MATH 465 Notes

Contents

1 Eigenvalues

1.1 Eigenvalues

Let $A \in \mathbb{C}^{n \times n}$ and $\lambda \in \mathbb{C}$. Then λ is an eigenvalue of A if there exists a vector $u \in \mathbb{C}^n \backslash \{0^n\}$ such that $Au = \lambda u$

- The vector u is an eigenvector of A associated with λ
- Let λ be an eigenvalue of A. Then the following are equivalent
  - $\lambda I - A$ is not invertible
  - $\det(\lambda I - A) = 0$
- Real matrices may have complex eigenvalues and eigenvectors

1.2 Dominant Eigenvalues

Let $A \in \mathbb{C}^{n \times n}$ with eigenvalues $\lambda_1, ..., \lambda_n$. Then the eigenvalue λ_1 is the dominant eigenvalue of A if $|\lambda_1| > |\lambda_i|$ for all $i \in \{2, ..., n\}$

- The dominant eigenvalue is the eigenvalue with the uniquely largest magnitude
- A matrix may have a dominant eigenvalue even if it does not have a complete set of eigenvectors

1.3 Characteristic Polynomials

A matrix $A \in \mathbb{C}^{n \times n}$ has characteristic polynomial $p_A(\lambda) = \det(\lambda I - A)$

- $p_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + ... + a_1\lambda + a_0$
- The eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ of A are the values of λ where $p_A(\lambda) = 0$

1.4 Complete Set of Eigenvectors

A matrix $A \in \mathbb{C}^{n \times n}$ has a complete set of eigenvectors if the eigenvectors of A form a basis of $\mathbb{C}^n$

- An $n \times n$ matrix A has a complete set of eigenvectors if there are n linearly independent eigenvectors

1.5 Similar Matrices

Matrices $A \in \mathbb{C}^{n \times n}$ and $B \in \mathbb{C}^{n \times n}$ are similar if there exists an invertible matrix S such that $B = S^{-1}AS$

- $B = S^{-1}AS$ if and only if $A = SBS^{-1}$
- B represents the same linear transformation as A with respect to the basis consisting of the columns of S
- Similar matrices have the same characteristic polynomial and the same eigenvalues
- The eigenvectors of similar matrices transform using the transition matrix S

1.6 Diagonalizable Matrices

A matrix $A \in \mathbb{C}^{n \times n}$ is diagonalizable if there exists an invertible matrix S and a diagonal matrix Λ such that $A = S\Lambda S^{-1}$

- A matrix A is diagonalizable if and only if A has a complete set of eigenvectors
- A matrix A is diagonalizable if and only if for every eigenvalue λ, its algebraic multiplicity is equal to its geometric multiplicity

1.7 Null Space

Let $A \in \mathbb{C}^{n \times n}$. Then the null space $N(A)$ is the set of vectors x such that $Ax = 0$

- The null space is the subspace of the vector space consisting of vectors which are mapped onto zero

1.8 Algebraic Multiplicity

The algebraic multiplicity of an eigenvalue of A is the number of times it appears as a zero of the characteristic polynomial $p_A(\lambda)$

- The algebraic multiplicity of an eigenvalue of $A \in \mathbb{C}^{n \times n}$ is at most n

1.9 Geometric Multiplicity

The geometric multiplicity of an eigenvalue λ of A is given by $\dim(N(\lambda I - A))$

- The geometric multiplicity of an eigenvalue λ is the number of linearly independent eigenvectors associated with λ
- The geometric multiplicity of an eigenvalue λ is at most its algebraic multiplicity

1.10 Spectral Mapping Theorem

Let $A \in \mathbb{C}^{n \times n}$ with eigenvalues $\lambda_1, ..., \lambda_n$ and eigenvectors $u_1, ..., u_n$ and let $\alpha \in \mathbb{C}$. Then $A - \alpha I$ has eigenvalues $\lambda_1 - \alpha, ..., \lambda_n - \alpha$ and eigenvectors $u_1, ..., u_n$

- If $\alpha \neq \lambda_i$ for all $i \in \{1, ..., n\}$, then $(A - \alpha I)^{-1}$ has eigenvalues $\frac{1}{\lambda_1 - \alpha}, ..., \frac{1}{\lambda_n - \alpha}$ and eigenvectors $u_1, ..., u_n$

2 Power Methods

2.1 Rayleigh Quotient

Let $A \in \mathbb{C}^{n \times n}$ and let $x \neq 0$ be an approximate eigenvector. Then the best guess for the eigenvalue is given by the Rayleigh quotient

$$\beta = \frac{x^T A x}{x^T x}$$

2.2 Basic Power Method

Given $A \in \mathbb{C}^{n \times n}$, let x_0 be a non-zero vector. Then the basic power method is defined by the iterative equation

$$\begin{aligned} x_{n+1} &= A x_n \\ \beta_{n+1} &= \frac{{v_n}^T x_{n+1}}{{v_n}^T x_n} \end{aligned}$$

where v_n is defined in one of the following ways

- v_n is x_n itself
- v_n is $e_r = (0_1 \ldots 1_r \ldots 0_n)$, where r is the index of the largest element in magnitude of x_n
- v_n is some fixed vector v

Properties of the basic power method

- If A has an dominant eigenvalue, then (x_n) converges to an eigenvector associated with the dominant eigenvalue
- If x_0 and v_n are not orthogonal to the eigenvector associated with the dominant eigenvalue, then (β_n) converges to the dominant eigenvalue
- Let λ_1 be the dominant eigenvalue and λ_2 be the second largest eigenvalue. Then the asymptotic error constant of the basic power method is given by $\left|\frac{\lambda_2}{\lambda_1}\right|$

