Properties of Singular Value Decomposition
Matrix Computations — CPSC 5006 E

Julien Dompierre

Department of Mathematics and Computer Science
Laurentian University

Sudbury, November 19, 2009

Properties of Singular Value Decomposition

- Four fundamental subspaces
- Condition number
- Rank estimation
- Thin SVD
- Pseudo-inverse (Moore-Penrose)
- Least-squares problems
- Best approximation
- Sections 2.5.1 – 2.5.5 and 5.5.1 – 5.5.4 of the textbook

The singular Value Decomposition (p. 475)

The SVD decomposition of A involves an $m \times n$ "diagonal" matrix Σ of the form

$$\Sigma_{m\times n} = \begin{bmatrix} D_{r\times r} & 0_{r\times n-r} \\ 0_{m-r\times r} & 0_{m-r\times n-r} \end{bmatrix}$$

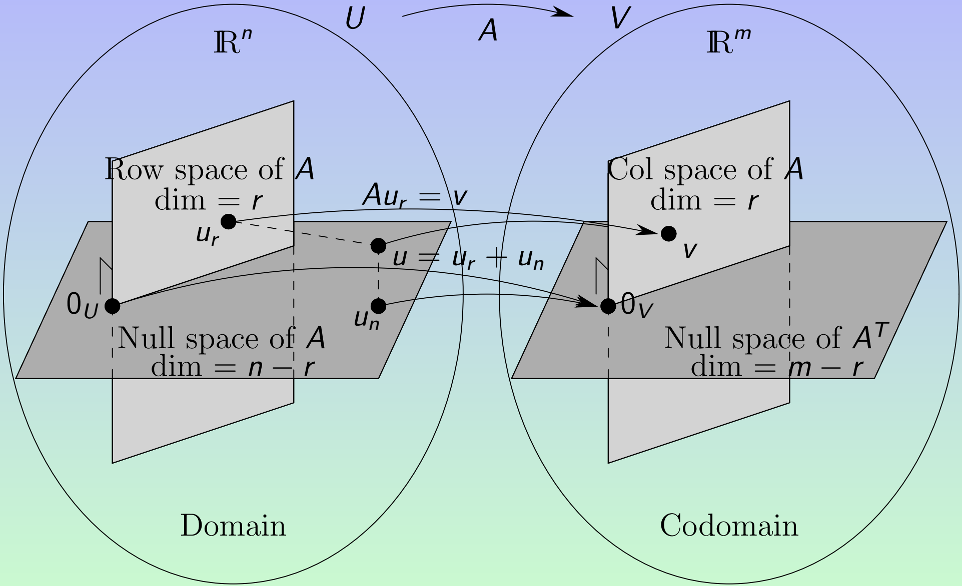
where D is an $r \times r$ diagonal matrix for some r not exceeding the smaller of m and n.

Theorem (Theorem 10, The singular Value Decomposition)

Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ "diagonal" matrix Σ for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

$$A = U\Sigma V^T.$$

The Four Fundamental Subspaces



The Four Fundamental Subspaces (p. 478–479)

Given an SVD for an $m \times n$ matrix A, let $u_1, \ldots, u_m$ be the left singular vectors, $v_1, \ldots, v_n$ the right singular vectors, $\sigma_1, \ldots, \sigma_n$ the singular values, and let r be the rank of A. Then

- $\{u_1, \ldots, u_r\}$ is an orthonormal basis for Col A;
- $\{u_{r+1}, \ldots, u_m\}$ is an orthonormal basis for Nul A^T;
- $\{v_{r+1}, \ldots, v_n\}$ is an orthonormal basis for Nul A;
- $\{v_1, \ldots, v_r\}$ is an orthonormal basis for Row A.

The Four Fundamental Subspaces (p. 478–479)

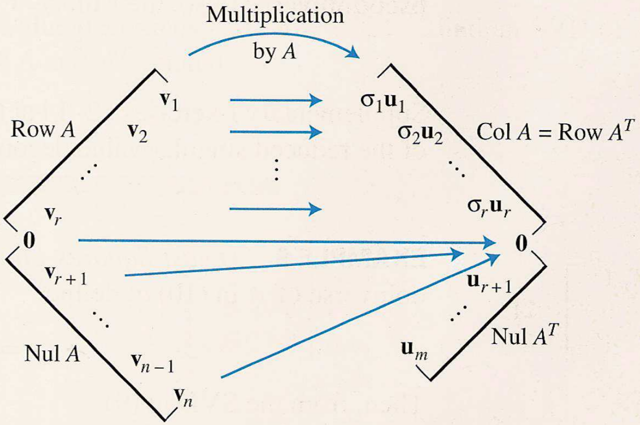


FIGURE 4 The four fundamental subspaces and the action of A.

