Miscellaneous Notes on Linear Algebra

Simon Xiang

March 20, 2021

Who ever suffered from learning too much linear algebra? These notes will seek to fill in my linear algebra gaps. New inclusion: these notes will also cover any miscellaneous material I should have learned in my undergraduate analysis, abstract algebra, topology, or whatever classes but didn't. Source files: https://git.simonxiang.xyz/math_notes/files.html

Contents

Lecture 1

Basic linear algebra

Here we review things like how to multiply matrices.

1.1 Basics

A set of vectors $\{v^i\}$ **linearly independent** if $\sum_i c_i v^i = 0$ implies $c_i = 0$ for all i. A **basis** is a linearly independent *spanning set*, that is, for a basis $\{e_i\}$, every vector $v \in V$ can be written as a linear combination $v = \sum_i v^i e_i$. A map $T\colon V \to W$ is **linear** (or a **homomorphism**) if for $v^1, v^2 \in V$ and $a_1, a_2 \in \mathbb{F}$, $T(a_1 v^1 + a_2 v^2) = a_1 T(v^1) + a_2 T(v^2)$. For $U := \{u^1, u^2, \cdots\}$ a finite subset of vectors in V, any map $T\colon U \to W$ induces a linear map $T\colon V \to W$ by the rule

$$T\left(\sum_i a_i u^i\right) := \sum_i a_i T(u^i).$$

The original map is said to have been **extended by linearity**[1]. The set of $v \in V$ such that $Tv = 0$[2] is the **kernel** of T, and $\dim \ker T$ is called the **nullity** of T. The **rank** of T is defined as $\dim \operatorname{im} T$. If T is bijective then it is an **isomorphism**, where V and W are said to be **isomorphic**. A linear map from a space to itself is an **endomorphism**, and a self-bijection is an **automorphism**.

Consider the short exact sequence

$$0 \longrightarrow \ker T \overset{\iota}{\hookrightarrow} V \xrightarrow{T} W \longrightarrow 0$$

for $T\colon V \to W$ surjective.

Theorem 1.1. *For the short exact sequence above, there exists a linear map $S\colon W \to V$ such that $T \circ S = 1$. We say the exact sequence* ***splits****.*

To see this, by surjectivity each basis element of W gets mapped onto by some element in V. Extend the inverse map by linearity, then this new map S satisfies $T \circ S = 1$. This map S is called a **section** of T.

Rank-Nullity Theorem. *For the short exact sequence above, let S be a section of T. Then*

$$V = \ker T \oplus S(W).$$

In particular, $\dim V = \dim \ker T + \dim S(W)$.

Proof. By the first isomorphism theorem, we have the short exact sequence $0 \to \ker T \hookrightarrow V \to \operatorname{im} T \to 0$. Then since $V \to \ker T$ is a retract, apply the splitting lemma to get that the middle map is an isomorphism in the diagram below.

$$\begin{tikzcd}
0 \arrow[r] \arrow[d] & \ker T \arrow[r, hook] \arrow[d] & V \arrow[r, "T"] \arrow[d, "\text{iso}"] & \operatorname{im} T \arrow[r] \arrow[d] & 0 \arrow[d] \\
0 \arrow[r] & \ker T \arrow[r] & \ker T \oplus \operatorname{im} T \arrow[r] & \operatorname{im} T \arrow[r] & 0
\end{tikzcd}$$

The rank nullity theorem follows. ⊠

[1] Doesn't this only work when U is a spanning set for V?

[2] We use the notation $T(v) := Tv$ from now on.

1.2 Fiddling with indices (without explanation)

For an endomorphism $T: V \to V$ with a basis $\{e_i\}$ of V, we can construct an $n \times n$ matrix whose entries T^i_j are given by

$$Te_j = \sum_i e_i T^i_j.$$

We write (T^i_j) or $\mathbf{T}$ to indicate the matrix with entries T^i_j. The map $T \to \mathbf{T}$ is a **representation** of T in the basis $\{e_i\}$. A different basis leads to a different matrix, but they represent the same endomorphism. Here's how I visualize the indices (with $j = 3$ as an example):

$$T\left(e_j\right) = T\begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix} = T_{13}\begin{pmatrix}1\\0\\0\\0\\0\end{pmatrix} + T_{23}\begin{pmatrix}0\\1\\0\\0\\0\end{pmatrix} + T_{33}\begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix} + T_{43}\begin{pmatrix}0\\0\\0\\1\\0\end{pmatrix} + T_{53}\begin{pmatrix}0\\0\\0\\0\\1\end{pmatrix} = \sum_i e_i T^i_j.$$

