Lecture 20: SVD Application - Image Compression

July 18, 2025

Outline

1. Best Approximation to A
2. Application of SVD to Image Compression
   1. Image Compression Demo

Image Compression - Introduction

- The singular value decomposition (SVD) can be thought of as representing A as the sum of rank-one matrices.
- In this lecture, we discuss approximating A using a **truncation** of this sum (i.e. omitting some terms at the end).

Image Compression - Best Approximation to A

Theorem 1

Let $m \geq n$. Let A be an $m \times n$ matrix, having rank r. Then A is the sum of r rank-one matrices, i.e. for $1 \leq j \leq r$, there exist scalars σ_j and vectors $u_j \in \mathbb{R}^m, v_j \in \mathbb{R}^n$ such that each $u_j v_j^T$ has rank 1, and

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^T.$$

Image Compression - Best Approximation to A

Proof.

From the definition of the (full, not reduced) SVD of A:

$$\underbrace{A}_{m\times n} = \underbrace{\begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}}_{m\times m} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}}_{n\times n}$$

$$= \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 v_1^T \\ \vdots \\ \sigma_r v_r^T \\ 0 \end{bmatrix},$$

$$= \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T = \sum_{j=1}^{r} \sigma_j u_j v_j^T.$$

□

Image Compression - Best Approximation to A

Exercise: What is the reduced SVD of the rank-1 matrix $A = xy^T$?
We can construct an approximate version of A, denoted A_k, using only the first k singular values as follows:

$$
\begin{aligned}
A_k &= \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} u_1 & \cdots & u_k \end{bmatrix}}_{U_k} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}}_{\Sigma_k} \underbrace{\begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix}}_{V_k}.
\end{aligned}
$$

So, as above, we may write $A_k = U_k \Sigma_k V_k^T$.

Image Compression - Best Approximation to A

Theorem 2 gives the following results:

1. Among all matrices B with rank $\leq k$, A_k minimizes $\|A - B\|_2$. In other words, A_k provides the best rank k approximation of A.
2. The approximation error is given by the singular value σ_{k+1}.

Image Compression - Best Approximation to A

Theorem 2

Let $m \geq n$. Let A be an $m \times n$ matrix, having rank r, and SVD:

$$A = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}.$$

For any $1 \leq k \leq r$, define

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T.$$

Then

$$\begin{aligned} \|A - A_k\|_2 &= \inf_{rank(B) \leq k} \|A - B\|_2 \\ &= \sigma_{k+1}. \end{aligned}$$

Image Compression - Best Approximation to A

Proof. $\|A - A_k\|_2 = \sigma_{k+1}$:
We will establish $\|A - A_k\|_2 = \sigma_{k+1}$, using the definition of SVD. We know that

$$
\begin{aligned}
A - A_k &= \left(\sum_{j=1}^{r} \sigma_j u_j v_j^T\right) - \left(\sum_{j=1}^{k} \sigma_j u_j v_j^T\right) \\
&= \sum_{j=k+1}^{r} \sigma_j u_j v_j^T \\
&= \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} 0 & & & & \\ & \sigma_{k+1} & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix},
\end{aligned}
$$

gives an SVD for $A - A_k$ (subject to reordering). We showed in an earlier Theorem that $\|A\|_2 = \sigma_1$, so we have $\|A - A_k\|_2 = \sigma_{k+1}$ (i.e. the largest remaining singular value).

Image Compression - Best Approximation to A

Part Two: Optimality:

We will show $\|A - A_k\|_2 = \inf_{\text{rank}(B) \leq k} \|A - B\|_2$ using a proof by contradiction. Towards a contradiction, suppose there exists B such that $\text{rank}(B) \leq k$ and $\|A - B\|_2 < \sigma_{k+1}$. That is, B is a strictly better approximation to A, with rank $\leq k$.

Recall that B is $m \times n$, i.e. we can view left multiplication by B as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. By the rank-nullity theorem,

$$\begin{aligned} rank(B) + nullity(B) &= n \\ \Rightarrow nullity(B) &= n - rank(B). \end{aligned}$$

So $null(B)$ has dimension $\geq n - k$, and contains non-zero vectors v (such that $Bv = 0$, by definition).

If there are non-zero vectors in $null(B)$, then B kills them.

Further, if $k = r = n$, then $A_k = A$. ($\|A - A_k\| = 0$, as small as possible).

Image Compression - Best Approximation to A

Observe that $null(B)$ and $span\{v_1, \ldots, v_{k+1}\}$ are subspaces of $\mathbb{R}^n$, with

- $nullity(B) \geq n - k$, and
- $\dim(span\{v_1, \ldots, v_{k+1}\}) = k + 1$.