Theorem

$\|A\|_2 = \sigma_1$. If A is square and non singular, then $\|A^{-1}\|_2 = 1/\sigma_n$ and $\kappa_2(A) = \|A\|_2 \times \|A^{-1}\|_2 = \sigma_1/\sigma_2$.

Proof.

It is clear from its definition that the two-norm of a diagonal matrix is the largest absolute entry on its diagonal. Thus,
$\|A\|_2 = \|U^T AV\|_2 = \|\Sigma\|_2 = \sigma_1$ and
$\|A^{-1}\|_2 = \|V^T A^{-1} U\|_2 = \|\Sigma^{-1}\|_2 = \sigma_n^{-1}$. □

Theorem

$\|A\|_F = \left(\sum_{i=1}^{r} \sigma_i^2\right) 1/2$.

Numerical Rank and the SVD (sec. 5.5.8)

Assume that the original matrix A is exactly of rank k.

The computed SVD of A will be the SVD of a nearby matrix $A + E$.

Easy to show that $|\hat{\sigma}_i - \sigma_i| \le \alpha \; \sigma_1 u$

Result: zero singular values will yield small computed singular values

Determining the "numerical rank": Treat singular values below a certain threshold δ as zero.

Practical problem: Need to set δ.

Numerical Rank

`[U, S, V, rk] = svd(X, tol)` gives in addition `rk`, the numerical rank of `X` i.e. the number of singular values larger than `tol`.

`rank(X)` is the numerical rank of `X` i.e. the number of singular values of `X` that are larger than `norm(size(X),'inf') * norm(X) * %eps`.

Reduced Singular Value Decomposition

When Σ contains rows or columns of zeros, a more compact decomposition of A is possible. Let $r = \operatorname{rank}(A)$ and partition U and V into submatrices whose first blocks contains r columns:

$$\begin{aligned} U &= \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}, \quad \text{where } U_r = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \\ V &= \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}, \quad \text{where } V_r = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}. \end{aligned}$$

Then U_r is $m \times r$ and V_r is $n \times r$. Then partitioned matrix multiplication shows that

$$A = U\Sigma V^T = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T.$$

This factorization of A is called a **reduced singular value decomposition** of A (or **thin SVD**). Since the diagonal entries in D are non zero, we can for the following matrix, called the **pseudoinverse** (also, the **Moore-Penrose inverse**) of A:

$$A^+ = V_r D^{-1} U_r^T.$$

Pseudo-Inverse of an Arbitrary Matrix

The pseudo-inverse of A is given by

$$A^+ = V \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i}$$

Moore-Penrose conditions:

The pseudo inverse X of a matrix A is uniquely determined by these four conditions:

a. $AXA = A$
b. $XAX = X$
c. $(AX)^H = AX$
d. $(XA)^H = XA$

In the full-rank overdetermined case, $A^+ = (A^T A)^{-1} A^T$

Properties of the Moore-Penrose Inverse

Theorem (Property a)

$$AA^+A = A$$

Proof.

Using the reduced singular values decomposition, the definition of A^+, and the associativity of matrix multiplication gives:

$$\begin{aligned} AA^+A &= (U_r D V_r^T)(V_r D^{-1} U_r^T)(U_r D V_r^T) \\ &= (U_r D D^{-1} U_r^T)(U_r D V_r^T) \\ &= U_r D D^{-1} D V_r^T \\ &= U_r D V_r^T \\ &= A \end{aligned}$$

□

Properties of the Moore-Penrose Inverse

Theorem (Property b)

$$A^+AA^+ = A^+$$

Proof.

