

Notes on Linear Algebra*

Jay R. Walton
Department of Mathematics
Texas A& M University

October 3, 2007

1 Introduction

Linear algebra provides the foundational setting for the study of multivariable mathematics which in turn is the bedrock upon which most modern theories of mathematical physics rest including classical mechanics (rigid body mechanics), continuum mechanics (the mechanics of deformable material bodies), relativistic mechanics, quantum mechanics, etc. At the heart of linear algebra is the notion of a (linear) vector space which is an abstract mathematical structure introduced to make rigorous the classical, intuitive concept of vectors as physical quantities possessing the two attributes of *length* and *direction*.

In these brief notes, vector spaces and linear transformations are introduced in a three step presentation. First they are studied as algebraic objects and a few important consequences of the concept of linearity are explored. Next the algebraic structure is augmented by introducing a topological structure (via a metric or a norm, for example) providing a convenient framework for extending key concepts from the calculus to the vector space setting permitting a rigorous framework for studying nonlinear, multivariable functions between vector spaces. Finally, an inner product structure for vector spaces is introduced in order to define the geometric notion of angle (and as a special case, the key concept of orthogonality) augmenting the notion of length or distance provided by previously by a norm or a metric.

2 Vector Space

The central object in the study of linear algebra is a *Vector Space* which is defined here as follows.

Definition 1 *A Vector Space over a field \mathcal{F} ¹ is a set, \mathcal{V} , whose elements are called vectors, endowed with two commutative and associative binary operations, called vector addition and scalar multiplication, subject to the following axioms.*

*Copyright © 2006 by Jay R. Walton. All rights reserved.

¹While in general \mathcal{F} can be any field, in these notes it will denote either the real numbers \mathbb{R} or the complex numbers \mathbb{C} .

1. If $\mathbf{a}, \mathbf{b} \in \mathcal{V}$, then their sum $\mathbf{a} + \mathbf{b}$ is also in \mathcal{V} .
2. Vector addition is commutative and associative, i.e. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ and all $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ and $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$ for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}$.
3. There exists a zero element, $\mathbf{0}$, in \mathcal{V} satisfying $\mathbf{0} + \mathbf{a} = \mathbf{a}$ for all $\mathbf{a} \in \mathcal{V}$.
4. Every element $\mathbf{a} \in \mathcal{V}$ possesses a (unique) additive inverse, $-\mathbf{a}$, satisfying $-\mathbf{a} + \mathbf{a} = \mathbf{0}$.
5. If $\mathbf{a} \in \mathcal{V}$ and $\alpha \in \mathcal{F}$, then $\alpha\mathbf{a} \in \mathcal{V}$.
6. For all $\mathbf{a} \in \mathcal{V}$, $1\mathbf{a} = \mathbf{a}$.
7. If $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ and $\alpha, \beta \in \mathcal{F}$, then $\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}$, $(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}$, $\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}$.²

Remark 1 It follows readily from these axioms that for any $\mathbf{a} \in \mathcal{V}$, $0\mathbf{a} = \mathbf{0}$ and $(-1)\mathbf{a} = -\mathbf{a}$. (Why?)

Remark 2 The above axioms are sufficient to conclude that for any collection of vectors $\mathbf{a}_1, \dots, \mathbf{a}_N \in \mathcal{V}$ and scalars $\alpha_1, \dots, \alpha_N \in \mathcal{F}$, the *Linear Combination* $\alpha_1\mathbf{a}_1 + \dots + \alpha_N\mathbf{a}_N$ is an unambiguously defined vector in \mathcal{V} . (Why?)

Example 1 The prototype examples of vector spaces over the real numbers \mathbb{R} and the complex numbers \mathbb{C} are the Euclidean spaces \mathbb{R}^N and \mathbb{C}^N of N -tuples of real or complex numbers. Specifically, the elements $\mathbf{a} \in \mathbb{R}^N$ are defined by

$$\mathbf{a} := \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}$$

where $a_1, \dots, a_N \in \mathbb{R}$. Vector addition and scalar multiplication are defined componentwise, that is

$$\alpha\mathbf{a} + \beta\mathbf{b} = \alpha \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} + \beta \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} = \begin{pmatrix} \alpha a_1 + \beta b_1 \\ \vdots \\ \alpha a_N + \beta b_N \end{pmatrix}.$$

Analogous relations hold for the complex case.

Example 2 The vector space \mathbb{R}^∞ is defined as the set of all (infinite) sequences, \mathbf{a} , of real numbers, that is, $\mathbf{a} \in \mathbb{R}^\infty$ means

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \end{pmatrix}$$

with vector addition and scalar multiplication done componentwise.

²The symbol “+” has been used to denote both the addition of vectors and the addition of scalars. This should not generate any confusion since the meaning of “+” should be clear from the context in which it is used.

Definition 2 A subset \mathcal{U} of a vector space \mathcal{V} is called a *Subspace* provided it is closed with respect to vector addition and scalar multiplication (and hence is also a vector space).

Example 3 One shows easily that \mathbb{R}^N is a subspace of \mathbb{R}^M whenever $N \leq M$. (Why?)

Example 4 The vector space \mathbb{R}_f^∞ is the set all of infinite sequences of real numbers with only finitely many non-zero components. It follows readily that \mathbb{R}_f^∞ is a subspace of \mathbb{R}^∞ . (Why?)

Definition 3 Let A be a subset of a vector space \mathcal{V} . Then the *Span* of A , denoted $\text{span}(A)$, is defined to be the set of all (finite) linear combinations formed from the vectors in A . A subset $A \subset \mathcal{V}$ is called a *Spanning Set* for \mathcal{V} if $\text{span}(A) = \mathcal{V}$. A spanning set $A \subset \mathcal{V}$ is called a *Minimal Spanning Set* if no proper subset of A is a spanning set for \mathcal{V} .

Remark 3 One readily shows that $\text{span}(A)$ is a subspace of \mathcal{V} . (How?)

2.1 Linear Independence and Dependence

The most important concepts associated with vector spaces are *Linear Independence* and *Linear Dependence*.

Definition 4 A set of vectors $A \subset \mathcal{V}$ is said to be *Linearly Independent* if given any vectors $\mathbf{a}_1, \dots, \mathbf{a}_N \in A$, then the only way to write the zero vector as a linear combination of these vectors is if all coefficients are zero, that is,

$$\alpha_1 \mathbf{a}_1 + \dots + \alpha_N \mathbf{a}_N = \mathbf{0} \quad \text{if and only if} \quad \alpha_1 = \dots = \alpha_N = 0.$$

Remark 4 If a set of vectors A is linearly independent and $B \subset A$, then B is also linearly independent. (Why?)

Remark 5 A minimal spanning set is easily seen to be linearly independent. Indeed, suppose A is a minimal spanning set and suppose it is not linearly independent. Then there exists a subset $\{\mathbf{a}_1, \dots, \mathbf{a}_N\}$ of A such that \mathbf{a}_N can be written as a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_{N-1}$. It follows that $A \sim \{\mathbf{a}_N\}$ (set difference) is a proper subset of A that spans \mathcal{V} contradicting the assumption that A is a minimal spanning set.

Definition 5 If $\mathbf{a} \neq \mathbf{0}$, then $\text{span}(\{\mathbf{a}\})$ is called a *Line* in \mathcal{V} . If $\{\mathbf{a}, \mathbf{b}\}$ is linearly independent, then $\text{span}(\{\mathbf{a}, \mathbf{b}\})$ is called a *Plane* in \mathcal{V} .

Definition 6 A set of vectors $A \subset \mathcal{V}$ is said to be *Linearly Dependent* if there exists a set of vectors $\mathbf{a}_1, \dots, \mathbf{a}_N \in \mathcal{V}$ and scalars $\alpha_1, \dots, \alpha_N \in \mathcal{F}$, not all zero, with

$$\alpha_1 \mathbf{a}_1 + \dots + \alpha_N \mathbf{a}_N = \mathbf{0}.$$

Remark 6 A set of vectors that contains the zero vector is linearly dependent. (Why?)

Remark 7 If the set of vectors A is linearly dependent and $A \subset B$, then B is also linearly dependent. (Why?)

Remark 8 If $A = \{\mathbf{a}_1, \dots, \mathbf{a}_N\} \subset \mathcal{V}$ is linearly dependent, then one of the \mathbf{a}_j can be written as a non-trivial (not all scalar coefficients zero) linear combination of the others, that is, one of the \mathbf{a}_j is in the span of the others.

2.2 Basis

The notion of *Basis* for a vector space is intimately connected with spanning sets of minimal cardinality. Here attention is restricted to vector spaces with spanning sets containing only finitely many vectors.

Definition 7 A vector space \mathcal{V} is called *Finite Dimensional* provided it has a spanning set with finitely many vectors.

Theorem 1 Suppose \mathcal{V} is finite dimensional with minimal spanning set $A = \{\mathbf{a}_1, \dots, \mathbf{a}_N\}$. Let B be another minimal spanning set for \mathcal{V} . Then $\#(B) = N$, where $\#(B)$ denotes the cardinality of the set B .

Proof: It suffices to assume that $\#(B) \geq N$. (Why?) Then choose $B' := \{\mathbf{b}_1, \dots, \mathbf{b}_N\} \subset B$. Clearly, B' is linearly independent. (Why?) Since A is spanning, there exist scalars α_{ij} such that $\mathbf{b}_i = \sum_{j=1}^N \alpha_{ij} \mathbf{a}_j$. Then the matrix $\mathbf{L} := [\alpha_{ij}]$ is invertible. (Why?) Suppose $\mathbf{L}^{-1} = [\beta_{ij}]$. Then $\mathbf{a}_i = \sum_{j=1}^N \beta_{ij} \mathbf{b}_j$, that is, the set $\{\mathbf{a}_1, \dots, \mathbf{a}_N\}$ is in the span of B' . But since A spans \mathcal{V} , it follows that B' must also span \mathcal{V} . But since B was assumed to be a minimal spanning set, it must be that $B = B'$. (Why?) Hence, $\#(B) = N$. \square^3

This theorem allows one to define unambiguously the notion of *dimension* for a finite dimensional vector space.

Definition 8 Let \mathcal{V} be a finite dimensional vector space. Then its *Dimension*, denoted $\dim(\mathcal{V})$, is defined to be the cardinality of any minimal spanning set.

Definition 9 A *Basis* for a finite dimensional vector space is defined to be any minimal spanning set.

It follows that given a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ for a finite dimensional vector space \mathcal{V} , every vector $\mathbf{a} \in \mathcal{V}$ can be written uniquely as a linear combination of the base vectors, that is

$$\mathbf{a} = \sum_{j=1}^N \alpha_j \mathbf{b}_j \quad (1)$$

with the coefficients $\alpha_1, \dots, \alpha_N$ being uniquely defined. These coefficients are called the *coordinates* or *components* of \mathbf{a} with respect to the basis B . It is often convenient collect

³The symbol \square is used in these notes to indicate the end of a proof.

these components into an N -tuple denoted

$$[\mathbf{a}]_B := \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}.$$

Notational Convention: It is useful to introduce simplifying notation for sums of the sort (1). Specifically, in the *Einstein Summation Convention*, expressions in which an index is repeated are to be interpreted as being summed over the range of the index. Thus (1) reduces to

$$\mathbf{a} = \sum_{j=1}^N \alpha_j \mathbf{b}_j = \alpha_j \mathbf{b}_j.$$

With the Einstein summation convention in effect, if one wants to write an expression like $\alpha_j \mathbf{b}_j$ for a particular, but unspecified, value of j , then one must write

$$\alpha_j \mathbf{b}_j \text{ (no sum)}$$

Example 5 In the Euclidean vector space \mathbb{R}^N , it is customary to define the *Natural Basis* $\mathcal{N} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ to be

$$\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad \mathbf{e}_N := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Finding components of vectors in \mathbb{R}^N with respect to the natural basis is then immediate.

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \in \mathbb{R}^N \implies [\mathbf{a}]_{\mathcal{N}} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}$$

that is,

$$\mathbf{a} = a_j \mathbf{e}_j.$$

In general, finding the components of a vector with respect to a prescribed basis usually involves solving a linear system of equations. For example, suppose a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ in \mathbb{R}^N is given as

$$\mathbf{b}_1 := \begin{pmatrix} b_{11} \\ \vdots \\ b_{N1} \end{pmatrix}, \quad \dots, \quad \mathbf{b}_N := \begin{pmatrix} b_{1N} \\ \vdots \\ b_{NN} \end{pmatrix}.$$

Then to find the components

$$[\mathbf{a}]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}$$

of a vector

$$\mathbf{a} := \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}$$

with respect to B , one solves the linear system of equations $a_i = b_{ij}\alpha_j$, $i = 1, \dots, N$ which in expanded form is

$$\begin{aligned} a_1 &= b_{11}\alpha_1 + \dots + b_{1N}\alpha_N \\ &\vdots \\ a_N &= b_{N1}\alpha_1 + \dots + b_{NN}\alpha_N. \end{aligned}$$

2.3 Change of Basis

Let $B = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ and $\tilde{B} = \{\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_N\}$ be two bases for the vector space \mathcal{V} . The question considered here is how to change components of vectors with respect to the basis B into components with respect to the basis \tilde{B} ? To answer the question, one needs to specify some relation between the two sets of base vectors. Suppose, for example, one can write the base vectors $\tilde{\mathbf{b}}_j$ as linear combinations of the base vectors \mathbf{b}_j , that is, one knows the components of each N -tuple $[\tilde{\mathbf{b}}_j]_B$, $j = 1, \dots, N$. To be more specific, suppose

$$\tilde{\mathbf{b}}_j = t_{ij}\mathbf{b}_i, \quad \text{for } j = 1, \dots, N. \quad (2)$$

Form the matrix $T := [t_{ij}]$ called the *Transition Matrix* between the bases B and \tilde{B} . Then T is invertible. (Why?) Let its inverse have components $T^{-1} = [s_{ij}]$. By definition of matrix inverse, one sees that

$$s_{ik}t_{kj} = \delta_{ij}$$

where the Kronecker symbol δ_{ij} is defined as

$$\delta_{ij} := \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{when } i \neq j. \end{cases}$$

Suppose now that a vector \mathbf{a} has components

$$[\mathbf{a}]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}.$$

Then one shows readily (How?) that the components of \mathbf{a} with respect to the basis \tilde{B} are given by

$$[\mathbf{a}]_{\tilde{B}} = \begin{pmatrix} \tilde{\alpha}_1 \\ \vdots \\ \tilde{\alpha}_N \end{pmatrix}$$

with

$$\tilde{\alpha}_i = s_{ij}\alpha_j.$$

In matrix notation this change of basis formula becomes

$$[\mathbf{a}]_{\tilde{B}} = T^{-1}[\mathbf{a}]_B. \quad (3)$$

3 Linear Transformations

Linear transformations are at the heart of the subject of linear algebra. Important in their own right, they also are the principal tool for understanding nonlinear functions between subsets of vector spaces. They will be studied here first from an algebraic perspective. Subsequently, they will be studied geometrically.

Definition 10 A linear transformation, $\mathbf{L} : \mathcal{V} \longrightarrow \mathcal{U}$, between the vector spaces \mathcal{V} and \mathcal{U} is a function satisfying

$$\mathbf{L}(\alpha \mathbf{a} + \beta \mathbf{b}) = \alpha \mathbf{L}\mathbf{a} + \beta \mathbf{L}\mathbf{b}, \quad \text{for every } \alpha, \beta \in \mathcal{F} \text{ and } \mathbf{a}, \mathbf{b} \in \mathcal{V}.^4 \quad (4)$$

Remark 9: The symbol “+” on the left hand side of (4) denotes vector addition in \mathcal{V} whereas on right hand side of (4) it denotes vector addition in \mathcal{U} .

Remark 10: From the definition (4) it follows immediately that once a linear transformation has been specified on a basis for the domain space, its values are determined on all of the domain space through linearity.

Two important subspaces associated with a linear transformation are its *Range* and *Nullspace*. Consider a linear transformation

$$\mathbf{L} : \mathcal{V} \longrightarrow \mathcal{U}. \quad (5)$$

Definition 11 The Range of a linear transformation (5) is the subset of \mathcal{U} , denoted $\text{Ran}(\mathbf{L})$, defined by $\text{Ran}(\mathbf{L}) := \{\mathbf{u} \in \mathcal{U} \mid \text{there exists } \mathbf{v} \in \mathcal{V} \text{ such that } \mathbf{u} = \mathbf{L}\mathbf{v}\}$.

Definition 12 The Nullspace of a linear transformation (5) is the subset of \mathcal{V} , denoted $\text{Nul}(\mathbf{L})$, defined by $\text{Nul}(\mathbf{L}) := \{\mathbf{v} \mid \mathbf{L}\mathbf{v} = \mathbf{0}\}$.

Remark 9 It is an easy exercise to show that $\text{Ran}(\mathbf{L})$ is a subspace of \mathcal{U} and $\text{Nul}(\mathbf{L})$ is a subspace of \mathcal{V} . (How?)

Example 6 Let A be an $M \times N$ matrix

$$A := \left(\begin{pmatrix} a_{11} \\ \vdots \\ a_{M1} \end{pmatrix} \cdots \begin{pmatrix} a_{1N} \\ \vdots \\ a_{MN} \end{pmatrix} \right) = (\mathbf{a}_1 \cdots \mathbf{a}_N)$$

where $\mathbf{a}_1, \dots, \mathbf{a}_N$ denote the *column vectors* of A , that is, A is viewed as consisting of N -column vectors from \mathbb{R}_M .

⁴In analysis it is customary to make a notational distinction between the name of a function and the values the function takes. Thus, for example, one might denote a function by f and values the function takes by $f(x)$. For a linear transformation, \mathbf{L} , it is customary to denote its values by $\mathbf{L}\mathbf{a}$. This notation is suggestive of matrix multiplication.

Definition 13 *The Column Rank of A is defined to be the dimension of the subspace of \mathbb{R}^M spanned by its column vectors.*

One can define a linear transformation $\mathbf{A} : \mathbb{R}^N \longrightarrow \mathbb{R}^M$ by

$$\mathbf{A}\mathbf{v} := A\mathbf{v} \text{ (matrix multiplication).} \quad (6)$$

It is readily shown that the range of \mathbf{A} is the span of the column vectors of A ,

$$\text{Ran}(\mathbf{A}) = \text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_N\}$$

and the nullspace of \mathbf{A} is the set of N -tuples

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \in \mathbb{R}^N$$

satisfying

$$v_1\mathbf{a}_1 + \dots + v_N\mathbf{a}_N = \mathbf{0}.$$

It follows that $\dim(\text{Ran}(\mathbf{A}))$ equals the number of linearly independent column vectors of A .

One can also view the matrix A as consisting of M -row vectors from \mathbb{R}^N and define its *row rank* by

Definition 14 *The Row Rank of A is defined to be the dimension of the subspace of \mathbb{R}^N spanned by its row vectors.*

A fundamental result about matrices is

Theorem 2 *Let A be an $M \times N$ matrix. Then*

$$\text{Row Rank of } A = \text{Column Rank of } A. \quad (7)$$

Proof: Postponed.

