Math 54 Notes

Leo Villani

Summer 2024

1 Chapter 1

1.1 Systems of Linear Equations

A **linear equation** in the variables $x_1, ..., x_n$ is an equation that can be written in the form:

$$a_1x_1 + ... + a_nx_n = b$$

where b and the **coefficients** $a_1, ..., a_n$ are real or complex numbers.
This leads to a **system of linear equations (linear system)** a collection of one or more linear equations involving the same variables.
A solution is naturally defined as a list $(s_1, ..., s_n)$ which "solves" the linear system.
The set of all possible solutions is called the **solution set** and two linear systems are called **equivalent** if they have the same solution set.
An important fact is:
A system of linear equations has no solution, exactly one solution or infinitely many solutions. A system is said to be **consistent** if it has either one solution or infinitely many solutions and **inconsistent** otherwise.
The language of linear systems naturally develops the need for a simpler formulation which gives us the matrix.
For example the system:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 - 5x_3 &= 10 \end{aligned}$$

gives rise to the **coefficient matrix**:

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

and the **augmented matrix**:

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

The size of the matrix is m by n meaning m rows and n columns.

The matrix gives us a more compact way to look at the linear system and with the following algorithm to solve linear systems will prove very useful.
The three operations to simplify a linear system are:

-Replace one equation by the sum of itself and a multiple of another equation
-Interchange two equations
-Multiply all terms in an equation by a nonzero constant

These are called the **elementary row operations**.
Two matrices are called **row equivalent** if via elementary row operations we can transform one into the other.
Note: these operations are reversible.
This yields the important fact:

If the augmented matrices of two linear systems are row equivalent then the two systems have the same solution set.

1.2 Row Reduction and Echelon Forms

The elementary row operations lead us to the **row reduction algorithm**.
Note: a **leading entry** of a row refers to the leftmost nonzero entry in a nonzero row.
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**. (RREF)

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

Any nonzero matrix may be row reduced into more than one matrix in echelon form using different sequences of row operations. **However the RREF one obtains from a matrix is unique.**
Pivot positions are the positions of the leading entries once in echelon form. Further operations to turn the echelon form into RREF do not move the leading entry (pivot) positions.
Note: the pivot columns are the columns associated with a pivot position.
The row reduction algorithm contains the forward phase which makes the matrix "triangular" and the backward phase which then scales pivots to 1 and "makes more zeros".
After the row reduction algorithm is applied it leads naturally to a solution set. For example say we have row reduced a matrix to:

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

There are three variables because the augmented matrix has 4 columns, so the associated system is:

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

The variables corresponding to pivot columns are called **basic variables** (x_1, x_2) and the others are called **free variables** (x_3)
We can solve the basic variables in terms of free variables for a solution set i.e.:

$$x_1 = 1 + 5x_3$$
$$x_2 = 4 - x_3$$
$$x_3 \text{ is free}$$

x_3 being free means we pick a value of x_3 and a numerical solution pops out after plugging into the basic variable equations.
The above format is called the parametric description of the solution set with free variables acting as parameters.
Note: we use the convention of free variables as parameters even though we could use the basic variables.
A nonreduced echelon form is enough to dewterine existence and uniqueness of solutions.

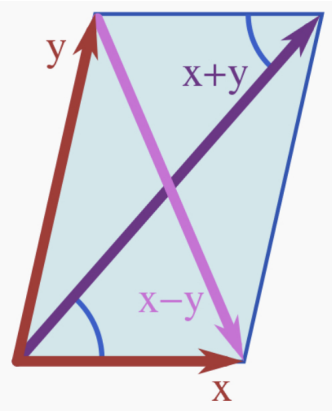
Theorem 1 *(Existence and Uniqueness Theorem) A linear system is consistent if and only if the rightmost column of the augmented matrix is not a pivot column that is if and only if an echelon form of the augmented matrix has no row of the form*
[0...0b] *with* b *nonzero*
If a linear system is consistent then the solution set contains either (i) a unique solution when there are no free variables, or (ii) infinitely many solutions when there is at least one free variable.

1.3 Vector Equations

A matrix with only one column is called a **column vector** or simply a **vector**. Vectors are ordered unlike sets so for vectors to be equivalent the corresponding entries must be equal.

We can identify normal points in R^2 with the column vector $\begin{bmatrix} a \\ b \end{bmatrix}$.

We can think of vectors as points or as arrows, but when thinking of them as arrows it leads us to the parallelogram rule for addition (and subtraction).
Basically if x and y are vectors we draw the parallelogram associated with x and y and we then get $x + y$.



A **linear combination** of the vectors $v_1, ..., v_p$ with weights $c_1, ..., c_p$ is defined as: $c_1v_1 + ... + c_pv_p$.
The set of all possible linear combinations of $v_1, ..., v_p$ is denoted by $Span v_1, ..., v_p$.
It is the collection of all vectors that can be written in the form $c_1v_1 + ... + c_pv_p$ with $c_1, ..., c_p$ being scalars.
We can shift our row reduction talk to the language of vectors naturally as asking whether a vector $\mathbf{b}$ is in $Span v_1, ..., v_p$ amounts to asking whether the vector equation:

$$x_1v_1 + ... + x_pv_p = \mathbf{b}$$

has a solution or equivalently whether the augmented matrix $[v_1...v_p\mathbf{b}]$ has a solution.
Geometrically speaking the span of a single vector denotes the line which intersects the vector and the origin or mathematically all scalar multiples of the vector. Similarly the span of two vectors denotes the plane that contains the two vectors and the origin.

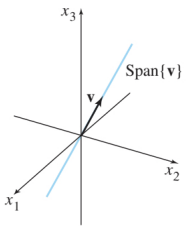


FIGURE 10 Span $\{\mathbf{v}\}$ as a line through the origin.

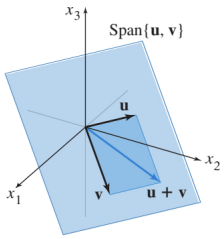


FIGURE 11 Span $\{\mathbf{u}, \mathbf{v}\}$ as a plane through the origin.

1.4 The Matrix Equation $Ax = b$

Matrix vector multiplication is easy and will match up with our framework of solution finding:

If A is an m by n matrix and x is in $\mathbb{R}^n$ then $Ax = [a_1, ..., a_n] \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{bmatrix}$

$$= x_1a_1 + ... + x_na_n$$

Thus the matrix equation $Ax = \mathbf{b}$ has the same solution set as $x_1a_1+...+x_na_n = \mathbf{b}$ and the augmented matrix $[a_1, ..., a_n, \mathbf{b}]$
So a linear system is a matrix equation, a vector equation or an augmented matrix.

Theorem 2 *If A is an m by n matrix:*

1. *For each $\boldsymbol{b}$ in $\mathbb{R}^m$ the equation $Ax = \boldsymbol{b}$ has a solution*
2. *Each $\boldsymbol{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A*
3. *The columns of A span $\mathbb{R}^m$*
4. *A has a pivot position in every row*

Intuitively these are just all conditions to make sure no matter what $\mathbf{b}$ is we stay consistent.
Note the following properties: $A(\mathbf{a}+\mathbf{b}) = A\mathbf{a} + A\mathbf{b}$, $A(c\mathbf{b}) = c(A\mathbf{b})$, where $\mathbf{a}, \mathbf{b}$ are vectors, c is a scalar and A is a matrix with appropriate dimensions.

1.5 Solution Sets of Linear Systems

A **homogeneous** linear system is one that can be written as $Ax = 0$.
Note: Homogeneous linear systems are always consistent by the trivial solution $x = 0$.
Thus we look for nontrivial solutions, but matrix equations can only have no solutions, exactly one solution or infinitely many solutions. So if we find a nontrivial solution we know there are infinite and that there must be ≥ 1 free variable.
For nonhomogeneous systems ($Ax = \mathbf{b}$, where $\mathbf{b} \neq 0$), the general solution can be written as one vector plus the solutions of the homogeneous system. Keep this picture of a nonhomogeneous systems solutions being a shift of the solution of the homogeneous system as it will come back in the differential equations unit.
More formally if the equation $Ax = \mathbf{b}$ is consistent for some given $\mathbf{b}$ and let p be a solution. Then the solution set of $Ax = \mathbf{b}$ is the set of all vectors of the form $w = p + v_h$ where v_h is any solution of the homogeneous system $Ax = 0$.

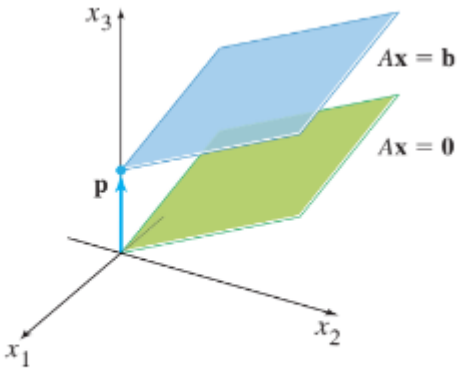


FIGURE 6 Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

Basically if $Ax = \mathbf{b}$ has a solution then the solution set is obtained by translating the solution set of $Ax = 0$ using any particular solution p of $Ax = \mathbf{b}$ for the translation.
Note: we could translate by any other solution vector on the blue plane it just might be harder to visualize.

1.7 Linear Independence

Looking at $Ax = 0$ from the vector perspective $x_1a_1 + ... + x_na_n = 0$ we can think about whether there are non-trivial linear combinations of vectors that make 0. This leads us to the following two definitions:
An indexed set of vectors $v_1, ..., v_p$ in $\mathbb{R}^n$ is said to be **linearly independent** if the vector equation

$$x_1v_1 + ... + x_pv_p = 0$$

has only the trivial solution.
The set $v_1, ..., v_p$ in $\mathbb{R}^n$ is said to be **linearly dependent** if there exist weights $c_1, ..., c_p$, not all zero, such that

$$c_1v_1 + ... + c_pv_p = 0$$

If the c-weights exist then the equation above is called a **linear dependence relation**.
Thus we can think back to $Ax = 0$ or equivalently $x_1a_1 + ... + x_na_n = 0$ and see each linear dependence relation among the columns of A corresponds to a nontrivial solution of $Ax = 0$.
Hence, the columns of A are linearly independent if and only if $Ax = 0$ only has the trivial solution.
To build some intuition for independence consider a single vector. A single vector is only linearly independent if its not the 0 vector as the equation $xv = 0$ will have the trivial solution if $v \neq 0$.
For two vectors we get $x_1v_1 + x_2v_2 = 0$ or $v_1 = -\frac{x_2}{x_1}v_2$. Basically we know 2 vectors are linearly independent if they arent scalar multiples of each other.
Formally, a set of two vectors v_1, v_2 is linearly dependent if at least one of the vectors is a multiple of the other. Geometrically we can think of the two vectors being on the same line through the origin.
This extends naturally to more vectors. To be linearly independent we want to in some way be gaining new information. If a vector in the set can already be made up of the other vectors then it doesn't "add new information" to the set which leaves us with a linearly dependent set.
Specifically if at least one of the vectors is a linear combination of the others than we have a linear dependent set.

