EECS 16B Spring 2021 — Designing Information Devices and Systems II — UC Berkeley

# Homework 12

**This homework is due on Friday, April 16, 2021, at 11:00PM. Self-grades and HW Resubmission are due on Tuesday, April 20, 2021, at 11:00PM.**

1. **Reading Lecture Notes**

Staying up to date with lectures is an important part of the learning process in this course. Here are links to the notes that you need to read for this week: Note 13, Note 14

Consider a matrix $A \in \mathbb{R}^{m\times n}$ with $rank(A) = r$ and $r \leq m < n$ (wide matrix with full row rank). We will consider the representations of the full, compact, and outer product forms of the SVD of $A$:

$$A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix}$$

$$A = U_r \Sigma_r V_r^T$$

$$A = \sum_{i=1}^{r} \vec{u_i}\sigma_i\vec{v_i}^T$$

Where $\vec{u_i}$ and $\vec{v_i}$ are the $i^{th}$ column vectors of $U$ and $V$ respectively. **Explain what spaces the columns of $U_r, U_{m-r}, V_r,$ and $V_{n-r}$ span. What is the rank of the matrix $\vec{u_i}\sigma_i\vec{v_i}^T$?**

**Solution:** The columns of $U_r$ form a basis for the column space of $A$, the columns of $U_{m-r}$ form a basis for the left nullspace of $A$. The columns of $V_r$ form a basis for the row space of $A$, the columns of $V_{n-r}$ form a basis for the nullspace of $A$.

The rank of $\vec{u_i}\sigma_i\vec{v_i}^T$ is equal to 1. To show this, let $M = \vec{u_i}\sigma_i\vec{v_i}^T \in \mathbb{R}^{m\times n}$:

$$\begin{aligned} col(M) &= \{M\vec{x} \mid \vec{x} \in \mathbb{R}^n\} \\ &= \{\vec{u_i}\sigma_i\vec{v_i}^T\vec{x} \mid \vec{x} \in \mathbb{R}^n\} \\ &= \{\vec{u_i}\sigma_i\langle\vec{v_i}, \vec{x}\rangle \mid \vec{x} \in \mathbb{R}^n\} \\ &= span\{\vec{u_i}\} \end{aligned}$$

And since the span of the columns of $M$ only has one vector, the rank of $M$ must be one.

## 2. SVD

(a) Given the matrix

$$A = \frac{1}{\sqrt{50}}\begin{bmatrix}3\\4\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix} + \frac{3}{\sqrt{50}}\begin{bmatrix}-4\\3\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix},$$

**write out a singular value decomposition of matrix $A$ in the form $U\Sigma V^\top$.** Note the ordering of the singular values in $\Sigma$ should be from the largest to smallest.

*HINT: You don't have to compute any eigenvalues for this. Some useful observations are that*

$$\begin{bmatrix}3,4\end{bmatrix}\begin{bmatrix}-4\\3\end{bmatrix} = 0, \quad \begin{bmatrix}1,-1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = 0, \quad \left\|\begin{bmatrix}3\\4\end{bmatrix}\right\| = \left\|\begin{bmatrix}-4\\3\end{bmatrix}\right\| = 5, \quad \left\|\begin{bmatrix}1\\-1\end{bmatrix}\right\| = \left\|\begin{bmatrix}1\\1\end{bmatrix}\right\| = \sqrt{2}.$$

**Solution:**

Observe that

$$\begin{bmatrix}3\\4\end{bmatrix}\cdot\begin{bmatrix}-4\\3\end{bmatrix} = 0$$

and

$$\begin{bmatrix}1\\-1\end{bmatrix}\cdot\begin{bmatrix}1\\1\end{bmatrix} = 0.$$

The singular value decomposition can be written in the form

$$A = \sum_{i=1}^{2}\sigma_i\vec{u}_i\vec{v}_i^\top,$$

with unit orthonormal vectors $\{\vec{u}_i\}$ and $\{\vec{v}_i\}$. From the given observations, we can see that the vectors we were provided are orthogonal, so we can just normalize them to get the desired answer. Taking it step by step:

$$\begin{align}
A &= \frac{1}{\sqrt{50}}\begin{bmatrix}3\\4\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix} + \frac{3}{\sqrt{50}}\begin{bmatrix}-4\\3\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} \tag{1}\\
&= \frac{5}{\sqrt{50}}\begin{bmatrix}\frac{3}{5}\\\frac{4}{5}\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix} + \frac{3\cdot 5}{\sqrt{50}}\begin{bmatrix}-\frac{4}{5}\\\frac{3}{5}\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} \tag{2}\\
&= \frac{5\sqrt{2}}{\sqrt{50}}\begin{bmatrix}\frac{3}{5}\\\frac{4}{5}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix} + \frac{3\cdot 5\sqrt{2}}{\sqrt{50}}\begin{bmatrix}-\frac{4}{5}\\\frac{3}{5}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix} \tag{3}\\
&= 1\begin{bmatrix}\frac{3}{5}\\\frac{4}{5}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix} + 3\begin{bmatrix}-\frac{4}{5}\\\frac{3}{5}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}. \tag{4}
\end{align}$$

From this, we can derive

$$\vec{u_1} = \begin{bmatrix}-\frac{4}{5}\\\frac{3}{5}\end{bmatrix}, \vec{u_2} = \begin{bmatrix}\frac{3}{5}\\\frac{4}{5}\end{bmatrix}, \vec{v_1} = \begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}, \vec{v_2} = \begin{bmatrix}\frac{1}{\sqrt{2}}\\-\frac{1}{\sqrt{2}}\end{bmatrix}$$

and corresponding singular values $\sigma_1 = 3, \sigma_2 = 1$ because we need to order them by size in decreasing order. This gives the singular value decomposition

$$A = \begin{bmatrix} -\frac{4}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^\top$$

(b) Define the matrix

$$A = \begin{bmatrix} -1 & 4 \\ 1 & 4 \end{bmatrix}.$$

Find the SVD of $A$. Then, find the eigenvectors and eigenvalues of $A$. Is there a relationship between the eigenvalues or eigenvectors of $A$ with the SVD of $A$?

**Solution:** Since we have a square matrix, we will arbitrarily use $A^\top A$ for our SVD:

$$\begin{aligned} A^\top A &= \begin{bmatrix} -1 & 1 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} -1 & 4 \\ 1 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 32 \end{bmatrix} \end{aligned}$$

Next, we find the eigenvalues of the above matrix.

$$\det(A^\top A - \lambda I) = (2 - \lambda)(32 - \lambda) = 0$$

Hence, the eigenvalues are $\lambda_1 = 32$ and $\lambda_2 = 2$, and the singular values are $\sigma_1 = \sqrt{32} = 4\sqrt{2}$ and $\sigma_2 = \sqrt{2}$.

Next, we find the right singular vectors (i.e. the columns of V). Finding $\text{null}(A^\top A - \lambda_1 I)$ and $\text{null}(A^\top A - \lambda_2 I)$ will give us $\vec{v}_1$ and $\vec{v}_2$ respectively.

Hence, $\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (the eigenvectors are already normalized here).