A closely related result is

Theorem 3 *Let A be an $M \times N$ matrix and \mathbf{A} the linear transformation (6). Then*

$$\dim(\text{Ran}(\mathbf{A})) + \dim(\text{Nul}(\mathbf{A})) = \dim(\mathbb{R}^N) = N. \quad (8)$$

The result (8) is a special case of the more general

Theorem 4 *Let $\mathbf{L} : \mathcal{V} \longrightarrow \mathcal{U}$ be a linear transformation. Then*

$$\dim(\text{Ran}(\mathbf{L})) + \dim(\text{Nul}(\mathbf{L})) = \dim(\mathcal{V}). \quad (9)$$

The proof of this last theorem is facilitated by introducing the notion of *Direct Sum Decomposition* of a vector space \mathcal{V} .

Definition 15 Let \mathcal{V} be a finite dimensional vector space. A *Direct Sum Decomposition* of \mathcal{V} , denoted $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$, consists of a pair of subspaces, \mathcal{U}, \mathcal{W} of \mathcal{V} with the property that every vector $\mathbf{v} \in \mathcal{V}$ can be written uniquely as a sum $\mathbf{v} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in \mathcal{U}$ and $\mathbf{w} \in \mathcal{W}$.

Remark 10 It follows easily from the definition of direct sum decomposition that

$$\mathcal{V} = \mathcal{U} \oplus \mathcal{W} \implies \dim(\mathcal{V}) = \dim(\mathcal{U}) + \dim(\mathcal{W}). \text{ (Why?)}$$

Example 7 Two non-colinear lines through the origin form a direct sum decomposition of \mathbb{R}^2 (Why?), and a plane and non-coincident line through the origin form a direct sum decomposition of \mathbb{R}^3 (Why?).

Proof of Theorem 4: Let $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$ be a basis for $\text{Nul}(\mathbf{L})$ where $n = \dim(\text{Nul}(\mathbf{L}))$. Fill out this set to form a basis for \mathcal{V} , $\{\mathbf{g}_1, \dots, \mathbf{g}_{N-n}, \mathbf{f}_1, \dots, \mathbf{f}_n\}$ (How?). Define the subspace $\tilde{\mathcal{V}} := \text{span}(\{\mathbf{g}_1, \dots, \mathbf{g}_{N-n}\})$. Then $\tilde{\mathcal{V}}$ and $\text{Nul}(\mathbf{L})$ form a direct sum decomposition of \mathcal{V} (Why?), that is

$$\mathcal{V} = \tilde{\mathcal{V}} \oplus \text{Nul}(\mathbf{L}).$$

Moreover, one can easily show that \mathbf{L} restricted to $\tilde{\mathcal{V}}$ is a one-to-one transformation onto $\text{Ran}(\mathbf{L})$ (Why?). It follows that $\{\mathbf{L}\mathbf{g}_1, \dots, \mathbf{L}\mathbf{g}_{N-n}\}$ forms a basis for $\text{Ran}(\mathbf{L})$ (Why?) and hence that $\dim(\text{Ran}(\mathbf{L})) = \dim(\tilde{\mathcal{V}}) = N - n$ (Why?) from which one concludes (9). \square

3.1 Matrices and Linear Transformations

Let $\mathbf{L} : \mathcal{V} \longrightarrow \mathcal{U}$ be a linear transformation and let $B := \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ be a basis for \mathcal{V} and $C := \{\mathbf{c}_1, \dots, \mathbf{c}_M\}$ a basis for \mathcal{U} . Then the action of the linear transformation \mathbf{L} can be carried out via matrix multiplication of coordinates in the following manner. Given $\mathbf{v} \in \mathcal{V}$, define $\mathbf{u} := \mathbf{L}\mathbf{v}$. Then the *Matrix of \mathbf{L} with respect to the bases B and C* is defined to be the unique matrix $L = [\mathbf{L}]_{B,C}$ satisfying

$$[\mathbf{u}]_C = L[\mathbf{v}]_B \quad (\text{matrix multiplication})$$

or in more expanded form

$$[\mathbf{L}\mathbf{v}]_C = [\mathbf{L}]_{B,C}[\mathbf{v}]_B. \tag{10}$$

To see how to construct $[\mathbf{L}]_{B,C}$, first construct the N column vectors (M -tuples) $\mathbf{l}_j := [\mathbf{L}\mathbf{b}_j]_C$, that is the coordinates with respect to the basis C in \mathcal{U} of the image vectors $\mathbf{L}\mathbf{b}_j$ of the base vectors in B . Then, one easily shows that defining

$$[\mathbf{L}]_{B,C} := [\mathbf{l}_1 \dots \mathbf{l}_N] \quad (\text{matrix whose columns are } \mathbf{l}_1, \dots, \mathbf{l}_N) \tag{11}$$

gives the unique $M \times N$ matrix satisfying (10). Indeed, if $\mathbf{v} \in \mathcal{V}$ has components $[\mathbf{v}]_B = [v_j]$, that is $\mathbf{v} = v_j \mathbf{b}_j$, then

$$[\mathbf{L}\mathbf{v}]_C = [v_j \mathbf{L}\mathbf{b}_j]_C = [\mathbf{l}_1 \dots \mathbf{l}_N][\mathbf{v}]_B$$

as required to prove (11).

Remark 11 If $\mathcal{U} = \mathcal{V}$ and $B = C$, then the matrix of \mathbf{L} with respect to B is denoted $[\mathbf{L}]_B$ and is constructed by using the B -coordinates of the vectors $\mathbf{L}\mathbf{b}_j$, $[\mathbf{L}\mathbf{b}_j]_B$, as its columns.

Example 8 The identity transformation $\mathbf{I} : \mathcal{V} \longrightarrow \mathcal{V}$ ($\mathbf{I}\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathcal{V}$) has

$$[\mathbf{I}]_B = [\delta_{ij}]$$

where δ_{ij} is the Kronecker symbol. (Why?)

3.2 Change of Basis

Suppose $\mathbf{L} : \mathcal{V} \longrightarrow \mathcal{V}$ and $B = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ and $\tilde{B} = \{\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_N\}$ are two bases for \mathcal{V} . The question of interest here is: How are the matrices $[\mathbf{L}]_B$ and $[\mathbf{L}]_{\tilde{B}}$ related? The question is easily answered by using the change of basis formula (3). From the definition of $[\mathbf{L}]_B$ and $[\mathbf{L}]_{\tilde{B}}$ one has

$$[\mathbf{L}\mathbf{v}]_B = [\mathbf{L}]_B[\mathbf{v}]_B \quad \text{and} \quad [\mathbf{L}\mathbf{v}]_{\tilde{B}} = [\mathbf{L}]_{\tilde{B}}[\mathbf{v}]_{\tilde{B}}.$$

It follows from (3) that

$$[\mathbf{L}\mathbf{v}]_{\tilde{B}} = T^{-1}[\mathbf{L}\mathbf{v}]_B = T^{-1}[\mathbf{L}]_B[\mathbf{v}]_B = T^{-1}[\mathbf{L}]_B T[\mathbf{v}]_{\tilde{B}}$$

from which one concludes

$$[\mathbf{L}]_{\tilde{B}} = T^{-1}[\mathbf{L}]_B T. \tag{12}$$

Remark 12 Square matrices A and B satisfying

$$A = T^{-1} B T$$

for some (non-singular) matrix T are said to be *Similar*. Thus, matrices related to a given linear transformation through a change of basis are similar and conversely similar matrices correspond the same linear transformation under a change of basis.

3.3 Trace and Determinant of a Linear Transformation

Consider linear transformations on a given finite dimensional vector space \mathcal{V} , i.e. $\mathbf{L} : \mathcal{V} \longrightarrow \mathcal{V}$. Then its *Trace* is defined by

Definition 16 The trace of \mathbf{L} , denoted $\text{tr}(\mathbf{L})$, is defined by

$$\text{tr}(\mathbf{L}) := \text{tr}([\mathbf{L}]_B)$$

for any matrix representation of \mathbf{L} .

To see that this notion of trace of a linear transformation is well (unambiguously) defined, let $[\mathbf{L}]_B$ and $[\mathbf{L}]_{\tilde{B}}$ denote two different matrix representations of \mathbf{L} . Then they must be similar matrices, i.e. there exists a non-singular matrix T such that

$$[\mathbf{L}]_{\tilde{B}} = T^{-1}[\mathbf{L}]_B T.$$

But then from the elementary property of the trace of a square matrix, $\text{tr}(AB) = \text{tr}(BA)$ for any two square matrices A, B (Why true?), one has

$$\text{tr}([\mathbf{L}]_{\tilde{B}}) = \text{tr}(T^{-1}[\mathbf{L}]_B T) = \text{tr}(T T^{-1}[\mathbf{L}]_B) = \text{tr}([\mathbf{L}]_B).$$

In similar fashion, one can define the *Determinant* of a linear transformation on \mathcal{V} .

Definition 17 The determinant of a linear transformation \mathbf{L} on a finite dimensional vector space \mathcal{V} is defined by

$$\det(\mathbf{L}) := \det([\mathbf{L}]_B)$$

where $[\mathbf{L}]_B$ is any matrix representation of \mathbf{L} .

That this notion of the determinant of a linear transformation is well defined can be seen from an argument similar to that above for the trace. The relevant property of the determinant of square matrices needed in the argument is $\det(AB) = \det(A)\det(B)$ (Why true?).

4 Metric, Norm and Inner Product

In this section, the algebraic structure for vector spaces presented above is augmented through the inclusion of geometric structure. Perhaps the most basic geometric notion is that of distance or length. A convenient way to introduce a notion of distance into a vector space is through use of a *Metric*.

Definition 18 A *Metric* on a vector space \mathcal{V} is a real valued function of two variables $d(\cdot, \cdot) : \mathcal{V} \longrightarrow \mathbb{R}$ satisfying the following axioms. For any $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}$,

$$\begin{aligned} d(\mathbf{a}, \mathbf{b}) &\geq 0, \quad \text{with} \quad d(\mathbf{a}, \mathbf{b}) = 0 \Leftrightarrow \mathbf{a} = \mathbf{b} && (\text{Positivity}) \\ d(\mathbf{a}, \mathbf{b}) &= d(\mathbf{b}, \mathbf{a}) && (\text{Symmetry}) \\ d(\mathbf{a}, \mathbf{b}) &\leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b}). && (\text{Triangle Inequality}) \end{aligned} \tag{13}$$

Remark 13 Given a metric on a vector space, one can define the important concept of a *metric ball*. Specifically, an (open) metric ball, $B(\mathbf{c}, r)$, centered at \mathbf{c} and of radius r is defined to be the set

$$B(\mathbf{c}, r) := \{\mathbf{v} \in \mathcal{V} \mid d(\mathbf{c}, \mathbf{v}) < r\}.$$

One can also define a notion of *metric convergent sequence* as follows. A sequence $\{\mathbf{a}_j, j = 1, 2, \dots\} \subset \mathcal{V}$ is said to be metric convergent to $\mathbf{a} \in \mathcal{V}$ provided

$$\lim_{j \rightarrow \infty} d(\mathbf{a}_j, \mathbf{a}) = 0.$$

An equivalent definition in terms of metric balls states that for all $\epsilon > 0$, there exists $N > 0$ such that

$$\mathbf{a}_j \in B(\mathbf{a}, \epsilon) \quad \text{for all} \quad j > N.$$

An important class of metrics on a vector space comes from the notion of a *Norm*.

Definition 19 A *Norm* on a vector space is a non-negative, real valued function $|\cdot| : \mathcal{V} \longrightarrow \mathbb{R}$ satisfying the following axioms. For all $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ and $\alpha \in \mathcal{F}$,

$$\begin{aligned} |\mathbf{a}| &\geq 0, \quad \text{with} \quad |\mathbf{a}| = 0 \Leftrightarrow \mathbf{a} = \mathbf{0} && (\text{Positivity}) \\ |\alpha \mathbf{a}| &= |\alpha| |\mathbf{a}| && (\text{Homogeneity}) \\ |\mathbf{a} + \mathbf{b}| &\leq |\mathbf{a}| + |\mathbf{b}|. && (\text{Triangle Inequality}) \end{aligned} \tag{14}$$

Remark 14 A norm has an associated metric defined by

$$d(\mathbf{a}, \mathbf{b}) := |\mathbf{a} - \mathbf{b}|.$$

An important class of norms on a vector space comes from the notion of *Inner Product*. Here a distinction must be made between real and complex vector spaces.

Definition 20 An inner product on a real vector space is a positive definite, symmetric bilinear form $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \longleftrightarrow \mathbb{R}$ satisfying the following axioms. For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}$ and $\alpha \in \mathbb{R}$,

$$\begin{aligned} \langle \mathbf{a}, \mathbf{a} \rangle &\geq 0 \quad \text{with} \quad \langle \mathbf{a}, \mathbf{a} \rangle = 0 \Leftrightarrow \mathbf{a} = \mathbf{0} && (\text{Positivity}) \\ \langle \mathbf{a}, \mathbf{b} \rangle &= \langle \mathbf{b}, \mathbf{a} \rangle && (\text{Symmetry}) \\ \langle \alpha \mathbf{a}, \mathbf{b} \rangle &= \alpha \langle \mathbf{a}, \mathbf{b} \rangle && (\text{Homogeneity}) \\ \langle \mathbf{a} + \mathbf{c}, \mathbf{b} \rangle &= \langle \mathbf{a}, \mathbf{b} \rangle + \langle \mathbf{c}, \mathbf{b} \rangle. && (\text{Additivity}) \end{aligned} \tag{15}$$

Definition 21 An inner product on a complex vector space is a positive definite, conjugate-symmetric form $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \longleftrightarrow \mathbb{C}$ satisfying the following axioms. For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}$ and $\alpha \in \mathbb{C}$,

$$\begin{aligned} \langle \mathbf{a}, \mathbf{a} \rangle &\geq 0 \quad \text{with} \quad \langle \mathbf{a}, \mathbf{a} \rangle = 0 \Leftrightarrow \mathbf{a} = \mathbf{0} && (\text{Positivity}) \\ \langle \mathbf{a}, \mathbf{b} \rangle &= \overline{\langle \mathbf{b}, \mathbf{a} \rangle} && (\text{Conjugate-Symmetry}) \\ \langle \alpha \mathbf{a}, \mathbf{b} \rangle &= \alpha \langle \mathbf{a}, \mathbf{b} \rangle && (\text{Homogeneity}) \\ \langle \mathbf{a} + \mathbf{c}, \mathbf{b} \rangle &= \langle \mathbf{a}, \mathbf{b} \rangle + \langle \mathbf{c}, \mathbf{b} \rangle && (\text{Additivity}) \end{aligned} \tag{16}$$

where in (16), $\overline{\langle \mathbf{b}, \mathbf{a} \rangle}$ denotes the complex conjugate of $\langle \mathbf{b}, \mathbf{a} \rangle$.

It is evident from the definition that a complex inner product is linear in the first variable but conjugate linear in the second. In particular, for all $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ and $\alpha \in \mathbb{C}$

$$\langle \alpha \mathbf{a}, \mathbf{b} \rangle = \alpha \langle \mathbf{a}, \mathbf{b} \rangle \quad \text{whereas} \quad \langle \mathbf{a}, \alpha \mathbf{b} \rangle = \overline{\alpha} \langle \mathbf{a}, \mathbf{b} \rangle.$$

Remark 15 Given an inner product on either a real or complex vector space, one can define an associated norm by

$$|\mathbf{a}| := \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}.$$

(Why does this define a norm?)

Remark 16 In \mathbb{R}^N , the usual inner product, defined by

$$\langle \mathbf{a}, \mathbf{b} \rangle := a_1 b_1 + \dots + a_N b_N = a_j b_j \quad \text{where} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix},$$

is often denoted by $\mathbf{a} \cdot \mathbf{b}$, the so-called “dot product”. In \mathbb{C}^N , the usual inner product is given by

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a} \cdot \overline{\mathbf{b}} = a_1 \overline{b_1} + \dots + a_N \overline{b_N}$$

where now $a_j b_j \in \mathbb{C}$ and the over-bar denotes complex conjugation.

Example 9 Let A be a positive definite, symmetric, real, $N \times N$ matrix.⁵ Then A defines an inner product on \mathbb{R}^N through

$$\langle \mathbf{a}, \mathbf{b} \rangle_A := \mathbf{a} \cdot (A\mathbf{b}).$$

⁵Recall that *symmetric* means that $A^T = A$, where A^T denotes the transpose matrix, and *positive definite* means $\mathbf{a} \cdot (A\mathbf{a}) > 0$ for all $\mathbf{a} \in \mathbb{R}^N$.

Example 10 An important class of norms on \mathbb{R}^N and \mathbb{C}^N are the p -norms defined for $p \geq 1$ by

$$\|\mathbf{a}\|_p := \left(\sum_{j=1}^N |a_j|^p \right)^{1/p}. \quad (17)$$

One must show that (17) satisfies the axioms (14) for a norm. The positivity and symmetry axioms from (14) are straight forward to show, but the triangle inequality is not. A proof that (17) satisfies the triangle inequality in (14) can be found in the supplemental *Notes on Inequalities*. The special case $p = 2$ is the famous *Minkowski Inequality*

$$\|\mathbf{a} + \mathbf{b}\|_2 \leq \|\mathbf{a}\|_2 + \|\mathbf{b}\|_2$$

for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$.

Example 11 The ∞ -norm is defined by

$$\|\mathbf{a}\|_\infty := \max\{|a_1|, \dots, |a_N|\}. \quad (18)$$

The task of showing that (18) satisfies the axioms (14) for a norm is left as an exercise. An important observation about the p -norms is that

$$\lim_{p \rightarrow \infty} \|\mathbf{a}\|_p = \|\mathbf{a}\|_\infty$$

whose proof is also left as an exercise. That the ∞ -norm is the limit of the p -norms as $p \rightarrow \infty$ is what motivates the definition (18).

Remark 17 The 2-norm is very special among the class of p -norms being the only one coming from an inner product. The 2-norm is also called the *Euclidean Norm* since it comes from the usual inner product on the Euclidean space \mathbb{R}^N .

Remark 18 In a real vector space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$, one can define a generalized notion of *angle between vectors*. Specifically, given $\mathbf{a}, \mathbf{b} \in \mathcal{V}$, one defines the angle, θ , between \mathbf{a} and \mathbf{b} to be the unique number $0 < \theta < \pi$ satisfying

$$\cos \theta = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \|\mathbf{b}\|}.$$

As a special case, two vectors $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ are said to be *Orthogonal* (in symbols $\mathbf{a} \perp \mathbf{b}$) if

$$\langle \mathbf{a}, \mathbf{b} \rangle = 0. \quad (19)$$

Remark 19 Analytically or topologically all norms on a finite dimensional vector space are *equivalent* in the sense that if a sequence is convergent with respect to one norm, it is convergent with respect to all norms on the space. This observation follows directly from the following important

Theorem 5 Let \mathcal{V} be a finite dimensional vector space and let $\|\cdot\|_I$ and $\|\cdot\|_{II}$ be two norms on \mathcal{V} . Then there exist positive constants m, M such that for all $\mathbf{v} \in \mathcal{V}$

$$m\|\mathbf{v}\|_I \leq \|\mathbf{v}\|_{II} \leq M\|\mathbf{v}\|_I.$$

Proof: The proof of this result is left as an exercise.

