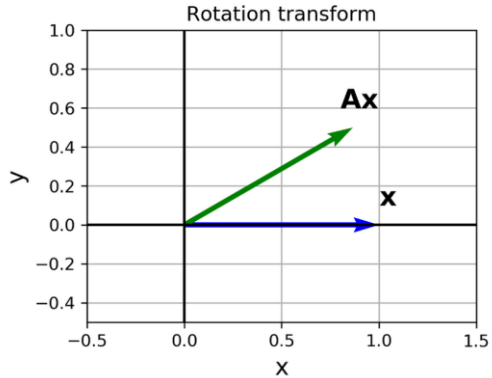
Principal Components and the Best Low Rank Matrix

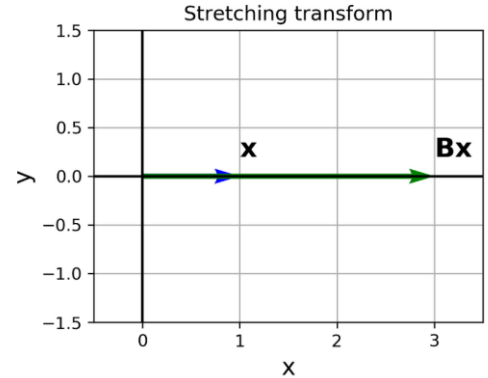
CE40282-1: Linear Algebra

Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

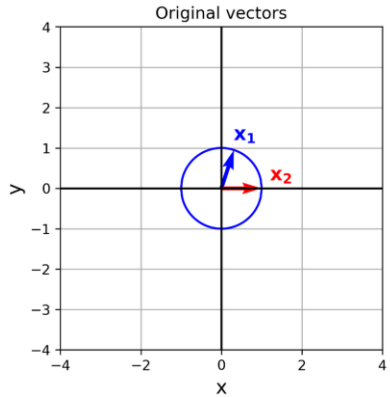
Review

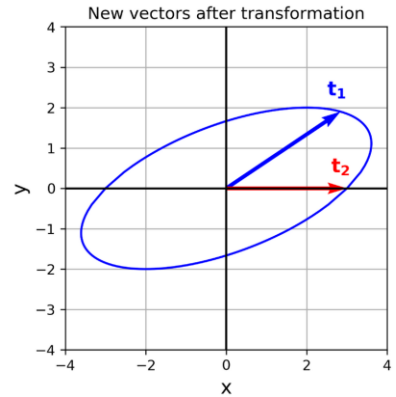
- matrix A as a transformation that acts on a vector x





- a circle that contains all the vectors that are one unit away from the origin $\mathbf{x} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad where \;\; x_i^2 + y_i^2 = 1$