The splitting happens because that's how matrix multiplication is defined. For $v = \sum_i v^i e_i \in V$, we have

$$v' := Tv = \sum_j v^j T e_j = \sum_{ij} v^j e_i T^i_j = \sum_i \left(\sum_j T^i_j v^j\right) e_i = \sum_i v^{i'} e_i,$$

so the components of v' are related to the components of v by the rule $v^{i'} = \sum_j T^i_j v^j$. It is time to introduce Einstein summation notation, where flipping the indices means an implicit sum. So our equation above becomes

$$v' := Tv = v^j T e_j = v^j e_i T^i_j = T^i_j v^j e_i = v^{i'} e_i \implies v^{i'} = T^i_j v^j.$$

For S and T two endomorphisms of V, if $ST := S \circ T$, matrix multiplication is defined as $ST_{ij} = \sum_k S_{ik} T_{kj}$. In Einstein summation notation, this is notated $ST^i_j = S^i_k T^k_j$.

Note. Indices are confusing. From Wikipedia, some mnemonics: the *up*per indices go *up* to down, *l*ower indices go *l*eft to right. Covariant tensors are row vectors with lower indices (but they sum over an upper index). The lower index indicates which *column* you are in, hence why the indeed go left to right. Similarly, the upper index indicates which *row* you are in. This is the picture to keep in mind:

$$\alpha = \begin{pmatrix} & \alpha & \end{pmatrix}, \quad v = \begin{pmatrix} \\ v \\ \\ \end{pmatrix}, \qquad \phi^j = \begin{pmatrix}0 & 0 & 1 & 0 & 0\end{pmatrix}, \quad e_i = \begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix}.$$

$$\begin{pmatrix} & \alpha & \end{pmatrix}\begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix} = \alpha_i, \qquad \begin{pmatrix}0 & 0 & 1 & 0 & 0\end{pmatrix}\begin{pmatrix} \\ v \\ \\ \end{pmatrix} = v^j.$$

Note that the only things you should be looking at are ϕ^j and e_i, since they're the actual vectors, while α_j and v^i are coordinate functions with flipped indices so we can sum over them. If you think of a covector $\alpha = \begin{pmatrix}w_1 & w_2 & \cdots\end{pmatrix}$, you can see why we say they have *lower* indices. However, when you write the implicit sum $\alpha = \alpha_j \phi^j$, the ϕ^j (which are covectors) have an upper index because that's what we're summing over: the actual entries have lower indices. For multi-index sums like $v^j e_i T^i_j$, we sum left to right.

The **row rank** (resp **column rank**) of a matrix T is the maximum number of LI rows (resp columns) when considered as vectors in $\mathbb{R}^n$. These concepts are equal, and we call this the **rank** of T, denoted $\operatorname{rank} T$. If $\operatorname{rank} T = n$, then T has **maximal rank**, otherwise T is **rank deficient**. For $\{e_i\}$ and $\{e_i'\}$ two bases of V, we can write $e_j' = e_j A^i_j$ for some nonsingular A, called the **change of basis matrix**. If $v = v^i e_i = v^{i'} e_i'$, then

$$v^{j'} e_j' = v^{j'} e_i A^i_j = A^i_j v^{j'} e_i = v^i e_i.$$

So $v_j = A^i_j \{v_j\}'$, or $v_j' = (A^{-1})^i_j v^j$. We write $\langle v, f\rangle$ or $\langle f, v\rangle$ to denote $f(v)$. Then for $\{\phi^j\}$ a dual basis for $\{e_i\}$, we have $\langle e_i, \phi^j\rangle = \delta^j_i$. For $\{\phi^{i'}\}$ a dual basis corresponding to $\{e_i'\}$, write $\phi^{j'} = \phi^i B^j_i$. Then

$$\delta^j_i = \langle e_i', \phi^{j'}\rangle = \langle e_k A^k_i, \phi^\ell B^j_\ell\rangle = A^k_i B^j_\ell \langle e_k, \phi^\ell\rangle = A^k_i B^j_\ell \delta^\ell_k = A^k_i B^j_k.$$

