Homework 12

This homework is due on Saturday, April 20, 2024 at 11:59PM.

1. SVD Proofs

(a) **Show, for a general matrix** $A \in \mathbb{R}^{m \times n}$ **with** $\text{rank}(A) = r$ **and** $A = U\Sigma V^{\top}$**, that** $\text{Null}(A) = \text{Col}(V_{n-r})$**.**

(HINT: How do we show that two spaces are equal? Consider using the outer product form of SVD. It may be helpful to recall the rank-nullity theorem: $\dim \text{Col}(A) + \dim \text{Null}(A) = n$*.)*

(b) **Show, for a general matrix** $A \in \mathbb{R}^{m \times n}$ **with** $\text{rank}(A) = r$ **and** $A = U\Sigma V^{\top}$**, that** $\text{Col}(A) = \text{Col}(U_r)$**.**

(c) **Show, for a general matrix** $A \in \mathbb{R}^{m \times n}$ **with** $\text{rank}(A) = r$ **and** $A = U\Sigma V^{\top}$**, that** $\text{Null}\left(A^{\top}\right) = \text{Col}(U_{m-r})$ **and** $\text{Col}\left(A^{\top}\right) = \text{Col}(V_r)$**. Use these facts and the previous subparts to show that** $\dim \text{Col}(A) + \dim \text{Null}\left(A^{\top}\right) = m$**, and that** $\text{Col}(A)$ **and** $\text{Null}\left(A^{\top}\right)$ **are orthogonal.**

2. SVD System ID

Previously, we saw instances for how to solve system ID problems when $D \in \mathbb{R}^{m\times n}$ is full rank (for $m > n$). Now, let us consider more generally the following problem of estimating $\vec{p}$ in

$$D\vec{p} = \vec{s} \tag{1}$$

where $\vec{p} \in \mathbb{R}^n$, $\vec{s} \in \mathbb{R}^m$, and $D \in \mathbb{R}^{m\times n}$. We assume that $\text{rank}(D) = r < \min(m, n)$, and we do not make any further assumptions on the relationship between m and n. Let's assume that D has an SVD given by

$$D = U\Sigma V^\top \tag{2}$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthonormal matrices. $\Sigma \in \mathbb{R}^{m\times n}$ has the following form:

$$\Sigma = \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \tag{3}$$

where $\Sigma_r = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$ is a $r \times r$ diagonal matrix with nonzero elements along its diagonal.

Using this problem setup, we can rewrite our original system ID problem as

$$U\Sigma V^\top \vec{p} = \vec{s} \tag{4}$$

Our goal is to find $\vec{p}$ with smallest norm that best estimates $\vec{s}$.

For notational convenience, denote $U := \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ where $U_r \in \mathbb{R}^{m\times r}$ is a matrix with the first r columns of U and $U_{m-r} \in \mathbb{R}^{m\times(m-r)}$ is a matrix with the last $m - r$ columns of U. Also, denote $V := \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ where $V_r \in \mathbb{R}^{n\times r}$ is a matrix that has the first r columns of V, and $V_{n-r} \in \mathbb{R}^{n\times(n-r)}$ is a matrix that has the last $n - r$ columns of V. From SVD properties, we know that the columns of U_r form an orthonormal basis for $\text{Col}(D)$ and that the columns of V_{n-r} form an orthonormal basis for $\text{Null}(D)$.

(a) **Using the fact that U is orthonormal, show that $\Sigma V^\top \vec{p} = U^\top \vec{s}$.**

(b) **Find $\vec{\alpha} \in \mathbb{R}^r$ and $\vec{\beta} \in \mathbb{R}^{n-r}$ such that $\vec{p} = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix} \begin{bmatrix} \vec{\alpha} \\ \vec{\beta} \end{bmatrix}$ ($\vec{\alpha}$ and $\vec{\beta}$ can be in terms of $\vec{p}$ and one or more other matrices). Show that changing $\vec{\beta}$ will not affect the result of $D\vec{p}$.** Since $\vec{\beta}$ does not affect the result of $D\vec{p}$ ($V_{n-r}\vec{\beta}$ is the component of $\vec{p}$ that is in the nullspace of D), we can set $\vec{\beta}$ to be whatever we want. To choose the $\vec{p}$ with smallest norm, we will set $\vec{\beta} = \vec{0}$ for future subparts.