Lastly, we find the right singular vectors (the columns of U)

$$\begin{aligned} \vec{u}_1 &= \frac{1}{\sigma_1} A v_1 \\ &= \frac{1}{4\sqrt{2}} \begin{bmatrix} -1 & 4 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \end{aligned}$$

Similarly, we get $\vec{u}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

So the full SVD representation of $A$ is given below

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{5}$$

Now that we have found the SVD of $A$, we will find the eigenvalues and eigenvectors of $A$. Let us start with the eigenvalues:

$$\det(A - \lambda I) = (-1-\lambda)(4-\lambda) - 4 = \lambda^2 - 3\lambda - 8 = 0$$

Using the quadratic formula, the eigenvalues are $\lambda_1 = \frac{3+\sqrt{41}}{2} \approx 4.7$ and $\lambda_2 = \frac{3-\sqrt{41}}{2} \approx -1.7$.

Since we already used $\vec{v}_1, \vec{v}_2$ for the SVD, let us denote the eigenvectors of $A$ as $\vec{r}_1, \vec{r}_2$.
Finding $\mathrm{null}(A - \lambda_1 I)$ and $\mathrm{null}(A - \lambda_2 I)$ will give us $\vec{r}_1$ and $\vec{r}_2$ respectively.
Hence, the normalized eigenvectors of $A$ are $\vec{r}_1 \approx \begin{bmatrix} -0.98 \\ 0.17 \end{bmatrix}$ and $\vec{r}_2 \approx \begin{bmatrix} -0.57 \\ -0.82 \end{bmatrix}$.

We notice that there is no relationship between the eigenvalues or eigenvectors of $A$ with the SVD of $A$.

**3. More SVD Proofs (PRACTICE)**

Let $A \in \mathbb{R}^{m\times n}$ be an arbitrary matrix. Let the SVD of $A = U\Sigma V^T$. Let $r = \text{rank}(A)$. Let the columns of $U$ be $\vec{u}_1, \dots, \vec{u}_m$, the columns of $V$ be $\vec{v}_1, \dots, \vec{v}_n$, and the singular values of $A$ be $\sigma_1, \dots, \sigma_r, 0, \dots 0$ since the remaining $\min\{m, n\} - r$ singular values are 0.

(a) **Show that $A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T$.**
*HINT: It might be helpful to use the outer product version of matrix multiplication, i.e. $XY = \sum_{i=1}^{k} \vec{x}_i \vec{y}_i^T$ where $\vec{x}_i$ are the columns of $X$ and $\vec{y}_i^T$ are the rows $Y$.*
**Solution:**

$$
\begin{aligned}
A &= U\Sigma V^T && (6)\\
&= \begin{bmatrix} \vec{u}_1 & \dots & \vec{u}_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r\times(n-r)} \\ & & \sigma_r & \\ \hline & 0_{(m-r)\times r} & & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix} && (7)\\
&= \begin{bmatrix} \sigma_1\vec{u}_1 & \dots & \sigma_r\vec{u}_r & \vec{0} & \dots & \vec{0} \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix} && (8)\\
&= \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T + \sum_{i=r+1}^{n} \vec{0}\vec{v}_i^T && (9)\\
&= \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T && (10)
\end{aligned}
$$

(b) **Prove that if $\vec{x}$ is a linear combination of $\vec{v}_{r+1}, \dots, \vec{v}_n$, then $A\vec{x} = 0$ so $\vec{x} \in Null(A)$.**

**Solution:** By definition of $\vec{x}$, we can write $\vec{x} = \alpha_{r+1}\vec{v}_{r+1} + \dots + \alpha_n\vec{v}_n$ where the $\alpha_i$ are some real coefficients for the linear combination. We then will consider $A\vec{x}$ when $A$ is written in outer product SVD form.

$$
\begin{aligned}
A\vec{x} &= \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T (\alpha_{r+1}\vec{v}_{r+1} + \dots + \alpha_n\vec{v}_n) && (11)\\
&= \sum_{i=1}^{r} \sigma_i \vec{u}_i (\alpha_{r+1}\vec{v}_i^T\vec{v}_{r+1} + \dots + \alpha_n\vec{v}_i^T\vec{v}_n) && (12)
\end{aligned}
$$

Since the vectors $\vec{v}_i$ are orthonormal, and since $i$ only goes from 1 to $r$, we know that $\vec{v}_i^T\vec{v}_{r+1} = \vec{v}_i^T\vec{v}_{r+2} = \dots = \vec{v}_i^T\vec{v}_n = 0$ for all $i$. Thus,

$$
A\vec{x} = \sum_{i=1}^{r} \sigma_i \vec{u}_i (\alpha_{r+1}\cdot 0 + \dots + \alpha_n \cdot 0) = 0 \qquad (13)
$$

Then by definition, $\vec{x} \in Null(A)$.

(c) **Now we prove the opposite of the previous part, so if $\vec{x} \in Null(A)$, then $\vec{x}$ is a linear combination of the vectors $\vec{v}_{r+1}, \dots, \vec{v}_n$.**

**Solution:** We will try to find the null space directly. Let $\vec{x} \in Null(A)$.

$$A\vec{x} = (\sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T)\vec{x} = \sum_{i=1}^{r} \sigma_i \vec{u}_i (\vec{v}_i^T \vec{x}) = 0$$

To have $A\vec{x} = 0$, since $\sigma_i \neq 0$ for $i = 1, \ldots r$ and $\vec{u}_i \neq \vec{0} \quad \forall i$, we will need $\vec{v}_i^T \vec{x} = 0$ for all $i = 1, \ldots r$. In other words, we want the space of vectors $\vec{x}$ that are orthogonal to $\vec{v}_1, \ldots, \vec{v}_r$. Since the matrix $V$ is orthonormal, this space of vectors is exactly spanned by $\vec{v}_{r+1}, \ldots, \vec{v}_n$. Thus, the null space is spanned by $\vec{v}_{r+1}, \ldots, \vec{v}_n$.

The last two parts together prove that $Null(A) = \text{span}\{\vec{v}_{r+1}, \ldots, \vec{v}_n\}$. Moreover, since the $\vec{v}_i$ are orthonormal (which implies linear independence), the vectors $\vec{v}_{r+1}, \ldots, \vec{v}_n$ form an orthonormal basis for $Null(A)$.

(d) **Show that $U_r = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_r \end{bmatrix}$ is a basis for the column space of $A$.**

**Solution:** One way to see this is to use the outer product form of $A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T$. Consider the vector $A\vec{x} = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T \vec{x} = \sum_{i=1}^{r} (\sigma_i \vec{v}_i^T \vec{x})\vec{u}_i$. Note that for any $\vec{x}$, $A\vec{x}$ is just a summation of some scalars $\sigma_i(\vec{v}_i^T \vec{x})$ times the vectors $\vec{u}_i$ from $i = 1, \ldots r$. Thus by definition, the column space of $A$ is exactly spanned by the vectors $\vec{u}_i$. Since the $\vec{u}_i$ are additionally orthonormal and thus linearly independent, we have that $U_r$ forms a basis for the column space of $A$.

