University of California, Berkeley

Literally everything I know about

Linear Algebra

Warren Alphonso

A very reductionist summary of Linear Algebra and its Applications by Lay, Lay, and McDonald, as well as Linear Algebra Done Wrong by Treil.

May 30, 2019

Contents

Chapter 1

Basic Notations

1.1 Vector Spaces

Definition 1.1.1. A **vector space** V is a collection of vectors, along with vector addition and scalar multiplication defined such that for vectors u, v, and w:

1. Commutative: $v + w = w + v$
2. Associative: $(u + v) + w = u + (v + w)$
3. Zero vector: $v + 0 = v$
4. Additive inverse: $v + w = 0$
5. Multiplicative identity: $1v = v$
6. Multiplicative associative: $(\alpha\beta)v = \alpha(\beta v)$
7. Distribution of scalars: $\alpha(u + v) = \alpha u + \alpha v$
8. Distribution of vectors: $(\alpha + \beta)u = \alpha u + \beta u$

These properties are simply to ensure that vector spaces are **abelian groups**.

Definition 1.1.2. An $m \times n$ **matrix** is an array with m rows and n columns. Elements of a matrix are called *entries*. Given a matrix A, its **transpose** is defined as the matrix whose columns are A's rows, so A^T is an $n \times m$ matrix.

1.2 Linear Combinations

Definition 1.2.1. A **linear combination** of vectors $v_1, \cdots, v_p \in V$ is a sum of the form

$$\alpha_1 v_1 + \cdots + \alpha_p v_p = \sum_{k=1}^{p} \alpha_k v_k$$

Definition 1.2.2. A set of vectors $v_1, \cdots, v_n$ is said to be **linearly independent** if the equation

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$$

has only the trivial solution.

Definition 1.2.3. A **basis** is a set of vectors $v_1, \cdots, v_n \in V$ such that any vector $u \in V$ has a *unique* representation as a linear combination

$$u = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

The coefficients $\alpha_1, \cdots, \alpha_n$ are called *coordinates* of u.

Fundamentally, our definition of basis states that a basis must be spanning and unique. In order for a representation to be unique, we know the basis must be linearly independent.

Theorem 1.2.1. *A set of vectors $v_1, \cdots, v_p \in V$ is a basis if and only if it is linearly independent and complete (spanning).*

Proof. We already know basis must be linearly independent and spanning, so we just need to prove the other direction.
Suppose the set $v_1, \cdots, v_p$ is linearly independent. Then we know for some vector $u \in V$:

$$u = \sum_{k=1}^{n} \alpha_k v_k$$

All that is remaining is to prove this representation is unique.
Suppose there is another representation, $u = \sum_{k=1}^{n} \beta_k v_k$. Then,

$$\sum_{k=1}^{n} (\alpha_k - \beta_k) v_k = v - v = 0$$

Since the set is linearly independent, we know $\alpha_k - \beta_k = 0$. Thus, the representation is unique. ■

1.3 Linear Transformations

Definition 1.3.1. A **transformation** T from set X to set Y assigns a value $y \in Y$ for every value $x \in X$: $y = T(x)$. X is called the *domain* of T, Y is called the *codomain* of T, and the set of all $T(x)$ is called the *range* of T.
Let V, W be vector spaces. A transformation $T : V \to W$ is **linear** if:

1. $T(u + v) = T(u) + T(v)$
2. $T(\alpha v) = \alpha T(v)$

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

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

We can represent linear transformations with matrices. To represent a transformation $T : \mathbb{F}^n \to \mathbb{F}^m$, we need to only how our n basis vectors are transformed. To see this, note that any vector $u = \alpha_1 v_1 + \cdots + \alpha_n v_n$. So $T(x) = \alpha 1 T(v_1) + \cdots + \alpha_n T(v_n)$. If we join the vectors $T(v_1), \cdots, T(v_n)$ in a matrix $A = \begin{bmatrix} T(v_1) & \cdots & T(v_n) \end{bmatrix}$, we have captured all the information about T.

