\square$

20 Ortho-diagonalization; Self-adjoint and Normal Operators; Spectral Theorems

20.1 Towards the Spectral Theorem

We make a few observations before diving into the good stuff:

1. Suppose $\mathcal{M} \in \mathbb{M}_n$, and $\mathcal{M} = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}$ then $\mathcal{M}^* = \begin{bmatrix} \mathcal{A}^* & \mathcal{C}^* \\ \mathcal{B}^* & \mathcal{D}^* \end{bmatrix}$.

2. Suppose $\mathcal{M}$ is normal, then $\mathcal{M}\mathcal{M}^* = \mathcal{M}^*\mathcal{M}$, which means that if $\mathcal{M} = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}$ then (by the fact that $\operatorname{tr}(\mathcal{A}^*\mathcal{A}) = \operatorname{tr}(\mathcal{A}\mathcal{A}^*)$ and $\operatorname{tr}(\mathcal{A}^*\mathcal{A}) = 0 \iff \mathcal{A} = 0$) we have $\operatorname{tr}(\mathcal{B}\mathcal{B}^*) = \operatorname{tr}(\mathcal{C}^*\mathcal{C})$.

3. Suppose that $\mathcal{A}$ is block-upper triangular i.e., $\mathcal{C} = 0$, and that $\mathcal{A}$ is normal, then $\mathcal{A}$ has to be diagonal. And so every normal operator has a matrix representation that is diagonal.

4. What about the converse? - we might ask ourselves.

20.2 The Spectral Theorem

The following statements are equivalent:

1. $\mathcal{A}$ is normal then there is an orthonormal basis of the underlying space made up entirely of the eigenvectors of $\mathcal{A}$.

2. $\mathcal{A} : \mathbf{V} \xrightarrow{\text{linear}} \mathbf{V}$, then there exist an appropriate orthonormal basis Γ such that $[\mathcal{A}]_{\Gamma \leftarrow \Gamma}$ is diagonal.

3. $\mathbf{V} = \bigoplus_{i=1}^{k} \mathbf{E}_{\lambda_i}$ where $\mathbf{E}_{\lambda_i}$ are mutually orthogonal λ_i-eigenspaces of $\mathcal{A}$.

4. **Spectral Resolution:** $\mathcal{A} = \sum_{i=1}^{k} \lambda_i \mathcal{E}_i$, where $\mathcal{E}_i$'s are **ortho-projections** (which are non-negative self-adjoint operators, i.e., $\mathcal{E}_i^2 = \mathcal{E}_i = \mathcal{E}_i^*$ and $\sigma_{\mathbb{C}}(\mathcal{E}_i) \subseteq \{0, 1\}$) and the distinct λ_i's are exactly the eigenvalues of $\mathcal{A}$.

It's probably crucial that we make the following distinction:

1. Diagonalizable $\implies$ operator can be written as a linear combination of idempotents which resolve identity.

2. Normal $\implies$ operator can be written as a linear combination of **ortho-projections**

Recall that for idempotents, $\mathbf{V} = \operatorname{Im}(\mathcal{E}) \oplus \ker(\mathcal{E})$. For ortho-projections, $\operatorname{Im}(\mathcal{E}) \perp \ker(\mathcal{E})$.

20.3 Spectral Theorem for Matrices

It follows naturally that

Theorem 20.1. Matrix $\mathcal{M}$ is normal $\iff$ $\mathcal{M}$ is unitarily equivalent to a diagonal matrix. $\square$

In fact, the right-to-left direction holds by a simple check. Suppose that $\mathcal{A}^*\mathcal{M}\mathcal{A} = \mathcal{D}$ is diagonal, and $\mathcal{A}$ is unitary. Does this mean $\mathcal{M}$ is normal? The answer is YES. $\mathcal{M} = \mathcal{A}\mathcal{D}\mathcal{A}^*$ then $\mathcal{M}^* = \mathcal{A}\mathcal{D}^*\mathcal{A}^*$. And so,

$$\mathcal{M}^*\mathcal{M} = \mathcal{A}\mathcal{D}^*\mathcal{A}^*\mathcal{A}\mathcal{D}\mathcal{A}^* = \mathcal{A}\mathcal{D}^*\mathcal{D}\mathcal{A}^* = \mathcal{A}\mathcal{D}\mathcal{D}^*\mathcal{A}^* = \mathcal{A}\mathcal{D}\mathcal{A}^*\mathcal{A}\mathcal{D}^*\mathcal{A}^* = \mathcal{M}\mathcal{M}^*.$$

So $\mathcal{M}$ is indeed normal. $\square$

Theorem 20.2. Schur's Theorem - extended Every $n \times n$ matrix is unitarily equivalent to a triangular matrix. $\square$

21 Positive (semi-)definite operators

Now, if we go back to normal matrices. If $\mathcal{A}$ is normal, then $\mathcal{A} = \sum_{i=1}^k \lambda_i \mathcal{E}_i$ where λ_i's are the eigenvalues of $\mathcal{A}$ and $\mathcal{E}_i$'s are ortho-projections that resolve identity. Now, what else can we "mine" from the spectral theorem? Well, if A is normal, then it is unitarily equivalent to a diagonal matrix, so $\mathcal{A} = \mathcal{U}^*\mathcal{D}\mathcal{U}$ where $\mathcal{D}$ is diagonal whose entries are the eigenvalues of $\mathcal{A}$. Then it follows that $\mathcal{A}^2 = \mathcal{U}^*\mathcal{D}^2\mathcal{U}$. In fact, if we define

$$f(\mathcal{A}) := \sum_{i=1}^k f(\lambda_i)\mathcal{E}_i$$

then we have

$$f(\mathcal{A}) = \mathcal{U}^* \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_k) \end{bmatrix}.$$

Now, let's go a bit further. Suppose that $\mathcal{A}$ is not only normal but also has $\sigma_{\mathbb{C}}(\mathcal{A}) \subseteq [0, \infty)$ (so $\mathcal{A}$ is now non-negative). Then

$$\sqrt{\mathcal{A}} \cdot \sqrt{\mathcal{A}} = \left(\sum_{i=1}^k \sqrt{\lambda_i} E_i\right)\left(\sum_{j=1}^k \sqrt{\lambda_j}\mathcal{E}_j\right) = \sum_{i=1}^k \lambda_i \mathcal{E}_i = \mathcal{A}.$$

But if we look at $\sqrt{\mathcal{A}}$, we can see that $\sqrt{\mathcal{A}}$ is also non-negative, hence self-adjoint, hence normal. And so we have

Theorem 21.1. Every non-negative normal matrix $\mathcal{A}$ has a **unique** non-negative normal square root $\sqrt{\mathcal{A}}$. □

The uniqueness proof is provided in the PRACTICE problem set. I will reproduce the proof here later.

21.1 Underlying Space Resolution For Normal Operators

Theorem 21.2. Suppose that $\mathbf{V}$ is an fdips, and $\mathcal{L}$ is a linear operator on $\mathbf{V}$, then

$$\begin{aligned} \ker(\mathcal{L}^*) &= \mathrm{Im}(\mathcal{L})^\perp \\ \ker(\mathcal{A})^\perp &= \ker(\mathcal{A}^{**})^\perp = \mathrm{Im}(\mathcal{A}^*). \end{aligned}$$

Proof. Then consider the following string of arguments (for any $x \in \mathbf{V}$)

$$\begin{aligned} y \in \mathrm{Im}(\mathcal{L})^\perp &\iff \langle \mathcal{L}x, y\rangle = 0 \\ &\iff \langle x, \mathcal{L}^* y\rangle = 0 \\ &\iff \mathcal{L}^* y = \mathbf{0} \\ &\iff y \in \ker(\mathcal{L}^*). \end{aligned}$$

□ □

Now, it turns out that

$$\ker(\mathcal{A}^*\mathcal{A}) = \ker(\mathcal{A})$$
$$\ker(\mathcal{A}\mathcal{A}^*) = \ker(\mathcal{A}^*)$$

Proof.

$$\langle \mathcal{A}^*\mathcal{A}x, x\rangle = \langle \mathcal{A}x, \mathcal{A}x\rangle = \|\mathcal{A}x\|^2.$$

If $\mathcal{A}^*\mathcal{A}x = \mathbf{0}$, then $\|\mathcal{A}x\|^2 = 0$, then $\mathcal{A}x = 0$. And if $\mathcal{A}x = \mathbf{0}$, then $\mathcal{A}^*\mathcal{A}x = \mathbf{0}$, so $\langle \mathcal{A}^*\mathcal{A}x, x\rangle = 0$. ☐

And so we have the following:

$$\mathcal{A}x = \mathbf{0} \iff \mathcal{A}^*\mathcal{A}x = \mathbf{0}, \text{ i.e., } \ker(\mathcal{A}) = \ker(\mathcal{A}^*\mathcal{A}).$$

So now, we automatically have that (using the fact that $\mathcal{A}^*\mathcal{A}$ is self-adjoint - one can check this by inspection)

$$\begin{aligned} \ker(\mathcal{A}^*\mathcal{A}) &= \ker(\mathcal{A}) \\ (\ker(\mathcal{A}^*\mathcal{A}))^\perp &= \ker(\mathcal{A})^\perp \\ \operatorname{Im}((\mathcal{A}^*\mathcal{A})^*) &= \operatorname{Im}(\mathcal{A}^*) \\ \operatorname{Im}(\mathcal{A}^*\mathcal{A}) = \operatorname{Im}(\mathcal{A}^*). \end{aligned}$$

Similarly, we also have

$$\operatorname{Im}(\mathcal{A}\mathcal{A}^*) = \operatorname{Im}(\mathcal{A}).$$

Now, $x \in \ker(\mathcal{A}^*) \iff \mathcal{A}^*x = \mathbf{0} \iff \|\mathcal{A}^*x\|^2 = 0 \iff \langle \mathcal{A}^*x, \mathcal{A}^*x\rangle = 0 \iff \langle \mathcal{A}\mathcal{A}^*x, x\rangle = 0$. Similarly, $x \in \ker(\mathcal{A}) \iff \langle \mathcal{A}^*\mathcal{A}x, x\rangle = 0$. And so, if $\mathcal{A}$ is normal, then $\ker(\mathcal{A}) = \ker(\mathcal{A}^*)$. Thus

Theorem 21.3. $\mathcal{A}$ normal $\implies \ker(\mathcal{A}) = \ker(\mathcal{A}^*) = \operatorname{Im}(\mathcal{A})^\perp$.

Now, since we also know that $\mathcal{A}^*\mathcal{A}$ is non-negative (so it's self-adjoint), there is **exactly** one $\sqrt{\mathcal{A}^*\mathcal{A}}$ non-negative. And similarly, $\mathcal{A}\mathcal{A}^*$ is non-negative, so there is **exactly** one $\sqrt{\mathcal{A}\mathcal{A}^*}$ non-negative. Now, we observe that

$$\begin{aligned} x \in \ker(\sqrt{\mathcal{A}^*\mathcal{A}}) &\iff \left\|\sqrt{\mathcal{A}^*\mathcal{A}x}\right\|^2 = 0 \\ &\iff \langle \sqrt{\mathcal{A}^*\mathcal{A}}x, \sqrt{\mathcal{A}^{\mathcal{A}}}x\rangle = 0 \\ &\iff \langle \sqrt{\mathcal{A}^*\mathcal{A}}^*\sqrt{\mathcal{A}^*\mathcal{A}}x, x\rangle = 0 \\ &\iff \langle \mathcal{A}^*\mathcal{A}x, x\rangle = 0 \\ &\iff \|\mathcal{A}x\|^2 = 0 \\ &\iff x \in \ker(\mathcal{A}), \end{aligned}$$

since $\mathcal{A}$ is nn-negative. Therefore, we have just shown that

$$\ker(\mathcal{A}^*\mathcal{A}) = \ker(\sqrt{\mathcal{A}^*\mathcal{A}}) = \ker(\mathcal{A}).$$

And similarly, of course,

$$\ker(\mathcal{A}\mathcal{A}^*) = \ker(\sqrt{\mathcal{A}\mathcal{A}^*}) = \ker(\mathcal{A}^*).$$

Lastly,

$$\operatorname{Im}(\sqrt{\mathcal{A}^*\mathcal{A}}) = \left(\ker((\sqrt{\mathcal{A}^*\mathcal{A}})^*)\right)^\perp = \ker(\sqrt{\mathcal{A}^*\mathcal{A}})^\perp = \ker(\mathcal{A})^\perp = \operatorname{Im}(\mathcal{A}^*).$$

So we just showed that

$$\begin{aligned}\operatorname{Im}(\sqrt{\mathcal{A}^*\mathcal{A}}) &= \operatorname{Im}(\mathcal{A}^*\mathcal{A}) = \operatorname{Im}(\mathcal{A}^*)\\ \operatorname{Im}(\sqrt{\mathcal{A}\mathcal{A}^*}) &= \operatorname{Im}(\mathcal{A}\mathcal{A}^*) = \operatorname{Im}(\mathcal{A}).\end{aligned}$$

Now, as we will show in one of the practice problems, we have the following amazing theorem:

Theorem 21.4.

$$\left\|\sqrt{\mathcal{A}^*\mathcal{A}}x\right\| = \|\mathcal{A}x\|.$$

□

21.2 Classification of inner products

21.3 Positive square roots

22 Polar decomposition

Let $\mathcal{A}$ any linear function $\mathcal{A} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{W}$ be given. Let us define a new function: $\phi : \ker(\mathcal{A})^{\perp} \to \mathrm{Im}(\mathcal{A})$ as

$$\phi\left[\sqrt{\mathcal{A}^*\mathcal{A}}x\right] := \mathcal{A}x.$$

It can be easily shown that ϕ is not only linear but it is also an isometry (length-preserving). The key here is to recognize that

$$\phi = \mathcal{A} \circ \left[\sqrt{\mathcal{A}^*\mathcal{A}}\Big|_{\ker(\mathcal{A})^{\perp}}\right]^{-1}.$$

Now, let us turn $\mathcal{A}$ into an operator by requiring the $\mathbf{W}$ is $\mathbf{V}$ itself, then $\dim(\ker(\mathcal{A})) = \dim(\mathrm{Im}(\mathcal{A})^{\perp})$. Then, we can define another isometry $\psi : \ker(\mathcal{A}) \to \mathrm{Im}(\mathcal{A})^{\perp}$. To do this, we simply pick an orthonormal basis Γ of $\ker(\mathcal{A})$ and an orthonormal basis Ω of $\mathrm{Im}(\mathcal{A})^{\perp}$, then send one basis element to another one in the other space. It is very easy to show that ψ is an isometry.

Now, by rank-nullity theorem, $\dim(\ker(\mathcal{A})) = \dim(\mathrm{Im}(\mathcal{A})^{\perp}) = \dim(\mathbf{V}) - \dim(\mathrm{Im}(\mathcal{A}))$. Now, pick any isometry ψ and define $\mathcal{U} : \mathbf{V} \to \mathbf{V}$ by

$$\mathcal{U}(v_1 + v_2) = \phi(v_1) + \psi(v_2).$$

And hence, we have just shown that

$$\mathcal{A} = \mathcal{U} \circ \sqrt{\mathcal{A}^*\mathcal{A}}.$$

This $\mathcal{U}$ is an isometry, and it is surjective, hence it is injective, i.e, it is **unitary**.

Similarly, if we start with the adjoint of $\mathcal{A}$, then we will end up with

$$\mathcal{A} = \sqrt{\mathcal{A}\mathcal{A}^*} \circ \hat{\mathcal{U}},$$

where $\hat{(\mathcal{U})}$ is also unitary. So...

Theorem 22.1. For any $\mathcal{A} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$, there exists a unitary $\mathcal{U}_1 : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ such that $\mathcal{A} = \sqrt{\mathcal{A}\mathcal{A}^*} \circ \mathcal{U}_1$ and there exists a unitary $\mathcal{U}_2 : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ such that $\mathcal{A} = \mathcal{U}_2 \circ \sqrt{\mathcal{A}^*\mathcal{A}}$. □

Now, if we think back to complex numbers where we have $z = |z|e^{i\theta}$, this is the analog of that, but in matrices and operators. Hence it makes sense to call this decomposition the **polar decomposition** of $\mathcal{A}$.

But there's more to the story. It turns out that we can use the same $\mathcal{U}$ for $\mathcal{U}_1$ and $\mathcal{U}_2$. Here's the proof:

Proof.

$$\mathcal{A} = \mathcal{U} \circ \sqrt{\mathcal{A}^*\mathcal{A}} = \left(\mathcal{U} \circ \sqrt{\mathcal{A}^*\mathcal{A}}\mathcal{U}^*\right)\mathcal{U} = \mathcal{N}\mathcal{U}.$$

Recognize that

$$\mathcal{N}^2 = (\mathcal{U}\mathcal{A}^*)(\mathcal{A}\mathcal{U}^*) = \mathcal{U}\sqrt{\mathcal{A}^*\mathcal{A}}\mathcal{U}^*\mathcal{U}\sqrt{\mathcal{A}^*\mathcal{A}}\mathcal{U}^* = \mathcal{A}\mathcal{A}^*,$$

which says $\mathcal{N}$ is exactly $\sqrt{\mathcal{A}^*\mathcal{A}}$. So,

$$\mathcal{A} = \mathcal{N}\mathcal{U} == \sqrt{\mathcal{A}^*\mathcal{A}}\mathcal{U} \implies \mathcal{A} = \mathcal{U}\sqrt{\mathcal{A}^*\mathcal{A}} = \sqrt{\mathcal{A}\mathcal{A}^*}\mathcal{U}.$$

So we can use the same unitary in the power decomposition. But note that the unitary $\mathcal{U}$ itself is not necessarily unique. $\mathcal{U}$ is only unique if $\mathcal{A}$ is invertible. $\square$

23 Singular value decomposition

Now, using our results from the polar decomposition of operators, we can imagine a case for $\mathcal{A}$ a square matrix. Then, $\mathcal{A} = \mathcal{U}\sqrt{\mathcal{A}^*\mathcal{A}}$ as before, where $\sqrt{\mathcal{A}^*\mathcal{A}}$ is non-negative, self-adjoint, normal, etc. And so, we can write $\mathcal{A}$ as

$$\mathcal{A} = \mathcal{U}\left(\hat{\mathcal{U}}\mathcal{D}\hat{\mathcal{U}}^*\right)$$

where $\mathcal{D}$ is a diagonal matrix. This says that

$$\mathcal{A} = (\mathcal{U}\hat{\mathcal{U}})\mathcal{D}\hat{\mathcal{U}}^{-1} = \mathcal{U}_1\mathcal{D}\mathcal{U}_2$$

and

$$\mathcal{D} = \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_k \end{bmatrix}$$

where s_i's are exactly the eigenvalues of $\sqrt{\mathcal{A}^*\mathcal{A}}$. They are also called the **singular values** of $\mathcal{A}$. And thus, this is the **Singular Value Decomposition (SVD)** for square matrices.

23.1 Spectral/operator norm

23.2 Singular values and approximation

23.3 Singular values and eigenvalues

24 Problems and Solutions

24.1 Problem set 1 (under corrections)

Problem. 1. Equivalent formulations of directness of a subspace sum: Suppose that $\mathbf{V}_1, \mathbf{V}_2 \ldots, \mathbf{V}_{315}$ are subspaces of a vector space $\mathbf{W}$. Argue that the following claims are equivalent.

1. The subspace sum $\mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_{315}$ is direct.

2. If $x_i \in \mathbf{V}_i$ and $x_1 + x_2 + \cdots + x_{315} = \mathbf{0_W}$, then $x_i = \mathbf{0_W}$, for every i.

3. If $x_i, y_i \in \mathbf{V}_i$ and $x_1 + x_2 + x_3 + \cdots + x_{315} = y_1 + y_2 + y_3 + \cdots + y_{315}$, then $x_i = y_i$, for every i.

4. For any i, no non-null element of $\mathbf{V}_i$ can be express as a sum of the elements of the other V_j's.

5. For any i, no non-null element of $\mathbf{V}_i$ can be expressed as a sum of the elements of the preceding V_j's.

Problem. 2. Sub-sums of direct sums are direct: Suppose that $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{315}$ are subspaces of a vector space $\mathbf{W}$, and the subspace sum $\mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_{315}$ is direct.

1. Suppose that for each i, $\mathbf{Z}_i$ is a subspace of $\mathbf{V}_i$. Argue that the subspace sum $\mathbf{Z}_1 + \mathbf{Z}_2 + \ldots \mathbf{Z}_{315}$ is direct.
2. Argue that the sum $\mathbf{V}_2 + \mathbf{V}_5 + \mathbf{V}_7 + \mathbf{V}_{12}$ is also direct.

Problem. 3. Associativity of directness of subspace sums: Suppose that

$$\mathbf{Y}, \mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_5, \mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_{12}, \mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_4, \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{53}$$

are subspaces of a vector space $\mathbf{W}$. Argue that the following claims are equivalent.

1. The subspace sum

   $$\mathbf{Y} + \mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_5 + \mathbf{U}_1 + \mathbf{U}_2 + \cdots + \mathbf{U}_{12} + \mathbf{Z}_1 + \mathbf{Z}_2 + \cdots + \mathbf{Z}_4 + \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_{53}$$

   is direct.

2. The subspace sums

   $$\begin{array}{l} [\mathbf{V} :=]\, \mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_5 \\ [\mathbf{U} :=]\, \mathbf{U}_1 + \mathbf{U}_2 + \cdots + \mathbf{U}_{12} \\ [\mathbf{Z} :=]\, \mathbf{Z}_1 + \mathbf{Z}_2 + \cdots + \mathbf{Z}_4 \\ [\mathbf{X} :=]\, \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_{53} \\ \mathbf{Y} + \mathbf{V} + \mathbf{U} + \mathbf{X} + \mathbf{Z} \end{array}$$

   are all direct.

Problem. 4. Direct sums preserve linear independence: Suppose that $\mathbf{U}, \mathbf{V}, \mathbf{W}$ are subspaces of a vector space $\mathbf{Z}$, and the sum $\mathbf{U}, \mathbf{V}, \mathbf{W}$ is direct.

1. Suppose that $U_1, U_2, \ldots, U_{13}$ is a linearly independent list in $\mathbf{U}$, $V_1, V_2, \ldots, V_6$ is a linearly independent list in $\mathbf{V}$, and $W_1, W_2, \ldots, W_{134}$ is a linearly independent list in $\mathbf{W}$. Argue that the concatenated list

$$U_1, U_2, \ldots, U_{13}, V_1, V_2, \ldots, V_6, W_1, W_2, \ldots, W_{134}$$

   is linearly independent.

2. Suppose that $U_1, U_2, \ldots, U_{13}$ is a basis of $\mathbf{U}$, $V_1, V_2, \ldots, V_6$ is a basis of $\mathbf{V}$, and $W_1, W_2, \ldots, W_{134}$ is a basis of $\mathbf{W}$. Argue that the concatenated list

$$U_1, U_2, \ldots, U_{13}, V_1, V_2, \ldots, V_6, W_1, W_2, \ldots, W_{134}$$

   is a basis of $\mathbf{U} \oplus \mathbf{V} \oplus \mathbf{W}$.

Problem. 5. Suppose that $x_1, x_2, \ldots, x_{14}$ are non-null elements of a linear space $\mathbf{W}$.

1. Argue that the following claims are equivalent.

   (a) $x_1, x_2, \ldots, x_{14}$ are linearly independent.

   (b) The subspace sum

   $$\operatorname{span}(x_1) + \operatorname{span}(x_2) + \cdots + \operatorname{span}(x_{14})$$

   is direct.

2. Argue that the following claims are equivalent.

   (a) $x_1, x_2, \ldots, x_{14}$ is a basis of $\mathbf{W}$.

   (b) $\mathbf{W} = \operatorname{span}(x_1) \oplus \operatorname{span}(x_2) \oplus \cdots \oplus \operatorname{span}(x_{14})$

24.2 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, each $\mathbf{W}_i$ contains an element w_i that does not belong to any other $\mathbf{W}_j$'s.

  Take $w_2 \in \mathbf{W}_2 \setminus \mathbf{W}_1$ and $w_1 \in \mathbf{W}_1$ 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 $\mathbf{W}_1, \mathbf{W}_2 \subseteq \mathbf{S}$. And since $\mathbf{S}$ is a subspace, $mw_1 + w_2 \in \mathbf{S}$ for any $m \in \mathbb{C}$ by closure under addition.

  Consider a collection of n_0 linear combinations of w_1 and w_2, defined as

$$T = \{w_1 + w_2, 2w_1 + w_2, \ldots, n_0 w_1 + w_2\} \subseteq \mathbf{S}.$$

  Consider a typical element of T: $nw_1 + w_2 \in \mathbf{S}$, $n \in \{1, 2, \ldots, n_0\}$. Suppose $nw_1 + w_2 \in \mathbf{W}_1$. By closure under addition,

$$(nw_1 + w_2) - nw_1 = w_2 \in \mathbf{W}_1.$$

  But this contradicts the choice of w_2. Therefore, $nw_1 + w_2 \notin \mathbf{W}_1$, for any $n \in \{1, 2, \ldots, n_0\}$. It follows that $T \subseteq \bigcup_{i=2}^{n_0} \mathbf{W}_i$.

  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 $\mathbf{W}_2, \ldots, \mathbf{W}_{n_0}$ must contain $nw_1 + w_2$ for two distinct values of n. Let these values be n_a and n_b. By closure under addition, we have

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

  But since $n_a \neq n_b$, $w_1 \in \mathbf{W}_k$. This contradicts our initial choice of w_1 and thus rules out the possibility that the implication $1. \implies 2.$ fails to hold. Therefore, the implication $1. \implies 2.$ must be true.

- [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 $\mathbf{W}$,

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\},$$

has two elements. Next, consider the vector space $\mathbb{C}^3$ over the complex numbers. Since $\mathbf{W}$ is a subspace and any element of $\mathbf{W}$ is contained in $\mathbb{C}^3$ by construction, $\mathbf{W} \prec \mathbb{C}^3$. In fact, $\mathbf{W} \subsetneq \mathbb{C}^3$ because $\mathbb{C}^3 \ni (0 \;\; 0 \;\; 1)^\top \notin \mathbf{W}$

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

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

Now, since $\dim(\mathbf{W}) = 2$ and $\mathcal{A}$ is a 3×3 matrix (i.e., neither $\dim(\ker(\mathcal{A}))$ nor $\dim(\mathrm{Im}(A))$ can exceed $3^{(\dagger)}$), 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,

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

It follows, by fact ($\dagger$), that

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

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

So, such a 3×3 matrix $\mathcal{A}$ that satisfies the given condition is

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

First, $\mathbf{W}$ is invariant under $\mathcal{A}$ because take any $(x\ \ y\ \ 0)^\top \in \mathbf{W}$, $\mathcal{A}\left((x\ \ y\ \ 0)^\top\right) = (x\ \ 0\ \ 0)^\top \in \mathbf{W}$. This 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 $\mathcal{A}\left((x\ \ y\ \ z)^\top\right) = (x\ \ 0\ \ 0)^\top \in \mathbf{W}$. But observe that $\mathrm{Im}(\mathcal{A})$ is a 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, because $\dim\left(A^{-1}\left[\mathcal{A}[\mathbf{W}]\right]\right) = 3$ and $A^{-1}\left[\mathcal{A}[\mathbf{W}]\right] \prec \mathbb{C}^3$, $A^{-1}\left[\mathcal{A}[\mathbf{W}]\right] = \mathbb{C}^3 \supsetneq \mathbf{W}$ (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. 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 $v \in \mathbf{V}, w \in \mathbf{W}$. Then $v + w \in \mathbf{V} + \mathbf{W}$. Next, consider $l \in \mathcal{L}[\mathbf{U} + \mathbf{W}]$ given by

   $$l = \mathcal{L}(v + w) = \mathcal{L}(v) + \mathcal{L}(w) = v' + w',$$

   where $v' \in \mathbf{V}$ and $w' \in \mathbf{W}$ because $\mathbf{V}, \mathbf{W}$ are invariant under $\mathcal{L}$. So $l \in \mathbf{V} + \mathbf{W}$. This implies $\mathcal{L}[\mathbf{V} + \mathbf{W}] \subseteq \mathbf{V} + \mathbf{W}$; i.e., $\mathfrak{Lat}(\mathcal{L})$ is closed under subspace sums.

3. Let $\mathbf{V} \in \mathfrak{Lat}(\mathcal{L})$ be given. By definition, $\mathcal{L}[\mathbf{V}] \subseteq \mathbf{V}$. It follows that $\mathcal{L}[\mathcal{L}[\mathbf{V}]] \subseteq \mathcal{L}[\mathbf{V}]$. 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, $\mathcal{L}^{-1}[\mathcal{L}[\mathbf{W}]] = \{v \in \mathbf{V} | \mathcal{L}(v) \in \mathcal{L}[\mathbf{W}]\}$, 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

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

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 that 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 $k \geq l \geq 0$. It is clear that $\mathbf{P}(\mathcal{L}, v_0)$ is closed under addition because,

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

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 $P(\mathcal{L}, v_0) \in \mathbf{P}(\mathcal{L}, v_0)$.

