Advanced Linear Algebra

Professor David Surowski
Notes by Ali Mohammad

Preface/Acknowledgement

The present set of notes represents, pretty faithfully in fact, the content of a course *Advanced Linear Algebra* that I gave during the Summer Term of 1997. The main objective was to present a proof of the *Spectral Theorem* for normal operators on a finite-dimensional complex Hilbert space. Indeed, I feel that this subject matter is missing from our "basic" undergraduate or graduate curricula. To my knowledge, the only systematic approach to this important theorem comes in our Functional Analysis course, where the student is expected to have a good command of both measure theory and the general theory of operators on an infinite-dimensional Hilbert space. The present treatment, given that the ambient spaces are all finite-dimensional, avoids all of the analytic subtleties that necessarily occur in the infinite-dimensional case. Furthermore, a student who has mastered the Spectral Theorem in the present setting should have no difficulty in taking the next logical step and treat the Spectral Theorem for normal compact operators on a Hilbert space, where the analytic prerequisites are not nearly as intimidating as in the general (non-compact) case. It is in this sense that I hope that the present treatment will serve as a bridge between algebra and some of the functional analysis that is central to the research of many of my distinguished colleagues.

Apart from the Spectral Theorem itself, it is my hope that the students will have increased their own levels of sophistication in linear algebra in general. To this end, I have tried to emphasize quotient spaces from very early on, with the hope and expectation that the students will eventually embrace their utility and intrinsic beauty. Furthermore, linear algebra is probably the simplest context in which to study quotients, owing primarily to the "splitting" property of vector spaces. Since quotient structures occur naturally in virtually every sub-discipline within mathematics, a student who starts to appreciate which situations naturally call for the construction or analysis of a quotient structure is a student who has already started to think on a higher plane.

Other "high points" in the development include *linear transformations* and their *matrix representations* (this is the bridge between "linear algebra" and "matrix theory"); *dual spaces*, *minimal polynomials*, and the *Primary Decomposition Theorem.* I've tried several times to emphasize that the Primary Decomposition Theorem is really almost a trivial application of the "Euclidean trick;" the more advanced students have throughout been encouraged to think of this result as being a direct analog of the splitting of a finite abelian group into the direct product of Sylow subgroups. Despite the simplicity of the Primary Decomposition Theorem, it is my hope that the students see the embryonic stages of the Spectral Theorem – especially the appearance of the orthogonal idempotents.

We all owe a great debt to Ali Mohammad, who not only took painstakingly clear notes, he invested countless additional hours to convert the written notes to a LaTeX file, thus making a record of this Summer's activities available to all. Furthermore, I am indebted to my friend and colleague, Bob Burckel, who invested the better part of his August, 1999, transatlantic flight to Germany in a critical reading of my notes. As if this weren't enough, he gave the chapter on Fourier analysis and quadratic reciprocity a valuable critical reading. In all of this, not only did he identify and correct some of my careless mathematical mistakes, he also took great pains to correct a large portion of my "stream of consciousness" writing style. The present set of notes is far better as a result of Professor Burckel's efforts.

David Surowski, July 30, 1997
Second revision, July 19, 1999
Third revision, January 9, 2000
Fourth revision, July 31, 2001
Fifth revision, July 24, 2002

Course Outline for Summer 1997 Course

It is assumed that the present course represents for each student at least the second exposure to linear algebra. Some of the students will have had multiple excursions into linear algebra – this is not a bad thing, as it is my firm conviction that a student cannot have“too much” linear algebra, owing to its fundamental importance in virtually all branches of mathematics as well as to a very wide range of “applied disciplines.”

That this course is being run under the title “Applied Matrix Theory,” is probably a misnomer, as it is my intention to convert it into a customized course in linear algebra, to be offered every other Summer, with sufficiently varying subject matter that students will be allowed to retake the course for credit. The present course (Summer, 1997) has the primary objective of providing the students with a carefully laid-out proof of the *Spectral Theorem for Normal Operators on a Finite-Dimensional Complex Hilbert Space.* I can think of at least two reasons motivating this particular choice at this particular time. The first, quite simply, is that I cannot locate this very fundamental result anywhere in our graduate curriculum, unless of course, one counts the vastly more general “Spectral Theorems” that are routinely discussed in treatments of functional analysis. However, much of the purely algebraic flavor can be gleaned in the finite-dimensional case, without the technical analytical and topological subtleties that occur in the infinite-dimensional case. This leads me to the second reason: many of my colleagues have been quietly (and not-so-quietly) grousing about the graduate students' lack of background in the study of a single linear transformation on a finite-dimensional vector space. The Spectral Theorem, which can be thought of in the more general context of “Jordan Canonical Forms,” is a particularly beautiful and complete result, made possible by the introduction into the vector space of a seemingly incidental[1] additional structure: a Hermitian inner product.

Since I have never tried to teach this material before in a Summer course, I don't know how the time allotment will work out. I'm confident, however, that we'll have enough time to cover the Spectral Theorem. If we have time left over, I hope to make a brief sally into a discussion of normed and Banach spaces and maybe give a brief introduction to multilinear constructions.

The brief outline of the course (at least for the material through the “Spectral Theorem”) is as follows:

[1] This is only a perception; one of my favorite manifestations of this structure “in nature” is the mathematical study of quantum mechanics in physics.

I. Basics

1. Vector Spaces and Subspaces
2. Basis and Dimension
3. Linear Transformations and Matrix Representation
4. Quotient Spaces, Isomorphism Theorems and the Rank-Nullity Theorem
5. Dual Spaces
6. Bilinear Forms and Duality

II. Eigenvalues and Eigenvectors

1. Basic Definitions
2. Characteristic Polynomial and Minimal Polynomial
3. Diagonalizability of Matrix Representation

III. Inner Product Spaces

1. Real-Symmetric and Complex-Hermitian Inner Product Spaces
2. Riesz Representation Theorem
3. Self-Adjoint and Normal Operators
4. The Spectral Theorem for Normal Operators

Added, July 1999: The notes themselves indicate that the above objective was pretty much met. Actually, there was some time left over, which allowed for the inclusion of material on tensor products in *Chapter* 4, giving perhaps a novel proof of the (unrestricted) *Cayley-Hamilton Theorem.* Unfortunately, there was not so much time left over to allow also for an excursion into Banach spaces, as I indicated might be possible above. Perhaps some other time ...

Contents

Chapter 1

Basics

1.1 Vector Spaces and Subspaces

1.1.1 Basic Definitions

We assume that the reader has already had some exposure to elementary linear algebra; thus, many of our definitions will be informal or incomplete.

DEFINITION. A *field* (*e.g.,* $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and others) is an algebraic system satisfying the usual associative, commutative, and distributive laws with addition and multiplication, and having multiplicative inverses for each non-zero element.

DEFINITION. A *vector space* over the field $\mathbb{F}$ is a set of objects called vectors that can be added subject to the usual rules (*e.g.*, commutativity, associativity, existence of additive identity, denoted 0. Please note that we shall not distinguish typographically between the *field scalar* 0 and the *vector* 0. This should not cause significant confusion.) In addition, there is the operation of *scalar multiplication* which satisfies $\alpha v \in V$ for all $\alpha \in \mathbb{F}, v \in V$, such that

$$\alpha(v_1 + v_2) = \alpha v_1 + \alpha v_2$$

$$(\alpha_1 + \alpha_2)v = \alpha_1 v + \alpha_2 v$$

$$0 \cdot v = 0^1$$

$$1 \cdot v = v \ (\textit{the unital property})$$

[1] Note that the "0" on the left hand side is $0 \in \mathbb{F}$; that on the right hand side is $0 \in V$.

We note that $0 \cdot v = 0$ because

$$\begin{aligned} 0 &= 0+0, \qquad \text{so} \\ 0v &= (0+0)v \\ &= 0v+0v \\ 0v-0v &= 0v+0v-0v \\ 0 &= 0v. \end{aligned}$$

1.1.2 Important Examples

The vector spaces $\mathbb{F}_m$ and $\mathbb{F}^n$. Let $\mathbb{F}$ be a field. Denote

$$\mathbb{F}_n = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) | \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}\}$$

(the set of ordered n-tuples)

$$\mathbb{F}^n = \left\{ \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \mid \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F} \right\}$$

(the set of $1 \times n$ matrices in $\mathbb{F}$)

Operations:

$$(\alpha_1, \alpha_2, \ldots, \alpha_n) + (\beta_1, \beta_2, \ldots, \beta_n) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n)$$

$$\alpha(\alpha_1, \alpha_2, \ldots, \alpha_n) = (\alpha\alpha_1, \alpha\alpha_2, \ldots, \alpha\alpha_n).$$

With the above operations, $\mathbb{F}_n$ becomes a vector space over $\mathbb{F}$.

Sample Axiom Confirmation: Let $\alpha \in \mathbb{F}$ and $v_1, v_2 \in \mathbb{F}_n$. We shall prove that:

$$\alpha(v_1 + v_2) = \alpha v_1 + \alpha v_2.$$

Denote

$$v_1 = (\alpha_1, \alpha_2, \ldots, \alpha_n)$$

and

$$v_2 = (\beta_1, \beta_2, \ldots, \beta_n).$$

Then:

$$\begin{aligned}
\alpha(v_1 + v_2) &= \alpha((\alpha_1, \alpha_2, \ldots, \alpha_n) + (\beta_1, \beta_2, \ldots, \beta_n)) \\
&= \alpha(\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n) \\
&= (\alpha(\alpha_1 + \beta_1), \alpha(\alpha_2 + \beta_2), \ldots, \alpha(\alpha_n + \beta_n)) \\
&= (\alpha\alpha_1 + \alpha\beta_1, \alpha\alpha_2 + \alpha\beta_2, \ldots, \alpha a_n + \alpha\beta_n) \\
&= (\alpha\alpha_1, \alpha\alpha_2, \ldots, \alpha\alpha_n) + (\alpha\beta_1, \alpha\beta_2, \ldots, \alpha\beta_n) \\
&= \alpha(\alpha_1, \alpha_2, \ldots, \alpha_n) + \alpha(\beta_1, \beta_2, \ldots, \beta_n) \\
&= \alpha v_1 + \alpha v_2.
\end{aligned}$$

The remaining axioms are similarly verified.

Using essentially the same arguments, $\mathbb{F}^n$ can be shown to be a vector space over $\mathbb{F}$.

The vector space of matrices $M_{mn}(\mathbb{F})$

Let $\mathbb{F}$ be a field. Set

$$M_{mn}(\mathbb{F}) = \left\{ \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{bmatrix} \mid \alpha_{ij} \in \mathbb{F} \right\}$$

(the set of $m \times n$ matrices over $\mathbb{F}$)

Addition

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{bmatrix} =$$

$$\begin{bmatrix} \alpha_{11} + \beta_{11} & \alpha_{12} + \beta_{12} & \cdots & \alpha_{1n} + \beta_{1n} \\ \alpha_{21} + \beta_{21} & \alpha_{22} + \beta_{22} & \cdots & \alpha_{2n} + \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} + \beta_{m1} & \alpha_{m2} + \beta_{m2} & \cdots & \alpha_{mn} + \beta_{mn} \end{bmatrix}$$

However, as you might have already noticed, this notation gets to be a bit tedious, so we condense it to:

$$[\alpha_{ij}] + [\beta_{ij}] = [\alpha_{ij} + \beta_{ij}]$$

$$\alpha[\alpha_{ij}] = [\alpha\alpha_{ij}].$$

As in the previous example, $M_{mn}(\mathbb{F})$ is a vector space over $\mathbb{F}$.

Notation: Write

$$M_n(\mathbb{F}) = M_{nn}(\mathbb{F}).$$

Let S be a set and let V be a vector space over $\mathbb{F}$. Denote

$$V^S = \{\text{functions } f : S \longrightarrow V\}.$$

We define addition and scalar multiplication by using "point-wise operations," as follows. If $f, g \in V^S$, $f + g$ is the function $S \longrightarrow V$ defined by setting

$$(f+g)(s) = f(s) + g(s) \qquad \text{for all } s \in S.$$

Similarly, if α is a scalar and $f \in V^S$ define $\alpha f : S \longrightarrow V$ by setting

$$(\alpha f)(s) = \alpha(f(s)) \qquad \text{for all } s \in S.$$

Let $f, g \in V^S$. If $s \in S$, we have

$$\begin{array}{rcl} (f+g)(s) & = & f(s) + g(s) \\ & = & g(s) + f(s) \\ & = & (g+f)(s) \\ \therefore f + g & = & g + f. \end{array}$$

As for the distributive property, let $f, g \in V^S, \alpha \in \mathbb{F}$. Then, for all $s \in S$,

$$\begin{array}{rcl} \alpha(f+g)(s) & = & \alpha \cdot ((f+g)(s)) \\ & = & \alpha \cdot (f(s) + g(s)) \\ & = & \alpha \cdot f(s) + \alpha \cdot g(s) \\ & = & (\alpha f)(s) + (\alpha g)(s) \\ & = & (\alpha f + \alpha g)(s), \end{array}$$

and so $\alpha(f+g) = \alpha f + \alpha g$. Note that in V^S, the 0-vector is the function

$$\theta : S \longrightarrow V$$

where $\theta(s) = 0$ for all $s \in S$. To check this, we must show $f + \theta = f$ for all $f \in V^S$.

Let $s \in S$,

$$\begin{aligned}(f+\theta)(s) &= f(s)+\theta(s)\\ &= f(s)+0\\ &= f(s).\end{aligned}$$

Definition. Let V be a vector space over $\mathbb{F}$ and let W be a *nonempty* subset. If W is closed under addition and scalar multiplication, then W is called a *subspace* of V. More precisely, W is a subspace if:

1. $w_1, w_2 \in W \Rightarrow w_1 + w_2 \in W$;
2. $w_1 \in W$ and $\alpha \in \mathbb{F} \Rightarrow \alpha w_1 \in W$.

Note that if W is a subspace, then $0 \in W$. Indeed, If $w \in W$, then $0 = 0 \cdot w \in W$ by closure with respect to scalar multiplication.

1.1.3 Description of Homogeneous "Linear Problems"

Many problems (*e.g.*, solving equations, differential equations) are formulated in the context of a vector space. Indeed, we call a problem a *homogeneous linear problem* if the set of solutions is a subspace of some vector space.

Example The O.D.E.

$$\frac{d^2y}{dx^2} + y = 0.$$

It can be thought of as a problem where the background or "context" space is the vector space $\mathcal{C}^\infty(\mathbb{R})$ of infinitely differentiable functions $\mathbb{R} \longrightarrow \mathbb{R}$.

Note that by application of Calc. I, one can show that $\mathcal{C}^\infty(\mathbb{R})$ is closed under addition and multiplication, thus is a subspace of $\mathbb{R}^{\mathbb{R}}$.

The solution of $y'' + y = 0$ consists of all functions of the form

$$\{\alpha \sin x + \beta \cos x | \alpha, \beta \in \mathbb{R}\} \subseteq \mathcal{C}^\infty(\mathbb{R}),$$

which turns out to be a subspace of $\mathcal{C}^\infty(\mathbb{R})$. This motivates the following:

Definition. Let V be a vector space over $\mathbb{F}$, and let $v_1, v_2, \ldots, v_n$ be vectors in V. A *linear combination* of $v_1, v_2, \ldots, v_n$ is an expression of the form

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \sum_{i=1}^{n} \alpha_i v_i \in V,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$. More generally, if $S \subseteq V$ (S might be infinite!) a linear combination of elements of S is an exprssion of the form

$$\alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_m s_m,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{F}$ and $s_1, s_2, \ldots, s_m \in S$.
NOTE: All linear combinations involve only a finite number of terms!! We don't consider expressions of the form:

$$\sum_{i=1}^{\infty} \alpha_i s_i;$$

such expressions are meaningful only in the context of analysis.

1.1.4 The Span of a Set

DEFINITION. Let $S \subseteq V$ be a subset. We denote $\langle S \rangle = \{$all linear combinations of elements of $S\}$, and call this the *span* of the subset $S \subseteq V$. We set $\langle \phi \rangle = \{0\}$. If $S = \{v_1, v_2, \ldots, v_n\}$, then we often write $\langle v_1, v_2, \ldots, v_n \rangle$ in place of $\langle \{v_1, v_2, \ldots, v_n\} \rangle$.

Proposition 1.1.1. *If S is a subset of V, then $\langle S \rangle$ is a subspace of V.*

PROOF. If $S = \phi$, this is clear. Otherwise, note that a typical element of $\langle S \rangle$ is of the form $\sum \alpha_s s$, where $s \in S$ ranges over the elements of S (which might be infinite), and where the coefficients $\alpha_s \in \mathbb{F}$ are non-zero only for finitely many $s \in S$. Therefore, if $v = \sum \alpha_s s$, $w = \sum \beta_s s$, then $v + w = \sum (\alpha_s + \beta_s) s$. Since $\alpha_s + \beta_s \neq 0$ for only finitely many $s \in S$, it follows that $v + w \in \langle S \rangle$. Similarly, if $\alpha \in \mathbb{F}$, then $\alpha v = \sum \alpha\alpha_s s$; again, as $\alpha\alpha_s \neq 0$ for only finitely many $s \in S$, it follows also that $\alpha v \in \langle S \rangle$. Therefore $\langle S \rangle$ is a subspace of V. ■

Returning to the linear problem, $y'' + y = 0$. Note that the set of solutions is

$$\langle \{\sin x, \cos x\} \rangle$$

Therefore, $y'' + y = 0$ is indeed a linear problem.

Lemma 1.1.2. *If V is an $\mathbb{F}$-vector space and $S \subseteq V$, then the span $\langle S \rangle$ satisfies*

(i) $S \subseteq \langle S \rangle$

(ii) $\langle\langle S\rangle\rangle = \langle S\rangle$

PROOF. (i) is obvious. (ii) follows since for any subspace $W \subseteq V$, one has $\langle W\rangle = W$. ■

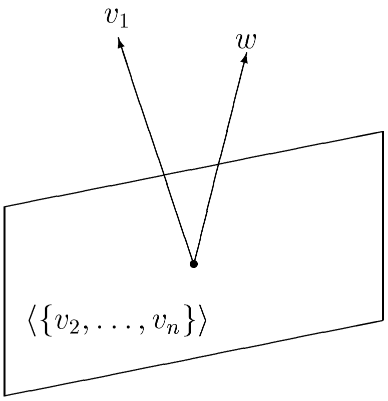
1.1.3 (Exchange Lemma). *Let V be a vector space over $\mathbb{F}$ and let $v_1, v_2, \ldots, v_n \in V$. If*

$$w \in \langle\{v_1, v_2, \ldots, v_n\}\rangle - \langle\{v_2, \ldots, v_n\}\rangle$$

then,

$$v_1 \in \langle\{w, v_2, \ldots, v_n\}\rangle - \langle\{v_2, \ldots, v_n\}\rangle.$$

PROOF. We picture the situation below:



We have $w = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ where $\alpha_1 \neq 0$. Therefore, $\alpha_1 v_1 = w - \sum\limits_{j=2}^{n} \alpha_j v_j$ which implies that

$$\begin{aligned} v_1 &= \frac{1}{\alpha_1}(w - \sum_{j=2}^{n} \alpha_j v_j) \\ &= \frac{1}{\alpha_1} w - \frac{1}{\alpha_1} \sum_{j=2}^{n} \alpha_j v_j \in \langle w, v_2, \ldots, v_n\rangle. \end{aligned}$$

Finally, if $v_1 \in \langle v_2, \ldots, v_n\rangle$, then

$$\langle v_2, \ldots, v_n\rangle = \langle v_1, v_2, \ldots, v_n\rangle$$

$$\langle v_1, v_2, \ldots, v_n\rangle - \langle v_2, \ldots, v_n\rangle = \emptyset.$$

However,

$$w \in \langle v_1, v_2, \ldots, v_n\rangle - \langle v_2, \ldots, v_n\rangle = \emptyset,$$

a contradiction.

■

We say that the vector space V is *finitely generated* if there exists a finite set of vectors $\{v_1, v_2, \ldots, v_k\}$ with

$$V = \langle v_1, v_2, \ldots, v_k\rangle.$$

1.2 Basis and Dimension

Definition. Let V be a vector space and let $S \subseteq V$ be a subset of vectors. We say that S is *linearly independent* if any finite set of vectors $\{s_1, s_2, \ldots, s_k\} \subseteq S$ satisfies

$$\sum_{i=1}^{k} \alpha_i s_i = 0,$$

then

$$\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0.$$

Otherwise, we call S *linearly dependent.*

If $S \subseteq V$ is a linearly independent set of vectors in V such that S also spans V (*i.e.*, $V = \langle S\rangle$), then we say that S is a *basis* of V.

Proposition 1.2.1. *Suppose that* $V = \langle v_1, v_2, \ldots, v_m\rangle = V$. *Then* $\{v_1, v_2, \ldots, v_m\}$ *contains a basis of* V.

Proof. We shall argue by induction on m. If $\{v_1, v_2, \ldots, v_m\}$ is linearly independent, then we're done. Otherwise, we may assume that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m = 0$$

for suitable $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{F}$. Without loss of generality, we may assume that $\alpha_1 \neq 0$ and so

$$v_1 = -\frac{1}{\alpha_1}\sum_{j=2}^{m} \alpha_j v_j \in \langle v_2, \ldots, v_m\rangle,$$

which clearly implies that $V = \langle v_2, v_3, \ldots, v_n \rangle$. By the induction hypothesis, V has a finite basis contained in $\{v_2, v_3, \ldots, v_m\}$, and we're done.

■

The following is immediate.

Corollary 1.2.1.1. *Any finitely-generated vector space has a basis.*

REMARK: The above results are true whether or not V is finitely-generated. Indeed, if $V = \langle S \rangle$, let $\mathcal{A} \subseteq S$ be a maximal linearly independent subset (which exists by Zorn's Lemma). Then $\mathcal{A}$ is a basis.

1.2.2 (Invariance of Dimension). *Let V be a finitely generated vector space and let $\{v_1, \ldots, v_m\}$, $\{w_1, \ldots, w_n\}$ be bases of V. Then $m = n$.*

PROOF. Assume that $m < n$. We may write

$$v_1 = \alpha_1 w_1 + \alpha_2 w_2 + \cdots + \alpha_n w_n$$

because $\{w_1, w_2, \ldots, w_n\}$ spans V. Without loss of generality, assume $\alpha_1 \neq 0$. Then,

$$v_1 \in \langle w_1, w_2, \ldots, w_n \rangle - \langle w_2, \ldots, w_n \rangle.$$

For if $v_1 = \beta_2 w_2 + \cdots + \beta_n w_n$, then

$$\alpha_1 w_1 + \alpha_2 w_2 + \cdots + \alpha_n w_n = \beta_2 w_2 + \cdots + \beta_n w_n,$$

and so

$$w_1 = -\frac{1}{\alpha_1}((\beta_2 - \alpha_2) w_2 + \cdots + (\beta_n - \alpha_n) w_n).$$

This clearly contradicts the linear independence of $\{w_1, w_2, \ldots, w_n\}$. Therefore, we must have

$$w_1 \in \langle v_1, w_2, \ldots, w_n \rangle - \langle w_2, \ldots, w_n \rangle.$$

Notice that $\{v_1, w_2, \ldots, w_n\}$ is a basis for V. Indeed,

$$w_1, w_2, \ldots, w_n \in \langle v_1, w_2, \ldots, w_n \rangle$$

implies that

$$\langle v_1, w_2, \ldots, w_n \rangle \supseteq \langle w_1, w_2, \ldots, w_n \rangle = V.$$

Next, if

$$\beta_1 v_1 + \beta_2 w_2 + \cdots + \beta_n w_n = 0$$

and if $\beta_1 = 0$ then,

$$\beta_2 w_2 + \cdots + \beta_n w_n = 0$$

and so

$$\beta_2 = \beta_3 = \cdots = \beta_n = 0$$

as $\{w_1, w_2, \ldots, w_n\}$ is linearly independent. If $\beta_1 \neq 0$, then $v_1 \in \langle w_2, \ldots, w_n \rangle$, also a contradiction.

Since $\{v_1, w_2, \ldots, w_n\}$ is a basis, we can write

$$v_2 = \beta_1 v_1 + \beta_2 w_2 + \cdots + \beta_n w_n.$$

If $\beta_2 = \beta_3 = \cdots = \beta_n = 0$, then v_1, v_2 are linearly dependent, a contradiction. Therefore, some $\beta_i \neq 0$. Re-index if necessary so that $\beta_2 \neq 0$. This gives the following:

$$v_2 \in \langle v_1, w_2, \ldots, w_n \rangle - \langle v_1, w_3, \ldots, w_n \rangle;$$

For if $v_2 \in \langle v_1, w_3, \ldots, w_n \rangle$,

$$v_2 = \gamma_1 v_1 + \gamma_3 w_3 + \cdots + \gamma_n w_n,$$

so,

$$\gamma_1 v_1 + \gamma_3 w_3 + \cdots + \gamma_n w_n = v_2 = \beta_1 v_1 + \beta_2 w_2 + \cdots + \beta_n w_n.$$

Thus, $w_2 \in \langle v_1, w_3, \ldots, w_n \rangle$, contrary to the linear independence of $\{v_1, w_2, \ldots, w_n\}$. So, we apply the *Exchange Lemma* 1.1.3 and get

$$w_2 \in \langle v_1, v_2, w_3, \ldots, w_n \rangle - \langle v_1, w_3, \ldots, w_n \rangle.$$

As above, $\{v_1, v_2, w_3, \ldots, w_n\}$ is a basis for V. Continue in this way and eventually obtain a basis of the form

$$\{v_1, v_2, \ldots, v_m, w_{m+1}, \ldots, w_n\}.$$

But as

$$\{v_1, v_2, \ldots, v_m\}$$

is a basis, $w_{m+1}, \ldots, w_n \in V = \langle v_1, v_2, \ldots, v_m \rangle$ contrary to linear independence of

$$\{v_1, v_2, \ldots, v_m, w_{m+1}, \ldots, w_n\}.$$

Thus $m = n$.

■

Example. We note here that $\dim \mathbb{F}^n = n$, since it is clear that

$$\left\{ v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, v_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

is a basis of $\mathbb{F}^n$. (See also *Exercise* B.7.)

1.3 Operations on Subspaces

1.3.1 (Basis Extension Theorem). *Let V be a finite dimensional vector space and let $\{v_1, v_2, \ldots, v_k\}$ be a linearly independent subset of V. Then, there exist vectors*

$$v_{k+1}, \ldots, v_n \in V \qquad (n = dim V)$$

such that $\{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ is a basis of V.

Proof. Let $\{w_1, w_2, \ldots, w_n\}$ be a basis of V. Let S be a maximal subset

$$S \subseteq \{w_1, w_2, \ldots, w_n\}$$

such that

$$\{v_1, v_2, \ldots, v_k\} \cup S$$

is linearly independent. We re-name the elements of S so that

$$S = \{v_{k+1}, v_{k+2}, \ldots, v_m\}.$$

Therefore, $\{v_1, \ldots, v_m\}$ is linearly independent. If $w_j \notin \langle v_1, \ldots, v_m \rangle$, then $\{v_1, \ldots, v_m\} \cup \{w_j\}$ is linearly independent, violating the maximality of S. This implies that

$$w_1, \ldots, w_n \in \langle v_1, \ldots, v_m \rangle$$

and so

$$V = \langle w_1, \ldots, w_n \rangle \subseteq \langle v_1, \ldots, v_m \rangle.$$

It follows that

$$\langle v_1, \ldots, v_m \rangle = V$$

and hence, $\{v_1, \ldots, v_m\}$ is a basis. By *Invariance of Dimension* (1.2.2), $m = n$.

■

Definition. Let $W_1, W_2 \subseteq V$ be subspaces. We define subspaces of V as follows:

(i) *Intersection* of W_1 and W_2:

$$W_1 \cap W_2 = \{w \in V | w \in W_1 \text{ and } w \in W_2\}.$$

It should be clear that $W_1 \cap W_2$ is again a subspace of V.

(ii) *Sum* of W_1 and W_2:

$$W_1 + W_2 = \{w_1 + w_2 | w_1 \in W_1 \text{ and } w_2 \in W_2\}.$$

To see that $W_1 + W_2$ is a subspace of V, let

$$w_1 + w_2, w_1' + w_2' \in W_1 + W_2.$$

Then,

$$\begin{aligned} (w_1 + w_2) + (w_1' + w_2') &= (w_1 + w_1') + (w_2 + w_2') \\ &\in W_1 + W_2, \end{aligned}$$

as $w_1 + w_1' \in W_1$ and $w_2 + w_2' \in W_2$. Similarly, we have closure with respect to scalar mulitiplication.

Definition. Let $W_1, W_2 \subseteq V$ be subspaces such that $W_1 \cap W_2 = \{0\}$. In this case we write

$$W_1 + W_2 = W_1 \oplus W_2$$

and call $W_1 \oplus W_2$ the *direct sum* of W_1 and W_2.

Note: If $W_1 + W_2 = W_1 \oplus W_2$, then every element of $W_1 + W_2$ can be written *uniquely* as $w_1 + w_2$, $w_1 \in W_1$ and $w_2 \in W_2$. Indeed, If $w_1 + w_2 = w_1' + w_2'$ with

$$w_1, w_1' \in W_1 \text{ and } w_2, w_2' \in W_2,$$

then,

$$w_1 - w_1' = w_2' - w_2 \quad \in \quad W_1 \cap W_2 = \{0\},$$

$$\begin{aligned} w_1 - w_1' &= w_2' - w_2 = 0 \\ w_1 &= w_1' \\ w_2 &= w_2' \end{aligned}$$

Conversely, if every element of $W_1 + W_2$ can be written uniquely as $w_1 + w_2$, $w_1 \in W_1$, $w_2 \in W_2$, then it is easy to check that $W_1 + W_2 = W_1 \oplus W_2$.

More generally, if $W_1, W_2, \ldots, W_k \subseteq V$ are subspaces then setting

$$W_1 + W_2 + \ldots + W_k = \{\sum_{i=1}^{k} w_i : w_i \in W_i,\ i = 1, 2, \ldots, k\}$$

also gives a subspace of V. Finally, if $w_1 + w_2 + \cdots + w_k = 0,\ w_i \in W_i,\ i = 1, 2, \ldots, k$ implies that $w_1 = 0,\ i = 1, 2, \ldots, k$, then we say that the sum $W_1 + W_2 + \cdots + W_k$ is *direct* and write

$$W_1 + W_2 + \cdots + W_k = W_1 \oplus W_2 \oplus \cdots \oplus W_k.$$

Note that this is equivalent to insisting that each element of $W_1 + W_2 + \cdots + W_k$ can be *uniquely* expressed as $w_1 + w_2 + \cdots + w_k,\ w_i \in W_i,\ i = 1, 2, \ldots, k$.

Proposition 1.3.2 (Complete Splitting or Complementation). *Let $W \subseteq V$ be a subspace. Then W has a* complement *in V,* i.e., *there exists a subspace $W' \subseteq V$ with $V = W \oplus W'$.*

Proof. Let $\{v_1, \ldots, v_k\}$ be a basis of W and extend it to a basis

$$\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$$

of V. Set

$$W' = \langle v_{k+1}, \ldots, v_n \rangle \subseteq V.$$

Clearly,

$$V = W + W'.$$

We need only show $W \cap W' = \{0\}$. To this end, if $v \in W \cap W'$ then

$$v = \alpha_1 v_1 + \cdots + \alpha_k v_k \text{ for suitable scalars} \alpha_1, \ldots, \alpha_k \in \mathbb{F}.$$

By the same token,

$$v = \alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n \text{ for suitable scalars } \alpha_{k+1}, \ldots, \alpha_n \in \mathbb{F}.$$

Therefore,

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = \alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n,$$

forcing

$$\alpha_1 v_1 + \cdots + \alpha_k v_k - \alpha_{k+1} v_{k+1} - \cdots - \alpha_n v_n = 0.$$

Since $\{v_1, \ldots, v_n\}$ is linearly independent, it follows that $\alpha_i = 0, i = 1, \ldots, n$. Thus, $v = 0$, *i.e.*, $W \cap W' = \{0\}$.

■

1.4 Linear Transformations

Definition. Let V, W be vector spaces over $\mathbb{F}$ and let $T : V \longrightarrow W$ be a mapping. We say that T is a *linear transformation* if

(i) $T(v_1 + v_2) = T(v_1) + T(v_2)$, for all $v_1, v_2 \in V$, and

(ii) $T(\alpha v) = \alpha T(v)$, for all $\alpha \in \mathbb{F}$, and for all $v \in V$

Note that conditions (i) and (ii) above are equivalent to the single condition

$$T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T(v_1) + \alpha_2 T(v_2) \text{ for all } v_1, v_2 \in V \text{ and } \alpha_1, \alpha_2 \in \mathbb{F}.$$

More generally, we see that a linear transformation maps linear combinations to linear combinations:

$$T(\sum_{i=1}^{k} \alpha_i v_i) = \sum_{i=1}^{k} \alpha_i T(v_i).$$

Definition. Let $T : V \longrightarrow W$ be a linear transformation. Define the *kernel* of T:

$$\begin{aligned} \ker(T) &= \{v \in V \mid T(v) = 0\} \\ &= T^{-1}(0) \qquad (0 \in W) \end{aligned}$$

Proposition 1.4.1. *Let $T : V \longrightarrow W$ be a linear transformation. Then,*

(1) $\ker(T)$ *is a subspace of V;*

(2) *If $V_1 \subseteq V$ is a subspace, then $T(V_1) \subseteq W$ is also a subspace;*

(3) *If $W_1 \subseteq W$ is a subspace, then*

$$T^{-1}(W_1) = \{v \in V \mid T(v) \in W_1\} \subseteq V$$

is a subspace of V.