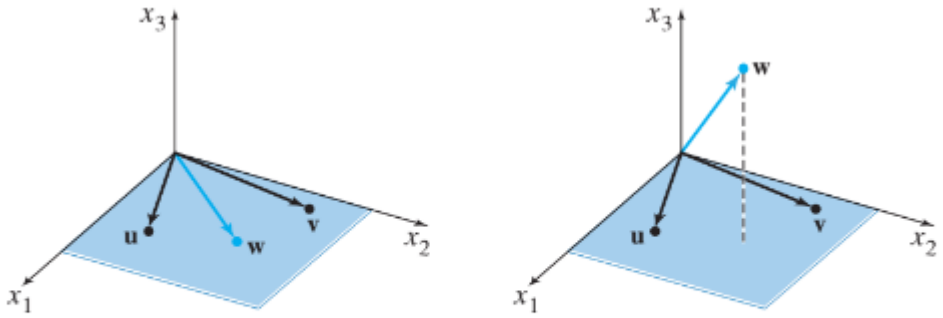


FIGURE 2 Linear dependence in $\mathbb{R}^3$.

Theorem 3 *If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is any set $v_1, ..., v_p$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.*

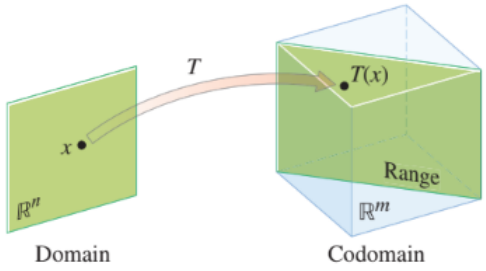
This theorem makes sense because if we go back to the matrix equation for these vectors the coefficient matrix will be forced to have free variables as there will be more columns then rows.

Theorem 4 *If a set $S = v_1, ..., v_p$ in $\mathbb{R}^n$ contains the zero vector, then the set is linearly dependent.*

The proof is just that if the zero vector is labeled v_1 we can set c_1 to be anything nonzero and all other weights to be 0, and we get a nontrivial linear dependence relation.

1.8 Introduction to Linear Transformations

Instead of the vector linear combination approach to the equation $Ax = b$ we can now think about a matrix A multiplying an input vector x to generate b, which gives rise to A-multiplication being a function / **linear transformation**. By the equation $Ax = b$ we are looking for vectors x when transformed by the function induced by A-multiplication result in b.
A **transformation** T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector x in $\mathbb{R}^n$ a vector $T(x)$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T, and $\mathbb{R}^m$ is called the **codomain** of T. The notation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ indicates that the domain of T is $\mathbb{R}^n$ and the codomain is $\mathbb{R}^m$. For x in $\mathbb{R}^n$, the vector $T(x)$ in $\mathbb{R}^m$ is called the **image of** x. The set of all images $T(x)$ is called the **range** of T.



We can finally answer why we speak of "linear" algebra. We have been analyzing matrices with corresponding matrix multiplication functions which induce linear transformations. Linearity is described by the following property:
A transformation T is linear if:

1. $T(u + v) = T(u) + T(v)$ for all u, v in the domain of T
2. $T(cu) = cT(u)$ for all scalars c and all u in the domain of T

Note: These properties also imply $T(0) = 0$ and $T(cu + dv) = cT(u) + dT(v)$ for all vectors u, v in the domain of T and all scalars c, d.
Linear transformations decompose linear combinations:
$T(c_1v_1 + ... + c_pv_p) = c_1T(v_1) + ... + c_pT(v_p)$

1.9 The Matrix of a Linear Transformation

The linear transformation T and its associated matrix A share important properties via some definitions we have learned earlier in the chapter. The way to find the associated A is to see what T does to the columns of the n by n sized identity matrix.
Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that:

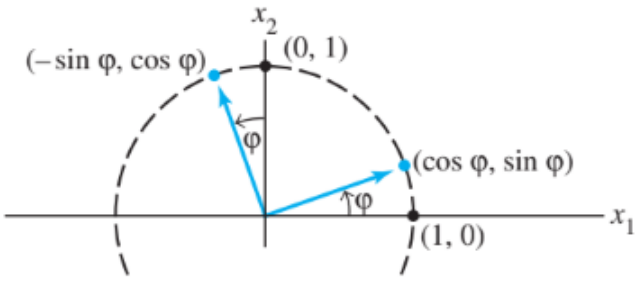
$$T(x) = Ax \text{ for all } x \text{ in } \mathbb{R}^n$$

In fact A is the m by n matrix whose jth column is the vector $T(e_j)$ where e_j is the jth column of the identity matrix in $\mathbb{R}^n$

$$A = [T(e_1)...T(e_n)]$$

As an example lets consider $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ which rotates each points counterclockwise in $\mathbb{R}^2$ about the origin by angle α. Our goal is to find the matrix A associated with this transformation.
This is easily done by seeing where the vectors $\begin{bmatrix}1\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\end{bmatrix}$ move to.
Notice $\begin{bmatrix}1\\0\end{bmatrix}$ rotates into $\begin{bmatrix}cos(\alpha)\\sin(\alpha)\end{bmatrix}$, and $\begin{bmatrix}0\\1\end{bmatrix}$ rotates into $\begin{bmatrix}-sin(\alpha)\\cos(\alpha)\end{bmatrix}$, yielding the matrix A:

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



Just for fun we can take this further and uncover something about matrix multiplication and function composition.
Suppose now I want to rotate first by θ and then by φ. I could also do that by just rotating $\theta + \varphi$. This would look like a composition of two linear transformations:

$$T_\varphi \circ T_\theta$$

but we can encode this in terms of a matrix product:

$$A_\phi A_\theta$$

where A_ϕ, and A_θ are the corresponding standard matrix representations of the rotation transformations. If we take the matrix product:

$$A_\phi A_\theta = \begin{bmatrix} cos(\varphi) & -sin(\varphi) \\ sin(\varphi) & cos(\varphi) \end{bmatrix} \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix}$$

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

which if we remember our cosine and sine "angle-adding" identites

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

Notice how this is just the original transformation rotating by angle $\varphi + \theta$ exactly what was expected from the composition operator. Isnt that cool! :)

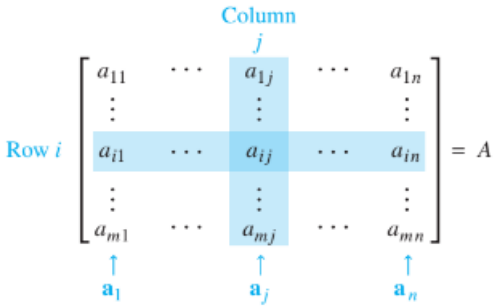
Moving on we will deal with some function definitions:
A mapping $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each b in $\mathbb{R}^m$ is the image of at least one x in $\mathbb{R}^n$.
A mapping $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is said to be **one-to-one** if each b in $\mathbb{R}^m$ is the image of at most one x in $\mathbb{R}^n$.
In terms of the induced matrix A we can check onto and one-to-one via free variables and whether or not we have unique solutions, or being always consistent.
Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if the equation $T(x) = 0$ has only the trivial solution.
Thus we are left with the theorem:
Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation, and let A be the standard matrix for T. Then:

1. T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$
2. T is one-to-one if and only if the columns of A are linearly independent

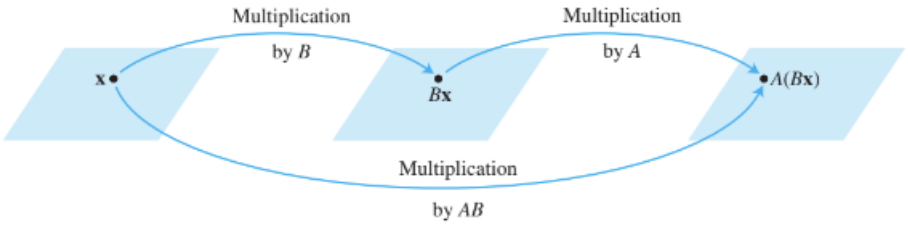
2 Chapter 2

2.1 Matrix Operations

We now need to talk about some basic matrix operations to then understand conceptually later. First when speaking of matrices we like to speak of their entries conveniently which is best summarized by the following diagram:



Note: **Diagonal entries** are just entries on the main diagonal from top left to bottom right, and a **diagonal matrix** is a matrix with only diagonal entries being nonzero. The **identity matrix** is just a diagonal matrix with all 1s on the main diagonal and a **zero matrix** is just an entirely 0-filled matrix.
We add matrices entrywise (given that they have the same dimension), and we can scale matrices i.e. $2A$ by just multiplying each entry in A by 2.
Note: These operations are associative, commutative and distributive in the way you could expect.
Moving on to matrix multiplication, in terms of linear transformations we can think of the composition of functions as multiplication of their respective matrices.



There are two ways to compute the product:

1. We can compute p matrix vector products $AB = A[b_1...b_p] = [Ab_1...Ab_p]$

2. Or we can compute entry wise via the row column rule $(AB)_{ij} = a_{i1}b_{1j} + ... + a_{in}b_{nj}$

A good way to remember the row column rule is that if I want $(AB)_{23}$ then I just need to dot product the 2nd row of A with the 3rd column of B.
Matrix multiplication is associative and left and right distributive as well given the correct dimensions. For a matrix A and a matrix B to be "multipliable" you need the number of columns of A to match with the number of rows of B. (mxn, nxp would work)
Note: Multiplying a matrix A by the identity matrix gives back A i.e. $AI_n = A = I_m A$.
Matrix multiplication is not commutative in general.
$A^k = A * ... * A$, where A is multiplied by itself k times.
A^T is called the transpose of a matrix where the columns are formed by the rows changing from a mxn matrix to a nxm matrix.
Transposing obeys the following properties:

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. For any scalar r, $(rA)^T = rA^T$
4. $(AB)^T = B^T A^T$

2.2 The Inverse of a Matrix

The last operation which hasnt been covered is division. We can think of the analogue of the multiplicative inverse of a nonzero number as the inverse of a matrix.
The matrix inverse of A defined as A^{-1} when multiplied by A yields the identity matrix.

