Lecture Notes on

Linear Algebra

William Huang

Contents

1 Linear Equations in Linear Algebra

1.1 Systems of Linear Equations

Example: Example System of Linear Equations

Consider the following example system of linear equations

$$2x + y - z = 1$$

$$-x - 3y + 4z = 0$$

$$3x + 2y - 5z = -7$$

Definition: Augmented Matrix

We can convert this into the **augmented matrix**

$$A = \begin{bmatrix} 2 & 1 & -1 & 1 \\ -1 & -3 & 4 & 0 \\ 3 & 2 & -5 & -7 \end{bmatrix}$$

An augmented matrix consists of a coefficient matrix (defined below) as well as an extra column that contains the constants on the right hand sides of the equations. It is clear that augmented matrices always have size $n \times (n+1)$.

Definition: More Terms

- The left three columns in this matrix is called the **coefficient matrix** for this system of equations.
- Each element in a matrix is called an **entry**.
- A matrix's **size** is denoted in general by $n \times m$, where n and m are the number of rows and columns, respectively.

Note that matrices can be expressed as variables such as the capital letter A (it must be expressed as a capital letter). For the above matrix, the size is 3×4, because there are 3 rows and 4 columns. Each entry of the matrix A can be expressed as a_{ij}, where i is the row number and j is the column number (e.g. $a_{23} = 4$ in the above example).

Theorem: Adding and Subtracting Matrices.

Adding and subtracting can occur *only for matrices of the same size.* To add or subtract matrices, simply add or subtract corresponding entries. For example:

$$M = \begin{bmatrix} -1 & 0 & 4 \\ 2 & 3 & 1 \end{bmatrix} \quad N = \begin{bmatrix} 0 & 1 & 2 \\ -3 & 4 & 5 \end{bmatrix} \quad L = \begin{bmatrix} 1 & 2 \\ -8 & 0 \\ 4 & 6 \end{bmatrix}$$

L can't be added or subtracted from the other matrices, but the following operations are possible:

$$M + N = \begin{bmatrix} -1 & 1 & 6 \\ -1 & 7 & 6 \end{bmatrix}$$

$$M - N = \begin{bmatrix} -1 & -1 & 2 \\ 5 & -1 & -4 \end{bmatrix}$$

Theorem: Multiplying Matrices.

In order for multiplication to be possible, the number of columns in the first matrix must be equal to the number of rows in the second matrix. The resulting matrix has the number of rows of the first matrix and the number of columns of the second matrix. In the previous example, $M \times L$ is possible, and it results in a 2×2 matrix. The resulting value of $M \times L$ is:

$$ML = \begin{bmatrix} -1 \cdot 1 + 0 \cdot -8 + 4 \cdot 4 & -1 \cdot 2 + 0 \cdot 0 + 4 \cdot 6 \\ 2 \cdot 1 + 3 \cdot -8 + 1 \cdot 4 & 2 \cdot 2 + 3 \cdot 0 + 1 \cdot 6 \end{bmatrix} = \begin{bmatrix} 15 & 22 \\ -18 & 10 \end{bmatrix}$$

Essentially, the summation of pairwise multiplication for row i of the first matrix and column j of the second matrix results in the entry a_{ij} of the resulting matrix.

Definition: Square Matrix

A **square matrix** has the same number of rows as columns.

1.2 Row Reduction and Echelon Forms

Definition: Leading Entry

The *leading entry* is the first non-zero entry in a row going from left to right.

Definition: Row Echelon Form (REF) and Reduced Row Echelon Form (RREF)

In order for a matrix to be in **REF**:

1. All rows consisting of 0s must be at the bottom
2. A leading entry must be strictly to the right of the leading entry of the row above.

Note, to visually see if a matrix is in REF, check the "stairs," which means everything underneath the leading entries (the "stairs") should be 0.

For **RREF**, there are two additional requirements on top of the REF requirements:

1. The leading entries must each be 1.
2. Each column containing a leading entry has zeros in all other entries.

For example, L is in RREF:

$$L = \begin{bmatrix} 1 & 0 & 3 & 0 & 7 \\ 0 & 1 & -1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

Theorem: Uniqueness of Reduced Echelon Form

Every matrix is row equivalent to one and only one reduced echelon matrix.

Because of the above theorem, it is often useful to think about a matrix's reduced echelon matrix. In particular, to say that a matrix A reduces to an RREF B, the notation for that is

$$A \sim B$$

How do we transform matrices into their REF or RREF? There are three *elementary row operations*:

1. *Interchange*: switching any two rows. Written as $r_i \leftrightarrow r_j$.
2. *Scaling*: taking any row and multiplying each entry by a constant $c \neq 0$.
3. *Replacement*: multiplying a row by a nonzero number then adding it to another row.

Example: Solving System of Equations

Suppose we wanted to solve the following system of equations:

$$x - 3z = 8$$

$$2x + 2y + 9z = 7$$

$$y + 5z = -2$$

We can write this set of equations in the following augmented matrix:

$$A = \begin{bmatrix} 1 & 0 & -3 & 8 \\ 2 & 2 & 9 & 7 \\ 0 & 1 & 5 & -2 \end{bmatrix}$$

We can make the first term of the second row become 0 by replacing r_2 with $-2r_1 + r_2$ (the operation can be written as $-2r_1 + r_2 \to r_2$):

$$A = \begin{bmatrix} 1 & 0 & -3 & 8 \\ 0 & 2 & 15 & -9 \\ 0 & 1 & 5 & -2 \end{bmatrix}$$

We can do another replacement operation on the third row, using $r_2 - 2r_3 \to r_3$:

$$A = \begin{bmatrix} 1 & 0 & -3 & 8 \\ 0 & 2 & 15 & -9 \\ 0 & 0 & 5 & -5 \end{bmatrix}$$

Now A is in REF. To solve the system of equations, we use *back-substitution*, starting from the last row and going up. The last row simply tells us $5z = -5$ or $\boxed{z = -1}$. The second row tells us $2y + 15z = -9$. Substituting $z = -1$ yields $\boxed{y = 3}$. Finally, the first row is equivalent to $x - 3z = 8$, which tells us $\boxed{x = 5}$. The solution can alternatively be written as $(5, 3, -1)$.

Notes:

- If the last row is $[0\ 0\ 0 \cdots b]$, if $b \neq 0$ (the last column is a pivot column), then there is no solution. This is the only scenario in which a system of linear equations is *inconsistent*. All other systems are *consistent*.
- If the last row is the same but $b = 0$, there are infinitely many solutions. In this class, we start assigning *free variables* from the bottom to the top (i.e. for 4 variables and 2 unique equations, in which any two of the variables can be free variables, we call x_4 and x_3 the free variables and express x_1 and x_2 in terms of the free variables).

These special cases yield the following theorem:

Theorem: Existence and Uniqueness Theorem

If a linear system is consistent (last row is not $[0\ 0\ 0 \cdots b]$ where $b \neq 0$), the solution set either contains

1. One unique solution
2. Infinitely many solutions (at least one free variable)

On the topic of free variables, the following is important information:

Definition: Pivot Position

A **pivot position** is simply a leading entry in a row. A **pivot column** is a column that contains a pivot position.

Definition: Types of Variables

A **basic variable** is any variable that corresponds with a pivot column, and a **free variable** is any variable that is not a basic variable.

