Applied Linear Algebra Notes, Spring 2022

Contents

Fun Stuff **2**

Course Introduction **3**
.1 Data and Linear Algebra 3

Chapter 1: Linear Equations in Linear Algebra **3**
.1 1.1 Systems of linear equations 3
.2 1.2 Row reduction and echelon form 5
.3 1.3 Vector equations 6
.4 1.4 The matrix equation $A\vec{x} = \vec{b}$ 7
.5 1.5 Solution sets of linear equations 9
.6 1.6 Applications of linear systems 10
.7 1.7 Linear independence 10
.8 1.8 Introduction to the linear transformation 11
.9 1.9 The matrix of a linear transformation 13
.10 1.10 Linear models in business, science, and engineering 14

Chapter 2: Matrix algebra **15**
.1 2.1 Matrix operations 15
.2 2.2 The inverse of a matrix 17
.3 2.3 Characterizations of invertible matrices 20
.4 2.4 Partitioned matrices 21
.5 2.5 Matrix factorizations 22
.6 2.6 The Leontief input-output model 22
.7 2.7 Applications to computer graphics 23
.8 2.8 Subspaces of $\mathbb{R}^n$ 23
.9 2.9 Dimension and rank 24

Chapter 3: Determinants **26**
.1 3.1 Introduction to determinants 26
.2 3.2 Properties of determinants 27
.3 3.3 Cramer's rule, volume, and linear transformations 29

Chapter 4: Vector spaces **29**

Chapter 5: Eigenvalues and eigenvectors **29**
.1 5.1 Eigenvectors and eigenvalues 29
.2 5.2 The characteristic equation 31
.3 5.3 Diagonalization 33
.4 5.4 Eigenvectors and linear transformations 34
.5 5.5 Complex eigenvalues 34
.6 5.6 Discrete dynamical systems 36

.7 5.7 Applications to differential equations 36
.8 5.8 Iterative estimates to eigenvalues 38

Chapter 6: Orthogonality and least squares **38**
.1 6.1 Inner product, length, and orthogonality 39
.2 6.2 Orthogonal sets 40
.3 6.3 Orthogonal projections 42
.4 6.4 The Gram-Schmidt process 42
.5 6.5 Least-squares problems 44
.6 6.6 Applications to linear models 44
.7 6.7 Inner product spaces 44
.8 6.8 Applications of inner product spaces 44

Chapter 7: Symmetric matrices and quadratic forms **45**
.1 7.1 Diagonalization of symmetric matrices 45
.2 7.2 Quadratic forms 45
.3 7.3 Constrained optimization 45
.4 7.4 The singular value decomposition 45
.5 7.5 Applications to image processing and statistics 45

Fun Stuff

1. Feynman Method: `https://www.youtube.com/watch?v=FrNqSLPaZLc`
2. Bad math writing: `https://lionacademytutors.com/wp-content/uploads/2016/10/sat-math-section.jpg`
3. Google AI experiments: `https://experiments.withgoogle.com/ai`
4. Babylonian tablet: `https://www.maa.org/press/periodicals/convergence/the-best-known-old-babyl`
5. Parabola in real world: `https://en.wikipedia.org/wiki/Parabola#Parabolas_in_the_physical_world`
6. Parabolic death ray: `https://www.youtube.com/watch?v=TtzRAjW6KO0`
7. Parabolic solar power: `https://www.youtube.com/watch?v=LMWIgwvbrcM`
8. Robots: `https://www.youtube.com/watch?v=mT3vfSQePcs`, riding bike, kicked dog, cheetah, backflip, box hockey stick
9. Cat or dog: `https://www.datasciencecentral.com/profiles/blogs/dogs-vs-cats-image-classificat`
10. History of logarithm: `https://en.wikipedia.org/wiki/History_of_logarithms`
11. Log transformation: `https://en.wikipedia.org/wiki/Data_transformation_(statistics)`
12. Log plot and population: `https://www.google.com/publicdata/explore?ds=kf7tgg1uo9ude_&met_y=population&hl=en&dl=en#!ctype=l&strail=false&bcs=d&nselm=h&met_y=population&scale_y=lin&ind_y=false&rdim=country&idim=state:12000:06000:48000&ifdim=country&hl=en_US&dl=en&ind=false`
13. Yelp and NLP: `https://github.com/skipgram/modern-nlp-in-python/blob/master/executable/Modern_NLP_in_Python.ipynb` `https://www.yelp.com/dataset/challenge`

14. Polynomials and splines: https://www.youtube.com/watch?v=O0kyDKu8K-k, Yoda / matlab, https://www.google.com/search?q=pixar+animation+math+spline&espv=2&source=lnms&tbm=isch&sa=X&ved=0ahUKEwj474fQja7TAhUB3YMKHY8nBGYQ_AUIBigB&biw=1527&bih=873#tbm=isch&q=pixar+animatio mesh+spline, http://graphics.pixar.com/library/
15. Polynomials and pi/taylor series: Matlab/machin https://en.wikipedia.org/wiki/Chronology_of_computation_of_%CF%80 https://en.wikipedia.org/wiki/Approximations_of_%CF%80#Machin-lik formula https://en.wikipedia.org/wiki/William_Shanks
16. Deepfake: face https://www.youtube.com/watch?v=ohmajJTcpNk
dancing https://www.youtube.com/watch?v=PCBTZh41Ris
17. Pi digit calculations: https://en.wikipedia.org/wiki/Chronology_of_computation_of_%CF%80, poor shanks...https://en.wikipedia.org/wiki/William_Shanks

Course Introduction

.1 Data and Linear Algebra

1. Image pixel: LINK
2. Sports ranking: LINK
3. Word2Vec: LINK
4. Recommender system: LINK
5. Dimension reduction: LINK

Chapter 1: Linear Equations in Linear Algebra

.1 1.1 Systems of linear equations

1. Definition: A *linear equation* is of the form

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

where x_i are unknown variables with a_i known constant coefficients and b known constant. Only powers of 1 per variable. No other products or quotients.

2. Fundamental problem of linear algebra:
   - Solve a system of linear equations (rich theory can completely study).
   - Key questions: Existence and uniqueness.
3. Familiar example, new ideas.
   (a) Solve for x and y.

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

   Linear equations, graphs are lines in 2d.

   (b) Three perspectives of this class:
   - Row picture (familiar)
   - Column picture (new)
   - Matrix representation (maybe new)

(c) Row picture:

- Graph in xy-plane. Solution is intersection of two lines. How to find? Substitute or elimination.
- In general, can see thee possibilities: Unique solution (lines differ in slope), infinite solutions (2 lines overlap), no solution (2 parallel non intersecting lines). No solution is called *inconsistent*. One or infinite many solutions called *consistent*.

(d) Column picture: Vector representation

- Remind of 2D vector geometry, scalar multiplication, vector addition, graph, and linear combination.
- Rewrite in vector form. How to think of this? What linear combination of column vectors $\vec{v}_1$ and $\vec{v_2}$ result in vector $\vec{b}$? Draw in the plane and sketch solution.
- Verify that solutions $x = 1, y = 2$ from before work.
- Again, three possibilities. What are the vector analogies regarding column vectors and RHS vector?
- Generalize: If we change the RHS vector, will we always have a solution? In this case yes since $\vec{v}_1$ and $\vec{v_2}$ span $\mathbb{R}^2$. Change for parallel column vectors to see not always.

(e) Matrix representation:

- Rewrite as coefficient matrix times unknown vector equal a RHS vector.
- Notation: Note text uses bold face letters for vectors.

$$A, \quad \vec{x}, \quad \vec{b}$$

- Can also write short hand as an augmented matrix.
- Solve using the same elimination strategy as with linear equations. Think of this as a computational view. Next section covers this.
- Matrix A can be thought of as an operator on solution vector $\vec{x}$ with resulting vector $\vec{b}$. Studying this linear system equations to studying properties of matrix A.

4. Higher dimensions:

(a) 3 equations, 3 unknowns:

$$\begin{cases} x + 2y + 3z = 5 \\ 2x + 5y + 2z = 7 \\ 6x - 3y + z = -2 \end{cases}$$

Solution is $x = 0, y = 1, z = 1$.

(b) Row picture

- Ask graph of each linear equation. Graph in Geogebra 3d to see. Can anyone solve? Plot solution point as well.
- Again 3 cases here, but a bit richer. 1 solution, infinite solutions (plane or line of intersection), no solution (2 planes parallel but not the same).
- Solve by row reduction and backwards substitution. Goal is to replace system with equivalent, though simpler system. Summarize 3 elementary row operations (swap, scale, replace with row plus multiple of another). Why bother swap or scale? Take advantage of zeros and nice numbers. Computers care for high dimension to avoid roundoff error. Mention could eliminate all the way to Gauss Jordan form.

(c) Column picture: Linear combination of three vectors giving RHS vector. Use Geogebra 3d again. Again, think of three cases. Key is all three vectors are linearly independent.

(d) Matrix picture: Easy to write down? Now what?

- Can see columns of A are column vectors.
- What about row vectors? Will develop this.
- Augmented matrix. Algorithm in next section.

(e) Advantages / disadvantages of each picture: Combined they offer a complete theory.

- Row picture: Lots of info and intuition, cannot extend beyond 3d, will think in analogies.
- Column picture: Easy to extend, hard to solve, lots of info and intuition.
- Easy to adapt as algorithm, little intuition.

5. Homework: 3, 7, 13, 18, 19, 23, 25, 33, 34

.2 1.2 Row reduction and echelon form

1. 2 algorithms for solving linear systems of equations:
   - Gaussian elimination and backwards substitution (saw last time).
   - Gauss-Jordan elimination.
2. Example, 2×2: Solve the system using equation form.

$$\begin{cases} x - 2y = 1 \ (R_1) \\ 3x + 2y = 11 \ (R_2) \end{cases}$$

(a) Use the same forward reduction and back substitution idea as in last section.

$$R_2 \to -3R_1 + R_2$$

Check solution works. Recall 3 elementary row operations.

(b) Generalize: Use augmented matrix and aim towards a standard form.

- Row echelon form (GE)

$$\left[\begin{array}{cc|c} 1 & -2 & 1 \\ 3 & 2 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 8 & 8 \end{array}\right]$$

- Reduced row echelon form (G-JE)

$$\left[\begin{array}{cc|c} 1 & -2 & 1 \\ 3 & 2 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 1 \end{array}\right]$$

- Pivot entries correspond to locations of 1's in RREF. Pivot columns are columns which contain a pivot entry.
- Note, for any matrix REF is not unique but RREF is. Will prove the latter later.

(c) What if...

- No solution:

$$\left[\begin{array}{cc|c} 1 & -2 & 1 \\ 3 & -6 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 0 & 8 \end{array}\right]$$

- Infinitely many solutions:

$$\left[\begin{array}{cc|c} 1 & -2 & 1 \\ 3 & -6 & 3 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 0 & 0 \end{array}\right]$$

Here y is a free variable and all solutions are

$$\begin{cases} x = 1 + 2y \\ y \text{ free} \end{cases}$$

or written parametrically as

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

for parameter t.

(d) Careful definitions

- Row echelon form (REF) requires
  - All nonzero rows are above rows of all zeros
  - Each leading entry of each row is in a column to the right of the leading entry of the row above it
  - All entries in column below a leading entry are zeros
- Reduced row echelon form (RREF) requires REF and also
  - The leading entry in each nonzero row is 1
  - Each leading 1 is the only nonzero entry in its column

(e) Pivot position in matrix A is the location in A that corresponds to a leading 1 in the RREF of A. A pivot column contains a pivot entry.

(f) Example: Find REF and RREF of random 3×2 matrix.

3. Example: Higher dimension, try on own:

$$\begin{cases} 2x + 4y - 2z = 2 \\ 4x + 9y - 3z = 8 \\ -2x - 3y + 7z = 10 \end{cases}$$

REF and backwards sub vs RREF.

4. Example: General solution in high dim. Find the general solution to the corresponding linear system.

$$\left[\begin{array}{cccc|c} 1 & -7 & 0 & 6 & 5 \\ 0 & 0 & 1 & -2 & -3 \\ -1 & 7 & -4 & 2 & 7 \end{array}\right]$$

Consistent vs inconsistent terminology.

5. Homework: 1, 3, 5, 7, 11, 13, 15, 17, 21, 23, 33-34

.3 1.3 Vector equations

1. 3 view of linear algebra:
   - Equation (row picture)
   - Matrix
   - Vector (column picture): This section, this is where we get geometric reasoning with math rigor.
2. Definition: The vector space $\mathbb{R}^n$ consists of all column vectors $\vec{u}$ with n real valued components.
   - Notation: $\vec{u} = [u_1, u_2, \ldots, u_n]^T$, each entry is called a component.
   - Special case: $\vec{0}$.
3. Examples: Geometry of vectors, imagine displacement.
   - $\vec{u} = [1, 2]^T \in \mathbb{R}^2$. Note not the same as (1,2). Vectors are location independent. Other examples in 4 quadrants. Sad zero vector.

