Linear Algebra

Dhruva Karkada

Spring 2018

Contents

1 Vector Spaces

Consider some objects, which we'll call vectors, on which we define two operations: addition between two vectors, and multiplication of a vector with a scalar. In this document, "vector" most commonly refers to a column of scalar quantities, but in a broad sense, a vector can be other mathematical objects, such as matrices, polynomials, and real-valued functions.

A vector space V is a set of vectors, subject to some definition of addition and scalar multiplication, that satisfies the following properties:

1. Closure. Adding any two vectors in the set yields another vector in the set; multiplying any vector in the set with any scalar yields another vector in the set.

2. Existence of the zero vector $\mathbf{0}$. This vector satisfies $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for any vector $\mathbf{v} \in V$.

3. Existence of additive inverse. This vector satisfies $\mathbf{v} + (-\mathbf{v}) = 0$ for any vector $\mathbf{v} \in V$.

4. Commutative and Associative properties for vector addition.

5. $1 \cdot \mathbf{u} = \mathbf{u}$

6. Commutative and Associative properties for scalar multiplication.

7. Distributive properties (scalar multiplied over vector sum, and vector multiplied over scalar sum).

Some well-known vector spaces are $\mathbb{R}^n$ (which is the vector space we'll focus on here), the set of $m \times n$ matrices, $\mathbb{P}^n$ (the set of polynomials with deg$\leq n$), and the set of real-valued functions.

A subspace of a vector space is a subset of a vector space which satisfies two properties: closure within the subspace, and the existence of the zero vector. An important result is that any subspace of a vector space is also a vector space.

This is great and all, but a good visual always helps the abstraction. Let's consider the vector space $\mathbb{R}^3$, where every vector is a column of 3 scalar quantities. We can visualize this associating each quantity with a coordinate axis; the result is a point in $\mathbb{R}^3$, which we often think of as an arrow from the origin to that point. Vector addition can be visualized as head-to-tail addition, and scalar multiplication just scales any arrow vector.

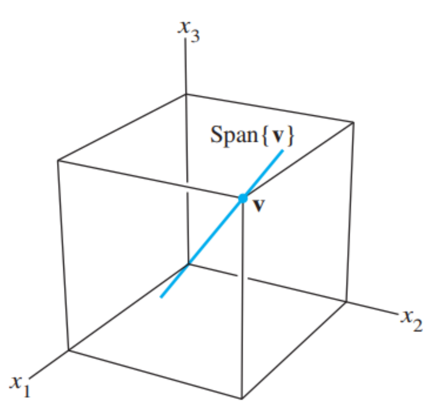
Clearly, the vector space is closed; when you add two vectors in 3-dimensional space, you get another vector in 3-dimensional space. Same goes with scaling a vector. The zero vector is the vector of zero length. The other properties can be easily verified.

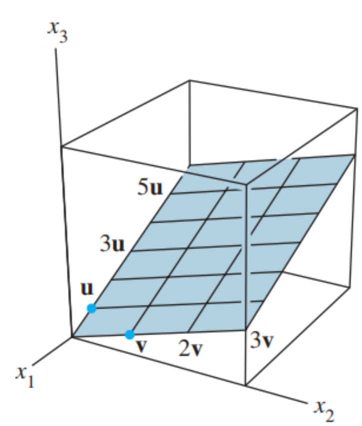
What about subspaces? We need a set of vectors that contains the zero vector (i.e. contains the origin of $\mathbb{R}^3$) and is closed. This corresponds to any plane or line which passes through the origin. One can verify that such subsets of $\mathbb{R}^3$ are closed, and satisfy the other properties of vector spaces.

1.1 Span and Linear Independence

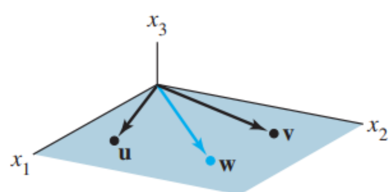
Let's start with one nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$. If we consider all the scaled versions of this vector, we get a line. We say that this vector *spans* this line. Notice that this line is a subspace of $\mathbb{R}^3$.

Now let's consider an additional vector $\mathbf{u}$, and all the *linear combinations* $c_1\mathbf{v} + c_2\mathbf{u}$ of the two vectors. What we're doing is adding all possible sums between all possible combinations of scalings of our two vectors, and considering the resulting set of vectors. Here, we have two cases: if $\mathbf{u}$ is on the line we considered originally, then the new set of vectors is just that same line. However, if $\mathbf{u}$ is NOT on the line, then the new set is actually expanded; the linear combinations now span a plane. This plane is also a subspace of $\mathbb{R}^3$.

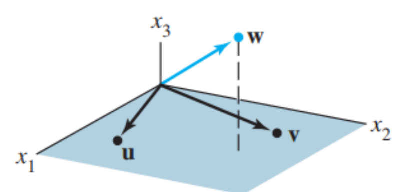




Now, we can repeat the process with a third vector $\mathbf{w}$, and consider their linear combinations $c_1\mathbf{v} + c_2\mathbf{u} + c_3\mathbf{w}$. Again we have two cases. If $\mathbf{w}$ lies on the same plane, then the new span of the three vectors is the same plane. However, if it's NOT on the plane, then the new set expands yet again, and the vectors span the full space of $\mathbb{R}^3$. Adding more vectors to our linear combination cannot further extend the span; we've already covered the whole vector space.



