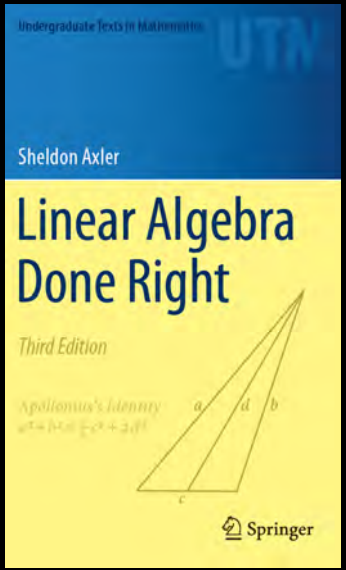
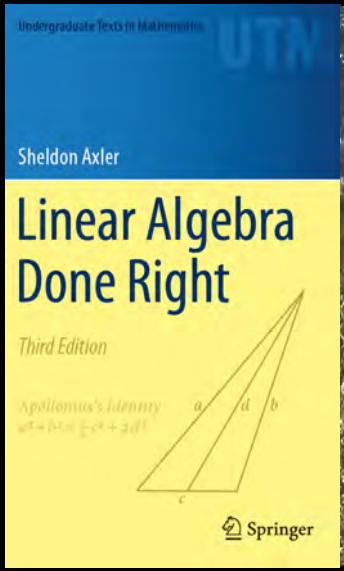
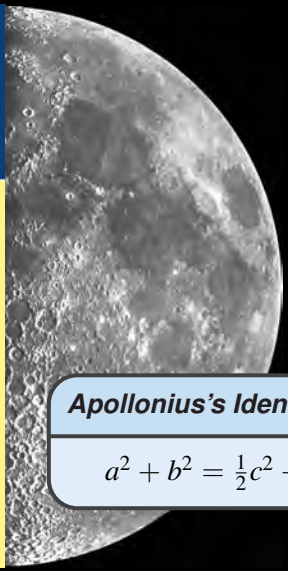
Introduction



The Front Cover





Apollonius's Identity

$$a^2 + b^2 = \tfrac{1}{2}c^2 + 2d^2$$

Definition Boxes and Result Boxes

Definition: *basis*

A *basis* of V is a list of vectors in V that is linearly independent and spans V.

Notation: $\mathbf{F}$

Throughout this book, $\mathbf{F}$ stands for either $\mathbf{R}$ or $\mathbf{C}$.

Dimension of a sum

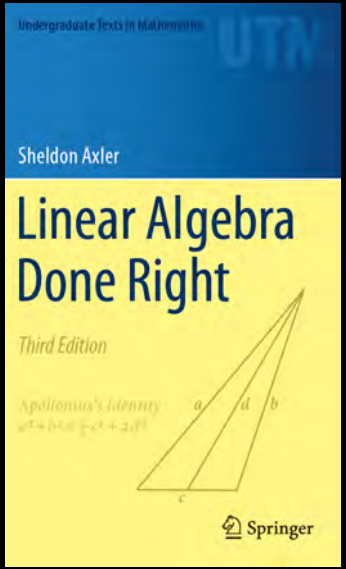
If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Contact Information

- Sheldon Axler's website: `www.axler.net`
- *Linear Algebra Done Right* website: `www.linear.axler.net`
- Suggestions and comments are welcome via email: `linear@axler.net`
- Follow *Linear Algebra Done Right* on Twitter: `@AxlerLinear`

$\mathbf{R}^n$ and $\mathbf{C}^n$



Real Numbers

Notation: $\mathbf{R}$

$\mathbf{R}$ denotes the field of real numbers.

Some nonconstant polynomials with real coefficients have no real zeros. Example: the equation

$$x^2 + 1 = 0$$

has no real solutions.

Thus we invent a solution, called i, with the property that $i^2 = -1$.

Complex Numbers

Definition: *complex numbers*

- A *complex number* is an ordered pair (a, b), where $a, b \in \mathbf{R}$, but we will write this as $a + bi$.
- The set of all complex numbers is denoted by $\mathbf{C}$:

$$\mathbf{C} = \{a + bi : a, b \in \mathbf{R}\}.$$

- *Addition and multiplication* on $\mathbf{C}$ are defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$
$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

If $a \in \mathbf{R}$, we identify $a + 0i$ with the real number a. Thus we think of $\mathbf{R}$ as a subset of $\mathbf{C}$. We also usually write $0 + bi$ as just bi, and we usually write $0 + 1i$ as just i. The definition of multiplication shows that $i^2 = -1$.

Properties of Complex Arithmetic

- **commutativity**
  $\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbf{C}$;
- **associativity**
  $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ and $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$;
- **identities**
  $\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$ for all $\lambda \in \mathbf{C}$;
- **additive inverse**
  for every $\alpha \in \mathbf{C}$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha + \beta = 0$;
- **multiplicative inverse**
  for every $\alpha \in \mathbf{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha\beta = 1$;
- **distributive property**
  $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbf{C}$.

Notation: $\mathbf{F}$

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

Elements of $\mathbf{F}$ are sometimes called *scalars*, which is just a fancy word for numbers.

$\mathbf{R}^2$ and $\mathbf{R}^3$

Definition: $\mathbf{R}^2$ *and* $\mathbf{R}^3$

- The set $\mathbf{R}^2$, which you can think of as a plane, is the set of all ordered pairs of real numbers:

$$\mathbf{R}^2 = \{(x, y) : x, y \in \mathbf{R}\}.$$

- The set $\mathbf{R}^3$, which you can think of as ordinary space, is the set of all ordered triples of real numbers:

$$\mathbf{R}^3 = \{(x, y, z) : x, y, z \in \mathbf{R}\}.$$

René Descartes and Queen Christina

Lists

Fix a positive integer n.

Definition: *list, length*

A *list* of *length* n is an ordered collection of n numbers separated by commas and surrounded by parentheses.

Example: $(7, 3)$ is a list of length 2. Thus $(7, 3) \in \mathbf{R}^2$.

Example: $(5, 9, -2)$ is a list of length 3. Thus $(5, 9, -2) \in \mathbf{R}^3$.

A list of length n looks like this:

$$(x_1, \ldots, x_n).$$

Two lists are equal if and only if they have the same length and the same elements in the same order.

Definition: $\mathbf{F}^n$

$\mathbf{F}^n$ is the set of all lists of length n of elements of $\mathbf{F}$:

$$\mathbf{F}^n = \{(x_1, \ldots, x_n) : x_j \in \mathbf{F} \text{ for } j = 1, \ldots, n\}.$$

Example: $\mathbf{R}^4$ is the set of all lists of four real numbers:

$$\mathbf{R}^4 = \{(x, y, z, w) : x, y, z, w \in \mathbf{R}\}.$$

Example: $\mathbf{C}^4$ is the set of all lists of four complex numbers:

$$\mathbf{C}^4 = \{(z_1, z_2, z_3, z_4) : z_1, z_2, z_3, z_4 \in \mathbf{C}\}.$$

Elements of $\mathbf{F}^n$ are often called *points* or *vectors*.

Addition and Scalar Multiplication

Definition: *addition in* $\mathbf{F}^n$

Addition in $\mathbf{F}^n$ is defined by adding corresponding coordinates:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n).$$

Example: $(3, 4, -1, -2) + (5, 0, 6, -7) = (8, 4, 5, -9)$

Definition: *scalar multiplication in* $\mathbf{F}^n$

The *product* of a number $\lambda \in \mathbf{F}$ and a vector in $\mathbf{F}^n$ is computed by multiplying each coordinate of the vector by λ:

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

Example: $i(i, 2, 5 + i, -3i) = (-1, 2i, -1 + 5i, 3)$

Single Letters Can Denote Elements of $\mathbf{F}^n$

Using a single letter to denote elements of $\mathbf{F}^n$ is often efficient.

Commutativity of addition in $\mathbf{F}^n$

If $x, y \in \mathbf{F}^n$, then $x + y = y + x$.

Definition: 0

Let 0 denote the list of length n whose coordinates are all 0:

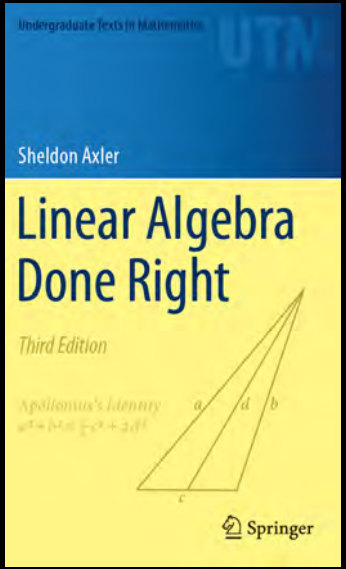
$$0 = (0, \ldots, 0).$$

Multiplication by 0

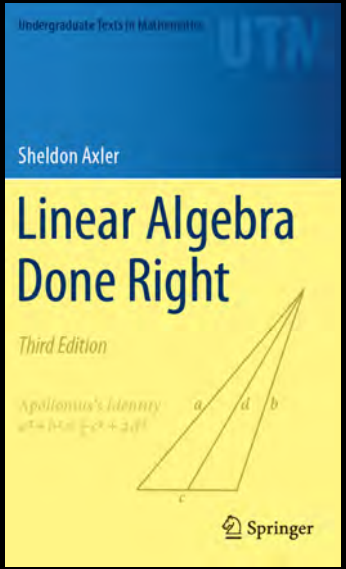
If $x \in \mathbf{F}^n$, then $0x = 0$.

The 0 above on the left is the number 0.
The 0 above on the right is the vector 0 in $\mathbf{F}^n$.

Linear Algebra Done Right, by Sheldon Axler



Definition of Vector Space



Motivation for Vector Space Definition

Reminder: $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

The motivation for the definition of a vector space comes from properties of addition and scalar multiplication in $\mathbf{F}^n$:

- Addition is commutative, associative, and has an identity.
- Every element has an additive inverse.
- Scalar multiplication is associative.
- Scalar multiplication by 1 acts as expected.
- Addition and scalar multiplication are connected by distributive properties.

Addition and Scalar Multiplication

Definition: *addition, scalar multiplication*

- An *addition* on a set V is a function that assigns an element $u + w \in V$ to each pair of elements $u, w \in V$.
- A *scalar multiplication* on a set V is a function that assigns an element $\lambda u \in V$ to each $\lambda \in \mathbf{F}$ and each $u \in V$.

Example: Suppose V is the set of real-valued functions on the interval $[0, 1]$. For $f, g \in V$ and $\lambda \in \mathbf{R}$, define $f + g$ and λf by

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

and

$$(\lambda f)(x) = \lambda f(x).$$

Thus $f + g \in V$ and $\lambda f \in V$.

Definition of Vector Space

Definition: *vector space*

A *vector space* is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:

commutativity
- $u + w = w + u$ for all $u, w \in V$;

associativity
- $(u + v) + w = u + (v + w)$ and $(ab)u = a(bu)$ for all $u, v, w \in V$ and all $a, b \in \mathbf{F}$;

additive identity
- there exists $0 \in V$ such that $u + 0 = u$ for all $u \in V$;

additive inverse
- for every $u \in V$, there exists $w \in V$ such that $u + w = 0$;

multiplicative identity
- $1u = u$ for all $u \in V$;

distributive properties
- $a(u + w) = au + aw$ and $(a + b)u = au + bu$ for all $a, b \in \mathbf{F}$ and all $u, w \in V$.

Examples

- $\mathbf{F}^n$ with the usual operations of addition and scalar multiplication is a vector space.
- $\mathbf{F}^\infty$ is defined to be the set of all sequences of elements of $\mathbf{F}$:

$$\mathbf{F}^\infty = \{(x_1, x_2, \dots) : x_j \in \mathbf{F} \text{ for } j = 1, 2, \dots\}.$$

Addition and scalar multiplication on $\mathbf{F}^\infty$ are defined as expected:

$$(x_1, x_2, \dots) + (y_1, y_2, \dots) = (x_1 + y_1, x_2 + y_2, \dots),$$
$$\lambda(x_1, x_2, \dots) = (\lambda x_1, \lambda x_2, \dots).$$

With these definitions, $\mathbf{F}^\infty$ becomes a vector space.

Another example

If S is a set, then $\mathbf{F}^S$ denotes the set of functions from S to $\mathbf{F}$.

For $f, g \in \mathbf{F}^S$, the *sum* $f + g \in \mathbf{F}^S$ is the function defined by

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

for all $x \in S$.

For $\lambda \in \mathbf{F}$ and $f \in \mathbf{F}^S$, the *product* $\lambda f \in \mathbf{F}^S$ is the function defined by

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in S$.

With these definitions of addition and scalar multiplication, $\mathbf{F}^S$ becomes a vector space.

Our First Theorem

The number 0 times a vector

If V is a vector space, then $0u = 0$ for every $u \in V$.

Proof For $u \in V$, we have

$$\begin{aligned} 0u &= (0+0)u \\ &= 0u + 0u. \end{aligned}$$

Adding the additive inverse of $0u$, denoted $-0u$, to both sides of the equation above gives

$$0u + (-0u) = 0u + 0u + (-0u),$$

which can be rewritten as

$$0 = 0u,$$

as desired. ∎

Why Abstraction?

Advantages of the abstract approach:

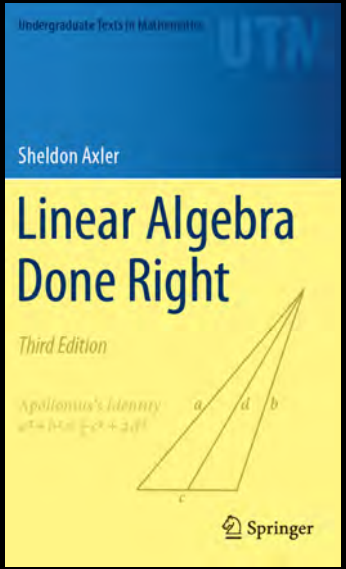
- Can apply in multiple new situations.
- Stripping away inessential properties leads to greater understanding.

If V is a vector space, it would be incorrect to prove that $0u = 0$ for $u \in V$ by writing: Let $u = (x_1, \ldots, x_n)$, thus

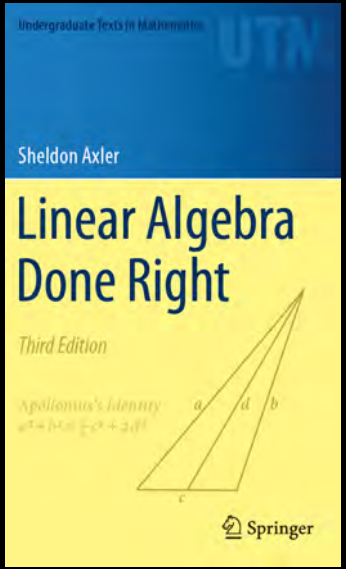
$$\begin{aligned} 0u &= (0x_1, \ldots, 0x_n) \\ &= (0, \ldots, 0) \\ &= 0. \end{aligned}$$

An element of V is not necessarily of the form $(x_1, \ldots, x_n)$.

Linear Algebra Done Right, by Sheldon Axler



Subspaces


Undergraduate Texts in Mathematics
Sheldon Axler
Linear Algebra Done Right
Third Edition
Apollonius's Identity
a
d
b
c
Springer

Definition of Subspace

Reminder: $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

From now on, V denotes a vector space over $\mathbf{F}$.

Definition: *subspace*

A subset U of V is called a *subspace* of V if U is also a vector space (using the same addition and scalar multiplication as on V).

Example: $\{(x_1, x_2, 0) : x_1, x_2 \in \mathbf{F}\}$ is a subspace of $\mathbf{F}^3$.

Conditions for a Subspace

additive identity

closed under addition

closed under scalar multiplication

Conditions for a subspace

A subset U of V is a subspace of V if and only if U satisfies the following three conditions:

- $0 \in U$;
- $u, w \in U$ implies $u + w \in U$;
- $\lambda \in \mathbf{F}$ and $u \in U$ implies $\lambda u \in U$.

Examples of Subspaces

Examples of subspaces:

- If $b \in \mathbf{F}$, then
$$\{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_3 = 5x_4 + b\}$$
is a subspace of $\mathbf{F}^4$ if and only if $b = 0$.
- The set of continuous real-valued functions on the interval $[0, 1]$ is a subspace of $\mathbf{R}^{[0,1]}$.
- The set of differentiable real-valued functions on $\mathbf{R}$ is a subspace of $\mathbf{R}^{\mathbf{R}}$.
- The set of differentiable real-valued functions f on the interval $(0, 3)$ such that $f'(2) = b$ is a subspace of $\mathbf{R}^{(0,3)}$ if and only if $b = 0$.
- The set of all sequences of complex numbers with limit 0 is a subspace of $\mathbf{C}^{\infty}$.
- The subspaces of $\mathbf{R}^2$ are precisely $\{0\}$, $\mathbf{R}^2$, and all lines in $\mathbf{R}^2$ through the origin.
- The subspaces of $\mathbf{R}^3$ are precisely $\{0\}$, $\mathbf{R}^3$, all lines in $\mathbf{R}^3$ through the origin, and all planes in $\mathbf{R}^3$ through the origin.

Sums of Subspaces

Definition: *sum of subsets*

Suppose $U_1, \ldots, U_m$ are subsets of V. The *sum* of $U_1, \ldots, U_m$, denoted $U_1 + \cdots + U_m$, is the set of all possible sums of elements of $U_1, \ldots, U_m$. More precisely,

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Sum of subspaces is the smallest containing subspace

Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is the smallest subspace of V containing $U_1, \ldots, U_m$.

Analogy: sums of subspaces $\longleftrightarrow$ unions of subsets in set theory

Direct Sums

Definition: *direct sum*

Suppose $U_1, \ldots, U_m$ are subspaces of V.

- The sum $U_1 + \cdots + U_m$ is called a *direct sum* if each element of $U_1 + \cdots + U_m$ can be written in only one way as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j.
- If $U_1 + \cdots + U_m$ is a direct sum, then $U_1 \oplus \cdots \oplus U_m$ denotes $U_1 + \cdots + U_m$, with the $\oplus$ notation serving as an indication that this is a direct sum.

Example: Suppose

$$U = \{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\} \quad \text{and} \quad W = \{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\}.$$

Then $\mathbf{F}^3 = U \oplus W$.

Direct Sums

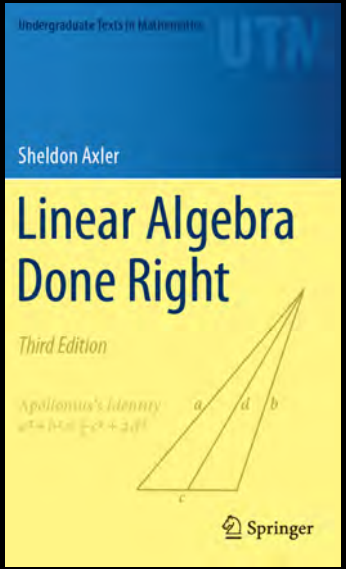
Condition for a direct sum

Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each u_j equal to 0.

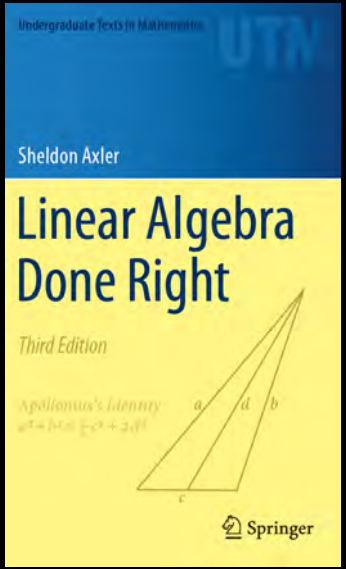
Direct sum of two subspaces

Suppose U and W are subspaces of V. Then $U + W$ is a direct sum if and only if $U \cap W = \{0\}$.

Linear Algebra Done Right, by Sheldon Axler



Span and Linear Independence



Notation

Reminder:

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V denotes a vector space over $\mathbf{F}$.

Fix a positive integer n.

Definition: *list, length*

A *list* of *length* n is an ordered collection of n elements (which might be numbers, other lists, or more abstract entities) separated by commas (and perhaps surrounded by parentheses).

Example: $(6, -4, 7) \in \mathbf{R}^3$.

Example: $(6, -4, 7), (5, -9, 8)$ is a list of length 2 of vectors in $\mathbf{R}^3$.

Linear Combination

Definition: *linear combination*

A *linear combination* of a list $v_1, \ldots, v_m$ of vectors in V is a vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where $a_1, \ldots, a_m \in \mathbf{F}$.

Example: In $\mathbf{F}^3$,

- $(13, -1, 7)$ is a linear combination of $(2, 1, -1), (1, -2, 4)$ because
$$5(2, 1, -1) + 3(1, -2, 4) = (13, -1, 7);$$

- $(13, -1, 6)$ is not a linear combination of $(2, 1, -1), (1, -2, 4)$ because there do not exist numbers $a_1, a_2 \in \mathbf{F}$ such that
$$a_1(2, 1, -1) + a_2(1, -2, 4) = (13, -1, 6).$$

Span

Definition: *span*

The set of all linear combinations of a list of vectors $v_1, \ldots, v_m$ in V is called the *span* of $v_1, \ldots, v_m$, denoted $\operatorname{span}(v_1, \ldots, v_m)$. In other words,

$$\operatorname{span}(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : a_1, \ldots, a_m \in \mathbf{F}\}.$$

Example: The previous example shows that in $\mathbf{F}^3$,

- $(13, -1, 7) \in \operatorname{span}\big((2, 1, -1), (1, -2, 4)\big)$;
- $(13, -1, 6) \notin \operatorname{span}\big((2, 1, -1), (1, -2, 4)\big)$.

Span is the smallest containing subspace

The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.

Finite-Dimensional Vector Space

Definition: *finite-dimensional vector space*

A vector space is called *finite-dimensional* if the span of some list of vectors in it is the entire vector space.

Example: $\mathbf{F}^3$ is finite-dimensional because

$$\mathbf{F}^3 = \operatorname{span}\big((1,0,0),(0,1,0),(0,0,1)\big).$$

Definition: *infinite-dimensional vector space*

A vector space is called *infinite-dimensional* if it is not finite-dimensional.

Example: $\mathbf{F}^\infty$ is infinite-dimensional.

Paul Halmos

Linear Independence

Definition: *linearly independent*

A list $v_1, \ldots, v_m$ of vectors in V is called *linearly independent* if the only choice of $a_1, \ldots, a_m \in \mathbf{F}$ that makes $a_1 v_1 + \cdots + a_m v_m$ equal 0 is $a_1 = \cdots = a_m = 0$.

Examples of linearly independent lists:

- A list v of one vector $v \in V$ is linearly independent if and only if $v \neq 0$.
- A list of two vectors in V is linearly independent if and only if neither vector is a scalar multiple of the other.
- $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)$ is linearly independent in $\mathbf{F}^4$.
- The list $1, x, \ldots, x^m$ is linearly independent in $\mathbf{R}^{\mathbf{R}}$ for each nonnegative integer m.

Linear Dependence

Definition: *linearly dependent*

- A list of vectors in V is called *linearly dependent* if it is not linearly independent.
- In other words, a list $v_1, \ldots, v_m$ of vectors in V is linearly dependent if there exist $a_1, \ldots, a_m \in \mathbf{F}$, not all 0, such that $a_1 v_1 + \cdots + a_m v_m = 0$.

Examples of linearly dependent lists:

- $(2,3,1), (1,-1,2), (7,3,8)$ is linearly dependent in $\mathbf{F}^3$ because
$$2(2,3,1) + 3(1,-1,2) + (-1)(7,3,8) = (0,0,0).$$
- Every list of vectors in V containing the 0 vector is linearly dependent.
- If some vector in a list of vectors in V is a linear combination of the other vectors, then the list is linearly dependent.

Linear Dependence Lemma

Linear Dependence Lemma

Suppose $v_1, \ldots, v_m$ is a linearly dependent list in V. Then there exists $j \in \{1, 2, \ldots, m\}$ such that the following hold:

- $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$;
- if the j^{th} term is removed from $v_1, \ldots, v_m$, the span of the remaining list equals $\operatorname{span}(v_1, \ldots, v_m)$.

Example: $(2, 3, 1), (1, -1, 2), (7, 3, 8)$ is linearly dependent in $\mathbf{F}^3$.

- $(7, 3, 8) \in \operatorname{span}\big((2, 3, 1), (1, -1, 2)\big)$
- $\operatorname{span}\big((2, 3, 1), (1, -1, 2)\big) = \operatorname{span}\big((2, 3, 1), (1, -1, 2), (7, 3, 8)\big)$

Length of Linearly Independent List $\leq$ Length of Spanning List

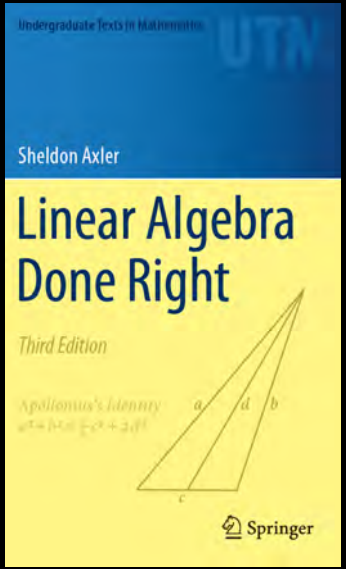
Length of linearly independent list $\leq$ length of spanning list

In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

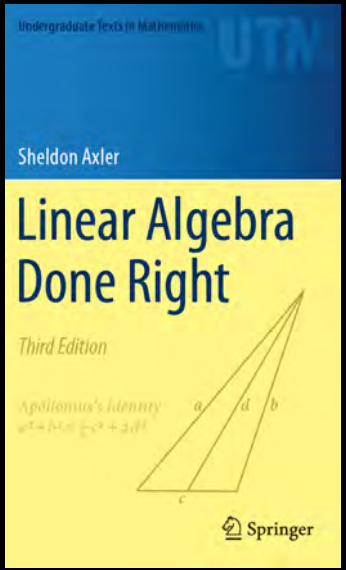
Example: The list $(1, 2, 3), (4, 5, 8), (9, 6, 7), (-3, 2, 8)$ is not linearly independent in $\mathbf{R}^3$ because the list $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ spans $\mathbf{R}^3$.

Example: The list $(1, 2, 3, -5), (4, 5, 8, 3), (9, 6, 7, -1)$ does not span $\mathbf{R}^4$ because the list $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)$ is linearly independent in $\mathbf{R}^4$.

Linear Algebra Done Right, by Sheldon Axler



Bases


Undergraduate Texts in Mathematics
UTM
Sheldon Axler
Linear Algebra Done Right
Third Edition
Apollonius's Identity
a
d
b
c
Springer

Notation

Reminder:

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V denotes a vector space over $\mathbf{F}$.

Definition: $\mathcal{P}_m(\mathbf{F})$

For m a nonnegative integer, $\mathcal{P}_m(\mathbf{F})$ denotes the set of polynomials with coefficients in $\mathbf{F}$ and degree at most m.

Examples:

- $2x^3 - 7x + 5 \in \mathcal{P}_3(\mathbf{R})$.
- $(3 + 2i)z^2 + 4iz + 9 \in \mathcal{P}_{20}(\mathbf{C})$

Definition of Basis

Definition: *basis*

A *basis* of V is a list of vectors in V that is linearly independent and spans V.

Examples of bases:

- The list $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ is a basis of $\mathbf{F}^n$, called the *standard basis* of $\mathbf{F}^n$.
- The list $(1, 1, 0), (0, 0, 1)$ is a basis of $\{(x, x, y) \in \mathbf{F}^3 : x, y \in \mathbf{F}\}$.
- The list $(1, 2), (3, 5)$ is a basis of $\mathbf{F}^2$.
- The list $(1, -1, 0), (1, 0, -1)$ is a basis of
$$\{(x, y, z) \in \mathbf{F}^3 : x + y + z = 0\}.$$
- The list $1, z, \ldots, z^m$ is a basis of $\mathcal{P}_m(\mathbf{F})$.

Non-examples of Bases

- The list $(1, 2, -4), (7, -5, 6)$ is linearly independent in $\mathbf{F}^3$ but is not a basis of $\mathbf{F}^3$ because it does not span $\mathbf{F}^3$.
- The list $(1, 2), (3, 5), (4, 13)$ spans $\mathbf{F}^2$ but is not a basis of $\mathbf{F}^2$ because it is not linearly independent.

Why Bases Are Useful

Basis gives unique representation as linear combination

A list $v_1, \ldots, v_n$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form

$$v = a_1 v_1 + \cdots + a_n v_n,$$

where $a_1, \ldots, a_n \in \mathbf{F}$.

Example: Let $V = \{(x, y, z) \in \mathbf{F}^3 : x + y + z = 0\}$.

The list $(1, -1, 0), (1, 0, -1)$ is a basis of V.

If $(x, y, z) \in V$, then

$$(x, y, z) = -y(1, -1, 0) + (-z)(1, 0, -1).$$

Existence of Bases

Every spanning list contains a basis

Every spanning list in a vector space can be reduced to a basis of the vector space.

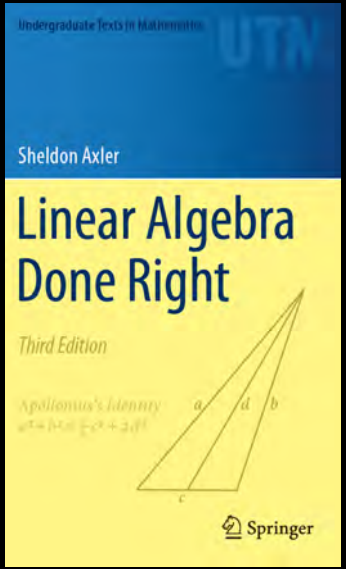
Basis of finite-dimensional vector space

Every finite-dimensional vector space has a basis.

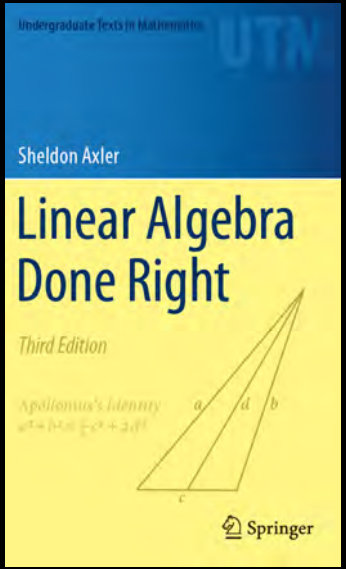
Every linearly independent list extends to a basis

Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.

Linear Algebra Done Right, by Sheldon Axler



Dimension


Undergraduate Texts in Mathematics
Sheldon Axler
Linear Algebra Done Right
Third Edition
Apollonius's Identity
a
d
b
c
Springer

Review

Reminder:

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V denotes a vector space over $\mathbf{F}$.

Definition: *basis*

A *basis* of V is a list of vectors in V that is linearly independent and spans V.

Example: The list $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ is a basis of $\mathbf{F}^n$, called the *standard basis* of $\mathbf{F}^n$.

How should we define the dimension of a vector space?

Basis Length is the Same for All Bases of V

Basis length does not depend on basis

Any two bases of a vector space have the same length.

Proof Suppose B_1 and B_2 are two bases of V.

Then B_1 is linearly independent in V and B_2 spans V, so the length of B_1 is less than or equal to the length of B_2.

Interchanging the roles of B_1 and B_2, we also see that the length of B_2 is less than or equal to the length of B_1.

Thus the length of B_1 equals the length of B_2, as desired. ∎

Definition of Dimension

Definition: *dimension*, $\dim V$

- The *dimension* of a finite-dimensional vector space is the length of any basis of the vector space.
- The dimension of V (if V is finite-dimensional) is denoted by $\dim V$.

Examples:

- $\dim \mathbf{F}^n = n$ because the standard basis of $\mathbf{F}^n$ has length n.
- $\dim \mathcal{P}_m(\mathbf{F}) = m + 1$ because the basis $1, z, \ldots, z^m$ of $\mathcal{P}_m(\mathbf{F})$ has length $m + 1$.

Can Check Just Linear Independence.

Linearly independent list of the right length is a basis

Suppose V is finite-dimensional. Then every linearly independent list of vectors in V with length $\dim V$ is a basis of V.

Proof

Suppose $\dim V = n$ and $v_1, \ldots, v_n$ is linearly independent in V.

The list $v_1, \ldots, v_n$ can be extended to a basis of V.

However, every basis of V has length n, so in this case the extension is the trivial one, meaning that no elements are adjoined to $v_1, \ldots, v_n$.

In other words, $v_1, \ldots, v_n$ is a basis of V, as desired. ∎

Example Checking Just Linear Independence

The list $(5, 7), (4, 3)$ is a basis of $\mathbf{F}^2$.

Proof

This list of two vectors in $\mathbf{F}^2$ is linearly independent because neither vector is a scalar multiple of the other.

Note that $\mathbf{F}^2$ has dimension 2.

Thus this linearly independent list of length 2 is a basis of $\mathbf{F}^2$ (we do not need to bother checking that it spans $\mathbf{F}^2$). ∎

Can Check Just Span

Spanning list of the right length is a basis

Suppose V is finite-dimensional. Then every spanning list of vectors in V with length $\dim V$ is a basis of V.