Using the reduced singular values decomposition, the definition of A^+, and the associativity of matrix multiplication gives:

$$\begin{aligned}
A^+AA^+ &= (V_rD^{-1}U_r^T)(U_rDV_r^T)(V_rD^{-1}U_r^T) \\
&= (V_rD^{-1}DV_r^T)(V_rD^{-1}U_r^T) \\
&= V_rD^{-1}DD^{-1}U_r^T \\
&= V_rD^{-1}U_r^T \\
&= A^+
\end{aligned}$$

$\square$

Properties of the Moore-Penrose Inverse

Theorem (Property 3)

For each y in $\mathbb{R}^m$, AA^+y is the orthogonal projection of y onto Col A.

Proof.

Because the columns of V_r are orthonormal,

$$AA^+y = (U_r D V_r^T)(V_r D^{-1} U_r^T)y = (U_r D D^{-1} U_r^T)y = U_r U_r^T y$$

Since $U_r U_r^T y$ is the orthogonal projection of y onto Col U_r, and since Col U_r = Col A, AA^+y is the orthogonal projection of y onto Col A. $\square$

Properties of the Moore-Penrose Inverse

Theorem (Property 4)

For each x in $\mathbb{R}^n$, A^+Ax is the orthogonal projection of x onto Row A.

Proof.

Because the columns of U_r are orthonormal,

$$A^+Ax = (V_rD^{-1}U_r^T)(U_rDV_r^T)x = (V_rD^{-1}DV_r^T)x = V_rV_r^Tx$$

Since $V_rV_r^Tx$ is the orthogonal projection of x onto Col V_r, and since Col V_r = Row A, A^+Ax is the orthogonal projection of x onto Row A. □

Properties of the Moore-Penrose Inverse

Theorem

Suppose the equation $Ax = b$ is consistent (have one or more solutions), and let $x^+ = A^+b$. Then, x^+ is the **minimum length solution** *of $Ax = b$.*

Proof.

- $x^+ = A^+b = A^+Ax$. Then, x^+ is the orthogonal projection of x onto Row A, using property 3. This shows that x is in Row A.
- $Ax^+ = A(A^+Ax)$ using the previous point. $Ax^+ = (AA^+A)x = Ax = b$ from property (a). This shows that x^+ is a solution of $Ax = b$.
- Let u be any solution of $Ax = b$. Since x^+ is the orthogonal projection of x onto Row A, the vector $u - x^+$ is orthogonal to x^+, and using the Pythagorean Theorem $\|u\|^2 = \|x^+\|^2 + \|u - x^+\|^2$. This shows that $\|x^+\| \leq \|u\|$, with equality only if $u = x^+$.

SVD can give much information about solving overdetermined and underdetermined linear systems.

Let A be an $m \times n$ matrix and $A = U\Sigma V^T$ its SVD with $r = \text{rank}\,(A)$, $V = [v_1, \ldots, v_n]$ $U = [u_1, \ldots, u_m]$. Then

$$x_{LS} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} \, v_i$$

minimizes $\|b - Ax\|_2$ and has the smallest 2-norm among all possible minimizers. In addition,

$$\rho_{LS} \equiv \|b - Ax_{LS}\|_2 = \|z\|_2 \text{ with } z = [u_{r+1}, \ldots, u_m]^T b$$

A restatement of the first part of the previous result:

Consider the general linear least-squares problem

$$\min_{x \in S} \|x\|_2, \quad S = \{x \in \mathbb{R}^n \mid \|b - Ax\|_2 \min\}.$$

This problem always has a unique solution given by

$$x = A^+ b$$

Taylor Series Expansion

The **Taylor series** is a representation of a function as an infinite sum of terms calculated from the values of its derivatives at a single point. It may be regarded as the limit of the Taylor polynomials. Taylor series are named in honour of English mathematician Brook Taylor (18 August 1685 - 30 November 1731).

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!} f^{(n)}(a) \\ &= \frac{(x-a)^0}{0!} f^{(0)}(a) + \frac{(x-a)^1}{1!} f^{(1)}(a) + \frac{(x-a)^2}{2!} f^{(2)}(a) \\ &\quad + \frac{(x-a)^3}{3!} f^{(3)}(a) + \frac{(x-a)^4}{4!} f^{(4)}(a) + \cdots \\ &= f(a) + (x-a) f'(a) + \frac{(x-a)^2}{2} f''(a) + \frac{(x-a)^3}{6} f'''(a) \\ &\quad + \frac{(x-a)^4}{24} f^{(4)}(a) + \frac{(x-a)^5}{120} f^{(5)}(a) + \cdots \end{aligned}$$

