Table of Contents and Results

1. Vector Spaces

A set F with two binary operations $+: (\mathbb{R} \times \mathbb{R}) \to \mathbb{R}$ and $\cdot : (\mathbb{R} \times \mathbb{R}) \to \mathbb{R}$ satisfying the following criterial for all $f_1, f_2, f_3 \in F$ is a field:

a. Operations commute, i.e. $f_1 \cdot f_2 = f_2 \cdot f_1$ and $f_1 + f_2 = f_2 + f_1$
b. Operations associate, i.e. $f_1 \cdot (f_2 \cdot f_3) = (f_1 \cdot f_2) \cdot f_3$ and . $f_1 + (f_2 + f_3) = (f_1 + f_2) + f_3$
c. Multiplication distributes over addition, i.e. $f_1 \cdot (f_2 + f_3) = f_1 \cdot f_2 + f_1 \cdot f_3$
d. Neutral elements exist, i.e. $\exists 0,1 \in F$ with $0 \neq 1$ such that $f + 0 = f$ and $1 \cdot f = f$
e. Inverse elements exist, i.e. $\exists (-f), f^{-1} \in F$ such that $f + (-f) = 0$ and $f \cdot f^{-1} = 1$ for $f \neq 0$

We use symbols to preserve generality, though oftentimes regular addition and multiplication will be the operations. Fields can be finite (like $\mathbb{Z}_5$) or infinite (like the Real and Rational Numbers). Notably, the integers don't form a field (there are no multiplicative inverses). In the broad topic of algebra, fields, along with groups and rings, are primary sources of interest in their own right. In this paper however, we introduce the concept of a field only in order to make clear the concept of a vector space.

A set V is a vector space over the field F provided there exists the two mappings $(F \times V) \to V$ (which permits scalar multiplication of all vectors in V and is denoted $f \cdot \vec{v}$), and $(V \times V) \to V$, (which permits vector addition for all vectors in V and is denoted $\vec{v} + \vec{w}$), that each satisfy the following criteria for all field elements $f_1, f_2 \in F$ and for all vectors $\vec{v}, \vec{w} \in V$:

a. Multiplication associates, i.e. $f_1 \cdot (f_2 \cdot \vec{v}) = (f_1 \cdot f_2) \cdot \vec{v}$.
b. Addition commutes, i.e. $\vec{v} + \vec{w} = \vec{w} + \vec{v}$.
c. V and F have neutral elements, i.e. $\exists 0 \in V$ and $\exists 1 \in F$ so $0 + \vec{v} = \vec{v}$ and $1 \cdot \vec{v} = \vec{v}$.
d. V has additive inverses, i.e. $\forall \vec{v} \in V, \exists (-\vec{v}) \in V$ so that $\vec{v} + (-\vec{v}) = 0$.
e. Elements distribute, i.e. $f_1 \cdot (\vec{v} + \vec{w}) = f_1 \cdot \vec{v} + f_1 \cdot \vec{w}$ and $(f_1 + f_2) \cdot \vec{v} = f_1 \cdot \vec{v} + f_2 \cdot \vec{v}$.

By convention, members of the vector space are called vectors and members of the field are called scalars. Geometrically, vectors are added "by tail": the vector resulting from $\vec{v} + \vec{w}$ is a vector whose tail is the base of $\vec{v}$, and whose head is at the point where $\vec{w}$ would be if it's tail were at the head of $\vec{v}$. Since the resultant vector is the same whether one "starts" at $\vec{v}$ or $\vec{w}$, this is sometimes called the parallelogram rule. A Vector Space shares some of the criteria of a field, but notably lacks the necessity of a multiplication-like operation between elements within the vector space—this prevents the necessity of having a "1" element in the vector space, as well as commutative multiplication between vectors, among other properties. The essence of a vector means different things depending on the context (a physics student and statistics student both use the term to describe different things), but mathematicians generalize the concept to be applicable to a variety of situations.

2. Basis and Dimension

It is natural to think about how vectors interact with each other. A **linear combination** of n vectors $\{\overrightarrow{v_1}, \dots, \overrightarrow{v_n}\} \in V$ presents in the form $f_1 \cdot \overrightarrow{v_1} + \cdots + f_n \cdot \overrightarrow{v_n}$ for generic field elements $\{f_1, \dots, f_n\} \in F$. The **span** of these vectors is the set of all their possible linear combinations; they span their vector space if for any $\vec{v} \in V$, there exist scalars $\{f_1, \dots, f_n\} \in F$ so that $\vec{v} = f_1 \cdot \overrightarrow{v_1} + \cdots + f_n \cdot \overrightarrow{v_n}$.

It is just as natural to think about how to most succinctly represent space with these vectors. After all, all of the vectors in a vector space clearly span the whole space, but that is not very interesting. To get a better sense of which vectors "add" something to the span, we need the idea of independence. Geometrically, vectors whose tail sits on the origin are **linearly independent** if they lie on distinct lines. If some group of vectors lied on the same line with different magnitudes, then the same space that would be reached by the scaled version of one of the vectors could be reached by the scaled version of the other vector. So to say that vectors are linearly independent is to say that whenever a combination of the vectors results in the null vector, the scalars are all zero. Adding a vector that is dependent (i.e. a scaled version) of another that is already in the span does not expand the span.

A motivating question is how to generate a vector space with a linear combination of some vectors in the space. The minimum number of vectors which span a space is called the **dimension** of the space. This aligns with our conceptual understanding of dimension-- to describe a line, one just needs width. To describe a plane, one needs width and height. To describe a cube, one needs width, height, and depth. Of course, one benefit of linear algebra is its relevance to higher dimensional spaces that do not have easy geometric interpretations.

The actual vectors whose linear combinations compose the space are called a **basis** of the space. To be concrete, a set of linearly independent vectors $\{\overrightarrow{v_1}, \dots, \overrightarrow{v_n}\} \in V$ form a basis for the vector space V if they span the space. In the plane, $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ are classic choices for the basis of the space (this choice is not at all unique). For simplicity, call $\begin{bmatrix}1\\0\end{bmatrix} = \vec{\imath}$ and call $\begin{bmatrix}0\\1\end{bmatrix} = \vec{\jmath}$. Every horizontal vector is a scaled version of $\vec{\imath}$, every vertical vector is a scaled version of $\vec{\jmath}$, and every other vector on the plane is a linear combination of two. In this view, the basis of a space serves as its "measuring stick". Just as how a man's weight could be described in both pounds and kilograms, a space can be described by different basis choices. Each choice is valid, but certain choices make more sense than others for a given situation (describing the weight of a truck in milligrams would probably be less valuable than describing it's weight in tons). Extending the metaphor, scaled versions of a basis vector are like describing a person's weight in kilograms and grams—they are different in terms of their magnitude but not in terms of their direction.

