Math 2270-6

Notes of 4/19/19

Announcements

- We are done with our subject. The remaining three days of the semester will be review.
- The last day of classes is Tuesday. Wednesday is a reading day.
- I'm planning to be in the office Wednesday, 8:00-2:00 and you are welcome to drop by if you want to discuss our subject. However, if you need make a special trip to see me let's set an appointment.
- Our final exam will take place Thursday, April 25, 8:00-10:00, in our regular classroom.
- It will cover the semester about evenly. (There will be no particular emphasis on Chapter 7).

Quotes

- My first thought when I find myself in a room full of manure is "there must be a pony here somewhere". (Ronald Reagan)
- The purpose of computing is insight, not numbers (Hamming.)
- It ain't so much the things we don't know that get us into trouble. It's the things we do know that just ain't so. (Artemus Ward.)
- In theory, theory and praxis are the same. In praxis, they aren't. (Richard Nixon.)
- Don't worry about your difficulties with mathematics. I can assure you that mine are still greater. (Albert Einstein, to his neighbor's young daughter.)
- A Vulgar Mechanick can practice what he has been taught or seen done, but if he is in an error he knows not how to find it out and correct it, and if you put him out of his road, he is at a stand; Whereas he that is able to reason nimbly and judiciously about figure, force and motion, is never at rest till he gets over every rub. (Isaac Newton)

Principles

- Focus on understanding concepts, facts, and connections. Numerical calculations are for computers.
- Understanding any piece of mathematics means being able to
  - explain it in terms of simpler mathematics,
  - to make many and redundant connections between facts and concepts,
  - to recognize and verbalize the underlying principles,
  - to solve mathematical problems,
  - to apply the mathematics to problems outside of mathematics.
- Understanding a piece of mathematics does **not** mean the ability
  - to apply a formula to a specific type of problem,
  - to read an example, and then do a similar example,
  - to google a word or phrase and apply what you find, correctly or incorrectly
  - to recite formulas, definitions, and theorems,
  - to pass a test, and then forget the material,
  - to operate a calculator,
  - to use Maple or Matlab.

- **Language matters!** If you do not understand the technical language of your subject you will not be able to think about it, to make progress in understanding it, and to use it to solve problems. So make sure you know the precise meaning of the technical words and phrases in your subject.

- Here is an incomplete list of words and phrases that we introduced in our class. You want to know what each of them means:

adjugate matrix, algebraic multiplicity of an eigenvalue, augmented matrix, basic variable, basis, basis conversion, best approximation, block matrix, characteristic equation, characteristic polynomial, coefficient matrix, cofactor, cofactor expansion, columns of a matrix, column space of a matrix, consistent linear system, coordinates of a vector with respect to a basis, Cramer's Rule, defective matrix, determinant of a matrix, diagonal entries of a matrix, diagonal of a matrix, diagonalizable matrices, diagonal matrix, dimension of a vector space, eigenspace of a matrix, eigenvalue of a matrix, eigenvector of a matrix, elementary row operations, entries of a matrix, equivalent linear system, Fourier Coefficients, Fourier Series, free variable, geometric multiplicity of an eigenvalue, Gershgorin Circle, Gershgorin Theorem, Hermitian matrix, homogeneous linear system, indefinite matrix, inner product, inner product space, inverse matrix, invertible matrix, isomorphism, Jordan block, Jordan canonical form, kernel of a matrix, least squares, least squares solution of an overdetermined system, left singular vectors of a matrix, linear combination of a set of vectors, linear function, linearly independent set of vectors, linear space, linear transformation, lower triangular matrix, LU factorization of a matrix, matrix, matrix multiplication, negative definite matrix, negative semidefinite matrix, norm of a vector, normal equations, null space of a matrix, one-to-one mapping, onto mapping, orthogonal basis of a linear space, orthogonal complement of a subspace, orthogonal decomposition of a vector, orthogonally similar matrices, orthogonal matrix, orthogonal projection, orthogonal set, orthogonal vectors, orthonormal basis, orthonormal set, orthonormal vectors, partitioned matrix, permutation matrix, pivot column, pivoting, pivot position, pivot row, positive definite matrix, positive semidefinite matrix, QR factorization of a matrix, quadratic form, rank of a matrix, rectangular matrix, reduced row echelon form of a matrix, right hand side of a linear system, right singular vectors of a matrix, row echelon form of a matrix, rows of a matrix, row space of a matrix, scalar, similarity transform, similar matrices, singular matrix, singular value decomposition, singular values, six views of matrix multiplication, solution of a linear system, solution set of a linear system, solving a linear system, span of a set of vectors, spanning set of a vector space, square matrix, standard basis of $\mathbb{R}^n$, standard basis vectors in $\mathbb{R}^n$, standard matrix of a linear transformation, subspace of a vector space, symmetric matrix, triangular matrix, unit lower triangular matrix, upper triangular matrix, vector, weighted least squares.

- Matrices define linear transformations between finite dimensional vector spaces, and for every linear transformation and given bases of domain and range there is a unique matrix that defines that transformation.
- In short, matrices and linear transformations are synonymous.
- Multiplying matrices means composing linear functions.
- To generalize a concept we ask what are its key properties, and then investigate what else has these properties. We used this procedure in our class to generalize $\mathbb{R}^n$ to linear spaces, and the dot product to inner products.
- While we may not carry out actual computations that way, it is often useful to think of numerical operations in terms of multiplying with certain matrices.
- Block matrices work like ordinary matrices.
- Whenever you see a minimization problem look for a positive definite matrix.
- Multiplying with orthogonal matrices does not amplify errors.

- A real matrix has an orthonormal set of eigenvectors if and only if it is symmetric.
- Trying to compute eigenvalues by computing the characteristic polynomial and finding its roots is futile except for very small matrices. By comparison, finding roots of a polynomial by finding the eigenvalues of its companion matrix works beautifully.
- The general solution of any linear problem is any particular solution, plus the general solution of the associated homogeneous problem.

Semester Subject Review

This list is neither complete nor self contained. Rather, the individual points should stir your memory of related concepts, facts, and connections. If you draw a blank you should review the relevant parts of your notes or the textbook.

- The focus of chapter 1 is on linear systems

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

  where A is an $m \times n$ matrix, $\mathbf{x}$ is in $\mathbb{R}^n$, and $\mathbf{b}$ is in $\mathbb{R}^m$.

- A vector $\mathbf{x}$ satisfying the equation $A\mathbf{x} = \mathbf{b}$ is a **solution** of the linear system.

- **Solving** the linear system means figuring out whether there is a solution, and if so, how many, and what they are.

- If $m = n$ then the matrix A, and the system $A\mathbf{x} = \mathbf{b}$, are said to be **square.**

- $A\mathbf{x}$ is a **linear combination** of the columns of A with the coefficients being given by the entries of $\mathbf{x}$. If

$$A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n\,] \tag{2}$$

where the $\mathbf{a}_i$ are the columns of A, and

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \tag{3}$$

then

$$A\mathbf{x} = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n\,] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i \mathbf{a}_i = \mathbf{b}. \tag{4}$$

- In general, a **linear combination** of vectors $\mathbf{v}_1$, $\mathbf{v}_2$, ... $\mathbf{v}_n$ is an expression of the form

$$\sum_{i=1}^{n} c_i \mathbf{v}_i \tag{5}$$

where the **coefficients** or **weights** c_i are real numbers.

- A function

$$\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^m \tag{6}$$

is a function whose domain is $\mathbb{R}^n$ and whose codomain is $\mathbb{R}^m$. We also write it in a more familiar form as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) \tag{7}$$

where $\mathbf{x}$ is in $\mathbb{R}^n$ and $\mathbf{y}$ is in $\mathbb{R}^m$,

- $\mathbf{f}$ is **linear** if

$$\mathbf{f}(\mathbf{u}+\mathbf{v}) = \mathbf{f}(\mathbf{u}) + \mathbf{f}(\mathbf{v}) \quad \text{and} \quad \mathbf{f}(c\mathbf{u}) = c\mathbf{f}(\mathbf{u}) \tag{8}$$

  for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ and **scalars** (real numbers) c.

- The matrix transformation

$$\mathbf{f}(\mathbf{x}) = A\mathbf{x} \tag{9}$$

  is linear.

- Actually, and most amazingly, given a linear function $\mathbf{T}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, there is a matrix A such that

$$\mathbf{T}(\mathbf{x}) = A\mathbf{x}. \tag{10}$$

- To construct the matrix A we use the columns of the identity matrix

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{11}$$

- The standard notation for the i-th column of I is $\mathbf{e}_i$ which is the vector in $\mathbb{R}^n$ all of whose entries are zero, except that the i-th entry equals 1.

- The $\mathbf{e}_i$ are called the **standard basis vectors** in $\mathbb{R}^n$. The value of n is not part of the notation, it should be clear from the context.
- For any vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{12}$$

  we have

$$\mathbf{x} = I\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i. \tag{13}$$

- Suppose now that we are given a linear function (or transformation) T from $\mathbb{R}^n$ to $\mathbb{R}^m$. Let

$$\mathbf{a}_i = T(\mathbf{e}_i), \quad i = 1, 2, \ldots, n \tag{14}$$

  Moreover, let A be the $m \times n$ matrix

$$A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_n\,] \tag{15}$$

Then

$$
\begin{aligned}
T(\mathbf{x}) &= T\left(\sum_{i=1}^{n} x_i \mathbf{e}_i\right) \\
&= \sum_{i=1}^{n} T\left(x_i \mathbf{e}_i\right) \qquad \text{(by part 1 of linearity!)} \\
&= \sum_{i=1}^{n} x_i T(\mathbf{e}_i) \qquad \text{(by part 2 of linearity!)} \\
&= \sum_{i=1}^{n} x_i \mathbf{a}_i \\
&= A\mathbf{x}.
\end{aligned} \tag{16}
$$

- In other words, with our choice of A,

$$T(\mathbf{x}) = A\mathbf{x} \tag{17}$$

- In this context, A is called the **standard matrix** of the transformation T.

