MATH H215-A002/C

Samuel E. Ross[1]

[1] *Haverford College, Haverford, PA 19041, USA*

Fall 2021

Lecture 01: A Review of Equations and Their Systems

August 30, 2021

We begin by examining two linear equations.

$$\begin{aligned} A &: 3x + 2y = 8 \\ B &: 3x - 4y = 2 \end{aligned}$$

To obtain a solution we can combine these equations such that:

$$(-1)A + B$$

This gives the following solution for y:

$$\begin{aligned} 0x - 6y &= -6 \\ y &= 1 \end{aligned}$$

Which finally gives the following solution for x using back substitution.

$$\begin{aligned} 3x + 2(1) &= 8 \\ 3x &= 6 \\ x &= 2 \end{aligned}$$

This example gives a single, unique solution. However, not all sets of equations must have solutions. Where two equations describe parallel lines, they have no solutions and their solution set is $\emptyset$.

Some systems of equations describe the same mathematical object:

$$\begin{aligned} 4x + y &= 1 \\ 8x + 2y &= 2 \end{aligned}$$

In this example, there are infinite solutions to this system and its solution set is $\{x, 1-4x\}$.

As it turns out, the only three possibilities for a system of $n \in \mathbb{R}$ linear equations are a single unique solution, no solutions at all, or an infinite number of solutions.

Notation. We use the variables m and n to describe linear equations in the manner: a system of m linear equations in n unknowns.

Theorem 1 (The Great Trichotomy)**.** A system of m linear equations with n variables has either no solutions, one unique solution, or an infinite number of solutions.

Notation. A system of equations having no solutions is considered inconsistent while those with any non-zero number of solutions are considered consistent.

In relation to this notation are a few fundamental mathematical questions. First, that of existence: does a solution or solutions to a system of equations exist? Second, that of uniqueness: are these solutions unique?

Notation. $\mathbb{R}$ is the set of all real numbers.

Notation. $\mathbb{R}^n$ is the set of all n tuples of real numbers. This satisfies the expression $\{(x_1, \ldots, x_n) \mid x_I \in \mathbb{R}\}$.

Now, pivoting a bit, we examine the equation following.

$$x + 2y + 3z = 6$$

This equation describes a plane and has an infinite number of solutions. We should note that two planes that are non-parallel must intersect at a line, three non-parallel planes must intersect at a point, and any number of planes in which at least one is parallel to another has no solution.

Definition 1. A linear equation in n variables is an expression that can be expressed as $a_1x_1, \ldots, a_nx_n = b$ where $a_I, b \in \mathbb{R}$ and $n \in \mathbb{N}$.

Finally, we'll finish by overviewing some elementary operations use to simplify linear systems. First, we can replace one equation with the sum of a multiple of another equation and itself. Second, we can multiply one equation by the sum of a non-zero constant. Third, we can simply swap a pair of equations.

Lecture 02: Elementary Row Operations, REF, & RREF

September 1, 2021

Definition 2. An $m \times n$ matrix is a rectangular array of real numbers with m rows and n columns.

Definition 3. Two $m \times n$ matrices are row-equivalent where a sequence of elementary row operations can transform one into the other.

Theorem 2. Linear systems with row-equivalent augmented matrices have the same solution set.

Definition 4. Matrices in Row Echelon Form (REF) are defined by three qualities:

1. Any non-zero rows are above fully-zero rows.
2. The leading entry of each row (leftmost non-zero element in a given row) is in a column to the right of leading entries above it.
3. All entries below each leading entry are zeroes (follows from above).

Example. Consider a 6×9 matrix in REF where b represents the non-zero pivot and * represents any number:

$$\begin{bmatrix} 0 & 0 & b & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & b & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & b & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\diamond$

Definition 5. Matrices in Reduced Row Echelon Form (RREF) are defined by two further qualities:

4. All leading entries are one.
5. All entries above the leading entries are zero.

Example. Consider a 6×9 matrix in RREF where 1 represents the pivot and * represents any number:

$$\begin{bmatrix} 0 & 0 & 1 & * & 0 & * & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 1 & * & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\diamond$

Lecture 03: REF and Existence and Uniqueness

September 2, 2021

This discussion session concerns the use of REF and RREF to determine the nature of the solution(s) of a linear system.

Consider a linear system of 3 equations in 3 unknowns. This is the matrix after row-reduction:

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

By each variable appearing in its own unique equation, this system is guaranteed to have a unique solution.

On the other hand, consider the following system:

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

Due to the contradiction in the bottom row, this system is guaranteed to have no solutions.

Finally consider this variation:

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

Because the bottom row describes nothing about the third variable, x_3 is considered a free variable and this system has infinite solutions. Further. the solution set of this system can be described $\{(1-t, -3t, t) \mid t \in \mathbb{R}\}$.

Lecture 04: More Existence & Uniqueness & an Introduction to Vectors

September, 3, 2021

Theorem 3 (Existence & Uniqueness)**.** Given a system of m linear equations in n unknowns, form an $m \times (n+1)$ augmented matrix and reduce to RREF. This system is inconsistent if and only if there is a row in the augmented matrix of the form:

$$\begin{bmatrix} 0 & 0 & \dots & 0 & b \end{bmatrix} : b \neq 0$$

The system is consistent and has a unique solution if there are n pivots in the first n columns of the augmented matrix. If there are fewer than n pivots then the system has **free variables** and thus has infinite solutions.

Example. Consider the matrix following:

$$\begin{bmatrix} 1 & * & * & * & * & * & * & * \\ 0 & 1 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Here, we are presented with a matrix representing a consistent system with infinite solutions and the free variables x_3 and x_4 where each column represents $x_1 - x_6$ and a constant. $\diamond$

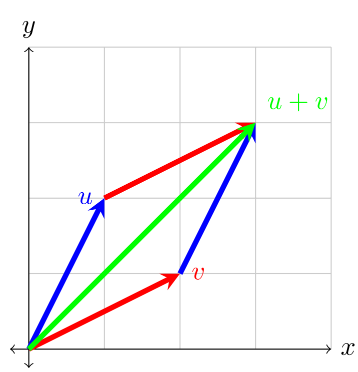
Definition 6. A vector in $\mathbb{R}^n$ is an ordered tuple of n real numbers that can be represented as follows:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} : v_i \in \mathbb{R}$$

Definition 7. Vector addition is defined algebraically as follows:

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

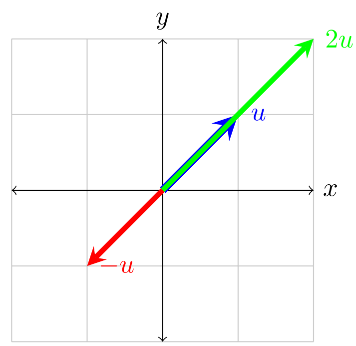
It can additionally be interpreted as the fourth vertex of a parallelogram formed by the two vectors to be added when placed tip-to-tail:



Definition 8. The scalar multiplication of vectors is defined algebraically as follows:

$$c\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} ca \\ cb \end{bmatrix}$$

Graphically, this can be interpreted as a "stretching" of the vector c times:



Definition 9. A linear combination of vectors $\mathbf{v_1}, \ldots, \mathbf{v_p} \in \mathbb{R}^n$ with weights $c_1, \ldots, c_p$ is the vector $\mathbf{w} = c_1\mathbf{v_1} + \ldots + c_p\mathbf{v_p}$.

Definition 10. The span of p vectors is defined as the set of all possible linear combinations of these vectors. This set is denoted $\text{span}(\mathbf{v_1}, \ldots, \mathbf{v_p})$.

Lecture 05: The Matrix Equation & Matrix Multiplication

September 8, 2021

We begin with a question—give the definition of $\text{span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$: This expression represents the set of all linear combinations of $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$ defined as $\{c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3} \mid c_i \in \mathbb{R}\}$.

Theorem 4. The vector equation $x_1\mathbf{a_1} + \ldots + x_n\mathbf{a_n} = \mathbf{b} : \mathbf{a_i}, \mathbf{b} \in \mathbb{R}^m$ has the same set of solutions as the linear system whose augmented matrix is:

$$\begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \ldots & \mathbf{a_n} & \mathbf{b} \end{bmatrix}$$

Theorem 5. The linear system given by $x_1\mathbf{a_1} + \ldots + x_n\mathbf{a_n} = \mathbf{b} : \mathbf{a_i}, \mathbf{b} \in \mathbb{R}^m$ is consistent if and only if $\mathbf{b}$ is a linear combination of $\mathbf{a_1}, \ldots, \mathbf{a_n}$ which is to say that $\mathbf{b} \in \text{span}(\mathbf{a_1}, \ldots, \mathbf{a_n})$.

Notation. Suppose A is a $m \times n$ matrix:

$$A = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \ldots & \mathbf{a_n} \end{bmatrix} : \mathbf{a_i} \in \mathbb{R}^m$$

Additionally consider the vector:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

The matrix multiplication of these elements gives:

$$A\mathbf{x} = x_1\mathbf{a_1} + \ldots + x_n\mathbf{a_n} : A\mathbf{x} \in \mathbb{R}^m$$

This gives us an alternate mode to express the linear system $x_1\mathbf{a_1} + \ldots + x_n\mathbf{a_n} = \mathbf{b} : \mathbf{a_i}, \mathbf{b} \in \mathbb{R}^m$ as $A\mathbf{x} = \mathbf{b}$.

Now, we give a proof of the statement $\text{span}(\mathbf{v_1}, \mathbf{v_2}) \subseteq \text{span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$:

Notation. The set-describing expression $A \subseteq B$ indicates that $\forall_{x \in A} x \in B$. Thus, to prove that $A \subseteq B$, the proof must be of the form "suppose $\forall_x x \in A \ldots \forall_x x \in B$."

Proof. Suppose $\mathbf{x} \in \text{span}(\mathbf{v_1}, \mathbf{v_2})$. Then $\mathbf{x} = c_1\mathbf{v_1} + c_2\mathbf{v_2}$ by the definition of span. So, $\mathbf{x} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + 0\mathbf{v_3}$. Therefore, $\mathbf{x} \in \text{span}(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$. □

Now, we give a proof of the statement $\text{span}(\mathbf{v_1}, \mathbf{v_2}) = \text{span}(2\mathbf{v_1}, 3\mathbf{v_2})$:

Notation. The set-describing expression $A = B$ indicates that $A \subseteq B \wedge B \subseteq A$. Therefore a proof that $A = B$ involves proving these two relationships.

Proof. Suppose $\mathbf{x} \in \text{span}(\mathbf{v_1}, \mathbf{v_2})$. So $\mathbf{x} = c_1\mathbf{v_1} + c_2\mathbf{v_2}$. Then $\mathbf{x} = \frac{1}{2}c_1(2\mathbf{v_1}) + \frac{1}{3}c_2(3\mathbf{v_2})$. Therefore $\mathbf{x} \in \text{span}(2\mathbf{v_1}, 3\mathbf{v_2})$. □

Proof. Suppose $\mathbf{x} \in \text{span}(2\mathbf{v_1}, 3\mathbf{v_2})$. So $\mathbf{x} = c_1(2\mathbf{v_1}) + c_3(3\mathbf{v_2})$. Then $\mathbf{x} = 2c_1(\mathbf{v_1}) + 3c_2\mathbf{v_2}$. Therefore $\mathbf{x} \in \text{span}(\mathbf{v_1}, \mathbf{v_2})$. □

By the previous two proofs, $\text{span}(\mathbf{v_1}, \mathbf{v_2}) = \text{span}(2\mathbf{v_1}, 3\mathbf{v_2})$.

Finally, we wrap up with some solutions to the in-class worksheet:

$$\begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 26 \\ -2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 2 \\ 17 & 4 & -1 \\ \pi & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ \sqrt{2} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 5 & 7 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} a + 3b + 5c + 7d \end{bmatrix}$$

The result of multiplying a 4×6 matrix A by a vector in $\mathbb{R}^4$ is NOT a vector in $\mathbb{R}^6$ because the matrix multiplication operation would only be defined where the input vector was in $\mathbb{R}^6$ and the resultant vector was in $\mathbb{R}^4$.

The linear system:

$$\begin{cases} 3x + 2y & = 1 \\ x - 2y + z & = 5 \\ 2x - z & = 0 \end{cases}$$

Can be expressed via the vector equation:

$$\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} x + \begin{bmatrix} 2 \\ -2 \\ 0 \end{bmatrix} y + \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} z = \begin{bmatrix} 1 \\ 5 \\ 0 \end{bmatrix}$$

Or the matrix equation:

$$\begin{bmatrix} 3 & 2 & 0 \\ 1 & -2 & 1 \\ 2 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 0 \end{bmatrix}$$

Lecture 06: The Matrix Equation, Sets, and Basic Proofs

September 9, 2021

Where does the matrix equation originate?

Consider the system:

$$\begin{cases} 3x + 2y + z = 1 \\ y + z = 0 \\ x - z = 5 \end{cases}$$

This can further be expressed as the vector equation:

$$x\begin{bmatrix}3\\0\\1\end{bmatrix}+y\begin{bmatrix}2\\1\\0\end{bmatrix}+z\begin{bmatrix}1\\1\\-1\end{bmatrix}=\begin{bmatrix}1\\0\\5\end{bmatrix}$$

Then given the definition of matrix multiplication that follows:

$$\begin{bmatrix}\mathbf{a_1} & \dots & \mathbf{a_n}\end{bmatrix}\begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix}=x_1\mathbf{a_1}+\ldots+x_n\mathbf{a_n}$$

We can then determine that this can further be expressed as:

$$\begin{bmatrix}3 & 2 & 1\\0 & 1 & 1\\1 & 0 & -1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}1\\0\\5\end{bmatrix}$$

Now, we consider some in-class problems:

Given the set $S = \{a, \{b, c\}, d\}$ with three elements, the following statements are true:

$$a \in S$$
$$\{a\} \subseteq S$$
$$\emptyset \subseteq S$$
$$S \subseteq S$$

While the following statements are false:

$$b \in S$$
$$\{a\} \in S$$
$$a \subseteq S$$
$$\{a, b\} \subseteq S$$
$$\emptyset \in S$$

We can prove that for any sets A and B such that $A \subseteq B$ then $A \cap C \subseteq B \cap C$:

Proof. Let $x \in A \cap C$. Then $x \in A \wedge x \in C$. Since $A \subseteq B$, we know that $x \in B$. So $x \in B \wedge x \in C$. Therefore, $x \in B \cap C$. $\square$

We can also prove that $(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$:

Proof. Let $x \in (A \cup B) \setminus (A \cap B)$. Then $x \in A \cup B \wedge x \notin A \cap B$. Further, $(x \in A \wedge x \notin A \cap B) \vee (x \in B \wedge x \notin A \cap B)$. Therefore, $x \in A \setminus B \vee x \in B \setminus A$. Then, $x \in (A \setminus B) \cup (B \setminus A)$. $\square$

Now, to prove that $\text{span}(\mathbf{u}, \mathbf{v}) = \text{span}(2\mathbf{u}, \mathbf{u} + \mathbf{v})$:

Proof. Suppose that $x \in \text{span}(\mathbf{u}, \mathbf{v})$. Then $\mathbf{x} = c_1\mathbf{u} + c_2\mathbf{v}$. Then, let $c_1 = 2c_3 + c_2$. This gives $\mathbf{x} = (2c_3 + c_2)\mathbf{u} + c_2\mathbf{v}$ which further gives $\mathbf{x} = 2c_3\mathbf{u} + c_2\mathbf{u} + c_2\mathbf{v}$. By simple factoring, we then have $\mathbf{x} = 2c_3\mathbf{u} + c_2(\mathbf{u} + \mathbf{v})$. Therefore, $\mathbf{x} \in \text{span}(2\mathbf{u}, \mathbf{u} + \mathbf{v})$. $\square$

Proof. Suppose that $\mathbf{x} \in \text{span}(2\mathbf{u}, \mathbf{u} + \mathbf{v})$. Then $\mathbf{x} = c_1(2\mathbf{u}) + c_2(\mathbf{u} + \mathbf{v})$. This gives $\mathbf{x} = 2c_1\mathbf{u} + c_2\mathbf{u} + c_2\mathbf{v}$. So $\mathbf{x} = (2c_1 + c_2)\mathbf{u} + c_2\mathbf{v}$. Therefore, $\mathbf{x} \in \text{span}(\mathbf{u}, \mathbf{v})$. $\square$

The proof of these two facts thus implies that $\text{span}(\mathbf{u}, \mathbf{v}) = \text{span}(2\mathbf{u}, \mathbf{u} + \mathbf{v})$.

Lecture 07: More Matrix Multiplication & Introducing Homogeneous Linear Systems

September 10, 2021

During the last lecture, we discussed the multiplication of an $m \times n$ matrix $\mathbf{A}$ where $\mathbf{A} = \begin{bmatrix} \mathbf{a_1} & \dots & \mathbf{a_n} \end{bmatrix}$ where $\mathbf{a_i} \in \mathbb{R}^m$:

$$\mathbf{A} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a_1} + \ldots + x_n\mathbf{a_n} \in \mathbb{R}^m$$

Today, we will investigate a simpler way to hand-calculate a matrix/vector product via the solutions to **homogeneous** linear systems:

$$\mathbf{Ax} = \mathbf{0}$$

Example. Consider this demonstration of matrix multiplication as a refresher:

$$\begin{bmatrix}1&1&0\\2&5&1\\3&1&7\\4&6&2\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix} = a\begin{bmatrix}1\\2\\3\\4\end{bmatrix} + b\begin{bmatrix}1\\5\\1\\6\end{bmatrix} + c\begin{bmatrix}0\\1\\7\\4\end{bmatrix}$$
$$= \begin{bmatrix}a+b\\2a+5b+c\\3a+b+7c\\4a+6b+2c\end{bmatrix}$$

◇

Now, let's examine some properties of matrix/vector multiplication:

Theorem 6. $\forall_{m\times n \text{ matrices } \mathbf{A}}$ and $\forall_{\mathbf{u},\mathbf{v}\in\mathbb{R}^n}$ and $\forall_{c\in\mathbb{R}}$:

1. $\mathbf{A}(\mathbf{u}+\mathbf{v}) = (\mathbf{A}\mathbf{u}+\mathbf{A}\mathbf{v})$
2. $\mathbf{A}(c\mathbf{u}) = c(\mathbf{A}\mathbf{u})$
3. $I_n\mathbf{u} = \mathbf{u}$ where I_n is the $n\times n$ identity matrix $I_n = \begin{bmatrix}1&0&\dots&0&0\\0&1&\dots&0&0\\\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\dots&1&0\\0&0&\dots&0&1\end{bmatrix}$

Example. Demonstrating the third criteria above:

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix} = \begin{bmatrix}a\\b\\c\end{bmatrix}$$

◇

Now, let's consider a way to describe the solutions of homogeneous linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ as the span of vectors:

Remark. $\mathbf{A}\mathbf{x} = \mathbf{0}$ is guaranteed to have either a unique solution or infinite solutions (i.e., it must be consistent) because $\mathbf{x} = \mathbf{0}$ is always a (trivial) solution.

Example. Take for example the homogeneous linear system:

$$\begin{bmatrix}1&3&5\\2&4&6\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

Whose coefficient matrix can be row-reduced to:

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

From this, we gain the equations:

$$\begin{aligned} x_1 &= x_3 \\ x_2 &= -2x_3 \end{aligned}$$

Which can further be used to represent the solutions as:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ -2x_3 \\ x_3 \end{bmatrix} \mid x_3 \in \mathbb{R}$$

This factors to:

$$x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Which is a linear combination of one vector, so the solution set can be expressed:

$$\text{span}(\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix})$$

◇

Generally, we can express the solutions to a homogeneous linear system as the span of one or more vectors, this number of vectors corresponding to the number of free variables in the equation.

Example. To demonstrate this, consider the following coefficient matrix of a homogeneous linear system:

$$\begin{bmatrix} 1 & 5 & 0 & -6 & 9 & 0 \\ 0 & 0 & 1 & -7 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This system has the free variables x_2, x_4, and x_5 which can be expressed in terms of the basic variables as:

$$\begin{aligned} x_1 &= -5x_2 + 6x_4 - 9x_5 \\ x_3 &= 7x_4 - 4x_5 \\ x_6 &= 0 \end{aligned}$$

Which gives us the solutions:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} -5x_2 + 6x_4 - 9x_5 \\ x_2 \\ 7x_4 - 4x_5 \\ x_4 \\ x_5 \\ 0 \end{bmatrix}$$

Which can finally be expressed as the linear combination:

$$\text{span}(\begin{bmatrix} -5 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 0 \\ 7 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -9 \\ 0 \\ -4 \\ 1 \\ 0 \\ 0 \end{bmatrix}) = x_2 \begin{bmatrix} -5 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 6 \\ 0 \\ 7 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -9 \\ 0 \\ -4 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

◇

Lecture 08: Linear Independence & Dependence

September 13, 2021

Last time, we examined solutions of linear systems of the homogeneous form $\mathbf{Ax} = \mathbf{0}$. Systems of this form:

1. Are always consistent, because of the trivial solution $\mathbf{x} = \mathbf{0}$.
2. Where there are infinite solutions, can always be represented as the span of some number of vectors, which is the number of free variables in the system.

Today, we will examine the concepts of linear dependence and independence as well as how they are connected to span.

Definition 11. A set of vectors $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\} \in \mathbb{R}^n$ is **linearly independent** if the vector equation (specifically referred to as the "dependence relation") $c_1\mathbf{v_1} + \ldots + c_p\mathbf{v_p} = \mathbf{0}$ has only the trivial solution. If the dependence relation has non-trivial solutions, then the vectors are **linearly dependent**. The dependence relation can be expressed by the matrix equation:

$$\begin{bmatrix} \mathbf{v_1} & \ldots & \mathbf{v_p} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = \mathbf{0}$$

Which is a homogeneous linear system. Therefore, determining whether a set of vectors $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\} \in \mathbb{R}^n$ is linearly independent amounts to finding the solution set of the above homogeneous linear system.

Example. Consider the set of vectors:

$$\{ \begin{bmatrix} 1 \\ -4 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -9 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -6 \end{bmatrix} \}$$

Are these vectors linearly independent or dependent? This can be addressed by determining whether the following system has non-trivial solutions:

$$c_1 \begin{bmatrix} 1 \\ -4 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ -9 \\ -3 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 2 \\ -6 \end{bmatrix} = \mathbf{0}$$

Which we can discover by forming the system's augmented matrix and row-reducing:

$$\begin{bmatrix} 1 & 3 & 1 & 0 \\ -4 & -9 & 2 & 0 \\ 0 & -3 & -6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From this, we know that the system has c_3 as a free variable, which means there are infinite non-trivial solutions and thus the system is linearly dependent. We can further determine a counter-example by solving for the basic variables in terms of the free variable:

$$\begin{aligned} c_1 &= 5c_3 \\ c_2 &= -2c_3 \end{aligned}$$

Therefore, $c_1 = 5$, $c_2 = -2$, and $c_3 = 1$ is one non-trivial solution to the system that—by definition—satisfies:

$$5 \begin{bmatrix} 1 \\ -4 \\ 0 \end{bmatrix} + -2 \begin{bmatrix} 3 \\ -9 \\ -3 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 2 \\ -6 \end{bmatrix} = \mathbf{0}$$

$\diamond$

Now, let's examine a few more examples:

Example. Is the following set of vectors linearly independent or dependent?

$$\{ \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \}$$

Because the augmented matrix for the dependence relation $\mathbf{Ax} = \mathbf{0}$ can be row-reduced as follows:

$$\begin{bmatrix} 3 & 1 & 0 & 0 \\ 4 & 0 & 1 & 0 \\ 5 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Then the dependence relation has the unique, trivial solution $\mathbf{x} = \mathbf{0}$ and the given set is thus linearly independent. $\diamond$

Example. Is the following set of vectors linearly independent or dependent?

$$\{ \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \}$$

Because this set contains the zero vector, we can immediately recognize that a subset of the solution set of the dependence relation can be expressed as $\{(0, 0, c_3) \in \mathbb{R}^3 \mid c_3 \in \mathbb{R}\}$ (one member is $(0, 0, 5)$, for example). This set contains more than just the trivial solution and thus this set is linearly dependent. This example alludes to the theorem that we will later prove that any set of vectors containing the zero vector must be linearly dependent. $\diamond$

Example. Is the following set of vectors linearly independent or dependent?

$$\{ \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \}$$

Because the augmented matrix for the dependence relation $\mathbf{Ax} = \mathbf{0}$ can be row-reduced as follows:

$$\begin{bmatrix} 3 & 6 & 0 & 0 \\ 4 & 8 & 1 & 0 \\ 5 & 10 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Then the dependence relation has an infinite number of non-trivial solutions (because c_3 is a free variable) and the given set is thus linearly dependent. This example alludes to the theorem that any set of vectors that contains a scalar multiple of another vector in the set must be linearly dependent. $\diamond$

Example. Is the following set of vectors linearly independent or dependent?

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

Because the augmented matrix for the dependence relation $\mathbf{Ax} = \mathbf{0}$ can be row-reduced as follows:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Then the dependence relation has the unique, trivial solution $\mathbf{x} = \mathbf{0}$ and the given set is thus linearly independent. $\diamond$

Example. Must a set of 4 vectors be linearly dependent in $\mathbb{R}^3$?

Yes, because the coefficient matrix formed by the vectors' dependence relation is 3×4 and therefore it cannot contain a pivot in every column. Thus, the system is guaranteed to have a free variable and thus infinite solutions, meaning that the set is linearly dependent. $\diamond$

> **Theorem 7.** The columns of a matrix $A = \begin{bmatrix} \mathbf{v_1} & \dots & \mathbf{v_p} \end{bmatrix}$ are linearly independent if and only if the system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, which means that $\mathbf{Ax} = \mathbf{0}$ has no free variables, which means that $\mathbf{A}$ has a pivot in every column.

Now, let's consider the following question: What conditions on $\mathbf{A} = \begin{bmatrix} \mathbf{v_1} & \dots & \mathbf{v_p} \end{bmatrix}$ will ensure that its columns span $\mathbb{R}^n$? In other words, what conditions on $\mathbf{A}$ will guarantee that every $\mathbf{b} \in \mathbb{R}^n$ can be formed by a linear combination of the columns of $\mathbf{A}$? This is further asking whether the system $\mathbf{Ax} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{R}^n$.

> **Theorem 8.** The system $\mathbf{Ax} = \mathbf{b}$ is consistent for all $\mathbf{b}$ if $\mathbf{A}$ contains a pivot in every **row**.

We'll pick up here during the next lecture.

Lecture 09: Proofs Involving Linear Independence & Dependence

September 15, 2021

> **Definition 12.** The set of vectors $\{\mathbf{v_1}, \dots, \mathbf{v_p}\}$ is linearly independent if the dependence relation $c_1\mathbf{v_1} + \dots + c_p\mathbf{v_p} = \mathbf{0}$ has only the trivial solution (i.e., $c_1, \dots, c_p = 0$). If more solutions exist, then the set is linearly dependent.