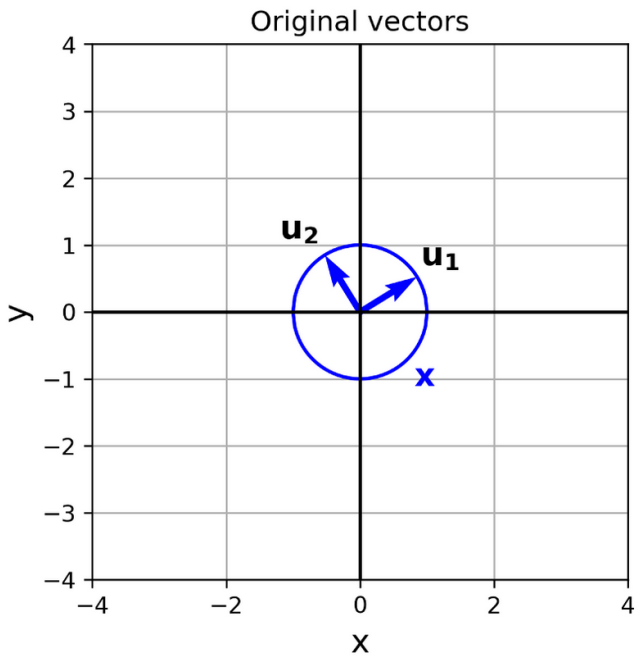

```
A =[ 3, 2 ]
    0, 2
```

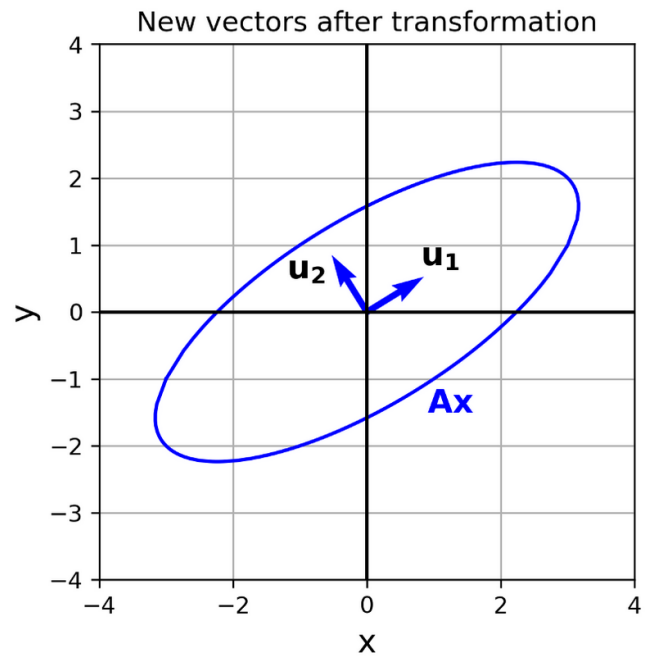
Eigen vector

- For matrix $\mathbf{B} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$

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

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



