Linear Algebra

Sartaj Ul Hasan

विद्याधनं सर्वधन प्रधानम्

Department of Mathematics
Indian Institute of Technology Jammu
Jammu, India - 181221

Email: sartaj.hasan@iitjammu.ac.in

Lecture 38

(Nov 08, 2019)

Inner Products

Orthogonal Decomposition

- **Theorem 5 (Orthogonal Decomposition Theorem (ODT)):** Let W be any finite-dimensional subspace of V. Then each vector y in V can be written **uniquely** in the form $y = \hat{y} + z$, where $\hat{y}$ is in W, and z is in $W^{\perp}$. In fact, if $\{u_1, u_2, \ldots, u_p\}$ is any orthogonal basis of W, then $\hat{y} = c_1 u_1 + c_2 u_2 + \cdots + c_p u_p$ with $c_j = \frac{<y, u_j>}{<u_j, u_j>}$ for $j = 1, \ldots, p$ and $z = y - \hat{y}$.
- **Theorem 5 (Alternative Statement):** Given any finite-dimensional subspace W of V, then we can express $V = W + W^{\perp}$, with $W \cap W^{\perp} = \{0\}$, i.e., $V = W \oplus W^{\perp}$.
- **Remark 1:** Hence, every vector can be *uniquely* expressed as the sum of a vector in W and a vector in $W^{\perp}$, i.e., as the sum of two vectors *which are orthogonal to each other* !!

- **Note 1:** The vector $\hat{y}$ is called the **orthogonal projection of y onto W**, written $\text{proj}_W y$. In case $W = \text{Span}\{u\}$ is a one-dimensional subspace, the expression is simplified to: $\hat{y} = \frac{<y,u>}{<u,u>} u$, which is simply called the orthogonal projection of y onto u.
- **Note 2:** In case y belongs to W, its orthogonal projection onto W is itself, i.e. $\hat{y} = y$ for $y \in W$. This follows from Proposition 48 along with ODT (Theorem 5).

Proof of Theorem 5 (ODT)

We will prove the original version, from which the alternative version can be derived immediately.
We assume that any finite-dimensional subspace W of an inner product space has an orthogonal basis.
This assumption is due to Theorem 6 (Gram-Schmidt Process) which will be done later – however, the proof of Theorem 6 does not require Theorem 5, so assumption is logically valid.
Uniqueness: We will first prove uniqueness, i.e., any vector $y \in V$ can not be expressed in more than one way as a sum of a vector in W and a vector in $W^{\perp}$. Suppose BWOC that:

$$y = \hat{y} + z$$
$$\text{and } y = \hat{y}_1 + z_1$$

where $\hat{y}, \hat{y}_1 \in W$ and $z, z_1 \in W^{\perp}$.
Subtracting $0 = (\hat{y} - \hat{y}_1) + (z - z_1)$ or $(\hat{y} - \hat{y}_1) = -(z - z_1)$

Proof of Theorem 5 (ODT)(Cont'd)

LHS $\in W$, RHS $\in W^{\perp}$, i.e., they both are in $W \cap W^{\perp} = \{0\}$ by Prop 47. Therefore, $\hat{y} - \hat{y_1} = 0 \implies \hat{y} = \hat{y_1}$ and $z - z_1 = 0 \implies z = z_1$. This proves the uniqueness.

Existence: We next prove that a decomposition exists, i.e., that $y = \hat{y} + z$, where $\hat{y} \in W$ and $z \in W^{\perp}$. In fact, put $\hat{y} = c_1 u_1 + \cdots + c_p u_p$, where $c_j = \frac{<y, u_j>}{<u_j, u_j>}$ for $j = 1, 2, \ldots, p$. Now, put $z = y - \hat{y}$ so abviously $y = \hat{y} + z$.

It only remains to show that $z \in W^{\perp}$. We use Prop 47(a) for this, so it suffices to show that $< z, u_j >= 0$ for $j = 1, 2, \ldots, p$. But,

$$
\begin{aligned}
< z, u_j > &=< y - \hat{y}, u_j > \\
&=< y, u_j > - < \hat{y}, u_j > \\
&=< y, u_j > -[c_1 < u_1, u_j > + \cdots + c_p < u_1, u_p >] \\
&=< y, u_j > -c_j < u_j, u_j > \\
&=< y, u_j > -\frac{< y, u_j >}{< u_j, u_j >} < u_j, u_j >= 0, \text{ as desired.}
\end{aligned}
$$

Orthogonal Bases (Cont'd)

