Mathematics 54

Linear Algebra and Differential Equations

Lecture Notes

Textbook: David Lay's *Introduction to Linear Algebra* and Kent Nagle's *Elementary Differential Equations*

Druv Pai

druv@berkeley.edu

ID: 3033848822

Professor: Constantin Teleman

GSI: Ritwik Ghosh

Contents

1 August 22, 2018

We review the syllabus.

Listening to math lectures is a skill, that one can hone. In three lines, one can write with clever notation what intuition is gained over one hundred examples. The information density is really high. It is imperative to follow each step of a calculation. Two readings of the textbook section are suggested: one before the lecture, and one while during the homework. Now to the material.

The basic problem of linear algebra is solving systems of linear equations. The general system of linear equations is

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

and can be expressed in terms of a **coefficient matrix**, because the variables are only functions of the positions.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

The **augmented matrix** has the solutions of each linear equation at the end of each row, so in this case it is

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix}$$

Let A be the coefficient matrix, $\mathbf{x} = (x_1, x_2, \dots, x_n)$, and $\mathbf{b} = (b_1, b_2, \dots, b_m)$ be vectors; then $A\mathbf{x} = \mathbf{b}$ is a concise way to rewrite the system of linear equations.

There are two parts of the solution method for linear systems: elimination and substitution.

Example 1.1

Solve the system

$$\begin{cases} x + 2y = 3 \\ 4x + 5y = 6 \end{cases}$$

Solution For eliminiation:

$$\begin{cases} x + 2y = 3 \\ 0x - 3y = -6 \end{cases}$$

Thus,

$$\begin{cases} x + 2y = 3 \\ y = 2 \end{cases}$$

Substitution yields

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

□

Another question is what happens when we have two equivalent equations, as in the following case.

Example 1.2

Solve

$$\begin{cases} x + 2y = 3 \\ 4x + 8y = 12 \end{cases}$$

Solution Doing elimination, we get

$$\begin{cases} x + 2y = 3 \\ 0x + 0y = 0 \end{cases}$$

Thus we have $x = 3 - 2y$ as our solution. □

Our task is to find a description for the solution sets of systems. A good description should allow you to decide whether two systems have the same solution, to compare the solutions of two systems (including whether or not one is strictly included in the other, implying that the first has strictly more constraints than the second for the same system), and to combine the systems to find simultaneous solutions.

Our first tool is Row Reduction (Gaussian Elimination) and it will help us answer all these questions.

Another possible problem classification is as follows.

Example 1.3

Solve the system

$$\begin{cases} x + 2y = 3 \\ 4x + 8y = 10 \end{cases}$$

Solution Doing elimination, we get

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

This system is **inconsistent** because there is a contradiction in the solution set. Consistent systems have no contradiction in the solution set.

□

When we are trying to develop linear models, it is not safe to measure the minimum number of values; we need to measure way more than we need. However, this introduces many constraints into the system, so it is likely that such a system will have no solutions. We cannot just give up and go home – we need to take our best guess as to what the actual linear model governing the true state actually is.

In short, for inconsistent systems, we might be interested in the **least squares solution** - that is, the $\mathbf{x}$ such that $A\mathbf{x}$ is as close to $\mathbf{b}$ as possible. To do this we need to introduce orthogonality, inner products, vector spaces, and similar abstract notions. Later in the course, we will use such linear algebraic concepts to solve differential equations, and a famous physics problem.

Now we introduce the systematic way of row reduction – elimination and substitution.

Definition 1.1. Two linear systems are equivalent if and only if they have the same solution set.

What we will do is keep transforming the system of equations into an equivalent system until it is so simple that we can obviously write down the solutions. We will need to get the system in reduced row echelon form. Thus the solution will manifest. We will thereafter work in matrices.

We call the eliminiation steps **elementary row operations**. They are:

Lemma 1.1 (Elementary Row Operations)

1. Swapping rows
2. Adding a multiple of a row to another
3. Multiplying a row by a nonzero number.

Do not swap columns. Swapping columns of the associated matrix changes the labeling of the variables and does change the solution set.

Algorithm 1.1 (Row Reduction)

1. Always have a **current entry**. Start with a_{11}.
2. If the current entry is 0, then find a nonzero entry below the current entry, and swap rows. If not, use the current row to zero out all the entries below the current entry.
3. Move one column to the right and one row down.
4. If on the last row or column, stop.
5. Repeat.

2 August 23, 2018

The first recitation! The grading breakdown for the section grade is 10% for homework (completion), 25% for weekly 20 minute quizzes, and 5% participation. Quizzes will be curved to a lower bound of 70% average.

Up to midterm 1, the fundamental question is why $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ are good axes. Up to midterm 2, the fundamental question is if we come up with good axes for a linear problem. This section of the course is significantly more abstract than the previous section, which is geometric. Finally, up to the final exam, we ask ourselves how to apply these methods to differential equations.

Note that the differential equations textbook is bad because it doesn't involve linear algebra. Use the other book for this!

A geometric vector has a direction and magnitude, and can be described by a collection of numbers. But an actual vector is an element in a vector space, which is a set with special properties. For example, $\sin(t)$ is a vector in the vector space of cyclic functions.

Note that GSI notation implies that vectors are underlined and matrices are double underlined.

Let's start with a system of equations.

Example 2.1

Solve the system

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & + & x_3 & = & 4 \\ x_1 & + & x_2 & - & x_3 & = & 1 \\ -x_1 & + & 2x_2 & + & 5x_3 & = & 0 \end{array}$$

Solution First, we rewrite in matrix-vector form:

$$\underbrace{\begin{bmatrix} 4 & 2 & 1 \\ 1 & 1 & -1 \\ -1 & 2 & 5 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{b}}$$

Note that the number of columns of the matrix has to be the same as the number of rows of the vector.

As an aside, note that linear equations in linear algebra have to go through the origin (lines that do not go through the origin are called **affine**) – that is, if we put a zero input, the output is zero.

The systematic method for solving this system is to do Gaussian elimination. We create a matrix

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

We do the operations $R_2 \leftarrow R_2 + R_3$ and $R_3 \leftarrow R_1 + 4R_3$ to get

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 4 & 2 & 1 & 4 \\ 0 & 3 & 4 & 1 \\ 0 & 10 & 21 & 4 \end{bmatrix}$$

Note that we are only adding and multiplying by constants, and also not outright replacing equations (because that would reduce the amount of information in the system).

Now we do the operation $R_3 \leftarrow 10R_2 - 3R_3$ to get

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 4 & 2 & 1 & 4 \\ 0 & 3 & 4 & 1 \\ 0 & 0 & -23 & -2 \end{bmatrix}$$

Note that this is equivalent to the system of equations

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & + & x_3 & = & 4 \\ & & 3x_2 & + & 4x_3 & = & 1 \\ & & & & -23x_3 & = & -2 \end{array}$$

This matrix is in echelon form, the target of Gaussian Elimination. This matrix is in triangular form; it has zeros below the main diagonal of indices (i, i) for $i = 1, 2, 3$.

Thus we have $x_3 = 2/23$, $x_2 = 5/23$, and $x_1 = 20/23$. $\square$

Note that there are a large amount of echelon forms. In general, the key idea is that the number of zeros on each row as the row number increases is monotonically increasing. Some echelon forms of 3×3 matrices are, for $\blacksquare$ a nonzero number (which we call our pivot) and $*$ a real number,

$$\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & \blacksquare \end{bmatrix}, \quad \begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \blacksquare & * & * \\ 0 & 0 & \blacksquare \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \blacksquare & * & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \blacksquare & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & \blacksquare \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

For such a system to have one solution, the planes described by the equations must intersect at one point, so the number of pivots is equal to the number of rows is equal to number of columns of the matrix.

For such a system to have infinite solutions, the planes described by the equations must intersect at a line, or all concur. The number of pivots does not equal the number of rows and number of columns. We must have

$$\begin{bmatrix} 0 & \dots & 0 & 0 \end{bmatrix}$$

For such a system to have no solution, the planes described by the equations must not intersect, so the number of pivots cannot equal the number of rows and the number of columns; alternatively, we must have

$$\begin{bmatrix} 0 & \dots & 0 & c \end{bmatrix} \quad \text{for } c \text{ nonzero}$$

Example 2.2

Solve the system

$$\begin{array}{rcrcrcl} x_1 & + & 4x_2 & + & 3x_3 & = & 2 \\ 2x_1 & + & 2x_2 & + & x_3 & = & 3 \\ -x_1 & & & + & 4x_3 & = & 7 \end{array}$$

Solution We convert it to augmented matrix form:

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

The operation $C_1 \Leftrightarrow C_2$ yields

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

Note that we are now solving for (x_2, x_1, x_3) now instead of the regular permutation. The operation $R_2 \Leftarrow R_1 - 2R_2$ yields

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

The operation $R_3 \leftarrow R_2 - 3R_3$ yields

$$\begin{bmatrix} 4 & 1 & 3 & 2 \\ 0 & -3 & 1 & -4 \\ 0 & 0 & -11 & -25 \end{bmatrix}$$

We thus have $x_3 = 25/11$, giving us the equation $-3x_1 + 25/11 = -4$, so $x_1 = 23/11$. From here we can solve for x_2. □

Example 2.3

Solve the system

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 3 \\ x_1 & - & 2x_2 & + & x_3 & = & 1 \\ 2x_1 & - & x_2 & + & 2x_3 & = & 4 \end{array}$$

Solution We convert to augmented matrix form:

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

The operations $R_2 \leftarrow R_1 - R_2$ and $R_3 \leftarrow 2R_1 - R_3$ yield

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

The obvious operation $R_3 \leftarrow R_2 - R_3$ yields

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

Our pivots are in the positions $(1, 1)$ and $(2, 2)$. Let $x_3 = t$, $x_2 = 2/3$, then $x_1 = 3 - \frac{2}{3} - t = \frac{7}{3} - t$. Thus our solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{x} = \begin{bmatrix} 7/3 \\ 2/3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

If we shifted the third equation's constant, we would get an inconsistent system, so there would be no solution. The three planes would all pairwise intersect but there would not be a triple intersection. □

Example 2.4

Solve the system

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & 3x_3 & = & 2 \\ 2x_1 & + & 2x_2 & + & 6x_3 & = & 4 \\ -x_1 & - & x_2 & - & 3x_3 & = & -2 \end{array}$$

Solution These three are the same equation, so the planes all coincide. Thus the parametric solution set is, from the first equation and letting $x_3 = t$, $x_2 = s$, so $x_1 = 2 - 3t - s$,

$$\mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

□

3 August 24, 2018

We will go over, today:

- Elimination
- Reduced echelon form
- Parametric solutions
- Vectors

We will do an example of elimination and highlight all the important steps and include the terminology.

Example 3.1

Row reduce the augmented matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 6 & 7 \\ 3 & 5 & 7 & 5 & 3 \end{bmatrix}$$

Solution Note that adding additional columns to the right hand side of the augmented matrix allows you to do several row reductions at once, on several different values of the constants.

We start with a_{11}. We do the operations $R_2 \leftarrow R_2 - 2R_1$ and $R_3 \leftarrow R_3 - 3R_1$ to get

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & -1 & -2 & -3 \\ 0 & -1 & -2 & -7 & -12 \end{bmatrix}$$

We move to a_{22} and see that this entry is a zero, so we swap rows: $R_2 \leftrightarrow R_3$.

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & -1 & -2 & -7 & -12 \\ 0 & 0 & -1 & -2 & -3 \end{bmatrix}$$

There is nothing below a_{22}, so we are done with elimination. The matrix is in **upper echelon form**, where the echelon is on a_{11}, a_{22}, a_{33}; these indices are pivots, and columns one, two, and three are pivot columns. The number of columns minus the number of pivot columns is the number of free variables; that is, each non-pivot column is a free variable.

The second part of the reduction is to:

- Rescale each row (divide by its pivot) so that each row starts with a "leading 1"
- Create zeros above the "leading 1's" by subtracting lower rows from the upper rows. Start with the bottom; it reduces the number of operations.

We do the first part, dividing by the pivots, so we have

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 7 & 12 \\ 0 & 0 & 1 & 2 & 3 \end{bmatrix}$$

The row operations $R_1 \leftarrow R_1 + 3R_3$ and $R_2 \leftarrow R_2 - 2R_1$ lead to

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

Finally, we do $R_1 \leftarrow R_1 - 2R_2$ to get

$$\begin{bmatrix} 1 & 0 & 0 & -8 & -16 \\ 0 & 1 & 0 & 3 & 6 \\ 0 & 0 & 1 & 2 & 3 \end{bmatrix}$$

This matrix is in reduced echelon form. □

A matrix is in upper echelon form if:

- Every nonzero row starts with a nonzero element called the **pivot**.
- Every pivot is strictly to the right of all the pivots above it.
- If there exist completely zero'd out rows, they are at the bottom.

Note that we do not want pivots on the right side of the augmented matrix, only in the coefficient matrix.

In a reduced echelon form, in addition:

- All pivots are leading 1's
- All the entries above a leading 1 are zero.

Note that these all apply to a coefficient matrix.

The above example had a system with variables $x_1, \ldots, x_4$. The reduced echelon form leads to the system

$$\begin{array}{ccccccc} x_1 & & & - & 8x_4 & = & -16 \\ & x_2 & & & 3x_4 & = & 6 \\ & & x_3 & + & 2x_4 & = & 3 \end{array}$$

Rewriting this leads to

$$\begin{cases} x_2 = 8x_4 - 16 \\ x_2 = -3x_4 + 6 \\ x_3 = -2x_4 + 3 \\ x_4 \text{ is free} \end{cases}$$

In general, in reduced row echelon form, you can write the solution set in parametric form. The parameters are the free variables, which are independent. Since every equation has one pivot variable and every pivot variable is in at least one equation, each equation determines one of the pivot variables.

In the case of inconsistent systems, there might be more rows at the bottom that don't have a leading row or a pivot, so we have $0 = c$ for c nonzero, drawing a contradiction. In other words, a system is inconsistent if and only if row reduction leads to a row $\begin{bmatrix} 0 & 0 & \dots & c \end{bmatrix}$, which corresponds to "zero equals nonzero".

Theorem 3.1

Two systems are equivalent (they have the same solution set) if and only if row reduction leads to the same reduced echelon form. Note that inconsistent systems are equivalent.

Proof For the "only if" direction, note that the echelon form leads to a parametric solution. For the "if" direction, from the parametric expression of the pivot variables in terms of free variables, we can recover the entries of the reduced echelon form. □

The gap in this proof is that we don't necessarily know that equivalent systems have the same pivot values. However, this is true. There is a clever characterization of pivot and free variables in terms of the solution, and that is that in the parameterization the equations only contain a pivot variable and the variables after it, so the parameterization must correspond to one unique reduced echelon form matrix.

Some terminology:

- A system $A\mathbf{x} = \mathbf{0}$ is called **homogeneous**. It is always consistent because it is always solved by $\mathbf{x}$ being the zero vector.
- The **general solution** of the system $A\mathbf{x} = \mathbf{b}$ can be written as the sum of a particular solution (free variables set to zero plus the general solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

From this second point, we get the fact that the solution to $A\mathbf{x} = \mathbf{b}$ (if it exists; if the system is consistent) is unique if and only if the solution to $A\mathbf{x} = \mathbf{0}$ is uniquely the zero vector if and only if there are no free variables and a pivot in every column. This is the answer to the uniqueness problem.

On the other hand, the fact that there may or may not be pivots in every row has to do with the existence problem. We have that the system $A\mathbf{x} = \mathbf{b}$ is consistent for every choice of $\mathbf{b}$ if and only if there is a pivot in every row (this rules out a row $\begin{bmatrix} 0 & 0 & \dots & 0 & c \end{bmatrix}$.)

We now deal with vectors in R^2, R^3, and R^n. The point of introducing these is to define some efficient notation for linear operations on collections of numbers; later, it will be a useful abstract notation.

A vector in R^n is an $n \times 1$ matrix ("column") of real numbers:

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

We also have row vectors in R^m, which are $1 \times m$ matrices ("rows") of real numbers:

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix}$$

Two operations defined on vectors are:

- Vector addition (componentwise, each component in two vectors are summed to create the sum of the vectors)
- Multiplication by a scalar, each component in the vector is multiplied by the scalar

4 August 27, 2018

Errata: two consistent linear systems are equivalent if and only if they can be transformed into each other through a series of elementary row operations. All inconsistent systems are equivalent.

Today we will cover vectors and linear operations; linear combinations and spans (which is related to the existence of solutions to $A\mathbf{x} = \mathbf{b}$); parametric vector form of solutions; and the matrix-vector product $A\mathbf{x}$.

Definition 4.1. A "vector" – a "column vector" of rank n is a point in R^n (a $n \times 1$ array, n rows, 1 column).

Definition 4.2. A "row vector" of rank m is a $1 \times m$ array of numbers.

A vector $\mathbf{v} \in \mathsf{V}$ has two operations:

- Addition: $\mathsf{R}^n \times \mathsf{R}^n \to \mathsf{R}^n$ – defined by

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}$$

- Scalar multiplication: $\mathbb{R} \times \mathsf{R}^n \to \mathsf{R}^n$ – defined by:

$$k\mathbf{a} = k \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} ka_1 \\ ka_2 \\ \vdots \\ ka_n \end{bmatrix}$$

Some obvious properties of vectors are:

- $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ (commutativity)
- $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (associativity)
- $\mathbf{a} + \mathbf{0} = \mathbf{a}$ $\left(\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right)$
- For all $\mathbf{v}$ there exists a vector $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ (existence of additive inverse)
- $(m)n\mathbf{a} = m(n\mathbf{a}) = (mn)\mathbf{a}$ (associativity)
- $m(\mathbf{a} + \mathbf{b}) = m\mathbf{a} + m\mathbf{b}$ (distributivity)

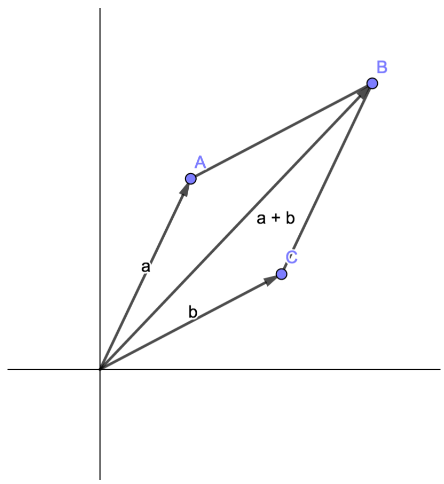
- $(m+n)\mathbf{a} = m\mathbf{a} + n\mathbf{a}$ (distributivity)
- $1\mathbf{v} = \mathbf{v}$

Remark We could use $\mathbb{C}$ instead of $\mathbb{R}$ as our scalars and C^n instead of R^n for our vectors, and nothing would change.

Remark The properties listed above are defined properties on an a real abstract vector space.

From here all theorems and results that we check for R^n apply to (finite-dimensional) abstract vector space, and with a little bit of work apply to infinite-dimensional vector spaces as well. We will use this for differential equations.

Now we focus on the geometric description. The sum of two vectors $\mathbf{a}, \mathbf{b} \in \mathsf{R}^n$ is the fourth vertex of the parallelogram with the other three vertices at $\mathbf{0}, \mathbf{u}, \mathbf{v}$.



Scaling a vector is even easier to express; the coordinates are scaled by the factor.

Definition 4.3. A linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is the value $\mathbf{w}$ given by the expression

$$\mathbf{w} = \sum_{i=1}^{p} k_i \mathbf{v}_i, \quad k_i \in \mathbb{R}$$

Note that the elements of this linear combination need to be from the same vector space i.e. operations between vectors from different vector spaces are not defined.

Let $\mathbf{u}$ and $\mathbf{v}$ be two non-geometric and non-parallel vectors in R^2 (that is, $\mathbf{u}$ and $\mathbf{v}$ are not scalar multiples of each other). Then any vector in R^2 is a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

Proof Pick a vector $\mathbf{w} \in \mathsf{R}^2$. Then draw $\mathbf{u}$ and $\mathbf{v}$. Draw lines out of $\mathbf{u}$ parallel to the lines from the origin to $\mathbf{u}$ and $\mathbf{w}$. The intersections of the lines should form the vertices of a parallelogram, showing that the sum is given by some combination of $\mathbf{u}$ and $\mathbf{v}$ by the parallelogram rule. $\square$

We can do the same in R^3. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be three vectors not lying in the same plane through the origin. Then every vector in R^3 is a linear combination of $\mathbf{u}, \mathbf{v}, \mathbf{w}$.

Proof We adopt the parallelogram rule to R^3, calling it instead the parallelepiped rule (with a similar application). Using this rule in the same way as above yields a similar result. $\square$

It is fairly clear how to prove the general result through induction.

> **Definition 4.4.** The span S of a collection of vectors S is the set of all vectors that may be expressed as linear combinations of the vectors in S, plus the zero vector.

In particular, using the above notation $\mathsf{S}\{\mathbf{u}, \mathbf{v}\} = \mathsf{R}^2$, and $\mathsf{S}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} = \mathsf{R}^3$.

Fact 4.1

The possible spans of collections of vectors in R^2 are the following:

$$\{\mathbf{0}\}, \quad \text{a line through } \mathbf{0}, \quad \text{all of } \mathsf{R}^2$$

Similarly for R^3, the possible spans are

$$\{\mathbf{0}\}, \quad \text{a line through } \mathbf{0}, \quad \text{a plane through } \mathbf{0}, \quad \text{all of } \mathsf{R}^3$$

We now cover the matrix-vector product. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathsf{R}^n$, $A = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_n \end{bmatrix} \in \mathsf{R}^{m \times n}$, then $A\mathbf{x} = \mathbf{b} \in \mathsf{R}^m$. In particular

$$A\mathbf{x} = \sum_{i=1}^{n} \mathbf{A_i} x_i = \mathbf{b}$$

is a linear combination of the columns of A with the coefficients being the entries of $\mathbf{x}$.

Again, let $A \in \mathsf{R}^{m \times n}$, $\mathbf{x} \in \mathsf{R}^n$, and $\mathsf{b} \in \mathsf{R}^m$. Then the equation $A\mathbf{x} = \mathbf{b}$ represents a mapping $f(\mathbf{x}) = A\mathbf{x} \colon \mathsf{R}^n \to \mathsf{R}^m$.

Quickly: the parametric solution to a linear system involves parameterizing the free variables and writing everything in terms of the parameters. Specifically, let t_i be a free variable for all i. Then a parametric equation is of the form

$$\mathbf{w} = \mathbf{x} + \sum_{i=1}^{p} t_i \mathbf{v}_i$$

for some vector $\mathbf{x}$ and set of vectors $\mathbf{v}_i$ that are determined by the system.

5 August 28, 2018

We define the existence of solutions to a system as the value representing whether there are either one or infinite solutions. We define the uniqueness of solutions as the value representing whether there is one solution to the system.

There are two types of Gaussian elimination problems: underconstrained, in which the number of equations is less than the number of unknowns; and overconstrained, in which the number of equations is greater than the number of unknowns.

Example 5.1

Solve the underconstrained system given by the augmented matrix

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

Solution We do $C_1 \leftrightarrow C_2$, noting that we are now solving for (x_2, x_1, x_3) instead of (x_1, x_2, x_3). We have

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

We can read off the solution set, swapping back, as

$$\begin{cases} x_1 = -x_3 \\ x_2 = 2 + 3x_3 \\ x_3 \text{ is free} \end{cases}$$

Parameterizing, the solution set is

$$\mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$$

If we were to solve $A\mathbf{x} = \mathbf{0}$ then it is obvious that $\mathbf{x} = t \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$ since the solutions of $A\mathbf{x} = \mathbf{b}$ are a translation of $A\mathbf{x} = \mathbf{0}$ (in this case by $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$). □

Example 5.2

Solve the overconstrained system given by the augmented matrix

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

Solution We do $R_2 \leftarrow R_1 - R_2$ and $R_4 \leftarrow R_1 - R_4$ to get

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

The pivots are in the positions $(1, 1)$, $(2, 2)$, and $(4, 3)$, so there are no non-pivot columns in the coefficient matrix. There are no free variables so there are either one or zero solutions.

Doing $R_3 \leftarrow R_2 - R_3$ yields

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

Doing $R_4 \leftarrow R_3 - 3R_4$ yields

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 12 \end{bmatrix}$$

Hence the system is inconsistent since it has a row of the form $\begin{bmatrix} 0 & 0 & 0 & 12 \end{bmatrix}$ which suggests that $0 = 12$. $\square$

The formula for the number of free variables f is the number of columns n minus the number of pivots p:

$$f = n - p$$

If a solution exists and the number of columns equals the number of pivots, the solution is unique.

Definition 5.1. A vector space is a set V where the elements of the set must satisfy the following properties. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathsf{V}$, and $c, d \in \mathbb{R}$. Then

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

For these to make sense we need a sense of vector addition and scalar multiplication. For example, for a geometric vector, the vector addition is given by the parallelogram/parallelepiped/parallelotope rule, and scalar multiplication is simply scaling a vector by the scalar. For a tuple of numbers, using the axes, we define the addition and multiplication componentwise. In fact, we can prove this from the geometric rules if we pick an axis system - this is not necessarily an axiom.

When we get to vector spaces, prove using non-axial references i.e. prove using geometric vectors.

Example 5.3 (Polynomials of Order 2 ($\mathsf{P}_2(t)$))
An element of $\mathsf{P}_2(t)$ is given by

$$\mathbf{p}(t) = \sum_{i=0}^{2} a_i t^i$$

Let $\mathbf{a} = \sum_{i=0}^{2} a_i t^i$ and $\mathbf{b} = \sum_{i=0}^{2} b_i t^i$. We define addition as just the addition of each term of $\mathbf{a}[t^i]$ and $\mathbf{b}[t^i]$ and scalar multiplication as multiplying each term of $\mathbf{b}[t^i]$. It is easy to see that $\mathsf{P}_2(t)$ is a vector space under these operations.

Elementary row operations either look like

$$R_j \leftarrow \sum_{i=1}^{n} k_i R_i \quad \text{if} \quad 1 \leq j \leq n \quad \text{and} \quad k_j \neq 0$$

which is reversible by the operation

$$R_j \leftarrow \frac{1}{k_i}\left(R_j - \sum_{i \neq j}^{n} k_i R_i\right)$$

or

$$R_i \leftrightarrow R_j$$

which is reversible by applying the operation again.

6 August 29, 2018

The set of linear combinations of a set of vectors - the span - which can also be expressed as a matrix-vector product, is related to the existence of solutions to this product. Today we see that the notion of linear independence is related to the uniqueness of solutions. We will cover the linear subspaces of R^n and their translates.

We quickly review parametric forms of solutions to linear systems. Recall that the solution to $A\mathbf{x} = \mathbf{b}$ is the solution set to $A\mathbf{x} = \mathbf{0}$ translated over by a particular solution of $A\mathbf{x} = \mathbf{b}$. For each free variable, we have a homogeneous solution to the system. Set a particular free variable to 1, set the others to 0, and solve for the pivot variables. If a system is inconsistent, the homogeneous solutions will still exist, but there is no particular solution.

Recall that the matrix vector product

$$A\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{A}_i = \mathbf{b}$$

where everything exists. In particular A has n columns $\mathbf{A}_i$ and $\mathbf{x}$ has n entries x_i.

We have a linear combination of the columns of A. As $\mathbf{x}$ varies with a fixed A, this sweeps out the span of the columns of A.

Definition 6.1. The **column space** $\mathsf{C}(A)$ is the span of the columns of A. In particular, if A is $m \times n$, then $\mathsf{C}(A) \subseteq \mathsf{R}^m$. The equation $A\mathbf{x} = \mathbf{b} = \sum_{i=1}^{n} x_i \mathbf{A}_i = \mathbf{b}$ expresses $\mathbf{b}$ as a linear combination of the columns of A.