- $\vec{u} = [-3, 1, 2]^T \in \mathbb{R}^3$

4. Definitions: Vector operations
   - Addition: $\vec{u} + \vec{v} = [u_1 + v_1, \dots, u_n + v_n]^T$ in $\mathbb{R}^n$. Note need vectors of same length.
   - Scalar multiplication: $c\vec{u} = [cu_1, \dots, cu_n]^T$ for scalar c.
   - Subtraction (triangular law): $\vec{u} - \vec{v}$
   - Bonus (dot product to compare direction, more later): $\vec{u} \cdot \vec{v}$
   - Bonus (norm or length, more later): $\|\vec{u}\|_n = \sqrt{u_1^2 + \cdots + u_n^2}$
5. Examples: $\vec{u} = [1, 2]^T, \vec{v} = [3, 1]^T$
   - $2\vec{u}, -\vec{u}, 4\vec{u}, 0\vec{u}, c\vec{u}$, set of all scalar multiples results in a line (rescaling gives name to scalar)
   - $\vec{u} + \vec{v}, \vec{v} + \vec{u}$ (Parallelogram law)
   - $\vec{u} - \vec{v} = \vec{u} + (-\vec{v})$ (Triangular law)
6. Theorem (these mirror familiar algebraic properties, some proofs in HW): For all $\vec{u}, \vec{v} \in \mathbb{R}^n$ and scalars
   (a) $\vec{u} + \vec{v} = \vec{v} + \vec{u}$ (Commutative)
   (b) $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$ (Associative)
   (c) $\vec{u} + \vec{0} = \vec{u}$ (Identity)
   (d) $\vec{u} + (-\vec{u}) = \vec{0}$ for $-\vec{u} = (-1)\vec{u}$ (Inverse)
   (e) $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$ (Distribution)
   (f) $(c + d)\vec{u} = c\vec{u} + d\vec{u}$ (Distribution)
   (g) $c(d\vec{u}) = (cd)\vec{u}$ (Compatibility)
   (h) $1\vec{u} = \vec{u}$ (Identity)
7. Definition (the linear of linear algebra): Vector $\vec{y} \in \mathbb{R}^n$ is a linear combination of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ if there exists scalars $c_1, \dots, c_n$ (called weights) such that
$$\vec{y} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$$
8. Example (Vector equation): Show that $\vec{b} = [3, 1, -1]^T$ is a linear combination of vectors $\vec{a_1} = [2, 0, -1]^T$ and $\vec{a_2} = [-1, 1, 1]^T$.
   - This is equivalent to solving a linear system via GE.
   - Geogebra and geometric interpretation.
   - Is the same true for any $\vec{b}$? No, only if it lies in the plane generated by all linear combinations of $\vec{a_1}$ and $\vec{a_2}$. Consider a $\vec{b}$ which does not.
9. Definition: The collection of all linear combinations of $\vec{v_1}, \dots, \vec{v_p} \in \mathbb{R}^n$ is called the Span$\{\vec{v_1}, \dots, \vec{v_p}\}$ and is a subset of $\mathbb{R}^n$.
10. Homework: 1, 3, 5, 7, 9, 11, 13, 15, 17, 21, 23, 27

.4 1.4 The matrix equation $A\vec{x} = \vec{b}$

1. 3 views of linear algebra:
   - Row picture (lines and planes, done)
   - Column picture (vectors, done)
   - Matrix picture (now, idea is to capture linear combination as an operation)

2. Definition: For A a $m \times n$ matrix with columns $\vec{a_1}, \ldots, \vec{a_n}$ and $\vec{x} \in \mathbb{R}^n$, the produce $A\vec{x}$ is the linear combination of the columns of A with weights as entries in $\vec{x}$. That is,

$$A\vec{x} = [\vec{a_1} \ldots \vec{a_n}] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{a_1} + \ldots x_n\vec{a_n}$$

Note, the number of columns in A must match the number of entries of $\vec{x}$.

3. Example: Multiply a random $A_{2\times3}$ matrix by a $\vec{x}_{3\times1}$ vector.

   3 linear algebra POVs are here. For general $\vec{x}$, write

   - 2 equations (planes, geometry)
   - Linear combinations of 3 vectors (vectors, geometry)
   - Matrix equation $A\vec{x} = \vec{b}$ (operation on a vector, similar to idea of function). Important question is given A, can we solve $A\vec{x} = \vec{b}$ for any RHS vector $\vec{b}$.

   We will readily switch between these views to gain insight and perspective.

4. Example (entry-wise matrix multiplication): Multiply a random $A_{3\times3}$ matrix by a $\vec{x}_{3\times1}$ vector.

   - Linear combination of 3 row vectors. Important concept.
   - Dot product of rows and $\vec{x}$. This version is more convenient for hand calculation.

   Replace A with identity matrix $I_{3\times3}$ and ask them to guess result.

5. Theorem (linearity of matrix multiplication): For matrix A $m \times n$, vectors $\vec{u}, \vec{v}$ $n \times 1$, and scalar c, we have

   (a) $A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v}$ (distributive)

   (b) $A(c\vec{u}) = c(A\vec{u})$ (associative)

   Proof (of (a), $n = 3$ case, (b) in text): All we need is the corresponding result from vectors in previous section.

$$\begin{aligned} A(\vec{u} + \vec{v}) &= A \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix} \\ &= (u_1 + v_1)\vec{a_1} + (u_2 + v_2)\vec{a_2} + (u_3 + v_3)\vec{a_3} \\ &= (u_1\vec{a_1} + u_2\vec{a_2} + u_3\vec{a_3}) + (v_1\vec{a_1} + v_2\vec{a_2} + v_3\vec{a_3}) \\ &= A\vec{u} + A\vec{v} \end{aligned}$$

6. Theorem (big result for entire course, will grow this list): For A a $m \times n$ matrix, the following statements are either all true or all false.

   (a) For each $\vec{b} \in \mathbb{R}^n$, equation $A\vec{x} = \vec{b}$ has a solution.

   (b) Each $\vec{b} \in \mathbb{R}^n$ is a linear combination of the columns of A.

   (c) The columns of A span $\mathbb{R}^m$.

   (d) A has a pivot position in every row.

7. Homework: 5, 7, 9, 11, 13, 15, 17, 23, 29, 30

.5 1.5 Solution sets of linear equations

1. We want to characterize solutions to a linear system of equations $A\vec{x} = \vec{b}$ for A and $\vec{b}$ given and $\vec{x}$ unknown thru two perspectives:

   - Geometrically (picture, intuition)
   - Explicitly (formula, practical)

   Our approach will be to consider two related cases:

   - Homogeneous linear system: $A\vec{x} = \vec{0}$
   - Nonhomogeneous linear system: $A\vec{x} = \vec{b}$

2. Homogeneous linear system: $A\vec{x} = \vec{0}$

   (a) For any A, $\vec{x} = \vec{0}$ is always a solution (called the trivial solution). We seek nontrivial solutions $\vec{x} \neq \vec{0}$. Will there always be a nontrivial solution? Only if the GE solution has at least one free variable.

   (b) Solve the homogeneous linear system:

   $$\begin{bmatrix} 1 & 3 & -5 \\ 1 & 4 & -8 \\ -3 & -7 & 9 \end{bmatrix} \vec{x} = \vec{0}$$

   Solving by GE gives x_3 a free variable with

   $$\vec{x} = x_3 \begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix} = x_3\vec{v} = span\{\vec{v}\}$$

   The set of these solutions are a line thru the origin parallel to $\vec{v}$.

   (c) Change above example so three rows are multiples of eachother giving 2 free variables.

   $$\begin{bmatrix} 1 & 3 & -5 \\ 1 & 3 & -5 \\ 1 & 3 & -5 \end{bmatrix} \vec{x} = \vec{0}$$

   Solving by GE gives x_2, x_3 free variables with

   $$\vec{x} = \begin{bmatrix} -3x_2 + 5x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2\vec{v_2} + x_3\vec{v_3} = span\{\vec{v_2}, \vec{v_3}\}$$

   generating a plane thru the origin. View in Geogebra.

3. Nonhomogenous linear system: $A\vec{x} = \vec{b}$

   (a) Example as from before:

   $$\begin{bmatrix} 1 & 3 & -5 \\ 1 & 4 & -8 \\ -3 & -7 & 9 \end{bmatrix} \vec{x} = \begin{bmatrix} 4 \\ 7 \\ 6 \end{bmatrix}$$

   gives

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

Again x_3 is free and we have

$$\vec{x} = \begin{bmatrix} -4x_3 - 5 \\ 3x_3 + 3 \\ x_3 \end{bmatrix} = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix} = \vec{p} + x_3\vec{v}$$

for the same $\vec{v}$ as in the homogenous case. Graph same lines as before but first shifted by vector $\vec{p}$ away from the origin.

(b) Solution to nonhomogenous equation is the same as the homogenous case but translated.

(c) Theorem: For $A\vec{x} = \vec{b}$ consistent and $\vec{p}$ a particular solution, then the solution set of all $A\vec{x} = \vec{b}$ is all vectors of the form

$$w = \vec{p} + \vec{v_h}$$

where $\vec{v_h}$ is any solution to the homogeneous equation $A\vec{x} = \vec{0}$. (sketch the plane case in $\mathbb{R}^3$)

4. Homework: 1, 5, 7, 9, 11, 13, 17, 19, 21, 23, 27, 29, 31

.6 1.6 Applications of linear systems

1. Skip. Possible lab material.

.7 1.7 Linear independence

1. Here we rephrase homogeneous systems of linear equations as vector equations instead. So our example homogeneous linear system

$$\begin{bmatrix} 1 & 3 & -5 \\ 1 & 4 & -8 \\ -3 & -7 & 9 \end{bmatrix} \vec{x} = \vec{0}$$

is equivalent to

$$x_1 \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ -7 \end{bmatrix} + x_3 \begin{bmatrix} -5 \\ -8 \\ 9 \end{bmatrix} = \vec{0}$$

which brings us to an important definition for this course.

2. Definition: The set of vectors $\{\vec{v_1}, \ldots, \vec{v_p}\}$ in $\mathbb{R}^n$ is linearly independent if the vector equation

$$x_1\vec{v_1} + \cdots + x_p\vec{v_p} = \vec{0}$$

has only the trivial solution. If there are weights $x_1, \ldots, x_p$ not all zero such that

$$x_1\vec{v_1} + \cdots + x_p\vec{v_p} = \vec{0}$$

then $\{\vec{v_1}, \ldots, \vec{v_p}\}$ is linearly dependent.

3. Example: Previous work on

$$\begin{bmatrix} 1 & 3 & -5 \\ 1 & 4 & -8 \\ -3 & -7 & 9 \end{bmatrix} \vec{x} = \vec{0}$$

gave solution set

$$\vec{x} = x_3 \begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix} = x_3\vec{v} = span\{\vec{v}\}$$

meaning that there are infinitely many solutions. Choosing $x_3 = 1$ gives $\vec{x} \neq 0$ so that

$$-4\vec{v_1} + 3\vec{v_2} + \vec{v_3} = \vec{0}$$

and so these three column vectors are linearly dependent. Alternatively,

$$\vec{v_3} = 4\vec{v_1} - 3\vec{v_2}$$

and there is redundant information in these columns. This points towards the following results.

4. Theorem: The columns of matrix A are linearly independent if and only if the equation $A\vec{x} = \vec{0}$ has only the trivial solution.

5. Theorem: The set of vectors $\{\vec{v_1}, \ldots, \vec{v_p}\}$ is linearly dependent if one vector can be written as a linear combination of the others.

6. Intuition of linear dependence / independence:

   (a) One vector: Is the set of one vector linearly independent or dependent? Only if that vector is not the zero vector.
   $$\vec{v_1} = [1, 2]^T$$

   (b) Two vectors, $n = 2$: When are two vectors linearly dependent? If one is a scalar multiple of the other.
   $$\vec{v_1} = [1, 2]^T, \vec{v_2} = [5, 10]^T, \text{ on the same line, same direction of information}$$
   $$\vec{v_1} = [1, 2]^T, \vec{v_2} = [1, 10]^T, \text{ not on the same line, separate direction of information}$$

   (c) Three vectors, $n = 2$: When are three vectors linearly dependent? Always. GE always yields a free variable. Graph example to show one vector as a linear combination of the other. Redundant information. This generalizes to the following result.

7. Theorem: The set $\{\vec{v_1}, \ldots, \vec{v_p}\}$ in $\mathbb{R}^n$ with $p > n$ is linearly dependent.

8. Note: With this section especially, we start to see the wide range of terminology in this course, much of it is a different perspective on the same root concept. Keeping this all straight is essential to avoid confusion.

9. Homework: 1, 3, 5, 7, 9, 15, 17, 21, 23, 25, 27, 31

.8 1.8 Introduction to the linear transformation

1. New perspective: Think of $A\vec{x} = \vec{b}$ as a matrix operation.

   (a) Similar to $f(x) = y$, function f acting on x to result in y.

   (b) Matrix A acts on vector $\vec{x}$ resulting in vector $\vec{y}$.

2. Def and terminology: A random 2×3 matrix times $\vec{x}$ giving $\vec{b}$.

   (a) Picture: Mapping of inputs to outputs

   (b) Inputs (domain) any vector in $\mathbb{R}^3$

   (c) Outputs (range) some vectors in $\mathbb{R}^2$ (codomain)

   (d) Linear transformation A mapping inputs to outputs

   (e) Notation: Matrix transformation $T(\vec{x}) = A\vec{x} = \vec{b}$ where $\vec{b}$ is the image of $\vec{x}$

   (f) Just as we try to understand a function for any input, we will try to understand a matrix transformation in general.

3. Example: Same 2×3 matrix as above. Define $T(\vec{x}) = A\vec{x}$.

   (a) Find the image of random vector $\vec{x}$.

(b) For random vector $\vec{b}$, find input $\vec{x}$ if possible. Is it unique? If no, transformation is not invertible (reversible) as with function inverses.

4. Linear transformations: Defined and alternate forms.

   (a) Def: A transformation $T(\vec{x})$ is linear if

   $$T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v}), \quad \text{and} \quad T(c\vec{u}) = cT(\vec{u})$$

   for all vectors $\vec{u}, \vec{v}$ in the domain of T and all scalars c.

   (b) We have from before that all matrix transformations are linear transformations, but there are other linear transformations to be seen later on.

   (c) Theorem: If $T(\vec{x})$ is a linear transformation, then

   $$T(c\vec{u}+d\vec{v}) = cT(\vec{u}) + dT(\vec{v}), \quad \text{and} \quad T(\vec{0}) = T(\vec{0})$$

   for all vectors $\vec{u}, \vec{v}$ in the domain of T and all scalars c, d.