Proof. If $v_1, v_2 \in \ker T$ and if $\alpha_1, \alpha_2 \in \mathbb{F}$, then $T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T(v_1) + \alpha_2 T(v_2) = \alpha_1 \cdot 0 + \alpha_2 \cdot 0 = 0$ and so $\alpha_1 v_2 + \alpha_2 v_2 \in \ker T$, *i.e.*, $\ker T$ is a subspace of V, proving (1). For (2), let $v_1, v_1' \in V_1$, $\alpha_1, \alpha_1' \in \mathbb{F}$. Then $\alpha_1 T(v_1) + \alpha_1' T(v_1') = T(\alpha_1 v_1 + \alpha_1' v_1') \in T(V_1)$, since $V_1 \subseteq V$ is a subspace, proving (2). Finally, if $v_1, v_1' \in T^{-1}(W_1)$, and if $\alpha_1, \alpha_1' \in \mathbb{F}$, then $T(\alpha_1 v_1 + \alpha_1' v_1') = \alpha_1 T(v_1) + \alpha_1' T(v_1') \in W_1$, since $W_1 \subseteq W$ is a subspace.

■

Definition. Let $T : V \longrightarrow W$ be a linear transformation, where V and W are both finite dimensional. Define the *nullity* of T to be the dimension of the kernel:

$$\text{nullity}(T) = \dim(\ker(T)),$$

and define the *rank* of T to be the dimension of the image:

$$\text{rank}(T) = \dim(T(V)).$$

We say that T is *injective* if T is a one-to-one function. We say that T is *surjective* if T is onto (*i.e.*, $T(V) = W$). We say T is an *isomorphism* if T is both injective and surjective. In this case, we write

$$V \cong W,$$

or

$$T : V \overset{\cong}{\longrightarrow} W.$$

Lemma 1.4.2. *Let $T : V \longrightarrow W$ be a linear transformation.*

(i) *T is injective if and only if* $\ker(T) = \{0\}$

(ii) *If T is an isomorphism, then $T^{-1} : W \longrightarrow V$ is defined and is also a linear transformation (in which case we call T* invertible*).*

Proof. For (i), T is injective and if $v \in \ker(T)$ then $T(v) = 0 = T(0)$, and so $v = 0$, *i.e.*, $\ker(T) = \{0\}$. Conversely, assume $\ker(T) = \{0\}$ and that $v_1, v_2 \in V$ with $T(v_1) = T(v_2)$. Then,

$$\begin{aligned} T(v_1 - v_2) &= T(v_1) - T(v_2) \\ &= 0 \ \text{ and so} \\ v_1 - v_2 &\in \ker(T) = \{0\} \\ v_1 - v_2 &= 0 \quad \textit{i.e.,} \\ v_1 &= v_2. \end{aligned}$$

For (ii) we need to show that

$$T^{-1}(w_1 + w_2) = T^{-1}(w_1) + T^{-1}(w_2)$$

$$T^{-1}(\alpha w) = \alpha T^{-1}(w),$$

for all $w, w_1, w_2 \in W$ and $\alpha \in \mathbb{F}$. But,

$$\begin{array}{rcl} T(T^{-1}(w_1 + w_2)) & = & w_1 + w_2; \\ & = & T(T^{-1}(w_1) + T^{-1}(w_2)); \end{array}$$

since T is injective, we infer that $T^{-1}(w_1 + w_2) = T^{-1}(w_1) + T^{-1}(w_2)$. Similarly,

$$\begin{array}{rcl} T(T^{-1}(\alpha w)) & = & \alpha w \\ & = & \alpha T(T^{-1}(w)) \\ & = & T(\alpha T^{-1}(w)); \end{array}$$

again, as T is injective, we get $T^{-1}(\alpha w) = \alpha T^{-1}(w)$.

■

Lemma 1.4.3. *If $T : V_1 \overset{\cong}{\to} V_2$ and V_1 is finite dimensional, then V_2 is finite dimensional and $\dim V_1 = \dim V_2$.*

Proof. Let $\dim V_1 = n$ and let $\{v_1, v_2, \ldots, v_n\}$ be a basis of V_1. Then as T is surjective it is clear that $\{T(v_1), T(v_2), \ldots, T(v_n)\}$ spans V_2. Thus, it suffices to show that $\{T(v_1), T(v_2), \ldots, T(v_n)\}$ is linearly independent. If $\sum_{i=1}^{n} \alpha_i T(v_i) = 0$, then since T is linear, we have $T(\sum_{i=1}^{n} \alpha_i v_i) = 0$. Since T is injective, we infer that $\sum_{i=1}^{n} \alpha_i v_i = 0$; since $\{v_1, v_2, \ldots, v_n\}$ is linearly independent, infer that $\alpha_i = 0$, $i = 1, 2, \ldots, n$.

■

1.4.4 (Rank-Nullity Theorem). *Let $T : V \longrightarrow W$ be a linear transformation. Then,*

$$\text{rank}(T) = \dim(V) - \text{nullity}(V).$$

Proof. We may use *Proposition* 1.3.2 to find a subspace $V_1 \subseteq V$ such that

$$V = \ker(T) \oplus V_1.$$

We have the "restriction" of T to V_1:

$$T|_{V_1} : V_1 \longrightarrow T(V_1).$$

We claim that the above is an isomorphism. If $v_1 \in \ker(T|_{V_1})$ then $T(v_1) = 0$ implies that $v_1 \in V_1 \cap \ker T = \{0\}$ and so $v_1 = 0$. That is, $\ker(T|_{V_1}) = \{0\}$, forcing $T|_{V_1}$ to be injective.

Let $T(v) \in T(V)$ for some $v \in V$. Then $v = x + v_1$, for suitable $x \in \ker T$, $v_1 \in V_1$. Then,

$$\begin{aligned} T(v) &= T(k + v_1) \\ &= T(k) + T(v_1) \\ &= 0 + T(v_1) = T(v_1), \end{aligned}$$

and so $T|_{V_1}$ is surjective, proving that $V_1 \cong T(V)$. Next, note that

$$\dim V_1 = \dim V - \dim(\ker T)$$

Indeed, if $\{x_1, x_2, \ldots, x_r\}$ is a basis of $\ker T$ and if $\{v_1, \ldots, v_m\}$ is a basis of V_1, then it is clear from the splitting $V = \ker(T) \oplus V_1$ that

$$\{x_1, \ldots, x_r, v_1, \ldots, v_m\}$$

is a basis of V. Thus,

$$\begin{aligned} \dim V &= r + m \\ &= \dim \ker\, T + \dim V_1 \end{aligned}$$

However, $V_1 \cong T(V)$ and so $\dim V_1 = \operatorname{rank} T$. Thus, $\dim V = \operatorname{nullity} T + \operatorname{rank} T$, and we are done.

■

1.4.5 (Extension by Linearity Theorem). *Let V be a vector space with basis $\{v_1, \ldots, v_n\}$ (so $\dim V = n$). Let W be a vector space and let there be given vectors $w_1, w_2, \ldots, w_n \in W$. There exists a unique linear transformation which satisfies $T(v_i) = w_i, i = 1, 2, \ldots, n$.*

Proof. Since $\{v_1, \ldots, v_n\}$ forms a basis, every vector $v \in V$ can be written uniquely as

$$v = \sum_{i=1}^{n} \alpha_i v_i.$$

Thus, we have a well-defined mapping $T : V \longrightarrow W$ given by

$$T(\sum_{i=1}^{n} \alpha_i v_i) = \sum_{i=1}^{n} \alpha_i w_i.$$

To show that T is linear, let $u, v \in V$. Then we can write

$$u = \sum_{i=1}^{n} \alpha_i v_i, \quad v = \sum_{i=1}^{n} \beta_i v_i.$$

Thus,

$$u + v = \sum_{i=1}^{n} (\alpha_i + \beta_i) v_i$$

so

$$\begin{aligned} T(u+v) &= T(\sum_{i=1}^{n} (\alpha_i + \beta_i) v_i) \\ &= \sum_{i=1}^{n} (\alpha_i + \beta_i) w_i \\ &= \sum_{i=1}^{n} \alpha_i w_i + \sum_{i=1}^{n} \beta_i w_i \\ &= T(u) + T(v). \end{aligned}$$

Similarly, $T(\alpha v) = \alpha T(v)$. Thus, T is linear.

Next, note that

$$T(v_i) = T(1 \cdot v_i) = 1 \cdot w_i = w_i, \qquad i = 1, \ldots, n.$$

Finally, we prove that T is unique. That is, if $S : V \longrightarrow W$ is another linear transformation satisfying

$$S(v_1) \;=\; w_1, \qquad i = 1, 2, \ldots, n$$

then we wish to show that $T(v) = S(v)$ for all $v \in V$. We write

$$v = \sum_{i=1}^{n} \alpha_i v_i$$

and so

$$\begin{aligned} T(v) &= T(\sum_{i=1}^{n} \alpha_i v_i) \\ &= \sum_{i=1}^{n} \alpha_i T(v_i) = \sum_{i=1}^{n} \alpha_i w_i \\ &= \sum_{i=1}^{n} \alpha_i S(v_i) \\ &= S(v). \end{aligned}$$

■

Proposition 1.4.6. *Let* $\dim V = \dim W$. *Then* $V \cong W$.

Proof. Let $\{v_1, \ldots, v_n\}$ be a basis of V and $\{w_1, \ldots, w_n\}$ be a basis of W. By the above result, there exists a linear transformation $T : V \longrightarrow W : T(v_i) = w_i \; i = 1, \ldots, n$. T is certainly surjective as

$$\begin{aligned} T(V) &= T(\langle v_1, \ldots, v_n \rangle) \\ &= \langle T(v_1), \ldots, T(v_n) \rangle \\ &= \langle w_1, \ldots, w_n \rangle \\ &= W. \end{aligned}$$

By the *Rank-Nullity Theorem* 1.4.4 we have

$$\operatorname{rank} T = \dim V - \operatorname{nullity}(T),$$

so

$$n = n - \operatorname{nullity}(T),$$

which implies that

$$\operatorname{nullity}(T) \;=\; 0$$

i.e., $\ker T = \{0\}$, and so T is injective by 1.4.2 (i).

■

Definition. Let $(v_1, v_2, \ldots, v_n)$ be a sequence of vectors in the vector space V. If $\{v_1, v_2, \ldots, v_n\}$ is a basis of V, we say that $(v_1, v_2, \ldots, v_n)$ is an *ordered basis* of V.

1.5 Matrices and the Representation Picture

Let V be a vector space and let $\mathcal{A} = (v_1, v_2, \ldots, v_n)$ be an ordered basis. Define the map

$$(\cdot)_{\mathcal{A}} : V \longrightarrow \mathbb{F}^n$$

as follows: if $v \in V$ and v is written (uniquely) as

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n,$$

set

$$\begin{aligned} (v)_{\mathcal{A}} &= v_{\mathcal{A}} \\ &= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{F}^n \end{aligned}$$

Proposition 1.5.1. *The map $(\cdot)_{\mathcal{A}} : V \longrightarrow \mathbb{F}^n$ is a linear isomorphism.*

Proof:

We shall prove that this is a linear transformation, i.e.,

(i) $(v + w)_{\mathcal{A}} = v_{\mathcal{A}} + w_{\mathcal{A}}$

(ii) $(\alpha v)_{\mathcal{A}} = \alpha v_{\mathcal{A}}$

for all $v, w \in V, \alpha \in \mathbb{F}$. To this end, let v, w be expressed as linear combinations of the basis vectors

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n,$$

$$w = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n.$$

Therefore,

$$\begin{aligned} v + w &= (\alpha_1 + \beta_1)v_1 + (\alpha_2 + \beta_2)v_2 + \cdots + (\alpha_n + \beta_n)v_n, \\ \alpha v &= (\alpha\alpha_1)v_1 + (\alpha\alpha_2)v_2 + \cdots + (\alpha\alpha_n)v_n, \end{aligned}$$

and so

$$(v + w)_{\mathcal{A}} = \begin{bmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \\ \vdots \\ \alpha_n + \beta_n \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = v_{\mathcal{A}} + w_{\mathcal{A}},$$

and

$$(\alpha v)_{\mathcal{A}} = \begin{bmatrix} \alpha\alpha_1 \\ \alpha\alpha_2 \\ \vdots \\ \alpha\alpha_n \end{bmatrix} = \alpha \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \alpha v_{\mathcal{A}}.$$

Finally, we show that $(\cdot)_{\mathcal{A}}$ is an isomorphism:

If $v \in \ker(\cdot)_{\mathcal{A}}$, then

$$v_{\mathcal{A}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

which implies that $v = 0 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_n = 0$. From this it follows that the nullity of the linear transformation $(\cdot)_{\mathcal{A}}$ is 0; by the *Rank-Nullity Theorem* 1.4.4, we get $\mathrm{rank}(\cdot)_{\mathcal{A}} = \mathrm{dim}V - 0 = n$ and so $(\cdot)_{\mathcal{A}}$ is surjective.

■

Definition. Let $T : V \longrightarrow W$ be a linear transformation from V to W, where V and W are both finite dimensional vector spaces over $\mathbb{F}$. Let $\mathcal{A} = (v_1, v_2, \ldots, v_n)$ be an ordered basis for V, and let $\mathcal{B} = (w_1, w_2, \ldots, w_m)$ be an ordered basis for W. We define the *matrix representation* of T *relative to the ordered bases* $\mathcal{A}$ and $\mathcal{B}$ by setting $(T)_{\mathcal{BA}} = T_{\mathcal{BA}} = [\alpha_{ij}] \in M_{mn}(\mathbb{F})$, where $T(v_j) = \sum_{i=1}^{m} \alpha_{ij} w_i$.

Example. We can think of the complex number field $\mathbb{C}$ as a two-dimensional vector space over $\mathbb{R}$. Let $\mathcal{A} = (1, i)$. Define

$$T : \mathbb{C} \longrightarrow \mathbb{C}$$

by

$$T(z) = (2 - 3i)z, \qquad z \in \mathbb{C}.$$

Then, the distributive and associative laws in $\mathbb{C}$ imply that $T(z_1 + z_2) = T(z_1) + T(z_2)$ and that $T(\alpha z) = \alpha T(z)$ for all $\alpha \in \mathbb{R}$.

We compute $T_{\mathcal{AA}}$

$$\begin{aligned} T(1) &= (2 - 3i)1 = 2 \cdot 1 - 3 \cdot i \\ T(i) &= (2 - 3i)i = 2 \cdot i + 3 \cdot 1. \end{aligned}$$

Therefore,

$$T_{\mathcal{A}\mathcal{A}} = \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}$$

1.5.2 (The Representation Picture). *Let $T : V \longrightarrow W$ be a linear transformation of finite dimensional vector spaces. Let $\mathcal{A}, \mathcal{B}$ be ordered bases for V, W respectively. Then, the following diagram commutes:*

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "(\cdot)_{\mathcal{A}}"] & W \arrow[d, "(\cdot)_{\mathcal{B}}"] \\ \mathbb{F}^n \arrow[r, "T_{\mathcal{A}\mathcal{B}}=A"] & \mathbb{F}^m \end{tikzcd}$$

where each $A \in M_{mn}(\mathbb{F})$ defines a linear transformation from $\mathbb{F}^n$ into $\mathbb{F}^m$ by

$$A\left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}\right) = A\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix},$$

the right-hand side being ordinary matrix product. To say the above diagram "commutes" is to say that for all $v \in V$, $(T(v))_{\mathcal{B}} = A(v_{\mathcal{A}})$.

Proof: Let $\mathcal{A} = (v_1, v_2, \ldots, v_n)$, $\mathcal{B} = (w_1, w_2, \ldots, w_m)$, and assume that

$$T_{\mathcal{A}\mathcal{B}} = A = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{bmatrix}$$

where

$$T(v_j) = \sum_{i=1}^{m} \alpha_{ij} w_i, \; j = 1, \ldots, n.$$

If $v = \sum\limits_{j=1}^{n} \alpha_j v_j$, we have

$$\begin{aligned}
(T(v))_{\mathcal{B}} &= (T(\sum_{j=1}^{n} \alpha_j v_j))_{\mathcal{B}} \\
&= (\sum_{j=1}^{n} \alpha_j T(v_j))_{\mathcal{B}} \\
&= (\sum_{j=1}^{n} \alpha_j (\sum_{i=1}^{m} \alpha_{ij} w_i))_{\mathcal{B}} \\
&= (\sum_{i=1}^{m} (\sum_{j=1}^{n} \alpha_j \alpha_{ij}) w_i)_{\mathcal{B}} \\
&= \begin{bmatrix} \sum\limits_{j=1}^{n} \alpha_j \alpha_{1j} \\ \sum\limits_{j=1}^{n} \alpha_j \alpha_{2j} \\ \vdots \\ \sum\limits_{j=1}^{n} \alpha_j \alpha_{mj} \end{bmatrix}. \\
&= \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \\
&= A v_{\mathcal{A}}.
\end{aligned}$$

■

Next we shall consider what effect changing the ordered bases has on the matrix representation of a linear transformation. Let $T : V \longrightarrow W$ with $\mathcal{A}, \mathcal{B}$ being ordered bases for V, W respectively. If $\mathcal{A}', \mathcal{B}'$ are new ordered bases for V, W respectively, we wish to determine the relationship between the $m \times n$ matrices

$$T_{\mathcal{B}\mathcal{A}} \text{ and } T_{\mathcal{B}'\mathcal{A}'}.$$

To this end, let $\mathcal{A} = (v_1, v_2, \ldots, v_n)$, $\mathcal{A}' = (v_1', v_2', \ldots, v_n')$, $\mathcal{B} = (w_1, w_2, \ldots, w_m)$, and $\mathcal{B}' = (w_1', w_2', \ldots, w_m')$. We have an $n \times n$ matrix:

$$P = [p_{ij}] \ (= C_{\mathcal{A}\mathcal{A}'})$$

defined by

$$v_j' = \sum_{i=1}^{n} p_{ij} v_i, \qquad j = 1, 2, \ldots, n.$$

Similarly, we have an $m \times m$ matrix

$$Q = [q_{kl}] \ (= C_{\mathcal{B}'\mathcal{B}})$$

defined by

$$w_l = \sum_{k=1}^{m} q_{kl} w'_k, \qquad l = 1, 2, \ldots, m.$$

We note that both P and Q are *invertible* matrices. To see that, for example, P is invertible, we write

$$v_l = \sum_{k=1}^{n} p'_{kl} v'_k, \qquad l = 1, 2, \ldots, n,$$

and set $P' = [p'_{kl}]$. We shall show that $PP' = I = P'P$.

Watch this:

$$\begin{aligned} v'_j &= \sum_{i=1}^{n} p_{ij} v_i \\ &= \sum_{i=1}^{n} p_{ij} \sum_{k=1}^{n} p'_{ki} v'_k \\ &= \sum_{k=1}^{n} \left(\sum_{i=1}^{n} p'_{ki} p_{ij} \right) v'_k. \end{aligned}$$

Since $(v'_1, \ldots, v'_n)$ is an ordered basis, we must have

$$\sum_{i=1}^{n} p'_{ki} p_{ij} = \begin{cases} 1 & k = j \\ 0 & k \neq j. \end{cases}$$

In other words,

$$\sum_{i=1}^{n} p'_{ki} p_{ij} = \delta_{kj}. \qquad \text{(Kronecker } \delta)$$

Note that $[\delta_{ij}] = I$, the identity matrix. So, we have shown $P'P = I$. Equivalently, in terms of the above notation, we have shown that $C_{\mathcal{A}'\mathcal{A}} C_{\mathcal{A}\mathcal{A}'} = I$, i.e., that $C_{\mathcal{A}\mathcal{A}'} = C^{-1}_{\mathcal{A}'\mathcal{A}}$. Similarly, $C_{\mathcal{B}\mathcal{B}'} = C^{-1}_{\mathcal{B}'\mathcal{B}}$.

Conversely, assume that $\mathcal{A} = (v_1, v_2, \ldots, v_n)$ is an ordered basis of V, and that $P = [p_{ij}]$ is an invertible matrix. If we define the vectors $v'_1, v'_2, \ldots, v'_n$ by the equations

$$v'_j = \sum_{i=1}^{n} p_{ij} v_i, \ j = 1, 2, \ldots, n,$$

then we claim that $\mathcal{A}' = (v_1', v_2', \ldots, v_n')$ is an ordered basis of V. Indeed if $P^{-1} = [q_{kl}]$, then

$$\begin{aligned} v_i &= \sum_{k=1}^{n} \delta_{ki} v_k \\ &= \sum_{k=1}^{n} \sum_{l=1}^{n} p_{kl} q_{li} v_k \\ &= \sum_{l=1}^{n} q_{li} \sum_{k=1} p_{kl} v_k \\ &= \sum_{l=1}^{n} q_{li} v_l', \end{aligned}$$

and so each vector v_i is a linear combination of the vectors in $\mathcal{A}'$. From this it is clear that $\mathcal{A}'$ spans V. To prove that it is a basis, note that by *Proposition* 1.2.1, the set $\{v_1', v_2', \ldots, v_n'\}$ must contain a basis of V. Since $\dim V = n$ we infer that $\{v_1', v_2', \ldots, v_n'\}$ must already be a basis.

Proposition 1.5.3 (Change of Basis). *Let $T : V \longrightarrow W$ and let $\mathcal{A}, \mathcal{A}', \mathcal{B}, \mathcal{B}', P$ and Q be as above. Then,*

$$T_{\mathcal{B}'\mathcal{A}'} = C_{\mathcal{B}'\mathcal{B}} T_{\mathcal{B}\mathcal{A}} C_{\mathcal{A}\mathcal{A}'},$$

i.e.,

$$T_{\mathcal{B}'\mathcal{A}'} = Q T_{\mathcal{B}\mathcal{A}} P.$$

Proof. We know that

$$T_{\mathcal{B}'\mathcal{A}'} = [\beta_{ij}]$$

where $T(v_j') = \sum_{i=1}^{m} \beta_{ij} w_i'$. Thus, if $C_{\mathcal{A}\mathcal{A}'} = [p_{ij}]$, $T_{\mathcal{B}\mathcal{A}} = [\alpha_{ij}]$, and $C_{\mathcal{B}'\mathcal{B}} = [q_{ij}]$, we must show that

$$\begin{aligned} \beta_{ij} &= (i,j)^{\underline{\text{th}}} \text{ entry of } C_{\mathcal{B}'\mathcal{B}} T_{\mathcal{B}\mathcal{A}} C_{\mathcal{A}\mathcal{A}'} \\ &= \sum_{l=1}^{m} \sum_{k=1}^{n} q_{il} \alpha_{lk} p_{kj}. \end{aligned}$$

Now,

$$\begin{aligned} T(v_j') &= T(\sum_{k=1}^{n} p_{kj} v_k) \\ &= \sum_{k=1}^{n} p_{kj} T(v_k) \\ &= \sum_{k=1}^{n} p_{kj} \sum_{l=1}^{m} \alpha_{lk} w_l \\ &= \sum_{k=1}^{n} p_{kj} \sum_{l=1}^{m} \alpha_{lk} \sum_{i=1}^{m} q_{il} w_i' \\ &= \sum_{i=1}^{n} (\sum_{l=1}^{m} \sum_{k=1}^{n} q_{il} \alpha_{lk} p_{kj}) w_i'. \end{aligned}$$

■

Conversely, assume that $T : V \longrightarrow V$ is a given linear transformation, that $\mathcal{A} = (v_1, ..., v_n)$ is an ordered basis and that there exists an invertible matrix $B = [\beta_{ij}]$, such that

$$P^{-1} T_{\mathcal{A}} P = B,$$

where $P = [p_{ij}]$. If we define $\mathcal{A}' = (v_1', v_2', ..., v_n')$ via the equations $v_j' = \sum\limits_{i=1}^{n} p_{ij} v_i$, $j = 1, 2, \ldots, n$, then by the above discussion $\mathcal{A}'$ is an ordered basis and $T_{\mathcal{A}'} = B$.

Definition. Let $A, B \in M_n(\mathbb{F})$. We say that A, B are *similar* matrices (and write $A \sim B$) if there exists an invertible matrix P such that

$$B = P^{-1} A P.$$

As a result of *Proposition* 1.5.3 and the ensuing paragraph, we see that two matrices A, B are similar if and only if they represent the same linear transformation, but possibly relative to different ordered bases.

One of the basic problems of matrix theory is to find, for a given matrix A, an invertible matrix P such that $P^{-1}AP$ is “simple”.

For example, if we can find P so that $P^{-1}AP$ is a diagonal matrix, i.e.,

$$P^{-1}AP = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{bmatrix}$$

we say that A is *diagonalizable.* Likewise, we say that a linear transformation $T : V \to V$ is *diagonalizable* if and only if V has an ordered basis $\mathcal{A}$ such that $T_{\mathcal{A}}$ is a diagonal matrix. We will take up this important topic in more detail later.

From the above, we see that if $T : V \longrightarrow V$ is a linear transformation and $\mathcal{A}, \mathcal{A}'$ are ordered bases of V, then the matrices $T_{\mathcal{A}}$ and $T_{\mathcal{A}'}$ are similar. This says that similar matrices are just representations of the *same* linear transformation relative to different ordered bases. Thus, one of the basic problems concerning the given linear transformation $T : V \longrightarrow V$ is to find an ordered basis $\mathcal{A}'$ such that $T_{\mathcal{A}'}$ is "simple".

1.6 Quotient Spaces

Let V be a vector space and let $W \subseteq V$ be a subspace. We wish to construct a new vector space V/W having the property that if V is finite dimensional, then

$$\dim V/W = \dim V - \dim W.$$

For a vector $v \in V$, set

$$v + W = \{v + w | w \in W\} \subseteq V.$$

We call this the *coset* determined by v.

Lemma 1.6.1. *Let $v_1, v_2 \in V$. Then $v_1 + W = v_2 + W$ if and only if $v_1 - v_2 \in W$.*

Proof. If $v_1 + W = v_2 + W$, then there exist $w_1, w_2 \in W$ with $v_1 + w_1 = v_2 + w_2$. But then $v_1 - v_2 = w_2 - w_1 \in W$. Conversely, if $v_1 - v_2 = w \in W$, then for any $w' \in W$, we have $v_1 + w' = v_2 + w + w' \in v_2 + W$, and so $v_1 + W \subseteq v_2 + W$. Similarly, we have $v_2 + W \subseteq v_1 + W$, proving the lemma. ■

We wish to give V/W the structure of a vector space as follows. If $v_1 + W, v_2 + W, v + W \in V/W, \alpha \in \mathbb{F}$, we set

$$(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W,$$

$$\alpha(v + W) = \alpha v + W.$$

We must show that these operations are well-defined, *i.e.*, if $v + W = v' + W, v_1 + W = v_1' + W, v_2 + W = v_2' + W$ then

$$(v_1 + W) + (v_2 + W) = (v_1' + W) + (v_2' + W)$$

and

$$\alpha(v + W) = \alpha(v' + W).$$

Indeed, we have, by *Lemma* 1.6.1, that $v_1 - v_1', v_2 - v_2' \in W$, and so $(v_1 + v_2) - (v_1' + v_2') = v_1 - v_1' + v_2 - v_2' \in W$, forcing

$$(v_1 + v_2) + W = (v_1' + v_2') + W$$

(again by *Lemma* 1.6.1), and so $(v_1 + W) + (v_2 + W) = (v_1' + W) + (v_2' + W)$, as required. Similarly, one can show that $\alpha(v + W) = \alpha(v' + W)$.

All the usual (and necessary) properties hold (e.g. associativity and commutativity of addition, the associativity of scalar multiplication, the distributive laws and the unital property). Therefore V/W is an $\mathbb{F}$-vector space, as claimed, called the *quotient space* of V by W.

Rationale: Let $T : V \longrightarrow W$ be a linear transformation and set

$$K = \ker T.$$

We show that if $w \in T(V) \subseteq W$, then

$$\begin{aligned} T^{-1}(w) &= \{v \in V | T(v) = w\} \\ &= v_0 + K \end{aligned}$$

where v_0 is any *fixed* vector with $T(v_0) = w$. That is to say, the inverse image under a linear transformation of any vector is a coset relative to the kernel of this linear transformation. Indeed, if $v \in v_0 + K$, then

$$v = v_0 + k \qquad (k \in K)$$

so,

$$\begin{aligned} T(v) &= T(v_0 + k) \\ &= T(v_0) + T(k) \\ &= T(v_0) + 0 \\ &= w, \end{aligned}$$

i.e.,

$$v_0 + K \subseteq T^{-1}(w).$$

Conversely, if $v \in T^{-1}(w)$, then $T(v) = w$ and so $v - v_0 \in K$, which says that $v \in v_0 + K$, and so

$$T^{-1}(w) \subseteq v_0 + K.$$

The result therefore follows.

A basic problem in linear algebra is to solve

$$T(x) = w$$

given a linear transformation $T : V \to W$ and a vector $w \in W$. If $w = 0$, then the solution set is the kernel of T, which we have already seen to be a subspace of V, in which case we call this a *homogeneous linear problem.* If $w \neq 0$, we call the problem $T(x) = w$ an *inhomogeneous* linear problem. If $v_0 \in V$ is a particular solution of $T(x) = w$, then the complete set of solutions is $v_0 + K$. This description should be familiar to students having studied non-homogeneous linear ordinary differential equations.

Proposition 1.6.2. *Suppose that*

$$T : V \longrightarrow W$$

is a linear transformation and that $V_1 \subseteq \ker T$. *If we define*

$$\overline{T} : V/V_1 \longrightarrow W, \text{ by } \overline{T}(V_1 + v) = T(v), \ v \in V$$

then $\overline{T}$ *is a well-defined linear transformation.*

Proof: For the well-definedness, we must show that if

$$v + V_1 = v' + V_1$$

then

$$\overline{T}(v + V_1) = \overline{T}(v' + V_1).$$

But, $v + V_1 = v' + V_1$ implies that $v - v' = v_1 \in V_1$. Thus,

$$\begin{aligned} T(v) - T(v') &= T(v - v') \\ &= T(v_1) \\ &= 0, \end{aligned}$$

i.e.,

$$T(v) = T(v'),$$

and so

$$\begin{aligned} \overline{T}(v + V_1) &= T(v) \\ &= T(v') \\ &= \overline{T}(v' + V_1), \end{aligned}$$

proving that $\overline{T}$ is well-defined. Finally, we show that

$$\overline{T} : V/V_1 \longrightarrow W$$

is linear. To this end, we have

$$\begin{aligned} \overline{T}((v + V_1) + (v' + V_1)) &= \overline{T}((v + v') + V_1) \\ &= T(v + v') \\ &= T(v) + T(v') \\ &= \overline{T}(v + V_1) + \overline{T}(v' + V_1). \end{aligned}$$

Similarly,

$$\begin{aligned} \overline{T}(\alpha(v + V_1)) &= \overline{T}(\alpha v + V_1) \\ &= T(\alpha v) \\ &= \alpha T(v) \\ &= \alpha T(v + V_1). \end{aligned}$$

■

Whenever $V \subseteq W$ is a subspace, we can define the *canonical projection*

$$\pi_W : V \longrightarrow V/W$$

by setting

$$\pi_W(v) = v + W, \qquad v \in V.$$

Proposition 1.6.3 (Canonical Projection). *With the notation as above,*

(i) $\pi_W : V \longrightarrow V/W$ *is a linear transformation,*

(ii) $\ker \pi_W = W$, *and*

(iii) π_W *is surjective.*

PROOF: We have

$$\begin{aligned}\pi_W(v_1 + v_2) &= (v_1 + v_2) + W \\ &= (v_1 + W) + (v_2 + W) \\ &= \pi_W(v_1) + \pi_W(v_2)\end{aligned}$$

Likewise,

$$\begin{aligned}\pi_W(\alpha v) &= (\alpha v) + W \\ &= \alpha(v + W) \\ &= \alpha\pi_W(v)\end{aligned}$$

So, π_W is a linear transformation, proving (i).

Next, note that the 0-vector in V/W is $0 + W$. Thus,

$$\ker \pi_W = \{v \in V | v + W = 0 + W\}$$

But,

$$v + W = W \Leftrightarrow v \in W,$$

forcing

$$\ker \pi_W = W.$$

This proves (ii).

Finally, if $v + W \in V/W$, the $\pi_W(v) = v + W$, giving (iii).

■

Corollary 1.6.3.1. *If* $dim\,\mathrm{V} < \infty$ *then* $dim\,(\mathrm{V/W}) = dim\,\mathrm{V} - dim\,\mathrm{W}$.

PROOF: By the *Rank-Nullity Theorem* (1.4.4), we have

$$\begin{aligned}\dim(V/W) &= \dim \pi_W(V) \\ &= \mathrm{rank}\pi_W \\ &= \dim V - \text{nullity } \pi_W \\ &= \dim V - \dim W.\end{aligned}$$

■

Corollary 1.6.3.2. *If V is a vector space and $W \subseteq V$ is a subspace then W is the kernel of some linear transformation.*

Proof: $\ker(\pi_W) = W$.

