Everything about AM 21b 2017

Mathematical Methods in the Sciences (Really, Linear Algebra)

- For an updated version, see Everything about AM 21b 2018

- **Kiffen Loomis** suggests that I call this google doc “linear algebra in 3000 bullets.”
- **Lucas Guzman** suggests that I place these images here.


- The solutions were prepared by teaching fellows. Please contact them if you have any questions.

- Homework 1 due February 3. Solution 1
- Homework 2 due February 10. Solution 2
- Getting ready for Midterm I, Thursday, 7:30-9:00 pm, 16 February 2017
- Solution to supplementary problem set I
- Midterm I February February 16. Solution to Midterm I
- Homework 3 due February 24. Solution 3
- Homework 4 due March 3. Solution 4
- Homework 5 due March 10. Solution 5
- Homework 6 due March 24. Solution 6
- Getting ready for Midterm II,Thursday, 7:30-9:00 pm, 30 March 2017
- Solution to Supplementary Problem Set II
- Midterm II March 30. Solution to Midterm II
- Homework 7 due April 7. Solution 7
- Homework 8 due April 14. Solution 8
- Homework 9 due April 21. Solution 9
- Homework 10, not collected, but required for final. Solution 10
- Getting ready for Final
- Final, Tuesday, 9:00 am, 9 May, Emerson 105

Lectures are mandatory. Northwest Building B103. Monday, Wednesday, Friday, 11:00am - 11:59am.

Sections
Monday 4:00-5:30 pm, Maxwell Dworkin 223 (**Sijie Sun**)
Tuesday 3:00-4:30 pm, Robinson 106 (**Emily Venable**)
Wednesday 1:30-3:00 pm, Geological Museum 418 (**Nicky Charles**)
Thursday 4:30-6:00 pm, Maxwell Dworkin 119 (**Jacob Scherba**)

Office Hours
Monday 5:30-7:00 pm, Maxwell Dworkin 223 (**Sijie Sun**)
Tuesday 4:30-6:30 pm, Kirkland Dining Hall (**Emily Venable** and **Jacob Scherba**)
Wednesday 3:00-5:00 pm, Sever 211 (**Nicky Charles**)
Thursday 3:00-5:00 pm, Pierce 309 (**Professor Suo**)
Friday 1:30 - 3:30pm, Boylston Hall G02 (**Sijie Sun**)

Instructor
Professor **Zhigang Suo** (suo@seas.harvard.edu)

Teaching fellows:
- **Nickolas Charles** <ncharles@seas.harvard.edu> Head TF
- **Jacob Scherba** <jscherba@college.harvard.edu>
- **Sijie Sun** <sijiesun@g.harvard.edu>
- **Emily Venable** <emilyvenable@college.harvard.edu>

Students in the class

Course description. This course is devoted to linear algebra. Basic topics include linear algebraic equations, matrices, determinants, and eigenvalues. Foundational ideas include numbers, scalars, vectors, and tensors, as well as linear maps, linear forms, dual spaces, quadratic forms, and inner-product spaces. Applications draw upon everyday experience, engineering, science, and economics.

Textbook. D.L. Lay, Linear Algebra and Its Applications, 4th edition, 2012.

Grades. Your grade for the course will be determined as follows:
- Homework (10 problem sets) 20%
- Midterm I (7:30-9:00 pm, Thursday, February 16, Northwest Building B103) 20%
- Midterm II (7:30-9:00 pm, Thursday, March 30, Northwest Building B103) 20%
- Final Exam (three hours) 40%

How to do well in this class?
- All problems on the exams will be similar to those from lectures, sections, textbook, and homework.
- Return to this google doc (everything about am 21b) frequently.
- Mark the midterms in your calendar now.
- Do not overcommit yourself in these two weeks.
- Keep good notes in lectures and sections.
- Master homework problems. They are similar to exam problems.

MATLAB. You will need MATLAB to solve some homework problems. You will also likely use MATLAB in your future life as well. Please attend the MATLAB Boot Camp. The first session starts on Monday January 30.

Coming to lectures is mandatory. Linear algebra became a college course only in recent decades, after computers are widely available. As computers, software and applications develop, the selection of materials and presentation also evolve. **I'll try my best to follow the textbook, but will draw upon current experience in using linear algebra**. I'll be explicit whenever I deviate from the textbook. Of course, coming to lectures and sections provides some structure to your learning. Learning is social. Don't be a loner.

Reading and watching. I posted my notes on linear algebra online. I wrote these notes for myself, and for people who have similar mathematical experience like me. These notes are *not* intended for this course. On occasions, I will draw upon bits and pieces of these notes in class. I will be explicit about what you need to read from them.

Your primary reading should be **this google doc**, the **textbook by Lay** and, most importantly, **your own notes** of lectures and sections. Available time is not enough for me to say everything in lectures. You will have to do some reading on your own.

Occasionally I'll post links to videos found online. Watching videos is supplementary, not mandatory.

Section. The Teaching Fellows will run weekly recitation sections. You may attend any section. Although section attendance is optional, it is in your best interest to attend, especially considering the quick pace of this course.

Homework. Homework will be posted on Canvas on Friday morning and due the **following Friday, at 5 pm**, **on Canvas**. Scan handwritten work using an app like “camScanner”. Canvas will only accept pdf and word document file types. Late homework will not be accepted; however, we will drop your homework set of the lowest grade when computing your final course grade. There will be no homework due during exam weeks.

Discussing homework is encouraged, but you must write your solutions and Matlab codes independently. Homework problems may have pre-existing solutions online. It is wrong to copy the pre-existing solutions.

Copying homework also hurts you in another way. Homework itself only contributes 20% of your grade for the course. There will be about 100 homework problems. Each problem only contributes 0.02% to your grade. But all exams will be similar to homework. Thus, doing homework is a significant way to prepare for exams. Copying homework means that you waste opportunities to prepare for exams.

A few general mathematical terms

- These terms are not specific to linear algebra, but are used extensively in linear algebra.
- You may have learned these terms before, but in studying linear algebra, **you must master them**.

Week 1 (January 23, 25, 27)

Reading and watching

- Video on row reduction.
- Video on using the MATLAB command rref. You only need to watch the beginning 2 minutes.
- Video on drawing equations in a plane.
- Suo notes on linear equation
- Lay 1.1 A system of linear equations
- Lay 1.2 Row reduction and echelon forms
- J.F. Grcar, Mathematicians of Gaussian Elimination. Notices of the American Mathematical Society 58, 782-792 (2011).

Lay 1.1 linear equation

- Solve a real-world problem in two steps:
- **Step 1. Set up equations**.
- **Step 2. Solve the equations**.

Chickens and rabbits

- A farm has chickens and rabbits. The farmer counts 26 heads and 82 feet. How many chickens and rabbits are there?

- **Translate this "real-world problem" (i.e., a word problem) to equations**.
- Let x be the number of chickens on the farm
- Let y be the number of rabbits on the farm.
- Write the equation of heads: $x + y = 26$.
- Write the equation of feet: $2x + 4y = 82$.
- The coefficients of these equations come from your worldly knowledge of chickens and rabbits, not from the statement of the word problem.
- This worldly knowledge consists of four facts:
- Each chicken has one head,
- Each chicken has two feet,

- Each rabbit has one head, and
- Each rabbit has four feet.
- We have just translated the word problem to **a system of two linear algebraic equations in two variables**.

- **Solve the equations by the method of elimination of variables**.
- The solution consists of two numbers:
- x = 11 chickens
- y = 15 rabbits

- **Graphical representation in the chicken-rabbit plane**.
- In a plane, such as a piece of paper, draw one line to represent various numbers of chickens, and draw another line to represent various numbers of rabbits.
- Each point in the plane represents a pair of numbers: a number of chickens, x, and a number of rabbits, y.
- Represent the equation of heads, x + y = 26, by one line.
- Represent the equation of feet, 2x + 4y = 82, by another line.
- The two lines intersect at one point, x = 11 chickens, y = 15 rabbits.
- This point gives the solution of the system, the number of chickens and the number of rabbits.

Rabbits and hamsters

- **First try**. A farm has rabbits and hamsters. The farmer counts 26 heads and 82 feet. How many rabbits and hamsters are there?
- Let y be the number of rabbits on the farm
- Let z be the number of hamsters on the farm.
- Write the equation of heads: y + z = 26.
- Write the equation of feet: 4y + 4z = 82.
- The system has no solution.

- Graphical representations of the equation of heads and the equation of feet, both in the **rabbit-hamster plane**.
- The two equations correspond to two parallel lines.
- The two parallel lines do not intersect.

- **Second try**. A farm has rabbits and hamsters. The farmer counts 26 heads and 104 feet. How many rabbits and hamsters are there?
- Let y be the number of rabbits on the farm, and z be the number of hamsters on the farm.
- Write the equation of heads: y + z = 26.
- Write the equation of feet: 4y + 4z = 104.
- The system has many solutions.

- Graphical representations of the equation of heads and the equation of feet, both in the **rabbit-hamster plane**.

- The two equations correspond to a single line.
- Every point on the line is a solution.

Chickens, rabbits, and hamsters

- A farm has chickens, rabbits, and hamsters. The farmer counts 26 heads and 82 feet. How many chickens, rabbits, and hamsters are there?
- Let x be the number of chickens on the farm, y be the number of rabbits on the farm, and z be the number of hamsters on the farm.
- Write the equation of heads: $x + y + z = 26$.
- Write the equation of feet: $2x + 4y + 4z = 82$.
- The system has many solutions.
- Graphical representation in the **chicken-rabbit-hamster space**.
- The equation of heads is a plane in the space.
- The equation of feet is another plane in the space.
- The two planes intersect at a line.
- Each point on the line is a solution to the system of two equations in three variables.

Summary in abstract terms

- A system of m linear algebraic equations in n variables.
- m = the number of **equations**.
- n = the number of **variables**.
- In general, $m \neq n$.
- **Coefficients of the system** are an m x n table of numbers.
- **Constant terms** are an m-tuple of numbers.
- **Variables** are an n-tuple of numbers.

- A **solution** to a system of m equations in n variables is an n-tuple of numbers that satisfy the m equations.
- The set of all solutions to a system is called the **solution set** of the system.

- **Solution set of three types**.
- **Single-element set** (unique solution). Chickens and rabbits.
- **Empty set** (no solution). Rabbits and hamsters, first try.
- **Many-element set** (many solutions). Rabbits and hamsters, second try. Chickens, rabbits, and hamsters.

- For a system of equations having many solutions, we divide the variables into two kinds.
- **Free variables** can take any values.
- **Basic variables** are expressed in terms of the free variables.

- Graphical representations are helpful for a system of equations in two variables, hard to draw for a system of equations in three variables, and impossible for a system of equations in more than three variables.

- An important skill is to represent data on a plane.

Lay 1.2 row reduction algorithm

- **Matrix of coefficients**.
- **Constant terms**.
- **Augmented matrix**.

- **Row operations of three types**

1. Swap two rows.
2. Multiply a row by a nonzero number.
3. Replace a row by the addition of the row and a multiple of another row.

- A matrix is in a **reduced row echelon form (rref)** if the following conditions hold.

1. In each nonzero row, the leftmost nonzero entry is one. This entry is called a **pivot**, its position is called a **pivot position**, and the corresponding column is called a **pivot column**.
2. In each pivot column, all entries above and below the pivot are zero.
3. The pivot in each nonzero row is to the right of the pivot in the row above.
4. All zero rows are at the bottom of the augmented matrix.

- Row operations do not change the solution set of a system of equations.
- A matrix A has a **unique** rref, independent of the sequence of row operations.
- Write the rref of A as **rref (A)**.
- The **row reduction algorithm** uses row operations to convert the augmented matrix to its reduced row echelon form.

- **Row reduction by hand**
- In exams, you perform the row reduction algorithm by hand.
- Master the steps.
- In class, I use an example to illustrate the algorithm.
- The row reduction algorithm is also called **Gaussian elimination**.
- Watch this video for another example of row reduction.

- **Row reduction using computers**
- In real life, you perform the row reduction algorithm using computers.
- Learn the MATLAB command, rref, by watching the first two minutes of this video.

- **From a reduced row echelon form of the augmented matrix, read out the solution set of three types.**
- When the rightmost column of the augmented matrix is a pivot column, the system has **no solution**.
- When the rightmost column is a nonpivot column, and when the number of pivots equals the number of variables, the system has a **unique solution**.
- When the rightmost column is a nonpivot column, and when the number of pivots is less than the number of variables, the system has **many solutions**.

- In the last case, **write the complete solution set** in the following steps.

1. In the rref of the matrix of coefficients, identify all pivot columns. Each pivot column corresponds to a basic variable, and each nonpivot column corresponds to a free variable.
2. Translate the rref of the augmented matrix back to a system of equations.
3. Move the free variables to the right side.

- The free variables can take any values.
- Each choice of the values of the free variables gives a solution to the system of equations.
- In class, I use an example to illustrate the steps.

- **Adil Bhatia's question**: "In high school I wrote equations one by one. Can I keep avoiding matrices?"
- Answer: You can get away without writing matrices for a while, but soon you will see the joy of matrices.

Week 2 (January 30, February 1, 3)

Reading and watching

- Video on vector
- Video on row picture and column picture of a system of equations.
- Suo notes on number
- Suo notes on scalar
- Suo notes on vector
- Suo notes on Sankey diagram (energy flow charts of the US and China)
- Wiki unit of measurement
- Lay 1.3 Vector equation
- Lay 1.4 Matrix equation
- Lay 1.8 Linear map (linear transformation)

- **The row reduction algorithm is a hammer looking for nails**.
- You, and certainly computers, know how to solve equations.
- But where do equations come from?
- They come from chickens, rabbits, hamsters, heads, and feet, as well as other things in the world.
- How do these things become equations?

Number field

- A **number field** is a set F that obeys some axioms.
- Each element in F is called a **number**.
- **The number field is closed under two operations**:
- The addition of two elements in F gives another element in F,

- The multiplication of two elements in F gives another element in F.
- Subtraction and division are defined by addition and multiplication.
- These statements are explained in my notes on number.
- The notes list the axioms of number field, but you do not need to memorize them.
- The axioms define number field as an **algebraic structure.**
- You have been conditioned to manipulate numbers since childhood.
- **This course will only use two number fields**:
- the field of real numbers, R, and
- the field of complex numbers, C.
- The field of real numbers, R, should be thoroughly familiar to you.
- For much of the course, we will use real numbers only.
- But complex numbers will appear in significant ways in a few places.
- **When a statement is valid for both R and C, we often denote the number field by F**.

Gold set

- Pieces of gold of all sizes form a set, called the **gold set**.
- Here are some elements in the gold set:
- (1 oz of gold),
- (1.3 oz of gold),
- (100 gold atoms),
- (1 kg of gold), and
- (a really large piece of gold).
- (-1.5 oz of gold) is also in this set, and means that we are in deficit of 1.5 oz of gold.
- **Is the gold set a number field?**
- Let's see.
- The gold set is closed under addition: the addition of two pieces of gold gives another piece of gold.
- The gold set, however, does not have a meaningful definition of multiplication.
- What does the multiplication of two pieces of gold even mean?
- **The gold set is *not* a number field**.
- Rather, the gold set has a different algebraic structure, called scalar set, as defined below.

Scalar set

- A collection of things S is called **a scalar set over a number field F** if the things in S are additive to one another, scalable by numbers in F, and proportional to one another.
- Each thing in S is called a **scalar**.
- The defining attributes of scalar set are described in the following three bullets.
- **Additivity**. If x and y are elements in S, then x + y is also an element in S.
- **Scalability**. If x is an element in S and c is an element in F, then cx is also an element in S.
- **Proportionality**. If u is a nonzero element in S, then every element x in S is in the form x = cu, where c is a number in F.

- **Proportionality implies scalability**. Thus, there is no need to list scalability in defining scalar set.
- Here I list scalability for later comparison with vector space. As you will see, vectors, in general, are not proportional to one another, but are scalable.

- **A unit for a scalar set**
- We call u a **unit** for the scalar set S, and c the **magnitude** of the scalar x relative to the unit u.
- For example, various amounts of dollars form a scalar set, called the dollar set.
- One dollar, denoted $, serves as a unit for the dollar set.
- 100 $ is an element in the dollar set, where the number 100 is the magnitude of this element relative to the unit ($) for the dollar set.

- **Examples**.
- We do not become good citizens by reading legal documents.
- We become good citizens by following good examples.
- Here are some examples of scalar sets.
- The collection of pieces of gold of all sizes is a scalar set, called the gold set.
- Piles and piles of apples form a scalar set, called the apple set.
- Arrows in a line.
- Piles and piles of US dollars.
- Various amounts of a commodity.
- Various numbers of chickens.
- Various numbers of feet.
- Various amounts of a pure substance, like water.
- Durations of time.
- Displacements in a direction.
- Mass.
- Charge.
- Energy.
- Entropy.

- **Things that do not form scalar sets**.
- Temperature does not form a scalar set.
- Nor does happiness, nor intelligence, nor love.
- Scales of earthquakes do not form a scalar set.
- Pieces of goldsilver of all sizes do not form a scalar set.

- **Change of unit**.
- Any nonzero element in a scalar set S is a unit for S.
- Let u and v be two nonzero scalars in S.
- The two scalars are proportional. Write
- v = pu

- The nonzero number p in F is called the **factor of conversion** from the unit u to the unit v.

- **Change of magnitude under a change of unit**
- Let u be a nonzero element in S.
- Write any element x in S as x = cu.
- The number c in F is the magnitude of the scalar x in S relative to the unit u for S.
- Any nonzero number p in F is a factor of conversion.
- When unit for S changes from u to pu, the magnitude of a scalar x in S changes from c to c/p.
- The magnitude of every scalar in S is **contravariant** with the unit for S.

- A scalar set S over the field of real numbers R is called a **real scalar set**.
- **Examples of real scalar sets**.
- Pieces of gold of all sizes form a real scalar set.
- Each type of commodity forms a real scalar set.
- Money is a real scalar set.
- Time, length, mass, charge, energy, entropy each is a real scalar set.

- **The discovery of each basic scalar set in nature is a fundamental advance in science**.
- Time, length, area, volume, and amount of substance were discovered in antiquity.
- The concept of mass was formulated by Galileo and Newton.
- Charge, energy, and entropy were discovered over a hundred years ago.
- The concepts of atoms and molecules took the modern forms about a hundred years ago.
- No new basic scalar set of practical significance has been discovered since.

- Sankey diagram is a visual display of the **flow** of a real scalar, from many sources, via many branches, into many sinks.
- The width of each branch is proportional to the magnitude of the scalar.
- The branches are additive.
- The best known Sankey diagram is the **energy flow chart** of the United States.

- **The field of real numbers R is a real scalar set**.
- The elements in R, the real numbers, are additive, scalable by real numbers, and proportional to one another.
- Thus, R is a real scalar set.

- **Arrows in a line form a real scalar set**. We confirm this statement by confirming that the set of arrows in a line has the defining attributes of a real scalar set: additivity, scalability, and proportionality.
- We do so in the next three bullets.

- **Additivity: The addition of two arrows in the line gives an arrow in the line**. Let x and y be two arrows in the line. Translate the arrow y so that the tail of y coincides with the head of x. The arrow from the tail of x to the head of y is x + y.
- **Scalability: The multiplication of an arrow and a real number gives another arrow**. Let x be an arrow in a line, and c be a real number. cx means an arrow of length c times the length of x.
- **Proportionality**: Let u be a nonzero arrow in the line. Any arrow x in the line is some real number c times u. That is, x = cu.

- A scalar set over the field of complex numbers is called a **complex scalar set**.
- Complex scalar sets are used to model oscillating electrical and mechanical systems.
- For the time being, we focus on real scalar sets.
- We will study complex scalar sets later.

Number field and scalar set are distinct algebraic structures

- My notes on scalar describe the defining attributes (i.e., the axioms of scalar set) in detail, but you do not need to memorize them.
- Study the examples, and learn to identify scalar sets over number fields.
- Homework will help you understand and differentiate the three algebraic structures: number field, scalar set, and vector space.
- The concept of scalar set is the foundation of linear algebra, but somehow the concept has not been isolated before, and its foundational role has remained obscure.
- I use the phrase "scalar set" to mimic the established phrase "vector space". Indeed, as we will see later, **a scalar set is the same as a one-dimensional vector space**.

- **Warning**. As noted at the end of my notes on scalars, textbooks of linear algebra and Wikipedia confuse the two distinct algebraic structures: number field and scalar set.
- In textbooks, and in the Wikipedia entry on scalar, when a number c is used to multiply a vector, the number c is called a scalar.
- This usage differs from the second usage of the word "scalar", which also appears in every textbook in linear algebra. In the second usage, the addition of two scalars in a set gives a scalar in the set, but multiplication of two scalars does not give another scalar in the set. Indeed the multiplication of two scalars are often meaningless.
- My notes on scalar explains the different usages further.
- **Being right**. In class, we will not use the word "scalar" to mean two different algebraic structures. We will call a number a number: a number needs no new name. **We will reserve the word "scalar" for an element in a scalar set**. I have two sons, and I give them different names, Daniel and Michael.
- **Being pragmatic.** Of course, in reading the literature and in talking to people outside the class, from the context, you can differentiate if the word "scalar" means an element in a number field, or an element in a scalar set.

- **The same name for two things**. It is OK to use one word to name two things, but it is bad to confuse two things just because they happen to have the same name. I have met many people named Daniel, but have never confused them with my son. There is no reason to confuse two distinct algebraic structures just because textbooks and Wikipedia use one word to name both, or leave one algebraic structure obscure. We can debate the wisdom of using one word to name two algebraic structures, but should not let this practice confuse them.
- Wiki units of measurement.
- My definition of scalar set is related to units of measurement, but difference is also significant.
- Scalar set is an abstract mathematical idea, specified by attributes: additivity, scalability, and proportionality.
- Temperature does not form a scalar set, but has many units of measurement.

Linear map from one scalar set to another scalar set

- **Map**. A map is also called a function, mapping, or transformation.
- If the idea of map is hazy to you, please look at the google doc on a few general mathematical terms.

- **Scalar-scalar linear map.**
- Let S and T be two scalar sets over a number field F.
- A map f: S → T is called a **linear map** if
- f(cs) = cf(s)
- for every scalar s in S and every number c in F.
- That is, if f sends a scalar s in S to a scalar t in T, then f also sends the scalar cs in S to the scalar ct in T for any number c in F.
- We often speak of linear map f using a phrase like **the amount in T per unit amount in S**.

- Individual scalar sets are the building blocks of all vector spaces.
- Individual scalar sets are the building blocks of all vector-vector linear maps.

- **Examples of scalar-scalar linear maps**.
- f(x) = 2x is a linear map from R to R.

- A linear map from the chicken set to the foot set:
- the number of feet per chicken.
- One chicken has two feet, three chickens have six feet, etc.

- A linear map from water-molecule set to hydrogen-atom set.
- Each water molecule has two hydrogen atoms.
- Three water molecules have six hydrogen atoms.

- A linear map from the scalar set of beef to the scalar set of protein: the mass of protein per unit mass of beef, also known as the protein content of beef.

- A linear map from the gold set to the dollar set: an amount of dollars per unit mass of gold, also called a price of gold.
- You can find the price online. It is about $40 per gram of gold.

- A linear map from the scalar set of US dollars to the scalar set of Euros: an exchange rate.

- For a pure substance, such as water, mass, volume, energy, entropy are all proportional to the number of molecules.
- For instance, the mass per water molecule defines a linear map from the water set to mass set.
- For a given pure substance, the mass per unit volume, known as the density of the substance, is a linear map from the volume set to the mass set.

- **Examples of nonlinear maps**
- $f(x) = x^2$.
- This function sends a real number to a real number, $f: R \to R$.

- Even $f(x) = 2x + 1$ is *not* a linear map.
- Linear map must satisfy $f(0) = 0$.

- **Graphical representation of a linear map from one real scalar set S to another real scalar set T**.
- You use such a graphical representation every day. Here is a reminder of details.
- Use the chicken set and foot set to illustrate this graphical representation of the linear map
- The number of feet per chicken is 2. Two chickens have four feet, three chickens have six feet, etc.
- In a plane, such as a piece of paper, draw a line to represent the real scalar set S, the chicken set.
- Choose a point in the line to represent the zero scalar in S.
- Choose a scalar u in S as a unit for S. We choose 1 chicken as a unit for the chicken set.
- Choose another point in the line to represent u.
- Every point in the line represents a scalar in S.
- Mark in the line the scalars 2u, 3u, -u, -2u, etc.
- In the same plane, draw another line to represent the other scalar set T, the foot set.
- We call the plane the S-T plane, the chicken-foot plane.
- Often we draw the line for T to be perpendicular to the line for S, but this preference is merely a convention, and is insignificant.
- Make sure the intersection of the two lines represents the zero scalar in S and the zero scalar in T. Call the intersection the origin of the S-T plane.
- Choose a scalar v in T as a unit for T. We choose 1 foot as a unit for the foot set.
- Choose another point in the line for T to represent a unit v of T.
- Every point in the line represents a scalar in T.

- Mark in the line the scalars 2v, 3v, -v, -2v, etc.
- The two units, u for S and v for T, need not be of equal length.
- Often we draw a grid, that is, two sets of lines parallel to the lines for S and T.
- Each point on the grid represents an ordered pair of elements in S and T, some number of chickens and some number of feet.
- The linear map A is the collection of ordered pairs of elements in S and T.
- The line contains the points (1 chicken, 2 feet), (2 chickens, 4 feet),...
- **The linear map A from a scalar set S to another scalar set T is a line that passes the origin in the S-T plane**.

- **A nonzero scalar-scalar linear map is bijective**. A nonzero linear map is a one-to-one correspondence between two scalar sets, and the correspondent elements in the two sets are in proportion.

Isomorphism

- Two scalar sets S and T are called **isomorphic**, written S ~ T, if there exists a nonzero linear map between them.
- **All real scalar sets are isomorphic**.
- This statement is understood as follows.
- Let s be a nonzero scalar in S, t be a nonzero scalar in T, and a linear map A maps s to t, A(s) = t. For any real number c, cs is a scalar in S, ct is a scalar in T, and the linear map A maps cs to ct, A(cs) = ct.
- A real scalar set is not isomorphic to a complex scalar set.
- The set of dollars is isomorphic to the set of euros. Say the dollars and euros exchange at the rate of 1 dollar to 0.9 euros. For any real number c, this exchange rate maps c dollars to 0.9c euros. This exchange rate is a linear map from the set of dollars to the set of euros.
- **Isomorphism is the foundation** for exchange of currencies, trading of commodities, buying commodities with money, and graphical representation of scalars.
- **Our ancestors were bean counters**. In antiquity, before the notions of real numbers and arrows were isolated, people used things like beans and marbles to represent quantities of other things.

- **Two important ways to represent scalar sets**.
- We have seen many real scalar sets. We next select **two special scalar sets**: the field of real numbers R, and the set of arrows in a line.
- Because all real scalar sets are isomorphic, we can represent any real scalar set S by any other real scalar set. .

Represent a real scalar set S by R

- Every nonzero scalar u in S is a unit for S.
- Every scalar x in S is proportional to u. That is, x = cu, where c is a real number.
- Thus, **each choice of a unit for S creates a nonzero linear map from S to R.**
- We say that S and R are isomorphic, written S ~ R.

- This isomorphism allows us to state mathematical results almost exclusively using R, even though the results are valid for all real scalar sets.

Represent a real scalar set S by the set of arrows in a line

- All real scalar sets are isomorphic to the set of arrows in a line.
- We use this isomorphism to graph any real scalar set S.
- Let's use the chicken set as an example.
- In a plane, such as a piece of paper, draw a line to represent the real scalar set S, the chicken set.
- Choose a point in the line to represent the zero scalar in S.
- Choose nonzero scalar u in S as a unit for S. we choose 1 chicken as a unit for the chicken set.
- Choose another point in the line to represent u.
- Every point in the line represents a scalar in S.
- Mark in the line the scalars 2u, 3u, -u, -2u, etc.
- **Each choice of a unit for S and choice of a unit arrow define a linear map between the two scalar sets**.
- For any real number c, the arrow from 0 to c represents the scalar cu.
- This drawing is a linear map from the set of arrows in the line to any real scalar set.
- The set of arrows in the line is, of course, different from the set of chickens, but both sets are real scalar sets, and the two scalar sets are **isomorphic**.
- We write (chicken set) ~ (arrows in a line).
- **The elements in the two scalar sets correspond one to one, in proportion.**

Gold-silver space

- Various amounts of gold form a scalar set, called the **gold set**.
- Various amounts of silver form another scalar set, called the **silver set**.
- The Cartesian product of two scalar sets is called the gold-silver space:
- **(gold-silver space) = (gold set) x (silver set)**.

- The gold-silver space is also a set.
- (1 oz of gold, 1.3 oz of silver) is an element in the gold-silver space.
- You get all elements in the gold-silver space by independently changing the amounts of gold and silver.

- Is the gold-silver space a scalar set? Let's see.
- Gold-silver pieces are additive. The addition of two gold-silver pieces gives another gold-silver piece.
- Gold-silver pieces are scalable by real numbers. A gold-silver piece multiplied by a real number gives another gold-silver piece.
- Gold-silver pieces, however, are not proportional to one another.
- Different pieces may contain different ratios of gold and silver.

- Thus, **the gold-silver space is not a scalar set**.
- Rather, the gold-silver space has a different algebraic structure, called a two-dimensional vector space, as defined below.

Vector space

- An **n-dimensional vector space V over a number field F** is the Cartesian product of n scalar sets S_1, ...,S_n over F:
- $V = S_1$ x...x S_n.
- Each **vector** v in V is an n-tuple of scalars from the scalar sets S_1,...,S_n.
- $v = (s_1,...,s_n)$, where s_i is a scalar in the scalar set S_i.
- **(a vector) = (an n-tuple of scalars)**
- We call this process of stacking a list of scalar sets into a vector space **vectorizing scalar sets**.

- A vector space V over the field of real numbers R is called a **real vector space**.
- A vector space V over the field of complex numbers C is called a **complex vector space**.
- For the time being, let us focus on real vector spaces.

- **Jeanette Jin's question: Isn't a scalar set just a one-dimensional vector space**?
- Answer: Yes.
- A scalar set is a one-dimensional vector space.
- An n-dimensional vector space is a Cartesian product of n scalar sets.

- **Example**. Various amounts of gold form a real scalar set, called the **gold set**. Various amounts of silver form another real scalar set, called the **silver set**. The Cartesian product of the two scalar sets defines a two-dimensional real vector space, called the **gold-silver space**:
- **(gold-silver space) = (gold set) x (silver set)**.
- **Specify a vector by an n-tuple of scalars**. We can list the n scalars in a **row**, or in a **column**.
- For example, a mixture of 1.9 ounce of gold and 200 grams of silver is a vector in the gold-silver space.
- We write this vector as a tuple in a row, (1.9 ounce of gold, 200 grams of silver).
- We also write this vector as a column

1.9 ounce of gold
200 grams of silver

- We often write this vector as a column of numbers:

1.9
200

- Of course, the numbers are meaningful only if we know the unit for the gold set, and the unit for the silver set.

- **The units for different scalar sets can never be the same**.
- 1 ounce of gold and 1 ounce of silver are two different things, even if they have the same weight.
- $100 worth of gold and $100 worth of silver are two different things, even if they cost the same amount of money.

- **Examples of vector spaces**.
- A **one-dimensional real vector space** is, by definition, a real scalar set. We have seen many examples of real scalar sets:
- gold set, chicken set, foot set, arrows in a line.
- In particular, the field of real numbers is a one-dimensional real vector space, denoted R.
- Thus, R is a number field, a real scalar set, and a one-dimensional real vector space.

- **Examples of two-dimensional real vector space**.
- The chicken-rabbit space is a two-dimensional real vector space.
- In this space, (11 chickens, 15 rabbits) is a vector, and (2 chickens, 3 rabbits) is another vector.
- The head-foot space is also a two-dimensional real vector space.
- In this space, (26 heads, 82 feet) is a vector.
- All arrows in a plane form a two-dimensional real vector space.
- All ordered pairs of real numbers form a two-dimensional real vector space, denoted R^2.

- **Examples of three-dimensional real vector space**.
- Gold-silver-platinum pieces of all sizes.
- Chicken-rabbit-hamster space.
- Arrows in the physical space.
- All 3-tuples of real numbers form a three-dimensional real vector space, denoted R^3.
- All n-tuples of real numbers form an n-dimensional real vector space, denoted R^n.

- **A vector is a soup (tuple) of scalars**.
- Indeed, all food forms a vector space, which we call the **food vector space**.
- Each piece of food is a vector, a package (tuple) of water, fats, proteins, carbohydrates, and other molecules and atoms.
- Each species of molecules and atoms forms a scalar set.
- **A burger is a real vector**. We can list the ingredients of the burger and their amounts by a column. Each ingredient forms a real scalar set. A burger is a tuple of these ingredients. A cup of milk is another vector in the food vector space. What vectors will you have for dinner?
- In fact, anything made of atoms and molecules is a real vector. All things made of atoms and molecules form a real vector space. Each species of atoms forms a scalar set. This real

vector space is called **matter**. These simple facts underlie the applications of linear algebra to chemistry.

- [My notes on vector](#) discuss the appleorange and spacetime in parallel. But the notes are not written for this course, and contain more than what we will need for this course. You need not read them if you have attended lectures. Rather, **you should read and update your own notes of the lectures**.

- **We build n-dimensional vector space using n scalar sets and one number field**.
- Our textbook, however, defines vector space by a list of axioms, properties that are not derived from other facts.
- For us, all these axioms of vector space are facts obvious from the Cartesian product of scalars.
- For example, in our definition, the **dimension n** of a vector space V is a simple idea: n is just the number of scalar sets used in the Cartesian product.
- In particular, a 4-dimensional vector space should never be a myth. Here is a 4-dimensional vector space: piles and piles of apples, oranges, peaches, and pears. In this vector space, each vector is a pile containing some quantities of apples, oranges, peaches, and pears.
- The myth of four-dimensional space created after Einstein should be dispelled. Just think about the apple-orange-peach-pear space.
- **Additivity**. If u and v are vectors in an n-dimensional real vector space V, then u + v is also a vector in V. Example: (2 chickens, 3 rabbits) + (5 chickens, 7 rabbits) = (7 chickens, 10 rabbits). All three ordered pairs are vectors in the chicken-rabbit vector space. We add chickens to chickens, and add rabbits to rabbits. We do not add chickens to rabbits.
- **Scalability**. If u is a vector in V and c is a real number, then cu is also a vector in V. Example: 7(2 chickens, 3 rabbits) = (14 chickens, 21 rabbits).
- **Non-proportionality**. Let u and v be two vectors in a real vector space V. The two vectors are called proportional if there exists a real number c, such that u = cv. Example: (2 chickens, 3 rabbits) is proportional to (10 chickens, 15 rabbits), but is not proportional to (2 chickens, 5 rabbits). For a one-dimensional vector space (i.e., a scalar set), all scalars in the set are proportional to one another. For a vector space V of two dimensions or more, most vectors in V are not proportional to one another.

- **R^n is an n-dimensional real vector space**. By definition R^n is the Cartesian product of R by n times. Each R is a real scalar. Of course, R^n is an n-dimensional real vector space.
- **Arrows in a plane form a two-dimensional real vector space**. We know that arrows in a line form a real scalar set. Arrows in a plane is the Cartesian product of two sets of arrows, one set of arrows in one line, and the other set of arrows in another line. Consequently, arrows in a plane form a two-dimensional real vector space.
- **Add two arrows in a plane**. Let u and v be two arrows in a plane. Translate the arrow v so that the tail of v coincides with the head of u. The arrow from the tail of u to the head of v is u + v.
- **Multiply an arrow in a plane by a real number**. Let u be an arrow in a plane, and c be a real number. cu means an arrow, of length c times the length of u, in parallel with u.

- **To boldface or not to**? Our textbook use boldfaced letters to represent vectors.
- Many people follow this practice.
- Many other people don't.
- This practice is impossible to follow consistently.
- We have already seen that a scalar set is just a one-dimensional vector space. Do we boldface scalars?
- Even a number field is a one-dimensional vector space. Do we boldface numbers?
- We will also see that matrices of the same size form a vector space. Do we boldface matrices?
- If the answer is yes in all these cases, then all things in linear algebra are boldfaced.
- We will just create an ugly text but differentiate nothing.
- In class, I'll be explicit about the nature of each symbol, and will not boldface vectors, or put arrows over vectors. They look pretentious and distracting, and serve no purpose.
- In this google doc, I use boldface to draw your attention to important words.
- You decide what you like, boldface or not.
- You don't even need be consistent, so long as you are clear what type of things you are dealing with.
- Use the words numbers, scalars, vectors frequently.

Lay 1.3, 1.4, 1.8. A system of linear algebraic equations in alternative forms

- **One system, many alternative forms**.
- We have just considerably expanded our mathematical vocabulary.
- Let's use this expanded vocabulary to talk about familiar things.
- That is, let's use familiar things to illustrate the new vocabulary.
- Consider the problem of chickens and rabbits again. A farm has chickens and rabbits. The farmer counts 26 heads and 82 feet. How many chickens and rabbits are in the farm?

- **Write the linear map in alternative forms**.
- One simple idea, written in many ways.
- In linear algebra, the same linear map is written in many alternative forms, as listed below.
- All these alternative forms mean the same thing as the two scalar equations above.
- You must learn to translate one form to another.

- **A story of four scalar sets and four scalar-scalar linear maps**
- Let x_1 be the number of chickens, x_2 be the number of rabbits, y_1 be the number of heads, and y_2 be the number of feet.
- The equation of heads: $y_1 = x_1 + x_2$.
- The equation of feet: $y_2 = 2x_1 + 4x_2$.
- This story involves four **real scalar sets**: the chicken set, rabbit set, head set, foot set.
- Also note the four **scalar-scalar linear maps**: each chicken has 1 head, each chicken has 2 feet, each rabbit has 1 head, and each rabbit has 4 feet.

- The two **scalar equations** completely describe the problem.
- The row reduction algorithm can determine the solution set of any system of linear algebraic equations.

- But somehow people have developed a large vocabulary to paraphrase the problem.
- We next learn this vocabulary.
- Of course, we will also use this vocabulary later to describe things other than solving systems of equations.

- **List the data by a table**.
- Row 1 lists the numbers of heads for a chicken, a rabbit, and the farm.
- Row 2 lists the numbers of feet for a chicken, a rabbit, and the farm.
- The table is simply the **augmented matrix** of the system.
- We have already learned how to solve this system using the row reduction algorithm.

	chicken	rabbit	farm
head	1	1	26
foot	2	4	82

- **Scalars.** The system involves four real scalar sets: the chicken set, rabbit set, head set, and foot set.
- **Scalar-scalar linear maps**. Our worldly knowledge supplies four scalar-scalar linear maps.
- Chicken-head linear map: each chicken has one head.
- Chicken-foot linear map: each chicken has two feet.
- Rabbit-head linear map: each rabbit has one head.
- Rabbit-feet linear map: each rabbit has two feet.
- These four scalar-scalar linear maps appear as the entries of the **matrix of coefficients**.
- The total number of heads and feet on the farm appear as the **constant terms.**

- **Two scalar equations**.
- Each row lists things in one scalar set, and corresponds to an equation of scalars. Let x_1 be the number of chickens, and x_2 be the number of rabbits.
- One equation is for heads: $x_1 + x_2 = 26$**.**
- Another equation is for feet: $2x_1 + 4x_2 = 82$**.**
- These equations are familiar to us, but we can now identify four scalar sets and four scalar-scalar linear maps.

- We can plot the two equations in the **chicken-rabbit space**.
- This two-dimensional real vector space is the Cartesian product of two real scalar sets:
- **(chicken-rabbit space) = (chicken set) x (rabbit set)**.

- **One vector equation**.

- The augmented matrix has three columns.
- Each column in the augmented matrix lists things in different scalar sets.
- Each column is a vector in the head-foot space.
- This two-dimensional real vector space is the Cartesian product of two real scalar sets:
- **(head-foot space) = (head set) x (foot set)**.
- In head-foot space, each chicken is a vector, described by an ordered pair
- a_1 = (1 head, 2 feet).
- We also write this vector as a column,

1 head
2 feet

- Each rabbit is another vector, described by an ordered pair
- a_2 = (1 head, 4 feet).
- We also write this vector as a column,

1 head
4 feet

- The constant terms give the third vector, b = (26 heads, 82 feet).
- We also write this vector as a column,

26 head
82 feet

- These three columns appear in the augmented matrix.
- Let x_1 be the number of chickens, and x_2 be the number of rabbits.
- We now write the system of scalar equations as a single vector equation:
- $x_1a_1 + x_2a_2 = b$.
- In this equation, each column is a vector in the head-foot space, and x_1 and x_2 are numbers (i.e., magnitudes of the two scalars, the number of chickens and the number of rabbits).
- We do not add heads to feet.
- But we add head-foot vectors.
- That is, we add entries in columns in parallel.