Proof

Suppose $\dim V = n$ and $v_1, \ldots, v_n$ spans V.

The list $v_1, \ldots, v_n$ can be reduced to a basis of V.

However, every basis of V has length n, so in this case the reduction is the trivial one, meaning that no elements are deleted from $v_1, \ldots, v_n$.

In other words, $v_1, \ldots, v_n$ is a basis of V, as desired. ∎

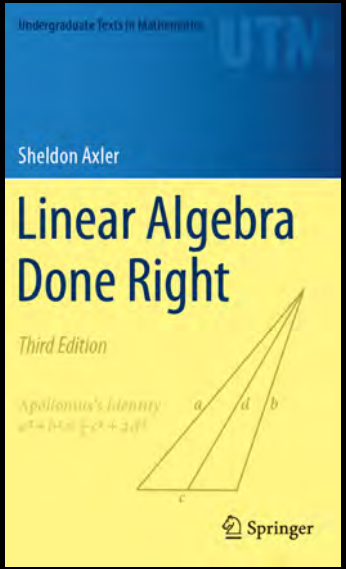
Dimension of a Sum

Dimension of a sum

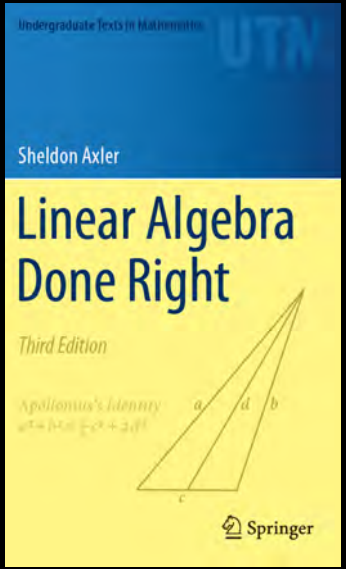
If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Linear Algebra Done Right, by Sheldon Axler



The Vector Space of Linear Maps



Notation

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V and W denote vector spaces over $\mathbf{F}$.

Notation: $\mathcal{P}(\mathbf{F})$

$\mathcal{P}(\mathbf{F})$ is the vector space of all polynomials with coefficients in $\mathbf{F}$.

Definition of Linear Map

Definition: *linear map*

A *linear map* from V to W is a function $T\colon V \to W$ with the following properties:

- additivity: $T(u+v) = Tu + Tv$ for all $u, v \in V$;
- homogeneity: $T(\lambda u) = \lambda(Tu)$ for all $\lambda \in \mathbf{F}$ and all $u \in V$.

Definition of Linear Map

Definition: *linear map*

A *linear map* from V to W is a function $T\colon V \to W$ with the following properties:

- additivity: $T(u_1 + u_2) = Tu_1 + Tu_2$ for all $u_1, u_2 \in V$;
- homogeneity: $T(\lambda u) = \lambda(Tu)$ for all $\lambda \in \mathbf{F}$ and all $u \in V$.

For linear maps, we often use the notation Tu as well as the more standard functional notation $T(u)$.

Notation: $\mathcal{L}(V, W)$

The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$.

Examples of Linear Maps

- Zero: Define $0 \in \mathcal{L}(V, W)$ by $0u = 0$ for all $u \in V$.
- Identity map: Define $I \in \mathcal{L}(V, V)$ by $Iu = u$ for all $u \in V$.
- Differentiation: Define $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ by $Dp = p'$.
- Integration: Define $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathbf{R}\big)$ by $Tp = \int_0^1 p(x)\,dx$.
- Multiplication by x^2: Define $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ by
$$(Tp)(x) = x^2 p(x)$$
for $x \in \mathbf{R}$.
- Backward shift: Define $T \in \mathcal{L}(\mathbf{F}^\infty, \mathbf{F}^\infty)$ by
$$T(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots).$$
- From $\mathbf{R}^3$ to $\mathbf{R}^2$: Define $T \in \mathcal{L}(\mathbf{R}^3, \mathbf{R}^2)$ by
$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z).$$

Linear Map Determined By What It Does On a Basis

Linear maps and basis of domain

Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_n \in W$. Then there exists a unique linear map $T\colon V \to W$ such that

$$Tv_j = w_j$$

for each $j = 1, \ldots, n$.

Proof

Define $T\colon V \to W$ by

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n,$$

where $c_1, \ldots, c_n$ are arbitrary elements of $\mathbf{F}$. ∎

Algebraic Operations on $\mathcal{L}(V, W)$

Definition: *addition and scalar multiplication on* $\mathcal{L}(V, W)$

Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbf{F}$. The *sum* $S + T$ and the *product* λT are the linear maps from V to W defined by

$$(S + T)(u) = Su + Tu \quad \text{and} \quad (\lambda T)(u) = \lambda(Tu)$$

for all $u \in V$.

$\mathcal{L}(V, W)$ *is a vector space*

With the operations of addition and scalar multiplication as defined above, $\mathcal{L}(V, W)$ is a vector space.

Composition of Linear Maps

Definition: *Product of Linear Maps*

If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the *product* $ST \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu)$$

for $u \in U$.

Algebraic properties of products of linear maps

- Associativity: $(T_1 T_2)T_3 = T_1(T_2 T_3)$
- Identity: $TI = IT = T$
- Distributive Properties:
  $(S_1 + S_2)T = S_1 T + S_2 T$ and $S(T_1 + T_2) = ST_1 + ST_2$

0 Goes to 0 Under a Linear Map

Linear maps take 0 _to_ 0

Suppose T is a linear map from V to W. Then $T(0) = 0$.

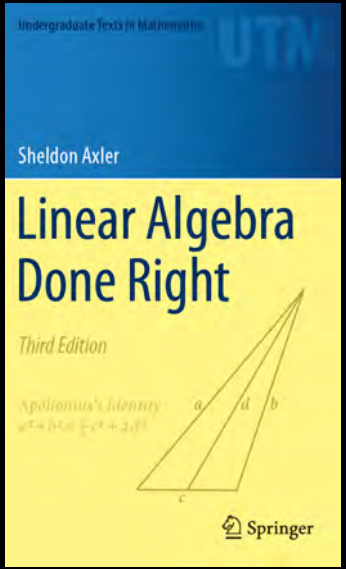
Proof

By additivity, we have

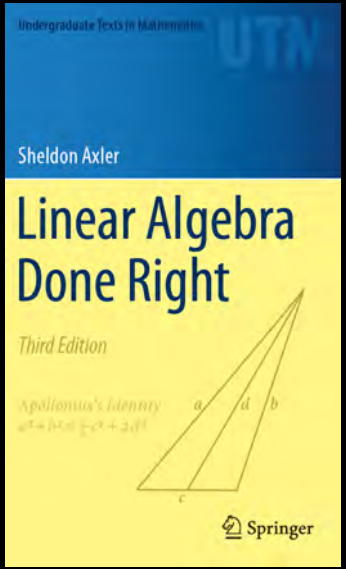
$$T(0) = T(0 + 0) = T(0) + T(0).$$

Add the additive inverse of $T(0)$ to each side of the equation above to conclude that $0 = T(0)$. ∎

Linear Algebra Done Right, by Sheldon Axler



Null Spaces and Ranges



Notation

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V and W denote vector spaces over $\mathbf{F}$.

Null Space

Definition: *null space,* $\operatorname{null} T$

For $T \in \mathcal{L}(V, W)$, the *null space* of T, denoted $\operatorname{null} T$, is the subset of V consisting of those vectors that T maps to 0:

$$\operatorname{null} T = \{u \in V : Tu = 0\}.$$

- Suppose T is the zero map from V to W; in other words, $Tu = 0$ for every $u \in V$. Then $\operatorname{null} T = V$.
- Suppose $\varphi \in \mathcal{L}(\mathbf{C}^3, \mathbf{C})$ is defined by $\varphi(z_1, z_2, z_3) = z_1 + 2z_2 + 3z_3$. Then $\operatorname{null} \varphi = \{(z_1, z_2, z_3) \in \mathbf{C}^3 : z_1 + 2z_2 + 3z_3 = 0\}$.
- Suppose $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is the differentiation map defined by $Dp = p'$. The only functions whose derivative equals the zero function are the constant functions. Thus the null space of D equals the set of constant functions.

Injectivity

The null space is a subspace

Suppose $T \in \mathcal{L}(V, W)$. Then $\operatorname{null} T$ is a subspace of V.

Definition: ***injective***

A function $T\colon V \to W$ is called *injective* if $Tu = Tv$ implies $u = v$.

Injectivity is equivalent to null space equals $\{0\}$

Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if $\operatorname{null} T = \{0\}$.

Range

Definition: *range*

For $T \in \mathcal{L}(V, W)$, the *range* of T is the subset of W consisting of those vectors that are of the form Tu for some $u \in V$:

$$\operatorname{range} T = \{Tu : u \in V\}.$$

- Suppose T is the zero map from V to W; in other words, $Tu = 0$ for every $u \in V$. Then $\operatorname{range} T = \{0\}$.
- Suppose $T \in \mathcal{L}(\mathbf{R}^2, \mathbf{R}^3)$ is defined by $T(x, y) = (2x, 5y, x + y)$, then $\operatorname{range} T = \{(2x, 5y, x + y) : x, y \in \mathbf{R}\}$. A basis of $\operatorname{range} T$ is $(2, 0, 1), (0, 5, 1)$.
- Suppose $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is the differentiation map defined by $Dp = p'$. Because for every polynomial $q \in \mathcal{P}(\mathbf{R})$ there exists a polynomial $p \in \mathcal{P}(\mathbf{R})$ such that $p' = q$, the range of D is $\mathcal{P}(\mathbf{R})$.

Surjectivity

The range is a subspace

If $T \in \mathcal{L}(V, W)$, then $\operatorname{range} T$ is a subspace of W.

Definition: ***surjective***

A function $T\colon V \to W$ is called *surjective* if its range equals W.

Fundamental Theorem of Linear Maps

Fundamental Theorem of Linear Maps

Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Proof

Let $u_1, \ldots, u_m$ be a basis of $\operatorname{null} T$; thus $\dim \operatorname{null} T = m$.
The linearly independent list $u_1, \ldots, u_m$ can be extended to a basis

$$u_1, \ldots, u_m, v_1, \ldots, v_n$$

of V. Thus $\dim V = m + n$.
To complete the proof, we need only show that $\dim \operatorname{range} T = n$.
We do this by proving that $Tv_1, \ldots, Tv_n$ is a basis of $\operatorname{range} T$. ∎

Consequence of Fundamental Theorem

A map to a smaller dimensional space is not injective

Suppose V and W are finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective.

Proof Suppose $T \in \mathcal{L}(V, W)$. Because

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$$

and

$$\begin{aligned} \dim V &> \dim W \\ &\geq \dim \operatorname{range} T, \end{aligned}$$

we have $\dim \operatorname{null} T > 0$.
Thus T is not injective. ∎

Consequence of Fundamental Theorem

A map to a larger dimensional space is not surjective

Suppose V and W are finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map from V to W is surjective.

Proof Suppose $T \in \mathcal{L}(V, W)$. Because

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T,$$

we have

$$\begin{aligned} \dim \operatorname{range} T &\leq \dim V \\ &< \dim W. \end{aligned}$$

Thus T is not surjective. ∎

Homogeneous Equations

Fix positive integers m and n, and let $A_{j,k} \in \mathbf{F}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$. Consider the homogeneous system of linear equations

$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= 0 \\ &\vdots \\ \sum_{k=1}^{n} A_{m,k} x_k &= 0. \end{aligned}$$

Homogeneous system of linear equations

A homogeneous system of linear equations with more variables than equations has nonzero solutions.

Do there exist solutions other than $x_1 = \cdots = x_n = 0$?
Define $T \colon \mathbf{F}^n \to \mathbf{F}^m$ by

$$T(x_1, \ldots, x_n) = \Big(\sum_{k=1}^{n} A_{1,k} x_k, \ldots, \sum_{k=1}^{n} A_{m,k} x_k\Big).$$

The equation $T(x_1, \ldots, x_n) = 0$ is the same as the homogeneous system of linear equations above. Is T injective?

Inhomogeneous Equations

Fix positive integers m and n, and let $A_{j,k} \in \mathbf{F}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$. Consider the inhomogeneous system of linear equations

$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= c_1 \\ &\vdots \\ \sum_{k=1}^{n} A_{m,k} x_k &= c_m. \end{aligned}$$

Inhomogeneous system of linear equations

An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of the constant terms.

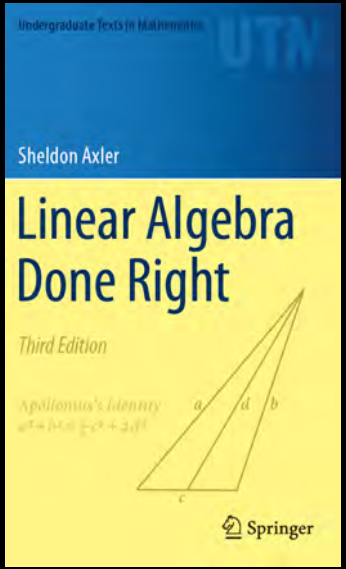
Is there some choice of $c_1, \ldots, c_m \in \mathbf{F}$ such that no solution exists?
Define $T \colon \mathbf{F}^n \to \mathbf{F}^m$ by

$$T(x_1, \ldots, x_n) = \big(\sum_{k=1}^{n} A_{1,k} x_k, \ldots, \sum_{k=1}^{n} A_{m,k} x_k\big).$$

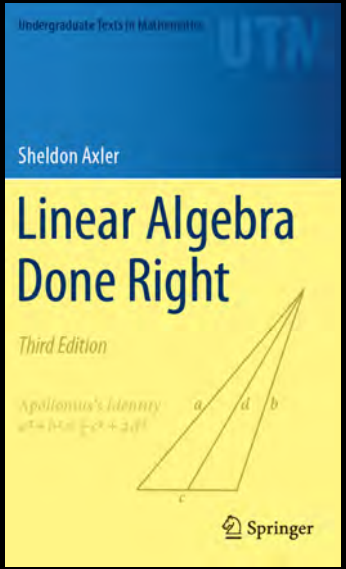
The equation $T(x_1, \ldots, x_n) = (c_1, \ldots, c_m)$ is the same as the inhomogeneous system of linear equations above. Is T surjective?

Carl Friedrich Gauss (1777 – 1855)

Linear Algebra Done Right, by Sheldon Axler



Matrices, part 1: The Matrix of a Linear Map



Notation

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V and W denote vector spaces over $\mathbf{F}$.

Definition and Notation for Matrix

Definition: *matrix,* $A_{j,k}$

Let m and n denote positive integers. An m-by-n *matrix* A is a rectangular array of elements of $\mathbf{F}$ with m rows and n columns:

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix}.$$

The notation $A_{j,k}$ denotes the entry in row j, column k of A.

The first index refers to the row number and the second index refers to the column number.

Thus $A_{2,3}$ refers to the entry in the second row, third column of A.

Example: Suppose $A = \begin{pmatrix} 8 & 4 & 5 \\ 1 & 9 & 7 \end{pmatrix}$. Then $A_{2,3} = 7$.

Matrix of a Linear Map

Definition: *matrix of a linear map,* $\mathcal{M}(T)$

Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. The *matrix of* T with respect to these bases is the m-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

If the bases are not clear from the context, then the notation $\mathcal{M}\big(T, (v_1, \ldots, v_n), (w_1, \ldots, w_m)\big)$ is used.

Understanding the Matrix of a Linear Map

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & \dots & v_k & \dots & v_n \end{array} \\ \begin{pmatrix} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{pmatrix} \end{array}.$$

The k^{th} column of $\mathcal{M}(T)$ consists of the scalars needed to write Tv_k as a linear combination of $w_1, \dots, w_m$:

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j.$$

The picture above should remind you that Tv_k can be computed from $\mathcal{M}(T)$ by multiplying each entry in the k^{th} column by the corresponding w_j from the left column, and then adding up the resulting vectors.

Examples of the Matrix of a Linear Map

Example: Suppose $T \in \mathcal{L}(\mathbf{F}^2, \mathbf{F}^3)$ is defined by

$$T(x, y) = (x + 3y, 2x + 5y, 7x + 9y).$$

Because $T(1, 0) = (1, 2, 7)$ and $T(0, 1) = (3, 5, 9)$, the matrix of T with respect to the standard bases is the 3-by-2 matrix

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}.$$

Example: Suppose $D \in \mathcal{L}\big(\mathcal{P}_3(\mathbf{R}), \mathcal{P}_2(\mathbf{R})\big)$ is the differentiation map defined by $Dp = p'$. Because $(x^n)' = nx^{n-1}$, the matrix of D with respect to the standard bases of $\mathcal{P}_3(\mathbf{R})$ and $\mathcal{P}_2(\mathbf{R})$ is the 3-by-4 matrix

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Addition of Matrices

Definition: *matrix addition*

The *sum of two matrices of the same size* is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix} + \begin{pmatrix} C_{1,1} & \dots & C_{1,n} \\ \vdots & & \vdots \\ C_{m,1} & \dots & C_{m,n} \end{pmatrix}$$

$$= \begin{pmatrix} A_{1,1} + C_{1,1} & \dots & A_{1,n} + C_{1,n} \\ \vdots & & \vdots \\ A_{m,1} + C_{m,1} & \dots & A_{m,n} + C_{m,n} \end{pmatrix}.$$

In other words, $(A + C)_{j,k} = A_{j,k} + C_{j,k}$.

In the following result, the assumption is that the same bases are used for $\mathcal{M}(S + T)$, $\mathcal{M}(S)$, and $\mathcal{M}(T)$.

The matrix of the sum of linear maps

Suppose $S, T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$.

Scalar Multiplication of Matrices

Definition: *scalar multiplication of a matrix*

The product of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \dots & \lambda A_{1,n} \\ \vdots & & \vdots \\ \lambda A_{m,1} & \dots & \lambda A_{m,n} \end{pmatrix}.$$

In other words, $(\lambda A)_{j,k} = \lambda A_{j,k}$.

In the following result, the assumption is that the same bases are used for $\mathcal{M}(\lambda T)$ and $\mathcal{M}(T)$.

The matrix of a scalar times a linear map

Suppose $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.

The Vector Space of Matrices

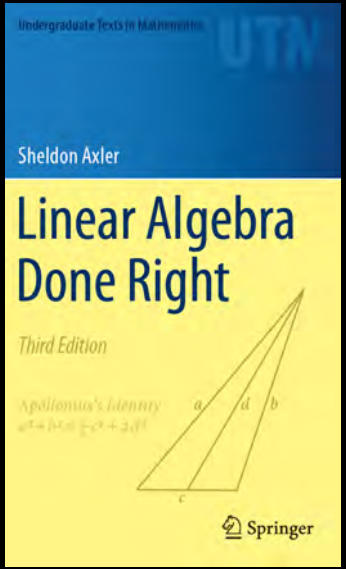
Notation: $\mathbf{F}^{m,n}$

For m and n positive integers, the set of all m-by-n matrices with entries in $\mathbf{F}$ is denoted by $\mathbf{F}^{m,n}$.

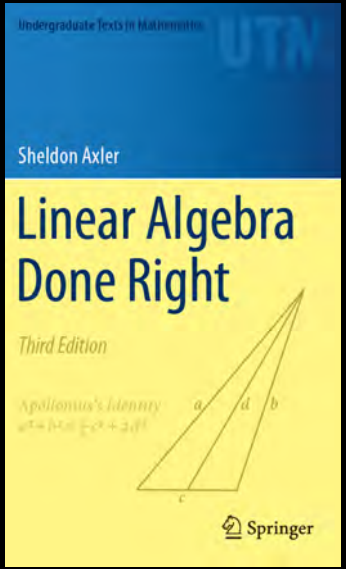
$\dim \mathbf{F}^{m,n} = mn$

Suppose m and n are positive integers. With addition and scalar multiplication defined as above, $\mathbf{F}^{m,n}$ is a vector space with dimension mn.

Linear Algebra Done Right, by Sheldon Axler



Matrices, part 2: Matrix Multiplication



Motivation for Definition of Matrix Multiplication

Suppose $v_1, \ldots, v_n$ is a basis of V
and $w_1, \ldots, w_m$ is a basis of W.
Suppose $u_1, \ldots, u_p$ is a basis of U.

Consider linear maps $T\colon U \to V$
and $S\colon V \to W$.

Does $\mathcal{M}(ST)$ equal $\mathcal{M}(S)\mathcal{M}(T)$?

Suppose $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$.

For $1 \le k \le p$, we have

$$\begin{aligned}(ST)u_k &= S\Big(\sum_{r=1}^{n} C_{r,k} v_r\Big)\\ &= \sum_{r=1}^{n} C_{r,k} S v_r\\ &= \sum_{r=1}^{n} C_{r,k} \sum_{j=1}^{m} A_{j,r} w_j\\ &= \sum_{j=1}^{m} \Big(\sum_{r=1}^{n} A_{j,r} C_{r,k}\Big) w_j.\end{aligned}$$

Thus $\mathcal{M}(ST)$ is the m-by-p matrix whose entry in row j, column k, is

$$\sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Definition of Matrix Multiplication

Definition: *matrix multiplication*

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then AC is defined to be the m-by-p matrix whose entry in row j, column k, is given by the following equation:

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Thus the entry in row j, column k, of AC is computed by taking row j of A and column k of C, multiplying together corresponding entries, and then summing.

Example: $$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{pmatrix}$$

In the following result, the same basis of V is used in considering $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, the same basis of W is used in considering $S \in \mathcal{L}(V, W)$ and $ST \in \mathcal{L}(U, W)$, and the same basis of U is used in considering $T \in \mathcal{L}(U, V)$ and $ST \in \mathcal{L}(U, W)$.

The matrix of the product of linear maps

If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.

The proof of the result above is the calculation that was done as motivation before the definition of matrix multiplication.

Notation for Row and Column of Matrix

Notation: $A_{j,\cdot}$, $A_{\cdot,k}$

Suppose A is an m-by-n matrix.

- If $1 \leq j \leq m$, then $A_{j,\cdot}$ denotes the 1-by-n matrix consisting of row j of A.
- If $1 \leq k \leq n$, then $A_{\cdot,k}$ denotes the m-by-1 matrix consisting of column k of A.

Example: If $A = \begin{pmatrix} 8 & 4 & 5 \\ 1 & 9 & 7 \end{pmatrix}$, then $A_{2,\cdot}$ is row 2 of A and $A_{\cdot,2}$ is column 2 of A. In other words,

$$A_{2,\cdot} = \begin{pmatrix} 1 & 9 & 7 \end{pmatrix} \quad \text{and} \quad A_{\cdot,2} = \begin{pmatrix} 4 \\ 9 \end{pmatrix}.$$

Alternative Ways to Think about Matrix Multiplication

Example: $\begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 26 \end{pmatrix} = 26.$

Entry of matrix product equals row times column

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{j,k} = A_{j,\cdot}\, C_{\cdot,k}$$

for $1 \leq j \leq m$ and $1 \leq k \leq p$.

Alternative Ways to Think about Matrix Multiplication

Column of matrix product equals matrix times column

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{\cdot,k} = AC_{\cdot,k}$$

for $1 \leq k \leq p$.

Row of matrix product equals row times matrix

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{j,\cdot} = A_{j,\cdot}C$$

for $1 \leq k \leq p$.

Matrix Times a Column Vector

Linear combination of columns

Suppose A is an m-by-n matrix and $c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ is an n-by-1 matrix.

Then $Ac = c_1 A_{\cdot,1} + \cdots + c_n A_{\cdot,n}$.

In other words, Ac is a linear combination of the columns of A, with the scalars that multiply the columns coming from c.

Example:

$$\begin{pmatrix} 8 & 4 & 5 \\ 1 & 9 & 7 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} = 3 \begin{pmatrix} 8 \\ 1 \end{pmatrix} + (-1) \begin{pmatrix} 4 \\ 9 \end{pmatrix} + 2 \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

$$= \begin{pmatrix} 30 \\ 8 \end{pmatrix}$$

Row Vector Times a Matrix

Linear combination of rows

Suppose $a = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$ is a 1-by-n matrix and C is an n-by-p matrix. Then

$$aC = a_1 C_{1,\cdot} + \cdots + a_n C_{n,\cdot} .$$

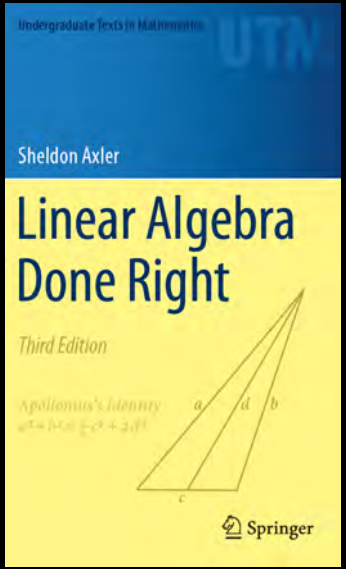
In other words, aC is a linear combination of the rows of C, with the scalars that multiply the rows coming from a.

Example:

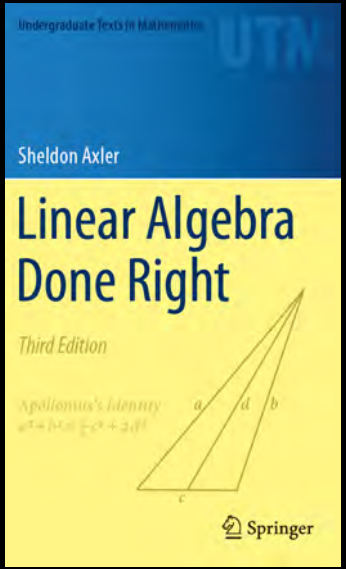
$$\begin{pmatrix} 3 & -1 \end{pmatrix} \begin{pmatrix} 8 & 4 & 5 \\ 1 & 9 & 7 \end{pmatrix} = 3 \begin{pmatrix} 8 & 4 & 5 \end{pmatrix} + (-1) \begin{pmatrix} 1 & 9 & 7 \end{pmatrix}$$

$$= \begin{pmatrix} 23 & 3 & 8 \end{pmatrix}$$

Linear Algebra Done Right, by Sheldon Axler



Invertibility and Isomorphic Vector Spaces



Notation

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V and W denote vector spaces over $\mathbf{F}$.

Invertible Linear Maps

Definition: *invertible*

A linear map $T \in \mathcal{L}(V, W)$ is called *invertible* if there exists a linear map $S \in \mathcal{L}(W, V)$ such that ST equals the identity map on V and TS equals the identity map on W.

Definition: *inverse*

A linear map $S \in \mathcal{L}(W, V)$ satisfying $ST = I$ and $TS = I$ is called an *inverse* of T (note that the first I is the identity map on V and the second I is the identity map on W).

Inverse is unique

An invertible linear map has a unique inverse.

Invertibility

Notation: T^{-1}

If T is invertible, then its inverse is denoted by T^{-1}. In other words, if $T \in \mathcal{L}(V, W)$ is invertible, then T^{-1} is the unique element of $\mathcal{L}(W, V)$ such that $T^{-1}T = I$ and $TT^{-1} = I$.

Invertibility is equivalent to injectivity and surjectivity

A linear map is invertible if and only if it is injective and surjective.

Dimension Determines Whether Vector Spaces Are Isomorphic

Definition: *isomorphism, isomorphic*

- An *isomorphism* is an invertible linear map.
- Two vector spaces are called *isomorphic* if there is an isomorphism from one vector space onto the other one.

Think of an isomorphism as a relabeling.

Dimension shows whether vector spaces are isomorphic

Two finite-dimensional vector spaces over $\mathbf{F}$ are isomorphic if and only if they have the same dimension.

The Dimension of $\mathcal{L}(V, W)$

$\mathcal{L}(V, W)$ *and* $\mathbf{F}^{m,n}$ *are isomorphic*

Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Then $\mathcal{M}$ is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$.

$\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$

Suppose V and W are finite-dimensional. Then $\mathcal{L}(V, W)$ is finite-dimensional and

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Matrix of a Vector

Definition: *matrix of a vector,* $\mathcal{M}(u)$

Suppose $u \in V$ and $v_1, \ldots, v_n$ is a basis of V. The *matrix of* u with respect to this basis is the n-by-1 matrix

$$\mathcal{M}(u) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

where $c_1, \ldots, c_n$ are the scalars such that

$$u = c_1 v_1 + \cdots + c_n v_n.$$

The matrix of $(5, 8, 2)$ with respect to the standard basis of $\mathbf{F}^3$ is

$$\begin{pmatrix} 5 \\ 8 \\ 2 \end{pmatrix}.$$

The matrix of $3 - 7x + 5x^2$ with respect to the standard basis of $\mathcal{P}_2(\mathbf{R})$ is

$$\begin{pmatrix} 3 \\ -7 \\ 5 \end{pmatrix}.$$

Linear Maps Act Like Matrix Multiplication

Linear maps act like matrix multiplication

Suppose $T \in \mathcal{L}(V, W)$ and $u \in V$. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Then

$$\mathcal{M}(Tu) = \mathcal{M}(T)\mathcal{M}(u).$$

Operators

Definition: *operator,* $\mathcal{L}(V)$

- A linear map from a vector space to itself is called an *operator*.
- The notation $\mathcal{L}(V)$ denotes the set of all operators on V. In other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

For $T \in \mathcal{L}(V)$ and $\dim V < \infty$, Injectivity $\iff$ Surjectivity

Injectivity is equivalent to surjectivity in finite dimensions

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent:

- T is invertible;
- T is injective;
- T is surjective.

Proof The Fundamental Theorem of Linear Maps states that

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

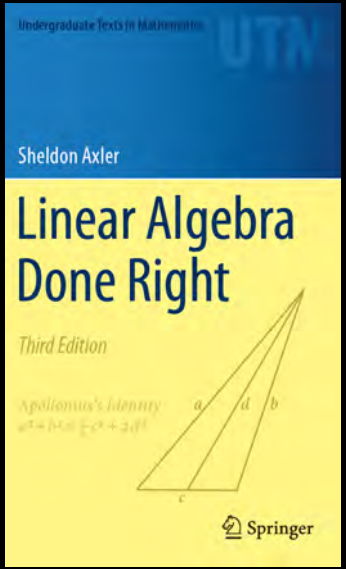
Thus

$$\dim \operatorname{null} T = 0 \iff \dim \operatorname{range} T = \dim V$$

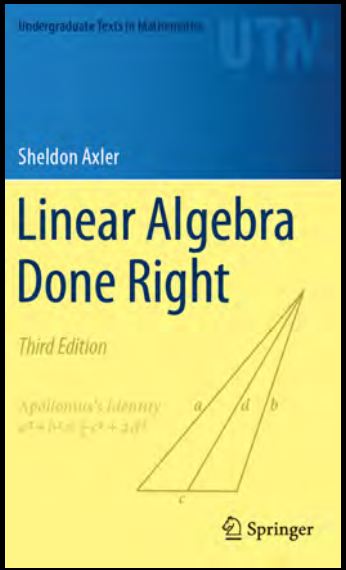
or

$$T \text{ is injective} \iff T \text{ is surjective.}$$

Linear Algebra Done Right, by Sheldon Axler



Products and Quotients of Vector Spaces, part 1: Products



Notation

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V and W denote vector spaces over $\mathbf{F}$.

Product of Vector Spaces

Definition: *product of vector spaces*

Suppose $V_1, \ldots, V_m$ are vector spaces over $\mathbf{F}$.

- The *product* $V_1 \times \cdots \times V_m$ is defined by
$$V_1 \times \cdots \times V_m = \{(u_1, \ldots, u_m) : u_1 \in V_1, \ldots, u_m \in V_m\}.$$
- Addition on $V_1 \times \cdots \times V_m$ is defined by
$$(u_1, \ldots, u_m) + (w_1, \ldots, w_m) = (u_1 + w_1, \ldots, u_m + w_m).$$
- Scalar multiplication on $V_1 \times \cdots \times V_m$ is defined by
$$\lambda(u_1, \ldots, u_m) = (\lambda u_1, \ldots, \lambda u_m).$$

Product of vector spaces is a vector space

Suppose $V_1, \ldots, V_m$ are vector spaces over $\mathbf{F}$. Then $V_1 \times \cdots \times V_m$ is a vector space over $\mathbf{F}$.

Example of Product of Vector Spaces

$$\mathbf{R}^2 \times \mathbf{R}^3 = \big\{\big((x_1, x_2), (x_3, x_4, x_5)\big) : x_1, x_2, x_3, x_4, x_5 \in \mathbf{R}\big\}$$

$$\mathbf{R}^5 = \{(x_1, x_2, x_3, x_4, x_5) : x_1, x_2, x_3, x_4, x_5 \in \mathbf{R}\}$$

Is $\mathbf{R}^2 \times \mathbf{R}^3 = \mathbf{R}^5$? No!

The linear map $T\colon \mathbf{R}^2 \times \mathbf{R}^3 \to \mathbf{R}^5$ defined by

$$T\big((x_1, x_2), (x_3, x_4, x_5)\big) = (x_1, x_2, x_3, x_4, x_5)$$

is clearly an isomorphism of $\mathbf{R}^2 \times \mathbf{R}^3$ onto $\mathbf{R}^5$. Thus these two vector spaces are isomorphic.