Simplifying things slightly we can say that **linear functions are synonymous with matrices**.

- That's why matrices are so important!
- A set

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \tag{18}$$

of vectors is **linearly independent** if the only way the zero vector can be written as a linear combination of the given vectors is to make all coefficients zero.

- In other words:

$$\sum_{i=1}^{n} c_i \mathbf{v}_i = \mathbf{0} \qquad \Longrightarrow \qquad c_1 = c_2 = \ldots = c_n = 0. \tag{19}$$

- Here are a number of true statements about linear independence. (We assume that all sets contain only vectors from the same space, i.e, they all have the same number of entries.)
  - A set containing just one vector $\mathbf{v}$ is linearly independent if and only if $\mathbf{v}$ is nonzero.
  - A set containing exactly two vectors is linearly independent if and only if neither vector is a multiple of the other.
  - If a set of (more than one) vectors is linearly dependent than at least one of those vectors can be written as a linear combination of the others.
  - No vector in a linearly independent set can be written as a linear combination of the others.
  - A linearly dependent set may (or may not) contain vectors that cannot be written as a linear combination of vectors.

- Any set that contains the zero vector is linearly dependent.
- The matrix transformation

$$\mathbf{y} = A\mathbf{x} \tag{20}$$

  is one-to-one if and only if the columns of A are linearly independent.
- The columns of A are linearly independent if and only if the homogeneous problem

$$A\mathbf{x} = \mathbf{0} \tag{21}$$

  has only the trivial solution $\mathbf{x} = \mathbf{0}$.

- The **span** of a set of vectors is the set of all linear combinations of those vectors.
- Here are some true statements about the span of a set of vectors.
  - Any vector in the span of a linearly independent set can be written in only one way as a linear combination of the given vectors.
  - Any vector in the span of a linearly dependent set can be written in more than one way as a linear combination of the given vectors.
  - The linear system

$$A\mathbf{x} = \mathbf{b} \tag{22}$$

has a solution if and only if $\mathbf{b}$ is in the span of the columns of A.

  - The linear system has a solution for all right hand sides $\mathbf{b}$ if and only if the span of the columns of A is all of $\mathbb{R}^m$.
  - The linear system has a unique solution if $\mathbf{b}$ is in the span of the columns of A and those columns are linearly independent.
  - The linear system has a unique solution for all right hand sides $\mathbf{b}$ if the span of the columns of A is all of $\mathbb{R}^m$ and the columns are linearly independent.

- The linear system $A\mathbf{x} = \mathbf{b}$ is **homogeneous** if $\mathbf{b} = 0$.
- Here are some more true statements about linear systems:
  - if $\mathbf{u}$ and $\mathbf{v}$ are solutions of $A\mathbf{x} = \mathbf{0}$ then so are $\mathbf{u}+\mathbf{v}$ and any other linear combination of $\mathbf{u}$ and $\mathbf{v}$.
  - If $\mathbf{u}$ and $\mathbf{v}$ are solutions of the inhomogeneous system $A\mathbf{x} = \mathbf{b}$ then $\mathbf{u} - \mathbf{v}$ is a solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$.
  - A linear system may have no solutions, a unique solution, or infinitely many solutions. (It may not have precisely 17 solutions, for example.)
  - **The general solution of the linear system**

$$A\mathbf{x} = \mathbf{b} \tag{23}$$

can be written as any particular solution plus the general solution of the homogeneous system

$$A\mathbf{x} = \mathbf{0}. \tag{24}$$

- The last statement is one of the most central principles in mathematics. It applies to all linear problems, not just linear algebraic equations.

Computations

- Given a matrix A you want to understand clearly how to answer the following questions about the linear system $A\mathbf{x} = \mathbf{b}$:
  - Given $\mathbf{b}$ is there at least one solution? In other words, is the system **consistent** for that vector $\mathbf{b}$?
  - Is there a solution for all right hand sides $\mathbf{b}$? In other words, is the system consistent for all possible $\mathbf{b}$? In yet other words, is the matrix transformation $\mathbf{T}(\mathbf{x}) = A\mathbf{x}$ **onto**?
  - If the system is consistent, is there only one solution? If there is we call the solution **unique**. In other words, is the matrix transformation $\mathbf{T}(\mathbf{x}) = A\mathbf{x}$ **one-to-one**?
- There are many ways to compute the answers to these problems. Those we discussed are

based on the reduced and unreduced **row echelon forms** of the **augmented matrix**

$$M = [\, A \quad \mathbf{b} \,] \tag{25}$$

of the linear system $A\mathbf{x} = \mathbf{b}$.

- These are obtained by applying **elementary row operations** to the augmented matrix. There are three such operations:

1. **Add a multiple of one row to another row.**
2. **Interchange two rows.**
3. **Multiply a row by a non-zero constant.**

- It is clear that these operations do not change the solution of the linear system and generate augmented matrices corresponding to **equivalent linear systems.**

- Two linear systems are **equivalent** if they have the same solution sets.

- A rectangular matrix is in **echelon form** if it has the following three properties:

1. **All nonzero rows are above any rows of all zeros.**
2. **Each leading entry of a row is in a column to the right of the leading entry of the row above it.**
3. **All entries in a column below a leading entry are zeros.**

- If a matrix in echelon form satisfies the following additional conditions, then it is in **reduced echelon form**.

4. **The leading entry in each non-zero row is 1.**
5. **Each leading 1 is the only nonzero entry in its column.**

- A **pivot position** in a matrix M is a location in M that corresponds to a leading 1 in the reduced row echelon from of M. A **pivot row** of A is a row that contains a pivot position and a **pivot column** of A is a column that contains a pivot position.

- The **reduced row echelon form** of a matrix is unique. (We will see precisely why this is true when we get to chapter 4.)

- A variable is **basic** if it corresponds to a pivot column.

- Otherwise it is **free**. The free variables can assume any values in the solutions of the linear systems. The values of the basic variables are determined uniquely by those of the free variables.

- Here are some true statements about the solution of linear systems

$$A\mathbf{x} = \mathbf{b} \tag{26}$$

- The linear system has a solution if and only if the last column of the augmented matrix is not a pivot column.
- The linear system has infinitely many solutions only if it has free variables. (For solutions to exist the last column of the augmented matrix still must not be a pivot column.)

- The general solution of the linear system can be written down most easily given the **reduced row echelon form** of the augmented matrix.
- However, some effort can be saved by just computing the (unreduced) row echelon form and then applying backward substitution.
- The computations of row echelon forms in the textbook are written as sequences of matrices, for clarity. Doing the calculations that way involves a lot of copying, however.
- If the purpose of the calculation is strictly the solution of the linear system, rather than theoretical insight, a more streamlined procedure can be used. Everything is written down only once. However, I recommend that to guard against errors you keep track of row sums and compute them redundantly in two ways. As long as they are equal you can be reasonably confident that your calculations so far are accurate.

- Example: Solve the linear system:

$$\begin{array}{ccccccccccc} a & + & b & + & c & & d & = & 2^1 & = & 2 \\ a & + & 2b & + & 4c & + & 8d & = & 2^2 & = & 4 \\ a & + & 3b & + & 9c & + & 27d & = & 2^3 & = & 8 \\ a & + & 4b & + & 16c & + & 64d & = & 2^4 & = & 16 \end{array} \tag{27}$$

- Here is the detailed calculation

$$\begin{array}{ccccccc} \# & a & b & c & d & \textbf{RHS} & \textbf{RS} \\ & & & & & & \\ \mathbf{1} & 1 & 1 & 1 & 1 & 2 & 6 \\ \mathbf{2} & 1 & 2 & 4 & 8 & 4 & 19 \\ \mathbf{3} & 1 & 3 & 9 & 27 & 8 & 48 \\ \mathbf{4} & 1 & 4 & 16 & 64 & 16 & 101 \\ & & & & & & \\ \mathbf{5} = \mathbf{2} - \mathbf{1} & & 1 & 3 & 7 & 2 & 13 \\ \mathbf{6} = \mathbf{3} - \mathbf{1} & & 2 & 8 & 26 & 6 & 42 \\ \mathbf{7} = \mathbf{4} - \mathbf{1} & & 3 & 15 & 63 & 14 & 95 \\ & & & & & & \\ \mathbf{8} = \mathbf{6} - 2 \times \mathbf{5} & & & 2 & 12 & 2 & 16 \\ \mathbf{9} = \mathbf{7} - 3 \times \mathbf{5} & & & 6 & 42 & 8 & 56 \\ & & & & & & \\ \mathbf{10} = \mathbf{9} - 3 \times \mathbf{8} & & & & 6 & 2 & 8 \end{array} \tag{28}$$

$$\begin{array}{ccrcccccc} \mathbf{10} & \Longrightarrow & 6d & = & 2 & \Longrightarrow & d & = & \frac{1}{3} \\ \mathbf{8} & \Longrightarrow & 2c + 4 & = & 2 & \Longrightarrow & c & = & -1 \\ \mathbf{5} & \Longrightarrow & b - 3 + \frac{7}{3} & = & 2 & \Longrightarrow & b & = & \frac{8}{3} \\ \mathbf{1} & \Longrightarrow & a + \frac{8}{3} - 1 + \frac{1}{3} & = & 2 & \Longrightarrow & a & = & 0 \end{array} \tag{29}$$

- We add two $m \times n$ matrices entry by entry, just like we add vectors.
- Similarly, we multiply any matrix by a scalar by multiplying each entry with that scalar.

