Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 1: Introduction to linear algebra

Lecturer: Alexander Paulin *January 17* *Aditya Sengupta*

Note: *LaTeXformat adapted from template courtesy of UC Berkeley EECS dept.*

1.1 What is linear algebra?

Linear algebra is difficult to describe succinctly; officially it's the "study of linear transformations on vector spaces". That doesn't mean anything when you compare it to something like calculus.

Linear algebra can be better defined as the mathematics that emerges when trying to solve systems of linear equations (linear systems). This leads to the question of what a linear system is. An example of a linear system would be something like:

Example 1

$$\text{Find all } x \text{ and } y \text{ s.t. } x - y = -1 \text{ and } 4x + 2y = 8.$$

This can be solved using a number of methods:

Method 1: solve first equation in y and substitute into the second equation.

The problem with this approach is it isn't easy to generalise to more equations or more variables/unknowns.

Method 2: combine equations to simplify

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

To eliminate the x, we can scale the first equation:

$$x - y = -1 \implies 4x - 4y = -4$$

With the scaled version of the first equation, we can subtract the first equation from the second:

$$2y - (-4y) = 8 - (-4) = 12 \implies 6y = 12 \implies y = 2$$

We can then substitute this back into the first equation,

$$x - y = -1 \implies x - 2 = -1 \implies x = 1$$

We can also think about this system of equations geometrically, by thinking about the equations as straight lines in the $x - y$ plane that intersect at the point $(1, 2)$. This shows us the other possibilities for linear systems: the lines could run parallel and not intersect, i.e. no solutions exist, or the lines could actually be the same, which would yield infinitely many solutions.

Example 2

Find x, y, z such that $x + 3y + 5z = 1$ and $x + y + 7z = 2$.

The first method in this case breaks down, as we will be left with one linear equation in two variables. We try the second method,

$$\begin{aligned} x + 3y + 5z &= 1 \\ x + y + 7z &= 2 \end{aligned}$$

We subtract the first from the second,

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

There is no way to proceed (e.g. to eliminate the y) without reintroducing the x and once again getting one equation in two variables. We can proceed by choosing a particular value of z. We get

$$y = z - \frac{1}{2}$$

Then, the first equation gives us

$$x = 1 - 3y - 5z = 1 - 3\Big(z - \frac{1}{2}\Big) - 5z = \frac{5}{2} - 8z$$

Therefore, the selection of a particular value of z gives us the corresponding value of x and y. This means the general solution is

$$\Big(\frac{5}{2} - 8z, z - \frac{1}{2}, z\Big)$$

Geometrically, this represents two planes intersecting in a line.

1.2 General situation

A linear equation in n variables (a fixed number of unknowns, $n > 0$) is usually represented in general as

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where a_i and b are constants.

A system of linear equations is a collection of these linear equations, which can be represented as follows:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}
$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 2: Gaussian elimination

Lecturer: Alexander Paulin *January 19* *Aditya Sengupta*

2.1 Solving linear systems, contd.

We can represent the linear system from last time in **matrix notation**, in which a matrix is a rectangular array of numbers.

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_n \end{array}\right]$$

This is called the *augmented matrix of a linear system.* The augmented matrix without the b_is is called the *coefficient matrix*, denoted A.

The strategy to solve the linear system in this representation is to combine the equations to systematically eliminate variables. The operations that are allowed for these combinations, in this representation, are:

1. Scale an equation or row by a nonzero number.
2. Add or subtract one row from another.
3. Rearrange rows.

None of these row operations will change the solutions to the linear system, because all of them are reversible.

2.2 Echelon Form

Definition 1. *A matrix is in echelon form if*

1. *all non-zero rows (those with at least one non-zero element) are above zero rows*
2. *each non-zero leading entry of a row is to the left of any non-zero leading entries of lower rows.*

For example,

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

is a matrix in echelon form.

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

is also in echelon form.

Definition 2. *A matrix is in reduced echelon form (REF) if*

3. *non-zero leading entries are 1*
4. *there are zeros above all leading non-zero entries.*

For example,

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

is in reduced echelon form.

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

is also in REF.

Theorem 2.1. *Every matrix can be put into unique REF using row operations.*

Example

$$\begin{bmatrix} 0 & 0 & 1 & 4 & -1 & 1 \\ 2 & -2 & 4 & 6 & 3 & 7 \\ 1 & -1 & 1 & -1 & 2 & 1 \end{bmatrix}$$

The first step in the row reduction is to get a non-zero term in the top-left corner, by $R_1 \iff R_3$:

$$\begin{bmatrix} 1 & -1 & 1 & -1 & 2 & 1 \\ 2 & -2 & 4 & 6 & 3 & 7 \\ 0 & 0 & 1 & 4 & -1 & 1 \end{bmatrix}$$

Now, we eliminate the 2 by scaling the first row by 2 and subtracting: $R_2 \to R_2 - 2R_1$.

$$\begin{bmatrix} 1 & -1 & 1 & -1 & 2 & 1 \\ 0 & 0 & 2 & 8 & -1 & 5 \\ 0 & 0 & 1 & 4 & -1 & 1 \end{bmatrix}$$

We move to the third column, which is the next one not in echelon form. First, we do $R_2 \iff R_3$:

$$\begin{bmatrix} 1 & -1 & 1 & -1 & 2 & 1 \\ 0 & 0 & 1 & 4 & -1 & 1 \\ 0 & 0 & 2 & 8 & -1 & 5 \end{bmatrix}$$

To eliminate the 2 in the bottom row, we scale the second row by 2 and subtract: $R_3 \to R_3 - 2R_2$

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

This matrix is now in EF. We can continue to get to REF, by finding a method to change all the elements above each row's leading element to zeros.

$R_2 \to R_2 + R_3$
$R_1 \to R_1 - 2R_3$

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

Finally, to change the element above the second row's leading element, $R_1 \to R_1 - R_2$:

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

This is now in reduced echelon form.

2.3 Calculating General Solutions to Linear Systems

Let the augmented matrix representing a linear system be in echelon form.

Definition 3. *A pivot column is a column with a non-zero leading entry.*

Definition 4. *A pivot position is the location of a non-zero leading entry.*

Definition 5. *A free column is a non-pivot column which is in the coefficient matrix.*

For example,

$$\left[\begin{array}{ccccc|c} 1 & -1 & 0 & -5 & 0 & -9 \\ 0 & 0 & 1 & 4 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{array}\right]$$

Columns 1, 3, and 5 are pivot columns. Columns 2 and 4 are free columns. Column 6 is not a pivot column but it cannot be a free column as it is not in the coefficient matrix.

There are three possibilities:

1. The last column of the augmented matrix is a pivot column. This means the linear system is inconsistent, i.e. no solutions exist.

2. The last column of the augmented matrix is not a pivot, and there are no free columns. This happens if and only if there is a single unique solution. For example,

$$\left[\begin{array}{cc|c} 1 & -1 & -1 \\ 0 & 6 & 12 \end{array}\right]$$

This corresponds to linear equations $x_1 - x_2 = -1$ and $6x_2 = 12$, which implies $x_1 = 1, x_2 = 2$.

3. The last column of the augmented matrix is not a pivot, and there are free columns. This corresponds to infinitely many solutions. Here, we can write pivot column variables in terms of free column variables. For example,

$$\left[\begin{array}{ccccc|c} 1 & -1 & 0 & -5 & 0 & -9 \\ 0 & 0 & 1 & 4 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{array}\right]$$

This corresponds to linear equations $x_1 - x_2 - 5x_4 = -9$, $x_3 + 4x_4 = 4$, and $x_5 = 3$, which gives us a general solution:

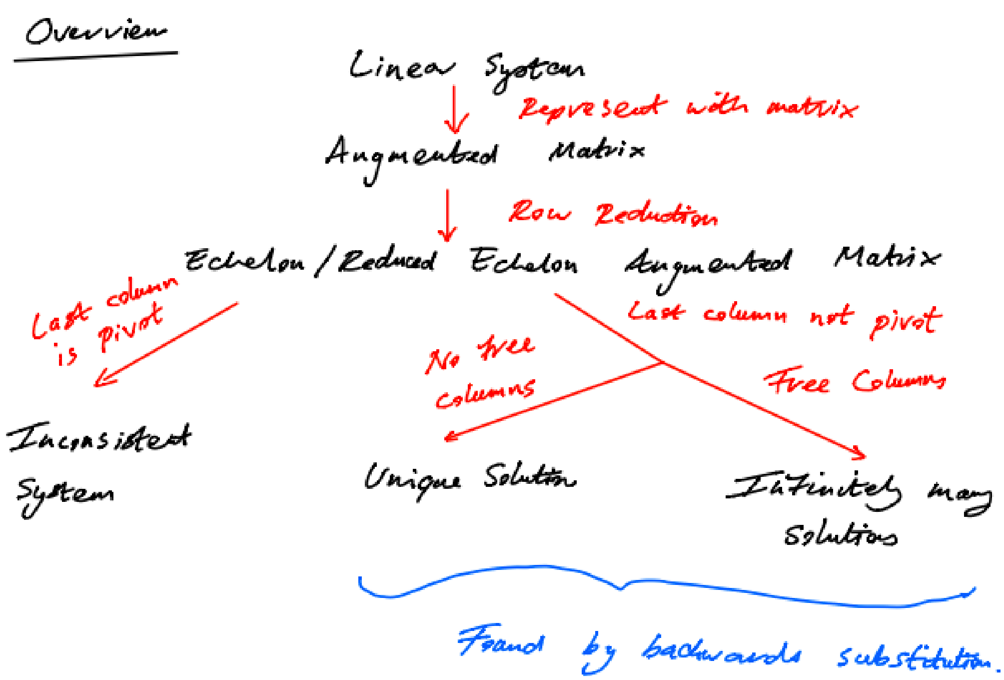
$$\begin{gathered} x_1 = -9 + x_2 + 5x_4 \\ x_2 \text{ is free} \\ x_3 = 4 - 4x_4 \\ x_4 \text{ is free} \\ x_5 = 3 \end{gathered}$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 3: Vectors, Addition and Scaling

Lecturer: Alexander Paulin *January 22* *Aditya Sengupta*

3.1 Overview



3.2 Vectors

Now that we know how to solve linear systems in general, we wish to find a conceptual way of understanding how linear systems work. This can be found in vectors in $\mathbb{R}^n$, since we can draw the analogy that both involve working with ordered collections of numbers.

If $\mathbb{R}$ is the set of real numbers, and n is the set of natural numbers $(1, 2, 3, ...)$, it follows that

$$\mathbb{R}^n := \text{ the set of ordered n-tuples of real numbers}$$

The solution to any linear system in which the n variables correspond to $x_1, x_2, \ldots, x_n$ can therefore be represented as a vector in $\mathbb{R}^n$. For this course, elements of $\mathbb{R}^n$ will always be written as columns.

$$\mathbb{R}^n = \text{set of all } \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ where } x_1, x_2, \ldots, x_n \in \mathbb{R}.$$

We call an element of $\mathbb{R}^n$ a **vector in** $\mathbb{R}^n$. We can look at specific values of n,

$$\mathbb{R}^1 = \mathbb{R} = \text{number line}$$
$$\mathbb{R}^2 = \text{plane } (x = x_1, y = x_2)$$
$$\mathbb{R}^3 = \text{3D space } (x = x_1, y = x_2, z = x_3)$$

Our geometric intuition breaks down for $n > 3$.

Notation:

$$\underline{\mathrm{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ in } \mathbb{R}^n$$

We aren't using the $\vec{x}$ notation because that relies on the idea of an arrow going between the origin and a point, which doesn't work at high dimensions.

3.3 Important Observations about Vectors

- We can add vectors in $\mathbb{R}^n$
- We can scale vectors in $\mathbb{R}^n$

Definition 6. *If* $\underline{x}$ =

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

and $\underline{y}$ =

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

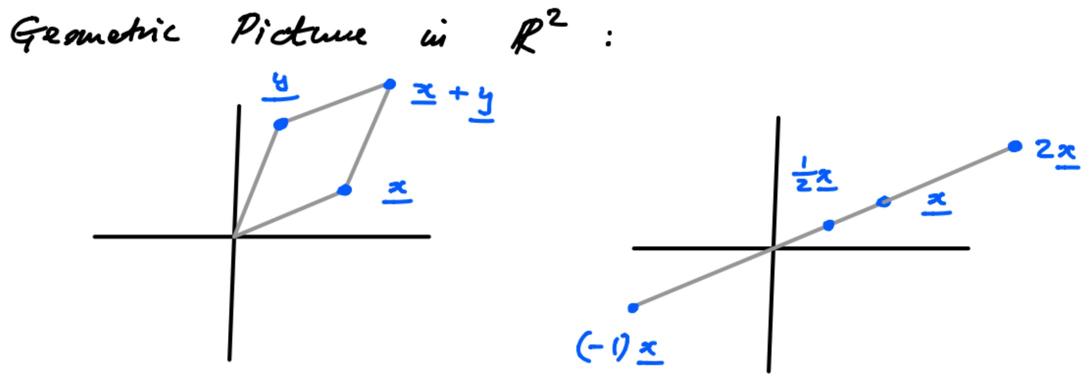
then we define

$$\underline{x} + \underline{y} := \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

and

$$\lambda \underline{x} := \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix}$$

The geometric picture of this in $\mathbb{R}^2$ is below:



Addition and scalar multiplication of vectors in $\mathbb{R}^n$ satisfy familiar rules of arithmetic, e.g.

$$(\underline{u} + \underline{v}) + \underline{w} = \underline{u} + (\underline{v} + \underline{w})$$

$$\underline{0} + \underline{v} = \underline{v}$$

$$\lambda(\underline{u} + \underline{v}) = \lambda\underline{u} + \lambda\underline{v}$$

$$0\underline{v} = \underline{0}$$

Definition 7. *Let $\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}$ be vectors in $\mathbb{R}^n$. For constant values $\lambda_1, \ldots, \lambda_k$,*

$$\lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} + \cdots + \lambda_k \underline{v_k}$$

is a linear combination of vectors $\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}$.

Definition 8. *The span of $\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}$ is the set of all linear combinations of $\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}$.*

It is intuitive that $\mathrm{span}(\underline{v_1}, \ldots, \underline{v_k})$ is everywhere we can get to in $\mathbb{R}^n$ by travelling only in directions $\underline{v_1}, \ldots, \underline{v_k}$.

3.3.1 Examples

1:

$$\underline{v_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ in } \mathbb{R}^2$$

The span of $\underline{v_1}$ is the set of all $\lambda \underline{v_1}$ for $\lambda \in \mathbb{R}$.

2:

$$\underline{v_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\underline{v_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Both are vectors in $\mathbb{R}^3$. We make an arbitrary linear combination,

$$\lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ 0 \end{bmatrix}$$

therefore the span is an xy-plane in 3D space.

3:

$$\underline{v_1} = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$$

$$\underline{v_2} = \begin{bmatrix} -1 \\ \frac{1}{2} \\ -2 \end{bmatrix}$$

Notice that $\underline{v_1} = -2\underline{v_2}$, therefore when we make the linear combination, we get:

$$\lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} = (-2\lambda_1 + \lambda_2)\underline{v_2}$$

Therefore the span of both vectors is only the span of $\underline{v_2}$, which is a straight line rather than a plane.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 4: Existence of Linear System Solutions

Lecturer: Alexander Paulin *January 24* *Aditya Sengupta*

4.1 Span of sets of vectors

The span of a single vector in $\mathbb{R}^2$ is the straight line representing the set of all points that can be reached by moving in that direction (positive or negative) by any amount.

The span of two linearly independent vectors in $\mathbb{R}^2$ is $\mathbb{R}^2$ itself, as we can reach any point in the plane as a combination of these two vectors.

The span of two linearly dependent vectors (in which $\underline{v_2}$ is a multiple of $\underline{v_1}$) is the same as that of a single vector, i.e. a straight line. With the addition of the second vector, we cannot reach any new points.

There are only three possibilities for the span of vectors in $\mathbb{R}^2$, which are the entire space $\mathbb{R}^2$, a straight line, or a single point at the origin (where the vector is a zero vector). In $\mathbb{R}^3$, this extends to four possibilities: the entire space, a plane, a line, and the origin.

4.2 Relating span to linear systems

Example: Can we write $\begin{bmatrix}1\\4\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\3\end{bmatrix}$ and $\begin{bmatrix}1\\2\end{bmatrix}$ in $\mathbb{R}^2$?

We need to find x_1, x_2 such that

$$x_1 \begin{bmatrix}1\\3\end{bmatrix} + x_2 \begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}1\\4\end{bmatrix}$$

This is true if and only if

$$x_1 + x_2 = 1, 3x_1 + 2x_2 = 4$$

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

Through a linear transformation, we bring this to echelon form:

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

$$x_2 = -1 \implies x_1 = 2$$

Therefore

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

4.3 General case of span relating to a linear system

Let a linear system have the following augmented matrix:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_n \end{array}\right]$$

which we can simplify into a series of vectors in $\mathbb{R}^m$:

$$(\underline{a_1}, \underline{a_2}, \dots, \underline{a_m} | \underline{b})$$

Let the solution to $(A|\underline{b})$ be the vector $\underline{x} = \begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$

Then, this yields the linear equations

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

which can be rearranged:

$$x_1\begin{bmatrix}a_{11}\\a_{21}\\\vdots\\a_{m1}\end{bmatrix} + x_2\begin{bmatrix}a_{12}\\a_{22}\\\vdots\\a_{m2}\end{bmatrix} + \cdots + x_n\begin{bmatrix}a_{1n}\\a_{2n}\\\vdots\\a_{mn}\end{bmatrix} = \begin{bmatrix}b_1\\b_2\\\vdots\\b_m\end{bmatrix}$$

i.e. the $\underline{b}$ vector is in the span of $\underline{a_1}, \dots, \underline{a_n}$.

In matrix equation notation, if $A = (\underline{a_1}, \dots, \underline{a_n})$ where $\underline{a_i}$ is a vector in $\mathbb{R}^m$, and $\underline{x} = \begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}$ in $\mathbb{R}^n$, then

$$A\underline{x} := x_1\underline{a_1} + x_2\underline{a_2} + \cdots + x_n\underline{a_n}$$

4.4 Conclusion

The linear system $(A|\underline{b})$ admits a solution $\iff$ $\underline{b}$ in $\text{Span}(\underline{a_1}, \dots, \underline{a_n})$ $\iff$ last column of reduced $(A|\underline{b})$ not a pivot.

4.5 Example

Is $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ in $\text{Span}\left(\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right)$?

This can be expressed as the augmented matrix

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

Start by scaling the first row by -1,

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

Then, we eliminate the first-column coefficients from the second and third rows,

$$\left[\begin{array}{ccc|c} -1 & 2 & 1 & 1 \\ 0 & 3 & 3 & 3 \\ 0 & 4 & 4 & 5 \end{array}\right]$$

We scale the second row and subtract 4 times the second row from the third,

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

The last column is a pivot, therefore the system is inconsistent. This means $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is not in $\text{Span}\left(\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right)$

A linear system $(A|\underline{b})$ has a solution for any b $\iff$ $\text{Span}(\underline{a_1}, \dots, \underline{(}a_n)) = \mathbb{R}^m$ $\iff$ the last column of the reduced $(A|\underline{b})$ is never a pivot for any $\underline{\text{b}}$.

This suggests the reduced A has no zero rows.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 5: Vectors and Uniqueness of Linear System Solutions

Lecturer: Alexander Paulin *26 January* *Aditya Sengupta*

We previously covered how to find whether a linear system $(A|\underline{b})$ has a solution for any $\underline{b}$. **What properties of $\underline{a_1}$, . . . , $\underline{a_n}$, $\underline{\mathbf{b}}$ determine whether the solution is unique?**

5.1 Homogeneous case

We approach this by examining a related problem, the *homogeneous linear system* in which $A\underline{x} = \underline{0}$.

$$A\underline{x} = \underline{0} \iff x_1\underline{a_1} + x_2\underline{a_2} + \cdots + x_n\underline{a_n} = \underline{0}$$

Note that this is always consistent as the trivial solution $\underline{x} = \underline{0}$ exists.

Definition 9. *Let $\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}$ be vectors in $\mathbb{R}^m$. We say that $\{\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}\}$ is* <u>*linearly independent*</u> *if*

$$\lambda_1\underline{v_1} + \lambda_2\underline{v_2} + \cdots + \lambda_k\underline{v_k} = \underline{0} \implies \lambda_i = 0 \forall i \in [0, k]$$

i.e. there is no linear combination of the vectors that sums to zero. If this is not the case, we say $\{\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}\}$ is <u>*linearly dependent*</u>*.*

Take $\{\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}\}$ linearly dependent. This implies that there exist $\lambda_1, \ldots, \lambda_k$ not all zero such that

$$\lambda_1\underline{v_1} + \cdots + \lambda_k\underline{v_k} = \underline{0}$$

.

For example, take $\lambda_1 \neq 0$. Then,

$$\underline{v_1} + \frac{\lambda_2}{\lambda_1}\underline{v_2} + \cdots + \frac{\lambda_k}{\lambda_1}\underline{v_k} = \underline{0}$$

$$\underline{v_1} = \frac{-\lambda_2}{\lambda_1}\underline{v_2} + \cdots + \frac{-\lambda_k}{\lambda_1}\underline{v_k}$$

which is a linear combination of $\{\underline{v_2}, \ldots, \underline{v_k}\}$. Therefore $\underline{v_1}$ is in the span of the other vectors.

Example

Take the vectors

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

They are linearly dependent because

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

Therefore $\begin{bmatrix}1\\2\\3\end{bmatrix}$ is in $\text{Span}\left(\begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}\right)$.

The intuition here is that if $\{\underline{v_1}, \underline{v_2}, \ldots, \underline{v_k}\}$ are linearly independent, then none of them are in the span of the others, i.e. they are all in totally independent directions.

Example

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

These are linearly independent as all of them are going in independent directions.

Conclusion

$A\underline{x} = \underline{0}$ has unique solution (i.e. $\underline{x} = \underline{0}$) $\iff$ $\underline{a_1}, \ldots, \underline{a_n}$ linearly independent $\iff$ reduced $(A|\underline{0})$ has no free columns.

5.2 The non-homogeneous case

Here, we have $A\underline{x} = \underline{b}$ where $\underline{b} \neq 0$. Let $\underline{v_p}$ be a particular solution i.e. $A\underline{v_p} = \underline{b}$. Then

$A\underline{v} = \underline{b} \implies A(\underline{v} - \underline{v_p}) = A\underline{v} - A\underline{v_p} = \underline{b} - \underline{b} = 0$. Therefore we can, in general, write the general solution as $\underline{v_p} + \underline{v_h}$ where $\underline{v_h}$ is a general solution to the homogeneous case.

We can conclude that $A\underline{x} = \underline{b}$ has a unique solution $\iff$ it has a solution **and** $A\underline{x} = \underline{0}$ has a unique solution.

$A\underline{x} = \underline{b}$ has a unique solution $\iff$ $\underline{b}$ in $\text{Span}(\underline{a_1}, \ldots, \underline{a_n})$ **and** $\{\underline{a_1}, \ldots, \underline{a_n}\}$ linearly independent. The last column of the reduced $(A|\underline{b})$ should not be a pivot and it should have no free columns.

