Midterm 3

Section 5.3: Diagonalization

Definitions

- Diagonal matrix
  - A matrix where the only **non-zero entries** are on the **main diagonal**
  - Everywhere else is 0's
- Similar matrices
  - A matrix A is **similar** to a matrix D if: $A = PDP^{-1}$
  - P is an invertible matrix
  - A & D have the **same eigenvalues** and **determinant**
  - **IMPORTANT NOTE:**
    - If two matrices are similar (same characteristic polynomial), then they have the **same eigenvalues**
    - **CONVERSE IS NOT TRUE:**
      - If two matrices have the same eigenvalues, **that does not necessarily mean they are similar to each other**
- Diagonalization
  - Splitting up a matrix A into a **diagonal** matrix D and an invertible matrix P
  - Useful to compute $\mathbf{A^k}$ for **large k**
- Algebraic multiplicity
  - The number of **repeats** for an eigenvalue
  - $a_i = 2$: eigenvalue appears **twice**
- Geometric multiplicity
  - The number of **eigenvectors** for a given eigenvalue
  - **Dimension** of Nul $(A - \lambda I)$ for a **specific** λ
- Singular = Not Invertible
  - Free variables
  - Linearly dependent columns
- Nonsingular = Invertible

Remarks

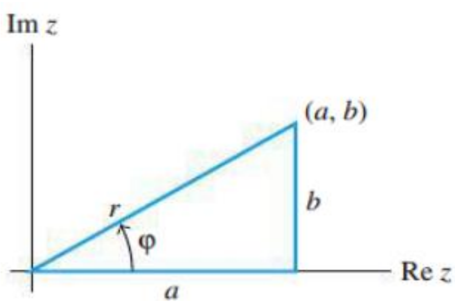
- **Diagonalization Formula**
  - $\mathbf{A = PDP^{-1}}$
    - P: the set of all **linearly independent eigenvectors**
    - D: the corresponding **eigenvalues** (in order)
  - $$A = (\vec{v}_1\ \vec{v}_2\ \dots\ \vec{v}_n)\left(\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \dots & \\ & & & \lambda_n \end{bmatrix}\right)(\vec{v}_1\ \vec{v}_2\ \dots\ \vec{v}_n)^{-1}$$

- Allows us to solve A^k for large k
    - $A^2 = PD(P^{-1}P)DP^{-1} \Rightarrow PD^2P^{-1}$
    - $\mathbf{A^k = PD^kP^{-1}}$
- **The Diagonalization Theorem (Theorem 5)**
  - An n x n matrix A is diagonalizable if and only if A has n linearly independent eigenvectors
    - Dimension of A = Dimension of P
  - A is diagonalizable if and only if there are **enough eigenvectors** to form a **basis of R^n**
    - Eigenvector basis
- **Steps to Diagonalize a Matrix**
  - Step 1: find the eigenvalues
    - det(A - λI) = 0
  - Step 2: find linearly independent eigenvectors of A
    - (A - λI)v = 0
      - Solve the **null space**
      - **Parametric vector form**
    - If # of total eigenvectors ≠ # of columns in A, then A is not diagonalizable (Theorem 5)
  - Step 3: construct P from vectors in Step 2
    - P = $\{v_1\ v_2 \dots v_n\}$
  - Step 4: construct D from corresponding eigenvalues
    - D = $\{\lambda_1\ \lambda_2 \dots \lambda_n\}$
- **Theorem 6**
  - An n x n matrix with n distinct eigenvalues is diagonalizable
  - **Note:**
    - It is not necessary for an n x n matrix to have n distinct eigenvalues in order to be diagonalizable
    - $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$: only 1 distinct eigenvalue but still has 2 eigenvectors
- **Theorem 7: Matrices whose Eigenvalues are Not Distinct**
  - Geometric multiplicity of λ must be **less than or equal to** the algebraic multiplicity of λ
    - $\mathbf{g}_i(\lambda) \leq \mathbf{a}_i(\lambda)$
  - A matrix is diagonalizable if and only if the **sum** of the dimensions of the eigenspaces equals n (the number of columns)
    - Total geometric multiplicity == number of columns in matrix A
    - Characteristic polynomial of A **factors completely** into linear factors
    - Geometric multiplicity for each eigenvalue = algebraic multiplicity for each eigenvalue

- **Diagonalizability and Invertibility have NO CORRELATION with each other**
    - **NEVER** associate the word **linearly independent, column space, null space, free variables, etc.** with diagonalizable