If we write $A^T := A^j_i$, we can write the result above as $A^T B = I$, equivalently $B = (A^T)^{-1} = (A^{-1})^T$, the **contragredient matrix** of A. For $f \in V^*$ a covector, under a change of basis we have

$$f' = f_j' \phi^{j'} = f_j' \phi^i B^j_i = B^j_i f_j' \phi^{i'} = f_i \phi^i = f, \quad \Longrightarrow \quad f_i = B^j_i f_j', \quad f_i' = (B^{-1})^j_i f_j.$$

Rewriting in terms of A, we have

$$\phi^{i'} = \phi^j B^i_j = (B^T)^i_j \phi^j = (A^{-1})^i_j \phi^j, \quad f_j' = (B^{-1})^i_j f_i = (A^T)^i_j f_i = f_i A^i_j.$$

1.3 Upstairs or downstairs?

Let's talk about what just happened. If we use standard notation, the symbol a_j is ambiguous: are they components of vectors, covectors, or neither? How can we tell? We can't, we can only guess (you can tell when they're paired with the corresponding basis elements e_i or ϕ^i, but sometimes those are omitted for brevity). Introducing up down indices allows us to differentiate the two.

Under a change of basis, the components of a covector transform like basis vectors, while the components of a vector transform like cobasis vectors. We say the components of a covector transform **covariantly** (with the basis vectors), while the components of a vector transform **contravariantly** (against the basis vectors). Because of this, we write e_i for a basis vector as normal, but we use a raised index ϕ^i to denote the basis covectors. Then vector components are written with upstairs (contravariant) indices and covector components are written with downstairs (covariant indices).

Writing $v = v^i e_i$ and $f = f_i \phi^i$ allows us to quickly pair the up indices and down indices to see what is being summed. When this happens, we say the indices have been **contracted**. Avoid things like $a_i = b^i$. To summarize our results, we have $\langle e_j, \phi^j\rangle = \delta^j_i$, $e_j' = e_i A^i_j$, $v'^i = (A^{-1}) i_j v^j$. This notation also leads to much pedanticism and confusion as you may have already noticed. Introducing the shorthand

$$\mathbf{A} = \begin{pmatrix} A^i_j \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 & e_2 & \cdots & e_n \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta^1 \\ \theta^2 \\ \vdots \\ \theta^n \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_1 & f_2 & \cdots & f_n \end{pmatrix}$$

helps, since the above equations become $\mathbf{e}' = \mathbf{eA}$, $\mathbf{v}' = \mathbf{A}^{-1}\mathbf{v}$, $\theta' = \mathbf{A}^{-1}\theta$, $\mathbf{f}' = \mathbf{fA}$. The invariance of v and f under a change of basis become easy to see, for example $v' = \mathbf{e}'\mathbf{v}' = \mathbf{eAA}^{-1}\mathbf{v} = \mathbf{ev} = v$.

1.4 Inner product spaces

This is a more mature treatment of the material later in this paper thing. Let $\mathbb{F}$ be a subfield of $\mathbb{C}$, and V and $\mathbb{F}$-vector space. A **sesquilinear form** on V is a map $g\colon V \times V \to \mathbb{F}$ satisfying the following properties: for all $u, v, w \in V$ and $a, b \in \mathbb{F}$, the map g is

(i) **linear on the second entry**: $g\colon (u, av + bw) \to= ag(u, v) + bg(u, w)$, and
(ii) **Hermitian**: $g(v, u) = \overline{g(u, v)}$.

These two properties imply that g is **antilinear** on the first entry, that is, $g(au+bv,w)=\overline{a}g(u,w)+\overline{b}g(v,w)$. If $\mathbb{F}$ is a real field (subfield of $\mathbb{R}$), then this just says that g is a **symmetric bilinear form**. If a sequilinear form g is **nongenerate**, where $g(u,v)=0$ for all v implies $u=0$, then g is an **inner product.** A space equipped with an inner product is an **inner product space**.

Note that $g(u,u)$ is real by Hermiticity. If $g(u,u)\geq 0$ (resp $g(u,u)\leq 0$), then g is **nonnegative definite** (resp **nonpositive definite**). If $g(u,u)=0$ implies that $u=0$, then g is **positive definite** (resp **negative definite**).

