EECS 16A Designing Information Devices and Systems I
Spring 2019 Discussion 13A

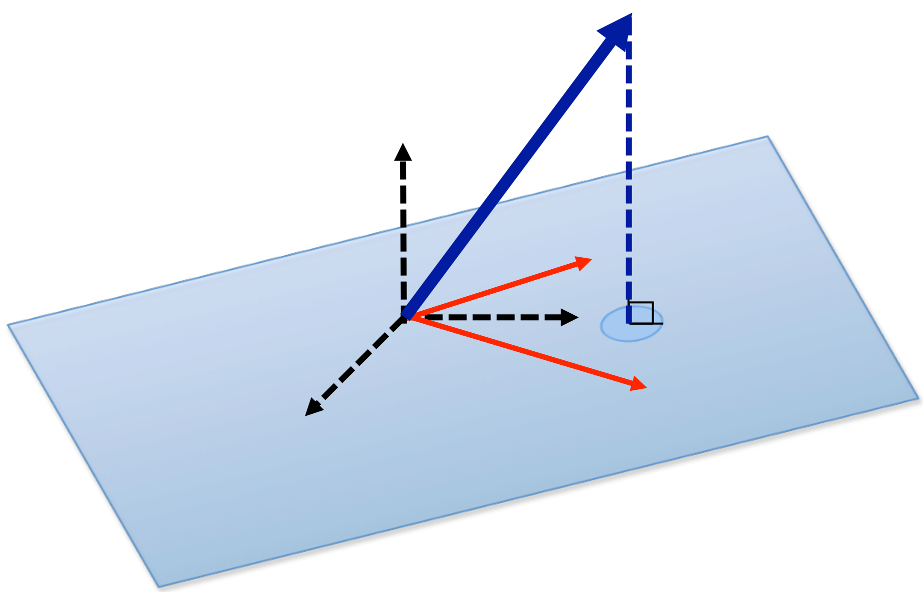
1. Linear Least Squares with Orthogonal Columns

(a) **Geometric Interpretation of Linear Least Squares**
Consider a linear least squares problem of the form

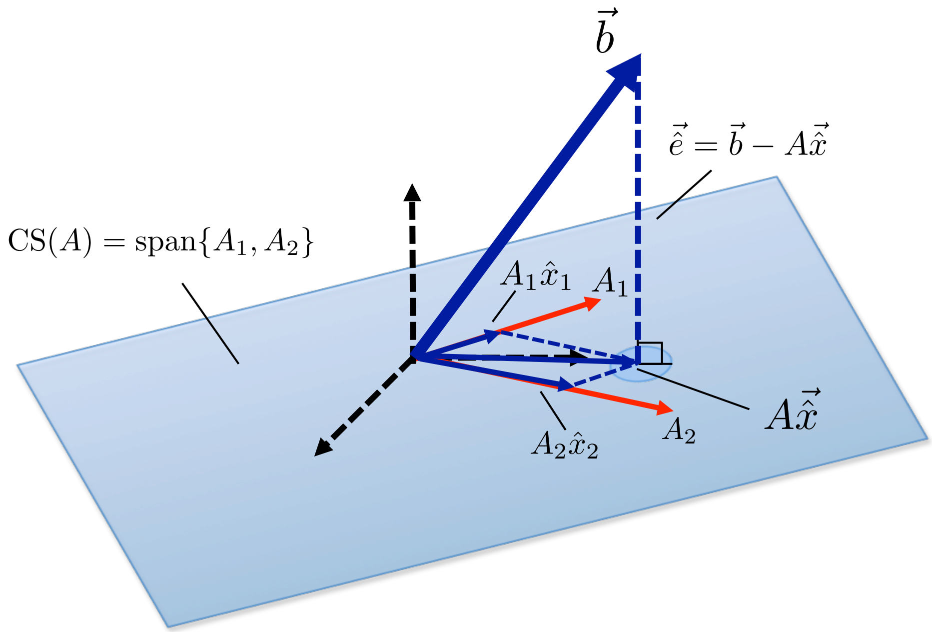
$$\min_{\vec{x}} \left\| \vec{b} - \mathbf{A}\vec{x} \right\|^2 = \min_{\vec{x}} \left\| \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} - \begin{bmatrix} | & | \\ A_1 & A_2 \\ | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2$$

Let the solution be $\vec{\hat{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$.
Label the following elements in the diagram below.

$$\vec{b}, \quad A_1,\ A_2, \quad \text{span}\{A_1, A_2\}, \quad \vec{e} = \vec{b} - \mathbf{A}\vec{\hat{x}}, \quad \mathbf{A}\vec{\hat{x}}, \quad A_1\hat{x}_1,\ A_2\hat{x}_2$$