3. Linear Transformations

A linear transformation is a function between vector spaces $T: V \to W$ that preserves linearity; for any $f_1, f_2 \in F_1$ and any vectors $\overrightarrow{v_1}, \overrightarrow{v_2} \in V$ we have $T(f_1\overrightarrow{v_1} + f_2\overrightarrow{v_2}) = f_1 \cdot T(\overrightarrow{v_1}) + f_2 \cdot T(\overrightarrow{v_2})$. The range of this function (the span of the transformation) is called the **column space**, and the dimension of the column space is called the **rank**. A key takeaway from linear algebra is that any linear transformation between finite-dimensional vector spaces can be represented as a matrix if the bases of the two vector spaces are fixed.

Pf: Consider the linear transformation $T: V \to W$, and fix $\{\overrightarrow{v_1}, \dots, \overrightarrow{v_n}\}$ to be the basis of V and $\{\overrightarrow{w_1}, \dots, \overrightarrow{w_m}\}$ to be the basis of W. Then for all $\vec{x} \in V$, there are unique scalars $\{a_1, \dots, a_n\} \in F$ so that $\vec{x} = a_1 \cdot \overrightarrow{v_1} + \cdots + a_n \cdot \overrightarrow{v_n}$. Likewise, for all $\vec{y} \in W$, there are unique scalars $\{b_1, \dots, b_m\} \in F$ so that $\vec{y} = b_1 \cdot \overrightarrow{w_1} + \cdots + b_m \cdot \overrightarrow{w_m}$.

So for all $\vec{x} \in V$, $T(\vec{x}) = T(a_1 \cdot \overrightarrow{v_1} + \cdots + a_n \cdot \overrightarrow{v_n}) = a_1 \cdot T(\overrightarrow{v_1}) + \cdots + a_n \cdot T(\overrightarrow{v_n})$ by the definition of linear transformations (where each collection of $\{a_1, \dots, a_n\}$ is of course dependent on the given $\vec{x} \in V$). Further, for all basis vectors $\overrightarrow{v_i} \in V$, we have $T(\overrightarrow{v_i}) \in W$, so $T(\overrightarrow{v_i})$ can be represented as a unique linear combination of the scalars in F_2 and the basis of W.

To distinguish between the scalars of W used to identify the basis vectors of V, we elect to use dual subscript. The first subscript refers to the relative position of the scalar in defining the element $T(\overrightarrow{v_i}) \in W$. The second subscript refers to the i^{th} basis vector of V that the linear transformation is being performed on. So for a basis vector $\overrightarrow{v_i} \in V$, $T(\overrightarrow{v_i}) = \{b_{1_i}, \dots, b_{m_i}\}^T$.

Recall that for a given $\vec{x} \in V$, we have $T(\vec{x}) = a_1 \cdot T(\overrightarrow{v_1}) + \cdots + a_n \cdot T(\overrightarrow{v_n})$. Specifically, $T(\vec{x}) = a_1 \cdot (b_{1_1} \cdot \overrightarrow{w_1} + \cdots + b_{m_1} \cdot \overrightarrow{w_m}) + \cdots + a_n \cdot (b_{1_n} \cdot \overrightarrow{w_1} + \cdots + b_{m_n} \cdot \overrightarrow{w_m})$. It is important to acknowledge that the only part of this equality unique to the element the transformation is being performed on are the a_i's; each b is fixed based on the given basis vector in V, and each $\overrightarrow{w_i}$ is a basis for W.

So $T(\vec{x}) = (a_1 \cdot b_{1_1} \cdot \overrightarrow{w_1} + \cdots + a_1 \cdot b_{m_1} \cdot \overrightarrow{w_m}) + \cdots + (a_n \cdot b_{1_n} \cdot \overrightarrow{w_1} + \cdots + a_n \cdot b_{m_n} \cdot \overrightarrow{w_m})$. Grouping terms, $T(\vec{x}) = (a_1 \cdot b_{1_1} + \cdots + a_n \cdot b_{1_n}) \cdot \overrightarrow{w_1} + \cdots + (a_1 \cdot b_{m_1} + \cdots + a_n \cdot b_{m_n}) \cdot \overrightarrow{w_m}$.

To show that the linear transformation can be performed with matrix multiplication, we must show that the multiplication of a matrix with the vector representation of a given element in V leads to the vector representation of the transformed element in W. Consider the matrix where

the i^{th} column in the matrix is the vector of unique scalars for the i^{th} basis vector of V, $T(\overrightarrow{v_i})$.
Call this fixed $m \times n$ matrix A. So $A = \begin{bmatrix} b_{1_1} & \cdots & b_{1_n} \\ \vdots & \cdots & \vdots \\ b_{m_1} & \cdots & b_{m_n} \end{bmatrix}$.

We showed above that the $n \times 1$ matrix (or column vector representation) of $\vec{x}$ is a set of scalars unique to the $\vec{x}$, $\{a_1, \dots, a_n\}^T$. Call this matrix v. So $A \cdot v$ is a $m \times 1$ matrix, call it w.

Observe $w = \left\{\left(a_1 \cdot b_{1_1} + \cdots + a_n \cdot b_{1_n}\right), (\dots), \left(a_1 \cdot b_{m_1} + \cdots + a_n \cdot b_{m_n}\right)\right\}^T$, precisely the unique matrix of scalars used to define $T(\vec{x})$.

