Linear Algebra Notes

anastasia

Spring 2025

Contents

1 Linear Equations in Linear Algebra

1.1 Systems of Linear Equations

A linear equation in the variables $x_1, x_2, \ldots, x_n$ is an equation that can be written in the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ where b and the coefficients $a_1, a_2, \ldots, a_n$ are real or complex numbers.

Example: $4x_1 - 5x_2 + 2 = x_1, x_2 = 2(6^{1/2} - x_1) + x_3$ are both linear. Not linear examples are $4x_1 - 5x_2 = x_1x_2, x_2 = 2(x_1^{1/2}) - 6$, and $2x_1^{-1} + \sin x_2 = 0$.

Systems of Linear Equations: an $(m \times n)$ system of linear equations is a system of m linear equations with n unknowns.

Example: the 2×3 system of equations below has a solution $x_1 = 5, x_2 = 6.5$, and $x_3 = 3$:

$$\begin{aligned} 2x_1 - x_2 + 1.5x_3 &= 8 \\ x_1 - 4x_3 &= -7 \end{aligned}$$

Two linear systems are called equivalent if they have the same solution set.

A system of linear equations can have: infinitely many solutions, no solution, or a unique solution. Coincident lines have infinitely many solutions, parallel lines have no solution, and intersecting lines have a unique solution.

Matrix Notation:

Let's say we have $x_1 - 2x_2 + x_3 = 0, 2x_2 - 8x_3 = 8$, and $5x_1 - 5x_3 = 10$.

The coefficient matrix is $\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$ and the augmented matrix is $\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$. The augmented matrix is 3×4 (3 rows, 4 columns).

To solve a linear system: if one of the following elementary operations is applied to a system of linear equations, the resulting system is equivalent, that is the resulting system has the same set of solutions as the original:

1. interchange two equations
2. multiply an equation by a non-zero scalar
3. add a constant multiple of one equation to another

Let's use the system

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

So using row operations and rearranging the rows, we can do $R_1 \leftrightarrow R_2$ to swap the first and second row. Now we can multiply the second line by $1/2$ so $1/2R_2$. The next step is to do $R_2 + R_1$.

The matrix that results from this is $\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$

Doing the operations $R_3 - 5R_1$, $R_3 - 5R_2$, and $1/2R_2$ and $1/30R_3$, we end up getting $\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$, so from this we have the equations

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

, so the solution set is $(1, 0, -1)$. Since a solution exists, the system is consistent.

Let's see if this one is consistent.

$$\begin{aligned} x_2 - 4x_3 &= 8 \\ 2x_1 - 3x_2 + 2x_3 &= 1 \\ 4x_1 - 8x_2 + 12x_3 &= 1 \end{aligned}$$

Doing the row operations $R_1 \leftrightarrow R_2$, $R_3 - 2R_1$, $R_3 + 2R_2$ we get that $0 = 15$, so this is inconsistent.

This last example has infinitely many solutions:

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

Doing the row operations $R_2 - 2R_1$ and $R + 3 + 3R_1$, then $R_3 + R_2$ and then $1/5R_2$ results in $\begin{bmatrix} 1 & -2 & 1 & -2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Solving this, we have x_3 with no restrictions, it is a "free parameter" and $x_2 = 1 - x_3$ and $x_1 = -x_3$ so x_1 and x_2 are parameterized by x_3.

1.2 Row Reduction and Echelon Forms

Leading entry of a row: the first (counting from left to right) non-zero entry (in a nonzero row)

Echelon Form: upper right-hand stair-case, triangle

1. all rows that consist entirely of zeros are grouped together at the bottom of the matrix
2. the first (counting from left to right) non-zero entry in the $(i + 1)$st row must appear in a column to the right of the first non-zero entry in the ith row.
3. all entries in a column below a leading entry are zeros

Reduced Echelon Form: an echelon Form matrix that also has the following properties:

1. the leading entry in each nonzero row is 1
2. each leading one is the only nonzero entry in its column

A pivot point: a location in a matrix A that corresponds to a leading 1 in a reduced echelon form of A. A pivot column is a column of A that contains a pivot position.

Steps to solving a system of linear equations:

1. begin with the leftmost nonzero column. This is the pivot column, the pivot position is at the top.
2. select a non zero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.
3. Use row replacement operations to create zeros in all positions below the pivot.
4. Cover (or ignore) the row containing the pivot position and cover all rows (if any) above it. Apply previous steps to the sub matrix that remains. Repeat until there are no more nonzero rows to modify.
5. Beginning with the rightomst pivot, create zeros above each pivot. Make each pivot equal to 1 by scaling.

Example

Determine the existence and uniqueness of the solution to

$$\begin{aligned} 3x_2 - 6x_3 + 6x_4 + 4x_5 &= -5 \\ 3x_1 - 7x_2 + 8x_3 - 5x_4 + 8x_5 &= 9 \\ 3x_1 - 9x_2 + 12x_3 - 9x_4 + 6x_5 &= 15 \end{aligned}$$

We can find that $x_5 = 4$, and x_3, x_4 are of infinite number of solutions.

Example

Find the general solution of the linear system whose augmented matrix has been reduced to $\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$

We have that $x_5 = 7$, $x_3 = 1/2(3 + 8x_4 + 7)$ and $x_1 = -4 - 6x_2 - 2(1/2)(3 + 8x_4 + 7) + 14$

Theorem 1.1: Existence and Uniqueness

A linear system is consistent if and only if the right most 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 $[0, \ldots, 0b]$ with b nonzero. If the system if consistent in the solution contains either a unique solution when there are no free variables or infinitely many solutions when there is at least one free variable.

1.3 Vector Equations

Vectors in $\mathbf{R}^2$: a matrix with only 1 column is called a vector. The set of all vectors with 2 entries is $\mathbf{R}^2$

$\vec{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ for example.

The vectors are odered pairs of real numbers:

$\vec{u}_1 = \begin{bmatrix} 3 \\ -1 \end{bmatrix} = (3, -1) \neq (-1, 3) = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

Vector addition: $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$

Geometric interpretation: parallelogram law.

$(2, 4) + (3, 1) = (2 + 3, 4 + 1) = (5, 5)$. We can create a parallelogram (google for review).

Let's say we subtract, $-a$ has the same magnitude as a but points in the opposite direction in this case. $(1, 3) - (2, 1) = (-1, 2)$. Drawing the line from the origin to the tip of the vectors give you what you get algebraically.

Vector multiplication: scalar multiplication: $a(x, y) = (ax, ay)$ where a is a scalar.

Geometric interpretation: $a(x, y, z)$ points in the same direction as (x, y, z) but is scaled by a factor of a.

Example

Let $\vec{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$ Find $4\vec{u}$, $-3\vec{v}$ and $4\vec{u} + (-3)\vec{v}/$

$4\vec{u} = (4, -8)$, $-3\vec{v} = (-6, 15)$ and $4\vec{u} + (-3)\vec{v} = (-2, 7)$

Vectors start at the origin and have magnitude and direction.

Representing vectors in $\mathbf{R}^3$. We add the z-axis.

