MATH 73: Linear Algebra

Connor Neely, Spring 2023

1 Vectors **2**

1.1 The Geometry and Algebra of Vectors . . . 2

1.2 The Dot Product . . . 3

2 Systems of Linear Equations **5**

2.1 Systems of Linear Equations . . . 5

2.2 Solving Linear Systems . . . 5

2.3 Spanning Sets and Linear Independence . . . 7

3 Matrices **9**

3.1 Matrix Operations . . . 9

3.2 Matrix Algebra . . . 10

3.3 Matrix Inverses . . . 11

3.4 Subspaces, Basis, Dimension, and Rank . . . 13

4 Eigenvalues and Eigenvectors **17**

4.1 Determinants . . . 17

4.2 Eigenvalues and Eigenvectors . . . 18

4.3 Similarity and Diagonalization . . . 20

5 Orthogonality **23**

5.1 Orthogonality in $\mathbb{R}^n$ . . . 23

5.2 Orthogonal Complements and Projections . . . 25

5.3 The Gram-Schmidt Process and the QR Factorization . . . 26

6 Vector Spaces **27**

6.1 Vector Spaces and Subspaces . . . 27

6.2 Linear Independence, Basis, and Dimension . . . 28

6.3 Change of Basis . . . 30

6.4 Linear Transformations . . . 31

6.5 Kernel and Range . . . 32

6.6 The Matrix of a Linear Transformation . . . 34

7 Distance and Approximation **36**

7.1 Inner Product Spaces . . . 36

7.2 Least Squares Approximation . . . 37

7.3 The Singular Value Decomposition . . . 38

A The Fundamental Theorem of Linear Algebra **40**

* Adapted from David Poole, *Linear Algebra: A Modern Introduction* (2015) and SP23 lectures.

1 Vectors

1.1 The Geometry and Algebra of Vectors

One of the fundamental objects of linear algebra is the vector. Vectors can take a variety of forms; for now, we will consider the simplest type, defined below.

Definition: Vectors in $\mathbb{R}^n$

$\mathbb{R}^n$ is the set of all ordered n-tuples of real numbers. These tuples are called vectors, and they are written in the form

$$\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

A vector all of whose components are zero is called the zero vector $\mathbf{0}$.

Rather than considering vectors in two- or three-dimensional space, as was the case in previous courses, we now think of vectors as objects in n-dimensional space. We'll define the basic operations in the same way.

Definition: Vector operations

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then addition and scalar multiplication are defined componentwise:

$$\mathbf{u}+\mathbf{v} = \begin{bmatrix} u_1+v_1 \\ \vdots \\ u_n+v_n \end{bmatrix}, \quad c\mathbf{v} = \begin{bmatrix} cv_1 \\ \vdots \\ cv_n \end{bmatrix}.$$

Since we've defined these two operations in familiar ways, it might not be surprising that many of their properties are also familiar. Their proofs are straightforward applications of the definition, so we omit them.

Theorem 1.1: Algebraic properties of vectors

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and let c and d be scalars. Then

(a) $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$.

(b) $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$.

(c) $\mathbf{u}+\mathbf{0} = \mathbf{u}$.

(d) $\mathbf{u}+(-\mathbf{u}) = \mathbf{0}$.

(e) $c(\mathbf{u}+\mathbf{v}) + c\mathbf{u} + c\mathbf{v}$.

(f) $(c+d)\mathbf{u} = c\mathbf{u}+d\mathbf{u}$.

(g) $c(d\mathbf{u}) = (cd)\mathbf{u}$.

(h) $1\mathbf{u} = \mathbf{u}$.

Familiar properties aside, there's one very important thing we can do with these operations. We can combine vectors in a very particular way to create a new vector—this leads us to some powerful new results, which we'll investigate later.

Definition: Linear combination

A vector $\mathbf{v}$ is a linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ if there are scalars $c_1, c_2, \dots, c_k$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k.$$

The scalars $c_1, c_2, \dots, c_k$ are called the coefficients of the linear combination.

1.2 The Dot Product

We now define a new vector operation. This time, we'll take a pair of two vectors and associate it with a scalar in the way defined below.

Definition: Dot product

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. The dot product of $\mathbf{u}$ and $\mathbf{v}$ is defined by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

The dot product turns out to have some very natural properties, many of which are reminiscent of the multiplication of real numbers. We once again omit the proofs.

Theorem 1.2: Algebraic properties of the dot product

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$.

(c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$.

(d) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

The real importance of the dot product is that it us to define things like length in higher dimensions.

Definition: Length

The length (or norm) of a vector $\mathbf{v}$ in $\mathbb{R}^n$ is the nonnegative scalar $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

Vectors that have a length of 1 are, in general, nice to work with. These vectors are especially nice when they describe the typical coordinate axes.

Definition: Unit vector

A vector of length 1 is called a unit vector. The standard unit vectors in $\mathbb{R}^n$ are denoted by $\mathbf{e}_1, \ldots, \mathbf{e}_n$, where $\mathbf{e}_k$ has a one in its kth component and zeros elsewhere.

Given how natural the notion of length is, a couple of its basic properties might be expected.

Theorem 1.3: Properties of the norm

Let $\mathbf{v}$ be a vector in $\mathbb{R}^n$ and let c be a scalar. Then

(a) $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(b) $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$.

Now, we have a pair of perhaps less expected but still very important inequalities. These proofs are not trivial, so we'll detail them.

Theorem 1.4: Cauchy-Schwarz inequality

For all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

Proof. Consider the vector $\|\mathbf{u}\|\mathbf{v} - \|\mathbf{v}\|\mathbf{u}$. We have

$$0 \leq (\|\mathbf{u}\|\mathbf{v} - \|\mathbf{v}\|\mathbf{u}) \cdot (\|\mathbf{u}\|\mathbf{v} - \|\mathbf{v}\|\mathbf{u}) = 2\|\mathbf{u}\|^2\|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|(\mathbf{u} \cdot \mathbf{v});$$

we can rearrange to get $\mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\|\|\mathbf{v}\|$. We could use a similar argument with the addition of scaled vectors to show that $-\mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\|\|\mathbf{v}\|$, and thus $\|\mathbf{u} \cdot \mathbf{v}\| \leq \|\mathbf{u}\|\|\mathbf{v}\|$, as desired. $\square$

Theorem 1.5: Triangle inequality

For all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Proof. We'll expand the left-hand side using nonnegativity and the Cauchy-Schwarz inequality:

$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) \\ &= \mathbf{u}\cdot\mathbf{u} + 2(\mathbf{u}\cdot\mathbf{v}) + \mathbf{v}\cdot\mathbf{v} \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2, \end{aligned}$$

as desired. $\square$

Moving along, not only does the dot product allow us to define the length of a vector, but it also allows us to define such other geometric concepts as the distance or angle between two vectors.

Definition: Distance

The distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is defined by

$$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|.$$

Definition: Angle

For nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$,

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

Two vectors are perpendicular to one another if the angle between them is $\pi/2$. Given how we've defined this angle, we can use the dot product to determine whether two vectors are perpendicular to each other (and arrive at an n-dimensional analog for the Pythagorean theorem).

Definition: Orthogonality

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal (or perpendicular) to each other if $\mathbf{u}\cdot\mathbf{v} = 0$.

Theorem 1.6: Pythagorean theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if and only if

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Proof. We have $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2$ with $\mathbf{u}\cdot\mathbf{v} = 0$ by perpendicularity. $\square$

With all this, we can define an object that encapsulates what happens when one vector $\mathbf{v}$ casts a "shadow" onto another vector $\mathbf{v}$. This definition will come in handy later.

Definition: Projection onto a vector

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^2$ and $\mathbf{u} \neq \mathbf{0}$, then the projection of $\mathbf{v}$ onto $\mathbf{u}$ is the vector defined by

$$\operatorname{proj}_{\mathbf{u}} \mathbf{v} = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|}\right)\frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

2 Systems of Linear Equations

2.1 Systems of Linear Equations

We now turn our attention to systems of simultaneous linear equations. (Vectors will make a return later.) First, we need to be quite a bit more precise about what a linear equation is, what a system of such equation entails, and what it means to solve one.

Definition: Linear equation

A linear equation in the n variables $x_1, x_2, \ldots, x_n$ is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where the coefficients $a_1, a_2, \ldots, a_n$ and the constant term b are constants. A solution of this equation is a vector $\begin{bmatrix} s_1 & s_2 & \cdots & s_n \end{bmatrix}$ whose components satisfy the equation when we substitute $x_i = s_i$ for all $i = 1, 2, \ldots, n$.

Definition: System of linear equations

A system of linear equations (or a linear system) is a finite set of linear equations, each with the same variables. A solution of a linear system is a vector that is simultaneously a solution of each equation in the system; the solution set of a linear system is the set of all solutions of the system.

Definition: Consistent system

A linear system is consistent if it has at least one solution. A system with no solutions is inconsistent.

2.2 Solving Linear Systems

Our general technique for solving systems of linear equations will be to gradually simplify the system to the point where the solution is easy to find. We first give a condition that will guide this simplification.

Definition: Equivalent systems

Two linear systems are equivalent if they have the same solution set.

Now, when we're working with a linear system, all that really matters are its coefficients and constants. Extracting these numbers and arranging them in a grid makes solution a bit less cumbersome. It also allows us to analyze linear systems in a more sophisticated way, which we'll see later.

Definition: Matrix of a linear system

A linear system's coefficient matrix contains the coefficients of the variables. A system's augmented matrix consists of its coefficient matrix and an extra column containing the system's constant terms.

There's a couple of ways we can define the "simplest" form of a system's matrix, coefficient or augmented. The first of these, along with steps that we might take to get there based on our knowledge of linear systems, is given below.

Definition: Row echelon form

A matrix is in row echelon form (REF) if it satisfies the following properties.

- Any rows consisting entirely of zeros are at the bottom.
- In each nonzero row, the first nonzero entry (called the leading entry) is in a column to the left of any leading entries below it.

Definition: Elementary row operations

The following elementary row operations (EROs) can be performed on a matrix.

- Interchange two rows.
- Multiply a row by a nonzero constant.
- Add a multiple of a row to another row.

The process of applying elementary row operations to bring a matrix into REF is called row reduction.

These rules help us reason what it means for two matrices to represent equivalent systems.

Definition: Row equivalent matrices

Two matrices are row equivalent if there is a series of EROs that converts one into the other.

We don't need to find such a series every time we want to determine whether two matrices are row equivalent, though. Rather, we can exploit the reversibility of each ERO to intuit the following condition.

Theorem 2.1: Condition for row equivalence

Matrices A and B are row equivalent if and only if they can be reduced to the same row echelon form.

All this gives us a powerful way to solve linear systems. The algorithm is described below.

Definition: Gaussian elimination

The process of Gaussian elimination is as follows.

1. Write the augmented matrix of the linear system.
2. Use elementary row operations to reduce the augmented matrix to row echelon form.
3. Using back substitution, solve the equivalent system that corresponds to the row-reduced matrix.

In many cases, a linear system will have infinitely many solutions. When this happens, the system's solution set is defined in terms of a certain number of arbitrary parameters. The following association between leading entries and nonzero rows leads us naturally to a simple relationship between nonzero rows and free variables.

Definition: Leading and free variables

The leading variables of a linear system correspond to the leading entries of its augmented matrix. The other variables are the free variables.

