Theorem

Any A an $m \times n$ matrix can be written as a product

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

where:

- $U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix}$ is an $m \times m$ orthogonal matrix.
- $V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ is an $n \times n$ orthogonal matrix.
- Σ is an $m \times n$ matrix of the following form:

$$\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{bmatrix}$$

(if $n \leq m$) (if $n \geq m$)

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$.

Note.

- The numbers $\sigma_1, \sigma_2, \ldots$ are called *singular values* of A.
- The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are called *left singular vectors* of A.
- Then the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are called *right singular vectors* of A.
- The formula $A = U\Sigma V^T$ is called a *singular value decomposition (SVD)* of A.
- The matrix Σ is uniquely determined, but U and V depend on some choices.

Theorem

Let A be a matrix with a singular value decomposition

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

If

$$U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix} \qquad V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$$

and $\sigma_1, \dots, \sigma_r$ are singular values of A then then

$$A = \sigma_1(\mathbf{u}_1\mathbf{v}_1^T) + \sigma_2(\mathbf{u}_2\mathbf{v}_2^T) + \dots + \sigma_r(\mathbf{u}_r\mathbf{v}_r^T)$$

e.g.:

$$\underset{3\times 2}{A} = \underset{3\times 3}{U} \cdot \underset{3\times 2}{\Sigma} \cdot \underset{2\times 2}{V^T}$$

$$U = [u_1 \; u_2 \; u_3]$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix}$$

$$V = [v_1 \; v_2]$$

$$A = [u_1 \; u_2 \; u_3] \cdot \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

$$= [\sigma_1 u_1 \; \sigma_2 u_2] \cdot \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

$$= \sigma_1(u_1 v_1)^T + \sigma_2(u_2 v_2^T)$$

Application: Image compression

- The size of this image is 700×800 pixels.
- The color of each pixel is represented by an integer between 0 (black) and 255 (white).
- The whole image is described by a ~~(symmetric)~~ matrix A consisting of $700 \times 800 = 560,000$ numbers
- Each number is stored in 1 byte, so the image file size is 560,000 bytes (≈ 0.53 MB).

How to make the image file smaller:

1) Compute SVD of the matrix A:

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

where

$$U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix} \qquad V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$$

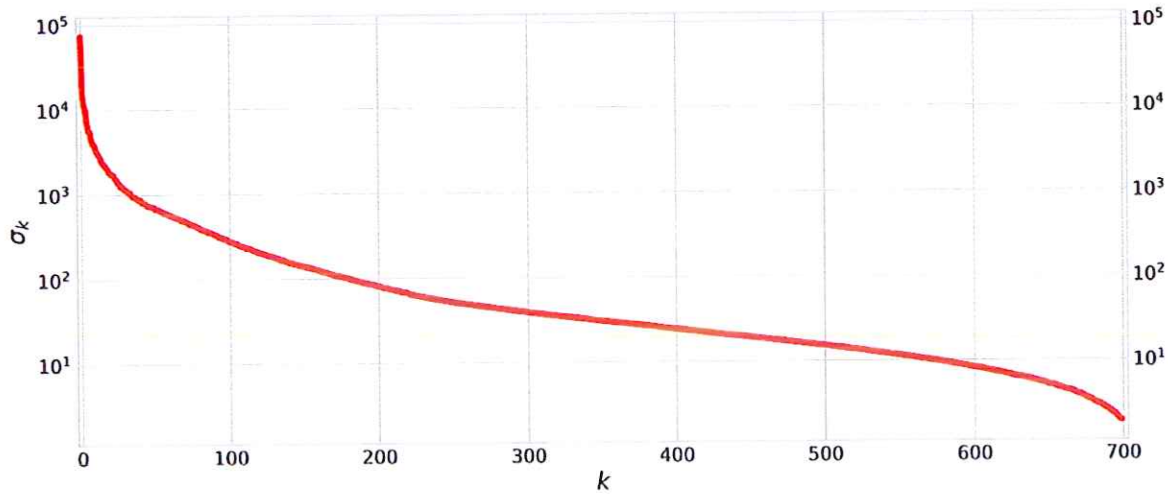
and $\sigma_1, \dots, \sigma_r$ are singular values of A.