Vectors in $\mathbf{R}^n$ we have that $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = (u_1, u_2, \dots, u_n)$ where $u_1, u_2, \dots \in \mathbb{R}$.

The $\vec{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ is the zero vector.

Algebraic Properties of $\mathbf{R}^n$: for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbf{R}^n$ and all scalars c and d:

- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- $\mathbf{u} + 0 = 0 + \mathbf{u} = \mathbf{u}$
- $\mathbf{u} + (-\mathbf{u}) = 0$
- $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
- $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
- $c(d\mathbf{u}) = (cd)\mathbf{u}$
- $1\mathbf{u} = \mathbf{u}$

Linear Combinations: given vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ in $\mathbf{R}^n$ and scalars $c_1, c_2, \dots, c_p$, the vector $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$ is called a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_p$ with weights $c_1, c_2, \dots, c_p$.

For example, if we have $3^{1/2}\mathbf{v}_1 + \mathbf{v}_2$ this can be written as $\vec{y} = \sqrt{3}\vec{v_1} + \vec{v_2}$ with $c_1 = \sqrt{3}$ and $c_2 = 1$.

Example

If $\vec{a_1} = (1, -2, -5), \vec{a_2} = (2, 5, 6)$, and $\vec{a_3} = (7, 4, -3)$ then determine if $\vec{b}$ can be written as a linear combination of $\vec{a_1}$ and $\vec{a_2}$. That is determine if there exists weights x_1, x_2 such that $x_1\vec{a_1} + x_2\vec{a_2} = \vec{b}$.

Using elementary row operations, we can determine that $x_1 = 3$, $x_2 = 2$ which is the linear combination of $\vec{a_1}$ and $\vec{a_2}$.

A vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots x_n\mathbf{x}_n = \mathbf{b}$ has the same solution set as the linear system with augmented matrix $[\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_n\mathbf{b}]$. In particular $\mathbf{b}$ can be generated by a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \dots \mathbf{a}_n$ if and only if there exists a solution to the linear system corresponding to the matrix $[\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_n\mathbf{b}]$

Span: if $\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_p$ are in $\mathbf{R}^n$ then the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ is denoted Span$\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_p\}$ and is called the subset of $\mathbf{R}^n$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_p$. That is, Span$\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_n\}$ is the collection of all vectors that can be written in the form: $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$ with $c_1, c_2, \dots, c_p$ scalars.

Asking if a vector $\mathbf{b}$ is in Span$\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_p\}$ amounts to asking whether the vector equation $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_p = \mathbf{b}$ has a solution, or equivalently whether the linear system with augmented matrix $[\mathbf{v}_1\mathbf{v}_2\mathbf{v}_p\mathbf{b}]$ has a solution.

Note Span$\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_p\}$ contains every scalar multiple of $\mathbf{v}_1$.

The span of a single vector is a line. The span of 2 linearly independent vectors is a plane (not scalar multiples of each other).

1.4 The Matrix Equation Ax = b

The Matrix Equation: if A is an $m \times n$ matrix with columns, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and if $\mathbf{x}$ is in $\mathbf{R}^n$, then the product of A and $\mathbf{x}$ denoted $A\mathbf{x}$ is the linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights.

Note: $\mathbf{A}x$ is defined only if the number of columns of A equals the numbers of entries in x.

For example: $\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix} = 4(1,0) + 3(2,-5) + 7(-1,3) = (3,6)$

Theorem 1.2: Matrix Equation, Vector Equation, System of Linear Equations

If A is an $m \times n$ matrix, with columns, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and if $\mathbf{b}$ is in $\mathbf{R}^m$, the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution as the vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ which in turn has the same solution as the system of linear equations represented by the augmented matrix $[\mathbf{a}_1\mathbf{a}_2 \ldots \mathbf{a}_n\mathbf{b}]$.

Existence of Solutions: the equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Example

Is $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & 7 \end{bmatrix}$ $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ Is the equation $A\vec{x} = \vec{b}$ consistent for all $\vec{b}$. Using rref, we get that $0 = -2b_1 + b_2 - 2b_3$, so it is not consistent for every $\vec{b}$. It is only consistent if $b_2 = 2b_1 + 2b_3$.

So let $\vec{b} = (1,4,1)$ and then do rref again and we get that x_3 is free, $x_2 = 1/7(4 - 5x_3)$ and $x_1 = 1 - 3(x_2) - 4x_3$ and this basically gives us $(1,4,1)$ too.

Theorem 1.3: Existence of soultion for $A\mathbf{x} = \mathbf{b}$

Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, they are all true statements or they are all false:

1. for each $\mathbf{b}$ in $\mathbf{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution
2. each $\mathbf{b}$ in $\mathbf{R}^m$ is a linear combination of the columns of A
3. The columns of A span $\mathbf{R}^m$
4. A has a pivot position in every row. Note: A is a coefficient matrix, not an augmented matrix.

Computation of $A\mathbf{x}$ - an efficient method (matrix multiplication): if the product $A\mathbf{x}$ is defined, then the ith entry in $A\mathbf{x}$ is the sum of the products of the corresponding entries from row i of A and from vector $\mathbf{x}$.

The above is trivial.

Properties of the Matrix-Vector Product $A\mathbf{x}$

Theorem 1.4

if A is an $m \times n$ matrix, $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbf{R}^n$, and c is a scalar, then :

1. $A(\mathbf{u} + \mathbf{v})$
2. $A(c\mathbf{u}) = c(A\mathbf{u})$

Algebraic Properties of $\mathbf{R}^n$: for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbf{R}^n$ and all scalars c and d. (This was from an above topic)

1.5 Solution Sets of Linear Systems

Parametric Vector Form of Solutions:

- Parametric Vector Form of a Plane: a plane can be expressed in explicit form, such as $10x_1 - 3x_2 - 2x_3 = 0$ or implicit form: $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$, for s and t scalars.
- Parametric Form of a Line containing point $\mathbf{p}$ in direction of $\mathbf{v}$: $l(t) = \mathbf{p} + t\mathbf{v}$

The parametric equation of a plane in $\mathbf{R}^2$: $\mathbf{x} = a\mathbf{v} + b\mathbf{s}$.

The span of 2 non-colinear vectors is a plane. Span$\{\mathbf{v}, \mathbf{s}\}$ = the $\mathbf{R}^2$ plane.

In $\mathbf{R}^3$, the Span is still a plane, just in $\mathbf{R}^3$.

The parametric equation of a line in $\mathbf{R}^2$: $\mathbf{l} = \mathbf{p} + t\mathbf{v}$.

Homogeneous Linear Equation - $A\mathbf{x} = 0$.

The homogeneous equation always has at least 1 solution, $\mathbf{x} = 0$ (the trivial solution).

Recall that a system of linear equation either has infinitely many solutions, no solution, or a unique solution.

The question is whether there exists a nontrivial solution (in which case there are infinitely many solutions).