2.3 Scaled Power Method

Given $A \in \mathbb{C}^{n \times n}$, let x_0 be a vector such that ${x_0}^T x_0 = 1$. Then the scaled power method is defined by the iterative equation

$$\begin{aligned} y_{n+1} &= A x_n \\ \beta_{n+1} &= {x_n}^T y_{n+1} && \leftarrow \text{let } v_n \text{ be } x_n \text{ itself} \\ x_{n+1} &= \frac{y_{n+1}}{\sqrt{{y_{n+1}}^T y_{n+1}}} && \leftarrow \text{scale } y_{n+1} \end{aligned}$$

Properties of the scaled power method

- (β_n) converges to the dominant eigenvalue under the same conditions as the basic power method
- The scaled power method scales x_n to prevent it from causing overflow or underflow errors at high iterations
- ${x_n}^T x_n = 1$ for all $n \in \mathbb{N}$

2.4 Shifted Power Method

Given $A \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1 > ... > \lambda_n$ and eigenvectors $u_1, ..., u_n$, let $\alpha \in \mathbb{C}$ such that $|\lambda_1 - \alpha| < |\lambda_n - \alpha|$. Then the power method on $A - \alpha I$ converges to the eigenvector associated with the smallest eigenvalue λ_n

- The shifted power method parabolically improves the asymptotic error constant

2.5 Inverse Power Method

Given $A \in \mathbb{C}^{n\times n}$, let x_0 be a vector such that ${x_0}^T x_0 = 1$, and let $\alpha \in \mathbb{C}$. Then the inverse power method is defined by the iterative equation

$$
\begin{aligned}
&(A - \alpha I)y_{n+1} = x_n && \leftarrow \text{solve for } y_{n+1} \\
&\beta_{n+1} = {x_n}^T y_{n+1} && \leftarrow \text{let } v_n \text{ be } x_n \text{ itself} \\
&\Theta_{n+1} = \frac{1}{\beta_{n+1}} + \alpha && \leftarrow \text{reverse the inverse and the shift} \\
&x_{n+1} = \frac{y_{n+1}}{\sqrt{{y_{n+1}}^T y_{n+1}}} && \leftarrow \text{scale } y_{n+1}
\end{aligned}
$$

Properties of the inverse power method

- The inverse power method is the scaled power method for $(A - \alpha I)^{-1}$
- (Θ_n) converges to the dominant eigenvalue of $(A - \alpha I)^{-1}$ under similar conditions to the basic power method
- The dominant eigenvalue of $(A - \alpha I)^{-1}$ is $\frac{1}{\lambda_i - \alpha}$ where λ_i is the closest eigenvalue of A to α
- Let λ_i be the closest eigenvalue of A to α and λ_j be the second closest eigenvalue of A to α. Then the asymptotic error constant of the inverse power method is given by $\left|\frac{\lambda_i - \alpha}{\lambda_j - \alpha}\right|$

2.6 Rayleigh Quotient Iteration

Given $A \in \mathbb{C}^{n\times n}$, let x_0 be a vector such that ${x_0}^T x_0 = 1$. Then Rayleigh quotient iteration is defined by the iterative equation

$$
\begin{aligned}
&\alpha_n = {x_n}^T A x_n && \leftarrow \text{select } \alpha_n \text{ as close as possible to the dominant eigenvalue} \\
&(A - \alpha_n I)y_{n+1} = x_n && \leftarrow \text{solve for } y_{n+1} \\
&\beta_{n+1} = {x_n}^T y_{n+1} && \leftarrow \text{let } v_n \text{ be } x_n \text{ itself} \\
&\Theta_{n+1} = \frac{1}{\beta_{n+1}} + \alpha_n && \leftarrow \text{reverse the inverse and the shift} \\
&x_{n+1} = \frac{y_{n+1}}{\sqrt{{y_{n+1}}^T y_{n+1}}} && \leftarrow \text{scale } y_{n+1}
\end{aligned}
$$

Properties of Rayleigh quotient iteration

- Rayleigh quotient iteration is the inverse power method where α is set to the best guess for the eigenvalue

2.7 Deflating Matrices

Given a normalized eigenvector u_1 of $A \in \mathbb{C}^{n\times n}$ associated with the eigenvalue λ_1, let $S \in \mathbb{C}^{n\times n}$ be any invertible matrix where the first column is u_1. Then $B = S^{-1}AS$ is similar to A with eigenvalues $\lambda_B = \lambda_A$ and eigenvectors $u_B = S^{-1}u_A$. This means that

$$B = \begin{bmatrix} \lambda_1 & b_{12} & \dots & b_{1n} \\ 0 & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & c_{n2} & \dots & c_{nn} \end{bmatrix} \qquad C = \begin{bmatrix} c_{22} & \dots & c_{2n} \\ \vdots & \ddots & \vdots \\ c_{n2} & \dots & c_{nn} \end{bmatrix}$$

Then C is an $(n-1)\times(n-1)$ matrix with eigenvalues $\lambda_C = \lambda_A$ and eigenvectors $u_C = S^{-1}u_A$

- Deflating matrices after running the power method enables us to determine all the eigenvalues of a matrix

2.8 Symmetric Matrices

If $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix, then A has real eigenvalues $\lambda_1, ..., \lambda_n$ and has a complete set of eigenvectors $u_1, ..., u_n$

