Image Compression Via Singular Value Decomposition Method

By: Anthony Silva

June 9, 2019

1 Introduction

In our modern society, efficiently transferring data is essential. In this paper we will particularly focus on how the singular value decomposition of real matrices can be used to compress images, allowing us to efficiently transfer them while using substantially fewer amounts of data. We can think of an image as a large matrix $A \in \mathcal{R}^{m \times n}$ of grayscale (or other scaled) real values, one for each pixel based on its shade in its respective position.

(a) Image

$$A = \begin{bmatrix} 1 & .25 \\ .75 & 0 \end{bmatrix}$$

(b) Grayscale Matrix

The values of the matrix A can range from $0 \leq a_{ij} \leq 1$. An all back pixel has the value $a_{ij} = 1$, while an all white pixel has the value $a_{ij} = 0$. Gray pixel values range anywhere in between. In its essence, singular value decomposition separates any matrix A into a sum of rank one matrices $\sum_{i=1}^{r} \sigma_i u_i v_i^T$ for $r = rank(A)$, $\sigma_i \in \mathcal{R}$, $u_i \in \mathcal{R}^m$, and $v_i \in \mathcal{R}^n$. However, rather than summing from 1 to the rank of the grayscale matrix $r = rank(A)$, we can closely approximate the matrix A by only summing the rank 1 matrices up to $k << r$. For a carefully chosen k, we can obtain an image that very closely resembles the original image with great accuracy for substantially fewer amounts of data.

2 What is the singular value decomposition of a matrix?

The singular value decomposition of a matrix $A \in \mathcal{R}^{m\times n}$ is a matrix factorization that separates a matrix into the product of 3 matrices, $U\Sigma V^T$, such that $U \in \mathcal{R}^{m\times m}$, $V \in \mathcal{R}^{n\times n}$ are orthogonal, and $\Sigma \in \mathcal{R}^{m\times n}$ is a diagonal matrix containing the singular values of A of (The square root of the eigenvalues of AA^T or A^TA, which are equivalent as we will show later on) on its main diagonal. The columns of U are the orthonormal singular column vectors of A (the normalized eigenvectors of AA^T) that form an orthonormal basis in $\mathcal{R}^m$, while the rows of V^T are the orthonormal singular row vectors of A (The normalized eigenvectors of A^TA) that form an orthonormal basis in $\mathcal{R}^n$.

Lemma 2. 1. *A^TA and AA^T contain the same eigenvalues.*

Proof. For $A \in \mathcal{R}^{m\times n}$, note that AA^T and A^TA are both symmetric and semi-positive definite. Thus, they each contain a diagonalization such that $AA^T = UZU^T$, and $A^TA = VSV^T$, with U forming an orthonormal basis in $\mathcal{R}^m$, and V forming an orthonormal basis in $\mathcal{R}^n$. The values in Z and S, are non-negative and real, and contain the eigenvalues of AA^T and A^TA on their main diagonal. Assume that A has a singular value decomposition. Then $A = U\Sigma V^T$. It follows that $AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma^2U^T$, which is the diagonlization of AA^T, such that the diagonal of Σ^2 contains its eigenvalues. Similarly, $A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^2V^T$, which is the diagonlization of A^TA such that the diagonal of Σ^2 contains its eigenvalues. Thus, AA^T and A^TA contain the same eigenvalues and $S = Z = \Sigma^2$. □

When constructing the SVD of a matrix, it is practical to arrange the order of the singular values in the diagonal matrix $\Sigma \in \mathcal{R}^{m\times n}$ in descending order such that $\sigma_{1,1} > \sigma_{2,2} \geq \sigma_{3,3} ... \geq \sigma_{p,p} \geq 0$ for p=$min(m, n)$. We also rearrange the singular vectors in U and V^T correspondingly. This allows us to rank the dominant column and row singular vectors of A in decreasing order with their corresponding singular values, such that we can closely approximate A using only the first $k < p$ singular vectors and values. When k is much smaller than p, and p is very large, we can reconstruct a matrix that approximates A with only $k(m+n+1)$ elements, as opposed to mn elements, which can be substantially less for a large matrix. Images today in 720p HD resolution that are 1280x720 pixels contain up to 921,600 pixels. A rank 100 approximation of this same image would allow us to closely approximate the same image for only 200,100 pixels!

