Homework 5

Gram-Schmidt

1. Use the Gram-Schmidt algorithm to compute QR decompositions of the following matrices.

   (a) $A_1 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

   (b) $A_2 = \begin{pmatrix} 0 & 1 & 1 \\ 3 & 0 & 0 \\ 4 & 0 & 1 \end{pmatrix}$

Solution:

(a)

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{a_1} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ r_{11} &= \|\mathbf{v}_1\| = \sqrt{10} \\ \mathbf{q_1} &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{pmatrix} \end{aligned}$$

$$\begin{aligned} r_{12} &= \mathbf{q_1}^T \mathbf{a_2} = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \frac{2}{\sqrt{10}} + \frac{12}{\sqrt{10}} = \frac{14}{\sqrt{10}} \\ \mathbf{v}_2 &= \mathbf{a_2} - r_{12}\mathbf{q_1} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} - \begin{pmatrix} \frac{14}{10} \\ \frac{42}{10} \end{pmatrix} = \begin{pmatrix} \frac{3}{5} \\ -\frac{1}{5} \end{pmatrix} \\ r_{22} &= \|\mathbf{v}_2\| = \sqrt{\frac{10}{25}} = \frac{\sqrt{10}}{5} \\ \mathbf{q_2} &= \frac{\mathbf{v}_2}{r_{22}} = \begin{pmatrix} \frac{3}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} \end{pmatrix} \end{aligned}$$

Summary:

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} \sqrt{10} & \frac{14}{\sqrt{10}} \\ 0 & \frac{\sqrt{10}}{5} \end{pmatrix}$$

(b)

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{a_1} \\ r_{11} &= \|\mathbf{v}_1\| = 5 \\ \mathbf{q_1} &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{pmatrix} 0 \\ \frac{0}{5} \\ \frac{4}{5} \end{pmatrix} \end{aligned}$$

$$r_{12} = \mathbf{q_1}^T\mathbf{a_2} = 0$$

$$\mathbf{v}_2 = \mathbf{a_2} - (\mathbf{q_1}^T\mathbf{a_2})\mathbf{q_1} = \begin{pmatrix}1\\0\\0\end{pmatrix}$$

$$r_{22} = \|\mathbf{v}_2\| = 1$$

$$\mathbf{q_2} = \begin{pmatrix}1\\0\\0\end{pmatrix}$$

$$r_{13} = \mathbf{q_1}^T\mathbf{a_3} = \frac{4}{5}$$

$$r_{23} = \mathbf{q_2}^T\mathbf{a_3} = 1$$

$$\mathbf{v}_3 = \mathbf{a_3} - r_{13}\mathbf{q_1} - r_{23}\mathbf{q_2} = \begin{pmatrix}1\\0\\1\end{pmatrix} - \frac{4}{5}\begin{pmatrix}0\\\frac{3}{5}\\\frac{4}{5}\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\-\frac{12}{25}\\\frac{9}{25}\end{pmatrix}$$

$$r_{33} = \|\mathbf{v}_3\| = \sqrt{\frac{12^2}{25^2} + \frac{9^2}{25^2}} = \frac{15}{25} = \frac{3}{5}$$

$$\mathbf{q_3} = \frac{\mathbf{v}_3}{r_{33}} = \begin{pmatrix}0\\-\frac{4}{5}\\\frac{3}{5}\end{pmatrix}$$

Summary:

$$A = \begin{pmatrix}0 & 1 & 0\\ \frac{3}{5} & 0 & -\frac{4}{5}\\ \frac{4}{5} & 0 & \frac{3}{5}\end{pmatrix}\begin{pmatrix}5 & 0 & \frac{4}{5}\\ 0 & 1 & 1\\ 0 & 0 & \frac{3}{5}\end{pmatrix}$$

2. Let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k\}$ be an orthonormal set in $\mathbb{R}^n$.

   (a) Show that
   $$P_1 = (I - q_1q_1^T)(I - q_2q_2^T)\ldots(I - q_kq_k^T)$$
   is a projection matrix.

   (b) Show that
   $$P_2 = I - q_1q_1^T - q_2q_2^T - \ldots - q_kq_k^T$$
   is a projection matrix.

   (c) Show that $P_1 = P_2$. Though mathematically equivalent, the first form of projection is more numerically stable and is used in modified Gram-Schmidt, whereas the second from is used in classical Gram-Schmidt.

Solution:

It is okay to first show (b), then (c), and then conclude that (a) follows from (b) and (c). It is also okay to do the 3 parts in order directly.