   (d) Theorem: The superposition principle holds for any linear transformation $T(\vec{x})$. That is,

   $$T(c_1\vec{u_1} + \cdots + c_p\vec{u_p}) = c_1T(\vec{u_1} + \cdots + T(c_p\vec{u_p})$$

   (e) These two theorems are often more convenient.

5. Examples: Geometry of linear transformations. For vectors $\vec{u} = [3,1]^T$, $\vec{v} = [1,2]^T$ and $\vec{u}+\vec{v}$, what does transformation $T(\vec{x}) = A\vec{x}$ do? Use linearity for $\vec{u}+\vec{v}$. Draw the parallelogram to see effect.

   - Dilation

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

   - Contraction

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

   - Reflection

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

   - Shear

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

   - 90 degree rotation

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

   - Projection

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

6. Homework: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 29, 31

.9 1.9 The matrix of a linear transformation

1. In the last section, we looked at a matrix transformation and saw geometry. Here we reverse. Given a geometric description, we will derive the needed linear transformation.

2. Unit basis in $\mathbb{R}^2$:

   - $\vec{e_1} = [1, 0]^T, \vec{e_2} = [0, 1]^T$, all other vectors in $\mathbb{R}^2$ are linear combinations of these two. Show example.
   - Amounts to geometric transformation of the unit square.
   - Using linearity, understanding $T(\vec{x}) = A\vec{x}$ action on these two unit basis will determine A. This is because for any $\vec{x} \in \mathbb{R}^2$,

   $$\vec{x} = x_1\vec{e_1} + x_2\vec{e_2}$$

3. Example: Find linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^4$ such that

   $$T(\vec{e}_1) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad \text{and} \quad T(\vec{e}_2) = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

   Since we have for any $\vec{x}$ that

   $$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1\vec{e}_1 + x_2\vec{e}_2,$$

   then

   $$T(\vec{x}) = T(x_1\vec{e}_1 + x_2\vec{e}_2,) = x_1T(\vec{e}_1) + x_2T(\vec{e}_2) = [T(\vec{e}_1) + T(\vec{e}_2)]\,\vec{x} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 3 & 0 \\ 4 & 0 \end{bmatrix} \vec{x}$$

   This holds for higher dimensional space as well.

4. Theorem: For linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, there exists a unique matrix A such that

   $$T(\vec{x}) = A\vec{x} = [T(\vec{e_1}) \cdots T(\vec{e_n})]\,\vec{x}$$

   for unit basis vectors $\vec{e_1}, \ldots, \vec{e_n}$.

   Matrix A is called the standard matrix for the linear transformation T. Also see that any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is also a matrix transformation.

5. Example: Use the above theorem to find the linear transformation which

   - Projects $\vec{x}$ onto the main diagonal.
   - Rotates $\vec{x}$ 180 degrees about the origin.

6. Example: Use the above theorem to find the linear transformation $T(\vec{x})$ which rotates vector $\vec{x}$ by θ radians counter clockwise.

   - Draw $\vec{e_1}$ and $\vec{e_2}$ in the plane and resulting rotated vectors.
   - Use trig to find resulting vectors:

   $$T(\vec{e_1}) = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}, \quad T(\vec{e_2}) = \begin{bmatrix} \cos(\theta + \pi/2) \\ \sin(\theta + \pi/2) \end{bmatrix} = \begin{bmatrix} \cos(\pi/2 - (-\theta)) \\ \sin(\pi/2 - (-\theta)) \end{bmatrix} = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix}$$

   - Theorem result says

   $$T(\vec{x}) = A\vec{x} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

7. Catalog of geometric transformations:
   - Thinking of what a transformation does to unit basis vectors $\vec{e_1}$ and $\vec{e_2}$ is equivalent to picturing its action on the unit square.
   - See text for list of common transformations.
   - Know these, do not memorize. Just think about what happens to $\vec{e_1}$ and $\vec{e_2}$

8. $A\vec{x} = \vec{b}$, existence and uniqueness rephrased in terms of linear transformations.
   (a) Definition: A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is onto $\mathbb{R}^m$ if each $\vec{b}$ in $\mathbb{R}^m$ is the image of at least one (though maybe more) $\vec{x}$ in $\mathbb{R}^n$. This is existence. Draw picture to illustrate.
   (b) Definition: A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one $\mathbb{R}^m$ if each $\vec{b}$ in $\mathbb{R}^m$ is the image of at most one (though maybe none) $\vec{x}$ in $\mathbb{R}^n$. This is uniqueness. Draw picture to illustrate.
   (c) Return to textbook basic linear transformations. Which are onto? One-to-one? Both? Neither?
   (d) Example: Random 3×4 matrix A in REF. Is A onto? Yes, full set of pivots. One-to-one? No, free variable. So we can answer these questions via row reduction, but there is an easier way.
   (e) Theorem: Linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one if and only if the equation $T(\vec{x}) = \vec{0}$ has only the trivial solution.
      - If and only if means if statement P is true, then statement Q is also true. Further if Q is true, then P is also true.
      - Here we prove this theorems in two steps. (1) Assume P is true, show Q is also true. (2) Assume P is false, then show Q also false (contrapositive of reverse direction).
      - Proof of (1): Assume T is one-to-one. Then $T(\vec{x} = \vec{0}$ has only one solution. We know matrix transformations are such that $T(\vec{0}) = \vec{0}$. Then $\vec{x} = \vec{0}$.
      - Proof of (2): Assume T is not one-to-one. Then for some $\vec{b}$ in $\mathbb{R}^m$ there are two vectors $\vec{u} \neq \vec{v}$ such that map to $\vec{b}$. But since T is linear
      $$T(\vec{u} - \vec{v}) = T(\vec{u}) - T(\vec{v}) = \vec{b} - \vec{b} = \vec{0}$$
      and hence $T(\vec{x}) = \vec{0}$ has a nontrivial solution.
   (f) Theorem: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. Then,
      - T is onto if and only if the columns of A span $\mathbb{R}^m$.
      - T is one-to-one if and only if the columns of A are linearly independent.

      Why does this theorem make intuitive sense?
   (g) Show T is a one-to-one linear transformation. Is T onto?
   $$T(\vec{x}) = \begin{bmatrix} x_1 - x_2 \\ -2x_1 + x_2 \\ x_1 \end{bmatrix}$$
   Show a matrix transformation with linearly indep columns, hence a one-to-one linear transformation. Two columns cannot span $\mathbb{R}^3$, so not onto.

9. Homework: 1, 3, 5, 7, 13, 15, 17, 23, 25, 27, 29, 30

.10 1.10 Linear models in business, science, and engineering

1. Possible lab material. Especially difference equations.

Chapter 2: Matrix algebra

.1 2.1 Matrix operations

1. Goal of this chapter: Treating A as an operator, we get a new view on $A\vec{x} = \vec{b}$.
   - Similar to $\frac{d}{dx}$ as an operator on $f(x)$
   - What are the properties of operator A?
   - How to reverse this operation (will call inverse)?
2. Basic matrix operations (easy): Arithmetic (addition and scalar multiplication)
   (a) Random 2×3 matrices A and B.
   (b) $2A$, entry-wise scalar multiplication
   (c) $A + B$, as with vectors, need dimensions to agree, entry-wise addition (and subtraction)
   (d) Theorem: For A, B, C matrices of the same dimension and scalars r, s,
   - $A + B = B + A$ (commutative)
   - $(A + B) + C = A + (B + C)$ (associative for addition)
   - $A + 0 = A$ (identity for addition)
   - $r(A + B) = rA + rB$ (scalar distribution)
   - $(r + s)A = rA + sA$ (matrix distribution)
   - $r(sA) = (rs)A$ (associative for mult)
3. Matrix multiplication:
   (a) Recall: $B\vec{x}$ as a linear combination of the column vectors of $n \times p$ matrix B

   $$B\vec{x} = x_1\vec{b}_1 + \cdots + x_p\vec{b}_p$$

   (b) Matrix composition: $A(B\vec{x})$ for A $m \times n$ and B $n \times p$.
   - Draw diagram: $\vec{x} \to B\vec{x} \to A(B\vec{x})$
   - One step arc on diagram: Think of AB as the new matrix operation for which $\vec{x} \to (AB)\vec{x}$.
   - Similar to function composition: $f(g(x)) = (f \circ g)(x)$
   - How to compute?

   $$B\vec{x} = x_1\vec{b}_1 + \cdots + x_p\vec{b}_p$$
   $$A(B\vec{x}) = A(x_1\vec{b}_1 + \cdots + x_p\vec{b}_p) = x_1A\vec{b}_1 + \cdots + x_pA\vec{b}_p = [A\vec{b}_1 \ldots A\vec{b}_p]\vec{x}$$

   - What is the dimension of AB? $m \times p$
   - Definition: For A $m \times n$ and B $n \times p$, then

   $$AB = A[\vec{b}_1 \ldots \vec{b}_p] = [A\vec{b}_1 \ldots A\vec{b}_p]$$

   where matrix AB is $m \times p$.

   (c) Example: Random matrices A (2×3) and B (3×2).
   - AB column view (can get just a column this way):

   $$AB = A[\vec{b}_1\vec{b}_2] = [A\vec{b}_1 A\vec{b}_2]$$

   - AB computational view (can get just an entry this way): Each row as row dot column
   - AB row view (can get just a row this way):

   $$AB = \begin{bmatrix} row_1(A) \\ row_2(A) \end{bmatrix} B = \begin{bmatrix} row_1(A)B \\ row_2(A)B \end{bmatrix}$$

   where this last step is done entry-wise.

- Show $AB \neq BA$. Makes sense thinking of function composition.

4. Matrix multiplication in general:

   (a) Summary of matrix multiplication: For A $(m \times n)$, B $(n \times m)$, and $C = AB$ $(m \times p)$,

   - Column-wise in general
   $$C = AB = [A\vec{b}_1 \dots A\vec{b}_p]$$
   - Computational in general
   $$C = [c_{ij}], \quad c_{ij} = row_i(A) \cdot \vec{b}_j = \sum_{k=1}^{n} a_{ik} b_{kj}$$
   - Row-wise in general
   $$C = AB = \begin{bmatrix} row_1(A)B \\ \vdots \\ row_m(A)B \end{bmatrix}$$

   (b) Theorem (matrix multiplication properties): For A, B, C matrices of suitable dimension

   - $A(BC) = (AB)C$ (associative)
   - $A(B + C) = AB + AC$ (right distributive)
   - $(B + C)A = BA + CA$ (left distributive)
   - $r(AB) = (rA)B = A(rB)$ (scalar commutative)
   - $IA = A = AI$ (identity matrix multiplication, explain what I is)

   Proofs in homework and book. These follow from vector properties shown previously.

   (c) Warning: Matrix multiplication does not follow the intuition of scalar multiplication. In general

   - $AB \neq BA$, not surprising since linear combos of cols of A need not equal linear combinations of cols of B.
   - $AB = AC$ need not imply $B = C$.
   - $AB = 0$ need not imply $A = 0$ or $B = 0$ for 0 the zero matrix.
   - Construct you own examples for fun.

5. Powers of a matrix A

   (a) Def: $A^k = A \cdot A \cdot \dots \cdot A$, repeated multiplication k times

   (b) Note, need a square matrix A $(n \times n)$.

   (c) Think if repeating an operation over and over. Similar to repeat function composition.

   (d) Will revisit this notion for important applications later.

6. Matrix transpose:

   (a) Def: For $(m \times n)$ matrix A, the transpose of A written A^T is the $(n \times m)$ matrix whose columns are the rows of A

   - Example: Random (2×3) matrix.
   - Draw general picture of row and column vectors switching
   - Entry-wise: $A_{m \times n} = [a_{ij}]$ gives $A^T_{n \times m} = [a_{ji}]$

   (b) Theorem: Properties of matrix transpose. For matrices A and B of suitable dimensions and scalar r,

   - $(A^T)^T = A$
   - $(A + B)^T = A^T + B^T$
   - $(rA)^T = rA^T$

- $(AB)^T = A^T B^T$ (only surprising result, shown in HW)

(c) Example: Random matrices and vectors $A_{3\times2}, B_{2\times2}, \vec{b}_3, \vec{c}_2$, find all possible products which are defined.

7. Homework: 1, 3, 5, 10, 11, 12, 15, 17, 19, 21, 23, 27, 33

.2 2.2 The inverse of a matrix

1. Reversing $A\vec{x} = \vec{b}$.

   (a) Draw picture: Thinking of $A\vec{x} = \vec{b}$ as an operation $A : \vec{x} \to \vec{b}$, how to invert this process? Same idea as inverting a function. We need the operation to be one-to-one.

   (b) Definition: Square matrix $A_{n\times n}$ is invertible if there exists matrix $A^{-1}_{n\times n}$ such that

   $$A \cdot A^{-1} = A^{-1} \cdot A = I$$

   for $I_{n\times n}$ the identity matrix. Note this only makes sense for square matrices.

   (c) Connection: Think of as composition of linear operators.

   $$\vec{x} = I\vec{x} = (A^{-1}A)\vec{x} = A^{-1}(A\vec{x})$$

   Draw picture. Similar to function inverses and composition, $(f \circ f^{-1})(x) = x$.

   (d) Not all matrices A are invertible. If invertible, called non singular. If not invertible, called singular (alone and without a counterpart). Singular terminology may also refer to unusual. In face most square matrices randomly generated are invertible (non-singular), for (2×2) case, need both columns to be colinear which is less common than not. Singular may also reference troublesome. Last reason may referr to the determinant being zero resulting in zero division (singularity).

   (e) Example: Show that

   $$A = \begin{bmatrix} 3 & 2 \\ 7 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 1 \\ 7/2 & -3/2 \end{bmatrix}$$

   are inverses of eachother. Just need to check that $AB = BA = I$ to show $B = A^{-1}$.

