Coding Project 3: Principal Component Analysis of a Mass-on-a-spring System

Lucas Liu

Abstract

This project focused on using Principal Component Analysis to approximate the motion of a Mass-on-a-spring System. Data was collected from the manual location of points from videos recorded of the motion from different perspectives. After being reshaped to the correct dimensions, the matrices were joined and the Singular Value Decomposition of the matrix was taken. The energies of the singular values was then examined to determine the best rank-n approximation to be used.

1 Introduction

Princiapal Compont Analysis (PCA) is a technique that transforms a large number of potentially correlated variables to a small number of uncorrelated variables. PCA has found applications in various fields, including computer vision, signal processing, genetics, finance, and social sciences. In computer vision, PCA is used for facial recognition, image compression, and object recognition. The widespread use of PCA reflects its utility in dealing with high-dimensional datasets and its ability to provide insights into complex systems.

The mass-on-spring system is a commonly used example of a simple harmonic oscillator that can by described by a second order linear ordinary differential equation. Stemming from Newton's second law and Hooke's law, the derivation of the ODE is as follows:

$$F = ma = m\frac{d^2x}{dt^2} = mx''$$

$$F_{spring} = -kx$$

$$F_{friction} = -b\frac{dx}{dt}$$

$$F_{total} = mx'' + bx' + kx$$

The remainder of this report will outline the background and process of applying the PCA to estimate the motion of a mass-on-spring from data retrieved from the manual locating of points from videos from different perspectives. The results and analysis of the findings will be included afterwards.

2 Theoretical Background

This section covers a brief overview of the analytic concepts and techniques used in the project.

2.1 Singular Value Decomposition

The Singular Value Decomposition (SVD) is a matrix decomposition technique that decomposes a matrix into three matrices.

$$M = U \times \Sigma \times V$$

The first matrix in the SVD, U, represents the orthogonal basis vectors that make up the original matrix M's column space. The second matrix Σ includes the singular values decreasing in size along it's diagonal. Singular values represent the importance of the dimensions in the input matrix and can be used to reduce the dimensionality of the data. The third matrix V represents orthogonal basis vectors that span the original matrix M's row space.

2.2 Rank-n Approximation

Upon obtaining the singular values of an input matrix M through the SVD, one can approximate the input matrix using the singular values. As the singular values are position in decreasing order along the diagonal of the Σ, the first singular value represents the most important basis and largely captures the most important features in the input matrix.

A rank-1 approximation only uses the first singular value a rank-2 approximation uses the first two singular values, and so on.

2.3 Principal Component Analysis

The analysis of higher order datasets can be exceptionally challenging. Principal Component Analysis (PCA), is a technique that transforms the data into a coordinate system that can describe the variation in data with less dimensions.

The initial step of PCA is to standardize the data. This can be done by subtracting the mean from all components in the data set. Following this, the covariance matrix ,$\tilde{X}$ of data set X is calculated:

$$\tilde{X} = \frac{1}{n-1} X X^T$$

The covariance matrix tells us the variance between different variables in the data set. Afterwards, the SVD of the covariance matrix can be taken to determine the principal components - linear combinations of the original variables that capture the most variance in the dataset

3 Numerical Methods

Below is the code used in the project separated and explained in snippets.

```
load('Xt1_1.mat');
load('Xt2_1.mat');
load('Xt3_1.mat');

Xt2_1 = Xt2_1(:,1:226);
Xt3_1 = Xt3_1(:,1:226);
```

The code above loads in the files and resizes them so that dimension errors do not occur when conducting calculations.

```
   means1 = mean(Xt1_1(1,:));
means2 = mean(Xt1_1(2,:));
Xt1_1(1,:)  = Xt1_1(1,:)  - repmat(means1,1,226);
Xt1_1(2,:)  = Xt1_1(2,:)  - repmat(means2,1,226);

means1 = mean(Xt2_1(1,:));
means2 = mean(Xt2_1(2,:));
Xt2_1(1,:)  = Xt2_1(1,:)  - repmat(means1,1,226);
Xt2_1(2,:)  = Xt2_1(2,:)  - repmat(means2,1,226);


means1 = mean(Xt3_1(1,:));
means2 = mean(Xt3_1(2,:));
Xt3_1(1,:)  = Xt3_1(1,:)  - repmat(means1,1,226);
Xt3_1(2,:)  = Xt3_1(2,:)  - repmat(means2,1,226);
```

Next, the data is regularized by subtracting the mean from all components.

```
   CombMtrx = [Xt1_1; Xt2_1; Xt3_1]; [U,S,V] = svd(CombMtrx);
Y = U*CombMtrx;
   A1 = Y;
```