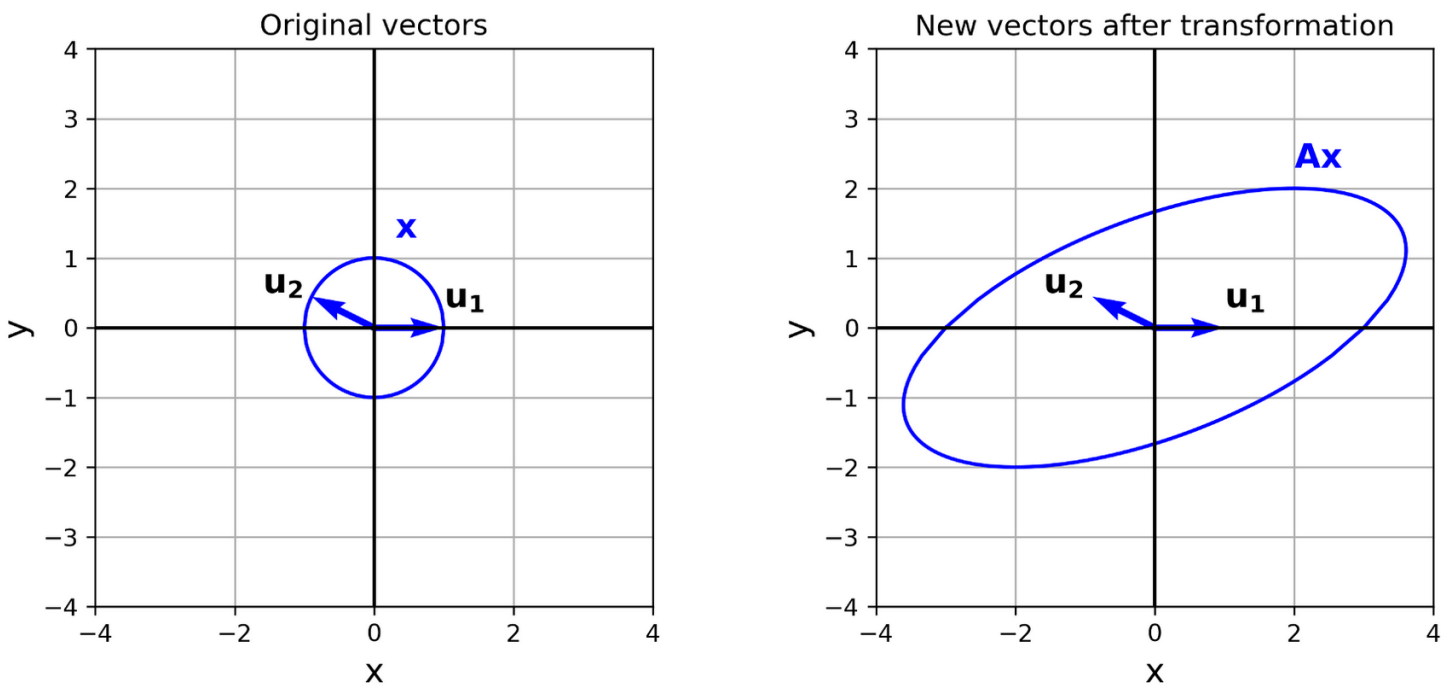


Eigen vector

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

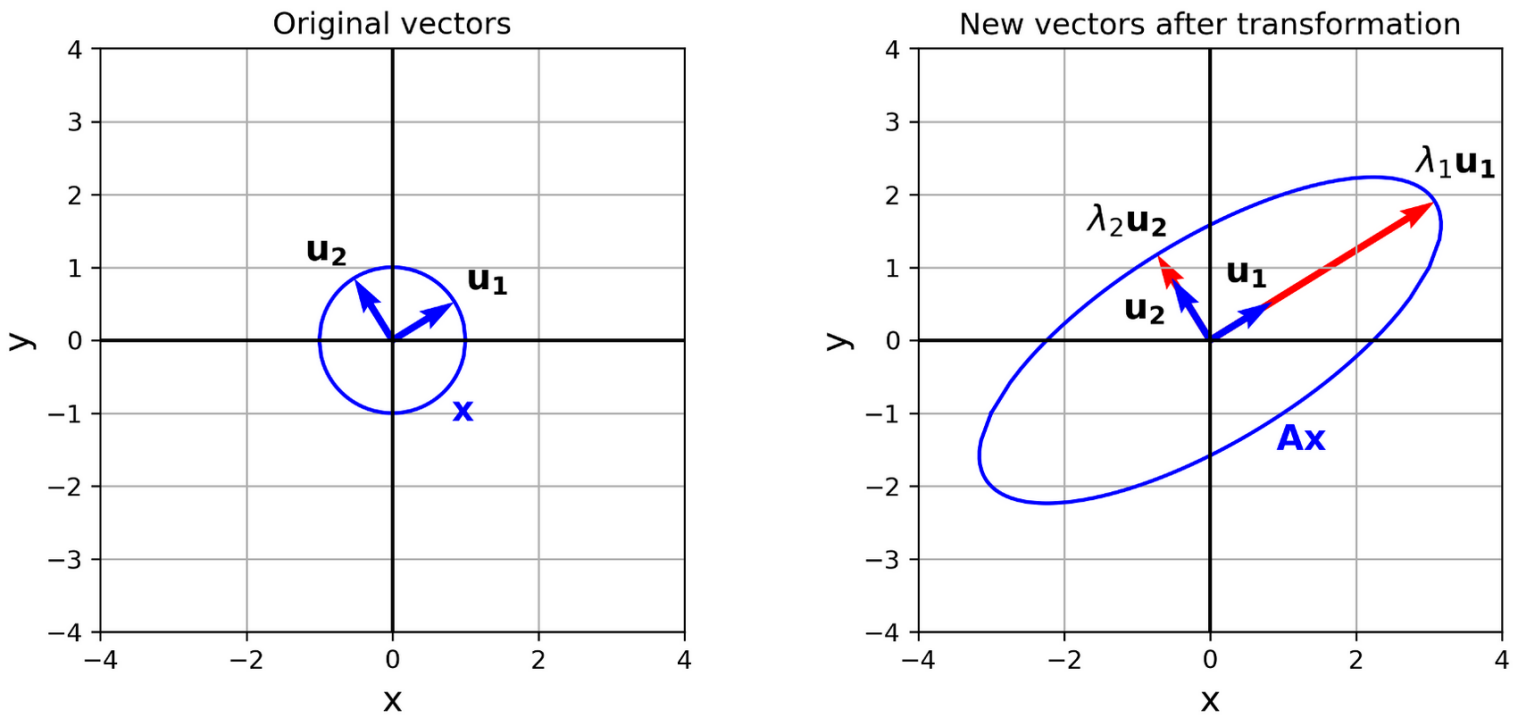
$$\mathbf{u_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{u_2} = \begin{bmatrix} -0.8944 \\ 0.4472 \end{bmatrix}$$