Note that we can have the case where A contains some 0 singular values. This happens when A is not full rank, and in fact, we have exactly $min(m,n) - r$ zero singular values. Because Σ is a diagonal matrix, we can reduce U, V^T, Σ such that we only look at the first r columns of U, the first r rows of V^T, and the first r diagonal entries of Σ. This is because $U\Sigma V^T$ can be thought of as the following sum: $\sum_{i=1}^{min(m,n)} \sigma_i u_i v_i^T$. For $\sigma_i = 0$, we can remove every $\sigma_j u_j v_j^T$, for $i \leq j \leq min(m,n)$, since the outer product merely produces a zero matrix (assuming our $\sigma_i' s$ are ranked in decreasing order). We will only consider this reduced form of the SVD decomposition from now on.

Example 2. 1.

$$A = \begin{bmatrix} 1 & 2 \\ 1 & -2 \\ 5 & 2 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 5 & -3 & 9 \\ -3 & 5 & 1 \\ 9 & 1 & 29 \end{bmatrix}$$

has eigenvalues $\lambda_1 = 32, \lambda_2 = 7, \lambda_3 = 0$, and produces the orthogonal matrix

$$U = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{35}} & \frac{-3}{\sqrt{14}} \\ 0 & \frac{5}{\sqrt{35}} & \frac{-2}{\sqrt{14}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{35}} & \frac{1}{\sqrt{14}} \end{bmatrix}$$

whose columns form an orthonormal basis in $\mathcal{R}^3$. The columns of U are the normalized eigenvectors of AA^T corresponding to the singular values of A.

$$A^TA = \begin{bmatrix} 27 & 10 \\ 10 & 12 \end{bmatrix}$$

has eigenvalues $\lambda_1 = 32$, $\lambda_2 = 7$, and produces the orthogonal matrix

$$V^T = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \end{bmatrix}$$

whose rows form an orthonormal basis in $\mathcal{R}^2$. The columns of V are the normalized eigenvectors of A^TA corresponding to the singular values of A.

Both AA^T and A^TA produce the same singular diagonal matrix

$$\Sigma = \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{7} \\ 0 & 0 \end{bmatrix}$$

Allowing us to reduce the matrix U to

$$U = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{35}} \\ 0 & \frac{5}{\sqrt{35}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{35}} \end{bmatrix}$$

and Σ to

$$\Sigma = \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{7} \end{bmatrix}$$

Thus, A $= U\Sigma V^T$ which we can verify via matrix multiplication. Another way to think of the SVD of A is as the following sum of rank 1 matrices: $\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$.

3 Applying SVD to the Image Compression Problem

Example 3. 1. Suppose that we were to construct a grayscale matrix for an all black image containing 5x3 pixels. A grayscale matrix representation of this image would be the following:

$$G = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Suppose you wanted to send this image to a friend efficiently, using minimal amounts of data. A more efficient way to represent G would be as the following outerproduct:

$$G = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

which requires sending 8 total elements, as opposed to G which contains 15 total elements.

Although this is an oversimplified example, we can apply the same idea to any grayscale matrix by taking its SVD, since it can be thought of as a sum of outerproducts that produce the original matrix. However, it gets even better, as for $k < rank(A) = r$, we can closely approximate A using its k rank SVD approximation, which requires us to know $(r-k)(m+n+1)$ fewer elements. If $k << r$, this difference could be substantial. Our method of ordering the singular values in decreasing order when constructing the SVD allows us to make informed decisions on a good k approximation by looking at a plot of the singular values. Below is an example.

Example 3. 1. Consider the following grayscale image of a kitten.

We can construct its grayscale matrix according to the shade of each pixel using software. Once we find its grayscale matrix, we can run an algorithm such as cholesky factorization to calculate the eigenvalues and eigenvectors of AA^T and A^TA (i.e. the singular values and singular vectors of A). Doing so would allow us to calculate its k rank SVD approximations by applying the sum $\sum_{i=1}^{k} \sigma_i u_i v_i^T$. Below are $k = 10, 50, 100$, and 200 rank SVD image approximations.