- The homogeneous equation $A\mathbf{x} = 0$ has nontrivial solution if and only if the equation has at least 1 free variable

Description of solutions: if the solutions consists of:

- the 0 vector: Span$\{0\}$
- 1 free variable: Span$\{\mathbf{v}\}$, the solutions are a line through the origin
- 2 free variables, Span$\{\mathbf{v}_1, \mathbf{v}_2\}$ is a plane through the origin

Example

Determine if the following system has a nontrivial solution. Then describe the solution set

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0 \\ 6x_1 + x_2 - 8x_3 &= 0 \end{aligned}$$

x_3 is free. And everything is in the form $(4/3, 0, 1)$.

Nonhomogeneous Equation: $A\mathbf{x} = \mathbf{b}$.

For example, in the previous example when we let $x_3 = 0$ we get $(-1, 2, 0)$.

Theorem 1.5

Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some given $\mathbf{b}$ and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the Form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$ where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = 0$.

1.6 Applications of Linear Systems

There are three examples here: economics, chemical equations and network flow.

Start with economics. There exist equilibrium prices that can be assigned to the total outputs of the various sectors in an economy in such a way that the income of each sector exactly balances its expenses.

You can use row operations to find an equilibrium price.

Other examples run similarly (sorry for bad note taking today I'm sick)

1.7 Linear Independence

Linear Independence: an indexed set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ in $\mathbf{R}^n$ is said to be

- linearly independent if the vector equation $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = 0$ has only the trivial solution
- linearly dependent if there exists weights $c_1, c_2, \ldots, c_p$ not all zero such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = 0$

I'm too lazy to write matrices so much.

Linear Independence of Matrix Columns: the columns of matrix A are linearly independent if and only if the equation $A\mathbf{x} = 0$ has only the trivial solution.

Sets of One or Two Vectors

- A set with 1 vector is linearly independent iff $\mathbf{v}$ is not the 0 vector because $x_1\mathbf{v} = 0$ has only the trivial solution
- the zero vector, $\mathbf{0}$ is linearly dependent because $x_1\mathbf{0} = \mathbf{0}$ has many nontrivial solutions
- two vector $\{\mathbf{v}_1, \mathbf{v}_2\}$ are linearly dependent iff at least one of the vectors is a multiple of the other

Theorem 1.6

Characterization of Linearly Dependent Sets: An indexed set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ of 2 or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1$ is not $\mathbf{0}$, then some $\mathbf{v}_j$ with $j > 1$ is a linear combination of the preceding vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}$.

Note: the theorem does not say every vector in a linearly dependent set is a linear combination of preceding vectors

Theorem 1.7

If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ in $\mathbf{R}^n$ is linearly dependent if $p > n$.

Theorem 1.8

If a set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ in $\mathbf{R}^n$ contains the zero vector, then the set is linearly dependent.

1.8 Introduction to Linear Transformations

Linear Transformations: we can view $A\mathbf{x} = \mathbf{b}$ as a mapping: the $m \times n$ matrix A is the transform, $A : \mathbf{R}^n \to \mathbf{R}^m$.

From this point of view, solving the equation $A\mathbf{x} = \mathbf{b}$ amounts to finding all the vectors $\mathbf{x}$ in $\mathbf{R}^n$ that are transformed to $\mathbf{b}$ in $\mathbf{R}^m$. The correspondence from $\mathbf{x}$ to $A\mathbf{x}$ is a function from one set of vectors to another.

Definition

A transform (or function or mapping) T from $\mathbf{R}^n$ to $\mathbf{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbf{R}^n$ a vector $T(x)$ in $\mathbf{R}^m$. The set $\mathbf{R}^n$ is called the domain of T, and $\mathbf{R}^m$ is called the codomain. The notion $T : \mathbf{R}^n \to \mathbf{R}^m$ indicates that the domain of T is $\mathbf{R}^n$ and the codomain is $\mathbf{R}^m$. For $\mathbf{x}$ in $\mathbf{R}^n T(\mathbf{x})$ in $\mathbf{R}^m$ is called the image of $\mathbf{x}$. The set of all images $T(\mathbf{x})$ is called the range of T.

Matrix Transformations: $T(\mathbf{x})$ is computed as $A\mathbf{x}$ where A is an $m \times n$ matrix. Note: the domain of T is $\mathbf{R}^n$ and the codomain of T is $\mathbf{R}^m$. The range of T is the set of all linear combinations of the columns of A.

Linear Transformations: a transformation (or mapping) T is linear if

1. $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$ and $\mathbf{v}$ in the domain of T
2. $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and all $\mathbf{u}$ in the domain of T.

- every matrix transformation is a linear transformation: $A(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}) + A\mathbf{v}$ and $A(c\mathbf{u}) = cA(\mathbf{u})$
- linear transformations preserve the operations of vector addition and scalar multiplication

If T is a linear transformation, then

1. $T(\mathbf{0}) = \mathbf{0}$
2. $T(c\mathbf{u}+d\mathbf{v}) = cT(\mathbf{u})+dT(\mathbf{v})$ for all scalars c, d and all vectors $\mathbf{u}, \mathbf{v}$ in the domain of T. The generalization $T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_pT(\mathbf{v}_p)$ is known in engineering as the superposition principle: whenever an input is expressed as a linear combination of signals the systems response is the same linear combination of the responses to the individual signals.

1.9 The Matrix of a Linear Transformation

Goal: given a geometric desciprtion of a transformation, T, we want to find a "formula" for T

- Every linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ can be represented by a matrix transformation $A(\mathbf{x})$.
- The key to finding matrix A is to that T is completely determined by what it does to the columns of the $n \times n$ identiy matrix, I_n.

Theorem 1.9

Standard Matrix for a Linear Transformation: let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Then there exists a unique matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x}$ in $\mathbf{R}^n$. And, A is the $m \times n$ matrix whose jth column is the vector $T(\mathbf{e}_j)$ where $\mathbf{e}_j$ is jth column of the identity matrix in $\mathbf{R}^n$. $A = [T(\mathbf{e}_1)T(\mathbf{e}_2)\ldots T(\mathbf{e}_n)]$. A is caleld the standard matrix for the linear transformation T.

Onto/Existence: a mapping $T : \mathbf{R}^n \to \mathbf{R}^m$ is said to be onto $\mathbf{R}^m$ if each $\mathbf{b}$ in $\mathbf{R}^m$ is the image of at least 1 $\mathbf{x}$ in $\mathbf{R}^n$.

- T is onto $\mathbf{R}^m$ when the range of T is all of the codomain $\mathbf{R}^m$; for each $\mathbf{b}$ in $\mathbf{R}^m$, there exists at least one solution of $T(\mathbf{x}) = \mathbf{b}$. The mapping T is not onto when there is some $\mathbf{b}$ in $\mathbf{R}^m$ for which $T(\mathbf{x}) = \mathbf{b}$ has no solution.

T is one-to-one if for each $\mathbf{b}$ in $\mathbf{R}^n$, the equation $T(\mathbf{x}) = \mathbf{b}$ has either unique solution or no solution. The mapping is not one-to-one when some $\mathbf{b}$ in $\mathbf{R}^m$ is the image of more than one vector in $\mathbf{R}^n$.