- **Is a chicken a scalar or a vector**?
- The answer depends on which set you put the chicken in.
- A chicken is a scalar in the chicken set, in which each element is a certain number of chickens.
- The chicken has one head and two feet, and is a one-head, two-feet vector in the head-foot space.

- **One matrix equation**.

- As another alternative, we regard the ordered pair (x_1, x_2) as a vector in the chicken-rabbit space.
- Denote the vector by a column x.
- Denote the matrix of coefficient by A.
- The system now becomes $Ax = b$.
- The rule for matrix-column multiplication is such that this matrix equation recovers the two scalar equations.
- Go over this multiplication in class.

- **Linear map from one vector space to another vector space.**
- Vector-vector linear map.
- As noted above, this system involves four scalar sets and four scalar-scalar linear maps.
- We have just defined the Cartesian product two scalar sets as the chicken-rabbit vector space, and defined the Cartesian product of the other two scalar sets as the head-foot vector space.
- The matrix of coefficients A has four entries, each being a scalar-scalar linear map.
- We now regard the matrix A as a vector-vector linear map.
- That is, A is a linear map from the chicken-rabbit vector space into the head-foot vector space.
- Column 1 of A maps the vector (1 chicken, 0 rabbit) in the chicken-rabbit space to the vector (1 head, 2 feet) in the head-foot space.
- Column 2 of A maps the vector (0 chicken, 1 rabbit) in the chicken-rabbit space to the vector (1 head, 4 feet) in the head-foot space.
- We now view the system as a vector-vector linear map, $Ax = b$.

- **A linear map from the chicken-rabbit space to the head-foot space.** These two scalar equations define a **linear map** from an ordered pair of numbers, (x_1, x_2), to another ordered pair of numbers, (y_1, y_2).
- $y_1 = x_1 + x_2$,
- $y_2 = 2x_1 + 4x_2$.
- These two equations define a **map**.
- That is, given the numbers of chickens and rabbits, (x_1, x_2), the equations calculate the numbers of heads and feet, (y_1, y_2).
- This map is **linear,** because both equations are linear.
- In the above, the linear map is defined by two scalar equations.
- The system of scalar equations is perhaps the most familiar form to us.

- **A equation of columns (or tuples)**
- The ordered pair (x_1, x_2) is a vector in the chicken-rabbit space, which is a two-dimensional real vector space.
- The ordered pair (y_1, y_2) is a vector in the head-foot space, which is another two-dimensional real vector space.
- The chicken-rabbit space is the **domain** of the linear map.
- The head-foot space is the **codomain** of the linear map.

- We say that **the linear map sends a vector in the domain to a vector in the codomain**.
- Write the above two equations (i.e., the linear map) as a single **vector equation**:
- $(y_1, y_2) = x_1(1,2) + x_2(1,4)$.
- Here we write each vector as an ordered pair. In class we wrote each vector as a column.

- **Vector-vector linear map**
- Each entry of the matrix means a linear map from one scalar set to another scalar set.
- Write the linear map in yet another way.
- $T: R^2 \rightarrow R^2$.
- That is, $T(x_1, x_2) = (x_1 + x_2, 2x_1 + 4x_2)$.
- Thus, T(11,15) = (11 + 15, 2x11 + 4x15) = (26, 82).
- We can also write the right side of the above equation as a column of two entries.
- Remember that all these alternative forms mean the same as the two scalar equations we began with. We learn these alternative forms because other people use them.

- The writing can get even more **abstract**.
- Denote the ordered pair (x_1, x_2) by another letter x, and call x a vector in the domain. Denote the ordered pair (y_1, y_2) by a single letter, y, and call y a vector in the codomain. Write the linear map as $y = Tx$.
- Designate the domain by X, and the codomain by Y.
- Write the linear map as $T: X \rightarrow Y$.

Graphical representation of a vector-vector linear map

- Video on row picture and column picture of a system of equations.
- The graphical representation works well when both the domain and the codomain are two-dimensional real vector spaces.
- Represent the chicken-rabbit space as a plane with two axes.
- Represent the head-foot space as another plane with two axes.
- For every point (x_1, x_2) in the chicken-rabbit plane, the linear map (i.e., the two linear equations) calculates a point (y_1, y_2) in the head-foot plane.
- In class, we mapped several points from the chicken-rabbit space to the head-foot space.
- In particular, the linear map sends a **rectangle** in the chicken-rabbit plane to a **parallelogram** in the head-foot plane.
- For more graphical representations of linear maps, see **Tables 1-4 starting on p. 73 of the textbook by Lay.**
- For the system of **chickens and rabbits**, the linear map sends the chicken-rabbit plane to the head-foot plane.
- Now look at the system of **rabbits and hamsters.** In this case, the linear map sends the rabbit-hamster plane to a line in the head-foot plane. Think about rabbits and hamsters, and you will see why.
- What happens to the system of **chickens, rabbits, and hamsters**? Write the two scalar equations that define the linear map from the chicken-rabbit-hamster space to the head-foot space. The linear map sends the three-dimensional real vector space to the two-dimensional real vector space.

Week 3 (February 6, 8, 10)

Reading and watching

- Video on linear combination, span, and basis
- Video on solving homogeneous equation Ax = 0.
- Video on solving inhomogeneous Ax = b.
- Yet another video on solving Ax = b.
- Lay 1.5 Solution set of three types
- Lay 1.3 Vector equation
- Lay 1.7 Linear independence

Lay 1.5 Solution set written in a parametric vector form

- **Vector-vector linear map Ax = b**.
- This is one of several alternative forms of the same system.
- x is a vector in the vector space R^n, and each entry in the column x is a scalar.
- b is a vector in the vector space R^m, and each entry in the column b is a scalar.
- A is a matrix of m rows and n columns, and each entry in the matrix A is a scalar-scalar linear map.
- We call A the matrix of the vector-vector linear map, or simply the best linear map.

- **Write solution set to the system Ax = b as a single equation in terms of vectors in space** R^n, and regard the free variables as parameters.
- In class I use an example to illustrate the steps.

1. Reduce the the augmented matrix [A b] to its reduced row echelon form (rref).
2. In the rref, identify nonpivot columns. Each nonpivot column corresponds to a free variable.
3. Translate the rref back to a system of equations.
4. Move the free variables to the right side.
5. For each free variable x_f, insert an identity $x_f = x_f$ into the system of equations.
6. Translate the equations back to a single equation in terms of vectors in space R^n.

- Any choice of values of the free variables gives a solution to the system. Each free variable is a parameter.
- The above steps give the **full solution set** of the system Ax = b in the **parametric vector form**:
- $x = p + c_1v_1 + \dots + c_kv_k$.
- Here k is the number of free variables,
- p is a **particular solution** to the system Ax = b,
- $c_1 \dots c_k$ are arbitrary numbers, called **parameters**,
- $v_1 \dots v_k$ are **homogeneous solutions** to Ax = 0.

- **Homogeneous equation**. The matrix equation Ax = 0 is called a homogeneous equation.

- A **homogeneous solution** v is a solution that satisfies Av = 0.
- The solution v = 0 is called the **trivial solution**.
- We find the set of all homogeneous solutions by the row reduction algorithm.

- **Nonhomogeneous equation**. The matrix equation Ax = b is called a nonhomogeneous equation.
- A **particular solution** p is a solution that satisfies the equation, Ap = b.
- **The full solution set of the nonhomogeneous equation Ax = b** is the sum of one particular solution to Ax = b and all solutions of the homogeneous equation Ax = 0: $x = p + c_1v_1 + \ldots + c_kv_k$.

- Watch this video for an example of solving Ax = b.
- Another video on solving Ax = b.
- Yet another video on solving Ax = b.

Head set, foot set, and head-foot space

- Let's illustrate the abstract, mathematical language using concrete, everyday language.
- Recall the chicken-rabbit system:

	chicken	rabbit	farm
head	1	1	26
foot	2	4	82

- **Head set** = {1 head, 26 heads,...}, modeled as a real scalar set.
- **Foot set** = {2 feet, 4 feet, 82 feet,...}, modeled as another real scalar set.
- **Head-foot space** is the Cartesian product of the head set and the foot set:
- **(head-foot space) = (head set) x (foot set)**.
- Head-foot space is a two-dimensional real vector space.
- Some head-foot vectors: (1 head, 2 feet), (1 head,4 feet), (2 heads, 6 feet).
- Each column of the augmented matrix lists a **vector** in the head-foot space:
- a chicken = (1 head, 2 feet),
- a rabbit = (1 head, 4 feet), and
- the farm = (26 heads, 82 feet).
- The head-foot space is **two-dimensional** because each vector is a tuple of two scalars: (a number of heads, and a number of feet).
- The head-foot space is **real** because a head-foot vector multiplied by a **real number** is still a head-foot vector. For example, 3 chickens = 3(1 head, 2 feet) = (3 head, 6 feet). It makes no sense to use complex numbers in this context.

- **Numerical representation of a vector.**
- For each scalar set, choose a nonzero scalar as a unit.
- Then, every scalar in the scalar set is a real number times the unit.

- Choose the scalar (1 head) as a unit for the head set.
- Choose the scalar (1 foot) as a unit for the foot set.
- Write the vector (1 head, 2 feet) as (1,2), the vector (1 head, 4 feet) as (1,4), and the vector (26 heads, 82 feet) as (26, 82).
- We say that (1,2) is a **numerical representation** of the vector (1 head, 2 feet).

- **Enxhi Buxheli's question**: **What if we make a different choice of unit for a scalar set**? For example, choose (1 head) as a unit for the head set, and (2 feet) as a unit for the foot set.
- **Answer**. We will then write the vector (1 head, 2 feet) as (1,1), the vector (1 head, 4 feet) as (1,2), and the vector (26 heads, 82 feet) as (26, 41).
- **Given a vector, its numerical representation depends on the choice of units for the scalar sets**. For example, for the two alternative choices of the unit for the foot set, the same vector (26 heads, 82 feet) has different numerical representations: (26, 82) and (26, 41).

- Once we chose a unit for each scalar set, every vector in the head-foot space corresponds (by a one-one correspondent linear map, i.e., an **isomorphism**) to an ordered pair of real numbers. We say that the head-foot space is **isomorphic** to R^2, and write (head-foot space) ~ R^2.
- In general, **any n-dimensional real vector space is isomorphic to R^n**. Each choice of units for the scalar sets defines a one-one correspondent linear map (i.e., an isomorphism). The isomorphism allows our textbook to focus almost exclusively on a particular n-dimensional real vector space, R^n, even though almost no applications really happen in this space. They happen in places like head-foot space, gold-silver space, and spacetime.

Isomorphism

- Two vector spaces V and W over a number field F are called **isomorphic** if there exists a bijective (one-one correspondent) linear map between them.
- In Greek, iso means "the same", and morph means "form".
- From the index of the textbook, you can find several locations the concept appears.
- All vector spaces, of the same dimension, over the same number field, are isomorphic.
- Real vector spaces of different dimensions are not isomorphic.
- A real vector space and a complex vector space are not isomorphic.

Graphical representation of a two-dimensional vector space

- The head-foot space is a two-dimensional real vector space.
- Arrows in a plane form another two-dimensional vector space.
- A head-foot vector is different from an arrow in a plane.
- But all two-dimensional real vector spaces are isomorphic.

- We can create a linear bijection between the head-foot space and the arrows in a plane.
- This linear bijection sends every head-foot vector to one and only one arrow in a plane.
- This linear bijection is called a **graphical representation** of the head-foot space.
- Here are the steps to create this linear bijection.
- In a plane, represent a unit for the head set, (1 head), as an arrow, called the 1-head arrow, or the unit head arrow.
- The length and the direction of this arrow are arbitrary.
- For any real number c, the scalar (c heads) corresponds to an arrow parallel to the 1-head arrow, and of length c times the 1-head arrow.
- Next represent a unit for the foot set, (1 foot), as another arrow, called the 1-foot arrow, or the unit foot arrow.
- The length and the direction of this arrow are also arbitrary, so long as the 1-head arrow and the 1-foot arrow are in **different directions**.
- For any real number b, the scalar (b feet) corresponds to an arrow parallel to the 1-foot arrow, and of length b times the 1-foot arrow.
- Parallel to the 1-head arrow draw a set of lines, called the head lines.
- Parallel to the 1-foot arrow draw another set of lines, called the foot lines.
- The two sets of lines form a grid, called **a system of coordinates** of the head-foot space.
- Pick any point in the plane as the origin, representing the zero vector, (0 head, 0 foot).
- Each head-foot vector corresponds to a point in the plane, and to the arrow from the origin to the point.
- In the head-foot plane, mark the three vectors: a chicken, a rabbit, and the farm.
- We can also mark a hamster, a snake, and an ant.

- **Lucas Guzman's question (2017): does the 1-head arrow have to be of the same length as the 1-foot arrow**?
- Answer: no. Here is why.

- **Ignore distractions. Examine your prejudices.**
- The set of arrows in a plane appeals to our intuition.
- The set is a graphical illustration of the general concept of vector space.
- Like any concrete example of a general concept, however, this example has distracting features (i.e., structures) that do not belong to the general concept of vector space.
- For example, the definition of vector space does not let us compare the lengths of the 1-head arrow and the 1-foot arrow, or calculate the angle between the two arrows.
- There is no reason to compare the length of the arrow representing a head and the length of the arrow representing a foot.
- There is no sense to talk about the angle between a head and a foot.
- Getting rid of prejudices liberates us, enabling much broader applications of linear algebra.
- Indeed, the vast majority of applications of vector spaces have nothing to do with arrows, lengths, and angles.

- Most applications deal with chickens, rabbits, and other commodities. They are merely scalars.
- A vector space is merely the Cartesian product of scalar sets. Perhaps a lot of them.

- **Do not let graphics blind our algebraic vision.**
- Seeing is believing.
- Seeing is deceiving.
- All two-dimensional real vector spaces are isomorphic to the set of arrows in a plane.
- We use this isomorphism to represent any two-dimensional real vector space by arrows in a plane.
- This graphical representation is perhaps the most visible representation of vector spaces, but is **limited to only two-dimensional real vector space**. (Henry Ford was reported to say, you can order a Model T of any color, so long as it is black.)
- We deceive ourselves by drawing **fake diagrams** to represent three-dimensional objects in two dimensions.
- Draw the fake diagrams if they are useful to you.
- Ignore fake diagrams if they confuse.
- Build a three-dimensional physical objects if you wish to see, and feel, the shape in three dimensions.
- No need to imagine four dimensions geometrically.
- Contrary to the popular myth, Einstein did not talk about four-dimensional geometry in his 1905 paper on relativity.
- By comparison, **the numerical representation is applicable to vector spaces of any dimension, over any number field**.
- Computers represent vectors by numbers, not arrows.

Linear combination

- Let $v_1,...,v_p$ be vectors in an n-dimensional vector space V over a number field F.
- Let $c_1,...c_p$ be numbers in F.
- $c_1v_1 + \dots + c_pv_p$ is a vector, and is called a **linear combination** of the vectors $v_1,...,v_p$ with **weights** $c_1,...c_p$.

- You of course recognize the equation
- 11(1 head, 2 feet) + 15(1 head, 4 feet) = (26 heads, 82 feet).
- We interpret this equation In the head-foot space.
- (1 head, 2 feet) is a head-foot vector corresponding to a chicken
- (1 head, 4 feet) is a head-foot vector corresponding to a rabbit
- (26 heads, 82 feet) is a head-foot vector corresponding to the farm.

- Graphical representation of the linear combination 11(1 head, 2 feet) + 15(1 head, 4 feet) = (26 heads, 82 feet).
- **Examples of linear combinations**. A linear combination of (2 chickens, 3 rabbits) and (2 chickens, 5 rabbits) is

- 2(2 chickens, 3 rabbits) + 3(2 chickens, 5 rabbits) = (10 chickens, 21 rabbits).
- This equation looks cleaner when we write each vector as a column.

Span

- Let $v_1,...,v_p$ be vectors in an n-dimensional vector space V over a number field F. The span of these vectors, written Span $(v_1,...,v_p)$, is the collection of all vectors of the form $c_1v_1 + \ldots + c_pv_p$, where $c_1,...c_p$ are any numbers in F.
- Thus, a linear combination of vectors produces a vector, but the span of vectors produces a **set of vectors**.
- **Span (u)**. Let u be a nonzero vector in an n-dimensional real vector space V. The **span** of u, written Span (u), is the collection of all vectors of the form cu, where c is any real number.
- Span (u) is a one-dimensional vector space, i.e., a scalar set. u serves as a unit for the scalar set Span (u). All elements in Span (u) are proportional to each other.
- **Example of Span (u)**. Consider the two-dimensional real vector space, the head-foot space. Consider a vector in the head-foot space, u = (1 head, 2 feet). 3u = (3 head, 6 feet). Span (u) is the set of all vectors of the form cu, where c is any real number. in this example, u represents a chicken, and Span (u) represents the chicken set.
- Span (u) corresponds (is isomorphic) to arrows in a line. **The vector u is said to span the line**.
- **Is Big Mac a vector or a scalar**? A Big Mac is a vector in the vector space of all food, but is a scalar in the scalar set in which each element is a certain number of Big Macs.
- **Span (u,v)**. Let u and v be two vectors in an n-dimensional real vector space V. The span of the two vectors, written Span (u,v), is the collection of all vectors of the form au + bv, where a and b are any real numbers.
- If u and v are proportional, then Span (u,v) = Span (u) = Span (v), and they correspond (is isomorphic) to all arrows in a line.
- If u and v are disproportional, then Span (u,v) corresponds (is isomorphic to) all arrows in a plane. **The vectors u and v are said to span the plane**.
- **Example of Span (u,v)**. Consider the three-dimensional real vector space, R^3. Consider two vectors, u = (1,2,3) and v = (4,5,6). Thus, Span (u,v) is the collection of all vectors of the form a(1,2,3) + b(4,5,6), where a and b are arbitrary real numbers.
- **A question about Span ($v_1,...,v_p$)**. Given vectors $v_1,...,v_p$ and b in an n-dimensional real vector space V, is b in Span $(v_1,...,v_p)$? This question is the same as asking, does the equation $x_1v_1 + ... + x_pv_p = b$ have a solution? We answer this question by the row reduction algorithm.

Lay 1.7 Linear independence

- Let V be an n-dimensional vector space over a number field F. A list of vectors $v_1,...,v_p$ in V is said to be **linearly independent** if, for any numbers $x_1,..., x_p$ in F, $x_1v_1 + ... + x_pv_p = 0$ implies that $x_1 = ... = x_p = 0$.
- **Linear dependence**. Vectors $v_1,...,v_p$ in V are said to be linearly dependent if there exist numbers $x_1,..., x_p$ in F, not all zero, such that $x_1v_1 + ... + x_pv_p = 0$.

- Ways to ascertain linear dependence are listed below.
- A single nonzero vector is linearly independent.
- A list of vectors containing the zero vector is linearly dependent.
- Two vectors u and v are linearly independent if they are disproportional. That is, for any number c in F, u = cv does not hold.
- If p > n, the list of vectors are linearly dependent.
- In general, **given a list of vectors (v_1,...,v_p), we ascertain its linear independence by the row reduction algorithm**. We collect the vectors in the list as the columns of a matrix, A = [v_1,...,v_p].
- The task of ascertaining the linear independence of the list (v_1,...,v_p) is the same as asking if the equation Ax = 0 has any nontrivial solution.
- If every column of A is a pivot column, the list of the vectors is linearly independent.
- If some columns are nonpivot columns, the list of vectors are linearly dependent.
- In class, I use an example to illustrate this general procedure.

Week 4 (February 13, 15, 17)

Reading and watching

- Video on linear map
- Video on linear map between vector spaces of different dimensions
- Lay 1.6 Applications of linear systems.
- Lay 1.10 Linear models in business, science, and engineering.
- Lay 1.8 Linear map (linear transformation)
- Lay 1.9 The matrix of a linear map
- Lay 2.1 Matrix operations

Lay 1.6, 1.10 Applications of linear equations

- These two sections give excellent examples of setting up equations from real-world problems.
- We will not go over these two sections in lectures.
- Read the sections and work on supplementary problem set I.
- **Exams will draw upon these problems**.

- **An algorithm to set up equations from a real-world problem.**
- All real scalars are isomorphic.
- All applications of linear equations are “isomorphic”, more or less, to the chicken-rabbit problem.
- We list the algorithm to set up equations from a real-world problem in following bullets.

- Identify n scalar sets like chicken set and rabbit set.
- Idensity m scalar sets like head set and foot set.
- Identify m x n scalar-scalar linear maps like chicken-to-head map, chicken-to-foot map, rabbit-to-head map, and rabbit-to-foot map.
- Tabulate data as an augmented matrix.

- **Solving these problems requires no knowledge of vectors**.
- The knowledge of vectors and vector-vector linear maps was unnecessary to the inventors of the row reduction algorithm 2000 years ago.
- It is still unnecessary to us today in solving similar problems.
- We were fully ready to solve these problems after we learned the row reduction algorithm in Lay 1.2.
- What matters is that many things in the world are scalar sets and scalar-scalar linear maps.
- We next go through several "real-world problems".

- **Input-output economics**. Wassily Leontief, of Harvard University, won the 1973 Nobel Prize in Economic Science for his work on input-output economics. Watch a video of Leontief explaining the input-output table.
- In the example in Lay 1.6, the economy is divided into three sectors: coal, electric, and steel. The output of each sector becomes the input of all sectors. For example, the 40%, 10%, and 50% of the output of the electric are used to make coal, electric, and steel, as listed in a column under "electric".

	coal	electric	steel
coal	0.0	0.4	0.6
electric	0.6	0.1	0.2
steel	0.4	0.5	0.2

- Question: Determine the prices for coal, electric, and steel, P_c, P_e, and P_s, that make the income of each sector matches the expenditure of the sector.
- Income-expenditure match for coal: $P_c = 0P_c + 0.4P_e + 0.6P_s$
- Income-expenditure match for electric: $P_e = 0.6P_c + 0.1P_e + 0.2P_s$
- Income-expenditure match for steel: $P_s = 0.4P_c + 0.5P_e + 0.2P_s$
- This problem gives a system of three **homogeneous equations** for three variables. We look for a **nontrivial solution**. Looking at the equations, we note that the above table is not the augmented matrix yet. Some modification is needed to obtain the augmented matrix. Once the augmented matrix is obtained, the row reduction algorithm will produce the the solution set.

- **Balancing chemical equations**. Balance the following chemical equation:
- $C_3H_8 + O_2 = CO_2 + H_2O$

	C_3H_8	O_2	CO_2	H_2O
Atom C	3	0	-1	-0
Atom H	8	0	-0	-2
Atom O	0	2	-2	-1

- The principle of chemical reaction was discovered about 200 hundred years ago. A chemical reaction changes molecules, but does not change atoms. That is, **in a chemical reaction, the number of each atom remains unchanged**.
- People have to discover the atomic composition of each molecule. This information is embodied in the **chemical formula** of the molecule. For example, a molecule of C_3H_8 contains three carbon atoms and eight hydrogen atoms.
- The reaction then corresponds to a table. We put a negative sign in front of each molecule on the right side. This table is the matrix of coefficients for a system of **homogeneous equations**.
- Individual molecules are analogous to chicken and rabbit. Individual atoms are analogous to head and foot.
- A water molecule has one oxygen atom and two hydrogen atoms, and is written H_2O. A chicken has one head and two feet, and may be written $(\text{foot})_2(\text{head})$.
- A methane molecule has one carbon atom and four hydrogen atoms, and is written CH_4. A rabbit has one head and four feet, and may be written $(\text{head})(\text{foot})_4$.

- **Network flow**. Looking at the flow diagram in Fig. 2 on p. 52 of Lay, we can directly construct the augmented matrix:

	x_1	x_2	x_3	x_4	x_5	
A	1	1	0	0	0	800
B		-1	1	-1	0	-300
C	0	0	0	-1	-1	-500
D	-1	0	0	0	-1	-600

- **Construct a diet to supply certain amounts of nutrients**. You wish to design a diet using three ingredients: nonfat milk, soy flour, and whey.
- The object of the design is to find a diet that supplies certain amounts of nutrients: protein, carbohydrate, and fat.

- Listed in the table are grams of nutrients supplied by 100 grams of each ingredient, along with the grams of nutrients in the diet. How much nonfat milk, soy flour, and whey should the diet use?
- The table is exactly the augmented matrix.
- Individual ingredients are analogous to chicken and and rabbit. Individual nutrients are analogous to head and foot.

	nonfat milk	soy flour	whey	diet
protein	36	51	13	33
carbohydrate	52	34	76	45
fat	0	7	1.1	3

- **Electrical networks.** A few laws of physics: voltage V forms a scalar set, current I forms another scalar set, Ohm's law is V = RI, where R is the resistance. Thus, R is a linear map from current to voltage.
- **Kirchhoff's voltage law**. The algebraic sum of the RI voltage drops in one direction around a loop equals the algebraic sum of the **voltage sources** in the same direction around the loop.
- Thus, Kirchhoff's voltage law is another way to say that "voltage forms a scalar set, and every point in the circuit has a unique value of voltage".
- For the electrical network on p. 82 of Lay, the network consists of three **closed loops**. Let I_1, I_2, I_3 be the three **loop currents**. Translate the schematic of the network to an augmented matrix. Each row corresponds to the application of Kirchhoff's voltage law to one closed loop.

	I_1	I_2	I_3	voltage source
loop 1	11	-3	0	30
loop 2	-3	6	-1	5
loop 3	0	-1	3	-25

- **City-suburb migration.** Each year, 5% of city people move to the suburbs, 95% of city people remain in the city, 3% suburbans move to the city, and 97% suburbans remain in the suburbs.
- Given the number of city people and the number of suburbans in one year, predict these numbers in subsequent years.
- Let x_c and x_s be the numbers of city people suburbans in one year, and y_c and y_s be the numbers of city people and suburbans in the next year.

- The number of city people next year: $y_c = 0.95x_c + 0.3x_s$.
- The number of suburbans next year: $y_s = 0.5x_c + 0.97x_s$.
- This pair of equations let us predict the future migration.

Lay 1.8, 1.9. Linear map (i.e., linear transformation) from one vector space to another vector space

- **Consider a general linear map T: $R^n \to R^m$.** This linear map is represented by a matrix of m rows and n columns.
- The domain of the linear map is an n-dimensional real vector space, $(x_1,..x_n)$.
- The codomain of the linear map is an m-dimensional real vector space, $(y_1,..y_m)$.
- **Example**. Let T(x,y,z) = (x + 2y, 3y + 4z). This equation defines a linear map T: $R^3 \to R^2$. Write the matrix of the linear map. Write the map as a system of equations. Write the map as a single vector equation.
- This linear map sends a 3-tuple of real numbers to an ordered pair of numbers. For example, T(1,3,5) = (1 + 2x3, 3x3 + 4x5) = (7,29).

- **Review the basic mathematical terms.** Before studying linear maps, you should know **maps**. You should have mastered the concept of maps in high school. You may or may not have heard of a few terms related to maps. These terms are listed in the section above, titled "**A few general mathematical terms**".

- **A formal definition of linear map**.
- In linear algebra, we focus on maps of a particular type:
- The **domain** is an n-dimensional vector space V over a number field F.
- The **codomain** is an m-dimensional vector space W over the same number field F.
- A map T: V → W is called a **linear map** if
- T(au + bv) = aT(u) + bT(v)
- for any vectors u and v in V and any numbers a and b in F.

- A linear map T preserves linear combination.
- That is, $T(c_1v_1 + \dots + c_pv_p) = c_1T(v_1) + \dots + c_pT(v_p)$ for any vectors $v_1,...,v_p$ in V and any numbers $c_1,...,c_p$ in F.

- The vector space V is just the Cartesian product of n scalar sets.
- The vector space W is just the Cartesian product of m scalar sets.
- The vector-vector linear map T is just an m x n table of scalar-scalar linear maps
- **But many scalar sets tell stories that cannot be told by one or two scalar sets**.

- **Lay pp. 75-77. Onto, one-to-one**.
- A map T: X → Y is called **onto (subjective)** if at least one solution **exists** to the equation T(x) = y for **every** y in Y.
- **Theorem.** A linear map is onto if and only if every row of its matrix is a pivot row.

- A map T: X → Y is called **one-to-one (one-one, 1-1, injective)** if the solution to the equation T(x) = y is **unique** whenever the solution exists.
- **Theorem.** A linear map is one-to-one if and only if every column of its matrix is a pivot column.
- A map T: X → Y is called **bijective (one-to-one correspondent)** if it is both onto and one-to-one**.** That is, the map is bijective if the equation T(x) = y has a **unique** solution x in X for **every** y in Y.
- **Theorem.** A linear map is bijective if and only if its matrix is a square matrix, every row is a pivot row, and every column is a pivot column. That is, the rref of the matrix is an identity matrix.

- **Allison Lee's question**. **Do we need to remember these legalistic definitions of onto and one-to-one**?
- Yes. These definitions are short, but significant.

- Starting from Lecture 1, we have been building up the ideas of **onto** and **one-to-one** in various ways.
- Here are some old examples.
- There **exists** a solution to the nonhomogeneous matrix equation Ax = b for every b if and only if every row of A is a pivot row.
- The homogeneous matrix equation Ax = 0 has a **unique** solution (i.e., the trivial solution x = 0) if and only if every column of A is a pivot column.
- Any nonzero linear map from one scalar set to another scalar set is bijective, and is therefore both onto and one-to-one.
- Any two vector spaces, of the same dimension, over the same number field, are **isomorphic**. That is, there exists a bijection between such two vector spaces.
- For example, **all two-dimensional real vector spaces are isomorphic**. (Chicken-rabbit space) ~ (head-foot space) ~ (rabbit-hamster space) ~ (gold-silver space) ~ (arrows in a plane) ~ R^2.

- **Examples of surjection, injection, and bijection**.
- **Chickens and rabbits**. The linear map from chicken-rabbit space to the head-foot space is both onto and one-to-one (i.e., is bijective). Given any number of heads and any number of feet, (y_1, y_2), we can find a unique solution to the number of chickens and the number of rabbits, (x_1, x_2).
- **Rabbits and hamsters**. The linear map from the rabbit-hamster space to the head-foot space is neither onto nor one-to-one. Not onto: the linear map sends the rabbit-hamster place to a line in the head-foot plane. Not one-to-one: the linear map sends multiple points in the rabbit-hamster space to one point in the head-foot space. For example, both one rabbit and one hamster have one head and four feet. The linear map sends the two points (1,0) and (0,1) in the rabbit-hamster space to one point (1,4) in the head-foot space.

- **Chickens, rabbits, and hamsters**. The linear map from chicken-rabbit-hamster space to the head-foot space is onto, but not one-to-one. We went through this example in detail in class. Make sure you study the notes.
- **Example 4 on p. 76 of Lay.** This example asks the same questions as we asked in class.

- **General procedure to find if a linear map is a surjection (onto), injection (one-to-one), or bijection**. The steps are as follows.
- Given a linear map T: $R^n \rightarrow R^m$, write the matrix A of the linear map T.
- Use the row reduction algorithm to obtain rref (A).
- If every row is a pivot row, T is surjective. Write an expression to show that there **exists** at least one solution x in X to the equation T(x) = y for every y in Y. For example, you can set all free variables to be zero.
- If every column is a pivot column, T is injective. If some columns are nonpivot columns, T is not injective. The corresponding free variables give many solutions to the homogeneous equation T(x) = 0.
- If A is a square matrix, every row is a pivot row, and every column is a pivot column, then T is bijective.

High-level view of Chapter 1

- **Solving a real-world problem in two steps**.

- **Step 1**: Translate the real-world problem to a system of linear algebraic equations.
- Tools: Scalar set, scalar-scalar linear map.
- Examples: chickens, rabbits, and hamsters. Lay 1.6 and 1.10.

- **Step 2**: Solve the system of linear algebraic equations.
- Tools: Row reduction algorithm, translation from rref to solution set. Matlab.
- Examples: Lay 1.1, 1.2, and 1.5.

- **Fundamentals.**

- A system of m linear algebraic equations in n variables
- Row reduction algorithm
- Solution set

- Number field
- Scalar set
- Scalar-scalar linear map

- Vector space

- Vector-vector linear map

- **Algebra is an art to structure sets**.
- Three algebraic structures:
- number field
- scalar set
- vector space.
- Elements in these sets are called, respectively, numbers, scalars, and vectors.
- The structure of a set means operations that relate elements in the set.
- The three structures have different operations, even though they have the same names, addition and multiplication.
- For example, the number-number multiplication follows rules different from number-scalar multiplication.

- **Many things in the world form scalar sets**.
- Just think of water, commodities, money, chickens, feet, voltage, current, atoms, molecules…
- But not everything forms a scalar set. Happiness does not scale. Nor does love. Nor does intelligence. Scales of earthquakes do not form a scalar set.

- **Different scalar sets are often related through scalar-scalar linear maps**.
- Money buys commodities.
- A chicken has two feet.
- A water molecule has two hydrogen atoms.
- V = RI.
- Not all scalar-scalar maps are linear. For example, the map $f(x) = x^3$ is a nonlinear map from R to R.

- Thus, money forms a scalar set, gold forms another scalar set, and using money to buy gold creates a scalar-scalar linear map.
- The scalar sets and scalar-scalar linear maps lead to **systems of linear algebraic equations**.

- Solve linear equations using the **row reduction algorithm**.
- This algorithm converts augmented matrix to the **reduced row echelon form (rref)**.
- In practice, this hard labor is done by computers.
- In exams, you do row reduction by hand.
- The rref of the augmented matrix classifies **solution sets into three types**: no solution, unique solution, and many solutions.
- In the last case, **write the solution set in a parametric form**.

- **A vector is just shorthand for a tuple of scalars.**
- **A vector space is just the Cartesian product of scalar sets**.

- The number field and scalar set are two fundamental and powerful algebraic structures. They model much of our empirical experience.
- By contrast, a vector space is a derivative concept: it is the Cartesian product of scalars.
- We do not need vectors to solve systems of equations, or to classify solution sets. That problem was solved over 2000 years ago, by the invention of the row reduction algorithm. Relabeling the algorithm using the vocabulary of vectors add nothing but confusion to the algorithm.
- For some time, homework will consist of problems of two basic varieties: name calling and row reducing. The former is annoying; the latter, boring.
- If you are searching your soul for new insight, search no more.
- So far vectors do not tell you anything new, except for calling old things by new names, and writing old equations in alternative forms.
- Indeed, the idea of vector space contributes nothing until later in the course.

Lay 2.1 Matrix operations

- An m x n matrix A over a number field F is a table of m rows and n columns. Every entry of the table is a number in F.
- A single number is a 1 x 1 matrix.
- A single row is a 1 x n matrix.
- A single column is an m x 1 matrix.
- Up to this point in this course, the m x n matrix A has appeared in two places.
- **System of linear algebraic equations**. Given A and b, find the solution set to the equation Ax = b.
- **Linear map**. Given A and x, calculate y from y = Ax.
- Later in the course, the matrix will appear in other places.
- But for now, these two examples suffice for us to discuss **operations of matrices**.

- **Operation 1: matrix-matrix addition.** To be additive, two matrices A and B must have the same number of rows, m, and the same number of columns, n. The addition A + B results in a m x n matrix, with each entry being the sum of the corresponding entries of A and B.
- **Operation 2: number-matrix multiplication.** For any nonzero number c in a number field F, and any m x n matrix A over F, cA is a m-by-n matrix, with every entry being c times the corresponding entry of A.
- **Theorem**. The set of all m x n matrices over a number field F is an mn-dimensional vector space over F. Denote this set of matrices by $M_{m,n}(F)$.
- **Proof**. Each m x n matrix over F is simply an mn-tuple of numbers in F. Of course, in writing the matrix, we list the numbers by a table, not by a single row. But how we list numbers does not alter that the matrix is just an ordered list of numbers. The set of all such tuples is the Cartesian product of mn copies of F. Recall that F is a scalar set, and that the Cartesian product of any number of scalar sets is a vector space.

- **Operation 3: Matrix-matrix multiplication**
- x is a **row**, with entries $x_1,\ldots, x_m$.
- y is a **column**, with entries $y_1,\ldots, y_m$.
- The row x and the column y have the same number of entries.
- Define the **row-column multiplication** by $xy = x_1y_1+\ldots+x_my_m$. The row-column multiplication results in a number.
- Generalize this definition to matrix-matrix multiplication as follows.
- A is an l x m matrix.
- B is an m x n matrix.
- AB is an l x n matrix, defined as follows.
- (row i of A)(column j of B) = (element ij of AB)
- The matrix-matrix multiplication AB is defined if and only if the inner dimensions of A and B are the same.
- **Matrix-matrix multiplication corresponds to successive linear maps, also known as composition of linear maps.**

- A linear map from the chicken-rabbit space to the head-foot space: y = Bx. The matrix B represents the familiar knowledge: each chicken has a head and two feet, and each rabbit has one head and four feet.
- A head-and-foot worship tribe buys heads and feet with gold, silver, and platinum. A linear map from the head-foot space to the gold-silver-platinum space: z = Ay. The matrix A represents the prices for head and foot paid by gold, silver, and platinum.
- A **composite map**, from the chicken-rabbit space to the gold-silver-platinum space, is z = Ay = A(Bx) = (AB)x. The matrix AB represents the prices for chicken and rabbit paid by gold, silver, and platinum. In class we go over numbers for this example.

- **Why do matrices multiply in this way**?
- The rule of matrix-matrix multiplication is derived from successive linear maps.
- The textbook gives a proof using matrices and columns (p. 94).
- Here is an alternative proof using scalars.
- Consider two linear maps: y = Bx and z = Ay. In this proof, we will use the linear maps in the previous example.
- We now write the linear maps as scalar equations. (When in doubt, return to scalars.)
- The linear map y = Bx means two scalar equations:
- $y_1 = B_{11}x_1 + B_{12}x_2$
- $y_2 = B_{21}x_1 + B_{22}x_2$
- The linear map z = Ay means three scalar equations:
- $z_1 = A_{11}y_1 + A_{12}y_2$
- $z_2 = A_{21}y_1 + A_{22}y_2$
- $z_3 = A_{31}y_1 + A_{32}y_2$
- We see why the number of rows in B must equal to the number of columns in A. Both must match the number of ys.
- Successive linear maps z = Ay = A(Bx) mean substitution y_1 and y_2.

- For example, $z_1 = A_{11}y_1 + A_{12}y_2 = A_{11}(B_{11}x_1 + B_{12}x_2) + A_{12}(B_{21}x_1 + B_{22}x_2)$. Collecting terms, we obtain that
- $z_1 = (A_{11}B_{11} + A_{12}B_{21})x_1 + (A_{11}B_{12} + A_{12}B_{21})x_2$.
- Similarly, we obtain that
- $z_2 = (A_{21}B_{11} + A_{22}B_{21})x_1 + (A_{21}B_{12} + A_{22}B_{21})x_2$
- $z_3 = (A_{31}B_{11} + A_{32}B_{21})x_1 + (A_{31}B_{12} + A_{32}B_{21})x_2$
- These three scalar equations mean the composite linear map z = (AB)x. Thus, we identify the entries of AB. They satisfy the rule of matrix-matrix multiplication.

- In general, AB is not equal to BA.
- Often, when AB is defined, BA is not even defined.
- In class, we go over an example using A as an 3-entry column, and B as a 3-entry row. AB is a 3-by-3 matrix, but BA is a single-entry matrix. Clearly, AB is not equal to BA.
- If AB = BA, we say that A and B **commute**.
- **A few identities**:
- A(BC) = (AB)C
- A(B + C) = AB + AC
- (A + B)C = AC + BC
- The m-by-m **identity matrix** I_m.
- If A is a m-by-n matrix, then $I_mA = A = AI_n$.
- **Power of a square matrix**. A square matrix A can multiply itself. Thus, we write AA = A^2, AAA = A^3...

- **Operation 4: The transpose of a matrix**. To transpose a matrix A, write the ith row as the ith column. Denote the transpose of A by A^T
- **Maximilian Richter's question: When we transpose the transpose of a matrix, do we get the original matrix back**? Yes.
- Thus, $(A^T)^T = A$. A few other identities follow.
- $(A + B)^T = A^T + B^T$
- $(AB)^T = B^TA^T$
- $c^T = c$ for any number c.
- **Alaisha Sharma's question**: **Where does transpose appear in practice**? So far in the course, matrices only appear in two places: as the matrix of coefficients in linear equations, and as the matrix of linear map. In both places, we do not need the transpose of a matrix.
- In a very special case, we write a tuple of numbers (i.e., a vector in R^n) sometimes as a column and sometimes as a row. The two ways of writing are transpose of each other.
- Later we will see the transpose of matrices in some other places.