Example

$$\left[\begin{array}{ccc|c}1 & 3 & 5 & 1\\0 & -2 & 2 & 1\end{array}\right]$$

We note that the last column is not a pivot, but the third column is free and is not a pivot. So we expect infinitely manyy solutions. We rearrange linear equations yielded from this matrix,

$x_2 = x_3 - \frac{1}{2}$ $x_1 = \frac{5}{2} - 8x_3$

Therefore the general solution is $\begin{bmatrix}\frac{5}{2} - 8x_3\\x_3 - \frac{1}{2}\\x_3\end{bmatrix} = \begin{bmatrix}\frac{5}{2}\\-\frac{1}{2}\\0\end{bmatrix} + x_3\begin{bmatrix}-8\\1\\1\end{bmatrix} = \underline{v_p} + \underline{v_h}$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 6: Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

Lecturer: Alexander Paulin *29 January* *Aditya Sengupta*

6.1 Linear transformation

Consider an $m \times n$ matrix A, which consists of n columns, each of which is a vector in $\mathbb{R}^m$: $\underline{a_1}, \dots, \underline{a_n}$. We can also consider this as the set of solutions to linear systems with coefficient matrix A, a vector in $\mathbb{R}^n$:

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

We can combine these two interpretations to see the linear combination we saw before:

$$A\underline{x} := x_1\underline{a_1} + x_2\underline{a_2} + \cdots + x_n\underline{a_n}$$

This gives us a link between the $\mathbb{R}^n$ interpretation and the $\mathbb{R}^m$ one. We can build a function to formalise this link.

Summary: Given A, an $m \times n$ matrix, we can construct a <u>function</u> $T_A : \mathbb{R}^n \to \mathbb{R}^m$. We call T_A the *linear transformation* associated to A.

6.2 Example of a linear transformation

Let $A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$. Then from the definition, $T_A : \mathbb{R}^3 \to \mathbb{R}^2$.

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

For example,

$$T_A\left(\begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} + 3 \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 13 \\ 16 \end{bmatrix}$$

6.3 Properties of T_A

1. $A(\underline{u} + \underline{v})$ for all $\underline{u}, \underline{v} \in \mathbb{R}^n \implies T_A(\underline{u} + \underline{v}) = T_A(\underline{u}) + T_A(\underline{v})$ This is interesting because it suggests a sum of vectors in $\mathbb{R}^n$ is equivalent to one in $\mathbb{R}^m$.

2. $A(\lambda\underline{u}) = \lambda(A\underline{u})$ for all $\underline{u} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. This suggests the above property holds for scalar multiplication too.

This is non-trivial. Most functions will not preserve addition and scalar multiplication like this.

E.g. $f : \mathbb{R}^2 \to \mathbb{R}^2$ is characterised by $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to \begin{bmatrix} x_1^2 + 1 \\ x_1 + x_2 \end{bmatrix}$.

Try

$$f\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}\Big) = f\Big(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\Big) = \begin{bmatrix} 10 \\ 5 \end{bmatrix}$$

and contrast with

$$f\Big(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Big) + f\Big(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\Big) = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$$

So addition is not preserved, as it is with linear transformations.

Definition 10. *We say a function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if*

1. $T(\underline{u} + \underline{v}) = T(\underline{u}) + T(\underline{v})$
2. $T(\lambda\underline{u}) = \lambda T(\underline{u})$

for all $\underline{u}, \underline{v} \in \mathbb{R}^n, \lambda \in \mathbb{R}$

6.4 Finding matrices corresponding to linear transformations

Given a function $T : \mathbb{R}^n \to \mathbb{R}^m$ linear, can we find a matrix A such that $T = T_A$?

Definition 11. $\underline{e_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \underline{e_2} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \underline{e_n} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$

Given a general $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, we can say

$$\underline{x} = x_1\underline{e_1} + x_2\underline{e_2} + \ldots x_n\underline{e_n}$$

This implies

$$T(\underline{x}) = T(x_1\underline{e_1} + x_2\underline{e_2} + \cdots + x_n\underline{e_n}) = T(x_1\underline{e_1}) + T(x_2\underline{e_2}) + \cdots + T(x_n\underline{e_n})$$

$$= x_1 T(\underline{e_1}) + x_2 T(\underline{e_2}) + \cdots + x_n T(\underline{e_n})$$

$$= \left(T(\underline{e_1}), \ldots, T(\underline{e_n}) \right) \underline{x}$$

which is an $m \times n$ matrix.

Definition 12. *Given $T : \mathbb{R}^n \to \mathbb{R}^m$ linear, the standard matrix associated to T is the $m \times n$ matrix*

$$A_T = \left(T(\underline{e_1}), \ldots, T(\underline{e_n}) \right)$$

$$T(\underline{x}) = A_T \underline{x} \forall \underline{x} \in \mathbb{R}$$

Therefore we can freely move between the $m \times n$ matrix representation and the linear transformation.

6.5 Example

Consider the function

$ID_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n, \underline{x} \to \underline{x}$

This is linear, therefore its standard matrix exists.

$$A_{ID_{\mathbb{R}^n}} = \left(ID_{\mathbb{R}^n}(e_1) ID_{\mathbb{R}^n}(e_2) \ldots ID_{\mathbb{R}^n}(e_n) \right) = (\underline{e_1 e_2} \ldots \underline{e_n})$$

which corresponds to a matrix

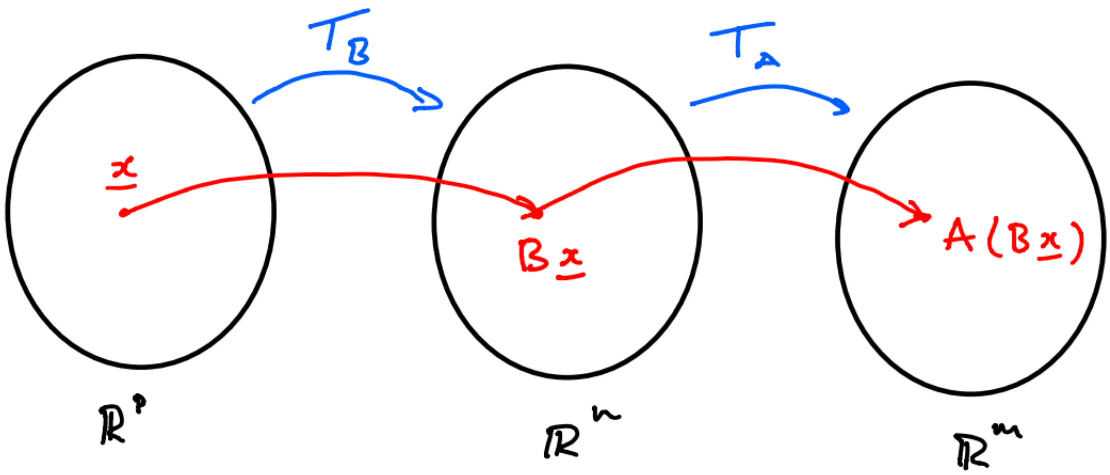
$$\begin{bmatrix} 1 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 7: Linear Transformations and Linear System Solutions

Lecturer: Alexander Paulin *31 January* *Aditya Sengupta*

7.1 Graphical interpretation of linear transformation-matrix relation



7.2 Example of transformation-matrix relation

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a transformation that rotates a vector about the origin by angle θ anticlockwise. Then, T is linear. This means T has a corresponding standard matrix.

Here is an advanced graph that requires the use of complicated tools:

Consider what happens to the basis vectors $\underline{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\underline{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. When T is applied to $\underline{e_1}$ it is rotated to another point on the circle, yielding $T(\underline{e_1}) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$. Similarly $T(\underline{e_2}) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$. This gives us the columns of the matrix.

$$T(\underline{x}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \underline{x} = \begin{bmatrix} x_1 \cos\theta - x_2 \sin\theta \\ x_1 \sin\theta + x_2 \cos\theta \end{bmatrix}$$

7.3 Examples

Find the standard matrix associated to the following:

1. $T : \mathbb{R}^2 \to \mathbb{R}^2$ - reflection in the line $y = -x$.
   Answer:
   We examine what reflection will do to the basis vectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$.
   $T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}0\\-1\end{bmatrix}$ and $T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}-1\\0\end{bmatrix}$
   Therefore in general it transforms a vector such that
   $T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}-x_2\\-x_1\end{bmatrix}$
   or in a pure matrix form
   $$A_T = \begin{bmatrix}0 & -1\\-1 & 0\end{bmatrix}$$
2. $T : \mathbb{R}^2 \to \mathbb{R}^2$ - linear such that
   $$T\left(\begin{bmatrix}1\\-1\end{bmatrix}\right) = \begin{bmatrix}2\\4\end{bmatrix}, T\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix}-2\\6\end{bmatrix}$$
   Add the transformed forms and inputs to get
   $$T\left(\begin{bmatrix}2\\0\end{bmatrix}\right) = \begin{bmatrix}0\\10\end{bmatrix} \implies T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}0\\5\end{bmatrix}$$
   Then subtract this from the first given transformation matrix
   $$T\left(\begin{bmatrix}1\\-1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}2\\4\end{bmatrix} - \begin{bmatrix}0\\5\end{bmatrix}$$
   therefore
   $$T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}-2\\1\end{bmatrix}$$
   So we can combine these in general to get the transform,
   $$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}5x_2\\-2x_1 + x_2\end{bmatrix}$$
   or in a pure matrix form
   $$A_T = \begin{bmatrix}0 & -2\\5 & 1\end{bmatrix}$$

7.4 Range of Linear Transformations

An $m \times n$ matrix is associated with a transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ which maps $x \to x_1\underline{a_1} + x_2\underline{a_2} + \cdots + x_n\underline{a_n}$.

The range of T_A is the image of $\mathbb{R}^n$ under T_A which is $\text{span}\{\underline{a_1}, \ldots, \underline{a_n}\}$.

$\underline{b}$ is in the range of $T_A \iff \underline{b} \in \text{span}\{\underline{a_1}, \ldots, \underline{a_n}\} \iff \left(A|\underline{b}\right)$ is consistent.

Definition 13. *T_A is said to be onto if and only if the range of $T_A = \mathbb{R}^m$.*

This suggests that T_A is onto if and only if $\left(A|\underline{b}\right)$ is consistent for all possible $\underline{b}$. This, in turn, is true if and only if there is a pivot position in every row of the reduced A.

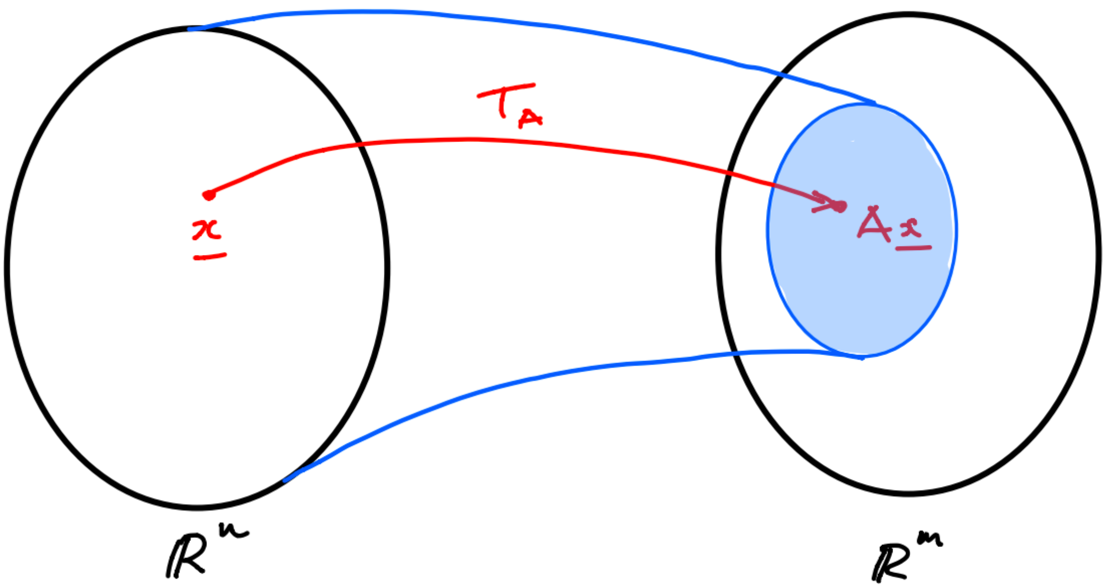
Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 8: Matrix Algebra

Lecturer: Alexander Paulin *2 February* *Aditya Sengupta*

8.1 Linear Transformation Definitions and relation to other concepts

Definition 14. *T_A is onto iff its range is $\mathbb{R}^m$.*



The graph shown here is not onto as it shows a range that is less than $\mathbb{R}^m$.

Definition 15. *T_A is one-to-one iff*

$$T_A(\underline{x}) = T_A(\underline{y}) \implies \underline{x} = \underline{y}$$

.

We now have a set of related concepts to express similar basic ideas, which can be summarised in this table:

Linear Transformations	Vectors	Linear Systems	Reduced Echelon Matrix
T_A is onto	$\text{span}(\underline{a_1}, \ldots, \underline{a_n}) = \mathbb{R}^n$	A$\underline{x}$ = $\underline{b}$ always has a solution	Pivot position in every row
T_A is one-to-one	$\{\underline{a_1}, \ldots, \underline{a_n}\}$ linearly independent	A$\underline{x}$ = $\underline{b}$ admits at most one solution	Pivot position in every column of reduced A

8.2 Adding, Scaling, Composing

We have established that a matrix and a linear transformation are essentially the same. This means we can translate the idea of adding, scaling, and composing functions to matrices.

Given $A = (\underline{a_1}, \dots, \underline{a_n})$ and $B = (\underline{b_1}, \dots, \underline{b_n})$, we can define

$$A + B := (\underline{a_1} + \underline{b_1}, \underline{a_2} + \underline{b_2}, \dots, ula_n + \underline{b_n})$$

the term by term sum of the two matrices. In terms of linear transformations, this is

$$T_A + T_B = T_{A+B}$$

where $T_A, T_B : \mathbb{R}^n \to \mathbb{R}^m$.

By the linearity of transforms, this also means

$$(T_A + T_B)(\underline{x}) = T_A(\underline{x}) + T_B(\underline{x})$$

Similarly with scaling, for matrices we have

$$\lambda A = (\lambda \underline{a_1}, \dots, \lambda \underline{a_n})$$

and for linear transforms,

$$\lambda T_A = T_{\lambda A}$$

For the sum of two matrices to be meaningful, their sizes must be the same.

Composition of matrices will be more difficult, as the domain of one has to be the same as the codomain as the other. Let A consist of n vectors in $\mathbb{R}^m$, $\underline{a_i}$ and let B consist of p vectors in $\mathbb{R}^n$, $\underline{b_i}$. We can define corresponding linear transforms,

$$T_A : \mathbb{R}^n \to \mathbb{R}^m, \underline{v} \to A\underline{v}$$
$$T_B : \mathbb{R}^p \to \mathbb{R}^n, \underline{x} \to B\underline{x}$$

So we can start in $\mathbb{R}^p$, undergo transform T_B, then end up in $\mathbb{R}^n$ whereupon we can apply transform T_A and reach $\mathbb{R}^m$. We can introduce the notion of composed linear transforms, in which a vector $\underline{x}$ goes to

$$\underline{x} \to B\underline{x} \to A(B\underline{x})$$

The composed linear transform $T_A \circ T_B : \mathbb{R}^p \to \mathbb{R}^m$ is also linear. We can try and find the standard matrix of this composed transform by tracking what happens to the basis vectors.

$$(T_A \circ T_B)(\underline{e_1}) = A(B\underline{e_1}) = A(B\underline{e_1}) = A(1 \cdot \underline{b_1} + 0 \cdot \underline{b_2} + \cdots + 0 \cdot \underline{b_p}) = A\underline{b_1}$$

and in general,

$$(T_A \circ T_B)(\underline{e_i}) = A\underline{b_i}, 0 < i \leq p$$

so the standard matrix of the transformation is

$$(T_A \circ T_B) : \mathbb{R}^p \to \mathbb{R}^m = (A\underline{b_1}, \ldots, A\underline{b_p})$$

This gives rise to the idea of matrix "multiplication".

Definition 16.

$$A_{m \times n} B_{n \times p} := (A\underline{b_1}, \ldots, A\underline{b_p})$$

The individual vectors in the above matrix, denoted by $A\underline{x}$, can also be interpreted as matrix multiplication where $p = 1$. The key takeaway from this is that matrix multiplication is the same as the composition of linear transformations.

i.e. $T_{AB} = T_A \circ T_B$

8.3 Row-Column Rule to compute AB

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2} + b_{2j} + \cdots + a_{in} + b_{nj}$$

where i and j represent, respectively, the number of the row and of the column entry of the multiplied matrix.

add in his graph

e.g.

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

This makes sense to do because the first matrix is 2×3 and the second is 3×2. So we expect a 2×2 matrix out.

The product is

$$\begin{bmatrix} 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 1 & 1 \cdot (-1) + 2 \cdot 2 + 3 \cdot 1 \\ 4 \cdot 1 + 5 \cdot 0 + 6 \cdot 1 & 4 \cdot (-1) + 5 \cdot 2 + 6 \cdot 1 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 10 & 12 \end{bmatrix}$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 9: Invertible Matrices

Lecturer: Alexander Paulin *5 February* *Aditya Sengupta*

9.1 Matrix Multiplication contd.

Matrix multiplication has some strange properties. For example, it is not commutative: in general, $AB \neq BA$, and $AB = 0$ does not imply $A = 0$ or $B = 0$. This can be understood if we interpret matrices as a composition of two linear transforms. We see in functions from $\mathbb{R}$ to $\mathbb{R}$ that composition is also noncommutative.

9.2 Invertibility

To avoid the issue of the domain and codomain not matching up, we define an $n \times n$ matrix A. Then

Definition 17. *A is invertible if there exists B, an $n \times n$ matrix such that $AB = BA = I_n$.*

If such a B exists it is unique. We write $B = A^{-1}$.

Proof. We prove the above fact by contradiction. Let B and B' be two inverses of A with the property that $B \neq B'$. Then

$$B = BI = BAB' = IB' = B'$$

Therefore we have a contradiction, so $B \neq B'$ cannot be true. This means a matrix has a unique inverse. $\square$

Example

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

and

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

We can interpret this in terms of linear transformations,

$$T_A \circ T_{A^{-1}} = T_{AA^{-1}} = ID_{\mathbb{R}^n}$$

which is the identity transform that sends $\underline{x}$ to $\underline{x}$.

Similarly,

$$T_{A^{-1}} \circ T_A = T_{A^{-1}A} = ID_{\mathbb{R}^n}$$

This tells us that $T_{A^{-1}}$ is the inverse function of T_A. So we can interpret matrix inversion in terms of function inversion. Recall that a function has an **inverse** if and only if it is both **one-to-one** and **onto**.

Therefore, given an $n \times n$ matrix A, it is invertible if and only if $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one and onto. This, in turn, is true if and only if the columns $\{\underline{a_1}, \ldots, \underline{a_n}\}$ are linearly independent and span$\{\underline{a_1}, \ldots, \underline{a_n}\} = \mathbb{R}^n$.

An interesting coincidence occurs in the $n \times n$ case, in which span$\{\underline{a_1}, \ldots, \underline{a_n}\} = \mathbb{R}^n$ implies the reduced A has a pivot position in every row and $\{\underline{a_1}, \ldots, \underline{a_n}\}$ are linearly independent means the reduced A has a pivot position in every column. Therefore, for an $n \times n$ matrix, these two conditions are the same: a pivot position in every row implies there is one in every column.

For a 3×3 matrix, the only possible echelon form with a pivot in every row is

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

where p is a pivot element. Therefore there is a pivot in every row and in every column. This means that for $n \times n$ matrices, T_A is onto if and only if it is one-to-one.

9.3 Summary

Given an invertible $n \times n$ matrix, $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one and onto, and we can show the converse too, in any direction: given T_A is either one-to-one or onto, we can show its matrix is invertible and that T_A has the other property as well.

The one-to-one and onto conditions can be further interpreted in terms of standard matrices. One-to-one implies $\{\underline{a_1}, \ldots, \underline{a_n}\}$ are linearly independent, i..e there is a pivot position in every column, and onto implies span$\{\underline{a_1}, \ldots, \underline{a_n}\} = \mathbb{R}^n$, i.e. there is a pivot position in every row. Both of these combine to give us the condition that the reduced echelon form of the original matrix has a row equivalent to I_n.

Every one of the above statements guarantees all the others.

9.4 General Inverse of a Matrix

While it is possible to take general expressions for a matrix and its inverse, multiply them, set them equal to I_n and come up with constraints on the inverse, this becomes algebraically complicated.

Instead, it turns out that A being invertible happens if and only if $det(A) := ad - bc \neq 0$, for a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and the inverse matrix is

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

Example

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}^{-1} = \frac{1}{1 \cdot 3 - 2 \cdot 1} \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}$$

For $n > 2$, a simple formula is no longer applicable, just because of how complicated it is. Instead, we observe that A is invertible iff A is row equivalent to I_n.

Therefore, we construct an inverse matrix using linear transformations.$(A|I_n)$ is row equivalent to $(I_n|A^{-1})$. So we have the following algorithm to compute A^{-1}:

1. Write down $(A|I_n)$
2. Put in reduced echelon form $(I_n|B)$
3. $A^{-1} = B$.

Example

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

We write the augmented matrix,

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

Then we perform row operations to get

$$A = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & \frac{1}{2} & \frac{-1}{2} \\ 0 & 1 & 0 & \frac{1}{3} & \frac{-1}{6} & \frac{1}{6} \\ 0 & 0 & 1 & \frac{-1}{3} & \frac{1}{6} & \frac{5}{6} \end{array}\right]$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 10: Determinants

Lecturer: Alexander Paulin *7 February* *Aditya Sengupta*

10.1 Generalising the determinant in $n \times n$ matrices, $n > 2$

Given an $n \times n$ matrix A, let A_{ij} be the $(n-1) \times (n-1)$ matrix formed by removing the $i-$th row and $j-$th column of A.

Then, we define

$$det(A) := a_{11} det(A_{11}) - a_{12} det(A_{12}) + a_{13} det(A_{13}) \tag{10.1}$$

where a_{ij} is the element at the $i-$th row and $j-$th column of A. This can be expanded fully for a 3×3 matrix, into a horrific mess:

$$det(A) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \tag{10.2}$$

This is a terrible formula, but fortunately, we have the recursive definition to simplify this. Mathematically, the $n \times n$ determinant is expressed as follows:

Definition 18.

$$det(A_{n\times n}) := a_{11} det(A_{11}) - a_{12} det(A_{12}) + \cdots + (-1)^{n+1} a_{1n} det(A_{1n}) \tag{10.3}$$

This is also absolutely diabolical, because it requires you to compute n $(n-1) \times (n-1)$ determinants, each of which requires you to do $(n-1)$ $(n-2) \times (n-2)$ determinants, and so on until a base case is reached. We need to find a simpler method.

We can calculate $|A|$ using alternating sums using any row or column. The following pattern must be followed:

$$\begin{bmatrix} + & - & + & - & \dots \\ - & + & - & + & \dots \\ \vdots & & & & \end{bmatrix}$$

Expanding along a different row or column helps in case a lot of values are zero along one of them.

Example

Let $A = \begin{bmatrix} 2 & 1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$. We see the first column has two zero values, so expanding along that gives us

$$\left|A\right| = 2\begin{vmatrix} 2 & 1 \\ 0 & 3 \end{vmatrix} - 0\begin{vmatrix} 1 & 3 \\ 0 & 3 \end{vmatrix} + 0\begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} = 12 - 0 + 0 = 12$$

Notce that in this case the determinant is equal to the product of the leading diagonal. This leads us to a definition,

Definition 19. *An $n \times n$ matrix is upper triangular if it has zero entries below the diagonal.*

An upper triangular matrix has a determinant equal to $\left|A\right| = a_{11}a_{22}\dots a_{nn}$. This can be proven by induction.

Proof. Base case: let $A_{2\times 2}$ be an upper triangular matrix of the form

$$A = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$$

Then $\left|A\right| = ad - b(0) = ad$, which is the product of upper diagonal elements.