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

Matrices can be or not be invertible, invertible matrices are called **singular matrices** and non-invertible are called **nonsingular**.
Note: the inverse of a $2x2$ matrix has an easy formula:

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$,$ad - bc \neq 0$ then A is invertible and $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. If $ad - bc = 0$ then A is not invertible.

Note: $ad - bc$ is called the determinant, $detA$ and will come up later.
If A is an invertible nxn matrix then for each b in $\mathbb{R}^n$ the equation $Ax = b$ has the unique solution $x = A^{-1}b$.
Here are a couple of useful properties:

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

You can see that the product of invertible matrices is invertible and the formula is given above.

There is an algorithm to find the inverse of a nxn matrix A. A matrix A is invertible if and only if it is row equivalent to I_n and we can use the row operations to go from A to I_n to construct A^{-1}.
Algorithm:

Row reduce the augmented matrix $[AI]$. If A is row equivalent to I then $[AI]$ is row equivalent to $[IA^{-1}]$. Otherwise A does not have an inverse.

Computerwise/numerically we rarely compute $A^{-1}b$ to find the solution to $Ax = b$ as the row reduction may be more accurate and it takes less operations.

2.3 Characterizations of Invertible Matrices

The Invertible Matrix Theorem:

Let A be a square nxn matrix. Then the following statements are equivalent.

1. A is an invertible matrix
2. A is row equivalent to I_n
3. A has n pivot positions
4. The equation $Ax = 0$ has only the trivial solution
5. The columns of A form linearly independent set
6. The linear transformation $x \to Ax$ is one-to-one
7. The equation $Ax = b$ has at least one solution for each b in $\mathbb{R}^n$
8. The columns of A span $\mathbb{R}^n$
9. The linear transformation $x \to Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$
10. There is an nxn matrix C such that $CA = I$
11. There is an nxn matrix D such that $AD = I$
12. A^T is an invertible matrix

This theorem gives a lot of different ways to figure out if a matrix A is invertible, including matrix methods we have covered in previous sections i.e. counting pivot positions, etc.
Note: The Invertible Matrix Theorem applies only to square matrices.
Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear transformation and let A be the standard matrix for T. Then T is invertible if and only if A is an invertible matrix. In that case, the linear transformation S given by $S(x) = A^{-1}(x)$ is the unique function satisfying $S(T(x)) = x$, $T(S(x)) = x$ for all x in $\mathbb{R}^n$.
We can see the power of the previous theorem via an example:
What can you say about a one-to-one linear transformation T from $\mathbb{R}^n$ into $\mathbb{R}^n$?
We know that the columns of the associated matrix A are linearly independent if T is one-to-one. So A is invertible by the Invertible Matrix Theorem, and thus T is invertible by the theorem.

3 Chapter 3

3.1 Introduction to Determinants

We want to extend our determinant definition and formula for general matrices. This leads us to the cofactor expansion which yields an alternating sum of submatrices to take determinants on.
For $n \geq 2$ the determinant of an nxn matrix $A = [a_{ij}]$ is the sum of the n terms of the form $\pm a_{1j} det(A_{1j})$ with plus and minus signs alternating where the entries $a_{11}, ..., a_{1n}$ are from the first row of A. In symbols:
$detA = a_{11}det(A_{11}) - a_{12}det(A_{12}) + ... + (-1)^{1+n}a_{1n}det(A_{1n}) = \sum_{j=1}^{n}(-1)^{1+j}a_{1j}det(A_{1j})$
Note: A_{11} denotes the matrix A after removing the first row and first columns entries.
We can rewrite this long sum in terms of cofactors where $C_{ij} = (-1)^{i+j}det(A_{ij})$ leaving $detA = a_{11}C_{11} + ... + a_{1n}C_{1n}$. It turns out this sum can be computed on any row or any column so just pick the row or the column that is the easiest (a lot of zeros).
The alternating term for C_{ij} can be thought of creating a checkerboard pattern:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

Example:
Compute $detA$ where

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

Because we can pick any row or column the first column is a good choice. The cofactor expansion will be $detA = 3 * det(\begin{bmatrix} 2 & -5 & 7 & 3 \\ 0 & 1 & 5 & 0 \\ 0 & 2 & 4 & -1 \\ 0 & 0 & -2 & 0 \end{bmatrix}) + 0 * C_{21} + 0 * C_{31} + 0 * C_{41} + 0 * C_{51}$ where all 0 multiplied cofactors will be 0 in the sum. So we can find the determinant of the $4x4$ submatrix ideally again using the first column. Try finishing this example on your own the answer is -12. *(Does that mean A is invertible or not invertible? We'll see in the next section.)*
The cofactor expansion can prove a useful theorem:
If A is a triangular matrix, then $detA$ is the product of the entries on the main diagonal of A.

3.2 Properties of Determinants

The secret of determinants is how they change when row operations are performed.

Let A be a square matrix.

1. If a multiple of one row of A is added to another row to produce a matrix B, then $detB = detA$.
2. If two rows of A are interchanged to produce B then $detB = -detA$
3. If one row of A is multiplied by k to produce B then $detB = kdetA$

These determinant rules for elementary row operations can help find determinants efficiently. Especially because if we reduce to a triangular matrix the theorem in the last section can tell us the determinant instantly.
The main theorem of this section is:
A square matrix A is invertible if and only if $detA \neq 0$
We can also perform operations on the columns of the matrix the same way we have used row operations. This is backed by the fact that if A is an nxn matrix then $detA^T = detA$. Thus every statement about elementary row operations for $detA$ can be replaced with elementary column operations.
Determinants of multiplied nxn matrices can also be found easily as: $detAB = detA * detB$. This kind of rule is not true in general for the sum of matrices so be careful. ($det(A + B) \neq detA + detB$ in general)

3.3 Cramer's Rule, Volume, and Linear Transformations

Cramers rule is used only for theoretical calculations given its inefficiency but it can be used to study solutions of $Ax = b$ given changes to b.
First we define $A_i(b) = [a_1...b...a_n]$
Cramers Rule:
Let A be an invertible nxn matrix. For any b in $\mathbb{R}^n$, the unique solution x of $Ax = b$ has entries given by

$$x_i = \frac{detA_i(b)}{detA}, i = 1, 2, ..., n$$

We'll touch briefly on the adjugate matrix of a square matrix A who like Cramers rule is usually only used in theoretic applications.
Using the cofactor definition from before we construct the adjugate matrix denoted $adjA$

$$adjA = \begin{bmatrix} C_{11} & C_{21} & ... & C_{n1} \\ C_{21} & C_{22} & ... & C_{n2} \\ ... & ... & ... & ... \\ C_{n1} & C_{2n} & ... & C_{nn} \end{bmatrix}$$

then if A is an invertible nxn matrix

$$A^{-1} = \frac{1}{detA} adjA$$

We can also verify the interpretation of a determinants as an area or a volume.
If A is a $2x2$ matrix, the area of the parallelogram determined by the columns of A is $|detA|$. If A is a $3x3$ matrix, the volume of the parallelepiped determined by the columns of A is $|detA|$.
Determinants can also be used to describe an important geometric property of linear transformations in the plane and $\mathbb{R}^3$.
First denote $T(S)$ as the set of images of points in S, for some set S in the domain of T.
Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be the linear transformation determined by a $2x2$ matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$\text{area of } T(S) = |detA| * \text{ area of } S$$

If T is determined by a $3x3$ matrix A, and if S is a parallelepiped in $\mathbb{R}^3$ then

$$\text{volume of } T(S) = |detA| * \text{ volume of } S$$

4 Chapter 4

4.1 Vector Spaces and Subspaces

A **vector space** is a nonempty set V of objects, called vectors, on which are defined two operations, called addition and multiplication by scalars, subject to the ten axioms listed below. The axioms must hold for all vectors u, v, w in V and for all scalars c, d.

1. The sum of u and v denoted $u + v$ is in V.
2. $u + v = v + u$
3. $(u + v) + w = u + (v + w)$
4. There is a zero vector 0 in V such that $u + 0 = u$
5. For each u in V there is a vector $-u$ in V such that $u + (-u) = 0$
6. The scalar multiple of u by c denoted by cu is in V
7. $c(u + v) = cu + cv$
8. $(c + d)u = cu + du$
9. $c(du) = (cd)u$
10. $1u = u$

We use $\mathbb{R}^n$ as a "nice" vector space to visualize many of the results through the chapter but there are many other vector spaces which are relevant that can be treated the same.
In many problems, a vector space consists of an appropriate subset of vectors from a larger vector space. If we satisfy only 3 of the 10 axioms that the subset is also a vector space called a **subspace**.
A subspace of a vector space V is a subset H of V that has three properties:

1. The zero vector of V is in H
2. H is closed under vector addition. That is, for each u, v in H the sum $u + v$ is in H
3. Similarly H must be closed under multiplication by scalars. That is for each u in H and each scalar c, the vector cu is in H

The most common way for constructing a subspace is by picking vectors in V and considering their Span.
Vector Subspaces Theorem:
If $v_1, ..., v_p$ are in a vector space V, then $Span(v_1, ..., v_p)$ is a subspace of V.
This result makes sense as the span considers all possible linear combinations

of the vectors (including the trivial 0 combination) so of course it satisfies the 3 properties of a subspace.
We call $Span(v_1, ..., v_p)$ the **subspace spanned** by $(v_1, ..., v_p)$. Given any subspace H of V a **spaning set** for H is a set $(v_1, ..., v_p)$ in H such that $H = Span(v_1, ..., v_p)$.
Note: a spanning set is not necessarily of minimal size for that we should think to linear independence and optimality of information

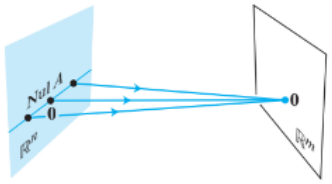
4.2 Null Spaces, Column Spaces, and Linear Transformations