Week 5 (February 22, 24)

Reading and watching

- Watch our TF Jacob Scherba sing.
- Video on matrix-matrix multiplication. This kind of geometric interpretation seems to appeal to some people, but confuses me. Take a look. No need to follow the intricacies if they do not appeal to you.
- Video on nonsquare matrices
- A video of Leontief explaining the input-output table.
- Supplementary reading: Wassily Leontief, Input-output economics, Scientific American, 1951.
- Supplementary reading: Wassily Leontief, Academic economics. Science 217, 104, 1982. An essay on empirical economics and academic economics.
- Supplementary reading: Leontief's, Input-Output Economics, 2nd edition, 1986.
- Lay 2.2 The inverse of a matrix
- Lay 2.3 Characterization of inverse matrices
- Lay 2.4 Partitioned matrices
- Lay 2.6 The Leontief input-output model
- Suo notes on input-output economics
- Wiki input-output model
- Lay 2.7 Computer graphics.
- A video on inverting matrix by row reduction.

Lay 2.2, 2.3 The inverse of a matrix

- **The inverse of a number.**
- The inverse of 2 is ½, also written 2^{-1}.
- In general, the inverse of a number is defined in the axioms of number field.
- Of course, the idea is familiar to us.
- Let F be a number field, and 1 be the identity element in F for multiplication.
- For **every** nonzero element a in F, there exists a **unique** element c in F, such that ac = 1.
- We call c the **inverse** of a, and write c as a^{-1}.
- We next generalize this definition of the inverse of a number to the inverse of a matrix.

- **Operation 5: The inverse of a matrix**.
- First generalize the number 1 to identity matrices.
- Let A be an n-by-n matrix over a number field F, and I be the n-by-n identity matrix.
- The matrix A is called **invertible** if an n-by-n matrix C exists, such that AC = I.

- We call C the **inverse** of A, and write C as A^{-1}.
- This generalization seems natural, but legalistic.
- Let's look at the inverse in a familiar setting.

- **Invert the linear map from the chicken-rabbit space to the head-foot space**.
- Recall that a matrix appears in a linear map from one vector space to another vector space.
- The linear map means two linear scalar equations that calculate the numbers of heads and feet using the numbers of chickens and rabbits.
- **The equation of heads: $y_1 = x_1 + x_2$.**
- **The equation of feet: $y_2 = 2x_1 + 4x_2$.**
- Here is the "high-school method" to invert this linear map.
- To invert the this linear map means solving the numbers of chickens and rabbits in terms of the numbers of heads and feet.
- Go over this process of inversion in class to obtain the following
- **The equations of chickens: $x_1 = 2y_1 - y_2/2$.**
- **The equations of rabbits: $x_2 = -y_1 + y_2/2$.**
- Write the linear map as a matrix equation.
- Write the inverse linear map as another matrix equation.
- Verify that the two matrices are inverse of each other.
- Verify our old story of chickens and rabbits on a farm.
- Set $y_1 = 26$ and $y_2 = 82$.
- The inverse map gives us $x_1 = 11$ and $x_2 = 15$.

- **Paul Lei's question: what does it mean to have a total of three heads and four feet?** Let's see. Set $y_1 = 3$ and $y_2 = 4$. The map gives us $x_1 = 4$ and $x_2 = -1$. The answer makes sense so long as we interpret $x_2 = -1$ as being "in deficit of one rabbit". Thus, four chickens and one "negative rabbit" give us three heads and four feet.
- In general, we choose to work with real numbers, rather than positive integers, because the former form a **number field**. The number field is closed under arithmetic operations. This closedness simplifies the theory of equations.
- We can always try to interpret the answer after we obtain it.

- The textbook gives a formula for the inverse of a 2 x 2 matrix. No need to memorize the formula for any of our exams. You can always work it out as above, or look it up in practice.

- **Calculate the inverse of a matrix using the row reduction algorithm**. The "high-school method" used above is too creative for computers. We need an **algorithm** to invert a large matrix.
- Given A, view AC = I as the linear algebraic equations that determine C. C has n columns. Thus, we solve for the n columns in parallel.
- Put the matrices A and I side by side to form an augmented matrix [A I].
- Row-reduce the n-by-2n matrix [A I].

- If the n columns on the left reduce to the identity matrix I, then the n columns on the right give the inverse of A.
- **Chickens and rabbits**. Use the row reduction algorithm to invert the linear map from the chicken-rabbit space to the head-foot space. We went through this example in class.
- For a large matrix A, you can use the rref command in matlab to determine if A is invertible, and calculate the inverse of A if it is invertible.
- Watch A video on inverting matrix by row reduction.

- **Robert Collins's question: How do we invert a non-square matrix**? Answer: The inverse is only defined for square matrices. Non-square matrices are not invertible.

- **Rabbits and hamsters**.
- The linear map from the rabbit-hamster space to the head-foot space is given by two scalar equations:
- The equation of heads: $y_1 = x_1 + x_2$.
- The equation of feet: $y_2 = 4x_1 + 4x_2$.
- Identify the matrix A.
- Form the augmented matrix (A I).
- Apply the row reduction algorithm.
- The rref (A) misses a pivot.
- A is not invertible.

- Thus, the linear map from the chicken-rabbit space to the head-foot space is invertible.
- But the linear map from the rabbit-hamster space to the head-foot space is not invertible.

- Every nonzero number is invertible.
- **But not all nonzero, square matrices are invertible**.
- An n x n matrix A is invertible if and only if the matrix has n pivots.
- That is, A is invertible if and only if A is row-equivalent to the identity matrix, A ~ I. This is the same as saying rref(A) = I.
- **Theorem**. For a square matrix, the following conditions are equivalent: invertible, surjective, injective, bijective.
- **Proof**. Each condition means that the n x n matrix has n pivots.
- An invertible matrix is also called **nonsingular** matrix. A noninvertible matrix is also called a **singular** matrix.

- **Theorem**. For an invertible matrix A, the solution to the equation Ax = b is unique for every b. The solution is $x = A^{-1}b$.
- **Proof**. Multiply both sides of the equation Ax = b by A^{-1}.

- **Theorem**. If AC = I, then CA = I. That is, if C is an inverse of A, then A is an inverse of C.

- **Proof**. The row operations are reversible. Start from A, the row reduction algorithm calculates the inverse of A, the matrix C. Then start from C, the row reduction algorithm gives its inverse, A.
- **An alternative proof that does not rely on the row reduction algorithm**. Start with AC = I. Assume that the inverse of C is Z, so that CZ = I. We wish to show that A = Z. **Proof.** A = AI = A(CZ) = (AC)Z = IZ = Z.

- **Theorem**. If a square matrix A is invertible, then its inverse is unique.
- **Proof**. This statement is obvious from the row reduction algorithm.
- Here is an alternative proof that does not rely on the row reduction algorithm. Assume that A has two inverses, B and C. Then, B = BI = B(AC) = (BA)C = IC = C.

- **The inverse of a product**. Here is an identity: $(AB)^{-1} = B^{-1}A^{-1}$.
- **Proof**. $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$. Thus, AB is the inverse of $B^{-1}A^{-1}$.
- A remark attributed to Coxeter: The "reversal of order" becomes clear when we think of putting on our socks and shoes, and then removing them.

- Two more identities:
- $(A^{-1})^{-1} = A$
- $(A^T)^{-1} = (A^{-1})^T$

Lay 2.4 Partitioned matrices

- Read this section on your own, and do the homework problems.
- What is a partitioned matrix?
- How do matrix operations work for partitioned matrices?
- Find the inverse of a partitioned matrix (**Example 5 on p. 119**).

- **Lay 2.5 Matrix factorization**. Skip this section on computational linear algebra. Also skip elementary matrices in Lay 2.2. No homework or exam problems will draw upon these ideas.
- Computational linear algebra has become a sophisticated subject, which belongs to a different course.

Lay 2.6 The Leontief input-output model

- Suo notes on input-output economics
- Wiki input-output model
- Wassily Leontief, Professor of Economics, of Harvard University, won the 1973 Nobel Prize in Economic Science for his work on input-output economics.
- Supplementary video: Leontief explains the input-output table.
- Supplementary reading: Wassily Leontief, Input-output economics, Scientific American, 1951.

- **Big data met computers in the 1940s.** Read Leontielf's 1951 article, [Input-output Economics], for the excitement of a historical moment. A sentence from Leontief's article: "...the method has had to await the modern high speed computing machine as well as the present propensity of government and private agencies to accumulate mountains of data."
- This historical moment has continued to this day, in most aspects of our lives, far beyond economics.
- We live in the **Age of Big Data**.

- **Algebra of scalars**.
- The Leontief input-output model is an excellent illustration of the algebra of scalars.
- Like the chicken-rabbit problem, the input-output model is **a pure play of scalars**, containing no distracting details unrelated to scalars.
- That is why I devote one full lecture and multiple homework problems to this model.
- Of course, you expect more: all exam problems will be similar to homework problems.
- Besides, the model was created at Harvard, the work won Leontief a Nobel Prize, and the Big Data will be a main story played out in your life.
- These considerations are, of course, minor compared to our immortal love for scalar sets and scalar-scalar linear maps.
- Big data will be fashionable in your lifetime. But scalars are immortal.
- **In Leontief's model, "big data" means "many scalars and many scalar-scalar maps".**
- Various amounts of a product form a **scalar set**. For example, various amounts of steel form a scalar set. Various numbers of cars form another scalar set.
- The flow of a product into the making of another product defines a **scalar-scalar linear map**. For example, making a car needs a certain amount of steel.
- The story has some similarity to that each chicken has two feet. Or each water molecule has two hydrogen atoms.

- **No reason for us to dress up scalars into vectors**.
- Vectoring scalars only obscures the economics of the Leontief model.
- So, for the time being, ignore all the rambling about vectors in this section of the textbook.
- Unlearn vectors.
- Return to Leontief's story of products and their flows.

- **Total output and final demand**
- **Products**. The Leontief model divides an economy into n **products**, labeled as product 1,2,...,n.
- Various amounts of product i form a **scalar set**, S_i.
- Let x_i be the **total output** of product i in a year.
- Let d_i be the **final demand** for product i in the year.

- x_i and d_i are two elements in the scalar set S_i.
- The model aims to predict the total outputs $x_1, x_2, \ldots, x_n$, given the final demands $d_1, d_2, \ldots, d_n$.
- In a **trivial economy**, each product is made from scratch, and the total output of the product matches the final demand for the product:
- $x_1 = d_1, x_2 = d_2, \ldots, x_n = d_n$.

- **Making a product needs many products**.
- In a **nontrivial economy**, Each product is not made from scratch, but requires many products.
- For example, making cars needs steel. Making cars even needs some cars.
- Let z_{ij} be the amount of product i needed for the total output x_j of product j.
- Remember the distinct meanings of the two subscripts.
- z_{ij} is an element in the scalar set S_i.

- **One for all**.
- Each row of the input-output table states a fact of economics: each product fulfills the needs of the making of all products and the final demand.
- Each row lists scalars in the same scalar set, i.e, the different amounts of the same product.
- Scalars in the same scalar set are additive:
- (total output of product i) = (sum of amounts of product i needed to make all products) + (final demand for product i).
- Translate the above economic statement to a system of equations.
- Product 1: $x_1 = z_{11} + z_{12} + \ldots + z_{1n} + d_1$,
- Product 2: $x_2 = z_{21} + z_{22} + \ldots + z_{2n} + d_2$,
- ……
- Product n: $x_n = z_{n1} + z_{n2} + \ldots + z_{nn} + d_n$.
- Note the order of the subscripts.
- Each equation corresponds to a row in the input-output table, that is, to different amounts of one product.

- **All for one**.
- Each column of the input-output table states another fact of economics: the making of each product requires the input of other products.
- Each column lists the amounts of products needed to make one product.
- Every entry in a column is an amount of a different product.
- Scalars in different scalar sets are not additive.

Input-output table of a two-product economy

	product 1	product 2	final demand	total output

product 1	z_{11}	z_{12}	d_1	x_1
product 2	z_{21}	z_{22}	d_2	x_2

Input-output table of the wheat-cloth economy in physical units

	wheat	cloth	final demand	total output
wheat	25 bushels	20 bushels	55 bushels	100 bushels
cloth	14 yards	6 yards	30 yards	50 yards

- **A wheat-cloth economy.** Here is an example adapted from Leontief's book, Input-Output Economics, 2nd edition, 1986.
- An economy makes two and only two products: wheat and cloth.
- Various amounts of wheat form one scalar set, called the **wheat set**.
- The unit for the wheat set is **1 bushel of wheat**.
- Various amounts of cloth form another scalar set, called the **cloth set**.
- The unit for the cloth set is **1 yard of cloth**.

- **The rows of the input-output table**
- The first row of the table lists some amounts of **wheat**.
- Of the total output of 100 bushels of wheat, 25 bushels go to the making of wheat, 20 bushels go to the making of cloth, and 55 bushels go to the final demand.
- Scalars in a scalar set (the wheat set) are additive:
- The total output of 100 bushels of wheat = 25 + 20 + 55.
- The second row of the table lists some amounts of **cloth.**
- Of the total output of 50 yards of cloth, 14 yards go to the making of wheat, 6 yards go to the making of cloth, and 30 yards go to the final demand.
- Scalars in a scalar set (the cloth set) are additive:
- The total output of 50 yards of cloth = 14 + 6 + 30.

- **The columns of the input-output table**
- The first column lists the amounts of wheat and cloth needed to make the total output of 100 bushels of wheat: 25 bushels of wheat, and 14 yards of cloth.
- The second column lists the amounts of wheat and cloth needed to make the total output of 50 yards of cloth: 20 bushels of wheat, and 6 yards of cloth.
- The third column lists the final demands: 55 bushels of wheat, and 30 yards of cloth.
- The last column lists the total outputs: 100 bushels of wheat, and 50 yards of cloth.
- The entries in any column are elements in different scalar sets. They are not additive.
- Wheat and cloth do not add.

- **Record past. Predict future**.

- The entries in the input-output table are, in practice, recorded economic data of a past year.
- We want to predict future.
- If we know the final demands in coming years, how do we predict the total outputs needed in coming years?
- Here is the key insight that enabled Leontief to use the recorded data of the past to predict future.
- **Whereas the number of cars made varies from year to year, the amount of steel needed to make each car is nearly constant**.
- This insight reflects a fact of the car-making technology.
- To abstract this insight, we say that the scalars change from year to year, but the scalar-scalar linear maps remain nearly constant.
- Leontief made this insight precise as follows.

- **Input coefficients**. Let C_{ij} be the amount of product i needed to make a unit amount of product j.
- C_{ij} is called the **input coefficient of product i into product j**.
- Remember the distinct meanings of the two subscripts.
- $z_{ij} = C_{ij}x_j$.
- z_{ij} is an element in scalar set S_i, and x_j is an element in the scalar set S_j.
- C_{ij} is a **linear map** from the scalar set of product j to the scalar set of product i.
- C_{ij}: $S_j \rightarrow S_i$.
- Leontief called C_{ij} the **input coefficients**, or the technology coefficients.
- Leontief listed the input coefficients as a table.
- His convention for the table: **column j** lists the amounts of all products needed to make a unit amount of product j.

Table of input coefficients of a two-product economy

	product 1	product 2
product 1	$C_{11} = z_{11}/x_1$	$C_{12} = z_{12}/x_2$
product 2	$C_{21} = z_{21}/x_1$	$C_{22} = z_{22}/x_2$

- We construct the table of input coefficients for the wheat-cloth economy.
- Each input coefficient has a distinct unit.
- The table is calculated using existing economic data.

Table of input coefficients of the wheat-cloth economy in physical units

	wheat	cloth
wheat	(25 bushels)/(100 bushels)	(20 bushels)/(50 yards)

cloth	(14 yards)/(100 bushels)	(6 yards)/(50 yards)

- **Output-demand equation**
- Now we are ready to answer the question: given the final demands of future years, how do we predict the total outputs?
- **Use past data to predict future**.
- The table of input coefficients is calculated using the past economic data.
- The same table is then assumed to be constant, and used to relate the future outputs to the future demands.
- Each product fulfills the needs of the making of all products and the final demand.
- That is, scalars in each scalar set are additive:
- **(total output of product i) = (sum of amounts of product i needed to make all products) + (final demand for product i).**
- Translate the above economic statement to a system of equations.
- Product 1: $x_1 = C_{11}x_1 + C_{12}x_2 + \ldots + C_{1n}x_n + d_1$,
- Product 2: $x_2 = C_{21}x_1 + C_{22}x_2 + \ldots + C_{2n}x_n + d_2$,
- ……
- Product n: $x_n = C_{n1}x_1 + C_{n2}x_2 + \ldots + C_{nn}x_n + d_n$.
- We call this system of equations the **output-demand equations**.
- **Each equation corresponds to a row in the input-output table**.
- Note the order of the subscripts.
- We now have n equations for n variables, $x_1, x_2, \ldots, x_n$.
- $(1-C_{11})x_1 - C_{12}x_2 - \ldots - C_{1n}x_n = d_1$,
- $-C_{21}x_1 + (1-C_{22})x_2 + \ldots - C_{2n}x_n = d_2$,
- ……
- $-C_{n1}x_1 - C_{n2}x_2 - \ldots + (1- C_{nn})x_n = d_n$.
- Assume that the final demands $d_1, d_2, \ldots, d_n$ are known, and the input coefficients C_{ij} are also known.

- **Relate outputs to demands in the wheat-cloth economy.**
- Using the table of input coefficients for this economy, the output-demand equations become
- $(1-0.25)x_1 - 0.40x_2 = d_1$ bushels of wheat
- $-0.14x_1 + (1-0.12)x_2 = d_2$ yards of cloth
- Given the total outputs, the above equations calculate the final demands.
- Conversely, given the final demands, we can solve these equations to predict the total outputs:
- $x_1 = 1.457d_1 + 0.6623d_2$ bushels of wheat
- $x_2 = 0.2318d_2 + 1.241d_2$ yards of cloth
- For example, set d_1 = 55 bushels of wheat and d_2 = 30 yards of cloth, and the equations will give x_1 = 100 bushels of wheat and x_2 = 50 yards of cloth.

- If the final demands in the future years are known to change to new values, the above equations predict necessary total outputs.

- **Price and value added**
- In the input-output table in physical units, entries in a column have different units, and are not additive.
- We now recall another ancient idea: **money is a universal scalar**.
- Various amounts of money (e.g., the US dollars) form a scalar set, M.
- A price is a linear map from the scalar set of a product to the scalar set of money.
- (amount of money) = (price of product i) (amount of product i)

- Let p_i be the **price** of product i.
- p_i is a linear map p_i: $S_i \to M$.
- $p_i x_i$ is the value of the total output of product i.
- $p_i d_i$ is the value of the final demand of product i.
- $p_i z_{ij}$ is the payment for the product j to make the total output x_i of product i.
- Let v_i be the **value added** per unit amount of product i.
- $v_i x_i$ is the value added of the total output of product i.
- Two rows are added to the input-output table, one for the values added of individual products, and the other for the total values of individual products.

Input-output table of a two-product economy in $

	product 1	product 2	final demand	total output
product 1	$p_1 z_{11}$	$p_1 z_{12}$	$p_1 d_1$	$p_1 x_1$
product 2	$p_2 z_{21}$	$p_2 z_{22}$	$p_2 d_2$	$p_2 x_2$
value added	$v_1 x_1$	$v_2 x_2$		
total value	$p_1 x_1$	$p_2 x_2$		

- We construct the input-output table for the wheat-cloth economy in the monetary unit.
- Assume the following prices
- p_{wheat} = 2 $/(bushel of wheat)
- p_{cloth} = 5 $/(yard of cloth)
- Every entry in the table has the same unit: $.

Input-output table of the wheat-cloth economy in $

	wheat	cloth	final demand	total output
wheat	50	40	110	200

cloth	70	30	150	250
value added	80	180		
total value	200	250		

- **Price-value equation**.
- Now the column under the heading “product i” lists scalars in the same scalar set, the money set M.
- These entries are additive:
- **(value of the total output of product i) = (payments for all products needed to make product i) + (value added to make product i)**
- Value of product 1: $p_1x_1 = p_1z_{11} + p_2z_{21} + \ldots + p_nz_{n1} + v_1x_1$,
- Value of product 2: $p_2x_2 = p_1z_{12} + p_2z_{22} + \ldots + p_nz_{n2} + v_2x_2$,
- …...
- Value of product n: $p_nx_n = p_1z_{1n} + p_2z_{2n} + \ldots + p_nz_{nn} + v_nx_n$.
- **Each equation corresponds to a column in the input-output table**.
- Note the order of the subscripts.

- Dividing equation i by the total output of product i, we obtain that
- Price of product 1: $p_1 = C_{11}p_1 + C_{21}p_2 + \ldots + C_{n1}p_n + v_1$,
- Price of product 2: $p_2 = C_{12}p_1 + C_{22}p_2 + \ldots + C_{n2}p_n + v_2$,
- …...
- Price of product n: $p_n = C_{1n}p_1 + C_{2n}p_2 + \ldots + C_{nn}p_n + v_n$.
- We call this system of equations the **price-value equations.**
- p_iC_{ij} is the payment for the product j to make a unit amount of product i.
- We now have n equations to relate the prices $p_1, p_2, \ldots, p_n$ and the value added per unit output $v_1, v_2, \ldots, v_n$:
- $(1-C_{11})p_1 - C_{21}p_2 - \ldots - C_{n1}p_n = v_1$,
- $-C_{12}p_1 + (1-C_{22})p_2 + \ldots - C_{n2}p_n = v_2$,
- ……
- $-C_{1n}p_1 - C_{2n}p_2 - \ldots + (1- C_{nn})p_n = v_n$.

- **Gross domestic product** (GDP).
- The gross domestic product is defined in two ways:
- $GDP = p_1d_1 + p_2d_2 + \ldots + p_nd_n$. That is, GDP is the monetary value of the final demand summed over all products.
- $GDP = v_1x_1 + v_2x_2 + \ldots + v_nx_n$. That is, GDP is value added of the total output summed over all products.

- **Theorem**. The two definitions of GDP give the same result.

- **Proof**. Multiply the **output-demand equation** of each product by the price of the product, and then sum over all products, giving
- $p_1x_1 + p_2x_2 + ... + p_nx_n$
- $= p_1C_{11}x_1 + p_1C_{12}x_2 + \ldots + p_1C_{1n}x_n + p_1d_1$
- $+ p_2C_{21}x_1 + p_2C_{22}x_2 + \ldots + p_2C_{2n}x_n + p_2d_2,$
- ……
- $+ p_nC_{n1}x_1 + p_nC_{n2}x_2 + \ldots + p_nC_{nn}x_n + p_nd_n.$

- Multiply the **price-value equation** of each product by the output of the product, and then sum over all products, giving
- $x_1p_1 + x_2p_2 + \ldots + x_np_n$
- $= x_1C_{11}p_1 + x_1C_{21}p_2 + \ldots + x_1C_{n1}p_n + x_1v_1,$
- $+ x_2C_{12}p_1 + x_2C_{22}p_2 + \ldots + x_2C_{n2}p_n + x_2v_2,$
- …...
- $+ x_nC_{1n}p_1 + x_nC_{2n}p_2 + \ldots + x_nC_{nn}p_n + x_nv_n.$

- A comparison of the two equations shows that
- $p_1d_1 + p_2d_2 + ... + p_nd_n = v_1x_1 + v_2x_2 + ... + v_nx_n$

- To develop some intuition, look up the two sums in the input-output table in monetary unit.
- Focus on the terms marked in red in the following two tables.

Input-output table of a two-product economy in $

	product 1	product 2	final demand	total output
product 1	p_1z_{11}	p_1z_{12}	p_1d_1	p_1x_1
product 2	p_2z_{21}	p_2z_{22}	p_2d_2	p_2x_2
value added	v_1x_1	v_2x_2		
total value	p_1x_1	p_2x_2		

Input-output table of the wheat-cloth economy in $

	wheat	cloth	final demand	total output
wheat	50	40	110	200
cloth	70	30	150	250
value added	80	180		
total value	200	250		

- **Matrix equations**
- We can, of course, translate the two systems of scalar equations into two matrix equations.

- **Output-demand equation**: $x = Cx + d$.
- Also write this equation as $(I - C)x = d$
- x = column of total outputs of products.
- d = column of final demands for products
- Cx = column of all products needed to produce total outputs (also called the intermediate demands for products).

- **Price-value equation**: $p = C^T p + v$.
- Also write this equation as $(I - C^T)p = v$
- p = column of prices of products.
- v = column of values added per unit amounts of products
- $C^T p$ = column of payments for making unit amounts of products

- Two definitions of gross domestic product:
- $GDP = p^T d$
- $GDP = v^T x$

- **Do not let vectors obscure economics.**
- In formulating the model above, we have made two types of translation:
- (Economics of products) → (Scalar equations).
- (Scalar equations) → (Vector equations)

- (Economics of products) → (Scalar equations)
- Various amounts of each product form a scalar set.
- Making one product consumes other products.
- Each product-product relation defines a scalar-scalar linear map.
- We translate the economics of products to equations of scalars.

- (Scalar equations) → (Vector equations)
- We then translate the scalar equations into vector equations
- Vector equations make the writing more concise.
- Concise writing takes less time to write and less time to think, letting us manipulate equations more readily.
- Facts of vectors and matrices are at your service.

- When the concise writing of vectors obscure economics, always return to the Leontief's original story of products.
- After all, vector equations are just a concise way of writing the same scalar equations.

- Concise writing of the same thing cannot possibly tell you anything really new.

- **Warning: the confusing transpose**
- The textbook by Lay adopts **two different conventions to construct the input-output table**:
- In the table on p. 50, **row j** lists the purchases by sector j from all sectors to produce a unit amount of goods in sector j. For example, the bottom row lists the purchases of the steel sector.

	coal	electric	steel
coal	0.0	0.4	0.6
electric	0.6	0.1	0.2
steel	0.4	0.5	0.2

- In the table on p. 133, **column j** lists the purchases by sector j from all sectors to produce a unit amount of product of sector j. For example, the last column lists the purchases of the services sector.

	manufacturing	agriculture	services
manufacturing	0.5	0.4	0.2
agriculture	0.2	0.3	0.1
services	0.1	0.1	0.3

- This **transpose** of the input-output table is confusing, but reflects different preferences in the literature.
- You just need be careful in constructing the input-output tables. State explicitly your convention in homework problems.

- **Lay 2.7 Computer graphics**. Read this section, and solve homework problems.

- **Lay 2.8, 2.9**. Skip these sections. The same content appears in Chapter 4, which we will study thoroughly.

Week 6 (February 27, March 1, 3)

Reading and watching

- A video on determinant.
- S. Axler, Down with determinants! American Mathematical Monthly 102, 139-154 (1995). This paper received the Lester R. Ford Award for expository writing from the Mathematical Association of America. The paper is rather technical for this class, but the introduction of the paper gives some perspective.
- Lay 3.1 Introduction to determinants.
- Lay 3.2 Properties of determinants.
- Lay 3.3 Cramer's rule. Area, volume, and linear maps.

Lay 3.1 Introduction to determinants

- My lecture deviates from Lay 3.1.
- **The invertible matrix theorem**.
- This theorem makes a large number of equivalent statements.
- Look for "invertible matrix theorem" in the index of our textbook.
- Here are two equivalent statements.
- A system of linear algebraic equations, Ax = b, has a **unique** solution for **every** b if and only if A is an invertible matrix.
- A square matrix A is invertible if and only if every column of A is a pivot column.
- We next develop yet another equivalent statement:
- Ax = b has a unique solution for every b if and only if **det A is not zero.**

- **Definition of det A**
- Determinant is only defined for square matrices.
- Let's work out a definition of determinant.

- 1 x 1 matrix
- A = [a_{11}]
- Ax = b has a unique solution for every b if and only if a_{11} is not zero.
- Define **det A = a_{11}**.

- 2 x 2 matrix A

a_{11}	a_{12}
a_{21}	a_{22}

- Row reduce the matrix.

a_{11}	a_{12}	
$a_{11}a_{21}$	$a_{11}a_{22}$	$a_{11} \times R_2$

a_{11}	a_{12}	
0	$a_{11}a_{22} - a_{21}a_{12}$	$R_2 - a_{21} \times R_1$

- We discover that
- Ax = b has a unique solution for every b if and only if $a_{11}a_{22} - a_{21}a_{12}$ is not zero.
- In the the first row operation above, we have assumed that a_{11} is not zero.
- If $a_{11} = 0$, inspecting the original matrix, to have pivots in both rows and both columns, we must require that a_{12} be not zero and a_{21} be not zero. In this case, we can still state that
- Ax = b has a unique solution for every b if and only if $a_{11}a_{22} - a_{21}a_{12}$ is not zero.
- Define det A = $a_{11}a_{22} - a_{21}a_{12}$.
- Thus, **Ax = b has a unique solution for every b if and only if det A is not zero**.
- A way to memorize the formula det A = $a_{11}a_{22} - a_{21}a_{12}$. A cross.

- 3 x 3 matrix A

a_{11}	a_{12}	a_{13}
a_{21}	a_{22}	a_{23}
a_{31}	a_{32}	a_{33}

- Row reduce the matrix.

a_{11}	a_{12}	a_{13}	
$a_{11}a_{21}$	$a_{11}a_{22}$	$a_{11}a_{23}$	$a_{11} \times R_2$
$a_{11}a_{31}$	$a_{11}a_{32}$	$a_{11}a_{33}$	$a_{11} \times R_3$

a_{11}	a_{12}	a_{13}	
0	$a_{11}a_{22} - a_{21}a_{12}$	$a_{11}a_{23} - a_{21}a_{13}$	$R_2 - a_{21} \times R_1$
0	$a_{11}a_{32} - a_{31}a_{12}$	$a_{11}a_{33} - a_{31}a_{13}$	$R_3 - a_{31} \times R_1$

a_{11}	a_{12}	a_{13}	
0	$a_{11}a_{22} - a_{21}a_{12}$	$a_{11}a_{23} - a_{21}a_{13}$	
0	0	$a_{11}D$	$(a_{11}a_{22} - a_{21}a_{12})R_3 - (a_{11}a_{32} - a_{31}a_{12})R_2$

- where D = $a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$.
- We discover that
- Ax = b has a unique solution for every b if and only if D is not zero.
- In the first row operation above, we have assumed that a_{11} is not zero. In the last row operation, we have assumed that $(a_{11}a_{22} - a_{21}a_{12})$ is not zero.
- If these assumptions are invalid, to ensure pivots for every row and column, we can still show that
- Ax = b has a unique solution for every b if and only if D is not zero.
- Define det A = $a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$.
- Jacob Bindman. noted that I wrote some products wrong on the blackboard in class. Please correct your notes.
- Thus, **Ax = b has a unique solution for every b if and only if det A is not zero.**
- A way to memorize the formula for the determinant of a 3 x 3 matrix. A cross and two stars of David.
- You can find another way to memorize the formula on p. 168 of Lay (In class, a number of students said that this is the way they learned in high school.)
- It makes no difference how you remember the definition. Just remember it!
- Example: Calculate the determinant of a 3 x 3 matrix.

- We can go on to row reduce a 4 x 4 matrix.
- Alternatively, let's look at the cases of 2 x 2 and 3 x 3 matrices again.
- Try to discover a definition of determinant that can generalize to any n.

- **A is a square matrix.**
- **det A is a number**.
- Let A be an n x n square matrix over a number field F.
- Each entry of A, written a_{ij}, is a number in F.
- The first index indicates the row and the second index the column in which a_{ij} appear in the matrix A.
- The determinant of A, written det A, is a **function** that sends a square matrix A to a number in F.
- det: A → F.
- How is this function defined?
- That is, given a square matrix A, how do we calculate the number det A?
- We next describe the determinant of a 3 x 3 matrix A in a way that generalizes to any n.

- **det A is an algebraic sum of certain products of the entries of A**.
- To define det A, we need to specify the factors in each product and the sign of each product.
- **Factors in each product**.
- Each product contains one entry from each row and each column.
- For a 3 x 3 matrix, write a product as $a_{i1}a_{j2}a_{k3}$.
- Arrange the column indices in the increasing order.

- The sequence of the row indices i, j, k is a **permutation** of the numbers 1, 2, 3.
- The total number of permutations for n = 3 is 3! = 6.
- For an n x n matrix A, there are a total of n! distinct products.
- Look at the expression of det A for the 3 x 3 matrix.
- Look at the stars of David for the 3 x 3 matrix.
- Each product contains one entry from each row and each column.
- **Sign of each product**.
- By an **inversion** in the sequence of numbers i, j, k we mean an arrangement of two indices such that the larger index appears before the smaller index.
- Denote the **number of inversions** in the sequence i,j,k by N(i,j,k).
- For example,
- N(1,2,3) = 0,
- N(3,2,1) = 3.
- **In the determinant, a product is positive if the number of inversions is an even number, and a product is negative if the number of inversions is an odd number**.
- For the 3 x 3 matrix A, we count the number of inversions for each of the 6 permutations:
- N(1,2,3) = 0,
- N(2,3,1) = 2,
- N(3,1,2) = 2,
- N(3,1,2) = 3,
- N(1,3,2) = 1,
- N(2,1,3) = 1.
- They correspond to positive and negative signs of the six products in det A.
- A picture is worth a thousand words. An equation is worth a thousand pictures.
- Here is an equation that defines the determinant of a 3 x 3 matrix A:
- $\det A = \text{Sum } (-1)^{N(i,j,k)} a_{i1} a_{j2} a_{k3}$.
- The sum is taken as i, j, k go through all permutations of 1, 2, 3.
- This definition generalizes to square matrices of any size n.
- For example, for a 4 x 4 matrix A,
- $\det A = \text{Sum } (-1)^{N(i,j,k,p)} a_{i1} a_{j2} a_{k3} a_{p4}$.
- The sum is taken as i, j, k, p go through all permutations of 1, 2, 3, 4.

- **Example**. 4 x 4 matrix A
- det A is an algebraic sum of 4! = 24 products.
- Each product takes the form $a_{i1}a_{j2}a_{k3}a_{p4}$, where i, j, k, p are a permutation of 1, 2, 3, 4.
- Because N(1,2,4,3) = 1, the sign of the product $a_{11}a_{22}a_{43}a_{34}$ in det A is negative.

- **An alternative definition of determinant** (Lay p. 165)
- $\det A = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$
- $\det A = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$
- Let A_{ij} be the **submatrix** of A obtained by deleting row i and column j.
- $\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}$
- This expression is called the expansion across row 1

- You can also write the expression across row 2:
- $\det A = -a_{21}\det A_{21} + a_{22}\det A_{22} - a_{23}\det A_{23}$
- You can also write an expansion across any row.
- You can also write an expansion down any column.
- Remember the sign using the following table.

+	-	+
-	+	-
+	-	+

- You can do the same for n x n matrix. For a 4 x 4 matrix, the table of signs is as follows:

+	-	+	-
-	+	-	+
+	-	+	-
-	+	-	+

- **The rule for the sign**. For element a_{ij}, the sign is positive if i + j is even, and the sign is negative if i + j is odd.
- The textbook goes on to introduce cofactor. We will ignore it. In case you hear people talking about it, ask them to explain, or just look it up on Wikipedia.

- **Triangular matrix**.
- A square matrix is invertible if and only if every column is a pivot column.
- Such a matrix is row-equivalent to a triangular matrix, with all the diagonal elements being nonzero.
- The determinant of a triangular matrix is the product of the diagonal elements.

Lay 3.2 Properties of determinants

- How does a **row operation** affect a determinant?
- Adding a multiple of a row to another row does not change the determinant.
- Switching two rows changes the sign of determinant.
- If one row of A is multiplied by a number c to produce B, then det B = c det A.

- Example: **calculate det A using the row reduction algorithm.**
- Write det A as |A|.

- **A few other properties of determinant**
- A square matrix A is invertible if and only if det A is not zero.
- $\det (A^T) = \det A$.
- det (AB) = (det A)(det B)

Lay 3.3 Cramer's rule, area, volume, and linear maps

- **Cramer's rule**.
- A system of equations, $Ax = b$, where A is an invertible square matrix.
- Thus, $Ax = b$ has a unique solution for every b.
- Let $A_i(b)$ be the matrix obtained from A by replacing column i with b.
- **Theorem**. The solution to $Ax = b$ is
- $x_i = \det(A_i(b)) / \det A$.
- **Proof**. Write A as n columns, $A = [a_1,\dots, a_i,\dots,a_n]$.
- Then, $A_i(b) = [a_1,\dots, b,\dots,a_n]$, where a_i, is replaced with b.
- Recall that $a_k = Ae_k$
- Thus, $A_i(b) = [Ae_1,\dots, Ax,\dots,Ae_n]$
- This equation is the same as
- $A_i(b) = A[e_1,\dots, x,\dots,e_n]$
- Take det on both sides, and we obtain that
- $\det(A_i(b)) = (\det A)(\det [e_1,\dots, x,\dots,e_n])$
- Note that $\det [e_1,\dots, x,\dots,e_n] = x_i$.
- Thus, $x_i = \det(A_i(b)) / \det A$.
- **Example**. Chicken-rabbit problem.
- Cramer's rule is costly in computation, and is not used by computers.

- **Area**
- Let A be a 2 x 2 matrix.
- The two columns of A define a parallelogram.
- The area of the parallelogram is $|\det A|$.
- What does "area" mean in the chicken-rabbit plane?
- Nothing!
- Thus, the theorem does not pass the chicken-rabbit test.
- Use this theorem with caution.
- Make sure you have a reason to calculate this "area".

- **Volume**
- Let A be a 3x 3 matrix.
- The three columns of A define a parallelepiped.
- The volume of the parallelogram is $|\det A|$.
- What does "volume" mean in the chicken-rabbit-hamster space?
- Nothing!
- Thus, the theorem does not pass the chicken-rabbit-hamster test.
- Use this theorem with caution.
- Make sure you have a reason to calculate this "volume".

- **Linear map from plane to plane**

- T maps one region S in a plane to a region T(S) in another plane.
- area of T(S) = |det T| (area of S).
- **Example**. The following linear map T sends a circle of unit radius to an ellipse of semiaxes a and b.

a	0
0	b

- (area of the ellipse) = |det T| (area of unit circle) = (ab)(pi).

- **Linear map from 3D space to 3D space**
- T maps one region S in a 3D space to a region T(S) in another 3D space.
- volume of T(S) = |det T| (volume of S).
- **Example**. The following linear map T sends a sphere of unit radius to an ellipsoid of semiaxes a, b, and c.

a	0	0
0	b	0
0	0	c

- (volume of the ellipsoid) = |det T| (volume of unit sphere) = (abc)(4pi/3).

Week 7 (March 6, 8, 10)

Reading and watching

- A video on subspace
- A video on change of basis
- A video on abstract vector space
- A video on change of basis for the polynomial space
- Supplementary reading. A. Tucker. The growing importance of linear algebra in undergraduate mathematics. The College Mathematics Journal 24, 3-9 (1993). This article sketches the historical development of linear algebra, and its importance in undergraduate curriculum.
- Lay 4.1-4.9

Lay 4.1, 4.5 Vector space. Subspace. Dimension

- **The textbook defines an n-dimensional real vector space by axioms**.
- Two essential pages are
- p. 190: definition of real vector space of any dimension.
- p. 226: definition of dimension.

- Take a look at these axioms, but no need to master them.
- We will derive these axioms in simple steps.
- The textbooks gives many good examples, which do not rely on the legalistic definition of vector spaces.
- Read examples with care.

- **We build a vector space from scalar sets:**
- An n-dimensional real vector space is the Cartesian product of n real scalar sets.
- We call the number of scalar sets in the Cartesian product, n, the **dimension of the vector space**.
- Thus, in our definition, the dimension of a vector space is a transparent idea, and requires no legalistic rambling.

- **Gold set and silver set**.
- Let's look at a familiar vector space.
- Various amounts of gold form a real scalar set, called the gold set, S_1.
- Various amounts of silver form a real scalar set, called the silver set, S_2.
- Any nonzero amount of gold u_1 is a piece of for gold set S_1.
- Any nonzero amount of silver u_2 is a unit for the silver set S_2.
- Any piece of gold is a scalar in S_1, and is a number x_1 times the unit u_1. Write the piece of gold as x_1u_1.
- Any piece of gold is a scalar in S_2, and is a number x_2 times the unit u_2. Write the piece of silver as x_2u_2.
- The numbers x_1 and x_2 are the **magnitudes** of the two scalars relative to the units.

- **We cannot choose the same unit for different scalar sets**.
- Choose 1 gram of gold as a unit for the gold set, S_1.
- u_1 = 1 gram of gold.
- Choose 100 atoms of silver as a unit for the silver set, S_2.
- u_2 = 100 atoms of silver.
- A unit for a scalar set is a thing, such as a piece of gold, or a piece of silver.
- 1 gram of gold and 1 gram silver have the same mass, but they are two different things.