**Alternative solution:** You can also use the approach from lecture to show this. We consider the vector $A\vec{w}$ where $\vec{w}$ is some vector in $\mathbb{R}^n$. Since we know that $V$ is a basis for $\mathbb{R}^n$, we can write $\vec{w} = \sum_{i=1}^{n} \alpha_i \vec{v}_i$ for some constants $\alpha_i$. Then,

$$A\vec{w} = A\sum_{i=1}^{n} \alpha_i \vec{v}_i = \sum_{i=1}^{n} \alpha_i A\vec{v}_i \tag{14}$$

From part (b), we know that the last $n - r$ vectors of $V$ will be in the null space of $A$. Thus,

$$A\vec{w} = \sum_{i=1}^{n} \alpha_i A\vec{v}_i = \sum_{i=1}^{r} \alpha_i A\vec{v}_i \tag{15}$$

Thus, for any $\vec{w}$, $A\vec{w}$ can be represented as a linear combination of $A\vec{v}_1, \ldots, A\vec{v}_r$ meaning those vectors span the column space. Now if we plug in the SVD $U\Sigma V^T$ in for $A$,

$$A\vec{v}_i = U\Sigma V^T \vec{v}_i = U\Sigma \begin{bmatrix} \vec{v}_1^T \vec{v}_i \\ \vdots \\ \vec{v}_r^T \vec{v}_i \end{bmatrix} \tag{16}$$

$$= U\Sigma \vec{e}_i = U(\sigma_i \vec{e}_i) = \sigma_i \vec{u}_i \tag{17}$$

where $\vec{e}_i$ is the $i$th standard basis vector, i.e. the $i$th column of a $n \times n$ identity matrix. Thus, $\sigma_1 \vec{u}_1, \ldots, \sigma_r \vec{u}_r$ span the column space of $A$. Since these vectors are orthogonal, they are also linearly independent. Then we can normalize these vectors to get an orthonormal basis $\vec{u}_1, \ldots, \vec{u}_r$ for the column space of $A$.

**4. Generalizing Least Squares using the Pseudoinverse**

This question will show you how to generalize the orthogonal projection from normal least squares to work for any set of vectors. On the way, we will show some powerful properties of the Moore-Penrose pseudoinverse.

(a) Suppose we have a matrix $A \in \mathbb{R}^{m \times n}$. Now consider a vector $\vec{y} \in \mathbb{R}^m$. We are interested in projecting $\vec{y}$ onto the column space of $A$, i.e. finding an $\vec{x}$ such that $||A\vec{x} - \vec{y}||^2$ is minimized. For that $\vec{x}$, $A\vec{x}$ is known as the orthogonal projection of $\vec{y}$ onto the column space of $A$ (this was actually how we derived $\vec{\hat{x}}$ in 16A). **If $m \geq n$ and $A$ has linearly independent columns, find $\vec{x}$ and then the projection $A\vec{x}$.**

*HINT: Remember what least squares does.*

**Solution:** We recall that least squares does exactly this projection since it tries to minimize the error between $A\vec{x}$ and $\vec{y}$. The least squares formula gives us $\vec{x} = (A^T A)^{-1} A^T \vec{y}$. However, remember that $\vec{x}$ is in $\mathbb{R}^m$ and is just the coefficients of the columns of $A$. So the actual projection of $\vec{y}$ onto the column space is $A\vec{x} = A(A^T A)^{-1} A^T \vec{y}$.

(b) The previous part only worked when restricting $A$ to have linearly independent columns so that $A^T A$ is invertible. Thus it does not work for a wide $A$ with more columns than rows. (Try to think why a wide matrix can't have linearly independent columns on your own.)

We want to remove these restrictions so now we consider an arbitrary $A$. In this case, $A$ can have a nontrivial nullspace, and so there can be infinite such solutions for $\vec{x}$ that minimize the objective. Thus, we want to impose an extra condition that out of all the minimizing vectors $\vec{x}$, we want to find the one with the minimum squared norm $||\vec{x}||^2$. This is called the **minimum-norm least squares problem**:

$$\underset{\vec{x}}{\operatorname{argmin}} ||\vec{x}||^2 \quad \text{s.t.} \quad ||A\vec{x} - \vec{y}||^2 = \min_{\vec{s}} ||A\vec{s} - \vec{y}||^2 \tag{18}$$

(Note that if $A$ doesn't have a nullspace, then the least squares problem has a unique solution, so it is automatically the minimum norm solution.)

To solve this problem, we will use the SVD and the Moore-Penrose pseudoinverse. Let the SVD of $A = U\Sigma V^T$. We define the Moore-Penrose pseudoinverse $A^\dagger = V\Sigma^\dagger U^T$ from the last homework where

$$\Sigma^\dagger = \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & 0_{r \times (m-r)} \\ & & \frac{1}{\sigma_r} & \\ \hline & 0_{(n-r) \times r} & & 0_{(n-r) \times (m-r)} \end{array}\right].$$

**First show that $||A\vec{x} - \vec{y}||^2 = ||\Sigma V^T \vec{x} - U^T \vec{y}||^2$.**

*HINT: Remember that orthonormal transformations don't affect norms of vectors (discussion 11A).*

**Solution:** We first replace $A$ with $U\Sigma V^T$. We then use the fact that multiplying by orthonormal matrices won't change the norm, which was proven in discussion 11A. Thus, we will multiply the inside vector by $U^T$:

$$\begin{aligned} ||A\vec{x} - \vec{y}||^2 &= ||U\Sigma V^T \vec{x} - \vec{y}||^2 & (19) \\ &= ||\Sigma V^T \vec{x} - U^T \vec{y}||^2 & (20) \end{aligned}$$

(c) We now let $\vec{z} = V^T\vec{x}$ and $\vec{w} = U^T\vec{y}$. Since $V^T$ is orthonormal, we know that $||\vec{x}||^2 = ||V^T\vec{x}||^2 = ||\vec{z}||^2$. Additionally since $V^T$ is invertible, we can convert between $\vec{x}$ and $\vec{z}$, which means we can equivalently optimize over $\vec{z}$ instead of $\vec{x}$ in (18). Combining this with the result from part (b), our minimum-norm least squares problem simplifies to

$$\underset{\vec{z}}{\operatorname{argmin}} ||\vec{z}||^2 \quad \text{s.t.} \quad ||\Sigma\vec{z} - \vec{w}||^2 = \min_{\vec{s}} ||\Sigma\vec{s} - \vec{w}||^2. \tag{21}$$

Using the fact that $\Sigma$ has a simple form with singular values on its diagonal, **find the minimum-norm least squares solution for (21). Write your answer in terms of $\vec{w}$ and $\Sigma^\dagger$ (not $\Sigma$).**

*HINT: Sbow that*

$$\Sigma\vec{z} - \vec{w} = \begin{bmatrix} \sigma_1 z_1 - w_1 \\ \vdots \\ \sigma_r z_r - w_r \\ -w_{r+1} \\ \vdots \\ -w_m \end{bmatrix}$$

*where $r = rank(A)$, and see which entries you can minimize by choosing $\vec{z}$.*

**Solution:** We first calculate

$$\Sigma\vec{z} - \vec{w} = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r\times(n-r)} \\ & & \sigma_r & \\ \hline & 0_{(m-r)\times r} & & 0_{(m-r)\times(n-r)} \end{array}\right] \vec{z} - \vec{w} \tag{22}$$

$$= \begin{bmatrix} \sigma_1 z_1 - w_1 \\ \vdots \\ \sigma_r z_r - w_r \\ -w_{r+1} \\ \vdots \\ -w_m \end{bmatrix} \tag{23}$$

as described in the hint. We can see that only the first $r$ elements of $\Sigma\vec{z} - \vec{w}$ contain any $z_i$, and thus only the first $r$ elements can be modified. Since we want to minimize the norm of this vector, we will want to first $r$ elements to all be 0 so the optimal choices will be $z_i = \frac{1}{\sigma_i} w_i$ for $i = 1, \dots, r$. Note that changing the last $n - r$ values of $\vec{z}$ won't actually affect the least squares objective. However since we still want to minimize the norm of $\vec{z}$, we will want to choose the last $n - r$ values of $\vec{z}$ to be 0. Putting this together gives us the minimum-norm least squares solution:

$$\vec{z} = \begin{bmatrix} \frac{1}{\sigma_1} w_1 \\ \vdots \\ \frac{1}{\sigma_r} w_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & 0_{r\times(m-r)} \\ & & \frac{1}{\sigma_r} & \\ \hline & 0_{(n-r)\times r} & & 0_{(n-r)\times(m-r)} \end{array}\right] \vec{w} = \Sigma^\dagger \vec{w} \tag{24}$$