Theorem 1.10

Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation, then T is one-to-one iff $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Theorem 1.11

Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation and let A be the standard matrix for T. Then:

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

1.10 Linear Models in Business, Science, and Engineering

Linear Equations can be done in electrical networks.

Current flow in a simple electrical network can be described by a system of linear equations. Consider Ohm's Law, $V = IR$, which describes the current which passes through a resistor. The algebraic sum of IR voltage drops in one direction around a loop equals the algebraic of the voltage sources in the same direction around the loop.

The model for current flow is linear since the voltage drop across a resistor is proportional to the current flowing through it, and the sum of the voltage drops in a loop equals the sum of the voltage sources in the loop.

For difference equations, if there is a matrix A such that $\mathbf{x}_1 = A\mathbf{x}_0, x_2 = A\mathbf{x}_1$, and in general $\mathbf{x}_{k+1} = A\mathbf{x}_k$ for $k = 0, 1, 2, \ldots.$ then this is called a linear difference equation (or recurrence relation).

Ok whatever just use logic.

2 Matrix Algebra

2.1 Matrix Operations

Sums and Scalar Multiples of Matrices: if A and B are $m \times n$ matrices, $A+B$ is the $m \times n$ whose columns are the sums of the corresponding columns in A and B, the scalar multiple rA is the matrix whose columns are r times the corresponding columns in A.

Theorem 2.1

Matrix addition and scalar multiplication: Let A, B, and C be matrices of the same size, and let r and s be scalars.

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $A + 0 = A$
4. $r(A + B) = rA = rB$
5. $(r + s)A = rA = sA$
6. $r(sA) = (rs)A$

Matrix Multiplication: if A is an $m \times n$ matrix and B is an $n \times p$ matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \dots \mathbf{b}_p$, then the product AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p$, i.e, $AB = A[\mathbf{b}_1\mathbf{b}_2 \dots \mathbf{b}_p] = [A\mathbf{b}_1 A\mathbf{b}_2 \dots A\mathbf{b}_p]$. Matrix multiplication corresponds to composition of linear transformations.

- An efficient Matrix Multpilcation: if the product AB is defined, then the entry in row i and column j of AB is the sum of the products of corresponding entries from row i of A and column j of B. If $(AB)_{ij}$ denotes the (i, j)th entry in AB, and if A is $m \times n$, then $(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$.

Properties of Matrix Multpilcation: Let A be $m \times n$ and let B, C have sizes such that the sums and products are defined:

1. $A(BC) = (AB)C$ associative law
2. $A(B + C) = AB + AC$ left distributive law
3. $(B + C)A = BA + CA$ right distributive law
4. $r(AB) = (rA)B = A(rB)$ for any scalar r
5. $I_m A = A = AI_n$ Identity matrix for multiplication

- Matrix mutiplication is not commutative. In general AB does not equal BA.
- Cancellation laws do not hold for matrix multiplication.
- If $AB = 0$, you cannot conclude either $A = 0$ or $B = 0$.

Powers of a Matrix: If A is $n \times n$ and k is a positive integer, A^k denote the product of k copies of A.

The Transpose of a Matrix: given an $m \times n$ matrix A, the transpose of A is the $n \times m$ matrix whose columns are formed from the corresponding rows of A.

Theorem 2.2: Transpose

Let A, B denote matrices whose sizes are appropriate for the following:

1. $(A^T)^T = A$

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

2.2 The Inverse of a Matrix

The Matrix Inverse is the matrix analogue of the multiplicative inverse of in real numbers.

- Invertible: an $n \times n$ matrix A is said to be invertible if there is an $n \times n$ matrix A^{-1} such that $A^{-1}A = AA^{-1} = I_n$. In this case A^{-1} is said to be the unique inverse of A.

Notice: because matrix multipilcation is not commutative, both equations are needed.

Singular Matrix: A matrix that is not invertible is a single matrix. An invertible matrix is nonsingular.

Theorem 2.3

Inverse of a 2×2: Let A be the 2×2 matrix shown. If $ab - dc$ is not zero, then A is invertible with A^{-1} as shown:

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

Determinant: det $A = ad - bc$. The theorem says that a 2×2 matrix is invertible iff det A is not zero.

Theorem 2.4

If A is an invertible matrix, then for each $\mathbf{b}$ in $\mathbf{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique soultion $\mathbf{x} = A^{-1}\mathbf{b}$.

Theorem 2.5

1. If A is an invertible matrix, then A^{-1} is invertible and $(A^{-1})^{-1} = A$
2. If A and B are $n \times n$ invertible matrices, then so is AB and $(AB)^{-1} = B^{-1}A^{-1}$. Generalization: the product of $n \times n$ invertible matrices is invertible, and the inverse is the product of the their inverse in the reverse order.
3. If A is an invertible matrix, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$

Theorem 2.6

An $n \times n$ matrix is invertible iff it is row equivalent to I_n, and any sequence of elementary row operations that reduces A to I_n also transforms I_n to A^{-1}.

2.3 Characterizations of Invertible Matrices

The Invertible Matrix Theorem: let A be an $n \times n$ matrix. Then the following statements are equivalent.

- A is an invertible matrix.
- A is row equivalent to the $n \times n$ identity matrix.
- A has n pivot positions
- the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution
- the columns of A form a linearly independent set
- the linear transform $\mathbf{x} \to A\mathbf{x}$ is one-to-one

- the equation $A\mathbf{x} = \mathbf{b}$ has at least 1 soln for each $\mathbf{b}$ in $\mathbf{R}^n$
- the columns of A span $\mathbf{R}^n$
- the linear transformation $\mathbf{x} \to A\mathbf{x}$ maps $\mathbf{R}^n$ onto $\mathbf{R}^n$
- there is an $n \times n$ matrix C such that $CA = I$
- there is an $n \times n$ matrix D such that $AD = I$
- A^T is an invertible matrix

Note that this only applies to square matrices.

Theorem 2.7: Inverse Transformation

Let $T : \mathbf{R}^n \to \mathbf{R}^n$ 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 $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique solution satisfying $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbf{R}^n$.

Recall that matrix multipilcation corresponds to composition of linear transformations. When a matrix A is invertible, the equation $A^{-1}A\mathbf{x} = \mathbf{x}$ can be viewed as a statement about linear transformations. A linear transformation $T : \mathbf{R}^n \to \mathbf{R}^n$ is said to be invertible if there exists a function $S : \mathbf{R}^n \to \mathbf{R}^n$ such that $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbf{R}^n$.

2.4 Matrix Factorizations

A factorization of a matrix A is an equation that expresses A as a product of two or more matrices.

Whereas matrix multiplication involves a synthesis of data (combining the effects of two or more linear transformations into a single matrix), matrix factorization is an analysis of data.

The LU factorization:

At first assume that A is an $m \times n$ matrix that can be row reduced to echelon form, without row interchanges. Then A can be written in the form $A = LU$, where L is an $m \times m$ lower triangular matrix with 1's on the diagonal and U is an $m \times n$ echelon form of A.

