EECS 16B Designing Information Devices and Systems II
Fall 2021 Discussion Worksheet Discussion 14B

The following notes are useful for this discussion: Note 2j.

1. Gram Schmidt on Complex Vectors

(a) Consider the three complex vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ \mathrm{j} \\ 0 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \qquad \vec{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{1}$$

Compute an orthonormal basis from this list of vectors with Gram Schmidt.

Solution: We know that in the Gram Schmidt process, we remove projections of a vector from the original vector to ensure that we get a residual that is orthogonal to the subspace that we are projecting onto. We then normalize the residual to get the next basis element for our larger, expanded subspace. We know how to do projections and norm calculations for complex vectors. Specifically, the projection operation is $P_{\vec{u}} = \frac{\vec{u}\vec{u}^*}{\vec{u}^*\vec{u}} = \frac{\vec{u}\vec{u}^*}{\|u\|^2} = \vec{u}\vec{u}^*$, since we are only going to project onto unit vectors (with norm 1 in the denominator). Notationally, $\vec{u}_i$ was what we called $\vec{q}_i$, and $\vec{r}_i$ was $\vec{z}_i$.
Recall that $\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}$ since we only have 1 vector at the start. We symbolically calculate subsequent vectors as follows:

$$\vec{r}_2 = \vec{v}_2 - P_{\vec{u}_1}\vec{v}_2 \tag{2}$$

$$\vec{u}_2 = \frac{\vec{r}_2}{\|\vec{r}_2\|} \tag{3}$$

$$\vec{r}_3 = \vec{v}_3 - P_{\vec{u}_2}\vec{v}_3 - P_{\vec{u}_1}\vec{v}_3 \tag{4}$$

$$\vec{u}_3 = \frac{\vec{r}_3}{\|\vec{r}_3\|} \tag{5}$$

Notice that symbolically, this is exactyl equivalent to the real case we performed in discussion. Now, let's calculate specific numbers.

First, we calculate the first item in our orthonormal basis $\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{\mathrm{j}}{\sqrt{2}} \\ 0 \end{bmatrix}$.

Now, we will look for the second orthogonal vector $\vec{u}_2$. We know we will be projecting onto the first vector $\vec{u}_1$, so we need

$$P_{\vec{u}_1} = \vec{u}_1\vec{u}_1^* = \begin{bmatrix} \vec{u}[1] \\ \vec{u}[2] \\ \vec{u}[3] \end{bmatrix} \begin{bmatrix} \overline{\vec{u}}[1] & \overline{\vec{u}}[2] & \overline{\vec{u}}[3] \end{bmatrix} = \begin{bmatrix} \vec{u}[1]\overline{\vec{u}}[1] & \vec{u}[1]\overline{\vec{u}}[2] & \vec{u}[1]\overline{\vec{u}}[3] \\ \vec{u}[2]\overline{\vec{u}}[1] & \vec{u}[2]\overline{\vec{u}}[2] & \vec{u}[2]\overline{\vec{u}}[3] \\ \vec{u}[3]\overline{\vec{u}}[1] & \vec{u}[3]\overline{\vec{u}}[2] & \vec{u}[3]\overline{\vec{u}}[3] \end{bmatrix} \tag{6}$$

$$P_{\vec{u}_1} = \begin{bmatrix} \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}}0 \\ \frac{j}{\sqrt{2}}\frac{1}{\sqrt{2}} & \frac{j}{\sqrt{2}}\frac{-j}{\sqrt{2}} & \frac{j}{\sqrt{2}}0 \\ 0\frac{1}{\sqrt{2}} & 0\frac{-j}{\sqrt{2}} & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{-j}{2} & 0 \\ \frac{j}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{7}$$

We use this to compute

$$\vec{r}_2 = \vec{v}_2 - P_{\vec{u}_1}\vec{v}_2 \tag{8}$$
$$= \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{-j}{2} & 0 \\ \frac{j}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \tag{9}$$
$$= \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{j}{2} \\ \frac{-1}{2} \\ 0 \end{bmatrix} \tag{10}$$
$$= \begin{bmatrix} \frac{-j}{2} \\ \frac{-1}{2} \\ 0 \end{bmatrix} \tag{11}$$