While all norms on a finite dimensional vector space might be topologically equivalent, they are not geometrically equivalent. In particular, the “shape” of the (metric) unit ball can vary greatly for different norms. That fact is illustrated in the following

Exercise 1 Sketch the unit ball in \mathbb{R}^2 for the p -norms, $1 \leq p \leq \infty$.

4.1 Orthonormal Basis

In a finite dimensional inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$, an important class of bases consists of the *Orthonormal Bases* defined as follows.

Definition 22 In an inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$, a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ is called *Orthonormal* provided

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = \delta_{ij} \quad \text{for } i, j = 1, \dots, N. \quad (20)$$

Thus, an orthonormal basis consists of pairwise orthogonal unit vectors. Finding coordinates of arbitrary vectors with respect to such a basis is as convenient as it is for the natural basis in \mathbb{F}^N . Indeed, what makes the natural basis so useful is that it is an orthonormal basis with respect to the usual inner product on \mathbb{R}^N .

Remark 20 Suppose $B = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ is an orthonormal basis for $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ and $\mathbf{a} \in \mathcal{V}$. Then one easily deduces that

$$\mathbf{a} = a_1 \mathbf{b}_1 + \dots + a_N \mathbf{b}_N \implies a_j = \langle \mathbf{a}, \mathbf{b}_j \rangle$$

as can be seen by taking the inner product of the first equation with each base vector \mathbf{b}_j . Moreover, one has

$$\langle \mathbf{a}, \mathbf{c} \rangle = [\mathbf{a}]_B \cdot [\mathbf{c}]_B \quad (21)$$

where the right hand side is the usual dot-product of N -tuples in \mathbb{R}^N

$$[\mathbf{a}]_B \cdot [\mathbf{c}]_B = a_1 c_1 + \dots + a_N c_N.$$

Thus, computing the inner product of vectors using components with respect to an orthonormal basis uses the familiar formula for the dot-product on \mathbb{R}^N with respect to the natural basis.

In similar fashion one can derive a simple formula for the components of the matrix representation of a linear transformation on an inner product space with respect to an orthonormal basis. Specifically, suppose $\mathbf{L} : \mathcal{V} \longrightarrow \mathcal{V}$ and $B = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ is an orthonormal basis. Then it is straightforward to show that

$$L := [\mathbf{L}]_B = [l_{ij}] \quad \text{with} \quad l_{ij} = \langle \mathbf{b}_i, \mathbf{L} \mathbf{b}_j \rangle. \quad (22)$$

One way to derive (22) is to first recall the formula

$$l_{ij} = \mathbf{e}_i \cdot (L\mathbf{e}_j) \quad (23)$$

for the components of an $N \times N$ matrix, $L := [l_{ij}]$, where $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ is the natural basis for \mathbb{R}^N . It follows from (21) that for all $\mathbf{a}, \mathbf{c} \in \mathcal{V}$

$$\langle \mathbf{a}, L\mathbf{c} \rangle = [\mathbf{a}]_B \cdot ([L\mathbf{c}]_B) = [\mathbf{a}]_B \cdot ([L]_B [\mathbf{c}]_B). \quad (24)$$

Now letting $\mathbf{a} = \mathbf{b}_i$, $\mathbf{c} = \mathbf{b}_j$ and noting that $[\mathbf{b}_j]_B = \mathbf{e}_j$ for all $j = 1, \dots, N$, one deduces (22) from (23,24).

4.1.1 General Bases and the Metric Tensor

More generally, one would like to be able to conveniently compute the inner product of two vectors using coordinates with respect to some general (not necessarily orthonormal) basis. Suppose $B := \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ is a basis for $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ and let $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ with coordinates $[\mathbf{a}]_B = [a_j]$, $[\mathbf{b}]_B = [b_j]$, respectively. Then one has

$$\langle \mathbf{a}, \mathbf{b} \rangle = \left\langle \sum_{j=1}^N a_j \mathbf{f}_j, \sum_{k=1}^N b_k \mathbf{f}_k \right\rangle = \sum_{j,k=1}^N a_j b_k \langle \mathbf{f}_j, \mathbf{f}_k \rangle = [\mathbf{a}]_B \cdot (F[\mathbf{b}]_B)$$

where the matrix

$$F := [\langle \mathbf{f}_i, \mathbf{f}_j \rangle] = [f_{ij}] \quad (25)$$

is called the *Metric Tensor* associated with the basis B . It follows that the familiar formula (21) for the dot-product in \mathbb{R}^N holds in general if and only if the basis B is orthonormal.

The matrix F is called a metric tensor because it is needed to compute lengths and angles using coordinates with respect to the basis B . In particular, if $\mathbf{v} \in \mathcal{V}$ has coordinates $[\mathbf{v}]_B = [v_j]$, then $|\mathbf{v}|$ (the length of \mathbf{v}) is given by

$$|\mathbf{v}|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = f_{ij} v_i v_j.$$

4.1.2 The Gram-Schmidt Orthogonalization Procedure

Given any basis $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ for an inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$, there is a simple algorithm, called the *Gram-Schmidt Orthogonalization Procedure*, for constructing from B an orthonormal basis $\tilde{B} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ satisfying

$$\text{span}\{\mathbf{f}_1, \dots, \mathbf{f}_K\} = \text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_K\} \quad \text{for each } 1 \leq K \leq N. \quad (26)$$

The algorithm proceeds inductively as follows.

Step 1: Define $\mathbf{e}_1 := \mathbf{f}_1 / \|\mathbf{f}_1\|$, where $\|\cdot\|$ denotes the norm associated with the inner product $\langle \cdot, \cdot \rangle$.

Step 2: Subtract from \mathbf{f}_2 its component in the direction of \mathbf{e}_1 and then normalize. More specifically, define

$$\mathbf{g}_2 := \mathbf{f}_2 - \langle \mathbf{f}_2, \mathbf{e}_1 \rangle \mathbf{e}_1$$

and then define

$$\mathbf{e}_2 := \frac{\mathbf{g}_2}{\|\mathbf{g}_2\|}.$$

It is easily seen that $\{\mathbf{e}_1, \mathbf{e}_2\}$ is an orthonormal pair satisfying (26) for $K = 1, 2$.

Inductive Step: Suppose orthonormal vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_J\}$ have been constructed satisfying (26) for $1 \leq K \leq J$ with $J < N$. Define

$$\mathbf{g}_{J+1} := \mathbf{f}_{J+1} - \langle \mathbf{f}_{J+1}, \mathbf{e}_1 \rangle \mathbf{e}_1 - \dots - \langle \mathbf{f}_{J+1}, \mathbf{e}_J \rangle \mathbf{e}_J$$

and then define

$$\mathbf{e}_{J+1} := \frac{\mathbf{g}_{J+1}}{\|\mathbf{g}_{J+1}\|}.$$

One easily sees that $\{\mathbf{e}_1, \dots, \mathbf{e}_{J+1}\}$ is an orthonormal set satisfying (26) for $1 \leq K \leq J+1$ thereby completing the inductive step.

4.2 Reciprocal Basis

Let $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space with general basis $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$. Then B has an associated basis, called its *Reciprocal or Dual Basis* defined through

Definition 23 Given a basis $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ on the finite dimensional inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$, there is a unique Reciprocal Basis $B^* := \{\mathbf{f}^1, \dots, \mathbf{f}^N\}$ satisfying

$$\langle \mathbf{f}_i, \mathbf{f}^j \rangle = \delta_{ij} \quad \text{for } i, j = 1, \dots, N. \quad (27)$$

Thus, every vector $\mathbf{v} \in \mathcal{V}$ has the two representations

$$\mathbf{v} = \sum_{i=1}^N v^i \mathbf{f}_i \quad (\text{Contravariant Expansion}) \quad (28)$$

$$= \sum_{i=1}^N v_i \mathbf{f}^i. \quad (\text{Covariant Expansion}) \quad (29)$$

The N -tuples $[\mathbf{v}]_B = [v^i]$ and $[\mathbf{v}]_{B^*} = [v_i]$ are called the *Contravariant Coordinates* and *Covariant Coordinates* of \mathbf{v} , respectively. It follows immediately from (27,28,29) that the contravariant and covariant coordinates of \mathbf{v} can be conveniently computed via the formulae

$$v^i = \langle \mathbf{v}, \mathbf{f}^i \rangle \quad \text{and} \quad v_i = \langle \mathbf{v}, \mathbf{f}_i \rangle. \quad (30)$$

Exercise 2 Let $B = \{\mathbf{f}_1, \mathbf{f}_2\}$ be any basis for the Cartesian plane \mathbb{R}^2 . Show how the reciprocal basis $B^* = \{\mathbf{f}^1, \mathbf{f}^2\}$ can be constructed graphically from B .

Exercise 3 Let $B = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ be any basis for the Cartesian space \mathbb{R}^3 . Show that the reciprocal basis $B^* = \{\mathbf{f}^1, \mathbf{f}^2, \mathbf{f}^3\}$ is given by

$$\mathbf{f}^1 = \frac{\mathbf{f}_2 \times \mathbf{f}_3}{\mathbf{f}_1 \cdot (\mathbf{f}_2 \times \mathbf{f}_3)}, \quad \mathbf{f}^2 = \frac{\mathbf{f}_1 \times \mathbf{f}_3}{\mathbf{f}_2 \cdot (\mathbf{f}_1 \times \mathbf{f}_3)}, \quad \mathbf{f}^3 = \frac{\mathbf{f}_1 \times \mathbf{f}_2}{\mathbf{f}_3 \cdot (\mathbf{f}_1 \times \mathbf{f}_2)}$$

where $\mathbf{a} \times \mathbf{b}$ denotes the vector cross product of \mathbf{a} and \mathbf{b} .

5 Linear Transformations Revisited

The theory of linear transformations between finite dimensional vector spaces is revisited in this section. After some general considerations, several important special classes of linear transformations will be studied from both an algebraic and a geometric perspective. The first notion introduced is the vector space $Lin(\mathcal{V}, \mathcal{U})$ of linear transformations between two vector spaces.

5.1 The Vector Space $Lin(\mathcal{V}, \mathcal{U})$

Definition 24 *Given two vector spaces, \mathcal{V}, \mathcal{U} , $Lin(\mathcal{V}, \mathcal{U})$ denotes the collection of all linear transformations between \mathcal{V} and \mathcal{U} . It is a vector space under the operations of addition and scalar multiplication of linear transformations defined as follows. Given two linear transformations $\mathbf{L}_1, \mathbf{L}_2 : \mathcal{V} \longrightarrow \mathcal{U}$ their sum, $(\mathbf{L}_1 + \mathbf{L}_2)$, is defined to be the linear transformation satisfying*

$$(\mathbf{L}_1 + \mathbf{L}_2)\mathbf{v} := \mathbf{L}_1\mathbf{v} + \mathbf{L}_2\mathbf{v} \quad \text{for all } \mathbf{v} \in \mathcal{V}.$$

Moreover, scalar multiplication of a linear transformation is defined by

$$(\alpha\mathbf{L})\mathbf{v} := \alpha\mathbf{L}\mathbf{v} \quad \text{for all } \alpha \in \mathcal{F} \quad \text{and } \mathbf{v} \in \mathcal{V}.$$

For the special case $\mathcal{V} = \mathcal{U}$, one denotes the vector space of linear transformations on \mathcal{V} by $Lin(\mathcal{V})$.

5.1.1 Adjoint Transformation

Let $\mathbf{L} \in Lin(\mathcal{V}, \mathcal{U})$ be a linear transformation between two inner product spaces $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ and $(\mathcal{U}, \langle \cdot, \cdot \rangle)$. The *Adjoint Transformation*, \mathbf{L}^* , is defined to be the unique transformation in $Lin(\mathcal{U}, \mathcal{V})$ satisfying

$$\langle \mathbf{L}\mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{L}^*\mathbf{u} \rangle \quad \text{for all } \mathbf{v} \in \mathcal{V} \quad \text{and } \mathbf{u} \in \mathcal{U}. \quad (31)$$

Remark 21 The right hand side of (31) is the inner product on \mathcal{V} whereas the left hand side is the inner product on \mathcal{U} .

Remark 22 This definition of adjoint makes sense due to the

Theorem 6 (*Riesz Representation Theorem*) *Suppose $\mathbf{l} : \mathcal{V} \longrightarrow \mathcal{F}$ is a linear, scalar valued transformation on the inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$. Then there exists a unique vector $\mathbf{a} \in \mathcal{V}$ such that*

$$\mathbf{l}\mathbf{v} = \langle \mathbf{v}, \mathbf{a} \rangle \quad \text{for all } \mathbf{v} \in \mathcal{V}.$$

Note that for fixed $\mathbf{u} \in \mathcal{U}$, $\mathbf{l}\mathbf{v} := \langle \mathbf{L}\mathbf{v}, \mathbf{u} \rangle$ is a scalar valued linear transformation on \mathcal{V} . By the Riesz Representation Theorem, there exists a unique vector $\mathbf{a} \in \mathcal{V}$ such that $\mathbf{l}\mathbf{v} = \langle \mathbf{v}, \mathbf{a} \rangle$ for all $\mathbf{v} \in \mathcal{V}$. One now defines $\mathbf{L}^*\mathbf{u} := \mathbf{a} \in \mathcal{V}$.

Example 12 Suppose $\mathcal{V} = \mathbb{R}^N$, $\mathcal{U} = \mathbb{R}^M$ and $\mathbf{L} \in \mathcal{M}^{M,N}(\mathbb{R})$, the space of all $M \times N$ (real) matrices. Then $\mathbf{L}^* = \mathbf{L}^T$, the transpose of \mathbf{L} . (Why?) On the other hand, if $\mathcal{V} = \mathbb{C}^N$, $\mathcal{U} = \mathbb{C}^M$ and $\mathbf{L} \in \mathcal{M}^{M,N}(\mathbb{C})$, the space of all $M \times N$ (complex) matrices. Then $\mathbf{L}^* = \overline{\mathbf{L}}^T$, the conjugate transpose of \mathbf{L} . (Why?)

Remark 23 Consider the special case $\mathcal{V} = \mathcal{U}$ (over $\mathcal{F} = \mathbb{R}$) and let $B := \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ be any basis for \mathcal{V} . Then one can show that for any $\mathbf{A} \in \text{Lin}(\mathcal{V})$

$$[\mathbf{A}^*]_B = F^{-1}[\mathbf{A}]_B^T F \quad (32)$$

where the matrix F is given by

$$F := [f_{ij}]$$

with

$$f_{ij} := \langle \mathbf{f}_i, \mathbf{f}_j \rangle.$$

The proof of (32) is left as an exercise.

5.1.2 Two Natural Norms for $\text{Lin}(\mathcal{V}, \mathcal{U})$

One can define a natural inner product on $\text{Lin}(\mathcal{V})$ by

$$\mathbf{L}_1 \cdot \mathbf{L}_2 := \text{tr}(\mathbf{L}_1^* \mathbf{L}_2) \quad (33)$$

where \mathbf{L}_1^* denotes the adjoint of \mathbf{L}_1 . (Why does this satisfy the axioms (15) or (16)?)

Remark 24 In the case $\mathcal{V} = \mathbb{R}^N$ with \mathbf{L}_1 and \mathbf{L}_2 given by $N \times N$ - matrices $\mathbf{L}_1 = [l_{ij}^{(1)}]$, $\mathbf{L}_2 = [l_{ij}^{(2)}]$, respectively, then one readily shows (How?) that

$$\mathbf{L}_1 \cdot \mathbf{L}_2 = \sum_{i,j=1}^N l_{ij}^{(1)} l_{ij}^{(2)}.$$

One can readily show that (33) also defines an inner product on $\text{Lin}(\mathcal{V}, \mathcal{U})$ where \mathcal{V} and \mathcal{U} are any two finite dimensional vector spaces. (How?) The norm associated to this inner product, called the *trace norm*, is then defined by

$$|\mathbf{L}| := \sqrt{\mathbf{L} \cdot \mathbf{L}} = \sqrt{\text{tr}(\mathbf{L}^* \mathbf{L})}. \quad (34)$$

Another natural norm on $\text{Lin}(\mathcal{V}, \mathcal{U})$, where $(\mathcal{V}, |\cdot|)$ and $(\mathcal{U}, |\cdot|)$ are normed vector spaces, called the *operator norm*, is defined by

$$\|\mathbf{L}\| := \sup_{|\mathbf{v}|=1} |\mathbf{L}\mathbf{v}|. \quad (35)$$

Question: What is the geometric interpretation of (35)? An important relationship between the trace norm and operator norm is given by

Theorem 7 Let $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ and $(\mathcal{U}, \langle \cdot, \cdot \rangle)$ be two finite dimensional, inner product spaces. Then for all $\mathbf{L} \in \text{Lin}(\mathcal{V}, \mathcal{U})$,

$$\|\mathbf{L}\| \leq |\mathbf{L}| \leq \sqrt{N} \|\mathbf{L}\|. \quad (36)$$

Proof: The proof uses ideas to be introduced in the up coming study of spectral theory. To see how spectral theory plays a role in proving (36), first notice that

$$\|\mathbf{L}\| = \sup_{|\mathbf{v}|=1} |\mathbf{L}\mathbf{v}| = \sup_{|\mathbf{v}|=1} \sqrt{\langle \mathbf{L}\mathbf{v}, \mathbf{L}\mathbf{v} \rangle} = \sup_{|\mathbf{v}|=1} \sqrt{\langle \mathbf{v}, \mathbf{L}^* \mathbf{L} \mathbf{v} \rangle}.$$

Next observe that $\mathbf{L}^* \mathbf{L}$ is a self-adjoint, positive linear transformation in $\text{Lin}(\mathcal{V})$. As part of spectral theory for self-adjoint operators on a finite dimensional inner product space, it is shown that

$$\sup_{|\mathbf{v}|=1} \langle \mathbf{v}, \mathbf{L}^* \mathbf{L} \mathbf{v} \rangle = \max_{j=1, \dots, N} |\lambda_j|^2$$

where $\{|\lambda_j|^2, j = 1, \dots, N\}$ are the (positive) eigenvalues of $\mathbf{L}^* \mathbf{L}$. Thus, one has

$$\|\mathbf{L}\| = \max_{j=1, \dots, N} |\lambda_j|. \quad (37)$$

On the other hand, spectral theory will show that for the trace norm, one has

$$|\mathbf{L}|^2 = \text{tr}(\mathbf{L}^* \mathbf{L}) = \sum_{j=1}^N |\lambda_j|^2. \quad (38)$$

Appealing to the obvious inequalities

$$\max_{j=1, \dots, N} |\lambda_j|^2 \leq \sum_{j=1}^N |\lambda_j|^2 \leq N \max_{j=1, \dots, N} |\lambda_j|^2,$$

one readily deduces (36) from (37,38). \square

5.1.3 Elementary Tensor Product

Among the most useful linear transformations is the class of *Elementary Tensor Products* which can be used as the basic building blocks of all other linear transformations. First consider the case of $\text{Lin}(\mathcal{V})$ where $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ is an inner product space.