Since $(n-k) + (k+1) > n$, therefore $null(B)$ and $\text{span}\{v_1, \ldots, v_{k+1}\}$ must have a non-zero intersection, i.e., $\exists z \neq 0$ such that

$$z \in null(B) \cap \text{span}\{v_1, \ldots, v_{k+1}\}.$$

Without loss of generality, let $\|z\|_2 = 1$. We will obtain a contradiction by showing $\|A - B\|_2 \geq \sigma_{k+1}$.

Image Compression - Best Approximation to A

Note $\|A-B\|_2^2 \geq \|(A-B)z\|_2^2$ (Recall the definition of the matrix 2-norm, $\|A\|_2 = \max\|Ax\|_2$ with $\|x\|_2 = 1$). Since $z \in \text{null}(B)$, $Bz = 0$, and therefore

$$\begin{aligned}
\|(A-B)z\|_2^2 &= \|Az - Bz\|_2^2 \\
&= \|Az - 0\|_2^2 \\
&= \|Az\|_2^2 \\
&= \left\| \left(\sum_{i=1}^{n} \sigma_i u_i v_i^T \right) z \right\|_2^2 .
\end{aligned}$$

For an arbitrary $0 \leq i \leq n$, the i^{th} term of the sum equals $\sigma_i u_i v_i^T z$.

Image Compression - Best Approximation to A

We also have $z \in \text{span}\{v_1, \ldots, v_{k+1}\} \subseteq \mathbb{R}^n$. The above i^{th} term therefore equals

$$\begin{aligned} & \sigma_i u_i \underbrace{\left(v_i^T z\right)}_{\text{scalar}} \\ = \quad & \sigma_i \left(v_i^T z\right) u_i, \end{aligned}$$

Image Compression - Best Approximation to A

and therefore the above squared 2-norm expression equals

$$
\begin{aligned}
& \left(\sum_{i=1}^{n} \sigma_i \left(v_i^T z\right) u_i\right)^T \left(\sum_{j=1}^{n} \sigma_j \left(v_j^T z\right) u_j\right) \\
= \; & \sum_{i=1}^{n} \sigma_i^2 \left(v_i^T z\right)^2, \text{ using orthogonality of the } u_i s \\
= \; & \sum_{i=1}^{k+1} \sigma_i^2 \left(v_i^T z\right)^2, \text{ since } z \in \operatorname{span}\{v_1, \ldots, v_{k+1}\} \\
\geq \; & \sigma_{k+1}^2 \sum_{i=1}^{k+1} \left(v_i^T z\right)^2, \text{ by the ordering of the } \sigma_i s.
\end{aligned}
$$

Now I claim that $\sum_{i=1}^{k+1} \left(v_i^T z\right)^2 = 1$. We have assumed that $\|z\|_2 = 1$. Since $z \in \operatorname{span}\{v_1, \ldots, v_{k+1}\}$, we may write $z = \sum_{\ell=1}^{k+1} c_\ell v_\ell$, for some c_ℓs.

Image Compression - Best Approximation to A

Then we have

$$\begin{aligned}
1 &= \|z\|_2 \\
&= \sqrt{z^T z} \\
&= \sqrt{\left(\sum_{\ell=1}^{k+1} c_\ell v_\ell\right)^T \left(\sum_{j=1}^{k+1} c_j v_j\right)} \\
&= \sqrt{\sum_{\ell=1}^{k+1} c_\ell^2 \left(v_\ell^T v_\ell\right)}, \text{ by orthogonality of the } v_\ell s \\
&= \sqrt{\sum_{\ell=1}^{k+1} c_\ell^2}, \text{ since each } \|v_\ell\|_2 = 1, \text{ and so} \\
1 &= \sum_{\ell=1}^{k+1} c_\ell^2.
\end{aligned}$$

Image Compression - Best Approximation to A

Now we can compute

$$
\begin{aligned}
\sum_{i=1}^{k+1} \left(v_i^T z \right)^2 &= \sum_{i=1}^{k+1} \left(v_i^T \left(\sum_{\ell=1}^{k+1} c_\ell v_\ell \right) \right)^2 \\
&= \sum_{i=1}^{k+1} c_i^2 \left(v_i^T v_i \right)^2, \text{ by orthogonality of the } v_i s \\
&= \sum_{i=1}^{k+1} c_i^2, \text{ since each } \|v_i\|_2 = 1 \\
&= 1, \text{ as claimed.}
\end{aligned}
$$

Image Compression - Best Approximation to A

Putting everything together, we finally get

$$\begin{aligned} \|A - B\|_2^2 &\geq \|(A - B)z\|_2^2 \\ &\geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \left(v_i^T z\right)^2 \\ &= \sigma_{k+1}^2, \end{aligned}$$

implying $\|A - B\|_2 \geq \sigma_{k+1}$ and contradicting the fact that $\|A - B\|_2 < \sigma_{k+1}$.