Linearly dependent,
w in Span{**u**, **v**}



Linearly independent,
w *not* in Span{**u**, **v**}

We've seen that *span* describes all possible linear combinations of a set of vectors. We use the term *linear independence* to describe a set of vectors that don't have any "redundant" vectors (like in the examples where the new vector was already within the old line/plane). If a set of vectors contains a vector which is already within the span of the other vectors, then that set is *linearly dependent*. Another way to test linear dependence is to ask: can any of the vectors in this set be expressed as a sum of the other vectors? If yes, then that vector is redundant, or linearly dependent on the other vectors. Formally, a set of n vectors is linearly dependent iff there are scalars $c_1 \dots c_n$, not all zero, such that $c_1 v_1 + \dots + c_n v_n = 0$.

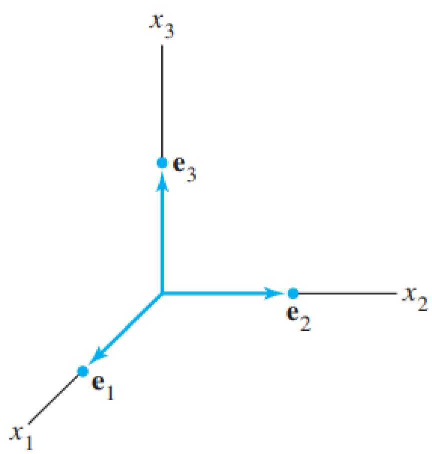
1.2 Basis and Dimension

Let's push this idea a bit further. Let's consider again our three vectors which span the full $\mathbb{R}^3$. Notice that removing any vector will necessarily decrease the span; adding any vector will necessarily make our linearly independent set linearly dependent. Since these three vectors are linearly independent and span R^3, any other vector in $\mathbb{R}^3$ can be written as a unique linear combination of these three vectors (this follows from the definition of span). Because this set of vectors has this property, we say that these vectors form a *basis* of $\mathbb{R}^3$. Notice that being a basis isn't all that special; if you pick any three vectors from $\mathbb{R}^3$, they're pretty likely to be linearly independent and span $\mathbb{R}^3$, and thus form a basis.

However, there is one “best” basis: the basis that is the simplest. We call this the standard basis, and in $\mathbb{R}^3$ the standard basis vectors are

$$\hat{\imath} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \hat{\jmath} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \hat{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

One can easily check that you can write any vector in $\mathbb{R}^3$ as some linear combination of $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$. In fact, the coefficients for each basis vector will exactly correspond with the scalar entries in the target vector.



We've seen that any basis of $\mathbb{R}^3$ has 3 vectors. Is it true that $\mathbb{R}^n$ has n basis vectors? Yep. The number of vectors needed to form a basis for any vector space is the *dimension* of the vector space. So, $\mathbb{R}^3$ is 3-dimensional, $\mathbb{R}^n$ is n-dimensional, and this allows us to extend the concept of dimension to a wide variety of other vector spaces.

It turns out that any vector space V of dimension n is *isomorphic* with $\mathbb{R}^n$. This means that they have the same algebraic structure; we can think of vectors in V as points in $\mathbb{R}^n$, and the math is exactly the same. We just have to remember that the points actually represent something else. This is useful because we can visualize operations in abstract vector spaces as operations in $\mathbb{R}^n$, which is much easier.

Some summarizing theorems:

1. Suppose vector space V with $\dim V = p$, and a set of vectors in V, $S = \{v_1 \dots v_p\}$. Then

$$S \text{ is linearly independent} \leftrightarrow S \text{ is basis of } V \leftrightarrow S \text{ spans } V.$$

   As a corollary, if the length of $S > p$ then the set must be linearly dependent. If the length of $S < p$ it cannot span V. Equivalently, any basis of V is the smallest spanning set, or the largest linearly independent set.

2. Any set of vectors that span V has a subset which is a basis of V. (Not necessarily strict subset.)

2 Linear Transformations

A transformation between vector spaces is a function which turns vectors into other vectors. Specifically, a transformation $T : A \mapsto B$ takes as input vectors from A and produces vectors in B. In this case A is the domain, B is the codomain. It is possible that the set of all possible outputs of T doesn't encompass all of B; so we call the set of output vectors the range of T. The range is always a subset of the codomain (not necessarily strict).

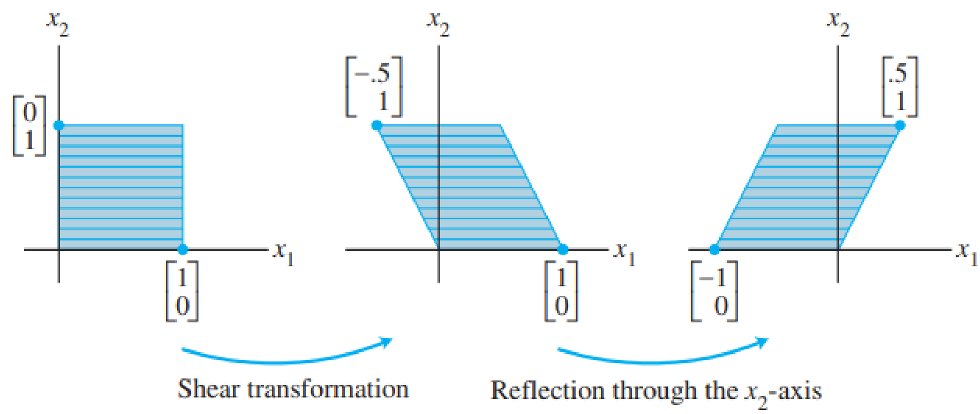
A *linear transformation* T satisfies the following properties:

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

Linear algebra deals exclusively with linear transformations.