In linear algebra applications subspaces of $\mathbb{R}^n$ usually arise in one of two ways:

1. the set of all solutions to a system of homogeneous linear equations
2. the set of all linear combinations of certain specified vectors

Recall the set of all x that satisfy $Ax = 0$ is called the solution set of the system. Because of $Ax = 0$ importance we denote this solution set the **null space** of the matrix A.
The null space of a mxn matrix A, written as $NulA$ is the set of all solutions of the homogeneous equation $Ax = 0$. In set notation:

$$NulA = (x : x \text{ is in } \mathbb{R}^{\kappa} \text{ and } Ax = 0)$$



The null space is actually a subspace of $\mathbb{R}^n$. For a proof notice by linearity if you take a linear combination of vectors in the $NulA$ each output will be 0.
Mathematically if v_1, v_2 are in $NulA$:

$$A(c_1v_1 + c_2v_2) = c_1A(v_1) + c_2A(v_2) = c_1 * 0 + c_2 * 0 = 0$$

Note: we cannot consider the solution set of $Ax = b$ with $b \neq 0$ as a subspace because it does not contain the zero vector
To solidify some concepts from ealier units, the number of free variables is equal to the number of vectors in the spanning set of $NulA$ because in a sense they control how large the subspace is
Another important subspace is called the column space.
The **column space** of a matrix A, written as $ColA$ is the set of all linear combinations of the columns of A. If $A = [a_1...a_n]$ then

$$ColA = Span(a_1, ..., a_n) = (b : b = Ax \text{ for some } x \text{ in } \mathbb{R}^n)$$

$ColA$ is really just the range of the linear transformation $x \longrightarrow Ax$ and it gives us a new way to discuss $Ax = b$.
$Ax = b$ has a solution if and only if b is in $ColA$.
Similarly $ColA$ is all of $\mathbb{R}^m$ if and only if the equation $Ax = b$ has a solution for each b in $\mathbb{R}^m$.

Its important to note where $NulA$ and $ColA$ "live". $NulA$ is a subset of the domain as its a constrained version of all possible input vectors. Conversely $ColA$ "lives" in the codomain as its a constrained version of the codomain.

Contrast Between Nul *A* and Col *A* for an *m* x *n* Matrix *A*

Nul A	Col A
1. Nul A is a subspace of $\mathbb{R}^n$.	**1.** Col A is a subspace of $\mathbb{R}^m$.
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	**2.** Col A is explicitly defined; that is, you are told how to build vectors in Col A.
3. It takes time to find vectors in Nul A. Row operations on $[\,A \quad \mathbf{0}\,]$ are required.	**3.** It is easy to find vectors in Col A. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A.	**4.** There is an obvious relation between Col A and the entries in A, since each column of A is in Col A.
5. A typical vector $\mathbf{v}$ in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	**5.** A typical vector $\mathbf{v}$ in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector $\mathbf{v}$, it is easy to tell if $\mathbf{v}$ is in Nul A. Just compute $A\mathbf{v}$.	**6.** Given a specific vector $\mathbf{v}$, it may take time to tell if $\mathbf{v}$ is in Col A. Row operations on $[\,A \quad \mathbf{v}\,]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	**7.** Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	**8.** Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

Moving back towards linear transformations we can define similar terms for a general $T : V \longrightarrow W$.
The **kernel (or null space)** of T is the set of all u in V such that $T(u) = 0$.
The **range** of T is the set of all vectors in W of the form $T(x)$ for some x in V.
Note: the proof kernel is a subspace of V is identical to the one for the null space of a matrix A just replace A with T, and the range is obviously a subspace of W

4.3 Linearly Independent Sets; Bases

In this section we talk about sets that efficiently describe the space they span. Recall the definitions for linear independence, dependence, and linear dependence relations in $\mathbb{R}^n$ as they will help describe the efficiency of a spanning set for a general V.
Let H be a subspace of a vector space V. An indexed set of vectors $B = (b_1, ..., b_p)$ in V is a **basis** for H if

1. B is a linearly independent set
2. the subspace spanned by B coincides with H; that is $H = Span(b_1, ..., b_p)$

The basic intuition is that a basis is an efficient spanning set that contains no unnecessary vectors. In fact we can construct a basis from any spanning set by removing unneeded vectors.
The Spanning Set Theorem:
Let $S = (v_1, ..., v_p)$ be a set in V, and let $H = Span(v_1, ..., v_p)$

1. If one of the vectors in S say v_k is a linear combination of the remaining vectors in S then the set formed from S by removing v_k still spans H
2. If $H \neq (0)$ some subset of S is a basis for H

For $NulA$ when the spanning set contains nonzero vectors its already linearly independent so it is a basis. For finding a basis for $ColA$ we can just row reduce and consider only pivot columns as the other vectors contain no new information for the image space.
Simply, the pivot columns of a matrix A form a basis for $ColA$ (these are the columns associated with the pivot positions in the original matrix not the already reduced column vectors).
Basis are the "perfect" size. For example in $\mathbb{R}^3$ a spanning set must have 3 or more vectors but a linearly independent can only have up to 3 vectors. Think of the basis size as a Goldilocks-type not too big or small set.

4.4 Coordinate Systems

A basis B for a vector space V imposes a coordinate system on V.
The Unique Representation Theorem:
Let $B = (b_1, ..., b_n)$ be a basis for a vector space V. Then for each x in V, there exists a unique set of scalars $c_1, ..., c_n$ such that

$$x = c_1 b_1 + ... + c_n b_n$$

This makes sense because there should be c's that due this do to B "spanning-ness" and they should be unique by its linear independence.
Suppose $B = (b_1, ..., b_n)$ is a basis for V and x is in V. The coordinates of x relative to the basis B are the weights $c_1, ..., c_n$ such that $x = c_1 b_1 + ... + c_n b_n$
If $c_1, ..., c_n$ are the $B-$coordinates of x then the vector in $\mathbb{R}^n$

$$[x]_B = \begin{bmatrix} c_1 \\ ... \\ c_n \end{bmatrix}.$$

is the **coordinate vector of x (relative to B)** or the **$B-$coordinate vector of x**. The mapping $x \longrightarrow [x]_B$ is the **coordinate mapping (determined by B)**.
When a basis B for $\mathbb{R}^n$ is fixed the $B-$coordinate vector of a specified x is easily found, heres an example:
Let $b_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $b_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $x = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, and $B = (b_1, b_2)$. Find the coordinate vector $[x]_B$ of x relative to B.
To do this we need to find c_1, c_2 such that $c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$. Turning this into a matrix problem and row reducing we can easily find $c_1 = 3, c_2 = 2$.

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

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

If we look at the matrix above $\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$ it changes the $B-$coordinates of a vector x into the standard coordinates for x. A similar change of coordinates can be carried out for a general $B = (b_1, ..., b_n)$ in $\mathbb{R}^n$.
Let

$$P_B = [b_1 b_2 ... b_n]$$

So

$$x = P_B [x]_B$$

We call P_B the **change of coordinates matrix** from B to the standard basis in $\mathbb{R}^n$. Also, because the columns of P_B form a basis for $\mathbb{R}^n$, P_B is invertible so we can change the above equation to get:

$$P_B^{-1}x = [x]_B$$

This gives us a way to go from x to $[x]_B$.
Hence, by choosing a basis $B = (b_1, ..., b_n)$ for a vector space V we can treat V as the familiar space $\mathbb{R}^n$.
Let $B = (b_1, ..., b_n)$ be basis for a vector space V. Then the coordinate mapping $x \longrightarrow [x]_B$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.
The previous theorem reveals an important result about isomorphisms. In general a one-to-one linear transformation from V to W is called an isomorphism from V onto W. The notation may be different but they are indistinguishable as vector spaces. Every vector space calculation in V is accurately reproduced in W and vice versa. So any vector space with a basis of n vectors is indistinguishable to $\mathbb{R}^n$.

4.5 The Dimension of a Vector Space

In the previous section we realized a vector space B containing n vectors is isomorphic to $\mathbb{R}^n$. The number n is actually an important property of the space V called the dimension that does not depend on B.
Note the following two theorems:
If a vector space V has a basis $B = (b_1, ..., b_n)$ then any set in V containing more than n vectors must be linearly dependent.
If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.

If V is spanned by a finite set, then V is said to be **finite dimensional** and the **dimension** of V written $dimV$ is the number of vectors in a basis for V. The dimension of the zero vector space is defined to be 0. If V is not spanned by a finite set then V is said to be **infinite dimensional**.
Note: $dim\mathbb{R}^n = n$, and $dim\mathbb{P}^n = n + 1$, where $\mathbb{P}^n$ is the vector space of polynomials with degree $\leq n$.
We expect subspaces to contain less than or equal information to V. Thus
Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded if necessary to a basis for H. Also H is finite dimensional and

$$dimH \leq dimV$$

When the dimension of a vector space or subspace is known, the search for a basis is simplified by the next theorem. It says that if a set has the right number of elements, then one has only to show either that the set is linearly independent or that it spans the space. The theorem is important for applications where linear independence is easier to prove than spanning.
The Basis Theorem:
Let V be a $p-$dimensional vector space, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.
Coming back to the ubiquitous subspaces $NulA$, $ColA$ the dimension of $NulA$ is the number of free variables in the equation $Ax = 0$ and the dimension of $ColA$ is the number of pivot columns in A.

4.6 Rank

Similar to the column space the row space is just all possible linear combinations of the rows of a matrix treated as vectors, denoted $RowA$.
Note: Since the rows are just the columns of A^T we know $RowA = ColA^T$.
It turns out we can actually row reduce matrices when we want to find something out about the row space.
If two matrices A and B are row equivalent then their row spaces are the same.
If B is in echelon form the nonzero rows of B form a basis for the row space of A as well as for that of B.
Thus if asked to find bases for $RowA$, $ColA$, and $NulA$ you can follow the following steps:

1. Row reduce A to an echelon form B
2. The nonzero rows of the B form a basis for $RowA$ (yes the reduced rows not the original rows)
3. Identify the pivot columns and thus the associated columns vectors in the original A to form the basis for $ColA$
4. Solve $Ax = 0$ using the echelon form or a further reduced echelon form and after putting solution into parametric vector form the vectors spanning $NulA$ appear. These vectors are also guaranteed linearly independent so they form a basis for $NulA$

Note: Unlike the basis for $ColA$ the bases for $RowA$ and $NulA$ have no simple connection with the entries in A itself.

$$\textbf{rank of } A = dimColA$$

Since $RowA$ is the same as $ColA^T$ the dimension of the row space is the rank of A^T. The dimension of the null space is sometimes called the **nullity** of A.
The Rank-Nullity Theorem:
The dimensions of the column space and the row space of an mxn matrix A are equal. This common dimension, the rank of A, also equals the number of pivot positions in A and satisfies the equation

$$rank(A) + Nullity(A) = n$$

The intuition behind the theorem is that:

(# pivot columns) + (# non-pivot columns/ # free variables) = (# columns)

With our new tools we can continue the Invertible Matrix Theorem:
The Continued Invertible Matrix Theorem:

Let A be a square nxn matrix. Then the following statements are equivalent.