Theorem 2.2: Number of free variables from REF

Let A be the coefficient matrix of a linear system with n variables. If the system is consistent, then

$$\#\text{ free variables} = n - (\#\text{ nonzero rows in REF}).$$

So far, we have been working only with a matrix's row echelon forms. A certain class of row echelon forms is particularly useful in our study of systems.

Definition: Reduced row echelon form

A matrix is in reduced row echelon form (RREF) if it satisfies the following properties.

- It is in row echelon form.
- The leading entry in each nonzero row is a 1.
- Each column containing a leading 1 has zeros everywhere else.

One advantage that comes with working with a reduced row echelon form is that it is unique. Not only that, but reducing a matrix to its RREF gives us an immediate solution to the corresponding linear system.

Definition: Gauss-Jordan elimination

The process of Gauss-Jordan elimination is the same as that of Gaussian elimination, but the matrix is converted into its reduced row echelon form rather than a row echelon form.

Given either of these simplified forms for a matrix, if the corresponding system takes a certain form, it is easy to tell whether or not a system will have infinitely many solutions. The following is an immediate consequence of the relationship between leading and free variables.

Definition: Homogeneous system of linear equations

A system of linear equations is homogeneous if the constant term in each equation is zero.

Theorem 2.3: Solutions of a homogeneous system

A homogeneous linear system of m equations in n variables has infinitely many solutions if $m < n$.

2.3 Spanning Sets and Linear Independence

The study of vectors frequently intersects with that of matrices. We'll begin our overview of the introductory definitions and results of this intersection with a condition for the consistency of a linear system, a trivial consequence of how we might translate systems of equations into vector equations.

Theorem 2.4: Condtion for consistency

A linear system with augmented matrix $\left[A \mid \mathbf{b}\right]$ is consistent if and only if $\mathbf{b}$ is a linear combination of the columns of A.

We'll continue with the idea of linear combinations, using it to define a couple of important concepts. The first of these gives a name to the vectors that a set of vectors can "reach" via a linear combination.

Definition: Span

If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a set of vectors in $\mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is called the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and is denoted by $\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ or $\mathrm{span}(S)$. If $\mathrm{span}(S) = \mathbb{R}^n$, then the S is called a spanning set for $\mathbb{R}^n$.

Building off of this, the next definition and the following theorem together characterize what it means for a vector to be entirely "separate" from another set of vectors.

Definition: Linear dependence

A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is linearly dependent if there are scalars $c_1, c_2, \ldots, c_k$, at least one of which is not zero, such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$. A set of vectors that is not linearly dependent is called linearly independent.

Theorem 2.5: Condition for linear dependence

Vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ in $\mathbb{R}^n$ are linearly dependent if and only if at least one of the vectors can be expressed as a linear combination of the others.

This immediately brings us to a couple of new connections between linear systems, matrices, and sets of vectors. The first theorem works with a matrix's column vectors, while the second works with row vectors.

Theorem 2.6: Condition for linearly independent columns

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be vectors in $\mathbb{R}^n$ and let A be the $n \times m$ matrix with these vectors as its columns. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly dependent if and only if the homogeneous linear system with augmented matrix $\left[A \mid \mathbf{0}\right]$ has a nontrivial solution.

Proof. The $\mathbf{v}_i$ are linearly dependent if and only if there are scalars c_i, not all zero, such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0}$. This is equivalent to saying that the vector whose components are the c_i is a solution of the system whose augmented matrix has columns $\mathbf{v}_i$ and $\mathbf{0}$. □

Theorem 2.7: Condition for linearly independent rows

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be vectors in $\mathbb{R}^n$ and let A be the $m \times n$ matrix with these vectors as its rows. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly dependent if and only if the row echelon form of A has less than m nonzero rows.

Proof. ($\Rightarrow$) Suppose the $\mathbf{v}_i$ are linearly dependent and, without loss of generality, that $\mathbf{v}_m$ can be written as a linear combination of the other $m-1$ vectors with coefficients c_i. Then the elementary row operations $R_m - c_1R_1, \ldots, R_m - c_{m-1}R_{m-1}$ applied to A will create a zero row in row m. Thus A has less than m nonzero rows.

($\Leftarrow$) If A has less than m nonzero rows, then there exists a sequence of row operations that creates a zero row. Thus $\mathbf{0}$ is a nontrivial linear combination of the $\mathbf{v}_i$, and the vectors are linearly dependent. □

Finally, we have a powerful sufficient (but not necessary!) condition for linear dependence.

Theorem 2.8: Condition for linear dependence

Any set of m vectors in $\mathbb{R}^n$ is linearly dependent if $m > n$.

Proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ be column vectors and let A be the matrix with these vectors as its columns. If A is a coefficient matrix, then the corresponding homogeneous linear system has more equations than unknowns and thus has infinitely many solutions. So the $\mathbf{v}_i$ are linearly dependent, as desired. □

3 Matrices

3.1 Matrix Operations

Now we turn our attention to matrices in their own right, removed from the context of vectors and linear systems. They come in several different forms, each of which has its own special properties.

Definition: Matrix

A matrix is a rectangular array of numbers called the entries, or elements, of the matrix. A matrix all of whose entries are zero is called the zero matrix O.

Definition: Special matrices

Let A be an $m \times n$ matrix.

- If $m = n$, then A is called a square matrix.
- A square matrix whose nondiagonal entries are all zero is called a diagonal matrix.
- A diagonal matrix all of whose diagonal entries are the same is called a scalar matrix.
- If the scalar on the diagonal is 1, the scalar matrix is called an identity matrix I_n.

Like vectors, matrices can be added and scaled.

Definition: Matrix operations

Let $A = [a_{ij}]$ and B_{ij} be $m \times n$ matrices, and let c be a scalar. Then addition and scalar multiplication are defined componentwise:

$$A + B = [a_{ij} + b_{ij}], \quad cA = [ca_{ij}].$$

Matrices can also be multiplied with each other, albeit in an at-first unintuitive way. Suppose we want to multiply $A\mathbf{x}$, where A is a matrix and $\mathbf{x}$ is a vector; we can think of A as a function that maps one vector to another. The first column of A tells us what happens to the first component $\mathbf{x}$, the second column determines the second component, and so on. If $\mathbf{a}_i$ is the ith column of A, then the resulting vector is the linear combination

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

Since a matrix is comprised of several vectors, this provides good motivation for the matrix product.

Definition: Matrix multiplication

If A is an $m \times n$ matrix and B is an $n \times r$ matrix, then the product $C = AB$ is an $m \times r$ matrix where each entry is given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

The matrix-column and row-matrix representations of the product are, respectively,

$$AB = \begin{bmatrix} A\mathbf{b}_1 & \cdots & A\mathbf{b}_r \end{bmatrix} \text{ and } AB = \begin{bmatrix} \mathbf{A}_1 B \\ \vdots \\ \mathbf{A}_m B \end{bmatrix}.$$

Sometimes it'll be useful to extract the ith row or column from a matrix. This is easy to accomplish using the left- or right-multiplication of the ith standard unit vector, respectively.

We'll define two more operations we can perform on matrices: exponentiation and transposition.

Definition: Matrix exponentiation and transposition

If A is a square matrices, then $A^n = AA\cdots A$, where there are n factors in this product.

The transpose of an $m \times n$ matrix A is the $n \times m$ matrix A^T obtained by interchanging the rows and columns of A. That is, the ith column of A^T is the ith row of A for all i, or $(A^T)_{ij} = A_{ji}$ for all i, j.

Visually, the transpose involves flipping a matrix over its main diagonal. This allows us to define another special, usually non-diagonal type of matrix that exhibits symmetry with respect to transposition.

Definition: Symmetric matrix

A square matrix A is symmetric if $A^T = A$, that is, if $A_{ij} = A_{ji}$ for all i and j.

3.2 Matrix Algebra

The definitions from the previous section lead naturally to some important properties for matrix algebra. We begin with the basic operations of addition and scalar multiplication, noting the similarities to their corresponding vector operations.

Theorem 3.1: Properties of matrix addition and scalar multiplication

Let A, B, and C be matrices of the same size and let c and d be scalars. Then

(a) $A + B = B + A$

(b) $(A + B) + C = A + (B + C)$

(c) $A + O = A$

(d) $A + (-A) = O$

(e) $c(A + B) = cA + cB$

(f) $(c + d)A = cA + dA$

(g) $c(dA) = (cd)A$

(h) $1A = A$

Now we'll look at matrix multiplication and transposition, noting that multiplication is not commutative!

Theorem 3.2: Properties of matrix multiplication

Let A, B, and C be matrices (such that all indicated operations are defined) and let k be a scalar. Then

(a) $A(BC) = (AB)C$

(b) $A(B + C) = AB + AC$

(c) $(A + B)C = AC + BC$

(d) $k(AB) = (kA)B = A(kB)$

(e) $I_m A = A = AI_n$

Theorem 3.3: Properties of matrix transposition

Let A and B be matrices (such that all indicated operations are defined) and let k be a scalar. Then

(a) $(A^T)^T = A$

(b) $(A + B)^T = A^T + B^T$

(c) $(kA)^T = k(A^T)$

(d) $(AB)^T = B^T A^T$

(e) $(A^r)^T = (A^T)^r$

We can use some of these to show what happens when we add or multiply a matrix by its own transpose.

Theorem 3.4: Adding or multiplying by a transpose

We have two relationships.

(a) If A is a square matrix, then $A + A^T$ is a symmetric matrix.

(b) For any matrix A, AA^T and A^TA are symmetric matrices.

3.3 Matrix Inverses

The additive inverse of a matrix can be found by negating each of its entries, but it might not be surprising that finding a multiplicative inverse is a bit more complicated! Thankfully, at least one property is familiar.

Definition: Inverse of a matrix

If A is a square matrix, an inverse of A is a matrix A^{-1} of the same size with the property that

$$AA^{-1} = I = A^{-1}A.$$

If such an A^{-1} exists, then A is called invertible.

Theorem 3.5: Uniqueness of the matirx inverse

If A is an invertible matrix, then its inverse is unique.

Proof. Suppose A' and A'' are both inverses of A. Then

$$A' = A'I = A'(AA'') = (A'A)A'' = IA'' = A'',$$

so $A' = A''$ and the inverse is unique. □

Since we can use matrix inverses to uniquely "undo" the effects of matrix multiplication, we can use them to uniquely solve certain equations involving matrices.

Theorem 3.6: Unique solution of a linear system

If A is an invertible $n \times n$ matrix, then the system of linear equations given by $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^n$.

In order to apply this to actual systems, we must be able to actually compute the matrix inverse. Thankfully, the 2×2 case is relatively straightforward and can be verified by a laborious computation.

Theorem 3.7: Inverse of a two-by-two matrix

The matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if $ad - bc \neq 0$, in which case

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

If $ad - bc = 0$, then A is not invertible.

To get a more general result, we'll need to look at some more easily-verified properties of inverses.

Theorem 3.8: Properties of matrix inverses

If A and B are invertible matrices of the same size, then the following are true:

(a) A^{-1} is invertible and $(A^{-1})^{-1} = A$.

(b) If c is a nonzero scalar, then cA is invertible and $(cA)^{-1} = (1/c)A^{-1}$.

(c) AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

(d) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

(e) A^n is invertible for all nonnegative integers n and $(A^n)^{-1} = (A^{-1})^n$.