Proposition 6.1

Let $A\mathbf{x} = \mathbf{b}$ be a linear system. Then it is consistent if and only if $\mathbf{b} \in \mathsf{C}(A)$.

We now turn to linear independence and uniqueness of solutions.

Definition 6.2. A collection $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of vectors in R^m is linearly independent if and only if the only linear combination $\sum_{i=1}^{p} k_i\mathbf{v}_i = k_1\mathbf{v}_1 + \cdots + k_p\mathbf{v}_p$ summing to $\mathbf{0}$ is the trivial one i.e. $k_i = 0$ for all i.

Proposition 6.2

The homogeneous linear system

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

has only the trivial solution if the columns of A are linearly independent.

Proposition 6.3

The equation

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

is equivalent to the equation

$$\sum_{i=1}^{n} x_i\mathbf{A}_i = x_1\mathbf{A}_1 + \cdots + x_n\mathbf{A}_n = \mathbf{0}$$

where the $\mathbf{A}_i$ are the columns of A.

Note that this notion of linear independence applies to a collection of vectors, not an individual vector.

Example 6.1

When is $\{\mathbf{v}\}$ linearly independent?

Solution The condition is that $k\mathbf{v} = \mathbf{0}$ which implies that either $k = 0$ or $\mathbf{v} = 0$. The first condition implies that only the trivial solution exists – implying linear independence, but the second condition shows that there are infinitely many solutions. In this case $\{\mathbf{v}\}$ is linearly dependent; in all other cases $\{\mathbf{v}\}$ is linearly independent. $\square$

Fact 6.1

Any collection containing the $\mathbf{0}$ vector is linearly dependent.

Example

We can write a linearly dependent relation for $\{\mathbf{0}, \mathbf{v}\}$ by $1\mathbf{0} + 0\mathbf{v}$. This applies for all $\mathbf{v}$.

They do some examples which are pretty trivial. We omit them here.

Fact 6.2

Two vectors in R^2 are linearly dependent if and only if one of them is a scalar multiple of the other.

Proof Linear dependence is equivalent to saying that we can find k_1 and k_2, not both zero, such that $k_1\mathbf{v}_1 + k_2\mathbf{v}_2 = \mathbf{0}$. Divide by a nonzero k_i, and subtract over; thus we have $\mathbf{v}_1$ as a scalar multiple of $\mathbf{v}_2$. $\square$

Proposition 6.4

A collection of non-zero vectors is linearly dependent if and only if at least one of the vectors (which implies all of them) can be expressed as a linear combination of the others. (Same proof as above, but with a longer sum.)

Example 6.2

Check whether the set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Is linearly independent.

Solution Note that $\mathbf{v}_2 - \mathbf{v}_1 = 2\mathbf{v}_4$. Thus the set is linearly dependent. However, $\mathbf{v}_3$ cannot be expressed in terms of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$. $\square$

Fact 6.3

In R^3, a collection of three vectors, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent if and only if the three arrows drawn from the origin to those points are not coplanar.

Definition 6.3. A linear subspace in R^n is a subset of vectors which:

- Contains the $\mathbf{0}$ vector;
- Is closed under linear combinations (that is, for any vectors in the set, a linear combination of these vectors is also in the set.)

It suffices to check that the subspace is closed under sums and scalar products.

7 August 30, 2018

We repeat Definition 6.3. Recall that a subspace that is a plane in R^3 isn't necessarily isomorphic to R^2.

Let $\mathsf{S} = \mathsf{S}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$. Then S is a subspace. It's easy to check that it fulfills the subspace axioms.

Some examples where we try to find whether a vector is in a span follow.

Then a small conversation on linear independence and equivalence of definitions.

8 August 31, 2018

Today we introduce the notions of basis, dimension, and coordinates.

Some concrete objects of a linear system $A\mathbf{x} = \mathbf{b}$ (A is $m \times n$) are:

- $\mathsf{C}(A)$ – a subspace of R^n
- The solution set of $A\mathbf{x} = \mathbf{0}$, a subspace of R^n
- THe solution set of $A\mathbf{x} = \mathbf{b}$

A central problem is: given $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathsf{R}^m$, can we form a specific linear combination

$$\sum_{i=1}^{r} k_i \mathbf{v}_i = k_1 \mathbf{v}_i + \cdots + k_r \mathbf{v}_r$$

The sum of this is a vector in R^m. As the k_i range over all real values, the sums sweep out the span $\mathsf{S}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$.

We repeat, again, definition 6.3.

Proposition 8.1

The span of any collection of vectors in R^m is a linear subspace.

Proof Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be the collection. We need to show that $\mathbf{0} \in \mathsf{S}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ and $\sum_{i=1}^{r} k_i \mathbf{x}_i \in \mathsf{S}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ for all $k_i \in \mathbb{R}$ and $\mathbf{x}_i \in \mathsf{S}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$.

The first part is trivial – no, really, the trivial linear combination shows that $\mathbf{0} \in \mathsf{S}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$.

The second part is also easy. Write the double sum composing $\mathbf{x}_i$ as a linear combination of $\mathbf{v}_i$.

$$\sum_{i=1}^{n} k_i \mathbf{x}_i = \sum_{i=1}^{n} k_i \sum_{j=1}^{r} k_i' \mathbf{v}_i = \sum_{i=1}^{n} \sum_{j=1}^{r} k_i k_j' \mathbf{v}_i = \sum_{i=1}^{r} k_i'' \mathbf{v}_i$$

by collecting terms. Thus we have that it is in the span. □

Proposition 8.2

Every linear subspace is the span of a collection of vectors.

Proof As it stands, take your (infinite) collection of vectors to be all vectors in the subspace. Silly. □

Proposition 8.3 (Revised)

Every linear subspace $\mathsf{L} \subseteq \mathsf{R}^m$ is the span of some finite collection of vectors.

Proof Much harder! We can generate parallograms to enumerate the line, spaces to enumerate the plane, and so on. □

Definition 8.1. The span of the columns of A is a subspaceof R^m, $\mathsf{C}(A)$. Any linear subspace in R^m is the column space of some matrix with m rows.

Proposition 8.4

The set of vectors $\mathbf{b} \in \mathsf{R}^m$ for which $A\mathbf{x} = \mathbf{b}$ is consistent is a linear subspace.

Proposition 8.5

The set of solutions to the system $A\mathbf{x} = \mathbf{0}$ is a linear subspace of R^n.

Caution: The solution to $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} \neq \mathbf{0}$ do not form a linear subspace (since it does not contain the zero vector).

Proof The set of homogeneous solutions is the span of the fundamental homogeneous solution:

$$\sum_i x_i \mathbf{h}_i \qquad \text{is the general homogeneous solution}$$

□

Proof We want to show that if $\mathbf{x}, \mathbf{x}'$are homogeneous solutions, then so are $\mathbf{x} + \mathbf{x}'$ and $k\mathbf{x}$ for $k \in \mathbb{R}$. We have

$$A\mathbf{x} = A\mathbf{x}' = \mathbf{0} \quad \text{so} \quad A(\mathbf{x} + \mathbf{x}') = \mathbf{0} \quad \text{and} \quad A(k\mathbf{x}) = \mathbf{0}$$

where the distributive property is able to be used by the argument of writing down the matrix-vector products as a sum of linear combinations. □

Remark In practice a linear subspace will be produced either as a span of a collection of vectors or a solution set of a homogeneous linear system.

We now turn to bases and candidates. Observe that in the case of homogeneous solutions, the coefficients x_i in the linear combination are uniquely determined by the sum $\sum_i x_i \mathbf{h}_i$. This expresses a fact that the $\mathbf{h}_i$ are linearly independent; that is, that they form a linearly independent collection.

Definition 8.2. A linearly independent collection $\mathbf{v}_1, \ldots \mathbf{v}_r$ of vectors spanning a subspace L is called a basis of L. Given a basis $\mathcal{L}$, we can define the coordinates of a vector $\mathbf{a}$ in that basis, with the entries a_i of the vector in that basis the coefficients of the linear combination of the basis vectors:

$$\mathbf{a} = \sum_i a_i \mathbf{v}_i$$

9 September 4, 2018

	Solution Exists	Solution is Unique
Pivot in Every Row (of coefficient matrix)	x	
Pivot in Every Column (of coefficient matrix)		x
$\mathsf{S}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathsf{R}^n$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent		x
$A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$		x

A matrix is a linear transform. A linear transform $\mathscr{T}$ between two vector spaces U and V has the properties

- $\mathscr{T}(\mathbf{0}) = \mathbf{0}$
- $\mathscr{T}(\sum_i c_i \mathbf{u}_i) = \sum_i c_i \mathscr{T}(\mathbf{u}_i)$

Example 9.1

Describe the linear transform

$$\mathscr{T}(a_0 + a_1 t + a_2 t^2) = (a_0 + a_1, a_2, -a_1)$$

Solution The transformation is $\mathscr{T} : \mathsf{P}_2(t) \to \mathsf{R}^3$. We have $\mathscr{T}(\mathbf{0}) = \mathbf{0}$ and $\mathscr{T}(\sum_i c_i \mathbf{u}_i) = \sum_i c_i \mathscr{T}(\mathbf{u}_i)$. $\square$

Although we don't necessarily have to deal with matrices in talking about linear transformations, we will argue that the properties of matrices apply in general to linear transformations. For the moment we will discuss the linear transformation $\mathscr{T}(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$. If A is $\mathsf{R}^{m\times n}$, then $\mathscr{T}$ takes vectors in R^n to R^m. We know that $\mathscr{T}(\mathbf{0}) = A\mathbf{0} = \mathbf{0}$. Additionally, we know that $\mathscr{T}(c\mathbf{u} + d\mathbf{v}) = cA\mathbf{u} + dA\mathbf{v} = c\mathscr{T}(\mathbf{u}) + d\mathscr{T}(\mathbf{v})$ which can be extensible to arbitrary linear combinations.

We now discuss the concept of basis vectors (or axes). We have $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\} \in \mathsf{V}$ is a basis vector set if and only if

- $\{\mathbf{b}_1, \ldots \mathbf{b}_n\}$ is linearly independent
- $\mathsf{S}\{\mathbf{b}_1, \ldots, \mathbf{b}_n\} = \mathsf{V}$

We define $\dim \mathsf{V}$ as the number of basis vectors of V. In the above example we have $\dim \mathsf{V} = n$.

The standard basis for R^n is $\{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Note that $\mathbf{e}_i$ takes less assumptions than the column vector form because for the column vector form we need to identify bases and a coordinate system. In particular, the vectors exist outside of axes; the axes allow us to put vectors in component form.

Let A be $m \times n$ and $\mathscr{T}(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$ be from $\mathsf{R}^n \to \mathsf{R}^m$. If

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{bmatrix}$$

then

$$A\mathbf{x} = A\left[\textstyle\sum_{i=1}^n x_i \mathbf{e}_i\right] = \sum_{i=1}^n x_i A\mathbf{e}_i = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n \mathbf{a}_i x_i$$

so $\mathbf{a}_i = \mathscr{T}(\mathbf{e}_i)$.

We get a new definition of a matrix. A matrix is a linear transformation that takes vectors in R^n to a vector in R^m.

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \end{bmatrix}$$

Let's take another look at the column space and null space. We have

$$\mathsf{C}(A) = \mathsf{S}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$$

and

$$\mathsf{N}(A) = \mathsf{S}\{\{\mathbf{x}_i\}\}$$

where $\mathbf{x}_h = \sum_i c_i \mathbf{x}_i$ and $A\mathbf{x}_h = \mathbf{0}$.

If A is $m \times n$, then $\mathsf{C}(A) \subseteq \mathsf{R}^m$ and $\mathsf{N}(A) \subseteq \mathsf{R}^n$.

Example 9.2

Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 1 & 2 \end{bmatrix}$. Find $\mathsf{C}(A)$ and $\mathsf{N}(A)$.

Solution The column space is the span of the pivot columns, so

$$\mathsf{C}(A) = \mathsf{S}\left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

Note that the vectors are linearly dependent, so we cut one off and get

$$\mathsf{C}(A) = \mathsf{S}\left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

These vectors are basis vectors for the column space. Since the space has dimension 2 and each vector has two entries, the column space is R^2.

To find $\mathsf{N}(A)$, we set up $A\mathbf{x} = \mathbf{0}$ and get

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

We have $R_1 \leftarrow 3R_1 - 2R_2$ and get

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

So $\mathsf{N}(A) = \mathsf{S}\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\}$, by parameterization. □

	Solution Exists	Solution is Unique
Pivot in every row	x	
Pivot in every column		x
$\mathsf{S}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathsf{R}^n$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent		x
$A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$		x
$\mathsf{C}(A) = \mathsf{R}^n$	x	
$\mathsf{N}(A) = \{\mathbf{0}\}$		x
$\dim \mathsf{C}(A) = m$	x	
$\dim \mathsf{N}(A) = 0$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a basis for R^m $(m = n)$	x	x

We introduce the rank-nullity theorem.

Theorem 9.1

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

$$\dim \mathsf{C}(A) + \dim \mathsf{N}(A) = \operatorname{rank} A + \operatorname{nullity} A = n$$

10 September 5, 2018

We cover the basics of a matrix as a linear transformation. First we discuss the concept of a basis.

Definition 10.1. A basis of a vector space V is an ordered collection $\{\mathbf{v}_1, \dots \mathbf{v}_n\}$ of vectors which is linearly independent and spans V,

Example 10.1

The standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of R^n is a basis, with $\mathbf{e}_{ij} = \delta_{ij}$ for index j in vector $\mathbf{e}_i$.

The coordinates of a vector $\mathbf{x} \in \mathsf{V}$ in a given basis are the coefficients of the unique expression $\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{v}_i$.

Example 10.2

In the standard basis of R^n we have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i \mathbf{e}_i$$

Example 10.3

In R^2, the condition for $\mathbf{v}_1$ and $\mathbf{v}_2$ to be a basis is that the vectors are not scalar multiples of each other.

A linear transformation relates vectors in different vector spaces i.e. $\mathsf{U} \to \mathsf{V}$. We need a map $\mathscr{T} : \mathsf{U} \to \mathsf{V}$ for this. We want this map to interact well with linear operations.

Definition 10.2. Such a $\mathscr{T}$ is a map that obeys the following properties as a function or operator:

- $\forall \mathbf{v}_1, \mathbf{v}_2 \in \mathsf{V}$, we have $\mathscr{T}(\mathbf{v}_1 + \mathbf{v}_2) = \mathscr{T}(\mathbf{v}_1) + \mathscr{T}(\mathbf{v}_2)$
- $\forall \mathbf{v} \in \mathsf{V}, k \in \mathbb{R}$, we have $\mathscr{T}(k\mathbf{v}) = k\mathscr{T}(\mathbf{v})$

From these properties we have

$$\mathscr{T}\left(\sum_i k_i \mathbf{v}_i\right) = \sum_i k_i \mathscr{T}(\mathbf{v}_i)$$

The space U is called the domain of $\mathscr{T}$ while V is called the codomain of $\mathscr{T}$.

Example 10.4

A rotation about O in R^2 is a linear transformation. So is a rotation about an axis. This is proved by the parallelogram rule. So is a reflection about the origin or a line. Note that translations are not linear transformations.

Proposition 10.1

If $\mathscr{T}$ is linear then $\mathscr{T}(\mathbf{0}) = \mathbf{0}$.

Proof We write $\mathbf{0} = 0\mathbf{a}$. So $\mathscr{T}(\mathbf{0}) = 0\mathscr{T}(\mathbf{a}) = \mathbf{0}$. □

Example 10.5 (Matrix Transformations)

For any fixed $m \times n$ matrix A, the map $\mathbf{x} \mapsto \mathbf{y} = A\mathbf{x}$ is a linear transformation $\mathscr{T} \colon \mathsf{R}^n \to \mathsf{R}^m$.

Note that solving the linear system $A\mathbf{x} = \mathbf{b}$ is equivalent to finding the vectors $\mathbf{x}$ which map to $\mathbf{b}$ under $\mathscr{T}_A$.

Definition 10.3. The kernel of a linear transformation $\mathscr{T}$ is the set

$$\{\mathbf{x} \in \mathsf{R}^n \colon \mathscr{T}(\mathbf{x}) = \mathbf{0}\}$$

Definition 10.4. The range of a linear transformation $\mathscr{T}$ is the set

$$\{\mathscr{T}(\mathbf{x}) \colon \mathbf{x} \in \mathsf{R}^n\} \subseteq \mathsf{R}^m$$

Proposition 10.2

The kernel and range of a linear transformation are linear subspaces of its domain and codomain (respectively).

Definition 10.5. A linear map is called onto, or surjective, if its range is exactly its codomain.

Definition 10.6. A linear map is called one-to-one, or injective, if for all $\mathbf{y} \in \mathsf{R}^m$, $\mathscr{T}(\mathbf{x}) = \mathbf{y}$ for at least one $\mathbf{x} \in \mathsf{R}^n$.

For the linear transformation $\mathscr{T}_A$, the statement that $\mathscr{T}_A$ is a surjection is equivalent to saying $\mathsf{C}(A) = \mathsf{R}^m$, or the system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$, or A has a pivot in every row.

For the linear transformation $\mathscr{T}_A$, the statement that $\mathscr{T}_A$ is an injection is equivalent to saying that $\mathsf{N}(A) = \{\mathbf{0}\}$, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, or A has a pivot in every column.

Proposition 10.3

If $m < n$ then $\mathscr{T}$ cannot be an injection; if $m > n$ then $\mathscr{T}$ cannot be a surjection.

Proof For matrix transformations, check the limitations on the number of pivots. $\square$

Theorem 10.1

Every linear transformation $\mathscr{T}\colon \mathsf{R}^n \to \mathsf{R}^m$ is equivalent to a matrix transformation $\mathscr{T}(\mathbf{x}) = A\mathbf{x}$.

Proof Exploit the linearity condition. Let $\mathbf{x} \in \mathsf{R}^n$, and $\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i$. Then, $\mathscr{T}(\mathbf{x}) = \mathscr{T}\left(\sum_{i=1}^{n} x_i \mathbf{e}_i\right) = \sum_{i=1}^{n} x_i \mathscr{T}(\mathbf{e}_i)$. Thus is just $A\mathbf{x}$ where $A = \begin{bmatrix} \mathscr{T}(\mathbf{e}_1) & \mathscr{T}(\mathbf{e}_2) & \ldots & \mathscr{T}(\mathbf{e}_n) \end{bmatrix}$. The columns of A are the images of the standard basis vectors. $\square$

11 September 6, 2018

A linear transform $\mathscr{T}\colon \mathsf{U} \to \mathsf{V}$ has domain U and codomain V. If $\mathscr{T}$ is represented by a matrix A, we say $\mathsf{C}(A)$ is the range of $\mathscr{T}(\mathbf{x})$, and $\mathsf{N}(A)$ is the kernel of $\mathscr{T}(\mathbf{x})$. Note that the range is a subspace of the codomain, while the kernel is a subspace of the domain. The solution set for $\mathscr{T}(\mathbf{x}) = \mathbf{b}$ is not a subspace. A function is onto/surjective when every element of the codomain has at least one element of the range mapping to it.

	Solution Exists	Solution is Unique
Pivot in every row	x	
Pivot in every column		x
$\mathsf{S}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathsf{R}^n$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent		x
$A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$		x
$\mathsf{C}(A) = \mathsf{R}^n$	x	
$\mathsf{N}(A) = \{\mathbf{0}\}$		x
$\dim \mathsf{C}(A) = m$	x	
$\dim \mathsf{N}(A) = 0$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a basis for R^m $(m = n)$	x	x
Range of $\mathscr{T}(\mathbf{x})$ is codomain	x	
Kernel of $\mathscr{T}(\mathbf{x}) = \{\mathbf{0}\}$		x
$\mathscr{T}(\mathbf{x})$ is onto	x	
$\mathscr{T}(\mathbf{x})$ is one-to-one		x

We introduce the notion of coordinates. Say we have a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Then if $\mathbf{x} = 2\mathbf{v}_1 + \mathbf{v}_2$ the explicit coordinates of $\mathbf{x}$ under $\mathcal{B}$ are $\left[\mathbf{x}\right]_{\mathcal{B}} = (2, 1)$ while the implicit coordinates are $\mathbf{x} = (2, 1)$.

We now discuss matrix operations. The first four are valid if we treat the matrix as a vector.

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $A + 0 = A$
4. $c(A + B) = cA + cB$
5. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
6. $A(c\mathbf{u}) = cA\mathbf{u}$
7. $A(BC) = (AB)C$
8. $A(B + C) = AB + AC$
9. $(B + C)A = BA + CA$
10. $c(AB) = (cA)B = A(cB)$
11. $I_m A = AI_n = A$

We say that $0 = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$. We say that I_m is an $m \times m$ matrix with $\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$.

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{bmatrix}$. Then

$$A + B = \begin{bmatrix} \mathbf{a}_1 + \mathbf{b}_1 & \mathbf{a}_2 + \mathbf{b}_2 & \dots & \mathbf{a}_n + \mathbf{b}_n \end{bmatrix}$$

where A and B have to both be $m \times n$ matrices. Let A be an $m \times n$ matrix, and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{bmatrix}$ be an $n \times p$ matrix. Then

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_n \end{bmatrix}$$

Since A is $m \times n$ and B is $n \times p$, AB is $m \times p$.

Matrix multiplication is a composition of the associated linear transformations. Let $\mathscr{T}(\mathbf{x}) = A\mathbf{x}$ and $\mathscr{S}(\mathbf{x}) = B\mathbf{x}$. Then $BA\mathbf{x}$ corresponds to $\mathscr{S}(\mathscr{T}(\mathbf{x}))$.

12 September 7, 2018

We begin with an example where linear map language is useful.

Example 12.1

Say we have a linear map $\mathscr{T} : \mathsf{R}^4 \to \mathsf{R}^4$. We know that $\mathscr{T}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$ and $\mathscr{T}\left(\begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 0 \\ 4 \\ 3 \end{bmatrix}$. We try to find $\mathscr{T}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\right)$.

Solution We cannot convert $\mathscr{T}$ to matrix form because there is not enough information to find the standard matrix of $\mathscr{T}$. But we can solve the problem by exploiting linearity. Note that

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{5}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}\right)$$

So

$$\mathscr{T}\left(\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right) = \frac{1}{5}\left(\mathscr{T}\left(\begin{bmatrix}1\\2\\3\\4\end{bmatrix}\right) + \mathscr{T}\left(\begin{bmatrix}4\\3\\2\\1\end{bmatrix}\right)\right) = \begin{bmatrix}1\\0\\1\\1\end{bmatrix}$$

□

Recall $m \times n$ matrices correspond to linear maps from R^n to R^m. Note that $\mathscr{T}_A(\mathbf{x}) = A\mathbf{x}$.

Now let's compose the maps. Say we have $\mathsf{R}^m \xrightarrow{\mathscr{S}} \mathsf{R}^l$ and $\mathsf{R}^n \xrightarrow{\mathscr{T}} \mathsf{R}^m$. Then $\mathscr{T} \circ \mathscr{S}$ is $\mathsf{R}^m \to \mathsf{R}^l$. Note that $\mathscr{S}$ corresponds to an $l \times m$ matrix, and $\mathscr{T}$ corresponds to an $m \times n$ matrix.

Proposition 12.1

The composition of two linear transformations is a linear transformation.

Proof Check that the composition works with linear transformations. □

Definition 12.1. The matrix product $\underbrace{B}_{l\times m}\underbrace{A}_{m\times n}$ is the $l \times n$ matrix corresponding to $\mathscr{S} \circ \mathscr{T}$.

Caution: the order of matrix multiplication matters. Even if $l = n$ and so AB is also defined, it would usually differ from BA.

We have three computations of BA:

- entry-by-entry (dot product):
$$(BA)_{ij} = \sum_{k=1}^{m} b_{ik}a_{kj}$$
- column-by-column
$$BA = \begin{bmatrix} B\mathbf{a}_1 & B\mathbf{a}_2 & \dots & B\mathbf{a}_n \end{bmatrix}$$
- row-by-row
$$BA = \begin{bmatrix} \mathbf{b}_1^T A & \mathbf{b}_2^T A & \dots & \mathbf{b}_m^T A \end{bmatrix}$$

Note that row-vector-matrix multiplication is done by using the weights of the row vector as weights in the linear combination of the rows of the matrix.

Some other operations involve matrices.

- addition, entry by entry (note that the matrices have to be the same size); written as $A + B$
- scalar multiplication, entry by entry; written as kA
- transposition: making the swap $(A)_{ij} \leftrightarrow (A)_{ji}$; written as A^T

Note that $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$.

Some properties of matrix operations:

- $(AB)C = A(BC)$ (proof: composition of linear maps is associative.)

- $(A+B)C = AC + BC$
- $D(A+B) = DA + DB$
- $(A+B)(C+E) = AC + BC + BC + BE$
- $(kA)B = k(AB) = A(kB)$
- additive properties give a vector space structure
- $(AB)^T = B^T A^T$

13 September 10, 2018

We start with the topic of the inverse matrix. Consider an $m \times n$ matrix A that corresponds to a linear map $\mathscr{T}_A$ from R^n to R^m. In the case that $\mathscr{T}_A$ is both injective (one-to-one) and surjective (onto), then $\mathscr{T}$ is bijective and we can define an inverse $\mathscr{T}_A^{-1}$ from R^m to R^n. The inverse $\mathscr{T}_A^{-1}$ has the property that for all $\mathbf{y} \in \mathsf{R}^m$ and $\mathbf{x} \in \mathsf{R}^n$, $\mathscr{T}_A^{-1}(\mathbf{y})$ is the unique vector $\mathbf{x}$ such that $\mathscr{T}_A(\mathbf{x}) = \mathbf{y}$. In particular $\mathscr{T}^{-1}(\mathscr{T}(\mathbf{x})) = \mathbf{x}$.

Fact 13.1

We have for some linear map $\mathscr{T}$ that $\mathscr{T}^{-1}$ is also a linear map. We can check easily additivity and multiplicativity by setting up the checks and applying $\mathscr{T}$ to everything.

If $\mathscr{T}$ corresponds to a matrix A then $\mathscr{T}^{-1}$ must correspond to a matrix B such that $AB = BA = I$, given that the inverse exists. This matrix B is called the inverse of A and is written A^{-1}.

Recall that a linear map $\mathscr{T}$ represented by a matrix A is injective when there is a pivot in every column and surjective when there is a matrix in every row. When A has pivots in every row and column, then A is bijective (in particular, if A is $m \times n$, then $m = n$).

Proposition 13.1

The matrix A is invertible if $\operatorname{rref}(A) = I_n$ (in particular A is a square matrix).

Remark A non-square matrix A can have left inverses B such that $BA = I_n$ (if A is injective) or right inverses C such that $AC = I_m$ (if A is surjective). If A is square then $B = C = A^{-1}$.

Example 13.1

We have

$$A = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}$$

representing a rotation by φ radians counterclockwise around the origin. The inverse is just the opposite rotation:

$$A^{-1} = \begin{bmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{bmatrix}$$

Trying, we have $AA^{-1} = A^{-1}A = I$.

Theorem 13.1 (Magic Method For Finding A^{-1})
If A is invertible and $n \times n$, then
$$\operatorname{rref}\begin{bmatrix} A & I_n \end{bmatrix} = \begin{bmatrix} I_n & A^{-1} \end{bmatrix}$$

Proof Write the equation $A\mathbf{x} = \mathbf{b}$. Then we have the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$. The solution appears on the last column of the reduced echelon form of $\operatorname{rref}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} I_n & \mathbf{x} \end{bmatrix}$. Write out solutions of the system $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$. Then we have $A\mathbf{s}_1 = \mathbf{e}_1, A\mathbf{s}_2 = \mathbf{e}_2, \ldots, A\mathbf{s}_n = \mathbf{e}_n$. Write out the row reduction finding all these solutions at once: $\begin{bmatrix} A & \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \end{bmatrix}$. Doing these n system computations at once yields $\begin{bmatrix} I_n & \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_n \end{bmatrix}$. Well, $A\begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{s}_1 & A\mathbf{s}_2 & \ldots & A\mathbf{s}_n \end{bmatrix} = I_n$. $\square$

If A is not invertible then $\operatorname{rref}\begin{bmatrix} A & I_m \end{bmatrix} = \begin{bmatrix} \operatorname{rref} A & M \end{bmatrix}$ and M has the property that $MA = \operatorname{rref} A$.