- The eigenvectors of A can be chosen to be orthonormal such that

$$u_i{}^T u_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

- The asymptotic error constant of the basic power method for symmetric matrices is given by $\left|\frac{\lambda_2}{\lambda_1}\right|^2$

2.9 Orthogonal Matrices

A real $(n\times n)$ matrix $U = [u_1 \; ... \; u_n]$ is orthogonal if its column vectors $u_1, ..., u_n$ are orthonormal

- U^TU has ij-element $u_i{}^T u_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$

If U is a real $(n\times n)$ matrix, then the following are equivalent

- U is orthogonal
- The columns of U are orthonormal
- The rows of U are orthonormal
- $U^TU = I$
- $UU^T = I$
- $||Ux||_2{}^2 = ||x||_2{}^2$ for all $x \in \mathbb{R}^n$

2.10 Compatible Matrix Norms

A matrix norm $||\cdot||_m$ is compatible with a vector norm $||\cdot||_v$ if $||Ax||_v \leq ||A||_m \cdot ||x||_v$ for every $(n\times n)$ matrix A and n-vector x

- If $\lambda \in \mathbb{C}$ is an eigenvalue of $A \in \mathbb{R}^{n\times n}$, then $|\lambda| \leq ||A||_m$
  - $\lambda \leq ||A||_1$, $\lambda \leq ||A||_F$, and $\lambda \leq ||A||_\infty$

2.11 Gershgorin Circle Theorem

Let $A \in \mathbb{C}^{n\times n}$ and define the absolute off-diagonal row and column sums as

$$r_k = \sum_{\substack{j=1 \\ j\neq k}}^{n} |a_{kj}| \qquad c_k = \sum_{\substack{i=1 \\ i\neq k}}^{n} |a_{ik}|$$

For $k = 1, 2, ..., n$, let $R_k = \{z : |z - a_{kk}| \leq r_k\}$ and $C_k = \{z : |z - a_{kk}| \leq c_k\}$ be circular disks in $\mathbb{C}$ centered at a_{kk} with radius r_k and c_k. If λ is an eigenvalue of A, then $\lambda \in \mathbb{R}_k$ for some k and $\lambda \in C_m$ for some m

- If m Gershgorin row-disks are disjoint from the other row-disks, then exactly m eigenvalues of A lie in the union of these m disks

3 Householder Transformations

3.1 Householder Transformations

For any non-zero vector $u \in \mathbb{R}^n$, define $Q_u \in \mathbb{R}^{n \times n}$ as

$$Q_u = I - \frac{2}{||u||_2{}^2} uu^T$$

Then Q_u is a Householder transformation

- Q_u is a symmetric and orthogonal matrix
- Q_u is a reflection transformation about the hyperplane consisting of vectors orthogonal to u
  - $Q_u u = -u$
  - $Q_u x = -\alpha u + y$ for some $\alpha \in \mathbb{R}$ and $y \in \mathbb{R}^n$ satisfying $u^T y = 0$
- Let $u \in \mathbb{R}^n$ be a non-zero vector where $u_1 = u_2 = ... = u_{k-1} = 0$.
  - For any $x \in \mathbb{R}^n$, the first $k-1$ elements of $Q_u x$ are the same as the first $k-1$ elements of x
  - If $y \in \mathbb{R}^n$ satisfies $y_k = y_{k+1} = ... = y_n = 0$, then $Q_u y = y$

3.2 Householder Transformations Between Two Vectors

Given any two non-zero vectors $v \neq w$ with $||v||_2 = ||w||_2$, we want to find Q_u such that $Q_u v = w$. Let $u = v - w$, then

$$Q_{v-w} v = v - \frac{2(v-w)^T v}{(v-w)^T (v-w)} (v - w) = w$$

- To avoid cancellation error, we can either map $v \mapsto w$ such that $u = v - w$ or map $v \mapsto -w$ such that $u = v + w$

3.3 Householder Transformation Between Vector Multiples

Given vectors $v \neq 0$ and $w = ||v||_2 \cdot e_1$ where e_1 is the standard basis vector, we want to find Q_u such that $Q_u v = ||v||_2 \cdot e_1$. Let $u = v - ||v||_2 \cdot e_1$, then

$$Q_{v-||v||_2 \cdot e_1} v = v - \frac{2(v - ||v||_2 \cdot e_1)^T v}{(v - ||v||_2 \cdot e_1)^T (v - ||v||_2 \cdot e_1)} (v - ||v||_2 \cdot e_1) = ||v||_2 \cdot e_1$$

- To avoid cancellation error, we can either map $v \mapsto ||v||_2 \cdot e_1$ such that $u = v - ||v||_2 \cdot e_1$ or map $v \mapsto -||v||_2 \cdot e_1$ such that $u = v + ||v||_2 \cdot e_1$

3.4 Householder Transformation Preserving First Column Vector

Given a vector v with $||v||_2 = 1$, we want to find Q_u such that the first column of Q_u is v. Let $u = v - e_1$. Then

$$Q_{v-e_1} v = v - \frac{2(v - e_1)^T v}{(v - e_1)^T (v - e_1)} (v - e_1) = e_1$$

- $Q_{v-e_1} v = e_1$ is equivalent to $Q_{v-e_1} e_1 = v$
- To avoid cancellation error, if $v_1 > 0$, then we can map $v \mapsto -e_1$ such that $u = v + e_1$; if $v_1 < 0$, then we can map $v \mapsto e_1$ such that $u = v - e_1$