(c) From the previous part, we can rewrite $\vec{p} = V \begin{bmatrix} \vec{\alpha} \\ \vec{0} \end{bmatrix}$. This simplifies our system ID problem as follows:

$$\Sigma V^\top V \begin{bmatrix} \vec{\alpha} \\ \vec{0} \end{bmatrix} = U^\top \vec{s} \tag{5}$$

$$\Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0} \end{bmatrix} = U^\top \vec{s} \tag{6}$$

 All Rights Reserved. This may not be publicly shared without explicit permission.

Expand the left hand side of eq. (6) using eq. (3). Rewrite $U^\top \vec{s}$ **as** $\begin{bmatrix} U_r^\top \vec{s} \\ U_{m-r}^\top \vec{s} \end{bmatrix}$ **and find an expression for** $\vec{\alpha}$**.** *(HINT: Block matrix multiplication will work like normal matrix-vector multiplication here since* $\Sigma_r \in \mathbb{R}^{r \times r}$ *and* $\vec{\alpha} \in \mathbb{R}^r$*.)*

(d) **Use the previous part to come up with a solution for** $\vec{p}$**.**

(e) From the concept of projections, we know that the optimal solution for $\vec{p}$ satisfies the property that the projection error, namely $\vec{s} - D\vec{p}$, is orthogonal to the projection itself, namely $D\vec{p}$. **Find** $\vec{w} \in \mathbb{R}^r$ **and** $\vec{z} \in \mathbb{R}^{m-r}$ **such that** $\vec{s} := \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{z} \end{bmatrix}$**. Using this, show that our solution for** $\vec{p}$ **is optimal.**

© UCB EECS 16B, Spring 2024. All Rights Reserved. This may not be publicly shared without explicit permission.

3. Rank 1 Decomposition and Image Compression With SVD

In this problem, we will introduce an important application of the singular value decomposition (SVD) called low-rank approximation in the context of image compression. Given a matrix $A \in \mathbb{R}^{m\times n}$ of rank $r \leq \min\{m,n\}$, we know that we can write A using the outer-product form of the SVD:

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top. \tag{7}$$

If our matrix is high-rank, i.e., $r \approx \min\{m,n\}$, then almost all the σ_i will be nonzero and non-negligible. However, if the data has some linear, low-rank essential structure, as is usually the case with real data such as images, most of our singular values will be very small (but usually nonzero due to noise or disturbances). If, say, the data has intrinsic linear rank ℓ, then the first ℓ singular values are large, and the remaining $r-\ell$ are small:

$$\begin{aligned} A &= \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top && (8)\\ &= \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^\top + \sum_{i=\ell+1}^{r} \underbrace{\sigma_i}_{\approx 0} \vec{u}_i \vec{v}_i^\top && (9)\\ &\approx \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^\top. && (10) \end{aligned}$$

This motivates approximating the data as

$$A_\ell := \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^\top \tag{11}$$

and using this compressed data for further analysis. As images are readily represented as matrices, in this problem, we will see how to make use of this low rank approximation technique to compress images without severely deteriorating the image quality.

We will start by decomposing a few images into linear combinations of rank 1 matrices. Remember that outer product of two vectors $\vec{s}\vec{g}^\top$ gives a rank 1 matrix. It has rank 1 because the column span is one-dimensional — multiples of $\vec{s}$ only — and the row span is also one dimensional — multiples of $\vec{g}^\top$ only. More generally, this exercise should provide intuition into how the outer product form of the SVD decomposes the underlying matrix into a sum of constituent rank 1 matrices with different weightings given by the singular values.

(a) Consider a standard 4×4 checkerboard shown in Figure 1. Assume that black colors represent -1 and that white colors represent 1.

 All Rights Reserved. This may not be publicly shared without explicit permission.