An abstract vector space is a set V with two operations:

- "+"
  - $\mathsf{V} \times \mathsf{V} \to \mathsf{V}$
  - $(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} + \mathbf{w}$
- "·"
  - $\mathbb{R} \times \mathsf{V} \to \mathsf{V}$
  - $(k, \mathbf{v}) \mapsto k\mathbf{v}$

They also have a set of properties under these operations that must hold. These properties ensure that the linear algebraic techniques hold.

Example 13.2
Some examples of vector spaces: R^m of row vectors, R^n of column vectors, $\mathsf{M}_{m \times n}(\mathbb{R})$ of $m \times n$ matrices. Note that addition and scalar multiplication are the only allowed operations.

Example 13.3
We can associate vectors in $\mathsf{M}_{2\times 2}(\mathbb{R})$ with R^4, and vice-versa:
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \leftrightarrow \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$
This allows vector addition and scalar multiplication to hold. If this correspondence is injective and surjective, then it is a bijection (in this case it is called a linear isomorphism).

In R^4, we have the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$, which are the standard coordinates. Under the isomorphism with $\mathsf{M}_{2\times 2}(\mathbb{R})$, this goes to the basis $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. We must check that this is a linearly independent collection of matrices which span $\mathsf{M}_{2\times 2}$, which is easy to verify. In this way we see that a, b, c, d, the weights of the constructed linear combination in the span, are coordinates.

Example 13.4

One example is P, the space of polynomials. We have each element of P is a vector $\mathbf{p}(x) = \sum_{i=0}^{k} a_i x^i$ for some k. It is easy to check that the aforementioned vector space properties hold. There is an isomorphism between P and "R^∞" where the ith coordinate of the column vector is a_i. In particular P_n is isomorphic to R^{n+1}.

14 September 11, 2018

Some basic matrix operations, composition of linear maps, etc. These are already known/covered in lecture.

We now cover the matrix power. If we have

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

then A is square and the exponent is valid. We have the matrix exponent:

$$\exp(A) = \sum_{i=0}^{\infty} \frac{A^i}{i!}$$

We now cover the transpose, which flips a matrix across its diagonal. We have some properties of the transpose:

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

An elementary matrix E is equivalent to an elementary row operation if we perform EA. It is easy to see that E has to be invertible, and if A is $m \times n$, then E is $m \times m$. There are also column operations, obtained by right-multiplication by E.

Note that if we have a problem $AB = C$ to find B, we can row reduce the matrix $\begin{bmatrix} A & C \end{bmatrix}$ to get $\begin{bmatrix} I & B \end{bmatrix}$.

If $AB = BA = I$ then B is the inverse of A; $B = A^{-1}$. To find the inverse we can row reduce $\begin{bmatrix} A & I \end{bmatrix}$ to get $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.

15 September 12, 2018

Definition 15.1. A linear subspace of a vector space is a subset which is closed under addition and scalar multiplication.

We say $\mathsf{W} \subset \mathsf{V}$ is a linear subspace if

- $\forall \mathbf{w}_1, \mathbf{w}_2 \in \mathsf{W}, \mathbf{w}_1 + \mathbf{w}_2 \in \mathsf{W}$
- $\forall k \in \mathbb{R}, \mathbf{w} \in \mathsf{W}, k\mathbf{w} \in \mathsf{W}$

- $\mathsf{W} \neq \emptyset \leftrightarrow \mathbf{0} \in \mathsf{W}$

Proposition 15.1

A linear subspace of a vector space is also a vector space.

Our goal is to identify, if possible, a given vector space, through an isomorphism with R^n, so we can use matrix techniques. (This is possible when the vector space is spanned by a finite collection of vectors; then n will be the dimension of V.) Finish the story with the "change of coordinates" formula, completing the isomorphism.

The basis of a vector space is a linearly independent (ordered) collection which spans the space. An example is the standard basis of R^n. The coordinates of a vector $\mathbf{v}$ in the basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ are the coefficients of the unique expression of $\mathbf{v}$ as a linear combination of the basis vectors.

Example 15.1

The monomials $1, x, x^2, \ldots$ form a basis for P. Every polynomial can be expressed as a linear combination of these monomials.

Theorem 15.1

Let $\mathscr{T} : \mathsf{R}^n \to \mathsf{V}$ have an inverse map $\mathscr{T}^{-1}\mathsf{V} \to \mathsf{R}^n$ and be a linear isomorphism. Then

- The vectors $\mathscr{T}(\mathbf{e}_1), \ldots, \mathscr{T}(\mathbf{e}_n)$ form a basis of V.
- The entries of $\mathscr{T}^{-1}(\mathbf{v})$ are the coordinates of $\mathbf{v}$ in that basis.

In effect, there are two ways to indicate a subspace of R^n. They are:

1. as a solution set of a homogeneous linear system (null space, kernel)
2. as a span of a finite collection of vectors (column space, range)

With row vectors, one would think of the second term as the row space of a matrix.

Definition 15.2. The left nullspace $\mathsf{LN}(A)$ of an $m \times n$ matrix A in the set of vectors $\mathbf{y}^T$ in R^m satisfying $\mathbf{y}^T A = \mathbf{0}$.

Remark We have

$$\left(\mathbf{y}^T A\right)^T = A^T \mathbf{y}$$

so

$$(\mathsf{LN}(A))^T = \mathsf{N}\left(A^T\right)$$

We have four linear subspaces associated with an $m \times n$ matrix A:

- $\mathsf{C}(A) \subseteq \mathsf{R}^m$
- $\mathsf{R}(A) \subseteq \mathsf{R}^n$ (rows)
- $\mathsf{N}(A) \subseteq \mathsf{R}^n$
- $\mathsf{LN}(A) \subseteq \mathsf{R}^m$ (rows)

Recall that the pivot columns of A form a basis for $\mathsf{C}(A)$ (caution: not the pivot columns of rref(A)).

The fundamental homogeneous solutions of $A\mathbf{x} = \mathbf{0}$ form a basis for $\mathsf{N}(A)$.

The nonzero rows of rref(A) form a basis of $\mathsf{R}(A)$.

To find a basis of $\mathsf{LN}(A)$ row reduce $\begin{bmatrix} A & I_m \end{bmatrix}$. Each row vector in the right square corresponding to a zero row of rref A gives you a basis vector for $\mathsf{LN}(A)$.

Fact 15.1

The $\mathsf{LN}(A)$ gives you a complete set of conditions for testing consistency of $A\mathbf{x} = \mathbf{b}$. The system is consistent if and only if

$$\mathbf{y}^T\mathbf{b} = 0 \quad \text{for all} \quad \mathbf{y} \in \mathsf{LN}(A)$$

16 September 12, 2018

Note that when we're doing column operations, we have to right-multiply.

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A^T)^{-1} = (A^{-1})^T = A^{-T}$

Example 16.1

If $C = AB$ and $D = A + B$ and C and D are invertible:

- If C^{-1} exists then A and B are both invertible.
- If D^{-1} exists then we don't have information about A and B.

	Solution Exists	Solution is Unique
Pivot in every row	x	
Pivot in every column		x
$\mathsf{S}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathsf{R}^n$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent		x
$A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$		x
$\mathsf{C}(A) = \mathsf{R}^n$	x	
$\mathsf{N}(A) = \{\mathbf{0}\}$		x
$\dim \mathsf{C}(A) = m$	x	
$\dim \mathsf{N}(A) = 0$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a basis for R^m $(m = n)$	x	x
Range of $\mathscr{T}(\mathbf{x})$ is codomain	x	
Kernel of $\mathscr{T}(\mathbf{x}) = \{\mathbf{0}\}$		x
$\mathscr{T}(\mathbf{x})$ is onto	x	
$\mathscr{T}(\mathbf{x})$ is one-to-one		x
$\mathsf{C}\left(A^T\right) = \mathsf{R}^n$		x
$\mathsf{N}\left(A^T\right) = \{\mathbf{0}\}$	x	
A^{-1} exists $(m = n)$	x	x
A is row equivalent to I $(m = n)$	x	x
A^T is invertible $(m = n)$	x	x
$CA = AC = I$ where C exists	x	x

The determinant indicates the ratio of the measure of the new space generated by the sums of the n columns of the $n \times n$ matrix.

Say we have $x = f(u, v)$ and $y = g(u, v)$. Then by chain rule we have

$$\mathrm{d}x = f_u\,\mathrm{d}u + f_v\,\mathrm{d}v$$

$$\mathrm{d}v = g_u\,\mathrm{d}u + g_v\,\mathrm{d}v$$

so

$$\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = \begin{bmatrix} f_u & f_v \\ g_u & g_v \end{bmatrix} \begin{bmatrix} \mathrm{d}u \\ \mathrm{d}v \end{bmatrix}$$

We define the Jacobian as the measure scaling form:

$$J = \begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix}$$

This scales to arbitrary dimensions.

If A is $n \times n$, fix a column i or a row j. Then

$$\det A = \sum_{\substack{i=1 \\ \text{or} \\ j=1}}^{n} (-1)^{i+j} a_{ij} \det A_{ij}$$

where A_{ij} is the submatrix generated by eliminating row i and column j from A.

Note for a two by two matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we have $\det A = ad - bc$.

17 September 14, 2018

A theorem of left null spaces is

Theorem 17.1

When row reducing $\begin{bmatrix} A & I_m \end{bmatrix}$, each zero row in rref$(A)$ gives you a basis vector for $\mathsf{LN}(A)$.

Remark When A is invertible, $\mathsf{LN}(A) = \{\mathbf{0}^T\}$.

Proof We have

$$\mathbf{y}^T A = \mathbf{0}^T \to \mathbf{y}^T A A^{-1} = \mathbf{y}^T = \mathbf{0}^T A^{-1} = \mathbf{0}^T$$

□

A basis for $\mathsf{LN}(A)$ is a complete and minimal set of defining equations for $\mathsf{C}(A)$ if and only if the system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{y}^T \mathbf{b} = \mathbf{0}$ for every $\mathbf{y}^T$ in a basis for $\mathsf{N}(A)$.

Example 17.1

Find a basis for $\mathsf{LN}\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right)$.

Solution We row reduce $\begin{bmatrix} A & I \end{bmatrix}$:

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 4/3 & -1/3 & 0 \\ 0 & 0 & 0 & -1/6 & 1/3 & -1/6 \end{bmatrix}$$

We spot the zero row in A which corresponds to $\begin{bmatrix} -1/6 & 1/3 & -1/6 \end{bmatrix}$. Thus a basis for $\mathsf{LN}(A)$ is $\left\{\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}\right\}$. Note that this shows that $\mathbf{y}^T \mathbf{b} = b_1 - 2b_2 + b_3 = 0$ and the second row is the average of the first and last rows, which is true. □

Now we move onto dimension of a vector space.

Theorem 17.2

Let V be a vector space spanned by a finite collection of vectors. Then,

1. V has a finite basis.
2. Any two bases have the same number of elements. This number is the dimension $\dim \mathsf{V}$.

Example 17.2

We know that R^n has dimension n. We also know that P, the space of polynomials, is not spanned by any finite set of vectors, so it is called infinite-dimensional. On the other hand, we have that P_n is n-dimensional (with basis $\{1, x, x^2, \ldots, x^n\}$).

Theorem 17.3

Fix some n. Let V be a vector space with dimension n.

1. Any linearly independent collection in V has no more than n vectors.
2. Any spanning collection has at least n vectors.
3. Any linearly independent collection with exactly n vectors spans V.
4. Any spanning collection with exactly n vectors is also linearly independent (i.e. a basis).
5. Any linearly independent collection can be completed to a basis.
6. From any spanning collection, a basis can be extracted.
7. Every basis of V has exactly n vectors.

Proofs are done by looking at a sample set of vectors, making a matrix out of them, and doing matrix operations.

Definition 17.1. The rank of a matrix A is the dimension of $\mathsf{C}(A)$. The nullity of a matrix is the dimension of $\mathsf{N}(A)$.

Theorem 17.4 (Rank-Nullity Theorem)

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

$$\dim \mathsf{C}(A) + \dim \mathsf{N}(A) = \operatorname{rank} A + \operatorname{nullity} A = n$$

18 September 17, 2018

Sometimes we denote

$$\text{row-rank}(A) = \dim \mathsf{R}(A) = \text{number of pivots in } A = \operatorname{rank} A$$

and

$$\dim \mathsf{LN}(A) + \text{row-rank}(A) = m$$

We now introduce a change of basis formula, which is used to compare calculations done in different bases.

Recall in general that the $\mathcal{B}$-coordinates of a vector $\mathbf{v}$ are given by $P_{\mathcal{B}}^{-1}\mathbf{v}$, where $P_{\mathcal{B}}$ is the matrix given by the columns of $\mathcal{B}$ in the standard basis.

The change of coordinates from the standard coordinates to $\mathcal{B}$-coordinates is performed by the inverse matrix. More generally, when changing from the standard basis to a new basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, the coordinates change as follows:

$$[\mathbf{x}]_{\mathcal{B}} = B^{-1}\mathbf{x}$$

where $B = \begin{bmatrix} \mathbf{b}_1 & \ldots & \mathbf{b}_n \end{bmatrix}$, because hte coordinates of $[\mathbf{x}]_{\mathcal{B}}$ must satisfy:

$$\sum_{i=1}^{n} \left([\mathbf{x}]_{\mathcal{B}}\right)_i \mathbf{b}_i = B[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$$

In general, if we have two bases $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_k\}$, they must be related by a change of basis matrix A from $\mathcal{B}$ to $\mathcal{C}$. We know that A expresses the vectors of $\mathcal{C}$ in terms of the old basis $\mathcal{B}$. The columns of A are the combination of the $\mathbf{c}_i$ in the basis $\mathcal{B}$. Equivalently,

$$\begin{bmatrix}\mathbf{c}_1 & \dots & \mathbf{c}_k\end{bmatrix} = \begin{bmatrix}\mathbf{b}_1 & \dots & \mathbf{b}_k\end{bmatrix} A$$

If we write $\mathbf{c}_i = \mathbf{b}_i a$ then a is the coordinate of $\mathbf{c}_i$ in the basis $\mathbf{b}_i$.

Proposition 18.1

The coordinates of a vector $\mathbf{v}$ in the bases $\mathcal{B}$ and $\mathcal{C}$ are related by the formula

$$[\mathbf{v}]_{\mathcal{C}} = A^{-1}[\mathbf{v}]_{\mathcal{B}}$$

When do two linear systems $A\mathbf{x} = \mathbf{b}$ and $A'\mathbf{x} = \mathbf{b}'$ have the same solution sets? If both systems are consistent, then this occurs when

$$\operatorname{rref}\begin{bmatrix}A & \mathbf{b}\end{bmatrix} = \operatorname{rref}\begin{bmatrix}A' & \mathbf{b}'\end{bmatrix}$$

When are all solutions of one equation also solutions of the other equation?

Assume first that $\mathbf{b} = \mathbf{b}' = \mathbf{0}$, then find the homogeneous solutions of the first equation and check that they solve the second equation. This checks that $\mathsf{N}(A) \subseteq \mathsf{N}(A')$.

Also we can check that $\mathsf{R}(A') \subseteq \mathsf{R}(A)$ by row reducing $\begin{bmatrix}A \\ A'\end{bmatrix}$ and checking that the A' block has all zeros.

If $\mathbf{b}, \mathbf{b}' \neq \mathbf{0}$ we must also check that the particular solution of the first equation also solves the second equation.

To ensure that $A\mathbf{x} = \mathbf{b}$ is consistent for some $\mathbf{b}$ we can either row reduce until we get the row $\begin{bmatrix}0 & 0 & \dots & 0 & b\end{bmatrix}$ or check that the left nullspace is zero.

19 September 18, 2018

For any triangular matrix T, we have $\det T = \prod_i t_{ii}$. Some properties of determinants:

- $\det AB = \det A \det B$
- $\det A^T = \det A$
- A row switch or column switch leads to $\det E = -1$.
- A row replacement $R_j \leftarrow \sum_{i=1}^{n} c_i R_i$ leads to $\det E = c_j$
- $\det A^{-1} = (\det A)^{-1}$ (so $\det A \neq 0$)

	Solution Exists	Solution is Unique
Pivot in every row	x	
Pivot in every column		x
$\mathsf{S}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \mathsf{R}^n$	x	
$\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is linearly independent		x
$A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$		x
$\mathsf{C}(A) = \mathsf{R}^n$	x	
$\mathsf{N}(A) = \{\mathbf{0}\}$		x
$\dim \mathsf{C}(A) = m$	x	
$\dim \mathsf{N}(A) = 0$	x	
$\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is a basis for R^m $(m = n)$	x	x
Range of $\mathscr{T}(\mathbf{x})$ is codomain	x	
Kernel of $\mathscr{T}(\mathbf{x}) = \{\mathbf{0}\}$		x
$\mathscr{T}(\mathbf{x})$ is onto	x	
$\mathscr{T}(\mathbf{x})$ is one-to-one		x
$\mathsf{C}\left(A^T\right) = \mathsf{R}^n$		x
$\mathsf{N}\left(A^T\right) = \{\mathbf{0}\}$	x	
A^{-1} exists $(m = n)$	x	x
A is row equivalent to I $(m = n)$	x	x
A^T is invertible $(m = n)$	x	x
$CA = AC = I$ where C exists	x	x
$\det A \neq 0$ (so $m = n$)	x	x
$\det A^T \neq 0$ (so $m = n$)	x	x
row vectors of A lin. indep.	x	

The last entry is shown by demonstrating that the columns of A^T are linearly independent, so there is a pivot in every column of A^T, and so there is a pivot in every row of A.

Let $A_i(\mathbf{b}) = \begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n\end{bmatrix}$. Then given $A\mathbf{x} = \mathbf{b}$, we have

$$\begin{aligned} AI_i(\mathbf{x}) &= A\begin{bmatrix}\mathbf{e}_1 & \dots & \mathbf{e}_{i-1}, \mathbf{x} & \mathbf{e}_{i+1} & \dots, \mathbf{e}_n\end{bmatrix} \\ &= \begin{bmatrix}A\mathbf{e}_1 & \dots & A\mathbf{e}_{i-1} & A\mathbf{x} & A\mathbf{e}_{i+1} & \dots & A\mathbf{e}_n\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n\end{bmatrix} \end{aligned}$$

Thus we have

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

In this way we can find the inverse. Let $AB = I$. Then

$$A\mathbf{b}_i = \mathbf{e}_i$$

so $\mathbf{b}_{ii} = \det A_i(\mathbf{e_i}) / \det A$. In particular for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ being 2×2 we have $A^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} / (ad - bc)$.

The change of coordinates matrix

$$\underset{\mathcal{C} \leftarrow \mathcal{B}}{P} = \begin{bmatrix}[\mathbf{b}_1]_{\mathcal{C}} & \dots & [\mathbf{b}_n]_{\mathcal{C}}\end{bmatrix}$$

is used in the following way:

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

where we can either do the straight multiplication to find $[\mathbf{x}]_{\mathcal{C}}$ or row reduce the augmented matrix $\begin{bmatrix}\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} & [\mathbf{x}]_{\mathcal{C}}\end{bmatrix}$ to find $[\mathbf{x}]_{\mathcal{B}}$.

20 September 19, 2018

The determinant of an $n \times n$ matrix A is a mapping from $\mathsf{M}_{n\times n} \to \mathbb{R}$. It denotes the measure of the parallelotope generated by having the edges of A as its column.

For a 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, we have $\det A = a_{11}a_{22} - a_{12}a_{21}$.

We need to determine what it means to establish a measure function with domain being a set of n vectors. Some properties of this function are:

1. Multilinearity. The measure function $M(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is linear in each vector variable $\mathbf{x}_1, \ldots, \mathbf{x}_n$. That is,
$$M(k\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = kM(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$
and
$$M(\mathbf{x}_1 + \mathbf{x}_1', \mathbf{x}_2, \ldots, \mathbf{x}_n) = M(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) + M(\mathbf{x}_1', \mathbf{x}_2, \ldots, \mathbf{x}_n)$$
This generalizes to summing the permutations of the measure function if two or more entries are added (and can obviously get really complicated really quickly).
2. If two or more of the vectors agree, the measure is zero.
3. The unit parallelotope is 1:
$$M(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = 1$$

Theorem 20.1

There exists a unique function, denoted by det and called the determinant, on the set of $n \times n$ matrices with the properties above. Namely, $\det A$ is multilinear in the rows of A, $\det A = 0$ if two rows of A agree, and $\det I_n = 1$.

We have three formulas for $\det A$.

$$\det A = \begin{cases} 0 \text{ if } A \text{ is not invertible (“singular”)} \\ \left(\prod \text{pivots}\right)\left(-1^{\#\text{ of row exchanges}}\right) \text{ if } A \text{ is invertible} \end{cases}$$

Moreover, det has the following additional properties:

a. It is “skew-symmetric” in the rows. (A row exchange switches the sign).

b. Test for invertibility: A is invertible $\leftrightarrow \det A \neq 0$

c. $\det A^T = \det A$ (what goes for rows for columns).

d. $\det AB = \det A \det B$

e. $\det(A+B) \neq \det A + \det B$ in general.

f. Big formula for the determinant:

$$\det A = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^{n} A_{i\sigma(i)}$$

where S is the set of permutations of $1, \ldots, n$ and σ is one such permutation. In particular $\sigma(i)$ maps i to j where i is the original index and j is the index in the permuted list. We have $\epsilon(\sigma)$ is the Levi-Civita symbol, and we determine its value by:

$$\epsilon(\sigma) = (-1)^{\# \text{ of out-of-order pairs}}$$

g Cofactor expansion: fix a row i or a column j. Then, sum over the non-fixed element to get

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \quad \text{or} \quad \det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}$$

where A_{ij} is the matrix obtained by deleting row i and column j from A. We denote the cofactor $C_{ij} = (-1)^{i+j} \det A_{ij}$.

21 September 20, 2018

There are ten properties of vector spaces. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathsf{V}$. Then

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

Example 21.1 (Sobolev Space)

Let L_n be the set of functions $\mathbf{f}$ that are differentiable everywhere n times. These spaces are important in differential equations which linearly map one Sobolev space to another, giving us a method to solve them.

Example 21.2 (Cyclic Space)

Let C_k be the set of functions $\mathbf{f}$ that satisfy $\mathbf{f}(t+k) = \mathbf{f}(t)$, where k is the period. Then we argue that

$$S = \left\{1, \sin\left(\frac{2\pi t}{k}\right), \sin\left(\frac{4\pi t}{k}\right), \ldots, \sin\left(\frac{2\pi n t}{k}\right), \cos\left(\frac{2\pi t}{k}\right), \cos\left(\frac{4\pi t}{k}\right), \ldots, \cos\left(\frac{2\pi n t}{k}\right)\right\}$$

for $n \in \mathbb{N}$. We later expand this to Fourier series of the form $g(t) = a_0 + \sum_k a_k \sin(kt) + b_k \cos(kt)$.

Example 21.3 (Polynomial Space)

Let P_n be the set of polynomials $\mathbf{p} = \sum_{i=0}^{n} a_i x^i$ up to order n.

22 September 21, 2018

We discuss more about determinants. First, we discuss a formula for the inverse matrix.

Let C be the cofactor matrix with entries C_{ij}. Then $AC^T = C^T A = I \det A$. If A is invertible then $A^{-1} = C^T / \det A$.

We review Cramer's rule. See above. Note that this follows from the definition of the cofactor matrix.

23 September 24, 2018

To check linear independence of a set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$, write them as columns of a matrix A and check whether there is a pivot in every row (or whether there is a maximal square submatrix with nonvanishing determinant); to check that the set has a maximal span, check whether there is a pivot in every column.

To check whether multiple vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ are in the span of the set, write them as the augments of the augmented matrix $\begin{bmatrix} A & B \end{bmatrix}$ using A and check that no row has zero in the matrix A and nonzero augments.

How do we check that the null space of A is a subset of the span of a set of vectors (that is, that $\mathsf{N}(A) \subseteq \mathsf{S}\{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$)?

1. Find a basis $\{\mathbf{h}_1, \ldots, \mathbf{h}_k\}$ of the homogeneous solution to $A\mathbf{x} = \mathbf{0}$. Check that $\mathsf{S}\{\mathbf{h}_1, \ldots, \mathbf{h}_k\} \subseteq \mathsf{S}\{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ as before: row reduce $\begin{bmatrix} \mathbf{b}_1 & \ldots & \mathbf{b}_m & \mathbf{h}_1 & \ldots & \mathbf{h}_k \end{bmatrix}$.
2. Note that the $\mathsf{S}\{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ denoted by the linear equations come from a basis of $\mathsf{LN}(\mathbf{b}_1, \ldots, \mathbf{b}_m)$. We must check either that each equation holds for each $\mathbf{h}_i, 1 \leq i \leq k$, or check that $\mathsf{LN}(B) \subseteq \mathsf{R}(A)$ by row reducing the matrix $\begin{bmatrix} A \\ \text{basis for } \mathsf{LN}(B) \end{bmatrix}$ to find $\begin{bmatrix} \operatorname{rref}(A) \\ 0 \end{bmatrix}$

How do we find the inverse of a 3×3 matrix?

1. $\operatorname{rref} \begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} I & A^{-1} \end{bmatrix}$ if A is invertible
2. Cramer's rule using the cofactor matrix: $A^{-1} = C^T / \det A$ (applies for all square matrices, but the determinants are computationally intensive).

How do we find a linear transformation $\mathscr{T} \colon \mathsf{R}^3 \to \mathsf{R}^3$ with $\mathscr{T}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathscr{T}\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}$? Write the problem as $A\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 6 \end{bmatrix}$, so we get 6 equations in 9 unknowns, which we can solve. Alternatively recognize that the map $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x \\ x+y \\ x+y+z \end{bmatrix}$ works and that's all we need.

24 September 27, 2018

Midterm review: 100%, or $(100-\epsilon)$%. Whoo.

Some random determinant stuff, regarding Cramer's rule and the cofactor expansion. This is all stuff we've written before, so nothing of note.

An introduction to tensors. Tensors are a superset of matrices; in particular matrices are second order tensors. Let $\mathscr{T} \colon \mathsf{V} \to \mathsf{U}$ be a linear transform, where a basis for V is $\mathcal{B}$ and a basis for U is $\mathcal{C}$, then the matrix representing $\mathscr{T}$ is given by $\mathbf{b}_i \otimes \mathbf{c}_j$ and the matrix is $A_{\mathcal{B}\times\mathcal{C}}$.

25 September 28, 2018

We start by discussing the Fibonacci numbers in terms of a linear recurrence:

$$F_{n+2} = F_{n+1} + F_n, \quad F_0 = F_1 = 1$$

Let a_i and b_i be the number of adult and baby rabbits in the classical Fibonacci rabbit example. Then we have

$$\begin{cases} a_{n+1} = a_n + j_n = a_n + a_{n-1} \\ j_{n+1} = a_n \end{cases}$$

We have converted a two-step recurrence into a one-step recurrence; the price to pay is that we now need to know a pair of numbers, so we use a vector. In particular, let $\mathbf{r}_i$ be the vector of rabbits (a_i, b_i). Then we see that

$$\mathbf{r}_{n+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{r}_n = F\mathbf{r}_n$$

This method is generalizable; we can convert a k-step recursion into a one-step recursion by converting entries into a k-dimensional vector and using techniques of linear algebra.