Q.E.D

Consider $T: \mathbb{R}^3 \to \mathbb{R}^2,\ T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. We showed previously that both $\mathbb{R}^3$ and $\mathbb{R}^2$ were vector spaces over the field of Real Numbers. T is a linear transformation as for any real numbers r_1 and r_2, and any vectors $(x_1, x_2, x_3)^T$ and $(x_4, x_5, x_6)^T$, we have $T\left(r_1\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + r_2\begin{pmatrix} x_4 \\ x_5 \\ x_6 \end{pmatrix}\right) =$

$$T\begin{pmatrix} r_1x_1 + r_2x_4 \\ r_1x_2 + r_2x_5 \\ r_1x_3 + r_2x_6 \end{pmatrix} = \begin{pmatrix} r_1x_1 + r_2x_4 \\ r_1x_2 + r_2x_5 \end{pmatrix} = r_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + r_2\begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = r_1T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + r_2T\begin{pmatrix} x_4 \\ x_5 \\ x_6 \end{pmatrix}.$$

Fix $\{(1,0,0), (0,1,0), (0,0,1)\}$ to be the basis of $\mathbb{R}^3$ and $\{(1,0), (0,1)\}$ to be the basis for $\mathbb{R}^2$—it should be immediately apparent that these sets can act as bases for the vector spaces in question. Then if the linear transformation is to be represented by a matrix, call it A, it must be a 2×3 one, since any given vector $\vec{x} \in \mathbb{R}^3$ is represented by a unique 3×1 matrix of scalars that is dependent on the basis chosen for $\mathbb{R}^3$. This allows for the product of the matrix representation of the linear transformation A and the matrix representation of a given vector $\vec{x}$ to be 2×1. We would like to see this 2×1 matrix be the unique matrix of scalars representing the element $T(\vec{x})$.

If we let A be the collection of transformed basis vectors for $\mathbb{R}^3$, we see that the matrix is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. This becomes apparent when one breaks down the basis vectors of $\mathbb{R}^3$. We see that

$T(\overrightarrow{v_1}) = T\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $T(\overrightarrow{v_2}) = T\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, and $T(\overrightarrow{v_3}) = T\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

So select any vector in $\mathbb{R}^3$. Preserving generality, we have $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$.

Then $A \cdot \vec{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} (1 \cdot x_1) + (0 \cdot x_2) + (0 \cdot x_3) \\ (0 \cdot x_1) + (1 \cdot x_2) + (0 \cdot x_3) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. We saw above that $T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, and thus have verified our results.

This example helps illustrate two properties of a linear transformation. First, each column in the $m \times n$ matrix represents where the basis vectors of the space are moved to. In the above example, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ took the basis vector $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$; in a 2×2 case, $\begin{bmatrix} 3 & 0 \\ 2 & -1 \end{bmatrix}$ takes the basis vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and the basis vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$. By following how the linear transformation changes the position of the basis vectors, we get a complete understanding of how the linear transformation changes the position of all vectors in the space; for a transformation between two-dimensional spaces, the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ takes the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ to $x\begin{bmatrix} a \\ c \end{bmatrix} + y\begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$. This explains why the **identity matrix** (usually denoted I), with 1's on the upper-left to bottom-right diagonal and 0's elsewhere, keeps any vector in it's original position.

Next, we see that when $m < n$ (when the number of rows is less than the number of columns, or when the amount of equations is less than the number of unknown variables), space becomes condensed— three space is reduced to a plane, a plane reduced to a line, a line reduced to a point, etc. Such a transformation is called a projection. In general, projections have many distinct inputs leading to the same output. In the opposite case, when $m > n$, one would instead see many different outputs being mapped to by the same input. In general, for a $m \times n$ matrix representing the linear transformation $T: V \to W$, $dim(V) = n$ and $dim(T(V)) = m$.

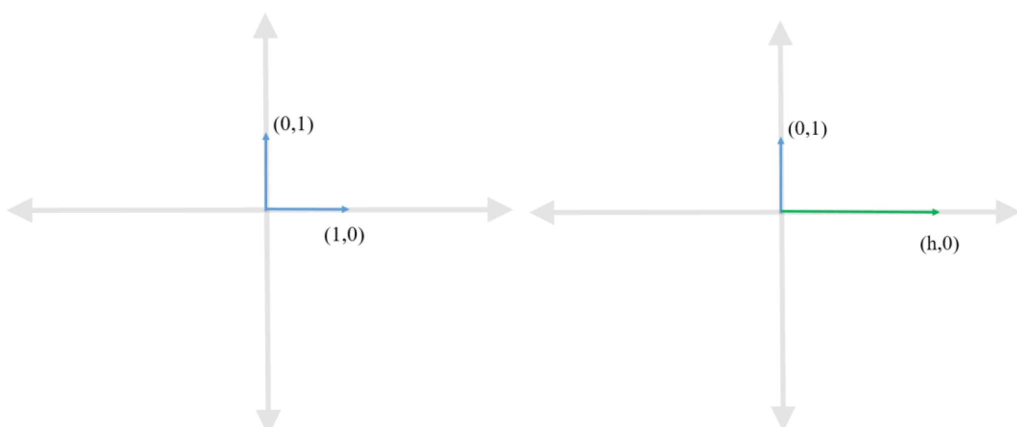
We can also combine transformations. Consider the transformation $T: V \to W$ and the transformation $L: W \to Z$. Applying L after first applying T to a vector $\vec{v}$ would be equivalent to taking $L(T(\vec{v}))$. Let A be the $m \times n$ matrix representing T and let B be a $p \times m$ matrix representing L. Then we can also represent the transformation as BA, where the product, call it the $p \times n$ matrix C, will be a sum of the matrices achieved by multiplying B and the first column of A, then B and the second column of A, etc. This gives a geometric interpretation for why matrix multiplication is not commutative in general—the order of linear transformations matter (e.g. rotating then reflecting a plane in three space is not the same as reflecting the plane then rotating the plane).

4. Determinant

Geometrically, the absolute value of the determinant represents the scalar multiple of how the area or volume or hyper-volume between basis vectors change after a transformation. Since the transformation is linear, the area/volume/hyper-volume of any section in the vector space changes by the same multiple after the transformation. This helps explain why the determinant only applies to square matrices: for a $m \times n$ matrix with $m > n$, space "expands" from n dimensions to m dimensions while when $m < n$, space "contracts" from m dimensions to n dimensions—in either case, what the determinant is measuring (the change in area or volume or hyper-volume) is illogical.

The above reasoning accurately describes the magnitude of a determinant, but it excludes one point about the direction of the determinant. The determinant of a linear transformation is negative when the basis vectors change their relative positions (if a basis vector was originally to the right of another basis vector and is now on the left for example).

Imagine the following scenarios for the vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to illustrate this point. The space expands horizontally by h if $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ goes to $\begin{bmatrix} h \\ 0 \end{bmatrix}$ with the transformation $\begin{bmatrix} h & 0 \\ 0 & 1 \end{bmatrix}$ (see below).