Inductive step: let $\left|A_{k\times k}\right|$ be the product of its upper diagonal elements. Then, we consider the upper triangular matrix $A_{(k+1)\times(k+1)}$, which has a first column of the form

$$\begin{bmatrix} a \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

We expand this determinant along the first column, which from the above definition yields

$$\left|A_{k+1\times k+1}\right| = a\left|A_{11}\right| - 0\left|A_{21}\right| + \dots + (-1)^{k+2} \cdot 0 \cdot \left|A_{n-1,1}\right| = a\left|A_{11}\right| \tag{10.4}$$

which is the top-left element multiplied by the upper diagonal elements of the $k \times k$ submatrix, i.e. the product of the upper diagonal elements of $A_{(k+1)\times(k+1)}$

Therefore by induction, the determinant of an upper triangular matrix is the product of its upper diagonal elements. □

Now, we notice that an $n \times n$ matrix in echelon form has to be upper triangular. We supplement this observation by two useful facts that let us reduce the matrix to echelon form and preserve information about the determinant,

1. Switching two rows in a matrix multiplies the determinant by -1.
2. Adding a scalar multiple of one row to another does not change the determinant.

10.2 Algorithm to calculate $det(A)$

1. Put A in echelon form U using only the above two observations.
2. Compute $det(U)$ as the product of diagonal entries.
3. $det(A) = (-1)^r det(U)$, where r is the number of operations that switch two rows.

Example

$$|A| = \begin{vmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 2 & 3 \\ 1 & 0 & 3 & 1 \\ 0 & 2 & 3 & 1 \end{vmatrix}$$

By row operations, including one row switch ($r = 1$), this matrix can be brought to

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

Therefore

$$|A| = (-1) \times |U| = -1 \cdot 1 \cdot 1 \cdot 1 \cdot 2 = -2$$

10.3 Conclusion

$|A| \neq 0$ if and only if $|U| \neq 0$, i.e. there are nonzero entries down the diagonal, which means A is row equivalent to a matrix with a pivot in every column/row and zero entries below, which in turn is row equivalent to I_n. Therefore A is invertible.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 11: Vector Spaces and Linear Transformations

Lecturer: Alexander Paulin *9 February* *Aditya Sengupta*

11.1 What do a matrix and differentiation have in common?

Let A be an $m \times n$ matrix associated to a transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$

Recall the definition of a linear transformation,

$$T_A(\underline{u} + \underline{v}) = T_A(\underline{u}) + T_A\underline{v}$$

$$T_A(\lambda\underline{u}) = \lambda T_A(\underline{u})$$

So, isn't differentiation living in a completely different world?

Let $\mathbb{C}(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \text{ continuous}\}$, and $\mathbb{C}^1(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \text{ with continuous derivative}\}$.

Then we can think of differentiation as a function itself,

$$\frac{d}{dx} : \mathbb{C}'(R) \to \mathbb{C}(\mathbb{R}), f \to \frac{df}{dx}$$

Consider the sum rule of derivatives which states that $\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx}$.

Also, recall that $\frac{d}{dx}(\lambda f) = \lambda\frac{df}{dx}$.

These are the same as the linearity conditions on a transformation above. This allows us to make the important observation that the conditions make sense because we can add and scalar multiply in $\mathbb{R}^n, \mathbb{R}^m, \mathbb{C}^1(\mathbb{R}), \mathbb{C}(\mathbb{R})$.

11.2 Vector Spaces

Definition 20. *Informally, a real vector space is a set V that comes with a concept of "addition" and "real scalar multiplication" satisfying some nice properties.*

1. Polynomials with real coefficients
2. Infinite series with real numbers
3. $\mathbb{R}^n$
4. $\{m \times n \text{ matrices}\}$
5. Planes or lines containing $\underline{0}$ in $\mathbb{R}^3$
6. $\mathbb{R}-$valued functions in more than one variable

7. Random variables and sample spaces
8. $\mathbb{R}-$valued functions on any set: $\{f : [a, b] \to R\}$
9. Sequences of real numbers
10. Complex numbers

A <u>vector</u> in a vector space V is just an element of V.

For example, $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is a vector in $\mathbb{R}^3$, $\sin x$ is a vector in $\mathbb{C}(\mathbb{R})$.

In all the above examples, the following properties hold:

1. $(\underline{u} + \underline{v}) + \underline{w} = \underline{u} + (\underline{v} + \underline{w})$
2. $\exists \underline{0}$ where $\underline{0} + \underline{v} = \underline{v} + \underline{0} = \underline{v}$.
3. Given $\underline{v}, \exists \underline{-v}$ s.t. $\underline{v} + \underline{-v} = \underline{-v} + \underline{v} = \underline{0}$
4. $\underline{u} + \underline{v} = \underline{v} + \underline{u}$
5. $\lambda(\underline{u} + \underline{v}) = \lambda\underline{u} + \lambda\underline{v}$
6. $(\lambda + \mu)\underline{u} = \lambda\underline{u} + \mu\underline{u}$
7. $(\lambda\mu)\underline{u} = \lambda(\mu\underline{u})$
8. $1 \cdot \underline{u} = \underline{u}$

We are now able to give a precise definition of a vector space,

Definition 21. *A real vector space is a set that comes with "addition" and "real scalar multiplication" such that the eight properties above hold.*

A non-example is the upper right quadrant of $\mathbb{R}^2$, i.e.

$$V = x, y | x, y \in R, x, y \geq 0$$

This does not satisfy the requirement that given a vector, its negative exists such that $\underline{v} + \underline{-v} = 0$. Therefore it is not a vector space even though it has addition and scalar multiplication.

Definition 22. *A linear transformation between two vector spaces V and W is a function $T : V \to W$ such that*

1. $T(\underline{u} + \underline{v}) = T(\underline{u}) + T(\underline{v})$
2. $T(\lambda\underline{u}) = \lambda T(\underline{u})$

for all $\underline{u}, \underline{v}$ in V and λ in $\mathbb{R}$.

11.3 Examples

1. $A - m \times n$ matrix $\rightarrow T_A$ linear transform from $\mathbb{R}^n$ to $\mathbb{R}^m$.
2. $\frac{d}{dx}$ = linear transformation from $\mathbb{C}^1(\mathbb{R})$ to $\mathbb{C}(\mathbb{R})$
3. $\int_a^b$ = linear transformation from $\{f : [a, b] \rightarrow \mathbb{R}\}$ to $\mathbb{R}$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 12: Linear Transformations Between Vector Spaces

Lecturer: Alexander Paulin *14 February* *Aditya Sengupta*

12.1 Definition

A linear transformation between two vector spaces V and W is a function $T : V \to W$ such that

1. $T(\underline{u} + \underline{v}) = T(\underline{u}) + T(\underline{v})$
2. $T(\lambda \underline{u}) = \lambda T(\underline{u})$

This clarifies the definition of linear algebra from the start of the course: *the study of linear transformations between vector spaces.*

For example, we can define a linear transformation between $\mathbb{C}^1(\mathbb{R})$ and $\mathbb{C}(R)$:

$$\frac{d}{dx} : f \to \frac{df}{dx}$$

.

Definition 23. *A subspace of a vector space V is a subset $U \subseteq V$ such that*

1. $\underline{0}$ *in* U
2. $\underline{x}, \underline{y}$ *in* $U \implies \underline{x} + \underline{y}$ *in* U.
3. $\underline{x}$ *in* $U \implies \lambda \underline{x}$ *in* $U \forall \lambda \in \mathbb{R}$

A subspace is also its own vector space.

12.1.1 Examples

1. $\mathbb{C}^1(\mathbb{R}) \subset \mathbb{C}(\mathbb{R})$
2. $\mathbb{P}_n(\mathbb{R}) \subset \mathbb{P}(\mathbb{R})$
3. All solutions to $y'' + y = 0$ in $\mathbb{C}^1(\mathbb{R})$

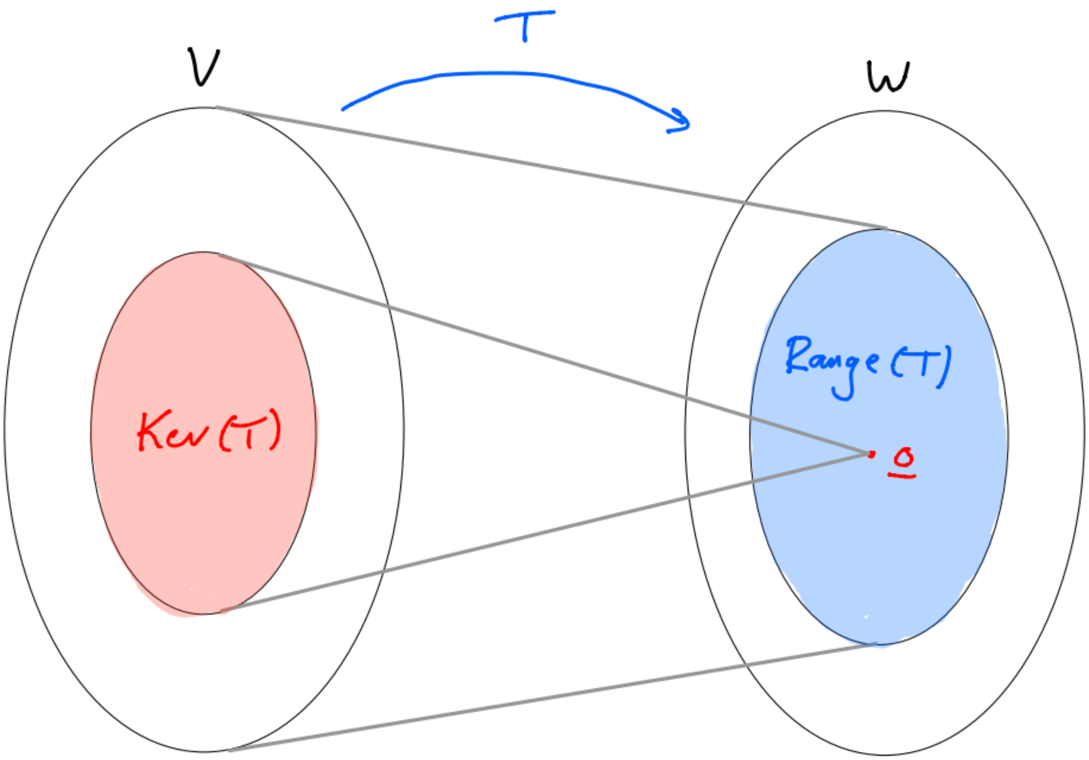
Is this a vector space? We see that the zero function is a solution, trivially as $\underline{0} + \underline{0} = \underline{0}$, that this space has addition because $(f + g)'' + (f + g) = 0$ if f and g are both solutions, and that this space has scalar multiplication because $\lambda(f'' + f)$ is a solution if f is a solution. Therefore this is a vector space.

12.2 Relation between Linear Transformations and Vector Spaces

Let $T : V \to W$ be a linear transformation, and let V, W be any vector spaces.

We define the kernel $\text{Ker}(T) := \{\underline{x} \in V s.t. T(\underline{x}) = \underline{0}\}$ where $\underline{0}$ is the zero vector in W.

$\text{Range}(T) := \{T(\underline{x}) \in W s.t. \underline{x} \in V\}$



Theorem 12.1.

$$\text{Ker}(T) \subset V \ \textit{and} \ \text{Range}(T) \subset W$$

are subspaces.

Proof. $\text{Ker}(T) \subset V$ is a subspace:

1. $T(\underline{0}) = T(\underline{0} \cdot \underline{0}) = 0T(\underline{0}) = \underline{0}$ Therefore the kernel of T has a zero vector.
2. $\underline{x}, \underline{y}$ in kernel of T $\implies T(\underline{x}) = T(\underline{y}) = \underline{0}$

$$T(\underline{x}) + T(\underline{y}) = \underline{0}$$

$$T(\underline{x} + \underline{y}) = \underline{0} \implies \underline{x} + \underline{y} \in \text{Ker}(T)$$

3. $\underline{x} \in \text{Ker}(T), \lambda \in \mathbb{R} \implies T(\underline{x}) = 0 \implies T(\lambda \underline{x}) = 0 \implies \lambda \underline{x} \in \text{Ker}(T)$.

$\text{Range}(T) \subset W$ is a subspace:

1. $T(\underline{0}) \implies \underline{0} \in \text{Range}(T)$

2. $\underline{u}, \underline{v} \in \text{Range}(T) \implies T(\underline{x}) = \underline{u}, T(\underline{y}) = \underline{v}$ for some $\underline{x}, \underline{y} \in V$

$$\underline{u} + \underline{v} = T(\underline{x}) + T(\underline{y}) = T(\underline{x} + \underline{y})$$

   therefore $\underline{u} + \underline{v}$ is in the range of T

3. $\underline{u}$ in Range(T), $\lambda \in \mathbb{R}$

$$T(\underline{x}) = \underline{u}$$

   for some $\underline{x} \in V$

$$\implies \lambda\underline{u} = \lambda T(\underline{x}) = T(\lambda\underline{x})$$

   therefore $\lambda\underline{u}$ is in the range of T

□

12.3 Specific Examples (or not)

Let A be an $m \times n$ matrix associated to a linear transform $T_A : \mathbb{R}^n \to \mathbb{R}^m$. Then

$$\text{Kernel}(T_A) = \{\text{All } \underline{x} \in \mathbb{R}^n \text{ such that } A\underline{x} = \underline{0}\}$$

This is associated with all solutions to the homogeneous linear system $A\underline{x} = \underline{0}$.

$$\text{Range}(T) = \{A\underline{x}\text{such that } \underline{x} \in \mathbb{R}^n\} = \text{span}(\underline{a_1}, \ldots, \underline{a_n}) = \text{Col}(A)$$

This is called the *column space* of A.

Define $T : y \to y'' + y$. We previously checked that this was a vector space, i.e. it has addition and scalar multiplication. We can see that the kernel is the solution set to the output differential equation.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 13: Spanning, Linear Independence and Dimension

Lecturer: Alexander Paulin *17 February* *Aditya Sengupta*

13.1 Span of Vector Spaces

Let V be any vector space. Then define a set of vectors $\{\underline{v_1}, \dots, \underline{v_p}\} \subset V$. We can define an arbitrary linear combination of these vectors,

$$\lambda_1 \underline{v_1} + \cdots + \lambda_n \underline{v_n}$$

where the linear combination is also a vector in V.

Definition 24. *The span of* $\{\underline{v_1}, \dots, \underline{v_p}\}$ *is the subset of all linear combinations of* $\{\underline{v_1}, \dots, \underline{v_p}\}$.

The span is a subspace, which means it has a zero vector, has addition and has scalar multiplication. We say $\{\underline{v_1}, \dots, \underline{v_p}\}$ is a *spanning set* of V if $\text{span}\{\underline{v_1}, \dots, \underline{v_p}\} = V$. For example $\{\underline{e_1}, \underline{e_2}, \dots, \underline{e_n}\} \subset \mathbb{R}^n$, i.e.

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \implies \underline{x} = x_1 \underline{e_1} + \cdots + x_n \underline{e_n}$$

$\{1, x, x^2, \dots, x^n\} \subset \mathbb{P}_n(\mathbb{R})$ means all polynomials are linear combinations of the above set of powers of x.

13.2 Dimension

Definition 25. *We say that V is a finite-dimensional (f.d.) vector space if there exists a finite spanning set.*

For example, $\mathbb{R}^n$ and $\mathbb{P}_n(\mathbb{R})$ is f.d. But the list of all continuous functions is not finite dimensional.

So what is the dimension of a f.d. vector space?

Definition 26. $\underline{v_1}, \dots, \underline{v_p}$ *is L.I* $\iff$ *the only linear combination that sums to zero is the trivial solution, and it is L.D.* $\iff$ *there exists* λ_i *not all zero such that* $\sum \lambda_i \underline{v_i} = \underline{0}$.

$\underline{v_1}, \dots, \underline{v_p}$ L.D. $\iff$ $\underline{v_j}$ in $\text{span}\{\underline{v_1}, \dots, \underline{v_{j-1}}, \underline{v_{j+1}}, \dots, \underline{v_p}\}$ for some j.

Therefore $\underline{v_j}$ does not affect the span of the set of vectors. Because the set is linearly dependent, $\underline{v_j}$ can be expressed as a linear combination of the other vectors, therefore it can be removed and a linear combination of all the other vectors to reach any vector in the previous span still exists.

13.3 Fundamental Definition

Definition 27. *A basis for V is a subset $ulv_1, \ldots, \underline{v_p} \subset V$ such that*

1. $\text{span}(\underline{v_1}, \ldots, \underline{v_p}) = V$
2. $\{\underline{v_1}, \ldots, \underline{v_p}\}$ *L.I*

For example, $\underline{e_1}, \ldots, \underline{e_n}$ is a *standard basis* for $\mathbb{R}^n$.

Let's define a vector space $V = \{\text{solution to } y'' + y = 0\}$. This has a basis $\{\sin x, \cos x\}$.

Note that bases are not unique. Any n linearly independent vectors in $\mathbb{R}^n$ could be bases for $\mathbb{R}^n$, for example. This tells us that if $\{\underline{v_1}, \ldots, \underline{v_p}\}$ is a spanning set of V, then it contains a basis. Also, if $\{\underline{v_1}, \ldots, \underline{v_p}\}$ is linearly independent, then it can be extended to a basis.

If we have a linearly independent set with p elements, and a spanning set k, then $p \leq k$.

Theorem 13.1. *Let V be a f.d. vector space. Any two bases have the same size.*

Proof. Let $\{\underline{v_1}, \ldots, \underline{v_p}\}$, $\{\underline{u_1}, \ldots, \underline{u_k}\}$ be bases for V. Then $\{v_i\}$ is L.I and $\{u_i\}$ is spanning, i.e. $p \leq k$. However if they are both bases then $\{v_i\}$ is spanning and $\{u_i\}$ is L.I., therefore $k \leq p$. This means $p = k$. $\square$

Definition 28. *Let V be a f.d. vector space. Then* $\dim V$ *is the size of any basis.*

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 14: Bases, Dimension, Rank and Nullity

Lecturer: Alexander Paulin *21 February* *Aditya Sengupta*

14.1 Bases, continued

Any linearly independent set is contained in a basis, and a basis by definition has size $\dim V$. Therefore any L.I. set has this property:

$$\{\underline{v_1}, \ldots, \underline{v_p}\} \text{ L. I. } \implies p \leq \dim(V)$$

Similarly,

$$\{\underline{v_1}, \ldots, \underline{v_p}\} \text{ spanning } \implies p \geq \dim(V)$$

Let $p = \dim V$ and let $\{\underline{v_1}, \ldots, \underline{v_p}\}$ be linearly independent. This happens if and only if $\{\underline{v_1}, \ldots, \underline{v_p}\}$ is a basis, because there are $\dim V$ linearly independent vectors. Similarly, let $\{\underline{v_1}, \ldots, \underline{v_p}\}$ be spanning. This also happens if and only if they are a basis, if $p = \dim V$.

For example, let $V = \mathbb{R}^n$ and take vectors $\{\underline{a_1}, \ldots, \underline{a_n}\} \subset \mathbb{R}^n$ L. I $\iff$ reduced A has a pivot position in every column. This is a square matrix, therefore there is also a pivot in every row. This implies $\{\underline{a_1}, \ldots, \underline{a_n}\}$ spans $\mathbb{R}^n$.

14.2 Subspaces

Theorem 14.1. *Let $U \subset V$ be a subspace. Then*

1. $\dim V < \infty \implies \dim U < \infty$
2. $\dim U \leq \dim V$
3. $\dim U = \dim V \implies U = V$

14.3 Rank and Nullity

Let V and W be vector spaces, and let $T : V \to W$ be linear. Then,

$$\text{Ker}(T) = \{\underline{x} \in V s.t. T(\underline{x}) = 0\} \subset V$$

$$\text{Range}(T) = \{T(\underline{x}) \in W s.t. \underline{x} \in V\} \subset W$$

This can be made more concrete with an example using a matrix. Let A be an $m \times n$ matrix associated to the transform $T_A : \mathbb{R}^n \to \mathbb{R}^m$. Then the kernel is defined as

$$\mathrm{Ker}(T_A) = \{\underline{x} \in \mathbb{R}^n s.t. A\underline{x} = \underline{0}\} = \mathrm{Nul}(A)$$

and the range is

$$\mathrm{Range}(T_A) = \{A\underline{x} = x_1\underline{a_1} + \cdots + x_n\underline{a_n}\} = \mathrm{Span}(\underline{a_1}, \ldots, \underline{a_n}) = Col(A)$$

Then, we see that

$$\mathrm{Ker}(T_A) = \{\underline{0}\} \iff \{\underline{a_1}, \ldots, \underline{a_n}\} \text{ L. I.}$$

i.e. the bigger $\mathrm{Ker}(T_A)$ is, the more linearly dependent $(\underline{a_1}, \ldots, \underline{a_n})$.

For a finite-dimensional vector space, the dimensions of the kernel and range are also both finite.

Definition 29. *The rank of a linear transformation T is the dimension of the range of T.*

Definition 30. *The nullity of T is the dimension of the kernel of T.*

For example, consider the matrix

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

Row reduce to get

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

Let the $i-$th column be $\underline{a_i}$. Then, we see that $\underline{a_2}$ is in the span of $\underline{a_1}$ and $\underline{a_4}$ is in the span of $\{\underline{a_1}, \underline{a_2}, \underline{a_3}\}$.

$$\implies \mathrm{Range}(T_A) = \mathrm{Col}(A) = \mathrm{span}(\underline{a_1}, \ldots, \underline{a_5}) = \mathrm{span}(\underline{a_1}, \underline{a_3}, \underline{a_5})$$

We can verify this by noticing that there are pivots in columns 1, 3, 5, therefore the first, third and fifth vectors are L.I. Put another way, $\underline{a_1}, \underline{a_3}, \underline{a_5}$ form a basis for $\mathrm{Col}(A)$ and the rank of T_A is 3.

In general, we can say that the pivot columns of A form a basis for $\mathrm{Range}(T_A)/\mathrm{Col}(A)$. This means the rank of T_A is the number of pivot columns.

The kernel of T_A is the set of $\underline{x} \in \mathbb{R}^5 s.t. A\underline{x} = \underline{0}$ We make an augmented matrix out of the above row reduced matrix, with the last column all zeros. This gives us a general solution,

$$\underline{x} = \begin{bmatrix} -2x_2 - x_4 \\ x_2 \\ -x_4 \\ x_4 \\ 0 \end{bmatrix}$$

This is the set of input vectors for which the output of the linear transform associated to A is the zero vector, i.e. this is the kernel.

In parametric form this becomes

$$x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

which gives us a spanning set for the kernel.

$$\implies \text{Ker}(T_A) = \text{span}\left(\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right)$$

These are linearly independent, therefore the nullity of T_A is 2. This process always gives a basis for the kernel or null space.

Also notice that the nullity is equal to the number of free columns of A.

Major consequence of this

(not major enough to get its own numbered subsection, but major)

Given an $m \times n$ matrix A,

$$n = \text{rank}(T_A) + \text{Nullity}(T_A) = \dim(\text{span}(\underline{a_1}, \ldots, \underline{a_n})) + \dim(\{\underline{x} \in \mathbb{R}^n s.t. A\underline{x} = \underline{0}\})$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 15: Bases and Coordinate Systems

Lecturer: Alexander Paulin *23 February* *Aditya Sengupta*

Let B be a basis for V. This means the vectors $\{\underline{b_1}, \ldots, \underline{b_n}\}$ span V and are linearly independent. We want to be able to use B to think about V in a more concrete way. We observe that each $\underline{x}$ in V can be written as a linear combination of the basis vectors in a unique way. This gives rise to a definition,

Definition 31. *The coordinate vector of $\underline{x}$ with respect to basis B is* $\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}$ *in $\mathbb{R}^n$, where $\underline{x} = \lambda_1 \underline{b_1} + \cdots + \lambda_n \underline{b_n}$.*

This allows us to translate between the abstract notion of a vector space to the concrete notion of $\mathbb{R}^n$ that we are used to.

