October 26 Math 3260 sec. 58 Fall 2017

Section 6.3: Orthogonal Projections

Equating points with position vectors, we may wish to find the point $\hat{\mathbf{y}}$ in a subspace W of $\mathbb{R}^n$ that is *closest* to a given point $\mathbf{y}$.

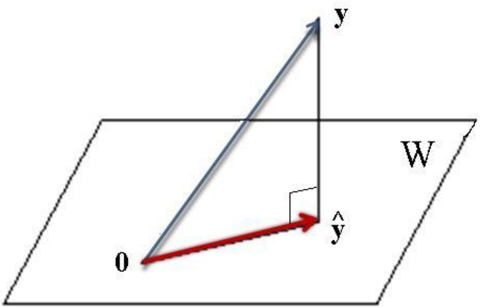

Figure: Illustration of an orthogonal projection. Note that $\text{dist}(\mathbf{y}, \hat{\mathbf{y}})$ is the shortest distance between $\mathbf{y}$ and the points on W.

Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Each vector $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely as a sum

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^\perp$.

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis for W, then

$$\hat{\mathbf{y}} = \sum_{j=1}^{p} \left(\frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j} \right) \mathbf{u}_j, \quad \text{and} \quad \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}.$$

Remark: Note that the basis must be orthogonal, but otherwise the vector $\hat{\mathbf{y}}$ is **independent** of the particular basis used!

Remark: The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of y onto** W. We can denote it

$$\mathrm{proj}_W\, \mathbf{y}.$$

Example

Let $\mathbf{y} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix}$ and

$$W = \operatorname{Span}\left\{ \begin{array}{c} \vec{v}_1 \\ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \end{array}, \begin{array}{c} \vec{v}_2 \\ \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \end{array} \right\}.$$

(a) Verify that the spanning vectors for W given are an orthogonal basis for W.

$$\vec{v}_1 \cdot \vec{v}_2 = 2(-2) + 1(2) + 2(1) = -4 + 2 + 2 = 0$$

So they are orthogonal. Since $\vec{v}_1 \neq c\vec{v}_2$ for all possible scalars c, they are lin. independent.

Hence $\{\vec{v}_1, \vec{v}_2\}$ is an orthogonal basis.

Example Continued...

$$W = \text{Span}\left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix}$$

(b) Find the orthogonal projection of **y** onto W.

$$\hat{y} = \frac{\vec{y}\cdot\vec{v}_1}{\|\vec{v}_1\|^2}\vec{v}_1 + \frac{\vec{y}\cdot\vec{v}_2}{\|\vec{v}_2\|^2}\vec{v}_2$$

$$\vec{y}\cdot\vec{v}_1 = 2(4)+1(8)+2(1) = 8+8+2 = 18$$

$$\vec{y}\cdot\vec{v}_2 = -2(4)+2(8)+1(1) = -8+16+1 = 9$$

$$\|\vec{v}_1\|^2 = 4+1+4 = 9$$

$$\|\vec{v}_2\| = 4 + 4 + 1 = 9$$

So

$$\hat{y} = \frac{18}{9}\vec{v}_1 + \frac{9}{9}\vec{v}_2 = 2\vec{v}_1 + \vec{v}_2$$

$$= 2\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4-2 \\ 2+2 \\ 4+1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

(c) Find the shortest distance between **y** and the subspace W.

If $\vec{y} = \hat{y} + \vec{z}$ then $\vec{z} = \vec{y} - \hat{y}$

$$\vec{z} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ -4 \end{bmatrix}$$

The distance between $\vec{y}$ and W is

$$\|\vec{z}\| = \sqrt{4+16+16} = 6$$

Computing Orthogonal Projections

Theorem: If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis of a subspace W of $\mathbb{R}^n$, and $\mathbf{y}$ is any vector in $\mathbb{R}^n$ then

$$\text{proj}_W\, \mathbf{y} = \sum_{j=1}^{p} (\mathbf{y} \cdot \mathbf{u}_j)\, \mathbf{u}_j.$$

And, if U is the matrix $U = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_p]$, then the above is equivalent to

$$\text{proj}_W\, \mathbf{y} = UU^T\mathbf{y}.$$

Remark: In general, U is not square; it's $n \times p$. So even though UU^T will be a square matrix, it is not the same matrix as U^TU and it is not the identity matrix.

Example

$$W = \text{Span}\left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \right\}$$

Find an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for W. Then compute the matrices $U^T U$ and UU^T where $U = [\mathbf{u}_1 \;\; \mathbf{u}_2]$.