Definition 25 Let $\mathbf{a}, \mathbf{b} \in \mathcal{V}$. Then the elementary tensor product of \mathbf{a} and \mathbf{b} , denoted $\mathbf{a} \otimes \mathbf{b}$, is defined to be the linear transformation on \mathcal{V} satisfying

$$(\mathbf{a} \otimes \mathbf{b}) \mathbf{v} := \langle \mathbf{v}, \mathbf{b} \rangle \mathbf{a} \quad \text{for all } \mathbf{v} \in \mathcal{V}. \quad (39)$$

Remark 25 It follows immediately from the definition that $\text{Ran}(\mathbf{a} \otimes \mathbf{b}) = \text{span}(\mathbf{a})$ and $\text{Nul}(\mathbf{a} \otimes \mathbf{b}) = \mathbf{b}^\perp$, where \mathbf{b}^\perp denote the orthogonal complement of \mathbf{b} , i.e. the $(N - 1)$ -dimensional subspace of all vectors orthogonal to \mathbf{b} (where $\dim(\mathcal{V}) = N$). The elementary tensor products are thus the rank one linear transformations in $\text{Lin}(\mathcal{V})$.

Remark 26 The following identities can also be derived easily from the definition of elementary tensor product. Given any $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathcal{V}$ and $A \in \text{Lin}(\mathcal{V})$

$$\begin{aligned} (\mathbf{a} \otimes \mathbf{b})^* &= \mathbf{b} \otimes \mathbf{a} \\ (\mathbf{a} \otimes \mathbf{b}) \cdot (\mathbf{c} \otimes \mathbf{d}) &= \langle \mathbf{a}, \mathbf{c} \rangle \langle \mathbf{b}, \mathbf{d} \rangle \\ A \cdot (\mathbf{a} \otimes \mathbf{b}) &= \langle \mathbf{a}, A\mathbf{b} \rangle \\ (\mathbf{a} \otimes \mathbf{b})(\mathbf{c} \otimes \mathbf{d}) &= \langle \mathbf{b}, \mathbf{c} \rangle \mathbf{a} \otimes \mathbf{d} \\ A(\mathbf{a} \otimes \mathbf{b}) &= (A\mathbf{a}) \otimes \mathbf{b} \\ (\mathbf{a} \otimes \mathbf{b})A &= \mathbf{a} \otimes (A^*\mathbf{b}) \\ \text{tr}(\mathbf{a} \otimes \mathbf{b}) &= \langle \mathbf{a}, \mathbf{b} \rangle. \end{aligned}$$

Example 13 When $\mathcal{V} = \mathbb{R}^N$ and $\mathbf{a} = (a_i)$, $\mathbf{b} = (b_i)$, then the matrix of $\mathbf{a} \otimes \mathbf{b}$ with respect to the natural basis on \mathbb{R}^N is

$$[\mathbf{a} \otimes \mathbf{b}]_{\mathcal{N}} = [a_i b_j]. \quad (\text{Why?}). \quad (40)$$

Remark 27 The result (40) has a natural generalization to a general finite dimensional inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$. Let $B := \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ be an orthonormal basis for \mathcal{V} and $\mathbf{a}, \mathbf{b} \in \mathcal{V}$. Then

$$[\mathbf{a} \otimes \mathbf{b}]_B = [a_i b_j] \quad (41)$$

where $[\mathbf{a}]_B = [a_i]$ and $[\mathbf{b}]_B = [b_i]$. Hence, the formula (40) generalizes to the matrix representation of an elementary tensor product with respect to any orthonormal basis for \mathcal{V} . More generally, if $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ is any (not necessarily orthonormal) basis for \mathcal{V} , then

$$[\mathbf{a} \otimes \mathbf{b}]_B = [a_i b_j] F \quad (\text{matrix product})$$

where F is the metric tensor associated with the basis B defined by (25). Verification of this last formula is left as an exercise.

The notion of elementary tensor product is readily generalized to the setting of linear transformations between two inner product spaces $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ and $(\mathcal{U}, \langle \cdot, \cdot \rangle)$ as follows.

Definition 26 Let $\mathbf{a} \in \mathcal{U}$ and $\mathbf{b} \in \mathcal{V}$. Then the elementary tensor product $\mathbf{a} \otimes \mathbf{b}$ is defined to be the linear transformation in $\text{Lin}(\mathcal{V}, \mathcal{U})$ satisfying

$$(\mathbf{a} \otimes \mathbf{b})\mathbf{v} := \langle \mathbf{v}, \mathbf{b} \rangle \mathbf{a} \quad \text{for all } \mathbf{v} \in \mathcal{V}.$$

5.1.4 A Natural Orthonormal Basis for $\text{Lin}(\mathcal{V}, \mathcal{U})$

Using the elementary tensor product, one can construct a natural orthonormal basis for $\text{Lin}(\mathcal{V}, \mathcal{U})$. First consider the special case of $\text{Lin}(\mathcal{V})$. Suppose $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ is an inner product space with orthonormal basis $B := \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$. Then one easily shows that the set

$$\mathcal{B} := \{\mathbf{e}_i \otimes \mathbf{e}_j, i, j = 1, \dots, N\}$$

provides an orthonormal basis for $\text{Lin}(\mathcal{V})$ equipped with the inner product (33). It follows immediately that

$$\dim(\text{Lin}(\mathcal{V})) = N^2.$$

Remark 28 If $\mathbf{L} \in \text{Lin}(\mathcal{V})$, then

$$\mathbf{L} = l_{ij} \mathbf{e}_i \otimes \mathbf{e}_j \quad \text{with} \quad l_{ij} = \langle \mathbf{L}, \mathbf{e}_i \otimes \mathbf{e}_j \rangle = \langle \mathbf{e}_i, \mathbf{L} \mathbf{e}_j \rangle. \quad (\text{Why?}) \quad (42)$$

More generally, suppose $\mathcal{B}_{\mathcal{U}} = \{\mathbf{b}_1, \dots, \mathbf{b}_M\}$ is an orthonormal basis for \mathcal{U} and $\mathcal{B}_{\mathcal{V}} = \{\mathbf{d}_1, \dots, \mathbf{d}_N\}$ is an orthonormal basis for \mathcal{V} . Then a natural orthonormal basis for $\text{Lin}(\mathcal{V}, \mathcal{U})$ (with respect to the trace inner product on $\text{Lin}(\mathcal{V}, \mathcal{U})$) is

$$\{\mathbf{b}_i \otimes \mathbf{d}_j \mid i = 1, \dots, M, j = 1, \dots, N\}.$$

Indeed, that these tensor products are pairwise orthogonal and have trace norm one follows from

$$\begin{aligned} \langle \mathbf{b}_i \otimes \mathbf{d}_j, \mathbf{b}_k \otimes \mathbf{d}_l \rangle &= \text{tr}((\mathbf{b}_i \otimes \mathbf{d}_j)^* \mathbf{b}_k \otimes \mathbf{d}_l) \\ &= \text{tr}(\mathbf{d}_j \otimes \mathbf{b}_i \mathbf{b}_k \otimes \mathbf{d}_l) \\ &= \langle \mathbf{b}_i, \mathbf{b}_k \rangle \text{tr}(\mathbf{d}_j \otimes \mathbf{d}_l) \\ &= \langle \mathbf{b}_i, \mathbf{b}_k \rangle \langle \mathbf{d}_j, \mathbf{d}_l \rangle = \delta_{ik} \delta_{jl}. \end{aligned}$$

The generalization of (42) for $\mathbf{L} \in \text{Lin}(\mathcal{V}, \mathcal{U})$ is

$$\mathbf{L} = l_{ij} \mathbf{b}_i \otimes \mathbf{d}_j \quad \text{with} \quad l_{ij} := \langle \mathbf{L}, \mathbf{b}_i \otimes \mathbf{d}_j \rangle = \langle \mathbf{b}_i, \mathbf{L} \mathbf{d}_j \rangle. \quad (\text{Why?})$$

Remark 29 Suppose $B := \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ is an orthonormal basis for \mathcal{V} . Then every $\mathbf{L} \in \text{Lin}(\mathcal{V})$ has the associated matrix $[\hat{l}_{ij}]$ defined through

$$\mathbf{L} = \sum_{i,j=1}^N \hat{l}_{ij} \mathbf{e}_i \otimes \mathbf{e}_j.$$

On the other hand, \mathbf{L} also has the associated matrix representation $[\mathbf{L}]_B = [l_{ij}]$ defined through the relation

$$[\mathbf{L}\mathbf{v}]_B = [\mathbf{L}]_B [\mathbf{v}]_B$$

which must hold for all $\mathbf{v} \in \mathcal{V}$ with

$$l_{ij} = \langle \mathbf{e}_i, \mathbf{L} \mathbf{e}_j \rangle.$$

It is straight forward to show (How?) that these two matrices are equal, i.e.

$$[\mathbf{L}]_B = [l_{ij}] = [\hat{l}_{ij}].$$

5.1.5 General Tensor Product Bases for $\text{Lin}(\mathcal{V})$ and Their Duals

Let $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ be a general basis for $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ with associated reciprocal basis $B^* = \{\mathbf{f}^1, \dots, \mathbf{f}^N\}$. Then $\text{Lin}(\mathcal{V})$ has four associated tensor product bases and corresponding reciprocal bases given by

$$\{\mathbf{f}_i \otimes \mathbf{f}_j\} \longleftrightarrow \{\mathbf{f}^i \otimes \mathbf{f}^j\} \quad (43)$$

$$\{\mathbf{f}^i \otimes \mathbf{f}^j\} \longleftrightarrow \{\mathbf{f}_i \otimes \mathbf{f}_j\} \quad (44)$$

$$\{\mathbf{f}_i \otimes \mathbf{f}^j\} \longleftrightarrow \{\mathbf{f}^i \otimes \mathbf{f}_j\} \quad (45)$$

$$\{\mathbf{f}^i \otimes \mathbf{f}_j\} \longleftrightarrow \{\mathbf{f}_i \otimes \mathbf{f}^j\} \quad (46)$$

where $i, j = 1, \dots, N$ in each set. Every $\mathbf{L} \in \text{Lin}(\mathcal{V})$ has four matrix representations defined through

$$\mathbf{L} = \sum_{i,j=1}^N \hat{l}^{ij} \mathbf{f}_i \otimes \mathbf{f}_j \quad (\text{Pure Contravariant}) \quad (47)$$

$$= \sum_{i,j=1}^N \hat{l}_{ij} \mathbf{f}^i \otimes \mathbf{f}^j \quad (\text{Pure Covariant}) \quad (48)$$

$$= \sum_{i,j=1}^N \hat{l}_j^i \mathbf{f}_i \otimes \mathbf{f}^j \quad (\text{Mixed Contravariant-Covariant}) \quad (49)$$

$$= \sum_{i,j=1}^N \hat{l}_i^j \mathbf{f}^i \otimes \mathbf{f}_j \quad (\text{Mixed Covariant-Contravariant}). \quad (50)$$

It follows easily (How?) that one can compute these various matrix components of \mathbf{L} from the formulae

$$\begin{aligned} \hat{l}^{ij} &= \langle \mathbf{f}^i, \mathbf{L} \mathbf{f}^j \rangle \\ \hat{l}_{ij} &= \langle \mathbf{f}_i, \mathbf{L} \mathbf{f}_j \rangle \\ \hat{l}_j^i &= \langle \mathbf{f}^i, \mathbf{L} \mathbf{f}_j \rangle \\ \hat{l}_i^j &= \langle \mathbf{f}_i, \mathbf{L} \mathbf{f}^j \rangle. \end{aligned}$$

Exercise 4 Let $\mathbf{L} = \mathbf{I}$, the identity transformation on \mathcal{V} . Then \mathbf{I} has the four matrix representations (47,48,49,50) with

$$\begin{aligned} \hat{l}^{ij} &= \langle \mathbf{f}^i, \mathbf{f}^j \rangle = f^{ij} \\ \hat{l}_{ij} &= \langle \mathbf{f}_i, \mathbf{f}_j \rangle = f_{ij} \\ \hat{l}_j^i &= \langle \mathbf{f}^i, \mathbf{f}_j \rangle = \delta_j^i \\ \hat{l}_i^j &= \langle \mathbf{f}_i, \mathbf{f}^j \rangle = \delta_j^i \end{aligned}$$

where δ_j^i denotes the usual Kronecker symbol. Thus, the matrix for the identity transformation when using a general (non-orthonormal) basis has the accustomed form only using a mixed covariant and contravariant representation.

Remark 30 Equivalent Matrix Representations.

Given a general basis $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ and its dual basis $B^* = \{\mathbf{f}^1, \dots, \mathbf{f}^N\}$, one also has the four matrix representations of a linear transformation $\mathbf{L} \in \text{Lin}(\mathcal{V})$ defined through

$$\begin{aligned} [\mathbf{w}]_B &= [\mathbf{L}\mathbf{v}]_B = [\mathbf{L}]_{(B,B)}[\mathbf{v}]_B \\ [\mathbf{w}]_{B^*} &= [\mathbf{L}\mathbf{v}]_{B^*} = [\mathbf{L}]_{(B^*,B^*)}[\mathbf{v}]_{B^*} \\ [\mathbf{w}]_{B^*} &= [\mathbf{L}\mathbf{v}]_{B^*} = [\mathbf{L}]_{(B,B^*)}[\mathbf{v}]_B \\ [\mathbf{w}]_B &= [\mathbf{L}\mathbf{v}]_B = [\mathbf{L}]_{(B^*,B)}[\mathbf{v}]_{B^*} \end{aligned}$$

which in component form becomes

$$\begin{aligned} w^i &= l_j^i v^j \\ w_i &= l_i^j v_j \\ w_i &= l_{ij} v^j \\ w^i &= l^{ij} v_j. \end{aligned}$$

One now readily shows that

$$\hat{l}_{ij} = l_{ij}, \quad \hat{l}^{ij} = l^{ij}, \quad \hat{l}_j^i = l_j^i \quad \text{and} \quad \hat{l}_i^j = l_i^j.$$

Indeed, consider the following calculation:

$$\begin{aligned} \mathbf{L}\mathbf{v} &= \sum_{i,j} \hat{l}^{ij} \mathbf{f}_i \otimes \mathbf{f}_j (v_k \mathbf{f}^k) \\ &= \sum_{i,j} \hat{l}^{ij} \mathbf{f}_i v_k \langle \mathbf{f}_j, \mathbf{f}^k \rangle \\ &= \sum_{i,j} \hat{l}^{ij} \mathbf{f}_i v_k \delta_j^k \\ &= \sum_i \mathbf{f}_i \left(\sum_j \hat{l}^{ij} v_j \right) \\ &= \sum_i w^i \mathbf{f}_i \end{aligned}$$

where

$$w^i = \hat{l}^{ij} v_j.$$

It now follows from the definition of \hat{l}^{ij} that $\hat{l}^{ij} = l^{ij}$. The remaining identities can be deduced by similar reasoning. Henceforth, the matrices $[\hat{l}^{ij}]$, $[l^{ij}]$, etc, will be used interchangeably.

Remark 31 Raising and Lowering Indices.

The metric tensor $F = [f_{ij}] = [\langle \mathbf{f}_i, \mathbf{f}_j \rangle]$ and its dual counterpart $[f^{ij}] = [\langle \mathbf{f}^i, \mathbf{f}^j \rangle]$ can be used to switch between covariant and contravariant component forms of vectors and linear transformations (and indeed, tensors of all orders). This process is called “raising and lowering indices” in classical tensor algebra. Thus, for example, if $\mathbf{v} = v^i \mathbf{f}_i = v_i \mathbf{f}^i$, then

$$v^i = \langle \mathbf{v}, \mathbf{f}^i \rangle = \langle v_k \mathbf{f}^k, \mathbf{f}^i \rangle = f^{ik} v_k.$$

Thus, $[f^{ij}]$ is used to transform the covariant components $[v_i]$ of a vector \mathbf{v} into the contravariant components $[v^i]$, a process called “raising the index”. By similar reasoning one shows that

$$v_i = f_{ik} v^k,$$

that is, the metric tensor $F = [f_{ij}]$ is used to transform the contravariant components of a vector into the covariant components, a process called “lowering the index”.

Analogous raising/lowering index formulas hold for linear transformations. In particular, suppose $\mathbf{w} = \mathbf{L}\mathbf{v}$ and consider the string of identities

$$\begin{aligned}\mathbf{L}\mathbf{v} &= \sum_{ij} l^{ij} \mathbf{f}_i \otimes \mathbf{f}_j (v^k \mathbf{f}_k) \\ &= \sum_{ij} l^{ij} \mathbf{f}_i v^k \langle \mathbf{f}_j, \mathbf{f}_k \rangle \\ &= \sum_i \mathbf{f}_i \left(\sum_k l^{ik} f_{kj} \right) v^j.\end{aligned}$$

From the definition of l_j^i , it follows that

$$l_j^i = l^{ik} f_{kj}.$$

Thus, the tensor $[f_{ij}]$ can be used to “lower” the column index of $[l^{ij}]$ by right multiplication of matrices. Similarly, $[f^{ij}]$ can be used to raise the column index by right multiplication. Moreover, they can be used to lower and raise the row index of the matrices $[l^{ij}]$ as illustrated by the following string of identities.

$$\begin{aligned}\mathbf{L} &= \sum_{ij} l^{ij} \mathbf{f}_i \otimes \mathbf{f}_j \\ &= \sum_{ij} l^{ij} (f_{iq} \mathbf{f}^q) \otimes \mathbf{f}_j \\ &= \sum_{qj} \left(\sum_i f_{iq} l^{ij} \right) \mathbf{f}^q \otimes \mathbf{f}_j\end{aligned}$$

which from the definition of the matrix $[l_q^j]$ implies that

$$l_i^j = f_{ik} l^{kj}. \quad (51)$$

The following identities are readily obtained by similar arguments.

$$l^{ij} = f^{ik} l_k^j = l_k^i f_{ik}, \quad l_{ij} = f_{ik} l_j^i = l_i^k f_{ik}, \quad \text{etc.} \dots$$

Applying (51) to the identity transformation for which $l_j^i = \delta_j^i$, $l_i^j = \delta_i^j$, $l^{ij} = f^{ij}$ and $l_{ij} = f_{ij}$, one sees that

$$\delta_i^j = f_{ik} f^{kj}.$$

It follows that the metric tensor $[f_{ij}]$ and the matrix $[f^{ij}]$ are inverses of one another, i.e.

$$[f^{ij}] = [f_{ij}]^{-1}.$$

5.1.6 Eigenvalues, Eigenvectors and Eigenspaces

Invariance principles are central tools in most areas of mathematics and they come in many forms. In linear algebra, the most important invariance principle resides in the notion of an *Eigenvector*.