If both basis vectors grow by h in the same direction, then the space expands by h^2. Fix the magnitudes of the basis vectors and the vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, then rotate the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ counter-clockwise. This rotation decreases the determinant of the matrix (the area between the vectors) from 1 (when the vectors are perpendicular, and the transformation is the identity matrix) to $\frac{\sqrt{2}}{2}$ (when the vectors are at a 45-degree angle), to zero (when the vectors are aligned with the transformation $\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$), to -1 (when the vectors change position with the transformation $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$).

A determinant is zero when the transformation brings space down a dimension (a two-dimensional plane becomes a line on the plane, a three-dimensional cube becomes a plane in three-space, etc.) In the case of a zero determinant, the transformation is not linearly independent, and the matrix is **singular**, meaning it has no multiplicative inverse. The opposite of a singular matrix is an **invertible matrix**; A is invertible if there exists a matrix A^{-1} such that $AA^{-1} = I = A^{-1}A$. From this definition, it is clear that non-square matrices are not invertible (though may have either a left or right inverse), and that most square matrices will be invertible (unless a square matrix has zero determinant and so happens to condense space down a dimension, it will be invertible).

We can state a recursive computation for the determinant of a $n \times n$ matrix A by acknowledging that the determinant of a 1×1 matrix is the element itself. For symbolic ease, we let a_{i_j} represent the element in the i^{th} row and j^{th} column of A, and let A_{ij} represent the $(n-1) \times (n-1)$ matrix formed by taking all but the i^{th} row and j^{th} column of A. Then by fixing a row i, the determinant of A is computed to be $\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$. Each matrix A_{ij} is called the **minor** of a_{ij}, and the determinant of each the signed minors, $A_{ij} = (-1)^{i+j} \det(A_{ij})$ is called the **cofactor.** A nice property of this definition is that it is generalized to be a linear combination of the cofactors of any row i and row i itself. This property could come in useful when a certain row has many zeros.

For thoroughness, we compute two examples:

For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(A) = [(-1)^2 \cdot a \cdot \det(d)] + [(-1)^3 \cdot b \cdot \det(c)] = ad - bc$.

For $A = \begin{bmatrix} abc \\ def \\ ghi \end{bmatrix}$, $\det(A) = \left[(-1)^2 \cdot a \cdot \det\left(\begin{bmatrix} ef \\ hi \end{bmatrix}\right)\right] + \left[(-1)^3 \cdot b \cdot \det\left(\begin{bmatrix} df \\ gi \end{bmatrix}\right)\right] + \left[(-1)^4 \cdot c \cdot \det\left(\begin{bmatrix} de \\ gh \end{bmatrix}\right)\right] = [a \cdot (ei - fh)] + [-b \cdot (di - fg)] + [c \cdot (dh - eg)] = aei - afh - bdi + bfg + cdh - ceg$. We would arrive at the same answer by taking the second row: $\det(A) = \left[(-1)^3 \cdot d \cdot \det\left(\begin{bmatrix} bc \\ hi \end{bmatrix}\right)\right] + \left[(-1)^4 \cdot e \cdot \det\left(\begin{bmatrix} ac \\ gi \end{bmatrix}\right)\right] + \left[(-1)^5 \cdot f \cdot \det\left(\begin{bmatrix} ab \\ gh \end{bmatrix}\right)\right] = [-d \cdot (bi - ch)] + [e \cdot (ai - cg)] + [-f \cdot (ah - bg)]$

So far we've talked about the geometric interpretation of the determinant and it's calculation, but we've left out it's defining properties. We mention them here for a matrix A:

1) The determinant is linearly dependent on the rows. If B is a matrix of zeros in all but one row, then $\det(A+B) = \det(A) + \det(B)$. If C is a matrix of all ones but for one column of constants c, then $det(AC) = c \cdot det(A)$.
2) Exchanging two rows of a matrix switches the sign of the determinant.
3) The determinant of the identity matrix is one.
4) The determinant of the product is the product of the determinant, $\det(AB) = \det(A)\det(B)$.
5) The determinant of the transpose is the determinant of the matrix, $\det(A^T) = \det(A)$.

The determinant function has more utility than just quantifying how linear transformations distort space; it can also be used to calculate the inverse of a matrix. To show this fact for an invertible $n \times n$ matrix A, we create a matrix D that has the determinant of A on the diagonal and zero's elsewhere. In this sense the matrix is $D = I \cdot \det(A)$. If D was a product of two matrices, say $D = D_1 D_2$, then each entry i, j in D would be a linear combination of the i^{th} row of D_1 and the j^{th} column of D_2. Every diagonal entry in D (d_{11}, d_{22}, d_{33}, etc.) is the determinant of A, which we have defined to be $\sum_{j=1}^{n} (-1)^{i+j} a_{ij} A_{ij}$ for a fixed row i of A. So every i^{th} row of D_1 will be identical to the i^{th} row of A ($D_1 = A$ because of the a_{ij} term in the summation), and every j^{th} column of D_2 will be the j^{th} row of the cofactor matrix, or equivalently the j^{th} column of the transpose of the cofactor matrix ($D_2 = A_{cof}^T$ because of the A_{ij} term in the summation). The transpose of the cofactor matrix is called the **adjugate matrix**. Recall that the entries to the cofactor/adjugate matrix come in the form A_{ij} representing the signed determinant of the matrix A less its i^{th} row and j^{th} column. So we can write $AA_{cof}^T = I \det(A)$, then $A^{-1}AA_{cof}^T = A^{-1}I \det(A)$, and finally $\frac{1}{\det(A)} A_{cof}^T = A^{-1}$. In the 3×3 case, we have $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix} = \begin{bmatrix} det(A) & 0 & 0 \\ 0 & det(A) & 0 \\ 0 & 0 & det(A) \end{bmatrix}$.

To see why every non-diagonal entry is zero, notice that the calculation for a non-diagonal entry is essentially computing the determinant of a matrix that is the same as A except for one row, which is a copy of another in the matrix. For example, the entry in the first row and second column of the above example is the sum-product $a_{11}A_{21} + a_{12}A_{22} + a_{13}A_{23}$. This is the formula for the determinant of A if $A_{1n} = A_{2n}$, which can only be the case when the second row is an exact replica of the first. We showed that a defining characteristic of the determinant was that a row exchange flipped the sign of the determinant. So if one were to flip the first and second row of the example, then the sign of the determinant would need to change. But switching the rows would not change the matrix since the rows are identical, and the determinant of the matrix must be zero.

