Principle Component Analysis: Yale Eigen Faces

Zhaokai Wu, Department of Mechanical Engineering, University of Washington

Abstract

Singular value decomposition (SVD) is one of the most power technique in data analysis. SVD is a universal tool in feature extraction by reducing the sample space into a feature space that can be represented using very few feature vectors, namely principle components. The importance of each principle components can be determined through mode analysis, a technique using singular values of a factor of data matrix. In this project, we explore the potentials of SVD on Yale cropped face images. We extracted six principle components, also referred as eigen faces, from 2414 pmg images. The result of our mode analysis suggests that the first three eigen faces condenses most of the data features, but the other three are also important.

Sec. I. Introduction and Overview

In data analysis, large scale data is often collected and stored in a matrix form. In real world application, theses data are complex. The goal of data analysis is to extract meaningful information and turns it into insights. For instance, we could extract the Newton Second Law by collecting digital data from cameras at different position filming a mass-spring system. Obviously, the massive image data embedded in a data matrix captures the oscillation movements. However, the redundancies and noises entangled in the data sets are obtrusive. Starring at an unreasonably large chunk of matrix can get us nowhere near having the same euphoric appreciation on the elegantly compact formula

$$F = m \cdot a$$

Fortunately, in the era of data there are techniques so elegant too that we can use on these data matrix and produce undiscovered knowledge that never be mined by a human brain. It takes a leap of imagination and some linear algebra knowledge to think about these data lives in an abstract space, namely the sample space, where there exists some fundamental molecule that by freely combining(addition) or scaling (scalar multiplication) can form every piece of initial unprocessed data in the matrix.

One of the most powerful tool in data analysis is called singular value decomposition (SVD). It guarantees to condense any form of data matrix by transforming the sample space into a feature space which normally can be described by very few elements. They are also called the bases of the feature space. The process of this condensation of key information is often referred as dimension reduction. Among bases of feature space, there are always some more dominant than others. Usually, the dominant bases are the attributes we as data analysts aim to mine. When data collection is done correctly, the insignificant bases are often subtle and contains noise information. We use technique called mode analysis to decide to keep or truncate certain molecule.

This project is applied SVD on face image stack. Modern digital images are recorded in pixel intensity and therefore can be quantified. The immediate result is that large sets of images can be arranged in data matrix and analyzed. The share power of computer allows us to reduce the dimension of the data and find the patterns across the data set. The following write-up includes two major block: theoretical background around the theorem of SVD, algorithm in analyzing the data and finally discussion on the results of SVD analysis.

<u>Sec. II. Theoretical Background</u>

1. <u>Statement of the Theorem</u>

For arbitrary matrix M with shape $m \times n$ from filed $\mathbb{F}$, where $\mathbb{F}$ represents either real or complex field, there always exists a factorization, called a singular value decomposition of M in form

$$\mathrm{M} = U \Sigma \mathrm{V}^*$$

where U is $m \times m$ unitary matrix; Σ is a diagonal matrix shaped in $m \times n$, where the non-negative diagonal entries σ_i are referred as singular values of M; $V^*$is a $n \times n$ unitary matrix.

2. <u>Geometric Interpretation</u>

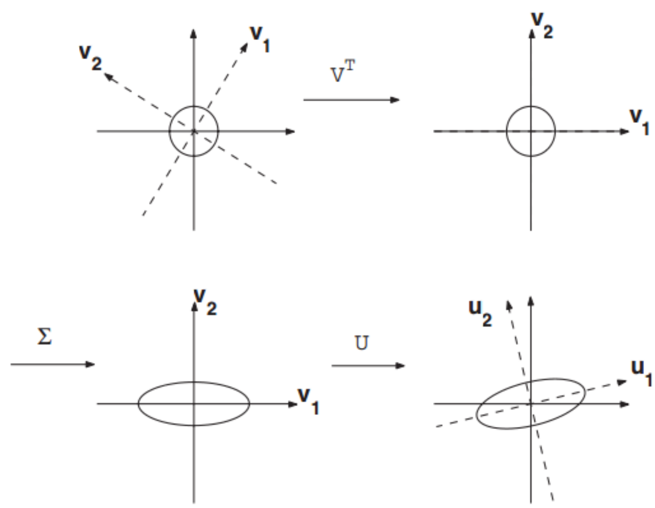

figure 1. Geometric meaning of SVD: The image of a circle under matrix multiplication is a ellipse

A high-level observation to make about the SVD theory is that any linear transformation can be viewed as a composition of rotation or reflection, a scaling, and another rotation or reflection. To illustrate this geometric interpretation, we will provide a demonstrates the SVD decomposition of a matrix multiplication on a 2D circle. Then, we claim that this geometric interpretation can be extended to any dimensions.