Theorem 9. Given an $m \times n$ matrix $\mathbf{A}$, the columns of $\mathbf{A}$ are linearly independent if and only if:

1. $\mathbf{Ax} = \mathbf{0}$ has $\mathbf{x} = \mathbf{0}$ as a unique solution.
2. $\mathbf{Ax} = \mathbf{0}$ has no free variables.
3. $\mathbf{A}$ has a pivot in every column.

These are three logically equivalent statements.

Theorem 10. Given an $m \times n$ matrix $\mathbf{A}$, the columns of $\mathbf{A}$ span $\mathbb{R}^m$ if and only if:

1. Every $\mathbf{b} \in \mathbb{R}^m$, given $\mathbf{Ax} = \mathbf{b}$, is a linear combination of the columns of $\mathbf{A}$.
2. $\forall \mathbf{b} \in \mathbb{R}^m \mid \mathbf{Ax} = \mathbf{b}$ has a solution.
3. $\mathbf{A}$ has a pivot in every row.

These are also three logically equivalent statements.

Now, let's work on some proofs of linear independence:

Suppose that $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{a}\}$ is linearly independent. Prove that $\{\mathbf{v_1} + \mathbf{a}, \mathbf{v_2} + \mathbf{a}, \mathbf{v_3} + \mathbf{a}\}$ is also linearly independent.

Proof. Suppose that $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{a}\}$ is linearly independent. Suppose also that $\{\mathbf{v_1} + \mathbf{a}, \mathbf{v_2} + \mathbf{a}, \mathbf{v_3} + \mathbf{a}\}$ is linearly independent. Then, the dependence relation $c_1\mathbf{v_1} + \mathbf{a} + c_2\mathbf{v_2} + \mathbf{a} + c_3\mathbf{v_3} + \mathbf{a} = \mathbf{0}$ has only the trivial solution. This dependence relation can be algebraically manipulated to yield $c_1\mathbf{v_1} + c_1\mathbf{a} + c_2\mathbf{v_2} + c_2\mathbf{a} + c_3\mathbf{v_3} + c_3\mathbf{a} = \mathbf{0}$ and can be further factored into the form $c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3} + (c_1 + c_2 + c_3)\mathbf{a} = \mathbf{0}$. By hypothesis, we know that $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{a}\}$ satisfies the dependence relation $c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3} + c_4\mathbf{a} = \mathbf{0}$ with only the trivial solution; therefore, we can state that $c_1 = c_2 = c_3 = 0$ and further that $c_1 + c_2 + c_3 = 0$. Thus, the set of vectors $\{\mathbf{v_1} + \mathbf{a}, \mathbf{v_2} + \mathbf{a}, \mathbf{v_3} + \mathbf{a}\}$ is linearly independent. $\square$

Suppose that $\{\mathbf{v_1}, \mathbf{v_2}\}$ is linearly dependent. Prove that one of the vectors must be a scalar multiple of the other.

Proof. Suppose that the set of vectors $\{\mathbf{v_1}, \mathbf{v_2}\}$ is linearly dependent. So, the dependence relation $c_1\mathbf{v_1} + c_2\mathbf{v_2} = \mathbf{0}$ has non-trivial solutions. We can express $c_1\mathbf{v_1} = -c_2\mathbf{v_2}$. Where $c_1 \neq 0$, then $\mathbf{v_1} = -\frac{c_2}{c_1}\mathbf{v_2}$ and conversely where $c_2 \neq 0$, $\mathbf{v_2} = -\frac{c_1}{c_2}\mathbf{v_1}$. Thus, these vectors are scalar multiples of each other. □

Example. Now, a brief side-note: where is $\{\mathbf{v}\}$ linearly independent or dependent?

Where $\mathbf{v} = \mathbf{0}$, then $\{\mathbf{0}\}$ is dependent because $c_1\mathbf{0} = \mathbf{0}$ holds for all $c_1 \in \mathbb{R}$, demonstrating non-trivial solutions. Where $\mathbf{v} \neq \mathbf{0}$, then $\{\mathbf{v}\}$ is linearly independent because $c_1 = 0$, the trivial solution, is the only solution. ◇

Now, let's work through four true-false questions from the in-class worksheet:

Example.

- The statement "4 vectors in $\mathbb{R}^3$ must span $\mathbb{R}^3$" is false. One counterexample is the case in which all vectors are the zero vector.
- The statement "2 vectors in $\mathbb{R}^3$ can never span $\mathbb{R}^3$" is true. This is because their linear combination has only two variables at maximum, less than the requisite three.
- The statement "4 vectors in $\mathbb{R}^3$ can be linearly independent" is false. This is because their dependence relation's coefficient matrix is guaranteed to have a free variable and thus non-trivial solutions.
- The statement "2 vectors in $\mathbb{R}^3$ must be linearly independent" is false. A counterexample is found where these vectors are scalar multiples of each other.

◇

Theorem 11. A set of vectors $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\} \in \mathbb{R}^n$ is linearly dependent if and only if one of the vectors is a linear combination of the others.

The proof of this theorem contains two conditions:

1. If one of the vectors is a linear combination of the others then $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is linearly dependent.

   Proof. Suppose one of the vectors is a linear combination of the others. Then, renumber the vectors such that $\mathbf{v_1}$ is a linear combination of $\mathbf{v_2}, \ldots, \mathbf{v_p}$. Then $\mathbf{v_1} = c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}$ and further $-\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p} = \mathbf{0}$. This is a non-trivial linear combination of $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ equalling $\mathbf{0}$; thus, $\{\mathbf{v_1}, \ldots \mathbf{v_p}\}$ is dependent. □

2. The second condition of this proof will be explored during Lecture 11.

Lecture 10: Logic and Linear Independence Proofs

September 16, 2021

Let us first examine the sixth in-class problem:

Example. Consider the linear system:

$$\begin{cases} x + 2y - z & = 1 \\ y + 2z & = 2 \end{cases}$$

This system's augmented matrix can be represented in REF and RREF as:

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

Which tells us that z is a free variable and that it can be represented in terms of the basic variables as:

$$\begin{aligned} x &= 5z - 3 \\ y &= -2z + 2 \end{aligned}$$

Which yields the solution set in vector form:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5z - 3 \\ 2 - 2z \\ z \end{bmatrix}$$

Which can be factored into:

$$z\begin{bmatrix}5\\-2\\1\end{bmatrix}+\begin{bmatrix}-3\\2\\0\end{bmatrix}=\text{span}(\begin{bmatrix}5\\-2\\1\end{bmatrix})+\begin{bmatrix}-3\\2\\0\end{bmatrix}$$

Note that the latter span term represents the solution set of the homogeneous linear system:

$$\begin{cases}x+2y-z & =0\\ y+2z & =0\end{cases}$$

$\diamond$

Next, let's consider the fourth question:

Example. The statement "If $m > n$ then there exists a vector $\mathbf{b} \in \mathbb{R}^m$ such that $\mathbf{Ax} = \mathbf{b}$ is inconsistent" can be logically represented as:

$$\forall\, m \times n\, \mathbf{A}, m > n \Rightarrow \exists\, \mathbf{b} \in \mathbb{R}^m,\ \forall \mathbf{x} \in \mathbb{R}^n,\ \mathbf{Ax} \neq \mathbf{b}$$

This is true because if we put such a matrix into RREF, then we could insert a non-zero value into the last row of $\mathbf{b}$ and this would yield a contradiction, making the system inconsistent.

Now, given the rules:

- $\neg(P \Rightarrow Q) \Leftrightarrow P \wedge Q$
- $\neg(\forall x P(x)) \Leftrightarrow \exists x \neg P(x)$
- $\neg(\exists x P(x)) \Leftrightarrow \forall x \neg P(x)$

This can be negated as:

$$\exists\, m \times n\, \mathbf{A},\ m > n \wedge \forall\, \mathbf{b} \in \mathbb{R}^m,\ \exists\, \mathbf{x} \in \mathbb{R}^n,\ \mathbf{Ax} = \mathbf{0}$$

As the original statement is true, the negated statement is false. $\diamond$

Now, let's consider the first question:

Example.

1. Where $\mathbf{A}$ is an 4×5 matrix, it is impossible for the equation $\mathbf{Ax} = \mathbf{b}$ to have a unique solution because a pivot cannot be placed into every column; however, the equation may have no solutions (in the case of a row of form $\begin{bmatrix}0 & \dots & 0 & b\end{bmatrix} : b \neq 0$ in the augmented matrix) or infinite solutions (every other case).

2. Where $\mathbf{A}$ is an 5×4 matrix, it is possible for the equation $\mathbf{Ax} = \mathbf{b}$ to have no solutions (in the case of a row of form $\begin{bmatrix} 0 & \dots & 0 & b \end{bmatrix} : b \neq 0$ in the augmented matrix), a unique solution (where a pivot can be found in every column), and infinite solutions (where one or more free variables exist).

$\diamond$

Finally, we consider the fifth question:

Example.

1. Is $\{\mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3}, \mathbf{v_1} + \mathbf{v_2}, \mathbf{2v_3}\}$ linearly independent?

   No, the dependence relation of this set, $c_1(\mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3}) + c_2(\mathbf{v_1} + \mathbf{v_2}) + c_3(\mathbf{2v_3}) = 0$, reveals that the vector $\mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3}$ can be expressed as a linear combination of the other two vectors (i.e., $(\mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3}) = (\mathbf{v_1} + \mathbf{v_2}) + \frac{1}{2}(2\mathbf{v_3}))$ and therefore the given set is linearly dependent.

2. Prove that $\{\mathbf{v_1} + \mathbf{v_2}, \mathbf{v_1} + \mathbf{v_3}, \mathbf{v_2} + \mathbf{v_3}\}$ is linearly independent.

   Proof. Suppose $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly independent. Also suppose that $c_1(\mathbf{v_1} + \mathbf{v_2}) + c_2(\mathbf{v_1} + \mathbf{v_3}) + c_3(\mathbf{v_2} + \mathbf{v_3}) = \mathbf{0}$. Then, $c_1\mathbf{v_1} + c_1\mathbf{v_2} + c_2\mathbf{v_1} + c_2\mathbf{v_3} + c_3\mathbf{v_2} + c_3\mathbf{v_3} = \mathbf{0}$. So, $(c_1 + c_2)\mathbf{v_1} + (c_1 + c_3)\mathbf{v_2} + (c_2 + c_3)\mathbf{v_3} = \mathbf{0}$. Since, by hypothesis $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ are linearly independent, then we can form the following system:

$$\begin{cases} c_1 + c_2 & = 0 \\ c_1 + c_3 & = 0 \\ c_2 + c_3 & = 0 \end{cases}.$$

   This yields the unique solution $c_1 = c_2 = c_3 = 0$ and thus $\{\mathbf{v_1} + \mathbf{v_2}, \mathbf{v_1} + \mathbf{v_3}, \mathbf{v_2} + \mathbf{v_3}\}$ is linearly independent. $\square$

$\diamond$

Lecture 11: Introducing Linear Transformations

September 17, 2021

Today, we will explore linear transformations (or functions, or maps).

Notation. The notation

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

denotes a function T that transforms inputs in $\mathbb{R}^n$, the domain, to outputs $\mathbb{R}^m$, the co-domain.

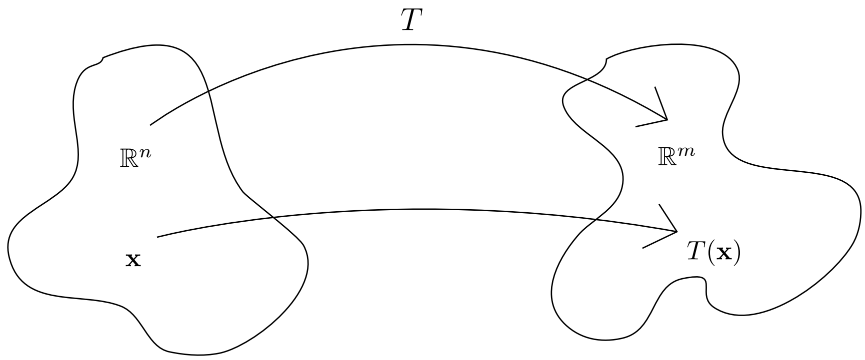


Figure 1: A Graphical Representation of T

Definition 13. A linear transformation T is a function that satisfies:

- $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n, T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
- $\forall \mathbf{u} \in \mathbb{R}^n \wedge c \in \mathbb{R}, T(c\mathbf{u}) = cT(\mathbf{u})$

For some examples of linear transformations, we can consider the derivative and the function of projection onto the x-axis (Figure 2).

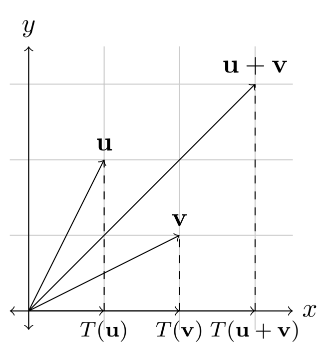


Figure 2: The Linearity of Projection

Now, let's determine if some transformations are linear:

1. For $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} 2x_2 \\ 0 \\ 3x_1 \end{bmatrix},$$

we can verify the first condition by

$$\begin{aligned}
T(\mathbf{u}+\mathbf{v}) &= T(\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \end{bmatrix}) \\
&= \begin{bmatrix} 2u_2+2v_2 \\ 0 \\ 3u_1+3v_1 \end{bmatrix} \\
&= \begin{bmatrix} 2u_2 \\ 0 \\ 3u_1 \end{bmatrix} + \begin{bmatrix} 2v_2 \\ 0 \\ 3v_1 \end{bmatrix} \\
&= T(\mathbf{u}) + T(\mathbf{v}),
\end{aligned}$$

and the second condition by

$$\begin{aligned}
T(c\mathbf{u}) &= T(\begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}) \\
&= \begin{bmatrix} 2cu_2 \\ 0 \\ 3cu_1 \end{bmatrix} \\
&= c\begin{bmatrix} 2u_2 \\ 0 \\ 3u_1 \end{bmatrix} \\
&= cT(\mathbf{u}).
\end{aligned}$$

Therefore, the given transformation is linear.

2. For $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} 2x_2 \\ 1 \\ 3x_1 \end{bmatrix},$$

we can find that the first condition does not hold by

$$\begin{aligned}
T(\mathbf{u}+\mathbf{v}) &= T(\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \end{bmatrix}) \\
&= \begin{bmatrix} 2u_2+2v_2 \\ 1 \\ 3u_1+3v_1 \end{bmatrix} \\
&\neq \begin{bmatrix} 2u_2 \\ 1 \\ 3u_1 \end{bmatrix} + \begin{bmatrix} 2v_2 \\ 1 \\ 3v_2 \end{bmatrix} \\
&= T(\mathbf{u}) + T(\mathbf{v}),
\end{aligned}$$

Therefore, the given transformation cannot be linear.

3. For $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix},$$

we can verify the first condition by

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}) \\ &= \begin{bmatrix} u_2 + v_2 \\ -u_1 - v_1 \end{bmatrix} \\ &= \begin{bmatrix} u_2 \\ -u_1 \end{bmatrix} + \begin{bmatrix} v_2 \\ -v_1 \end{bmatrix} \\ &= T(\mathbf{u}) + T(\mathbf{v}), \end{aligned}$$

and the second condition by

$$\begin{aligned} T(c\mathbf{u}) &= T(\begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}) \\ &= \begin{bmatrix} cu_2 \\ -cu_1 \end{bmatrix} \\ &= c\begin{bmatrix} u_2 \\ -u_1 \end{bmatrix} \\ &= cT(\mathbf{u}). \end{aligned}$$

Therefore, the given transformation is linear.

4. For $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} (x_1)^2 \\ 0 \end{bmatrix},$$

we can find that the first condition does not hold by

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}) \\ &= \begin{bmatrix} (u_1 + v_1)^2 \\ 0 \end{bmatrix} \\ &\neq \begin{bmatrix} (u_1)^2 \\ 0 \end{bmatrix} + \begin{bmatrix} (v_1)^2 \\ 0 \end{bmatrix} \\ &= T(\mathbf{u}) + T(\mathbf{v}), \end{aligned}$$

Therefore, the given transformation cannot be linear.

Consider the first example above. Because this system is linear, there exists a 3×2 matrix $\mathbf{A}$ such that $\mathbf{Ax} = \mathbf{b}$ is equivalent to $T(\mathbf{x}) = \mathbf{b}$. This matrix is represented in the equation

$$\begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ 0 \\ 3x_1 \end{bmatrix}.$$

Further, for the third example, this transformation can be represented

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix}.$$

Interestingly, in the latter case, this transformation can be interpreted as rotating a vector $\mathbf{u} \in \mathbb{R}^2$ by 90-degrees clockwise (see Figure 3).

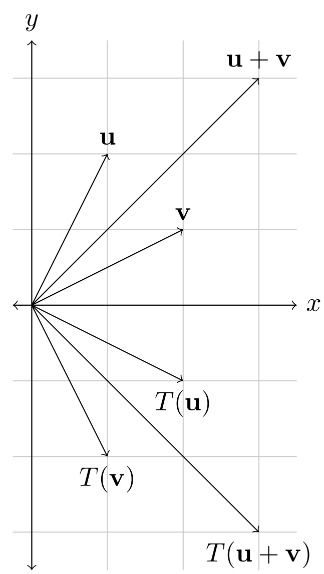


Figure 3: Some Vectors Under the Third Transformation

Does it make sense for this sort of transformation to be linear? Consider the rotation of the vector addition parallelogram that is implied in Figure 3—the entire shape is rotating by the same amount.
Finally, let's consider the final in-class problem:

Given a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ and the knowledge that $T(\begin{bmatrix} 1 \\ 0 \end{bmatrix}) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $T(\begin{bmatrix} 0 \\ 1 \end{bmatrix}) = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, we can easily find $T(\begin{bmatrix} 2 \\ 3 \end{bmatrix})$ by representing it as

$T(2\begin{bmatrix}1\\0\end{bmatrix}+3\begin{bmatrix}0\\1\end{bmatrix})$ and leveraging the linearity to find that

$$\begin{aligned}T(\begin{bmatrix}2\\3\end{bmatrix}) &= T(2\begin{bmatrix}1\\0\end{bmatrix}+3\begin{bmatrix}0\\1\end{bmatrix})\\ &= 2T(\begin{bmatrix}1\\0\end{bmatrix})+3T(\begin{bmatrix}0\\1\end{bmatrix})\\ &= 2\begin{bmatrix}1\\2\\3\end{bmatrix}+3\begin{bmatrix}0\\1\\-1\end{bmatrix}\\ &= \begin{bmatrix}2\\7\\3\end{bmatrix}.\end{aligned}$$

We leave off with a final theorem, the second stipulation of which will not be proved here.

Theorem 12. If $T:\mathbb{R}^m \to \mathbb{R}^n$ is linear, then

1. $T(\mathbf{0}) = \mathbf{0}$
2. $T(c_1\mathbf{v_1}+\ldots+c_p\mathbf{v_p}) = c_1T(\mathbf{v_1})+\ldots+c_pT(\mathbf{v_p})$

Proof. 1. $T(\mathbf{0}+\mathbf{0}) = T(\mathbf{0})+T(\mathbf{0})$, but $T(\mathbf{0}+\mathbf{0}) = T(\mathbf{0})$. So, $T(\mathbf{0}) = 2T(\mathbf{0})$ and then $\mathbf{0} = T(\mathbf{0})$.

□

Lecture 12: Standard Matrices & One-to-One/Onto Functions

September 20, 2021

Today, we will cover two main ideas—the matrix representation of linear equations and one-to-one/onto functions.

Theorem 13. If $T:\mathbb{R}^n \to \mathbb{R}^m$ is linear, then there exists a unique $m\times n$ matrix $\mathbf{A}$ such that $\forall \mathbf{x}\in\mathbb{R}^n, T(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Further, this matrix can be represented $\mathbf{A} = \begin{bmatrix}T(\mathbf{e_1}) & \ldots & T(\mathbf{e_n})\end{bmatrix}$.

Notation. The vector $\mathbf{e_i}\in\mathbb{R}^n$ refers to the ith column of the identity matrix $\mathbf{I_n}$.

Example. Let's consider the worksheet's applied problems.

1. Given the linear function $T:\mathbb{R}^2\to\mathbb{R}^3$ represented

$$T(\begin{bmatrix}x_1\\x_2\end{bmatrix}) = \begin{bmatrix}3x_1+2x_2\\-x_1\\3x_2\end{bmatrix},$$

we can determine this transformation's standard matrix $\mathbf{A}$ such that $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ by finding that

$$T(\mathbf{e_1}) = T(\begin{bmatrix} 1 \\ 0 \end{bmatrix}) = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$$
$$T(\mathbf{e_2}) = T(\begin{bmatrix} 0 \\ 1 \end{bmatrix}) = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}.$$

Using the theorem above, we can then state that this function's standard matrix is

$$\begin{bmatrix} T(\mathbf{e_1}) & T(\mathbf{e_2}) \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ -1 & 0 \\ 0 & 3 \end{bmatrix}.$$

2. Given the linear function $T : \mathbb{R}^3 \to \mathbb{R}^3$ represented

$$T(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} 2x_2 + x_3 \\ 0 \\ 3x_1 \end{bmatrix},$$

we can determine this transformation's standard matrix $\mathbf{A}$ such that $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ by finding that

$$T(\mathbf{e_1}) = T(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}) = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$
$$T(\mathbf{e_2}) = T(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}) = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$
$$T(\mathbf{e_3}) = T(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Using the theorem above, we can then state that this function's standard matrix is

$$\begin{bmatrix} T(\mathbf{e_1}) & T(\mathbf{e_2}) & T(\mathbf{e_3}) \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix}.$$

3. Given the linear function $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T(\mathbf{x})$ represents a reflection of $\mathbf{x}$ across the x-axis and a subsequent 90° anticlockwise rotation, we can determine this transformation's standard matrix $\mathbf{A}$ such that

$T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ by using either visual intuition or the definitions of these transformations to find that

$$T(\mathbf{e_1}) = T(\begin{bmatrix} 1 \\ 0 \end{bmatrix}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$T(\mathbf{e_2}) = T(\begin{bmatrix} 0 \\ 1 \end{bmatrix}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Using the theorem above, we can then state that this function's standard matrix is

$$\begin{bmatrix} T(\mathbf{e_1}) & T(\mathbf{e_2}) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

$\diamond$

Now, with some intuition established, we can go about proving Theorem 13.

Proof. If $\mathbf{x} \in \mathbb{R}^n$, then

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ v_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \ldots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = x_1\mathbf{e_1} + \ldots + x_n\mathbf{e_n}.$$

Applying a linear transformation T to both sides then yields

$$T(\mathbf{x}) = T(x_1\mathbf{e_1} + \ldots + x_n\mathbf{e_n}).$$

Then, using the linearity of T, we can express this as

$$\begin{aligned} T(\mathbf{x}) &= x_1 T(\mathbf{e_1}) + \ldots + x_n T(\mathbf{e_n}) \\ &= \begin{bmatrix} T(\mathbf{e_1}) & \ldots & T(\mathbf{e_n}) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \end{aligned}$$

This completes the proof that a matrix $\mathbf{A}$ such that $\forall \mathbf{x} \in \mathbb{R}^n, T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ exists for all linear transformations T. Now, we will show that this is the only such matrix which has this property. Suppose $\forall \mathbf{x} \in \mathbb{R}^n, T(\mathbf{x}) = \mathbf{B}\mathbf{x}$. Then

$$\begin{aligned} T(\mathbf{e_1}) &= \mathbf{B}\mathbf{e_1} = \text{1st column of } \mathbf{B} \\ &\vdots \\ T(\mathbf{e_n}) &= \mathbf{B}\mathbf{e_n} = n\text{th column of } \mathbf{B}. \end{aligned}$$

This demonstrates that this representation is unique, thus completing the proof. $\square$

Now, let's examine some properties of functions $T : \mathbb{R}^n \to \mathbb{R}^m$.

Definition 14. A function T is one-to-one (or injective) if $T\mathbf{x}$ has at most one solution for every $b \in \mathbb{R}^m$ (i.e., unique solution for every output). In other words, $T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow \mathbf{x} = \mathbf{y}$. This property and it's lack thereof can further be expressed graphically as in Figure 4 and Figure 5.

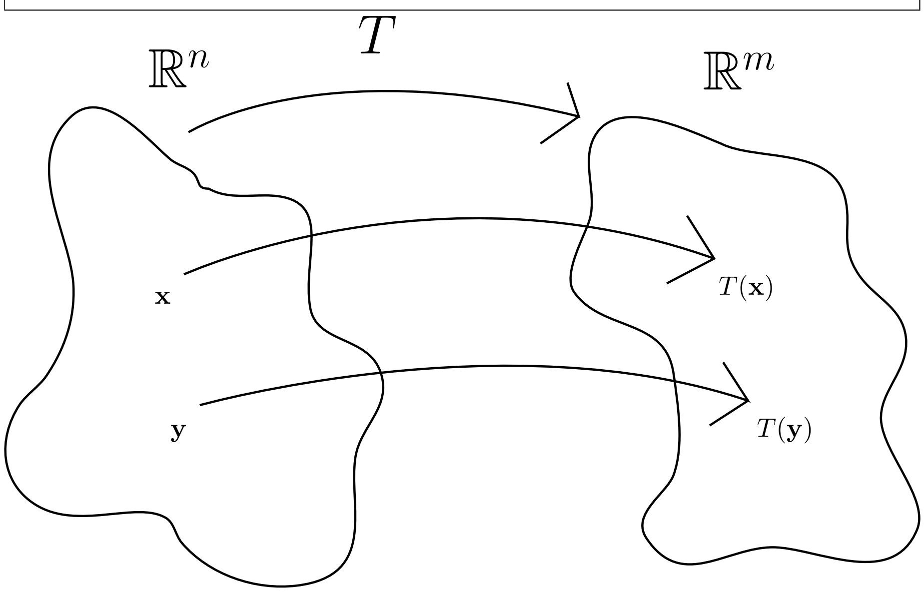


Figure 4: A One-to-One Function T

Definition 15. A function T is onto ($\mathbb{R}^m$) (or surjective) if $T(\mathbf{x}) = \mathbf{b}$ has solutions for all $b \in \mathbb{R}^m$ (i.e., a solution exists for every possible output). This can be expressed graphically as in Figure 6.

Example. Let's consider some functions $f : \mathbb{R} \to \mathbb{R}$.

- $f(x) = \sin(x)$ is not one-to-one because $\sin(0) = \sin(\pi) \not\Rightarrow 0 = \pi$ and is not onto $\mathbb{R}$ because $\sin(x) = 5$ has no solution.
- $g(x) = 3x+2$ is one-to-one because $g(x) = g(y) \Rightarrow 3x+2 = 3y+2 \Rightarrow x = y$ and is onto $\mathbb{R}$ because $\forall b \in \mathbb{R}$, the equation $3x + 2 = b$ is consistent.
- $h(x) = x^3 - x$ is not one-to-one because $0 = h(0) = h(1) \not\Rightarrow 0 = 1$ but is onto $\mathbb{R}$ because $\forall b \in \mathbb{R}$, the equation $x^3 + x = b$ is consistent.
- $j(x) = 1-\frac{1}{e^x}$ is one-to-one because $j(x) = j(y) \Rightarrow 1-\frac{1}{e^x} = 1-\frac{1}{e^y} \Rightarrow x = y$ but is not onto because the equation $1 - \frac{1}{e^x} = 2$ has no solution.