15.1 Examples of coordinate bases

1. Trivially, if we take $V = \mathbb{R}^n$, we see the coordinate vector of $\underline{x}$ is $\underline{x}$.

2. A more interesting example is $\mathbb{P}_n(\mathbb{R})$, where the basis is $\{1, x, x^2, \ldots, x^n\}$. To write the arbitrary vector in this space

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

in coordinates, we employ the above definition and see the coordinate vector is

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

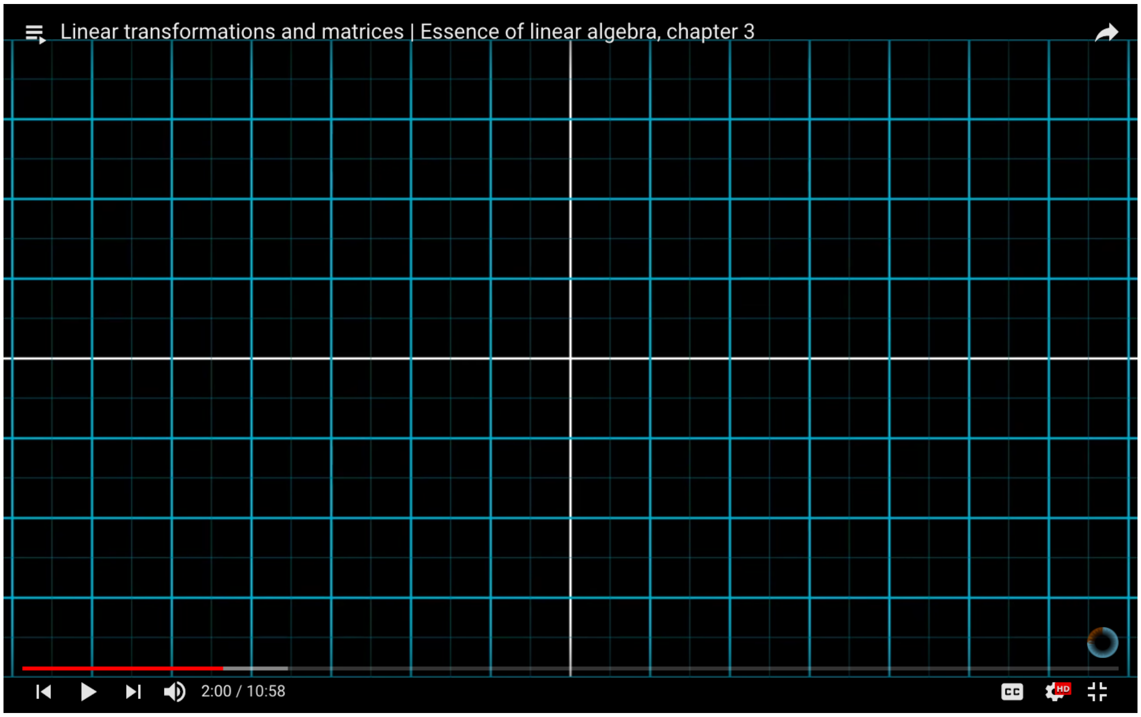
which is a vector in $\mathbb{R}^{n+1}$.

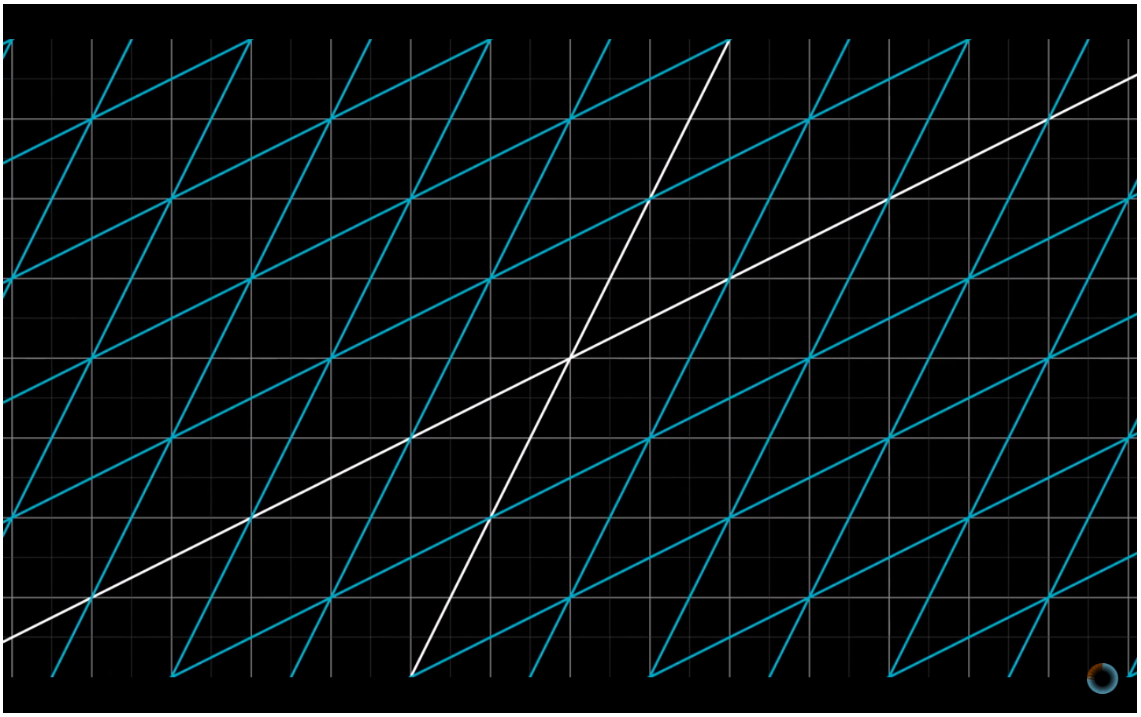
3. $V = \mathbb{R}^2$, $B = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$

Suppose we want to write $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$ in B coordinates. We row reduce the augmented matrix consisting of the basis vectors and see that the vector we get is $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Therefore

$$\begin{bmatrix} 5 \\ 3 \end{bmatrix}_B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The geometric intuition for this is that the $x - y$ plane is warped. I'm stealing a picture from 3Blue1Brown brb:


Linear transformations and matrices | Essence of linear algebra, chapter 3
2:00 / 10:58



15.2 General Situation

Let B be a basis in $\mathbb{R}^n$. Then in general we have an $n \times n$ matrix augmented by an arbitrary vector $\underline{x}$, that we row reduce:

$$(\underline{b_1}, \ldots, \underline{b_n} | \underline{x}) \to (I_n | (\underline{x})_B)$$

Choosing a basis B allows us to identify V with $\mathbb{R}^n$. Because the transform $\underline{x} \to \underline{x}_B$ is linear, one-one and onto, so is the general transform from V to $\mathbb{R}^n$ that is characterised by the above transform on a generic vector.

The structural properties of V are preserved when we switch it to $\mathbb{R}^n$.

Example

Is $\{x^2 + 2x + 1, x^2, x + 1\}$ linearly independent in $\mathbb{P}_2(\mathbb{R})$?

We switch these vectors into their coordinate representation, and get

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

Row reduction shows us that this set is linearly independent in $\mathbb{R}^3$. This implies the original set of vectors is linearly independent in $\mathbb{P}_2(\mathbb{R})$.

15.3 Timely Warning from UCPD

(yes I'm stealing my own bad not-even-jokes from 53 notes)

Different bases give different coordinate systems. Even if two different bases are valid, i.e. they are linearly independent sets and therefore an arbitrary vector can be expressed as a linear combination of them, the coefficients of the linear combination will naturally be different and so the coordinate vector related to each of them will be different.

We can convert between coordinate systems by using the fact that all transformations in this case are linear. Say there are two bases B and C. Given the coordinates in B, we want to find those in C, i.e. we need $T(\underline{x}_B) = \underline{x}_C$ for all $\underline{x} \in V$.

Because T is linear, $A(\underline{x}_B) = \underline{x}_C$ for some $n \times n$ matrix A. Recall that $T = (T(\underline{e_1}), \ldots, T(\underline{e_n}))$ and $(b_i)_B = \underline{e_i}$. Therefore

$$T(e_i) = T((\underline{b_i})_B) = (\underline{b_i})_C$$

which gives rise to a definition,

Definition 32.

$$P_{B \to C} := ((\underline{b_1}_C), \ldots, \underline{b_n}_C)$$

therefore the B to C conversion can be found just by knowing what happens to the B basis vectors when converted to C.

This is ridiculously abstract.

15.4 General Change of Basis for $\mathbb{R}^n$

$$(\underline{c_1}, \ldots, \underline{c_n} | \underline{b_i}) \to (I_n | (\underline{b_i}_C))$$

$$\implies (\underline{c_1}, \ldots, \underline{c_n} | \underline{b_1}, \ldots, \underline{b_n}) \to (I_n | P_{B \to C})$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 16: Eigenvalues and Eigenvectors

Lecturer: Alexander Paulin *26 February* *Aditya Sengupta*

16.1 Studying One Specific Linear Transformation

Given an $n \times n$ matrix A, we want to understand $T_A : \mathbb{R}^n \to \mathbb{R}^n$ in more depth.

Example

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

We can understand what this means by multiplying by the basis vectors,

$$A\underline{e_1} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2\underline{e_1}$$

$$A\underline{e_2} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} = 3\underline{e_2}$$

Therefore, we have the general expression

$$A\underline{x} = A(x_1\underline{e_1} + x_2\underline{e_2}) = x_1A\underline{e_1} + x_2A\underline{e_2} = 2x_1\underline{e_1} + 3x_2\underline{e_2} = \begin{bmatrix} 2x_1 \\ 3x_2 \end{bmatrix}$$

Definition 33. *Let A be an $n \times n$ matrix. An eigenvector of A is a vector $\underline{v}$ in $\mathbb{R}^n$ such that*

1. $\underline{v} \neq \underline{0}$
2. $A\underline{v} = \lambda\underline{v}$ *for some* $\lambda \in \mathbb{R}$

We call any such λ an eigenvalue of A.

For A as defined above, $\underline{e_1}$ is an eigenvector with eigenvalue 2, and $\underline{e_2}$ is an eigenvector with eigenvalue 3.

Definition 34. *We say that an $n \times n$ matrix A is diagonal if $(A)_{ij} = 0$ for $i \neq j$.*

An $n \times n$ matrix is diagonal if and only if all of the basis vectors are eigenvectors of A. We can show this based on the fact that $A\underline{e_i}$ is the $i-$th column of A, which is equal to $\lambda_i\underline{e_i}$ for some $\lambda_i \in \mathbb{R}$.

16.2 Eigenvectors Corresponding to an Eigenvalue

Given A is an $n \times n$ matrix, $\lambda \in \mathbb{R}$,

Definition 35. *The λ eigenspace of A is the subset of $\mathbb{R}^n$ given by all eigenvectors of A with eigenvalue λ and the zero vector.*

We observe that $\underline{v}$ in $\lambda-$eigenspace $\iff A\underline{v} = \lambda\underline{v} \iff (A - \lambda I_n)\underline{v} = 0$

which means $\underline{v}$ is in $\mathrm{Nul}(A - \lambda I_n)$.

We conclude that the $\lambda-$eigenspace of A is equal to the null space of $A - \lambda I_n$. The $\lambda-$eigenspace of A is a subspace of something.

Observe that λ is an eigenvalue of A if and only if the $\lambda-$eigenspace $\neq \{\underline{0}\}$.

Example

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

We want to find the $2-$eigenspace of A. This is equivalent to the null space of $A - 2I_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

This has the null space $\left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \right\}$, which is the span of $\underline{e_1}$ and $\underline{e_2}$.

More generally, if we have A as a diagonal matrix in which $A_{ii} = \lambda_i$, then the $\lambda-$eigenspace is the span of all $\underline{e_i}s$ such that $\lambda_i = \lambda$.

16.3 Finding Eigenvalues/Eigenvectors of Non-Diagonal Matrices

λ is an eigenvalue of A if and only if there exists $\underline{v} \neq \underline{0}$ such that $A\underline{v} = \lambda\underline{v}$. This in turn is true if and only if there exists $\underline{v} \neq \underline{0}$ such that $(A - \lambda I_n)\underline{v} = \underline{0}$. This implies the columns of $A - \lambda I_n$ are linearly dependent, which means it is not invertible. Therefore the determinant of $A - \lambda I_n$ is 0.

Definition 36. *The characteristic polynomial of A is given by $\left|\left(\right| A - xI_n\right)$ where x is a variable.*

Example

Let A be

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

Then, the characteristic polynomial is the determinant of

$$A - xI_3 = \begin{bmatrix} 1-x & 3 & 3 \\ -3 & -5-x & -3 \\ 3 & 3 & 1-x \end{bmatrix}$$

which is

$$det(A - xI_3) = (1-x)\begin{vmatrix} -5-x & -3 \\ 3 & 1-x \end{vmatrix} - 3\begin{vmatrix} -3 & -3 \\ 3 & 1-x \end{vmatrix} + 3\begin{vmatrix} -3 & -5-x \\ 3 & 3 \end{vmatrix}$$

The characteristic polynomial comes out to be $-x^3 - 3x^2 + 4$.

We notice that the degree of the characteristic polynomial is equal to the size of the matrix. Solving for the zeros of this polynomial, we get $-(x-1)(x+2)^2$, i.e. the eigenvalues are 1 and -2.

Definition 37. *The algebraic multiplicity of λ is the number of times $x - \lambda$ divides the characteristic polynomial.*

In the above example, 1 has algebraic multiplicity 1 and -2 has algebraic multiplicity 2.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 17: The Characteristic Equation

Lecturer: Alexander Paulin *28 February* *Aditya Sengupta*

17.1 Problem

Translating a matrix to a characteristic polynomial whose zeros are the eigenvalues sounds great, but in general, solving polynomials of degree greater than 3 becomes quite difficult.

In addition to this, it is possible that the matrix has no eigenvectors/eigenvalues at all. Consider the matrix

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

The characteristic polynomial is then

$$det(A - xI_2) = \begin{vmatrix} -x & -1 \\ 1 & -x \end{vmatrix} = x^2 + 1$$

which has no roots in $\mathbb{R}$.

We understand why this has no roots if we recognise that this is a rotation matrix, that rotates its elements by $\frac{\pi}{2}$ anticlockwise about $\underline{0}$. No eigenvectors are possible here because it necessarily moves you off the line.

Theorem 17.1. *The eigenvalues of an upper triangular matrix are just the entries of the main diagonal.*

Proof. Let A be a general upper triangular matrix with diagonal elements a_{ii}, $1 \leq i \leq n$. Then the determinant of $(A - xI_n)$ is the product of the diagonal elements $a_{ii} - x$. Therefore the characteristic polynomial is

$$det(A - xI_n) = (a_{11} - x)(a_{22} - x) \dots (a_{nn} - x)$$

□

17.2 Warning

The section names sure are cheerful today, aren't they?

The characteristic polynomial is not preserved by row operations. Therefore we cannot row reduce an arbitrary matrix to upper triangular form to easily yield the characteristic polynomial. "Life is not as simple as you might hope."

Example

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \implies \begin{vmatrix} 1-x & 1 \\ 1 & 2-x \end{vmatrix} = 1 - 3x + x^2$$

whereas if we row reduced, we would have

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \implies 1 - 2x + x^2$$

We see that the characteristic equation is not preserved.

17.3 Basis of eigenvectors - extended application

Here we will examine *discrete dynamic systems* as an illustration of the utility of eigenvectors. A discrete dynamic system is any system whcih changes at discrete time intervals.

1. The state of the system can be described at each time k using a vector $\underline{x}_k$ in $\mathbb{R}^n$.
2. Each time transition is governed by a fixed $n \times n$ matrix A, that is

$$A\underline{x}_k = \underline{x}_{k+1}$$

We are interested in finding the long term behaviour of a system like this, e.g. the population of owls and rats measured annually. Let the vector representing this be

$$\underline{x}_k = \begin{bmatrix} o_k \\ r_k \end{bmatrix}$$

and let the transformation matrix be

$$A = \begin{bmatrix} -1 & 3 \\ \frac{-3}{2} & \frac{7}{2} \end{bmatrix}$$

Given the starting point $\underline{x}_0 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. To predict long-range behaviour without doing a lot of computation, we find the eigenvalues of A.

We find that the characteristic equation is

$$det(A - xI_2) = \frac{1}{2}(2x-1)(x-2)$$

therefore the eigenvalues are $\{\frac{1}{2}, 2\}$.

The null space of $A - \frac{1}{2}I_2$ is $\text{Nul}\left(\begin{bmatrix} \frac{-3}{2} & 3 \\ \frac{-3}{2} & 3 \end{bmatrix}\right)$

which we can see is the span of $\begin{bmatrix}2\\1\end{bmatrix}$, if we go through a single step of row reduction.

Similarly, the null space of $A - 2I_2$ is

$$\text{Nul}\Big(\begin{bmatrix}-3 & 3\\ \frac{-3}{2} & \frac{3}{2}\end{bmatrix}\Big)$$

which is the span of $\begin{bmatrix}1\\1\end{bmatrix}$.

Therefore the basis of eigenvalues is $\{\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}\}$.

We know that

$$\underline{x}_0 = \begin{bmatrix}3\\2\end{bmatrix} = \begin{bmatrix}2\\1\end{bmatrix} + \begin{bmatrix}1\\1\end{bmatrix}$$

and we can calculate

$$\underline{x}_1 = A\begin{bmatrix}3\\2\end{bmatrix} = A\Big(\begin{bmatrix}2\\1\end{bmatrix} + \begin{bmatrix}1\\1\end{bmatrix}\Big) = \frac{1}{2}\begin{bmatrix}2\\1\end{bmatrix} + 2\begin{bmatrix}1\\1\end{bmatrix}$$

$$\underline{x}_2 = \frac{1}{2}\cdot\frac{1}{2}\begin{bmatrix}2\\1\end{bmatrix} + 2\cdot 2\begin{bmatrix}1\\1\end{bmatrix}$$

(which we can calculate by applying A to $\underline{x}_1$ as calculated above).

In general,

$$\underline{x}_k = A^k\begin{bmatrix}3\\2\end{bmatrix} = \frac{1}{2^k}\begin{bmatrix}2\\1\end{bmatrix} + 2^k\begin{bmatrix}1\\1\end{bmatrix}$$

"Good luck doing that by multiplying matrices repeatedly."

In the limit as $k \to \infty$, the $\frac{1}{2^k}$ term drops out and therefore

$$\underline{x}_k \approx \begin{bmatrix}2^k\\2^k\end{bmatrix}$$

17.4 Follow up questions

1. Is there $\underline{x}_0$ such that $\underline{x}_0 = \underline{x}_1 = \ldots$?

   The answer turns out to be no, as that would require that $\underline{x}_0 = A\underline{x}_0$ which would require an eigenvector with eigenvalue 1.

2. What property of $\underline{x}_0$ guarantees that $\underline{x}_k \to 0$ as $k \to \infty$?

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 18: Diagonalization

Lecturer: Alexander Paulin *2 March* *Aditya Sengupta*

18.1 Basis consisting of eigenvectors of a matrix

Suppose we have a basis $\{\underline{b_1}, \ldots, \underline{b_n}\}$. This corresponds to a matrix in which the columns are the basis vectors. This matrix is necessarily invertible. This means

$$P\underline{e_i} = \underline{b_i}$$

by construction. But because the matrix is invertible, we can also say

$$\underline{e_i} = P^{-1}\underline{b_i}$$

Theorem 18.1. *$\{\underline{b_1}, \ldots, \underline{b_n}\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A if and only if the matrix $P^{-1}AP$ is diagonal, i.e. any element off the left diagonal is zero.*

Proof. Suppose there exists a basis consisting of eigenvectors. This is characterised by

$$A\underline{b_i} = \lambda_i \underline{b_i} \forall i$$

Then,

$$(P^{-1}AP)\underline{e_i} = P^{-1}A\underline{b_i} = \lambda_i P^{-1}\underline{b_i} = \lambda_i \underline{e_i}$$

This means $\underline{e_i}$ is an eigenvector of $P^{-1}AP$ with eigenvalue λ_i. Therefore each of the columns of $P^{-1}AP$ has zero elements not along the leading diagonal, and the appropriate eigenvalues along the leading diagonal. Therefore the matrix is diagonal and the proof is complete.

To show an if-and-only-if relationship, we prove this in the other direction too. Let $P^{-1}AP$ be a diagonal matrix with elements along the leading diagonal λ_i and other elements zero. We multiply this matrix by the stadnard basis vector $\underline{e_i}$, and we get:

$$(P^{-1}AP)\underline{e_i} = \lambda_i \underline{e_i} \implies AP\underline{e_i} = \lambda_i P\underline{e_i}$$

Therefore

$$A\underline{b_i} = \lambda_i \underline{b_i}$$

which means $\underline{b_i}$ is an eigenvector of A with eigenvalue λ_i. $\square$

Example

$$A = \begin{bmatrix} -1 & 3 \\ -\frac{3}{2} & \frac{7}{2} \end{bmatrix}$$

Solving the characteristic polynomial gives us eigenvalues $\frac{1}{2}$ and 2, which gives us the basis vectors

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

which means something that I couldn't get because he moved to the next page.

Definition 38. *$n \times n$ matrices A and B are similar if there exists an invertible matrix P such that $P^{-1}AP = B \iff A = PBP^{-1}$.*

We can conclude that there exists a basis of $\mathbb{R}^n$ consisting of eigenvectors of A if and only if A is similar to a diagonal matrix. We say that A is diagonalizable. We shouldn't say it because that's a horrible word, but we say it.

This gives us the following useful consequence:

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

where B is a diagonal matrix. Then,

$$A^k = PBP^{-1}PBP^{-1} \dots PBP^{-1}$$

which can cancel to get

$$A^k = PB^kP^{-1}$$

As B is diagonal, B^k is just the matrix we get by raising every element of B to the $k-th$ power.

18.2 Is every square matrix diagonalizable?

No.

1. There may be no real eigenvalues or eigenvectors.
   We *can* develop the theory of complex vector spaces, but it's beyond our scope right now.
2. The dimension of the eigenspaces may not be big enough.

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

   This has 0 as its only eigenvalue. Then the null space of $A - \lambda I_2$ is just the null space of A. A is already in RREF, so the null space is the span of $\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\}$. This is one dimensional, but we need two linearly independent eigenvectors.

Theorem 18.2. *Let A be an $n \times n$ matrix with distinct real eigenvalues $\lambda_1, \ldots, \lambda_p$.*

1. *$dim(\lambda_i$ eigenspace$) \leq$ algebraic multiplicity of λ_i.*

   Therefore the sum of the dimension of the λ_i eigenspace for $1 \leq i \leq n$ is n, the degree of the characteristic polynomial.

2. *A diagonalizable $\iff$ $dim(\lambda_i$ eigenspace$) =$ algebraic multiplicity of λ_i for all i.*

3. *If A is diagonalizable and β_i is a basis of the λ_i eigenspace, then $\beta = B_1 \bigcup B_2 \cdots \bigcup B_p$ is a basis of $\mathbb{R}^n$.*

18.3 Important Consequence

If A has exactly n distinct real eigenvalues $\lambda_1, \ldots, \lambda_n$, then A is diagonalizable.

Justification:

$$dim(\lambda_i \text{ } eigenspace)) \geq 1$$

$$\implies dim(\lambda_1 \text{ } eigenspace) + \cdots + dim(\lambda_n \text{ } eigenspace) \geq n$$

but we know this is also $\leq n$ from the first property. Therefore it is equal to n, which means it is diagonalizable.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 19: Matrices and Abstract Linear Transformations

Lecturer: Alexander Paulin *5 March* *Aditya Sengupta*

Let B and C be bases of vector spaces V and W respectively, such that $dimV = n$ and $dimW = m$. Recal that we can use a coordinate system to translate between the abstract V and the concrete $\mathbb{R}^n$. This relation is one-to-one, onto, and linear. Therefore

$$V \iff \mathbb{R}^n \implies \underline{x} \iff \underline{x_B} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}, \underline{x} = \lambda_1 \underline{b_1} + \cdots + \lambda_n \underline{b_n}$$

.

and

$$W \iff \mathbb{R}^m \implies \underline{x} \iff \underline{x_C} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}, \underline{x} = \mu_1 \underline{c_1} + \cdots + \mu_n \underline{c_m}$$

.

