Linear Algebra

Warren Alphonso

May 18, 2019

A very reductionist summary of key Linear Algebra concepts from *Linear Algebra and its Applications* by Lay, Lay, and McDonald.

1 Systems of Linear Equations

Definition 1.1. A **linear equation** is an equation that can be written in the form

$$a_1x_1 + ... + a_nx_n = b$$

where b and the coefficients a_k are real or complex numbers.

We can record the important information of a system of linear equations in a matrix. Given the system

$$x_1 - 2x_2 + x_3 = 0$$

$$2x_2 - 8x_3 = 8$$

$$5x_1 - 5x_3 = 10$$

we place the coefficients of each variable aligned in columns

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$$

This is called the **coefficient matrix** and

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

is called the **augmented matrix**. The size of a matrix is described as **m x n** where m denotes the number of rows and n the number of columns.

Definition 1.2. Elementary Row Operations:

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a nonzero constant.

Definition 1.3. Row Echelon form denotes a matrix with:

1. All nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in column below a leading entry are zeros.

Reduced Row Echelon form means the leading entry in nonzero rows is 1.

Parallelogram Rule for Addition: If $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$ are represented as points on the plane, then $\mathbf{u} + \mathbf{v}$ corresponds to the fourth vertex of the parallelogram whose other vertices are 0, $\mathbf{u}$, and $\mathbf{v}$.

Definition 1.4. Span$\{v_1, ..., v_p\}$ denotes the set of all vectors formed by $c_1 v_1 + ... + c_p v_p$.

Definition 1.5. A set of vectors $\{v_1, ..., v_p\}$ is said to be **linearly independent** if the equation

$$c_1 v_1 + ... + c_p v_p = 0$$

has only the trivial solution.

Theorem 1. *If a set contains more vectors than there are entries in each vector, then the set is linearly independent. That is, the set $\{v_1, ..., v_p\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.*

Proof. Let $A = [v_1 \cdots v_p]$. Then A is $n \times p$, and the equation $Ax = 0$ corresponds to a system of n equations in p unknowns. In $Ax = b$, the x vector must have dimension p, so if $p > n$, then there are more variables than equations, so $Ax = 0$ has a nontrivial solution, and the columns of A are linearly dependent. ∎

An alternate way to conceptualize matrix multiplication: A **transformation** (or **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector x in $\mathbb{R}^n$ a vector $T(x)$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T and $\mathbb{R}^m$ is called the **codomain** of T. The set of all images $T(x)$ is called the **range** of T.

Definition 1.6. A transformation T is **linear** if they preserve vector addition and scalar multiplication. That is:

1. $T(u+v) = T(u) + T(v)$
2. $T(cu) = cT(u)$ for all scalars c

Every matrix transformation is a linear transformation. These two requirements mean that $T(0) = 0$ for linear transformations.

Theorem 2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a* ***unique*** *matrix A such that*

$$T(x) = Ax$$

In fact, A is the $m \times n$ matrix whose jth column is the vector $T(e_j)$, where e_j is the jth column of the identity matrix in $\mathbb{R}^n$:

$$A = [T(e_1) \cdots T(e_n)]$$

Proof. Write $x = I_n x = [e_1 \cdots e_n]x = x_1 e_1 + \cdots + x_n e_n$, and use the linearity of T to compute

$$T(x) = T(x_1 e_1 + \cdots x_n e_n) = x_1 T(e_1) + \cdots + x_n T(e_n)$$

$$\begin{bmatrix} T(e_1) \cdots T(e_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = Ax$$

■

Definition 1.7. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **onto** $\mathbb{R}^m$ if each b in $\mathbb{R}^m$ is the image of at least one x in $\mathbb{R}^n$.

Definition 1.8. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **one-to-one** if each b in $\mathbb{R}^m$ is the image of at most one x in $\mathbb{R}^n$.

2 Matrix Algebra

Definition 2.1. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then

$$AB = A\begin{bmatrix} b_1 & b_2 & \cdots & b_p \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_p \end{bmatrix}$$

Definition 2.2. Let A be an $m \times n$ matrix and let B and C have sizes for which the indicated sums and products are defined.