$$\lambda_1 = 3 \quad \lambda_2 = 2$$



Eigen Vector and Eigen Value

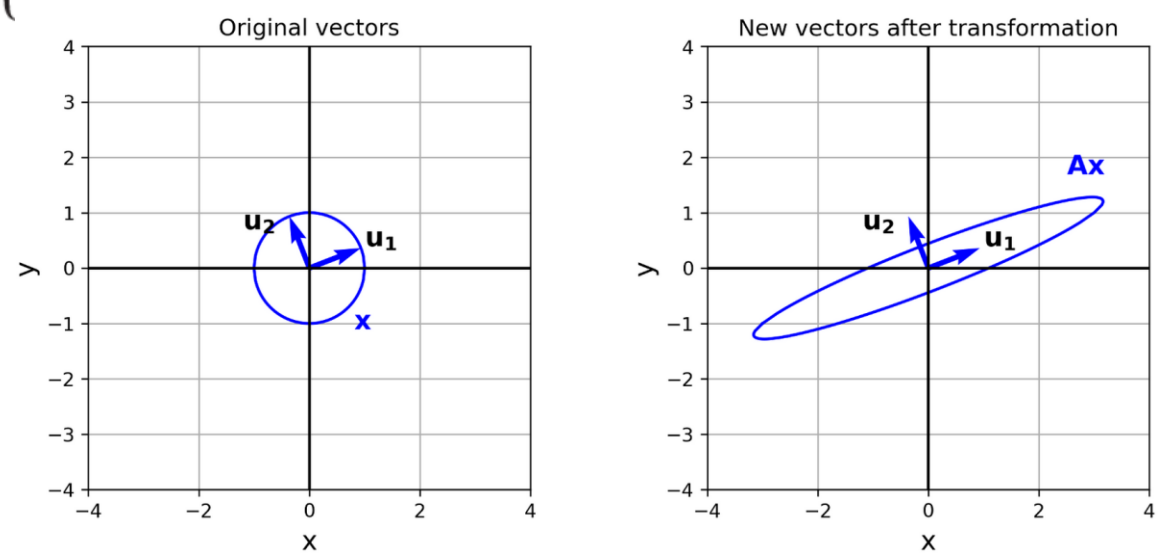
- A symmetric matrix transforms a vector by stretching or shrinking it along its eigenvectors.



Eigen Vector and Eigen Value

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

$$\mathbf{C} = \begin{bmatrix} 3 & 1 \\ 1 & 0.8 \end{bmatrix} \qquad \mathbf{u_1} = \begin{bmatrix} 0.9327 \\ 0.3606 \end{bmatrix} \qquad \mathbf{u_2} = \begin{bmatrix} -0.3606 \\ 0.9327 \end{bmatrix}$$