Although the above is the definition of a basic variable/free variable used, in practice one might think of a free variable as a variable that can take any value, and a basic variable is a variable that is pre-determined once all the free variables have been decided. However, implicit in the definition above is the convention that we assign free variables from bottom to top. To understand what this means, see the following example:

Example: Free and Basic Variables

Consider the following augmented matrix in RREF:

$$\begin{bmatrix} 1 & 0 & -5 & 0 & -2 & 6 \\ 0 & 1 & 2 & 0 & 11 & 7 \\ 0 & 0 & 0 & 1 & -16 & -9 \end{bmatrix}$$

From our definition, the basic variables are x_1, x_2, x_4 and the free variables are x_3, x_5. That is because if we work from the bottom row and work upwards, we can always solve for the variable in the pivot column in terms of free variables and the basic variables below. In other words, the general solution can be written as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 6 \\ 7 \\ 0 \\ -9 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ -11 \\ 0 \\ 16 \\ 1 \end{bmatrix}$$

So, it is always possible for the general solution can be written in a form such that the free variables (under our definition) can be varied freely. However, that is NOT to say the general solution cannot be written in a different form where other variables are to be varied freely. For example, the general solution can also be written as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 57/8 \\ 13/16 \\ 0 \\ 0 \\ 9/16 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1/8 \\ -11/16 \\ 0 \\ 1 \\ 1/16 \end{bmatrix}$$

Thus, the definition of free and basic variables presented above serves as a way to implicitly incorporate the convention to assign free variables to the highest numbers first.

1.3 Vector Equations

Know the difference between lines, rays, and segments. Add vectors tip to tail. Vector properties: for all $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^n$ and all scalars c, d:

- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$
- $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (where $-\mathbf{u} = (-1)\mathbf{u}$
- $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
- $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
- $c(d\mathbf{u}) = cd\mathbf{u}$
- $1\mathbf{u} = \mathbf{u}$

Definition: Linear Combinations.

Let $\mathbf{v}_1, \mathbf{v}_2 \dots, \mathbf{v}_p \in \mathbb{R}^n$ and $c_1, c_2, \dots, c_p \in \mathbb{R}$. The vector $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p$ is called a *linear combination* of $\mathbf{v}_1, \mathbf{v}_2 \dots, \mathbf{v}_p$ with weights $c_1, c_2, \dots c_p$.

Linear combinations are just another way of writing a system of linear equations:

$$c_1 \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + c_3 \begin{bmatrix} 5 \\ -6 \\ 8 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 5 & 2 \\ -2 & 1 & -6 & -1 \\ 0 & 2 & 8 & 6 \end{bmatrix}$$

Some matrix operations show that this system of equations has infinitely many solutions. Assuming c_3 to be the free variable, $c_1 = 2 - 5c_3$ and $c_2 = 3 - 4c_3$. In vector form, the solutions can be written as

$$\mathbf{c} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} -5 \\ -4 \\ 1 \end{bmatrix}$$

Definition: Span.

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^n$, the set of all linear combinations of these vectors is denoted by $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ and is the subset of $\mathbb{R}^n$ spanned (or generated) by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. In other words, the span is the collection of all vectors that can be written in the form of $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$.

1.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$

The matrix equation $A\mathbf{x} = \mathbf{b}$ is a common equation where A is an $m \times n$ matrix and $\mathbf{x}$ and $\mathbf{b}$ are column vectors of size $n \times 1$ and $m \times 1$ respectively. The bold face represents a vector, with size $a \times 1$ where a is any integer. The matrix A can be notated in the following manner:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

So the equation $A\mathbf{x} = \mathbf{b}$ can be written as

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

Therefore, the matrix equation has an equivalent vector equation, which in turn has an equivalent augmented matrix, revealing a similar way to solve it as in Section 1.3. This matrix equation is subject to the following important theorem:

Theorem: Spanning $\mathbb{R}^m$

The columns of A span $\mathbb{R}^m$ $\iff$ the matrix A in the equation $A\mathbf{x} = \mathbf{b}$ has a pivot position in every row.

Proof: Spanning $\mathbb{R}^m$

The columns of A span $\mathbb{R}^m$ if for any vector $\mathbf{b} \in \mathbb{R}^m$ there is a solution for $\mathbf{x}$. This is true if the augmented matrix $[A\ \mathbf{b}]$ is consistent, meaning its last row is not $[0\ 0\ \cdots\ b]$ (where $b \neq 0$) for any value of $\mathbf{b}$. Since $\mathbf{b}$ can be anything, there is no shot of obtaining a last row of $[0\ 0\ \cdots\ 0]$ for all $\mathbf{b}$ upon row reduction, so it must be that the RREF form of A has a pivot in the last row, which implies it has a pivot in every row.

What does this mean, intuitively? By row reducing $A\mathbf{x} = \mathbf{b}$, we preserve the underlying solutions but convert the general A to a more tractable set of equations to use. A pivot in a row of this new set of equations essentially tells us that one can form the *standard basis vector* of that row. Consider the following as an example of standard basis vectors in $\mathbb{R}^3$:

$$\begin{bmatrix}1\\0\\0\end{bmatrix} \quad \begin{bmatrix}0\\1\\0\end{bmatrix} \quad \begin{bmatrix}0\\0\\1\end{bmatrix}$$

What this theorem is saying is that if the equivalent set of equations $\mathrm{rref}(A)$ can form these above vectors, then they can form any vector in the vector space via a linear combination of the solutions.

Definition: Identity Matrix (for square matrices).

$$I_1 = [1]$$

$$I_2 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

$$I_n = \begin{bmatrix}1 & 0 & 0 & \cdots & 0\\0 & 1 & 0 & \cdots & 0\\0 & 0 & 1 & \cdots & 0\\\vdots & \vdots & \vdots & \ddots & \vdots\\0 & 0 & 0 & \cdots & 1\end{bmatrix}$$

Identity matrices only exist for square matrices. Note that for any vector $\mathbf{x} \in \mathbb{R}^n$:

$$I_n\mathbf{x} = \mathbf{x}$$

Notes on matrix operations. If A is an $m \times n$ scalar, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, c is a scalar, then:

- $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
- $A(c\mathbf{u}) = c(A\mathbf{u})$
- $A\mathbf{v} \neq \mathbf{v}A$

1.5 Solution Sets of Linear Systems

Definition: Homogenous Linear System

A **homogenous linear system** is a system of the form

$$A\mathbf{x} = \mathbf{0}$$

A **nonhomogenous linear system** is simply $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} \neq \mathbf{0}$.

$\mathbf{x} = \mathbf{0}$ is always a solution to homogenous systems, called the **trivial solution**. Depending on the system, there may also exist a **nontrivial solution** that is $\mathbf{x} \neq \mathbf{0}$.

Theorem: Nontrivial Solutions

The homogenous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.

Now, we consider **parametric vector equations**, which emphasize that the parameters vary over all real numbers.

Definition: Parametric Vector Equation

The parametric vector equation for a line passing through the origin, for example, is

$$\mathbf{x} = t\mathbf{v}$$

The parametric vector equation for a plane passing through the origin is

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v}$$

Thinking in terms of parametric vector equations helps us reach the following important theorem:

Theorem: Solutions to Nonhomogenous Linear System

The solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form

$$\mathbf{w} = \mathbf{p} + \mathbf{v}_h$$

where $\mathbf{v}_h$ is any solution of the homogenous equation $A\mathbf{x} = 0$ and $\mathbf{p}$ is a constant, particular solution. Note this theorem only applies if $A\mathbf{x} = \mathbf{b}$ has at least one solution.

Note that the parametric vector equation helps us think of the intuition. Suppose the solution space for the homogenous system is a plane passing through the origin. The solution space for the nonhomogenous system is simply a translation of that plane by some constant vector $\mathbf{p}$.

1.6 Applications of Linear Systems

Applications of linear systems: balancing chemical equations, network flow (straightforward linear systems of equations).

1.7 Linear Independence

Definition: Linearly Independent

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\} \in \mathbb{R}^n$ is called linearly independent if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution $x_1 = x_2 = \cdots = x_p = 0$. If the set is not linearly independent, it is called *linearly dependent*.

Some cases to think about:

- *only one vector.* If $\mathbf{v} \neq 0$, it is independent; otherwise, it is dependent.
- *two vectors.* For linear independence, both $\mathbf{v}_1, \mathbf{v}_2 \neq \mathbf{0}$, and $\mathbf{v}_1$ must not be a multiple of $\mathbf{v}_2$.
- *more than two vectors.* If the set of vectors contains the zero vector, then the set S is linearly *dependent*.
- If a set has more vectors than there are entries in each vector, then the set is linearly dependent.