As a side note, property (e) allows for some convenient notation for packaging inverses and exponents together.

Definition: Negative powers of a matrix

If a is an invertible matrix and n is a positive integer, then A^{-n} is defined by

$$A^{-n} = (A^{-1})^n = (A^n)^{-1}.$$

Our upcoming result regarding the inverse of a general square matrix will rely on the formalization of elementary row operations (EROs) using matrix multiplication. Specifically, each ERO corresponds to a matrix which, when multiplied, is equivalent to performing that ERO. (Of course, these matrices are invertible since the EROs themselves are easily reversible.)

Definition: Elementary matrix

An elementary matrix is any matrix that can be obtained by performing an elementary row operation on an identity matrix.

Theorem 3.9: Multiplication by an elementary matrix

Let E be the elementary matrix obtained by performing an elementary row operation on I_n. If the same elementary row operation is performed on an $n \times r$ matrix A, the result is the same as the product EA.

Theorem 3.10: Inverse of an elementary matrix

Each elementary matrix is invertible, and its inverse is an elementary matrix of the same type.

From here, we can begin formulating a theorem that will not only be useful for our immediate purposes, but which will also reoccur plenty throughout the course: the fundamental theorem of linear algebra (FTLA). It's essentially a list of useful conditions that are both necessary and sufficient for matrix invertibility.

Theorem 3.11: Additions to the FTLA

Let A be an $n \times n$ matrix and let $\mathbf{b}$ be a vector in $\mathbb{R}^n$. The following statements are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{b}$ has a unique solution.

(c) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(d) The reduced row echelon form of A is I_n.

(e) A is a product of elementary matrices.

For the full theorem, see Appendix A.

Proof. We'll use an implication chain. Note that already know (a) $\Rightarrow$ (b), and (c) follows trivially.

(c) $\Rightarrow$ (d). If the system $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = 0$, then we can reduce the corresponding augmented matrix to find $x_1 = \cdots = x_n = 0$. The same reduction A turns into the identity matrix.

(d) $\Rightarrow$ (e). If A reduces to I_n, then there are elementary matrices such that $E_k \cdots E_1 A = I_n$. Elementary operations are also invertible, so $A = E_k^{-1} \cdots E_1^{-1} I_n = E_k^{-1} \cdots E_1^{-1}$.

(e) $\Rightarrow$ (a). A is a product of elementary matrices, all of which are invertible. Thus A is invertible. $\square$

This is a powerful set of results! We'll use it in a a variety of cases to prove invertibility and, at present, to find a general algorithm for computing matrix inverses.

Theorem 3.12: Condition for matrix invertibility

Let A be a square matrix. If B is a square matrix such that either $AB = I$ or $BA = I$, then A is invertible and $B = A^{-1}$.

Proof. Consider the equation $A\mathbf{x} = \mathbf{0}$, and suppose $BA = I$. Then $BA\mathbf{x} = B\mathbf{0}$ and $I\mathbf{x} = \mathbf{0}$, meaning $A\mathbf{x} = \mathbf{0}$ has only the trivial solution and A is invertible. The $AB = I$ case is similar, and $B = A^{-1}$ follows from basic algebra. $\square$

Theorem 3.13: Computing the inverse of a matrix

Let A be a square matrix. If a series of elementary row operations reduces A to I, then the same series of elementary row operations transforms I into A^{-1}.

Proof. Suppose there exist elementary matrices such that $(E_k \cdots E_1)A = I$. By the previous theorem, the inverse is $A^{-1} = E_k \cdots E_1 = E_k \cdots E_1 I$. $\square$

So in order to invert a matrix, we simply reduce it to the identity matrix while simultaneously performing the same sequence of row operations on the identity.

3.4 Subspaces, Basis, Dimension, and Rank

Now, we'll introduce the ideas of spaces and subspaces, and we'll see how they relate to matrices.

Definition: Subspace

A subspace of $\mathbb{R}^n$ is any collection S of vectors in $\mathbb{R}^n$ such that

- the zero vector $\mathbf{0}$ is in S and
- S is closed under both addition and scalar multiplication.

A very simple example of a subspace is the set of all vectors that can be reached by just a few vectors. Those reachable by the rows or columns of a matrix are particularly relevant.

Theorem 3.14: Span as a subspace

Let S be a set of vectors in $\mathbb{R}^n$. Then $\mathrm{span}(S)$ is a subspace of $\mathbb{R}^n$.

Definition: Row space and column space

Let A be an $m \times n$ matrix.

- The row space $\mathrm{row}(A)$ of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A.
- The column space $\mathrm{col}(A)$ of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A.

Noting that a matrix's row space is invariant under elementary row operations, we get the following.

Theorem 3.15: Row spaces of row equivalent matrices

If B is row equivalent to A, then $\text{row}(B) = \text{row}(A)$.

As for the third space of interest, we also note that the solution set of a homogeneous linear system is a subspace. This can be shown via some simple algebra.

Theorem 3.16: Solution space of a homogeneous linear system

Let A be an $m \times n$ matrix and let S be the set of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. Then S is a subspace of $\mathbb{R}^n$.

Definition: Null space

Let A be an $m \times n$ matrix. The null space $\text{null}(A)$ of A is the subspace of A consisting of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$.

This notion is interesting: due to the properties of subspaces null spaces must either be empty, singleton, or infinite. That illuminates something important about systems!

Theorem 3.17: Number of solutions to a linear system

Let A be a real-entry matrix. The solution set to $A\mathbf{x} = \mathbf{b}$ is either empty, singleton, or infinite.

Proof. If (a) or (b) are true, then we are done. Otherwise, suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ are both solutions to the system $A\mathbf{x} = \mathbf{b}$. Their difference $\mathbf{x}_0$ solves the associated homogeneous system. Now, if the solutions are distinct, then $\mathbf{x}_0 \neq \mathbf{0}$ and $\text{null}(A)$ is infinite. Thus $\mathbf{x}_1 + c\mathbf{x}_0$ solves the system for all $c \in \mathbb{R}$. □

Let's move our discussion along with a new definition.

Definition: Basis of a subspace

A basis for a subspace S of $\mathbb{R}^n$ is a set of vectors in S that

- spans S and
- is linearly independent.

Bases are valuable because they provide the minimum information required to completely describe a subspace. Thankfully, there is a straightforward way to find a basis for each subspace we've discussed.

Theorem 3.18: Determining bases for row, column, and null spaces

Let R be the reduced row echelon form for A. To find bases for the row, column, and null spaces of a matrix A, we do the following.

(a) For $\text{row}(A)$, take the nonzero row vectors of R.

(b) For $\text{col}(A)$, take the column vectors of A that correspond to the pivot columns of R.

(c) For $\text{null}(A)$, solve the system $A\mathbf{x} = R\mathbf{x} = \mathbf{0}$ and write the solution as a linear combination of vectors times free variables.

No subspace (except for the trivial space $\{\mathbf{0}\}$) has a unique basis. However, every basis for a subspace is comprised by the same number of vectors, making it an intrinsic quantity for the subspace.

Theorem 3.19: Sizes of distinct bases

Let S be a subspace of $\mathbb{R}^n$. Then any two bases for S have the same number of vectors.

Proof. Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ and $\mathcal{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$ be bases for S such that $r < s$, and let

$$c_1\mathbf{v}_1 + \cdots + c_s\mathbf{v}_s = 0.$$

Each $\mathbf{v}_i$ can be written in terms of the $\mathbf{u}_j$; specifically, $\mathbf{v}_i = a_{i1}\mathbf{u}_1 + \cdots + a_{ir}\mathbf{u}_r$. Substituting this into the above and applying the linear independence of $\mathcal{B}$ gives a system of r equations in $c_1, \ldots, c_s$. Since there are more variables than equations, this system has infinitely many solutions and the above equation can be satisfied with nontrivial $c_1, \ldots, c_s$, contradicting the linear independence of $\mathcal{C}$.

Doing the same for $r > s$ yields a similar contradiction, meaning $r = s$. □

Definition: Dimension

If S is a subspace of $\mathbb{R}^n$ then its dimension, denoted $\dim(S)$, is the number of vectors in a basis for S.

It's clear that, for a matrix in row echelon form, the row space has the same dimension as the column space. But the dimensions of the row and column spaces are invariant under elementary row operations, meaning the dimensions are also the same for any other matrix!

Theorem 3.20: Row and column space dimensions

The row and column spaces of a matrix A have the same dimension.

Definition: Rank

The rank of a matrix A, denoted $\operatorname{rank}(A)$, is the common dimension of its row and column spaces.

Since the row and column spaces of a matrix just get swapped upon transposition, we get the following.

Theorem 3.21: Rank of a transpose

For any matrix A, $\operatorname{rank}(A^T) = \operatorname{rank}(A)$.

Now, the rank of a matrix is very closely related to its null space. In terms that we already know, the rank r of a matrix is given by the number of nonzero rows in a row echelon form; its null space has $n - r$ free variables, corresponding to the dimension of the null space. Adding these together gives the full dimension of the space being acted upon.

Definition: Nullity

The nullity $\operatorname{nullity}(A)$ of a matrix A is the dimension of its null space.

Theorem 3.22: Rank theorem

If A is an $m \times n$ matrix, then

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

We're now in a position to add statements about the row, column, and null spaces to the FTLA.

Theorem 3.23: Additions to the FTLA

Let A be an $n \times n$ matrix. The following statements are equivalent.

(a) A is invertible.

$\vdots$

(f) $\operatorname{rank}(A) = n$

(g) $\operatorname{nullity}(A) = 0$

(h) The column vectors of A are linearly independent.

(i) The column vectors of A span $\mathbb{R}^n$.

(j) The column vectors of A form a basis for $\mathbb{R}^n$.

(k) The row vectors of A are linearly independent.

(l) The row vectors of A span $\mathbb{R}^n$.

(m) The row vectors of A form a basis for $\mathbb{R}^n$.

For the full theorem, see Appendix A.

As an application, we have a nice little result that will be useful later.

Theorem 3.24: Rank and invertibility of $A^T A$

Let A be an $m \times n$ matrix. Then

(a) $\operatorname{rank}(A^T A) = \operatorname{rank}(A)$ and

(b) $A^T A$ is invertible if and only if $\operatorname{rank}(A) = n$.

Proof. We only prove (a) since part (b) follows trivially. Since A and $A^T A$ both have n columns, it suffices to show that the two matrices have the same null space.

If $\mathbf{x} \in \operatorname{null}(A)$ then $A^T A\mathbf{x} = A^T \mathbf{0} = \mathbf{0}$, meaning $\mathbf{x} \in \operatorname{null}(A^T A)$. If instead $\mathbf{x} \in \operatorname{null}(A^T A)$, then

$$(A\mathbf{x}) \cdot (A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \mathbf{0} = \mathbf{0},$$

so $(A\mathbf{x}) \cdot (A\mathbf{x}) = \mathbf{0}$ and $A\mathbf{x} = \mathbf{0}$. Thus $\mathbf{x} \in \operatorname{null}(A)$ if and only if $\mathbf{x} \in \operatorname{null}(A^T A)$. $\square$

There's one more thing that will come in handy later—unique vector representations under a given basis. At a high level, we'd take two representations $\mathbf{u}$ and $\mathbf{v}$ and show that each coefficient in $\mathbf{u} - \mathbf{v}$ must be zero.