The solution to this equation is

$$\mathbf{r}_n = F\mathbf{r}_{n-1} = F^2\mathbf{r}_{n-2} = \cdots = F^n\mathbf{r}_0 = F^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The problem just reduces to finding F^n. Let's look at the one case when this problem is easy: where F is a diagonal matrix.

Let $F = \operatorname{diag}(a_1, a_2, \ldots, a_k)$. Then $F^n = \operatorname{diag}(a_1^n, a_2^n, \ldots, a_k^n)$. However, F is not diagonal, so we need to look for another way. There is really another easy case, where $F = SDS^{-1}$ with D diagonal and S invertible. Then $F^n = SDS^{-1}SDS^{-1} \cdots SDS^{-1} = SD^nS^{-1}$. Turns out that we can find S and D!

Let $F = SDS^{-1}$, then $FS = SD$. If we let $S = \begin{bmatrix} \mathbf{s}_1, \ldots, \mathbf{s}_k \end{bmatrix}$ and $D = \operatorname{diag}(d_1, \ldots, d_k)$, then we have

$$F\mathbf{s}_i = d_i\mathbf{s}_i$$

Definition 25.1. For a square matrix A, the equation

$$A\mathbf{v} = \lambda\mathbf{v} \tag{1}$$

in the eigenvalue equation λ is an eigenvalue of A. If this has a solution with $\mathbf{v} \neq \mathbf{0}$, such a $\mathbf{v}$ is called a λ-eigenvalue for A.

Remark We have that λ is only called an eigenvalue of A if we can solve the equation (1) with some $\mathbf{v} = \mathbf{0}$. (Some texts call $\mathbf{v}$ an eigenvector even if $\mathbf{v} = \mathbf{0}$.)

Now how do we solve the eigenvalue problem? Rewrite (1) as $(A - \lambda I)\mathbf{v} = \mathbf{0}$. We want a nontrivial solution. This happens if and only if $A - \lambda I$ is not invertible.

Definition 25.2. The characteristic polynomial $\chi_A(\lambda)$ of the $k \times k$ matrix A is the determinant of the matrix $\det(\lambda I_k - A)$. The solutions of the characteristic polynomial are exactly the eigenvalues of A. Note that there are at most k eigenvalues of A since $\chi_A(\lambda)$ is of degree n.

Now we return to the Fibonacci problem. We have that $F = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. Thus $\chi_F(\lambda) = \det \begin{bmatrix} t-1 & -1 \\ -1 & t \end{bmatrix} = t^2 - t - 1$. Thus the roots are $\lambda_{\pm} = (1 + \sqrt{5})/2$. We have that $\lambda_+ > 0, \lambda_- < 0$, and $\lambda_+\lambda_- = -1$. Now we have to find the eigenvectors, which are just the elements of $\mathsf{N}(F - \lambda_{\pm} I)$. Thus $\mathbf{v}_+ = \begin{bmatrix} \lambda_+ \\ 1 \end{bmatrix}$, $\mathbf{v}_- = \begin{bmatrix} \lambda_- \\ 1 \end{bmatrix}$ (or multiples thereof), so $S = \begin{bmatrix} \lambda_+ & \lambda_- \\ 1 & 1 \end{bmatrix}$ and $S^{-1} = \begin{bmatrix} 1 & -\lambda_- \\ -1 & \lambda_+ \end{bmatrix} / \sqrt{5}$. Then

$$\begin{aligned}
F^n &= S \begin{bmatrix} \lambda_+^n & 0 \\ 0 & \lambda_-^n \end{bmatrix} S^{-1} \\
&= \begin{bmatrix} \lambda_+ & \lambda_- \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_+^n & 0 \\ 0 & \lambda_-^n \end{bmatrix} \begin{bmatrix} 1 & -\lambda_- \\ -1 & \lambda_+ \end{bmatrix} / \sqrt{5} \\
&= \begin{bmatrix} \lambda_+^{n+1} & \lambda_-^{n+1} \\ \lambda_+^n & \lambda_-^n \end{bmatrix} \begin{bmatrix} 1 & -\lambda_- \\ -1 & \lambda_+ \end{bmatrix} / \sqrt{5} \\
&= \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_+^{n+1} - \lambda_-^{n+1} & \lambda_+^n - \lambda_-^n \\ \lambda_+^n - \lambda_-^n & \lambda_+^{n-1} - \lambda_-^{n-1} \end{bmatrix}
\end{aligned}$$

Thus $\mathbf{r}_n = F^n \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{5}} \left(\lambda_+^{n+1} - \lambda_-^{n+1} \right)$.

There is another alternative method for doing this without computing SDS^{-1}.

1. Express $\mathbf{r}_0$ as a linear combination of the eigenvectors.
2. Apply F^n to this linear combination, setting up a system of equations which is solvable.

Proposition 25.1

Let $\lambda_1, \ldots, \lambda_k$ be pairwise distinct eigenvalues for an $n \times n$ matrix A. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be (nonzero) eigenvectors for them. Then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is a linearly independent set.

Theorem 25.1

If A has n distinct eigenvalues then it is diagonalizable. $A = SDS^{-1}$, where D is the diagonal matrix of eigenvalues and S is a matrix where the columns are the nonzero eigenvectors for the respective eigenvalues. In particular the ith column of S is the eigenvector corresponding to the eigenvalue D_{ii}.

26 October 1, 2018

Today we review the geometry of diagonalization. Hereafter let A be an $n \times n$ matrix.

Note that in general,

$$\chi_A(\lambda) = \lambda^n - \lambda^{n-1}\operatorname{Tr}(A) + \lambda^{n-2}? + \cdots + (-1)^n \det A$$

The intermediate terms are hard to compute. To find the constant term set $\lambda = 0$ and get $\det(-A)$.

Having found $\chi_A(\lambda)$ we have theoretically determined the eigenvalues which are the roots. They can be real or complex, and there are exactly n of them if we count repeated roots. In particular

$$\chi_A(\lambda) = \prod_{i=1}^{n}(\lambda - \lambda_i)$$

where $\lambda_1, \ldots, \lambda_n$ are the roots; they may or may not be repeated.

Remark For a generic A, we get n distinct roots. In fact, $\chi_A(\lambda)$ will have multiple roots if and only if it has a root in common with its derivative. (Exercise: prove this.)

Theorem 26.1

If the n roots of $\chi_A(\lambda)$ are distinct, then we get an eigenbasis (a basis consisting of eigenvectors) for F^n by choosing a nonzero eigenvector for each eigenvalue.

This relies on the following proposition:

Proposition 26.1

Eigenvectors for distinct eigenvalues are linearly independent.

Proof Let $\lambda_1, \ldots, \lambda_k$ be eigenvalues such that $\lambda_i \neq \lambda_j$ if $i \neq j$. Let $\mathbf{v}_i$ be a nonzero eigenvector for λ_i. Assume that

$$\sum_{i=1}^{k} c_i \mathbf{v}_i = \mathbf{0}$$

we want to show that c_i is necessarily 0 for all i, satisfying linear independence.

Assume for the sake of contradiction that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent. Let p be the least index such that $\mathbf{v}_{p+1}$ is a linear combination of the preceding (linearly independent) vectors. Then there exist scalars $c_1, \ldots, c_p$ such that

$$\sum_{i=1}^{p} c_i \mathbf{v}_i = \mathbf{v}_{p+1}$$

Multiplying both sides by A and using $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for all $\mathbf{v}_i$, we get

$$\sum_{i=1}^{p} c_i A\mathbf{v}_i = \sum_{i=1}^{p} c_i \lambda_i \mathbf{v}_i = A\mathbf{v}_{p+1} = \lambda_{p+1}\mathbf{v}_{p+1}$$

Multiplying both sides of the first equation by λ_{p+1} and subtracting from the last equation, we have

$$\sum_{i=1}^{p} (\lambda_i - \lambda_{p+1})\mathbf{v}_i = \mathbf{0}$$

Because of linear independence, the weights are all zero, but none of the factors are zero, because of distinctness. Hence $c_i = 0$ for $i = 1, \ldots, p$. But then we have that $\mathbf{v}_{p+1} = \mathbf{0}$ because of the first equation, which is impossible. Hence a contradiction is reached, so $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent. $\square$

We also have the following result.

Remark In the case of n distinct eigenvalues, we can conclude that each eigenspace is 1-dimensional.

Proof We use an algebraic argument. Write $A = SDS^{-1}$ where $S = \begin{bmatrix} \mathbf{s}_1 & \ldots & \mathbf{s}_n \end{bmatrix}$ and $D = \operatorname{diag} \lambda_1, \ldots, \lambda_n$. Then if $\mathbf{v} \in \mathsf{N}(A - \lambda_i)$,

$$\begin{aligned}
\mathbf{0} &= (A - \lambda_i I_n)\mathbf{v} \\
&= (SDS^{-1} - \lambda_i I_n)\mathbf{v} \\
&= S(D - \lambda_i I_n)S^{-1}\mathbf{v} \\
&= (D - \lambda_i I_n)(S^{-1}\mathbf{v})
\end{aligned}$$

Hence $\mathbf{v} \in \lambda_i$-eigenspace of A is equivalent to $S^{-1}\mathbf{v} \in \mathsf{N}(D - \lambda_i I_n) = t\mathbf{e}_i$ for $t \in \mathbb{R}$, where $D - \lambda_i I_n = \operatorname{diag}(\lambda_1 - \lambda_i, \ldots, 0, \ldots, \lambda_n - \lambda_i)$. Hence $\mathbf{v}$ is on the line $St\mathbf{e}_i$ for $t \in \mathbb{R}$. $\square$

The geometric reasoning of $A = SDS^{-1}$ is found in the language of linear transformations.

Let $\mathscr{S}$ be an isomorphism between R^n and a vector space V given by $\mathbf{x} \mapsto S\mathbf{x}$, $\mathscr{T}$ be a linear transformation from V to V given by $\mathbf{x} \mapsto A\mathbf{x}$, and $\mathscr{D}$ be a linear transformation between R^n and R^n given by $\mathbf{x} \mapsto D\mathbf{x}$. Note that they retain their original meanings given by diagonalization. Now we map $\mathbf{y} \mapsto \mathscr{S}\mathbf{y} \mapsto \mathscr{T} \circ \mathscr{S}\mathbf{y} \mapsto \mathscr{S}^{-1} \circ \mathscr{T} \circ \mathscr{S}\mathbf{y}$. Then we compute $\mathscr{T}$ as

$$\mathbf{y} \mapsto S^{-1}AS\mathbf{y} = D\mathbf{y}$$

This is a possible way to do anisotropic scaling in "eigencoordinates":

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} \lambda_1 x_1 \\ \vdots \\ \lambda_n x_n \end{bmatrix}$$

Now we cover the case of repeated eigenvalues. The algebraic multiplicity of an eigenvalue is its multiplicity as a root of χ_A. The geometric multiplicity of an eigenvalue is the dimension of its eigenspace. Earlier we have shown that an algebraic multiplicity of 1 implies a geometric multiplicity of 1. Indeed,

Fact 26.1

The algebraic multiplicity of an eigenvalue is greater than or equal to its geometric multiplicity.

Theorem 26.2
A matrix A is diagonalizable if and only if for each eigenvalue the algebraic multiplicity equals the geometric multiplicity.

Then, get an eigenbasis by choosing a basis in each eigenspace.

So what do we do if A is not diagonalizable? We write A as $S(D+N)S^{-1}$, where N is nilpotent.

27 October 2, 2018

We rehash the eigenvalue problem except in the language of linear transformations $\mathscr{T}$ instead of matrices A.

Observe that

1. $\dim \operatorname{domain} \mathscr{T} = \dim \operatorname{codomain} \mathscr{T}$
2. There are a maximum of n possible λ_i
3. If λ_i is repeated p times (with $p \leq n$), then $\dim \mathsf{N}(A - \lambda_i I) \leq p$ - that is, we can get anywhere from 1 to p corresponding eigenvectors

If we have a matrix A with n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ and each one has a corresponding eigenvector $\mathbf{v}_1, \ldots, \mathbf{v}_n$, then if we have

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

and $\mathbf{b} = \sum_{i=1}^{n} c_i \mathbf{v}_i$, then $\mathbf{x}_i = \sum_{i=1}^{n} c_i \mathbf{v}_i / \lambda_i$.

These problems are first order eigenvalue problems - that is, the power of each eigenvalue is 1. In more advanced physics there are second- and up order eigenvalue problems. These are just reducible to first-order problems using partitioned matrices.

	Solution Exists	Solution is Unique
Pivot in every row	x	
Pivot in every column		x
$\mathsf{S}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathsf{R}^n$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent		x
$A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$		x
$\mathsf{C}(A) = \mathsf{R}^n$	x	
$\mathsf{N}(A) = \{\mathbf{0}\}$		x
$\dim \mathsf{C}(A) = m$	x	
$\dim \mathsf{N}(A) = 0$	x	
$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a basis for R^m $(m = n)$	x	x
Range of $\mathscr{T}(\mathbf{x})$ is codomain	x	
Kernel of $\mathscr{T}(\mathbf{x}) = \{\mathbf{0}\}$		x
$\mathscr{T}(\mathbf{x})$ is onto	x	
$\mathscr{T}(\mathbf{x})$ is one-to-one		x
$\mathsf{C}\left(A^T\right) = \mathsf{R}^n$		x
$\mathsf{N}\left(A^T\right) = \{\mathbf{0}\}$	x	
A^{-1} exists $(m = n)$	x	x
A is row equivalent to I $(m = n)$	x	x
A^T is invertible $(m = n)$	x	x
$CA = AC = I$ where C exists	x	x
$\det A \neq 0$ (so $m = n$)	x	x
$\det A^T \neq 0$ (so $m = n$)	x	x
row vectors of A lin. indep.	x	
A is square and all eigenvectors are nonzero	x	x

To solve an arbitrary eigenvalue problem, pick a basis, construct a matrix for the transform, and solve from there.

28 October 3, 2018

We solve the real case of eigenvalue geometry.

Example 28.1

Find the geometry of the matrix

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

Solution We have $\chi_A(\lambda) = (\lambda - 3)(\lambda - 2)$ so $\sigma = \{2, -3\}$. The corresponding eigenvectors are $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Thus our change of basis matrix $S = \begin{bmatrix} 1 & 2 \\ -3 & -1 \end{bmatrix}$ and the inverse change of coordinates matrix $S^{-1} = \frac{1}{5}\begin{bmatrix} -1 & -2 \\ 3 & 1 \end{bmatrix}$. We have the transform $\begin{bmatrix} a \\ b \end{bmatrix} \mapsto \begin{bmatrix} 2^n a \\ (-3)^n b \end{bmatrix}$ upon n applications of the matrix A.

Drawing this out we obtain that the origin is an unstable/repelling fixed point i.e. any perturbation sends it to infinity; the graph looks like parabolas pointing both up and down. Thus they lie upon the curve $|y_1|^{\log 2} = c|y_2|^{\log 3}$ for any c. ☐

Example 28.2

Find the geometry of the matrix

$$B = \begin{bmatrix} 1 & 1/5 \\ 1/5 & 1 \end{bmatrix}$$

Solution We have $\chi_A(\lambda) = (5x-4)(5x-6)/25$. Thus our eigenvalues are $4/5$ and $6/5$, so $\sigma = \{4/5, 6/5\}$. The corresponding eigenvalues are $\mathbf{v}_1 = \begin{bmatrix} 1// -1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ so $S = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ and $S^{-1} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}/2$. We obtain the mapping $\begin{bmatrix} a \\ b \end{bmatrix} \mapsto \begin{bmatrix} (4/5)^n a \\ (6/5)^n b \end{bmatrix}$. Drawing this out we obtain that the origin is a semistable fixed point; any perturbation in the b-direction sends it to infinity, but any perturbation in the a-direction sends it back to zero; these curves are pushed to infinity. The curve lied upon is $|y_1|^{\log(6/5)} = c|y_2|^{\log(4/5)}$. ☐

Note 1 Some differently characterized real spectra have different behavior with regards to discrete evolutions. In particular

1. if $|\lambda_1|, |\lambda|_2 > 1$, we have a repelling fixed point but the curves are parabolic and point outwards
2. if $|\lambda 1| < 1$ but $|\lambda_2| > 1$, we have an repelling fixed point but the curves are hyperbolic and point outwards
3. if $|\lambda_1|, |\lambda_2| < 1$, we have an attracting/stable fixed point but the curves are parabolic and point inwards
4. if exactly one of $|\lambda_1|$ or $|\lambda_2|$ is 1, we have a repelling fixed point; the curves are straight and point outwards

Consider the matrix $E = \begin{bmatrix} p & -q \\ q & p \end{bmatrix}$, where p and q are real numbers. Define $r = \sqrt{p^2 + r^2}$, and write $E = r\begin{bmatrix} p/r & -q/r \\ q/r & p/r \end{bmatrix} = r\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. This matrix is a rotation by angle θ and a scaling operation by r.

There are three cases for the geometry of this matrix as it applies to discrete evolution:

1. $r > 1$: points on a spiral that goes out to infinity
2. $r = 1$: points stay on a circle
3. $r < 1$: points on a spiral that goes inwards to zero

With regards to movement, the central angular velocity is constant, and the distance to zero has exponential behavior.

The continuous version is easy to show by DeMoivre's theorem as $E_t = r^t \begin{bmatrix} \cos t\theta & -\sin t\theta \\ \sin t\theta & \cos t\theta \end{bmatrix}$.

Why so different from the real case? The matrix E usually has complex eigenvalues, specifically $p \pm \mathrm{i}q$. In particular $\chi_E(\lambda) = \lambda^2 - 2p\lambda + (p^2 + q^2)$. The corresponding eigenvalues are $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix}$. This is easily obtainable once we have one of the pair because of the following fact.

Fact 28.1

The complex eigenvalues of a real matrix come in complex conjugate pairs. If $\mathbf{v}$ is an eigenvector for λ, then $\overline{\mathbf{v}}$ is an eigenvector for $\overline{\lambda}$.

Note: we cannot represent $\mathbf{v}_1$ and $\mathbf{v}_2$ in R^2 because they have complex entries. However, we can represent $\operatorname{Re} \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\operatorname{Im} \mathbf{v}_1 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$;they are basis vectors of R^2 – $\mathbf{e}_1$ and $-\mathbf{e}_2$.

Theorem 28.1

If you use the basis consisting of $\operatorname{Re} \mathbf{v}_1, -\operatorname{Im}\ vbv_1$ for a 2×2 matrix with complex eigenvalues $p \pm \mathrm{i}q$, then in that basis the matrix becomes $\begin{bmatrix} p & -q \\ q & p \end{bmatrix}$.

If $R = \begin{bmatrix} \operatorname{Re} \mathbf{v}_1 & -\operatorname{Im} \mathbf{v}_1 \end{bmatrix}$, then $R^{-1}ER = \begin{bmatrix} p & -q \\ q & p \end{bmatrix}$.

29 October 4, 2018

Note that diagonalization takes a vector to R^n, does an operation (multiplying by matrix of eigenvalues), and then converts back.

If we have $\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y}$, then $y = \exp(At) = \sum_{j=0}^{\infty} A^j t^j/(j!)$. If A is diagonalizable, then $\exp(At) = P(\sum_{j=0}^{\infty} D^j t^j/(j!))P^{-1}$. Some more about this during differential equations.

Properties of symmetric matrices where $A = A^{\mathrm{T}}$:

1. The eigenvalues of A are real
2. A is diagonalizable
3. The matrix P of eigenvectors of A has the property that $P^{\mathrm{T}} = P^{-1}$.

30 October 5, 2018

For a 2×2 real matrix F of non-real eigenvalues $\lambda_{\pm} = p + \mathrm{i}q$, with $\mathbf{v}_+$ a λ_+-eigenvector, then in the basis $R = \begin{bmatrix} \operatorname{Re} \mathbf{v}_+ & -\operatorname{Im} \mathbf{v}_+ \end{bmatrix}$, F becomes $\begin{bmatrix} p & -q \\ q & p \end{bmatrix} = M$, then $F = RMR^{-1}$.

Remark You will NOT find real eigenvalues for a non-real eigenvalue of a real matrix.

Say a square $n \times n$ matrix A is

- diagonalizable over $\mathbb{R}$ if there exists a real eigenbasis of R^n.

- diagonalizable over $\mathbb{C}$ if there exists a real eigenbasis of C^n.

Let $\mathsf{F}^n = \mathsf{R}^n$ or C^n and V be a vector space with dimension n. Let $\mathscr{T}$ be a map from V to V, A be a matrix mapping F^n to F^n, let S be a change-of-basis matrix in F^n, and D be a matrix mapping F^n to F^n.

Then $A = SDS^{-1}$ and the columns of S are the basis of real eigenvectors. If $\mathsf{F} = \mathsf{C}$, then D has a diagonal block with real eigenvalues, and the rest of the matrix on the diagonal are 2×2 blocks of rotation matrices represented by complex eigenvalues $p_j + \mathrm{i}q_j$. Caution: the complex eigenvalues come in complex conjugate pairs. You must choose one eigenvalue from each pair.

How do we find S? The first k columns of S are the vectors from bases of all the real eigenspaces (corresponding to real diagonalization). For the complex eigenvalues, choose one from each complex conjugate pair. Choose a complex basis $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ of each eigenspace. The remaining columns of S are $\operatorname{Re}\mathbf{v}_1, -\operatorname{Im}\mathbf{v}_1, \dots, \operatorname{Re}\mathbf{v}_r, -\operatorname{Im}\mathbf{v}_r$.

We denote the inner product of R^n of two vectors $\mathbf{x}$ and $\mathbf{y}$ as

$$\langle \mathbf{x} | \mathbf{y} \rangle = \mathbf{x}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

We have

$$\langle \mathbf{x} | \mathbf{x} \rangle = \|\mathbf{x}\|^2 = \sum_{i=1}^{n} x_i^2$$

By Pythagoras we have $\|\mathbf{x}\|$ as the length of $\mathbf{x}$, so in general we define the length of $\mathbf{x} \in \mathsf{R}^n$ by $\sqrt{\langle \mathbf{x} | \mathbf{x} \rangle}$. The distance between two points $\mathbf{x}$ and $\mathbf{y}$ is $\|\mathbf{x} - \mathbf{y}\|$. The angle between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\alpha = \arccos\left(\frac{\langle \mathbf{x} | \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}\right)$$

Proposition 30.1 (Cauchy-Schwarz)
We have that, for two vectors $\mathbf{x}$ and $\mathbf{y}$,

$$|\langle \mathbf{x} | \mathbf{y} \rangle| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

with equality holding if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are called orthogonal if $\langle \mathbf{x} | \mathbf{y} \rangle = 0$; it is okay for $\mathbf{x}$ or $\mathbf{y}$ to be zero. (That is, the zero vector is orthogonal to every vector). A collection of vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ is called orthonormal if they are pairwise orthogonal (that is, $\langle \mathbf{u}_i | \mathbf{u}_j \rangle = 0$ whenever $i \neq j$) and normalized (that is, $\langle \mathbf{u}_i | \mathbf{u}_i \rangle = 1$ for all i). For the collection to be simply orthogonal, the first condition is all that is required.

Proposition 30.2
A collection of vectors is orthonormal if the matrix U formed by the vectors satisfies

$$U^{\mathrm{T}}U = I$$

and orthogonal if $U^{\mathrm{T}}U$ is diagonal (with square lengths in the diagonal).

31 October 8, 2018

One goal of orthogonality: find approximation methods.

For example: say $\mathsf{L} \subset \mathsf{R}^n$ is a linear subspace, and $\mathbf{x} \in \mathsf{R}^n$, not necessarily in L. Then the goal is to find the best approximation to $\mathbf{x}$ within L, that is, $\mathbf{y} \in \mathsf{L}$ with $\|\mathbf{x} - \mathbf{y}\|$ minimal.

From geometry we know that we can drop a perpendicular and figure out the length by a Pythagorean theorem bash. Note that $\mathbf{x} - \mathbf{y}$ is orthogonal to any vector in L. We need a good way to find $\mathbf{y}$ without any geometry because it gets really complicated in high dimensions. There is a very useful approximation called the "least squares approximation" which minimizes $\|\mathbf{x} - \mathbf{y}\|$.

Say that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is an orthogonal collection of nonzero vectors if $\mathbf{v}_i \perp \mathbf{v}_j$ whenever $i \neq j$. Then if the collection is orthogonal and $\mathbf{v}_i \neq 0$ for all i, then $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ is an orthonormal collection.

If $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{bmatrix}$ then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is orthogonal if and only if $V^{\mathrm{T}} V$ is diagonal and square, while the collection is orthonormal if and only if $V^{\mathrm{T}} V = I_k$.

Definition 31.1. The orthogonal complement $\mathsf{L}^\perp$ of a linear subspace $\mathsf{L} \in \mathsf{R}^n$ is the set of vectors orthogonal to every vector in L. In particular

$$\mathsf{L}^\perp = \{\mathbf{x} \mid \mathbf{x} \perp \boldsymbol{\ell} \;\forall\, \boldsymbol{\ell} \in \mathsf{L}\}$$

Note in particular that $\mathsf{L} = \{\mathbf{0}\}$ if and only if $\mathsf{L}^\perp = \mathsf{R}^n$.

Proposition 31.1

We have $\mathsf{L}^\perp$ is a linear subspace ofR^n such that $\dim \mathsf{L} + \dim \mathsf{L}^\perp = \dim \mathsf{R}^n = n$ and $\left(\mathsf{L}^\perp\right)^\perp = \mathsf{L}$.

Proof for this is long, tedious, and full of tautologies.

Theorem 31.1

Every vector $\mathbf{x}$ in R^n can be uniquely expressed as $\mathbf{p} + \mathbf{q}$ where $\mathbf{p} \in \mathsf{L}$ and $\mathbf{q} \in \mathsf{L}^\perp$. In particular $\mathbf{p}$ is called the orthogonal projection of $\mathbf{x}$ onto L.

The proof is to use dimensions to show that $\mathcal{B}_{\mathsf{L}} \cup \mathcal{B}_{\mathsf{L}^\perp} = \mathcal{B}_{\mathsf{R}^n}$.

Proposition 31.2

The map $\mathbf{x} \mapsto \mathbf{p}$ is a linear map; it must be given by a matrix. If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is an orthogonal basis of L, then $\mathbf{p}$ is given by

$$\mathbf{p} = \sum_{i=1}^{r} \frac{\langle \mathbf{x} | \mathbf{v}_i \rangle}{\langle \mathbf{v}_i | \mathbf{v}_i \rangle}$$

If $\mathbf{u}_1, \ldots, \mathbf{u}_r$ is an orthonormal basis, then $\mathbf{p} = U^{\mathrm{T}} U \mathbf{x}$, where

$$U = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_r \end{bmatrix} = V(V^{\mathrm{T}} V)^{-1} V^{\mathrm{T}} \mathbf{x}$$

hence UU^{T} is the matrix representation of the orthogonal projection onto L.

What if we don't have an orthogonal basis of L? Then The Gram-Schmidt process will create an orthogonal basis for you. There also exists a matrix formula in any basis $B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_r \end{bmatrix}$; this formula is

$$A = B\left(B^{\mathrm{T}}B\right)^{-1}B^{\mathrm{T}}$$

32 October 9, 2018

We cover the Jordan Canonical Form.

If A is diagonalizable, $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for R^n. The transform $\mathscr{T}(\mathbf{x}) = A\mathbf{x}$ becomes $\mathscr{T}([\mathbf{x}]_{\mathsf{E}}) = D[\mathbf{x}]_{\mathsf{E}}$, and $D = PDP^{-1}$. The condition for this is that there are n linearly independent eigenvectors. If we don't have this, there is a second best alternative, which is using triangular matrices.