  Proof: Too Many Vectors

  Let $A = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_p]$ $(n \times p)$, where $p > n$. Since this creates a set of p unknowns and n equations, where the number of unknowns is greater than the number of equations, there must be at least $p - n$ free variables in order for there to be a solution. Therefore, there are infinite non-trivial solutions.

  The intuition is that maintaining linear independence is like having all vectors orthogonal to each other (more accurately, having at least one component of a vector orthogonal to the span of the other vectors). One can only have as many vectors orthogonal to each other as there are dimensions.

Theorem: Characterization of Linearly Dependent Sets

A set of two or more vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_p\}$ is linearly dependent if and only if *at least one* of the vectors is a linear combination of the others. (If at least one of the vectors is the zero vector, then the set is linearly dependent.)

Example: Is this system linearly independent?

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -5 \\ -3 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 10 \\ 6 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$$

If it is not, it will have nontrivial solutions to the following matrix:

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ -5 & 10 & 0 & 0 \\ -3 & 6 & 4 & 0 \end{bmatrix}$$

The non-trivial solution is:

$$\mathbf{x} = \begin{bmatrix} 2x_2 \\ x_2 \\ 0 \end{bmatrix}$$

Therefore this system is *not* linearly independent.

1.8 Introduction to Linear Transformations

Definition: Transformation

A transformation (or function or mapping) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns each vector $\mathbf{x} \in \mathbb{R}^n$ to a vector $\mathbf{T}(\mathbf{x}) \in \mathbb{R}^m$.

A transformation can be written as $\mathbf{x} \mapsto T(\mathbf{x})$.

Definition: Domain, Codomain, Range

For a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the set $\mathbb{R}^n$ is called the **domain**, the set $\mathbb{R}^m$ is the **codomain**, and the set of images $T(\mathbf{x})$ is called the **range**.

Intuitively, the domain is the set of possible inputs, the range is the set of possible outputs, and the codomain is the smallest vector space of the form $\mathbb{R}^n$ that contains the range.

We will almost exclusively focus on linear transformations for the rest of Linear Algebra:

Definition: Linear Transformations

A transformation T is **linear** if

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in the domain of T.
2. $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and all $\mathbf{u}$ in the domain of T.

Possible linear transformations: contraction, dilation, shear, rotation.

1.9 The Matrix of a Linear Transformation

Theorem: Standard Matrix for Linear Transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that $T(\mathbf{x}) = A\mathbf{x}, \forall \mathbf{x} \in \mathbb{R}^n$ ($\forall$ means for all).

Proof: Existence of Standard Matrix

$\mathbf{x} = I_n\mathbf{x}$, where $I_n = \begin{bmatrix} \mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_n} \end{bmatrix}$ ($\mathbf{e}_i$ is the unit (column) vector in the ith dimension). Therefore $\mathbf{x} = x_1\mathbf{e_1} + x_2\mathbf{e_2} + \cdots + x_n\mathbf{e_n}$. Then a linear transformation $T(\mathbf{x})$ can be transformed into the following:

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e_1} + x_2\mathbf{e_2} + \cdots + x_n\mathbf{e_n}) \\ &= x_1T(\mathbf{e_1}) + x_2T(\mathbf{e_2}) + \cdots x_nT(\mathbf{e_n}) \end{aligned}$$

where the second line is possible since $T(\mathbf{x})$ is *linear*. Therefore, the operation can be written as:

$$T(\mathbf{x}) = \begin{bmatrix} T(\mathbf{e_1}) & T(\mathbf{e_2}) & \cdots & T(\mathbf{e_n}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x}$$

The above proof is really important in conceptualizing what a matrix (in a linear transformation) represents: each column represents where the standard basis vectors, $\mathbf{e}_i$, end up after the transformation, and the transformation of any vector is completely determined upon knowing where the basis vectors end up.

Now, what is a linear transformation geometrically? (Taken from 3B1B Lin Alg Series Ep 3) A linear transformation must satisfy the following requirements:

- All lines in space stay lines after the transformation. (In general, the first geometric requirement implies that a set of evenly spaced grid lines should remain parallel and evenly spaced, because a transformation that changes the grid lines to be unevenly spaced would transform a diagonal line into something non-linear.)
- The origin stays in place after the transformation.

Note that rotations are one type of linear transformation. They are linear because you can see geometrically that the two conditions of linearity apply. They have a transformation matrix that is

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

where θ is the counterclockwise angle a vector is increased by upon transformation.

Example: Linear Transformation

Let $T : \mathbb{R}^2 \to \mathbb{R}^4$. If

- $T(\mathbf{e_1}) = (3, 1, 3, 1)$
- $T(\mathbf{e_2}) = (-5, 2, 0, 0)$
- $\mathbf{e_1} = (1, 0)$
- $\mathbf{e_2} = (0, 1)$

Then what is the standard matrix A of T? It is simply

$$A = \begin{bmatrix} 3 & -5 \\ 1 & 2 \\ 3 & 0 \\ 1 & 0 \end{bmatrix}$$

Since

$$A\mathbf{x} = x_1 \begin{bmatrix} 3 \\ 1 \\ 3 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -5 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

Definition: Onto, One-to-one.

- A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of at least one $\mathbf{x}$ in $\mathbb{R}^n$
- A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **one-to-one** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of at most one $\mathbf{x}$ in $\mathbb{R}^n$

In other words, onto is a matter of the existence of a solution for all of the codomain, whereas one-to-one is a matter of the uniqueness of a solution for a given vector in the range.

Theorem: Onto, One-to-one.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let A be the standard matrix for T Then

- T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ $\iff$ the columns of A span $\mathbb{R}^m$. As stated earlier, this implies that there must be as many pivot positions as there are rows in A.
- T is one-to-one $\iff$ the columns of A are linearly independent (OR the equation $T(\mathbf{x}) = A\mathbf{x} = 0$ has only trivial solutions.) This implies there must be a pivot in every column, since there may be no free variables. A corollary is that any matrix A with more columns than rows $m < n$ cannot satisfy this condition.

1.10 More Applications

Some applications, like electrical network. In electrical networks, the class's convention is to consider each loop's independent current, and by convention have + mean counterclockwise.

2 Matrix Algebra

2.1 Matrix Operations

Theorem: Addition and Subtraction

A, B, and C are matrices of the same size, and r and s are scalars. The following identities are true:

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $A + 0 = A$
4. $r(A + B) = rA + rB$
5. $(r + s)A = rA + sA$
6. $r(sA) = r(sA)$

Theorem: Matrix Multiplication

Let A be an $m \times n$ matrix, and let B and C be matrices with sizes such that the indicated sums and products are defined. Let r be a scalar. Then:

1. $A(BC) = (AB)C$ (associative property)
2. $A(B + C) = AB + AC$ (left distributive law)
3. $(B + C)A = BA + CA$ (right distributive law)
4. $r(AB) = (rA)B = A(rB)$
5. $I_m A = A = AI_n$ (identity for matrix multiplication)

Some good intuition for thinking about how the above properties might be true: matrix multiplication can be thought of as a composition of linear transformations. For example:

Example: Two Transformations

Imagine applying a 90° CCW rotation, then a horizontal shear. The two matrices that describe these motions are R and S, respectively:

$$R = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Definition: Commute

If $AB = BA$, we say that A and B **commute** with each other.

Note that matrices do not commute in general. In fact, the following warnings are in order:

- $AB \neq BA$ in general.
- If $AB = AC$, this does not imply $B = C$.
- If AB is the zero matrix, this does not imply $A = 0$ or $B = 0$.