3.5 Householder Transformation Creating Zeros

Given a vector $v \neq 0$ and an index $k \leq n$, we want to find Q_u such that $w = Q_u v$ has $w_i = v_i$ for $1 \leq i \leq k-1$ and $w_i = 0$ for $k+1 \leq i \leq n$. Let $u = v - w$ where

$$w = \begin{bmatrix} v_1 \\ \vdots \\ v_{k-1} \\ \sqrt{{v_k}^2 + {v_{k+1}}^2 + ... + {v_n}^2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{stays the same} \\ \leftarrow k^{\text{th}} \text{ entry} \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{becomes zero} \end{matrix}$$

Then $Q_{v-w} v = v - \dfrac{2(v-w)^T v}{(v-w)^T (v-w)}(v-w) = w$

- To avoid cancellation error, if $v_k > 0$, then choose $w_k = -\sqrt{{v_k}^2 + {v_{k+1}}^2 + ... + {v_n}^2}$; if $v_k < 0$, then choose $w_k = \sqrt{{v_k}^2 + {v_{k+1}}^2 + ... + {v_n}^2}$

3.6 QR Factorization

QR factorization generates a sequences of factored matrices

$$A^{(k)} = Q^{(k)} ... Q^{(1)} A^{(0)} \text{ for } 1 \leq k \leq n-1$$

where $A^{(0)} = A$

1. Assume by induction that $A^{(k-1)} = Q^{(k-1)} ... Q^{(1)} A^{(0)}$ has zeros below the diagonal in columns $1, 2, ..., k-1$
2. Let v be the k^{th} column of $A^{(k-1)}$ and find the Householder transformation $Q^{(k)} = Q_u$ that creates zeros along v below the row index k
3. Then ${a_{ik}}^{(k)} = 0$ for $i = k+1, k+2, ..., n$

Since $Q^{(k)}$ are orthogonal matrices for all $1 \leq k \leq n-1$, $Q^T = Q^{(n-1)} ... Q^{(1)}$ is an orthogonal matrix. Let $R = A^{(n-1)}$ such that R is an upper triangular matrix with non-zero diagonal elements. Then

$$A^{(0)} = QR$$

Since $A = A^{(0)}$, we can write $A = QR$

3.7 Solution By QR Factorization

Let $y = Rx$ be the solution of $Qy = b$. Since $Q^T = Q^{(n-1)} ... Q^{(1)}$, then $y = Q^{(n-1)} ... Q^{(1)} b$. Use forward substitution to solve for y, and then use back substitution to solve for x

- Once Q and R is calculated for a matrix A, we can easily solve for x in $Ax = b$ for any b

3.8 Basic Unshifted QR Algorithm

Given $A \in \mathbb{R}^{n\times n}$, let $A^{(0)} = A$. Then the basic unshifted QR algorithm is defined by the iterative equation

$$A^{(n)} = Q^{(n)}R^{(n)}$$
$$A^{(n+1)} = R^{(n)}Q^{(n)}$$

4 Reduction to Hessenberg Form

4.1 Transformation Methods

Transformation methods seek to find a similarity transformation that transforms an $(n \times n)$ matrix A to a similar matrix B

4.2 Hessenberg Matrices

An $(n \times n)$ matrix H is a Hessenberg matrix if it has the form

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1,n-1} & h_{1n} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2,n-1} & h_{2n} \\ 0 & h_{32} & h_{33} & \dots & h_{3,n-1} & h_{3n} \\ 0 & 0 & h_{43} & \dots & h_{4,n-1} & h_{4n} \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & h_{n,n-1} & h_{nn} \end{bmatrix}$$

- H is a Hessenberg matrix if $h_{ij} = 0$ for all $i > j + 1$

4.3 Reduction to Hessenberg Form

Reduction to Hessenberg Form iteratively reduces a matrix into a Hessenberg matrix

$$A^{(k)} = Q^{(k)} ... Q^{(1)} A^{(0)} Q^{(1)} ... Q^{(k)} \text{ for } 1 \le k \le n-2$$

where $A^{(0)} = A$

1. Assume by induction that $A^{(k-1)} = Q^{(k-1)} ... Q^{(1)} A^{(0)} Q^{(1)} ... Q^{(k-1)}$
2. Let v be the k^{th} column of $A^{(k-1)}$ and find the Householder transformation $Q^{(k)} = Q_u$ that creates zeros along v below the row index $k+1$
3. Then ${a_{ik}}^{(k)} = 0$ for $i = k+2, k+3, ..., n$

4.4 Hessenberg Matrices and Real Symmetric Matrices

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and let S be an orthogonal matrix. Then $S^{-1}AS$ is symmetric and tridiagonal if $S^{-1}AS$ is a Hessenberg matrix

4.5 Krylov's Method

Let H be a Hessenberg matrix, let $w_0 = [1\ 0\ ...\ 0]^T$, and define

$$w_k = H w_{k-1} \qquad \text{for } k = 0, 1, ..., n$$

Then the coefficients $a_1, ..., a_n$ of the monic characteristic polynomial $p(\lambda) = \lambda^n + a_1 \lambda^{n-1} + ... + a_n$ are given by the system

$$a_n w_0 + a_{n-1} w_1 + ... + a_1 w_{n-1} = -w_n$$

5 Tridiagonal Matrices

5.1 Tridiagonal Matrices

A tridiagonal matrix is a real $(n \times n)$ symmetric matrix of the form

$$H = \begin{bmatrix} d_1 & b_1 & 0 & \dots & 0 \\ b_1 & d_2 & b_2 & \ddots & \vdots \\ 0 & b_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} \\ 0 & \dots & 0 & b_{n-1} & d_n \end{bmatrix}$$