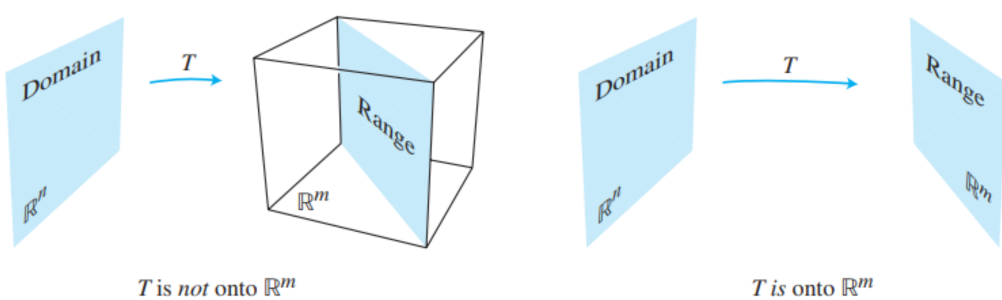
Let's suppose we're working with $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$. Each vector in R^2 is taken to another vector, and we can visualize this as a morphing of the space itself. Some transformations may stretch vectors out; we can visualize this as the gridlines getting stretched while the coordinate axes stay the same. Some transformations may rotate vectors around the origin; we can visualize this as the gridlines rotating around while the coordinate axes stay the same. Some transformations may take every vector to a single point or line; we can visualize this as the gridlines collapsing to that point or line while the coordinate axes stay the same.

Linear transformations place a special restriction: when we morph the grid lines, the origin must stay fixed, and parallel gridlines must remain parallel. Equivalently, if T is linear, then $T(\mathbf{0}) = \mathbf{0}$, and T maps lines to lines.

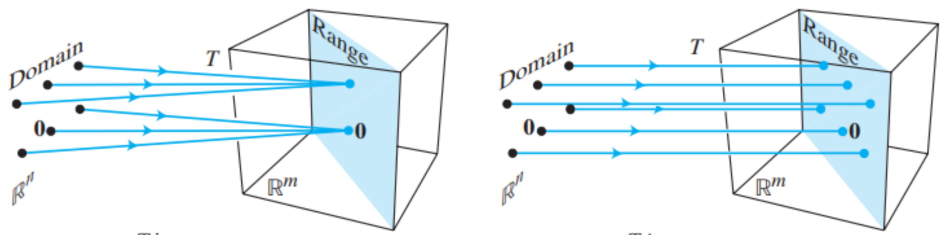


It's a little harder to visualize transformations between vector spaces of different dimension. Instead, it's easier to just visualize the domain and range of the transformation. For example, a linear transformation such as $T : \mathbb{R}^2 \mapsto \mathbb{R}^3$ can map $\mathbb{R}^2$ to a plane in $\mathbb{R}^3$ (or a line, or a point). Importantly, the range of a linear transformation is always a subspace of the codomain, so the plane/line/point must contain the origin.

Since $\mathbb{R}^3$ is larger than $\mathbb{R}^2$, there is no way for the range of T to encompass all of the codomain. The largest subspace T can span is a plane in $\mathbb{R}^3$. So, T is not *onto*; its range is not equal to its codomain.



However, it can be *one-to-one*: this occurs when there is a one-to-one correspondence between the domain and range; every vector in the range is mapped to by only one vector in the domain (of course since T is a function, every element in the domain maps to only one element in the range). So when the range is a plane, T is one-to-one, but if the range is a line or a point, T is not. T can never be onto.



However, if we consider another transformation $L : \mathbb{R}^2 \mapsto \mathbb{R}$, then we can see that L can be onto, but can never be one-to-one. Since $\mathbb{R}$ is smaller than $\mathbb{R}^2$, there's no way to map all of a 2-dimensional space into a cramped one-dimensional space without overlap. In the case that L maps every vector in the domain to a single vector in the codomain, L is neither onto nor one-to-one.

Formally, a one-to-one transformation is called an injection, and an onto transformation is called a surjection. A transformation which is both onto and one-to-one is called a bijection; this can only happen if the domain and the codomain have the same dimension (and even this doesn't guarantee a bijection). Bijections are very important in linear algebra.

2.1 Matrix Representation

So we have a visual picture of linear transformations; what about a mathematical representation? This is where matrices come in. Let's consider a transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^m$. Since T is linear, all the information of the transformation is encoded in how the T affects a basis. The details of how T acts on any other vector can be found by expressing the vector as a linear combination of basis vectors, transforming each of them separately, and adding the weighted results (which is valid because of the linearity of T).

Conventionally, we use the standard basis. So now we find out how T transforms the standard basis vectors. There are n basis vectors for $\mathbb{R}^n$, and they are each transformed into vectors with m scalar elements (since vectors in the codomain have m coordinates). If we represent each transformed vector as a column vector, and put the vectors side-by-side in order of the standard basis, and smush them together, we'll end up with a grid of numbers with height m and width n. We call this an $m \times n$ matrix, and this matrix represents the transformation T.

This matrix acts on a vector in $\mathbb{R}^n$ by multiplication. Note that multiplying an $m \times n$ matrix by a vector is only defined if the vector has n elements. In the case that $n = 2$ and $m = 3$, a possible transformation is

$$\begin{bmatrix} 1 & 4 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix}.$$

This definition of matrix multiplication makes sense given the explanation of matrices given above; we're just constructing the resulting vector from the linear combination of transformed basis vectors. Notice that the input vector is in $\mathbb{R}^2$ and the resulting vector is in $\mathbb{R}^3$, just as we expected.

When we consider a linear transformation acting on a vector, we often express it as

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

Clearly, $\mathbf{x}$ is from the domain and $\mathbf{b}$ is from the range (which is in the codomain). Since $\mathbf{b}$ must always be a linear combination of columns of A (with the weights being given by $\mathbf{x}$), the columns of A must span the range of the linear transformation. For this reason, the range (which, as mentioned, is a subspace of the codomain) is often also called the *column space* of A, denoted col A. The columns of A are not necessarily a basis of the column space, since the columns of A are not necessarily linearly independent.