◇

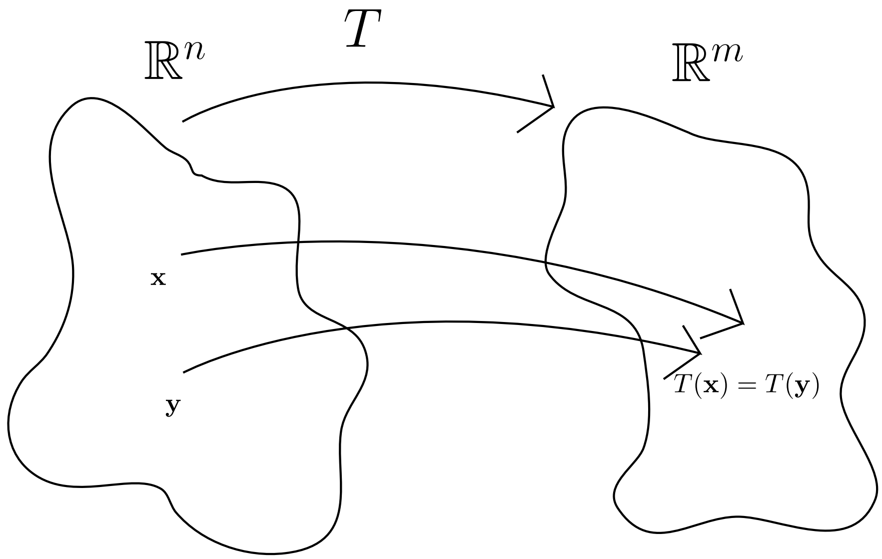


Figure 5: A Non-One-to-One Function T

> **Theorem 14.** Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear with standard matrix A. T is one-to-one if A has a pivot in every column. T is onto if A has a pivot in every row.

> **Theorem 15.** Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear. T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$.

We will give a proof of Theorem 15 during the next class.

Lecture 13: More on One-to-One & Onto Functions

September 22, 2021

This is the last lecture that covers material that will be present on the first exam.

Now, let's explore more about one-to-one and onto functions.

One-to-one-ness and onto-ness are fundamental properties of all functions. In this course, we have learned to encode linear systems in equations of form $\mathbf{Ax} = \mathbf{b}$, where questions of existence were interpreted as "$\forall \mathbf{b} \in \mathbb{R}^n$ does $\mathbf{Ax} = \mathbf{b}$ have a solution?" and questions of uniqueness were interpreted as "if given a consistent system, does $\mathbf{Ax} = \mathbf{b}$ have more than one solution?" Onto-ness and one-to-one-ness are analogous concepts when applied to functions, with existence corresponding to onto-ness and uniqueness corresponding to one-to-oneness. With onto-ness we can ask "$\forall \mathbf{b} \in \mathbb{R}^n$ does $T(\mathbf{x}) = \mathbf{b}$ have a solution where

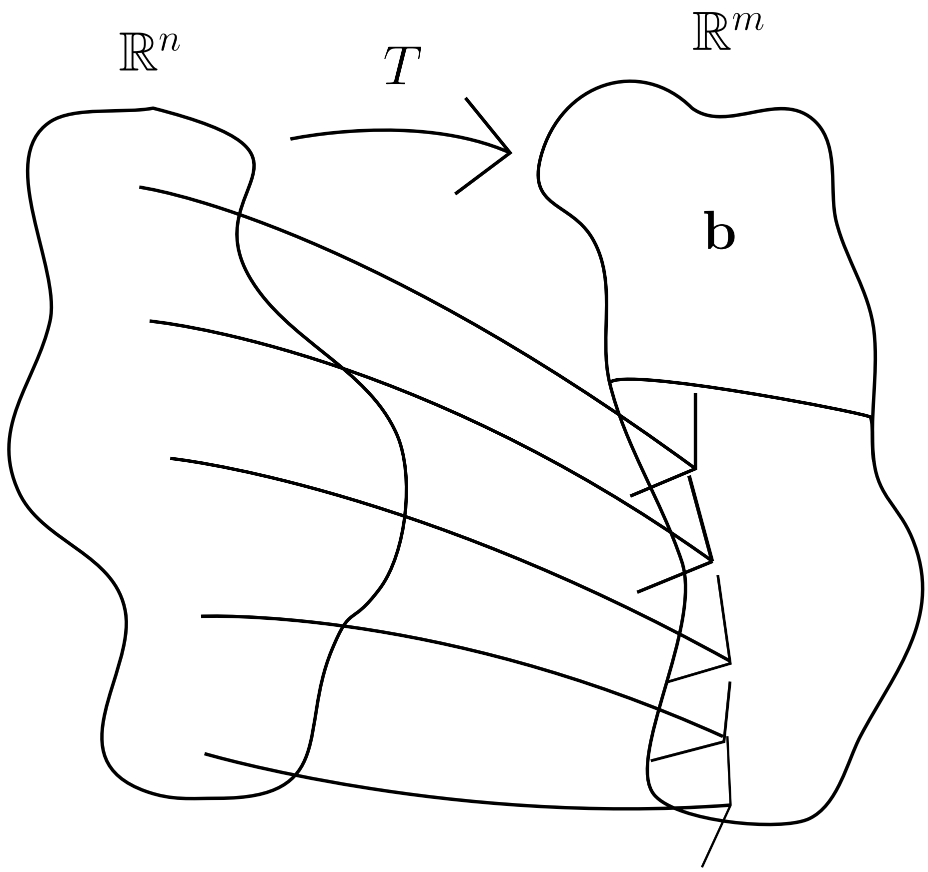


Figure 6: A Non-Onto Function T

$\mathbf{A}\mathbf{x} = T(\mathbf{x})$?" We could also further ask if $\mathbf{A}$ has a pivot in every row. With one-to-one-ness we can ask "if $T(\mathbf{x}) = \mathbf{b}$ and $T(\mathbf{y}) = \mathbf{b}$ does this imply that $\mathbf{x} = \mathbf{y}$?" We could also ask whether $\mathbf{A}$ has a pivot in every column.

Now, let's prove the theorem presented during the last lecture.

Theorem 16. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear. T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$.

As this theorem involves the bi-conditional logical connective, the proof follows in two parts.

1. We begin by proving the first direction: If T is one-to-one, then $T(\mathbf{0}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$

   Proof. Suppose T is one-to-one. Suppose also $T(\mathbf{x}) = \mathbf{0}$. Since T is linear, we know that $T(\mathbf{0}) = \mathbf{0}$. Since T is one-to-one, then we must have $\mathbf{x} = \mathbf{0}$. This completes the proof. $\square$

2. We finish by proving the other direction: If $\forall \mathbf{x} \in \mathbb{R}^n, T(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$, then T is one-to-one.

   Proof. Suppose $T(\mathbf{x}) = T(\mathbf{y})$. Then, $T(\mathbf{x}) - T(\mathbf{y}) = \mathbf{0}$. So $T(\mathbf{x} - \mathbf{y}) = \mathbf{0}$, by the linearity of T. This then gives $\mathbf{x} - \mathbf{y} = \mathbf{0}$ by hypothesis. Therefore, $\mathbf{x} = \mathbf{y}$, which means that T is one-to-one, by definition. This completes the proof. $\square$

Now, let's consider some in-class problems.

Example. Are the following linear transformations one-to-one? Onto?

1. $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} 3x_2 \\ x_2 \end{bmatrix}$$

   is neither one-to-one nor onto. It is not one-to-one because $T(\begin{bmatrix} 1 \\ 0 \end{bmatrix}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ —this follows from Theorem 16. It is not onto because $T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ does not have a solution.

2. $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} x_2 \\ 0 \\ x_1 \end{bmatrix}$$

   is one-to-one but not onto. It is onto because the only way to get the zero vector as an output is where $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{0}$. It is not onto because $T(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ does not have a solution.

3. $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$T(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} x_1 + x_3 \\ x_2 \end{bmatrix}$$

is onto but not one-to-one. It is not one-to-one because $T(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}) = \mathbf{0}$. It is onto because inputs of form $\begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$ can form every possible output $\begin{bmatrix} a \\ b \end{bmatrix}$ in the co-domain.

4. $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\mathbf{x}) =$ "reflection of $\mathbf{x}$ across the x-axis" is both onto and one-to-one because indeed $T(\mathbf{0}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$ and because the solution to every vector in the co-domain can be identified as itself, with the y-component multiplied by -1.

5. $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\mathbf{x}) = \mathbf{0}$ is neither onto nor one-to-one because all inputs yield $\mathbf{0}$ and thus $T(\mathbf{0}) = \mathbf{0} \not\Rightarrow \mathbf{x} = \mathbf{0}$ as well as because any non-origin output will lack a solution.

6. $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\mathbf{x}) = \mathbf{x}$ is both onto and one-to-one because indeed $T(\mathbf{0}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$ and because the solution to every vector in the co-domain can be identified as itself.

◇

Now, some questions:

Example. What can we say about a linear transformation T such that

1. $T : \mathbb{R}^5 \to \mathbb{R}^{10}$?

   Here, T is not onto because the transformation can be represented $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is a 10×5 matrix, and this matrix necessarily cannot have a pivot in every row.

2. $T : \mathbb{R}^{10} \to \mathbb{R}^5$?

   Here, T is not one-to-one because the transformation can be represented $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is a 5×10 matrix, and this matrix necessarily cannot have a pivot in every column.

3. $T : \mathbb{R}^{10} \to \mathbb{R}^{10}$?

   Here, T is one-to-one if and only if T is onto because the transformation can be represented $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is a 10×10 matrix, and this matrix necessarily cannot have a pivot in every column without having a pivot in every row.

◇

Lecture 14: A Review of Proof Techniques & Linear Transformation Specifics

September 23, 2021

Today, we'll address some student questions and work through a worksheet.

We've demonstrated in class the following theorem.

Theorem 17. If $T : \mathbb{R}^m \to \mathbb{R}^n$ is linear, then

1. $T(\mathbf{0}) = \mathbf{0}$
2. $T(c_1\mathbf{v_1} + \ldots + c_p\mathbf{v_p}) = c_1T(\mathbf{v_1}) + \ldots + c_pT(\mathbf{v_p})$

Within our homework proofs, we can reference any theorem proved in class without proving them ourselves.

With this disclaimer out of the way, let's consider a few of the notions that we have developed in class.

- A set of vectors $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ may span $\mathbb{R}^n$.
- A set of vectors $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ is linearly independent or dependent.
- A function $T : \mathbb{R}^n \to \mathbb{R}^m$ may be onto.
- A function $T : \mathbb{R}^n \to \mathbb{R}^m$ may be one-to-one.

Each of these ends demands different proof techniques. Respectively, these goals are set forth for the former three cases as follows.

- Let $\mathbf{w} \in \mathbb{R}^n$. Show that $\exists c_1, c_2, \ldots, c_3 : \mathbf{w} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}$.
- Suppose $c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}$. Show that $c_1 = c_2 = \ldots = c_p = 0$ is the unique solution.
- Let $\mathbf{w} \in \mathbb{R}^m$. Show that $T(\mathbf{v}) = \mathbf{w}$ where $\mathbf{v} \in \mathbb{R}^n$.

Now, let's work through a proof of the second case.

Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation. Prove that if $\{T(\mathbf{v_1}), T(\mathbf{v_2}), T(\mathbf{v_3})\}$ is linearly independent then $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly independent.

Proof. Suppose that $c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3} = \mathbf{0}$. Applying T to both sides gives $T(c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3}) = T(\mathbf{0})$. By the linearity of T, this simplifies to $c_1T(\mathbf{v_1}) + c_2T(\mathbf{v_2}) + c_3T(\mathbf{v_3}) = \mathbf{0}$. Since $\{T(\mathbf{v_1}), T(\mathbf{v_2}), T(\mathbf{v_3})\}$ is independent, $c_1 = c_2 = c_3 = 0$ by hypothesis. Therefore, $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is independent. This completes the proof. $\square$

Now demonstrated is an example of an incorrect proof that may be formed by approaching with the opposite supposition.

Proof. Suppose that $c_1T(\mathbf{v_1}) + c_2T(\mathbf{v_2}) + c_3T(\mathbf{v_3}) = \mathbf{0}$. Then $c_1 = c_2 = c_3 = 0$ since $\{T(\mathbf{v_1}), T(\mathbf{v_2}), T(\mathbf{v_3})\}$, by hypothesis. Then, by the linearity of T, $T(c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3}) = \mathbf{0}$. So, $c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3} = \mathbf{0}$. Therefore, since $c_1 = c_2 = c_3 = 0$, $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is independent. □

This proof is flawed because it assumes that T is one-to-one. In order to move from $T(c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3}) = \mathbf{0}$ to $c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3} = \mathbf{0}$, then $T(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$. This is not guaranteed for all T, and thus the proof is logically fallacious.

Now, let's work through another proof.

Given a linear function $T : \mathbb{R}^3 \to \mathbb{R}^3$, prove that if $\{T(\mathbf{v_1}), T(\mathbf{v_2}), T(\mathbf{v_3})\}$ spans $\mathbb{R}^3$, then T is onto.

Proof. Let $\mathbf{w} \in \mathbb{R}^3$. By hypothesis, since $\{T(\mathbf{v_1}), T(\mathbf{v_2}), T(\mathbf{v_3})\}$ spans $\mathbb{R}^3$, we know that $\mathbf{w} = c_1T(\mathbf{v_1}) + c_2T(\mathbf{v_2}) + c_3T(\mathbf{v_3})$. Since T is linear, then $\mathbf{w} = T(c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3})$. Let $\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + c_3\mathbf{v_3}$. Then $T(\mathbf{v}) = \mathbf{w}$. Therefore, T is onto. This completes the proof. □

Now, pivoting a bit, let's determine a standard matrix for rotations of θ degrees. To this end, we simply need to determine how such a rotation effects the standard basis vectors in $\mathbb{R}^2$,

$$\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$\mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Figure 7 quite easily gives the standard rotation matrix that we desire,

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

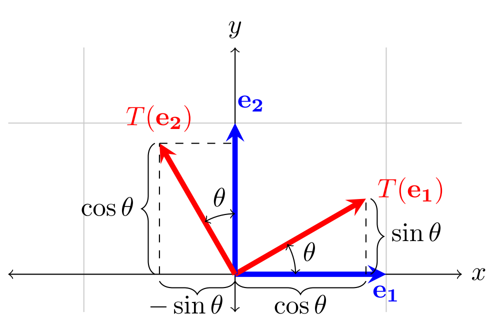


Figure 7: Rotation of the Standard Basis Vectors in $\mathbb{R}^2$ by θ Degrees

Now, let's brainstorm some examples of certain functions.

Example.

1. Two examples of non-linear functions $T : \mathbb{R}^2 \to \mathbb{R}^3$ are

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} (x_1)^2 \\ 0 \\ 0 \end{bmatrix}$$

   and

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} x_1 \\ 1 \\ x_2 \end{bmatrix}.$$

2. A linear function $T : \mathbb{R}^2 \to \mathbb{R}^3$ that is one-to-one but not onto may be

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}.$$

3. A linear function $T : \mathbb{R}^3 \to \mathbb{R}^2$ that is onto but not one-to-one may be

$$T(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} x_1 \\ x_3 \end{bmatrix}.$$

$\diamond$

We'll wrap up with some clarification about the fifth homework problem.

The proof sought—given an onto linear function $T : \mathbb{R}^3 \to \mathbb{R}^2$, prove that if $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ spans $\mathbb{R}^3$ then $\{T(\mathbf{v_1}), T(\mathbf{v_2}), T(\mathbf{v_3})\}$ spans $\mathbb{R}^2$—should follow the following form.

Proof. Let $\mathbf{w} \in \mathbb{R}^2$. Since T is onto, by hypothesis, then $\exists \mathbf{v} \in \mathbb{R}^3$ such that $T(\mathbf{v}) = \mathbf{w}$.

$$\vdots$$

Therefore, $\mathbf{w} = c_1 T(\mathbf{v_1}) + c_2 T(\mathbf{v_2}) + c_3 T(\mathbf{v_3})$. This completes the proof. $\square$

Lecture 15: Some Applications of Linear Systems

September 24, 2021

Let's consider the applied problem of the Global Positioning System, or GPS.

Our assumptions are as follows:

- Distances are measured in units of Earth's radius.

- Time is measured in centiseconds.
- The time for a GPS signal to travel to earth from space is 0.47 Earth radii per centisecond.
- We use an xyz-coordinate system where the z-axis intersects the poles.
- Points (x, y, z) on earth's surface satisfy $x^2 + y^2 + z^2 = 1$.

Imagine that a hiker is lost in the woods at position (x, y, z) at time t when their GPS receives the following signals in Table 1.

Table 1: GPS Data

Satellite	Position $[(x_s, y_s, z_s)]$	Time $[t_s]$
1	$(1.11, 2.55, 2.14)$	1.29
2	$(2.87, 0, 1.43)$	1.31
3	$(0, 0.08, 2.29)$	2.75
4	$(1.54, 1.01, 1.23)$	1.29

For each of these satellites, we can quite easily find the distance using the given speed and time sent

$$d = 0.47(t - t_s).$$

We can also calculate the distance using the distance formula as

$$d = \sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2}.$$

Setting these two distances equal yields the non-linear equation in (x, y, z, t)

$$(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2 = 0.47^2(t - t_s)^2.$$

Plugging in the given values and simplifying yields the system

$$\begin{cases} 2.22x + 5.10y + 4.28z - 0.57t & = x^2 + y^2 + z^2 - 0.22t^2 + 11.95 \\ 5.74x + 2.86z - 0.58t & = x^2 + y^2 + z^2 - 0.22t^2 + 9.90 \\ 2.16y + 4.58z - 1.12t & = x^2 + y^2 + z^2 - 0.22t^2 + 4.74 \\ 3.08x + 2.02y + 2.46z - 1.79t & = x^2 + y^2 + z^2 - 0.22t^2 + 1.26 \end{cases}.$$

Subtracting the first equation from the latter three then yields the linear system

$$\begin{cases} 3.52x - 5.20y - 1.42z - 0.01t & = -2.05 \\ -2.22x - 2.94y - 0.30z - 0.64t & = -7.21 \\ 0.86x - 3.08y - 1.82z - 1.22t & = -10.69 \end{cases},$$

whose augmented matrix then reduces to

$$\begin{bmatrix} 1 & 0 & 0 & 0.36 & 2.97 \\ 0 & 1 & 0 & 0.03 & 0.81 \\ 0 & 0 & 1 & 0.79 & 5.91 \end{bmatrix},$$

yielding the following solution, with t as a free variable,

$$\begin{aligned} x &= 297 - 0.36t \\ y &= 0.81 - 0.03t \\ z &= 5.91 - 0.79t. \end{aligned}$$

Plugging this solution into the initial non-linear equation yields the quadratic

$$0.54t^2 - 6.65t + 20.32 = 0,$$

with solutions

$$t = 5.60, 6.74.$$

These two solutions give coordinates $(0.96, 0.65, 1.46)$ and $(0.55, 0.61, 0.56)$ respectively. As only the second set of coordinates lies on the unit sphere (i.e., $0.55^2 + 0.61^2 + 0.56^2 = 0.9998$), then this is the feasible solution and these are the hiker's coordinates.

Now let's introduce the concept of discrete dynamical systems (DDS).

Consider two geographical centres, a city and a suburb. Every year, a certain percentage of people will stay in the city, leave the city, stay in the suburbs, or leave the suburbs. Take these values to be $0.95, 0.05, 0.97$, and 0.03 respectively. Let c_0 be the initial population of the city, and s_0 be the initial population of the suburbs. We can give equations for c_1 and s_1 as

$$\begin{aligned} c_1 &= 0.95c_0 + 0.3s_0 \\ s_1 &= 0.05c_0 + 0.97s_0. \end{aligned}$$

Let

$$\mathbf{x_i} = \begin{bmatrix} c_i \\ s_i \end{bmatrix}.$$

Then, we can describe the aforementioned system as

$$\mathbf{x_i} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} c_{i-1} \\ s_{i-1} \end{bmatrix}.$$

This matrix is often referred to as the "migration matrix." Further, a matrix whose columns each sum to 1 can be referred to as the "probability" or "stochastic" matrix.

Now, with this iterative description in-hand, we can calculate some concrete values.

Example. If we wish to determine the population values after one year, then we can simply calculate

$$\mathbf{x_1} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} c_0 \\ s_0 \end{bmatrix}.$$

And further after two years we can find

$$\mathbf{x_2} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} (\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} c_0 \\ s_0 \end{bmatrix}).$$

Finally, in the general case we can determine that

$$\mathbf{x_n} = \underbrace{\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} (\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \dots (\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}}_{n} \begin{bmatrix} c_0 \\ s_0 \end{bmatrix}) \dots).$$

◇

In we calculate a series of values for this system in Mathematica, we can find that this system settles into a stable equilibrium, but, depending on the values in the migration matrix, a system of this form could create oscillatory or chaotic behaviour.

Finally, let's introduce a notion of matrix multiplication.

Let

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

and

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Trivially,

$$\mathbf{Ax} = x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix}.$$

Now, consider the case

$$\begin{aligned} \mathbf{A}^2\mathbf{x} &= \mathbf{A}(\mathbf{Ax}) \\ &= \mathbf{A}(x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix}). \end{aligned}$$

By the linearity of matrix multiplication, we can then state that

$$\begin{aligned}\mathbf{A}(x\begin{bmatrix}a\\c\end{bmatrix}+y\begin{bmatrix}b\\d\end{bmatrix}) &= x\mathbf{A}\begin{bmatrix}a\\c\end{bmatrix}+y\mathbf{A}\begin{bmatrix}b\\d\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{A}\begin{bmatrix}a\\c\end{bmatrix} & \mathbf{A}\begin{bmatrix}b\\d\end{bmatrix}\end{bmatrix}.\end{aligned}$$

This thus can define the multiplication of two identical matrices $\mathbf{A}^2$. Next lecture, we'll examine this notion using two different matrices rather than one.

Lecture 16: Matrix Multiplication & Other Matrix Operations

September 27, 2021

We consider matrix multiplication as an operation that satisfies the relationship

$$\mathbf{A}(\mathbf{B}\mathbf{x}) = \mathbf{A}\mathbf{B}(\mathbf{x}).$$

Now, to review. Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, this transformation can be represented

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x},$$

where $\mathbf{A}$ is an $m \times n$ matrix.

Consider a linear transformation $S : \mathbb{R}^p \to \mathbb{R}^n$ such that

$$S(\mathbf{x}) = \mathbf{B}\mathbf{x}.$$

where $\mathbf{B}$ is an $n \times p$ matrix.

The composition of these two transformations yields a function

$$\begin{aligned}&\mathbb{R}^p \xrightarrow{S} \mathbb{R}^n \xrightarrow{T} \mathbb{R}^m\\ &\mathbf{x} \mapsto \ \mathbf{B}\mathbf{x} \mapsto \mathbf{A}(\mathbf{B}\mathbf{x})\\ &\mathbf{x} \quad \underset{\mathbf{AB}}{\longmapsto} \ \mathbf{A}(\mathbf{B}\mathbf{x}).\end{aligned}$$

Here, we would like $\mathbf{AB}$ to be an $m \times p$ matrix such that $\forall \mathbf{x} \in \mathbb{R}^p, (\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{B}\mathbf{x})$.

Consider representing $\mathbf{B}$ as

$$\mathbf{B} = \begin{bmatrix}\mathbf{b_1} & \dots & \mathbf{b_p}\end{bmatrix}.$$

Then, define $\mathbf{AB}$ as

$$\mathbf{AB} = \begin{bmatrix}\mathbf{Ab_1} & \dots & \mathbf{Ab_p}\end{bmatrix}.$$

By the linearity of matrix multiplication, it thus follows that

$$(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx}).$$

To be clear, this operation is only defined where the number of columns in $\mathbf{A}$ is equivalent to the number of rows in $\mathbf{B}$.

Now, let's work through an example.

Example. Let

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}. \end{aligned}$$

We can thus find the left-multiplication of these matrices as

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} \mathbf{A} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} & \mathbf{A} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 6 & 3 \\ 8 & 4 \end{bmatrix}. \end{aligned}$$

In this case, the right-multiplication of these matrices also happens to be defined, because of their size. This gives

$$\begin{aligned} \mathbf{BA} &= \begin{bmatrix} \mathbf{B} \begin{bmatrix} 1 \\ 2 \end{bmatrix} & \mathbf{B} \begin{bmatrix} 3 \\ 4 \end{bmatrix} & \mathbf{B} \begin{bmatrix} 5 \\ 6 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ 1 & 3 & 5 \end{bmatrix}. \end{aligned}$$

As is made clear by this example, the multiplication of matrices is not necessarily commutative—that is, $\mathbf{AB} \neq \mathbf{BA}$ in all cases. $\diamond$

Now, let's work through some in-class problems.

Example.

1. Given the matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 3 & 0 \end{bmatrix}$$
$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 2 \\ 1 & -1 & 0 \end{bmatrix}$$
$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

we can find the product of all pairs for which matrix multiplication is defined as follows.

Because $\mathbf{A}$ is a 3×2 matrix and $\mathbf{B}$ is a 2×3 matrix, then we can find

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

Because $\mathbf{B}$ is a 2×3 matrix and $\mathbf{A}$ is a 3×2 matrix, then we can find

$$\mathbf{BA} = \begin{bmatrix} 7 & 0 \\ 1 & 3 \end{bmatrix}.$$

Because $\mathbf{C}$ is a 3×3 matrix and $\mathbf{A}$ is a 3×2 matrix, then we can find

$$\mathbf{CA} = \begin{bmatrix} 1 & 1 \\ 4 & 3 \\ 3 & 0 \end{bmatrix}.$$

Finally, because $\mathbf{B}$ is a 2×3 matrix and $\mathbf{C}$ is a 3×3 matrix, then we can find

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

2. Given the matrices

$$A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$$
$$B = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix},$$

we can show that these are inverses of each other by

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In this case, this gives

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

Therefore, these matrices are inverses.

3. Given the matrix

$$\mathbf{A} = \begin{bmatrix}1 & 2\\ 1 & 2\end{bmatrix},$$

a non-zero matrix $\mathbf{C}$ such that $\mathbf{AC} = \mathbf{0}$ is

$$\mathbf{C} = \begin{bmatrix}1 & -1\\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}.$$

This matrix also happens to give $\mathbf{CA} = \mathbf{0}$. Therefore, given a matrix equation $\mathbf{AB} = \mathbf{0}$, we cannot conclude that either $\mathbf{A}$ or $\mathbf{B}$ are the zero matrix.