Suppose A can be reduced to an echelon form U using only row replacements that add a multiple of one row to another below it. In this case, there exist unit lower triangular elementary matrices, $E_1, \ldots, E_p$ such that $E_p \cdots E_1 A = U$.

Then $A = (E_p \cdots E_1)^{-1}U = LU$ where $L = (E_p \cdots E_1)^{-1}$.

It can be shown that products and inverses of unit lower triangular matrices are also unit lower triangular. Thus L is unit lower triangular.

Algorithm:

- Reduce A to an echelon form U by a sequence of row replacement operations, if possible.
- Place entries in L such that the same sequence of row operations reduces L to I.

3 Determinants

3.1 Introduction to Determinants

Definition of the Determinant: given an $n \times n$ matrix $A = [a_{ij}]$, the determinant is det A $= a_{11}a_{22} - a_{12}a_{21}$.

For $n \geq 2$, the determinant of an A is the sum of n terms of the form $+/-a_{1j}A_{1j}$ with the plus and minus signs alternating, where the entries $a_{11}, a_{12}, \ldots, a_{1n}$ are from the first row of A.

The (i,j)-cofactor $= C_{ij} = (-1)^{i+j}$ det A$_{ij}$ s.t. det A $= a_{11}C_{11} + a_{12} + C_{12} + \cdots + a_{1n} + C_{1n}$.

Theorem 3.1

The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion across any row or down any column/

The expansion across the ith row using the cofactors is: det A $= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}Cin$

The expansion across the jth column is: det A $= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + n_{nj}C_{nj}$.

This theorem is helpful for computing determinants of a matrix that contains many zeros. For example, if 1 row contains many zeros, than a cofactor expansion across that row will be easier to calculate.

Theorem 3.2

If a is a triangular matrix, then det A is the product of the entries on the main diagonal of A.

3.2 Properties of Determinants

"The secret of determinants lies in how they change when row operations are performed"

Theorem 3.3

Let A be a square matrix.

- A multiple of one row of A is added to another row to produce a matrix B, then det $B =$ det A
- if two rows of A are interchanged to produce B then det $B =$ -det A
- if 1 row of A is multiplied by k to produce B, then det $B = k$det A

we can use a strategy to reduce a matrix to echelon form and then use the fact that the determinant of a triangular matrix is the product of its diagonal entries.

Theorem 3.4

A square matrix A is invertible if and only if det A is not zero. If A is an $n \times n$ matrix, then det $A^T =$ det A.

Theorem 3.5

If A and B are $n \times n$ matrices, then det $AB =$ (det A)(det B)

3.3 Cramer's Rule, Volume, and Linear Transformations

Cramer's Rule: Cramers rule is needed for a variety of theoretical calculations. However the formula is inefficient for hand calculations except for 2×2.

For any $n \times n$ matrix A and any **b** in $\mathbf{R}^n$ let $A_1(\mathbf{b})$ be the matrix obtained from A by replacing the ith column by **b**, $A_i(\mathbf{b}) = [\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_{i-1}\mathbf{b}\mathbf{a}_{i+1} \dots \mathbf{a}_n]$

Theorem 3.6

Let A be an intertible invertible $n \times n$ matrix. For any **b** in $\mathbf{R}^n$, the unique solution **x** of $A\mathbf{x} = \mathbf{b}$ has entries given by: $\mathbf{x}_i = (\text{det } A_i(\mathbf{B})/(\text{det } A))$ $i = 1, 2, \dots, n$.

Application to engineering: A number of important engineering problems can be analyzed by Laplace transformations. This approach converts an appropriate system of linear differential equations into a system of linear algebraic equations whose coefficient involve a parameter.

Formula for A^{-1}: Cramer's Rule leads to a general formula for the inverse of an $n \times n$ matrix.

Theorem 3.7

Let A be an invertible $n \times n$ matrix. Then A^{-1} is given by: (Google This)

Determinants as Area of Volume

Theorem 3.8

If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is —det A—. If A is a 3×3 matrix, the volume of the paralellepiped determined by the columns is —det A—.

Theorem 3.9

Let $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be the linear the transformation obtained by a 2×2 matrix A. If S is a parallelogram in $\mathbf{R}^2$ then the area of $T(S)$ is equal to —det A— times the area of S.

If T is determined by a 3×3 matrix A, and S is a paralellepiped in $\mathbf{R}^3$, then the volume of $T(S)$ is equal to the area of —det A— times the volume of S.

4 Vector Spaces

4.1 Vector Spaces and Subspaces

A vector space is a nonempety set V of objects, called vectors, on which are defined two operations called addition and scalar multiplication to the 10 axioms listed below. The axioms must hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and for all scalars a and b.

Closure Properties:

- $\mathbf{u} + \mathbf{v}$ is a vector in V
- $a\mathbf{v}$ is a vector in V

Properties of Addition:

- commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- associative: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- additive identity: there is a vector $\mathbf{0}$ in V such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v}$ in V
- additive inverse: given a vector $\mathbf{v}$ in V, there is a vector $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$

Properties of Scalar Multiplication:

- associative: $a(b\mathbf{v}) = (ab)\mathbf{v}$
- distributive: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
- distributive: $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$
- multiplicative identity: $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v}$ in V

A subspace of a vector space, V, is a subset H of V that has three properties:

1. the zero vector of V is H
2. H is closed under vector addition
3. H is closed under scalar multiplication

A subspace H of V itself is a vector space. The other properties of a vector space are "inherited" since H is a subset of V.

Theorem 4.1

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are in a vector space V, then Span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is a subspace of V.

4.2 Null Spaces, Column Spaces, Row Spaces, and Linear Transformations

The Null Space of a $m \times n$ matrix A, denoted Nul A, is the set of solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation: Nul $A = \{\mathbf{x} : \mathbf{x}$ is in $\mathbf{R}^n$ and $A\mathbf{x} = \mathbf{0}\}$.

Theorem 4.2

The null space of an $m \times n$ matrix A is a subspace of $\mathbf{R}^n$. Equivalently, the set of all solutions to a system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknown is a subspace of $\mathbf{R}^n$.

The Column Space of a $m \times n$ matrix A, denoted Col A, is the set of all linear combinations of the columns of A. If $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ then Col A = Span$\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$.

Theorem 4.3

The column space of a $m \times n$ matrix A is a subspace of $\mathbf{R}^m$.

The Row Space of an $m \times n$ matrix A, denoted Row A, is the set of all linear combinations of the row vectors. Each row has n entries, so the Row A is a subspace of $\mathbf{R}^n$.

The null and column space are related. They are very different: When A Is not square, the column space and the null space exist entirely in different "universes". For an $m \times n$ matrix, the Column Space is in m-dimensional space, where the Null space is in n-dimensional space.