1. $A(BC) = (AB)C$
2. $A(B + C) = AB + AC$
3. $(B + C)A = BA + CA$
4. $r(AB) = (rA)B = A(rB)$
5. $I_m A = A = AI_n$

Warnings:

1. In general, $AB \neq BA$
2. If $AB = AC$, then it is **not true** in general that $B = C$
3. If $AB = 0$, then it is **not true** always that $A = 0$ or $B = 0$

Definition 2.3. The **transpose** of A is the matrix whose columns are formed from the corresponding rows of A, denoted by A^T.

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(cA)^T = rA^T$
- $(AB)^T = B^T A^T$

Definition 2.4. An $n \times n$ matrix A is **invertible** if there is an $n \times n$ matrix A^{-1} such that $A^{-1}A = I$.

- $(A^{-1})^{-1} = A$
- If A and B are $n \times n$ invertible matrices then so is AB. And $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$

To compute the inverse, solve the equation $AB = I$, by row-reducing the augmented matrix [A I], until you get [I B].

3 Determinants

Definition 3.1. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

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

The quantity $ad - bc$ is the **determinant** of the matrix. If the determinant is 0, the matrix A is not invertible.

Definition 3.2. To generalize, the determinant of an $n \times n$ matrix A can be computed using a **cofactor expansion** across any row or down any column. The expansion across the ith row is

$$det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

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

Theorem 3. *If A is an upper triangular matrix, then det(A) is the product of the entries on the main diagonal.*

Proof. Cofactoring an upper triangular matrix by using the first column ultimately leads to continuously multiplying the upper left item by the determinant of the smaller matrix. For example,

$$A = \begin{bmatrix} 3 & 2 & 9 \\ 0 & 4 & -1 \\ 0 & 0 & -8 \end{bmatrix}$$

Then,

$$det(A) = 3 \cdot det(\begin{bmatrix} 4 & -1 \\ 0 & -8 \end{bmatrix}) = 3 \cdot -32 = -96 = 3 \cdot 4 \cdot -8$$

∎

Definition 3.3. Determinants after Row Operations

1. If a multiple of a row in matrix A is added to another row to produce matrix B, then $det(B) = det(A)$
2. If two rows in A are swapped to produce B, then $det(B) = -det(A)$
3. If one row in A is multiplied by k to produce B, then $det(B) = k \cdot det(A)$

These identities can be used to easily find determinants of square matrices. Once we reduce a matrix A to upper triangular form B, we know $det(B) = (-1)^r det(A)$ if r is the number of row swaps we performed. If we cannot reduce to row echelon form, we know the determinant must be 0 since A must not be invertible.

Theorem 4. *If A is an $n \times n$ matrix, then $det(A^T) = det(A)$.*

Proof. We proceed by induction. The theorem is trivially true for $n = 1$. Assume the theorem is true for $k \times k$ matrices. We will show it holds for $n = k + 1$. The cofactor of a_{1j} in A equals the cofactor of a_{j1} in A^T because it is a $k \times k$ determinant. Thus, the cofactor of $det(A^T)$ down the first *column* equals the cofactor of $det(A)$ across the first *row*, so A and A^T have equal determinants. Thus, the statement is true for all n. ∎

Theorem 5. *If A and B are $n \times n$ matrices, then $det(AB) = det(A)det(B)$.*

Theorem 6. *Cramer's Rule: Let A be an invertible $n \times n$ matrix. For any b in $\mathbb{R}^n$, the unique solution x of $Ax = b$ has entries given by*

$$x_i = \frac{det(A_i(b))}{det(A)} \text{ for } i = 1, 2, ..., n$$

where $A_i(b)$ denotes the matrix obtained by replacing A's ith column with b.