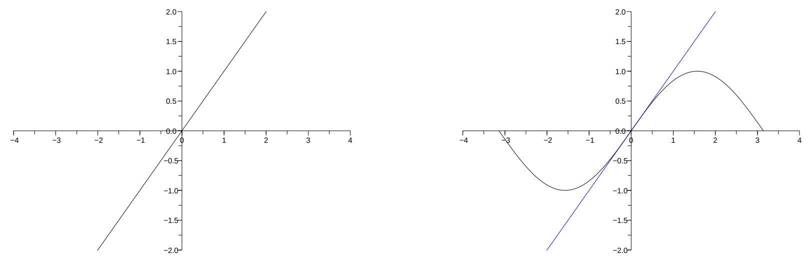
Taylor Series of $\sin(x)$ at $a = 0$

$$
\begin{aligned}
\sin(x) &= \sin(0) + (x-0)\sin'(0) + \frac{(x-0)^2}{2}\sin''(0) \\
&\quad + \frac{(x-0)^3}{6}\sin'''(0) + \frac{(x-0)^4}{24}\sin^{(4)}(0) \\
&\quad + \frac{(x-0)^5}{120}\sin^{(5)}(0) + \cdots
\end{aligned}
$$

As $\sin' = \cos$, $\cos' = -\sin$, $\sin(0) = 0$ and $\cos(0) = 1$, then

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \frac{x^{13}}{13!} - \cdots$$