However, we multiply matrices so that the product of two matrices is the standard matrix of the composition of the corresponding linear functions.

Matrix Multiplication

- The composition $f \circ g$ of two linear functions f and g is linear, and its matrix is the product of the matrices of the constituent functions.

$$\begin{array}{ccccc} & & f \circ g & & \\ \mathbb{R}^p & \xrightarrow[B]{g} & \mathbb{R}^n & \xrightarrow[A]{f} & \mathbb{R}^m \\ & n \times p & & m \times n & \\ & & C = AB & & \\ & & C \text{ is } m \times p & & \end{array} \tag{30}$$

- Note the switch in the sequence. B comes first in the diagram and second in the product, just like g comes first in the diagram and second in the composition.

Six Views of $C = AB$

- We'll look at six different ways of thinking about matrix multiplication. All of them are useful!

- For any matrix A let $\mathbf{r}_i(A)$ denote the i-th row of A, interpreted as a matrix with one row, and let $\mathbf{c}_i(A)$ denote the i-th column, interpreted as matrix with one column. We also identify 1×1 matrices with their single scalar entry.

- Let's illustrate the descriptions with the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}. \tag{31}$$

1. The Formula. Here is what you might find in a textbook or mathematical dictionary: The product of an $m \times n$ matrix A and an $n \times p$ matrix B is an $m \times p$ matrix $C = AB$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \qquad i = 1, \ldots, m, \quad j = 1, \ldots, p. \tag{32}$$

- For our example,

$$
\begin{aligned}
AB &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 1 \times 2 + 2 \times 1 & 1 \times 1 + 2 \times 2 \\ 2 \times 3 + 1 \times 4 & 1 \times 3 + 2 \times 4 \end{bmatrix} \\
&= \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}.
\end{aligned}
\tag{33}
$$

2. Writing it. We saw in class that it is advantageous to write the second factor to the upper right of the first factor. The product fits into the corner made by the two factors, and the $i-j$ entry of C sits at the intersection of the i-th row of A and the j-th column of B. In our example we get

$$
\begin{array}{cc}
 & \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\
\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} & \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}
\end{array}
\tag{34}
$$

- More generally, we get:

$$
\begin{array}{cc}
 & B \quad n \times p \\
 & \begin{bmatrix} b_{11} & \dots & b_{1j} & \dots & b_{1p} \\ \vdots & & \vdots & & \vdots \\ b_{i1} & \dots & b_{ij} & \dots & b_{ip} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \dots & b_{nj} & \dots & b_{np} \end{bmatrix} \\
\begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix} & \begin{bmatrix} & & & & \\ & & \vdots & & \\ & \dots & c_{ij} & \dots & \\ & & \vdots & & \\ & & & & \end{bmatrix} \\
A \quad m \times n & C = AB \quad m \times p
\end{array}
\tag{35}
$$

It is evident from this picture that

- the $i - j$ entry of C is the (dot) product of the i-th row of A and the j-th column of B,
- the j-th column of C is the product of A and the j-th column of B,
- the i-th row of C is the product of the i-th row of A and B.

3. The entry by entry view.

$$c_{ij} = \mathbf{r}_i(A)\mathbf{c}_j(B). \tag{36}$$

In our example

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} & \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}. \end{aligned} \tag{37}$$

4. The Column View. This is actually how we first derived our formula for matrix multiplication: the j-th column of C equals A multiplied with the j-column of B. As a formula:

$$\mathbf{c}_j(C) = A\mathbf{c}_j(B), \qquad j = 1, \ldots, p. \tag{38}$$

In our example:

$$\begin{aligned} C &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} A \begin{bmatrix} 2 \\ 1 \end{bmatrix} & A \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}. \end{aligned} \tag{39}$$

- Note that in this view every column of the product is a linear combination of the columns of A. The coefficients of the linear combination are in the corresponding column of B.

5. **The Row View**. The i-th row of C is the i-th row of A multiplied with B:

$$r_i(C) = r_i(A)B. \tag{40}$$

In our example:

$$\begin{aligned} C &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} B \\ \begin{bmatrix} 3 & 4 \end{bmatrix} B \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix} \end{aligned} \tag{41}$$

- Note that in this view every row of the product is a linear combination of the rows of B. The coefficients of the linear combination are in the corresponding row of A.

6. **The matrix view**. Note that the product of the k-th column of A and the k-th row of B is an $m \times p$ matrix, the product of an $m \times 1$ matrix and a $1 \times p$ matrix. The $i - j$ entry of $\mathbf{c}_k(A)\mathbf{r}_k(B)$ is $a_{ik}b_{kj}$. So we get, by our formula

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}, \tag{42}$$

that

$$C = \sum_{k=1}^{n} \mathbf{c}_k(A)\mathbf{r}_k(B). \tag{43}$$

In our example

$$\begin{aligned} C &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix} \end{aligned} \tag{44}$$

in general, the product of an $m \times 1$ matrix A and a $1 \times p$ matrix B is an $m \times p$ matrix C which has rank 1. Every row of C is a multiple of A and every column of C is a multiple of B.

The Inverse of a Matrix

- A square $n \times n$ matrix A is said to be **invertible** if there is an $n \times n$ matrix C such that
$$CA = I \quad \text{and} \quad AC = I \tag{45}$$
where $I = I_n$, the $n \times n$ identity matrix.
- If A is not invertible it is said to be **singular**.
- the terms invertible and singular only apply to square matrices.
- If A is invertible its inverse is unique.
- There are many equivalent properties of a square matrix that are equivalent to invertibility. Some of them are expressed in what the textbook calls the

Invertibility Theorem Let A be a square $n \times n$ matrix. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

a. A is an invertible matrix.

b. A is row equivalent to the identity matrix.

c. A has n pivot positions.

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

e. The columns of A form a linearly independent set.

f. The linear transformation $\mathbf{x} \longrightarrow A\mathbf{x}$ is one-to-one.

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

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

i. The linear transformation maps $\mathbb{R}^b$ onto $\mathbb{R}^n$.

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

k. There is an $n \times n$ matrix D such that $CA = I$. (Of course, the left and right inverses C and D are actually equal.)

l. A^T is an invertible matrix.

m. $\text{Col}A = \mathbb{R}^n$.

n. $\dim \text{Col}A = n$.

p. $\text{rank}A = n$.

q. $\text{Nul}A = \{\mathbf{0}\}$.

r. $\dim \text{Nul}A = 0$.

More Properties of Inverse Matrices

- If A is invertible and $A\mathbf{x} = \mathbf{b}$ then $\mathbf{x} = A^{-1}\mathbf{b}$.
- If A is invertible and $AB = C$ then $B = A^{-1}C$.
- If A is invertible and $BA = C$ then $B = CA^{-1}$.
- The process of inverting and transposing a matrix commute:

$$\left(A^T\right)^{-1} = \left(A^{-1}\right)^T. \tag{46}$$

- Assuming A and B are invertible and have the same size,

$$(AB)^{-1} = B^{-1}A^{-1} \tag{47}$$

In general

$$B^{-1}A^{-1} \neq A^{-1}B^{-1} \tag{48}$$

since matrix multiplication does not commute.

Partitioned Matrices

- A **partitioned matrix**, or **block matrix**, is a matrix whose entries are matrices. Block matrices have many useful properties but we discussed only one

Major Principle: Provided the partitions are conformable, multiplying block matrices works exactly like multiplying matrices.

- As an application, we computed the inverse of a 2×2 block matrix:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{49}$$

where A_{11} is $p \times p$, A_{22} is $q \times q$, and $p + q = n$. Of course, A_{12} is $p \times q$ and A_{21} is $q \times p$.

Matrix Factorizations

- A **factorization** of a matrix A is obtained by writing A as a product of several (usually 2 or 3) matrices.
- In particular we discussed the LU factorization of a square matrix

$$A = LU \tag{50}$$

where

L is **unit lower triangular**, i.e.,

$$l_{ii} = 1, \quad i = 1, 2, \dots n \quad \text{and} \quad j < i \quad \Longrightarrow \quad a_{ij} = 0 \tag{51}$$

U is **upper triangular**, i.e.,

$$l_{ii} = 1, \quad i = 1, 2, \dots n \quad \text{and} \quad j > i \quad \Longrightarrow \quad a_{ij} = 0 \tag{52}$$

- In other words, denoting possibly non-zero entries by x, L and U are of the form:

$$L = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ x & 1 & 0 & \dots & 0 & 0 \\ x & x & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x & x & x & \dots & 1 & 0 \\ x & x & x & \dots & x & 1 \end{bmatrix} \tag{53}$$

and

$$U = \begin{bmatrix} x & x & x & \dots & x & x \\ 0 & x & x & \dots & x & x \\ 0 & 0 & x & \dots & x & x \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & x & x \\ 0 & 0 & 0 & \dots & 0 & x \end{bmatrix} \tag{54}$$

- A major application of the LU factorization is to solve the linear system $A\mathbf{x} = \mathbf{b}$ by solving the two triangular systems

$$L\mathbf{y} = \mathbf{b} \quad \text{and} \quad U\mathbf{x} = y. \tag{55}$$

Computing the LU factorization is equivalent to applying row operations to convert A to upper triangular form.

- Frequently we need to interchange rows during Gaussian Elimination. The process is called **pivoting**. Pivoting can be expressed in terms of a **permutation matrix** which is a matrix that has been obtained from the identity matrix by permuting its rows or columns.

It usually is a bad idea actually to compute an inverse matrix.