Theorem 3.25: Unique representation of a vector

Let S be a subspace of $\mathbb{R}^n$ and let $\mathcal{B}$ be a basis for S. For every vector $\mathbf{v}$ in S, there is exactly one way to write $\mathbf{v}$ as a linear combination of the basis vectors in $\mathcal{B}$.

4 Eigenvalues and Eigenvectors

4.1 Determinants

We have already encountered determinants in our study of inverse matrices, albeit not by name. Here, we generalize to square matrices of any size.

Definition: Determinant

Let A be an $n \times n$ matrix, and let A_{ij} be the submatrix of A with its ith row and jth column deleted. Then the quantity

$$C_{ij} = (-1)^{i+j} \det A_{ij}$$

is called the (i, j)-cofactor of A. (The quantity $\det A_{ij}$ is the (i, j)-minor of A.) If $n \geq 2$, then the determinant of A is the scalar

$$\det A = |A| = \sum_{j=1}^{n} a_{1j} C_{1j}.$$

This gives us a way to compute determinants via a "cofactor expansion" along the top row, but it turns out we can do it on any row or column we'd like! This can save us a lot of work in certain situations.

Theorem 4.1: Computing determinants

The determinant of an $n \times n$ matrix A with $n \geq 2$ is given, for any $i, j \leq n$, by

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij} = \sum_{i=1}^{n} a_{ij} C_{ij}.$$

Theorem 4.2: Determinant of a triangular matrix

The determinant of a triangular matrix is the product of the entries on its main diagonal.

Sometimes it will be useful to use row reduction as an alternative to cofactor expansion. We do this by taking advantage of the following rules, which simply follow from cofactor expansion.

Theorem 4.3: Determinants via row reduction

Let A be a square matrix.

(a) If A has a zero row, then $\det A = 0$.

(b) If B is obtained by interchanging two rows of A, then $\det B = -\det A$.

(c) If A has two identical rows, then $\det A = 0$.

(d) If B is obtained by multiplying a row of A by k, then $\det B = k \det A$.

(e) If A, B, and C are identical except the ith row of C is the sum of the ith rows of A and B, then $\det C = \det A + \det B$.

(f) If B is obtained by adding a multiple of one row of A to another row, then $\det B = \det A$.

We'll use these to connect determinants and invertibility, using elementary matrices as our bridge.

Theorem 4.4: Determinant of an elementary matrix

Let E be an $n \times n$ elementary matrix.

(a) If E results from interchanging two rows of I_n, then $\det E = -1$.

(b) If E results from multiplying one row of I_n by k, then $\det E = k$.

(c) If E results from adding a multiple of one row of I_n to another row, then $\det E = 1$.

This allows us to succinctly summarize parts (b), (d), and (f) of Theorem 4.3 in another theorem.

Lemma 4.5: Determinant after an ERO

Let B be a square matrix and let E be an elementary matrix of the same size. Then

$$\det(EB) = (\det E)(\det B).$$

Finally, with the power of the fundamental theorem, we get the following.

Theorem 4.6: Condition for invertibility

A square matrix A is invertible if and only if $\det A \neq 0$.

Proof. Let A be an $n \times n$ matrix and let R be the reduced row echelon form of A. If $E_1, E_2, \ldots, E_r$ are the elementary matrices that reduce A to R then $E_r \cdots E_2 E_1 A = R$ and

$$(\det E_r) \cdots (\det E_2)(\det E_1)(\det A) = \det R.$$

The elementary matrices have nonzero determinants, so $\det A \neq 0$ if and only if $\det R \neq 0$.

Now suppose A is invertible; by the FTLA $R = I_n$, meaning $\det R \neq 0$ and so $\det A \neq 0$. Conversely, suppose $\det A \neq 0$; then $\det R \neq 0$, so R cannot contain a zero row, $R = I_n$, and A is invertible. □

Many matrix operations come with their own ways to calculate the determinants of their resulting matrices. We'll consider scalar multiplication, matrix multiplication, inverses, and transposes; we exclude addition because there is no clear relationship between $\det(A + B)$ and the individual determinants of A and B.

Theorem 4.7: Determinant after matrix operations

Let A and B be $n \times n$ matrices.

(a) $\det kA = k^n \det A$.

(b) $\det(AB) = (\det A)(\det B)$.

(c) $\det(A^{-1}) = (\det A)^{-1}$ for invertible A.

(d) $\det A = \det A^T$.

4.2 Eigenvalues and Eigenvectors

We now begin investigating one of the most central problems to linear algebra: the eigenvalue problem.

Definition: Eigen-stuff

Let A be a square matrix. A scalar λ is called an eigenvalue of A if there is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Such a vector $\mathbf{x}$ is called an eigenvector of A corresponding to λ. The set of all eigenvectors corresponding to λ, including the zero vector, is called the eigenspace of λ and is denoted by E_λ.

The study of eigenvectors is the study of vectors whose directions are invariant under a matrix transformation; we'll now discuss how to compute them. From the definition, $(\lambda, \mathbf{x})$ is an "eigenpair" of A if they satisfy the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$. This gives us an easy way to immediately compute the eigenvalues of A.

Definition: Characteristic polynomial

The polynomial in λ resulting from the expansion of $\det(A - \lambda I)$ is called the characteristic polynomial $c_A(\lambda)$ of A. The equation $\det(A - \lambda I) = 0$ is called the characteristic equation of A.

Theorem 4.8: Computing eigenvalues

The eigenvalues of a square matrix A are precisely the roots λ of the characteristic polynomial of A.

Once we have our eigenvalues, we use them in the equation $(A - \lambda I)\mathbf{x} = 0$ to determine the eigenvectors and, thus, eigenspaces.

Theorem 4.9: Computing eigenvectors

The eigenspace of a matrix A corresponding to the eigenvalue λ is precisely $\text{null}(A - \lambda I)$.

The remainder of this chapter will be dedicated to investigating this eigenstuff, both in their own right and in applications. Firstly, just like with determinants, our job is easy when our matrix has certain properties.

Theorem 4.10: Eigenvalues of a triangular matrix

The eigenvalues of a triangular matrix are the entries on its main diagonal.

Theorem 4.11: Eigenvalues after exponentiation

Let A be a square matrix with an eigen-pair $(\lambda, \mathbf{x})$.

(a) For any positive n, $(\lambda^n, \mathbf{x})$ is an eigen-pair of A^n.

(b) Is A is invertible, then $(1/\lambda, \mathbf{x})$ is an eigen-pair of A^{-1}.

(c) If A is invertible, then for any integer n, $(\lambda^n, \mathbf{x})$ is an eigen-pair of A^n.

We can now draw a connection to invertible matrices. Just like with determinants, for a matrix to be invertible, its eigenvalues must be nonzero.

Theorem 4.12: Condition for invertibility

A square matrix is invertible if and only if it does not have 0 as an eigenvalue.

Proof. Let A be a square matrix, so A is invertible if and only if $\det A \neq 0$. But this is equivalent to $\det(A - 0I) \neq 0$, meaning 0 is not a root of the characteristic equation of A. □

With this, we can add two more statements to our fundamental theorem.

Theorem 4.13: Additions to the FTLA

Let A be an $n \times n$ matrix. The following statements are equivalent.

(a) A is invertible.

⋮

(n) $\det A \neq 0$.

(o) 0 is not an eigenvalue of A.

For the full theorem, see Appendix A.

We can use the properties discussed so far to show what happens when we multiply a vector by the same matrix multiple time in succession. We take advantage of the case in which the vector can be represented as a linear combination of eigenvectors.

Theorem 4.14: Repeated matrix transformations

Suppose the $n \times n$ matrix A has eigenpairs $(\lambda_1, \mathbf{x}_1), (\lambda_2, \mathbf{x}_2), \ldots, (\lambda_m, \mathbf{x}_m)$. If $\mathbf{x}$ is a vector in $\mathbb{R}^n$ that can be expressed as a linear combination of these eigenvectors, then for every integer k,

$$A^k \mathbf{x} = c_1 \lambda_1^k \mathbf{v}_1 + c_2 \lambda_2^k \mathbf{v}_2 + \cdots + c_m \lambda_m^k \mathbf{v}_m.$$

We won't always be able to apply this, though, since it's possible that $\mathbf{x}$ is not a linear combination of eigenvectors. As a step toward seeing when we can apply the theorem, we have the following.

Theorem 4.15: Condition for linearly independent eigenvectors

Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. Then these eigenvectors are linearly independent.

Proof. Suppose, for contradiction, that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly dependent, and let $\mathbf{v}_{k+1}$ be the first of these vectors that can be expressed as a linear combination of the previous ones. Then there exist $c_1, c_2, \ldots, c_k$ such that

$$\mathbf{v}_{k+1} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_k \mathbf{v}_k.$$

Multiplying both sides of this equation by A gives

$$\lambda_{k+1} \mathbf{v}_{k+1} = A\mathbf{v}_{k+1} = c_1 \lambda_1 \mathbf{v}_1 + c_2 \lambda_2 \mathbf{v}_2 + \cdots + c_k \lambda_k \mathbf{v}_k;$$

instead multiplying by λ_{k+1} gives

$$\lambda_{k+1} \mathbf{v}_{k+1} = c_1 \lambda_{k+1} \mathbf{v}_1 + c_2 \lambda_{k+1} \mathbf{v}_2 + \cdots + c_k \lambda_{k+1} \mathbf{v}_k.$$

After subtracting these we get

$$0 = c_1(\lambda_1 - \lambda_{k+1})\mathbf{v}_1 + \cdots + c_k(\lambda_k - \lambda_{k+1})\mathbf{v}_k,$$

and since these $\mathbf{v}_i$ are linearly independent (and the eigenvalues are distinct) the c_i must be zero. But this means $\mathbf{v}_{k+1}$ is the zero vector, a contradiction. So $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent. □

4.3 Similarity and Diagonalization

We've already considered a couple of different ways to transform a matrix, namely Gaussian and Gauss-Jordan elimination. Here, we consider another type of transformation that preserves eigenvalues and other key properties.

Definition: Similar matrices

Let A and B be square matrices of the same size. We say that A is similar to B if there is an invertible matrix P such that $P^{-1}AP = B$. If A is similar to B, we write $A \sim B$.

Similarity can be interpreted as a kind of equivalence between matrices. We thus call similarity an equivalence relation, meaning it has the following familiar properties.

Theorem 4.16: Similarity as an equivalence relation

Let A, B, and C be square matrices of the same size. Then:

(a) $A \sim B$.

(b) If $A \sim B$, then $B \sim A$.

(c) If $A \sim B$ and $B \sim C$, then $A \sim C$.

Objects that are related via an equivalence relation usually share important properties. This is true of similar matrices, and some of these properties are below.

Theorem 4.17: Shared properties of similar matrices

Let A and B be square matrices of the same size with $A \sim B$. Then A and B have the same determinant, invertibility, rank, characteristic polynomial, and eigenvalues.

Also, $A^m \sim B^m$ for all integers $m \geq 0$, and if A is invertible then this is true for all integers m.

The best possible situation is when a square matrix is similar to a diagonal matrix, because these are really nice to work with. The idea of finding a similar diagonal matrix is intimately related to eigenstuff; not only do they give us a condition for diagonalizability, but they help us do the actual diagonalization.

