Lecture 22: Applications of Orthogonalization

Comment

Recall that if $W \subseteq \mathbb{R}^n$ is a subspace with orthonormal basis $u_1, \ldots, u_r$, then for any $\mathbf{x} \in U$,

$$\mathbf{x} = (\mathbf{x} * u_1) \cdot u_1 + \cdots + (\mathbf{x} * u_r) \cdot u_r.$$

If instead, we take an orthogonal basis $w_1, \ldots, w_r$, then the expression for $\mathbf{x}$ becomes

$$\mathbf{x} = \frac{\mathbf{x} * w_1}{w_1 * w_1} \cdot w_1 + \cdots + \frac{\mathbf{x} * w_r}{w_r * w_r} \cdot w_r.$$

To see why: Suppose $\mathbf{x} = \alpha_1 w_1 + \cdots \alpha_r w_r$. Then $\mathbf{x} * w_1 =$

$(\alpha_1 w_1 + \cdots + \alpha_r w_r) * w_1 = \alpha_1(w_1 * w_1) + \cdots + \alpha_r(w_r * w_1) = \alpha_1(w_1 * w_1)$.

Thus, $\mathbf{x} * w_1 = \alpha_1(w_1 * w_1)$, which shows that $\alpha_1 = \frac{\mathbf{x} * w_1}{w_1 * w_1}$.

A similar calculation shows that each $\alpha_i = \frac{\mathbf{x} * w_i}{w_i * w_i}$, for all i.

Example

Let $U \subset \mathbb{R}^3$ have basis $v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, and take $\mathbf{x} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, a vector in U. Find an orthogonal basis for U and express $\mathbf{x}$ in terms of that basis.

Solution. We first apply Gram-Schmidt to the original basis. Take $w_1 = v_1$. Calculate $w_1 * w_2 = 1$ and $w_1 * w_1 = 2$. Therefore,

$$w_2 = v_2 - \frac{1}{2} w_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix}$$

To avoid a fractions, we may take $w_2 = 2 \cdot \begin{bmatrix} -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

Example continued

Now: $x * w_1 = 8$ and $x * w_2 = 12$. In addition: $w_1 * w_1 = 2$ and $w_2 * w_2 = 6$. Thus:

$$\frac{\mathbf{x} * w_1}{w_1 * w_1} \cdot w_1 + \frac{\mathbf{x} * w_2}{w_2 * w_2} \cdot w_2 = \frac{8}{2} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{12}{6} \cdot \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = \mathbf{x},$$

as required.

What if $\mathbf{x}$ is not in the subspace U?

Definition

Suppose $w_1, \ldots, w_r$ is an orthogonal basis for the subspace $W \subseteq \mathbb{R}^n$. Take any vector $\mathbf{x}$ in $\mathbb{R}^n$. The **orthogonal projection of x onto** U is the vector:

$$\mathbf{p}_U\mathbf{x} = \frac{\mathbf{x} * w_1}{w_1 * w_1} \cdot w_1 + \cdots + \frac{\mathbf{x} * w_r}{w_r * w_r} \cdot w_r.$$

Note. If $u_1, \ldots, u_r$ is an orthonormal basis for W, then the orthogonal projection of $\mathbf{x}$ onto U is just:

$$\mathbf{p}_U\mathbf{x} = (\mathbf{x} * u_1) \cdot u_1 + \cdots + (\mathbf{x} * u_r) \cdot u_r.$$

For Example: If U is the subspace in the previous example and $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$,

then:

$$\mathbf{p}_U\mathbf{x} = \frac{\mathbf{x} * w_1}{w_1 * w_1} \cdot w_1 + \frac{\mathbf{x} * w_2}{w_2 * w_2} \cdot w_2 = \frac{2}{2} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{2}{6} \cdot \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{4}{3} \end{bmatrix} = \frac{1}{3} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}.$$

Theorem

Important properties of the orthogonal projection.