Figure 2: Kitten

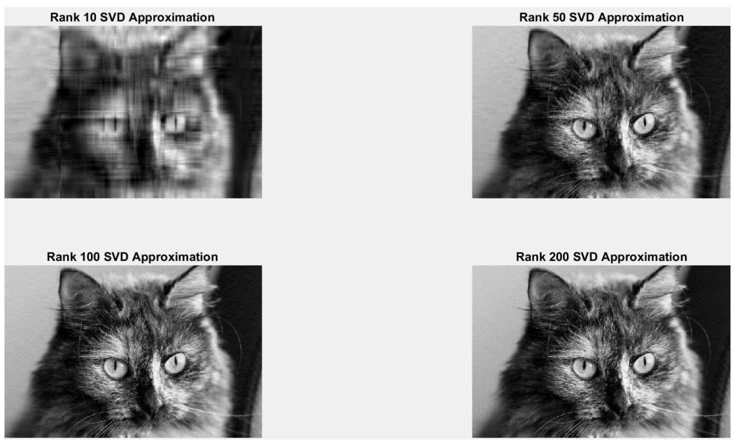

Figure 3: Kitten Image Approximations

Note that at a rank 100 SVD approximation, the image is indistinguishable to the original image to the naked eye. Our original kitten image grayscale matrix was of full rank with dimensions 400x600 for a total of 240,000 pixels. Its rank 100 SVD approximation only requires us to know $100(400 + 600 + 1) = 100,100$ pixels, accounting for a 58.29% reduction in the amount of data required to construct the photo. However, the question arises whether a rank 100 SVD is the most efficient image approximation. To find out we can look at the plot of the singular values below in Figure 4.

Recall that the SVD of the image is the sum $\sum_{i=1}^{400} \sigma_i u_i v_i^T$. We want to look for sudden drops between the magnitudes of one singular value and the next, followed by miniscule singular values in decreasing order. In our case, a rank 75 approximation appears to be a good rank SVD approximation since most

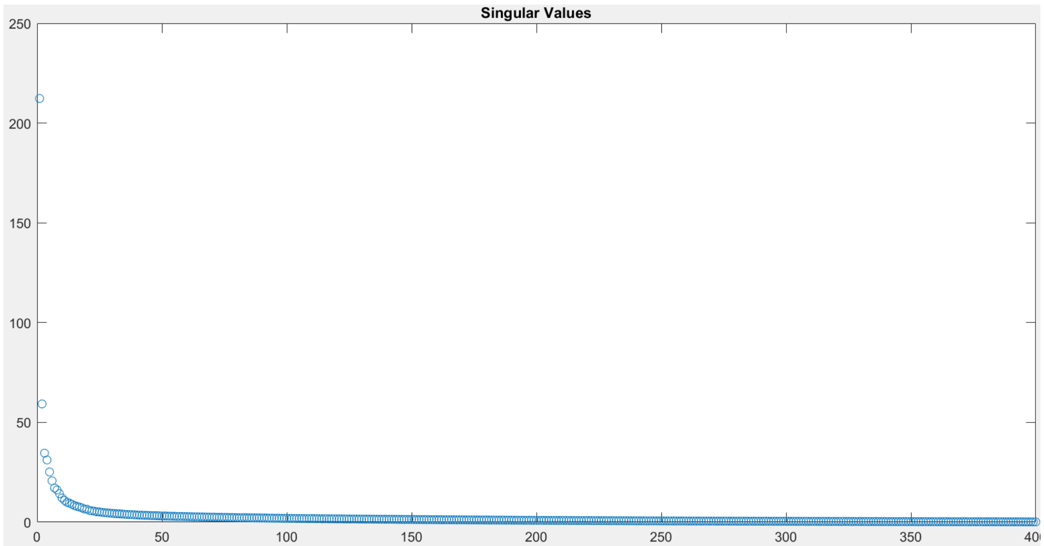

Figure 4: Kitten Image Singular Values

singular values after it are close to zero and don't contribute much to the overall sum when multiplied with the outerproduct (our outerproduct values are bounded between 0 and 1, so multiplying each term by a number close to zero, would produce a number even closer to zero). This approximation produces the following image in Figure 5.