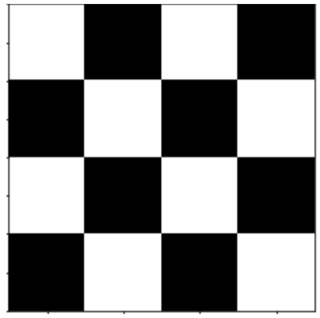

Figure 1: 4×4 checkerboard.

The checkerboard is given by the following 4×4 matrix L:

$$L = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \tag{12}$$

Express L as a linear combination of outer products.

(HINT: In order to determine how many rank 1 matrices you need to combine to represent the matrix, find the rank of the matrix you are trying to represent.)

(b) We next consider the same checkerboard given by the following 4×4 matrix C where black colors now represent 0 and that white colors represent 1.

$$C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \tag{13}$$

Express C as a linear combination of outer products.
(HINT: Note that C is a rank 2 matrix and therefore you need 2 outer products to represent it. To find these individual outer products consider the columns of C.)

(c) Now consider the Swiss flag shown in Figure 2. Assume that red colors represent 0 and that white colors represent 1.

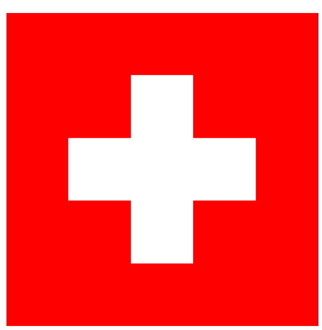

Figure 2: Swiss flag.

© UCB EECS 16B, Spring 2024. All Rights Reserved. This may not be publicly shared without explicit permission.

Assume that the Swiss flag is given by the following 5×5 matrix S:

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{14}$$

Express S as a linear combination of outer products.

(d) To represent the full $A \in \mathbb{R}^{m \times n}$, we need $m \cdot n$ numbers. If we instead take the rank-k approximation $A_k =:= \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^\top$, **in terms of k, m, and n, how many real numbers do we require to describe A_k?**

(e) Here we show two images of a kitten: An uncompressed A with dimensions 900×1200 (right), and a compressed rank 200 approximation of A (left). **Can you easily spot any difference between the two images even though we are using only about 40% of the data?**

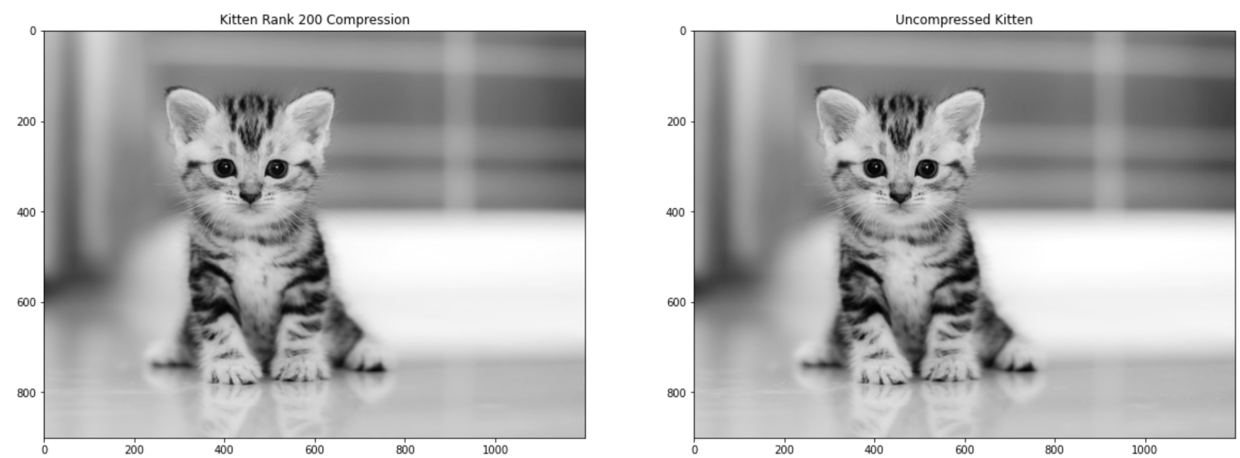

Figure 3: Compressed vs uncompressed image of a kitten.