Definition: Powers of a Matrix

If A is an $n \times n$ matrix and k is a positive integer, then A^k denotes the product of k copies of A:

$$A^k = \underbrace{A \cdots A}_{k \text{ times}}$$

Definition: Transpose of a Matrix

Given an $m \times n$ matrix, the **transpose** of A is the $n \times m$ matrix, denoted by A^T. For example:

$$A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \qquad A^T = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$$

Theorem: Transpose Operations

Let A and B indicate matrices whose size are appropriate for the following operations, and let r be a scalar:

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(rA)^T = rA^T$
4. $(AB)^T = B^T A^T$. In general, $(ABCD \cdots)^T = \cdots D^T C^T B^T A^T$

2.2 Inverse of a Matrix

The notation for the inverse of the matrix A is

$$A^{-1}$$

Note that $A^{-1} \neq 1/A$. What is the inverse of a matrix?

Definition: Inverse of a (Square) Matrix

The inverse of a matrix A satisfies the following equality:

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

where I is an identity matrix. Note that A *must be a square*; a non-square matrix does not have an inverse. A square matrix does not necessarily have an inverse, but it can have one.

Theorem: Calculating Inverse of a 2×2 Matrix

For a 2×2 matrix A:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then its inverse is

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

where $\det A = ad - bc$. Visually, note that other than the determinant in the denominator, the inverse swaps a and d while slapping a negative sign on b and c. If $\det A = 0$, *the inverse matrix does not exist.*

Note that for the equation $A\mathbf{x} = \mathbf{b}$. If we left-multiply both sides by A^{-1} (note that left multiplying and right multiplying are different), we have that

$$A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}$$

$$\mathbf{x} = A^{-1}\mathbf{b}$$

This presents a new method of solving a system of linear equations.

Definition: Invertible, Singular

A matrix that has an inverse is **invertible** or **nonsingular**, whereas a matrix without an inverse is **singular**.

The following presents a general way to find A^{-1} when the size of A is $n \times n$ with $n > 2$:

Theorem: General Method of Finding Inverse

We can make the $n \times 2n$ matrix in the following way:

$$[A \mid I_n]$$

If we row reduce the above matrix into RREF, then it will have the following form:

$$[I_n \mid A^{-1}]$$

Note that it is possible that upon row reduction to RREF, I_n is not on the left side (e.g. a row of all zeros). If the left hand side of the matrix is not I_n, then A^{-1} does not exist.

Proof: General Inverse

If A is an $m \times n$ matrix, an elementary row operation can be written as EA, where E is the $m \times m$ matrix that results from performing the operation on I_m. For example, the matrix

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

is the result of $4r_1 + r_3 \to r_3$ on I_3, and one can quickly check that E performs the same operation on any $3 \times n$ matrix A. An important property of these **elementary matrices** is that they *must be invertible*, as row operations clearly invertible.

Now, suppose A is an $n \times n$ matrix that is invertible. Then, since $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x} = A^{-1}\mathbf{b}$ for every $\mathbf{b}$, it must mean that A has a pivot position in every row. Since A is square, the pivot positions must all be along the diagonal, which implies the reduced row echelon form of A is I_n, or $A \sim I_n$.

Given this, there must be a set of elementary row operations such that

$$A \sim E_1 A \sim E_2(E_1 A) \sim \cdots \sim E_p(E_{p-1} \cdots E_1 A) = I_n$$

Calling upon the property of elementary matrices, we know that $E_p \cdots E_1$ is invertible, so

$$A = (E_p \cdots E_1)^{-1} I_n = (E_p \cdots E_1)^{-1}$$

Therefore,

$$A^{-1} = ((E_p \cdots E_1)^{-1})^{-1} = E_p \cdots E_1$$

The final line means that A^{-1} can be constructed by performing the same elementary row operations E_1 through E_p that reduces A to I_n. ■

2.3 Characterizations of Invertible Matrices

Theorem: Invertible Matrix theorem

Let A be a square $n \times n$ matrix. Then the following statements are equivalent:

- A is an invertible matrix.
- A is row equivalent to I_n
- A has n pivot positions
- The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. (since if A is invertible, $\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$)
- The columns form a linearly independent set.
- The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one to one.
- The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$
- The columns of A span $\mathbb{R}^n$
- The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$

- There is an $n \times n$ matrix C such that $CA = I$
- There is an $n \times n$ matrix D such that $AD = I$
- A^T is an invertible matrix.

Theorem: Square Inverses

If A and B are square matrices and $AB = I$, then A and B are invertible with

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

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

Theorem: Invertible Linear Transformations: T^{-1}

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be invertible if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

- $S(T(\mathbf{x})) = \mathbf{x}, \forall \mathbf{x} \in \mathbb{R}^n$
- $T(S(\mathbf{x})) = \mathbf{x}, \forall \mathbf{x} \in \mathbb{R}^n$

Note: T is invertible $\iff$ A is invertible, where A is the standard matrix of T.

Theorem: Invertible Inverses

- If A is invertible, then A^{-1} is invertible and
$$(A^{-1})^{-1} = A$$
- If A and B are invertible, then so is AB and
$$(AB)^{-1} = B^{-1}A^{-1}$$
- If A is invertible, so is A^T, and
$$(A^T)^{-1} = (A^{-1})^T$$

3 Determinants

3.1 Introduction to Determinants

Theorem: Calculating Determinant

Suppose we have the matrix M

$$M = \begin{bmatrix} 3 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -1 \end{bmatrix}$$

Choosing any row or any columns you can do the $+ - +$ method and split any 3×3 matrix into 2×2 matrices. For example, suppose we chose the top row. Then, we could compute the determinant as the following:

$$\det M = 3\begin{vmatrix} 3 & 2 \\ 5 & -1 \end{vmatrix} - 0\begin{vmatrix} 2 & 2 \\ 0 & -1 \end{vmatrix} + 4\begin{vmatrix} 2 & 3 \\ 0 & 5 \end{vmatrix} = 3(3 \cdot (-1) - 2 \cdot 5) + 4(2 \cdot 5) = 3(-13) + 40 = 1$$

The $+ - +$ method requires that the coefficients have $+$ and $-$ in the following manner:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

Alternatively, note that the sign is simply $(-1)^{r+c}$, where r and c are the row and column numbers. Note that this method, called **cofactor expansion,** can be used for general matrices by applying it successively: 4×4 matrices can be split into three 3×3 matrices, which can each be split into two 2×2 matrices, and so on. It helps to pick a row or column that has many zeros.

Definition: Cofactor

The **(i,j) cofactor** of the matrix A is the matrix that is A but excludes row i and column j. It is also written as C_{ij}.

For a special case: for a triangular matrix A, $\det A$ is the product of the entries in the main diagonal of A:

- An example **upper triangular** matrix is

$$\begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & -1 \\ 0 & 0 & 7 \end{bmatrix}$$

  The determinant is $2 \cdot 5 \cdot 7 = 70$

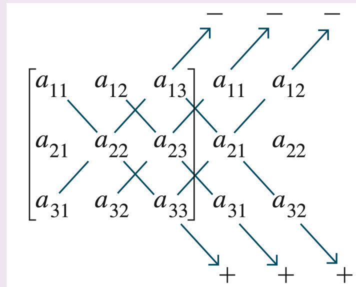
- An example **lower triangular** matrix is

$$\begin{bmatrix} 3 & 0 & 0 \\ 1 & -4 & 0 \\ 2 & -5 & 6 \end{bmatrix}$$

  The determinant is $3 \cdot -4 \cdot 6 = -72$.

Theorem: Shoelace Method of Finding Determinant

For a 3×3 matrix only, the determinant can be found in the following way:



3.2 Properties of Determinants

Theorem: Elementary Row Operations

Let A be a square matrix. The determinant does the following under elementary row operations:

- *scaling:* if one row of A is multiplied by a value $k \neq 0$ to produce the matrix B, then

$$\det B = k \det A$$

  Note that the formula still works with $k = 0$, but B would no longer be equivalent to A.

- *interchange:* if one row of A is switched with another row to produce B, then

$$\det B = -\det A$$

- *replacement:* if a multiple of one row of A is added to another row to produce a matrix B, then

$$\det B = \det A$$

Proof: Elementary Row Operations

The proofs, in my opinion, are unenlightening, because they are either really obvious (in the case of scaling) or involve $\det AB = \det A \cdot \det B$ with one of the letters being the identity matrix.