Dimension of a Product

Dimension of a product is the sum of dimensions

Suppose $V_1, \ldots, V_m$ are finite-dimensional vector spaces. Then $V_1 \times \cdots \times V_m$ is finite-dimensional and

$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m.$$

Products and direct sums

Suppose that $U_1, \ldots, U_m$ are subspaces of V. Define a linear map $\Gamma : U_1 \times \cdots \times U_m \to U_1 + \cdots + U_m$ by

$$\Gamma(u_1, \ldots, u_m) = u_1 + \cdots + u_m.$$

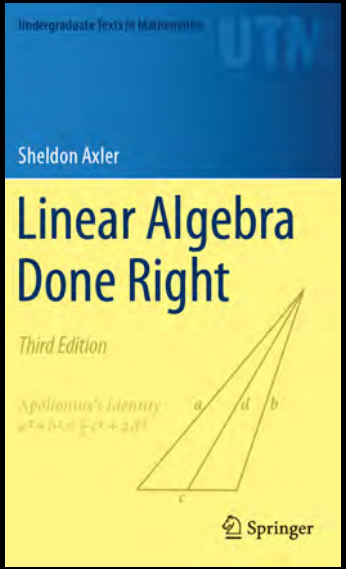
Then $U_1 + \cdots + U_m$ is a direct sum if and only if Γ is injective.

A sum is a direct sum if and only if dimensions add up

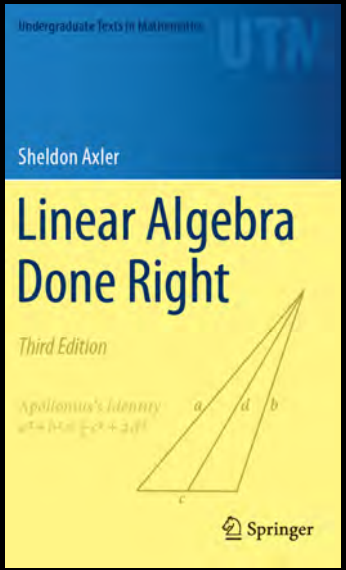
Suppose V is finite-dimensional and $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m.$$

Linear Algebra Done Right, by Sheldon Axler



Products and Quotients of Vector Spaces, part 2: Quotients



The Sum of a Vector and a Subspace

Definition: $v + U$

Suppose $v \in V$ and U is a subspace of V. Then $v + U$ is the subset of V defined by

$$v + U = \{v + u : u \in U\}.$$

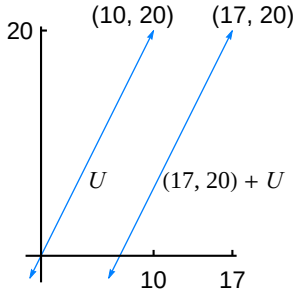
Example: Suppose

$$U = \{(x, 2x) \in \mathbf{R}^2 : x \in \mathbf{R}\}.$$

Then U is the line in $\mathbf{R}^2$ through the origin with slope 2. Thus

$$(17, 20) + U$$

is the line in $\mathbf{R}^2$ that contains the point $(17, 20)$ and has slope 2.



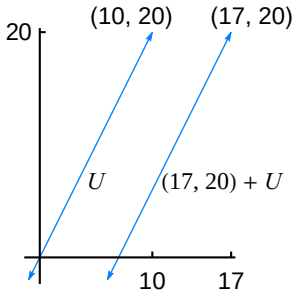
Affine Subsets

Definition: *affine subset, parallel*

- An *affine subset* of V is a subset of V of the form $v + U$ for some $v \in V$ and some subspace U of V.
- For $v \in V$ and U a subspace of V, the affine subset $v + U$ is said to be *parallel* to U.

If $U = \{(x, 2x) \in \mathbf{R}^2 : x \in \mathbf{R}\}$, then all the lines in $\mathbf{R}^2$ with slope 2 are parallel to U.

If $U = \{(x, y, 0) \in \mathbf{R}^3 : x, y \in \mathbf{R}\}$, then the affine subsets of $\mathbf{R}^3$ parallel to U are the planes in $\mathbf{R}^3$ that are parallel to the xy-plane U in the usual sense.



Quotient Spaces

Definition: *quotient space,* V/U

Suppose U is a subspace of V. Then the *quotient space* V/U is the set of all affine subsets of V parallel to U. In other words,

$$V/U = \{v + U : v \in V\}.$$

Examples:

- If $U = \{(x, 2x) \in \mathbf{R}^2 : x \in \mathbf{R}\}$, then $\mathbf{R}^2/U$ is the set of all lines in $\mathbf{R}^2$ that have slope 2.
- If U is a line in $\mathbf{R}^3$ containing the origin, then $\mathbf{R}^3/U$ is the set of all lines in $\mathbf{R}^3$ parallel to U.
- If U is a plane in $\mathbf{R}^3$ containing the origin, then $\mathbf{R}^3/U$ is the set of all planes in $\mathbf{R}^3$ parallel to U.

Parallel Affine Subsets Are Either Equal or Disjoint

Two affine subsets parallel to U are equal or disjoint

Suppose U is a subspace of V and $v, w \in V$. Then the following are equivalent:

(a) $v - w \in U$;

(b) $v + U = w + U$;

(c) $(v + U) \cap (w + U) \neq \varnothing$.

Making a Quotient Space a Vector Space

Definition: *addition and scalar multiplication on* V/U

Suppose U is a subspace of V. Then *addition* and *scalar multiplication* are defined on V/U by

$$(v+U)+(w+U)=(v+w)+U$$
$$\lambda(v+U)=(\lambda v)+U$$

for $v, w \in V$ and $\lambda \in \mathbf{F}$.

Quotient space is a vector space

Suppose U is a subspace of V. Then V/U, with the operations of addition and scalar multiplication as defined above, is a vector space.

Quotient Map

Definition: *quotient map,* π

Suppose U is a subspace of V. The *quotient map* π is the linear map $\pi : V \to V/U$ defined by

$$\pi(v) = v + U$$

for $v \in V$.

Dimension of a quotient space

Suppose V is finite-dimensional and U is a subspace of V. Then

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

Proof

$$\begin{aligned} \dim V &= \dim \operatorname{null} \pi + \dim \operatorname{range} \pi \\ &= \dim U + \dim V/U \end{aligned}$$

The Induced Map on $V/(\operatorname{null} T)$

Definition: $\tilde{T}$

Suppose $T \in \mathcal{L}(V, W)$. Define $\tilde{T}\colon V/(\operatorname{null} T) \to W$ by

$$\tilde{T}(v + \operatorname{null} T) = Tv.$$

Null space and range of $\tilde{T}$

Suppose $T \in \mathcal{L}(V, W)$. Then

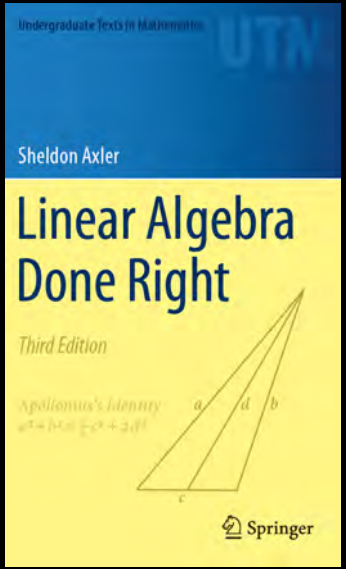
(a) $\tilde{T}$ is a linear map from $V/(\operatorname{null} T)$ to W;

(b) $\tilde{T}$ is injective;

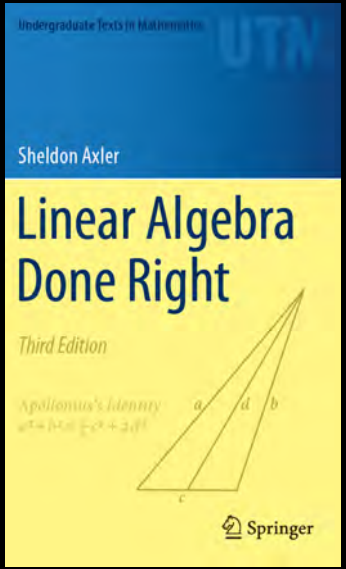
(c) $\operatorname{range} \tilde{T} = \operatorname{range} T$;

(d) $V/(\operatorname{null} T)$ is isomorphic to $\operatorname{range} T$.

Linear Algebra Done Right, by Sheldon Axler



Duality, part 1: Dual Bases and Dual Maps



Notation

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

V and W denote vector spaces over $\mathbf{F}$.

Linear Functionals

Definition: *linear functional*

A *linear functional* on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

Examples:

- Define $\varphi\colon \mathbf{R}^3 \to \mathbf{R}$ by $\varphi(x, y, z) = 4x - 5y + 2z$. Then φ is a linear functional on $\mathbf{R}^3$.
- Fix $(b_1, \ldots, b_n) \in \mathbf{C}^n$. Define $\varphi\colon \mathbf{C}^n \to \mathbf{C}$ by
$$\varphi(z_1, \ldots, z_n) = b_1 z_1 + \cdots + b_n z_n.$$
Then φ is a linear functional on $\mathbf{C}^n$.
- Define $\varphi\colon \mathcal{P}(\mathbf{R}) \to \mathbf{R}$ by $\varphi(p) = 3p''(5) + 7p(4)$. Then φ is a linear functional on $\mathcal{P}(\mathbf{R})$.
- Define $\varphi\colon \mathcal{P}(\mathbf{R}) \to \mathbf{R}$ by $\varphi(p) = \int_0^1 p(x)\, dx$. Then φ is a linear functional on $\mathcal{P}(\mathbf{R})$.

The Dual Space and Its Dimension

Definition: *dual space,* V'

The *dual space* of V, denoted V', is the vector space of all linear functionals on V. In other words, $V' = \mathcal{L}(V, \mathbf{F})$.

$\dim V' = \dim V$

Suppose V is finite-dimensional. Then V' is also finite-dimensional and $\dim V' = \dim V$.

Proof

$$\begin{aligned} \dim V' &= \dim \mathcal{L}(V, \mathbf{F}) \\ &= (\dim V)(\dim \mathbf{F}) \\ &= \dim V \end{aligned}$$

The Dual Basis

Definition: *dual basis*

If $v_1, \ldots, v_n$ is a basis of V, then the *dual basis* of $v_1, \ldots, v_n$ is the list $\varphi_1, \ldots, \varphi_n$ of elements of V', where each φ_j is the linear functional on V such that

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

Example: Consider the standard basis $e_1, \ldots, e_n$ of $\mathbf{F}^n$. For $1 \leq j \leq n$, define φ_j by

$$\varphi_j(x_1, \ldots, x_n) = x_j$$

for $(x_1, \ldots, x_n) \in \mathbf{F}^n$. Clearly

$$\varphi_j(e_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

Thus $\varphi_1, \ldots, \varphi_n$ is the dual basis of the standard basis $e_1, \ldots, e_n$ of $\mathbf{F}^n$.

Dual basis is a basis of the dual space

Suppose $\dim V < \infty$. Then the dual basis of a basis of V is a basis of V'.

The Dual Map

Definition: *dual map,* T'

If $T \in \mathcal{L}(V, W)$, then the *dual map* of T is the linear map $T' \in \mathcal{L}(W', V')$ defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$.

Example: Define $D\colon \mathcal{P}(\mathbf{R}) \to \mathcal{P}(\mathbf{R})$ by $Dp = p'$.

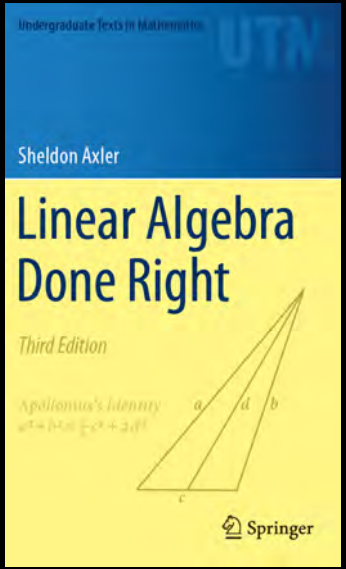
- Suppose φ is the linear functional on $\mathcal{P}(\mathbf{R})$ defined by $\varphi(p) = p(3)$. Then $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ given by
$$\big(D'(\varphi)\big)(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = p'(3).$$
Thus $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ that takes p to $p'(3)$.
- Suppose φ is the linear functional on $\mathcal{P}(\mathbf{R})$ defined by $\varphi(p) = \int_0^1 p$. Then $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ given by
$$\big(D'(\varphi)\big)(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = \int_0^1 p' = p(1) - p(0).$$
Thus $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ that takes p to $p(1) - p(0)$.

Algebraic Properties of Dual Maps

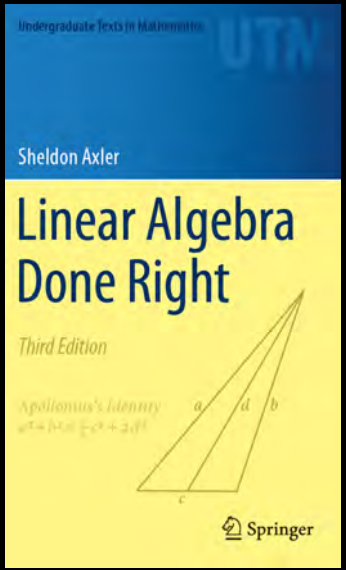
Algebraic properties of dual maps

- $(S+T)' = S' + T'$ for all $S, T \in \mathcal{L}(V, W)$.
- $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbf{F}$ and all $T \in \mathcal{L}(V, W)$.
- $(ST)' = T'S'$ for all $T \in \mathcal{L}(U, V)$ and all $S \in \mathcal{L}(V, W)$.

Linear Algebra Done Right, by Sheldon Axler



Duality, part 2: Annihilators and the Matrix of a Dual Map



Notation

Definition: *annihilator,* U^0

For $U \subset V$, the *annihilator* of U, denoted U^0, is defined by

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\}.$$

Example: $\{0\}^0 = V'$ and $V^0 = \{0\}$.
Example: Suppose $U = x^2\mathcal{P}(\mathbf{R}) \subset \mathcal{P}(\mathbf{R})$. Let φ be the linear functional on $\mathcal{P}(\mathbf{R})$ defined by $\varphi(p) = p'(0)$. Then $\varphi \in U^0$.

The annihilator is a subspace

Suppose $U \subset V$. Then U^0 is a subspace of V'.

Dimension of the annihilator

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U + \dim U^0 = \dim V.$$

The Null Space and Range of a Dual Map

The null space and range of T'

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

- $\operatorname{null} T' = (\operatorname{range} T)^0$;
- $\operatorname{range} T' = (\operatorname{null} T)^0$.

T surjective is equivalent to T' injective

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

- T is surjective if and only if T' is injective;
- T is injective if and only if T' is surjective.

$\dim \operatorname{range} T' = \dim \operatorname{range} T$

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

$$\dim \operatorname{range} T' = \dim \operatorname{range} T.$$

Proof

$$\begin{aligned}
\dim \operatorname{range} T' &= \dim W' - \dim \operatorname{null} T' \\
&= \dim W - \dim(\operatorname{range} T)^0 \\
&= \dim \operatorname{range} T
\end{aligned}$$

The Transpose of a Matrix

Definition: *transpose,* A^{t}

The *transpose* of a matrix A, denoted A^{t}, is the matrix obtained from A by interchanging the rows and columns. More specifically, if A is an m-by-n matrix, then A^{t} is the n-by-m matrix whose entries are given by the equation

$$(A^{\mathrm{t}})_{k,j} = A_{j,k}.$$

Example: If $A = \begin{pmatrix} 5 & -7 \\ 3 & 8 \\ -4 & 2 \end{pmatrix}$,

then $A^{\mathrm{t}} = \begin{pmatrix} 5 & 3 & -4 \\ -7 & 8 & 2 \end{pmatrix}$.

The transpose of a product

If A is an m-by-n matrix and C is an n-by-p matrix, then

$$(AC)^{\mathrm{t}} = C^{\mathrm{t}}A^{\mathrm{t}}.$$

The Matrix of the Dual of a Linear Map

The matrix of T' is the transpose of the matrix of T

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

$$\mathcal{M}(T') = \big(\mathcal{M}(T)\big)^{\mathrm{t}}.$$

Row Rank and Column Rank

Definition: *row rank, column rank*

Suppose A is an m-by-n matrix with entries in $\mathbf{F}$.

- The *row rank* of A is the dimension of the span of the rows of A in $\mathbf{F}^{1,n}$.
- The *column rank* of A is the dimension of the span of the columns of A in $\mathbf{F}^{m,1}$.

Example:

Suppose $A = \begin{pmatrix} 4 & 7 & 1 & 8 \\ 3 & 5 & 2 & 9 \end{pmatrix}$. The row rank of A is the dimension of

$$\operatorname{span}\left(\begin{pmatrix} 4 & 7 & 1 & 8 \end{pmatrix}, \begin{pmatrix} 3 & 5 & 2 & 9 \end{pmatrix} \right)$$

in $\mathbf{F}^{1,4}$. Thus the row rank of A is 2.

The column rank of A is the dimension of

$$\operatorname{span}\left(\begin{pmatrix} 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 7 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 8 \\ 9 \end{pmatrix} \right)$$

in $\mathbf{F}^{2,1}$. Thus the column rank of A is 2.

Column Rank Equals Dimension of Range

Dimension of $\operatorname{range} T$ equals column rank of $\mathcal{M}(T)$

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $\dim \operatorname{range} T$ equals the column rank of $\mathcal{M}(T)$.

Row Rank Equals Column Rank

Row rank equals column rank

Suppose $A \in \mathbf{F}^{m,n}$. Then the row rank of A equals the column rank of A.

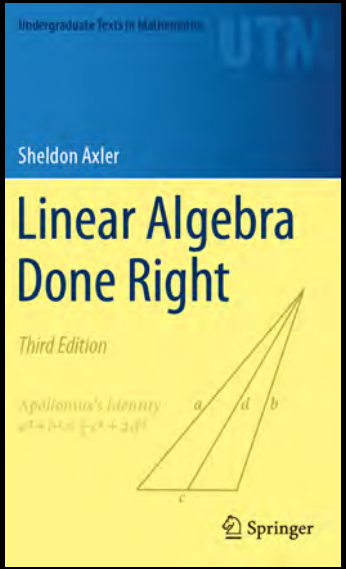
Proof Define $T\colon \mathbf{F}^{n,1} \to \mathbf{F}^{m,1}$ by $Tx = Ax$. Thus $\mathcal{M}(T) = A$ (with respect to standard bases). Now

$$\begin{aligned}
\text{column rank of } A &= \text{column rank of } \mathcal{M}(T) \\
&= \dim \operatorname{range} T \\
&= \dim \operatorname{range} T' \\
&= \text{column rank of } \mathcal{M}(T') \\
&= \text{column rank of } A^{\mathrm{t}} \\
&= \text{row rank of } A.
\end{aligned}$$

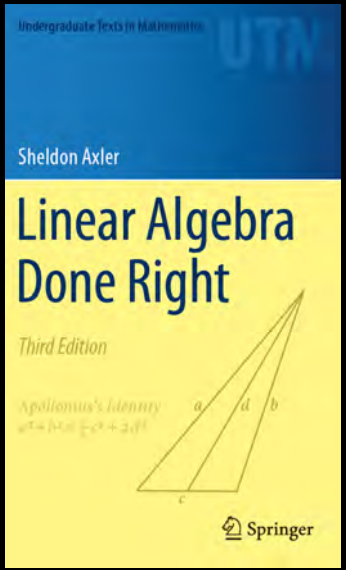
Definition: ***rank***

The *rank* of a matrix $A \in \mathbf{F}^{m,n}$ is the column rank of A.

Linear Algebra Done Right, by Sheldon Axler



Polynomials



Notation

$\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.

Real Part, Imaginary Part

Definition: $\operatorname{Re} z$, $\operatorname{Im} z$

Suppose $z = a + bi$, where a and b are real numbers.

- The *real part* of z, denoted $\operatorname{Re} z$, is defined by $\operatorname{Re} z = a$.
- The *imaginary part* of z, denoted $\operatorname{Im} z$, is defined by $\operatorname{Im} z = b$.

Thus for every complex number z, we have

$$z = \operatorname{Re} z + (\operatorname{Im} z)i.$$

Example: Suppose $z = 3 + 2i$. Then

$$\operatorname{Re} z = 3 \quad \text{and} \quad \operatorname{Im} z = 2.$$

Complex Conjugate, Absolute Value

Definition: *complex conjugate,* $\bar{z}$*, absolute value,* $|z|$

Suppose $z \in \mathbf{C}$.

- The *complex conjugate* of $z \in \mathbf{C}$, denoted $\bar{z}$, is defined by
$$\bar{z} = \operatorname{Re} z - (\operatorname{Im} z)i.$$
- The *absolute value* of a complex number z, denoted $|z|$, is defined by
$$|z| = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}.$$

Example: Suppose $z = 3 + 2i$. Then

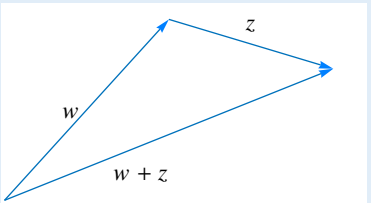
- $\bar{z} = 3 - 2i$;
- $|z| = \sqrt{3^2 + 2^2} = \sqrt{13}$.

Properties of Complex Conjugation and Absolute Value

Properties of complex numbers

Suppose $w, z \in \mathbf{C}$. Then

- $z + \bar{z} = 2 \operatorname{Re} z$;
- $z - \bar{z} = 2(\operatorname{Im} z)i$;
- $z\bar{z} = |z|^2$;
- $\overline{w + z} = \bar{w} + \bar{z}$ and $\overline{wz} = \bar{w}\bar{z}$;
- $\bar{\bar{z}} = z$;
- $|\operatorname{Re} z| \leq |z|$ and $|\operatorname{Im} z| \leq |z|$
- $|\bar{z}| = |z|$;
- $|wz| = |w|\,|z|$;
- Triangle Inequality: $|w + z| \leq |w| + |z|$.



Zero Polynomials

Definition: *polynomial,* $\mathcal{P}(\mathbf{F})$

- A function $p\colon \mathbf{F} \to \mathbf{F}$ is called a *polynomial* with coefficients in $\mathbf{F}$ if there exist $a_0, \ldots, a_m \in \mathbf{F}$ such that

  $$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

  for all $z \in \mathbf{F}$.
- $\mathcal{P}(\mathbf{F})$ is the set of all polynomials with coefficients in $\mathbf{F}$.

If a polynomial is the zero function, then all its coefficients are 0

Suppose $a_0, \ldots, a_m \in \mathbf{F}$. If

$$a_0 + a_1 z + \cdots + a_m z^m = 0$$

for every $z \in \mathbf{F}$, then $a_0 = \cdots = a_m = 0$.

Division Algorithm for Polynomials

Division Algorithm for Polynomials

Suppose that $p, s \in \mathcal{P}(\mathbf{F})$, with $s \neq 0$. Then there exist unique polynomials $q, r \in \mathcal{P}(\mathbf{F})$ such that

$$p = sq + r$$

and $\deg r < \deg s$.

Zeros of Polynomials

Definition: *zero of a polynomial*

A number $\lambda \in \mathbf{F}$ is called a *zero* (or *root*) of a polynomial $p \in \mathcal{P}(\mathbf{F})$ if

$$p(\lambda) = 0.$$

Each zero of a polynomial corresponds to a degree-1 factor

Suppose $p \in \mathcal{P}(\mathbf{F})$ and $\lambda \in \mathbf{F}$. Then $p(\lambda) = 0$ if and only if there is a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbf{F}$.

A polynomial has at most as many zeros as its degree

Suppose $p \in \mathcal{P}(\mathbf{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct zeros in $\mathbf{F}$.

Omar Khayyám (1048 – 1131)

Fundamental Theorem of Algebra

Fundamental Theorem of Algebra

Every nonconstant polynomial with complex coefficients has a zero.

Factorization of a polynomial over $\mathbf{C}$

If $p \in \mathcal{P}(\mathbf{C})$ is a nonconstant polynomial, then p has a unique factorization (except for the order of the factors) of the form

$$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m),$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbf{C}$.

Complex Zeros of Polynomials with Real Coefficients

Polynomials with real coefficients have zeros in pairs

Suppose $p \in \mathcal{P}(\mathbf{C})$ is a polynomial with real coefficients. If $\lambda \in \mathbf{C}$ is a zero of p, then so is $\bar{\lambda}$.

Factorization of Polynomials over $\mathbf{R}$

Factorization of a quadratic polynomial

Suppose $b, c \in \mathbf{R}$. Then there is a polynomial factorization of the form

$$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

with $\lambda_1, \lambda_2 \in \mathbf{R}$ if and only if $b^2 \geq 4c$.

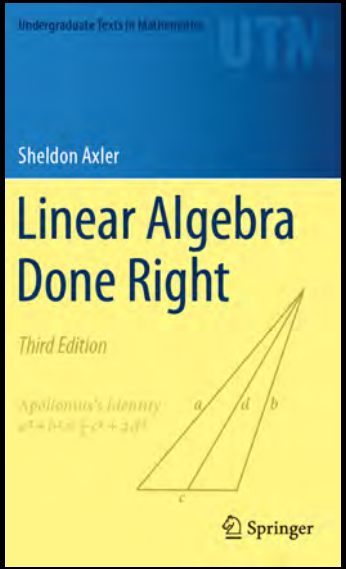
Factorization of a polynomial over $\mathbf{R}$

Suppose $p \in \mathcal{P}(\mathbf{R})$ is a nonconstant polynomial. Then p has a unique factorization (except for the order of the factors) of the form

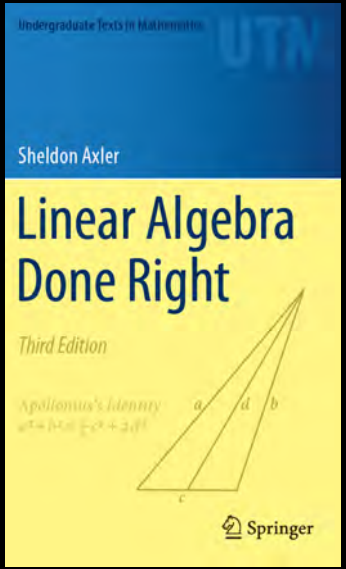
$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbf{R}$, with ${b_j}^2 < 4c_j$ for each j.

Linear Algebra Done Right, by Sheldon Axler



Invariant Subspaces



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a vector space over $\mathbf{F}$.
- operator = linear map from a vector space to itself
- $\mathcal{L}(V) = \mathcal{L}(V, V)$

Invariant Subspace

Suppose $T \in \mathcal{L}(V)$. We will try to investigate T by decomposing V as

$$V = U_1 \oplus \cdots \oplus U_m,$$

and then looking at each $T|_{U_j}$. However, to use results about operators, we need for $T|_{U_j}$ to map U_j into itself.

Definition: *invariant subspace*

Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called *invariant* under T if $u \in U$ implies $Tu \in U$.

Example: Each of these subspaces of V is invariant under $T \in \mathcal{L}(V)$:

- $\{0\}$;
- V;
- $\operatorname{null} T$;
- $\operatorname{range} T$.

Example: Suppose that $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R})\big)$ is defined by $Tp = p'$. Then $\mathcal{P}_4(\mathbf{R})$ is invariant under T because if $p \in \mathcal{P}(\mathbf{R})$ has degree at most 4, then p' also has degree at most 4.

Invariant Subspaces of Dimension 1

Suppose $v \in V$ and $v \neq 0$. Let

$$U = \{\lambda v : \lambda \in \mathbf{F}\} = \operatorname{span}(v).$$

Then U is a one-dimensional subspace of V. U is invariant under T if and only if

$$Tv = \lambda v$$

for some $\lambda \in \mathbf{F}$.

Definition: *eigenvalue*

Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbf{F}$ is called an *eigenvalue* of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$.

Equivalent conditions to be an eigenvalue

Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in F$. Then the following are equivalent:

- λ is an eigenvalue of T;
- $T - \lambda I$ is not injective;
- $T - \lambda I$ is not surjective;
- $T - \lambda I$ is not invertible.

An Operator with No Eigenvalues

Example: Suppose $T \in \mathcal{L}(\mathbf{R}^2)$ is defined by

$$T(x, y) = (-y, x).$$

T is a counterclockwise rotation by 90° about the origin in $\mathbf{R}^2$.

An operator has an eigenvalue if and only if there exists a nonzero vector in its domain that gets sent by the operator to a scalar multiple of itself.

A 90° counterclockwise rotation of a nonzero vector in $\mathbf{R}^2$ obviously never equals a scalar multiple of itself.

Conclusion: T has no eigenvalues.

Complex Eigenvalues

Example: Suppose $T \in \mathcal{L}(\mathbf{C}^2)$ is defined by

$$T(w, z) = (-z, w).$$

Then

$$\begin{aligned} T(1, -i) &= (i, 1) \\ &= i(1, -i). \end{aligned}$$

Thus i is an eigenvalue of T.

Also

$$\begin{aligned} T(1, i) &= (-i, 1) \\ &= -i(1, i). \end{aligned}$$

Thus $-i$ is an eigenvalue of T.

Eigenvectors

Definition: *eigenvector*

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$ is an eigenvalue of T. A vector $v \in V$ is called an *eigenvector* of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

Example:
If $T \in \mathcal{L}(\mathbf{C}^2)$ is defined by $T(w, z) = (-z, w)$, then $(1, -i)$ is an eigenvector corresponding to the eigenvalue i because

$$T(1, -i) = i(1, -i).$$

If $b \in \mathbf{C}$ and $b \neq 0$, then $(b, -bi)$ is also an eigenvector corresponding to the eigenvalue i because

$$T(b, -bi) = i(b, -bi).$$

Linearly Independent Eigenvectors

Linearly independent eigenvectors

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

Proof Suppose $v_1, \ldots, v_m$ is linearly dependent. Let k be the smallest positive integer such that

$$v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

Thus there exist $a_1, \ldots, a_{k-1} \in \mathbf{F}$ such that

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}.$$

Apply T to both sides of this equation, getting

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}.$$

Multiply both sides of the first equation by λ_k and then subtract the equation above, getting

$$0 = a_1(\lambda_k - \lambda_1)v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1})v_{k-1}.$$

However, $v_1, \ldots, v_{k-1}$ is linearly independent. Thus all the a's are 0. However, this means that v_k equals 0, contradicting our hypothesis that v_k is an eigenvector. Therefore our assumption that $v_1, \ldots, v_m$ is linearly dependent was false. ∎

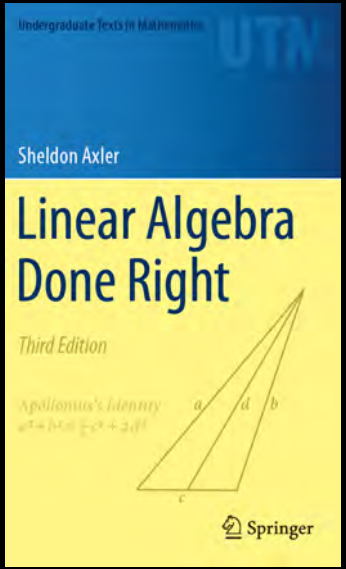
Number of Eigenvalues

Number of eigenvalues

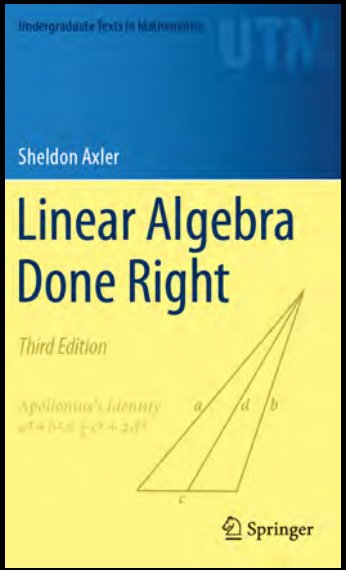
Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Proof Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Let $v_1, \ldots, v_m$ be corresponding eigenvectors. Then the list $v_1, \ldots, v_m$ is linearly independent. Thus $m \leq \dim V$, as desired. ∎

Linear Algebra Done Right, by Sheldon Axler



Existence of Eigenvalues



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a vector space over $\mathbf{F}$.
- operator = linear map from a vector space to itself
- $\mathcal{L}(V) = \mathcal{L}(V, V)$

Powers of an Operator

Definition: T^m

Suppose $T \in \mathcal{L}(V)$ and m is a positive integer.

- T^m is defined by
$$T^m = \underbrace{T \cdots T}_{m \text{ times}}.$$
- T^0 is defined to be the identity operator I on V.
- If T is invertible with inverse T^{-1}, then T^{-m} is defined by
$$T^{-m} = (T^{-1})^m.$$

Example: $T^3 = TTT$

If $T \in \mathcal{L}(V)$, then
$$T^m T^n = T^{m+n} \quad \text{and} \quad (T^m)^n = T^{mn},$$
where m and n are allowed to be arbitrary integers if T is invertible and nonnegative integers if T is not invertible.

Polynomials Applied to an Operator

Definition: $p(T)$

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$ is a polynomial given by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for $z \in \mathbf{F}$. Then $p(T)$ is the operator defined by

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

Example: If $p(z) = z^3$, then $p(T) = T^3$.