Subspaces of $\mathbb{R}^n$

- a **subspace of $\mathbb{R}^n$** is a non-empty subset of $\mathbb{R}^n$ that is closed under addition and scalar multiplication.
- Examples of subspaces of $\mathbb{R}^n$ include
  - The set $\{\mathbf{0}\}$,
  - A line through the origin,
  - A plane containing the origin,
  - The span of any set of vectors in $\mathbb{R}^n$,
  - $\mathbb{R}^n$ itself,
  - The column space of a matrix,
  - the null space or kernel of a matrix.
- Suppose A is an $m \times n$ matrix. Then its **column space** is the span of its set of columns. In other words,

$$\mathrm{Col}A = \{\mathbf{y} : \mathbf{y} = A\mathbf{x}\}. \tag{56}$$

  It is a subspace of $\mathbb{R}^m$.

- The **null space** or **kernel** of A is the set of all solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. In other words:

$$\mathrm{Nul}A = \{\mathbf{x} : A\mathbf{x} = \mathbf{0}\}. \tag{57}$$

  It is a subspace of $\mathbb{R}^n$.

- A **spanning set** β of a subspace H is a set of vectors in H which is such that every vector in H can be written as a linear combination of vectors in β.

- A **basis** of a subspace is a linearly independent spanning set of that subspace.
- All bases of a given subspace H have the same number of vectors. That number is the **dimension** of H.
- The **rank** of a matrix is the dimension of its column space. It equals the number of pivots (rows or columns).
- The dimension of the column space of an $m \times n$ matrix is

$$\dim \mathrm{Col} A = \mathrm{rank} A \tag{58}$$

  The pivot columns of A form a basis of its column space.
- The dimension of the kernel of an $m \times n$ matrix A is the number of free variables in $A\mathbf{x} = \mathbf{0}$. It is given by

$$\dim \mathrm{Nul} A = n - \mathrm{rank} A \tag{59}$$

- To obtain a basis for the kernel obtain one basis vector for each free variable, by setting that variable equal to 1 and the other free variables to zero. In each case, compute the basic variables from the equation $A\mathbf{x} = \mathbf{0}$ using (reduced) row echelon form of A.

- Any spanning set of p vectors in a p-dimensional space is linearly independent, i.e., it's a basis.
- Any linearly independent set of p vectors in a p-dimensional space is a spanning set, i.e., it's a basis.
- When expressing a vector $\mathbf{v}$ as a linear combination of basis vectors the coefficients of the linear combination are sometimes called the **coordinates** of $\mathbf{v}$ with respect to the given basis.

Chapter 3: Determinants

- Determinants are numbers associated with **square** matrices. Throughout these notes let A be an $n \times n$ square matrix.
- We denote the determinant of A by $|A|$ or $\det A$. Note that in this context the vertical bars do not mean absolute values.
- The determinant of a 1×1 matrix equals its unique entry.
- For a 2×2 matrix we define

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}. \qquad (60)$$

- For $n > 2$ the determinant of A is defined recursively. Suppose A is an $n \times n$ matrix where $n > 2$. We define A_{ij} to be the $(n-1)\times$

$(n-1)$ matrix obtained from A by removing the i-th row and the j-th column.

- For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [\, a_{ij} \,]$ is

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \ldots + (-1)^{1+n} a_{1n} \det A_{1n} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \end{aligned} . \tag{61}$$

However, a corresponding expansion gives the same numerical value for **any row or column**.

- The corresponding formula is usually expressed in terms of **cofactors**:

$$C_{ij} = (-1)^{i+j} \det A_{ij}. \tag{62}$$

- The factor

$$(-1)^{i+j} \tag{63}$$

creates the familiar checkerboard pattern

$$\begin{bmatrix} + & - & + & - & \dots \\ - & + & - & + & \dots \\ + & - & + & - & \dots \\ - & + & - & + & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{64}$$

- **Theorem 1** on page on page 168 in the textbook gives the general formulas

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij} \tag{65}$$

for any choice of i and

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij} \tag{66}$$

for any choice of j.

- There is no proof of this fact in the textbook. One way to see this, with a bit of work, is to show that formula (61) implies that

$$\det A = \sum_{\sigma} \operatorname{sign}(\sigma) \prod_{i=1}^{n} a_{i\sigma_i} \tag{67}$$

where the sum goes over all $n!$ permutations of the set $\{1, 2, \ldots, n\}$ and the large symbol π indicates the product of n factors, one from each row i, and the column σ_i.

- The formula (67) is symmetric in the rows and columns, and so any row or column can be used to compute the determinant.

- Clearly, in the cofactor expansion (65) or (66) you want to pick rows or columns that contain many zeros.

- An extreme case of the exploitation of zero entries is provided by triangular matrices. **The determinant of a triangular matrix is the product of the diagonal entries.** (This is **Theorem 2** on page 169 of the textbook.)

- The formula (67) gives the determinant as the sum of $n!$ terms, each of which has n factors. Computing the determinant by that formula is prohibitively expensive even for small values on n.

- Determinants can be computed much more efficiently by row operations. The relevant facts are contained in Theorem 3 on page 171 of the textbook:

Theorem 3. (Row Operations) Let A be a square matrix. Then

a. If a multiple of a row of A is added to another row to produce a matrix B, then

$$\det B = \det A. \tag{68}$$

b. If two rows of A are interchanged to produce B then

$$\det B = -\det A. \tag{69}$$

c. If one row of A is multiplied by a scalar k to produce B then

$$\det B = k \det A. \tag{70}$$

- All of these statements can be proved by observing that they are true for $n = 2$ and then using induction based on the cofactor expansion.

Theorem 4, page 173.

s. A square matrix A is invertible if and only if

$$\det A \neq 0. \tag{71}$$

- This can be seen by reducing the matrix to row echelon form and applying Theorem 2 on triangular matrices.
- (The Label **s.** indicates the position in the invertible matrix theorem.)

Theorem 5, page 174. If A is a square matrix then

$$\det A = \det A^T. \tag{72}$$

- This is obvious by the fact that any row or column can be used for the computation of the determinant.

Theorem 6, page 175. If A and B are both $n \times n$ matrices then

$$\det(AB) = \det A \times \det B. \tag{73}$$

- We saw that this is true by observing that the statement holds for elementary matrices (those implementing row operations) and writing one of the matrices as a product of elementary matrices and the identity matrix.

Linearity: The determinant is a linear function of each row or column separately.

- This follows immediately from the cofactor expansion.
- **Cramer's Rule** rule states that for the linear system
$$A\mathbf{x} = \mathbf{b} \tag{74}$$
the i-th entry of $\mathbf{x}$ is given by
$$x_i = \frac{|A_i(\mathbf{b})|}{|A|} \tag{75}$$
where $A_i(\mathbf{b})$ is the matrix formed by replacing the i-th column of A with $\mathbf{b}$.
- To see that this is true let I_i be the matrix obtained from the identity matrix by replacing the i-th column with $\mathbf{x}$. Thus
$$I_i = [\, \mathbf{e}_1 \quad \mathbf{e}_2 \quad \ldots \mathbf{e}_{i-1} \quad \mathbf{x} \quad \mathbf{e}_{i+1} \quad \ldots \quad \mathbf{e}_n \,] \tag{76}$$
Then, by the way we defined matrix multiplication,
$$AI_i = A_i. \tag{77}$$
The determinant of the product equals the product of the determinants:
$$|A||I_i| = |A_i|, \quad \text{i.e.,} \quad |I_i| = \frac{|A_i|}{|A|}. \tag{78}$$

Moreover, by expanding about the i-th row we see that

$$|I_i| = x_i. \tag{79}$$

Cramer's rule follows.

- In particular, Cramer's Rule gives a formula for A^{-1}

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix} \tag{80}$$

- The matrix

$$\mathrm{adj} A = \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix} \tag{81}$$

is called the **adjugate** of A.

note that the adjugate is the **transpose** of the matrix of cofactors!

- The determinant can be interpreted geometrically as the volume V of a parallelotope defined by the columns of A:

$$V = |\det A| \tag{82}$$

where the vertical bars in this case do denote the absolute values.

Chapter 4: Vector Spaces

- **Definition:** A **vector space**[−1−] is a nonempty set V of objects, called vectors, on which are defined two operations, called **addition** and **multiplication by scalars (real numbers)**, subject to the ten axioms (or rules) listed below. The axioms must hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V, and for all scalars c and d.

1. The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u}+\mathbf{v}$, is in V.

2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

3. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.

4. There is a zero vector $\mathbf{0}$ in V such that

$$\mathbf{u}+\mathbf{0}=\mathbf{u}. \tag{83}$$

5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V such that $u+(-\mathbf{u})=\mathbf{0}$.

6. The scalar multiple of $\mathbf{u}$ by c, denoted by $c\mathbf{u}$, is in V.

7. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$.

8. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$.

9. $c(d\mathbf{u})=(cd)\mathbf{u}$.

10. $1\mathbf{u}=\mathbf{u}$.

−1− Also called a **linear space**

- A **subspace** of a vector space V is a non-empty subset of V that is closed under addition and scalar multiplication.

every subspace is a vector space itself.