2. Finding matrix inverses:

   (a) For A a given matrix,

   - How to check if A is invertible? For functions can check if $f(x)$ is one-to-one.
   - How to compute A^{-1}? Method for functions as well, key is $(f^{-1} \circ f)(x) = x$, the inverse relation.

   (b) General 2×2 case:

   $$AB = A[\vec{b}_1\vec{b}_2] = [A\vec{b}_1 A\vec{b}_2] = I$$

   requires

   $$A\vec{b_1} = \vec{e_1}, \quad A\vec{b_2} = \vec{e_2}.$$

   These are two linear systems to solve. Likewise 3 linear systems for (3×3), and so on.

   (c) Example: Find the inverse of

   $$A = \begin{bmatrix} 3 & 2 \\ 7 & 4 \end{bmatrix}.$$

   Previous example lets us know what to expect here.

   - Solve two systems as separate augmented matrices.

     $$A\vec{b_1} = \vec{e_1}, \quad A\vec{b_2} = \vec{e_2}$$

     by using backwards substitution.

- Note redundancy and combine into a single augmented matrix

$$[A|I] \rightarrow [I|B] = [I|A^{-1}]$$

then use full Gauss-Jordan elimination.
- Elementary row operations are a key ingredient here. More shortly.
- Note: This approach of using Gaussian elimination extends to 3 or higher dimensions as well.

(d) Theorem: Can complete the (2×2) case in general. For any matrix

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

A is invertible if $ad - bc \neq 0$ and

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

If $ad - bc = 0$ then A is not invertible. In the (2×2) case, $ad - bc$ is called the determinant of A (note zero division singularity). Derive and verify on own.

(e) Validate for previous example.

(f) Above theorem generalizes to higher dimensions to a certain extent. Namely the idea of determinant generalizes via recursion. More later.

3. Using inverses to solve linear systems $A\vec{x} = \vec{b}$.

(a) Theorem: If $A_{n\times n}$ is invertible, then for each $\vec{b} \in \mathbb{R}^n$, $A\vec{x} = \vec{b}$ has a unique solution

$$\vec{x} = A^{-1}\vec{b}.$$

This isn't a practical method to solve (see previous example work), but it is important in reach (general, existence, uniquiness).

Proof: Two steps:
- Existence: Check that $\vec{x} = A^{-1}\vec{b}$ works. Key is inverse relation $AA^{-1} = I$.
- Uniquiness: If $\vec{x}$ and $\vec{y}$ are two solutions, then $A\vec{x} = \vec{b}$ and $A\vec{y} = \vec{b}$. Then we have $A\vec{x} = A\vec{y}$ and so $A^{-1}A\vec{x} = A^{-1}A\vec{y}$ implying $\vec{x} = \vec{y}$.

(b) Example: Solving a linear system via inverse.

$$\begin{cases} 3x_1 + 2x_2 = 3 \\ 7x_1 + 4x_2 = 2 \end{cases}$$

Use the above calculation where $\vec{x} = A^{-1}\vec{b} = [-4, 11]^T$. Note the Gaussian elimination work as before was packaged into the inverse function calculation.

4. Properties of inverses

(a) Theorem: For invertible matrices A and B of the same dimension,

i. $(A^{-1})^{-1} = A$ (makes sense with respect to reversing an operator)

ii. $(AB)^{-1} = B^{-1}A^{-1}$ (note the reverse of multiplication order, this is the reverse of operator composition)

iii. $(A^T)^{-1} = (A^{-1})^T$ (note inverse of a symmetric matrix also symmetric)

(b) Proofs of each, just need to check each works. Multiply to the identify.

i. Need matrix C such that

$$A^{-1}C = I, \quad CA^{-1} = I.$$

By definition $C = A$ is what we have.

ii. Compute $(AB)(B^{-1}A^{-1}) = \cdots = I$ and $(B^{-1}A^{-1})(AB) = \cdots = I$

iii. This one relies on the reversing of multiplication for transpose.

$$(A^T)(A^{-1})^T = (A^{-1}A)^T = I^T = I$$

$$(A^{-1})^T(A^T) = (AA^{-1})^T = I^T = I$$

5. Elementary matrices and decomposing Gaussian elimination

(a) Example: Think of three elementary row operations on matrix

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$$

i. $R_1 \leftrightarrow R_3$: We seek matrix E_1 such that

$$E_1 A = E_1 \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} = \begin{bmatrix} 4 & -3 & 8 \\ 1 & 0 & 3 \\ 0 & 1 & 2 \end{bmatrix}$$

Thinking bout the row picture for matrix multiplication,

$$\begin{bmatrix} row_1(E_1) \\ row_2(E_1) \\ row_3(E_1) \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} \begin{bmatrix} 4 & -3 & 8 \\ 1 & 0 & 3 \\ 0 & 1 & 2 \end{bmatrix}$$

So doing the same elem row operation on the identity matrix is the multiplier we need to swap rows 1 and 3.

ii. $R_1 \to 2R_1$. Thinking of the same row picture,

$$E_2 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

iii. $R_3 \to R_3 + 2R_2$

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$$

(b) Elementary matrices:

- Definition: An elementary matrix is the matrix resulting from performing a single elementary row operation on the identity matrix I.
- So each elementary row operation can be performed as multiplication of an elementary matrix.
- Turns out all elementary matrices are invertible. The inverse can be found by construction (reversing the elementary row operation) and validating the inverse relation. Illustrate for above 3 examples.

$$E_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad E_1^{-1} = ?, \quad E_2 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2^{-1} = ?, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \quad E_3^{-1} = ?.$$

(c) Example: Use Gaussian elimination to find the inverse of

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}.$$

Perform Gauss-Jordan elimination on

$$[A \mid I] \quad \to \quad \dots \quad \to \quad [I \mid A^{-1}]$$

resulting in

$$A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}.$$

Easy to check this is correct: $AA^{-1} = I$.

(d) Thinking of elementary matricies, we must have

$$A^{-1} = E_5 E_4 E_3 E_2 E_1$$

and so

$$A = E_1^{-1} E_2^{-1} E_3^{-1} E_4^{-1} E_5^{-1}.$$

Easy to check. This leads to a general result.

(e) Theorem: Square matrix A is invertible if and only if A is row equivalent to the identity matrix I.

6. Homework: 1, 5, 7, 9, 21, 25, 27, 29, 31, 35

.3 2.3 Characterizations of invertible matrices

1. Theorem: (Invertible matrix theorem)

   For A a square $n \times n$ matrix, the following statements are equivalent (either all true or all false).

   - A is an invertible matrix
   - A is row equivalent to the $n \times n$ identity matrix
   - A has n pivots positions
   - The equation $A\vec{x} = \vec{0}$ has only the trivial solution
   - The columns of A for a linearly independent set
   - The linear transformation $T(\vec{x}) = A\vec{x}$ is one-to-one
   - The equation $A\vec{x} = \vec{b}$ has at least one solution for each $\vec{b}$ in $\mathbb{R}^n$
   - The columns of A span $\mathbb{R}^n$
   - The linear transformation $T(\vec{x}) = A\vec{x}$ is onto
   - There is a $n \times n$ matrix C such that $CA = I$
   - There is a $n \times n$ matrix D such that $AD = I$
   - A^T is an invertible matrix

2. Example: Problems 1-6 in the exercises. Decide if invertible or not.

   (a) Yes, LI columns

   (b) No, LD columns

   (c) Yes, 3 pivots after row reduction $(5, -7, -1)$. Don't need to do the row reduction here.

(d) No, LD columns since zero vector included.
(e) No after swapping rows 1 and 2 and doing row reduction, only 2 pivots
(f) Do row reduction to see.
(g) Note, 8 easy to see

3. Inverse of linear transformations:

   (a) Def: A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if there exists a transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ such that
   $$S(T(\vec{x})) = \vec{x}, \quad T(S(\vec{x})) = \vec{x}$$
   for all $\vec{x} \in \mathbb{R}^n$. S is called the inverse of T and we denote $S = T^{-1}$.
   (b) Theorem: For $T : \mathbb{R}^n \to \mathbb{R}^n$ a linear transformation with $T(\vec{x}) = A\vec{x}$, T is invertible if and only if A is an invertible matrix. In which case, $T^{-1}(\vec{x}) = A^{-1}\vec{x}$.

4. Homework: 1-7 odd, 11, 15-23 odd, 33

.4 2.4 Partitioned matrices

1. Idea: Generalize matrix multiplication to block multiplication
   - Certain problems naturally lead to symmetry / block structure of a matrix.
   - Can also block matrices to distribute computation to speed up compute time (parallel computing). High performance computing especially uses this approach.

2. Example:
$$A = \left[\begin{array}{cc|ccc} 2 & -3 & 1 & 0 & -4 \\ \hline 1 & 5 & -2 & 3 & -1 \\ 0 & -4 & -2 & 7 & -1 \end{array}\right], \quad B = \left[\begin{array}{cc} 6 & 4 \\ -2 & 1 \\ \hline -3 & 7 \\ -1 & 3 \\ 5 & 2 \end{array}\right]$$

   (a) Note these are compatible for multiplication.
   (b) 2 approaches already: Row and columns
   $$AB = A[\vec{b}_1 \ \vec{b}_2] = [A\vec{b}_1 \ A\vec{b}_2]$$
   $$AB = \begin{bmatrix} row_1(A) \\ row_2(A) \\ row_3(A) \end{bmatrix} B = \begin{bmatrix} row_1(A)B \\ row_2(A)B \\ row_3(A)B \end{bmatrix}$$
   (c) New idea: Partition A and B into blocks.
   $$AB = [A_1 | A_2] \left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right] = [A_1B_1 + A_2B_2]$$
   for A_1 (3×2), A_2 (3×3) and B_1 (2×2) and B_2 (3×2). Compare to the above 2 forms.
   (d) Note: We need submatrices to be compatible for multiplication.
   (e) Can partition further as
   $$AB = \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right] \left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right] = [A_1B_1 + A_2B_2]$$
   for A_1 (1×2), A_2 (1×3), A_3 (2×2), A_4 (2×3) and B_1 (2×2) and B_2 (3×2). Compare to the above.

(f) Think of other ways to partition:
   - A into 4 parts, B into 2
   - B into 3 parts
   - Repeat partitioning leads to the below theorem.

3. Theorem: For A $(m \times n)$ and B $(n \times p)$,

$$AB = [\vec{a}_1 \ \vec{a}_2 \ \ldots \ \vec{a}_n] \begin{bmatrix} row_1(B) \\ row_2(B) \\ \vdots \\ row_n(B) \end{bmatrix}$$

4. Example: Inverses of partitioned matrices.

   (a) Find A^{-1} for A $(n \times n)$ where

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

   where A_{11} $(p \times p)$, A_{12} $(p \times q)$, A_{22} $(q \times q)$, and 0 $(q \times p)$.

   (b) Find matrix B such that

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = I_{n \times n}$$

   Multiplying, we have that
   - $A_{11}B_{11} + A_{12}B_{21} = I_p$
   - $A_{11}B_{12} + A_{12}B_{22} = 0$
   - $A_{22}B_{21} = 0$
   - $A_{22}B_{22} = I_q$

   (c) The last bullet says $B_{22} = A_{22}^{-1}$ from the invertible matrix theorem.

   (d) From the third bullet, $B_{21} = 0$ since A_{22} is invertible and multiplying by A_{22}^{-1}.

   (e) The first bullet then gives $B_{11} = A_{11}^{-1}$.

   (f) Finally, the second bullet gives

$$B_{12} = -A_{11}^{-1}A_{12}A_{22}^{-1}.$$

   (g) Finally,

$$A^{-1} = B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

   for B as derived.

   (h) Note, this approach is especially nice if we can get down to (2×2) matrices where the inverse has a simple formula.

5. Homework: 1, 3, 7, 9, 11, 13, 15.

.5 2.5 Matrix factorizations

1. Lab content

2. Homework: 22-26

.6 2.6 The Leontief input-output model

1. Lab content

2. Homework:

.7 2.7 Applications to computer graphics

1. Lab content
2. Homework:

.8 2.8 Subspaces of $\mathbb{R}^n$

1. Here we generalize to the main theory of linear algebra as a way to delve deeper into $A\vec{x} = \vec{b}$.
2. Subspaces of $\mathbb{R}^n$.
   (a) Definition: A *subspace* of $\mathbb{R}^n$ is any set H in $\mathbb{R}^n$ such that
      - The zero vector is in H
      - For each $\vec{u}, \vec{v} \in H$, we have $(\vec{u} + \vec{v}) \in H$
      - For each $\vec{u} \in H$, we have $(c\vec{u}) \in H$

      This is says subspaces are *closed* under vector addition and scalar multiplication.

   (b) Example: For $\vec{u}, \vec{v}, \vec{w}$ as

$$\vec{u} = \begin{bmatrix} -3 \\ 0 \\ 6 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} 0 \\ -6 \\ 3 \end{bmatrix},$$

   show $Span\{\vec{u}, \vec{v}, \vec{w}\}$ is a subspace of $\mathbb{R}^3$. Is

$$\vec{p} = \begin{bmatrix} 1 \\ 14 \\ -9 \end{bmatrix}$$

   in this subspace?
      - Solution: For $Span\{\vec{u}, \vec{v}, \vec{w}\}$, need to check three properties. Note, $\vec{y} \in Span\{\vec{u}, \vec{v}, \vec{w}\}$ means

$$\vec{y} = x_1\vec{u} + x_2\vec{v} + x_3\vec{w}$$

        for some scalars x_1, x_2, x_3. Then the zero vector is there. Also, show addition and scalar multiplication are preserved.
      - Note, spans will always be a subspace. Sometimes say subspace spanned by these vectors or subspace generate by these vectors.
      - For $\vec{p}$, this amounts to solving a linear system via Gaussian elimination.

$$\left[\begin{array}{ccc|c} -3 & -2 & 0 & 1 \\ 0 & 2 & -6 & 14 \\ 6 & 3 & 3 & -9 \end{array}\right] \sim \left[\begin{array}{ccc|c} -3 & -2 & 0 & 1 \\ 0 & 2 & -6 & 14 \\ 0 & -1 & 3 & -7 \end{array}\right] \sim \left[\begin{array}{ccc|c} -3 & -2 & 0 & 1 \\ 0 & 2 & -6 & 14 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

        The system is consistent and so $\vec{p}$ is in the subspace.