Theorem 6 (The Gram-Schmidt Orthogonalization Process): Given a basis $\{x_1, x_2, \ldots, x_p\}$ for a subspace W of V, we can generate an orthogonal basis $\{v_1, v_2, \ldots, v_k\}$ for W such that Span $\{v_1, v_2, \ldots, v_k\} =$ Span $\{x_1, x_2, \ldots, x_k\}$ for $k = 1, 2, \ldots, p$. In fact, the vectors v_j are defined as follows:

$$
\begin{aligned}
v_1 &= x_1 \\
v_2 &= x_2 - \frac{< x_2, v_1 >}{< v_1, v_1 >} v_1 \\
v_3 &= x_3 - \frac{< x_3, v_1 >}{< v_1, v_1 >} v_1 - \frac{< x_3, v_2 >}{< v_2, v_2 >} v_2 \\
&\vdots \\
v_p &= x_p - \frac{< x_p, v_1 >}{< v_1, v_1 >} v_1 - \frac{< x_p, v_2 >}{< v_2, v_2 >} v_2 - \ldots \\
&\quad \ldots - \frac{< x_p, v_{p-1} >}{< v_{p-1}, v_{p-1} >} v_{p-1}
\end{aligned}
$$

Obtaining an Orthogonal Basis

- **Description of the Gram-Schmidt Process:** At each stage, subtract from the original basis vector x_i its projection onto the span of the previously obtained orthogonal vectors $v_1, v_2, \ldots, v_{i-1}$.
- **Remark 1:** The process uses the idea we already used in ODT, of subtracting the orthogonal projection onto a subspace from the original vector. A formal proof that the vectors $\{v_1, v_2, \ldots, v_k\}$ form an orthogonal set and that Span $\{v_1, v_2, \ldots, v_k\} =$ Span $\{x_1, x_2, \ldots, x_k\}$ can be done by induction on k.
- **Remark 2:** We can obtain an orthonormal basis for every subspace W of V by normalizing each vector in an orthogonal basis (dividing each of the vectors by its norm). Each vector in an orthonormal basis has norm 1.

  Note: This step is usually left to the end, because square roots can emerge.

Example for Gram-Schmidt Process

Construct an orthonormal basis for $\mathbb{R}^3$ starting with the basis:

$$x_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, x_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\text{Put} \quad v_1 = x_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \tag{1}$$

$$\text{Then} \quad v_2 = x_2 - \frac{< v_1, x_2 >}{< v_1, v_1 >} v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{13}{14} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} -6/7 \\ 15/14 \\ 3/14 \end{bmatrix} \tag{2}$$

Example for Gram-Schmidt Process (Cont'd)

$$\text{Then} \quad v_3 = x_3 - \frac{< v_1, x_3 >}{< v_1, v_1 >} v_1 - \frac{v_2, x_3}{v_2, v_2} v_2$$
$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{7} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \frac{2}{9} \begin{bmatrix} -6/7 \\ 15/14 \\ 3/14 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ -1/3 \end{bmatrix} \tag{3}$$

$\longrightarrow$ We can check that v_1, v_2, v_3 are orthogonal (where as the original basis vectors were not!)

$$< v_1, v_2 > = -12/7 + 15/14 + 9/14 = 0$$
$$< v_1, v_3 > = 2/3 + 1/3 - 3/3 = 0$$
$$< v_2, v_3 > = -6/21 + 15/42 - 3/42 = 0$$

Example for Gram-Schmidt Process (Cont'd)

$\longrightarrow$ If we desire an orthonormal basis, we divide each vector v_i by its length to get:

$$v_1' = \frac{1}{\sqrt{14}}\begin{bmatrix}2\\1\\3\end{bmatrix}, v_2' = \frac{1}{\sqrt{42}}\begin{bmatrix}-4\\5\\1\end{bmatrix}, v_3' = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\-1\end{bmatrix}$$

Some Other Results Related to Orthogonality

Proposition 49 (Pythagorean Theorem): u and v are orthogonal to each other if and only if $||u+v||^2 = ||u||^2 + ||v||^2$.

Proof:

$$\begin{aligned} ||u+v||^2 = < u+v, u+v > &= < u,u > + < v.v > +2 < u,v > \\ &= ||u||^2 + ||v||^2 + 2 < u,v > \end{aligned}$$

$$\therefore ||u+v||^2 = ||u||^2 + ||v||^2 \iff 2 < u,v >= 0 \iff u \perp v$$

Proposition 50 (Best Approximation Theorem): Let W be any finite-dimensional subspace of V, y any vector in V, and $\hat{y}$ be the orthogonal projection of y onto W. Then $||y - \hat{y}|| < ||y - v||$ for all v in W distinct from $\hat{y}$, in other words, $\hat{y}$ is the closest vector (point) in W to y.

Some Other Results Related to Orthogonality (Cont'd)

Proof of Proposition 50: Let $v \in W$. Then:

$$\begin{aligned} ||y-v||^2 &= < y-v, y-v > \\ &= < (y-\hat{y}) + (\hat{y}-v), (y-\hat{y}) + (\hat{y}-v) > \\ &= < y-\hat{y}, y-\hat{y} > + < \hat{y}-v, \hat{y}-v > + \\ & \ 2 < y-\hat{y}, \hat{y}-v > \end{aligned} \tag{4}$$

Now, $y - \hat{y} \in W^{\perp}$ whereas $\hat{y} - v \in W$. Hence, the 3rd term on RHS of (4) is 0. Therefore,

$$||y-v||^2 = ||y-\hat{y}||^2 + ||\hat{y}-v||^2 \tag{5}$$

Now, if $y = v$, then $y \in W \implies y = \hat{y} = v$, which is not allowed. Hence, $||\hat{y} - v||^2 > 0$ whence $||y - v|| > ||y - \hat{y}||$ from (5).