1. A is an invertible matrix
2. The columns of A form a basis for $\mathbb{R}^n$
3. $ColA = \mathbb{R}^n$
4. $dimColA = n$
5. $rankA = n$
6. $NulA = (0)$, where (0), is the 0-set
7. $dimNulA = 0$, where $dimNulA = Nullity(A)$

5 Chapter 5

5.1 Eigenvectors and Eigenvalues

In general when speaking of a matrix there are certain special vectors that reveal something important about the underlying transformations. While there are plenty of directions the matrix moves its input vectors towards there are some special directions where the transformation only stretches and dilates the input.
An **eigenvector** of an nxn matrix A is a nonzero vector x such that $Ax = \lambda x$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution x of $Ax = \lambda x$; such an x is called an eigenvector corresponding to λ.
To find eigenvectors for a matrix we look at the equation $Ax = \lambda x$. Rewriting we get to $(A - \lambda I)x = 0$ and we look for nontrivial solutions x. The set corresponding to a fixed λ is called the **eigenspace** of A corresponding to λ which is clearly a null space meaning it is also a subspace of $\mathbb{R}^n$.
Just like in the case of determinants triangular matrices make things easy:
The eigenvalues of a triangular matrix are the entries on the main diagonal.
Note: in the special case of the $0-$eigenvalue we are looking at nontrivial solutions to $Ax = 0x = 0$ which gives us a new condition for invertibility.
Given that eigenvectors denote "special directions" the next theorem should make intuitive sense given our knowledge of bases:
If $v_1, ..., v_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, ..., \lambda_r$ of an nxn matrix A then the set $(v_1, ..., v_r)$ is linearly independent.

5.2 The Characteristic Equation

Because eigenvalues are scalars λ such that $(A - \lambda I)x = 0$ has a nontrivial solution by the invertible matrix theorem we are looking for the matrix to not be invertible. An easy way to ensure this is that we need $det(A - \lambda I) = 0$.
With some new tools we can again continue the Invertible Matrix Theorem:
The Continued Continued Invertible Matrix Theorem:

Let A be a square nxn matrix. Then the following statements are equivalent.

1. A is an invertible matrix
2. The number 0 is not an eigenvalue of A.
3. The determinant of A is not zero.

Before looking onwards make sure you know how to compute determinants and determinant properties from chapter 3.
The scalar equation $det(A - \lambda I) = 0$ is called the **characteristic equation** of A.
A scalar λ is an eigenvalue of A if and only if λ satisfies $det(A - \lambda I) = 0$.
When A is an nxn matrix then $det(A - \lambda I)$ is a polynomial of degree n called the **characteristic polynomial** of A.
Here is a quick example:
Find the characteristic equation of

$$A = \begin{bmatrix} 5 & -2 & 6 & 1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

After setting up the matrix $A - \lambda I$ and finding its determinant easily (upper triangular) you get

$$det(A - \lambda I) = (5 - \lambda)(3 - \lambda)(5 - \lambda)(1 - \lambda)$$

Yielding characteristic equation

$$(\lambda - 5)^2(3 - \lambda)(1 - \lambda) = 0$$

where $(\lambda - 5)^2(3 - \lambda)(1 - \lambda)$ is the characteristic polynomial.
We can read off the solutions as $\lambda = 1, 3, 5$ with (algebraic) multiplicities $1, 1, 2$ respectively corresponding to how many times they appear in the polynomial.
The characteristic polynomial also comes up in talk of similar matrices.
If A and B are nxn matrices then A **is similar to** B if there is an invertible matrix P such that $P^{-1}AP = B$ or equivalently $A = PBP^{-1}$. Using $Q = P^{-1}$ we see B is also similar to A so A and B **are similar**.
If A and B are similar then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).

For the proof just take the appropriate determinants and see the P and P^{-1} cancel.
Note: Having the same eigenvalues doesnt mean you're similar. Also similarity is not the same as row equivalence. (row operations can change eigenvalues)

5.3 Diagonalization

In many cases the λ-value and vector information for A can be displayed in a useful factorization $A = PDP^{-1}$, where D is diagonal. Diagonal matrices are easy to exponentiate:

$$D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \text{ then } D^k = \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}$$

This allows easy computation of A^k because intermediary P and P^{-1} terms cancel, $A^k = PD^kP^{-1}$.
A square matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix.
The Diagonalization Theorem:
An nxn matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.
In fact $A = PDP^{-1}$ with D a diagonal matrix if and only if the columns of P are n linearly independent eigenvectors of A. In this case the diagonal entries of D are eigenvalues of A that correspond respectively to the eigenvectors in P.

In other words A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an **eigenvector basis** of $\mathbb{R}^n$.

Diagonalizing a Matrix "Step-by-Step Tutorial":

1. Find the eigenvalues of A (characteristic polynomial solutions)
2. Find n linearly independent eigenvectors of A (normal null space calculation for each eigenspace)
   Note: if step 2 is not possible than A is not diagonalizable
3. Construct P from the eigenvectors found in step 2, $P = [v_1, ..., v_n]$
4. Construct D from the corresponding eigenvalues

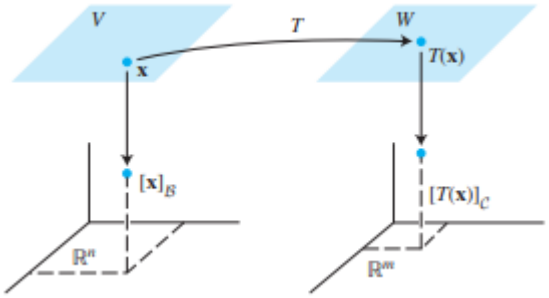
To verify instead of computing P^{-1} we can simply verify $AP = PD$.
Note: An nxn matrix with n distinct eigenvalues is diagonalizable.
For matrices whose eigenvalues are not distinct we just need the dimension of each eigenspace of λ_k to equal the mulitiplicity of λ_k.

5.4 Eigenvectors and Linear Transformations

Continuing our talk of $A = PDP^{-1}$ we will see that the transformation $x \longrightarrow Ax$ is essentially the same as the map $u \longrightarrow Du$ given the right perspective.

Recall for any linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ we can implement it as left-multiplication by the standard matrix A. We need to generalize this idea for a linear map between any two finite vector spaces.

Let V be an n-dimensional vector space, let W be an $m-$dimensional vector space and let T be any linear transformation from V to W. To associate a matrix with T chooses bases B, C for V, W respectively. Given any x in V the coordinate vector $[x]_B$ is in $\mathbb{R}^n$ and the coordinate vector of its image, $[T(x)]_C$ is in $\mathbb{R}^m$.



The connection between $[x]_B$ and $[T(x)]_C$ is easy to find. Let $(b_1, ..., b_n)$ be the basis B. If $x = r_1 b_1 + ... + r_n b_n$, then

$$[x]_B = \begin{bmatrix} r_1 \\ ... \\ r_n \end{bmatrix}$$

and

$$T(x) = T(r_1 b_1 + ... + r_n b_n) = r_1 T(b_1) + ... + r_n T(b_n)$$

because T is linear. As the coordinate mapping from W to $\mathbb{R}^m$ is linear we get

$$[T(x)]_C = r_1 [T(b_1)]_C + ... + r_n [T(b_n)]_C$$

Since $C-$coordinates vectors are in $\mathbb{R}^m$ the vector equation can be written as

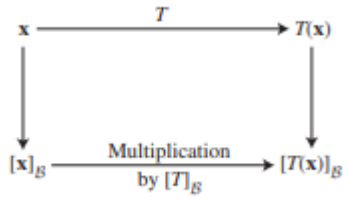
$$[T(x)]_C = M[x]_B$$

where

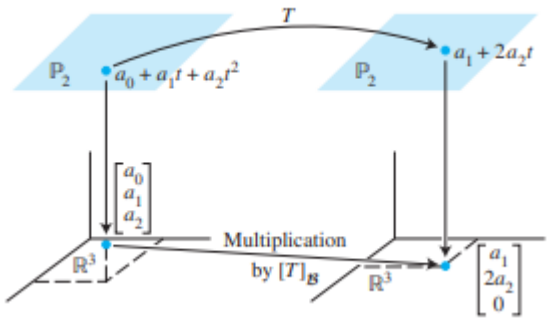
$$M = [[T(b_1)]_C ... [T(b_n)]_C]$$

The matrix M is a matrix representation of T called the **matrix for T relative to the bases B and C**
Simply said discounting "coordinate stuff", T can be viewed as left multiplication by M.
Note: If B and C are bases for the same space V and if T is the identity transformation $T(x) = x$ then the matrix M is just the change-of-coordinates matrix in chapter 4.
In the common case where $W = V$ and the basis $C = B$ the matrix M is called the **matrix for T relative to B** or simply the **B-matrix for T** denoted $[T]_B$.
The $B-$matrix for $T : V \longrightarrow V$ satisfies

$$[T(x)]_B = [T]_B[x]_B \text{ for all } x \text{ in } V$$

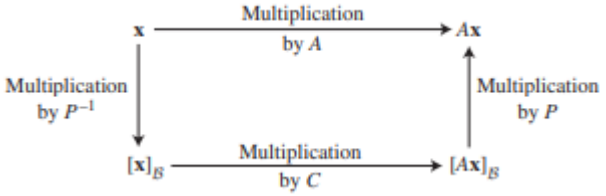


For the derivative linear transformation $T : \mathbb{P}^2 \longrightarrow \mathbb{P}^2$ we get the following diagram:



Diagonal Matrix Representation Theorem
Suppose $A = PDP^{-1}$, where D is a diagonal nxn matrix. If B is the basis for $\mathbb{R}^n$ formed from the columns of P then D is the $B-$matrix for the transformation $x \longrightarrow Ax$.