We will first show **(c)**. Let
$$S_j = \sum_{i=1}^{j} q_iq_i^T,$$
and note that $S_jq_{j+1}q_{j+1}^T = 0$ because of the orthogonality of the q_j's. Then
$$(I - S_{j-1})(I - q_jq_j^T) = I - S_{j-1} - q_jq_j^T + S_{j-1}q_jq_j^T = I - S_{j-1} - q_jq_j^T = I - S_j. \tag{1}$$
Applying this to the defition of P_1, it follows that $P_1 = P_2$. Next, we show **(b)** that P_2 is a projection matrix. Let $S = S_k$, so that $P_2 = I - S$, and note that S is a projection matrix since
$$S^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} q_iq_i^Tq_jq_j^T = \sum_{i=1}^{k}\sum_{j=1}^{k} q_i\delta_{ij}q_j^T = \sum_{i=1}^{k} q_iq_i^T = S.$$

Then P_2 is a projection, since

$$P_2^2 = (I - S)(I - S) = I - 2S + S^2 = I - 2S + S = I - S = P_2.$$

Finally, **(c)** and **(b)** together imply **(a)**.

In the next two problems, you will implement QR factorization by both the classical and the modified Gram-Schmidt algorithms, and study the instability of classical Gram-Schmidt. Note: You may do this assignment in Python if you prefer (in which case you should convert the code skeleton below to Python).

3. Write Matlab/Octave code to compute matrices Q and R such that $A = QR$ using the Gram-Schmidt process, or the modified Gram-Schmidt process, by filling in the function below. Include your code.

```
function [Q,R] = MyGS(A,modified)

[m,n] = size(A);
R = zeros(n);

for k=1:n
    v_k = A(:,k);
    % orthogonalize against previous columns
    for j=1:k-1
        if (modified)
           R(j,k) = Q(:,j)'*v_k; % modified
        else
           R(j,k) = Q(:,j)'*A(:,k); % classical
        end
        v_k = v_k - Q(:,j)*R(j,k);
    end
    R(k,k) = norm(v_k);
    Q(:,k) = v_k ./ R(k,k);
end

end
```

4. Study the instability of Classical Gram-Schmidt by running the function below.

```
function GSInstability(lo,hi)

mGS_orthogonality = [];
mGS_factorization = [];
cGS_orthogonality = [];
cGS_factorization = [];

for i = lo:hi
    A = hilb(i) + eye(i) * 1e-6;
    [Q,R] = MyGS(A,false);
    cGS_orthogonality(i-lo+1) = norm(Q'*Q-eye(i));
    cGS_factorization(i-lo+1) = norm(Q*R - A);
    [Q,R] = MyGS(A,true);
    mGS_orthogonality(i-lo+1) = norm(Q'*Q-eye(i));
    mGS_factorization(i-lo+1) = norm(Q*R - A);
end

figure
hold on
plot(lo:hi,log(cGS_orthogonality),'--k');
plot(lo:hi,log(mGS_orthogonality),'-b');
title('GS orthogonality');

figure
hold on
plot(lo:hi,cGS_factorization,'--k');
plot(lo:hi,mGS_factorization,'-b');
title('GS factorization');
end
```

(a) What is being plotted by the code?

(b) Include the plots generated.

(c) Are Modified Gram-Schmidt and Classical Gram-Schmidt computing accurate factorizations? I.e., how close are A and $Q * R$? Explain based on the plot generated.

(d) Are Modified Gram-Schmidt and Classical Gram-Schmidt computing an orthogonal Q? Explain based on the plot generated.

Solution:

(a) The first plot compares the orthogonality of the Q matrices obtained through Classical Gram-Schmidt and Modified Gram-Schmidt.

GSInstability.png

(b)

(c) Both Modified Gram-Schmidt and Classical Gram-Schmidt compute factorizations where $\|QR - A\|$ is very small, so both satisfy $A = QR$ up to some small numerical error.

(d) Modified Gram-Schmidt and Classical Gram-Schmidt differ significantly in the orthogonality of the Q that they compute. For Modified Gram-Schmidt, the orthogonality error $\|Q^T Q - I\|$ is very small as desired, but for Classical Gram-Schmidt, the orthgonality error is many orders of magnitude larger and increases as the size of the matrix increases.

Householder transformations

5. (Strang II.2 7) Find a Householder reflection matrix H such that $H \begin{pmatrix} 4 \\ 2 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ 0 \\ 0 \end{pmatrix}$, for some $r_1, r_2 \in \mathbb{R}$.

