Principal Components and the Best Low Rank Matrix

CE282: Linear Algebra

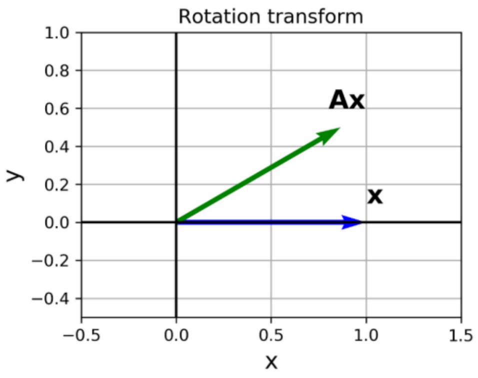
Computer Engineering Department

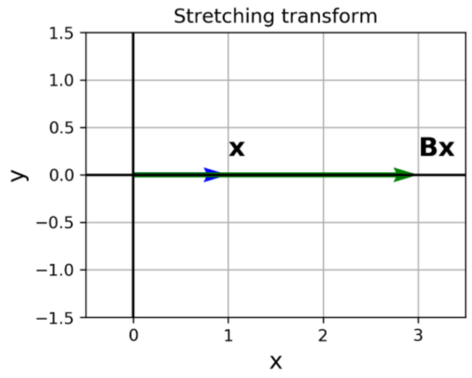
Sharif University of Technology

Hamid R. Rabiee

Maryam Ramezani

- Matrix A as a transformation that acts on a vector x.

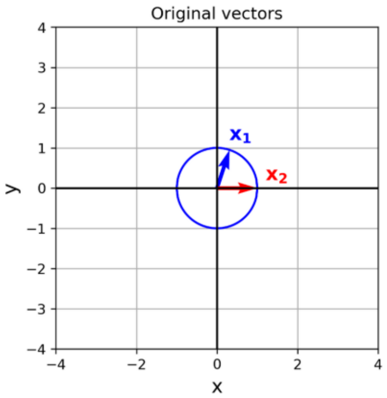


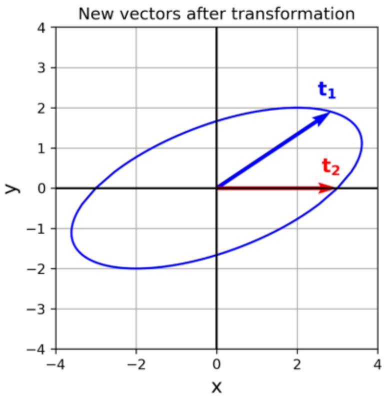


- A circle that contains all the vectors that are on unit away from the origin

$$x = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \; where \; x_i^2 + y_i^2 = 1$$

$$A = \begin{bmatrix} 3 & 2 \\ 0 & 2 \end{bmatrix}$$

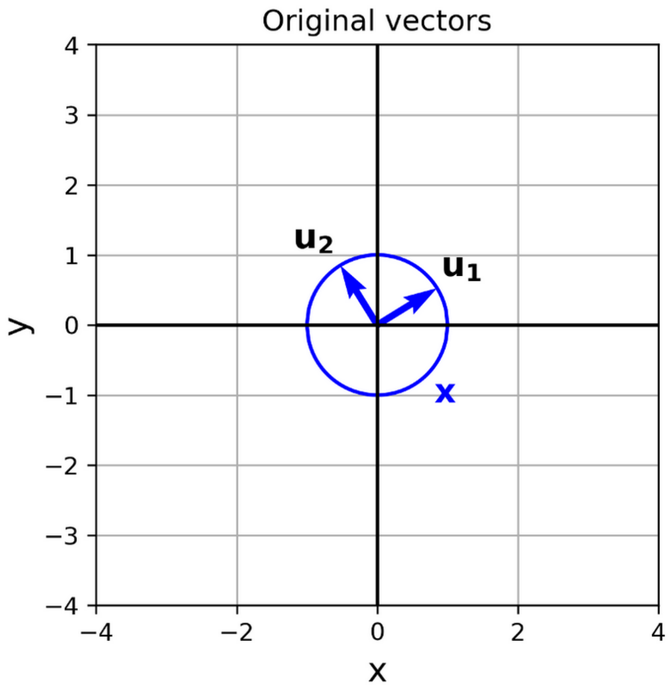


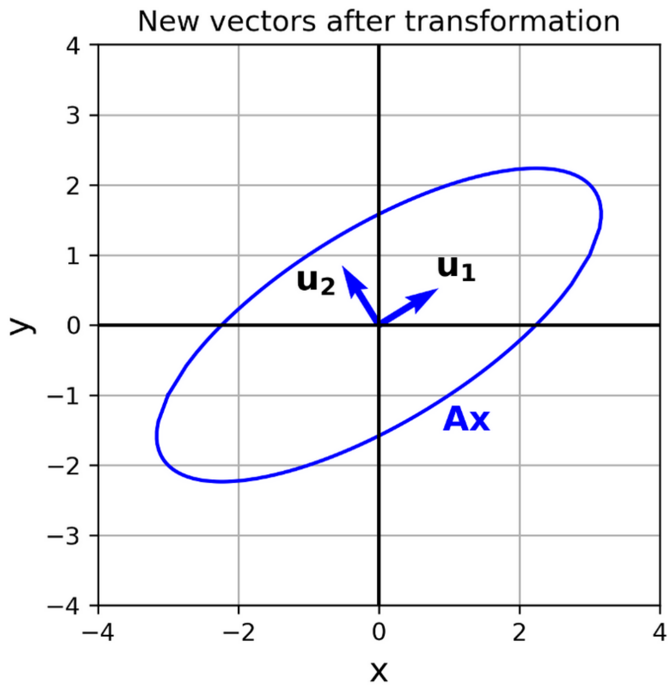


- For matrix $\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$

$$\boldsymbol{u_1} = \begin{bmatrix} 0.8507 \\ 0.5257 \end{bmatrix} \qquad \boldsymbol{u_2} = \begin{bmatrix} -0.5257 \\ 0.8507 \end{bmatrix}$$