(d) Since we found the $\vec{z}$ that solved (21), we can convert this solution back to $\vec{x}$ and it should solve (18). **Do this and show that the minimum-norm least squares solution for (18) is $\vec{x} = A^\dagger \vec{y}$.**

**Solution:** Substituting $\vec{x} = V\vec{z}$ and $\vec{w} = U^T\vec{y}$ gives us $\vec{x} = V\vec{z} = V\Sigma^\dagger\vec{w} = V\Sigma^\dagger U^T\vec{y} = A^\dagger\vec{y}$.

(e) Thus, we just proved that applying the pseudoinverse exactly gives the minimum-norm least squares solution which works for all $A$ matrices!

Now, we can finally go back to part (a) and address the restrictions on $A$. Using your new-found tool, **determine the projection of $\vec{y}$ onto the column space of $A$, where now $A$ can be any arbitrary matrix.**

**Solution:** We just proved that the pseudoinverse minimizes the squared error between $\vec{y}$ and $A\vec{x}$, which means it does an orthogonal projection. Then $\vec{x} = A^\dagger\vec{y}$ will be the coefficients for the orthogonal projection onto the column space of $A$. To get the actual projected vector, we need to multiply the coefficients by $A$ which gives us $A\vec{x} = AA^\dagger\vec{y}$.

(f) Let $r = \text{rank}(A)$. We will define $U_r = \begin{bmatrix}\vec{u}_1 & \dots & \vec{u}_r\end{bmatrix}$ which is the first $r$ columns of $U$. **Show that the expression for the projection from part (e) can be simplified to $U_rU_r^T\vec{y}$ by plugging in the SVD and pseudoinverse.**

*HINT: Write out both $\Sigma$ and $\Sigma^\dagger$ and multiply them out.*

**Solution:** Plugging in the SVD and pseudoinverse,

$$P = AA^\dagger = U\Sigma V^T V\Sigma^\dagger U^T = U\Sigma\Sigma^\dagger U^T \tag{25}$$

$$= U \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r\times(n-r)} \\ & & \sigma_r & \\ \hline & 0_{(m-r)\times r} & & 0_{(m-r)\times(n-r)} \end{array}\right] \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & 0_{r\times(m-r)} \\ & & \frac{1}{\sigma_r} & \\ \hline & 0_{(n-r)\times r} & & 0_{(n-r)\times(m-r)} \end{array}\right] U^T \tag{26}$$

$$= \begin{bmatrix}\vec{u}_1 & \dots & \vec{u}_m\end{bmatrix} \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0_{r\times(m-r)} \\ & & 1 & \\ \hline & 0_{(m-r)\times r} & & 0_{(m-r)\times(m-r)} \end{array}\right] \begin{bmatrix}\vec{u}_1^T \\ \vdots \\ \vec{u}_m^T\end{bmatrix} \tag{27}$$

$$= \begin{bmatrix}\vec{u}_1 & \dots & \vec{u}_r & \vec{0} & \dots & \vec{0}\end{bmatrix} \begin{bmatrix}\vec{u}_1^T \\ \vdots \\ \vec{u}_m^T\end{bmatrix} \tag{28}$$

$$= \sum_{i=1}^{r} \vec{u}_i\vec{u}_i^T = U_rU_r^T \tag{29}$$

In (25) we used the property that $V$ is an orthonormal matrix and so $V^TV = I$. In (29) we use the outer product form of matrix multiplication to simplify our result.

This simplification of the projected vector is very interesting, and there actually is an alternative reasoning to arrive at it. We know that $U_r$ is a basis for the column space of $A$, which was proved in lecture. Then projecting onto the column space of $A$ should be the same as projecting onto $U_r$. But now, note that $U_r$ has linearly independent columns, so we can just use part (a) replacing $A$ with $U_r$. Thus, the projected vector is $U_r(U_r^TU_r)^{-1}U_r^T\vec{y} = U_rI^{-1}U_r^T\vec{y} = U_rU_r^T\vec{y}$, where $U_r^TU_r = I$ since $U_r$ has orthonormal columns.

**5. Brain-machine interface and neuron sorting**

The iPython notebook pca_brain_machine_interface.ipynb will guide you through the process of analyzing brain machine interface data leveraging the SVD. This will help you to prepare for the project.

For SIXT33N (your robot car), you will need to use the SVD and PCA to classify your sound inputs so that your car does what you tell it to do.

Brain-Machine interfaces (BMIs) are a way for a brain to directly communicate to an external device. They're often used for research, mapping, assisting or repairing human cognitive or sensory-motor function.

Data was collected that shows waveform traces of (simulated) brain waves of a subject whose arm is pointing in certain directions over time. These waveform traces are gathered from electrodes inserted into the brain, but the electrodes are generally recording more than 1 neuron at the same time — the brain is crowded after all. To make predictions based on neuron firing rates, we need to be able to distinguish waveforms that come from different neurons.

Each neuron has its own waveform "signature" shape unique to that neuron. This is due to the physical characteristics of neurons, such as their physical shape and structure. It is impossible to know beforehand how many neurons an electrode measures or what each neuron's waveform looks like, so we must first separate the waveforms from the different neurons near the electrode.

The goal of this problem is to see how we can use the SVD (in its PCA manifestation) and clustering to decide which neuron fired. We will first look at a case where there are only two neurons, then a case where there are three neurons. The neurons have also been presorted using a professional software package, so we can check our model against presorted data.

Please complete the notebook by following the instructions given.

**Task 1: Two Neuron Waveform Sorting**

(a) The data you have are traces of length 32 timesteps. The data is arranged along the rows of the matrix, each row is one trace. Import the data sets and see the average waveform for each presorted neuron by running the corresponding cells in the .ipynb. **What is the shape of the training data matrix three_neurons_training?**

**Solution:** Part a) of the notebook pca_brain_machine_interface_sol.ipynb contains solutions to this exercise. Three_neurons_training.shape = (4398, 32)

(b) **What do the columns and rows of the training data matrix: three_neurons_training represent?**

**Solution:** Each row of the training data matrix corresponds to a captured waveform of a (possibly different) neuron. The columns of the training data set correspond to the times at which the neuron's waveform was sampled. Here, the times are all relative to the start" of each waveform.

In the real world, this data would have come from a segmentation (i.e. cutting into pieces) of a long recording of a trace from an electrode embedded in the brain. We have not taught you how to do such segmentation in this course. However, even from 16A, you might have a rough idea of how to do it. You could simply look for time-periods in which there is some activity and then cut those out. In the real-world, the problem of segmentation is done iteratively with understanding what the signals of interest look like. This is because we would ideally like to align all the waveforms to start at the right starting point. (This should remind you a bit of the positioning lab that you did in 16A.) For this knowing the waveforms is useful. These kinds of chicken-egg" problems are solved iteratively using the same philosophy that is done in k-means clustering (which you should have seen in 61A's yelp-like project.).