Considering the three factors of the SVD separately, note that V^* is a pure rotation of the circle. Figure 1 shows how the axes $v1, v2$ are first rotated by $V^*$to coincide with the coordinate axes. Secondly, the circle is stretched by Σ in the directions of the coordinate axis to form an ellipse. Lastly, the ellipse is rotated by U into its final position. Note how $v1, v2$ are rotated to end up as $u1, u2$, the principal axes of the final ellipse. A direct calculation shows that $Mv_j = \sigma_j u_j$. In other words, v_j is first rotated to coincide with the j coordinate axis, stretched by a factor σ_j , and then rotated to point in the direction of u_j.

It is not to illustrate this composition of transformation is also applicable in 3D space, where the object of examination is replaced with a unit circle. However, as human we are limited to construct our imagination around objects that's less or equal to three dimension. Therefore, it is only through the power of mathematics, particularly linear algebra, that we reached the conclusion that this decomposition is also true in higher dimension.

3. Singular Value and Mode Analysis

Singular values are contained in matrix Σ. It determines the magnitude of stretching in the transformed sample space V^*. In the example of a 2D unit circle, the largest singular value is the length of the longest principle axis. In high dimension data analysis, Σ stretches the bases of V^*. If M is normalized and zero centered with mean, singular values can be interpreted as the importance of each base vector. In other words, if some singular values are significantly smaller than others, their contribution to the characteristics of the data is minimal and often neglectable. The technique to distinguish the importance of the bases and selectively neglect unimportant information (often turns out to be noise) is a called mode analysis.

4. Relationship to Eigen Decomposition

SVD is a generalization of the eigen decomposition, which works only on positive semidefinite normal matrix. In this section, we will discuss their connections and differences respectively.

We note one connection between SVD and eigen decomposition is that the singular values equal to the square of eigen values when these the target matrix allows both decompositions. Since matrix M is not necessarily square which means that eigen decomposition cannot be done, we turn to examine the covariance data matrix

$$C_M = MM^*$$

where C_y is positively definite which means both decomposition can be applied on C_y. Suppose

$$MM^*S = S\Lambda$$

$$M = U\Sigma V^*$$

where S, Λ contains eigen vectors $S = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ and eigen values $\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$ respectively.

We first consider transferring the sample space into eigen sample space Y and consider the covariance matrix C_Y:

$$Y = S^*M$$

$$C_Y = \frac{1}{n-1}YY^* = \frac{1}{n-1}(S^*M)(S^*M)^* = \frac{1}{n-1}\Lambda$$

then, we perform similar operations with feature space F

$$F = U^*M$$

$$C_F = \frac{1}{n-1} FF^* = \frac{1}{n-1}(U^*M)(U^*M)^* = \frac{1}{n-1}\mathrm{U}^*XX^*U$$

where $XX^* = U\Sigma U^*$. Hence,

$$C_F = \frac{1}{n-1}\Sigma^2$$

Note that C_F, C_Y both examine the similarities of bases vector of data matrix M, we obtain that,

$$\Lambda = \Sigma^2$$

In words, the singular values of the data matrix equal to the square of the eigen values of the same matrix.

While related, the eigen decomposition and SVD differ except for positive semi-definite matrices. Firstly, eigen matrix S and eigen value matrix Λ is not necessarily unitary, while the U, Σ, V^* are guaranteed to be unitary.

Sec. III. Algorithm Implementation and Development

The coding tasks for this project are divided in two three categories: image loading, applying SVD and image reconstruction. To achieve

1. Image Loading

Our image analysis uses the cropped Yaleface database provided on USCD CS website. The data is consisted of 38 folders with each represent one person. Within each folder, there are around 64 images in pgm format of various shadings. Each pgm image has a uniform resolution of 192×168.

We would like to point out couple observations made while processing theses image. First, on the folder level, "yaleB14" folder is missing. Secondly, for image file, many folders contain files with extension "*.pgm.bad". These files are filtered out from the sample set.