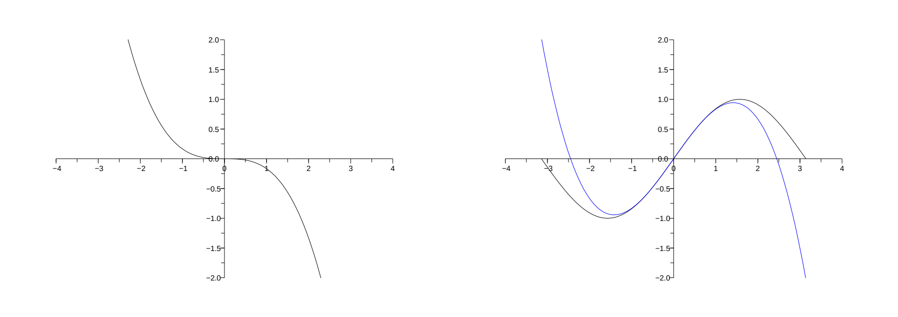
Taylor Series of $\sin(x)$ at $a = 0$



The first term x of the Taylor series and the sum

$$\sin(x) \approx x.$$

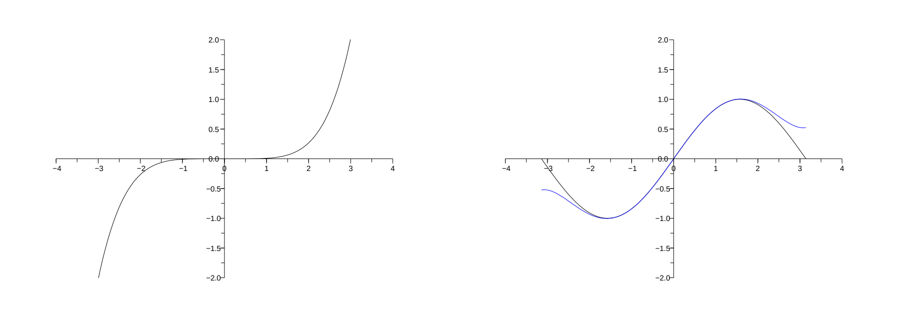
Taylor Series of $\sin(x)$ at $a = 0$



The second term $-x^3/6$ of the Taylor series and the sum

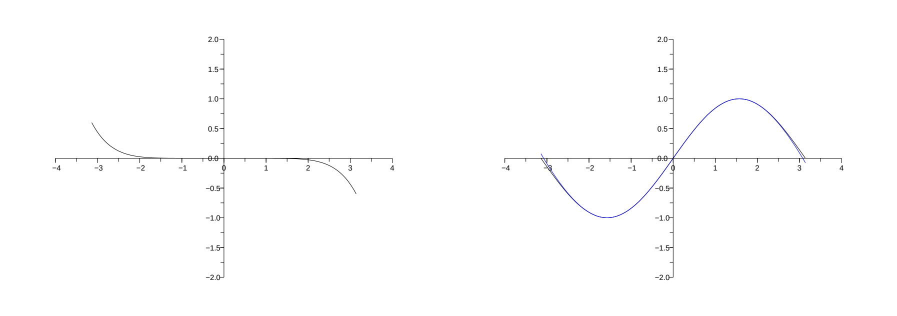
$$\sin(x) \approx x - \frac{x^3}{3!}.$$

Taylor Series of $\sin(x)$ at $a = 0$



The third term $x^5/120$ of the Taylor series and the sum

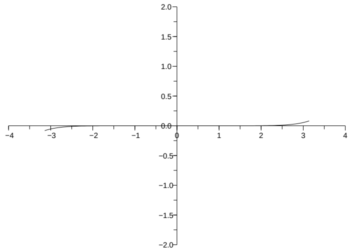
$$\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

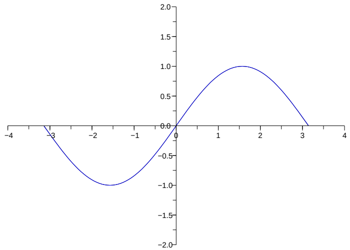


The fourth term $-x^7/5040$ of the Taylor series and the sum

$$\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}.$$

Taylor Series of $\sin(x)$ at $a = 0$

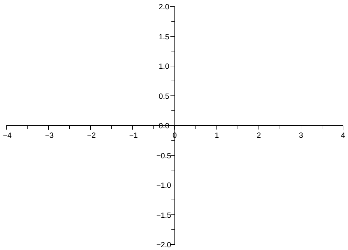


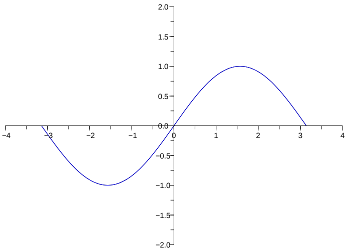


The fifth term $x^9/362880$ of the Taylor series and the sum

$$\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!}.$$

Taylor Series of $\sin(x)$ at $a = 0$





The sixth term $-x^{11}/39916800$ of the Taylor series and the sum

$$\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!}.$$

Spectral Decomposition

Given an SVD for an $m \times n$ matrix A. If $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}$ and $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, then the **spectral decomposition** of A is given by

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T.$$

Proof.

$U\Sigma = \begin{bmatrix} \sigma_1 u_1 & \cdots & \sigma_r u_r & 0 & \cdots & 0 \end{bmatrix}$. The column-row expansion of the product $(U\Sigma)V^T$ shows that

$$A = (U\Sigma)V^T = (U\Sigma)\begin{bmatrix} v_1^T \\ \vdots \\ v_n^t \end{bmatrix} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

where r is the rank of A. □

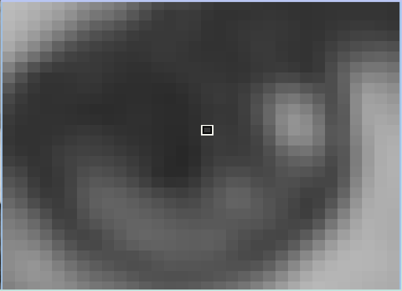
Truncated SVD

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Truncate the SVD by only keeping the $\sigma_i's$ that $\geq \tau$, where τ is a threshold. This gives the Truncated SVD solution:

$$A_\tau = \sum_{\sigma_i \geq \tau} \sigma_i u_i v_i^T$$

Many applications [e.g., Image processing,..]



The initial picture is a 512×512 matrix. Each entry is an integer in the interval $[0, 255]$ that represents the gray level, 0 for black, 255 for white.

The Lena Test Case

Lenna or Lena is the name given to a standard test image originally cropped from a Playboy magazine centerfold picture of Lena Söderberg, a Swedish model who posed naked for the November 1972 issue. The image is probably the most widely used test image for all sorts of image processing algorithms (such as compression and denoising) and related scientific publications.

Let A be the matrix that stores the gray levels of the Lena's picture. The SVD of A gives matrices U, Σ and V such that $A = U\Sigma V^T$. The following slides shows the pictures obtained with partial spectral decompositions of A.