   (c) Example: $\{\vec{0}\}$ is a subspace of $\mathbb{R}^n$. Check three items. Called the zero subspace.

3. Column and null space of a matrix A. The two fundamental subspaces concerning $A\vec{x} = \vec{b}$.
   (a) Definitions:
      - The column space of matrix A, written $col(A)$, is the set of all linear combinations of the columns of A
      - The null space of matrix A, written $nul(A)$, is the set of all vectors which solve the homogeneous system $A\vec{x} = \vec{0}$.

- Note the dimension of each relies on the number of rows and columns of A.

(b) Theorems:

- $col(A)$ is a subspace of $\mathbb{R}^n$. This holds since it is a span of vectors as discussed above.
- $nul(A)$ is a subspace of $\mathbb{R}^n$. This requires careful proof.

  Proof: Check the three items. $A\vec{0} = \vec{0}$. For $\vec{u}, \vec{v} \in nul(A)$, we have

  $$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} = \vec{0} + \vec{0} = \vec{0}$$

  $$A(c\vec{u}) = cA\vec{u} = \vec{0}$$

  which leverages linearity of a matrix transformation.

(c) Example: From previous example, $\vec{p}$ is in $col(A)$.

4. Basis for a subspace.

(a) Definition: A *basis* for a subspace H of $\mathbb{R}^n$ is linearly independent set H which spans H.

(b) Find a basis for the $nul(A)$ for above example.

(c) Find a basis for the $col(A)$ for the above example.

- Set of all columns would span $col(A)$, but they need not be linearly independent. In this case they aren't due to free variables.
- Eliminating to reduced row echelon form, we see how a column is a linear combination of the others.
- While the columns change thru row reduction, the system has the same solution and hence the linear dependence relation does not change.

(d) Theorem: The pivot columns of matrix A form a basis for $col(A)$.

- Note: These columns are the original columns of A, not the echelon form columns. Can see why echelon form wouldn't work with zeros in row entries.

(e) Columns of I form the standard unit basis for $\mathbb{R}^n$. Check that every system is consistent. Full set of pivots says the columns are a basis. Can now see where this terminology comes from.

(f) Find a basis for $col(A)$ and $nul(A)$. Note the difference in dimension of these spaces.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ -5 & -1 & 0 \\ 2 & 7 & 11 \end{bmatrix}$$

5. Homework: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25

.9 2.9 Dimension and rank

1. Coordinate systems

(a) Given a basis $B = \{\vec{b_1}, \ldots, \vec{b_p}\}$ of subspace H in $\mathbb{R}^n$, each element $\vec{x}$ of H can be written uniquely as a linear combination of basis elements of B. Reason: If there were two ways,

$$\vec{x} = c_1\vec{b_1} + \ldots, \quad \vec{x} = d_1\vec{b_1} + \ldots$$

then

$$\vec{0} = \vec{x} - \vec{x} = (c_1 - d_1)\vec{b}_1, \ldots$$

Due to linear independence of the basis vectors, $c_1 - d_1 = 0$ gives $c_1 = d_1$ and likewise for the remaining weights.

(b) Definition: For basis $B = \{\vec{b}_1, \ldots, \vec{b}_p\}$ of subspace H, each $\vec{x}$ in H can be expressed as the *coordinate vector* in $\mathbb{R}^p$ as

$$[\vec{x}]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

where $\vec{x} = c_1\vec{b}_1 + \cdots + c_p\vec{b}_p$.

(c) Example: For

$$\vec{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix},$$

show $B = \{\vec{v}_1, \vec{v}_2\}$ is a basis for the subspace H spanned by these vectors. Then show $\vec{x}$ is in H and find its coordinate vector. In the end

$$[\vec{x}]_B = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Show picture of the book of the intuition of two dimensional subspace H and coordinates with respect to B.

(d) Note that H in the above example resembles $\mathbb{R}^2$ geometrically and with the coordinate vector. Further, $\vec{x} \to [\vec{x}]_B$ has a one-to-one correspondence since coordinate vectors are unique. Call this an *isomorphism* and say H is *isomorphic* to $\mathbb{R}^2$.

(e) Definition: The *dimension* of nonzero subspace H, denoted by $dim(H)$, is the number of vectors in any basis of H. The dimension of $\{\vec{0}\}$ is defined to be zero.

(f) Note: Can show if H has a basis with p vectors, then all possible basis for H must also have p vectors.

2. Return to linear system $A\vec{x} = \vec{b}$.

   (a) Definition: The *rank* of matrix A, denoted $rank(A)$, is the dimension of the column space of A.

   (b) Example: Number 12 in the book on screen. A has 5 total columns with 3 pivot columns. Then $rank(A) = 3$. Note, the dimension of the null space of A is the count of the remaining columns (free variables), 2 in this case. It is always the case that $rank(A) + dim(Nul(A)) = n$ where n is the total column number of A.

   (c) Theorem: For A matrix with n columns, we have that $rank(A) + dim(Nul(A)) = n$.

   (d) Terminology of *rank* is an important number which measures the "singularness" of a matrix. If full rank, nonsingular. If less than full rank, singular. Lower rank means "more" linear independence. Can also show the rank of A is the same as A^T. That is the number of LI columns matches the number of LI rows. Can see this from RREF form of a matrix.

   (e) Theorem: For H a p-dimensional subspace of $\mathbb{R}^n$, any linearly independent set of p elements in H is a basis for H. Likewise, any set of p elements which spans H is a basis for H.

3. Theorem: Last, we add to the Invertible Matrix Theorem

   (a) The columns of A form a basis for $\mathbb{R}^n$.

   (b) $Col(A) = \mathbb{R}^n$

   (c) $dim(Col(A)) = n$

   (d) $rank(A) = n$

   (e) $Nul(A) = \{\vec{0}\}$

   (f) $dim(Nul(A)) = 0$

4. Homework: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23

Chapter 3: Determinants

.1 3.1 Introduction to determinants

1. We return to the question of invertibility for square matrix A.

   (a) 2×2 case from before:

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

   We denote $\det(A) = ad - bc$. Then A is invertible only if $det(A) \neq 0$.

   (b) Notes:

   - $\det(A) = ad - bc = 0$ implies $\frac{a}{c} = \frac{b}{d}$ and columns are parallel (linearly dependent).
   - $\det(A) = a(d - \frac{c}{a}b)$ is the product of the pivots of A. Turns out this will generalize.
   - $det(A)$ not only determinies invertibility, but it also plays a role in the entries of A^{-1}.

   (c) 3×3 case: Perform row reduction on a general matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{11}a_{21} & a_{11}a_{22} & a_{11}a_{23} \\ a_{11}a_{31} & a_{11}a_{32} & a_{11}a_{33} \end{bmatrix} \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & a_{11}a_{32} - a_{12}a_{31} & a_{11}a_{33} - a_{13}a_{31} \end{bmatrix}$$

$$\sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{bmatrix}$$

   where for the last step row 3 is multiplied by $(a_{11}a_{22} - a_{12}a_{21})$ and this new row is added to $-(a_{11}a_{32} - a_{12}a_{31})$. Also,

$$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

   We need these pivots to be non-zero for A to be invertible.

   (d) Notes:

   - If A is invertible, $\Delta \neq 0$. Will show later if $\Delta \neq 0$, then A is invertible.
   - $\det(A) = \Delta$ is the formula for the 3×3 case.
   - The product has 1 entry from each row and 1 from each column. By row / column swapping, there are $6 = 3!$ total ways for the 3×3 case.
   - The sign of each product is determined by the column (or row) permutations. More on this later.
   - Grouping terms reveals (2×2) determinants leading to a recursive computational formula.

$$\begin{aligned} \Delta &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{12}a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| \\ &= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \end{aligned}$$

     where A_{ij} is the submarix of A formed by deleting the ith row of A and the jth column of A. Also,

$$C_{ij} = (-1)^{i+j}|A_{ij}|$$

     is called the (i,j)th cofactor of A.
   - This formula will extend to $n \times n$ matrices.
   - We can group by any row / column we wish here.

2. Example: Compute $\det(A)$ for
$$A = \begin{bmatrix} 2 & 3 & -3 \\ 4 & 0 & 3 \\ 6 & 1 & 5 \end{bmatrix}.$$
Do first row.

3. Definition: For $n \geq 2$, the determinant of $n \times n$ matrix A is
$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{1+n}\det(A_{1n}) = \sum_{j=1}^{n}(-1)^{i+j}a_{1j}det(A_{1j}) = \sum_{j=1}^{n}(-1)^{i+j}a_{1j}C_{1j}$$

4. Theorem: The determinant can be computed by expansion across any row or column of matrix A.
$$\det(A) = a_{i1}\det(A_{i1}) - a_{i2}\det(A_{i2}) + \cdots + (-1)^{1+n}\det(A_{in}) = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}det(A_{ij})$$
$$\det(A) = a_{1j}\det(A_{1j}) - a_{2j}\det(A_{2j}) + \cdots + (-1)^{1+n}\det(A_{nj}) = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}det(A_{ij})$$
Ask them to repeat above example with better row and column.

5. Examples:
   (a) Number 14 in text.
   (b) Determinant of a random 4×4 triangular matrix. Expand about the column of all zeros repeatively.
   (c) Determinant of I.
   (d) Determinant of matrix with column of zeros.
   (e) $\det(A) = \det(A^T)$.

6. Theorem: The determinant of a triangular matrix is the product of the diagonals.

7. Homework: 1, 3, 5, 9, 13, 15, 19, 21, 23, 25, 29, 33, 39, 41

.2 3.2 Properties of determinants

1. Cofactor expansion is easy, but it's recursive nature makes it very inefficient computationally. Instead, we will leverage Gaussian elmination to get to triangular form. This requires two parts:
   (a) Understand the effect of EROPs on $\det(A)$ (as well as elementary matrices).
   (b) Compute $\det(A)$ for an upper triangular matrix.

2. Example: Compare the determinants of each.
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \begin{bmatrix} c & d \\ a & b \end{bmatrix}, \quad \begin{bmatrix} ka & kb \\ c & d \end{bmatrix}, \quad \begin{bmatrix} a+kc & b+kc \\ c & d \end{bmatrix}$$
Generalize this to 3×3.

3. Theorem: Let A be a square matrix.
   - If a multiple of one row of A is added to another to row to produce matrix B, then $\det(B) = \det(A)$.
   - If two rows of A are interchanged to produce B, then $\det(B) = -\det(A)$.

- If one of A is multiplied by k to produce B, then $\det(B) = k\det(A)$.
- Note: This last result (row scaling) shows determinants can be made big or small via rescaling. So the size of the determinant isn't necessarily a good indicator of interest on its own.

4. Example: Accelerated calculation

$$\begin{vmatrix} 0 & 4 & 0 \\ 0 & 0 & 1 \\ 3 & 0 & 5 \end{vmatrix} = - \begin{vmatrix} 3 & 0 & 5 \\ 0 & 0 & 1 \\ 0 & 4 & 0 \end{vmatrix} = \begin{vmatrix} 3 & 0 & 5 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 12$$

5. Example: Use row reduction to accelerate calculation of

$$\begin{vmatrix} 1 & 3 & 2 & -4 \\ 0 & 1 & 2 & -5 \\ 2 & 7 & 6 & -3 \\ -3 & -10 & -7 & 2 \end{vmatrix}.$$

Strive for triangular form then use product of pivots. Can also use cofactor expansion to simplify some. Take out common factors of rows. What if one row all zeros?

6. Note: Getting to triangular form U via row reduction, we have

$$\det(A) = (-1)^r \det(U) = (-1)^r(\text{product of pivots of } U)$$

where r is the number of row swaps used and no row scaling was done.

7. Theorems: For A a square matrix,

- $$\det(A) = \begin{cases} (-1)^r(\text{product of pivots of } U), & \text{if } A \text{ is invertible} \\ 0, & \text{if } A \text{ is not invertible} \end{cases}$$
- A is invertible if and only if $\det(A) \neq 0$. (Use above bullet)
- $\det(A) = \det(A^T)$. (Use the above bullet and IVT. Can now do column operations as we do for row operations)

8. Example: Use the determinant to determine if vectors are linearly independent.

$$\begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} -7 \\ 0 \\ 7 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ -5 \\ -2 \end{bmatrix}$$

9. Elementary matrices:

   (a) Recall, there are three types of elementary matrices each result from EROPs applied to the identity matrix I.

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

   (b) $\det(E)$ is easy to compute in each case.

$$\det(E) = \begin{cases} 1, & \text{for } E \text{ row replacement} \\ -1, & \text{for } E \text{ row swap} \\ k, & \text{for } E \text{ row scaling by } k \end{cases}$$

   This is another argument that the first theorem of this section holds true since we now see

$$\det(EA) = \det(E)\det(A)$$

   . See text for careful proof via induction.

(c) Theorem: $\det(AB) = \det(A)\det(B)$. Prove by writing A as a product of elementary matrices and repeatedly using the above result.

10. Homework: 1, 3, 5, 7, 11, 15, 17, 19, 21, 25, 27, 29, 37, 39

.3 3.3 Cramer's rule, volume, and linear transformations

1. Skip / lab?

Chapter 4: Vector spaces

Skip, content in chapter 2

Chapter 5: Eigenvalues and eigenvectors

We return to the view of $A\vec{x}$ as a transformation $\vec{x} \to A\vec{x}$ and aim to understand the key characteristics of this transformation.