- Examples of vector spaces:
  - The primary examples of vector spaces are of course $\mathbb{R}^n$ and subspaces of $\mathbb{R}^n$.
  - The column space of a matrix.
  - The null space of a matrix.
  - The column space of A^T (called the **row space** of A).
  - The null space of A^T, i.e., the set of all $\mathbf{x}$ such that
$$A^T x = \mathbf{0}. \tag{84}$$
  - The set of all quadratic polynomials.
  - The set of all polynomials of degree n
  - The set of all polynomials.
  - The set of all real valued functions defined on some set (domain).
  - The set of all functions that are continuous on $[a, b]$, usually denoted by $C^0[a, b]$ or $C[a, b]$.
  - The set of all functions that are square integrable on $\mathbb{R}$:
$$V = \left\{ f : \int_{-\infty}^{\infty} f^2(x) \mathrm{d}x < \infty \right\}. \tag{85}$$

- The set of all solutions of the differential equation
$$y'' = k^2 y \tag{86}$$
- The set of all $m \times n$ matrices.
- The set of all upper triangular $n \times n$ matrices.
- The set of all diagonal matrices.
- The set of all symmetric $n \times n$ matrices (those that satisfy $A = A^T$.)
- The set of all sequences
$$x_0, x_1, x_2, x_3, \dots \tag{87}$$
- The set of all sequences $x_0, x_1, x_2, \dots$ that satisfy the infinitely many equations
$$x_{n+2} - x_{n+1} - x_n = 0, \qquad n = 0, 1, 2, \dots. \tag{88}$$
- The set of all convergent sequences.
- The range of a linear transformation
- The null space of a linear transformation

Here are some examples of sets that are **not** vector spaces:

- A line or plane in $\mathbb{R}^n$ not containing the origin.
- The set of all triangular matrices.

- The set of all non-singular (square) matrices
- The set of all singular (square) matrices.
- The set of all sequences $x_0, x_1, x_2, \ldots$ that satisfy the infinitely many equations

$$x_{n+2} - x_{n+1} - x_n = 1, \qquad n = 0, 1, 2, \ldots. \tag{89}$$

- The set of all divergent sequences.
- The solution set of a linear system $A\mathbf{x} = \mathbf{b}$ (unless $\mathbf{b} = \mathbf{0}$).

- A **linear combination** of a (finite) set of vectors is obtained by multiplying each vector with some scalar and adding up the products.
- The **span** of a set of vectors is the set of all linear combinations of those vectors.
- A **spanning set of a vector space** is a subset of the vector space whose span is the space.
- **linear independence of a set of vectors** means that the only way to get the zero vector as a linear combination of the vectors is by picking all coefficients equal to zero.
- A **basis of a vector space** is a linearly independent spanning set of the space.
- All bases of a specific vector space have the same number of elements. That number is the **dimension** of the vector space.

- We saw that this is true by showing that if $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ is a basis with k elements than any set of k elements is linearly dependent. To do that we expressed every vector in the larger set in terms of the basis, and obtained a homogeneous rectangular matrix problem that was certain to have a non-trivial solution.

- Two vector spaces V and W are **isomorphic** if there is a linear transformation from V to W that is one-to-one and onto.

- Two isomorphic vector spaces have the same structure. Essentially they are the same. They differ only in notation or interpretation. As my linear algebra teacher said long ago, one space is painted green, the other is painted red.

- An isomorphism is invertible!

- Given a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of V and a basis $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ of W an isomorphism C can be defined by

$$C\left(\sum_{i=1}^{n} \alpha_i \mathbf{v}_i\right) = \sum_{i=1}^{n} \alpha_i \mathbf{w}_i. \tag{90}$$

- Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.

In particular, all n-dimensional vector spaces are isomorphic to $\mathbb{R}^n$.

- Thus in a sense the only finite dimensional vector spaces are $\mathbb{R}^n$ for $n = 1, 2, 3, \ldots$.
- Suppose

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\} \tag{91}$$

  is a basis of a vector space V and

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{b}_i \tag{92}$$

  is a vector in V. Then the vector

$$[\mathbf{x}]_{\mathrm{V}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \tag{93}$$

  is the **coordinate vector of x with respect to the basis V.**

- We can convert between bases. Suppose we have three bases of $\mathbb{R}^n$.

$$\begin{aligned} \mathrm{I} &= \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\} \\ \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} &= \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\} \\ \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix} &= \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\} \end{aligned} \tag{94}$$

- I is the **standard basis**.
- As usual, we associate the matrices

$$
\begin{aligned}
B &= [\,\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_n\,] \\
\text{and} & \\
C &= [\,\mathbf{c}_1 \quad \mathbf{c}_2 \quad \dots \quad \mathbf{c}_n\,]
\end{aligned}
\tag{95}
$$

  with the bases $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}$.

- B and C are square and invertible.
- A vector

$$
\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{96}
$$

  can be expressed variously as

$$
\mathbf{x} = [\mathbf{x}]_{\mathbf{I}} = B[\mathbf{x}]_{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}} = C[\mathbf{x}]_{\mathbf{c}} \tag{97}
$$

- It follows that

$$
[\mathbf{x}]_{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}} = B^{-1}\mathbf{x} \quad \text{and} \quad [\mathbf{x}]_{\begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}} = C^{-1}\mathbf{x}. \tag{98}
$$

- We can convert between the bases $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}$ by the formulas

$$[\mathbf{x}]_{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}} = B^{-1}C[\mathbf{x}]_{\begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}} \tag{99}$$

and

$$[\mathbf{x}]_{\begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}} = C^{-1}B[\mathbf{x}]_{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}}. \tag{100}$$

- Suppose A is an $m \times n$ matrix. It defines a linear transformation

$$\mathbf{y} = A\mathbf{x} \tag{101}$$

from $\mathbb{R}^n$ to $\mathbb{R}^m$. Suppose we want to express **the same linear transform** in terms of a basis

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\} \tag{102}$$

of $\mathbb{R}^n$ and a basis

$$\begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m\} \tag{103}$$

of $\mathbb{R}^m$.

- In other words, we want to find a matrix T such that

$$[\mathbf{y}]_{\begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix}} = T[\mathbf{x}]_{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}} \tag{104}$$

- The situation is illustrated in this "commuting diagram":

$$\begin{array}{cccccc}
 & \mathbf{x} & \longrightarrow & A\mathbf{x} & \\
B\mathbf{x} & \mathbb{R}^n & \longrightarrow & \mathbb{R}^m & \mathbf{x} \\
 & & & & \\
\Big\uparrow & \Big\uparrow & & \Big\downarrow & \Big\downarrow \\
 & & & & \\
\mathbf{x} & \mathbb{R}^n & \longrightarrow & \mathbb{R}^m & C^{-1}\mathbf{x} \\
 & \mathbf{x} & \longrightarrow & T\mathbf{x} &
\end{array} \tag{105}$$

- start in the lower left corner. Move to the lower right corner either by going directly to the right, or in three steps by going up, right, and then down. We want T to be such that in either way we get to the same vector.

- Clearly,

$$T = C^{-1}AB. \tag{106}$$

- By the same token,

$$A = CTB^{-1}. \tag{107}$$

- check the dimensions.

In the special case that $m = n$ and $B = C$ we get that

$$T = B^{-1}AB. \tag{108}$$

- In this case A and B are said to be similar, and the formula (or the matrix B) is called a **similarity transform**.

6. Eigenvalues and Eigenvectors

- Unless stated otherwise, in this section A is a real square $n \times n$ matrix.
- An **eigenvector** of a square ($n \times n$) matrix A is a **non-zero** vector $\mathbf{x}$ such that

$$A\mathbf{x} = \lambda\mathbf{x} \tag{109}$$

  for some scalar λ. λ is called the **eigenvalue** of A corresponding to the eigenvector $\mathbf{x}$. $\mathbf{x}$ is an eigenvector corresponding to the eigenvalue λ.

- The pair $(\lambda, \mathbf{x})$ is sometimes called an **eigenpair** of A.

Note than any non-zero scalar multiple of an eigenvector is also an eigenvector, with the same eigenvalue.

the main difference between linear systems and eigenvalue problems is that eigenvalue problems are **nonlinear**!

- More insight can be gained by writing

$$Ax = \lambda \mathbf{x} \tag{110}$$

as

$$A\mathbf{x} - \lambda x = (A - \lambda I)\mathbf{x} = \mathbf{0}. \tag{111}$$

- Any eigenvector is a non-trivial solution of the homogeneous linear system

$$(A - \lambda I)\mathbf{x} = 0. \tag{112}$$

- Every eigenvector is in the nullspace of $A - \lambda I$.
- Every non-zero vector in the nullspace of $A - \lambda I$ is an eigenvector of A.
- A square homogeneous linear system has a non-trivial solution if and only if the coefficient matrix is singular.

thus λ is an eigenvalue of A if and only if $A - \lambda I$ is singular.

Upshot: we have one more characterization of singularity. A square matrix A is singular if and only if 0 is an eigenvalue of A. It is invertible if and only if all eigenvalues of A are non-zero.

- Suppose $\mathbf{x}_i, i = 1, \ldots, m$ are eigenvectors corresponding to the same eigenvalue λ. Then any (non-zero) linear combination of the eigenvectors is also an eigenvector:

$$A \sum_{i=1}^{m} \alpha_i \mathbf{x}_i = \sum_{i=1}^{n} \alpha_i A \mathbf{x}_i = \sum_{i=1}^{n} \alpha_i \lambda \mathbf{x}_i = \lambda \sum_{i=1}^{n} \alpha_i \mathbf{x}_i. \tag{113}$$

- Thus, if we add the zero vector to the set of eigenvectors corresponding to a specific eigenvalue, that set is a linear space, the nullspace of $A-\lambda I$. That space is also called the **eigenspace of A corresponding to λ**.

- Important example: The eigenvalues of a **triangular** matrix are the **diagonal entries**, because if A is triangular and λ is an eigenvalue then $A - \lambda I$ is a triangular matrix with at least one zero entry on the diagonal. It is thus singular.

Row operations do not preserve eigenvalues or eigenvectors!