Let U be a subspace of $\mathbb{R}^n$. For any vector $\mathbf{x}$:

(i) $\mathbf{x} - \mathbf{p}_U\mathbf{x}$ is orthogonal to every vector in U.

(ii) $\mathbf{p}_U\mathbf{x}$ is the closest vector in U to $\mathbf{x}$.

In other words, for all $\mathbf{y}$ in U, $||\mathbf{p}_U\mathbf{x} - \mathbf{x}|| \leq ||\mathbf{y} - \mathbf{x}||$.

That is, the distance from $\mathbf{x}$ to $\mathbf{p}_U\mathbf{x}$ is less than or equal the distance from $\mathbf{x}$ to any other vector in U.

Why: (i) Take w_1 the first element in the orthogonal basis. Then

$$(\mathbf{x} - \mathbf{p}_U\mathbf{x}) * w_1 = \mathbf{x} * w_1 - (\frac{\mathbf{x} * w_1}{w_1 * w_1} \cdot w_1 + \cdots + \frac{\mathbf{x} * w_r}{w_r * w_r} \cdot w_r) * w_1$$

$$= x * w_1 - \frac{\mathbf{x} * w_1}{w_1 * w_1} \cdot (w_1 * w_1) + \cdots + \frac{\mathbf{x} * w_r}{w_r * w_r} \cdot (w_r * w_1) = x * w_1 - x * w_1 = 0.$$

A similar calculation shows that $\mathbf{x} - \mathbf{p}_U\mathbf{x}$ is orthognonal to every basis vector for U and this implies $\mathbf{x} - \mathbf{p}_U\mathbf{x}$ is orthogonal to every vector in U.

(ii) Take $\mathbf{y}$ any vector in U. Set $\mathbf{p} = \mathbf{p}_U\mathbf{x}$. Then $\mathbf{y} - \mathbf{p}$ belongs to U and is thus orthogonal to $\mathbf{p} - \mathbf{x}$.

Since $\mathbf{y} - \mathbf{x} = (\mathbf{y} - \mathbf{p}) + (\mathbf{p} - \mathbf{x})$, and these vectors form a right triangle, applying the Pythagorean theorem we get:

$$||\mathbf{y} - \mathbf{x}||^2 = ||\mathbf{y} - \mathbf{p}||^2 + ||\mathbf{p} - \mathbf{x}||^2 \geq ||\mathbf{p} - \mathbf{x}||^2,$$

Thus, $||\mathbf{p} - \mathbf{x}|| \leq ||\mathbf{y} - \mathbf{x}||$, which is what we want.

For Example: In the example above, $\mathbf{p}_U\mathbf{x} = \frac{1}{3} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$ is the closest vector in the subspace $U = \text{span}\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \}$ to the vector $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Moreover $u = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$ in in U and

$$u * (\mathbf{p}_U\mathbf{x} - \mathbf{x}) = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} * (\frac{1}{3} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}) = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} * \frac{1}{3} \cdot \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = 0.$$

IMPORTANT APPLICATION

Let A be any $m \times n$ matrix, and suppose $A \cdot \mathbf{X} = \mathbf{b}$ is a system of equations with no solutions.

Let U denote the column space of A, i.e., the subspace of $\mathbb{R}^m$ spanned by the columns of A.

Then any solution $\mathbf{z}$ to the system $A \cdot \mathbf{X} = \mathbf{p}_U \mathbf{b}$ is the best approximation to a solution of the given system.

WHY: Recall that the system has a solution if and only if there is a vector $\mathbf{z}$ such that $A \cdot \mathbf{z} = \mathbf{b}$. In other words, if and only if $\mathbf{b}$ belongs to the column space of A.