Definition 27 Let \mathcal{V} be a vector space and $\mathbf{L} \in \text{Lin}(\mathcal{V})$ a linear transformation on \mathcal{V} . Then a non-zero vector $\mathbf{v} \in \mathcal{V}$ is called an *Eigenvector* and $\lambda \in \mathcal{F}$ its associated *Eigenvalue* provided

$$\mathbf{L}\mathbf{v} = \lambda\mathbf{v}. \quad (52)$$

From the definition (52) it follows that $\mathbf{L}(\text{span}(\mathbf{v})) \subset \text{span}(\mathbf{v})$, that is the line spanned by the eigenvector \mathbf{v} is invariant under \mathbf{L} in that \mathbf{L} maps the line back onto itself. The set of all eigenvectors corresponding to a given eigenvalue λ (with the addition of the zero vector) forms a subspace of \mathcal{V} (Why?), denoted $E(\lambda)$, called the *Eigenspace* corresponding to the eigenvalue λ . One readily shows (How?) from the definition (52) that eigenvectors corresponding to distinct eigenvalues of a given linear transformation $\mathbf{L} \in \text{Lin}(\mathcal{V})$ are independent, from which one concludes

$$E(\lambda_1) \cap E(\lambda_2) = \{\mathbf{0}\} \quad \text{if} \quad \lambda_1 \neq \lambda_2. \quad (53)$$

Re-writing the definition (52) as

$$(\mathbf{L} - \lambda I)\mathbf{v} = \mathbf{0}$$

one sees that \mathbf{v} is an eigenvector with eigenvalue λ if and only if

$$\mathbf{v} \in \text{Nul}(\mathbf{L} - \lambda I).$$

It follows that λ is an eigenvalue for \mathbf{L} if and only if

$$\det(\mathbf{L} - \lambda I) = 0.$$

Recalling that $\det(\mathbf{L} - \lambda I)$ is a polynomial of degree N , one defines the *Characteristic Polynomial* $p_{\mathbf{L}}(\lambda)$ of the transformation \mathbf{L} to be

$$p_{\mathbf{L}}(\lambda) := \det(\mathbf{L} - \lambda I) = (-\lambda)^N + i_1(\mathbf{L})(-\lambda)^{N-1} + \dots + i_{N-1}(\mathbf{L})(-\lambda) + i_N(\mathbf{L}). \quad (54)$$

The coefficients $\mathcal{I}_{\mathbf{L}} := \{i_1(\mathbf{L}), \dots, i_N(\mathbf{L})\}$ of the characteristic polynomial are called the *Principal Invariants* of the transformation \mathbf{L} because they are invariant under similarity transformation. More specifically, let $\mathbf{T} \in \text{Lin}(\mathcal{V})$ be non-singular. Then the characteristic polynomial for $\mathbf{T}^{-1}\mathbf{L}\mathbf{T}$ is identical to that for \mathbf{L} as can be seen from the calculation

$$\begin{aligned} p_{\mathbf{T}^{-1}\mathbf{L}\mathbf{T}}(\lambda) &= \det(\mathbf{T}^{-1}\mathbf{L}\mathbf{T} - \lambda I) \\ &= \det(\mathbf{T}^{-1}\mathbf{L}\mathbf{T} - \lambda\mathbf{T}^{-1}\mathbf{T}) \\ &= \det(\mathbf{T}^{-1}(\mathbf{L} - \lambda I)\mathbf{T}) \\ &= \det(\mathbf{L} - \lambda I) \\ &= p_{\mathbf{L}}(\lambda). \end{aligned}$$

An explanation of why the elements of $\mathcal{I}_{\mathbf{L}}$ are called the *Principal Invariants* of \mathbf{L} must wait until the upcoming section on spectral theory.

The principle invariants $i_1(\mathbf{L})$ and $i_N(\mathbf{L})$ have simple forms. Indeed one can readily show (How?) that

$$i_1(\mathbf{L}) = \text{tr}(\mathbf{L}) \quad \text{and} \quad i_N(\mathbf{L}) = \det(\mathbf{L}).$$

Moreover, if the characteristic polynomial has the factorization

$$p_{\mathbf{L}}(\lambda) = \prod_{j=1}^N (\lambda_j - \lambda),$$

then one also has that

$$i_1(\lambda) = \text{tr}(\mathbf{L}) = \lambda_1 + \dots + \lambda_N \quad \text{and} \quad i_N(\mathbf{L}) = \det(\mathbf{L}) = \lambda_1 \dots \lambda_N.$$

5.2 Classes of Linear Transformations

An understanding of general linear transformations in $\text{Lin}(V)$ can be gleaned from a study of special classes of transformations and representation theorems expressing general linear transformations as sums or products of transformations from the special classes. The discussion is framed within the setting of a finite dimensional, inner product vector space. The first class of transformations introduced are the projections.

5.2.1 Projections

Definition 28 A linear transformation P on $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ is called a *Projection* provided

$$P^2 = P. \tag{55}$$

Remark 32 It follows readily from the definition (55) that P is a projection if and only if $I - P$ is a projection. Indeed, assume P is a projection, then

$$(I - P)^2 = I^2 - 2P + P^2 = I - 2P + P = I - P$$

which shows that $I - P$ is a projection. The reverse implication follows by a similar argument since if $I - P$ is a projection, then one need merely write $P = I - (I - P)$ and apply the above calculation.

Remark 33 The geometry of projections is illuminated by the observation that there is a simple relationship between their range and nullspace. In particular, one can show easily from the definition (55) that

$$\text{Ran}(P) = \text{Nul}(I - P) \quad \text{and} \quad \text{Nul}(P) = \text{Ran}(I - P). \tag{56}$$

For example, to prove that $\text{Ran}(P) = \text{Nul}(I - P)$, let $\mathbf{v} \in \text{Ran}(P)$. Then $\mathbf{v} = P\mathbf{u}$ for some $\mathbf{u} \in \mathcal{V}$ and $(I - P)\mathbf{v} = (I - P)P\mathbf{u} = P\mathbf{u} - P^2\mathbf{u} = P\mathbf{u} - P\mathbf{u} = \mathbf{0}$. It follows that $\mathbf{v} \in \text{Nul}(P)$ and hence that

$$\text{Ran}(P) \subset \text{Nul}(I - P). \quad (57)$$

Conversely, if $\mathbf{v} \in \text{Nul}(I - P)$, then $(I - P)\mathbf{v} = \mathbf{0}$, or equivalently $\mathbf{v} = P\mathbf{v}$. Thus, $\mathbf{v} \in \text{Ran}(P)$ from which it follows that

$$\text{Nul}(I - P) \subset \text{Ran}(P). \quad (58)$$

From (57,58) one concludes that $\text{Ran}(P) = \text{Nul}(I - P)$ as required. The second equality in (56) follows similarly.

Remark 34 Another important observation concerning projections is that \mathcal{V} is a direct sum of the subspaces $\text{Ran}(P)$ and $\text{Nul}(P)$

$$\mathcal{V} = \text{Ran}(P) \oplus \text{Nul}(P). \quad (59)$$

To see (59), one merely notes that for any $\mathbf{v} \in \mathcal{V}$

$$\mathbf{v} = P\mathbf{v} + (I - P)\mathbf{v}, \quad \text{with } P\mathbf{v} \in \text{Ran}(P) \quad \text{and} \quad (I - P)\mathbf{v} \in \text{Nul}(P).$$

This fact and (9) show (59). Because of (59) one often refers to a projection P as “projecting \mathcal{V} onto $\text{Ran}(P)$ along $\text{Nul}(P)$.”

An important subclass of projections are the *Perpendicular Projections*.

Definition 29 A projection P is called a *Perpendicular Projection* provided $\text{Ran}(P) \perp \text{Nul}(P)$.

A non-obvious fact about perpendicular projections is the

Theorem 8 A projection P is a perpendicular projection if and only if P is self-adjoint, i.e. $P = P^*$.

Proof: First assume P is self-adjoint and let $\mathbf{v} \in \text{Ran}(P)$ and $\mathbf{w} \in \text{Nul}(P)$. Then, $\mathbf{v} = P\mathbf{u}$ for some $\mathbf{u} \in \mathcal{V}$ and

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle P\mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{u}, P^*\mathbf{w} \rangle = \langle \mathbf{u}, P\mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{0} \rangle = 0$$

from which it follows that $\mathbf{v} \perp \mathbf{w}$ and hence that $\text{Ran}(P) \perp \text{Nul}(P)$.

Conversely, suppose P is a perpendicular projection. One must show that P is self-adjoint, i.e. that

$$\langle P\mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, P\mathbf{u} \rangle \quad \text{for all } \mathbf{v}, \mathbf{u} \in \mathcal{V}. \quad (60)$$

To that end, notice that

$$\langle P\mathbf{v}, \mathbf{u} \rangle = \langle P\mathbf{v}, P\mathbf{u} + (I - P)\mathbf{u} \rangle = \langle P\mathbf{v}, P\mathbf{u} \rangle + \langle P\mathbf{v}, (I - P)\mathbf{u} \rangle = \langle P\mathbf{v}, P\mathbf{u} \rangle. \quad (\text{Why?})$$

Since $\langle P\mathbf{v}, P\mathbf{u} \rangle$ is symmetric in \mathbf{v} and \mathbf{u} , the same must be true for $\langle P\mathbf{v}, \mathbf{u} \rangle$, from which one concludes (60). \square

Question: Among the class of elementary tensor products, which (if any) are projections? Which are perpendicular projections?

Answer: An elementary tensor product $\mathbf{a} \otimes \mathbf{b}$ is a projection if and only if $\langle \mathbf{a}, \mathbf{b} \rangle = 1$. It is a perpendicular projection if and only if $\mathbf{b} = \mathbf{a}$ and $|\mathbf{a}| = 1$. (Why?)

Thus, if \mathbf{a} is a unit vector, then $\mathbf{a} \otimes \mathbf{a}$ is perpendicular projection onto the line spanned by \mathbf{a} . More generally, if \mathbf{a} and \mathbf{b} are perpendicular, unit vectors, then $\mathbf{a} \otimes \mathbf{a} + \mathbf{b} \otimes \mathbf{b}$ is a perpendicular projection onto the plane (2-dimensional subspace) spanned by \mathbf{a} and \mathbf{b} . (Exercise.) Generalizing further, if $\mathbf{a}_1, \dots, \mathbf{a}_K$ are pairwise orthogonal unit vectors, then $\mathbf{a}_1 \otimes \mathbf{a}_1 + \dots + \mathbf{a}_K \otimes \mathbf{a}_K$ is a perpendicular projection onto $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_K)$.

5.2.2 Diagonalizable Transformations

A linear transformation \mathbf{L} on a finite dimensional vector space \mathcal{V} is called *Diagonalizable* provided it has a matrix representation that is diagonal, that is, there is a basis B for \mathcal{V} such that

$$[\mathbf{L}]_B = \text{diag}(l_1, \dots, l_N). \quad (61)$$

Notation: An $N \times N$ diagonal matrix will be denoted $A = \text{diag}(a_1, \dots, a_N)$ where $A = [a_{ij}]$ with $a_{11} = a_1, \dots, a_{NN} = a_N$ and $a_{ij} = 0$ for $i \neq j$.

The fundamental result concerning diagonalizable transformations is

Theorem 9 *A linear transformation \mathbf{L} on \mathcal{V} is diagonalizable if and only if \mathcal{V} has a basis consisting entirely of eigenvectors of \mathbf{L} .*

Proof: Assume first that \mathcal{V} has a basis $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ consisting of eigenvectors of \mathbf{L} , that is $\mathbf{L}\mathbf{f}_j = \lambda_j \mathbf{f}_j$ for $j = 1, \dots, N$. Recall that the column N -tuples comprising the matrix $[\mathbf{L}]_B$ are given by $[\mathbf{L}\mathbf{f}_j]_B$, the coordinates with respect to B of the vectors $\mathbf{L}\mathbf{f}_j$ for $j = 1, \dots, N$. But

$$[\mathbf{L}\mathbf{f}_1]_B = [\lambda_1 \mathbf{f}_1]_B = \begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad [\mathbf{L}\mathbf{f}_2]_B = [\lambda_2 \mathbf{f}_2]_B = \begin{pmatrix} 0 \\ \lambda_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{etc.}$$

It now follows immediately that

$$[\mathbf{L}]_B = \text{diag}(\lambda_1, \dots, \lambda_N). \quad (\text{Why?})$$

Conversely, suppose \mathbf{L} is diagonalizable, that is, there exists a basis $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ for \mathcal{V} such that $[\mathbf{L}]_B = \text{diag}(\lambda_1, \dots, \lambda_N)$. But then for each $j = 1, \dots, N$

$$[\mathbf{L}\mathbf{f}_j]_B = [\mathbf{L}]_B [\mathbf{f}_j]_B = \text{diag}(\lambda_1, \dots, \lambda_N) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (j^{\text{th}}\text{-component}=1) = [\lambda_j \mathbf{f}_j]_B.$$

It follows that for each $j = 1, \dots, N$, $\mathbf{L}\mathbf{f}_j = \lambda_j \mathbf{f}_j$ (Why?), and hence that $B = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ is a basis of eigenvectors of \mathbf{L} . \square

5.2.3 Orthogonal and Unitary Transformations

Let $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ be a finite dimensional, inner product vector space. The the class of *Orthogonal Transformations*, in the case of a vector space over \mathbb{R} , and the class of *Unitary Transformation*, in the case of a vector space over \mathbb{C} , is defined by

Definition 30 A transformation $\mathbf{Q} \in \text{Lin}(\mathcal{V})$ is called *Orthogonal (real case) or Unitary (complex case)* provided

$$\mathbf{Q}^{-1} = \mathbf{Q}^* \quad \text{or equivalently} \quad I = \mathbf{Q}^* \mathbf{Q}. \quad (62)$$

That is, the inverse of \mathbf{Q} equals the adjoint of \mathbf{Q} .

Remark 35 It follows immediately from (62) that orthogonal transformations preserve the inner product on \mathcal{V} and hence leave lengths and angles invariant. Specifically, one has

$$\langle \mathbf{Q}\mathbf{a}, \mathbf{Q}\mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{Q}^* \mathbf{Q}\mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{Q}^{-1} \mathbf{Q}\mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{b} \rangle. \quad (63)$$

In particular, from (63) it follows that

$$|\mathbf{Q}\mathbf{a}|^2 = \langle \mathbf{Q}\mathbf{a}, \mathbf{Q}\mathbf{a} \rangle = \langle \mathbf{a}, \mathbf{a} \rangle = |\mathbf{a}|^2$$

showing that \mathbf{Q} preserves lengths.

Remark 36 The class of orthogonal transformations will be denoted Orth . One easily shows that

$$\det(\mathbf{Q}) = \pm 1. \quad (\text{Why?})$$

It is useful to define the subclasses

$$\text{Orth}^+ := \{\mathbf{Q} \in \text{Orth} \mid \det(\mathbf{Q}) = 1\} \quad \text{and} \quad \text{Orth}^- := \{\mathbf{Q} \in \text{Orth} \mid \det(\mathbf{Q}) = -1\}.$$

Example 14 If \mathcal{V} is the Euclidean plane \mathbb{R}^2 , then $\text{Orth}^+(\mathbb{R}^2)$ is the set of (rigid) *Rotations*, i.e. given $\mathbf{Q} \in \text{Orth}^+(\mathbb{R}^2)$, there exists a $0 < \theta < 2\pi$ such that

$$\mathbf{Q} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \quad (64)$$

Also, the transformation

$$\mathbf{R}_M := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is in $\text{Orth}^-(\mathbb{R}^2)$ (Why?) (How does \mathbf{R}_M act geometrically?). More generally,

$$\text{Orth}^-(\mathbb{R}^2) = \{\mathbf{R} \in \text{Orth}(\mathbb{R}^2) \mid \mathbf{R} = \mathbf{R}_M \mathbf{Q} \text{ for some } \mathbf{Q} \in \text{Orth}^+(\mathbb{R}^2)\}. \quad (\text{Why?})$$

Example 15 If \mathcal{V} is the Euclidean space \mathbb{R}^3 , then Orth^+ is called the set of *Proper Orthogonal Transformations* or the set of *Rotations*. To motivate calling $\mathbf{Q} \in \text{Orth}^+(\mathbb{R}^3)$, one can prove that

Theorem 10 If $Q \in \text{Orth}^+(\mathbb{R}^3)$, then there exists a non-zero vector \mathbf{a} such that $Q\mathbf{a} = \mathbf{a}$ (i.e. \mathbf{a} is an eigenvector corresponding to the eigenvalue $\lambda = 1$) and Q restricted to the plane orthogonal to \mathbf{a} is a two dimensional rotation of the form (64). The line spanned by \mathbf{a} is called the *Axis of Rotation*.

Proof: Exercise. (Hint: Use the following facts to show that $1 \in \sigma(Q)$).

1. $1 = \det(\mathbf{Q}) = \lambda_1 \lambda_2 \lambda_3$ where the (complex) spectrum of $\mathbf{Q} = \{\lambda_1, \lambda_2, \lambda_3\}$.
2. At least one of the eigenvalues must be real (Why?), say λ_1 , and the other two must be complex conjugates (Why?), i.e. $\lambda_2 \lambda_3 = |\lambda_2|^2$.
3. $|\lambda_j| = 1$ for $j = 1, 2, 3$ (Why?) and hence $\lambda_2 \lambda_3 = 1$.
4. Since $\lambda_1 = 1$ is an eigenvalue for \mathbf{Q} , it has an eigenvector \mathbf{a} , i.e. $\mathbf{Q}\mathbf{a} = \mathbf{a}$.
5. \mathbf{Q} is in Orth^+ when restricted to the plane orthogonal to \mathbf{a} (Why?).

The conclusion of the theorem now follows readily.) \square

It then follows that

$$\text{Orth}^-(\mathbb{R}^3) = \{R \in \text{Orth}(\mathbb{R}^3) \mid R = -Q \text{ for some } Q \in \text{Orth}^+(\mathbb{R}^3)\}. \quad (\text{Why?})$$

Remark 37 Let $\mathcal{V} = \mathbb{R}^N$ equipped with the Euclidean inner product and natural basis. One can now make the identification of $\text{Lin}(\mathbb{R}^N)$ with the set of $N \times N$ -matrices, \mathcal{M}^N . Then $\text{Orth}(\mathbb{R}^N)$, the set of orthogonal transformations on \mathbb{R}^N , corresponds to the set of orthogonal matrices in \mathcal{M}^N . It follows from the definition (62) that $Q \in \text{Orth}(\mathbb{R}^N) \subset \mathcal{M}^N$ if and only if the columns (and hence rows) of Q form an orthonormal basis for \mathbb{R}^N . (Why?)