4. (a) The standard matrix $\mathbf{A}$ for the linear transformation of $\mathbb{R}^2$ that reflects vectors across the x-axis is

$$\mathbf{A} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}.$$

(b) The standard matrix $\mathbf{B}$ for the linear transformation of $\mathbb{R}^2$ that rotates vectors $\frac{\pi}{2}$ radians anti-clockwise about the origin is

$$\mathbf{B} = \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}.$$

(c) The left and right-multiplication of these matrices is computed as

$$\begin{aligned}\mathbf{AB} &= \begin{bmatrix}0 & -1\\ -1 & 0\end{bmatrix}\\ \mathbf{BA} &= \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.\end{aligned}$$

Obviously, $\mathbf{AB} \neq \mathbf{BA}$.

(d) These multiplied matrices represent the respective compositions of the two given transformations, now given as a reflection of vectors across the lines $y = -x$ and $y = x$ respectively.

◇

Notation. Given an $m \times n$ matrix $\mathbf{A}$, we can refer to a specific entry in $\mathbf{A}$ by the notation a_{ij} where i is the row of the element and j is the column.

For example, where

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

then

$$\begin{aligned} a_{32} &= 2 \\ a_{12} &= -7. \end{aligned}$$

Now, let's consider an algorithm for calculating the matrix product $\mathbf{AB}$ in terms of a_{ij} and b_{ij}, the elements of $\mathbf{A}$ and $\mathbf{B}$.

Let

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} a_{ij} \end{bmatrix} : \mathbf{A} \text{ is } m \times n \\ \mathbf{B} &= \begin{bmatrix} b_{ij} \end{bmatrix} : \mathbf{B} \text{ is } n \times p. \end{aligned}$$

A given element of their left-multiplication can thus be determined

$$\begin{aligned} (\mathbf{AB})_{ij} &= \begin{bmatrix} a_{i1} & \dots & a_{in} \end{bmatrix} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{bmatrix} \\ &= a_{i1}b_{1j} + \ldots + a_{in} + b_{nj} \\ &= \sum_{k=1}^{n} a_{ik}b_{kj}. \end{aligned}$$

Now, let's consider two more operations on matrices.

Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ matrices such that

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} a_{ij} \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} b_{ij} \end{bmatrix}. \end{aligned}$$

The addition of these matrices is defined

$$(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$$

and the multiplication of these matrices is defined

$$(c\mathbf{A})_{ij} = \begin{bmatrix} ca_{ij} \end{bmatrix} : c \in \mathbb{R}.$$

Theorem 18. The properties of matrix addition and scalar multiplication are as follows

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
3. $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$
4. $(c + d)\mathbf{A} = c\mathbf{A} + d\mathbf{A}$
5. $\mathbf{A} + \mathbf{0} = \mathbf{A}$
6. $(cd)\mathbf{A} = c(d\mathbf{A})$

where $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ matrices, $\mathbf{0}$ is the $m \times n$ matrix whose entries are all zero, and $c, d \in \mathbb{R}$.

Theorem 19. If $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are matrices for which the product $\mathbf{ABC}$ is defined, then the associative law $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ holds.

Lecture 17: Introducing Matrix Inverses

September 29, 2021

Last time, we defined matrix multiplication such that given an $m \times n$ matrix $\mathbf{A}$ and an $n \times p$ matrix $\mathbf{B}$ then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{Ab_1} & \dots & \mathbf{Ab_p} \end{bmatrix}.$$

This definition satisfies

$$\forall \mathbf{x} \in \mathbb{R}^p,\ \mathbf{A}(\mathbf{Bx}) = (\mathbf{AB})\mathbf{x}.$$

Today, we'll consider matrix inverses. Note that a matrix can only have an inverse if it is a square matrix.

Definition 16. An $n \times n$ matrix $\mathbf{A}$ is invertible with inverse $\mathbf{B}$ if

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I_n}.$$

From this definition, we gain the properties

$$\mathbf{ABx} = \mathbf{x}$$
$$\mathbf{BAx} = \mathbf{x}.$$

Let's consider an example.

Example. Given the matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$
$$\mathbf{B} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix},$$

is **B** the inverse of **A**? Because

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

B is certainly not the inverse of **A**. ◇

Theorem 20. Given a matrix

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

A is invertible if and only if

$$ad - bc \neq 0.$$

This value $ad - bc$ is known as the **determinant** of **A**. Further, if **A** is invertible, then

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

Let's continue our example.

Example. The matrix

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

has a determinant of -1, so it is invertible. Using the formula developed in Theorem 20, then

$$\begin{aligned} \mathbf{A}^{-1} &= - \begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}. \end{aligned}$$

This matrix is indeed the inverse of **A**, because

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

◇

Now, let's prove the theorem presented during the last lecture.

Theorem 21. Given matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ for which the product $\mathbf{ABC}$ is defined,

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}).$$

Proof. Suppose

$$\mathbf{C} = \begin{bmatrix} \mathbf{c_1} & \dots & \mathbf{c_p} \end{bmatrix}.$$

Then

$$(\mathbf{AB})\mathbf{C} = \begin{bmatrix} (\mathbf{AB})\mathbf{c_1} & \dots & (\mathbf{AB})\mathbf{c_p} \end{bmatrix}.$$

By the definition of matrix multiplication, this gives

$$\begin{aligned} (\mathbf{AB})\mathbf{C} &= \begin{bmatrix} \mathbf{A}(\mathbf{Bc_1}) & \dots & \mathbf{A}(\mathbf{Bc_p}) \end{bmatrix} \\ &= \mathbf{A} \begin{bmatrix} \mathbf{Bc_1} & \dots & \mathbf{Bc_p} \end{bmatrix} \\ &= \mathbf{A}(\mathbf{BC}). \end{aligned}$$

This completes the proof. $\square$

Example. This theorem can be used to show that

$$\mathbf{A}(((\mathbf{BC})\mathbf{D})\mathbf{E}) = (\mathbf{ABCD})\mathbf{E} = \mathbf{A}(\mathbf{BCDE}).$$

$\diamond$

Now, let's consider some further properties of matrix algebra.

Theorem 22. Given $m \times n$ matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, then

$$\begin{gathered} \mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC} \\ (\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA} \\ (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B}) = c(\mathbf{AB}) \\ \mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}. \end{gathered}$$

Now, to work through some in-class problems.

Example.

1. Given the matrix

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

such that $ad - bc \neq 0$, the matrix

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

is the inverse of $\mathbf{A}$ because

$$\begin{aligned}
\mathbf{AB} &= \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\mathbf{BA} &= \frac{1}{ad - bc}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\end{aligned}$$

2. *Proof.* Suppose that $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ invertible matrices. We can consider the products

$$\begin{aligned}
(\mathbf{AB})\mathbf{B}^{-1}\mathbf{A}^{-1} &= \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} \\
&= \mathbf{AIA}^{-1} \\
&= \mathbf{AA}^{-1} \\
&= \mathbf{I}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{B}^{-1}\mathbf{A}^{-1}(\mathbf{AB}) &= \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} \\
&= \mathbf{B}^{-1}\mathbf{IB} \\
&= \mathbf{B}^{-1}\mathbf{B} \\
&= \mathbf{I}.
\end{aligned}$$

This thus demonstrates both that

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

and, by the definition of matrix inverses, that $(\mathbf{AB})$ is invertible. □

3. Given the invertible $n \times n$ matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, the matrix $\mathbf{ABC}$ is invertible with the matrix $\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$ because

$$\begin{aligned}
(\mathbf{ABC})(\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}) &= \mathbf{AB}(\mathbf{CC}^{-1})\mathbf{B}^{-1}\mathbf{A}^{-1} \\
&= \mathbf{ABIB}^{-1}\mathbf{C}^{-1} \\
&= \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} \\
&= \mathbf{AIA}^{-1} \\
&= \mathbf{AA}^{-1} \\
&= \mathbf{I}
\end{aligned}$$

and

$$\begin{aligned}(\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{ABC}) &= \mathbf{C}^{-1}\mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{BC} \\ &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{IBC} \\ &= \mathbf{C}^{-1}(\mathbf{B}^{-1}\mathbf{B})\mathbf{C} \\ &= \mathbf{C}^{-1}\mathbf{IC} \\ &= \mathbf{C}^{-1}\mathbf{C} \\ &= \mathbf{I}.\end{aligned}$$

$\diamond$

Lecture 18: Midterm Review Questions

September 30, 2021

During today's discussion session, we will work through a set of true/false questions that review material covered on the midterm exam, which is open during this period. These questions, along with their solutions, will be available on the Moodle page separate from these notes.

Lecture 19: More On Invertiblilty

October 1, 2021

Last time, we defined the notion of an inverse $\mathbf{B}$ for an $n \times n$ matrix $\mathbf{A}$ satisfying

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n.$$

We also examined a theorem regarging invertibility for 2×2 matrices which stated that a matrix

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

is invertible if and only if this matrix's determinant is not equal to zero, or

$$ad - bc \neq 0.$$

Finally, we considered the so-called "shoes and socks theorem" which stated that given two $n \times n$ invertible matrices $\mathbf{A}, \mathbf{B}$, then $\mathbf{AB}$ is invertible and its inverse is

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

This theorem extends to more $n \times n$ invertible matrices $\{\mathbf{M}_1, \ldots, \mathbf{M}_p\}$ as

$$(\mathbf{M}_1\mathbf{M}_2 \ldots \mathbf{M}_p)^{-1} = \mathbf{M}_p^{-1} \ldots \mathbf{M}_2^{-1}\mathbf{M}_1^{-1}.$$

Now, a couple of questions. Consider $\mathbf{A}, \mathbf{B}$ to be $n \times n$ invertible matrices.

- Is $\mathbf{A} + \mathbf{B}$ invertible?
- If $c \in \mathbb{R}$, is $c\mathbf{A}$ invertible?

To the first question, consider the case

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

Here, $\mathbf{A}$ and $\mathbf{B}$ are invertible yet their sum is the zero matrix, which is trivially singular. This disproves the first guarantee. As for the second question, where $c \neq 0$, then

$$\begin{aligned}(c\mathbf{A})(\frac{1}{c}\mathbf{A}^{-1}) &= (c\frac{1}{c})\mathbf{A}\mathbf{A}^{-1} \\ &= \mathbf{I}.\end{aligned}$$

Therefore, this statement holds as long as c is non-zero. Now, a final question.

- If $\mathbf{A}$ is invertible, is $\mathbf{A}^{-1}$ invertible?

The answer to this question is, of course, yes. This inverse is defined by the aforementioned definition of matrix inverses as

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}.$$

In summary, the inverse of an invertible matrix is always invertible.

Today, we'll consider an algorithm for finding $\mathbf{A}^{-1}$ where $\mathbf{A}$ is invertible.

Notation. An invertible matrix $\mathbf{A}$ is also called non-singular and a non-invertible matrix $\mathbf{B}$ is also called singular.

Theorem 23. An $n \times n$ matrix $\mathbf{A}$ is invertible if and only if $\mathbf{A}$ is row-equivalent to $\mathbf{I}_n$.

From this theorem, we can determine that we can find $\mathbf{A}^{-1}$ if $\mathbf{A}$ is row-equivalent to $\mathbf{I}_n$ by performing the augmented matrix reduction

$$\begin{bmatrix} \mathbf{A} & \mathbf{I}_n \end{bmatrix} \sim \begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1} \end{bmatrix}.$$

We'll now complete some in-class problems whose answers will be available on Moodle.

Now, to introduce elementary matrices. Elementary matrices model the elementary row operations used to reduce augmented matrices. There are three of these operations.

- Replacement

- Swap
- Scale

To demonstrate, the expression **EA** represents one of these operations being performed on **A**. The given worksheet contains a few examples of these matrices. For example, the equation

$$\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ -3a+d & -3b+e & -3c+f \\ g & h & i \end{bmatrix}$$

represents the performance of the elementary row operation $-3R_1 + R_2$ on R_2.

As it turns out, we can find the elementary matrix for a given row operation by simply performing that operation on the identity matrix. Using these elementary matrices, we can model the reduction of a matrix **A** into RREF using matrix multiplication. For example,

$$\mathrm{RREF}(\mathbf{A}) = \mathbf{E}_p \mathbf{E}_{p-1} \dots \mathbf{E}_2 \mathbf{E}_1 \mathbf{A}.$$

We'll pick up with a proof of the aforementioned algorithm that uses these elementary matrices during the next lecture.

Lecture 20: The Invertible Matrix Theorem

October 4, 2021

Last time, we introduced elementary matrices, which are matrices that represent elementary row operations. In today's pre-lecture video, we proved that an $n \times n$ matrix **A** is invertible if and only if **A** is row-equivalent to the identity matrix $\mathbf{I}_n$. This proof further showed that the same elemenary matrices that reduce **A** to the identity matrix also transform the identity matrix into $\mathbf{A}^{-1}$, giving us the algorithm that we examined during the last class.

After completing some in-class problems whose solutions can be found on the Moodle page, let's consider the Invertible Matrix Theorem.

Theorem 24 (The Invertible Matrix Theorem)**.** Suppose $\mathbf{A}$ is an $n \times n$ matrix. Then the following statements are logically equivalent.

a. $\mathbf{A}$ is invertible.

b. $\mathbf{A}$ is row-equivalent to the identity matrix $\mathbf{I}_n$.

c. $\mathbf{A}$ has n pivots.

d. The equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

e. The columns of $\mathbf{A}$ form a linearly independent set.

f. The linear transformation $T(\mathbf{x}) = \mathbf{Ax}$ is one-to-one.

g. The equation $\mathbf{Ax} = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in \mathbb{R}^n$.

h. The columns of $\mathbf{A}$ span $\mathbb{R}^n$.

i. The linear transformation $T(\mathbf{x}) = \mathbf{Ax}$ is onto.

j. There is an $n \times n$ matrix $\mathbf{C}$ such that $\mathbf{CA} = \mathbf{I}_n$.

k. There is an $n \times n$ matrix $\mathbf{D}$ such that $\mathbf{AD} = \mathbf{I}_n$.

A proof of this theorem involves a new technique. Rather than proving that each pair of these statements is logically equivalent with an "if and only if" proof, we'll prove two cycles of implications and cite a few established theorems to establish

$$\begin{aligned}
&a \Rightarrow j \Rightarrow d \Rightarrow c \Rightarrow b \Rightarrow a \\
&a \Rightarrow k \Rightarrow g \Rightarrow a \\
&d \Leftrightarrow e \Leftrightarrow f \\
&g \Leftrightarrow h \Leftrightarrow i.
\end{aligned}$$

With these relationships proved, we can find any bi-conditional relationship in the cycle to be proved as well. For example, the first condition of $a \Leftrightarrow c$ is obtained simply by the first loop of the former cycle, and the second is given by starting at c and proceeding around the cycle to a.

Now, we'll give a proof of the theorem.

Proof. First, we will prove the first circle of implications.

- $a \Rightarrow j$: Suppose $\mathbf{A}$ is invertible. Then, by the definition of invertibility, there exists a matrix $\mathbf{C}$ such that $\mathbf{CA} = \mathbf{I}$ and $\mathbf{AC} = \mathbf{I}$. This completes the proof.

- $j \Rightarrow d$: Suppose that there exists a matrix $\mathbf{C}$ such that $\mathbf{CA} = \mathbf{I}$. Then, the equation $\mathbf{Ax} = \mathbf{0}$ can be manipulated as

$$\begin{aligned} \mathbf{Ax} &= \mathbf{0} \\ \mathbf{CAx} &= \mathbf{C0} \\ \mathbf{x} &= \mathbf{0}, \end{aligned}$$

showing that this equation has only the trivial solution. This completes the proof.

- $d \Rightarrow c$: Suppose that the equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution. Then, by the uniqueness of this solution, $\mathbf{A}$ must have a pivot in every column. $\mathbf{A}$ has n columns, so it has n pivots. This completes the proof.

- $c \Rightarrow b$: Suppose that $\mathbf{A}$ has n pivots. Then, $\mathbf{A}$ must be able to be reduced into the identity matrix. Therefore, $\mathbf{A}$ is row-equivalent to the identity matrix.

- $b \Rightarrow a$: By the theorem proved during this lecture's pre-lecture video, where $\mathbf{A}$ is row-equivalent to the identity matrix, it is invertible.

Now, we'll prove the second circle.

- $a \Rightarrow k$: Suppose $\mathbf{A}$ is invertible. Then, by the definition of invertibility, there exists a matrix $\mathbf{C}$ such that $\mathbf{CA} = \mathbf{I}$ and $\mathbf{AC} = \mathbf{I}$. This completes the proof.

- $k \Rightarrow g$: Suppose that there exists a matrix $\mathbf{D}$ such that $\mathbf{AD} = \mathbf{I}$. Then, we can take $\mathbf{x} = \mathbf{Db}$ as a trivial solution to the equation $\mathbf{Ax} = \mathbf{b}$ for all $\mathbf{b} \in \mathbb{R}^n$. This completes the proof.

- $g \Rightarrow a$: Suppose that the equation $\mathbf{Ax} = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in \mathbb{R}^n$. Then, $\mathbf{A}$ has a pivot in every row and, by the fact that it's a square matrix, a pivot in every column. By part c in the first circle, this implies a. This completes the proof.

Finally, we can use theorems previously proven in class to show that $d \Leftrightarrow e \Leftrightarrow f$ and $g \Leftrightarrow h \Leftrightarrow i$. This completes the entire proof. $\square$

We should note that this is not the ultimate version of this theorem; as we progress through the rest of the course, we will prove that more statements are logically equivalent to those presented here. For example, we could use the fact that

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$$

to show that $\mathbf{A}$ is invertible if and only if $\mathbf{A}^T$ is invertible. We can further show that

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T.$$

Lecture 21: Subspaces of $\mathbb{R}^n$

October 6, 2021

Last time, we proved the Invertible Matrix Theorem. Contained within this theorem was a surprise: If $\mathbf{A}, \mathbf{B}$ are $n \times n$ matrices, then $\mathbf{AB} = \mathbf{I_n}$ if and only if $\mathbf{BA} = \mathbf{I_n}$. We additionally touched upon adding the line

- $\mathbf{A^T}$ is invertible and $(\mathbf{A}^{-1})^T$ is its inverse.

to the Invertible Matrix Theorem. Proving this requires a theorem that states

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T,$$

which will be proved for homework. Taking this as a given, though, we can show

$$\begin{aligned}(\mathbf{A}^T)(\mathbf{A}^{-1})^T &= (\mathbf{A}^{-1}\mathbf{A})^T \\ &= \mathbf{I}^T \\ &= \mathbf{I},\end{aligned}$$

and thus the given line holds.

Why would we define the notion of a matrix transpose in the first place? Some further notions that we may explore later in the course rely upon the transpose, such as

- A symmetric matrix $\mathbf{A}$ is defined by satisfying $\mathbf{A}^T = \mathbf{A}$.
- A skew-symmetric matrix $\mathbf{A}$ is defined by satisfying $\mathbf{A}^T = -\mathbf{A}$.
- An orthogonal matrix $\mathbf{A}$ is defined by satisfying $\mathbf{A}^T = \mathbf{A}^{-1}$.

Today, we'll introduce the concept of subspaces of $\mathbb{R}^n$.

Definition 17. Let H be a subset of $\mathbb{R}^n$. We say that H is a subspace of $\mathbb{R}^n$ if

a. $\mathbf{0} \in H$.

b. $\forall \mathbf{u}, \mathbf{v} \in H, \mathbf{u} + \mathbf{v} \in H$.

c. $\forall \mathbf{u} \in H, \forall c \in \mathbb{R}, c\mathbf{u} \in H$.

The final two conditions describe the subspace H being "closed" under vector addition and scalar multiplication.

Example. As an example of a subspace of $\mathbb{R}^n$, consider the set

$$H = \{ \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \in \mathbb{R}^3 \mid x, y \in \mathbb{R}\}.$$

This set essentially represents the xy-plane embedded in $\mathbb{R}^3$. This trivially satisfies the first condition of the definition, satisfies the second condition by

$$\begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} x_3 \\ x_4 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 + x_3 \\ x_2 + x_4 \\ 0 \end{bmatrix},$$

since this sum is contained within H, and satisfies the final condition by

$$c \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} cx_1 \\ cx_2 \\ 0 \end{bmatrix},$$

because this product is again within H. Interestingly, performing these verifications with sets of form

$$H = \{ \begin{bmatrix} x \\ y \\ a \end{bmatrix} \in \mathbb{R}^3 \mid x, y \in \mathbb{R}\} : a \in \mathbb{R}_{\neq 0},$$

which represent the xy-plane translated along the z-axis by a non-zero scalar, results in a failure to satisfy the definition of a subspace. Sets that represent translations of legitimate subspaces by a vector, like the example just presented, are referred to as "affine subspaces." $\diamond$

Some further examples of subspaces include $\mathbb{R}^n$, which trivially satisfies the definition of a subspace, the set $\{\mathbf{0}\}$, and the set

$$\begin{aligned} H &= \{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x = 2y\} \\ &= \{ \begin{bmatrix} 2y \\ y \end{bmatrix} \in \mathbb{R}^2 \mid y \in \mathbb{R}\}, \end{aligned}$$

which satisfies the first condition where $y = 0$, satisfies the second condition by

$$\begin{aligned} \begin{bmatrix} 2x_1 \\ x_1 \end{bmatrix} + \begin{bmatrix} 2x_2 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 2x_1 + 2x_2 \\ x_1 + x_2 \end{bmatrix} \\ &= \begin{bmatrix} 2(x_1 + x_2) \\ x_1 + x_2 \end{bmatrix}, \end{aligned}$$

and the third by

$$\begin{aligned} c \begin{bmatrix} 2x_1 \\ x_1 \end{bmatrix} &= \begin{bmatrix} 2cx_1 \\ cx_2 \end{bmatrix} \\ &= \begin{bmatrix} 2(cx_1) \\ cx_1 \end{bmatrix}. \end{aligned}$$

Now, we'll work through some in-class problems whose solutions are available on the Moodle page. From examining these problems, we come across the following theorems.

Theorem 25. Let $\mathbf{v_1}, \ldots, \mathbf{v_p} \in \mathbb{R}^n$. Then $\operatorname{span}(\mathbf{v_1}, \ldots, \mathbf{v_p})$ is a subspace of $\mathbb{R}^n$.

Theorem 26. Let $\mathbf{A}$ be an $m \times n$ matrix. The set of solutions of $\mathbf{Ax} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$.

These theorems will be further examined during the next lecture.

Lecture 22: Null & Column Spaces of a Matrix

October 8, 2021

During the last lecture, we introduced a notion of subspaces of $\mathbb{R}^n$. Subspaces of $\mathbb{R}^n$ "look like" miniature versions of $\mathbb{R}^m$ where $m \leq n$. For example, in $\mathbb{R}^3$, the possible subspaces are $\mathbb{R}^m$, any plane containing $\mathbf{0}$, any line containing $\mathbf{0}$, and $\{\mathbf{0}\}$ itself. We also discussed two theorems.

Theorem 27. If $\mathbf{A}$ is an $m \times n$ matrix, then $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}$ is a subspace called the "null space of $\mathbf{A}$" and is denoted $\operatorname{Null}(\mathbf{A})$.

Proof. Let $\mathbf{A}$ be an $m \times n$ matrix. Trivially, the vector $\mathbf{x} = \mathbf{0}$ satisfies the equation $\mathbf{Ax} = \mathbf{0}$, verifying the first subspace condition. Then, suppose that vectors $\mathbf{u}, \mathbf{v}$ satisfy $\mathbf{Ax} = \mathbf{0}$. By the linearity of $\mathbf{A}$, we can find that

$$\begin{aligned} \mathbf{0} &= \mathbf{A}(\mathbf{u}) + \mathbf{A}(\mathbf{v}) \\ &= \mathbf{A}(\mathbf{u} + \mathbf{v}), \end{aligned}$$

verifying the second condition. And, finally,

$$\begin{aligned} \mathbf{0} &= c\mathbf{A}(\mathbf{u}) \\ &= \mathbf{A}(c\mathbf{u}), \end{aligned}$$

verifying the third. This completes the proof. $\square$

Theorem 28. If $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p} \in \mathbb{R}^n$ then $\operatorname{span}(\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p})$ is a subspace of $\mathbb{R}^n$.

Definition 18. If $\mathbf{A}$ is an $m \times n$ matrix

$$A = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \ldots & \mathbf{a_n} \end{bmatrix},$$

then the "column space of $\mathbf{A}$," denoted $\operatorname{Col}(\mathbf{A})$, is

$$\operatorname{span}(\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_n}).$$

Now, we can make an important observation. If $\mathbf{A}$ is an $m \times n$ matrix then $\operatorname{Null}(\mathbf{A})$ is a subspace of $\mathbb{R}^n$ and $\operatorname{Col}(\mathbf{A})$ is a subspace of $\mathbb{R}^m$. So, we can visualize the null space as a subspace of the domain of the transformation represented $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ which is mapped onto $\mathbf{0}$ in the co-domain, while we can consider the column space as being the subspace of the co-domain created by the span of the columns of $\mathbf{A}$. These visualizations are realized in Figure 8. Now, let's consider an example.

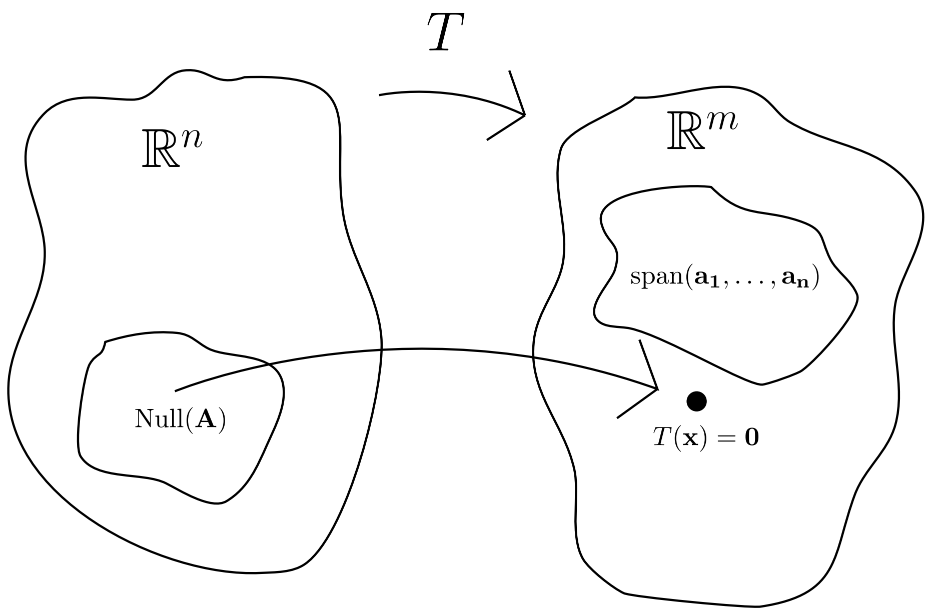


Figure 8: Visualized Null and Column Spaces

Example. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}.$$

And further

$$\text{RREF}(\mathbf{A}) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

The null space of $\mathbf{A}$ represents all of the solutions to $\mathbf{A}\mathbf{x} = \mathbf{0}$, which, in this case, resolves to the set

$$x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} : x_3 \in \mathbb{R} = \text{span}(\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}),$$

because the RREF form gives

$$x_1 = x_3$$
$$x_2 = -2x_3$$
$$x_3 = x_3.$$

Further, the column space of $\mathbf{A}$ can be expressed as

$$\text{span}(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}7\\8\\9\end{bmatrix}) = \text{span}(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}),$$

which is notably a non-unique representation and is dependent on whether $\mathbf{A}$ is in RREF or not (i.e., the column space of a matrix changes with elementary operations). $\diamond$

Definition 19. If H is a subspace of $\mathbb{R}^n$, we say that

$$B = \{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$$

is a basis of H if $\text{span}(\mathbf{b_1}, \ldots, \mathbf{b_p}) = H$ and $\{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$ is linearly independent.