The sample sets turns out to contain 2414 pgm files. A reader algorithm is developed to concentrate all information into a raw data matrix M of shape 2414×32256. Note that 32256 is the product of 192 and 168. Reorganizing the pixels of each image in a single row or column is a standard way in image process due to its compact form. With this set-up interpreting the data matrix M is easy. Each row represents an image. In addition, the images are loaded sequentially from folder "yaleB01" to "yaleB39".

2. Applying SVD

Due to the large size of the data matrix, we used a sped-up SVD algorithm included in SciPy package. The wall time of this computation is around 6-9 seconds. A test on using standard Numpy SVD method returns memory error.

After obtaining the SVD factor of the raw data matrix, we calculated the rank of the feature space represented by U using the "matrix_rank()" method provide in the linear algebra subpackage with in Numpy library.

3. <u>Image Reconstruction</u>

We practiced principle component analysis on the data matrix by mode analysis and generated eigen faces with different modes. Finally, we compared superposed faces consisting of various combination of principle directions.

As for mode analysis, we plotted the normalized singular values in percentage for all modes on a equally-spaced and log scale. The motivation is to decide if there are any principle directions are resulted from noise and therefore can be truncated.

For image reconstruction, we pulled out all six bases u_i of the feature space U and left multiple each with $\sigma_i V^*$. Note that doing so where result in 2124 images in each principle directions. Since we are not interested in recognizing, we average all the 2124×32256 matrix of each principle direction and condense it into a single 1×32256 matrix and finally reshaped back to image resolution for printing.

4. <u>Mode Analysis</u>

The objective to perform mode analysis is to decide if any principle directions are result from noise from data collection and thus subject to neglect. An empirical decision can be made based on examining the singular values and how they compare to each other in terms of magnitude.

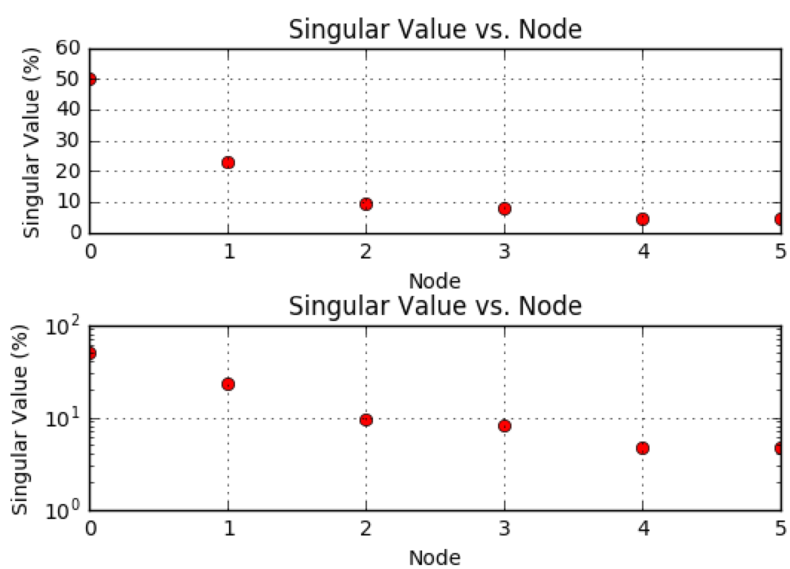

figure 2.Singular value vs. node

Figure 2 provided normed singular values plotted on equally-spaced and log scale. In our case, the slope of decrease is not steep nut rather mellowly dies out. Therefore, we generally have hesitation to drop out any of the modes during reconstruction. The intuition to justify this phenomenon is that the data we have doesn't include much noise, which matches our expectation on cropped data.