5.2.4 Self-Adjoint Transformations

Let $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space. An important subspace of $\text{Lin}(\mathcal{V})$ is the collection of *Hermitian* or *Self-Adjoint* transformations denoted by $\text{Herm}(\mathcal{V})$. The self-adjoint transformations are just those transformations equal to their adjoints, i.e. $S \in \text{Herm}(\mathcal{V})$ means $S^* = S$. The goal of this section is to present the celebrated *Spectral Theorem* for self-adjoint transformations. The presentation proceeds in steps through a sequence of lemmas. For much of the discussion, the scalar field can be either \mathbb{R} or \mathbb{C} .

Definition 31 The *Spectrum* of a transformation in $\text{Lin}(\mathcal{V})$, denoted $\sigma(\mathbf{L})$, is the set of all eigenvalues of \mathbf{L} .

Lemma 1 Let \mathbf{L} be a self-adjoint transformation on the inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$. Then $\sigma(\mathbf{L}) \subset \mathbb{R}$, i.e. the spectrum of \mathbf{L} contains only real numbers.

Proof: Let $\lambda \in \sigma(\mathbf{L})$ with associated unit eigenvector \mathbf{e} , i.e. $|\mathbf{e}| = 1$ and $\mathbf{L}\mathbf{e} = \lambda\mathbf{e}$. Then

$$\lambda = \lambda \langle \mathbf{e}, \mathbf{e} \rangle = \langle \lambda \mathbf{e}, \mathbf{e} \rangle = \langle \mathbf{L}\mathbf{e}, \mathbf{e} \rangle = \langle \mathbf{e}, \mathbf{L}\mathbf{e} \rangle = \langle \mathbf{e}, \lambda \mathbf{e} \rangle = \bar{\lambda} \langle \mathbf{e}, \mathbf{e} \rangle = \bar{\lambda}.$$

Thus, $\lambda = \bar{\lambda}$ from which it follows that $\lambda \in \mathbb{R}$. \square

Lemma 2 Let \mathbf{L} be a self-adjoint transformation on the inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$. Then distinct eigenvalues of \mathbf{L} have orthogonal eigenspaces.

Proof: Let $\lambda_1, \lambda_2 \in \sigma(\mathbf{L})$ be distinct (real) eigenvalues of \mathbf{L} with associated eigenvectors $\mathbf{e}_1, \mathbf{e}_2$, respectively. What must be shown is that $\mathbf{e}_1 \perp \mathbf{e}_2$, i.e. that $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0$. To that end,

$$\lambda_1 \langle \mathbf{e}_1, \mathbf{e}_2 \rangle = \langle \lambda_1 \mathbf{e}_1, \mathbf{e}_2 \rangle = \langle \mathbf{L}\mathbf{e}_1, \mathbf{e}_2 \rangle = \langle \mathbf{e}_1, \mathbf{L}\mathbf{e}_2 \rangle = \langle \mathbf{e}_1, \lambda_2 \mathbf{e}_2 \rangle = \lambda_2 \langle \mathbf{e}_1, \mathbf{e}_2 \rangle.$$

It follows that

$$(\lambda_1 - \lambda_2) \langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0.$$

But since $\lambda_1 \neq \lambda_2$, it must be that $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0$, i.e. that $\mathbf{e}_1 \perp \mathbf{e}_2$, as required. \square

Lemma 3 Let \mathbf{L} be a self-adjoint transformation on the inner product space $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ and let $\lambda \in \sigma(\mathbf{L})$ with associated eigenspace $E(\lambda)$. Then $E(\lambda)$ and its orthogonal complement $E(\lambda)^\perp$ are invariant subspaces of \mathbf{L} , i.e.

$$\mathbf{L}(E(\lambda)) \subset E(\lambda) \tag{65}$$

$$\mathbf{L}(E(\lambda)^\perp) \subset E(\lambda)^\perp. \tag{66}$$

Proof: Assertion (65) follows immediately from the definitions of eigenvalue and eigenspace. For the second assertion (66), one must show that if $\mathbf{v} \in E(\lambda)^\perp$, the $\mathbf{L}\mathbf{v}$ is also in $E(\lambda)^\perp$. To that end, let $\mathbf{e} \in E(\lambda)$ and $\mathbf{v} \in E(\lambda)^\perp$. Then $\mathbf{v} \perp \mathbf{e}$ and

$$\langle \mathbf{L}\mathbf{v}, \mathbf{e} \rangle = \langle \mathbf{v}, \mathbf{L}^*\mathbf{e} \rangle = \langle \mathbf{v}, \mathbf{L}\mathbf{e} \rangle = \langle \mathbf{v}, \lambda \mathbf{e} \rangle = \lambda \langle \mathbf{v}, \mathbf{e} \rangle = 0.$$

It follows that $\mathbf{L}\mathbf{v}$ is perpendicular to every eigenvector in $E(\lambda)$ from which one concludes that $\mathbf{L}\mathbf{v} \in E(\lambda)^\perp$ as required. \square

With the aid of these lemmas, one can now prove the *Spectral Theorem* for self-adjoint transformations.

Theorem 11 Let $\mathbf{L} \in \text{Lin}(\mathcal{V})$ be self-adjoint. Then there is an orthonormal basis for \mathcal{V} consisting entirely of eigenvectors of \mathbf{L} . In particular, the eigenspaces of \mathbf{L} form a direct sum decomposition of \mathcal{V}

$$\mathcal{V} = E(\lambda_1) \oplus \dots \oplus E(\lambda_K) \tag{67}$$

where $\sigma(\mathbf{L}) = \{\lambda_1, \dots, \lambda_K\}$ is the spectrum of \mathbf{L} .

Proof: Since all of the eigenvalues of \mathbf{L} are real numbers, the characteristic polynomial factors completely over the real numbers, i.e.

$$\det(\mathbf{L} - \lambda I) = p_{\mathbf{L}}(\lambda) = \prod_{j=1}^K (\lambda_j - \lambda)^{n_j}$$

with

$$n_1 + \dots + n_K = N$$

where $\dim(\mathcal{V}) = N$. The exponent n_j is called the *algebraic multiplicity* of the eigenvalue λ_j . The *geometric multiplicity*, m_j , of each eigenvalue λ_j is defined to be $\dim(E(\lambda_j))$. To prove the theorem, it suffices to show that

$$m_1 + \dots + m_K = N. \quad (68)$$

Indeed, (68) shows that there is a basis for \mathcal{V} consisting of eigenvectors of \mathbf{L} , called an *eigenbasis*, and Lemma 2 shows one can find an orthonormal eigenbasis for \mathcal{V} . Since the eigenspaces $E(\lambda_j)$, $j = 1, \dots, K$ are pairwise orthogonal, one can form the direct sum subspace

$$\mathcal{V}_1 := E(\lambda_1) \oplus \dots \oplus E(\lambda_K).$$

The problem is to show that $\mathcal{V}_1 = \mathcal{V}$. Suppose not. Then one can write

$$\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_1^\perp.$$

First note that

$$\mathcal{V}_1^\perp = \cap_{j=1}^K E(\lambda_j)^\perp. \quad (\text{Why?})$$

It follows from Lemma 3, that $\mathbf{L}(\mathcal{V}_1^\perp) \subset \mathcal{V}_1^\perp$. (Why?) Let \mathbf{L}_1 denote the restriction of \mathbf{L} to the subspace \mathcal{V}_1^\perp

$$\mathbf{L}_1 : \mathcal{V}_1^\perp \longrightarrow \mathcal{V}_1^\perp,$$

that is, if $\mathbf{v} \in \mathcal{V}_1^\perp$, then $\mathbf{L}_1 \mathbf{v} := \mathbf{L} \mathbf{v} \in \mathcal{V}_1^\perp$. Notice that \mathbf{L}_1 is self-adjoint. Hence, the characteristic polynomial of \mathbf{L}_1 factors completely over the real numbers. Let μ be an eigenvalue of \mathbf{L}_1 with associated eigenvector \mathbf{e} , i.e. $\mathbf{L}_1 \mathbf{e} = \mu \mathbf{e}$. Thus, \mathbf{e} is also an eigenvector of \mathbf{L} with eigenvalue μ . It follows that μ must be one of the λ_j , $j = 1, \dots, K$ and $\mathbf{e} \in E(\lambda_j)$ for some $j = 1, \dots, K$. But \mathbf{e} must be orthogonal to all of the eigenspaces $E(\lambda_j)$, $j = 1, \dots, K$. This contradiction implies that $\mathcal{V}_1^\perp = \{\mathbf{0}\}$ and that $\mathcal{V}_1 = \mathcal{V}$ as required. \square

Let \mathbf{P}_j denote the perpendicular projection operator onto the eigenspace $E(\lambda_j)$. Then since the eigenspaces are pairwise orthogonal, it follows that (Why?)

$$\mathbf{P}_i \mathbf{P}_j = \mathbf{0} \quad \text{if } i \neq j, \quad \text{and} \quad \mathbf{P}_j^2 = \mathbf{P}_j. \quad (69)$$

The spectral theorem implies that a self-adjoint transformation \mathbf{L} has a *Spectral Decomposition* defined by

$$\mathbf{L} = \lambda_1 \mathbf{P}_1 + \dots + \lambda_K \mathbf{P}_K \quad (70)$$

where the spectrum of \mathbf{L} is $\sigma(\mathbf{L}) = \{\lambda_1, \dots, \lambda_K\}$. Moreover, for each positive integer n , \mathbf{L}^n is also self-adjoint (Why?), and appealing to (70) one easily shows (How?) that \mathbf{L}^n has the spectral decomposition

$$\mathbf{L}^n = \lambda_1^n \mathbf{P}_1 + \dots + \lambda_K^n \mathbf{P}_K.$$

Thus, \mathbf{L}^n has the same eigenspaces as \mathbf{L} with spectrum $\sigma(\mathbf{L}^n) = \{\lambda_1^n, \dots, \lambda_K^n\}$. More generally, let $f(x) = \sum_{k=0}^n a_k x^k$ be a polynomial. Then $f(\mathbf{L})$ is self-adjoint (Why?) and has the spectral decomposition

$$f(\mathbf{L}) = \sum_{j=1}^K f(\lambda_j) \mathbf{P}_j. \quad (71)$$

It follows that $\sigma(f(\mathbf{L})) = \{f(\lambda_1), \dots, f(\lambda_K)\}$ and $f(\mathbf{L})$ has the same eigenspaces as \mathbf{L} .

Generalizing further, let $f(x) = \sum_{k=0}^{\infty} a_k x^k$ be an entire function, i.e. $f(x)$ has a power series representation convergent for all x . Then, again one concludes that $f(\mathbf{L})$ (Why is the power series with \mathbf{L} convergent?) is self-adjoint and has the spectral decomposition (71).

Example 16 Consider the entire function $\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Then, $\exp(\mathbf{L})$ is self-adjoint (if \mathbf{L} is) and has the spectral representation

$$\exp(\mathbf{L}) = \sum_{j=1}^K \exp(\lambda_j) \mathbf{P}_j.$$

Example 17 Let $\mathcal{V} = \mathbb{R}^N$ with the usual inner product and let $\mathbf{A} \in \mathcal{M}^N(\mathbb{R})$ be a self-adjoint $N \times N$ real matrix. Consider the system of ordinary differential equations

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) \tag{72}$$

with initial condition $\mathbf{x}(0) = \mathbf{x}_0$. Then the solution to the initial value problem for the system (72) is

$$\mathbf{x}(t) = \exp(\mathbf{A}t) \mathbf{x}_0$$

which by the spectral theorem can be written

$$\mathbf{x}(t) = \sum_{j=1}^K \exp(\lambda_j t) \mathbf{P}_j$$

where \mathbf{A} has the spectrum $\sigma(\mathbf{A}) = \{\lambda_1, \dots, \lambda_K\}$ and spectral projections \mathbf{P}_j , $j = 1, \dots, K$.

5.2.5 Square-root of Positive Definite Hermitian Transformations

Suppose $\mathbf{L} \in \text{Herm}$ has only positive eigenvalues. Then \mathbf{L} has a unique positive definite Hermitian square-root. Indeed, if \mathbf{L} is positive definite Hermitian, then it has the spectral decomposition

$$\mathbf{L} = \mu_1 \mathbf{P}_1 + \dots + \mu_K \mathbf{P}_K$$

with $\mu_j \in \mathbb{R}^+$ and \mathbf{P}_j the perpendicular projection onto the eigenspace $E(\mu_j)$. Define

$$\sqrt{\mathbf{L}} := \sqrt{\mu_1} \mathbf{P}_1 + \dots + \sqrt{\mu_K} \mathbf{P}_K.$$

Then clearly $\sqrt{\mathbf{L}}$ is positive definite Hermitian and $(\sqrt{\mathbf{L}})^2 = \mathbf{L}$.

Remark 38 The spectral theorem shows that a necessary and sufficient condition for a transformation $\mathbf{L} \in \text{Lin}(\mathcal{V})$ to have a real spectrum $\sigma(\mathbf{L}) = \{\lambda_1, \dots, \lambda_N\} \subset \mathbb{R}$ and a set of eigenvectors forming an orthonormal basis for \mathcal{V} , is for \mathbf{L} to be Hermitian or self-adjoint. A natural question to ask is under what conditions on \mathbf{L} does this result hold when the spectrum is permitted to be complex. The desired condition on \mathbf{L} is called *Normality*.

5.2.6 Normal Transformations

Definition 32 A transformation $\mathbf{L} \in \text{Lin}(\mathcal{V})$ is called *Normal* provided it commutes with its adjoint, i.e. $\mathbf{L}^* \mathbf{L} = \mathbf{L} \mathbf{L}^*$.

Remark 39 It follows immediately from the definitions that unitary and Hermitian transformations are also normal. Moreover, if \mathbf{L} has a spectral decomposition

$$\mathbf{L} = \lambda_1 \mathbf{e}_1 \otimes \mathbf{e}_1 + \dots + \lambda_N \mathbf{e}_N \otimes \mathbf{e}_N \quad (73)$$

with $\lambda_j \in \mathbb{C}$, $j = 1, \dots, N$ and $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ an orthonormal basis for \mathcal{V} , then \mathbf{L} is normal. The following generalization of the spectral theorem says that this condition is both necessary and sufficient for \mathbf{L} to be normal.

Theorem 12 A necessary and sufficient condition for $\mathbf{L} \in \text{Lin}(\mathcal{V})$ to be normal is that it have a Spectral Decomposition of the form (73).

Proof: The proof follows exactly as in the case for self-adjoint transformations with the three key lemmas being modified as follows.

Lemma 4 If $\mathbf{L} \in \text{Lin}(\mathcal{V})$ is normal, then $\mathbf{L} \mathbf{e}_j = \lambda_j \mathbf{e}_j$ if and only if $\mathbf{L}^* \mathbf{e}_j = \overline{\lambda_j} \mathbf{e}_j$, and hence $\sigma(\mathbf{L}^*) = \overline{\sigma(\mathbf{L})}$ and $E_{\mathbf{L}^*}(\overline{\lambda_j}) = E_{\mathbf{L}}(\lambda_j)$ for each $\lambda_j \in \sigma(\mathbf{L})$.

Proof: Note first that if \mathbf{L} is normal then so is $\mathbf{L} - \lambda I$ since $(\mathbf{L} - \lambda I)^* = \mathbf{L}^* - \overline{\lambda} I$. Next observe that if \mathbf{L} is normal, then $|\mathbf{L} \mathbf{v}| = |\mathbf{L}^* \mathbf{v}|$ for all $\mathbf{v} \in \mathcal{V}$. This last fact follows from

$$|\mathbf{L} \mathbf{v}|^2 = \langle \mathbf{L} \mathbf{v}, \mathbf{L} \mathbf{v} \rangle = \langle \mathbf{L}^* \mathbf{L} \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{L} \mathbf{L}^* \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{L}^* \mathbf{v}, \mathbf{L}^* \mathbf{v} \rangle = |\mathbf{L}^* \mathbf{v}|^2.$$

One now deduces that

$$|(\mathbf{L} - \lambda I) \mathbf{v}| = |(\mathbf{L}^* - \overline{\lambda} I) \mathbf{v}|$$

from which the conclusion of the lemma readily follows. (Why?) \square

Lemma 5 If \mathbf{L} is normal and λ_i, λ_j are distinct eigenvalues, then their eigenspaces are orthogonal, i.e. $E_{\mathbf{L}}(\lambda_j) \perp E_{\mathbf{L}}(\lambda_i)$.

Proof: Suppose $\mathbf{e}_i \in E_{\mathbf{L}}(\lambda_i)$ and $\mathbf{e}_j \in E_{\mathbf{L}}(\lambda_j)$. Then

$$\lambda_i \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \langle \lambda_i \mathbf{e}_i, \mathbf{e}_j \rangle = \langle \mathbf{L} \mathbf{e}_i, \mathbf{e}_j \rangle = \langle \mathbf{e}_i, \mathbf{L}^* \mathbf{e}_j \rangle = \langle \mathbf{e}_i, \overline{\lambda_j} \mathbf{e}_j \rangle = \overline{\lambda_j} \langle \mathbf{e}_i, \mathbf{e}_j \rangle$$

from which one concludes that $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ if $\lambda_i \neq \lambda_j$ thereby proving the lemma. \square

Lemma 6 If \mathbf{L} is normal with eigenvalue λ and associated eigenspace $E_{\mathbf{L}}(\lambda)$, then $\mathbf{L}(E_{\mathbf{L}}(\lambda)^\perp) \subset E_{\mathbf{L}}(\lambda)^\perp$.

Proof: Let $\mathbf{v} \in E_{\mathbf{L}}(\lambda)^\perp$ and $\mathbf{e} \in E_{\mathbf{L}}(\lambda)$. Then $\langle \mathbf{v}, \mathbf{e} \rangle = 0$ and

$$\langle \mathbf{L} \mathbf{v}, \mathbf{e} \rangle = \langle \mathbf{v}, \mathbf{L}^* \mathbf{e} \rangle = \langle \mathbf{v}, \overline{\lambda} \mathbf{e} \rangle = \overline{\lambda} \langle \mathbf{v}, \mathbf{e} \rangle = 0$$

from which one concludes that $\mathbf{L} \mathbf{v} \in E_{\mathbf{L}}(\lambda)^\perp$ thus proving the lemma. \square

Using these lemmas, the proof of the spectral theorem for self-adjoint transformations now readily applies to the case of normal transformations. (How?) \square

5.2.7 Skew Transformation

The *Skew Transformations*, also called *Anti-Self-Adjoint*, are defined as follows.

Definition 33 A transformation $\mathbf{W} \in \text{Lin}(\mathcal{V})$ is called *Skew* provided it satisfies $\mathbf{W}^* = -\mathbf{W}$.

It follows immediately that skew transformations are normal, i.e. they commute with their adjoints, and therefore can be diagonalized (over the complex numbers) by a unitary transformation. For example, in \mathbb{R}^2

$$\mathbf{W} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

is skew. It's spectrum is $\sigma(\mathbf{w}) = \{i, -i\}$ where $i = \sqrt{-1}$ (Why?) and

$$\mathbf{Q}^* \mathbf{W} \mathbf{Q} = \text{diag}(i, -i)$$

where

$$\mathbf{Q} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$$

is unitary (Why?).