2.2 Composition

We can also multiply two matrices, which has the effect of composing two transformations. For example, if matrix A represents a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and matrix B represents a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^p$, then the matrix product BA represents the composition of the two transformations. Notice that the codomain of one transformation must be the domain of other transformation. In terms of matrices, this means that the width of the first matrix must equal the height of the second.

If the column vectors of A are denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_n$, then

$$BA = \begin{bmatrix} B\mathbf{a}_1 & B\mathbf{a}_2 & \cdots & B\mathbf{a}_n \end{bmatrix}$$

which is a $p \times n$ matrix as expected. This multiplication formula follows from the composition interpretation of matrix multiplication. Some important caveats: $AB \neq BA$; $AB = AC$ does not imply $B = C$; and $AB = 0$ does not imply either $A = 0$ or $B = 0$.

3 Linear Systems as Matrix Equations

Consider the following matrix equation in the form $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 4 & 3 \\ 2 & 0 & -2 \\ 1 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix}.$$

When you expand this, you get the following system of equations:

$$\begin{aligned} x_1 + 4x_2 + 3x_3 &= 6 \\ 2x_1 - 2x_3 &= -2 \\ x_1 + x_2 + 6x_3 &= 3. \end{aligned}$$

So matrix equations can be used to represent and solve systems of linear equations. By performing row operations on the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$, we can get the matrix into *row-echelon form*, which allows us to solve for each variable. This technique won't be explained here, but the idea is that the final matrix will

- Represent a system of equations with the same solution set as the original matrix (even though other properties such as the column space may not be the same)
- Be arranged in such a way that zeroes accumulate at the bottom left corner
- Have *pivot columns* and *free variables*.

After getting the augmented matrix into echelon form, we have three cases, which correspond to different statements about the existence and uniqueness of solutions:

- There is a zeroed row: a row with all zeros except in the augmented column (containing the row-reduced $\mathbf{b}$). Then there is no solution. The zeroed row yields an equation of the form $0x_1 + \cdots + 0x_n = b_n$ which is unsatisfiable for any values of $x_1, \ldots, x_n$. Because there is no solution, the system is *inconsistent*.

- There are no free variables and no zeroed row; every column in A has a pivot. Then there is a single solution to the system, which is easily found by converting the echelon form into equations. The system is *consistent*.

- There are free variables and no zeroed row; not every column in A has a pivot. Then there are infinitely many solutions, and the system is consistent. The general solution is written as a linear combination of vectors, which can be found by converting the echelon form into equations, solving for the pivot variables in terms of the free variables, and factoring out the free variables as undetermined coefficients for the resulting set of vectors.

Keep in mind that we're using matrices here to represent systems of linear equations. Nonetheless, the notation is the same as for linear transformations between vector spaces: $A\mathbf{x} = \mathbf{b}$. So, even this equation represents different phenomena, the underlying algebra turns out to be the same.

4 Existence and Uniqueness Questions

Questions about existence and uniqueness are important in linear algebra – both in solving systems, and in analyzing transformations. For example: given coefficients in matrix A, is $\mathbf{x} = \mathbf{0}$ the only solution to $A\mathbf{x} = \mathbf{0}$? If not, what are the others? This is a uniqueness question. Or, given A, does every possible $\mathbf{b}$ have a solution $\mathbf{x}$ that satisfies $A\mathbf{x} = \mathbf{b}$? This is an existence question. We'll discuss both of these questions.

4.1 The Homogeneous Equation

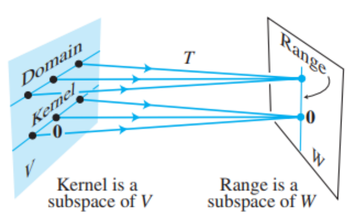
It is sometimes useful to solve the system $A\mathbf{x} = \mathbf{0}$, known as the homogeneous equation (since the right side is $\mathbf{0}$). Of course $\mathbf{x} = \mathbf{0}$ is a solution; this is called the *trivial solution*. We're looking for nontrivial solutions.

We solve this using the same method described in the previous section. Since the augmented column is comprised entirely of zeros, the row operations don't change it at all. Therefore, we're safe from the first solution case (zeroed row); as mentioned, there will always be a (trivial) solution. We want case three: infinitely many solutions (only one of which is trivial). This arises when there's at least one free variable in A.

When we expand the matrix equation with a nontrivial solution, we end up with a set of vectors whose sum (coefficients not all zero) is $\mathbf{0}$. This is exactly the formal definition for linear dependence. Therefore, if we have nontrivial solutions, then the columns of A are linearly dependent. The inverse is also true: if we have only the trivial solution, then the columns of A are linearly independent.

In terms of linear transformations between vector spaces, solutions to the homogeneous equation represent vectors which are taken to $\mathbf{0}$ in the codomain. Since it's linear, we know that $\mathbf{0}$ maps to $\mathbf{0}$; this is just another rehash of the idea that there's always a trivial solution. What we're looking for is some subspace of vectors in the domain which collapse to the origin in the codomain. In other words, we don't want the transformation to be one-to-one; if it is, then we know that we have only the trivial solution (and vice versa). In fact, the inverse is true; any linear transformation which is NOT one-to-one must have nontrivial solutions (infinitely many).

It turns out that any solution set to $A\mathbf{x} = \mathbf{0}$ actually comprises a subspace in the domain; for this reason, we call the general solution the *null space* (or *kernel*) of A, denoted nul A. The null space always contains $\mathbf{0}$; if the dimension of nul A is greater than 0, then we have nontrivial solutions (and vice versa). In fact, dim nul A is given by the number of free variables in A.