.1 5.1 Eigenvectors and eigenvalues

1. Example: Discrete dynamical systems. Return to the La Crosse / Onalaska population model of our past lab.

$$P = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}, \quad \vec{x}_0 = \begin{bmatrix} 50000/80000 \\ 30000/80000 \end{bmatrix} = \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix}, \quad \vec{x}_n = P\vec{x}_{n-1}$$

where P contains probability transitions and $\vec{x}_0$ is the proportion of populations in La Crosse and Onalaska respectfully.

(a) Question of end behavior: As $n \to \infty$, $\vec{x}_n \to ?$

(b) One way: Iterate via computer. Solution we approach is called the steady state vector of this system which satisfies

$$\vec{x} = P\vec{x}$$

(c) Solving the steady state equation leads to a homogeneous system.

$$\vec{x} = P\vec{x} \to (P - I)\vec{x} = \vec{0} \to \left[\begin{array}{cc|c} 0.6-1 & 0.3 & 0 \\ 0.4 & 0.7-1 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} -0.4 & 0.3 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

with x_2 a free variable giving

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 3/4 \\ 1 \end{bmatrix}$$

Since there are infinite vectors in this null space (with basis $B = \{[3,4]^T\}$), we normalize to make a probability vector.

$$\vec{x} = \begin{bmatrix} 3/7 \\ 4/7 \end{bmatrix}$$

which agrees with our R calculation.

(d) Observation: For matrix P, multiplication by nice vector $[3,7]^T$ results in just a scaling of the original vector. Graph in $x_1 - x_2$ plan to illustrate. Entire line ($Span\{\vec{x}\}$) also has this feature, but vectors not on this line do not.

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

2. Definition: An eigenvector of $n \times n$ matrix A is a nonzero vector $\vec{x}$ such that $A\vec{x} = \lambda\vec{x}$ for λ some scalar (called an eigenvalue).

   - Note: These special directions for a matrix give:
     - insight into the matrix multiplication by A
     - a way to efficiently compute A^k

3. Find eigenvectors and eigenvalues of
$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}.$$

   (a) Require a nonzero vector $\vec{x}$ such that
$$A\vec{x} = \lambda\vec{x} \to (A - \lambda I)\vec{x} = \vec{0}.$$
   This is a homogeneous equation and we seek nontrivial solutions. Then,
$$\det(A - \lambda I) = 0 \quad \text{(Why?)}$$
   must hold and this gives a way to find eigenvalues λ.
$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 6 \\ 5 & 2-\lambda \end{vmatrix} = (1-\lambda)(2-\lambda) - 30 = \lambda^2 - 3\lambda - 28 = 0$$
   giving $\lambda = 7, \lambda = -4$. So there are two systems, $A\vec{x} = 7\vec{x}$ and $A\vec{x} = -4\vec{x}$.

   (b) $\lambda = -4$ gives
$$\vec{x} = x_2 \begin{bmatrix} -6/5 \\ 1 \end{bmatrix} = c \begin{bmatrix} -6 \\ 5 \end{bmatrix} = c\vec{x}$$
   and so eigenvector $\vec{x}$ had corresponding eigenvalue $\lambda = -4$.

   (c) $\lambda = 7$ gives
$$y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

   (d) Graph these eigenvectors in the x_1x_2 plane. $Span\{\vec{x}\}$ and $Span\{\vec{y}\}$ are both subspaces of $\mathbb{R}^2$ since both are $Nul(A - \lambda I)$ for each eigenvalue λ. These are called eigenspaces of A.

4. Process for finding eigenvectors and eigenvalues for $n \times n$ matrix A:

   (a) $A\vec{x} = \lambda\vec{x}$ having a nonzero solution implies homogeneous equation $(A - \lambda I)\vec{x} = \vec{0}$ has nontrivial solutions.

   (b) Require the characteristic equation
$$\det(A - \lambda I) = 0.$$
   This is a polynomial equation in λ of degree n.

   (c) For each solution λ to the characteristic equation, solve $(A - \lambda I)\vec{x} = \vec{0}$ via Gaussian elimination.

5. Examples: Eigenvalue special cases. Find the eigenvalues of each.

   (a) Triangular matrices: Eigenvalues are diagonal entries
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

   (b) Zero eigenvalues imply non-zero eigenvector. So A is singular.
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

(c) Repeat eigenvalues

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

(d) Complex eigenvalues

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

6. Theorem: If $\{\vec{v}_1, \ldots, \vec{v}_p\}$ are eigenvectors of $n \times n$ matrix A with distinct eigenvalues $\lambda_1, \ldots, \lambda_p$, then $\{\vec{v}_1, \ldots, \vec{v}_p\}$ is linearly independent.

   - Proof strategy: To show P (if I am in WI) implies Q (then am in US), it is enough to show P (if I am in WI) and not Q (then not in US) leads to contradiction.
   - Proof: Assume distinct eigenvalues and $\{\vec{v}_1, \ldots, \vec{v}_p\}$ is linearly dependent. The one eigenvector is a linear combination of the preceding ones.

$$c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} = \vec{v}_k \quad \to \quad c_1 A\vec{v}_1 + \cdots + c_{k-1}A\vec{v}_{k-1} = A\vec{v}_k$$

$$\to \quad c_1\lambda_1\vec{v}_1 + \cdots + c_{p-1}\lambda_{k-1}\vec{v}_{k-1} = \lambda_k\vec{v}_k$$

   Multiply this original equation by λ_k and subtract from the final line to get

$$c_1(\lambda_1 - \lambda_k)\vec{v}_1 + \cdots + c_{k-1}(\lambda_{k-1} - \lambda_k)\vec{v}_{k-1} = \vec{0}$$

   Since these eigenvalues are distinct, it must be that c_i's be all zero and so $\vec{v}_k$ is $\vec{0}$. This is not possible for an eigenvector. Contradiction.

7. Homework: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 25, 27

.2 5.2 The characteristic equation

1. Characteristic equation:

   (a) The easiest way to find eigenvalues and eigenvectors of matrix A is to find the eigenvalues first via the characteristic equation.

$$A\vec{x} = \lambda\vec{x}, \quad \vec{x} \neq \vec{0} \quad \to \quad (A - \lambda I)\vec{x} = \vec{0} \quad \to \quad \det(A - \lambda I) = 0$$

   This last determinant gives a polynomial to allow solving for λ. Eigenvectors come from Gaussian elimination.

   (b) Example: Markov chain from last time. Find the eigenvalues.

$$A = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}$$

   (c) Definition: Scalar λ is an eigenvalue of $n \times n$ matrix A if and only if λ satisfies the characteristic equation

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

   Note: This will always be a degree n polynomial in λ.

   (d) Example: Find the eigenvalues of

$$A = \begin{bmatrix} 5 & -2 & 3 \\ 0 & 1 & 0 \\ 6 & 7 & -2 \end{bmatrix}$$

   Note: Row operations aren't useful here, but choosing a smart cofactor expansion helps.

2. Application of eigenvalues: Markov chains

   (a) Previous example:
$$\vec{x}_{k+1} = A\vec{x}_k, \quad A = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}, \quad \vec{x}_0 = \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix}$$

   (b) Already found eigenvalues $\lambda_1 = 1, \lambda_2 = 0.3$ with eigenvectors
$$\vec{v}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

   (c) Theorem from previous section say distinct eigenvalues give linearly independent eigenvectors. True in this case, though easy to see by inspection.

   (d) Write $\vec{x}_0$ as a linear combination of these eigenvectors.
$$\vec{x}_0 = c_1\vec{v}_1 + c_2\vec{v}_2$$

   (e) Can find weights c_1, c_2 as
$$\vec{c} = [\vec{v}_1 \ \vec{v}_2]^{-1}\,\vec{x}_0 = \begin{bmatrix} 1/7 \\ 11/56 \end{bmatrix}$$

   (f) Apply A repeatedly to the eigenvector representation of $\vec{x}_0$.
$$\vec{x}_1 = A\vec{x}_0 = c_1A\vec{v}_1 + c_2A\vec{v}_2 = c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2$$
$$x_2 = c_1\lambda_1^2\vec{v}_1 + c_2\lambda_2^2\vec{v}_2$$
$$x_n = c_1\lambda_1^n\vec{v}_1 + c_2\lambda_2^n\vec{v}_2$$

   For our case,
$$x_n = \frac{1}{7}1^n\vec{v}_1 + \frac{11}{56}(3/10)^n\vec{v}_2 = \begin{bmatrix} 3/7 \\ 4/7 \end{bmatrix} + \frac{11}{56}(3/10)^n\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

   (g) Then as $n \to \infty$, as we saw computationally $\vec{x}_n \to [3/7, 4/7]$ since $(3/10)^n \to 0$. Note, for any initial vector
$$\vec{x}_n = c_1(1)^n\begin{bmatrix} 3 \\ 4 \end{bmatrix} + c_2(3/10)^n\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \to \quad c_1\begin{bmatrix} 3 \\ 4 \end{bmatrix}$$
   unless $c_1 = 0$ in which case $\vec{x}_n \to \vec{0}$.

   (h) Note, so long as $\vec{x}_0$ is such that components are positive and sum to 1 (always with proportion vectors), the result of this system will remain the same. Check simulations in R.

3. Computational challenge of eigenvalues: Finding determinants and zeros of resulting characteristic polynomial is not easy. Solution is to reformulate the problem in an eigenvalue equivalent way.

   (a) Definition: $n \times n$ matrices A and B are similar if there is an invertible matrix P such that
$$B = P^{-1}AP$$
   or equivalently
$$A = P^{-1}BP.$$

   (b) Theorem: If $n \times n$ matrices A and B are similar, then they have the same characteristic equation and hence the same eigenvalues (with the same multiplicities).

   Proof: If $B = P^{-1}AP$, then
$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(A - \lambda I)P$$
   and so
$$\det(B-\lambda I) = \det(P^{-1}(A-\lambda I)P) = \det(P^{-1})\det(A-\lambda I)\det(P) = \det(P^{-1}P)\det(A-\lambda I) = \det(A-\lambda I)$$

(c) Notes:

- Matrices can have the same eigenvalues yet not be similar.
- Similarity is not the same as row equivalence. Row operations usually change eigenvalues.

(d) Jacobi's method: Eigenvalues via iteration and simlarity. For $A = A^T$ (symmetric matrices),

$$A_1 = A, \quad A_{k+1} = P_k^{-1} A_k P_k$$

for rotation matrices P_k. This results in the non-diagonal entries of A_k tending to zero and eigenvalues of A being the limiting diagonals of A_k.

4. Homework: 1, 3, 5, 7, 9, 13, 15, 17, 19, 21, 23, 25, 27

.3 5.3 Diagonalization

1. Motivation:

(a) We saw previously matrix powers show up in applications (dynamical systems), so computing A^k for potentially large values of k is useful.

(b) For matrix A similar to a diagonal matrix D, A^k becomes easy.

$$A = PDP^{-1} \quad \rightarrow \quad A^2 = (PDP^{-1})^2 = PD^2P^{-1} \quad \rightarrow \quad A^k = PD^kP^{-1}$$

Note, D^k is also diagonal with entries of D to the power of k.

(c) Such diagonalizable matrices are connected to eigenvalues and eigenvectors.

2. Diagonalization Theorem: $n \times n$ matrix A is diagonlizable if and only if it has n linearly independent eigenvectors. Specifically,

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

for $P = [\vec{v}_1, \ldots, \vec{v}_n$ a matrix of eigenvectors and $D = diag(\lambda_1, \ldots, \lambda_n)$ a diagonal matrix of corresponding eigenvalues.

- Proof idea:

$$AP = A[\vec{v}_1 \ldots \vec{v}_n] = [A\vec{v}_1 \ldots A\vec{v}_n] = [\lambda_1\vec{v}_1 \ldots \lambda_n\vec{v}_n] = PD$$

Because P is invertible due to a set of linearly independent columns,

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

3. Example: For previous example

$$A = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}$$

we already found eigenvalues $\lambda_1 = 1, \lambda_2 = 3/10$ with corresponding eigenvectors

$$\vec{v}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Then,

$$A = PDP^{-1} \quad \rightarrow \quad A^k = PD^kP^{-1}$$

agrees with previous calculation.

4. Diagonlize the below matrix:

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

- Can show $\lambda_1 = 1, \lambda_2, 3 = 2$. Then,

$$\vec{v}_1 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

5. Homework: 1, 3, 5, 9, 11, 13, 15, 19, 21

.4 5.4 Eigenvectors and linear transformations

1. Skip. Not much interest without general vector spaces. Possible interest for diagonlizable matrix tranformations. Might be worthwhile to see idea of change of basis.
2. Homework:

.5 5.5 Complex eigenvalues

1. Before, eigenvalue and eigenvector problem

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

was though of as directions in which applying A amounts to a scaling of $\vec{x}$. There are cases where this is not the case.

2. 90° rotation:

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

   (a) From the characteristic equation, there are no real eigenvalues. If we allow imaginary numbers, we get $\lambda_1 = i,\ \lambda_2 = -i$.
   (b) Corresponding eigenvectors are then

$$\vec{v}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}$$

   (c) Verify these pairs work.

3. Complex eigenvalues: Full story.

   (a) Example:

$$A = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix}$$

   (b) Eigenvalues:

$$\det(A - \lambda I) = \lambda^2 - 1.6\lambda + 1 = 0 \quad \rightarrow \quad \lambda = 0.8 \pm 0.6i$$

   (c) Eigenvectors:
      - $\lambda_1 = 0.8 - 0.6i$

$$\vec{v}_1 = \begin{bmatrix} -2 - 4i \\ 5 \end{bmatrix}$$

      Note, hard to solve the resulting system with

$$A - \lambda_1 I = \begin{bmatrix} -0.3 + 0.6i & -0.6 \\ 0.75 & 0.3 + 0.6i \end{bmatrix}$$

      but since we assume nontrivial eigenvector, there must be a free variable. Then the second equation yields

$$x_1 = (-0.4 - 0.8i)x_2$$

      giving the above eigenvector.