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

where $a_i \in \mathbb{C}$ and $k \geq 0$. Define new 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 $\mathrm{span}(v_0)$. We know that $\dim(\mathrm{span}(v_0)) = 1$. Consider $v \in \mathrm{span}(v_0)$. 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, $\mathrm{span}(v_0) \subseteq \mathbf{P}(\mathcal{L}, v_0)$. But because $\dim(\mathbf{P}(\mathcal{L}, v_0)) = 1 = \dim(\mathrm{span}(v_0))$, $\mathbf{P}(\mathcal{L}, v_0) = \mathrm{span}(v_0)$; 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 $P(\mathcal{L}, v_0) = \mathcal{L}^0(v_0) + \mathcal{L}(v_0) \in \mathbf{P}(\mathcal{L}, v_0) = \mathrm{span}(v_0)$, $v_0 \neq \mathbf{0_V}$. By closure under addition,

   $$(\mathcal{L}^0(v_0)+\mathcal{L}(v_0))-v_0 = (\mathcal{L}^0(v_0)+\mathcal{L}(v_0))-\mathcal{L}^0(v_0) = \mathcal{L}(v_0) \in \mathrm{span}(v_0).$$

   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, $\mathbf{P}(\mathcal{L}, v_0) = \mathrm{span}(v_0)$; 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 true 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 $\mathbf{P_i}$ denotes $\mathbf{P}(\mathcal{L}, v_i)$ 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 $\mathbf{P_i}$, $1 \leq i \leq n$, must contain all the other $\mathbf{P_j}$'s. Without loss of generality, assume that 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, $\text{Im}(\mathcal{M}^k)$ 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 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>: $\mathcal{L}\left[\text{Im}(\mathcal{M}^k)\right] \subseteq \text{Im}(\mathcal{M}^k)$.

   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., $\mathcal{L}[\text{Im}(\mathcal{M}^k)] = \mathcal{M}^k[\text{Im}(\mathcal{L})]$. But since $\text{Im}(\mathcal{L}) \subseteq \mathbf{V}$ (because $\mathcal{L} \in \mathfrak{L}(\mathbf{V})$),

$$\mathcal{M}^k[\text{Im}(\mathcal{L})] \subseteq \mathcal{M}^k[\mathbf{V}] = \text{Im}(\mathcal{M}^k).$$

   It follows that $\mathcal{L}[\text{Im}(\mathcal{M}^k)] \subseteq \text{Im}(\mathcal{M}^k)$. So, $\text{Im}(\mathcal{M}^k)$ 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 $\mathcal{A}[\mathbf{J}] = \{\mathbf{0}\} \subseteq \mathbf{J}$. 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 $n > 1$ based on the 2×2 example above. Let $\mathcal{A}$ be an $n \times n$ zero matrix, and $\mathcal{B}$ be an $n \times n$ matrix of the form

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

It is clear that $\mathcal{A}$ and $\mathcal{B}$ commute, since $\mathcal{A}$ is the zero matrix:

$$\mathcal{A}\mathcal{B} = [0]_{n\times n} = \mathcal{B}\mathcal{A}.$$

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 $\mathcal{A}(j') = \mathbf{0} \in \mathbf{J}$ 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:

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

   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) <u>To show</u>: $\cdots \subseteq \mathcal{L}^3[\mathbf{W}] \subseteq \mathcal{L}^2[\mathbf{W}] \subseteq \mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$.

   Assume that $\mathcal{L}^{k+1}[\mathbf{W}] \subseteq \mathcal{L}^k[\mathbf{W}]$ for all $k \geq 0$. The base case $k = 0$ is true because $\mathbf{W}$ is an invariant subspace under $\mathcal{L}$. 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, $\cdots \subseteq \mathcal{L}^3[\mathbf{W}] \subseteq \mathcal{L}^2[\mathbf{W}] \subseteq \mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$.

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

   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 $\mathcal{L}^{-1}[\mathbf{W}] = \{v \in \mathbf{V} | \mathcal{L}(v) \in \mathbf{W}\}$ and $\mathcal{L}[\mathbf{W}] \subseteq \mathbf{W}$, $w \in \mathcal{L}^{-1}[\mathbf{W}]$. Therefore, $\mathbf{W} \subseteq \mathcal{L}^{-1}[\mathbf{W}]$ (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, $\mathbf{W} \subseteq \mathcal{L}^{-1}[\mathbf{W}] \subseteq \mathcal{L}^{-2}[\mathbf{W}] \subseteq \dots$ (ii).

   From (i) and (ii),

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

2. (a) Assume the hypothesis that $\mathcal{L}^{k+1}[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}]$ for some $k \geq 0$. We can prove by induction that $\mathcal{L}^{k+j}[\mathbf{W}] = \mathcal{L}^k[\mathbf{W}]$ for all non-negative j. The base case where $j = 0$ is 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 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

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

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 minimal. It 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 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, $\mathcal{L}^k[\mathbf{W}] = \mathcal{L}^l[\mathbf{W}]$ for any $k \geq l$, which follows from (a). Therefore, the chain in part 1. stabilizes in the "inclusion by" direction as well.

We claim that 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:

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

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 satisfied and consider the next proper inclusion in the chain, a similar argument shows that the invariant subspace of interest has 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 $\{\mathbf{0}_{\mathbf{V}}\}$. 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 nonnegative integer k. $\mathbf{M}$ is a subspace. We have that

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

Let $\mathcal{L}^{-h}[\mathbf{W}] \subseteq \mathbf{M} \subseteq \mathcal{L}^{-(h+1)}[\mathbf{W}]$ be given for some nonnegative integer h. We have that

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

But because $\mathcal{L}^{-1}[\mathbf{M}] = \{v \in \mathbf{V} | \mathcal{L}(v) \in \mathbf{M}\}$, $\mathcal{L}[\mathcal{L}^{-1}[\mathbf{M}]] \subseteq \mathbf{M}$. Therefore, we have $\mathcal{L}[\mathbf{M}] \subseteq \mathbf{M}^{(ii)}$.

From (i) and (ii), $\mathbf{M}$ is invariant under $\mathcal{L}$ if $\mathbf{M}$ is a subspace that falls between two consecutive subspaces in the chain shown in part 1.

24.3 Problem set 3

Problem. 1. Suppose that $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ are non-trivial vector spaces, and for each $i, j \in \{1, 2, 3\}$,

$$\mathcal{L}_{ij} : \mathbf{V}_j \overset{\text{linear}}{\longrightarrow} \mathbf{V}_i.$$

1. Argue that

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times$$

   is a linear function.

2. Argue that every linear function $\mathcal{T} : \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3 \to \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$ can be expressed as

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times ,$$

   with $\mathcal{L}_{ij} : \mathbf{V}_j \overset{\text{linear}}{\longrightarrow} \mathbf{V}_i$, in exactly one way.

3. If the function

$$\begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\times : \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3 \to \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$$

   is defined similarly, find (with proof) the block-matrix form of the functions

$$2 \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times + 3 \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\times$$

   and

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \circ \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\times .$$

Solution. 1.

1. (a) It suffices to show that $\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times$ satisfies the linearity conditions.

Consider

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$$

where $a_j, b_j \in \mathbf{V}_j$ for $j \in \{1, 2, 3\}$. By definition,

$$\begin{aligned}
&\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{L}_{11}(a_1) + \mathcal{L}_{12}(a_2) + \mathcal{L}_{13}(a_3) \\ \mathcal{L}_{21}(a_1) + \mathcal{L}_{22}(a_2) + \mathcal{L}_{23}(a_3) \\ \mathcal{L}_{31}(a_1) + \mathcal{L}_{32}(a_2) + \mathcal{L}_{33}(a_3) \end{pmatrix} + \begin{pmatrix} \mathcal{L}_{11}(b_1) + \mathcal{L}_{12}(b_2) + \mathcal{L}_{13}(b_3) \\ \mathcal{L}_{21}(b_1) + \mathcal{L}_{22}(b_2) + \mathcal{L}_{23}(b_3) \\ \mathcal{L}_{31}(b_1) + \mathcal{L}_{32}(b_2) + \mathcal{L}_{33}(b_3) \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{L}_{11}(a_1) + \mathcal{L}_{12}(a_2) + \mathcal{L}_{13}(a_3) + \mathcal{L}_{11}(b_1) + \mathcal{L}_{12}(b_2) + \mathcal{L}_{13}(b_3) \\ \mathcal{L}_{21}(a_1) + \mathcal{L}_{22}(a_2) + \mathcal{L}_{23}(a_3) + \mathcal{L}_{21}(b_1) + \mathcal{L}_{22}(b_2) + \mathcal{L}_{23}(b_3) \\ \mathcal{L}_{31}(a_1) + \mathcal{L}_{32}(a_2) + \mathcal{L}_{33}(a_3) + \mathcal{L}_{31}(b_1) + \mathcal{L}_{32}(b_2) + \mathcal{L}_{33}(b_3) \end{pmatrix} \\
&= \begin{pmatrix} [\mathcal{L}_{11}(a_1) + \mathcal{L}_{11}(b_1)] + [\mathcal{L}_{12}(a_2) + \mathcal{L}_{12}(b_2)] + [\mathcal{L}_{13}(a_3) + \mathcal{L}_{13}(b_3)] \\ [\mathcal{L}_{21}(a_1) + \mathcal{L}_{21}(b_1)] + [\mathcal{L}_{22}(a_2) + \mathcal{L}_{22}(b_2)] + [\mathcal{L}_{23}(a_3) + \mathcal{L}_{23}(b_3)] \\ [\mathcal{L}_{31}(a_1) + \mathcal{L}_{31}(b_1)] + [\mathcal{L}_{32}(a_2) + \mathcal{L}_{32}(b_2)] + [\mathcal{L}_{33}(a_3) + \mathcal{L}_{33}(b_3)] \end{pmatrix}, \quad \text{by the linearity of } \mathcal{L}_{ij} \\
&= \begin{pmatrix} \mathcal{L}_{11}(a_1 + b_1) + \mathcal{L}_{12}(a_2 + b_2) + \mathcal{L}_{13}(a_2 + b_2) \\ \mathcal{L}_{21}(a_1 + b_1) + \mathcal{L}_{22}(a_2 + b_2) + \mathcal{L}_{23}(a_2 + b_2) \\ \mathcal{L}_{31}(a_1 + b_1) + \mathcal{L}_{32}(a_2 + b_2) + \mathcal{L}_{33}(a_2 + b_2) \end{pmatrix} \\
&= \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{pmatrix} \\
&= \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \left(\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right).
\end{aligned}$$

Therefore, the additivity condition is satisfied.

(b) Next, we consider the scaling condition. Let $c \in \mathbb{C}$ be given,

$$
\begin{aligned}
&\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \left(c \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \right) \\
=&\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \begin{pmatrix} ca_1 \\ ca_2 \\ ca_3 \end{pmatrix} \\
=&\begin{pmatrix} \mathcal{L}_{11}(ca_1) + \mathcal{L}_{12}(ca_2) + \mathcal{L}_{13}(ca_3) \\ \mathcal{L}_{21}(ca_1) + \mathcal{L}_{22}(ca_2) + \mathcal{L}_{23}(ca_3) \\ \mathcal{L}_{31}(ca_1) + \mathcal{L}_{32}(ca_2) + \mathcal{L}_{33}(ca_3) \end{pmatrix} \\
=&\begin{pmatrix} c\mathcal{L}_{11}(a_1) + c\mathcal{L}_{12}(a_2) + c\mathcal{L}_{13}(a_3) \\ c\mathcal{L}_{21}(a_1) + c\mathcal{L}_{22}(a_2) + c\mathcal{L}_{23}(a_3) \\ c\mathcal{L}_{31}(a_1) + c\mathcal{L}_{32}(a_2) + c\mathcal{L}_{33}(a_3) \end{pmatrix}, \quad \text{by the linearity of } \mathcal{L}_{ij} \\
=&c \begin{pmatrix} \mathcal{L}_{11}(a_1) + \mathcal{L}_{12}(a_2) + \mathcal{L}_{13}(a_3) \\ \mathcal{L}_{21}(a_1) + \mathcal{L}_{22}(a_2) + \mathcal{L}_{23}(a_3) \\ \mathcal{L}_{31}(a_1) + \mathcal{L}_{32}(a_2) + \mathcal{L}_{33}(a_3) \end{pmatrix}.
\end{aligned}
$$

Therefore, the scalar multiplication condition is also satisfied.

From (a) and (b), $\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times}$ is a linear function.

2. (a) <u>Existence:</u> Let Π_i be a coordinate projection on $\mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$ onto $\mathbf{V}_i$ defined in Definition 0.2. Let γ_i be a coordinate injection of $\mathbf{V}_i$ into $\mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$ also defined in Definition 0.2.

Consider the block-matrix $\mathcal{L}_\times$,

$$\mathcal{L}_\times = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times .$$

Let the linear function $\mathcal{T} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ be given. Let each $\mathcal{L}_{ij} : \mathbf{V}_j \overset{\text{linear}}{\longrightarrow} \mathbf{V}_i$ be defined as

$$\mathcal{L}_{ij} = \Pi_i \circ \mathcal{T} \circ \gamma_j \Big|_{\mathbf{V}_j} .$$

Then we have

$$\begin{aligned} \mathcal{L}_\times &= \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \\ &= \begin{bmatrix} \Pi_1 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} & \Pi_1 \circ \mathcal{T} \circ \gamma_2 \Big|_{\mathbf{V}_2} & \Pi_1 \circ \mathcal{T} \circ \gamma_3 \Big|_{\mathbf{V}_3} \\ \Pi_2 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} & \Pi_2 \circ \mathcal{T} \circ \gamma_2 \Big|_{\mathbf{V}_2} & \Pi_2 \circ \mathcal{T} \circ \gamma_3 \Big|_{\mathbf{V}_3} \\ \Pi_3 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} & \Pi_3 \circ \mathcal{T} \circ \gamma_2 \Big|_{\mathbf{V}_2} & \Pi_3 \circ \mathcal{T} \circ \gamma_3 \Big|_{\mathbf{V}_3} \end{bmatrix}_\times . \end{aligned}$$

Consider $\begin{pmatrix} w_1 & \mathbf{0}_\mathbf{V} & \mathbf{0}_\mathbf{V} \end{pmatrix}^\top \in \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$ with $w_1 \in \mathbf{V}_1$.

$$\begin{aligned} &\begin{bmatrix} \Pi_1 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} & \Pi_1 \circ \mathcal{T} \circ \gamma_2 \Big|_{\mathbf{V}_2} & \Pi_3 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_3} \\ \Pi_2 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} & \Pi_2 \circ \mathcal{T} \circ \gamma_2 \Big|_{\mathbf{V}_2} & \Pi_3 \circ \mathcal{T} \circ \gamma_2 \Big|_{\mathbf{V}_3} \\ \Pi_3 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} & \Pi_3 \circ \mathcal{T} \circ \gamma_2 \Big|_{\mathbf{V}_2} & \Pi_3 \circ \mathcal{T} \circ \gamma_3 \Big|_{\mathbf{V}_3} \end{bmatrix}_\times \begin{pmatrix} w_1 \\ \mathbf{0}_\mathbf{V} \\ \mathbf{0}_\mathbf{V} \end{pmatrix} \\ &= \begin{bmatrix} \Pi_1 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} (w_1) \\ \Pi_2 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} (w_1) \\ \Pi_3 \circ \mathcal{T} \circ \gamma_1 \Big|_{\mathbf{V}_1} (w_1) \end{bmatrix} \\ &= \begin{bmatrix} \Pi_1 \circ \mathcal{T} \begin{pmatrix} w_1 & \mathbf{0}_\mathbf{V} & \mathbf{0}_\mathbf{V} \end{pmatrix}^\top \\ \Pi_2 \circ \mathcal{T} \begin{pmatrix} w_1 & \mathbf{0}_\mathbf{V} & \mathbf{0}_\mathbf{V} \end{pmatrix}^\top \\ \Pi_3 \circ \mathcal{T} \begin{pmatrix} w_1 & \mathbf{0}_\mathbf{V} & \mathbf{0}_\mathbf{V} \end{pmatrix}^\top \end{bmatrix} . \end{aligned}$$

Let $\mathcal{T}\left(\begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^\top\right)$ be $\begin{pmatrix} a & b & c \end{pmatrix}^\top \in \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$. By definition,

$$\begin{aligned}
&\begin{bmatrix} \Pi_1 \circ \mathcal{T}\begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^\top \\ \Pi_2 \circ \mathcal{T}\begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^\top \\ \Pi_3 \circ \mathcal{T}\begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^\top \end{bmatrix} \\
&= \begin{bmatrix} \Pi_1\left(\begin{pmatrix} a & b & c \end{pmatrix}^\top\right) \\ \mathbf{0_V} \\ \mathbf{0_V} \end{bmatrix} + \begin{bmatrix} \mathbf{0_V} \\ \Pi_2\left(\begin{pmatrix} a & b & c \end{pmatrix}^\top\right) \\ \mathbf{0_V} \end{bmatrix} + \begin{bmatrix} \mathbf{0_V} \\ \mathbf{0_V} \\ \Pi_3\left(\begin{pmatrix} a & b & c \end{pmatrix}^\top\right) \end{bmatrix} \\
&= \begin{bmatrix} a \\ \mathbf{0_V} \\ \mathbf{0_V} \end{bmatrix} + \begin{bmatrix} \mathbf{0_V} \\ b \\ \mathbf{0_V} \end{bmatrix} + \begin{bmatrix} \mathbf{0_V} \\ \mathbf{0_V} \\ c \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \mathcal{T}\begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}
\end{aligned}$$

Therefore,

$$\mathcal{L}_\times \begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix} = \mathcal{T}\begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}.$$

In general, the equality above also holds for any $v = w_1 + w_2 + w_3 \in \mathbf{V}$ with $w_i \in \mathbf{V}_i$, since v can be expressed as

$$v = \begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix} + \begin{pmatrix} \mathbf{0_V} \\ w_2 \\ \mathbf{0_V} \end{pmatrix} + \begin{pmatrix} \mathbf{0_V} \\ \mathbf{0_V} \\ w_3 \end{pmatrix},$$

and the linearity of $\mathcal{L}_\times$ and $\mathcal{T}$ ensures that $\mathcal{L}_\times(v) = \mathcal{T}(v)$. Hence, $\mathcal{T}$ can be expressed as the matrix $\mathcal{L}_\times$ whose elements $\mathcal{L}_{ij}$'s are defined as

$$\mathcal{L}_{ij} = \Pi_i \circ \mathcal{T} \circ \gamma_j \Big|_{\mathbf{V}_j}.$$

.

(b) Uniqueness: We observe that all of the steps in the "Existence part" are reversible. This means if we were given the linear map $\mathcal{T}$, we could consider applying $\mathcal{T}$ to $\begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^\top$, $\begin{pmatrix} \mathbf{0_V} & w_2 & \mathbf{0_V} \end{pmatrix}^\top$, and $\begin{pmatrix} \mathbf{0_V} & \mathbf{0_V} & w_3 \end{pmatrix}^\top \in \mathbf{V}$; and by reversing the steps in part (a), we know that $\mathcal{T}$ can be expressed as a 3×3 block-matrix whose elements must have the form $\Pi_i \circ \mathcal{T} \circ \gamma_j \Big|_{\mathbf{V}_j} = \mathcal{L}_{ij}$.

From parts (a) and (b), the linear function $\mathcal{T}$ can be expressed as the matrix $\mathcal{L}_\times$ in exactly one way.

3. Let

$$\mathcal{L} := \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$$

and

$$\mathcal{M} := \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$$

be given.

(a) Let $\mathcal{L}$ be the matrix of the linear function $\mathcal{T}$ above and $\mathcal{M}$ be the matrix of a linear function $\mathcal{K} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$. Consider the expression $2\mathcal{L}_{ij} + 3\mathcal{M}_{ij}$. Let $\mathcal{L}_{ij}$ be constructed as in Part 1 for the linear function $\mathcal{T}$. Let $\mathcal{M}$ be constructed similarly for $\mathcal{K}$. Then, we have

$$\begin{aligned} 2\mathcal{L}_{ij} + 3\mathcal{M}_{ij} &= 2\Pi_i \circ \mathcal{T} \circ \gamma_j \Big|_{\mathbf{W}_j} + 3\Pi_i \circ \mathcal{K} \circ \gamma_j \Big|_{\mathbf{W}_j} \\ &= \Pi_i \circ 2\mathcal{T} \circ \gamma_j \Big|_{\mathbf{W}_j} + \Pi_i \circ 3\mathcal{K} \circ \gamma_j \Big|_{\mathbf{W}_j} \\ &= \Pi_i \circ \left(2\mathcal{T} \circ \gamma_j \Big|_{\mathbf{W}_j} + 3\mathcal{K} \circ \gamma_j \Big|_{\mathbf{W}_j} \right) \\ &= \Pi_i \circ (2\mathcal{T} + 3\mathcal{K}) \circ \gamma_j \Big|_{\mathbf{W}_j} \\ &= (2\mathcal{T} + 3\mathcal{K})_{ij}, \ \text{ by definition.} \end{aligned}$$

This shows that each element of the matrix of the linear function $(2\mathcal{L} + 3\mathcal{K})$ is given by

$$\boxed{(2\mathcal{T} + 3\mathcal{K})_{ij} = 2\mathcal{L}_{ij} + 3\mathcal{M}_{ij}}$$

i.e.,

$$\begin{aligned} &2 \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} + 3 \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_{\times} \\ &= \begin{bmatrix} 2\mathcal{L}_{11} + 3\mathcal{M}_{11} & 2\mathcal{L}_{12} + 3\mathcal{M}_{12} & 2\mathcal{L}_{13} + 3\mathcal{M}_{13} \\ 2\mathcal{L}_{21} + 3\mathcal{M}_{21} & 2\mathcal{L}_{22} + 3\mathcal{M}_{22} & 2\mathcal{L}_{23} + 3\mathcal{M}_{23} \\ 2\mathcal{L}_{31} + 3\mathcal{M}_{31} & 2\mathcal{L}_{32} + 3\mathcal{M}_{32} & 2\mathcal{L}_{33} + 3\mathcal{M}_{33} \end{bmatrix}_{\times}. \end{aligned}$$

(b) Let $\mathcal{T}$, $\mathcal{M}_{ij}$, $\mathcal{K}$, and $\mathcal{L}_{ij}$ be defined similarly: $\mathcal{L}$ is the matrix of the linear function $\mathcal{T}$ and $\mathcal{M}$ is the matrix of the linear function $\mathcal{K}$. Consider the expression $\sum_{k=1}^{3} \mathcal{L}_{ik} \circ \mathcal{M}_{kj}$, we have

$$\begin{aligned}
\sum_{k=1}^{3} \mathcal{L}_{ik} \circ \mathcal{M}_{kj} &= \sum_{k=1}^{3} \Pi_i \circ \mathcal{T} \circ \gamma_k\Big|_{\mathbf{W}_k} \circ \Pi_k \circ \mathcal{K} \circ \gamma_j\Big|_{\mathbf{W}_j} \\
&= \sum_{k=1}^{3} \Pi_i \circ \mathcal{T} \circ \left(\gamma_k\Big|_{\mathbf{W}_k} \circ \Pi_k\right) \circ \mathcal{K} \circ \gamma_j\Big|_{\mathbf{W}_j} \\
&= \Pi_i \circ \mathcal{T} \circ \left(\sum_{k=1}^{3} \gamma_k\Big|_{\mathbf{W}_k} \circ \Pi_k\right) \circ \mathcal{K} \circ \gamma_j\Big|_{\mathbf{W}_j}.
\end{aligned}$$

Consider $v = \begin{pmatrix} w_1 & w_2 & w_2 \end{pmatrix}^\top \in \mathbf{V}$ such that $w_i \in \mathbf{W}_i$ for $i \in \{1, 2, 3\}$. We can show that the summation in the expression above is simply the identity function:

$$\begin{aligned}
\left(\sum_{k=1}^{3} \gamma_k\Big|_{\mathbf{W}_k} \circ \Pi_k\right)(v) &= \gamma_1\Big|_{\mathbf{W}_1} \circ \Pi_1(v) + \gamma_2\Big|_{\mathbf{W}_2} \circ \Pi_2(v) + \gamma_3\Big|_{\mathbf{W}_3} \circ \Pi_3(v) \\
&= \gamma_1(w_1) + \gamma_2(w_2) + \gamma_3(w_3) \\
&= \begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix} + \begin{pmatrix} \mathbf{0_V} \\ w_2 \\ \mathbf{0_V} \end{pmatrix} + \begin{pmatrix} \mathbf{0_V} \\ \mathbf{0_V} \\ w_3 \end{pmatrix} \\
&= \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \\
&= v.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{k=1}^{3} \mathcal{L}_{ik} \circ \mathcal{M}_{kj} &= \Pi_i \circ \mathcal{T} \circ \mathcal{K} \circ \gamma_j\Big|_{\mathbf{W}_j} \\
&= \Pi_i \circ (\mathcal{T} \circ \mathcal{K}) \circ \gamma_j\Big|_{\mathbf{W}_j} \\
&= (\mathcal{T} \circ \mathcal{K})_{ij}.
\end{aligned}$$

This shows that each element of the matrix of the linear function $(\mathcal{T} \circ \mathcal{K})$ is given by

$$\boxed{(\mathcal{T} \circ \mathcal{K})_{ij} = \sum_{k=1}^{3} \mathcal{L}_{ik} \circ \mathcal{M}_{kj}}$$

which gives the full block-matrix form:

$$
\begin{aligned}
&\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \circ \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \\
&= \begin{bmatrix} \sum_{k=1}^{3} \mathcal{L}_{1k} \circ \mathcal{M}_{k1} & \sum_{k=1}^{3} \mathcal{L}_{1k} \circ \mathcal{M}_{k2} & \sum_{k=1}^{3} \mathcal{L}_{1k} \circ \mathcal{M}_{k3} \\ \sum_{k=1}^{3} \mathcal{L}_{2k} \circ \mathcal{M}_{k1} & \sum_{k=1}^{3} \mathcal{L}_{2k} \circ \mathcal{M}_{k2} & \sum_{k=1}^{3} \mathcal{L}_{2k} \circ \mathcal{M}_{k3} \\ \sum_{k=1}^{3} \mathcal{L}_{3k} \circ \mathcal{M}_{k1} & \sum_{k=1}^{3} \mathcal{L}_{3k} \circ \mathcal{M}_{k2} & \sum_{k=1}^{3} \mathcal{L}_{3k} \circ \mathcal{M}_{k3} \end{bmatrix}_{\times}
\end{aligned}
$$

Problem. 2. Suppose that $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3$ and the $\mathbf{W}_i$'s are non-trivial subspaces of $\mathbf{V}$. Suppose that for each $i, j \in \{1, 2, 3\}$,

$$\mathcal{L}_{ij} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i.$$

1. Verify for formula

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{\maltese} = \left(\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right)_{\maltese},$$

   where we have used as few brackets as possible without losing clarity.

2. Argue that every $T : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ can be expressed as $\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus}$,

   with $\mathcal{L}_{ij} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i$, in exactly one way.

3. If the function

$$\begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_{\oplus} : \mathbf{V} \longrightarrow \mathbf{V}$$

   is defined similarly, find (with proof) the corresponding black-matrix form of the functions

$$2 \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} + 3 \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_{\oplus}$$

   and

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} \circ \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_{\oplus}.$$

4. Suppose that $\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 = \mathcal{I}_{\mathbf{V}}$ is a resolution of the identity, where $\mathrm{Im}(\mathcal{E}_i) = \mathbf{W}_i$. If

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus},$$

   express (with proof, of course) $\mathcal{L}_{ij}$ in terms of $\mathcal{L}, \mathcal{E}_k$'s and $\mathbf{W}_k$'s.

5. Continuing with the set-up of part 4, find $\mathcal{M}_{ij} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i$ such that

$$\mathcal{E}_1 = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_{\oplus}.$$

   Then do the same for $\mathcal{E}_2$ and $\mathcal{E}_3$.

Solution. 2.

1. By definition,

$$
\begin{aligned}
\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{\maltese}
&= \left[\maltese \circ \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \circ \maltese^{-1} \right] \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{\maltese} \\
&= \left[\maltese \circ \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \circ \maltese^{-1} \circ \maltese \right] \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
&= \maltese \circ \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
&= \maltese \left(\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right) \\
&= \left(\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right)_{\maltese} .
\end{aligned}
$$

2. (a) Existence: Let the linear functions $\mathcal{E}_i := \Pi_i \circ \maltese^{-1} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i$ and $\mathcal{K}_i := \maltese \circ \gamma_i : \mathbf{W}_i \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ be given, where the Π_i's and γ_i's are defined the same way as in Problem 1.