$$\lambda_1 = 3.618 \qquad \lambda_2 = 1.382$$



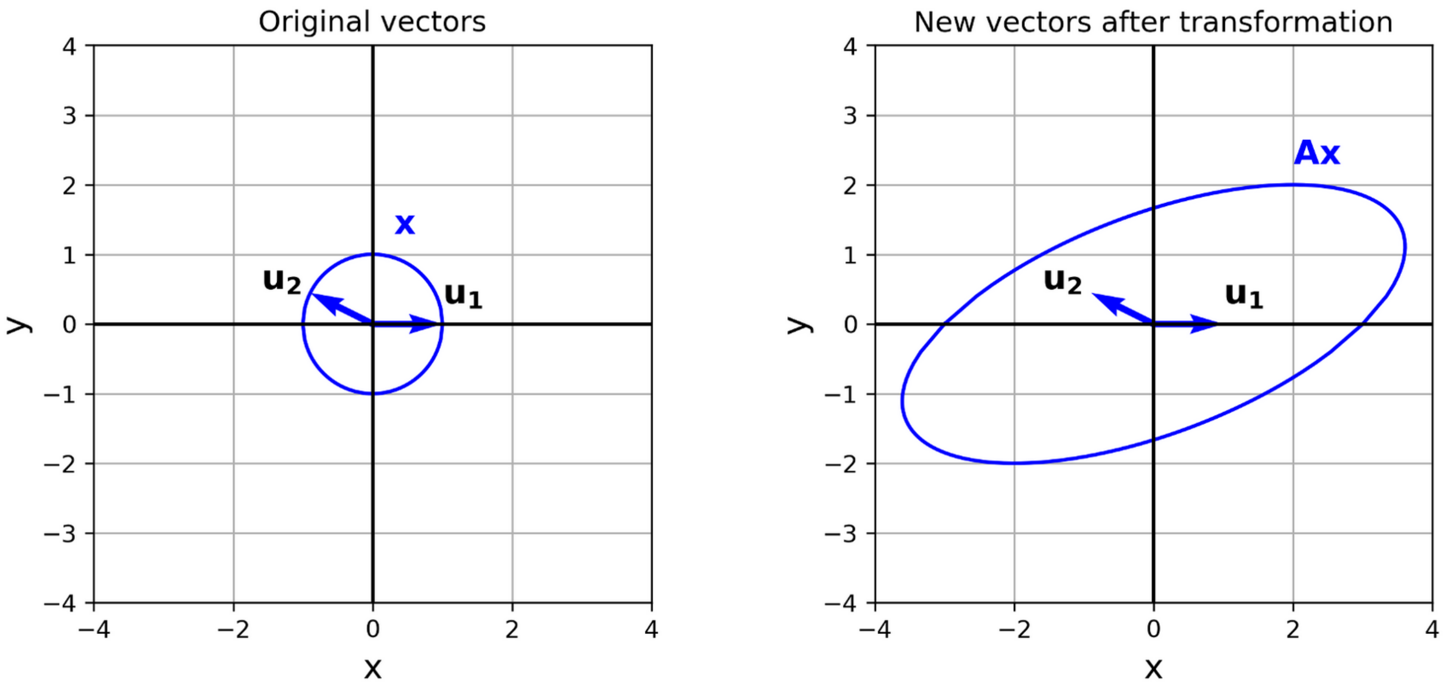


Eigenvector

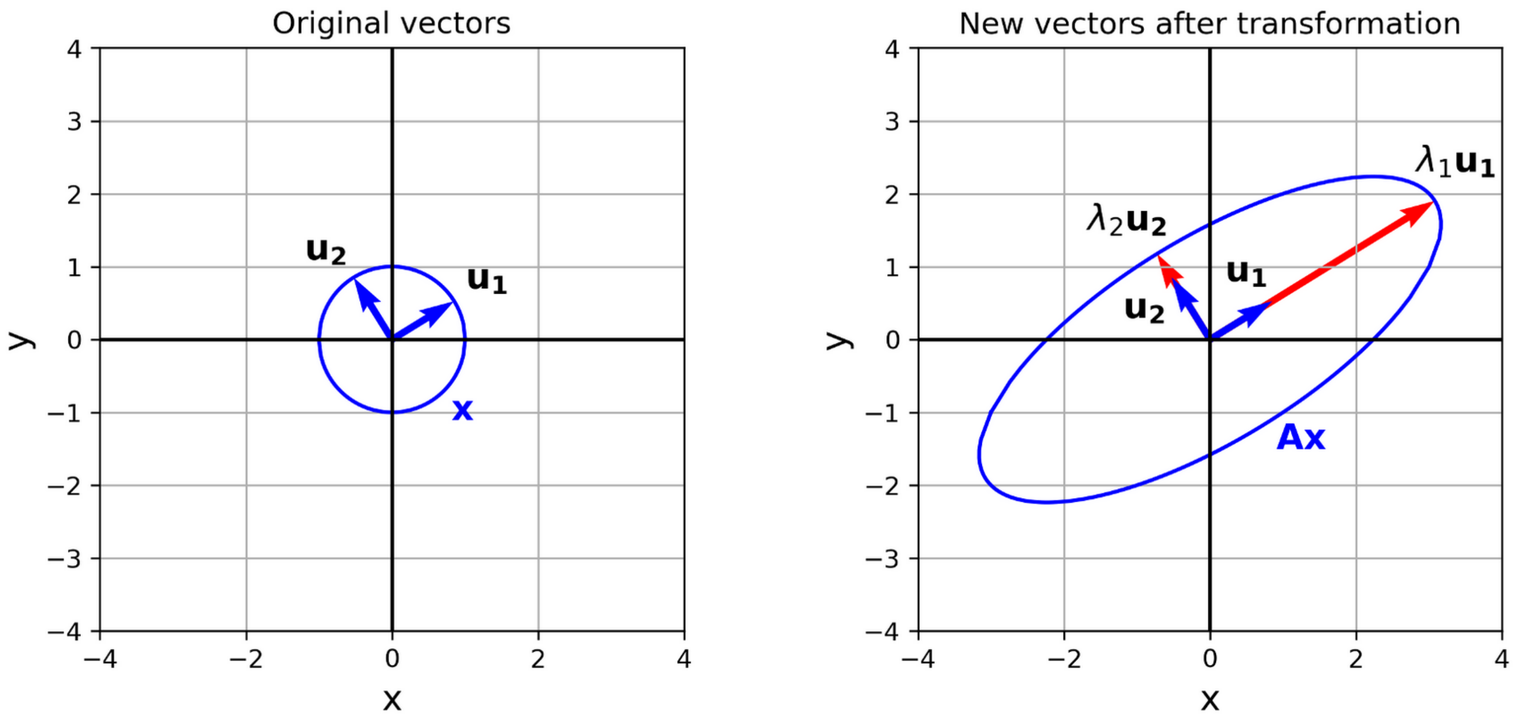
- For matrix B= $\begin{bmatrix} 3 & 2 \\ 0 & 2 \end{bmatrix}$

$$\boldsymbol{u_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \boldsymbol{u_2} = \begin{bmatrix} -0.8944 \\ 0.4472 \end{bmatrix}$$

$$\lambda_1 = 3 \qquad \lambda_2 = 2$$



- A symmetric matrix transforms a vector by stretching or shrinking it along its eigenvectors

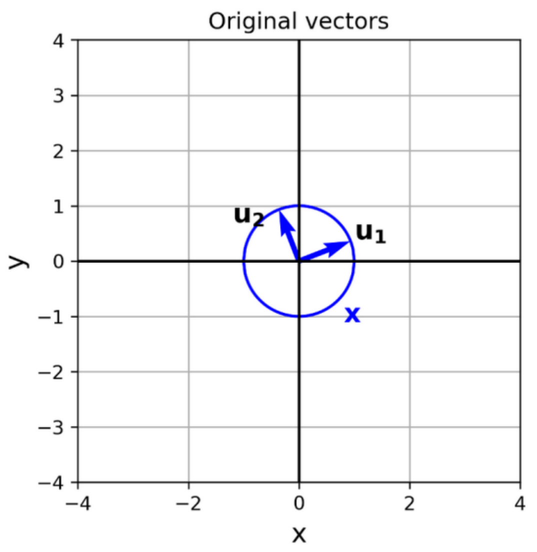


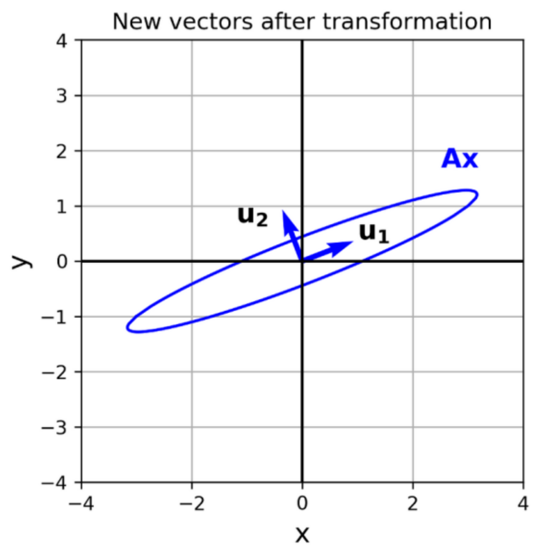
- If the absolute value of an eigenvalue is greater that 1, the circle x stretches along it, and if the absolute value is less than 1, it shrinks along it.

$$\boldsymbol{C} = \begin{bmatrix} 3 & 2 \\ 0 & 2 \end{bmatrix}$$

$$\boldsymbol{u_1} = \begin{bmatrix} 0.9327 \\ 0.3606 \end{bmatrix} \qquad \boldsymbol{u_2} = \begin{bmatrix} -0.3606 \\ 0.9327 \end{bmatrix}$$

$$\lambda_1 = 3.3866 \qquad \lambda_2 = 0.4134$$





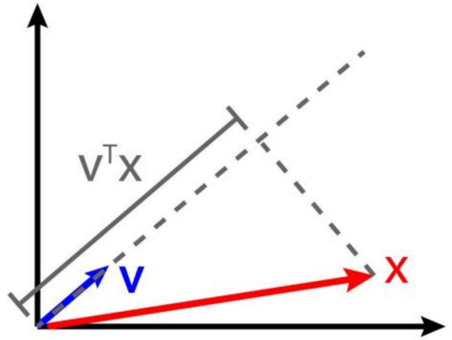
Geometrical Interpretation of Eigendecomposition

- A symmetric matrix is orthogonally diagonalizable. It means that if we have an $n \times n$ symmetric matrix $\boldsymbol{A}$, we can decompose it as

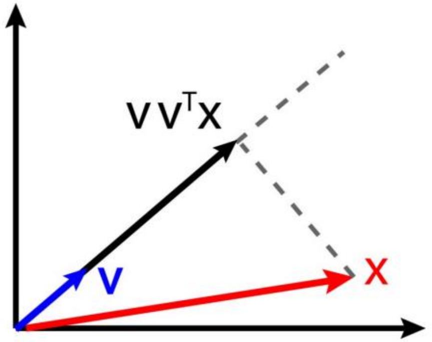
$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^T$$

$$\boldsymbol{A} = [\boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \cdots \quad \boldsymbol{u_n}] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} [\boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \cdots \quad \boldsymbol{u_n}]^T$$