Determinants can also be used as a tool to solve systems of linear equations. For n linear equations and n unknown variables, we can arrange an $n \times n$ matrix A of all the coefficients, a column matrix x of all the unknown variables, and a column matrix b of the solutions. There are many methods for solving such systems (outlined later in the paper), but one way involves the use of determinants via **Cramer's Rule**.

We saw $A^{-1} = \frac{1}{\det(A)} A^T_{cof}$, and so by multiplying each side by b we can say $A^{-1}b = \frac{1}{\det(A)} A^T_{cof} b$. We are looking for the solution x to $Ax = b$, which is just $A^{-1}b$, so $x = A^{-1}b = \frac{1}{\det(A)} A^T_{cof} b$. If B_j is the matrix A whose j^{th} column is replaced by b, then the j^{th} element in x is $x_j = \frac{\det(B_j)}{\det(A)}$. We have shown earlier that $\det(A) = \det(A^T)$, and can then compute $\det(B_j)$ as a linear combination of the j^{th} column of B_j (which is b) and the cofactors of the j^{th} column, $\det(B_j) = b_1 A_{1j} + \cdots + b_n A_{nj}$. This quantity is the j^{th} element of $A^T_{cof} b$, so we have shown that the j^{th} element of the solution is a ratio of the aforementioned determinants and are done. It should be mentioned that applying Cramer's rule is generally a much slower and more computationally demanding way of solving systems of linear equations than other methods; its usefulness is its ability to find individual components to the solution relatively quickly.

5. Eigenvectors

So far, we have been interested in how linear transformations change vectors in a space. We know that given a basis for a n dimensional vector space, a $m \times n$ matrix exists that represents the transformation of the space from $\mathbb{R}^n$ to $\mathbb{R}^m$. When the linear transformation in question preserves the dimension of the space, the determinant represents the scaling factor of the generalized volume between any vectors in the space. The determinant was used to find the vector $\vec{x}$ that lands on a vector $\vec{b}$ after the transformation A. This is a natural problem to solve given our interest in how linear transformations change space.

We now move our attention to another problem which seems less natural but is never the less equally important to $A\vec{x} = \vec{b}$. This problem is finding the vectors that are only scaled in magnitude by a linear transformation; given a linear transformation A we want to find the vectors $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$ for some scalar quantity λ. Said a different way, we are looking for vectors that remain on their span after a linear transformation is applied. Unless the transformation is a multiple of the identity matrix, such vectors are rare or even non-existent (for instance we can immediately say that any linear transformation between different dimensions cannot possibly alter a vector only in terms of its magnitude). We call the vectors which possess this special property eigenvectors, and we call the scaling factor λ their eigenvalue.

Notice that we can write $A\vec{v} = \lambda\vec{v}$ as $A\vec{v} = I\lambda\vec{v}$ (where I has the same number of rows and columns as A) and then $A\vec{v} - I\lambda\vec{v} = 0$ and finally $(A - I\lambda)\vec{v} = 0$. The matrix $I\lambda$ shifts the diaganol enteries of A by λ, and we are trying to find the non-trivial solutions to this equation (the zero vector is always a solution because the origin of a space remains stationary after a linear transformation). In general, the collection of vectors which land on the origin after a linear transformation are called **null space** of the transformation, and the null space of $A - I\lambda$ is called the **eigenspace**. Only singular matrices have a non-trivial null space, so we need to find when the determinant of the transformation is zero; we are looking for the solutions to $det(A - I\lambda) = 0$. In this sense, we are first choosing the eigenvalues which allows $A - I\lambda$ to have a non-trivial null space before solving for the null space.

Since there are n diagonal entries effected by the eigenvalue, we know that the determinant of the eigenmatrix will be a n^{th} degree polynomial (called the **characteristic polynomial**) and thus have at most n real roots. One can then solve for the roots to determine the values of the (at most) n eigenvalues before plugging each value in to the matrix one-by-one and solving for the line of eigenvectors associated with its eigenvalue.

In order to understand a key application of eigenvectors, we must return again to the idea of basis vectors. Recall that vectors form a basis for a space provided they are linearly independent and span the space. Basis vectors are far from unique; for an n-dimensional space, any n linearly independent vectors will span the space. The important realization here is that for two distinct sets of basis vectors $\langle \overrightarrow{v_1}, \ldots, \overrightarrow{v_n} \rangle$ and $\langle \overrightarrow{w_1}, \ldots, \overrightarrow{w_n} \rangle$ each basis vector in either set *defines* unit length in the given direction.

An example may help clarify. Consider the basis vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Any vector in the space is a combination of the scaled versions of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Now consider the vectors that would be described in the original basis as $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$. These vectors also span the space, but in light of this new basis they would be described as $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. So the choice of basis is really a shift in perspective; the same vectors in a space may be described differently depending on the choice of basis. A natural question then becomes how we can translate between bases in a space.

We already have enough background to answer this question. Call $b_1 = \langle \overrightarrow{v_1}, \ldots, \overrightarrow{v_n} \rangle$ and $b_2 = \langle \overrightarrow{w_1}, \ldots, \overrightarrow{w_n} \rangle$. Then let S be the linear transformation describing b_2 in terms of b_1. This is called the **change of basis matrix**. Each column i in S is then the column vector $\overrightarrow{w_i}$ in terms of b_1. Then given a vector $\vec{v}$ described by the basis vectors b_2, we can describe $\vec{v}$ in terms of the basis vectors b_1 by obtaining the product $S\vec{v}$. Since S translates vectors described in b_2 to b_1, S^{-1} does the opposite and translates vectors described in b_1 to vectors described in b_2. Consider the same change of basis as described above, $\overrightarrow{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\overrightarrow{v_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to $\overrightarrow{w_1} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $\overrightarrow{w_2} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ (from the perspective of the b_1 basis). Then a vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ in b_2 terms can be described in b_1 terms by taking the product $S\vec{v} = \begin{bmatrix} 20 \\ 02 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. We showed how to calculate an inverse matrix previously. $S^{-1} = \frac{1}{\det(S)} S_{cof}^{T} = \frac{1}{4} \begin{bmatrix} S_{11}S_{21} \\ S_{12}S_{22} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 20 \\ 02 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$. So to desribe the vector $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ (from the perspective of b_1) in terms of b_2, we take $\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and have succeeded in showing an example of our results.