Consider the block matrix $\mathcal{L}_{\oplus}$

$$
\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} .
$$

Let the linear function $\mathcal{T} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ be given. Let each $\mathcal{L}_{ij} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i$ be defined as:

$$
\mathcal{L}_{ij} = \mathcal{E}_i \circ \mathcal{T} \circ \mathcal{K}_j \Big|_{\mathbf{W}_j} .
$$

By definition,

$$
\begin{aligned}
\mathcal{L}_{ij} &= \mathcal{E}_i \circ \mathcal{T} \circ \mathcal{K}_j \Big|_{\mathbf{W}_j} \\
&= \Pi_i \circ \maltese^{-1} \circ \mathcal{T} \circ \maltese \circ \gamma_j \Big|_{\mathbf{W}_j} .
\end{aligned}
$$

Applying $\mathcal{L}_\oplus$ to $\begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}_{\maltese}^{\top} \in \mathbf{V}$, we get

$$\begin{aligned}
\mathcal{L}\begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}_{\maltese} &= \begin{bmatrix} \Pi_1 \circ \maltese^{-1} \circ \mathcal{T} \circ \maltese \circ \gamma_1\Big|_{\mathbf{W}_1} (w_1) \\ \Pi_2 \circ \maltese^{-1} \circ \mathcal{T} \circ \maltese \circ \gamma_1\Big|_{\mathbf{W}_1} (w_1) \\ \Pi_3 \circ \maltese^{-1} \circ \mathcal{T} \circ \maltese \circ \gamma_1\Big|_{\mathbf{W}_1} (w_1) \end{bmatrix}_{\maltese} \\
&= \begin{bmatrix} \Pi_1 \circ \maltese^{-1} \circ \mathcal{T} \circ \maltese \begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^{\top} \\ \Pi_2 \circ \maltese^{-1} \circ \mathcal{T} \circ \maltese \begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^{\top} \\ \Pi_3 \circ \maltese^{-1} \circ \mathcal{T} \circ \maltese \begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}^{\top} \end{bmatrix}_{\maltese} \\
&= \begin{bmatrix} \Pi_1 \circ \maltese^{-1} \circ \mathcal{T}(w_1) \\ \Pi_2 \circ \maltese^{-1} \circ \mathcal{T}(w_1) \\ \Pi_3 \circ \maltese^{-1} \circ \mathcal{T}(w_1) \end{bmatrix}_{\maltese} \\
&= \maltese\left(\begin{bmatrix} \Pi_1 \circ \maltese^{-1} \circ \mathcal{T}(w_1) \\ \Pi_2 \circ \maltese^{-1} \circ \mathcal{T}(w_1) \\ \Pi_3 \circ \maltese^{-1} \circ \mathcal{T}(w_1) \end{bmatrix}\right) \\
&= \maltese \circ \maltese^{-1} \circ \mathcal{T}(w_1), \quad \text{by Part 2(a) of Problem 1} \\
&= \mathcal{T}(w_1)
\end{aligned}$$

By the definition of $\maltese$, we can write this as

$$\mathcal{L}(w_1 + \mathbf{0_V} + \mathbf{0_V}) = \mathcal{L}(w_1) = \mathcal{T}(w_1).$$

In general, the equality above also holds if we start with any $v = w_1 + w_2 + w_3 \in \mathbf{V}$ with $w_i \in \mathbf{W}_i$ for $i = \{1, 2, 3\}$, because as we have shown in Part 2(a) of Problem 1, v can be written as

$$v = \begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}_{\maltese} + \begin{pmatrix} \mathbf{0_V} \\ w_2 \\ \mathbf{0_V} \end{pmatrix}_{\maltese} + \begin{pmatrix} \mathbf{0_V} \\ \mathbf{0_V} \\ w_3 \end{pmatrix}_{\maltese},$$

and the linearity of $\mathcal{L}$ and $\mathcal{T}$ ensures $\mathcal{L}(v) = \mathcal{T}(v)$. Hence, $\mathcal{T}$ can expressed as the matrix $\mathcal{L}$ whose elements are defined as

$$\mathcal{L}_{ij} = \mathcal{E}_i \circ \mathcal{T} \circ \mathcal{K}_j\Big|_{\mathbf{W}_j}.$$

(b) Uniqueness: Once again, we observe that all of the steps in the "Existence" part are reversible. We can consider applying $\mathcal{T}$ to v, expressed in Part 3(a), and by reversing the steps in Part (a) show that the 3×3 block-matrix of $\mathcal{T}$ must consist elements defined by $\mathcal{E}_i \circ \mathcal{T} \circ \mathcal{K}_j\Big|_{\mathbf{W}_j} = \mathcal{L}_{\oplus ij}$ for $i, j \in \{1, 2, 3\}$.

From parts (a) and (b), the linear function $\mathcal{T}$ can be expressed as the matrix $\mathcal{L}_\oplus$ defined in the problem statement in exactly one way.

3. (a) Let $\mathcal{L}_\oplus = [\mathcal{L}_{\oplus_{ij}}]$ denote the matrix $\mathcal{L}_\oplus : \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3 \xrightarrow{\text{linear}} \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3$. Let $\mathcal{M}_\oplus = [\mathcal{M}_{\oplus_{ij}}]$ be defined similarly for $\mathcal{M}$. By part 1, we can re-express the matrix

$$[(2\mathcal{L}_\oplus + 3\mathcal{M}_\oplus)_{ij}] = 2\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\oplus + 3\begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\oplus$$

as

$$\begin{aligned}
&2\maltese \circ \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \circ \maltese^{-1} + 3\maltese \circ \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\times \circ \maltese^{-1} \\
= {}& \maltese \circ 2\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \circ \maltese^{-1} + \maltese \circ 3\begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\times \circ \maltese^{-1} \\
= {}& \maltese \circ \left(2\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times \circ \maltese^{-1} + 3\begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\times \circ \maltese^{-1} \right) \\
= {}& \maltese \circ \left(2\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\times + 3\begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\times \right) \circ \maltese^{-1} \\
= {}& \maltese \circ \left(2[\mathcal{L}_{\times_{ij}}] + 3[\mathcal{M}_{\times_{ij}}] \right) \circ \maltese^{-1},
\end{aligned}$$

where the notation $[A_{ij}]$ denotes the block-matrix of elements A_{ij} where A is a linear function, and the last relation:

$$[(2\mathcal{L}_\times + 3\mathcal{M}_\times)_{ij}] = 2[\mathcal{L}_{\times_{ij}}] + 3[\mathcal{M}_{\times_{ij}}].$$

follows from Problem 1. Then, we have

$$\begin{aligned}
[(2\mathcal{L}_\oplus + 3\mathcal{M}_\oplus)_{ij}] &= \maltese \circ \left(2[\mathcal{L}_{\times_{ij}}] + 3[\mathcal{M}_{\times_{ij}}] \right) \circ \maltese^{-1} \\
&= \maltese \circ [2\mathcal{L}_{\times_{ij}}] \circ \maltese^{-1} + \maltese \circ [3\mathcal{M}_{\times_{ij}}] \circ \maltese^{-1} \\
&= 2[\mathcal{L}_{\oplus_{ij}}] + 3[\mathcal{M}_{\oplus_{ij}}].
\end{aligned}$$

So, in the block-matrix form:

$$\begin{aligned}
&2\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_\oplus + 3\begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_\oplus \\
= {}& \begin{bmatrix} 2\mathcal{L}_{11} + 3\mathcal{M}_{11} & 2\mathcal{L}_{12} + 3\mathcal{M}_{12} & 2\mathcal{L}_{13} + 3\mathcal{M}_{13} \\ 2\mathcal{L}_{21} + 3\mathcal{M}_{21} & 2\mathcal{L}_{22} + 3\mathcal{M}_{22} & 2\mathcal{L}_{23} + 3\mathcal{M}_{23} \\ 2\mathcal{L}_{31} + 3\mathcal{M}_{31} & 2\mathcal{L}_{32} + 3\mathcal{M}_{32} & 2\mathcal{L}_{33} + 3\mathcal{M}_{33} \end{bmatrix}_\oplus .
\end{aligned}$$

(b) Let $\mathcal{L}_\oplus$ and $\mathcal{M}_\oplus$ denote the same matrices in part (a), and let $\mathcal{L}_\oplus$ represent the linear function $\mathcal{T}$, $\mathcal{M}_\oplus$ represent the linear function $\mathcal{N}$. Let the matrix $[(\mathcal{L}_\oplus \circ \mathcal{M}_\oplus)_{ij}]$ express the composition $\mathcal{T} \circ \mathcal{N}$. Its the elements $[\mathcal{L}_\oplus \circ \mathcal{M}_\oplus]$ can be constructed as

$$(\mathcal{L}_\oplus \circ \mathcal{M}_\oplus)_{ij} = \mathcal{E}_i \circ (\mathcal{L}_\oplus \circ \mathcal{M}_\oplus) \circ \mathcal{K}_j \Big|_{\mathbf{W}_j} = \mathcal{E}_i \circ (\mathcal{T} \circ \mathcal{N}) \circ \mathcal{K}_j \Big|_{\mathbf{W}_j}, \quad (\dagger)$$

as we have done in Part 1. Consider the expression $\sum_{k=1}^{3} \mathcal{L}_{\oplus_{ik}} \circ \mathcal{M}_{\oplus_{kj}}$. By part 1, we have

$$\mathcal{L}_{\oplus_{ik}} = \mathcal{E}_i \circ \mathcal{T} \circ \mathcal{K}_k \Big|_{\mathbf{W}_k} \text{ and } \mathcal{M}_{\oplus_{kj}} = \mathcal{E}_k \circ \mathcal{N} \circ \mathcal{K}_j \Big|_{\mathbf{W}_j}.$$

This gives

$$\begin{aligned} \sum_{k=1}^{3} \mathcal{L}_{\oplus_{ik}} \circ \mathcal{M}_{\oplus_{kj}} &= \sum_{k=1}^{3} \mathcal{E}_i \circ \mathcal{T} \circ \left(\mathcal{K}_k \Big|_{\mathbf{W}_k} \circ \mathcal{E}_k \right) \circ \mathcal{N} \circ \mathcal{K}_j \Big|_{\mathbf{W}_j} \\ &= \mathcal{E}_i \circ \mathcal{T} \circ \left(\sum_{k=1}^{3} \mathcal{K}_k \Big|_{\mathbf{W}_k} \circ \mathcal{E}_k \right) \circ \mathcal{N} \circ \mathcal{K}_j \Big|_{\mathbf{W}_j} \end{aligned}$$

By definition, we have

$$\mathcal{K}_k \Big|_{\mathbf{W}_k} = \maltese \circ \gamma_k \Big|_{\mathbf{W}_k} \text{ and } \mathcal{E}_k = \Pi_k \circ \maltese^{-1},$$

which imply

$$\begin{aligned} \sum_{k=1}^{3} \mathcal{K}_k \Big|_{\mathbf{W}_k} \circ \mathcal{E}_k &= \sum_{k=1}^{3} \maltese \circ \gamma_k \Big|_{\mathbf{W}_k} \circ \Pi_k \circ \maltese^{-1} \\ &= \maltese \circ \left(\sum_{k=1}^{3} \gamma_k \Big|_{\mathbf{W}_k} \circ \Pi_k \right) \maltese^{-1} \\ &= \maltese \circ \maltese^{-1}, \text{ by Part 3(b) of Problem 1} \\ &= \text{Identity}. \end{aligned}$$

It follows that,

$$\sum_{k=1}^{3} \mathcal{L}_{\oplus_{ik}} \circ \mathcal{M}_{\oplus_{kj}} = \mathcal{E}_i \circ \mathcal{T} \circ \mathcal{N} \circ \mathcal{K}_j \Big|_{\mathbf{W}_j} = (\mathcal{L}_\oplus \circ \mathcal{M}_\oplus)_{ij}, \text{ by } (\dagger).$$

Therefore, the matrix elements of $[\mathcal{L}_\oplus \circ \mathcal{M}_\oplus]$ are given by

$$\boxed{(\mathcal{L}_\oplus \circ \mathcal{M}_\oplus)_{ij} = \sum_{k=1}^{3} \mathcal{L}_{\oplus_{ik}} \circ \mathcal{M}_{\oplus_{kj}}}$$

Since the full block-matrix expression for $[\mathcal{L}_\oplus \circ \mathcal{M}_\oplus]$ is very similar to what we have found with $\mathcal{L}_\times \circ \mathcal{M}_\times$, we will not produce it here.

4. Let $\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 = \mathcal{I}_{\mathbf{V}}$ be a resolution of identity, with $\mathrm{Im}(\mathcal{E}_i) = \mathbf{W}_i$, be given. Furthermore, let the linear map $\mathcal{L} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ be given where

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus}.$$

<u>Claim</u>:

$$\boxed{\mathcal{L}_{ij} = \mathcal{E}_i \circ \mathcal{L}\Big|_{\mathbf{W}_j} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i}$$

Consider applying $\mathcal{L}$ to $\begin{pmatrix} w_1 & \mathbf{0_V} & \mathbf{0_V} \end{pmatrix}_{✠}^{\top} \in \mathbf{V}$, we have

$$\mathcal{L}\begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}_{✠} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} \begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}_{✠} = \begin{pmatrix} \mathcal{L}_{11}(w_1) \\ \mathcal{L}_{21}(w_1) \\ \mathcal{L}_{31}(w_1) \end{pmatrix}_{✠}.$$

To extract an element $\mathcal{L}_{i1}$, we can apply $\mathcal{E}_i$ to this result:

$$\mathcal{E}_i \circ \mathcal{L}\begin{pmatrix} w_1 \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}_{✠} = \mathcal{E}_i \begin{pmatrix} \mathcal{L}_{11}(w_1) \\ \mathcal{L}_{21}(w_1) \\ \mathcal{L}_{31}(w_1) \end{pmatrix}_{✠} = \mathcal{L}_{i1}(w_1).$$

This shows that by considering the restricted map $\mathcal{L}\Big|_{\mathbf{W}_1}$, we are able to obtain the $\mathcal{L}_{i1}$ element of $\mathcal{L}$ by composing the idempotent $\mathcal{E}_i$ with $\mathcal{L}\Big|_{\mathbf{W}_1}$. So in general, if we consider the restricted map $\mathcal{L}\Big|_{\mathbf{W}_j}$, the same procedure will give us the element $\mathcal{L}_{ij}$. Hence, we have verified the claim that

$$\mathcal{L}_{ij} = \mathcal{E}_i \circ \mathcal{L}\Big|_{\mathbf{W}_j}.$$

5. Let

$$\mathcal{E}_1 = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} \end{bmatrix}_{\oplus}.$$

be given, where $\mathrm{Im}(\mathcal{E}_i) = \mathbf{W}_i$ and $\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 = \mathcal{I}_{\mathbf{V}}$ is a resolution of identity. Continuing with the set-up of part 4, we have

$$\mathcal{M}_{ij} = \mathcal{E}_i \circ \mathcal{E}_1\Big|_{\mathbf{W}_j}.$$

Since $\mathcal{E}_1+\mathcal{E}_2+\mathcal{E}_3 = \mathcal{I}_{\mathbf{V}}$ is a resolution of identity, and $\mathrm{Im}(\mathcal{E}_i) = \mathbf{W}_i$ where $\mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3 = \mathbf{V}$, $\ker(\mathcal{E}_1) = \mathbf{W}_2 \oplus \mathbf{W}_3$. This implies the restricted map $\mathcal{E}_1\Big|_{\mathbf{W}_j}$ gives $\mathbf{0}_{\mathbf{V}}$ for $j \neq 1$ and acts as an identity on $\mathbf{W}_j$ if $j = 1$. Therefore, for $j \neq 1$, M_{ij} is a zero function.

If $j = 1$, then

$$\mathcal{M}_{ij} = \mathcal{E}_i \circ \mathcal{E}_1\Big|_{\mathbf{W}_1} = \mathcal{E}_i\Big|_{\mathbf{W}_1},$$

since $\mathcal{E}_1$ acts as identity on $\mathbf{W}_1$. The same argument for the restricted map $\mathcal{E}_i\Big|_{\mathbf{W}_1}$ shows that $\mathcal{M}_{ij}$ is the zero function if $i \neq 1$ and the identity on $\mathbf{W}_1$ if $i = 1$. Therefore,

$$\text{For } \mathcal{E}_1 : \qquad \mathcal{M}_{ij} = \begin{cases} Id_{\mathbf{W}_1}, & i,j = 1 \\ \mathcal{O}, & i \text{ or } j \neq 1 \end{cases}$$

The exact same procedure can be used to find the elements M_{ij} of $\mathcal{E}_2$ and $\mathcal{E}_3$, simply replacing all 1's with 2's and 3's. So,

$$\text{For } \mathcal{E}_2 : \qquad \mathcal{M}_{ij} = \begin{cases} Id_{\mathbf{W}_2}, & i,j = 2 \\ \mathcal{O}, & i \text{ or } j \neq 3 \end{cases}$$

$$\text{For } \mathcal{E}_3 : \qquad \mathcal{M}_{ij} = \begin{cases} Id_{\mathbf{W}_3}, & i,j = 3 \\ \mathcal{O}, & i \text{ or } j \neq 3 \end{cases}$$

Problem. 3.

1. Using the set-up of Definition 0.7, verify the identity
$$[\mathcal{T}(\mathcal{X})]_\Omega = [\mathcal{T}]_{\Omega\leftarrow\Gamma}[\mathcal{X}]_\Gamma.$$

2. Use part 1 to prove that
$$[\mathcal{T}]_{\Omega\leftarrow\Gamma} = \begin{bmatrix}[\mathcal{T}(v_1)]_\Omega & [\mathcal{T}(v_2)]_\Omega & \dots & [\mathcal{T}(v_n)]_\Omega\end{bmatrix}.$$

3. If $\mathbf{U}$ is a finite-dimensional vector space with a coordinate system Δ, and $\mathcal{S} : \mathbf{Z} \xrightarrow{\text{linear}} \mathbf{U}$, argue that
$$[\mathcal{S} \circ \mathcal{T}]_{\Delta\leftarrow\Gamma} = [\mathcal{S}]_{\Delta\leftarrow\Omega}[\mathcal{T}]_{\Omega\leftarrow\Gamma}.$$

4. Verify that in the case $\mathcal{T}$ is invertible (in which case $m = n$), so is $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$, and
$$\left[\mathcal{T}^{-1}\right]_{\Gamma\leftarrow\Omega} = [\mathcal{T}]^{-1}_{\Omega\leftarrow\Gamma}.$$

Solution. 3.

1. By the set-up in Definition 0.3, we have $[\mathcal{T}]_{\Omega\leftarrow\Gamma} = [\ ]_\Omega \circ \mathcal{T} \circ \mathcal{A}_\Gamma$, and $[\mathcal{X}]_\Gamma = [\ ]_\Gamma(\mathcal{X})$. Hence,
$$\begin{aligned}[\mathcal{T}]_{\Omega\leftarrow\Gamma}[\mathcal{X}]_\Gamma &= [\ ]_\Omega \circ \mathcal{T} \circ \mathcal{A}_\Gamma \circ [\mathcal{X}]_\Gamma \\ &= [\ ]_\Omega \circ \mathcal{T} \circ \mathcal{A}_\Gamma \circ [\ ]_\Gamma(\mathcal{X}) \\ &= [\ ]_\Omega \circ \mathcal{T} \circ (\mathcal{A}_\Gamma \circ [\ ]_\Gamma)(\mathcal{X}) \\ &= [\ ]_\Omega \circ \mathcal{T}(\mathcal{X}) \\ &= [\ ]_\Omega(\mathcal{T}(\mathcal{X})) \\ &= [\mathcal{T}(\mathcal{X})]_\Omega.\end{aligned}$$

2. <u>Claim</u>:
$$[\mathcal{T}]_{\Omega\leftarrow\Gamma} = \begin{bmatrix}[\mathcal{T}(v_1)]_\Omega & [\mathcal{T}(v_2)]_\Omega & \dots & [\mathcal{T}(v_n)]_\Omega\end{bmatrix}.$$

   Let $\mathcal{X}$ be $v_i \in \{v_1, v_2, \dots, v_n\}$ a coordinate system of Γ. Then we can create a bijective atrix
$$A_\Gamma = \begin{bmatrix}v_1 & v_2 & \dots & v_n\end{bmatrix}$$
   whose inverse is $[\ ]_\Gamma$. To find the columns of the matrix $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$, we want to apply $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$ to the standard basis vectors. To generate these standard basis vectors, we simply coordinatize $\mathcal{X}$:
$$[\mathcal{X}]_\Gamma = [v_i]_\Gamma = [\ ]_\Gamma(v_i) = A_\Gamma^{-1}(v_i) = \begin{bmatrix}v_1 & v_2 & \dots & v_n\end{bmatrix}^{-1}(v_i) = \vec{e}_i \in \mathbb{C}^n.$$

Thus, the i^{th} column of $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$ is given by

$$[\mathcal{T}]_{\Omega\leftarrow\Gamma}(\vec{e}_i) = [\mathcal{T}]_{\Omega\leftarrow\Gamma}[v_i]_\Gamma = [\mathcal{T}(v_i)]_\Gamma, \text{ by Part 1.}$$

This verifies the claim.

3. Let the finite-dimensional vector space $\mathbf{U}$ with coordinate system Δ be given. Let $\mathcal{S} : \mathbf{Z} \overset{\text{linear}}{\longrightarrow} \mathbf{U}$ be given. <u>Claim</u>:

$$[\mathcal{S} \circ \mathcal{T}]_{\Delta\leftarrow\Gamma} = [\mathcal{S}]_{\Delta\leftarrow\Omega}[\mathcal{T}]_{\Omega\leftarrow\Gamma}.$$

Let $\Delta = (w_1, w_2, \ldots, w_l)$ be a coordinate system for $\mathbf{U}$. We also define the bijective atrix

$$A_\Delta := \begin{bmatrix} w_1 & w_2 & \ldots & w_l \end{bmatrix} : \mathbb{C}^l \to \mathbf{U}.$$

Since $\mathcal{A}_\Delta$ s invertible, we can write $[\ ]_\Delta$ for $\mathcal{A}_\Delta^{-1}$.

By the set-up in definition 0.7, we have the following

$$\begin{aligned} [\mathcal{T}]_{\Omega\leftarrow\Gamma} &= [\ ]_\Omega \circ \mathcal{T} \circ \mathcal{A}_\Gamma \\ [\mathcal{S}]_{\Delta\leftarrow\Omega} &= [\ ]_\Delta \circ \mathcal{S} \circ \mathcal{A}_\Omega. \end{aligned}$$

It follows that

$$\begin{aligned} [\mathcal{S}]_{\Delta\leftarrow\Omega}[\mathcal{T}]_{\Omega\leftarrow\Gamma} &= [\ ]_\Delta \circ \mathcal{S} \circ \mathcal{A}_\Omega \circ [\ ]_\Omega \circ \mathcal{T} \circ \mathcal{A}_\Gamma \\ &= [\ ]_\Delta \circ \mathcal{S} \circ (\mathcal{A}_\Omega \circ [\ ]_\Omega) \circ \mathcal{T} \circ \mathcal{A}_\Gamma \\ &= [\ ]_\Delta \circ \mathcal{S} \circ \mathcal{T} \circ \mathcal{A}_\Gamma \\ &= [\ ]_\Delta \circ (\mathcal{S} \circ \mathcal{T}) \circ \mathcal{A}_\Gamma \\ &= [\mathcal{S} \circ \mathcal{T}]_{\Delta\leftarrow\Gamma}, \end{aligned}$$

where we have used the fact that $A_\Omega[\ ]_\Omega = A_\Omega A_\Omega^{-1}$ is the identity function. This completes the argument.

4. (a) <u>To show</u>: $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$ is invertible when $\mathcal{T}$ is invertible.

   The matrix $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$ can be expressed as

   $$[\mathcal{T}]_{\Omega\leftarrow\Gamma} = [\ ]_\Omega \circ \mathcal{T} \circ \mathcal{A}_\Gamma.$$

   Since T, $[\ ]_\Omega$, and $\mathcal{A}_\Gamma$ are invertible, the above composition is also invertible, i.e., $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$ is invertible.

(b) Let $\mathcal{T}$ be invertible (in which case $m = n$). <u>Claim</u>:

$$\left[\mathcal{T}^{-1}\right]_{\Gamma\leftarrow\Omega} = [\mathcal{T}]^{-1}_{\Omega\leftarrow\Gamma}.$$

Since $\mathcal{T}$, $\mathcal{A}_\Omega$, $\mathcal{A}_\Gamma$ are all invertible, their inverses are defined. By the set-up in definition 0.7, the matrix $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$ can be written as the composition $[\ ]_\Omega \circ \mathcal{T} \circ A_\Gamma$ for the linear function $T : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{Z}$. Since $\mathcal{T}^{-1} : \mathbf{Z} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ is defined, we can express its matrix $[\mathcal{T}^{-1}]_{\Gamma\leftarrow\Omega}$ as

$$[\mathcal{T}^{-1}]_{\Gamma\leftarrow\Omega} = [\ ]_\Gamma \circ \mathcal{T}^{-1} \circ \mathcal{A}_\Omega. \ (\dagger)$$

Consider the inverse of the matrix $[\mathcal{T}]_{\Omega\leftarrow\Gamma}$:

$$\begin{aligned} [\mathcal{T}]^{-1}_{\Omega\leftarrow\Gamma} &= ([\ ]_\Omega \circ \mathcal{T} \circ \mathcal{A}_\Gamma)^{-1} \\ &= \mathcal{A}_\Gamma^{-1} \circ ([\ ]_\Gamma \circ \mathcal{T})^{-1} \\ &= \mathcal{A}_\Gamma^{-1} \circ \mathcal{T}^{-1} \circ [\ ]_\Omega^{-1} \\ &= [\ ]_\Gamma \circ \mathcal{T}^{-1} \circ \mathcal{A}_\Omega. \ (\dagger\dagger) \end{aligned}$$

From $(\dagger)$ and $(\dagger\dagger)$, we have

$$[\mathcal{T}^{-1}]_{\Gamma\leftarrow\Omega} = [\mathcal{T}]^{-1}_{\Omega\leftarrow\Gamma}$$

as claimed.

Problem. 4. Let us write Γ_0 for the standard coordinate system $(1, x, x^2, x^3)$ of the vector space $\mathbb{P}_3$ of all polynomials of degree at most 3.

1. Use *Mathematica* and the fact that $\left[\quad\right]_{\Gamma_0}$ is an isomorphism (and isomorphisms map bases to bases) to verify that

$$\Delta := (1 + x + x^2 + x^3, 1 + 2x + 4x^2 + 8x^3,$$
$$1 + 3x + 9x^2 + 27x^3, 1 + 4x + 16x^2 + 64x^3)$$

and

$$\Omega := (x + x^2, x - 6x^3, 1 + 4x^2, 1 + 8x^3)$$

are also coordinate systems of $\mathbb{P}_3$.

2. Verify that general identity

$$\begin{aligned} [\mathcal{T}]_{\Delta\leftarrow\Omega} &= \left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Delta\leftarrow\Gamma_0} [\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0} \left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Omega} \\ &= \left(\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Delta}\right)^{-1} [\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0} \left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Omega}, \end{aligned}$$

for $\mathcal{T} : \mathbb{P}_3 \overset{\text{linear}}{\longrightarrow} \mathbb{P}_3$.

3. Consider $\mathcal{T} : \mathbb{P}_3 \to \mathbb{P}_3$ defined by

$$\mathcal{T}(p(x)) := xp'(2x) + p(x).$$

Prove that $\mathcal{T}$ is a linear function and use part 2 to find $[\mathcal{T}]_{\Delta\leftarrow\Omega}$.

Solution. 4.

1. (a) Let

$$\Delta := (1 + x + x^2 + x^3, 1 + 2x + 4x^2 + 8x^3,$$
$$1 + 3x + 9x^2 + 27x^3, 1 + 4x + 16x^2 + 64x^3)$$

be given. To show that Δ is a coordinate system of $\mathbf{V}$, it suffices to show there exists an isomorphism $[\mathcal{II}]_{\Gamma_0\leftarrow\Delta} : \mathbb{C}^4_{\Delta_0} \to \mathbb{C}^4_{\Gamma_0}$.

We can construct $[\mathcal{II}]_{\Gamma_0\leftarrow\Delta}$ column-by-column by letting $[\ ]_{\Gamma_0}$ act on the elements of Δ, by Part 2 of Problem 3:

$$[\mathcal{II}]_{\Gamma_0\leftarrow\Delta} = \begin{bmatrix} [(1 + x + x^2 + x^3)]_{\Gamma_0} \\ [(1 + 2x + 4x^2 + 8x^3)]_{\Gamma_0} \\ [(1 + 3x + 9x^2 + 27x^3)]_{\Gamma_0} \\ [(1 + 4x + 16x^2 + 64x^3)]_{\Gamma_0} \end{bmatrix}^\top = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}.$$

To show $[\mathcal{I}\mathcal{I}]_{\Gamma_0 \leftarrow \Delta}$ is an isomorphism, it suffices to check if its determinant is nonzero, which can be done in *Mathematica*:

$$\det\left([\mathcal{I}\mathcal{I}]_{\Gamma_0 \leftarrow \Delta}\right) = 12 \neq 0.$$

Therefore, Δ is a coordinate system of $\mathbb{P}_3$.