Kernel and Range if a Linear Transformation

- a Linear Transformation T from a vector space V to be vector space W is a rule that assigns each vector $\mathbf{x}$ in V a unique vector $T(\mathbf{x})$ in W such that $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ and $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u}, \mathbf{v}$ in V and c in $\mathbf{R}$.
- The Kernel of T is the set of all $\mathbf{u}$ in V such that $T(\mathbf{u}) = \mathbf{0}$
- The Range of T is the set of all vectors in W of the form $T(\mathbf{x})$ for some $\mathbf{x}$ in V
- The kernel and range of T are subspaces of V

4.3 Linearly Independent Sets; Bases

A basis spans a vector space as "efficiently" as possible.

- linearly independent: an indexed set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ in V is linearly independent if $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ has only the trivial solution.
- linearly dependent: an indexed set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ in V is linearly dependent if $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ has a nontrivial solution

Theorem 4.4

An indexed set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ of 2 or more vectors with $\mathbf{v}_1$ not zero, is linearly dependent if and only if some $\mathbf{v}_j (j > 1)$ is a linear combination of the preceding vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{j-1}$.

This definition is for general vectors where we may be unable to write the vectors as columns of a matrix A.

Basis: Let H be a subspace of a vector space V. A set of vectors $\mathcal{B}$ in V is a basis of H if:

1. $\mathcal{B}$ is a linearly independent set and
2. the subspace spanned by $\mathcal{B} = H$ (or H = span $\mathcal{B}$)

Theorem 4.5

$S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ is a set in a vector space, and let H = Span$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$.

- if one of the vectors in S, say $\mathbf{c}_k$, is a linear combination of the remaining vectors in S, then the set formed from S by removing $\mathbf{v}_k$ still spans H.
- if H is not equal to $\{\mathbf{0}\}$, some subset of S is a basis for H

The basis if the smallest possible spanning set (because all vectors are linearly independent)

The basis is the largest possible linearly independent set that spans (an additional vector will make the set dependent)

4.4 Coordinate Systems

Theorem 4.6

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Then for each $\mathbf{x}$ in V, there exists a unique set of scalars $c_1, c_2, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$.

Coordinates: suppose $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. The coordaintes of $\mathbf{x}$ relative to the basis $\mathcal{B}$ are the weights $c_1, c_2, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$.

The change of coordinates matrix, $P_{\mathcal{B}} = [\mathbf{b}_1\mathbf{b}_2 \ldots \mathbf{b}_n]$ change coordinates from $\mathcal{B}$ to the standard basis in $\mathbf{R}^n$. $P_{\mathcal{B}}^{-1}$ transforms $\mathbf{x}$ to $[\mathbf{x}]_{\mathcal{B}}$.

Theorem 4.7

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Then the coordinate mapping $\mathbf{x} - > [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one, onto linear transformation from V to $\mathbf{R}^n$.

The coordinate mapping in this theorem is an important example of isomorphism from V onto $\mathbf{R}^n$. The notation and terminology for the two vector space may be different, but the two spaces are indistinguishable. Every vector space calculation in V is accurately reproduced in $\mathbf{R}^n$ and vice versa.

4.5 The Dimension of a Vector Space

Theorem 4.8

If a vector space V has a basis $\mathcal{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n]$ then any set in V containing more than n vectors must be linearly dependent.

Theorem 4.9

If a vector space V has a basis $\mathcal{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n]$ then every basis of V must contain exactly n vectors.

Dimension: if a vector space V is spanned by a finite space, then V is said to be finite-dimensional, and the dimension of V, dim V is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be infinite-dimensional.

Theorem 4.10

Let H be a subspace of a finite dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to be a basis for H. Also, H is finite dimensional and dim H $\leq$ dim V

Theorem 4.11

Let V be a p-dimensional vector space with $p \geq 1$. Any linearly independent set of exactly p elements in V is a basis for V. Any set of exactly p elements that spans V is a basis for V

The rank of an $m \times n$ matrix A is the dimension of the column space, and the nullity of A is the dimension of the null space.

Theorem 4.12

The dimension of the column space and the null space of an $m \times n$ matrix A satisfies the equation: rank A + nullity A = number of columns in A

The rank of an $m \times n$ matrix A is the number of pivot columns and nullity of A is the number of free variables. Since the dimension of the row space is the number of pivot rows, dim row space = rank A.

The following are equivalent to the statement A is an invertible matrix:

- the columns of A for a basis for $\mathbf{R}^n$
- Col $A = \mathbf{R}^n$
- rank $A = n$
- nullity $A = 0$
- Nul A $= \{0\}$

4.6 Change of Basis

Theorem 4.13

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ and $\mathcal{B}' = \{\mathbf{b}'_1, \mathbf{b}'_2, \ldots, \mathbf{b}'_n\}$ be bases of a vector space V.

Then there is a unique $n \times n$ matrix $\mathbf{P}_{\mathcal{B} \leftarrow \mathcal{B}}$ such that $[\mathbf{x}]_{\mathcal{B}} = \mathbf{P}_{\mathcal{B} \leftarrow \mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$. The columns of $\mathbf{P}_{\mathcal{B} \leftarrow \mathcal{B}}$ are the $\mathcal{B}'$ coordinates vectors of the vectors in the basis $\mathcal{B}$.

If $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ and $\mathcal{E}$ is the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ then $[\mathbf{b}_1]_{\mathcal{E}} = \mathbf{b}_1$, and likewise for the other vectors in $\mathcal{B}$.

5 Eigenvalues and Eigenvectors

5.1 Eigenvectors and Eigenvalues

Eigenvectors and Eigenvalues: viewing a matrix as a linear transformation $x \to A\mathbf{x}$ vectors that are only scaled (not rotated) are called eigenvectors. The scaling factor is the associated eigenvalue.

Definition

An eigenvector of an $n \times n$ matrix A is a nonzero vector $\vec{x}$ such that

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

is true for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution $\vec{x}$ of the above equation. We often state that $\vec{x}$ is an eigenvector corresponding to λ.

The equation you will solve to find eigenvectors: $A\vec{x} = \lambda\vec{x} \implies A\vec{x} - \lambda\vec{x} = \vec{0} \implies (A - \lambda I)\vec{x} = \vec{0}$.

The set of solutions to this homogeneous equation is just the nullspace of the matrix, so this set is a subspace of $\mathbf{R}^n$ and is caleld the eigenspace corresponding to λ.

Theorem 5.1

The eigenvalues of a triangular matrix are the entries on its main diagonal.

Theorem 5.2

If $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ are eigenvectors that correspond to distinct eigenvalues of an $n \times n$ matrix, then the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ is linearly independent.

Eigenvectors and Difference Equations: constructing a solution of the first-order difference equation: $x_{k+1} = A\mathbf{x}_k(k = 0, 1, 2, \dots)$.

If A is $n \times n$ then $\mathbf{x}_{k+1}(k = 0, 1, 2, \dots)$ is a recursive description of a sequence $\{\mathbf{x}_k\}$ in $\mathbf{R}^n$.

A solution is an explicit description of $\{\mathbf{x}_k\}$ whose formula for each $\mathbf{x}_k$ does not depend directly on A or the preceding terms in the sequence other than the initial term $\mathbf{x}_0$.

The simplest way to build a solution is to take an eigenvector $\mathbf{x}_0$ and it's corresponding eigenvalue and let $\vec{x}_k = \lambda^k x$.