Therefore, we know we can translate between $\mathbb{R}^n$ and V, between V and W, and between V and $\mathbb{R}^m$. This means that we can define a transformation between $\mathbb{R}^n$ and $\mathbb{R}^m$. Call this transformation T_{BC}. This takes the vector from $\underline{x_B}$ to $\underline{x}$ to $T(\underline{x})$ to $T(\underline{x})_C$.

$T_{BC} : \mathbb{R}^n \to \mathbb{R}^m$ is linear. This means there exists an $m \times n$ matrix A_{BC} such that

$$T_{BC}((\underline{x})_B) = (T(\underline{x}))_C = A_{BC}\underline{x_B}$$

To work out what A_{BC} is, we need $\underline{x_B}$ to be the $\underline{b_i}$s, that is, the basis vectors. Recall that $(\underline{b_{i_\beta}}) = \underline{e_i}$.

The $i - th$ column of A_{BC} is $A_{BC}\underline{e_i} = A_{BC}(\underline{b_{i_\beta}}) = (T(\underline{b_i})_C)$.

Definition 39. *The matrix of* $T : V \to W$ *with respect to the bases* B *and* C *is* $A_{BC} = \{T(\underline{b_1})_C, \ldots, T(\underline{b_n})_C\}$

This matrix has the key property that $A_{BC}\underline{x_B} = (T(x)_C) \forall x \in B$.

Example

$V = \mathbb{R}^n, W = \mathbb{R}^m, \beta = \varepsilon_n, C = \varepsilon_m, T : \mathbb{R}^n \to \mathbb{R}^m$ linear.

Then, the standard matrix we're looking for is

$$A_{\varepsilon_n, \varepsilon_m} := \Big((T(\underline{e_1})_{\varepsilon_m}, T(\underline{e_n})_{\varepsilon_m})\Big)$$

$$= (T(\underline{e_1}), \ldots, T(\underline{e_n}))$$

which is the standard matrix of T.

Example

$$V = W = \mathbb{P}_2(\mathbb{R}), \beta = C = \{1, x, x^2\}$$

$$(a_0 + a_1 x + a_2 x^2)_\beta = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

This is a transformation between $\mathbb{P}_2(\mathbb{R})$ and $\mathbb{R}^3$.

Let the linear transformation be $p(x) \to p'(x)$. Then the transformation matrix is

$$A_{BB} = \{(T(1))_\beta, T(x)_\beta, T(x^2)_\beta\}$$

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

which is the derivatives of each of the basis elements in B coordinates.

Example

one of the main reasons we're doing this

$$T : \mathbb{R}^2 \to \mathbb{R}^2$$

$$\underline{x} \to \begin{bmatrix} -1 & 3 \\ \frac{-3}{2} & \frac{7}{2} \end{bmatrix} \underline{x}$$

Recall that we found the basis of eigenvectors for this matrix, which translates to

$$B = C = \{\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$$

with eigenvalues $\frac{1}{2}$ and 2 respectively.

Then,

$$A_{BB} = \left(T(\begin{bmatrix} 2 \\ 1 \end{bmatrix})_B, T(\begin{bmatrix} 1 \\ 1 \end{bmatrix})_B\right) = \left((\frac{1}{2}\begin{bmatrix} 2 \\ 1 \end{bmatrix})_B, (2\begin{bmatrix} 1 \\ 1 \end{bmatrix})_B\right)$$

$$= (\frac{1}{2}\underline{e_1}, 2\underline{e_2}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}$$

19.1 A Really Awesome Fact For Us

Take a general case in which A is an $n \times n$ matrix, $B \subset \mathbb{R}^n$ is a basis. Then A_{BB} is diagonal if and only if B is a basis of $\mathbb{R}^n$ consisting of eigenvectors.

If $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ and $T = T_A$ for $A_{n\times n}$, can we find a more direct way to determine A_{BC}?

Recall that $\underline{v}_{\varepsilon_n} = \underline{v}$, and $P_{B\to\varepsilon_n} = (\underline{b_1}, \ldots, \underline{b_n})$. Therefore,

$$P_{B\to\varepsilon_n}(\underline{x_B}) = \underline{x}_{\varepsilon_n} = \underline{x}$$

Then, we can invert the transformation between B and ε_n to get

$$\underline{x}_B = (P_{B\to\varepsilon_n})^{-1}\underline{x}$$

Therefore $P^{-1}_{B\to\varepsilon_n} = P_{\varepsilon_n\to B}$. This lets us write down the first transformation in our chain from $\mathbb{R}^n \to \mathbb{R}^m$ as a matrix multiplication, i.e.

$$\underline{x_B} \to P_{B\to\varepsilon_n}(\underline{x_B}) \to AP_{B\to\varepsilon_n}(\underline{x_B}) \to P_{\varepsilon_m\to C}AP_{B\to\varepsilon_n}(\underline{x_B})$$

Conclusion:

$$A_{BC} = P_{\varepsilon_m\to C}AP_{B\to\varepsilon_n}$$

$$= (P_{C\to\varepsilon_m})^{-1}A(\underline{b_1}, \ldots, \underline{b_n})$$

$$= (\underline{c_1}, \ldots, \underline{c_m})^{-1}A(\underline{b_1}, \ldots, \underline{b_n})$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 20: Examples of using abstract linear transformations

Lecturer: Alexander Paulin *7 March* *Aditya Sengupta*

20.1 Recall

Given $T : V \to W$, we cannot make an associated matrix based on this alone. We need to define a basis in V and another in W. So, let $B = \{\underline{b_i}\}, 1 \le i \le n$ be a basis for V and let $C = \{\underline{c_i}\}, 1 \le i \le m$ be a basis for V. Then we can define the transformation based on what happens to the basis vectors in the domain:

$$A_{BC} = \Big((T(\underline{b_i}))_C\Big), 1 \le i \le n$$

This has the key property that

$$\Big(T(\underline{x})_C\Big) = A_{BC}\underline{x}_B$$

20.1.1 Special Cases

1. A - $m \times n$ matrix, $T = T_A$

$$\implies A_{BC} = (\underline{c_1}, \ldots, \underline{c_m})^{-1} A(\underline{b_1}, \ldots, \underline{b_n})$$

2. A - $n \times n$ matrix, $T = T_A, B = C, P = (\underline{b_1}, \ldots, \underline{b_n})$

$$\implies A_{BB} = P^{-1}AP$$

3. A - $n \times n$ matrix, $B = C$, then

$$A_{BB} \text{ diagonal } \iff B \text{ basis of eigenvectors}$$

20.1.2 Important Future Goal

In general, we want to be able to find bases B and C such that A_{BC} is as simple as possible.

Definition 40. *A and M are similar if and only if there exists P invertible such that $P^{-1}AP = M$.*

20.2 Examples

1. $T : V \to W$ linear

$$B = \{\underline{b_1}, \underline{b_2}, \underline{b_3}\}, C = \{\underline{c_1}, \underline{c_2}, \underline{c_3}, \underline{c_4}\}$$

$$T(\underline{b_1}) = \underline{c_1} - 2\underline{c_2}, T(\underline{b_2}) = \underline{c_1} + \underline{c_2} + \underline{c_3} + \underline{c_4}, T(\underline{b_3}) = \underline{c_1} + 2\underline{c_3}$$

Find A_{BC}. Is T one to one? What is its rank?

Answer:

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

T is not one-to-one as that requires a pivot in every column. So we row reduce to get

$$A_{RREF} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & \frac{4}{3} \\ 0 & 0 & 0 \end{bmatrix}$$

or something. (recheck)

Therefore there is a pivot in every column, so T is one-to-one. The rank of T is 3.

2. $T : \mathbb{P}_2(\mathbb{R}) \to \mathbb{R}^2$

$$p(x) \to \begin{bmatrix} p(0) \\ p(1) \end{bmatrix}$$

Find A_{BC} when

(a)

$$B = \{1, x, x^2\}, C = \{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$$

Answer

$$A = [(T(\underline{b_1}))_C (T(\underline{b_2}))_C (T(\underline{b_3}))_C]$$

$$= \left[T \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_C \; T \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_C \; T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_C \right]$$

$$= [\begin{bmatrix} 1 \\ 1 \end{bmatrix}_C \begin{bmatrix} 0 \\ 1 \end{bmatrix}_C \begin{bmatrix} 0 \\ 1 \end{bmatrix}_C]$$

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

(b)

$$B = \{1, x+1, x^2-1\}, C = \{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}\}$$

Answer

$$A_{BC} = [(T(\underline{b_1}))_C(T(\underline{b_2}))_C(T(\underline{b_3}))_C]$$

$$= [T(1)_C T(x+1)_C T(x^2-1)_C]$$

$$= [\begin{bmatrix}1\\1\end{bmatrix}_C \begin{bmatrix}1\\2\end{bmatrix}_C \begin{bmatrix}-1\\0\end{bmatrix}_C]$$

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

3. We have to think about this. (The horror.)

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

and $T = T_A : \mathbb{R}^3 \to \mathbb{R}^3$.

Find bases B and C such that

$$A_{BC} = I_3$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 21: Inner Products, Lengths and Orthogonality

Lecturer: Alexander Paulin *March 9* *Aditya Sengupta*

21.1 Lengths and Angles

We want to introduce the familiar concepts of lengths and angles into the theory of vector spaces, because we know how to deal with those.

We know what the length of a vector is in $\mathbb{R}^n$. For example, in $\mathbb{R}^2$, we can represent the length of a vector $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ as $||\underline{u}||$. We can also find an angle between two vectors, and the distance between two vectors $||\underline{u} - \underline{v}||$.

From Pythagoras' theorem, we know that $||\underline{u}|| = \sqrt{u_1^2 + u_2^2}$. This tells us that the distance between two vectors $\underline{u}$ and $\underline{v}$ is

$$||\underline{u} - \underline{v}|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$$

Trigonometry tells us

$$||\underline{u} - \underline{v}||^2 = ||\underline{u}||^2 + ||\underline{v}||^2 - 2||\underline{u}|| \cdot ||\underline{v}|| \cos\theta$$

This gives us the result

$$u_1 v_1 + u_2 v_2 = ||\underline{u}||||\underline{v}|| \cos\theta$$

which gives us a relation between side lengths and the angle θ.

21.2 Basic formulas that hold in $\mathbb{R}^3$

Definition 41. *Given* $\underline{u} = \begin{bmatrix} u_1 \\ \dots \\ u_n \end{bmatrix}$, $\underline{v} = \begin{bmatrix} v_1 \\ \dots \\ v_n \end{bmatrix}$, *in* $\mathbb{R}^n$, *the scalar product is the number*

$$\underline{u} \cdot \underline{v} = \sum_{i=1}^{n} u_i v_i \tag{21.1}$$

We can think about a vector in $\mathbb{R}^n$ as an $n \times 1$ matrix, which we can switch to a $1 \times n$ matrix. Then, the inner product (which is the same as the dot product) can be expressed as a matrix multiplication:

$$\underline{u} \cdot \underline{v} = u_1 v_1 + \cdots + u_n v_n = (u_1 \dots u_n) \cdot \begin{bmatrix} v_1 \\ \dots \\ v_n \end{bmatrix} = \underline{u}^T \underline{v}$$

In $\mathbb{R}^2$ or $\mathbb{R}^3$, we know the following:

1. length of $\underline{u}$ is equal to $\sqrt{\underline{u} \cdot \underline{u}}$.
2. $\underline{u} \cdot \underline{v} =$ (length of $\underline{u}$)$\times$(length of $\underline{v}$) $\times \cos\theta$

21.3 Properties of Standard Inner Product on $\mathbb{R}^n$

1. $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$
2. $(\underline{u} + \underline{v}) \cdot \underline{w} = \underline{u} \cdot \underline{w} + \underline{v} \cdot \underline{w}$
3. $(\lambda\underline{u}) \cdot \underline{w} = \lambda(\underline{u} \cdot \underline{w})$
4. $\underline{u} \cdot \underline{u} \geq 0$
5. $\underline{u} \cdot \underline{u} = 0 \iff \underline{u} = 0$

21.4 Super Obvious Definitions

Definition 42. *Given* $\underline{u} \in \mathbb{R}^n$, $||\underline{u}|| := \sqrt{\underline{u} \cdot \underline{u}}$ *is the length (norm) of* $\underline{u}$.

Definition 43. $\underline{u}$ *is called a unit vector if* $||\underline{u}|| = 1$

For example, $\{\underline{e_1}, \ldots, \underline{e_n}\} \subset \mathbb{R}^n$ are all unit vectors.

We can infer that

$$||\lambda\underline{u}|| = \sqrt{(\lambda\underline{u}) \cdot (\lambda\underline{u})} = |\lambda| \cdot ||\underline{u}||$$

We can construct a unit vector in any direction by *normalising* it: for any $\underline{u}$ such that $||\underline{u}|| \neq 0$, we know that

$$\frac{\underline{u}}{||\underline{u}||}$$

is a unit vector.

For example, if $\underline{u} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}$, we can see that its magnitude is 6, therefore the normalised unit vector is

$$\begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 0 \\ -1/\sqrt{6} \end{bmatrix}$$

21.5 Distance

Given $\underline{u}$, $\underline{v}$ in $\mathbb{R}^n$, $dist(\underline{u}, \underline{v}) = ||\underline{u} - \underline{v}||$.

For example, $\underline{u} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \underline{v} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 3 \end{bmatrix}$. Then the distance is

$$dist(\underline{u}, \underline{v}) = \sqrt{(1-(-1))^2 + 2^2 + 0^2 + (-1-3)^2} = \sqrt{24}$$

21.6 Orthogonality

Given $\underline{u}, \underline{v} \in \mathbb{R}^n$, we say that $\underline{u}, \underline{v}$ are orthogonal if and only if $\underline{u} \cdot \underline{v} = 0$. In $\mathbb{R}^2$ and $\mathbb{R}^3$, this implies that $\cos\theta = 0 \implies \theta = \frac{\pi}{2}$

This gives us a useful property. For orthogonal vectors $\underline{u}$ and $\underline{v}$,

$$||\underline{u} + \underline{v}||^2 = \underline{u} \cdot \underline{u} + \underline{u} \cdot \underline{v} + \underline{v} \cdot \underline{u} + \underline{v} \cdot \underline{v} = ||\underline{u}||^2 + ||\underline{v}||^2$$

This is a generalisation of Pythagoras' Theorem.

21.7 More definitions yayyy

Given $W \subset \mathbb{R}^n$, we define

$$W^\perp := \{\underline{u} \in \mathbb{R}^n s.t. \underline{u} \cdot \underline{w} = 0 \forall \underline{w} \in W\}$$

For example, if W is a line which is a subspace of $\mathbb{R}^2$, $W^\perp$ is the line at 90 degrees to it.

21.7.1 Properties of $W^\perp$

1. $W^\perp$ is a subspace.
2. If $W = Span(\underline{v_1} \dots \underline{v_p})$, $\underline{u} \in W^\perp \iff \underline{u} \cdot \underline{v_i} = 0 \forall i$

Theorem 21.1. *Let A be an $m \times n$ matrix. Then*

$$(Col(A))^\perp = Nul(A^T)$$

$$(Nul(A))^\perp = Col(A^T)$$

Proof. $A = (a_1 \dots a_n) \implies A^T = \begin{bmatrix} \underline{a_1}^T \\ \vdots \\ \underline{a_n}^T \end{bmatrix}$

$\underline{u} \in (Col(A))^{\perp} \iff \underline{a_i} \cdot \underline{u} = 0 \quad \forall i \iff \underline{a_i}^T \underline{u} = 0 \quad \forall i \iff A^T \underline{u} = \underline{0} \iff \underline{u} \in Nul(A^T)$

Therefore,

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

$\square$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 22: Orthogonal Sets and Matrices

Lecturer: Alexander Paulin *12 March* *Aditya Sengupta*

22.1 Definition of orthogonal and orthonormal sets

Definition 44. $\{\underline{u_1}, \ldots, \underline{u_p}\} \subseteq \mathbb{R}^n$ *is an orthogonal set* $\iff$ $\underline{u_i} \cdot \underline{u_j} = 0$ *if* $i \neq j$.

Definition 45. $\{\underline{u_1}, \ldots, \underline{u_p}\} \subseteq \mathbb{R}^n$ *is an orthonormal set* $\iff$ *it is orthonormal and* $||\underline{u_i}|| = 1 \, \forall i$.

Any subset of the standard basis $\{\underline{e_1}, \ldots, \underline{e_n}\} \subset \mathbb{R}^n$ is an orthonormal set, for example.

Theorem 22.1. *Let* $\{\underline{u_1}, \ldots, \underline{u_p}\} \subseteq \mathbb{R}^n$ *be an orthogonal set of nonzero vectors.*

$$\underline{v} = \lambda_1 \underline{u_1} + \cdots + \lambda_p \underline{u_p} \implies \lambda_i = \frac{\underline{v} \cdot \underline{u_i}}{\underline{u_i} \cdot \underline{u_i}}$$

Proof.

$$\underline{v} = \lambda_1 \underline{u_1} + \cdots + \lambda_p \underline{u_p}$$

$$\implies \underline{v} \cdot \underline{u_i} = \lambda_1 (\underline{u_1} \cdot \underline{u_i}) + \cdots + \lambda_i (\underline{u_i} \cdot \underline{u_i}) + \lambda_p (\underline{u_p} \cdot \underline{u_i})$$

$$= \lambda_i (\underline{u_i} \cdot \underline{u_i})$$

$$\therefore \lambda_i = \frac{\underline{v} \cdot \underline{u_i}}{\underline{u_i} \cdot \underline{u_i}}$$

□

22.2 A pretty serious consequence

Lemma 22.2. *Any orthogonal set of nonzero vectors is automatically linearly independent.*

Proof.

$$\lambda_1 \underline{u_1} + \cdots + \lambda_p \underline{u_p} = \underline{0} \implies \lambda_i = \frac{\underline{0} \cdot \underline{u_i}}{\underline{u_i} \cdot \underline{u_i}} = 0$$

□

Example

Determine $\begin{bmatrix}1\\2\end{bmatrix}_\beta$ where $\beta = \left\{ \begin{bmatrix}3\\-1\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix} \right\}$.

Observe that the vectors in the basis of β are orthogonal, but not orthonormal.

$$\begin{bmatrix}1\\2\end{bmatrix} = \lambda_1 \begin{bmatrix}3\\-1\end{bmatrix} + \lambda_2 \begin{bmatrix}1\\3\end{bmatrix} \iff \begin{bmatrix}1\\2\end{bmatrix}_\beta = \begin{bmatrix}\lambda_1\\\lambda_2\end{bmatrix}$$

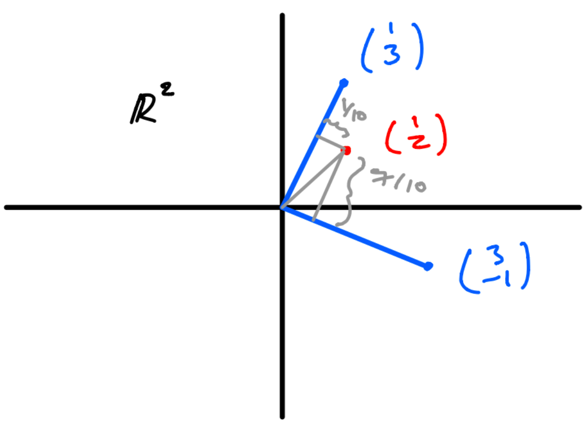
Then we can find the λs using the above method:

$$\lambda_1 = \frac{\begin{bmatrix}1\\2\end{bmatrix} \cdot \begin{bmatrix}3\\-1\end{bmatrix}}{\begin{bmatrix}3\\-1\end{bmatrix} \cdot \begin{bmatrix}3\\-1\end{bmatrix}} = \frac{1}{10}$$

$$\lambda_2 = \frac{\begin{bmatrix}1\\2\end{bmatrix} \cdot \begin{bmatrix}1\\3\end{bmatrix}}{\begin{bmatrix}1\\3\end{bmatrix} \cdot \begin{bmatrix}1\\3\end{bmatrix}} = \frac{7}{10}$$

Therefore $\begin{bmatrix}1\\2\end{bmatrix}_\beta = \begin{bmatrix}1/10\\7/10\end{bmatrix}$.

Visualization



Definition 46. $\{\underline{u_1}, \ldots, \underline{u_n}\}$ *is an orthogonal basis* $\iff$ *it is an orthogonal set and it is a basis.*

omg what a shock

bet you can't guess what an orthonormal basis is

Definition 47. $\{\underline{u_1}, \ldots, \underline{u_n}\}$ *is an orthonormal basis* $\iff$ *it is an orthonormal set and it is a basis.*

For instance, in the above example, $\left\{ \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\}$ is an orthogonal basis, and it can be scaled by $\sqrt{10}$ to make it orthonormal. The standard basis for a subset of $\mathbb{R}^n$ is orthonormal.

22.3 Something Incredibly Neat

Let $A = \{\underline{a_1}, \dots, \underline{a_n}\}$, where $\underline{a_i}$ are vectors in $\mathbb{R}^m$, so A is an $m \times n$ matrix. Then,

$$A^T = \begin{bmatrix} \underline{a_1}^T \\ \vdots \\ \underline{a_n}^T \end{bmatrix}$$

This is an $n \times m$ matrix, which can be multiplied by the original matrix. This gives us a matrix,

$$A^T A = \begin{bmatrix} \underline{a_1}^T \underline{a_1} & \underline{a_1}^T \underline{a_2} & \dots & \underline{a_1}^T \underline{a_n} \\ \underline{a_2}^T \underline{a_1} & \underline{a_2}^T \underline{a_2} & \dots & \underline{a_2}^T \underline{a_n} \\ \vdots & \vdots & \vdots & \vdots \\ \underline{a_n}^T \underline{a_1} & \underline{a_n}^T \underline{a_2} & \dots & \underline{a_n}^T \underline{a_n} \end{bmatrix}$$

By definition, this is equal to

$$\begin{bmatrix} \underline{a_1} \cdot \underline{a_1} & \underline{a_1} \cdot \underline{a_2} & \dots & \underline{a_1} \cdot \underline{a_n} \\ \underline{a_2} \cdot \underline{a_1} & \underline{a_2} \cdot \underline{a_2} & \dots & \underline{a_2} \cdot \underline{a_n} \\ \vdots & \vdots & \vdots & \vdots \\ \underline{a_n} \cdot \underline{a_1} & \underline{a_n} \cdot \underline{a_2} & \dots & \underline{a_n} \cdot \underline{a_n} \end{bmatrix}$$

This gives us the following:

$U = (\underline{u_1}, \dots, \underline{u_n})$ - $m \times n$ matrix.

1. the set of the columns of U is an orthogonal set if and only if $\underline{u_i} \cdot \underline{u_j} = 0 \, \forall i \neq j$, which in turn is true if and only if $U^T U$ is diagonal.
2. the set of the columns of U is an orthonormal set if and only if $U^T U = I_n$.

22.4 I don't know how to write in red in LaTeXsend help pls

If $m \neq n$, U is not square, therefore $UU^T \neq I_m$.

3. Let the set of the columns of U be an orthonormal set, and take $\underline{x}, \underline{y} \in \mathbb{R}^n$, then

$$(U\underline{x}) \cdot (U\underline{y}) = (U\underline{x})^T (U\underline{y}) = \underline{x}^T U^T U \underline{y} = \underline{x}^T \underline{y} = \underline{x} \cdot \underline{y}$$

That is, multiplication by U preserves the standard inner product.

