Math 260 Summary of Least Squares, etc.
"One formula to rule them all."

Notation for today:

General case	Orthogonal columns
$A = [\vec{a}_1 \cdots \vec{a}_n]$	$U = [\vec{u}_1 \cdots \vec{u}_n]$ where $\vec{u}_i \cdot \vec{u}_j = 0$ for $i \neq j$
$Col\ A = span\{\vec{a}_1, \dots, \vec{a}_n\}$	$Col\ U = span\{\vec{u}_1, \dots, \vec{u}_n\}$

Goal: given linearly independent columns $\vec{a}_1, \dots, \vec{a}_n$ in R^m, where $m \geq n$ (that is, the number of equations $\geq$ the number of unknowns; and note that if $m < n$ then the system generally has an infinite number of solutions) and a vector $\vec{b}$ in R^m, we want to build $\vec{b}$ using $\vec{a}_1, \dots, \vec{a}_n$. That is, we want to find $x_1, \dots, x_n$ so that

$$x_1\vec{a}_1 + \cdots x_n\vec{a}_n = \vec{b}$$

(or $x_1\vec{a}_1 + \cdots x_n\vec{a}_n \approx \vec{b}$ as well as possible if we can't build $\vec{b}$ exactly), that is,

$$[\vec{a}_1 \cdots \vec{a}_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \vec{b}$$

that is

$$A\vec{x} = \vec{b}\,.$$

The general (least squares) solution $\hat{\vec{x}}$ to $A\vec{x} = \vec{b}$ is

$$\begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_n} \end{bmatrix} = \hat{\vec{\boldsymbol{x}}} = (\boldsymbol{A^T A})^{-1}(\boldsymbol{A^T \vec{b}})$$

This tells us how to best build $\vec{b}$, but it doesn't actually build $\vec{b}$. To build $\vec{b}$:

$$\hat{\vec{\boldsymbol{b}}} = \boldsymbol{A}\hat{\vec{\boldsymbol{x}}} = \boldsymbol{A}(\boldsymbol{A^T A})^{-1}(\boldsymbol{A^T \vec{b}})$$

Four cases

Case 1: $n < m$ (not enough columns in A to span R^m, so A^{-1} does not exist)

Columns of A are not orthogonal

$$\hat{\vec{x}} = (A^T A)^{-1} A^T \vec{b}$$

$$\vec{b} \approx \hat{\vec{b}} = A\hat{\vec{x}} = A(A^T A)^{-1} A^T \vec{b}$$

Case 2: $n = m$ (enough columns in A to span R^m, so A^{-1} exists)

Columns of A are not orthogonal

$$\vec{x} = (A^T A)^{-1} A^T \vec{b} = A^{-1}(A^T)^{-1} A^T \vec{b} = A^{-1}\vec{b}$$

$$\vec{b} = A\vec{x} = AA^{-1}\vec{b} = \vec{b}$$

Case 3: $n < m$ (not enough columns in U to span R^m, so U^{-1} does not exist)

Columns of $U = [\, \vec{u}_1 \, \vec{u}_2 \cdots \vec{u}_n \,]$ are orthogonal

First recall that

$$U^T U = \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} [\, \vec{u}_1 \, \vec{u}_2 \cdots \vec{u}_n \,] = \begin{bmatrix} \vec{u}_1^T\vec{u}_1 & \vec{u}_1^T\vec{u}_2 & \cdots & \vec{u}_1^T\vec{u}_n \\ \vec{u}_2^T\vec{u}_1 & \vec{u}_2^T\vec{u}_2 & \cdots & \vec{u}_2^T\vec{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{u}_n^T\vec{u}_1 & \vec{u}_n^T\vec{u}_2 & \cdots & \vec{u}_n^T\vec{u}_n \end{bmatrix} = \begin{bmatrix} \vec{u}_1^T\vec{u}_1 & 0 & \cdots & 0 \\ 0 & \vec{u}_2^T\vec{u}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \vec{u}_n^T\vec{u}_n \end{bmatrix}$$

so that

$$(U^T U)^{-1} = \begin{bmatrix} \vec{u}_1^T\vec{u}_1 & 0 & \cdots & 0 \\ 0 & \vec{u}_2^T\vec{u}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \vec{u}_n^T\vec{u}_n \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{\vec{u}_1^T\vec{u}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\vec{u}_2^T\vec{u}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\vec{u}_n^T\vec{u}_n} \end{bmatrix}$$

then

$$\hat{\vec{x}} = (U^T U)^{-1} U^T \vec{b} = \begin{bmatrix} \frac{1}{\vec{u}_1^T\vec{u}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\vec{u}_2^T\vec{u}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\vec{u}_n^T\vec{u}_n} \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} \vec{b} = \begin{bmatrix} \frac{1}{\vec{u}_1^T\vec{u}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\vec{u}_2^T\vec{u}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\vec{u}_n^T\vec{u}_n} \end{bmatrix} \begin{bmatrix} \vec{u}_1^T\vec{b} \\ \vec{u}_2^T\vec{b} \\ \vdots \\ \vec{u}_n^T\vec{b} \end{bmatrix} = \begin{bmatrix} \frac{\vec{u}_1^T\vec{b}}{\vec{u}_1^T\vec{u}_1} \\ \frac{\vec{u}_2^T\vec{b}}{\vec{u}_2^T\vec{u}_2} \\ \vdots \\ \frac{\vec{u}_n^T\vec{b}}{\vec{u}_n^T\vec{u}_n} \end{bmatrix} = \begin{bmatrix} \frac{\vec{b}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1} \\ \frac{\vec{b}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2} \\ \vdots \\ \frac{\vec{b}\cdot\vec{u}_n}{\vec{u}_n\cdot\vec{u}_n} \end{bmatrix}$$