5.2 Strum Sequences

Given a tridiagonal matrix H, its Strum sequence is defined as

$$\begin{aligned} &p_0(t) = 1 \\ &p_1(t) = d_1 - t \\ &p_2(t) = (d_2 - t)p_1(t) - {b_1}^2 p_0(t) \\ &\quad\vdots \\ &p_k(t) = (d_k - t)p_{k-1}(t) - {b_{k-1}}^2 p_{k-2}(t) \\ &\quad\vdots \\ &p_n(t) = (d_n - t)p_{n-1}(t) - {b_{n-1}}^2 p_{n-2}(t) \end{aligned}$$

$p_n(t)$ is $(-1)^n$ times the characteristic polynomial for H

5.3 Sign Patterns

Given a real number c, find $p_0 = p_0(c)$, $p_1 = p_1(c)$, ..., $p_n = p_n(c)$ and let $N(c)$ be the number of sign agreements in adjacent terms. Then there are $N(c)$ roots of $p_n(t) = 0$ in the interval $[c, \infty)$

- If $p_k(c) = 0$ for some k, then we take the sign of $p_k(c)$ to be that of $p_{k-1}(c)$
- i.e. if $\{p_0(c), p_1(c), p_2(c), p_3(c), p_4(c)\}$ has the sign pattern $\{+, -, +, -, -\}$, then $N(c) = 3$
- i.e. if $\{p_0(c), p_1(c), p_2(c), p_3(c), p_4(c)\}$ has the sign pattern $\{+, -, 0, +, +\}$, then $N(c) = 2$

6 Complex Matrices

6.1 Conjugate/Hermitian Transpose

Given a complex matrix $(A)_{ij} = a_{ij}$, its conjugate/Hermitian transpose is given by $(A^H)_{ij} = \bar{a}_{ji}$, where $\bar{a}_{ji}$ is the conjugate of a_{ji}

- If A is real, then $A^H = A^T$
- If $A^H = A$, then A is called Hermitian

6.2 Unitary Matrices

A complex $(n \times n)$ matrix $U = [u_1 \text{ ... } u_n]$ is unitary if its column vectors $u_1, ..., u_n$ are orthonormal

- $U^H U$ has ij-element ${u_i}^H u_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$

If U is a complex $(n \times n)$ matrix, then the following are equivalent

- U is unitary
- The columns of U are orthonormal
- The rows of U are orthonormal
- $U^H U = I$
- $U U^H = I$
- $||Ux||_2{}^2 = ||x||_2{}^2$ for all $x \in \mathbb{R}^n$

6.3 Unitarily Similar Matrices

Matrices $A \in \mathbb{C}^{n \times n}$ and $B \in \mathbb{C}^{n \times n}$ are unitarily similar if there exists a unitary matrix U such that $B = U^H A U$

- $B = U^H A U$ if and only if $A = U B U^H$

6.4 Normal Matrices

A matrix $A \in \mathbb{C}^{n \times n}$ is normal if A and A^H commute such that $A^H A = A A^H$

- If $A \in \mathbb{C}^{n \times n}$ is normal and $U \in \mathbb{C}^{n \times n}$ is unitary, then $U^H A U$ is normal
- If $T \in \mathbb{C}^{n \times n}$ is normal and upper triangular, then T is diagonal

6.5 Quasi-Upper Triangular Matrices

An $(n \times n)$ matrix W is quasi-upper triangular if it is block upper triangular with only (1×1) or (2×2) blocks

$$\text{i.e. } W = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ 0 & 0 & b_{33} & b_{34} & b_{35} & b_{36} \\ 0 & 0 & 0 & b_{44} & b_{45} & b_{46} \\ 0 & 0 & 0 & 0 & b_{55} & b_{56} \\ 0 & 0 & 0 & 0 & b_{65} & b_{66} \end{bmatrix}$$

6.6 Schur's Theorem for Real Matrices

Let $A \in \mathbb{R}^{n \times n}$ with real eigenvalues. Then there exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that $U^T AU = T$, where T is upper triangular

6.7 Schur's Theorem for Real Matrices With Complex Eigenvalues

Let $A \in \mathbb{R}^{n \times n}$ with complex eigenvalues. Then there exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that $U^T AU = W$, where $W \in \mathbb{R}^{n \times n}$ is quasi-upper triangular

6.8 Schur's Theorem for Complex Matrices

Let $A \in \mathbb{C}^{n \times n}$. Then there exists a unitary matrix $U \in \mathbb{C}^{n \times n}$ such that $U^H AU = T$, where T is upper triangular

6.9 Spectral Mapping Theorem for Hermitian Matrices

Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. Then A is unitarily similar to a real diagonal matrix Λ

- If $A \in \mathbb{R}^{n \times n}$ is symmetric, then A is orthogonally similar to a real diagonal matrix Λ
- Let $A \in \mathbb{C}^{n \times n}$ and let $U \in \mathbb{C}^{n \times n}$ be unitary. Then $U^H AU = \Lambda$ if and only if $A = U\Lambda U^H$ such that $A = A^H$ for some real diagonal matrix Λ
- Let $A \in \mathbb{R}^{n \times n}$ and let $U = \mathbb{R}^{n \times n}$ be orthogonal. Then $U^T AU = \Lambda$ if and only if $A = U\Lambda U^T$ such that $A = A^T$ for some real diagonal matrix Λ

6.10 Spectral Mapping Theorem for Normal Matrices

Let $A \in \mathbb{C}^{n \times n}$ be a normal matrix. Then A is unitarily similar to a complex diagonal matrix Λ