4. $$(U\underline{x}) \cdot (U\underline{x}) = \underline{x} \cdot \underline{x} \implies ||U\underline{x}||_{\mathbb{R}^m} = ||\underline{x}||_{\mathbb{R}^n}$$

   This means the underlying linear transformation $T_u : \mathbb{R}^n \to \mathbb{R}^m, \underline{x} \to U\underline{x}$ preserves the standard inner product, lengths, distances, and orthogonality.

5. If U is a square matrix, with dimensions $n \times n$, then its columns form an orthonormal basis if and only if $U^T U = I_n$, therefore $U^T = U^{-1}$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 23: π Day (Orthogonal Projection)

Lecturer: Alexander Paulin *14 March* *Aditya Sengupta*

23.1 Minimum distance between a point and W

Let $\hat{y}$ be a point in W such that $\underline{y} - \hat{\underline{y}}$ is orthogonal to every $\underline{w} \in W$. Then the minimum distance to $\underline{y}$ is $||\underline{y} - \hat{\underline{y}}||$. Therefore $\underline{y} - \hat{\underline{y}} \in W^\perp$. So we want to find $\hat{\underline{y}} \in W$ such that $\underline{y} - \hat{\underline{y}} \in W^\perp$.

Let $\{\underline{u_1}, \dots, \underline{u_p}\}$ be an orthogonal basis for W. We want to find $\lambda_1, \dots, \lambda_p$ such that

1. $\hat{\underline{y}} = \lambda_1 \underline{u_1} + \cdots + \lambda_p \underline{u_p}$
2. $(\underline{y} - \hat{\underline{y}}) \cdot \underline{u_i} = 0, \ \forall i$

Then,

$$(\underline{y} - \lambda_1 \underline{u_1} - \cdots - \lambda_p \underline{u_p}) \cdot \underline{u_i} = \underline{y} \cdot \underline{u_i} - \lambda_1 (\underline{u_1} \cdot \underline{u_i}) - \cdots - \lambda_i (\underline{u_i} \cdot \underline{u_i}) - \cdots - \lambda_p (\underline{u_p} \cdot \underline{u_i})$$

By orthogonality, this is

$$\underline{y} \cdot \underline{u_i} - \lambda_i (\underline{u_i} \cdot \underline{u_i}) = 0$$

$$\therefore \lambda_i = \frac{\underline{y} \cdot \underline{u_i}}{\underline{u_i} \cdot \underline{u_i}}$$

Definition 48. *If $W \subset \mathbb{R}^n$ is a subspace with orthogonal basis $\{\underline{u_1}, \dots, \underline{u_p}\}$, the orthogonal projection of $\underline{y}$ onto W is the vector*

$$\text{proj}_W(\underline{y}) := \sum_{i=1}^{p} \frac{\underline{y} \cdot \underline{u_i}}{\underline{u_i} \cdot \underline{u_i}} \underline{u_i}$$

23.2 Important Facts

1. Given $W \subset \mathbb{R}^n$ we can always find an orthogonal basis for W.
2. $\text{proj}_W(\underline{y})$ is independent of our choice of orthogonal basis.
3. There is one and only one way to write $\underline{y} = \underline{z_1} + \underline{z_2}$ where $\underline{z_1} \in W$, $\underline{z_2} \in W^\perp$.
   We can visually see that $\underline{z_1} = \text{proj}_W(\underline{y})$ and $\underline{z_2} = \text{proj}_W^\perp(\underline{y})$.
4. The projection operation ($\mathbb{R}^n \to W$, $\underline{y} \to \text{proj}_W(\underline{y})$) is linear and onto.
5. $||\underline{y} - \text{proj}_W(\underline{y})|| \leq ||\underline{y} - \underline{w}|| \forall \underline{w} \in W$.

23.3 omg numbers what a weird idea

$\underline{y} = \begin{bmatrix}1\\2\\3\end{bmatrix}$, $\underline{u_1} = \begin{bmatrix}1\\-1\\1\end{bmatrix}$, $\underline{u_2} = \begin{bmatrix}-1\\0\\1\end{bmatrix}$. Let W be the span of $\underline{u_1}$ and $\underline{u_2}$.

We can calculate $\mathrm{proj}_W(\underline{y}) = \frac{\underline{y}\cdot\underline{u_1}}{\underline{u_1}\cdot\underline{u_1}}\underline{u_1} + \frac{\underline{y}\cdot\underline{u_2}}{\underline{u_2}\cdot\underline{u_2}}\underline{u_2} = \frac{2}{3}\begin{bmatrix}1\\-1\\1\end{bmatrix} + \frac{2}{2}\begin{bmatrix}-1\\0\\1\end{bmatrix} = \begin{bmatrix}-1/3\\-2/3\\5/3\end{bmatrix}$

Then the minimum distance is

$$\left|\left|\begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}-1/3\\-2/3\\5/3\end{bmatrix}\right|\right| = 7/3$$

(maybe)

Observation

If $\underline{u_1}, \dots, \underline{u_p}$ is an orthonormal basis of W, the projection formula becomes simpler:

$$\mathrm{proj}_W(\underline{y}) = \sum_{i=1}^{p}(\underline{y}\cdot\underline{u_i})\underline{u_i}$$

Let $U = (\underline{u_1}, \dots, \underline{u_p})$. Then $\mathrm{proj}_W(\underline{y}) = U\left(\begin{bmatrix}\underline{u_1}\cdot\underline{y}\\ \vdots\\ \underline{u_p}\cdot\underline{y}\end{bmatrix}\right) = U\begin{bmatrix}\underline{u_1}^T\underline{y}\\ \vdots\\ \underline{u_p}^T\underline{y}\end{bmatrix} = UU^T\underline{y}$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 24: The Gram-Schmidt Process

Lecturer: Alexander Paulin *16 March* *Aditya Sengupta*

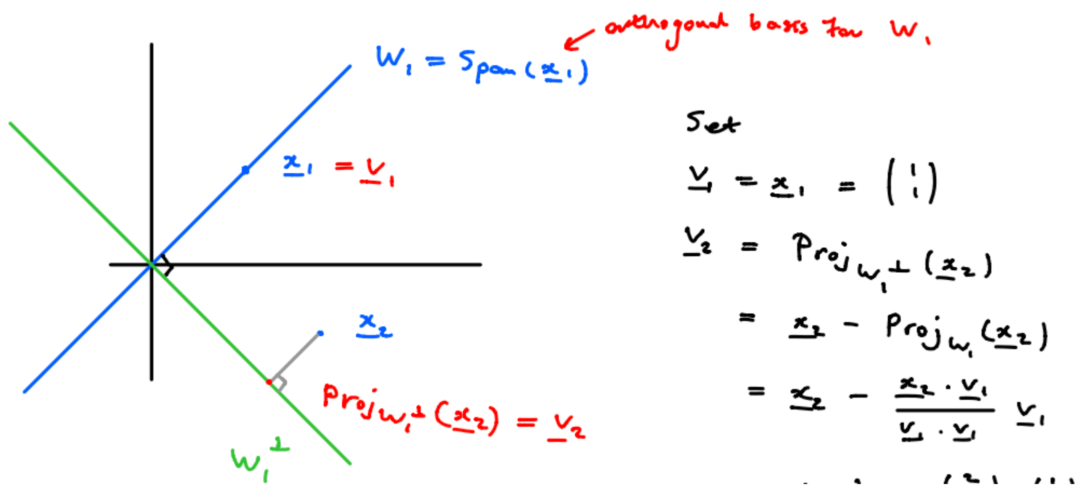
24.1 Aim

Given $W \subset \mathbb{R}^n$, find $\{\underline{v_1}, \ldots, \underline{v_p}\}$, an orthogonal basis for W.

The algorithm to construct this basis is called the Gram-Schmidt Process.

24.2 Example

Take $W = \mathbb{R}^2$, $W = \text{span}(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix}) = \underline{x_1}, \underline{x_2}$.



We can take $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ as the first vector in the orthogonal basis. Then, by construction, we can say that $\text{proj}_{W_1^\perp}(\begin{bmatrix} 2 \\ -1 \end{bmatrix}) = \underline{x_2} - Proj_{W_1}(\underline{x_2})$ is perpendicular to $\underline{x_1}$ and the two form an orthogonal basis.

In this case, this is

$$\underline{x_2} - \frac{\underline{x_2} \cdot \underline{v_1}}{\underline{v_1} \cdot \underline{v_1}} \underline{v_1} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/2 \\ -3/2 \end{bmatrix}$$

We note that $Span(x_1) = Span(v_1)$, and that $Span(\underline{x_1}, \underline{x_2}) = Span(\underline{v_1}, \underline{v_2}) = W$.

24.3 Generalising

In general, we want to build $\{\underline{v_1}, \ldots, \underline{v_p}\}$ starting from $\{\underline{x_1}, \ldots, \underline{x_p}\}$. We start with $\underline{v_1} = \underline{x_1}$, construct $\underline{v_k}$ using $\underline{v_1}, \ldots, \underline{v_{k-1}}$ and $\underline{x_k}$.

1. Let $W_{k-1} = Span(\underline{v_1}, \ldots, \underline{v_{k-1}})$.
2. Let $\underline{v_k} := Proj_{W_{k-1}^{\perp}}(\underline{x_k}) = \underline{x_k} - \sum_{i=1}^{k-1} \frac{\underline{x_k} \cdot \underline{v_i}}{\underline{v_i} \cdot \underline{v_i}}$

24.4 Facts

1. $\{\underline{x_1}, \ldots, \underline{x_p}\}$ basis for $W \implies \{\underline{v_1}, \ldots, \underline{v_p}\}$ basis for W
2. The span of the x vectors and that of the v vectors are equal.
3. To apply the Gram-Schmidt process, we must start with a linearly independent set. If $\underline{x_k}$ is in the span of the first $k-1 x-$vectors, then the projection on W_{k-1} of $\underline{x_k}$ is zero, therefore we have a zero vector in the basis and we cannot continue the process.

Theorem 24.1. *Any subspace $W \subset \mathbb{R}^n$ has an orthogonal basis.*

Proof. Choose any basis, and apply the Gram-Schmidt process. By construction, an orthogonal basis exists. ☐

Remark 24.2. *Once we have an orthogonal basis, we can scale each vector to get an orthonormal basis.*

24.5 Interesting Consequence

Suppose we carry out the GS process on some basis A, to get an orthonormal basis Q. The two have the same span, therefore we can write

$$\underline{x_k} = \lambda_{1k}\underline{u_1} + \cdots + \lambda_{kk}\underline{u_k} + 0\underline{u_{k+1}} + \cdots + 0\underline{u_p}$$

which we can put into a matrix:

$$R = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & \ldots & \lambda_{1p} \\ 0 & \lambda_{22} & \lambda_{23} & \ldots & \vdots \\ \vdots & 0 & \lambda_{33} & & \vdots \\ \vdots & \vdots & 0 & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_{pp} \end{bmatrix}$$

Then observe that $Q\underline{r_k} = \lambda_{1k}\underline{u_1} + \cdots + \lambda_{kk}\underline{u_k} + 0\underline{u_{k+1}} + \cdots + 0\underline{u_p} = \underline{x_k}$.

So,

$$QR = (Q\underline{r_1}Q\underline{r_2}\dots Q\underline{r_p}) = (\underline{x_1 x_2}\dots \underline{x_p}) = A$$

Remark 24.3. *The columns of Q being an orthonormal set implies $Q^T Q = I_p$*

$$A = QR = Q^T A = Q^T QR = R$$

In conclusion, given $A_{n\times p}$ (L.I.) we can find $Q_{n\times p}$ with orthonormal columns and $R_{p\times p}$ upper triangular such that $A = QR$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 25: Least Squares Solutions

Lecturer: Alexander Paulin *19 March* *Aditya Sengupta*

Definition 49. *A least-squares solution to the system $A\underline{x} = \underline{b}$, where A is an $m \times n$ matrix and $\underline{b} \in \mathbb{R}^m$, is a vector $\underline{\hat{x}} \in \mathbb{R}^n$ such that*

$$\| \underline{b} - A\underline{\hat{x}} \| \leq \| \underline{b} - A\underline{x} \| \; \forall \underline{x} \in \mathbb{R}^n$$

Remark 25.1. *If $A\underline{x} = \underline{b}$ consistent, choose $\underline{\hat{x}}$ such that $A\underline{\hat{x}} = \underline{b}$, i.e. we choose an actual solution. This reduces the least-squares error to zero.*

Recall:

1. $W \subset \mathbb{R}^n$ subspace, $\underline{\hat{b}} \in W$ and $\underline{b} \in \mathbb{R}^m$, then

$$||\underline{b} - \underline{\hat{b}}|| \leq ||\underline{b} - \underline{w}|| \forall \underline{w} \in W \iff \underline{\hat{b}} = \text{proj}_W(\underline{b}) \iff \underline{b} - \underline{\hat{b}} \in W^{\perp}$$

2. $Col(A) := \{A\underline{x} \text{ such that } \underline{x} \in \mathbb{R}^n\}$

Therefore,

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

for all $\underline{x} \in \mathbb{R}^n$. Both of these vectors are in $Col(A)$.

This means $A\underline{\hat{x}} = Proj_{Col(A)}\underline{b}$, which is true if and only if $\underline{b} - A\underline{\hat{x}} \in (Col(A))^{\perp}$.

25.1 Concreteness

Let $A = (\underline{a_1}, \ldots, \underline{a_n})$. This implies $Col(A) = Span(\underline{a_1}, \ldots, \underline{a_n})$.

This means, for a vector to be in $(Col(A))^{\perp}$, it has to be orthogonal to each basis vector. Therefore

$$\underline{a_i} \cdot (\underline{b} - A\underline{\hat{x}}) = 0 \forall i$$

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

which is true if and only if

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

This gives us

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

Remark 25.2. *This also follows from the fact that*

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

25.2 Conclusion

$\underline{\hat{x}}$ is a least-squares solution to $A\underline{x} = \underline{b}$ if and only if $\underline{\hat{x}}$ is a solution to $(A^TA)\underline{x} = A^T\underline{b}$.

25.3 Example

Let $A = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix}$ and $\underline{b} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$. Find the least-squares solution.

We find the normal equations,

$$\left(\begin{bmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{bmatrix} \cdot \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix} \right) \underline{\hat{x}} = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 6 & -11 \\ -11 & 22 \end{bmatrix} \underline{\hat{x}} = \begin{bmatrix} -4 \\ 11 \end{bmatrix}$$

Then, we can solve this by row reduction, but because the A matrix is invertible, we could also multiply by its inverse:

$$\underline{\hat{x}} = \begin{bmatrix} 2 & 1 \\ 1 & \frac{6}{11} \end{bmatrix} \cdot \begin{bmatrix} 4 \\ -11 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

(double check calculation here)

Therefore, $\underline{\hat{x}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ is a least squares solution to $A\underline{x} = \underline{b}$.

Remark 25.3. *$A\underline{x} = \underline{b}$ has a unique least squares solution if and only if $A\underline{\hat{x}} = \underline{b}$ has a unique solution. This in turn is true if and only if the columns of A are linearly independent, which means A^TA is invertible.*

In this case $\underline{\hat{x}} = (A^TA)^{-1}A^T\underline{b}$.

Recall from the Gram-Schmidt process that $A_{m\times n} = Q_{m\times n}R_{n\times n}$, where Q is orthonormal and R is upper triangular. Therefore we can say

$$A^T = R^TQ^T \implies A^TA = R^TQ^TQR = R^TR$$

which implies

$$\underline{\hat{x}} = (A^TA)^{-1}A^T\underline{b} = (R^TR)^{-1}R^TQ^T\underline{b} = R^{-1}(R^T)^{-1}R^TQ^T\underline{b} = R^{-1}Q^T\underline{b}$$

Example

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

We apply the Gram-Schmidt process and express A as the product of matrices Q and R:

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

Then, to find $\underline{\hat{x}}$, we find $Q^T\underline{b}$:

$$Q^T\underline{b} = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

Then, $\underline{\hat{x}} = R^{-1}Q^T\underline{b}$ implies that $\underline{\hat{x}}$ is the solution to the system $R\underline{\hat{x}} = Q^T\underline{b}$. So we row reduce the following system:

$$\left[\begin{array}{ccc|c} 2 & 4 & 5 & 6 \\ 0 & 2 & 3 & -6 \\ 0 & 0 & 2 & 4 \end{array}\right]$$

which yields $\underline{\hat{x}} = \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 26: Inner Product Spaces

Lecturer: Alexander Paulin *March 21* *Aditya Sengupta*

26.1 Extending the Standard Inner Product

Recall that the standard inner product on $\mathbb{R}^n$ is a function that, to each pair of vectors $\underline{u}, \underline{v} \in \mathbb{R}^n$, assigns a real number $\underline{u} \cdot \underline{v}$.

Useful Properties:

1. $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$
2. $(\underline{u} + \underline{v}) \cdot \underline{w} = \underline{u} \cdot \underline{w} + \underline{v} \cdot \underline{w}$
3. $(\lambda \underline{u}) \cdot \underline{v} = \lambda(\underline{u} \cdot \underline{v})$
4. $\underline{u} \cdot \underline{u} \geq 0$, and $\underline{u} \cdot \underline{u} = 0 \iff \underline{u} = 0$.

If a function satisfies these properties, it is a valid inner product. More concretely, we can define an inner product as follows:

Definition 50. *Let V be a vector space. An inner product on V is a function that, to each pair of vectors $\underline{u}, \underline{v} \in V$, assigns a real number $\langle \underline{u}, \underline{v} \rangle$ such that*

1. $\langle \underline{u}, \underline{v} \rangle = \langle \underline{v}, \underline{u} \rangle$
2. $\langle (\underline{u} + \underline{v}), \underline{w} \rangle = \langle \underline{u}, \underline{w} \rangle + \langle \underline{v}, \underline{w} \rangle$
3. $\langle (\lambda \underline{u}), \underline{v} \rangle = \lambda \langle \underline{u}, \underline{v} \rangle$
4. $\langle \underline{u}, \underline{u} \rangle \geq 0$, *and* $\langle \underline{u}, \underline{u} \rangle = 0 \iff \underline{u} = 0$.

We refer to any vector space that is equipped with some inner product as an *inner product space.*

So far, we've been working mostly with $\mathbb{R}^n$ equipped with the standard inner product, but we can extend all of the familiar terminology to any inner product space. The **length**, or norm, of a vector v in any inner product space is $||\underline{v}|| = \sqrt{\langle \underline{u}, \underline{u} \rangle}$. The **distance** between two vectors $\underline{u}$ and $\underline{v}$ is $||\underline{u} - \underline{v}||$, where length is defined as above. $\underline{u}$ and $\underline{v}$ are **orthogonal** if their inner product is zero.

All of the concepts based on these in $\mathbb{R}^n$, such as orthogonal complements, projections, and the Gram-Schmidt process, turn out to work in any inner product space with the above definitions.

26.2 Examples of Other Inner Products

26.2.1 Polynomials

Let $V = \mathbb{P}_n(\mathbb{R})$, and let $t_0, t_1, \ldots, t_n$ be distinct, fixed real numbers. Then, we can define

$$\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \cdots + p(t_n)q(t_n)$$

We can see this is an inner product. It can easily be shown that it is commutative, distributive, and supports scalar multiplication. For the fourth property, we consider

$$\langle p, p \rangle = (p(t_0))^2 + (p(t_1))^2 + \cdots + (p(t_n))^2$$

These are all positive numbers, therefore the inner product of a polynomial with itself is greater than 0, as required. The inner product is only equal to zero if each component is zero, i.e. $p(t_0) = p(t_1) = \cdots = p(t_n) = 0$. Since $p \in \mathbb{P}_n(\mathbb{R})$, if it has $n+1$ distinct zeroes then it must be the zero polynomial.

26.2.2 Nonstandard Inner Products on $\mathbb{R}^n$

Using the first three properties of inner products, we can find $\langle u, v \rangle$ for any $\underline{u}$ and $\underline{v}$ if we know $\langle e_i, e_j \rangle$ for all i, j. For example,

$$\left\langle \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \left\langle 2\underline{e_1} + 3\underline{e_2}, \underline{e_1} + \underline{e_2} \right\rangle = 2\left\langle \underline{e_1}, \underline{e_1} \right\rangle + 2\left\langle \underline{e_1}, \underline{e_2} \right\rangle + 3\left\langle \underline{e_2}, \underline{e_1} \right\rangle + 2\left\langle \underline{e_1}, \underline{e_1} \right\rangle$$

An inner product can therefore be expressed in terms of matrix multiplication. Let

$$A = \begin{bmatrix} \left\langle \underline{e_1}, \underline{e_1} \right\rangle & \cdots & \left\langle \underline{e_1}, \underline{e_n} \right\rangle \\ \vdots & & \\ \left\langle \underline{e_n}, \underline{e_1} \right\rangle & \cdots & \left\langle \underline{e_n}, \underline{e_n} \right\rangle \end{bmatrix}$$

A is symmetric because inner products are always commutative. Then, we can write

$$\langle \underline{u}, \underline{v} \rangle = \underline{u}^T A \underline{v}$$

where, for the standard inner product, A is the identity matrix.

The above equation defines an inner product if and only if $A = A^T$ and the eigenvalues of A are strictly positive.

26.2.3 Inner Product on Functions

Let $V = C[a, b]$, the space of functions that are continuous on the interval $[a, b]$. Then, we can define

$$\langle f(x), g(x) \rangle := \int_a^b f(x)g(x)dx$$

26.3 Using Nonstandard Inner Products

If $W = Span\,\{ 1, x^2 \} \subset C[-1, 1]$, find the projection of x^3 onto W.

We first need to apply Gram-Schmidt to the basis for W, to make it orthogonal. We use the above integral definition of an inner product to do this.

$$\underline{v_1} = 1$$

$$\underline{v_2} = x^2 - \frac{\int_{-1}^{1} x^2 \cdot 1 dx}{\int_{-1}^{1} 1 \cdot 1 dx} 1 = x^2 - \frac{1}{3}$$

Then, we can calculate the projection:

$$\text{proj}_W(x^3) = \frac{\int_{-1}^{1} x^3 \cdot 1 dx}{\int_{-1}^{1} x^3 \cdot 1 dx} - \frac{\int_{-1}^{1} x^3 (x^2 - 1/3) dx}{\int_{-1}^{1} (x^2 - 1/3)^2 \, dx} = \frac{90}{56} \left(x^2 - \frac{1}{3} \right)$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 27: Diagonalization of Symmetric Matrices

Lecturer: Alexander Paulin *April 2* *Aditya Sengupta*

27.1 Symmetric Matrices

Let A be an $n \times n$ matrix. We say that A is symmetric if and only if $A = A^T$. That is, $A_{ij} = A_{ji} \forall i, j \leq n$.

For example, the 2×2 matrix $\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ is symmetric. But the 3×3 matrix $\begin{bmatrix} 2 & 1 & 2 \\ 1 & 3 & 3 \\ 2 & 4 & -1 \end{bmatrix}$ is not symmetric.

Any diagonal matrix, in which the nondiagonal elements are all zero, is automatically symmetric, which is easy to see considering that $A_{ij} = 0 \forall i \neq j$ and $A_{ii} = \lambda_i$ is the definition of a diagonal matrix.

27.2 Properties of Symmetric Matrices

Definition 51. *Let A be an $n \times n$ matrix. We say that A is orthogonally diagonalizable if and only if there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.*

Example

A diagonal matrix, $\begin{bmatrix} \lambda_1 & 0 & 0 \\ \dots & \vdots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix}$, is orthogonally diagonalizable with an orthonormal basis of eigenvectors $\{\underline{e_1}, \dots, \underline{e_n}\}$, which is the standard basis.

Recall that $\{\underline{u_1}, \dots, \underline{u_n}\} \subset \mathbb{R}^n$ is an orthonormal basis if and only if $P = (\underline{u_1}, \dots, \underline{u_n})$ is invertible with $P^{-1} = P^T$. P is an orthogonal matrix.

We can observe that A is orthogonally diagonalizable if and only if there exists P, an orthogonal basis, such that $P^T AP$ is diagonal.