Representing a linear transformation with a matrix is dependent on the basis chosen. Each i^{th} column in the transformation matrix represents where the i^{th} basis vector lands after the transformation, so it is clear that the matrix representing a transformation in one basis will not always be the same as the matrix representing the same transformation in a different basis. We saw that given $\vec{v}$ in terms of b_2, $S\vec{v}$ describes $\vec{v}$ in terms of b_1. Then if A is the transformation matrix in terms of b_1, $AS\vec{v}$ describes the transformed vector in terms of b_1, and then $S^{-1}AS\vec{v}$ describes the transformated vector in terms of b_2. So we conclude that if a transformation A is given in terms of b_1 then $S^{-1}AS$ describes the same transformation in terms of b_2. In general, we say that A is **similar** to B if there exists an invertible matrix S such that $B = S^{-1}AS$.

Returning to the problem of diagonalization, if we let S be the matrix of eigenvectors, appropriately called the **eigenbasis matrix**, then writing the transformation A as $\Lambda = S^{-1}AS$ is an equivelant transformation to A from the perspective of basis vectors which lie on the eigenvectors. In doing so, we guarantee the transformation Λ, called the **eigenvalue matrix**, is diagonal, with each entry equal to the eigenvalues of the eigenvectors! To see this, let A be $n \times n$ with n eigenvectors in the form $\vec{x_i} = \langle x_{i_1}, \dots, x_{i_n} \rangle$ and corresponding eigenvalues in the form λ_i. Then $AS = \begin{bmatrix} a_{1_1} & \cdots & a_{n_1} \\ \vdots & \ddots & \vdots \\ a_{1_n} & \cdots & a_{n_n} \end{bmatrix} \begin{bmatrix} x_{1_1} & \cdots & x_{n_1} \\ \vdots & \ddots & \vdots \\ x_{1_n} & \cdots & x_{n_n} \end{bmatrix} = \begin{bmatrix} \lambda_1 x_{1_1} & \cdots & \lambda_n x_{n_1} \\ \vdots & \ddots & \vdots \\ \lambda_1 x_{1_n} & \cdots & \lambda_n x_{n_n} \end{bmatrix}$ by the definition of eigenvalue (recall that the initial problem was to find $A\vec{v} = \lambda\vec{v}$, and by multiplying by columns, we see A multiplying each eigenvector on the left). By the properties of matrix multiplication, we see that this is exactly SΛ, $\begin{bmatrix} \lambda_1 x_{1_1} & \cdots & \lambda_n x_{n_1} \\ \vdots & \ddots & \vdots \\ \lambda_1 x_{1_n} & \cdots & \lambda_n x_{n_n} \end{bmatrix} = \begin{bmatrix} x_{1_1} & \cdots & x_{n_1} \\ \vdots & \ddots & \vdots \\ x_{1_n} & \cdots & x_{n_n} \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$. So we have $AS =$ SΛ, and then $S^{-1}AS = \Lambda$ as required.

For a non-diagonal square matrix A, calculating A^n for large n is computationally difficult. It is almost always preferable to first transform A to the eigenvalue matrix Λ, then calculate the n^{th} power as $\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}^m = \begin{bmatrix} {\lambda_1}^m & & \\ & \ddots & \\ & & {\lambda_n}^m \end{bmatrix}$, then transform back to A. See $\Lambda^n = (S^{-1}AS)^n = (S^{-1}AS)(S^{-1}AS) \dots (S^{-1}AS) = (S^{-1})A(SS^{-1})A(SS^{-1}) \dots A(S) = S^{-1}A^nS$. We want A^n, so we take $S\Lambda^nS^{-1} = S(S^{-1}A^nS)S^{-1} = (SS^{-1})A^n(SS^{-1}) = A^n$. This is the general procedure for finding A^n: we first find the eigenvectors and eigenvalues, then form the eigenvalue matrix Λ, then calculate Λ^n, and finally compute $S\Lambda^nS^{-1}$.

Take the following example: $\begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}^{21}$. This is a bonafide nightmare without the help of Wolfram Alpha or something similar. Instead we elect to use the method of diagonalization described above. First, we calculate the eigenvalues and eigenvectors: $\det\left(\begin{bmatrix} (7-\lambda) & 2 \\ -4 & (1-\lambda) \end{bmatrix}\right) = \lambda^2 - 8\lambda + 15$. The quadratic formula tells us the roots of the characteristic polynomial are $\frac{-(-8) \pm \sqrt{(-8)^2 - [4(1)(15)]}}{2(1)} = \frac{8 \pm 2}{2}$; $\lambda_1 = 5$ and $\lambda_2 = 3$. Solving for the eigenvectors, we have $\begin{bmatrix} 7-(5) & 2 \\ -4 & 1-(5) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ so $x \begin{bmatrix} 2 \\ -4 \end{bmatrix} + y \begin{bmatrix} 2 \\ -4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\vec{v_1} = c \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ for any scalar c. Further $\begin{bmatrix} 7-(3) & 2 \\ -4 & 1-(3) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ so $x \begin{bmatrix} 4 \\ -4 \end{bmatrix} + y \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\vec{v_2} = c \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ for any scalar c. So we arrange $\vec{v_1}$ and $\vec{v_2}$ into a change of basis matrix, $S = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$. We know that $S^{-1} = \frac{1}{\det(S)} S_{cof}^{T} = \frac{1}{-1}\begin{bmatrix} +S_{11} & -S_{21} \\ -S_{12} & +S_{22} \end{bmatrix} = \frac{1}{-1}\begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$. Then the eigenvalue matrix is $S^{-1}AS = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$. We are interested in applying the transformation 21 times over, so the transformation is $\Lambda^{21} = \begin{bmatrix} 5^{21} & 0 \\ 0 & 3^{21} \end{bmatrix}$. To write this in terms of the original basis, we transform back as $S\Lambda^{21}S^{-1} =$

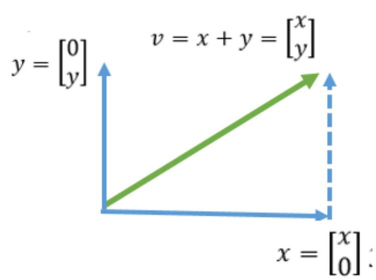
$\begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^{21} & 0 \\ 0 & 3^{21} \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} [(2\cdot 5^{21}) - 3^{21}] & [5^{21} - 3^{21}] \\ [(-2\cdot 5^{21}) + (2\cdot 3^{21})] & [-5^{21} + (2\cdot 3^{21})] \end{bmatrix}$. The number of steps has been dramatically reduced—to calculate A^{21} by brute force, we'd need one computation for A^2, another for A^4, a third for A^8, a fourth for A^{16}, a fifth for A^{20}, and a sixth for A^{21}. Here we just diagonalized and calculated an inverse before performing one set of matrix multiplication.