Example 32.1

Let A have eigenvalue λ and eigenvector $\mathbf{v}_1$. Then

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

so

$$A\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} = \mathbf{v}_1 + \lambda\mathbf{v}_2$$

Hence $(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1$, so $(A - \lambda I)^2\mathbf{v}_2 = (A - \lambda I)\mathbf{v}_1 = \mathbf{0}$. We call $\mathbf{v}_2$ a second order eigenvector of A. In general, the nth corder eigenvectors are given by $(A - \lambda I)^n\mathbf{v}_n = \mathbf{0}$.

If A is $m \times n$, then the singular value decomposition breaks A down into

$$A = \underbrace{U}_{m\times m}\underbrace{\Sigma}_{m\times n}\underbrace{V^*}_{n\times n}$$

where the columns of U are a set of orthonormal eigenvectors of AA^*, the columns of V are a set of orthonormal eigenvectors of A^*A, and the elements of the diagonal of Σ are the singular values (square roots of the nonzero eigenvalues) of A.

There are two properties of eigenvalues:

1. $\det A = \det D$ or $\det J$, where D is the diagonal matrix of the eigenvalues of A, and J is the Jordan canonical form of A. We have $\det A = \prod_i \lambda_i$.
2. $\operatorname{tr} A = \sum_i \lambda_i$.

We define the inner product operation on $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathsf{V}$ as a binary operation that satisfies the following:

1. $\langle \mathbf{u}|\mathbf{v}\rangle = \langle \mathbf{v}|\mathbf{u}\rangle$
2. $\langle \mathbf{u} + \mathbf{v}|\mathbf{w}\rangle = \langle \mathbf{u}|\mathbf{w}\rangle + \langle \mathbf{v}|\mathbf{w}\rangle$
3. $\langle c\mathbf{u}|\mathbf{v}\rangle = c\langle \mathbf{u}|\mathbf{v}\rangle$
4. $\langle \mathbf{u}|\mathbf{u}\rangle \geq 0$ and $\langle \mathbf{u}|\mathbf{u}\rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

If W, $\mathsf{W}^\perp$ are subspaces of V, then $\forall \mathbf{x} \in \mathsf{W}$ and $\forall \mathbf{v} \in \mathsf{W}^\perp$, we have $\mathbf{x} \perp \mathsf{v}$. Note that

$$(\mathsf{R}(A))^\perp = \mathsf{N}(A), \quad (\mathsf{LN}(A))^\perp = \mathsf{C}(A)$$

The proof of this is to write the matrix multiplication as a dot product.

A mild discussion on orthogonality proofs follows.

33 October 10, 2018

Recall that if $\mathsf{L} \subseteq \mathsf{R}^n$ is a linear subspace: assume $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is an orthogonal basis, let $V = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_r \end{bmatrix}$, and $D = V^\mathrm{T} V$, then the projection $\hat{\mathbf{p}}$ onto L is

$$\hat{\mathbf{p}} = \sum_{i=1}^{r} \frac{\langle \mathbf{p} | \mathbf{v}_i \rangle}{\langle \mathbf{v}_i | \mathbf{v}_i \rangle} \mathbf{v}_i = \left(V D^{-1} V^\mathrm{T} \right) \mathbf{p} = P\mathbf{p}$$

Note that $\hat{\mathbf{p}} \in \mathsf{L}$, and $\mathbf{p} - \hat{\mathbf{p}}$ is perpendicular to L. Note that $\mathbf{p} - \hat{\mathbf{p}}$ is the projection of $\mathbf{p}$ onto $\mathsf{L}^\perp$, implying that (I_P) is the projection matrix onto $\mathsf{L}^\perp$. Note that P satisfies two properties, which can be checked algebraically from the definition of P: $P^2 = P$, and $P^\mathrm{T} = P$.

Proposition 33.1

A real matrix P satisfying $P = P^\mathrm{T}$ and $P = P^2$ is the orthogonal projection onto $\mathsf{C}(P)$.

Proof Given $\mathbf{x} \in \mathsf{R}^n$, let $\mathbf{p} = P\mathbf{x}$ and $\mathbf{q} = (I - P)\mathbf{x}$. If $\mathbf{p} \in \mathsf{C}(P)$ and $\mathbf{q} \perp \mathsf{C}(P)$, we are done. We also claim that the second property is equal to the first. Write $P^\mathrm{T}\mathbf{q} = \mathbf{0}$ (checking that $\mathbf{q} \perp \mathbf{c}$ for $\mathbf{c} \in \mathsf{C}(P)$). Then

$$P^\mathrm{T}(I - P)\mathbf{x} = P(I0P)\mathbf{x} = \left(P - P^2 \right) \mathbf{x} = \mathbf{0}$$

as claimed. □

Intuitively, the equation for orthogonal projection is shown by taking the components of a vector along many orthogonal lines. This projects the vector onto a hyperplane of order r.

Caution: with complex vectors, the inner product hides a complex conjugation:

$$\langle \mathbf{z} | \mathbf{w} \rangle = \sum_{i=1}^{r} \overline{z_i} w_i$$

(that is, $\langle \cdot | \cdot \rangle$ is a sesquilinear form).

THe Gram-Schmidt process produces an orthogonal basis $\mathbf{v}_1, \ldots, \mathbf{v}_r$ of a subspace L given a not-necessarily-orthogonal basis $\mathbf{a}_1, \ldots, \mathbf{a}_r$ of L such that $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_r \end{bmatrix}$ and $\mathsf{L} = \mathsf{C}(A)$. We can of course normalize this basis to an orthonormal basis. In general,

$$\mathbf{v}_j = \mathbf{a}_j - \sum_{i=1}^{j-1} \frac{\langle \mathbf{v}_i | \mathbf{a}_j \rangle}{\langle \mathbf{v}_i | \mathbf{v}_i \rangle} \mathbf{v}_i$$

Note that at any stage where we have processed k terms,

$$\mathsf{S}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathsf{S}\{\mathbf{a}_1, \ldots, \mathbf{v}_k\}$$

This computation can be captured by a matrix identity expressing the $\mathbf{a}_i$ in terms of the $\mathbf{v}_i$:

$$A = VR$$

where $A = \begin{bmatrix}\mathbf{a}_1 & \ldots & \mathbf{a}_r\end{bmatrix}$, $V = \begin{bmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_r\end{bmatrix}$, and R is the upper triangular matrix with entries $\langle \mathbf{v}_i | \mathbf{a}_j \rangle / \langle \mathbf{v}_i | \mathbf{v}_i \rangle$.

We can go on to normalize the $\mathbf{v}_i$ to $\mathbf{u}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$ to get $A = UR'$, where R' is the upper triangular matrix with $\|\mathbf{v}_i\|$ as the entry in the index (i, i). If the $\mathbf{v}_i$ is a basis of R^n, we get the QR factorization, where Q is the square matrix with orthonormal columns, and R is the matrix R' above.

A square matrix Q is orthogonal if it has orthonormal columns, that is, if $Q^{\mathrm{T}} Q = I \leftrightarrow Q^{\mathrm{T}} = Q^{-1}$.

A square matrix A is orthogonal if and only if the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$ is a conformal isometry (distance-and-angle-preserving).

34 October 11, 2018

We cover the Least Squares solution as a modification of Gram-Schmidt where the resulting set has to be orthonormal instead of orthogonal. We can convert the process into a matrix and get the classical solution.

35 October 12, 2018

Recall the approximation property of projection: if $\mathbf{x} \in \mathsf{R}^n$, and $\mathsf{L} \subseteq \mathsf{R}^n$ is a linear subspace, and $\mathbf{p} = \operatorname{proj}_{\mathsf{L}} \mathbf{x}$, then $\|\boldsymbol{\ell} - \mathbf{x}\|$ as $\boldsymbol{\ell} \in \mathsf{L}$ is strictly minimized at $\boldsymbol{\ell} = \mathbf{p}$. This is easy to prove:

$$\begin{aligned}
\|\mathbf{x} - \boldsymbol{\ell}\|^2 &= \|\mathbf{q} - \mathrm{d}\mathbf{p}\|^2 = \langle \mathbf{q} - \mathrm{d}\mathbf{p} | \mathbf{q} - \mathrm{d}\mathbf{p} \rangle \\
&= \|\mathbf{q}\| - 2 \langle \mathrm{d}\mathbf{p} | \mathbf{q} \rangle + \|\mathrm{d}\mathbf{p}\|^2 \\
&= \|\mathbf{q}\|^2 + \|\mathrm{d}\mathbf{p}\|^2
\end{aligned}$$

This error quantity is minimized at $\|\mathbf{q}\|^2$; as we move away from $\mathbf{p}$, the error increases by $\|\mathrm{d}\mathbf{p}\|^2$.

The Least Squares problem is to find the best approximate solution to the system $A\mathbf{x} = \mathbf{b}$ in the sense of minimizing the error $\|A\mathbf{x} - \mathbf{b}\|^2$.

Remark Any actual solution to the system answers this. The question is only interesting for an inconsistent system.

Example 35.1

We suspect a linear relation between a quantity y and some other quantities $x_1, \ldots, x_n$:

$$\sum_{i=1}^{n} k_i x_i = y$$

We use past data points for the x's and y and set up a linear system in the k_i. We have

$$\begin{cases} x_{11}k_1 + x_{12}k_2 + \cdots + x_{1n}k_n = y_1 \\ \vdots \\ x_{p1}k_1 + x_{p2}k_2 + \cdots + x_{pn}k_n = y_p \end{cases}$$

Statistics can prove that if the errors are i.i.d. and normally distributed, then the least squares solution gives the expectation for k_i – the most likely value.

To do this, we can solve the system $A\mathbf{x} = \mathbf{p}$, where $\mathbf{p} = \text{proj}_{\mathsf{C}(A)} \mathbf{b}$. One way from here is to use Gram-Schmidt on a basis of $\mathsf{C}(A)$ – numerically, this is very good, but overall, there is a shortcut. Rewrite $\mathbf{b} - \mathbf{p}$ as $\mathbf{b} - A\mathbf{x}$, and state that it is orthogonal to $\mathsf{C}(A)$ - and so orthogonal to every column of A, so their dot products are 0. In matrix form, this says that $A^{\mathrm{T}}(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$, so

$$A^{\mathrm{T}}A\mathbf{x} = A^{\mathrm{T}}\mathbf{b}$$

(which is called the normal equation.)

We know that this system is consistent. Is this solution unique? Clearly not, if $\mathsf{N}(A) \neq \{\mathbf{0}\}$.

Proposition 35.1

If A has linearly independent columns, then $A^{\mathrm{T}}A$ is invertible.

If $A^{\mathrm{T}}A$ is invertible, then the unique solution to the least squares problem is

$$\mathbf{x} = \left(A^{\mathrm{T}}A\right)^{-1} A^{\mathrm{T}}\mathbf{b}$$

If A has linearly independent columns, then the orthogonal projection matrix onto $\mathsf{C}(A)$ is $A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$.

Remark If A is an orthogonal matrix, then the projection matrix simplifies to AA^{T}. If $A = QR$, where Q is an orthogonal matrix, and R is upper triangular, then $A^{\mathrm{T}}A = R^{\mathrm{T}}R$; this is less computationally intensive and more numerically simple.

To solve a least squares problem (say, curve fitting), simply find an appropriate linear function to fit, write the inconsistent system, write the normal equations, and solve for your coefficients.

36 October 15, 2018

Today, we discuss the notion of abstract inner product spaces. An example of when this is useful is a weighted least squares problem. Say we have an appropriate data set and are trying to fit a linear regression. If the measurements are not equally reliable, we need to weight the data points differently. The correct weighting for uncorrelated errors is to use σ^{-2} at each data pont when computing the error function. This involves redefining the inner product to

$$\left\|\begin{bmatrix} v_1 & \vdots & v_n \end{bmatrix}\right\|^2 = \sum_{i=1}^{n} \sigma_i^2 v_i^2$$

Equivalently, we scale the basis vectors so $\mathbf{e}_i \to \sigma \mathbf{e}_i$. This process leads to the weighted normal equations.

Previously, the normal equation was

$$A^{\mathrm{T}} A\mathbf{x} = A^{\mathrm{T}}\mathbf{b}$$

Now the weighting matrix comes in. Let $W = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Then

$$A^{\mathrm{T}} W A = A^{\mathrm{T}} W\mathbf{b}$$

Remark If the errors are correlated, then the matrix W is the inverse of the covariance matrix Σ. Hence we have

$$A^{\mathrm{T}}\Sigma^{-1}A = A^{\mathrm{T}}\Sigma^{-1}A$$

Now we can fully define the inner product.

Definition 36.1. An inner product on a vector space V is a bilinear function $\mathsf{V} \times \mathsf{V} \to \mathbb{R}$ denoted $\langle \mathbf{u}|\mathbf{v}\rangle$, with the properties:

- Symmetry: $\langle \mathbf{u}|\mathbf{v}\rangle = \langle \mathbf{v}|\mathbf{u}\rangle$
- Bilinearity: $\langle \mathbf{u_1} + \mathbf{u_1}|\mathbf{v}\rangle = \langle \mathbf{u}_1|\mathbf{v}\rangle + \langle \mathbf{u}_2|\mathbf{v}\rangle$ and $\langle k\mathbf{u}|\mathbf{v}\rangle = k\,\langle \mathbf{u}|\mathbf{v}\rangle$ (same for $\mathbf{v}$ for fixed $\mathbf{u}$
- Positivity: $\langle \mathbf{u}|\mathbf{u}\rangle = \|\mathbf{u}\|^2 \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$.

Example 36.1

Let C be a symmetric $n \times n$ matrix (that is, $C = C^{\mathrm{T}}$). We say C is positive definite if for every $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^{\mathrm{T}} C\mathbf{x} > 0$ (equivalently, a symmetric matrix which has only positive eigenvalues). If so, then $\langle \mathbf{x}|\mathbf{y}\rangle_C = \mathbf{x}^{\mathrm{T}} C\mathbf{y}$ is an inner product in R^n. The proof of this is easy.

Proposition 36.1

If V is an n-dimensional inner product space, then it has an orthonormal basis of n elements, $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Proof Take any basis of V and use the Gram-Schmidt process to turn it orthonormal. $\square$

Some fun facts about the space of periodic continuous functions. Consider the functions:

$$1, \quad \cos x, \quad \cos 2x, \quad \cos 3x, \quad \ldots$$

$$\sin x, \quad \sin 2x, \quad \sin 3x, \quad \ldots$$

Any two distinct functions, one from each set, are orthogonal under the L^2 inner product; that is,

$$\int_{-\pi}^{\pi} \sin(nx)\cos(mx)\,\mathrm{d}x = 0$$

We also have, for nonzero n,

$$\|\sin(nx)\|^2 = \|\cos(nx)\|^2 = \frac{1}{2}$$

so the set is not orthogonal.

For any f, the orthogonal projection onto $\text{Span}\{1, \cos x, \ldots, \cos nx, \sin x, \ldots, \sin nx\}$ is given by

$$\sum_{k=0}^{n} (a_k \cos kx + b_k \sin kx)$$

where

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(kx) f(x) \, \mathrm{d}x$$

and similar for b_k, where $k > 0$. This is the "Fourier inversion formula."

37 October 16, 2018

We review the least squares solution as the solution to the problem $A\mathbf{x} = \mathbf{b}$, when $\mathbf{b} \notin \mathsf{C}(A)$. We have that $A^{\mathrm{T}} A \hat{\mathbf{x}} = A^{\mathrm{T}} \mathbf{b}$. If $A = QR$, then $R\hat{\mathbf{x}} = Q^{\mathrm{T}} \mathbf{b}$. This is guaranteed to have a solution, and if $\hat{\mathbf{x}}$ is unique, then $\det A = 0$; hence the columns of A are linearly independent.

We cover Laguerre, Lagrange, and Chebyshev polynomials.

38 October 17, 2018

Today we are covering two topics: the spectral theorem for real symmetric matrices, and Fourier series.

The spectral theorems combine diagonalization with orthogonality. In the real case, the question answered by the spectral theorem is: when is a real square matrix A diagonalizable in a real orthonormal basis?

Recall that diagonalization involves writing $A = SDS^{-1}$, where D is diagonal, and the columns of S are the eigenvectors of A. The columns of S form an eigenbasis for the diagonalization. The question reduces to asking: when can we choose an orthogonal S (that is, S such that the columns of S are orthonormal)?

If we can do that, then $S^{\mathrm{T}} = S^{-1}$, so $A = SDS^{\mathrm{T}}$ and $A^{\mathrm{T}} = SD^{\mathrm{T}}S^{\mathrm{T}} = SDS^{\mathrm{T}} = A$, so A has to be symmetric. Thus $A = A^{\mathrm{T}}$. The spectral theorem is the converse of that; that is, if A is symmetric.

Theorem 38.1

If A is a real symmetric matrix, then A is diagonalizable, and we can find a real orthonormal eigenbasis (that is, $A = SDS^{-1}$). In particular, all eigenvalues are real.

If we expand the range of A, we can generate alternate spectral theorems which apply to more classes of matrices.

Some partial results hold:

1. Every eigenvalue of A is real.

Proof Let λ be an eigenvalue with eigenvector $\mathbf{v}$. For the sake of contradiction λ is complex. Write $A\mathbf{v} = \lambda\mathbf{v}$, so

$$\overline{\mathbf{v}}^{\mathrm{T}} A\mathbf{v} = \overline{\mathbf{v}}^{\mathrm{T}} \lambda\mathbf{v} = \lambda\overline{\mathbf{v}}^{\mathrm{T}}\mathbf{v} = \overline{\lambda}\overline{\mathbf{v}}^{\mathrm{T}}\mathbf{v}$$

by additional algebra, so $\lambda = \overline{\lambda}$ and λ is real. ■

The result applies to complex matrices which satisfy $A = \overline{A}^{\mathrm{T}} = A^*$. If $A = A^*$, we have that A is an adjoint (Hermitian) matrix.

2. Eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ for different eigenvalues $\lambda_1 \neq \lambda_2$ are orthogonal.

   Proof Begin in the same way as before. We have $A\mathbf{v}_1 = \lambda\mathbf{v}_1$. Then

$$\mathbf{v}_2^{\mathrm{T}} A\mathbf{v}_1 = \mathbf{v}_2^{\mathrm{T}} \lambda_1\mathbf{v}_1 = \lambda_1 \langle \mathbf{v}_1 | \mathbf{v}_2 \rangle$$

   and

$$\mathbf{v}_2^{\mathrm{T}} A^{\mathrm{T}}\mathbf{v}_1 = (A\mathbf{v}_2)^{\mathrm{T}}\mathbf{v}_1 = \lambda 2_2 \mathbf{v}_2^{\mathrm{T}}\mathbf{v}_1 = \lambda_2 \langle \mathbf{v}_2 | \mathbf{v}_1 \rangle$$

   Since λ_1 and λ_2 are distinct, $\langle \mathbf{v}_1 | \mathbf{v}_2 \rangle = 0$ This proves the theorem when A has n distinct eigenvalues: they are real and every eigenbasis is orthogonal. We can rescale the columns to make them orthonormal, and at this point we are done. □

The general result (including repeated eigenvalues) uses the following lemma. Let A be an $n \times n$ symmetric matrix, with $A = A^{\mathrm{T}}$. Then let $\mathscr{T}_A \colon \mathsf{R}^n \to \mathsf{R}^n$ be a map from R^n to itself through $\mathbf{x} \mapsto A\mathbf{x}$, and $P \colon \mathsf{R}^n \to \mathsf{R}^n$ be a change-of-coordinates isomorphism. Then the columns of P form a new basis for the map $\mathbf{x} \mapsto A\mathbf{x}$. The whole chain can be represented by $\mathbf{y} = P^{-1}AP\mathbf{y}$. Let $B = P^{-1}AP$.

Lemma 38.1

If P is an orthogonal matrix (so the new basis $\mathsf{C}(P)$ is orthonormal), then B is also symmetric: $B^{\mathrm{T}} = P^{\mathrm{T}}A^{\mathrm{T}}(P^{-1})^{\mathrm{T}} = P^{-1}AP = B$.

Now we can prove the general spectral theorem.

Proof We know A has at least one real eigenvalue. Choose one, λ_1, and a unit length eigenvector $\mathbf{u}_1$. Continue $\mathbf{u}_1$ to an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ of R^n. Let $P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{bmatrix}$ as in the lemma. Because $A\mathbf{u}_1 = \lambda_1\mathbf{u}_1$, B is symmetric and looks like

$$B = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & C & \\ 0 & & & \end{bmatrix}$$

We know that C is an $(n-1) \times (n-1)$ matrix. So we can repeat such a process for C. □

Say we have an orthogonal basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ for a space L. Then the best approximation (orthogonal projection) of $\mathbf{p}$ onto L is

$$\hat{\mathbf{p}} = \sum_{i=1}^{k} \frac{\langle \mathbf{v}_i | \mathbf{p} \rangle}{\langle \mathbf{v}_i | \mathbf{v}_i \rangle} \mathbf{v}_i$$

The benefit of an orthogonal basis is that if $\mathsf{L} \subseteq \mathsf{V}$, where $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_m$ are additional orthogonal basis vectors of V, the orthogonal projection is just gained by adding the term

$$\sum_{i=k+1}^{m} \frac{\langle \mathbf{v}_i | \mathbf{p} \rangle}{\langle \mathbf{v}_i | \mathbf{v}_i \rangle} \mathbf{v}_i$$

to $\hat{\mathbf{p}}$, so

$$\hat{\mathbf{p}} == \sum_{i=1}^{m} \frac{\langle \mathbf{v}_i | \mathbf{p} \rangle}{\langle \mathbf{v}_i | \mathbf{v}_i \rangle} \mathbf{v}_i$$

as desired. Hence, the formula stabilizes. This is not the case with the projection formula in a general basis.

Example 38.1

Let F be the space of 2π-periodic real continuous functions that are subsets of $\mathsf{C}^0[-\pi, \pi]$. We have an orthogonal system of trigonometric functions $1, \cos x, \cos 2x, \ldots, \cos mx, \sin x, \sin 2x, \ldots, \sin nx$. The inner product is given by

$$\langle \mathbf{f} | \mathbf{g} \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{f}(x) \mathbf{g}(x) \, \mathrm{d}x$$

We note that $\|\sin nx\|^2 = \|\cos mx\|^2 = 1/2$ for all integers m, n, and $\|1\|^2 = 1$. Let $\mathsf{F}_n = \mathsf{S}\{1, \cos x, \ldots, \cos nx, \sin x, \ldots$ Then the projection formula for any $\mathbf{f}$ yields

$$\operatorname{proj}_{\mathsf{F}_n} \mathbf{f}(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

where

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{f}(x) \, \mathrm{d}x \,, \quad a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(kx) \mathbf{f}(x) \, \mathrm{d}x \,, \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin(kx) \mathbf{f}(x) \, \mathrm{d}x$$

We have

$$a_0^2 + \frac{1}{2} \sum_{k=1}^{n} \left(a_k^2 + b_k^2 \right) \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} (\mathbf{f}(x))^2 \, \mathrm{d}x$$

which is equal to $\|\mathbf{f}(x) - \operatorname{proj}_{\mathsf{F}_n} \mathbf{f}(x)\|^2$ by Pythagoras. As $n \to \infty$, we have that $\operatorname{proj}_{\mathsf{F}_n} \mathbf{f}(x)$ becomes a series. This series converges.

Theorem 38.2

This mean square limit is f itself; that is

$$\|\mathbf{f}(x) - \operatorname{proj}_{\mathsf{F}_n} \mathbf{f}(x)\|^2 \to 0 \quad \text{as} \quad n \to \infty$$

39 October 18, 2018

For a symmetric matrix A such that $A = A^{\mathrm{T}}$:

1. The eigenvectors of A are real.

Proof We use contradiction. Let $\lambda \in \mathbb{C}$ and $\mathbf{v} \in \mathsf{C}^n$. Then we have

$$A\mathbf{v} = \lambda\mathbf{v} \to \overline{\mathbf{v}}^{\mathrm{T}}(A\mathbf{v}) = \overline{\mathbf{v}}^{\mathrm{T}}\lambda\mathbf{v} = \lambda\overline{\mathbf{v}}^{\mathrm{T}}\mathbf{v}$$

by definition, and

$$\overline{(\lambda\overline{\mathbf{v}}^{\mathrm{T}}\mathbf{v})}^{\mathrm{T}} = \overline{(\overline{\mathbf{v}}^{\mathrm{T}}A\mathbf{v})} = \overline{\mathbf{v}}^{\mathrm{T}}A\mathbf{v}$$

so

$$\overline{\lambda}\overline{\mathbf{v}}^{\mathrm{T}}\mathbf{v} = \lambda\overline{\mathbf{v}}^{\mathrm{T}}\mathbf{v} \to \overline{\lambda} = \lambda \to \lambda \in \mathbb{R}$$

as claimed. ■

2. A is diagonalizable.

   Proof We use contradiction. Without loss of generality suppose repetition only occurs once (so one eigenvalue has 2 occurrences). Suppose that for each eigenvector, one eigenvector exists. Then

   $$(A - \lambda I)^2\mathbf{v} = \mathbf{0} \quad \text{and} \quad (A - \lambda I)\mathbf{v} \neq \mathbf{0}$$

   so

   $$\overline{\mathbf{v}}^{\mathrm{T}}(A - \lambda I)^2\mathbf{v} = \left[\overline{\mathbf{v}}^{\mathrm{T}}(A - \lambda I)\right]\left[(A - \lambda I)\mathbf{v}\right] = \mathbf{0}$$

   contradicting that $(A - \lambda I)\mathbf{v} \neq \mathbf{0}$ as claimed. ■

3. A has orthogonal eigenvectors.

   Proof Write $A = PDP^{-1}$, with $P^{\mathrm{T}} = P^{-1}$. Then $A^{\mathrm{T}} = A = \left(P^{-1}\right)^{\mathrm{T}} DP^{\mathrm{T}}$, so $PP^{\mathrm{T}} = I$, so the eigenvectors are orthogonal. □

For a symmetric matrix, if λ_k and $\overline{\lambda}_k$ are eigenvalues where $\lambda_k = a + b\mathrm{i}$, $\mathbf{v}_k$ and $\overline{\mathbf{v}}_k$ are eigenvectors. Then $P = \begin{bmatrix}\mathbf{v}_1 & \ldots & \operatorname{Re}\mathbf{v}_k & \operatorname{Im}\mathbf{v}_k & \ldots & \mathbf{v}_n\end{bmatrix}$, with $D = \operatorname{diag}(\lambda_1, \ldots, \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \ldots, \lambda_n)$.

Two matrices A and B are similar if they share the same eigenvalues and have the same number of eigenvectors.

If we have an eigenbasis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of a matrix A and we are trying to solve $A\mathbf{x} = \mathbf{b}$, then we can write $A = PDP^{-1}$, so $\mathbf{x} = PD^{-1}P^{-1}$, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Write $\mathbf{c} = P^{-1}\mathbf{b}$ Since

$$\mathbf{b} = \sum_{i=1}^{n} c_i\mathbf{v}_i \to \mathbf{x} = \sum_{i=1}^{n} \frac{c_i}{\lambda_i}\mathbf{v}_i$$

If we have an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ for A and we are trying to solve $A\mathbf{x} = \mathbf{b}$, write

$$\mathbf{b} = \sum_{i=1}^{n} \frac{\langle \mathbf{b} | \mathbf{w}_i \rangle}{\langle \mathbf{w}_i | \mathbf{w}_i \rangle}\mathbf{w}_i$$

if $\mathbf{u}_i = \mathbf{w}_i / \|\mathbf{w}_i\|$, then $U\mathbf{x} = \mathbf{b}$ where $U = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix}$, and so $\mathbf{x} = U^{\mathrm{T}}\mathbf{b}$.

We find an eigenbasis by taking the kernel of the mapping $\mathbf{x} \mapsto (A - \lambda I)\mathbf{x}$. Special cases are when A has n distinct eigenvectors or is symmetric.