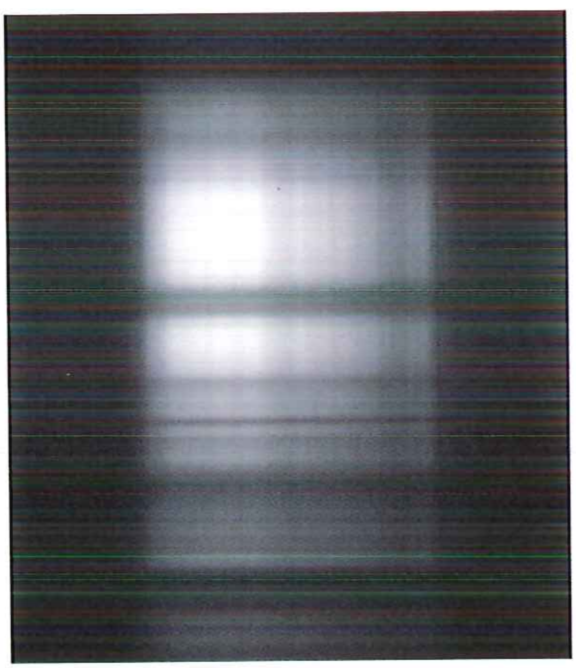
2) Replace A by the matrix

$$B_k = \sigma_1(\mathbf{u}_1\mathbf{v}_1^T) + \ldots + \sigma_k(\mathbf{u}_k\mathbf{v}_k^T)$$

for some $1 \leq k \leq 700$. This matrix can be stored using $k \cdot (700 + 800 + 1)$ numbers.

Singular values of the matrix A





matrix B_1
1501 bytes
compression 374:1

matrix B_5
7905 bytes
compression 75:1

matrix B_{10}
15,010 bytes
compression 37:1

matrix B_{20}
30,020 bytes
compression 18:1

matrix B_{50}
75,050 bytes
compression 7:1

matrix B_{100}
150,100 bytes
compression 4:1

How to compute SVD of a matrix A

Assume: $A = \underbrace{U}_{\text{orthogonal}} \cdot \underbrace{\Sigma}_{\text{diagonal}} \cdot \underbrace{V^T}_{\text{orthogonal}}$

Then: $A^TA = (U\Sigma V^T)^T \cdot (U\Sigma V^T) = (V^T)^T \Sigma^T \cancel{U^T U} \Sigma V = V(\Sigma^T\Sigma)V^T$

Note: $\Sigma^T\Sigma$ is a diagonal matrix with squares of singular values on the diagonal.

e.g.: $\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}$

$$\Sigma^T \cdot \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We obtain: $\underbrace{A^TA}_{\text{symmetric}} = \underbrace{V}_{\text{orthogonal}} \underbrace{(\Sigma^T\Sigma)}_{\text{diagonal}} V^T$ ← This is an orthogonal diagonalization of A^TA. We know how to compute.

This gives matrices V and Σ:

(columns of V) = (orthogonal eigenvectors of A^TA)

$\left(\begin{array}{l}\text{diagonal entries of } \Sigma \\ \text{i.e. singular values of } A\end{array}\right) = \sqrt{\text{eigenvalues of } A^TA}$

It remains to compute the matrix U

$A = U\Sigma V^T$ gives: $AV = U\Sigma$

Note: If $U = [u_1 \dots u_m]$ $V = [v_1 \dots v_n]$ $\sigma_1, \dots, \sigma_r$ - non-zero singular values of A

then: $AV = [Av_1\ Av_2 \dots Av_n]$

$U\Sigma = [\sigma_1 u_1\ \sigma_2 u_2 \dots \sigma_r u_r\ 0 \dots 0]$

So: $u_1 = \frac{1}{\sigma_1}(Av_1)$ $u_2 = \frac{1}{\sigma_2}(Av_2), \dots, u_r = \frac{1}{\sigma_r}(Av_r)$

Vectors $u_{r+1}, \dots, u_m$ can be chosen in an arbitrary way so that $\{u_1, \dots, u_r, u_{r+1}, \dots, u_m\}$ is an orthonormal basis of $\mathbb{R}^m$.