Recall that we said that generally, since col A is a subspace of the codomain, we don't know its basis. If the column vectors are all linearly independent, then they form a basis; but if they aren't, then we have some redundant columns. It turns out that the free variable columns are these redundant columns; if we take the columns of A which have pivots, these columns form a basis of col A.

This leads us to a crucial relationship between the columns of A. There are n columns, each of which is either a pivot column or free variable column, corresponding to the column space and null space respectively. So

$$\text{dim col } A + \text{dim nul } A = n.$$

Additionally, since each column of A represents a transformed basis vector, n is the dimension of the domain. This is known as the *rank theorem*, since the dimension of col A is sometimes called the rank of A. In terms of transformations, it basically says that the dimension of the range plus the dimension of the vectors which collapse to $\mathbf{0}$ must equal the dimension of the domain (which makes intuitive sense).

So, we've seen that the following are logically equivalent:

- $A\mathbf{x} = \mathbf{0}$ has no nontrivial solutions
- A has no free variables (i.e. every column has pivot)
- Columns of A are linearly independent.
- The transformation encoded by A is one-to-one
- dim nul $A = 0$

4.2 The Inhomogeneous Equation

What about every other $\mathbf{b}$ in the codomain? Is $A\mathbf{x} = \mathbf{b}$ consistent? If the system is consistent for every $\mathbf{b}$, then we know that the range of A coincides with the codomain, and the transformation is onto. Since the columns of A span the range of A, we can say that the columns of A span the entirety of the codomain.

The question then becomes: how can we determine if the range (a.k.a. column space) of A coincides with the codomain? It turns out A is onto iff

A has a pivot in every row (contrast with the previous subsection, where we checked if A has a pivot in every column).

So this answers the question of existence for all $\mathbf{b}$. Now let's consider the uniqueness of solutions for a single $\mathbf{b}$ (assuming there is a solution). We have two cases:

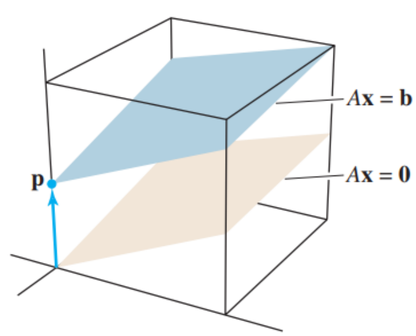
- There is a single solution. This happens when dim nul $A = 0$. The reasoning is explained shortly.

- There are infinitely many solutions. This happens when dim nul $A > 0$. Then the general solution is given by

$$\mathbf{x}_{gen} = \mathbf{x}_p + \mathbf{x}_h$$

  where $\mathbf{x}_p$ is a *particular solution* (any solution to the inhomogeneous equation), and $\mathbf{x}_h$ is the general solution to the associated homogeneous equation $A\mathbf{x} = \mathbf{0}$. This relationship is especially important in the study of differential equations.

So, case 1 is really just a special case of the result in case 2, where $\mathbf{x}_h = \mathbf{0}$. In fact, there is a visual representation of this idea: the solutions of $A\mathbf{x} = \mathbf{b}$ are merely the solutions of $A\mathbf{x} = \mathbf{0}$, translated by a single vector (the particular solution). When we use row reduction on the augmented matrix to find the general solution, the reduced augmented column gives a particular solution.



So, we've seen that the following are logically equivalent:

- $A\mathbf{x} = \mathbf{b}$ has solutions for all $\mathbf{b}$ in the codomain
- A has pivots in every row
- Columns of A span the codomain
- The transformation encoded by A is onto

5 Inverses, Determinants, and Transposes

Earlier, we said that bijections are important in linear algebra. If a function is both onto and one-to-one, then we know that it has an inverse function which undoes the effect of the function. In other words, if $y = f(x)$, then $x = f^{-1}(y)$ and therefore $x = f^{-1}(f(x))$.

Similar logic applies to linear transformations which are bijective. If a matrix A represents a bijective transformation, then there exists a matrix A^{-1} which the property that $\mathbf{x} = A^{-1}A\mathbf{x}$; we call A *invertible.* Invertible matrices must be square, and their inverses are also square; the product $A^{-1}A = I$ the identity matrix, which is the following square matrix:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

The diagonal of a square matrix refers to the diagonal from the top left to the bottom right; so the identity matrix is a matrix whose diagonal entries are 1 and non-diagonal entries are 0. Clearly, the identity matrix represents a transformation which does not change vectors.

If we're solving $A\mathbf{x} = \mathbf{b}$ and A is invertible, then we can left-multiply both sides by the inverse to easily get the solution $\mathbf{x} = A^{-1}\mathbf{b}$. The question now becomes, how do we generate A^{-1} from A? This requires a useful number called the *determinant* of the matrix.

5.1 Determinants

The determinant is defined only for square matrices, and has some useful properties. For example, if a square matrix A has $\det A = 0$, then we know that the matrix is not invertible. In addition, $|\det A|$ gives the volume of the parallelopiped whose edges are given by the columns of A. Importantly, the determinant helps us find the inverse of A (if it exists).