Example: Suppose $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R})\big)$ is the differentiation operator defined by $Dq = q'$ and p is the polynomial defined by

$$p(x) = 7 - 3x + 5x^2.$$

Then

$$p(D) = 7I - 3D + 5D^2;$$

thus

$$\big(p(D)\big)q = 7q - 3q' + 5q''$$

for every $q \in \mathcal{P}(\mathbf{R})$.

Algebraic Properties of $p(T)$

$p \mapsto p(T)$ *is linear*

Fix $T \in \mathcal{L}(V)$. The function from $\mathcal{P}(\mathbf{F})$ to $\mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear.

Multiplicativity

Suppose $T \in \mathcal{L}(V)$ and $p, q \in \mathcal{P}(\mathbf{F})$. Then $(pq)(T) = p(T)q(T)$.

Example: Suppose $p(z) = z + 2$, $q(z) = z + 3$. Then

$$(pq)(z) = z^2 + 5z + 6.$$

We have $p(T) = T + 2I$, $q(T) = T + 3I$, and

$$(pq)(T) = T^2 + 5T + 6I.$$

Thus

$$\begin{aligned} p(T)q(T) &= (T + 2I)(T + 3I) \\ &= T^2 + 5T + 6I \\ &= (pq)(T). \end{aligned}$$

Commutativity

Suppose $T \in \mathcal{L}(V)$ and $p, q \in \mathcal{P}(\mathbf{F})$. Then $p(T)q(T) = q(T)p(T)$.

Proof

$$\begin{aligned} p(T)q(T) &= (pq)(T) \\ &= (qp)(T) \\ &= q(T)p(T) \end{aligned}$$

Existence of Eigenvalues

Operators on complex vector spaces have an eigenvalue

Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

This result is false on real vector spaces.
Example: Define $T \in \mathcal{L}(\mathbf{R}^2)$ by

$$T(x, y) = (-y, x).$$

Then T has no eigenvalues.

This result is false on infinite-dimensional complex vector spaces.
Example: Define $T \in \mathcal{L}(\mathcal{P}(\mathbf{C}))$ by

$$(Tp)(z) = zp(z).$$

Then T has no eigenvalues.

Existence of Eigenvalues

Operators on complex vector spaces have an eigenvalue

Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

Proof Suppose V is a complex vector space with dimension $n > 0$ and $T \in \mathcal{L}(V)$. Choose $v \in V$ with $v \neq 0$. Then

$$v, Tv, T^2v, \ldots, T^nv$$

is not linearly independent, because V has dimension n and we have a list of length $n+1$. Thus there exist complex numbers $a_0, \ldots, a_n$, not all 0, such that

$$0 = a_0v + a_1Tv + \cdots + a_nT^nv.$$

Note that $a_1, \ldots, a_n$ cannot all be 0, because otherwise the equation above would become $0 = a_0v$, which would force a_0 also to be 0.

Make the a's the coefficients of a polynomial, which has a factorization

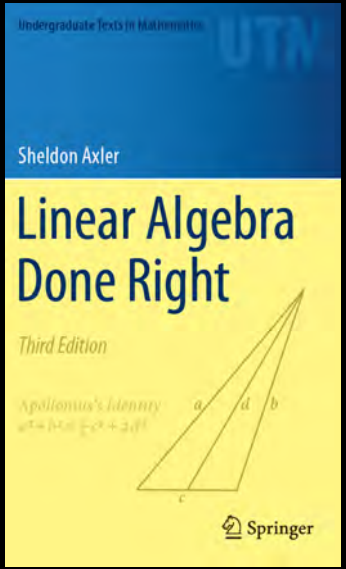
$$a_0 + a_1z + \cdots + a_nz^n = c(z-\lambda_1)\cdots(z-\lambda_m),$$

where c is a nonzero complex number and each λ_j is in $\mathbf{C}$. We then have

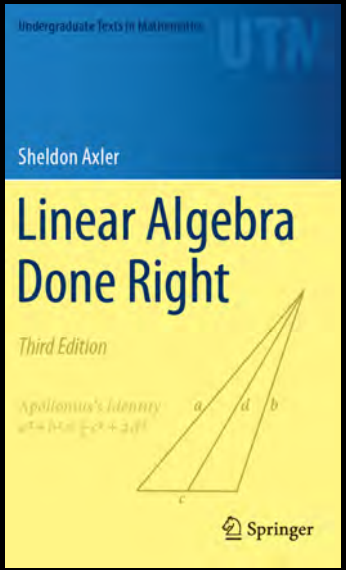
$$\begin{aligned} 0 &= a_0v + a_1Tv + \cdots + a_nT^nv \\ &= (a_0I + a_1T + \cdots + a_nT^n)v \\ &= c(T-\lambda_1I)\cdots(T-\lambda_mI)v. \end{aligned}$$

Thus $T - \lambda_jI$ is not injective for at least one j. Hence T has an eigenvalue. ∎

Linear Algebra Done Right, by Sheldon Axler



Upper-Triangular Matrices



The Matrix of an Operator

Definition: *matrix of an operator,* $\mathcal{M}(T)$

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. The *matrix of* T with respect to this basis is the n-by-n matrix

$$\mathcal{M}(T) = \begin{pmatrix} A_{1,1} & \ldots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \ldots & A_{n,n} \end{pmatrix}$$

whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}v_1 + \cdots + A_{n,k}v_n.$$

If the basis is not clear from the context, then the notation $\mathcal{M}\big(T, (v_1, \ldots, v_n)\big)$ is used.

Suppose $T \in \mathcal{L}(V)$, then

- $\mathcal{M}(T)$ is computed using just one basis;
- $\mathcal{M}(T)$ is a ***square*** matrix.

Example: Define $T \in \mathcal{L}(\mathbf{R}^3)$ by

$$T(x, y, z) = (2x+y, 5y+3z, 8z).$$

Then

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

with respect to the standard basis of $\mathbf{R}^3$.

Upper-Triangular Matrices

Example:

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

Definition: *diagonal of matrix*

The *diagonal* of a square matrix consists of the entries along the line from the upper left corner to the bottom right corner.

Definition: *upper-triangular matrix*

A matrix is called *upper triangular* if all the entries below the diagonal equal 0.

Conditions for upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then the following are equivalent:

- the matrix of T with respect to $v_1, \ldots, v_n$ is upper triangular;
- $Tv_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each $j = 1, \ldots, n$;
- $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each $j = 1, \ldots, n$.

Over $\mathbf{C}$, Every Operator has Upper-Triangular Matrix

Over $\mathbf{C}$, every operator has an upper-triangular matrix

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & * \\ & \lambda_2 & \\ & & \ddots \\ 0 & & \lambda_n \end{pmatrix}$$

The first basis vector v_1 must be an eigenvector of T with eigenvalue λ_1.

Eigenvalues Along the Diagonal

Determination of eigenvalues from upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & * \\ & \lambda_2 & \\ & & \ddots \\ 0 & & \lambda_n \end{pmatrix}$$

Example: Define $T \in \mathcal{L}(\mathbf{R}^3)$ by

$$T(x, y, z) = (2x+y, 5y+3z, 8z).$$

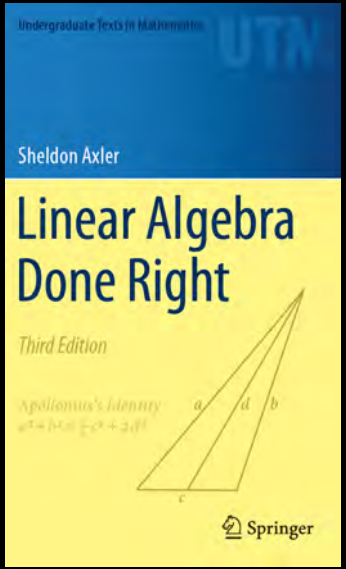
Then

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

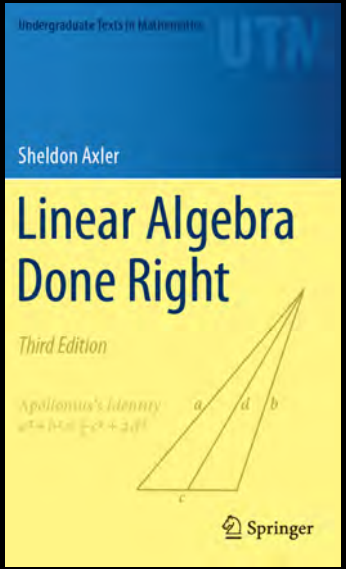
with respect to the standard basis of $\mathbf{R}^3$.

Thus the eigenvalues of T are 2, 5, and 8.

Linear Algebra Done Right, by Sheldon Axler



Eigenspaces and Diagonal Matrices



Diagonal Matrices

Definition: *diagonal matrix*

A *diagonal matrix* is a square matrix that is 0 everywhere except possibly along the diagonal.

Example: $\begin{pmatrix} 8 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$ is a diagonal matrix.

Every diagonal matrix is upper triangular.

If an operator has a diagonal matrix with respect to some basis, then the entries along the diagonal are precisely the eigenvalues of the operator.

Eigenspaces

Definition: *eigenspace,* $E(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The *eigenspace* of T corresponding to λ, denoted $E(\lambda, T)$, is defined by

$$E(\lambda, T) = \operatorname{null}(T - \lambda I).$$

$$v \in \operatorname{null}(T - \lambda I) \iff Tv = \lambda v.$$

Thus $E(\lambda, T)$ is the set of eigenvectors of T corresponding to λ, along with the 0 vector.

For $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, the eigenspace $E(\lambda, T)$ is a subspace of V.

λ is an eigenvalue of T if and only if $E(\lambda, T) \neq \{0\}$.

Example: Suppose the matrix of an operator $T \in \mathcal{L}(V)$ with respect to a basis v_1, v_2, v_3 of V is the matrix

$$\begin{pmatrix} 8 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

Then

$$E(8, T) = \operatorname{span}(v_1), \quad E(5, T) = \operatorname{span}(v_2, v_3).$$

If λ is an eigenvalue of an operator $T \in \mathcal{L}(V)$, then T restricted to $E(\lambda, T)$ is the operator of multiplication by λ.

Sum of Eigenspaces

Sum of eigenspaces is a direct sum

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum. Furthermore,

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

Proof To show that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum, suppose

$$u_1 + \cdots + u_m = 0,$$

where each u_j is in $E(\lambda_j, T)$.

Because eigenvectors corresponding to distinct eigenvalues are linearly independent, this implies that each u_j equals 0. This implies that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum, as desired.

Diagonalizable Operators

Definition: *diagonalizable*

An operator $T \in \mathcal{L}(V)$ is called *diagonalizable* if the operator has a diagonal matrix with respect to some basis of V.

Example: Define $T \in \mathcal{L}(\mathbf{R}^2)$ by

$$T(x, y) = (41x + 7y, -20x + 74y).$$

The matrix of T with respect to the standard basis of $\mathbf{R}^2$ is

$$\begin{pmatrix} 41 & 7 \\ -20 & 74 \end{pmatrix},$$

which is not a diagonal matrix.

However, T is diagonalizable, because the matrix of T with respect to the basis $(1, 4), (7, 5)$ is

$$\begin{pmatrix} 69 & 0 \\ 0 & 46 \end{pmatrix},$$

as you should verify.

Conditions equivalent to diagonalizability

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:

- T is diagonalizable;
- V has a basis consisting of eigenvectors of T;
- there exist 1-dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that
$$V = U_1 \oplus \cdots \oplus U_n;$$
- $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$;
- $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Not Every Operator is Diagonalizable

Define $T \in \mathcal{L}(\mathbf{C}^2)$ defined by

$$T(w, z) = (z, 0).$$

Then 0 is the only eigenvalue of T and furthermore

$$E(0, T) = \{(w, 0) \in \mathbf{C}^2 : w \in \mathbf{C}\}.$$

Thus T is not diagonalizable.

Enough Eigenvalues Implies Diagonalizability

Enough eigenvalues implies diagonalizability

If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable.

Proof Suppose $T \in \mathcal{L}(V)$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, where $n = \dim V$.
For each j, let $v_j \in V$ be an eigenvector corresponding to the eigenvalue λ_j.
Because eigenvectors corresponding to distinct eigenvalues are linearly independent, $v_1, \ldots, v_n$ is linearly independent.

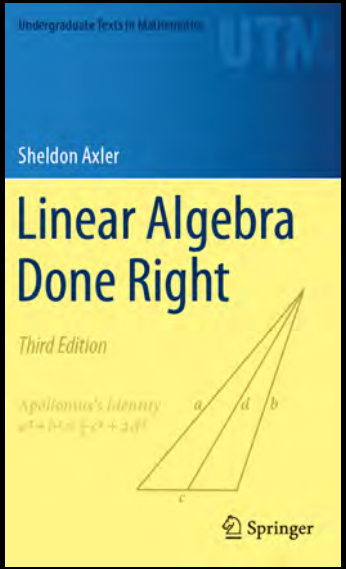
A linearly independent list of n vectors in V is a basis of V; thus

$$v_1, \ldots, v_n$$

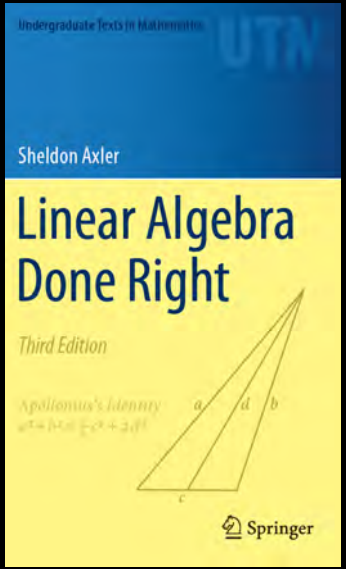
is a basis of V.
With respect to this basis consisting of eigenvectors, T has a diagonal matrix. ∎

Fibonacci [Leonardo of Pisa] (1170 – 1250)

Linear Algebra Done Right, by Sheldon Axler



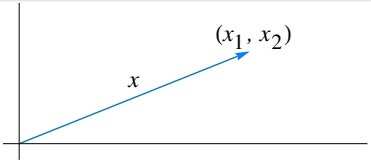
Inner Products and Norms, part 1: Inner Products



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a vector space over $\mathbf{F}$.

Dot Product on $\mathbf{R}^n$



This vector x has length $\sqrt{{x_1}^2 + {x_2}^2}$.

For $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$, define the norm of x by $\|x\| = \sqrt{{x_1}^2 + \cdots + {x_n}^2}$.

Definition: *dot product*

For $x, y \in \mathbf{R}^n$, the *dot product* of x and y, denoted $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n,$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

$x \cdot x = \|x\|^2$ for all $x \in \mathbf{R}^n$.

The dot product on $\mathbf{R}^n$ has the following properties:

- $x \cdot x \geq 0$ for all $x \in \mathbf{R}^n$;
- $x \cdot x = 0$ if and only if $x = 0$;
- for $y \in \mathbf{R}^n$ fixed, the map from $\mathbf{R}^n$ to $\mathbf{R}$ that sends $x \in \mathbf{R}^n$ to $x \cdot y$ is linear;
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbf{R}^n$.

Complex Complication

Recall that if $\lambda = a + bi$, where $a, b \in \mathbf{R}$, then

- the absolute value of λ, denoted $|\lambda|$, is defined by $|\lambda| = \sqrt{a^2 + b^2}$;
- the complex conjugate of λ, denoted $\bar{\lambda}$, is defined by $\bar{\lambda} = a - bi$;
- $|\lambda|^2 = \lambda\bar{\lambda}$.

For $z = (z_1, \ldots, z_n) \in \mathbf{C}^n$, define the norm of z by

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

This suggests that the inner product of $w = (w_1, \ldots, w_n) \in \mathbf{C}^n$ with z should equal

$$w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

Inner Product

Definition: *inner product*

An *inner product* on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v \rangle \in \mathbf{F}$ and has the following properties:

positivity
: $\langle v, v \rangle \geq 0$ for all $v \in V$;

definiteness
: $\langle v, v \rangle = 0$ if and only if $v = 0$;

additivity in first slot
: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$;

homogeneity in first slot
: $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$ for all $\lambda \in \mathbf{F}$ and all $u, v \in V$;

conjugate symmetry
: $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in V$.

Examples of Inner Products

- The *Euclidean inner product* on $\mathbf{F}^n$ is defined by
$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = w_1 \overline{z_1} + \cdots + w_n \overline{z_n}.$$
- If $c_1, \ldots, c_n$ are positive numbers, then an inner product can be defined on $\mathbf{F}^n$ by
$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = c_1 w_1 \overline{z_1} + \cdots + c_n w_n \overline{z_n}.$$
- An inner product can be defined on the vector space of continuous real-valued functions on the interval $[-1, 1]$ by
$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x)\, dx.$$
- An inner product can be defined on $\mathcal{P}(\mathbf{R})$ by
$$\langle p, q \rangle = \int_{0}^{\infty} p(x) q(x) e^{-x}\, dx.$$

Inner Product Spaces

Definition: *inner product space*

An *inner product space* is a vector space V along with an inner product on V.

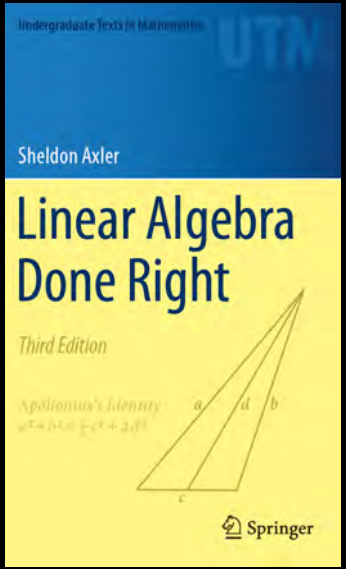
Notation: V

Until further notice in these videos, V denotes an inner product space over $\mathbf{F}$.

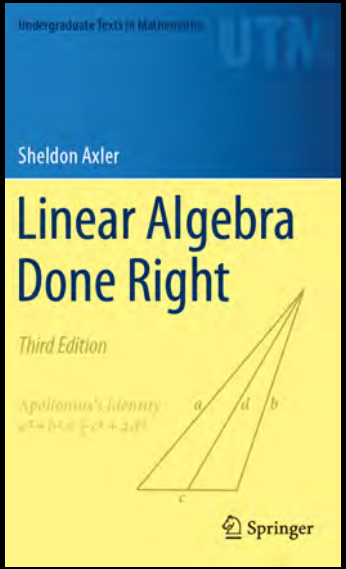
Properties of an inner product

- For each fixed $u \in V$, the function that takes v to $\langle v, u \rangle$ is a linear map from V to $\mathbf{F}$.
- $\langle 0, u \rangle = 0$ for every $u \in V$.
- $\langle u, 0 \rangle = 0$ for every $u \in V$.
- $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all $u, v, w \in V$.
- $\langle u, \lambda v \rangle = \bar{\lambda} \langle u, v \rangle$ for all $\lambda \in \mathbf{F}$ and $u, v \in V$.

Linear Algebra Done Right, by Sheldon Axler



Inner Products and Norms, part 2: Norms



Norms

Definition: *norm*, $\|v\|$

For $v \in V$, the *norm* of v, denoted $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

- If $(z_1, \ldots, z_n) \in \mathbf{F}^n$, then

$$\|(z_1, \ldots, z_n)\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

- In the vector space of continuous real-valued functions on $[-1, 1]$ with inner product $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dx$, we have

$$\|f\| = \sqrt{\int_{-1}^{1} \big(f(x)\big)^2\,dx}.$$

Basic properties of the norm

Suppose $v \in V$.

- $\|v\| = 0$ if and only if $v = 0$.
- $\|\lambda v\| = |\lambda|\,\|v\|$ for all $\lambda \in \mathbf{F}$.

Proof:Suppose $\lambda \in \mathbf{F}$. Then

$$\begin{aligned} \|\lambda v\|^2 &= \langle \lambda v, \lambda v \rangle \\ &= \lambda \langle v, \lambda v \rangle \\ &= \lambda \bar{\lambda} \langle v, v \rangle \\ &= |\lambda|^2 \|v\|^2. \end{aligned}$$

Taking square roots now gives the desired equality. ∎

Orthogonality

Definition: *orthogonal*

Two vectors $u, v \in V$ are called *orthogonal* if $\langle u, v \rangle = 0$.

Orthogonality and 0

- 0 is orthogonal to every vector in V.
- 0 is the only vector in V that is orthogonal to itself.

Pythagorean Theorem

Suppose u and v are orthogonal vectors in V. Then

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

Proof

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= \|u\|^2 + \|v\|^2 \end{aligned}$$

Woman Teaching Geometry

Cauchy–Schwarz Inequality

Cauchy–Schwarz Inequality

Suppose $u, v \in V$. Then

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

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

Examples:

- If $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbf{R}$, then
$$|x_1 y_1 + \cdots + x_n y_n|^2 \leq (x_1{}^2 + \cdots + x_n{}^2)(y_1{}^2 + \cdots + y_n{}^2).$$
- If f, g are continuous real-valued functions on $[-1, 1]$, then
$$\left| \int_{-1}^{1} f(x) g(x) \, dx \right|^2 \leq \left(\int_{-1}^{1} \big(f(x)\big)^2 \, dx \right) \left(\int_{-1}^{1} \big(g(x)\big)^2 \, dx \right).$$

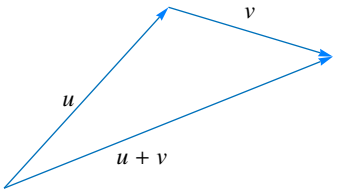
Triangle Inequality

Triangle Inequality

Suppose $u, v \in V$. Then

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

This inequality is an equality if and only if one of u, v is a nonnegative multiple of the other.



Proof

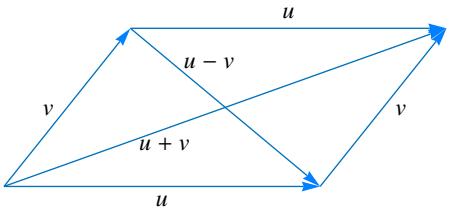
$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v \rangle \\
&= \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \langle v, u \rangle \\
&= \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \overline{\langle u, v \rangle} \\
&= \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}\langle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

Parallelogram Equality

Parallelogram Equality

Suppose $u, v \in V$. Then

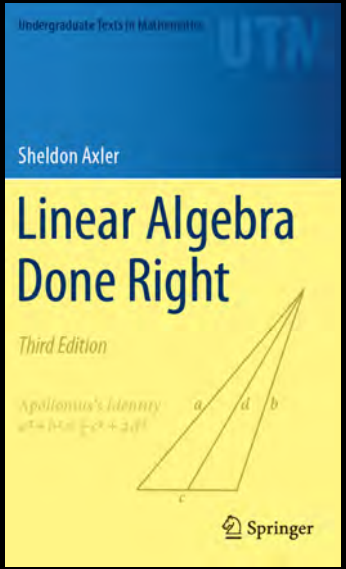
$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2).$$



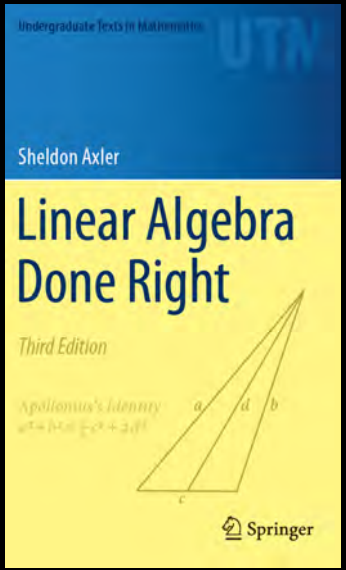
Proof

$$\begin{aligned}
&\|u+v\|^2 + \|u-v\|^2 \\
&\quad = \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\
&\quad = \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \\
&\qquad + \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle \\
&\quad = 2(\|u\|^2 + \|v\|^2)
\end{aligned}$$

Linear Algebra Done Right, by Sheldon Axler



Orthonormal Bases



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes an inner product space over $\mathbf{F}$.

Orthonormal Lists

Definition: *orthonormal*

- A list of vectors in V is called *orthonormal* if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.
- In other words, a list $e_1, \ldots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k \rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Examples:

- The standard basis in $\mathbf{F}^n$ is an orthonormal list.
- $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$ is an orthonormal list in $\mathbf{F}^3$.
- $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right)$ is an orthonormal list in $\mathbf{F}^3$.

Orthonormal Linear Combinations

The norm of an orthonormal linear combination

If $e_1, \ldots, e_m$ 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 $a_1, \ldots, a_m \in \mathbf{F}$.

Proof Pythagorean Theorem

Orthonormal list is linearly independent

Every orthonormal list of vectors is linearly independent.

Proof Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in V and $a_1, \ldots, a_m \in \mathbf{F}$ are such that

$$a_1 e_1 + \cdots + a_m e_m = 0.$$

Then $|a_1|^2 + \cdots + |a_m|^2 = 0$, which means that all the a_j's are 0. Thus $e_1, \ldots, e_m$ is linearly independent. ∎

Orthonormal Bases

Definition: *orthonormal basis*

An *orthonormal basis* of V is an orthonormal list of vectors in V that is also a basis of V.

An orthonormal list of the right length is an orthonormal basis

Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.

Example:

$\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right), \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right), \left(-\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)$

is an orthonormal basis of $\mathbf{F}^4$.

Writing a vector as linear combination of orthonormal basis

Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $v \in V$. Then

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

and

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

Proof There exist scalars $a_1, \ldots, a_n$ such that $v = a_1 e_1 + \cdots + a_n e_n$. Take inner product with e_j.

Gram–Schmidt Procedure

Gram–Schmidt Procedure

Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2, \ldots, m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}.$$

Then $e_1, \ldots, e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(e_1, \ldots, e_j)$$

for $j = 1, \ldots, m$.

Example:
Suppose $V = \mathcal{P}_2(\mathbf{R})$, where the inner product is given by

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

Applying the Gram–Schmidt Procedure to the basis $1, x, x^2$ produces the orthonormal basis

$$\sqrt{\tfrac{1}{2}}, \sqrt{\tfrac{3}{2}}x, \sqrt{\tfrac{45}{8}}\left(x^2 - \tfrac{1}{3}\right).$$

Orthonormal Bases

Existence of orthonormal basis

Every finite-dimensional inner product space has an orthonormal basis.

Proof Suppose V is finite-dimensional. Choose a basis of V. Apply the Gram–Schmidt Procedure to it, producing an orthonormal list with length $\dim V$. This orthonormal list is an orthonormal basis of V. ∎

Orthonormal list extends to orthonormal basis

Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Upper-Triangular Matrices

Schur's Theorem

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof T has an upper-triangular matrix with respect to some basis $v_1, \ldots, v_n$ of V:

$$\mathcal{M}(T) = \begin{pmatrix} * & & * \\ & \ddots & \\ 0 & & * \end{pmatrix}.$$

Thus $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each $j = 1, \ldots, n$.

Apply the Gram–Schmidt Procedure to $v_1, \ldots, v_n$, producing an orthonormal basis $e_1, \ldots, e_n$ of V.
Because

$$\operatorname{span}(e_1, \ldots, e_j) = \operatorname{span}(v_1, \ldots, v_j)$$

for each j, we conclude that $\operatorname{span}(e_1, \ldots, e_j)$ is invariant under T for each $j = 1, \ldots, n$. Thus T has an upper-triangular matrix with respect to the orthonormal basis $e_1, \ldots, e_n$. ∎

Linear Functionals

Definition: *linear functional*

A *linear functional* on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

Example: Suppose $u \in V$. Define $\varphi : V \to \mathbf{F}$ by

$$\varphi(w) = \langle w, u \rangle.$$

Then φ is a linear functional on V.

Riesz Representation Theorem

Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that

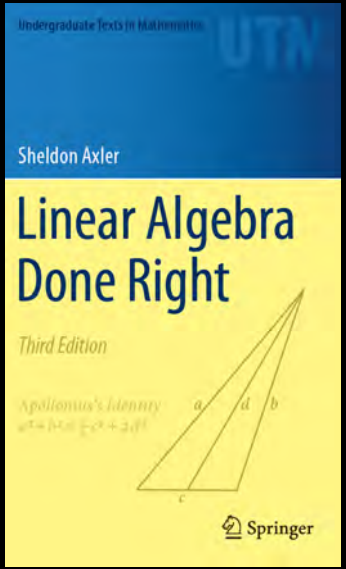
$$\varphi(w) = \langle w, u \rangle$$

for every $w \in V$.

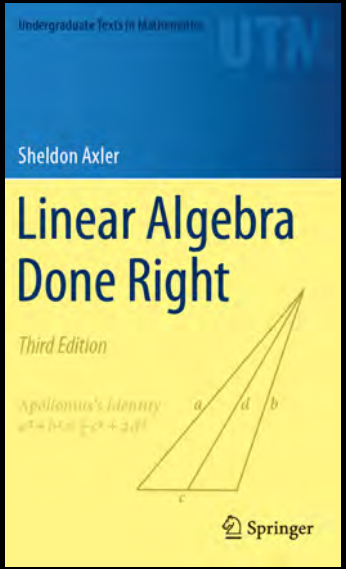
Proof Let $e_1, \ldots, e_n$ be an orthonormal basis of V. Then

$$u = \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n.$$

Linear Algebra Done Right, by Sheldon Axler



Orthogonal Complements



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes an inner product space over $\mathbf{F}$.

Orthogonal Complements

Definition: *orthogonal complement,* $U^\perp$

If U is a subset of V, then the *orthogonal complement* of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle v, u \rangle = 0 \text{ for every } u \in U\}.$$

- If U is a line in $\mathbf{R}^3$ containing the origin, then $U^\perp$ is the plane containing the origin that is perpendicular to U.
- If U is a plane in $\mathbf{R}^3$ containing the origin, then $U^\perp$ is the line containing the origin that is perpendicular to U.

Basic properties of orthogonal complement

- If U is a subset of V, then $U^\perp$ is a subspace of V.
- $\{0\}^\perp = V$.
- $V^\perp = \{0\}$.
- If U is a subset of V, then $U \cap U^\perp \subset \{0\}$.
- If U and W are subsets of V and $U \subset W$, then $W^\perp \subset U^\perp$.

Direct Sum

Direct sum of a subspace and its orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^{\perp}.$$

Proof Suppose $v \in V$. Let $e_1, \ldots, e_m$ be an orthonormal basis of U. Let

$$u = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Then $\langle v - u, e_j \rangle = \langle v, e_j \rangle - \langle u, e_j \rangle = 0$ for $j = 1, \ldots, m$. Thus $v - u \in U^{\perp}$. Now

$$v = u + (v - u),$$

showing that $v \in U + U^{\perp}$. Thus $V = U + U^{\perp}$.

Dimension of the orthogonal complement

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^{\perp} = \dim V - \dim U.$$

The orthogonal complement of the orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$U = (U^{\perp})^{\perp}.$$

Orthogonal Projection

Definition: *orthogonal projection,* P_U

Suppose U is a finite-dimensional subspace of V. The *orthogonal projection* of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows:
For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then $P_U v = u$.

If $e_1, \ldots, e_m$ is an orthonormal basis of U, then

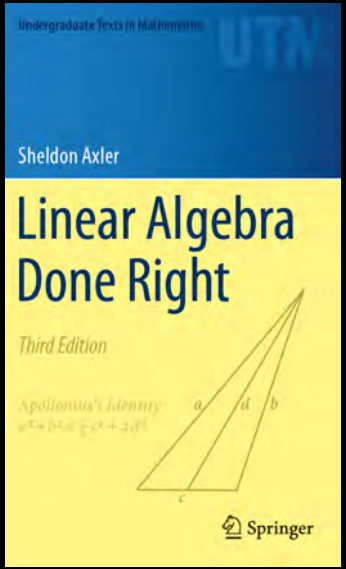
$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Properties of the orthogonal projection P_U

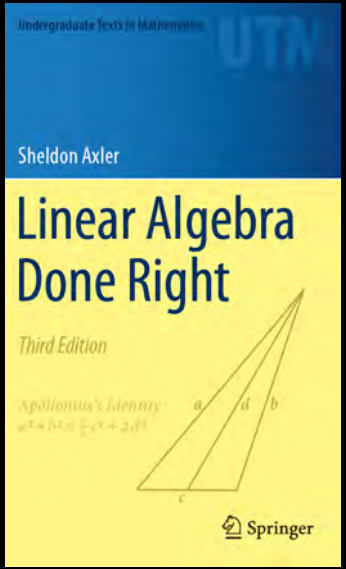
Suppose U is a finite-dimensional subspace of V. Then

- $P_U \in \mathcal{L}(V)$;
- $P_U u = u$ for every $u \in U$;
- $P_U w = 0$ for every $w \in U^\perp$;
- $\operatorname{range} P_U = U$;
- $\operatorname{null} P_U = U^\perp$;
- $v - P_U v \in U^\perp$ for every $v \in V$;
- ${P_U}^2 = P_U$;
- $\|P_U v\| \leq \|v\|$ for every $v \in V$.

Linear Algebra Done Right, by Sheldon Axler



Minimization Problems



Minimization Problems

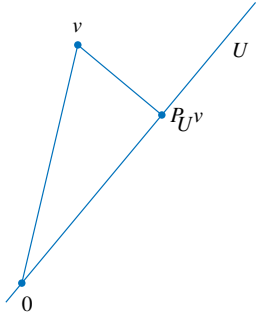
Given a subspace U of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible.