How to compute SVD of a matrix A

1) Compute an orthogonal diagonalization of the symmetric $n \times n$ matrix $A^T A$:

$$A^T A = QDQ^T$$

such that eigenvalues on the diagonal of the matrix D are arranged from the largest to the smallest. We set $V = Q$.

2) If

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

then $\sigma_i = \sqrt{\lambda_i}$. This gives the matrix Σ.

Note: if $n > m$ then we use only $\lambda_1, \dots, \lambda_m$. The remaining eigenvalues $\lambda_{m+1}, \dots, \lambda_n$ of D will be equal to 0 in this case.

3) Let $V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$, and let $\sigma_1, \dots, \sigma_r$ be non-zero singular values of A. The first r columns of the matrix $U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix}$ are given by

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

The remaining columns $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ can be added arbitrarily so that U is an orthogonal matrix (i.e.$\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$) is an orthonormal basis of $\mathbb{R}^m$.

Example. Find SVD of the following matrix:

$$A = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$$

$\underset{3\times 2}{A} = U\Sigma V^T$

$\underset{3\times 3}{U}$ $\quad \underset{3\times 2}{\Sigma} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix}$ $\quad \underset{2\times 2}{V}$

① Compute an orthogonal diagonalization of A^TA:

$$A^TA = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$P(\lambda) = \det(A^TA - \lambda I) = \lambda^2 - 4\lambda + 3$ eigenvalues of A^TA: $\lambda_1 = 3, \lambda_2 = 1$

$(\text{basis of eigensp. for } \lambda_1 = 3) = \left\{ \overset{w_1}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}} \right\}$ $\quad \|w_1\| = \sqrt{2}$

$(\text{basis of eigensp for } \lambda_1 = 1) = \left\{ \underset{w_2}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}} \right\}$ $\quad \|w_2\| = \sqrt{2}$

We get: $A^TA = VDV^T$ where $V = \left[\frac{w_1}{\|w_1\|} \ \frac{w_2}{\|w_2\|} \right] = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

$$D = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

σ_1^2 $\quad \sigma_2^2$

② We obtain:

$$V = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

③ Compute U:

Let $U = [u_1 \ u_2 \ u_3]$

$$\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ -2/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \underbrace{\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{V} = [u_1 \ u_2 \ u_3] \cdot \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = [\sqrt{3}u_1 \ \ 1 \cdot u_2]$$

So: $u_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1/\sqrt{2} \\ -2/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$ $\quad u_2 = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$

$u_3 = ?$

- Start with a vector z_3 lin. indep. of u_1, u_2.
  In this example we can use e.g. $z_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
- $\{u_1, u_2, z_3\}$ is an basis of $\mathbb{R}^3$. Use Gram-Schmidt process to make it into an orthogonal basis.
  Since u_1, u_2 are already orthogonal it suffices to modify z_3:

$$w_3 = z_3 - \left(\frac{z_3 \cdot u_1}{u_1 \cdot u_1}\right) u_1 - \left(\frac{z_3 \cdot u_2}{u_2 \cdot u_2}\right) u_2$$

$z_3 \cdot u_1 = 1/\sqrt{6}$, $u_1 \cdot u_1 = 1$, $z_3 \cdot u_2 = -\frac{1}{\sqrt{2}}$, $u_2 \cdot u_2 = 1$

so:

$$w_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{6}} \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} - \left(-\frac{1}{\sqrt{2}}\right) \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/3 \\ -1/3 \\ 1/3 \end{bmatrix}$$

Take $u_3 = \frac{w_3}{\|w_3\|}$

$\|w_3\| = \frac{1}{\sqrt{3}}$ so: $u_3 = \sqrt{3} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$

We obtain:

$$A = \underbrace{\begin{array}{c} \begin{array}{ccc} u_1 & u_2 & u_3 \end{array} \\ \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix} \end{array}}_{U} \cdot \underbrace{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \cdot \underbrace{\begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{V^T}$$