Section 5.5: Complex Eigenvalues

Definitions

- Complex number: $a + bi$
    - Any number of the form: $a + bi$
    - $i = \sqrt{-1}$
- Complex eigenvalue: λ
    - An eigenvalue that is a complex number: $a + bi$
        - Note: if $b = 0$, then λ is a **real eigenvalue**
- Complex eigenvector: x
    - An eigenvector subsisting of a complex eigenvalue
- Complex number space: $\mathbb{C}^n$
    - The space of all complex numbers
- $\mathbb{C}^2$
    - A complex number space with **2 entries**
    - At least **one entry is a complex number**
- Conjugate of a complex number
    - The conjugate for $(a + bi)$ is $(a - bi)$
- Complex conjugate of a vector **x**
    - $\bar{x}$
- Re x
    - The **real** parts of a complex vector **x**
    - An entry **can** be 0
- Im x
    - The **imaginary** parts of a complex vector **x**
    - An entry **can** be 0
- Argument of $\lambda = a + bi$
    - The **angle** ϕ produced by a and b on their respective Re x and Im x axis
    - 

Remarks

- **Finding complex eigenvalues and complex eigenvectors**
  - Step 1: det(A - λ) = 0
    - Getting the eigenvalues: λ
    - If the **characteristic equation** produces **complex roots**, then those roots are the complex eigenvalues
  - Step 2: Solve (A - λI)x = 0 for x
    - Getting the eigenvectors: x
    - Will get something with the form:
      - $\begin{bmatrix} -.3+6i & -.6 \\ .75 & .3+.6i \end{bmatrix} \to \begin{bmatrix} .75 & .3+.6i \\ 0 & 0 \end{bmatrix}$
    - x:
      - $\begin{bmatrix} -2-4i \\ 5 \end{bmatrix}$
  - Step 3: Find the other eigenvector
    - Find the **conjugate** of the other eigenvector:
      - $\begin{bmatrix} -2+4i \\ 5 \end{bmatrix}$
- **Re x & Im x**
  - $\overline{x}$: vector whose entries are the **complex conjugates** of the entries in x

    $$\begin{bmatrix} 3-i \\ i \\ 2 \end{bmatrix} \implies \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + i \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

    $$Re\, x = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} \quad Im\, x = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

  - $$\overline{x} = \begin{bmatrix} 3+i \\ -i \\ 2 \end{bmatrix}$$
- **Properties of Complex Conjugate Matrices**
  - Where
    - r: scalar
    - x: vector
    - B: matrix

    $\overline{rx} = \bar{r}\,\overline{x}$

    $\overline{Bx} = \overline{B}\,\overline{x}$

    $\overline{BC} = \overline{B}\,\overline{C}$
  - $\overline{rB} = \bar{r}\,\overline{B}$
    - Basically, you can **find the conjugates first**, then multiply them together
- **Complex Eigenvalues and Complex Eigenvectors Come in Pairs**

$$v_1 = \begin{bmatrix} -2-4i \\ 5 \end{bmatrix} \quad v_2 = \begin{bmatrix} -2+4i \\ 5 \end{bmatrix}$$

- $v_1 = \overline{v_2}$

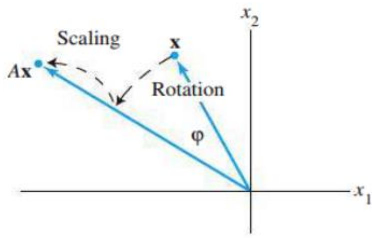
- **The Meaning of a Matrix that Acts on $\mathbb{C}^n$**
  - A transformation matrix that **rotates then scales**
  - 

    FIGURE 3 A rotation followed by a scaling.
- **Theorem 9**
  - For A = real 2 x 2 matrix with ($\lambda = a - bi$, where $b \neq 0$) and associated eigenvector **v** in $\mathbb{C}^2$:
    - **A = PCP^{-1}**
      - P = $[Re\ v \quad Im\ v]$
      - C = $\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{bmatrix}$
  - Why does this work?
    - A is **2 x 2** and has **two eigenvalues** (complex eigenvalues come in **pairs**)
    - C must be a **2 x 2** matrix as a result