We find an orthogonal basis, we use the Gram Schmidt process. Start with a basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, then:

$$\mathbf{w}_i = \mathbf{b}_i - \operatorname{proj}_{\mathsf{H}_{i-1}} \mathbf{b}_i$$

where $\mathsf{H}_{i-1} = \mathsf{S}\{\mathbf{w}_1, \ldots, \mathbf{w}_i - 1\}$.

40 October 19, 2018

Today we cover quadratic forms and spectral theorem variants.

Let $f(\mathbf{x}) \in \mathsf{C}^2[\mathsf{R}^n]$. Then if f has a local extremum at $\mathbf{c}$, $\boldsymbol{\nabla} f(\mathbf{c}) = \mathbf{0}$, but to know more, we need the second derivatives. The matrix of second derivatives $H_f = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]$ is symmetric.

In a neighborhood of $\mathbf{x} = \mathbf{c}$, we have

$$f(\mathbf{x}) = f(\mathbf{c}) + \frac{1}{2}(\mathbf{x} - \mathbf{c})^{\mathrm{T}} H_f (\mathbf{x} - \mathbf{c}) + \epsilon(\mathbf{x})$$

where $\epsilon(\mathbf{x})$ is an error term such that $\epsilon(\mathbf{x})/\|\mathbf{x} - \mathbf{c}\|^2 \to 0$ as $\mathbf{x} \to \mathbf{c}$.

The Hessian term generalizes the second degree term in the single variable Taylor polynomial.

A quadratic form on R^n is a homogeneous quadratic function $Q \colon \mathsf{R}^n \to \mathsf{R}$ is a homogeneous quadratic function

$$Q(\mathbf{x}) = \sum_{i=1}^{n} \left(Q_{ii} x_i^2\right) + \sum_{\substack{i \neq j \\ 1 \leq i,j \leq n}} \left(Q_{ij} x_i x_j\right)$$

where Q_{ij} is a symmetric matrix uniquely defined by the function

$$Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}} M_Q \mathbf{x}$$

There exists an orthonormal basis of R^n in which Q takes the form

$$\mathbf{x} \mapsto \sum_{i=1}^{n} \lambda_i x_i^2$$

where λ_i is the i-th eigenvalues of M_Q. The basis is an orthonormal basis of eigenvectors.

There exists an orthogonal basis where Q has the form

$$\mathbf{x} \mapsto \left(\sum_{i=1}^{p} x_i^2\right) + \mathcal{O}\left(\sum_{i=p+1}^{q} x_i^2\right) - \left(\sum_{i=q+1}^{n} x_i^2\right)$$

We know that the number p and $(n - q)$ are invariants with respect to the coordinates; they are properties of the original quadratic form.

We now look at change of coordinates in a quadratic form:

$$\mathbf{x} \mapsto Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}} M_Q \mathbf{x}$$

in a new basis $P = \left[\mathbf{p}_1, \ldots, \mathbf{p}_n\right]$. The function taking $\mathsf{R}^n \xrightarrow{P} \mathsf{R}^n \xrightarrow{Q} \mathsf{R}$ describes the mapping $\mathbf{y} \to \mathbf{x} = P\mathbf{y} \to Q(P\mathbf{y})$. If P is orthogonal, this becomes $P^{-1} M_Q P$, which can be made for an orthonormal choice of an eigenbasis P.

A quadratic form Q is called positive definite if all its eigenvalues are positive (so $Q(\mathbf{x}) \geq 0$ with equality occurring if $\mathbf{x} = \mathbf{0}$). It is negative definite if all its eigenvalues are negative (so $Q(\mathbf{x}) \leq 0$ with equality occurring if $\mathbf{x} = \mathbf{0}$). It is indefinite if it has both positive and negative eigenvalues. We also have positive semidefinite if all its eigenvalues are nonnegative, and negative semidefinite if all its eigenvalues are nonpositive.

From this we can see that f has a strict local minimum if $H_f(\mathbf{c})$ is positive definite; a strict local maximum if $H_f(\mathbf{c})$ is negative definite; neither a local maximum nor minimum if $H_f(\mathbf{c})$ is indefinite. If $H_f(\mathbf{c})$ is semidefinite, the test is inconclusive.

Now we cover various spectral theorems. Take a real orthogonal matrix $A^{\mathrm{T}} = A^{-1}$. Then the matrix is complex-diagonalizable in a complex orthogonal eigenbasis. The eigenvalues have modulus 1 and are of the form $\cos\theta + \mathrm{i}\sin\theta$, with θ the angle of rotation. If $\mathbf{v}$ is a complex eigenvector, then $\operatorname{Re}\mathbf{v}$ and $\operatorname{Im}\mathbf{v}$ are orthogonal and have the same length. We can find an orthonormal basis of R^n consisting of real eigenvectors with eigenvalues ± 1, and pairs of vectors $\operatorname{Re}\mathbf{v}, \operatorname{Im}\mathbf{v}$ on which A acts as a rotation.

41 October 22, 2018

We go over some problems that are useful for picking the midterm. One problem is to find a matrix of a linear transformation with the preferred values on a basis of R^3. We set up those basis vectors as columns of a matrix S, and their images under the map as columns of a matrix B; then our solution is given by $AS = B$ so $A = BS^{-1}$.

We can deal with limits by diagonalization.

To find the orthogonal complement of the column space of a matrix, find the left nullspace. In addition, if we are trying to find it for an orthogonal matrix, we can exploit the fact that all the rows and columns are normalized.

42 October 26, 2018

Today we cover the singular value decomposition.

Let A be a matrix with rows as one run each of observations on the same set of quantities and columns as runs of observations of a given quantity. The goal is to make sense of the data contained in A.

Does diagonalization help? No, diagonalization only makes sense for matrices that map a space to itself.

Statistics tells us to look for a correlation. Assume the sum of entries in each column is 0; else, subtract the column average from each column of A.

Then $A^{\mathrm{T}}A$ is an $n \times n$ symmetric matrix; in particular, $A^{\mathrm{T}}A/m$ is the covariance matrix, where each entry demonstrates the correlation between the two observations. We know that $A^{\mathrm{T}}A$ is square, symmetric, and positive semidefinite (that is, $\mathbf{x}^{\mathrm{T}}A^{\mathrm{T}}A\mathbf{x} \geq 0$ for all $\mathbf{x}$). Then A has an orthonormal basis of eigenvectors:

$$V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}, \quad \lambda_i \geq 0 \ \forall\ i$$

The quantities corresponding to the eigencoordinates are uncorrelated (statistically independent) with variances $\sigma_i^2 = \lambda_i$.

Now we can diagonalize $A^{\mathrm{T}}A$:

$$A^{\mathrm{T}}A = VDV^{\mathrm{T}}$$

Here $\lambda_i = \sigma_i^2$, with $\sigma_i \geq 0$. We have that each σ_i is a singular value of A. Let $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_m)$. Then

$$A^{\mathrm{T}}A = V\Sigma\Sigma^{\mathrm{T}}V^{\mathrm{T}}, \quad \left(V^{\mathrm{T}}A^{\mathrm{T}}\right)(AV) = D$$

which shows that the columns of AV form a orthogonal collection with lengths σ_i.

We can construct an orthonormal basis of R^m from the orthogonal set by rescaling the columns by their lengths, forming the orthonormal basis $U = \begin{bmatrix}\mathbf{u}_1, \dots, \mathbf{u}_m\end{bmatrix}$ of R^m. Thus the singular value decomposition is given by

$$A = U\Sigma V^{\mathrm{T}}$$

where $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_{\operatorname{rank} A}, 0, \dots, 0)$, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\operatorname{rank} A} \geq 0$. If A is $m \times n$, then U is $m \times m$, Σ is $m \times n$, and V^{T} is $n \times n$, with U and V orthonormal matrices.

In this configuration $\mathbf{u}_i$ are the left singular values and $\mathbf{v}_i$ are the right singular values. The $\mathbf{v}_i$ are unique up to sign for n distinct singular values, but there are ambiguities for repeated singular values including zeros. The $\mathbf{u}_i$ for nonzero singular values are unique up to sign, the rest are an orthonormal basis of $\mathsf{S}\{\mathbf{u}_1, \dots, \mathbf{u}_{\operatorname{rank} A}\}^{\perp}$. We can compress the singular value decomposition to

$$A = U_r \Sigma_r V_r^{\mathrm{T}}$$

where $r = \operatorname{rank} A$, U_r is $m \times r$, $\Sigma_r = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$ is $r \times r$, and V_r^{T} is $r \times n$. We know that U_r and V_r have orthonormal columns:

$$U_r^{\mathrm{T}} U_r = I_r = V_r^{\mathrm{T}} V_r$$

but they are not square in general.

Theorem 42.1

Let $r = \operatorname{rank} A$, and write $A = U\Sigma V^{\mathrm{T}}$. The vectors $\mathbf{u}_1, \dots, \mathbf{u}_r$ are an orthonormal basis of $\mathsf{C}(A)$. The vectors $\mathbf{u}_{r+1}^{\mathrm{T}}, \dots, \mathbf{u}_m^{\mathrm{T}}$ are an orthonormal basis of $\mathsf{LN}(A)$. The vectors $\mathbf{v}_1^{\mathrm{T}}, \dots, \mathbf{v}_r^{\mathrm{T}}$ are an orthonormal basis for $\mathsf{R}(A)$. The vectors $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ are an orthonormal basis for $\mathsf{N}(A)$.

43 October 29, 2018

Today we start talking about differential equations. We begin with the study of ordinary differential equations, or functions of one variable. Partial differential equations, on the other hand, are functions of multiple variables. The basic problem is something we've encountered before, but it's a continuous version. In particular, we concern ourselves with the time evolution of a system.

Before, we had seen a linear system

$$\mathbf{x}_n = A_n \mathbf{x}_{n-1}$$

A linear system had

$$\mathbf{x}_n - \mathbf{x}_{n-1} = (A - I)\mathbf{x}_{n-1}$$

If A is "close" to I, then the evolution is slow. This is the same as writing the difference quotient

$$\frac{\Delta \mathbf{x}}{\Delta t} = (A - I)\mathbf{x}_n$$

The continuous version is thus

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x}$$

where A is a square matrix. This is a linear homogeneous ordinary differential equation with constant coefficients.

The source of this type of problem is a physics system. Take, for example, a spring-mass system, with mass m, spring constant k. Set the position where the mass is at equilibrium to be the origin. Hooke's law says that

$$F(t) = -kx(t)$$

while Newton says that

$$F(t) = m\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2}$$

Thus, we have

$$\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} + \frac{k}{m}x(t) = 0$$

We can solve this. Let $\mathbf{x}(t) = \begin{bmatrix} x(t) \\ \mathrm{d}x/\mathrm{d}t \end{bmatrix}$; then

$$\frac{\mathrm{d}\mathbf{x}(\mathbf{t})}{\mathrm{d}t} = \begin{bmatrix} \mathrm{d}x(t)/\mathrm{d}t \\ \mathrm{d}^2x(t)/\mathrm{d}t^2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k/m & 0 \end{bmatrix}\begin{bmatrix} x(t) \\ \mathrm{d}x(t)/\mathrm{d}t \end{bmatrix}$$

In this way we can turn an arbitrary n-th order differential equation into a first order differential vector equation.

We have another example, that of an RLC circuit. Let $V(t)$ be a voltage source, C be the capacitance, L be the inductance, and R the electrical resistance. If $Q(t)$ is the charge on the capacitor and $I(t)$ is the current, then we have

$$L\frac{\mathrm{d}^2Q(t)}{\mathrm{d}t^2} + R\frac{\mathrm{d}Q(t)}{\mathrm{d}t} + CQ(t) = V(t)$$

This is an inhomogeneous second order linear ordinary differential equation with constant coefficients in $Q(t)$. Turning this into a matrix differentail equation yields

$$\mathbf{Q}(t) = \begin{bmatrix} Q(t) \\ \mathrm{d}Q(t)/\mathrm{d}t \end{bmatrix}, \quad \frac{\mathrm{d}Q}{\mathrm{d}t} = \begin{bmatrix} \mathrm{d}Q(t)/\mathrm{d}t \\ \mathrm{d}^2Q(t)/\mathrm{d}t^2 \end{bmatrix} = \begin{bmatrix} \mathrm{d}Q(t)/\mathrm{d}t \\ -(\mathrm{d}Q/\mathrm{d}t)R/L - Q(t)C/L + V(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -C/L & -R/L \end{bmatrix}\mathbf{Q}(t) + \begin{bmatrix} 0 \\ V(t) \end{bmatrix}$$

Hence, this is of the form $\mathbf{x} = A\mathbf{x} + \mathbf{b}$.

Our task is to solve ordinary differential equations when we can.

Theorem 43.1 (Cauchy, Picard)

Let $F(t; x_0, x_1, \ldots, x_n)$ be a continuously differential function of $(t, x_0, \ldots, x_n) \in (a, b) \times (\text{box in the } x_i)$. Let $\mathbf{p}$ be a specific value for $(x_0, \ldots, x_n)$. Then F gives rise to a differntial equation

$$F\left(t; \mathbf{p}(t), \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}, \frac{\mathrm{d}^2\mathbf{p}}{\mathrm{d}t^2}, \ldots, \frac{\mathrm{d}^n\mathbf{p}}{\mathrm{d}t^n}\right)$$

so that

$$x(t) \mapsto x_0 = x(t), x_1 = \frac{\mathrm{d}x}{\mathrm{d}t}, \ldots, x_n = \frac{\mathrm{d}^nx}{\mathrm{d}t^n}$$

The theorem says there is an interval $(t_0 - \epsilon, t_0 + \epsilon)$ where the equation has a unique solution $x(t)$ with initial data $\mathbf{p}$ at t_0:

$$\mathbf{p} = \begin{bmatrix} x(t_0) \\ \mathrm{d}x(t_0)/\mathrm{d}t \\ \vdots \\ \mathrm{d}^n x(t_0)/\mathrm{d}t^n \end{bmatrix}$$

We now tackle the way to solve second order linear differential equations with constant coefficients. First we examine the homogeneous case.

$$a\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + b\frac{\mathrm{d}x}{\mathrm{d}t} + cx(t) = 0$$

where a, b, c are constants with a nonzero. Consider the auxiliary equation

$$a\lambda^2 + b\lambda + c = 0$$

This is a quadratic equation with three phases.

1. There are two distinct real roots. Then the general solution has the form

   $$A\mathrm{e}^{\lambda_1 t} + B\mathrm{e}^{\lambda_2 t}$$

   with A and B arbitrary constants.

   Remark The solution forms a 2 dimensional vector space. Note that the linearity of the equation implies that the solution set forms a vector space.

   Remark A and B are uniquely determined by the value of the function $x(t_0)$ and the derivative $\mathrm{d}x(t_0)/\mathrm{d}t$ at any particular time t_0. If we write out the function and its derivatives, we get a 2×2 linear system which always has a unique solution in A and B. This is the case for an nth order differential equation by the same methodology.

2. We have a double root; that is, $\lambda_1 = \lambda_2$. Here we have a problem because both the terms of the two-λ case agree. So we find the solution that has the form

   $$x(t) = A\mathrm{e}^{\lambda_1 t} + Bt\mathrm{e}^{\lambda_1 t}$$

   where $\lambda_1 = \lambda_2$. Note that remark 2 still applies.

3. We have one or more complex λ_i. They have to be complex conjugates, so $\lambda_\pm = p \pm \mathrm{i}q$. The general solution has the form

   $$x(t) = A\mathrm{e}^{pt}\cos(qt) + B\mathrm{e}^{pt}\sin(qt)$$

   remark 2 still applies; we can always solve uniquely for A and B from $x(t_0)$ and $\mathrm{d}x(t_0)/\mathrm{d}t$.

44 October 30, 2018

In the singular value decomposition, we need to find the direction where $\|A\mathbf{x}\|$ is optimized. If $\|A\mathbf{x}\|$ is optimized, then $\|A\mathbf{x}\|^2 = \mathbf{x}^\mathrm{T} A^\mathrm{T} A\mathbf{x}$ is optimized. Along $\mathbf{v}_i$ we solve the eigenvalue problem:

$$\|A\mathbf{v}_i\|^2 = \lambda_i \mathbf{v}_i^\mathrm{T}\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i^\mathrm{T}\mathbf{v}_i$$

for $\sigma_i = \sqrt{\lambda_i}$, since $A^\mathrm{T} A$ is positive semidefinite.

Write

$$A = U\Sigma V^{\mathrm{T}}$$

where $V = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_n\end{bmatrix}$ and $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$ with blocks of empty rows and columns to make it $m \times n$, where r is the rank of A.

If $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthogonal basis for R^n, and A has r nonzero singular values, then $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthonormal basis for $\mathsf{C}(A)$. It is easy to show linear independence, and takes only a bit of algebraic manipulation to find that the span matches.

We can define our U by $\mathbf{u}_i A\mathbf{v}_i/\|A\mathbf{v}_i\| = A\mathbf{v}_i/\sigma_i$ for $i \in \{1, \dots, r\}$. For $i \in \{r+1, m\}$, just apply Gram-Schmidt.

We write a linear homonogeneous ordinary differential equation by

$$\sum_{k=0}^{n} a_0 \frac{\mathrm{d}^k y}{\mathrm{d}t^k} = 0$$

If we write

$$\mathbf{y} = \begin{bmatrix} y \\ \mathrm{d}y/\mathrm{d}t \\ \mathrm{d}^2 y/\mathrm{d}t^2 \\ \vdots \\ \mathrm{d}^{n-1} y/\mathrm{d}t^{n-1} \end{bmatrix}, \quad \mathbf{y} \in \mathsf{L}_\infty^n$$

, then

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \begin{bmatrix} \mathrm{d}y/\mathrm{d}t \\ \mathrm{d}^2 y/\mathrm{d}t^2 \\ \vdots \\ \mathrm{d}^n y/\mathrm{d}t^n \end{bmatrix} = \begin{bmatrix} \mathrm{d}y/\mathrm{d}t \\ \mathrm{d}^2 y/\mathrm{d}t^2 \\ \vdots \\ -(\sum_{k=0}^{n-1} a_k \frac{\mathrm{d}y}{\mathrm{d}t})/a_n \end{bmatrix}$$

We can write the problem as

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_0/a_n & -a_1/a_n & -a_2/a_n & \dots & -a_{n-1}/a_n \end{bmatrix} \mathbf{y}$$

or more generally as $\mathscr{D}(\mathbf{y}) = A\mathbf{y}$. If $\mathbf{y} = \exp(At)\mathbf{c}$, then $D(\mathbf{y}) = A\exp(At)\mathbf{c}$. If $A = PDP^{-1}$, then $\mathbf{y} = P\operatorname{diag}(\exp(\lambda_1 t), \dots, \exp(\lambda_n t))P^{-1}\mathbf{c}$. In particular we have

$$\left(\sum_{k=0}^{n} \mathscr{D}^k\right) y = 0$$

If we try to solve the eigenvalue problem, we have

$$\mathscr{D}(y) = \lambda I y \to (\mathscr{D} - \lambda I)y = 0 \to y = \mathrm{e}^{\lambda t}$$

where $\mathrm{e}^{\lambda t}$ is the eigenvector (eigenfunction).

We can find the characteristic polynomial of the eigenvalue problem:

$$\left(\sum_{k=0}^{n} a_k \lambda^k\right) \mathrm{e}^{\lambda t} = 0$$

Obviously, the exponential is never 0, so our $\chi_A(\lambda) = \sum_{k=0}^{n} a_k \lambda^k$.

Assume that λ_i are distinct. Then $\{\mathrm{e}^{\lambda_i t}\}$ is an eigenbasis for the kernel of the differential equation.

45 October 31, 2018

Last time we covered the case of second order ordinary differential equations with constant coefficients. The solutions always form a two-dimensional vector space with bases of eigenfunctions and coordinates A, B. We can pick out the unique solution by specifying the function value and the derivative value at a given time.

Note the similarity between the differential equation

$$ay'' + by' + c = 0$$

and the systems $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{x} = \mathbf{b}$. Think of A as a linear transformation $\mathscr{T}: \mathsf{R}^n \to \mathsf{R}^m$. The solution to the homogeneous equation forms the kernel $\operatorname{Ker} \mathscr{T}$, which is a linear subspace of R^n. The inhomogeneous solution has $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_h$ is a solution to the homogeneous system, and $\mathbf{x}_p$ is a particular solution to the system. Consider $\mathsf{V} = \mathsf{C}^{infty}(\mathbb{R})$ (smooth functions of a single variable). We thus have $y \in \mathsf{V}$, so $ay'' + by' + cy \in \mathsf{V}$ as well.

We also can consider it from an operator viewpoint. Consider the operator

$$\mathscr{D}_2 = a\frac{\mathrm{d}^2}{\mathrm{d}t^2} + b\frac{\mathrm{d}}{\mathrm{d}t} + c\mathrm{Id}$$

This sends y to the differential equation $ay'' + by' + cy$. We can thus write that the set of solutions to the homogeneous equations is $\operatorname{Ker} \mathscr{D}$, which is obviously a vector space. The solution of the inhomogeneous version would have the form $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_p$ is a single particular solution, and $\mathbf{x}_h$ is an element of $\operatorname{Ker} \mathscr{D}$.

We now seek to explain what happens as λ_2 approaches λ_1. We have that the two-dimensional eigenspace of solutions to the differential equation is given by

$$\lim_{\lambda_2 \to \lambda_1} \mathsf{E} = \lim_{\lambda_2 \to \lambda_1} \mathsf{S}\left\{\mathrm{e}^{\lambda_1 t}, \mathrm{e}^{\lambda_2 t}\right\} = \lim_{\lambda_2 \to \lambda_1} \mathsf{S}\left\{\mathrm{e}^{\lambda_1 t}, \frac{\mathrm{e}^{\lambda_2 t} - \mathrm{e}^{\lambda_1 t}}{\lambda_2 - \lambda_1}\right\} = \mathsf{S}\left\{\mathrm{e}^{\lambda_1 t}, t\mathrm{e}^{\lambda_1 t}\right\}$$

since the limit is exactly the definition of the derivative.

To explain the complex roots leading to trigonometric functions in the eigenspace, we simply utilize the complex exponential in context of the differential equation and expand. The first claim is that the two coordinates are complex conjugates.

We can thus combine the real and complex forms at the expense of using complex numbers.

To explain uniqueness of solutions, we first examine the one-dimensional case.

$$x'(t) + bx(t) = 0$$

The characteristic polynomial is

$$\chi(\lambda) = \lambda + b = 0$$

So the root of this equation is $\lambda = -b$. The only solution is thus $x(t) = A\mathrm{e}^{-bt}$.

Actually, this problem has another solution with a small hole. Write $x'(t) = -bx(t)$, so

$$\frac{\mathrm{d}}{\mathrm{d}t}(\log|x(t)|) = \frac{\mathrm{d}x/\mathrm{d}t}{x(t)} = -b \to \log|x(t)| = -bt + C \to x(t) = A\mathrm{e}^{-bt}$$

The problem occurs when $x(t_0) = 0$ for some t_0. The fix is the variation of parameters method.

Set $x(t) = A(t)\mathrm{e}^{-bt}$, so no assumptions have been made. We have

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{\mathrm{d}A}{\mathrm{d}t}\mathrm{e}^{-bt}0bA(t)\mathrm{e}^{-bt} = A'(t)\mathrm{e}^{-bt} - bx(t)$$

Hence

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} + bx(t) = A(t)\mathrm{e}^{-bt} \to \frac{\mathrm{d}A}{\mathrm{d}t} = 0$$

This allows us to solve the inhomogeneous equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} + bx(t) = g(t)$$

The same steps as above give

$$\frac{\mathrm{d}A}{\mathrm{d}t}\mathrm{e}^{-bt} = g(t) \to \frac{\mathrm{d}A}{\mathrm{d}t} = g(t)\mathrm{e}^{bt}$$

This gives the general solution

$$x(t) = \underbrace{A(0)\mathrm{e}^{-bt}}_{\text{gen. homo. soln.}} + \underbrace{\mathrm{e}^{-bt}\int_0^t g(s)\mathrm{e}^{bs}\,\mathrm{d}s}_{\text{particular inhomog. soln.}}$$

This works for any $g(t)$. A special case is when $g(t) = p(t)\mathrm{e}^{\lambda t}$ for $p(t)$ a polynomial of degree k, and λ is a real or complex number.

Question: what is

$$\mathrm{e}^{-bt}\int_0^t p(s)\mathrm{e}^{\lambda s + bs}\,\mathrm{d}s$$

If $\lambda \neq -b$, then the integrand has the form $q(t)\mathrm{e}^{(\lambda+b)t}$. The particular solution is thus of the form

$$x(t) = q(t)\mathrm{e}^{(\lambda+b)t}$$

with q a polynomial with the same degree as p. But if $\lambda = -b$, then the integrand is $q(t)$ for q a polynomial with degree one greater than the degree of p. We get

$$x(t) = q(t)\mathrm{e}^{-bt}$$

This hints that the general solution is to write the differential equation as linear equations in the parameters.

46 November 1, 2018

Recall that now we're dealing with problems of the form

$$\mathscr{D}[y] = \sum_{k=0}^{n} a_k \frac{\mathrm{d}^k y}{\mathrm{d}t^k} = f(t)$$

with the associated linear differential operator

$$\mathscr{D} = \sum_{k=0}^{n} a_k \frac{\mathrm{d}^k}{\mathrm{d}t^k}$$

If $\mathbf{y} = \left(y, y', \ldots, y^{(n-1)}\right)$, then

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \\ -a_0/a_n & -a_1/a_n & -a_2/a_n & \ldots & -a_{n-1}/a_n \end{bmatrix} \mathbf{y} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ f(t) \end{bmatrix}$$

From linear algebra we have $y = y_h + y_p$, with $\mathscr{D}(y_h) = 0$ and $\mathscr{D}(y_p) = f(t)$.

Since $\frac{\mathrm{d}y}{\mathrm{d}t} - \lambda y = 0$ is solved by $y = \mathrm{e}^{\lambda t}$ (the eigenvector/eigenfunction), we have

$$\mathscr{D}\left(\mathrm{e}^{\lambda t}\right) = \mathrm{e}^{\lambda t} \sum_{k=0}^{n} a_0 \lambda^k = 0$$

If we have n distinct λ, the eigenbasis is

$$\left\{\mathrm{e}^{\lambda_1 t}, \mathrm{e}^{\lambda_2 t}, \ldots, \mathrm{e}^{\lambda_n t}\right\}$$

If we have repeated λ, we take those entries and put a term t^k in front of them to ensure linear independence:

$$\left\{\mathrm{e}^{\lambda_1 t}, \mathrm{e}^{\lambda_2 t}, \ldots, \mathrm{e}^{\lambda_{n-p} t}, t\mathrm{e}^{\lambda t}, t^2\mathrm{e}^{\lambda t}, t^{p-1}\mathrm{e}^{\lambda t}\right\}$$

We can construct this as an eigenvector problem of the pth type and use proof by induction to get our solution.

Now we introduce the method of undetermined coefficients. Our original transformation is $\mathscr{D}\colon \mathsf{L}_\infty \to \mathsf{L}_\infty$, then any function f that belongs to L_∞ gets mapped also to an element L_∞ and furthermore its image is also in all vector spaces that f is an element of.

For this method, we guess what the output vector space is, and write a generic linear combination of the basis vectors of this vector space. Then we plug it into the differential equation and match coefficients.

47 November 4, 2018

We discuss linear ordinary differential equations with constant coefficients. Take the second order equation

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + b\frac{\mathrm{d}x}{\mathrm{d}t} + cx = g(t)$$

We have already found solutions to this at length.

Consider the first order euqation

$$\frac{\mathrm{d}x}{\mathrm{d}t} + bx = g(t)$$

We can solve this completely with

$$x(t) = c\mathrm{e}^{-bt} + \mathrm{e}^{-bt}\int_0^t \mathrm{e}^{bs} g(s)\,\mathrm{d}s$$

This is really only a solution if we're able to compute the exact antiderivative. There is a special case where $g(t) = \mathrm{e}^{(\lambda + \mathrm{i}\omega)t}p(t)$ for a polynomial $p(t)$.