However, below you will see that $\det A = \det A^T$, which can be used to see some intuition. Because $\det A = \det A^T$, clearly row operations will have the same effect as column operations, which combined with the below theorem about how determinant is related to volume, can quickly result in intuition about each of the row operations:

- For scaling, that is equivalent to simply scaling one column vector up or down; clearly, the volume of the unit parallelopiped is proportional to one column vector.

- Switching any two column vectors is like switching the orientation of the linear transformation, which results in the negative sign. For example, in the two dimensional case, flipping the $\hat{i}$ and $\hat{j}$ vectors is like an inversion.
- In the two-dimensional case, one can see that the volume of the unit parallelopiped is unchanged when adding the multiple of one vector to another vector; it is like performing a shear on the original transformation. This intuition extends to the general multidimensional case.

In practice, using the above properites to find the determinant of a general matrix A will be much faster than cofactor expansion.

Theorem: Other Properties

1. $\det A^T = \det A$
2. $\det AB = \det A \cdot \det B$

Next, we will discuss the important matrix $A_i(\mathbf{x})$:

Theorem: Linearity of Determinant

Let us define $A_i(\mathbf{x})$ as the matrix with the ith column of A substituted with $\mathbf{x}$. Then, the function $\det A_i(\mathbf{x})$ is a linear function in $\mathbf{x}$.

Proof: Linearity

Let $T(\mathbf{x})$ be the transformation from $\mathbb{R}^n \to \mathbb{R}$ that is defined by $\det A_i(\mathbf{x})$. The two conditions for linearity are met, and therefore T is a linear transformation:

1. $T(c\mathbf{x}) = cT(\mathbf{x})$, since this is simply the scaling of a column vector.
2. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, which can be seen through the cofactor expansion of the determinant along the ith column.

3.3 Cramer's Rule, Volume, and Linear Transformations

Theorem: Cramer's Rule

Suppose you want to solve the equation $A\mathbf{x} = \mathbf{b}$, where A is an invertible matrix. Let $A_i(\mathbf{b})$ be the matrix A such that column i is replaced by $\mathbf{b}$. Then the solution $\mathbf{x}$ is given by the entries

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

Proof: Cramer's Rule

Let us define $I_i(\mathbf{x})$ to be the $n \times n$ identity matrix with its ith column replaced by $\mathbf{x}$. If $A\mathbf{x} = \mathbf{b}$, then:

$$\begin{aligned} A(I_i(\mathbf{x})) &= A[\mathbf{e}_1 \quad \cdots \quad \mathbf{x} \quad \cdots \quad \mathbf{e}_n] \\ &= [A\mathbf{e}_1 \quad \cdots \quad A\mathbf{x} \quad \cdots \quad A\mathbf{e}_n] \\ &= [\mathbf{a}_1 \quad \cdots \quad \mathbf{b} \quad \cdots \quad \mathbf{a}_n] = A_i(\mathbf{b}) \end{aligned}$$

By the multiplicative property of determinants

$$\det A \det I_i(\mathbf{x}) = \det A_i(\mathbf{b})$$

But $\det I_i(\mathbf{x})$ is simply x_i, which can be seen through the cofactor expansion along the ith row (not column). The desired result is thus proven.

Note that Cramer's Rule easily leads to a general formula for A^{-1}. Since $AA^{-1} = I$, the jth column of A^{-1} must be the vector $\mathbf{x}$ that satisfies

$$A\mathbf{x} = \mathbf{e}_j$$

By Cramer's Rule, the ith entry of $\mathbf{x}$, in other words the (i, j)th entry of A^{-1}, is

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

The cofactor expansion along the ith column of $A_i(\mathbf{e}_j)$ shows that the determinant in the numerator is

$$\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji}$$

where C_{ji} is a cofactor of A. Therefore:

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

where the right matrix is called the **adjugate** or **classical adjoint**. Note that the subscripts on C_{ji} are the reverse of (i, j); in other words, the adjugate is the transpose of the matrix of cofactors.

Theorem: Volume

$\det A$ gives the n-dimensional volume of the parallelopiped determined by the columns of A.

For a linear transformation T determined by the matrix A, the volume of any finite region S is scaled by a factor of $\det A$.

4 Vector Spaces

4.1 Vector Spaces and Subspaces

Definition: Vector Space

A vector space is a nonempty set V of vectors which are defined on two operations: addition and multiplication by scalars, subject to the following 10 axioms:

1. $\mathbf{u}+\mathbf{v}$ is in V
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
3. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$
4. $\mathbf{u}+\mathbf{0}=\mathbf{u}$
5. $\mathbf{u}+(-\mathbf{u})=0$
6. $c\mathbf{u}$ is in V
7. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$
8. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$
9. $c(d\mathbf{u})=(cd)\mathbf{u}$
10. $1\mathbf{u}=\mathbf{u}$

Definition: Subspace

A subspace of a vector space V is a subset H of V that has three properties:

1. The zero vector of V is in H
2. H is closed under vector addition (i.e. $\forall \mathbf{u},\mathbf{v}\in H$, $\mathbf{u}+\mathbf{v}\in H$)
3. H is closed under multiplication by scalars (i.e. $\forall \mathbf{u}\in H$, $c\mathbf{u}\in H$)

What do these mean, intuitively? Vector spaces and subspaces simply formalize intuition about lines, planes, and higher dimensional equivalents. A vector space (in flimsy, non-rigorous terms) is basically an infinite, n-dimensional space such as a line, a plane, and higher dimensional equivalents (it could also be simply the $\mathbf{0}$ vector). Example vector spaces include $\mathbf{0}$, $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, etc. Technically there are others but intuitively they are similar to these. A subspace is simply a “smaller” or equal-sized vector space, but still containing an infinite number of vectors (unless, again, it is just $\mathbf{0}$). For example, for the vector space $\mathbb{R}^3$, example subspaces would be itself, infinite planes that pass through the origin, infinite lines that pass through the origin, and $\mathbf{0}$.

Theorem: Span

If $\mathbf{v}_1,\cdots\mathbf{v}_p$ are in a vector space V, then $\text{Span}\{\mathbf{v}_1,\cdots\mathbf{v}_p\}$ is a subspace of V.

4.2 Null/Column/Row Spaces, Linear Transformations

Definition: Null Space

The **null space** or **kernel** of a $m \times n$ matrix A, written as Nul A, is the set of all solutions $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\text{Nul } A = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}$$

Note that Nul A is a subspace of $\mathbb{R}^n$.

Theorem: Finding Null Space

In practice to find the null space of A, reduce $[A\ \mathbf{0}]$ to its RREF, find the basic variables in terms of free variables, then decompose the vector into its general solution.

Definition: Column Space

The **column space** of an $m \times n$ matrix A, written as Col A, is

$$\text{Col } A = \text{Span}\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$$

where $\mathbf{a}_i$ is the ith column of A. In set notation,

$$\text{Col } A = \{\mathbf{b} : \mathbf{b} = A\mathbf{x} \text{ for some x} \in \mathbb{R}^n\}$$

A row space is similarly defined, but for rows. In essence, Row A = Col A^T

Theorem: Finding Column Space

To find the column space, reduce A to its RREF, then use the pivot columns of the original A as a basis of the column space. For the row space, you can directly use the pivot rows of the RREF of A, since row operations do not alter the row space.

4.3 Linearly Independent Sets, Bases

Definition: Bases

Let H be a subspace of a vector space V. A set of vectors $\mathcal{B} = \{b_1, b_2, \cdots, b_p\} \in V$ is a **basis** for H if

1. $\mathcal{B}$ is a linearly independent set
2. The subspace spanned by $\mathcal{B}$ coincides with H. (i.e. $H = \text{Span}\{b_1, \cdots, b_p\}$

Theorem: Spanning Set Theorem

Let $S = \{v_1, v_2, \cdots, v_p\}$ be a set in V. Let $H = \text{Span}\{v_1, v_2, \cdots, v_p\}$.

