Math 285 Homework 1, SJSU, Fall 2015. (Due: Tuesday, 9/22, in class)

(1) Find, by hand, the economic SVD of the following matrix

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

What are the different norms (Frobenius, Spectral and Nuclear) of this matrix?

(2) Now for the matrix in Question 1, use MATLAB to find the full SVD. Submit both your script and the results.

(3) Find the best-fit line (under the orthogonal error criterion) to the points in Question 1 (i.e., the rows of $\mathbf{A}$) and plot it with the data (by hand or computer). What are the coordinates of the projections of the original data onto the best-fit line? Find also the principle components of the data. What do they mean?

(4) Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square, invertible matrix and the SVD is

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

Show that the inverse of $\mathbf{A}$ is

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

(5) Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a matrix and $\vec{\sigma}$ the vector of its singular values. Which of the following norms are equal to each other? Draw a line between those that are equal.

$\|\mathbf{A}\|_F$	$\|\vec{\sigma}\|_\infty$
$\|\mathbf{A}\|_2$	$\|\vec{\sigma}\|_1$
$\|\mathbf{A}\|_*$	$\|\vec{\sigma}\|_2$

(6) First show that the product of two orthogonal matrices (of the same size) is also an orthogonal matrix. Then use this fact to show that

(a) If $\mathbf{L} \in \mathbb{R}^{m\times m}$ is orthogonal and $\mathbf{A} \in \mathbb{R}^{m\times n}$ is arbitrary, then the product $\mathbf{LA}$ has the same singular values and right singular vectors with $\mathbf{A}$.

(b) If $\mathbf{A} \in \mathbb{R}^{m\times n}$ is arbitrary and $\mathbf{R} \in \mathbb{R}^{n\times n}$ is orthogonal, then the product $\mathbf{AR}$ has the same singular values and left singular vectors with $\mathbf{A}$.

Note that an immediate consequence of the above results is that

$$\|\mathbf{LA}\| = \|\mathbf{A}\| = \|\mathbf{AR}\|$$

regardless of which norm (Frobenius/spectral/nuclear) is used. You don't need to prove this part.

(7) Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be any matrix. Denote $r = \text{rank}(\mathbf{A})$. Show that

(a) $\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_F \le \sqrt{r}\|\mathbf{A}\|_2$

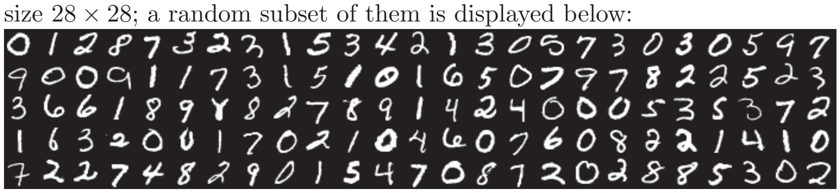
(b) $\|\mathbf{A}\|_F \le \|\mathbf{A}\|_*$

In fact, it is also true that $\|\mathbf{A}\|_* \le \sqrt{r}\|\mathbf{A}\|_F$, *but the proof requires using Cauchy-Schwarz Inequality:*

$$\left(\sum a_i b_i\right)^2 \le \left(\sum a_i^2\right)\left(\sum b_i^2\right).$$

You don't need to prove this part.

(8) The MNIST database contains 70,000 images of handwritten digits (i.e., 0,1,...,9) collected from about 250 writers. The images all have the same size 28×28; a random subset of them is displayed below:



More details about this dataset can be found at `http://yann.lecun.com/exdb/mnist/`.

In this homework we focus on the handwritten digit 1 in the training set; there are still 6,742 of them. We are going to convert each of these images to a 784 dimensional vector so that our data can be stored in a $6,742 \times 784$ matrix (see *mnist_digit1.mat*). Now you are asked to use MATLAB to perform principal component analysis on such data.

Specifically, you need to show

- The center of the handwritten 1's as an image of size 28×28 (this is how the "average" writer writes the digit 1)
- The first 50 singular values and their explained variances (this can help select k)
- The top 20 principal directions (i.e., right singular vectors) as images of size 28×28
- The top two principal components of the data (in order to visualize the data)

Include your MATLAB script with your submission.

HW1 answers.