4. PCA Introduction

Let $X \in \mathbb{R}^{m \times n}$ be defined as $X := \begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \end{bmatrix}$ where each $\vec{x}_i \in \mathbb{R}^m$. Let X have an SVD $X = U\Sigma V^\top$. Now, let $U_\ell := \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_\ell \end{bmatrix}$ where $\vec{u}_i$ is the ith column of U. In other words, U_ℓ is the first ℓ columns of U. In this problem, we will go about showing that

$$U_\ell \in \underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \vec{x}_i - WW^\top \vec{x}_i \right\|^2 \tag{15}$$

where $W^\top W = I_\ell$ (i.e., it is a matrix with orthonormal columns). This is an important result for deriving PCA, as you will see in lecture.

(a) **First, show that**

$$\left\| \vec{x}_i - WW^\top \vec{x}_i \right\|^2 = \|\vec{x}_i\|^2 - \left\| W^\top \vec{x}_i \right\|^2 \tag{16}$$

(HINT: Expand the left hand side of the equation above using transposes. That is, use the fact that $\|\vec{v}\|^2 = \vec{v}^\top \vec{v}$.)

(b) Using the result from the previous part, we can simplify the original optimization problem to say

$$\underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \vec{x}_i - WW^\top \vec{x}_i \right\|^2 = \underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\|\vec{x}_i\|^2 - \left\| W^\top \vec{x}_i \right\|^2 \right) \tag{17}$$

$$\iff \underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(- \left\| W^\top \vec{x}_i \right\|^2 \right) \tag{18}$$

$$\iff \underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmax}} \sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 \tag{19}$$

where we get the second line from noticing that we cannot change $\vec{x}_i$, so we remove it from the optimization problem. Then, we pull out the negative to turn the minimization problem into a maximization problem. Now, let $W := \begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_\ell \end{bmatrix}$. **Show that**

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{k=1}^{\ell} \vec{w}_k^\top \left(XX^\top \right) \vec{w}_k \tag{20}$$

You may use the fact that $\sum_{i=1}^{n} \vec{x}_i \vec{x}_i^\top = XX^\top$.
(HINT: Start by expanding out the norm squared expression as the sum of squares of the individual entries of $W^\top \vec{x}_i$. Also, is the inner product commutative?)

(c) **Use the result of the previous part to show that**

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{k=1}^{\ell} \vec{\tilde{w}}_k^\top \Sigma\Sigma^\top \vec{\tilde{w}}_k \tag{21}$$

where $\vec{\tilde{w}}_k = U^\top \vec{w}_k$. Then, argue that $\Sigma\Sigma^\top$ can be written as

$$\Sigma\Sigma^\top = \begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \tag{22}$$

 All Rights Reserved. This may not be publicly shared without explicit permission.

where $r = \text{rank}(X)$
(HINT: Use the SVD of X to simplify the $XX^\top$ term from the previous part.)

(d) From the previous part, we have the following expression:

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{k=1}^{\ell} \vec{\tilde{w}}_k^\top \begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \vec{\tilde{w}}_k \tag{23}$$

One may show (via Cauchy-Schwarz) that

$$\sum_{k=1}^{\ell} \vec{\tilde{w}}_k^\top \begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \vec{\tilde{w}}_k \leq \sum_{k=1}^{\ell} \sigma_k^2 \tag{24}$$

if $\vec{\tilde{w}}_k$ are required to be orthonormal (you are not required to show this). **Using this fact, find some specific values of $\vec{\tilde{w}}_i$ such that we attain eq. (24) with equality. Then, use this to argue that U_ℓ maximizes $\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2$ and is thus a solution to the original optimization problem.**

Contributors:

- Anish Muthali.
- Daniel Abraham.
- Ashwin Vangipuram.
- Tanmay Gautam.
- Ryan Hoque.

© UCB EECS 16B, Spring 2024. All Rights Reserved. This may not be publicly shared without explicit permission.