- **Gold-silver space**.
- The Cartesian product of the two real scalar sets defines a two-dimensional real vector space, called the gold-silver space,
- $V = S_1 \times S_2$.
- **(gold-silver space) = (gold set) x (silver set)**
- Any gold-silver vector v has some amount of gold and some amount of silver.
- That is a gold-silver vector v is an ordered pair of some amount of gold and some amount of silver:
- $v = (x_1u_1, x_2u_2)$.

- Given the units u_1 and u_2, we can label each gold-silver vector v by an ordered pair of real numbers, (x_1, x_2).
- That is, a choice of units for the scalar sets defines an **isomorphism** between the gold-silver vector space V and R^2.

- For example, here is a gold-silver vector:
- v = (12 grams of gold, 700 atoms of silver).
- Then, $v = (12u_1, 7u_2)$.
- The magnitudes of the two scalars relative to the units are $x_1 = 12$ and $x_2 = 7$.
- Once we choose the units, we label a piece of 12 grams of gold and 700 atoms of silver by an ordered pair of numbers, (12,7).
- To understand what the pair of numbers mean, of course, we need to know the two units.

Lay 4.3, 4.4 A basis for a vector space. Components of a vector relative to a basis

- We will develop the theory using a two-dimensional vector space, and will then summarize the results for an n-dimensional vector space.
- Let u_1 be a unit for a real scalar set S_1.
- Let u_2 be a unit for another real scalar set S_2.
- The Cartesian product of the two real scalar sets defines a two-dimensional real vector space, $V = S_1 \times S_2$.
- A vector v in V is an ordered pair of scalars, $v = (x_1u_1, x_2u_2)$
- The numbers x_1 and x_2 are called the **magnitudes** of the scalars relative to the units u_1 and u_2.

- We next formulate the idea of **a basis for the vector space V**.

- A **standard basis** for V:
- $a_1 = (u_1, 0)$
- $a_2 = (0, u_2)$
- Any vector v in V is a **linear combination** of the standard basis vectors:
- $v = x_1a_1 + x_2a_2$.
- The numbers x_1 and x_2 are called the **components of the vector v relative to the basis a_1 and a_2**.
- The numbers x_1 and x_2 are often written as a column x, called the **column of the vector v relative to the basis a_1 and a_2**.

- In a two-dimensional vector space V, there are many pairs of **linearly independent vectors**.
- Any two linearly independent vectors b_1 and b_2 are called **a basis for V**.
- Any vector v in V is a linear combination of b_1 and b_2.

- That is, for any vector v in V, there exist numbers y_1 and y_2, such that
- $v = y_1b_1 + y_2b_2$.
- The numbers y_1 and y_2 are called the **components of the vector v relative to the basis b_1 and b_2**.
- The numbers y_1 and y_2 are often written as a column y, called the **column of the vector v relative to the basis b_1 and b_2**.

Lay 4.7 Change of basis

- Any two linearly independent vectors in a two-dimensional vector space V are a basis for V.
- Start with a basis a_1 and a_2 for V, and generate any other basis b_1 and b_2 for V by using a 2 x 2 invertible matrix P.
- Here a_1 and a_2 do not have to be a standard basis.
- We will call a_1 and a_2 an old basis
- We will call b_1 and b_2 a new basis.
- We will call P the change-of-basis matrix from the old basis a_1 and a_2 to the new basis b_1 and b_2.

- To illustrate the steps, let us use an invertible matrix P

2	5
3	7

- Confirm that the matrix P is invertible:
- The two columns are linearly independent.
- Because a_1 and a_2 are a basis for V, each vector in V is a linear combination of the basis vectors a_1 and a_2.
- In particular, b_1 and b_2 are vectors in V. Write
- $b_1 = 2a_1 + 3a_2$,
- $b_2 = 5a_1 + 7a_2$.
- We adopt the following convention:
- Column 1 of the matrix P lists the components of the new basis vector b_1 relative to the old basis a_1 and a_2.
- Column 2 of the matrix P lists the components of the new basis vector b_2 relative to the old basis a_1 and a_2.
- By this convention, the matrix P appears transposed in relating the bases.

- For instance, let a_1 and a_2 be a standard basis,
- $a_1 = (u_1, 0)$
- $a_2 = (0, u_2)$
- The matrix P changes the standard basis a_1 and a_2 to a new basis:
- $b_1 = (2u_1, 3u_2)$,

- $b_2 = (5u_1, 7u_2)$.

- As a second example, let a_1 and a_2 be a nonstandard basis,
- $a_1 = (u_1,u_2)$
- $a_2 = (4u_1, u_2)$
- The matrix P changes the standard basis a_1 and a_2 to a new basis:
- $b_1 = 2a_1 + 3a_2 = 2(u_1,u_2) + 3\ (4u_1, u_2) = (14u_1,5u_2)$
- $b_2 = 5a_1 + 7a_2 = 5(u_1,u_2) + 7(4u_1, u_2) = (33u_1,12u_2)$.

- **Theorem**. For any linearly independent vectors a_1 and a_2 in a two-dimensional vector space V, the vectors b_1 and b_2 in V are linearly independent if and only if P is invertible.
- That is, a basis a_1 and a_2 for V generates any other basis for V by using a 2 x 2 invertible matrix P.

Change of the column of a vector under a change of basis

- Let a_1 and a_2 be an old basis for V, not necessarily a standard basis.
- Let b_1 and b_2 be a new basis for V.
- To show the basic points, we still use the invertible matrix P

2	5
3	7

- by convention, write
- $b_1 = 2a_1 + 3a_2$,
- $b_2 = 5a_1 + 7a_2$.
- A vector v in V has the old components x_1 and x_2 relative to the old basis a_1 and a_2, and has the new components y_1 and y_2 relative to new basis b_1 and b_2:
- $v = x_1a_1 + x_2a_2$
- $v = y_1b_1 + y_2b_2$
- Thus,
- $v = y_1b_1 + y_2b_2$
- $= y_1(2a_1 + 3a_2) + y_2(5a_1 + 7a_2)$
- $= (2y_1 + 3y_2)a_1 + (5y_1 + 7y_2)a_2$
- Comparing this equation with $v = x_1a_1 + x_2a_2$, we obtain that
- $x_1 = 2y_1 + 5y_2$
- $x_2 = 3y_1 + 7y_2$
- The change of the components of a vector v in V is in a way inverse to the change of basis for V.
- The components of a vector v are **contravariant** with the basis for V.
- P is called the **change-of-basis matrix** from the old basis a_1 and a_2 to the new basis b_1 and b_2.

- P is also called the **change-of-components matrix** from the new components y_1 and y_2 of a vector v to the old components x_1 and x_2 of the vector v.

- **Question by Maximilian Richter and Allison Lee**: What do you mean that "components of a vector v in V are **contravariant** with the basis for V"?
- **Answer**. An invertible matrix P changes both the basis for V and components of each vector v in V, but in opposite directions.
- The matrix P changes an **old** basis a_1 and a_2 to a **new** basis b_1 and b_2.
- The same matrix P changes the **new** components y_1 and y_2 to **old** components x_1 and x_2.

Generalize to any n-dimensional vector space V

- Let $S_1,...,S_n$ be n scalar sets over a number field F.
- The Cartesian product of $S_1,...,S_n$ defines an **n-dimensional vector space**:
- $V = S_1 \times \ldots \times S_n$.
- A vector v in V is an n-tuple of scalars:
- $v = (s_1,...,s_n)$, where s_i is a scalar in the scalar set S_i.

- **Basis for V**
- There are many lists of **n linearly independent vectors** in V.
- Any n linearly independent vectors $a_1,\ldots, a_n$ are a **basis** for V.
- Each vector in the basis is called a **basis vector**.

- **Column of a vector v relative to a basis for V**
- Any vector v in V is a **linear combination** of the basis vectors:
- $v = x_1a_1 +...+ x_na_n$.
- The numbers $x_1, \ldots, x_n$ in F are called the **components** of the vector v relative to the basis $a_1, \ldots, a_n$.
- Write the components $x_1, \ldots, x_n$ as a column x, called the **column** of the vector v relative to the basis $a_1, \ldots, a_n$.
- A vector v is **absolute**, not relative to any basis for V.
- The column x of the vectors v is **relative** to a basis for V.

- **Standard basis for V**
- Let u_i be a **nonzero scalar** in the scalar set S_i.
- Here is a list of n vectors in V:
- $a_1 = (u_1, 0, \ldots, 0)$,
- $a_2 = (0, u_2, \ldots, 0)$,
- ……
- $a_n = (0, 0, \ldots, u_n)$.
- These n vectors are linearly independent, and are therefore a basis for V.
- We call these n vectors a **standard basis** for V.
- Each scalar set has infinitely many nonzero scalars.

- There are infinitely many standard bases for a vector space V.

- **Standard column of a vector v**
- Recall that a vector v in V is an n-tuple of scalars:
- $v = (s_1,...,s_n)$, where s_i is a scalar in the scalar set S_i.
- Recall that any scalar s_i in S_i is proportional to a nonzero scalar u_i in S_i:
- $s_1 = x_1u_1$
- ……
- $s_n = x_nu_n$
- The number x_i in F is the magnitude of the scalar s_i in S_i relative to the nonzero scalar u_i in S_i.
- Write the vector $v = (s_1,...,s_n)$ as $v = (x_1u_1,...,x_nu_n)$. Thus,
- $v = x_1a_1 +...+ x_na_n$
- We call the numbers $x_1,...x_n$ the **standard components** of the vector v relative to a standard basis $a_1, …, a_n$.
- We call the column formed by $x_1,...x_n$, written x, a **standard column** of the vector v relative to a standard basis $a_1, …, a_n$.

- **Change of basis**
- Let $a_1,…, a_n$ be an old basis for V, not necessarily a standard basis.
- Let $b_1, …, b_n$ be a new basis for V.
- Each vector in V is a linear combination of basis vectors. In particular, b_j is a vector in V. Write
- $b_j = P_{1j}a_1 + … + P_{nj}a_j$.
- The numbers $P_{1j}, ..., P_{nj}$ are the components of a new basis vector b_j relative to the old basis $a_1, …, a_n$.
- Convention: list the components of the new basis vector b_j relative to the old basis $a_1, …, a_n$ as **column** j of a matrix P.
- By this convention, the matrix P appears transposed in relating the bases.
- The matrix P is called the **change-of-basis matrix**.
- The change-of-basis matrix P maps the old basis $a_1, …, a_n$ to the new basis $b_1, …, b_n$.
- Every change-of-basis matrix is **invertible**, and every invertible matrix is a change-of-basis matrix.
- Here is a procedure to generate any basis for V: start with a basis for V, and map this basis to another basis for V by using an invertible matrix.
- For example, we can start with a standard basis for V.

- **Change of the column of a vector under change of basis**
- Any vector v in V is a linear combination of the vectors in a basis for V:
- $v = x_1a_1 +...+ x_na_n$.
- The components $x_1,...x_n$ of the vector v relative to the basis $a_1,...,a_n$ form a column x.
- $v = y_1b_1 +...+ y_nb_n$.
- The components $y_1,…,y_n$ of the vector v relative to the basis $b_1,…,b_n$ form a column y.

- The same matrix P maps the components $y_1, \dots, y_n$ to the components $x_1, \dots, x_n$:
- $x_i = P_{i1}y_1 + \dots + P_{in}y_n$.
- Also write $x = Py$.
- P is also called a **change-of-components matrix**.
- That is, the same matrix P changes an old basis $a_1,\dots,a_n$ to a new basis $b_1,\dots,b_n$, and also changes the new column y of a vector v to the old column of the vector v.
- The column of a vector v in V is contravariant with the basis for V.

- **Inverse change**
- The inverse matrix P^{-1} maps the basis $b_1,\dots,b_n$ to the basis $a_1,\dots,a_n$, and maps the components $x_1,\dots,x_2$ to components $y_1,\dots,y_2$.
- $a_j = (P^{-1})_{1j}b_1 + \dots + (P^{-1})_{nj}b_j$.
- $y = P^{-1}x$.

- **Example 3, p. 242 of Lay**.
- Study this example on your own.
- The statement of the problem gives two bases for a vector space using standard columns.
- Write the vectors in one basis as a linear combinations of the vectors in another basis.
- That is, find the change-of-basis matrix.
- Use the row reduction algorithm.

Scalar vs. vector

Scalar set S 1-dimensional	**Vector space** V n-dimensional
Scalar s	**Vector** v
Any nonzero scalar b is a **unit** for scalar set.	Any n linearly independent vectors $b_1,\dots, b_n$ are a **basis** for vector space.
Magnitude of a scalar s relative to a unit, x	**Components** of a vector v relative to a basis, $x_1,\dots, x_n$
Proportionality $s = xb$	**Linear combination** $v = x_1b_1 + \dots + x_nb_n$
Scalar-scalar addition:	**Vector-vector addition**:

If s ∈ S and t ∈ S, then s + t ∈ S.	If v ∈ V and w ∈ V, then v + w ∈ V.
Number-scalar multiplication: If c ∈ R and s ∈ S, then cs ∈ S.	**Number-vector multiplication**: If c ∈ R and v ∈ V, then cv ∈ V.

- A scalar set is a one-dimensional vector space.
- An n-dimensional vector space is the Cartesian product of n scalar sets.
- Each scalar is a thing.
- Each vector is a thing.
- One thing can be a scalar or a vector, depending on the set to which the thing belongs.
- A chicken is a scalar in the chicken set.
- A chicken is the (1-head, 2 feet) vector in the head-foot space.

- **Change of unit for a scalar set S**.
- Any **nonzero scalar** in a scalar set S is a unit for S.
- Any two units a and b for a scalar set S are related by a **nonzero number** p:
- b = pa.
- A scalar s in S is proportional to either unit.
- s = xa, where x is **the magnitude of the scalar s relative to the unit a**.
- s = yb, where y is the magnitude of the scalar s relative to the unit b.
- The same number p also relates the two magnitudes of the scalar s:
- x = py.
- The magnitude of a scalar is contravariant with the unit for the scalar set.

Unit and basis do not affect real things, but affect magnitudes and components

- A piece of gold is a real thing, and remains to be the same thing after we change the unit for the gold set from 1 gram of gold to 100 atoms of gold. But the change of unit does change the magnitude of the piece of gold.
- A piece of gold-silver is a real thing, and remains to be the same thing after we change the basis for the gold-silver space. But the change of basis does change the components of the gold-silver vector.
- The displacement vector from this lecture hall to Harvard Square is a real thing, and remains to be the same thing after we rotate coordinate axes. But the change of basis does change the components of the displacement vector.

- **Unit and basis are necessary evils**.

- Necessary because a choice of a unit or basis maps real things (scalars and vectors) to virtual things (numbers), and numbers are much easier to work with for computers (and people) than real things.
- Evil because the magnitude of the same scalar change when unit changes, and the components of the same vector change when basis changes.

- So long as we know the unit and basis, we can translate the numbers back to the real things.
- Even a coordinate-free thinker cannot live without coordinates.

- **Coordinate-free thinker**
- A professor called himself a coordinate-free thinker.
- He went for sailing.
- A storm came.
- He radioed coast guards for help.
- “What are your coordinates?” A coast guard asked.
- “I cannot tell you. I'm a coordinate-free thinker.” Answered the professor.
- To this day we don't know the last coordinates of the professor.

- When you have a piece of gold, you can enjoy it without knowing any scalar set or unit. But how do you write to a friend to tell her what you have?
- You have to make sure that she knows what gold set is and what a gram is.

- **A vector as a thing itself, or as a tuple of scalars**.
- Specifying an n-dimensional vector space by the Cartesian product of n scalar sets means choosing a basis for the vector space.
- We even call the basis the standard basis for the vector space!
- Each vector is an n-tuple of scalars.
- Is this practice a necessary evil?
- Who can resist the temptation to use (1 chicken, 0 rabbit) and (0 chicken, 1 rabbit) as a basis for the chicken-rabbit space?
- Do you really wish to talk about the chicken-rabbit space with some other basis, or with no basis?

The matrix of a linear map

- **Chicken-rabbit column, head-foot column, and the matrix of a linear map**
- The Cartesian product of the chicken set and the rabbit set defines a two-dimensional vector space, called the chicken-rabbit space.
- An ordered pair of some number of chickens and some number of rabbits is called a chicken-rabbit column x.
- The Cartesian product of the head set and the foot set defines a two-dimensional vector space, called the head-foot space.

- An ordered pair of some number of heads and some number of feet is called a head-foot column y.
- Four facts of nature:
- A chicken has 1 head.
- A chicken has 2 feet.
- A rabbit has 1 head.
- A rabbit has 4 feet.
- The four numbers define four scalar-scalar linear maps.
- Tabulate the four numbers as a matrix A:

1	1
2	4

- The matrix A maps a chicken-rabbit column x to a head-foot column y:
- Ax = y.
- In writing the columns x and y and the matrix A, we have assumed a standard basis for the chicken-rabbit space and a standard basis for the head-foot space.
- The columns and the matrix come out “naturally”, with no fuss.

- **The matrix of a linear map relative to spa standard basis**
- We now express this idea in general terms.
- The Cartesian product of n scalar sets defines an n-dimensional vector space V
- Let x be a column of a vector v in V.
- The Cartesian product of m scalar sets defines an m-dimensional vector space W.
- Let y be a column of a vector w in W.
- Consider a linear map from one of the n scalar sets to one of the m scalar sets.
- Tabulate these m x n scalar-scalar linear maps as a matrix A.
- The matrix A maps the column x to the column y:
- Ax = y.
- In writing the columns x and y and the matrix A, we have assumed a standard basis for V and a standard basis for W.
- We call x the **standard column** of a vector v in V relative to a standard basis for V.
- We call y the **standard column** of a vector w in W relative to a standard basis for W.
- We call the matrix A the **standard matrix** of the linear map T: V → W relative to a standard basis for V and a standard basis for W.

- **The matrix of a linear map relative to arbitrary bases.**
- This idea will appear in section 5.4 in Lay, but seems natural to appear here.
- **The column of a vector v** in a vector space V is relative to a basis for V.
- Similarly, **the matrix of a linear map T** from a vector space V into a vector space W is relative to a basis for V and a basis for W.
- We now define the matrix of T relative to an arbitrary basis for V and an arbitrary basis for W.

- Let V be an n-dimensional vector space.
- Let W be an m-dimensional vector space.
- Let T: V → W be a linear map.
- Let $b_1,...,b_n$ be a basis for V.
- Let $c_1,...,c_m$ be a basis for W.
- Observe that b_j is a vector in V, and $T(b_j)$ is a vector in W.
- Any vector in W is a linear combination of the basis $c_1,...,c_m$. Write
- $T(b_j) = A_{1j}c_1 + \ldots + A_{mj}c_m$ for j = 1,..., n.
- This equation defines the components of the vector $T(b_j)$ relative to the basis $c_1,...,c_m$ for W.
- Tabulate the m x n numbers A_{ij} as a matrix A, called the **matrix of the linear map** T: V → W relative to the basis $b_1,...,b_n$ for V and the basis $c_1,...,c_m$ for W.
- Convention: list the components of the vector $T(b_j)$ relative to the basis $c_1, \ldots, c_m$ as **column** j of the matrix A.
- A linear map T: V → W is **absolute**, independent of any basis for V and any basis for W.
- The matrix A of the linear map T: V → W is **relative** to a basis for V and a basis for W.

- **Columns of two vectors. Matrix of a linear map**.
- T maps a vector v in V to a vector w in W:
- w = T(v).
- The vector v in V is a linear combination of the basis vectors $b_1,...,b_n$ for V:
- $v = x_1b_1 + \ldots + x_nb_n$, where $x_1,..., x_n$ are components of the vector v in V relative to the basis $b_1,...,b_n$.
- The vector w in W is a linear combination of the basis vectors $c_1,...,c_m$ for W:
- $w = y_1c_1 + \ldots + y_mc_m$, where $y_1,..., y_m$ are components of the vector w in W relative to the basis $c_1,...,c_m$.
- w = T(v)
- $= T(x_1b_1 + \ldots + x_nb_n)$
- $= x_1(A_{11}c_1 + \ldots + A_{m1}c_m) + \ldots + x_n(A_{1n}c_1 + \ldots + A_{mn}c_m)$
- Collecting terms for each basis vector, we obtain that
- $w = (A_{11}x_1 + \ldots + A_{1n}x_n)c_1$
- $+\ldots$
- $+ (A_{m1}x_1 + \ldots + A_{mn}x_n)c_m$
- Comparing this equation with $w = y_1c_1 + \ldots + y_mc_m$, we obtain that
- $y_1 = A_{11}x_1 + \ldots + A_{1n}x_n$
- …
- $y_m = A_{m1}x_1 + \ldots + A_{mn}x_n$
- These m equations map the components $(x_1,..., x_n)$ of the vector v in V to the components $(y_1,..., y_m)$ of the vector w in W.
- Write $x_1,..., x_n$ as a column x.
- Write $y_1,..., y_m$ as a column y.
- Write A_{ij} as a matrix A.
- The above m scalar equations become a matrix equation y = Ax.

- That is, the matrix A of a linear map T sends the column x of a vector v in V to the column y of a vector in W.
- a choice of a basis for V and a basis for W turns the linear map w = T(v) to a matrix equation y = Ax.

Change of the matrix of a linear map under changes of bases

- **Theorem**. Let A be the matrix of a linear map T: V → W relative to a basis for W and a basis for V. When the basis for W changes by an invertible matrix Q, and the basis for V changes by an invertible matrix P, the matrix of the linear map T changes from A to $Q^{-1}AP$.
- The matrix of a linear map T: V → W is **contravariant** with the basis for W, but is **covariant** with the basis for V.

- **Proof**.
- The linear map T sends a vector v in V to a vector w in W:
- w = T(v).
- Let x be the column of the vector v in V relative to an old basis for V.
- Let X be the column of the vector v in V relative to a new basis for V.
- An n x n invertible matrix P maps the old basis for V to the new basis for V, so that
- x = PX.
- Let y be the column of the vector w in W relative to an old basis for W.
- Let Y be the column of the vector w in W relative to a new basis for W.
- An m x m invertible matrix Q maps the old basis for W to the new basis for W, so that
- y = QY.
- Let A be the matrix of the linear map T relative to the old basis for V and the old basis for W.
- The matrix A maps the column x to the column y:
- y = Ax
- Thus, QY = APX.
- Solving for Y, we obtain that
- $Y = Q^{-1}APX$
- Thus, when the basis for V changes by a matrix P, and the basis for W changes by a matrix Q, the matrix of the linear map T changes from A to $Q^{-1}AP$.

- **Example**. **Gold-silver jewelry**.
- A goldsmith makes two types of jewelry, called Alice and Bob.
- Each Alice contains 7 grams of gold and 3 grams of silver.
- Each Bob contains 5 grams of gold and 2 grams of silver.

- **The data define a linear map from one vector space to another vector space**.
- T: (Alice-Bob space) → (gold-silver space).
- V = Alice-Bob space

- W = gold-silver space
- Let x_1 be the number of Alices, and x_2 be the number of Bobs.
- Let y_1 be the number of grams of gold, and y_2 be the number of grams of silver.
- Equation of gold: $y_1 = 7x_1 + 5x_2$
- Equation of silver: $y_2 = 3x_1 + 2x_2$
- These two scalar equations define the linear map T.
- The matrix A of the linear map T is

7	5
3	2

- **A standard Alice-Bob basis, and a standard gold-silver basis**
- In writing the linear map T as the two scalar equations, we have chosen a standard basis for the Alice-Bob space and the gold-silver space.
- The chosen standard Alice-Bob basis is
- (1 Alice, 0 Bob),
- (0 Alice, 1 Bob).
- The numbers x_1 are x_2 are the components of a Alice-Bob vector v relative to this standard Alice-Bob basis.
- The chosen standard gold-silver basis is
- (1 gram of gold, 0 silver),
- (0 gold, 1 gram of silver).
- The numbers y_1 are y_2 are the components of a gold-silver vector w relative to this standard gold-silver basis.
- The matrix A given above is the matrix of the linear map T relative to the standard Alice-Bob basis and the standard gold-silver basis.

- **A new Alice-Bob basis:**
- (2 Alices, 3 Bobs),
- (4 Alices, 5 Bobs).
- Here we keep the old gold-silver basis.
- The new Alice-Bob basis relates to the standard (old) Alice-Bob basis as
- (2 Alices, 3 Bobs) = 2(1 Alice, 0 Bob) + 3(0 Alice, 1 Bob)
- (4 Alices, 5 Bobs) = 4(1 Alice, 0 Bob) + 5(0 Alice, 1 Bob)
- This change of basis corresponds to the change-of-basis matrix P

2	4
3	5

- Verify that P is invertible.
- We follow the convention:
- Column 1 of P lists components of the first new basis vector relative to the old basis vectors.

- Column 2 of P lists components of the second new basis vector relative to the old basis vectors.
- Let X_1 and X_2 be the components of a Alice-Bob vector v relative to the new Alice-Bob basis.
- The change of Alice-Bob basis causes the change of the components of every Alice-Bob vector v:
- $x_1 = 2X_1 + 4X_2$
- $x_2 = 3X_1 + 5X_2$
- We next express the components of a gold-silver vector w relative to the old gold-silver basis in terms of the components of the Alice-Bob vector v relative to the new Alice-Bob basis:
- $y_1 = 7x_1 + 5x_2 = 7(2X_1 + 4X_2) + 5(3X_1 + 5X_2) = (7 \times 2 + 5 \times 3)X_1 + (7 \times 4 + 5 \times 5)X_2$
- $y_2 = 3x_1 + 2x_2 = 3(2X_1 + 4X_2) + 2(3X_1 + 5X_2) = (3 \times 2 + 2 \times 3)X_1 + (3 \times 4 + 2 \times 5)X_2$
- Thus, the linear map T sends the components of the Alice-Bob vector relative to the new basis to the components of the gold-silver vector relative to the old basis as
- $y_1 = 29X_1 + 53X_2$
- $y_2 = 12X_1 + 22X_2$
- The coefficients in the above two equations give the matrix of linear map T relative to the new Alice-Bob basis and the old gold-silver basis.
- You can confirm that the change of basis changes the matrix of the linear map T from A to AP.
- The above calculation also lets you see where this change A → AP comes from.

- **A new gold-silver basis:**
- (1 grams of gold, 2 grams of silver),
- (3 grams of gold, 4 grams of silver).
- Here we keep the old Alice-Bob basis.
- The new gold-silver basis relates to the standard (old) gold-silver basis as
- (1 grams gold, 2 grams of silver) = 1 (grams gold, 0 silver) + 2 (0 gold, 1 gram of silver)
- (3 grams gold, 4 grams of silver) = 3 (grams gold, 0 silver) + 4 (0 gold, 1 gram of silver)
- This change of basis corresponds to a matrix Q

1	3
2	4

- Verify that this matrix is invertible.
- Let Y_1 and Y_2 be the components of a gold-silver vector w relative to the new basis.
- The change of gold-silver basis causes the change of the components of every gold-silver vector w:
- $y_1 = 1Y_1 + 3Y_2$
- $y_2 = 2Y_1 + 4Y_2$
- Invert this linear relation, and we obtain that
- $Y_1 = -2y_1 + 1.5y_2$

- $Y_2 = y_1 - 0.5y_2$
- We next express the components of a gold-silver vector relative to the new basis in terms of the components of Alice-Bob vector relative to the old basis:
- $Y_1 = -2(7x_1 + 5x_2) + 1.5(3x_1 + 2x_2) = (-2 \times 7 + 1.5 \times 3)x_1 + (-2 \times 5 + 1.5 \times 2)x_2$
- $Y_2 = (7x_1 + 5x_2) - 0.5(3x_1 + 2x_2) = (7 - 0.5 \times 3)x_1 + (5 - 0.5 \times 2)x_2$
- Thus, the linear map T sends the components of the Alice-Bob vector relative to the old basis to the components of the gold-silver vector relative to the new basis as
- $Y_1 = -9.5x_1 - 7x_2$
- $Y_2 = 5.5x_1 + 4x_2$
- The coefficients in the above two equations give the matrix of linear map T relative to the old Alice-Bob basis and the new gold-silver basis.
- You can confirm that the change of basis changes the matrix of T from A to $Q^{-1}A$.
- The above calculation also lets you see where this change $A \rightarrow Q^{-1}A$ comes from.

- **Allison Lee** pointed out an error in my previous calculation, which I now corrected. She also asked me what would happen if bases for both spaces are new. I added the following bullets.
- **New Alice-Bob basis and new gold-silver basis**
- We now need to relate Y_1 and Y_2 to X_1 and X_2 .
- $Y_1 = -9.5(2X_1 + 4X_2) - 7(3X_1 + 5X_2)$
- $Y_2 = 5.5(2X_1 + 4X_2) + 4(3X_1 + 5X_2)$
- Completing the calculation, we obtain that
- $Y_1 = -40X_1 - 73X_2$
- $Y_2 = 23X_1 + 42X_2$
- The coefficients in the above two equations give the matrix of the linear map T relative to the new Alice-Bob basis and the new gold-silver basis.
- You can confirm that the matrix of the linear map T relative to the new Alice-Bob basis and the new gold-silver basis is $Q^{-1}AP$.

- **Map things to numbers**.
- Write the column of x_1 are x_2 as x.
- Write the column of y_1 are y_2 as y.
- Once we choose the bases for the two vector spaces, a Alice-Bob vector v becomes a column x, a gold-silver vector w becomes a column y, and the linear map w = T(v) becomes a matrix equation y = Ax.
- w, v, T are things.
- y, x, A are entries of numbers.
- A choice of a basis for V and a basis for W maps things to numbers.
- A different choice of basis for V and a basis for W maps the same things to different numbers.

Lay 2.8, 4.1 Subspace

- Let V be a real vector space.
- A subset H of V is called a **subspace** of V if H is a vector space.
- For every x and y in H, x + y is also in H. That is, H is closed under vector-vector addition.
- For every vector x in H and every real number c, cx is also a vector in H. That is, H is closed under number-vector multiplication.

- Subspace H is a vector space in its own right.
- We should identify its dimension m, and identify m linearly independent vectors as a basis for H.

- **Examples of subspace of V**
- The zero vector in V is a subspace of V. The dimension of the subspace is 0.
- The space V itself is a subspace of V.
- A powerful way to build subspaces is to use spans.
- Let $v_1,\ldots, v_k$ be a list of k linearly independent vectors in V. Span $(v_1,\ldots, v_k)$ is a k-dimensional subspace of V.
- Span $(v_1,\ldots, v_k)$ means all linear combinations of $y = c_1v_1 + c_2v_2 + \ldots + c_kv_k$, where $c_1,\ldots, c_k$ are any real numbers.
- Given a set of vectors $v_1,\ldots, v_k$ as columns of numbers, use the row reduction algorithm to see if these vectors form a linearly independent set.
- See below for a discussion of the column space.
- Let v be a nonzero vector in V. We have defined Span (v) as the collection of all vectors of the form cv, where c is any real number. Span (v) is a scalar set, with v as a unit for the scalar set. Thus, span (v) is a one-dimensional subspace of V.
- In the head-foot space, v = (1 head, 2 feet) is a vector, identified with a chicken. Span (v) means the set of all vectors cv = c(1 head, 2 feet), where c is any real number. The subspace is a line in the head-foot space, identified with the chicken set.
- In a three-dimensional vector space V, two linearly independent vectors u and v span a plane. Span (u,v) is a two-dimensional subspace of V.

- Watch this video for an example on subspaces.
- Watch this video for another example on subspaces.
- Video: find basis for a subspace

Week 8 (March 20, 22, 24)

Reading and watching

- A video on inverse, column space, and null space
- Video by Gilbert Strang on independence, basis, and dimension.
- Wiki Row and column spaces
- Video on Kernel T and Image T (Kernel means null space. Image means column space)
- A video on eigenvalues and eigenvectors
- last minute of this video to learn the MATLAB command [V,L]=eig(A).
- Lay Chapter 4
- Lay 5.1-5.2

Column space, null space, row space

- **Lay 2.9 Dimension and rank**
- **Lay 4.2 Null space, column space, linear map**
- **Lay 4.3 Basis**
- **Lay 4.5 Dimension**
- **Lay 4.6 Rank and nullity**

- A is a matrix of real numbers.
- A is also a linear map A: $R^n \to R^m$.
- Ax = y.
- Domain of the linear map is R^n.
- Codomain of the linear map is R^m.

- **Column space**
- Each column of A is a vector in R^m (i.e., the codomain of the linear map A).
- The **span of all columns of A** is a subspace of R^m.
- This subspace is called the **column space** of A, written **Col A**.
- Column space is also called the **range** of the linear map A.
- The span of all columns of A, namely, Col A, means the set of all linear combinations:
- x_1(column 1) +...+ x_n(column n),
- where $x_1,..., x_n$ are any real numbers.
- Row operations do not change the linear dependence relations among the columns of A.
- But columns in rref A may not be vectors in Col A.
- The pivot columns of A are a basis for the Col A.
- The **dimension** of the Col A equals the number of pivot columns in A.
- The number of pivot columns in A is called the **rank**.

- **Null space.**
- Consider the homogeneous equations, Ax = o.
- Each solution x to the equation Ax = o is a vector in R^n (i.e., the domain of the linear map A).
- The solution set of the homogeneous equations, Ax = o, is a subspace of R^n.
- This subspace is called the **null space** of A, written **Nul A**.
- The null space is called the **kernel** of A.
- The solution set to the homogeneous equations, Ax = o, written in the parametric form, gives a basis for Nul A.
- The **dimension** of Nul A equals the number of nonpivot columns in A
- The dimension of Nul A is called the **nullity**.

- **Rank-nullity theorem**
- (number of pivot columns) + (number of nonpivot columns) = (number of columns)
- (dimension of Col A) + (dimension of Nul A) = (number of columns in A).
- rank A + nullity A = n

- **Row space.**
- Each row of the matrix A is a vector in R^n (i.e., the domain of the linear map A).
- The **span of all rows of the matrix A** is a subspace of R^n.
- The subspace is called the **row space** of A, written **Row A**.
- Row operations keep all rows in Row A.
- But row operations may change linear dependence relations among rows.
- The pivot rows of rref A give a basis for Row A.
- The **dimension** of Row A equals the number of pivots, which is the same as the rank, and is the same as the dimension of Col A.
- (dimension of Row A) = (dimension of Col A)

- **Example**. Consider a 3 x 4 matrix A

1	3	3	2
2	6	9	7
-1	-3	3	4

- Using the row reduction algorithm, we find rref A:

1	3	0	-1
0	0	1	1
0	0	0	0

- **Col A**. Column 1 and column 3 are pivot columns.
- The dimension of Col A is **rank A = 2.**

- The two pivot columns of A give a **basis for Col A**.
- We write each basis vector as a column in R^3.

1	3
2	9
-1	3

- Row operations retain linear dependence relations among **columns**.
- Denote the columns of A by a_1, a_2, a_3, a_4. Denote the columns of rref A by b_1, b_2, b_3, b_4.
- In rref A, we see that $b_2 = 3b_1$ and $b_4 = -b_1 + b_3$.
- In A, we see that $a_2 = 3a_1$ and $a_4 = -a_1 + a_3$.

- **Important! Use the pivot columns of A as a basis for Col A.**
- Columns of rref A may **not** be in Col A, and cannot be used as a basis for A.
- For example, the first column of rref A is not in Col A.
- Indeed, the three vectors b_1, a_1, a_3 are linearly independent.
- Confirm this linear independence by row reduction of the following matrix:

1	1	3
0	2	9
0	-1	3

- **Nul A**. Column 2 and column 3 correspond to two free variables.
- The dimension of Nul A is **nullity A = 2**.
- Translate the rref A back to a system of equations, and insert one equation for every free variable:
- $x_1 + 3x_2 - x_4 = 0$
- $x_2 = x_2$
- $x_3 + x_4 = 0$
- $x_4 = x_4$
- Write the solution set in the parametric vector form, and we obtain the **basis for Nul A.**
- We write each basis vector for Nul A as a column in R^4.

-3	0
1	0
0	-1
0	1

- **Important! Use the solutions in the parametric vector form as a basis for Nul A.**

- **Row A**. Row 1 and row 2 of rref A are nonzero rows.

- These two rows give **a basis for Row A**.
- The dimension of Row A is 2.
- We write each basis vector for Row A as a column in R^4.

1	0
3	0
0	1
-1	1

- **Important**! **Use the nonzero rows of rref A as a basis for Row A**.
- Row operations may change linear dependence relations among **rows**. (After all, row operation can even swap rows.)
- See **Example 2 on page 231 of Lay.**
- Study this example on your own.
- Very testable material!

- **Lay 4.8 Applications to difference equations**
- **Lay 4.9 Applications to Markov chains**
- These two sections illustrate linear algebra in some applications.
- Read the two sections on your own.
- Work through supplementary problems before Midterm II.

- **Jacob Bindman's comment**. When I read the textbook in preparing for Midterm II, I see the same ideas repeated many times. Sometimes I am not even sure if I have read a certain part or not.
- **Response**: Indeed, the first four chapters all come from a single motivation (solving a system of linear algebraic equations) and a single algorithm (row reduction).
- It is remarkable that the single motivation and single algorithm lead to abstract ideas like vector space, linear map, linear combination, span, subspace, linear independence, basis, column, matrix,…
- The basic motivation and the algorithm came to people at least 2000 years ago. The abstract ideas, however, were formulated only in last 150 years.
- These ideas became standard curriculum in college only after computers were developed, after 1950s.
- It is also remarkable that these abstract ideas lead to other useful ideas, as we will explore in the next three chapters.

Operator

- When the domain and the codomain of a linear map T are the same vector space V, the linear map T is called a **linear operator** on V.
- Write T: V → V.

- This course seldom talks about nonlinear operators. For brevity, unless otherwise specified, an operator means a linear operator.
- We next consider several examples.

- **An n x n matrix of real numbers**
- A: $R^n \to R^n$.

- In many applications, an n x n matrix A is the matrix of an operator T on an n-dimensional real vector space V relative to a basis for V.

- The linear map **T: (chicken-rabbit space) → (head-foot space)** is not an operator, because the domain, the chicken-rabbit space, is different from the codomain, the head-foot space.
- This linear map is not an operator even though both the domain and the codomain are two-dimensional real vector spaces.

- **An operator on the product space of an economy**
- This example is described in Lay 1.6 and 2.6, as well as in Suo Notes on Input-output economics.
- An economy makes n products.
- Various amounts of product i form a scalar set, called a **product set** S_i.
- The Cartesian product of the n scalar sets, $V = S_1 \times \ldots \times S_n$, is an n-dimensional vector space, called the **product space** of the economy.
- Let x_i be the total output of product i in a year.
- x_i is a scalar in the scalar set S_i.
- Denote the total outputs of all products by $x = (x_1,...,x_n)$.
- We call x the **total-output vector**, which is a vector in the product space V.
- Let d_i be the final demand for product i in the year.
- d_i is a scalar in the scalar set S_i.
- Denote the final demands of all products by $d = (d_1,...,d_n)$.
- We call d the **final-demand vector**, which is also a vector in the product space V.
- **The Leontief input-output model creates a linear operator on the product space,**
- A: $V \to V$.
- The operator A maps the total-output vector x of a year to the final-demand vector d of the year:
- $Ax = d$.
- The Leontief input-output model relates the matrix A to the matrix of input coefficients C:
- $A = I - C$.
- C is typically known, obtained from the economic data of a past year.

- **Difference equations**
- Example 7 on p. 250 of Lay.
- See the operator on the owl-rat space later.

Lay 5.1, 5.2 Eigenvalues and eigenvectors. Characteristic polynomial

- Let V be an n-dimensional vector space over a number field F.
- Let T: V → V be an operator on V.
- A number λ in F and a nonzero vector v in V are called corresponding **eigenvalue** and **eigenvector** of T if
- T(v) = λv**.**
- (λ, v) is called an **eigenpair** of T.
- The set of all eigenvalues of T is called the **spectrum** of T.

- **Eigenspace**
- If (λ, v) is an eigenpair of T, then, for any number c in F, (λ, cv) is also an eigenpair of T.
- If (λ, v) and (λ, w) are two eigenpairs of T, then (λ, v + w) is also an eigenpair of T.
- Thus, the set of all eigenvectors corresponding to an eigenvalue λ is a subspace of V, called the **eigenspace** corresponding to λ.

- **Basis for a vector space**. Let V be a vector space over a number field F.
- Any n linearly independent vectors $a_1,...,a_n$ in V are a basis for V.

- **Column of a vector**. A vector v in V is a linear combination of the basis vectors for V:
- $v = x_1a_1 + \ldots + x_na_n$
- The numbers $x_1,...,x_n$ in F are the **components** of the vector v relative to the basis basis $a_1,...,a_n$.
- Write the numbers $x_1,...,x_n$ in a column x, called the **column of the vector** v relative to the basis basis $a_1,...,a_n$.
- Thus, $x \in F^n$.