In \mathbb{R}^3 , every skew matrix has the form (Why?)

$$\mathbf{W} = \begin{pmatrix} 0 & -\gamma & \beta \\ \gamma & 0 & -\alpha \\ -\beta & \alpha & 0 \end{pmatrix}.$$

One then shows (How?) that \mathbf{W} acts on a general vector $\mathbf{v} \in \mathbb{R}^3$ through vector cross product with a vector \mathbf{w} , called the *axial vector* of \mathbf{W} ,

$$\mathbf{W}\mathbf{v} = \mathbf{w} \times \mathbf{v}$$

with

$$\mathbf{w} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}.$$

(What is the spectrum of \mathbf{W} ?)

5.3 Decomposition Theorems

Decomposition theorems give insight into the structure and properties of general linear transformations by expressing them as multiplicative or additive combinations of special classes of transformations whose structure and properties are easily understood. The spectral decomposition is one such result expressing self-adjoint or normal as linear combinations of perpendicular projections. Perhaps the most basic of decompositions expresses a general real transformation as the sum of a symmetric and a skew-symmetric transformation, and a general complex transformation as the linear combination of two Hermitian transformations.

5.3.1 Sym-Skew and Hermitian-Hermitian Decompositions

Theorem 13 *Let $\mathbf{L} \in \text{Lin}(\mathcal{V})$ be a transformation on the real vector space \mathcal{V} . Then \mathbf{L} can be written uniquely as the sum*

$$\mathbf{L} = \mathbf{S} + \mathbf{W} \quad (74)$$

with

$$\mathbf{S} \in \text{Sym}(\mathcal{V}) \quad \text{and} \quad \mathbf{W} \in \text{Skew}(\mathcal{V}) \quad (75)$$

where $\text{Sym}(\mathcal{V})$ denotes the set of symmetric transformations (i.e. $\mathbf{S} \in \text{Sym}(\mathcal{V}) \iff \mathbf{S}^T = \mathbf{S}$) and $\text{Skew}(\mathcal{V})$ denotes the set of skew-symmetric transformations (i.e. $\mathbf{W} \in \text{Skew}(\mathcal{V}) \iff \mathbf{W}^T = -\mathbf{W}$).

Proof: The proof is by construction. Define

$$\mathbf{S} := \frac{1}{2}(\mathbf{L} + \mathbf{L}^T) \quad \text{and} \quad \mathbf{W} := \frac{1}{2}(\mathbf{L} - \mathbf{L}^T).$$

Then clearly (74,75) hold. Conversely, if (74,75) hold, then

$$\mathbf{L} + \mathbf{L}^T = (\mathbf{S} + \mathbf{W}) + (\mathbf{S} + \mathbf{W})^T = 2\mathbf{S}$$

and

$$\mathbf{L} - \mathbf{L}^T = (\mathbf{L} + \mathbf{W}) - (\mathbf{S} + \mathbf{W})^T = 2\mathbf{W}$$

thereby proving uniqueness of the representation (74,75). \square

Theorem 14 *Let $\mathbf{L} \in \text{Lin}(\mathcal{V})$ be a transformation on the complex vector space \mathcal{V} . Then \mathbf{L} can be written uniquely as the sum*

$$\mathbf{L} = \mathbf{S} + i\mathbf{W} \quad (76)$$

with

$$\mathbf{S}, \mathbf{W} \in \text{Herm}(\mathcal{V}). \quad (77)$$

Proof: The proof is identical to that above for the real transformation case except that

$$\mathbf{S} := \frac{1}{2}(\mathbf{L} + \mathbf{L}^*) \quad \text{and} \quad \mathbf{W} := -i\frac{1}{2}(\mathbf{L} - \mathbf{L}^*).$$

Then

$$\mathbf{W}^* = i\frac{1}{2}(\mathbf{L}^* - \mathbf{L}) = -i\frac{1}{2}(\mathbf{L} - \mathbf{L}^*) = \mathbf{W}$$

and

$$\mathbf{S} + i\mathbf{W} = \frac{1}{2}(\mathbf{L} + \mathbf{L}^*) + \frac{1}{2}(\mathbf{L} - \mathbf{L}^*) = \mathbf{L}. \square$$

Remark 40 Notice that (76,77) is the analog of the Cartesian representation of a complex number $z = x + iy$ with $x, y \in \mathbb{R}$.

5.3.2 Polar Decomposition

The *Polar Decomposition* gives a multiplicative decomposition of a nonsingular linear transformation into the product of a positive definite Hermitian transformation and a unitary transformation. It is useful to define the classes of transformations $\text{Orth}^+ := \{\mathbf{Q} \in \text{Orth} \mid \det(\mathbf{Q}) = 1\}$, $\text{PHerm} := \{\mathbf{S} \in \text{Herm} \mid \sigma(\mathbf{S}) \subset \mathbb{R}^+\}$ and $\text{PSym} := \{\mathbf{S} \in \text{Sym} \mid \sigma(\mathbf{S}) \subset \mathbb{R}^+\}$.

Theorem 15 *Let $\mathbf{L} \in \text{Lin}(\mathcal{V})$ for the vector space \mathcal{V} with $\det(\mathbf{L}) \neq 0$. Then there exist $\mathbf{U}, \mathbf{V} \in \text{PHerm}(\mathcal{V})$ and $\mathbf{Q} \in \text{Unit}(\mathcal{V})$ with*

$$\mathbf{L} = \mathbf{Q}\mathbf{U} \quad \text{Right Polar Decomposition} \quad (78)$$

$$= \mathbf{V}\mathbf{Q} \quad \text{Left Polar Decomposition.} \quad (79)$$

Proof: Note first that $\mathbf{L}^*\mathbf{L} \in \text{PHerm}(\mathcal{V})$ (Why?). Define \mathbf{U} to be the (unique) positive definite Hermitian square-root of $\mathbf{L}^*\mathbf{L}$, i.e. $\mathbf{U} := \sqrt{\mathbf{L}^*\mathbf{L}}$ and then define

$$\mathbf{Q} := \mathbf{L}\mathbf{U}^{-1}. \quad (80)$$

It follows that $\mathbf{Q}^* = \mathbf{U}^{-1}\mathbf{L}^*$ (Why?) and hence that

$$\mathbf{Q}^*\mathbf{Q} = \mathbf{U}^{-1}\mathbf{L}^*\mathbf{L}\mathbf{U}^{-1} = \mathbf{U}^{-1}\mathbf{U}^2\mathbf{U}^{-1} = \mathbf{I}.$$

Therefore, \mathbf{Q} is unitary and (78) holds. One now defines

$$\mathbf{V} := \mathbf{Q}\mathbf{U}\mathbf{Q}^*$$

and observes that \mathbf{V} is Hermitian and positive definite thereby proving (79). Note also that

$$\mathbf{V}^2 = \mathbf{L}\mathbf{L}^*.\square$$

Remark 41 An important application of the Polar Decomposition is to the definition of strain in nonlinear elasticity. To that end, let \mathbf{F} denote a deformation gradient.⁶ Then one defines the Cauchy-Green strains by

$$\mathbf{C} := \mathbf{F}^*\mathbf{F} \quad \text{Right Cauchy-Green Strain}$$

$$\mathbf{B} := \mathbf{F}\mathbf{F}^* \quad \text{Left Cauchy-Green Strain.}$$

Defining the *Left and Right Stretches* by

$$\mathbf{U} := \sqrt{\mathbf{C}} \quad \text{Right Stretch}$$

$$\mathbf{V} := \sqrt{\mathbf{B}} \quad \text{Left Stretch,}$$

the left and right polar decompositions of \mathbf{F} give

$$\mathbf{F} = \mathbf{V}\mathbf{Q} = \mathbf{Q}\mathbf{U}$$

which gives a multiplicative decomposition of the deformation gradient into a product of a stretch tensor and a rigid rotation.

⁶The term *deformation gradient* will be defined rigorously in the *Notes on Tensor Analysis*.

5.4 Singular Value Decomposition

As noted previously, not all square (real or complex) matrices can be diagonalized, that is become diagonal relative to a suitable basis. Indeed, only those $(N \times N)$ -matrices possessing a basis for \mathbb{C}^N consisting of eigenvectors can be diagonalized. The *Singular Value Decomposition* gives a representation reminiscent of the spectral decomposition for any (not necessarily square) matrix.

Theorem 16 *Let $\mathbf{A} \in \mathcal{M}^{M,N}(\mathbb{C})$ be an arbitrary $M \times N$ matrix. Then there exists unitary matrices $\mathbf{V} \in \mathcal{M}^M(\mathbb{C})$ and $\mathbf{W} \in \mathcal{M}^N(\mathbb{C})$ and a diagonal matrix $\mathbf{D} \in \mathcal{M}^{M,N}(\mathbb{C})$ satisfying*

$$\mathbf{A} = \mathbf{VDW}^* \quad (81)$$

with $\mathbf{D} = \text{diag}(\sigma_j)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > \sigma_{k+1} = \dots = \sigma_L = 0$ where k is the rank of \mathbf{A} and $L := \min\{M, N\}$.⁷

Remark 42 The nonnegative numbers $\{\sigma_1, \dots, \sigma_L\}$ are called the *Singular Values* for the matrix \mathbf{A} .

Proof: It suffices to assume $M \leq N$ since otherwise, one need merely prove the result for \mathbf{A}^* and take the adjoint of the resulting decomposition. The first step in the proof is to take the adjoint of the desired decomposition (81)

$$\mathbf{A}^* = \mathbf{WD}^*\mathbf{V}^*$$

and form the $M \times M$ Hermitian matrix

$$\mathbf{AA}^* = \mathbf{VDD}^*\mathbf{V}^*. \quad (82)$$

The Hermitian matrix \mathbf{AA}^* has nonnegative eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_k^2 > \sigma_{k+1}^2 = \dots = \sigma_M^2 = 0$ and (82) is its spectral decomposition. Define $\mathbf{D} := \text{diag}(\sigma_1, \dots, \sigma_M)$ and \mathbf{V} to be the unitary matrix in the spectral decomposition (82). The unitary matrix \mathbf{W}^* must satisfy

$$\mathbf{V}^*\mathbf{A} = \mathbf{DW}^*$$

or equivalently

$$\mathbf{A}^*\mathbf{V} = \mathbf{WD}^T. \quad (83)$$

The $N \times N$ matrix \mathbf{W} will be constructed by specifying its column vectors

$$\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_N]. \quad (84)$$

The left hand side of (83) has the form

$$\mathbf{A}^*\mathbf{V} = [(\mathbf{A}^*\mathbf{v}_1) \ \dots \ (\mathbf{A}^*\mathbf{v}_M)] \quad (85)$$

where \mathbf{V} has the column vectors \mathbf{v}_j , $j = 1, \dots, M$. On the other hand, the right hand side of (83) has the column structure

$$\mathbf{WD}^T = [(\sigma_1\mathbf{w}_1) \ \dots \ (\sigma_k\mathbf{w}_k) \ \mathbf{0} \ \dots \ \mathbf{0}] \quad (86)$$

⁷A general matrix $\mathbf{D} = [d_{ij}] \in \mathcal{M}^{M,N}$ is called *diagonal* if $d_{ij} = 0$ for $i \neq j$.

in which the last $M - k$ columns are the (N -dimensional) zero vector. Note that

$$\mathbf{A}^* \mathbf{v}_j = \mathbf{0}, \quad \text{for } j = k + 1, \dots, M$$

as can be seen from

$$0 = (\mathbf{A} \mathbf{A}^* \mathbf{v}_j) \cdot \mathbf{v}_j = (\mathbf{A}^* \mathbf{v}_j) \cdot (\mathbf{A}^* \mathbf{v}_j) = |\mathbf{A}^* \mathbf{v}_j|^2 \quad \text{for } j = k + 1, \dots, M$$

which follows from

$$\mathbf{A} \mathbf{A}^* \mathbf{v}_j = \mathbf{0} \quad \text{for } j = k + 1, \dots, M.$$

Equating columns in (85) and (86), one defines the first k -columns of \mathbf{W} by

$$\mathbf{w}_j := \frac{1}{\sigma_j} \mathbf{A}^* \mathbf{v}_j, \quad \text{for } j = 1, \dots, k.$$

These k vectors $\mathbf{w}_1, \dots, \mathbf{w}_k$ are easily seen to form an orthonormal set. (Why?) One now defines the remaining $N - k$ column vectors of \mathbf{W} in such a way that the set $\{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ is orthonormal thereby proving the theorem. \square

5.5 Solving Equations

An important application of linear algebra is to the problem of solving simultaneous systems of linear algebraic equations. Specifically, let $\mathbf{A} \in \mathcal{M}^{M,N}(\mathbb{C})$ and consider the system of linear equations

$$\mathbf{A} \mathbf{x} = \mathbf{b} \tag{87}$$

where $\mathbf{b} \in \mathbb{C}^M$ is given and $\mathbf{x} \in \mathbb{C}^N$ is sought. Thus, (87) is a system of M linear equations in N unknowns. In general, the system (87) can have no solutions, a unique solution or infinitely many solutions. If \mathbf{A} is regarded as defining a linear transformation from \mathbb{C}^N into \mathbb{C}^M , then (87) has solutions if and only if $\mathbf{b} \in \text{Ran}(\mathbf{A})$ and such solutions are unique if and only if $\text{Nul}(\mathbf{A}) = \{\mathbf{0}\}$. Moreover, $\text{Ran}(\mathbf{A})$ is the span of the columns of \mathbf{A} . However, testing whether or not $\mathbf{b} \in \text{Ran}(\mathbf{A})$ can be computationally expensive. The *Fredholm Alternative Theorem* gives another characterization of $\text{Ran}(\mathbf{A})$ that is often much easier to check.

Theorem 17 *Let $\mathbf{A} \in \mathcal{M}^{M,N}(\mathbb{C})$. Then*

$$\text{Ran}(\mathbf{A}) = (\text{Nul}(\mathbf{A}^*))^\perp. \tag{88}$$

Proof: First, note that if $\mathbf{b} \in \text{Ran}(\mathbf{A})$ and $\mathbf{a} \in \text{Nul}(\mathbf{A}^*)$, then $\mathbf{b} = \mathbf{A} \mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^N$ and

$$\langle \mathbf{b}, \mathbf{a} \rangle = \langle \mathbf{A} \mathbf{x}, \mathbf{a} \rangle = \langle \mathbf{x}, \mathbf{A}^* \mathbf{a} \rangle = 0.$$

Thus,

$$\text{Ran}(\mathbf{A}) \subset \text{Nul}(\mathbf{A}^*)^\perp. \tag{89}$$

But one also has

$$\begin{aligned}
\dim(\text{Nul}(\mathbf{A}^*)^\perp) &= M - \dim(\text{Nul}(\mathbf{A}^*)) \quad (\text{Why?}) \\
&= \dim(\text{Ran}(\mathbf{A}^*)) \quad (\text{Why?}) \\
&= \dim(\text{Ran}(\mathbf{A})). \quad (\text{Why?})
\end{aligned} \tag{90}$$

Combining (89) and (90) one concludes that

$$\text{Ran}(\mathbf{A}) = \text{Nul}(\mathbf{A}^*)^\perp \quad (\text{Why?})$$

as required. \square

Remark 43 The Fredholm Alternative Theorem says that (87) has solutions if and only if \mathbf{b} is orthogonal to the null-space of \mathbf{A}^* . Usually, the null-space of \mathbf{A}^* has small dimension, so testing for $\mathbf{b} \in \text{Nul}(\mathbf{A}^*)^\perp$ is often much easier than testing to see if it is in the span of the columns of \mathbf{A} .

In many applications, the system (87) is badly over determined (i.e. far more independent equations than unknowns) and one does not expect there to be a solution. In such cases, one might want to find best approximate solutions. A typical such example arises in linear regression in which one wishes to find a linear function that “best fits” a large number of data points, where best is usually in the least square sense. These ideas are made rigorous through the concept of *Minimum Norm Least Square Solution* to (87).

5.5.1 Minimum Norm Least Squares Solution

A *Least Squares Solution* to (87) is any $\mathbf{x} \in \mathbb{R}^N$ that minimizes the *Residual Error* $|\mathbf{Ax} - \mathbf{b}|$. Let $\mathcal{Q} = \{\mathbf{x} \in \mathbb{R}^N \mid |\mathbf{Ax} - \mathbf{b}| = \min_{\mathbf{z} \in \mathbb{R}^N} |\mathbf{Az} - \mathbf{b}|\}$. The set \mathcal{Q} is always non-empty but may contain many vectors \mathbf{x} . Indeed, form the direct sum decomposition

$$\mathbb{R}^N = \text{Ran}(\mathbf{A}) \oplus \text{Ran}(\mathbf{A})^\perp$$

and write

$$\mathbf{b} = \mathbf{b}_r + \mathbf{b}_0 \tag{91}$$

where

$$\mathbf{b}_r \in \text{Ran}(\mathbf{A}) \quad \text{and} \quad \mathbf{b}_0 \in \text{Ran}(\mathbf{A})^\perp = \text{Nul}(\mathbf{A}^T). \tag{92}$$

Then, $\mathbf{Ax} - \mathbf{b}_r \in \text{Ran}(\mathbf{A})$ and

$$|\mathbf{Ax} - \mathbf{b}|^2 = |\mathbf{Ax} - \mathbf{b}_r|^2 + |\mathbf{b}_0|^2 \quad (\text{Why?}). \tag{93}$$

Clearly, $\mathbf{x} \in \mathcal{Q}$ if and only if $\mathbf{Ax} = \mathbf{b}_r$ and

$$|\mathbf{b}_0| = \min_{\mathbf{z} \in \mathbb{R}^N} |\mathbf{Az} - \mathbf{b}|. \tag{94}$$

Since $\mathbf{b}_r \in \text{Ran}(\mathbf{A})$, \mathcal{Q} is non-empty. (Why?) Another characterization of \mathcal{Q} is that $\mathbf{x} \in \mathcal{Q}$ if and only if $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}$. To see why, observe that if $\mathbf{x} \in \mathcal{Q}$, then $\mathbf{Ax} = \mathbf{b}_r$ from which it follows that

$$\begin{aligned}
\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) &= \mathbf{A}^T(\mathbf{Ax} - \mathbf{b}_r) + \mathbf{A}^T\mathbf{b}_0 \\
&= \mathbf{A}^T(\mathbf{Ax} - \mathbf{b}_r) = \mathbf{0}.
\end{aligned}$$

Conversely, if $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) = 0$ then $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}_r) = 0$. Also $\text{Nul}(\mathbf{A}^T) = \text{Ran}(\mathbf{A})^\perp$. Hence, $\mathbf{Ax} - \mathbf{b}_r \in \text{Ran}(\mathbf{A})^\perp$. But since it is obvious that $(\mathbf{Ax} - \mathbf{b}_r) \in \text{Ran}(\mathbf{A})$, one concludes that $(\mathbf{Ax} - \mathbf{b}_r) = 0$.

