Outline

- Orthogonal Vectors and Subspaces
- Projection
- Least Squares Solution
- Orthonormal Bases and Orthogonal Matrices
- Gram-Schmidt Orthogonalization
- QR Decomposition
- Review

Orthogonal Vectors and Subspaces

Length, Inner Product and Orthogonal Vectors

- *What is "perpendicular"?* Given 2 vectors a and b, how does one decide if they are perpendicular?
- generalization of properties in R^2 (or R^3) to R^n
  - length: $||x||^2 = x_1^2 + \cdots + x_n^2 = x^T x$
  - inner product of two vectors: $(x, y) = x^T y$
  - two vectors are perpendicular if $x^T y = 0$ (Pythagoras)
- (theorem) non-zero vectors are independent if they are orthogonal

Orthogonal Subspaces

- two subspaces U and V of S are orthogonal if $u^T v = 0$ for every $u \in U$ and $v \in V$
  - only need to check spanning sets of U and V
- Given a matrix A, the row space is orthogonal to the nullspace and the column space is orthogonal to the left nullspace. Let $x \in \mathcal{N}(A)$, then

$$v \in \mathcal{R}(A^T) \Rightarrow v = A^T z \Rightarrow v^T x = z^T A x = 0.$$

Orthogonal complements

- The *orthogonal complement* of a subspace V of a vector space S, denoted by $V^\perp$, is the largest subset of S that is orthogonal to V.
- $V^\perp$ is a subspace of S.
- Furthermore.

$$dim(V) + dim(V^\perp) = dim(S)$$

$$V = W^\perp \Rightarrow W = V^\perp$$

Fundamental Theorem of Linear Algebra

- *Fundamental theorem of linear algebra: Part II*
  Given a $m \times n$ matrix,
  - the row space is the orthogonal complement of the nullspace in R^n.
  - the column space is the orthogonal complement of the left nullspace

Decomposition of Vectors

- given orthogonal complements V and W of a space S, every vector $x \in S$ can be written as $x = v + w$, where $v \in V$ and $w \in W$
- v(resp. w) is called the projection of x onto V(resp. W)

What Is Matrix Multiplication?

- A vector $x \in R^n$ can be decomposed into $x = x_r + x_n$, where x_r is in the row space and x_n is in the nullspace of an $m \times n$ matrix A. Then

$$Ax = A(x_r + x_n) = Ax_r$$

- the mapping from the row space to the column space is invertible. That is, every b in the column space comes from only one x_r in the row space.
- *A matrix transforms its row space to its column space*

Projections

Projection onto a Line

- Given b, the vector to be projected, and a, the direction of the line to be projected onto
- Let $p = \bar{x}a$ be the projection point, then $a^T(b - \bar{x}a) = 0$. It follows that

$$p = \bar{x}a = a\frac{a^T b}{a^T a} = Pb, \text{ where } P = \frac{aa^T}{a^T a}$$

- P is a projection matrix
  - P is symmetric
  - $P^2 = P$
  - P is invariant w.r.t. the length of a

Why is a projection matrix symmetric?

- For any vectors x, y and matrix A

$$(Ax)^T y = x^T A^T y = x^T (A^T y)$$

  That is, the inner product of Ax and y is the same as the inner product of x and $A^T y$.

- It is easy to see that $P^2 = P$ and P is invariant to the length of a. To see that P is symmetric, note that $\forall x, y$

$$\begin{aligned} x^T(Py) = y^T(Px) &= (Px)^T y \\ \Rightarrow x^T P y &= x^T P^T y \\ \Rightarrow P &= P^T \end{aligned}$$

Schwarz Inequality

- Schwarz inequality

$$|a^T b| \leq ||a|| \, ||b||$$

follows from

$$||b - p||^2 = ||b - \frac{a^T b}{a^T a} a||^2 \geq 0$$

- the angle between vectors a and b

$$\cos\theta = \frac{a^T b}{||a|| \, ||b||}$$

Least Squares Solutions

Single-variable Case

- Assume that the relationship between two quantities a and b is governed by $ax = b$. We have noisy data of $(a_i, b_i), i = 1 \dots m$. Unless $a = [a_1 \dots a_m] = cb$, there is no solution for x. Instead, we want to decide the $\bar{x}$ that minimizes $E^2 = \sum_i (b_i - a_i x)^2$? Such $\bar{x}$ is called the *least squares sultion* to $ax = b$.

- It can be shown, by taking derivative of E^2, that

$$\bar{x} = \frac{a^T b}{a^T a}$$

- $a\bar{x}$ is the projection point of b onto a!

Multiple-variable Case

- Let A be a $m \times n$ matrix, where $m > n$. $Ax = b$ is very likely to be inconsistent.
- We want to find $\bar{x}$ that minimizes $E = ||Ax - b||$, the distance of b to a point Ax in the column space of A.
- The minimum is achieved when $A\bar{x}$ is the projection point of b. I.e., $(b - A\bar{x})$ is in the left nullspace of A,

$$A^T(b - A\bar{x}) = 0 \Rightarrow A^T A\bar{x} = A^T b$$

  The above eqation is also known as the normal equation.

An Example

- 2 variables, 3 equations

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

Apparently the projection of b onto the column space of A is $[4\ 5\ 0]^T$. This is consistent with the formula for p.