$\|\vec{v}_1\| = \|\vec{v}_2\| = 3$

$$\vec{u}_1 = \frac{1}{\|\vec{v}_1\|}\vec{v}_1 = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$$

$$U = \begin{bmatrix} 2/3 & -2/3 \\ 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$$

$$\vec{u}_2 = \frac{1}{\|\vec{v}_2\|}\vec{v}_2 = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix}$$

$$U^T U = \begin{bmatrix} 2/3 & 1/3 & 2/3 \\ -2/3 & 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 2/3 & -2/3 \\ 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$U U^T = \begin{bmatrix} 2/3 & -2/3 \\ 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 2/3 & 1/3 & 2/3 \\ -2/3 & 2/3 & 1/3 \end{bmatrix} = \begin{bmatrix} \frac{8}{9} & \frac{-2}{9} & \frac{2}{9} \\ \frac{-2}{9} & \frac{5}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{5}{9} \end{bmatrix}$$

Note the symmetry

$$\left(UU^T\right)^T = \left(U^T\right)^T U^T = UU^T$$

Example

$$W = \text{Span}\left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix}$$

Use the matrix formulation to find $\text{proj}_W \mathbf{y}$.

$$\hat{y} = UU^T \vec{y} = \begin{bmatrix} \frac{8}{9} & \frac{-2}{9} & \frac{2}{9} \\ \frac{-2}{9} & \frac{5}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{5}{9} \end{bmatrix} \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix} = \begin{bmatrix} 18/9 \\ 36/9 \\ 45/9 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

Best Approximation Theorem

Suppose W is a subspace of $\mathbb{R}^n$ and $\mathbf{y}$ is a vector in $\mathbb{R}^n$. If $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto W, then $\hat{\mathbf{y}}$ is the *closest* point in W to $\mathbf{y}$. That is

$$\|\mathbf{y}-\hat{\mathbf{y}}\| \le \|\mathbf{y}-\mathbf{v}\|$$

for every point $\mathbf{v}$ in W, and equality occurs if and only if $\mathbf{v}=\hat{\mathbf{y}}$.

Approximate Solution to Inconsisitent System

Suppose we wish to solve a system $A\mathbf{x} = \mathbf{b}$ but it is inconsistent. Note that this means

$$\mathbf{b} \quad \text{is not in Col}A.$$

We seek an approximate solution $\hat{\mathbf{x}}$ by considering instead a system

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}}.$$

To be consistent, we insist that $\hat{\mathbf{b}}$ is in Col$A$.

Question: Of all possible vectors in Col$A$, how should we choose $\hat{\mathbf{b}}$?

Using the Orthogonal Projection

The best approximation $\hat{\mathbf{b}}$ to $\mathbf{b}$ in Col$A$ is the orthogonal projection of $\mathbf{b}$ onto Col$A$!

Recall: The orthogonal complement to the column space of A is the null space of A^T. That is

$$[\mathrm{Col}A]^\perp = \mathrm{Nul}A^T.$$

Least Squares Problem

Suppose we wish to best approximate a solution to

$$A\mathbf{x} = \mathbf{b}.$$

$$\text{Let} \quad \mathbf{b} = \hat{\mathbf{b}} + \mathbf{z}$$

where $\hat{\mathbf{b}}$ is in Col$A$ and $\mathbf{z}$ is orthogonal to Col$A$.

$\vec{z}$ is in $[\text{Col}\,A]^{\perp}$

Show that the system $A^T A\mathbf{x} = A^T\mathbf{b}$ is consistent.

$A\vec{x}$ is never equal to $\vec{b}$.

Letting $\vec{b} = \hat{b} + \vec{z}$ with $\hat{b}$ in Col A and $\vec{z}$ in $[\text{Col} A]^\perp$.

$$A^T\vec{b} = A^T(\hat{b} + \vec{z}) = A^T\hat{b} + A^T\vec{z}$$

$$= A^T\hat{b} + \vec{0}$$

$$= A^T\hat{b}$$

Since $\hat{b}$ is in Col A there is an x, say $\hat{x}$ such that

$$A\hat{x} = \hat{b}$$

So

$$A^T A\vec{x} = A^T\vec{b} \quad \text{has solution} \quad \hat{x}$$

since

$$A^T A\hat{x} = A^T(A\hat{x}) = A^T\hat{b}$$

Least Squares Problem

The system $A^TA\mathbf{x} = A^T\mathbf{b}$ is called the **normal equations** for the system $A\mathbf{x} = \mathbf{b}$.

Theorem: If the columns of A are linearly independent, then there is a unique least squares solution $\hat{\mathbf{x}}$ to the equation $A\mathbf{x} = \mathbf{b}$ that minimizes the error in the sense that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$$

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

Example: Best Fit Line

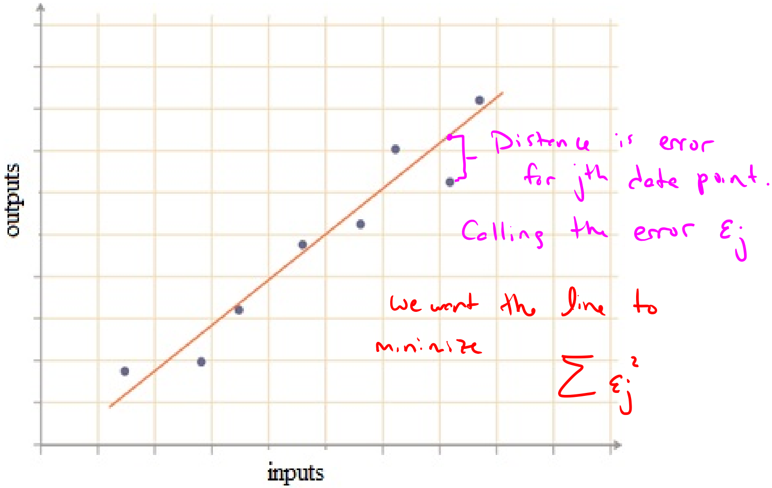

Figure: Given a set of data, we wish to determine the line $y = mx + b$ of best fit.