Definition 1.3.2. There are two ways to approach **matrix-vector multiplication**:
Column by coordinate rule: Multiply each column of the matrix by the corresponding coordinate of the vector and add.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 2 \end{bmatrix} + 3 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$$

Row by column rule: To get entry k of the result, multiply row k of the matrix with the column.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 \\ 3 \cdot 1 + 2 \cdot 2 + 1 \cdot 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$$

Definition 1.3.3. The natural extension to **matrix multiplication** of two matrices AB is to multiply A by each column of B.

$$AB = A \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix} = \begin{bmatrix} Ab_1 & \cdots & Ab_n \end{bmatrix}$$

Using the *row by columns rule*, we can see that the $(AB)_{j,k} =$ (row j of A)$\cdot$(column k of B). This also means AB is only defined if A is $m \times n$ and B is $n \times r$.
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. $\alpha(AB) = (\alpha A)B = A(\alpha B)$
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 1.3.4. The **transpose** of A is the matrix whose columns are formed from the corresponding rows of A, denoted by A^T.

1. $(A^T)^T = A$
2. $(A+B)^T = A^T + B^T$
3. $(cA)^T = cA^T$
4. $(AB)^T = B^T A^T$

To understand the final property, let AB denote a $n \times m$ matrix so that

$$AB = \begin{bmatrix} A_{1*} \cdot B_1 & A_{1*} \cdot B_2 & \cdots & A_{1*} \cdot B_m \\ A_{2*} \cdot B_1 & A_{2*} \cdot B_2 & \cdots & A_{2*} \cdot B_m \\ \vdots & \vdots & & \vdots \\ A_{n*} \cdot B_1 & A_{n*} \cdot B_2 & \cdots & A_{n*} \cdot B_m \end{bmatrix}$$

$$(AB)^T = \begin{bmatrix} A_{1*} \cdot B_1 & A_{2*} \cdot B_1 & \cdots & A_{n*} \cdot B_1 \\ A_{1*} \cdot B_2 & A_{2*} \cdot B_2 & \cdots & A_{n*} \cdot B_2 \\ \vdots & \vdots & & \vdots \\ A_{1*} \cdot B_m & A_{2*} \cdot B_m & \cdots & A_{n*} \cdot B_m \end{bmatrix}$$

where $A_{\alpha*}$ denotes the αth row of a A. Note that $(AB)^T_{jk}$ = (row k of A) $\cdot$ (column j of B) = (row j of B^T) $\cdot$ (column k of A^T).

Definition 1.3.5. For a *square* matrix A, its **trace** is the sum of its diagonal entries.

$$trace(A) = \sum_{k=1}^{n} a_{k,k}$$

Theorem 1.3.1. *Let A and B be size $m \times n$ and $n \times m$, respectively. Then*

$$trace(AB) = trace(BA)$$

Proof. We need only show that the diagonal entries of AB are the same as the diagonal entries of BA. Using the *row by columns rule*, we know $(AB)_{j,k}$ = (row j of A)$\cdot$(column k of B), so when applying to diagonal entries we get $(AB)_{k,k} = (BA)_{k,k}$ = (row k of A) $\cdot$ (column k of B) ■

1.4 Invertible Transformations and Isomorphisms

Definition 1.4.1. An $n \times n$ matrix A is **invertible** if there is an $n \times n$ matrix A^{-1} such that $A^{-1}A = I$.

Theorem 1.4.1. *If A and B are invertible (and such that AB is defined), then the product AB is invertible and*

$$(AB)^{-1} = B^{-1}A^{-1}$$

Proof. Direct computation:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$$

and

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I$$

■

Theorem 1.4.2. *If A is invertible, then A^T is also invertible and*

$$(A^{-1})^T = (A^T)^{-1}$$

Proof. Using $(AB)^T = B^T A^T$,

$$(A^{-1})^T(A^T) = (AA^{-1})^T = I$$

and

$$A^T(A^{-1})^T = (A^{-1}A)^T = I$$

■

Definition 1.4.2. An invertible linear transformation $A : V \to W$ is called an **isomorphism**. The two vector spaces V and W for which A is defined are called **isomorphic**, denoted $V \cong W$.