Minimizing the distance to a subspace

Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

$$\|v - P_U v\| \leq \|v - u\|.$$

for all $u \in U$.



$P_U v$ is the closest point in U to v.

Proof Suppose $u \in U$. Then

$$\begin{aligned}\|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2.\end{aligned}$$

Example Minimization Problem

Find $u \in \mathcal{P}_5(\mathbf{R})$ that approximates $\sin x$ as well as possible on the interval $[-\pi, \pi]$, in the sense that

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$$

is as small as possible.

Here $V = C_{\mathbf{R}}[-\pi, \pi]$ with inner product $\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x) \, dx$. Also, $v \in C_{\mathbf{R}}[-\pi, \pi]$ is given by $v(x) = \sin x$, and $U = \mathcal{P}_5(\mathbf{R})$. Thus

$$\begin{aligned} u &= P_U v \\ &= \sum_{j=1}^{6} \langle v, e_j \rangle e_j, \end{aligned}$$

where $e_1, \ldots, e_6$ is an orthonormal basis of U.

Find $e_1, \ldots, e_6$ by applying the Gram–Schmidt procedure to the basis $1, x, x^2, x^3, x^4, x^5$ of U.

Then use the last equation to get

$$\begin{aligned} u = {} & \frac{105(1485 - 153\pi^2 + \pi^4)}{8\pi^6} x \\ & - \frac{315(1155 - 125\pi^2 + \pi^4)}{4\pi^8} x^3 \\ & + \frac{693(945 - 105\pi^2 + \pi^4)}{8\pi^{10}} x^5 \\ \approx {} & 0.987862x - 0.155271x^3 + 0.00564312x^5. \end{aligned}$$

Example Minimization Problem

Find $u \in \mathcal{P}_5(\mathbf{R})$ that approximates $\sin x$ as well as possible on the interval $[-\pi, \pi]$, in the sense that

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$$

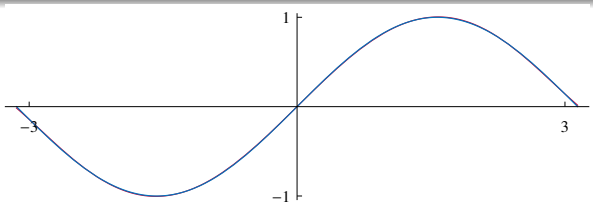
is as small as possible.

Solution:

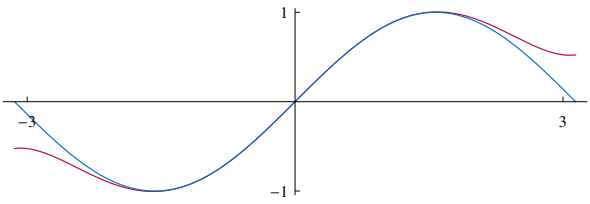
$$u(x) \approx 0.987862x - 0.155271x^3 + 0.00564312x^5$$

Compare to Taylor polynomial:

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} \approx x - 0.166667x^3 + 0.00833333x^5$$

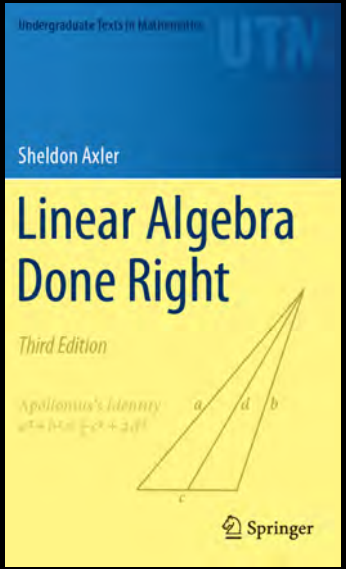


Graphs on $[-\pi, \pi]$ *of* $\sin x$ (*blue*) *and its approximation* $u(x)$ (*red*).

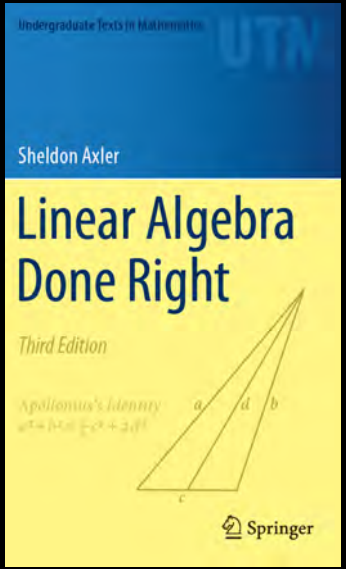


Graphs on $[-\pi, \pi]$ *of* $\sin x$ (*blue*) *and the Taylor polynomial* (*red*).

Linear Algebra Done Right, by Sheldon Axler



Self-Adjoint and Normal Operators, Part 1: Adjoints



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V and W denote finite-dimensional inner product spaces over $\mathbf{F}$.

Adjoints

Definition: *adjoint,* T^*

Suppose $T \in \mathcal{L}(V, W)$. The *adjoint* of T is the function $T^*\colon W \to V$ such that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for every $v \in V$ and every $w \in W$.

Example: Define $T\colon \mathbf{R}^3 \to \mathbf{R}^2$ by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

To compute T^*, fix a point $(y_1, y_2) \in \mathbf{R}^2$.

Then for every $(x_1, x_2, x_3) \in \mathbf{R}^3$ we have

$$\begin{aligned}\langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle &= \langle T(x_1, x_2, x_3), (y_1, y_2)\rangle \\ &= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle \\ &= x_2y_1 + 3x_3y_1 + 2x_1y_2 \\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle.\end{aligned}$$

Thus

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1).$$

Properties of the Adjoint

Definition: *adjoint,* T^*

Suppose $T \in \mathcal{L}(V, W)$. The *adjoint* of T is the function $T^* \colon W \to V$ such that

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for every $v \in V$ and every $w \in W$.

The adjoint is a linear map

If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$.

Properties of the adjoint

- $(S + T)^* = S^* + T^*$ for all $S, T \in \mathcal{L}(V, W)$;
- $(\lambda T)^* = \bar{\lambda} T^*$ for all $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$;
- $(T^*)^* = T$ for all $T \in \mathcal{L}(V, W)$;
- $I^* = I$, where I is identity operator on V;
- $(ST)^* = T^* S^*$ for all $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$ (here U is an inner product space over $\mathbf{F}$).

The Adjoint of a Composition

- $(ST)^* = T^*S^*$ for all $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$.

Proof Suppose $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$. If $v \in V$ and $u \in U$, then

$$\begin{aligned}\langle v, (ST)^* u\rangle &= \langle STv, u\rangle \\ &= \langle Tv, S^* u\rangle \\ &= \langle v, T^*(S^* u)\rangle.\end{aligned}$$

Thus $(ST)^* u = T^*(S^* u)$, as desired. ∎

Null Space and Range of T^*

Null space and range of T^*

Suppose $T \in \mathcal{L}(V, W)$. Then

(a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;

(b) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$;

(c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$;

(d) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

Condition for Surjectivity

Suppose $T \in \mathcal{L}(V, W)$. Then T is surjective if and only if T^* is injective.

Proof We begin by proving (a). Let $w \in W$. Then

$$\begin{aligned} w \in \operatorname{null} T^* &\iff T^* w = 0 \\ &\iff \langle v, T^* w \rangle = 0 \text{ for all } v \in V \\ &\iff \langle Tv, w \rangle = 0 \text{ for all } v \in V \\ &\iff w \in (\operatorname{range} T)^\perp. \end{aligned}$$

Thus $\operatorname{null} T^* = (\operatorname{range} T)^\perp$, proving (a).

If we take the orthogonal complement of both sides of (a), we get (d).

Replacing T with T^* in (a) gives (c).

Finally, replacing T with T^* in (d) gives (b). ∎

Conjugate Transpose

Definition: *conjugate transpose*

The *conjugate transpose* of an m-by-n matrix is the n-by-m matrix obtained by interchanging the rows and columns and then taking the complex conjugate of each entry.

Example: The conjugate transpose of the matrix $\begin{pmatrix} 2 & 3+4i & 7 \\ 6 & 5 & 8i \end{pmatrix}$ is the matrix

$$\begin{pmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{pmatrix}.$$

The matrix of T^*

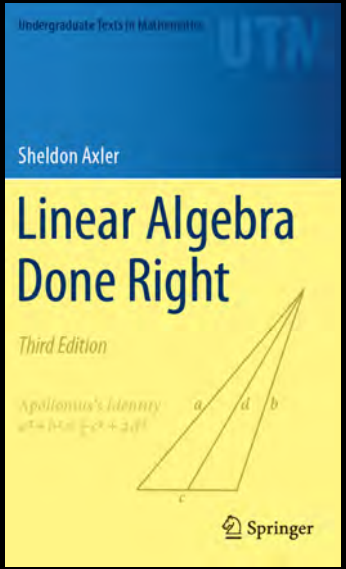
Let $T \in \mathcal{L}(V,W)$. Suppose $e_1,\dots,e_n$ is an orthonormal basis of V and $f_1,\dots,f_m$ is an orthonormal basis of W. Then

$$\mathcal{M}\big(T^*,(f_1,\dots,f_m),(e_1,\dots,e_n)\big)$$

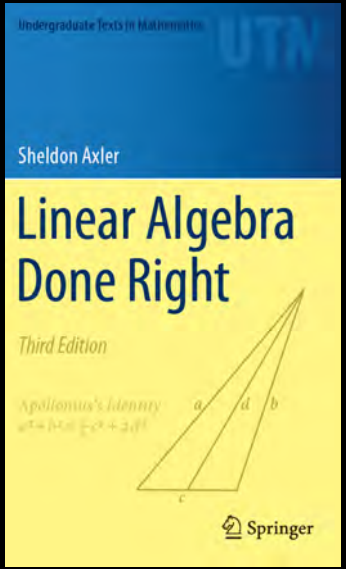
is the conjugate transpose of

$$\mathcal{M}\big(T,(e_1,\dots,e_n),(f_1,\dots,f_m)\big).$$

Linear Algebra Done Right, by Sheldon Axler



Self-Adjoint and Normal Operators, Part 2: Self-Adjoint Operators



Definition and Example of Self-Adjoint Operator

Definition: *self-adjoint*

An operator $T \in \mathcal{L}(V)$ is called *self-adjoint* if $T = T^*$. In other words, $T \in \mathcal{L}(V)$ is self-adjoint if and only if

$$\langle Tv, w \rangle = \langle v, Tw \rangle$$

for all $v, w \in V$.

Example: Suppose T is the operator on $\mathbf{F}^2$ whose matrix is

$$\begin{pmatrix} 2 & b \\ 3 & 7 \end{pmatrix}.$$

Recall that $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$.

Thus the operator T is self-adjoint if and only if $b = 3$.

Eigenvalues of Self-Adjoint Operators are Real

Eigenvalues of self-adjoint operators are real

Every eigenvalue of a self-adjoint operator is real.

Proof Suppose T is a self-adjoint operator on V.
Let λ be an eigenvalue of T, and let v be a nonzero vector in V such that $Tv = \lambda v$. Then

$$\begin{aligned}
\lambda\|v\|^2 &= \lambda\langle v, v\rangle \\
&= \langle \lambda v, v\rangle \\
&= \langle Tv, v\rangle \\
&= \langle v, Tv\rangle \\
&= \langle v, \lambda v\rangle \\
&= \bar{\lambda}\|v\|^2.
\end{aligned}$$

Thus $\lambda = \bar{\lambda}$, which means that λ is real, as desired. ∎

Over $\mathbf{C}$, $\langle Tv, v\rangle = 0$ for all v only for the 0 operator

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. If

$$\langle Tv, v\rangle = 0$$

for all $v \in V$, then $T = 0$.

Over $\mathbf{C}$, characterization of self-adjoint operators

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then T is self-adjoint if and only if

$$\langle Tv, v\rangle \in \mathbf{R}$$

for every $v \in V$.

Proof Let $v \in V$. Then

$$\begin{aligned}\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} &= \langle Tv, v\rangle - \langle v, Tv\rangle \\ &= \langle Tv, v\rangle - \langle T^*v, v\rangle \\ &= \langle (T - T^*)v, v\rangle.\end{aligned}$$

If $\langle Tv, v\rangle \in \mathbf{R}$ for every $v \in V$, then the left side of the first equation above equals 0, so $\langle (T - T^*)v, v\rangle = 0$ for every $v \in V$. This implies that $T - T^* = 0$. Hence T is self-adjoint.

Conversely, if T is self-adjoint, then the right side of the last equation above equals 0, so $\langle Tv, v\rangle = \overline{\langle Tv, v\rangle}$ for every $v \in V$. This implies that $\langle Tv, v\rangle \in \mathbf{R}$ for every $v \in V$, as desired. ∎

Over $\mathbf{C}$, $\langle Tv, v \rangle = 0$ for all v only for the 0 operator

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. If

$$\langle Tv, v \rangle = 0$$

for all $v \in V$, then $T = 0$.

Over $\mathbf{C}$, characterization of self-adjoint operators

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then T is self-adjoint if and only if

$$\langle Tv, v \rangle \in \mathbf{R}$$

for every $v \in V$.

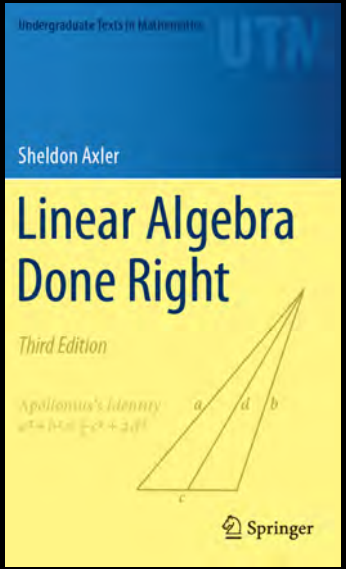
If $T = T^*$ and $\langle Tv, v \rangle = 0$ for all v, then $T = 0$

Suppose T is a self-adjoint operator on V such that

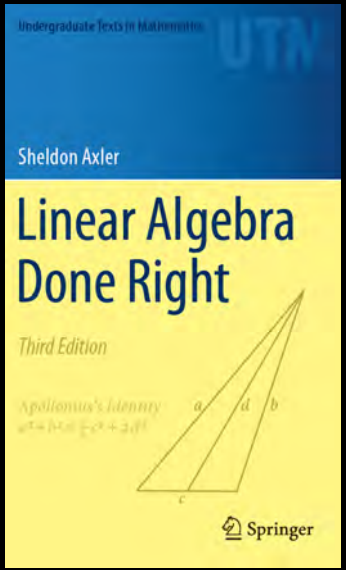
$$\langle Tv, v \rangle = 0$$

for all $v \in V$. Then $T = 0$.

Linear Algebra Done Right, by Sheldon Axler



Self-Adjoint and Normal Operators, Part 3: Normal Operators



Definition and Example of Normal Operator

Definition: *normal*

- An operator on an inner product space is called *normal* if it commutes with its adjoint.
- In other words, $T \in \mathcal{L}(V)$ is normal if

$$TT^* = T^*T.$$

Obviously every self-adjoint operator is normal, because if T is self-adjoint then $T^* = T$.

Example: Let T be the operator on $\mathbf{F}^2$ whose matrix is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

This operator is not self-adjoint.
The matrix of TT^* equals

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Similarly, the matrix of T^*T equals

$$\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Because TT^* and T^*T have the same matrix, we see that $TT^* = T^*T$. Thus T is normal.

Normality and Norms

T *is normal if and only if* $\|Tv\| = \|T^*v\|$ *for all* v

An operator $T \in \mathcal{L}(V)$ is normal if and only if

$$\|Tv\| = \|T^*v\|$$

for all $v \in V$.

The result above implies that $\operatorname{null} T = \operatorname{null} T^*$ for every normal operator T.

Proof Let $T \in \mathcal{L}(V)$. Note that

$$\begin{aligned} T \text{ is normal} &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0 \\ &\qquad \text{for all } v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \\ &\qquad \text{for all } v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \\ &\qquad \text{for all } v \in V. \end{aligned}$$

The equivalence of first and last conditions above gives the desired result. ∎

A Normal Operator and Its Adjoint Have the Same Eigenvectors

For T normal, T and T^* have the same eigenvectors

Suppose $T \in \mathcal{L}(V)$ is normal and $v \in V$ is an eigenvector of T with eigenvalue λ. Then v is also an eigenvector of T^* with eigenvalue $\bar{\lambda}$.

Proof Because T is normal, so is $T - \lambda I$, as you should verify.
Thus

$$\begin{aligned} 0 &= \|(T - \lambda I)v\| \\ &= \|(T - \lambda I)^* v\| \\ &= \|(T^* - \bar{\lambda} I)v\|. \end{aligned}$$

Hence v is an eigenvector of T^* with eigenvalue $\bar{\lambda}$, as desired. ∎

Orthogonal Eigenvectors for Normal Operators

Orthogonal eigenvectors for normal operators

Suppose $T \in \mathcal{L}(V)$ is normal. Then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof Suppose α, β are distinct eigenvalues of T, with corresponding eigenvectors u, v. Thus $Tu = \alpha u$ and $Tv = \beta v$.

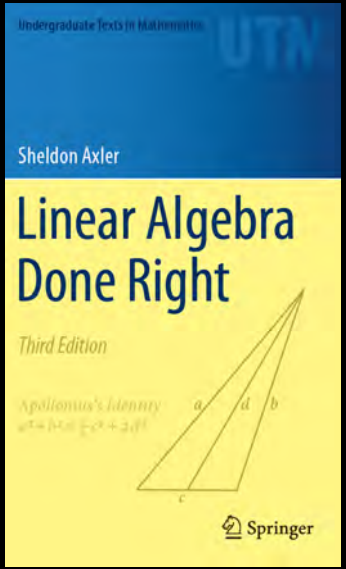
We have $T^*v = \bar{\beta}v$. Thus

$$\begin{aligned}(\alpha - \beta)\langle u, v\rangle &= \langle \alpha u, v\rangle - \langle u, \bar{\beta}v\rangle \\ &= \langle Tu, v\rangle - \langle u, T^*v\rangle \\ &= 0.\end{aligned}$$

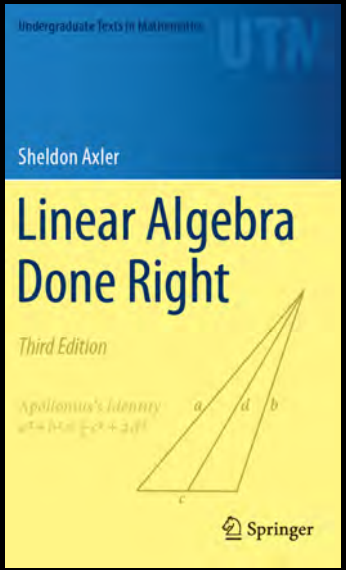
Because $\alpha \neq \beta$, the equation above implies that $\langle u, v\rangle = 0$. Thus u and v are orthogonal, as desired. ∎

Isaac Newton (1642 – 1727)

Linear Algebra Done Right, by Sheldon Axler



The Spectral Theorem



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional inner product space over $\mathbf{F}$.

Diagonalization by an Orthonormal Basis?

- A diagonal matrix is a square matrix that is 0 everywhere except possibly along the diagonal.
- An operator on V has a diagonal matrix with respect to a basis if and only if the basis consists of eigenvectors of the operator.
- For which operators on V is there an *orthonormal* basis of V with respect to which the operator has a diagonal matrix?
- For which operators on V is there an *orthonormal* basis of V consisting of eigenvectors of T?
- The Spectral Theorem will answer these questions.
- The answer is depends upon whether the scalar field is real or complex.

The Complex Spectral Theorem

Complex Spectral Theorem

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

Proof First suppose (c) holds.

The matrix of T^* is obtained by taking the conjugate transpose of the matrix of T; hence T^* also has a diagonal matrix.

Any two diagonal matrices commute; thus T commutes with T^*.

Thus T is normal. In other words, (a) holds.

We have proved that (c) implies (a). The equivalence of (b) and (c) is easy.

We will complete the proof by showing that (a) implies (c).

The Complex Spectral Theorem

Complex Spectral Theorem

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

Proof Suppose (a) holds, so T is normal. By Schur's Theorem , there is an orthonormal basis $e_1, \ldots, e_n$ of V with respect to which T has an upper-triangular matrix. Thus

$$\mathcal{M}\big(T, (e_1, \ldots, e_n)\big) = \begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}.$$

The condition $\|Te_1\| = \|T^*e_1\|$ implies that all entries in the first row of the matrix, except possibly the first entry $a_{1,1}$, equal 0.

The condition $\|Te_2\| = \|T^*e_2\|$ implies that all entries in the second row of the matrix, except possibly the diagonal entry $a_{2,2}$, equal 0.

Continuing in this fashion, we see that all the nondiagonal entries in the matrix equal 0. Thus the matrix above is diagonal, and (c) holds. ∎

The Complex Spectral Theorem

Complex Spectral Theorem

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

Example: Consider the normal operator $T \in \mathcal{L}(\mathbf{C}^2)$ whose matrix is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

As you can verify,

$$\frac{(i,1)}{\sqrt{2}}, \frac{(-i,1)}{\sqrt{2}}$$

is an orthonormal basis of $\mathbf{C}^2$ consisting of eigenvectors of T.

With respect to this basis the matrix of T is the diagonal matrix

$$\begin{pmatrix} 2+3i & 0 \\ 0 & 2-3i \end{pmatrix}.$$

The Real Spectral Theorem

Real Spectral Theorem

Suppose $\mathbf{F} = \mathbf{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is self-adjoint.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

Example: Consider the self-adjoint operator T on $\mathbf{R}^3$ whose matrix is

$$\begin{pmatrix} 14 & -13 & 8 \\ -13 & 14 & 8 \\ 8 & 8 & -7 \end{pmatrix}.$$

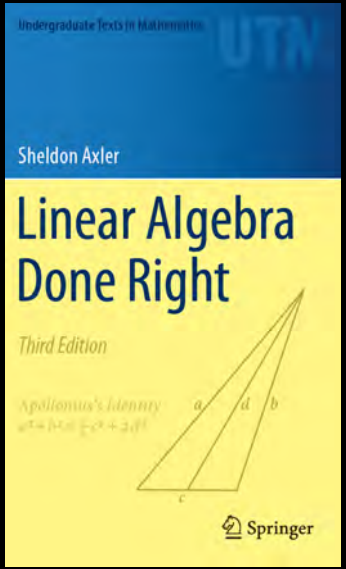
As you can verify,

$$\frac{(1,-1,0)}{\sqrt{2}}, \frac{(1,1,1)}{\sqrt{3}}, \frac{(1,1,-2)}{\sqrt{6}}$$

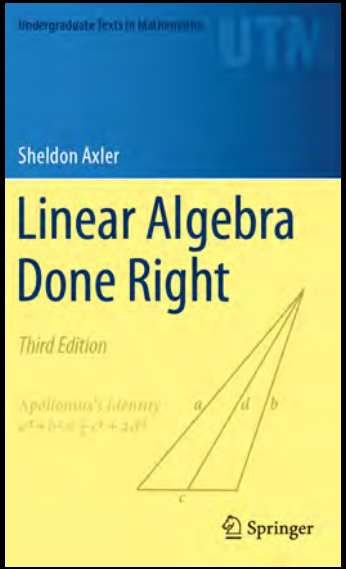
is an orthonormal basis of $\mathbf{R}^3$ consisting of eigenvectors of T. With respect to this basis, the matrix of T is the diagonal matrix

$$\begin{pmatrix} 27 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & -15 \end{pmatrix}.$$

Linear Algebra Done Right, by Sheldon Axler



Positive Operators and Isometries, part 1: Positive Operators



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional inner product space over $\mathbf{F}$.

Definition and Examples of Positive Operators

Definition: *positive operator*

An operator $T \in \mathcal{L}(V)$ is called *positive* if T is self-adjoint and

$$\langle Tv, v \rangle \geq 0$$

for all $v \in V$.

If V is a complex vector space, then the requirement that T is self-adjoint can be dropped from the definition above.

Examples:

- If U is a subspace of V, then the orthogonal projection P_U is a positive operator.
- If $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbf{R}$ are such that $b^2 \leq 4c$, then $T^2 + bT + cI$ is a positive operator.
- If $\mathcal{M}(T)$ is a diagonal matrix with nonnegative entries on the diagonal, then T is a positive operator.

Square Roots of Operators

Definition: *square root*

An operator R is called a *square root* of an operator T if $R^2 = T$.

Example: If $T \in \mathcal{L}(\mathbf{F}^3)$ is defined by

$$T(z_1, z_2, z_3) = (z_3, 0, 0),$$

then the operator $R \in \mathcal{L}(\mathbf{F}^3)$ defined by

$$R(z_1, z_2, z_3) = (z_2, z_3, 0)$$

is a square root of T.

Example: If $T \in \mathcal{L}(\mathbf{F}^3)$ has matrix

$$\mathcal{M}(T) = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix},$$

then the operator $R \in \mathcal{L}(\mathbf{F}^3)$ with matrix

$$\mathcal{M}(R) = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

is a square root of T.

Characterization of Positive Operators

Characterization of positive operators

Let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is positive;

(b) T is self-adjoint and all eigenvalues of T are nonnegative;

(c) T has a positive square root;

(d) T has a self-adjoint square root;

(e) there exists $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Proof We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (a).

First suppose (a) holds, so that T is positive. Obviously T is self-adjoint (by the definition of a positive operator). To prove the other condition in (b), suppose λ is an eigenvalue of T. Let v be an eigenvector of T corresponding to λ. Then

$$\begin{aligned} 0 &\leq \langle Tv, v\rangle \\ &= \langle \lambda v, v\rangle \\ &= \lambda\langle v, v\rangle. \end{aligned}$$

Thus λ is a nonnegative number. Hence (b) holds.

Characterization of Positive Operators

Characterization of positive operators

Let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is positive;

(b) T is self-adjoint and all eigenvalues of T are nonnegative;

(c) T has a positive square root;

(d) T has a self-adjoint square root;

(e) there exists $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Now suppose (b) holds, so that T is self-adjoint and all eigenvalues of T are nonnegative. By the Spectral Theorem, there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. Each $\lambda_j \geq 0$. Let $R \in \mathcal{L}(V)$ be such that

$$Re_j = \sqrt{\lambda_j} e_j$$

for $j = 1, \ldots, n$. Then R is a positive operator. Furthermore, $R^2 e_j = \lambda_j e_j = Te_j$ for each j, which implies that $R^2 = T$.
Thus R is a positive square root of T.
Hence (c) holds.

Characterization of Positive Operators

Characterization of positive operators

Let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is positive;

(b) T is self-adjoint and all eigenvalues of T are nonnegative;

(c) T has a positive square root;

(d) T has a self-adjoint square root;

(e) there exists $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Clearly (c) implies (d) (because, by definition, every positive operator is self-adjoint).

Now suppose (d) holds, meaning that there exists a self-adjoint operator R on V such that

$$T = R^2.$$

Then

$$T = R^*R$$

because $R^* = R$. Hence (e) holds.

Characterization of Positive Operators

Characterization of positive operators

Let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is positive;

(b) T is self-adjoint and all eigenvalues of T are nonnegative;

(c) T has a positive square root;

(d) T has a self-adjoint square root;

(e) there exists $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Finally, suppose (e) holds. Let $R \in \mathcal{L}(V)$ be such that $T = R^*R$. Now

$$\begin{aligned} T^* &= (R^*R)^* \\ &= R^*(R^*)^* \\ &= R^*R \\ &= T. \end{aligned}$$

Hence T is self-adjoint. To complete the proof that (a) holds, note that

$$\begin{aligned} \langle Tv, v\rangle &= \langle R^*Rv, v\rangle \\ &= \langle Rv, Rv\rangle \\ &\geq 0 \end{aligned}$$

for every $v \in V$. Thus T is positive. ∎

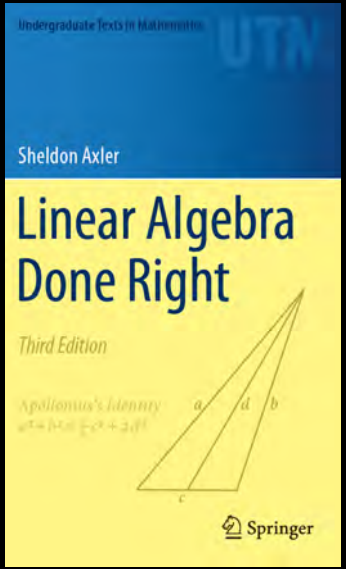
Uniqueness of Positive Square Root

Each positive operator has only one positive square root

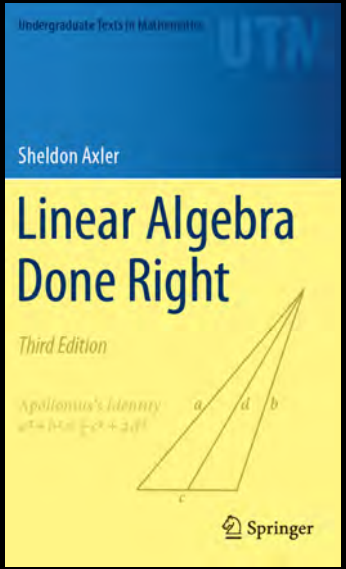
Every positive operator on V has a unique positive square root.

A positive operator can have infinitely many square roots, although only one of them can be positive. For example, the identity operator on V has infinitely many square roots if $\dim V > 1$.

Linear Algebra Done Right, by Sheldon Axler



Positive Operators and Isometries, part 2: Isometries



Definition and Example of Isometries

Definition: *isometry*

- An operator $S \in \mathcal{L}(V)$ is called an *isometry* if
$$\|Sv\| = \|v\|$$
for all $v \in V$.
- In other words, an operator is an isometry if it preserves norms.

Example: Suppose $\lambda \in \mathbf{F}$. Then λI is an isometry if and only if $|\lambda| = 1$.

Example: Suppose $\lambda_1, \ldots, \lambda_n$ are scalars with $|\lambda_j| = 1$ and $S \in \mathcal{L}(V)$ satisfies $Se_j = \lambda_j e_j$ for some orthonormal basis $e_1, \ldots, e_n$ of V. Suppose $v \in V$. Then

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

and

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

Now

$$\begin{aligned} Sv &= \langle v, e_1 \rangle Se_1 + \cdots + \langle v, e_n \rangle Se_n \\ &= \lambda_1 \langle v, e_1 \rangle e_1 + \cdots + \lambda_n \langle v, e_n \rangle e_n. \end{aligned}$$

Thus

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

Hence $\|v\| = \|Sv\|$. Thus S is an isometry.

Characterization of Isometries

Characterization of isometries

For $S \in \mathcal{L}(V)$, the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \ldots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \ldots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_1, \ldots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

Proof We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f) $\Rightarrow$ (g) $\Rightarrow$ (h) $\Rightarrow$ (a).

First suppose (a) holds, so S is an isometry. If $\mathbf{F} = \mathbf{R}$, then for every $u, v \in V$ we have

$$\begin{aligned}\langle u, v \rangle &= \frac{\|u+v\|^2 - \|u-v\|^2}{4} \\ &= \frac{\|S(u+v)\|^2 - \|S(u-v)\|^2}{4} \\ &= \frac{\|Su+Sv\|^2 - \|Su-Sv\|^2}{4} \\ &= \langle Su, Sv \rangle,\end{aligned}$$

and thus (b) holds.

Characterization of isometries

For $S \in \mathcal{L}(V)$, the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \ldots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \ldots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_1, \ldots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

Now suppose (b) holds, so S preserves inner products. Suppose that $e_1, \ldots, e_n$ is an orthonormal list of vectors in V. Then the list $Se_1, \ldots, Se_n$ is orthonormal because

$$\langle Se_j, Se_k \rangle = \langle e_j, e_k \rangle.$$

Thus (c) holds.

Clearly (c) implies (d).

Characterization of Isometries

Characterization of isometries

For $S \in \mathcal{L}(V)$, the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \ldots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \ldots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_1, \ldots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

Now suppose (d) holds.
Let $e_1, \ldots, e_n$ be an orthonormal basis of V such that $Se_1, \ldots, Se_n$ is orthonormal. Thus

$$\begin{aligned} \langle S^*Se_j, e_k \rangle &= \langle Se_j, Se_k \rangle \\ &= \langle e_j, e_k \rangle \end{aligned}$$

for $j, k = 1, \ldots, n$. All vectors $u, v \in V$ can be written as linear combinations of $e_1, \ldots, e_n$, and thus the equation above implies that

$$\langle S^*Su, v \rangle = \langle u, v \rangle$$

for all $u, v \in V$. Hence $S^*S = I$; in other words, (e) holds.

Characterization of isometries

For $S \in \mathcal{L}(V)$, the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \ldots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \ldots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_1, \ldots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

That (e) implies (f) is a special case of the result that if $S, T \in \mathcal{L}(V)$ and $TS = I$ implies $ST = I$.

Now suppose (f) holds, so $SS^* = I$. If $v \in V$, then

$$\begin{aligned} \|S^*v\|^2 &= \langle S^*v, S^*v \rangle \\ &= \langle SS^*v, v \rangle \\ &= \langle v, v \rangle \\ &= \|v\|^2. \end{aligned}$$

Thus S^* is an isometry, showing that (g) holds.