Some Other Results Related to Orthogonality (Cont'd)

Corollary 50.1: If v is any vector, and W is a finite-dimensional subspace, then: $||\text{proj}_W v|| \leq ||v||$.

Proof: We have that $v = \text{Proj}_W v + z$, where $z \in W^\perp$

Applying Pythagorean Theorem yields the following,

$$||v||^2 = ||\text{Proj}_W v + z||^2 = ||\text{Proj}_W v||^2 + ||z||^2$$

Since $||z||^2 \geq 0$, the result follows.

Proposition 51: (The Cauchy-Schwarz Inequality): For all u, v in V, $| < u, v > | \leq ||u||||v||$.

Proof: Clearly result holds if either u or $v = 0$. So we may assume both u and v are non-zero, and apply Cor 50.1 above, taking $W = \text{Span}\{v\}$.

$$\therefore ||\text{Proj}_W u|| \leq ||u|| \tag{6}$$

$$\text{In (6), LHS} = ||\text{Proj}_W u|| = ||\frac{< u, v >}{< v, v >} v|| = \frac{| < u, v > | \ \ ||v||}{||v||^2} \tag{7}$$

From (6) and (7), $| < u, v > | \leq ||u||||v||$ as required.

Some Other Results Related to Orthogonality (Cont'd)

Proposition 52 (The Triangle Inequality): For all u, v in V, $||u+v|| \leq ||u|| + ||v||$.

Proof: We have,

$$\begin{aligned} ||u+v||^2 &= < u+v, u+v > \\ &= ||u||^2 + ||v||^2 + 2 < u, v > \\ &\leq ||u||^2 + ||v||^2 + 2| < u, v > | \\ &\leq ||u||^2 + ||v||^2 + 2||u|| \ \ ||v||, \text{using C-S inequality} \\ &= (||u|| + ||v||)^2. \end{aligned}$$

Hence the result follows.

Best Approximation; Least squares

- First we shall observe and prove the following facts about a rectangular matrix $A_{m\times n}$:
  - $(\text{Row } A)^{\perp} = \text{Nul } A$.
  - $(\text{Col } A)^{\perp} = \text{Nul } A^T$.
  - The matrix A^TA is symmetric matrix and Nul $(A^TA) = \text{Nul } A$.
    Proof: Certainly if $AX = 0$ then $A^TAX = 0$. Vectors X in the nullspace of A are also in the nullspace of A^TA. To go in the other direction, start by supposing that $A^TAX = 0$, and take the inner product with X to show that
    $$< X, A^TAX >= X^TA^TAX = 0, \quad \text{or } ||AX||^2 = 0, \quad \text{or } AX = 0.$$
    Thus, the two nullspaces are identical.
  - If the columns of A are linearly independent, then Nul $A = \{0\}$.
  - A has independent columns $\iff$ A^TA is square, symmetric, and invertible.
    Proof: Since columns of A are LI, we have Nul $A = \{0\}$. Thus, Nul $A^TA = \{0\}$, and hence A^TA is invertible by TIMT.

Best Approximation; Least squares

- Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Suppose the system $AX = b$ is an *inconsistent* system. So no exact solution is possible, so we will look for a vector X that comes "as close as possible" to being a solution in the sense that it minimizes $||b - AX||$ with respect to the usual inner product (dot product) on $\mathbb{R}^m$. One may think of AX as an approximation to b and $||b - AX||$ as the *error* in that approximation– the smaller the error, the better the approximation.

Definition

If $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, **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 $\mathbb{R}^n$.

- The adjective "least-squares" arises from the fact that $||b - AX||$ is the square root of a sum of squares.

Best Approximation; Least squares

- **Proposition 53**: The set of least-squares solutions of $AX = b$ coincides with the non-empty set of solution of the **normal equations** $A^T AX = A^T b$.
  - **Example:** Find the least-squares solution of the inconsistent system $AX = b$ for

$$A = \begin{bmatrix} 4 & 2 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Ans: $\hat{X} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

- **Proposition 54**: The matrix $A^T A$ is invertible if and only if the columns of A are linearly independent. In this case, the equation $AX = b$ has only one least-squares solution $\hat{X}$, and it is given by

$$\hat{X} = (A^T A)^{-1} A^T b.$$

  (Proof of this proposition is left as an exercise!)

- **Definition**: 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.
  - **Example**: Given A and b as below, determine the least-squares error in the least-square solution of $AX = b$.

$$A = \begin{bmatrix} 4 & 2 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$