Solution:

Let

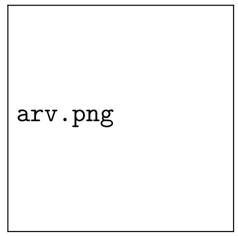
$$H = \begin{pmatrix} 1 & 0 \\ 0 & \hat{H} \end{pmatrix},$$

with $\hat{H} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} r_2 \\ 0 \\ 0 \end{pmatrix}$. Then $r_1 = 4$, and $r_2 = \sqrt{4+1+4} = 3$. The Householder vector is $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} =$

$\begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix}$, so

$$\hat{H} = I - 2\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} = I - \frac{1}{3}\begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & 2 \\ -2 & 2 & -1 \end{pmatrix}$$

6. (Strang II.2 6) A Householder reflection matrix has the form $H = I - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2}$. Let $\mathbf{v} = \mathbf{a} - \mathbf{r}$, where $\|\mathbf{a}\|_2 = \|\mathbf{r}\|_2$, as illustrated in the figure. Confirm that $H\mathbf{a} = \mathbf{r}$. This shows how to construct a Householder reflection matrix that reflects one vector to another, as in the case of Householder QR, where $\mathbf{r} = \alpha\mathbf{e}_1$.



Solution:

We will use the fact that $\mathbf{a}^T\mathbf{a} = \mathbf{r}^T\mathbf{r}$.

$$\begin{aligned} H\mathbf{a} &= \left(I - \frac{2\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2}\right)\mathbf{a} = \left(I - \frac{2(\mathbf{a}-\mathbf{r})(\mathbf{a}-\mathbf{r})^T}{(\mathbf{a}-\mathbf{r})^T(\mathbf{a}-\mathbf{r})}\right)\mathbf{a} \\ &= \mathbf{a} - \frac{2(\mathbf{a}-\mathbf{r})(\mathbf{a}^T\mathbf{a} - \mathbf{r}^T\mathbf{a})}{(\mathbf{a}-\mathbf{r})^T(\mathbf{a}-\mathbf{r})} = \mathbf{a} - \frac{2(\mathbf{a}-\mathbf{r})(\mathbf{a}^T\mathbf{a} - \mathbf{r}^T\mathbf{a})}{(\mathbf{a}^T\mathbf{a} - 2\mathbf{r}^T\mathbf{a} + \mathbf{r}^T\mathbf{r})} \\ &= \mathbf{a} - \frac{2(\mathbf{a}-\mathbf{r})(\mathbf{a}^T\mathbf{a} - \mathbf{r}^T\mathbf{a})}{2(\mathbf{a}^T\mathbf{a} - \mathbf{r}^T\mathbf{a})} = \mathbf{a} - (\mathbf{a}-\mathbf{r}) = \mathbf{r}. \end{aligned}$$

Singular Value Decomposition

7. (T&B 4.1) Determine SVDs of the following matrices (by hand calculation):

(a) $\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$, (b) $\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$, (c) $\begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$, (d) $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, (e) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

Solution:

In general, the SVD of A is written $A = U\Sigma V^T$, with

- U, V orthogonal matrices, and
- Σ diagonal matrix with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.

(a) This is almost in SVD form. We just to negate the -2 since all singular values should be non-negative.

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Note that the left-most matrix acts to row-scale the middle matrix.

(b) Here we just need to symmetrically permute the matrix so that the positions of the diagonal elements are flipped. The left-most matrix applies a row permutation while the right-most marix applies a column permutation.

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

(c) Here we only require a column permutation.

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}2&0\\0&0\\0&0\end{pmatrix}\begin{pmatrix}0&1\\1&0\end{pmatrix}$$

For parts (d) and (e), we note that for a 2×2 matrix, the SVD is

$$U\Sigma V^T = \begin{pmatrix} | & | \\ \mathbf{u}_1 & \mathbf{u}_2 \\ | & | \end{pmatrix}\begin{pmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{pmatrix}\begin{pmatrix} -\mathbf{v}_1^T- \\ -\mathbf{v}_2^T- \end{pmatrix} = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T.$$

Therefore we can construct A out of two rank-one matrices $\sigma_1\mathbf{u}_1\mathbf{v}_1^T$ and $\sigma_2\mathbf{u}_2\mathbf{v}_2^T$.
(d) We note that $\begin{pmatrix}1&1\\0&0\end{pmatrix}$ is rank one, because it has a repeated column, so we write it as $\sigma_1\mathbf{u}_1\mathbf{v}_1^T$:

$$\begin{pmatrix}1&1\\0&0\end{pmatrix} = 1\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1&1\end{pmatrix}$$

This is not yet in the desired form, because $\begin{pmatrix}1&1\end{pmatrix}$ is not normalized. Normalizing, we get $\sqrt{2}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix} = \begin{pmatrix}1&1\end{pmatrix}$, so

$$1\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1&1\end{pmatrix} = \sqrt{2}\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix} = \sigma_1\mathbf{u}_1\mathbf{v}_1^T.$$