■

1.6.4 (Fundamental Homomorphism Theorem). *Let $T : V \longrightarrow V'$ be a surjective linear transformation. Then*

$$V' \cong V/W$$

where $W = \ker \mathrm{T}$. More precisely, we have a commutative diagram:

$$\begin{tikzcd}
V \arrow[rr, "T"] \arrow[dr, "\pi_{\ker T}"'] & & V' \\
& V/\ker T \arrow[ur, "\overline{T}"', "\cong"] &
\end{tikzcd}$$

Proof: We already have the well-defined linear transformation

$$\overline{T} : V/W \longrightarrow V'$$

satisfying $\overline{T}(v + W) = T(v)$. If $v' \in V'$, then because T is surjective, there exists $v \in V$ with $T(v) = v'$. Therefore, $\overline{T}(v + W) = T(v) = v'$ and so $\overline{T}$ is surjective. Finally, we know

$$\begin{aligned}
\ker \overline{T} &= \{v + W \in V/W \mid \overline{T}(v + W) = 0 \in V'\} \\
&= \{v + W | T(v) = 0 \in V'\} \\
&= \{v + W | v \in \ker T = W\} \\
&= \{W\},
\end{aligned}$$

and so

$$\ker \overline{T} = \{W\};$$

Since W is the 0-vector in V/W, we see that $\overline{T}$ is injective.

■

Proposition 1.6.5 (Correspondence). *Let $T : V \longrightarrow V'$ be a surjective linear transformation. Then there is a one-to-one correspondence between the subspaces of V' and the subspaces of V which contain $W = \ker \mathrm{T}$. In fact, the correspondence is given by*

$$T^{-1} : \{\textit{subspaces of } V'\} \longrightarrow \{\textit{subspaces of } V \textit{ containing } W\}.$$

PROOF. Let $V_1', V_2' \subset V'$ be subspaces and assume that $T^{-1}(V_1') = T^{-1}(V_2')$. Then, as T is surjective, we have $V_1' = T(T^{-1}(V_1')) = T(T^{-1}(V_2')) = V_2'$, and so T^{-1} is injective. If $V_1 \subseteq V$ is a subspace containing W, then it is clear that if $V_1' = T(V_1)$, then $T^{-1}(V_1') \supseteq V_1$. But then V_1 and $T^{-1}(V_1')$ both contain W and so V_1/W and $T^{-1}(V_1')/W$ are subspaces of V/W that both map to $V_1' \subseteq V'$ under $\overline{T}$. As $\overline{T}$ is an isomorphism, we conclude that $V_1/W = T^{-1}(V_1')/W$. Thus if $v \in T^{-1}(V_1')$, we have that $v + W \in V_1/W$ and so $v \in V_1$, forcing $V_1 = T^{-1}(V_1')$. The result follows.

■

Corollary 1.6.5.1. *The subspaces of a quotient space are those of the form V_0/W, where V_0 is a subspace of V which contains W.*

PROOF: If $\overline{V_0} \subseteq V/W$ is a subspace, then by (1.6.5),

$$\overline{V_0} = \pi_W \pi_W^{-1}(\overline{V_0}).$$

If we set

$$V_0 = \pi_W^{-1}(\overline{V_0}),$$

then V_0 is a subspace of V containing W, and $\pi_W(V_0) = V_0/W$.

■

1.7 Dual Spaces

Let V, V' be vector spaces. Recall that the set

$$V'^V = \{\text{mappings } V \longrightarrow V'\}$$

is a vector space relative to point-wise operations. Set

$$L(V, V') = \{T \in V'^V : T \text{ is linear}\}.$$

Proposition 1.7.1. *$L(V, V')$ is a subspace of V'^V.*

PROOF: Let $T_1, T_2 \in L(V, V'), \alpha \in \mathbb{F}, v_1, v_2 \in V$. Then

$$\begin{aligned}
(T_1 + T_2)(v_1 + v_2) &= T_1(v_1 + v_2) + T_2(v_1 + v_2) \\
&= T_1(v_1) + T_1(v_2) + T_2(v_1) + T_2(v_2) \\
&= (T_1(v_1) + T_2(v_1)) + (T_1(v_2) + T_2(v_2)) \\
&= (T_1 + T_2)(v_1) + (T_1 + T_2)(v_2),
\end{aligned}$$

and,

$$\begin{aligned}(T_1+T_2)(\alpha v_1) &= T_1(\alpha v_1)+T_2(\alpha v_1)\\ &= \alpha T_1(v_1)+\alpha T_2(v_1)\\ &= \alpha(T_1(v_1)+T_2(v_1))\\ &= \alpha(T_1+T_2)(v_1),\end{aligned}$$

so, $T_1+T_2 \in L(V,V')$. Similarly, if $T \in L(V,V')$, and $\alpha \in \mathbb{F}$, then $\alpha T \in L(V,V')$.

■

Proposition 1.7.2. *If dim* $\mathrm{V}=\mathrm{n}$, *dim* $\mathrm{V}'=\mathrm{m}$ *then dim* $\mathrm{L(V,V')}=\mathrm{nm}$.

Proof: Let $\mathcal{A}=(v_1,\ldots,v_n)$, $\mathcal{A}'=(v_1',\ldots,v_m')$ be ordered bases for V, V' respectively. Define

$$L(V,V')\xrightarrow{(\cdot)_{\mathcal{A}'\mathcal{A}}}M_{mn}(\mathbb{F}),\quad T\mapsto T_{\mathcal{A}'\mathcal{A}}$$

We shall show that $(\cdot)_{\mathcal{A}'\mathcal{A}}$ is

(i) a linear transformation;

(ii) an isomorphism.

For (i), we show first that

$$(T_1+T_2)_{\mathcal{A}'\mathcal{A}}=(T_1)_{\mathcal{A}'\mathcal{A}}+(T_2)_{\mathcal{A}'\mathcal{A}}$$

Let

$$(T_1)_{\mathcal{A}'\mathcal{A}}=[\alpha_{ij}^{(1)}],\ (T_2)_{\mathcal{A}'\mathcal{A}}=[\alpha_{ij}^{(2)}].$$

This means that

$$T_1(v_j)=\sum_{i=1}^{m}\alpha_{ij}^{(1)}v_i',\quad j=1,2,\ldots,n$$

and

$$T_2(v_j)=\sum_{i=1}^{m}\alpha_{ij}^{(2)}v_i',\quad j=1,2,\ldots,n.$$

Therefore,

$$\begin{aligned}(T_1+T_2)(v_j) &= T_1(v_j)+T_2(v_j)\\ &= \sum_{i=1}^{m}\alpha_{ij}^{(1)}v_i'+\sum_{i=1}^{m}\alpha_{ij}^{(2)}v_i'\\ &= \sum_{i=1}^{m}(\alpha_{ij}^{(1)}+\alpha_{ij}^{(2)})v_i'.\end{aligned}$$

This implies that

$$\begin{aligned}(T_1+T_2)_{\mathcal{A}'\mathcal{A}} &= [\alpha_{ij}^{(1)}+\alpha_{ij}^{(2)}]\\ &= [\alpha_{ij}^{(1)}]+[\alpha_{ij}^{(2)}]\\ &= (T_1)_{\mathcal{A}'\mathcal{A}}+(T_2)_{\mathcal{A}'\mathcal{A}}\end{aligned}$$

Similarly, if $T \in L(V,V')$ and $\alpha \in \mathbb{F}$, then

$$(\alpha T)_{\mathcal{A}'\mathcal{A}} = \alpha T_{\mathcal{A}'\mathcal{A}}$$

Thus,

$$(\cdot)_{\mathcal{A}'\mathcal{A}} : L(V,V') \longrightarrow M_{mn}(\mathbb{F})$$

is a linear transformation.

Finally, let $[\alpha_{ij}] \in M_{mn}(\mathbb{F})$. By the *Extension by Linearity Theorem* (1.4.5), there exists, for each $j = 1, 2, \ldots, n$, a unique linear transformation satisfying

$$T(v_j) = \sum_{i=1}^{m} \alpha_{ij} v_i' \in V',$$

which says that $T_{\mathcal{A}'\mathcal{A}} = [\alpha_{ij}]$. This shows both that $(\cdot)_{\mathcal{A}'\mathcal{A}}$ is surjective and injective, and so

$$L(V,V') \cong M_{mn}(\mathbb{F}).$$

■

Definition. Let V be a vector space over $\mathbb{F}$ and set

$$V^* = L(V,\mathbb{F})$$

($\mathbb{F}$ is a vector space over itself with basis $\{1\}$.) V^* is called the *dual space* of V, and elements of V^* are called *linear functionals*.

Remarks:

(1) If $\dim V = n$ then $\dim V^* = 1 \cdot n = n$. Thus, $V \cong V^*$; however, this isomorphism is not “natural” since any non-trivial isomorphism $V \overset{\cong}{\rightarrow} V^*$ depends on choosing bases in both V and V^*.

(2) We may iterate the dual construction and form V^{**}, the *double dual* of V. Note, that in this case there is a “natural” mapping $\eta : V \to V^{**}$ given as follows. Let $v \in V$, $f \in V^*$, and set

$$\eta(v)(f) \;=\; f(v) \in \mathbb{F}.$$

> One easily checks that η is a linear transformation; note that since its definition does not depend on the choice of any bases, it is said to be "natural."
>
> We may prove that $\eta : V \to V^*$ is injective, as follows. If $0 \neq v \in V$, then we may extend v to a basis $\{v = v_1, v_2, \ldots, v_n\}$ of V, using the *Basis Extension Theorem* (1.3.1). Now apply the *Extension by Linearity Theorem* (1.4.5) to infer the existence of a linear functional $f : V \to \mathbb{F}$ such that $f(v_i) = 1$, $i = 1, 2, \ldots, n$. Therefore, $\eta(v)(f) = f(v) = 1$, so in particular, $\eta(v) \neq 0$. Thus, $\eta : V \to V^{**}$ is injective, as claimed. Finally, note that η must be an isomorphism since V and V^{**} have the same dimension.

Definition. Let V have dimension n with ordered basis $\mathcal{A} = (v_1, v_2, \ldots, v_n)$. We define elements $v_i^* \in V^*$ by setting

$$v_i^*(v_j) = \delta_{ij} \in \mathbb{F}, \qquad j = 1, \ldots, n.$$

By the *Extension by Linearity Theorem* (1.4.5), the above recipe uniquely defines a linear transformation (functional) $v_i^* : V \longrightarrow \mathbb{F}$.

Claim: $\mathcal{A}' = (v_1^*, \ldots, v_n^*)$ is an ordered basis of V^*. Indeed, if

$$\sum_{i=1}^{n} \alpha_i v_i^* = 0 \in V^*$$

then

$$(\sum_{i=1}^{n} \alpha_i v_i^*)(v) = 0,$$

for all $v \in V$. In particular,

$$\begin{aligned} 0 &= (\sum_{i=1}^{n} \alpha_i v_i^*)(v_j) \\ &= \sum_{i=1}^{n} \alpha_i v_i^*(v_j) \\ &= \alpha_j \cdot 1 = \alpha_j. \end{aligned}$$

Since j was arbitrary, we see that $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$, and so $v_1^*, v_2^*, \ldots, v_n^*$ are linearly independent. Since $\dim V^* = n$, we conclude that $(v_1^*, v_2^*, ..., v_n^*)$ is an ordered basis, called the *dual basis* of $\mathcal{A} = (v_1, v_2, ..., v_n)$.

By the same token, starting with an ordered basis $(f_1, f_2, \ldots, f_n)$ of V^*, we have the dual basis $(f_1^*, f_2^*, \ldots, f_n^*)$ of V^{**}. In view of the isomorphism $\eta : V \to V^{**}$ discussed above, we have vectors $v_i^* \in V$ defined by $\eta(v_i^*) = f_i^*$, $i = 1, 2, \ldots, n$ and so $f_i(v_j^*) = \eta(v_j^*)(f_i) = f_j^*(f_i) = \delta_{ij}$, and so $(v_1^*, v_2^*, \ldots, v_n^*)$ is dual to $(f_1, f_2, \ldots, f_n)$.

Remark: Assume that V is a vector space of countably infinite dimension with basis $\{v_1, v_2, \ldots, \}$. We may define functionals $v_1^*, v_2^*, \ldots \in V^*$ exactly as above by setting

$$v_i^*(v_j) = \delta_{ij}$$

(apply the *Extension by Linearity Theorem* (1.4.5) [which is valid here, also]). However, we can show that:

$$V^* \neq \langle v_1^*, v_2^*, \ldots, \rangle$$

To this end, define $f \in V^*$ by

$$f(v_i) = 1 \qquad i = 1, 2, \ldots,$$

Then we cannot express

$$f = \sum_{i=1}^{m} \alpha_i v_i^*.$$

In fact, if V is not finite-dimensional, then in the sense of cardinal numbers,

$$\dim V < \dim V^*.$$

Definition. Let V be a vector space and let $W \subseteq V$ be a subspace. Set

$$\mathrm{Ann}(W) \;=\; \{f \in V^* | f(W) = 0\} \subseteq V^*$$

Then $\mathrm{Ann}(W)$ is a subspace of V^*, called the *annihilator* of W.
Note that $W_1 \subseteq W_2 \subseteq V$ implies

$$\mathrm{Ann}(W_1) \supseteq \mathrm{Ann}(W_2),$$

and so $\mathrm{Ann}(\cdot)$ is an "inclusion-reversing" map. Similarly, if $L \subseteq V^*$ is a subspace, we set

$$\mathrm{Ann}^*(L) = \{v \in V | f(v) = 0 \text{ for all } f \in L\} \subseteq V.$$

Definition. Let V be a vector space (possibly infinite dimensional). If $H \subseteq V$ is a subspace of V such that

$$\dim V/H = 1,$$

we call H a *hyperplane* of V.

Proposition 1.7.3. *Assume that* $\mathrm{dimV} = \mathrm{n} < \infty$ *and that* W *is a* k*-dimensional subspace of* V*. Then,* $\dim Ann(\mathrm{W}) = \mathrm{n} - \mathrm{k}$*. Likewise, if* $L \subseteq V^*$ *has dimension* m*, then* $\dim Ann^*(\mathrm{L}) = \mathrm{n} - \mathrm{m}$.

Proof: Let $(v_1, ..., v_k)$ be an ordered basis of W and extend it to a basis $(v_1, ..., v_k, v_{k+1}, ..., v_n)$ of V. Let $(v_1^*, ..., v_n^*)$ be the dual basis in V^*. If

$$f = \sum_{i=1}^{n} \alpha_i v_i^* \in \mathrm{Ann}(W),$$

then for all $i = 1, ..., k$, we have $0 = f(v_i) = \alpha_i$. Therefore, $f \in \langle v_{k+1}^*, ..., v_n^* \rangle$. Clearly,

$$\langle v_{k+1}^*, ..., v_n^* \rangle \subseteq \mathrm{Ann}(W),$$

and so $\dim \mathrm{Ann}(W) = n - k$ as claimed. The second statement follows from the isomorphism $V \cong V^{**}$.

■

Corollary 1.7.3.1. *If* $dim \mathrm{V} = \mathrm{n} < \infty$ *then* $Ann : \{$*subspaces of* $\mathrm{V}\} \longrightarrow \{$*subspaces of* $\mathrm{V}^*\}$ *is a bijection with inverse* Ann^*.

Proof: Note that if $W \subseteq V$ is a subspace, then $W \subseteq \mathrm{Ann}^*\mathrm{Ann}(W)$ and apply (1.7.3).

■

Example. If $0 \neq f \in V^*$, then $\ker f = H$ is a hyperplane of V. Indeed,

$$V \xrightarrow{f} \mathbb{F}$$

is surjective and so by the *Fundamental Homomorphism Theorem* (1.6.4), we have $V/H \cong \mathbb{F}$. Therefore $\dim (V/H) = 1$, which implies that H is a hyperplane. Conversely, if $H \subseteq V$ is a hyperplane, then $\dim (V/H) = 1$, forcing $V/H \cong \mathbb{F}$. Let $T : V/H \xrightarrow{\cong} \mathbb{F}$. Now, set $f = T \circ \pi_H$:

$$\begin{tikzcd}
V \arrow[rr, "\pi_H"] \arrow[dr, "f"'] & & V' \arrow[dl, "T"] \\
& V/H &
\end{tikzcd}$$

Note that $\ker f = \ker \pi_H = H$. More generally, if $W \subseteq V$ and if $\dim(V/W) = m$, we say that W has *codimension* m in V. Therefore, hyperplanes have codimension 1.

Proposition 1.7.4. *If $W \subseteq V$ has codimension m, then there exist hyperplanes $H_1, H_2, \ldots, H_m$ such that $W = \bigcap_{i=1}^{m} H_i$.*

Proof. By *Proposition* 1.7.3, we have that $\dim \operatorname{Ann}(W) = m$ and so $\operatorname{Ann}(W)$ has a basis $\{f_1, f_2, ..., f_m\}$. By (1.7.3.1) we have

$$W = \operatorname{Ann}^*(f_1, ..., f_n) = \bigcap_{i=1}^{m} \operatorname{Ann}^*(f_i).$$

Setting $H_i = \operatorname{Ann}^*(f_i)$ $(= \ker f_i)$, $i = 1, ..., m$, we see that each H_i is a hyperplane and

$$W = \bigcap_{i=1}^{m} H_i.$$

■

Let $T : V \longrightarrow V'$ be a linear transformation. We define a linear transformation $T^* : V^* \longrightarrow V^*$ by the following diagram:

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[dr, "T^*(f')"'] & V' \arrow[d, "f' \in V'^*"] \\
& \mathbb{F}
\end{tikzcd}$$

In other words, if $f' \in V'^*$, set

$$T^*(f') = f' \circ T : V \longrightarrow \mathbb{F}.$$

Note that $T^*(f') \in V^*$ since the composition of two linear transformations is linear. T^* is linear as

$$\begin{aligned}
T^*(f'_1 + f'_2) &= (f'_1 + f'_2) \circ T \\
&= f'_1 \circ T + f'_2 \circ T \\
&= T^*(f'_1) + T^*(f'_2).
\end{aligned}$$

Also, it is clear that

$$T^*(\alpha f') = \alpha T^*(f').$$

Note that if $v \in V, f' \in V'^*$,

$$T^*(f')(v) = f'(T(v)) = (f' \circ T)(v),$$

so $T^*(f') = f' \circ T$, proving that the diagram commutes.

■

Proposition 1.7.5. *Let $T : V \longrightarrow V'$ be a linear transformation. Let $\mathcal{A} = (v_1, \ldots, v_n)$, $\mathcal{A}' = (v'_1, \ldots, v'_m)$ be ordered bases for V and V', respectively, and let $\mathcal{A}^*$, $\mathcal{A}'^*$ be the corresponding dual bases of V^*, V'^*, respectively. Then,*

$$T^*_{\mathcal{A}^*\mathcal{A}'^*} = (T_{\mathcal{A}'\mathcal{A}})^\intercal$$

where, for any matrix A, $A^\intercal$ denotes the transpose *of A.*

Proof: Let $T_{\mathcal{A}'\mathcal{A}} = [\alpha_{ij}]$, so

$$T(v_j) = \sum_{i=1}^{m} \alpha_{ij} v'_i, \qquad j = 1, 2, \ldots, m.$$

Let $T^*_{\mathcal{A}^*\mathcal{A}'^*} = [\beta_{ij}] \in M_{nm}(\mathbb{F})$, that is,

$$T^*(v'^*_j) = \sum_{i=1}^{n} \beta_{ij} v^*_i, \qquad j = 1, \ldots, m.$$

Then,

$$\begin{aligned} \beta_{ij} &= (\sum_{k=1}^{n} \beta_{kj} v^*_k)(v_i) \\ &= T^*(v'^*_j)(v_i) \\ &= v'^*_j(T(v_i)) \\ &= v'^*_j(\sum_{k=1}^{m} \alpha_{ki} v'_k) \\ &= \alpha_{ji}. \end{aligned}$$

■

Finally, let $T : V \to W$ be a linear transformation of $\mathbb{F}$-vector spaces. We define the *adjoint* of T to be the linear transformation $T^* : W^* \to V^*$ by setting

$$T^*(f)(v) = f(T(v)), \qquad f \in W^*, v \in V.$$

In other words, $T^*(f) = f \circ T$, and so $T^*(f)$ is certainly a functional defined on V. To see that T^* is itself linear, we proceed in the obvious fashion: if $\alpha_1, \alpha_2 \in \mathbb{F}, f_1, f_2 \in W^*$, and if $v \in V$, then

$$\begin{aligned}
T^*(\alpha_1 f_1 + \alpha_2 f_2)(v) &= (\alpha_1 f_1 + \alpha_2 f_2)(T(v)) \\
&= \alpha_1 f_1(T(v)) + \alpha_2 f_2(T(v)) \\
&= \alpha_1 T^*(f_1)(v) + \alpha_2 T^*(f_2)(v) \\
&= (\alpha_1 T^*(f_1) + \alpha_2 T^*(f_2))(v).
\end{aligned}$$

Therefore $T^* : W^* \to V^*$ is a linear transformation. We shall study adjoints in much more detail in *Section* 3.2.

Chapter 2

Eigenvalues and Eigenvectors

2.1 Basic Definitions

Definition. Let $T : V \longrightarrow V$ be a linear transformation of the vector space V into itself. The vector $0 \neq v \in V$ is said to be an *eigenvector* of T if there exists $\lambda \in \mathbb{F}$ with $T(v) = \lambda v$. The scalar λ is called the *eigenvalue* of T corresponding to v.

Example. Let V be the set of all infinitely differentiable functions and let

$$T = \frac{d}{dx} : V \longrightarrow V$$

Then, $v(x) = e^{\lambda x}$ is an eigenvector of T because

$$\frac{d}{dx} v(x) = \lambda v(x).$$

Definition. Let $T : V \longrightarrow V$ be a linear transformation on the finite dimensional space V. Define the *determinant* of T:

$$\det(T) = \det(T_{\mathcal{A}})$$

where $\mathcal{A}$ is an ordered basis of V. Note that this is well-defined, for if $\mathcal{A}'$ is another ordered basis of V, and if

$$C_{\mathcal{A}\mathcal{A}'} = P$$

is the change of basis matrix then

$$T_{\mathcal{A}'} = P^{-1} T_{\mathcal{A}} P.$$

By familiar properties of the determinant, we have

$$\begin{aligned}\det(T_{\mathcal{A}'}) &= \det(P^{-1}T_{\mathcal{A}'}P)\\ &= \det(P^{-1})\det(T_{\mathcal{A}})\det(P)\\ &= \det(T_{\mathcal{A}})\det(P^{-1})\det(P)\\ &= \det(T_{\mathcal{A}})\det(P^{-1}P)\\ &= \det(T_{\mathcal{A}}).\end{aligned}$$

Recall that a matrix A is invertible if and only if

$$\det A \neq 0.$$

Correspondingly, if $T : V \longrightarrow V$ is a linear transformation on the finite dimensional vector space V, then T is invertible if and only if

$$\det T \neq 0.$$

This implies the following.

Proposition 2.1.1. *Let $T : V \longrightarrow V$ be a linear transformation on the finite-dimensional vector space V. Then $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if* $\det(T - \lambda I) = 0$.

Proof. If λ is an eigenvalue of T with eigenvector $v \in V$, then it is clear that $v \in \ker(T - \lambda I)$, and so $\det(T - \lambda I) = 0$. The converse is entirely similar.

■

Calculation of eigenvalues (and eigenvectors) of $T : V \longrightarrow V$:

(1) Relative to the ordered basis $\mathcal{A}$, set $A = T_{\mathcal{A}}$ and solve the (polynomial) equation

$$\det(xI - A) = 0$$

for x. (If $\dim V = n$, then $\det(xI - A)$ is a polynomial of degree n. This polynomial is called the *characteristic polynomial* of T and is denoted $c_T(x)$).

(2) For each solution $x = \lambda \in \mathbb{F}$, solve the equation

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

for v. In terms of matrices and solutions of homogeneous linear systems, one solves

$$(A - \lambda I) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

a system of n equations and n unknowns. If $x_1 = \alpha_1, x_2 = \alpha_2, ..., x_n = \alpha_n$ is a nontrivial solution, then

$$v = \sum_{i=1}^{n} \alpha_i v_i \in V$$

is an eigenvector of T corresponding to λ.

Theorem 2.1.2. *Let $T : V \longrightarrow V$ be a linear transformation of the finite dimensional vector space V. Then T is diagonalizable if and only if V has an ordered basis consisting entirely of eigenvectors.*

Proof. Assume T is diagonalizable. Then (see *page* 27), there exists an ordered basis

$$\mathcal{A} = (v_1, v_2, ..., v_n)$$

such that

$$T_{\mathcal{A}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

This means that

$$\begin{aligned} T(v_1) &= \lambda_1 v_1 \\ T(v_2) &= \lambda_2 v_2 \\ &\vdots \\ T(v_n) &= \lambda_n v_n \end{aligned}$$

and so each v_i is an eigenvector, which is to say that $\mathcal{A}$ consists entirely of eigenvectors. Conversely, if $\mathcal{A}$ consists entirely of eigenvectors,

$$\begin{aligned} T(v_1) &= \lambda_1 v_1 \\ &\vdots \end{aligned}$$

$$T(v_n) = \lambda_n v_n$$

for suitable $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{F}$. Thus,

$$T_{\mathcal{A}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix},$$

and so T is diagonalizable.

■

In terms of matrices, a matrix A is diagonalizable if and only if there exists an invertible matrix P such that

$$P^{-1}AP = D,$$

a diagonal matrix.

If

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix},$$

and if

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

then

$$AP = PD = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix}$$

$$= \left[\lambda_1 \begin{pmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{n1} \end{pmatrix} \quad \lambda_2 \begin{pmatrix} p_{12} \\ p_{22} \\ \vdots \\ p_{n2} \end{pmatrix} \quad \cdots \quad \lambda_n \begin{pmatrix} p_{1n} \\ p_{2n} \\ \vdots \\ p_{nn} \end{pmatrix} \right].$$

As for the left hand side:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

$$= \left[A \begin{pmatrix} p_{11} \\ \vdots \\ p_{n1} \end{pmatrix} \quad \cdots \quad A \begin{pmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{pmatrix} \right].$$

Comparing both sides, we have

$$\begin{aligned} A \begin{pmatrix} p_{11} \\ \vdots \\ p_{n1} \end{pmatrix} &= \lambda_1 \begin{pmatrix} p_{11} \\ \vdots \\ p_{n1} \end{pmatrix} \\ \vdots \quad & \quad \vdots \\ A \begin{pmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{pmatrix} &= \lambda_n \begin{pmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{pmatrix} \end{aligned}$$

i.e., the *i-th* column of P is an eigenvector in $\mathbb{F}^n$ with eigenvalue λ_i.

Example. Let

$$A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

If possible, find invertible P with $P^{-1}AP = D$ is a diagonal matrix.

Solution:

$$\begin{aligned} \det(A - xI) &= 0, \text{ so} \\ \det \begin{bmatrix} -2-x & 1 \\ 1 & -2-x \end{bmatrix} &= 0 \\ (2+x)^2 - 1 &= 0 \\ 2 + x &= \pm 1 \\ x &= -2 \pm 1 \\ &= -3, -1. \end{aligned}$$

If $\lambda_1 = -3$, then

$$(A - \lambda_1 I) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

if and only if

$$x_1 = -x_2.$$

Therefore, corresponding to $\lambda_1 = -3$ we have an eienvector $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

If $\lambda_2 = -1$, then

$$(A - \lambda_1 I) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

if and only if

$$-x_1 + x_2 = 0$$

and so

$$x_1 = x_2.$$

Thus, corresponding to $\lambda_2 = -1$ we may take

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Set

$$P = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}.$$

We get

$$P^{-1}AP = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}.$$

Occasionally, we need to determine P^{-1}:

$$P^{-1} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

2.1.1 The Matrix Exponential

Let $A \in M_n(\mathbb{R})$. We define

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

This is a convergent series. If we introduce a parameter, we have

$$\begin{aligned} e^{tA} &= \sum_{k=0}^{\infty} \frac{1}{k!}(tA)^k \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \end{aligned}$$

with convergence for every $t \in \mathbb{R}$.

Example. If

$$A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix},$$

as in the section above, we compute e^{tA}. Note first that the individual powers of tA are not easy to calculate, and so a direct approach is not feasible. However, we have already seen that $P^{-1}AP = D$, where

$$P = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} e^{tD} &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}^k \\ &= \begin{bmatrix} \sum_{k=0}^{\infty} \frac{(-3t)^k}{k!} & 0 \\ 0 & \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} \end{bmatrix} \\ &= \begin{bmatrix} e^{-3t} & 0 \\ 0 & e^{-t} \end{bmatrix}. \end{aligned}$$

But $P^{-1}AP = D = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix}$,
so $A = PDP^{-1}$, forcing

$$\begin{aligned} e^{tA} &= \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} (PDP^{-1})^k. \end{aligned}$$

Note, however, that

$$\begin{aligned}(PDP^{-1})^k &= (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})\\ &= PD^kP^{-1},\end{aligned}$$

and so,

$$\begin{aligned}e^{tA} &= \sum_{k=0}^{\infty}\frac{t^k}{k!}(PD^kP^{-1})\\ &= P(\sum_{k=0}^{\infty}\frac{t^k}{k!}D^k)P^{-1}\\ &= \begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}e^{-3t} & 0\\ 0 & e^{-t}\end{bmatrix}\begin{bmatrix}-\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}.\end{aligned}$$

Applications. Let $\mathbb{F} = \mathbb{R}$, the field of real numbers.

Consider the first-order ODE:

$$X'(t) = AX(t),$$

where

$$X(0) = X_0 = \begin{bmatrix}x_1(0)\\ \vdots\\ x_n(0)\end{bmatrix}, \qquad A \in M_n(\mathbb{R}),\ X(t) = \begin{bmatrix}x_1(t)\\ \vdots\\ x_n(t)\end{bmatrix}.$$

The general solution is given by

$$X(t) = e^{tA}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$$

where $c_1, ..., c_n$ are arbitrary parameters. The particular solution is

$$X(t) = e^{tA}X_0.$$

As observed above, this can be solved easily if we can find matrix P satisfying

$$P^{-1}AP = D(= \text{diagonal matrix}).$$

Next, we consider systems of second-order ODE. First of all, if $\lambda > 0$, the ODE

$$x''(t) = -\lambda x(t)$$

has a general solution of the form

$$x(t) = a_0 \sin(\sqrt{\lambda} t + b_0),$$

a purely sinusoidal solution. A particular instance of this type of problem comes from elementary physics.

Linear Harmonic Oscillator. Consider the linear harmonic oscillator with equal masses m and spring constants k. It is a simple application of Hooke's Law to determine the equation of motion for the evolution of the system, given that the initial positions of the masses are $x_1(0) = x_1^{(0)}$, $x_2(0) = x_2^{(0)}$ and the initial velocities are zero.

```
/|                                          |\
/|      k             k             k       |\
/|----oooooo---[]---oooooo---[]---oooooo----|\
/|             m             m              |\
/|                                          |\
   ------------|-------------|------------
```

$x_1 = 0$ $x_2 = 0$

The relevant equations are

$$\begin{array}{rcrcr} x_1''(t) & = & -2kx_1(t) & + & kx_2(t) \\ x_2''(t) & = & kx_1(t) & - & 2kx_2(t). \end{array}$$

The above system can be written in matrix form as:

$$X''(t) = \begin{bmatrix} -2k & k \\ k & -2k \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix},$$

$$X'(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad X(0) = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix}.$$

We continue the above analysis in the following more general context. Suppose we have the second-order system of ODE

$$X''(t) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

$$X'(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad X(0) = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix}.$$

Suppose there exists an invertible matrix P with

$$P^{-1}AP = \begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}, \qquad \lambda_1, \lambda_2 > 0.$$

Define new variables y_1, y_2 by the equation

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = P \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}.$$

Then, the ODE becomes

$$P \begin{bmatrix} y_1''(t) \\ y_2''(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} P \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix},$$

and so,

$$\begin{aligned} \begin{bmatrix} y_1''(t) \\ y_2''(t) \end{bmatrix} &= P^{-1}AP \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} \\ &= \begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}, \end{aligned}$$

thereby decoupling the system into two second-order ODEs:

$$\begin{aligned} y_1''(t) &= -\lambda_1 y_1(t) \\ y_2''(t) &= -\lambda_2 y_2(t) \end{aligned}$$

with solutions that are purely sinusoidal.

2.2 Eigenvalues and the Minimal Polynomial

2.2.1 Some recollections about polynomials

Let $\mathbb{F}$ be our field and let $\mathbb{F}[x]$ be the ring of polynomials with coefficients in $\mathbb{F}$. We shall state without proof some familiar facts about polynomials, starting with the following:

2.2.1 (Division Algorithm). *Let $f(x), g(x) \in \mathbb{F}[x]$,$g(x) \neq 0$. Then there exist (unique) polynomials $q(x), r(x) \in \mathbb{F}[x]$ such that*

$$f(x) = q(x)g(x) + r(x)$$

where

$$\deg r(x) < \deg g(x).$$

■

2.2.2 (Unique Factorization). *If $f(x) \in \mathbb{F}[x]$ then $f(x)$ can be uniquely factored as*

$$f(x) = p_1(x)^{e_1} p_2(x)^{e_2} \cdots p_r(x)^{e_r}$$

where $p_1(x), ..., p_r(x)$ are distinct irreducible polynomials and $e_1, e_2, \ldots, e_r$ are positive integers.