- $\lambda_2 = 0.8 + 0.6i$

$$\vec{v}_2 = \begin{bmatrix} -2+4i \\ 5 \end{bmatrix}$$

- Note the conjugate pair relationship of eigenvalues and eigenvectors.

(d) How to make sense of complex eigenvalues and eigenvectors? Think of this as a rotation. Plot trajectories for this system.

(e) We see rotation, why does rotation occur?

- Recall the general rotation linear transformation:

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

- We generalize to any matrix

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

with eigenvalues $\lambda = a \pm bi$. Drawing a right triangle with legs a, b in the complex plane with hypotenuse $r = \sqrt{a^2 + b^2}$, we have

$$C = r \begin{bmatrix} a/r & -b/r \\ b/r & a/r \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos(\theta) & -sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

and we see the rotation then a scale.

(f) Example: For above matrix

$$A = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix}$$

and eigenvalue $\lambda_1 = 0.8 - 0.6i$ with eigenvector

$$\begin{bmatrix} -2-4i \\ 5 \end{bmatrix}$$

construct matrix P as

$$P = [Re(\vec{v}_1 \quad IM(\vec{v}_1)] = \begin{bmatrix} -2 & -4 \\ 5 & 0 \end{bmatrix}$$

Can show such a P always has linearly independent columns.

(g) Compute matrix C similar to A such that

$$C = P^{-1}AP = \begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}$$

(h) Then, $A = PCP^{-1}$ with C the rotation and P, P^{-1} acting as a change of variable (draw the rectangular transformation map, note same map applies to general diagonalization). Note that $r = |\lambda_1| = 1$ and so can compute θ from

$$\cos(\theta) = 0.8, \quad \sin(\theta) = 0.6$$

giving $\theta \approx 36.87°$.

4. Above ideas generalize to the following theorem:

- Theorem: For A a 2×2 real matrix with complex eigenvalue $\lambda = a - bi$ ($b \neq 0$) and associate eigenvector $\vec{v}$ in $\mathbb{C}^2$, then

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

for

$$P = [Re(\vec{v} \quad Im(\vec{v}], \quad C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

5. Note: Euler's identity is worth mentioning here.

6. Homework: 1, 3, 7, 9, 13, 15, 17, 21

.6 5.6 Discrete dynamical systems

1. Project material. Intro covered before. Skip.
2. Homework:

.7 5.7 Applications to differential equations

1. ODEs:
   (a) Examples: Note infinitely many solutions for each.
   $$x' = x, \quad x' = -2x, \quad x' = 0.5x$$
   for $x = x(t)$ an unknown function of time t.
   (b) Initial value problems:
   $$\begin{cases} x' = 0.5x \\ x(0) = 80 \end{cases}$$
   where $x(0)$ is called the initial condition and this $x(t)$ is a particular solution to this ODE.
   (c) General initial value problem
   $$\begin{cases} x' = \lambda x \\ x(0) = x_0 \end{cases}$$
   giving general solution
   $$x(t) = x_0 e^{\lambda t}.$$
2. Example: Real eigenvalues
   (a) MORE INTERESTING EXAMPLE:
   $$A = \begin{bmatrix} -1.5 & 0.5 \\ 1 & -1 \end{bmatrix}, \quad \vec{x}(0) = [5, 4]^T$$
   (b) Coupled system of ODEs:
   $$\begin{cases} x_1' = -2x_1 + x_2 \\ x_2' = x_1 - 2x_2 \end{cases}$$
   with initial conditions
   $$\begin{cases} x_1(0) = 6 \\ x_2(0) = 2 \end{cases}$$
   (c) Matrix formulation:
   $$\frac{d\vec{x}}{dt} = A\vec{x}, \quad A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$
   with initial condition
   $$\vec{x}(0) = \begin{bmatrix} 6 \\ 0 \end{bmatrix}.$$
   (d) Idea: For any eigenvector of A, $\vec{v}$, with corresponding eigenvalue λ, write $\vec{x} = e^{\lambda t}\vec{v}$. Then,
   $$\frac{d\vec{x}}{dt} = \lambda\vec{x}, \quad A\vec{x} = \lambda\vec{x}.$$
   Noting both A and $\frac{d}{dt}$ are linear operators on $\vec{x}$, we apply the superposition principle to get
   $$\vec{x} = c_1 e^{\lambda_1 t}\vec{v}_1 + c_2 e^{\lambda_2 t}\vec{v}_2$$

where c_1, c_2 satisfy the initial condition

$$\vec{x}(0) = c_1\vec{v}_1 + c_2\vec{v}_2.$$

Note this initial condition can be written this way so long as eigenvectors are linearly independent and form a basis for $\mathbb{R}^2$.

(e) For this example, $\lambda_1 = -1$ and $\lambda_2 = -3$ with eigenvectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

giving

$$\vec{x} = c_1 e^{-t}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 e^{-3t}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Applying the initial condition $\vec{x}(0) = [6, 2]^T$ gives $c_1 = 4$ and $c_2 = 2$ and so

$$\vec{x} = \begin{bmatrix} 4e^{-t} + 2e^{-3t} \\ 4e^{-t} - 2e^{-3t} \end{bmatrix}.$$

(f) Plot this trajectory in R. Note the tending to zero.

(g) Use pplane.html to plot many trajectories. This is known as a sink with the origin as the attractor. Note can see the eigenvectors and power of the eigenvalue.

(h) Imagine if the sign of the eigenvalues change. Leads to sink, source, saddle.

3. General linear systems story: Diagonlization is the key.

(a) System

$$\frac{d\vec{x}}{dt} = A\vec{x}, \quad \vec{x}(0) = \vec{x}_0$$

has solution

$$\vec{x}c_1 e^{\lambda_1 t}\vec{v}_1 + \ldots c_n e^{\lambda_n t}\vec{v}_n$$

for λ_i eigenvalues of A with eigenvectors $\vec{v}_i$ with constants c_i such that

$$\vec{x}_0 = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n.$$

(b) Reason: Diagonalization is the key.

- For A diagonlizable write

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

  and perform the change of variable

$$\vec{y} = P^{-1}\vec{x}, \quad (P\vec{y} = \vec{x})$$

  and so

$$P\frac{d\vec{y}}{dt} = PD\vec{y} \quad \rightarrow \quad \frac{d\vec{y}}{dt} = D\vec{y}.$$

- This system is a decoupled linear systems which can be solved as a single ODE.

$$\vec{y} = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix}$$

  and finally

$$\vec{x} = P\vec{y} = c_1 e^{\lambda_1 t}\vec{v}_1 + \cdots + c_n e^{\lambda_n t}\vec{v}_n.$$

4. Complex eigenvalues and orbits:

   (a) For A real and $\vec{x}_0$ real, we expect real solutions to

   $$\frac{d\vec{x}}{dt} = A\vec{x}.$$

   For complex eigenvalues, we will see Euler's identity again.

   $$e^{i\theta} = \cos(\theta) + i\sin(\theta)$$

   (b) Example:

   $$\frac{d\vec{x}}{dt} = A\vec{x}, \quad A = \begin{bmatrix} -2 & 1 \\ -1 & -2 \end{bmatrix}, \quad \vec{x}(0) = [6, 2]^T$$

   - We have complex eigenvalues $\lambda = 2 \pm i$ with corresponding eigenvectors

   $$\vec{v} = \begin{bmatrix} 1 \\ \pm i \end{bmatrix}$$

   - Solving for the initial condition $\vec{x}(0) = [6, 2]^T$ gives

   $$\vec{x} = c_1 e^{\lambda_1 t}\vec{v}_1 + c_2 e^{\lambda_2 t}\vec{v}_2$$

5. Homework: 1, 3, 5

.8 5.8 Iterative estimates to eigenvalues

1. Homework:

Chapter 6: Orthogonality and least squares

1. End goal: Least squares approximation

   (a) The main goal of this chapter is to solve an overdetermined system where $A\vec{x} = \vec{b}$ has no solution. That is $\vec{b}$ is not in the column space of A.

   $$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

   (b) Linear regression, housing example, square foot vs price. If $y = mx + b$ for m price per square foot and b the intercept, we have

   $$\begin{bmatrix} 1 & sf_1 \\ 1 & sf_2 \\ 1 & sf_3 \\ 1 & sf_4 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}$$

   (c) The resulting line can be used to

   - Better understand the data
   - Predict on new data

   (d) How to find a best approximation in $Col(A)$? Project onto this space. Key ingredients:

   - Angle (inner product)
   - Length (norm)
   - Orthogonality (unique directions)

.1 6.1 Inner product, length, and orthogonality

1. Inner production: Comparing direction
   (a) $\mathbb{R}^2$: For $\vec{u} = [u_1, u_2]^T, \vec{v} = [v_1, v_2]^T$,
   $$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 = \vec{u}^T\vec{v}$$
   (b) $\mathbb{R}^n$: For $\vec{u} = [u_1, u_2, \ldots, u_n]^T, \vec{v} = [v_1, v_2, \ldots, v_n]^T$,
   $$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \vec{u}^T\vec{v}$$
   (c) Example
   (d) Properties of dot product:
   - $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
   - $(\vec{u} \cdot \vec{v}) \cdot \vec{w} = \vec{u} \cdot (\vec{v} \cdot \vec{w})$
   - $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v})$
   - $\vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$
2. Length: AKA Euclidean norm
   (a) $\mathbb{R}^2$: Draw picture, Pythagorean theorem
   $$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2} \to \|\vec{u}\|^2 = \vec{u} \cdot \vec{u}$$
   (b) $\mathbb{R}^n$:
   $$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} \to \|\vec{u}\|^2 = \vec{u} \cdot \vec{u}$$
   (c) Norm property: For scalar c,
   $$\|c\vec{u}\| = \sqrt{(c\vec{u}) \cdot (c\vec{u})} = \sqrt{c^2}\sqrt{\vec{u} \cdot \vec{u}} = |c|\|\vec{u}\|$$
   (d) Unit vector: Any $\vec{u}$ such that $\|\vec{u}\| = 1$. For any $\vec{u}$, $\frac{\vec{u}}{\|\vec{u}\|}$ is a unit vector.
   (e) Example: Vector in $\mathbb{R}^3$, find length, make a unit vector, not two options for $\pm$.
3. Distance:
   (a) $\mathbb{R}^2$: Distance formula, draw picture, tip-to-tip length
   $$dist(\vec{u}, \vec{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 + v_2)^2} = \|\vec{u} - \vec{v}\|$$
   (b) $\mathbb{R}^n$:
   $$dist(\vec{u}, \vec{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 + v_2)^2 + \cdots + (u_n + v_n)^2} = \|\vec{u} - \vec{v}\|$$
4. Orthogonality: How to tell if $\vec{u}$ and $\vec{v}$ are opposite directions?
   (a) Definition: $\vec{u}$ and $\vec{v}$ are orthogonal if and only if $\vec{u} \cdot \vec{v} = 0$.
   (b) Reason 1: Geometrically want $dist(\vec{u}, \vec{v}) = dist(\vec{u}, -\vec{v})$.
   $$\|\vec{u} - \vec{v}\|^2 = \|\vec{u} + \vec{v}\|^2 \quad \to \quad (\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) = (\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v}) \quad \to \quad 4\vec{u} \cdot \vec{v} = 0$$
   (c) Reason 2: Want Pythagorean theorem to hold:
   $$\|\vec{u}\|^2 + \|\vec{v}\|^2 = \|\vec{u} - \vec{v}\|^2 \quad \to \quad -2\vec{u} \cdot \vec{v} = 0$$

(d) Reason 3: Consequence of law of cosines: $c^2 = a^2 + b^2 - 2ab\cos(\theta)$

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\|\cos(\theta)$$

results in Pythagoras again for $\theta = \frac{\pi}{2}$.

(e) Consequentially by comparison from the law of cosines

$$\vec{u} \cdot \vec{v} = \|\vec{u}\|\|\vec{v}\|\cos(\theta)$$

and the dot product says something about the angle between vectors. This is known as cosine similarity of two vectors.

(f) Note, $\vec{0} \cdot \vec{u} = 0$ always so $\vec{0}$ is orthogonal to every vector.

5. Vector spaces and orthogonal compliments:

(a) Definitions

- If $\vec{z}$ is orthogonal to every vector in subspace W, then we say $\vec{z}$ is orthogonal to subspace W.
- The set of all vectors $\vec{z}$ orthogonal to subspace W is call the orthogonal compliment of W written $W^\perp$.

(b) Theorems:

- Can show if W is a subspace, then $W^\perp$ is also.
- Fundamental theorem of linear algebra: For A an $m \times n$ matrix, we have

$$(Row(A))^\perp = Nul(A), \quad \text{and} \quad (Col(A))^\perp = Nul(A^T)$$

Proof: If $\vec{x} \in Nul(A)$ we have $A\vec{x} = \vec{0}$. Then for any row of A,

$$\vec{r_i} \cdot \vec{x} = 0.$$

Because $Row(A)$ is all linear combinations of rows of A, we have $\vec{x} \in Row(A)^\perp$. Considering this first result, apply it to A^T and the second result follows.

6. Homework: 1, 3, 5, 7, 11, 13, 15, 17, 19, 23, 25, 27, 31

.2 6.2 Orthogonal sets

1. Orthogonal sets:

(a) Definition: A set of vectors $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$ is an orthogonal set of vectors in $\mathbb{R}^n$ if

$$\vec{u}_i \cdot \vec{u}_j = 0$$

for all distinct pairs $i \neq j$. If all vectors are unit vectors (length 1), then is called an orthonormal set.