Since the matrix is rank one, $\sigma_2 = 0$. Now it just remains to find unit vectors $\mathbf{u}_2$ and $\mathbf{v}_2$ that are normalized and orthogonal to $\mathbf{u}_1$ and $\mathbf{v}_1$, respectively. These are given by

$$\mathbf{u}_2 = \begin{pmatrix}0\\1\end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}.$$

Combining all the terms, we have

$$\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}\sqrt{2}&0\\0&0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix}.$$

(e) We take a similar approach as in (d). The matrix has rank one and

$$\begin{pmatrix}1&1\\1&1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}1&1\end{pmatrix}.$$

Normalizing, we have

$$= 2\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix},$$

giving

$$\sigma_1 = 2,\ \mathbf{u}_1 = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix},\ \mathbf{v}_1 = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}.$$

Since the matrix is rank one, $\sigma_2 = 0$. Completing the orthonomal bases for U and V, we get

$$\mathbf{u}_2 = \begin{pmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix},\ \mathbf{v}_2 = \begin{pmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}.$$

So the SVD is

$$\begin{pmatrix}\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}2&0\\0&0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix}.$$

8. Let A be an $m \times n$ singular matrix of rank r with SVD

$$A = U\Sigma V^T = \left(\begin{array}{c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_m \end{array}\right) \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \left(\begin{array}{c} \mathbf{v}_1^T \\ \hline \mathbf{v}_2^T \\ \hline \vdots \\ \hline \mathbf{v}_n^T \end{array}\right)$$

$$= \begin{pmatrix} \hat{U} & \tilde{U} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \begin{pmatrix} \hat{V}^T \\ \tilde{V}^T \end{pmatrix}$$

where $\sigma_1 \geq \ldots \geq \sigma_r > 0$, $\hat{U}$ consists of the first r columns of U, $\tilde{U}$ consists of the remaining $m - r$ columns of U, $\hat{V}$ consists of the first r columns of V, and $\tilde{V}$ consists of the remaining $n - r$ columns of V. Give bases for the spaces range(A), null(A), range(A^T) and null(A^T) in terms of the components of the SVD of A, and a brief justification.

Solution:

A summary of the relationships between the columns of U and V and the spaces associated with A is given in the following table:

range(A)	$\hat{U}$
null(A)	$\tilde{V}$
range(A^T)	$\hat{V}$
null(A^T)	$\tilde{U}$

These results are explained below. First note that A can also be written as

$$A = \sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Given any vector $\mathbf{x} \in \mathbb{R}^n$, we have using the above sum,

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

This shows that any $\mathbf{y} = A\mathbf{x}$ in the range of A is a linear combination of $\mathbf{u}_i, 1 \leq i \leq r$. Therefore, the columns of $\hat{U}$ form a basis for range(A).
Now let $\mathbf{z}$ be such that $A\mathbf{z} = \mathbf{0}$, i.e., $\mathbf{z} \in$ null(A). Then

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

Since $\sigma_i \neq 0, 1 \leq i \leq r$, and the $\mathbf{u}_i$ are linearly independent, we must have $\mathbf{v}_i^T \mathbf{z} = 0, 1 \leq i \leq r$ or $\mathbf{z} \perp \hat{V}$. Therefore, $\mathbf{z} \in$ span$(\tilde{V})$, so the columns of $\tilde{V}$ form a basis for null(A). The SVD of A^T is $A^T = V\Sigma^T U^T$. Therefore, by the same arguments above for A, the columns of $\hat{V}$ form a basis for range(A), and the columns of $\tilde{U}$ form a basis for null(A^T).

9. Use the SVD of A to show that for an $m \times n$ matrix of full column rank n, the matrix $A(A^TA)^{-1}A^T$ is an orthogonal projector onto range(A).

Solution:

Let $P = A(A^TA)^{-1}A^T$. First, P is a <u>projector</u> if P is idempotent, i.e., $P^2 = P$. We check this:

$$\begin{aligned} P^2 &= \left[A(A^TA)^{-1}A^T\right]\left[A(A^TA)^{-1}A^T\right] \\ &= A\left[(A^TA)^{-1}A^TA\right](A^TA)^{-1}A^T \\ &= A(A^TA)^{-1}A^T = P. \end{aligned}$$

Next, P is an <u>orthogonal projector</u> if $P = P^T$. We check this:

$$\begin{aligned} P^T &= \left[A(A^TA)^{-1}A^T\right]^T = (A^T)^T(A^TA)^{-T}A^T \\ &= A\left[(A^TA)^T\right]^{-1}A^T = A(A^TA)^{-1}A^T = P. \end{aligned}$$