This sequence is a solution because: $A\vec{x}_k = A(\lambda^k\vec{x}_0) = \lambda^k(A\vec{x}_0) = \lambda^k(\lambda\vec{x}_0) = \lambda^{k+1}\vec{x}_0 = \vec{x}_{k+1}$.

5.2 The Characteristic Equation

Theorem 5.3

An $n \times n$ matrix A is invertible if and only if 0 is not an eigenvalue of A,

The Characteristic Equation: a scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the characteristic equation: det $(A - \lambda I) = 0$.

The characteristic equation of an $N \times n$ matrix is an nth degree polynomial. We expect exactly n roots, counting multiplicities, provided complex roots are allowed.

Similarity: If A and B are $n \times n$ matrices, then A is similar to B if there is an invertible matrix P such that $P^{-1} = AP = B$ or equivalently $A = PBP^{-1}$. Changing A to $P^{-1}AP$ is called a similarity transformation.

Theorem 5.4

If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues with the same multiplicities.

Similarity is not the same as row equivalence. If matrices have the same eigenvalues, they may not be similar.

5.3 Diagonalization

A square matrix A is said to be diagonalizable if A is similar to a diagonal matrix, that is if $A = PDP^{-1}$ for some invertible matrix P and some diagonal, matrix D.

Theorem 5.5

An $n \times n$ 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. We call such a basis an eigenvector basiso f.

Theorem 5.6

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Theorem 5.7

Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \dots, \lambda_p$.

For $1 \leq k \leq p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.

The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if the characteristic polynomial factors completely into linear factors and the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

If A is diagonalizable and B_k is a bsis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $B_1, \dots, B_p$ forms and eigenvector basis.

5.4 Applications to Differential Equations

Systems of Differential Equations: Several quantities are varying continuously in time and related by a linear system of differential equations:

Linear $\mathbf{x}'(t) = A\mathbf{x}(t)$ where $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$.

A solution of $\mathbf{x}'(t) = A\mathbf{x}(t)$ is a vector-valued function that satisfies $\mathbf{x}'(t) = A\mathbf{x}(t)$ for all t in some interval of real numbers such as $t \geq 0$.

From Differential Equations, we know there always exist what is called a Fundamental Set of Solutions to $\mathbf{x}'(t) = A\mathbf{x}(t)$ A fundamental set if a basis for the set of all solutions of $\mathbf{x}'(t) = A\mathbf{x}(t)$, and the solution set is an n-dimensional vector space of functions.

If a vector $\mathbf{x}_0$ specified, then the initial value problem is to construct the unique function $\mathbf{x}(t)$ such that $\mathbf{x}'(t) = A\mathbf{x}(t)$ and $\mathbf{x}(0) = \mathbf{x}_0$.

The 2 real solutions to $\mathbf{x}'(t) = A\mathbf{x}(t)$ are: $\mathbf{y}_1(t) = \text{Re}\mathbf{x}_1(t) = [(\text{Re}\mathbf{v})\cos(bt) - (\text{Im}\mathbf{v})\cos(bt)]e^{at}$. and $\mathbf{y}_2(t)$ is basically the same but it is Im $\mathbf{x}_1(t)$.

6 Orthogonality and Least Sqaures

6.1 Inner Product, Length, and Orthogonality

Inner Product (generalization of the dot product of vectors in $\mathbf{R}^n$)

Inner Product/Dot Product of vectors in $\mathbf{R}^n$: $\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$

Theorem 6.1

Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be vectors, let c be a scalar, then an inner product is a function assigns a scalar to each pair of vector $\mathbf{u}$ and $\mathbf{v}$ satisfies:

- $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
- $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$
- $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (c\vec{v})$
- $\vec{u} \cdot \vec{u} = 0$ iff $\vec{u} = \vec{0}$

The Length/Norm of a vector of $\mathbf{v}$ is the nonnegative scalar $||\vec{v}|| = \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$ and $||\vec{v}||^2 = \vec{v} \cdot \vec{v}$

- A vector whose length is 1 is called a unit vector. If we divide a nonzero vector $\mathbf{v}$ by its length, that is, multiply by $1/||\mathbf{v}||$, we obtain a unit vector this process is called normalizing (direction is preserved)

Distance in $\mathbf{R}^n$: for $\mathbf{u}$ and $\mathbf{v}$ in $\mathbf{R}^n$, the distance between u and v, written dist$(\mathbf{u}, \mathbf{v})$ is the length of the vector $\mathbf{u} - \mathbf{v}$, that is dist$(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$

- In $\mathbf{R}^2$ and $\mathbf{R}^3$ this definition coincides with the usual formulas for Euclidean distance between 2 points

Orthogonality of vectors in $\mathbf{R}^n$ is the generalization of the concept of perpendicular lines in ordinary Euclidean geometry.

Def: two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbf{R}^n$ are orthogonal (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

Note the zero vector is ortohgonal to every vector in $\mathbf{R}^n$

Pythagorean theorem: 2 vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$.

Orthogonal Complements: used in SVD

If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbf{R}^n$, then $\mathbf{z}$ is said to be ortohgonal to W

The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the ortohgonal complement of W denoted $W^\perp$ read as "W perpendicular" or "W perp"

Properties/facts

- a vector $\mathbf{x}$ is in W perp if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans W
- W perp is a subspace of $\mathbf{R}^n$

Theorem 6.2

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. (Row $A)^\perp$ = Nul A and (Col $A)^\perp$ = Nul A^T.

6.2 Orthogonal Sets

Orthogonal Sets: a set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, dots, \mathbf{u}_p\}$ in $\mathbf{R}^n$ is said to be an orthogonal set if each pair of distinct vectors from theset is orthogonal.

Theorem 6.3

If $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ is an orthogonal set of non-zero vectors in $\mathbf{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

Definition

An Orthogonal Basis for a subspace W of $\mathbf{R}^n$ is a basis for W that is also an orthogonal set.

Theorem 6.4

Let $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbf{R}^n$. For each vector $\mathbf{y}$ in W, the weights in the linear combination $\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p$ are:

$$c_j = \frac{\vec{v} \cdot \vec{u}_j}{\vec{u}_j \cdot \vec{u}_j} \qquad j = 1, 2, \dots, p$$

This formula is why an orthogonal basis is much nicer than others.

Orthogonal projection: given a nonzero vector $\mathbf{u}$ in $\mathbf{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbf{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$.

$\hat{y}$ ios the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$, the vector $|textbfz$ is the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$.

Geometric Interpreation THeorem for finding coordinates of an orthogonal basis: The theorem decomposes vector $\mathbf{y}$ into a sum of orthogonal productions onto one-dimensiona subspaces.

Decomposing a Force into Component Forces: occurs in physics

Orthonormal Sets: a set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ is an orthonormal set if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ is an orthonormal basis for W¡ since the set is automatically linearly independent.