Section 10.2: Google PageRank

Definitions

- Stochastic matrix
  - A matrix whose individual columns have an **entry sum of 1**
  - **Always** has **at least one** steady state
- Steady-state vectors
  - A probability vector ***q*** such that **P*q* = *q***
- Regular stochastic matrix
  - A stochastic matrix where for some power ***k***, **P^k** contains **entries all > 0**
  - **Always** has a **unique** steady state
- Dangling nodes
  - Any column that represents a web page that is a **dead end**
  - Usually is the form of an **elementary column:** $\{e_1, e_2, \ldots, e_n\}$

Remarks

- If **P** is a stochastic matrix, then a steady-state vector for **P** is a probability vector ***q*** such that

  $$\mathbf{P}\boldsymbol{q} = \boldsymbol{q}$$

- **Notes about stochastic matrices**
  - **Every** stochastic matrix **P** has a steady-state vector **q**
  - **1** must be an **eigenvalue** of any stochastic matrix
  - A steady-state vector is a probability vector which is also an **eigenvector** of **P** associated with the **eigenvalue 1**
  - Non-regular stochastic matrices can have **multiple steady state vectors**
- **Theorem 1**
  - If **P** is a **regular** $m \times m$ stochastic matrix with $m \geq 2$, then the following statements are true:
    a. There is a stochastic matrix Π such that $\lim_{n\to\infty} P^n = \Pi$
    b. Each column of Π is the **same probability vector *q***
       i. Would look something like this:

       $$\Pi = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

    c. For **any** initial probability vector x_0, $\lim_{n\to\infty} P^n x_0 = q$
    d. The vector ***q*** is the **unique** probability vector which is an **eigenvector** of **P** associated with the eigenvalue 1
    e. The eigenvalues of **P** satisfy $|\lambda| \leq 1$
- **PageRank**
  - **Adjustment 1:**
    - Replace all **dangling node** columns with $\begin{bmatrix} 1/n \\ 1/n \\ \cdots \\ 1/n \end{bmatrix}$ where n is the number of columns/rows
    - **P**$_*$ = **P** but with all dangling nodes replaced with the adjustment
  - **Adjustment 2:**
    - **K** = $\begin{bmatrix} 1/n \\ 1/n \\ \cdots \\ 1/n \end{bmatrix}$
  - **Google Matrix Formula:**

    $$\boldsymbol{G} = \mathbf{0.85}\boldsymbol{P}_* + \mathbf{0.15}\boldsymbol{K}$$

Section 6.1: Inner Product, Length, and Orthogonality

Definitions

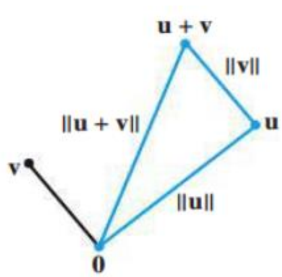
- Inner product (dot product)
    - If u and v are vectors in $\mathbb{R}^n$, then the **inner (dot) product** of u and v is:
        - $u^T v$ or:
        - $u \cdot v$
    - $$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$
- Vector length: $||v||$
    - $||v|| = \sqrt{v \cdot v} = \sqrt{v_1^{\ 2} + v_2^{\ 2} + \ldots + v_n^{\ 2}}$
- Unit vector
    - A vector whose length is 1
- Vector normalization
    - Dividing a nonzero vector by its length to make it a unit vector
- Distance between two vectors
    - $dist(u, v) = ||u - v||$
- Orthogonal vectors
    - Two vectors are orthogonal if their **dot product equals 0**
- Orthogonal complements
    - A **set** of vectors that are all **orthogonal to a subspace W**
    - Representation as a **line** or **plane** depends on the **null space of W**
- What does it mean for a subspace to be in $\mathbb{R}^n$?
    - Subspace (contains zero vector and is closed under addition and multiplication) has **n entries** for each vector in it **(dimension n)**
    - Note: R^1 means that the vectors have **one entry**
        - **Span** of just **[1]**

Remarks

- **Dot Product and Cross Product are Different**
    - $\boldsymbol{u} \cdot v = u^{\ T} v$
    - Dot product gives you a **number**
    - Cross product gives you a **vector**
- **Theorem 1: Dot Product Properties**
    - Where
        - u, v, and w are vectors in $\mathbb{R}^n$
        - c is a scalar in $\mathbb{R}$