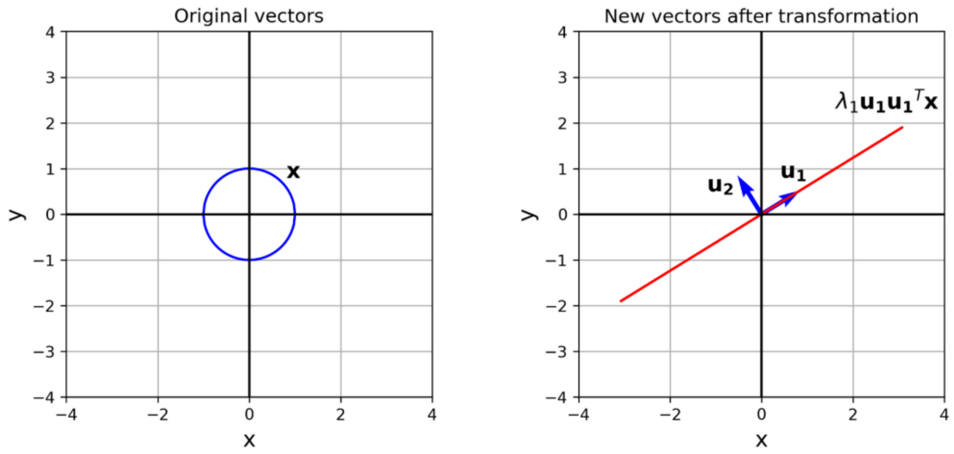
Projection of x onto v



Orthogonal projection of x onto v

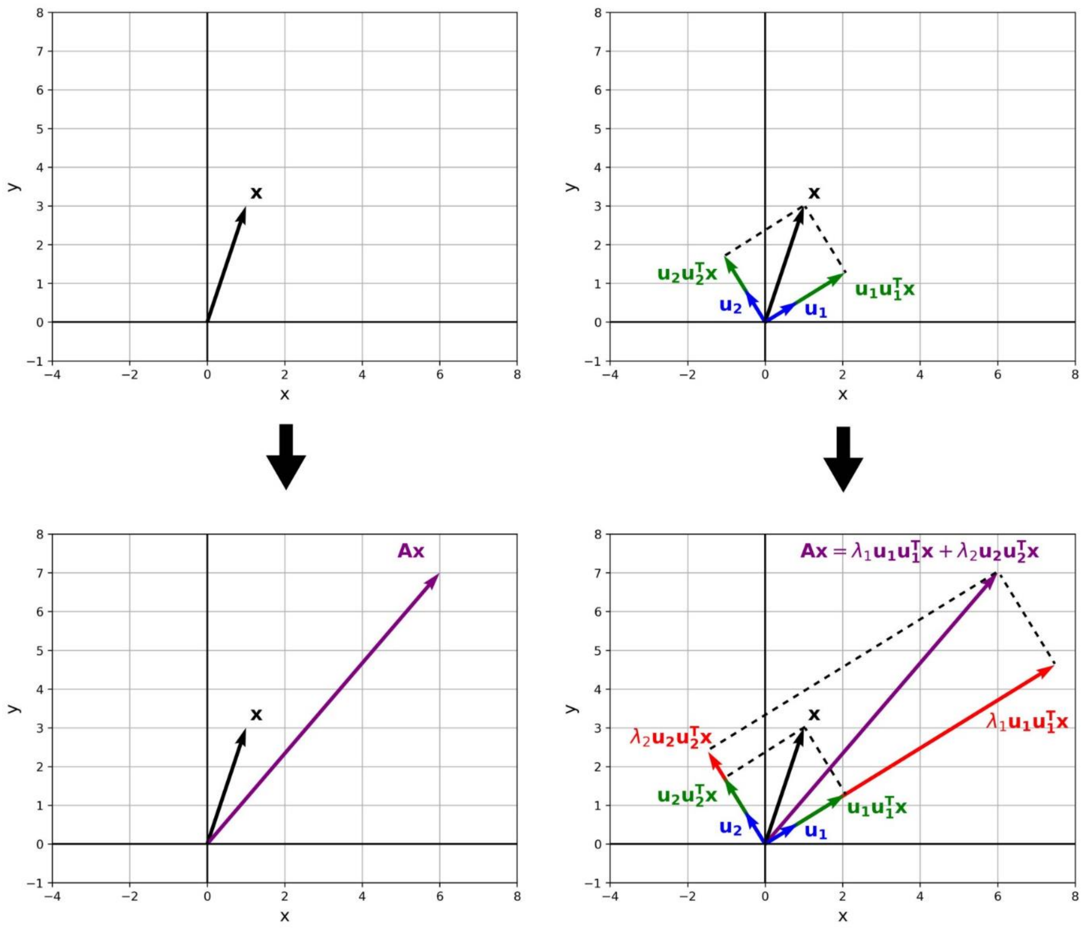


- All the projection matrices in the eigendecomposition equation are symmetric with rank 1.
- So we conclude that each matrix $\lambda_i \boldsymbol{u_i} \boldsymbol{u_i}^T$ in the eigendecomposition equation is a symmetric $n \times n$ matrix with n eigenvectors. The eigenvectors are the same as the original matrix $\boldsymbol{A}$ which are $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_n}$. The corresponding eigenvalue of $\boldsymbol{u_i}$ is λ_i (which is the same as $\boldsymbol{A}$), but all the other eigenvalues are zero.



- Symmetric matrix A

$$\boldsymbol{A}x = \lambda_1 \boldsymbol{u_1}\boldsymbol{u_1^T}x + \lambda_2 \boldsymbol{u_2}\boldsymbol{u_2^T}x + \cdots + \lambda_n \boldsymbol{u_n}\boldsymbol{u_n^T}x$$



Approximation

- $\boldsymbol{A} = \lambda_1 \boldsymbol{u_1 u_1^T} + \lambda_2 \boldsymbol{u_2 u_2^T} + \cdots + \lambda_n \boldsymbol{u_2 u_2^T} + \cdots + \lambda_n \boldsymbol{u_n u_n^T}$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda\_n$$

- $\boldsymbol{A} \approx \boldsymbol{A}_k = \lambda_1 \boldsymbol{u_1 u_1} + \lambda_n \boldsymbol{u_2 u_2^T} + \cdots + \lambda_n \boldsymbol{u_k u_k^T}$

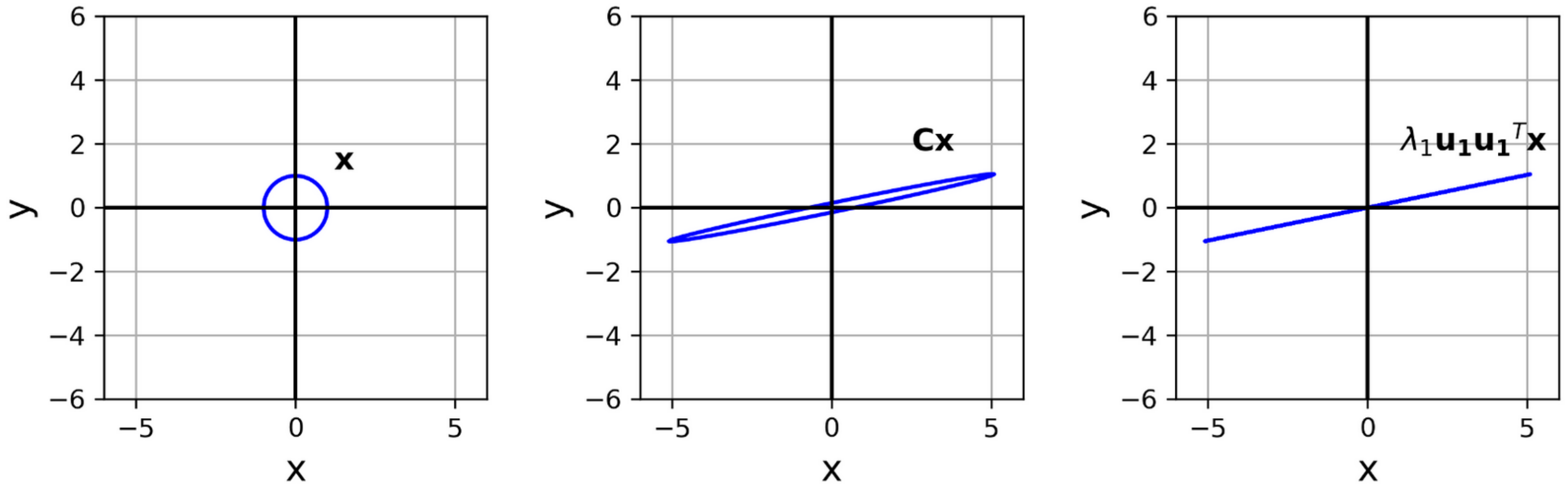
- $\boldsymbol{A} \approx \boldsymbol{A}_k = P_k D_k P_k^T = [\boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \ldots \quad \boldsymbol{u_k}] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u_1^T} \\ \boldsymbol{u_2^T} \\ \vdots \\ \boldsymbol{u_k^T} \end{bmatrix}$