Mathematica code:

```
In[4]:= Det[{{1, 1, 1, 1}, {1, 2, 3, 4}, {1, 4, 9, 16},
                        {1, 8, 27, 64}}]

Out[4]= 12
```

(b) For $\Omega := (x + x^2, x - 6x^3, 1 + 4x^2, 1 + 8x^3)$, we repeat the procedure laid out above to determine whether Ω is a coordinate system for $\mathbb{P}_3$.

$$[\mathcal{I}\mathcal{I}]_{\Gamma_0 \leftarrow \Omega} = \begin{bmatrix} [(x + x^2)]_{\Gamma_0} \\ [(x - 6x^3)]_{\Gamma_0} \\ [(1 + 4x^2)]_{\Gamma_0} \\ [(1 + 8x^3)]_{\Gamma_0} \end{bmatrix}^{\top} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 4 & 0 \\ 0 & -6 & 0 & 8 \end{pmatrix}.$$

We can calculate the determinant of $[\mathcal{I}\mathcal{I}]_{\Gamma_0 \leftarrow \Omega}$ in *Mathematica*:

$$\det\left([\mathcal{I}\mathcal{I}]_{\Gamma_0 \leftarrow \Omega}\right) = -32 \neq 0.$$

Therefore, Ω is a coordinate system of $\mathbb{P}_3$.

Mathematica code:

```
In[6]:= Det[{{0, 0, 1, 1}, {1, 1, 0, 0},
                         {1, 0, 4, 0}, {0, -6, 0, 8}}]

Out[6]= -32
```

2. Let $\mathcal{T} : \mathbb{P}_3 \overset{\text{linear}}{\longrightarrow} \mathbb{P}_3$ be given.

(a) First identity:

$$\begin{aligned} & \left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Delta \leftarrow \Gamma_0} [\mathcal{T}]_{\Gamma_0 \leftarrow \Gamma_0} \left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0 \leftarrow \Omega} \\ = \ & (\left[\ \right]_{\Delta} \circ \mathcal{I}_{\mathbb{P}_3} \circ \mathcal{A}_{\Gamma_0}) \circ (\left[\ \right]_{\Gamma_0} \circ \mathcal{T} \circ \mathcal{A}_{\Gamma_0}) \circ (\left[\ \right]_{\Gamma_0} \circ \mathcal{I}_{\mathbb{P}_3} \circ \mathcal{A}_{\Omega}) \\ = \ & \left[\ \right]_{\Delta} \circ \mathcal{I}_{\mathbb{P}_3} \circ \mathcal{T} \circ \mathcal{I}_{\mathbb{P}_3} \circ \mathcal{A}_{\Omega} \\ = \ & \left[\ \right]_{\Delta} \circ \mathcal{T} \circ \mathcal{A}_{\Omega} \\ = \ & [\mathcal{T}]_{\Delta \leftarrow \Omega}. \end{aligned}$$

(b) The second identity follows from the fact that

$$\begin{aligned} \left([\mathcal{I}_{\mathbb{P}_3}]_{\Gamma_0 \leftarrow \Delta}\right)^{-1} &= \left(\left[\ \right]_{\Gamma_0} \circ \mathcal{I}_{\mathbb{P}_3} \circ \mathcal{A}_{\Delta}\right)^{-1} \\ &= \left[\ \right]_{\Delta} \circ \mathcal{I}_{\mathbb{P}_3}^{-1} \circ \mathcal{A}_{\Gamma_0} \\ &= \left[\ \right]_{\Delta} \circ \mathcal{I}_{\mathbb{P}_3} \circ \mathcal{A}_{\Gamma_0}. \end{aligned}$$

So we have

$$\begin{aligned}
[\mathcal{T}]_{\Delta\leftarrow\Omega} &= \left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Delta\leftarrow\Gamma_0}[\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0}\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Omega} \\
&= ([\ \ ]_\Delta \circ \mathcal{I}_{\mathbb{P}_3} \circ \mathcal{A}_{\Gamma_0}) \circ [\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0}\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Omega} \\
&= ([\mathcal{I}_{\mathbb{P}_3}]_{\Gamma_0\leftarrow\Delta})^{-1}[\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0}\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Omega}
\end{aligned}$$

3. Let $\mathcal{T} : \mathbb{P}_3 \to \mathbb{P}_3$ be given. T is defined as

$$\mathcal{T}(p(x)) := xp'(2x) + p(x).$$

(a) Claim: $\mathcal{T}$ is a linear function.
$\mathcal{T}$ is a linear function if it satisfies the linearity conditions. Consider $p(x), q(x) \in \mathbb{P}_3$. Then we have

$$\begin{aligned}
\mathcal{T}(p(x)+q(x)) &= x(p(2x)+q(2x))' + (p(x)+q(x)) \\
&= xp'(2x) + xq'(2x) + p(x) + q(x) \\
&= (xp'(2x)+p(x)) + (xq'(2x)+q(x)) \\
&= \mathcal{T}(p(x)) + \mathcal{T}(q(x)).\ \ (\dagger)
\end{aligned}$$

Consider $c \in \mathbb{C}$ and $p(x) \in \mathbb{P}_3$,

$$\begin{aligned}
\mathcal{T}(cp(x)) &= x(cp(2x))' + cp(x) \\
&= cxp'(2x) + cp(x) \\
&= c(xp'(2x)+p(x)) \\
&= c\mathcal{T}(p(x)).\ \ (\dagger\dagger)
\end{aligned}$$

From (†) and (†), $\mathcal{T}$ is a linear function.

(b) Find $[\mathcal{T}]_{\Delta\leftarrow\Omega}$.

By part 2 of Problem 3

$$[\mathcal{T}]_{\Delta\leftarrow\Omega} = \left(\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Delta}\right)^{-1}[\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0}\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Omega}.$$

In part 1, we have found the isomorphisms:

$$\left(\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Delta}\right)^{-1} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}^{-1} = \begin{pmatrix} 4 & -13/3 & 3/2 & -1/6 \\ -6 & 19/2 & -4 & 1/2 \\ 4 & -7 & 7/2 & -1/2 \\ -1 & 11/6 & -1 & 1/6 \end{pmatrix}$$

$$\left[\mathcal{I}_{\mathbb{P}_3}\right]_{\Gamma_0\leftarrow\Omega} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 4 & 0 \\ 0 & -6 & 0 & 8 \end{pmatrix}.$$

To find $[\mathcal{T}]_{\Delta\leftarrow\Omega}$ we only need to find $[\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0}$. To do this, we use part 2 of Problem 3, which says,

$$[\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0} = \begin{bmatrix}[\mathcal{T}(1)]_{\Gamma_0} & [\mathcal{T}(x)]_{\Gamma_0} & [\mathcal{T}(x^2)]_{\Gamma_0} & [\mathcal{T}(x^3)]_{\Gamma_0}\end{bmatrix},$$

where

$$\begin{aligned}
\mathcal{T}(1) &= x\frac{d}{dx}2 + 1 = 1\\
\mathcal{T}(x) &= x\frac{d}{dx}(2x) + x = 3x\\
\mathcal{T}(x^2) &= x\frac{d}{dx}(2x)^2 + x^2 = 9x^2\\
\mathcal{T}(x^3) &= x\frac{d}{dx}(2x)^3 + x^3 = 25x^3.
\end{aligned}$$

Therefore,

$$\begin{aligned}
[\mathcal{T}]_{\Gamma_0\leftarrow\Gamma_0} &= \begin{bmatrix}[T(1)]_{\Gamma_0} & [T(x)]_{\Gamma_0} & [T(x^2)]_{\Gamma_0} & [T(x^3)]_{\Gamma_0}\end{bmatrix}\\
&= \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & 9 & 0\\ 0 & 0 & 0 & 25\end{pmatrix}.
\end{aligned}$$

Finally, we can compute $[\mathcal{T}]_{\Delta\leftarrow\Omega}$:

$$\begin{aligned}
[\mathcal{T}]_{\Delta\leftarrow\Omega} &= \begin{pmatrix} 4 & -13/3 & 3/2 & -1/6\\ -6 & 19/2 & -4 & 1/2\\ 4 & -7 & 7/2 & -1/2\\ -1 & 11/6 & -1 & 1/6\end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & 9 & 0\\ 0 & 0 & 0 & 25\end{pmatrix}\begin{pmatrix} 0 & 0 & 1 & 1\\ 1 & 1 & 0 & 0\\ 1 & 0 & 4 & 0\\ 0 & -6 & 0 & 8\end{pmatrix}\\
&= \begin{pmatrix} 1/2 & 12 & 58 & -88/3\\ -15/2 & -93/2 & -150 & 94\\ 21/ & 54 & 130 & -96\\ -7/2 & -39/2 & -37 & 97/3\end{pmatrix}.
\end{aligned}$$

Mathematica code:

```
A =
Inverse[{{1, 1, 1, 1}, {1, 2, 3, 4}, {1, 4, 9, 16}, {1, 8, 27, 64}}]

{{4, -(13/3), 3/2, -(1/6)}, {-6, 19/2, -4, 1/2}, {4, -7, 7/
            2, -(1/2)}, {-1, 11/6, -1, 1/6}}

B = {{1, 0, 0, 0}, {0, 3, 0, 0}, {0, 0, 9, 0}, {0, 0, 0, 25}}


CC = {{0, 0, 1, 1}, {1, 1, 0, 0}, {1, 0, 4, 0}, {0, -6, 0, 8}}

In[7]:= A.B.CC

Out[7]= {{1/2, 12, 58, -(88/3)}, {-(15/2), -(93/2), -150, 94},
{21/2, 54, 130, -96}, {-(7/2), -(39/2), -37, 97/3}}
```

Problem. 5. Suppose that $\mathbf{V} = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3$ and Γ_i is a coordinate system of $\mathbf{W}_i$. Suppose that for each $i, j \in \{1, 2, 3\}$,

$$\mathcal{L}_{ij} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i.$$

Let

$$\mathcal{L} := \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus}.$$

Let Δ be the concatenation $\Gamma_1 || \Gamma_2 || \Gamma_3$ of the coordinate systems Γ_i. As we know, Δ is a coordinate system of $\mathbf{V}$.

Prove that $[\mathcal{L}]_{\Delta \leftarrow \Delta}$ equals the partitioned matrix

$$\begin{bmatrix} [\mathcal{L}_{11}]_{\Gamma_1 \leftarrow \Gamma_1} & [\mathcal{L}_{12}]_{\Gamma_1 \leftarrow \Gamma_2} & [\mathcal{L}_{13}]_{\Gamma_1 \leftarrow \Gamma_3} \\ [\mathcal{L}_{21}]_{\Gamma_2 \leftarrow \Gamma_1} & [\mathcal{L}_{22}]_{\Gamma_2 \leftarrow \Gamma_2} & [\mathcal{L}_{23}]_{\Gamma_2 \leftarrow \Gamma_3} \\ [\mathcal{L}_{31}]_{\Gamma_3 \leftarrow \Gamma_1} & [\mathcal{L}_{32}]_{\Gamma_3 \leftarrow \Gamma_2} & [\mathcal{L}_{33}]_{\Gamma_3 \leftarrow \Gamma_3} \end{bmatrix}.$$

Solution. 5. Consider a 3×3 block-matrix, which consists of just the linear function $\mathcal{L}_{ij} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i$ at the i^{th} row and j^{th} column and the zero function everywhere else. Let $[\mathcal{L}_{ij}]_{\oplus} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}$ denote this matrix. By this construction, we have

$$\mathcal{L} = \sum_{i=1}^{3} \sum_{j=1}^{3} [\mathcal{L}_{ij}]_{\oplus} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus}.$$

It follows that the matrix $[\mathcal{L}]_{\Delta \leftarrow \Delta}$ of $\mathcal{L}$ can be written as

$$\begin{aligned} \left[\mathcal{L}\right]_{\Delta \leftarrow \Delta} &= \left[\textstyle\sum_{i=1}^{3} \sum_{j=1}^{3} [\mathcal{L}_{ij}]_{\oplus}\right]_{\Delta \leftarrow \Delta} \\ &= \sum_{i=1}^{3} \sum_{j=1}^{3} \left[[\mathcal{L}_{ij}]_{\oplus}\right]_{\Delta \leftarrow \Delta} \end{aligned}$$

Consider the coordinate system Γ_k where $k \in \{1, 2, 3\}$. Let Γ_k be formed by a (finite) collection of basis elements, called g_{lk}'s. Consider the mapping $[\mathcal{L}_{ij}]_{\oplus}(g_{lk})_{\oplus}$ and the matrix $\left[[\mathcal{L}_{ij}]_{\oplus}\right]_{\Delta \leftarrow \Delta}$.

We observe that $[\mathcal{L}_{ij}]_{\oplus}(g_{lk})_{\oplus}$ is $\mathbf{0_V}$ if $k \neq j$ and $\maltese\left([\mathcal{L}_{ij}]_{\oplus}(g_{lj}) + \mathbf{0_V} + \mathbf{0_V}\right)$ if $j = k$, where $\maltese$ is defined in Definition 0.4. What this says is that $[\mathcal{L}_{ij}]_{\oplus}(g_{lk})_{\oplus}$ gives a column vector with $\mathcal{L}_{ij}(g_{lj})$ on the i^{th} row and $\mathbf{0_V}$ everywhere else. This means that if we consider the coordinatization $\left[[\mathcal{L}_{ij}]_{\oplus}(g_{lj})_{\oplus}\right]_{\Delta}$, where $\Delta = \Gamma_1 || \Gamma_2 || \Gamma_3$, we will end up with a column vector with the zero function everywhere except $\left[[\mathcal{L}_{ij}]_{\oplus}(g_{lj})\right]_{\Gamma_i}$ on the i^{th} row. (†)

Now, we can construct $\big[[\mathcal{L}_{ij}]_{\oplus}\big]_{\Delta\leftarrow\Delta}$ column-by-column as

$$\big[[\mathcal{L}_{ij}]_{\oplus}\big]_{\Delta\leftarrow\Delta} = \begin{bmatrix} \begin{bmatrix} \big[[\mathcal{L}_{ij}]_{\oplus}(g_{11})_{\oplus}\big]_{\Delta} \\ \cdots \\ \big[[\mathcal{L}_{ij}]_{\oplus}(g_{a1})_{\oplus}\big]_{\Delta} \end{bmatrix}^{\top} & \begin{bmatrix} \big[[\mathcal{L}_{ij}]_{\oplus}(g_{12})_{\oplus}\big]_{\Delta} \\ \cdots \\ \big[[\mathcal{L}_{ij}]_{\oplus}(g_{b2})_{\oplus}\big]_{\Delta} \end{bmatrix}^{\top} & \begin{bmatrix} \big[[\mathcal{L}_{ij}]_{\oplus}(g_{13})_{\oplus}\big]_{\Delta} \\ \cdots \\ \big[[\mathcal{L}_{ij}]_{\oplus}(g_{c3})_{\oplus}\big]_{\Delta} \end{bmatrix}^{\top} \end{bmatrix}$$

where a, b, c are the number of elements in Γ_1, Γ_2, Γ_3, respectively. From (†), we know that two of three matrix-blocks in the expression above will be the zero-block because j has a fixed value of either 1,2, or 3. Consider the only non-zero block (located at the j^{th} column)

$$\begin{bmatrix} \big[[\mathcal{L}_{ij}]_{\oplus}(g_{1j})_{\oplus}\big]_{\Delta} \\ \cdots \\ \big[[\mathcal{L}_{ij}]_{\oplus}(g_{aj})_{\oplus}\big]_{\Delta} \end{bmatrix}^{\top} = \begin{bmatrix} \big[[\mathcal{L}_{ij}]_{\oplus}(g_{1j})_{\oplus}\big]_{\Delta} & \cdots & \big[[\mathcal{L}_{ij}]_{\oplus}(g_{aj})_{\oplus}\big]_{\Delta} \end{bmatrix}. \qquad (††)$$

Recalling that $\big[[\mathcal{L}_{ij}]_{\oplus}(g_{lj})_{\oplus}\big]_{\Delta}$ is a column vector with $\mathcal{L}_{ij}(g_{lj})$ on the i^{th} row and $\mathbf{0_V}$ everywhere else, it follows that we can represent this matrix block as one with two rows of zeros and an i^{th} row of

$$\begin{bmatrix} \mathcal{L}_{ij}(g_{1j}) & \mathcal{L}_{ij}(g_{1j}) & \ldots & \mathcal{L}_{ij}(g_{aj}) \end{bmatrix},$$

which precisely expresses the matrix $\big[\mathcal{L}_{ij}\big]_{\Gamma_i\leftarrow\Gamma_j}$. Therefore, we can treat the expression (††) as a 3-element column of the matrix $\big[[\mathcal{L}_{ij}]_{\oplus}\big]_{\Delta\leftarrow\Delta}$ whose element on the i^{th} row and j^{th} column is $\big[\mathcal{L}_{ij}\big]_{\Gamma_i\leftarrow\Gamma_j}$, and the zero function everywhere else. Hence,

$$\begin{aligned} \big[\mathcal{L}\big]_{\Delta\leftarrow\Delta} &= \sum_{i=1}^{3}\sum_{j=1}^{3} \big[[\mathcal{L}_{ij}]_{\oplus}\big]_{\Delta\leftarrow\Delta} \\ &= \begin{bmatrix} \big[\mathcal{L}_{11}\big]_{\Gamma_1\leftarrow\Gamma_1} & \big[\mathcal{L}_{21}\big]_{\Gamma_2\leftarrow\Gamma_1} & \big[\mathcal{L}_{31}\big]_{\Gamma_3\leftarrow\Gamma_1} \\ \big[\mathcal{L}_{12}\big]_{\Gamma_1\leftarrow\Gamma_2} & \big[\mathcal{L}_{22}\big]_{\Gamma_2\leftarrow\Gamma_2} & \big[\mathcal{L}_{32}\big]_{\Gamma_3\leftarrow\Gamma_2} \\ \big[\mathcal{L}_{13}\big]_{\Gamma_1\leftarrow\Gamma_3} & \big[\mathcal{L}_{23}\big]_{\Gamma_2\leftarrow\Gamma_3} & \big[\mathcal{L}_{33}\big]_{\Gamma_3\leftarrow\Gamma_3} \end{bmatrix}. \end{aligned}$$

This completes the argument.

An example can help justify this argument. Consider $[\mathcal{L}_{11}]_{\oplus}$,

$$[\mathcal{L}_{11}] = \begin{bmatrix} \mathcal{L}_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{\oplus},$$

where we are using the 0's to denote the zero functions $\mathbf{W}_i \to \mathbf{W}_j$. Then consider g_{11} the first vector in the coordinate system Γ_1. We have

$$\begin{bmatrix} \mathcal{L}_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{\oplus} \begin{pmatrix} g_{11} \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}_{\maltese} = \begin{pmatrix} \mathcal{L}_{11}(g_{11}) \\ \mathbf{0_V} \\ \mathbf{0_V} \end{pmatrix}_{\maltese}.$$

Now consider $[[\mathcal{L}_{11}]_{\oplus}]_{\Delta\leftarrow\Delta}$. This matrix can be constructed column-by-column with each column being $[\mathcal{L}_{11}(g_{ij})]_{\Delta\leftarrow\Delta}$, where j indicates from which Γ_j g_{ij} is taken. We observe that if j is not 1, then $\mathcal{L}_{11}(g_{ij}) = \mathbf{0_V}$. This means the all columns of $[[\mathcal{L}_{11}]_{\oplus}]_{\Delta\leftarrow\Delta}$ are the zero columns unless $j = 1$. Consider the $j = 1$ columns. We know that the coordinatization

$$\left[\begin{pmatrix}\mathcal{L}_{11}(g_{11})\\ \mathbf{0_V}\\ \mathbf{0_V}\end{pmatrix}_{\maltese}\right]_{\Delta\leftarrow\Delta} = \begin{pmatrix}[\mathcal{L}_{11}(g_{11})]_{\Gamma_1}\\ \mathbf{0_V}\\ \mathbf{0_V}\end{pmatrix},$$

because the $\mathcal{L}_{11}(g_{11}) \in \mathbf{W}_1$. So it follows that the concatenation of the columns

$$\begin{pmatrix}[\mathcal{L}_{11}(g_{11})]_{\Gamma_1} & \dots & \mathcal{L}_{11}(g_{a1})]_{\Gamma_1}\\ \mathcal{O} & \dots & \mathcal{O}\\ \mathcal{O} & \dots & \mathcal{O}\end{pmatrix}$$

in the matrix $[[\mathcal{L}_{11}]_{\oplus}]_{\Delta\leftarrow\Delta}$ is nothing but

$$\begin{pmatrix}[\mathcal{L}_{11}]_{\Gamma_1\leftarrow\Gamma_1}\\ \mathcal{O}\\ \mathcal{O}\end{pmatrix}.$$

So we have

$$[[\mathcal{L}_{11}]_{\oplus}]_{\Delta\leftarrow\Delta} = \begin{bmatrix}[\mathcal{L}_{11}]_{\Gamma_1\leftarrow\Gamma_1} & \mathcal{O} & \mathcal{O}\\ \mathcal{O} & \mathcal{O} & \mathcal{O}\\ \mathcal{O} & \mathcal{O} & \mathcal{O}\end{bmatrix},$$

where $\mathcal{O}$ denotes the zero function.

We can equivalently generate other examples for $[\mathcal{L}_{ij}]_{\oplus}$, $i, j \in \{1, 2, 3\}$, and in the end add up the results (with different combinations of i, j's of course) to get $[\mathcal{L}]_{\Delta\leftarrow\Delta}$, expressed in the form given in the problem statement as desired.

24.4 Problem set 4

Problem. 1.

Suppose that $T \in \mathfrak{L}(\mathbf{V})$ and

$$\mathrm{Im}(T) = \mathbf{W} + \mathbf{Z},$$

where $\mathbf{W}$ and $\mathbf{Z}$ are subspaces of $\mathbf{V}$. Argue that

$$\mathbf{V} = T^{-1}[\mathbf{W}] + T^{-1}[\mathbf{Z}].$$

Solution. *removed for corrections*

Problem. 2.

1. Suppose that $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ are non-trivial vector spaces, and for each $i, j \in \{1, 2, 3\}$,

$$\mathcal{L}_{ij} : \mathbf{V}_j \overset{\text{linear}}{\longrightarrow} \mathbf{V}_i.$$

Let $\mathcal{L}$ be the block-matrix function:

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} : \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3 \overset{\text{linear}}{\longrightarrow} \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3.$$

Suppose that it turns out that $\mathbf{V}_2 = \mathbf{V}_{2.1} \times \mathbf{V}_{2.2}$. Then $\mathcal{L}$ maybe considered as a linear function on

$$\mathbf{V}_1 \times \mathbf{V}_{2.1} \times \mathbf{V}_{2.2} \times \mathbf{V}_3.$$

What is the corresponding block-matrix form of $\mathcal{L}$ and how does it relate to $\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times}$? Justify your claims.

2. Suppose that $\mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3 = \mathbf{V}$ and the $\mathbf{W}_i$'s are non-trivial subspaces of $\mathbf{V}$. Suppose that for each $i, j \in \{1, 2, 3\}$:

$$\mathcal{L}_{ij} : \mathbf{W}_j \overset{\text{linear}}{\longrightarrow} \mathbf{W}_i.$$

Let $\mathcal{L}$ be the block-matrix function

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} : \mathbf{V} \overset{\text{linear}}{\longrightarrow} \mathbf{V}.$$

Suppose it turns out that $\mathbf{W}_1 = \mathbf{W}_{1.1} \oplus \mathbf{W}_{1.2}$. Then

$$\mathbf{V} = \mathbf{W}_{1.1} \oplus \mathbf{W}_{1.2} \oplus \mathbf{W}_2 \oplus \mathbf{W}_3.$$

What is the block-matrix form of $\mathcal{L}$ with respect to this direct sum decomposition and how does it relate to $\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus}$? Justify your claims.

Solution.

1. Recall from the previous problem set that the elements of the block matrix function

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} : \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3 \xrightarrow{\text{linear}} \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3.$$

can be expressed as

$$\mathcal{L}_{ij} = \Pi_i \circ \mathcal{L} \circ \gamma_j,$$

where the Π_i's and γ_j's are coordinate projections from $\mathbf{V}$ onto $\mathbf{V}_i$ and injections from $\mathbf{V}_j$ into $\mathbf{V}$. Let $\mathbf{V}_2 = \mathbf{V}_{2.1} \times \mathbf{V}_{2.2}$. We claim that the matrix-block form of $\mathcal{L}$ in this case is

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{1|2.1} & \mathcal{L}_{1|2.2} & \mathcal{L}_{13} \\ \mathcal{L}_{2.1|1} & \mathcal{L}_{2.1|2.1} & \mathcal{L}_{2.1|2.2} & \mathcal{L}_{2.1|3} \\ \mathcal{L}_{2.2|1} & \mathcal{L}_{2.2|2.1} & \mathcal{L}_{2.2|2.2} & \mathcal{L}_{2.2|3} \\ \mathcal{L}_{31} & \mathcal{L}_{3|2.1} & \mathcal{L}_{3|2.2} & \mathcal{L}_{33} \end{bmatrix}_{\times}.$$

In particular, we claim the following relations between the matrix elements $\mathcal{L}_{ij}$ in the former expression and the block-elements in the latter expression:

$$\mathcal{L}_{12} = \begin{bmatrix} \mathcal{L}_{1|2.1} & \mathcal{L}_{1|2.2} \end{bmatrix}, \mathcal{L}_{32} = \begin{bmatrix} \mathcal{L}_{3|2.1} & \mathcal{L}_{3|2.2} \end{bmatrix},$$

$$\mathcal{L}_{21} = \begin{bmatrix} \mathcal{L}_{2.1|1} \\ \mathcal{L}_{2.2|1} \end{bmatrix}, \mathcal{L}_{23} = \begin{bmatrix} \mathcal{L}_{2.1|3} \\ \mathcal{L}_{2.2|3} \end{bmatrix}, \mathcal{L}_{22} = \begin{bmatrix} \mathcal{L}_{2.1|2.1} & \mathcal{L}_{2.1|2.2} \\ \mathcal{L}_{2.2|2.1} & \mathcal{L}_{2.2|2.2} \end{bmatrix}.$$

We can mimic what we have done to define $\mathcal{L}_{ij}$ to define these new block-matrix elements. Let $\delta_{n.m}$ be defined as the coordinate injection from $\mathbf{V}_{n.m}$ into $\mathbf{V}_2$, $n, m \in \{1, 2\}$. Also, define $\rho_{n.m}$ as the coordinate projection from $\mathbf{V}$ onto $\mathbf{V}_{n.m}$, $n, m \in \{1, 2\}$. Consider the linear function $\mathcal{L}_{22} : \mathbf{V}_2 \to \mathbf{V}_2$, which we claim to be

$$\mathcal{L}_{22} = \begin{bmatrix} \mathcal{L}_{2.1|2.1} & \mathcal{L}_{2.1|2.2} \\ \mathcal{L}_{2.2|2.1} & \mathcal{L}_{2.2|2.2} \end{bmatrix}.$$

By exactly the same way we have defined the $\mathcal{L}_{ij}$'s of $\mathcal{L}$ using the coordinate projections and injections, we have

$$\begin{aligned} \mathcal{L}_{2.1|2.1} &= \rho_{2.1} \circ \mathcal{L}_{22} \circ \delta_{2.1} \\ \mathcal{L}_{2.2|2.1} &= \rho_{2.2} \circ \mathcal{L}_{22} \circ \delta_{2.1} \\ \mathcal{L}_{2.1|2.2} &= \rho_{2.1} \circ \mathcal{L}_{22} \circ \delta_{2.2} \\ \mathcal{L}_{2.2|2.2} &= \rho_{2.2} \circ \mathcal{L}_{22} \circ \delta_{2.2} \end{aligned}$$

In terms of the linear function $\mathcal{L}$,

$$\begin{aligned}
\mathcal{L}_{2.1|2.1} &= \rho_{2.1} \circ \Pi_2 \circ \mathcal{L} \circ \gamma_2 \circ \delta_{2.1} \\
\mathcal{L}_{2.2|2.1} &= \rho_{2.2} \circ \Pi_2 \circ \mathcal{L} \circ \gamma_2 \circ \delta_{2.1} \\
\mathcal{L}_{2.1|2.2} &= \rho_{2.1} \circ \Pi_2 \circ \mathcal{L} \circ \gamma_2 \circ \delta_{2.2} \\
\mathcal{L}_{2.2|2.2} &= \rho_{2.2} \circ \Pi_2 \circ \mathcal{L} \circ \gamma_2 \circ \delta_{2.2}.
\end{aligned}$$

Next, consider $\mathcal{L}_{i2} = \Pi_i \circ \mathcal{L} \circ \gamma_2$. We define

$$\begin{aligned}
\mathcal{L}_{i|2.1} &= \Pi_i \circ \mathcal{L} \circ \gamma_2 \circ \delta_{2.1} \\
\mathcal{L}_{i|2.2} &= \Pi_i \circ \mathcal{L} \circ \gamma_2 \circ \delta_{2.2},
\end{aligned}$$

for $i \in \{1,3\}$. Finally, consider $\mathcal{L}_{2j} = \Pi_2 \circ \mathcal{L} \circ \gamma_j$. We define

$$\begin{aligned}
\mathcal{L}_{2.1|j} &= \rho_{2.1} \circ \Pi_2 \circ \mathcal{L} \circ \gamma_j \\
\mathcal{L}_{2.2|j} &= \rho_{2.2} \circ \Pi_2 \circ \mathcal{L} \circ \gamma_j,
\end{aligned}$$

for $j \in \{1,3\}$.