Proof. Denote the columns of A by $a_1, ..., a_n$ and the columns of the $n \times n$ identity matrix by $e_1, ..., e_n$. If $Ax = b$, the definition of matrix multiplication tells us

$$\begin{aligned} A \cdot I_i(x) = A\begin{bmatrix} e_1 & \cdots & x & \cdots & e_n \end{bmatrix} &= \begin{bmatrix} Ae_1 & \cdots & Ax & \cdots & Ae_n \end{bmatrix} \\ &= \begin{bmatrix} a_1 & \cdots & b & \cdots & a_n \end{bmatrix} = A_i(b) \end{aligned}$$

Using the multiplicative property of determinants,

$$(det(A))(det(I_i(x))) = det(A_i(b))$$

Since $det(I_i(x))$ is x, we just divide by $det(A)$. ∎

4 Vector Spaces

Some of this is from Chapter 1, but I think it makes more sense to define these concepts here.

Definition 4.1. A **subspace** of $\mathbb{R}^n$ is any set H in $\mathbb{R}^n$ that is closed under addition and scalar multiplication. That is:

1. The zero vector is in H
2. For each u and v in H, the sum $u + v$ is in H
3. For each u in H and each scalar c, the vector cu is in H

Definition 4.2. The **column space** of an $m \times n$ matrix A is the set of all linear combinations of the columns of A, denoted by $Col(A)$. Since the columns of A are in $\mathbb{R}^m$, the columns space is in $\mathbb{R}^m$.

Definition 4.3. The **null space** of a matrix A is the set of all solutions for $Ax = 0$, denoted by $Nul(A)$. When $Nul(A)$ contains nonzero vectors, the number of vectors in the nullspace equals the number of free variables in $Ax = 0$.

Definition 4.4. A **basis** for a subspace H in $\mathbb{R}^n$ is a linearly independent set in H that spans H.

Using the basis for a subspace H is preferable because any vector in H can only be written in one way as a linear combination of the basis vectors.

Proof. Suppose $\mathbb{B} = \{ b_1 \ , ..., b_p\}$ is a basis for H, and suppose a vector x in H can be generated in two ways:

$$x = c_1 b_1 + \cdots + c_p b_p \text{ and } x = d_1 b_1 + \cdots + d_p b_p$$

Subtracting gives us:

$$0 = (c_1 - d_1)b_1 + \cdots + (c_p - d_p)b_p$$

Since $\mathbb{B}$ is linearly independent, the weights must all be zero, so $c_j = d_j$ so the two representations are really just the same. ■

Theorem 7. *The pivot columns of a matrix A form a basis for $Col(A)$.*

Proof. Let B be the reduced echelon form of A. The set of pivot columns of B is linearly independent, since no vector is a linear combination of the vectors that precede it. Since A is *row equivalent* to B, the pivot columns of A are linearly independent as well. Thus, the nonpivot columns of A can be discarded from the spanning set of $Col(A)$.

Warning: The pivot columns of A are only evident when A has been reduced to *echelon* form. After reducing, make sure to use the **pivot columns of A itself** for the basis of $Col(A)$. The columns of an echelon form of A are often not in the column space of A. ■

Theorem 8. *Unique Representation Theorem: Let $\mathbb{B} = \{b_1, ..., b_n\}$ be a basis for a vector space V. Then for each x in V, there exists a unique set of scalars $c_1, ..., c_n$ such that*

$$x = c_1 b_1 + \cdots + c_n b_n$$

Proof. Since $\mathbb{B}$ spans V, we know there exist scalars such that we can form x. Assume x also has the representation

$$x = d_1 b_1 + \cdots + d_n b_n$$

Then, after subtracting we have

$$0 = (c_1 - d_1)b_1 + \cdots + (c_n - d_n)b_n$$

Since $\mathbb{B}$ is linearly independent, these weights must all be zero so $c_j = d_j$. ■

Because of the unique representation of each vector x in a basis, we can define the coordinates of x relative to the basis $\mathbb{B}$ as the weights $c_1, ..., c_n$.

Definition 4.5. Changing coordinates:

$$x = \mathbb{B}[x]_{\mathbb{B}}$$

where $\mathbb{B}$ denotes the matrix whose columns are basis vectors, and $x_{\mathbb{B}}$ denotes the x vector represented by basis coordinates.