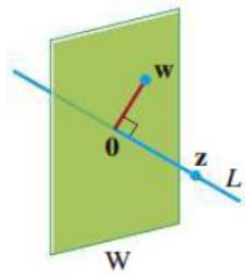
    a. $u \cdot v = v \cdot u$
        i. Symmetry

b. $(u+v)\cdot w = u\cdot w + v\cdot w$
   i. Linearity
c. $(cu)\cdot v = c(u\cdot v) = u\cdot(cv)$
   i. Scalars
   ii. Easy method: just find the dot product of the two vectors first, then multiply by the scalar
d. $u\cdot u \geq 0$
   i. Positivity
   ii. $u\cdot u = 0$ if and only if $u = 0$

- **Vector Length Properties**
  - Vector length is **always positive**
  - $||cv|| = |c|\,||v||$
  - $||cv||^2 = c^2||v||^2$
- **Normalizing a Vector**
  - $v(\frac{1}{||v||}) = u$
  - u: a unit vector
  - u is in the **same direction** as v, but u has **different magnitude** than v
- **Finding the Distance between Two Vectors**
  - Step 1: **subtract** the two vectors
    - $u - v$
  - Step 2: find the **length** of the resultant vector
    - $||u-v||$
- **Rudimentary Notes about Orthogonality**
  - Two vectors are orthogonal = two vectors are **perpendicular to each other**
  - $||u-v|| = ||u-(-v)||$
  - $u\cdot v = 0$
  - Zero vector is orthogonal to **every vector** in $\mathbb{R}^n$
- **Theorem 2: The Pythagorean Theorem**
  - Two vectors are orthogonal if and only if $||u+v||^2 = ||u||^2 + ||v||^2$
  - 

- **Rudimentary Notes about Orthogonal Complements**
  - What is an orthogonal complement?

- It is a **set of vectors** where each vector is orthogonal to a **subspace W**



- **Orthogonal Complement of W = W$^\perp$**
  - A vector **x** is in **W$^\perp$** if and only if **x** is orthogonal to every vector in a set that **spans W**
    - Must calculate **every single dot product pair** to prove orthogonality
  - **W$^\perp$** is a subspace of $\mathbb{R}^n$ ↔ **W** is also a subspace of $\mathbb{R}^n$
    - Both subspaces have **n entries**
    - **They do not necessarily have the same dimension**
      - **dim(Row W$^\perp$) = n - dim(Col W)**
      - Could be 2,2 or 1,3 where n = 4

- **Theorem 3**
  - Let A be an m x n matrix:
    - (Row A)$^\perp$ = Nul A
      - The row space of the orthogonal complement of A is the **null space** of A
    - (Col A)$^\perp$ = Nul A^T
      - The column space of the orthogonal complement of A is the **null space** of A **transpose**
  - **Proof**

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

$$\overline{x}=\begin{bmatrix}0\\0\\1\end{bmatrix}\, is\, orthogonal\, to\, matrix\, A$$

    -
    - **What is null space?**
      - **Av = 0**
      - Essentially taking the **dot product** of **every row of A** with the vector **v** and seeing that **v** is **orthogonal** to **A**

- **Rank Theorem**
  - *Row A*
    - The space spanned by the rows of matrix A
    - Given by the **pivot rows of A**

            - dim(*Row* A) = dim(*Col* A)
                - # of pivot columns = # of pivot rows
            - *Row* A^T = *Col* A
    - **N** = # of columns in a matrix
        - N = dim(*Col* A) + dim(*Nul* A)
        - N = dim(*Row* A) + dim(*Nul* A)

Section 6.2: Orthogonal Sets

Definitions

- Orthogonal set
    - A set of vectors $\{u_1, ..., u_p\}$ in $\mathbb{R}^n$ where each pair of distinct vectors from the set is orthogonal
    - $u_i \cdot u_j = 0, u \neq j$
- Orthogonal basis
    - A basis for a subspace **W** that is also an orthogonal set
- Orthogonal projection
    - Essentially projecting a vector onto a line/plane to get its **orthogonal complement**
    - $\hat{y} = proj_L\ y = (\frac{y \cdot u}{u \cdot u})u$ L: subspace spanned by u
- Orthonormal set
    - An orthogonal set where every vector is a unit vector
- Orthonormal basis
    - A basis for a subspace **W** that is also an orthonormal set
- Orthogonal matrix
    - A **square** matrix whose columns form an **orthonormal set**