- **Matrix of an operator**. Let T: V → V be an operator on V.
- $T(a_j)$ is a vector in V, and is a linear combination of the basis vectors for V:
- $T(a_j) = A_{1j}a_1 + \ldots + A_{nj}a_n$
- The numbers A_{ij} in F form a matrix A, called the **matrix of the operator** T relative to the basis basis $a_1,...,a_n$.
- Thus, $A: F^n \to F^n$.

- Relative to the basis $a_1,...,a_n$, T(v) = λv becomes Ax = λx.

- **Find eigenvalues and eigenvectors by hand**
- Ax = λx is a homogeneous equation, written as (A - λI)x = 0.
- (A - λI) is the matrix of coefficients of the homogeneous equation.
- A nontrivial solution exists if and only if

- det (A - λI) = 0, called the **characteristic equation** of the matrix A.
- det (A - λI) is an n^{th} order polynomial of λ, called the **characteristic polynomial**.
- Solve the polynomial equation to find eigenvalues, $\lambda_1,...,\lambda_r$.
- For each eigenvalue λ_i, find any corresponding eigenvector x by solving the homogeneous equation $(A - \lambda_i I)x = 0$.
- The solution set is the **eigenspace** corresponding to the eigenvalue λ_i.
- The eigenspace is the **null space** of the matrix $(A - \lambda_i I)$, and is a **subspace** in V.

- **Example**. Let a matrix A be

4	-3
2	-1

- Ax = λx is a homogeneous equation, with the matrix of coefficients:

4 - λ	-3
2	-1 - λ

- The homogeneous equation has nontrivial solutions if and only if det (A - λI) = 0.
- The characteristic polynomial is det (A - λI) = (4 - λ)(-1 - λ) - (-3)(2)
- Find roots for the quadratic equation, $\lambda^2 - 3\lambda + 2 = 0$.
- We obtain two eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 2$.
- For $\lambda_1 = 1$, solve the homogeneous equation (A - λI)x = 0, and we find the corresponding eigenvector v_1:

1
1

- For $\lambda_2 = 2$, solve the homogeneous equation (A - λI)x = 0, and we find the corresponding eigenvector v_2:

3
2

- In this example, each eigenvalue corresponds to a one-dimensional eigenspace.
- On p. 268 of Lay, Example 4 shows a two-dimensional eigenspace.

- **Find eigenvalues and eigenvectors by matlab**.
- Watch the last minute of this video to learn the MATLAB command [V,L]=eig(A).

- **Graphical interpretation of eigenvalues and eigenvectors**.
- **Jacob Bindman** asked me to draw figures for him.
- I drew this graph in class. If you missed the lecture, please talk to someone who came to the lecture.
- R^2 means the set of all ordered pairs of real numbers.

- R^2 is a two-dimensional real vector space.
- Represent R^2 by a plane.
- Represent the **zero vecto**r in R^2 by a point in the plane, and call the point the **origin**.

- **The standard basis for R^2**
- Draw two arrows from the origin to represent the standard basis for R^2, $e_1 = (1,0)$ and $e_2 = (0,1)$.
- We often draw e_1 and e_2 to be of the same length, orthogonal to each other.
- These practices are merely habits. Nether is necessary.
- That is, we can represent e_1 and e_2 by any two arrows, so long as they are not colinear.
- In class, I followed our habits.

- **A vector v in R^2**
- Represent a vector v in R^2 by an arrow from the origin.
- Draw a parallelogram, with its diagonal being v, and with its edges co-linear with two basis vectors e_1 and e_2.
- Each vector v is a linear combination of the basis vectors:
- $v = x_1e_1 + x_2e_2$
- The numbers x_1 and x_2 are the components of the vector v relative to the basis vectors e_1 and e_2.
- In the graph, x_1 and x_2 are the lengths of the edges of the parallelogram relative to the lengths of the basis vectors e_1 and e_2.
- Denote the numbers x_1 and x_2 by a column x:

x_1
x_2

- Because x_1 and x_2 are the components of the vector v relative to the standard basis, we call x the **standard column** of the vector v.

- **An operator T on R^2**
- T: $R^2 \rightarrow R^2$.
- Let A be the matrix of the operator T relative to the standard basis e_1, e_2.
- We are given the matrix A

4	-3
2	-1

- A is the **standard matrix** of the linear operator T.
- The first column of A means the components of the vector $T(e_1) = 4e_1 + 2e_2$. Draw this vector as an arrow.
- The second column of A means the components of the vector $T(e_2) = - 3e_1 - e_2$. Draw this vector as an arrow.

- **The eigenvectors of the operator**.
- Draw the two eigenvectors, $v_1 = (1,1)$ and $v_2 = (3,2)$, as two arrows.
- The corresponding eigenvectors and eigenvalues of the linear operator T:
- $T(v_1) = \lambda_1 v_1$, $\lambda_1 = 1$. Draw the vector $T(v_1)$ as an arrow.
- $T(v_2) = \lambda_2 v_2$, $\lambda_2 = 2$. Draw the vector $T(v_2)$ as an arrow.
- When v is an arbitrary vector, the vectors v and T(v) have different directions and different lengths.
- When v is a real eigenvector, $Tv = \lambda v$, the vectors v and T(v) are along the same line, but their lengths differ by a factor of λ.

Week 9 (March 27, 29, 31)

Reading and watching

- Lay sections 5.3, 5.4, 5.6
- Wiki diagonalizable matrix
- Wiki linear difference equation
- Video: An exam problem on eigenvalues
- Video: More on finding eigenvalues
- Video: similar matrices

Lay 5.3 Diagonalization

- **Linearly independent eigenvectors**
- **Theorem**. When an operator T on V has p distinct eigenvalues $\lambda_1,...\lambda_p$, the corresponding eigenvectors $v_1,..., v_p$ are linearly independent vectors in V.
- This condition is sufficient, but not necessary, for an operator T to have linearly independent eigenvectors.
- See Example 4 on p. 268 of Lay.

- Assume that an n x n matrix A has n linearly independent eigenvectors $v_1,...,v_n$.
- Let $\lambda_1,...\lambda_n$ be the corresponding eigenvalues.
- Thus, $Av_1 = \lambda_1 v_1,...,Av_n = \lambda_n v_n$.
- List the vectors as columns of matrices:
- $[Av_1,...,Av_n] = [\lambda_1 v_1,...,\lambda_n v_n]$.
- This equation is the same as
- AP = PD.
- P is an invertible matrix, and lists the n eigenvectors as columns:
- $P = [v_1,...,v_n]$.
- D is a diagonal matrix, and lists the n eigenvalues on the diagonal:

λ_1	...	0

...	...	...
0	...	λ_n

- Because the n eigenvectors are linearly independent, P is an invertible matrix.
- Right-multiply P^{-1} to the equation AP = PD, and we obtain that $APP^{-1} = PDP^{-1}$. This equation is the same as
- $A = PDP^{-1}$.

- **Definition**. A matrix A is called **diagonalizable** if a diagonal matrix D and an invertible matrix P exist, such that $A = PDP^{-1}$.

- **Theorem**. An n x n matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.
- **Proof**. See above.

- **Return to the numerical example**.
- The matrix D lists the eigenvalues along its diagonal:

1	0
0	2

- The matrix P lists eigenvectors as columns:

1	3
1	2

- The inverse matrix P^{-1} is

-2	3
1	-1

- The matrix A is diagonalizable:
- $A = PDP^{-1}$**.**

- **A property of a diagonalizable matrix**
- $A^2 = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PDIDP^{-1} = PD^2P^{-1}$.
- For any positive integer k, we have
- $A^k = PD^kP^{-1}$.
- D^k is easy to calculate:

$(\lambda_1)^k$	0
0	$(\lambda_2)^k$

Lay 4.7, 5.4 Change of basis

- We have covered the two sections in previous lectures.
- The key results are listed here.

- **Change-of-basis matrix**
- Let $a_1,\dots, a_n$ be an old basis for V.
- Let $b_1, \dots, b_n$ be a new basis for V.
- b_j is a vector in V, and is a linear combination of the old basis vectors. Write
- $b_j = P_{1j}a_1 + \dots + P_{nj}a_j$.
- The numbers P_{ij} in F form a matrix P, called the **change-of-basis matrix** from the old basis $a_1,\dots, a_n$ to the new basis $b_1, \dots, b_n$.
- The matrix P is invertible.

- **Change of the column of a vector under change of basis**.
- Let x be the column of a vector v relative to the basis $a_1,\dots, a_n$.
- Let y be the column of the same vector v relative to the basis $b_1,\dots, b_n$.
- The two columns x and y of the same vector v relate to each other:
- $y = P^{-1}x$.

- **Change of the matrix of an operator under change of basis**.
- Let A be the matrix of an operator T relative to the basis $a_1,\dots, a_n$.
- Let B be the matrix of the same operator T relative to the basis $b_1,\dots, b_n$.
- The two matrices A and B of the same operator T relate to each other:
- $B = P^{-1}AP$.

- **Similar matrices**.
- Two n x n matrices A and B are called **similar** if an invertible n x n matrix P exists, such that $B = P^{-1}AP$.
- Similar matrices A and B are the matrices of the same operator T on a vector space V relative to two bases for V.
- The invertible matrix P is a change-of-basis matrix.
- Video: examples of similar matrices.

From an arbitrary basis to an eigenbasis

- **Eigenvalue and eigenvector of an operator T on a vector space V, defined with no mention of basis for V**
- Let T be an operator on an n-dimensional vector space V over a number field F.
- A number λ in F and a nonzero vector v in V are called corresponding eigenvalue and eigenvector of the operator T if
- $T(v) = \lambda v$.

- The above definition is **absolute**, not relative to any basis for V.
- The eigenvalue λ is a number in F, and is **invariant** with the basis for V.
- The eigenvector v is a vector in V, and its column x is **contravariant** with the basis for V.
- T is an operator on V, and its matrix A is **partly contravariant and partly covariant** with the basis for V.
- These statements are understood as follows.

- **Relative to a basis for V, the column of v is x, the matrix of T is A, and T(v) = λv becomes Ax = λx.**
- Let $a_1,...,a_n$ be a basis for V.
- Let x be the column of a vector v in V relative to the basis, so that
- $v = x_1a_1 + … + x_na_n$.
- Let A be the matrix of the operator T on V relative to the basis, so that
- $T(a_j) = A_{1j}a_1 + … + A_{nj}a_n$.
- Thus,
- $T(v) = T(x_1a_1 + … + x_na_n)$
- $= x_1T(a_1) + … + x_nT(a_n)$
- $= x_1(A_{11}a_1 + … + A_{n1}a_n) + … + x_n(A_{1n}a_1 + … + A_{nn}a_n)$
- $= (A_{11}x_1 + … + A_{1n}x_n)a_1 + … + (A_{n1}x_1 + … + A_{nn}x_n)a_n$
- Consequently, the equation T(v) = λv becomes that
- $(A_{11}x_1 + … + A_{1n}x_n)a_1 + … + (A_{n1}x_1 + … + A_{nn}x_n)a_n = λ(x_1a_1 + … + x_na_n)$
- The component relative to each basis vector must be equal, so that
- $A_{11}x_1 + … + A_{1n}x_n = λx_1$
- …..
- $A_{n1}x_1 + … + A_{nn}x_n = λx_n$
- These n equations mean that
- Ax = λx.

- **Change of basis**
- Relative to a basis $a_1,...,a_n$ for V, the matrix of the operator T is A, the column of a vector v is x, and T(v) = λv becomes Ax = λx.
- Relative to another basis $b_1,...,b_n$ for V, the matrix of the operator T is B, the column of the vector v is y, and T(v) = λv becomes By = λy.
- Let P be the change-of-basis matrix from the basis $a_1,...,a_n$ to the basis $b_1,...,b_n$.
- The two columns of the same vector v relate to each other by
- $y = P^{-1}x$.
- The two matrices of the same operator T relate to each other by
- $B = P^{-1}AP$.
- Ax = λx and By = λy are equivalent, one implying the other.
- For example, start with the equation By = λy, insert $y = P^{-1}x$ and $B = P^{-1}AP$, and we obtain that $(P^{-1}AP)(P^{-1}x) = λP^{-1}x$. Thus, $P^{-1}Ax = λP^{-1}x$. Right-multiply this equation by P, and we obtain the other equation Ax = λx.

- The two equations, $Ax = \lambda x$ and $By = \lambda y$, express the same equation $T(v) = \lambda v$, relative to different bases for V.
- The equation $T(v) = \lambda v$ is absolute, not relative to any basis for V.
- The equation $Ax = \lambda x$ is relative to the basis $a_1,...,a_n$ for V.
- The equation $By = \lambda y$ is relative to the basis $b_1,...,b_n$ for V.

- **Eigenbasis**
- Assume that the operator T on V has n linearly independent eigenvectors $v_1,..., v_n$.
- Thus, the eigenvectors $v_1,..., v_n$ are a basis for V, called an **eigenbasis**.
- The corresponding eigenvalues $\lambda_1,...\lambda_n$ need not be distinct.
- Recall $T(v_1) = \lambda_1 v_1, \ldots, T(v_n) = \lambda_n v_n$.
- Thus, the matrix of the operator T relative to the eigenbasis $v_1,..., v_n$ is a **diagonal matrix** D**,** with $\lambda_1,...\lambda_n$ listed on the diagonal.

- **From an arbitrary basis to an eigenbasis**
- Let $a_1,...,a_n$ be an arbitrary basis for V.
- Let A be the matrix of an operator T on V relative to the basis $a_1,...,a_n$.
- We have solved the eigenvalue problem $Ax = \lambda x$, and obtained the columns of the eigenvectors $v_1,..., v_n$ relative to the basis $a_1,...,a_n$.
- List these columns of eigenvectors as a matrix:
- $P = [v_1,...,v_n]$.
- P is the change-of-basis matrix from the arbitrary basis $a_1,...a_n$ to the eigenbasis $v_1,..., v_n$.
- A is the matrix of the operator T relative to the arbitrary basis $a_1,...,a_n$.
- D is the matrix of the operator T relative to the eigenbasis $v_1, \ldots, v_n$.
- Recall the change of the matrix of an operator T under a change of the basis for V:
- $P^{-1}AP = D$.
- Rearrange the above equation, and we obtain that
- $A = PDP^{-1}$.
- Thus, diagonalization of a matrix means changing from an arbitrary basis to an eigenbasis.

Markov chain

- Markov chain appears in Lay 4.9, but the existence of steady state is unexplained.
- Wiki examples of Markov chains.
- Repeat an **experiment** for many **trials**.
- Each trial of the experiment produces one and only one **outcome** in the **sample space**.
- That is, the sample space is the set of all possible outcomes.
- Consider a sample space of n outcomes:
- (sample space) = {outcome 1, …, outcome n}.
- A probability distribution, also called a **probability vector,** is an n-tuple of nonnegative numbers that sum to 1.

- A probability vector is a vector in R^n.
- A Markov chain is characterized by a sequence of trials of the experiment.
- x_k = the probability vector on the k-th trial.
- A **stochastic matrix** P maps the probability vector x_k to the probability vector x_{k+1}:
- $x_{k+1} = Px_k$
- That is, the stochastic matrix P is a linear operator on the probability-vector space.
- **Convention adopted by Lay and by this course**: Entries in each **column** of P sum to 1.
- Most other textbooks and Wikipedia adopt a different convention: Entries in each **row** of P sum to 1. Wiki stochastic matrix.
- The two conventions list the same stochastic matrix by transpose.

- **Example 1, p.254 of Lay**
- Every year, some people migrate from city to suburb, some other people migrate from suburb to city. Model this migration year by year as a Markov chain, with a stochastic table:

city → city	suburb → city
city → suburb	Suburb → suburb

- For example,

0.95	0.03
0.05	0.97

- We follow the convention that entries in each column sum to 1.

Steady state of a Markov chain

- **Theorem** (Lay p. 259) If P is an n x n regular stochastic matrix, then P has a unique steady-state vector q. Further, if x_0 is any initial state and $x_{k+1} = Px_k$ for k = 0, 1, 2,..., then the Markov chain $\{x_k\}$ converges to q as k approaches infinity.

- Let us try to understand this theorem.
- Entries of each column of a stochastic matrix sum to 1.
- In general, a 2 x 2 stochastic matrix P takes the form

a	1 - b
1 - a	b

- where 0 < a < 1 and 0 < b < 1.
- The characteristic polynomial of the stochastic matrix P is det $(P - \lambda I)$
- $= (a - \lambda)(b - \lambda) - (1 - a)(1 - b)$
- $= ab - (a + b)\lambda + \lambda^2 - 1 + (a + b) - ab$
- $= (a + b)(-\lambda + 1) + (\lambda - 1)(\lambda + 1)$

- $= (\lambda - 1) (\lambda + 1 - a - b)$.
- Consequently, the characteristic equation det $(P - \lambda I) = 0$ gives two eigenvalues:
- $\lambda_1 = 1$, $\lambda_2 = a + b - 1$.
- Denote the corresponding eigenvectors by v_1 and v_2.
- Because $0 < a < 1$ and $0 < b < 1$, we note that $|\lambda_2| < 1$.
- Let the initial probability vector be
- $x_0 = c_1v_1 + c_2v_2$
- Thus, $x_k = c_1(1)^kv_1 + c_2(\lambda_2)^kv_2$
- As k increases, the first term prevails, and the second term approaches zero.

- We now generalize the above situation to n x n stochastic matrix.
- Note that $x_k = P^kx_0$
- For any stochastic matrix P, 1 is an eigenvalue.
- For a stochastic matrix of positive entries, the absolute values of all other eigenvalues are smaller than 1.
- These facts explain why a Markov chain approaches a steady state probability vector q determined by
- $Pq = q$.
- Wiki Perron-Frobenius theorem for more information.

- **Lucas Guzman's question**: What are eigenvalues for?
- **Answer**. We have just used eigenvalues to understand why every Markov chain converges.
- We next use eigenvalues to study difference equations and differential equations.
- You will see how eigenvalues appear in the rise and fall of populations, as well as in the oscillation of a pendulum.

Lay 4.8, 5.6 Difference equations

- **Specifying a discrete dynamical system**
- k = discrete time.
- V = n-dimensional real vector space, called the **state space** of the dynamical system.
- x = a vector in V, called a **state vector**. Represent a **state of the system**.
- x_k = the state vector at time k.
- The system evolves in time according to
- $x_{k+1} = Ax_k$, k = 0,1,2,...
- A is an n x n real matrix, and is given.
- x_0 is an initial condition, and is given.
- We want to determine the state of system at subsequent time, x_1, x_2,...

- **Brute force solution**
- Given x_0, compute
- $x_1 = Ax_0$,

- $x_2 = Ax_1$,
- ……
- The computer does this computation rapidly to any specified time k.

- **Solution relative to eigenbasis**
- In most applications, the state space V is an n-dimensional **real** vector space.
- Let us focus on the real vector space.
- Assume that the matrix A has **n linearly independent**, **real** eigenvectors $v_1, \ldots, v_n$.
- The corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ must be real numbers.
- Use the eigenvectors as a basis for the state space V.
- The given initial condition x_0 is a vector in V, and is a linear combination of the eigenvectors:
- $x_0 = c_1 v_1 + \ldots + c_n v_n$.
- The numbers $c_1, \ldots, c_n$ are the components of the vector x_0 relative to the eigenbasis.
- Given the vectors $x_0, v_1, \ldots, v_n$, the above system of linear equations determines $c_1, \ldots, c_n$.
- $x_1 = Ax_0 = A(c_1 v_1 + \ldots + c_n v_n) = c_1 A(v_1) + \ldots + c_n A(v_n) = c_1 \lambda_1 v_1 + \ldots + c_n \lambda_n v_n$.
- Similarly,
- $x_k = c_1(\lambda_1)^k v_1 + \ldots + c_n(\lambda_n)^k v_n$.
- That is, the numbers $c_1(\lambda_1)^k, \ldots, c_n(\lambda_n)^k$ are the components of the vector x_k relative to the eigenbasis.

- **Long-time behavior**
- This solution relative to the eigenbasis let us see the long-time behavior.
- As k increases, component i grows if $|\lambda_i| > 1$, and component i decays if $|\lambda_i| < 1$.
- **Attractor** (Lay p. 304, Fig. 1)
- **Repeller** (Lay p. 305, Fig. 2)
- **Saddle point** (Lay p. 305, Fig. 3. Lay p. 307, Fig. 4)

- **Max Richter's question**: How do we determine $c_1, \ldots, c_n$?
- Answer: x_0 is a known vector in V, and $v_1, \ldots, v_n$ are a known basis for V.
- The components $c_1, \ldots, c_n$ are determined by solving the linear algebraic equations:
- $x_0 = c_1 v_1 + \ldots + c_n v_n$.
- See the the example of an owl-rat system.

Owl-rat system

- p. 302 of Lay.
- Owls and rats in a region form a dynamical system, called the owl-rat system.
- Various numbers of owls form a scalar set, called the owl set.
- Various numbers of rats form another scalar set, called the rat set.
- k = discrete time.
- O_k = the number of owls at time k
- R_k = the number of rats at time k

- The unit for the owl set is 1 owl, the unit for the rat set is 100 rats, and the unit for time is one month.
- The populations of the two species evolve according to the equations
- $O_{k+1} = (0.5)O_k + (0.4)R_k$
- $R_{k+1} = -(0.104)O_k + (1.1)R_k$
- A story for each term:
- $(0.5)O_k$: Without feeding on rats, half of the owls will survive per month.
- $(0.4)R_k$: Rats help to increase the number of owls.
- $-(0.104)O_k$: owls eat rats.
- $(1.1)R_k$: Without owls, the number of rats increases 10% per month.
- List these four numbers by a matrix A:

0.5	0.4
- 0.104	1.1

- owl-rat space V = (owl set) x (rat set).
- A **state of the system**, called a **state vector,** is an ordered pair of numbers.
- Let x_k be the column

O_k
R_k

- The matrix A updates the owl-rat state from time k to time k + 1:
- $x_{k+1} = Ax_k$.
- This equation **evolves** the populations of the two species in time.
- We know the initial condition, i.e., the number of owls and and the number of rats at time zero:
- $x_0 = (20, 15)$.

- **Eigenvalue analysis**.
- The matrix A has two linearly independent, real eigenvectors:
- $v_1 = (10,13)$, corresponding to $\lambda_1 = 1.02$.
- $v_2 = (5,1)$, corresponding to $\lambda_2 = 0.58$.
- Each eigenvector is an owl-rat vector. For example, v_1 is an owl-rat vector of 10 owls and 1300 rats.
- The vector x_0 is the linear combination of the two eigenvectors:
- $x_0 = c_1v_1 + c_2v_2$.
- The numbers c_1 and c_2 are the components of the vector x_0 relative to the eigenvectors.
- For, $x_0 = (20, 15)$, $v_1 = (10,13)$ and $v_2 = (5,1)$, we find that $c_1 = 1$ and $c_2 = 2$.
- The owl-rat vector at time k is given by
- $x_k = c_1(1.02)^k v_1 + c_2(0.58)^k v_2$
- The numbers $c_1(1.02)^k$ and $c_2(0.58)^k$ are the components of the vector x_k relative to the eigenvectors.
- As k increases, the first component will amplify, and the second component will decay.

- Thus, for a large k, both species will increase by 2% per month.
- The region will maintain a fixed ratio, (10 units of owls):(13 units of rats).
- This conclusion is independent of the initial condition, so long as c_1 is nonzero.

Week 10 (April 3, 5, 7)

Reading and watching

- Lay Chapter 5.5, 5.7
- Lay Chapter 6
- Wiki linear dynamical system
- Wiki complex number

Lay 5.7 Differential equations

- Differential equations are a subject of great beauty broad applications.
- If you are a applied mathematics concentrator, you will take a separate course on differential equations.
- If you are an engineering concentrator, you will study differential equations in many courses.
- The same comment is applicable for many other concentrators.
- Here we will focus on the role of **linear algebra** in solving **differential equations of constant coefficients**.

- **Owl-rat system. Difference equations.**
- We have studied an owl-rat system using two **difference equations**:
- $O_{k+1} = (0.5)O_k + (0.4)R_k$
- $R_{k+1} = -(0.104)O_k + (1.1)R_k$
- k = discrete time. (Unit for time is 1 month).
- **Initial conditions**:
- $O_0 = 20$. (Unit for the owl set = 1 owl).
- $R_0 = 15$. (Unit for the rat set = 100 rats).
- Starting from the initial conditions, the difference equations update the number of owls and the number of rats monthly.
- The choice of monthly update seems to be arbitrary.

- **Owl-rat system. Differential equations.**
- We may wish to model the owl-rat system continuously in time t.
- Regard the number of owls as a function of time, O(t), and the number of rats as another function of time, R(t).
- We can convert the difference equations into differential equations.

- The above equations are written for changes in numbers per month.
- Translate the above sentences into the expressions:
- $(O_{k+1} - O_k)/(1 \text{ month}) = dO/dt$,
- $(R_{k+1} - R_k)/(1 \text{ month}) = dR/dt$.
- Translate the difference equations into **ordinary differential equations** (ODE):
- dO/dt = - (0.5)O + (0.4)R
- dR/dt = - (0.104)O + (0.1)R
- The **initial conditions** are prescribed. We know the number of owls and the number of rats at time zero:
- O(0) = 20,
- R(0) = 15.
- The two differential equations evolve the number of owls and the number of rats in time, starting at the initial populations.

- The two differential equations are **coupled**: owls eat rats.
- This fact causes two-way coupling:
- (0.4)R : Rats increase the number of owls.
- - (0.104)O : Owls reduce the number of rats.
- The coupling makes the system interesting, but makes mathematics difficult.

- **Specifying a continuous dynamical system**
- Let us abstract the owl-rat example in the language of **dynamical systems**.
- **System**: a real-world system under study. For example, the owl-rat system.
- **State of a system**: a vector v in an n-dimensional real vector space V.
- V is called a **state space,** or **phase space**. For example, V is the owl-rat space, the Cartesian product of the owl set and the rat set.
- v is called a **state vector**, or just **state**. For example, v is an ordered pair of a number of owls and a number of rats.
- t = continuous time.
- v(t) = the state of the system at time t.
- v(0) = the state of the system at time 0. v(0) is given, and is called the **Initial condition**.
- The physical system evolves according to **a system of linear, homogeneous, ordinary differential equations**:
- dv/dt = T(v).
- T = an operator T on V, which I will call the **rate-of-change operator**.
- **Problem 1**. Translate a real-world system to a system of differential equations.
- **Problem 2**. Given the initial condition v(0) and the rate-of-change operator T, determine the state of the system as a function of time,v(t).

- **State of equilibrium**
- v(t) = 0 satisfies the equation dv/dt = T(v).
- This solution says that the state is fixed at the zero vector over time.
- We call the zero vector the **state of equilibrium**.

- If the given initial state v(0) is exactly the zero vector, then the state of system will remain fixed subsequently.
- How about the given initial state v(0) is not the zero vector?
- Will the state evolve toward the state of equilibrium or away from it?

- **Phase portrait of a two-dimensional system**.
- Represent the state space V by a plane. For example, the horizontal axis represents the number of owls, and the vertical axis represents the number of rats.
- Each point in the plane represents a state of system v.
- The origin represents the state of equilibrium.
- Start from v(0), Plot a sequence of states v(t) in the plane as a curve.
- We do this incrementally in time. Knowing v(t), use the rate of change operator to obtain v(t + dt):
- v(t + dt) = v(t) + T(v)dt.
- The curve represents a path of evolution.
- In plotting the curve, we use the time t as a parameter. We cannot read the time from the curve.
- But we can mark the time on points on the curve, such as the initial time 0, 1 month, 2 month, etc.
- We can plot many paths of evolution, starting from different initial conditions.
- The phase portrait is great, but only works well for two-dimensional systems.

- **One-dimensional case**
- The region just has rats, and has no owls.
- R(t) = number of rats at time t.
- Find the function R(t) to satisfy
- dR/dt = 0.1 R, R(0) = 15.
- 0.1 = the **rate of growth** per unit of time and per unit number of rats. Rats breed, and rats die.
- 15 units of rats at time zero.
- Recall a result in calculus:
- d[exp(at)]/dt = a exp(at).
- The solution of the ODE is
- R(t) = 15 exp(0.1t)
- The number of rats grows **exponentially**.

- **n-dimensional case.**
- Now return to the general case of the n-dimensional state space V.
- Let $a_1,...,a_n$ be a basis for V.
- x = the column of a vector v in V relative to the basis $a_1,...,a_n$.
- A = the matrix of an operator T on V relative to the basis $a_1,...,a_n$.
- The differential equation dv/dt = T(v) becomes
- dx/dt = Ax.

- The matrix A is known, and the initial condition x(0) is known.
- We solve for x(t).
- **The nonzero off-diagonal entries of A couple the differential equations**.

- **Eigenbasis decouples the differential equations**
- Assume that the matrix A has **n linearly independent, real** eigenvectors $v_1, \dots, v_n$, and n corresponding **real** eigenvalues $\lambda_1, \dots, \lambda_n$.
- The real eigenvectors are a basis for the state space V.
- The state at time zero x(0) is known. Write x(0) as a linear combination of the eigenvectors:
- $x(0) = c_1v_1 + \dots + c_nv_n$.
- The numbers $c_1,\dots,c_n$ are the components of the initial state vector x(0) relative to the eigenbasis. Given the columns x(0), $v_1,\dots,v_n$, the above vector equation determines the numbers $c_1,\dots,c_n$.
- $v_1, \dots, v_n$ are the columns of the eigenvectors relative the basis $a_1,\dots,a_n$.
- Thus, the matrix $P = [v_1 \dots v_n]$ is the change-of-basis matrix from the basis $a_1,\dots,a_n$ to the eigenbasis $v_1, \dots, v_n$.
- Let y be the column of a vector v in V relative to the eigenbasis $v_1, \dots, v_n$.
- The two columns x and y of the same vector relate to each other as
- x(t) = Py(t). This equation means that
- $x(t) = y_1(t)v_1 + \dots + y_n(t)v_n$.
- In particular, at time zero,
- $x(0) = y_1(0)v_1 + \dots + y_n(0)v_n$.
- Comparing this equation with the red equation above, we obtain that
- $y_1(0) = c_1,\dots, y_n(0) = c_n$.
- The diagonal matrix D is the matrix of the operator T relative to the eigenbasis $v_1, \dots, v_n$.
- Recall that $A = PDP^{-1}$.
- The differential equation dx/dt = Ax becomes
- dy/dt = Dy
- Because D is a diagonal matrix, the n equations are **decoupled**:
- $dy_1/dt = \lambda_1y_1, \dots, dy_n/dt = \lambda_ny_n$.
- Their solutions are
- $y_1(t) = c_1 \exp(\lambda_1t), \dots, y_n(t) = c_n \exp(\lambda_nt)$.
- The solution to the system of ODEs is
- $x(t) = c_1v_1\exp(\lambda_1t) + \dots + c_nv_n\exp(\lambda_nt)$.
- The state moves away from the state of equilibrium any one of the eigenvalues is positive..
- The state moves towards the state of equilibrium if all eigenvalues are negative.

- **Summary of the idea**
- The n linearly independent eigenvectors are a basis for the state space V.
- Each component relative to the eigenbasis is decoupled from other components, and evolves in the same one as in the one-dimensional case.

Steps to solve a system of ODEs with distinct, real eigenvalues

- **Set up ODEs**
- Translate a real-world phenomenon to a system of linear ODEs: $dx/dt = Ax$.
- Prescribe a rate-of-change operator: an $n \times n$ matrix A
- Prescribe a initial state: an n-entry column $x(0)$.
- You have seen the owl-rat system.
- You will see more examples in other courses.

- **Solve eigenvalue problem**
- Remember $x(t) = v \exp(\lambda t)$, where v is an n-entry column, and λ is a number.
- Insert this form into the ODE dx/dt = Ax, and we obtain that
- $Ax = \lambda x$
- The eigenvalue problem comes out naturally!
- Here we assume that the matrix A has n real eigenvectors $v_1,...,v_n$, and n corresponding real eigenvalues $\lambda_1,..., \lambda_n$.
- Find the n corresponding eigenpairs by hand in exams, and using computers on other occasions.

- **Linear combination**
- The general solution to the ODE is a linear combination
- $x(t) = c_1v_1\exp(\lambda_1 t) + \dots + c_nv_n\exp(\lambda_n t)$.
- The numbers $c_1,...c_n$ are determined by the initial condition.
- Each eigenpair represents a **mode** of exponential growth or decay, with the eigenvalue being the rate of growth or decay, and the eigenvector being the relative magnitude of entries in the column x.

- **Match with the initial condition**
- When t = 0, the above equation becomes
- $x(0) = c_1v_1 + \dots + c_nv_n$.
- This initial condition is a system of linear algebraic equations to determine the numbers $c_1,...c_n$ for given columns $x(0)$, $v_1,...,v_n$.
- These steps solve the ODE by solving an eigenvalue problem, and a system linear algebraic equation.

Example

- **Statement of the problem**. A dynamical system is specified by a rate-of-change operator A,

4	-3

2	-1

- and by an initial state vector x(0),

5
4

- The system evolves according to the differential equation:
- dx/dt = Ax.

- **Eigenvalue problem**
- Remember x(t) = v exp(λt), where v is an n-entry column, and λ is a number.
- Insert this form into the ODE dx/dt = Ax, and we obtain that
- Ax = λx
- We have solved the eigenvalue problem for the matrix A, and find that A has two distinct, real eigenvalues.
- For the eigenvalue $\lambda_1 = 1$, the corresponding eigenvector v_1 is

1
1

- For the eigenvalue $\lambda_2 = 2$, the corresponding eigenvector v_2 is

3
2

- The initial state vector is a linear combination of the two eigenvectors:
- $x(0) = c_1v_1 + c_2v_2$.
- This vector equation corresponds to two linear algebraic equations, giving
- $c_1 = 2$, $c_2 = 1$.
- The solution of the differential equation is
- $x(t) = c_1v_1\exp(1t) + c_2v_2\exp(2t) =$

2 exp(1t) + 3 exp(2t)
2 exp(1t) + 2exp(2t)

- Both eigenvalues are positive.
- As t increases, the state vector x(t) goes away from the origin.
- The origin is a **repeller**.

- **Graphical representation. Phase portrait**
- Represent the state space by a plane.
- Represent the state of equilibrium by the origin.
- Represent a standard basis by two axes.
- Draw the two eigenvectors.

- Start with the initial state x(0), and plot a sequence of states.
- Start with another initial state, and plot another sequence of states.

- **Attractor** (Lay p. 313, Fig. 2).
- **Saddle point** (Lay p. 314, Fig. 3)

Lay 5.5 Complex eigenvalues

- If you need a reminder for complex numbers, wiki complex number.
- The following examples will illustrate the key operations of complex numbers.
- Let a matrix A be

1	-2
1	3

- The matrix A is real, but its eigenvalues can be complex numbers.
- $Ax = \lambda x$ is a homogeneous equation, with the matrix of coefficients:

$1 - \lambda$	-2
1	$3 - \lambda$

- The homogeneous equation has nontrivial solution if and only if $\det(A - \lambda I) = 0$.
- The characteristic polynomial is $\det(A - \lambda I) = (1 - \lambda)(3 - \lambda) - (1)(-2) = \lambda^2 - 4\lambda + 5$
- The characteristic equation is $\lambda^2 - 4\lambda + 5 = 0$.
- Complete square: $(\lambda - 2)^2 + 1 = 0$.
- Write $(\lambda - 2)^2 = - 1$.
- We obtain two eigenvalues:
- $\lambda_1 = 2 + i$,
- $\lambda_2 = 2 - i$.
- Because the eigenvalues are complex numbers, we now let A be an operator on C^2.
- For $\lambda_1 = 2 + i$, solve the homogeneous equation $(A - \lambda I)x = 0$ by row reduction, and we find the corresponding eigenvector v_1:

$-1 + i$
1

- For $\lambda_2 = 2 - i$, solve the homogeneous equation $(A - \lambda I)x = 0$ by row reduction, and we find the corresponding eigenvector v_2:

$-1 - i$
1

- Confirm that they are corresponding eigenvalues and eigenvectors.

Steps to solve a system of two ODEs with complex-conjugate eigenvalues

- **Set up ODEs**
- Translate a real-world phenomenon to a system of ODEs, $dx/dt = Ax$.
- Prescribe a 2 x 2 matrix A.
- Prescribe initial 2-entry column $x(0)$.
- We will go over spring-mass system.
- You will study more phenomena in later courses.

- **Solve eigenvalue problem**
- Remember $x(t) = v \exp(\lambda t)$, where v is an 2-entry column, and λ is a number.
- Insert this form into the ODE $dx/dt = Ax$, and we obtain that
- $Ax = \lambda x$
- The eigenvalue problem comes out naturally!
- Here we assume that A has complex-conjugate eigenpairs: (λ, v) and (λ^*, v^*).
- Determine the eigenvalue λ by solving the characteristic equation $\det (A - \lambda I) = 0$.
- Determine the eigenvector v by solving the homogeneous equation $(A - \lambda I)v = 0$.

- **Linear combination**
- The two eigenpairs correspond to two solutions of $dx/dt = Ax$:
- $v \exp(\lambda t)$ and $v^*\exp(\lambda^* t)$.
- But they are not real.
- The two solutions are linearly independent. Any linear combination of the two solutions gives another solution.
- In particular, look at two linear combinations:
- $\mathrm{Re}\, [v \exp(\lambda t)] = [v \exp(\lambda t) + v^*\exp(\lambda^* t)]/2$
- $\mathrm{Im}\, [v \exp(\lambda t)] = [v \exp(\lambda t) - v^*\exp(\lambda^* t)]/(2i)$
- $\mathrm{Re}\, [v \exp(\lambda t)]$ and $\mathrm{Im}\, [v \exp(\lambda t)]$ are linearly independent, real solutions of the ODE.
- Because $x(t)$ is real, write
- $x(t) = G\, \mathrm{Re}\, [v \exp(\lambda t)] + H\, \mathrm{Im}\, [v \exp(\lambda t)]$.
- where G and H are **real** numbers determined by the initial condition.
- Write $\lambda = \alpha + i \beta$, $\beta > 0$. Thus, α and β are **real** numbers.
- Write $v = g + i h$. Thus, g and h are **real** vectors in V.
- Recall Euler's formula: $\exp (i\theta) = \cos (\theta) + \sin (\theta)$.
- $v \exp (\lambda t)$
- $= (g + i h) \exp (\alpha t + i\beta t)$
- $= (g + i h) [\cos (\beta t) + i \sin (\beta t)] \exp (\alpha t)$
- $= [g \cos (\beta t) - h \sin (\beta t)] \exp (\alpha t) + i [g \sin (\beta t) + h \cos (\beta t)] \exp (\alpha t)$.
- Thus,
- $\mathrm{Re}\, [v \exp(\lambda t)] = [g \cos (\beta t) - h \sin (\beta t)] \exp (\alpha t)$

- Im [v exp(λt)] = [g sin (βt) + h cos (βt)] exp (αt)
- The solution of the differential equation is
- x(t) = G [g cos (βt) - h sin (βt)] exp (αt) + H [g sin (βt) + h cos (βt)] exp (αt).

- **Determine G and H using the initial conditions**
- When t = 0, the above equation reduces to
- x(0) = G g + H h
- This system of linear algebraic equations determines the numbers G and H for given columns x(0), g and h.

Week 11 (April 10, 12, 14)

Reading and watching

- Lay Section 5.5, 5.6, 5.7
- Video on dot product
- Lay Section 6.1

Example

- The matrix A is

1	-2
1	3

- The matrix A has two complex-conjugate eigenpairs, (λ, v) and (λ*, v*).
- An eigenvalue is λ = 2 + i, with the real part being α = 2 and the imaginary part being β = 1.
- The corresponding eigenvector v is

-1 + i
1

- The real part of v is g =

-1
1

- The imaginary part of v is h =

1
0

- Assume the initial condition x(0) is

1
1

- Inserting the columns x(0), g and h into x(0) = G g + H h, we obtain linear algebraic equations for G and H, giving
- G = 1, H = 2.
- The solution of the ODE is
- x(t) = G [g cos (βt) - h sin (βt)] exp (αt) + H [g sin (βt) + h cos (βt)] exp (αt)
- Inserting numbers, we obtain that x(t) =

(cos t - 3 sin t) exp(2t)
(cos t + 2 sin t) exp(2t)

- Because of the exponential growth, the origin is a repeller.
- See examples of **phase portrait**.

Steps to solve a system of two difference equations with complex-conjugate eigenvalues

- **Set up a system of two difference equations in two variables**
- Model a real-world problem with $x_{k+1} = Ax_k$, k = 0,1,2,...
- Prescribe the 2 x 2 real matrix A.
- Prescribe initial condition as a 2-entry column x_0.

- **Eigenvalue problem**
- The solution to the difference equations takes the form
- $x_k = \lambda^k v$,
- where λ is a number and v is a nonzero vector.
- Inserting $x_k = \lambda^k v$ into $x_{k+1} = Ax_k$, we obtain that
- Av = λv.
- The eigenvalue problem comes out naturally.
- A has complex-conjugate eigenpairs, (λ, v) and (λ*, v*).