Now for $\vec{u}_2$, we normalize this residual to find $\vec{u}_2 = \begin{bmatrix} \frac{-j}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 0 \end{bmatrix}$. We need to now compute $P_{\vec{u}_2}$,

$$P_{\vec{u}_2} = \begin{bmatrix} \frac{-j}{\sqrt{2}} \cdot \frac{j}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \cdot \frac{-1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \cdot 0 \\ \frac{-1}{\sqrt{2}} \cdot \frac{j}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \cdot \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \cdot 0 \\ 0 \cdot \frac{j}{\sqrt{2}} & 0 \cdot \frac{-1}{\sqrt{2}} & 0 \cdot 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{j}{2} & 0 \\ \frac{-j}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{12}$$

Now, to compute $\vec{r}_3$:

$$\vec{r}_3 = \vec{v}_3 - P_{\vec{u}_1}\vec{v}_3 - P_{\vec{u}_2}\vec{v}_3 \tag{13}$$
$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{-j}{2} & 0 \\ \frac{j}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{j}{2} & 0 \\ \frac{-j}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{14}$$
$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1-j \\ j+1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1+j \\ 1-j \\ 0 \end{bmatrix} \tag{15}$$
$$\vec{r}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{16}$$

This is already normalized, so we finish the orthonormal set with $\vec{u}_3 = \vec{r}_3$.

For finality, the set is:

$$\vec{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\ j\\ 0\end{bmatrix} \qquad \vec{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-j\\ -1\\ 0\end{bmatrix} \qquad \vec{u}_3 = \begin{bmatrix}0\\ 0\\ 1\end{bmatrix}. \tag{17}$$

(b) **Derive the least-squares solution for the case of a complex tall matrix of data and a tall matrix of values.** We want to find the best (complex) linear combination of the columns for predicting the observed values in a least-squares sense — we want to minimize the norm of the residual.

This can be formulated as having a feature matrix of data $D \in \mathbb{C}^{m\times n}$ where $m > n$ and measurements $\vec{y} \in \mathbb{C}^m$. In this case, feel free to assume that the columns of D are linearly independent even when we allow complex linear combinations. **First assume that the columns of D are orthonormal.** *Hint: You may find it useful to define a matrix $U = \begin{bmatrix}D & D_+\end{bmatrix}$, obtained via Gram-Schmidt.* Recall that the procedure here resembles that used in the SVD derivation (one portion of U contains the data/info we "care about", and the remainder is there via extension, to span the space.)

Solution: Note that when we want to solve $D\vec{p} = \vec{y}$, and the columns of D are orthonormal, then we can lean on the reasoning used in the real case; there, the solution is $\widehat{\vec{p}} = D^\top\vec{y}$. The heuristic principle is that to solve problems involving complex vectors, we should replace transposes with conjugate transposes. That would suggest that $D^*\vec{y}$ is the answer we're looking for. This isn't actually a proof yet, but having a direction to aim towards is quite useful.

Recall the least-squares problem is trying to find a solution vector $\vec{x}$ to solve the solution of $D\vec{x} = y$.

$$\arg\min_{\vec{x}} \|D\vec{x} - \vec{y}\|^2 = \arg\min_{\vec{x}} (D\vec{x} - \vec{y})^*(D\vec{x} - \vec{y}) \tag{18}$$

So, what can we do to proceed? Well, we know we can change variables by multiplying by any orthonormal square matrix U since $\|U\vec{x}\|^2 = \|\vec{x}\|^2$ as shown in a previous discussion. Now D is a tall orthonormal matrix. This means that we can extend it to a square orthonormal matrix by using Gram-Schmidt, just as in the real case (by augmenting D with the identity and then running Gram-Schmidt, throwing away any $\vec{0}$ vectors. The span of the resulting collection has to be the span of the identity and so we will get a full n set of vectors). Call the resulting orthonormal square matrix $U = \begin{bmatrix}D & D_+\end{bmatrix}$. If U is orthonormal, so is U^*.