Matrices whose columns form an orthonormal set are important in applications and computer algorithms for matrix computations. Their main properties are given in the following two theorems:

Theorem 6.5

An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

Theorem 6.6

Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in R^n then:

- $||U\vec{x}|| = ||\vec{x}||$
- $(U\vec{x}) \cdot (U\vec{y}) = \vec{x} \cdot \vec{y}$
- $(U\vec{x}) \cdot (U\vec{y}) = 0$ iff $\vec{x} \cdot \vec{y} = 0$

An orthogonal matrix is a square invertible matrix U, $U^{-1} = U^T$.

6.3 Orthogonal Projections

The Orthogonal Projection: Given a vector $\mathbf{y}$ and a subspace W in $\mathbf{R}^n$ there is a vector $\hat{\mathbf{y}}$ in W such that

- $\hat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$
- $\hat{\mathbf{y}}$ is the unique vector for which $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W.

These two properties of $\vec{\mathbf{y}}$ provide the key to finding the least squares solution to linear systems.

Theorem 6.7

Let W be a subspace of $\mathbf{R}^n$. Then each $\mathbf{y}$ in $\mathbf{R}^n$ can be written uniquely in the form: $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\vec{\mathbf{y}}$ is in W and z is in $W^{\perp}$. In fact, if $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{y} = \frac{\vec{v} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \cdots + \frac{\vec{v} \cdot \vec{u}_p}{\vec{u}_p \cdot \vec{u}_p}\vec{u}_p \text{ and } \vec{z} = \vec{y} \cdot \hat{y}$$

The theorem tells us decomposition of $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$ can be computed without having an orthogonal basis for $\mathbf{R}^n$. It is enough to haev an orthogonal basis only for W.

Geometric Interpretation of the Orthogonal Projection: the orthogonal projection $\hat{y}$ of $\mathbf{y}$ onto W is the sum of the projections of $\mathbf{y}$ onto one-dimensional subspaces that are orthogonal to each other.

Properties of Orthogonal Projections

Theorem 6.8

Let W be a subspace of $\mathbf{R}^n$, let $\mathbf{y}$ be any vector in $\mathbf{R}^n$ and $\hat{y}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{y}$ is the closest point in W to $\mathbf{y}$ in the sense that $||\mathbf{y} - \hat{\mathbf{y}}|| < ||\mathbf{y} - \mathbf{v}||$ for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

Theorem 6.9

If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbf{R}^n$, then $\text{proj}_w\mathbf{y} = (\vec{y} \cdot \vec{u}_1)\vec{u}_1 + (\vec{y} \cdot \vec{u}_2)\vec{u}_2 + \cdots + (\vec{y} \cdot \vec{u}_p)\vec{u}_p$.

If $U = [\mathbf{u}_1\mathbf{u}_2 \ldots \mathbf{u}_p]$ then $\text{proj}_w\mathbf{y} = UU^T\mathbf{y}$ for all $\mathbf{y}$ in $\mathbf{R}^n$.

6.4 The Gram-Schmidt Process

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

Theorem 6.10

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a nonzero subspace of $\mathbf{R}^n$, define:

- $\vec{v}_1 = \vec{x}_1$
- $\vec{v}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1$
- $\vec{v}_3 = \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \frac{\vec{x}_2 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2$

Orthonormal Basis: when working problems by hand, it is easier to normalize each $\mathbf{v}_k$ as they are foud.

QR Factorization of Matrices: if an $m \times n$ matrix A has linearly independent columns $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ then applying the Gram-Schmidt process with normalizations to $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ amounts to factoring A as described in the following theorem and is used widely in computer algorithms.

Theorem 6.11

If A is an $m \times n$ matrix with linearly independent columns then A can be factored as $A = QR$ where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular matrix with positive entities on the diagonal.

When the Gram Schmidt process is run on the computer, a round off error can build up as the vectors are calculated, one by one. For j and k large but unequal, the innter product may not be sufficiently close to zero. A different computer based QR factorization is usually preferred to the modified Graham Schmidt Method because it yields a more accurate orthogonal basis, even though the factorization requires about twice as much arithmetic.

6.5 Least-Squares Problems

The Least Squares Problem: given $A\mathbf{x} = \mathbf{b}$ that is possibly inconsistent, find an $\mathbf{x}$ that makes $||\mathbf{b} - A\mathbf{x}||$ as small as possible.

Definition

If A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbf{R}^n$, a least-squares solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x}$ in $\mathbf{R}^n$ such that $||\mathbf{b} - A\mathbf{x}|| \leq ||\mathbf{b} - A\mathbf{x}'||$ for all $\mathbf{x}'$ in $\mathbf{R}^n$

Notice: $A\mathbf{x}$ is in the column space of A, Col A, so we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point to in Col A to $\mathbf{b}$.

Theorem 6.12

The set of least squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation $A^T A\hat{\mathbf{x}} = A^T\hat{\mathbf{b}}$

Note: if there is a free variable, the least squares solution may not be unique

Deriving the Normal Equations for $A\mathbf{x} = \mathbf{b}$:

1. Note $A\mathbf{x}$ is in the Col A, therefore the $\mathbf{b}$ associated least squares solution of $A\mathbf{x} = \mathbf{b}$ is in Col A
2. Use the Best Approximation Theorem to determine the solution of the Least Square Problem is the orthogonal projection of $\mathbf{b}$ onto Col A, $\hat{b} = \text{proj}_{\text{Col} A}\mathbf{b}$
3. Note: for the Least Squares Problem: we are looking for $\hat{\mathbf{x}}$ that satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$
4. use the Orthogonal Decomposition Theorem to find the vector $\mathbf{z}$ orthogonal to $\hat{\mathbf{b}}$, $\mathbf{z} = \mathbf{b} - \hat{\mathbf{b}}$
5. use the fact that $\mathbf{z}$ must be orthogonal to the Col A, and therefore any column vector of A, $\mathbf{a}_i$ is orthogonal to $\mathbf{z} = \mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}} \implies a_i \cdot (\mathbf{b} - A\hat{\mathbf{x}} = 0)$

Theorem 6.13

Let A be an $m \times n$ matrix. The following statements are logically equivalent

1. the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-square solution for each $\mathbf{b}$ in $\mathbf{R}^n$
2. the columns of A are linearly independent
3. the matrix $A^T A$ is invertible

When these statements are true, the least squares solution $\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}$

6.6 Machine Learning and Linear Models

Machine learning uses linear models in sitautions where the machine is being trained to predict the outcome based on the values of hte inputs.

The machine is given a set of training data where the values of the independent and dependent variables are known.

The least squares line is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals.

This line is also called a line of regression of y on x. The coefficients β_0, β_1 of the line are called regression coefficients.

In general a linear model will arise whenever y is to be predicted by an equation of the form

$$y = \beta_0 f(0)(u,v) + \beta_1 f_1(u,v) + \cdots + \beta_k f_k(u,v)$$

with $f_0, \ldots, f_k$ any sort of known functions and $\beta_0, \beta_1, \ldots, \beta_k$ unknown weights.

6.7 Inner Product Spaces

Definition

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

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

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

7 Symmetric Matrices and Quadratic Forms

7.1 Diagonalization of Symmetric Matrices

7.2 Quadratic Forms