■

Definition. Let $f(x), g(x) \in \mathbb{F}[x]$ and assume that $d(x)$ is a monic polynomial which is a divisor of both $f(x)$ and $g(x)$ (write $d(x)|f(x), d(x)|g(x)$) such that if $d_0(x)$ is another common divisor of $f(x), g(x)$ then

$$d_0(x)|d(x)$$

In this case, call $d(x)$ the *greatest common divisor* of $f(x), g(x)$, and write

$$d(x) = \mathrm{GCD}(f(x), g(x)).$$

Similarly, one defines the *least common multiple*, denoted $\mathrm{LCM}(f(x), g(x))$.

Remark: If $f(x), g(x)$ have been factored into irreducibles, then

$$\mathrm{GCD}(f(x), g(x))$$

$$\mathrm{LCM}(f(x), g(x))$$

are easy to calculate.

Example. Let $\mathbb{F} = \mathbb{R}$ and assume that

$$\begin{aligned} f(x) &= (x^2+1)^4(x^2-x+1)(x-2)^4(x+5), \text{and} \\ g(x) &= (x^2-x+1)^3(x-2)^3(x-64)^2. \end{aligned}$$

Then

$$\begin{aligned} \mathrm{GCD}(f,g) &= (x^2-x+1)(x-2)^3, \text{ and} \\ \mathrm{LCM}(f,g) &= (x^2+1)^4(x^2-x+1)^3(x-2)^4(x+5)(x-64)^2. \end{aligned}$$

Definition. Polynomials $f_1(x), f_2(x) \in \mathbb{F}[x]$ are said to be *relatively prime* if they share no common nontrivial factors, *i.e.*, if $\mathrm{GCD}(f_1(x), f_2(x)) = 1$.

Example. Let $\mathbb{F} = \mathbb{R}$,

$$\begin{aligned} f_1(x) &= (x-2)^2(x-4)(x^2+x+1)^2 \\ f_2(x) &= (x-3)(x^2+2x+2). \end{aligned}$$

Then $f_1(x), f_2(x)$ are relatively prime.

More generally, polynomials $f_1(x), ..., f_k(x)$ are said to be relatively prime if there is no common factor in the $f_1(x), ..., f_k(x)$.

Example. Suppose we have $f(x) \in \mathbb{F}[x]$ with factorization

$$f(x) = p_1(x)^{e_1} p_2(x)^{e_2} \cdots p_k(x)^{e_k}$$

where $p_1(x), ..., p_k(x) \in \mathbb{F}[x]$ are irreducible, pairwise distinct polynomials.

Form new polynomials

$$\begin{aligned} q_1(x) &= \widehat{p_1(x)^{e_1}} p_2(x)^{e_2} \cdots p_k(x)^{e_k}, \\ &\vdots \qquad\qquad\qquad \vdots \\ q_k(x) &= p_1(x)^{e_1} p_2(x)^{e_2} \cdots \widehat{p_k(x)^{e_k}}. \end{aligned}$$

The hat indicates we are not to include the factor beneath it.

Note that $q_1(x), ..., q_k(x)$ are relatively prime but no proper subset of these polynomials is relatively prime.

2.2.3 (Euclidean trick). *Let $f_1(x), ..., f_k(x)$ be relatively prime. Then there exist polynomials*

$$s_1(x), ..., s_k(x) \in \mathbb{F}[x]$$

such that

$$s_1(x)f_1(x) + s_2(x)f_2(x) + \cdots + s_k(x)f_k(x) = 1.$$

■

Example. Let $\mathbb{F} = \mathbb{R}$

$$\begin{aligned} f_1(x) &= x^2+x+1 \\ f_2(x) &= x^2-x+1. \end{aligned}$$

Then,

$$-\frac{1}{2}(x-1)f_1(x)+\frac{1}{2}(x+1)f_2(x)=1.$$

Let $T : V \longrightarrow V$ be a linear transformation. As we have seen, the $\mathbb{F}$-vector space $L(V,V)$ (the set of linear transformations $V \longrightarrow V$) has dimension n^2, if $\dim V = n$. Therefore, the linear transformations

$$I_V = T^0, T^1, ..., T^{n^2}$$

are $\mathbb{F}$-linearly dependent Therefore, there exist $\alpha_0, \alpha_1, ..., \alpha_{n^2} \in \mathbb{F}$ (not all 0) with

$$\sum_{i=0}^{n^2} \alpha_i T^i = 0.$$

In other words, if we set

$$0 \neq f(x) = \sum_{i=0}^{n^2} \alpha_i x^i \in \mathbb{F}[x]$$

then $f(T) = 0$.

Definition. Let $m_T(x)$ be the monic polynomial of *least* degree in the set of polynomials

$$\{0 \neq g(x) \in \mathbb{F}[x] | g(T) = 0\}.$$

Call $m_T(x)$ the *minimal polynomial* of T.

At least two questions ought to be asked now:

(1) How do we compute $m_T(x)$? (This question is generally difficult. However, there is an algorithm for doing this.)

(2) What is the relationship between $m_T(x)$ and $c_T(x)$ (the characteristic polynomial)?

Lemma 2.2.4. *Let $m_T(x)$ be the minimal polynomial of T and let $f(x) \in \mathbb{F}[x]$ with $f(T) = 0$. Then*

$$m_T(x) | f(x).$$

Proof. Use the division algorithm (2.2.1) and write

$$f(x) = q(x)m_T(x) + r(x);$$

we must show that $r(x) = 0$. But,

$$0 = f(T) = q(T)m_T(T) + r(T),$$

which implies that $r(T) = 0$. However, $\deg r(x) < \deg m_T(x)$. This contradicts the minimality of $m_T(x)$ unless $r(x) = 0$.

■

Definition. As with the minimal polynomial of T, we can define the *minimal polynomial of a vector relative to* T: if $0 \neq v \in V$, set $m_{T,v}(x) =$ the monic polynomial of least degree such that

$$m_{T,v}(T)(v) = 0.$$

The existance of $m_{T,v}(x)$ is guaranteed since $m_T(T)(v) = 0$.

Exactly as with the previous lemma, we have the following.

Lemma 2.2.5. *Let $0 \neq v \in V$ and let $f(x) \in \mathbb{F}[x]$ with $f(T)(v) = 0$. Then $m_{T,v}(x)|f(x)$. In particular, $m_{T,v}(x)|m_T(x)$.*

■

As we have already seen,

$$\deg m_T(x) \leq n^2.$$

However, we can do much better than this; it turns out that $\deg m_T(x) \leq n$, as we shall show below. First, however, we need some preliminary results.

Lemma 2.2.6. *If $0 \neq v \in V$ then $\deg m_{T,v}(x) \leq n = dim\,\mathrm{V}$.*

Proof. Note that

$$v = v_0 = T^0(v), v_1 = T(v), ..., v_n = T^n(v)$$

are linearly dependent, so there exist scalars $\alpha_0, ..., \alpha_n \in \mathbb{F}$ not all 0 with

$$\sum_{i=0}^{n} \alpha_i T^i(v) = 0.$$

If

$$f(x) = \sum_{i=0}^{n} \alpha_i x^i,$$

then

$$f(T)(v) = 0,$$

so

$$m_{T,v}(x) | f(x),$$

i.e.,

$$\deg m_{T,v}(x) \leq n.$$

■

Lemma 2.2.7. *Let $0 \neq v \in V$ and assume that $m_{T,v}(x) = f_1(x) \cdot f_2(x)$ where both factors are monic. If $w = f_1(T)(v)$, then $m_{T,w}(x) = f_2(x)$.*

Proof. Note that

$$\begin{aligned} f_2(T)(w) &= f_2(T)f_1(T)(v) \\ &= m_{T,v}(T)(v) = 0 \end{aligned}$$

by definition. This implies that $m_{T,w}(x) | f_2(x)$:

$$f_2(x) = q(x) m_{T,w}(x)$$

But

$$\begin{aligned} 0 &= m_{T,w}(T)(w) \\ &= m_{T,w}(T) f_1(T) v, \end{aligned}$$

which implies that

$$m_{T,v}(x) | m_{T,w}(x) f_1(x),$$

which of course implies that $f_2(x) | m_{T,w}(x)$.

■

Lemma 2.2.8. *Let $v_1, v_2 \in V$ and assume that*

$$\begin{aligned} m_{T,v_1}(x) &= f_1(x), \\ m_{T,v_2}(x) &= f_2(x) \end{aligned}$$

are relatively prime. Then

$$m_{T,v_1+v_2}(x) = f_1(x) f_2(x).$$

Proof. The polynomial $f(x) = f_1(x)f_2(x)$ satisfies

$$\begin{array}{rcl} f(T)(v_1+v_2) & = & f(T)(v_1) + f(T)(v_2) \\ & = & f_2(T)f_1(T)(v_1) + f_1(T)f_2(T)(v_2) \\ & = & f_2(T)(0) + f_1(T)(0) \\ & = & 0 + 0 = 0. \end{array}$$

Thus $m_{T,v_1+v_2}(x)|f_1(x)f_2(x)$. Next, assume that $g(x) \in \mathbb{F}[x]$ satisfies

$$g(T)(v_1+v_2) = 0.$$

We wish to show that $f(x)|g(x)$; this will say, in particular, that $f(x)|m_{T,v_1+v_2}(x)$. Since $f(x) = f_1(x)f_2(x)$ and since $f_1(x), f_2(x)$ are relatively prime, it is enough to show that

$$f_1(x), f_2(x)|g(x).$$

By hypothesis,

$$g(T)v_1 = -g(T)v_2.$$

Therefore,

$$f_1(T)(g(T)v_1) = g(T)f_1(T)v_1 = 0$$

and

$$\begin{array}{rcl} f_2(T)(g(T)(v_1)) & = & f_2(T)(-g(T)v_2) \\ & = & -g(T)f_2(T)v_2 = 0. \end{array}$$

Likewise, $f_1(T), f_2(T)$ both kill $g(T)v_2$.

By the Euclidean Trick (2.2.3), there exist

$$s_1(x), s_2(x) \in \mathbb{F}[x],$$

$$s_1(x)f_1(x) + s_2(x)f_2(x) = 1,$$

which implies that upon substitution by T we have

$$s_1(T)f_1(T) + s_2(T)f_2(T) = I_V,$$

where I_V is the identity transformation on V. Therefore,

$$\begin{array}{rcl} g(T)v_1 & = & I_V \cdot g(T)v_1 \\ & = & (s_1(T)f_1(T) + f_2(T)s_2(T))g(T)v_1 \\ & = & 0. \end{array}$$

Similarly, $g(T)v_2 = 0$. Therefore, as $f_1(x) = m_{T,v_1}(x)$, we get $f_1(x)|g(x)$; likewise $f_2(x)|g(x)$.

■

Proposition 2.2.9. *Let $T : V \longrightarrow V$ be a linear transformation. Then there exists a non-zero vector $v \in V$ with $m_{T,v}(x) = m_T(x)$.*

Proof. Since $0 \neq v \in V$ implies that $m_{T,v}(x)|m_T(x)$, we may choose $v \in V$ so that $\deg m_{T,v}(x)$ is maximal. Now let $w \in V$ be an arbitrary non-zero vector. We claim that

$$m_{T,w}(x)|m_{T,v}(x)$$

Factor both polynomials into products of irreducible polynomials.

$$\begin{array}{rcll} m_{T,w}(x) & = & p_1(x)^{e_1}p_2(x)^{e_2}\cdots p_r(x)^{e_r} & \quad e_i \geq 0 \\ m_{T,v}(x) & = & p_1(x)^{f_1}p_2(x)^{f_2}\cdots p_r(x)^{f_r} & \quad f_i \geq 0 \end{array}$$

If $m_{T,w}(x) \not| m_{T,v}(x)$, then there exists an index $i :$ such that

$$e_i > f_i.$$

Let

$$w' = p_1(T)^{e_1}\cdots\widehat{p_i(T)^{e_i}}\cdots p_r(T)^{e_r}w.$$

Then by *Lemma* 2.2.7

$$m_{T,w'}(x) = p_i(x)^{e_i}.$$

If $v' = p_i(T)^{f_i}v$, then

$$m_{T,v'}(x) = p_1(x)^{f_1}\cdots\widehat{p_i(x)^{f_i}}\cdots p_r(x)^{f_r}.$$

Thus, by *Lemma* 2.2.8

$$m_{T,v'+w'}(x) \;=\; p_1(x)^{f_1}\cdots p_{i-1}(x)^{f_{i-1}}p_i(x)^{e_i}p_{i+1}(x)^{f_{i+1}}\cdots p_r(x)^{f_r}$$

which has degree greater than $\deg m_{T,v}(x)$, contradicting the maximality of the degree of the latter, thereby proving our claim. It follows that, for all $w \in V$, $m_{T,v}(T)w = 0$, that is, $m_{T,v}(T) = 0$ and so $m_T(x)|m_{T,v}(x)$. Clearly, then, $m_T(x) = m_{T,v}(x)$, and we are done.

■

Corollary 2.2.9.1. $\deg m_T(x) \leq n = dim$ V.

Proof. Choose $0 \neq v \in V$ with

$$m_{T,v}(x) = m_T(x).$$

By *Lemma* 2.2.6

$$\deg m_{T,v}(x) \leq n,$$

so we're done.

■

Example. Let

$$V = \mathbb{R}^2 = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} : a, b \in \mathbb{R} \right\}, \quad T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Then $T^2 = I$, and so $m_T(x) = x^2 - 1$. Therefore if $0 \neq v \in V$, then

$$m_{T,v}(x) | x^2 - 1.$$

(1) If $m_{T,v}(x) = x - 1$, then

$$(T - I)v = 0,$$

forcing

$$Tv = Iv = v,$$

and so v is an eigenvector with eigenvalue 1.

(2) If $m_{T,v}(x) = x + 1$, then

$$(T + I)v = 0,$$

and so

$$Tv = -Iv = -v,$$

i.e., v is an eigenvector with eigenvalue -1.

(3) Here, if

$$v = \begin{bmatrix} 3 \\ 2 \end{bmatrix},$$

then, of course,

$$m_{T,v}(x) | x^2 - 1.$$

Note, however that

$$(T - I)v = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and that

$$(T + I)v = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

It follows, therefore, that

$$m_{T,v}(x) = x^2 - 1.$$

Watch this : if $v_1 = (T + I)v$, then

$$m_{T,v_1}(x) = \frac{x^2 - 1}{x + 1} = x - 1.$$

That is, v_1 is an eigenvector with eigenvalue 1.

$$v_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}.$$

$$(T\left(\begin{bmatrix} 5 \\ 5 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 5 \end{bmatrix})$$

Likewise, if

$$v_{-1} = (T - I)v,$$

then

$$m_{T,v-1}(x) = \frac{x^2 - 1}{x - 1} = x + 1$$

i.e., v_{-1} is an eigenvector with eigenvalue -1.

Example. $V = \mathbb{R}^2$; let

$$T = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}.$$

Note that

$$\begin{aligned} T^2 + T + I \quad &= \quad \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} + \\ &\qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

so,

$$m_T(x) | x^2 + x + 1.$$

But as $x^2 + x + 1$ is irreducible over $\mathbb{R}$, we must have

$$m_T(x) = x^2 + x + 1.$$

Likewise, if $0 \neq v \in V$, then

$$m_{T,v}(x) | x^2 + x + 1$$

which implies that

$$m_{T,v}(x) = x^2 + x + 1,$$

for every nonzero vector $v \in V$.

Example. Let

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

If we define

$$T = \begin{bmatrix} 0 & 0 & \cdots & \cdots & -a_0 \\ 1 & 0 & \cdots & \cdots & -a_1 \\ 0 & \vdots & \ddots & \ddots & -a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -a_{n-1} \end{bmatrix},$$

(often called the *companion matrix* of T) then setting

$$v = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

a pretty easy calculation reveals that

$$m_{T,v}(x) = f(x).$$

Since

$$\dim V = \dim \mathbb{R}^n = n = \deg(f(x))$$

we must have, by *Corollary* 2.2.9.1

$$m_T(x) = f(x).$$

As a result of the above example, we have the following corollary:

Corollary 2.2.9.2. *For any monic polynomial $f(x) \in \mathbb{F}[x]$, there is a linear transformation $T : V \longrightarrow V$ with*

$$m_T(T) = f(x).$$

■

The following is fundamental, both for the proof of the "Restricted Cayley-Hamilton Theorem" below, as well as for the proof of the "Spectral Theorem" in the next section. First of all, if $T : V \to V$ is a linear transformation and if $W \subseteq V$ is a subspace, we say that W is T-*invariant* if $T(W) \subseteq W$.

2.2.10 (Primary Decomposition Theorem, Part I). *Let $T : V \longrightarrow V$, dim*V $< \infty$*, $m_T(x) = p_1(x)^{e_1} \cdots p_k(x)^{e_k}$, where $p_1(x), ..., p_k(x)$ are distinct monic irreducible polynomials in $\mathbb{F}[x]$. For each $i = 1, ..., k$, set*

$$V_i = \ker\, p_i(T)^{e_i}.$$

Then each V_i is T-invariant and

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k.$$

Furthermore, if

$$T|_{V_i} : V_i \longrightarrow V_i$$

is the restriction of T to V_i, then

$$m_{T|_{V_i}}(x) = p_i(x)^{e_i}, \qquad i = 1, ..., k.$$

Proof. That each V_i is T-invariant is easy. Next, define polynomials

$$\begin{aligned} q_i(x) &= m_T(x)/p_i(x)^{e_i} \\ &= p_1(x)^{e_1} \cdots \widehat{p_i(x)^{e_i}} \cdots p_k(x)^{e_k}. \end{aligned}$$

Note that if $v \in V$, then $q_i(T)v \in V_i$ because

$$\begin{aligned} p_i(T)^{e_i} q_i(T)v &= m_T(T)v \\ &= 0. \end{aligned}$$

Next, as $q_1(x), ..., q_k(x)$ are relatively prime, then by the *Euclidean Trick* (2.2.3), we have $s_1(x), ..., s_k(x) \in \mathbb{F}[x]$ with

$$s_1(x)q_1(x) + s_2(x)q_2(x) + \cdots + s_k(x)q_k(x) = I_V.$$

Therefore,

$$s_1(T)q_1(T) + \cdots + s_k(T)q_k(T) = I_V,$$

and so for all $v \in V$,

$$\begin{aligned} v &= I_V(v) \\ &= s_1(T)q_1(T)v + \cdots + s_k(T)q_k(T)v \end{aligned}$$

But, $q_i(T)v \in V_i$, and V_i is T-invariant, so

$$s_i(T)(q_i(T)v) \in V_i,$$

also. In other words,

$$V = V_1 + V_2 + \cdots + V_k.$$

Finally, we show that the above sum is *direct.* To this end, let $w_i \in V_i, i = 1, 2, ..., k$ satisfy

$$w_1 + w_2 + \cdots + w_k = 0.$$

Now fix i, $1 \leq i \leq k$ and use the *Euclidean trick* (2.2.3) to obtain polynomials $s(x), t(x) \in \mathbb{F}[x]$ with

$$s(x)p_i(x)^{e_i} + t(x)q_i(x) = 1.$$

Here, we have

$$\begin{aligned} w_i &= I_V w_i \\ &= s(T)p_i(T)^{e_i}w_i + t(T)q_i(T)w_i \\ &= 0 + t(T)q_i(T)w_i. \end{aligned}$$

Since $t(T)q_i(T)w_j = 0$ for each $j \neq i$ we conclude also that $t(T)q_i(T)w_i = 0$, since $w_i = -\sum\limits_{j \neq i} w_j$. Therefore, the above implies that each $w_i = 0$ and so the sum is direct, as claimed.

Finally, note that

$$m_{T|_{V_i}}(x)|p_i(x)^{e_i},$$

and so we may write $m_{T|_{V_i}}(x) = p_i(x)^{f_i}$, where $f_i \leq e_i$. We set

$$f(x) = p_1(x)^{f_1} \cdots p_k(x)^{f_k}.$$

If $v \in V$ is arbitrary, we may write

$$v = v_1 + v_2 + \cdots + v_k$$

for suitable $v_i \in V_i$, $i = 1, ..., k$. Then

$$f(T)v = f(T)v_1 + \cdots + f(T)v_k;$$

but

$$\begin{aligned} f(T)v_i &= (\text{something})p_i(T)^{f_i}v_i \\ &= (\text{something})(0) \\ &= 0, \end{aligned}$$

as $p_i(x)^{f_i} = m_{T|_{V_i}}(x)$ and $v_i \in V_i$, $i = 1, ..., k$. Thus $f(T)v = 0$, and so $m_T(x)|f(x)$. Therefore, it follows that no f_i can be strictly less than e_i from which it follows that $m_{T|_{V_i}} = p_i(x)^{e_i}$ as claimed.

■

Definition. A linear transformation $P : V \to V$ is called an *idempotent* if and only if $P^2 = P$. Note that any idempotent must be a root of the polynomial $x(x-1)$ and hence must be diagonalizable (with all eigenvalues being 0 or 1). A family $\{P_1, P_2, \ldots, P_k\}$ of idempotents is called *orthogonal* if $P_iP_j = 0$ whenever $i \neq j$.

Now let $T : V \to V$ be a linear transformation with minimal polynomial $m_T(x) = p_1(x)^{e_1}p_2(x)^{e_2}\cdots p_k(x)^{e_k}$. As usual define the polynimials $q_i(x) = m_T(x)/p_i(x)^{e_i}$, $i = 1, 2, \ldots, k$ and let $s_1(x), s_2(x), \ldots, s_k(x)$ be determined so that

$$s_1(x)q_1(x) + s_2(x)q_2(x) + \cdots + s_k(x)q_k(x) = 1.$$

Set $P_i = s_i(T)q_i(T)$; obviously we have

$$P_1 + P_2 + \cdots + P_k \;=\; I_V.$$

Furthermore, note that if $i \neq j$, then

$$P_iP_j \;=\; s_i(T)q_i(T)s_j(T)q_j(T) = 0,$$

since obviously $m_T(x)|q_i(x)q_j(x)$ whenever $i \neq j$. Therefore, it follows that for each $i = 1, 2, \ldots, k$

$$P_i^2 \;=\; P_i(P_1 + P_2 + \cdots + P_k) = P_i,$$

and so $P_1, P_2, \ldots, P_k$ are orthogonal idempotents. Note that each of these idempotents clearly commute with T as each is a polynomial in T. In particular, each subspace P_iV is a T-invariant subspace of V, $i = 1, 2, \ldots, k$.

2.2.11 (Primary Decomposition Theorem, Part II). *Let $T : V \longrightarrow V$, dim V $< \infty$, $m_T(x) = p_1(x)^{e_1}\cdots p_k(x)^{e_k}$, where $p_1(x), ..., p_k(x)$ are distinct monic irreducible polynomials in $\mathbb{F}[x]$. Then there exist orthogonal idempotents $P_1, P_2, \ldots, P_k$ commuting with T such that*

$$V = P_1V \oplus P_2V \oplus \cdots \oplus P_kV;$$

furthermore, we have

$$P_iV \;=\; \ker p_i(T)^{e_i},\; i = 1, 2, \ldots, k.$$

PROOF. Since $P_1 + P_2 + \cdots + P_k = I_V$, we have

$$V \;=\; P_1V + P_2V + \cdots + P_kV.$$

If $v \in P_iV \cap \sum\limits_{j\neq i} P_jV$, then $v = \sum\limits_{i\neq j} P_j(v_j)$, where each $v_j \in V$, and so $v = P_i(v) = P_i \sum\limits_{j\neq i} P_j(v_j) = 0$. Therefore the above sum is direct. Finally, note that as $P_i = s_1(T)q_i(T)$, where the polynomials $s_i(x), q_i(x)$, $i = 1, 2, \ldots, k$ are constructed as usual, then $p_i(T)^{e_i}P_i = 0$, and so $P_iV \subseteq \ker\, p_i(T)^{e_i}$, $i = 1, 2, \ldots, k$. Since

$$P_1V \oplus P_2V \oplus \cdots \oplus P_kV \;=\; V \;=\; \ker\, p_1(T)^{e_1} \oplus \ker\, p_2(T)^{e_2} \oplus \cdots \oplus \ker\, p_k(T)^{e_k},$$

we infer that $P_iV = \ker\, p_i(T)^{e_i}$, $i = 1, 2, \ldots, k$.

■

Corollary 2.2.11.1 (Spectral Decomposition). *Let $T : V \to V$ be a diagonalizable linear transformation. Then there exist orthogonal idempotents $P_1, P_2, \ldots, P_k$, each commuting with T, and scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$ (the eigenvalues of T) such that*

$$T \;=\; \lambda_1P_1 + \lambda_2P_2 + \cdots + \lambda_kP_k.$$

PROOF. We have that $m_T(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_k)$; from (2.2.11) we have orthogonal idempotents $P_1, P_2, \ldots, P_k$ such that

$$V \;=\; P_1V \oplus P_2V \oplus \cdots \oplus P_kV,$$

where $P_iV = \ker\,(T - \lambda_iI_V)$, $i = 1, 2, \ldots, k$. That is to say, T acts as scalar multiplication by λ_i on P_iV, $i = 1, 2, \ldots, k$. The result follows.

■

2.2.12 (Restricted Cayley-Hamilton Theorem). *Let V be a finite-dimensional vector space over the algebraically closed field $\mathbb{F}$.*[1] *(For example, take $\mathbb{F} = \mathbb{C}$.) If $T : V \to V$ is a linear transformation, then*

$$m_T(x) | c_T(x).$$

[1] A field $\mathbb{F}$ is called *algebraically closed* if every polynomial $f(x) \in \mathbb{F}[x]$ splits completely into linear factors in $\mathbb{F}[x]$. Actually, the present theorem is valid even when the field in not algebraically closed. A proof of this more general version will be given in *Section* 4.3 in *Theorem* 4.3.3 as an application of tensor products.

Proof. We factor

$$m_T(x) = (x-\lambda_1)^{e_1}(x-\lambda_2)^{e_2}\cdots(x-\lambda_k)^{e_k}$$

for suitable $\lambda_1, ..., \lambda_k \in \mathbb{F}$. By the *Primary Decomposition Theorem* (2.2.10), we can decompose V as a direct sum:

$$V = V_1 \oplus \cdots \oplus V_k, \quad V_i = \ker{(T-\lambda_i I)^{e_i}}, \quad i = 1, ..., k.$$

We have that V_i is T-invariant and

$$m_{T|_{V_i}}(x) = (x-\lambda_i)^{e_i}.$$

From (2.3.9.1),

$$\begin{aligned} \deg m_{T|_{V_i}} = e_i &\leq \dim V_i \\ &= \deg c_{T|_{V_i}}(x). \end{aligned}$$

Claim: $c_{T|_{V_i}}(x) = (x-\lambda_i)^{f_i}$, where $f_i = \dim V_i$. If not, then $(x-\lambda)|c_{T|_{V_i}}(x)$ for some $\lambda \neq \lambda_i$, *i.e.*, the restriction of T to V_i has an eigenvalue other than λ_i Thus $v_i \in V_i$, $v_i \neq 0$ with $T(v_i) = \lambda v_i$, *i.e.*,

$$(T-\lambda I)v_i = 0.$$

By the *Euclidean Trick* (2.2.3) applied to $(x-\lambda_i)^{e_i}$ and $(x-\lambda)$ we get polynomials $s(x), t(x)$ with

$$s(x)(x-\lambda_i)^{e_i} + t(x)(x-\lambda) = 1.$$

Therefore,

$$\begin{aligned} v_i &= Iv_i \\ &= s(T)(T-\lambda_i I)^{e_i}v_i + t(T)(T-\lambda I)v_i \\ &= 0+0=0 \end{aligned}$$

a contradiction. This proves the claim. Finally, if we choose ordered bases $\mathcal{A}_i \subseteq V_i$, then we can obtain an ordered basis of V:

$$\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k = \mathcal{A}.$$

Relative to this basis, we have

$$T_{\mathcal{A}} = \begin{bmatrix} (T|_{V_1})_{\mathcal{A}_1} & 0 & \cdots & 0 \\ 0 & (T|_{V_2})_{\mathcal{A}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & (T|_{V_k})_{\mathcal{A}_k} \end{bmatrix}$$

and so

$$\begin{aligned}
c_T(x) &= \det(xI - T) \\
&= \det(xI - T_{\mathcal{A}}) \\
&= \det \begin{bmatrix} xI - (T|_{V_1})_{\mathcal{A}_1} & 0 & \cdots & 0 \\ 0 & xI - (T|_{V_2})_{\mathcal{A}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & xI - (T|_{V_k})_{\mathcal{A}_k} \end{bmatrix} \\
&= \det(xI - (T|_{V_1})_{\mathcal{A}_1}) \cdots \det(xI - (T|_{V_k})_{\mathcal{A}_k}) \\
&= c_{T|_{V_1}}(x) \cdots c_{T|_{V_k}}(x) \\
&= (x - \lambda_1)^{f_1} \cdots (x - \lambda_k)^{f_k}
\end{aligned}$$

Since $m_T(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}$, $e_i \leq f_i$, $i = 1, 2, ..., k$, and consequently

$$m_T(x) | c_T(x).$$

■

The following is immediate.

Corollary 2.2.12.1. *If* $\mathbb{F}$ *is algebraically closed, and* $T : V \to V$ *is a linear transformation of the* $\mathbb{F}$*-vector space, then* $c_T(T) = 0$, i.e., T *satisfies its own characteristic polynomial.*

■

Again, the above is valid even without assuming that the field $\mathbb{F}$ is algebraically closed. We defer the proof of this until *Section* 4.3.

Chapter 3

The Spectral Theorem

3.1 The Geometry of Hilbert Spaces

Unless specified otherwise, V is a finite-dimensional vector space over $\mathbb{C}$.

Definition. A *Hermitian Inner Product* $(\cdot,\cdot) : V \times V \longrightarrow \mathbb{C}$ is an inner product satisfying the following conditions: for all $v, w, v_1, v_2 \in V, \lambda \in \mathbb{C}$:

(1) $(v, v_1 + v_2) = (v, v_1) + (v, v_2)$

(2) $(v, \lambda w) = \lambda(v, w)$

(3) $(v, w) = \overline{(w, v)}$

(4) $(v, v) \geq 0$ with equality if and only if $v = 0$.

Note that

$$\begin{aligned}(v_1 + v_2, v) &= \overline{(v, v_1 + v_2)} \\ &= \overline{(v, v_1) + (v, v_2)} \\ &= \overline{(v, v_1)} + \overline{(v, v_2)} \\ &= (v_1, v) + (v_2, v)\end{aligned}$$

So $(\cdot,\cdot)$ is additive in the first coordinate. Note also that

$$\begin{aligned}(\lambda v, w) &= \overline{(w, \lambda v)} \\ &= \overline{\lambda(w, v)} \\ &= \overline{\lambda}\,\overline{(w, v)} \\ &= \overline{\lambda}(v, w).\end{aligned}$$

A finite-dimensional complex vector space V together with a Hermitian inner product is called a (finite-dimensional) complex *Hilbert Space.*

Lemma 3.1.1 (Cauchy-Schwarz Inequality). *If $v, w \in V$, then*

$$|(v,w)| \leq \|v\| \cdot \|w\|,$$

where

$$\|v\|^2 = (v,v).$$

PROOF. Define the quadratic function of $t \in \mathbb{R}$ as follows:

$$q(t) = \|v - tw\|^2 \geq 0.$$

Then

$$\begin{aligned} q(t) &= (v - tw, v - tw) \\ &= (v,v) - t(w,v) - t(v,w) + t^2(w,w) \\ &= \|v\|^2 - t((w,v) + (v,w)) + t^2\|w\|^2. \end{aligned}$$

From this we see that the discriminant of $q(t)$ satisfies

$$\mathrm{disc}(q(t)) = 4(\mathrm{Re}(w,v))^2 - 4\|w\|^2\|v\|^2 \leq 0.$$

Case 1: If $(w,v) \in \mathbb{R}$ then

$$\begin{aligned} 0 &\geq \mathrm{disc}(q(t)) \\ &= 4(v,w)^2 - 4\|w\|^2\|v\|^2, \\ 4\|w\|^2\|v\|^2 &\geq 4(v,w)^2 \\ \|w\|^2\|v\|^2 &\geq (v,w)^2 \\ \|w\|\|v\| &\geq |(v,w)|, \end{aligned}$$

proving the result in case $(v,w) \in \mathbb{R}$.

Case 2: $(v,w) \in \mathbb{C} - \mathbb{R}$

Set $u = \frac{1}{(w,v)}v$; then

$$\begin{aligned} (u,w) &= (\frac{1}{(w,v)} \cdot v, w) \\ &= \frac{1}{\overline{(w,v)}}(v,w) \\ &= \frac{(v,w)}{(v,w)} = 1 \in \mathbb{R}. \end{aligned}$$

By Case 1, we get

$$\begin{array}{rcl} 1 = |(u,w)| & \leq & \|u\| \cdot \|w\| \\ 1 \leq & & \|\frac{v}{(w,v)}\| \cdot \|w\| \end{array}$$

But, for any vector $x \in V$, and $\alpha \in \mathbb{C}$, $\|\alpha x\| = |\alpha|\|x\|$. Therefore,

$$1 \leq \frac{\|v\|}{|(w,v)|} \cdot \|w\|$$

$$|(w,v)| \leq \|v\|\|w\|.$$

Because $|(w,v)| = |(v,w)|$,

$$|(v,w)| \leq \|v\| \cdot \|w\|.$$

■

The Cauchy-Schwarz inequality can be used to show that the inner product defines a complex-valued *continuous* function $V \times V \to \mathbb{C}$; see *Exercise* 3 of *Appendix* H.