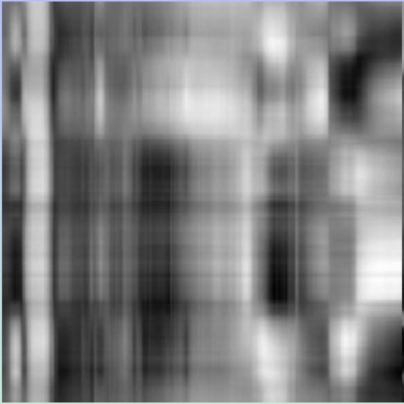
Spectral Decomposition of Degree 1

Left, spectral decomposition with 1 terms: $\sum_{i=1}^{1} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 2

Left, spectral decomposition with 2 terms: $\sum_{i=1}^{2} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 3



Left, spectral decomposition with 3 terms: $\sum_{i=1}^{3} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 4

Left, spectral decomposition with 4 terms: $\sum_{i=1}^{4} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 5

Left, spectral decomposition with 5 terms: $\sum_{i=1}^{5} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 6

Left, spectral decomposition with 6 terms: $\sum_{i=1}^{6} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 7


Left, spectral decomposition with 7 terms: $\sum_{i=1}^{7} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 8

Left, spectral decomposition with 8 terms: $\sum_{i=1}^{8} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 9

Left, spectral decomposition with 9 terms: $\sum_{i=1}^{9} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 10

Left, spectral decomposition with 10 terms: $\sum_{i=1}^{10} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 12

Left, spectral decomposition with 12 terms: $\sum_{i=1}^{12} \sigma_i u_i v_i^T$. Right, the initial picture.

Left, spectral decomposition with 15 terms: $\sum_{i=1}^{15} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 20

Left, spectral decomposition with 20 terms: $\sum_{i=1}^{20} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 25

Left, spectral decomposition with 25 terms: $\sum_{i=1}^{25} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 30

Left, spectral decomposition with 30 terms: $\sum_{i=1}^{30} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 40

Left, spectral decomposition with 40 terms: $\sum_{i=1}^{40} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 50

Left, spectral decomposition with 50 terms: $\sum_{i=1}^{50} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 60

Left, spectral decomposition with 60 terms: $\sum_{i=1}^{60} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 80

Left, spectral decomposition with 80 terms: $\sum_{i=1}^{80} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 100

Left, spectral decomposition with 100 terms: $\sum_{i=1}^{100} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 150

Left, spectral decomposition with 150 terms: $\sum_{i=1}^{150} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 200

Left, spectral decomposition with 200 terms: $\sum_{i=1}^{200} \sigma_i u_i v_i^T$. Right, the initial picture.

Spectral Decomposition of Degree 300

Left, spectral decomposition with 300 terms: $\sum_{i=1}^{300} \sigma_i u_i v_i^T$. Right, the initial picture.

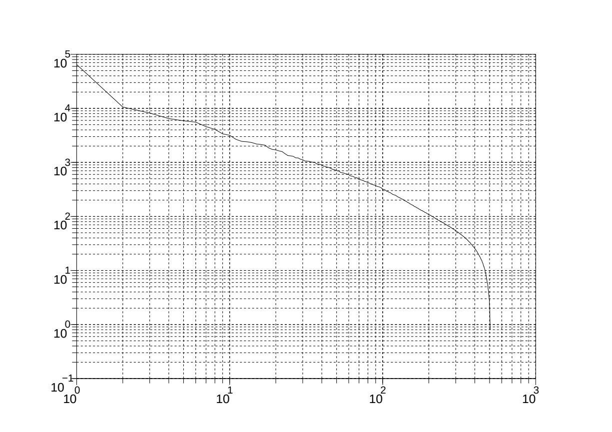
SVD for Picture Compression

The initial picture has 512×512 pixels, which means $512 \times 512 = 262\,144$ numbers to store.
If a spectral decomposition with 100 terms is enough to represent a picture with sufficient accuracy, i.e.,