- A matrix is singular if and only if its determinant is zero. Thus we get the key result:

$$\lambda \text{ is an e.v.} \quad \Longleftrightarrow \quad \det(A - \lambda I) = 0. \tag{114}$$

- The equation

$$\det(A - \lambda I) = 0. \tag{115}$$

  is the **characteristic equation** of A.

- The function $f(\lambda) = |A - \lambda I|$ is a polynomial of degree n with leading coefficient $(-1)^n$

- We can see this using a cofactor expansion or the formula

$$\det A = \sum_{\sigma} \operatorname{sign}(\sigma) \prod_{i=1}^{n} a_{i\sigma_i} \tag{116}$$

  where the sum goes over all $n!$ permutations of the set $\{1, 2, \ldots, n\}$ and the large symbol π indicates the product of n factors, one from each row i, and the column σ_i.

- The polynomial

$$p(\lambda) = \det(A - \lambda I) \tag{117}$$

  is the **characteristic polynomial** of A

The eigenvalues of A are the **roots** of the characteristic polynomial.

- This means

1. There are precisely n of them, properly counting multiplicity.

2. They may be repeated.

3. They may be complex.

4. If there are complex eigenvalues then they occur in conjugate complex pairs.

- The natural way to compute eigenvalues and eigenvectors by hand proceeds in two steps:

1. Compute the characteristic polynomial and find its roots.

2. For each eigenvalue λ find the nullspace of $A - \lambda I$.

- This works well only for small matrices with exactly known entries.

- However, the opposite process, computing roots of polynomials by computing the eigenvalues of a suitable matrix works very well.

- For every polynomial p of degree n with leading term $(-1)^n$ there exists a matrix A whose characteristic polynomial is p. Check:

$$\det\left(\begin{bmatrix} \alpha_{n-1} & \alpha_{n-2} & \cdots & \alpha_1 & \alpha_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} - \lambda I\right) = (-1)^n \left[\lambda^n - \sum_{j=0}^{n-1} \alpha_j \lambda^j\right]. \tag{118}$$

- Central in eigenvalue calculations is the concept of **similarity**.

- **Definition:** Two matrices A and B are **similar** it there is a non-singular matrix P such that
$$B = P^{-1}AP. \tag{119}$$

- Similar matrices have the same eigenvalues, and their eigenvectors are related in a straightforward way. To see this suppose that
$$A\mathbf{x} = \lambda x \tag{120}$$
and note that
$$B(P^{-1}\mathbf{x}) = P^{-1}APP^{-1}\mathbf{x} = P^{-1}\lambda\mathbf{x} = \lambda(P^{-1}\mathbf{x}). \tag{121}$$

- In other words, the eigenvectors of B are those of A, multiplied with P^{-1}.

- Another way to see that similar matrices have the same eigenvalues is to observe that their characteristic polynomials are the same. Using the multiplicative property of determinants and the fact that the determinant of the inverse is the reciprocal of the determinant of the original matrix we see
$$\begin{aligned} |B - \lambda I| &= |P^{-1}AP - \lambda P^{-1}IP| \\ &= |P^{-1}(A - \lambda I)P| \\ &= |P^{-1}||A - \lambda I||P| \\ &= |A - \lambda I| \end{aligned} \tag{122}$$

- **Definition:** A matrix is **diagonalizable** if it is similar to a diagonal matrix.
- In other words, A is diagonalizable if there exists a diagonal matrix D and a non-singular matrix P such that

$$D = P^{-1}AP. \tag{123}$$

- This equation can be rewritten as

$$AP = PD. \tag{124}$$

- Suppose

$$P = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n\,] \tag{125}$$

and

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{126}$$

- Note that the equation for the i-th column in (124) is precisely the eigenvector equation

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i. \tag{127}$$

A matrix is diagonalizable if and only if it has n linearly independent eigenvectors. The similarity transform to diagonal form is the matrix of eigenvectors and the similar diagonal matrix has the eigenvalues along the diagonal.

- A matrix that is not diagonalizable is called **defective**.
- A matrix is not defective if an only if it has a set of n linearly independent eigenvectors.

Invertibility is unrelated to Diagonalizability.

$$\begin{array}{rcc} & \textbf{singular} & \textbf{invertible} \\ \textbf{defective:} & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\ \textbf{diagonalizable:} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array} \tag{128}$$

- It is sometimes useful to be able to construct a matrix with given eigenvalues and eigenvectors. Note that

$$D = P^{-1}AP \tag{129}$$

is equivalent to

$$A = PDP^{-1}. \tag{130}$$

Suppose you want to construct a matrix A with given eigenvalues and given eigenvectors. Proceed as follows:

1. Collect the eigenvectors into the matrix P as before.
2. Compute P^{-1}.
3. Compute
$$A = PDP^{-1}. \tag{131}$$

- It is not always possible to diagonalize a matrix. However, for all matrices A there exists a similarity transform to its **Jordan Canonical Form**[-2-] (named after Camille Jordan, 1838-1922).
- The JCF is a block diagonal matrix

$$P^{-1}AP = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_k \end{bmatrix} \tag{132}$$

[-2-] The textbook mentions the Jordan Canonical Form in a footnote on page 294.

where each diagonal block is of the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_i & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda_i & \dots & 0 & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_i \end{bmatrix} \tag{133}$$

- Apart from reordering the diagonal blocks the JCF is unique.
- Each Jordan block J_i corresponds to one eigenvector with eigenvalue λ_i.
- A matrix is diagonalizable if and only if all of its Jordan blocks are 1×1.
- The **algebraic multiplicity** of an eigenvalue is its order as a root of the characteristic polynomial. Its **geometric multiplicity** is the dimension of its eigenspace.
- Here is an example. Suppose the Jordan form

of a matrix is given by

$$J = \begin{bmatrix} 2 & 1 & . & . & . & . & . & . & . & . \\ . & 2 & . & . & . & . & . & . & . & . \\ . & . & 2 & . & . & . & . & . & . & . \\ . & . & . & 3 & 1 & . & . & . & . & . \\ . & . & . & . & 3 & 1 & . & . & . & . \\ . & . & . & . & . & 3 & . & . & . & . \\ . & . & . & . & . & . & 3 & . & . & . \\ . & . & . & . & . & . & . & 4 & . & . \\ . & . & . & . & . & . & . & . & 4 & . \\ . & . & . & . & . & . & . & . & . & 5 \end{bmatrix} \tag{134}$$

- Entries indicated by dots are zero.
- The characteristic polynomial of this matrix is

  $$p(\lambda) = |J - \lambda I| = (2-\lambda)^3(3-\lambda)^4(4-\lambda)^2(5-\lambda). \tag{135}$$

  The number 2 is an eigenvalue of algebraic multiplicity 3 and geometric multiplicity 2, 3 is an eigenvalue of algebraic multiplicity 3 and geometric multiplicity 2, 4 is an eigenvalue of algebraic and geometric multiplicity 2, and 5 is an eigenvalue of algebraic and geometric multiplicity 1. The dimension of the space spanned by all eigenvectors is the sum of the geometric multiplicities which is 7. The matrix is defective.
- A set of eigenvectors corresponding to distinct eigenvalues is linearly independent. (The word

"distinct" means that no two of the eigenvalues are equal.)

- Recall that a matrix is diagonalizable if it has a set of n linearly independent eigenvectors.

- Thus a matrix with distinct eigenvalues is diagonalizable.

- This implies, for example, that the JCF can be computed only in exact arithmetic.

- A non-diagonalizable matrix must have multiple eigenvalues.

The most important thing to know about complex eigenvalues is that **symmetric real matrices don't have any!** The textbook addresses this issue in problem 24 on page 303 (and later in chapter 7).

- But the argument is quite simple.

- For any matrix A or vector $\mathbf{x}$ let

$$A^H = \bar{A}^T \quad \text{and} \quad \mathbf{x}^H = \bar{\mathbf{x}}^T \tag{136}$$

  where the bar denotes conjugate complex.

- A complex matrix A is **Hermitian**[3] if

$$A = \bar{A}^T. \tag{137}$$

[3] named after Charles Hermite, 1822–1901.

- We will show that the eigenvalues of a Hermitian matrix are real.

Note that symmetric real matrices are special cases of Hermitian matrices.

- Suppose

$$A\mathbf{x} = \lambda \mathbf{x} \tag{138}$$

where $A = A^H$, and A, λ, and $\mathbf{x}$ are all possibly complex. Taking the conjugate complex on both sides turns this into

$$\mathbf{x}^H A^H = x^H A = \bar{\lambda}\mathbf{x}^H. \tag{139}$$

Left multiplying with $\mathbf{x}^H$ in (138) and right multiplying with $\mathbf{x}$ in (139) gives

$$\mathbf{x}^H A\mathbf{x} = \lambda \mathbf{x}^H \mathbf{x} \quad \text{and} \quad \mathbf{x}^H A\mathbf{x} = \bar{\lambda}\mathbf{x}^H \mathbf{x}. \tag{140}$$

Thus

$$\lambda \mathbf{x}^H \mathbf{x} = \bar{\lambda}\mathbf{x}^H \mathbf{x}. \tag{141}$$

This implies that $\lambda = \bar{\lambda}$, i.e., λ is real.

- **Gershgorin Theorem.** Suppose A is an $n \times n$ matrix, and λ is one of its eigenvalues. Then, for some $i \in \{1, 2, \ldots, n\}$

$$|a_{ii} - \lambda| \leq \sum_{j \neq i} |a_{ij}|. \tag{142}$$

- In other words, every eigenvalue lies in some circle whose center is a diagonal entry of A, and whose radius equals the sum of the absolute values of the off-diagonal entries in that row.

