Homework 6

Due date: April 17, 2017, 11:55PM.

Bibliography: Course lecture notes Lessons 23-26.

Note: Save this file as `Homework6Solution.tm` in your `/home/student/courses/MATH547/homework` directory. Then carry out your work on solving the problems in TeXmacs. Submit your final version through Sakai. If you've understood the theoretical concepts reviewed in the Background section and lecture notes, each problem should require about 20 minutes to complete, hence the entire homework can be completed within 3 hours.

1 Background

This assignment investigates the multiple uses of the singular value decomposition of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times m}$, $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, with , $\boldsymbol{U} \in \mathbb{R}^{m \times m}$, $\boldsymbol{V} \in \mathbb{R}^{n \times n}$, $\boldsymbol{\Sigma} \in \mathbb{R}_+^{m \times n}$, $\boldsymbol{U}$, $\boldsymbol{V}$ orthogonal matrices $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_m$, $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_n$, $\boldsymbol{\Sigma}$ a diagonal matrix. If $r = \operatorname{rank}(\boldsymbol{A})$, it is often the case that $r \ll m$, and it is more economical to work with the reduced SVD defined as $\boldsymbol{A} = \hat{\boldsymbol{U}}\boldsymbol{\Sigma}\hat{\boldsymbol{V}}^T$, with $\hat{\boldsymbol{U}} \in \mathbb{R}^{m \times r}$, $\boldsymbol{\Sigma} \in \mathbb{R}_+^{r \times r}$, $\hat{\boldsymbol{V}} \in \mathbb{R}^{n \times r}$.

2 Theory problems

2.1 Symmetric matrices

The SVD of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is constructed from eigenproblems of the normal matrices $\boldsymbol{A}\boldsymbol{A}^T \in \mathbb{R}^{m \times m}$ and $\boldsymbol{A}^T\boldsymbol{A} \in \mathbb{R}^{n \times n}$. It is useful to become familiar with properties of symmetric matrices to understand the SVD.

Ex8.1.16 Solution.

Ex8.1.24 Solution.

2.2 Geometry of the SVD for 2×2 matrices

Ex8.3.1 Solution.

Ex8.3.2 Solution.

Ex8.3.3 Solution.

Ex8.3.4 Solution.

2.3 SVD solution of least squares problems

Both the $\boldsymbol{Q}\boldsymbol{R}$ and the SVD decompositions can be used to solve the least squares problem: given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ find $\boldsymbol{x} \in \mathbb{R}^n$ the minimizes the norm of the difference between $\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{b}$

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|.$$

QR solution.

1. Compute decomposition $\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{A}$
2. The desired best approximation of $\boldsymbol{b}$ is the projection onto $C(\boldsymbol{A})$, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{b}$ with $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^T$, $(\boldsymbol{Q}\boldsymbol{R})\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{b}$, hence $\boldsymbol{x}$ is found as the solution of the system $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^T\boldsymbol{b}$

SVD solution.

1. Compute decomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{A}$

2. The projector onto $C(\boldsymbol{A})$ is $\boldsymbol{P}=\boldsymbol{U}\boldsymbol{U}^T$, and $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{P}\boldsymbol{b}$, leads to $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}=\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{b}$. Introduce notation $\boldsymbol{y}=\boldsymbol{V}^T\boldsymbol{x}$. Solve system $\boldsymbol{\Sigma}\boldsymbol{y}=\boldsymbol{U}^T\boldsymbol{b}$
3. Solution is $\boldsymbol{x}=\boldsymbol{V}\boldsymbol{y}$

Ex8.3.17 Solution.

Ex8.3.18 Solution.

2.4 Properties of the SVD

Ex8.3.32 Solution.

Ex8.3.33 Solution.

Ex8.2.34 Solution.

