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

Discussion 9B

The following notes are useful for this discussion: Note 14

1. Orthonormality and Least Squares

(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 $m < n$?**

(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 $\langle \vec{x}_1, \vec{x}_2 \rangle = \vec{x}_2^\top \vec{x}_1 = \vec{x}_1^\top \vec{x}_2$.

(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 $\|\vec{x}_2\|_2^2$.**

(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.

(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:

$$\tilde{\vec{x}} = V^\top \vec{x} \tag{1}$$

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

$$y_i = \tilde{\vec{a}}^\top \tilde{\vec{x}}_i \tag{2}$$

$$= \vec{\tilde{a}}^\top V^\top \vec{x}_i \tag{3}$$

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

(f) Now suppose that we have the matrix

$$\begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_m^\top \end{bmatrix} \triangleq X = U\Sigma V^\top. \tag{4}$$

where U is an $m \times m$ matrix, and V is an $n \times n$ matrix. Here, $\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n & 0 & \cdots & 0 \end{bmatrix}$.

Here we assume 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 factorized representation.

Set up a least squares formulation for estimating $\vec{\tilde{a}}$ 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 factorized representation of X into the answer of the previous part.

Contributors:

- Neelesh Ramachandran.
- Kuan-Yun Lee.
- Anant Sahai.
- Kumar Krishna Agrawal.