We can approximate each waveform in a lower dimensional space using PCA. In your implementation of PCA you will decide how to choose these components.

(c) You will be using the SVD in your PCA function. **What matrices does the SVD return? What are the dimensions of these matrices for the SVD of three_neurons_training?**

**Solution:** Our training data: three_neurons_training, was of shape: (4398, 32). When we take the SVD our three matrices $U$, $\Sigma$ and $V$ will be of shapes 4398 x 4398, 4398 x 32, 32 x 32.

(d) To represent each waveform in a lower dimensional space we want a basis for the time signals of the waveforms of the neurons. To construct this basis we start by taking the SVD (of three_neurons_training). **Which of the $U$ and $V$ matrices can we use to construct our desired basis?**

**Solution:** Each of our waveforms is a time series of length 32 and thus can be represented as a 32 x 1 vector. Since our data is along the rows of our data matrix, vectors in matrix $V$ corresponds to our basis.

(e) **Read through PCA_train and impliment PCA_project**. We will use these functions throughout the rest of the notebook, so make sure you understand them.

**Solution:** In .ipynb

(f) **Call PCA_train on two_neurons_training and plot the 2 principal components.** Note that since the dataset is randomized, you might get different plots every time you run the second code cell of this notebook (the _make_training_set function). Note that these components have no real 'physical' relationship to the actual shape of the neuron plots. They are a part of a basis, not the representation of exemplar traces.

**Solution:** In .ipynb

(g) Before we project onto our principal components, let us project our data onto 2 random 32 length vectors. **Run the corresponding part in the .ipynb file and comment on the separation of the 2 neuron's distinctive trace shapes in this projected basis.** Do you think that it would be easy to tell them apart just by looking at their projection into these two random dimensions?

**Solution:** We can see that since there is a larger overlap between the presorted points this random basis does not create a good separation.

(h) **Project two_neurons_test onto the two principal components found using the SVD and produce a 2D scatter plot in the new basis.** We will also try projecting the presorted data containing 2 neurons so we can see how the model behaves on the 2 neurons. **Comment on the difference between the projections here and part (g) where you projected onto random directions.**

**Solution:** In .ipynb. The projections created using PCA have more separation.

(i) Now we will repeat this process for three principal components. **Do you see a significant improvement with 3 principal components?**

**Solution:** In .ipynb. It is not necessary to use 3 principal components since the 3-D plots show us that there are similar clusters in comparison to 2-D.

**Task 2: Three Neuron Sorting**

(j) In the previous part we sorted 2 neurons using two or three principal components. In this part we are now dealing with 3 neurons. **Replicate what we did in the previous parts by finding and projecting our data on the first two principal components of this three neuron dataset in the .ipynb.**

**Solution:** In .ipynb

(k) Now **call PCA_train on three_neurons_training and plot the 3 principal components. Do the same classification process as the 2 neuron data, but now with the 3 neuron data. Compare your model's behavior with that of the presorted data.** The plot_3D function will be useful to view the results.

**Solution:** In .ipynb

(l) **How many principal components do you actually need to cluster the 3 neurons? (Hint: Think about whether there was an increase in separability between the last two parts)**

**Solution:** 2 - we can see that the clusters are relatively similar at each Z location, so we only need to worry about their location in the x-y plane.

**Task 3: Determining Neurons (OPTIONAL)** Now that we know how to project our data onto a basis where it is clearly separable we can classify our points and determine which neuron fired. (This is what we need to know if we want to use the firings of different neurons to build a BMI.)

To classify our points we use the following algorithm:

i. Find centroids: points that represent the average waveform of a particular neuron firing, in the basis of principal components.
ii. Classify each incoming point by assigning it the value of the centroid closest to it (i.e. declare that the neuron waveform that a point represents is the same as the neuron waveform of the centroid the point is closest to).

(m) **OPTIONAL** When we are doing such problems in practice, we will not have a presorted data set. In such cases we can use algorithms like $k$-means clustering to determine our centroids (You should have seen this algorithm in CS61A during the Maps project). As a refresher, $k$-means will do the same thing as in (l)i. but with a couple modifications. A step is added at the beginning to randomly assign each data point to some centroid and a step is added to the end such that if the classifications remained the same as last time, exit. Else, go back to step 2.

Scipy happens to have a built in function to compute $k$-means given clustered data, which we have conveniently formatted for you.

Run the code that we have provided to do k-means for you and find the centroids. **Then use 'which_neuron' on the two_classified data set and count the number of times each neuron fired. Are these values the same as the previous part ?**

**Solution:** In the .ipynb file

In later courses, you will build on the basic techniques taught here. But in reality, the ideas that you have learned in 16AB so far already form a conceptual core for machine learning. You can go very far with just the ideas you have already learned.

**6. Movie Ratings and PCA**

Recall from the lecture on PCA that we can think of movie ratings as a structured set of data. For every person $i$ and movie $j$, we have that person's rating $R_{i,j}$ (thought of as a real number).

Suppose that there are $m$ movies and $n$ people. Let's think about arranging this data into a big $n \times m$ matrix $R$ with people corresponding to rows and movies corresponding to columns. So the entry in the $i$-th row and $j$-th column should be $R_{i,j}$ above. Note that this is organized differently from how it was in lecture. Each row corresponds to a unique person and each column to a unique movie.

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nm} \end{bmatrix}$$

(a) Suppose we believe that there is actually an underlying pattern to this rating data and that a separate study has revealed that every movie is characterized by a set of characteristics: say action and comedy. This means that every movie $j$ has a pair of numbers $a[j], c[j]$ that define it. At the same time, every person $i$ has a pair of sensitivities $s_a[i], s_c[i]$ that essentially define that person's preferences. A person $i$ will rate the movie $j$ as $R_{i,j} = s_a[i]a[j] + s_c[i]c[j]$.

If we arrange the sensitivities into a pair of $n$-dimensional vectors $\vec{s_a}, \vec{s_c}$, and the movie characteristics into a pair of $m$-dimensional vectors $\vec{a}, \vec{c}$, **use outer products to express the rating matrix $R$ in terms of these vectors $\vec{s_a}, \vec{s_c}, \vec{a}, \vec{c}$.**

**Solution:**

Here $R = \vec{s_a}\vec{a}^\top + \vec{s_c}\vec{c}^\top$. The $\vec{s_a}$ has one entry for every person, and the people are arranged as rows So to get the sizes to work out, it has to be that we do it in this order. If it helps, imagine that there is exactly 1 movie. In that case, $R$ is a tall matrix that is only 1 wide. The only tall things we have are $\vec{s_a}$ and $\vec{s_b}$.

The outer product $\vec{s_a}\vec{a}^\top$ is therefore a matrix of the same size as $R$ that is filled with $s_a[i]a[j]$ in the $i$-th row and $j$-th column. Similarly $\vec{s_c}\vec{c}^\top$ is a matrix of the same size as $R$ that is filled with $s_c[i]c[j]$ in the $i$-th row and $j$-th column. Adding them together gives us a matrix whose $i$-th row and $j$-th column contains $R_{i,j} = s_a[i]a[j] + s_c[i]c[j]$ as desired.