(1) We start by calculating

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 14 & 6 \\ 6 & 14 \end{pmatrix}.$$

Its eigenvalues are $\lambda_1 = 20, \lambda_2 = 8$ with associated eigenvectors

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,1)^T, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}}(1,-1)^T.$$

From this, we can already get the singular vaues $\sigma_1 = \sqrt{20}, \sigma_2 = \sqrt{8}$ and corresponding left singular vectors of $\mathbf{A}$:

$$\mathbf{u}_1 = \frac{1}{\sigma_1}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{10}}(1,1,2,2)^T, \quad \mathbf{u}_2 = \frac{1}{\sigma_2}\mathbf{A}\mathbf{v}_2 = \frac{1}{2}(-1,1,-1,1)^T.$$

In sum, the economic SVD of $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{1}{2} \\ \frac{1}{\sqrt{10}} & \frac{1}{2} \\ \frac{2}{\sqrt{10}} & -\frac{1}{2} \\ \frac{2}{\sqrt{10}} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \sqrt{20} & \\ & \sqrt{8} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}^T$$

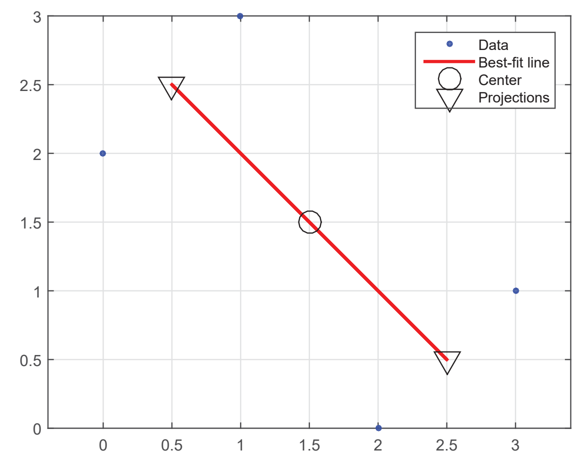
The different norms are $\sqrt{28}$ (Frobenius), $\sqrt{20}$ (Spectral) and $\sqrt{20} + \sqrt{8}$ (Nuclear).

(2) Type in Matlab command window

$$A = [0,2;2,0;1,3;3,1]$$
$$[U,S,V] = svd(A)$$

and you will be able to see the full SVD.

(3) See figure below for best-fit line and projections of the original data:



These are obtained from the SVD of **centered data**: $\widetilde{\mathbf{X}} = \mathbf{U}\Sigma\mathbf{V}^T$, or

$$\begin{pmatrix} -\frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{3}{2} \\ -\frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \sqrt{8} & \\ & \sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T$$

The top 2 principal components of the points in $\mathbf{X}$ of Question 1 (following the same order) are the rows of

$$\mathbf{U}\Sigma = \begin{pmatrix} -\sqrt{2} & -\frac{\sqrt{2}}{2} \\ \sqrt{2} & -\frac{\sqrt{2}}{2} \\ -\sqrt{2} & \frac{\sqrt{2}}{2} \\ \sqrt{2} & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

They represent the coordinates of the original data (i.e., the rows of $\mathbf{X}$ in Question 1) relative to the new coordinate axes along the principal directions $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,-1)^T$, $\mathbf{v}_2 = \frac{1}{\sqrt{2}}(1,1)^T$, with origin placed at the center of the original data $(\frac{3}{2}, \frac{3}{2})$.

(4) Proof: Let $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ be the full SVD. Since $\mathbf{A}$ is invertible, Σ is also invertible. It follows that

$$\mathbf{A}^{-1} = (\mathbf{U}\Sigma\mathbf{V}^T)^{-1} = \mathbf{V}\Sigma^{-1}\mathbf{U}^T = \sum_{i=1}^{n} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^T.$$

(5) $\|\mathbf{A}\|_F = \|\vec{\sigma}\|_2, \quad \|\mathbf{A}\|_2 = \|\vec{\sigma}\|_\infty, \quad \|\mathbf{A}\|_* = \|\vec{\sigma}\|_1$

(6) Let $\mathbf{Q}_1, \mathbf{Q}_2$ be two orthogonal matrices of the same size. Then

$$(\mathbf{Q}_1\mathbf{Q}_2)^{-1} = \mathbf{Q}_2^T\mathbf{Q}_1^T = (\mathbf{Q}_1\mathbf{Q}_2)^T.$$

This shows that $\mathbf{Q}_1\mathbf{Q}_2$ is also an orthogonal matrix. To prove part (b), assume $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ is the full SVD. Then $\mathbf{LA} = (\mathbf{LU})\Sigma\mathbf{V}^T$. This represents the SVD of $\mathbf{LA}$ because $\mathbf{LU}$ as a product of two orthogonal matrices is also orthogonal. From this we see that $\mathbf{LA}$ has the same singular values and right singular vectors with $\mathbf{A}$. Part (b) is proved similarly.

(7) Proof: Express all the matrix norms in terms of vector norms, using the results of Question 5. It will be straightforward to prove the inequalities.

(8) Omitted.