To understand this, let $b_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, b_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, x = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, and $\mathbb{B} = \{b_1, b_2\}$. To find $[x]_{\mathbb{B}}$ of x relative to $\mathbb{B}$,

$$c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

or

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

Since we know the columns of $\mathbb{B}$ are linearly independent, it must be invertible so we can multiply x by $\mathbb{B}^{-1}$ to get $[x]_{\mathbb{B}}$.

Definition 4.6. Change of basis: We can generalize the above further. Let $\mathbb{B} = \{b_1, ..., b_n\}$ and $\mathbb{C} = \{c_1, ..., c_n\}$ be bases of a vector space V. Then there is a *unique* $n \times n$ matrix $\underset{\mathbb{C}\to\mathbb{B}}{P}$ such that

$$[x]_{\mathbb{C}} = \underset{\mathbb{C}\to\mathbb{B}}{P}[x]_{\mathbb{B}}$$

The columns of $\underset{\mathbb{C}\to\mathbb{B}}{P}$ are the $\mathbb{C}$-coordinate vectors of the vectors in the basis $\mathbb{B}$, that is

$$\underset{\mathbb{C}\to\mathbb{B}}{P} = \Big[[b_1]_{\mathbb{C}} \cdots [b_n]_{\mathbb{C}}\Big]$$

Definition 4.7. An **isomorphism** from V to W is a one-to-one linear transformation.

Definition 4.8. The **dimension** of a nonzero subspace H is the number of vectors in any basis for H. The dimension of the zero subspace $\{0\}$ is defined to be zero.

Definition 4.9. The **rank** of a matrix A is the dimension of the column space of A.

Theorem 9. *If a matrix A has n columns, then $Rank(A) + Dim(Nul(A)) = n$.*

Proof. An intuitive understanding for this can be achieved by restating the theorem as follows:

$$\Big(\text{num of pivot columns}\Big) + \Big(\text{num of nonpivot columns}\Big) = \Big(\text{num of columns}\Big)$$

■

Definition 4.10. If A is an $m \times n$ matrix, each row has n entries and can be understood as a vector in $\mathbb{R}^n$. The set of all linear combinations of the row vectors is called the **row space** of A, denoted by $Row(A)$. Note that $Row(A) = Col(A^T)$.

5 Eigenvalues and Eigenvectors

Definition 5.1. An **eigenvector** of an $n \times n$ matrix A is a nonzero vector x such that $Ax = \lambda x$. The scalar λ is called an **eigenvalue** of A if and only if the equation

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

has a nontrivial solution. The set of all solutions to this equation is the null space of the matrix $A - \lambda I$; this subspace of $\mathbb{R}^n$ is called the **eigenspace** of A.

Theorem 10. *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

Proof. Consider the 3×3 case. If A is upper triangular, then

$$A - \lambda I = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix}$$

The scalar λ is an eigenvalue if and only if $(A - \lambda I)x = 0$ has a nontrivial solution, which means the equation must have a free variable, which would only occur if at least one of the values on the diagonal is zero. ■

Theorem 11. *The eigenvectors of A, $v_1, \ldots, v_r$, that correspond to distinct eigenvalues, $\lambda_1, \ldots, \lambda_r$, are linearly independent.*

Proof. Suppose $\{v_1, \ldots, v_r\}$ is linearly dependent. Let p be the least index such that v_{p+1} is a linear combination of the preceding linearly independent eigenvectors. Then there exist scalars such that

$$c_1 v_1 + \cdots + c_p v_p = v_{p+1}$$

Multiply both sides by A, using the fact that $Av_k = \lambda_k v_k$, to get

$$c_1 \lambda_1 v_1 + \cdots + c_p \lambda_p v_p = \lambda_{p+1} v_{p+1}$$

We can also multiply both sides of our first equation by λ_{p+1} and then subtract to get

$$c_1(\lambda_1 - \lambda_{p+1})v_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})v_p = 0$$