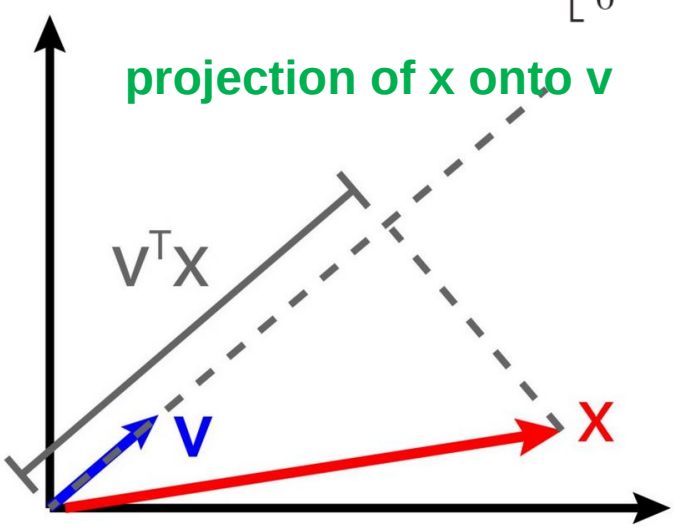


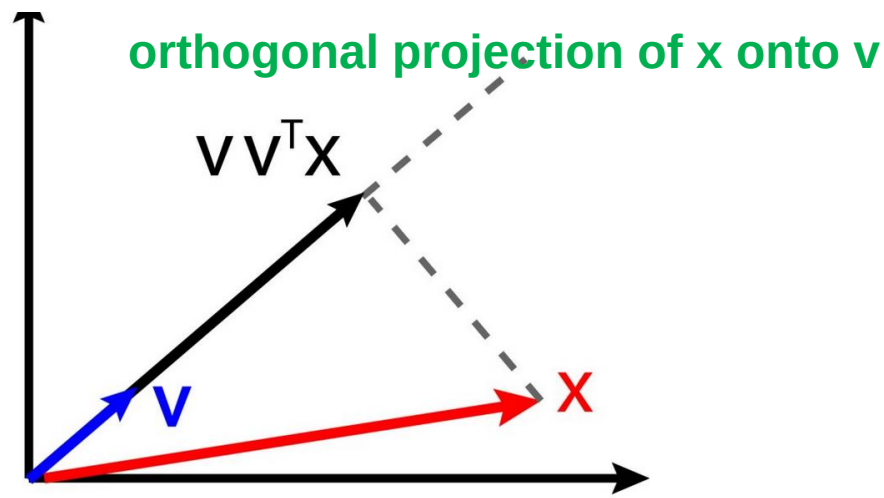
Geometrical interpretation of eigendecomposition

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

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

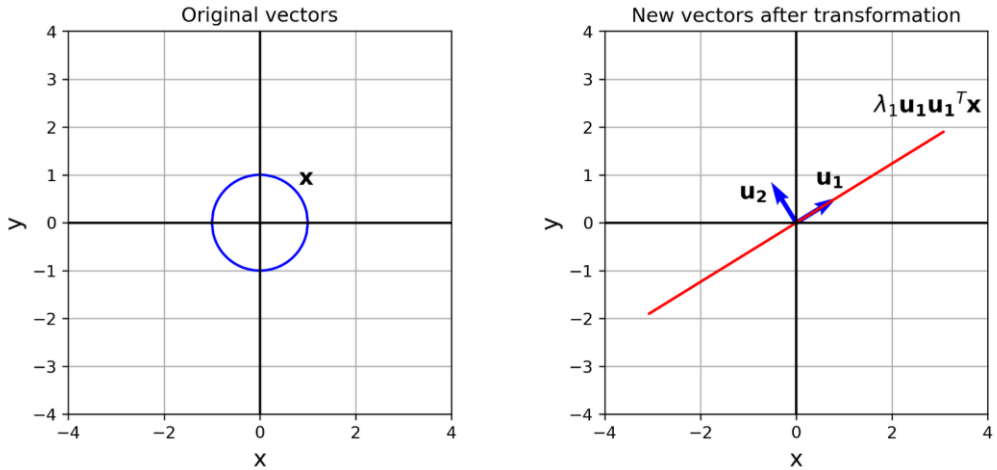
$$\mathbf{A} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \dots & \mathbf{u_n} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \dots & \mathbf{u_n} \end{bmatrix}^T$$