- **Linear combination**
- The two eigenpairs correspond to two linearly independent solutions for x^k:
- $\lambda^k v$ and $(\lambda^*)^k v^*$.
- Any linear combination of the two solution gives another solution.
- Because x^k a real vector, write
- $x_k = G\ \mathrm{Re}(\lambda^k v) + H\ \mathrm{Im}(\lambda^k v)$,
- Re $(\lambda^k v)$ and Im $(\lambda^k v)$ are real vectors.
- G and H are real numbers determined by the initial condition.

- **Making the solution real**

- Write $v = g + i h$. Thus, g and h are real vectors.
- Write $\lambda = \alpha + i \beta$, $\beta > 0$. Thus, α and β are real numbers.
- $r = (\alpha^2 + \beta^2)^{½}$, $\cos \phi = \beta/r$, $\sin \phi = \alpha/r$.
- $\lambda = r (\cos \phi + i \sin \phi)$.
- Recall Euler's formula: $\exp(i \phi) = \cos \phi + i \sin \phi$
- Thus, $\exp(i k\phi) = \cos k\phi + i \sin k\phi$.
- $\lambda^k v$
- $= r^k (\cos k\phi + i \sin k\phi) (g + i h)$
- $= r^k [g \cos k\phi - h \sin k\phi] + i r^k [g \sin k\phi + h \cos k\phi]$.
- The solution is
- $x_k = G r^k [g \cos k\phi - h \sin k\phi] + H r^k [g \sin k\phi + h \cos k\phi]$.
- When k = 0, the above equation reduces to
- $x_0 = G g + H h$.
- This equation determines G and H for given columns x_0, g and h.

Example

- This example appears on Lay p. 296, with an interesting phase portrait.
- Here we explain why we should have this phase portrait.
- Matrix A is

0.5	- 0.6
0.75	1.1

- The matrix A has two complex-conjugate eigenpairs, (λ, v) and (λ*, v*).
- λ = 0.8 + 0.6 i, and v is

-2 +4 i
5

- Note that $0.8^2 + 0.6^2 = 1.0$. That is, r = 1.
- $\lambda = \cos \phi + i \sin \phi$,
- where $\cos \phi = 0.8$ and $\sin \phi = 0.6$.
- g is

-2
5

- h is

4
0

- The initial condition x_0 is

2

0

- x_0 = G g + H h.
- G = 0, H = 0.5.
- The solution $x_k = H r^k [g \sin k\phi + h \cos k\phi]$ =

- sin kϕ + 2 cos kϕ
2.5 sin kϕ

- As k increases, the state vector moves along an ellipse shown in Fig, 1, Lay p. 297.

- **Three types of behavior**
- Represent the states of the dynamical system by points in a plane.
- When r = 1, the state vectors move on an ellipse in the plane (Lay p. 297, Fig 1).
- When r < 1, the state vectors spiral inward to the origin. The origin is an **attractor** (Lay p. 308, Fig. 5).
- When r > 1, the state vectors spiral outward away the origin. The origin is a **repeller**.

Oscillation of a mass-spring system

- We now translate a real-world problem to a differential equation.

- **Empirical observations**
- Set up a **mass** and a **spring** on a **frictionless table**.
- When the spring is undeformed and the mass has zero velocity, the system is in a **state of equilibrium**.
- When a hand pulls the mass, the spring exerts a force on the mass in the opposite direction.
- After the hand is released, the mass **oscillates**.
- **Amplitude** of the oscillation is the maximum displacement of the mass.
- **Frequency** of the oscillation is the inverse of the time per cycle.

- **Physics**
- Let s be the **displacement** of the mass relative to the state of equilibrium. Adopt the sign convention that s is positive when the spring is pulled, and s is negative when the spring is pressed.
- A **free-body diagram** exposes forces acting on the mass. The spring exerts a **force** f on the mass.
- **Hooke's law of elasticity** says that the force f exerted by the spring on the mass is proportional to the displacement:
- f = - ks.
- Here k is a constant specific to the spring, and is called the **stiffness** of the spring.
- The negative sign means that the force acts on the mass in the direction opposite to that of the displacement.

- Let u be the **velocity** of the mass relative to the table.
- Both the displacement and the velocity are functions of time, s(t) and u(t).
- Velocity is defined as the change in displacement per unit time:
- u = ds/dt.
- **Acceleration** is defined as the change in velocity per unit time:
- a = du/dt.
- **Newton's law of inertia** says that the force equals the mass times acceleration:
- f = ma,
- where m is the **mass**.

- **Mathematics**
- You might have seen this example in other courses. There, the textbooks may have told you a trick to find a solution, and the class quickly moves on to study the physics of oscillation, leaving linear algebra unmentioned.
- Here we use this example to test the general method in linear algebra.
- Collecting the equations, we write a system two differential equations in two functions x(t) and u(t):
- ds/dt = u
- du/dt = - ks/m
- **The displacement s and the velocity u describe the state of the mass-spring system**.
- The initial conditions are given by the initial displacement $s(0) = s_0$ and the initial velocity $u(0) = u_0$.
- Let x be the column of s and u:

s
u

- Let A be the matrix

0	1
- k/m	0

- Given an initial condition x(0) (i.e., the initial displacement and the initial velocity), the mass-spring system evolves according to the differential equations dx/dt = Ax.
- Let us first find the eigenvalues of the matrix A.
- det (A - λI) = λ^2 + (k/m).
- The characteristic equation λ^2 + (k/m) = 0 has two imaginary roots:
- $\lambda_1 = i (k/m)^{1/2}$ and $\lambda_2 = - i (k/m)^{1/2}$.
- You will continue this problem in homework.
- You will discover that the eigenvalue is related to the **frequency** of the oscillation.

High-order ordinary differential equations of constant coefficients

- This type of equations appear in many applications.
- The method of their solutions is simple, and is directly related to what we have just learned.
- Take a look at this example, and you will be ready to do the homework problem.

- **Statement of a problem**
- Find a function y(t) that satisfies the a third-order ODE,
- $d^3y/dt^3 - 2d^2y/dt^2 - dy/dt + 2y = 0$,
- and the initial conditions
- y(0) = 2,
- (dy/dt) (0) = 2,
- (d^2y/dt^2) (0) = 2.
- The ODE is linear, each term of the ODE has a **constant coefficient**.

- **Solution**
- Here is the key idea: the solution takes the form y(t) = exp(λt).
- The number λ is also called an eigenvalue of the ODE.
- Note here we do not need to have a vector in front of the exponential function.
- Inserting y(t) = exp(λt) into the ODE, and we obtain that
- $\lambda^3 - 2\lambda^2 - \lambda + 2 = 0$.
- The ODE reduces to a polynomial equation.
- This equation is also called the characteristic equation for the ODE.
- Factor the polynomial, and we obtain that
- (λ + 1)(λ - 1)(λ - 2) = 0.
- The three roots are λ = - 1, 1, 2.
- The solution to the ODE is a linear combination of the three solutions:
- y(t) = a exp(-t) + b exp(t) + c exp (2t)
- The three numbers a, b and c are determined by the three initial conditions.
- y(0) = 2 becomes a + b + c = 2
- (dy/dt) (0) = 2 becomes - a + b + 2c = 2
- (d^2y/dt^2) (0) = 2 becomes a + b + 4 c = 2
- These equations are a system of three linear algebraic equation in three variables a, b, c.
- The solution to the system of algebraic solution is a = 0, b = 2, c = 0.
- The solution to the ODE is y(t) = 2 exp (t).

An n-th order ODE can be written as n first-order ODEs

- This fact establishes the equivalence between a high-order ODE and a system of first-order ODEs.

- (This part is optional material.)
- Still use the above ODE as an example.
- Write $y = x_1$,
- $dx_1/dt = x_2$
- $dx_2/dt = x_3$.
- Write $d^3y/dt^2 - 2d^2y/dt^2 - dy/dt + 2y = 0$ as
- $dx_3/dt = - 2x_1 + x_2 + 2x_3$.
- Write x_1, x_2, and x_3 as a column x.
- The original 3rd ODE $d^3y/dt^2 - 2d^2y/dt^2 - dy/dt + 2y = 0$ is equivalent to a system of three first order ODEs:
- $dx/dt = Ax$,
- with the matrix A given by

0	1	0
0	0	1
-2	1	2

- The characteristic polynomial of A is $\det(A - \lambda I) = (2 - \lambda)(\lambda + 1)(\lambda - 1)$. Not surprisingly, we obtain the same eigenvalues as we did before.
- Here is a video on solving such a problem using this approach, but you don't need to follow this approach.

Real operator having distinct eigenvalues, real or complex

- We now list the general results for a real operator having distinct eigenvalues, real or complex.

- **Fundamental theorem of algebra**. Any algebraic equation,
- $a_0 + a_1z + a_2z^2 + \ldots + a_{n-1}z^{n-1} + z^n = 0$,
- where $a_0, a_1, a_2, \ldots, a_{n-1}$ are complex numbers, has at least one solution z in the field of complex numbers C.
- Proof of this theorem is beyond the scope of linear algebra.

- **Theorem**. If λ and v are corresponding eigenvalue and eigenvector of a real matrix A, then their complex conjugate λ* and v* are also corresponding eigenvalue and eigenvector of A.
- **Proof**. Because λ and v are corresponding eigenvalue and eigenvector, $Av = \lambda v$. Taking complex conjugate on both sides of the equation, and recalling that A is real, we obtain that $Av^* = \lambda^*v^*$. That is, if (λ, v) is an eigenpair, so is (λ*, v*).
- Consequently, **the eigenvalues of a real matrix A are either real numbers or form complex conjugates**.

- **Eigenvalues and eigenvectors of a real matrix**

- Let T be an operator on an n-dimensional real vector space V.
- Assume that T has n distinct eigenpairs.
- p real eigenvalues $\lambda_1,\ldots, \lambda_p$ and corresponding real eigenvectors $v_1,\ldots, v_p$.
- q eigenvalues $\alpha_1 + i\,\beta_1, \ldots, \alpha_q + i\,\beta_q$ and corresponding complex eigenvectors are $g_1 + i\, h_1,\ldots,g_q + i\, h_q$.
- q eigenvalues $\alpha_1 - i\,\beta_1, \ldots, \alpha_q - i\,\beta_q$ and corresponding complex eigenvectors are $g_1 - i\, h_1,\ldots,g_q - i\, h_q$.
- As a convention, we choose positive imaginary part for each complex eigenvalue $\beta_1 > 0, \ldots, \beta_q > 0$.
- Warning: In Section 5.5, Lay adopts the sign convention to make α - i β and g + i h corresponding eigenvalue and eigenvector. But in Section 5.7 he switches to the common sign convention to make α + i β and g + i h corresponding eigenvalue and eigenvector. I will use the common sign convention that α + i β and g + i h are corresponding eigenvalue and eigenvector everywhere.

- **Eigenbasis**
- Because the eigenvalues are assumed to be distinct, n = p + 2q.
- Further, the real eigenvectors $v_1,\ldots, v_p$ and the real and imaginary parts of the complex eigenvectors $g_1,h_1\ldots, g_q,h_q$ are n linearly independent real vectors in V.
- These n real vectors are a basis for V:
- $v_1,\ldots, v_p, g_1,h_1\ldots, g_q,h_q$.
- Although the real and imaginary parts of a complex eigenvector are not eigenvectors, we will still call this basis the eigenbasis for brevity.
- For each real eigenpair v and λ,
- T(v) = λv,
- The vector T(v) has a single component λ relative to the eigenbasis.
- For each complex eigenpair v and λ,
- Write v = g + i h. g and h are **real** vectors in V
- Write λ = α + i β, β > 0. α and β are **real** numbers.
- T(v) = λv becomes that
- T(g + i h) = (α + i β) (g + i h).
- Separating the real and imaginary parts, we obtain that
- T(g) + i T(h) = (αg - βh) + i (βg + αh)
- Matching the real and imaginary parts separately, we obtain that
- T(g) = αg - βh
- T(h) = βg + αh
- g and h are real vectors in V, so are T(g) and T(h).
- The vector T(g) has two components α and - β relative to the eigenbasis, and the two components are in the directions g and h.
- The vector T(h) has two components β and α relative to the eigenbasis.
- Consequently, the matrix of the operator T relative to the eigenbasis is nearly a diagonal matrix, denoted matrix E:

λ_1								
	...							
		λ_p						
			α_1	β_1				
			$-\beta_1$	α_1				
					...			
						...		
							α_q	β_q
							$-\beta_q$	α_q

- **Arbitrary basis**
- Let $a_1,\ldots,a_n$ be an arbitrary basis for V.
- Let A be the matrix of the operator T relative to the basis $a_1,\ldots,a_n$.

- **From arbitrary basis to eigenbasis**
- When we solve the eigenvalue problem $Ax = \lambda x$, we obtain $v_1,\ldots, v_p, g_1,h_1\ldots,g_q,h_q$ as the columns of vectors relative to the basis $a_1,\ldots,a_n$.
- Consequently, the matrix
- $P = [v_1 \ \ldots \ v_p \ \ g_1 \ \ h_1 \ldots . \ g_q \ \ h_q]$
- is the change-of-basis matrix from the basis $a_1,\ldots,a_n$ to the eigenbasis.
- A is the matrix of the operator T relative to the basis $a_1,\ldots,a_n$.
- E is the matrix of the operator T relative to the eigenbasis.
- Recall the change of the matrix of an operator T under a change of basis:
- $A = PEP^{-1}$.
- If an operator T on an n-dimensional real vector space V has n distinct eigenvalues, real or complex, the matrix A of T relative to an arbitrary basis is similar to the nearly diagonal matrix E.

Differential equations having real and complex eigenvalues

- **Coupled differential equations**
- $dx/dt = Ax$, with A and $x(0)$ prescribed.
- x is the column of a state vector v relative to an arbitrary basis $a_1,\ldots,a_n$.
- A is the matrix of an operator T relative the arbitrary basis $a_1,\ldots,a_n$.
- Typically, A is a matrix with many nonzero entries, which couples various components of x.

- For examples, two coupled ODEs model the **co-evolution** of the populations of owls and rats.

- **Eigenbasis decouples differential equations**
- Solve the eigenvalue problem of the matrix $Ax = \lambda x$.
- Construct the eigenbasis of the rate-of-change operator T.
- Let P be the change-of-basis matrix from the basis $a_1,\ldots,a_n$ to the eigenbasis.
- $x = Py$
- $x(0) = Py(0)$
- $A = PEP^{-1}$
- $dy/dt = Ey$
- y is the column of the state vector v relative to the eigenbasis.
- E is the matrix of the rate-of-change operator T relative to the eigenbasis.
- E is nearly a diagonal matrix.
- Thus, the ODE in y(t) is **nearly decoupled**!
- We have studied two cases:
- A single real eigenvalue.
- A single pair of complex-conjugate eigenvalues.
- The general solution of the ODE is a **linear combination**:
- $x(t) = c_1v_1 \exp(\lambda_1 t) + \ldots + c_pv_p \exp(\lambda_p t) +$
- $+ [(G_1g_1 + H_1h_1) \cos(\beta_1 t) + (- G_1h_1 + H_1g_1) \sin(\beta_1 t)] \exp(\alpha_1 t)$
- $+\ldots$
- $+ [(G_qg_q + H_qh_q) \cos(\beta_q t) + (- G_qh_q + H_qg_q) \sin(\beta_q t)] \exp(\alpha_q t)$.
- The numbers $c_1,\ldots,c_p$, G_1, $H_1,\ldots,G_q$, H_q are determined by the initial condition:
- $x(0) = c_1v_1 + \ldots + c_pv_p + G_1g_1 + H_1h_1 + \ldots + G_qg_q + H_qh_q$.
- Recall that for each pair of complex-conjugate eigenvalues, we write
- $v = g + i\,h$.
- $\lambda = \alpha + i\,\beta$, $\beta > 0$.

Discrete dynamical system having real and complex eigenvalues

- **Coupled discrete dynamical system**
- $x_{k+1} = Ax_k$, $k = 0,1,2,\ldots$, with A and x_0 prescribed.
- x is the column of a vector v relative to an arbitrary basis $a_1,\ldots,a_n$.
- A is the matrix of an operator T relative the arbitrary basis $a_1,\ldots,a_n$.
- If A is a matrix with many nonzero entries, various components of x are coupled.

- **Eigenbasis decouples a discrete dynamical system**
- Solve the eigenvalue problem of the matrix A.
- Construct the eigenbasis of the operator T.
- Let P be the change-of-basis matrix from the basis $a_1,\ldots,a_n$ to the eigenbasis.
- $x = Py$
- $x_0 = Py_0$.

- $A = PEP^{-1}$
- $y_{k+1} = Ey_k$, $k = 0,1,2,...$
- The matrix E has a lot of zero entries.
- We have studied two cases:
- A single real eigenvalue.
- A single pair of complex-conjugate eigenvalues.
- The general solution is a linear combination:
- $x_k = c_1 v_1 \lambda_1^k + \ldots + c_p v_p \lambda_1^p +$
- $+ [(G_1 g_1 + H_1 h_1) \cos k\phi_1 + (- G_1 h_1 + H_1 g_1) \sin k\phi_1] r_1^k$
- $+ \ldots$
- $+ [(G_q g_q + H_q h_q) \cos k\phi_q + (- G_q h_q + H_q g_q) \sin k\phi_q] r_q^k$.
- The numbers $c_1,...,c_p$, G_1, $H_1,...,G_q$, H_q are determined by the initial condition:
- $x(0) = c_1 v_1 + \ldots + c_p v_p + G_1 g_1 + H_1 h_1 + \ldots + G_q g_q + H_q h_q$.
- Recall that for each pair of complex-conjugate eigenvalues, we write
- $v = g + i h$.
- $\lambda = \alpha + i \beta$, $\beta > 0$.
- $r = (\alpha^2 + \beta^2)^{1/2}$, $\cos \phi = \beta/r$, $\sin \phi = \alpha/r$.

Chicken-rabbit space has no length and angle

- The geometry of two and three dimensions has two conspicuous concepts: the length of an arrow, and the angle between two arrows.
- Length and angle, however, are absent in general vector spaces.
- Indeed, we regard them as pesky distractions. We have trained ourselves to avoid them.
- For example, in the chicken-rabbit space, we compare the magnitudes of scalars in the **same** scalar set. The scalar of 14 chickens is twice the scalar of 7 chickens.
- But we do not compare “lengths” of different chicken-rabbit vectors.
- Which vector is longer, (1 chicken, 0 rabbit) or (0 chicken, 1 rabbit)?
- This question has no commonly useful meaning in real life.
- We choose to leave this question meaningless in the chicken-rabbit space.
- We do not define length in the chicken-rabbit space.
- Also meaningless is the “angle” between two chicken-rabbit vectors.
- Indeed, most applications of vector spaces do not need the concepts of length and angle.
- For instance, the Leontief input-output model requires no concepts of length and angle in the product space.
- General vector spaces have no length and angle.

- **The importance of length and angle**
- This said, we do appreciate the importance of length and angle in plane and solid geometry, and in our life.

- In statistics, now called data science, we often speak of the **distance** between two data points.
- Linear algebra accommodates these needs by generalizing the concepts of length and angle, and adding a new thing to a vector space: inner product.
- A general vector space without inner product is one algebraic structure (Chapters 1-5).
- A vector space with inner product is another algebraic structure (Chapters 6 and 7).

- **Anam Javed's questions**. What is an algebraic structure? Can you give me examples?
- Answer: An algebraic structure is a set whose elements follow some rules of operation.
- We have studied several algebraic structures: number field, scalar set, vector space.
- We now study another algebraic structure: **inner-product space**.

Lay 6.1 Inner product, length, and orthogonality

- **Vectors in R^m**.
- In geometry of two and three dimensions, the length of a vector and the angle between two vectors play important roles.
- We wish to generalize these ideas to R^m.
- The standard basis for R^m consists of m vectors:
- e_1 = (1,0,..,0),
- e_2 = (0,1,..,0),
- …..
- e_m = (0,0,..,1).
- Unless otherwise specified, we list the components of a vector u in R^m relative to the standard basis by a column:

u_1
...
u_m

- Similarly, we list the components of another vector v in R^m relative to the standard basis by a column:

v_1
...
v_m

Standard inner product

- The **standard inner product** on R^m is defined as
- $u \cdot v = u_1 v_1 + \dots + u_m v_m$.

- The inner product is also called a **dot product**.
- A vector space with an inner product is called an **inner-product space.**

- A picture is worth a thousand words.
- An equation is worth a thousand pictures.
- We have just given such an equation, the definition of the inner product.

- **Theorem.** Let u, v, w be vectors in R^m, and let c be a number in R. From the definition of the inner product, we can confirm the following properties of the inner product.
- $u \cdot v = v \cdot u$.
- $(u + v) \cdot w = u \cdot w + v \cdot w$.
- $(cu) \cdot v = c(u \cdot v)$.
- $u \cdot u > o$ if u is a nonzero vector.

- The inner product is a **symmetric, bilinear, positive-definite map**.
- The map sends two vectors in R^m to a number in R.
- (inner product): $R^m \times R^m \rightarrow R$.

- We also write the standard inner product in the matrix notation:
- $u^T v = u_1 v_1 + \ldots + u_m v_m$.

Length

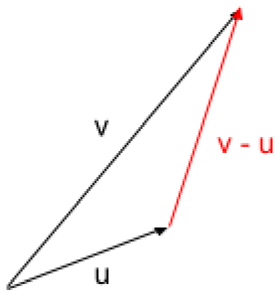
- Define the **length** of a vector v in R^m by
- $||v|| = (v \cdot v)^{½} = (v_1^2 + \ldots + v_m^2)^{1/2}$.

- A **unit vector** is a vector whose length is 1.

- For a nonzero vector v in R^n, its length ||v|| is a nonzero real number, and v/||v|| is a unit vector in the same direction as v.
- This method to obtain a unit vector is called to **normalize a vector**.

- **Example**
- Let v = (3, 4) be a vector in R^2. Determine the unit vector u in the direction v.
- The length of the vector v is
- $||v|| = (3^2 + 4^2)^{½} = 5$.
- The unit vector in the direction of v is
- v/||v|| = (3,4)/5 = (3/5, 4/5).
- Represent R^2 by points in a plane.
- Draw the vector v and the unit vector v/||v||.

- The **distance** between two vectors u and v is the length of the vector v - u:
- $||v - u||$.



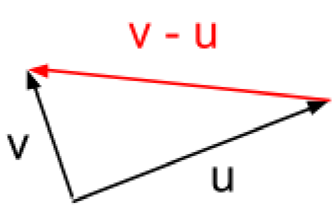
- **Example**
- Let u = (1, 2) and v = (3, 4) be two vectors in R^2.
- Plot the two vectors in a plane.
- The distance between the two vectors u and v is
- $||v - u|| = [(3 - 1)^2 + (4 - 2)^2]^{½} = 2\sqrt{2}$.

Orthogonality

- Two vectors u and v are called **orthogonal** if
- $u \cdot v = 0$.

- **Example**
- The standard basis vectors are unit vectors:
- $e_1 \cdot e_1 = e_2 \cdot e_2 = \ldots e_n \cdot e_n = 1$.
- The standard basis vectors are orthogonal to one another:
- $e_i \cdot e_j = 0$ when $i \neq j$.

- **Pythagorean theorem**
- Two vectors u and v are orthogonal if and only if
- $||u - v||^2 = ||u||^2 + ||v||^2$.



- **Proof**. $||u - v||^2 = (u - v) \cdot (u - v) = u \cdot u - 2\, u \cdot v + v \cdot v = ||u||^2 - 2\, u \cdot v + ||v||^2$. End of the proof.
- Geometric interpretation: the vectors u, v and v - u are three sides of a **right triangle**.

- **Example**. Consider two vectors:
- u = (1, 1, 1, 1),
- v = (1, 1, - 1, - 1).

- The two vectors are in R^4.
- We cannot draw the two vectors to see if they are orthogonal, but we can calculate their inner product:
- $u^Tv = 1 \times 1 + 1 \times 1 - 1 \times 1 - 1 \times 1 = 0$.
- Thus, the two vectors are orthogonal.
- We can confirm that they satisfy the Pythagorean theorem:
- $||u||^2 = ||v||^2 = 4$
- u - v = (0, 0, 2, 2), so that that $||u - v||^2 = 8$.
- Thus, $||u - v||^2 = ||u||^2 + ||v||^2$.

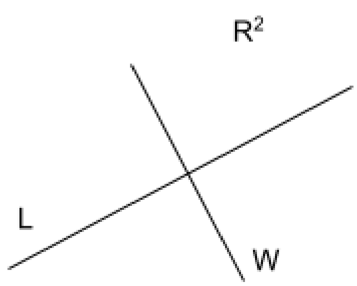
Orthogonal complement

- Let W be a subspace of R^m.
- A vector z in R^m is called orthogonal to the subspace W if the vector z is orthogonal to **every** vector in W.
- All vectors in R^m orthogonal to W form a subspace, denoted Z, and called the **orthogonal complement** of W.

- If Z is the orthogonal complement of W, then W is the orthogonal complement of Z.

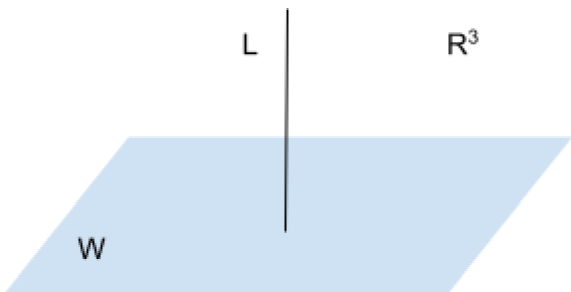
- To summarise, W and Z are called orthogonal complements in R^m if
- (1) W and Z are subspaces in R^m,
- (2) every vector in Z is orthogonal to every vector in W, and
- (3) dim W + dim Z = m.

- **Example**
- In R^2, a line through the origin is a subspace, denoted W.
- The orthogonal complement of W is the line L orthogonal to the line W and through the origin.
- Every vector in L is orthogonal to every vector in W.
- dim W = 1, dim L = 1.



- **Example**
- In R^3, a plane through the origin is a subspace, denoted W.

- The orthogonal complement of W is the line L orthogonal to the plane W and through the origin.
- Every vector in the plane W is orthogonal to every vector in the line L.
- dim W = 2, dim L = 1.



- **Example.** Consider vectors
- a = (1, 0, 1, 0, 1),
- b = (0, 1, -2, 0, 3),
- c = (0, 0, 0, 1, -5),
- x = (-1, 2, 1, 0, 0),
- y = (-1, -3, 0, 5, 1).
- Let W = Span {a, b, c}
- Let Z = Span {x, y}.
- The five vectors are in R^5.
- We cannot draw pictures to see the vectors and subspaces, but we can do algebra.
- Verify that a, b, c are linearly independent, so that dim W = 3.
- Verify that x and y are linearly independent, so that dim Z = 2.
- Verify that each of a, b, c are orthogonal to each of x and y, so that W and Z are orthogonal complements in R^5.

- **Example**
- The previous examples appeal to our intuition in plane and solid geometry. Here is an example makes us use the abstract definition in high dimensions.
- A subspace W in R^4 is spanned by two vectors (1,2,2,3) and (1,3,3,2). Find a basis for the orthogonal complement of W.
- Let x be a vector in R^4 in the orthogonal complement of W.
- By definition, x must be orthogonal to every vector in the spanning set of W.
- That is, x must be orthogonal to the two vectors (1,2,2,3) and (1,3,3,2).
- The above sentence means that x is a solution to the equation Ax = 0, where the matrix A lists the two vectors (1,2,2,3) and (1,3,3,2) as rows.
- The solution set of the equation Ax = 0 gives a basis for the orthogonal complement of W.
- We summarize the process as two septs
- 1. List the vectors that span the subspace W as rows in a matrix A.

- 2. Find the solution set of Ax = 0 to give a basis for the orthogonal complement of W.
- Thus, W is the row space of A, Row A
- The orthogonal complement of W is the null space of A, Nul A.
- Please find the basis for the orthogonal complement of W on your own.
- Compare your solution to that described in this video.

Week 12 (April 17, 19, 21)

Reading and watching

- Lay Chapter 6
- Video by Gilbert Strang on The Big Picture of Linear Algebra.
- Video by Andrew Ng on regression as part of his course on machine learning.
- Wiki linear regression

Four subspaces created by a linear map

- Let A be an m x n matrix.
- The matrix A is a linear map,
- A: $R^n \to R^m$.
- The linear map sends every vector x in R^n to a vector y in R^m:
- y = Ax.
- We now adopt the standard inner product for R^n, and the standard inner product for R^m.

- **Theorem**
- Row A and Nul A are orthogonal complements in R^n.

- **Proof**
- Row A is the subspace in R^n spanned by the rows of A.
- Nul A is the subspace in R^n spanned by all the solution vectors to the homogeneous equation Ax = 0.
- The equation Ax = 0 lists m scalar equations.
- Each scalar equation is the inner product of a row vector in A and the solution vector x.
- The equation Ax = 0 means that each row in A is orthogonal to each solution vector x.
- Thus, Row A and Nul A are orthogonal complements in R^n.

- **Theorem**
- Col A and Nul A^T are orthogonal complements in R^m.

- This theorem is analogous to the previous theorem.
- Just transpose the matrix.
- Similarly, Nul A^T is the solution set of $A^Ty = 0$.

- This equation means that each row of A^T is orthogonal to every solution vector y.
- Also recall that Row A^T = Col A.

- **Rank of a matrix**
- Recall that the rank of A, r, is the number of pivots in A.
- dim Col A = dim Row A = r.
- dim Nul A = n - r.
- dim Nul A^T = m - r.

- **Example**
- Watch a video by Gilbert Strang on The Big Picture of Linear Algebra.
- Consider a linear map A: $R^3 \rightarrow R^2$, where the matrix A is

1	2	3
4	5	6

- n = 3, m = 2.
- The two rows are independent, and span Row A.
- The three columns are dependent.
- For example, the combination of 1 times the first column , -2 times the second column, and 1 times the third column gives the zero vector.
- That is, a solution to the equation Ax = 0 is the vector x

1
-2
1

- This vector spans Nul A.
- By construction, x is orthogonal to the two row vectors of A.
- dim Row A = 2, dim Nul A = 1.
- Any two columns in A span Col A.
- dim Col A = 2, dim Nul A^T = 0.

Lay 6.2 Orthogonal sets

- A set of vectors $u_1,...,u_p$ in R^m are called an **orthogonal set** if each pair of the vectors in the set are orthogonal:
- $u_i^T u_j = 0$ if $i \neq j$.

- **Orthogonal basis**
- Any orthogonal set of nonzero vectors $u_1,...,u_p$ in R^m are linearly independent.
- An orthogonal set $u_1,...,u_p$ is a basis for a p-dimensional subspace W in R^n, and is called an **orthogonal basis** for W.

- What is special about an orthogonal basis?

- **Components of a vector relative to an orthogonal basis**
- Any vector w in W is a linear combination of the basis vectors for W:
- $w = c_1u_1 + ... + c_pu_p$.
- The numbers $c_1,...,c_p$ are the components of w relative to the basis $u_1,...,u_p$.
- Calculate $c_1,...,c_p$ as follows. Important and interesting!
- $u_1^Tw = u_1^T(c_1u_1 + ... + c_pu_p)\cdot u_j = c_1u_1^Tu_1$.
- Because the basis vectors are an orthogonal set, only one inner product survives.
- This calculation gives that
- $c_1 = (u_1^Tw)/(u_1^Tu_1)$.
- c_1 is the component of the vector w **projected onto** the basis vector u_1
- Similarly,
- $c_2 = (u_1^Tw)/(u_2^Tu_2)$,
- ……
- $c_p = (u_p^Tw)/(u_p^Tu_p)$.
- In the past, finding components of a vector relative to a basis requires us to solve a system of equations.
- With an orthogonal basis, we no longer solve equations. We just calculate inner products.

- Write
- $w = u_1(u_1^Tw)/(u_1^Tu_1) + … + u_p(u_p^Tw)/(u_p^Tu_p)$.
- This is a complicated formula generated from a simple idea.

- **Example**
- Consider three columns u_1, u_2, and u_3 listed below.

3	- 1	-1
1	2	- 4
1	1	7

- Verify that the three columns are orthogonal to one another:
- $u_1^Tu_2 = u_2^Tu_3 = u_3^Tu_1 = 0$.
- Consequently, the three vectors u_1, u_2 and u_3 are an orthogonal basis for R^3.

- We next calculate the components of the vector y = (1, 1, 1) relative to the basis u_1, u_2 and u_3.
- Write $y = c_1u_1 + c_2u_2 + c_3u_3$.
- $c_1 = (u_1^Ty)/(u_1^Tu_1) = 5/11$
- $c_2 = (u_2^Ty)/(u_2^Tu_2) = 2/6$
- $c_3 = (u_3^Ty)/(u_3^Tu_3) = 2/66$

Orthonormal set

- Vectors $u_1,...,u_n$ in R^m are called an **orthonormal set** if
- $u_i^T u_j = 0$ when $i \neq j$, and
- $u_i^T u_i = 1$.

- **A matrix of orthonormal columns**
- Put the n vectors as columns of a matrix:
- $U = [u_1 \ldots u_n]$.
- U is an m x n matrix.
- The matrix U has orthonormal columns if and only if $U^T U = I$.
- Let U be an m x n matrix U with orthonormal columns, and x and y be vectors in R^n. Then
- $(Ux)^T(Uy) = x^T y$.
- Thus, the linear map $U: R^n \rightarrow R^m$ preserve value of the inner product.
- Here are two consequences:
- The linear map U preserves the length of a vector:
- $||Ux|| = ||x||$.
- The linear map U preserves the orthogonality of two vectors:
- $(Ux)^T(Uy) = 0$ if and only if $x^T y = 0$.

- **Orthogonal matrix**
- A matrix U is called an **orthogonal matrix** if
- (1) U is a **square** matrix, and
- (2) the columns of U are an orthonormal set.

- Verify that $U^T U = I$
- Because any orthonormal set are linearly independent, and because U is a square matrix, U is invertible.
- Right-multiply U^{-1} to $U^T U = I$, and we obtain that
- $U^T = U^{-1}$.
- Left-multiply U, and we obtain that
- $UU^T = I$
- Consequently, the rows of an orthogonal matrix U also form an orthonormal set.

- **Example**. Consider the matrix:

3	- 1	-1
1	2	- 4
1	1	7

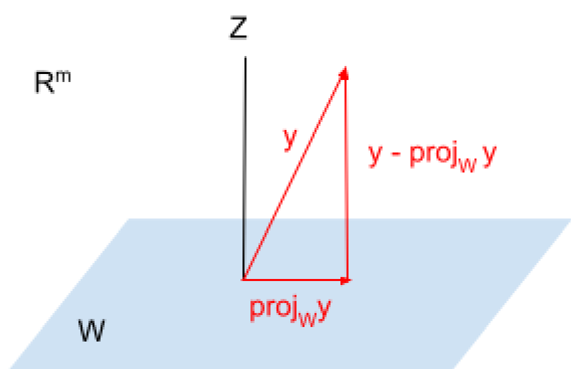
- The columns of the above matrix are an orthogonal set, but the rows of the above matrix are not an orthogonal set.
- We now normalize every column of the matrix, and obtain

3/√11	- 1/√6	-1/√66
1/√11	2/√6	- 4√66
1/√11	1/√6	7√66

- The columns of this matrix are an orthonormal set.
- The rows of this matrix are also an orthonormal set!

Lay 6.3 Orthogonal projection

- Let y be a vector in R^m.
- Let W be a subspace in R^m.
- The **orthogonal projection of the vector y onto the subspace W** is a vector in R^m, written $proj_W$ y, satisfying properties:
- (1) $proj_W$ y is in W, and
- (2) y - $proj_W$ y is orthogonal to every vector in W.



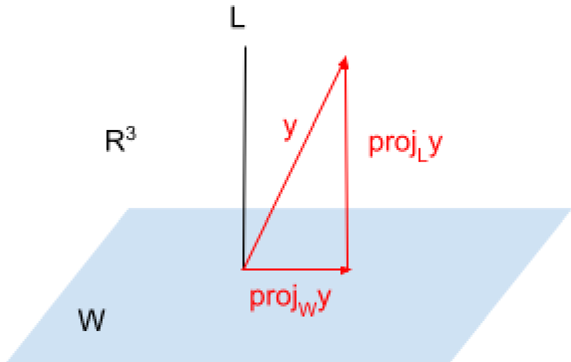
- Condition (2) is equivalent to that the vector y - $proj_W$ y is in the subspace Z, the orthogonal complement of W.
- $proj_W$ is a linear operator on R^m, sending any vector y in R^m to the orthogonal projection of y onto W.

Orthogonal decomposition theorem

- Let W and Z be orthogonal complements in R^m.
- Every vector y in R^m is the sum of the orthogonal projections onto W and Z:
- y = $proj_W$ y + $proj_Z$ y.

- **Example**

- In R^3, a plane W through the origin is a subspace, and a line L through the origin is also a subspace.
- When W and L are orthogonal, they are orthogonal complements.
- A vector y is the sum of two orthogonal vectors, denoted $\text{proj}_W y$ and $\text{proj}_L y$. The former lies in the plane W, and the latter in the line L. Write
- $y = \text{proj}_W y + \text{proj}_L y$.



Given a subspace W in R^n and a vector y in R^n, find proj_W y

- Let $v_1,\dots,v_p$ be an **orthogonal basis** for the subspace W in R^m.
- proj_W y is a vector in W, and is therefore a linear combination of the basis vectors for W:
- $\text{proj}_W\, y = v_1(v_1^T y)/(v_1^T v_1) + \dots + v_p\,(v_p^T y)/(v_p^T v_p)$.

- **Example**
- Consider two vectors:
- p = (1, 1, 1, 1),
- q = (1, 1, - 1, - 1).

- Verify that p and q are an **orthogonal set** in R^4:
- $p^T q = 0$.

- Find the orthogonal projection of a vector y = (1, 2, 3, 4) onto Span {p,q}.
- $\text{proj}_{\text{Span}\,\{p,q\}}\, y$
- $= p(p^T y)/(p^T p) + q(q^T y)/(q^T q)$
- = 2.5p - q
- = (1.5, 1.5, 3.5, 3.5).

- Let $u_1,\dots,u_p$ be an **orthonormal basis** for the subspace W in R^m.
- $\text{proj}_W\, y = u_1(u_1^T y) + \dots + u_p\,(u_p^T y)$.
- List the orthonormal basis vectors as columns in a matrix:
- $U = [u_1 \dots u_p]$.
- The above expression for proj_W y reduces to
- $\text{proj}_W\, y = UU^T y$.

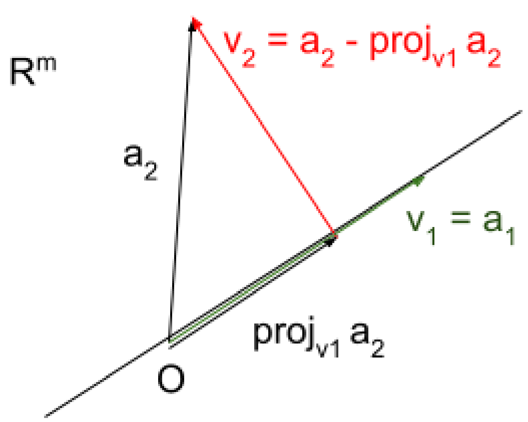
- Thus, UU^T is a **linear operator** on R^n, sending any vector y in R^n to the orthogonal projection of y onto W.

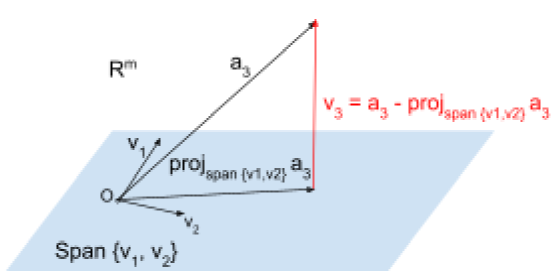
- Let Z be the orthogonal complement of W in R^m.
- $proj_Z\, y = y - UU^Ty = (I - UU^T)y$.
- Thus, $I - UU^T$ is a linear operator on R^n, sending any vector y in R^n to the orthogonal projection of y onto Z.

Lay 6.4 The Gram-Schmidt process

- Let W be a subspace in R^m.
- The Gram-Schmidt process changes an arbitrary basis $a_1,\ldots,a_n$ for W to an orthogonal basis $v_1,\ldots,v_n$ for W.
- $v_1 = a_1$,
- $v_2 = a_2 - v_1(v_1^Ta_2)/(v_1^Tv_1)$,
- $v_3 = a_3 - v_1(v_1^Ta_3)/(v_1^Tv_1) - v_2(v_2^Ta_3)/(v_2^Tv_2)$,
- ……
- $v_n = a_n - v_1(v_1^Ta_n)/(v_1^Tv_1) - v_2(v_2^Ta_n)/(v_2^Tv_2) - \ldots. - v_{n-1}(v^T_{n-1}a_n)/(v^T_{n-1}v_{n-1})$.

