

1 Frobenius Norm

In this problem we will investigate the basic properties of the Frobenius norm.

Much like the norm of a vector $\vec{x} \in \mathbb{R}^N$ is $\|x\| = \sqrt{\sum_{i=1}^N x_i^2}$, the Frobenius norm of a matrix $A \in \mathbb{R}^{N \times N}$ is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^N |A_{ij}|^2}.$$

This is basically the norm that comes from treating a matrix like a big vector filled with numbers. Note that matrices have other types of natural norms as well, like the induced norm.

a) With the above definitions, **show that**

$$\|A\|_F = \sqrt{\text{Tr}\{A^T A\}}.$$

Note: The trace of a matrix is the sum of its diagonal entries. For example, let $A \in \mathbb{R}^{N \times N}$, then,

$$\text{Tr}\{A\} = \sum_{i=1}^N A_{ii}$$

b) **Show that if U and V are orthonormal matrices, then**

$$\|UA\|_F = \|AV\|_F = \|A\|_F.$$

(HINT: You can proceed directly, or you can think about the relationship of the Frobenius norm of a matrix to the norms of each of its columns or rows. If you take the latter path, then the fact that orthonormal matrices don't change the norms of vectors might prove handy.)

c) **Use the SVD decomposition to show that $\|A\|_F = \sqrt{\sum_{i=1}^N \sigma_i^2}$, where $\sigma_1, \dots, \sigma_N$ are the singular values of A .**

(HINT: The previous part might be quite useful.)

2 Minimum Energy Control

Consider the system

$$\vec{x}(t+1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \vec{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t).$$

Our goal is to reach the target state

$$\vec{x}(5) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

starting at $\vec{x}(0) = 0$.

- a) Find the input sequence $u(0), u(1), u(2), u(3), u(4)$ that achieves this with the least possible "energy," as defined by

$$E = u(0)^2 + u(1)^2 + u(2)^2 + u(3)^2 + u(4)^2.$$

Find the value of E for the sequence you computed.

- b) Show that, if we reach the target state $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ at $t = 2$ and apply a zero input henceforth ($u(2) = u(3) = u(4) = \dots = 0$) then

$$\vec{x}(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad t = 2, 3, 4, 5, \dots$$

- c) With the result of part (b) in mind, find inputs $u(0), u(1)$ such that

$$\vec{x}(2) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and apply $u(2) = u(3) = u(4) = 0$ so

$$\vec{x}(5) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

as well. Find the energy E for the resulting input sequence and compare it to the one in part (a).

3 PCA

We collect temperature ($^{\circ}\text{F}$) and humidity (%) data each day for four days. We can organize our data into 4×2 matrix A , where the each column of A is a type of data (that is, temperature and humidity) and each row is a particular day's measurement.

$$A = \begin{bmatrix} 64 & 39 \\ 66 & 45 \\ 70 & 41 \\ 64 & 39 \end{bmatrix}$$

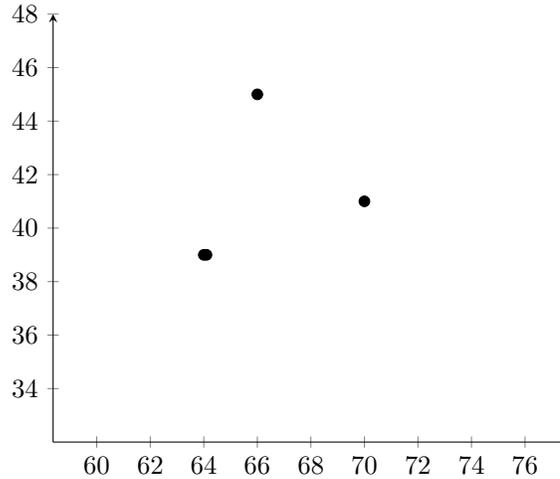
We're going to work through the principal component analysis of A to illustrate how we can analyze the data that A represents.

- a) **Find the covariance matrix of A .** We define the covariance matrix, C , of A to be $\frac{1}{m-1}\tilde{A}^T\tilde{A}$. The diagonal terms of C represent the variance of each measurement type, while the off-diagonal terms represent the covariance between each different measurement type. Remember that you must subtract the mean of each column of A in order to mean-center the data before finding C .

Since C is a symmetric matrix, we can eigendecompose it into the form $C = P\Lambda P^T$, where P has columns that contain the eigenvectors of C , and Λ is a diagonal matrix with the eigenvalues of C along the diagonal. The eigenvectors of C are equivalent to the principal components of A , for which C is the covariance matrix.

These eigenvalues along the diagonal of Λ are equivalent to the squared weights of the corresponding principal components of A in the columns of P .

- b) **Find the eigenvalues of C and order them from largest to smallest, $\lambda_1 > \lambda_2$.**
- c) **Find the orthonormal eigenvectors \vec{p}_i of C .** All eigenvectors are mutually orthogonal, and can be normalized to have unit length.
- d) Given what you have found above, **what are the principal components (direction along which the variance of the data is maximized) of the matrix A ? What is the principal component matrix, P ?**
- e) **What are the weights of each principal component?** These weights relate to the variance of the data in A along each principal component and indicate the relative strength of each component as a representation of the data.
- f) We have plotted the data points from A in the graph below, with column 1 data on the x axis and column 2 data on the y axis. **Plot the two principal components scaled by their weights on the following graph.** Remember that we subtracted the column means from each column before doing other computations. Therefore, the vectors you draw should originate from the mean. Also, in case you were wondering, they visually look like there are three points because two of them are the same — so they land on top of each other.



4 Netflix recommendation using SVD

On Maruf's recommendation, the EECS16B TAs hang out all the time outside of work. Every Friday night, we watch movies on Netflix and collect ratings for all the movies we've watched. We give star ratings (between 1 and 5 stars) for each of the movies we've watched. These data are saved in the file `data_TAs.csv`. Professors Arcak and Sanders sometimes crash movie night, and when they do we also collect their ratings. These data are saved in `data_arcak.json` and `data_sanders.json`.

In this problem, we will use SVD to build a system that will predict ratings for unrated movies based on a small sample of rated movies. This will allow us to make customized movie recommendations for the professors, like Netflix does for its viewers. Before starting, we recommend you review the lecture notes to see how to interpret the U , S , and V matrices of the SVD for this problem.

- a) Refer to the iPython notebook `recommender_system.ipynb` to complete this question.