Then, we can multiply through by U^* to get:

$$\begin{aligned}
\arg\min_{\vec{x}} \|D\vec{x} - \vec{y}\|^2 &= \arg\min_{\vec{x}} \left\|U^*D\vec{x} - U^*\vec{y}\right\|^2 && (19)\\
&= \arg\min_{\vec{x}} \left\|\begin{bmatrix}D^*\\ D_+^*\end{bmatrix} D\vec{x} - \begin{bmatrix}D^*\\ D_+^*\end{bmatrix}\vec{y}\right\|^2 && (20)\\
&= \arg\min_{\vec{x}} \left\|\begin{bmatrix}D^*D\\ D_+^*D\end{bmatrix}\vec{x} - \begin{bmatrix}D^*\vec{y}\\ D_+^*\vec{y}\end{bmatrix}\right\|^2 && (21)\\
&= \arg\min_{\vec{x}} \left\|\begin{bmatrix}\vec{x}\\ \vec{0}\end{bmatrix} - \begin{bmatrix}D^*\vec{y}\\ D_+^*\vec{y}\end{bmatrix}\right\|^2 && (22)\\
&= \arg\min_{\vec{x}} \left(\left\|\vec{x} - D^*\vec{y}\right\|^2 + \left\|-D_+^*\vec{y}\right\|^2\right) && (23)
\end{aligned}$$

$$= \arg\min_{\vec{x}} (\|\vec{x} - D^*\vec{y}\|^2) \quad (24)$$

$$= D^*\vec{y}. \quad (25)$$

Here, we do block-matrix manipulations along the way, expand out the norm square of a long vector as the sum of the norm squares of its two halves, and then realize that the second half is contributing a constant term that doesn't have any dependence on $\vec{x}$ (and so is irrelevant for the argmin). We can't get a norm smaller than 0.

(c) **Repeat the previous part without the assumption of orthonormality for the columns of D.** You can keep the assumption of linear independence.

Solution: Again, we are trying to solve

$$\arg\min_{\vec{x}} \|\vec{y} - D\vec{x}\|^2 \quad (26)$$

Where D is a tall matrix, with linearly-independent columns.
We know that in the real case, the answer is going to be $(D^\top D)^{-1}D^\top\vec{y}$. We derived this in 16A. By the heuristic of replacing transposes by conjugate-transposes, we know that the answer we want is $(D^*D)^{-1}D^*\vec{y}$. But how do we show this?
There are two approaches we can use. The first approach, which is more aligned to this discussion, is to argue that we can orthonormalize D using Gram-Schmidt, solve the system exactly as above, and then convert the coordinates back to verify the solution we obtained. This approach is clean, and leans on the 16B theme of reducing problems to what has been derived and seen before.
What's the second approach? Well, we could redo the proof/derivation that was done in 16A. Recall that we called $\hat{y} = D\hat{\vec{x}}$ the projection of $\vec{y}$ onto the subspace spanned by the columns of D and observed that $\vec{r} = \vec{y} - \hat{y}$ is the residual that is left. For the residual's length to be minimal, by the Pythagorean theorem, the residual vector must be orthogonal to the entire subspace spanned by the columns of D. Otherwise, we could choose new $\vec{x}$ that would only remove any component of $\vec{r}$ that was in the subspace spanned by the columns of D and find ourselves with a strictly smaller residual. (Notice that this was also the spirit of what was proved in the last part.)
Consequently, by the definition of orthogonality for complex vectors, $D^*\vec{r} = \vec{0}$ and so $D^*\vec{y} - D^*D\vec{x} = \vec{0}$. This means that the optimal $\vec{x}$ must satisfy $D^*D\vec{x} = D^*\vec{y}$.
At this point, we are almost done. We must re-consider the matrix D^*D, is it invertible? With our assumption of linear independence, the answer is yes. Here is a quick proof as to why (there are multiple correct proofs). If it is not invertible, it is because it must have a nullspace. If it has a nullspace, then there exists a nonzero $\vec{s}$ so that $D^*D\vec{s} = \vec{0}$, which implies $0 = \vec{s}^*D^*D\vec{s} = (D\vec{s})^*(D\vec{s}) = \|D\vec{s}\|^2$. So $D\vec{s} = \vec{0}$ which is impossible since D has linearly independent columns.
Invertibility gives us the solution:

$$\vec{x} = \left(D^*D\right)^{-1}D^*\vec{y}. \quad (27)$$

Notice that along the way, what we ended up having to do was just replicating arguments that we had already had to make in the real case for the complex case. The heuristic of replacing transposes with conjugate transposes works in guiding us in doing this. The homework is there to make sure that you internalize how to do this.

2. Q&A time! [$\approx$ 20 minutes]

This time is here for you all to ask any questions from discussion 14A and 14B to the TAs to review the material on complex vectors. If there are no further questions, then feel free to discuss anything else related to the course content, as this is the last non-review discussion.

Contributors:

- Neelesh Ramachandran.
- Nathan Lambert.
- Anant Sahai.