- $A \in \mathbb{C}^{n \times n}$ is Hermitian if and only if $\Lambda = U^H AU$ is also Hermitian
- $A \in \mathbb{C}^{n \times n}$ is skew Hermitian such that $A^H = -A$ if and only if $\Lambda = U^H AU$ is also Hermitian
  - This means that $\lambda_i = -\bar{\lambda}_i$ such that each eigenvalue is pure imaginary
- $A \in \mathbb{C}^{n \times n}$ is unitary if and only if $\Lambda = U^H AU$ is also unitary
  - This means that $\bar{\lambda}_j \lambda_j = |\lambda_j|^2 = 1$ such that each eigenvalue has magnitude 1

6.11 Spectral Radius

The spectral radius of $A \in \mathbb{C}^{n\times n}$ is given by $\rho(A) = \max_{1\le i\le n} |\lambda_i|$ where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A

- $\lim_{n\to\infty} A^n = 0$ if and only if $\rho(A) < 1$

6.12 Complex Householder Transformations

For any non-zero vector $u \in \mathbb{C}^n$, define $Q_u \in \mathbb{C}^{n\times n}$ as

$$Q_u = I - \frac{2}{u^H u} u u^H$$

Then Q_u is a complex Householder transformation

- Q_u is symmetric, Hermitian, and unitary such that $Q^{-1} = Q^H = Q$

6.13 Deflating Real Matrices With Complex Eigenvalues

Given an eigenvector $z = v + iw$ of $A \in \mathbb{R}^{n\times n}$ associated with the eigenvalue $\lambda_1 = \alpha + i\beta$, let $S \in \mathbb{C}^{n\times n}$ be any invertible matrix where the first two columns are v, w. Then $B = S^{-1}AS$ is similar to A with eigenvalues $\lambda_B = \lambda_A$ and eigenvectors $u_B = S^{-1}u_A$. This means that

$$B = \begin{bmatrix} \alpha & \beta & b_{13} & \dots & b_{1n} \\ -\beta & \alpha & b_{23} & \dots & b_{2n} \\ 0 & 0 & b_{33} & \dots & b_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & b_{n3} & \dots & b_{nn} \end{bmatrix} \qquad C = \begin{bmatrix} b_{33} & \dots & b_{3n} \\ \vdots & \ddots & \vdots \\ b_{n3} & \dots & b_{nn} \end{bmatrix}$$

Then C is an $(n-2)\times(n-2)$ matrix with eigenvalues $\lambda_C = \lambda_A$ and eigenvectors $u_C = S^{-1}u_A$

7 Companion Matrices

7.1 Companion Matrices

Given a monic polynomial $p(x) = x^n + a_1 x^{n-1} + ... + a_{n-1}x + a_n$ of degree n with possibly complex coefficients $a_1, ..., a_n$, the companion matrix A of p has the form

$$A = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ -a_n & -a_{n-1} & \dots & -a_1 \end{bmatrix}$$

- The characteristic polynomial of the companion matrix A of the polynomial p is the polynomial itself
- The companion matrix A of the polynomial p is diagonalizable if and only if p has n distinct zeros

8 Inner Product Spaces and Least Squares

8.1 Inner Product Spaces

An inner product space is a real vector space V that can undergo inner product operation $V \times V$

- Inner product spaces satisfy the following properties
  1. Positive-definiteness: if $x \neq 0$, then $\langle x, x \rangle > 0$
  2. Linearity: $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$
  3. Symmetry: $\langle x, y \rangle = \langle y, x \rangle$

8.2 Inner Product Space Norm

The norm of a vector $x \in V$ is $||x|| = \sqrt{\langle x, x \rangle}$

- $||x|| \geq 0$ for all $x \in V$
- $||\alpha x|| = |\alpha| \cdot ||x||$
- $||x + y|| \leq ||x|| + ||y||$

8.3 Cauchy-Schwartz Inequality

The Cauchy-Schwartz inequality states that $| \langle x, y \rangle | \leq ||x||_2 \cdot ||y||_2$

- Equality holds if and only if x and y are linearly independent
- $\langle x, y \rangle = ||x||_2 \cdot ||y||_2 \cdot \cos\theta$

8.4 Orthogonal Vectors

Let x, y be vectors in V. Then x, y are orthogonal if $\langle x, y \rangle = 0$, denoted $x \perp y$

8.5 Orthogonal Subspaces

Let S be a subspace of V and let x be a vector in V. Then x is orthogonal to S if $x \perp y$ for all $y \in S$, denoted $x \perp S$

- $S^\perp = \{x \in V \mid x \perp S\}$
  - $S^\perp$ is a subspace of V
  - $S \cap S^\perp = \{0\}$

8.6 Orthogonal Systems

Let $\varphi_1, ..., \varphi_n \in V$. Then $\{\varphi_1, ..., \varphi_n\}$ is an orthogonal system in V if

$\langle \varphi_i, \varphi_j \rangle \neq 0$ for $i = j$

$\langle \varphi_i, \varphi_j \rangle = 0$ for $i \neq j$

- If $\langle \varphi_i, \varphi_j \rangle = 1$ for $i = j$, then $\{\varphi_1, ..., \varphi_n\}$ is an orthonormal system in V
- If $\{\varphi_1, ..., \varphi_n\}$ is an orthogonal system, then $\varphi_1, ..., \varphi_n$ is linearly independent
- If $\{\varphi_1, ..., \varphi_n\}$ is an orthogonal system, then $||\sum_{i=1}^n c_i \varphi_i||^2 = \sum_{i=1}^n |c_i|^2 \langle \varphi_i, \varphi_i \rangle$