Because we assumed v_{p+1} was the first linearly dependent eigenvector, the set $\{v_1, \ldots, v_p\}$ must be linearly independent. That means all $(\lambda_k - \lambda_{p+1})$ should be 0, but because the eigenvalues are distinct they cannot be. Thus, we arrive at a contradiction. ■

Remember that to find eigenvalues, we need to find scalars λ such that

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

has a nontrivial solution. This is equivalent to the matrix $A - \lambda I$ being not invertible, which is equivalent to $det(A - \lambda I) = 0$. Writing the determinant as a polynomial involving only λ is called the characteristic equation of a matrix.

Theorem 12. *Diagonalization Theorem: An $n \times n$ matrix A is* ***diagonalizable*** *if and only if A has n linearly independent eigenvectors. If this condition is met, we can write*

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

where P is a matrix whose columns are n linearly independent eigenvectors of A, and D is a diagonal matrix whose diagonal entries are corresponding eigenvalueus of A.

Proof. Right multiplying both sides by P gives us $AP = PD$.

$$AP = A\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} Av_1 & \cdots & Av_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \cdots & \lambda_n v_n \end{bmatrix}$$

$$PD = P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \cdots & \lambda_n v_n \end{bmatrix}$$

Since these are equal and P has an inverse because its columns are linearly independent eigenvectors, $A = PDP^{-1}$. ■

6 Orthogonality and Least Squares

Definition 6.1. The **inner product** of u and v is

$$u \cdot v = \begin{bmatrix} u_1 \cdots u_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \cdots + u_n v_n$$

Properties of inner products: Let u, v, and w be vectors in $\mathbb{R}^n$. Then

1. $u \cdot v = v \cdot u$
2. $(u + v) \cdot w = u \cdot w + v \cdot w$
3. $u \cdot u \geq 0$

Definition 6.2. The **norm** of v is $||v||$ defined by

$$||v|| = \sqrt{v \cdot v}$$

The distance between v and u is $||u - v||$.

Definition 6.3. Two vectors u and v are **orthogonal** if $u \cdot v = 0$.
Note the zero vector is orthogonal to every other vector.
The Pythagorean Theorem tells us that if u and v are orthogonal then

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

.

The set of all vectors z that are orthogonal to W is called the **orthogonal complement** of W, denoted by $W^{\perp}$.

Theorem 13. *Let A be an $m \times n$ matrix. Then*

$$(Row(A))^{\perp} = Nul(A) \text{ and } (Col(A))^{\perp} = Nul(A^T)$$

Proof. If x is in $Nul(A)$, then $\begin{bmatrix} Ax_1 & Ax_2 & \cdots & Ax_n \end{bmatrix}$, which implies that x is orthogonal to all the rows in A. Conversely, if x is orthogonal to $Row(A)$, then clearly $Ax = 0$. A similar proof can be shown for the second statement. ∎

Definition 6.4. In $\mathbb{R}^2$ or $\mathbb{R}^3$, $u \cdot v = ||u||||v||cos(\theta)$. To see this, we can use the law of cosines,

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

Definition 6.5. The set of vectors $\{u_1, ..., u_p\}$ is called an **orthogonal set** if each pair of distinct vectors is orthogonal.

Theorem 14. *If $S = \{u_1, ..., u_p\}$ is an orthogonal set of nonzero vectors, then S is linearly independent.*

Proof. We know

$$0 = c_1 u_1 + \cdots + c_p u_p$$

Multiplying by u_1,

$$0 = (c_1 u_1 + \cdots + c_p u_p) \cdot u_1$$

$$0 = (c_1 u_1) \cdot u_1 + (c_2 u_2) \cdot u_1 + \cdots + (c_p u_p) \cdot u_1$$

$$0 = c_1 (u_1 \cdot u_1)$$

Since u_1 is nonzero, $(u_1 \cdot u_1)$ must be nonzero, so c_1 must be 0. A similar proof can be used to show $c_2, ..., c_p$ must be zero. Thus, S is linearly independent. ∎