The determinant is given by a recursive process called cofactor expansion. We pick any row or column of A, and for each entry in chosen row, we take a sum of products:

$$\det A = \sum_{\text{row}} a_{ij} C_{ij} \quad \text{where} \quad C_{ij} = \pm \det A_{\cancel{ij}}.$$

Here, a_{ij} refers to the entry in the chosen row/column, and C_{ij} refers to the cofactor at i, j. The cofactor is found by finding the determinant of the remaining matrix after you remove the i^{th} row and j^{th} column, and then applying a sign corresponding to the "checkerboard rule": positive when $i+j$ is even, negative when odd. This creates a checkboard pattern in the signs. Because you need to find a smaller determinant in order to calculate the big determinant, this process is often recursive and tedious. For this reason, we try to pick rows and columns which have many zeros, so the product automatically vanishes.

In some lucky cases, we have what's called a triangular matrix: every entry below (or above) the diagonal is zero. In this case, the cofactor expansion simplifies dramatically, and the determinant of A is just the product of the diagonal entries. If a diagonal entry of a triangular matrix is zero, then $\det A = 0$ and A is not invertible.

We may want to perform some row operations to get a matrix into triangular form, and then use the determinant shortcut. The row operations, transforming from matrix A to B, have the following effect on the determinant:

- Multiply a row by k: $\det B = k \det A$
- Add a multiple of another row: $\det B = \det A$
- Interchange two rows: $\det B = -\det A$

5.2 Finding Inverses

There are two main ways to find the inverse of a matrix: the adjugate formula, and row reduction of an augmented matrix. The adjugate formula is based on a result known as Cramer's rule, while row reduction is generally faster (especially if working by hand).

Cramer's rule tells us that the entries of $\mathbf{x}$ for $A\mathbf{x} = \mathbf{b}$ is given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}$$

where $A_i(\mathbf{b})$ is the matrix resulting when you replace the i^{th} column of A with $\mathbf{b}$. By manipulating the definition $\mathbf{A_j^{-1}} = A^{-1}\mathbf{e}_j$ (where $\mathbf{e}_j$ is the j^{th}

standard basis vector), we know that $A\mathbf{A_j^{-1}} = \mathbf{e}_j$. Then, using Cramer's rule, we can find the i, j element of A^{-1}:

$$A_{ij}^{-1} = \frac{\det A_i(\mathbf{e}_j)}{\det A} = \frac{C_{ji}}{\det A}$$

where C_{ji} is the j, i cofactor as described above. Since the indices are reversed, we'll have to flip everything over the diagonal in the end (called a transpose). This leads us to the adjugate formula:

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{1,1} & \dots & C_{1,n} \\ \vdots & \ddots & \vdots \\ C_{n,1} & \dots & C_{n,n} \end{bmatrix}^T$$

where the T represents the transpose. The transpose of the cofactor matrix is called the adjugate of A. For 2×2 matrices, this yields the simple formula

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

The row reduction method is a bit more straightforward. We create the $n \times 2n$ matrix $[A\ I]$, and row reduce the left side down to I. The operations that bring A to I are the same operations which bring I to A^{-1}; therefore, the final matrix is $[I\ A^{-1}]$, yielding the inverse.

5.3 Transposes

The transpose of a matrix A is the matrix whose rows are the columns of A. This is equivalent to a diagonal reflection. The product of a matrix and its transpose is always square, even if the original matrix is not.

Some important properties of determinants, inverses, and transposes are shown below.

Determinant	Inverse	Transpose
$\det AB = \det A \det B$	$(AB)^{-1} = B^{-1}A^{-1}$	$(AB)^T = B^T A^T$
$\det(rA) = r^n \det A$	$(rA)^{-1} = \frac{1}{r}A^{-1}$	$(rA)^T = rA^T$
$\det A^T = \det A$	$(A^T)^{-1} = (A^{-1})^T$	$(A^T)^T = A$

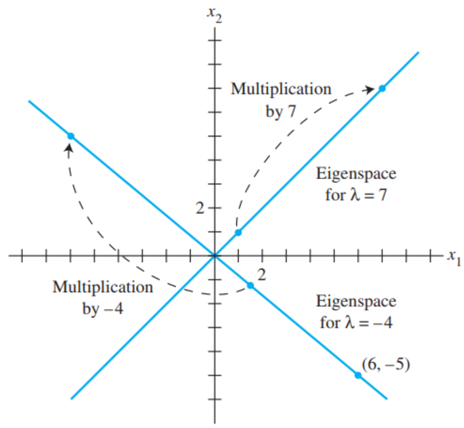
5.4 Invertible Matrix Theorem

Some of the most fundamental results of linear algebra are summarized by the invertible matrix theorem, which relates properties of invertible matrices and bijective transformations. The theorem states that if A is an $n \times n$ matrix, all the following statements are equivalent (either they're all true or all false):

- A is invertible
- A can be row-reduced to I
- A has n pivots
- $A\mathbf{x} = \mathbf{0}$ has only the trivial solution
  - dim col $A = n$
  - dim nul $A = 0$
  - Columns are linearly independent
  - One-to-one
- $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{R}^n$
  - Columns span $\mathbb{R}^n$
  - Onto
- $\det A \neq 0$
- 0 is not an eigenvalue

6 Eigenvalues and Eigenvectors

The linear transformations associated with square matrices can be visualized as a transformation which takes vectors in a space and applies some combination of rotations (around the origin) and stretches. However, for many such transformations, there are a few special vectors which do not get rotated; they are only scaled by the transformation. Consider a horizontal shear; most vectors are rotated towards from the x-axis; but vectors which are already on the x-axis are only stretched.



For some given square matrix, the vectors which are only stretched are called *eigenvectors*, while an eigenvector's scaling factor is called its *eigenvalue*, typically denoted λ. The zero vector is trivial and is therefore never considered an eigenvector. By their definitions, to find eigenvectors and eigenvalues, we want to find pairs which satisfy

$$A\mathbf{x} = \lambda\mathbf{x}.$$

After some manipulation, we can get this equation to the form $(A - \lambda I)\mathbf{x} = \mathbf{0}$. Unfortunately, we still only have one equation and two unknowns. Thankfully, we can use previous ideas here. If we want nontrivial $\mathbf{x}$ for square $(A - \lambda I)$, then this new matrix has to have free variables; in other words, it cannot be invertible. By the invertible matrix theorem, this means that the determinant of $(A - \lambda I)$ should be 0.