It should be noted that there are some transformations which do not permit such an easy process. Any rotation of space for example can not possibly have any real eigenvectors, as no vector in the original space remains on it's span. And matrices with an **algebraic multiplicity**, that is those matrices whose characteristic polynomial permits the same eigenvalues, lack the nessecary eigenvectors to diaganolize. Take $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. The eigenvalues solve $(2-\lambda)(2-\lambda) = 0$, so are $\lambda_1 = \lambda_2 = 2$. But then there is just one line of non-trivial solutions for $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

6. Orthogonality

The idea of orthogonality necessitates the idea of direction. A set of vectors are orthogonal if no component of one vector points in the same direction as another. On a plane, orthogonality is akin to perpendicularity; we know two vectors are orthogonal if there is a 90-degree angle between them. We would like to find an explicit test for orthogonality in any dimension. To do so, we introduce the idea of a vector **norm**. "Norm" is practically synonymous with length and magnitude.

In one dimension, the norm is clear. It is simply the vector's component. In two dimensions, the norm follows from the Pythagorean Theorem. The first component of the vector gives the length along one basis vector, and the second component gives the height along the second basis vector. The vector is calculated as the sum of the scaled magnitudes of each basis vector. In the illustration below, we see that a vector $\vec{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ is the geometric sum of x multiples of one basis vector and y multiples of the other. The length of $\vec{v}$, denoted $\|\vec{v}\|$, is known to be $\|\vec{v}\|^2 = \mathrm{x}^2 + y^2$ (see the right triangle formed by $\vec{x}, \vec{y}$, and $\vec{v}$).

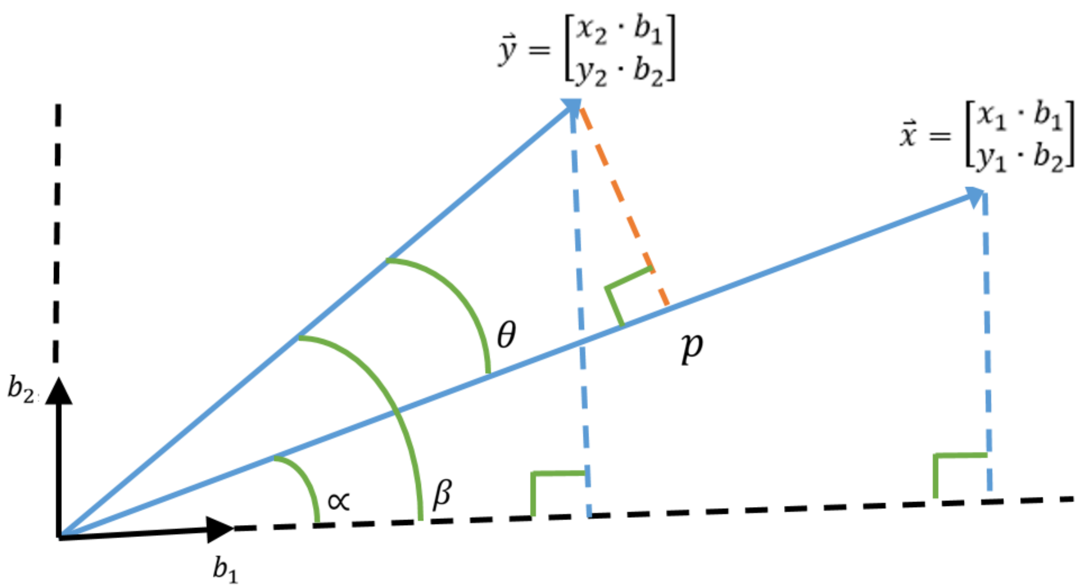


This is generalized in higher dimensions by repeating the process in two space. Imagine a vector in three-space $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Then a box is formed with each basis vector giving length, width, and height, and the vector determining the endpoints in each direction. Applying the theorem to first the length and width, we have the squared length of the base of a triangle, $\mathrm{x}^2 + y^2$, and the height of the box z. We can then reapply the theorem to see the length of the hypotenuse is $x^2 + y^2 + z^2$. In n dimensions, the same procedure is done: first we get length in one plane, then use that length and reapply the formula to get length in another plane, then use that length to calculate the length in another plane, etc. So for a vector $\vec{x}$, $\|\vec{x}\|^2 = \vec{x}^T\vec{x}$.

It is immediately clear how to test for orthogonality between individual vectors. They are orthogonal if and only if the square of their magnitudes is the squared norm of their difference; for vectors $\vec{x}$ and $\vec{y}$, we need $\|\vec{x}\|^2 + \|\vec{y}\|^2 = \|\vec{x} - \vec{y}\|^2$. Expanding this equation, we have $(x_1^2 + x_2^2 + \cdots + x_n^2) + (y_1^2 + y_2^2 + \cdots + y_n^2) = (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2$. The right side is

$(x_1{}^2 - 2x_1y_1 + y_1{}^2) + \cdots + (x_n{}^2 - 2x_ny_n + y_n{}^2)$. Grouping terms, this becomes $(x_1^2 + \cdots + x_n^2) + (y_1^2 + \cdots + y_n^2) - 2(x_1y_1 + \cdots + x_ny_n)$. The left side of the equation cancels the first two parts of the right side and we are left with $0 = -2(x_1y_1 + \cdots + x_ny_n)$. This can only be the case when the sum $(x_1y_1 + \cdots + x_ny_n)$, called the **scalar product** or the **dot product** and usually written x^Ty, is zero. When we choose basis vectors that are of equal length and orthogonal, we call them an **orthonormal basis.** When every vector in a subspace is orthogonal to every vector in another subspace, we can call the whole subspaces **orthogonal.** One of the most beautiful realizations in linear algebra deals with this concept, but we save specifics for another paper.