Next, the three individaul data sets are combined to one data set. The SVD is then taken of the combined data set and a coordinate transformation is done on the principal components. This matrix is saved as A1.

```
   A2 = zeros(6,1);
sig = diag(S);
A2(1,1) = sig(1).2 div( sum(sig.2));
A2(2,1) = sum(sig(1:2).2)/sum(sig.2);
A2(3,1) = sum(sig(1:3).2)/sum(sig.2);
A2(4,1) = sum(sig(1:4).2)/sum(sig.2);
A2(5,1) = sum(sig(1:5).2)/sum(sig.2);
A2(6,1) = sum(sig(1:5).2)/sum(sig.2);
```

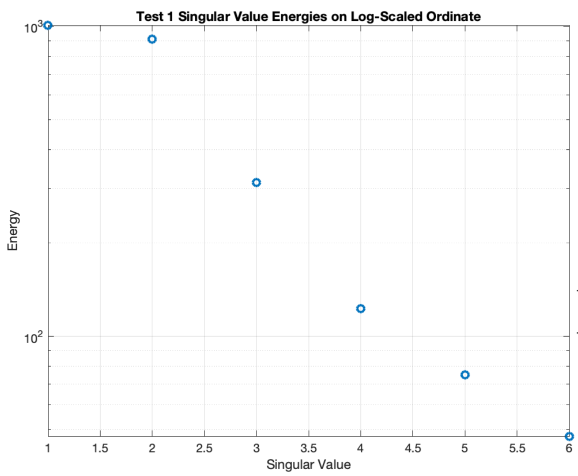
The energies of the singular values are calculated and saved in matrix A2. The energies are calculated by the sum of all the energies leading up to the current energy divided by the total sum of all energies.

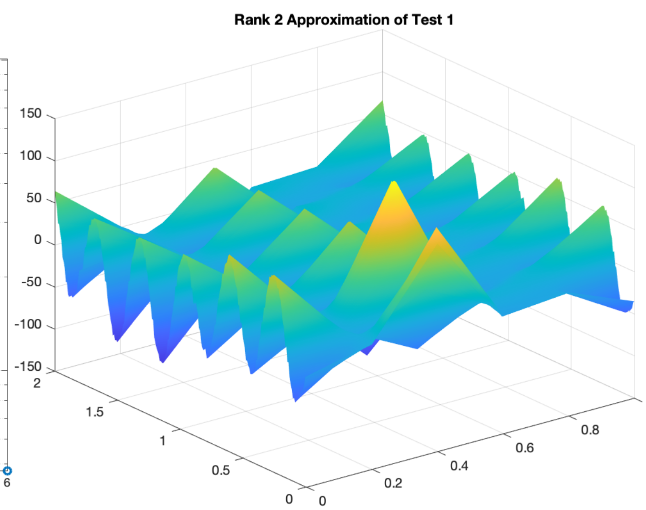
```
A3 = U(:,1)*S(1,1)*V(:,1)' + U(:,2)*S(2,2)*V(:,2)';
```

Finally, the rank 2 approximation of the data sets is created and saved in variable A3. The rank of the approximation is determined by looking at the principal components and their magnitudes found in the previous step. This process was done for four different sets of three input files.

4 Results

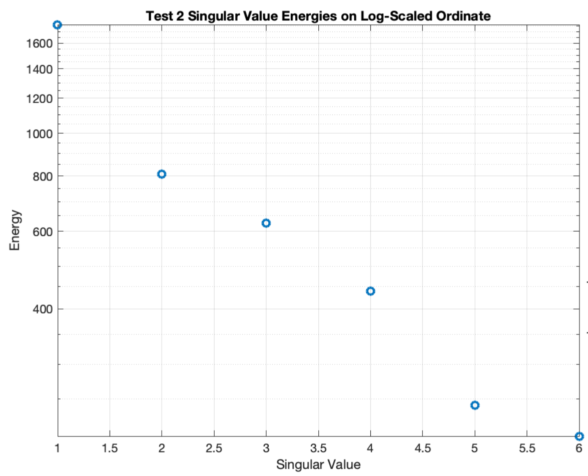
4.1 Test 1

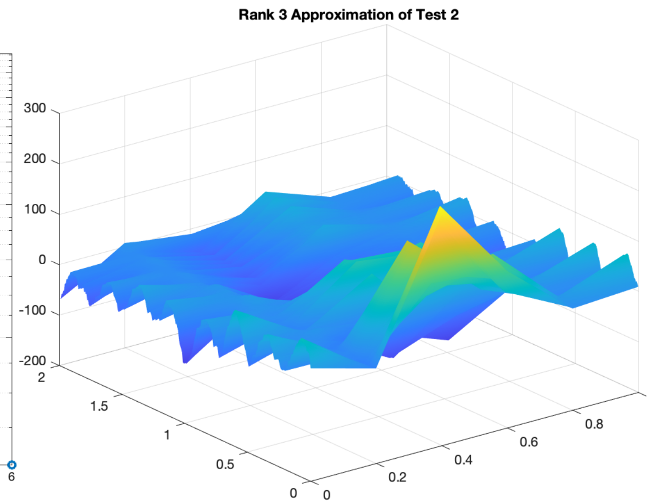




For Test 1, a log-scaled Ordinate plot of the singular value energies of the SVD of the data matrix reveals very high energies for the first two singular values, followed by significant drop offs in the following singular values. As such, this implies that most of the variance in the data can be captured in the first two principle components. Thus, a rank 2 approximation was used and is shown on the right.