$$= \lambda_1 \boldsymbol{u_1 u_1} + \lambda_n \boldsymbol{u_2 u_2^T} + \cdots + \lambda_n \boldsymbol{u_k u_k^T}$$

$$\boldsymbol{C} = \begin{bmatrix} 5 & 1 \\ 0 & 2 \end{bmatrix}$$

$$\boldsymbol{u_1} = \begin{bmatrix} 0.9794 \\ 0.2017 \end{bmatrix} \qquad \boldsymbol{u_2} = \begin{bmatrix} -0.2017 \\ 0.9794 \end{bmatrix}$$

$$\lambda_1 = 5.2059 \qquad \lambda_2 = 0.1441$$



Non-symmetric matrix

- No real eigenvalues

$$\begin{bmatrix} 3 & -1 \\ 1 & 2 \end{bmatrix}$$

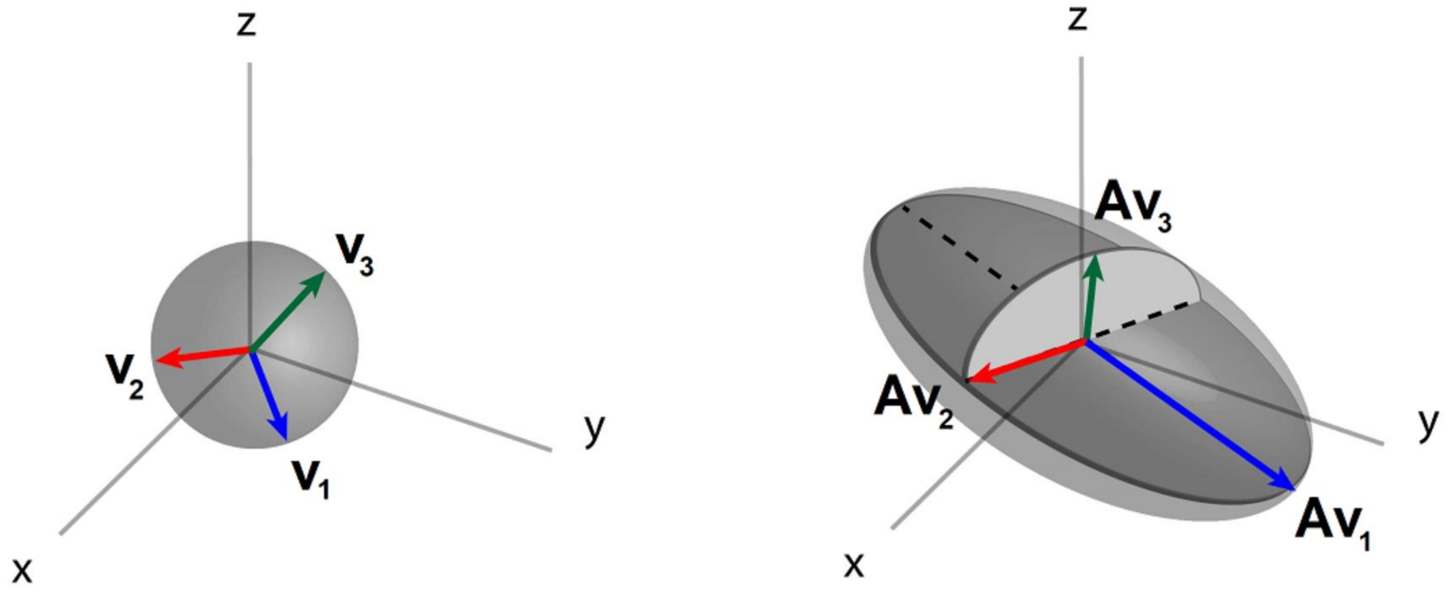
- The eigenvectors are not linearly independent

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$$

- The eigenvectors are linearly independent, but they are not orthogonal

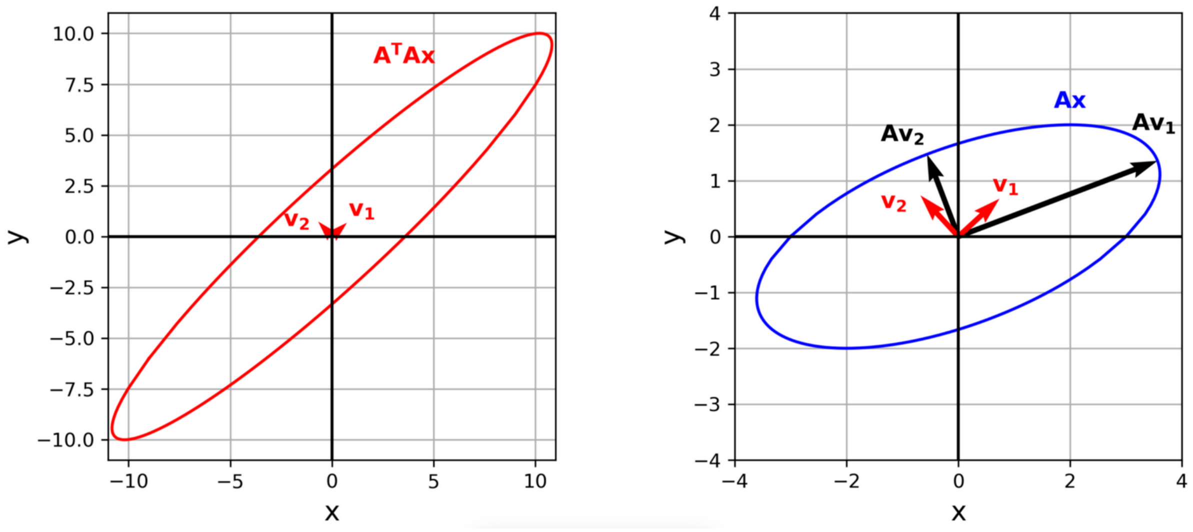
$$\begin{bmatrix} 3 & 2 \\ 0 & 2 \end{bmatrix}$$

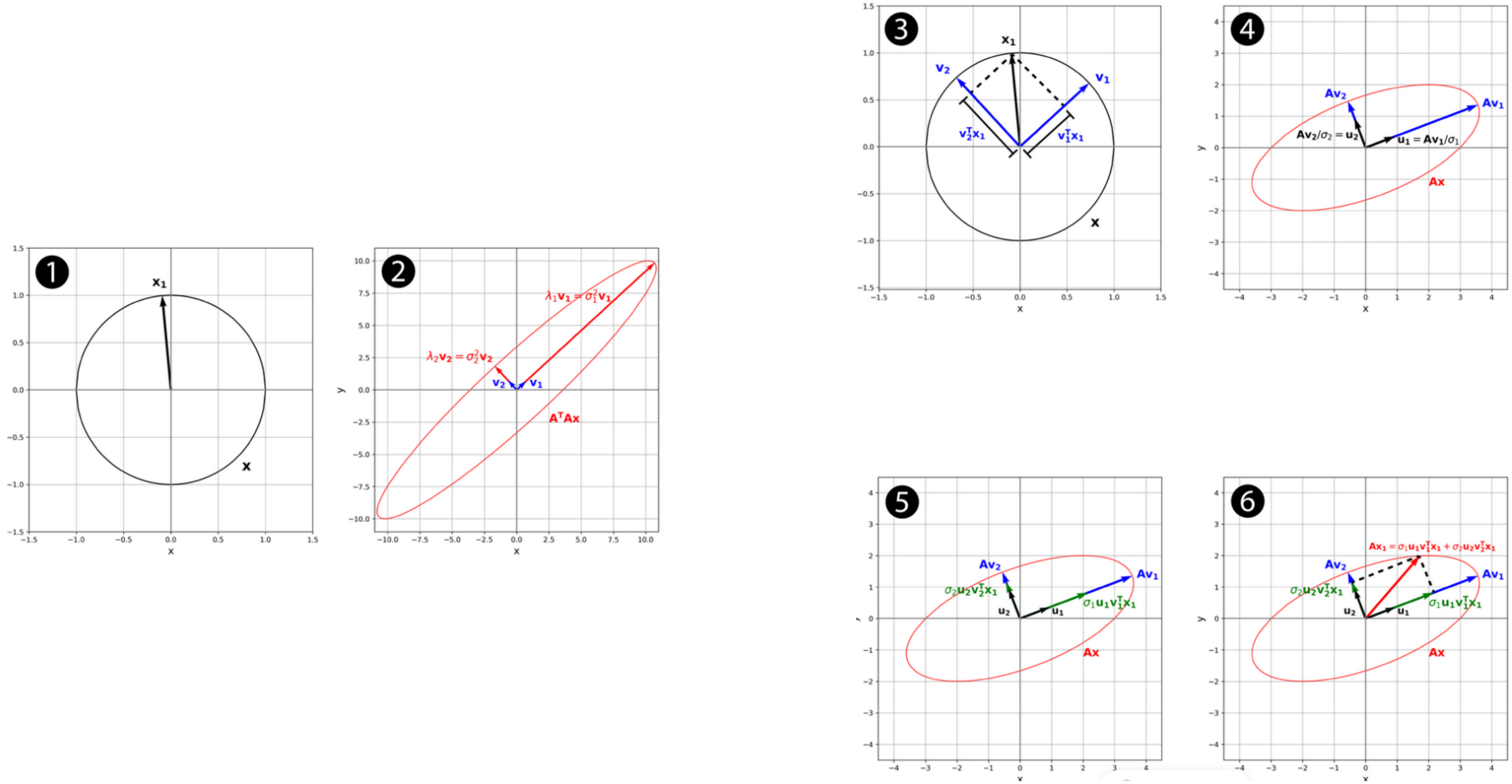
- For non-symmetric matrix, gram matrix is symmetric.