$$A \approx \sum_{i=1}^{100} \sigma_i u_i v_i^T ,$$

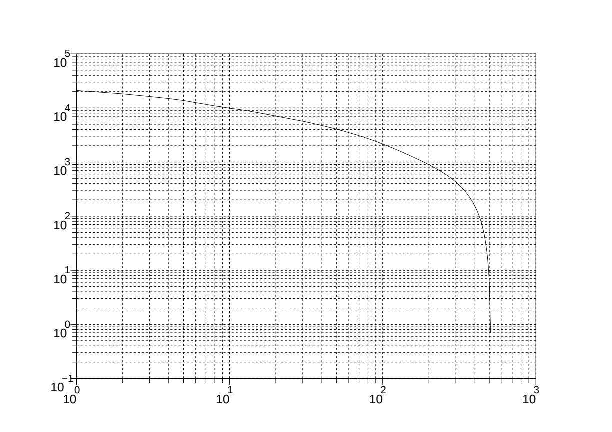
then this spectral decomposition requires 100 vectors u and 100 vectors v. This represents $200 \times 512 = 102\,400$ numbers to store instead of the initial 262 144, which is a compression rate of 39%.
The SVD decomposition is then a compression algorithm with loss of information.

Distribution of the Singular Values



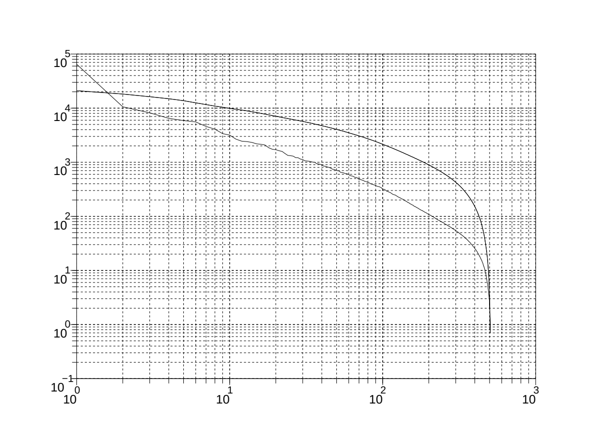
Here is the plot of the log in base 10 of the singular values in function of the log in base 10 of n. There are 507 singular values from 64735.683 to 1.829E-30. The last five singular values are zeros. The rank of the matrix is 507.

Frobenius Norm of the Error



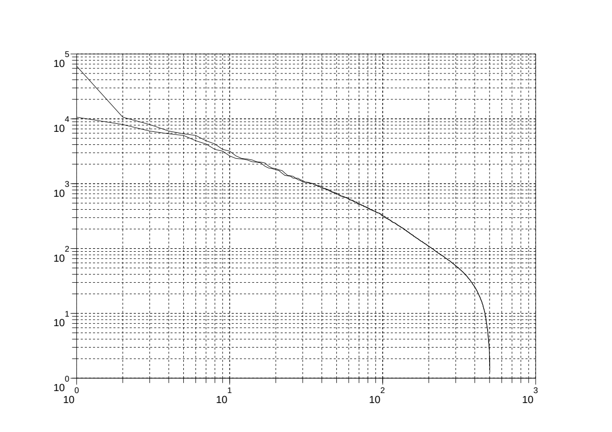
Here is the plot of the log in base 10 of the Frobenius norm of the difference between the initial picture and the spectral decomposition of degree n in function of the log in base 10 of n.

Frobenius Error versus Singular Value



The error between the initial picture and the spectral decomposition is controlled by the singular value distribution.

2-Norm Error versus Singular Value



The 2-norm error between the initial picture and the spectral decomposition is exactly the same as the singular values.

Best Approximation Theorem

Theorem

Let $A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$. Then a matrix of rank $k < n$ closest to A (measured with $\|\cdot\|_2$) is $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$, and $\|A - A_k\|_2 = \sigma_{k+1}$. We may also write $A_k = U \Sigma_k V^T$, where $\Sigma_k = \mathrm{diag}(\sigma_1, ..., \sigma_k, 0, ..., 0)$.

In other words,

$$\min_{\mathrm{rank}\,(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

A_k has rank k by construction and

$$\begin{aligned}
\|A - A_k\|_2 &= \left\| \sum_{i=k+1}^{n} \sigma_i u_i v_i^T \right\|_2 \\
&= \left\| U \begin{bmatrix} 0_k & & & \\ & \sigma_{k+1} & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} V^T \right\|_2 \\
&= \sigma_{k+1}.
\end{aligned}$$