1. If one of the vectors in S, say v_k, is a linear combination of the remaining vectors in S, then the set formed by S by removing v_k still spans H.
2. If $H \neq \{\mathbf{0}\}$, some subset of S is a basis for H.

4.4 Coordinate Systems

Definition: Coordinate System

Suppose $\mathcal{B} = \{b_1, b_2 \cdots b_n\}$ is a basis for a vector space V and $\mathbf{x} \in V$. The coordinates of $\mathbf{x}$ relative to the basis $\mathcal{B}$ (or the $\mathcal{B}$-coordinate of $\mathbf{x}$), notated

$$[x]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

are the weights $c_1, c_2, \cdots, c_n$ such that

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$$

The above can also be written as

$$x = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]\,[x]_{\mathcal{B}} = P_B[x]_{\mathcal{B}}$$

where P_B is the change of coordinates matrix.

Note that a coordinate mapping from $\mathbf{x} \mapsto [x]_{\mathcal{B}}$ is a one to one linear transformation from V onto $\mathbb{R}^n$. Such a transformation is also known as an **isomorphism** from V onto $\mathbb{R}^n$, because V and $\mathbb{R}^n$ are essentially indistinguishable.

4.5 Dimension

Definition: Dimension

If the vector space V is spanned by a finite set, V is **finite dimensional**. The **dimension** of V is the number of vectors in a basis of V.

Definition: Rank

The **rank** of a matrix A is the dimension of the column space (and equivalently, the dimension of the row space).

The pivot columns of a matrix A form the basis of Col A; the number of pivot columns therefore specifies the dimension of Col A. Since a basis for Row A can be found by taking the pivot rows of the reduced echelon form of A, the dimension of Row A and Col A are the same.

Definition: Nullity

The **nullity** of a matrix A is just the dimension of its null space.

If the equation $A\mathbf{x} = \mathbf{0}$ has k free variables, that means the basis of Nul A has exactly k vectors; in other words, the nullity of A is k. This yields the following important theorem:

Theorem: Rank-Nullity Theorem

For any $m \times n$ matrix A:

$$\text{rank } A + \text{nullity } A = \text{num cols } A$$

The above theorem is true because the k free variables that make up the basis of the null space are exactly the things that are not the pivot columns of A; in total, pivot + non-pivot columns add up to the total number of columns.

Theorem: Invertible Matrix Theorem (cont.)

The following statements are equivalent to having an invertible $n \times n$ matrix A:

1. The columns of A form a basis of $\mathbb{R}^n$
2. Col $A = \mathbb{R}^n$ (intuitively, a linear transformation is invertible if it doesn't reduce the dimension, because an inverse linear transformation can never map a smaller dimension to a larger dimension).
3. rank $A = n$
4. nullity $A = 0$
5. Nul $A = \{\mathbf{0}\}$

4.6 Change of Basis

Theorem: Change of Basis Matrix

Let $\mathcal{B} = \{\mathbf{b}_1, \cdots, \mathbf{b}_n\}$, $\mathcal{C} = \{\mathbf{c}_1, \cdots, \mathbf{c}_n\}$. Then there is a unique matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ such that

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

$$[\mathbf{x}]_\mathcal{C} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_\mathcal{B}$$

Note that

$$(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$$

Theorem: Finding $P_{\mathcal{C}\leftarrow\mathcal{B}}$

To find the matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$, augment all of the c vectors with all of the b vectors, then row reduce:

$$\begin{aligned} &[\mathbf{c}_1\ \mathbf{c}_2\ \cdots\ \mathbf{c}_n\ \mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_n] \\ \sim &[\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_n\ [\mathbf{b}_1]_{\mathcal{C}}\ [\mathbf{b}_2]_{\mathcal{C}}\ \cdots\ [\mathbf{b}_n]_{\mathcal{C}}] \\ = &[I\ P_{\mathcal{C}\leftarrow\mathcal{B}}] \end{aligned}$$

5 Chapter 5

5.1 Eigenvectors & Eigenvalues

Definition: Eigenvalue, Eigenvector

An eigenvector of an $n \times n$ matrix is a *nonzero vector* $\mathbf{x}$ such that

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

where λ is a scalar and is called an eigenvalue of A.

Note we could also express the above equation as

$$(A - \lambda I)\mathbf{x} = 0$$

λ is an eigenvalue if the above equation has a nontrivial solution. Note that there are only nontrivial solutions if $A - \lambda I$ is singular, or if $\det(A - \lambda I) = 0$. Therefore:

Theorem: Finding Eigenvectors and Eigenvalues

- To find the eigenvalues of a matrix, set

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

  and solve for λ.

- Upon finding eigenvalues, you can find the eigenvector corresponding to each eigenvalue by solving the linear set of equations

  $$(A - \lambda I)\mathbf{v} = 0$$

Definition: Eigenspace

The **eigenspace** of a matrix A is just the set of all solutions to $(A - \lambda I)\mathbf{x} = 0$. Since it is the null space of $A - \lambda I$, the eigenspace of a matrix is a subspace of $\mathbb{R}^n$.

Theorem: Eigenvalues of a Triangular Matrix

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

Proof: Eigenvalues of a Triangular Matrix

The equation $(A - \lambda I)\mathbf{x} = 0$ only has nontrivial solutions if there are free variables in the matrix $A - \lambda I$. If we write out $A - \lambda I$:

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

It is clear that there will be free variables if and only if λ is one of a_{11}, a_{22}, or a_{33}.

Theorem: Distinct Eigenvectors

If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$ of an $n \times n$ matrix, then the set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is linearly independent.

Proof: Distinct Eigenvectors

Suppose $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is linearly dependent. Then, there exists one vector that is a linear combination of a linearly independent group of preceding vectors. Let p be the smallest index such that $\mathbf{v}_{p+1}$ is a linear combination of the preceding (linearly independent) vectors. Then:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1}$$

If we multiply each side by the matrix $A - \lambda_{p+1}I$, we get

$$c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + c_2(\lambda_2 - \lambda_{p+1})\mathbf{v}_2 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p = 0$$

However, if all the eigenvectors are distinct, then this equation can only be true if $c_1 = c_2 = \cdots = c_p = 0$, which can only be true if $\mathbf{v}_{p+1} = \mathbf{0}$, which cannot be true by the definition of an eigenvector. Therefore, by contradiction, distinct eigenvalues implies linearly independent eigenvectors.

5.2 Characteristic Equation

Note it is possible for 0 to be an eigenvalue, if there is a nontrivial vector $\mathbf{x}$ that satisfies $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$. Therefore, the following extension of the invertible matrix theorem follows:

Theorem: Invertible Matrix Theorem (cont.)

A is invertible iff 0 is *not* an eigenvalue of A.

The characteristic equation is the equation that follows from taking $\det(A - \lambda I) = 0$. Thus:

Definition: Characteristic Equation

The **characteristic equation** or **characteristic polynomial** is a polynomial in λ that satisfies $\det(A - \lambda I) = 0$. It has the following properties:

- For an $n \times n$ matrix, the characteristic polynomial will have degree n.
- The **(algebraic) multiplicity** of an eigenvalue λ is its multiplicity as the root of the characteristic equation, e.g. in $\lambda^4(\lambda - 6)(\lambda + 2)$ the eigenvalue 0 has multiplicity 4.

Next, we discuss the topic of similarity:

Definition: Similarity

Two matrices A and B are similar if A can be expressed in terms of B and an invertible matrix P such that

$$A = P^{-1}BP$$

If the above statement holds, it follows that B can be written as

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

where $Q = P^{-1}$

It turns out that similar matrices have the same eigenvalues:

Theorem: Similar Matrices Means Same Eigenvalues

If two matrices A and B are similar, they have the same eigenvalues.