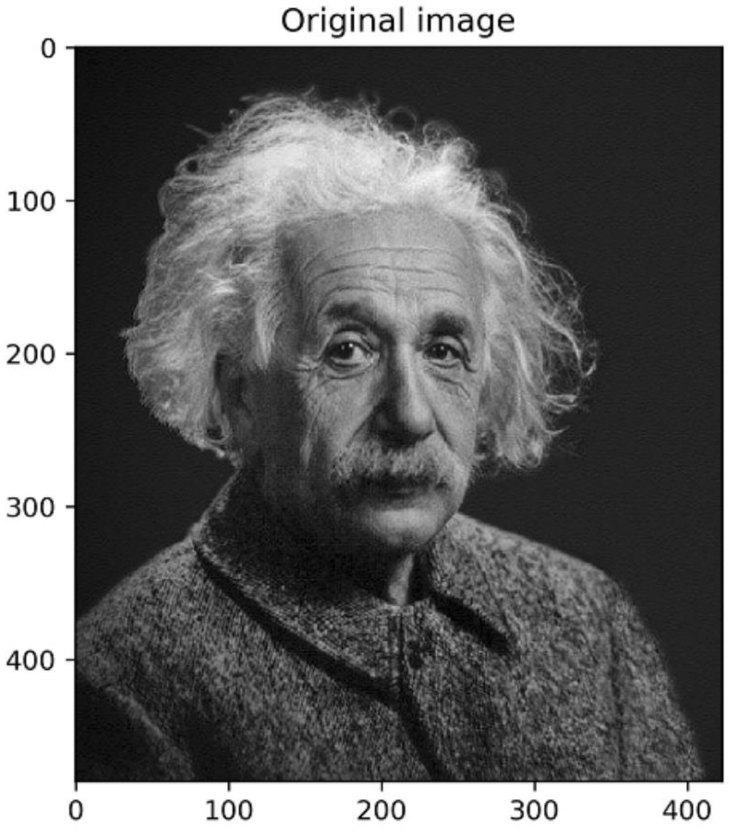
- Now let $\boldsymbol{A}$ be an $m \times n$ matrix. We showed that $\boldsymbol{A}^T\boldsymbol{A}$ is a symmetric matrix, so it has n eigenvalues and n linearly independent and orthogonal eigenvectors which can form a basis for the n-element vectors that it can transform (in R^n space). We call these eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ and we assume they are normalized.