Example 1.1 (The **Lorentizan inner product** on $\mathbb{R}^n$)**.** Let $u=(u_0,u_1,\cdots,u_{n-1})$ and $v=(v_0,v_1,\cdots,v_{n-1})$, and define

$$g(u,v):=-u_0v_0+\sum_{i=1}^{n-1}u_iv_i.$$

The vector space $\mathbb{R}^n$ equipped with this inner product is denoted $\mathbb{M}^n$ and is called **Minkowski space** (or **Minkowski spacetime**). Note that while the Lorentzian inner product is an indeed an inner product, it is not positive definite.

A set $\{v_i\}$ of vectors is **orthogonal** if $g(v_i,v_j)=0$ for $i\neq j$, and is **orthonormal** if $g(v_i,v_j)=\pm\delta_{ij}$. A vector v satisfying $g(v,v)=\pm 1$ is a **unit vector**.

Theorem 1.2. *Every inner product space has an orthonormal basis.*

First proof of Theorem 1.2. We use induction on $k=\dim V$. If todo:some algebra ⊠

Second proof of Theorem 1.2. todo:grammian, spectral theorem, diagonalization, sylvester's law of inertia ⊠

todo:the reisz lemma

1.5 The tensor product

What are tensors? Define a new vector product called the **tensor product**, denoted by $v\otimes w$[3]. The product is a **tensor of order 2** or a **second-order tensor** or a **2-tensor**. The tensor product is *noncommutative* in general, and we form higher order tensors by repeated iteration. Order-0 tensors are scalars, while order-1 tensors are vectors. In older literature $v\otimes w$ becomes vw and is called a *dyadic* product.

The set $\mathscr{T}^r$ of order r tensors forms a natural vector space: for S and T order r tensors, $aT+bS$ is another order r tensor. We write $\mathscr{T}^r:=V\otimes V\otimes\cdots\otimes V=V^{\otimes r}$. The set $\mathscr{T}=\bigcup_r\mathscr{T}^r$ forms an **algebra**, basically a ringed vector space satisfying homogeneity. The multiplication says that for R a tensor of order r and S an s-tensor, then $R\otimes S$ is an $(r+s)$-tensor. Let us write the (graded) algebra conditions in tensor language:

(1) **left distributivity**: $R\otimes(S+T)=R\otimes S+R\otimes T$,
(2) **right distributivity**: $(S+T)\otimes R=S\otimes R+T\otimes R$,
(3) **homogeneity**: $T\otimes(aS)=(aT)\otimes S=a(T\otimes S)$.

A tensor also has components in some basis. For e_i a basis of $\mathbb{R}^n$, the canonical basis for $\mathbb{R}^n\otimes\mathbb{R}^m$ is given by the nm elements of $\{e_i\otimes e_j\}$ as i varies over n and j varies over m. A general second-order tensor on $\mathbb{R}^n$ is a linear combination of these basis vectors of the form $T=\sum_{ij}T^{ij}e_i\otimes e_j=T^{ij}e_i\otimes e_j$. Usually the basis is understood, so T^{ij} is called a tensor, when it actually gives the components of a tensor with respect to some basis. To find the components of $v\otimes w$, observe that

$$v\otimes w=v^ie_i\otimes w^je_j=v^iw^i(e_i\otimes e_j).$$

Example 1.2. Given a rigid body consisting of a bunch of point masses m_α at positions $\mathbf{r}_\alpha=(x_{\alpha,1},x_{\alpha,2},x_{\alpha,3})$, its **inertia tensor** is given by

$$I_{ij}=\sum_\alpha m_\alpha(r_\alpha^2\delta_{ij}-x_{\alpha,i},x_{\alpha,j}),$$

where $r_\alpha^2=\mathbf{r}_\alpha\cdot\mathbf{r}_\alpha$. There is a lot of sloppiness going on with indices and denoting components as tensors.

[3]These are actually defined by a *universal property* in category theory, but let's brush over the details.