We can now verify that $\mathcal{L}$ can be represented as the 4-block-by-4-block as above. Consider

$$v = \begin{pmatrix} a \\ \begin{pmatrix} b \\ c \end{pmatrix} \\ d \end{pmatrix} \in \mathbf{V}_1 \times \mathbf{V}_2 \times \mathbf{V}_3$$

with $\mathbf{V}_2 = \mathbf{V}_{2.1} \times \mathbf{V}_{2.2}$. v can be written as

$$v = \begin{pmatrix} a \\ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \begin{pmatrix} b \\ \mathbf{0} \end{pmatrix} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \begin{pmatrix} \mathbf{0} \\ c \end{pmatrix} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \\ d \end{pmatrix}.$$

We can check that $\mathcal{L}(v)$ is the same as applying the 4-block-by-4-block matrix expression to v by going through every term in the above expression of v. For example, we can consider $\begin{pmatrix} a & \begin{pmatrix} \mathbf{0} & \mathbf{0} \end{pmatrix}^\top & \mathbf{0} \end{pmatrix}^\top$. Then

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{1|2.1} & \mathcal{L}_{1|2.2} & \mathcal{L}_{13} \\ \mathcal{L}_{2.1|1} & \mathcal{L}_{2.1|2.1} & \mathcal{L}_{2.1|2.2} & \mathcal{L}_{2.1|3} \\ \mathcal{L}_{2.2|1} & \mathcal{L}_{2.2|2.1} & \mathcal{L}_{2.2|2.2} & \mathcal{L}_{2.2|3} \\ \mathcal{L}_{31} & \mathcal{L}_{3|2.1} & \mathcal{L}_{3|2.2} & \mathcal{L}_{33} \end{bmatrix}_\times \begin{pmatrix} a \\ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathcal{L}_{11}(a) \\ \begin{pmatrix} \mathcal{L}_{2.1|1}(a) \\ \mathcal{L}_{2.2|1}(a) \end{pmatrix} \\ \mathcal{L}_{31}(a) \end{pmatrix}.$$

We can look at

$$\begin{aligned}
\begin{pmatrix} \mathcal{L}_{2.1|1}(a) \\ \mathcal{L}_{2.2|1}(a) \end{pmatrix} &= \begin{pmatrix} \rho_{2.1} \circ \mathcal{L} \circ \gamma_1(a) \\ \rho_{2.2} \circ \mathcal{L} \circ \gamma_1(a) \end{pmatrix} \\
&= \begin{pmatrix} \rho_{2.1} \circ \Pi_2 \circ \mathcal{L}_{21}(a) \\ \rho_{2.2} \circ \Pi_2 \circ \mathcal{L}_{21}(a) \end{pmatrix} \\
&= \mathcal{L}_{21}(a), \qquad \text{by the construction of } \rho_{2.1}, \rho_{2.2}.
\end{aligned}$$

Hence

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{1|2.1} & \mathcal{L}_{1|2.2} & \mathcal{L}_{13} \\ \mathcal{L}_{2.1|1} & \mathcal{L}_{2.1|2.1} & \mathcal{L}_{2.1|2.2} & \mathcal{L}_{2.1|3} \\ \mathcal{L}_{2.2|1} & \mathcal{L}_{2.2|2.1} & \mathcal{L}_{2.2|2.2} & \mathcal{L}_{2.2|3} \\ \mathcal{L}_{31} & \mathcal{L}_{3|2.1} & \mathcal{L}_{3|2.2} & \mathcal{L}_{33} \end{bmatrix}_{\times} \begin{pmatrix} a \\ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathcal{L}_{11}(a) \\ \begin{pmatrix} \mathcal{L}_{2.1|1}(a) \\ \mathcal{L}_{2.2|1}(a) \end{pmatrix} \\ \mathcal{L}_{31}(a) \end{pmatrix} = \begin{pmatrix} \mathcal{L}_{11}(a) \\ \mathcal{L}_{21}(a) \\ \mathcal{L}_{31}(a) \end{pmatrix}.$$

We can repeat this checking process and find that the middle two "columns" of the 4-by-4 block expression correspond to the second column of the 3-by-3 block expression, and that the last column of the 4-by-4 expression corresponds to the the third column of the 3-by-3 expression.

This will complete our justification.

2. From the previous problem set, we know that

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus} = \maltese \circ \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times} \circ \maltese^{-1}$$

where $\maltese$ is the isomorphism $\mathbf{W}_1 \times \mathbf{W}_2 \times \mathbf{W}_3 \overset{\text{linear}}{\longrightarrow} \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3$ defined in the last problem set. Since $\mathbf{W}_1 = \mathbf{W}_{1.1} \oplus \mathbf{W}_{1.2}$, we can define a similar isomorphism $\clubsuit : \mathbf{W}_{1.1} \times \mathbf{W}_{1.2} \overset{\text{linear}}{\longrightarrow} \mathbf{W}_{1.1} \oplus \mathbf{W}_{1.2}$ where for $w_{1.1} \in \mathbf{W}_1$ and $w_{1.2} \in \mathbf{W}_{1.2}$,

$$\clubsuit \begin{pmatrix} w_{1.1} \\ w_{1.2} \end{pmatrix} = w_{1.1} + w_{1.2}.$$

Also, define the linear functions $\sigma_{1.m} := \rho_{1.m} \circ \clubsuit^{-1} : \mathbf{W}_1 \to \mathbf{W}_{1.m}$ and $\omega_{1.m} := \clubsuit \circ \delta_{1.m} : \mathbf{W}_{1.m} \to \mathbf{W}_1$, where $\rho_{1.m}$ and $\delta_{1.m}$ are linear functions defined similarly as in Part 1, $m = \{1, 2\}$.

Considering the previous part of this problem, we know that the block-matrix expression for $\mathcal{L}_{\times}$ corresponding to $\mathbf{W}_1 = \mathbf{W}_{1.1} \times \mathbf{W}_{1.2}$ is

$$\begin{bmatrix} \mathcal{L}_{1.1|1.1} & \mathcal{L}_{1.1|1.2} & \mathcal{L}_{1.1|2} & \mathcal{L}_{1.1|3} \\ \mathcal{L}_{1.2|1.1} & \mathcal{L}_{1.2|1.2} & \mathcal{L}_{1.2|2} & \mathcal{L}_{1.2|3} \\ \mathcal{L}_{2|1.1} & \mathcal{L}_{2|1.2} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{3|1.1} & \mathcal{L}_{3|1.2} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\times},$$

where the matrix elements are defined in the same way as in Part 1. We claim that the 4-by-4 block-matrix expression for $\mathcal{L}_{\oplus}$ is

$$\begin{bmatrix} \mathcal{L}_{1.1|1.1} & \mathcal{L}_{1.1|1.2} & \mathcal{L}_{1.1|2} & \mathcal{L}_{1.1|3} \\ \mathcal{L}_{1.2|1.1} & \mathcal{L}_{1.2|1.2} & \mathcal{L}_{1.2|2} & \mathcal{L}_{1.2|3} \\ \mathcal{L}_{2|1.1} & \mathcal{L}_{2|1.2} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{3|1.1} & \mathcal{L}_{3|1.2} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix}_{\oplus}.$$

Recall the linear functions $\mathcal{E}_i := \Pi_i \circ \maltese^{-1} : \mathbf{V} \to \mathbf{V}_i$ and $\mathcal{K}_j := \maltese \circ \gamma_j : \mathbf{W}_j \to \mathbf{V}$ from the last problem set. We have also defined

$$\mathcal{L}_{ij} = \mathcal{E}_i \circ \mathcal{L}_\oplus \circ \mathcal{K}_j.$$

We claim that

$$\mathcal{L}_{11} = \begin{bmatrix} \mathcal{L}_{1.1|1.1} & \mathcal{L}_{1.1|1.2} \\ \mathcal{L}_{1.2|1.1} & \mathcal{L}_{1.2|1.2} \end{bmatrix}, \mathcal{L}_{12} = \begin{bmatrix} \mathcal{L}_{1.1|2} \\ \mathcal{L}_{1.2|2} \end{bmatrix}, \mathcal{L}_{13} = \begin{bmatrix} \mathcal{L}_{1.1|3} \\ \mathcal{L}_{1.2|3} \end{bmatrix},$$

$$\mathcal{L}_{21} = \begin{bmatrix} \mathcal{L}_{2|1.1} & \mathcal{L}_{2|1.2} \end{bmatrix}, \mathcal{L}_{31} = \begin{bmatrix} \mathcal{L}_{3|1.1} & \mathcal{L}_{3|1.2} \end{bmatrix}.$$

The definitions of these matrix elements are similar to those in Part 1, except that we have to use the isomorphism $\clubsuit$ to obtain the appropriate inputs for the $\mathcal{L}_{ij}$'s. Consider the elements of $\mathcal{L}_{11}$. We claim that

$$\mathcal{L}_{1.i|1.j} = \sigma_{1.i} \circ \mathcal{E}_1 \circ \mathcal{L}_\oplus \circ \mathcal{K}_1 \circ \omega_{1.j}.$$

For the elements of $\mathcal{L}_{21}$ and $\mathcal{L}_{31}$, we claim that

$$\mathcal{L}_{i|1.j} = \mathcal{E}_i \circ \mathcal{L}_\oplus \circ \mathcal{K}_1 \circ \omega_{1.j}.$$

For the elements of $\mathcal{L}_{12}$ and $\mathcal{L}_{13}$, we claim that

$$\mathcal{L}_{1.i|j} = \sigma_{1.i} \circ \mathcal{E}_1 \circ \mathcal{L}_\oplus \circ \mathcal{K}_j.$$

The justification for these definitions can be reproduced from the example in Part 1., except instead of consider a v in the Cartesian product, we now considering a v in the direct sum:

$$v = \begin{pmatrix} a & \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}_\clubsuit & \mathbf{0} \end{pmatrix}_\maltese^\top,$$

which can be expressed as

$$\begin{pmatrix} a & \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}_\clubsuit & \mathbf{0} \end{pmatrix}_\maltese^\top + \begin{pmatrix} \mathbf{0} & \begin{pmatrix} b \\ \mathbf{0} \end{pmatrix}_\clubsuit & \mathbf{0} \end{pmatrix}_\maltese^\top + \begin{pmatrix} \mathbf{0} & \begin{pmatrix} \mathbf{0} \\ c \end{pmatrix}_\clubsuit & \mathbf{0} \end{pmatrix}_\maltese^\top + \begin{pmatrix} \mathbf{0} & \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}_\clubsuit & d \end{pmatrix}_\maltese^\top.$$

Verifying the above definitions is now only a matter of evaluating $\mathcal{L}_\oplus$ applied to each term in this expression for v and compare the result to the 4-by-4 block representation applied to the same term. This will complete our justification.

Problem. 3.

Suppose that relatively prime polynomials p_1 and p_2 have degrees 6 and 11 respectively. Consider the function:

$$\psi : \mathbb{P}_{10} \times \mathbb{P}_5 \to \mathbb{P}_{16}$$

defined by

$$\psi \begin{pmatrix} f \\ g \end{pmatrix} := f \cdot p_1 - g \cdot p_2.$$

Verify each o the following claims.

1. ψ is a linear function.
2. ψ is injective.
3. ψ is surjective.
4. There exist polynomials q_1 and q_2 such that

   $$q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1},$$

   where $\mathbb{1}$ is the constantly 1 polynomial.

Solution.

1. (a) Consider

   $$\begin{pmatrix} f \\ g \end{pmatrix}, \begin{pmatrix} h \\ k \end{pmatrix} \in \mathbb{P}_{10} \times \mathbb{P}_5.$$

   Then

   $$\begin{aligned} \psi \begin{pmatrix} f \\ g \end{pmatrix} + \psi \begin{pmatrix} h \\ k \end{pmatrix} &= (f \cdot p_1 - g \cdot p_2) + (h \cdot p_1 - k \cdot p_1) \\ &= (f + h) \cdot p_1 - (g + k) \cdot p_2 \\ &= \psi \begin{pmatrix} f + h \\ g + k \end{pmatrix} \\ &= \psi \left(\begin{pmatrix} f \\ g \end{pmatrix} + \begin{pmatrix} h \\ k \end{pmatrix} \right) \qquad (\dagger) \end{aligned}$$

   (b) Consider $c \in \mathbb{C}$,

   $$\psi \begin{pmatrix} cf \\ cg \end{pmatrix} = cf \cdot p_1 - cg \cdot p_2 = c(f \cdot p_1 - g \cdot p_2) = c\psi \begin{pmatrix} f \\ g \end{pmatrix}. \qquad (\dagger\dagger)$$

   From $(\dagger)$ and $(\dagger\dagger)$, ψ is a linear function.

2. Assume (to get a contradiction) that $\begin{pmatrix} f \\ g \end{pmatrix}$ is a non-zero element in $\ker(\psi)$. Then

$$f \cdot p_1 - g \cdot p_2 = \mathbb{0}.$$

Since p_1, p_2 are both non-zero (they are relatively prime), if either one of f or g were zero, it would force the other polynomial to be zero so that $f \cdot p_1 - g \cdot p_2 = \mathbb{0}$, contradicting the requirement that $\begin{pmatrix} f \\ g \end{pmatrix}$ is non-zero element. Therefore, neither f nor g can be zero.

Since p_1 and p_2 are relatively prime (i.e., they have no common roots), the roots of p_2 must also be roots of f. Furthermore, the roots of p_2 must also be roots of f with equal or greater multiplicities, because otherwise there will at least one factor $(x - \lambda_k)$ in the factorization of p_2 that has a higher power (or multiplicity) than that in f (and not in p_1), resulting in a non-zero $f \cdot p_1 - g \cdot p_2$.

This implies that the degree of f has to be at least $\deg(p_2) = 11$. But since $\deg(f) \leq 10$, we have a contradiction. By symmetry, we will also have a contradiction if we repeat this argument for g and p_1.

Hence, there is no such non-zero $\begin{pmatrix} f \\ g \end{pmatrix}$ in the kernel of ψ, i.e., ψ is injective.

3. Since both $\mathbb{P}_{10} \times \mathbb{P}_5$ and $\mathbb{P}_{16}$ are finite-dimensional vector spaces, we can use the Rank-Nullity Theorem:

$$\dim(\ker(\psi)) + \dim(\operatorname{Im}(\psi)) = \dim(\mathbb{P}_{10} \times \mathbb{P}_5).$$

From part 1., we know $\dim(\ker(\psi)) = 0$, so

$$\dim(\operatorname{Im}(\psi)) = \dim(\mathbb{P}_{10} \times \mathbb{P}_5) = \dim(\mathbb{P}_{10}) + \dim(\mathbb{P}_5) = 11 + 6 = 17 = \dim(\mathbb{P}_{16}).$$

Furthermore, $\operatorname{Im}(\psi) \prec \mathbb{P}_{16}$. Hence, $\operatorname{Im}(\psi) = \mathbb{P}_{16}$. So, ψ is surjective.

4. Since ψ is surjective, there exist polynomials $f \in \mathbb{P}_{10}$ and $g \in \mathbb{P}_5$ such that

$$\psi \begin{pmatrix} f \\ g \end{pmatrix} = \mathbb{1} \in \mathbb{P}_{16} = \operatorname{Im}(\psi),$$

i.e., the given f and g satisfy

$$f \cdot p_1 - g \cdot p_2 = \mathbb{1}.$$

Let $q_1 = f \in \mathbb{P}_{10}$ and $q_2 = -g \in \mathbb{P}_5$, then

$$q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}.$$

This verifies the existence of q_1 and q_2.

Problem. 4.

1. What would you do to prove the last claim of problem 2 in a general case of non-zero relatively prime polynomials p_1 and p_2? You do not need to carry out the proof, but you DO need to set it all up along the lines of Problem 2. Make sure you cover all of the cases!

2. Prove that the following claims are equivalent.

   (a) Non-zero polynomials p_1 and p_2 are relatively prime.

   (b) There exist polynomials q_1 and q_2 such that

   $$q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}.$$

3. Argue that for any non-zero polynomials f and g there exist polynomials q_1 and q_2 such that

   $$q_1 \cdot f + q_2 \cdot g = \gcd(f, g).$$

4. Argue that the following claims are equivalent for any non-zero polynomials f, g, and h.

   (a) There exist polynomials q_1 and q_2 such that

   $$q_1 \cdot f + q_2 \cdot g = h.$$

   (b) h is a polynomial multiple of $\gcd(f, g)$.

Solution.

1. (a) Case 1: Both p_1 and p_2 are zero-degree polynomials.

   In this case, p_1 and p_2 are two constant, non-zero polynomials. Let $p_2(z) = C \in \mathbb{C}$ for any $z \in \mathbb{C}$, $C \neq 0$. If we let $q_1(z)$ be the constant 0 and $q_2(z)$ be the constant $1/C$ for any $z \in \mathbb{C}$ then

   $$q_1(z)p_1(z) + q_2(z)p_2(z) = 0 + \frac{1}{C}C = 0 + 1 = 1,$$

   for any $z \in \mathbb{C}$, i.e., $q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}$. This shows there are polynomials q_1 and q_2 such that $q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}$ is satisfied.

   (b) Case 2: Exactly one of p_1 and p_2 is a zero-degree polynomial.

   Assume that p_1 is the non-zero zero-degree polynomial of the two, i.e., it is a non-zero constant polynomial. Let $p_1(z) = C \in \mathbb{C}$, $C \neq 0$, for any $z \in \mathbb{C}$. If we let $q_1(z)$ be the constant $1/C$ and $q_2(z)$ be the constant 0 for any $z \in \mathbb{C}$ then

$$q_1(z)p_1(z) + q_2(z)p_2(z) = \frac{1}{C}C + 0 = 1 + 0 = 1,$$

for any $z \in \mathbb{C}$, i.e., $q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}$. This shows there are polynomials q_1 and q_2 such that $q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}$ is satisfied.

(c) Case 3: Both p_1 and p_2 have degree greater than 0.

To show the existence of q_1 and q_2 such that $q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}$ we can construct a similar proof to that in Problem 3. Let the polynomials p_1 and p_2 be given with degrees $m, n > 0$ respectively. Consider the function:

$$\psi : \mathbb{P}_{n-1} \times \mathbb{P}_{m-1} \to \mathbb{P}_{m+n-1}, \tag{13}$$

defined by

$$\psi \begin{pmatrix} f \\ g \end{pmatrix} := f \cdot p_1 + g \cdot p_2. \tag{14}$$

We can then show that ψ is a linear function. Next, we show that it is injective, which implies surjectivity by the Rank-Nullity Theorem. By the surjectivity of ψ, we know that given $\mathbb{1} \in \mathrm{Im}(\psi)$, there must exist a pair $f \in \mathbb{P}_{n-1}, -g \in \mathbb{P}_{m-1}$ such that $\psi \begin{pmatrix} f \\ -g \end{pmatrix} = \mathbb{1}$.

From parts (a), (b), and (c), we can conclude that there exist polynomials q_1, q_2 such that for non-zero relatively prime polynomials p_1, p_2, $q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1}$.

2.
- [$a. \implies b.$] This implication is true by Part 1. of this problem.
- [$b. \implies a.$] Let polynomials q_1 and q_2 be given such that

$$q_1 \cdot p_1 + q_2 \cdot p_2 = \mathbb{1},$$

where p_1 and p_2 are also polynomials. Claim: Non-zero p_1 and p_2 are relatively prime.

Assume that p_1 and p_2 are non-zero. From TYC 2.4, we can write p_1 and p_2 as

$$p_1 = \hat{p}_1 \cdot \gcd(p_1, p_2)$$
$$p_2 = \hat{p}_2 \cdot \gcd(p_1, p_2),$$

where $\hat{p}_1$ and $\hat{p}_2$ are non-zero relatively prime polynomials. It follows that

$$q_1 \cdot \hat{p}_1 \cdot \gcd(p_1, p_2) + q_2 \cdot \hat{p}_2 \cdot \gcd(p_1, p_2) = \mathbb{1},$$

i.e.,

$$(q_1 \cdot \hat{p}_1 + q_2 \cdot \hat{p}_2) \cdot \gcd(p_1, p_2) = \mathbb{1},$$

which implies

$$\deg(q_1 \cdot \hat{p}_1 + q_2 \cdot \hat{p}_2) = \deg(\gcd(p_1, p_2)) = 0.$$

Therefore, $\gcd(p_1, p_2)$ has to be a constant polynomial. Moreover, since $\gcd(p_1, p_2)$ is monic by definition, it has to be the constantly 1 polynomial. Hence, p_1 and p_2 are relatively prime by definition.

3. Let non-zero polynomials f, g be given. Then we have

$$\begin{aligned} f &= \hat{f} \cdot \gcd(f, g) \\ g &= \hat{g} \cdot \gcd(f, g), \end{aligned}$$

where $\hat{f}, \hat{g}$ are relatively prime. It follows that

$$\begin{aligned} q_1 \cdot f + q_2 \cdot g &= q_1 \cdot \hat{f} \cdot \gcd(f, g) + q_2 \cdot \hat{g} \cdot \gcd(f, g) \\ &= (q_1 \cdot \hat{f} + q_2 \cdot \hat{g}) \cdot \gcd(f, g). \end{aligned}$$

From Part 2., we know that for relatively prime polynomials $\hat{f}$ and $\hat{g}$, there exist polynomials q_1, q_2 such that

$$q_1 \cdot \hat{f} + q_2 \cdot \hat{g} = \mathbb{1}.$$

This means such q_1 and q_2 will give

$$(q_1 \cdot \hat{f} + q_2 \cdot \hat{g}) \cdot \gcd(f, g) = \mathbb{1} \cdot \gcd(f, g) = \gcd(f, g).$$

This completes the argument.

4. • [$a. \implies b.$] Let non-zero polynomials f, g, h be given such that

$$q_1 \cdot f + q_2 \cdot g = h,$$

for some polynomials q_1, q_2. From Part 3., we know that

$$\begin{aligned} h = q_1 \cdot f + q_2 \cdot g &= q_1 \cdot \hat{f} \cdot \gcd(f, g) + q_2 \cdot \hat{g} \cdot \gcd(f, g) \\ &= (q_1 \cdot \hat{f} + q_2 \cdot \hat{g}) \cdot \gcd(f, g), \end{aligned}$$

for relatively prime $\hat{f}$ and $\hat{g}$. This implies h is a polynomial multiple of $\gcd(p_1, p_2)$.

- [$b. \implies a.$] Let h be written as $k \cdot \gcd(f, g)$ where k is some polynomial. From Part 3., there exist polynomials s_1, s_2 such that

  $$s_1 \cdot f + s_2 \cdot g = \gcd(f, g).$$

  Multiplying both sides by the polynomial k, we have

  $$k \cdot s_1 \cdot f + k \cdot s_2 \cdot g = k \cdot \gcd(f, g) = h.$$

  Hence, there exist polynomials q_1 and q_2 such that

  $$q_1 \cdot f + q_2 \cdot g = h,$$

  namely,

  $$\begin{aligned} q_1 &= k \cdot s_1 \\ q_2 &= k \cdot s_2. \end{aligned}$$

Problem. 5.

Suppose that $\mathbf{W}$ is a non-$\{\mathbb{0}\}$ ideal in $\mathbb{P}$. Then $\mathbf{W}$ contains some monic polynomials (why?) and among these there must be some of the smallest degree, say n_0. Let p_0 be one such. (It is entirely possible that $p_0 = \mathbb{1}$.)

1. Suppose that p is a non-zero polynomial in $\mathbf{W}$, and using the Division Algorithm for Polynomials (see Axler p.121) we write

$$p = q \cdot p_0 + r,$$

   where $q, r \in \mathbb{P}$ and $\deg(r) < \deg(p_0)$. Argue that $r \in \mathbf{W}$.

2. Use the result of part 1 to argue that every polynomial p in $\mathbf{W}$ is a polynomial multiple of p_0.

3. Argue that p_0 is the only monic polynomial of the smallest degree n_0 in $\mathbf{W}$, and that

$$\mathbf{W} = \left\{ q \cdot p_0 \middle| q \in \mathbb{P} \right\}.$$

This polynomial p_0 is said to be **the generator** of the ideal $\mathbf{W}$.

Solution.

1. Let $p \in \mathbf{W}$ a non-zero polynomial be given. Let $p_0 \in \mathbf{W}$ be the monic polynomial with the smallest degree. $\mathbf{W}$ is an ideal in $\mathbb{P}$, so for $q \in \mathbb{P}$, $q \cdot p_0 \in \mathbf{W}$. $\mathbf{W}$ is also subspace, so

$$r = p - q \cdot p_0 \in \mathbf{W}.$$

2. Let the above q, p_0, p, r be given again. We know that

$$p = q \cdot p_0 + r,$$

   where $\deg(r) < \deg(p_0)$. Since $\deg(p_0)$ minimal and p is non-zero, r has to be the constantly zero polynomial. Therefore,

$$p = q \cdot p_0 \in \mathbf{W},$$

   i.e., every $p \in \mathbf{W}$ is a polynomial multiple of p_0.

3. Let some p_0' be a monic polynomial of the smallest degree n_0 in $\mathbf{W}$ be given. Since $p_0' \in \mathbf{W}$, p_0' is a polynomial multiple of p_0, by the previous part. Let us write $p_0' = q \cdot p_0$ for some $p \in \mathbb{P}$. Since $\deg(p_0') = \deg(p_0) = n_0$, $\deg(q) = 1$, i.e., p is a constant polynomial. Furthermore, since both p_0 and p_0' are monic, q has to be $\mathbb{1}$. Hence, $p_0' = p_0$, i.e., p_0 is unique.

   It follows from the previous part that $\mathbf{W}$ is a collection of elements that are polynomial multiples of a unique monic polynomial of the smallest degree, called p_0, i.e.,

$$\mathbf{W} = \{q \cdot p_0 | p \in \mathbb{P}\}.$$

24.5 Problem set 5

Problem. 1. Suppose that $\mathfrak{F}$ is a commutative collection of linear operators on a (not necessarily finite-dimensional) vector space $\mathbf{V}$. Suppose that λ is an eigenvalue for some $\mathcal{A} \in \mathfrak{F}$, and let $\mathbf{E}_{\mathcal{A}}(\lambda)$ be the corresponding eigenspace of $\mathcal{A}$. Argue that this $\mathbf{E}_{\mathcal{A}}(\lambda)$ is an invariant subspace for every operator in $\mathfrak{F}$.

Solution. 1.

Since $\mathbf{E}_{\mathcal{A}}(\lambda)$ is subspace, it suffices to show that $\mathcal{B}[\mathbf{E}_{\mathcal{A}}(\lambda)] \subset \mathbf{E}_{\mathcal{A}}(\lambda)$.

Consider $v \in \mathbf{E}_{\mathcal{A}}(\lambda)$, then $\mathcal{A}v = \lambda v$. Also, consider $\mathcal{B} \in \mathfrak{F}$. Since $\mathcal{A}, \mathcal{B} \in \mathfrak{F}$, $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$, so

$$\mathcal{A}\mathcal{B}v = \mathcal{B}\mathcal{A}v = \mathcal{B}(\lambda v) = \lambda \mathcal{B}v.$$

Therefore, $\mathcal{B}v$ is a λ-eigenvector of $\mathcal{A}$, i.e., $\mathcal{B}(v) \in \mathbf{E}_{\mathcal{A}}(\lambda)$ for any $v \in \mathbf{E}_{\mathcal{A}}(\lambda)$. So, $\mathcal{B}[\mathbf{E}_{\mathcal{A}}(\lambda)] \subseteq \mathbf{E}_{\mathcal{A}}(\lambda)$. Hence $\mathbf{E}_{\mathcal{A}}(\lambda)$ is an invariant subspace for any $\mathcal{B} \in \mathfrak{F}$. $\square$

Problem. 2. Let us use the same set-up as in Problem 1, and let $\mathbf{W}$ be a subspace of $\mathbf{V}$ such that

$$\mathbf{V} = \mathbf{E}_{\mathcal{A}}(\lambda) \oplus \mathbf{W}.$$

Argue that with respect to this decomposition, operators in $\mathfrak{F}$ have block matrix representation of the form

$$\begin{bmatrix} \mathcal{L} & \mathcal{M} \\ \mathcal{O} & \mathcal{K} \end{bmatrix},$$

where the $\mathcal{L}$'s form a commutative family in $\mathfrak{L}(\mathbf{E}_{\mathcal{A}}(\lambda), \mathbf{E}_{\mathcal{A}}(\lambda))$, and the $\mathcal{K}$'s form a commutative family in $\mathfrak{L}(\mathbf{W}, \mathbf{W})$.

Solution. 2.