- Those circles are referred to as the **Gershgorin Circles**.
- To see this suppose $\mathbf{x}$ is an eigenvector of the $n \times n$ matrix A, with corresponding eigenvalue λ. Thus
$$A\mathbf{x} = \lambda \mathbf{x}. \tag{143}$$
- Since an eigenvector is determined only up to a non-zero factor we may assume that $\mathbf{x}$ is normalized such that
$$\max_{j=1,\ldots,n} |x_j| = x_i = 1 \tag{144}$$
for some i in $\{1, 2, \ldots, n\}$. This fixes i. If there are several such indices i we pick any particular one of them.
- The i-th component of the vector equation (144) is
$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i = \lambda. \tag{145}$$
- Subtracting $a_{ii} x_i = a_{ii}$ on both sides gives the equation
$$\lambda - a_{ii} = \sum_{j \neq i} a_{ij} x_j \tag{146}$$

- Taking absolute values on both sides, applying the triangle inequality, and observing that $|x_j| \le 1$ for all j shows that λ lies in the Gershgorin Circle centered at x_i:

$$\begin{aligned} |\lambda - a_{ii}| &= \left| \sum_{j \ne i} a_{ij} x_j \right| \\ &\le \sum_{j \ne i} |a_{ij} x_j| \\ &= \sum_{j \ne i} |a_{ij}||x_j| \\ &\le \sum_{j \ne i} |a_{ij}| \end{aligned} \tag{147}$$

It's not true in general that every Gershgorin Circle contains an eigenvalue.

On the other hand, it is true that any union of k Gershgorin Circles that does not overlap with any of the remaining Gershgorin Circles contains precisely k eigenvalues, counting multiplicity.

6. Orthogonality and Least Squares

- The **inner product**, previously called the **dot product**, of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, is defined to be

$$\mathbf{u}\bullet\mathbf{v} = \mathbf{u}^T\mathbf{v} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \sum_{i=1}^{n} u_i v_i. \tag{148}$$

- **Theorem 1, p. 333**. Let **u**, **v** and **w** be vectors in $\mathbb{R}^n$, and c be a scalar. Then

  a. $\mathbf{u} \bullet \mathbf{v} = \mathbf{v} \bullet \mathbf{u}$

  b. $(\mathbf{u} + \mathbf{v}) \bullet \mathbf{w} = \mathbf{u} \bullet \mathbf{w} + \mathbf{v} \bullet \mathbf{w}$

  c. $(c\mathbf{u}) \bullet \mathbf{v} = c(\mathbf{u} \bullet \mathbf{v})$

  d. $\mathbf{u} \bullet \mathbf{u} \geq 0, \quad \text{and} \quad \mathbf{u} \bullet \mathbf{u} = 0 \Longrightarrow \mathbf{u} = \mathbf{0}.$

- The **length** or **norm**[4] of a vector **v** is defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \bullet \mathbf{v}}. \tag{149}$$

- Definition: Two vectors **u** and **v** are **orthogonal** (or **perpendicular**) if

$$\mathbf{u} \bullet \mathbf{v} = 0. \tag{150}$$

the zero vector is orthogonal to all vectors in $\mathbb{R}^n$.

---

[4] also called **Standard Norm**, **Euclidean Norm**, or **2-norm**.

- Suppose W is a subspace of $\mathbb{R}^n$. Then the set

$$W^\perp = \{\mathbf{z} : \mathbf{z} \text{ is orthogonal to all vectors in } W\} \tag{151}$$

is a linear space, called the **orthogonal complement** of W.

- $W^\perp$ is read as "W-perpendicular" or, more commonly, just "W-perp".

- Example: line and plane in $\mathbb{R}^3$.

- **Theorem 3**, p. 337: Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T

$$(\text{Row}A)^\perp = \text{Nul} \quad \text{and} \quad (\text{Col}A)^\perp = \text{Nul}A^T. \tag{152}$$

- A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ from $\mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors from that set is orthogonal, i.e.,

$$i \neq j \quad \Longrightarrow \quad \mathbf{u}_i \bullet \mathbf{u}_j = 0. \tag{153}$$

- **Theorem 4**, p. 340, textbook. If

$$S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\} \tag{154}$$

is an orthogonal set of **nonzero** vectors in $\mathbb{R}^n$, then S is linearly independent. (Hence S is a basis of span(S).)

- Naturally, an **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

- Orthogonal Bases are nice! You can compute coefficients without solving a linear system.

- Suppose

$$\mathrm{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\} \tag{155}$$

is a basis of a subspace W of $\mathbb{R}^n$,

$$B = [\,\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\,], \tag{156}$$

and $\mathbf{y}$ is a vector in W. Then, in general, computing the coordinate vector

$$[\mathbf{y}]_{\mathrm{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} \tag{157}$$

of $\mathbf{y}$ requires the solution of the linear system

$$B[\mathbf{y}]_{\mathrm{B}} = \mathbf{y}. \tag{158}$$

- However, if B is an orthogonal basis we can compute the components of $[\mathbf{y}]_{\mathrm{B}}$ directly:

$$c_j = \frac{\mathbf{y} \bullet \mathbf{u}_j}{\mathbf{u}_j \bullet \mathbf{u}_j}. \tag{159}$$

- **Theorem 6**, p. 345. An $m \times n$ matrix U has orthonormal columns if and only if

$$U^T U = I \tag{160}$$

(where I is the $n \times n$ identity matrix.).

- **Theorem 7**, p. 345. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and y be vectors in $\mathbb{R}^n$. Then:

  a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$

  b. $(U\mathbf{x}) \bullet (U\mathbf{y}) = \mathbf{x} \bullet \mathbf{y}$

  c. $(U\mathbf{x}) \bullet (U\mathbf{y}) = 0 \quad \text{if and only if} \quad \mathbf{x} \bullet y = 0$

- The Pythagorean Theorem states that

$$\|\mathbf{u} \pm \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \quad \Longleftrightarrow \quad \mathbf{u} \bullet \mathbf{v} = 0. \tag{161}$$

- The **orthogonal projection** of a vector $\mathbf{v}$ onto a vector $\mathbf{u}$ is given by

$$\text{proj}_{\mathbf{u}}\mathbf{v} = \frac{\mathbf{u} \bullet \mathbf{v}}{\mathbf{u}\mathbf{u}}\mathbf{u}. \tag{162}$$

- **Theorem 8.** Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{163}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^\perp$.

- This is the **orthogonal Decomposition theorem.** The vector $\hat{\mathbf{y}}$ in (163) is called the **orthogonal projection of y onto** W.

- The textbook uses the notation

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}. \tag{164}$$

- **Best Approximation Theorem** (Theorem 9, p. 352) Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \mathbf{v}\| < \|\mathbf{y} - \hat{\mathbf{y}}\| \tag{165}$$

  for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

- **Theorem 10**, p. 353. If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y} = \sum_{i=1}^{p} (\mathbf{y} \bullet \mathbf{u}_i)\mathbf{u}_i. \tag{166}$$

  If $U = [\,\mathbf{u}_1 \quad \mathbf{u})2 \quad \ldots \mathbf{u}_p\,]$, then

$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y} \tag{167}$$

  for all $\mathbf{y}$ in $\mathbb{R}^n$.

- We considered three versions of the Gram-Schmidt Process.

- Version 1: is described by **Theorem 11**, page 357, textbook: Given a basis

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\} \tag{168}$$

for a non-zero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 - \ldots - \frac{\mathbf{x}_p \bullet \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \bullet \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned} \tag{169}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition,

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \quad \text{for} \quad k = 1, 2, \ldots, p. \tag{170}$$

- Version 2 is just a more compact notation for the process. For $k = 1, \ldots, p$ define

$$\mathbf{v}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\mathbf{x}_k \bullet \mathbf{v}_i}{\mathbf{v}_i \bullet \mathbf{v}_i}\mathbf{v}_i. \tag{171}$$

- Version 3 combines normalization with orthogonalization: For $k = 1, \ldots, p$ define

$$\left\{ \begin{array}{rcl} \mathbf{w}_k &=& \mathbf{x}_k - \displaystyle\sum_{i=1}^{k-1} (\mathbf{x}_k \bullet \mathbf{v}_i)\mathbf{v}_i \\ \mathbf{v}_k &=& \dfrac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \end{array} \right. \tag{172}$$

- Definition: A square matrix Q is **orthogonal** if its columns form an **orthonormal** set.
- This means that

$$Q^T Q = I, \tag{173}$$

i.e., Q is invertible, and

$$Q^{-1} = Q^T. \tag{174}$$

(see textbook, page 346.)

- **Theorem 12**, page 359, textbook. If A is and $m \times n$ matrix with linearly independent columns, then A can be factored a

$$A = QR \tag{175}$$

where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\mathrm{Col}(A)$ and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

- Suppose we have an overdetermined linear system

$$A\mathbf{x} = \mathbf{b} \tag{176}$$

- Here A is $m \times n$, $\mathbf{x}$ is in $\mathbb{R}^n$, $\mathbf{b}$ is in $\mathbb{R}^m$, and $m \geq n$ (and typically, $m > n$).
- Usually, the system (176) will not have a solution. In that case, the next best thing is to solve the alternative problem

$$\|A\mathbf{x} - \mathbf{b}\| = \min \tag{177}$$

- In other words (the words of our textbook), we want to find a vector $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \tag{178}$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

- The textbook calls such an $\hat{\mathbf{x}}$ a **Least Squares Solution** of

$$A\mathbf{x} = \mathbf{b}. \tag{179}$$

- I would call it a solution of

$$\|A\mathbf{x} - \mathbf{b}\| = \min. \tag{180}$$

- First: **Theorem 13** (p. 363) The set of least square solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the **normal equations**

$$A^T A\mathbf{x} = A^T\mathbf{b}. \tag{181}$$

- **Theorem 14** (p. 365) Let A be an $m \times n$ matrix. The following statements are logically equivalent. (This means they are either all true or all false):

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix $A^T A$ is invertible.