3.1.2 (Triangle Inequality). *If $v, w \in V$, then*

$$\|v+w\| \leq \|v\| + \|w\|.$$

PROOF. We have

$$\begin{array}{rcl} \|v+w\|^2 & = & (v+w, v+w) \\ & = & (v,v) + (v,w) + (w,v) + (w,w) \\ & = & \|v\|^2 + (v,w) + \overline{(v,w)} + \|w\|^2 \\ & \leq & \|v\|^2 + 2|(v,w)| + \|w\|^2 \\ & \leq & \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\ & = & (\|v\| + \|w\|)^2, \end{array}$$

and so

$$\|v+w\| \leq \|v\| + \|w\|.$$

■

The above allows us to define a "distance function" on V:

$$\mathrm{d}(v,w) = \|v-w\|.$$

This distance function satisfies the following conditions:

(i) $\mathrm{d}(v,w) \geq 0$ with equality if and only if $v = w$,

(ii) $\mathrm{d}(v,w) = \mathrm{d}(w,v)$, and

(iii) $\mathrm{d}(v,w) \leq \mathrm{d}(v,u) + \mathrm{d}(u,w)$,

for all $u, v, w \in V$. Thus, one can define "basic open sets" $U_\epsilon(v)$ in V by setting

$$U_\epsilon(v) = \{w \in V | \mathrm{d}(v,w) < \epsilon\}.$$

These sets form a base for a topology on V called the *metric topology* relative to d. The following pleasant features emerge:

1. $V \times V \xrightarrow{(,)} \mathbb{C}$ is continuous, when $V \times V$ is given the product topology and $\mathbb{C}$ carries the usual metric topology ($\mathrm{d}(z_1, z_2) = |z_1 - z_2|$).

2. Any linear transformation $T : V \longrightarrow V$ is continuous.

3. Let $\mathcal{B} = \{v \in V : \|v\| \leq 1\}$, and let $: V \longrightarrow V$ be a linear transformation. Then $T(\mathcal{B})$ is compact. (In the present setting, this is quite trivial: T is continuous and $\mathcal{B}$ is closed and bounded, hence is compact.)

4. If $f \in V^*$ then $f : V \longrightarrow \mathbb{C}$ is continuous.

5. If $W \subseteq V$ is a subspace, then W is closed (false in every infinite dimensional Hilbert space).

6. V is *complete*, *i.e.*, if (v_n) is a Cauchy sequence in V (given $\epsilon > 0$, there exists N such that for all $n, m \geq N$, $\|v_n - v_m\| < \epsilon$), then there exists $v \in V$ with
$$\lim_{n \longrightarrow \infty} v_n = v.$$

3.1.1 Orthogonal complements

Definition. If $W \subseteq V$ is a subspace, we set

$$W^\perp = \{v \in V | (w,v) = 0 \ \text{ for all } w \in W\},$$

i.e., the set of all vectors in V orthogonal to each vector in W.

Since $(w, \cdot)$ is linear, we see that $W^\perp$ is a subspace of V. In fact

$$W^\perp = \bigcap_{w \in W} \ker\,(w, \cdot).$$

If $w \neq 0, \ker(w, \cdot)$ is a hyperplane, so the above equation represents $W^\perp$ as an intersection of hyperplanes, *cf. Proposition* 1.7.4. If $w \in V$ is a fixed vector, set

$$w^\perp = \langle w \rangle^\perp.$$

Definition. An *orthonormal basis* is a basis $\{v_1, ..., v_n\}$ where

$$(v_i, v_j) = \delta_{ij}.$$

If W is a subspace of V, we have, by *Lemma* 3.1.4.3, that $V = W \oplus W^\perp$. If $v \in V$, we can write v as $v = w + w'$, $w \in W, w' \in W^\perp$. We set $w = \text{proj}_W(v)$, the *projection* of v onto w.

Lemma 3.1.3. *An orthonormal basis exists.*

Proof. Let $0 \neq v_1 \in V$, and set

$$u = \frac{v_1}{||v_1||}.$$

Then

$$\begin{aligned} ||u_1||^2 = (u_1, u_1) &= (\frac{v_1}{||v_1||}, \frac{v_1}{||v_1||}) \\ &= \frac{1}{||v_1||^2}(v_1, v_1) \\ &= \frac{||v_1||^2}{||v_1||^2} = 1 \end{aligned}$$

Next, since $u_1 \perp$ is a hyperplane and $\langle u_1 \rangle \cap u_1 \perp = \{0\}$, we infer that $V = \langle u_1 \rangle \oplus u_1 \perp$. By induction, $(u_1)^\perp$ has an orthonormal basis $\{u_2, ..., u_n\}$; it is then obvious that $\{u_1, ..., u_n\}$ is an orthonormal basis of V.

■

The following is a rather more explicit version of the above.

3.1.4 (Gram-Schmidt Process). *Let $(v_1, ..., v_k)$ be an ordered linearly independent set. Then, there exists an ordered set $(u_1, ..., u_k)$ of orthonormal vectors such that*

$$\begin{aligned} \langle u_1 \rangle &= \langle v_1 \rangle \\ \langle u_1, u_2 \rangle &= \langle v_1, v_2 \rangle \\ &\vdots \\ \langle u_1, ..., u_k \rangle &= \langle v_1, ..., v_k \rangle. \end{aligned}$$

Proof. Set $u_1 = \frac{v_1}{||v_1||}$. Then $||u_1||^2 = 1$ and $\langle u_1 \rangle = \langle v_1 \rangle$. Set

$$u_2 = \frac{v_2 - \overline{(v_2, u_1)}u_1}{||v_2 - \overline{(v_2, u_1)}u_1||}$$

Then,

$$\begin{aligned} (u_2, u_1) &= \frac{(v_2, u_1) - (v_2, u_1)(u_1, u_1)}{||v_2 - \overline{(v_2, u_1)}u_1||} \\ &= 0 \end{aligned}$$

and so $\{u_1, u_2\}$ is orthonormal, and $\langle u_1, u_2 \rangle = \langle v_1, v_2 \rangle$.

Set

$$u_3 = \frac{v_3 - \overline{(v_3, u_2)}u_2 - \overline{(v_3, u_1)}u_1}{||v_3 - \overline{(v_3, u_2)}u_2 - \overline{(v_3, u_1)}u_1||}$$

$$(u_3, u_2) = (u_3, u_1) = 0,$$

and $||u_3||^2 = 1$. Continue in this way, obtaining vectors $u_1, u_2, \ldots, u_k$ with the stated properties.

■

Corollary 3.1.4.1. *Let $W \subseteq V$ be a subspace of the finite-dimensional complex Hilbert space, and let $(u_1, u_2, \ldots, u_k)$ be an ordered orthonormal basis of W. Then this can be extended to an ordered orthonormal basis $(u_1, u_2, \ldots, u_k, u_{k+1}, \ldots, u_n)$ of V.*

■

Corollary 3.1.4.2 (Fourier Analysis). *If $(u_1, ..., u_n)$ is an ordered orthonormal basis of V, and if $v \in V$ then the coefficients of v relative to $(u_1, ..., u_n)$ are simply inner products:*

$$v = \sum_{i=1}^{n} (u_i, v)u_i.$$

Proof. If $1 \leq j \leq n$,

$$\begin{aligned} (v - \sum_{i=1}^{n} (u_i, v)u_i, u_j) &= (v, u_j) - \sum_{i=1}^{n} \overline{(u_i, v)}(u_i, u_j) \\ &= (v, u_j) - (v, u_j) = 0. \end{aligned}$$

Thus,

$$v - \sum_{i=1}^{n}(u_i, v)u_i \in u_1^\perp \cap u_2^\perp \cap \cdots \cap u_n^\perp = V^\perp = \{0\}.$$

■

Corollary 3.1.4.3. *If $W \subseteq V$ is a subspace, then*

$$V = W \oplus W^\perp.$$

Proof. If $w \in W \cap W^\perp$, then $(w, w) = 0 \Rightarrow w = 0$. Thus $W + W^\perp = W \oplus W^\perp$. Therefore, we need only show $W + W^\perp = V$. Let $\dim W = k$, with ordered orthonormal basis $(w_1, ..., w_k)$ (possible by *Lemma* 3.1.3). For any vector $v \in V$, set $v' = \sum_{j=1}^{k}(w_j, v)w_j \in W$. Note that for any $j = 1, 2, \ldots, k$ we have

$$\begin{aligned} (v - v', w_j) &= (v, w_j) - (v', w_j) \\ &= (v, w_j) - \sum_{j=1}^{k}((v_j, v)w_j, w_i)) \\ &= (v, w_j) - \sum_{i=1}^{k}\overline{(w_i, v)}(w_j, w_i) \\ &= (v, w_j) - (v, w_j) = 0. \end{aligned}$$

Therefore, $v - v' \in W^\perp$ and so $v \in v' + W^\perp \subseteq W + W^\perp$.

■

Corollary 3.1.4.4. *If W is a subspace of V, then $W^{\perp\perp} = W$.*

Proof. If $\dim W = k$ the above shows that $\dim W^\perp = n - k$. Thus, $\dim W^{\perp\perp} = k$. Since $W \subseteq W^{\perp\perp}$, we are done.

■

Definition. If W is a subspace of V, we have, by *Corollary* 3.1.4.3, that $V = W \oplus W^\perp$. If $v \in V$, we can write v as $v = w + w'$, $w \in W, w' \in W^\perp$. We set $w = \mathrm{proj}_W(v)$, the *projection* of v onto w.

As a result of *Corollary* 3.1.4.1, the following is immediate:

Corollary 3.1.4.5. *If W is a subspace of the finite-dimensional complex Hilbert space V and if $(u_1, u_2, \ldots, u_k)$ is an ordered orthonormal basis of W, then the projection map* $\mathrm{proj}_W : V \to W$ *is given by*

$$\mathrm{proj}_W(v) = \sum_{i=1}^{k} (u_i, v)u_i.$$

3.2 Adjoints and Self-Adjoint Operators

Let V be a finite-dimensional complex Hilbert space and let $T : V \longrightarrow V$ be a linear transformation. (In this context, we shall frequently refer to T as a *linear operator.*) We have already seen in *Section* 1.7 how to define the adjoint $T^* : V^* \longrightarrow V^*$. In the present context, however, we can actually define $T^* : V \longrightarrow V$ in a very reasonable way. This makes use of the following theorem.

3.2.1 (Riesz Representation Theorem). *If V is a finite-dimensional complex inner product space, then the mapping $\varphi : V \longrightarrow V^*$ given by $\varphi(v) = (v, \cdot)$ is an isomorphism of $\mathbb{R}$-vector spaces.*

Proof. Recall that since V is a finite-dimensional complex vector space, it is, *afortiori*, a real vector space. In fact, if $\dim_{\mathbb{C}} V = n$ then $\dim_{\mathbb{R}} V = 2n$. (If $\{v_1, ..., v_n\}$ is a $\mathbb{C}$-basis, then $\{v_1, ..., v_n, iv_1, ..., iv_n\}$ is an $\mathbb{R}$-basis) Therefore, the above map

$$\varphi : V \longrightarrow V^*$$

given by $\varphi(v) = (v, \cdot)$ is $\mathbb{R}$-linear.

$$\begin{aligned}
\varphi(v_1 + v_2) &= (v_1 + v_2, \cdot) \\
&= (v_1, \cdot) + (v_2, \cdot) \\
&= \varphi(v_1) + \varphi(v_2), \\
\varphi(av) &= (av, \cdot) \\
&= a(v, \cdot) \\
&= a\varphi(v)
\end{aligned}$$

for all $v, v_1, v_2 \in V$ and $a \in \mathbb{R}$. Since

$$\begin{aligned}
\dim_{\mathbb{R}} V &= 2\dim_{\mathbb{C}} V \\
&= 2\dim_{\mathbb{C}} V^* \\
&= \dim_{\mathbb{R}} V^*,
\end{aligned}$$

it suffices, by the rank-nullity theorem to show that $\ker\varphi = \{0\}$. If $\varphi(v) = 0$, then

$$(v, \cdot) : V^* \longrightarrow \mathbb{C}$$

is the 0-functional. In particular,

$$(v, v) = 0$$

so $v = 0$.

■

Summary: The *Riesz Representation Theorem* says that any linear functional $f : V \longrightarrow \mathbb{C}$ is "represented" as an inner-product

$$f(v) = (v_f, v)$$

for some unique vector $v_f \in V$ (dependent upon f). The map $\varphi : v \mapsto (v, \cdot)$ is not a $\mathbb{C}$-linear transformation since

$$\begin{aligned} \varphi(\alpha v) &= (\alpha v, \cdot) \\ &= \overline{\alpha}(v, \cdot) \\ &= \overline{\alpha}\varphi(v). \end{aligned}$$

As a result, we often say that $\varphi : v \mapsto (v, \cdot)$ is an *anti-linear* isomorphism.

Now, let $T : V \longrightarrow V$ be a linear transformation. Recall that in *Section* 1.7, *page* 40, the *adjoint* of T, $T^* : V^* \longrightarrow V^*$ was defined by setting

$$T^*(f) = f \circ T : V \longrightarrow \mathbb{C}.$$

That is, $T^*(f)(v) = f(T(v))$, for all $v \in V$. In the present setting, we can define $T^* : V \longrightarrow V$ as follows: let $v \in V$; we have $T^*(v, \cdot) = (v, \cdot) \circ T$. If $w \in V$, then

$$\begin{aligned} (v, \cdot) \circ T(w) &= (v, \cdot)(T(w)) \\ &= (v, T(w)). \end{aligned}$$

Thus

$$T^*(v, \cdot) = (v, T(\cdot)).$$

By the *Riesz Representation Theorem* (3.2.1), there is a vector $T^*(v) \in V$ such that

$$(v, T(\cdot)) = (T^*(v), \cdot).$$

■

Proposition 3.2.2. *$T^* : V \longrightarrow V$ so defined is a linear transformation.*

Proof. If $v_1, v_2 \in V$ and if $w \in V$, we have

$$\begin{aligned}(T^*(v_1+v_2), w) &= (v_1+v_2, T(w)) \\ &= (v_1, T(w)) + (v_2, T(w)) \\ &= (T^*(v_1), w) + (T^*(v_2), w) \\ &= (T^*(v_1) + T^*(v_2), w),\end{aligned}$$

that is,

$$T^*(v_1+v_2) = T^*(v_1) + T^*(v_2).$$

Next, let $v \in V$, $\alpha \in \mathbb{F}$. If w is arbitrary,

$$\begin{aligned}(T^*(\alpha v), w) &= (\alpha v, T(w)) \\ &= \overline{\alpha}(v, T(w)) \\ &= \overline{\alpha}(T^*(v), w) \\ &= (\alpha T^*(v), w).\end{aligned}$$

As w is arbitrary,

$$T^*(\alpha v) = \alpha T^*(v).$$

■

Lemma 3.2.3. *Let $T : V \longrightarrow V$, $T^* : V \longrightarrow V$ be as above, and let $\mathcal{A} = (v_1, ..., v_n)$ be a basis of V. Define $\mathcal{A}^* = (v_1^*, v_2^*, ..., v_n^*)$ to be a basis dual to that of $\mathcal{A}$, where $v_i^* \in V$ are chosen according to the Riesz Representation Theorem:*

$$(v_i^*, v_j) = \delta_{ij}$$

If $A = T_{\mathcal{A}}$, then $A^* = T^*_{\mathcal{A}^*}$, where A^* is the "Hermitian Adjoint" of A (i.e., *if $A = [\alpha_{ij}]$, then $A^* = [\overline{\alpha_{ji}}]$.)*

Proof. We have, by definition of A:

$$T(v_j) = \sum_{i=1}^{n} \alpha_{ij} v_i$$

Write

$$T^*(v_j^*) = \sum_{i=1}^{n} \alpha_{ij}^* v_i^*;$$

We shall show that for all i and j that

$$\alpha^*_{ij} = \overline{\alpha_{ji}}.$$

We have

$$\begin{aligned}\overline{\alpha^*_{ij}} &= (T^*(v^*_j), v_i)\\ &= (v^*_j, T(v_i))\\ &= (v^*_j, \sum_{k=1}^{n} \alpha_{ki} v_k)\\ &= \sum_{k=1}^{n} \alpha_{ki}(v^*_j, v_k)\\ &= \alpha_{ji},\end{aligned}$$

which finishes the proof.

■

Remark. Note that if $(u_1, ..., u_n)$ is an orthonormal basis, then $u^*_i = u_i,\ i = 1, 2, \ldots, n$.

3.2.1 Self-adjoint operators

A linear operator $T : V \longrightarrow V$ is called *self-adjoint* or *Hermitian* if $T^* = T$. In terms of matrices, this says that if $A = T_{\mathcal{A}}$, then $A = A^* = T^*_{\mathcal{A}^*}$.

Recall that a linear transformation is diagonalizable if and only if V has a basis consisting of eigenvectors for the linear transformation.

Proposition 3.2.4. *Let $T : V \longrightarrow V$ be self-adjoint. Then T is diagonalizable. Furthermore, V has an orthonormal basis consisting of eigenvectors of T.*

Proof. Since the field $\mathbb{C}$ is algebraically closed, there exists an eigenvector $v'_1 \in V$ of T with eigenvalue $\lambda_1 \in \mathbb{C}$. If we set $v_1 = \frac{v'_1}{\|v'_1\|}$, we have that $T(v_1) = \lambda v_1$ and $\|v_1\| = 1$. Next, we have

$$V = \langle v_1 \rangle \oplus v_1^{\perp}.$$

Let $w \in v_1^{\perp}$ (so $(v_1, w) = 0$). We have the following:

$$\begin{aligned}(v_1, T(w)) &= (T^*(v_1), w)\\ &= (T(v_1), w)\\ &= (\lambda_1 v_1, w)\\ &= \overline{\lambda_1}(v_1, w)\\ &= 0.\end{aligned}$$

Thus, $w \in v_1^\perp \Rightarrow T(w) \in v_1^\perp$, *i.e.*, $V_1 = v_1^\perp$ is a T-invariant subspace of V. Since $T|_{V_1}$ is obviously self-adjoint, induction provides an orthonormal basis consisting of eigenvectors of T.

■

If T is self-adjoint, then the eigenvalues of T are all real. Indeed, if v is an eigenvector of T with eigenvalue λ, then

$$\begin{aligned}\lambda(v, v) &= (v, \lambda v)\\ &= (v, T(v))\\ &= (T(v), v)\\ &= (\lambda v, v)\\ &= \overline{\lambda}(v, v).\end{aligned}$$

Thus, $\lambda = \overline{\lambda}$. Also, if $\mathcal{U} = (u_1, ..., u_n)$ is an orthonormal basis (not necessarily one consisting of eigenvectors) and if

$$A = [\alpha_{ij}] = T_{\mathcal{U}},$$

then since $\mathcal{U} = \mathcal{U}^*$, A is Hermitian, so $a_{ij} = \overline{a_{ji}}$ for all i, j.

Suppose now that $\mathcal{U}' = (u'_1, ..., u'_n)$ be an orthonormal basis consisting of eigenvectors of T:

$$Tu'_i = \lambda_i u'_i \qquad i = 1, ..., n.$$

Consider the change-of-basis matrix $U = [\mu_{ij}]$ where

$$u'_j = \sum_{i=1}^{n} \mu_{ij} u_i, \qquad j = 1, ..., n.$$

We have seen that $U^{-1}AU = D$ where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

But

$$\begin{aligned}\delta_{ij} &= (u'_i, u'_j)\\ &= (\sum_{k=1}^{n} \mu_{ki} u_k, \sum_{l=1}^{n} \mu_{lj} u_l)\\ &= \sum_{k=1}^{n} \overline{\mu_{ki}} \sum_{l=1}^{n} \mu_{lj} (u_k, u_l)\\ &= \sum_{k=1}^{n} \overline{\mu_{ki}} \mu_{kj}\end{aligned}$$

which is the (i, j)th entry of U^*U. That is,

$$U^*U = I,$$

$$U^* = U^{-1},$$

that is to say,

$$U^*AU = D.$$

Definition. A matrix U satisfying $U^* = U^{-1}$ is called a *unitary* matrix. (In case $\mathbb{F} = \mathbb{R}$ a matrix satisfying $U^\intercal = U^{-1}$ is called an *orthogonal* matrix.)

As a result of *Proposition* 3.2.4, we have

Corollary 3.2.4.1. *If A is an $n \times n$ Hermitian matrix, then there exists a unitary matrix U satisfying $U^*AU = D$, a diagonal matrix (with real entries).*

■

3.2.2 Idempotents and orthogonal projections

Let V be a finite-dimensional complex Hilbert space. Let $W \subseteq V$ and define $P = \text{proj}_W : V \longrightarrow W$ as on *page* 73. Then $\ker P = W^\perp$, $\operatorname{im} P = W$, $P|_W = I_W$ and $W = W \oplus W^\perp$ jointly insure that P is idempotent and that $\ker P \perp \operatorname{im} P$.

Definition. Let $P : V \longrightarrow V$ be an idempotent transformation on the Hilbert Space V. We say that P is an *orthogonal projection* if and only if

$$\ker P \perp \operatorname{im} P.$$

Lemma 3.2.5. *Let P be an orthogonal projection. Then $V = \ker P \oplus im\ P$.*

Proof. $\ker P \perp \operatorname{im} P$ implies

$$\ker P \cap \operatorname{im} P = \{0\};$$

by the *Rank-Nullity Theorem* (1.4.4)

$$\dim(\ker P \oplus \operatorname{im} P) = \dim V,$$

so

$$V = \ker P \oplus \operatorname{im} P.$$

■

Lemma 3.2.6. *Let P be an idempotent on V. Then P is self-adjoint if and only if P is an orthogonal projection.*

Proof. Assume

$$P = P^*.$$

Let $v \in \ker P, w = P(v') \in \operatorname{im} P$. Then

$$\begin{aligned}(v, w) &= (v, Pv')\\ &= (Pv, v')\\ &= (0, v')\\ &= 0.\end{aligned}$$

So, $\ker P \perp \operatorname{im} P$. Conversely, assume that P is an orthogonal projection. Thus $\ker P \perp \operatorname{im} P$ and

$$V = \ker P \oplus \operatorname{im} P.$$

Let v, w be arbitrary vectors:

$$v = v_1 + v_2$$

$$w = w_1 + w_2$$

where $v_1, w_1 \in \ker P$, $v_2, w_2 \in \operatorname{im} P$. It suffices to show

$$(P^*(v), w) = (P(v), w).$$

We have

$$\begin{aligned}
(P(v), w) &= (v_2, w) \\
&= (v_2, w_1 + w_2) \\
&= (v_2, w_2) \\
&= (v_1 + v_2, w_2) \\
&= (v, w_2) \\
&= (v, P(w)) \\
&= (P^*(v), w).
\end{aligned}$$

■

Corollary 3.2.6.1. *For any subspace $W \subseteq V$, $P = \mathrm{proj}_W$ is self-adjoint.*

Recall that if $P_1, P_2 : V \longrightarrow V$ are idempotents with

$$P_1 P_2 = P_2 P_1 = 0$$

we say that P_1, P_2 are *orthogonal idempotents.*

Lemma 3.2.7. *The self-adjoint transformations P_1, P_2 are orthogonal if and only if their images are orthogonal, that is, if and only if*

$$P_1(V) \perp P_2(V).$$

Proof. For all $v_1, v_2 \in V$,

$$\begin{aligned}
(P_1(v_1), P_2(v_2)) &= (v_1, P_1^* P_2(v_2)) \\
&= (v_1, P_1 P_2(v_2)).
\end{aligned}$$

From this, the result follows easily.

■

Let $T : V \longrightarrow V$ be a self-adjoint operator. We have already seen that T is diagonalizable; let $(v_1, ..., v_n)$ be an ordered basis of V consisting of eigenvectors, and assume that $\lambda_1, ..., \lambda_k$ are the distinct eigenvalues of T. We have that

$$m_T(x) = (x - \lambda_1) \cdots (x - \lambda_k)$$

as follows: for any i, we have $(T - \lambda_j I)v_i = 0$ for some j and so

$$(T - \lambda_1 I)\cdots(T - \lambda_k I)(v_i) = 0.$$

Since any $v \in V$ is a linear combination of $v_1, ...v_n$, we conclude that

$$(T - \lambda_1 I)\cdots(T - \lambda_k I)(v) = 0,$$

and so $m_T(x)|(x - \lambda_1)\cdots(x - \lambda_k)$. But by *Lemma* 2.2.5 each $(x - \lambda_i)|m_T(x)$ so we are done.

3.2.8 (Spectral Theorem for a Self-Adjoint Operator). *Let $T : V \longrightarrow V$ be a self-adjoint operator. There exist pairwise orthogonal self-adjoint idempotents $P_1, ..., P_k$ and distinct real scalars $\lambda_1, ..., \lambda_k$ such that*

(i) $I_V = P_1 + P_2 + \cdots + P_k$;

(ii) $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_k P_k$;

(iii) $\operatorname{im} P_i = P_i(V) = \ker(T - \lambda_i I) = \lambda_i$*-eigenspace of* T*;*

(iv) $V = P_1(V) \oplus \cdots \oplus P_k(V)$.

PROOF. We have already proved all of the above except for the self-adjointness of the idempotents $P_1, P_2, \ldots, P_k$ (see (2.2.11) and (2.2.11.1)). However, since T is self-adjoint, it has *real* eigenvalues and so the minimal polynomial of T has real coefficients. This implies that each of the projections P_i is a polynomial in T *with real coefficients* and hence must also be self-adjoint. The result follows.

■

3.3 The Spectral Theorem for Normal Operators

DEFINITION. Let $T : V \longrightarrow V$, where V is a finite-dimensional complex Hilbert space. We call T *normal* if $TT^* = T^*T$.

EXAMPLES:

1. Clearly, any self-adjoint operator is normal.

2. We say that the linear operator T is *unitary* if and only if $(T(v), T(w)) = (v, w)$ for all $v, w \in V$. In this case, we see that for all $v, w \in V$,

   $$(v, w) = (T(v), T(w)) = (v, T^*T(w)).$$

   Thus,

   $$(v, w - T^*T(w)) = 0$$

   for all $v, w \in V$, and so $T^*T = I_V$, *i.e.*, $T^* = T^{-1}$. So a unitary operator is clearly normal.

3. The linear operator $T : V \to V$ is called *skew-Hermitian* if $T^* = -T$. Thus, a skew-Hermitian is normal.

Examples 2 and 3 are related by the exponential:

$$e^{\text{skew-Hermitian}} = \text{unitary}$$

(See *Exercise* 6 of *Appendix* I.) Indeed, if $T^* = -T$, then

$$\begin{aligned}
(e^T)^* &= (\sum_{k=0}^{\infty} \frac{T^k}{k!})^* \\
&= \sum_{k=0}^{\infty} \frac{1}{k!}(T^*)^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!}(-T)^k \\
&= e^{-T} \\
&= (e^T)^{-1}
\end{aligned}$$

(for justifications of some of the steps, see *Exercises* 7, 8 of *Appendix* I.)

Digression: Just as the unitary operators form a group under multiplication, the skew-Hermitian operators organize into an interesting algebraic structure called a *Lie Algebra.* Here, set

$$u_V = \{\text{skew-Hermitian operators } A : V \longrightarrow V\}.$$

Note that if $A_1, A_2 \in u_V$,

$$\begin{aligned}
(A_1 + A_2)^* &= A_1^* + A_2^* \\
&= -A_1 - A_2 \\
&= -(A_1 + A_2),
\end{aligned}$$

and if $a \in \mathbb{R}$ and $A \in u_V$,

$$\begin{aligned} (aA)^* &= aA^* \\ &= a(-A) \\ &= -(aA), \end{aligned}$$

so u_V is a real vector space. Finally, there is a multiplication called the *Lie bracket*:

$$[A_1A_2] = A_1A_1 - A_2A_1$$

$$\begin{aligned} [A_1A_2]^* &= (A_1A_2 - A_2A_1)^* \\ &= (A_1A_2)^* - (A_2A_1)^* \\ &= A_2^*A_1^* - A_1^*A_2^* \\ &= A_2A_1 - A_1A_2 \\ &= -(A_1A_2 - A_2A_1). \\ &= -[A_1A_2]. \end{aligned}$$

However, in general $(A_1A_2)^* \neq -(A_2A_1)$. Thus, u_V is closed under bracket, but not under ordinary multiplication.

The Lie bracket is *not* an associative multiplication:

$$[A_1[A_2A_3]] \neq [[A_1A_2]A_3].$$

Instead, there is the "Jacobi Identity" :

$$[A_1[A_2A_3]] + [A_2[A_3A_1]] + [A_3[A_1A_2]] = 0.$$

This being the case, we say that u_V is a *real Lie algebra.*

Lemma 3.3.1.

(a) *If $T : V \longrightarrow V$ is a self-adjoint operator with $T^l(v) = 0$ for some $l \geq 0$, then $T(v) = 0$.*

(b) *If $T : V \longrightarrow V$ is normal and $T^l(v) = 0$ for some $l \geq 0$, then $T(v) = 0$.*

PROOF. First assume that T is self-adjoint. We have seen by the *Spectral Theorem for Self-Adjoint Operators* (3.2.8) that $V = V_1 \oplus \cdots \oplus V_k$ where $V_i = \ker(T - \lambda_i I)$ and $\lambda_1, ..., \lambda_k$ are the distinct eigenvalues of T. Now, let $v \in V$ with $T^l(v) = 0$. We may write

$$v = v_1 + \cdots + v_k, \qquad v_i \in V_i,\ i = 1, ..., k.$$

Then

$$\begin{aligned} 0 &= T^l(v) \\ &= T^l(v_1) + \cdots + T^l(v_k) \\ &= \lambda_1^l v_1 + \cdots + \lambda_k^l v_k. \end{aligned}$$

This says that each $\lambda_i^l v_i = 0$, but as $\lambda_1, ..., \lambda_k$ are distinct, at most one of the λ_is can be zero. Say $\lambda_1 = 0$, and so $v_2, v_3, ..., v_k = 0$. Then $v = v_1 \in V_1$, so $T(v) = \lambda_1 v_1 = 0 \cdot v_1 = 0$, proving part (a).

More generally, if T is normal, set $S = T^*T$. Then as T and T^* commute,

$$\begin{aligned} S^l(v) &= (T^*T)^l(v) \\ &= (T^*)^l(T^l)(v) = 0. \end{aligned}$$

Note that S is self-adjoint:

$$\begin{aligned} S^* &= (T^*T)^* \\ &= T^*T^{**} \\ &= T^*T = S. \end{aligned}$$

By (a), $S(v) = 0$. Therefore,

$$\begin{aligned} (T(v), T(v)) &= (T^*T(v), v) \\ &= (S(v), v) \\ &= (0, v) \\ &= 0. \end{aligned}$$

This says that $\|T(v)\| = 0$, and so $T(v) = 0$.

■

Lemma 3.3.2. *If T is normal, then for any $v \in V$,*

$$\|T(v)\| = \|T^*(v)\|.$$

Proof.

$$\begin{aligned}\|T(v)\|^2 &= (T(v),T(v)) \\ &= (T^*T(v),v) \\ &= (TT^*(v),v) \\ &= (T^*(v),T^*(v)) \\ &= \|T^*(v)\|^2.\end{aligned}$$

■

In fact, *Lemma* 3.3.2 is an *if and only if* statement; see *Exercise* 11 of *Appendix* I.

Corollary 3.3.2.1. *If T is normal, then* $\ker T = \ker T^*$.

■

Lemma 3.3.3. *Let $P : V \longrightarrow V$ be a normal idempotent operator ($P^2 = P$). Then P is an orthogonal projection, hence is self-adjoint.*

Proof. We need only show that

$$\ker P \perp \operatorname{im} P.$$

Let $v_1 \in \ker P, v \in V, v_2 = P(v) \in \operatorname{im} P$. Then by (3.3.2.1) $v_1 \in \ker P^*$ and so

$$\begin{aligned}(v_1,v_2) &= (v_1,Pw) \\ &= (P^*v_1,w) \\ &= (0,w) = 0.\end{aligned}$$

■

3.3.4 (Spectral Theorem for a Normal Operator). *Let $T : V \longrightarrow V$ be a normal operator on a finite-dimensional complex Hilbert Space. There exist pairwise orthogonal self-adjoint idempotents $P_1, ..., P_k$ and distinct scalars $\lambda_1, ..., \lambda_k$ such that*

(i) $I_V = P_1 + P_2 + \cdots + P_k$;

(ii) $T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_k P_k$;

(iii) $\operatorname{im} P = P_i(V) = \ker(T - \lambda_i I) = \lambda_i$*-eigenspace of* T;

(iv) $V = P_1(V) \oplus \cdots \oplus P_k(V)$.