(b) Now suppose that we want to discover the underlying nature of movies from the data $R$ itself.

Suppose for this part, that you have four observed rating data vectors (corresponding to four different movies being rated by six individuals).

All of the movie data vectors just happened to be multiples of the following 6-dimensional vector $\vec{w} = \begin{bmatrix} 2 \\ -2 \\ 3 \\ -4 \\ 4 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{w}\| = 7$.)

You arrange the movie data vectors as the columns of a matrix $R$ given by:

$$R = \begin{bmatrix} | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} \\ | & | & | & | \end{bmatrix} \tag{30}$$

You want to perform PCA (for movies) using the SVD of the matrix $R$ to better understand the pattern in your data.

The first "principal component vector" is a *unit vector* that tells which direction we would want to project the columns of $R$ onto to get the best rank-1 approximation for $R$.

**Find this first principal component vector of the columns of $R$ to explain the nature of your movie data points.**

*(HINT: You don't need to actually compute any SVDs in this simple case. Also, be sure to think about what size vector you want as the answer. Don't forget that you want a unit vector!)*

**Solution:**

Principal component analysis is in general about understanding how best to approximate our (potentially) high-dimensional data (like recordings from a microphone, or in this case, a movie's ratings by lots of people) with its lower-dimensional essence. The first principal component is about seeing which one-dimensional line best approximates the data points — i.e. which is the line for which projecting the data points onto it results in "estimates" that are as close as possible to the data points.

In the case of this problem, every point is explicitly given as a multiple of a single vector $\vec{w}$ and so the data already lies on such a straight line going through the origin. So, the first principal component is just along the direction of $\vec{w}$. Because a principal component represents a direction, it is conventional to normalize the vector to have unit length. In this case, we are told that the vector $\vec{w}$ has length 7, and so the answer is $\frac{\vec{w}}{7}$.

(Because the line is all that matters, you could also have used the negative of this $-\frac{\vec{w}}{7}$.)

We can also do this using the SVD.

The singular value decomposition of a matrix $R$ is a way of decomposing $R$ into a sum of rank 1 matrices. In this sum the $i^{th}$ rank 1 matrix is formed from taking the outer product of normalized column vectors $\vec{u}_i$ and normalized row vectors $\vec{w}_i^\top$, scaled by their respective singular values $\sigma_i$.

Looking at our given $R$, we can see that the matrix itself is rank 1 as the columns are all multiples of the same vector: $\vec{w}$. Seeing this we realize we can rewrite the matrix $R$ as the following outer product:

$$R = \begin{bmatrix} | \\ \vec{w} \\ | \end{bmatrix} \begin{bmatrix} -1 & -2 & 2 & 4 \end{bmatrix} \tag{31}$$

However the SVD requires we normalize the vectors $\vec{u_1}$ and $\vec{v_1}^\top$. In order to reconstruct $A$ properly we must scale back with the norms that we divided out to normalize.

$\left\|[-1,-2,2,4]\right\| = \sqrt{25} = 5$ and $\|\vec{w}\| = 7$. Consequently, when we pull that out, we get $\sigma_1 = 35$ as the singular value that corresponds to the first (and only) principal component.

Thus we can write the SVD of $R$ as:

$$R = \begin{bmatrix} | \\ \frac{\vec{w}}{7} \\ | \end{bmatrix} 35 \begin{bmatrix} \frac{-1}{5} & \frac{-2}{5} & \frac{2}{5} & \frac{4}{5} \end{bmatrix} \tag{32}$$

Now we just have to pick which normalized vector to deem the principal component. Since our data (the movie ratings) are collected as columns we choose $\frac{\vec{w}}{7}$ as the principal component.

**Alternate Solution** To find the first principal component along the columns, we can use $\Sigma$ and $U$. This is because our data is stored in the columns of $R$. We know that

$$R = U\Sigma V^T$$
$$\implies RR^T = U\Sigma\Sigma^T U^T$$

where $\Sigma\Sigma^T$ represents the square diagonal matrix with $min(m,n)$ singular values squared on the diagonal. Plugging in for R gives

$$RR^T = \begin{bmatrix} -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} \end{bmatrix} \begin{bmatrix} -\vec{w} \\ -2\vec{w} \\ 2\vec{w} \\ 4\vec{w} \end{bmatrix}$$
$$RR^T = \vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 16\vec{w}\vec{w}^T$$
$$RR^T = 25\vec{w}\vec{w}^T$$

By the spectral decomposition, we know that the eigenvectors of $RR^T$ correspond to the columns of $U$ and the eigenvalues of $RR^T$ correspond to the singular values squared. By inspection, we can see that the eigenvector is $\vec{w}$. Solving for the eigenvalue and singular value:

$$RR^T\vec{w} = 25\vec{w}\vec{w}^T\vec{w}$$
$$RR^T\vec{w} = (25\vec{w}^T\vec{w})\vec{w}$$
$$\implies \lambda = 25\vec{w}^T\vec{w} = 5^2 7^2$$
$$\implies \sigma = \sqrt{\lambda} = 35$$

For PCA, we require normalized vectors, so for that reason our first principal component is $\frac{\vec{w}}{\|\vec{w}\|} = \frac{\vec{w}}{7}$ with a corresponding singular value of $\sigma = 35$.

(c) Suppose that now, we have two more data points (corresponding to two more movies being rated by the same set of six people, i.e. we added two columns to our matrix) that are multiples of a different vector $\vec{p}$ where:

$\vec{p} = \begin{bmatrix} 6 \\ 3 \\ -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{p}\| = 7$ and that $\vec{p}^\top \vec{w} = 0$.)

We augment our ratings data matrix with these two new data points to get:

$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix} \tag{33}$$

Find the first two principal components corresponding to the nonzero singular values of $R$. This is what we would use to best project the movie data points onto a two-dimensional subspace.

**What is the first principal component vector? What is the second principal component vector? Justify your answer.**

*(Hint: Think about the inner product of $\vec{w}$ and $\vec{p}$ and what that implies for being able to appropriately decompose $R$. Again, very little computation is required here.)*

**Solution:** The solution to the previous part tells you what we need to do. We need to find the best two-dimensional subspace that best represents our data.

We start by taking the SVD of $R$.

The columns of $R$ are all multiples of two vectors: $\vec{w}$ and $\vec{p}$. Each of these can be used to create a rank 1 matrix, and these can be summed together to form $R$.

Since $\vec{w}$ and $\vec{p}$ are orthogonal to one another, our life is easier. This problem's R matrix is made especially nice by seeing that a data point is either purely in the $\vec{w}$ direction, or the $\vec{p}$ direction.

Using this knowledge we rewrite $R$ as:

$$R = \begin{bmatrix} | \\ \vec{w} \\ | \end{bmatrix} \begin{bmatrix} -1 & -2 & 2 & 4 & 0 & 0 \end{bmatrix} + \begin{bmatrix} | \\ \vec{p} \\ | \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & -3 & 3 \end{bmatrix}.$$

The orthogonality relationships demanded by the SVD are clearly satisfied since the row-vectors involved above have disjoint support (i.e. when one is nonzero, the other is zero) and the columns are orthogonal since we've been told so.