Theorem 47.1

We can find the particular solution for such $g(t)$ of the form

$$g(t) = q(t)\mathrm{e}^{\lambda t}(\cos \omega t + \mathrm{i} \sin \omega t)$$

where $\deg q = \deg p$, unless $\lambda = -b$, in which case $\deg q = \deg p + 1$.

The method of undetermined coefficients allows us to find g from p without integration. This method generalizes to higher order equations.

Reduction of this method can solve ordinary differential equations of any order (as long as they have constant coefficients).

Remark By superposition, we can solve

$$g(t) = \sum_{k=1}^{n} g_k(t) = \sum_{k=1}^{n} p_k(t)\mathrm{e}^{\lambda_k t}$$

If $x_k(t)$ solves $x_k'(t) + bx_k(t) = g_k(t)$ for all k, then $\sum_{k=1}^{n} x_i(t)$ solves $\sum_{k=1}^{n} g(t)$.

So what is the method of undetermined coefficients?

Let $p(t) = \sum_{k=0}^{n} p_k t^k$ and $q(t) = \sum_{k=0}^{n} q_k t^k$ if $\lambda \neq -b$ and $\sum_{k=0}^{n+1} q_k t^k$ if $\lambda = -b$. Plug in $x(t) = g(t)\mathrm{e}^{\lambda t}$ in the equation and solve for the q_k. In particular,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mathrm{e}^{\lambda t}\left(\lambda \sum_{k=0}^{n} q_k t^k + \sum_{k=1}^{n} k q_k t^{k-1}\right)$$

and

$$bx(t) = b\mathrm{e}^{\lambda t}\sum_{k=0}^{n} q_k t^k$$

Hence

$$\mathrm{e}^{\lambda t}p(t) = (b+\lambda)\mathrm{e}^{\lambda t}\sum_{k=0}^{n} q_k t^k$$

We get a huge linear system:

$$(b+\lambda)q_i + (n-i+1)q_{n+1} = p_i$$

for $0 \le i \le n$, and $(b+\lambda)q_{n+1} = p_{n+1} = 0$.

If $\lambda \neq -b$, there is a unique solution by substitution from the top. There is no solution if $b + \lambda = 0$ but $p_{k \neq 0}$.

The correct system becomes $(n - i + 1)q_{n-i+1} = p_{n-i}$ for $0 \leq i \leq n$. This system is consistent but q_0 is undetermined because $q_0 \mathrm{e}^{\lambda t} = q_0 \mathrm{e}^{-bt}$, and is a homogeenous solution, so we can add any homogeneous solution to it and still get the same correct system.

Theorem 47.2

Let

$$\mathscr{D} = \sum_{k=0}^{n} c_k \frac{\mathrm{d}^k}{\mathrm{d}t^k}$$

and $\mathscr{D}[x] = 0$ be an ordinary differential equation. Then the space of solutions is spanned by the following basis.

Factor the auxiliary polynomial

$$\chi(\lambda) = \sum_{k=0}^{n} c_0 \lambda^k = \prod_{k=0}^{r} (\lambda - \lambda_k)^{m_i}$$

for m_i the multiplicity of the term, such that $\sum_i m_i = n$, and have $\lambda_i \neq \lambda_j$ for $i \neq j$. The basis of homogeenous solutions is

$$\left\{ \mathrm{e}^{\lambda_1 t}, t\mathrm{e}^{\lambda_1 t}, \ldots, t^{m_1 - 1}\mathrm{e}^{\lambda_1 t}, \ldots, \mathrm{e}^{\lambda_i t}, t\mathrm{e}^{\lambda_i t}, \ldots, t^{m_i - 1}\mathrm{e}^{\lambda_i t}, \ldots \right\}$$

Theorem 47.3

A particular solution

$$q(t)\mathrm{e}^{\mu t}$$

for the inhomogeneous equation $\mathscr{D}[x] = p(t)\mathrm{e}^{\mu t}$, can be found with a polynomial q with $\deg q = \deg p + m_i$ if $\mu = \lambda_i$ for some i (and if μ is distinct from all the roots, $m_i = 0$).

48 November 5, 2018

Today we discuss the method of variation of parameters.

First, we discuss the reduction of order in a linear ordinary differential equation. We can find the complete solution of the constant coefficient case by reduction to a system of first order differential equations, or solve systems with non-constant coefficients.

For an equation

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + b(t)\frac{\mathrm{d}x}{\mathrm{d}t} + c(t)x(t) = 0 \quad \text{or} \quad g(t)$$

knowledge of one homogeneous solution allows the construction of a second solution (by reducing the order of the equation by 1). Knowledge of two independent homogeneous solutions allows solving the inhomogeneous equation.

The first order equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} + b(t)x(t) = g(t)$$

can be solved completely. The principle is that knowing one solution of a linear ordinary differential equation allows you to reduce the order of the equation by 1. We now approach this method.

We start with the equation

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + b\frac{\mathrm{d}x}{\mathrm{d}t} + cx(t) = g(t)$$

The auxiliary equation is

$$\lambda^2 + b\lambda + c = (\lambda - \lambda_1)(\lambda - \lambda_2) = 0$$

We end up getting that the homogeneous solution of the differential equation is

$$f(t) = \beta \mathrm{e}^{\lambda_2 t} + \mathrm{e}^{\lambda_2 t}\int_0^t \mathrm{e}^{-\lambda_2 s}\left(\frac{\mathrm{d}f}{\mathrm{d}s} - \lambda_2 f(s)\right)\mathrm{d}s$$

This allows us to factor the problem of solving an nth order (inhomogeneous) ordinary differential equation with constant coefficients into a sequence of n first order equations. When $g(t)$ is a linear combination of exponentials and polynomials, we can solve by undetermined coefficients, knowing that the solution is of the form $g(t)\mathrm{e}^{\mu t}$.

Example 48.1

Solve the differential equation

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\frac{\mathrm{d}x}{\mathrm{d}t} + x = \mathrm{e}^{-t}$$

Solution The characteristic polynomial is

$$\lambda^2 + 2\lambda + 1 = (\lambda + 1)^2 = 0 \to \lambda = -1$$

We look for, as a particular solution

$$(pt^2 + qt + r)\mathrm{e}^{-t}$$

but since $t\mathrm{e}^{-t}$ and e^{-t} are homogeneous solutions, we only care about $pt^2\mathrm{e}^{-t}$. Write

$$x(t) = pt^2\mathrm{e}^{-t}$$

and

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -pt^2\mathrm{e}^{-t} - 2pt\mathrm{e}^{-t}$$

and

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = pt^2\mathrm{e}^{-t} - 2pt\mathrm{e}^{-t} - 2pt\mathrm{e}^{-t} + 2p\mathrm{e}^{-t}$$

Hence we get $2p\mathrm{e}^{-t} = \mathrm{e}^{-t}$. Hence $p = 1/2$ so our particular solution is $t^2\mathrm{e}^{-t}/2$. Our general solution is

$$x(t) = A\mathrm{e}^{-t} + Bt\mathrm{e}^{-t} + \frac{1}{2}t^2\mathrm{e}^{-t}$$

□

Example 48.2

Solve the differential equation

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + x(t) = t\sin t$$

Solution The characteristic polynomial is

$$\lambda^2 + 1 = (\lambda + \mathrm{i})(\lambda - \mathrm{i}) = 0$$

Hence $\lambda = \pm\mathrm{i}$. We look for $x(t)$ as a second order equation, but we don't need the constant term since there is no constant on the right hand side. We write

$$x(t) = (at + bt^2)\mathrm{e}^{it} + (ct + dt^2)\mathrm{e}^{-it}$$

Differentiating this and plugging it in yields our solution. □

49 November 6, 2018

Today we go over some more differential equations using undetermined coefficients and variation of parameters.

Note that when we're doing variation of parameters with the complex exponential, it's much easier to use sin and cos functions.

We now cover existence and uniqueness theorems. Let

$$\mathscr{D} = \sum_{k=0}^{n} a_k(t)\frac{\mathrm{d}^k}{\mathrm{d}t^k} \to \mathscr{D}[y] = \sum_{k=0}^{n} a_k(t)\frac{\mathrm{d}^k y}{\mathrm{d}t^k}$$

and concern ourselves with the differential equation $\mathscr{D}[y] = b(t)$. Dividing through by $a_n(t)$, we have

$$\mathscr{D}^* = \frac{\mathrm{d}^n}{\mathrm{d}t^n} + \sum_{k=0}^{n-1} p_k(t)\frac{\mathrm{d}^k}{\mathrm{d}t^k} \to \mathscr{D}^*[y] = \frac{\mathrm{d}^n y}{\mathrm{d}t^n} + \sum_{k=0}^{n-1} p_k(t)\frac{\mathrm{d}^k y}{\mathrm{d}t^k}$$

and concern ourselves with the differential equation $\mathscr{D}^*[y] = g(t)$. If for all i, $p_i(t) \in \mathsf{L}_1$, and $g(t) \in \mathsf{L}_1$, then there is a guaranteed unique solution to the initial value problem, and a guaranteed (possibly nonunique) solution to the boundary value problem.

In the special case that $p_i \in \mathsf{R}$ for all i, $g(t) = 0$, and $y = \sum_{k=0}^{n} c_i y_i$, the initial value problem at t_0 can be represented as a matrix

$$W(t_0)\mathbf{c} = \mathbf{k}$$

where $\mathbf{c} = (c_0, c_1, \ldots, c_n)$ and $\mathbf{k} = (k_0, k_1, \ldots, k_n)$, where $y^{(i)}(t_0) = k_i$.

The Wronskian is defined as

$$W = \begin{bmatrix} \mathbf{y}^{\mathrm{T}} \\ \left(\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t}\right)^{\mathrm{T}} \\ \vdots \\ \left(\frac{\mathrm{d}^{n-1}\mathbf{y}}{\mathrm{d}t^{n-1}}\right)^{\mathrm{T}} \end{bmatrix}$$

where $\mathbf{y}(t) = (y_1, y_2, \ldots, y_n)$.

Write

$$\mathbf{y} = \begin{bmatrix} y \\ y' \\ \vdots \\ y^{n-1} \end{bmatrix}$$

Then we can write $\mathbf{y}'$ as the same off-diagonal matrix times $\mathbf{y}$ summed with a vector whose last element is $f(t)$. We have $\mathbf{y}' = A\mathbf{y} + \mathbf{f}$; the solution to this is $\mathbf{y}_h = \mathbf{c}e^{At}$.

50 November 7, 2018

Today we concern ourselves with linear ordinary differential equations with variable coefficients. We start with first order equations, which we can solve completely. Then we can discuss reduction of order (specifically, we will cover second order ordinary differential equations) to the order one case. If we know one homogeneous solution, we can find the second homogeneous solution. Finally, we will use the Lagrange Formula to solve the inhomogeneous equation.

Start with the equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} + b(t)x(t) = g(t)$$

where $b(t)$ and $g(t)$ are $\mathsf{L}_\infty^{[a,b]}$. We solve the homogeneous equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} + b(t)x(t) = 0$$

Dividing, we have

$$\frac{x'(t)}{x(t)} = -b(t) \to \frac{\mathrm{d}}{\mathrm{d}t}(\log|x(t)|) = b(t)$$

Integrating, we have

$$\int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}s}(\log|x(s)|) = \int_{t_0}^{t} b(s)\,\mathrm{d}s = \log\left|\frac{x(t)}{x(t_0)}\right| \to x(t) = x(t_0)\exp\left(-\int_{t_0}^{t} b(s)\,\mathrm{d}s\right)$$

Thus, we have

$$x(t) = C\exp\left(-\int_{t_0}^{t} b(s)\,\mathrm{d}s\right)$$

what if $x(t_0) = 0$ for some t_0? Construct this function as the solution to the differential equation and verify that it works, avoiding the issue with dividing by zero.

Hence,

$$\frac{\mathrm{d}x}{\mathrm{d}t} + b(t)x(t) = \frac{\mathrm{d}C}{\mathrm{d}t} \underbrace{\exp\left(-\int_{t_0}^{t} b(s)\,\mathrm{d}s\right)}_{W(t)=\text{ homogeneous solution}}$$

Hence

$$\frac{\mathrm{d}C}{\mathrm{d}t} = g(t)\exp\left(\int_{t_0}^{t} b(s)\,\mathrm{d}s\right)$$

A particular solution is given by

$$C(t) = \int_{t_0}^{t} g(s) \exp\left(\int_{s_0}^{t} b(r)\,\mathrm{d}r\right) \mathrm{d}s$$

and

$$x(t) = W(t) \int_{t_0}^{t} g(s) W^{-1}(s)\,\mathrm{d}s$$

gives a particular inhomogeneous solution.

In particular the basic homogeneous solution is given by $W(t)$:

$$W(t) = \exp\left(-\int_{t_0}^{t} b(s)\,\mathrm{d}s\right)$$

Now we deal with second order equations. Let $\mathscr{D}$ be the differentail operator described by

$$\mathscr{D} = \frac{\mathrm{d}^2}{\mathrm{d}t^2} + b(t)\frac{\mathrm{d}}{\mathrm{d}t} + c(t)\mathrm{Id}$$

and let H be the differential equation $\mathscr{D}[y] = 0$; I be the differential equation $\mathscr{D}[y] = g(t)$. Let $u(t)$ be a homogeneous solution. We are looking for the kernel of $\mathscr{D}$.

We can construct $\mathscr{D}$ by applying the differentiation directly, or factorize it into

$$\mathscr{D} = \left(\left(\frac{\mathrm{d}}{\mathrm{d}t} + \left(b(t) + \frac{u'(t)}{u(t)}\right)\mathrm{Id}\right) \circ \left(\frac{\mathrm{d}}{\mathrm{d}t} - \frac{u'(t)}{u(t)}\mathrm{Id}\right)\right)$$

By checking the arithmetic, it is clear that this works. Note that $u'' = -bu' - cu$, and do the cancellation.

Note that $\mathscr{D} \colon \mathsf{L}_\infty^{[a,b]} \to \mathsf{L}_\infty^{[a,b]}$, and each of the factors is also $\mathsf{L}_\infty^{[a,b]} \to \mathsf{L}_\infty^{[a,b]}$.

Remark If $\mathscr{D}$ is a differential operator of order n and u was a solution to $\mathscr{D}[u] = 0$, then we can factor $\mathscr{D}$ into

$$\mathscr{D} = \mathscr{E} \circ \left(\frac{\mathrm{d}}{\mathrm{d}t} - \frac{u'(t)}{u(t)}\mathrm{Id}\right)$$

where $\mathscr{E}$ is a differential operator of order $n - 1$.

The conclusion is that solving the inhomogeneous equation $\mathscr{D}[y] = g(t)$ amounts to solving the two first order ordinary differential equations

$$\begin{cases} \phi'(t) + \left(b(t) + \frac{u'(t)}{u(t)}\right)\phi(t) = g(t) \\ x'(t) - \frac{u'(t)}{u(t)}x(t) = \phi(t) \end{cases}$$

Now we find the homogeneous solution. This is the same as the Wronskian $W(t)$:

$$\begin{aligned} \phi(t) &= \exp\left(-\int_{t_0}^{t} b(s)\,\mathrm{d}s - \int_{t_0}^{t} \frac{u'(s)}{u(s)}\,\mathrm{d}s\right) \\ &= |u(t)|^{-1} \exp\left(-\int_{t_0}^{t} b(s)\,\mathrm{d}s\right) = |u(t)|^{-1} W(t) \end{aligned}$$

We are free to choose a constant in front. We choose $\phi(t) = u(t)^{-1}W(t)$.

Now we solve for $x(t)$. The first homogeneous solution is given by

$$x(t) = \exp\left(\int_{t_0}^{t} \frac{u'(s)}{u(s)}\,\mathrm{d}s\right) = u(t)$$

The second homogeneous solution is

$$\begin{aligned} x(t) &= u(t)\int_{t_0}^{t} u(s)^{-1}u(s)^{-1}W(s)\,\mathrm{d}s \\ &= u(t)\int_{t_0}^{t} u(s)^{-2}\exp\left(\int_{s_0}^{s} -b(r)\,\mathrm{d}r\right)\mathrm{d}s \\ &= u(t)\int_{t_0}^{t} u(s)^{-2}W(s)\,\mathrm{d}s \end{aligned}$$

We can repeat the work with $g \neq 0$ to find an inhomogeneous solution.

Theorem 50.1

If $u(t)$ and $v(t)$ are two solutions to the differential equation $\mathscr{D}[y] = 0$, then a particular solution $x(t)$ to the differential equation $\mathscr{D}[y] = g(t)$ can be written as

$$x(t) = \alpha(t)u(t) + \beta(t)v(t)$$

where

$$\alpha(t) = -\int_{t_0}^{t} \frac{v(s)g(s)}{uv' - vu'}\,\mathrm{d}s\,, \quad \beta(t) = -\int_{t_0}^{t} \frac{u(s)g(s)}{uv' - vu'}$$

where α and β are canonical.

Remark There are three methods to get this formula: the way the book does it, following through the reduction of order; and matrix methods.

We now introduce the Wronskian as above. To reiterate: let $\mathbf{y}(t) = (y_1(t), \ldots, y_n(t))$ be a vector of n solutions to a homogeneous differential equation.

$$W[\mathbf{y}](t) = \begin{bmatrix} \mathbf{y}^{\mathrm{T}} \\ \left(\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t}\right)^{\mathrm{T}} \\ \vdots \\ \left(\frac{\mathrm{d}^{n-1}\mathbf{y}}{\mathrm{d}t^{n-1}}\right)^{\mathrm{T}} \end{bmatrix}$$

The only important quantity of this matrix is the determinant, so we henceforth refer to W as $\det W[\mathbf{y}](t)$.

Theorem 50.2

Either W is identically zero (so the homogeneous solutions are linearly dependent) or W is never 0 for any t. Moreover, in the two-dimensional case,

$$W = C\exp\left(-\int_{t_0}^{t} b(s)\,\mathrm{d}s\right)$$

which agrees with our previous results.

51 November 8, 2018

Today we conver the formal concept of eigenfunctions.

The equation

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda\mathrm{Id}\right)y = 0$$

represents the eigenvalue problem for $\frac{\mathrm{d}}{\mathrm{d}t}$ with λ as an eigenvalue and $\mathrm{e}^{\lambda t}$ as the corresponding eigenvector.

Consider the general linear ordinary differential equation.

$$\mathscr{D}(y) = \sum_{k=0}^{n} a_k \frac{\mathrm{d}^k y}{\mathrm{d}t^k} = \left(\sum_{k=0}^{n} a_k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\right) = a_n\left(\prod_{k=1}^{n}\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_k\mathrm{Id}\right)\right)y$$

We can multiply these terms, from right to left, to compose the differentiatial operator by factoring.

For example, say we have

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\mathrm{Id}\right)\left[\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\mathrm{Id}\right)y\right] = \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\mathrm{Id}\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\mathrm{Id}\right)y = \left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} - (\lambda_1+\lambda_2)\frac{\mathrm{d}}{\mathrm{d}t} + \lambda_1\lambda_2\mathrm{Id}\right)y = \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} - (\lambda_1+\lambda_2)\frac{\mathrm{d}y}{\mathrm{d}t} + \lambda$$

If the λ are distinct, any solution y_i to

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_i\mathrm{Id}\right)y_i = 0$$

that is, $y_i = \mathrm{e}^{\lambda_i t}$, solves $\mathscr{D}[y] = 0$. If λ_{n-p} is repeated p times and the other λ are distinct,

$$\mathscr{D}[y] = \left(\prod_{k=1}^{n-p-1}\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_k\mathrm{Id}\right)\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_{n-p}\mathrm{Id}\right)^p y$$

Then functions y_i that satisfy $\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_{n-p}\mathrm{Id}\right)^p y_i = 0$ – the pth order eigenvalue problem, solved by $y_i = t^{p-1}\mathrm{e}^{\lambda_{n-p}t}$ – must also be solutions to $\mathscr{D}[y] = 0$. As an example, take the differential equation

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + 2\frac{\mathrm{d}y}{\mathrm{d}t} + y = t\mathrm{e}^{-t}$$

The differential operator is defined by

$$\mathscr{D}[y] = \left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 2\frac{\mathrm{d}}{\mathrm{d}t} + \mathrm{Id}\right)y = \left(\frac{\mathrm{d}}{\mathrm{d}t} + \mathrm{Id}\right)^2 y$$

The homogeneous solutions give

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)y_h &= 0\\ \left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)\mathrm{e}^{-t} &= 0 \to \left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)^2\mathrm{e}^{-t} = 0\\ \left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)y_2 = \left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)t\mathrm{e}^{-t} &= \mathrm{e}^{-t}\\ \left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)^2 t\mathrm{e}^{-t} &= 0\\ y_h &= c_1\mathrm{e}^{-t}+c_2t\mathrm{e}^{-t}\end{aligned}$$

Now to the inhomogeneous problem. We have

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)^2\left[\left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)^2 y_p\right] = \left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)^2 t\mathrm{e}^{-t} \to \left(\frac{\mathrm{d}}{\mathrm{d}t}+\mathrm{Id}\right)^4 y_p = 0$$

Hence y_p is a 4th order eigenvector, so it is $y_p = At^3\mathrm{e}^{-t}$. Some algebra figures out that the coefficient is $1/6$, so $y_p = t^3\mathrm{e}^{-t}/6$.

We can deal with the simple harmonic oscillator system with two masses which have positions y_1, y_2 given by

$$M\ddot{y}+C\dot{y}+Ky = 0$$

where M, C, K are $2 \times L$ matrices. Indeed, it is easy to show

$$\mathbf{y} = \begin{bmatrix} y_1\\ y_2\\ \dot{y}_1\\ \dot{y}_2\end{bmatrix} \to \dot{\mathbf{y}} = \begin{bmatrix} \dot{y}_1\\ \dot{y}_2\\ \ddot{y}_1\\ \ddot{y}_2\end{bmatrix} = \begin{bmatrix} 0 & I\\ M^{-1}K & M^{-1}C\end{bmatrix}\mathbf{y}$$

by using block matrices.

In general $\mathbf{y}' = A\mathbf{y}$ is a very general ordinary differential equation, for $m \times n$ matrix A.

The first method of solution is the matrix exponential. Write

$$\mathbf{y} = \mathrm{e}^{At}\mathbf{c}, \quad \mathbf{c} \in \mathsf{R}^n \to \frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathrm{e}^{At}\mathbf{c} = A\mathbf{y}$$

From Taylor series, we know that

$$\mathrm{e}^{At} = \sum_{k=0}^{\infty}\frac{(At)^k}{k!}$$

If A is diagonalizable, $A = PDP^{-1}$, so $A^n = PD^nP^{-1}$. So

$$\begin{aligned}\mathrm{e}^{At} &= \sum_{k=0}^{\infty}\frac{PD^kP^{-1}t^k}{k!} = P\left(\sum_{k=0}^{\infty}\frac{D^kt^k}{k!}\right)P^{-1}\\ &= P\left(\operatorname{diag}\left(\sum_{k=0}^{\infty}\frac{(\lambda_1t)^k}{k!},\ldots,\sum_{k=0}^{\infty}\frac{(\lambda_nt)^k}{k!}\right)\right)P^{-1} = P\Big(\operatorname{diag}(\mathrm{e}^{\lambda_1t},\ldots,\mathrm{e}^{\lambda_nt})\Big)P^{-1}\\ &= \begin{bmatrix}\mathbf{v}_1\mathrm{e}^{\lambda_1t} & \ldots & \mathbf{v}_n\mathrm{e}^{\lambda_nt}\end{bmatrix}P^{-1}\end{aligned}$$

Hence

$$\mathbf{y} = \mathrm{e}^{At}\mathbf{c} = \begin{bmatrix}\mathbf{v}_1\mathrm{e}^{\lambda_1 t} & \dots & \mathbf{v}_n\mathrm{e}^{\lambda_n t}\end{bmatrix} P^{-1}\mathbf{c} = \begin{bmatrix}\mathbf{v}_1\mathrm{e}^{\lambda_1 t} & \dots & \mathbf{v}_n\mathrm{e}^{\lambda_n t}\end{bmatrix}\begin{bmatrix}k_1\\ \vdots \\ k_n\end{bmatrix} = \sum_{i=1}^{n} k_i\mathbf{v}_i\mathrm{e}^{\lambda_i t}$$

so if A is diagonalizable, then $\mathbf{y}' = A\mathbf{y}$.

The second method of solution is to write

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y} \to \left(I\frac{\mathrm{d}}{\mathrm{d}t} - A\right)\mathbf{y} = \mathbf{0} \to \lambda_1\mathbf{v}_1\mathrm{e}^{\lambda_1 t} - A\mathbf{v}_1\mathrm{e}^{\lambda_1 t} = \mathbf{0}$$

52 November 9, 2018

We consider first order differential equations of a vector valued function $\mathbf{x}(t)$ in R^n. The prototypical example for a homogeneous equation is

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A(t)\mathbf{x}(t)$$

and for an inhomogeneous equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A(t)\mathbf{x}(t) + \mathbf{b}(t)$$

where A is an $m \times n$ matrix, and $\mathbf{b}(t)$ is a vector in R^n. Assume everything is smooth.

There is a special case of constant coefficients; that is, when A is independent of t. On the other hand, $\mathbf{b}$ may depend on t. This is the only case which allows a complete solution of the system; in the general case, there is no explicit solution.

Theorem 52.1

Let $F(t;\mathbf{x})$ be a continuously differentiable R^n-valued function of t and $\mathbf{x}$ defined on $[a,b]\times$(box in R^n). Then for any $t_0 \in (a,b)$ (the "initial time") and $\mathbf{x}_0 \in$ box (the "intial condition") there exists, for $t \in (t_0 - \epsilon, t_0 + \epsilon)$ with ϵ arbitrarily small, a unique and continuously differentiable function $\mathbf{x}(t)$ with

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = F(t;\mathbf{x}(t)) \quad \text{and} \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

Note that in the linear case $F(t;\mathbf{x}(t)) = A(t)\mathbf{x}(t)$.

For the $n = 1$ case, the equation is separable. In the nonlinear case, it is difficult to impossible to predict the long term behavior of the solution.

We now present an example where the uniquness of solutions condition fails, even where F is smooth.

Example 52.1

Consider

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sqrt{x}$$

then

$$\frac{1}{\sqrt{x}}\frac{\mathrm{d}x}{\mathrm{d}t} = 1$$

then

$$2\frac{\mathrm{d}}{\mathrm{d}t}\left(\sqrt{x(t)}\right) = 1$$

hence

$$\sqrt{x(t)} = \frac{t}{2} + C$$

and so

$$x(t) = \left(\frac{t}{2} + C\right)^2$$

We missed the solution that $x(t)$ is identically zero, but more than that, there are infinitely many solution functions that have $x(0) = 0$.

We can construct approximate solutions geometrically. Consider the two-dimensional graph of t against $\mathbf{x}(t)$. Then draw a velocity curve that points along with $\mathbf{x}'(t)$ at any point. Then drawing flow lines on the velocity curve yields an integral curve.

An integral curve of a vector field $(t, \mathbf{x}) \mapsto (1, F(t; \mathbf{x}))$ is the graph of a solution of the corresponding ordinary differential equation $F(t; \mathbf{x})$.

Basically, we can utilize Euler's method, and to achieve progressively higher accuracy we allow the timesteps to go to 0. In the limit, we get an actual (local) solution as long as $F \in \mathsf{C}_1$.

Why are we only looking at first order equations? Indeed, vectors allow reduction to first order equations.