PROOF. First we show that T is diagonalizable. If $m_T(x) = (x-\lambda_1)^{e_1}(x-\lambda_2)^{e_2}\cdots(x-\lambda_k)^{e_k}$, then since $T - \lambda_i I_V$ is normal for each $i = 1, 2, \ldots, k$, we conclude by *Lemma* 3.3.1 (b), that $\ker(T-\lambda_i I)^{e_i} = \ker(T - \lambda_i I)$, which clearly implies that, in fact, $m_T(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$, and so T is diagonalizable. Thus, everything above follows, save possibly for the self-adjointness of the idempotents $P_1, P_2, \ldots, P_k$. However, as each of the idempotents P_i is a polynomial (with complex coefficients) in T, we infer immediately that each P_i is normal, and so, by *Corollary* 3.3.2.1, $\ker P_i = \ker P_i^*$, $i = 1, 2, \ldots, k$. We now shall show that $\ker P_i \perp \operatorname{im} P_i$, $i = 1, 2, \ldots, k$. Thus, set $P = P_i$ and let $v \in \ker P$, $w = P(v')$, for some $v' \in V$. Then

$$\begin{aligned}(v, w) &= (v, P(v')) \\ &= (P^*(v), v') \\ &= (0, v') \\ &= 0,\end{aligned}$$

where we have used the fact that $\ker P = \ker P^*$. From this, everything follows.

■

Corollary 3.3.4.1. *Let T be a normal operator on V. Then, there exists an ordered orthonormal basis consisting of eignevectors of T.*

PROOF. Let $(u_{i1}, u_{i2}, ..., u_{ik_i})$ be an orthonormal basis of V_i. Note that $T(u_{i1}) = \lambda_i u_{i1}$, $T(u_{i2}) = \lambda_i u_{i2}$, ..., $T(u_{ik_i}) = \lambda_i u_{ik_i}$. Since $V_i \perp V_j$, $i \neq j$, we see that

$$\begin{array}{llll}(u_{11}, & u_{12}, & \cdots, & u_{1k_1}; \\ u_{21}, & u_{22}, & \cdots, & u_{2k_2}; \\ & \vdots & \vdots\ddots & \vdots \\ u_{k1}, & u_{k2}, & \cdots, & u_{kk_k})\end{array}$$

is an orthonormal basis of V consisting of eigenvectors of T.

■

Corollary 3.3.4.2. *Let A be a normal matrix ($A^*A = AA^*$). Then there exists a unitary matrix U with*

$$U^*AU = D,$$

where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix}.$$

Corollary 3.3.4.3. *T is a normal operator if and only if V has an orthonormal basis consisting of eigenvectors of T.*

Proof. If $\mathcal{A}$ is an orthonormal basis consisting of eigenvectors, then

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix} = T_{\mathcal{A}},$$

and so, since $\mathcal{A}^* = \mathcal{A}$ (*cf. Remark* on *page* 79)

$$\begin{aligned} T^*_{\mathcal{A}} = T^*_{\mathcal{A}^*} &= \overline{T_{\mathcal{A}}}^{\mathsf{T}} \quad \text{(by (3.2.3))} \\ &= \begin{bmatrix} \overline{\lambda_1} & 0 & \cdots & 0 \\ 0 & \overline{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \overline{\lambda_k} \end{bmatrix} \end{aligned}$$

Thus $T_{\mathcal{A}}$ and $T^*_{\mathcal{A}}$ commute. Hence, so do T and T^*.

■

Chapter 4

Tensor Products

4.1 Basic Definitions

The present section is a bit more advanced than the first three chapters and will make somewhat greater demands on the reader. For example, in *Section* 4.2 below, we assume at one point that the reader has already had exposure to the notion of "invariant factors" (or "elementary divisors") of a linear transformation, as well as to the "Smith canonical form," which explicitly gives invariant factors. However, this discussion is not central to the flow of the chapter, and can safely be omitted. The principal reason, however, for introducing the tensor product construction is to obtain a less restricted version of the *Cayley-Hamilton Theorem* (2.2.12) of *Chapter* 2.

Definition. Let V, W be vector spaces over the field $\mathbb{F}$. If U is an $\mathbb{F}$-vector space, then a *bilinear map*

$$B : V \times W \longrightarrow U$$

is one satisfying

(i) $B(v_1 + v_2, w) = B(v_1, w) + B(v_2, w)$

(ii) $B(v, w_1 + w_2) = B(v, w_1) + B(v, w_2)$

(iii) $B(\alpha v, w) = B(v, \alpha w) = \alpha B(v, w)$,

for all $v, v_1, v_2 \in V,\ w, w_1, w_2 \in W$ and all $\alpha \in \mathbb{F}$. The *tensor product* $V \otimes_{\mathbb{F}} W$ is the unique (up to isomorphism) vector space such that

(i) There exists a bilinear map

$$t : V \times W \longrightarrow V \otimes_{\mathbb{F}} W$$

given by

$$t(v, w) = v \otimes w$$

(ii) Any bilinear map

$$B : V \times W \longrightarrow U$$

factors uniquely through $V \otimes_{\mathbb{F}} W$:

$$\begin{tikzcd}
& V \otimes_{\mathbb{F}} W \arrow[dr, dotted, "\exists ! T"] & \\
V \times W \arrow[ur, "t"] \arrow[rr, "B"] & & U
\end{tikzcd}$$

The first result is "categorical" in nature, but quite important.

Proposition 4.1.1. *If the tensor product of $\mathbb{F}$-vector spaces V, W exists, then it is unique up to vector space isomorphism.*

Proof. Let $V \otimes_{\mathbb{F}} W$ and $V \otimes'_{\mathbb{F}} W$ be two tensor products of V and W with bilinear maps $t : V \times W \to V \otimes_{\mathbb{F}} W$ and $t' : V \times W \to V \otimes'_{\mathbb{F}} W$. This induces the diagram

$$\begin{tikzcd}
& V \otimes_{\mathbb{F}} W \arrow[dr, dotted, "\exists ! T"] & \\
V \times W \arrow[ur, "t"] \arrow[rr, "t'"] \arrow[dr, "t"] & & V \otimes'_{\mathbb{F}} W \arrow[dl, dotted, "\exists ! T'"] \\
& V \otimes_{\mathbb{F}} W &
\end{tikzcd}$$

At the same time, we have the diagram

$$\begin{tikzcd}
& V \otimes_{\mathbb{F}} W \arrow[dr, dotted, "\exists ! S(\text{which must be the identity } I_{V \otimes_{\mathbb{F}} W})"] & \\
V \times W \arrow[ur, "t"] \arrow[rr, "t"] & & V \otimes_{\mathbb{F}} W
\end{tikzcd}$$

Therefore, in the first diagram above, we must have $T' \circ T = I_{V \otimes_{\mathbb{F}} W} : V \otimes_{\mathbb{F}} W \to V \otimes_{\mathbb{F}} W$. In an entirely similar fashion, we have $T \circ T' = I_{V \otimes'_{\mathbb{F}} W} : V \otimes'_{\mathbb{F}} W \to V \otimes'_{\mathbb{F}} W$. It follows, therefore, that

$$T : V \otimes_{\mathbb{F}} W \overset{\cong}{\to} V \otimes'_{\mathbb{F}} W.$$

■

This leaves the question of existence of the tensor product, which we resolve as follows. Given $\mathbb{F}$-vector spaces V and W, let $\mathcal{X}$ be the $\mathbb{F}$-vector space with basis $V \times W$. Notice that when the field $\mathbb{F}$ is infinite, $\mathcal{X}$ is also infinite-dimensional. Let $\mathcal{Y} \subseteq \mathcal{X}$ be the subspace of $\mathcal{X}$ generated by vectors of the form

$$(v_1 + v_2, w) - (v_1, w) - (v_2, w),$$

$$(v, w_1 + w_2) - (v, w_1) - (v, w_2),$$

$$(\alpha v, w) - (v, \alpha w),$$

where $v, v_1, v_2 \in V,\ w, w_1, w_2 \in W,\ \alpha \in \mathbb{F}$. Write $V \otimes_{\mathbb{F}} W = \mathcal{X}/\mathcal{Y}$ and write $v \otimes w = (v, w) + \mathcal{Y} \in V \otimes_{\mathbb{F}} W$. Therefore, in $V \otimes_{\mathbb{F}} W$ we have the "bilinear relations"

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w,$$

$$v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2,$$

$$\alpha v \otimes w = v \otimes \alpha w,$$

where $v, v_1, v_2 \in V,\ w, w_1, w_2 \in W,\ \alpha \in \mathbb{F}$. Furthermore, $V \otimes_{\mathbb{F}} W$ is clearly generated by all "simple tensors" $v \otimes w,\ v \in V, w \in W$.

Define the map $t : V \times W \to V \otimes_{\mathbb{F}} W$ by setting $t(v, w) = v \otimes w,\ v \in V,\ w \in W$. Then, *by construction*, t is a bilinear map. In fact we have the following result.

Proposition 4.1.2. *The $\mathbb{F}$-vector space $V \otimes_{\mathbb{F}} W$, together with the balanced map $t : V \times W \to V \otimes_{\mathbb{F}} W$ is the tensor product of V and W.*

Proof. Let U be any $\mathbb{F}$-vector space and let $B : V \times W \to U$ be a bilinear map. By the obvious infinite-dimensional version of the *Extension by Linearity Theorem* (1.4.5), there is a unique $\mathbb{F}$-linear transformation $\tau : \mathcal{X} \to U$ satisfying $\tau(v, w) = B(v, w)$. Since $B : V \times W \to U$ is bilinear, we see that $\tau|_{\mathcal{Y}} = 0$, and so τ determines a unique $\mathbb{F}$-linear transformation $T : V \otimes_{\mathbb{F}} W = \mathcal{X}/\mathcal{Y} \to U$, such that $T \circ t = B : V \times W \to U$. The result follows.

■

The following shows that dimension is well-behaved under the formation of tensor products.

Proposition 4.1.3. *If* $dim\, V = n,\ dim\, W = m$, *then*

$$dim\,(V \otimes_{\mathbb{F}} W) = mn.$$

Proof. Let $\mathcal{A} = \{v_1, ..., v_n\},\ \mathcal{B} = \{w_1, ..., w_n\}$ be bases for $V,\ W$ respectively. Set

$$\begin{aligned} \mathcal{A} \otimes \mathcal{B} \ &= \ \{v_1 \otimes w_1, v_1 \otimes w_2, ..., v_1 \otimes w_m; \\ &\quad\ v_2 \otimes w_1, v_2 \otimes w_2, ..., v_2 \otimes w_m; \\ &\qquad\qquad \vdots \\ &\quad\ v_n \otimes w_1, v_n \otimes w_2, ..., v_n \otimes w_m\}. \end{aligned}$$

We shall show that $\mathcal{A} \otimes \mathcal{B}$ is a basis of $V \otimes_{\mathbb{F}} W$. Clearly, $\mathcal{A} \otimes \mathcal{B}$ spans $V \otimes_{\mathbb{F}} W$, so $\dim V \otimes_{\mathbb{F}} W \leq mn$. We need only to prove that $\mathcal{A} \otimes \mathcal{B}$ is linearly independent. If

$$\mathcal{A}^* = \{v_1^*, v_2^*, ..., v_n^*\}$$

$$\mathcal{B}^* = \{w_1^*, w_2^*, ..., w_m^*\}$$

are the corresponding dual bases of $V,\ W$, define

$$\varphi_{ij} : V \otimes_{\mathbb{F}} W \longrightarrow \mathbb{F}$$

by the universality property

$$\begin{tikzcd} & V \otimes_{\mathbb{F}} W \arrow[dr, dotted, "\exists!\, \phi_{ij}"] & \\ V \times W \arrow[ur, "t"] \arrow[rr, "B_{ij}"] & & \mathbb{F} \end{tikzcd}$$

$$\begin{tikzcd} (v, w) \arrow[r, mapsto, "B_{ij}"] & v_i^*(v) \cdot w_j^*(w) \end{tikzcd}$$

Thus,

$$\varphi_{ij}(v \otimes w) = v_i^*(v) \cdot w_j^*(w).$$

If

$$\sum_{k,l} \alpha_{kl}(v_k \otimes w_l) = 0,$$

then

$$\begin{aligned} 0 &= \varphi_{ij}(\sum_{k,l} \alpha_{kl} v_k \otimes w_l) \\ &= \sum_{k,l} \alpha_{kl} \varphi_{ij}(v_k \otimes w_l) \\ &= \sum_{k,l} \alpha_{kl} v_i^*(v_k) \cdot w_j^*(w_l) \\ &= \alpha_{ij}. \end{aligned}$$

Since i, j were arbitrary, $\mathcal{A} \otimes \mathcal{B}$ is a linearly independent set.

4.2 Functoriality

Let $T : V \longrightarrow V'$, $S : W \longrightarrow W'$ be linear transformations.

Proposition 4.2.1. *With notation as above, there exists a unique linear transformation*

$$T \otimes S : V \otimes_{\mathbb{F}} W \longrightarrow V' \otimes_{\mathbb{F}} W'$$

satisfying

$$(T \otimes S)(v \otimes w) = T(v) \otimes S(w).$$

Proof. Construct the diagram:

$$\begin{tikzcd} & V \otimes_{\mathbb{F}} W \arrow[dr, "T\otimes S"] & \\ V \times W \arrow[ur, "t"] \arrow[rr] & & V' \otimes_{\mathbb{F}} W' \\ (v, w) \arrow[rr, mapsto] & & T(v) \otimes S(w) \end{tikzcd}$$

■

Observe that if $V_1 \subseteq V$, $W_1 \subseteq W$ are subspaces, then we can form the subspace

$$V_1 \otimes_{\mathbb{F}} W_1 \subseteq V \otimes_{\mathbb{F}} W.$$

If $T : V \longrightarrow V$, $S : W \longrightarrow W$ are linear transformations and if $V_1 \subseteq V$ is T-invariant (i.e., $T(V_1) \subseteq V_1$) and $W_1 \subseteq W$ is S-invariant, then $V_1 \otimes_{\mathbb{F}} W_1 \subseteq V \otimes_{\mathbb{F}} W$ is a $T \otimes S$-invariant subspace of $V \otimes_{\mathbb{F}} W$.

Basic Problem: Calculate the *invariant factors* (or *elementary divisors*) of $T \otimes S$ in terms of those of T and S.

Proposition 4.2.2. *Let $T : V \longrightarrow V$, $S : W \longrightarrow W$ and let $\mathcal{A}, \mathcal{B}$ be ordered bases of V and W respectively. Set*

$$A = T_{\mathcal{A}} = [\alpha_{ij}], \ B = S_{\mathcal{B}} = [\beta_{ij}].$$

Then $(T \otimes S)_{\mathcal{A}\otimes\mathcal{B}} = A \otimes B$ (Kronecker Product), defined as the matrix

$$\begin{bmatrix} \alpha_{11}B & \alpha_{12}B & \cdots & \alpha_{1n}B \\ \alpha_{21}B & \alpha_{22}B & \cdots & \alpha_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{n1}B & \alpha_{n2}B & \cdots & \alpha_{nn}B \end{bmatrix}.$$

The proof of this is routine, and will be left to the reader.

■

Example.

$$\begin{bmatrix} a & 0 \\ 1 & a \end{bmatrix} \otimes \begin{bmatrix} b & 0 \\ 1 & b \end{bmatrix} = \begin{bmatrix} ab & 0 & 0 & 0 \\ a & ab & 0 & 0 \\ b & 0 & ab & 0 \\ 1 & b & a & ab \end{bmatrix}.$$

Example. Let $\dim V = \dim W = 2$. $T : V \longrightarrow V, S : W \longrightarrow W$ and assume that T has a single invariant factor ($f_T(x) = (x-a)^2$), which is therefore equal to the minimal polynomial $m_T(x)$ of T. Thus, there exists an ordered basis $\mathcal{A}$ of V with

$$T_{\mathcal{A}} = \begin{bmatrix} a & 0 \\ 1 & a \end{bmatrix},$$

a 2×2 Jordan Block. Assume also that S has a single invariant factor $f_S(x) = (x-b)^2$. Then there exists an ordered basis $\mathcal{B}$ of W with

$$S_{\mathcal{B}} = \begin{bmatrix} b & 0 \\ 1 & b \end{bmatrix}.$$

Thus, $(T \otimes S)_{\mathcal{A}\otimes\mathcal{B}} = T_{\mathcal{A}} \otimes S_{\mathcal{B}}$ as above.

Question: What are the invariant factors of $T \otimes S$?

Recall that the invariant factors can be calculated by reducing

$$xI - (T \otimes S)_{\mathcal{A}\otimes\mathcal{B}}$$

to Smith canonical form:

$$\begin{bmatrix} x-ab & 0 & 0 & 0 \\ -a & x-ab & 0 & 0 \\ -b & 0 & x-ab & 0 \\ -1 & -b & -a & x-ab \end{bmatrix} = \begin{bmatrix} A(x) & 0 \\ B & A(x) \end{bmatrix}.$$

We shall assume that $b \neq 0$ and so B^{-1} exists; thus if "$\sim$" denotes Smith equivalence, then

$$\begin{aligned} \begin{bmatrix} A(x) & 0 \\ B & A(x) \end{bmatrix} &\sim \begin{bmatrix} 0 & -B^{-1}(A(x))^2 \\ B & A(x) \end{bmatrix} \\ &\sim \begin{bmatrix} 0 & -B^{-1}(A(x))^2 \\ B & 0 \end{bmatrix} \\ &\sim \begin{bmatrix} I & 0 \\ 0 & (A(x))^2 \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (x-ab)^2 & 0 \\ 0 & 0 & -2a(x-ab) & (x-ab)^2 \end{bmatrix} \\ &\sim \begin{cases} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (x-ab)^3 & 0 \\ 0 & 0 & 0 & (x-ab) \end{bmatrix} & \text{if } 2a \neq 0 \\ \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (x-ab)^2 & 0 \\ 0 & 0 & 0 & (x-ab)^2 \end{bmatrix} & \text{if } 2a = 0. \end{cases} \end{aligned}$$

Therefore the invariant factors of $T \otimes S$ are $(x-ab)^3,\ x-ab$ if $2a \neq 0$, and are $(x-ab)^2, (x-ab)^2$ if $2a = 0$. The reader should investigate what happens when $b = 0$, *i.e.*, when B is not invertible.

4.3 Another Application of Tensor Products

Let V be an $\mathbb{F}$-vector space and let $\mathbb{F} \subseteq \mathbb{K}$ be an extension of fields. If the extension degree $[\mathbb{K} : \mathbb{F}]$ is finite, then

$$\mathbb{K} \otimes_{\mathbb{F}} V$$

is an $\mathbb{F}$-vector space of $\mathbb{F}$-dimension equal to

$$(\dim_{\mathbb{F}} V)[\mathbb{K} : \mathbb{F}],$$

where we have used (4.1.3). But we can also think of $\mathbb{K} \otimes_{\mathbb{F}} V$ as a $\mathbb{K}$-vector space with $\mathbb{K}$-scalar multiplication determined by

$$\alpha(B \otimes v) = (\alpha B) \otimes v \qquad \alpha, B \in \mathbb{K}, v \in V$$

Proposition 4.3.1. *If $\{v_1, v_2, \ldots, v_n\}$ is an $\mathbb{F}$-basis of V, and if $\mathbb{K}$ is an extension field of $\mathbb{F}$, then $\{1 \otimes v_1, ..., 1 \otimes v_n\}$ is a $\mathbb{K}$-basis of $\mathbb{K} \otimes_{\mathbb{F}} V$.*

Proof. Clearly, $\{1 \otimes v_1, ..., 1 \otimes v_n\}$ spans $\mathbb{K} \otimes_{\mathbb{F}} V$. Let

$$\varphi_i : \mathbb{K} \times V \longrightarrow \mathbb{K}$$

be the mapping determined by

$$\varphi_i(\alpha, v) = \alpha \cdot v_i^*(v)$$

where $\{v_1^*, ..., v_n^*\}$ is the basis dual to $\{v_1, ..., v_n\}$. We have, for all $a, \alpha, \alpha_1, \alpha_2 \in \mathbb{K}$, $v, v' \in V$, that

$$\varphi_i(\alpha_1 + \alpha_2, v) = \varphi_i(\alpha_1, v) + \varphi_i(\alpha_2, v)$$

$$\begin{aligned} \varphi_i(\alpha, v + v') &= \varphi_i(\alpha, v) + \varphi_i(\alpha, v') \\ \varphi_i(a\alpha, v) &= a\alpha v_i^*(v) \\ &= a(\alpha v_i^*(v)) \\ &= a\varphi_i(\alpha, v). \end{aligned}$$

Also,

$$\begin{aligned} \varphi_i(a\alpha, v) &= a(\alpha v_i^*(v)) \\ &= \alpha(a v_i^*(v)) \\ &= \alpha \cdot v_i^*(av) \\ &= \varphi_i(\alpha, av). \end{aligned}$$

Therefore, by the universality of $\otimes_{\mathbb{F}}$ we get a unique map

$$\theta_i : \mathbb{K} \otimes_{\mathbb{F}} V \longrightarrow \mathbb{K}$$

satisfying

$$\theta_i(\alpha \otimes v) = \alpha \cdot v_i^*(v)$$

for all $\alpha \in \mathbb{K}$, $v \in V$. If

$$0 = \sum_{i=1}^{n} \alpha_i (1 \otimes v_i),$$

then, by definition of $\mathbb{K}$-scalar multiplication,

$$0 = \sum_{i=1}^{n} \alpha_i \otimes v_i,$$

so

$$\begin{aligned} 0 = \theta_j(0) &= \theta_j(\sum_{i=1}^{n} \alpha_i \otimes v_i) \\ &= \sum_{i=1}^{n} \theta_j(\alpha_i \otimes v_i) \\ &= \sum_{i=1}^{n} \alpha_i v_j^*(v_i) \\ &= \alpha_j \cdot v_j^*(v_j) = \alpha_j. \end{aligned}$$

Since j was arbitrary,

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0,$$

i.e.,

$$1 \otimes v_1, ..., 1 \otimes v_n$$

are $\mathbb{K}$-linearly independent.

■

Proposition 4.3.2. *Let V, W be $\mathbb{F}$-vector spaces and let $T : V \to W$ be an $\mathbb{F}$-linear transformation. If $\mathbb{K} \supseteq \mathbb{F}$ is an extension field, then $1 \otimes T : \mathbb{K} \otimes_{\mathbb{F}} V \to \mathbb{K} \otimes_{\mathbb{F}} W$ is a $\mathbb{K}$-linear transformation.*

Proof. Form the diagram

$$\begin{tikzcd} & \mathbb{K} \otimes_{\mathbb{F}} V \arrow[dr, dotted, "\exists!\, 1\otimes T"] & \\ \mathbb{K} \times V \arrow[ur, "t"] \arrow[rr, "\phi_T"] & & \mathbb{K} \otimes_{\mathbb{F}} W \end{tikzcd}$$

$$\varphi_T(\alpha, v) = \alpha \otimes T(v).$$

φ_T is $\mathbb{F}$-bilinear, so

$$1 \otimes T : \mathbb{K} \otimes_{\mathbb{F}} V \longrightarrow \mathbb{K} \otimes_{\mathbb{F}} W$$

exists, satisfying

$$(1 \otimes T)(\alpha \otimes v) = \alpha \otimes T(v).$$

Furthermore, if $\beta \in \mathbb{K}$,

$$\begin{aligned} 1 \otimes T(\beta(\alpha \otimes v)) &= (1 \otimes T)(\beta\alpha \otimes v) \\ &= (\beta\alpha) \otimes T(v) \\ &= \beta(\alpha \otimes T(v)) \\ &= \beta(1 \otimes T)(\alpha \otimes v), \end{aligned}$$

i.e., $1 \otimes T$ is a $\mathbb{K}$-linear transformation.

$$\mathbb{K} \otimes_{\mathbb{F}} V \longrightarrow \mathbb{K} \otimes_{\mathbb{F}} W.$$

Corollary 4.3.2.1. *Let V be an $\mathbb{F}$-vector space and let $\mathbb{K} \supseteq \mathbb{F}$ be an extension field of $\mathbb{F}$. Let $\mathcal{A} = (v_1, v_2, \ldots, v_n)$ be an ordered $\mathbb{F}$-basis of V, and set $1 \otimes \mathcal{A} = (1 \otimes v_1, \ldots, 1 \otimes v_n)$ (an ordered $\mathbb{K}$-basis of $\mathbb{K} \otimes_{\mathbb{F}} V$ by (4.3.1)). If $T : V \to V$ is an $\mathbb{F}$-linear transformation, then*

$$T_{\mathcal{A}} \;=\; (1 \otimes T)_{1\otimes\mathcal{A}}$$

as $n \times n$ matrices over $\mathbb{F}$.

Proof. We simply note that for all indices i, if

$$T(v_i) = \sum_{j=1}^{n} \alpha_{ji} v_j, \qquad (\alpha_{ji} \in \mathbb{F}),$$

then

$$\begin{aligned} (1 \otimes T)(1 \otimes v_i) &= 1 \otimes T(v_i) \\ &= 1 \otimes \sum_{j=1}^{n} \alpha_{ji} v_j \\ &= \sum_{j=1}^{n} \alpha_{ji}(1 \otimes v_j). \end{aligned}$$

The result follows.

■

Corollary 4.3.2.2. *If $0 \neq T : V \to V$ is an $\mathbb{F}$-linear transformation, and $\mathbb{K} \supseteq \mathbb{F}$ is an extension field, then $0 \neq 1 \otimes T : \mathbb{K} \otimes_{\mathbb{F}} V \to \mathbb{K} \otimes_{\mathbb{F}} V$.*

Proof. This follows immediately from 4.3.2.1.

The next result relates the characteristic polynomial of a linear transformation $T : V \to V$ to that of $1 \otimes T : \mathbb{K} \otimes_{\mathbb{F}} V \to \mathbb{K} \otimes_{\mathbb{F}} V$.

Corollary 4.3.2.3. *If $0 \neq T : V \to V$ is an $\mathbb{F}$-linear transformation, and $\mathbb{K} \supseteq \mathbb{F}$ is an extension field, then $c_T(x) = c_{1\otimes T}(x) \in \mathbb{F}[x]$.*

Proof. Indeed, by 4.3.2.1, T and $1 \otimes T$ can be represented by the same matrix.

■

4.3.3 (Cayley-Hamilton Theorem). *Let $T : V \longrightarrow V$ be a linear transformation. Then,*

$$m_T(x) | c_T(x)$$

(equivalently, $c_T(T) = 0$).

Proof. Assume that $T : V \to V$ has characteristic polynomial

$$c_T(x) = \sum_{i=0}^{n} \alpha_i x^i.$$

It suffices to prove that $\sum_{i=0}^{n} \alpha_i T^i = 0$. Let $\mathbb{K} \supseteq \mathbb{F}$ be an algebraic closure of $\mathbb{F}$.[1] By the *Restricted Cayley-Hamilton Theorem* 2.2.12, we have that $m_{1\otimes T}(x) | c_{1\otimes T}(x)$. Using (4.3.2.3), we have

$$\begin{aligned} 0 &= \sum_{i=0}^{n} \alpha_i (1 \otimes T)^i \\ &= 1 \otimes \sum_{i=0}^{n} \alpha_i T^i, \end{aligned}$$

and so $\sum_{i=0}^{n} \alpha_i T^i = 0$, where we have used (4.3.2.2).

■

[1] Here, the reader must be willing to accept the fact that any field has an algebraic closure. Actually, the proof will work using only the weaker fact that any field can be embedded into an algebraically closed field.

Chapter 5

Fourier Analysis and Quadratic Reciprocity

The scope of the present chapter is quite unusual, at least as measured against "typical" treatments of linear algebra. The goal herein is to prove the celebrated *Quadratic Reciprocity Theorem* of Gauss,[1] The statement of this seemingly inocuous result is easy enough: if p and q are distinct odd primes then the congruences

$$\begin{aligned} x^2 &\equiv q \pmod{p} \\ x^2 &\equiv p \pmod{q} \end{aligned}$$

are either both solvable or both unsolvable unless $p, q \equiv 3 \pmod 4$, in which case exactly one of the congruences is solvable. However innocent sounding this result may seem, its applications are vast. What is perhaps surprising is that a proof of this result can be developed with very little number theory: the bulk of the work will be on the shoulders of linear algebra, although the reader will need to recall some very basic modern, especially as regards the ring $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n.

5.1 Fourier Analysis on $\mathbb{Z}/n\mathbb{Z}$

Thoughout this section, G shall denote $\mathbb{Z}/nZ$, the additive group of integers modulo n. We define $L^2(G) := \{\text{functions } f : G \to \mathbb{C}\}$, which is a complex vector space relative to pointwise addition and scalar multiplication. Furthermore, it is a $\mathbb{C}$-algebra relative to the following multiplications:

(i) **Pointwise multiplication**;

[1] This theorem was Gauss' personal favorite, and he dubbed it the *aureum theorema* or the "golden theorem."

(ii) **Convolution**; this is defined by setting

$$f * g(a) = \sum_{b\in G} f(b)g(a-b).$$

It is routine to check that convolution is associative, commutative, and gives $L^2(G)$ the structure of a $\mathbb{C}$-algebra with identity given by the "point mass function" $\delta_0 : G \to \mathbb{C}, \quad \delta_0(a) = \delta_{0a}$ (Kronecker δ). (Note that relative to pointwise multiplication, the identity is the constant function having value $1 \in \mathbb{C}$.)

A Hermitian inner product is defined on $L^2(G)$ in pretty much the usual way:

$$\langle f, g\rangle \;=\; \sum_{a\in G} f(a)\overline{g(a)} \in \mathbb{C}, \qquad (f, g \in L^2(G)).$$

This inner product $\langle\cdot,\cdot\rangle$ has the properties

- $\langle\cdot,\cdot\rangle$ is linear in the first coordinate and conjugate linear in the second;
- For all $f \in L^2(\mathbb{C})$, $\langle f, f\rangle \geq 0$ (hence is real), with equality if and only if $f = 0$;

We define the *norm* of $f \in L^2(G)$ by setting

$$\|f\| \;:= \sqrt{\langle f, f\rangle}.$$

Of particular interest are the "exponential functions" (or *characters*) on G, as follows. For any residue class $[a]_n \in \mathbb{Z}/n\mathbb{Z}$, define $e_{[a]} : G \to \mathbb{C}$ by setting

$$e_{[a]}([b]) = e^{2\pi abi/n} \in \mathbb{C}, \quad ([b] \in G).$$

We shall usually drop the "residue class notation" $[\,\cdot\,]$ and write the above as

$$e_a(b) = e^{2\pi abi/n} \in \mathbb{C},$$

where we understand that a, b are representatives of corresponding elements in G. With this understanding, then the exponential functions are of the form $e_0, e_1, \ldots, e_{n-1}$. Note that e_0 is the multiplicative identity of $L^2(G)$ relative to pointwise multiplication. Note also that the characters $e_a : G \to \mathbb{C}$ are actually **homomorphisms**:

$$e_a(x+y) \;=\; e_a(x)e_a(y), \quad (x, y \in G).$$

Another set of useful functions are the *point mass functions* $\delta_g : G \to \mathbb{C}$, $g \in G$ given by

$$\delta_g(h) \ := \ \delta_{gh} \ \text{ (Kronecker } \delta).$$

Again, it will sometimes be convenient to denote these functions as $\delta_0, \delta_1, \ldots, \delta_{n-1}$. As observed above, δ_0 is the multiplicative identity of $L^2(G)$ relative to convolution.

It should be obvious that $\{\delta_0, \ldots, \delta_{n-1}\}$ is an orthonormal basis of $L^2(G)$. The same is (almost) true of $\{e_0, \ldots, e_{n-1}\}$ once we understand the

5.1.1 (Principle of Cyclotomy). *Let $\omega \in \mathbb{C}$ be an nth root of unity, i.e., $\omega^n = 1$. Then*

$$\sum_{k=0}^{n-1} \omega^k \ = \ \begin{cases} n & \text{if } \omega = 1, \\ 0 & \text{if } \omega \neq 1. \end{cases}$$

As a result, we have this:

Lemma 5.1.2. *The set of functions $\{\frac{1}{\sqrt{n}}e_0, \frac{1}{\sqrt{n}}e_1, \ldots, \frac{1}{\sqrt{n}}e_{n-1}\}$ is an orthonormal basis of $L^2(G)$.*

We now define the *Fourier transformation* $\mathcal{F} : L^2(G) \to L^2(G)$ by setting

$$\begin{aligned} (\mathcal{F}f)(a) \ &:= \ \langle f, e_a \rangle \\ &= \ \sum_{x \in G} f(x)\overline{e_a(x)} \\ &= \ \sum_{x \in G} f(x) e^{-2\pi a x i/n}. \end{aligned}$$

We shall often find it convenient to use the notation $\hat{f} := \mathcal{F}(f)$, and call $\hat{f}$ the *Fourier transform* of f.

The following very useful facts are valid for the Fourier transformation:

1. $\mathcal{F} : L^2(G) \to L^2(G)$ is a bijective linear transformation.
2. $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$, $\mathcal{F}(fg) = \mathcal{F}(f) * \mathcal{F}(g)$, (Thus $\mathcal{F}$ is an algebra isomorphism $L^2(G)_* \leftrightarrow L^2(G)_\bullet$, where $L^2(G)_*$ is the algebra with convolution as multiplication and where $L^2(G)_\bullet$ is the algebra endowed with pointwise multiplication.

3. $(\mathcal{F}^2 f)(x) = nf(-x),\ f \in L^2(G),\ x \in G.$
4. If we set $\mathcal{M} = \frac{1}{\sqrt{n}}\mathcal{F}$ (the *Plancherel transformation* of $L^2(G)$), then $\mathcal{M}$ is an *isometry* of $L^2(G)$, i.e.,
$$\langle \mathcal{M}f, \mathcal{M}g\rangle \ = \ \langle f, g\rangle, \ \text{ for all } f, g \in L^2(G).$$
5. $\mathcal{F}\delta_a = e_{-a},\ \mathcal{F}e_a = n\delta_a,\ a \in G.$

As a result of (3), (4) above, we see that $(\mathcal{M}f)(x) = f(-x)$, for all $f \in L^2(G)$ and all $x \in G$. Therefore, it follows that $\mathcal{M}^4 = 1_{L^2(G)}$ and so the eigenvalues of $\mathcal{M}$ are among the complex numbers ± 1, $\pm i$. Below, we shall compute the dimension of the corresponding eigenspaces.