Proof: Similar Matrices, Same Eigenvalues

Suppose A can be written as $P^{-1}BP$. Then the following is true:

$$\begin{aligned}\det(A - \lambda I) &= \det(P^{-1}BP - \lambda P^{-1}IP) = \det(P^{-1}(B - \lambda I)P) \\ &= \det(P^{-1}) \cdot \det(B - \lambda I) \cdot \det(P)\end{aligned}$$

Since $\det(P^{-1}) \cdot \det(P) = \det(P^{-1}P) = \det(I) = 1$, we find that the characteristic equations of the two matrices A and B are equivalent, and therefore the eigenvalues for each matrix are the same.

Some brief warnings:

- Matrices need not be similar to have the same eigenvalues. Consider, for instance, the following set of matrices:

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

- Similarity is not the same as row equivalence. Row equivalence means for $B = EA$ for some invertible matrix E. In general, row operations change the eigenvalues of a matrix.

5.3 Diagonalization

Definition: Diagonalizable

A matrix is diagonalizable if A is similar to a diagonal matrix; that is, if $A = PDP^{-1}$ for some invertible matrix P and diagonal matrix D.

Theorem: Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable iff A has n linearly independent eigenvectors.

In fact, $A = PDP^{-1}$, with D a diagonal matrix, iff the columns of P are the n linearly independent eigenvectors of A, and the diagonal entries in D are the eigenvalues that correspond to the eigenvectors in A.

In other words, A is diagonalizable iff there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an **eigenvector basis** of $\mathbb{R}^n$.

Proof: Diagonalization Theorem

First, note that the condition $A = PDP^{-1}$ is equivalent to $AP = PD$. If we parameterize the columns of P to be $\mathbf{v}_1, \cdots, \mathbf{v}_n$, and parameterize the diagonal entries of D to be $\lambda_1, \cdots, \lambda_n$, then the following condition must be true:

$$A[\mathbf{v}_1 \ \cdots \ \mathbf{v}_n] = [\lambda_1 \mathbf{v}_1 \ \cdots \ \lambda_n \mathbf{v}_n]$$

Clearly, this condition is satisfied only if P is the matrix of eigenvectors and D is the diagonal matrix of eigenvalues. Additionally, since P is invertible, the columns $\mathbf{v}_1, \cdots, \mathbf{v}_n$ must be linearly independent.

Now, we look at an example problem:

Example: Using Eigenvectors and Eigenvalues

Find a P and diagonal D such that $A = PDP^{-1}$, for

$$A = \begin{bmatrix} -2 & 12 \\ -1 & 5 \end{bmatrix}$$

Because of the diagonalization theorem, we simply need to find the eigenvalues and eigenvectors of the matrix A to find P and D. Thus, we follow the following steps:

1. Find eigenvalues, $\det(A - \lambda I) = 0$. For the matrix A, that entails the following

$$\begin{vmatrix} -2-\lambda & 12 \\ -1 & 5-\lambda \end{vmatrix} = (-2-\lambda)(5-\lambda) - (12)(-1) = 0$$

$$\lambda = 1, 2$$

2. Find eigenvectors, solve system of equations given λ for $(A - I\lambda)\mathbf{x} = 0$. For $\lambda = 1$, solving the following augmented matrix:

$$\begin{bmatrix} -3 & 12 & 0 \\ -1 & 4 & 0 \end{bmatrix}$$

yields $x_1 = 4x_2$, or an eigenvector of

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

Using a similar method for $\lambda = 2$, the eigenvector obtained is

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

3. $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $D = I\boldsymbol{\lambda}$. For this example:

$$P = \begin{bmatrix} 1 & 1 \\ 4 & 3 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Notes:

1. An $n \times n$ matrix with n distinct eigenvalues is diagonizable.
2. If eigenvalues are *not* distinct:
   - The dimension of the eigenspace for a particular eigenvalue λ_k is less than or equal to the multiplicity of the eigenvalue.
   - The matrix A is diagonalizable iff the sum of the dimensions of the eigenspaces equals n. This only happens if 1) the polynomial factors completely into linear factors AND 2) the dimension of each eigenspace = the multiplicity.
3. If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k, then the total collection of vectors in $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

5.4 Eigenvectors and Linear Transformations

Theorem: Another Linear Transformation

Let V be an n-dimensional vector space. Let B be a basis for V. Let W be a m-dimensional vector space. Let C be a basis for W. Let T be any linear transformation from V to W ($T : V \to W$). Then:

$$[T(x)]_C = M[x]_B$$

$$M = \begin{bmatrix} [T(b_1)]_C & [T(b_2)]_C & \cdots & [T(b_n)]_C \end{bmatrix}$$

M is called the matrix for T relative to bases B and C.

If W is the same as V and the basis C is the same as B, then

$$[T(x)]_B = [T]_B[x]_B$$

where $[T]_B$ is called the **B-matrix**.

Theorem: Diagonal Matrix Representation

Suppose $A = PDP^{-1}$, where D is an $n \times n$ diagonal matrix. If B is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the B-matrix for the transformation $x \mapsto Ax$.

To find the B-matrix of A given A and the basis vectors (found in the columns of P), note that $A = PDP^{-1}$, so $D = P^{-1}AP$.

- Compute AP.
- Find P^{-1}, then compute $P^{-1}AP$ to get D where D is the B-matrix.

6 Orthogonality and Least Squares

6.1 Inner Product, Length, Orthogonality

Definition: Orthogonal Complement

The set of all vectors that are orthogonal to every vector in a subspace W is called the **orthogonal complement** of W and is denoted by $W^\perp$.

Theorem: Row, Column, Null Spaces of Orthogonal Complement

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

For a sketch of a proof, consider the following equation, which defines the vectors $\mathbf{x}$ in Nul A:

$$A\mathbf{x} = \begin{bmatrix} -- & \text{row 1} & -- \\ -- & \text{row 2} & -- \\ & \vdots & \\ -- & \text{row n} & -- \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This matrix equation clearly shows that every row vector in A, upon a dot product with $\mathbf{x}$, is 0. Therefore the row space of A is orthogonal to the null space.

6.2 Orthogonal Sets

Definition: Orthogonal Set

A set of vectors is an **orthogonal set** if each pair of distinct vectors is orthogonal.

Theorem: Orthogonal Sets are Bases

An orthogonal set of vectors S is linearly independent and hence is a basis for the subspace spanned by S.

Note that orthogonal bases are nice because for every vector $\mathbf{y}$ spanned by an orthognal set $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$, the weights in the linear combination $\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n$ can be written as

$$c_i = \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}$$

Geometrically, this consists of projecting $\mathbf{y}$ onto each vector $\mathbf{u}_i$.

Definition: Orthogonal Projection

For any vectors $\mathbf{y}$ and $\mathbf{u}$, $\mathbf{y}$ can be decomposed into a component $\hat{\boldsymbol{y}}$ that is a multiple of $\mathbf{u}$ and a component $\mathbf{z}$ that is orthogonal to $\mathbf{u}$. In other words:

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

$\hat{\boldsymbol{y}}$ is also known as the **orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$**;

$$\hat{\boldsymbol{y}} = \operatorname{proj}_L \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$

Now, we distinguish orthogonal sets with *orthonormal* sets:

Definition: Orthonormal Sets

An **orthonormal set** is an orthogonal set of unit vectors. The simplest of examples is the standard basis.

To check if a set of unit vectors is orthonormal, simply check that the dot product of every pair is 0.

Theorem: Orthonormal Columns

A matrix U has orthonormal columns iff $U^T U = I$.

The above theorem is essentially a compact way of expressing the dot product of every pair of vectors.

Theorem: Orthonormal Matrix Operations

Let U be an $m \times n$ matrix with orthonormal columns. Then:

- $||U\mathbf{x}|| = ||\mathbf{x}||$
- $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$

Essentially, the transformation $\mathbf{x} \mapsto U\mathbf{x}$ preserves length and angles.

Definition: Orthogonal Matrix

An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. Clearly, any square matrix with orthonormal columns is an orthogonal matrix.