Now, for a simple example.

Example. Trivially, the set

$$B = \{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$$
$$= \{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\}$$

is a basis of $\mathbb{R}^3$. $\diamond$

Now, let's introduce a bit of useful terminology.

Notation. To say that two mathematical objects are **isomorphic** is to say that they "look like" each other.

Finally, we'll complete a set of in-class problems whose solutions are available on the class Moodle page.

Lecture 23: Bases & Dimension

October 18, 2021

Today, we're going to talk about bases and dimension. Before the break, we discussed a basis for a matrix's column space. Recall that if $\mathbf{A}$ is an $m \times n$

matrix, then the column space of $\mathbf{A}$ is the span of the columns of $\mathbf{A}$. Further, recall that if H is a subspace of $\mathbb{R}^n$, then we say that $B = \{\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_p}\}$ is a basis of H if $\text{span}(\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_p}) = H$ and $\{\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_p}\}$ are linearly independent. Today, we will sketch a proof of the following theorem.

> **Theorem 29.** The pivot columns of a matrix $\mathbf{A} = \begin{bmatrix}\mathbf{a_1} & \mathbf{a_2} & \ldots & \mathbf{a_n}\end{bmatrix}$ form a basis of $\text{Col}(\mathbf{A})$.

The general idea behind this proof is that given the subspace $\text{Col}(\mathbf{A}) = \text{span}(\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_n})$, we will look for columns $\mathbf{a_i}$ which are dependent, then remove them to form a basis. Let's consider an example.

Example. Given the following matrices $\mathbf{A}$ and $\mathbf{B} = \text{RREF}(\mathbf{A})$,

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & 8 \end{bmatrix}$$
$$\mathbf{B} = \begin{bmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

then, trivially, the pivot columns of $\mathbf{B}$ form a basis for $\text{Col}(\mathbf{B})$. By the aforementioned theorem, the corresponding columns of $\mathbf{A}$—that is, the first, second, and fifth—form a basis of $\mathbf{A}$. Hence, for $\mathbf{A}$,

$$B = \{\begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ 1 \\ 8 \end{bmatrix}\}.$$

Note that in both $\mathbf{A}$ and $\mathbf{B}$, the third column can be conceived as a dependence relation of the first two columns with respective weights -3 and 2. This follows for the other non-pivot columns (the fourth) with other weights. $\diamond$

From this final observation arises the following claim. If $\mathbf{A}$ is row-equivalent to $\mathbf{B}$, then the columns of $\mathbf{A}$ have the same linear relations as the columns of $\mathbf{B}$. Given matrices $\mathbf{A} = \begin{bmatrix}\mathbf{a_1} & \ldots & \mathbf{a_n}\end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix}\mathbf{b_1} & \ldots & \mathbf{b_n}\end{bmatrix}$, then the following statements are logically equivalent.

$$
\begin{aligned}
\mathbf{b_3} = -3\mathbf{b_1} + 2\mathbf{b_2} &\Leftrightarrow \mathbf{0} = -3\mathbf{b_1} + 2\mathbf{b_2} - \mathbf{b_3} \\
&\Leftrightarrow \mathbf{0} = \begin{bmatrix} \mathbf{b_1} & \mathbf{b_2} & \mathbf{b_3} & \mathbf{b_4} & \mathbf{b_5} \end{bmatrix} \begin{bmatrix} -3 \\ 2 \\ -1 \\ 0 \\ 0 \end{bmatrix} \\
&\Leftrightarrow \mathbf{0} = \mathbf{B} \begin{bmatrix} -3 \\ 2 \\ -1 \\ 0 \\ 0 \end{bmatrix} \\
&\Leftrightarrow \mathbf{0} = \mathbf{A} \begin{bmatrix} -3 \\ 2 \\ -1 \\ 0 \\ 0 \end{bmatrix} \\
&\Leftrightarrow \mathbf{0} = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} & \mathbf{a_5} \end{bmatrix} \begin{bmatrix} -3 \\ 2 \\ -1 \\ 0 \\ 0 \end{bmatrix} \\
&\Leftrightarrow \mathbf{0} = -3\mathbf{a_1} + 2\mathbf{a_2} - \mathbf{a_3} \\
&\Leftrightarrow \mathbf{a_3} = -3\mathbf{a_1} + 2\mathbf{a_2},
\end{aligned}
$$

because the equations $\mathbf{Ax} = \mathbf{0}$ and $\mathbf{Bx} = \mathbf{0}$ must share the same solution set. Therefore, the claim holds and the proof is sketched.

Now, let's introduce a new theorem and a notion of dimension.

Theorem 30. All bases of a subspace H of $\mathbb{R}^n$ have the same number of vectors. This will be proved for homework.

Definition 20. The **dimension** of H is defined as the number of vectors in any of its bases. We further define $\dim(\mathbf{0}) = \mathbf{0}$.

Example. Let's consider the case $\dim(\mathbb{R}^4) = 4$. Since a basis of $\mathbb{R}^4$ is

$$
\begin{aligned}
B &= \{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}, \mathbf{e_4}\} \\
&= \{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \},
\end{aligned}
$$

and this set has 4 elements, then it follows that $\dim(\mathbb{R}^4) = 4$. $\diamond$

Example. Let's create a list of all subspaces of $\mathbb{R}^3$ by dimension. This gener-

Table 2

Dimension	Subspace
3	$\mathbb{R}^3$
2	Planes through $\mathbf{0}$
1	Lines through $\mathbf{0}$
0	$\{\mathbf{0}\}$

alizes to other dimensions with a similar—although more difficult to visualize—intuition. $\diamond$

Now, we'll work through a series of in-class problems whose solutions can be found on this class' Moodle page.

Finally, we'll consider a theorem that will be explored further in later lectures.

Theorem 31. If $\mathbf{A}$ is an $m \times n$ matrix, then

$$\dim(\mathrm{Col}(\mathbf{A})) + \dim(\mathrm{Null}(\mathbf{A})) = n.$$

Notation. Given a matrix $\mathbf{A}$, $\dim(\mathrm{Col}(\mathbf{A}))$ is also known as $\mathrm{rank}(\mathbf{A})$.

Lecture 24: Introducing Coordinates

October 20, 2021

To begin, let's review the definition of a matrix's column and null spaces.

Definition 21. Given an $m \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{a_1} & \dots & \mathbf{a_n} \end{bmatrix},$$

then the column space of this matrix is defined

$$\text{Col}(\mathbf{A}) = \text{span}(\mathbf{a_1}, \dots, \mathbf{a_n}).$$

The column space can be interpreted as the set of all possible outputs $\mathbf{b}$ of the multiplication of this matrix by a vector. To illustrate,

$$\text{Col}(\mathbf{A}) = \begin{bmatrix} \mathbf{a_1} & \dots & \mathbf{a_n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1 \dots, x_n \in \mathbb{R}.$$

This is also known as the image (or range) of the associated matrix transformation $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

Definition 22. Given an $m \times n$ matrix $\mathbf{A}$, then the null space of this matrix is defined

$$\text{Null}(\mathbf{A}) = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}.$$

This can be interpreted as the set of all vectors $\mathbf{x}$ in the matrix transformation's domain that are mapped to the zero vector in the co-domain.

Finally, recall the theorem presented last class.

Theorem 32. Given an $m \times n$ matrix $\mathbf{A}$, then

$$\dim(\text{Col}(\mathbf{A})) + \dim(\text{Null}(\mathbf{A})) = n,$$

or, using alternative notation,

$$\text{Rank}(\mathbf{A}) + \text{Nullity}(\mathbf{A}) = n.$$

Now, let's examine the notion of coordinates with respect to a basis.

Theorem 33. Let H be a subspace of $\mathbb{R}^n$ with basis $B = \{\mathbf{b_1}, \dots, \mathbf{b_p}\}$. Every vector $\mathbf{x} \in H$ can be written in **exactly** one way as a linear combination of these basis vectors.

Note that we could use the basis given in Theorem 33 to express every $\mathbf{x} \in H$ as

$$\mathbf{x} = c_1\mathbf{b_1} + \dots + c_p\mathbf{b_p}.$$

From this observation arises the following definition.

Definition 23. Given a basis $B = \{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$ of a subspace H, we can express every $\mathbf{x} \in H$ as a linear combination of these basis vectors

$$\mathbf{x} = c_1\mathbf{b_1} + \ldots + c_p\mathbf{b_p}.$$

Then, we could collect these weights into a vector

$$\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \in \mathbb{R}^p,$$

which we call the **coordinates of x** with respect to a basis B, further denoted by $[\mathbf{x}]_B$.

Now, let's provide a proof of Theorem 33.

Proof. Let H be a subspace of $\mathbb{R}^p$ and $B = \{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$ be a basis of H. We know that there is at least one way to write every $\mathbf{x} \in H$ as a linear combination of these basis vectors since $H = \operatorname{span}(\mathbf{b_1}, \ldots, \mathbf{b_p})$. To prove that such a linear combination is unique, suppose that we could express a vector $\mathbf{x} \in H$ in two different ways

$$\begin{aligned} \mathbf{x} &= c_1\mathbf{b_1} + \ldots + c_p\mathbf{b_p} \\ \mathbf{x} &= d_1\mathbf{b_1} + \ldots + d_p\mathbf{b_p}. \end{aligned}$$

Subtracting these expressions from each other gives

$$\mathbf{0} = (c_1 - d_1)\mathbf{b_1} + \ldots + (c_p - d_p)\mathbf{b_p}.$$

Since $\{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$ is independent, then

$$\begin{aligned} c_1 - d_1 &= 0 \\ &\vdots \\ c_p - d_p &= 0, \end{aligned}$$

and thus $c_i = d_i$. This completes the proof. □

Finally, we'll work through some in-class problems whose solutions are available on the class Moodle page.

Lecture 25: Proofs by Induction

October 21, 2021

Let's begin this discussion session by exploring the first part of the third question on the fifth homework assignment. This problem asks us to find a specific 4×4

matrix $\mathbf{A}$ such that $\text{Col}(\mathbf{A}) = \text{Null}(\mathbf{A})$. By the Rank-nullity Theorem, we know that because $\text{Rank}(\mathbf{A}) = \text{Nullity}(\mathbf{A})$, then $\text{Rank}(\mathbf{A}) = \text{Nullity}(\mathbf{A}) = 2$. This should be help enough for us to solve the rest of this problem on our own.

Next, let's examine the fourth question. In this problem, we are asked to prove that if B_1 and B_2 are bases of a subspace H of $\mathbb{R}^n$, then the number of vectors within each set of bases are equal. Ultimately, this will amount to showing that both

$$|B_1| \leq |B_2|,$$

and

$$|B_2| \leq |B_1|,$$

thus implying that

$$|B_1| = |B_2|.$$

In order to show each of these dependencies, we will need to show that given $W = \text{span}(\mathbf{b_1}, \dots, \mathbf{b_p})$ and $\{\mathbf{a_1}, \dots, \mathbf{a_m}\} \in W$ with $m > p$, then $\{\mathbf{a_1}, \dots, \mathbf{a_m}\}$ is linearly dependent. We can collect these sets of vectors into matrices

$$\begin{aligned} \mathbf{B} &= \begin{bmatrix} \mathbf{b_1} & \dots & \mathbf{b_p} \end{bmatrix} \\ \mathbf{A} &= \begin{bmatrix} \mathbf{a_1} & \dots & \mathbf{a_m} \end{bmatrix}. \end{aligned}$$

Then showing that there exists a vector $\mathbf{c_i}$ for every $\mathbf{a_i}$ with $\mathbf{a_i} = \mathbf{B}\mathbf{c_i}$ shows that every $\mathbf{a_i}$ is within $\text{Col}\mathbf{B}$.

Now, we'll work through a problem that demonstrates the method of induction. Prove that

$$\frac{1}{1 \times 2} + \frac{1}{2 \times 3} + \dots + \frac{1}{n(n+1)} = \frac{n}{n+1}.$$

Proof. Let's begin with our base case where $n = 1$. Trivially,

$$\frac{1}{1 \times 2} = \frac{1}{2},$$

so this is true. Next, we'll proceed with the inductive step. Suppose that

$$\frac{1}{1 \times 2} + \frac{1}{2 \times 3} + \dots + \frac{1}{n(n+1)} = \frac{n}{n+1}.$$

Then, by the inductive hypothesis, in the $(n+1)$st case

$$\begin{aligned}
\frac{1}{1\times 2}+\ldots+\frac{1}{n(n+1)}+\frac{1}{(n+1)(n+2)} &= \frac{n}{n+1}+\frac{1}{(n+1)(n+2)} \\
&= \frac{n(n+2)+1}{(n+1)(n+2)} \\
&= \frac{n^2 2n+1}{(n+1)(n+2)} \\
&= \frac{(n+1)^2}{(n+1)(n+2)} \\
&= \frac{n+1}{n+2} \\
&= \frac{(n+1)}{(n+1)+1}.
\end{aligned}$$

As the induction hypothesis implies the $(n+1)$st case, then, by induction, the given statement is true for all $n \geq 1$. This completes the proof. $\square$

Now, for another proof by induction. Prove that for all $n \in \mathbb{N}$ where $r \in \mathbb{R}_{\neq 1}$ then

$$1+r+r^2+\ldots+r^n = \frac{1-r^{n+1}}{1-r}.$$

Proof. Let's first consider the case $n=1$. This gives

$$\begin{aligned}
\frac{1-r^2}{1-r} &= \frac{(1+r)(1-r)}{(1-r)} \\
&= 1+r,
\end{aligned}$$

which thus holds. Next, consider the induction hypothesis

$$1+r+r^2+\ldots+r^n = \frac{1-r^{n+1}}{1-r},$$

which we suppose is true. In the $(n+1)$st case, we find by the hypothesis that

$$\begin{aligned}
1+r+r^2+\ldots+r^n+r^{n+1} &= \frac{1-r^{n+1}}{1-r}+r^{n+1} \\
&= \frac{1-r^{n+1}+(1-r)r^{n+1}}{n-1} \\
&= \frac{1-r^{n+2}}{1-r}.
\end{aligned}$$

As the induction hypothesis implies the $(n+1)$st case, then, by induction, the given statement is true for all $n \in \mathbb{N}$. This completes the proof. $\square$

Finally, we'll prove that given matrices $\mathbf{A}, \mathbf{P}$, for all $n \in \mathbb{N}$,

$$(\mathbf{PAP}^{-1})^n = \mathbf{PA}^n\mathbf{P}^{-1}.$$

Proof. Given the statement

$$(\mathbf{PAP}^{-1})^n = \mathbf{PA}^n\mathbf{P}^{-1},$$

the base case $n = 1$ holds as

$$\mathbf{PAP}^{-1} = \mathbf{PAP}^{-1}.$$

As the induction hypothesis, suppose that

$$(\mathbf{PAP}^{-1})^n = \mathbf{PA}^n\mathbf{P}^{-1}.$$

In the $(n+1)$st case, this gives

$$\begin{aligned}(\mathbf{PAP}^{-1})^{n+1} &= (\mathbf{PAP}^{-1})(\mathbf{PAP}^{-1})^n \\ &= (\mathbf{PAP}^{-1})\mathbf{PA}^n\mathbf{P}^{-1} \\ &= \mathbf{PA}^{n+1}\mathbf{P}^{-1}.\end{aligned}$$

As the induction hypothesis implies the $(n+1)$st case, then, by induction, the given statement is true for all $n \in \mathbb{N}$. This completes the proof. ☐

Lecture 26: Change of Basis

October 22, 2021

Last time, we examined a subspace $H \in \mathbb{R}^n$ with basis $B = \{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$. Given an arbitrary element $\mathbf{x} \in H$, we can represent this vector uniquely as

$$\mathbf{x} = c_1\mathbf{b_1} + \ldots + c_p\mathbf{b_p}.$$

This yields the coordinate vector

$$[\mathbf{x}]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \in \mathbb{R}^p.$$

Given this vector, we can thus further represent $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{P}_B[\mathbf{x}]_B,$$

where $\mathbf{P}_B = \begin{bmatrix}\mathbf{b_1} & \ldots & \mathbf{b_p}\end{bmatrix}$. In the case where $H = \mathbb{R}^n$, then $\mathbf{P}_B$ is an $n \times n$ invertible matrix, so this can also be rearranged to

$$[\mathbf{x}]_B = \mathbf{P}_B^{-1}\mathbf{x}.$$

Theorem 34. Consider a transformation $T : H \to \mathbb{R}^p$ given

$$T(\mathbf{x}) = [\mathbf{x}]_B.$$

Then T must be linear and bijective (both one-to-one and onto) (i.e., T is an isomorphism).

Today, let's suppose that we are given two bases for $H = \mathbb{R}^n$ as

$$\begin{aligned} B &= \{\mathbf{b_1}, \ldots, \mathbf{b_n}\} \\ C &= \{\mathbf{c_1}, \ldots, \mathbf{c_n}\}. \end{aligned}$$

We know that given a vector $\mathbf{x} \in \mathbb{R}^n$ we can find $[\mathbf{x}]_B$ by left-multiplying by $\mathbf{P}_B^{-1}$ and we can find $[\mathbf{x}]_C$ by left-multiplying by $\mathbf{P}_C^{-1}$. What if we wanted to take the vector $[\mathbf{x}]_B$ and find $[\mathbf{x}]_C$ directly? We could first convert $[\mathbf{x}]_B$ to $\mathbf{x}$ by left-multiplying by $\mathbf{P}_B$ then convert to $[\mathbf{x}]_C$ by left-multiplying by $\mathbf{P}_C^{-1}$, finally giving the operation $\mathbf{P}_C^{-1}\mathbf{P}_B[\mathbf{x}]_B$ for the direct change of basis. We term this matrix

$$\underset{C \leftarrow B}{\mathbf{P}} = \mathbf{P}_C^{-1}\mathbf{P}_B,$$

and, by the "shoes-and-socks" theorem, we can find the inverse matrix to be

$$\begin{aligned} \underset{C \leftarrow B}{\mathbf{P}}^{-1} &= (\mathbf{P}_C^{-1}\mathbf{P}_B)^{-1} \\ &= \mathbf{P}_B^{-1}\mathbf{P}_C \\ &= \underset{B \leftarrow C}{\mathbf{P}}. \end{aligned}$$

A visual depiction of this relationship can be found in Figure 9.

As a note, in more general cases, this matrix is defined

$$\underset{C \leftarrow B}{\mathbf{P}} = \begin{bmatrix} [\mathbf{b_1}]_C & \ldots & [\mathbf{b_p}]_C \end{bmatrix}.$$

Finally, we're going to pivot and discuss the application of matrices for encoding networks. Let's consider an example.

Example. Consider the network with 4 nodes given by Figure 10.
We can represent all possible movements around the network with a 4×4 **adjacency matrix** defined by the rule

$$a_{ij} = \begin{cases} 1 \text{ if there is a link from node } i \text{ to } j \\ 0 \text{ otherwise} \end{cases}.$$

In the given case, this gives matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

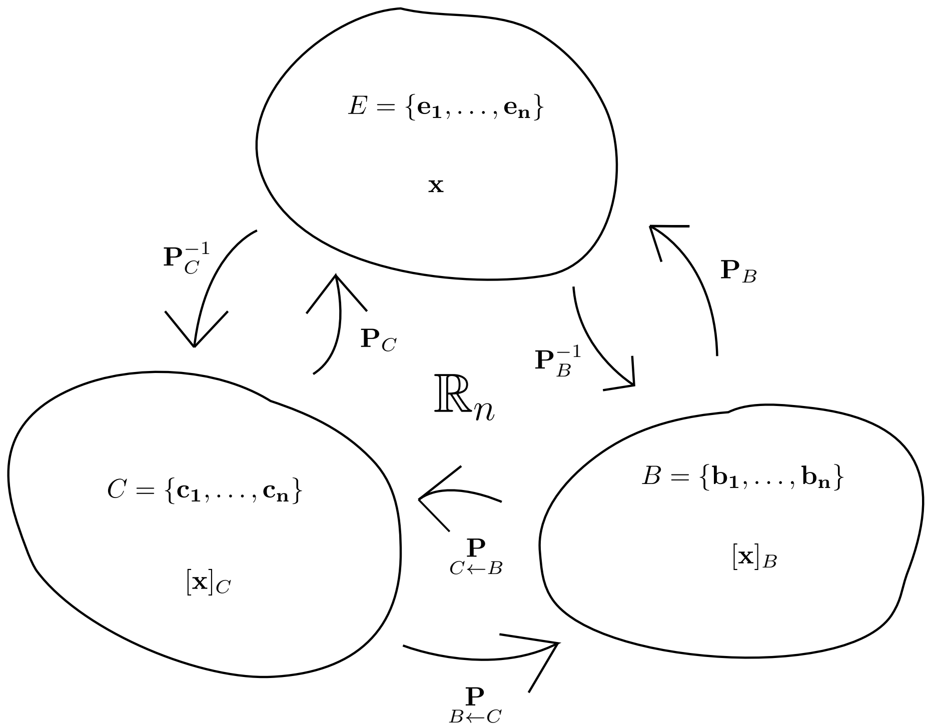


Figure 9: The Relationship Between Three Bases in $\mathbb{R}_n$

Interestingly, we can find that

$$a_{ij}^2 = \# \text{ of paths of length 2 from node } i \text{ to } j.$$

Using matrix multiplication to find $\mathbf{A}^2$, we can calculate $a_{11}^2 = 1$, which tells us that there is one way to get from node 1 to node 1 using two steps. $\diamond$

Lecture 27: The Determinant

October 25, 2021

Today, we're going to discuss the determinant. Despite the large amount of mathematics behind and surrounding the determinant, in the interest of time we will solely be considering how the determinant relates to the theorem that $\mathbf{A}$ is invertible if and only if $\det\mathbf{A} \neq 0$.

> Among the determinant's many applications, the determinant of a 2×2 matrix
>
> $$\mathbf{A} = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} \end{bmatrix}$$

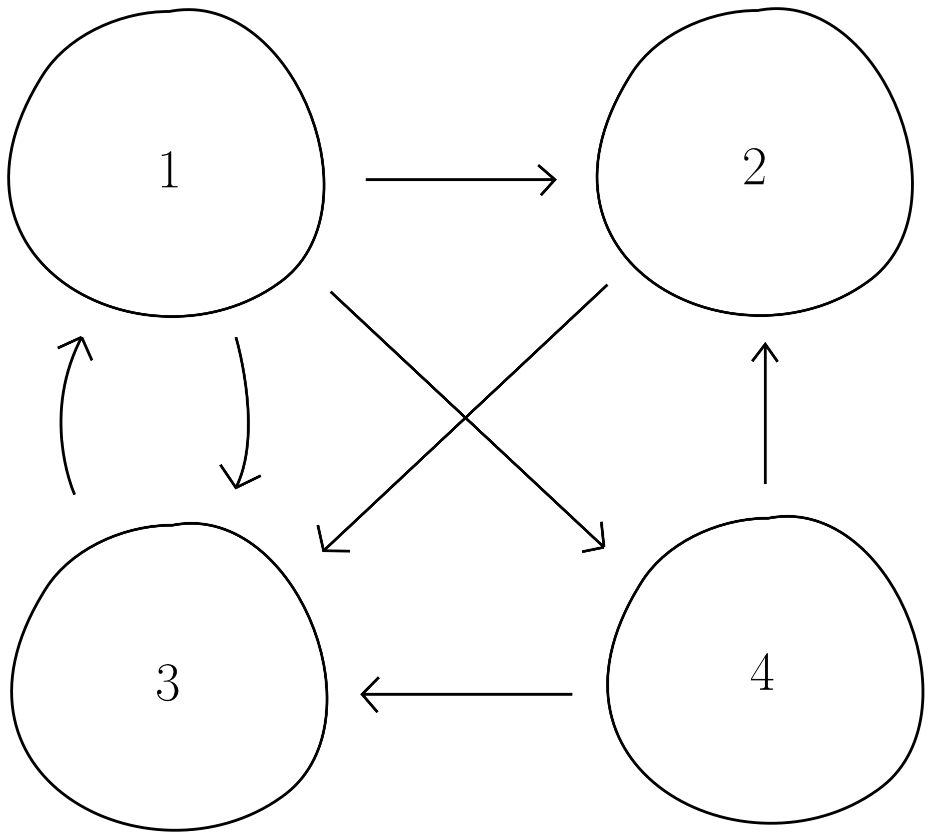


Figure 10: A Network with Four Nodes

is equivalent to the area of the parallelogram formed by vectors $\mathbf{v_1}, \mathbf{v_2}$, and further the determinant of a 3×3 matrix

$$\mathbf{B} = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} \end{bmatrix}$$

is the area of the parallelepiped formed by vectors $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$.

The determinant also appears in the calculation of multiple integrals. In this case, consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Parallelograms in this transformation's domain are transformed into co-domain parallelograms with area

$$\text{area of output parallelogram} = |\det\mathbf{A}| \text{ area of input parallelogram.}$$

This "area conversion factor" and its higher-dimensional analogues are used heavily in multiple-integration.

Now, let's define det$\mathbf{A}$ in terms of the entries of $\mathbf{A}$ in $\mathbb{R}^2$. Suppose that $a_{11} \neq 0$, we can reduce $\mathbf{A}$ as

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \sim \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix}.$$

Therefore, $\mathbf{A}$ is invertible if and only if $a_{11}a_{22} - a_{12}a_{21} \neq 0$. This is indeed the definition of the determinant of a 2×2 matrix. Given a 3×3 matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

performing this same row-reduction yields the condition that

$$\begin{aligned} &(a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &- a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}) \neq 0, \end{aligned}$$

for $\mathbf{A}$ to be invertible. We can factor this easily to

$$\begin{aligned} &(a_{11}(a_{22}a_{33} - a_{23}a_{32}) && (a_{11}\det(\mathbf{A}_{11}) \\ &- a_{12}(a_{23}a_{31} - a_{21}a_{33}) = && - a_{12}\det(\mathbf{A}_{12}) \\ &+ a_{13}(a_{21}a_{32} - a_{22}a_{31})) && + a_{13}\det(\mathbf{A}_{13})), \end{aligned}$$

where the matrices $\mathbf{A}_{ij}$ are the 2×2 minor matrices created by removing the ith row and jth column of $\mathbf{A}$. Hence, we can simplify the determinant of a 3×3 matrix to determinants of 2×2 matrices. With this, we can give the recursive definition of the determinant of an $n \times n$ matrix $\mathbf{A}$ as

$$\begin{aligned} \det\mathbf{A} &= |\mathbf{A}| \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(\mathbf{A}_{1j}) \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij}) \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij}). \end{aligned}$$

This definition gives an algorithm for calculating the determinant of an $n \times n$ matrix $\mathbf{A}$ via **cofactor expansion**, a weighted sum of the determinants of the $(n-1) \times (n-1)$ minors of $\mathbf{A}$. We'll now consider some examples of this method within the in-class problems, whose solutions are available on the class' Moodle page.