Geometrically, the dot product represents the projection of one vector onto the other. In the perpendicular case, this is easy to see: no component of one vector is in the same direction as the other, and so the dot product is zero. But we are not limited to just the perpendicular case. To think about other cases, we must first understand how to measure the angle between the vectors, and assume a basic understanding of trigonometry. Consider two-dimensional space. Fix a basis and let $\vec{x} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ be two vectors. Let $\propto$ be the angle between $\vec{x}$ and the first basis vector and β be the angle between $\vec{y}$ and the first basis vector (see below).



We are interested in finding $\theta = \beta - \propto$, the angle between the vectors, in terms of the vector norms and the scalar product. We know that the sine of an angle represents the ratio between the opposite side length and the hypotenuse. So $\sin(\propto) = \frac{y_1}{\|\vec{x}\|}$ and $\sin(\beta) = \frac{y_2}{\|\vec{y}\|}$. Additionally, we know that the cosine of an angle represents the ratio between the adjacent side length and the hypotenuse. So $\cos(\propto) = \frac{x_1}{\|\vec{x}\|}$ and $\cos(\beta) = \frac{x_2}{\|\vec{y}\|}$. Trigonometric identities tell us that $\cos(\theta) = \cos(\beta - \propto) = \cos(\beta)\cos(\propto) + \sin(\beta)\sin(\propto) = \frac{x_2}{\|\vec{y}\|} \cdot \frac{x_1}{\|\vec{x}\|} + \frac{y_2}{\|\vec{y}\|} \cdot \frac{y_1}{\|\vec{x}\|} = \frac{x_2 \cdot x_1 + y_2 \cdot y_1}{\|\vec{y}\| \cdot \|\vec{x}\|}$.

Notice that the numerator is exactly the dot product of $\vec{x}$ and $\vec{y}$. Then $\cos(\theta) = \frac{x^T y}{\|\vec{y}\| \cdot \|\vec{x}\|}$, and we see the dot product $x^T y = \|\vec{y}\| \cdot \|\vec{x}\| \cdot \cos(\theta)$.

The whole goal of this example was to get a sense of the dot product as the projection of one vector onto another. Suppose we want to project $\vec{y}$ onto $\vec{x}$, finding the component of $\vec{y}$ that stretches in the direction of $\vec{x}$. This entails finding the point p on $\vec{x}$ that is closest to the head of $\vec{y}$. Since the shortest distance between two points is a straight line, the line from $\vec{y}$ to p must be perpendicular to $\vec{x}$. As p lies on $\vec{x}$, it is some multiple c of $\vec{x}$, $p = c \cdot \vec{x}$, and the problem becomes solving for c. The line between p and $\vec{y}$ is perpendicular to $\vec{x}$, so the dot product $x^T(y - p) = 0$. Substituting and expanding, we see $x^T(y - cx) = x^T y - x^T cx$. The values $x^T y$ and $x^T cx$ are scalars, and we can pull c out on the right, so we are left with $x^T y = cx^T x$ and $\frac{x^T y}{x^T x} = c$. Solving for the projection, we know $p = c \cdot \vec{x}$, so $p = \frac{x^T y}{x^T x} \cdot \vec{x}$, and $p = \frac{xx^T y}{x^T x}$ as $\frac{x^T y}{x^T x}$ is a scalar. The numerator is the product of a matrix and a vector. Call the square matrix $xx^T = P$. Finally we have $p = (P\vec{y})\frac{1}{x^T x}$. It's unclear which form of the solution is easier to visualize: $p = \frac{x^T y}{x^T x} \cdot \vec{x}$ or $p = (P\vec{y})\frac{1}{x^T x}$. The important thing is that we found what the solution is. Algebraically, the dot product is $x^T y$. Geometrically, it is $\|\vec{y}\| \cdot \|\vec{x}\| \cdot \cos(\theta)$, or from the view of a projection $(P\vec{y})\frac{1}{x^T x}$.

We've seen how to transform between bases, and now direct our attention to creating orthonormal ones. We notice that since the dot product between orthonormal vectors is zero, for any vectors $\vec{v_i}, \vec{v_j}$ with $i \neq j$ it must be the case that $v_i^T v_j = 0$. Further, when $i = j$, $v_i^T v_j = 1$, as the lengths are normalized. If we create a matrix whose columns are the orthonormal vectors in the space, call it Q, then we see that $Q^T Q = I$, and so $Q^T = Q^{-1}$. Of course, Q can only be the matrix of basis vectors if its rank is the dimension of the space. Choosing an orthonormal basis is preferable many scenarios—in fact one might go so far as to say that if at all possible, transforming to an eigenbasis or orthonormal basis will always be preferable to a random selection. Whereas the strength of an eigenbasis laid in its ability to easily scale transformations, the strength of an orthonormal basis is its ability to maintain orthogonality—any vectors that are orthogonal prior to a transformation will also be orthogonal after the transformation.

The process of transforming to a normal basis is called **Gram-Schmidt Orthogonalization**. We have enough background to guess at this process. For ease we consider three basis vectors, but of course can apply the below steps to higher dimensional spaces. Suppose these three vectors are $\vec{v_1}$, $\vec{v_2}$, and $\vec{v_3}$. We would like to transform these vectors so that they are orthonormal and there combinations still span the space. The first vector is easy—we just divide by its magnitude to normalize the length and force the other vectors to be orthogonal to it. So $\overrightarrow{q_1} = \frac{\overrightarrow{v_1}}{\|\overrightarrow{v_1}\|}$. Now we consider $\overrightarrow{v_2}$, and so must subtract off any component which is in the same direction as $\overrightarrow{v_1}$. The projection of $\overrightarrow{v_2}$ onto $\overrightarrow{q_1}$ is $\frac{{q_1}^T v_2}{{q_1}^T q_1} \cdot \overrightarrow{q_1}$, but the denominator is one since q_1 is of unit length. The projection then simplifies to $({q_1}^T v_2)q_1$, which comes off $\overrightarrow{v_2}$ and then is divided by it's length. So $\overrightarrow{q_2} = \frac{v_2 - ({q_1}^T v_2)q_1}{\|v_2 - ({q_1}^T v_2)q_1\|}$. Finally we consider $\overrightarrow{v_3}$ and must subtract off any component in the same direction as either $\overrightarrow{q_1}$ or $\overrightarrow{q_2}$. So $\overrightarrow{q_3} = \frac{v_3 - ({q_1}^T v_3)q_1 - ({q_2}^T v_3)q_2}{\|v_3 - ({q_1}^T v_3)q_1 - ({q_2}^T v_3)q_2\|}$.