Isomorphic spaces can be considered different representations of the *same* space. The following theorem is an example.

Theorem 1.4.3. *Let $A : V \to W$ be an isomorphism, and let $v_1, \cdots, v_n$ be a basis in V. Then $Av_1, \cdots, Av_n$ is a basis in W.*

Proof. Because V and W are isomorphic, every $w \in W$ can be represented as some $v \in V$ by applying A^{-1}. For arbitrary $w \in W$

$$A^{-1}w = v = \sum_{k=1}^{n} \alpha_k v_k$$

Then we apply A to get

$$Av = w = \sum_{k=1}^{n} \alpha_k A v_k$$

■

Chapter 2

Systems of Linear Equations

2.1 Representations of Linear Systems

The first understanding of a *linear system* is simply a collection of m linear equations with n unknowns $x_1, \cdots, x_n$. To solve this system entails finding all n-tuples of numbers $x_1, \cdots, x_n$ which satisfy the m equations simultaneously. If we define

$$A = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,n} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,n} \\ \vdots & \vdots & & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \cdots & \alpha_{m,n} \end{bmatrix}$$

then we can summarize our linear system in matrix form

$$Ax = b$$

The above is the **coefficient matrix**. If we want to contain all the information in a single matrix, we can use an **augmented matrix**

$$\left[\begin{array}{cccc|c} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,n} & b_1 \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \cdots & \alpha_{m,n} & b_m \end{array}\right]$$

2.2 Solving Linear Systems

Linear systems are solved using **Gaussian elimination**. We can perform the following row operation on an augmented matrix:

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.

These operations belong to the *elementary matrices*: any operation can be described by applying the same operation to I to get E and then multiplying EA.

Definition 2.2.1. For an augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 3 & 2 & 1 & 7 \\ 2 & 1 & 2 & 1 \end{array}\right]$$

the **echelon form** is

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 2 & -4 \end{array}\right]$$

and the **reduced echelon form** is

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

Formally, the **echelon form** requires that all zero rows are below all nonzero rows and that any nonzero row's **pivot**, its leading entry, is strictly to the right of the leading entry in the previous row. The particular echelon form above is called **triangular form** and is only possible when we have a square matrix. The **reduced echelon form** requires echelon form in addition to maintaining that all pivot entries are 1 and that all entries above each pivot are 0.

The *existence* and *uniqueness* of a solution can be determined by analyzing pivots in the echelon form of a matrix.
When looking at the coefficient matrix:

1. A solution (if it exists) is unique if and only if there are no free variables, that is if the echelon form has a pivot in every *column*.

2. A solution is consistent if and only if the echelon form has a pivot in every *row*.

The first statement is trivial because free variables are responsible for all non-uniqueness. For the second statement, if we have a row with no pivots in the echelon form of a matrix, we have $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & b_k \end{array}\right]$, which certainly has no solution. Thus, in order for a solution to *exist* and be *unique*, the echelon form must have a pivot in *every column and every row*.

Theorem 2.2.1. *Any linearly independent system of vectors in $\mathbb{F}^n$ cannot have more than n vectors in it.*

Proof. Let a system $v_1, \cdots, v_m \in \mathbb{F}^n$ be linearly independent and let $A = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix}$ be $n \times m$. We must show that $x_1 v_1 + \cdots + x_m v_m = 0$, or equivalently $Ax = 0$, has unique solution $x = 0$. According to statement 1 above, a solution can only be unique if the echelon form has a pivot in every column. This is impossible if $m > n$. ∎

Theorem 2.2.2. *A matrix A is invertible if and only if its echelon form has pivot in every column and every row.*

Proof. Since a matrix must have unique solution for $Ax = b$ for any b in order to be invertible, it is necessary that the echelon form has pivot in every column and row, according to statements 1 and 2 above. ■

This directly implies that an invertible matrix **must be square**.
Since an invertible matrix must be square and must pivots in every row and column in echelon form, any invertible matrix is row equivalent to the identity matrix.
We can use this to get the following algorithm for finding A^{-1}:

1. Form an *augmented* $n \times 2n$ matrix $\left[A \mid I\right]$.
2. Perform row operations to transform A into I.
3. The matrix that was originally I will now be A^{-1}.

