M520 Ch. 6 Notes: Orthogonality and Least Squares

6.1

The Inner Product

- u and v are vectors in R^n, then u and v are $n \times 1$ matrices
- Definition: $u^T v$ is the inner product (or dot product) of u and v and written as $u \cdot v$
- $u =< u_1, ..., u_n >$ and $v =< v_1, ..., v_n >$, then $u \cdot v$ is $u_1 v_1 + ... + u_n v_n$

Inner Product Properties

- Let u, v, and w be vectors in R^n, and let c be a scalar, then:
- a: $u \cdot v = v \cdot u$
- b: $(u + v) \cdot w = u \cdot w + v \cdot w$
- c: $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$
- d: $u \cdot u \geq 0$, and $u \cdot u = 0$ if and only if $u = 0$
- another useful property: $(c_1 u_1 + ... + c_p u_p) \cdot w = c_1(u_1 \cdot w) + ... + c_p(u_p \cdot w)$

The Length of a Vector

- Definition: The length (or norm) of a vector v is the nonnegative scalar $||v||$ defined by: $||v|| = \sqrt{v \cdot v} = \sqrt{v_1^2 + ... + v_n^2}$ and $||v||^2 = v \cdot v$

Distance in $\mathbb{R}^n$

- Definition: For u and v in $\mathbb{R}^n$, the distance between u and v, written as dist(u, v), is the length of the vector $u - v$
- dist(u, v)= $||u - v||$

Orthogonal Vectors

- Definition: Two vectors u and v in $\mathbb{R}^n$ are orthogonal to each other if $u \cdot v = 0$
- zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $0^T v = 0$ for all v
- Theorem 2: The Pythagorean Theorem: Two vectors u and v are orthogonal if and only if $||u+v||^2 = ||u||^2 + ||v||^2$

Orthogonal Complements

- if vector z is orthogonal to every vector in a subspace W of $\mathbb{R}^n$ then z is said to be orthogonal to W
- Definition: the set of all vectors z that are orthogonal to W is called the orthogonal complement of W and is denoted by $W^\perp$ (W perpendicular)
- fact 1 about $W^\perp$ where W is a subspace of $\mathbb{R}^n$: A vector x is in $W^\perp$ if and only if x is orthogonal to every vector in a set that spans W
- fact 2: $W^\perp$ is a subspace of $\mathbb{R}^n$
- Theorem 3: Let A be an mxn matrix.
  - orthogonal complement of the row space of A is the null space of A: $(RowA)^\perp = NulA$
  - the orthogonal complement of the column space of A is the null space of A^T: $(ColA)^\perp = NulA^T$

Angles in $\mathbb{R}^2$ and $\mathbb{R}^3$

- $u \cdot v = ||u||||v||cos\theta$

6.2

Orthogonal Sets

- Definition: A set of vectors $\{u_1, ..., u_p\}$ in $\mathbb{R}^n$ is said to be an orthogonal set if each pair of distinct vector from the set is orthogonal; that is if $u_i \cdot u_j = 0$ whenever $i \neq j$
- Theorem 4: If $S = \{u_1, ..., u_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S

Orthogonal Basis

- Definition: An orthogonal basis for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set
- Theorem 5: Let $\{u_1, ..., u_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each y in W, the weights in the linear combination $y = c_1 u_1 + ... + c_p u_p$ are given by $c_j = \frac{y \cdot u_j}{u_j \cdot u_j}$ for $j = 1, ..., p$

Orthogonal Projection

- want to decompose a vector y in $\mathbb{R}^n$ into the sum of 2 vectors, one is a multiple of u and the other orthogonal to u (where u is nonzero)
- $y = \hat{y} + z$
- $\hat{y}$ is the scalar multiple of u; $\hat{y}$ is the orthogonal projection of y onto u; $y = \alpha u$ where $\alpha = \frac{y \cdot u}{u \cdot u}$
- z is a vector orthogonal to u; z is the component of y orthogonal to u

Orthonormal Sets

- Definition: A set $\{u_1, ..., u_p\}$ is an orthonormal set if it is an orthogonal set of unit vectors
- Orthonormal basis: if W is a subspace spanned by an orthonormal set then the set is the orthonormal basis for W
- Theorem 6: An mxn matrix U has orthonormal columns if and only if $U^T U = I$
- Theorem 7: Let U be an mxn matrix with orthonormal columns, and let x and y be in $\mathbb{R}^n$, then:
  - a $||Ux|| = ||x||$
  - b $(Ux) \cdot (Uy) = x \cdot y$
  - c $(Ux) \cdot (Uy) = 0$ if and only if $x \cdot y = 0$

- Orthogonal matrix: a square invertible matrix U such that $U^{-1} = U^T$; matrix has orthonormal columns and rows