Let $\mathcal{B}_1, \mathcal{B}_2 \in \mathfrak{F}$ be given. With respect to the decomposition $\mathbf{V} = \mathbf{E}_{\mathcal{A}}(\lambda) \oplus \mathbf{W}$, where $\mathbf{E}_{\mathcal{A}}(\lambda) \in \mathfrak{Lat}(\mathcal{B}_1)$ and $\mathbf{E}_{\mathcal{A}}(\lambda) \in \mathfrak{Lat}(\mathcal{B}_2)$, the block matrix representations of $\mathcal{B}_1, \mathcal{B}_2$ have the form

$$\begin{bmatrix} \mathcal{L}_1 & \mathcal{M}_1 \\ \mathcal{O}_1 & \mathcal{K}_1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathcal{L}_2 & \mathcal{M}_2 \\ \mathcal{O}_2 & \mathcal{K}_2 \end{bmatrix},$$

respectively, where

$$\begin{aligned} \mathcal{L}_j &: \mathbf{E}_{\mathcal{A}}(\lambda) \overset{\text{linear}}{\longrightarrow} \mathbf{E}_{\mathcal{A}}(\lambda) \\ \mathcal{K}_j &: \mathbf{W} \overset{\text{linear}}{\longrightarrow} \mathbf{W}. \end{aligned}$$

Since $\mathcal{B}_1, \mathcal{B}_2 \in \mathfrak{F}$, they commute, i.e.,

$$\begin{aligned} \begin{bmatrix} \mathcal{L}_1 & \mathcal{M}_1 \\ \mathcal{O}_1 & \mathcal{K}_1 \end{bmatrix} \begin{bmatrix} \mathcal{L}_2 & \mathcal{M}_2 \\ \mathcal{O}_2 & \mathcal{K}_2 \end{bmatrix} &= \begin{bmatrix} \mathcal{L}_1\mathcal{L}_2 & \square \\ \mathcal{O} & \mathcal{K}_1\mathcal{K}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{L}_2\mathcal{L}_1 & \triangle \\ \mathcal{O} & \mathcal{K}_2\mathcal{K}_1 \end{bmatrix} = \begin{bmatrix} \mathcal{L}_2 & \mathcal{M}_2 \\ \mathcal{O}_2 & \mathcal{K}_2 \end{bmatrix} \begin{bmatrix} \mathcal{L}_1 & \mathcal{M}_1 \\ \mathcal{O}_1 & \mathcal{K}_1 \end{bmatrix}. \end{aligned}$$

Therefore, it is necessary that $\mathcal{L}_1, \mathcal{L}_2$ commute and $\mathcal{K}_1, \mathcal{K}_2$ commute. Since this holds for any choice of $\mathcal{B}_1, \mathcal{B}_2 \in \mathfrak{F}$, the $\mathcal{L}$'s form a commutative family in $\mathfrak{L}\left(\mathbf{E}_{\mathcal{A}}(\lambda), \mathbf{E}_{\mathcal{A}}(\lambda)\right)$, and the $\mathcal{K}$'s form a commutative family in $\mathfrak{L}\left(\mathbf{W}, \mathbf{W}\right)$. $\square$

Problem. 3. Commuting collections of complex matrices are simultaneously $\triangle$-able
Argue that every commutative family of operators on a finite-dimensional vector space over the complex numbers is simultaneously upper-triangularizable, and simultaneously lower-triangularizable.

Solution. 3. Let $\mathfrak{F}$, a commuting collection of complex matrices on $\mathbb{C}^n$, be given. We notice that if $n = 1$ then any complex matrix in $\mathbb{M}_{1\times 1}$ is automatically triangular, so any such $\mathfrak{F}$ on $\mathbb{C}$ is triangularizable. In order for the statement above to fail, $n \geq 2$. Let us consider $n_0 \geq 2$ the smallest integer for which this statement fails.

Suppose that λ is an eigenvalue for some matrix $\mathcal{A} \in \mathfrak{F}$, and let $\mathbf{E}_{\mathcal{A}}(\lambda)$ be the corresponding eigenspace of $\mathcal{A}$. By Problem 2, with respect to the decomposition $\mathbf{V} = \mathbf{E}_{\mathcal{A}}(\lambda) \oplus \mathbf{W}$ where $\mathbf{W} \prec \mathbf{V}$, any $\mathcal{B} \in \mathfrak{F}$ has a block-matrix representation of the form

$$[\mathcal{B}] = \begin{bmatrix} \mathcal{L} & \mathcal{M} \\ \mathcal{O} & \mathcal{K} \end{bmatrix},$$

where the $\mathcal{L}$'s form a commutative family in $\mathfrak{L}(\mathbf{E}_{\mathcal{A}}(\lambda), \mathbf{E}_{\mathcal{A}}(\lambda))$ and the $\mathcal{K}$'s form a commutative family in $\mathfrak{L}(\mathbf{W}, \mathbf{W})$. Since the $\mathcal{L}$'s and $\mathcal{K}$'s are matrices on $\mathbb{C}^a, \mathbb{C}^b$ respectively where $a, b < n_0$, the statement above implies that both $\mathfrak{L}(\mathbf{E}_{\mathcal{A}}(\lambda), \mathbf{E}_{\mathcal{A}}(\lambda))$ and $\mathfrak{L}(\mathbf{W}, \mathbf{W})$ are simultaneously upper triangularizable. This means there exist a basis Γ_1 for $\mathbf{E}_{\mathcal{A}}(\lambda)$ with respect to which $[\mathcal{L}]_{\Gamma_1 \leftarrow \Gamma_1}$ of $\mathcal{L} \in \mathfrak{L}(\mathbf{E}_{\mathcal{A}}(\lambda), \mathbf{E}_{\mathcal{A}}(\lambda))$ is upper triangular, and a basis Γ_2 for $\mathbf{W}$ with respect to which $[\mathcal{K}]_{\Gamma_2 \leftarrow \Gamma_2}$ of $\mathcal{K} \in \mathfrak{L}(\mathbf{W}, \mathbf{W})$ is upper triangular.

Since $\mathbf{V} = \mathbf{E}_{\mathcal{A}}(\lambda) \oplus \mathbf{W}$, the concatenation $\Gamma = \Gamma_1 || \Gamma_2$ is a basis for $\mathbf{V}$ with respect to which both $[\mathcal{L}]_{\Gamma_1 \leftarrow \Gamma_1}$ and $[\mathcal{K}]_{\Gamma_2 \leftarrow \Gamma_2}$ are upper-triangular for any $\mathcal{L} \in \mathfrak{L}(\mathbf{E}_{\mathcal{A}}(\lambda), \mathbf{E}_{\mathcal{A}}(\lambda)), \mathcal{K} \in \mathfrak{L}(\mathbf{W}, \mathbf{W})$. It follows that with respect to this basis, any (block) matrix $[\mathcal{B}]_{\Gamma \leftarrow \Gamma}$ is upper triangular, and thus any $\mathcal{B} \in \mathfrak{F}$ is similar to an upper triangular matrix, i.e., upper triangularizable. Hence, $\mathfrak{F}$ on $\mathbb{C}^{n_0}$ is simultaneously upper triangularizable.

However, this contradicts the assumption that the statement fails at $\mathbb{C}^{n_0}$. By contraposition, every commutative family of operators on a finite-dimensional vector space over the complex numbers must be simultaneously upper-triangularizable.

To show that the "lower triangularizable" case also holds, we reproduce a similar argument as above, starting with writing $[\mathcal{B}]$ with respect to the decomposition $\mathbf{V} = \mathbf{W} \oplus \mathbf{E}_{\mathcal{A}}(\lambda)$ as

$$[\mathcal{B}] = \begin{bmatrix} \mathcal{L} & \mathcal{O} \\ \mathcal{M} & \mathcal{K} \end{bmatrix}$$

then showing (by induction and contraposition) that $\mathfrak{L}(\mathbf{E}_{\mathcal{A}}(\lambda), \mathbf{E}_{\mathcal{A}}(\lambda)), \mathfrak{L}(\mathbf{W}, \mathbf{W})$ both being simultaneously lower triangularizable implies $\mathfrak{F}$ simultaneously lower triangularizable. $\square$

Problem. 4. Transpose of a square matrix is similar to that matrix
Use Jordan Canonical Form Theorem to argue that the transpose $\mathcal{J}_{\lambda,n}^{\top}$ of a Jordan block $\mathcal{J}_{\lambda,n}$ is similar to $\mathcal{J}_{\lambda,n}$, and then use this fact to argue that every $k \times k$ complex matrix is similar to its transpose.

Solution. 4.

1. <u>To show</u>: $\mathcal{J}_{\lambda,n}^{\top} \sim \mathcal{J}_{\lambda,n}$. By Jordan Canonical Form Theorem, $\mathcal{J}_{\lambda,n}^{\top} \sim \bigoplus_{i=1}^{k} \mathcal{J}_{\lambda_i,m_i}$ for some k, and $\sum_{i=1}^{k} m_i = n$. But since the only eigenvalue of $\mathcal{J}_{\lambda,n}^{\top}$ is λ ($\mathcal{J}_{\lambda,n}^{\top}$ is lower triangular and has only λ's on the diagonal), $\mathcal{J}_{\lambda,n}^{\top} \sim \bigoplus_{i=1}^{k} \mathcal{J}_{\lambda,m_i}$. Next, we observe that

$$\mathcal{J}_{\lambda,n}^{\top} - \lambda\mathcal{I} = \begin{bmatrix} 0 & & & \\ 1 & & & \\ & \ddots & & \\ & & 1 & 0 \end{bmatrix}$$

is a nilpotent of order n, which implies $(\mathcal{J}_{\lambda,n}^{\top} - \lambda\mathcal{I})^{l} = \mathcal{O}$ only if $l \geq n$. On the other hand,

$$\begin{aligned}
(\mathcal{J}_{\lambda,n}^{\top} - \lambda\mathcal{I})^{\max(m_i)} &= \left(\bigoplus_{i=1}^{k} \mathcal{J}_{\lambda,m_i} - \lambda\mathcal{I}\right)^{\max(m_i)} \\
&= \left(\bigoplus_{i=1}^{k} \mathcal{J}_{0,m_i}\right)^{\max(m_i)} \\
&= \bigoplus_{i=1}^{k} (\mathcal{J}_{0,m_i})^{\max(m_i)} \\
&= \bigoplus_{i=1}^{k} (\mathcal{N}_{m_i})^{\max(m_i)} \\
&= \mathcal{O},
\end{aligned}$$

where each $\mathcal{N}_{m_i}$ is a nilpotent of order $m_i \leq n$, for any $i \in \{1, 2, \ldots, k\}$. Thus, $\max(m_i) \leq n \leq \max(m_i)$, which holds if and only if $\max(m_i) = n$. This means $\mathcal{J}_{\lambda,\max(m_i)}$ has size n, which implies it is the only summand in the direct sum with non-zero size. So,

$$\mathcal{J}_{\lambda,n}^{\top} \sim \bigoplus_{i=1}^{k} \mathcal{J}_{\lambda,m_i} = \mathcal{J}_{\lambda,n} \iff \mathcal{J}_{\lambda,n}^{\top} \text{ is similar to } \mathcal{J}_{\lambda,n}.$$

2. <u>To show</u>: $\mathcal{A} \sim \mathcal{A}^{\top}$ for any complex $k \times k$ matrix $\mathcal{A}$. Let $\mathcal{J} \sim \mathcal{A} = \mathcal{B}^{-1}\mathcal{J}\mathcal{B}$ be the Jordan form of $\mathcal{A}$. From the previous part, $\mathcal{J}_{\lambda,n}^{\top} \sim \mathcal{J}_{\lambda,n}$, so

$$\mathcal{J} = \bigoplus_{i=1}^{k} \mathcal{J}_{\lambda_i,m_i} = \bigoplus_{i=1}^{k} \mathcal{F}_i^{-1} \mathcal{J}_{\lambda_i,m_i}^{\top} \mathcal{F}_i = \bigoplus_{i=1}^{k} \mathcal{F}_i^{-1} \bigoplus_{i=1}^{k} \mathcal{J}_i^{\top} \bigoplus_{i=1}^{k} \mathcal{F}_i = \mathcal{F}^{-1}\mathcal{J}^{\top}\mathcal{F}.$$

So, we have $\mathcal{J} \sim \mathcal{J}^\top = (\mathcal{B}\mathcal{A}\mathcal{B}^{-1})^\top = (\mathcal{B}^{-1})^\top \mathcal{A}^\top \mathcal{B}^\top = (\mathcal{B}^\top)^{-1} \mathcal{A}^\top \mathcal{B}^\top$, which implies $\mathcal{J} \sim \mathcal{A}^\top$. Therefore, $\mathcal{A} \sim \mathcal{A}^\top$ by transitivity. $\square$

Problem. 5.

1. Argue that

$$\lim_{n\to\infty}\left(5\alpha^n+4n\alpha^{n-1}+3\frac{n(n-1)}{2!}\alpha^{n-2}+2\frac{n(n-1)(n-2)}{3!}\alpha^{n-3}\right.$$
$$\left.+\frac{n(n-1)(n-2)(n-3)}{4!}\alpha^{n-4}\right)=\begin{cases}0, & \text{if } 0\le\alpha<1\\ \infty, & \text{if } \alpha\ge 1\end{cases}.$$

2. Argue that for $m\ge 2$

$$\lim_{n\to\infty}\|(\mathcal{J}_{\lambda,n})^n\|_2=\begin{cases}0, & \text{if } |\lambda|<1\\ \infty, & \text{if } |\lambda|\ge 1\end{cases}$$

Note that $\mathcal{J}_{\lambda,m}=\lambda\mathcal{I}+\mathcal{N}$, where $\mathcal{N}$ is a nice cyclic nilpotent of order m, so that

$$(\mathcal{J}_{\lambda,m})^n=(\lambda\mathcal{I}+\mathcal{N})^n=\lambda^n\mathcal{I}+\binom{n}{1}\lambda^{n-1}\mathcal{N}+\binom{n}{2}\lambda^{n-2}\mathcal{N}^2+\ldots$$

You may want to start with small m first, and calculate some $(\mathcal{J}_{\lambda,m})^n$'s using *Mathematica*...

Solution. 5.

1. For $\alpha=0$,

$$\lim_{n\to\infty}\left(5\alpha^n+4n\alpha^{n-1}+3\frac{n(n-1)}{2!}\alpha^{n-2}+2\frac{n(n-1)(n-2)}{3!}\alpha^{n-3}+\frac{n(n-1)(n-2)(n-3)}{4!}\alpha^{n-4}\right)$$
$$=\lim_{n\to\infty}(0)=0.$$

For $0<\alpha<1$, for $\beta=1/\alpha$, $\beta>1$

$$\lim_{n\to\infty}\left(5\alpha^n+4n\alpha^{n-1}+\frac{3n(n-1)}{2!}\alpha^{n-2}+\frac{2n(n-1)(n-2)}{3!}\alpha^{n-3}+\frac{n(n-1)(n-2)(n-3)}{4!}\alpha^{n-4}\right)$$
$$=\lim_{n\to\infty}\left(\frac{5}{\beta^n}+\frac{4n}{\beta^{n-1}}+\frac{3n(n-1)}{2!\beta^{n-2}}+\frac{2n(n-1)(n-2)}{3!\beta^{n-3}}+\frac{n(n-1)(n-2)(n-3)}{4!\beta^{n-4}}\right)$$

We can rewrite this limit as a sum whose summands are of the form

$$\lim_{n\to\infty}\frac{p(n)}{\beta^{n-j}}=\lim_{n\to\infty}\frac{p(n)\beta^j}{\beta^n}$$

where j,β are fixed, $\beta>1$, and $p(n)$ is a polynomial in n with fixed powers. From Fact 0.5, each of these summand is zero. So, the limit being evaluated is zero. Hence, for $0\le\alpha<1$, the limit is zero.

For $\alpha \geq 1$,

$$\lim_{n\to\infty}\left(5\alpha^n + 4n\alpha^{n-1} + \frac{3n(n-1)}{2!}\alpha^{n-2} + \frac{2n(n-1)(n-2)}{3!}\alpha^{n-3} + \frac{n(n-1)(n-2)(n-3)}{4!}\alpha^{n-4}\right)$$
$$= \lim_{n\to\infty}\alpha^n\left(5 + \frac{4n}{\alpha} + \frac{3n(n-1)}{2!\alpha^2} + \frac{2n(n-1)(n-2)}{3!\alpha^3} + \frac{n(n-1)(n-2)(n-3)}{4!\alpha^4}\right)$$
$$= \lim_{n\to\infty}\alpha^n q(n), \qquad q(n) \text{ is the polynomial in } n \text{ from the line above.}$$

Let us consider $q(n)$. For $n \notin \{0, 1, 2, 3\}$,

$$q(n) = n(n-1)(n-2)(n-3)\left(\frac{5}{n(n-1)(n-2)(n-3)} + \frac{4}{(n-1)(n-2)(n-3)\alpha}\right.$$
$$\left. + \frac{3}{2!(n-2)(n-3)\alpha^2} + \frac{2}{3!(n-3)\alpha^3} + \frac{1}{4!\alpha^4}\right) = n(n-1)(n-2)(n-3)r(n).$$

So the limit we are evaluating becomes

$$\lim_{n\to\infty}\alpha^n \cdot n(n-1)(n-2)(n-3) \cdot r(n).$$

If $\alpha = 1$, then this limit is $\lim\limits_{n\to\infty} n(n-1)(n-2)(n-3) \cdot r(n)$. But since

$$\lim_{n\to\infty} n(n-1)(n-2)(n-3) = \infty$$
$$\lim_{n\to\infty} r(n) = \frac{1}{4!\alpha^4},$$

we have

$$\lim_{n\to\infty}\alpha^n \cdot n(n-1)(n-2)(n-3) \cdot r(n) = \infty.$$

If $\alpha > 1$, then $\lim\limits_{n\to\infty}\alpha^n = \infty$, so again,

$$\lim_{n\to\infty}\alpha^n \cdot n(n-1)(n-2)(n-3) \cdot r(n) = \infty.$$

Hence for $\alpha \geq 1$, the limit we are evaluating is ∞. $\square$

2. We first observe that the sum

$$(\mathcal{J}_{\lambda,m})^n = (\lambda\mathcal{I} + \mathcal{N})^n = \lambda^n\mathcal{I} + \binom{n}{1}\lambda^{n-1}\mathcal{N} + \binom{n}{2}\lambda^{n-2}\mathcal{N}^2 + \dots$$

is truncated at the term with $\mathcal{N}^m = \mathcal{O}$, since $\mathcal{N}$ is a nilpotent with order m. Furthermore, we recognize that this sum can be written as

$$(\mathcal{J}_{\lambda,m})^n = \begin{bmatrix} \lambda^n & \binom{n}{1}\lambda^{n-1} & & \binom{n}{m-1}\lambda^{n-(m-1)} \\ & \lambda^n & \ddots & \\ & & \ddots & \binom{n}{1}\lambda^{n-1} \\ & & & \lambda^n \end{bmatrix}$$

where the coefficients on the diagonal come only from the term with $\mathcal{I}$, the coefficients just above the diagonal come only from the term with $\mathcal{N}$, the coefficients two "diagonals" above the diagonal come only from the term with $\mathcal{N}^2$, and so on. This is because $\mathcal{N}$ has the form

$$\mathcal{N} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix},$$

and different non-zero powers of $\mathcal{N}$ have only 1's on different "diagonals."

So, by counting and collecting like terms, we can rewrite the norm as

$$\begin{aligned} \|(\mathcal{J}_{\lambda,m})^n\|_2 &= \left(m|\lambda^n|^2 + (m-1)\binom{n}{1}^2 \left|\lambda^{n-1}\right|^2 + \dots + \binom{n}{m-1}^2 \left|\lambda^{n-(m-1)}\right|^2\right)^{\frac{1}{2}} \\ &= \left(m|\lambda|^{2n} + (m-1)\binom{n}{1}^2 |\lambda|^{2(n-1)} + \dots + \binom{n}{m-1}^2 |\lambda|^{2(n-(m-1))}\right)^{\frac{1}{2}} \\ &= \left(m\left(|\lambda|^2\right)^n + (m-1)\binom{n}{1}^2 (|\lambda|^2)^{(n-1)} + \dots + \binom{n}{m-1}^2 (|\lambda|^2)^{(n-(m-1))}\right)^{\frac{1}{2}} \\ &\triangleq \sqrt{f(|\lambda|^2, n)}, \end{aligned}$$

where $f(|\lambda|^2, n)$ is a polynomial of (variable) degree n in $|\lambda|^2$. If $|\lambda| = |\lambda|^2 = 0$ then $f(|\lambda|^2, n) = 0$. Thus

$$\lim_{n\to\infty} \|(\mathcal{J}_{\lambda,m})^n\|_2 = \lim_{n\to\infty} \sqrt{0} = 0.$$

If $|\lambda| < 1$ then $|\lambda|^2 < 1$. And if $0 < |\lambda|^2 < 1$ then by following a similar argument as in Part 1., $\lim_{n\to\infty} f(|\lambda|^2, n) = 0$. Thus

$$\lim_{n\to\infty} \|(\mathcal{J}_{\lambda,m})^n\|_2 = \lim_{n\to\infty} \sqrt{\left(f(|\lambda|^2, n)\right)} = 0$$

Therefore, $\lim_{n\to\infty} \|(\mathcal{J}_{\lambda,m})^n\|_2 = 0$ for $|\lambda| < 1$.

If $|\lambda| \geq 1$ then $|\lambda|^2 \geq 1$. Again, by following a similar argument as in Part 1., we get $\lim_{n\to\infty} \left(f(|\lambda|^2, n)\right) = \infty$, thus

$$\lim_{n\to\infty} \|(\mathcal{J}_{\lambda,m})^n\|_2 = \lim_{n\to\infty} \sqrt{\left(f(|\lambda|^2, n)\right)} = \infty$$

for $|\lambda| \geq 1$. Note that if $m = 1$ then for $|\lambda| = 1$, $\lim_{n\to\infty} \|(\mathcal{J}_{\lambda,m})^n\|_2 = 1$. $\square$

Problem. 6.

1. Use logarithms and L'Hopital's Rule to argue that for any $\alpha > 0$,

$$\lim_{x\to\infty}\left(5+4x\alpha^{-1}+3\frac{x(x-1)}{2!}\alpha^{-2}+2\frac{x(x-1)(x-2)}{3!}\alpha^{-3}+\frac{x(x-1)(x-2)(x-3)}{4!}\alpha^{-4}\right)^{\frac{1}{x}}=1.$$

2. Argue that

$$\lim_{n\to\infty}\left(\|(\mathcal{J}_{\lambda,m})^n\|_2\right)^{\frac{1}{n}}=|\lambda|.$$

Solution. 6.

1. Let

$$f(x)=5+4x\alpha^{-1}+3\frac{x(x-1)}{2!}\alpha^{-2}+2\frac{x(x-1)(x-2)}{3!}\alpha^{-3}+\frac{x(x-1)(x-2)(x-3)}{4!}\alpha^{-4}.$$

It suffices to show

$$\lim_{x\to\infty}\ln\left((f(x))^{\frac{1}{x}}\right)=\lim_{x\to\infty}\frac{1}{x}\ln(f(x))=\ln(1)=0.$$

By l'Hopital's rule:

$$\lim_{x\to\infty}\frac{1}{x}\ln(f(x))=\lim_{x\to\infty}\frac{\frac{d}{dx}\ln(f(x))}{\frac{dx}{dx}}=\lim_{x\to\infty}\frac{d}{dx}\ln(f(x))=\lim_{x\to\infty}\frac{f'(x)}{f(x)}.$$

Since $f'(x)$ is a polynomial in x of degree 3, while $f(x)$ is a polynomial in x of degree 4,

$$\lim_{x\to\infty}\frac{f'(x)}{f(x)}=\lim_{x\to\infty}\frac{1}{x}=0.$$

Thus,

$$\lim_{x\to\infty}\ln\left((f(x))^{\frac{1}{x}}\right)=0,$$

i.e.,

$$\lim_{x\to\infty}(f(x))^{\frac{1}{x}}=1.$$

$\square$

2. From the previous problem, we have

$$\|(\mathcal{J}_{\lambda,m})^n\|_2 = \left(m(|\lambda|^2)^n + (m-1)\binom{n}{1}^2(|\lambda|^2)^{(n-1)} + \cdots + \binom{n}{m-1}^2(|\lambda|^2)^{(n-(m-1))}\right)^{\frac{1}{2}}.$$

If $|\lambda| = 0$ then

$$\|(\mathcal{J}_{\lambda,m})^n\|_2 = 0,$$

thus

$$\lim_{n\to\infty}\left(\|(\mathcal{J}_{\lambda,m})^n\|_2\right)^{\frac{1}{n}} = \lim_{n\to\infty} 0 = 0 = |\lambda|.$$

If $|\lambda| > 0$, by factoring out $|\lambda|^n$, we get

$$\|(\mathcal{J}_{\lambda,m})^n\|_2 = |\lambda|^n\left(m + \frac{(m-1)\binom{n}{1}^2}{|\lambda|^2} + \cdots + \frac{\binom{n}{m-1}^2}{|\lambda|^{2(m-1)}}\right)^{\frac{1}{2}}.$$

Therefore,

$$\begin{aligned}\left(\|(\mathcal{J}_{\lambda,m})^n\|_2\right)^{\frac{1}{n}} &= |\lambda|\left[\left(m + \frac{(m-1)\binom{n}{1}^2}{|\lambda|^2} + \cdots + \frac{\binom{n}{m-1}^2}{|\lambda|^{2(m-1)}}\right)^{\frac{1}{2}}\right]^{\frac{1}{n}}\\ &= |\lambda|\left[\left(m + \frac{(m-1)\binom{n}{1}^2}{|\lambda|^2} + \cdots + \frac{\binom{n}{m-1}^2}{|\lambda|^{2(m-1)}}\right)^{\frac{1}{n}}\right]^{\frac{1}{2}}.\end{aligned}$$

Let

$$f(n) = m + \frac{(m-1)\binom{n}{1}^2}{|\lambda|^2} + \cdots + \frac{\binom{n}{m-1}^2}{|\lambda|^{2(m-1)}}.$$

We recognize that $f(n)$ is a polynomial in n, similar to $f(x)$ in Part 1., where we have argued that $\lim_{x\to\infty}(f(x))^{1/x} = 0$. Thus,

$$\lim_{n\to\infty}(f(n))^{\frac{1}{n}} = 1,$$

which implies

$$\lim_{n\to\infty}\sqrt{(f(n))^{\frac{1}{n}}} = 1$$

It follows that

$$\lim_{n\to\infty}\left(\|(\mathcal{J}_{\lambda,m})^n\|_2\right)^{\frac{1}{n}} = |\lambda| \cdot \lim_{n\to\infty}\sqrt{(f(n))^{\frac{1}{n}}} = |\lambda| \cdot 1 = |\lambda|.$$

$\square$

Problem. Extra credit 1

1. Suppose that $[x_n], [y_n], [z_n], [u_n]$ are sequences of positive numbers such that

$$\begin{aligned}\left[(x_n)^{\frac{1}{n}}\right] &\longrightarrow \alpha\\ \left[(y_n)^{\frac{1}{n}}\right] &\longrightarrow \beta\\ \left[(z_n)^{\frac{1}{n}}\right] &\longrightarrow \gamma\\ \left[(u_n)^{\frac{1}{n}}\right] &\longrightarrow \delta.\end{aligned}$$

   (a) Evaluate the limit of

$$\left[(\alpha^n + \beta^n + \gamma^n + \delta^n)^{\frac{1}{n}}\right].$$

   (b) Evaluate the limit of

$$\left[(x_n + y_n + z_n + u_n)^{\frac{1}{n}}\right].$$

2. Suppose that $\mathcal{A} = \bigoplus_{i=1}^{23} \mathcal{J}_{\lambda_i, m_i}$. Evaluate the limit

$$\lim_{n\to\infty} \left(\|\mathcal{A}^n\|_2\right)^{\frac{1}{n}}.$$

Solution.