So, we first solve for λ by solving

$$\det(A - \lambda I) = 0$$

which yields a polynomial of degree n, called the *characteristic polynomial*[1]. We've converted a problem about eigenvalues and linear transformations to a problem about finding real roots of polynomials. By the fundamental theorem of algebra, we are guaranteed n roots, but some or all of them could be complex, and we're only interested in real roots. For example, consider a transformation on $\mathbb{R}^2$ which rotates every vector 90° clockwise. Visualizing,

[1]If the matrix is triangular, then the eigenvalues are just the diagonal entries.

we can see that there are no eigenvectors; this is because the characteristic polynomial has two non-real roots.

The roots of the characteristic polynomial are our eigenvalues; each eigenvalue can have multiplicities as well. Finally, we find the eigenvectors for each λ by solving the matrix equation. For a real-valued A, the number of linearly dependent eigenvectors corresponding to each λ is $\leq$ the multiplicity of λ, and span the *eigenspace* of that particular λ. Furthermore, a theorem tells us that eigenvectors from different λ are guaranteed to be linearly independent. Therefore, we should end up with $\leq n$ linearly independent eigenvectors. When we have exactly n linearly independent eigenvectors, A is diagonalizable; this is covered later.

7 Orthogonality

Let's consider the familiar vector space $\mathbb{R}^n$. The familiar dot product gives us some useful information about vectors; for example, $\mathbf{v} \cdot \mathbf{v} = |\mathbf{v}|^2$, where $|\mathbf{v}|$ represents the length, or norm, of the vector. We can use the same definition to calculate distances between vectors as well. Importantly, dot products also encode information about orientation; if two vectors have a dot product of zero, they are necessarily perpendicular, or *orthogonal*.

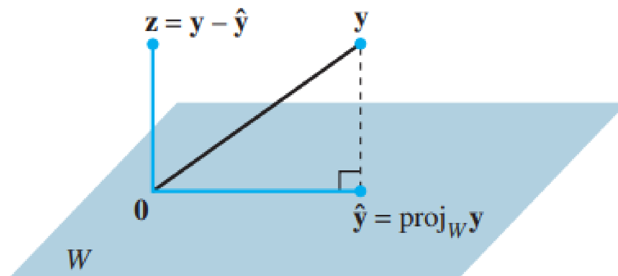
7.1 Least-Squares Approximations

The idea of orthogonality is important in both $\mathbb{R}^n$ and other vector spaces. For example, orthogonality implies linear independence, so a set of orthogonal vectors which span a subspace are necessarily a basis.

If we have some subspace W with orthogonal basis vectors $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$, then the set of all vectors which are orthogonal to all vectors in B span the "orthogonal complement" of W, which is a subspace denoted $W^\perp$. These mutually exclusive subspaces comprise the entirety of the vector space.

Now suppose we have a vector $\mathbf{y}$ which is in V but not in W. Now we want to find the vector in W which is closest to $\mathbf{y}$; in other words, we want to find the best approximation of $\mathbf{y}$ in W. The idea of distance is given by the dot product (more generally, the inner product). It turns out that we can get this approximation by "projecting" our vector onto W; the difference between the projection and the actual vector is a vector in $W^\perp$ which is orthogonal to the projection. Moreover, the projection of $\mathbf{y}$ onto W is the sum of the projections of $\mathbf{y}$ on the orthogonal basis vectors of W:

$$\hat{\mathbf{y}} \in W = \frac{\mathbf{y} \cdot \mathbf{b}_1}{\mathbf{b}_1 \cdot \mathbf{b}_1}\mathbf{b}_1 + \frac{\mathbf{y} \cdot \mathbf{b}_2}{\mathbf{b}_2 \cdot \mathbf{b}_2}\mathbf{b}_2 + \cdots + \frac{\mathbf{y} \cdot \mathbf{b}_m}{\mathbf{b}_m \cdot \mathbf{b}_m}\mathbf{b}_m$$



This idea that a projection represents the best approximation is useful in solving some systems of equations. Consider an inconsistent system $A\mathbf{x} = \mathbf{b}$. There is no solution because the column space of A does not contain $\mathbf{b}$. So, we want to approximate the solution; which vector in col A is closest to $\mathbf{b}$? More importantly, which $\mathbf{x}$ will transform into this approximation? This is known as the least-squares approximation, since we are minimizing the distance between the approximation and $\mathbf{b}$. The solution $\hat{\mathbf{x}}$ is given by

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

This result follows from the following theorem. If A is $m \times n$ then the following are equivalent:

- $A\mathbf{x} = \mathbf{b}$ has unique least-squares solutions for all $\mathbf{b}$
- Columns of A are linearly independent
- $A^T A$ is invertible

7.2 Inner Products

We can generalize the idea of orthogonality to abstract vector spaces. To do this, we use a metric called an *inner product*, which is a function defined on a vector space that maps two vectors $\mathbf{u}$ and $\mathbf{v}$ to a real number. The inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ has these properties:

- Commutativity: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
- Scalar associativity: $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
- Distribution over addition: $\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle$
- $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ iff $\mathbf{u} = \mathbf{0}$.