Finally, we show that <u>range(P) = range(A)</u>. Since A has full column rank n, then we can write

$$A = \sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad \sigma_1 \geq \ldots \geq \sigma_n > 0.$$

Substituting this expression for P, we get

$$\begin{aligned} P = A(A^TA)^{-1}A^T &= (\sum_{i=1}^{n}\sigma_i\mathbf{u}_i\mathbf{v}_i^T)\left[(\sum_{i=1}^{n}\sigma_i\mathbf{v}_i\mathbf{u}_i^T)(\sum_{i=1}^{n}\sigma_i\mathbf{u}_i\mathbf{v}_i^T)\right]^{-1}(\sum_{i=1}^{n}\sigma_i\mathbf{v}_i\mathbf{u}_i^T) \\ &= (\sum_{i=1}^{n}\sigma_i\mathbf{u}_i\mathbf{v}_i^T)(\sum_{i=1}^{n}\sigma_i^2\mathbf{v}_i\mathbf{v}_i^T)^{-1}(\sum_{i=1}^{n}\sigma_i\mathbf{v}_i\mathbf{u}_i^T) \\ &= (\sum_{i=1}^{n}\sigma_i\mathbf{u}_i\mathbf{v}_i^T)(\sum_{i=1}^{n}\sigma_i^{-2}\mathbf{v}_i\mathbf{v}_i^T)(\sum_{i=1}^{n}\sigma_i\mathbf{v}_i\mathbf{u}_i^T) \\ &= (\sum_{i=1}^{n}\sigma_i^{-1}\mathbf{u}_i\mathbf{v}_i^T)(\sum_{i=1}^{n}\sigma_i\mathbf{v}_i\mathbf{u}_i^T) = \sum_{i=1}^{n}\mathbf{u}_i\mathbf{u}_i^T. \end{aligned}$$

Note that this is an SVD form for P, and by the arguments in Problem 6, range(P) = span($\mathbf{u}_1, \ldots, \mathbf{u}_n$) = range(A).

10. (adapted from I.9 1) **(Eckart-Young theorem in ℓ^2)** Consider a matrix A with SVD $A = U\Sigma V^T$. The matrix A could have a large rank. Constructing **low rank** approximations to A can be very useful (e.g., data compression). One such low rank approximation, constructed from the SVD, is

$$A_k = \sum_{i=1}^{k}\sigma_i\mathbf{u}_i\mathbf{v}_i^T,$$

for small k, i.e., the approximation we get by keeping the largest k terms in the SVD of A. The Eckart-Young Theorem says that A_k is a **best** rank-k approximation to A (when "best" is measured in ℓ^2). That is, any other rank-k approximation, B, will be no better: $\|A - B\|_2 \geq \|A - A_k\|_2 = \sigma_{k+1}$. What are the singular values (in descending order) of $A - A_k$? Omit any zeros.

Solution:

We know that

$$A = \sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^T,$$

where r is the rank of A. Therefore,

$$A - A_k = \sum_{i=k+1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^T.$$

This tells us the SVD of $A - A_k$, so its singular values are $\sigma_{k+1} \geq \ldots \geq \sigma_r$.

11. (adapted from I.9 2) Find a closest rank-1 approximation to these matrices (L^2 or Frobenius norm $\|A\|_F = \sqrt{\mathrm{tr}\{(A^TA)\}}$):

$$A_1 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & 3 \\ 2 & 0 \end{pmatrix} \qquad A_3 = \begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix}$$

Solution:

Denote the closest rank-1 approximation to A_i as $\tilde{A}_i$. A_1 already has the form of Σ in its SVD, so $\tilde{A}_1 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

The SVD of A_2 is

$$A_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

so the closest rank-1 approximation is

$$\begin{aligned} \tilde{A}_2 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Finally, using Octave/Matlab we find that

$$A_3 = \begin{pmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{pmatrix} \begin{pmatrix} 6.7082 & 0 \\ 0 & 2.2361 \end{pmatrix} \begin{pmatrix} 0.3162 & 0.9487 \\ -0.9487 & 0.3162 \end{pmatrix},$$

so that

$$\begin{aligned} \tilde{A}_3 &= \begin{pmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{pmatrix} \begin{pmatrix} 6.7082 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0.3162 & 0.9487 \\ -0.9487 & 0.3162 \end{pmatrix} \\ &= \begin{pmatrix} 1.5000 & 4.5000 \\ 1.5000 & 4.5000 \end{pmatrix} \end{aligned}$$