However for the SVD the vectors: $\vec{u_1}$, $\vec{v_1}^\top$, $\vec{u_2}$ and $\vec{v_2}^\top$ must be normalized and each rank 1 matrix must be scaled by the appropriate $\sigma_i$ to allow the sum to properly reconstruct $R$. We also need to figure out which $\sigma_i$ is bigger so we can order them properly. In the previous part, we have already done the calculations for $\vec{w}$'s part in this story. So what remains is the $\vec{p}$ part. Clearly the norm of the relevant row is $3\sqrt{2}$ which the norm of the relevant column is 7. So the singular value in question is $21\sqrt{2}$.

Using this we can rewrite $R$ as:

$$= \begin{bmatrix} | \\ \frac{\vec{w}}{7} \\ | \end{bmatrix} 35 \begin{bmatrix} \frac{-1}{5} & \frac{-2}{5} & \frac{2}{5} & \frac{4}{5} & 0 & 0 \end{bmatrix} + \begin{bmatrix} | \\ \frac{\vec{p}}{7} \\ | \end{bmatrix} 21\sqrt{2} \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{-3}{3\sqrt{2}} & \frac{3}{3\sqrt{2}} \end{bmatrix}.$$

So, what is bigger 35 or $21\sqrt{2}$? This is the same as asking which is bigger 5 or $3\sqrt{2}$. Squaring both tells us that this is the same as asking which is bigger, 25 or 18. Clearly, the first term is larger.

From this we see that our singular values are $\sigma_1 = 35$ and $\sigma_2 = 21\sqrt{2}$ since $35 > 21\sqrt{2}$. Thus $\frac{\vec{w}}{7}$ which corresponds to $\sigma_1$ is still the first principal component vector and $\frac{\vec{p}}{7}$ which corresponds to $\sigma_2$ is the second principal component vector.

**Alternate Solution** Again, to find the first principal component along the columns, we can use $\Sigma$ and $U$. This is because our data is stored in the columns of $R$. We know that

$$R = U\Sigma V^T$$
$$\implies RR^T = U\Sigma\Sigma^T U^T$$

where $\Sigma\Sigma^T$ represents the square diagonal matrix with $min(m, n)$ singular values squared on the diagonal. Plugging in for R gives

$$RR^T = \begin{bmatrix} -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \end{bmatrix} \begin{bmatrix} -\vec{w} \\ -2\vec{w} \\ 2\vec{w} \\ 4\vec{w} \\ -3\vec{p} \\ 3\vec{p} \end{bmatrix}$$

$$RR^T = \vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 16\vec{w}\vec{w}^T + 9\vec{p}\vec{p}^T + 9\vec{p}\vec{p}^T$$

$$RR^T = 25\vec{w}\vec{w}^T + 18\vec{p}\vec{p}^T$$

By the spectral decomposition, we know that the eigenvectors of $RR^T$ correspond to the columns of $U$ and the eigenvalues of $RR^T$ correspond to the singular values squared. By inspection, we can see that the first eigenvector is $\vec{w}$ and the second one is $\vec{p}$. Starting with the first:

$$RR^T\vec{w} = 25\vec{w}\vec{w}^T\vec{w} + 18\vec{p}(\vec{p}^T\vec{w})$$

$$RR^T\vec{w} = (25\vec{w}^T\vec{w})\vec{w} + 0$$

$$\implies \lambda_1 = 25\vec{w}^T\vec{w} = 5^2 7^2$$

$$\implies \sigma_1 = \sqrt{\lambda_1} = 35$$

When computing the first eigenvalue, we used the fact that $\vec{p}^T\vec{w} = 0$. We can repeat the same process for the second eigenvalue:

$$RR^T\vec{p} = 25\vec{w}(\vec{w}^T\vec{p}) + 18\vec{p}(\vec{p}^T\vec{w})$$

$$RR^T\vec{p} = 0 + 18(\vec{p}^T\vec{p})\vec{p}$$

$$\implies \lambda_2 = 18\vec{p}^T\vec{p} = 3^2 7^2 2$$

$$\implies \sigma_2 = \sqrt{\lambda_2} = 21\sqrt{2}$$

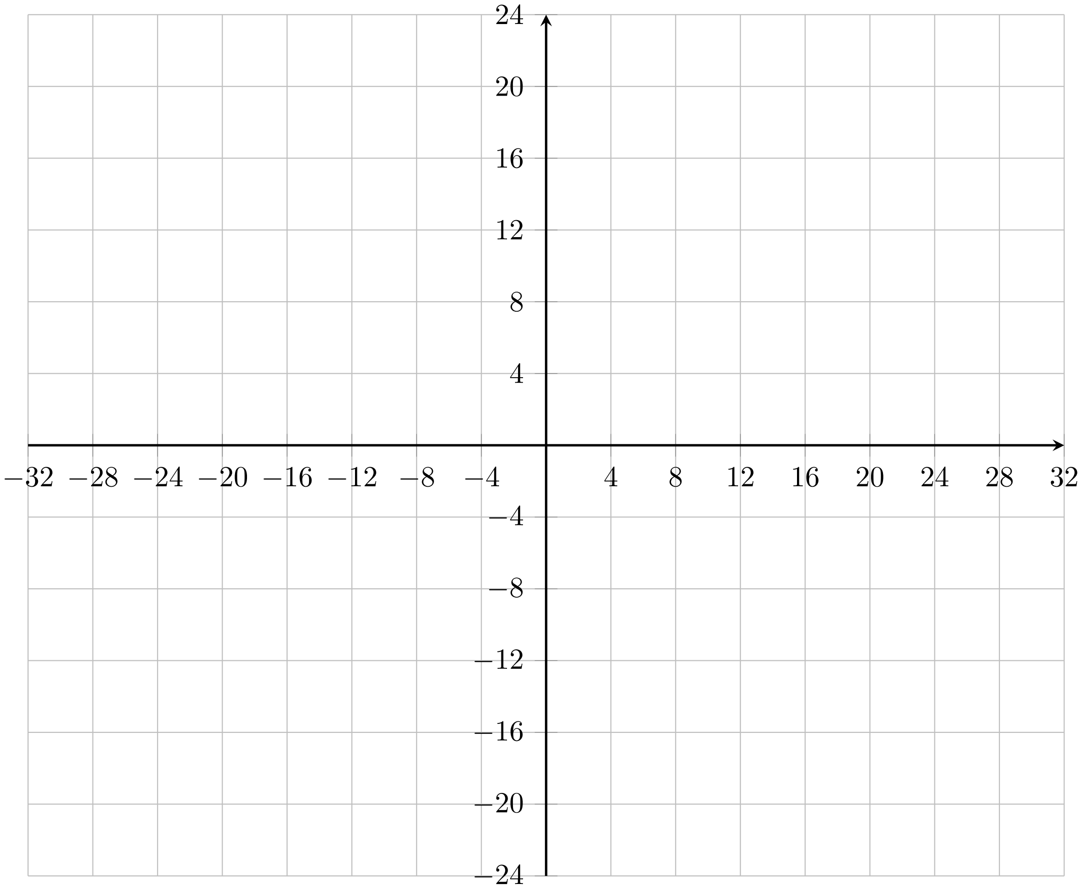
For PCA, we require normalized vectors, so for that reason our first principal component is $\frac{\vec{w}}{\|\vec{w}\|} = \frac{\vec{w}}{7}$ with a corresponding singular value of $\sigma_1 = 35$. Our second principal component is $\frac{\vec{p}}{\|\vec{p}\|} = \frac{\vec{p}}{7}$ with a corresponding singular value of $\sigma_2 = 21\sqrt{2}$. As mentioned in the solution above, we choose $\sigma_1 = 35$, $\sigma_2 = 21\sqrt{2}$ in this order because $35 > 21\sqrt{2}$.