From (5), we see that for $a = 0, 1, \ldots, n-1$,
$$\mathcal{F}\delta_a \ = \ \sum_{b=0}^{n-1} \zeta^{-ab}\delta_b,$$
where $\zeta = e^{2\pi i/n}$. This says that relative to the ordered basis $(\delta_0, \delta_1, \ldots, \delta_{n-1})$, $\mathcal{F}$ is represented by the matrix $V_n(\zeta^{-1}) = V_n(\overline{\zeta})$, where for an indeterminate T, $V_n(T)$ is given by
$$V_n(T) \ = \ \begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & 1 \\ 1 & T & T^2 & \cdot & \cdot & T^{n-1} \\ 1 & T^2 & T^4 & \cdot & \cdot & T^{2(n-1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & T^{n-1} & T^{2(n-1)} & \cdot & \cdot & T^{(n-1)^2} \end{bmatrix}.$$
One has the following:

Lemma.
$$\det V_n(\zeta) = \sqrt{n^n}\, i^{\binom{n}{2}} i^{n(n-1)^2}.$$

Proof. One starts by recalling the *Vandermonde* matrix
$$V_n \ = \ V_n(x_1, x_2, \ldots, x_n) \ = \ \begin{bmatrix} 1 & x_1 & x_1^2 & \cdot & \cdot & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdot & \cdot & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdot & \cdot & x_3^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 & \cdot & \cdot & x_n^{n-1} \end{bmatrix},$$

which has determinant $\det V_n = \prod_{k>j}(x_k - x_j)$. From this, it follows that

$$\det V_n(\zeta) \;=\; \det V_n(1,\zeta,\zeta^2,\ldots,\zeta^{n-1}) \;=\; \prod_{0\le j<k\le n-1}(\zeta^k-\zeta^j).$$

Next, note that if $z_1=\rho_1 e^{\theta_1 i}$, $z_2=\rho_2 e^{\theta_2 i}\in\mathbb{C}$, where ρ_1, ρ_2 are positive real numbers, then simple geometry tells us that

$$z_1+z_2=|z_1+z_2|e^{\frac{1}{2}(\theta_2+\theta_2)}.$$

Therefore, we conclude easily that for each j,k,

$$\zeta^k-\zeta^j \;=\; |\zeta^k-\zeta^j|e^{\frac{\pi i}{n}(k+j)+\frac{\pi i}{2}} \;=\; ie^{\pi(k+j)i/n}.$$

From this, it follows that

$$\prod_{0\le j<k\le n-1}(\zeta^k-\zeta^j) \;=\; \rho i^{\binom{n}{2}}e^{\frac{\pi i}{n}\sum_{0\le j<k\le n-1}(k+j)},$$

where $\rho=|\prod_{0\le j<k\le n-1}(\zeta^k-\zeta^j)|$. A routine calculation gives

$$\sum_{0\le j<k\le n-1}(k+j) \;=\; (n-1)^2 n/2.$$

Finally, note that by using 5.1.1, one has

$$(\det\ V_n(\zeta))^2 \;=\; \det\ V_n(\zeta)^2 \;=\; \det\begin{bmatrix} n & 0 & 0 & \cdot & \cdot & 0\\ 0 & 0 & 0 & \cdot & \cdot & n\\ 0 & 0 & 0 & \cdot & n & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & n & 0 & \cdot & \cdot & 0\end{bmatrix}.$$

It follows immediately from this that $|(\det V_n(\zeta))^2| \;=\; n^n$, from which it follows that $\det V_n(\zeta) =\!\sqrt{n^n}\, i^{\binom{n}{2}} i^{(n-1)^2}$.

From the above, we extract the following corollary:

COROLLARY. *The determinants of the Fourier and Plancherel transforms on $L^2(G)$, $G=\mathbb{Z}/n\mathbb{Z}$ are given by*

$$\det\mathcal{F} =\!\sqrt{n^n}\, i^{-\binom{n}{2}} i^{(n-1)^2},\quad \det\mathcal{M} = i^{-\binom{n}{2}} i^{-(n-1)^2}.$$

Note that if n is odd, then $(n-1)^2$ is divisible by 4 and so the above determinants simplify to:

$$\det\mathcal{F} = \sqrt{n^n}\, i^{-\binom{n}{2}}, \quad \det\mathcal{M} = i^{-\binom{n}{2}}.$$

5.2 Multiplicative Characters and Gauss Sums: The Case $n = p$ (Prime).

Throughout this section, G shall denote the additive group $\mathbb{Z}/p\mathbb{Z}$, where p is an odd prime. It is known in this case that the group $G^\times := G\backslash\{0\}$ is a cyclic group.[2] We fix a generator τ of $G^\times$ and define two types of *characters* of G:

Additive Character: This is a function $\pi : G \to \mathbb{C}$ such that

$$\pi(x+y) = \pi(x)\pi(y),$$

for all $x, y \in G$;

Multiplicative Character: This is a function $\chi : G \to \mathbb{C}$ such that

$$\chi(xy) = \chi(x)\chi(y), \text{ for all } x, y \in G^\times, \quad \chi(0) = 0.$$

As for examples, set $\zeta = e^{2\pi i/p}$, and for each $a = 0, 1, \ldots, p-1$, set $\pi_a(x) = \zeta^{ax}$. Then it's trivial to verify that $\pi_0, \pi_1, \ldots, \pi_{p-1}$ are additive characters of G. Note that if $\pi : G \to \mathbb{C}$ is any additive character, then as $\pi(1)$ must be a *p-th* root of unity, it follows that $\pi(1) = \zeta^a$ for some a, $0 \leq a \leq p-1$. This implies that $\pi = \pi_a$ and so we have accounted for all of the additive characters of G.

Next, set $\xi = e^{2\pi i/(p-1)}$ and define $\chi_0, \chi_1, \ldots, \chi_{p-2} : G \to \mathbb{C}$ by setting $\chi_a(0) = 0$, $\chi_a(\tau^b) = \xi^{ab}$. Again, it's trivial to verify that χ_a, $a = 0, 1, \ldots, p-2$

[2] The proof isn't too hard. First of all, one shows that if $\tau \in G^\times$ has maximal order in $G^\times$, then for all $\sigma \in G^\times$, the order of σ must divide the order of τ (else the product $\tau\sigma$ has larger order). Therefore, if we let k be the order of τ, then every element of $G^\times$ is a root of the polynomial $x^k - 1 \in \mathbb{F}[x]$, where $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$. Of course, this is a contradiction unless $k = p-1$, i.e., τ is a generator of $G^\times$, forcing $G^\times$ to be cyclic.

are multiplicative characters and account for all such. We set $r := (p-1)/2$, and note that $\chi_r(\tau^b) = e^{\pi i b} = (-1)^b$. Therefore, where $x \in G^\times$, we have

$$\chi_r(x) \;=\; \begin{cases} 1 & \text{if } x \text{ is a square in } G, \\ -1 & \text{otherwise.} \end{cases}$$

That is, for $x \neq 0$

$$\chi_r(x) \;=\; \left(\frac{x}{p}\right),$$

where $\left(\frac{x}{p}\right)$ is the familiar *Legendre symbol.*

Definition. A *Gauss Sum* is simply an inner product of a multiplicative character with an additive character. One often writes

$$G(\chi, \pi) \;:=\; \langle \chi, \pi \rangle.$$

Note that if $\pi = \pi_a,\ 0 \le a \le p-1$, then a quick computation reveals that $G(\chi, \pi_a) = \hat{\chi}(a)$.

One has that, if $x \neq 0$,

$$\begin{aligned} \widehat{\chi_a}(x) &= \sum_{y \in G^\times} \chi_a(y) e^{-2\pi i x y/p} \\ &= \sum_{y \in G^\times} \chi_a(yx^{-1}) e^{-2\pi i y/p} \\ &= \chi_a(x^{-1}) \sum_{y \in G^\times} \chi_a(y) e^{-2\pi i y/p} \\ &= \chi_a(x^{-1}) \widehat{\chi_a}(1) \\ &= \chi_{p-1-a}(x) \widehat{\chi_a}(1) \end{aligned}$$

Similarly, if $x = 0$, and if $a \neq 0$, then $\widehat{\chi_a}(x) = \sum_{y \in G^\times} \chi_a(y) = 0 = \widehat{\chi_a}(1)\chi_{p-1-a}(0)$. Therefore, we see that if $a \neq 0$, then

$$\widehat{\chi_a} \;=\; \widehat{\chi_a}(1) \chi_{p-1-a}.$$

The Principle of Cyclotomy reveals that if $\alpha = 1/\sqrt{p-1}$, then $(\delta_0, \alpha\chi_0, \ldots, \alpha\chi_{p-2})$ is an orthonormal ordered basis of $L^2(G)$. Since the above calculation shows that $\mathcal{M}(\chi_a) = \frac{1}{\sqrt{p}} \widehat{\chi_a}(1) \chi_{p-1-a},\ a \neq 0$, we infer that

$$\mathcal{B} = (\delta_0, \alpha\chi_0, \alpha\chi_1, \alpha\mathcal{M}\chi_1, \ldots, \alpha\chi_{r-1}, \alpha\mathcal{M}\chi_{r-1}, \alpha\chi_r)$$

is yet another ordered orthonormal basis of $L^2(G)$.[3] We shall compute the matrix representation of $\mathcal{M}$ relative to $\mathcal{B}$ in order to obtain another expression for $\det \mathcal{M}$.

So let's compute. Since $\mathcal{F}(\delta_0) = e_0$ (see page 106) $= \delta_0 + \chi_0$, it follows that

$$\mathcal{M}(\delta_0) \;=\; \frac{1}{\sqrt{p}}\delta_0 + \frac{\sqrt{p-1}}{\sqrt{p}}\alpha\chi_0.$$

Similarly, $\mathcal{F}(\chi_0) = \mathcal{F}(e_0 - \delta_0) = p\delta_0 - e_0$ (again, see page 106) $= (p-1)\delta_0 - \chi_0$, and so

$$\mathcal{M}(\alpha\chi_0) \;=\; \frac{\sqrt{p-1}}{\sqrt{p}}\delta_0 - \frac{1}{\sqrt{p}}\alpha\chi_0.$$

It follows, therefore, that the matrix $(\mathcal{M})_{\mathcal{B}}$ of $\mathcal{M}$ relative to $\mathcal{B}$ begins in the "northwest" with the 2×2 matrix block

$$\begin{bmatrix} \frac{1}{\sqrt{p}} & \frac{\sqrt{p-1}}{\sqrt{p}} \\ \frac{\sqrt{p-1}}{\sqrt{p}} & -\frac{1}{\sqrt{p}} \end{bmatrix};$$

note that this matrix block has determinant -1. Next, recalling that for all $f \in L^2(G)$, $\mathcal{M}^2 f(x) = f(-x)$, $x \in G$, and noting that $\chi_a(-1) = (-1)^a$, we have that $\mathcal{M}^2\chi_a(x) = \chi_a(-x) = \chi_a(-1)\chi_a(x) = (-1)^a\chi_a(x)$, from which it follows that the 2-dimensional subspace of $L^2(G)$ spanned by $\alpha\chi_a, \alpha\mathcal{M}\chi_a$ is invariant under $\mathcal{M}$; relative to the pair $(\alpha\chi_a, \alpha\mathcal{M}\chi_a)$ we get, for $a = 1, 2, \ldots, r-1$, 2×2 matrix blocks of the form

$$\begin{bmatrix} 0 & (-1)^a \\ 1 & 0 \end{bmatrix};$$

Finally, since we have already shown that $\mathcal{M}\chi_r = \frac{1}{\sqrt{p}}\widehat{\chi_r}(1)\chi_r$, we conclude that the matrix representing $\mathcal{M}$ looks like the *block diagonal direct sum*:

$$\left(\begin{bmatrix} \frac{1}{\sqrt{p}} & \frac{\sqrt{p-1}}{\sqrt{p}} \\ \frac{\sqrt{p-1}}{\sqrt{p}} & -\frac{1}{\sqrt{p}} \end{bmatrix} \bigoplus \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \bigoplus \cdots \bigoplus \begin{bmatrix} 0 & (-1)^{r-1} \\ 1 & 0 \end{bmatrix} \bigoplus [\widehat{\chi_r}(1)/\sqrt{p}]\right).$$

Since $\det \mathcal{M}$ is the product of the determinants of the above matrices, we infer that

$$\det \mathcal{M} \;=\; \begin{cases} (-1)^m\widehat{\chi_r}(1)/\sqrt{p} & \text{if } p = 4m+1, \\ (-1)^{m+1}\widehat{\chi_r}(1)/\sqrt{p} & \text{if } p = 4m+3. \end{cases}$$

[3] Since the Plancherel transformation is an isometry, it follows that $1 = |\alpha\chi_a| = |\mathcal{M}(\alpha\chi_a)| = |\frac{1}{\sqrt{p}}\widehat{\chi_a}(1)\alpha\chi_{p-1-a}| = |\frac{1}{\sqrt{p}}\widehat{\chi_a}(1)|$, which says that $|\widehat{\chi_a}(1)| = \sqrt{p}$.

Comparing this with the result on page 107, we may equate determinants and deduce that

$$\widehat{\chi_r}(1) \;=\; \begin{cases} \sqrt{p} & \text{if } p = 4m+1, \\ -i\sqrt{p} & \text{if } p = 4m+3. \end{cases}$$

From the above, we can determine the dimensions of the eigenspaces of the Plancherel transformation:

Corollary. *The eigenspaces of the Plancherel transformation $\mathcal{M} : L^2(G) \to L^2(G)$ have dimensions follows:*

$p = 4m + 1$:

$$\begin{aligned} &\dim(1 - \text{eigenspace}) = m+1, \\ &\dim((-1,\, \pm i) - \text{eigenspaces}) = m, \end{aligned}$$

$p = 4m + 3$:

$$\begin{aligned} &\dim((\pm 1,\, -i) - \text{eigenspaces}) = m+1, \\ &\dim(i - \text{eigenspace}) = m. \end{aligned}$$

5.3 The Case $n = pq$ (Distinct Primes)

We wish to try to generalize some of the preceding section to cover the composite case $n = pq$, where p and q are distinct odd primes. We begin with the following rather easy result:

Theorem 5.3.1. *(***Chinese Remainder Theorem***) Let p, q be distinct primes and let $s, t \in \mathbb{Z}$ satisfy $sp + tq = 1$. Then the mapping*

$$\mathbb{Z}/pq\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z},\ [x]_{pq} \mapsto ([x]_p, [x]_q)$$

is an isomorphism of rings, with inverse

$$([a]_p, [b]_q) \mapsto [atq + bsp]_{pq}.$$

Next, following Chapter 4, define a bilinear mapping

$$B : L^2(\mathbb{Z}/p\mathbb{Z}) \times L^2(\mathbb{Z}/q\mathbb{Z}) \to L^2(\mathbb{Z}/pq\mathbb{Z}),$$

where if $f \in L^2(\mathbb{Z}/p\mathbb{Z})$, $g \in L^2(\mathbb{Z}/p\mathbb{Z})$ then $B(f,g) \in L^2(\mathbb{Z}/pq\mathbb{Z})$ is defined by setting

$$B(f,g)([x]_{pq}) = f([x]_p)g([x]_q), \quad x \in \mathbb{Z}.$$

Proposition 5.3.2. *The above bilinear map realizes an isomorphism*

$$L^2(\mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{C}} L^2(\mathbb{Z}/q\mathbb{Z}) \cong L^2(\mathbb{Z}/pq\mathbb{Z}).$$

Proof. Note first that if $x \in \mathbb{Z}$, then the point mass function $\delta_{[x]_{pq}}$ is clearly given by $B(\delta_{[x]_p}, \delta_{[x]_q})$. It follows immediately that the induced linear mapping $L^2(\mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{C}} L^2(\mathbb{Z}/q\mathbb{Z}) \cong L^2(\mathbb{Z}/pq\mathbb{Z})$ is surjective. By 4.1.3 we know that $\dim L^2(\mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{C}} L^2(\mathbb{Z}/q\mathbb{Z}) = pq$; apply the rank-nullity theorem.

Via 5.3.2 we shall identify $L^2(\mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{C}} L^2(\mathbb{Z}/q\mathbb{Z})$ with $L^2(\mathbb{Z}/pq\mathbb{Z})$, and write $f \otimes g([x]_{pq}) = f([x]_p)g([x]_q)$, where $x \in \mathbb{Z}$. In order to simplify notation, if $x \in \mathbb{Z}$ and $f \in L^2(\mathbb{Z}/n\mathbb{Z})$ we shall write $f(x)$ in place of $f([x]_n)$. Note that if $x = atq + bsp$, then $f \otimes g(x) = f \otimes g(atq + bsp) = f(atq + bsp)g(atq + bsp) = f(a)g(b)$.

If $\chi \in L^2(\mathbb{Z}/p\mathbb{Z})$, $\psi \in L^2(\mathbb{Z}/q\mathbb{Z})$ are multiplicative characters, then the Fourier transform, $\widehat{\chi \otimes \psi}$ of $\chi \otimes \psi$ is computed below:

$$\begin{aligned}
\widehat{\chi \otimes \psi}(atq + bsp) &= \sum_{a'=0}^{p-1}\sum_{b'=0}^{q-1} (\chi \otimes \psi)(a'tq + b'sp)e^{-2\pi i(atq+bsp)(a'tq+b'sp)/pq} \\
&= \sum_{a'=0}^{p-1}\sum_{b'=0}^{q-1} \chi(a')\psi(b')e^{-2\pi i(aa'tq+bb'sp)/pq} \\
&= \sum_{a'=0}^{p-1} \chi(a')e^{-2\pi aa'ti/p} \sum_{b'=0}^{q-1} \psi(b')e^{-2\pi bb'si/q} \\
&= \sum_{a'=0}^{p-1} \chi(a't^{-1})e^{-2\pi aa'i/p} \sum_{b'=0}^{q-1} \psi(b's^{-1})e^{-2\pi bb'i/q} \\
&= \sum_{a'=0}^{p-1} \chi(a'q)e^{-2\pi aa'i/p} \sum_{b'=0}^{q-1} \psi(b'p)e^{-2\pi bb'i/q} \\
&= \chi(q)\psi(p)\widehat{\chi}(a)\widehat{\psi}(b),
\end{aligned}$$

where we have used the fact that $s^{-1} \equiv p\,(\text{mod } q)$, $t^{-1} \equiv q$, (mod p).

In the same fashion, we infer that

$$\widehat{\delta_0^{(p)} \otimes \psi} \;=\; \psi(p)\widehat{\delta_0^{(p)}} \otimes \widehat{\psi};$$

similarly,

$$\widehat{\chi \otimes \delta_0^{(q)}} \;=\; \chi(q)\widehat{\chi} \otimes \widehat{\delta_0^{(q)}}.$$

5.3.1 The Plancherel Transform on $L^2(\mathbb{Z}/pq\mathbb{Z})$

In this subsection, we shall identify $L^2(\mathbb{Z}/pq\mathbb{Z})$ with $L^2(\mathbb{Z}/q\mathbb{Z}) \otimes_{\mathbb{C}} L^2(\mathbb{Z}/q\mathbb{Z})$ and represent the Plancherel transform $\mathcal{M} : L^2(\mathbb{Z}/pq\mathbb{Z}) \to L^2(\mathbb{Z}/pq\mathbb{Z})$ as one of the form

$$\mathcal{M} = T \otimes S : L^2(\mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{C}} L^2(\mathbb{Z}/q\mathbb{Z}) \to L^2(\mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{C}} L^2(\mathbb{Z}/q\mathbb{Z}),$$

which, by *Exercise* 1 of Appendix J will allow us to compute the determinant of $\mathcal{M}$. To this end, we set $\alpha_p := 1/\sqrt{p-1}$, and let $(\delta_0^{(p)}, \alpha_p\chi_0^{(p)}, \alpha_p\chi_1^{(p)}, \ldots, \alpha_p\chi_{p-2}^{(p)})$ be the orthonormal basis of $L^2(\mathbb{Z}/p\mathbb{Z})$ given on page 109. We likewise set $\alpha_q = 1/\sqrt{q-1}$ and define the orthonormal basis $(\delta_0^{(q)}, \alpha_q\chi_0^{(q)}, \alpha_q\chi_1^{(q)}, \ldots, \alpha_q\chi_{q-2}^{(q)})$ in the obvious manner. Define the linear transformation $D_p : L^2(\mathbb{Z}/p\mathbb{Z}) \to L^2(\mathbb{Z}/p\mathbb{Z})$ by setting

$$D_p(\delta_0^{(p)}) = \delta_0^{(p)},\; D_p(\alpha_p\chi_a^{(p)}) = \chi_a^{(p)}(q)\alpha_p\chi_a^{(p)},\; a = 0, 1, \ldots, p-2.$$

Likewise define the linear transformation $D_q : L^2(\mathbb{Z}/q\mathbb{Z}) \to L^2(\mathbb{Z}/q\mathbb{Z})$ by setting

$$D_q(\delta_0^{(q)}) = \delta_0^{(q)},\; D_q(\alpha_q\chi_b^{(q)}) = \chi_b^{(q)}(p)\alpha_p\chi_a^{(p)},\; b = 0, 1, \ldots, q-2.$$

Let $\mathcal{M}_p$, $\mathcal{M}_q$ be, respectively, the Plancherel transforms on $L^2(\mathbb{Z}/p\mathbb{Z})$ and $L^2(\mathbb{Z}/q\mathbb{Z})$, respectively. We have

$$\mathcal{M}(\delta_0^{(p)} \otimes \alpha_q\chi_b^{(q)}) = \chi_b^{(q)}(p)\mathcal{M}_p(\delta_0^{(p)}) \otimes \mathcal{M}_q(\alpha_q\chi_b^{(q)}),\; b = 0, 1, \ldots, q-2,$$

$$\mathcal{M}(\alpha_p\chi_a^{(p)} \otimes \delta_0^{(q)}) = \chi_a^{(p)}(q)\mathcal{M}_p(\alpha_p\chi_a^{(p)}) \otimes \mathcal{M}_q(\delta_0^{(q)}),\; a = 0, 1, \ldots, p-1,$$

$$\mathcal{M}(\alpha_p\chi_a^{(p)} \otimes \alpha_q\chi_b^{(q)}) = \chi_a^{(p)}(q)\chi_b^{(q)}(p)\mathcal{M}_p(\alpha_p\chi_a^{(p)}) \otimes \mathcal{M}_q(\alpha_q\chi_b^{(q)}),$$
$$a = 0, 1, \ldots, p-1,\; b = 0, 1, \ldots, q-1.$$

The above calculations reveal that

$$\mathcal{M} = \mathcal{M}_p D_p \otimes \mathcal{M}_q D_q : L^2(\mathbb{Z}/pq\mathbb{Z}) \to L^2(\mathbb{Z}/pq\mathbb{Z}),$$

from which we deduce that

$$\begin{aligned} \det(\mathcal{M}) &= \det(\mathcal{M}_p D_p)^q \cdot \det(\mathcal{M}_q D_q)^p \\ &= \det(\mathcal{M}_p)\det(D_p)\det(\mathcal{M}_q)\det(D_q). \end{aligned}$$

From page 108 we have

$$\det(\mathcal{M}_p) = i^{-\binom{p}{2}}, \ \det(\mathcal{M}_q) = i^{-\binom{q}{2}}.$$

Therefore, our calculation will be complete as soon as we compute the determinants of D_p and D_q. To this end, note that for all $0 \neq x \in \mathbb{Z}/p\mathbb{Z}$, we have that $\chi_a^{(p)}(x)\chi_{p-1-a}^{(p)}(x) = 1$, $1 \leq a \leq p-2$; similarly, for all $0 \neq y \in \mathbb{Z}/p\mathbb{Z}$, we have that $\chi_b^{(q)}(y)\chi_{q-1-b}^{(q)}(y) = 1$, $1 \leq b \leq q-2$. From these observations, it follows immediately that

$$\det(D_p) = \chi_{r_p}^{(p)}(q), \quad \det(D_q) = \chi_{r_q}^{(q)}(p),$$

where $r_p = (p-1)/2$, $r_q = (q-1)/2$. Since $\chi_{r_p}^{(p)}(q) = \left(\frac{q}{p}\right)$, $\chi_{r_q}^{(q)}(p) = \left(\frac{p}{q}\right)$, we may write the final result as

$$\det(\mathcal{M}) = \left(\frac{p}{q}\right)\left(\frac{q}{p}\right) i^{-q\binom{p}{2}} i^{-p\binom{q}{2}}.$$

5.4 The Punch Line: Gauss' Quadratic Reciprocity

Since pq is odd, we may refer to page 108 and infer that $\det(\mathcal{M}) = i^{-\binom{pq}{2}}$; insert into the above equation:

$$\begin{aligned} \left(\frac{p}{q}\right)\left(\frac{q}{p}\right) &= i^{\binom{pq}{2} - q\binom{p}{2} - p\binom{q}{2}} \\ &= i^{\frac{pq(p-1)(q-1)}{2}}. \end{aligned}$$

Finally, one shows that

$$i^{\frac{pq(p-1)(q-1)}{2}} = (-1)^{\frac{p-1}{2}\frac{q-1}{2}},$$

which implies our sought-after result:

GAUSS' RECIPROCITY THEOREM. *If p and q are distinct odd primes, then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

Appendix A

Exercises

Basic Concepts:

Fields (especially $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$, the fields of rational, real and complex numbers, as well as the finite fields $\mathbb{F}_q$ of q elements), *vector spaces*, the vector space V^S, where S is an arbitrary set, *pointwise operations*, *linear span* $\langle S\rangle$ of a set $S \subseteq V$, *linear combinations* of vectors, *finitely generated vector spaces*, the vector spaces $\mathbb{F}^n$, $\mathbb{F}_n$, $\mathrm{M}_n(\mathbb{F})$, $\mathrm{M}_{mn}(\mathbb{F})$, sum and intersections of subspaces.

1. Let S be a set, $s \in S$, and let V be a vector space over the field $\mathbb{F}$. Prove that the set $W = \{f \in V^S |\ f(s) = 0\}$ is a subspace of V^S. What happens if we replace the scalar 0 above by a non-zero scalar $0 \neq \alpha \in \mathbb{F}$?

2. Prove the *modular law* for subspaces of a vector space V: If $W_1, W_2, W_3 \subseteq V$ are subspaces of V with $W_1 \supseteq W_2$, then
$$W_1 \cap (W_3 + W_2) = (W_1 \cap W_3) + W_2.$$

3. If S_1, S_2 are subsets of the vector space V, prove that $\langle S_1 \cup S_2\rangle = \langle S_1\rangle + \langle S_2\rangle$.

4. Let $\mathcal{C}^1(\mathbb{R})$ be the vector space of all differentiable real-valued funtions on the real line $\mathbb{R}$. Let $S \subseteq \mathcal{C}(\mathbb{R})$ consist entirely of *polynomial functions*, and prove that the function $f \notin \langle S\rangle$, where $f(x) = \sin x$, $x \in \mathbb{R}$.

5. Let
$$W \ = \ \{A \in \mathrm{M}_n(\mathbb{F})|\ A^t = -A\}.$$
Show that W is a subspace of $\mathrm{M}_n(\mathbb{F})$ and find a finite set of generators for W.

6. Let $\mathbb{RP}^n$ be the set of all "lines" in the vector space $\mathbb{R}_{n+1}$. (A *line* in a vector space V is a non-zero subspace generated by a single vector.) Try to find a surjective map $\rho : S^n \to \mathbb{RP}^n$, where $S^n \subseteq \mathbb{R}_{n+1}$ is the n-sphere $\{(a_0, a_1, \ldots, a_n) | \sum a_i^2 = 1\}$.

Appendix B

Exercises

Basic Concepts:

The *Exchange Lemma* (1.1.3), *linear dependence* and *linear independence* of a subset $S \subseteq V$, *basis* of a vector space, *Invariance of Dimension* (1.2.2), *Basis Extension Theorem* (1.3.1), existence of complements (1.3.2), direct sums of subspaces.

1. Let V be a vector space of dimension n over the field $\mathbb{F}$. If $S \subseteq V$ is a subset with more than n elements, then S is linearly dependent.

2. Consider the system of homogeneous equations over the field $\mathbb{F}$:

$$\begin{aligned} a_{11}\mathbf{x}_1 + a_{12}\mathbf{x}_2 + \cdots + a_{1n}\mathbf{x}_n &= 0 \\ a_{21}\mathbf{x}_1 + a_{22}\mathbf{x}_2 + \cdots + a_{2n}\mathbf{x}_n &= 0 \\ &\cdot\ \cdot\ \cdot \\ a_{m1}\mathbf{x}_1 + a_{m2}\mathbf{x}_2 + \cdots + a_{mn}\mathbf{x}_n &= 0, \end{aligned}$$

   where $m < n$. Prove that there is a non-trivial solution to this system, i.e., there exist numbers $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$, not all 0, such that $\mathbf{x}_i = \alpha_i,\ i = 1, 2, \ldots, n$ is a solution to the above system. (Hint: this follows immediately from *Exercise* 1.)

3. Let V be a finitely-generated vector space, and assume that $V = \langle S \rangle$. Prove that there exists a *finite* subset $\{s_1, s_2, \ldots, s_n\} \subseteq S$ such that $V = \langle s_1, s_2, \ldots, s_n \rangle$.

4. Let V be a finitely-generated vector space with $V = \langle S \rangle$. Prove that if $\{v_1, v_2, \ldots, v_k\}$ is a linearly independent subset of V, then there exists a finite subset $\{s_1, s_2, \ldots, s_h\} \subseteq S$ such that $\{v_1, v_2, \ldots, v_k, s_1, s_2, \ldots s_h\}$ is a basis of V.

5. Let $\dim V \geq 2$. Prove that V has more than one basis.

6. * Let V be an n-dimensional vector space over the field $\mathbb{F}_q$ of q elements. How many distinct bases does V have? (If $n = 2$ the correct answer is $\frac{1}{2}q(q-1)(q^2-1)$.)

7. Compute the dimensions of $\mathbb{F}_n$, $\mathbb{F}^n$, $\mathrm{M}_{mn}(\mathbb{F})$.

8. Let V be a vector space and assume that we can write V as a direct sum of subspaces: $V = V_1 \oplus V_2$. If $W \subseteq V$ is a subspace, either prove that $W = (W \cap V_1) \oplus (W \cap V_2)$ or give a counter-example.

9. Let $V = W_1 + W_2 + \cdots + W_k$, where $W_1, W_2, \ldots, W_k$ are subspaces of V. Show that this sum is direct iff for each $i = 1, 2, ..., k$, $W_i \cap (W_1 + W_2 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_n) = \{0\}$.

Appendix C

Exercises

Basic Concepts:

Linear Transformation, *kernel*, *null-space*, *nullity*, *rank*, *injective linear transformation*, *surjective linear transformation*, *Rank-Nullity Theorem* (1.4.4), *invertible linear transformation*, *inverse of a linear transformation*, *isomorphism*, *Extension-by-Linearity Theorem* (1.3.1), *ordered basis*, *matrix representations*, *the "Representation Picture"* (1.5.2).

1. Let $T : V_1 \to V_2$ be a linear transformation and let $W_1,\ W_1' \subseteq V_1$ be subspaces of V_1.

   (a) Prove that $T(W_1 + W_1') = T(W_1) + T(W_1')$.

   (b) Is it necessarily true that $T(W_1 \cap W_1') = T(W_1) \cap T(W_1')$? What's the most you can say here?

2. Let $T : V_1 \to V_2$ be an injective linear transformation. Prove that $\dim V_1 \leq \dim V_2$.

3. Let $T : V_1 \to V_2$ be a surjective linear transformation. Prove that $\dim V_1 \geq \dim V_2$.

4. Let V_1, V_2 be finite dimensional vector spaces and let $W_1 \subseteq V_1$, $W_2 \subseteq V_2$ be *isomorphic* subspaces, via a linear transformation $S : W_1 \overset{\cong}{\to} W_2$. Prove that there exists a linear transformation $T : V_1 \to V_2$ such that $T|_{W_1} = S$.

5. Using matrices, find examples as called for below:

   (a) Find a linear transformation $T : V \to V$ such that $\ker T \neq 0$ but that T is not surjective.

(b) Find a linear transformation $T : V \to V$ such that $T \neq I_V$, but $T^2 = T$.

(c) Find a linear transformation $T : V \to V$ such that $T \neq 0_V$, but $T^2 = 0_V$.

(d) Find a linear transformation $T : V \to V$ such that $T^k \neq 0_V$, but $T^{k+1} = 0_V$.

(e) Find a linear transformation $T : V \to V$ such that $T^4 = I_V$, but $T^k \neq I_V$ if $0 < k < 4$.

6. Let $\mathbb{C}$ be the field of complex numbers, regarded as a vector space of dimension 2 over $\mathbb{R}$. Let $T : \mathbb{C} \to \mathbb{C}$ be defined by $T(z) = (2 - i)z,\ z \in \mathbb{C}$. Prove that T is an $\mathbb{R}$-linear transformation, and compute $T_{\mathcal{A}}$, where $\mathcal{A} = (1, i)$.