Lecture 28: Properties of Determinants

October 27, 2021

In the pre-lecture video, we determined the following.

- The determinant of the $n \times n$ identity matrix $\mathbf{I}_n$ is

$$\det(\mathbf{I}_n) = 1.$$

- If $\mathbf{A}$ is an upper or lower triangular matrix—that is, it either has exclusively zeros below or above the diagonal—then the determinant of $\mathbf{A}$ is the product of the diagonal entries, or

$$\begin{aligned}\det\mathbf{A} &= a_{11}a_{22}\ldots a_{nn}\\ &= \prod_{i=1}^{n} a_{ii}.\end{aligned}$$

- Given an $n \times n$ matrix $\mathbf{A}$, then

$$\det(\mathbf{A}^T) = \det\mathbf{A}.$$

- The elementary row operations have the following effects on the determinant of an $n \times n$ matrix $\mathbf{A}$.

  1. For a replacement operation, the determinant is unchanged.
  2. For a swap operation, the determinant is multiplied by -1.
  3. For a scaling operation by $k \neq 0$, the determinant is multiplied by k.

Today, we will explore the following facts.

- An $n \times n$ matrix $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$.
- Given two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, then

$$\det(\mathbf{AB}) = \det\mathbf{A} + \det\mathbf{B}.$$

We'll now complete some in-class problems whose solutions are available on the class' Moodle page.

Theorem 35. An $n \times n$ matrix $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$.

Proof. Let $\mathbf{A}$ be an $n \times n$ matrix with echelon form $\mathbf{U}$ and let $\mathbf{A}$ be reducible to this matrix using only replacement and row-swap operations—such a matrix must exist for every matrix $\mathbf{A}$.

- For the forward case, suppose that $\mathbf{A}$ is invertible. Then, $\mathbf{A}$ has n pivots. Hence, $\mathbf{U}$ has exclusively non-zero entries in its main diagonal and therefore $\det\mathbf{U}$ is non-zero. Because $\mathbf{A}$ is reducible to $\mathbf{U}$ by row-swaps and replacement operations alone, then $\det\mathbf{A} = (-1)^h\det\mathbf{U}$ where h is the number of row-swap operations required. Thus, $\det\mathbf{A} \neq 0$.

- For the converse, suppose that $\det\mathbf{A} \neq 0$. Then, the product of the entries of the main diagonal of $\mathbf{U}$ is non-zero and thus each entry must be non-zero. This means that the matrix $\mathbf{A}$ has n pivots and therefore is invertible by the Invertible Matrix Theorem.

This completes the proof. $\square$

Theorem 36. Given two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, then

$$\det(\mathbf{AB}) = \det\mathbf{A} + \det\mathbf{B}.$$

In the interest of time, a proof of this second theorem can be found in the Lay textbook.

Lecture 29: Exam 2 Preparation

October 28, 2021

We'll spend this discussion session working through a set of true-false questions intended to prepare us for next week's exam. These questions and their solutions are available on the class' Moodle page.

Lecture 30: Introducing Eigenthings

October 29, 2021

Let's begin by recalling the Markov chain depicted in Figure 11. This relation-

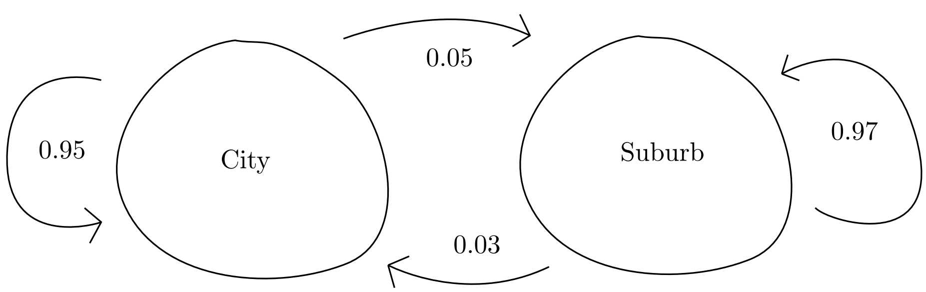


Figure 11: A simple Markov chain.

ship can also be represented by the stochastic matrix

$$\mathbf{M} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}.$$

When multiplied by a state vector, a vector that describes the population in each location at a given point in time, this matrix gives the change in this state over a given period, as in

$$\begin{aligned}\mathbf{M}\begin{bmatrix}c_n\\s_n\end{bmatrix} &= \begin{bmatrix}0.95c_n+0.03s_n\\0.05c_n+0.97s_n\end{bmatrix}\\ &= \begin{bmatrix}c_{n+1}\\s_{n+1}\end{bmatrix}.\end{aligned}$$

When this state vector represents a proportion of the population rather than the absolute population itself, it is termed a probability vector. To find an equilibrium state which is unchanging over time, we would thus need to find a state vector satisfying

$$\begin{aligned}\mathbf{Mx}=\mathbf{0} &\Leftrightarrow \mathbf{Mx}-\mathbf{x}=\mathbf{0}\\ &\Leftrightarrow (\mathbf{M}-\mathbf{I})\mathbf{x}=\mathbf{0}\\ &\Leftrightarrow \mathbf{x}\in \mathrm{Null}(\mathbf{M}-\mathbf{I}).\end{aligned}$$

Using this, we can row-reduce the matrix $\mathbf{M}-\mathbf{I}$ to find that

$$\begin{aligned}\mathbf{M}-\mathbf{I} &= \begin{bmatrix}0.95-1 & 0.03\\ 0.05 & 0.97-1\end{bmatrix}\\ &= \begin{bmatrix}-0.05 & 0.03\\ 0.05 & -0.03\end{bmatrix}\\ &\sim \begin{bmatrix}-0.05 & 0.03\\ 0 & 0\end{bmatrix},\end{aligned}$$

and thus the vector $\begin{bmatrix}0.03\\0.05\end{bmatrix}$ and any scalar multiple of this vector (e.g., the vectors $\begin{bmatrix}3\\5\end{bmatrix}$, $\begin{bmatrix}\frac{3}{8}\\\frac{5}{8}\end{bmatrix}$, $\begin{bmatrix}0.375\\0.625\end{bmatrix}$) is in the null space of $\mathbf{M}-\mathbf{I}$. Therefore, the steady-state of this Markov chain occurs where 37.5% of the population live in the city and 62.5% live in the suburbs. We can then use Mathematica to verify that

$$\mathbf{M}\begin{bmatrix}0.375\\0.625\end{bmatrix}=\begin{bmatrix}0.375\\0.625\end{bmatrix},$$

and further that

$$\lim_{p\to\infty}\mathbf{M}^p\mathbf{x}=\begin{bmatrix}0.375\\0.625\end{bmatrix},$$

for any probability vector $\mathbf{x}$. Now, let's generalize.

Definition 24. Given an $n\times n$ matrix $\mathbf{A}$, a non-zero vector $\mathbf{x}\neq 0\in\mathbb{R}^n$, and a scalar $\lambda\in\mathbb{R}$ such that they satisfy $\mathbf{Ax}=\lambda\mathbf{x}$, we say that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ.

Let's consider a simple example.

Example. Let

$$\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 3 & 6 \end{bmatrix}.$$

The vector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda = 3$ because

$$\mathbf{A}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 3\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is not an eigenvector of $\mathbf{A}$ because

$$\begin{aligned} \mathbf{A}\begin{bmatrix} 1 \\ 2 \end{bmatrix} &= \begin{bmatrix} 9 \\ 15 \end{bmatrix} \\ &\neq \lambda \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \end{aligned}$$

$\diamond$

Example. Let

$$\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

This matrix has eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ because

$$\begin{aligned} \mathbf{B}\begin{bmatrix} 1 \\ 1 \end{bmatrix} &= 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \mathbf{B}\begin{bmatrix} 3 \\ 3 \end{bmatrix} &= 2\begin{bmatrix} 3 \\ 3 \end{bmatrix} \\ \mathbf{B}\begin{bmatrix} 1 \\ -1 \end{bmatrix} &= 0\begin{bmatrix} 1 \\ -1 \end{bmatrix}. \end{aligned}$$

Hence, any scalar multiple of an eigenvector will also be an eigenvector with the same eigenvalue (in this case, $\lambda = 2$) and 0 may be an eigenvalue as well. $\diamond$

Example. The eigenvalue $\lambda = -3$ is an eigenvalue of the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 4 \\ 6 & 9 \end{bmatrix}.$$

This is because in order for -3 to be an eigenvalue, the following chain of logically equivalent statements would have to be satisfied.

$$\begin{aligned} \mathbf{A}\mathbf{x} = -3\mathbf{x} &\Leftrightarrow \mathbf{A}\mathbf{x} + 3\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow (\mathbf{A} + 3\mathbf{I})\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \mathrm{Null}(\mathbf{A} + 3\mathbf{I}) \neq \{\mathbf{0}\}. \end{aligned}$$

As we can easily find that

$$\begin{aligned}\mathbf{A}+3\mathbf{I} &= \begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix},\end{aligned}$$

and further

$$\begin{bmatrix} -2 \\ 1 \end{bmatrix} \in \text{Null}(\mathbf{A}+3\mathbf{I}),$$

then -3 is an eigenvalue of $\mathbf{A}$. $\diamond$

We'll now work on some in-class problems whose solutions are available on the class' Moodle page.

Lecture 31: Eigenspaces & The Characteristic Polynomial

November 1, 2021

Recall that if $\mathbf{A}$ is a $n \times n$ matrix and $\mathbf{v} \in \mathbb{R}^n_{\neq 0}$, we say that $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ if $\mathbf{Av} = \lambda\mathbf{v}$. Expanding this, the following logical statements are equivalent.

$$\begin{aligned}\mathbf{v} \text{ is an eigenvector of } \mathbf{A} \text{ with eigenvalue } \lambda &\Leftrightarrow \mathbf{Av} = \lambda\mathbf{v} \\ &\Leftrightarrow \mathbf{Av} - \lambda\mathbf{v} = \mathbf{0} \\ &\Leftrightarrow (\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \\ &\Leftrightarrow \text{Null}(\mathbf{A} - \lambda\mathbf{I}) \neq \{\mathbf{0}\} \\ &\Leftrightarrow (\mathbf{A} - \lambda\mathbf{I}) \text{ is not invertible} \\ &\Leftrightarrow \det(\mathbf{A} - \lambda\mathbf{I}) = 0.\end{aligned}$$

Definition 25. If λ is an eigenvalue of $\mathbf{A}$ we call $\text{Null}(\mathbf{A}-\lambda\mathbf{I})$ the eigenspace of $\mathbf{A}$ corresponding to λ denoted by E_λ.

Notably, $\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$ gives an n-degree polynomial in λ termed the "characteristic polynomial." The roots of this polynomial are the eigenvalues of $\mathbf{A}$. Let's consider an example.

Example. Let $\mathbf{A}$ be a 2×2 matrix given

$$\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 3 & 6 \end{bmatrix}.$$

This matrix's characteristic polynomial is thus

$$\begin{aligned}
0 &= \det(\mathbf{A} - \lambda \mathbf{I}) \\
&= \det(\begin{bmatrix} 5-\lambda & 2 \\ 3 & 6-\lambda \end{bmatrix}) \\
&= (5-\lambda)(6-\lambda) - 6 \\
&= (\lambda - 8)(\lambda - 3).
\end{aligned}$$

Hence, the eigenvalues of this matrix are $\lambda = 8, 3$. We can then use these values to find eigenvectors of this matrix. In the case $\lambda = 8$, we know that some of this matrix's eigenvectors satisfy

$$\mathbf{x} \in \text{Null}(\mathbf{A} - 8\mathbf{I}).$$

One such vector is trivially $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and because this subspace is one-dimensional we can thus state that

$$\begin{aligned}
E_8 &= \text{Null}(\mathbf{A} - 8\mathbf{I}) \\
&= \text{span}(\begin{bmatrix} 2 \\ 3 \end{bmatrix}).
\end{aligned}$$

In the other case, $\lambda = 3$, performing the same procedure yields

$$\begin{aligned}
E_3 &= \text{Null}(\mathbf{A} - 3\mathbf{I}) \\
&= \text{span}(\begin{bmatrix} -1 \\ 1 \end{bmatrix}).
\end{aligned}$$

These eigenspaces thus encompass all eigenvectors of $\mathbf{A}$. $\diamond$

We'll now complete some in-class problems whose solutions are available on the class' Moodle page.

Finally, we'll finish by observing that

$$(\mathbf{A} - \lambda \mathbf{I})^T = \mathbf{A}^T - \lambda \mathbf{I}$$

and hence $\mathbf{A}$ and $\mathbf{A}^T$ have the same eigenvalues.

Lecture 32: More on Eigenspaces

November 3, 2021

Today, we'll develop some intuition for next class' proof that eigenvectors with distinct eigenvalues are linearly independent. On last class' worksheet, we examined the matrix

$$\mathbf{A} = \begin{bmatrix} 7 & 4 \\ -3 & -1 \end{bmatrix},$$

with eigenspaces

$$E_1 = \operatorname{span}(\begin{bmatrix} 2 \\ -3 \end{bmatrix})$$
$$E_5 = \operatorname{span}(\begin{bmatrix} 2 \\ -1 \end{bmatrix}).$$

Because these basis vectors are linearly independent and thus ultimately span $\mathbb{R}^2$, we can describe the left-multiplication of an arbitrary vector by the matrix $\mathbf{A}$ as

$$\begin{aligned}
\mathbf{A}\begin{bmatrix} a \\ b \end{bmatrix} &= \mathbf{A}(c_1 \begin{bmatrix} 2 \\ -3 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ -1 \end{bmatrix}) \\
&= c_1\mathbf{A}\begin{bmatrix} 2 \\ -3 \end{bmatrix} + c_2\mathbf{A}\begin{bmatrix} 2 \\ -1 \end{bmatrix} \\
&= c_1 \begin{bmatrix} 2 \\ -3 \end{bmatrix} + c_2 \begin{bmatrix} 10 \\ -5 \end{bmatrix}.
\end{aligned}$$

Notably, though, this is not possible for all 2×2 matrices, as in the case of the matrix

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

which has no eigenvectors, or in the case of the matrix

$$\begin{bmatrix} 10 & -9 \\ 4 & -2 \end{bmatrix},$$

which has only one eigenvalue ($\lambda = 4$). Further, consider the matrix

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

which has the unique eigenvalue $\lambda = 2$ with eigenspace

$$E_2 = \operatorname{span}(\begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}).$$

Hence, singular eigenvalues do not necessarily imply one-dimensional eigenspaces.

We'll now spend the rest of class observing a graphical representation of eigenvectors and working on some in-class problems, the solutions of which are available on the class' Moodle page.

Lecture 33: Eigenbases

November 5, 2021

Today, we're going to prove by induction the following theorem.

Theorem 37. If $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}$ are eigenvectors of an $n \times n$ matrix $\mathbf{A}$ with the corresponding distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$, then the set $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ is linearly independent.

Proof. Suppose that $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ are eigenvectors of $\mathbf{A}$ with corresponding eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_p\}$ such that $\lambda_i \neq \lambda_j$ when $i \neq j$. In the base case $p = 2$, we have $\{\mathbf{v_1}, \mathbf{v_2}\}$ as eigenvalues of $\mathbf{A}$ and $\{\lambda_1, \lambda_2\}$ such that $\lambda_1 \neq \lambda_2$. Suppose that

$$c_1\mathbf{v_1} + c_2\mathbf{v_2} = \mathbf{0}.$$

Multiplying both sides by $\mathbf{A}$ then yields

$$\begin{aligned}\mathbf{0} &= \mathbf{A}(c_1\mathbf{v_1} + c_2\mathbf{v_2}) \\ &= c_1\mathbf{A}\mathbf{v_1} + c_2\mathbf{A}\mathbf{v_2} \\ &= c_1\lambda_1\mathbf{v_1} + c_2\lambda_2\mathbf{v_2}.\end{aligned}$$

Subtracting this final expression from that obtained by multiplying the original dependence relation by λ_1 then gives

$$\begin{aligned}\mathbf{0} &= (\lambda_1 c_1\mathbf{v_1} + \lambda_1 c_2\mathbf{v_2}) - (c_1\lambda_1\mathbf{v_1} + c_2\lambda_2\mathbf{v_2}) \\ &= c_2(\lambda_1 - \lambda_2).\end{aligned}$$

Because $\lambda_1 \neq \lambda_2$, then this implies that $c_2 = 0$. Plugging this back into the dependence relation then yields

$$\mathbf{0} = c_1 v_1.$$

Therefore, $c_1 = 0$ as well. As $c_1 = c_2 = 0$, then $\mathbf{v_1}$ and $\mathbf{v_2}$ are linearly independent. Therefore, the base case holds. As our inductive hypothesis, suppose that the given statement holds in the pth case. Now, for the $(p+1)$st case, suppose that

$$\mathbf{0} = c_1\mathbf{v_1} + \ldots + c_p\mathbf{v_p} + c_{p+1}\mathbf{v_{p+1}}.$$

Multiplying both sides by $\mathbf{A}$ then yields

$$\mathbf{0} = c_1\lambda_1\mathbf{v_1} + \ldots + c_p\lambda_p\mathbf{v_p} + c_{p+1}\lambda_{p+1}\mathbf{v_{p+1}}.$$

Subtracting from this expression that formed by multiplying the original dependence relation by λ_{p+1} then gives

$$\mathbf{0} = c_1(\lambda_1 - \lambda_{p+1})\mathbf{v_1} + \ldots + c_p(\lambda_p - \lambda_{p+1}).$$

Using the induction hypothesis, we can then state that $c_i(\lambda_i - \lambda_{p+1}) = 0$ where $i = 1, 2, \ldots, p$. Since the eigenvalues $\lambda_1, \ldots, \lambda_p, \lambda_{p+1}$ are distinct, it then follows that $c_1 = \ldots = c_p = 0$. Plugging this back into the original dependence relation yields

$$c_{p+1}\lambda_{p+1}\mathbf{v_{p+1}} = \mathbf{0},$$

and, because $\lambda_{p+1} \neq 0$, then $c_{p+1} = 0$. Therefore, the $(p+1)$st case holds. As we have shown the base case $p = 2$ and that the truth of the $(p+1)$st case follows from the truth of the pth case, then, by principle of induction, the given statement is true for $p \geq 2$. This completes the proof. $\square$

> **Corollary 1.** If an $n \times n$ matrix $\mathbf{A}$ has n distinct eigenvalues, then $\mathbf{A}$ has a basis of eigenvectors for $\mathbb{R}^n$.

Example. Consider the stochastic matrix

$$\mathbf{M} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}.$$

In the past, we experimentally found that

$$\lim_{p\to\infty} \mathbf{M}^p \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.375 \\ 0.625 \end{bmatrix}.$$

Let's find out why. Using the characteristic polynomial of this matrix, we can find that it has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0.92$ with respective eigenvectors $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. By today's corollary, we can then state that

$$B = \{\begin{bmatrix} 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\}$$

forms an eigenbasis of $\mathbb{R}^2$. This means that we can express the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ as

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 3 \\ 5 \end{bmatrix} + \frac{5}{8}\begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

and the multiplication of both sides by $\mathbf{M}$ then gives

$$\begin{aligned} \mathbf{M}\begin{bmatrix} 1 \\ 0 \end{bmatrix} &= \frac{1}{8}\mathbf{M}\begin{bmatrix} 3 \\ 5 \end{bmatrix} + \frac{5}{8}\mathbf{M}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \frac{1}{8}(1)\begin{bmatrix} 3 \\ 5 \end{bmatrix} + \frac{5}{8}(0.92)\begin{bmatrix} 1 \\ -1 \end{bmatrix}. \end{aligned}$$

Hence, we can express the powers of the matrix $\mathbf{M}$ multiplied by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ as

$$\mathbf{M}^p \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{8}(1)^p \begin{bmatrix} 3 \\ 5 \end{bmatrix} + \frac{5}{8}(0.92)^p \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

$\diamond$

Lecture 34: Diagonalization

November 8, 2021

Last time, we proved that eigenvectors of an $n \times n$ matrix $\mathbf{A}$ with distinct eigenvalues are linearly independent. We further showed that if we have n independent eigenvectors, then we can use these vectors as a basis of $\mathbb{R}^n$ and express any other vector in terms of this eigenbasis, simplifying many calculations. Today, we'll look at diagonalization, or factoring a square matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{PDP}^{-1}$, where $\mathbf{D}$ is a diagonal matrix. This procedure, again, can simplify computations. To clarify, a diagonal matrix is defined as a matrix with zeros in every position other than the main diagonal. Such matrices exhibit commutativity and their powers can be found as

$$\mathbf{D}^k = \begin{bmatrix} *^k & 0 & \dots & 0 & 0 \\ 0 & *^k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & *^k & 0 \\ 0 & 0 & \dots & 0 & *^k \end{bmatrix}.$$

Definition 26. We say that an $n \times n$ matrix $\mathbf{A}$ is diagonalizable if $\mathbf{A}$ is similar to a diagonal matrix. I.e., there exists an invertible matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$ such that $\mathbf{A} = \mathbf{PDP}^{-1}$.

Let's consider an example.

Example. Let $\mathbf{A}$ be defined as

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

We can diagonalize this matrix as

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} -2 & 1 \\ -12 & 5 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ -3 & 1 \end{bmatrix} \\ &= \mathbf{PDP}^{-1}. \end{aligned}$$

Using this factorization, we can then calculate powers of $\mathbf{A}$ as

$$\begin{aligned} \mathbf{A}^k &= \mathbf{PD}^k\mathbf{P}^{-1} \\ &= \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1^k & 0 \\ 0 & 2^k \end{bmatrix} \begin{bmatrix} 4 & -1 \\ -3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 4 - 3 \times 2^k & -1 + 2^k \\ 12 - 12 \times 2^k & -3 + 4 \times 2^k \end{bmatrix}. \end{aligned}$$

◇

Theorem 38. Given an $n \times n$ matrix $\mathbf{A}$, a matrix $\mathbf{P}$ whose columns are eigenvectors of $\mathbf{A}$, and a diagonal matrix $\mathbf{D}$ with the corresponding eigenvalues of $\mathbf{A}$, then $\mathbf{AP} = \mathbf{PD}$.

We'll now complete some in-class problems whose solutions are available on the class' Moodle page. From these problems we can make the observation that given an $n \times n$ matrix $\mathbf{A}$ with eigenvectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$ and corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, we can calculate $\mathbf{AP}$ to be

$$\begin{aligned}
\mathbf{AP} &= \mathbf{A}\begin{bmatrix}\mathbf{v_1} & \ldots & \mathbf{v_2}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{Av_1} & \ldots & \mathbf{Av_n}\end{bmatrix} \\
&= \begin{bmatrix}\lambda_1\mathbf{v_1} & \ldots & \lambda_n\mathbf{v_n}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{v_1} & \ldots & \mathbf{v_n}\end{bmatrix}\begin{bmatrix}\lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n\end{bmatrix} \\
&= \mathbf{PD}.
\end{aligned}$$

From this arises the following theorem.

Theorem 39. If an $n \times n$ matrix $\mathbf{A}$ has n distinct eigenvalues, then $\mathbf{A}$ is diagonalizable.

So, how do we know if an $n \times n$ matrix $\mathbf{A}$ is diagonalizable if it has fewer than n distinct eigenvalues? We'll answer this question during the next class.

Lecture 35: More on Diagonalization & Applications

November 10, 2021

Last time, we defined a notion of an $n \times n$ matrix $\mathbf{A}$ being diagonalizable where an invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{D}$ exist such that $\mathbf{A} = \mathbf{PDP}^{-1}$. We also established the following two theorems.

Theorem 40. An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ has n linearly independent eigenvectors.

Theorem 41. If an $n \times n$ matrix $\mathbf{A}$ has n distinct eigenvalues, then $\mathbf{A}$ is diagonalizable.

From the latter theorem arises the question of whether an $n \times n$ matrix $\mathbf{A}$ with fewer than n distinct eigenvalues can be diagonalizable.

Definition 27. If $p(\lambda)$ is a characteristic polynomial for an $n \times n$ matrix $\mathbf{A}$, then the linear factors of this polynomial give rise to the eigenvalues of $\mathbf{A}$. Take for example the characteristic polynomial

$$\begin{aligned} p(\lambda) &= (\lambda - 5)(\lambda - 5)(\lambda - 3)(\lambda - 2) \\ &= (\lambda - 5)^2(\lambda - 3)(\lambda - 2). \end{aligned}$$

Because the linear factor $(\lambda - 5)$ appears twice, the eigenvalue $\lambda_1 = 5$ has **algebraic multiplicity** of 2. This can be denoted by $\mu_{\mathbf{A}}(\lambda_1) = 2$. Similarly, the other two eigenvalues have algebraic multiplicities of 1.

Definition 28. The **geometric multiplicity** of an eigenvalue λ_i of an $n \times n$ matrix $\mathbf{A}$, denoted by $\gamma_{\mathbf{A}}(\lambda_i)$, is defined as the dimension of its associated eigenspace, or $\gamma_{\mathbf{A}}(\lambda_i) = \dim(E_{\lambda_i})$.

Associated with these definitions are two theorems.

Theorem 42. Given an eigenvalue λ_i of an $n \times n$ matrix $\mathbf{A}$, then the geometric multiplicity of λ must be less than or equal to this eigenvalue's algebraic multiplicity, or

$$\mu_{\mathbf{A}}(\lambda_i) \geq \gamma_{\mathbf{A}}(\lambda_i).$$

Theorem 43. An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if the characteristic polynomial can be described in terms of linear factors and the algebraic and geometric multiplicities of the eigenvalues that arise from this operation are equal, or

$$\mu_{\mathbf{A}}(\lambda_i) = \gamma_{\mathbf{A}}(\lambda_i).$$

Example. Suppose that $\mathbf{A}$ is a 6×6 matrix with characteristic polynomial

$$p(\lambda) = (\lambda - 3)^2(\lambda - 5)^3(\lambda - 7),$$

and that

$$\begin{aligned} \mathrm{Dim}(E_3) &= 2 \\ \mathrm{Dim}(E_5) &= 2 \\ \mathrm{Dim}(E_7) &= 1. \end{aligned}$$

The matrix $\mathbf{A}$ is thus not diagonalizable because

$$\begin{aligned} \mu_{\mathbf{A}}(5) &\neq \gamma_{\mathbf{A}}(5) \\ 3 &\neq 2. \end{aligned}$$

$\diamond$

We'll now complete some in-class problems whose solutions are available on the class' Moodle page.