Remarks

- **Theorem 4: Orthogonal Sets and Linear Independence**
    - If $S = \{u_1, \ldots, u_p\}$is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is **linearly independent** and is a basis for the subspace spanned by S
- All orthogonal sets are linearly independent sets
    - However, not all linearly independent sets are orthogonal
    - Remember to **omit the zero vector** for an orthogonal set
- **Theorem 5: Finding the Weights for a Linear Combination of an Orthogonal Basis**
    - Let$\{u_1, \ldots, u_p\}$be an orthogonal basis for a subspace W of $\mathbb{R}^n$:
      For every **y** in **W**, the weights in the linear combination

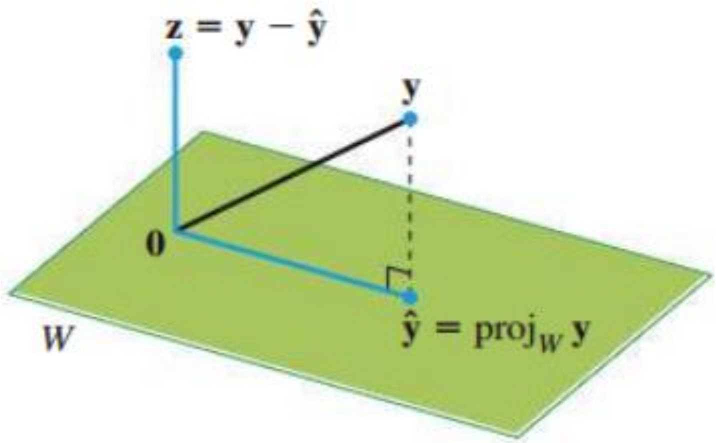
      $$\boldsymbol{y = c_1 u_1 + \ldots + c_p u_p}$$

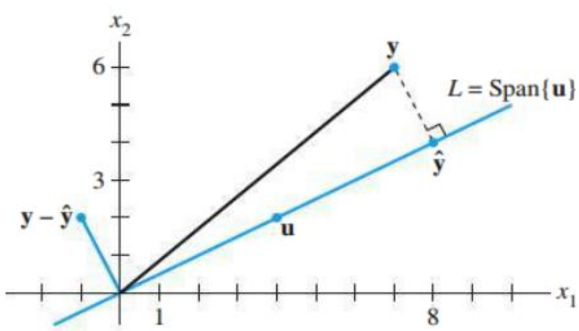
are given by:

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

- This method is better for finding the scalars than row reduction
  - **However,** this method is only applicable for **orthogonal bases**

- **How to find an Orthogonal Projection**
  - $$\hat{\mathbf{y}} = \text{proj}_L\, \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$
  - $$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$
    - **z**: the component of **y** orthogonal to **u**

- **Geometric Representations of an Orthogonal Projection**
  - 

  - 

- **Orthogonal Projections can be written as a Linear Combination of a Vector's Components**
  - $y = \left(\frac{y \cdot u_1}{u_1 \cdot u_1}\right) u_1 + \left(\frac{y \cdot u_2}{u_2 \cdot u_2}\right) u_2$

- All orthonormal sets are orthogonal
  - However, not all orthogonal sets are orthonormal

- **Theorem 6: Transpose of a Matrix with Orthonormal Columns**
  - An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$

- The **transpose** of a matrix with orthonormal columns multiplied by the original matrix always results in the **identity matrix**
    - Does it need to be square? **NO!**
- **Proof**
    - $$U^TU = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix}$$
    - Main diagonal: all **1's**
        - Remember, an orthonormal vector times itself is the square root of its length, which equals **1!!!**
    - Everywhere else: all **0's**
        - Remember, an orthonormal vector is **also orthogonal**, so two different vectors that are orthogonal to each other will have a product of **0**
- **A^TA** where A is a matrix with **orthogonal columns (DIFFERENT)**
    - Produces a **diagonal matrix** with all entries equal to **each vector's length squared**

- **Theorem 7: Properties of a Matrix with Orthonormal Columns**
    - Let U be an $m \times n$ matrix with orthonormal columns, and let **x** and **y** be in $\mathbb{R}^n$:
        - $||Ux|| = ||x||$
            - Linear mapping $x \to Ux$ **preserves length**
        - $(Ux) \cdot (Uy) = x \cdot y$
        - $(Ux) \cdot (Uy) = 0$ if and only if x and y are **orthogonal** to each other
            - Linear mapping $x \to Ux$ **preserves orthogonality**
- **Difference between Orthogonal Matrix and a Matrix with Orthonormal Columns**
    - Orthogonal matrix **must be square!!!**
- $U^{-1} = U^T$
    - The inverse of orthogonal matrices is its transpose
    - Orthogonal matrices have **linearly independent** columns
- **Determinant of an Orthogonal Matrix**
    - If A is an orthogonal matrix, then detA is equal to **1 or -1**
    - **Converse is NOT TRUE**
        - If the determinant of a square matrix = 1, then the matrix must be orthogonal. => **False**
        - $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