- Suppose we write

$$\boxed{A = QR} \tag{182}$$

where

$$Q = m \begin{array}{c} \\ \end{array} \begin{array}{cc} n & m-n \\ \left(Q_1 \right. & \left. Q_2 \right) \end{array} \tag{183}$$

is *orthogonal* and

$$R = \begin{array}{c} n \\ m-n \end{array} \begin{array}{c} n \\ \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \end{array} \tag{184}$$

with R_1 being upper triangular.

- Earlier we discussed how to obtain

$$A = Q_1 R_1, \tag{185}$$

for example by the Gram-Schmidt Process.

- To get Q from Q_1 we simply add vectors to the orthonormal basis of the column space of A to get an orthonormal bass of $\mathbb{R}^m$.

- We won't actually need Q_2, but it's useful to describe the idea.

- A significant property of an orthogonal matrix is that multiplying with it does not alter the norm of a vector:

$$\|Qx\|^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T x = \|x\|^2. \tag{186}$$

- Using

$$A = QR \quad \text{and} \quad Q^T A = R \tag{187}$$

we obtain

$$\begin{aligned} \|Ax - b\|^2 &= \|Q^T(Ax - b)\|^2 \\ &= \|Q^T Ax - Q^T b\|^2 \\ &= \left\| \begin{pmatrix} R_1 x \\ 0 \end{pmatrix} - \begin{pmatrix} Q_1^T b \\ Q_2^T b \end{pmatrix} \right\|^2 \\ &= \left\| R_1 x - Q_1^T b \right\|^2 + \left\| Q_2^T b \right\|^2 . \end{aligned} \tag{188}$$

- Of the two terms on the right we have no control over the second, and we can render the first one zero by solving (the square triangular $n \times n$ linear system)

$$R_1 x = Q_1^T b. \tag{189}$$

- **Definition** (p. 378, textbook): An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, associates a real number $< \mathbf{u}, \mathbf{v} >$ and satisfies the following axioms, for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and all scalars c:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$.

2. $< \mathbf{u} + \mathbf{v}, \mathbf{w} > = < \mathbf{u}, \mathbf{w} > + < \mathbf{v}, \mathbf{w} >$

3. $< c\mathbf{u}, \mathbf{v} > = c < \mathbf{u}, \mathbf{v} >$

4. $< \mathbf{u}, \mathbf{u} > \geq 0$ and $< \mathbf{u}, \mathbf{u} > = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

- A Vector space with an inner product is called an **inner product space**.
- The Cauchy-Schwarz Inequality says

$$| < \mathbf{u}, \mathbf{v} > | \leq \|\mathbf{u}\| \|\mathbf{v}\| \tag{190}$$

- The triangle inequality says

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|. \tag{191}$$

- One major application of inner product spaces is **weighted least squares**.
- The underlying space is $\mathbb{R}^n$ and the inner product is

$$< \mathbf{x}, \mathbf{y} > = \sum_{i=1}^{n} w_i x_i y_i \tag{192}$$

where the w_i are given positive weights.

- The normal equations for the weighted Least Squares Solution of

$$A\mathbf{x} = \mathbf{b} \tag{193}$$

are

$$A^T W A \mathbf{x} = A^T W \mathbf{b}. \tag{194}$$

- Another major example is **Fourier Series**. The underlying linear space is the set of 2π periodic functions that are square integrable over an interval of length 2π.
- The underlying inner product is

$$< f, g >= \int_{-\pi}^{\pi} f(t)g(t)\mathrm{d}t. \tag{195}$$

- The Fourier series of a 2π-periodic function f is

$$f(\mathbf{t}) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\mathbf{t}) + b_n \sin(nt) \tag{196}$$

where the **Fourier coefficients** are given by

$$\begin{aligned} a_n &= \frac{< f, \cos(nt) >}{\pi} = \frac{\int_{-\pi}^{\pi} f(t)\cos(nt)\mathrm{d}t}{\pi} \\ b_n &= \frac{< f, \sin(nt) >}{\pi} = \frac{\int_{-\pi}^{\pi} f(t)\sin(nt)\mathrm{d}t}{\pi} \end{aligned} \tag{197}$$

Ch.7 — Symmetric Matrices and Quadratic Forms

- Throughout this chapter A is a square real matrix.
- A matrix A is symmetric if

$$A = A^T \tag{198}$$

- This concept of symmetry applies only to square matrices.

- Two matrices A and B are **orthogonally similar** if there is an orthogonal matrix P such that
$$B = P^{-1}AB. \tag{199}$$

- A matrix A is **orthogonally diagonalizable** if it is orthogonally similar to a real diagonal matrix D. Thus if A is orthogonally diagonalizable then there exists an orthogonal matrix P such that
$$D = P^{-1}AP = P^TAP. \tag{200}$$

- This implies that A is symmetric since
$$A = PDP^T \tag{201}$$

- So if A is orthogonally diagonalizable then if is symmetric.

- The converse also holds: If A is symmetric then it is orthogonally diagonalizable.

- In class we saw a cool proof by induction of why this is true.

- We also proved the **Spectral Theorem for Symmetric Matrices** An $n \times n$ symmetric matrix A has the following properties:

a. A has n real eigenvalues, counting multiplicities.

b. The dimension of the eigenspace λ equals the multiplicity of λ as a root of the characteristic equation.

c. The eigenspace are mutually orthogonal, in the sense that eigenvectors corresponding different eigenvalues are orthogonal.

d. A is orthogonally diagonalizable.

Put a little more simply: As far as the eigenvalue problem, **symmetric matrices** are as nice as can be.

- **Definition:** A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \tag{202}$$

where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

- In terms of the entries of A and $\mathbf{x}$ the quadratic form is given by

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j. \tag{203}$$

- **The Principal Axes Theorem**. Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variables, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form into a quadratic form

$$\mathbf{y}^T D\mathbf{y} \tag{204}$$

with no cross product terms.

- Note that saying "with not cross product terms" is equivalent to saying that D is diagonal. We've seen of course, that P is the matrix of eigenvectors, and D is the diagonal matrix of eigenvalues.
- A quadratic form $Q = \mathbf{x}^T A\mathbf{x}$, and its matrix A, is

a. **positive definite** if all eigenvalues are positive

b. **negative definite** if all eigenvalues are negative if

c. **indefinite** if some eigenvalues are positive and some are negative

d. **positive semidefinite** if all eigenvalues are non-negative

e. **negative semidefinite** if all eigenvalues are non-positive

A scalar function f of a vector variable $\mathbf{x}$ assumes a minimum value at a point $\hat{x}$ if

$$\nabla f(\hat{x}) = \mathbf{0} \quad \text{and} \quad \nabla^2 f(\hat{x}) \text{ is positive defineite} \tag{205}$$

- It assumes a maximum value if

$$\nabla f(\hat{x}) = \mathbf{0} \quad \text{and} \quad \nabla^2 f(\hat{x}) \text{ is negative defineite} \tag{206}$$

- It has a saddle point at $\hat{x}$ if $\nabla^2 f$ is indefinite at $\hat{x}$.
- If $\nabla^2 f(\mathbf{x})$ is indefinite the second derivative test in inconclusive.

Positive Definiteness is the right generalization of the positivity of numbers.

- The minimum value of a quadratic form $\mathbf{x}^T A\mathbf{x}$ over all unit vectors $\mathbf{x}$ is the minimum eigenvalue of A.
- The maximum value of a quadratic form $\mathbf{x}^T A\mathbf{x}$ over all unit vectors $\mathbf{x}$ is the maximum eigenvalue of A.
- Applying the method of Lagrange multiplier to the constrained problem

$$\mathbf{x}^A\mathbf{x} = \min \quad \text{where} \quad \mathbf{x}^T\mathbf{x} = 1 \tag{207}$$

leads directly to the eigenvector equation

$$A\mathbf{x} = \lambda x. \tag{208}$$

- The *Singular Value Decomposition* of an $m \times n$ matrix A (where we assume for simplicity that $m \geq n$) is

$$A = U \Sigma V^T \tag{209}$$

where

- U is $m \times m$ orthogonal, i.e., $U^{-1} = U^T$,
- V is $n \times n$ orthogonal, i.e., $V^{-1} = V^T$, and
- Σ is $m \times n$ diagonal. Specifically,

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \tag{210}$$

where

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0. \tag{211}$$

- Using the SVD we can reduce many problems to a problem of the same type, but with A replaced with Σ.

- Our list included:

1. Computing the rank of A
2. Computing the determinant of A (provided A is square).
3. Finding all solutions of the linear system $A\mathbf{x} = \mathbf{b}$.
4. Solving the Least Square problem $\|A\mathbf{x}-\mathbf{b}\| = \min$.

- The transformation from problems involving A to problems involving Σ, and vice versa, can all be obtained by multiplying with orthogonal matrices which does not amplify errors.
- We also learned how to approximate A by a sum of rank 1 matrices, which has applications, for example, in image compression.
- We also discussed how to compute the SVD, at least in principle.
- The singular values are the square roots of the eigenvalues of AA^T or A^TA.
- The left singular vectors (the columns of U) are the eigenvectors of AA^T
- The right singular vectors (the columns of V) are the eigenvectors of A^TA.
- We also looked at the textbook procedure for compute the SVD.
- The actual procedure used, for example, in matlab is vastly more complicated.