EECS 16A Designing Information Devices and Systems I
Fall 2018 Discussion 14A

1. 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,

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

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

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

2. Orthogonal Matching Pursuit

Let's work through an example of the OMP algorithm. Suppose that we have a vector $\vec{x} \in \mathbb{R}^4$ that is sparse and we know that it has only 2 non-zero entries. In particular,

$$\mathbf{M}\vec{x} \approx \vec{y} \tag{15}$$

$$\begin{bmatrix} | & | & | & | \\ \vec{m}_1 & \vec{m}_2 & \vec{m}_3 & \vec{m}_4 \\ | & | & | & | \end{bmatrix} \vec{x} \approx \vec{y} \tag{16}$$

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \approx \begin{bmatrix} 5 \\ 1 \\ 1 \end{bmatrix} \tag{17}$$

where exactly 2 of x_1 to x_4 are non-zero. Use Orthogonal Matching Pursuit to estimate x_1 to x_4.

(a) Why can we not solve for $\vec{x}$ directly?

Answer:

We cannot solve for $\vec{x}$ directly because we have three measurements (or equations) but four unknowns. Since our system is underdetermined, we cannot solve for the unique $\vec{x}$ directly.

(b) Why can we not apply the least squares process to obtain $\vec{x}$?

Answer:

Recall the least squares solution: $\vec{\hat{x}} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\vec{y}$. $\mathbf{M}^T\mathbf{M}$ is only invertible if it has a trivial null space, i.e., if $\mathbf{M}$ has a trivial null space. However, in this case, $\mathbf{M}$ is a 3×4 matrix, so there is at least one free variable, which means that its null space is non-trivial. Therefore, $\mathbf{M}^T\mathbf{M}$ is not invertible, and we cannot use least squares to solve for $\vec{x}$.

(c) Let us start by reviewing the OMP procedure,

Inputs:

- A matrix $\mathbf{M}$, whose columns, $\vec{m}_i$, make up a set of vectors, $\{\vec{m}_i\}$, each of length n
- A vector $\vec{y}$ of length n
- The sparsity level k of the signal

Outputs:

- A vector $\vec{x}$, that contains k non-zero entries.
- A error vector $\vec{e} = \vec{y} - \mathbf{M}\vec{x}$

Procedure:

- Initialize the following values: $\vec{e} = \vec{y}$, $j = 1$, k, $\mathbf{A} = \begin{bmatrix} \ \end{bmatrix}$
- while ($j \le k$):
  i. Compute the inner product for each vector in the set, $\vec{m}_i$, with $\vec{e}$: $c_i = \langle \vec{m}_i, \vec{e} \rangle$.
  ii. Column concatenate matrix $\mathbf{A}$ with the column vector that had the maximum inner product value with $\vec{e}$, c_i: $\mathbf{A} = \begin{bmatrix} \mathbf{A} & | & \vec{m}_i \end{bmatrix}$
  iii. Use least squares to compute $\vec{x}$ given the $\mathbf{A}$ for this iteration: $\vec{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$
  iv. Update the error vector: $\vec{e} = \vec{y} - \mathbf{A}\vec{x}$
  v. Update the counter: $j = j + 1$

(d) Compute the inner product of every column with the $\vec{y}$ vector. Which column has the largest inner product? This will be the first column of the matrix $\mathbf{A}$.

Answer:

$$\left\langle \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}5\\1\\1\end{bmatrix} \right\rangle = 5$$

$$\left\langle \begin{bmatrix}0\\1\\2\end{bmatrix}, \begin{bmatrix}5\\1\\1\end{bmatrix} \right\rangle = 3$$

$$\left\langle \begin{bmatrix}2\\2\\0\end{bmatrix}, \begin{bmatrix}5\\1\\1\end{bmatrix} \right\rangle = 12$$

$$\left\langle \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}5\\1\\1\end{bmatrix} \right\rangle = 6$$

The third column has the largest inner product with $\begin{bmatrix}5\\1\\1\end{bmatrix}$, so $\mathbf{A} = \begin{bmatrix}2\\2\\0\end{bmatrix}$.

(e) Now, find the projection of $\vec{y}$ onto the columns of $\mathbf{A}$ (ie. $\text{proj}_{\text{Col}(\mathbf{A})}\vec{y} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$). Use this to update the error vector.

Answer:

$$\vec{\hat{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y} = \left(\begin{bmatrix}2 & 2 & 0\end{bmatrix} \begin{bmatrix}2\\2\\0\end{bmatrix} \right)^{-1} \begin{bmatrix}2 & 2 & 0\end{bmatrix} \begin{bmatrix}5\\1\\1\end{bmatrix} = \frac{1}{8} \cdot 12 = \frac{3}{2}$$

$$\text{proj}_{\text{Col}(\mathbf{A})}\vec{y} = \mathbf{A}\vec{\hat{x}} = \begin{bmatrix}2\\2\\0\end{bmatrix} \cdot \frac{3}{2} = \begin{bmatrix}3\\3\\0\end{bmatrix}$$

$$\vec{e} = \vec{y} - \text{proj}_{\text{Col}(\mathbf{A})}\vec{y} = \begin{bmatrix}5\\1\\1\end{bmatrix} - \begin{bmatrix}3\\3\\0\end{bmatrix} = \begin{bmatrix}2\\-2\\1\end{bmatrix}$$

(f) Now compute the inner product of every column with the new error vector. Which column has the largest inner product? This will be the second column of $\mathbf{A}$.

Answer:

$$\left\langle \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}2\\-2\\1\end{bmatrix} \right\rangle = 2$$

$$\left\langle \begin{bmatrix}0\\1\\2\end{bmatrix}, \begin{bmatrix}2\\-2\\1\end{bmatrix} \right\rangle = 0$$

$$\left\langle \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} \right\rangle = 0$$

$$\left\langle \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} \right\rangle = 3$$

The fourth column has the largest inner product with $\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$, so $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}$.

(g) We now have two non-zero entries for our vector, $\vec{x}$. Find the values of those two entries.

(Reminder: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$)

Answer:

$$\vec{\hat{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y} = \left(\begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 2 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Therefore, $x_3 = 1$ and $x_4 = 2$, so $\vec{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix}$.