(d) In the previous part, you had

$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix}$$

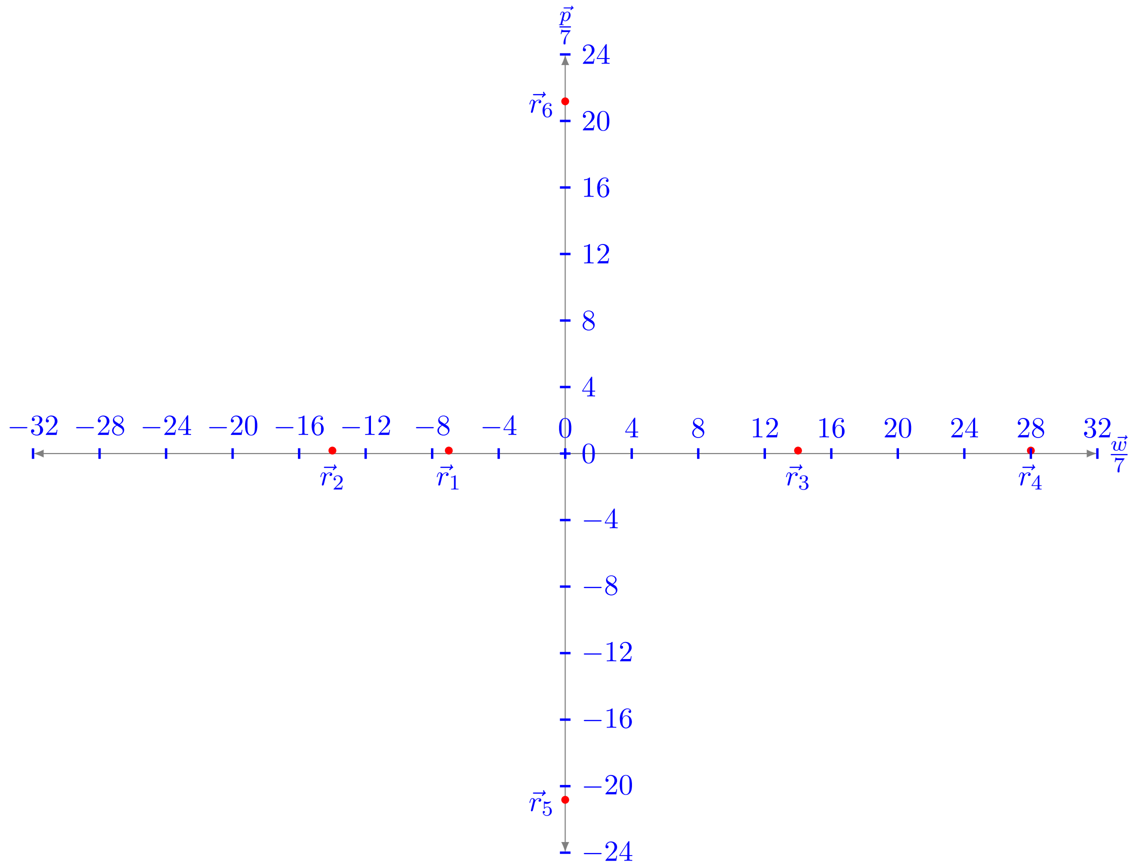
with $\|\vec{w}\| = 7$ and $\|\vec{p}\| = 7$, satisfying $\vec{p}^\top\vec{w} = 0$.

If we use $\vec{r}_i$ to denote the $i$-th column of $R$, **plot the movie data points $\vec{r}_i$ projected onto the first and second principal component vectors along the columns of R.** The coordinate along the first principal component should be represented by horizontal axis and the coordinate along the second principal component should be the vertical axis. **Label each point, and the axes. Remember that principal component vectors are normalized**.




**Solution:**

Once we know what the principal components are, we know that the first four data points are just multiples of the first principal component and the last two data points are just multiples of the second principal component. What multiples? For the first four, the multiples are clearly $-7, -14, 14, 28$ since the norm of $\vec{w}$ is 7. For the final two, the multiples are clearly $-21, +21$ since the norm of $\vec{p}$ is also 7. Plotting:



$\frac{\vec{p}}{7}$
$\frac{\vec{w}}{7}$
24
20
16
12
8
4
0
−4
−8
−12
−16
−20
−24
−32 −28 −24 −20 −16 −12 −8 −4 0 4 8 12 16 20 24 28 32
$\vec{r}_1$
$\vec{r}_2$
$\vec{r}_3$
$\vec{r}_4$
$\vec{r}_5$
$\vec{r}_6$

**7. Image Compression With SVD**

Open the **image_compression.ipynb** Jupyter Notebook. Make sure to read through the theory section to understand how we will use what we know about PCA and SVD in order to compress images! Once you have read through the math and are comfortable with it, go ahead and start running the coding cells and answer the questions below.

(a) Run all cells up to and throughout Part 1. Thoroughly read through the 'rank_k_approx' function. **Since we are now only using the top $k$ singular values, in terms of $k, m,$ and $n$, how many real numbers do we require to describe the rank $k$ approximation of $A$?**

**Solution:** We require storage of $k(m+n+1)$ real numbers.

(b) Run the cells in Part 2 of the notebook. Notice there are two images: One compressed, one uncompressed. **Can you easily visually detect any difference between the two images even though we are using only about 40% of the data?**

**Solution:** We just expect that you notice very little to no change between the images.

(c) Run the cells that plot the singular values. **Given the plots above, why did the image look so good even with only about 40% of the data?**

**Solution:** The compression worked well because the first $k = 200$ singular values are much larger than the remaining singular values.

(d) Play around with the slider for the kitten picture. **What is the lowest acceptable value of $k$ that you can go without losing too much image quality? What were the memory savings?**
*Note: Use your best judgment since this depends on eyesight, screen resolution, etc.*

**Solution:** Playing around with the slider, $k = 75$ was the lowest we could go without noticing too much image loss. The memory savings are $\approx 15$

(e) Plot the singular values for the image of the US flag. **What do you notice about the singular values here in comparison to the kitten's singular values? What does this mean for our low rank compression?**

**Solution:** The singular values have a much sharper cutoff. This means that we can compress this image much better than the kitten.

(f) Play around with the interactive slider for the US flag. **What is the lowest acceptable value of $k$ here? What is the memory saving?**

**Solution:** Here we can get away with $k = 20$ resulting in a memory saving of $\approx 5\%$! This is much better than the kitten! This is true because this image is much more structured than the kitten.

**8. Homework Process and Study Group**

Citing sources and collaborators are an important part of life, including being a student!
We also want to understand what resources you find helpful and how much time homework is taking, so we can change things in the future if possible.

(a) **What sources (if any) did you use as you worked through the homework?**

(b) **If you worked with someone on this homework, who did you work with?**
List names and student ID's. (In case of homework party, you can also just describe the group.)

**Contributors:**

- Elena Jia.
- Neelesh Ramachandran.
- Aditya Arun.
- Daniel Abraham.
- Ashwin Vangipuram.
- Nikhil Shinde.
- Gaoyue Zhou.
- Nathan Lambert.
- Anant Sahai.