Definition 6.6. An **orthogonal basis** for subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

Theorem 15. *Let $\{u_1, \ldots, u_p\}$ be an orthogonal basis for subspace W of $\mathbb{R}^n$. For each y in W, the weights c_k in*

$$y = c_1 u_1 + \cdots + c_p u_p$$

are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j} \textit{ for } (j = 1, \ldots, p)$$

Proof. Taking the dot product of u_1 on both sides,

$$y \cdot u_1 = (c_1 u_1 + c_2 u_2 + \cdots c_p u_p) \cdot u_1 = c_1(u_1 \cdot u_1)$$

Since u_1 is nonzero, we can divide by $(u_1 \cdot u_1)$ to solve for c_1. The same proof can be used to solve $c_2, \ldots, c_p$. ■

Definition 6.7. An set is **orthonormal** if all its vectors are unit vectors and are orthogonal to one another. Let U be an $m \times n$ matrix with orthonormal columns. Then,

1. $\|Ux\| = \|x\|$
2. $(Ux) \cdot (Uy) = x \cdot y$

An orthonormal basis can be constructed from an orthogonal basis $\{v_1, \ldots, v_p\}$ by simply normalizing all the v_k.

Theorem 16. *An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$. If U is a square matrix, then it is called an **orthogonal matrix** and has $U^{-1} = U^T$.*

Proof. We will prove with a simpler version of U with only three columns, but the proof can generalize. Let $U = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$. Then

$$U^T U = \begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{bmatrix}$$

By definition of orthonormal vectors, only the diagonal entries simplify to 1 and all other entries simplify to 0. ■

Definition 6.8. Orthogonal projection: Consider representing a nonzero vector u in $\mathbb{R}^n$ as the sum of two vectors, one a multiple of some vector y and the other orthogonal to y. That is,

$$u = \alpha y + (u - \alpha y)$$

This means $u - \alpha y$ is orthogonal to y if and only if

$$0 = (u - \alpha y) \cdot y = u \cdot y - \alpha(y \cdot y)$$

That is, in order for $(u - \alpha y)$ to be orthogonal to y,

$$\alpha = \frac{u \cdot y}{y \cdot y} \text{ and } y = \frac{u \cdot y}{y \cdot y} y$$

The vector y is called the **orthogonal projection of** u **onto** y.

$$y = proj_L y = \frac{y \cdot u}{y \cdot y} y$$

Definition 6.9. Orthogonal Decomposition We can extend projections to subspaces. Let W be a subspace of $\mathbb{R}^n$. Then each y in $\mathbb{R}^n$ can be written

$$y = \hat{y} + z$$

where $\hat{y}$ is in W and z is in $W^{\perp}$. If $\{u_1, ..., u_p\}$ is an orthogonal basis of W, then

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \cdot + \frac{y \cdot u_p}{u_p \cdot u_p} u_p$$

and $z = y - \hat{y}$.

The vector $\hat{y}$ is called the **orthogonal projection of** y **onto** W and is written as $proj_W y$.

Note the the denominator $u_k \cdot u_k = 1$ if W is an orthonormal basis. It follows that for orthonormal bases of W,

$$proj_W y = (y \cdot u_1)u_1 + (y \cdot u_2)u_2 + \cdots + (y \cdot u_p)u_p = UU^T y$$

Theorem 17. *The Gram-Schmidt Process: The Gram-Schmidt process is an algorithm for producing an orthogonal or orthonormal basis for a nonzero subspace of* $\mathbb{R}^n$.