Equivalently: If there is no solution, $\mathbf{b}$ is not in the subspace U. Since the closest vector in U to $\mathbf{b}$ is $\mathbf{p}_U \mathbf{b}$, any vector $\mathbf{z}$ satisfying $A \cdot \mathbf{z} = \mathbf{p}_U \mathbf{b}$ can be regarded as the best approximation to a solution of the original system.

Example

Consider the system

$$\begin{aligned} x &= 1 \\ y &= 1 \\ x + y &= 1. \end{aligned}$$

It has no solution. Find the best approximation to a solution.

The system is $A \cdot \mathbf{x} = \mathbf{b}$, for $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. We must solve the system $A \cdot \mathbf{x} = \mathbf{p}_U \mathbf{x}$.

The column space of A is generated by the vectors v_1, v_2 from the example above. Thus, $\mathbf{p}_U \mathbf{x} = \frac{1}{3} \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$.

Example continued

Using Guassian elimination:

$$\left[\begin{array}{cc|c} 1 & 0 & \frac{2}{3} \\ 0 & 1 & \frac{2}{3} \\ 1 & 1 & \frac{4}{3} \end{array}\right] \xrightarrow[-R_2+R_3]{-R_1+R_3} \left[\begin{array}{cc|c} 1 & 0 & \frac{2}{3} \\ 0 & 1 & \frac{2}{3} \\ 0 & 0 & 0 \end{array}\right].$$

Thus, $x = \frac{2}{3}$ and $y = \frac{2}{3}$ is the best approximation to a solution of the original solution.

Class Example

Find the best approximate solution to the system

$$\begin{aligned} x - y &= -1 \\ y &= 2 \\ x + y &= 1. \end{aligned}$$

Write down the augmented matrix you would solve to find the best approximate solution to the given system.

Note that the columns of the coefficient matrix are orthogonal.

Use the formula $\mathbf{p}_U\mathbf{x} = \frac{\mathbf{x} * w_1}{w_1 * w_1} \cdot w_1 + \frac{\mathbf{x} * w_2}{w_2 * w_2} \cdot w_2$.

Class Example continued

Solution. The given system becomes $A \cdot \mathbf{X} = \mathbf{b}$, for $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

We must solve the system $A \cdot \mathbf{X} = \mathbf{p}_U \mathbf{b}$, for U the column space of A.

$$\mathbf{p}_U \mathbf{b} = \frac{\mathbf{b} * w_1}{w_1 * w_1} \cdot w_1 + \frac{\mathbf{b} * w_2}{w_2 * w_2} \cdot w_2 = \frac{0}{2} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{4}{3} \cdot \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \frac{4}{3} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus, we need to work with the augmented matrix $\left[\begin{array}{cc|c} 1 & -1 & -\frac{4}{3} \\ 0 & 1 & \frac{4}{3} \\ 1 & 1 & \frac{4}{3} \end{array}\right]$.

Comment

An Alternate method. Suppose the system $A \cdot \mathbf{X} = \mathbf{b}$ deos not have a solution. Then any vector $\mathbf{z} \in \mathbb{R}^n$ satisfying

$$A^t A \cdot \mathbf{z} = A^t \mathbf{b}$$

is the best approximation to a solution of the original system.

Example. For the same example above, with $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$,

$$A^t A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

and $A^t \cdot \mathbf{b} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Example continued. Thus, a best approximation is a solution $\mathbf{z}$ to the system $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$. Using Gaussian elimination:

$$\left[\begin{array}{cc|c} 2 & 1 & 2 \\ 1 & 2 & 2 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 2 & 1 & 2 \end{array}\right] \xrightarrow{-2 \cdot R_1 + R_1} \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & -3 & -2 \end{array}\right]$$

$$\xrightarrow{-\frac{1}{3} \cdot R_2} \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 1 & \frac{2}{3} \end{array}\right] \xrightarrow{-2 \cdot R_2 + R_1} \left[\begin{array}{cc|c} 1 & 0 & \frac{2}{3} \\ 0 & 1 & \frac{2}{3} \end{array}\right]$$

Thus, $x = \frac{2}{3}$ and $y = \frac{2}{3}$ is the best approximation to a solution of the original system.