6.3 Orthogonal Projections

Theorem: Orthogonal Decomposition Theorem

Every $\mathbf{y} \in \mathbb{R}^n$ can be written in the form

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

where $\hat{\boldsymbol{y}} \in W$, a subspace of $\mathbb{R}^n$, and $\mathbf{z} \in W^\perp$. In fact:

$$\hat{\boldsymbol{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p$$

where p is the dimension of W.

Theorem: Best Approximation Theorem

Let $\hat{\boldsymbol{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\boldsymbol{y}}$ is the closest point from $\mathbf{y}$ to W.

Theorem: Orthonormal Basis Simplifications

Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ be an orthonormal basis for the subspace W of $\mathbb{R}^n$. Then:

$$\text{proj}_W\ \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p$$

If $U = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_p]$, then

$$\text{proj}_W\ \mathbf{y} = UU^T\mathbf{y}$$

The second statement above follows because each coefficient $(\mathbf{y} \cdot \mathbf{u}_i)$ can be written instead as $(\mathbf{u}_i^T\mathbf{y})$, showing they are the entries in $U^T\mathbf{y}$.

Note the difference between UU^T and U^TU. For an $n \times p$ matrix U with orthonormal columns, let W be the column space of U. Then:

- $U^TU\mathbf{x} = I_p\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^p$

- $UU^T\mathbf{y} = \text{proj}_W\ \mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$.

If $p = n$, then U is an orthogonal matrix, $W = \mathbb{R}^n$, and $UU^T\mathbf{y} = I\mathbf{y} = \mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$.

6.4 Gram-Schmidt Process

The Gram-Schmidt Process is a method of forming an orthogonal or orthonormal basis given the basis vectors of a subspace:

1. Set one of the basis vectors as one of the basis vectors of the orthogonal basis: $\mathbf{v}_1 = \mathbf{x}_1$.

2. To construct the next orthogonal vector, take the basis vector $\mathbf{x}_i$ and subtract the projection of $\mathbf{x}_i$ on Span$\{v_1, \cdots, \mathbf{v}_{i-1}\}$. For example:

$$\mathbf{v}_3 = \mathbf{x}_3 - \left(\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2\right)$$

To obtain an orthonormal basis from an orthogonal basis, simply scale all the vectors to have a length 1.

Theorem: QR Factorization

If A is an $m \times n$ matrix with linearly independent columns, A can be factored as

$$A = QR$$

- Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A
- R is an $n \times n$ upper triangular matrix with positive entries on the diagonal.

Proof: QR Factorization

Suppose Q is constructed in the Gram-Schmidt process or some other means. Then $Q = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n]$, such that for all $1 \leq k \leq n$:

$$\mathbf{x}_k \in \text{Span}\{\mathbf{x}_1, \cdots, \mathbf{x}_k\} = \text{Span}\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$$

Therefore, we can express $\mathbf{x}_k$ as a linear combination of just the first k vectors in Q. In other words, $\mathbf{x}_k = Q\mathbf{r}_k$ where

$$\mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

We may assume $r_{kk} \geq 0$ (because if not, we can multiply both r_{kk} and $\mathbf{u}_k$ by -1). Thus, we have that

$$A = [Q\mathbf{r}_1 \ \cdots \ Q\mathbf{r}_n] = QR$$

where R is clearly an upper triangular matrix with positive entries on the diagonal. R is additionally invertible, which is a consequence of the columns of A being linearly independent.

6.5 Least Squares Problems

The solution to the least squares problem is the vector $\hat{\boldsymbol{x}}$ such that $||\mathbf{b} - A\hat{\boldsymbol{x}}||$ is minimized. Clearly, the solution satisfies the following relation:

$$A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$$

where $\hat{\boldsymbol{b}}$ is the projection of $\mathbf{b}$ onto Col A.

Theorem: Least Squares Solution

$\mathbf{b} - \hat{\boldsymbol{b}} = \mathbf{b} - A\hat{\boldsymbol{x}}$ must be orthogonal to each column of A. Therefore:

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

The above equation represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$.

Why would one complicate the equation $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$? For one, calculating $\hat{\boldsymbol{b}}$ is a pain. Additionally, if $A^T A$ is invertible, one can directly find a solution for $\hat{\boldsymbol{x}}$:

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

The above system of equations is called the **normal equations** for $A\mathbf{x} = \mathbf{b}$.

Theorem: Unique Least Squares Solution

The following statements are equivalent:

- The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.
- The columns of A are linearly independent.
- The matrix $A^T A$ is invertible.

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

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

There are alternative methods to solve the least-squares problem.

- If the columns of A are orthogonal, $\hat{\boldsymbol{b}}$ is simply given by the projection of $\mathbf{b}$ onto the columns $\mathbf{a}_i$ of A:

$$\hat{\boldsymbol{b}} = \sum \frac{\mathbf{b} \cdot \mathbf{a}_i}{\mathbf{a}_i \cdot \mathbf{a}_i} \mathbf{a}_i$$

However, upon solving for $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$, we see that this equation has already been solved. The ith entry in $\mathbf{x}$ is simply the ith coefficient in $\hat{\boldsymbol{b}}$:

$$x_i = \frac{\mathbf{b} \cdot \mathbf{a}_i}{\mathbf{a}_i \cdot \mathbf{a}_i}$$

- Alternatively, if A can be already factored into QR, then

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

One can check that this solution works:

$$A\hat{\boldsymbol{x}} = QR(R^{-1}Q^T\hat{\boldsymbol{x}}) = QQ^T\mathbf{b}$$

Since the columns of Q form an orthonormal basis for Col A, $QQ^T\mathbf{b}$ is the orthogonal projection of $\mathbf{b}$ onto Col A; in other words, $QQ^T\mathbf{b} = \hat{\boldsymbol{b}}$.

6.6 Applications to Linear Models

Let's first investigate best-fit lines/linear regression lines/least-squares lines:

Definition: Least Squares Line

Instead of $A\mathbf{x} = \mathbf{b}$, $X\boldsymbol{\beta} = \mathbf{y}$, where X is the **design matrix**, $\boldsymbol{\beta}$ is the **parameter vector**, and $\mathbf{y}$ is the **observation vector**.

Theorem: Best-Fit Line/Linear Regression Line/Least-Squares Line

Suppose you had the data points $(x_1, y_1) \cdots (x_n, y_n)$. If they all fell on a line $y = \beta_0 + \beta_1 x$, that would mean

$$\begin{gathered}\beta_0 + \beta_1 x_1 = y_1 \\ \beta_0 + \beta_1 x_2 = y_2 \\ \vdots \\ \beta_0 + \beta_1 x_n = y_n\end{gathered}$$

We could write this in matrix form:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
$$X\boldsymbol{\beta} = \mathbf{y}$$

If we instead had data points and we wanted to find the least squares regression line, that is equivalent to finding the vector $\hat{\boldsymbol{\beta}}$ such that $||\mathbf{y} - X\hat{\boldsymbol{\beta}}||$ is minimized. Therefore, we can use our least squares method from before to find

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$$

General Linear Models: not covered in class.

7 Symmetric Matrices and Quadratic Forms

7.1 Diagonalization of Symmetric Matrices

Definition: Symmetric Matrix

A matrix is symmetric if $A^T = A$. Such a matrix is necessarily square. Visually, the main diagonal is a line of symmetry.

Symmetric matrices have the following properties:

Theorem: Symmetric Matrices

- If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.
- A symmetric $\iff$ A is orthogonal diagonalizable (in other words, $A = PDP^{-1}$ where P is square, orthogonal, orthonormal and $P^T = P^{-1}$).

Proof: Symmetric Matrix Properties

- Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors of the eigenvalues λ_1 and λ_2, respectively. Then:

$$\begin{aligned}\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = \lambda_1 \mathbf{v}_1^T \mathbf{v}_2 = (\lambda \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 \\ = \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) = \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) \\ = \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2\end{aligned}$$

  Therefore

$$(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$$

  So either vectors come from the space eigenspace ($\lambda_1 = \lambda_2$), or the eigenvectors are orthogonal.

- Since the columns of P are the eigenvectors of A, P is orthogonal.

7.2 Quadratic Forms