Characterization of isometries

For $S \in \mathcal{L}(V)$, the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \ldots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \ldots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_1, \ldots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

Now suppose (g) holds, so S^* is an isometry.
Using the implications (a) $\Rightarrow$ (e) and (a) $\Rightarrow$ (f) but with S replaced with S^* [and using the equation $(S^*)^* = S$], we conclude that

$$SS^* = I \quad \text{and} \quad S^*S = I.$$

Thus S is invertible and $S^{-1} = S^*$; in other words, (h) holds.

Characterization of isometries

For $S \in \mathcal{L}(V)$, the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \ldots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \ldots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_1, \ldots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

Now suppose (h) holds, so S is invertible and $S^{-1} = S^*$.
Thus $S^*S = I$.
If $v \in V$, then

$$\begin{aligned} \|Sv\|^2 &= \langle Sv, Sv \rangle \\ &= \langle S^*Sv, v \rangle \\ &= \langle v, v \rangle \\ &= \|v\|^2. \end{aligned}$$

Thus S is an isometry, showing that (a) holds. ∎

Description of isometries when $\mathbf{F} = \mathbf{C}$

Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$. The following are equivalent:

(a) S is an isometry.

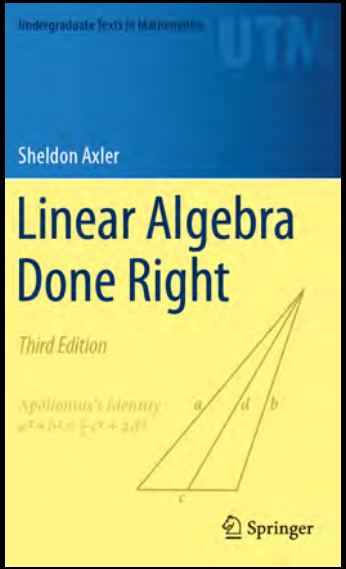
(b) There is an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have absolute value 1.

Proof Suppose (a) holds, so S is an isometry. By the Complex Spectral Theorem, there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of S. For $j \in \{1, \ldots, n\}$, let λ_j be the eigenvalue corresponding to e_j. Then

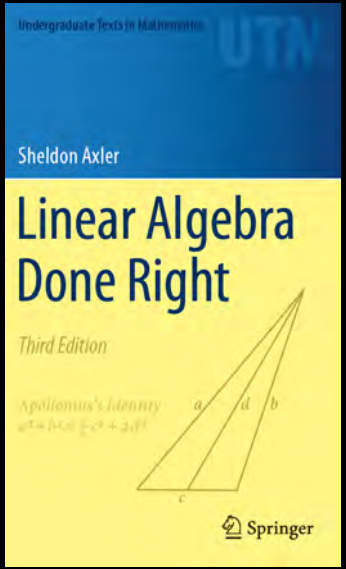
$$\begin{aligned} |\lambda_j| &= \|\lambda_j e_j\| \\ &= \|Se_j\| \\ &= \|e_j\| \\ &= 1. \end{aligned}$$

Thus each eigenvalue of S has absolute value 1. Hence (b) holds.
We earlier showed that (b) implies (a). ∎

Linear Algebra Done Right, by Sheldon Axler



Polar Decomposition and SVD, part 1: Polar Decomposition



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional inner product space over $\mathbf{F}$.

Positive Operators

Definition: *positive operator*

An operator $T \in \mathcal{L}(V)$ is called *positive* if T is self-adjoint and

$$\langle Tv, v \rangle \geq 0$$

for all $v \in V$.

Definition: *square root*

An operator R is called a *square root* of an operator T if $R^2 = T$.

Characterization of positive operators

Let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is positive;

(b) T is self-adjoint and all eigenvalues of T are nonnegative;

(c) T has a positive square root;

(d) T has a self-adjoint square root;

(e) there exists $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Square Roots

Each positive operator has only one positive square root

Every positive operator on V has a unique positive square root.

Notation: $\sqrt{T}$

If T is a positive operator, then $\sqrt{T}$ denotes the unique positive square root of T.

Analogy Between $\mathbf{C}$ and $\mathcal{L}(V)$

Polar decomposition on $\mathbf{C}$:

$$z = r(\cos\theta + i\sin\theta), \text{ with } r = |z|$$

$$= \left(\frac{z}{|z|}\right)|z|$$

$$= \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z}, \text{ with } \left|\frac{z}{|z|}\right| = 1$$

$\mathbf{C}$	$\mathcal{L}(V)$
z	T
$\bar{z}$	T^*
$z \in \mathbf{R}$ $(z = \bar{z})$	T is self-adjoint $(T = T^*)$
$z \geq 0$	T is positive
$(z = \overline{w}w$ for some $w)$	$(T = R^*R$ for some $R)$
$\lvert z\rvert = 1$ $(\bar{z}z = 1)$	T is an isometry $(T^*T = I)$
$z = \left(\frac{z}{\lvert z\rvert}\right)\sqrt{\bar{z}z}$	$T = S\sqrt{T^*T}$ for some isometry S

Polar Decomposition

Suppose $T \in \mathcal{L}(V)$. Then there exists an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Proof If $v \in V$, then

$$\begin{aligned}\|Tv\|^2 &= \langle Tv, Tv\rangle \\ &= \langle T^*Tv, v\rangle \\ &= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle \\ &= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle \\ &= \|\sqrt{T^*T}v\|^2.\end{aligned}$$

Thus

$$\|Tv\| = \|\sqrt{T^*T}v\|.$$

Define $S_1 : \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ by

$$S_1(\sqrt{T^*T}v) = Tv.$$

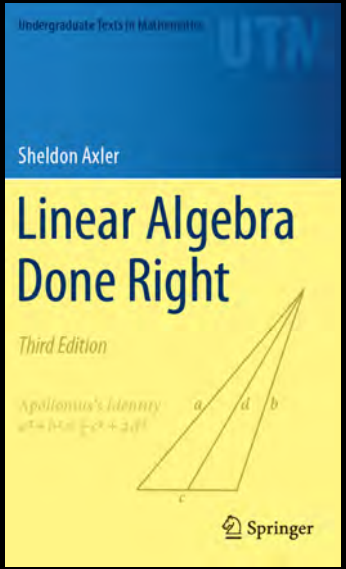
The equation $\|\sqrt{T^*T}v\| = \|Tv\|$ shows that

$$\|S_1 u\| = \|u\|$$

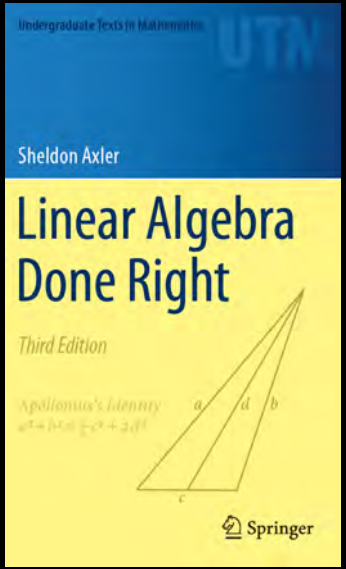
for all $u \in \operatorname{range}\sqrt{T^*T}$.

Extend S_1 to an isometry $S \in \mathcal{L}(V)$.
Do this by defining S to be an isometry from $(\operatorname{range}\sqrt{T^*T})^\perp$ to $(\operatorname{range} T)^\perp$, then extend by linearity. ∎

Linear Algebra Done Right, by Sheldon Axler



Polar Decomposition and SVD, part 2: Singular Value Decomposition



Singular Values

Definition: *singular values*

Suppose $T \in \mathcal{L}(V)$. The *singular values* of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ listed

$$\dim \operatorname{null}(\sqrt{T^*T} - \lambda I)$$

times.

The singular values of T are nonnegative numbers, because they are eigenvalues of the positive operator $\sqrt{T^*T}$.

Each operator on V has $\dim V$ singular values.

Example:Define $T \in \mathcal{L}(\mathbf{F}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4).$$

A calculation shows that

$$T^*T(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4).$$

Thus

$$\sqrt{T^*T}(z_1, z_2, z_3, z_4) = (3z_1, 2z_2, 0, 3z_4).$$

The eigenvalues of $\sqrt{T^*T}$ are $3, 2, 0$ and

$$\dim \operatorname{null}(\sqrt{T^*T} - 3I) = 2,$$

$$\dim \operatorname{null}(\sqrt{T^*T} - 2I) = 1,$$

$$\dim \operatorname{null} \sqrt{T^*T} = 1.$$

Hence the singular values of T are $3, 3, 2, 0$.

Singular Value Decomposition

Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \ldots, s_n$. Then there exist orthonormal bases $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ of V such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for every $v \in V$.

With notation as in the statement of the Singular Value Decomposition,

$$Te_j = s_j f_j$$

for each j. Thus

$$\mathcal{M}\big(T,(e_1,\ldots,e_n),(f_1,\ldots,f_n)\big) = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{pmatrix}.$$

Singular Value Decomposition

Singular Value Decomposition

Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \ldots, s_n$. Then there exist orthonormal bases $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ of V such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for every $v \in V$.

Proof By the Spectral Theorem, there is an orthonormal basis $e_1, \ldots, e_n$ of V such that $\sqrt{T^*T}e_j = s_j e_j$ for $j = 1, \ldots, n$. Now

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

for every $v \in V$.

Apply $\sqrt{T^*T}$ to both sides of this equation, getting

$$\sqrt{T^*T}v = s_1\langle v, e_1\rangle e_1 + \cdots + s_n\langle v, e_n\rangle e_n$$

for every $v \in V$. By the Polar Decomposition, there is an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. Apply S to both sides of the equation above, getting

$$Tv = s_1\langle v, e_1\rangle Se_1 + \cdots + s_n\langle v, e_n\rangle Se_n$$

for every $v \in V$. For each j, let $f_j = Se_j$. Because S is an isometry, $f_1, \ldots, f_n$ is an orthonormal basis of V. The equation above now becomes

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for every $v \in V$, completing the proof. ∎

Singular Values Without Taking Square Roots

Singular values without taking square root of an operator

Suppose $T \in \mathcal{L}(V)$. Then the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, with each eigenvalue λ listed

$$\dim \operatorname{null}(T^*T - \lambda I)$$

times.

Example:Define $T \in \mathcal{L}(\mathbf{F}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4).$$

A calculation shows that

$$T^*T(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4).$$

The eigenvalues of T^*T are $9, 4, 0$ and

$$\dim \operatorname{null}(T^*T - 9I) = 2,$$
$$\dim \operatorname{null}(T^*T - 4I) = 1,$$
$$\dim \operatorname{null} T^*T = 1.$$

Hence the singular values of T are $3, 3, 2, 0$.

Exercises

- Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Prove that the singular values of T equal the absolute values of the eigenvalues of T, repeated appropriately.
- Suppose $T \in \mathcal{L}(V)$. Prove that T and T^* have the same singular values.
- Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if 0 is not a singular value of T.
- Suppose $T \in \mathcal{L}(V)$. Prove that $\dim \operatorname{range} T$ equals the number of nonzero singular values of T.
- Suppose $S \in \mathcal{L}(V)$. Prove that S is an isometry if and only if all the singular values of S equal 1.
- Suppose $T \in \mathcal{L}(V)$. Let $\hat{s}$ denote the smallest singular value of T, and let s denote the largest singular value of T.
  (a) Prove that $\hat{s}\|v\| \le \|Tv\| \le s\|v\|$ for every $v \in V$.
  (b) Suppose λ is an eigenvalue of T. Prove that $\hat{s} \le |\lambda| \le s$.

Exercises

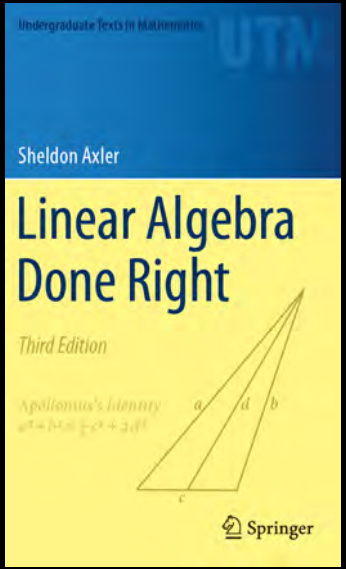
Suppose $T \in \mathcal{L}(V)$ has singular value decomposition given by

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

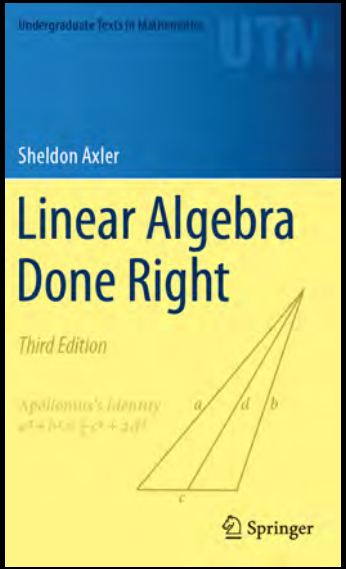
for every $v \in V$, where $s_1, \ldots, s_n$ are the singular values of T and $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ are orthonormal bases of V.

- Prove that if $v \in V$, then
$$T^*v = s_1\langle v, f_1\rangle e_1 + \cdots + s_n\langle v, f_n\rangle e_n.$$
- Prove that if $v \in V$, then
$$T^*Tv = {s_1}^2\langle v, e_1\rangle e_1 + \cdots + {s_n}^2\langle v, e_n\rangle e_n.$$
- Prove that if $v \in V$, then
$$\sqrt{T^*T}v = s_1\langle v, e_1\rangle e_1 + \cdots + s_n\langle v, e_n\rangle e_n.$$
- Suppose T is invertible. Prove that if $v \in V$, then
$$T^{-1}v = \frac{\langle v, f_1\rangle e_1}{s_1} + \cdots + \frac{\langle v, f_n\rangle e_n}{s_n}$$
for every $v \in V$.

Linear Algebra Done Right, by Sheldon Axler



Null Spaces of Powers of an Operator



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

Sequence of Null Spaces

Sequence of increasing null spaces

Suppose $T \in \mathcal{L}(V)$. Then

$$\{0\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^k \subset \operatorname{null} T^{k+1} \subset \cdots .$$

Proof Suppose k is a nonnegative integer and $v \in \operatorname{null} T^k$. Then

$$T^k v = 0.$$

Hence

$$\begin{aligned} T^{k+1} v &= T(T^k v) \\ &= T(0) \\ &= 0. \end{aligned}$$

Thus $v \in \operatorname{null} T^{k+1}$. Hence $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$, as desired. ∎

Sequence of Null Spaces

Sequence of increasing null spaces

Suppose $T \in \mathcal{L}(V)$. Then

$$\{0\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^k \subset \operatorname{null} T^{k+1} \subset \cdots .$$

Equality in the sequence of null spaces

Suppose $T \in \mathcal{L}(V)$. Suppose m is a nonnegative integer such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. Then

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \operatorname{null} T^{m+3} = \cdots .$$

Proof Let k be a positive integer. We want to prove that

$$\operatorname{null} T^{m+k} = \operatorname{null} T^{m+k+1}.$$

We already know that $\operatorname{null} T^{m+k} \subset \operatorname{null} T^{m+k+1}$.

To prove the inclusion in the other direction, suppose $v \in \operatorname{null} T^{m+k+1}$. Then

$$\begin{aligned} T^{m+1}(T^k v) &= T^{m+k+1} v \\ &= 0. \end{aligned}$$

Hence

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

Thus

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

which means that $v \in \operatorname{null} T^{m+k}$. This implies that $\operatorname{null} T^{m+k+1} \subset \operatorname{null} T^{m+k}$, completing the proof. ∎

Null Spaces Stop Growing

Null spaces stop growing

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$\operatorname{null} T^n = \operatorname{null} T^{n+1} = \operatorname{null} T^{n+2} = \cdots .$$

Proof We need only prove that $\operatorname{null} T^n = \operatorname{null} T^{n+1}$. Suppose this is not true. Then

$$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^n \subsetneq \operatorname{null} T^{n+1},$$

At each of the strict inclusions in the chain above, the dimension increases by at least 1. Thus $\dim \operatorname{null} T^{n+1} \geq n + 1$, a contradiction because a subspace of V cannot have a larger dimension than n. ∎

The Direct Sum of $\operatorname{null} T^{\dim V}$ and $\operatorname{range} T^{\dim V}$

V *is the direct sum of* $\operatorname{null} T^{\dim V}$ *and* $\operatorname{range} T^{\dim V}$

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n.$$

Proof First we show that

$$(\operatorname{null} T^n) \cap (\operatorname{range} T^n) = \{0\}.$$

Suppose $v \in (\operatorname{null} T^n) \cap (\operatorname{range} T^n)$. Then

$$T^n v = 0,$$

and there exists $u \in V$ such that

$$T^n u = v.$$

Applying T^n to both sides shows that

$$\begin{aligned} T^{2n} u &= T^n v \\ &= 0. \end{aligned}$$

Thus $T^n u = 0$. Thus $v = T^n u = 0$, showing that

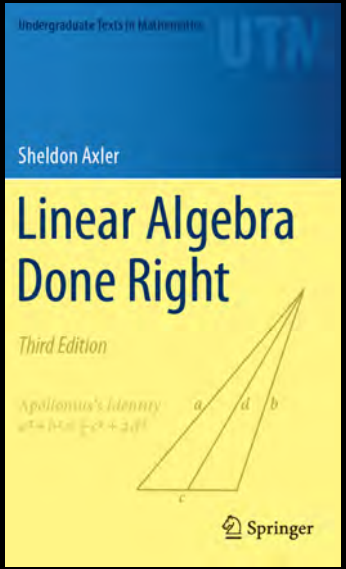
$$(\operatorname{null} T^n) \cap (\operatorname{range} T^n) = \{0\}.$$

Thus $\operatorname{null} T^n + \operatorname{range} T^n$ is a direct sum. Now

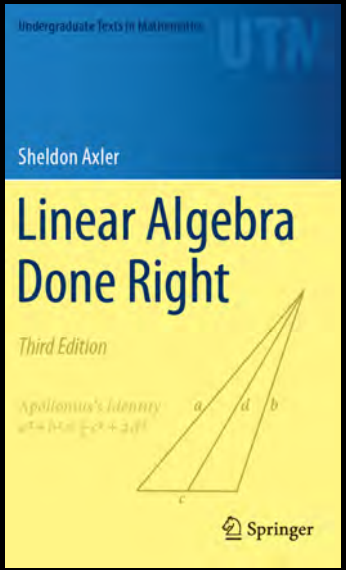
$$\begin{aligned} &\dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) \\ &\quad = \dim \operatorname{null} T^n + \dim \operatorname{range} T^n \\ &\quad = \dim V. \end{aligned}$$

The equation above implies that $\operatorname{null} T^n \oplus \operatorname{range} T^n = V$, as desired. ∎

Linear Algebra Done Right, by Sheldon Axler



Generalized Eigenvectors



Eigenvalue and Eigenvector Review

Definition: *eigenvalue*

Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbf{F}$ is called an *eigenvalue* of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$.

Definition: *eigenvector*

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$ is an eigenvalue of T. A vector $v \in V$ is called an *eigenvector* of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

Definition: *eigenspace,* $E(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The *eigenspace* of T corresponding to λ, denoted $E(\lambda, T)$, is defined by

$$E(\lambda, T) = \operatorname{null}(T - \lambda I).$$

Clearly

$$v \in \operatorname{null}(T - \lambda I) \iff Tv = \lambda v.$$

Thus $E(\lambda, T)$ is the set of eigenvectors of T corresponding to λ, along with the 0 vector.

Generalized Eigenvectors

Definition: *generalized eigenvector*

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

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

for some positive integer j.

Definition: *generalized eigenspace,* $G(\lambda, T)$

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

Description of generalized eigenspaces

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

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

Example of Generalized Eigenspaces

Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

The eigenvalues of T are 0 and 5. The corresponding eigenspaces are easily seen to be

$$E(0, T) = \{(z_1, 0, 0) : z_1 \in \mathbf{C}\}$$

and

$$E(5, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}.$$

Note that this operator T does not have enough eigenvectors to span its domain $\mathbf{C}^3$.

We have

$$T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$$

for all $z_1, z_2, z_3 \in \mathbf{C}$. Thus

$$G(0, T) = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{C}\}.$$

We have

$$(T-5I)^3(z_1, z_2, z_3) = (-125z_1+300z_2, -125z_2, 0).$$

Thus

$$G(5, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}.$$

Now

$$\mathbf{C}^3 = G(0, T) \oplus G(5, T).$$

Linearly Independent Generalized Eigenvectors

Linearly independent generalized eigenvectors

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

Proof Suppose $a_1, \ldots, a_m \in \mathbf{F}$ are such that

$$0 = a_1 v_1 + \cdots + a_m v_m.$$

Let k be the largest nonnegative integer such that $(T - \lambda_1 I)^k v_1 \neq 0$. Let

$$w = (T - \lambda_1 I)^k v_1.$$

Thus $(T - \lambda_1 I)w = (T - \lambda_1 I)^{k+1} v_1 = 0$.

Hence $Tw = \lambda_1 w$. Thus $(T - \lambda I)w = (\lambda_1 - \lambda)w$ for every $\lambda \in \mathbf{F}$. Hence

$$(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w$$

for every $\lambda \in \mathbf{F}$, where $n = \dim V$. Apply the operator

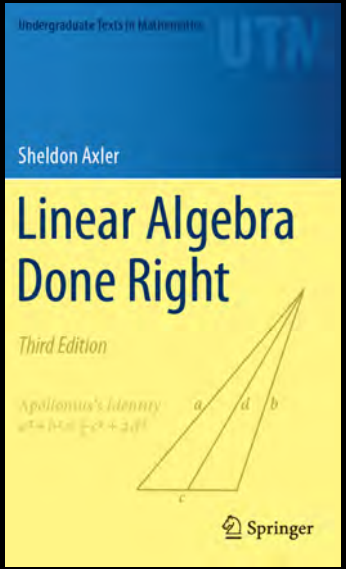
$$(T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n$$

to both sides, getting

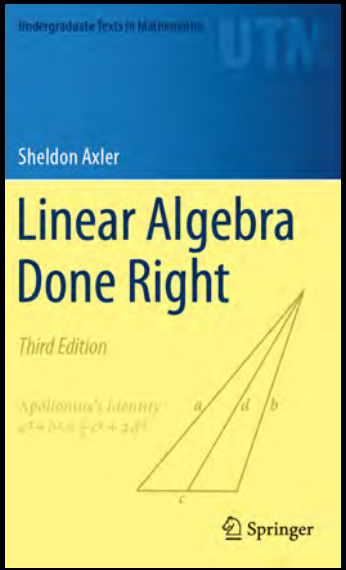
$$\begin{aligned} 0 &= a_1 (T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n v_1 \\ &= a_1 (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n w \\ &= a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n w. \end{aligned}$$

The equation above implies that $a_1 = 0$. In a similar fashion, $a_j = 0$ for each j. Thus $v_1, \ldots, v_m$ is linearly independent. ∎

Linear Algebra Done Right, by Sheldon Axler



Nilpotent Operators



Definition and Examples of Nilpotent Operator

Definition: *nilpotent*

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

Example: The operator $N \in \mathcal{L}(\mathbf{F}^4)$ defined by

$$N(z_1, z_2, z_3, z_4) = (z_3, z_4, 0, 0)$$

is nilpotent because $N^2 = 0$.

Example The operator of differentiation on $\mathcal{P}_m(\mathbf{R})$ is nilpotent because the $(m+1)^{\text{st}}$ derivative of every polynomial of degree at most m equals 0.

***Nilpotent operator raised to dimension of domain is* 0**

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then

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

Matrix of a Nilpotent Operator

Matrix of a nilpotent operator

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

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

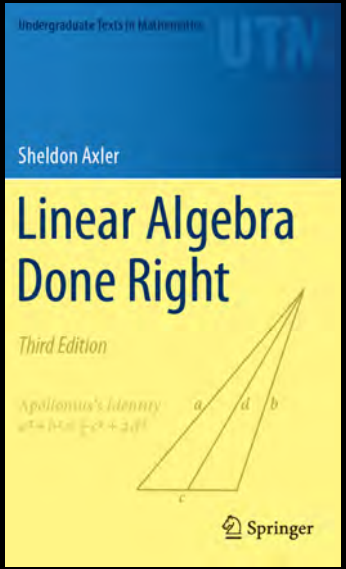
entries on and below diagonal are 0's.

Proof First choose a basis of $\operatorname{null} N$. Then extend this to a basis of $\operatorname{null} N^2$. Then extend to a basis of $\operatorname{null} N^3$. Continue in this fashion, eventually getting a basis of V.

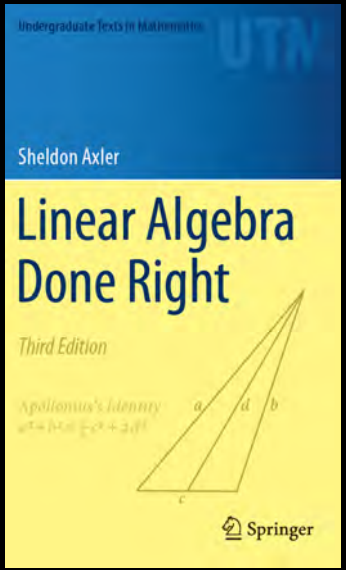
The first column of the matrix of N, and perhaps additional columns at the beginning, consists of all 0's. The next set of columns comes from basis vectors in $\operatorname{null} N^2$. Applying N to any such vector, we get a vector in $\operatorname{null} N$; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus all nonzero entries in these columns lie above the diagonal. The next set of columns comes from basis vectors in $\operatorname{null} N^3$. Applying N to any such vector, we get a vector in $\operatorname{null} N^2$; in other words, we get a vector that is a linear combination of the previous basis vectors. Continue in this fashion to complete the proof. ∎

Hypatia ($\approx 355 - 415$)

Linear Algebra Done Right, by Sheldon Axler



Decomposition via Generalized Eigenvectors



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

Null Space and Range of $p(T)$ Are Invariant

The null space and range of $p(T)$ are invariant under T

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$. Then $\operatorname{null} p(T)$ and $\operatorname{range} p(T)$ are invariant under T.

Proof Suppose $v \in \operatorname{null} p(T)$. Then $p(T)v = 0$. Thus

$$\begin{aligned}\big((p(T)\big)(Tv) &= T\big(p(T)v\big)\\ &= T(0)\\ &= 0.\end{aligned}$$

Hence $Tv \in \operatorname{null} p(T)$. Thus $\operatorname{null} p(T)$ is invariant under T, as desired.

Suppose $v \in \operatorname{range} p(T)$. Then there exists $u \in V$ such that $v = p(T)u$. Thus

$$\begin{aligned}Tv &= T\big(p(T)u\big)\\ &= p(T)(Tu).\end{aligned}$$

Hence $Tv \in \operatorname{range} p(T)$. Thus $\operatorname{range} p(T)$ is invariant under T, as desired. ∎

Decomposition Theorem

Description of operators on complex vector spaces

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then

(a) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$;

(b) each $G(\lambda_j, T)$ is invariant under T;

(c) each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.

Proof of (b): Recall that

$$G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^{\dim V}$$

and then use the previous result that $\operatorname{null} p(T)$ is invariant under T for every polynomial p.

Proof of (c): Clear from definition of $(T - \lambda_j I)|_{G(\lambda_j, T)}$.

Basis of Generalized Eigenvectors

Description of operators on complex vector spaces

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then

(a) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$;

(b) each $G(\lambda_j, T)$ is invariant under T;

(c) each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.

A basis of generalized eigenvectors

Suppose that V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.

Proof Choose a basis of each $G(\lambda_j, T)$. Put all these bases together to form a basis of V consisting of generalized eigenvectors of T. ∎

Multiplicity

Definition: *multiplicity*

- Suppose $T \in \mathcal{L}(V)$. The *multiplicity* of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$.
- In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null}(T - \lambda I)^{\dim V}$.

Sum of the multiplicities equals $\dim V$

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then the sum of the multiplicities of all the eigenvalues of T equals $\dim V$.

Proof $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$

algebraic multiplicity of $\lambda = \dim \operatorname{null}(T - \lambda I)^{\dim V} = \dim G(\lambda, T)$

geometric multiplicity of $\lambda = \dim \operatorname{null}(T - \lambda I) = \dim E(\lambda, T)$

Multiplicity Example

Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (3z_1 + 4z_2, 3z_2, 8z_3).$$

$$\mathcal{M}(T) = \begin{pmatrix} 3 & 4 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

The eigenvalues of T are 3 and 8.

The eigenspaces of T are

$$E(3, T) = \{(z_1, 0, 0) : z_1 \in \mathbf{C}\},$$
$$E(8, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}.$$

Thus the eigenvalue 3 has geometric multiplicity 1 and the eigenvalue 8 has geometric multiplicity 1.

The generalized eigenspaces of T are

$$G(3, T) = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{C}\},$$
$$G(8, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}.$$

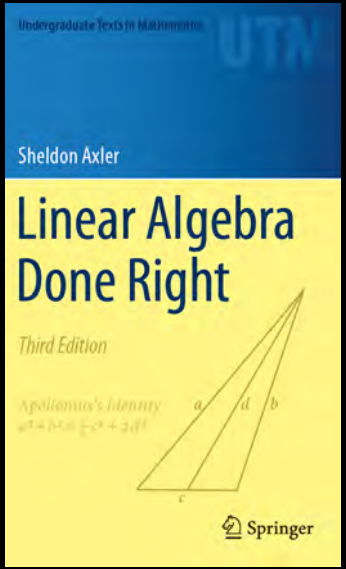
Thus the eigenvalue 3 has algebraic multiplicity 2 and the eigenvalue 8 has algebraic multiplicity 1.

We have

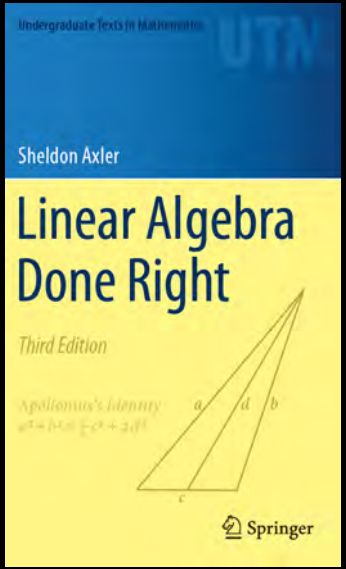
$$\mathbf{C}^3 = G(3, T) \oplus G(8, T),$$

as expected by the Decomposition Theorem.

Linear Algebra Done Right, by Sheldon Axler



Block Diagonal Matrices



Definition of Block Diagonal Matrix

Definition: *block diagonal matrix*

A *block diagonal matrix* is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where $A_1, \ldots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0.

Example: The 5-by-5 matrix

$$A = \begin{pmatrix} \begin{pmatrix} 4 \end{pmatrix} & 0 & 0 & 0 & 0 \\ 0 & \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix} & & 0 & 0 \\ 0 & & & 0 & 0 \\ 0 & 0 & 0 & \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix} & \\ 0 & 0 & 0 & & \end{pmatrix}$$

is a block diagonal matrix with

$$A = \begin{pmatrix} A_1 & & 0 \\ & A_2 & \\ 0 & & A_3 \end{pmatrix},$$

where

$$A_1 = \begin{pmatrix} 4 \end{pmatrix}, A_2 = \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix}, A_3 = \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix}.$$

Block Diagonal Matrix with Upper-Triangular Blocks

Block diagonal matrix with upper-triangular blocks

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then there is a basis of V with respect to which T has a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_j is a d_j-by-d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

Example of Block Diagonal

Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (6z_1+3z_2+4z_3, 6z_2+2z_3, 7z_3).$$

The matrix of T (with respect to the standard basis) is

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix},$$

which is an upper-triangular matrix but is not of the form promised by our result. The eigenvalues of T are 6 and 7 with corresponding generalized eigenspaces

$$G(6, T) = \operatorname{span}\big((1, 0, 0), (0, 1, 0)\big),$$
$$G(7, T) = \operatorname{span}\big((10, 2, 1)\big).$$

A basis of $\mathbf{C}^3$ consisting of generalized eigenvectors of T is

$$(1, 0, 0), (0, 1, 0), (10, 2, 1).$$

The matrix of T with respect to this basis is

$$\begin{pmatrix} \begin{pmatrix} 6 & 3 \\ 0 & 6 \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \begin{pmatrix} 7 \end{pmatrix} \end{pmatrix},$$

which is a matrix of the appropriate block diagonal form.

