M435 Projects in Applied Mathematics

Singular Value Decomposition Worksheet

Monday, September 15, 2003

This worksheet is to prepare students for a project in the application of the singular value decomposition. The problems in theory reinforce the material derived in class while the computing section develops the code that will be applied in the pattern recognition project. All of your answers to this worksheet should be included in the **Description of the Method** section in your project report on the SVD.

The application concerns the low-rank approximation of a digital image of the planet Saturn. This image will be loaded into a data matrix X.

1 Theory

1. Show that the non-zero singular values of the matrix X are the same as the positive square roots of the eigenvalues of XX^T and X^TX.

2. Describe the meaning of a plot of the eigenvalues of XX^T in decreasing order, i.e., largest to smallest.

3. Explain the meaning of zero singular values. What is the largest number of non-zero singular values possible for a $P \times N$ data matrix X? (Consider the decomposition $X = UDV^T$.

4. Show that the left (and right) singular vectors may be viewed as data-dependent by expressing them as sums of the columns (rows) of the data matrix X.

2 Computing

The Saturn image for this problem may be downloaded from the class web page. The format of the data is "uint8", which stands for unsigned integer, 8 bits. Before you apply the SVD, change it to "double" format:

```
A=imread('saturn.tif')
A=double(A);
```

To display the image use

```
imagesc(A);
colormap(gray);
```

Use the singular value decomposition to construct a low-rank approximation to Saturn.

```
[U,D,V] = svd(X);
```

1. Plot the squares of the singular values in decreasing order. Based on this graph can you comment on how many terms you might expect to need to accurately reconstruct Saturn?

2. Plot the rank one images for several different eigenvalues (large, medium, and small) and compare.

3. You may observe that the rings of Saturn are poorly approximated at the edges. Identify the left and right singular vectors that are most important (based on maximum mean squared projection) for "fitting" the rings and modify your approximation appropriately. This is an example of *feature selection* in that you are selecting the best modes to represent your data.

4. Write a subroutine to automatically identify whether a column of the picture contains only rings or both planet and rings. This is a two class classification problem. This algorithm is of the same general type we will use for classification of the EEG data.