1.6 Two ways to view general tensors

1: As an element of the tensor product space

We have been excluding covectors from the fun. A **tensor of type $(\mathbf{r}, \mathbf{s})$** is an element of the tensor product space

$$T_s^r = \overbrace{V \otimes V \otimes \cdots \otimes V}^{r \text{ times}} \otimes \overbrace{V^* \otimes V^* \otimes \cdots \otimes V^*}^{s \text{ times}} = V^{\otimes r} \otimes (V^*)^{\otimes s}.$$

An r-tensor previously is now a tensor of type $(r, 0)$. This space of all tensors forms a **multigraded algebra**, that is, multiplying a (r,s)-tensor and a (p,q)-tensor gives a tensor of type $(r+p, s+q)$. For a basis $\{e_i\}$ of V and dual basis $\{\phi^i\}$ of V^*, a basis for $\mathscr{T}_s^r$ is given by

$$e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_r} \otimes \phi^{j_1} \otimes \phi^{j_2} \otimes \cdots \otimes \phi^{j_s},$$

where the indices run from 1 to $\dim V$. A general tensor of type (r,s) is a linear combination

$$T = T^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_r} \otimes \phi^{j_1} \otimes \phi^{j_2} \otimes \cdots \otimes \phi^{j_s},$$

with an implicit sum over $i_1 \cdots i_r$, $j_1 \cdots j_s$. From before, we can see that upstairs indices transform contravariantly, while downstairs indices transform covariantly.

$$T^{i_1' \cdots i_r'}_{j_1' \cdots j_s'} = T^{i_1 \cdots i_r}_{j_1 \cdots j_s} (A^{-1})^{i_1'}_{i_1} \cdots (A^{-1})^{i_r'}_{i_r'} A^{j_1}_{j_1'} \cdots A^{j_s}_{j_s'}.$$

2: As a multilinear functional on the dual space

Consider the space of multilinear maps $\widetilde{\mathscr{T}}_s^r$. Recall the **natural pairing**, where $\langle f, v \rangle = \langle v, f \rangle$ denotes $f(v)$. We can view the tensor $e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes \phi^{j_1} \otimes \cdots \otimes \phi^{j_s}$ as a multilinear map on the space $(V^*)^{\times r} \times V^{\times s}$ that acts according to the rule

$$\begin{aligned}
&(e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes \phi^{j_1} \otimes \cdots \otimes \phi^{j_s})(\phi^{k_1}, \cdots, \phi^{k_r}, e_{\ell_1}, \cdots, e_{\ell_s}) \\
&\quad = \langle e_{i_1}, \phi^{k_1} \rangle \cdots \langle e_{i_r}, \phi^{k_r} \rangle \langle \phi^{j_1}, e_{\ell_1} \rangle \cdots \langle \phi^{j_s}, e_{\ell_s} \rangle \\
&\quad = \delta^{k_1}_{i_1} \cdots \delta^{k_r}_{i_r} \delta^{j_1}_{\ell_1} \cdots \delta^{j_s}_{\ell_s}.
\end{aligned}$$

If we view the tensor product this way, we have

$$\begin{aligned}
&T(\phi^{k_1}, \cdots, \phi^{k_r}, e_{\ell_1}, \cdots, e_{\ell_s}) \\
&\quad = T^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} \times (e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_r} \otimes \phi^{j_1} \otimes \phi^{j_2} \otimes \cdots \otimes \phi^{j_s})(\phi^{k_1}, \cdots, \phi^{k_r}, e_{\ell_1}, \cdots, e_{\ell_s}) \\
&\quad = T^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} \delta^{k_1}_{i_1} \cdots \delta^{k_r}_{i_r} \delta^{j_1}_{\ell_1} \cdots \delta^{j_s}_{\ell_s} \\
&\quad = T^{k_1 k_2 \cdots k_r}_{\ell_1 \ell_2 \cdots \ell_s}.
\end{aligned}$$

This gives an isomorphism between $\widetilde{\mathscr{T}}_s^r$ and $\mathscr{T}_s^r$. In essence, you can choose to view tensors *passively* as elements of a certain vector space (the tensor product space), or *actively* as multilinear functionals on the dual space. They are two sides of the same coin, so we can interchange the notations as we please.

Lecture 2

Miscellaneous topics

TODO: affine spaces, inverse function, change of variables for multiple integrals (spivak 34,67) or tu appendix, rank, nullity, binomial theorem, freed's thing, maybe topology bases, subspace/product, tychonoff, convergnece, etc

2.1 The Inverse Function Theorem

Lecture 3

Inner-Product Spaces

What is an inner product?? Let's find out.