Block Diagonal Matrix with Upper-Triangular Blocks

Block diagonal matrix with upper-triangular blocks

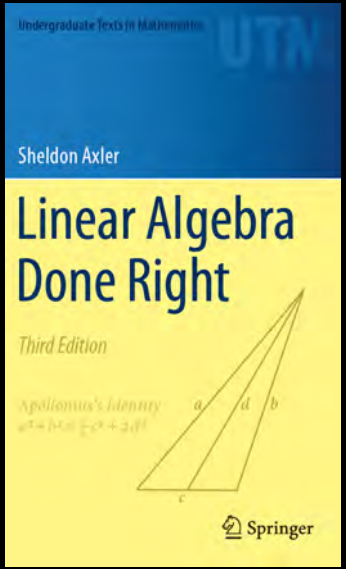
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then there is a basis of V with respect to which T has a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

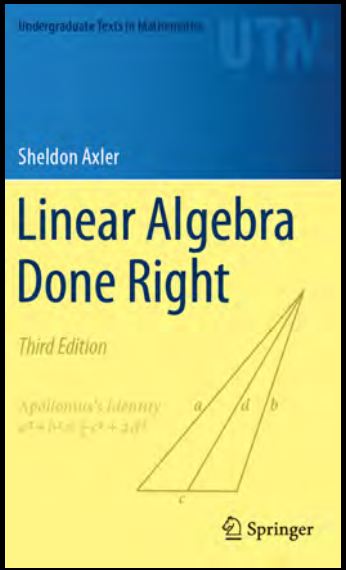
where each A_j is a d_j-by-d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

Linear Algebra Done Right, by Sheldon Axler



Square Roots of Operators



Example of Operator with No Square Root

Definition: *square root*

An operator R is called a *square root* of an operator T if $R^2 = T$.

Example: Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (z_2, z_3, 0).$$

Then T does not have a square root.

Identity Plus Nilpotent Has a Square Root

Identity plus nilpotent has a square root

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.

Proof Consider the Taylor series for the function $\sqrt{1+x}$:

$$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots .$$

Because N is nilpotent, $N^m = 0$ for some positive integer m. Thus we guess that there is a square root of $I + N$ of the form

$$I + a_1 N + a_2 N^2 + a_3 N^3 + \cdots + a_{m-1} N^{m-1}.$$

Now

$$\begin{aligned}
&(I + a_1 N + a_2 N^2 + a_3 N^3 + \cdots + a_{m-1} N^{m-1})^2 \\
&= I + 2a_1 N + (2a_2 + {a_1}^2) N^2 + (2a_3 + 2a_1 a_2) N^3 + \cdots \\
&\quad + (2a_{m-1} + \text{terms involving } a_1, \ldots, a_{m-2}) N^{m-1} \\
&= I + N
\end{aligned}$$

if $2a_1 = 1$ (thus $a_1 = 1/2$) and $2a_2 + {a_1}^2 = 0$ (thus $a_2 = -1/8$) and $a_3 = 1/16$ and .... Continue in this fashion for $j = 4, \ldots, m-1$, at each step solving for a_j so that the coefficient of N^j on the right side of the equation above equals 0. ∎

Over $\mathbf{C}$, Invertible Operators Have Square Roots

Over $\mathbf{C}$, invertible operators have square roots

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Then T has a square root.

Proof Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. For each j, there exists a nilpotent operator $N_j \in \mathcal{L}\big(G(\lambda_j, T)\big)$ such that $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$. Because T is invertible, none of the λ_j's equals 0, so we can write

$$T|_{G(\lambda_j,T)} = \lambda_j\Big(I + \frac{N_j}{\lambda_j}\Big).$$

Clearly N_j/λ_j is nilpotent, and so $I + N_j/\lambda_j$ has a square root. Multiplying a square root of the number λ_j by a square root of $I + N_j/\lambda_j$ gives a square root R_j of $T|_{G(\lambda_j,T)}$. A typical vector $v \in V$ can be written uniquely in the form

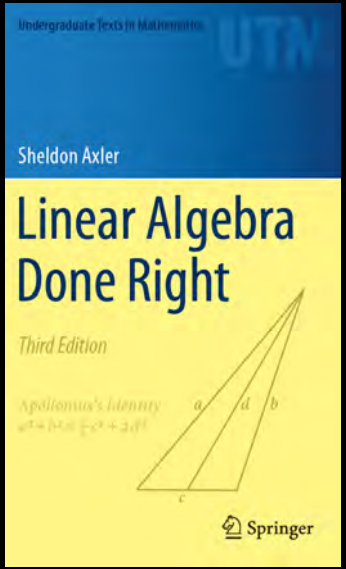
$$v = u_1 + \cdots + u_m,$$

where each u_j is in $G(\lambda_j, T)$. Using this decomposition, define an operator $R \in \mathcal{L}(V)$ by

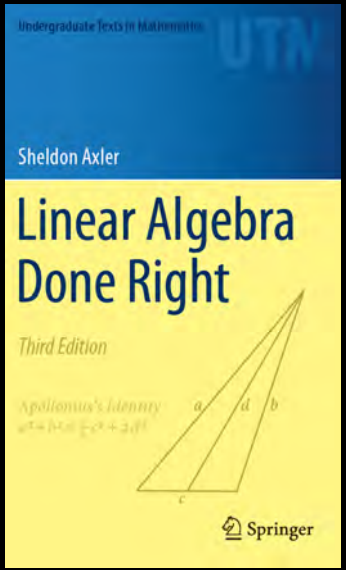
$$Rv = R_1 u_1 + \cdots + R_m u_m.$$

This operator R is a square root of T, completing the proof. ∎

Linear Algebra Done Right, by Sheldon Axler



The Characteristic Polynomial



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

Definition of Characteristic Polynomial

Definition: *characteristic polynomial*

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the *characteristic polynomial* of T.

Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

The matrix of T (with respect to the standard basis) is

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}.$$

Because the eigenvalues of T are 6, with multiplicity 2, and 7, with multiplicity 1, the characteristic polynomial of T is

$$(z - 6)^2(z - 7).$$

Definition of Characteristic Polynomial

Definition: *characteristic polynomial*

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the *characteristic polynomial* of T.

Degree and zeros of characteristic polynomial

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then

- the characteristic polynomial of T has degree $\dim V$;
- the zeros of the characteristic polynomial of T are the eigenvalues of T.

Cayley–Hamilton Theorem

Cayley–Hamilton Theorem

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.

Proof Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of the operator T.

Let $d_1, \ldots, d_m$ be the dimensions of the corresponding generalized eigenspaces $G(\lambda_1, T), \ldots, G(\lambda_m, T)$.

Each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent and

$$(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0.$$

Every vector in V is a sum of vectors in $G(\lambda_1, T), \ldots, G(\lambda_m, T)$. To prove that $q(T) = 0$, we need only show that $q(T)|_{G(\lambda_j, T)} = 0$ for each j.

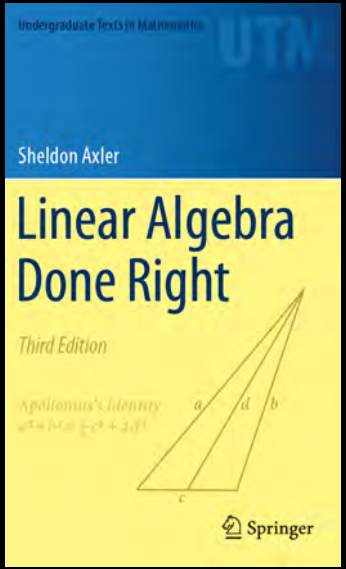
Thus fix $j \in \{1, \ldots, m\}$. We have

$$q(T) = (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_m I)^{d_m}.$$

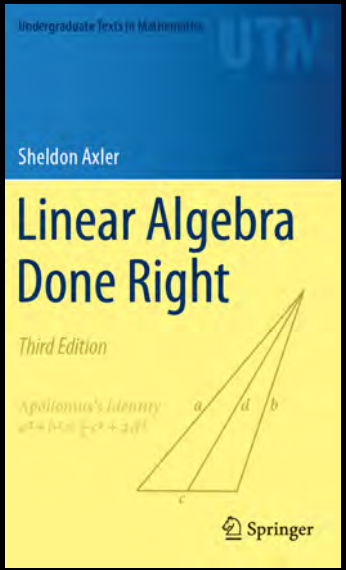
The operators on the right side of the equation above all commute, so we can move the factor $(T - \lambda_j I)^{d_j}$ to be the last term in the expression on the right.

Because $(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0$, we conclude that $q(T)|_{G(\lambda_j, T)} = 0$, as desired. ∎

Linear Algebra Done Right, by Sheldon Axler



The Minimal Polynomial



Monic Polynomials

Definition: *monic polynomial*

A *monic polynomial* is a polynomial whose highest-degree coefficient equals 1.

Example: The polynomial

$$2 + 9z^2 + z^7$$

is a monic polynomial of degree 7.

Existence of Minimal Polynomial

Minimal polynomial

Suppose $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial p of smallest degree such that $p(T) = 0$.

Proof Let $n = \dim V$. The list

$$I, T, T^2, \ldots, T^{n^2}$$

is not linearly independent in $\mathcal{L}(V)$, because $\mathcal{L}(V)$ has dimension n^2 and the list has length $n^2 + 1$. Let m be the smallest positive integer such that

$$I, T, T^2, \ldots, T^m$$

is linearly dependent.

The Linear Dependence Lemma implies that T^m is a linear combination of $I, T, T^2, \ldots, T^{m-1}$. Thus there exist scalars $a_0, a_1, a_2, \ldots, a_{m-1} \in \mathbf{F}$ such that

$$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0.$$

Define a monic polynomial $p \in \mathcal{P}(\mathbf{F})$ by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m.$$

Then $p(T) = 0$.
No monic polynomial $q \in \mathcal{P}(\mathbf{F})$ with degree smaller than m can satisfy $q(T) = 0$. Suppose $q \in \mathcal{P}(\mathbf{F})$ is a monic polynomial with degree m and $q(T) = 0$. Then $(p - q)(T) = 0$ and $\deg(p - q) < m$. Thus $q = p$, completing the proof. ∎

Definition of Minimal Polynomial

Definition: *minimal polynomial*

Suppose $T \in \mathcal{L}(V)$. The *minimal polynomial* of T is the unique monic polynomial p of smallest degree such that $p(T) = 0$.

Example: Let T be the operator on $\mathbf{C}^5$ whose matrix is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Find the minimal polynomial of T.

We have

$$3\mathcal{M}(I) - 6\mathcal{M}(T) = -\mathcal{M}(T)^5$$

with no solutions for lower powers.
Thus the minimal polynomial of T is

$$3 - 6z + z^5.$$

For general $T \in \mathcal{L}(V)$, consider the system of linear equations

$$a_0\mathcal{M}(I) + a_1\mathcal{M}(T) + \cdots + a_{m-1}\mathcal{M}(T)^{m-1} = -\mathcal{M}(T)^m$$

for successive values of $m = 1, 2, \ldots$ until this system of equations has a solution $a_0, a_1, a_2, \ldots, a_{m-1}$.
The scalars $a_0, a_1, a_2, \ldots, a_{m-1}, 1$ will then be the coefficients of the minimal polynomial of T.

Polynomials That Annihilate T

$q(T) = 0$ *implies* q *is a multiple of the minimal polynomial*

Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.

Proof Let p denote the minimal polynomial of T.

Suppose q is a polynomial multiple of p. Thus there exists a polynomial $s \in \mathcal{P}(\mathbf{F})$ such that $q = ps$. We have

$$q(T) = p(T)s(T) = 0,$$

as desired.

To prove the other direction, now suppose $q(T) = 0$. By the Division Algorithm for Polynomials, there exist polynomials $s, r \in \mathcal{P}(\mathbf{F})$ such that

$$q = ps + r$$

and $\deg r < \deg p$. We have

$$\begin{aligned} 0 &= q(T) \\ &= p(T)s(T) + r(T) \\ &= r(T). \end{aligned}$$

The equation above implies that $r = 0$. Thus $q = ps$. Hence q is a polynomial multiple of p, as desired. ∎

Polynomials That Annihilate T

$q(T) = 0$ *implies* q *is a multiple of the minimal polynomial*

Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.

Characteristic polynomial is a multiple of minimal polynomial

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

Zeros of the Minimal Polynomial

Eigenvalues are the zeros of the minimal polynomial

Let $T \in \mathcal{L}(V)$. Then the zeros of the minimal polynomial of T are precisely the eigenvalues of T.

Proof Let p be the minimal polynomial of T.
First suppose $\lambda \in \mathbf{F}$ is a zero of p.
Then p can be written in the form

$$p(z) = (z - \lambda)q(z),$$

where q is a monic polynomial.
Now for every $v \in V$ we have

$$\begin{aligned} 0 &= p(T)v \\ &= (T - \lambda I)(q(T)v). \end{aligned}$$

Because the degree of q is less than the degree of the minimal polynomial p, there exists at least one vector $v \in V$ such that $q(T)v \neq 0$.

Thus λ is an eigenvalue of T.

We have shown that every zero of p is an eigenvalue of T.

Zeros of the Minimal Polynomial

Eigenvalues are the zeros of the minimal polynomial

Let $T \in \mathcal{L}(V)$. Then the zeros of the minimal polynomial of T are precisely the eigenvalues of T.

Proof Let p be the minimal polynomial of T.
Suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T. Thus there exists $v \in V$ with $v \neq 0$ such that

$$Tv = \lambda v.$$

Repeated applications of T to both sides of this equation show that

$$T^j v = \lambda^j v$$

for every nonnegative integer j. Thus

$$p(T)v = p(\lambda)v.$$

Now

$$\begin{aligned} 0 &= p(T)v \\ &= p(\lambda)v. \end{aligned}$$

Because $v \neq 0$, the equation above implies that $p(\lambda) = 0$.
We have shown that every eigenvalue of T is a zero of p. ∎

Examples

Example: Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (6z_1+3z_2+4z_3, 6z_2+2z_3, 7z_3).$$

The eigenvalues of T are 6 and 7.
Thus the minimal polynomial of T is a polynomial multiple of $(z-6)(z-7)$.
The characteristic polynomial of T is

$$(z-6)^2(z-7).$$

Thus the minimal polynomial of T is either $(z-6)(z-7)$ or $(z-6)^2(z-7)$.
A simple computation shows that

$$(T-6I)(T-7I) \neq 0.$$

Thus the minimal polynomial of T is

$$(z-6)^2(z-7).$$

Example: Suppose $T \in \mathcal{L}(\mathbf{C}^3)$is defined by

$$T(z_1, z_2, z_3) = (6z_1, 6z_2, 7z_3).$$

For this operator T, the eigenvalues of T are 6 and 7.
The characteristic polynomial of T is

$$(z-6)^2(z-7).$$

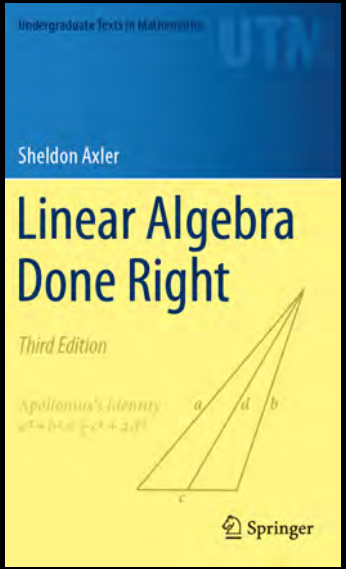
Thus the minimal polynomial of T is either $(z-6)(z-7)$ or $(z-6)^2(z-7)$.
A simple computation shows that

$$(T-6I)(T-7I) = 0.$$

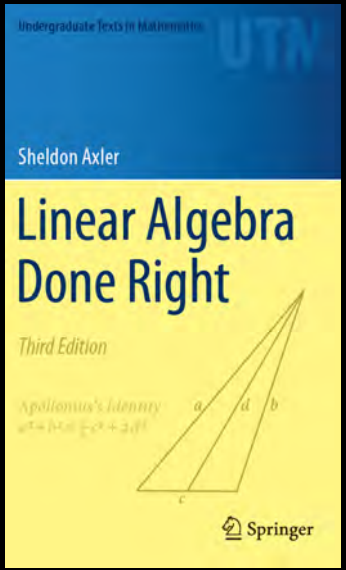
Thus the minimal polynomial of T is

$$(z-6)(z-7).$$

Linear Algebra Done Right, by Sheldon Axler



Jordan Form


Undergraduate Texts in Mathematics
UTM
Sheldon Axler
Linear Algebra Done Right
Third Edition
Apollonius's Identity
a
d
b
c
Springer

Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

Examples

Basis corresponding to a nilpotent operator

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then there exist vectors $v_1, \ldots, v_n \in V$ and nonnegative integers $m_1, \ldots, m_n$ such that

- $N^{m_1}v_1, \ldots, Nv_1, v_1, \ldots, N^{m_n}v_n, \ldots, Nv_n, v_n$ is a basis of V;
- $N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$.

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

Example: Let $N \in \mathcal{L}(\mathbf{F}^6)$ be the nilpotent operator defined by

$$N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0).$$

Let $v_1 = (1, 0, 0, 0, 0, 0)$,
$v_2 = (0, 0, 0, 1, 0, 0)$,
$v_3 = (0, 0, 0, 0, 0, 1)$.
With respect to the basis

$$N^2v_1, Nv_1, v_1, Nv_2, v_2, v_3,$$

the matrix of N is

$$\begin{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \end{matrix} & \begin{pmatrix} 0 \end{pmatrix} \end{pmatrix}.$$

Definition: *Jordan basis*

Suppose $T \in \mathcal{L}(V)$. A basis of V is called a *Jordan basis* for T if with respect to this basis T has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

where each A_j is of the form

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

Jordan Form

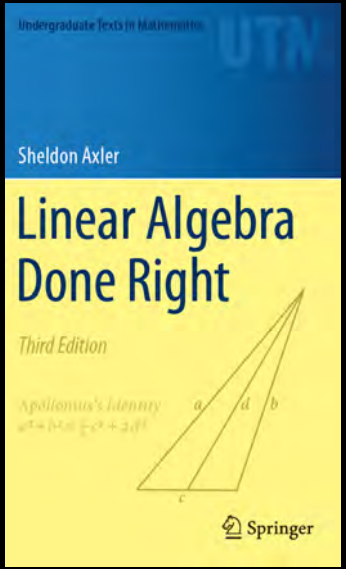
Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then there is a basis of V that is a Jordan basis for T.

Proof Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. We have the generalized eigenspace decomposition

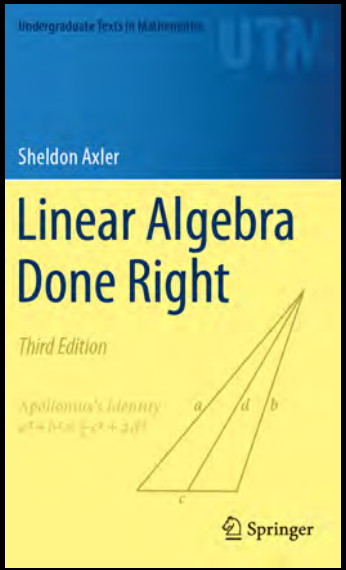
$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus some basis of each $G(\lambda_j, T)$ is a Jordan basis for $(T - \lambda_j I)|_{G(\lambda_j, T)}$. Put these bases together to get a basis of V that is a Jordan basis for T. ∎

Linear Algebra Done Right, by Sheldon Axler



Complexification



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

Complexification of a Vector Space

Definition: *complexification of V, $V_{\mathbf{C}}$*

Suppose V is a real vector space.

- The *complexification* of V, denoted $V_{\mathbf{C}}$, equals $V \times V$. An element of $V_{\mathbf{C}}$ is an ordered pair (u, v), where $u, v \in V$, but we will write this as $u + iv$.
- Addition on $V_{\mathbf{C}}$ is defined by
  $$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$
  for $u_1, v_1, u_2, v_2 \in V$.
- Complex scalar multiplication on $V_{\mathbf{C}}$ is defined by
  $$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$
  for $a, b \in \mathbf{R}$ and $u, v \in V$.

$V_{\mathbf{C}}$ *is a complex vector space.*

Suppose V is a real vector space. Then $V_{\mathbf{C}}$ is a complex vector space.

Basis of V is basis of $V_{\mathbf{C}}$

Suppose V is a real vector space.

- If $v_1, \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is a basis of $V_{\mathbf{C}}$ (as a complex vector space).
- The dimension of $V_{\mathbf{C}}$ (as a complex vector space) equals the dimension of V (as a real vector space).

Complexification of an Operator

Definition: *complexification of T, $T_{\mathbf{C}}$*

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. The *complexification* of T, denoted $T_{\mathbf{C}}$, is the operator $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$ defined by

$$T_{\mathbf{C}}(u + iv) = Tu + iTv$$

for $u, v \in V$.

Example: Suppose A is an n-by-n matrix of real numbers. Define $T \in \mathcal{L}(\mathbf{R}^n)$ by

$$Tx = Ax,$$

where elements of $\mathbf{R}^n$ are thought of as n-by-1 column vectors.

Identifying the complexification of $\mathbf{R}^n$ with $\mathbf{C}^n$, we then have

$$T_{\mathbf{C}} z = Az$$

for each $z \in \mathbf{C}^n$, where again elements of $\mathbf{C}^n$ are thought of as n-by-1 column vectors.

Matrix of $T_{\mathbf{C}}$ equals matrix of T

Suppose V is a real vector space with basis $v_1, \ldots, v_n$ and $T \in \mathcal{L}(V)$. Then $\mathcal{M}(T) = \mathcal{M}(T_{\mathbf{C}})$, where both matrices are with respect to the basis $v_1, \ldots, v_n$.

Invariant Subspaces

Every operator has an invariant subspace of dimension 1 or 2

Every operator on a nonzero finite-dimensional vector space has an invariant subspace of dimension 1 or 2.

Proof Every operator on a nonzero finite-dimensional complex vector space has an eigenvalue and thus has a 1-dimensional invariant subspace.

Hence assume V is a real vector space and $T \in \mathcal{L}(V)$.

The complexification $T_{\mathbf{C}}$ has an eigenvalue $a + bi$, where $a, b \in \mathbf{R}$. Thus there exist $u, v \in V$, not both 0, such that

$$T_{\mathbf{C}}(u + iv) = (a + bi)(u + iv).$$

The last equation can be rewritten as

$$Tu + iTv = (au - bv) + (av + bu)i.$$

Thus

$$Tu = au - bv \quad \text{and} \quad Tv = av + bu.$$

Let U equal the span in V of the list u, v. Then U is a subspace of V with dimension 1 or 2.
The equations above show that U is invariant under T, completing the proof. ∎

Minimal Polynomial of the Complexification

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Repeated application of the definition of $T_{\mathbf{C}}$ shows that

$$(T_{\mathbf{C}})^n(u + iv) = T^n u + iT^n v$$

for every positive integer n and all $u, v \in V$.

Minimal polynomial of $T_{\mathbf{C}}$ equals minimal polynomial of T

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the minimal polynomial of $T_{\mathbf{C}}$ equals the minimal polynomial of T.

Eigenvalues of the Complexification

Real eigenvalues of $T_{\mathbf{C}}$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{R}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if λ is an eigenvalue of T.

Nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if $\bar{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$.

Multiplicity of λ equals multiplicity of $\bar{\lambda}$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$ is an eigenvalue of $T_{\mathbf{C}}$. Then the multiplicity of λ as an eigenvalue of $T_{\mathbf{C}}$ equals the multiplicity of $\bar{\lambda}$ as an eigenvalue of $T_{\mathbf{C}}$.

Operators on Odd-Dimensional Vector Space Have Eigenvalues

Operator on odd-dimensional vector space has eigenvalue

Every operator on an odd-dimensional real vector space has an eigenvalue.

Proof Suppose V is a real vector space with odd dimension and $T \in \mathcal{L}(V)$.

Because the nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs with equal multiplicity, the sum of the multiplicities of all the nonreal eigenvalues of $T_{\mathbf{C}}$ is an even number.

Because the sum of the multiplicities of all the eigenvalues of $T_{\mathbf{C}}$ equals the (complex) dimension of $V_{\mathbf{C}}$, this implies that $T_{\mathbf{C}}$ has a real eigenvalue.

Every real eigenvalue of $T_{\mathbf{C}}$ is also an eigenvalue of T, giving the desired result. ∎

Characteristic Polynomial of Complexification

Characteristic polynomial of $T_{\mathbf{C}}$

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real.

Proof Suppose λ is a nonreal eigenvalue of $T_{\mathbf{C}}$ with multiplicity m. Then $\bar{\lambda}$ is also an eigenvalue of $T_{\mathbf{C}}$ with multiplicity m.

Thus the characteristic polynomial of $T_{\mathbf{C}}$ includes factors of $(z-\lambda)^m$ and $(z-\bar{\lambda})^m$.

Multiplying together these two factors, we have

$$\begin{aligned}(z-\lambda)^m(z-\bar{\lambda})^m &= \big((z-\lambda)(z-\bar{\lambda})\big)^m \\ &= \big(z^2-2(\operatorname{Re}\lambda)z+|\lambda|^2\big)^m,\end{aligned}$$

which has real coefficients.

The characteristic polynomial of $T_{\mathbf{C}}$ is the product of terms of the form above and terms of the form $(z-t)^d$, where t is a real eigenvalue of $T_{\mathbf{C}}$ with multiplicity d.

Thus the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real. ∎

Characteristic Polynomial

Definition: *Characteristic polynomial*

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the *characteristic polynomial* of T is defined to be the characteristic polynomial of $T_{\mathbf{C}}$.

Example: Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ is defined by

$$T(x_1, x_2, x_3) = (2x_1, x_2 - x_3, x_2 + x_3).$$

The eigenvalues of $T_{\mathbf{C}}$ are 2, $1 + i$, and $1 - i$, each with multiplicity 1.
Thus the characteristic polynomial of the complexification $T_{\mathbf{C}}$ is

$$(z - 2)\big(z - (1 + i)\big)\big(z - (1 - i)\big),$$

which equals

$$z^3 - 4z^2 + 6z - 4.$$

Hence the characteristic polynomial of T is also

$$z^3 - 4z^2 + 6z - 4.$$

Properties of the Characteristic Polynomial

Degree and zeros of characteristic polynomial

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then

- the coefficients of the characteristic polynomial of T are all real;
- the characteristic polynomial of T has degree $\dim V$;
- the eigenvalues of T are precisely the real zeros of the characteristic polynomial of T.

Cayley–Hamilton Theorem

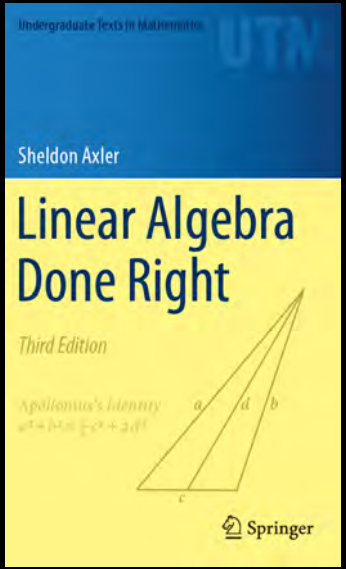
Suppose $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.

Characteristic polynomial is a multiple of minimal polynomial

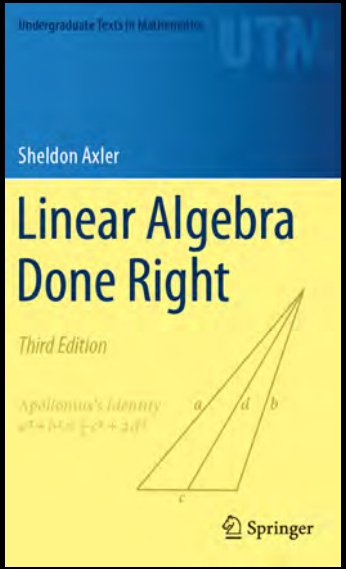
Suppose $T \in \mathcal{L}(V)$. Then

- the degree of the minimal polynomial of T is at most $\dim V$;
- the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

Linear Algebra Done Right, by Sheldon Axler



Normal Operators on Real Inner Product Spaces



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero inner product space over $\mathbf{F}$.

Normal but not self-adjoint operators

Suppose $\dim V = 2$, $\mathbf{F} = \mathbf{R}$, and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal but not self-adjoint.

(b) The matrix of T with respect to every orthonormal basis of V has the form

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

with $b \neq 0$.

(c) The matrix of T with respect to some orthonormal basis of V has the form

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

with $b > 0$.

Normal operators and invariant subspaces

Suppose $T \in \mathcal{L}(V)$ is normal and U is a subspace of V that is invariant under T. Then

(a) $U^\perp$ is invariant under T;

(b) U is invariant under T^*;

(c) $(T|_U)^* = (T^*)|_U$;

(d) $T|_U \in \mathcal{L}(U)$ and $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ are normal operators.

Proof Let $e_1, \ldots, e_m$ be an orthonormal basis of U. Extend to an orthonormal basis $e_1, \ldots, e_m, f_1, \ldots, f_n$ of V. Thus

$$\mathcal{M}(T) = \begin{array}{c} \\ e_1 \\ \vdots \\ e_m \\ f_1 \\ \vdots \\ f_n \end{array} \begin{array}{c} \begin{array}{cccccc} e_1 & \ldots & e_m & f_1 & \ldots & f_n \end{array} \\ \begin{pmatrix} & A & & & B & \\ & & & & & \\ & & & & & \\ & 0 & & & C & \\ & & & & & \end{pmatrix} \end{array}.$$

Because T is normal,

$$\sum_{j=1}^{m} \|Te_j\|^2 = \sum_{j=1}^{m} \|T^*e_j\|^2.$$

Thus B is the matrix of all 0's, proving (a).

Description of Normal Operators on Real Inner Product Spaces

Characterization of normal operators when $\mathbf{F} = \mathbf{R}$

Suppose V is a real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:

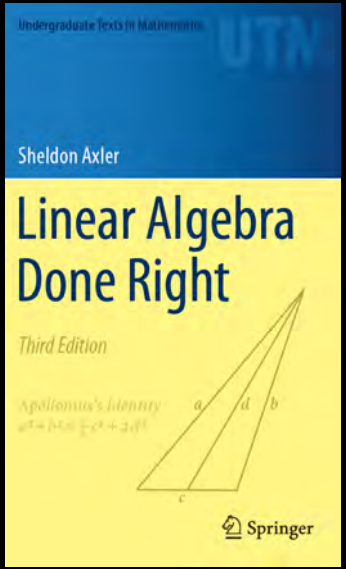
(a) T is normal.

(b) There is an orthonormal basis of V with respect to which T has a block diagonal matrix such that each block is a 1-by-1 matrix or a 2-by-2 matrix of the form

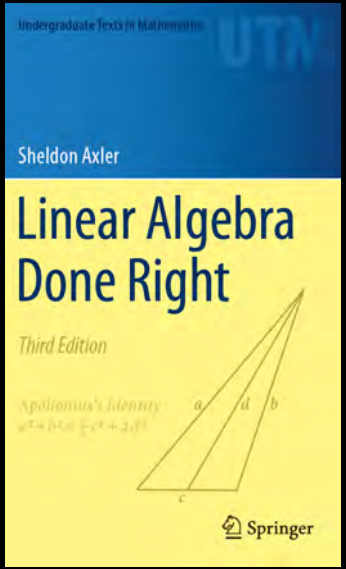
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b > 0$.

Linear Algebra Done Right, by Sheldon Axler



Isometries on Real Inner Product Spaces



Review of Isometries

Definition: *isometry*

- An operator $S \in \mathcal{L}(V)$ is called an *isometry* if
$$\|Sv\| = \|v\|$$
for all $v \in V$.
- In other words, an operator is an isometry if it preserves norms.

Every isometry is normal

Suppose $S \in \mathcal{L}(V)$. Then the following are equivalent:

- S is an isometry.
- $S^*S = I$.
- $SS^* = I$.

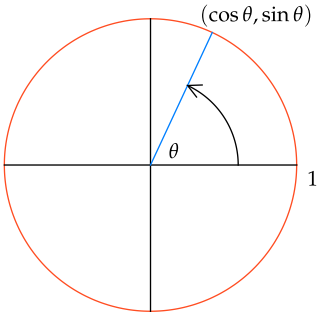
Examples of Isometries

Example: Let $\theta \in \mathbf{R}$. Then the operator on $\mathbf{R}^2$ of counterclockwise rotation by an angle of θ is an isometry, as is geometrically obvious.

The matrix of this operator with respect to the standard basis is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

If θ is not an integer multiple of π, then no nonzero vector of $\mathbf{R}^2$ gets mapped to a scalar multiple of itself, and hence the operator has no eigenvalues.



Description of Isometries on Real Inner Product Spaces

Description of isometries when $\mathbf{F} = \mathbf{R}$

Suppose V is a real inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent:

(a) S is an isometry.

(b) There is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block on the diagonal is a 1-by-1 matrix containing 1 or -1 or is a 2-by-2 matrix of the form

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

with $\theta \in (0, \pi)$.

Proof Suppose (a) holds, so S is an isometry. There is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block is a 1-by-1 matrix or a 2-by-2 matrix of the form

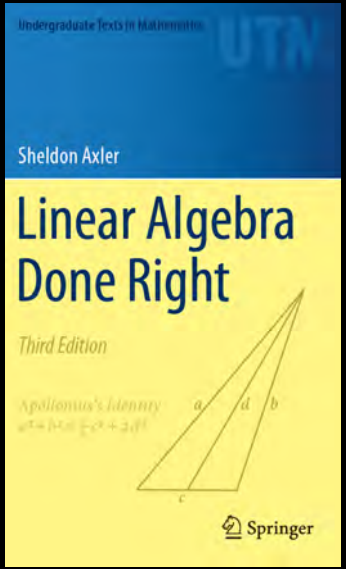
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b > 0$.