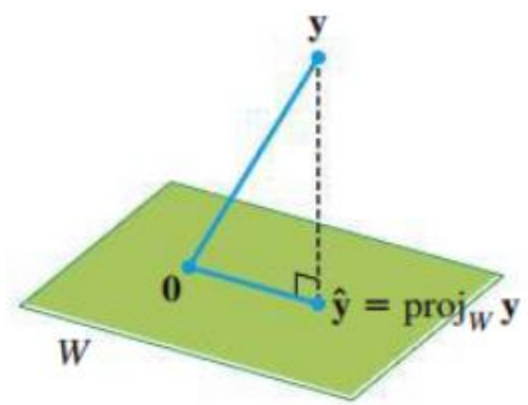
Section 6.3: Orthogonal Projections

Definitions

- **ŷ**: orthogonal **projection** of **y** onto **W**
  - **ŷ** = proj$_W$**y**
- **z**: orthogonal **component** of **y** onto **W**
  - **z = y - ŷ**
- Best approximation
  - $||y - \hat{y}|| < ||y - v||$
  - The **vertical distance** going straight up and down between a vector and its projection's space
    - Any distance between a vector and a subspace that is **not perpendicular to the space** is automatically **not the shortest distance**

Remarks

- Properties of an orthogonal projection onto $\mathbb{R}^n$



  -
  - Given a vector **y** and a subspace **W** in $\mathbb{R}^n$, there is a vector **ŷ** in **W** such that:
    - **ŷ** is the **unique** vector in **W** for which $y - \hat{y}$ is **orthogonal** to **W**
    - **ŷ** is the unique vector in **W closest to y**
  - Key to finding **least-squares solutions** (6.5)
- **Theorem 8: The Orthogonal Decomposition Theorem**
  - Let **W** be a subspace of $\mathbb{R}^n$. Then each **y** in $\mathbb{R}^n$ can be uniquely in the form

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

    where

    ŷ is in W

    z is in W$^\perp$
  - If **{u$_1$, ... , u$_p$}** is any **orthogonal basis** of **W**, then

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

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

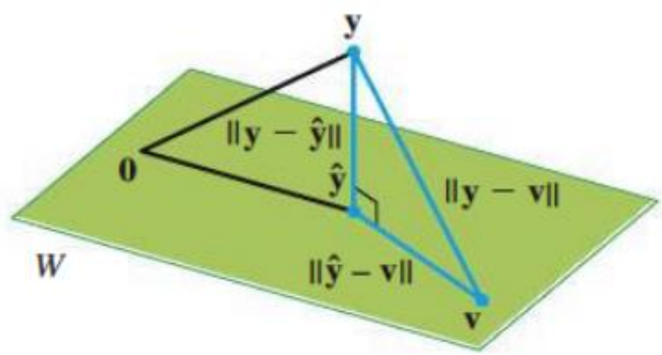
- We assume **W** is not the **zero subspace**
    - Otherwise, $\mathbf{W}^{\perp} = \mathbb{R}^n$
    - $\mathbf{y} = \mathbf{0} + \mathbf{y}$
        - Everything projected onto the zero subspace is just the **zero vector**
- **Properties of Orthogonal Projections**
    - If **y** is in $\mathbf{W} = \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then $\mathbf{proj}_W\mathbf{y} = \mathbf{y}$
    - If **y** is already in the subspace, then projecting it onto the same subspace **does not do anything**
- **Theorem 9: The Best Approximation Theorem**
    - Let **W** be a subspace of $\mathbb{R}^n$, let **y** be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of **y** onto **W**. Then $\hat{\mathbf{y}}$ is the **closest point** in **W** to **y**.
    - $||y - \hat{y}|| < ||y - v||$
      for all **v** in **W** distinct from $\hat{\mathbf{y}}$
    - 