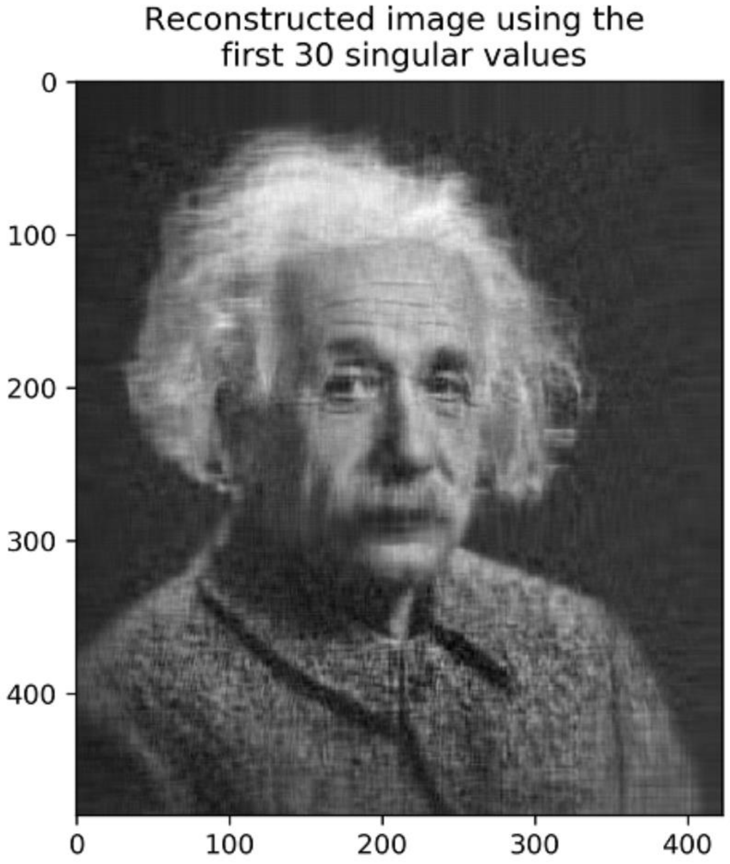
- Back to page 4



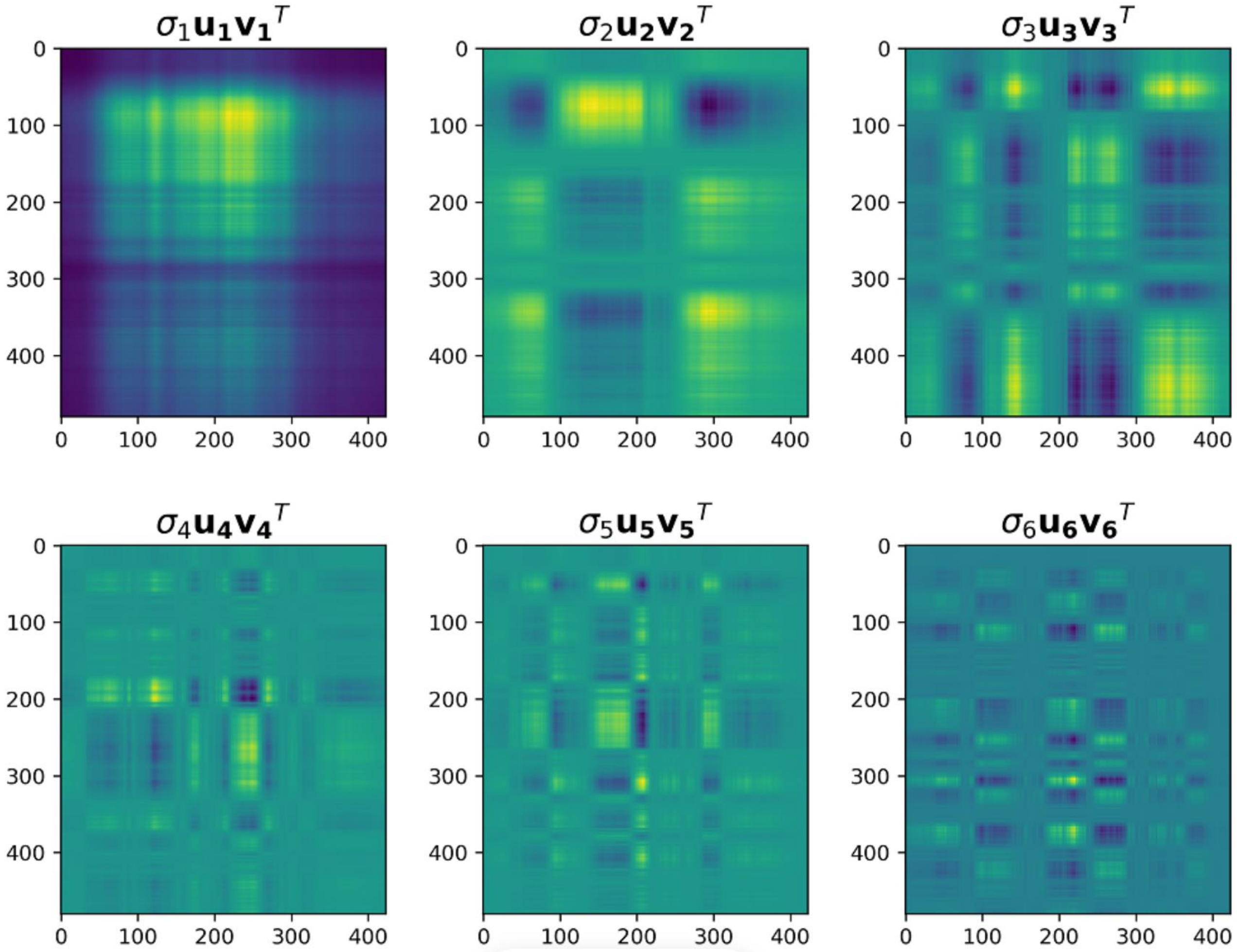

1
$\mathbf{x_1}$
$\mathbf{x}$
2
$\lambda_1\mathbf{v_1} = \sigma_1^2\mathbf{v_1}$
$\lambda_2\mathbf{v_2} = \sigma_2^2\mathbf{v_2}$
$\mathbf{v_2}$
$\mathbf{v_1}$
$\mathbf{A^TAx}$
3
$\mathbf{x_1}$
$\mathbf{v_2}$
$\mathbf{v_1}$
$\mathbf{v_2^Tx_1}$
$\mathbf{v_1^Tx_1}$
$\mathbf{x}$
4
$\mathbf{Av_2}$
$\mathbf{Av_1}$
$\mathbf{Av_2}/\sigma_2 = \mathbf{u_2}$
$\mathbf{u_1} = \mathbf{Av_1}/\sigma_1$
$\mathbf{Ax}$
5
$\mathbf{Av_2}$
$\mathbf{Av_1}$
$\sigma_2\mathbf{u_2v_2^Tx_1}$
$\mathbf{u_2}$
$\mathbf{u_1}$
$\sigma_1\mathbf{u_1v_1^Tx_1}$
$\mathbf{Ax}$
6
$\mathbf{Ax_1} = \sigma_1\mathbf{u_1v_1^Tx_1} + \sigma_2\mathbf{u_2v_2^Tx_1}$
$\mathbf{Av_2}$
$\mathbf{Av_1}$
$\sigma_2\mathbf{u_2v_2^Tx_1}$
$\mathbf{u_2}$
$\mathbf{u_1}$
$\sigma_1\mathbf{u_1v_1^Tx_1}$
$\mathbf{Ax}$

Application: Dimensionality Reduction

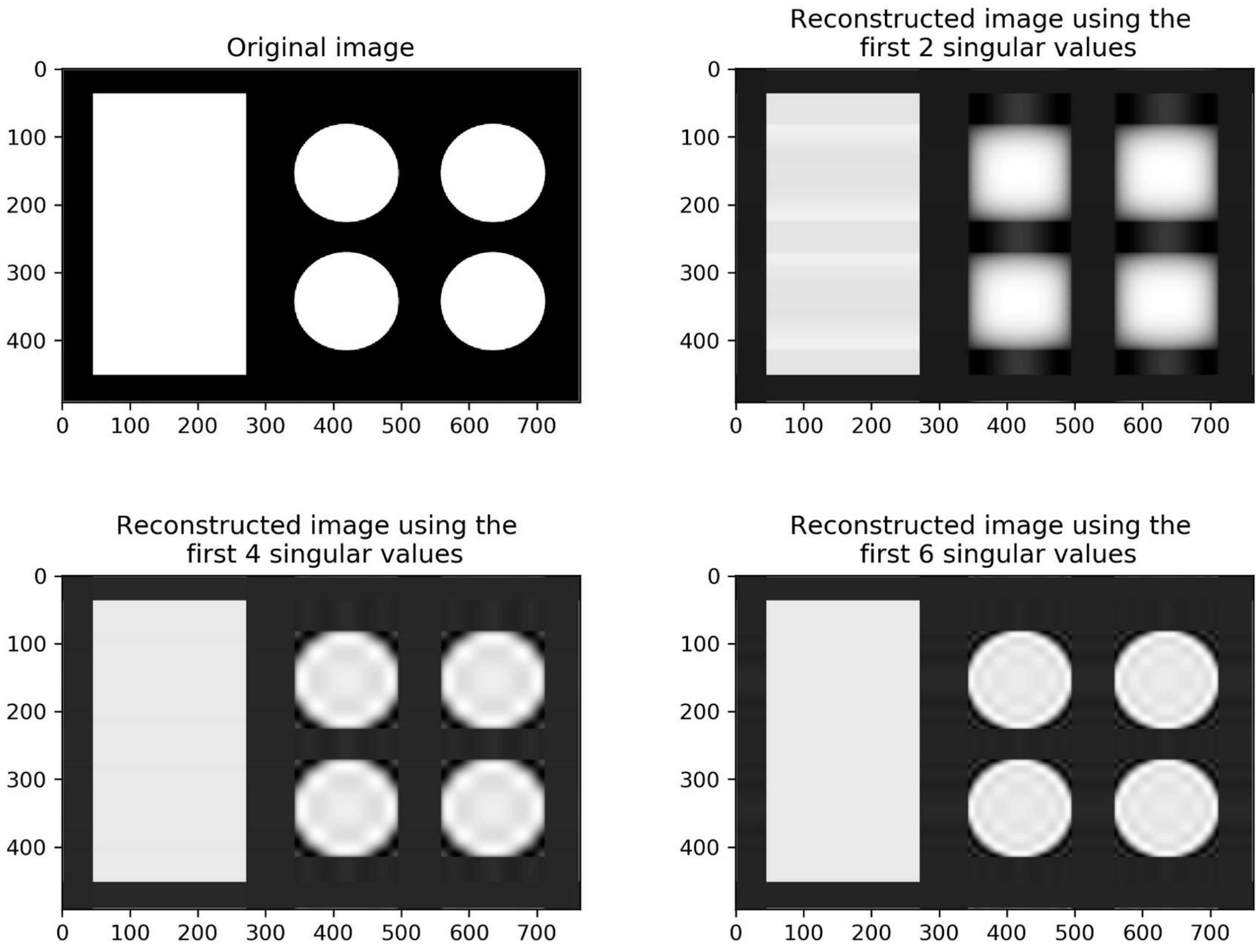




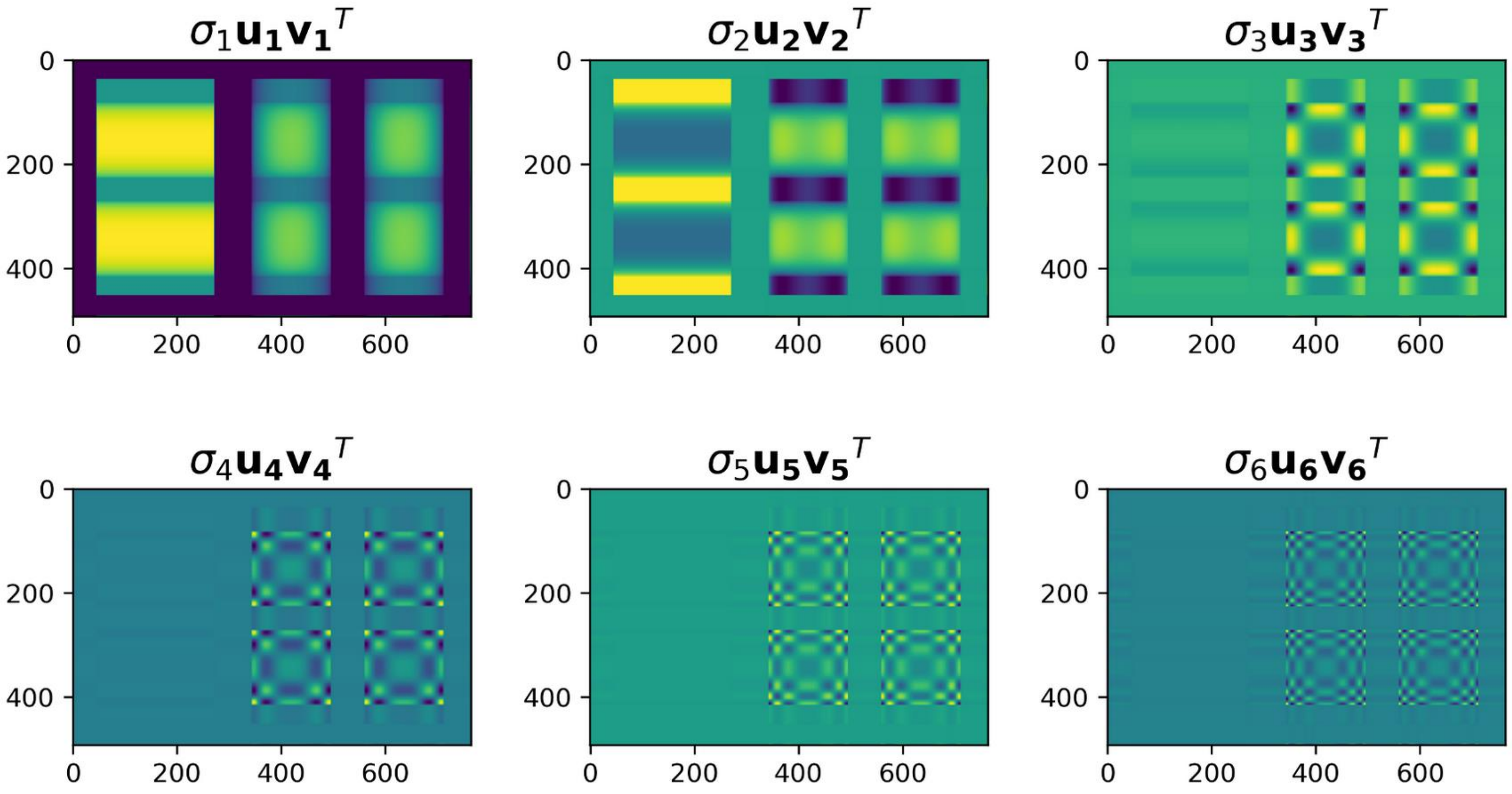
Application: Dimensionality Reduction



Application: Dimensionality Reduction



Application: Dimensionality Reduction



PCA

- PCA is a useful way to summarize high-dimensional data (*repeated observations of multiple variables*)