- Hence no such B can exist.
- Note that the analogous statement holds true for the Frobenius norm

$$\|A - A_k\|_F = \inf_{\text{rank}(B) \leq k} \|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_r^2}.$$

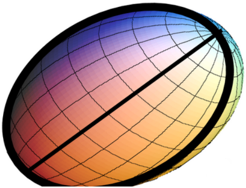
□

Image Compression - Best Approximation to A

- A geometric interpretation of the low rank approximation is as follows.
- Consider trying to determine the line segment that "best" approximates a (hyper)ellipsoid.
- The best approximation is the line segment along the longest axis of the (hyper)ellipsoid.

Image Compression - Best Approximation to A

- This example corresponds to approximating A with $k = 1$.
- With $k = 2$, we can ask what ellipse gives "best" approximation of the (hyper)ellipsoid?
- The best ellipse is the one spanning the two longest axes (as shown with black curves in the figure below for the ellipsoid in $\mathbb{R}^3$).
- With larger k the same idea holds.



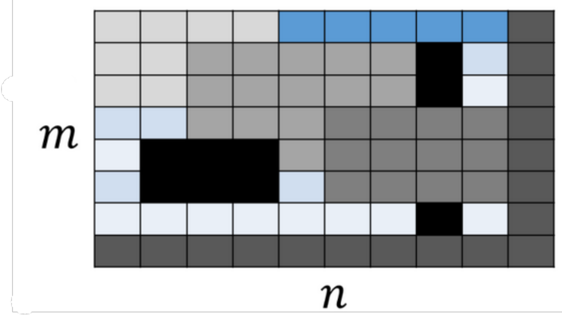
Application of SVD to Image Compression

The SVD can be used to produce a cheaper approximate version of an image (or other dataset) that captures the “most important” parts.

K=10 K=20 K=50 Original

Application of SVD to Image Compression

Consider and $m \times n$ pixel (grayscale) image as an $m \times n$ matrix A where A_{ij} is the intensity of the pixel (i, j). If we can store fewer than mn entries, we have a compressed representation.



Application of SVD to Image Compression

- Let $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ be the best rank-k approximation of A.
- Then A_k gives a compressed version of the image A using the first k singular values.
- For example, given an input image with $m = 320, n = 200$.
- For A_k, we need to only store vectors $u_1, \ldots, u_k$ and $\sigma_1 v_1, \ldots, \sigma_k v_k$.
- Thus, we have $(m+n)k$ entries to store in total.
- This gives a compression ratio of $\frac{(m+n)k}{mn}$.
- In our specific example we have

$$\frac{(320+200)k}{320 \cdot 200} \approx \frac{k}{123}.$$

Application of SVD to Image Compression

The table below gives values for different k. As can be seen from the relative error and the compression ratios, this is an effective approach with small k.

k	Rel. err. σ_{k+1}/σ_1	Comp. ratio
3	0.155	2.4%
10	0.077	8.1%
20	0.04	16.3%

Image Compression Demo

An example of code for image compression using the above ideas is given in Algorithm 24. Note that sample code for colour images is given in **SVDimageCompression.m** (which just computes an SVD for each colour channel separately).

Image Compression Demo

Algorithm 1 : Grayscale Image Compression

```
A=rgb2gray(imread('baboon.png'));
A=double(A);
[U,S,V]=svd(A);
k = 30;                                  ▷ try different choices
Ak=U(:,1:k)*S(1:k,1:k)*V(:,1:k)';
colormap('gray');
imagesc(Ak);
axis equal;
```

Image Compression Demo

Given the input figure on the left below, the code computes the compressed grayscale output on the right.

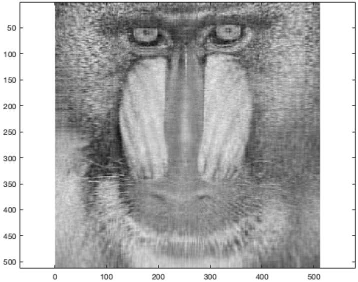
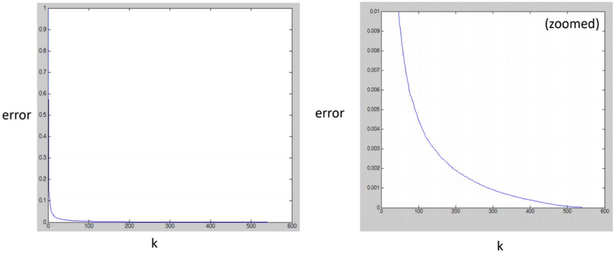


Image Compression Demo

We had previously shown that $\|A - A_k\|_2 = \sigma_{k+1}$ gives the approximation error. So we can plot the relative error as $\frac{\sigma_{k+1}}{\sigma_1}$ against the choice of k (see below).



As depicted in the plot, the greater the k, the closer to the original image, thus a lower approximation error.