Sec. IV. Computational Results

1. Principle Mode: mode 1 only

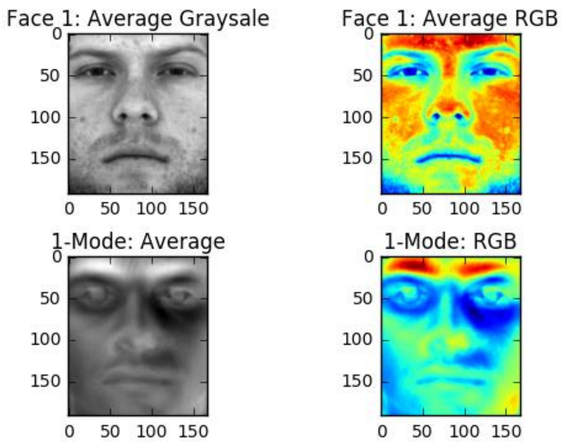

figure 3. Average face of person 1 in grey scale (top left); Average face of person 1 in RGB (top right); Eigen face 1 in grey scale (bottom left); Eigen face 1 in grey scale (bottom right)

From mode analysis, we calculated that there exists a dominant principle direction 1 the largest singular value contributes around 50% in terms of weight in describing the feature space. We would like to demonstrate the resemblance of this eigen face to our unprocessed images. We plotted figure 3 for a direct illustration. In figure 3, the top two images are the average face of the first person in the data set in grey scale and RGB. The bottom two images are eigen faces in principle direction number one in grey scale and RGB. Couple quick observations: first, this major eigen face looks very like the first person. This confirms our arguments that eigen face 1 is a dominant feature shared across the data set. Secondly, the obvious shading under the right eye of the eigen face may suggest that there exists an unbalanced composition of images with lighting from the left and right. In this case, we think there are more images with lighting from the left, which causes shading under right eye becomes a dominant feature.

2. Six Eigen Faces

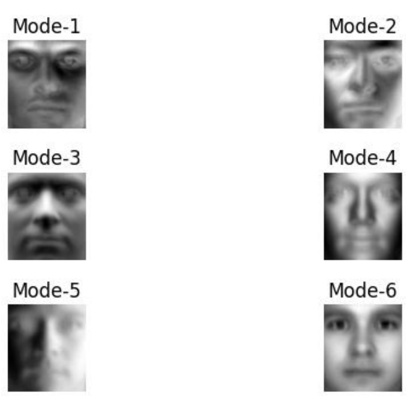

Figure 4. Six eigen faces number with corresponding singular value sorted in descending order

Figure 4 shows all six eigen faces. Note that eigen face 6 with the smallest singular value percentage (less than 5%) looks a lot like a normal image compared to the others. The intuition is that this principle direction is not representative across the entire sets. The underlying assumption is that the less representative feature accidently stimulates out brain in thinking that it looks like a person we see in daily life. In addition, the orthonormal nature of each principle direction suggests that it indeed extract certain characters of the data sets that are not picked up by any other direction.

3. <u>Superposition of faces</u>

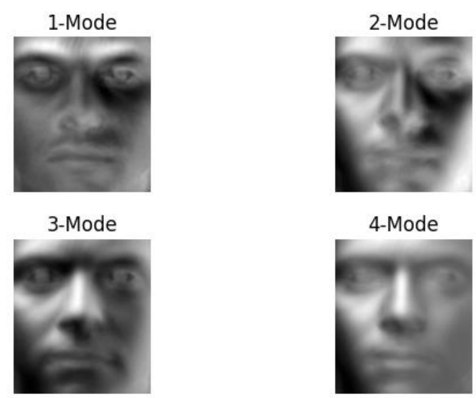

figure 5. Superposed faces: “2-mode” contains eigen face 1,2; “3-mode” contains eigen face 1,2,3; “4-mode” contains eigen faces 1,2,3,4

The six eigen faces we showed above can be thought as the smallest chemical molecule in the feature space. We can produce any sample faces with linear combination of these six simple eigen faces. The fact that the feature space is a linear space suggest that superposition is a valid operation. In this section, we provide three superimposed faces “2-mode”,”3-mode”,”4-mode”. There consisted of the two, three and four eigen faces with larger singular values respectively. “4-Mode” consists four eigen faces with gross singular value percentage exceeding 85%. We could see that the effect of shading is somewhat peeled off from the feature. Intuitively, the computer distinguishes the changes in shading across data set and determines that shading is not a part of human face.

Sec. V. Summary and Conclusions

This project aims to use a power data analysis technique, namely SVD, to extract the features on cropped face images provided by UCSD. The sample space is relatively large but well-constructed. By well-constructed, we mean that all image data are collected with the target facing directly into the camera and cropped uniformly. There are two results of this nature. Firstly, we can reduce the dimension of the sample spaces into only six principle directions. Secondly, the singular values of the data matrix are close to each other. In other word, all principle directions pick up some important features that we are interested in studying.