- The Gram-Schmidt process is **recursive**:
- Set $v_1 = a_1$.
- Set v_{j+1} = orthogonal projection of a_{j+1} onto the orthogonal complement of Span $\{v_1,\ldots,v_j\}$.





QR factorization

- **Theorem**. If A is an m x n matrix with linearly independent columns, then A can be factored as A = QR, where Q is an m x n matrix whose columns are an orthonormal basis for Col A, and R is an upper-triangular, invertible, n x n matrix, with positive entries on its diagonal.

- **Proof by construction**. Write A as n columns of linearly independent vectors:
- $A = [a_1\ a_2\ a_3 \dots a_n]$.
- The Gram-Schmidt process produces orthogonal vectors before normalization: $v_1,\dots,v_n$.
- Normalize each vector: $u_1 = v_1/(v_1^Tv_1)^{½},\dots, u_n = v_n/(v_n^Tv_n)^{½}$.
- Write Q as n columns of the orthonormal vectors:
- $Q = [u_1\ u_2\ u_3 \dots u_n]$.
- R is an n x n matrix of inner products calculated in the Gram-Schmidt process: R =

$(v_1^Tv_1)^{1/2}$	$u_1^Ta_2$	$u_1^Ta_3$	...	$u_1^Ta_n$
0	$(v_2^Tv_2)^{1/2}$	$u_2^Ta_3$	...	$u_2^Ta_n$
0	0	$(v_3^Tv_3)^{1/2}$	...	$u_3^Ta_n$
...	...	...	...	...
0	0	0	...	$(v_n^Tv_n)^{1/2}$

- **QR factorization by hand**
- Given an m x n matrix A, write the n columns of the matrix, $A = [a_1 \dots a_n]$.
- Use the Gram-Schmidt process converts the columns $a_1, \dots, a_n$ to an orthogonal set $v_1,\dots,v_n$.
- Normalize each vector: $u_1 = v_1/(v_1^Tv_1)^{½},\dots, u_n = v_n/(v_n^Tv_n)^{½}$.
- Write Q as n columns of the orthonormal vectors:
- $Q = [u_1\ u_2\ u_3 \dots u_n]$.
- Calculate $R = Q^TA$.

- **QR factorization by MATLAB**
- Command [Q,R]=qr(A).

- **Example**
- Given a matrix A

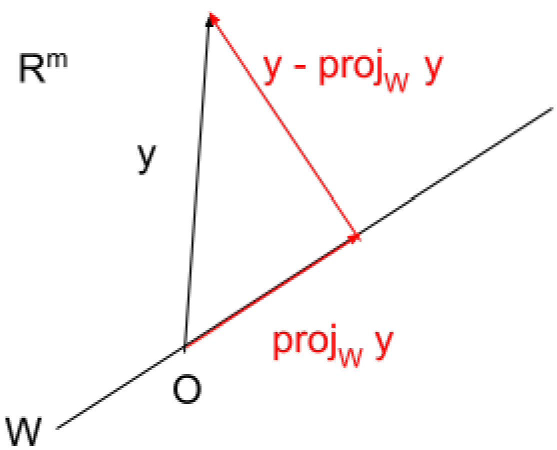
1	2	4
0	0	5

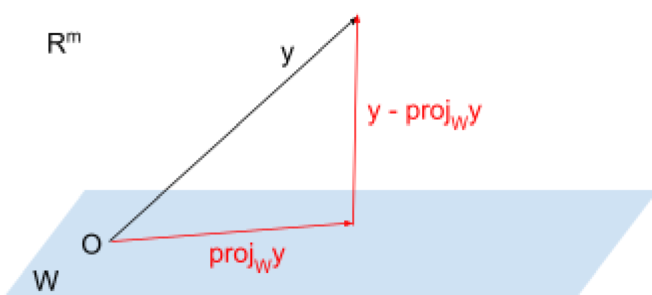
0	3	6

- From the three columns of A find three orthonormal vectors.
- Write A = QR, where Q = [u_1 … u_n] is an orthogonal matrix, and R an upper triangular matrix.
- The solution is described in this video.

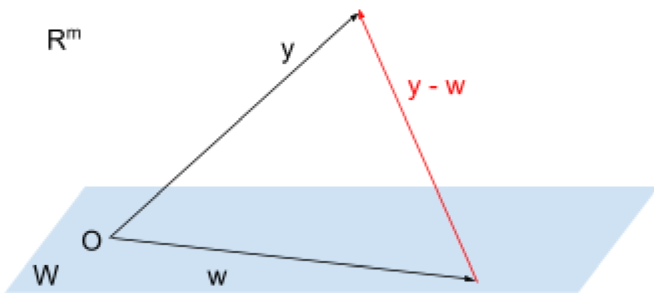
The best approximation theorem

- Given a vector y in R^m and a subspace W in R^m, we have learned to find the orthogonal projection of y onto W, $proj_W$ y.
- What can orthogonal projection do for us?



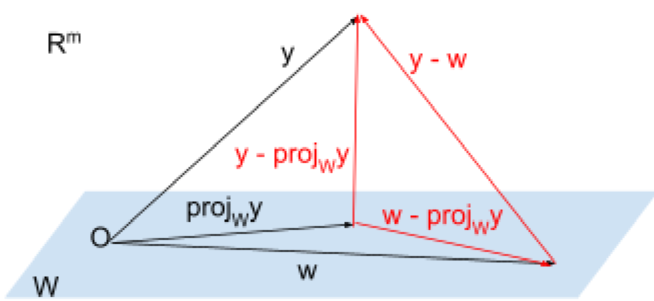


- **Problem**
- Let y be a vector in R^m, and W be a subspace in R^m.
- Find the element w in W that **best approximates** y.
- That is, find w in W that minimizes the distance ||y - w||.



- **Solution**
- The vector $\text{proj}_W\, y$ is the element in W that best approximates y.
- The vector $y - \text{proj}_W\, y$ is the **error vector (or residual vector) of the approximation**
- The number $||y - \text{proj}_W\, y||$ is called the **distance between the vector y and the subspace W**.

- **Theorem**
- Let y be a vector in R^m, and W be a subspace in R^m.
- Of all vectors in W, the vector $\text{proj}_W\, y$ best approximates y.
- That is, $||y - \text{proj}_W\, y|| \le ||y - w||$ for every vector w in W.



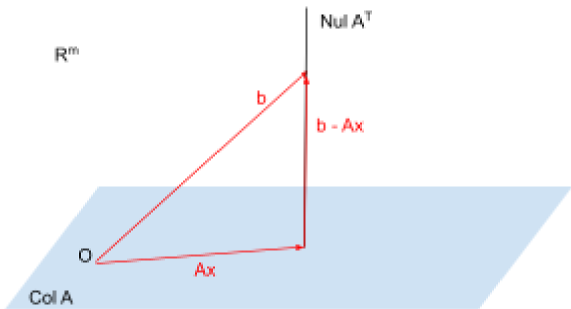
- **Proof**
- The three vectors marked in red form a right triangle, because the vector $y - \text{proj}_W\, y$ is orthogonal to every vector in W, and $w - \text{proj}_W\, y$ is a vector in W.
- Apply the Pythagorean theorem to this right triangle:
- $||y - w||^2 = ||y - \text{proj}_W\, y||^2 + ||w - \text{proj}_W\, y||^2$.
- Dropping the last term in the above equality, we obtain the desired inequality.

Lay 6.5 Least squares

- The theory of least squares is a place where geometry, algebra, and analysis (calculus) meet.

- **Algebra poses a problem**
- Ax = b is a system of **inconsistent** linear algebraic equation.

- That is, given a matrix A: $R^n \rightarrow R^m$ and given a column b in R^m, no column x in R^n satisfies the equation Ax = b.
- We lower our demand for solution:
- Find x that minimizes $(b - Ax)^T(b - Ax)$, a sum of squares of differences.
- Such a solution x is called a **least-squares solution** to the system of inconsistent equations.
- b - Ax is called the **residual vector**.



- **Geometry: seeing the least-squares solution**
- The least-squares problem is equivalent to a geometric problem:
- Given a linear map A: $R^n \rightarrow R^m$ and given a vector b in R^m, find a vector x in R^m to minimize the **length** of the vector b - Ax.
- We next try to "see" the geometry of this problem using words and equations, as well as a fake diagram.

- Write A as n columns in R^m: $A = [a_1 \ldots a_n]$.
- Write x as an n-tuple of numbers: $x = (x_1,\ldots,x_n)$.
- Given A, as x changes, Ax spans the column space, Col A.
- b is a vector in R^m, and Col A is a subspace in R^m.
- Ax = b means that b is a linear combination of the columns in R^m:
- $b = x_1a_1 + \ldots + x_na_n$.
- Ax = b is consistent if and only if b is in Col A.
- If Ax = b is inconsistent**,** b is not in Col A.
- We lower our demand for solution:
- Find x that minimizes the distance between the two vectors, b and Ax.
- This least-squares solution x makes b - Ax orthogonal to Col A.
- That is, the least-squares solution requires that the residual vector b - Ax be in the subspace Nul A^T.

- **Translate this geometry back to algebra**:
- $A^T(b - Ax) = 0$.
- Write this equation as
- $A^TAx = A^Tb$.

- This equation is called the **normal equation** for Ax = b.
- A solution x to $A^TAx = A^Tb$ is called a **least-squares solution**.

- **Calculus: minimize a function**
- We wish to find x to minimize ||Ax -b||.
- This is the same as to find x to minimize the function
- $f(x) = (Ax - b)^T(Ax - b)$.
- This function has n independent variables $x_1,...,x_n$.
- To minimize f(x), we set
- $\partial f/\partial x_1 = 0$,
- …....
- $\partial f/\partial x_n = 0$.
- This procedure, with some work, will also lead to the normal equation
- $A^TAx = A^Tb$.

- The method of least squares needs the following theorem.
- For an m x n matrix A, the matrix A^TA is invertible if and only if the columns of A are linearly independent.
- We will study the matrix A^TA in some depth when we study singular value decomposition.

Lay 6.6 Applications to linear regression

- **Sizes and prices of houses**.
- Watch a video by Andrew Ng as part of his course on machine learning.
- We have collected a set of data points.
- Each data point is an ordered pair (x, y).
- x = the size of a house, in the unit of square feet.
- y = the price of the house, in the unit of 1000 $.
- List the data points in a table

Size in square feet	Price in 1000 $
2104	460
1416	232
1534	315
852	178
...	...

- Now we wish to sell a house of a certain size, and need to set a price for the house.

- **Fit data to a straight line using the method of least squares**

- Video by Gilbert Strang on least squares.
- Wiki linear least squares
- We have a set of data points $(x_1, y_1), \ldots, (x_m, y_m)$.
- The data points do not lie on a straight line, but we wish to fit the data points to a straight line:
- $y = \beta_0 + \beta_1 x$
- That is, find β_0 and β_1 so that the line $y = \beta_0 + \beta_1 x$ best fits the data.
- $y_1 = \beta_0 + \beta_1 x_1$
- ……
- $y_m = \beta_0 + \beta_1 x_m$
- This is a system of m linear algebraic equations in two variables β_0 and β_1.
- Write the equations in the matrix notation:
- $X\beta = y$
- The matrix X is called the **design matrix**, given by

1	x_1
...	...
1	x_m

- The column β is called the **parameter vector in R^2**:

β_0
β_1

- The column y is called the **observation vector in R^m**:

y_1
...
y_m

- $X: R^2 \to R^m$
- (design matrix): (parameter space) → (observation space)
- The equation $X\beta = y$ is inconsistent.
- Find the least-squares solution β from the normal equation:
- $X^TX\beta = X^Ty$.
- Write $y = X\beta + \varepsilon$
- The column ε is called the **residual vector in R^m**:

ε_1
...
ε_m

- **Example**
- This example is taken from [Wikipedia]().
- I have changed it a little to conform to our notation.
- Given four data points

x_i	y_i
1	6
2	5
3	7
4	10

- The four data points do not lie on a straight line, but we wish to fit the data points to a straight line:
- $y = \beta_0 + \beta_1 x$
- That is, find β_0 and β_1 so that the line $y = \beta_0 + \beta_1 x$ best fits the data.

- Insert the four data points into the equation $y = \beta_0 + \beta_1 x$, and we obtain that
- $\beta_0 + 1\beta_1 = 6$
- $\beta_0 + 2\beta_1 = 5$
- $\beta_0 + 3\beta_1 = 7$
- $\beta_0 + 4\beta_1 = 10$
- These are a system of four equations in two variables β_0 and β_1.
- Write the equations in the matrix notation:
- $X\beta = y$
- The **design matrix** X is

1	1
1	2
1	3
1	4

- The **parameter vector** β is

β_0
β_1

- The **observation vector** y is

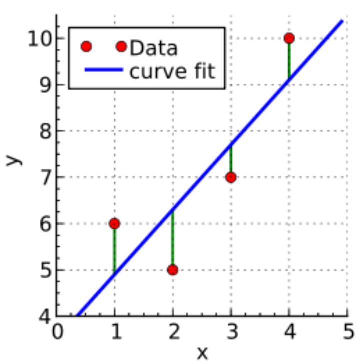
6
5

7
10

- The equation $X\beta = y$ is inconsistent.
- Find the least-squares solution β from the normal equation:
- $X^TX\beta = X^Ty$.
- The augmented matrix of this normal equation is

4	10	28
10	30	77

- The solution is
- $\beta_0 = 3.5$
- $\beta_1 = 1.4$
- The line best fit the four data points is
- $y = 3.5 + 1.4x$



- **(x,y) plane**
- Plot the four data points as red dots
- Plot the least-squares line $y = \beta_0 + \beta_1 x$ as a blue line.
- Plot the four residues as green segments.

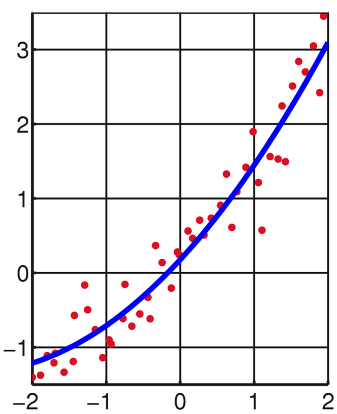
- Calculate the residuals $\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$:
- $\varepsilon_1 = y_1 - (\beta_0 + \beta_1 x_1) = 6 - (3.5 + 1.4 \times 1) = 1.1$
- $\varepsilon_2 = y_2 - (\beta_0 + \beta_1 x_2) = 5 - (3.5 + 1.4 \times 2) = - 1.3$
- $\varepsilon_3 = y_3 - (\beta_0 + \beta_1 x_3) = 7 - (3.5 + 1.4 \times 3) = - 0.7$
- $\varepsilon_4 = y_4 - (\beta_0 + \beta_1 x_4) = 10 - (3.5 + 1.4 \times 4) = 0.9$
- The sum of squares of the residuals is
- $(1.1)^2 + (-1.3)^2 + (-0.7)^2 + (0.9)^2 = 4.2$.

- **Least squares by MATLAB**
- Many MATLAB commands compute least-squares fit.
- Here is one: X\y

- Also see Wikipedia for a Matlab code for the above numerical example.

- **Least squares as a problem of minimization in calculus**
- Given m data points $(x_1, y_1),\ldots, (x_m, y_m)$, find the numbers β_0 and β_1 that best fit the data to the equation $y = \beta_0 + \beta_1 x$.
- To "best fit" is to minimize the **sum of squares**:
- $(\beta_0 + \beta_1 x_1 - y_1)^2 + \ldots + (\beta_0 + \beta_1 x_m - y_m)^2$.
- Write the sum as
- $m\beta_0^2 + 2(x_1 + \ldots + x_m)\beta_0\beta_1 + (x_1^2 + \ldots + x_m^2)\beta_1^2$
- $- 2(y_1 + \ldots + y_m)\beta_0 - 2(x_1y_1 + \ldots + x_my_m)\beta_1$
- $+ y_1^2 + \ldots + y_m^2$.
- The sum is a function of the parameter vector, $g(\beta_0, \beta_1)$.
- Minimize the function $g(\beta_0, \beta_1)$ by setting
- $\partial g(\beta_0, \beta_1)/\partial\beta_0 = 0$,
- $\partial g(\beta_0, \beta_1)/\partial\beta_1 = 0$.
- We obtain that
- $m\beta_0 + (x_1 + \ldots + x_m)\beta_1 = (y_1 + \ldots + y_m)$
- $(x_1 + \ldots + x_m)\beta_0 + (x_1^2 + \ldots + x_m^2)\beta_1 = (x_1y_1 + \ldots + x_my_m)$
- The above is a system of two linear algebraic equations in two variables β_0 and β_1.
- Verify these equations are the same as the normal equation $X^TX\beta = X^Ty$.

- **Least-squares fit to a curve**
- Wiki least squares.



- We have a set of data points $(x_1, y_1), \ldots, (x_m, y_m)$.
- We wish to fit the data points to curve of the form:
- $y = \beta_0 f_0(x) + \ldots + \beta_n f_n(x)$,
- where $f_0(x), \ldots, f_n(x)$ are known functions, and $\beta_0, \ldots, \beta_n$ are parameters to be determined.
- Now the design matrix X is

$f_0(x_1)$	…	$f_n(x_1)$
…	…	…

$f_0(x_m)$	....	$f_n(x_m)$

- The parameter vector β isy

β_0
...
β_n

- Determine the parameter vector β by solving the normal equation
- $X^TX\beta = X^Ty$.

- **Multiple regression**
- A variable y depends on two variables u and v.
- We have data points $(u_1, v_1, y_1),\ldots, (u_m, v_m, y_m)$.
- We wish to fit the data points to
- $y = \beta_0 f_0(u,v) + \ldots + \beta_n f_n(u,v)$,
- where $f_0(u,v), \ldots, f_n(u,v)$ are known functions, and $\beta_0, \ldots, \beta_n$ are parameters to be determined.
- Now the design matrix X is

$f_0(u_1, v_1)$	...	$f_n(u_1, v_1)$
...	...	...
$f_0(u_m, v_m)$	....	$f_n(u_m, v_m)$

- The parameter vector β is

β_0
...
β_n

- Determine the parameter vector β by solving the normal equation
- $X^TX\beta = X^Ty$.

Lay 6.7 Inner product spaces

- Wiki [inner product space]()

- **Binary map**
- Let V be a real vector space.
- Let S be a real scalar set.
- Let f: V × V → S be a map that sends each pair of vectors u and v in V to a scalar s in S.
- The standard notation for such a map is

- s = f(u,v).
- You have studied such maps in multivariable calculus.

- **Example.** A map f: $R^2 \times R^2 \to R$.
- Assume the standard basis for R^2.
- The components of a vector u in R^2 are u_1 and u_2, and the components of a vector v in R^2 are v_1 and v_2.
- Here is a map f: $R^2 \times R^2 \to R$, s = f(u,v).
- $s = \sin u_1 + \cos u_2 + \exp v_1 + \tan v_2$.
- This map is so silly that I have never seen it elsewhere.

Axioms of inner product

- Use a special notation for a binary map that sends each pair of vectors u and v in a vector space V to a scalar s in a scalar set S:
- $s = \langle u, v\rangle$.

- Let u, v, w be vectors in V, and let c be a real number.
- An **inner product** on V is a map V × V → S that satisfies the following axioms.
- **Symmetric**: $\langle u, v\rangle = \langle v, u\rangle$.
- **Bilinear**: $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$, $\langle cu, v\rangle = c\langle u, v\rangle$.
- **Positive-definite**: $\langle v, v\rangle > 0$ for every nonzero v.

- **Positive or negative**
- In defining positive-definiteness, we must specify a unit e for S.
- For any nonzero real number p, pe is a nonzero scalar in S, and is therefore also a unit for S.
- If $p > 0$, a scalar s having positive magnitude relative to e still has positive magnitude relative to pe.
- If $p < 0$, a scalar s having positive magnitude relative to e has negative magnitude relative to pe.

- **Number field vs. scalar set**
- In textbooks, and also in the Wikipedia entry on inner product space, an inner product on a vector space V is defined as a map V × V → R.
- In practice, the inner product is often a map V × V → S.
- Our definition of inner product invokes a scalar set S.
- R is a special scalar set.

- **Inner-product space**
- A vector space with an inner product is called an **inner-product space**.
- To specify an inner-product space, we need three ingredients:

1. a real vector space V,

2. a real scalar set S, and
3. a symmetric, bilinear, positive-definite map V × V → S.

- **Stingy notation ⟨u, v⟩**
- The notation does not indicate the vector space V and the scalar set S.
- Nor does the notation differentiate different inner products on the same vector space.
- We supplement the notation with words.
- Don't be stingy with words.
- Always be explicit about the three ingredients for an inner-product space: the vector space V, the scalar set S, and the binary map V × V → S.

Examples of inner-product spaces

- **The standard inner product on R^2**
- (1) The real vector space is R^2.
- (2) The real scalar set is R.
- (3) The map $R^2 \times R^2 \to R$ is
- $\langle u, v\rangle = u_1v_1 + u_2v_2$.
- In writing this map, we assume the standard basis for R^2.
- The components of a vector u in R^2 are u_1 and u_2, and the components of a vector v in R^2 are v_1 and v_2.
- This map is symmetric, bilinear, and positive-definite, and qualifies as an inner product on R^2.
- This inner product is called the **standard inner product** on R^2.
- This inner product, and its counterpart on R^3, is used to describe length and angle in our physical space.
- The standard inner product is assumed in Lay 6.1-6.6.

- **A (nonstandard) inner product on R^2**
- Keep everything the same as for the standard inner product, but change the map $R^2 \times R^2 \to R$ to
- $\langle u, v\rangle = 3u_1v_1 + 5u_2v_2$.
- This map is symmetric, bilinear, and positive-definite, and qualifies as an inner product on R^2.
- The map specifies a nonstandard inner product on R^2.
- An inner product like this is used in weighted least-squares (Lay 6.8).

- **An inner product on a polynomial space**
- Let V be the (n + 1)-dimensional vector space, in which each element is a polynomial of order up to n.
- A basis for V is a list of polynomials: 1, t, t^2,..., t^n.
- Each element in V is a linear combination of the basis polynomials:
- $p(t) = a_0 + a_1t + a_2t^2 + \dots + a_nt^n$.

- Let R be the scalar set.
- Let $t_0,\ldots,t_n$ be distinct real numbers.
- For two polynomials p and q in V, define an inner product on V by
- $\langle p, q\rangle = p(t_0)q(t_0) + \ldots + p(t_n)q(t_n)$.
- This map V × V → R is symmetric, positive-definite, and bilinear.
- An inner product like this is used in trend analysis of data (Lay 6.8).

- **An inner product on a function space**
- Consider all continuous functions on the interval [a,b].
- Each element in V is a function f: [a,b] → R.
- A special notation for this vector space V = C[a,b].
- Draw some functions in the f-t plane.
- The functions form an infinite-dimensional vector space V.
- Human mind is so malleable that we still benefit from the fake diagram of representing a vector by an arrow, even though a function is a vector in an infinite-dimensional vector space.
- **All diagrams are fake, but some are useful**.
- For each pair of functions f(t) and g(t), define an inner product on V as
- $\langle f, g\rangle = \int f(t)g(t)dt$.
- The integral is over the interval [a,b].
- This map V × V → R is symmetric, bilinear, and positive-definite, and qualifies as an inner product on V.
- Such an inner product is used in Fourier analysis (Lay 6.8), and in polynomial approximation (Lay p. 381, Example 8).

Orthogonality

- We have studied properties of the standard inner product on R^n.
- All properties are derived from the three "basic properties": symmetry, bilinearity, and positive-definiteness.
- We have used the three basic properties as axioms to define general inner-product spaces.
- We now use the axioms to derive all other properties analogous to other properties already familiar to us in R^n.
- This generalization greatly broadens the applications of inner products.

- Two vectors u and v in an inner-product space V are called **orthogonal** if $\langle u, v\rangle = 0$.

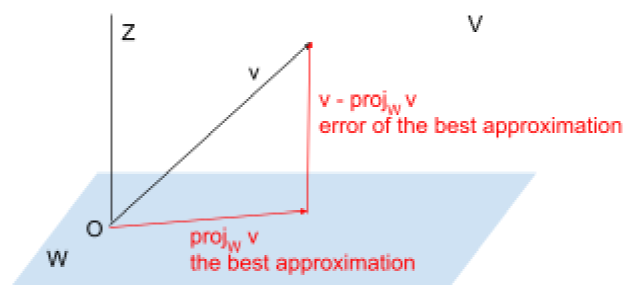
- **Example**
- Two functions sin t and cos t are elements in the function space C [0, 2π].
- The two functions are orthogonal with respect to the inner product
- $\langle f, g\rangle = \int f(t)g(t)dt$, with the integral over the interval [0, 2π].
- Verify $\int(\sin t)(\cos t)dt = 0$.

- This result is also obvious from the graphs of the two functions.

- **Pythagorean theorem**. Two vectors u and v are orthogonal if and only if
- $\langle u - v, u - v \rangle = \langle u, u \rangle + \langle v, v \rangle$.

- **Orthogonal complement**
- Let W be a subspace of an inner-product space V.
- A vector u in V is called orthogonal to W if it is orthogonal to every vector in W.
- The collection of all vectors orthogonal to W is a subspace of V, called the orthogonal complement of W.
- Thus, W and Z are orthogonal complements in a finite-dimensional vector space V if
- (1) W and Z are subspaces in V,
- (2) every vector in W is orthogonal to every vector in Z, and
- (3) dim W + dim Z = dim V.

- **Orthogonal projection**
- Let V be an inner-product space, v be a vector in V, and W be a subspace in V.
- The **orthogonal projection of v onto W** is a vector, written $\text{proj}_W v$, satisfying the following properties:
- (1) $\text{proj}_W v$ is in W, and
- (2) $v - \text{proj}_W v$ is orthogonal to all vectors in W, i.e., $v - \text{proj}_W v$ is in the orthogonal complement of W.



- **Orthogonal decomposition theorem**
- Let W and Z be orthogonal complements in V.
- Any vector v in V can be written as a sum of two vectors, one in W and the other in Z:
- $v = \text{proj}_W v + \text{proj}_Z v$.

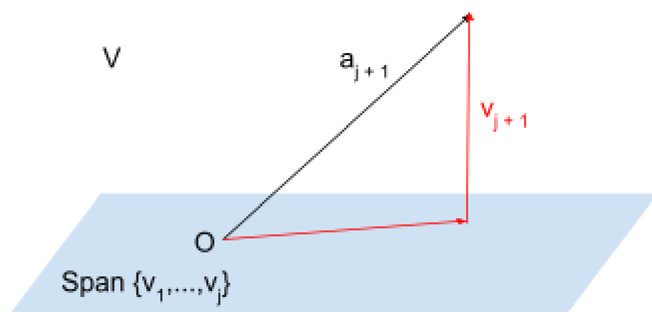
- **Best-approximation theorem**
- Let v be a vector in V, and W be a subspace in V.
- Of all vectors w in W, the vector $w = \text{proj}_W v$ minimizes $\langle v - w, v - w \rangle$.
- $\text{proj}_W v$ is called **the best approximation to a vector v by elements in W**.

- Define the **error (residual) vector** of the best approximation by

- $\text{error}_W\ v = v - \text{proj}_W\ v$.
- The error vector is an element in Z, the orthogonal complement of W.
- The number $||v - \text{proj}_W\ v||$ is called the **distance between the vector v and the subspace W**.

- **Compute orthogonal projection of a vector in V onto a subspace W.**
- Let $v_1,...,v_p$ be an **orthogonal basis** for the subspace W in V.
- That is, $\langle v_i, v_j\rangle = 0$ if $i \neq j$.
- Let v be a vector in V.
- $\text{proj}_W\ v$ is a vector in W, and is therefore a linear combination of the basis vectors for W:
- $\text{proj}_W\ v = c_1v_1 + ... + c_pv_p$.
- The components of the vector $\text{proj}_W\ v$ relative to the basis for W are
- $c_1 = \langle v_1, v\rangle / \langle v_1, v_1\rangle$,
- …..
- $c_p = \langle v_p, v\rangle / \langle v_p, v_p\rangle$.

- **Gram-Schmidt process**
- The process changes an arbitrary basis $a_1,...a_p$ for an inner-product space W to an orthogonal basis $v_1,...,v_p$ for W.
- The process is recursive:
- $v_1 = a_1$,
- …
- $v_{j+1} = \text{error}_{\text{Span}\ \{v1,...,vj\}}\ a_{j+1}$.



Function space and polynomial subspace

- **A vector space V**
- Consider all continuous functions on the interval [0,1].
- Each element in V is a function f: [0,1] → R.
- A special notation for this vector space V = C[0,1].
- The functions form an infinite-dimensional vector space V.

- **An inner product on V**
- For each pair of functions f(t) and g(t), define an inner product on V as

- $\langle f, g\rangle = \int f(t)g(t)dt$.
- The integral is over the interval [0,1].
- This map $V \times V \to R$ is symmetric, bilinear, and positive-definite.

- **A subspace W in V**
- Two functions $a_1(t) = 1$ and $a_2(t) = t$ are linearly independent vectors in V.
- The two functions span a two-dimensional subspace in V:
- $W = \text{Span}\{a_1, a_2\}$.
- a_1 and a_2 are a basis for W.
- Thus, W is a set of functions.
- Each element w in W is a linear combination of the basis vectors:
- $w = c_1 + c_2t$.
- The numbers c_1 and c_2 are the components of w relative to the basis a_1 and a_2.

- **An orthogonal basis for W**
- Use the Gram-Schmidt process to generate an orthogonal basis for W.
- Set $v_1 = a_1$.
- Set $v_2 = a_2 - v_1\langle v_1, a_2\rangle/\langle v_1, v_1\rangle$.
- Calculate the inner products:
- $\langle v_1, a_2\rangle = \int t dt = 0.5$.
- $\langle v_1, v_1\rangle = \int 1 dt = 1$.
- Thus, $v_2 = t - 0.5$.
- Verify that v_1 and v_2 are orthogonal:
- $\langle v_1, v_2\rangle = \int v_1(t)v_2(t)dt = \int(t - 0.5)dt = 0$.
- Thus, $v_1 = 1$ and $v_2 = t - 0.5$ are an orthogonal basis for W.

- **The best approximation to an element in V by elements in W**
- Let $y(t) = t^2$ be an element in V.
- The element in W that best approximates y is
- $\text{proj}_W\, y = v_1\langle v_1, y\rangle / \langle v_1, v_1\rangle + v_2\langle v_2, y\rangle /\langle v_2, v_2\rangle$.
- Calculate the inner products:
- $\langle v_1, y\rangle = \int v_1(t)y(t)dt = \int t^2dt = ⅓$,
- $\langle v_1, v_1\rangle = \int v_1(t)v_1(t)dt = \int 1dt = 1$,
- $\langle v_2, y\rangle = \int v_2(t)y(t)dt = \int(t - 0.5)t^2dt = 1/12$,
- $\langle v_2, v_2\rangle = \int v_2(t)v_2(t)dt = \int(t - 0.5)^2dt = 1/12$.
- $\text{proj}_W\, y = v_1/3 + v_2 = t - ⅙$.
- Compare the function $y(t) = t^2$ with its best approximation $\text{proj}_W\, y = t - ⅙$.

- **Distance between y and W**
- $||y - \text{proj}_W\, y|| = (\int(y - \text{proj}_W\, y)^2dt)^{½} = (\int(y - \text{proj}_W\, y)^2dt)^{½} = (\int(t^2 - t + ⅙)^2dt)^{½} =$ ____

Lay 6.8 Applications of inner-product spaces

- Weighted least squares. Read on your own.
- Trend analysis of data. Read on your own.

Fourier series

- **Function space**
- V = the set of "all functions" on the interval [0, 2π].
- Each element in V is a function f: [0, 2π] → R.
- That is, each function is a vector in V.
- On the f-t plane, draw a few functions, t, sin t...
- Draw a diagram for a v vector in V, and a subspace W in V.
- dim V = ∞.

- **Problem**
- Approximate any function f(t) in V by a linear combination of functions in a subspace W.

- **An inner product on V**
- For each pair of functions f(t) and g(t), define an inner product on V as
- ⟨f, g⟩ = ∫f(t)g(t)dt.
- The integral is over the the interval [0, 2π].
- This map V × V → R is symmetric, bilinear, and positive-definite, and qualifies as an inner product.

- **Subspace**
- A subspace in V is
- W = Span {1, sin t, cos t, sin 2t, cos 2t,...}.
- The trigonometric functions are an **orthogonal basis** for W.
- Verify the orthogonality:
- ∫(sin mt) (sin nt)dt = 0 when m ≠ n,
- ∫(cos mt) (cos nt)dt = 0 when m ≠ n.
- ∫(cos mt) (sin nt)dt = 0.

- **Fourier series**
- The Fourier approximation of a function f(t) in V is the best approximation to f(t) by elements in W:
- f(t) ≅ proj_W f.
- proj_W f is a function in W, and is therefore a linear combination of the basis functions for W:
- f(t) ≅ $a_0/2 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t + \ldots$
- a_k = ⟨f, cos kt⟩/⟨cos kt, cos kt⟩ = ∫f(t)(cos kt)/π,

- $b_k = \langle f, \sin kt\rangle / \langle \sin kt, \sin kt\rangle = \int f(t)(\sin kt)dt/\pi$.

Week 13 (April 24, 26)

Reading and watching

- Lay Chapter 7
- Video on positive-definite matrix
- Video on finding singular value decomposition
- The basic idea of using SVD in image processing is explained in this video.
- Two-part video on an application of SVD in data mining. Part 1 Singular value decomposition. Part 2 Dimensionality reduction.

Lay 7.1 Diagonalization of symmetric matrices

- A matrix A is called **symmetric** if $A^T = A$.

- A symmetric matrix must be a **square** matrix.
- $A_{ij} = A_{ji}$
- This section will focus on real, symmetric, n × n matrices.

- **Examples**
- Here is a real, symmetric, 2 × 2 matrix:

7	2
2	4

- Here is a real, symmetric, 3 × 3 matrix:

7	2	1
2	4	0
1	0	3

The primary theorems of this chapter

- A real and symmetric matrix A has the following properties:
- (a) Every eigenvalue of A is real.
- (b) Different eigenspaces of A are orthogonal.
- (c) The multiplicity of each eigenvalue of A equals the dimension of the corresponding eigenspace.

- **Example**. Consider the matrix A =

7	2
2	4

- The matrix A is 2 × 2, real, and symmetric.
- The characteristic polynomial of the matrix A is
- det (A - λI) = (7 - λ)(4 - λ) - 4 = λ^2 - 11λ + 24 = (λ - 8)(λ - 3).
- The roots of the polynomial are real: λ_1 = 8 and λ_2 = 3.
- For λ_1 = 8, the homogeneous equation is A - 8I = 0, and the corresponding eigenvector is v_1 = (2, 1).
- For λ_2 = 3, the homogeneous equation is A - 3I = 0, and the corresponding eigenvector is v_2 = (-1, 2).
- Verify that v_1 and v_2 are orthogonal, $v_1^T v_2 = 0$.

- The matrix A has two distinct real eigenvalues.
- Corresponding to each eigenvalue is a one-dimensional eigenspace.
- The two eigenspaces are orthogonal.

- **Example**. Consider the identity matrix I =

1	0	0
0	1	0
0	0	1

- The identity matrix is 3 × 3, real, and symmetric.
- The characteristic polynomial of the matrix is $(1 - \lambda)^3 = 0$.
- The matrix I has a single real eigenvalue λ = 1, of multiplicity three.
- Every vector v in R^3 is an eigenvector, Iv = v.
- The eigenspace is R^3.

- The dimension of the eigenspace equals the multiplicity of the eigenvalue.

- **Example**. Consider the matrix A =

1	0	0
0	2	0
0	0	1

- This matrix is 3 × 3, real, and symmetric.
- The matrix A has two real eigenvalues:
- λ_1 = 2, of multiplicity one, and
- λ_2 = 1, of multiplicity two.

- Corresponding to the eigenvalue $\lambda_1 = 2$ is a one-dimensional eigenspace. A basis for this eigenspace is the vector (0, 1, 0).
- Corresponding to the eigenvalue $\lambda_2 = 1$ is a two-dimensional eigenspace. A basis for this eigenspace is the vectors (1, 0, 0) and (0, 0, 1). The two vectors span a subspace in R^3. Every vector in this subspace is an eigenvector corresponding to the eigenvalue $\lambda_2 = 1$.

- The dimension of each eigenspace equals the multiplicity of the corresponding eigenvalue.
- Different eigenspaces are orthogonal.

- A comment by **Jacob Bindman**: "When I am in a hurry, I don't read proofs in the textbook. But when I do read proofs, they often let me understand the theorems."
- My response. Some proofs are tedious and pedantic; they do not shed light, and just weigh us down. They are necessary evils in building mathematics. But some proofs are just what you said. Our textbook is mindful in selecting good proofs and neglecting bad ones. Read the proofs when you do get time. This course is not about proofs, but proofs may aid this course.

Proofs of the primary theorems

- **Theorem (a)**
- If a matrix A is real and symmetric, then every eigenvalue of A is real.

- **Proof**
- For any real and square matrix A, if (λ, v) is an eigenpair of A, so is the complex conjugate (λ^*, v^*).
- Let us first prove this fact.
- If (λ, v) is an eigenpair of A, write
- $Av = \lambda v$.
- Take complex conjugate on both sides of the equation, and we obtain that
- $Av^* = \lambda^* v^*$.
- Here we have used the fact that A is real, so that $A^* = A$.
- The above equation shows that (λ^*, v^*) is an eigenpair of A.

- We next prove that the eigenvalue of A must be real.
- Form inner products between v^* and the vectors on both sides of the equation $Av = \lambda v$:
- $v^{*T}Av = \lambda v^{*T}v$.
- Form inner products between v and the vectors on both sides of the equation $Av^* = \lambda^* v^*$:
- $v^TAv^* = \lambda^* v^Tv^*$.
- Note that $v^{*T}v = v^Tv^*$.
- Given $A^T = A$, we have $v^{*T}Av = v^TAv^*$.
- Thus

- $0 = (\lambda - \lambda^*) v^{*T}v$.
- Note that $v^{*T}v$ is a real and positive number.
- $0 = (\lambda - \lambda^*) v^{*T}v$ implies that
- $\lambda - \lambda^* = 0$.
- This means that the eigenvalues are real.

- **Theorem (b)**
- If A is a symmetric matrix, then eigenvectors corresponding to different eigenvalues are orthogonal.

- **Proof**
- Let λ_1 and λ_2 be two different eigenvalues.
- Let v_1 and v_2 be the corresponding eigenvectors.
- $Av_1 = \lambda_1 v_1$
- $Av_2 = \lambda_2 v_2$
- Form the standard inner products:
- $v_2^T Av_1 = \lambda_1 v_2^T v_1$
- $v_1^T Av_2 = \lambda_1 v_1^T v_2$
- Note that $v_2^T v_1 = v_1^T v_2$.
- Given $A^T = A$, we have $v_2^T Av_1 = v_1^T Av_2$.
- Thus, $0 = (\lambda_1 - \lambda_2) v_1^T v_2$.
- For different eigenvalues, $\lambda_1 \neq \lambda_2$, so that
- $v_1^T v_2 = 0$.
- The two corresponding eigenvectors are orthogonal.

- **Theorem (c)**
- The multiplicity of each eigenvalue of A equals the dimension of the corresponding eigenspace.

- **Proof**
- Multiplicity of eigenvalues is nearly always a fake issue in applications, because a small perturbation of the entries of A will make all eigenvalues distinct.
- This insight gives us some feel for Theorem (c).
- Let A be a real and symmetric matrix, having an eigenvalue λ of multiplicity k.
- Slightly perturb the entries of A, so that the matrix is still real and symmetric, but the eigenvalue λ splits into k distinct eigenvalues.
- According to Theorems (a) and (b), the k eigenvalues are real, and the k corresponding eigenvectors are an orthogonal set.
- We then gradually reduce the perturbation, so that k eigenvalues gradually collapse to the single eigenvalue λ, of multiplicity k.
- As we gradually reduce the perturbation, the k corresponding eigenvectors only rotate slightly, and cannot collapse on top of one another.
- Consequently, the eigenvalue λ of multiplicity k has k corresponding eigenvectors.

Spectral decomposition

- Let A be a real, symmetric, n × n matrix.