Figure 5: 75 Rank SVD Approximation

4 Existence of SVD of Matrices

How can we be sure that a singular value decomposition exists for any grayscale matrix? To answer this question, we look first at how the SVD of any matrix $A \in \mathcal{R}^{m \times n}$ is constructed. Recall that in order to calculate the SVD of a matrix, we first calculated the eigenvalues and eigenvectors of $A^T A$ and AA^T. We know that these matrices are always symmetric and positive semi-definite. Therefore, they are each diagonalizable by an orthonormal matrix, and its non-zero eigenvalues are all strictly positive. This fact is key and is what allows the singular value decomposition of a matrix to happen. Similarly, we know that $r = rank(A) = rank(A^T A) = rank(AA^T)$, thus we end up with exactly r positive singular values. We can begin by first finding the diagonlization of the matrix $A^T A = V\Sigma^2 V^T$, which we now know exists. We then define U using the relationship $A = U\Sigma V^T$ and solve for it, to obtain $AV\Sigma^{-1} = U$ which we know exists since Σ is a diagonal matrix, and V is an orthonormal matrix. It follows that U must also be orthogonal.

5 Limitations of SVD for Image Compression Problems

Unfortunately, SVD is not always practical for image compression. There are computation expenses and some images cannot be approximated accurately for smaller ranks, such as photos without clearly defined borders or structure. Below is an example of an image that would be difficult to accurately approximate, even for a rank 100 SVD approximation:

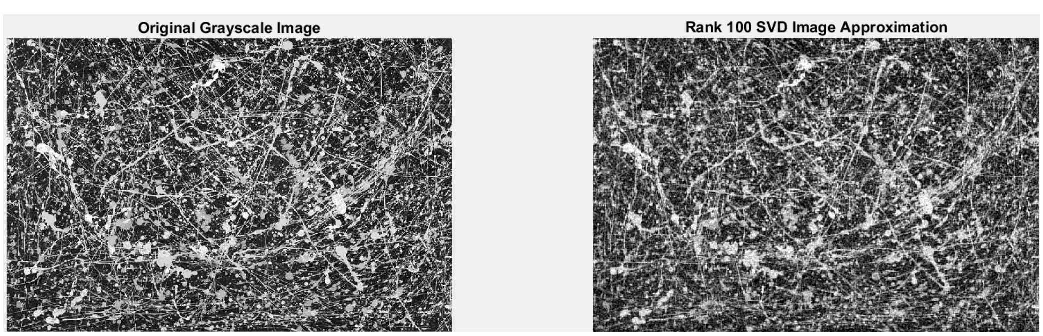

Figure 6

Looking at a plot of its singular values, they remain relatively high (greater than 1) up to about the 400th singular value. Our image is of rank 592. Thus, a low rank SVD approximation would not provide us with an accurate representation of the original photo. See Figure 7 for a plot of the image's singular values.

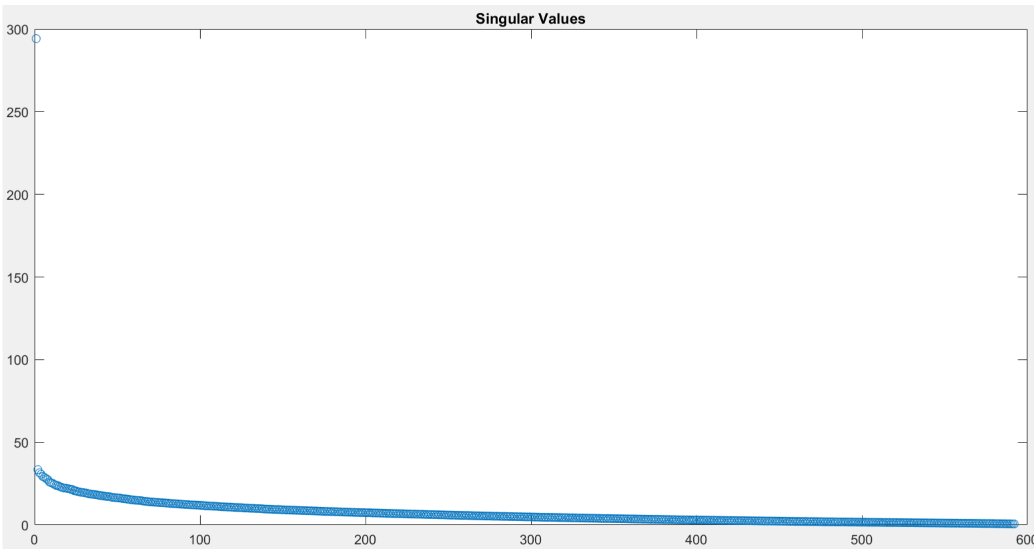

Figure 7

6 References

Strang, G. *Introduction to Linear Algebra, Fifth Edition.* Wellesley-Cambridge Press, Wellesley, MA, (2016)