Definition: Diagonalizable matrix

A square matrix A is diagonalizable if there is a diagonal matrix D such that A is similar to D—that is, if there is an invertible matrix P such that $P^{-1}AP = D$.

Theorem 4.18: Condition and method for diagonalization

Let A be an $n \times n$ matrix. A is diagonalizable if and only if it has n linearly independent eigenvectors.

More precisely, there exist an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$ if and only if the columns of P are n linearly independent eigenvectors of A and the diagonal entries of D are the eigenvalues of A corresponding to the eigenvectors in P in the same order.

Proof. Suppose A is similar to the diagonal matrix D. Let the columns of P be $\mathbf{p}_1, \dots, \mathbf{p}_n$ and let the diagonal entries of D be $\lambda_1, \dots, \lambda_n$. Then $AP = PD$ becomes

$$\begin{bmatrix} A\mathbf{p}_1 & \cdots & A\mathbf{p}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{p}_1 & \cdots & \lambda_n\mathbf{p}_n \end{bmatrix},$$
$$A\mathbf{p}_1 = \lambda_1\mathbf{p}_1, \; \dots, \; A\mathbf{p}_n = \lambda_n\mathbf{p}_n.$$

Since P is invertible, its columns are linearly independent. For the reverse direction we can simply trace this argument in reverse. □

Usually we'd have to verify that P is invertible before using it to find a matrix similar to A. The next theorem shows that this is not necessary for diagonalization problems.

Theorem 4.19: Basis eigenvectors are linearly independent

Let A be a square matrix. The set of basis vectors for all of the eigenspaces of A is linearly independent.

Proof. Let A have k distinct eigenvalues and let $\mathcal{B}_i = \{\mathbf{v}_{i1}, \dots, \mathbf{v}_{in}\}$. Suppose some nontrivial linear combination of the vectors in all the $\mathcal{B}_i$ is the zero vector. If $\mathbf{x}_i$ is the portion of this combination comprised of the vectors in $\mathcal{B}_i$ then we can write $\mathbf{x}_1 + \cdots + \mathbf{x}_k = \mathbf{0}$, meaning the $\mathbf{x}_i$ are linearly dependent. Now, since $\mathbf{x}_i \in E_{\lambda_i}$, it is either $\mathbf{0}$ or an eigenvector corresponding to λ_i. But since the λ_i are distinct, if any of the $\mathbf{x}_i$ is an eigenvector then they are linearly independent, a contradiction. □

This shows that linear independence is preserved when the bases of different eigenspaces are combined. So when we have n different eigenspaces, each of dimension one, we must have a diagonalizable matrix.

Theorem 4.20: Distinct eigenvalues imply diagonalizibility

If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.

When this theorem isn't applicable, there's still a way we can use eigenvalues to determine diagonalizability.

Definition: Algebraic and geometric multiplicity

The algebraic multiplicity of an eigenvalue λ is its multiplicity as a root of the characteristic polynomial. The geometric multiplicity of the eigenvalue is $\dim(E_\lambda)$

Lemma 4.21: Diagonalizibility from multiplicities

If A is a square matrix, then the geometric multiplicity of each eigenvalue is less than or equal to its algebraic multiplicity.

Theorem 4.22: Conditions for diagonalizibility

Let A be an $n \times n$ matrix. The following statements are equivalent.

(a) A is diagonalizable.

(b) The union $\mathcal{B}$ of the bases of the eigenspaces of A contains n vectors.

(c) The algebraic multiplicity of each eigenvalue equals its geometric multiplicity.

5 Orthogonality

5.1 Orthogonality in $\mathbb{R}^n$

Here, we extend the notion of orthogonality from pairs of vectors to sets of vectors.

Definition: Orthogonal set of vectors

A set of vectors in $\mathbb{R}^n$ is called an orthogonal set if all pairs of distinct vectors in the set are orthogonal.

There are many advantages to working with orthogonal sets. One of these is that they are necessarily linearly independent, meaning they make for convenient bases for subspaces; for this reason, they make convenient bases for subspaces.

Theorem 5.1: Orthogonal sets are linearly independent

Any orthogonal set of nonzero vectors in $\mathbb{R}^n$ is linearly independent.

Proof. If $c_1, \ldots, c_k$ are scalars such that $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$ then

$$c_1(\mathbf{v}_1 \cdot \mathbf{v}_i) + \cdots + c_k(\mathbf{v}_k \cdot \mathbf{v}_i) = 0.$$

All of the dot products here are zero except for $\mathbf{v}_i \cdot \mathbf{v}_i$, so we have $c_i(\mathbf{v}_i \cdot \mathbf{v}_i) = 0$. Thus $c_i = 0$, and since this is true for all $i = 1, \ldots, k$ implies that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly independent set. □

Definition: Orthogonal basis

An orthogonal basis for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthogonal set.

Of course, any vector in W can be written as a linear combination of these orthogonal vectors. Another convenience that comes with orthogonal sets is that the coefficients of such a combination are easy to find.

Theorem 5.2: Coefficients of an orthogonal linear combination

Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then the unique scalars $c_1, c_2, \ldots, c_k$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

are given by

$$c_1 = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \quad \text{for} \quad i = 1, \ldots, k.$$

Proof. Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for W, there are unique scalars such that $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$. So

$$\mathbf{w} \cdot \mathbf{v}_i = (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_i = c_i(\mathbf{v}_i \cdot \mathbf{v}_i),$$

and dividing by $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$ gives the equation we desire. □

Notice that, in the case of an orthogonal basis, we simply project the vector onto each of the basis vectors to find the coefficients of the linear combination. Things are even more convenient if our orthogonal set consists entirely of unit vectors.

Definition: Orthonormal set

A set of vectors in $\mathbb{R}^n$ is an orthonormal set if it is an orthogonal set of unit vectors. An orthonormal basis for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthonormal set.

Theorem 5.3: Coefficients of an orthonormal linear combination

Let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{w} \cdot \mathbf{q}_2)\mathbf{q}_2 + \cdots + (\mathbf{w} \cdot \mathbf{q}_k)\mathbf{q}_k,$$

and this representation is unique.

When a matrix has columns that form an orthonormal set, they turn out to have many very nice properties. The rest of this section is dedicated to examining these.

Notice that we can think of matrix multiplication as a grid of dot products between rows and columns, and when a row-column pair is orthogonal, they'll simply cancel. So if we have a matrix whose columns comprise an orthonormal set, multiplying by its transpose simply leaves behind an identity matrix.

Theorem 5.4: Transpose of a matrix with orthonormal columns

The columns of an $m \times n$ matrix Q form an orthonormal set if and only if $Q^T Q = I_n$.

For square Q we can characterize this behavior in terms of matrix inverses.

Definition: Orthogonal matrix

A square matrix Q whose columns form an orthonormal set is called an orthogonal matrix.

Theorem 5.5: Inverse of an orthogonal matrix

A square matrix Q is orthogonal if and only if $Q^{-1} = Q^T$.

Perhaps expectedly, orthogonal matrices have some convenient properties. One of the most important ones is that they preserve the lengths of the vectors they transform—that is, they are isometric.

Theorem 5.6: Orthogonal matrices are isometric

Let Q be a $n \times n$ matrix. The following statements are equivalent.

(a) Q is orthogonal.

(b) $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

(c) $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for every $\mathbf{x}$ in $\mathbb{R}^n$.

Proof. Suppose Q is orthogonal. Then for any $\mathbf{x}, y \in \mathbb{R}^n$ we have

$$Q\mathbf{x} \cdot Q\mathbf{y} = (Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}.$$

Now take $\mathbf{y} = \mathbf{x}$, so $\|Q\mathbf{x}\| = \sqrt{Q\mathbf{x} \cdot Q\mathbf{x}} = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \|\mathbf{x}\|$. This proves (a) $\implies$ (b) $\implies$ (c). Now, let $\mathbf{q}_i$ denote the ith column of Q; it could be shown that

$$\mathbf{x} \cdot \mathbf{y} = \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right) + \frac{1}{4}\left(\|Q\mathbf{x}+Q\mathbf{y}\|^2 - \|Q\mathbf{x}-Q\mathbf{y}\|^2\right) = Q\mathbf{x} \cdot Q\mathbf{y}.$$

If $\mathbf{e}_i$ is the ith standard basis vector, then $\mathbf{q}_i = Q\mathbf{e}_i$ and $\mathbf{q}_i \cdot \mathbf{q}_j = \mathbf{e}_i \cdot \mathbf{e}_j$, which is one for like indices and zero otherwise. Thus the columns of Q form an orthonormal set. $\square$

We'll finish off with a couple of theorems elaborating on some properties of orthogonal matrices.

Theorem 5.7: Orthogonal matrices have orthonormal rows

If Q is an orthogonal matrix then its rows form an orthonormal set.

Proof. Since $Q^{-1} = Q^T$ we have $(Q^T)^{-1} = Q = (Q^T)^T$, meaning Q^T is an orthogonal matrix. □

Theorem 5.8: Properties of orthogonal matrices

Let Q be an orthogonal matrix.

(a) Q^{-1} is orthogonal.

(b) $\det Q = \pm 1$.

(c) If λ is an eigenvalue of Q, then $|\lambda| = 1$.

(d) If Q_1 and Q_2 are orthogonal matrices of the same size, then so is $Q_1 Q_2$.

5.2 Orthogonal Complements and Projections

Now we generalize two more ideas: the vector normal to a plane, and the projection of one vector onto another. We begin with the first of these.

Definition: Orthogonal complement

Let W be a subspace of $\mathbb{R}^n$. We say that a vector $\mathbf{v}$ in $\mathbb{R}^n$ is orthogonal to W if $\mathbf{v}$ is orthogonal to every vector in W. The set of all such vectors is called the orthogonal complement of W, denoted $W^\perp$.

Theorem 5.9: Properties of orthogonal complements

Let W be a subspace of $\mathbb{R}^n$.

(a) $W^\perp$ is a subspace of $\mathbb{R}^n$.

(b) $(W^\perp)^\perp = W$.

(c) $W \cap W^\perp = \{\mathbf{0}\}$.

(d) If $W = \operatorname{span}(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k)$, then $\mathbf{v}$ is in $W^\perp$ if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ for all $i = 1, \ldots, k$.

Using orthogonal complements, we can make describe some fundamental relationships between the subspaces associated with a matrix. To relate the row space to the null space, notice that for $A\mathbf{x} = \mathbf{0}$ to be true, we must have $\mathbf{r} \cdot \mathbf{x} = 0$ for every row $\mathbf{r}$ in A; therefore, all rows in A are orthogonal to vectors in $\operatorname{null}(A)$. A similar line of reasoning relates $\operatorname{col}(A)$ to $\operatorname{null}(A^T)$.

Theorem 5.10: Fundamental spaces and orthogonality

Let A be a matrix. Then $(\operatorname{row}(A))^\perp = \operatorname{null}(A)$ and $(\operatorname{col}(A))^\perp = \operatorname{null}(A^T)$.

Now we'll move on to generalizing projections.