3 Computational problems

The previous assignment and this one illustrate the applications of eigenvalues and the singular value decomposition in the face recognition problem. The following instructions load a matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, containing $n=99$ columns, where adding the average $\boldsymbol{a}\in\mathbb{R}^m$ to columns of $\boldsymbol{A}$ gives an image with $m=2^{14}=16384$ pixels. Also loaded are matrices $\boldsymbol{R},\boldsymbol{S}\in\mathbb{R}^{n\times p}$ of $n=99$ coefficients of linear combinations of columns of $\boldsymbol{A}$ to give $p=2000$ face images.

```
octave> cd /home/student/courses/MATH547/homework/
octave> load ../lessons/mitfaces/faces.mat
octave> who
   Variables in the current scope:

   A         R         S         a         ans       dispans   prompt    r

octave> [size(A) size(a) size(R) size(S)]
   ans =

      16384      99   16384       1      99    2000      99    2000

octave> [m,n]=size(A); [n,p]=size(R);
octave> [m n p]
   ans =

      16384      99    2000

octave> norm(a)
   ans =  7063.6
octave>
```

A vector of length m, say the average face $\boldsymbol{a}$, is displayed through the instructions

```
octave> imagesc(flip(reshape(a,128,128)',1)); colormap(gray(256)); print -dpng aface.png
```

```
octave>
```

The reduced SVD of $\boldsymbol{A}$ is computed as

```
octave> [U,S,V]=svd(A,'0');
```

The columns of $\hat{\boldsymbol{U}}$ represent the deviations of the faces within $\boldsymbol{A}$ from the average face $\boldsymbol{a}$, ordered such that $\boldsymbol{u}_1$ is the most important deviation, $\boldsymbol{u}_2$ the next most important, and so on. Here are the changes of the verage face by displacement along the first singular vector by distances $10^{-k}\boldsymbol{u}_1$ for $k=3,2,1$.

```
octave> D=norm(a)*U(:,1)*10.^(-3:-1); F=a*[1 1 1]+D;
octave> imagesc(flip(reshape(F(:,1),128,128)',1)); colormap(gray(256)); print -dpng
        svdfm3.png
octave> imagesc(flip(reshape(F(:,2),128,128)',1)); colormap(gray(256)); print -dpng
        svdfm2.png
octave> imagesc(flip(reshape(F(:,3),128,128)',1)); colormap(gray(256)); print -dpng
        svdfm1.png
octave>
```

Figure 1. Changes in the average face along the dominant singlar vector

Consider now the problem of facial recognition or classification. Consider some face $\boldsymbol{f}\in\mathbb{R}^m$. The components of $\boldsymbol{f}$ are given in the basis $\boldsymbol{I}$, $\boldsymbol{f}=\boldsymbol{I}\boldsymbol{f}$. The standard basis $\boldsymbol{I}$ is however inefficient in this context (as in many others!). It is better to transform to coordinates in the basis $\boldsymbol{U}$ by solving the problem $\boldsymbol{U}\boldsymbol{x}=\boldsymbol{I}\boldsymbol{f}$. The coordinates of the face in this new basis are components of $\boldsymbol{x}=\boldsymbol{U}^T\boldsymbol{f}$. Since the vectors within $\boldsymbol{U}$ are ordered, a very good approximation is found by $\tilde{\boldsymbol{f}}=\tilde{\boldsymbol{U}}\,\tilde{\boldsymbol{x}}$, with $\tilde{\boldsymbol{U}}\in\mathbb{R}^{m\times s}$, $\tilde{\boldsymbol{x}}\in\mathbb{R}^s$, taking the first s columns of $\boldsymbol{U}$, with $s\ll m$. This forms the basis of facial recognition. The first s coordinates of the face $\boldsymbol{f}$ in the basis $\boldsymbol{U}$ are a compact representation, typically with $s=\mathcal{O}(10)$, many fewer components thn $m=16384$.

Choose some face $\boldsymbol{f}$ from within the database.

1. Compute the reduced SVDs of $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, $\boldsymbol{M}=\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^T$ from Homework 5.
2. Compute the coordinates $\boldsymbol{x},\boldsymbol{y}$ of the face on the bases $\boldsymbol{U},\boldsymbol{U}_1$ respectively.
3. How many components of $\boldsymbol{x},\boldsymbol{y}$ are required to give an approximation of the face to 99% certainty?
4. Is the chosen face $\boldsymbol{f}$ within those chosen in the construction of $\boldsymbol{A},\boldsymbol{M}$.

The same note as in Homework 5 applies here also: the above questions can be answered using very few lines of Octave code, no more that 5 lines per question. The main effort is properly conceptualizing the linear algebra operations, and thinking in terms of vectors and matrix operations. Pay particular attention to the dimension of the result from every matrix multiplication.