To fully understand this algorithm, remember that every row operation can be expressed as the left multiplication by an elementary matrix. Let $E = E_n \cdots E_2 E_1$ represent all the performed row operations. Since we know E transforms A to the identity matrix, we have $EA = I$, so $E = A^{-1}$. Since row operations affect the entire augmented matrix, we have $\left[A \mid I\right] \to \left[EA \mid EI\right] = \left[I \mid A^{-1}\right]$.

2.3 Fundamental Subspaces

Definition 2.3.1. A **subspace** of vector space V is a non-empty subset $V_0 \subset V$ which is also a vector space.

For any linear transformation $A : V \to W$, we can associate the following subspaces:

1. The *null space*, or *kernel*, of A which consists of all vectors $v \in V$ such that $Av = 0$.
2. The *range* of A which is the set of all vectors $w \in W$ which can be represented as $w = Av$ for $v \in V$.

By the *column by coordinate rule*, we know that any vector in $Range(A)$ can be represented as a weighted sum of the column vectors of A, which is why the term Column Space is sometimes used to refer to Range.
In addition, we can consider the corresponding subspaces of the transposed matrix. The term *row space* is used to denote $Range(A^T)$, and the term *left null space* is used to denote $Null(A^T)$. Together, these four subspaces are known as the **fundamental subspaces** of the matrix A.

Definition 2.3.2. The **dimension** of a vector space V, denoted $dim(V)$, is the number of vectors in a basis.

Theorem 2.3.1. *General solution of a linear equation Let a vector x_1 denote a solution to the equation $Ax = b$, and let H be the set of all solutions of $Ax = 0$. Then the set*

$$x = x_1 + x_h : x_h \in H$$

is the set of all solutions of the equation $Ax = b$.
In other words,

$$\Big(\text{General solution of } Ax = b\Big) = \Big(\text{A particular solution of } Ax = b\Big) + \Big(\text{General solution of } Ax = 0\Big)$$

Proof. We know $Ax_1 = b$ and $Ax_h = 0$. For $x = x_1 + x_h$,

$$Ax = A(x_1 + x_h) = Ax_1 + Ax_h = b + 0 = b$$

Therefore, any solution x for $Ax = b$ can be represented as $x = x_1 + x_h$ with some $x_h \in H$. ■

The power of this theorem is its generality - it applies to all linear equations. Aside from showing the structure of the solution set, this theorem allows us to separate investigation of uniqueness from existence. To study existence, we only need to analyze uniqueness of $Ax = 0$, which always has a solution.

Theorem 2.3.2. *In order to compute the fundamental subspaces, we need to do row reduction. Let A be the original matrix and let A_e be its echelon form.*

1. *The pivot columns of the original matrix A (ie the columns where after row operations we will have pivots in echelon form) give us a basis for $Range(A)$.*
2. *The pivot rows of A_e give us the basis in row space.*
3. *To find $Null(A)$, we need to solve $Ax = 0$.*

Proof. In turn,

1. We know the pivot columns of A_e form a basis for $Range(A_e)$. Since $A_e = EA$ (E is the matrix product of the elementary matrices representing the row operations completed), $A = E^{-1}A_e$. This means the corresponding columns in A of A_e is a basis of A.
2. We know that the pivot rows of the echelon form are linearly independent. Now we need only prove that they span the entirety of the row space. Notice that *row operations do not change the row space*. To prove this,

   $$A_e = EA$$

   where A is $m \times n$ and E is an $m \times m$ invertible matrix.

   $$Range(A_e^T) = Range(A^T E^T) = A^T(Range(E^T)) = A^T(\mathbb{R}^m) = Range(A^T)$$

where the final step follows from applying an $n \times m$ matrix to $\mathbb{R}^m$, which is just a transformation from $\mathbb{R}^m$ to $Range(A^T)$.