(b) Examples:

- $\left\{\vec{i}, \vec{j}, \vec{k}\right\}$ is an orthogonal set in $\mathbb{R}^3$.
- $\left\{[0,1,0]^T, [1,0,1]^T, [1,0,-1]\right\}$ is an orthogonal set in $\mathbb{R}^3$.

(c) Theorem: If $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and forms a basis for $Span\{S\}$.

Proof: Need to show

$$c_1\vec{u}_1 + \ldots c_p\vec{u}_p = \vec{0}$$

requires $c_1 = c_2 = \cdots = c_p = 0$. Take the dot product with $\vec{u}_1$ on both sides.

$$(c_1\vec{u}_1 + \ldots c_p\vec{u}_p) \cdot \vec{u}_1 = \vec{0} \cdot \vec{u}_1 \quad \to \quad c_1\vec{u}_1 \cdot \vec{u}_1 = 0 \quad \to \quad c_1 = 0$$

Likewise, all other $c_j = 0$.

2. Orthogonal basis: Best choice of a basis.

   (a) Theorem: For $S = \{\vec{u}_1, \dots, \vec{u}_p\}$ an orthogonal basis of subpace W of $\mathbb{R}^n$, each $\vec{y}$ in W can be written
   $$\vec{y} = c_1\vec{u}_1 + \cdots + c_p\vec{u}_p$$
   where
   $$c_j = \frac{\vec{y}\cdot\vec{u}_j}{\vec{u}_j\cdot\vec{u}_j}, \quad j = 1, 2, \dots, p.$$
   Proof: Use same dot product idea as above. Note that $\vec{u}_i \cdot \vec{u}_i = \|\vec{u}_i\|^2 \neq 0$.

   (b) Notes:
   - We no longer need to solve a linear system to write $\vec{y}$ as a linear combination of basis elements.
   - This is even nicer for orthonormal basis.
   $$\vec{y} = (\vec{y}\cdot\vec{u}_1)\vec{u}_1 + \cdots + (\vec{y}\cdot\vec{u}_p)\vec{u}_p$$
   - Relate above basis to standard unit basis in $\mathbb{R}^n$.

   (c) Example: Write $\vec{y} = [3, 1, 4]$ as a linear combination of basis $\left\{[0, 1, 0]^T, [1, 0, 1]^T, [1, 0, -1]\right\}$.

3. Orthogonal projections:

   (a) $\mathbb{R}^2$: Recall, one can project a vector onto another using basic trigonometry.
   - Project $\vec{v}$ onto vector $\vec{u}$. Draw right triangle with angle θ between two vectors..
   - Need to find length c of triangle leg parallel with $\vec{u}$. Use right triangle trigonometry.
   $$c = \|\vec{v}\|\cos(\theta) = \|\vec{v}\|\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|} = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|}$$
   - Then,
   $$proj_{\vec{u}}(\vec{v}) = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|}\vec{u}$$
   - The triangle leg orthogonal to $\vec{u}$ is called the component of $\vec{v}$ orthogonal to $\vec{u}$. Then,
   $$\vec{v} = proj_{\vec{u}}(\vec{v}) + comp_{\vec{u}}(\vec{v}) \quad \rightarrow \quad comp_{\vec{u}}(\vec{v}) = \vec{v} - proj_{\vec{u}}(\vec{v})$$
   - Think of $\vec{v} = proj_{\vec{u}}(\vec{v}) + comp_{\vec{u}}(\vec{v})$ as decomposing $\vec{v}$ as the sum of two orthogonal vectors. This is the same as the previous theorem
   $$\vec{v} = \frac{\vec{v}\cdot\vec{u_1}}{\vec{u_1}\cdot\vec{u_1}} + \frac{\vec{v}\cdot\vec{u_2}}{\vec{u_2}\cdot\vec{u_2}}$$

   (b) Example: Write $\vec{v} = [7, 6]^T$ as a sum of a vector parallel and perpendicular to $\vec{u} = [2, 1]^T$. Draw picture to illustrate. Verify that the two vectors are orthogonal via the dot product. In the end,
   $$proj_{\vec{u}}(\vec{v}) = [8, 4]^T, \quad comp_{\vec{u}}(\vec{v}) = [-1, 2]^T.$$

4. Orthogonal matrices: For U a square matrix with orthogonal columns, can show
   - $U^TU = I$ (inverse is its transpose)
   - $\|U\vec{x}\| = \|\vec{x}\|$ (maintains length)
   - $(U\vec{x})\cdot(U\vec{y}) = \vec{x}\cdot\vec{y}$ (maintains angle between)

5. Homework: 1, 3, 5, 7, 9, 11, 13, 17, 23, 25, 27

.3 6.3 Orthogonal projections

1. Here we extend the idea of projecting a vector onto another vector to projecting a vector onto a subspace.

2. Theorem: For W any subspace of $\mathbb{R}^n$, each $\vec{y}$ in $\mathbb{R}^n$ can be written uniquely as

$$\vec{y} = \vec{\hat{y}} + \vec{z}$$

where $\vec{\hat{y}} \in W$ and $\vec{z} \in W^{\perp}$. Further, for $\{\vec{u}_1, \dots, \vec{u}_p\}$ an orthogonal basis of W,

$$\vec{\hat{y}} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \cdots + \frac{\vec{y} \cdot \vec{u}_n}{\vec{u}_n \cdot \vec{u}_n}\vec{u}_n = proj_W(\vec{y})$$

and

$$\vec{z} = proj_{W^{\perp}}(\vec{y}) = \vec{y} - \vec{\hat{y}}.$$

   - Show graph of vector $\vec{y}$ projected onto a subspace.
   - This theorem is the first step towards solving the least squares problem.
   - Proof in text.

3. Example: Find the projection of $\vec{y}$ onto subspace W where $W = Span\{\vec{u}_1, \vec{u}_2\}$.

$$\vec{y} = \begin{bmatrix} 6 \\ 3 \\ -2 \end{bmatrix}, \quad \vec{u}_1 = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} -4 \\ 3 \\ 0 \end{bmatrix}$$

   - Note: Can think of this as just using the theorem of the previous section. First project $\vec{y}$ onto $\vec{u}_1$, then $\vec{y}$ onto $\vec{u}_2$ and so $proj_W(\vec{y})$ is the sum of those two projections using the parallelogram law. Draw a picture.

4. Theorem (Best approximation theorem): For W a subspace of $\mathbb{R}^n$ and any vector $\vec{y}$ in $\mathbb{R}^n$, let $\vec{\hat{y}} = Proj_W(\vec{y})$. Then,

$$\|\vec{y} - \vec{\hat{y}}\| < \|\vec{y} - \vec{v}\|$$

for all $\vec{v} \in W$, $\vec{v} \neq \vec{\hat{y}}$.

   - Proof in text. Rather big deal. Above picture gives intuition.
   - This solves the least squares problem for $W = Col(A)$.

5. Homework: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21

.4 6.4 The Gram-Schmidt process

1. Orthogonal basis have a huge advantage:
   (a) Linear combinations are easy
   (b) Projections are easy, essential for approximation
   (c) Matrices with orthogonal columns are slick
   (d) Question: How to find an orthogonal basis for a subspace? Gram-Schmidt process

2. Idea of Gram Schmidt: Use projection and component to get orthogonal basis.
   (a) Example: Consider the subspace W spanned by

$$\vec{x_1} = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}, \quad \vec{x_2} = \begin{bmatrix} 5 \\ 6 \\ -7 \end{bmatrix}$$

- Show basis is LI but not orthogonal. Draw picture in $\mathbb{R}^3$.
- New basis will be $\{\vec{v}_1, \vec{v}_2\}$ which spans W but is now orthogonal.
- Draw picture of idea

$$\vec{v}_1 = \vec{x}_1, \quad \vec{v}_2 = x_2 - proj_{x_1}(\vec{x}_2) = comp_{x_1}(\vec{x}_2)$$

- Check the result is orthogonal. How to check $\{\vec{v}_1, \vec{v}_2\}$ and $\{\vec{x}_1, \vec{x}_2\}$ both span the same subspace W? Enough to see $\{\vec{v}_1, \vec{v}_2\}$ are both linear combinations of $\{\vec{x}_1, \vec{x}_2\}$, but we proved this already in past section (orthonormal set is a basis).
- Easy to make an orthonormal basis.

$$\left\{ \frac{1}{\|\vec{v}_1\|}\vec{v}_1, \quad \frac{1}{\|\vec{v}_2\|}\vec{v}_2 \right\}$$

- Coding example.

(b) Example: For 3+ vectors, same idea. Just need to project onto a subspace.
- Random 4×3 matrix $X = [\vec{x}_1\ \vec{x}_2\ \vec{x}_3]$. We find an orthonormal basis for the $Col(A)$.
- First two steps same as before. Last step projects onto a plane.
- Draw picture of process.

$$\begin{aligned} \vec{v}_1 &= \vec{x}_1 \quad (W_1 = Span(\vec{v}_1)) \\ \vec{v}_2 &= \vec{x}_2 - proj_{W_2}(\vec{x}_2) \quad (W_2 = Span(\vec{v}_1, \vec{v}_2)) \\ \vec{v}_3 &= \vec{x}_3 - proj_{W_3}(\vec{x}_3) \quad (W_3 = Span(\vec{v}_1, \vec{v}_2, \vec{v}_3)) \end{aligned}$$

- Written in terms of dot product:

$$\begin{aligned} \vec{v}_1 &= \vec{x}_1 \\ \vec{v}_2 &= \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 \\ \vec{v}_3 &= \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 \end{aligned}$$

- Show theorem in text for p dimensional case.
- Normalize if wanted.
- Coding example.

3. QR decomposition: Can write the above Gram schmidt process as a matrix factorization.

(a) Can write the above as a matrix factorization

$$X = QR, \quad Q = [\vec{v}_1, \vec{v}_2, \vec{v}_3]$$

(b) Note, Q is an orthonormal matrix and so $Q^{-1} = Q^T$ and so

$$R = Q^T X = \begin{bmatrix} \vec{v}_1 \cdot \vec{x}_1 & \vec{v}_1 \cdot \vec{x}_2 & \vec{v}_1 \cdot \vec{x}_3 \\ \vec{v}_2 \cdot \vec{x}_1 & \vec{v}_2 \cdot \vec{x}_2 & \vec{v}_2 \cdot \vec{x}_3 \\ \vec{v}_3 \cdot \vec{x}_1 & \vec{v}_3 \cdot \vec{x}_2 & \vec{v}_3 \cdot \vec{x}_3 \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot \vec{x}_1 & \vec{v}_1 \cdot \vec{x}_2 & \vec{v}_1 \cdot \vec{x}_3 \\ 0 & \vec{v}_2 \cdot \vec{x}_2 & \vec{v}_2 \cdot \vec{x}_3 \\ 0 & 0 & \vec{v}_3 \cdot \vec{x}_3 \end{bmatrix}$$

(c) General proof in text

(d) Easy to solve $A\vec{x} = \vec{b}$ now.

(e) Easy to invert A now. Work was transferred to GS process.

(f) Eigenvalues and eigenvectors now easier.

(g) Coding example.

4. Homework: 1, 3, 5, 7, 9, 13, 15, 17

.5 6.5 Least-squares problems

1. Return to linear systems $A\vec{x} = \vec{b}$
   (a) For A $m \times n$ with $m >> n$, such systems usually have no solution.
   (b) Idea is to find the best fit solution by making
   $$\|\vec{b} - A\vec{x}\|$$
   as small as possible.
   (c) Because this error is the sum of squares, this is called the least squares problem.
   (d) Key idea is projection.
2. Least squares problem
   (a) Definition: For A $m \times n$ and $\vec{b} \in \mathbb{R}^m$, the least squares solution of $A\vec{\hat{x}} = \vec{b}$ is vector $\vec{x}$ in $\mathbb{R}^n$ such that
   $$\|\vec{b} - A\vec{\hat{x}}\| \leq \|\vec{b} - A\vec{x}\|$$
   for all $\vec{x} \in \mathbb{R}^n$. This is the same idea as the best approximation theorem.
   (b) Idea:
   - If $\vec{b}$ is not in $Col(A)$, project $\vec{b}$ onto the columns space as $\vec{\hat{b}} \in Col(A)$. Draw picture. Solve the resulting linear system.
   $$A\vec{\hat{x}} = \vec{\hat{b}}$$
   - Can approach this as in previous sections with
   $$\vec{\hat{b}} = proj_{Col(A)}(\vec{b}),$$
   but the fundamental theorem of linear algebra finds another way.
   - Since $\vec{b} \perp \vec{\hat{b}}$, we have $\vec{\hat{b}} = A\vec{\hat{x}}$ and
   $$(\vec{b} - A\vec{\hat{x}}) \in Col(A)^{\perp} = Nul(A^T)$$
   and so
   $$\vec{a}_j^T(\vec{b} - A\vec{\hat{x}}) = \vec{0} \quad \rightarrow \quad A^T(\vec{b} - A\vec{\hat{x}}) = \vec{0} \quad \rightarrow \quad A^TA\vec{\hat{x}} = A^T\vec{b}$$
   - No need to compute $\vec{\hat{b}}$ at all.
3. Homework: 1, 3, 5, 7, 17, 25

.6 6.6 Applications to linear models

1. Homework:

.7 6.7 Inner product spaces

1. Homework:

.8 6.8 Applications of inner product spaces

1. Homework:

Chapter 7: Symmetric matrices and quadratic forms

.1 7.1 Diagonalization of symmetric matrices

1. Homework:

.2 7.2 Quadratic forms

1. Homework:

.3 7.3 Constrained optimization

1. Homework:

.4 7.4 The singular value decomposition

1. Homework:

.5 7.5 Applications to image processing and statistics

1. Homework: