EECS 16B Designing Information Devices and Systems II
Spring 2021 Discussion Worksheet — Discussion 11A

In this discussion, we discuss orthonormal transformations in the context of setting up a least squares problem, and in the end, we make a connection to the SVD.

1. Orthonormality, Least Squares, and Intro to SVD

(a) Let U be an $m \times n$ matrix with orthonormal columns, with $m \geq n$. Compute $U^\top U$. How does this change if $m < n$?

Answer: In the case when U is square ($m = n$), each of these terms will give I. The diagonal entries are the only ones that will be 1, while all other (off-diagonal) entries will be 0 (as seen in dis10A, Q1c.).

We can compute $U^\top U$ for when U is "tall" (more rows than columns, $m > n$):

$$U^\top U = \underbrace{\begin{bmatrix} — & \vec{u}_1^\top & — \\ — & \vec{u}_2^\top & — \\ & \vdots & \\ — & \vec{u}_n^\top & — \end{bmatrix}}_{n \times m} \underbrace{\begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \\ | & | & & | \end{bmatrix}}_{m \times n} \tag{1}$$

$$= \underbrace{\begin{bmatrix} \vec{u}_1^\top \vec{u}_1 & \vec{u}_1^\top \vec{u}_2 & \cdots & \vec{u}_1^\top \vec{u}_n \\ \vec{u}_2^\top \vec{u}_1 & \vec{u}_2^\top \vec{u}_2 & \cdots & \vec{u}_2^\top \vec{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{u}_n^\top \vec{u}_1 & \vec{u}_n^\top \vec{u}_2 & \cdots & \vec{u}_n^\top \vec{u}_n \end{bmatrix}}_{n \times n} \tag{2}$$

$$= I_{n \times n} \tag{3}$$

As expected, the final dimension of $U^\top U$ is $(n \times m) \times (m \times n) = n \times n$ (small square). It is also the identity matrix.

When $m < n$, it is impossible for the columns to all be orthogonal! This is because each column vector is of dimension m, and taking the first m of them will already span $\mathbb{R}^m$ (assuming each vector is orthogonal to the previous ones).

(b) Suppose you have a real, square, $n \times n$ orthonormal matrix U (the columns of U are unit norm and mutually orthogonal). You also have real vectors $\vec{x}_1, \vec{x}_2, \vec{y}_1, \vec{y}_2$ such that

$$\vec{y}_1 = U\vec{x}_1$$
$$\vec{y}_2 = U\vec{x}_2$$

Calculate $\langle \vec{y}_1, \vec{y}_2 \rangle = \vec{y}_2^\top \vec{y}_1 = \vec{y}_1^\top \vec{y}_2$ in terms of $\langle \vec{x}_1, \vec{x}_2 \rangle = \vec{x}_2^\top \vec{x}_1 = \vec{x}_1^\top \vec{x}_2$.

Answer: Since we have defined the y vectors, we can substitute their expressions into $\vec{y}_2^\top \vec{y}_1$:

$$\langle \vec{y}_1, \vec{y}_2 \rangle = \vec{y}_2^\top \vec{y}_1 \tag{4}$$

$$= (U\vec{x}_2)^\top U\vec{x}_2 \tag{5}$$
$$= \vec{x}_2^\top \underbrace{U^\top U}_{I_{n\times n}} \vec{x}_1 \tag{6}$$
$$= \vec{x}_2^\top \vec{x}_1 \tag{7}$$
$$= \langle \vec{x}_1, \vec{x}_2 \rangle \tag{8}$$

Note that in going from eq. (6) to eq. (7), we used the fact that for square orthonormal matrices, the transpose in the inverse (see dis10A, Q1.c).

So we've shown that under an orthonormal matrix transformation, the inner product is preserved!

(c) Following the previous question, express $\|\vec{y}_1\|_2^2$ and $\|\vec{y}_2\|_2^2$ in terms of $\|\vec{x}_1\|_2^2$ and $\|\vec{x}_2\|_2^2$.

Answer: The 2-norm[1] of a vector is the same as the square root of the inner product of that vector with itself. That is, $\|\vec{v}\|_2 = \sqrt{\langle v, v\rangle} = \sqrt{v^\top v}$. So, we can write that:

$$\|\vec{y}_1\|_2^2 = \langle \vec{y}_1, \vec{y}_1 \rangle = \vec{y}_1^\top \vec{y}_1 = \vec{x}_1^\top U^\top U \vec{x}_1 = \vec{x}_1^\top \vec{x}_1 = \|\vec{x}_1\|_2^2$$
$$\|\vec{y}_2\|_2^2 = \langle \vec{y}_2, \vec{y}_2 \rangle = \vec{y}_2^\top \vec{y}_2 = \vec{x}_2^\top U^\top U \vec{x}_2 = \vec{x}_2^\top \vec{x}_2 = \|\vec{x}_2\|_2^2$$

We see here that vector norms are preserved under orthonormal transformations.

(d) Suppose you observe data coming from the model $y_i = \vec{a}^\top \vec{x}_i$, and you want to find the linear scale-parameters (each a_i). We are trying to learn the model $\vec{a}$. You have m data points $(\vec{x}_i, y_i)$, with each $\vec{x}_i \in \mathbb{R}^n$.

Note that $\vec{x}_i$ refers to the i-th vector, not the i-th element of a single vector. Each $\vec{x}_i$ is a different input vector that you take the inner product of with $\vec{a}$, giving a scalar y_i.

Set up a least squares formulation for estimating $\vec{a}$, and find the solution to the least squares problem.

Answer: Since $y = \vec{a}^\top \vec{x}$ means that $y = \vec{x}^\top \vec{a}$, we can stack the equations with the following definitions:

$$X \triangleq \begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_m^\top \end{bmatrix} \qquad \vec{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{9}$$

Then, we have $\vec{y} = X\vec{a}$. We want to solve for the vector $\vec{w}$ that minimizes the magnitude of the difference between $\vec{y}$ and $X\vec{w}$. Then, our estimate for $\vec{a}$, which we denote $\hat{\vec{a}}$, will be $\vec{w}$.

Using more formal notation, we write:

$$\hat{\vec{a}} = \underset{\vec{w}}{\operatorname{argmin}} \|\vec{y} - X\vec{w}\|_2^2 . \tag{10}$$

This has the form of a standard least-squares problem, which we know the solution to. So we can write that:

$$\hat{\vec{a}} = \left(X^\top X\right)^{-1} X^\top y. \tag{11}$$

(e) Now suppose V is an orthonormal square matrix, and rather than observing $\vec{a}^\top \vec{x}$ directly, we actually observe data points that result from our inputs being transformed by $V^\top$ as follows:

$$\tilde{\vec{x}} = V^\top \vec{x} \tag{12}$$

[1] We specify that it's a 2-norm because there's a general class of ℓ_p-norms but this 2-norm is the definition you're likely used to. $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots}$.

That is, our model acts on the modified input data $\tilde{\vec{x}}_i$, so the data points we collected are now $\left(\tilde{\vec{x}}, y\right)$. We must now consider the new model:

$$y = \tilde{\vec{a}}^\top \tilde{\vec{x}} \tag{13}$$

$$= \tilde{\vec{a}}^\top V^\top \vec{x} \tag{14}$$

Set up a least-squares formulation for $\hat{\tilde{\vec{a}}}$. How is $\hat{\tilde{\vec{a}}}$ related to $\hat{\vec{a}}$?

Answer: We first want to find how $\widetilde{X}$ relates to X and V as a result of the input transformation:

$$\widetilde{X} = \begin{bmatrix} \tilde{\vec{x}}_1^\top \\ \tilde{\vec{x}}_2^\top \\ \vdots \\ \tilde{\vec{x}}_m^\top \end{bmatrix} \tag{15}$$

$$= \begin{bmatrix} \left(V^\top \vec{x}_1\right)^\top \\ \left(V^\top \vec{x}_2\right)^\top \\ \vdots \\ \left(V^\top \vec{x}_m\right)^\top \end{bmatrix} \tag{16}$$

$$= \begin{bmatrix} \vec{x}_1^\top V \\ \vec{x}_2^\top V \\ \vdots \\ \vec{x}_m^\top V \end{bmatrix} \tag{17}$$

$$= \begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_m^\top \end{bmatrix} V \tag{18}$$

$$= XV \tag{19}$$

$$\tag{20}$$

Now, the least squares formulation is

$$\hat{\tilde{\vec{a}}} = \underset{\vec{w}}{\operatorname{argmin}} \left\| \vec{y} - \widetilde{X}\vec{w} \right\|_2^2. \tag{21}$$

The solution is then

$$\hat{\tilde{\vec{a}}} = \left(\widetilde{X}^\top \widetilde{X}\right)^{-1} \widetilde{X}^\top y \tag{22}$$

$$= \left(V^\top X^\top X V\right)^{-1} V^\top X^\top y \tag{23}$$

$$= V^{-1} \left(X^\top X\right)^{-1} \left(V^\top\right)^{-1} V^\top X^\top y \tag{24}$$

$$= V^{-1} \left(X^\top X\right)^{-1} V V^\top X^\top y \tag{25}$$

$$= V^\top \left(X^\top X\right)^{-1} X^\top y \tag{26}$$

$$= V^\top \hat{\vec{a}}. \tag{27}$$

Our least-squares solution is the same as it was before, but now there's a $V^\top$ in front! The transformation carried through the system linearly into our estimate.

(f) Now suppose that we have the matrix

$$\begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_m^\top \end{bmatrix} \triangleq X = U\Sigma V^\top. \tag{28}$$

where U is an $m \times m$ matrix, and V is an $n \times n$ matrix. Suppose that we have more data points than the dimension of our space (that is, $m > n$). Also, the transformation V in part e) is the same V in this full SVD representation. Set up a least squares formulation for estimating $\vec{\hat{a}}$ and find the solution to the least squares. Is there anything interesting going on?

Note: Don't worry about how we would find u, Σ, $V^\top$ for now; assume that X has the given form and that U and V are orthonormal.

Hint: Start by substituting the full SVD representation of X into the answer of the previous part.

Answer: We start from the previous part (eq. (26)), make substitutions, and simplify as much as we can. Note that since we're using the full SVD, each of the matrices U and V is orthonormal and square, so their tranposes are equal to their inverses.

$$\begin{aligned}
\vec{\hat{a}} &= V^\top \left(X^\top X\right)^{-1} X^\top y && (29)\\
&= V^\top \left(\left(U\Sigma V^\top\right)^\top U\Sigma V^\top\right)^{-1} \left(U\Sigma V^\top\right) y && (30)\\
&= V^\top \left(V\Sigma^\top \underbrace{U^\top U}_{I_{m\times m}} \Sigma V^\top\right)^{-1} V\Sigma^\top U^\top y && (31)\\
&= \underbrace{V^\top \left(V^\top\right)^{-1}}_{I_{n\times n}} \left(\Sigma^\top\Sigma\right)^{-1} \underbrace{V^{-1}V}_{I_{n\times n}} \Sigma^\top U^\top y && (32)\\
&= \left(\Sigma^\top\Sigma\right)^{-1}\Sigma^\top U^\top y && (33)\\
&= \left(\begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}\right)^{-1} \Sigma^\top U^\top y && (34)\\
&= \left(\begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}\right)^{-1} \Sigma^\top U^\top y. && (35)
\end{aligned}$$

$$= \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_n^2} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n & 0 & \cdots & 0 \end{bmatrix} U^\top y. \tag{36}$$

$$= \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_n} & 0 & \cdots & 0 \end{bmatrix} U^\top y. \tag{37}$$

The matrix inverse term $\left(\Sigma^\top \Sigma\right)^{-1}$ can be easily computed because it is a diagonal square matrix. We only need to invert each of the diagonal coordinates. Multiplying this with $\Sigma^\top$ adds the extra $\vec{0}$ columns.

The key takeaway is that even though our input underwent an orthonormal matrix transformation, the estimate that we get is independent of this transformation because the V matrix terms all cancelled in the process above. This was dependent on the transformation being the same as the V matrix of the full SVD decomposition.

Contributors:

- Neelesh Ramachandran.
- Kuan-Yun Lee.
- Anant Sahai.