Theorem 27.1. *If A is orthogonally diagonalizable, then A is symmetric.*

Proof. Let $\{\underline{u_1}, \dots, \underline{u_n}\} \subset \mathbb{R}^n$ be an orthonormal basis such that $A\underline{u_i} = \lambda_i \underline{u_i}$ for some $\lambda_i \in \mathbb{R}$ This implies that $P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 \\ \dots & \vdots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix}$, which further implies that

$$A = P \begin{bmatrix} \lambda_1 & 0 & 0 \\ \dots & \vdots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix} P^{-1} = P \begin{bmatrix} \lambda_1 & 0 & 0 \\ \dots & \vdots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix} P^T$$

Therefore, we take the transpose. Using the identity $(CB)^T = B^T C^T$:

$$A^T = (P^T)^T \begin{bmatrix} \lambda_1 & 0 & 0 \\ \dots & \vdots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix}^T P^T = P \begin{bmatrix} \lambda_1 & 0 & 0 \\ \dots & \vdots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix} P^T = A$$

$\square$

27.3 Spectral Theorem for Symmetric Matrices

Theorem 27.2. *A being symmetric implies A is orthogonally diagonalizable.*

That's so weird.

As consequences of this, A being symmetric implies the following:

1. All zeros of $det(A - xI_n)$ are real.
2. Eigenvectors from different eigenspaces are orthogonal
3. The dimension of the λ eigenspace is the algebraic multiplicity of λ for all eigenvalues

27.4 Orthogonally Diagonalizing a Symmetric Matrix

1. Calculate $det(A - xI_n)$ and find all zeroes $\lambda_1, \dots, \lambda_p$
2. Row reduce and find a basis for each of the eigenspaces $Nul(A - \lambda_i I_n)$
3. Apply Gram-Schmidt to each λ_i eigenspace basis, normalize and take the union.

   By construction, all of the resultant eigenvectors are orthonormal.

Example

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

Step 1:

The characteristic polynomial is $-x^2(x-3)$, which can be found by taking the determinant $\begin{vmatrix} 1-x & 1 & 1 \\ 1 & 1-x & 1 \\ 1 & 1 & 1-x \end{vmatrix}$.

So we can easily see that the eigenvalues are 0 and 3.

Step 2:

We row reduce $A - 0I_3$ to find the null space of the $0 - eigenspace$.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \implies \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies Nul(A - 0I_3) = Span\Bigg(\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \Bigg)$$

which we can see by taking x_2 and x_3 as free variables.

We row reduce $A - 3I_3$ to find the null space of the $3 - eigenspace$.

$$\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix} \implies x_1 = x_3, x_2 = x_3$$

which means $Nul(A - 3I_3) = Span\Bigg(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Bigg)$

Of course, we can't have nice things, so we have to do Gram-Schmidt on the first basis, and normalise the second.

Step 3:

The basis for $0 - eigenspace$ after applying Gram-Schmidt is $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix} \right\}$, which we can normalize and take the union with the normalized $3 - eigenspace$ to get the final answer:

$$P = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}$$

and this gives us $P^T AP = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 28: Singular Value Decompositions

Lecturer: Alexander Paulin *April 4* *Aditya Sengupta*

Given an $m \times n$ matrix A, we want to find orthonormal bases $B \subset \mathbb{R}^n$ and $C \subset \mathbb{R}^m$ such that $A_{B,C}$ is as simple as possible.

To do this, look at the matrix $A^T A$. This is an $n \times n$ matrix, making it square. Take its transpose:

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

Therefore, $A^T A$ is symmetric, which means it is orthogonally diagonalizable by the Spectral Theorem.

This allows us to choose a basis $B = \{\underline{v_1}, \ldots, \underline{v_n}\} \subset \mathbb{R}^n$, an orthonormal basis such that each v_i is an eigenvector, that is,

$$A^T A \underline{v_i} = \lambda_i \underline{v_i}$$

for some $\lambda_i \in \mathbb{R}$.

Left multiply this above expression by $\underline{v_i}^T$:

$$\underline{v_i}^T A^T A \underline{v_i} = \underline{v_i}^T \lambda_i \underline{v_i} \implies (A\underline{v_i})^T A\underline{v_i} = \lambda_i \underline{v_i}^T \underline{v_i}$$

We know that $\underline{x}^T \underline{y} = \underline{x} \cdot \underline{y}$, and that $||\underline{v_i}|| = 1$ as they are elements of an orthonormal basis. Therefore,

$$(A\underline{v_i}) \cdot (A\underline{v_i}) = \lambda_i \implies ||A\underline{v_i}||^2 = \lambda_i$$

This guarantees that $\lambda_i \geq 0 \forall i$. So we can reorder the eigenvalues, and assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

Definition 52. *The singular values of a matrix A are the numbers*

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

where $\sigma_i := \sqrt{\lambda_i}$.

By construction, $||A\underline{v_i}|| = \sigma_i$.

Theorem 28.1. *Assume $\sigma_1, \ldots, \sigma_r \neq 0$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$. Then, $Rank(A) = r$ and $\{A\underline{v_1}, \ldots, A\underline{v_r}\}$ is an orthogonal basis for $Col(A)$.*

Proof. We know that $\{\underline{v_1}, \ldots, \underline{v_n}\} \subset \mathbb{R}^n$ is a basis, which means that $\{A\underline{v_1}, \ldots, A\underline{v_n}\}$ spans Col(A).

We also know that $||A\underline{v_i}|| = \sigma_i \implies A\underline{v_i} = 0 \iff \sigma_i = 0$.

Therefore, $A\underline{v}_{r+1} = \cdots = A\underline{v_n} = \underline{0}$ because the corresponding singular values are zero.

This means they can be removed from the basis with no impact. Therefore, $\{A\underline{v_1}, \ldots, A\underline{v_r}\}$ spans $Col(A)$, so it is a basis.

Now, to prove that it is an orthogonal basis, we take a general inner product on all nonequal elements ($i \neq j$)

$$(A\underline{v_i}) \cdot (A\underline{v_j}) = \underline{v_i}^T A^T A\underline{v_j} = \underline{v_i}^T \lambda_j \underline{v_j} = \lambda_j(\underline{v_i} \cdot \underline{v_j}) = 0$$

Therefore, all basis elements are perpendicular to one another.

We can easily show that the rank of this matrix is r, because r is the size of the orthogonal basis constructed above. □

Example

Let $A = \begin{bmatrix} \sqrt{2} & -1/\sqrt{2} & 0 \\ \sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$. Multiply it by its transpose to get the symmetric matrix

$$A^T A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Choose the standard basis ($\underline{v_1} = \underline{e_1}$ etc). The eigenvalues of $A^T A$ are clearly $4, 1, 0$, because

$$A^T A\underline{v_1} = 4 \cdot \underline{v_1}, A^T A\underline{v_2} = 1 \cdot \underline{v_2}, A^T A\underline{v_3} = 0 \cdot \underline{v_3}$$

Therefore, the singular values are the square roots of these, $2, 1, 0$. This means $\{A\underline{v_1}, A\underline{v_2}\}$ form an orthogonal basis for $Col(A)$, taking only the nonzero singular values. Therefore

$$\left\{ \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$$

is an orthogonal basis for $Col(A)$.

Significance

$$A = \begin{bmatrix} \sqrt{2} & -1/\sqrt{2} & 0 \\ \sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

The above matrix represents a linear transformation $T_A : \mathbb{R}^3 \to \mathbb{R}^2$. Geometrically, this represents a mapping between a hollow sphere of radius 1 and a solid ellipse whose dimensions are given by the singular values.

In general, if we want to maximize $||A\underline{x}||$ where $||\underline{x}|| = 1$, this turns out to be the maximum singular value, which because of the reordering above is σ_1. This implies

$$Max\{||A\underline{x}||, where ||\underline{x}|| \leq d\} = d\sigma_1$$

and similarly

$$Min\{||A\underline{x}||, where ||\underline{x}|| = 1\} = \sigma_n$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 29: The Singular Value Decomposition continued

Lecturer: Alexander Paulin *6 April* *Aditya Sengupta*

Definition 53. *Let* $\underline{u_i} = \frac{1}{\sigma_i} A\underline{v_i}$ *for all nonzero* σ_i, $1 \leq i \leq r$.

This gives us the following:

1. $\{\underline{u_1}, \ldots, \underline{u_r}\}$ is an orthonormal basis for $Col(A)$.
2. $A\underline{v_i} = \sigma_i \underline{u_i}$ for $0 \leq i \leq r$

We can use this to extend $\{\underline{u_1}, \ldots, \underline{u_r}\}$ to an orthonormal basis for all $\mathbb{R}^m$ as follows:

1. Find a basis for $(Col(A))^\perp = Nul(A^T)$
2. Apply Gram-Schmidt and normalize to get an orthonormal basis for $(Col(A))^\perp$. This has dimension $m - r$, so we consider it to consist of elements $\{\underline{u_{r+1}}, \ldots, \underline{u_m}\}$
3. Take the union of these two to get an orthonormal basis for $\mathbb{R}^m$.

29.1 Recap

In general, an $m \times n$ matrix A represents a linear transformmation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Using the above procedure, we can make orthonormal bases for $\mathbb{R}^n$ (which turns out to be $Col(A^T) \bigcup Nul(A)$) and $\mathbb{R}^m$ (which is $Col(A) \bigcup Nul(A^T)$)

Each $\underline{v_i}$ gets transformed to $\sigma_i \underline{u_i}$ if $1 \leq i \leq r$, and 0 otherwise. This lets us write the change-of-bases transformation matrix $A_{B,C}$ explicitly as:

$$A_{B,C} = \Big((A\underline{v_1})_C \ldots (A\underline{v_r})_C (A\underline{v_{r+1}})_C \ldots (A\underline{v_n})_C \Big)$$

$$= \{(\sigma_1 \underline{u_1})_C, \ldots, (\sigma_r \underline{u_r})_C, (\underline{0})_C \underline{0}, \ldots, \underline{0}\}$$

which is something that I can't easily typeset: diagonal with elements $A_{B,Cii} = \sigma_i$ for $1 \leq i \leq r$, and every other element equal to 0.

As a result, if $U = (\underline{u_1}, \ldots, \underline{u_m})$ and $V = (\underline{v_1}, \ldots, \underline{v_n})$ are orthogonal bases for $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, then

$$U^{-1}AV = A_{BC}$$

(which is the above described diagonal-or-zero matrix, referred to as A_{SV})

Therefore,

$$A = UA_{SV}V^T = U\Sigma V^T$$

This is the *singular value decomposition* of a matrix.

29.2 Overview of finding SVD

1. Orthogonally diagonalize A^TA, which yields an orthonormal basis V, $A^TA\underline{v_i} = \lambda_i\underline{v_i}$
2. Reorder and take square roots so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} \geq \cdots \geq \sigma_n \geq 0$, and $\sigma_i = \sqrt{\lambda_i}$.
3. Define $\underline{u_i} = \frac{1}{\sigma_i}A\underline{v_i}$ for $1 \leq i \leq r$
4. Find an orthonormal basis for $(Col(A))^\perp = Nul(A^T)\{\underline{u_{r+1}}, \ldots, \underline{u_m}\}$, and take a union with the above-defined $\underline{u_i}$s to get an orthonormal basis for $\mathbb{R}^m$.
5. $U = (\underline{u_1}, \ldots, \underline{u_m})$, $V = (\underline{v_1}, \ldots, \underline{v_n})$, $\Sigma = \left[\begin{array}{cc|c} \sigma_1 & 0 & 0 \\ 0 & \ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \sigma_r & 0 \\ \hline 0 & \ldots & 0 \end{array}\right]$

   $A = U\Sigma V^T$

29.3 This is amazing. Why?

Image Processing

Imagine we have a grayscale image, with resolution 512×512 pixels. We can encode this image as a 512×512 matrix, where each entry represents the brightness of the corresponding pixel.

Then, we can decompose it into $A = U\Sigma V^T$. Since A is a square matrix, Σ will be a diagonal matrix whose entries are the singular values,

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_{512} \end{bmatrix}$$

In general, the columns are linearly independent, since not every image is completely black or completely white. So, for any $1 \leq k \leq 512$, let Σ_k be the following:

$$\Sigma_k = \begin{bmatrix} \sigma_1 & \ldots & 0 \\ \ldots & \sigma_k & \ldots \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix}$$

$k \leq Rank(A) \implies Rank(A_k) = k$

A_k is the best possible rank-k approximation to A. (add pic of compression)

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 30: Second-Order Linear Differential Equations

Lecturer: Alexander Paulin *April 9* *Aditya Sengupta*

Definition 54. *A differential equation is an equation in an unknown function y, and its derivatives $y', y'', \dots$*

Differential equations are of enormous importance to applied mathematics, such as all of physics. The general form of a linear second-order constant coefficient differential equation is as follows:

$$ay'' + by' + cy = f(t)$$

where a, b, c are constants and $a \neq 0$. The above is an equality of functions, i.e. this equation is true for all values of y we could come up with.

Our aim is to find a general solution to this. To do this, we first look at the homogeneous case where $ay'' + by' + cy = 0$.

We observe the following:

1. The zero function is a solution
2. y_1, y_2 solutions $\implies y_1 + y_2$ solution
3. y_1 solution, λ real $\implies \lambda y_1$ solution

Therefore, the solutions to $ay'' + by' + cy = 0$ are a subspace of all twice differentiable functions.

Theorem 30.1. *Fix any t_0 in $\mathbb{R}$. Then, the transformation between the subspace of all solutions to $ay'' + by' + cy = 0$ and $\mathbb{R}^2$, characterised by*

$$y(t) \to \begin{bmatrix} y(t_0) \\ y'(t_0) \end{bmatrix}$$

is a one-to-one, onto, linear transformation.

As a consequence of this, the solution subspace defined above is 2-dimensional. This means we need to find two linearly independent solutions to $ay'' + by' + cy = 0$.

Finding two linearly independent solutions, in general, is tougher than it seems. Analytically solving it requires integrals that are either difficult or provably impossible.

"Provably, everyone is screwed." - Paulin 2018

So let's guess. Try $y(t) = e^{rt}$. Then, the derivatives are

$$y'(t) = re^{rt}, y''(t) = r^2 e^{rt}$$

Therefore,

$$ay'' + by' + cy = (ar^2 + br + c)e^{rt} = 0$$

Therefore, $y(t) = e^{rt}$ is a solution to the homogeneous case if and only if $ar^2 + br + c = 0$. Call this the *auxiliary equation*, and call the polynomial $ar^2 + br + c$ the *auxiliary polynomial*. We now have three cases:

1. Two distinct real solutions r_1, r_2 ($b^2 - 4ac > 0$)

   This gives us the two linearly independent solutions we were looking for, $\{e^{r_1 t}, e^{r_2 t}\}$, which in turn gives us a general solution of $C_1 e^{r_1 t} + C_2 e^{r_2 t}$.

2. One repeated real root r_1 ($b^2 - 4ac = 0$)

   Consider $y(t) = te^{r_1 t}$. This implies that $y' = e^{r_1 t} + r_1 te^{r_1 t}$ and $y'' = 2r_1 e^{r_1 t} + r_1^2 te^{r_1 t}$. Therefore,

   $$ay'' + by' + cy = a(2r_1 e^{r_1 t} + r_1^2 + r_1^2 te^{r_1 t}) + b(e^{r_1 t} + r_1 te^{r_1 t}) + c(te^{r_1 t})$$

   $$= (2ar_1 + b)e^{r_1 t} + (ar_1^2 + br_1 + c)te^{r_1 t}$$

   $2ar_1 + b = 0$ because the discriminant of the auxiliary polynomial is 0, and $ar_1^2 + br_1 + c = 0$ because that is the original equation. Therefore $te^{r_1 t}$ is a solution, so our linearly independent set is $\{e^{r_1 t}, te^{r_1 t}\}$.

3. Two complex conjugate non-real solutions $\alpha \pm i\beta$ ($b^2 - 4ac < 0$)

   This means our proposed solution is the complex-valued function $y(t) = e^{(\alpha + i\beta)t}$. This is a bit problematic because we only want real-valued solutions. So we can do this:

   $$e^{(\alpha+i\beta)t} = e^{\alpha t} e^{i\beta t} = e^{\alpha t}\Big(\cos(\beta t) + i\sin(\beta t)\Big)$$

   From this, we get our general solution as $C_1 e^{\alpha t}\cos(\beta t) + C_2 e^{\alpha t}\sin(\beta t)$.

Example

$$y'' + y = 0 \implies r^2 + 1 = 0 \implies r = \pm i$$

This corresponds to $\alpha = 0, \beta = 1$.

This means our general solution is $C_1 \cos t + C_2 \sin t$.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 31: Linear Algebra Techniques for Second-Order ODEs

Lecturer: Alexander Paulin *April 11* *Aditya Sengupta*

31.1 Non-Homogeneous Second-Order ODEs

This is represented by

$$ay'' + by' + cy = f(t)$$

where $f(t)$ is not the zero function. To solve this, we fix y_p, a particular solution to the above. That is,

$$ay_p'' + by_p' + cy_p = f(t)$$

Let y be any other solution, i.e. $ay'' + by' + cy = f(t)$.

Then, subtracting these:

$$(ay'' + by' + cy) - (ay_p'' + by_p' + cy_p) = f(t) - f(t) = 0$$

$$\therefore a(y - y_p)'' + b(y - y_p)' + c(y - y_p) = 0$$

That is, $y_h := y - y_p$ is a solution to the homogeneous linear system. It follows that $y = y_p + y_h$. So we can conclude that:

1. A general solution to $ay'' + by' + cy = f(t)$ is $y_p + y_h$, where y_p is a particular solution and y_h is a general solution to the case $f(t) = 0$.

2. $y_h(t_0) = 0, y_h'(t_0) = 0 \implies y_h =$ zero function, so $y = y_p$.

Our problem now is that finding particular solutions for a general $f(t)$ is *really hard*. We can deal with this using the method of undetermined coefficients.

For $f(t) = P_m(t)e^{rt}$, where P_m is a polynomial of degree m, we try the following:

$$y_p(t) = t^s(A_0 + A_1 t + \cdots + A_m t^m)e^{rt}$$

If r is not a solution to the auxiliary equation, set $s = 0$. If r is a simple (not repeated) solution to the auxiliary equation, set $s = 1$. If r is a repeated solution to the auxiliary equation, set $s = 2$.

31.2 It Only Gets Worse

What about if we have a complex r?

To find a particular solution to

$$ay'' + by' + cy = P_m(t)e^{\alpha t}\cos(\beta t)$$

(or the same with sine), we try the solution

$$y_p(t) = t^s(A_0 + A_1 t + \cdots + A_m t^m)e^{\alpha t}\cos(\beta t) + t^s(B_0 + B_1 t + \cdots + B_m t^m)e^{\alpha t}\sin(\beta t)$$

yikes.

If $\alpha + i\beta$ is not a solution to the auxiliary equation, set $s = 0$. If $\alpha + i\beta$ is a solution to the auxiliary equation, set $s = 1$.

This has $2(m+1)$ unknowns. We want a general method to find suitable sets of $A_0, \ldots, A_m, B_0, \ldots, B_m$, which we ultimately do by solving linear systems.

Examples

Find general solutions to the following:

1. $y'' - 2y' + y = 0$
   Auxiliary equation: $r^2 - 2r + 1 = 0 \implies r = 1$
   So the solution is of the form $C_1 e^t + C_2 t e^t$.
2. $y'' - 2y' + y = t^2 - 5t + 5$
   $y = y_p + y_h$. y_h is given above, so to find y_p we set $s = 2$, so

   $$y_p = (A_0 + A_1 t + A_2 t^2)$$

   We take derivatives,

   $$y_p' = A_1 + 2A_2 t$$

   $$y_p'' = 2A_2$$

   and we compute the left side of the DE:

   $$y_p'' - 2y_p' + y_p = 2A_2 - 2A_1 - 4A_2 t + A_0 + A_1 t + A_2 t^2 = A_2 t^2 + (A_1 - 4A_2)t + (A_0 - 2A_1 + 2A_2) = t^2 - 5t + 5$$

   which is three linear equation in three unknowns.
   Note that r here is not the same r as in the homogeneous case.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 32: Linear Systems of Differential Equations

Lecturer: Alexander Paulin *13 April* *Aditya Sengupta*

32.1 Dealing with Slightly More Complicated Functions

For this, we introduce the *superposition principle*. This is the idea that if we have

$$ay_1'' + by_1' + cy_1 = f_1(t)$$

and

$$ay_2'' + by_2' + cy_2 = f_2(t)$$

then we can add them to find that $y_1 + y_2$ is a solution to

$$ay'' + by' + cy = f_1(t) + f_2(t)$$

This allows us to present a general overview of finding the solution to this DE:

1. Find a general solution to the homogeneous case
2. Break up $f(t)$ into the sum of $P_m(t)e^{kt}$, $P_m(t)e^{kt}\cos(lt)$, $P_m(t)e^{kt}\sin(lt)$
3. Find a particular solution for each piece using the method of undetermined coefficients.
4. Apply the Superposition Principle: add all of these solutions to get a total particular solution.
5. Given initial conditions, calculate $y(t_0)$ and $y'(t_0)$, and solve for C_1 and C_2.

32.2 Motivating Example

Let two tanks A and B consist of salty water. They are connected by two tubes and fresh water is pumped into tank A at the rate $30L/min$.

Let $x(t)$ represent the mass of salt in tank A, and let $y(t)$ represent the mass in tank B. Both tanks have a capacity of 10L, the flow rate into tank B from tank A is 40L/min, the flow rate into tank A from tank B is 10L/min, and the rate of outflow is $30L/min$. Let the initial concentrations of salt in A and B be x_0 and y_0 kg.

$$x'(t) = \text{input rate - output rate} = \frac{10y(t)}{10} - \frac{40}{10}x(t) = -4x(t) + y(t)$$

$$y'(t) = 4x(t) - y(t) - 3y(t) = 4x(t) - 4y(t)$$

32.2.1 Approach One

$$y' = 4x - 4y \implies 4x = y' + 4y$$

$$y'' = 4x' - 4y' = 4(-4x + y) - 4y' = 4(-y' - 4y + y) - 4y' = -8y' - 12y$$

Therefore $y'' + 8y' + 12y = 0$. This has the solution $c_1 e^{-2t} + c_2 e^{-6t}$. We take one derivative,

$$y'(t) = -2c_1 e^{-2t} - 6c_2 e^{-6t}$$

Then, we can find $x(t)$:

$$x(t) = \frac{1}{4} y'(t) + y(t) = \frac{1}{2} c_2 e^{-2t} - \frac{1}{2} c_2 e^{-6t}$$

Plugging in the initial conditions stated above by substituting $t = 0$, we get

$$x(t) = \frac{1}{2} \frac{2x_0 + y_0}{2} e^{-2t} - \frac{1}{2} \frac{y_0 - 2x_0}{2} e^{-6t}$$

$$y(t) = \frac{2x_0 + y_0}{2} e^{-2t} + \frac{y_0 - 2x_0}{2} e^{-6t}$$

There we go, done, walk away, never think about this again.

32.2.2 Approach Two

Let's use linear algebra.

We know that $x'(t) = -4x(t) + y(t)$ and $y'(t) = 4x(t) - 4y(t)$

This could be written in matrix notation as

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \begin{bmatrix} -4 & 1 \\ 4 & -4 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

Imagine $\underline{v}$ is an eigenvector for this coefficient matrix. Consider $\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = e^{\lambda t} \underline{v} = \begin{bmatrix} e^{\lambda t} v_1 \\ e^{\lambda t} v_2 \end{bmatrix}$

Then,

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \lambda e^{\lambda t} \underline{v} = e^{\lambda t} A \underline{v} = A e^{\lambda t} \underline{v} = A \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

so we've found that $e^{\lambda t} \underline{v}$ is a solution!