1. (a) Without loss of generality, let

$$\max(\alpha, \beta, \gamma, \delta) = \delta.$$

   Thus, for some $0 < j, k, l < 1$, we can write

$$\begin{aligned}\alpha &= j\delta\\ \beta &= k\delta\\ \gamma &= l\delta,\end{aligned}$$

   so that

$$\begin{aligned}(\alpha^n + \beta^n + \gamma^n + \delta^n)^{\frac{1}{n}} &= ((j\delta)^n + (k\delta)^n + (l\delta)^n + \delta^n)^{\frac{1}{n}}\\ &= \delta\,(j^n + k^n + l^n + 1)^{\frac{1}{n}}.\end{aligned}$$

   Since $0 < j < 1$, $\lim_{n\to\infty} j^n = 0$ (this follows from Fact 0.5, by writing j as $1/j^{-1}$ where $j^{-1} > 1$). Repeating this argument for k, l, we get

$$\lim_{n\to\infty} (j^n + k^n + l^n + 1) = 0 + 0 + 0 + 1 = 1. \qquad (\dagger)$$

Consider

$$\Lambda(n) = \delta \left(j^n + k^n + l^n + 1 \right)^{\frac{1}{n}}.$$

Since $\Lambda(n) > 0$

$$\begin{aligned}
\frac{\Lambda(n)}{\delta} &= \exp\left[\ln\left(\left(j^n + k^n + l^n + 1\right)^{\frac{1}{n}}\right)\right] \\
&= \exp\left[\frac{1}{n}\ln(j^n + k^n + l^n + 1)\right].
\end{aligned}$$

By Eq. (†), and the fact that $\lim_{n\to\infty} 1/n = 0$

$$\lim_{n\to\infty} \frac{1}{n} \ln\left(j^n + k^n + l^n + 1\right) = 0.$$

Thus,

$$\lim_{n\to\infty} \frac{\Lambda(n)}{\delta} = e^0 = 1,$$

i.e.,

$$\lim_{n\to\infty} \Lambda(n) = \lim_{n\to\infty} \delta \left(j^n + k^n + l^n + 1 \right)^{\frac{1}{n}} = \delta.$$

And so,

$$\lim_{n\to\infty} \left(\alpha^n + \beta^n + \gamma^n + \delta^n \right)^{\frac{1}{n}} = \lim_{n\to\infty} \delta \left(j^n + k^n + l^n + 1 \right)^{\frac{1}{n}} = \delta.$$

Therefore, the limit of $[(\alpha^n + \beta^n + \gamma^n + \delta^n)^{1/n}]$ is $\max(\alpha, \beta, \gamma, \delta)$. $\square$

(b) Let

$$\begin{aligned}
X_n &= (x_n)^{\frac{1}{n}} \\
Y_n &= (y_n)^{\frac{1}{n}} \\
Z_n &= (z_n)^{\frac{1}{n}} \\
U_n &= (u_n)^{\frac{1}{n}}.
\end{aligned}$$

Then

$$\begin{aligned}
[X_n] &\longrightarrow \alpha \\
[Y_n] &\longrightarrow \beta \\
[Z_n] &\longrightarrow \gamma \\
[U_n] &\longrightarrow \delta
\end{aligned}$$

and

$$\begin{aligned}(X_n)^n &= x_n\\(Y_n)^n &= y_n\\(Z_n)^n &= z_n\\(U_n)^n &= u_n.\end{aligned}$$

Then,

$$\Lambda(n) = (x_n + y_n + z_n + u_n)^{\frac{1}{n}} = ((X_n)^n + (Y_n)^n + (Z_n)^n + (U_n)^n)^{\frac{1}{n}}.$$

Without loss of generality, suppose $\max(X_n, Y_n, Z_n, U_n) = U_n$ as $n \to \infty$, then for $0 < a, b, c < 1$ we can write

$$\lim_{n\to\infty} \Lambda(n) = \lim_{n\to\infty} U_n \left(a^n + b^n + c^n + 1\right)^{\frac{1}{n}},$$

where $aX_n = bY_n = cZ_n = U_n$ as $n \to \infty$. From Part 1., we have argued that for such a, b, c,

$$\lim_{n\to\infty} \left(a^n + b^n + c^n + 1\right)^{\frac{1}{n}} = 1.$$

Thus, if $\max(X_n, Y_n, Z_n, U_n) = U_n$ as $n \to \infty$, i.e., $\max(\alpha, \beta, \gamma, \delta) = \delta$, then

$$\lim_{n\to\infty} \Lambda(n) = \lim_{n\to\infty} U_n = \delta.$$

Therefore, the limit of $\left[(x_n + y_n + z_n + u_n)^{\frac{1}{n}}\right]$ is $\max(\alpha, \beta, \gamma, \delta)$. $\square$

2. Let $\mathcal{A} = \bigoplus_{i=1}^{23} \mathcal{J}_{\lambda_i, m_i}$ be given, then

$$\begin{aligned}\lim_{n\to\infty} \left(\|\mathcal{A}^n\|_2\right)^{\frac{1}{n}} &= \lim_{n\to\infty} \left(\left\|\left(\bigoplus_{i=1}^{23} \mathcal{J}_{\lambda_i, m_i}\right)^n\right\|_2\right)^{\frac{1}{n}}\\&= \lim_{n\to\infty} \left(\left\|\bigoplus_{i=1}^{23} (\mathcal{J}_{\lambda_i, m_i})^n\right\|_2\right)^{\frac{1}{n}}\\&= \lim_{n\to\infty} \left(\sqrt{\sum_{i=1}^{23} \left(\|(\mathcal{J}_{\lambda_i, m_i})^n\|_2\right)^2}\right)^{\frac{1}{n}}\\&= \lim_{n\to\infty} \left(\left(\sum_{i=1}^{23} \left(\|(\mathcal{J}_{\lambda_i, m_i})^n\|_2\right)^2\right)^{\frac{1}{n}}\right)^{\frac{1}{2}}\end{aligned}$$

From Problem 6., we know that

$$\lim_{n\to\infty} \left(\|(\mathcal{J}_{\lambda_i, m_i})^n\|_2\right)^{\frac{1}{n}} = |\lambda_i|.$$

So,

$$\lim_{n\to\infty}\left(\|(\mathcal{J}_{\lambda_i,m_i})^n\|_2\right)^{\frac{2}{n}} = \lim_{n\to\infty}\left(\left(\|(\mathcal{J}_{\lambda_i,m_i})^n\|_2\right)^2\right)^{\frac{1}{n}} = |\lambda_i|^2.$$

If $\left\|\left(\mathcal{J}_{\lambda_j,m_j}\right)^n\right\|_2$ is zero for some j, then (i) $\lambda_j = 0$, and (ii) we can drop this term from the direct sum of operators (to $\mathcal{A}$). Then, we can treat the positive $\left(\|(\mathcal{J}_{\lambda_i,m_i})^n\|_2\right)^2$'s as elements of the sequences $\left[\left(\|(\mathcal{J}_{\lambda_i,m_i})^n\|_2\right)^2\right]$, each converges to a corresponding $|\lambda_i|^2$, $i = 1, 2, \ldots, k \leq 23$. Using the result from Part 1., we get

$$\begin{aligned}\lim_{n\to\infty}\left(\|\mathcal{A}^n\|_2\right)^{\frac{1}{n}} &= \lim_{n\to\infty}\left(\left(\sum_{i=1}^{23}\left(\|(\mathcal{J}_{\lambda_i,m_i})^n\|_2\right)^2\right)^{\frac{1}{n}}\right)^{\frac{1}{2}} \\ &= \sqrt{\max(|\lambda_i|^2)} \\ &= \max(|\lambda_i|)\end{aligned}$$

24.6 Problem set 6

Problem. 1.

1. **Problem 20 in Exercises 6.A in Axler.** Suppose $\mathbf{V}$ is a complex inner product space. Prove that

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}$$

for all $u, v \in \mathbf{V}$.

Solution. 1.1 We have

$$\|u+v\|^2 = \langle u+v, u+v\rangle = \langle u, u+v\rangle + \langle v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle$$
$$\|u-v\|^2 = \langle u-v, u-v\rangle = \langle u, u-v\rangle - \langle v, u-v\rangle = \langle u,u\rangle - \langle u,v\rangle - \langle v,u\rangle + \langle v,v\rangle$$

and

$$\begin{aligned} i\|u+iv\|^2 &= i\left[\langle u, u+iv\rangle + i\langle v, u+iv\rangle\right] \\ &= i\left[\langle u,u\rangle - i\langle u,v\rangle + i\langle v,u\rangle + \langle v,v\rangle\right] \\ &= i\langle u,u\rangle + \langle u,v\rangle - \langle v,u\rangle + i\langle v,v\rangle \end{aligned}$$

and

$$\begin{aligned} -i\|u-iv\|^2 &= i\left[\langle u, u-iv\rangle - i\langle v, u-iv\rangle\right] \\ &= -i\left[\langle u,u\rangle + i\langle u,v\rangle - i\langle v,u\rangle + \langle v,v\rangle\right] \\ &= -i\langle u,u\rangle + \langle u,v\rangle - \langle v,u\rangle - \langle v,v\rangle. \end{aligned}$$

Putting these terms together, we get

$$\|u+v\|^2 - \|u-v\|^2 + i\|u+iv\|^2 - i\|u-iv\|^2 = 4\langle u,v\rangle.$$

Thus,

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}$$

as desired. $\square$

2. **Problem 21 in Exercises 6.A in Axler.** A norm on a vector space $\mathbf{U}$ is a function $\| \ \| : \mathbf{U} \to [0, \infty)$ such that $\|u\| = 0$ if and only if $u = \mathbf{0}$, $\|\alpha u\| = |\alpha| \|u\|$ for all $\alpha \in \mathbb{F}$ and all $u \in \mathbf{U}$, and $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in \mathbf{U}$. Prove that a norm satisfying the parallelogram equality comes from an inner product (in other words, show that if $\| \ \|$ is a norm on $\mathbf{U}$ satisfying the parallelogram equality, then there is an inner product $\langle , \rangle$ on $\mathbf{U}$ such that $\|u\| = \langle u, u\rangle^{1/2}$ for all $u \in \mathbf{U}$).

   Solution. 1.2 We define $\langle \cdot, \cdot \rangle : \mathbf{U} \times \mathbf{U} \to \mathbb{F}$ such that for $u, v \in \mathbf{U}$,

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}.$$

   We will show that if $\| \ \|$ is a norm on $\mathbf{U}$ satisfying the parallelogram equality

$$\|u+v\|^2 + \|u-v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right),$$

   for any $u, v \in \mathbf{U}$, then $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbf{U}$ such that $\|u\| = \langle u, u\rangle^{1/2}$ for all $u \in \mathbf{U}$. To show that $\langle \cdot, \cdot \rangle$ is an inner product, we check the following conditions:

   (a) **Positivity:**

$$\begin{aligned}
\langle u, u\rangle &= \frac{\|2u\|^2 - \|\mathbf{0}\|^2 + \|(1+i)u\|^2 i - \|(1-i)u\|^2 i}{4} \\
&= \frac{4\|u\|^2 + i|1+i|^2\|u\|^2 - i|1-i|^2\|u\|^2 i}{4} \\
&= \frac{4\|u\|^2 + 2i\|u\|^2 - 2i\|u\|^2}{4} \\
&= \|u\|^2 \geq 0.
\end{aligned}$$

   Thus, if $\langle \cdot, \cdot \rangle$ is an inner product, then it is also one that gives $\|u\| = \langle u, u\rangle^{1/2}$. □

   (b) **Definiteness:**

$$\langle u, u\rangle = \|u\|^2 = 0 \iff u = \mathbf{0}.$$

   □

   (c) **Additivity in first slot:** Let $w \in \mathbf{U}$ be given. Here we want to use the parallelogram equality to show additivity in first slot. Let $\alpha \in \mathbb{C}$

be given, then

$$
\begin{aligned}
\|u+\alpha v\|^2+\|w+\alpha v\|^2 &= \left\|\frac{2\alpha v+u+w}{2}-\frac{w-u}{2}\right\|^2+\left\|\frac{2\alpha v+u+w}{2}+\frac{w-u}{2}\right\|^2 \\
&= 2\left(\left\|\frac{2\alpha v+u+w}{2}\right\|^2+\left\|\frac{w-u}{2}\right\|^2\right) \\
&= 2\left(\left\|\alpha v+\frac{u+w}{2}\right\|^2+\left\|\frac{w-u}{2}\right\|^2\right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
4\left(\langle u,v\rangle+\langle w,v\rangle\right) &= \|u+v\|^2-\|u-v\|^2+\|u+iv\|^2 i-\|u-iv\|^2 i \\
&\quad +\|w+v\|^2-\|w-v\|^2+\|w+iv\|^2 i-\|w-iv\|^2 i \\
&= \left(\|u+v\|^2+\|w+v\|^2\right)-\left(\|u-v\|^2+\|w-v\|^2\right) \\
&\quad + i\left(\|u+iv\|^2+\|w+iv\|^2\right)-i\left(\|u-iv\|^2+\|w-iv\|^2\right) \\
&= 2\left(\left\|v+\frac{u+w}{2}\right\|^2+\left\|\frac{w-u}{2}\right\|^2\right)-2\left(\left\|\frac{u+w}{2}-v\right\|^2+\left\|\frac{w-u}{2}\right\|^2\right) \\
&\quad + 2i\left(\left\|iv+\frac{u+w}{2}\right\|^2+\left\|\frac{w-u}{2}\right\|^2\right)-2i\left(\left\|-iv+\frac{u+w}{2}\right\|^2+\left\|\frac{w-u}{2}\right\|^2\right) \\
&= 2\left(\left\|\frac{u+w}{2}+v\right\|^2-\left\|\frac{u+w}{2}-v\right\|^2+i\left\|\frac{u+w}{2}+iv\right\|^2-i\left\|\frac{u+w}{2}-iv\right\|^2\right) \\
&= 8\left\langle\frac{u+w}{2},v\right\rangle. \qquad (\dagger)
\end{aligned}
$$

Thus,

$$
\langle u,v\rangle+\langle w,v\rangle = 2\left\langle\frac{u+w}{2},v\right\rangle.
$$

But we are not done yet. Since we don't have "homogeneity in first slot" yet, we have to show $8\left\langle\frac{u+w}{2},v\right\rangle = 4\langle u+w,v\rangle$ by applying the parallelogram equality again. For $\alpha\in\mathbb{C}$:

$$
\begin{aligned}
\|u+w+\alpha v\|^2+\|\alpha v\|^2 &= \left\|\left(\frac{u+w}{2}+\alpha v\right)+\frac{u+w}{2}\right\|^2+\left\|\left(\frac{u+w}{2}+\alpha v\right)-\frac{u+w}{2}\right\|^2 \\
&= 2\left(\left\|\frac{u+w}{2}+\alpha v\right\|^2+\left\|\frac{u+w}{2}\right\|^2\right).
\end{aligned}
$$

Thus,

$$\begin{aligned}
8\left\langle \frac{u+w}{2}, v\right\rangle &= 2\left(\left\|\frac{u+w}{2}+v\right\|^2 - \left\|\frac{u+w}{2}-v\right\|^2 + i\left\|\frac{u+w}{2}+iv\right\|^2 - i\left\|\frac{u+w}{2}-iv\right\|^2\right)\\
&= 2\left(\left\|\frac{u+w}{2}+v\right\|^2 + \left\|\frac{u+w}{2}\right\|^2 - \left\|\frac{u+w}{2}-v\right\|^2 - \left\|\frac{u+w}{2}\right\|^2\right.\\
&\quad \left. +i\left\|\frac{u+w}{2}+iv\right\|^2 + i\left\|\frac{u+w}{2}\right\|^2 - i\left\|\frac{u+w}{2}-iv\right\|^2 - i\left\|\frac{u+w}{2}\right\|^2\right)\\
&= \|u+w+v\|^2 + \|u+w-v\|^2 + i\|u+w+iv\|^2 - i\|u+w-iv\|^2\\
&= 4\langle u+w, v\rangle. \qquad (\dagger\dagger)
\end{aligned}$$

Hence, from (†) and (††),

$$\langle u, v\rangle + \langle w, v\rangle = \langle u+w, v\rangle$$

as desired. $\square$

(d) **Conjugate symmetry:** We want to show $\langle u, v\rangle = \overline{\langle v, u\rangle}$.

$$\begin{aligned}
\overline{\langle v, u\rangle} &= \overline{\frac{\|v+u\|^2 - \|v-u\|^2 + \|v+iu\|^2 i - \|v-iu\|^2 i}{4}}\\
&= \frac{\overline{\|u+v\|^2} - \overline{\|u-v\|^2} + \overline{\|v+iu\|^2 i} - \overline{\|v-iu\|^2 i}}{4}\\
&= \frac{\|u+v\|^2 - \|u-v\|^2 - i\overline{\|v+iu\|^2} + i\overline{\|v-iu\|^2}}{4}\\
&= \frac{\|u+v\|^2 - \|u-v\|^2 - i\|v+iu\|^2 + i\|v-iu\|^2}{4}\\
&= \frac{\|u+v\|^2 - \|u-v\|^2 - i\|(-i)(v+iu)\|^2 + i\|(-i)(v-iu)\|^2}{4}\\
&= \frac{\|u+v\|^2 - \|u-v\|^2 - i\|u-iv\|^2 + i\|u+iv\|^2}{4}\\
&= \langle u, v\rangle.
\end{aligned}$$

$\square$

(e) **Homogeneity in first slot:** For $\lambda \in \mathbb{C}$, we want to show that $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$.

Note to Leo: I was in the lab the entire Wednesday afternoon so I couldn't come to office hours for your advice on this part of the problem. I tried using the definition of the inner product given above, but I couldn't "pull the λ out of $\langle \lambda a, b\rangle$" to show that $\lambda\langle a, b\rangle = \langle \lambda a, b\rangle$. I figured that I could try showing homogeneity in first slot holds for λ's that are zero, positive integers, positive rationals, negative rationals, real numbers, and ultimately complex numbers. I don't know if this is a good idea. But at least I make some progress this way...

i. If $\lambda = 0$, then $\langle \lambda u, v\rangle = \langle \mathbf{0}, v\rangle = 0 = 0\langle u, v\rangle = \lambda\langle u, v\rangle$.

ii. If λ is a positive integer, then $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ by additivity.

iii. If λ is a positive rational number, then let $\lambda = p/q$ where $p, q \in \mathbb{N}, p, q \neq 0$. Then

$$\langle \lambda u, v\rangle = \left\langle \frac{p}{q}u, v\right\rangle = p\left\langle \frac{1}{q}u, v\right\rangle.$$

Let $u' = \frac{1}{q}u$, then

$$\left\langle u', v\right\rangle = \left\langle \frac{1}{q}u, v\right\rangle = \left\langle \frac{q}{q}u', v\right\rangle = q\left\langle \frac{1}{q}u', v\right\rangle,$$

so,

$$\frac{1}{q}\left\langle u', v\right\rangle = \left\langle \frac{1}{q}u', v\right\rangle.$$

Therefore,

$$\langle \lambda u, v\rangle = \lambda\langle u, v\rangle.$$

iv. If $\lambda = -1$, then since $\|a\|^2 = \|-a\|^2$:

$$\begin{aligned}\langle -a, -a\rangle &= \frac{\|-a-a\|^2 - \|-a+a\|^2 + i\|a+ia\|^2 - i\|a-ia\|^2}{4}\\ &= \frac{4\|a\|^2 + 2i\|u\|^2 - 2i\|a\|^2}{4}\\ &= \|a\|^2\\ &= \langle a, a\rangle,\end{aligned}$$

we have

$$\begin{aligned}\langle -u, v\rangle &= \frac{\|-u+v\|^2 - \|-u-v\|^2 + i\|-u+iv\|^2 - i\|-u-iv\|^2}{4}\\ &= \frac{\|u-v\|^2 - \|u+v\|^2 + i\|u-iv\|^2 - i\|u+iv\|^2}{4}\\ &= -\langle u, v\rangle.\end{aligned}$$

v. If $\lambda = i$, then

$$\begin{aligned}\langle iu, v\rangle &= \frac{\|iu+v\|^2 - \|iu-v\|^2 + i\|iu+iv\|^2 - i\|iu-iv\|^2}{4}\\ &= i\left(\frac{\|u+v\|^2 - \|u-v\|^2 + i\|u+iv\|^2 - i\|u-iv\|^2}{4}\right)\\ &= i\langle u, v\rangle.\end{aligned}$$

vi. The last step is to show that $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ still holds when λ is any real number. Once this is shown, showing $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ will be much simpler.
But this is where I'm stuck. I'm hoping to somehow express λ, a real number, in terms of rational numbers, but I don't know how.

vii. Assume that we have successfully completed the previous part, then let $\lambda = a + ib$ where $a, b \in \mathbb{R}$, then

$$\begin{aligned}\langle \lambda u, v\rangle &= \langle (a+ib)u, v\rangle \\ &= \langle au, v\rangle + \langle ibu, v\rangle \\ &= a\langle u, v\rangle + b\langle iu, v\rangle \\ &= a\langle u, v\rangle + ib\langle u, v\rangle \\ &= (a+ib)\langle u, v\rangle \\ &= \lambda\langle u, v\rangle.\end{aligned}$$

$\square$

(f) We can show that *additivity in second slot* can be derived from the properties above:

$$\langle a, b+c\rangle = \overline{\langle b+c, a\rangle} = \overline{\langle b, a\rangle + \langle c, a\rangle} = \overline{\langle b, a\rangle} + \overline{\langle c, a\rangle} = \langle a, b\rangle + \langle a, c\rangle.$$

$\square$

(g) Conjugate symmetry, homogeneity in first slot, and additivity in second slot can combine to give partial conjugate linear in second slot. Let $a, b \in \mathbb{C}$ and $u, v, w \in \mathbf{U}$ be given, then

$$\begin{aligned}\langle u, av + bw\rangle &= \overline{\langle av + bw, u\rangle} \\ &= \overline{a\langle v, u\rangle + b\langle w, u\rangle} \\ &= \overline{a}\overline{\langle v, u\rangle} + +\overline{b}\langle w, u\rangle \\ &= \overline{a}\langle u, v\rangle + \overline{b}\langle w, v\rangle.\end{aligned}$$

So if we could show item (e.vi), then we would show that there exists $\langle \cdot, \cdot\rangle$ on $\mathbf{U}$ that satisfies the hypothesis. This inner product is given by

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + i\|u+iv\|^2 - i\|u-iv\|^2}{4}.$$

$\square$

Problem. 2.

1. **Problem 24 in Exercises 6.A in Axler.** Suppose $S \in \mathfrak{L}(\mathbf{V})$ is an injective operator on $\mathbf{V}$. Define $\langle \cdot, \cdot \rangle_1$ by

$$\langle u, v \rangle_1 = \langle Su, Sv \rangle$$

for all $u, v \in \mathbf{V}$. Show that $\langle \cdot, \cdot \rangle_1$ is an inner product on $\mathbf{V}$.

Solution. 2.1 We check if $\langle Su, Sv \rangle$ satisfies the conditions:

(a) **Positiveness:** Let $u \in \mathbf{V}$ be given, then $Su = v \in \mathbf{V}$ since $S \in \mathfrak{L}(\mathbf{V})$. Thus,

$$\langle u, u \rangle_1 = \langle Su, Su \rangle = \langle v, v \rangle \geq 0$$

since $\langle \cdot, \cdot \rangle$ in an inner product on $\mathbf{V}$.

(b) **Definiteness:** Let $u \in \mathbf{V}$ be given, then

$$0 = \langle u, u \rangle_1 = \langle Su, Su \rangle \iff Su = \mathbf{0} \iff u = \mathbf{0},$$

where the last equivalence statement is due to the injectivity of S.

(c) **Additivity in first slot:** Let $u, w, v \in \mathbf{V}$ be given, then

$$\begin{aligned} \langle u + w, v \rangle_1 &= \langle S(u + w), Sv \rangle \\ &= \langle Su + Sw, Sv \rangle \\ &= \langle Su, Sv \rangle + \langle Sw, Sv \rangle \\ &= \langle u, v \rangle_1 + \langle w, v \rangle_1, \end{aligned}$$

where the third equality is due to $\langle \cdot, \cdot \rangle$ being an inner product on V.

(d) **Homogeneity in first slot:**

$$\langle \alpha u, v \rangle_1 = \langle S(\alpha u), Sv \rangle = \langle \alpha Su, Sv \rangle = \alpha \langle Su, Sv \rangle = \alpha \langle u, v \rangle_1,$$

where the last equality is due to $\langle \cdot, \cdot \rangle$ being an inner product on V.

(e) **Conjugate symmetry:**

$$\overline{\langle v, u \rangle_1} = \overline{\langle Sv, Su \rangle} = \langle Su, Sv \rangle = \langle u, v \rangle_1,$$

where the second equality is due to $\langle \cdot, \cdot \rangle$ being an inner product on V.

□

2. **Problem 25 in Exercises 6.A in Axler.** Suppose $S \in \mathfrak{L}(\mathbf{V})$ is not injective. Define $\langle \cdot, \cdot \rangle_1$ as in the exercise above. Explain why $\langle \cdot, \cdot \rangle_1$ is not an inner product on $\mathbf{V}$.

Solution. 2.2 If S is not injective, then there exists a nonzero $v \in \mathbf{V}$ such that $Sv = \mathbf{0}$, so that $\langle v, v \rangle_1 = \langle Sv, Sv \rangle = 0$ but $v \neq \mathbf{0}$. Thus, such S makes $\langle \cdot, \cdot \rangle$ fail to be an inner product. □

Problem. 3.

1. **Problem 27 in Exercises 6.A in Axler.** Suppose $u, v, w \in \mathbf{V}$. Prove that

$$\left\| w - \frac{1}{2}(u+v) \right\|^2 = \frac{\|w-u\|^2 + \|w-v\|^2}{2} - \frac{\|u-v\|^2}{4}$$

Solution. 3.1 Let an inner product be defined as in Problem 21 such that $\|u\| = \langle u, u\rangle^{1/2}$. Then we have

$$\begin{aligned}\left\| w - \frac{1}{2}(u+v) \right\|^2 &= \left\langle w - \frac{1}{2}(u+v), w - \frac{1}{2}(u+v) \right\rangle \\ &= \left\langle \frac{1}{2}(w-u) + \frac{1}{2}(w-v), \frac{1}{2}(w-u) + \frac{1}{2}(w-v), \right\rangle.\end{aligned}$$

Let $a = (1/2)(w-u), b = (1/2)(w-v)$. $a, b \in \mathbf{V}$. Then

$$\begin{aligned}\left\| w - \frac{1}{2}(u+v) \right\|^2 &= \langle a+b, a+b\rangle \\ &= \langle a,a\rangle + \langle a,b\rangle + \langle b,a\rangle + \langle b,b\rangle \\ &= \langle a,a\rangle + \langle a, b-a+a\rangle + \langle b, a-b+b\rangle + \langle b,b\rangle \\ &= 2\langle a,a\rangle + \langle a, b-a\rangle + \langle b, a-b\rangle + 2\langle b,b\rangle \\ &= 2\langle a,a\rangle + \langle a, b-a\rangle + \langle -b, b-a\rangle + 2\langle b,b\rangle \\ &= 2\langle a,a\rangle + \langle a-b, b-a\rangle + 2\langle b,b\rangle \\ &= 2\left\langle \frac{1}{2}(w-u), \frac{1}{2}(w-u) \right\rangle + \left\langle \frac{1}{2}(v-u), \frac{1}{2}(u-v) \right\rangle \\ &\qquad + 2\left\langle \frac{1}{2}(w-v), \frac{1}{2}(w-v) \right\rangle \\ &= \frac{1}{2}\langle w-u, w-u\rangle - \frac{1}{4}\langle u-v, u-v\rangle + \frac{1}{2}\langle w-v, w-v\rangle \\ &= \frac{\|w-u\|^2 + \|w-v\|^2}{2} - \frac{\|u-v\|^2}{4}.\end{aligned}$$

$\square$

We can also use the parallelogram equality to solve this problem. Again, let $a = (1/2)(w-u)$ and $b = (1/2)(w-v)$, then

$$\begin{aligned}
\frac{\|w-u\|^2 + \|w-v\|^2}{2} - \frac{\|u-v\|^2}{4} &= 2\left(\|a\|^2 + \|b\|^2\right) - \|b-a\|^2 \\
&= 2\left(\left\|\frac{a+b}{2} - \frac{b-a}{2}\right\|^2 + \left\|\frac{a+b}{2} + \frac{b-a}{2}\right\|^2\right) - \|b-a\|^2 \\
&= \|a+b\|^2 + \|b-a\|^2 - \|b-a\|^2 \\
&= \|a+b\|^2 \\
&= \left\|w - \frac{1}{2}(u-v)\right\|^2.
\end{aligned}$$

$\square$

2. **Problem 28 in Exercises 6.A in Axler.** Suppose C is a subset of $\mathbf{V}$ with the property that $u, v \in C$ implies $\frac{1}{2}(u+v) \in C$. Let $w \in \mathbf{V}$. Show that there is at most one point in C that is closest to w. In other words, show that there is at most one $u \in C$ such that

$$\|w-u\| \leq \|w-v\| \quad \text{for all } v \in C.$$

Hint: Use the previous exercise.

Solution. 3.2 We prove by contradiction. Suppose there are more than one such vector in C. Let these vectors be a, b, $a \neq b$, then we have

$$\begin{aligned} \|w-a\| \leq \|w-v\| &\quad \text{for all } v \in C \\ \|w-b\| \leq \|w-v\| &\quad \text{for all } v \in C. \end{aligned}$$

This implies

$$\|w-a\| \leq \|w-b\| \text{ and } \|w-b\| \leq \|w-a\| \iff \|w-a\| = \|w-b\|.$$

Now, since $a, b \in C$, $(1/2)(a+b) \in C$. Thus, by the previous problem

$$\begin{aligned} \left\|w - \frac{1}{2}(a+b)\right\|^2 &= \frac{\|w-a\|^2 + \|w-b\|^2}{2} - \frac{\|a-b\|^2}{4} \\ &= \|w-a\|^2 - \frac{\|a-b\|^2}{4} \\ &< \|w-a\|^2. \end{aligned}$$

So,

$$\left\|w - \frac{1}{2}(a+b)\right\| < \|w-a\|.$$

Clearly, for $v = (1/2)(a+b)$, $\|w-a\| > \|w-(1/2)(a+b)\|$. This contradicts the choice of a since $\|w-a\|^2$ must be less than or equal to $\|w-v\|^2$ for *any* $v \in C$. Thus, there is at most one vector in C that meets the hypothesis. $\square$

Problem. 4.

1. Use Cauchy-Schwarz inequality to show that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n \leq \sqrt{n} \cdot \sqrt{\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2}$$

for any non-negative α_i.