3.1 Inner Products

The length of a vector x is the **norm** of x, denoted $\|x\|$. If $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, we have $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$. Note that the norm is not linear. For $x, y \in \mathbb{R}^n$, the **dot product** of x and y, denoted $x \cdot y$, is defined by $x \cdot y = x_1 y_1 + \cdots + x_n y_n$. Note that this is a number, not a vector. Clearly $x \cdot x = \|x\|^2$ for all $x \in \mathbb{R}^n$, which implies $x \cdot x \geq 0$ for all $x \in \mathbb{R}^n$ ($x \cdot x = 0$ only if x is the zero vector). The map that sends $x \in \mathbb{R}^n$ to $x \cdot y$ in $\mathbb{R}$ for fixed y is linear since $\mathbb{R}$ is a field. The dot product is also commutative, since $\mathbb{R}$ is.

Inner products generalize dot products. Recall that $|\lambda|^2 = \lambda\overline{\lambda}$ for $\lambda \in \mathbb{C}$. For $z \in \mathbb{C}^n$, we define the norm of z by $\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}$. We take the modulus of z_i since we want the result to be nonnegative. Note that $\|z\|^2 = z_1\overline{z_1} + \cdots z_n\overline{z_n}$. We want to think of $\|z\|^2$ as the inner product of z with itself, like in $\mathbb{R}^n$. This suggests we define the inner product of $w = (w_1, \cdots, w_n) \in \mathbb{C}^n$ with z as $w_1\overline{z_1} + \cdots + w_n\overline{z_n}$. We expect the inner product of w with z equal the complex conjugate of the inner product of z with w. With this motivation in mind, let us define inner products.

Definition 3.1 (Inner product)**.** An **inner product** on an F-vector space V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in F$ such that

(i) $\langle v, v\rangle \geq 0$ for all $v \in V$; (**positivity**)

(ii) $\langle v, v\rangle = 0$ iff $v = 0$; (**definiteness**)

(iii) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$; (**additivity in first slot**)

(iv) $\langle av, w\rangle = a\langle v, w\rangle$ for all $a \in F$ and all $v, w, \in V$; (**homogeneity in first slot**)

(v) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$. (**conjugate symmetry**).

For real numbers, condition (v) simply becomes $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$. An **inner product space** is a vector space V along with an inner product on V.

Example 3.1. The most important example is the **Euclidian inner product** on $\mathbf{F}^n$ (Axler uses $\mathbf{F}$ to denote either $\mathbb{C}$ or $\mathbb{R}$). We define an inner product on $\mathbf{F}^n$ by

$$\langle (w_1, \cdots, w_n), (z_1, \cdots, z_n)\rangle = w_1\overline{z_1} + \cdots w_n\overline{z_n}.$$

An example of another inner product on $\mathbf{F}^n$ is defined by $\langle (w_1, \cdots, w_n), (z_1, \cdots, z_n)\rangle = c_1 w_1\overline{z_1} + \cdots + c_n w_n\overline{z_n}$ for c_i positive constants. The case where $c_i = 1$ for all i is simply the standard Euclidian inner product.

Example 3.2. Consider the vector space $\mathscr{P}_m(\mathbf{F})$, the polynomial ring over $\mathbf{F}$ of polynomials with degree at most m. We can define an inner product on $\mathscr{P}_m(\mathbf{F})$ by

$$\langle p, q\rangle = \int_0^1 p(x)\overline{q(x)}\,dx.$$

For fixed $w \in V$, the function that takes v to $\langle v, w\rangle$ is a linear map $V \to \mathbf{F}$. So $\langle 0, w\rangle = 0$, and by condition (v) $\langle w, 0\rangle = 0$ as well. Furthermore, $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$ and $\langle u, av\rangle = \overline{a}\langle u, v\rangle$ hold as well: This second condition is known as conjugate homogeneity in the second slot.

3.2 Norms

For $v \in V$, we define the **norm** of v, denoted $\|v\|$, by $\|v\| = \sqrt{\langle v, v\rangle}$. For example, if $p \in \mathscr{P}_m(\mathbf{F})$, then $\|p\| = \sqrt{\int_0^1 |p(x)|^2\,dx}$. Some properties: $\|v\| = 0$ iff $v = 0$, and $\|av\| = |a|\|v\|$. To see this, note that $\|av\|^2 = \langle av, av\rangle = a\langle v, av\rangle = a\overline{a}\langle v, v\rangle = |a|^2\|v\|^2$, taking square roots gives us our result. This illustrates a general idea: working with norms squared is easier than working directly with norms.