Example

Find the line of best fit to the data set

$$\{(-1,0),(0,1),(1,2),(2,4)\}.$$

We want a line $y=mx+b$. If they were on this line, we'd have

$$\begin{aligned} -1m + b &= 0 \\ 0m + b &= 1 \\ 1m + b &= 2 \\ 2m + b &= 4 \end{aligned}$$

$$\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

This is inconsistent.

We'll solve $A^T A \vec{x} = A^T \vec{b}$ $\quad \vec{x} = \begin{bmatrix} m \\ b \end{bmatrix}$ $\quad \vec{b} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix}$

$$A^T A = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

$$\det(A^T A) = 24 - 4 = 20$$

$$(A^T A)^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -2 \\ -2 & 6 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$$

$$\begin{bmatrix} m \\ b \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 10 \\ 7 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 13 \\ 11 \end{bmatrix}$$

The best fit line is

$$y = \frac{13}{10}x + \frac{11}{10}$$

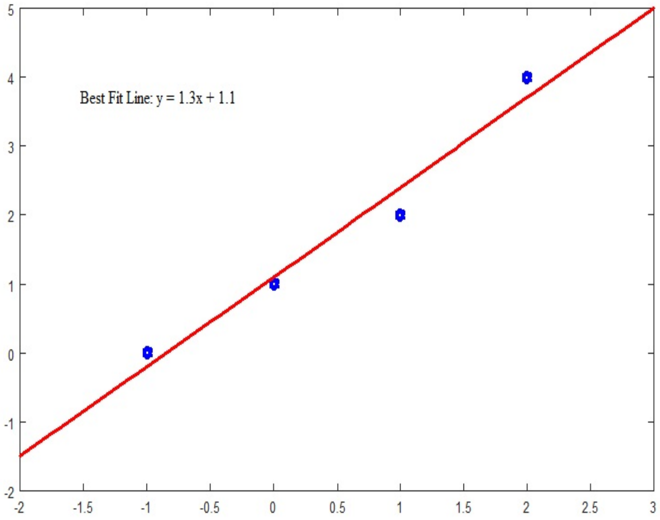

Figure: Our data set along with the least squares, best fit line.

Section 6.4: Gram-Schmidt Orthogonalization

Question: Given any-old basis for a subspace W of $\mathbb{R}^n$, can we construct an orthogonal basis for that same space?

Example: Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} \right\}$. Find an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ that spans W.

We need $\vec{v}_1, \vec{v}_2$ in W, so we take them to be linear combinations of $\vec{x}_1, \vec{x}_2$.

Let's take $\vec{v}_1 = \vec{x}_1$. Let $\vec{v}_2 = c_1\vec{x}_1 + c_2\vec{x}_2$.

c_2 must be nonzero, so we can take $c_2 = 1$.

$$\vec{v}_2 = \vec{x}_2 + c_1 \vec{x}_1 .$$

We require $\vec{v}_1 \cdot \vec{v}_2 = 0$.

$$0 = \vec{v}_1 \cdot (\vec{x}_2 + c_1 \vec{x}_1) = \vec{v}_1 \cdot \vec{x}_2 + c_1 \vec{v}_1 \cdot \vec{x}_1$$

Since $\vec{x}_1 = \vec{v}_1$,

$$0 = \vec{v}_1 \cdot \vec{x}_2 + c_1 \vec{v}_1 \cdot \vec{v}_1$$

$$\Rightarrow \quad c_1 = \frac{-\vec{v}_1 \cdot \vec{x}_2}{\|\vec{v}_1\|^2}$$

So $\vec{v}_1 = \vec{x}_1$ and $\vec{v}_2 = \vec{x}_2 - \dfrac{\vec{v}_1 \cdot \vec{x}_2}{\|\vec{v}_1\|^2} \vec{v}_1$

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}$$

$$\vec{v}_1 \cdot \vec{x}_2 = -2 \qquad \|\vec{v}_1\|^2 = 1+1+1 = 3$$

$$\vec{v}_2 = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -1/3 \\ -1/3 \end{bmatrix}$$

Our orthogonal basis is

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2/3 \\ -1/3 \\ -1/3 \end{bmatrix} \right\}$$

Theorem: Gram Schmidt Process

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ be any basis for the nonzero subspace W of $\mathbb{R}^n$. Define the set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ via

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \left(\frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \left(\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \sum_{j=1}^{p-1}\left(\frac{\mathbf{x}_p \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j}\right)\mathbf{v}_j.
\end{aligned}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. Moreover, for each $k = 1, \ldots, p$

$$\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}.$$