Given a basis $\{x_1, ..., x_p\}$ *for nonzero subspace* W *of* $\mathbb{R}^n$*, define*

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$
$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$
$$\vdots$$
$$v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \cdots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}$$

Then $\{v_1, ..., v_p\}$ *is an orthogonal basis for* W. *In addition*

$$Span(\{v_1, ..., v_k\}) = Span(\{x_1, ..., x_k\}) \text{ for } 1 \leq k \leq p$$

Proof. We proceed by induction. If we set $v_1 = x_1$, then $Span(\{v_1\}) = Span(\{x_1\})$. Suppose for some $k < p$, we construct $v_1, ..., v_k$ so that $\{v_1, ..., v_k\}$ is an orthogonal basis for W_k. Define

$$v_{k+1} = x_{k+1} - proj_{W_k} x_{k+1}$$

By Orthogonal Decomposition, v_{k+1} is orthogonal to W_k. Since $proj_{W_k} x_{k+1}$ is in W_k, it is also in W_{k+1}, which means v_{k+1} is also in W_{k+1} since W_{k+1} is a subspace which must be closed under subtraction. $v_{k+1} \neq 0$ because x_{k+1} is not in $W_k = Span(\{x_1, ..., x_k\})$. Thus, $\{v_1, ..., v_k\}$ is an orthogonal set of nonzero vectors in $k+1$-dimensional subspace W_{k+1}, so they must be a basis for W_{k+1}. Thus, the Gram-Schmidt algorithm yields an orthogonal basis by induction. ∎

Theorem 18. *If* A *is an* $m \times n$ *matrix with linearly independent columns, then we can factor* $A = QR$, *where* Q *is an* $m \times n$ *matrix whose columns form an orthonormal basis for* $Col(A)$ *and* R *is an* $n \times n$ *upper triangular invertible matrix with positive entries on its diagonal.*

Proof. The columsn of A form a basis $\{x_1, ..., x_n\}$ for $Col(A)$. We can use Gram-Schmidt to construct an orthogonal basis for $W = Col(A)$ and then scale this basis to get an orthonormal basis for $Col(A)$

$$Q = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$$

For $1 \leq k \leq n$, $x_k \in Span(Q)$, that is, there are constants such that

$$x_k = r_{1k} u_1 + \cdots + r_{kk} u_k + 0 \cdot u_{k+1} + \cdots + 0 \cdot u_n$$

We can assume $r_{kk} \geq 0$, if it isn't, just multiply both r_{kk} and u_k by -1. Thus,

$$r_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

So, $x_k = Qr_k$. Let $R = \begin{bmatrix} r_1 & \cdots & r_n \end{bmatrix}$. Finally,

$$A = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} Qr_1 & \cdots & Qr_n \end{bmatrix} = QR$$

■

Definition 6.10. If A is an $m \times n$ matrix and b is in $\mathbb{R}^m$, a **least-squares solution** of $Ax = b$ is an $\hat{x}$ in $\mathbb{R}^n$ such that

$$||b - A\hat{x}|| \leq ||b - Ax||$$

for all x in $\mathbb{R}^n$. This should hold even if b is outside $Col(A)$.

Theorem 19. *Let A be an $m \times n$ matrix. The following are logically equivalent:*

1. *$Ax = b$ has a unique least-squares solution for each b in $\mathbb{R}^m$.*

2. *The columns of A are linearly independent.*

3. *The matrix $A^T A$ is invertible*

When these are true, the least-squares solution $\hat{x}$ is given by

$$\hat{x} = (A^T A)^{-1} A^T b$$

Proof. To get the best approximation of b on A, we can take the projection

$$\hat{b} = proj_{Col(A)} b$$

Now we know the equation $A\hat{x} = \hat{b}$ is consistent, and that $\hat{x}$ would be the least-squares solution. We know $b - \hat{b}$ is orthogonal to $Col(A)$, so $b - A\hat{x}$ is orthogonal to each column of A, that is

$$A^T (b - A\hat{x}) = 0$$

Simplifying

$$A^T b - A^T A\hat{x} = 0$$

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

$$\hat{x} = (A^T A)^{-1} A^T b$$

■

Theorem 20. *When the columns of A are orthogonal, as they often are in linear regression problems, then we can use QR factorization to produce a computationally easier calculation.*
Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be the QR factorization. For any b in $\mathbb{R}^m$, the least-squares solution is given by

$$\hat{x} = R^{-1} Q^T b$$

Proof. Let $\hat{x} = R^{-1} Q^T b$. Then

$$A\hat{x} = QR\hat{x} = QRR^{-1} Q^T b = QQ^T b$$

Since the columns of Q form an orthonormal basis for $Col(A)$, $QQ^T b$ is the orthogonal projection of b onto $Col(A)$ by Definition 6.9. Hence, it is a least-squares solution. ■