Next, note that \mathcal{Q} is a convex subset of \mathbb{R}^N . To see this, let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{Q}$ and let $\alpha_1, \alpha_2 > 0$ with $\alpha_1 + \alpha_2 = 1$. Then,

$$\begin{aligned} \mathbf{A}^T \mathbf{A}(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2) &= \alpha_1 \mathbf{A}^T \mathbf{A} \mathbf{x}_1 + \alpha_2 \mathbf{A}^T \mathbf{A} \mathbf{x}_2 \\ &= \alpha_1 \mathbf{A}^T \mathbf{b}_r + \alpha_2 \mathbf{A}^T \mathbf{b}_r \\ &= \mathbf{A}^T \mathbf{b}_r (\alpha_1 + \alpha_2) \\ &= \mathbf{A}^T \mathbf{b}_r \end{aligned}$$

which implies that $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 \in \mathcal{Q}$ and hence that \mathcal{Q} is convex.

It follows that \mathcal{Q} has a unique vector of minimum norm, i.e. closest to the origin $\mathbf{0}$. (Why?) This motivates the following

Definition 34 *The Minimum Norm Least Squares Solution to the linear system (87) is the unique element $\mathbf{x} \in \mathcal{Q}$ such that $|\mathbf{x}| = \min_{\mathbf{z} \in \mathcal{Q}} |\mathbf{z}|$.*

While the minimum norm least squares solution is well-defined, there remains the question of how to find it. The answer lies in the *Moore-Penrose Pseudo-Inverse* of \mathbf{A} .

5.5.2 Moore-Penrose Pseudo-Inverse

The Moore-Penrose Pseudo-Inverse of \mathbf{A} is defined through its singular value decomposition. (81).

Definition 35 *Given any matrix $\mathbf{A} \in \mathcal{M}^{M,N}(\mathbb{C})$ with Singular Value Decomposition $\mathbf{A} = \mathbf{VDW}^*$, its Moore-Penrose Pseudo-Inverse $\mathbf{A}^\dagger \in \mathcal{M}^{N,M}$ is defined by*

$$\mathbf{A}^\dagger := \mathbf{WD}^\dagger \mathbf{V}^* \quad (95)$$

where $\mathbf{D}^\dagger \in \mathcal{M}^{N,M}$ is the $N \times M$ diagonal matrix

$$\mathbf{D}^\dagger := \text{diag}(\delta_j) \quad (96)$$

with

$$\delta_j := \frac{1}{\sigma_j} \quad \text{for } j = 1, \dots, k, \text{ and } \delta_j = 0 \text{ for } j = k + 1, \dots, L \quad (97)$$

and $L = \min\{M, N\}$, $k := \text{Rank}(\mathbf{A})$.

Remark 44 Note that if $M \leq N$ and $k = \text{Rank}(\mathbf{A}) = M$, then \mathbf{A}^\dagger is a right inverse of \mathbf{A} , i.e. $\mathbf{AA}^\dagger = I_M$ where I_M denotes the $M \times M$ identity matrix. Similarly, if $N \leq M$ and $k = \text{Rank}(\mathbf{A}) = N$, then \mathbf{A}^\dagger is a left inverse of \mathbf{A} , i.e. $\mathbf{A}^\dagger \mathbf{A} = I_N$ the $N \times N$ identity matrix.

The principal use of the Moore-Penrose pseudo-inverse is given in the following theorem.

Theorem 18 *The Minimum Norm Least Squares solution of the system of equations $\mathbf{Ax} = \mathbf{b}$ is given by*

$$\mathbf{x} = \mathbf{A}^\dagger \mathbf{b}.$$

Proof: Consider the singular value decomposition $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{W}^T$ and suppose $\text{Rank}(\mathbf{A}) = k$. Moreover, define $\mathbf{a} := \mathbf{W}^T \mathbf{x}$ and let \mathbf{a} have components $[\mathbf{a}] = [a_j]$. Also, assume \mathbf{V} has the column structure $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_M]$. Observe first that

$$\begin{aligned} |\mathbf{A}\mathbf{x} - \mathbf{b}|^2 &= |\mathbf{V}^T(\mathbf{A}\mathbf{x} - \mathbf{b})|^2 \text{ (Why?)} \\ &= |\mathbf{V}^T \mathbf{A} \mathbf{W} \mathbf{W}^T \mathbf{x} - \mathbf{V}^T \mathbf{b}|^2 \\ &= |\mathbf{D} \mathbf{W}^T \mathbf{x} - \mathbf{V}^T \mathbf{b}|^2 \\ &= \sum_{j=1}^k (\sigma_j a_j - \mathbf{v}_j \cdot \mathbf{b})^2 + \sum_{j=k+1}^M (\mathbf{v}_j \cdot \mathbf{b})^2. \end{aligned}$$

It follows that $|\mathbf{A}\mathbf{x} - \mathbf{b}|$ is minimized by choosing

$$a_j = \frac{1}{\sigma_j} \mathbf{v}_j \cdot \mathbf{b}, \quad \text{for } j = 1, \dots, k. \quad (98)$$

Also,

$$|\mathbf{x}|^2 = |\mathbf{W}^T \mathbf{x}|^2 = a_1^2 + \dots + a_k^2 + a_{k+1}^2 + \dots + a_N^2$$

which is minimized by choosing

$$a_{k+1} = \dots = a_N = 0.$$

One now sees that (How?)

$$\mathbf{x} = \mathbf{W}\mathbf{D}\mathbf{V}^T \mathbf{b} = \mathbf{A}^\dagger \mathbf{b}$$

thereby proving the theorem. \square

A natural question to ask is: How are the singular value decomposition and spectral decomposition related? This answer is in the following

Remark 45 Suppose $\mathbf{A} \in \mathcal{M}^N(\mathbb{C})$ is normal, i.e. $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$. Then \mathbf{A} has a spectral decomposition $\mathbf{A} = \mathbf{V}\text{diag}(\lambda_1, \dots, \lambda_N)\mathbf{V}^*$ with \mathbf{V} unitary and the spectrum $\sigma(\mathbf{A}) \subset \mathbb{C}$. Write $\lambda_j = |\lambda_j|e^{i\theta_j}$ and define the diagonal matrix $\mathbf{D} := \text{diag}(|\lambda_1|, \dots, |\lambda_N|)$ and unitary matrix $\mathbf{W} := \mathbf{V}\mathbf{D}$ (Why is it unitary?). Then \mathbf{A} has the singular value decomposition $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{W}^*$. (Verify.)

Remark 46 A bit more surprising is the following relationship between the polar decomposition and the singular value decomposition. Suppose $\mathbf{A} \in \mathcal{M}^N(\mathbb{R})$ with $\det(\mathbf{A}) \neq 0$ has polar decomposition $\mathbf{A} = \mathbf{U}\mathbf{Q}$ with $\mathbf{U} \in \text{PSym}$ and $\mathbf{Q} \in \text{Orth}^+$. Then \mathbf{U} has a spectral decomposition $\mathbf{U} = \mathbf{V}\mathbf{D}\mathbf{V}^T$ with $\mathbf{V} \in \text{Orth}$, $\mathbf{D} = \text{diag}(\sigma_1, \dots, \sigma_N)$ and spectrum $\{\sigma_1, \dots, \sigma_N\} \subset \mathbb{R}^+$. Then \mathbf{A} has the singular value decomposition $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{W}^T$ with $\mathbf{W} := \mathbf{Q}^T \mathbf{V}$. (Verify.)

Example 18 Let \mathbf{A} be the $1 \times N$ matrix $\mathbf{A} := [a_1 \dots a_N]$. Then show that the singular value decomposition of $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{W}^*$ with $\mathbf{V} = [1]$ being the 1×1 identity matrix, $\mathbf{D} = [|\mathbf{a}| \ 0 \dots 0]$ and $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_N]$ with $\mathbf{w}_1 := \mathbf{a}/|\mathbf{a}|$ and $\mathbf{w}_2, \dots, \mathbf{w}_N$ chosen so that $\{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ forms an orthonormal basis for \mathbb{C}^N and the N -tuple \mathbf{a} defined by

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}.$$

Exercise 5 Find the minimum norm least squares solution to the $N \times 1$ system of equations

$$\begin{aligned} a_1 x &= b_1 \\ &\vdots \\ a_N x &= b_N. \end{aligned}$$

(**Hint:** To find the singular value decomposition for the $N \times 1$ matrix

$$\mathbf{A} := \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}$$

first construct the SVD for its $1 \times N$ transpose $\mathbf{A}^T = [a_1 \ \dots \ a_N]$.)

6 Rayleigh-Ritz Theory

Thus far, eigenvalues and eigenvectors have been studied from an algebraic and geometric perspective. However, they also have a *variational* characterization that is revealed in the *Rayleigh-Ritz Theory*. The first result considered is the foundational *Rayleigh-Ritz Theorem*.

Theorem 19 Let $\mathbf{L} \in (\mathcal{V}, \langle \cdot, \cdot \rangle)^8$ be self-adjoint, i.e. $\mathbf{L}^* = \mathbf{L}$. Let its (real) eigenvalues be ordered $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$. Then

$$\lambda_1 \leq \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle \leq \lambda_N \quad \text{for all } \mathbf{v} \in \mathcal{V}, |\mathbf{v}| = 1 \quad (99)$$

$$\lambda_{\max} = \lambda_N = \sup_{|\mathbf{v}|=1} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle \quad (100)$$

$$\lambda_{\min} = \lambda_1 = \inf_{|\mathbf{v}|=1} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle. \quad (101)$$

Proof: By the Spectral Theorem, there is an orthonormal basis $\{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ for \mathcal{V} consisting of eigenvectors of \mathbf{L} satisfying $\mathbf{L}\mathbf{b}_j = \lambda_j \mathbf{b}_j$ for $j = 1, \dots, N$. It follows that for every unit vector $\mathbf{v} \in \mathcal{V}$,

$$\langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle = \sum_{j=1}^N \lambda_j |\langle \mathbf{b}_j, \mathbf{v} \rangle|^2 \quad (\text{Why?}) \quad (102)$$

and

$$|\langle \mathbf{b}_1, \mathbf{v} \rangle|^2 + \dots + |\langle \mathbf{b}_N, \mathbf{v} \rangle|^2 = 1. \quad (\text{Why?}) \quad (103)$$

Equation (102) says that the quadratic form $\langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle$ for unit vectors \mathbf{v} is a weighted average of the eigenvalues of \mathbf{L} . One then immediately deduces the inequalities (99). (How?) Moreover, choosing $\mathbf{v} = \mathbf{b}_N$ and \mathbf{b}_1 , gives (100,101), respectively. (How?) \square

The Rayleigh-Ritz theorem gives a variational characterization of the smallest and largest eigenvalues of a self-adjoint linear transformation. However, an easy extension of the theorem characterizes all of the eigenvalues as solutions to constrained variational problems.

⁸Unless otherwise indicated, throughout this section $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ is assumed to be a complex inner product space.

Theorem 20 *With the same hypotheses and notation as the previous theorem, the eigenvalues $\lambda_1 \leq \dots \leq \lambda_N$ of \mathbf{L} satisfy*

$$\lambda_k = \min_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \{\mathbf{b}_1, \dots, \mathbf{b}_{k-1}\}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle \quad (104)$$

and

$$\lambda_{N-k} = \max_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \{\mathbf{b}_N, \dots, \mathbf{b}_{N-k+1}\}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle. \quad (105)$$

Proof: The proof of the previous theorem is easily generalized to prove (104,105). To that end, suppose \mathbf{v} is a unit vector orthogonal to the first $k-1$ eigenvectors, $\{\mathbf{b}_1, \dots, \mathbf{b}_{k-1}\}$. Then one easily shows (How?) that

$$|\langle \mathbf{v}, \mathbf{b}_k \rangle|^2 + \dots + |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 = 1$$

and

$$\langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle = \sum_{j=1}^N \lambda_j |\langle \mathbf{v}, \mathbf{b}_j \rangle|^2 = \sum_{j=k}^N \lambda_j |\langle \mathbf{v}, \mathbf{b}_j \rangle|^2 \geq \lambda_k. \quad (\text{Why?}) \quad (106)$$

But one also has $\mathbf{b}_k \perp \{\mathbf{b}_1, \dots, \mathbf{b}_{k-1}\}$ and

$$\langle \mathbf{L}\mathbf{b}_k, \mathbf{b}_k \rangle = \lambda_k. \quad (107)$$

The result (104) now follows readily from (106,107). The proof of (105) is similar and left as an exercise. \square

The Rayleigh-Ritz theory is a powerful theoretical tool for studying eigenvalues for a self-adjoint linear transformation. However, the necessity of knowing explicitly the eigenvectors in order to estimate eigenvalues other than the smallest and largest seriously limits its use as a practical tool. This limitation can be avoided through the following generalization known as the *Courant-Fisher Theorem* which gives a min-max and max-min characterization of the eigenvalues of a self-adjoint linear transformation.

Theorem 21 *With the same hypotheses and notation of the previous two theorems, the eigenvalues $\lambda_1, \dots, \lambda_N$ of \mathbf{L} satisfy*

$$\lambda_k = \min_{\mathbf{w}_1, \dots, \mathbf{w}_{N-k}} \max_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \{\mathbf{w}_1, \dots, \mathbf{w}_{N-k}\}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle \quad (108)$$

and

$$\lambda_k = \max_{\mathbf{w}_1, \dots, \mathbf{w}_{k-1}} \min_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \{\mathbf{w}_1, \dots, \mathbf{w}_{k-1}\}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle. \quad (109)$$

Proof: From the Spectral Theorem it follows that for any unit vector $\mathbf{v} \in \mathcal{V}$

$$\begin{aligned} \mathbf{v} &= \mathbf{b}_1 \langle \mathbf{v}, \mathbf{b}_1 \rangle + \dots + \mathbf{b}_N \langle \mathbf{v}, \mathbf{b}_N \rangle \\ \mathbf{L}\mathbf{v} &= \lambda_1 \mathbf{b}_1 \langle \mathbf{v}, \mathbf{b}_1 \rangle + \dots + \lambda_N \mathbf{b}_N \langle \mathbf{v}, \mathbf{b}_N \rangle \\ 1 &= |\langle \mathbf{v}, \mathbf{b}_1 \rangle|^2 + \dots + |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \end{aligned}$$

and hence that

$$\langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle = \lambda_1 |\langle \mathbf{v}, \mathbf{b}_1 \rangle|^2 + \dots + \lambda_N |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2.$$

That (108,109) are true for $k = 1$ or N follows immediately from the Rayleigh-Ritz Theorem. As a special case, consider first $k = N - 1$. Then (108) takes the form

$$\lambda_{N-1} = \min_{\mathbf{w}} \max_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \mathbf{w}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle. \quad (110)$$

The idea now is to choose amongst all the unit vectors \mathbf{v} orthogonal to a given vector \mathbf{w} only those lying in the plane orthogonal to $\{\mathbf{b}_1, \dots, \mathbf{b}_{N-2}\}$. Such unit vectors \mathbf{v} must satisfy

$$\begin{aligned} 1 &= |\langle \mathbf{v}, \mathbf{b}_{N-1} \rangle|^2 + |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \quad (\text{Why?}) \\ \mathbf{v} &= \mathbf{b}_{N-1} \langle \mathbf{v}, \mathbf{b}_{N-1} \rangle + \mathbf{b}_N \langle \mathbf{v}, \mathbf{b}_N \rangle \quad (\text{Why?}) \\ \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle &= \lambda_{N-1} |\langle \mathbf{v}, \mathbf{b}_{N-1} \rangle|^2 + \lambda_N |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2. \quad (\text{Why?}) \end{aligned}$$

It follows that (Why?)

$$\max_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \mathbf{w}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle \geq \max_{\substack{1=|\langle \mathbf{v}, \mathbf{b}_{N-1} \rangle|^2 + |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \\ \mathbf{v} \perp \mathbf{w}}} \lambda_{N-1} |\langle \mathbf{v}, \mathbf{b}_{N-1} \rangle|^2 + \lambda_N |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \geq \lambda_{N-1}. \quad (111)$$

Since (111) holds for all vectors \mathbf{w} , one concludes that

$$\min_{\mathbf{w}} \max_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \mathbf{w}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle \geq \lambda_{N-1}. \quad (112)$$

The desired result (110) now follows from (112) and the observation that

$$\max_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \mathbf{b}_N}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle = \lambda_{N-1}$$

due to the Rayleigh-Ritz theorem.

The proof of (108) for all eigenvalues λ_k proceeds along the same lines as that just given for λ_{N-1} . More specifically, one chooses amongst all the unit vectors \mathbf{v} orthogonal to a given set of $N - k$ vectors $\{\mathbf{w}_1, \dots, \mathbf{w}_{N-k}\}$ only those orthogonal to $\{\mathbf{b}_1, \dots, \mathbf{b}_{k-1}\}$ (Why can this be done?). Then such \mathbf{v} must satisfy

$$\begin{aligned} 1 &= |\langle \mathbf{v}, \mathbf{b}_k \rangle|^2 + \dots + |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \\ \mathbf{v} &= \mathbf{b}_k \langle \mathbf{v}, \mathbf{b}_k \rangle + \dots + \mathbf{b}_N \langle \mathbf{v}, \mathbf{b}_N \rangle \\ \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle &= \lambda_k |\langle \mathbf{v}, \mathbf{b}_k \rangle|^2 + \dots + \lambda_N |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2. \end{aligned}$$

It follows that (Why?)

$$\max_{\substack{|\mathbf{v}|=1 \\ \mathbf{v} \perp \{\mathbf{w}_1, \dots, \mathbf{w}_{N-k}\}}} \langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle \geq \max_{\substack{1=|\langle \mathbf{v}, \mathbf{b}_k \rangle|^2 + \dots + |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \\ \mathbf{v} \perp \{\mathbf{w}_1, \dots, \mathbf{w}_{N-k}\}}} \lambda_k |\langle \mathbf{v}, \mathbf{b}_k \rangle|^2 + \dots + \lambda_N |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \geq \lambda_k. \quad (113)$$

The desired result (108) now follows from (113) and the fact that

$$\max_{\substack{1=|\langle \mathbf{v}, \mathbf{b}_k \rangle|^2 + \dots + |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 \\ \mathbf{v} \perp \{\mathbf{b}_{k+1}, \dots, \mathbf{b}_N\}}} \lambda_k |\langle \mathbf{v}, \mathbf{b}_k \rangle|^2 + \dots + \lambda_N |\langle \mathbf{v}, \mathbf{b}_N \rangle|^2 = \lambda_k$$

due to the Rayleigh-Ritz Theorem. The proof of (109) proceeds along similar lines and is left as an exercise. \square

Remark 47 The Rayleigh-Ritz Theorem (105) says that the k -th eigenvalue, λ_k , is the maximum value of the *Rayleigh Quotient*

$$\frac{\langle \mathbf{L}\mathbf{v}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \tag{114}$$

over the k -dimensional subspace $\{\mathbf{b}_{k+1}, \dots, \mathbf{b}_N\}^\perp = \text{span}\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$. The Courant-Fisher Theorem (108) says that the maximum of the Rayleigh quotient (114) over an arbitrary k -dimensional subspace is always at least as large as λ_k .