Solution. 4.1 Let the vectors $\alpha, \epsilon \in \mathbb{C}^n$ be given, where $\alpha = \begin{pmatrix}\alpha_1 & \alpha_2 & \ldots & \alpha_n\end{pmatrix}^\top$ and $\epsilon = \begin{pmatrix}1 & 1 & \ldots & 1\end{pmatrix}^\top$, α_i are non-negative. Consider the standard inner product $\langle \cdot, \cdot \rangle$ on $\mathbf{V}$. Then we have

$$\|\alpha\|^2 = |\alpha_1|^2 + |\alpha_2|^2 + \cdots + |\alpha_n|^2 = \alpha_1^2 + \alpha_2^2 + \ldots \alpha_n^2$$
$$\|e\|^2 = \underbrace{1^2 + 1^2 + \ldots 1^2}_{n \text{ times}} = n$$
$$|\langle \alpha, \epsilon \rangle| = \alpha_1 \cdot 1 + \alpha_2 \cdot 1 + \cdots + \alpha_n \cdot 1 = \alpha_1 + \alpha_2 + \cdots + \alpha_n$$

Hence, the Cauchy-Schwarz inequality

$$\left|\langle \alpha, \epsilon \rangle\right| \leq \|\alpha\| \cdot \|e\|$$

implies that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n \leq \sqrt{n} \cdot \sqrt{\alpha_1^2 + \alpha_2^2 + \ldots \alpha_n^2}$$

as desired. $\square$

2. **Problem 14 in Exercises 6.B in Axler** Suppose $e_1, \dots, e_n$ is an orthonormal basis of $\mathbf{V}$ and $v_1, \dots, v_n$ are vectors in $\mathbf{V}$ such that

$$\|e_j - v_j\| < \frac{1}{\sqrt{n}}$$

for each j. Prove that $v_1, \dots, v_n$ is a basis of $\mathbf{V}$.

Solution. 4.2 To show that $v_1, \dots, v_n$ is a basis of $\mathbf{V}$, we can show $v_1, \dots, v_n$ make an linearly independent list and span $\mathbf{V}$. However, this would be redundant since if $v_1, \dots, v_n$ is linearly independent then $\dim(\text{span}(v_1, \dots, v_n)) = n = \dim(\mathbf{V})$, and because $\text{span}(v_1, \dots, v_n) \prec \mathbf{V}$, this says $\text{span}(v_1, \dots, v_n) = \mathbf{V}$, hence $v_1, \dots, v_n$ is a basis for $\mathbf{V}$. So, it suffices to show $v_1, \dots, v_n$ is a linearly independent list. We will do this by contradiction.

Suppose for some $a_j \neq 0$, $j = 1, 2, \dots, n$,

$$\sum_{j=1}^{n} a_j v_j = \mathbf{0}.$$

This means

$$\begin{aligned} \left\| \sum_{j=1}^{n} a_j (e_j - v_j) \right\|^2 &= \left\| \sum_{j=1}^{n} a_j e_j - \sum_{j=1}^{n} a_j v_j \right\|^2 \\ &= \left\| \sum_{j=1}^{n} a_j e_j \right\|^2 \\ &= \left\langle \sum_{j=1}^{n} a_j e_j, \sum_{i=1}^{n} a_i e_i \right\rangle \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} \langle a_j e_j, a_i e_i \rangle \\ &= \sum_{j=1}^{n} |a_j|^2. \end{aligned}$$

But we also have

$$\begin{aligned}
\left\|\sum_{j=1}^{n} a_j(e_j - v_j)\right\|^2 &= \underbrace{\left\langle \sum_{j=1}^{n} a_j(e_j - v_j), \sum_{i=1}^{n} a_i(e_i - v_i)\right\rangle}_{\text{real}} \\
&= \underbrace{\sum_{j=1}^{n}\sum_{i=1}^{n} \left\langle a_j(e_j - v_j), a_i(e_i - v_i)\right\rangle}_{\text{real}} \\
&\leq \sum_{j=1}^{n}\sum_{i=1}^{n} \left|\left\langle a_j(e_j - v_j), a_i(e_i - v_i)\right\rangle\right|,
\end{aligned}$$

where

$$\begin{aligned}
\left|\left\langle a_j(e_j - v_j), a_i(e_i - v_i)\right\rangle\right| &\leq \|a_j(e_j - v_j)\| \cdot \|a_i(e_i - v_i)\| \qquad \text{Cauchy-Schwarz inequality} \\
&= |a_j| \cdot |a_i| \cdot \|e_j - v_j\| \cdot \|e_i - v_i\| \\
&< |a_j| \cdot |a_i| \cdot \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \qquad \text{by hypothesis} \\
&= |a_j| \cdot |a_i| \cdot \frac{1}{n}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\left\|\sum_{j=1}^{n} a_j(e_j - v_j)\right\|^2 &< \sum_{j=1}^{n}\sum_{i=1}^{n} |a_j| \cdot |a_i| \cdot \frac{1}{n} \\
&= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} |a_j| \cdot |a_i| \\
&= \frac{1}{n}\left(\sum_{j=1}^{n} |a_j|\right)^2.
\end{aligned}$$

By the previous problem, we know that

$$\left(\sum_{j=1}^{n} |a_j|\right)^2 \leq \left(\sqrt{n} \cdot \sqrt{\sum_{j=1}^{n} |a_j|^2}\right)^2 = n \cdot \sum_{j=1}^{n} |a_j|^2.$$

Therefore,

$$\left\|\sum_{j=1}^{n} a_j(e_j - v_j)\right\|^2 < \sum_{j=1}^{n} |a_j|^2.$$

But we have shown that

$$\left\| \sum_{j=1}^{n} a_j (e_j - v_j) \right\|^2 = \sum_{j=1}^{n} |a_j|^2,$$

so we have a contradiction. Hence, $\sum_{j=1}^{n} a_j v_j = \mathbf{0}$ if and only if $a_j = 0$ for all $j = 1, 2, \ldots, n$, i.e., $v_1, v_2 \ldots, v_n$ is a linearly independent list. So, $v_1, v_2, \ldots, v_n$ is a basis for $\mathbf{V}$.

Problem. 5.

1. Suppose that $z_1, z_2, \ldots, z_n$ is an orthonormal list of elements of a (not necessarily finite-dimensional) vector space $\mathbf{V}$, and w is an element of $\mathbf{V}$ that is not in the span of $z_1, z_2, \ldots, z_n$. **Use Gramians** to argue that there is an element $y \in \mathbf{V}$ such that $z_1, z_2, \ldots, z_n, y$ is an orthonormal list and

$$\operatorname{span}(z_1, z_2, \ldots, z_n, y) = \operatorname{span}(z_1, z_2, \ldots, z_n, w).$$

Solution. 5.1 Let w and the z_i's be given. We first want to show that there exists $y \in \mathbf{V}$ such that $\langle y, z_i\rangle = 0$ for any $i = 1, 2, \ldots, n$. Consider

$$y = w - \alpha_1 z_1 - \alpha_2 z_2 - \cdots - \alpha_n z_n.$$

We first observe that $y \neq \mathbf{0}$ because $w \notin \operatorname{span}(z_1, \ldots, z_n)$ and thus $\|y\| \neq 0$. Next, we want to show there exist $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$ such that

$$\langle y, z_i\rangle = 0 = \langle w - \alpha_1 z_1 - \alpha_2 z_2 - \cdots - \alpha_n z_n, z_i\rangle$$

for any $i \in \{1, 2, \ldots, n\}$. But this is equivalent to showing there exist $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$ such that

$$\langle w, z_i\rangle = \alpha_1\langle z_1, z_i\rangle + \ldots \alpha_n\langle z_n, z_i\rangle$$

for any $i \in \{1, 2, \ldots, n\}$, i.e.,

$$\text{Gramian}(\gamma) = \begin{bmatrix} \langle z_1, z_1\rangle & \ldots & \langle z_n, z_1\rangle \\ \vdots & \ddots & \vdots \\ \langle z_1, z_n\rangle & \ldots & \langle z_n, z_n\rangle \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \langle w, z_1\rangle \\ \vdots \\ \langle w, z_n\rangle \end{pmatrix},$$

where $\gamma = \begin{pmatrix}\alpha_1 & \ldots & \alpha_n\end{pmatrix}^\top$, has a unique solution, i.e. the Gramian matrix is invertible. But since the Gramian matrix is square, it suffices to show it is injective. Suppose Gramian$(\gamma) = \mathbf{0}$. Then,

$$\sum_{i=1}^n \alpha_i\langle z_i, z_1\rangle = \cdots = \sum_{i=1}^n \alpha_i\langle z_i, z_n\rangle = 0,$$

which means $\sum_{i=1}^n \alpha_i z_i \perp z_1, \ldots, z_n$. In particular, $\sum_{i=1}^n \alpha_i z_i \perp \sum_{i=1}^n \alpha_i z_i \iff \sum_{i=1}^n \alpha_i z_i = \mathbf{0}$. But since the z_i's are nonzero vectors (because each has length 1), $\alpha_i = 0$ for all $i \in \{1, 2, \ldots, n\}$. This shows the Gramian is injective and therefore invertible.

So, we have shown that there exist $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ such that y as given above is orthogonal to all the z_i's. Thus, the normalized y, denoted y_N and defined as

$$y_N = \frac{w - \alpha_1 z_1 - \alpha_2 z_2 - \cdots - \alpha_n z_n}{\|w - \alpha_1 z_1 - \alpha_2 z_2 - \cdots - \alpha_n z_n\|}$$

is not only orthogonal to all z_i's but also has length 1. Hence, $z_1, \ldots, z_n, y_N$ form an orthonormal list.

Now, consider the subspaces $\text{span}(z_1, \ldots, z_n, y_N)$ and $\text{span}(z_1, \ldots,_n, w)$. By definition,

$$\text{span}(z_1, \ldots, z_n, y_N) = \left\{ a_1 z_1 + \ldots a_n z_n + b y_N \middle| a_1, \ldots, a_n, b \in \mathbb{C} \right\}$$
$$\text{span}(z_1, \ldots, z_n, w) = \left\{ a_1 z_1 + \ldots a_n z_n + b w \middle| a_1, \ldots, a_n, b \in \mathbb{C} \right\}.$$

But for $y = \|y\| y_N = w - \alpha_1 z_1 - \alpha_2 z_2 - \cdots - \alpha_n z_n$,

$$w = \alpha_1 z_1 + \cdots + \alpha_n z_n + y = \alpha_1 z_1 + \cdots + \alpha_n z_n + \|y\| y_N.$$

So,

$$\begin{aligned} \text{span}(z_1, \ldots, z_n, w) &= \left\{ c_1 z_1 + \cdots + c_n z_n + d y_N \middle| c_1, \ldots, c_n, d \in \mathbb{C} \right\} \\ &= \text{span}(z_1, \ldots, z_n, y_N). \end{aligned}$$

Therefore, given $w \in \mathbf{V}$ and the z_i's, we have shown there exists a vector $y \in \mathbf{V}$ that satisfies all the requirements. In particular,

$$y = \frac{w - \alpha_1 z_1 - \alpha_2 z_2 - \cdots - \alpha_n z_n}{\|w - \alpha_1 z_1 - \alpha_2 z_2 - \cdots - \alpha_n z_n\|}$$

□

2. **Problem 2 in Exercises 6.B in Axler** Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in $\mathbf{V}$. Let $v \in \mathbf{V}$. Prove that

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2$$

if and only if $v \in \operatorname{span}(e_1, \ldots e_m)$.

Solution. 5.2

(a) To show: $\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2 \implies v \in \operatorname{span}(e_1, \ldots, e_m)$. Assume that $v \in \mathbf{V}$ is not in $\operatorname{span}(e_1, \ldots, e_m)$, then by the previous problem, there exists an element $y \in \mathbf{V}$ such that $e_1, e_2, \ldots, e_m, y$ is an orthonormal list and

$$\operatorname{span}(e_1, e_2, \ldots, e_m, y) = \operatorname{span}(e_1, e_2, \ldots, e_m, v).$$

Then we can write

$$v = \beta y + \sum_{i=1}^{m} \alpha_i e_i$$

for $\alpha_1, \ldots, \alpha_m, \beta \in \mathbb{C}$, and $\beta \neq 0$ since otherwise $v \in \operatorname{span}(e_1, e_2, \ldots, e_m)$. Then,

$$\langle v, e_i\rangle = \alpha_i$$

and thus

$$\begin{aligned}
\|v\|^2 &= \left\langle \beta y + \sum_{i=1}^{m} \alpha_i e_i, \beta y + \sum_{j=1}^{m} \alpha_j e_j \right\rangle \\
&= |\beta|^2 + \sum_{i=1}^{m} |\alpha_i|^2 \\
&= |\beta|^2 + \sum_{i=1}^{m} |\langle v, e_i\rangle|^2 \\
&= |\beta|^2 + \|v\|^2, \text{ by hypothesis.}
\end{aligned}$$

Therefore, $\beta = 0$. We have a contradiction. Thus, $v \in \operatorname{span}(e_1, \ldots e_m)$ as desired.

(b) To show: $v \in \operatorname{span}(e_1, \ldots, e_m) \implies \|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2$. Let $v \in \mathbf{V}$ be given, then

$$v = \sum_{i=1}^{m} \alpha_i e_i,$$

where $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$. Thus,

$$\|v\|^2 = \left\langle \sum_{i=1}^{m} \alpha_i e_i, \sum_{j=1}^{m} \alpha_j e_j \right\rangle = \sum_{i=1}^{m} |\alpha_i|^2.$$

We also know that

$$\langle v, e_i \rangle = \left\langle \sum_{j=1}^{m} \alpha_j e_j, e_i \right\rangle = \alpha_i.$$

Hence,

$$\|v\|^2 = \sum_{i=1}^{m} |\langle v, e_i \rangle|^2,$$

as desired. $\square$

24.7 Problem set PRACTICE

Problem. 1. Shuttling Inner Products

1. Suppose that $(\mathbf{V}, \langle \cdot, \cdot \rangle)$ is a finite-dimensional inner product space (fdips) with an orthogonal basis Γ. Let $\mathcal{A} : \mathbb{C}^n \to \mathbf{V}$ be the atrix corresponding to Γ. Prove that:

$$\langle v, w \rangle = \mathcal{A}^{-1}(v) \cdot \mathcal{A}^{-1}(w)$$

   for every $v, w \in \mathbf{V}$, where $\cdot$ is the standard inner product.

2. Suppose that $\mathbf{V}$ and $\mathbf{W}$ are fdips with ortho-bases Γ and Ω respectively, and $\text{Ł} : \mathbf{V} \to \mathbf{W}$ is a linear function. Prove that $[\mathcal{L}^*]_{\Gamma \leftarrow \Omega}$ is the conjugate transpose of $[\mathcal{L}]_{\Omega \leftarrow \Gamma}$.

Problem. 2. Isometries Suppose that $\mathbf{V}$ and $\mathbf{W}$ are inner product spaces, and $\mathcal{L} : \mathbf{V} \to \mathbf{W}$ is a linear function.

1. Argue that the following claims are equivalent.

   (a) $\mathcal{L}$ is an isometry; i.e., $\|\mathcal{L}(v)\|_{\mathbf{W}} = \|v\|_{\mathbf{V}}$, for all $v \in \mathbf{V}$.

   (b) $\langle \mathcal{L}(v), \mathcal{L}(z) \rangle_{\mathbf{W}} = \langle v, z \rangle_{\mathbf{V}}$ for all $v, z \in \mathbf{V}$.

   (c) $\mathcal{L}^* \mathcal{L} = \mathcal{I}_{\mathbf{V}}$.

   (d) $\|\mathcal{L}(v)\|_{\mathbf{W}} = 1$, for every unit vector $v \in \mathbf{V}$.

2. Argue that the following claims are equivalent.

   (a) $\mathcal{L}$ is a scalar multiple of an isometry.

   (b) $\|\mathcal{L}(v)\|_{\mathbf{W}} = \|\mathcal{L}(z)\|_{\mathbf{W}}$ for any unit vectors $v, z \in \mathbf{V}$.

   (c) $\mathcal{L}$ preserves orthogonality; i.e., if $\langle v, z \rangle_{\mathbf{V}} = 0$ then $\langle \mathcal{L}(v), \mathcal{L}(z) \rangle_{\mathbf{W}} = 0$.

3. Suppose that $\mathbf{V}$ and $\mathbf{W}$ are fdips with orthonormal bases Γ and Ω respectively. Argue that $\mathcal{L}$ is an isometry if and only if the columns of $[\mathcal{L}]_{\Omega \leftarrow \Gamma}$ are orthonormal.

Problem. 3. Unitaries Suppose that $\mathbf{V}$ is an inner product space, and $\mathcal{L} : \mathbf{V} \to \mathbf{V}$ is a linear function.

1. Argue that the following claims are equivalent.
   (a) $\mathcal{L}$ is unitary.
   (b) $\mathcal{L}$ is invertible, and $\mathcal{L}^{-1} = \mathcal{L}^*$.
2. Argue that the following are equivalent in the case that V is fdips, and Γ is an orthonormal basis of $\mathbf{V}$.
   (a) $\mathcal{L}$ is unitary.
   (b) The columns of $[\mathcal{L}]_{\Gamma\leftarrow\Gamma}$ form an orthonormal basis (of the appropriate $\mathbb{C}^n$).

Problem. 4. Unitary Equivalence for Matrices Let $\mathbf{V}$ be an $n-$dimensional inner product space.

1. Suppose that Γ and Ω are two orthonormal bases of $\mathbf{V}$. Argue that $[\mathcal{I}]_{\Omega\leftarrow\Gamma}$ is a unitary matrix.
2. Argue that for each unitary matrix $\mathcal{U}$ in $\mathbb{M}_n$ there exist orthonormal bases Γ and Ω of $\mathbf{V}$ such that $\mathcal{U} = [\mathcal{I}]_{\Omega\leftarrow\Gamma}$.
3. Suppose that $\mathcal{A}, \mathcal{B} \in \mathbb{M}_n$. Argue that the following claims are equivalent.
   (a) There exists a linear function $\mathcal{L} : \mathbf{V} \to \mathbf{V}$ and orthonormal bases Γ and Ω of $\mathbf{V}$ such that
   $$\mathcal{A} = [\mathcal{L}]_{\Gamma\leftarrow\Gamma} \quad \text{and} \quad \mathcal{B} = [\mathcal{L}]_{\Omega\to\Omega}.$$
   (b) There exists a unitary matrix $\mathcal{U}$ in $\mathbb{M}_n$ such that
   $$\mathcal{B} = \mathcal{U}^{-1}\mathcal{A}\mathcal{U}.$$

Problem. 5. Properties of Unitary Similarity/Equivalence

1. Argue that unitary equivalence is an equivalence relation on $\mathfrak{L}(\mathbf{V}, \mathbf{V})$.
2. Argue that unitary equivalence preserves normality, "unitariness" and "self-adjoint-ness."

Problem. 6. Spectral Theorem

1. Argue that the following claims are equivalent for a square matrix $\mathcal{A}$.
   (a) $\mathcal{A}$ is normal.
   (b) $\mathcal{A}$ is unitarily equivalent to a diagonal matrix.
2. Argue that the following claims are equivalent for a square matrix $\mathcal{B}$.
   (a) $\mathcal{B}$ is unitary.
   (b) $\mathcal{B}$ is unitarily equivalent to a diagonal matrix with diagonal entries of modulus 1.
3. Argue that the following claims are equivalent for a square matrix $\mathcal{C}$.
   (a) $\mathcal{C}$ is self-adjoint.
   (b) $\mathcal{C}$ is unitarily equivalent to a diagonal matrix with real entries.

Problem. 7. Spectral Theorem For Normal Linear Operators Suppose that $\mathbf{V}$ is an fdips, and $\mathcal{L}$ is a linear operator on $\mathbf{V}$. Prove each of the following claims.

1. $\mathcal{L}$ is normal if and only if $\mathbf{V}$ has an orthonormal basis comprised of the eigenvectors of $\mathcal{L}$.

2. $\mathcal{L}$ is unitarily if and only if $\mathcal{L}$ is normal, and all eigenvalues of $\mathcal{L}$ have modulus 1.

3. $\mathcal{L}$ is self-adjoint if and only if $\mathcal{L}$ is normal, and all eigenvalues of $\mathcal{L}$ are real.

Problem. 8. Diagonalizability and LInear Combinations of Mutually Annihilating Idempotents Suppose that $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_p$ are idempotent linear operators on a vector space $\mathbf{V}$, such that

$$\mathcal{G}_i \mathcal{G}_j = \mathcal{O},$$

whenever $i \neq j$. In this case we say that the $\mathcal{G}_i$'s are mutually annihilating idempotents.

1. Argue that $\mathcal{G}_1 + \mathcal{G}_2 + \ldots \mathcal{G}_p$ is an idempotent and infer that

   $$\alpha_1 \mathcal{G}_1 + \alpha_2 \mathcal{G}_2 + \ldots \alpha_p \mathcal{G}_p$$

   is a diagonalizable operator.

2. Argue that a linear operator on $\mathbf{V}$ is diagonalizable exactly when it is a linear combination of some mutually annihilating idempotent operators.

Problem. 9. Atomic Spectral Idempotents are Polynomials in $\mathcal{L}$ Suppose that $\mathbf{V}$ is a finite-dimensional vector space, and $\mathcal{L}$ is a diagonalizable linear operator on $\mathbf{V}$ with

$$\sigma_{\mathbb{C}}\{\mathcal{L}\} = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}.$$

Argue that each corresponding atomic spectral idempotent $\mathcal{E}_i$ of $\mathcal{L}$ can be expressed as $p_i(\mathcal{L})$, for the unique polynomial p_i of degree at most $k-1$ which maps λ_i to 1, and all other eigenvalues of $\mathcal{L}$ to zero.

Problem. 10. Uniqueness of Spectral Resolution Suppose that $\mathbf{V}$ is a finite-dimensional vector space, and $\mathcal{L}$ is a diagonalizable linear operator on $\mathbf{V}$ with

$$\sigma_{\mathbb{C}}\{\mathcal{L}\} = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}.$$

1. Suppose that

   $$\mathcal{L} = \gamma_1 \mathcal{F}_1 + \gamma_2 \mathcal{F}_2 + \ldots \gamma_m \mathcal{F}_m,$$

   for some idempotents $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_m$ which resolve the identity on $\mathbf{V}$, and some *distinct* complex numbers $\gamma_1, \gamma_2, \ldots, \gamma_m$. Argue that $m = k$, that

   $$\sigma_{\mathbb{C}}\{\mathcal{L}\} = \{\lambda_1, \lambda_2, \ldots, \lambda_k\} = \{\gamma_1, \gamma_2, \ldots, \gamma_k\}$$

   and that $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$ are the atomic spectral idempotents of $\mathcal{L}$.

2. Explain how one discerns the atomic spectral idempotents and the eigenspaces of $\mathcal{L}$ from any $\mathcal{L}-$eigenbasis of $\mathbf{V}$. Justify your claims.

Problem. 11. Spectral Resolutions of Normal Operators Suppose that $\mathbf{V}$ is an fdips, that $\mathcal{L}$ is a normal linear operator on $\mathbf{V}$ with

$$\sigma_{\mathbb{C}}\{\mathcal{L}\} = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}.$$

1. Argue that every atomic spectral idempotent of $\mathcal{L}$ is normal, and conclude that every normal spectral idempotent of $\mathcal{L}$ is a non-negative (i.e., positive semi-definite) operator.

   Non-negative idempotents are said to be **ortho-projections** or **projections**, for short.

2. Argue that

$$\mathbf{V} = \mathbf{E}_{\lambda_1} \oplus \mathbf{E}_{\lambda_2} \oplus \cdots \oplus \mathbf{E}_{\lambda_k},$$

   where $\mathbf{E}_{\lambda_i}$ is the eigenspace of $\mathcal{L}$ corresponding to the eigenvalue λ, where the $\mathbf{E}_{\lambda_i}$ are mutually orthogonal subspaces.

Problem. 12. Another Form of the Spectral Theorem for Normal Operators Suppose that $\mathbf{V}$ is an fdips, that $\mathcal{L}$ is a linear operator on $\mathbf{V}$. Argue that the following claims are equivalent.

1. $\mathcal{L}$ is normal.
2. $\mathcal{L}$ is a linear combination of ortho-projections that resolve the identity on $\mathbf{V}$.
3. $\mathcal{L}$ is a linear combination of mutually annihilating projections.

Problem. 13. Commuting with a Normal Suppose that $\mathbf{V}$ is an fdips, that $\mathcal{L}$ is a normal linear operator on $\mathbf{V}$. Argue that the following claims are equivalent for an operator $\mathcal{M}$ on $\mathbf{V}$.

1. $\mathcal{M}$ commutes with $\mathcal{L}$.
2. $\mathcal{M}$ commutes with every atomic spectral projection of $\mathcal{L}$.

Problem. 14. Another Polarization Identity Suppose that $\mathcal{L}$ is a linear operator on an inner project space $\mathbf{V}$.

1. Argue that

$$4\langle \mathcal{L}(x), y\rangle = \langle \mathcal{L}(x+y), x+y\rangle - \langle \mathcal{L}(x-y), x-y\rangle$$
$$+i\langle \mathcal{L}(x+iy), x+iy\rangle - i\langle \mathcal{L}(x-iy), x-iy\rangle.$$

2. Infer that

$$\langle \mathcal{L}(z), z\mathcal{L} = 0$$

for all $z \in \mathbf{V}$ if and only if $\mathcal{L} = 0$, and that for an operator $\mathcal{M}$ on $\mathbf{V}$,

$$\langle \mathcal{L}(z), z\rangle = \langle \mathcal{M}(z), z\rangle,$$

for all $z \in \mathbf{V}$ if and only if $\mathcal{L} = \mathcal{M}$.

Problem. 15. Non-negativity (and Positivity) Conditions I Suppose that $\mathbf{V}$ is an fdips, and $\mathcal{L}$ is a linear operator on $\mathbf{V}$. Prove each of the following claims.

1. $\mathcal{L}$ is non-negative if and only if $\mathcal{L}$ is normal, and all eigenvalues of $\mathcal{L}$ are non-negative.
2. $\mathcal{L}$ is positive if and only if $\mathcal{L}$ is normal, and all eigenvalues of $\mathcal{L}$ are positive.
3. $\mathcal{L}$ is non-negative if and only if $\mathcal{L}$ is self-adjoint, and all eigenvalues of $\mathcal{L}$ are non-negative.
4. $\mathcal{L}$ is positive if and only if $\mathcal{L}$ is non-negative and invertible.

Problem. 16. Non-negativity Conditions II Suppose that $\mathbf{V}$ is an fdips, and $\mathcal{L}$ is a linear operator on $\mathbf{V}$. Prove that the following claims are equivalent.

1. $\mathcal{L}$ is non-negative.
2. $\mathcal{L} = \mathcal{N}^2$ for some non-negative operator $\mathcal{N}$ on $\mathbf{V}$.*
3. $\mathcal{L} = \mathcal{M}^2$ for some self-adjoint operators $\mathcal{M}$ on $\mathbf{V}$.
4. $\mathcal{L} = \mathcal{T}^*\mathcal{T}$ for some operator $\mathcal{T}$ on $\mathbf{V}$.
5. $\mathcal{L} = \mathcal{K}\mathcal{K}^*$ for some operator $\mathcal{K}$ on $\mathbf{V}$.

*Such an $\mathcal{N}$ is said to be a **non-negative square root of $\mathcal{L}$**

Problem. 17. Suppose that $\mathbf{V}$ is an fdips, and $\mathcal{L}$ is a non-negative linear operator on $\mathbf{V}$. Argue that for any $x \in \mathbf{V}$ the following claims are equivalent.

1. $\mathcal{L}(x) = \mathbf{0}$.
2. $\langle \mathcal{L}(x), x \rangle = 0$.
3.

Problem. 18. Non-negativity Conditions for Matrices

1. Argue that the following claims are equivalent for a square matrix $\mathcal{P}$.
   (a) $\mathcal{P}$ is non-negative.
   (b) $\mathcal{P}$ is unitarily equivalent to a diagonal matrix with non-negative diagonal entries.
2. Argue that the following claims are equivalent for a square matrix $\mathcal{P}$.
   (a) $\mathcal{P}$ is positive.
   (b) $\mathcal{P}$ is unitarily equivalent to a diagonal matrix with positive diagonal entries.

Problem. 19. Properties of $\sqrt{\mathcal{L}^*\mathcal{L}}$ and $\sqrt{\mathcal{L}\mathcal{L}^*}$ Suppose that $\mathbf{V}$ and $\mathbf{W}$ are fdips, and $\mathcal{L} : \mathbf{V} \to \mathbf{W}$ is a linear operator. Argue that

1. $$\begin{aligned} \left\| \sqrt{\mathcal{L}^*\mathcal{L}}(x) \right\| &= \|\mathcal{L}(x)\|, \quad \text{for all } x \in \mathbf{V} \text{ and} \\ \left\| \sqrt{\mathcal{L}\mathcal{L}^*}(y) \right\| &= \|\mathcal{L}^*(y)\|, \quad \text{for all } y \in \mathbf{V} \end{aligned}$$

2. $$\begin{aligned} \ker(\sqrt{\mathcal{L}^*\mathcal{L}}) &= \ker(\mathcal{L}) \text{ and} \\ \ker(\sqrt{\mathcal{L}\mathcal{L}^*}) &= \ker(\mathcal{L}^*). \end{aligned}$$

3. $$\begin{aligned} \mathrm{Im}(\sqrt{\mathcal{L}^*\mathcal{L}}) &= \mathrm{Im}(\mathcal{L}^*) \text{ and} \\ \mathrm{Im}(\sqrt{\mathcal{L}\mathcal{L}^*}) &= \mathrm{Im}(\mathcal{L}). \end{aligned}$$