Important

The central ideas of PCA are **orthonormal coordinate** systems, the distinction between **variance and covariance**, and the possibility of choosing an orthonormal basis to **eliminate covariance**. Technically, PCA may be performed either by **eigenvector analysis** of the covariance matrix or by **singular value decomposition** of the original observation matrix.

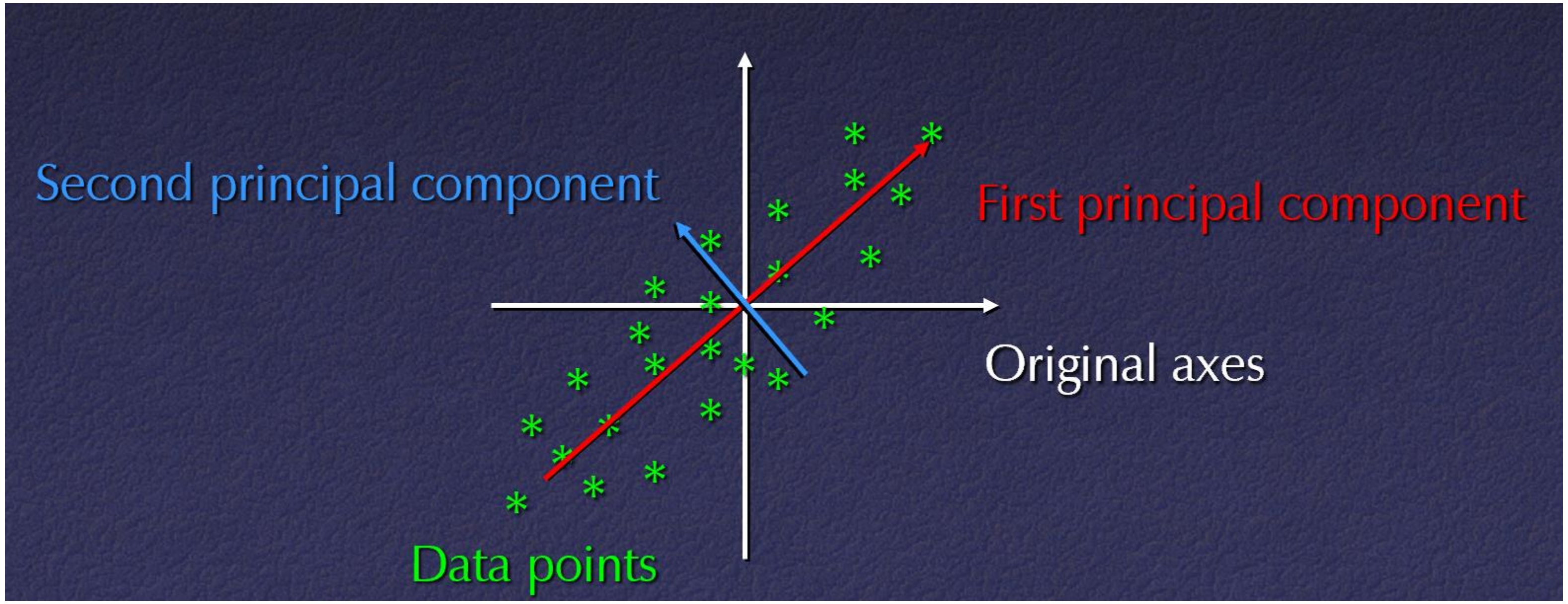
PCA and SVD

Note

https://towardsdatascience.com/pca-and-svd-explained-with-numpy-5d13b0d2a4d8

Principal Components Analysis (PCA)

Approximating a high-dimensional data set with a lower-dimensional subspace.



SVD and PCA

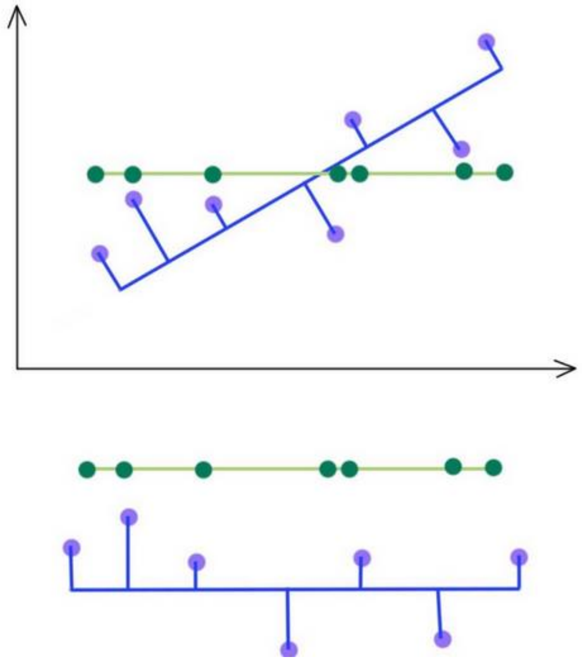
- Data matrix with points as rows, take SVD
  - Subtract out mean (“Whitening”)

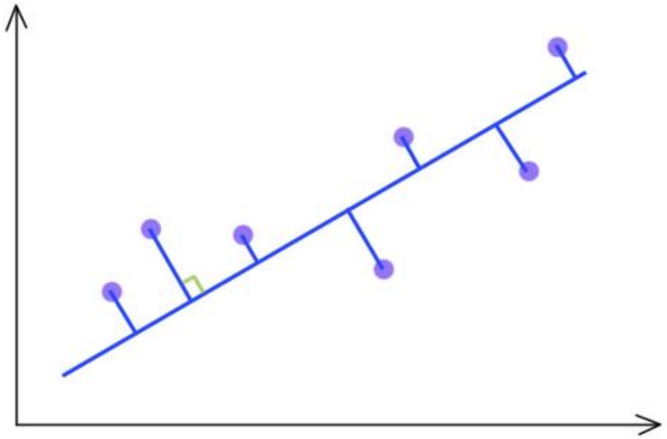
- Columns of V_k are principal components.

- Value of Σ_i give the importance of each component.

SVD and PCA

$$\begin{pmatrix} x_{11} & & x_{1n} \\ & \ddots & \\ x_{m1} & & x_{mn} \end{pmatrix} \rightarrow \begin{pmatrix} u_{11} & & u_{k1} & & u_{r1} \\ & \ddots & & \ddots & \\ u_{1m} & & u_{km} & & u_{rm} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_k & & \\ & & & \ddots & \\ 0 & & & & \sigma_r \end{pmatrix} \begin{pmatrix} v_{11} & & v_{1k} & & v_{1n} \\ & \ddots & & \ddots & \\ v_{r1} & & v_{rk} & & v_{rn} \end{pmatrix} \rightarrow \begin{pmatrix} u_{11} & & u_{k1} \\ & \ddots & \\ u_{1m} & & u_{km} \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ 0 & & \sigma_k \end{pmatrix} \begin{pmatrix} v_{11} & & v_{1k} \\ & \ddots & \\ v_{r1} & & v_{rk} \end{pmatrix}$$





PCA Proof

۱. (۱۵ نمره) در فضای R^n، m تا بردار $\{r_1, r_2, \dots, r_m\}$ داریم. اگر میانگین بردارهای r_i را $\bar{r}$ بنامیم، ماتریس $S_{n\times n}$ را به صورت $\frac{1}{m}\sum_{i=1}^{m}(r_i - \bar{r})(r_i - \bar{r})^T$ تعریف می‌کنیم. بردار یکه v را برحسب بردارهای ویژه و مقادیر ویژه ماتریس S به‌گونه‌ای به دست آورید که $v^T S v$ بیشینه شود.

Note

Proof PCA with variance and mean in attached video.