Then

$$\vec{b} \approx \hat{\vec{b}} = U\hat{\vec{x}} = U(U^T U)^{-1} U^T \vec{b} = [\, \vec{u}_1 \, \vec{u}_2 \cdots \vec{u}_n \,] \begin{bmatrix} \frac{\vec{b}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1} \\ \frac{\vec{b}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2} \\ \vdots \\ \frac{\vec{b}\cdot\vec{u}_n}{\vec{u}_n\cdot\vec{u}_n} \end{bmatrix} = \frac{\vec{b}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\vec{u}_1 + \frac{\vec{b}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\vec{u}_2 + \cdots + \frac{\vec{b}\cdot\vec{u}_n}{\vec{u}_n\cdot\vec{u}_n}\vec{u}_n$$

Does this look familiar?

Case 4: $n = m$ (enough columns in U to span R^m, so U^{-1} exists)
Columns of U are orthogonal

Same as Case 3, but now exactly (rather than approximately) we have

$$\vec{\boldsymbol{b}} = \frac{\vec{\boldsymbol{b}} \cdot \vec{\boldsymbol{u}}_1}{\vec{\boldsymbol{u}}_1 \cdot \vec{\boldsymbol{u}}_1} \vec{\boldsymbol{u}}_1 + \cdots + \frac{\vec{\boldsymbol{b}} \cdot \vec{\boldsymbol{u}}_n}{\vec{\boldsymbol{u}}_n \cdot \vec{\boldsymbol{u}}_n} \vec{\boldsymbol{u}}_n$$

In Cases 3 and 4, if the columns of U are orthonormal, then $U^T U = I$, and

$$\hat{\vec{b}} = proj_{Col\,U} \vec{b} = \left(\vec{b} \cdot \vec{u}_1\right) \vec{u}_1 + \cdots + \left(\vec{b} \cdot \vec{u}_n\right) \vec{u}_n = U U^T \vec{b}$$

and in Case 4 U is square so $U^T U = I$, which also means that $U U^T = I$ and

$$\hat{\vec{b}} = proj_{Col\,U} \vec{b} = U U^T \vec{b} = \vec{b}$$

exactly.

One final observation. Recall that

$$\hat{\vec{b}} = proj_{Col\,A} \vec{b} = A\,\hat{\vec{x}} = A\,(A^T A)^{-1} A^T \vec{b}.$$

Note that $\hat{\vec{b}} = proj_{Col\,A} \vec{b}$ is in $Col\ A$ since it is matrix A times a vector.

Suppose we were to project $\hat{\vec{b}}$ (rather than $\vec{b}$) onto the column space of A. Then we have

$$\begin{aligned} proj_{Col\,A} \hat{\vec{\boldsymbol{b}}} &= A\hat{\vec{x}} = A(A^T A)^{-1} A^T\, \hat{\vec{\boldsymbol{b}}} \\ &= A(A^T A)^{-1} A^T\, \boldsymbol{A(A^T A)^{-1} A^T \vec{b}} \\ &= A(A^T A)^{-1}\, I\, A^T \vec{b} \\ &= A(A^T A)^{-1} A^T \vec{b} \\ &= \hat{\vec{b}} \end{aligned}$$

Examples of using the least squares formula and the formula for the orthogonal projection of one vector onto others. The four cases are as discussed earlier. The shaded cases below are "bad"—the vectors are not orthogonal.

Matrix	Vector to build	How to build $\vec{b}$, that is, how much of each column of A	The vector we've built	How to build $\vec{b}$, if we assume columns of A are orthogonal	The vector we've built
$A = [\vec{a}_1\ \vec{a}_2]$	$\vec{b}$	$\hat{\vec{x}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (A^TA)^{-1}A^T\vec{b}$	$\hat{\vec{b}} = x_1\vec{a}_1 + x_2\vec{a}_2 = A\hat{\vec{x}}$	$\left(\frac{\vec{b}\cdot\vec{a}_1}{\vec{a}_1\cdot\vec{a}_1}, \frac{\vec{b}\cdot\vec{a}_2}{\vec{a}_2\cdot\vec{a}_2}\right)$	$\hat{\vec{b}} = \frac{\vec{b}\cdot\vec{a}_1}{\vec{a}_1\cdot\vec{a}_1}\vec{a}_1 + \frac{\vec{b}\cdot\vec{a}_2}{\vec{a}_2\cdot\vec{a}_2}\vec{a}_2$
$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$	$\left(\frac{6}{3}, \frac{10}{14}\right)$	$\begin{bmatrix} 2\frac{10}{14} \\ 3\frac{6}{14} \\ 4\frac{2}{14} \end{bmatrix}$
$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ -2 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	$\left(\frac{2}{10}, \frac{2}{20}\right)$	$\begin{bmatrix} \frac{8}{20} \\ 1 \end{bmatrix}$
$\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$	$\left(\frac{6}{3}, -\frac{2}{2}\right)$	$\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$
$\begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -1/5 \\ -3/5 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	$\left(-\frac{1}{5}, -\frac{3}{5}\right)$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

Notice that $\begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ is orthogonal to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, as well as $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.