8.7 Gram-Schmidt Process for Orthonormal Bases

Let S be an n-dimensional subspace of an inner product space V and let $\{\varphi_1, ..., \varphi_n\}$ be any basis of S. Then the Gram-Schmidt process for orthonormal bases is defined by the iterative equation

$$\begin{aligned}
\zeta_1 &= \varphi_1 \\
\psi_1 &= \tfrac{\zeta_1}{||\zeta_1||} \\
&\vdots \\
\zeta_i &= \varphi_i - \sum_{j=1}^{i-1} \langle \varphi_i, \psi_j \rangle \psi_j \\
\psi_i &= \tfrac{\zeta_i}{||\zeta_i||}
\end{aligned}$$

for $i = 1, ..., n$

- $\psi_1, ..., \psi_n$ is an orthonormal basis of S
- $\mathrm{span}(\varphi_1, ..., \varphi_i) = \mathrm{span}(\psi_1, ..., \psi_i)$ for each $i = 1, ..., n$

8.8 Gram-Schmidt Process for Orthogonal Bases

Let S be an n-dimensional subspace of an inner product space V and let $\{\varphi_1, ..., \varphi_n\}$ be any basis of S. Then the Gram-Schmidt process for orthogonal bases is defined by the iterative equation

$$\begin{aligned}
\eta_1 &= \varphi_1 \\
&\vdots \\
\eta_i &= \varphi_i - \sum_{j=1}^{i-1} \tfrac{\langle \varphi_i, \eta_j \rangle}{\langle \eta_j, \eta_j \rangle} \eta_j
\end{aligned}$$

for $i = 1, ..., n$

- $\eta_1, ..., \eta_n$ is an orthogonal basis of S

8.9 The Projection Theorem

Let V be an inner product space and let S be a finite dimensional subspace. Then given a vector $y \in V$, there exists unique $x^* \in S$ and $w^* \in S^\perp$ such that $y = x^* + w^*$

- This is denoted as $V = S \oplus S^\perp$
- x^* is the unique element of S which satisfies $\langle y - x^*, x \rangle = 0$ for all $x \in S$
- x^* is the unique element of S which minimizes $||y - x||^2$ over all $x \in S$

8.10 The Normal Equations

Let V be an inner product space, let S be a finite dimensional subspace, and let $\varphi_1, ..., \varphi_n \in S$ be a set of vectors which spans S. Then $x^* = c_1\varphi_1 + ... + c_n\varphi_n$ minimizes $||y - x||^2$ over all $x \in S$ if and only if the coefficients $c_1, ..., c_n$ satisfy

$$\begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}^T \begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}^T y$$

$$\begin{bmatrix} \langle\varphi_1, \varphi_1\rangle & \langle\varphi_1, \varphi_2\rangle & \cdots & \langle\varphi_1, \varphi_n\rangle \\ \langle\varphi_2, \varphi_1\rangle & \langle\varphi_2, \varphi_2\rangle & \cdots & \langle\varphi_2, \varphi_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\varphi_n, \varphi_1\rangle & \langle\varphi_n, \varphi_2\rangle & \cdots & \langle\varphi_n, \varphi_n\rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \langle\varphi_1, y\rangle \\ \langle\varphi_2, y\rangle \\ \vdots \\ \langle\varphi_n, y\rangle \end{bmatrix}$$

- If $\{\varphi_1, ..., \varphi_n\}$ is an orthogonal basis of S, then $x^* = \sum_{i=1}^{n} \frac{\langle y, \varphi_i\rangle}{\langle\varphi_i, \varphi_i\rangle}\varphi_i$
- If $\{\varphi_1, ..., \varphi_n\}$ is an orthonormal basis of S, then $x^* = \sum_{i=1}^{n} \langle y, \varphi_i\rangle \varphi_i$
- If V is a finite dimensional inner product space and $\psi_1, ..., \psi_n$ is an orthonormal basis of V, then $y = \sum_{i=1}^{n} \langle y, \psi_i\rangle \psi_i$

8.11 Bessel's Inequality

Let V be an inner product space and let $\{\psi_1, ..., \psi_n\}$ be an orthonormal set in V. Then

$$\sum_{i=1}^{n} |\langle y, \psi_i\rangle|^2 \leq ||y||^2 \qquad \text{for every } y \in V$$

8.12 Parseval's Equality

Let V be an inner product space and let $\{\psi_1, \psi_2, ...\}$ be an orthonormal set in V. Then the following are equivalent for all $y \in V$

- $\sum_{i=1}^{n} |\langle y, \psi_i\rangle|^2 = ||y||^2$
- The series $\sum_{i=1}^{\infty} \langle y, \psi_i\rangle \psi_i$ converges to y

8.13 Complete Orthonormal System

Let V be an inner product space and let $\{\psi_1, \psi_2, ...\}$ be an orthonormal set in V. Then $\{\psi_1, \psi_2, ...\}$ is a complete orthonormal system in V if it satisfies Parseval's inequality

8.14 Weighted Discrete Inner Product

Let $x_0 < x_1 < ... < x_m \in \mathbb{R}$ be fixed values, let $w_0, w_1, ..., w_m > 0$, and let GF be the vector space of all grid functions $f : (x_0, ..., x_m) \to \mathbb{R}$ where each $f \in \mathrm{GF}$ is represented by its graph vector $F = [f(x_0) \; ... \; f(x_m)] \in \mathbb{R}^{m+1}$. Then the weighted discrete inner product on GF is

$$\langle f, g \rangle_w = \sum_{i=0}^{m} w_i f(x_i) g(x_i)$$

- Given $f, g \in \mathrm{GF}$, $\langle f, g \rangle_w = F^T W G$ where $W = \begin{bmatrix} w_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_m \end{bmatrix}$
- $||f||_w = \sqrt{F^T W F}$
- A set of functions in GF is linearly independent if and only if their graph vectors are linearly independent in $\mathbb{R}^{m+1}$