3. Solving for $Ax = 0$ certainly gives us a spanning set for $Null(A)$. To prove the set is linearly independent, multiply each vector by its corresponding free variable and add. For every free variable x_k, the entry k is exactly x_k, so the only way the sum of the set is 0 is if all the free variables are 0.

■

As an example of these computations, consider the matrix

$$\begin{bmatrix} 1 & 1 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 2 \\ 1 & 1 & -1 & -1 & 0 \end{bmatrix}$$

Performing row operations, we get the echelon form

$$\begin{bmatrix} 1 & 1 & 2 & 2 & 1 \\ 0 & 0 & -3 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So the first and third columns of the *original matrix* give us a basis for $Range(A)$:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \\ -1 \end{bmatrix}$$

We also know the basis for $Row(A)$ is the first and second row of the *echelon form*:

$$\begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -3 \\ -3 \\ -1 \end{bmatrix}$$

To find $Null(A)$ we solve $Ax = 0$. The reduced echelon form is

$$\begin{bmatrix} 1 & 1 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This means

$$\begin{Bmatrix} x_1 = -x_2 - \frac{1}{3}x_5 \\ x_2 \text{ is free} \\ x_3 = -x_4 - \frac{1}{3}x_5 \\ x_4 \text{ is free} \\ x_5 \text{ is free} \end{Bmatrix} \longrightarrow x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -\frac{1}{3} \\ 0 \\ -\frac{1}{3} \\ 0 \\ 1 \end{bmatrix}$$

The vectors at each free variables form the basis for $Null(A)$.

Definition 2.3.3. The **rank** of a linear transformation A, denoted $rank(A)$, is the dimension of the range of A.

$$rank(A) := Dim(Range(A))$$

Theorem 2.3.3. *The Rank Theorem For a matrix A*

$$rank(A) = rank(A^T)$$

The proof of this is trivial since rank of both column space and row space are dependent on the number of pivots in echelon form.

Theorem 2.3.4. *Let A be an $m \times n$ matrix. Then*

1. $dim(Null(A)) + dim(Range(A)) = n$ *(dimension of domain)*
2. $dim(Null(A^T)) + dim(Range(A^T)) = dim(Null(A^T)) + rank(A) = m$ *(dimension of codomain)*

Proof. In turn,

1. The first equality is simply that the number of free variables + the number of pivots = the number of columns.
2. The second equality applies the Rank Theorem to prove the row counterpart to the first equality.

■

The following follows from the second statement in the above theorem.

Theorem 2.3.5. *Let A be an $m \times n$ matrix. Then the equation*

$$Ax = b$$

has a solution for every $b \in \mathbb{R}^m$ if and only if the dual equation

$$A^T x = 0$$

has only the trivial solution.

2.4 Change of Basis

Let V be a vector space with a basis $B := b_1, \cdots, b_n$. Recall that any vector $v \in V$ can be written

$$v = x_1 b_1 + \cdots + x_n b_n$$

where the numbers $x_1, \cdots, x_n$ are called the coordinates of v. We can write the *coordinate vector* as

$$[v]_B := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n$$

Note that $v \mapsto [v]_B$ is an isomorphism between V and $\mathbb{F}^n$.

Definition 2.4.1. Let $T : V \to W$ be a linear transformation, and let $A = \{a_1, \cdots, a_n\}$, $B = \{b_1, \cdots, b_m\}$ be bases in V, W respectively.
A **matrix of transformation** T **in bases** A **and** B is an $m \times n$ matrix, denoted by $[T]_{BA}$,

$$[Tv]_B = [T]_{BA}[v]_A$$

The matrix $[T]_{BA}$ is easy - its kth column is just $[Ta_k]_B$.

Definition 2.4.2. For the above two bases A and B, the **change of basis** is

$$[v]_B = [I]_{BA} v_A$$

where $[I]_{BA}$ is the **change of basis matrix** whose kth column is $[a_k]_B$.
Clearly, any change of basis is invertible and

$$[I]_{BA} = ([I]_{AB})^{-1}$$

Definition 2.4.3. We can use this to define **similar matrices** as matrices A, B such that

$$A = Q^{-1} B Q$$

This means we can treat similar matrices as different representations of the same linear operator.