- **A has n orthonormal eigenvectors**.
- Theorem (a): every eigenvalue is real.
- Theorem (b): eigenspaces are orthogonal.
- Theorem (c): an eigenvalue of multiplicity k corresponds to a k-dimensional eigenspace.
- Each vector in the k-dimensional eigenspace is an eigenvector corresponding to the same eigenvalue λ.
- There exist k linearly independent eigenvectors in the k-dimensional eigenspace.
- These eigenvectors may not be an orthogonal set, but we can always use the Gram -Schmidt process to find find an orthogonal set of k eigenvectors.

- Let $u_1,...,u_n$ be n **orthonormal eigenvectors** of A.
- Let $\lambda_1, \ldots,\lambda_n$ be the corresponding eigenvalues.
- $Au_1 = \lambda_1 u_1$,
- …...
- $Au_n = \lambda_n u_n$.
- The n eigenvectors are an orthogonal basis for R^n.

- Any vector v in R^n is a linear combination of the orthonormal basis vectors $u_1,...,u_n$:
- $v = u_1(u_1^T v) + \ldots + u_n(u_n^T v)$.
- The numbers $u_1^T v, \ldots, u_n^T v$ are the components of the vector v in R^n relative to the orthonormal basis $u_1,...,u_n$ for R^n.
- We have learned this fact in Chapter 6.

- Calculate Av:
- $Av = \lambda_1 u_1(u_1^T v) + \ldots + \lambda_n u_n(u_n^T v)$.
- The numbers $\lambda_1(u_1^T v),..., \lambda_n(u_n^T v)$ are the components of the vector Av in R^n relative to the orthonormal basis $u_1,...,u_n$ for R^n.

- The above equation is valid for every vector v in R^n.
- Dropping v on both sides of the equation, we write
- $A = \lambda_1 u_1 u_1^T + \ldots + \lambda_n u_n u_n^T$.
- This equation is called the **spectral decomposition** (or eigenvalue decomposition) of the matrix A.
- A real, symmetric, n × n matrix A is a sum of n matrices of rank 1.
- For example, every column of $\lambda_1 u_1 u_1^T$ is a multiple of u_1.

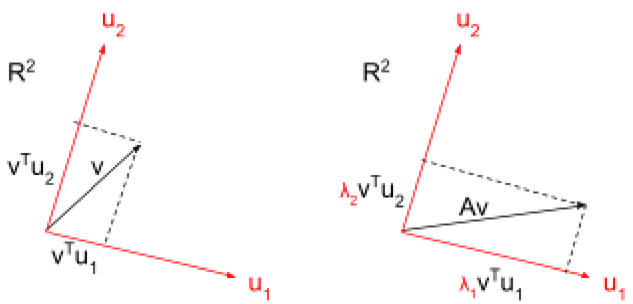
- Incidentally, uu^T is called an **outer product** of the vector u.
- The outer product of the vector u gives a matrix.

- The inner product of the vector u gives a number.

Graphical interpretation of the spectral decomposition in R^2

- Everything happens in a single vector space, R^2.
- But for visual clarity, I draw two copies of the same space.
- u_1 is a unit vector: an eigenvector of A.
- λ_1 is a number: an eigenvalue of A.
- Similar interpretation for u_2.
- The two eigenvectors are orthogonal, $u_2^T u_1 = 0$.

- In R^2, the equations reduce to
- $v = u_1(u_1^T v) + u_2(u_2^T v)$,
- $Av = \lambda_1 u_1(u_1^T v) + \lambda_2 u_2(u_2^T v)$.
- Let us interpret the terms piece by piece, using figures.
- In drawing the figures, I have taken $\lambda_1 \cong 2$ and $\lambda_2 \cong ½$.



- v is a vector in R^2.
- $(u_1^T v)$ is a number: the magnitude of the orthogonal projection of the vector v onto the subspace Span $\{u_1\}$.

- Av is another vector in R^2.
- $\lambda_1(u_1^T v)$ is a number: the magnitude of the orthogonal projection of the vector Av onto the subspace Span $\{u_1\}$.

- What does the operator A do to the vector v?
- Av is a vector, scales one component of v by λ_1, and scales the other component of v by λ_2.

Diagonalization

- For a square matrix A of n linearly independent real eigenvectors, list the eigenvectors as columns of a matrix,
- $P = [u_1 \ldots u_n]$.
- List the corresponding eigenvalues as a diagonal matrix,

- $D = \text{diag}[\lambda_1,\ldots,\lambda_n]$.
- Recall $Au_1 = \lambda_1 u_1, \ldots, Au_n = \lambda_n u_n$.
- Rewrite these n equations as
- $AP = PD$.
- Right-multiplying this equation by P^{-1} gives that
- $A = PDP^{-1}$.
- If A has n linearly independent real eigenvectors, then A can be **diagonalized** using the n eigenvectors.
- We have learned this fact in Chapter 5.

- P is the change-of-basis matrix from the standard basis for R^n to the eigenbasis for R^n.
- A is the matrix of an operator T relative to the standard basis.
- D is the matrix of the same operator T relative to the eigenbasis.
- To diagonalize a matrix A is to change from the standard basis to the eigenbasis.

Orthogonal diagonalization

- Now for a symmetric matrix A, we can make the n eigenvectors $u_1,\ldots,u_n$ be an **orthonormal set**.
- The matrix A is diagonalized using n orthonormal eigenvectors.
- A real and symmetric matrix is **orthogonally diagonalizable**.
- Recall that an orthogonal matrix P satisfies the identity
- $P^{-1} = P^T$.
- Write $A = PDP^{-1}$ as
- $A = PDP^T$.

Lay 7.2 Quadratic forms

- A **quadratic form** on R^n is a map that sends a vector x in R^n to a real number,
- $Q: R^n \to R$,
- $Q(x) = x^TAx$.
- A is a real, symmetric, $n \times n$ matrix, and is called **the matrix of the quadratic form** Q.

- **Example**
- Here is a quadratic form on R^2:
- $Q(x) = 7x_1^2 + 4x_1x_2 + 4x_2^2$.
- The two numbers x_1 and x_2 are the components of a vector x in R^2 relative to the standard basis for R^2.
- Write $Q(x) = x^TAx$.
- The matrix of the quadratic form is A =

7	2

2	4

- The coefficient for each cross-product term in Q is twice an off-diagonal entry in the matrix A.
- Verify this statement by multiplying x^TAx.

Eigenbasis

- The matrix A of a quadratic form on R^n is a real, symmetric, n × n matrix.
- Thus, A has n orthonormal eigenvectors, $u_1,\ldots,u_n$, and has n corresponding real eigenvalues $\lambda_1,\ldots,\lambda_n$, counting multiplicity.
- $P = [u_1 \ldots u_n]$.
- $D = \text{diag}\,[\lambda_1,\ldots,\lambda_n]$.
- $A = PDP^{-1} = PDP^T$.
- $Q = x^TAx = x^TPDP^Tx = (P^Tx)^TD(P^Tx)$
- Write $x = Py$.
- Recall $P^TP = I$.
- $y = P^Tx$.
- $Q = y^TDy$.
- $Q = \lambda_1y_1^2 + \ldots + \lambda_ny_n^2$.

- We next interpret these equations in terms of a change of basis for R^n.
- x is the column of a vector v in R^n relative to the standard basis for R^n.
- y is the column of the same vector v in R^n relative to the eigenbasis for R^n.
- A is the matrix of a quadratic form Q on R^n relative to the standard basis for R^n.
- D is the matrix of the same quadratic form Q on R^n relative to the eigenbasis for R^n.

- **Principal axes theorem**
- In an eigenbasis, the quadratic form has no cross-product terms.

Graphical interpretation of quadratic forms on R^2

- **See Lay pp. 404-405 for figures**.
- A quadratic form in the standard basis:
- $Q = A_{11}x_1^2 + 2A_{12}x_1x_2 + A_{22}x_2^2$.
- Change to the eigenbasis by setting $x = Py$, so that
- $Q = \lambda_1y_1^2 + \lambda_2y_2^2$.

- **Curves in R^2**
- c = constant
- $Q = \lambda_1y_1^2 + \lambda_2y_2^2 = c$ represents curves in R^2.
- Use y_1 and y_2 as axes in R^2.
- For example, $\lambda_1 = 1$, $\lambda_2 = 1/4$, $c = 1$, the curve is an ellipse.

- **Surfaces in R^3**
- $Q = \lambda_1 y_1^2 + \lambda_2 y_2^2$.
- Use y_1, y_2, and Q as three axes in R^3.

- **Positive-definite quadratic form**
- A quadratic form Q(x) is called positive-definite if Q(x) > 0 for all nonzero vector x.
- A quadratic form $x^T Ax$ is positive-definite if the eigenvalues of A are all positive.

- We can make similar statements about **negative-definite** and **indefinite** quadratic forms.

Lay 7.3 Constrained optimization

- **Theorem**
- Let A be a real, symmetric, n × n matrix.
- List the eigenvalues of A in the descending order: $\lambda_1 \geq \ldots \geq \lambda_n$.
- Let the corresponding orthonormal eigenvectors be $u_1,...,u_n$.
- Let x be a unit vector in R^n.
- $x^T Ax$ reaches maximum λ_1 when $x = u_1$, and reaches minimum λ_n when $x = u_n$.

- **Proof**
- On changing from the standard basis to an eigenbasis, the column of vector changes from x to y,
- x = Py.
- We obtain that
- $x^T Ax = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2$.
- Given a unit vector x, because P is a orthogonal matrix, y is also a unit vector.

Basis and change of basis

- I asked **Daniel Rosenblatt** for suggestions of topics to be dropped when I teach the course next time.
- “I don’t know,” he said. “Everything seems to be essential. But wait! How about change of basis?”
- Well, we need a basis for a vector space V to turn a v vector in V into a column of numbers, and turn an operator T on V into a matrix of numbers.
- Some bases are nicer than others: they make the matrix of an operator simpler. That is why we change basis.
- Chapter 4 introduces basis and change of basis.
- Chapter 5 changes an arbitrary basis to an eigenbasis.
- Chapter 6 Gram-Schmidt processes an arbitrary basis to an orthonormal basis.
- Chapter 7 changes bases for both the domain and the codomain to singular bases.

Email to the class before the last lecture

Our last lecture will be devoted to singular value decomposition (SVD), a singularly valuable decomposition.

SVD has become a workhorse for statistics, data science, and machine learning. Wiki singular value decomposition to have an overview of its range of applications.

SVD is becoming the climax of introductory linear algebra courses. SVD integrates every significant idea we have learned so far. Thus, to lecture on SVD is to review the entire course. We will revisit vector spaces, subspaces, linear maps, ranks, eigenvalues, inner products, and outer products. We will meet again chickens, rabbits, and hamsters, and let them tell us the meaning of SVD. We will learn how Netflix uses the rankings of movies by viewers.

Whenever you have a matrix, ask what the SVD of the matrix means.

Let me paraphrase our better known alumnus. And so my fellow algebraists, ask not what you can do for SVD, ask what SVD can do for you.

Come to the last lecture on Wednesday. We will celebrate the end of the course, and the beginning of a deeper understanding of linear algebra.

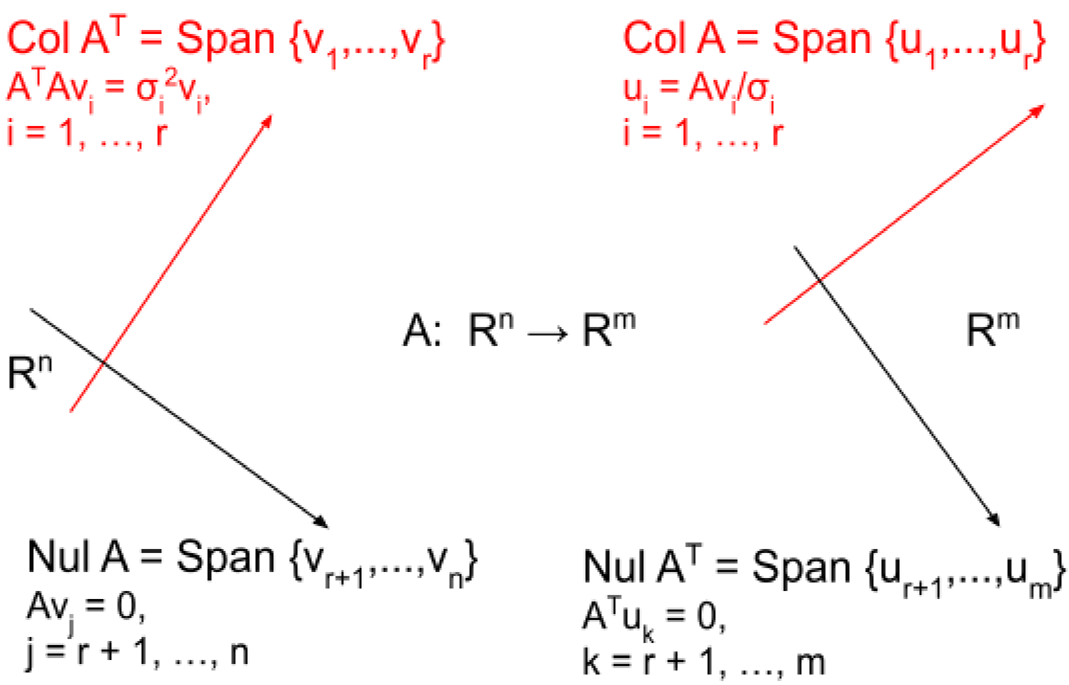
Lay 7.4 Singular value decomposition (SVD)

- **Linear map**
- Let A be an $m \times n$ real matrix.
- The linear map A sends a vector x in R^n to a vector y in R^m:
- y = Ax.

- **Rank**
- r = rank A = the number of pivots in A.
- $r \le n$.
- $r \le m$.

- **Subspaces in R^n and R^m**
- We have already learned the facts in black.
- We will next learn the facts in red: discover “nice” bases for the row space and column space of a linear map A.

Domain, R^n	Codomain, R^m
Nul A = $\{x \mid Ax = 0\}$	Nul $A^T = \{y \mid A^Ty = 0\}$
dim Nul A = n - r	dim Nul A^T = m - r
Nul A = Span $\{v_{r+1},\ldots v_n\}$	Nul A^T = Span $\{u_{r+1},\ldots u_m\}$
Col $A^T = \{x \mid x = A^Ty$ for all y in $R^m\}$	Col A = $\{y \mid y = Ax$ for all x in $R^n\}$
dim Col A^T = r	dim Col A = r
Col A^T and Nul A are orthogonal complements in R^n	Col A and Nul A^T are orthogonal complements in R^m
A^TA = symmetric n × n matrix	AA^T = symmetric m × m matrix
Nul A^TA = Nul A	Nul AA^T = Nul A^T
rank A^TA = r	rank AA^T = r
$A^TAv = \sigma^2 v$	$AA^Tu = \sigma^2 u$
Col A^T = Span $\{v_1,\ldots,v_r\}$	Col A = Span $\{u_1,\ldots,u_r\}$



- **All diagrams are fake, but some are useful**
- Many graphical representations of these facts exist online.

- Here is a diagram of the four subspaces associated with the linear map A: $R^n \to R^m$, along with a choice of basis for each subspace.
- Arrows in a plane is a two-dimensional vector space.
- I represent each subspace by a single arrow.
- The left side of the diagram represents R^n.
- The right side of the diagram represents R^m.
- Compare the diagram with the table.

Singular values and singular vectors

- A is an m × n real matrix.

- A^TA is a real, symmetric, n × n matrix.
- A^TA has n real eigenvalues (counting multiplicity), and n orthonormal eigenvectors in R^n.

- AA^T is a real, symmetric, m × m matrix
- AA^T has m real eigenvalues (counting multiplicity), and m orthonormal eigenvectors in R^m.

- **Let λ = 0 be an eigenvalue of A^TA**, and v be a corresponding eigenvector. Thus,
- $A^TAv = 0$.
- Left-multiplying v^T to the equation $A^TAv = 0$ gives $(Av)^T(Av) = 0$.
- Consequently, $A^TAv = 0$ if and only if $Av = 0$.
- Nul A^TA = Nul A.
- dim Nul A^TA = n - r,
- rank A^TA = r.
- All eigenvectors of A^TA corresponding to the eigenvalue λ = 0 form an eigenspace, which is a subspace in R^n, and is identical to Nul A^TA and Nul A.

- Similarly,
- Nul AA^T = Nul A^T.
- dim Nul AA^T = m - r,
- rank AA^T = r.
- All eigenvectors of AA^T corresponding to the eigenvalue λ = 0 form an eigenspace, which is a subspace in R^m, and is identical to Nul AA^T and Nul A^T.

- **Let λ ≠ 0 be an eigenvalue of A^TA**, and v be a corresponding unit eigenvector. Thus,
- $A^TAv = \lambda v$.
- Left-multiplying v^T to the equation $A^TAv = \lambda v$ gives that
- $(Av)^T(Av) = (Av)^T(Av) = v^T(A^TAv) = v^T\lambda v = \lambda$.
- Consequently, λ > 0, and the length of the vector Av is $\sqrt{\lambda}$.
- Left-multiplying A to the equation $A^TAv = \lambda v$ gives that $AA^TAv = \lambda Av$, so that (λ, Av) is an eigenpair of AA^T.

- We reach the following result:
- If a unit vector v in R^n is an eigenvector of A^TA corresponding to a nonzero eigenvalue λ, then λ > 0, and u = Av/√λ is a unit vector in R^m and is an eigenvector of AA^T corresponding to λ.

- The positive number $\sigma = \sqrt{\lambda}$, the unit eigenvector u of AA^T, and the unit eigenvector vector v of A^TA are called, respectively, the **corresponding singular value, singular vector in R^m, and singular vector in R^n**.

- Write
- $Av = \sigma u$,
- $A^Tu = \sigma v$.

Four fundamental subspaces created by a linear map

- A = m × n real matrix A
- r = rank of A.
- A has r positive singular values, $\sigma_1, \ldots \sigma_r$.

- Let $v_1, \ldots, v_r$ be the corresponding singular vectors in R^n.
- Col A^T = Span $\{v_1, \ldots, v_r\}$.
- $A^TAv_j = \sigma_j^2 v_j$, j = 1,...,r.

- Let $u_1, \ldots, u_r$ be the corresponding singular vectors in R^m.
- Col A = Span $\{u_1, \ldots, u_r\}$.
- $AA^Tu_j = \sigma_j^2 u_j$, j = 1,...,r.

- Let $v_1 \ldots v_r, v_{r+1}, \ldots, v_m$ be an orthonormal basis for R^n, which we call a **singular basis for R^n**.
- Nul A = Span $\{v_{r+1}, \ldots, v_n\}$.
- $Av_j = 0$, j = r + 1,...,n.

- Let $u_1 \ldots u_r, u_{r+1}, \ldots, u_m$ be an orthonormal basis for R^m, which we call a **singular basis for R^m**.
- Nul A^T = Span $\{u_{r+1}, \ldots, u_m\}$.
- $A^Tu_j = 0$, j = r + 1,...,m.

- $Av_1 = \sigma_1 u_1$,
- ……
- $Av_r = \sigma_r u_r$.
- $Av_{r+1} = 0$,
- …..

- $Av_n = 0$.

Singular value decomposition

- Any vector x in R^n is a linear combination of the orthonormal basis vectors $v_1,...,v_n$:
- $x = v_1v_1^Tx + \ldots + v_nv_n^Tx$.
- The numbers $v_1^Tx, \ldots, v_n^Tx$ are the components of the vector x in R^n relative to the orthonormal basis $v_1,...,v_n$ for R^n.

- Calculate Ax, and we obtain that
- $Ax = \sigma_1u_1v_1^Tx + \ldots + \sigma_ru_rv_r^Tx$.
- The numbers $\sigma_1(v_1^Tx), \ldots \sigma_r(v_r^Tx)$ are the components of the vector Ax in Col A relative to the orthonormal basis $u_1,...,u_r$ for Col A.
- Ax is a vector in Col A, which is a subspace in R^m.
- Consequently, the components of Ax relative to $u_{r+1},...,u_m$ vanish.

- The last equation is valid for any vector x in R^n.
- Thus, the matrix A itself must be
- $A = \sigma_1u_1v_1^T + \ldots + \sigma_ru_rv_r^T$.
- This equation is called the **reduced singular value decomposition** of the matrix A.
- r = rank A.
- $\sigma_1, \ldots, \sigma_r$ are the positive singular values.
- $u_1, \ldots, u_r$ are the orthonormal singular vectors in R^m.
- $v_1, \ldots, v_r$ are the orthonormal singular vectors in R^n.

- **Outer product vs. inner product**
- For a vector v in R^n and a vector u in R^m, uv^T is called the **outer product**.
- The outer product of the two vectors is an m × n matrix.
- The rank of the outer product is 1, because every column of the matrix uv^T is a multiple of u.

- By contrast, we can only form an inner product of two vectors in the same vector space, and the inner product of the two vectors is a number.

- **Singular value decomposition vs. spectral decomposition**
- For a real, symmetric, n × n matrix A, we regard A as a linear map on one vector space, $A: R^n \rightarrow R^n$.
- There exist n orthonormal eigenvectors $u_1,...,u_n$, and corresponding real eigenvalues $\lambda_1,...,\lambda_n$.
- The **spectral decomposition** of A is
- $A = \lambda_1u_1u_1^T + \ldots + \lambda_nu_nu_n^T$.

- By contrast, singular value decomposition works for any real m × n matrices.

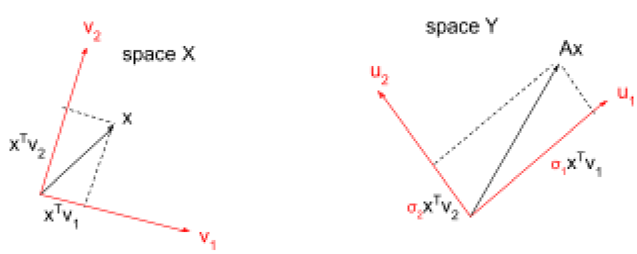
- The two sets of singular vectors are in different spaces.
- When the matrix A is symmetric, the singular value decomposition recovers the spectral decomposition.

A graphical interpretation of SVD

- A picture is worth a thousand words.
- But a picture without words is usually misleading.
- An equation ($A = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$) is worth [a thousand pictures](#).

- Here I add one more picture, along with words and equations.
- Consider the case m = n = r = 2,
- A: X → Y.
- X and Y are different two-dimensional real vector spaces.
- For example, X = (chicken-rabbit space), Y = (head-foot space).
- (It would be confusing if I called both vector spaces R^2.)

- v_1 and v_2 are orthonormal singular vectors in X, and are a basis for X.
- Because n = r = 2, Nul A = 0.
- u_1 and u_2 are orthonormal singular vectors in Y, and are a basis for Y.
- Because m = r = 2, Nul A^T = 0.
- σ_1 and σ_2 are positive singular values.



- **What does the linear map A do to a vector x in X**?
- The singular vectors of A define an orthonormal basis v_1 and v_2 for X, and an orthonormal basis u_1 and u_2 for Y.
- A vector x in X is a linear combination of the basis vectors for X:
- $x = v_1 v_1^T x + v_2 v_2^T x$.
- The numbers $v_1^T x$ and $v_2^T x$ are the components of the vector x in X relative to the orthonormal basis v_1, v_2 for X.
- Ax is a vector in Y, and is a linear combination of the basis vectors for Y:
- $Ax = \sigma_1 u_1 v_1^T x + \sigma_2 u_2 v_2^T x$.
- The numbers $\sigma_1 v_1^T x$ and $\sigma_2 v_2^T x$ are the components of the vector Ax in Y relative to the orthonormal basis u_1, u_2 for Y.

- That is, A scales one component of a vector by σ_1, and scales the other component of the vector by σ_2.

Low-rank matrix approximation

- This approximation underlies most applications of SVD.
- In the SVD of A,
- $A = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$,
- All singular vectors are unit vectors, so that the entries of the rank-one matrices are of similar magnitude.
- However, the singular values $\sigma_1, \ldots, \sigma_r$ may differ greatly.

- In the past, the rank r of a matrix A is just an integer.
- This idea of rank may not be as useful as it appears, because a slight perturbation of the entries of the matrix A can change the rank of A.
- Not all rank-one matrices contribute equally to A.
- Each singular value measures the **strength** of the contribution of a rank-one matrix to A.

- For example, if σ_1 is much larger than the other singular values, we way choose to approximate the matrix as a single rank-1 matrix:
- $A \cong \sigma_1 u_1 v_1^T$.
- For a given x in R^n, we have
- $Ax \cong \sigma_1 u_1 v_1^T x$.
- $v_1^T x$ is the magnitude of x projected onto the singular vector v_1.
- Ax is a vector in R^m in the direction u_1, with the magnitude $\sigma_1 v_1^T x$.

- More generally, a rank-k approximation is
- $A = \sigma_1 u_1 v_1^T + \ldots + \sigma_k u_k v_k^T$.
- The singular values help us to decide the rank of the approximation, the corresponding singular vectors give the bases for the subspaces in R^n and R^m.

An alternative expression of singular value decomposition

- Write the singular value decomposition,
- $A = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$,
- as
- $AV = U\Sigma$.
- $V = [v_1 \ldots v_n]$ = n × n orthogonal matrix

- $U = [u_1 \dots u_m]$ = m × m orthogonal matrix
- Σ = m × n matrix =

σ_1	0	0	0	0	0	0
0	…	0	0	0	0	0
0	0	σ_r	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

- Right-multiplying V^T gives
- $A = U\Sigma V^T$.
- This equation is called the **extended singular decomposition**.
- Adding zeros does not change A.
- But U lists a full orthonormal basis for R^m, and V lists a full orthonormal basis for R^n.
- The full orthonormal bases can be useful.

- **Singular value decomposition means the change of the matrix of a linear map T under a change of basis for R^m and a change of basis for R^n**
- $A = U\Sigma V^T$.
- U is the change-of-basis matrix from the standard basis for R^m to a singular basis for R^m.
- V is the change-of-basis matrix from the standard basis for R^n to a singular basis for R^n.
- A is the matrix of a linear map T relative to the standard basis for R^m and the standard basis for R^n.
- Σ is the matrix of the same linear map T relative to the singular basis for R^m and the singular basis for R^n.
- The singular basis for R^m and the singular basis for R^n are "nice" because the matrix of the linear map T becomes a simple matrix Σ.

Compute singular value decomposition

- **Singular value decomposition by MATLAB**
- Command: [U,S,V] = svd(A).

- **Singular value decomposition by hand**
- Watch video.
- Read Examples 3 and 4 on Lay pp. 418-420.

- **Steps to calculate the reduced singular decomposition**
- $A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$.

- Step 1. Solve the eigenvalue problem $A^TAv = \lambda v$. The eigenvalues of A^TA give the singular values: $\sigma_1 = \sqrt{\lambda_1}$, ..., $\sigma_r = \sqrt{\lambda_r}$. The corresponding eigenvectors of A^TA give the singular vectors in R^n: v_1,..., v_r.
- Step 2. Calculate the singular vectors in R^m: $u_1 = Av_1/\sigma_1$,..., $u_r = Av_r/\sigma_r$.

- **Additional steps to obtain the extended singular value decomposition**
- $A = U\Sigma V^T$ (**Example 4, Lay p. 419**).
- Step 3. List the n orthonormal eigenvectors of A^TA as $V = [v_1 \dots v_n]$.
- Step 4. Construct the m × n matrix Σ.
- Step 5. Find a basis for the orthogonal complement of the subspace Span $\{u_1,...u_r\}$. That is, find the solution set of $A^Ty = 0$.
- Step 6. Gram-Schmidt process the basis for the solution set to an orthonormal basis u_{r+1},...u_m. Write $U = [u_1 \dots u_m]$.

(chicken-rabbit space) → (head-foot space)

- Axiom 1 of linear algebra: All ideas must pass the chicken-rabbit test.
- Our worldly knowledge boils down to a matrix A =

	chicken	rabbit
head	1	1
foot	2	4

- Let x be a chicken-rabbit vector:

number of chickens
number of rabbits

- Let y be a head-foot vector:

number of heads
number of feet

- The matrix A maps a chicken-rabbit vector x to a head-foot vector y:
- y = Ax

- Use the MATLAB to find the SVD of the matrix A:

```
>> A = [1,1;2,4]
A =
    1     1
    2     4
```

```
>> [U,S,V] = svd(A)

U =
   0.2898   -0.9571
   0.9571    0.2898

S =
    4.6708         0
         0    0.4282

V =
   0.4719   -0.8817
   0.8817    0.4719
```

- Recall $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$
- In the chicken-rabbit test, $\sigma_1 = 4.6708$ and $\sigma_2 = 0.4282$.
- The two singular values differ by more than a factor of 10.
- Thus, the matrix A mainly comes from the rank-one matrix $\sigma_1 u_1 v_1^T$.
- MATLAB gives $\sigma_1 u_1 v_1^T$

0.6388	1.1935
2.1096	3.9416

- This rank-one matrix is close to our original matrix A

1	1
2	4

- Let us see how the low-rank approximation $A \cong \sigma_1 u_1 v_1^T$ will do.
- v_1 = (0.4719, 0.8817) is the **singular chicken-rabbit vector** corresponding to σ_1
- u_1 = (0.2898, 0.9571) is the **singular head-foot vector** corresponding to σ_1.
- For a chicken-rabbit vector x, the low-rank approximation gives a head-foot vector y:
- $y = Ax \cong \sigma_1 u_1 v_1^T x$.
- Thus, the rank-one approximation does three things in succession:
- projects a chicken-rabbit vector x onto v_1,
- maps to a head-foot vector in the direction u_1, and
- stretches by the factor σ_1.

- For the numbers of chickens and rabbits on the farm
- x = (11, 15),
- the low-rank approximation predicts that the numbers of heads and feet are
- y ≅ (24.9284, 82.3291).
- The approximate numbers are remarkably close to the exact numbers (26, 82).

(chicken-rabbit-hamster space) → (head-foot space)

- The matrix of this linear map is A =

	chicken	rabbit	hamster
head	1	1	1
foot	2	4	4

- Here is the SVD of this matrix obtained using MATLAB
- Try to interpret it in the language of the chicken-rabbit-hamster space and the head-foot space.

```
>> A = [1,1,1;2,4,4]

A =
    1     1     1
    2     4     4

>> [U,S,V] = svd(A)

U =
   0.2691   -0.9631
   0.9631    0.2691

S =
    6.2285         0         0
         0    0.4541         0

V =
   0.3525   -0.9358   -0.0000
   0.6617    0.2492   -0.7071
   0.6617    0.2492    0.7071
```

Viewers rank movies

- A movie-streaming service, say Netflix, records the rankings of movies by viewers.
- A New York Times article on the Netflix Prize: All competing teams use SVD.
- Wiki Netflix Prize.
- The following example appears in the video course from Stanford on data mining.

Rankings of five movies by seven viewers

	Matrix	Alien	Serenity	Casabanca	Amelie
Alice	1	1	1	0	0
Bob	3	3	3	0	0
Cate	4	4	4	0	0
Dan	5	5	5	0	0
Eric	0	2	0	4	4
Felicia	0	0	0	5	5
Greg	0	1	0	2	2

- By inspection, we see that
- Dan likes Matrix, Alien, and Serenity.
- Felicia likes Casablanca and Amelie.

- In this example, A is a matrix of numbers, rankings by viewers for movies.
- Each movie has a column vector, which we call a **movie vector**.
- Each viewer has a row vector, which we call a **viewer vector**.
- What can svd (A) do for us, or for Netflix?

- Let us first obtain svd (A) using MATLAB.

```
>> [U,S,V] = svd(A)

U =
   -0.1376    0.0236    0.0108    0.9901   -0.0000   -0.0000         0
   -0.4128    0.0708    0.0324   -0.0594   -0.8850    0.1916         0
   -0.5504    0.0944    0.0432   -0.0792    0.4243    0.7071         0
   -0.6880    0.1181    0.0540   -0.0990    0.1916   -0.6807         0
   -0.1528   -0.5911   -0.6537   -0.0000    0.0000   -0.0000   -0.4472
   -0.0722   -0.7313    0.6782    0.0000   -0.0000         0         0
   -0.0764   -0.2956   -0.3268   -0.0000   -0.0000   -0.0000    0.8944

S =
   12.4810         0         0         0         0
         0    9.5086         0         0         0
```

0	0	1.3456	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

V =

-0.5623	0.1266	0.4097	-0.7071	0
-0.5929	-0.0288	-0.8048	0.0000	0.0000
-0.5623	0.1266	0.4097	0.7071	-0.0000
-0.0901	-0.6954	0.0913	-0.0000	0.7071
-0.0901	-0.6954	0.0913	0.0000	-0.7071

- svd (A) gives three positive singular values:
- $\sigma_1 = 12.4810$
- $\sigma_2 = 9.5086$
- $\sigma_3 = 1.3456$

- **What does svd mean?**
- Let us look at the largest singular value σ_1, and its corresponding **singular movie vector** u_1 and **singular viewer vector** v_1:
- $u_1 = (0.1376, 0.4128, 0.5504, 0.6880, 0.1528, 0.0722, 0.0764)$
- $v_1 = (0.5623, 0.5929, 0.5623, 0.0901, 0.0901)$

- The **singular movie vector** u_1 does not represent the rankings of any particular movie.
- But we can calculate the component of each real movie vector projected onto the singular movie vector u_1:
- u_1^T(Matrix) = 7.0176
- u_1^T(Alien) = 7.3996
- u_1^T(Serenity) = 7.0176
- u_1^T(Casablenca) = 1.1250
- u_1^T(Amilie) = 1.1250
- A movie vector with a large magnitude projected on u_1 is “close” to u_1.

- Similarly, the **singular viewer vector** v_1 does not represent any real viewer.
- But we can calculate the component of each real viewer vector projected onto the singular viewer vector v_1:
- v_1^T(Alice) = 1.7175
- v_1^T(Bob) = 5.1525
- v_1^T(Cate) = 6.8700
- v_1^T(Dan) = 8.5875
- v_1^T(Eric) = 1.9066
- v_1^T(Felicia) = 0.9010

- v_1^T(Greg) = 0.9533
- A viewer vector with a large magnitude projected on v_1 is "close" to v_1.

- Thus, svd (A) discovers a movie concept, which we call SiFi.
- The movies Matrix, Alien, Serenity belong to this concept.
- The viewers Bob, Cate and Dan love this concept.

- The second largest singular value σ_2 = 9.5086 identifies another movie concept, which we call romance.
- The corresponding singular movie vector is
- u_2 = (0.0236, 0.0708, 0.0944, 0.1181, -0.5911, -0.7313, -0.2956)
- The corresponding singular viewer vector is
- v_2 = (0.1266, -0.0288, 0.1266, -0.6954, -0.6954)

- u_2^T(Matrix) = 1.2041
- u_2^T(Alien) = -0.2737
- u_2^T(Serenity) = 1.2041
- u_2^T(Casablenca) = -6.6121
- u_2^T(Amilie) = -6.6121

- v_2^T(Alice) = 0.2244
- v_2^T(Bob) = 0.6732
- v_2^T(Cate) = 0.8976
- v_2^T(Dan) = 1.1220
- v_2^T(Eric) = -5.6208
- v_2^T(Felicia) = -6.9540
- v_2^T(Greg) = -2.8104

- The moves Casablenca and Amilie belong to this concept.
- The viewers Eric and Felicia love this concept.

Pseudoinverse

- The **pseudoinverse** of a linear map A is defined by
- $A^+ = \sigma_1^{-1} v_1 u_1^T + \ldots + \sigma_r^{-1} v_r u_r^T$.

- Any vector y in R^m is a linear combination of the orthonormal basis vectors $u_1,\ldots,u_m$:
- $y = u_1(u_1^T y) + \ldots + u_m(u_m^T y)$.
- The numbers $u_1^T y, \ldots, u_m^T y$ are the components of the vector y in R^m relative to the orthonormal basis $u_1,\ldots,u_m$ for R^m.

- The linear map A^+ sends the vector y in R^m to a vector A^+y in Col A^T, which is a subspace in R^n:

- $A^+y = \sigma_1^{-1}v_1(u_1^Ty) + \ldots + \sigma_r^{-1}v_r(u_r^Ty)$.
- The vector A^+y is a linear combination of the orthonormal basis vectors $v_1,\ldots,v_r$ for Col A^T.
- The numbers $\sigma_1^{-1}(u_1^Ty), \ldots \sigma_r^{-1}(u_r^Ty)$ are the components of the vector A^+y in Col A^T relative to the orthonormal basis $v_1,\ldots,v_r$ for Col A^T.
- Each singular value contracts a component of the vector A^+y.

- $AA^+ = u_1u_1^T + \ldots + u_ru_r^T$.
- AA^+y is the orthogonal projection of a vector y in R^m onto Col A.

- $A^+A = v_1v_1^T + \ldots + v_rv_r^T$.
- A^+Ax is the orthogonal projection of a vector x in R^n onto Col A^T.

Lay 7.5 Principal component analysis (PCA)

- **A problem in statistics**
- The following table lists the weights and heights of five boys:

	Abe	Bob	Chad	Dan	Eric
Weight (lb)	120	125	125	135	145
Height (in)	61	60	64	68	72

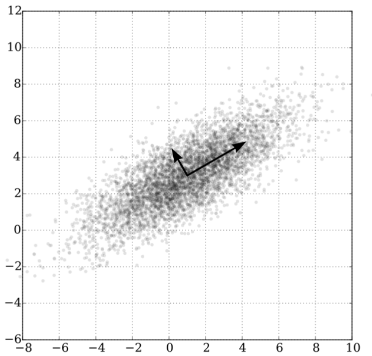
- Problem: Find a single **size index** that explains most of the variation in the data.
- This problem is solved on Lay p. 431.
- Here we use this problem to motivate the development of the principal component analysis.

- Various weights form a scalar set.
- Various heights form another scalar set.
- The Cartesian product of the two scalar sets is a two-dimensional vector space, which we call the **weight-height space**.
- The samples are five boys: Abe, Bob, Chad, Dan, and Eric.
- Each boy has a weight and a height, and gives a vector in the weight-height space.
- We next set up the problem in general terms.

- **Matrix of observation**
- The Cartesian product of the p scalar sets is a p-dimensional vector space, called the **observation space**.
- Denote the p scalars measured from a sample by $x_1,\ldots,x_p$, and list them as a column X, called a **observation vector**.
- Let N be the number of **samples**.
- Let $X_1,\ldots, X_N$ be the observation vectors measured from the N samples.
- The $p \times N$ matrix $[X_1 \ldots X_N]$ is called the **matrix of observation**.

- In the above example, p = 2 and N = 5.
- The observation space is the weight-height space
- The samples consist of the five boys.
- The matrix of observation is listed as the 2 × 5 table.

- **Scatter plot**
- When p = 2, we can plot the N observation vectors as N points in a plane, called the **observation plane**.
- Such a plot is called a **scatter plot**.



- For p = 2, the cloud of data is approximately an ellipse.
- The center of the ellipse corresponds to the sample mean.
- In one axis of the ellipse, the scatter of the data is the largest.
- In the other axis of the ellipse, the scatter of the data is the smallest.
- This behavior is understood in the principal component analysis.

- Define the **sample mean** by
- $M = (X_1 + \ldots + X_N)/N$.

- List the differences of the N observation vectors from the mean as columns of a p × N matrix:
- $B = [X_1 - M \ \ldots \ X_N - M]$.

- Define the **covariance matrix** by
- $S = BB^T/(N - 1)$.
- The covariance matrix S is a real, symmetric, p × p matrix.
- Denote the scalars in the p scalar sets by $x_1,\ldots,x_p$.
- S_{jj} is the **variance** of x_j.
- S_{ij} is the **covariance** of x_i and x_j.
- When $S_{ij} = 0$, x_i and x_j are called **uncorrelated**.

- **Principal component analysis**

- S has p nonnegative eigenvalues, $\lambda_1 \geq ... \geq \lambda_p \geq 0$, counting multiplicity.
- S has p orthonormal eigenvectors, $u_1,...,u_p$, called the **principal components** of the data.
- The **first principal component** of data, u_1, is the eigenvector corresponding to the largest eigenvalue.

- The orthonormal eigenvectors $u_1,...,u_p$ are an orthonormal basis for the p-dimensional observation space.
- $y_1 = u_1^T X$ is the component of X projected onto u_1.
- As X varies over the observation vectors $X_1,\ldots, X_N$, λ_1 is the **variance** of y_1.
- $\lambda_1 +...+ \lambda_p$ is called the total variance of the data.
- The ratio $\lambda_1/(\lambda_1 +...+ \lambda_p)$ is the fraction of the variance “explained” or “captured” by y_1.

- Let u be a unit vector in the observation space.
- Let $y = u^T X$.
- As X varies over the observation vectors $X_1,\ldots, X_N$, the variance of y is $u^T S u$.
- Of the unit vector u in the observation space, the eigenvector u_1 maximizes the variance, and the maximal variance is λ_1.

- **Example**: Lay p. 431 gives the solution to the five-sample, two-scalar problem.