6.3
Theorem 8: The Orthogonal Decomposition Theorem

- Let W be a subspace of $\mathbb{R}^n$
- then each y in $\mathbb{R}^n$ can be written uniquely in the form: $y = \hat{y} + z$ where $\hat{y}$ is in W and z is in $W^{\perp}$
- If $\{u_1, ..., u_p\}$ is any orthogonal basis of W, then:

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u1 + ... + \frac{y \cdot u_p}{u_p \cdot u_p} up$$

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

Properties of Orthogonal Projections

- If y is in W = Span$\{u_1, ..., u_p\}$, then $proj_w y = y$
- 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 $\hat{y}$ be the orthogonal projection of y onto W
  - Then $\hat{y}$ is the closest point in W to y, in the sense that

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

    for all v in W distinct from $\hat{y}$
- Theorem 10:
  - If $\{u_1, ..., u_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

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

  - similar to Theorem 8 except that denominators are 1 because two unit vectors dotted with each other is 1
  - If U = $[u_1 ... u_p]$, then

    $$proj_W y = UU^T y$$

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

6.4
The Gram-Schmidt Process

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

$$v_1 = x_1$$
$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$
$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$
$$...$$
$$v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - ... - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}$$

- Then $\{v_1, ..., v_p\}$ is an orthogonal basis for W; in addition: Span$\{v_1, ..., v_k\}$=Span$\{x_1, ..., x_k\}$ for $1 \leq k \leq p$

QR Factorization of Matrices

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

6.5: Least Squares Problems

General Least Squares Problem

- Sometimes $Ax = b$ has an inconsistent solution, want to find an x that makes Ax as close as possible to b; minimize the distance between b and Ax; want to make $||b - Ax||$ as small as possible
- Definition: If A is mxn and b is in $\mathbb{R}^m$, a least-squares solution of Ax=b is an $\hat{x}$ in $\mathbb{R}^n$ such that

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

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

Solution of General Least-Squares Problem

- apply Theorem 9 Best Approximation Theorem to subspace Col A to get: $\hat{b} = proj_{ColA} b$
- since $\hat{b}$ is the closest point in Col A to b, a vector $\hat{x}$ is a least-squares solution of Ax=b if and only if $\hat{x}$ satisfies $A\hat{x} = \hat{b}$
- Theorem 13: The set of least-squares solutions of Ax=b coincides with the nonempty set of solutions of the normal equations $A^T Ax = A^T b$; every least-squares solution of Ax=b satisfies the equation $A^T Ax = A^T b$
- Theorem 14: Let A be an mxn 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 $A^T A$ is invertible
  - When these statements are true, the least-squares solution $\hat{x}$ is given by $\hat{x} = (A^T A)^{-1} A^T b$
- Least-squares error: when a least-squares solution $\hat{x}$ is used to produce $A\hat{x}$ as an approximation to b, the distance from b to $A\hat{x}$ is called the least-squares error of this approximation
- Theorem 15:
  - Given an mxn matrix A with linearly independent columns, let A=QR be a QR factorization of A as in Theorem 12.
  - Then for each b in $\mathbb{R}^m$, the equation Ax=b has a unique least-squares solution given by $\hat{x} = R^{-1} Q^T b$

6.7: Inner Product Spaces

Inner Product Space

- An inner product on a vector space V is a function that, to each pair of vectors u and v in V, associates a real number $< u, v >$ and satisfies the following axioms, for all u, v, and w in V and all scalars c:
- 1: $< u, v >=< v, u >$
- 2: $< u + v, w >=< u, w > + < v, w >$
- 3: $< cu, v >= c < u, v >=< u, cv >$
- 4: $< u, u >\geq 0$ and $< u, u >= 0$ if and only if $u = 0$
- A vector space with an inner product is called an inner product space

Lengths, Distances, and Orthogonality

- Let V be an inner product space, with the inner product denoted by $< u, v >$
- Length (norm): length of a vector v is the scalar $||v|| = \sqrt{< v, v >}$ just as in $\mathbb{R}^n$
- Distance between u and v is $||u - v||$
- Orthogonal: vectors u and v are orthogonal if $< u, v >= 0$

The Gram-Schmidt Process

- existence of orthogonal bases for finite-dimensional subspaces of an inner product space can be established by the Gram-Schmidt process, just as in $\mathbb{R}^n$
- orthogonal projection of a vector onto a subspace W with an orthogonal basis has the properties described in the Orthogonal decomposition Theorem and the Best Approximation Theorem

Best Approximation in Inner Product Spaces

- vector space V whose elements are functions
- want to approximate a function f in V by a function g from a specified subspace W of V
- best approximation to f by functions in W is the orthogonal projection of f onto the subspace W

The Cauchy-Schwarz Inequality

- For all vectors u and v in an inner product space V:

$$| < u, v > | \leq ||u|| * ||v||$$

Theorem 17: The Triangle Inequality

- For all vectors u and v in an inner product space V:

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

An Inner Product for C[a, b]

-

6.8: Applications of Linear Product Spaces

Fourier Series

-