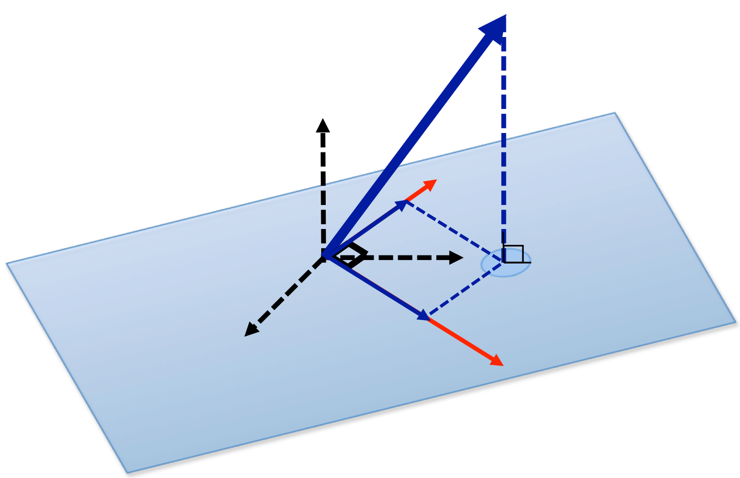


Answer:



(b) We now consider the special case of linear least squares where the columns of **A** are orthogonal (illustrated in the figure below). Use the linear least squares formula $\vec{\hat{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}$ to show that

$$\hat{x}_1 = \text{factor by which } A_1 \text{ is scaled to produce the projection of } \vec{b} \text{ onto } A_1,$$
$$\hat{x}_2 = \text{factor by which } A_2 \text{ is scaled to produce the projection of } \vec{b} \text{ onto } A_2.$$



Answer:

The projection of $\vec{b}$ onto A_1 and A_2 are given by:

$$\text{proj}_{A_1}(\vec{b}) = \frac{\langle A_1, \vec{b}\rangle}{\|A_1\|^2}A_1 \qquad \text{proj}_{A_2}(\vec{b}) = \frac{\langle A_2, \vec{b}\rangle}{\|A_2\|^2}A_2$$

Length: $\frac{\langle A_1, \vec{b}\rangle}{\|A_1\|}$ $\qquad$ $\frac{\langle A_2, \vec{b}\rangle}{\|A_2\|}$

The linear least squares solution is given by:

$$\begin{aligned}
\begin{bmatrix}\hat{x}_1\\ \hat{x}_2\end{bmatrix} &= \left(\begin{bmatrix}- & A_1^T & -\\ - & A_2^T & -\end{bmatrix}\begin{bmatrix}| & |\\ A_1 & A_2\\ | & |\end{bmatrix}\right)^{-1}\begin{bmatrix}- & A_1^T & -\\ - & A_2^T & -\end{bmatrix}\begin{bmatrix}b_1\\ b_2\end{bmatrix}\\
&= \begin{bmatrix}\frac{1}{\|A_1\|^2} & 0\\ 0 & \frac{1}{\|A_2\|^2}\end{bmatrix}\begin{bmatrix}- & A_1^T & -\\ - & A_2^T & -\end{bmatrix}\begin{bmatrix}b_1\\ b_2\end{bmatrix}\\
&= \begin{bmatrix}\frac{A_1^T\vec{b}}{\|A_1\|^2}\\ \frac{A_2^T\vec{b}}{\|A_2\|^2}\end{bmatrix}
\end{aligned}$$

(c) Compute the linear least squares solution to

$$\min_{\vec{x}} \quad \left\|\begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix} - \begin{bmatrix}1&0&0\\0&0&0\\0&1&0\\0&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right\|^2.$$

Answer:

$$\begin{bmatrix}1\\3\\5\end{bmatrix}$$

2. Orthonormal Matrices and Projections

An orthonormal matrix, $\mathbf{A}$, is a matrix whose columns, $\vec{a}_i$, are:

- Orthogonal (ie. $\langle\vec{a}_i,\vec{a}_j\rangle = 0$ when $i \neq j$)
- Normalized (ie. vectors with length equal to 1, $\|\vec{a}_i\| = 1$). This implies that $\|\vec{a}_i\|_2 = \langle\vec{a}_i,\vec{a}_i\rangle = 1$.

(a) Suppose that the matrix $\mathbf{A} \in \mathbb{R}^{N\times M}$ has linearly indpendent columns. The vector $\vec{y}$ in $\mathbb{R}^N$ is not in the subspace spanned by the columns of $\mathbf{A}$. What is the projection of $\vec{y}$ onto the subspace spanned by the columns of $\mathbf{A}$?

Answer: When finding a projection onto a subspace, we're trying to find the "closest" vector in that subspace. This can be found by first finding $\vec{x}$ that minimizes $\|\vec{y}-\mathbf{A}\vec{x}\|$. From least squares, we know that $\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$. The projection of $\vec{y}$ onto the columns of $\mathbf{A}$ is then $\hat{\vec{y}} = \mathbf{A}\hat{\vec{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$.