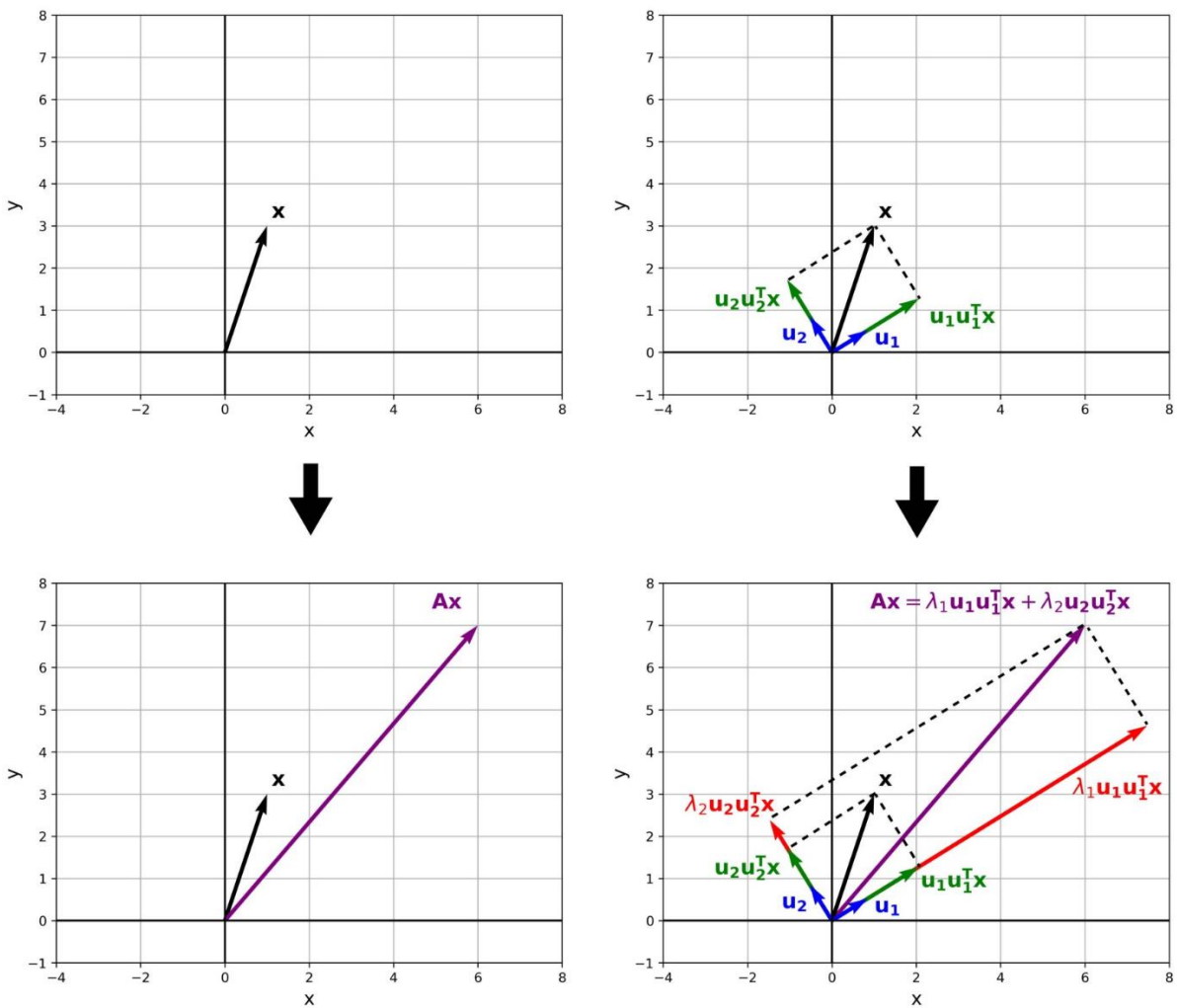
Geometrical interpretation of eigendecomposition

- All the projection matrices in the eigendecomposition equation are symmetric with rank 1.
- So we conclude that each matrix $\lambda_i \mathbf{u_i} \mathbf{u_i}^T$ in the eigendecomposition equation is a symmetric $n \times n$ matrix with n eigenvectors. The eigenvectors are the same as the original matrix **A** which are **u1, u2, … un**. The corresponding eigenvalue of **ui** is λi (which is the same as **A**), but all the other eigenvalues are zero.