In the familiar space $\mathbb{R}^n$, the standard inner product is defined as the dot product. In other spaces, there are other definitions; for example, in the infinite-dimensional vector space of continuous functions on the interval $[a, b]$, we define the inner product as

$$\langle f(x), g(x) \rangle = \int_a^b f(x)g(x)dx.$$

Using the definition of orthogonality, $\langle f(x), g(x) \rangle = 0$, we want to find a basis for the vector space. In other words, we want to be able to express any function $y(x)$ on the interval as a sum of basis functions. One such basis is given[2] by $B = \{1, \cos x, \sin x, \cos 2x, \sin 2x, \dots\}$. Using these trig functions to represent functions is called a Fourier series:

$$y(x) = \sum_{n=0}^{\infty} a_n \cos nx + b_n \sin nx$$

where the coefficients a_n and b_n are given by the usual projection formula:

$$b_n = \frac{\langle y(x), \sin nx \rangle}{\langle \sin nx, \sin nx \rangle} = \frac{1}{\pi} \int_a^b y(x) \sin(x) dx$$

and

$$a_n = \frac{1}{\pi} \int_a^b y(x) \cos(x) dx \quad \text{except} \quad a_0 = \frac{1}{2\pi} \int_a^b y(x) dx.$$

Unfortunately, our basis is infinite in length, since the dimension is infinite. If we want to make a Fourier approximation of $y(x)$ in a finite number of basis vectors, we're basically making a projection, and the coefficients are the same.

[2]Notice that even though these functions are orthogonal, there's nothing "perpendicular" about them.

7.3 Orthogonal Matrices

Going back to $\mathbb{R}^n$, let's consider a set of n orthogonal vectors. We already know that they're perpendicular; let's now impose the additional constraint that each has to be a unit vector. If we scale each vector accordingly, we're left with a set of *orthonormal* vectors (where ortho refers to orthogonality, and normal refers to the norm being 1).

We already discussed that orthogonal vectors are linearly independent. Now let's consider a square matrix U whose columns are orthonormal. Such a matrix is called an *orthogonal matrix*, and has special properties:

- $U^T = U^{-1}$
- The rows of U are also orthonormal
- $|U\mathbf{x}| = |\mathbf{x}|$ and $U\mathbf{x} \cdot U\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$. In other words, U preserves both length and orthogonality.

These properties, specifically the fact that the transpose is equal to the inverse, are important for studying diagonalization and quadratic forms.

8 Diagonalization

For $n \times n$ matrix A, if we have n linearly independent eigenvectors, A is diagonalizable; it can be written in the form

$$A = PDP^{-1}$$

where P is the $n \times n$ matrix whose columns are the eigenvectors of A, and D is a diagonal matrix whose entries are the corresponding λ for the columns of P. In this sense, diagonal means that every non-diagonal entry is zero. One advantage of such a factorization is that $A^k = PD^kP^{-1}$, which is much easier to compute. In addition, the inverse $A^{-1} = PD^{-1}P^{-1}$ is also easy to compute.

Intuitively, the matrices P and P^{-1} represent a sort of rotation; the inverse matrix takes eigenvectors to basis axes, and P takes them back. The diagonal matrix represents a pure stretch along the basis axes.

8.1 Coordinate Transformations

This is a special case of a more general factorization, $A = PQP^{-1}$, where Q is not necessarily diagonal. Here, we're considering a new coordinate system, where P is the change-of-coordinate matrix containing the new basis vectors. To convert between bases, we have

$$\mathbf{x} = P\mathbf{x}_B \quad \text{and} \quad \mathbf{x}_B = P^{-1}\mathbf{x}$$

where $\mathbf{x}_B$ is the representation of $\mathbf{x}$ in the new coordinate system.

Now if we consider a transformation in the standard basis, $A\mathbf{x}$, then there is an analogous matrix for the new basis, $A_B\mathbf{x}_B$, which is equivalent to the first transformation. This B-matrix for A is exactly Q. In other words, the factorization of A using the change of coordinate matrix P yields the B-matrix for A.

8.2 Symmetric Matrices and Quadratic Forms

A symmetric matrix A has the property that $A = A^T$; in other words, it is symmetric about its diagonal. Symmetric matrices have useful properties, which are given by the *spectral theorem*:

- We are guaranteed n real λ (counting multiplicities), and the dimension of the eigenspace for a given λ will be equal to its multiplicity.
- These eigenspaces are mutually orthogonal.
- Every symmetric matrix is orthogonally diagonalizable; this means that the matrix P is orthogonal, which means that the factorization can be rewritten as
$$A = PDP^T.$$

This factorization is called the *spectral decomposition* of A, since D consists of the eigenvalues, and the spectrum of a matrix refers to its eigenvalues.

Switching gears a bit, we'll move to quadratic forms, which are functions $Q : \mathbf{x} \mapsto \mathbb{R}$ given by

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

for symmetric A. Quadratic forms represent polynomials of degree 2 in many variables and many have special interpretations. For example, the quadratic form where $A = I$ yields $Q = |\mathbf{x}|^2$. Generally, we have A_{ij} is the coefficient of the $x_i x_j$ term divided by two, except the diagonal: A_{ii} is the coefficient of the x_i^2 term.

So, when A is more messy, the polynomial can have many "cross-terms": terms that include more than one variable. We're interested in an expression which has only the square terms; this corresponds to A being a diagonal matrix. Since A is symmetric, we can rewrite

$$Q = \mathbf{x}^T P D P^T \mathbf{x} = \mathbf{y}^T D \mathbf{y}$$

where $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ with basis P. Since this new basis yields a quadratic form with no cross-terms, the basis vectors in P are called the principal axes for Q. We can also classify quadratic forms; positive-definite iff all $\lambda > 0$, negative-definite iff all $\lambda < 0$, and indefinite otherwise.