Definition 6.11. Inner product: **Inner products** on a vector space V is a function that, to each pair of vectors u and v, associates a real number $\langle u, v \rangle$ and satisfies the following for vectors u, v, w in V and scalar c:

1. $\langle u, v \rangle = \langle v, u \rangle$
2. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
3. $\langle cu, v \rangle = c\langle u, v \rangle$
4. $\langle u, u \rangle \geq 0$

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

Theorem 21. *The Cauchy-Schwartz Inequality: For all u, v in V,*

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

Proof. If $u = 0$, both sides are zero, so the inequality is true. If $u \neq 0$, let W be the subspace spanned by u, then

$$\|proj_W v\| = \|\frac{\langle v, u\rangle}{\langle u, u\rangle} u\| = \frac{|\langle v, u\rangle|}{\|u\|^2}\|u\| = \frac{|\langle u, v\rangle|}{\|u\|}$$

Since $\|proj_W v\| \leq \|v\|$, we simplify the above to

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

, which proves the theorem. ∎

Theorem 22. *The Triangle Inequality: For all u, v in V,*

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

Proof.

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle = \langle u, u\rangle + 2\langle u, v\rangle + \langle v, v\rangle \\ &\leq \|u\|^2 + 2\,|\langle u, v\rangle| + \|v\|^2 \\ &\leq \|u\|^2 + 2\|u\|\,\|v\| + \|v\|^2 \\ &= (\,\|u\| + \|v\|\,)^2 \end{aligned} \tag{1}$$

where the second to last step follows from the Cauchy-Schwartz Inequality. ∎

7 Symmetric Matrices and Quadratic Forms

Definition 7.1. A **symmetric** matrix is a matrix A such that $A^T = A$. This condition necessitates A be a square matrix. Its main diagonal entries are arbitrary, but the other entries must occur in pairs on opposite sides of the main diagonal.

Theorem 23. *If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.*

Proof. Let v_1 and v_2 be eigenvectors that correspond to distinct eigenvalues λ_1 and λ_2.

$$\lambda_1 v_1 \cdot v_2 = (\lambda_1 v_1)^T v_2 = (Av_1)^T v_2 = (v_1^T A^T) v_2 = v_1^T (Av_2) = v_1^T(\lambda_2 v_2) = \lambda_2 v_1 \cdot v_2$$

Thus, $(\lambda_1 - \lambda_2) v_1 \cdot v_2 = 0$, so since the eigenvalues are distinct, the eigenvectors must be orthogonal. ∎

Note that this means when we diagonalize into $A = PDP^{-1}$, P is n orthogonal matrix so $P^{-1} = P^T$. If a matrix A can be written as

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

then it is called **orthogonally diagonalizable**.
A is orthogonally diagonalizable if and only if A is symmetric. Proving "if" is complicated but to prove "only if" note that

$$A^T = (PDP^T)^T = P^{TT}D^TP^T = PDP^T = A$$

Theorem 24. *The Spectral Theorem for Symmetric Matrices: The set of eigenvalues of a matrix A are sometimes referred to as the **spectrum** of A.*
An $n \times n$ symmetric matrix A has the following properties:

1. *A has n real eigenvalues, counting multiplicities.*
2. *The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of the λ.*
3. *The eigenspaces are mutually orthogonal.*
4. *A is orthogonally diagonalizable.*

Definition 7.2. A **quadratic form** of $\mathbb{R}^n$ is a function Q whose value at a vector x can be computed by an expression of the form $Q(x) = x^TAx$, where A is an $n \times n$ symmetric matrix called the **matrix of the quadratic form**.
The simplest form of this is $Q(x) = x^TIx = ||x||^2$.