Low-Rank Approximation

- Another way to write the SVD (assuming for now $m > n$ for simplicity)

$$\boldsymbol{A} = \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \dots & \boldsymbol{u}_m \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \\ & & \vdots \\ & & 0 \end{pmatrix} \begin{pmatrix} \vdots & \boldsymbol{v}_1^T & \vdots \\ \vdots & \dots & \vdots \\ \vdots & \boldsymbol{v}_n^T & \vdots \end{pmatrix}$$

$$= \begin{pmatrix} \vdots & \cdots & \vdots \\ \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \cdots & \sigma_1 \boldsymbol{v}_1^T & \cdots \\ \vdots & \dots & \vdots \\ \cdots & \sigma_n \boldsymbol{v}_n^T & \cdots \end{pmatrix}$$

$$= \sigma_1 \boldsymbol{u}_1^T \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2^T \boldsymbol{v}_2^T + \cdots + \sigma_n \boldsymbol{u}_n^T \boldsymbol{v}_n^T$$

- The SVD writes the matrix A as a sum of outer products (of the left and right singular vectors.)

Low-Rank Approximation

- The best **rank-k** approximation for a $m \times n$ matrix $\boldsymbol{A}$, (where $k \leq \min(\mathrm{m}, \mathrm{n})$) is the one that minimizes the following problem:

$$\min_{A_k} \|A - A_k\|$$
$$such\ that\ rank(A_k) \leq k$$

- When using the induced 2-norm, the best **rank-k** approximation is given by:

$$\boldsymbol{A}_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$$
$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq 0$$

- Note that $rank(\boldsymbol{A}) = n$ and $rank(\boldsymbol{A}_k) = k$ and the norm of the difference between the matrix and its approximation is

$$\|\boldsymbol{A} - \boldsymbol{A}_k\|_2 = \left\|\sigma_{k+1} \boldsymbol{u}_{k+1} \boldsymbol{v}_{k+1}^T + \sigma_{k+2} \boldsymbol{u}_{k+2} \boldsymbol{v}_{k+2}^T + \cdots + \sigma_n \boldsymbol{u}_n \boldsymbol{v}_n^T\right\| = \sigma_{k+1}$$

Class Activity

What is the best rank-1 approximation for $X = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$

Or go to the link below
https://forms.gle/PzrVhK5TpxrmhwVbA

Timer: (2:30 minutes)

Example

What is the best rank-1 approximation for $X = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$

Solution

$$X_1 = \sigma_1 u_1 v_1^T = \sqrt{2} \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

In this problem, the approximation error under either norm (spectral or Frobenius) is the same: $\|X - X_1\| = \sigma_2 = 1$.

Matrix norms

- The Euclidean norm of an orthogonal matrix is equal to 1

$$\|U\|_2 = \max_{\|x\|_2=1} \|Ux\|_2 = \max_{\|x\|_2=1} \sqrt{(Ux)^T(Ux)} = \max_{\|x\|_2=1} \sqrt{x^T x} = \max_{\|x\|_2=1} \|x\|_2 = 1$$

- The Euclidean norm of a matrix is given by the largest singular value.

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \max_{\|x\|_2=1} \|U\Sigma V^T x\|_2 = \max_{\|x\|_2=1} \|\Sigma V^T x\,\|_2$$

$$= \max_{\|V^T x\|_2=1} \|\Sigma V^T x\|_2 = \max_{\|y\|_2=1} \|\Sigma y\|_2 = \max(\sigma_i)$$

Where we used the fact that $\|U\|_2 = 1, \|V\|_2 = 1$ and Σ is diagonal.

$$\|A\|_2 = \max(\sigma_i) = \sigma_{\max}$$

Norm for the inverse of a matrix

The Euclidean norm of the inverse of a square-matrix is given by

Assume here A is full rank, so that A^{-1} exists

$$\|A^{-1}\|_2 = \max_{\|x\|_2=1} \|(U\Sigma V^T)^{-1}x\|_2 = \max_{\|x\|_2=1} \|V\Sigma^{-1}U^T x\|_2$$

Since $\|U\|_2 = 1, \|V\|_2 = 1$ and Σ is diagonal then

$$\|A^{-1}\|_2 = \frac{1}{\sigma_{\min}}$$

$\sigma_{\min}$ is the smallest singular value

Norm of the pseudo-inverse

The norm of the pseudo-inverse of a $m \times n$ matrix is

$$\|A^+\|_2 = \frac{1}{\sigma_r}$$

Where σ_r is the smallest **non-zero** singular value. This is valid for any matrix, regardless of the shape or rank.

Note that for a full rank square matrix, $\|A^+\|_2$ is the same as $\|A^{-1}\|_2$.

Zero matrix: If A is a zero matrix, then A^+ is also the zero matrix, and $\|A^+\| = 0$.

SVD Summary

- The SVD is a factorization of a $m \times n$ matrix into $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$ where $\boldsymbol{U}$ is a $m \times m$ orthogonal matrix, $\boldsymbol{V^T}$ is a $n \times n$ orthogonal matrix and $\boldsymbol{\Sigma}$ is a $m \times n$ diagonal matrix.

- **In reduced form:** $\boldsymbol{A} = \boldsymbol{U_R \Sigma_R V_R^T}$, where $\boldsymbol{U_R}$ is a $m \times k$ matrix, $\boldsymbol{\Sigma_R}$ is a k $\times\, k$ matrix, and $\boldsymbol{V_R}$ is a $n \times k$ matrix, and $k = \min(m, n)$.

- The columns of $\boldsymbol{V}$ are the eigenvectors of the matrix $\boldsymbol{A^T A}$, denoted the right singular vectors.

SVD Summary

- The columns of $\boldsymbol{U}$ are the eigenvectors of the matrix $\boldsymbol{AA^T}$, denoted the left singular vectors.

- The diagonal entries of $\boldsymbol{\Sigma^2}$ are the eigenvalues of $\boldsymbol{A^T A}$. $\sigma_i = \sqrt{\lambda_i}$ are called the singular values.

- The singular values are always non-negative (since $A^T A$ is a positive semi-definite matrix, the eigenvalues are always $\lambda \geq 0$)

Example

$$C = \begin{bmatrix} 3 & -2 \\ -1 & 3 \end{bmatrix}$$

Solution

$$\|C\|_1 = \max_{\|x\|_1} \|Cx\|_1$$

$$\|C\|_1 = \max_{1\le j\le 3} \sum_{i=1}^{3} |C_{ij}|$$

$$\|C\|_1 = \max(|3| + |-1|, |-2| + 3)$$

$$\|C\|_1 = \max_{\|x\|_1}(4, 5)$$

$$\|C\|_1 = 5$$

$$\|C\|_\infty = \max(|3| + |-2|, |-1| + |3|)$$

$$\|C\|_\infty = \max(5, 4)$$

$$\|C\|_\infty = 5$$

$$\|C\|_2 = \max_{\|x\|_2} \|Cx\|_2$$

$$\det(C^T C - \lambda I) = 0$$

$$\det\left(\begin{bmatrix} 3 & -1 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -1 & 3 \end{bmatrix} - \lambda I\right) = 0$$

$$\det\left(\begin{bmatrix} 9+1 & -6-3 \\ -3-6 & 4+9 \end{bmatrix} - \lambda I\right) = 0$$

$$\det\left(\begin{bmatrix} 10-\lambda & -9 \\ -9 & 13-\lambda \end{bmatrix}\right) = 0$$

$$(10-\lambda)(13-\lambda) - 81 = 0$$

$$\lambda^2 - 23\lambda + 130 - 81 = 0$$

$$\lambda^2 - 23\lambda + 49 = 0$$

$$\left(\lambda - \frac{1}{2}\left(23 + 3\sqrt{37}\right)\right)\left(\lambda - \frac{1}{2}\left(23 - 3\sqrt{37}\right)\right) = 0$$

$$\|C\|_2 = \sqrt{\lambda_{\max}} = \sqrt{\frac{1}{2}\left(23 + 3\sqrt{37}\right)} \approx 4.54$$

Matrix norm and low-rank approximation

Theorem

Let A be a square, symmetric matrix, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Show that

$$\|A\|_2 = \max(|\lambda_1|, |\lambda_n|)$$

Proof:

It suffices to show that the singular values of A are given by $|\lambda_1|, \ldots, |\lambda_n|$. To See this, consider the orthogonal diagonalization of $A = \mathrm{Q\Lambda Q^T}$. It follows that

$$AA^T = A^2 = Q\Lambda^2 Q^T$$

Therefore, the squared singular values of A (which are eigenvalues of AA^T) coincide with the eigenvalues of A^2, i.e., $\sigma_1 = |\lambda_1|, \ldots, \sigma_n = |\lambda_n|$ (still not sorted, but the largest singular value must be the larger one of $|\lambda_1|, |\lambda_n|$.)

Compressing images using Linear Algebra

https://medium.com/analytics-vidhya/compressing-images-using-linear-algebra-bdac64c5e7ef