More generally instead of a matrix D if A is similar to C with $A = PCP^{-1}$ then C is the $B-$matrix for the transformation $x \longrightarrow Ax$ when the basis B is formed from the columns of P.



Conversely if $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is defined by $T(x) = Ax$ and if B is any basis of $\mathbb{R}^n$ then the $B-$matrix for T is similar to A. In fact we can see that $[T]_B = P^{-1}AP$, hence the set of all matrices similar to a matrix A coincides with the set of all matrix representations of the transformation $x \longrightarrow Ax$

5.5 Complex Eigenvalues

Considering complex numbers polynomials of degree always have n roots counting multiplicities.
The matrix eigenvalue-eigenvector theory already developed for $\mathbb{R}^n$ applies equally well to $\mathbb{C}^n$, so we use the same definitions from before to define eigenvalues and eigenvectors.
Note: the complex conjugate of a vecotr x in $\mathbb{C}^n$ is the vector whose entries are the complex conjugates of the entries of x. ($a + bi$ goes to $a - bi$) Similarly the complex conjugate of a matrix is just the conjugate entrywise.
We can see that for a real valued matrix A if λ is an eigenvalue then its complex conjugate $\bar{\lambda}$ is also an eigenvalue, in other words these eigenvalues always come in pairs.
There is a quick discussion on a fact about a certain A that I wouldnt worry too much about, but here is the theorem.
Let A be a real $2x2$ matrix with a complex eigenvalue $\lambda = a - bi(b \neq 0)$ and an associated eigenvector v in $\mathbb{C}^2$. Then

$$A = PCP^{-1} \text{ where } P = [Re(v)Im(v)] \text{ and } C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

6 Chapter 6

6.1 Inner Product, Length, and Orthogonality

We now shift our attention to generalizing geometric concepts such as length, distance, and perpendicularity which are well known for $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$.
If u and v are vectors in $\mathbb{R}^n$ the "matrix" product $u^T v$ will result in a $1x1$ matrix or a scalar. This number $u^T v$ is called the **inner product** of u and v and is often written as $u \cdot v$. This specific inner product is also called the **dot product**.

Inner product of u and v:

$$[u_1 ... u_n] \begin{bmatrix} v_1 \\ ... \\ v_n \end{bmatrix} = u_1 v_1 + ... + u_n v_n$$

Here are some useful properties of the inner product.
Let u, v, and w be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

1. $u \cdot v = v \cdot u$
2. $(u + v) \cdot w = u \cdot w + v \cdot w$
3. $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$
4. $u \cdot u \geq 0$, and $u \cdot u = 0$ if and only if $u = 0$

We can combine properties 2 and 3 to get:

$$(c_1 u_1 + ... + c_p u_p) \cdot w = c_1(u_1 \cdot w) + ... + c_p(u_p \cdot w)$$

Note: The square root of $v \cdot v$ is always defined because $v \cdot v \geq 0$.
The length (or norm) of v is the nonnegative scalar $\|v\|$ defined by

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + ... + v_n^2}, \text{ and } \|v\|^2 = v \cdot v$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$ this corresponds to the normal pythagorean distance from the origin to v.
Note: $\|cv\| = |c|\|v\|$
A vector whose length is 1 is called a **unit vector**. If we divide a nonzero vector v by its length then we get a unit vector with v's direction preserved. This process is called **normalizing**.
We are now ready to describe how close one vector is to another. Recall on the real number line the distance between a and b is just $|a - b|$. Analogously for u and v in $\mathbb{R}^n$ the **distance between u and v**, written as $dist(u, v)$ is the length of the vector $u - v$. Equivalently

$$dist(u, v) = \|u - v\|$$

For a general form of perpendicularity we can again go back to the dot product.
Two vectors u and v in $\mathbb{R}^n$ are **orthogonal** (to each other) if $u \cdot v = 0$.
Note: the zero vector is orthogonal to every vector
The next theorem provides a useful fact about orthogonal vectors.
The Pythagorean Theorem
Two vectors u and v are orthogonal if and only if $\|u+v\|^2 = \|u\|^2 + \|v\|^2$
If a vector z is orthogonal to every vector in a subspace W of $\mathbb{R}^n$ then z is said to be **orthogonal to** W. The set of all vectors z that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^{\perp}$.
Here are a couple of facts on $W^{\perp}$

1. A vector x is in $W^{\perp}$ if and only if x is orthogonal to every vector in a set that spans W

2. $W^{\perp}$ is a subspace of $\mathbb{R}^n$

Row, Col, Nul Perp Theorem
Let A be an mxn 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:

$$(RowA)^{\perp} = NulA \text{ and } (ColA)^{\perp} = NulA^T$$

If u and v are nonzero vectors in either $\mathbb{R}^2$ or $\mathbb{R}^3$ then there is a nice connection between their inner product and the angle θ between the two line segments from the origin to the points identified with u and v. The formula is

$$u \cdot v = \|u\|\|v\|cos(\theta)$$

(in statistics for suitable u and v, $cos(\theta)$ is the *correlation coefficient*)

6.2 Orthogonal Sets

A set of vectors $(u_1, ..., u_p)$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is if $u_i \cdot u_j = 0$ whenever $i \neq j$.
Orthogonality Independence Theorem
If $S = (u_1, ..., u_p)$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$ then S is linearly independent and hence is a basis for the subspace spanned by S.
Note: an **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.
With an orthongonal basis we can compute the weights in a linear combination easily.
Let $(u_1, ..., u_p)$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each y in W, the weights in the linear combination

$$y = c_1 u_1 + ... + c_p u_p$$

are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j} \ (j = 1, ..., p)$$

Instead of solving a system of equations we can just calculate the weights directly.
Now consider the problem of decomposing a vector y into the sum of two vectors one being the multiple of u and the other orthogonal to u. $(y = \hat{y} + z)$

$$\hat{y} = proj_L y = \frac{y \cdot u}{u \cdot u} u$$

$\hat{y}$ is called the **orthogonal projection of** y **onto** u, and the vector z is called the **component of** y **orthogonal to** u

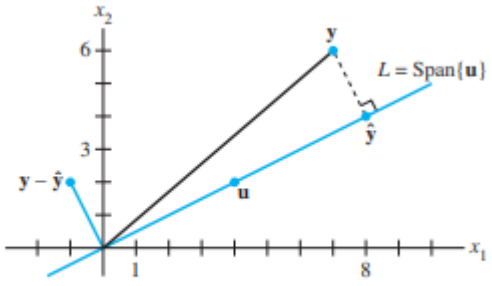


FIGURE 3 The orthogonal projection of **y** onto a line L through the origin.

Note: if asked for the distance from y to a line L spanned by a vector u you can just calculate the perpendicular section in the orthogonal decomposition i.e. the z vector and take its norm
We define an **orthonormal set** to be an orthogonal set made of unit vectors.

If W is the subspace spanned by such a set then the vectors form an **orthonormal basis** for W by the sets immediate independence.
Matrices whose columns form an orthonormal set are important in applications and in computer algorithms for matrix computations. Here are a couple of theorems on them:

1. An mxn matrix U has orthonormal columns if and only if $U^T U = I$
2. Let U be an mxn matrix with orthonormal columns, and let x and y be in $\mathbb{R}^n$. Then
   (a) $\|Ux\| = \|x\|$
   (b) $(Ux) \cdot (Uy) = x \cdot y$
   (c) $(Ux) \cdot (Uy) = 0$ if and only if $x \cdot y = 0$

6.3 Orthogonal Projections

The orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$. Given a vector y and a subspace W in $\mathbb{R}^n$ there is a vector $\hat{y}$ in W such that

1. $\hat{y}$ is the unique vector in W for which $y - \hat{y}$ is orthogonal to W
2. $\hat{y}$ is the unique vector in W closest to y (least squares next section)

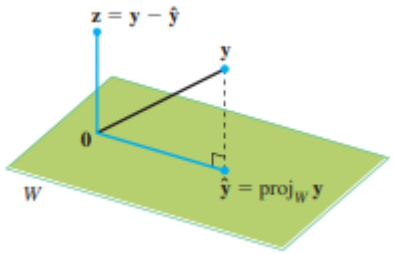
Given an orthogonal basis for $\mathbb{R}^n$ its easy to decompose y into a vector in a subspace W made up of some of the vectors in the basis and a vector in $W^{\perp}$ consisting of a combination of the others. It turns out in the next theorem we only need an orthogonal basis for W to make this decomposition.
The Orthogonal Decomposition Theorem
Let W be a subspace of $\mathbb{R}^n$. Then each y in $\mathbb{R}^n$ can be written uniquely in the form

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

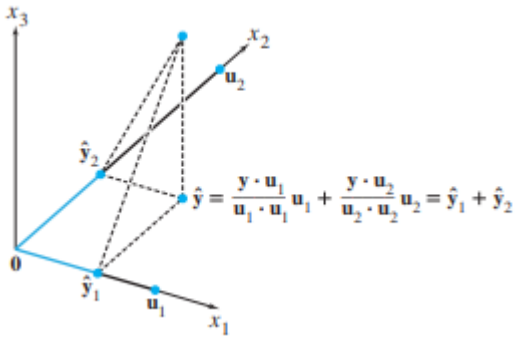
where $\hat{y}$ is in W and z is in $W^{\perp}$. In fact if $(u_1, ..., u_p)$ is any orthogonal basis of W then

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + ... + \frac{y \cdot u_1}{u_1 \cdot u_1} u_p$$

and $z = y - \hat{y}$
The vector $\hat{y}$ is called the **orthogonal projection of y onto W** and often is written as $proj_W y$



A good geometric understanding comes from looking at the formula above and understanding that the orthogonal projection of y is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal.



Note: If $(u_1, ..., u_p)$ is an orthogonal basis for W and if y happens to be in W, then the formula for $proj_W y$ is exactly the same as the representation of y given earlier in orthogonal sets section (the easy weights formula). In this case $proj_W y = y$.

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let y be any vector in $\mathbb{R}^n$ and let $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the closest point in W to y, in the sense that

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

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

The vector $\hat{y}$ is called the **best approximation to y by elements of W**. The distance from y to some fixed v in W is called the "error" in approximation and $\hat{y}$ minimizes it.

The formula for $proj_W y$ is simplified when the u_i are an orthonormal set as the bottom terms are just 1.