8.15 Linear Least Squares in GF

Let $g_0, ..., g_n$ be linearly independent functions in GF, let $S = \mathrm{span}(g_0, ..., g_n)$, and let $y \in \mathrm{GF}$. Then $f^* = c_0 g_0 + ... + c_n g_n$ minimizes $||y - f||_w^2$ over all $f \in S$ if and only if the coefficients $c_0, ..., c_n$ satisfy

$$\begin{bmatrix} \sqrt{w_0} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{w_m} \end{bmatrix} \begin{bmatrix} g_0(x_0) & \dots & g_n(x_0) \\ \vdots & \ddots & \vdots \\ g_0(x_m) & \dots & g_n(x_m) \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \sqrt{w_0} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{w_m} \end{bmatrix} \begin{bmatrix} y(x_0) \\ \vdots \\ y(x_m) \end{bmatrix}$$

- Let $x_0 < x_1 < ... < x_m \in \mathbb{R}$. Then $g_0(x_i) = 1, g_1(x_i) = x_i, ..., g_n(x_i) = x_i{}^n$ are linearly independent in GF if and only if $n \leq m$

8.16 Weighted Integral Inner Product

Let $a < b \in \mathbb{R}$ be fixed values, let $w(x) > 0$ be a continuous function on (a, b) with $\int_a^b w(x)\, dx < \infty$, and let $V = C[a, b]$ be the space of all continuous functions $f : [a, b] \to \mathbb{R}$. Then the weighted integral inner product on V is

$$\langle f, g \rangle_w = \int_a^b w(x) f(x) g(x)\, dx$$

8.17 Linear Least Squares in V

Let $g_0, ..., g_n$ be linearly independent functions in V, let $S = \mathrm{span}(g_0, ..., g_n)$, and let $y \in V$. Then $f^* = c_0 g_0 + ... + c_n g_n$ minimizes $||y - f||_w^2$ over all $f \in S$ if and only if the coefficients $c_0, ..., c_n$ satisfy

$$\begin{bmatrix} \langle g_0, g_0 \rangle_w & \dots & \langle g_0, g_n \rangle_w \\ \vdots & \ddots & \vdots \\ \langle g_n, g_0 \rangle_w & \dots & \langle g_n, g_n \rangle_w \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \langle g_0, y \rangle_w \\ \vdots \\ \langle g_n, y \rangle_w \end{bmatrix}$$

8.18 Monic Polynomials Orthogonal to Weighted Inner Product

Let $\langle f, g\rangle_w$ be a weighted inner product. Then the monic polynomials that are orthogonal with respect to this inner product are given by

$$q_0(x) = 1$$

$$q_1(x) = x - \frac{\langle x, q_0\rangle_w}{\langle q_0, q_0\rangle_w} q_0(x)$$

$$\vdots$$

$$q_k(x) = x^k - \sum_{j=0}^{k-1} \frac{\langle x^k, q_j\rangle_w}{\langle q_j, q_j\rangle_w} q_j(x)$$

$$\vdots$$

$$q_n(x) = x^n - \sum_{j=0}^{n-1} \frac{\langle x^n, q_j\rangle_w}{\langle q_j, q_j\rangle_w} q_j(x)$$

- $q_k(x)$ is a monic polynomial of degree k
- $\{q_0, ..., q_n\}$ is an orthogonal basis of $\mathscr{P}_n$
- Let $p_k(x) = \frac{q_k(x)}{||q_k||_w}$. Then $\{p_0, ..., p_n\}$ is an orthonormal basis of $\mathscr{P}_n$

8.19 Weierstrass Approximation Theorem

Let $q_0, q_1, q_2, ...$ be a sequence of monic orthogonal polynomials and $p_k = \frac{q_k}{||q_k||_w}$ be the corresponding orthonormal polynomials. Then given $f \in C[a, b]$ and $\varepsilon > 0$, there exists $N \in \mathbb{N}$ and $q_N \in \mathscr{P}_N$ such that

$$\max_{a \le x \le b} |f(x) - g_N(x)| = ||f - g_N||_\infty < \varepsilon$$

8.20 Chebyshev Polynomials

The Chebyshev polynomials are given by $T_k(x) = \cos(k \cos^{-1}(x))$ for $k = 0, 1, 2, ...$

- T_k is a polynomial of degree k
- $T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$

8.21 Linear Least Squares with Orthogonal Polynomials

Given the inner product $\langle f, g\rangle_w$, let $r_0, ..., r_n$ be orthogonal polynomials with each r_k having exact degree k such that $\{r_0, ..., r_n\}$ is an orthogonal basis of $\mathscr{P}_n$. Then the closest polynomial $p_n^*(x)$ of degree n orthogonal to a polynomial $f(x)$ is given by

$$p_n^* = \sum_{i=0}^{n} \frac{\langle f, r_i\rangle}{\langle r_i, r_i\rangle} r_i(x)$$

- $\{p_0, p_1, ...\}$ is a complete orthonormal system
- $||f - p_n^*||_w^2 = ||f||_w^2 - ||p_n^*||_w^2$ since f and p_n^* are orthogonal