Two vectors $u, v \in V$ are **orthogonal** if $\langle u, v\rangle = 0$. The zero vector is orthogonal to every vector, and the only vector orthogonal to itself. Assume $V = \mathbb{R}^2$, now let us state a 2500 year old theorem.

Pythagorean Theorem. *If u, v are orthogonal vectors in V, then*

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof. Exercise. ⊠

Suppose $u, v \in V$. We want to write u as a scalar multiple of v plus a vector w orthogonal to v. Let $a \in \mathbf{F}$ be a scalar, then $u = av + (u - av)$. We need to choose a such that v is orthogonal to $u - av$, in other words, we want $0 = \langle u - av, v\rangle = \langle u, v\rangle - a\|v\|^2$. So we should choose $a = \langle u, v\rangle / \|v\|^2$ (where $v \neq 0$). Then

$$u = \frac{\langle u, v\rangle}{\|v\|^2} v + \left(u - \frac{\langle u, v\rangle}{\|v\|^2} v\right).$$

Cauchy-Schwarz Inequality. *If $u, v \in V$, then*

$$|\langle u, v\rangle| \leq \|u\|\|v\|.$$

This inequality is an equality iff one of u, v is a scalar multiple of the other.

Proof. Let $u, v \in V$, and assume $v \neq 0$. Consider $u = \frac{\langle u,v\rangle}{\|v\|^2} v + w$, where w is orthogonal to v. By the Pythagorean theorem, we have

$$\|u\|^2 = \left\|\frac{\langle u, v\rangle}{\|v\|^2} v\right\|^2 + \|w\|^2 = \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}.$$

Multiply both sides, take a square root, and we are done. This is an equality iff $w = 0$, but this is true iff u is a multiple of v. ⊠

Triangle Inequality. *If $u, v \in V$, then*

$$\|u+v\| \leq \|u\| + \|v\|.$$

This is an equality iff one of u, v is a nonnegative multiple of the other.

Proof. Let $u, v \in V$. Then

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \overline{\langle u, v\rangle} = \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\langle u, v\rangle \leq \|u\|^2 + \|v^2\| + 2\|u\|\|v\| = (\|u\| + \|v\|)^2.$$

The inequality step frollows from Cauchy-Schwartz, where $2\,\mathrm{Re}\langle u, v\rangle \leq 2|\langle u, v\rangle|$. Taking square roots gives the triangle inequality. This is an equality iff the two inequalities above are equalities, which is true iff $\langle u, v\rangle = \|u\|\|v\|$. ⊠

Parallelogram Equality. *If $u, v \in V$, then*

$$\|u+v\|^2 + \|u-v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right).$$

Proof. Exercise. ⊠

3.3 Orthonormal Bases

A list $(e_1, \cdots, e_m)$ of vectors in V is orthonormal if $\langle e_j, e_k \rangle = 0$ when $j \neq k$ and equals 1 when $j = k$, for $j, k \in \{1, \cdots, m\}$. Orthonormal lists are nice.

Proposition 3.1. *If $(e_1, \cdots, e_m)$ is an orthonormal list of vectors in V, then*

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all $a_1, \cdots, a_m \in \mathbf{F}$.

Proof. Since each e_j has norm 1, this follows from repeated applications of the Pythagorean theorem. ⊠

Corollary 3.1. *Every orthonormal list of vectors is linearly independent.*

An **orthonormal basis** of V is an orthonormal list of vectors in V that forms a basis for V. The standard basis is a good example. If we find an orthonormal list of length $\dim V$, then this is automatically an orthonormal basis of V (since they must be LI). In general, given a basis $(e_1, \cdots, e_n)$ of V and a vector $v \in V$, we know there is some choice of scalars $a_1, \cdots, a_m$ such that $v = a_1 e_1 + \cdots a_n e_n$, but finding the a_j's can be difficult. This is not the case for an orthonormal basis.

Theorem 3.1. *Suppose $(e_1, \cdots, e_n)$ is an orthonormal basis of V. Then*

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$$

for every $v \in V$.

Proof. Let $v \in V$. Since $(e_1, \cdots, e_n)$ is a basis of V, there exist scalars $a_1, \cdots, a_n$ such that $v = a_1 e_1 + \cdots + a_n e_n$. Taking the inner product of both sides with e_j, we get $\langle v, e_j \rangle = a_j$. The second part follows from the first proposition and our previous result. ⊠