Note that this agrees with the answer obtained previously by the first technique.

Comment

Remark. Given the system $A \cdot \mathbf{X} = \mathbf{b}$ with no solution:

(i) That the second method works depends on the fact that $\mathbf{p}_U\mathbf{x} - \mathbf{x}$ is orthogonal to every vector in the column space of A. A proof can be found in our text.

(ii) Regardless of the method, a best approximate solution $\mathbf{z}$ need not be unique.

(iii) If the square matrix A^tA is invertible, then the system $(A^tA) \cdot \mathbf{X} = A^t\mathbf{b}$ has a unique solution $\mathbf{z} = (A^tA)^{-1}A^t\mathbf{b}$.

In this case, there is a unique best approximate solution.

SECOND IMPORTANT APPLICATION

Curve Fitting. Suppose we wish to find a quadratic function

$$f(x) = a_0 + a_1 x + a_2 x^2$$

that best fits given data (1,3), (-1, 2), (2, 6), (3, 10). To do this, we find the best approximation to a system that would have a solution if the given points were exactly on a quadratic polynomial.

Solution. If there were a polynomial $f(x)$ as above with these data points, then:

$$\begin{aligned} a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 &= 3 \\ a_0 + a_1 \cdot (-1) + a_2 \cdot (-1)^2 &= 2 \\ a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 &= 6 \\ a_0 + a_1 \cdot 3 + a_2 \cdot 3^2 &= 10. \end{aligned}$$

Solution continued. Thus, we consider the system $A \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \mathbf{b}$, where

$\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ 6 \\ 10 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$. We seek a best approximate solution to this system.

That is, a solution $\mathbf{z}$ to the system $(A^t A) \cdot \mathbf{X} = A^t \cdot \mathbf{b}$, where $\mathbf{X} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$.

Computer software shows we need a solution to the equation:

$$\begin{bmatrix} 4 & 5 & 15 \\ 5 & 15 & 35 \\ 15 & 35 & 99 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 21 \\ 43 \\ 119 \end{bmatrix}.$$

The coefficient matrix in this sytem has non-zero determinant, so there will exists a unique best approximation to the original system $A^t A \cdot \mathbf{X} = A^t \cdot \mathbf{b}$.

Example continued. Linear algebra software yields the solution

$$a_0 = \frac{20}{11}, a_1 = \frac{31}{55}, a_2 = \frac{8}{11}.$$

Thus, $f_0(x) = \frac{20}{11} + \frac{31}{55}x + \frac{8}{11}x^2$ best fits the given data.

Let's see how well $f_0(x)$ fits the data.

$$\begin{aligned}
f_0(1) &= \frac{20}{11} + \frac{31}{55} \cdot 1 + \frac{8}{11} \cdot 1^2 = \frac{171}{55} \approx 3.1 \\
f_0(-1) &= \frac{20}{11} + \frac{31}{55} \cdot (-1) + \frac{8}{11} \cdot (-1)^2 = \frac{109}{55} \approx 1.98 \\
f_0(2) &= \frac{20}{11} + \frac{31}{55} \cdot 2 + \frac{8}{11} \cdot 2^2 = \frac{312}{55} \approx 5.67 \\
f_0(3) &= \frac{20}{11} + \frac{31}{55} \cdot 3 + \frac{8}{11} \cdot 3^2 = \frac{553}{55} \approx 10.1
\end{aligned}$$

Note: These values are close to the exact values of 3, 2, 6, 10, respectively, so $f_0(x)$ fits the data well.

Comment

Given any collection of data points, (x_i, y_i), the method above can be used to find a polynomial of degree d best fitting the data as long as we have at least $d + 1$ data points. When we seek a a polynomial of degree one, i.e., a line of best fit, this method is called the Method of Least Squares.