      FIGURE 4 The orthogonal projection of **y** onto W is the closest point in W to **y**.
- **Theorem 10**
    - If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an **orthonormal basis** for a subspace **W** in $\mathbb{R}^n$, then
      $$proj_W y = (y \cdot u_1)u_1 + (y \cdot u_2)u_2 + \ldots + (y \cdot u_p)u_p$$
    - If $\mathbf{U} = [u_1 \ u_2 \ \ldots \ u_p]$, then
      $$proj_W y = UU^T y$$ for all **y** in $\mathbb{R}^n$
    - Remember, if u_1 is **a unit vector**, then $u_1 \cdot u_1 = 1$
- **Theorem 10 using Matrix with Orthonormal Columns vs. Orthogonal Matrix**
    - If **U** is an $n \times p$ matrix with orthonormal columns and **W** is the column space of **U**,
        - $\mathbf{U}^T\mathbf{U}\mathbf{x} = \mathbf{I}_p\mathbf{x} = \mathbf{x}$ for all **x** in $\mathbb{R}^p$
        - $\mathbf{U}\mathbf{U}^T\mathbf{y} = proj_W\mathbf{y}$ for all **y** in $\mathbb{R}^n$
    - If **U** is an $n \times n$ matrix with orthonormal columns, then **U** is an **orthogonal matrix**

- $\mathbf{UU^T y = Iy = y}$ for all **y** in $\mathbb{R}^n$
- See end of 6.2

Section 6.4: The Gram-Schmidt Process

Definitions

- Gram-Schmidt process
  - Algorithm for producing an **orthogonal/orthonormal** basis for any nonzero subspace of $\mathbb{R}^n$

Remarks

- **Theorem 11: The Gram-Schmidt Process**
  - Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

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

$$\operatorname{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \operatorname{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \tag{1}$$

  - **Remember**: a basis is a set of **linearly independent vectors** that span a subspace **W**
    - # of vectors in a basis = # of pivot columns/rows
    - Gram-Schmidt **requires a linearly independent basis (invertible/nonsingular)**
  - Any nonzero subspace **W** of $\mathbb{R}^n$ has an orthogonal basis because an ordinary basis $\{x_1, \dots, x_p\}$ is **always available**
- **Orthonormal Bases**
  - Simply **normalize** all vectors in an orthogonal basis $\{v_1, \dots, v_p\}$
- **Theorem 12: The QR Factorization**
  - If A is an m x n matrix with **linearly independent columns**, then A can be factored as **A = QR**
    - **Q**: an m x n matrix whose columns form an **orthonormal basis** for Col A

R: an $n \times n$ upper triangular matrix with positive entries on its diagonal

- **Process**
  1. Use **Gram-Schmidt** to find **Q**
  2. If needed, **normalize** the orthogonal basis given by **Q**
  3. Solve **A = QR** for **R**
     a. $R = Q^T A$
- If the columns of A were **linearly dependent**, then **R** would **not be invertible**

Section 6.5: Least-Squares Problems

Definitions

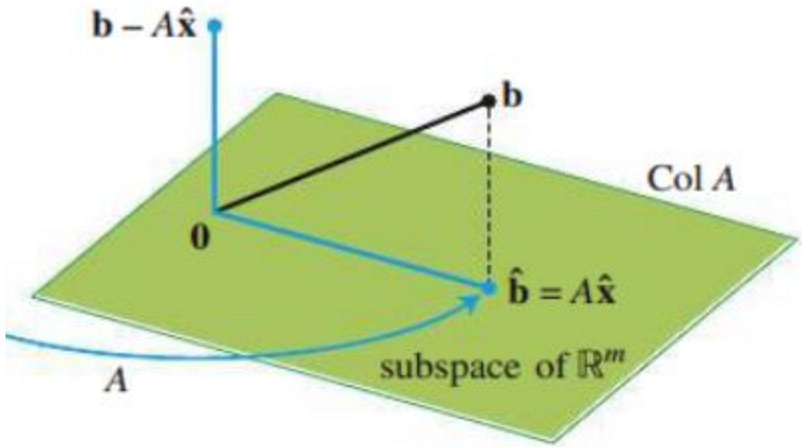
- General least-squares problem
  - Find **x** that makes $||b - Ax||$ **as small as possible**
- Normal equations
  - $A^T Ax = A^T b$
- Difference between **x** and $\hat{x}$
  - **x** just refers to **some general solution**
  - $\hat{x}$ is the solution that **solves the least-squares problem/normal equations**
- Least-squares error
  - Distance from **b** to $A\hat{x}$ where $\hat{x}$ is the least-squares solution to **b**
  - $||b - A\hat{x}||$