Simplified Theorem

If $(u_1, ..., u_p)$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$ then

$$proj_W y = (y \cdot u_1)u_1 + ... + (y \cdot u_p)u_p$$

If $U = [u_1, ..., u_p]$, then

$$proj_W y = UU^T y \text{ for all } y \text{ in } \mathbb{R}^n$$

6.4 The Gram-Schmidt Process

The Gram-Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of $\mathbb{R}^n$.

This is easiest to show by example so lets walk through one:
Let $x_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $x_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}$, and $x_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$. Then (x_1, x_2, x_3) is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

Step 1 Let $v_1 = x_1$ and $W_1 = Span(x_1) = Span(v_1)$ **Step 2** Let v_2 be the vector produced by subtracting from x_2 its projection onto the subspace W_1 (forcing its orthogonality). That is let

$$\begin{aligned} v_2 &= x_2 - proj_{W_1} x_2 \\ &= x_2 - \tfrac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \text{ (as } v_1 = x_1) \\ &= \begin{bmatrix}0\\1\\1\\1\end{bmatrix} - \tfrac{3}{4} \begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}\frac{-3}{4}\\\frac{1}{4}\\\frac{1}{4}\\\frac{1}{4}\end{bmatrix} \end{aligned}$$

Now (v_1, v_2) is an orthogonal basis for the subspace W_2 spanned by x_1 and x_2.

Step 2' (optional) Feel free to scale v_2 to make later computations easier i.e. multiply $\begin{bmatrix}\frac{-3}{4}\\\frac{1}{4}\\\frac{1}{4}\\\frac{1}{4}\end{bmatrix}$ by 4 to get $v_2' = \begin{bmatrix}-3\\1\\1\\1\end{bmatrix}$.

Step 3 Let v_3 be the vector produced by subtracting from x_3 its projection onto the subspace W_2. Again this will force v_3 to be orthogonal to W_2 and hence v_1 and v_2'.

$$proj_{W_2} x_3 = \tfrac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 + \tfrac{x_3 \cdot v_2'}{v_2' \cdot v_2'} v_2' = \begin{bmatrix}0\\\frac{2}{3}\\\frac{2}{3}\\\frac{2}{3}\end{bmatrix}$$

Then v_3 is the component fo x_3 orthogonal to W_2 namely

$$v_3 = x_3 - proj_{W_2} x_3 = \begin{bmatrix}0\\\frac{-2}{3}\\\frac{1}{3}\\\frac{1}{3}\end{bmatrix}$$

v_1, v_2', v_3 is our resulting orthogonal basis for W.

The Gram-Schmidt Process

Given a basis $(x_1, ..., x_p)$ for a nonzero subspace W of $\mathbb{R}^n$ define

$$\begin{gathered} v_1 = x_1 \\ v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\ ... \; v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - ... - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1} \end{gathered}$$

Then $(v_1, ..., v_p)$ is an orthogonal basis for W. In addition

$$Span(v_1, ..., v_k) = Span(x_1, ..., x_k) \text{ for } 1 \leq k \leq p$$

By the correctness of the Gram-Schmidt Process any nonzero subspace W of $\mathbb{R}^n$ has an orthogonal basis because an ordinary basis is always available.

Note: If asked for an orthonormal basis its easy to just scale all the v_i's at the end of the process.

The Gram-Schmidt Process also yields a matrix factorization presented in the next theorem.

The QR Factorization If A is an mxn matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an mxn matrix whose columns form an orthonormal basis for $ColA$ and R is an nxn upper triangular invertible matrix with positive entries on its diagonal.

Note: when asked for a QR factorization of a matrix A we can find Q via Gram-Schmidt and normalizing the column vectors and then as Q is orthonormal $Q^TQ = I$,

$$Q^TA = Q^T(QR) = IR = R$$

which gives us a way to find R.

6.5 Least-Squares Problems

"When a solution is demanded and none exists, the best one can do is to find an x that makes Ax as close as possible to b. Think of Ax as an approximation to b. The smaller the distance between b and Ax given by $\|b - Ax\|$ the better the approximation. The **general least-squares problem** is to find an x that makes $\|b - Ax\|$ as small as possible.

If A is mxn and b is in $\mathbb{R}^m$, a **least squares solution** of $Ax = b$ is an $\hat{x}$ in $\mathbb{R}^n$ such that

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

for all x in $\mathbb{R}^n$.
The best approximation theorem says $\hat{b} = proj_{ColA}b$ is the best we can do where we seek $\hat{x}$ such that $A\hat{x} = \hat{b}$.
By some orthogonality properties we get $A^T(b - A\hat{x}) = 0$ yielding the least square solution

$$A^TAx = A^Tb$$

This system is called the **normal equations** for $Ax = b$ and its solutions are often denoted as $\hat{x}$.
The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equations $A^TAx = A^Tb$.
Note: By multiplying both sides of $Ax = b$ by A^T we are now always guaranteed to have a solution $\hat{x}$
The next theorem gives useful criteria for determining when there is only one least-squares solution of $Ax = b$.
Uniqueness Theorem
Let A be an mxn matrix. The following statements are logically equivalent:

1. The equation $Ax = b$ has a unique least-squares solution for each b in $\mathbb{R}^m$
2. The columns of A are linearly independent
3. The matrix A^TA is invertible

When these statements are true, the least-squares solution $\hat{x}$ is given by

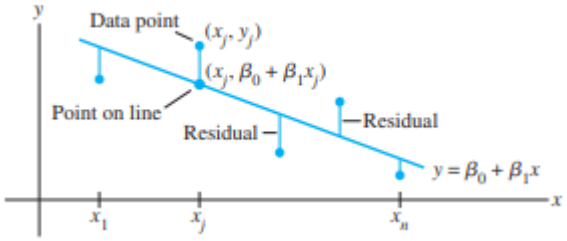
$$\hat{x} = (A^TA)^{-1}A^Tb$$

Note: the distance from b to $A\hat{x}$ is called the **least-squares error**
QR Least Squares Theorem
Given an mxn matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A. Then for each b in $\mathbb{R}^m$ the equation $Ax = b$ has a unique least-squares solution given by

$$\hat{x} = R^{-1}Q^Tb$$

6.6 Applications to Linear Models

A common task in science and engineering is to analyze and understand relationships among several quantities that vary. We quickly define some notation that is commonly used in statistical analysis of scientific and engineering data. Instead of $Ax = b$ we write $X\beta = y$ and refer to X as the **design matrix**, β as the **parameter vector** and y as the **observation vector**.
Sometimes when observing data we seek to relate two variables by a linear equation. We want to determine the weights in a linear equation that minimize the **residuals**. If we seek a solution that minimizes the square of the residuals (which actually is a pretty risk-averse error function go take CS 188 to learn more) we can rely on the **least squares line** as a solution.



The least squares line is actually equivalent to the least squares solution of $X\beta = y$ given appropriate X, β, y

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \; y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix},$$

as we are trying to find β_0, β_1 that would satisfy

Predicted y-value		**Observed y-value**
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

Note: Usually there will be no β_0, β_1 that will satisfy every data point hence the enitre point of using least-squares.
A common practice before computing a least-squares line is to compute the average $\bar{x}$ of the original $x-$values and form a new variable $x* = x - \bar{x}$. The new $x-$data are said to be in **mean-derivation form**. In this case, the two columns of the design matrix will be orthogonal and the solutions of the normal

equations will be simplified.
In some applications when not fitting with a line we just set up the equation using a **residual vector** ϵ

$$y = X\beta + \epsilon$$

Any equation of this form is referred to as a **linear model**. Once X and y are determined we seek to minimize the length of ϵ, which amounts to doing least squares on $X\beta = y$.
We can add nonlinear features to our design matrix by just modifying the variables. For example if $y = X\beta + \epsilon$ for a quadratic model in x then we want to find $\beta_0, \beta_1, \beta_2$ so that $y = \beta_0 + \beta_1 x + \beta_2 x^2$ we can use the following:

$$\begin{bmatrix} y_1 \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ ... & ... & ... \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ ... \\ \epsilon_n \end{bmatrix}$$

6.7 Inner Product Spaces

Notions of length, distance, and orthogonality are often important in applications involving a vector space. For $\mathbb{R}^n$, these concepts were based on the properties of the dot product, but for other spaces, we need analogues of the inner product with the same properties.

An **inner product** on a vector space V is a functionn that, to each pair of vectors u and v in V associates a real number $\langle u, v\rangle$ and satisfies the following axioms, for all u, v and w in V and all scalars c:

1. $\langle u, v\rangle = \langle v, u\rangle$
2. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$
3. $\langle cu, v\rangle = c\langle u, v\rangle$
4. $\langle u, u\rangle \geq 0$ and $\langle u, u\rangle = 0$ if and only if $u = 0$

A vector space with an inner product is called an **inner product space**.

Let V be an inner product space, with inner product denoted by $\langle u, v\rangle$. Just as in $\mathbb{R}^n$ we define the **length** or **norm** of a vector v to be the scalar

$$\|v\| = \sqrt{\langle v, v\rangle}$$

A **unit vector** is one whose length is 1. The **distance between u and v** is $\|u - v\|$. Vectors u and v are **orthogonal** if $\langle u, v\rangle = 0$.
The existence of orthogonal bases for finite-dimensional subspaces of an inner product space can be established by the Gram–Schmidt process, just as in $\mathbb{R}^n$.
A common problem in applied mathematics involves a vector space V whose elements are functions. The problem is to approximate a function f in V by a function g from a specified subspace W of V . The "closeness" of the approximation of f depends on the way $\|f - g\|$ is defined. (think about best approximating a 4-th degree polynomial with a polynomial of degree 2)
Now we will talk about two important inner product related inequalities
Given a vector v in an inner product space V and given a finite-dimensional subspace W we may apply the pythagorean theorem to the orthogonal decomposition of v with respect to W and obtain

$$\|v\|^2 = \|proj_W v\|^2 + \|v - proj_W v\|^2$$

Simply the norm of the projection of v onto W does not exceed the norm of v itself. This yields the following important inequality
The Cauchy-Schwarz Inequality
For all u, v in V,

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

The Cauchy-Schwarz inequality also helps prove another fundamental inequality.
The Triangle Inequality
For all u, v in V,

$$\|u + v\| \leq \|u\| + \|v\|$$

In terms of function spaces such as $C[a, b]$ the space of all continuous functions on an interval $a \leq t \leq b$ there is a widely used inner product that uses some calculus and is known to be on exams.
For f, g in $C[a, b]$ set

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt$$

As an exercise prove to yourself this is a well-defined inner product.