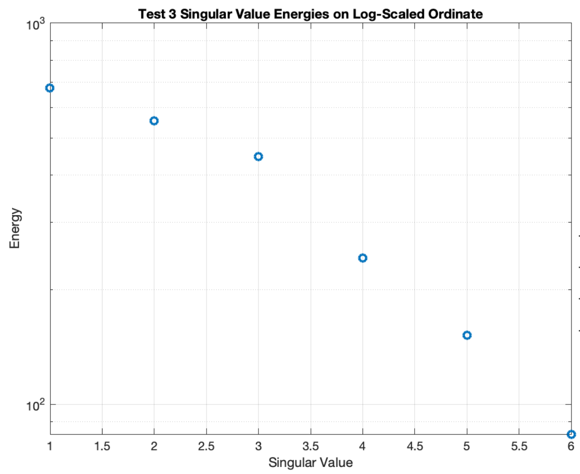
4.2 Test 2





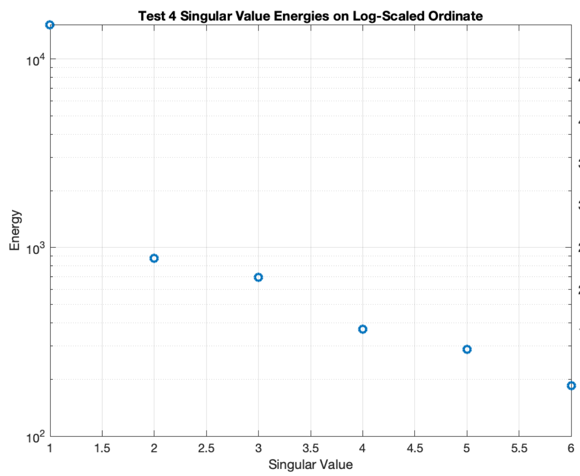
Unlike the energies of the SVD of the values in Test 1, the first singular value of Test 2 is very high, but there is a significant drop off between the first and second singular values. Cosequently, not as much of the information in the data can be kept with just two principal components. This is why a rank 3 approximation was taken instead. Contrary to only using two of the principal components in the approximation of Test 1, three principal components are needed in the approximation of Test 2 (shown above on the right).

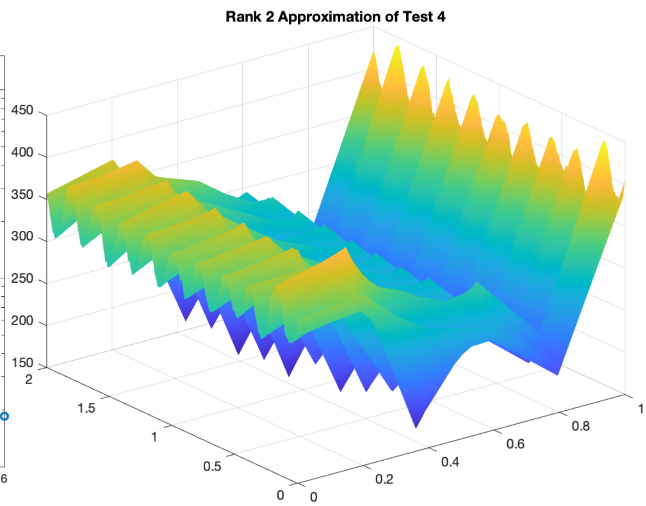
4.3 Test 3



In Test 3, the energy of the first singular value is significantly less than that of the first three tests. However, the second and third singular values have higher energies. Consequently, while a rank 2 approximation may not be sufficient in the approximation of Test 3, a rank 3 approximation can confidently be said to capture most of the variance in the data and to provide a good approximation.

4.4 Test 4





In Test 4, the first singular value is of very large magnitude, however, there is a significant decrease in the energies of the remaining singular values. For instance, the energy of the second singular value is less than half of that of the first singular value. This implies that the most of the important information can be captured in the first principal component. Consequently, a rank 2 approximation was chosen to approximate the motion of objects in Test 4.

5 Conclusion

In this project, principal component analysis was used to approximate the motion of a mass-on-a-spring system in four different tests each with varying perspectives and disturbances. By applying the singular value decomposition to a matrix containing values representing the three-dimensional location of the object's center of mass, we are able to create a matrix of basis that captures the information of the data. By analyzing the energies of the singular values, we can deduce the rank our approximation should take to best estimate the motion of the mass.

Acknowledgment

The sample code from the class was referenced in the project.