Remarks

- **What is the motivation for solving least-squares problems?**
  - Finding a **close enough** solution to **Ax = b** when it is an **inconsistent system**
- If A is $m \times n$ and **b** is in $\mathbb{R}^m$, a least-squares solution of A**x** = **b** is an $\hat{x}$ in $\mathbb{R}^n$ such that

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

  for all **x** in $\mathbb{R}^n$
  - If A is **already consistent**, then $||b - A\hat{x}|| = 0$
- **Solution of the General Least-Squares Problem**
  - Use the **Normal Equations!!!**
    - $A^T Ax = A^T b$
  - Derivation



- $$(Col\ A)^{\perp} = (Nul\ A)^T$$
  $$(b - A\hat{x}) \in (Nul\ A)^T$$
  $$\implies A^T(b - A\hat{x}) = 0$$
- $$\implies A^T A\hat{x} = A^T\vec{b}$$

- **Theorem 13**
  - The set of least-squares solutions of **Ax = b** coincides with the nonempty set of solutions of the normal equations $A^T Ax = A^T b$
  - Possible to have **more than one least-squares solution**
    - Existence of a **free variable** ⇔ columns of A are **linearly dependent**
- **Theorem 14**
  - Let A be an m x n matrix. The following statements are **logically equivalent**
    a. The equation **Ax = b** has a **unique** least-squares solution for each **b** in $\mathbb{R}^m$
    b. The columns of A are **linearly independent**
    c. The matrix $\mathbf{A^T A}$ is **invertible**
  - When these statements are true, the least-squares solution $\hat{x}$ is given by:
    $$\hat{x} = (A^T A)^{-1} A^T b$$
- **Calculating the Least-Squares Error**
  - $||b - A\hat{x}||$
- **Theorem 15: Finding the Least-Squares Solution using QR Factorization**
  - Given an m x n 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
    $$\hat{x} = R^{-1}Q^T b$$
    $$R\hat{x} = Q^T b$$
- What if **b** is **orthogonal** to the columns of **A**? What can we say about the least-squares solution of Ax = b?
  - If **b** is orthogonal to **A**, then the projection of **b** onto **A** is **0**

- A least-squares solution, $\hat{x}$, of **Ax = b** satisfies $A\hat{x} = 0$

Section 6.6: Applications to Linear Models

Definitions

- Least-Squares Lines
  - $y = \beta_0 + \beta_1 x$
- Residual
  - Difference between the **actual y-value** and the **predicted y-value**

Remarks

- **What is a Least-Squares Line?**
  - It is basically a **line of best-fit** for a set of data
  - Least-squares lines **minimize:**

    the **sum of the squares** of the residuals ⇔ the **least-squares solution**
- **Objective:**
  - Find β_0 *and* β_1(coefficients) that create the least-squares line
  - **Procedure using Normal Equations:**
    - $$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
    - Use the **normal equations** to solve
    - $$X^T X \boldsymbol{\beta} = X^T \mathbf{y}$$
  - **Procedure using Mean-Deviation Form:**
    - Find the **average** of all the **x-values:** $\underline{x}$
    - Calculate $x^* = x - \underline{x}$ for each **x**
    - $$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
      - Do this but with the new **x*** values
- **The General Linear Model**
  - $y = X\beta + \epsilon$
    - Solve the **normal equations:**
    - $X^T X\beta = X^T y$
  - **Example:**

ex: Second Order Polynomial

$$y = c_1 x_t + c_2 x^2$$

x	-1	0	0	1
y	2	1	0	6

model gives:

$$-c_1 + c_2 = 2$$
$$0c_1 + 0c_2 = 1$$
$$0c_1 + 0c_2 = 0$$
$$c_1 + c_2 = 6$$

$$\rightarrow \begin{pmatrix} -1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix} \vec{x} = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 6 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \hat{x} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

$$\rightarrow y = 2x + 4x^2$$

- **Multiple Regression**
  - Occurs when there are **2 or more independent variables**
  - **Example:**

ex.

[illegible]

x	-1	0	0	1
y	-1	-1	1	1
z	2	2	4	6

$$z = c_0 + c_1 x + c_2 y$$

$$c_0 - c_1 - c_2 = 2$$
$$c_0 + 0 - c_2 = 2$$
$$c_0 + 0 + cy = 4$$
$$c_0 + c_1 + c_2 = 6$$

$$\rightarrow \begin{pmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \vec{x} = \begin{pmatrix} 2 \\ 2 \\ 4 \\ 6 \end{pmatrix}$$