Finally, we'll examine the notion of a linear discrete dynamical system—specifically, a "predator-prey" system. Let c_k denote the number of coyotes (thousands) in month k and r_k denote the number of roadrunners (thousands) in month k. This system evolves over time such that

$$\begin{aligned} c_{k+1} &= 0.86c_k + 0.08r_k \\ r_{k+1} &= -0.12c_k + 1.14r_k. \end{aligned}$$

Now, let's define the state vector

$$\mathbf{x_k} = \begin{bmatrix} c_k \\ r_k \end{bmatrix} ,$$

so that we can now state

$$\begin{aligned} \mathbf{x_{k+1}} &= \mathbf{M}\mathbf{x_k} \\ &= \begin{bmatrix} 0.86 & 0.08 \\ -0.12 & 1.14 \end{bmatrix} \mathbf{x_k}. \end{aligned}$$

Now, we can ask how, given an initial population $\mathbf{x_0}$, this population will evolve in the long run (i.e., as $k \to \infty$). Let

$$\mathbf{x_0} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}.$$

We can calculate that $\mathbf{M}$ has eigenvalues $\lambda_1 = 1.1$ and $\lambda_2 = 0.9$ with respective eigenvectors $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Because these vectors form an eigenbasis of $\mathbb{R}^2$, we can now find that

$$\begin{aligned} \mathbf{M}\mathbf{x_0} &= \mathbf{M} \begin{bmatrix} 4 \\ 7 \end{bmatrix} \\ &= \mathbf{M}(2 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 1 \end{bmatrix}) \\ &= 2\mathbf{M} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1\mathbf{M} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ &= 2(1.1) \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1(0.9) \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \end{aligned}$$

Hence, we can now state that

$$\begin{aligned} \mathbf{x_{k+1}} &= \mathbf{M}^k \mathbf{x_0} \\ &= 2(1.1)^k \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1(0.9)^k \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \end{aligned}$$

Finally, this gives

$$\lim_{k\to\infty} \mathbf{M}^k \mathbf{x_0} \approx 2(1.1)^k \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

So, both populations will grow by about 10% per month with a ratio of roughly three roadrunners to one coyote in the long run.

Lecture 36: Similar Matrices & Diagonalization Practice

November 11, 2021

We begin this discussion session with a clarification of a few things about similar matrices. The concept of similar matrices comes from the change of basis. Consider a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, which can be expressed using the $n \times n$ matrix $\mathbf{A}$ as $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Under a new basis B, this transformation becomes $[T(\mathbf{x})]_B = \mathbf{A}_B[\mathbf{x}]_B$. As it turns out, $\mathbf{A}$ must be similar to $\mathbf{A}_B$; this is the aforementioned origin. Now, what are similar matrices' properties? Their shared features are as follows.

- Eigenvalues
- Characteristic Polynomial
- Determinant
- Trace

Here, "trace" denotes the sum of the elements on a matrix's main diagonal. We have shown the other features in class, but a proof of this trace characteristic is exceedingly similar to that for the determinant given by the following. Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ matrices and $\mathrm{Det}(\mathbf{AB}) = \mathrm{Det}(\mathbf{BA})$. We can thus state that

$$\begin{aligned} \det(\mathbf{PBP}^{-1}) &= \det(\mathbf{P}^{-1}\mathbf{PB}) \\ &= \det(\mathbf{B}). \end{aligned}$$

The final connection to question three of the homework assignment is left as an exercise to the reader. As for the second part of this question, note that trivially

$$\mathrm{tr}(\mathbf{D}) = \sum_{k=1}^{n} \lambda_k.$$

Hence, since $\mathbf{A}$ is similar to $\mathbf{D}$, we should be able to state something about the equation

$$\mathrm{tr}(\mathbf{A}) = \sum_{k=1}^{n} \lambda_k.$$

Again, the final connection here is left to the reader. Now, let's work on some discussion session problems.

1. Let $\mathbf{A} = \begin{bmatrix} 3/4 & 1/2 \\ 1/4 & 1/2 \end{bmatrix}$. This matrix has eigenvalues 1 and 0.25 with respective eigenvectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

   (a) The vector $\mathbf{x_0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ can be expressed as a linear combination of the eigenvectors of $\mathbf{A}$ as

$$\mathbf{x_0} = \frac{2}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

   (b) The long term behaviour of $\mathbf{A}^k\mathbf{x_0}$ is thus

$$\begin{aligned} \mathbf{A}^k\mathbf{x_0} &= \mathbf{A}^k(\frac{2}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ -1 \end{bmatrix}) \\ &= \frac{2}{3}\mathbf{A}^k\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{3}\mathbf{A}^k\begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \frac{2}{3}1^k\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{3}0.25^k\begin{bmatrix} 1 \\ -1 \end{bmatrix}. \end{aligned}$$

   And finally,

$$\lim_{k\to\infty} \mathbf{A}^k\mathbf{x_0} = \begin{bmatrix} 4/3 \\ 2/3 \end{bmatrix}.$$

   (c) We can diagonalize $\mathbf{A}$ using its eigenvectors and eigenvalues as

$$\begin{aligned} \mathbf{A} &= \mathbf{PDP}^{-1} \\ &= \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}. \end{aligned}$$

   And hence,

$$\mathbf{A}^k = \frac{1}{3}\begin{bmatrix} 2 + 0.25^k & 2 - 2(0.25)^k \\ 1 - 0.25^k & 1 + 2(0.25)^k \end{bmatrix}.$$

2. Let $\mathbf{A} = \begin{bmatrix} 2 & 1 & -1 & 1 \\ 0 & 1 & k & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

   (a) Where $k = 1$, the eigenspace E_1 resolves to

$$\text{Null}(\mathbf{A} - \mathbf{I}) = \text{Null}(\begin{bmatrix} 1 & 1 & -1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}),$$

which has 2 pivots and hence $\dim(E_1) = 2$. For the eigenspace E_2, we can similarly find that

$$\text{Null}(\mathbf{A} - 2\mathbf{I}) = \text{Null}(\begin{bmatrix} 0 & 1 & -1 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}),$$

which has 2 pivots and hence $\dim(E_2) = 2$.

(b) Where $k = 1$, because $\forall_{i \in \{1,2\}}\ \mu_{\mathbf{A}}(\lambda_i) = \gamma_{\mathbf{A}}(\lambda_i)$, $\mathbf{A}$ is diagonalizable.

(c) Where $k \neq 1$, the eigenspace E_2 resolves to

$$\text{Null}(\mathbf{A} - 2\mathbf{I}) = \text{Null}(\begin{bmatrix} 0 & 1 & -1 & 1 \\ 0 & -1 & k & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}),$$

which has three pivots and thus $\dim(E_2) = 3$.

(d) Where $k \neq 1$, because $\dim(E_2) = 3$, then $\gamma_{\mathbf{A}}(2) = 3$. Therefore, because $\mu_{\mathbf{A}}(2) = 2$, then $\mu_{\mathbf{A}}(2) \neq \gamma_{\mathbf{A}}(2)$ and thus $\mathbf{A}$ is not diagonalizable.

3. (a) A 2×2 matrix $\mathbf{A}$ with only one eigenvalue can be diagonalizable, by the case of

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

(b) The matrix $\begin{bmatrix} 5 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & 2 & 3 \end{bmatrix}$ is diagonalizable because it has three distinct eigenvalues: 5, 1, and 3.

(c) A 2×2 predator-prey matrix $\mathbf{M}$ with both eigenvalues less than 1 will obey $\lim\limits_{k \to \infty} \mathbf{M}^k \mathbf{x_0} = \mathbf{0}$ because each term, which necessarily contains the factor λ^k, will go to zero as k increases.

The three remaining true-false questions are left as an exercise to the reader.

Lecture 37: Introducing Orthogonality

November 12, 2021

Last time, we looked at the evolution of a system represented by the state vector

$$\mathbf{x_k} = \begin{bmatrix} c_k \\ r_k \end{bmatrix},$$

where this evolution was defined

$$\begin{aligned}\mathbf{x_{k+1}} &= \mathbf{M}\mathbf{x_k} \\ &= \begin{bmatrix} 0.86 & 0.08 \\ -0.12 & 1.14 \end{bmatrix} \mathbf{x_k}.\end{aligned}$$

This stochastic matrix has eigenvalues 1.1 and 0.9 with corresponding eigenvectors $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Using this and the fact that we can express the initial state vector in terms of these eigenvectors as

$$\mathbf{x_0} = d_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + d_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

we were then able to show that

$$\begin{aligned}\mathbf{M}^k \mathbf{x_0} &= d_1 \mathbf{M}^k \begin{bmatrix} 1 \\ 3 \end{bmatrix} + d_2 \mathbf{M}^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ &= d_1 (1.1)^k \begin{bmatrix} 1 \\ 3 \end{bmatrix} + d_2 (0.9)^k \begin{bmatrix} 2 \\ 1 \end{bmatrix},\end{aligned}$$

and finally that

$$\lim_{k \to \infty} \mathbf{M}^k \mathbf{x_0} = d_1 (1.1)^k \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

where d_1 and d_2 are both positive. We'll now observe a phase diagram of this system with various different initial states. Notably, only a narrow band of initial states creates the conditions for the above limit to be realized, with many others causing both populations to become negative. We'll then examine the phase diagram for the matrix

$$\mathbf{M} = \begin{bmatrix} 0.4 & 0.1 \\ -0.15 & 0.8 \end{bmatrix},$$

which, for all initial states, approaches the origin. In this case, $\mathbf{0}$ is called the "attractor" of the system. More about this and DDSs more generally can be found in section 5.6 of the Lay textbook. Now, let's examine the dot product and orthogonality.

Definition 29. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we define the dot product (or inner product, or scalar product) $\mathbf{x} \cdot \mathbf{y}$ as

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}.$$

More generally, if $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, then

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= x_1 y_1 + \ldots + x_n y_n \\ &= \sum_{i=1}^{n} x_i y_i. \end{aligned}$$

Definition 30. If $\mathbf{x} \in \mathbb{R}^n$, we define $\|\mathbf{x}\|$ (the **length of x** or magnitude, or norm) as

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}.$$

We say that $\mathbf{x} \in \mathbb{R}^n$ is a **unit vector** if $\|\mathbf{x}\| = 1$.

Theorem 44 (Properties of the Dot Product)**.** Given $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then the following properties are satisfied.

a. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ (Commutative)

b. $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = (\mathbf{x} \cdot \mathbf{z}) + (\mathbf{y} \cdot \mathbf{z})$ (Distributive)

c. $(c\mathbf{x}) \cdot \mathbf{y} = c(\mathbf{x} \cdot \mathbf{y})$

d. $\mathbf{x} \cdot \mathbf{x} = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

Definition 31. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we say that $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** if

$$\mathbf{x} \cdot \mathbf{y} = 0.$$

Theorem 45 (A Variation on the Pythagorean Theorem)**.** If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

if and only if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.

Proof. This is a proof of Theorem 45. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Suppose that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. Using the properties of the dot product outlined in Theorem 44, we can then find that

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})\\ &= \mathbf{x}\cdot\mathbf{x}+\mathbf{x}\cdot\mathbf{y}+\mathbf{y}\cdot\mathbf{x}+\mathbf{y}\cdot\mathbf{y}\\ &= \|\mathbf{x}\|^2+2\mathbf{x}\cdot\mathbf{y}+\|\mathbf{y}\|^2\\ &= \|\mathbf{x}\|^2+\|\mathbf{y}\|^2,\end{aligned}$$

the last step being justified by the orthogonality of the vectors. This completes the proof. □

We'll now finish up the lecture by working through some in-class problems whose solutions are available on the class' Moodle page.

Lecture 38: Orthogonal Projection

November 15, 2021

Today, we're going to continue our discussion of orthogonality. As a goal, consider a subspace W and a vector $x \notin W$; what is the vector $y \in W$ with the least distance from $\mathbf{x}$? To formalize, given a subspace W and a vector $\mathbf{x} \notin W$, we're looking for the orthogonal projection of $\mathbf{x}$ onto W. Recall the following definition for review.

Definition 32. The distance between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is $\|\mathbf{x}-\mathbf{y}\|$.

Let's also quickly review the pre-lecture video. In the video, we defined $W^\perp$ as the orthogonal complement of subspace W in $\mathbb{R}^n$, or

$$\{\mathbf{x}\in\mathbb{R}^n \mid \forall \mathbf{y}\in W,\ \mathbf{x}\cdot\mathbf{y}=0\}.$$

We also proved the following theorem.

Theorem 46. If $\{\mathbf{v_1},\ldots,\mathbf{v_p}\}$ is a set of orthogonal vectors and $\mathbf{v_i}\neq\mathbf{0}$, then this set is linearly independent.

Some properties of the orthogonal complement are as follows; these will be proved for homework.

Theorem 47. Given a subspace W of $\mathbb{R}^n$ and its orthogonal complement $W^\perp$, the following statements are true.

a. $W^\perp$ is a subspace of $\mathbb{R}^n$.

b. If $W=\operatorname{span}(\mathbf{v_1},\ldots,\mathbf{v_p})$, then $\mathbf{x}\in W^\perp$ if and only if $\forall i,\ \mathbf{x}\cdot\mathbf{v}_i=0$

We'll now work on some in-class problems whose solutions are available on the class' Moodle page. From this worksheet, we can conjecture the following theorem which will be proved for homework.

Theorem 48. If W is a subspace of $\mathbb{R}^n$, then

$$\dim(W) + \dim(W^\perp) = n.$$

Definition 33. If $B = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is a basis of W in $\mathbb{R}^n$ and B is an orthogonal set, we say that B is an **orthogonal basis** of W.

Theorem 49. If $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is an orthogonal basis for a subspace $W \subseteq \mathbb{R}^n$, then any vector $y \in W$ can be written as

$$\mathbf{y} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p},$$

where

$$c_i = \frac{\mathbf{y} \cdot \mathbf{v_i}}{\mathbf{v_i} \cdot \mathbf{v_i}}.$$

Proof. Let $B = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ be an orthogonal basis of a subspace $W \subseteq \mathbb{R}^n$. Since B is a basis of W, we can represent any vector $y \in W$ as a linear combination of the vectors in B, as

$$\mathbf{y} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}.$$

Taking the dot product of both sides of this equation with $\mathbf{v_1}$ then gives

$$\begin{aligned}
\mathbf{y} \cdot \mathbf{v_1} &= (c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}) \cdot \mathbf{v_1} \\
&= c_1\mathbf{v_1} \cdot \mathbf{v_1} + c_2\mathbf{v_2} \cdot \mathbf{v_1} + \ldots + c_p\mathbf{v_p} \cdot \mathbf{v_1} \\
&= c_1\mathbf{v_1} \cdot \mathbf{v_1}.
\end{aligned}$$

Therefore,

$$c_1 = \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}.$$

This same argument can be repeated for every vector $\mathbf{v_2}, \ldots, \mathbf{v_p}$, thus establishing that

$$c_i = \frac{\mathbf{y} \cdot \mathbf{v_i}}{\mathbf{v_i} \cdot \mathbf{v_i}}.$$

This completes the proof. □

Now let's finally address our opening question by considering the two-dimensional case. Consider a vector $\mathbf{u} \in \mathbb{R}^2_{\neq 0}$ and a vector $\mathbf{y} \in \mathbb{R}^2$. We can thus define a notion of orthogonal projection as $\text{proj}_\mathbf{u}\mathbf{y} = c\mathbf{u}$ such that $c \in \mathbb{R}$ and $\mathbf{y} - \text{proj}_\mathbf{u}\mathbf{y} \perp \mathbf{u}$.

We can then find an algebraic formula for c as

$$\begin{aligned}
\mathbf{y} - c\mathbf{u} &\perp \mathbf{u} \\
(\mathbf{y} - c\mathbf{u}) \cdot \mathbf{u} &= 0 \\
\mathbf{y} \cdot \mathbf{u} - c\mathbf{u} \cdot \mathbf{u} &= 0 \\
\mathbf{y} \cdot \mathbf{u} &= c\mathbf{u} \cdot \mathbf{u} \\
c &= \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}.
\end{aligned}$$

Therefore, we can define the orthogonal projection of a vector onto another vector in two dimensions as

$$\text{proj}_{\mathbf{u}}\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}.$$

We'll explore the higher-dimensional case during the next class.

Lecture 39: Orthogonal Decomposition

November 17, 2021

Today, we're going to discuss the orthogonal decomposition theorem. First, let's review a bit of the material that we covered last class. Last time, we showed that given an orthogonal basis $B = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ of a subspace $W \subseteq \mathbb{R}^3$ we could write any vector $\mathbf{y} \in W$ as a linear combination of these basis vectors using the form

$$\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} + \frac{\mathbf{y} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} + \ldots + \frac{\mathbf{y} \cdot \mathbf{v_p}}{\mathbf{v_p} \cdot \mathbf{v_p}}\mathbf{v_p}.$$

If B is an **orthonormal basis**, or an orthogonal basis with every vector normalized (having length one), then the formula resolves to

$$\mathbf{y} = (\mathbf{y} \cdot \mathbf{v_1})\mathbf{v_1} + (\mathbf{y} \cdot \mathbf{v_2})\mathbf{v_2} + \ldots + (\mathbf{y} \cdot \mathbf{v_p})\mathbf{v_p},$$

because the scalar product of a unit vector with itself is one. We also discussed the notion of the two-dimensional projection of a vector $\mathbf{y}$ onto another vector $\mathbf{u}$ as

$$\text{proj}_{\mathbf{u}}\mathbf{y} = (\frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}})\mathbf{y}.$$

Now, let's consider the higher-dimensional case: we can conceive of a projection of a vector $\mathbf{y}$ onto a subspace W as $\text{proj}_W\mathbf{y} = \hat{\mathbf{y}}$ such that $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ and $\mathbf{z} \in W^{\perp}$. From this intuition arises the following theorem.

Theorem 50 (The Orthogonal Decomposition Theorem)**.** Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$. Then, we can write $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ uniquely where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^{\perp}$. Given an orthogonal basis of W as $B = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$, then these vectors $\hat{\mathbf{y}}, \mathbf{z}$ resolve to

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} + \frac{\mathbf{y} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} + \ldots + \frac{\mathbf{y} \cdot \mathbf{v_p}}{\mathbf{v_p} \cdot \mathbf{v_p}}\mathbf{v_p},$$

and

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

Now, let's prove this theorem.

Proof. Let's first consider the case $W = \{\mathbf{0}\}$. Then, $W^{\perp} = \mathbb{R}^n$ and we can set $\hat{\mathbf{y}} = \mathbf{0} \in W$ and $\mathbf{z} = \mathbf{y} \in \mathbb{R}^n$ with $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, which gives $\mathbf{y} = \mathbf{0} + \mathbf{y}$. Therefore, this case holds. Now, suppose that $W \neq \{\mathbf{0}\}$. Let $B = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ be an orthogonal basis of W. If $\mathbf{y} \in \mathbb{R}^n$ then $\hat{\mathbf{y}} \in W$ such that

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} + \frac{\mathbf{y} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} + \ldots + \frac{\mathbf{y} \cdot \mathbf{v_p}}{\mathbf{v_p} \cdot \mathbf{v_p}}\mathbf{v_p}.$$

Then, we want to show that $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} \in W^{\perp}$ by showing that $\mathbf{z} \perp \mathbf{v_i}$ for all i. To do this, consider the vector $\mathbf{z}$ expressed as

$$\begin{aligned}\mathbf{z} &= \mathbf{y} - \hat{\mathbf{y}} \\ &= \mathbf{y} - \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} + \frac{\mathbf{y} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} + \ldots + \frac{\mathbf{y} \cdot \mathbf{v_p}}{\mathbf{v_p} \cdot \mathbf{v_p}}\mathbf{v_p}.\end{aligned}$$

Taking the scalar product of both sides with $\mathbf{v_1}$ then gives

$$\begin{aligned}\mathbf{z} \cdot \mathbf{v_1} &= (\mathbf{y} - \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} + \frac{\mathbf{y} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} + \ldots + \frac{\mathbf{y} \cdot \mathbf{v_p}}{\mathbf{v_p} \cdot \mathbf{v_p}}\mathbf{v_p}) \cdot \mathbf{v_1} \\ &= \mathbf{y} \cdot \mathbf{v_1} - \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} \cdot \mathbf{v_1} + \frac{\mathbf{y} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} \cdot \mathbf{v_1} + \ldots + \frac{\mathbf{y} \cdot \mathbf{v_p}}{\mathbf{v_p} \cdot \mathbf{v_p}}\mathbf{v_p} \cdot \mathbf{v_1} \\ &= \mathbf{y} \cdot \mathbf{v_1} - \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} \cdot \mathbf{v_1} \\ &= \mathbf{y} \cdot \mathbf{v_1} - \mathbf{y} \cdot \mathbf{v_1} \\ &= 0.\end{aligned}$$

Repeating this argument for all $\mathbf{v_i}$ thus shows that $\mathbf{z} \perp \mathbf{v_i}$ for all i. This proves the existence of vectors $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^{\perp}$ in the forms given such that $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$.

Lemma 1. Given a subspace W and its orthogonal compliment $W^\perp$, then

$$W \cap W^\perp = \{\mathbf{0}\}.$$

Proof. If $\mathbf{v} \in W$ and $\mathbf{v} \in W^\perp$, then

$$\mathbf{v} \cdot \mathbf{v} = 0,$$

which clearly implies that $\mathbf{v} = \mathbf{0}$. □

Now, let's prove the uniqueness of this orthogonal decomposition. Suppose that $\mathbf{y} \in \mathbb{R}^n$ can be expressed $\mathbf{y} = \hat{\mathbf{y}}_\mathbf{1} + \mathbf{z_1}$ such that $\hat{\mathbf{y}}_\mathbf{1} \in W$ and $\mathbf{z}_1 \in W^\perp$ and $\mathbf{y} = \hat{\mathbf{y}}_\mathbf{2} + \mathbf{z_2}$ such that $\hat{\mathbf{y}}_\mathbf{2} \in W$ and $\mathbf{z}_2 \in W^\perp$. Then we can state that

$$\begin{aligned}\hat{\mathbf{y}}_\mathbf{1} + \mathbf{z_1} &= \hat{\mathbf{y}}_\mathbf{2} + \mathbf{z_2}\\ \hat{\mathbf{y}}_\mathbf{1} - \hat{\mathbf{y}}_\mathbf{2} &= \mathbf{z_2} - \mathbf{z_1},\end{aligned}$$

where $\hat{\mathbf{y}}_\mathbf{1} - \hat{\mathbf{y}}_\mathbf{2} \in W$ and $\mathbf{z_2} - \mathbf{z_1} \in W^\perp$. So, by Lemma 1,

$$\begin{aligned}\mathbf{0} &= \hat{\mathbf{y}}_\mathbf{1} - \hat{\mathbf{y}}_\mathbf{2}\\ &= \mathbf{z_2} - \mathbf{z_1},\end{aligned}$$

and thus $\hat{\mathbf{y}}_\mathbf{1} = \hat{\mathbf{y}}_\mathbf{2}$ and $\mathbf{z_1} = \mathbf{z_2}$. Therefore, this orthogonal decomposition is unique. This completes the proof. □

Lecture 40: A Review of Ortho-things

November 18, 2021

Let's begin this discussion session with a review of some of the terms involving orthogonality.

- Orthogonal Vectors. These are vectors whose scalar products are zero; i.e., $\mathbf{u} \cdot \mathbf{v} = 0$.
- Orthogonal Set. This is a set of vectors which are all orthogonal to each other, or $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ such that $\mathbf{v_i} \cdot \mathbf{v_j} = 0 : i \neq j$.
- Orthogonal Basis. This is a basis $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ of a subspace W such that this set is orthogonal.
- Orthogonal Complement. The orthogonal complement of a subspace W is the set of all vectors that are orthogonal to all vectors in W, or $W^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \forall \mathbf{w} \in W,\ \mathbf{x} \cdot \mathbf{w} = 0\}$. By last class' lemma, we established that the only vector in the intersection of W and $W^\perp$ is $\mathbf{0}$, or $W \cap W^\perp = \{\mathbf{0}\}$. Also, for a subspace W in $\mathbb{R}^n$, we know that $\dim(W) + \dim(W^\perp) = n$.
- Orthogonal Decomposition. This describes the fact that we can represent any vector $y \in \mathbb{R}^n$ uniquely as $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$.

- Orthogonal Projection. The orthogonal projection of a vector $\mathbf{y} \in \mathbb{R}^n$ onto a subspace W, or $\text{proj}_W \mathbf{y}$, is the vector $\hat{\mathbf{y}} \in W$ as described above.
- Orthogonal Matrix. This is an $n \times n$ matrix $\mathbf{A}$ such that $\mathbf{A}^T\mathbf{A} = \mathbf{I}$.

We'll spend the rest of this discussion session working on some in-class problems whose solutions are available on the class' Moodle page.

Lecture 41: The Best Approximation Theorem & The Gram-Schmidt Process

November 19, 2021

Today, we're going to discuss the Best Approximation Theorem and the Gram-Schmidt Process for finding orthogonal bases. Recall that if W is a subspace of $\mathbb{R}^n$ with orthogonal basis $B = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$, then for some $\mathbf{y} \in \mathbb{R}^n$

$$\text{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} + \ldots + \frac{\mathbf{y} \cdot \mathbf{v_p}}{\mathbf{v_p} \cdot \mathbf{v_p}}\mathbf{v_p}.$$

Further consider an orthonormal basis of W denoted $B' = \{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$. In this case, the above formula simplifies to

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u_1})\mathbf{u_1} + \ldots + (\mathbf{y} \cdot \mathbf{u_p})\mathbf{u_p}.$$

Now, let the $n \times p$ matrix $\mathbf{U}$ be defined

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} & \ldots & \mathbf{u_p} \end{bmatrix}.$$

Then, we can easily show that for some $\mathbf{y} \in \mathbb{R}^n$ then

$$\begin{aligned} \mathbf{U}\mathbf{U}^T\mathbf{y} &= \begin{bmatrix} \mathbf{u_1} & \ldots & \mathbf{u_p} \end{bmatrix} \begin{bmatrix} \mathbf{u_1} \\ \vdots \\ \mathbf{u_p} \end{bmatrix} \mathbf{y} \\ &= \begin{bmatrix} \mathbf{u_1} & \ldots & \mathbf{u_p} \end{bmatrix} \begin{bmatrix} \mathbf{y} \cdot \mathbf{u_1} \\ \vdots \\ \mathbf{y} \cdot \mathbf{u_p} \end{bmatrix} \\ &= (\mathbf{y} \cdot \mathbf{u_1})\mathbf{u_1} + \ldots + (\mathbf{y} \cdot \mathbf{u_p})\mathbf{u_p} \\ &= \text{proj}_W \mathbf{y}. \end{aligned}$$

Hence, for such a matrix $\mathbf{U}$ whose columns are an orthonormal basis for a subspace W, we know that $\mathbf{U}\mathbf{U}^T\mathbf{y} = \text{proj}_W \mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}^n$. Notably, by the linearity of matrix multiplication, this tells us that the projection operation is a linear transformation. Further, can determine that this matrix $\mathbf{U}$ is orthogonal. An orthogonal matrix $\mathbf{A}$ is defined as a matrix such that $\mathbf{A}^T\mathbf{A} = \mathbf{I}$. For a matrix $\mathbf{U}$ constructed as described, this criterion resolves to

$$\begin{aligned}
\mathbf{U}^T\mathbf{U} &= \begin{bmatrix} \mathbf{u_1} \\ \vdots \\ \mathbf{u_p} \end{bmatrix} \begin{bmatrix} \mathbf{u_1} & \dots & \mathbf{u_p} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{u_1} \cdot \mathbf{u_1} & \mathbf{u_1} \cdot \mathbf{u_2} & \dots & \mathbf{u_1} \cdot \mathbf{u_{p-1}} & \mathbf{u_1} \cdot \mathbf{u_p} \\ \mathbf{u_2} \cdot \mathbf{u_1} & \mathbf{u_2} \cdot \mathbf{u_2} & \dots & \mathbf{u_2} \cdot \mathbf{u_{p-1}} & \mathbf{u_2} \cdot \mathbf{u_p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{u_{p-1}} \cdot \mathbf{u_1} & \mathbf{u_{p-1}} \cdot \mathbf{u_2} & \dots & \mathbf{u_{p-1}} \cdot \mathbf{u_{p-1}} & \mathbf{u_{p-1}} \cdot \mathbf{u_p} \\ \mathbf{u_p} \cdot \mathbf{u_1} & \mathbf{u_p} \cdot \mathbf{u_2} & \dots & \mathbf{u_p} \cdot \mathbf{u_{p-1}} & \mathbf{u_p} \cdot \mathbf{u_p} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \\
&= \mathbf{I}.
\end{aligned}$$

Therefore, any square matrix $\mathbf{U}$ whose columns are orthonormal is an orthogonal matrix. Now, we'll pivot a bit and discuss the Best Approximation Theorem.