7. Let $V = \mathrm{M}_2(\mathbb{F})$, and define the matrix A by setting

$$A = \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix}.$$

Now define $T : V \to V$ via $T(B) = AB,\ B \in V$. Compute $T_{\mathcal{A}}$ relative to the ordered basis $\mathcal{A} = (e_{11}, e_{12}, e_{21}, e_{22})$ ($e_{ij} \in \mathrm{M}_2(\mathbb{F})$ is defined to be the 2×2 matrix with 1 in the i, j-position, and 0 everywhere else). Is T an isomorphism?

Appendix D

Exercises

Basic Concepts:

Cosets relative to a subspace, *quotient spaces*, *well-defined mapping*, *Fundamental Homomorphism Theorem* (1.6.4), *Correspondence Theorem* (1.6.5).

1. Let $T : V \to W$ be a linear transformation. Assume that $V' \subseteq V$, $W' \subseteq W$ are subspaces and assume that $T(V') \subseteq W'$. Prove that the recipe $\bar{T} : V/V' \to W/W'$ given by $\bar{T}(v + V') = T(v) + W'$ is a well-defined linear transformation.

2. Prove the *Noether Isomorphism Theorem:* If V is an $\mathbb{F}$-vector space and if $U, W \subseteq V$ are subspaces, then $(U + W)/W \cong W/(U \cap W)$.

3. Let $T : V \longrightarrow V$ be a linear transformation. Define the T-*commutator subspace* of V by setting $[T, V] = \{T(v) - v | v \in V\}$. Show first that, in fact, $[T, V]$ is a subspace of V, and then show that if $W = [T, V]$, then $T(W) \subseteq W$ and the linear transformation $\overline{T} : V/W \longrightarrow V/W$ defined as in Problem 1 above satisfies $\overline{T}(v + W) = v + W$ for all $v \in V$.

4. Let $W' \subseteq W \subseteq V$ (subspaces); prove that $(V/W')/(W/W') \cong V/W$.

5. Let $V \xrightarrow{T} W \xrightarrow{S} U$ be a sequence of linear transformations. We say that the sequence is *exact* (at W) if $T(V) = \ker S$.

   (a) Show that the linear transformation $T : V \to W$ is injective if and only if the sequence $\{0\} \to V \xrightarrow{T} W$ is exact.

   (b) Show that the linear transformation $T : V \to W$ is surjective if and only if the sequence $V \xrightarrow{T} W \to \{0\}$ is exact.

(c) Use the *Fundamental Homomorphism Theorem* (1.6.4) to show that if the sequence $\{0\} \to V' \xrightarrow{T} V \xrightarrow{S} V'' \to \{0\}$ is exact (at all possible places), then $V'' \cong V/T(V')$.

6. Let

$$\{0\} \to V_1 \xrightarrow{T_1} V_2 \xrightarrow{T_2} V_3 \xrightarrow{T_3} V_4 \to \{0\}$$

be an exact sequence of finite-dimensional vector spaces. Prove that $\dim V_1 - \dim V_2 + \dim V_3 - \dim V_4 = 0$. Can you generalize this?

7. Let $T : V \to V$ be a linear transformation and let W be a T-*invariant* subspace of V. (This means that $T(W) \subseteq W$.) Assume that $\mathcal{A} = (v_1, \ldots, v_k, v_{k+1}, \ldots, v_n)$ is an ordered basis of V, where $(v_1, \ldots, v_k)$ is an ordered basis of W. Assume that

$$T_{\mathcal{A}} = \begin{bmatrix} A & B \\ O & D \end{bmatrix},$$

where $A \in \mathrm{M}_k(\mathbb{F})$, $B \in \mathrm{M}_{k,n-k}(\mathbb{F}), O \in \mathrm{M}_{n-k,k}(\mathbb{F})$ (O is the $(n-k) \times k$ 0-matrix), $D \in \mathrm{M}_{n-k}(\mathbb{F})$. Relative to the ordered basis $\mathcal{B} = (v_{k+1} + W, \ldots, v_n + W)$, compute $\bar{T}_{\mathcal{B}}$, where $\bar{T} : V/W \to V/W$ is defined as in *Exercise* 1.

Appendix E

Exercises

Basic Concepts:

$L(V, V')$ as a subspace of V'^V, *linear functionals*, *dual space*, *dual basis*, *hyperplanes*, Ann $W \subseteq V^*$ (the *annihilator* in V^* of the subspace $W \subseteq V$), *adjoint* of a linear transformation, the *double dual* V^{**}, the *natural* injection $V \to V^{**}$, the *unnatural* isomorphism $V \cong V^*$ (dim $V < \infty$).

1. Let V be a finite-dimensional vector space, and let $v_1 \neq v_2$ in V. Prove that there exists $f \in V^*$ with $f(v_1) \neq f(v_2)$.

2. Let V be a vector space and let $f, g \in V^*$ such that $f(v) = 0$ if and only if $g(v) = 0$. Prove that $f = \lambda g$ for some $\lambda \in \mathbb{F}$.

3. Let V be a vector space and let $H \subseteq V$ be a hyperplane in V. If $v \in V$, $v \notin H$ is a fixed vector, prove that there exists a unique $f \in V^*$ whose kernel contains H and which satisfies $f(v) = 1$.

4. Let V be a vector space and let $W, H \subseteq V$ be subspaces of V, where H is a hyperplane. Prove that dim $W - 1 \leq$ dim $(W \cap H) \leq$ dim W. Conclude that the intersection of k hyperplanes in V has dimension $\geq n - k$.

5. In what sense can the solution set of the system of homogeneous equations

$$\begin{array}{rcl} a_{11}\mathbf{x}_1 + a_{12}\mathbf{x}_2 + \cdots + a_{1n}\mathbf{x}_n & = & 0 \\ a_{21}\mathbf{x}_1 + a_{22}\mathbf{x}_2 + \cdots + a_{2n}\mathbf{x}_n & = & 0 \\ & \cdot\ \cdot\ \cdot & \\ a_{m1}\mathbf{x}_1 + a_{m2}\mathbf{x}_2 + \cdots + a_{mn}\mathbf{x}_n & = & 0, \end{array}$$

be thought of as the intersection of hyperplanes in $\mathbb{F}^n$?

6. Describe a natural isomorphism $\mathbb{F}_n \to (\mathbb{F}^n)^*$.

7. Let $V = \mathcal{C}([0,1])$, the $\mathbb{R}$-vector space of real-valued continuous functions on the interval $[0,1]$. Define the map

$$\phi : V \longrightarrow V^*, \ \phi(f)(g) \ = \ \int_0^1 f(x)g(x)dx.$$

Show that ϕ is a linear transformation whose kernel is trivial.

8. As we have seen, any two vector spaces V, W of the same dimension are isomorphic. However, such an isomorphism generally depends on a choice of bases in V and W, respectively. Let us say that vector spaces V and W are *naturally* (or *canonically*) isomorphic if there there is an isomorphism $T : V \to W$ that doesn't depend on any choice of bases. Prove that the following are natural isomorphisms:

   (a) $V \cong V^{**}$ ($\dim V < \infty$).

   (b) $(V/W)^* \cong \mathrm{Ann}(W)$, where V is a vector space and W is a subspace of V.

   (c) The *codual* of a vector space. Let V be a vector space over the field $\mathbb{F}$, and set $V_* = L(\mathbb{F}, V)$. Define the map $\epsilon : V \to V_*$ by setting $\epsilon(v)(\alpha) = \alpha \cdot v \in V$.

9. Let V be a finite-dimensional vector space with dual space V^*. If W_1, W_2 are subspaces of V, show that

   (a) $\mathrm{Ann}(W_1 \cap W_2) = \mathrm{Ann}(W_1) + \mathrm{Ann}(W_2)$.

   (b) $\mathrm{Ann}(W_1 + W_2) = \mathrm{Ann}(W_1) \cap \mathrm{Ann}(W_2)$.

   (Hint: Use the fact that $\mathrm{Ann} : \{\text{subspaces of } V\} \to \{\text{subspaces of } V^*\}$ is an inclusion-reversing bijection.)

10. Let $V = M_n(\mathbb{F})$, an n^2-dimensional vector space. Define the *trace* of $A \in V$ by

$$\tau(A) = \sum_{i=1}^{n} a_{ii} \in \mathbb{F}$$

where $A = [a_{ij}]$. Note that τ is a linear transformation and $\tau \in V^*$. Call τ the *trace functional.* Prove the following:

   (a) $\tau(AB) = \tau(BA) \qquad$ for all $A, B \in V$.

(b) Assume that $f \in V^*$ satisfies

$$f(AB) = f(BA) \qquad \text{for all } A, B \in V.$$

Show that $f = \alpha\tau$ for some $\alpha \in \mathbb{F}$.

11. Let $T : V \to V'$ be a linear transformation with adjoint $T^* : V'^* \to V^*$. Prove

(a) T is surjective if and only if T^* is injective;

(b) T is injective if and only if T^* is surjective.

(This exercise is not entirely trivial. You might wish to restrict your attention to the case in which both V, V' are finite-dimensional.)

Appendix F

Exercises

Basic Concepts:

Eigenvectors, eigenvalues, characteristic polynomial, characteristic equation, determinant of a linear transformation $T : V \to V$, diagonalizable linear transformation, real matrix exponential.

1. Let $T : V \to V$ be a linear transformation, and assume that $v_1, v_2, \ldots, v_k$ are eigenvectors of T, corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Prove that if $\lambda_i \neq \lambda_j$ for all $i \neq j$, then $v_1, v_2, \ldots, v_k$ are linearly independent.

2. Prove that $T : V \to V$ is a linear transformation such that the characteristic polynomial $c_T(x)$ splits into distinct linear factors, then T is diagonalizable.

3. Give an example of a linear transformation on a *real* vector space which has no real eigenvalues.

4. If we regard $\mathbb{C}$ as a 2-dimensional real vector space and if $T(z) = (2i - 3)z$, does T have real eigenvalues? In general, if $a \in \mathbb{C}$ and $T : \mathbb{C} \to \mathbb{C}$ is given by $T(z) = az$, what can you say about a in order that T have real eigenvalues?

5. Let V be a two-dimensional vector space over the real field $\mathbb{R}$, and let $T : V \to V$ be a linear transformation. Show that if the characteristic polynomial of T has a complex zero $a + bi$, then there exists an ordered basis $\mathcal{A}$ in V such that

   $$T_{\mathcal{A}} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

   (This takes a little work.)

6. Let V be a finite-dimensional vector space and let $T : V \longrightarrow V$ be a linear transformation on V. Show that 1 is an eigenvalue of T if and only if $[T, V] \neq V$, where $[T, V]$ is the T-commutator subspace of V (see *Exercise* 3 of Exercise Batch D).

7. Using the matrix exponential, find the general solution of the system of first order ODE

$$\begin{aligned} x_1'(t) &= 4x_1(t) - 3x_2(t) \\ x_2'(t) &= 6x_1(t) - 5x_2(t). \end{aligned}$$

8. Suitably modify your arguments above to find the general solution of

$$\begin{aligned} x_1'(t) &= -3x_1(t) + 2x_2(t) \\ x_2'(t) &= -8x_1(t) + 5x_2(t). \end{aligned}$$

9. Solve the initial-value problem

$$\mathbf{x}'(t) = \begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix} \mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}.$$

10. Recall that the equation of motion for the simple harmonic oscillator (drawn below) is given by *Hooke's Law*: $mx''(t) = -kx(t)$, where m is the mass of the body, and k is the spring constant.

```
/|
/|      k
/|----oooooo---[]
/|             m
/|
```

Now consider the coupled harmonic oscillator, depicted below.

```
/|                                               |\
/|     k_1            k_2             k_3        |\
/|----oooooo---[]---oooooo---[]---oooooo----|\
/|             m_1            m_2                |\
/|                                               |\
```

Show that the equations of motion are

$$\begin{aligned} m_1x_1''(t) &= -(k_1+k_2)x_1(t)+k_2x_2(t) \\ m_2x_2''(t) &= k_2x_1(t)-(k_2+k_3)x_2(t). \end{aligned}$$

Let us now assume, for simplicity's sake that $m_1 = m_2 = 1$ and that $k_1 = k_3$. The equations of motion, in matrix form, can now be written

$$\mathbf{x}''(t) = A\mathbf{x}(t), \quad A = \begin{bmatrix} -(k_1+k_2) & k_2 \\ k_2 & -(k_1+k_2) \end{bmatrix}.$$

Using the eigenvectors of A, we can "decouple" the above system, as follows. Let

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

be the change-of-basis matrix whose columns are eigenvectors of A, with corresponding eigenvalues λ_1, λ_2. Now introduce new variables $y_1(t), y_2(t)$ by the matrix equation

$$\mathbf{x}(t) = P\mathbf{y}(t), \quad \mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}.$$

Show that relative to the new coordinate system, the second-order system becomes uncoupled:

$$\begin{aligned} y_1''(t) &= \lambda_1 y_1(t) \\ y_2''(t) &= \lambda_2 y_2(t). \end{aligned}$$

Solve this and thereby obtain the general solution of the original second-order system of ODE.

Finally, note that if initial values are given along eigenvectors of the matrix A, then the resulting motion is *purely sinusoidal*, and its frequency is a function of the corresponding eigenvalue. Carry out everything in detail, obtaining answers in terms of the spring constants k_1 and k_2.

Appendix G

Exercises

Basic Concepts:

Polynomials, division algorithm (long division) (2.2.1), *greatest common divisor, least common multiple, relatively prime polynomials, the "Euclidean trick"* (2.2.3), *minimal polynomial, Primary Decomposition Theorem* (2.2.10), *Cayley-Hamilton Theorem* (2.2.12).

1. Give an example of a linear transformation $T : V \to V$, where V is 2-dimensional over the real field $\mathbb{R}$, and where $m_T(x)$ is irreducible of degree 2.

2. Let $T : V \to V$ be a linear transformation on the 2-dimensional vector space V over the field $\mathbb{F}$. If $m_T(x) = (x - a)^2$, $a \in \mathbb{F}$, what (if anything) can you say about T? Can you give an example of such a linear transformation?

3. Let $T : V \to V$ be a linear transformation, and let $W \subseteq V$ be a T-invariant subspace. Prove that if $T|_W$ denotes the restriction of T to the subspace W, then $m_{T|_W}(x)|m_T(x)$.

4. Let $T : V \to V$ be a linear transformation, and assume that $V = V_1 \oplus V_2$, where V_1, V_2 are T-invariant subspaces. Show that if $m_{T|_{V_1}}(x)$, $m_{T|_{V_2}}(x)$ are relatively prime, then $m_T(x) = m_{T|_{V_1}}(x)m_{T|_{V_2}}(x)$.

5. Assume that $T : V \to V$ is a linear transformation. Show that T is diagonalizable if and only if $m_T(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_k)$ for *pairwise distinct* $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$. (*Cf. Exercise Appendix* F, *Exercise* 1.)

6. Prove that if $T : V \to V$ is a linear transformation, then T is invertible if and only if 0 is not a zero of $m_T(x)$. (Hint: consider the equation

$T^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0I = 0$. Then $a_0 \neq 0$ and $T(T^{n-1} + a_{n-1}T^{n-2} + \cdots a_1) = -a_0I$. So what?)

7. Suppose $T : V \to V$ is a linear transformation on the vector space V over the real field $\mathbb{R}$. If $T^2 = T - I$, what can you say about the minimal polynomial of T? Can T be diagonalizable?

8. Compute $m_A(x)$ for each matrix below:

   (a) $A = \begin{bmatrix} \alpha & \beta \\ 0 & \alpha \end{bmatrix}$.

   (b) $A = \begin{bmatrix} \alpha_1 & \beta \\ 0 & \alpha_2 \end{bmatrix}$. $\alpha_1 \neq \alpha_2$

   (c) $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

   (d) $A = \begin{bmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & . \\ 0 & 0 & 0 & . & . \\ . & . & . & . & . \\ . & . & . & 0 & 1 \\ . & . & . & 0 & 0 \end{bmatrix}$.

   (e) $A = \begin{bmatrix} \alpha & 1 & 0 & . & 0 \\ 0 & \alpha & 1 & . & . \\ 0 & 0 & \alpha & . & . \\ . & . & . & . & 1 \\ . & . & . & 0 & \alpha \end{bmatrix}$.

   (f) $A = \begin{bmatrix} 0 & 0 & . & . & -a_0 \\ 1 & 0 & . & . & . \\ 0 & 1 & . & . & . \\ . & . & . & . & . \\ . & . & . & 0 & -a_{n-2} \\ . & . & . & 1 & -a_{n-1} \end{bmatrix}$.

9. Let $T : V \to V$ be a linear transformation and let $m_T(x) = p_1(x)^{e_1}p_2(x)^{e_2} \cdots p_k(x)^{e_k}$ be its minimal polynomial, where the factors $p_i(x)$, $i = 1, 2, \ldots, k$ are distinct, monic, irreducible polynomials.

   (a) Let $V_i = \ker\, p_i(T)^{e_i}$, $i = 1, 2, \ldots k$. Show that if $q_i(x) = m_T(x)/p_i(x)^{e_i}$ (so $p_i(x)^{e_i}$ and $q_i(x)$ are relatively prime), and if $s(x), t(x) \in \mathbb{F}[x]$ satisfy $s(x)p_i(x)^{e_i} + t(x)q_i(x) = 1$, then for all $v_i \in V_i$, $t(T)q_i(T)(v_i) = v_i$.

(b) Let $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$, where $V_i = \ker\, p_i(T)^{e_i},\ i = 1, 2, \ldots, k$ as in the *Primary Decomposition Theorem.* Prove that if $W \subseteq V$ is a T-invariant subspace, then $W = (W \cap V_1) \oplus (W \cap V_2) \oplus \cdots \oplus (W \cap V_k)$. (Warning: the T-invariance of W is crucial!)

Appendix H

Exercises

Basic Concepts:

Hermitian inner product, Hilbert Space, norm of a vector, *triangle inequality* (3.1.2), *standard inner product* on $\mathbb{C}^n$, $\mathbb{C}_n$, *orthonormal* basis, *Gram-Schmidt process* (3.1.4), *Grammian* of an inner product (relative to an ordered basis), *orthogonal direct sum* of subspaces, *Cauchy-Schwarz inequality* (3.1.1), *Fourier analysis* (3.1.4.2). Unless otherwise stated, in the exercises below, V denotes a finite-dimensional complex Hilbert space, with inner product denoted $(\cdot\,,\cdot)$.

1. Let $v_1, v_2 \in V$. Prove that if v_1 and v_2 are orthogonal, then $\|v_1+v_2\|^2 = \|v_1\|^2 + \|v_2\|^2$. Is this an if and only if statement?

2. Let $\{u_1, u_2, \ldots, u_k\}$ be an orthonormal set of vectors in V. Prove that $u_1, u_2, \ldots, u_k$ are linearly independent.

3. The norm $\|\cdot\|$ on the inner product space V can be used to define a *metric space* structure on V, with distance function

$$d(v, w) \;=\; \|v - w\|.$$

Show that the Cauchy-Schwarz inequality guarantees that the mapping

$$(\cdot\,,\cdot) : V \times V \to \mathbb{C}$$

is a *continuous* complex-valued function on V.

4. Let $\{u_1, u_2, \ldots, u_k\}$ be an orthonormal set of vectors in V. Let $v \in V$, and set $x_i = (v, u_i)$.

   (a) Prove Bessel's inequality: $\sum_{i=1}^{k} |x_i|^2 \leq \|v\|^2$.

(b) Prove Parseval's equation: $\sum_{i=1}^{k} x_i u_i = v$ if and only if $\sum_{i=1}^{k} |x_i|^2 = \|v\|^2$.

5. Define an inner product on $\mathbb{C}_2$ by setting

$$B((\alpha_1, \alpha_2), (\beta_1, \beta_2)) = \bar{\alpha_1}\beta_2 + \bar{\alpha_2}\beta_1.$$

Does this inner product give $\mathbb{C}_2$ a Hilbert space structure?

6. Let $v, w \in V$. Prove that $|(v, w)| = \|v\| \cdot \|w\|$ if and only if v and w are linearly dependent.

7. Let $v_1, v_2, \ldots, v_k \in V$. Show that $v_1^\perp \cap v_2^\perp \cap \cdots \cap v_k^\perp = \langle v_1, v_2, \ldots, v_k \rangle^\perp$.

8. Let V be a Hilbert space, let $\mathcal{A} = (v_1, v_2, \ldots, v_n)$ be an ordered basis, and let $A = [\alpha_{ij}]$ be the so-called *Grammian matrix* of $\mathcal{A}$ ($\alpha_{ij} = (v_i, v_j)$). Show that if $\mathcal{B} = (w_1, w_2, \ldots, w_n)$ is another ordered basis, with change-of-basis matrix $P = [p_{ij}]$, where $v_j = \sum_{i=1}^{n} p_{ij} w_i$, then the Grammian matrix of $\mathcal{B}$ is given by $P^t A P$.

9. Let $v_1, v_2, \ldots, v_k \in V$, and define the matrix $A = [\alpha_{ij}]$ by setting $\alpha_{ij} = (v_i, v_j)$. Show that $v_1, v_2, \ldots, v_k$ are linearly independent if and only if $\det A \neq 0$. (Hint: To prove $\Rightarrow$, use Gram-Schmidt to obtain an ordered orthonormal basis $(u_1, u_2, \ldots u_k)$ for $\langle v_1, v_2, \ldots, v_k \rangle$ and then apply *Exercise* 8, above. For $\Leftarrow$, if $\sum_{i=0}^{k} \alpha_i v_i = 0$, let $X = (\alpha_1, \alpha_2, \ldots, \alpha_k)^t$ and observe that $AX = 0$.)

10. Let $W \subseteq V$ be a subspace of V and let $v \in V$. Prove that $w_0 \in W$ satisfies $\|v - w_0\| \leq \|v - w\|$ for all $w \in W$ if and only if $v - w_0 \in W^\perp$. (Hint: To prove $\Rightarrow$, let $w \in W$ be arbitrary and form the quadratic $q(t) = \|v - w_0 + tw\|^2$, where $t \in \mathbb{R}$. Then $q(t) \geq \|v - w_0\|^2$ together with a discriminant argument reveals that $\text{Re}\,\langle v - w_0, w \rangle = 0$. Since $w \in W$ is arbitrary, this implies that $\langle v - w_0, w \rangle = 0$ for all $w \in W$. The converse is much easier: if $w \in W$, use *Exercise* 1 to get $\|v - w\|^2 = \|v - w_0 + w_0 - w\|^2 = \|v - w_0\|^2 + \|w_0 - w\|^2 \geq \|v - w_0\|^2$.) As a result, conclude that if $v \in V$, $w_0 = \text{proj}_W(v)$, then $\|v - w_0\| \geq \|v - w\|$ for all $w \in W$.

11. Show that if $W \subseteq V$ and $\{w_1, ..., w_k\}$ is an orthonormal basis of W, then

$$\text{proj}_W(v) = \sum_{i=1}^{k} (w_i, v) w_i,$$

for every $v \in V$.

Appendix I

Exercises

Basic Concepts:

Riesz Representation Theorem (3.2.1), *adjoint* of a linear transformation, *self-adjoint, unitary* and *normal* operators, *orthogonal projections*, *Spectral Theorem* (3.2.8), (3.3.4).

Unless otherwise stated, V shall denote a finite-dimensional complex Hilbert space, with Hermitian inner product $(\cdot\ ,\cdot)$.

1. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis of V, and let $w_1, w_2, \ldots, w_n$ be a vectors in V satisfying $(w_i, v_j) = \delta_{ij}$. Prove that $\{w_1, w_2, \ldots, w_n\}$ is a basis of V.

2. Let $f_1, ..., f_k \in V^*$ be linearly independent linear functionals and let $v_1^*, ..., v_k^* \in V$ correspond to $f_1, ..., f_k$ using the Riesz Representation Theorem:
$$f_i(v) = (v_i^*, v)$$
for all $v \in V$. Show that $v_1^*, ..., v_k^*$ are linearly independent.

3. Let $T : V \to V$ be a linear transformation satisfying $(T(v), w) = (v, T(w))$ for all $v, w \in V$. Show that T must be self-adjoint.

4. Let $T : V \to V$ be a skew-Hermitian operator ($T^* = -T$). What can you say about the eigenvalues of T?

5. Let $T : V \to V$ again be skew-Hermitian. Show that $e^T : V \to V$ is a unitary operator. (Just use formal properties of the exponential.)

6. For any linear operator T on the finite-dimensional Hilbert space V, prove that
$$(\sum_{k=0}^{\infty} \frac{T^k}{k!})^* = \sum_{k=0}^{\infty} \frac{1}{k!}(T^*)^k.$$

7. For any linear operator T on the finite-dimensional Hilbert space V, prove that $e^{-T} = (e^T)^{-1}$.

8. Let $T : V \to V$ be a normal operator. Show that

   (a) T is Hermitian if and only if all the eigenvalues of T are real;

   (b) T is unitary if and only if every eigenvalue of T has complex norm 1.

   (c) T is unitary if and only if T is "length-preserving," i.e., for all $v \in V$, $\|T(v)\| = \|v\|$. (Hint: Let $u, w \in V$ and apply the condition $\|T(v)\| = \|v\|$ to $v = u + w$ and to $v = u + iw$.)

9. Let V be a two-dimensional Hilbert space with ordered orthonormal basis $\mathcal{A} = (u_1, u_2)$. Let T be a Hermitian operator satisfying

$$T_{\mathcal{A}} = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}.$$

   (a) Find an ordered orthonormal basis consisting of eigenvectors of T;

   (b) Find orthogonal self-adjoint projections P_1, P_2 such that $I_V = P_1 + P_2$ and $T = \lambda_1 P_1 + \lambda_2 P_2$.

10. $T : V \to V$ be a linear transformation. Prove that the following conditions are equivalent:

   (a) T is normal.

   (b) $T^* = f(T)$, for some polynomial $f(x) \in \mathbb{C}[x]$.

   (c) $||T(v)|| = ||T^*(v)||$ for all $v \in V$.

   (d) Every T-invariant subspace of V is also T^*-invariant.

   (Hint: For (a)⇒ (b) recall that $T = \lambda_1 P_1 + \cdots + \lambda_k P_k$, where the projections $P_1, \ldots, P_k$ are self-adjoint and are polynomials in T. Thus $T^* = \bar{\lambda}_1 P_1 + \cdots + \bar{\lambda}_k P_k$, a polynomial in T. For (c)⇒ (a), form the self-adjoint operator $N = T^*T - TT^*$; if N has a non-zero eigenvalue λ with corresponding eigenvector v, then using (c), get $(T(v), T(v)) = (T^*(v), T^*(v))$ from which it follows that $0 = (v, (T^*T - TT^*)(v)) = (v, N(v)) = \lambda||v||^2$, a contradiction. Thus $T^*T - TT^* = 0$. If we assume (d), and if $v \in V$ is an eigenvector of T, then by assumption v is an eigenvector of T^*, also. Therefore, $(v, w) = 0$ implies that $(v, T(w)) = (T^*(v), w) = 0$, i.e., T (and T^*) both leave $v^\perp$ invariant. By induction on the dimension of the underlying Hilbert space, $T|_{v^\perp}$ is normal; the details are easy enough to finish. All remaining implications are pretty obvious.)

11. Show that the set of unitary operators on V forms a group under multiplication. If dim $V \geq 2$, show that this group is not abelian.

12. Let $T : V \to V$ be a unitary linear transformation. Show that $|\det T| = 1$.

13. Let V be two-dimensional and let $\mathrm{SU}_2 = \{T : V \to V|\ T^*T = I_V$ and $\det T = 1\}$. Show that

$$\mathrm{SU}_2 \cong \left\{ \begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix} \mid a, b \in \mathbb{C},\ |a|^2 + |b|^2 = 1 \right\}.$$

14. * Show that SU_2 is homeomorphic with the 3-sphere $S^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}_4|\ x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\}$, where SU_2 carries the subspace topology inherited from $\mathbb{C}^2 = \mathbb{C} \times \mathbb{C}$.

Appendix J

Exercises

Basic Concepts:
tensor product of vector spaces and of linear transformations; *Kronecker product* of matrices.

1. Suppose that $T : V \to V$ and $S : W \to W$ are linear transformations. Show that if $\dim V = n$ and $\dim W = m$, then

   (i) $\det T \otimes S = (\det T)^m (\det S)^n$.

   (ii) $\mathrm{trace}\, T \otimes S = (\mathrm{trace}\, T)(\mathrm{trace}\, S)$.

2. Let $T : V \to V$ and $S : W \to W$ are linear transformations. Show that if $v \in V$ is an eigenvector of T with corresponding eigenvalue λ, and $w \in W$ is an eigenvector of S with corresponding eigenvalue σ, then $v \otimes w$ is an eigenvector of $T \otimes S$ with eigenvalue $\lambda\sigma$.

3. Let $f_1, ..., f_k \in V^*$ be linearly independent linear functionals and let $v_1^*, ..., v_k^* \in V$ correspond to $f_1, ..., f_k$ using the Riesz Representation Theorem:
$$f_i(v) = (v_i^*, v)$$
for all $v \in V$. Show that $v_1^*, ..., v_k^*$ are linearly independent.

4. Let $T : V \to V$ be a linear transformation satisfying $(T(v), w) = (v, T(w))$ for all $v, w \in V$. Show that T must be self-adjoint.

5. Let $T : V \to V$ be a skew-Hermitian operator ($T^* = -T$). What can you say about the eigenvalues of T?

6. Let $T : V \to V$ again be skew-Hermitian. Show that $e^T : V \to V$ is a unitary operator. (Just use formal properties of the exponential.)

7. For any linear operator T on the finite-dimensional Hilbert space V, prove that
$$(\sum_{k=0}^{\infty} \frac{T^k}{k!})^* = \sum_{k=0}^{\infty} \frac{1}{k!}(T^*)^k.$$

8. For any linear operator T on the finite-dimensional Hilbert space V, prove that $e^{-T} = (e^T)^{-1}$.

9. Let $T : V \to V$ be a normal operator. Show that

   (a) T is Hermitian if and only if all the eigenvalues of T are real;

   (b) T is unitary if and only if every eigenvalue of T has complex norm 1.

   (c) T is unitary if and only if T is "length-preserving," i.e., for all $v \in V$, $\|T(v)\| = \|v\|$. (Hint: Let $u, w \in V$ and apply the condition $\|T(v)\| = \|v\|$ to $v = u + w$ and to $v = u + iw$.)

10. Let V be a two-dimensional Hilbert space with ordered orthonormal basis $\mathcal{A} = (u_1, u_2)$. Let T be a Hermitian operator satisfying
$$T_{\mathcal{A}} = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}.$$

    (a) Find an ordered orthonormal basis consisting of eigenvectors of T;

    (b) Find orthogonal self-adjoint projections P_1, P_2 such that $I_V = P_1 + P_2$ and $T = \lambda_1 P_1 + \lambda_2 P_2$.

11. $T : V \to V$ be a linear transformation. Prove that the following conditions are equivalent:

    (a) T is normal.

    (b) $T^* = f(T)$, for some polynomial $f(x) \in \mathbb{C}[x]$.

    (c) $||T(v)|| = ||T^*(v)||$ for all $v \in V$.

    (d) Every T-invariant subspace of V is also T^*-invariant.

    (Hint: For (a)$\Rightarrow$ (b) recall that $T = \lambda_1 P_1 + \cdots + \lambda_k P_k$, where the projections $P_1, \ldots, P_k$ are self-adjoint and are polynomials in T. Thus $T^* = \bar{\lambda}_1 P_1 + \cdots + \bar{\lambda}_k P_k$, a polynomial in T. For (c)$\Rightarrow$ (a), form the self-adjoint operator $N = T^*T - TT^*$; if N has a non-zero eigenvalue λ with corresponding eigenvector v, then using (c), get $(T(v), T(v)) = (T^*(v), T^*(v))$ from which it follows that $0 = (v, (T^*T - TT^*)(v)) =$

$(v, N(v)) = \lambda ||v||^2$, a contradiction. Thus $T^*T - TT^* = 0$. If we assume (d), and if $v \in V$ is an eigenvector of T, then by assumption v is an eigenvector of T^*, also. Therefore, $(v, w) = 0$ implies that $(v, T(w)) = (T^*(v), w) = 0$, i.e., T (and T^*) both leave $v^\perp$ invariant. By induction on the dimension of the underlying Hilbert space, $T|_{v^\perp}$ is normal; the details are easy enough to finish. All remaining implications are pretty obvious.)

12. Show that the set of unitary operators on V forms a group under multiplication. If dim $V \geq 2$, show that this group is not abelian.

13. Let $T : V \to V$ be a unitary linear transformation. Show that $|\text{det } T| = 1$.

14. Let V be two-dimensional and let $\mathrm{SU}_2 = \{T : V \to V |\ T^*T = I_V \text{ and det } T = 1\}$. Show that

$$\mathrm{SU}_2 \ \cong \ \left\{ \left[\begin{array}{cc} a & b \\ -\bar{b} & \bar{a} \end{array} \right] \mid a, b \in \mathbb{C}, \ |a|^2 + |b|^2 = 1 \right\}.$$

15. * Show that SU_2 is homeomorphic with the 3-sphere $S^4 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}_4 |\ x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\}$, where SU_2 carries the subspace topology inherited from $\mathbb{C}^2 = \mathbb{C} \times \mathbb{C}$.

Index