(b) Show if $\mathbf{A} \in \mathbb{R}^{N\times N}$ is an orthonormal matrix then the columns, $\vec{a}_i$, form a basis for $\mathbb{R}^N$.

Answer:

We want to show that the columns of $\mathbf{A}$ form a basis for $\mathbb{R}^N$. To show that the columns form a basis for $\mathbb{R}^N$ we need to show two things:

- The columns must form a set of N linearly independent vectors.
- Any vector $\vec{x} \in \mathbb{R}^N$ can be represented as a linear combination of the vectors in the set.

We already know we have N vectors, so first we will show they are linearly independent. We shall do this by showing that $\mathbf{A}\vec{\beta} = \vec{0}$ implies that $\vec{\beta}$ can be only $\vec{0}$.

$$\mathbf{A}\vec{\beta} = \vec{0} \tag{1}$$

$$\beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N = \vec{0} \tag{2}$$

Then to exploit the properties of orthogonal vectors, we consider taking the inner product of each side of the above equation with $\vec{a}_i$.

$$\langle \vec{a}_i, \beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N \rangle = \langle \vec{a}_i, \vec{0} \rangle = 0 \tag{3}$$

Now we apply the distributive property of the inner product and the definition of orthonormal vectors,

$$\langle \vec{a}_i, \beta_1\vec{a}_1 \rangle + \ldots + \langle \vec{a}_i, \beta_i\vec{a}_i \rangle + \ldots + \langle \vec{a}_i, \beta_N\vec{a}_N \rangle = 0 \tag{4}$$

$$0 + \ldots + \beta_i\langle \vec{a}_i, \vec{a}_i \rangle + \ldots + 0 = 0 \tag{5}$$

$$0 + \ldots + \beta_i\vec{a}_i^T\vec{a}_i + \ldots + 0 = 0 \tag{6}$$

Because $\vec{a}_i^T\vec{a}_i = 1$, $\beta_i = 0$ for the equation to hold. Then, since this is true for all i from 1 to N, all the elements of the vector beta must be zero ($\vec{\beta} = \vec{0}$). Because $\vec{x} = \vec{0}$ implies $\vec{\beta} = \vec{0}$, the columns of $\mathbf{A}$ are linearly independent.
Now, we will show that any vector $\vec{x} \in \mathbb{R}^N$ can be represented as a linear combination of the columns of $\mathbf{A}$.

$$\vec{x} = \mathbf{A}\vec{\beta} = \beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N \tag{7}$$

Because we know that the N columns of $\mathbf{A}$ are linearly independent, then there exists $\mathbf{A}^{-1}$. Applying the inverse to the equation above,

$$\mathbf{A}^{-1}\mathbf{A}\vec{\beta} = \mathbf{A}^{-1}\vec{x} \tag{8}$$

$$\vec{\beta} = \mathbf{A}^{-1}\vec{x}, \tag{9}$$

we find that there exists a unqiue β that allow us to represent any $\vec{x}$ as a linear combination of the columns of $\mathbf{A}$.

(c) When $\mathbf{A} \in \mathbb{R}^{N\times M}$ and $N \geq M$ (i.e. tall matrices), show that if the matrix is orthonormal, then $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{M\times M}$.
Answer: Want to show $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{M\times M}$.

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \vec{a}_1^T\vec{a}_1 & \vec{a}_2^T\vec{a}_1 & \ldots & \vec{a}_n^T\vec{a}_1 \\ \vec{a}_2^T\vec{a}_1 & \vec{a}_2^T\vec{a}_2 & \ldots & \vec{a}_n^T\vec{a}_2 \\ \vdots & \vdots & & \vdots \end{bmatrix} = \mathbf{I}_{M\times M} \tag{10}$$

When $\vec{a}_i^T\vec{a}_i = \|\vec{a}_i\|^2 = 1$ and when $i \neq j$, $\vec{a}_i^T\vec{a}_j = 0$ because the eigenvectors are orthogonal.

(d) Again, suppose $\mathbf{A} \in \mathbb{R}^{N \times M}$ where $N \geq M$ is an orthonormal matrix. Show that the projection of $\vec{y}$ onto the subspace spanned by the columns of $\mathbf{A}$ is now $\mathbf{A}\mathbf{A}^T\vec{y}$.

Answer:

Starting with the result from part a,

$$\mathbf{A}\vec{\hat{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}, \tag{11}$$

we can apply the result from part c,

$$\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\vec{y} \tag{12}$$

$$= \mathbf{A}\mathbf{I}\mathbf{A}^T\vec{y} \tag{13}$$

$$= \mathbf{A}\mathbf{A}^T\vec{y} \tag{14}$$