Theorem 51 (Best Approximation Theorem)**.** If $\hat{\mathbf{y}}$ is the projection of $\mathbf{y}$ onto a subspace W, then for all $\mathbf{v} \in W$

$$\|\mathbf{y} - \mathbf{v}\| \geq \|\mathbf{y} - \hat{\mathbf{y}}\|.$$

Proof. To prove this theorem, we'll begin by noting that

$$\mathbf{y} - \mathbf{v} = \mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \mathbf{v}.$$

where $\mathbf{y} - \hat{\mathbf{y}} \in W^\perp$ and $\hat{\mathbf{y}} - \mathbf{v} \in W$. Now, recall that the Pythagorean theorem for $\mathbb{R}^n$ states that $\mathbf{a} \cdot \mathbf{b} = 0$ if and only if

$$\|\mathbf{a} + \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2.$$

As $\mathbf{y} - \hat{\mathbf{y}} \in W^\perp$ and $\hat{\mathbf{y}} - \mathbf{v} \in W$, then we can state that

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2.$$

Because $\|\hat{\mathbf{y}} - \mathbf{v}\| \geq 0$, then

$$\|\mathbf{y} - \mathbf{v}\|^2 \geq \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$

Taking the square-root of each side finally gives

$$\|\mathbf{y} - \mathbf{v}\| \geq \|\mathbf{y} - \hat{\mathbf{y}}\|.$$

This completes the proof. $\square$

Now, let's discuss the Gram-Schmidt process. This algorithm takes a non-orthogonal basis $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ of W as an input and outputs an orthogonal basis $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ of this same subspace. Let's begin by setting

$$\mathbf{u_1} = \mathbf{v_1}.$$

Then, we can find $\mathbf{u_2}$ as

$$\begin{aligned}\mathbf{u_2} &= \mathbf{v_2} - \mathrm{proj}_{\mathrm{span}(\mathbf{u_1})}(\mathbf{v_2})\\ &= \mathbf{v_2} - \mathrm{proj}_{W_1}(\mathbf{v_2}),\end{aligned}$$

where $W_p = \mathrm{span}(\mathbf{v_1}, \ldots, \mathbf{v_p}) = \mathrm{span}(\mathbf{u_1}, \ldots, \mathbf{u_p})$. By the orthogonal decomposition theorem, we know that $\mathbf{u_2} \in W_1^{\perp}$ and hence $\{\mathbf{u_1}, \mathbf{u_2}\}$ forms an orthogonal basis of W_2. Repeating this for the next dimension gives

$$\mathbf{u_3} = \mathbf{v_3} - \mathrm{proj}_{W_2}\mathbf{v_3},$$

where $\{\mathbf{u_1}, \mathbf{u_2}, \mathbf{u_3}\}$ forms an orthogonal basis of W_3. And, more generally,

$$\mathbf{u_i} = \mathbf{v_i} - \mathrm{proj}_{W_i}\mathbf{v_i},$$

where $\{\mathbf{u_1}, \ldots, \mathbf{u_i}\}$ forms an orthogonal basis of W_i. After computing all of these vectors $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$, we have thus formed an orthogonal basis for W. Normalizing each of these vectors $\mathbf{u_i}$ then gives an orthonormal basis for this same subspace. We'll now spend the rest of the lecture working on in-class problems whose solutions are available on the class' Moodle page.

Lecture 42: The Least-Squares Method

November 22, 2021

Last time, we talked about the Gram-Schmidt process for finding orthogonal bases and the Best Approximation Theorem, which acts as an additional interpretation for the projection of a vector into a subspace. Today, we'll discuss the least-squares technique. Consider an inconsistent linear system $\mathbf{Ax} = \mathbf{b}$; the least-squares technique allows us to find $\hat{\mathbf{b}}$, which is in the column space of $\mathbf{A}$ (and thus makes $\mathbf{Ax} = \hat{\mathbf{b}}$ consistent) and is as close as possible to $\mathbf{b}$. To find this vector $\hat{\mathbf{b}}$, we can simply utilize the Best Approximation Theorem to find that

$$\hat{\mathbf{b}} = \mathrm{proj}_{\mathrm{Col}(\mathbf{A})}\mathbf{b}.$$

Any vector $\hat{\mathbf{x}}$ which satisfies $\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}$ is a solution which minimizes

$$\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|,$$

hence the term "least-squares." In the pre-lecture video, we worked through an example of this method with a matrix $\mathbf{A}$ with orthogonal columns; this nicety significantly simplified our calculations. Now, we'll find a more general set of equations $\hat{\mathbf{x}}$ must satisfy to be a least-squares solution of $\mathbf{Ax} = \mathbf{b}$. From our initial discussion, it's easy to see that $\mathbf{b} - \hat{\mathbf{b}} \in \mathrm{Col}(\mathbf{A})$. Assuming that $\hat{\mathbf{x}}$ satisfies $\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}$, we can extend this fact into the series of logically equivalent statements

$$\begin{aligned}
\mathbf{b} - \hat{\mathbf{b}} \in \mathrm{Col}(\mathbf{A})^{\perp} &\Leftrightarrow (\mathbf{b} - \hat{\mathbf{b}}) \cdot \mathbf{a_i} = 0 \\
&\Leftrightarrow \mathbf{A}^T(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0} \\
&\Leftrightarrow \mathbf{A}^T\mathbf{b} - \mathbf{A}^T\hat{\mathbf{b}} = \mathbf{0} \\
&\Leftrightarrow \mathbf{A}^T\mathbf{b} = \mathbf{A}^T\hat{\mathbf{b}} \\
&\Leftrightarrow \mathbf{A}^T\mathbf{b} = \mathbf{A}^T\mathbf{A}\hat{\mathbf{x}}.
\end{aligned}$$

We term this final system as a transformation's **normal equations**. Let's consider an example.

Example. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}$$
$$\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

We can then calculate that

$$\begin{aligned}
\mathbf{A}^T\mathbf{A} &= \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},
\end{aligned}$$

and that

$$\begin{aligned}
\mathbf{A}^T\mathbf{b} &= \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 3 \end{bmatrix}.
\end{aligned}$$

Therefore, this system's normal equations are

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$$
$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

which nets the least-squares solution

$$\hat{\mathbf{x}} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}.$$

Now, suppose that the matrix $\mathbf{A}$ actually represented the points $(1,1),(0,2),(-1,0)$. If we wanted to find a line-of-best-fit for these data, we would be looking for a least-squares solution to the system

$$\begin{cases} 1m+b &= 1 \\ 0m+b &= 2 \\ -1m+b &= 0 \end{cases}.$$

This question is encapsulated by the system

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix},$$

which, as we have demostrated above, has the least-squares solution

$$\hat{\mathbf{x}} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}.$$

Hence, the line $y = 0.5x + 1$ fits the given data best, demonstrating a useful application of this method. ◇

We'll now work on some in-class problems whose solutions are available on the class' Moodle page. Let's finish with a theorem.

Theorem 52. Given an $m \times n$ matrix $\mathbf{A}$ and a vector $\mathbf{b}$ such that $\mathbf{Ax} = \mathbf{b}$ is inconsistent, the following statements are logically equivalent.

- This system's least-squares solution $\hat{\mathbf{x}}$ is unique.
- The matrix $\mathbf{A}^T\mathbf{A}$ is invertible.
- The columns of $\mathbf{A}$ are linearly independent.

Lecture 43: Another Look at Changing Bases

November 29, 2021

We'll begin today's lecture by reviewing last week's work on the least-squares method. The general idea for this method is to begin with an inconsistent system $\mathbf{Ax} = \mathbf{b}$, then define a least-squares solution $\hat{\mathbf{x}}$ of the system such that for all $\mathbf{x} \in \mathbb{R}^n$ then $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\| \leq \|\mathbf{Ax} - \mathbf{b}\|$. To find this $\hat{\mathbf{x}}$, notice that $\mathbf{b} \notin \mathrm{Col}(\mathbf{A})$; thus, we can define $\hat{\mathbf{b}} = \mathrm{proj}_{\mathrm{Col}(\mathbf{A})}\mathbf{b}$ so that $\mathbf{Ax} = \hat{\mathbf{b}}$ is consistent. Hence, a solution to the system $\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}$ gives a least-squares solution. We can then extend this concept into the following chain of logically equivalent statements.

$$\begin{aligned}
\mathbf{b} - \hat{\mathbf{b}} \in \mathrm{Col}(\mathbf{A})^{\perp} &\Leftrightarrow (\mathbf{b} - \hat{\mathbf{b}}) \cdot \mathbf{a_i} = 0 \\
&\Leftrightarrow \mathbf{A}^T(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0} \\
&\Leftrightarrow \mathbf{A}^T\mathbf{b} - \mathbf{A}^T\hat{\mathbf{b}} = \mathbf{0} \\
&\Leftrightarrow \mathbf{A}^T\mathbf{b} = \mathbf{A}^T\hat{\mathbf{b}} \\
&\Leftrightarrow \mathbf{A}^T\mathbf{b} = \mathbf{A}^T\mathbf{A}\hat{\mathbf{x}}.
\end{aligned}$$

This final statement is the so-called "normal equations" for the system $\mathbf{Ax} = \mathbf{b}$, which gives a simplified method for finding least-squares solutions $\hat{\mathbf{x}}$ to such a system. If $\mathbf{A}^T\mathbf{A}$ is invertible, we find an even simpler statement:

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

For homework, we'll prove the related theorem that states that the columns of $\mathbf{A}$ are linearly independent if and only if $\mathbf{A}^T\mathbf{A}$ is invertible. Now, let's work through an example.

Example. Let $\mathbf{Ax} = \mathbf{b}$ be an inconsistent system defined by

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \\
\mathbf{b} &= \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}.
\end{aligned}$$

Given that $\mathbf{A}^T\mathbf{A}$ is invertible, we can use Mathematica to find that

$$\begin{aligned}
\hat{\mathbf{x}} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \\
&= \begin{bmatrix} -5/3 \\ 3/2 \end{bmatrix}.
\end{aligned}$$

This is the least-squares solution to the given system. $\diamond$

Now, let's review some material about change-of-basis. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation. We know that there exists an $n \times n$ matrix $\mathbf{A}$ such that for all $\mathbf{x} \in \mathbb{R}^n$ then $T(\mathbf{x}) = \mathbf{Ax}$; this matrix $\mathbf{A}$ is termed the **standard matrix** of T and is defined

$$\mathbf{A} = \begin{bmatrix} T(\mathbf{e_1}) & \dots & T(\mathbf{e_n}) \end{bmatrix}.$$

Now, let $B = \{\mathbf{b_1}, \mathbf{b_2}, \dots, \mathbf{b_n}\}$ be another basis of $\mathbb{R}^n$. We would like to define the **matrix representation of** T **with respect to** B, denoted $[T]_B$, such that

$$[T(\mathbf{x})]_B = [T]_B[\mathbf{x}]_B.$$

To aid in this, recall that given a matrix $\mathbf{P}_B$ whose columns are basis vectors of $\mathbb{R}^n$, we know that

$$\mathbf{x} = \mathbf{P}_B[\mathbf{x}]_B,$$

and the function that takes $\mathbf{x}$ to $[\mathbf{x}]_B$ is an isomorphism. Now, we can state the following theorem.

Theorem 53. Given $B = \{\mathbf{b_1}, \dots, \mathbf{b_n}\}$, a basis of $\mathbb{R}^n$, and a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, then the B-matrix of T is defined as

$$[T]_B = \begin{bmatrix} [T(\mathbf{b_1})]_B & \dots & [T(\mathbf{b_n})]_B \end{bmatrix},$$

so that

$$[T(\mathbf{x})]_B = [T]_B[\mathbf{x}]_B.$$

Proof. Let $B = \{\mathbf{b_1}, \dots, \mathbf{b_n}\}$ be a basis of $\mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Suppose that $\mathbf{x} \in \mathbb{R}^n$ can be expressed

$$\mathbf{x} = c_1\mathbf{b_1} + \dots + c_n\mathbf{b_n}.$$

Then, by the linearity of T, we can state that

$$T(\mathbf{x}) = c_1T(\mathbf{b_1}) + \dots + c_nT(\mathbf{b_n}).$$

Next, we can take the coordinates of both sides with respect to B and, using the linearity of the coordinate transformation, find the following identity

$$\begin{aligned}
[T(\mathbf{x})]_B &= [c_1T(\mathbf{b_1}) + \dots + c_nT(\mathbf{b_n})]_B \\
&= c_1[T(\mathbf{b_1})]_B + \dots + c_n[T(\mathbf{b_n})]_B \\
&= \begin{bmatrix} [T(\mathbf{b_1})]_B & \dots & [T(\mathbf{b_n})]_B \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \\
&= [T]_B[\mathbf{x}]_B.
\end{aligned}$$

Therefore, it is quite easy to see that

$$[T]_B = \begin{bmatrix} [T(\mathbf{b_1})]_B & \dots & [T(\mathbf{b_n})]_B \end{bmatrix}.$$

This completes the proof. □

We'll spend the rest of this lecture working on some in-class problems whose solutions are available on the class' Moodle page.

Lecture 44: A Final Look at Bases & Symmetric Matrices

December 1, 2021

During the last lecture, we recalled that for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ we can find a **standard matrix A of** T such that for all $\mathbf{x} \in \mathbb{R}^n$ then $\mathbf{A}\mathbf{x} = T(\mathbf{x})$. We then built upon this by considering a basis $B = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ of $\mathbb{R}^n$ and defining a matrix

$$[T]_B = \begin{bmatrix} [T(\mathbf{b_1})]_B & \ldots & [T(\mathbf{b_n})]_B \end{bmatrix},$$

such that for all $\mathbf{x} \in \mathbb{R}^n$ then

$$[T]_B[\mathbf{x}]_B = [T(\mathbf{x})]_B.$$

Now, to introduce some novel material, we can define eigenthings for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$. Given such a transformation, then every non-zero vector $\mathbf{v} \in \mathbb{R}^n_{\neq \mathbf{0}}$ that satisfies $T(\mathbf{v}) = \lambda \mathbf{v}$ is an eigenvector of T with eigenvalue $\lambda \in \mathbb{R}$. To make $[T]_B$ as simple as possible, we should thus use a basis of eigenvectors of T (if one exists), because given that such a basis would satisfy $T(\mathbf{b_i}) = \lambda_i \mathbf{b_i}$ then

$$\begin{aligned} [T]_B &= \begin{bmatrix} [T(\mathbf{b_1})]_B & [T(\mathbf{b_2})]_B & \ldots & [T(\mathbf{b_n})]_B \end{bmatrix} \\ &= \begin{bmatrix} [\lambda_1 \mathbf{b_1}]_B & [\lambda_2 \mathbf{b_2}]_B & \ldots & [\lambda_n \mathbf{b_n}]_B \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 \mathbf{e_1} & \lambda_2 \mathbf{e_2} & \ldots & \lambda_n \mathbf{e_n} \end{bmatrix}, \end{aligned}$$

which is a diagonal matrix with the eigenvalues on the diagonal. So then, how are $\mathbf{A}$ and $[T]_B$ related? Consider the matrix $\mathbf{P}_B = \begin{bmatrix} \mathbf{b_1} & \ldots & \mathbf{b_n} \end{bmatrix}$ which satisfies $\mathbf{x} = \mathbf{P}_B[\mathbf{x}]_B$. Using this matrix and its inverse, along with $\mathbf{A}$, we can construct a relationship between $\mathbf{x}, [\mathbf{x}]_B, T(\mathbf{x}), [T(\mathbf{x})]_B$ as depicted in Figure. As this figure makes clear, the operation of moving from $[\mathbf{x}]_B$ to $[T(\mathbf{x})]_B$ can be enacted both by the standard matrices $[T]_B$ and $\mathbf{P}_B^{-1}\mathbf{A}\mathbf{P}_B$, so

$$[T]_B = \mathbf{P}_B^{-1}\mathbf{A}\mathbf{P}_B,$$

and

$$\mathbf{A} = \mathbf{P}_B[T]_B\mathbf{P}_B^{-1}.$$

Hence, $\mathbf{A}$ and $[T]_B$ are similar, demonstrating the defining characteristic of similar matrices. We'll now spend some time working on in-class problems whose solutions are available on the class' Moodle Page.

Now, let's focus on symmetric matrices; that is, $n \times n$ matrices $\mathbf{A}$ such that $\mathbf{A}^T = \mathbf{A}$.

> **Theorem 54.** If $\mathbf{A}$ is a symmetric matrix, then the eigenvectors of this matrix with distinct eigenvalues are orthogonal.

Proof. Let $\mathbf{A}$ be an $n \times n$ matrix that satisfies $\mathbf{A}^T = \mathbf{A}$. Let us first establish the following identity.

$$\begin{aligned}\mathbf{A}\mathbf{x} \cdot \mathbf{y} &= (\mathbf{A}\mathbf{x})^T \mathbf{y} \\ &= \mathbf{x}^T \mathbf{A}^T \mathbf{y} \\ &= \mathbf{x}^T \mathbf{A}\mathbf{y} \\ &= \mathbf{x} \cdot \mathbf{A}\mathbf{y}.\end{aligned}$$

Now, suppose that $\mathbf{A}\mathbf{x} = \lambda_1 \mathbf{x}$ and $\mathbf{A}\mathbf{y} = \lambda_2$ with $\lambda_1 \neq \lambda_2$. Then, we can show that

$$\begin{aligned}\mathbf{A}\mathbf{x} \cdot \mathbf{y} &= (\lambda_1 \mathbf{x}) \cdot \mathbf{y} \\ &= \lambda_1 (\mathbf{x} \cdot \mathbf{y}),\end{aligned}$$

and further

$$\begin{aligned}\mathbf{x} \cdot \mathbf{A}\mathbf{y} &= \mathbf{x} \cdot (\lambda_2 \mathbf{y}) \\ &= \lambda_2 (\mathbf{x} \cdot \mathbf{y}).\end{aligned}$$

Therefore, by the first identity shown, we can determine that

$$\begin{aligned}\lambda_1 (\mathbf{x} \cdot \mathbf{y}) &= \lambda_2 (\mathbf{x} \cdot \mathbf{y}) \\ 0 &= (\lambda_1 - \lambda_2)(\mathbf{x} \cdot \mathbf{y}).\end{aligned}$$

Since $\lambda_1 \neq \lambda_2$, then we must have $\mathbf{x} \cdot \mathbf{y} = 0$; thus, $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. This completes the proof. □

Definition 34. We say that a matrix $\mathbf{A}$ is **orthogonally diagonalizable** if there exists an orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$ such that

$$\begin{aligned}\mathbf{A} &= \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \\ &= \mathbf{P}\mathbf{D}\mathbf{P}^T.\end{aligned}$$

Theorem 55. A matrix $\mathbf{A}$ is orthogonally diagonalizable if and only if $\mathbf{A}$ is symmetric.

Lecture 45: A Review of Orthogonal Diagonalizability

December 2, 2021

During this discussion session, no new material was introduced; we discussed a selection of homework problems and completed an in-class worksheet whose solutions are available on the class' Moodle page.

Lecture 46: A Review of Complex Numbers

December 3, 2021

Today, we're going to finish our discussion of symmetric matrices and begin a review of complex numbers. Last time, we proved that any symmetric $n \times n$ matrix $\mathbf{A}$ whose eigenvectors have distinct eigenvalues is orthogonal. In particular, if such a matrix has n distinct eigenvalues, then it is orthogonally diagonalizable, or can be represented as $\mathbf{A} = \mathbf{PDP}^T$.

> **Theorem 56.** If a matrix $\mathbf{A}$ is orthogonally diagonalizable, then $\mathbf{A}$ is symmetric.

Proof. Suppose $\mathbf{A} = \mathbf{PDP}^T$ where $\mathbf{D}$ is diagonal and $\mathbf{P}$ is orthogonal. Then

$$\begin{aligned}\mathbf{A}^T &= (\mathbf{PDP}^T)^T \\ &= (\mathbf{P}^T)^T\mathbf{D}^T\mathbf{P}^T \\ &= \mathbf{PDP}^T \\ &= \mathbf{A}.\end{aligned}$$

Hence, $\mathbf{A} = \mathbf{A}^T$ and $\mathbf{A}$ is symmetric. This completes the proof. □

The converse of this theorem is also true, but this will not be proven in class. When fully generalized, this gives the following theorem.

> **Theorem 57.** An $n \times n$ matrix $\mathbf{A}$ is symmetric if and only if $\mathbf{A}$ is orthogonally diagonalizable.

Now, we can talk a bit about complex numbers. The set of all complex numbers is defined as

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}, i = \sqrt{1}\}.$$

As it turns out, everything that we've done in this class so far in $\mathbb{R}$ can also be done in $\mathbb{C}$, with the exception of a slight modification to the scalar product. Now, let's define two new operations on complex numbers: the conjugate and the modulus. The conjugate of a complex number $z = a + bi$ is defined as

$$\overline{z} = a - bi.$$

This can be visualized as a reflection of z across the real axis. The modulus of a complex number $z = a + bi$ is defined as

$$\begin{aligned}|z| &= \sqrt{z\overline{z}} \\ &= \sqrt{(a + bi)(a - bi)} \\ &= \sqrt{a^2 - b^2}.\end{aligned}$$

This can be visualized as the norm (or magnitude, or length) of the complex number. We'll now spend some time working on a set of in-class problems whose solutions are available on the class' Moodle page. From these problems, the following set of properties about the complex conjugate and modulus were developed. Given two complex numbers $z, w \in \mathbb{C}$,

- $\overline{w+z} = \overline{w} + \overline{z}$
- $\overline{wz} = \overline{wz}$
- $|wz|^2 = |w|^2|z|^2$
- $|z| \geq 0$
- $|wz| = |w||z|$.

Using these properties, we were able to prove the following theorem.

Theorem 58. If $\mathbf{A}$ is a real matrix and $\lambda = a + bi$ is a complex eigenvalue of $\mathbf{A}$, then $\overline{\lambda} = a - bi$ is also an eigenvalue of $\mathbf{A}$.

Lecture 47: Complex Eigenvalues

December 6, 2021

Today, we're going to talk a bit more about 2×2 real matrices with complex eigenvalues. To begin, consider the matrix

$$\mathbf{A} = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix},$$

with $\lambda = 0.8 \pm 0.6i$ and $|\lambda| = 1$. If we examine the evolution of a DDS described by this system, we will interestingly find that the evolution is periodical, tracing an oval around the origin, depending on the seed conditions. This is due to the modulus of the eigenvalues being one. If we consider the matrix

$$\mathbf{B} = \begin{bmatrix} 0.8 & 0.5 \\ -0.1 & 1.0 \end{bmatrix},$$

with $\lambda = 0.9 \pm 0.2i$ and $|\lambda| < 1$, and perform the same visualization, we will find that the evolution results in a spiral inwards towards the origin. We will attempt to uncover the mechanism behind this relationship during this lecture. Now, for some more about complex numbers. Given a complex number

$$z = a + bi,$$

we can write this number in 'polar form' by letting

$$\begin{aligned} a &= r\cos\phi \\ b &= r\sin\phi, \end{aligned}$$

where r is the modulus of z, or $r = |z| = \sqrt{a^2 + b^2}$, and ϕ is the 'argument' of z, $\arg(z)$, or the angle between z and the positive real axis. Using Euler's Formula,

$$e^{i\phi} = \cos\phi + i\sin\phi,$$

we can thus finally write

$$\begin{aligned} z &= a + bi \\ &= r\cos\phi + ir\sin\phi \\ &= re^{i\phi}. \end{aligned}$$

Last time, we showed that a 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

has eigenvalues $\lambda = a \pm bi$. When such a matrix has $|\lambda| = 1$, it is called a **rotation matrix**, and for such matrices

$$\begin{aligned} a &= \cos\phi \\ b &= \sin\phi. \end{aligned}$$

Recall that the formula for a matrix $\mathbf{B}$ which rotates a vector by ϕ radians anticlockwise is

$$\mathbf{B} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}.$$

More generally, if

$$\mathbf{A} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

we can let $r = |z|$ and say that

$$\mathbf{A} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} a/r & -b/r \\ b/r & a/r \end{bmatrix},$$

where the former matrix is a scaling matrix and the latter is a rotation matrix.

Theorem 59. If $\mathbf{A}$ is a 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$ then we can factor $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{PCP}^{-1},$$

where

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \\ \mathbf{P} &= \begin{bmatrix} \mathrm{Re}(\mathbf{v}) & \mathrm{Im}(\mathbf{v}) \end{bmatrix}, \end{aligned}$$

and $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda = a - bi$.

We'll spend the remainder of this lecture working on in-class problems whose solutions are available on the class' Moodle page.

Lecture 48: True-False Review for the Third Exam

December 8, 2021

We'll spend this lecture working on a set of true-false questions whose solutions are available on the class' Moodle page.

Lecture 49: Cancelled Discussion Session

December 9, 2021

This discussion session was cancelled in the interest of independent review for the third exam.

Lecture 50: Final Review for the Third Exam

December 10, 2021

During this final lecture, we discussed a selection of review questions to prepare for the third exam.