Example 52.2

Take the differential equation

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + b(t)\frac{\mathrm{d}x}{\mathrm{d}t} + c(t)x = 0$$

Let $\mathbf{x} = (x, x')$. Then

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \begin{bmatrix} x' \\ x'' \end{bmatrix} = \begin{bmatrix} x' \\ -bx' - cx \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -c & -b \end{bmatrix}\begin{bmatrix} x \\ x' \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -c(t) & b(t) \end{bmatrix}}_{A(t)}\mathbf{x}$$

The vector equation and scalar equations are equivalent:

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -cx - b\frac{\mathrm{d}}{\mathrm{d}x}$$

For a scalar linear ordinary differential equation of order n, let $\mathscr{D}$ be the differential operator defined by

$$\mathscr{D} = \sum_{k=0}^{n} a_k(t) \frac{\mathrm{d}^k}{\mathrm{d}t^k}$$

The differential equation $\mathscr{D}[x]$ gives a special kind of matrix equation (where $\mathbf{x} = (x, x', \ldots, x^{(n-1)})$.

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A(t)\mathbf{x}, \quad A(t) = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_0/a_n & -a_1/a_n & -a_2/a_n & \ldots & -a_{n-1}/a_n \end{bmatrix}$$

Proposition 52.1

Assume that the a_i are constant. Then the auxiliary polynomial of $\mathscr{D}$, that is

$$\chi_{\mathscr{D}}(\lambda) = \sum_{k=0}^{n} a_k \lambda^k$$

is the characteristic polynomial of the matrix $A(t)$ given above. A consequence of this is that the exponents appearing in the solution $\mathrm{e}^{\lambda_i t}$ of the ordinary differential equation $\mathscr{D}[x] = 0$ are the eigenvalues of $A(t)$.

A reminder that all of this work is for constant coefficients. This will lead to the eigenvector method for solving the constant coefficient equation.

53 November 13, 2018

Recall the matrix problem

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y}$$

which is solved by

$$\mathbf{y} = \mathrm{e}^{At}\mathbf{c}$$

If A is diagonalizable,

$$\mathbf{y} = \sum_{k=1}^{n} c_i \mathrm{e}^{\lambda_i t}\mathbf{v}_i$$

If A is not diagonalizable,

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

where J is the Jordan canonical matrix.

Consider the case that $A \in \mathsf{M}_{2\times 2}$. Then

$$A = PJP^{-1} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}^{-1}$$

so

$$(A - \lambda I)\mathbf{v}_1 = 0, \quad (A - \lambda I)\mathbf{v}_2 = [v]_1$$

We have

$$A^n = PJ^nP^{-1}, \quad J^n = \begin{bmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{bmatrix} \to \mathrm{e}^{At} = P\begin{bmatrix} \mathrm{e}^{\lambda t} & t\mathrm{e}^{\lambda t} \\ 0 & \mathrm{e}^{\lambda t} \end{bmatrix}P^{-1}$$

We have $\mathbf{y}' = A\mathbf{y}$ so

$$\mathbf{y} = c_1\mathrm{e}^{\lambda t}\mathbf{v}_1 + c_2\Big(t\mathrm{e}^{\lambda t}\mathbf{v}_1 + \mathrm{e}^{\lambda t}\mathbf{v}_2\Big)$$

Say we have the system

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y}$$

given by the equation

$$\sum_{k=0}^{n} a_k \frac{\mathrm{d}^k y}{\mathrm{d}t^k} = 0$$

with corresponding matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_0/a_n & -a_1/a_n & -a_2/a_n & \dots & -a_{n-1}/a_n \end{bmatrix}$$

If $\lambda = \lambda_1$, then

$$y = c_1\mathrm{e}^{\lambda_1 t}, \quad \frac{\mathrm{d}^k y}{\mathrm{d}t^k} = c_1\lambda^k\mathrm{e}^{\lambda_1 t}, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ \lambda_1 \\ \lambda_1^2 \\ \vdots \\ \lambda_1^{n-1} \end{bmatrix}$$

If the eigenvalues are complex, we multiply by the conjugate and turn the problem into a real problem which we know how to solve. Also, we know that complex eigenvalues come in pairs.

If we have

$$\begin{aligned} \mathbf{y} &= c(\mathbf{a} + \mathbf{b}\mathrm{i})\mathrm{e}^{(\alpha+\beta\mathrm{i})t} + \bar{c}(\mathbf{a} - \mathbf{b}\mathrm{i})\mathrm{e}^{(\alpha-\beta\mathrm{i})t} \\ &= c(\mathbf{a} + \mathbf{b}\mathrm{i})\Big(\mathrm{e}^{\alpha t}(\cos(\beta t) + \mathrm{i}\sin(\beta t))\Big) + \bar{c}(\mathbf{a} - \mathbf{b}\mathrm{i})\Big(\mathrm{e}^{\alpha t}(\cos(\beta t) - \mathrm{i}\sin(\beta t))\Big) \\ &= (c + \bar{c})\mathrm{e}^{\alpha t}(\mathbf{a}\cos(\beta t) - \mathbf{b}\sin(\beta t)) + (c - \bar{c})\mathrm{e}^{\alpha t}(\mathbf{a}\sin(\beta t) + \mathbf{b}\cos(\beta t)) \end{aligned}$$

Now we deal with the differential equations which are of the form

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y} + \underbrace{\mathbf{f}(\mathbf{t})}_{\text{forcing term}}$$

We use the method of undetermined coefficients. Let

$$\mathbf{f}(t) \in \Big\{t^n\mathrm{e}^{\alpha t}, t^n\cos(\beta t), t^n\sin(\beta t)\Big\}$$

for some values of α, β. If $\mathbf{f}(t) \in \{t^n \mathrm{e}^{\alpha t}\}$, then

$$\mathbf{y}_p = \left(\sum_{k=0}^{n} A_k t^k\right)\mathrm{e}^{\alpha t}$$

If $\mathbf{f}(t) \in \{t^n \cos(\beta t), t^n \sin(\beta t)\}$, we have

$$\mathbf{y}_p = \sum_{k=0}^{n} (A_k \cos(\beta t) + B_k \sin(\beta t))t^k$$

If $\mathbf{f}'(t) = A\mathbf{f}(t)$ and A is diagonalizable, then

$$\mathbf{f}(t) = c\mathrm{e}^{\lambda_1 t}\mathbf{v}_1$$

and

$$\mathbf{y}_p = c_1\Big(t\mathrm{e}^{\lambda_1 t}\mathbf{v}_1 + \mathrm{e}^{\lambda_1 t}\mathbf{v}_2\Big)$$

where $(A - \lambda_1 I)\mathbf{v}_2 = \mathbf{v}_1$. Note that for this case both λ and $\mathbf{v}$ have to match.

54 November 14, 2018

Today we discuss the eigenvector method for solving constant coefficient homogeneous matrix ordinary diffferential equations.

Before we start here, recall the formal aspects of matrix ordinary differential equations from the perspective of linear algebra. Consider the homogeneous equation

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = A(t)\mathbf{x}(t)$$

with $A(t)$ given.

Some facts:

1. The set of solutions is a vector space V; in particular it is a subspace of L_1 functions of t with values in R^n.
2. The quantity $\dim \mathsf{V} = n$, which can be shown using the existence-uniquness theorem. More precisely, for any time t_0, the map $\varphi\colon \mathsf{V} \to \mathsf{R}^n$ given by $\mathbf{x}(t) \mapsto \mathbf{x}(t_0)$ is a linear isomorphism, which is to say that it is a bijection.
3. For each time t_0, this "evalulation at t_0" map φ gives a coordinate system on the space V of solutions to the homogeneous equation. This depends on t_0.
4. Varying t, we get a change-of-coordinates matrix $X(t)$ whose action on a vector is given by a functor which is $\mathsf{R}^n \to \mathsf{R}^n$. The matrix $X(t)$ is invertible since it is a change-of-coordinates matrix relating the coordinates coming from the function evaluated at t_0 to the function evaluated at t. (Note that $X(t)$ depends on the choice of t_0).

This $X(t)$ has the properties

1. $X(t_0) = I_n$.
2. $X'(t) = A(t)X(t)$ for all t.

3. $X(t)$ is invertible for all t.

The second property is equivalent to asserting that each column of $X(t)$ solves the ordinary differential equation. More precisely, the kth column of X is the solution to the ordinary differential equation which specializes to $\mathbf{e}_k$ at $t = t_0$.

$$\begin{tikzcd}
& \mathsf{V} \arrow[dl, "\mathrm{eval}(t_0)"'] \arrow[dr, "\mathrm{eval}(t)"] & \\
\mathsf{R}^n \arrow[rr, "X(t)"] & & \mathsf{R}^n
\end{tikzcd}$$

Remark Some ideas as to the solution set:

1. Any matrix $X(t)$ satisfying the second and third properties is called a fundamental solution matrix of the ordinary differential equation.
2. Such a matrix is unique up to right multiplication by a constant invertible matrix. If $Y(t)$ satisfies
$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = A(t)Y(t)$$
and $Y(t)$ is invertible, then $Y(t) = X(t)C$ where C is a constant invertible matrix.
3. The determinant of such a matrix $Y(t)$ satisfies the first order ordinary differential equation
$$\frac{\mathrm{d}}{\mathrm{d}t}(\det(Y(t))) = \mathrm{Tr}(A(t))\det(Y(t))$$

Checking the second property relies on on the Leibniz rule for matrix products. If $B(t)$ and $D(t)$ are matrix-valued differentiable functions and $B(t)D(t)$ is defined, then

$$\frac{\mathrm{d}}{\mathrm{d}t}(B(t)D(t)) = \frac{\mathrm{d}B(t)}{\mathrm{d}t}D(t) + B(t)\frac{\mathrm{d}D(t)}{\mathrm{d}t}$$

This reduces to the usual Leibniz rule using the entry-by-entry formula for the product. Now we can check that $Y(t) = X(t)C$ for C a constant invertible matrix. Because $X(t)$ is inveritble, we can write for some $C(t)$

$$Y(t) = X(t)C(t)$$

and so

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = \frac{\mathrm{d}X(t)}{\mathrm{d}t}C(t) + X(t)\frac{\mathrm{d}C(t)}{\mathrm{d}t}$$

Since $Y'(t) = A(t)Y(t)$, and $X'(t) = A(t)X(t)$, and writing $X(t)C(t) = Y(t)$, we get

$$A(t)Y(t) = A(t)X(t)C(t) + X(t)\frac{\mathrm{d}C(t)}{\mathrm{d}t} = A(t)Y(t) + X(t)\frac{\mathrm{d}C(t)}{\mathrm{d}t}$$

Hence

$$X(t)\frac{\mathrm{d}C(t)}{\mathrm{d}t} = 0$$

So, if X is invertible, then $C'(t) = 0$, so C is constant.

Remark Knowing a fundamental solution matrix $Y(t)$ allows you to solve the ordinary differential equation for any initial data $\mathbf{x}(t_0)$. If we had the preferred $X(t)$ with $X(t_0) = I_n$, write

$$\mathbf{x}(t_0) = \sum_{k=1}^{n} x_k(t_0)\mathbf{e}_k$$

Then $X(t)\mathbf{x}(t_0)$ is a solution to the ordinary differential equation. Note that it is a linear combination of the columns of X so it has to be a solution. The initial value given by $I_n\mathbf{x}(t_0)$ is our solution.

To get from $Y(t)$ to a solution which is I_n at t_0, just write

$$X(t) = Y(t)Y(t_0)^{-1}$$

The desired solution is

$$\underbrace{Y(t)Y(t_0)^{-1}}_{=I_n \text{ at } t_0}\mathbf{x}(t_0)$$

Now we tackle constant coefficient ordinary differential equations.

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x}(t) \tag{*}$$

This equation is easy to solve when A is diagonal; that is, when $A = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Break this up into

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \lambda_i x_i \quad \text{solution:} \quad x_i(t) = x_i(0)\mathrm{e}^{\lambda_i t}$$

with total solution

$$\mathbf{x} = \sum_{k=1}^{n} x_k(t)\mathbf{e}_k\mathrm{e}^{\lambda_k t}$$

Assume A is diagonalizable with eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. We can express any solution

$$\mathbf{x}(t) = \sum_{k=1}^{n} c_k(t)\mathbf{v}_k$$

Then

$$A\mathbf{x}(t) = \sum_{k=1}^{n} c_k(t)\lambda_k\mathbf{v}_k$$

and

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \sum_{k=1}^{n} \frac{\mathrm{d}c_k(t)}{\mathrm{d}t}\mathbf{v}_k$$

Hence $(*)$ is equivalent to

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = \lambda_i c_i(t) \leftrightarrow c_i(t) = c_i(0)\mathrm{e}^{\lambda_i t}$$

Hence the complete solution is given by

$$\mathbf{x}(t) = \sum_{k=1}^{n} c_k(0)\mathrm{e}^{\lambda_k t}\mathbf{v}_k$$

From the initial value $\mathbf{x}(0)$, work out the coefficients $c_i(0)$ in the eigenbasis.

$$\mathbf{c}(0) = S^{-1}\mathbf{x}(0)$$

where $S = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n\end{bmatrix}$.

55 November 15, 2018

We now deal with variation of parameters, starting with second order equations. We have the differentail operator

$$\mathscr{D} = \frac{\mathrm{d}^2}{\mathrm{d}t^2} + a_1\frac{\mathrm{d}}{\mathrm{d}t} + a_0\mathrm{Id}$$

Let $\{y_1, y_2\}$ be solutions to $\mathscr{D}[y] = 0$. Then

$$y_p = v_1y_1 + v_2y_2$$

Setting $v_1'y_1 + v_2'y_2 = 0$, we get

$$v_1'y_1' + v_2'y_2' = f(t)$$

So we have the system

$$\underbrace{\begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix}}_{W}\begin{bmatrix} v_1' \\ v_2' \end{bmatrix} = \begin{bmatrix} 0 \\ f(t) \end{bmatrix}$$

Define the Wronskian W. Then if $\det W \neq 0$, the variation of parameters has a solution. By Cramer's rule, we have

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{\det W_i(\mathbf{f})}{\det W}$$

In particular,

$$\frac{\mathrm{d}v_1}{\mathrm{d}t} = \frac{\det\begin{bmatrix} 0 & y_2 \\ f & y_2' \end{bmatrix}}{\det\begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix}} \to v_1 = -\int \frac{f(t)y_2(t)}{y_1y_2' - y_2y_1'}\,\mathrm{d}t$$

Similarly for v_2, we have:

$$\frac{\mathrm{d}v_2}{\mathrm{d}t} = \frac{\det\begin{bmatrix} y_1 & 0 \\ y_1' & f \end{bmatrix}}{\det\begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix}} \to v_2 = \int \frac{f(t)y_1(t)}{y_1y_2' - y_2y_1'}\,\mathrm{d}t$$

We could use reduction of order. In particular we could also define

$$W = \mathrm{e}^{-\int a_2(t)\mathrm{d}t}$$

and the reduction of order formula gives

$$\frac{\mathrm{d}W}{\mathrm{d}t} + a_1W(t) = 0$$

Now we do the same for higher order differential equations. Let the differential operator $\mathscr{D}$ be

$$\mathscr{D} = \sum_{k=0}^{n} a_k\frac{\mathrm{d}^k}{\mathrm{d}t^k}$$

and we are concerned with the differential equation $\mathscr{D}[y] = f(t)$. Let $\{y_1, \ldots, y_n\}$ be funcitons that satisfy $\mathscr{D}[y] = 0$. Then

$$y_p = \sum_{k=1}^{n} v_k(t)y_k(t)$$

Taking derivatives, we set all but the last of the sums of all terms with second derivatives in the parameters. So we get

$$\begin{array}{ccccccccc} v_1'y_1 & + & v_2'y_2 & + & \dots & + & v_n'y_n & = & 0 \\ v_1'y_1' & + & v_2'y_2' & + & \dots & + & v_n'y_n' & = & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & & & \\ v_1'y_1^{(n-2)} & + & v_2'y_2^{(n-2)} & + & \dots & + & v_n'y_n^{(n-2)} & & \\ v_1'y_1^{(n-1)} & + & v_2'y_2^{(n-1)} & + & \dots & + & v_n'y_n^{(n-1)} & = & f(t) \end{array}$$

which basically says that

$$\sum_{k=1}^{n} \frac{\mathrm{d}v_k}{\mathrm{d}t}\frac{\mathrm{d}^j y_k}{\mathrm{d}t^j} = 0$$

for $j = 1, 2, \dots, n-2$. Also

$$\frac{\mathrm{d}^j y_p}{\mathrm{d}t^j} = \sum_{k=1}^{n} v_k \frac{\mathrm{d}^j y_k}{\mathrm{d}t^j}$$

t for $j = 1, 2, \dots, n-1$.

Finally, we have

$$\sum_{k=1}^{n} \frac{\mathrm{d}v_k}{\mathrm{d}t}\frac{\mathrm{d}^{n-1} y_k}{\mathrm{d}t^{n-1}} + \sum_{k=1}^{n} v_k \frac{\mathrm{d}^n y_k}{\mathrm{d}t^n} + \sum_{j=1}^{n-1}\sum_{i=1}^{n} a_j v_i \frac{\mathrm{d}^j y_i}{\mathrm{d}t^j} = f(t)$$

Simplifying, we get

$$\sum_{k=1}^{n} v_k \left[\frac{\mathrm{d}^n y_k}{\mathrm{d}t^n} + \sum_{j=1}^{n-1} a_j \frac{\mathrm{d}^j y_k}{\mathrm{d}t^j}\right] + \sum_{k=1}^{n} \frac{\mathrm{d}v_k}{\mathrm{d}t}\frac{\mathrm{d}^{n-1} y_k}{\mathrm{d}t^{n-1}}$$

This gives a system

$$\begin{bmatrix} y_1 & y_2 & \dots & y_n \\ y_1' & y_2' & \dots & y_n' \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)} & y_2^{(n-1)} & \dots & y_n^{(n-1)} \end{bmatrix} \begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_n' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ f(t) \end{bmatrix}$$

Thus we have an explicit solution:

$$y = \sum_{k=1}^{n} c_k y_k + \sum_{k=1}^{n} \int \frac{y_k \det W_i(\mathbf{f})}{\det W}\,\mathrm{d}t = \sum_{k=1} y_k\left[c_k + \int \frac{W_k(\mathbf{f})}{W}\,\mathrm{d}t\right]$$

Define

$$\mathbf{y} = \begin{bmatrix} y \\ y' \\ \vdots \\ y^{(n-1)} \end{bmatrix}$$

Then the equation is

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_0/a_n & -a_1/a_n & -a_2/a_n & \dots & -a_{n-1}/a_n \end{bmatrix}\mathbf{y} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ f(t) \end{bmatrix}$$

The fundamental matrix of the equation $\mathbf{y}' = A\mathbf{y} + \mathbf{f}(t)$ is $X = \begin{bmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_n \end{bmatrix}$ where each $\mathbf{y}_k$ solves the differential equation.

Write
$$\mathbf{y}_p = X(t)\mathbf{v}(t)$$
then
$$\frac{\mathrm{d}\mathbf{y}_p}{\mathrm{d}t} = \frac{\mathrm{d}X}{\mathrm{d}t}\mathbf{v} + X\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}$$
Substituting in, we get
$$A\mathbf{y}_p + \mathbf{f} = AX\mathbf{v} + \mathbf{f}$$
and so
$$\frac{\mathrm{d}X}{\mathrm{d}t} = AX$$
and
$$X\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{f}$$
We get
$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = X^{-1}\mathbf{f}$$
and so
$$\mathbf{v} = \int X(t)^{-1}\mathbf{f}(t)\,\mathrm{d}t$$
Note that
$$\mathbf{y}_h = X\mathbf{c}$$
and
$$\mathbf{y}_p = X\int X^{-1}\mathbf{f}(t)\,\mathrm{d}t$$
so
$$\mathbf{y} = X\left[\mathbf{c} + \int X^{-1}\mathbf{f}\,\mathrm{d}t\right]$$
Since $X = \mathrm{e}^{At}$, we get
$$\mathbf{y} = \mathrm{e}^{At}\left[\mathbf{c} + \int \mathrm{e}^{-At}\mathbf{f}\,\mathrm{d}t\right]$$

56 November 26, 2018

Today we discuss the eigenvector method for matrix ordinary differential equations with constant coefficients. Consider the homogeneous equation
$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x} \tag{H}$$
and the corresponding inhomogeneous equation
$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x} + \mathbf{b} \tag{I}$$

Remark 1 The eigenvector method forks for diagoanlizable A. Otherwise, we need generalized eigenvectors.

The goal is to find a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of A-eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$. Write

$$\mathbf{x}(t) = \sum_{k=1}^{n} c_k(t)\mathbf{v}_k$$

where the c_i are coordinates in the eigenbasis. We need a change of coordinates formula to find the β_i from $\mathbf{b}$. Write the equations (H, I) in the basis, noting that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Equate the eigencoordinates, getting equivalently

$$\frac{\mathrm{d}c_k(t)}{\mathrm{d}t} = \lambda_k c_k \tag{H}$$

and

$$\frac{\mathrm{d}c_k(t)}{\mathrm{d}t} = \lambda_k c_k + \beta_k \tag{I}$$

The homogeneous solution is

$$c_k(t) = c_k(0)\mathrm{e}^{\lambda_k t}$$

A particular inhomogeneous solution is

$$c_k(t) = \mathrm{e}^{\lambda_k t} \int_0^t \mathrm{e}^{-\lambda_k s} \beta_k(s)\,\mathrm{d}s$$

so

$$\mathbf{x}(t) = \sum_{k=1}^{n} \left(\mathrm{e}^{\lambda_k t} \int_0^t \mathrm{e}^{-\lambda_k s} \beta(s)\,\mathrm{d}s \right) \mathbf{v}_k$$

Example 56.1

Solve the system

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix} \mathbf{x}$$

Solution Let $\mathbf{x} = (y, y')$. The normal form of the first order system is given by

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} - 3\frac{\mathrm{d}y}{\mathrm{d}t} + 2y = 0 \tag{H}$$

The roots of the auxiliary polynomial are $\lambda_1 = 1$ and $\lambda_2 = 2$. The eigenvalues of A are $\mathbf{v}_1 = (1, 1)$ and $\mathbf{v}_2 = (1, 2)$. The general homogeneous solution is given by

$$\mathbf{x}(t) = c_1(0)\begin{bmatrix} 1 \\ 1 \end{bmatrix}\mathrm{e}^t + c_2(0)\begin{bmatrix} 1 \\ 2 \end{bmatrix}\mathrm{e}^{2t} = \begin{bmatrix} c_1(0)\mathrm{e}^t + c_2(0)\mathrm{e}^{2t} \\ c_1(0)\mathrm{e}^t + 2c_2(0)\mathrm{e}^{2t} \end{bmatrix}$$

To find initial conditions, note that $\mathbf{x}(0) = \mathbf{v}$ for some $\mathbf{v}$, find $c_1(0)$ and $c_2(0)$, so that $c_1(0)\mathbf{v}_1 + c_2(0)\mathbf{v}_2 = \mathbf{v}$.

□

Example 56.2

Solve the system

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} \mathbf{x}$$

Solution The characteristic polynomial of A is $\chi_A(\lambda) = \lambda^2\lambda + 1$. This is the normal (matrix) form of a scalar equation $y'' + y' + y = 0$. Note that

$$\lambda^2 + \lambda - 1 = \frac{\lambda^3 - 1}{\lambda - 1}$$

so the eigenvalues are

$$\lambda_\pm = -\frac{1}{2} \pm \frac{\mathrm{i}\sqrt{3}}{2}$$

and the corresponding eigenvalues are given by

$$\mathbf{v}_\pm = \begin{bmatrix} 1 \\ \lambda_\pm \end{bmatrix}$$

Note that $\lambda_+ = \overline{\lambda_-}$, so $\mathbf{v}_+ = \overline{\mathbf{v}_-}$. The general homogeneous solution is given by

$$\mathbf{x}(t) = c_+ \mathrm{e}^{\lambda_+ t}\mathbf{v}_+ + c_- \mathrm{e}^{\lambda_- t}\mathbf{v}_- = \begin{bmatrix} c_+\mathrm{e}^{\lambda_+ t} + c_-\mathrm{e}^{\lambda_- t} \\ \lambda_+ c_+ \mathrm{e}^{\lambda_+ t} + \lambda_- c_- \mathrm{e}^{\lambda_- t} \end{bmatrix}$$

To get real values from this we need $\mathrm{Im}\,(c_+ + c_-) = 0$ and $\mathrm{Re}\,(c_+ - c_-) = 0$. Hence $c_+ = \overline{c_-}$. Let $c_+ = a + ib$ so $c_- = a - ib$. Then

$$\mathbf{x}(t) = 2\mathrm{e}^{-t/2}\left(a\left(\cos\left(\frac{\sqrt{3}}{2}t\right)\mathrm{Re}(\mathbf{v}_-) + \sin\left(\frac{\sqrt{3}}{2}t\right)\mathrm{Im}(\mathbf{v}_-)\right) + b\left(-\sin\left(\frac{\sqrt{3}}{2}t\right)\mathrm{Re}(\mathbf{v}_-) + \cos\left(\frac{\sqrt{3}}{2}t\right)\mathrm{Im}(\mathbf{v}_-)\right)\right)$$

If we commit to the basis given by the matrix $\beta = \begin{bmatrix}\mathrm{Re}(\mathbf{v}_-) & \mathrm{Im}(\mathbf{v}_-)\end{bmatrix}$, the solution is

$$[\mathbf{x}]_\beta = \mathrm{e}^{-t/2}\begin{bmatrix} \cos\left(\sqrt{3}t/2\right) & -\sin\left(\sqrt{3}t/2\right) \\ \sin\left(\sqrt{3}t/2\right) & \cos\left(\sqrt{3}t/2\right) \end{bmatrix}\begin{bmatrix} 2a \\ 2b \end{bmatrix}$$

□

Example 56.3

Solve the system

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}\mathbf{x} = \begin{bmatrix} \mathrm{e}^t \\ t\mathrm{e}^t \end{bmatrix}$$

Solution We diagonalize A. We get $\lambda_1 = 1, \lambda_2 = 2$, and $\mathbf{v}_1 = (1, 1)$, $\mathbf{v}_2 = (1, 2)$. We can follow the procedure above to find $\mathbf{x}_p$. We know what kind of answer to look for by undetermined coefficients. From the formula it's clear that $\mathbf{x}_p(t) = (f(t), g(t))$ where both f and g are the product of a polynomial and exponential function. We compute the integrals to find our coefficient functions:

$$c_k(t) = \mathrm{e}^{\lambda_k t}\int_0^t \mathrm{e}^{-\lambda_k s}\mathrm{e}^s p(s)\,\mathrm{d}s$$

where $p(s)$ is a single-variable polynomial in s. For $\lambda_1 = 1$, we get

$$c_1(t) = \mathrm{e}^t\int_0^t p(s)\,\mathrm{d}s$$

□

When solving

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x} + \mathbf{b}$$

when A is diagonalizable and each entry of $\mathbf{b}$ is a polynomial of degree d times $\mathrm{e}^{\mu t}$ for some μ, a particular solution has the form $\mathbf{x}_p$ where each of the entries is $p(t)\mathrm{e}^{\mu t}$ for the polynomial $p(t)$, with degree d if μ is not an eigenvalue or $d + 1$ if it is.

The vector method of undetermined coefficients requires a guess of coefficient vectors, then plug into the ordinary differential equation and solve.

We turn to the computation of the matrix exponential in the case where the matrix is diagonalizable. Let A be a square matrix. It satisfies

$$\frac{\mathrm{d}(\exp(At))}{\mathrm{d}t} = A\exp(At)$$

If $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then $\exp(D) = \mathrm{diag}(\exp(\lambda_1), \ldots, \exp(\lambda_n))$. If A is diagonalizable, then

$$A = PDP^{-1} \to \exp(A) = P\exp(D)P^{-1}$$