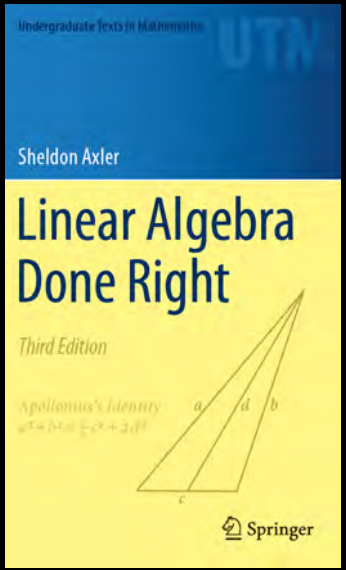
If λ is an entry in a 1-by-1 matrix along the diagonal of the matrix of S, then $\lambda = 1$ or $\lambda = -1$. Because $a^2 + b^2 = 1$ and $b > 0$, we can choose $\theta \in (0, \pi)$ such that $a = \cos\theta$ and $b = \sin\theta$. Thus the matrix has the required form, proving that (a) implies (b).

Euclid Explaining Geometry

Linear Algebra Done Right, by Sheldon Axler



Change of Basis



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

Identity Matrix; Inverse of a Matrix

Definition: *identity matrix, I*

Suppose n is a positive integer. The n-by-n diagonal matrix

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

is called the *identity matrix*.

Definition: *invertible, inverse,* A^{-1}

A square matrix A is called *invertible* if there is a square matrix B of the same size such that

$$AB = BA = I;$$

we call B the *inverse* of A and denote it by A^{-1}.

Matrix of an Operator

Definition: *matrix of an operator,* $\mathcal{M}(T)$

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are bases of V. The *matrix of T* with respect to these bases is the n-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{n,k}w_n.$$

If the bases are not clear from the context, then use the notation $\mathcal{M}\big(T, (v_1, \ldots, v_n), (w_1, \ldots, w_n)\big)$.

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_n \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & \ldots & v_k & \ldots & v_n \end{array} \\ \begin{pmatrix} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{n,k} & & \end{pmatrix} \end{array}.$$

The k^{th} column of $\mathcal{M}(T)$ consists of the scalars needed to write Tv_k as a linear combination of $w_1, \ldots, w_n$:

$$Tv_k = \sum_{j=1}^{n} A_{j,k}w_j.$$

$$\mathcal{M}\big(I, (v_1, \ldots, v_n), (v_1, \ldots, v_n)\big) = I$$

Matrix of the Identity with Respect to Two Bases

The matrix of the product of operators

$$\mathcal{M}(ST, (u_1, \ldots, u_n), (w_1, \ldots, w_n)) =$$
$$\mathcal{M}(S, (v_1, \ldots, v_n), (w_1, \ldots, w_n))\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Matrix of the identity with respect to two bases

Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then

$$\mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$$

and

$$\mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n))$$

are inverses of each other.

Proof In the result above, replace w_j with u_j, and replace S and T with I, getting

$$I = \mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n)) \cdot \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Now interchange the roles of the u's and v's, getting the product in the other order. ∎

Change of Basis Formula

The matrix of the product of operators

$$\mathcal{M}(ST, (u_1, \ldots, u_n), (w_1, \ldots, w_m)) = \mathcal{M}(S, (v_1, \ldots, v_n), (w_1, \ldots, w_m))\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Change of basis formula

Suppose $T \in \mathcal{L}(V)$. Let $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ be bases of V. Let

$$A = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Then

$$\mathcal{M}(T, (u_1, \ldots, u_n)) = A^{-1}\mathcal{M}(T, (v_1, \ldots, v_n))A.$$

Proof In the result above, replace w_j with u_j and replace S with I, getting

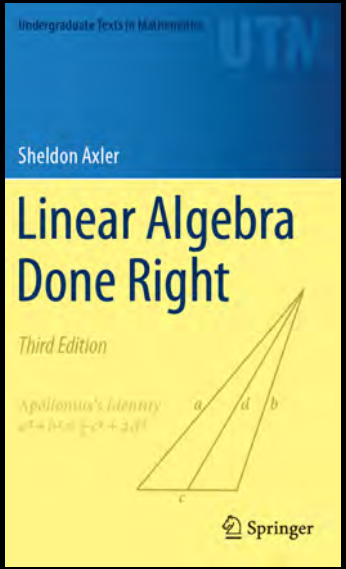
$$\mathcal{M}(T, (u_1, \ldots, u_n)) = A^{-1}\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Now in result above replace w_j with v_j. Also replace T with I and replace S with T, getting

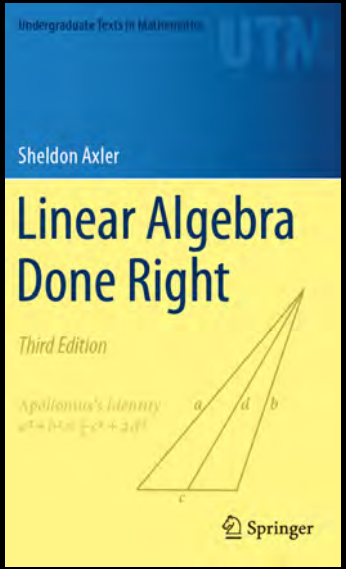
$$\mathcal{M}(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n)) = \mathcal{M}(T, (v_1, \ldots, v_n))A.$$

Substitute, getting the desired result. ∎

Linear Algebra Done Right, by Sheldon Axler



Trace of an Operator and of a Matrix



Review of Multiplicity

Definition: *Multiplicity*

Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. Then

multiplicity of $\lambda = \dim G(\lambda, T)$
$= \dim \operatorname{null}(T - \lambda I)^{\dim V}$.

Sum of multiplicities

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then the sum of the multiplicities of all the eigenvalues of T equals $\dim V$.

Definition: *Repeated accord to multiplicity*

Suppose $T \in \mathcal{L}(V)$. The phrase '*with each eigenvalue repeated according to its multiplicity* means that if $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T with multiplicities $d_1, \ldots, d_m$, then we create a list with λ_1 listed d_1 times, $\ldots$, λ_m listed d_m times. If $\mathbf{F} = \mathbf{C}$, then this list has length $\dim V$.

Trace of an Operator

Definition: *trace of an operator*

Suppose $T \in \mathcal{L}(V)$.

- If $\mathbf{F} = \mathbf{C}$, then the *trace* of T is the sum of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbf{F} = \mathbf{R}$, then the *trace* of T is the sum of the eigenvalues of $T_{\mathbf{C}}$, with each eigenvalue repeated according to its multiplicity.

The trace of T is denoted by $\operatorname{trace} T$.

Example: Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is the operator whose matrix is

$$\begin{pmatrix} 3 & -1 & -2 \\ 3 & 2 & -3 \\ 1 & 2 & 0 \end{pmatrix}.$$

Then the eigenvalues of T are 1, $2 + 3i$, and $2 - 3i$, each with multiplicity 1, as you can verify. Thus

$$\begin{aligned} \operatorname{trace} T &= 1 + (2 + 3i) + (2 - 3i) \\ &= 5. \end{aligned}$$

Trace and Characteristic Polynomial

Trace and characteristic polynomial

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $\operatorname{trace} T$ equals the negative of the coefficient of z^{n-1} in the characteristic polynomial of T.

Proof Suppose $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with each eigenvalue repeated according to its multiplicity. Then the characteristic polynomial of T equals

$$(z - \lambda_1) \cdots (z - \lambda_n).$$

Expand the polynomial above to write the characteristic polynomial of T in the form

$$z^n - (\lambda_1 + \cdots + \lambda_n) z^{n-1} + \cdots + (-1)^n (\lambda_1 \cdots \lambda_n).$$

The expression above gives the desired result. ∎

Trace of a Matrix

Suppose $T \in \mathcal{L}(V)$, $\mathbf{F} = \mathbf{C}$, and we choose a basis of V corresponding to the Decomposition Theorem. Then $\operatorname{trace} T$ equals the sum of the diagonal entries of that matrix.

Definition: *trace of a matrix*

The *trace* of a square matrix A, denoted $\operatorname{trace} A$, is defined to be the sum of the diagonal entries of A.

Example: Suppose

$$A = \begin{pmatrix} 3 & -1 & -2 \\ 3 & 2 & -3 \\ 1 & 2 & 0 \end{pmatrix}.$$

Then

$$\begin{aligned} \operatorname{trace} A &= 3 + 2 + 0 \\ &= 5. \end{aligned}$$

Trace of AB equals trace of BA

If A and B are square matrices of the same size, then

$$\operatorname{trace}(AB) = \operatorname{trace}(BA).$$

Trace Does Not Depend on Basis

Trace of matrix of operator does not depend on basis

Let $T \in \mathcal{L}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then

$$\operatorname{trace} \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) = \operatorname{trace} \mathcal{M}\big(T, (v_1, \ldots, v_n)\big).$$

Proof Let $A = \mathcal{M}\big(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)\big)$.
Then

$$\begin{aligned}
\operatorname{trace} \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) &= \operatorname{trace}\Big(A^{-1} \mathcal{M}\big(T, (v_1, \ldots, v_n)\big) A\Big) \\
&= \operatorname{trace}\Big(\mathcal{M}\big(T, (v_1, \ldots, v_n)\big) A A^{-1}\Big) \\
&= \operatorname{trace} \mathcal{M}\big(T, (v_1, \ldots, v_n)\big).
\end{aligned}$$

The last equality completes the proof. ∎

Trace of an Operator Equals Trace of Its Matrix

Trace of an operator equals trace of its matrix

Suppose $T \in \mathcal{L}(V)$. Then

$$\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T).$$

Proof To show that

$$\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T)$$

for every basis of V, we need only show that the equation above holds for some basis of V.

As we have already discussed, if V is a complex vector space, then choosing the basis given by the Decomposition Theorem gives the desired result.

If V is a real vector space, then applying the complex case to the complexification $T_{\mathbf{C}}$ (which is used to define $\operatorname{trace} T$) gives the desired result. ∎

Trace in Additive

Trace is additive

Suppose $S, T \in \mathcal{L}(V)$. Then

$$\operatorname{trace}(S+T) = \operatorname{trace} S + \operatorname{trace} T.$$

Proof Choose a basis of V. Then

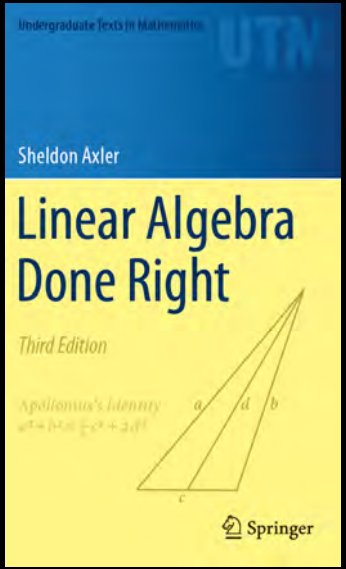
$$\begin{aligned}\operatorname{trace}(S+T) &= \operatorname{trace} \mathcal{M}(S+T)\\ &= \operatorname{trace}\big(\mathcal{M}(S)+\mathcal{M}(T)\big)\\ &= \operatorname{trace} \mathcal{M}(S) + \operatorname{trace} \mathcal{M}(T)\\ &= \operatorname{trace} S + \operatorname{trace} T,\end{aligned}$$

completing the proof. ∎

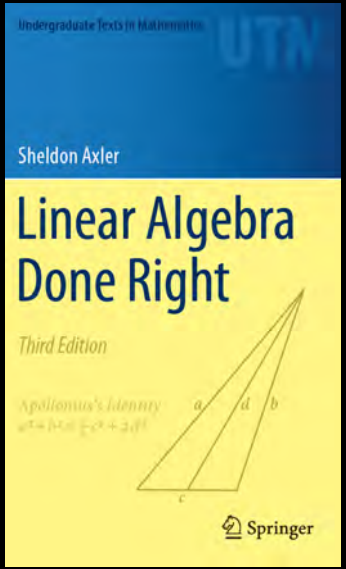
Exercises

- Suppose $P \in \mathcal{L}(V)$ satisfies $P^2 = P$. Prove that
$$\operatorname{trace} P = \dim \operatorname{range} P.$$
- Suppose V is an inner product space with orthonormal basis $e_1, \ldots, e_n$ and $T \in \mathcal{L}(V)$. Prove that
$$\operatorname{trace}(T^*T) = \|Te_1\|^2 + \cdots + \|Te_n\|^2.$$
Conclude that the right side of the equation above is independent of the orthonormal basis $e_1, \ldots, e_n$.
- Suppose V is an inner product space. Prove that
$$\langle S, T \rangle = \operatorname{trace}(ST^*)$$
defines an inner product on $\mathcal{L}(V)$.
- Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, repeated according to multiplicity. Suppose
$$\begin{pmatrix} A_{1,1} & \ldots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \ldots & A_{n,n} \end{pmatrix}$$
is the matrix of T with respect to some orthonormal basis of V. Prove that
$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \leq \sum_{k=1}^{n} \sum_{j=1}^{n} |A_{j,k}|^2.$$

Linear Algebra Done Right, by Sheldon Axler



Determinant of an Operator and of a Matrix



Notation

- $\mathbf{F}$ denotes either $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

Definition of Determinant of an Operator

Definition: *determinant of an operator,* $\det T$

Suppose $T \in \mathcal{L}(V)$.

- If $\mathbf{F} = \mathbf{C}$, then the *determinant* of T is the product of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbf{F} = \mathbf{R}$, then the *determinant* of T is the product of the eigenvalues of $T_{\mathbf{C}}$, with each eigenvalue repeated according to its multiplicity.

The determinant of T is denoted by $\det T$.

If $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with multiplicities $d_1, \ldots, d_m$, then the definition implies

$$\det T = \lambda_1^{d_1} \cdots \lambda_m^{d_m}.$$

Or if you prefer to list the eigenvalues with each repeated according to its multiplicity, then the eigenvalues could be denoted $\lambda_1, \ldots, \lambda_n$ (where the index n equals $\dim V$) and the definition implies

$$\det T = \lambda_1 \cdots \lambda_n.$$

Example

Definition: *determinant of an operator,* $\det T$

Suppose $T \in \mathcal{L}(V)$.

- If $\mathbf{F} = \mathbf{C}$, then the *determinant* of T is the product of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbf{F} = \mathbf{R}$, then the *determinant* of T is the product of the eigenvalues of $T_{\mathbf{C}}$, with each eigenvalue repeated according to its multiplicity.

The determinant of T is denoted by $\det T$.

Example: $\det I = 1$

Example: Suppose $T \in \mathcal{L}(\mathbf{F}^3)$ is the operator whose matrix is

$$\begin{pmatrix} 3 & -1 & -2 \\ 3 & 2 & -3 \\ 1 & 2 & 0 \end{pmatrix}.$$

Then the eigenvalues of T (or $T_{\mathbf{C}}$) are 1, $2 + 3i$, and $2 - 3i$, each with multiplicity 1.
Thus

$$\begin{aligned} \det T &= 1 \cdot (2 + 3i) \cdot (2 - 3i) \\ &= 13. \end{aligned}$$

Determinant and characteristic polynomial

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $\det T$ equals $(-1)^n$ times the constant term of the characteristic polynomial of T.

Proof Suppose $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with each eigenvalue repeated according to its multiplicity. Then the characteristic polynomial of T equals

$$(z - \lambda_1) \cdots (z - \lambda_n).$$

Expand the polynomial above to write the characteristic polynomial of T in the form

$$z^n - (\lambda_1 + \cdots + \lambda_n) z^{n-1} + \cdots + (-1)^n (\lambda_1 \cdots \lambda_n).$$

The expression above gives the desired result. ∎

Invertible is Equivalent to Nonzero Determinant

T *invertible* $\iff \det T \neq 0$

An operator on V is invertible if and only if its determinant is nonzero.

Proof First suppose V is a complex vector space and $T \in \mathcal{L}(V)$.

The operator T is invertible if and only if 0 is not an eigenvalue of T.

Clearly this happens if and only if the product of the eigenvalues of T is not 0. Thus T is invertible if and only if $\det T \neq 0$, as desired.

Now consider the case where V is a real vector space and $T \in \mathcal{L}(V)$.

Again, T is invertible if and only if 0 is not an eigenvalue of T, which happens if and only if 0 is not an eigenvalue of $T_{\mathbf{C}}$ (because $T_{\mathbf{C}}$ and T have the same real eigenvalues).

Thus again we see that T is invertible if and only if $\det T \neq 0$. ∎

Characteristic Polynomial of T Equals $\det(zI - T)$

Characteristic polynomial of T $= \det(zI - T)$

Suppose $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T equals $\det(zI - T)$.

Proof First suppose $\mathbf{F} = \mathbf{C}$. If $\lambda, z \in \mathbf{C}$, then

$$-(T - \lambda I) = (zI - T) - (z - \lambda)I.$$

Thus λ is an eigenvalue of T if and only if $z - \lambda$ is an eigenvalue of $zI - T$. Raising both sides of this equation to the $\dim V$ power and then taking null spaces of both sides shows that the multiplicity of λ as an eigenvalue of T equals the multiplicity of $z - \lambda$ as an eigenvalue of $zI - T$.

Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of T, repeated according to multiplicity.

Thus for $z \in \mathbf{C}$, the eigenvalues of $zI - T$ are $z - \lambda_1, \ldots, z - \lambda_n$, repeated according to multiplicity. The determinant of $zI - T$ is the product of these eigenvalues. Thus

$$\det(zI - T) = (z - \lambda_1) \cdots (z - \lambda_n).$$

The right side of the equation above is, by definition, the characteristic polynomial of T, completing the proof when $\mathbf{F} = \mathbf{C}$.

Now suppose $\mathbf{F} = \mathbf{R}$. Applying the complex case to $T_{\mathbf{C}}$ gives the desired result. ∎

Permutations

Definition: *permutation,* $\operatorname{perm} n$

- A *permutation* of $(1, \ldots, n)$ is a list $(m_1, \ldots, m_n)$ that contains each of the numbers $1, \ldots, n$ exactly once.
- The set of all permutations of $(1, \ldots, n)$ is denoted $\operatorname{perm} n$.

Interchanging two entries in a permutation

Interchanging two entries in a permutation multiplies the sign of the permutation by -1.

Definition: *sign of a permutation*

The *sign* of a permutation $(m_1, \ldots, m_n)$ is defined to be 1 if the number of pairs of integers (j, k) with $1 \leq j < k \leq n$ such that j appears after k in the list $(m_1, \ldots, m_n)$ is even and -1 if the number of such pairs is odd.

In other words, the sign of a permutation equals 1 if the natural order has been changed an even number of times and equals -1 if the natural order has been changed an odd number of times.

Definition of Determinant of a Matrix

Definition: *determinant of a matrix,* $\det A$

Suppose A is an n-by-n matrix

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix}.$$

The *determinant* of A, denoted $\det A$, is defined by

$$\det A = \sum_{(m_1,\dots,m_n)\in \operatorname{perm} n} \big(\operatorname{sign}(m_1,\dots,m_n)\big) A_{m_1,1}\cdots A_{m_n,n}.$$

Example: Suppose

$$A = \begin{pmatrix} A_{1,1} & & * \\ & \ddots & \\ 0 & & A_{n,n} \end{pmatrix}.$$

The permutation $(1,2,\dots,n)$ has sign 1 and thus contributes a term of $A_{1,1}\cdots A_{n,n}$ to the sum defining $\det A$. All other permutations $(m_1,\dots,m_n) \in \operatorname{perm} n$ contain at least one $m_j > j$, which means that $A_{m_j,j} = 0$. Thus all the other terms in the sum make no contribution. Hence

$$\det A = A_{1,1}\cdots A_{n,n}.$$

Determinant is Multiplicative

Determinant is multiplicative

Suppose A and B are square matrices of the same size. Then

$$\det(AB) = \det(BA) = (\det A)(\det B).$$

Determinant of matrix of operator does not depend on basis

Let $T \in \mathcal{L}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then

$$\det \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) = \det \mathcal{M}\big(T, (v_1, \ldots, v_n)\big).$$

Proof Let

$$A = \mathcal{M}\big(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)\big).$$

Then

$$\begin{aligned}
&\det \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) \\
&\quad = \det\Big(A^{-1}\mathcal{M}\big(T, (v_1, \ldots, v_n)\big)A\Big) \\
&\quad = \det\Big(\mathcal{M}\big(T, (v_1, \ldots, v_n)\big)AA^{-1}\Big) \\
&\quad = \det \mathcal{M}\big(T, (v_1, \ldots, v_n)\big).
\end{aligned}$$

The last equality completes the proof. ∎

Determinant of Operator Equals Determinant of Its Matrix

Determinant of operator equals determinant of its matrix

Suppose $T \in \mathcal{L}(V)$. Then $\det T = \det \mathcal{M}(T)$.

Proof To show that $\det T = \det \mathcal{M}(T)$ for every basis of V, we need only show that the result holds for some basis of V.

If $\mathbf{F} = \mathbf{C}$, then choosing the basis given by the Decomposition Theorem gives the desired result.

If $\mathbf{F} = \mathbf{R}$, then applying the complex case to the complexification $T_{\mathbf{C}}$ gives the desired result. ∎

Determinant is multiplicative

Suppose $S, T \in \mathcal{L}(V)$. Then

$$\det(ST) = \det(TS) = (\det S)(\det T).$$

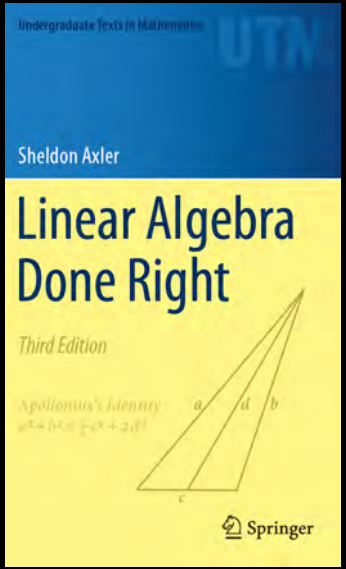
Proof Choose a basis of V. Then

$$\begin{aligned}\det(ST) &= \det \mathcal{M}(ST)\\ &= \det\big(\mathcal{M}(S)\mathcal{M}(T)\big)\\ &= \big(\det \mathcal{M}(S)\big)\big(\det \mathcal{M}(T)\big)\\ &= (\det S)(\det T).\end{aligned}$$

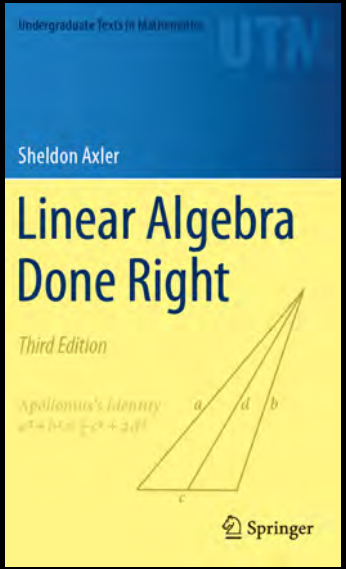
Interchanging S and T gives

$$\det(TS) = (\det T)(\det S)$$

Linear Algebra Done Right, by Sheldon Axler



Change of Variables in Integration in $\mathbf{R}^n$



If S is an Isometry, Then $|\det S| = 1$

Definition: *isometry*

Suppose V is an inner product space. An operator $S \in \mathcal{L}(V)$ is called an *isometry* if

$$\|Sv\| = \|v\|$$

for all $v \in V$.

Isometries have determinant with absolute value 1

Suppose V is an inner product space and $S \in \mathcal{L}(V)$ is an isometry. Then

$$|\det S| = 1.$$

Proof First suppose $\mathbf{F} = \mathbf{C}$.
Then all the eigenvalues of S have absolute value 1. Thus the product of the eigenvalues of S, counting multiplicity, has absolute value one. In other words,

$$|\det S| = 1,$$

as desired.

Now suppose $\mathbf{F} = \mathbf{R}$.
With respect to the inner product on the complexification $V_{\mathbf{C}}$, it is easy to see that $S_{\mathbf{C}}$ is an isometry on $V_{\mathbf{C}}$. Thus $|\det S_{\mathbf{C}}| = 1$. Hence $|\det S| = 1$, completing the proof. ∎

Definition: *isometry*

Suppose V is an inner product space. An operator $S \in \mathcal{L}(V)$ is called an *isometry* if

$$\|Sv\| = \|v\|$$

for all $v \in V$.

Isometries have determinant with absolute value 1

Suppose V is an inner product space and $S \in \mathcal{L}(V)$ is an isometry. Then

$$|\det S| = 1.$$

$|\det T| = \det \sqrt{T^*T}$

Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Then

$$|\det T| = \det \sqrt{T^*T}.$$

Proof By the Polar Decomposition, there is an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Thus

$$\begin{aligned} |\det T| &= |\det S| \det \sqrt{T^*T} \\ &= \det \sqrt{T^*T}, \end{aligned}$$

as desired. ∎

Volume

Definition: *box*

A *box* in $\mathbf{R}^n$ is a set of the form

$$\{(y_1, \ldots, y_n) \in \mathbf{R}^n : x_j < y_j < x_j + r_j \text{ for } j = 1, \ldots, n\},$$

where each $r_j > 0$ and $(x_1, \ldots, x_n) \in \mathbf{R}^n$. The numbers $r_1, \ldots, r_n$ are called the *side lengths* of the box.

Definition: *volume of a box*

The *volume* of a box B in $\mathbf{R}^n$ with side lengths $r_1, \ldots, r_n$ is defined to be $r_1 \cdots r_n$ and is denoted by $\text{volume}\, B$.

Definition: *volume*

Suppose $\Omega \subset \mathbf{R}^n$. The *volume* of Ω, denoted $\text{volume}\, \Omega$, is defined to be the infimum of

$$\text{volume}\, B_1 + \text{volume}\, B_2 + \cdots,$$

where the infimum is taken over all sequences $B_1, B_2, \ldots$ of boxes in $\mathbf{R}^n$ whose union contains Ω.

Notation: $T(\Omega)$

For $T \colon \Omega \to \mathbf{R}^n$, define $T(\Omega)$ by

$$T(\Omega) = \{Tx : x \in \Omega\}.$$

Positive Operators Change Volume by Determinant

Positive operators change volume by factor of determinant

Suppose $T \in \mathcal{L}(\mathbf{R}^n)$ is a positive operator and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume } T(\Omega) = (\det T)(\text{volume } \Omega).$$

Proof First consider the special case where $\lambda_1, \ldots, \lambda_n$ are positive numbers and

$$T(x_1, \ldots, x_n) = (\lambda_1 x_1, \ldots, \lambda_n x_n).$$

This operator stretches the j^{th} standard basis vector by a factor of λ_j. If B is a box in $\mathbf{R}^n$ with side lengths $r_1, \ldots, r_n$, then $T(B)$ is a box in $\mathbf{R}^n$ with side lengths $\lambda_1 r_1, \ldots, \lambda_n r_n$.

Note that $\det T = \lambda_1 \cdots \lambda_n$. Thus

$$\text{volume } T(B) = (\det T)(\text{volume } B)$$

for every box B in $\mathbf{R}^n$. Thus

$$\text{volume } T(\Omega) = (\det T)(\text{volume } \Omega).$$

Now consider an arbitrary positive operator $T \in \mathcal{L}(\mathbf{R}^n)$. By the Real Spectral Theorem, there exist an orthonormal basis $e_1, \ldots, e_n$ of $\mathbf{R}^n$ and nonnegative numbers $\lambda_1, \ldots, \lambda_n$ such that $Te_j = \lambda_j e_j$ for $j = 1, \ldots, n$. This operator stretches the j^{th} basis vector in an orthonormal basis by a factor of λ_j. Thus T multiplies volume by a factor of $\lambda_1 \cdots \lambda_n$, which again equals $\det T$. ∎

Isometries Do Not Change Volume

Isometries do not change volume

Suppose $S \in \mathcal{L}(\mathbf{R}^n)$ is an isometry and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume } S(\Omega) = \text{volume } \Omega.$$

Proof 1 For $x, y \in \mathbf{R}^n$, we have

$$\begin{aligned} \|Sx - Sy\| &= \|S(x - y)\| \\ &= \|x - y\|. \end{aligned}$$

In other words, S does not change the distance between points.
Thus S does not change volume. ∎

Proof 2 Consider our complete description of isometries on real inner product spaces.

The isometry S can be decomposed into pieces, each of which is the identity on some subspace (which clearly does not change volume) or multiplication by -1 on some subspace (which again clearly does not change volume) or a rotation on a 2-dimensional subspace (which again does not change volume). ∎

T Changes Volume by Factor of $|\det T|$

$$\text{volume}\, T(\Omega) = |\det T|(\text{volume}\, \Omega)$$

Suppose $T \in \mathcal{L}(\mathbf{R}^n)$ and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume}\, T(\Omega) = |\det T|(\text{volume}\, \Omega).$$

Proof By the Polar Decomposition, there is an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Thus $T(\Omega) = S\big(\sqrt{T^*T}(\Omega)\big)$. Hence

$$\begin{aligned}
\text{volume}\, T(\Omega) &= \text{volume}\, S\big(\sqrt{T^*T}(\Omega)\big) \\
&= \text{volume}\, \sqrt{T^*T}(\Omega) \\
&= (\det \sqrt{T^*T})(\text{volume}\, \Omega) \\
&= |\det T|(\text{volume}\, \Omega),
\end{aligned}$$

as desired. ∎

Integration

Definition: *integral,* $\int_\Omega f$

Suppose $\Omega \subset \mathbf{R}^n$ and $f\colon \Omega \to \mathbf{R}$. The *integral* of f over Ω, denoted

$$\int_\Omega f \quad \text{or} \quad \int_\Omega f(x)\,dx,$$

is defined by breaking Ω into pieces small enough that f is almost constant on each piece. On each piece, multiply the (almost constant) value of f by the volume of the piece, then add up these numbers for all the pieces, getting an approximation to the integral that becomes more accurate as Ω is divided into finer pieces.

Actually, Ω in the definition needs to be a reasonable set (for example, open or measurable) and f needs to be a reasonable function (for example, continuous or measurable).

Notice that the x in $\int_\Omega f(x)\,dx$ is a dummy variable and could be replaced with any other symbol.

Differentiation

Definition: *differentiable, derivative,* $\sigma'(x)$

Suppose $\Omega \subset \mathbf{R}^n$ and $\sigma\colon \Omega \to \mathbf{R}^n$. The function σ is called *differentiable* at a point $x \in \Omega$ if there exists an operator $\sigma'(x) \in \mathcal{L}(\mathbf{R}^n)$ such that

$$\lim_{y \to 0} \frac{\|\sigma(x+y) - \sigma(x) - \big(\sigma'(x)\big)y\|}{\|y\|} = 0.$$

The idea of the derivative is that for x fixed and $\|y\|$ small,

$$\sigma(x+y) \approx \sigma(x) + \big(\sigma'(x)\big)y.$$

We can write

$$\sigma(x) = \big(\sigma_1(x), \ldots, \sigma_n(x)\big),$$

where each σ_j is a function from Ω to $\mathbf{R}$. The partial derivative of σ_j with respect to the k^{th} coordinate is denoted $D_k\sigma_j$. Evaluating this partial derivative at a point $x \in \Omega$ gives $D_k\sigma_j(x)$. If σ is differentiable at x, then the matrix of $\sigma'(x)$ with respect to the standard basis of $\mathbf{R}^n$ contains $D_k\sigma_j(x)$ in row j, column k. In other words,

$$\mathcal{M}\big(\sigma'(x)\big) = \begin{pmatrix} D_1\sigma_1(x) & \ldots & D_n\sigma_1(x) \\ \vdots & & \vdots \\ D_1\sigma_n(x) & \ldots & D_n\sigma_n(x) \end{pmatrix}.$$

Change of Variables in an Integral

Change of variables in an integral

Suppose Ω is an open subset of $\mathbf{R}^n$ and $\sigma\colon \Omega \to \mathbf{R}^n$ is differentiable at every point of Ω. If f is a real-valued function defined on $\sigma(\Omega)$, then

$$\int_{\sigma(\Omega)} f(w)\,dw = \int_{\Omega} f\big(\sigma(x)\big)\,|\det \sigma'(x)|\,dx.$$

Proof Let $x \in \Omega$ and let Γ be a small subset of Ω containing x such that f is approximately equal to the constant $f\big(\sigma(x)\big)$ on the set $\sigma(\Gamma)$.

The approximation

$$\sigma(x+y) \approx \sigma(x) + \big(\sigma'(x)\big)y$$

shows that

$$\operatorname{volume} \sigma(\Gamma) \approx \operatorname{volume}\big[\big(\sigma'(x)\big)(\Gamma)\big].$$

Hence

$$\operatorname{volume} \sigma(\Gamma) \approx |\det \sigma'(x)|(\operatorname{volume} \Gamma).$$

Let $w = \sigma(x)$. Multiply the left side of the equation above by $f(w)$ and the right side by $f\big(\sigma(x)\big)$, getting

$$f(w) \operatorname{volume} \sigma(\Gamma) \approx f\big(\sigma(x)\big)|\det \sigma'(x)|(\operatorname{volume} \Gamma).$$

Now break Ω into many small pieces and add the corresponding versions of the equation above, getting the desired result. ∎

Ada Lovelace (1815 – 1852)

Linear Algebra Done Right, by Sheldon Axler