Definition: Orthogonal projection onto a subspace

Let W be a subspace of $\mathbb{R}^n$ and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for W. For any vector $\mathbf{v}$ in $\mathbb{R}^n$, the orthogonal projection of $\mathbf{v}$ onto W is defined as

$$\operatorname{proj}_W(\mathbf{v}) = \operatorname{proj}_{\mathbf{u}_1}(\mathbf{v}) + \operatorname{proj}_{\mathbf{u}_2}(\mathbf{v}) + \cdots + \operatorname{proj}_{\mathbf{u}_k}(\mathbf{v}).$$

The component of $\mathbf{v}$ orthogonal to W is the vector $\operatorname{perp}_W(\mathbf{v}) = \mathbf{v} - \operatorname{proj}_W(\mathbf{v})$.

We can use orthogonal projections to decompose a vector into orthogonal components.

Theorem 5.11: Orthogonal decomposition

Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there are unique vectors $\mathbf{w}$ in W and $\mathbf{w}^\perp$ in $W^\perp$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$.

Proof. To show that such a decomposition exists, take an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ for W and define $\mathbf{w} = \operatorname{proj}_W(\mathbf{v})$ with $\mathbf{w}^\perp = \operatorname{perp}_W(\mathbf{v})$. We can immediately see that $\mathbf{w} + \mathbf{w}^\perp = \mathbf{v}$ and $\mathbf{w} \in W$; to show that $\mathbf{w}^\perp \in W^\perp$ we could use a direct computation to get $\mathbf{u}_i \cdot \mathbf{v} = 0$ for all basis vectors $\mathbf{u}_i$.

To show uniqueness, suppose there is another decomposition $\mathbf{v} = \mathbf{w}' + \mathbf{w}'^\perp$. Then $\mathbf{w} + \mathbf{w}^\perp = \mathbf{w}' + \mathbf{w}'^\perp$ and $\mathbf{w} - \mathbf{w}' = \mathbf{w}'^\perp - \mathbf{w}^\perp$. But the left-hand side is in W while the right-hand side is in $W^\perp$, and because these subspaces intersect at $\mathbf{0}$, both sides must be zero. Thus $\mathbf{w}' = \mathbf{w}$ and $\mathbf{w}'^\perp = \mathbf{w}^\perp$. $\square$

Lastly, we have a relationship between the dimensions of a subspace and its orthogonal complement.

Theorem 5.12: Dimension of orthogonal subspace

If W is a subspace of $\mathbb{R}^n$, then
$$\dim W + \dim W^\perp = n.$$

Proof. Let $\mathcal{B}$ be the union of bases for W and $W^\perp$. We can immediately see that $\mathcal{B}$ is an orthogonal set and so is linearly independent. Now, by the orthogonal decomposition theorem, any $\mathbf{v} \in \mathbb{R}^n$ can be written as a sum of vectors in W and $W^\perp$, which can in turn be written as linear combinations of vectors in $\mathcal{B}$. Thus $\mathcal{B}$ is a basis for $\mathbb{R}^n$; it follows that $\dim W + \dim W^\perp = \dim \mathbb{R}^n$. $\square$

Note that the rank theorem is a corollary to this theorem.

5.3 The Gram-Schmidt Process and the QR Factorization

Given how useful orthogonal bases have proven to be, it might also be useful to know how to construct one from a non-orthogonal basis. We do this using a series of projections.

Theorem 5.13: Gram-Schmidt process

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ be a basis for a subspace W of $\mathbb{R}^n$, let $W_i = \operatorname{span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i)$, and define the following:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \operatorname{perp}_{W_1}(\mathbf{x}_2) &&= \mathbf{x}_k - \operatorname{proj}_{\mathbf{v}_1} \mathbf{x}_k \\
&\vdots \\
\mathbf{v}_k &= \operatorname{perp}_{W_{k-1}}(\mathbf{x}_k) &&= \mathbf{x}_k - \operatorname{proj}_{\mathbf{v}_1} \mathbf{x}_k - \operatorname{proj}_{\mathbf{v}_2} \mathbf{x}_k - \cdots - \operatorname{proj}_{\mathbf{v}_{k-1}} \mathbf{x}_k
\end{aligned}$$

Then for each $i = 1, \ldots, k$, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for W_i. In particular, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W.

We can use this to decompose any invertible square matrix into an orthogonal matrix and a triangular matrix—the vectors resulting from the process are the columns of Q, while their coefficients are the entries of R.

Theorem 5.14: QR factorization

Let A be an $m \times n$ matrix with linearly independent columns. Then A can be factored as $A = QR$, where Q is an $m \times n$ matrix with orthonormal columns and R is an invertible upper triangular matrix.

6 Vector Spaces

6.1 Vector Spaces and Subspaces

So far we have treated vectors as ordered tuples of real numbers. However, there are plenty of other objects (like matrices) that behave similarly. We use the familiar properties of vectors in $\mathbb{R}^n$ to motivate a more general definition for a vector, one that encapsulates this characteristic behavior.

Definition: Vector space

Let V be a set on which two operations, called addition and scalar multiplication, have been defined. If $\mathbf{u}$ and $\mathbf{v}$ are in V, the sum of $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u}+\mathbf{v}$. If c is a scalar, the scalar multiple of $\mathbf{u}$ by c is denoted by $c\mathbf{u}$. If the following axioms hold for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and for all scalars $c, d \in \mathbb{R}$, then V is called a vector space and its elements are called vectors.

1. $\mathbf{u}+\mathbf{v}$ is in V.
2. $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$.
3. $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$.
4. There exists a $\mathbf{0} \in V$ such that $\mathbf{u}+\mathbf{0} = \mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is an element $-\mathbf{u}$ in V such that $\mathbf{u}+(-\mathbf{u}) = \mathbf{0}$.
6. $c\mathbf{u}$ is in V.
7. $c(\mathbf{u}+\mathbf{v}) = c\mathbf{u}+c\mathbf{v}$.
8. $(c+d)\mathbf{u} = c\mathbf{u}+d\mathbf{u}$.
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$.
10. $1\mathbf{u} = \mathbf{u}$.

We'll also define the difference between $\mathbf{u}$ and $\mathbf{v}$, denoted $\mathbf{u}-\mathbf{v}$, as $\mathbf{u}+(-\mathbf{v})$.

This definition allows us to describe many of the ideas we've already seen, just in more general terms. The rest of this section is dedicated to these, omitting the more straightforward proofs.

Theorem 6.1: Basic properties of vectors

Let V be a vector space, $\mathbf{u}$ a vector in V, and c a scalar.

(a) $0\mathbf{u} = \mathbf{0}$.

(b) $c\mathbf{0} = \mathbf{0}$.

(c) $(-1)\mathbf{0} = -\mathbf{u}$.

(d) If $c\mathbf{u} = \mathbf{0}$, then $c = 0$ or $\mathbf{u} = \mathbf{0}$.

Proof. To prove (b), note that $c\mathbf{0} = c(\mathbf{0}+\mathbf{0}) = c\mathbf{0}+c\mathbf{0}$. Subtracting $c\mathbf{0}$ to either side gives $\mathbf{0} = c\mathbf{0}$. As for (d), suppose $c\mathbf{u} = \mathbf{0}$ and $c \neq 0$; then $\mathbf{u} = (1/c)(c\mathbf{u}) = \mathbf{0}$. □

Definition: Linear combination

Let V be a vector space, let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be vectors in V, and let $c_1, c_2, \ldots, c_n$ be scalars. Then $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ is a linear combination of these vectors.

Definition: Subspace

A subset W of a vector space V is called a subspace of V if W is itself a vector space with the same scalars and operations as V.

The criteria for a subset being a subspace are analogous to the ones we saw previously, and a full justification for them would require showing that the other eight axioms follow.

Theorem 6.2: Conditions for a subspace

Let V be a vector space and let W be a nonempty subset of V. Then W is a subspace of V if and only if the following conditions hold.

(a) If $\mathbf{u}$ and $\mathbf{v}$ are in W, then $\mathbf{u}+\mathbf{v}$ is in W.

(b) If $\mathbf{u}$ is in W and c is a scalar, then $c\mathbf{u}$ is in W.

Definition: Span

If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a set of vectors in a vector space V, then the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is called the span of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ and is denoted by $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ or $\operatorname{span}(S)$. If $V = \operatorname{span}(S)$, then S is called a spanning set for V and V is said to be spanned by S.

Theorem 6.3: Span as a subspace

Let S be a set of vectors in a vector space V.

(a) $\operatorname{span}(S)$ is a subspace of V.

(b) $\operatorname{span}(S)$ is the smallest subspace of V that contains the vectors in S.

Proof. To prove (b), let W be a subspace of V that has S as a subset. W contains every linear combination of the vectors in S, meaning $\operatorname{span}(S)$ is contained in W. □

6.2 Linear Independence, Basis, and Dimension

We'll continue to generalize many of the definitions and properties in $\mathbb{R}^n$ to all vector spaces. In most cases, the theorems (and their proofs) are identical, just replacing $\mathbb{R}^n$ with V. We begin with linear independence.

Definition: Linear dependence

A set S of vectors in a vector space V is linearly dependent if there is a nontrivial linear combination of the vectors in S that creates the zero vector. A set of vectors that is not linearly dependent is said to be linearly independent.

Theorem 6.4: Condition for linear dependence

A set of vectors in a vector space V is linearly dependent if and only if at least one of the vectors can be expressed as a linear combination of the others.

With linear independence defined, we can move onto bases.

Definition: Basis

A subset $\mathcal{B}$ of a vector space V is a basis for V if

- $\mathcal{B}$ spans V and
- $\mathcal{B}$ is linearly independent.

Theorem 6.5: Unique representation of a vector

Let V be a vector space and let $\mathcal{B}$ be a basis for V. For every vector $\mathbf{v}$ in V, there is exactly one way to write $\mathbf{v}$ as a linear combination of the basis vectors in $\mathcal{B}$.

Now we introduce a new concept, one that we alluded to earlier. Every basis for a vector space also provides a new coordinate system for that vector space; this coordinate system is defined as follows.

Definition: Coordinate vector

Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{v} \in V$, and write $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Then $c_1, \ldots, c_n$ are called the coordinates of $\mathbf{v}$ with respect to $\mathcal{B}$, and the column vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

is called the coordinate vector of $\mathbf{v}$ with respect to $\mathcal{B}$.

There is a couple of useful things that arise from the use of coordinate systems. First is a corollary to the previous theorem: coordinate vectors preserve linear combinations.

Theorem 6.6: Coordinates preserve linear combinations

Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V and let c be a scalar. Then

(a) $[\mathbf{u} + \mathbf{v}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}}$ and

(b) $[c\mathbf{u}]_{\mathcal{B}} = c[\mathbf{u}]_{\mathcal{B}}$.

Proof. (Sketch) We can write $\mathbf{u}$ and $\mathbf{v}$ as linear combinations of basis vectors in V and add or multiply them as needed. □

Secondly, coordinates allow us to transfer information from a general vector space to $\mathbb{R}^n$, a space that we have already studied thoroughly; we will explore this idea in detail in the next section. For now, we can see that a set's linear independence (or dependence) carries over to the set's $\mathbb{R}^n$ counterpart.

Theorem 6.7: Coordinates preserve linear independence

Let $\mathcal{B}$ be a basis for a vector space V and let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be vectors in V. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is linearly independent in V if and only if $\{[\mathbf{u}_1]_{\mathcal{B}}, \ \ldots, \ [\mathbf{u}_k]_{\mathcal{B}}\}$ is linearly independent in $\mathbb{R}^n$.