7 Symmetric Matrices and Quadratic Forms

7.1 Diagonalization of Symmetric Matrices

Recall a symmetric matrix is a matrix A such that $A^T = A$. This forces A to be square and while its main diagonal entries can be arbitrary if we reflect the entries on one side of the main diagonal to the other they should be the same.
Symmetric Matrix Eigenvectors
If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.
An nxn matrix A is said to be **orthogonally diagonalizable** if there are an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that

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

An nxn matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.
Note: when asked to orthogonally diagonalize a matrix, basis eigenvectors corresponding to the same eigenspace are linearly independent but not necessarily orthogonal. Instead just apply gram-schmidt to generate a new orthogonal eigenbasis.
The Spectral Theorem *The set of eigenvalues of a matrix A is sometimes called the spectrum of A, and the following description of the eigenvalues is called a* ***spectral theorem***
An nxn symmetric matrix A has the following properties:

1. A has n real eigenvalues, counting multiplicities
2. The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation (geometric multiplicity = algebraic multiplicity)
3. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal
4. A is orthogonally diagonalizable

Using $A = PDP^T$ where the columns of P are orthonormal eigenvectors we can derive

$$A = \lambda_1 u_1 u_1^T + ... + \lambda_n u_n u_n^T$$

This representation of A is called a **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum (eigenvalues). Each terms is an nxn matrix of rank 1 called a projection matrix in the sense that $(u_1 u_1^T)x$ is the orthogonal projection of x onto the subspace spanned by u_1.

7.2 Quadratic Forms

A **quadratic form** on $\mathbb{R}^n$ is a function Q defined of $\mathbb{R}^n$ whose value at a vector x in $\mathbb{R}^n$ can be computed by an expression of the form $Q(x) = x^T Ax$, where A is an nxn symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

Quadratic forms tend to look like: $Q(x) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$ where each term is a "degree 2" term hence the name quadratic form. The ax_i^2 terms imply the values of the diagonal entries and the cross terms $bx_ix_j, i \neq j$ are the off diagonal entries at $(i, j), (j, i)$ for A to be symmetric having value $\frac{b}{2}$.

In some cases, quadratic forms are easier to use when they have no cross-product terms—that is, when the matrix of the quadratic form is a diagonal matrix. Fortunately, the cross-product term can be eliminated by making a suitable change of variable.

Change of Variable in a Quadratic Form

If x represent a variable vector in $\mathbb{R}^n$ then a **change of variable** is an equation of the form

$$x = Py \text{ or equivalently } y = P^{-1}x$$

where P is an invertible matrix and y is a new variable vector in $\mathbb{R}^n$.

If the change of variable is made in a quadratic form $x^T Ax$ then

$$x^T Ax = (Py)^T A(Py) = y^T (P^T AP)y$$

and the new matrix of the quadratic form is $P^T AP$. Since A is symmetric we know there exists a P such that $P^T AP$ is a diagonal matrix D and the quadratic form would become $y^T Dy$.

Because $y = P^{-1}x = P^T x$ computing $y^T Dy$ is the same as going to x via P and computing $x^T Ax$.

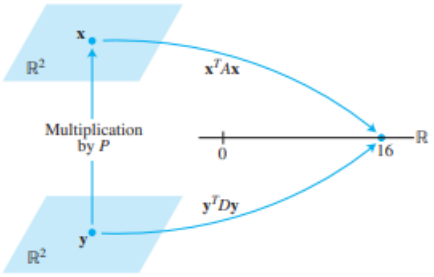


FIGURE 1 Change of variable in $\mathbf{x}^T A\mathbf{x}$.

The Principle Axes Theorem

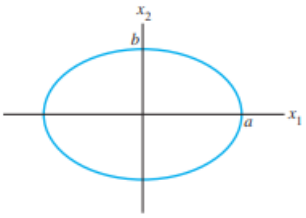
Let A be an nxn symmetric matrix. Then there is an orthogonal change of variable, $x = Py$, that transforms the quadratic form $x^T Ax$ into a quadratic form $y^T Dy$ with no cross-product term.

The columns of P in the theorem are called the **principal axes** of the quadratic form $x^T Ax$.

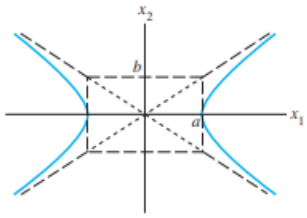
Geometric view of Principlal Axes
Suppose $Q(x) = x^T Ax$, where A is an invertible $2x2$ symmetric matrix, and let c be a constant. It can be shown that the set of all x in $\mathbb{R}^2$ that satisfy

$$x^T Ax = c$$

either corresponds to an ellipse (or circle), a hyperbola, two intersecting lines, or a single point, or contains no points at all. If A is diagonal the graph is in standard position but if not then it is rotated out of position with "correctly ordered" coordinate system now being the principal axes.
Standard Position (Diagonal A):

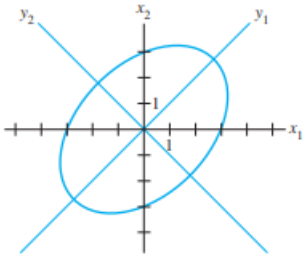


$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1,\ a > b > 0$
ellipse

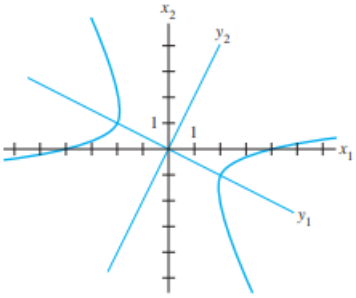


$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1,\ a > b > 0$
hyperbola

Rotated by Principal Axes (Nondiagonal A):



(a) $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$



(b) $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$

Classifying Quadratic Forms
A quadratic form Q is:

1. **positive definite** if $Q(x) > 0$ for all $x \neq 0$
2. **positive semidefinite** if $Q(x) \geq 0$ for all x
3. **negative definite** if $Q(x) < 0$ for all $x \neq 0$
4. **negative semidefinite** if $Q(x) \leq 0$ for all x
5. **indefinite** if $Q(x)$ assumes both positive and negative values

Quadratic Forms and Eigenvalues Theorem
Let A be an nxn symmetric matrix. Then a quadratic form $x^T Ax$ is:

1. positive definite if and only if the eigenvalues of A are all positive
2. negative definite if and only if the eigenvalues of A are all negative
3. indefinite if and only if A has both positive and negative eigenvalues

Note: The classification of a quadratic form is often carried over to the matrix of the form. Thus a positive definite matrix A is a symmetric matrix for which the quadratic form $x^T Ax$ is positive definite. Other terms, such as positive semidefinite matrix, are defined analogously.

7.3 The Singular Value Decomposition (SVD)

As we know now not all matrices can be factored as $A = PDP^{-1}$ with D diagonal. However a factorization $A = QDP^{-1}$ is possible for any mxn matrix A! A special factorization of this type is called the **singular value decomposition**. Eigenvalues for a matrix A represent how much A stretches certain vectors so the largest magnitude eigenvalue has the largest "stretching effect". There is a similar analogue for rectangular matrices.

The Singular Values of an mxn Matrix

Let A be an mxn matrix. Then A^TA is symmetric and can be orthogonally diagonalized. The singular values of A are the square roots of the eigenvalues of A^TA denoted by $\sigma_1, ...\sigma_n$ and they are arranged in decreasing order.

Note: the eigenvalues of A^TA are always non-negative so the square root is always well defined

Suppose $(v_1, ..., v_n)$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA arranged so that the corresponding eigenvalues of A^TA satsify $\lambda_1 \geq ... \geq \lambda_n$ and suppose A has r nonzero singular values. Then $(Av_1, ..., Av_n)$ is an orthogonal basis for $ColA$ and $rankA = r$.

The decomposition of A involves an mxn "diagonal matrix" Σ of the form

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

where D is an rxr diagonal matrix for some r not exceeding the smaller of m and n.

The Singular Value Decomposition

Let A be an mxn matrix with rank r. Then there exists an mxn matrix Σ for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geq ... \geq \sigma_r > 0$ and there exist an mxm orthogonal matrix U and an nxn orthogonal matrix V such that

$$A = U\Sigma V^T$$

Note: U and V are not uniquely determined by A but the diagonal entries of Σ are necessarily the singular values of A. The columns of U are the called **left singular vectors** and for V's columns **right singular vectors**.

How to SVD

1. **Find an orthogonal diagonlization of A^TA**
   Find the eigenvalues of A^TA and a corresponding orthonormal set of eigenvectors, $(v_1, ..., v_n)$. (note might need to do gram schmidt if not all distinct eigenvalues)

2. **Set up V and Σ**
   Arrange the eigenvalues of A^TA in decreasing order. The corresponding unit eigenvectors are the right singular vectors of A. This constructs V. Then taking square roots of eigenvalues consturct Σ.

3. **Construct** U
   When A has rank r the first r columns of U are the normalized vectors obtained from $Av_1, ..., Av_r$. Notice that $\|Av_i\| = \sigma_i$ so we can find the columns of U by $u_i = \frac{1}{\sigma_i} Av_i$. We then form an orthogonal basis for $\mathbb{R}^m$ by using the first few u vectors and then extending to a basis via gram schmidt and some null space calculations.

Using some new language we can again continue the invertible matrix theorem.
The Continued Invertible Matrix Theorem:

Let A be a square nxn matrix. Then the following statements are equivalent.

1. A is an invertible matrix
2. $(ColA)^{\perp} = (0)$
3. $(NulA)^{\perp} = \mathbb{R}^n$
4. $RowA = \mathbb{R}^n$
5. A has n nonzero singular values.

The Reduced SVD
A nice application of the SVD is its use as an approximator. Because we order the singular values and the columns in terms of how large the eigenvalues are which encode how significant of a stretch it provides by deleting coumns we get good approximation to what A represents.
When Σ contains rows or columns of zeros we get the following more compact definition if $r = rankA$.

$$A = U_r D V_r^T$$

Note: By removing lower eigenvalues and corresponding columns/rows we get good approximations to A. This is useful for things like pictures where we can delete less useful rows and columns of the factorization, less useful singular values, to scale down how large the image file/data is.