Properties of A^TA

- A^TA is symmetric
- A^TA has the same nullspace as A.
- If A has full column rank, then
  - A^TA is invertible.
  - A^TA is positive definite
  - From the normal equation, the projection point of b onto the column space of A is

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

Projection Matrice P

- We can view the point of projection p, as the result of applying a projection matrix P on the vector b. Since

$$p \triangleq P\,b = A(A^TA)^{-1}A^T\,b \Rightarrow P = A(A^TA)^{-1}A^T$$

- $P^2 = P$ and $P^T = P$ (symmetric).
- Conversely, if P is symmetric and $P^2 = P$, then P is a projection matrix onto the column space of P.

$$(b - Pb)^T\;Pa = b^TPa - b^TP^TPa = 0$$

Gram-Schmidt Orthogonalization

Orthonormal Vectors and Basis

- $orthonormal = orthogonal + normal$
- The vectors $q_1, \ldots, q_k$ are orthonormal if

$$q_i^T q_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

- A basis is orthonormal if the basis vectors are orthonormal
  - The standard basis is orthonormal
  - Given a basis, one can create an orthonormal one

Orthogonal Matrices

- A matrix is *orthogonal* if it is square and the column vectors are orthonormal.
  - $Q^TQ = QQ^T = I$. If the columns of a square matrix are orthonormal, so are the rows!
  - $||Qx|| = ||x||$. Length is preserved under orthogonal transformations.
  - Any vector can be written as $b = \sum_i (q_i^T b) q_i$ $(b = Qx \Leftrightarrow x = Q^T b)$.
  - Every vector b is the sum of the projetions onto the lines through the $q's$

Matrices with Orthonormal Column Vectors

- If $m \neq n$, $Q^T (m > n)$ is still the left inverse of Q, i.e., $Q^T Q = I$
- The least squares solution $\bar{x}$ to $Qx = b$, where Q has orthonormal columns, satisfies

$$
\begin{aligned}
Q^T Q\bar{x} &= Q^T b, \\
\Rightarrow \bar{x} &= Q^T b, \ (\bar{x}\text{: optimal coefficients}) \\
\Rightarrow p &\triangleq Q\bar{x}, \ (p\text{: projection point}) \\
&= QQ^T b \triangleq Pb, \ (P\text{: projection matrix})
\end{aligned}
$$

$p = \sum_i (q_i^T b) q_i$ still holds

Gram-Schmidt Process

- The projection to a space is the sum of projections to the vectors in an orthonormal basis of the space
- Given a set of independent vectors, convert it to a set of orthonormal vectors spanning the same space
- The basic idea of Gram-Schmidt process is to subtract from a_j the components in the directions already settled

$$a'_j = a_j - \sum_{i=1}^{j-1} (q_i^T a_j) q_i,$$

  and then normalize a'_j to $q_j = \frac{a'_j}{||a'_j||}$

Example of Gram-Schmidt Process

$$a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \; b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; c = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Applying the process, one by one

$$q_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \; q_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, \; q_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

QR Decomposition

- Every $m \times n$ matrix A with linearly independent columns can be written as
$$A = QR,$$
where Q contains orthonormal column vectors and R is an invertible upper-triangular matrix
- Gram-Schmidt process: columns of A = initial vectors, columns of Q = orthonormal vectors, and columns of R = the combinations from q_i's to a_i's
- $A^T A\bar{x} = A^T b \Rightarrow R\bar{x} = Q^T b$, is easy to solve since R is triangular

Example of QR Decomposition

$$
\begin{aligned}
A &= \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & \sqrt{2} \\ 0 & 1/\sqrt{2} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix} \\
&= QR
\end{aligned}
$$

Review

Systems of Linear Equations

- Solution for $Ax = b$ when A is invertible
  - Gauss elimination
- Solution for $Ax = b$ when A is rectangular
  - check if b is in the column space of A
- Least squares solution for $Ax = b$
  - solve $A^T A\bar{x} = A^T b$

Factorizations/Decompositions

- LU factorizations
- QR decompositions
- There are others
  - Cholesky factorizations
  - Reduced factorizations
  - Singular value decompositions

Intersection and Addition of Subspaces

- The intersection of two subspace V and W of S, defined by $V \cap W = \{x|x \in V \text{ and } x \in W\}$ is a subspace of S

$$dim(V \cap W) \leq \min(dim\, V, dim\, W)$$

- the sum of two subspace V and W of S, defined by $V + W = \{x|x = v + w, v \in V \text{ and } w \in W\}$ is a subspace of S.

$$dim(V + W) \leq dim\, V + dim\, W$$

- Theorem

$$dim(V + W) + dim(V \cap W) = dim\, V + dim\, W$$

Proof

Let's prove that

$$dim(V + W) + dim(V \cap W) = dim\ V + dim\ W$$

Consider bases of V and W and put them in a matrix $D = [S_V|S_W]$. A vector $y \in V + W$ can be written as $y = v + w = S_V c + S_W d$, so the dimension of $V + W$ is the same as the dimension of the column space of D. In addition, the dimension of $V \cap W$ is the same as the nullspace of D, since every vector $x \in \mathcal{N}(D)$ corresponds one-to-one to a vector $y \in V \cap W$. The above relation is a result of the fundamental theorem of linear algebra.

Fundamental Subspaces of AB

- $\mathcal{N}(B) \subset \mathcal{N}(AB)$
- $\mathcal{R}(AB) \subset \mathcal{R}(A)$
- $\mathcal{N}(A^T) \subset \mathcal{N}(B^T A^T)$
- $\mathcal{R}(B^T A^T) \subset \mathcal{R}(B^T)$
- $r(AB) \leq r(A), r(AB) \leq r(B),$
  $dim\, \mathcal{N}(AB) \geq dim\, \mathcal{N}(B)$