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 33: Linear Systems of Differential Equations, II

Lecturer: Alexander Paulin *16 April* *Aditya Sengupta*

33.1 The Matrix Method for Solving Linear Systems of ODEs

Last time, we found out that $e^{\lambda t}\underline{v}$ was a solution to the given system of linear ordinary differential equations, where $\underline{v}$ was an eigenvector of the coefficient matrix. So let's diagonalize the coefficient matrix:

$$A = \begin{bmatrix} -4 & 1 \\ 4 & -4 \end{bmatrix}$$

$$det(A - xI_2) = (-4-x)(-4-x) - 4 = (x+2)(x+6)$$

Therefore the eigenvalues are -2 and -6. Then, we find the null space of the eigenspaces:

$$Nul(A + 2I_2) = Nul \begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix} = Span\Big(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \Big)$$

$$Nul(A + 6I_2) = Nul \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix} = Span\Big(\begin{bmatrix} 1 \\ -2 \end{bmatrix} \Big)$$

Therefore, $e^{-2t}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $e^{-6t}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ are solutions. This implies any linear combination of these two is a solution, i.e.

$$d_1 e^{-2t}\begin{bmatrix} 1 \\ 2 \end{bmatrix} + d_2 e^{-6t}\begin{bmatrix} 1 \\ -2 \end{bmatrix} \text{ is a solution for any } d_1, d_2$$

Therefore, at $t = 0$, we can apply the constraint $x_0 = x(0)$ and the same with y to get:

$$x_0 = d_1 + d_2, y_0 = 2d_1 - 2d_2$$

Row reduction tells us that

$$d_1 = \frac{2x_0 + y_0}{4}, d_2 = \frac{2x_0 - y_0}{4}$$

Therefore, we can conclude that

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} \frac{2x_0+y_0}{4}e^{-2t} + \frac{2x_0-y_0}{4}e^{-6t} \\ \frac{2x_0+y_0}{2}e^{-2t} - \frac{2x_0-y_0}{2}e^{-6t} \end{bmatrix}$$

33.2 Notation

$$\underline{v}(t) = \begin{bmatrix} v_1(t) \\ \vdots \\ v_n(t) \end{bmatrix}$$

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \dots & a_{1n}(t) \\ \vdots & & & \\ a_{m1}(t) & \dots & & a_{mn}(t) \end{bmatrix}$$

where each element is a real-valued function.

Then, $\underline{v} : \mathbb{R} \to \mathbb{R}^n, t \to \underline{v}(t)$ is a vector valued function, and $A : \mathbb{R} \to m \times n$ matrix, $t \to A(t)$. We can define, say, differentiation, by differentiating each of the elements of a v matrix.

Definition 55. *A linear system of first-order differential equations in n functions $x_1(t), \dots, x_n(t)$ is*

$$x_1'(t) = a_{11}(t)x_1(t) + a_{12}(t)x_2(t) + \dots + a_{1n}(t)x_n(t) + f_1(t)$$

$$x_2'(t) = a_{21}(t)x_1(t) + a_{22}x_2(t) + \dots + a_{2n}(t)x_n(t) + f_2(t)$$

$$\vdots$$

$$x_m'(t) = a_{m1}(t)x_1(t) + \dots + \dots + a_{mn}(t)x_n(t) + f_m(t)$$

33.3 Cool Example

Let $A(t)$ and $f(t)$ be defined as follows:

$$A(t) = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 \\ -p_0(t) & -p_1(t) & \dots & \dots & -p_{n-1}(t) \end{bmatrix}$$

$$f(t) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ g(t) \end{bmatrix}$$

Then, for the system $x'(t) = A(t)x(t) + f(t)$, we can say the following:

$$x_1'(t) = x_2(t), x_2'(t) = x_3(t), \dots, x_{n-1}'(t) = x_n(t)$$

$$x_n'(t) = -p_0(t)x_1(t) - p_1(t)x_2(t) - \cdots - p_{n-1}(t)x_n(t) + g(t)$$

From the first statement, we can see that $x_i(t) = x_1^{(i-1)}(t)$. Therefore the second statement can be changed into a linear $n-$th order differential equation. So, solving this system is the same as solving

$$y^{(n)}(t) + P_{n-1}(t)y^{(n-1)}(t) + \cdots + P_0(t)y(t) = g(t)$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 34: Linear Systems of Differential Equations, III

Lecturer: Alexander Paulin *18 April* *Aditya Sengupta*

Theorem 34.1. *If $A(t)$ and $\underline{f}(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{bmatrix}$ are continuous on an open interval I, containing some t_0, then for any choice of $\underline{x_0}$, a vector in $\mathbb{R}^n$, there exists a unique x_t on I such that*

1. $\underline{x}'(t) = A(t)\underline{x}(t) + \underline{f}(t)$
2. $\underline{x}(t_0) = \underline{x_0}$

34.1 Homogeneous Case

Let $f(t) = 0$ for all t. Then, the following are true:

1. $\underline{0}(t)$ is a solution to $x'(t) = Ax(t)$.
2. $x(t), y(t)$ being solutions $\implies$ $x(t) + y(t)$ is a solution.
3. $x(t)$ being a solution implies $\lambda x(t)$ is a solution.

This means the solutions to $x'(t) = Ax(t)$ form a vector space. Therefore, in order to find the complete set of solutions, we need to find n linearly independent solutions, where n is the dimension of the solution space. In other words, if there exists a one-one, onto, linear transformation T : solution space $\to \mathbb{R}^n$, $x(t) \to x(t_0)$, we need to find n L.I. solutions.

To check whether a solution set is linearly independent, we introduce *the Wronskian*.

Definition 56. *For n $\mathbb{R}^n$ valued functions $\{\underline{x_1}, \dots, \underline{x_n}\}$, the Wronskian of the set is*

$$W\left[\underline{x_1}, \dots, \underline{x_n}\right](t) = det\left(\underline{x_1}(t), \dots, \underline{x_n}(t)\right)$$

The Wronskian of such an $\underline{x}-$matrix is

$$W[\underline{x_1}, \dots, \underline{x_n}] = det(\underline{x_1}(t)\underline{x_2}(t) \dots \underline{x_n}(t))$$

Theorem 34.2. *If $\underline{x_1}, \dots, \underline{x_n}$ are solutions to $\underline{x}'(t) = A(t)\underline{x}(t)$, then*

$$\{\underline{x_1}, \dots, \underline{x_n}\} L.D. \iff W[\underline{x_1}, \dots, \underline{x_n}](t_0) = 0$$

for some t_0 in I.

Proof. If the set of $x-$vectors is L.D., then we can find $c_1, \ldots, c_n$ such that

$$c_1\underline{x_1} + \cdots + c_n\underline{x_n} = \underline{0}(t)$$

for all t.

Therefore, $\{\underline{x_1}(t), \ldots, \underline{x_n}(t)\} \subset \mathbb{R}^n$ is linearly dependent for all t in I.

This means the Wronskian is 0 for all t in I. (Invertible Matrix Theorem).

To prove the other direction, assume there exists a t_0 in I such that $W[\underline{x_1}, \ldots, \underline{x_n}](t_0) = 0$.

Therefore, $\{\underline{x_1}(t_0), \ldots, \underline{x_n}(t_0)\}$ is linearly dependent, so there exist $c_1, \ldots, c_n$ not all zero, such that

$$c_1\underline{x_1}(t_0) + \cdots + c_n\underline{x_n}(t_0) = \underline{0}(t_0)$$

By the uniqueness of initial conditions,

$$c_1\underline{x_1}(t) + \cdots + c_n\underline{x_n}(t) = \underline{0}(t)\forall t \in I$$

□

34.2 Conclusion

If $\{\underline{x_1}, \ldots, \underline{x_n}\}$ are solutions to $\underline{x}'(t) = A(t)\underline{x}(t)$, then the x_is are L.I if and only if the Wronskian is nonzero on I, and if the Wronskian is nonzero on I, then $\underline{x}(t) = \lambda_1\underline{x_1}(t) + \cdots + \lambda_n\underline{x_n}(t)$ is a general solution to $\underline{x}'(t) = A(t)\underline{x}(t)$.

In this case, we call $\{\underline{x_1}, \ldots, \underline{x_n}\}$ a fundamental solution set.

To go back to the original example,

$$\underline{x}'(t) = \begin{bmatrix} -4 & 1 \\ 4 & -4 \end{bmatrix} \underline{x}(t)$$

$$\underline{x_1}(t) = \begin{bmatrix} e^{-2t} \\ 2e^{-2t} \end{bmatrix}, \underline{x_2}(t) = \begin{bmatrix} e^{-6t} \\ e^{-2t} \end{bmatrix}$$

The Wronskian is

$$W = det \begin{bmatrix} e^{-2t} & e^{-6t} \\ 2e^{-2t} & -2e^{-6t} \end{bmatrix} \neq 0$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 35: Constant Coefficient Linear Systems

Lecturer: Alexander Paulin *20 April* *Aditya Sengupta*

35.1 Real Eigenvalues

Let $\underline{x}$ and A be defined as follows:

$$\underline{x} = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

where each a_{ij} is a constant real number. We want to find the general solution to

$$x'(t) = Ax(t), t \in \mathbb{R}$$

We observe that if $\underline{v} = \begin{bmatrix} v_1(t) \\ \vdots \\ v_n(t) \end{bmatrix}$ is an eigenvector of A with eigenvalue λ, then $x(t) = e^{\lambda t}\underline{v}$ is a solution:

$$x'(t) = \begin{bmatrix} \lambda e^{\lambda t}\underline{v_1} \\ \vdots \\ \lambda e^{\lambda t}\underline{v_n} \end{bmatrix} = e^{\lambda t}\lambda\underline{v} = e^{\lambda t}A\underline{v} = A\underline{x(t)}$$

Theorem 35.1. *If* $\{\underline{v_1}, \dots, \underline{v_n}\} \subset \mathbb{R}^n$ *is a basis of eigenvectors of* A *with eigenvectors* $\lambda_1, \dots, \lambda_n$*, then*

$$x_1(t) = e^{\lambda_1 t}\underline{v_1}, x_2(t) = e^{\lambda_2 t}\underline{v_2}, \dots, x_n(t) = e^{\lambda_n t}\underline{v_n}$$

is a fundamental solution set.

Proof. A set is an FSS if and only if Wronskian is nonzero for all times, i.e.

$$W[x_1, \dots, x_n](t) \neq 0 \forall t \in (-\infty, \infty)$$

$$W(t) = det(e^{\lambda_1 t}\underline{v_1}, \dots, e^{\lambda_n t}\underline{v_n})$$

$$= e^{\lambda_1 t} e^{\lambda_2 t} \dots e^{\lambda_n t} det(\underline{v_1} \dots \underline{v_n})$$

which is nonzero for all times because the exponential function is never zero, and the determinant is that of a basis for $\mathbb{R}^n$, i.e. all the columns are linearly independent, making the determinant nonzero.

$\square$

Therefore, if A is diagonalizable, we can completely solve $x'(t) = Ax(t)$.

Example

$$x'(t) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} x(t)$$

This coefficient matrix has eigenvalues $\lambda = 1, 3$. Therefore we find the $\lambda-$eigenspaces:

$$A - I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \implies Nul(A - I) = Span \{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \}$$

$$A - 3I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \implies Nul(A - 3I) = Span \{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \}$$

This gives us the two linearly independent solutions

$$x_1(t) = e^t \begin{bmatrix} 1 \\ -1 \end{bmatrix}, x_2(t) = e^{3t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

which form a fundamental solution set.

35.2 Non-Real Eigenvalues

We run into a problem with this approach when the characteristic equation has non-real roots. For example,

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \implies det(A - xI_2) = x^2 + 1 = 0$$

This gives us $x = \pm i$. This actually still works, i.e. $\underline{x}(t) = e^{it} \begin{bmatrix} i \\ 1 \end{bmatrix}$ is a solution:

$$x'(t) = ie^{it} \begin{bmatrix} i \\ 1 \end{bmatrix} = e^{it} i \begin{bmatrix} i \\ 1 \end{bmatrix} = e^{it} A \begin{bmatrix} i \\ 1 \end{bmatrix} = A \left(e^{it} \begin{bmatrix} i \\ 1 \end{bmatrix} \right) = A\underline{x}(t)$$

However, we want to find real solutions. To do this, we can employ the same methods of linear algebra, which still work for complex numbers. For example, finding a null space works the same way:

$$Nul(A - iI_2) = Nul \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}$$

$$\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \to \begin{bmatrix} 1 & -i \\ -i & -1 \end{bmatrix} \to \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}$$

Therefore $Nul(A - iI_2) = Span \left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\}$

To find real solutions, we can do the following:

$$e^{it} = (\cos t + i \sin t) \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} + i \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \left(\cos t \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \sin t \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) + i \left(\cos t \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sin t \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)$$

This gives us the fundamental solution set we want:

$$\underline{x_1}(t) = \cos t \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \sin t \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}, \underline{x_2}(t) = \cos t \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sin t \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 36: Introduction to Fourier Series

Lecturer: Alexander Paulin *23 April* *Aditya Sengupta*

36.1 Complex Constant-Coefficient Linear Systems

In general, we have an $n \times n$ matrix A, with non-real eigenvalue $\alpha + i\beta$ and an eigenvector $\underline{a} + i\underline{b}$, and we want to solve the linear system of differential equations associated with this. We get

$$\underline{x_1}(t) = e^{\alpha t}\cos(\beta t)\underline{a} - e^{\alpha t}\sin(\beta t)\underline{b}$$

$$\underline{x_2}(t) = e^{\alpha t}\sin(\beta t)\underline{a} + e^{\alpha t}\cos(\beta t)\underline{b}$$

These are real solutions to the system $\underline{x}'(t) = A\underline{x}(t)$.

Example with Initial Conditions

$$\underline{x}'(t) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \underline{x}(t), \underline{x}(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

From before, the following is a fundamental solution set:

$$\left\{ \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}, \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} \right\}$$

Therefore, $\underline{x}(t)$ is a linear combination of these. Applying the initial condition gives us

$$\underline{x}(0) = \begin{bmatrix} C_2 \\ C_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Therefore

$$\underline{x}(t) = \begin{bmatrix} \cos t - \sin t \\ \cos t + \sin t \end{bmatrix}$$

36.2 Fourier Series

Recall that Taylor series look like this:

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Under nice circumstances,

$$f(x) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

Fourier series attempt to achieve the same goal of approximating functions well, but instead of using the function's derivatives, it does this using sine and cosine functions of increasing frequency. Conventionally, we let f be piecewise continuous on [-L, L].

Let a function g be $2L-$periodic if and only if $g(x) = g(x+2L)$ for all L. For example, $\sin\left(\frac{m\pi n}{L}\right), \cos\left(\frac{m\pi n}{L}\right)$ are $2L-$periodic. This can be easily verified with sine and cosine addition laws.

Here, m is the frequency, or the number of complete oscillations over [-L, L].

Recall that the vector space $V = \{\, f : [-L, L] \to \mathbb{R}, \text{piecewise-continuous}\,\}$ comes with the natural inner product

$$\langle f, g\rangle = \int_{-L}^{L} f(x)g(x)dx$$

Theorem 36.1.

$$\{\cos\left(\frac{\pi 0x}{L}\right), \sin\left(\frac{\pi x}{L}\right), \cos\left(\frac{\pi x}{L}\right), \sin\left(\frac{2\pi x}{L}\right), \dots\}$$

is an orthogonal set.

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 37: Fourier Series

Lecturer: Alexander Paulin *25 April* *Aditya Sengupta*

Proof. $\sin\left(\frac{m\pi x}{L}\right)$ is odd, and $\cos\left(\frac{m\pi x}{L}\right)$ is even. Therefore their product is odd.

$$\langle \sin\left(\frac{m\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)\rangle = \int_{-L}^{L} \sin\left(\frac{m\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 0$$

To calculate this for different sines and cosines:

$$\left\langle \sin\left(\frac{m\pi x}{L}\right)\sin\left(\frac{n\pi x}{L}\right)\right\rangle = \int_{-L}^{L} \sin\left(\frac{m\pi x}{L}\right)\sin\left(\frac{n\pi x}{L}\right)dx$$

Using integration by parts twice, we can show this is 0 for $m \neq n$. If $m = n \neq 0$, this integral evaluates to L. Similar logic applies for cosine, with the exception that the integral is $2L$ for $m = n = 0$.

Therefore, because the inner product is 0 for all nonequal functions, the set is orthogonal. □

Let

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^{k}\left(a_m \cos\left(\frac{m\pi x}{L}\right) + b_m \sin\left(\frac{m\pi x}{L}\right)\right)$$

For example, $f(x) = \frac{5}{2} + 6\cos\left(\frac{\pi x}{L}\right) - 4\sin\left(\frac{\pi x}{L}\right) + \sin\left(\frac{2\pi x}{L}\right)$. We want to determine the a_ms and b_ms directly. To do this, we consider the inner product of $f(x)$ and $\cos\left(\frac{n\pi x}{L}\right)$:

$$\left\langle \frac{a_0}{2} + \sum_{m=1}^{k}\left(a_m \cos\left(\frac{m\pi x}{L}\right) + b_m \sin\left(\frac{m\pi x}{L}\right)\right), \cos\left(\frac{n\pi x}{L}\right)\right\rangle$$

By the above theorem, every component of this except the $\cos\left(\frac{n\pi x}{L}\right)$ becomes zero. Formally, we have

$$\frac{a_0}{2}\left\langle \cos\left(\frac{0\pi x}{L}\right), \cos\left(\frac{n\pi x}{L}\right)\right\rangle + \sum_{m=1}^{k} a_m \left\langle \cos\left(\frac{m\pi x}{L}\right), \cos\left(\frac{n\pi x}{L}\right)\right\rangle + \sum_{m=1}^{k} b_m \left\langle \sin\left(\frac{m\pi x}{L}\right), \cos\left(\frac{n\pi x}{L}\right)\right\rangle$$

which splits into the piecewise result

$$= \begin{cases} \frac{a_0}{2}\cdot 2L & n = 0 \\ a_n L & n \neq 0, n \leq k \\ 0 & n > k \end{cases}$$

Similarly,

$$\left\langle f(x), \sin\left(\frac{n\pi x}{L}\right)\right\rangle = \begin{cases} b_n L & 0 < n \leq k \\ 0 & n > k \end{cases}$$

Conclusion

If $f(x) = \frac{a_0}{2} + \sum_{m=1}^{k}\left(a_m \cos\left(\frac{m\pi x}{L}\right) + b_m \sin\left(\frac{m\pi x}{L}\right)\right)$, then

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right) dx, 0 \leq n \leq k$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right) dx, 1 \leq n \leq k$$

Remark 37.1. *This is similar to expressing a function $f(x) = c_0 + c_1 x + \cdots + c_n x^n$, in which $c_n = \frac{f^{(n)}(0)}{n!}$.*

We are now ready to define a Fourier series.

Definition 57. *Let f be a piecewise continuous function on $[-L, L]$. The Fourier series of f is the infinite series*

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right)\right)$$

where

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right) dx, 0 \leq n \leq k, n = 0, 1, 2, \ldots$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right) dx, 1 \leq n \leq k, n = 1, 2, \ldots$$

Remark 37.2. *We have proven that if $f(x)$ is a trigonometric sum as above, then the Fourier series is a finite sum and is equal to $f(x)$ on $[-L, L]$.*

¯_(ツ)_/¯

Example

Compute the Fourier series of $|x|$ on the interval $[-1, 1]$.

Because $|x|$ is even, $|x|\sin(n\pi x)$ is odd. Therefore,

$$\int_{-1}^{1} f(x)\sin(n\pi x)\, dx = 0$$

so all the b_ns are zero.

$|x|\cos(n\pi x)$ is even, therefore

$$\int_{-1}^{1} |x|\cos(n\pi x)\,dx = 2\int_{0}^{1} x\cos(n\pi x)\,dx$$

We can calculate each of the a_ns this way.

$$a_0 = 2\int_0^1 x dx = x^2\Big|_0^1 = 1$$

$$a_n = 2\int_0^1 x\cos(n\pi x)\,dx = \frac{1}{n\pi}x\sin(n\pi x)\Big|_0^1 - \int_0^1 \frac{1}{n\pi}\sin(n\pi x)\,dx$$

The first part of the evaluation goes to 0, because $\sin pi = \sin 0 = 0$, and the second part becomes

$$a_n = \frac{2}{n^2\pi^2}\left((-1)^n - 1\right)$$

Therefore the Fourier series is

$$|x| = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{2}{n^2\pi^2}\left((-1)^n - 1\right)\cos(n\pi x)$$

Math 54: Linear Algebra and Differential Equations **Spring 2018**

Lecture 38: Fourier Series contd.

Lecturer: Alexander Paulin *27 April* *Aditya Sengupta*

Theorem 38.1. *If f and f' are piecewise continuous on $[-L, L]$ then for any x in $(-L, L)$, the Fourier series of the function is equal to*

$$\frac{1}{2}\left(\lim_{h\to 0^+} f(x+h) + \lim_{h\to 0^-} f(x+h)\right)$$

By this, if f is continuous at x, the function is *exactly equal* to its own Fourier series. On the endpoints, i.e. $x = \pm L$, the series converges to $\frac{1}{2}\left(f(L^-) + f(-L^+)\right)$ where $f(x^-)$ and $f(x^+)$ are defined as in the limits above.

As a consequence of this, if f is continuous on $[-L, L]$ and $f(-L) = f(L)$, then the Fourier series converges to $f(x)$ for all x in $[-L, L]$.

Also, if y is a continuous $2L-$periodic function on $(-\infty, \infty)$, then the Fourier series to converges to $f(x)$ for all $x \in \mathbb{R}$

Example

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ , -2 & -1 \le x < 0 \end{cases}$$

We want to know what the Fourier series converges to at $x = 0$ and $x = \frac{5}{2}$.

At $x = 0$, the Fourier series converges to

$$\frac{1}{2}\left(f(0^+) + f(0^-)\right) = \frac{-1}{2}$$

and at $x = \frac{5}{2}$, the Fourier series converges to 1.

38.1 Calculus on Fourier Series

We can integrate and differentiate Fourier series term by term.

$$F'(x) = \sum_{n=1}^{\infty}\left(-a_n \frac{n\pi}{L}\sin\left(\frac{n\pi x}{L}\right) + b_n \frac{n\pi}{L}\cos\left(\frac{n\pi x}{L}\right)\right)$$

and similarly, an integral becomes

$$\int F(x)dx = C + \frac{a_0}{2}x + \sum_{n=1}^{\infty}\left(\frac{L}{n\pi}a_n \sin\left(\frac{n\pi x}{L}\right) - \frac{L}{n\pi}b_n \cos\left(\frac{n\pi x}{L}\right)\right)$$

38.2 Fourier Sine and Cosine Series

If f is even on its symmetric domain, then $b_n = 0$ for all n. Similarly, if f is odd, then $a_n = 0$ for all n. That is, the Fourier series of an even function only has cosine terms, and that of an odd function only has sine terms.

Let f be piecewise continuous on the closed interval $[0, L]$. Then, we define

$$f_e(x) = \begin{cases} f(x) & 0 \leq x \leq L \\ f(-x) & -L \leq x \leq 0 \end{cases}$$

$$f_o(x) = \begin{cases} f(x) & 0 \leq x \leq L \\ -f(-x) & -L \leq x \leq 0 \end{cases}$$

By construction, these are respectively even and odd. Then,

Definition 58. *The Fourier cosine series of f on $[0, L]$ is the Fourier series of f_e on $[-L, L]$:*

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right)$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f_e(x)\cos\left(\frac{n\pi x}{L}\right) dx = \frac{2}{L}\int_{0}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right) dx$$

Definition 59. *The Fourier sine series of f on $[0, L]$ is the Fourier series of f_o on $[-L, L]$:*

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right)$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f_o(x)\sin\left(\frac{n\pi x}{L}\right) dx = \frac{2}{L}\int_{0}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right) dx$$