Proof. (Sketch) We can write a zero linear combination of the coordinate vectors, apply the previous theorem, and use the definition of linear independence to show that the coefficients of the combination are zero. □

Now, having defined basis, we can generalize notions of dimension. We begin with some motivating theorems, both of which establish the size of a basis as an important quantity.

Theorem 6.8: Necessary conditions for span and linear independence

Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V.

(a) Any set of more than n vectors in V must be linearly dependent.

(b) Any set of fewer than n vectors in V cannot span V.

Theorem 6.9: Sizes of distinct bases

If a vector space V has a basis with n vectors, then every basis for V has exactly n vectors.

Now that we've established that every basis for a vector space contains the same number of vectors, we can define dimension as an intrinsic property of vector spaces.

Definition: Dimension

A vector space V is called finite-dimensional if it has a basis containing finitely many vectors. The dimension of V, denoted by $\dim(V)$, is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. A vector space that has no finite basis is called infinite-dimensional.

Knowing the dimension of a vector space provides us with much information about V. Some of this information is described in the next two theorems.

Theorem 6.10: Dimension, span, and linear independence

Let V be a vector space with $\dim(V) = n$.

(a) Any linearly independent set in V contains at most n vectors.

(b) Any spanning set for V contains at least n vectors.

(c) Any linearly independent set of exactly n vectors in V is a basis for V.

(d) Any spanning set of exactly n vectors in V is a basis for V.

(e) Any linearly independent set in V can be extended to a basis for V.

(f) Any spanning set for V can be reduced to a basis for V.

Theorem 6.11: Dimension of a subspace

Let W be a subspace of a finite-dimensional vector space V. Then:

(a) W is finite-dimensional and $\dim(W) \leq \dim(V)$.

(b) $\dim(W) = \dim(V)$ if and only if $W = V$.

6.3 Change of Basis

Equipped with a working knowledge of coordinate systems in general vector spaces, we can now investigate how to convert between coordinate systems within a vector space. Most of our work here will center around a matrix that does this for us.

Definition: Change-of-basis matrix

Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\mathcal{C}$ be bases for a vector space V. The square matrix whose columns are the coordinate vectors $[\mathbf{u}_1]_\mathcal{C}, [\mathbf{u}_2]_\mathcal{C}, \ldots, [\mathbf{u}_n]_\mathcal{C}$ of the vectors in $\mathcal{B}$ with respect to $\mathcal{C}$ is denoted by $P_{\mathcal{C}\leftarrow\mathcal{B}}$ and is called the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$. That is,

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[\mathbf{u}_1]_\mathcal{C}, [\mathbf{u}_2]_\mathcal{C}, \cdots, [\mathbf{u}_n]_\mathcal{C}].$$

If we multiply a vector written with respect to $\mathcal{B}$ by this change-of-basis matrix, we get the same vector written with respect to $\mathcal{C}$. The change-of-basis matrix has this property and others, as described below.

Theorem 6.12: Properties of the change-of-basis matrix

Let $\mathcal{B}$ and $\mathcal{C}$ be bases for a vector space V and let $P_{\mathcal{C}\leftarrow\mathcal{B}}$ be the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$.

(a) $P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_\mathcal{B} = [\mathbf{x}]_\mathcal{C}$ for all $\mathbf{x}$ in V.

(b) $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is the unique matrix P with the above property.

(c) $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible and $(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$.

Proof. For (a) we can simply write

$$[\mathbf{x}]_{\mathcal{C}} = [c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n]_{\mathcal{C}} = \begin{bmatrix}[\mathbf{u}_1]_{\mathcal{C}} & \cdots & [\mathbf{u}_n]_{\mathcal{C}}\end{bmatrix}\begin{bmatrix}c_1\\ \vdots \\ c_n\end{bmatrix}.$$

For (b), let P be a matrix satisfying $P[\mathbf{x}]_{\mathcal{B}}$ for all $\mathbf{x}$ in V. If $\mathbf{x} = \mathbf{u}_i$ is the ith basis vector in $\mathcal{B}$ then the ith column of P is

$$\mathbf{p}_i = P\mathbf{e}_i = P[\mathbf{u}_i]_{\mathcal{B}} = [\mathbf{u}_i]_{\mathcal{C}}.$$

For (c), notice that the columns of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are linearly independent and that $[\mathbf{x}]_{\mathcal{B}} = (P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1}[\mathbf{x}]_{\mathcal{C}}$. □

There is another way we can compute the change of basis matrix, this time via row reduction. It is similar to how we compute the inverse of a matrix via row reduction.

Theorem 6.13: Change-of-basis matrix via row reduction

Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be bases for a vector space V. Also let $B = [[\mathbf{u}_1]_{\mathcal{E}} \cdots [\mathbf{u}_n]_{\mathcal{E}}]$ and $C = [[\mathbf{v}_1]_{\mathcal{E}} \cdots [\mathbf{v}_n]_{\mathcal{E}}]$, where $\mathcal{E}$ is any basis for V. Then row reduction applied to the $n \times 2n$ augmented matrix $[C \mid B]$ produces

$$[C \mid B] \to [I \mid P_{\mathcal{C}\leftarrow\mathcal{B}}].$$

6.4 Linear Transformations

Matrices can be interpreted as a transformation between two Euclidean vector spaces. Here, we extend this concept to linear transformations between arbitrary vector spaces.

Definition: Linear transformation

A linear transformation from a vector space V to a vector space W is a mapping $T : V \to W$ such that, for all $\mathbf{u}$ and $\mathbf{v}$ in V and for all scalars c,

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ and
- $T(c\mathbf{u}) = cT(\mathbf{u})$.

In other words, T is a linear transformation if and only if it preserves all linear combinations.

Definition: Zero and identity transformations

The zero transformation $T_0 : V \to W$ maps every vector in V to the zero vector in W. The identity transformation $I : V \to V$ maps every vector in V to itself.

Many of the properties of linear transformations are familiar. Some of these are listed below.

Theorem 6.14: Properties of linear transformations

Let $T : V \to W$ be a linear transformation. Then:

(a) $T(\mathbf{0}) = \mathbf{0}$.

(b) $T(-\mathbf{v}) = -T(\mathbf{v})$ for all $\mathbf{v}$ in V.

(c) $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in V.

The most important property of a linear transformation $T : V \to W$ is that T is completely determined by its effect on a basis for V.

Theorem 6.15: Linear transformation using a basis

Let $T : V \to W$ be a linear transformation and let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a spanning set for V. Then $T(\mathcal{B}) = \{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ spans the range of T.

Just like with matrix multiplication, we can compose linear transformations. The result is a new linear transformation that "skips" over the intermediate space.

Definition: Composition of linear transformations

If $T : U \to V$ and $S : V \to W$ are linear transformations, then the composition of S with T is the mapping $S \circ T : U \to W$, defined by $(S \circ T)(\mathbf{u}) = S(T(\mathbf{u}))$ where $\mathbf{u}$ is in U.

Theorem 6.16: Compositions of linear transformations are linear

If $T : U \to V$ and $S : V \to W$ are linear transformations, then $S \circ T : U \to W$ is a linear transformation.

Lastly, we draw one more generalization from matrices having to do with inverses.

Definition: Inverse transformation

A linear transformation $T : V \to W$ is invertible if there is a linear transformation $T^{-1} : W \to V$ such that

$$T^{-1} \circ T = I_V \quad \text{and} \quad T \circ T^{-1} = I_W.$$

In this case, T^{-1} is called an inverse for T.

Theorem 6.17: Condition for invertibility

If T is an invertible linear transformation, then its inverse is unique.

6.5 Kernel and Range

Now we'll generalize the notions of a matrix's null space and column space.

Definition: Kernel and range

Let $T : V \to W$ be a linear transformation. The kernel and range of T are, respectively,

$$\ker(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\} \quad \text{and} \quad \operatorname{range}(T) = \{T(\mathbf{v}) : \mathbf{v} \in V\}.$$

Like the null and column spaces of a matrix are subspaces of $\mathbb{R}^n$, the kernel and range of a linear transformation are subspaces of the domain and codomain.

Theorem 6.18: Kernel and range as subspaces

Let $T : V \to W$ be a linear transformation. Then:

(a) $\ker(T)$ is a subspace of V.

(b) $\operatorname{range}(T)$ is a subspace of W.

Like before, we assign names to the dimensions of these subspaces and draw a connection between them.

Definition: Rank and nullity

Let $T : V \to W$ be a linear transformation. The rank of T is $\operatorname{rank}(T) = \dim(\operatorname{range}(T))$ and the nullity of T is $\operatorname{nullity}(T) = \dim(\ker(T))$.

Theorem 6.19: Rank theorem

Let $T : V \to W$ be a linear transformation where V is finite-dimensional. Then

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim V.$$

Proof. (Sketch) This can be proved using the theory we currently have. It's more easily proved, however, in the language of matrices, which we'll develop in the next section. □

Now we have some vocabulary that will help us better describe the input and output spaces of a transformation, along with a couple of criteria which tell us when this vocabulary is valid.

Definition: One-to-one and onto

Let $T : V \to W$ be a linear transformation.

- T is one-to-one if, for all $\mathbf{u}, \mathbf{v}$ in V, $\mathbf{u} \neq \mathbf{v} \implies T(\mathbf{u}) \neq T(v)$.
- T is onto if $\operatorname{range}(T) = W$.

Theorem 6.20: Condition for one-to-one

A linear transformation T is one-to-one if and only if $\ker(T) = \{\mathbf{0}\}$

Proof. The forward direction follows immediately from the definition, so suppose $\ker(T) = \{0\}$. Let $\mathbf{u}, \mathbf{v} \in V$ satisfy $T(\mathbf{u}) = T(\mathbf{v})$, so $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}$ and $\mathbf{u} - \mathbf{v}$ is in the kernel of T. Thus $\mathbf{u} - \mathbf{v} = \mathbf{0}$ and $\mathbf{u} = \mathbf{v}$, meaning T is one-to-one. □

Theorem 6.21: Condition for onto

Let $\dim V = \dim W$. Then a linear transformation $T : V \to W$ is one-to-one if and only if it is onto.

Proof. By the rank theorem, $\operatorname{nullity}(T) = 0$ if and only if $\operatorname{rank}(T) = n$, where $\dim W = n$. □

In the previous section, we found that linear transformations, in a sense, preserve spanning sets. We now provide conditions under which a linear transformation will preserve linear independence and, thus, bases.

Theorem 6.22: One-to-one transformations preserve linear independence

Let $T : V \to W$ be a one-to-one linear transformation. If S is a linearly independent set in V, then $T(S)$ is a linearly independent set in W.

Corollary 6.23: Condition for the preservation of bases

Let $\dim V = \dim W$. Then a one-to-one linear transformation $T : V \to W$ maps a basis for V to a basis for W.

We can use all this to describe which linear transformations are invertible. (The proof of the following theorem is clunky, but that's just because there're lots of individual things to prove. Each part is relatively straightforward, so we omit the whole.)

Theorem 6.24: Condition for invertibility

A linear transformation T is invertible if and only if it is one-to-one and onto.

To finish the section, we concretely define what it means for two vector spaces to be "essentially the same" and when, exactly, this happens.

Definition: Isomorphism

A linear transformation $T : V \to W$ is called an isomorphism if it is one-to-one and onto. If V and W are two vector spaces such that there is an isomorphism from V to W, then we say that V is isomorphic to W and we write $V \cong W$.

Theorem 6.25: Condition for isomorphism

Let V and W be two finite-dimensional vector spaces. Then V is isomorphic to W if and only if $\dim V = \dim W$.

For our purposes the "field" of scalars we're pulling from is $\mathbb{R}$, but really we could build most of the same theory using any field (like $\mathbb{C}$ or $\mathbb{Z}_n$)—see an abstract algebra text for details on what this means.

6.6 The Matrix of a Linear Transformation

We can exploit the fact that all n-dimensional vector spaces are isomorphic to $\mathbb{R}^n$ to represent linear transformations as matrices. Rather than transform directly between vector spaces, we will transform between spaces of *coordinate* vectors, as the first theorem of this section shows.

Theorem 6.26: Matrix of a linear transformation

Let V and W be two finite-dimensional vector spaces with bases $\mathcal{B}$ and $\mathcal{C}$, respectively, where $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. If $T : V \to W$ is a linear transformation, then the $m \times n$ matrix A defined by

$$A = \begin{bmatrix} [T(\mathbf{v}_1)]_{\mathcal{C}} & \cdots & [T(\mathbf{v}_n)]_{\mathcal{C}} \end{bmatrix}$$

satisfies

$$A[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}}$$

for every vector $\mathbf{v}$ in V.

This matrix A is called the matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$; it is sometimes denoted by $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$. We can make some intuitive statements about the matrices of composite and inverse linear transformations.

Theorem 6.27: Composition of matrix transformations

Let U, V, and W be finite-dimensional vector spaces with bases $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$, respectively. Let $T : U \to V$ and $S : V \to W$ be linear transformations. Then

$$[S \circ T]_{\mathcal{D}\leftarrow\mathcal{B}} = [S]_{\mathcal{D}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}}.$$

Theorem 6.28: Inverse of a matrix transformation

Let $T : V \to W$ be a linear transformation between n-dimensional vector spaces V and W and let $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W, respectively. Then T is invertible if and only if the matrix $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible. In this case,

$$\left([T]_{\mathcal{C}\leftarrow\mathcal{B}}\right)^{-1} = \left[T^{-1}\right]_{\mathcal{B}\leftarrow\mathcal{C}}.$$

We can draw a relationship between the potential matrices representing a linear transformation from a vector space to itself. Namely, we can use the change-of-basis matrix between two bases to show that all of the possible matrices are similar to each other. This is stated more clearly in the following theorem.

Theorem 6.29: Matrix transformations are similar

Let V be a finite-dimensional vector space with bases $\mathcal{B}$ and $\mathcal{C}$ and let $T : V \to V$ be a linear transformation. Then

$$[T]_{\mathcal{C}} = P^{-1}[T]_{\mathcal{B}}P$$

where P is the change-of-basis matrix from $\mathcal{C}$ to $\mathcal{B}$.

This naturally leads to the notion of diagonalization.

Definition: Diagonalizible linear transformation

Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear transformation. Then T is called diagonalizable if there is a basis $\mathcal{C}$ for V such that the matrix $[T]_{\mathcal{C}}$ is a diagonal matrix.

We can now make some new additions to the Fundamental Theorem.

Theorem 6.30: Additions to the FTLA

Let A be an $n \times n$ matrix and let $T : V \to V$ be a linear transformation whose matrix $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ with respect to bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively, is A. The following statements are equivalent.

(a) A is invertible.

$\vdots$

(p) T is invertible.

(q) T is one-to-one.

(r) T is onto.

(s) $\ker(T) = \{\mathbf{0}\}$.

(t) $\mathrm{range}(T) = W$.

For the full theorem, see Appendix A.

7 Distance and Approximation

7.1 Inner Product Spaces

We begin this chapter by continuing our generalization of Euclidean vector spaces. Here, we define the inner product as a generalization of the dot product.

Definition: Inner product

An inner product on a vector space V is an operation that assigns to every pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ such that the following properties hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in W and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.
2. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$.
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

A vector space with an inner product is called an inner product space.

This technically only defines real inner product spaces, since we're assuming the scalars of our vector space are pulled from $\mathbb{R}$ and that the inner product yields a real number. For more general (say, complex) inner product spaces the definition would be slightly different.

Theorem 7.1: Properties of the inner product

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in an inner product space V and let c be a scalar.

(a) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$.

(b) $\langle \mathbf{u}, c\mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$.

(c) $\langle \mathbf{u}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle = 0$.

We can now define such notions as length, distance, and orthogonality in abstract vector spaces. The construction of all these objects is very similar to what we did with Euclidean spaces, so we won't rehash all the details.

Definition: Geometry in inner product spaces

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V.

- The length (or norm) of $\mathbf{v}$ is $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. (A unit vector is a vector of length 1.)
- The distance between $\mathbf{u}$ and $\mathbf{v}$ is $\mathrm{d}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.
- $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Theorem 7.2: Pythagorean theorem

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Definition: Orthogonal projection

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for a subspace W of an inner product space V, and let $\mathbf{v}$ be a vector in V. Then the orthogonal projection $\mathrm{proj}_W(\mathbf{v})$ of $\mathbf{v}$ onto W is

$$\mathrm{proj}_W(\mathbf{v}) = \frac{\langle \mathbf{u}_1, \mathbf{v}\rangle}{\langle \mathbf{u}_1, \mathbf{u}_1\rangle}\mathbf{u}_1 + \cdots + \frac{\langle \mathbf{u}_k, \mathbf{v}\rangle}{\langle \mathbf{u}_k, \mathbf{u}_k\rangle}.$$

The component of $\mathbf{v}$ orthogonal to W is the vector $\mathrm{perp}_W(\mathbf{v}) = \mathbf{v} - \mathrm{proj}_W(\mathbf{v})$.

Theorem 7.3: Cauchy-Schwarz inequality

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space v. Then

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|,$$

with equality holding if and only if $\mathbf{u}$ and $\mathbf{v}$ are scalar multiples of each other.

Theorem 7.4: Triangle inequality

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. Then

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

7.2 Least Squares Approximation

We can use this theory to find a "line of best fit" for a given set of data. We'll begin by rephrasing this search in terms of vectors.

Definition: Best approximation of a vector

If W is a subspace of a normed vector space V and if $\mathbf{v}$ is a vector in V, then the best approximation to $\mathbf{v}$ in W is the vector $\overline{\mathbf{v}}$ in W such that

$$\|\mathbf{v} - \overline{\mathbf{v}}\| < \|\mathbf{v} - \mathbf{w}\|$$

for every vector $\mathbf{w}$ in W different from $\overline{\mathbf{v}}$.

Provided that V is equipped with an inner product, our problem has a simple solution.

Theorem 7.5: Best approximation theorem

If W is a finite-dimensional subspace of an inner product space V and if $\mathbf{v}$ is a vector in V, then $\mathrm{proj}_W(\mathbf{v})$ is the best approximation to $\mathbf{v}$ in W.

We can use this pair of ideas to determine approximate solutions to inconsistent linear systems.

Definition: Least squares solution

If A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$, a least squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\overline{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\overline{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

Note that $\mathbf{b} - A\overline{\mathbf{x}} = \mathrm{perp}_{\mathrm{col}(A)}(\mathbf{b})$. Therefore, $\mathbf{b} - A\overline{\mathbf{x}}$ is in $(\mathrm{col}(A))^\perp = \mathrm{null}(A^T)$, and we have the following.

Theorem 7.6: Least squares theorem

Let A be an $m \times n$ matrix and let $\mathbf{b}$ be in $\mathbb{R}^m$. Then $A\mathbf{x} = \mathbf{b}$ always has at least one least squares solution $\overline{\mathbf{x}}$. Moreover:

(a) $\overline{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\overline{\mathbf{x}}$ is a solution of the normal equations $A^T A\overline{\mathbf{x}} = A^T\mathbf{b}$.

(b) A has linearly independent columns if and only if $A^T A$ is invertible. In this case, the least squares solution of $A\mathbf{x} = \mathbf{b}$ is unique and is given by

$$\overline{\mathbf{x}} = (A^T A)^{-1}A^T\mathbf{b}.$$

Suppose we have a set of points of the form (x_i, y_i), and we want to find a line of best fit $y = a + bx$ through them. To do this, we use the above theorem to find a least-squares solution to the linear system

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Computing $A^T A$ may take some time, and solving the resulting system of equations may take even longer. However, if we have a QR factorization of A, our job is made much easier.

Theorem 7.7: Least squares via QR factorization

Let A be an $m \times n$ matrix with linearly independent columns and let $\mathbf{b}$ be in $\mathbb{R}^m$. If $A = QR$ is a QR factorization of A, then the unique least square solution $\overline{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{v}$ is

$$\overline{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.$$

7.3 The Singular Value Decomposition

In this final section, we will briefly investigate one last matrix decomposition. These have to do with the singular values of a matrix, defined below.

Definition: Singular values

If A is an $m \times n$ matrix, the singular values of A are the square roots of the eigenvalues of $A^T A$ and are denoted by $\sigma_1, \ldots, \sigma_n$. It is conventional to arrange the singular values so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

If the eigenvalues of a matrix have to do with a lack of motion under a linear transformation, the singular values of a matrix quantifies the maximum and minimum motions of space under a linear transformation. We can use these singular values, along with the matrix's eigenvectors, to construct a relatively straightforward decomposition.

Theorem 7.8: Singular value decomposition

Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ matrix Σ such that

$$A = U\Sigma V^T.$$

The matrix Σ contains the ordered singular values of A along its main diagonal. The matrices other V

and U are of the following forms:

$$V = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$$
$$U = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_m\end{bmatrix}$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are the eigenvectors of $A^T A$, $\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$ for $i \leq r$, and $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ are vectors added to the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to make it a basis for $\mathbb{R}^m$.

We can now make one last addition to the Fundamental Theorem.

Theorem 7.9: Additions to the FTLA

Let A be an $n \times n$ matrix and let $T : V \to V$ be a linear transformation whose matrix $[T]_{\mathcal{C} \leftarrow \mathcal{B}}$ with respect to bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively, is A. The following statements are equivalent.

(a) A is invertible.

$\vdots$

(u) 0 is not a singular value of A.

For the full theorem, see Appendix A.

A The Fundamental Theorem of Linear Algebra

Let A be an $n \times n$ matrix and let $T : V \to V$ be a linear transformation whose matrix $[T]_{\mathcal{C} \leftarrow \mathcal{B}}$ with respect to bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively, is A. The following statements are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

(c) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(d) The reduced row echelon form of A is I_n.

(e) A is a product of elementary matrices.

(f) $\mathrm{rank}(A) = n$

(g) $\mathrm{nullity}(A) = 0$

(h) The column vectors of A are linearly independent.

(i) The column vectors of A span $\mathbb{R}^n$.

(j) The column vectors of A form a basis for $\mathbb{R}^n$.

(k) The row vectors of A are linearly independent.

(l) The row vectors of A span $\mathbb{R}^n$.

(m) The row vectors of A form a basis for $\mathbb{R}^n$.

(n) $\det A \neq 0$

(o) 0 is not an eigenvalue of A.

(p) T is invertible.

(q) T is one-to-one.

(r) T is onto.

(s) $\ker(T) = \{\mathbf{0}\}$.

(t) $\mathrm{range}(T) = W$.

(u) 0 is not a singular value of A.