Approximation

- symmetric matrix **A** $\mathbf{Ax} = \lambda_1 \mathbf{u_1}\mathbf{u_1}^T\mathbf{x} + \lambda_2 \mathbf{u_2}\mathbf{u_2}^T\mathbf{x} + \ldots + \lambda_n \mathbf{u_n}\mathbf{u_n}^T\mathbf{x}$



Approximation

- $$\mathbf{A} = \lambda_1 \mathbf{u_1}\mathbf{u_1}^T + \lambda_2 \mathbf{u_2}\mathbf{u_2}^T + \ldots + \lambda_n \mathbf{u_n}\mathbf{u_n}^T$$
  $$\lambda 1 \geq \lambda 2 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_n$$

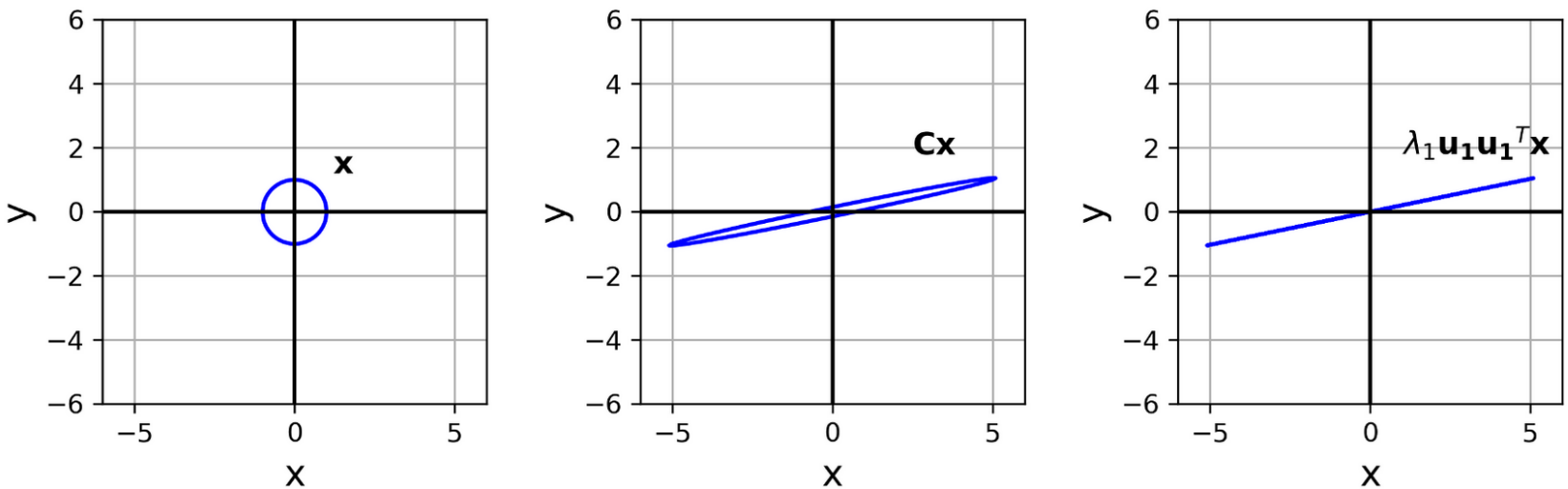
- $$\mathbf{A} \approx \mathbf{A}_k = \lambda_1 \mathbf{u_1}\mathbf{u_1}^T + \lambda_2 \mathbf{u_2}\mathbf{u_2}^T + \ldots + \lambda_k \mathbf{u_k}\mathbf{u_k}^T$$

- $$\mathbf{A} \approx \mathbf{A_k} = \mathbf{P_k}\mathbf{D_k}\mathbf{P_k^T} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \ldots & \mathbf{u_k} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_k \end{bmatrix} \begin{bmatrix} \mathbf{u_1}^T \\ \mathbf{u_2}^T \\ \vdots \\ \mathbf{u_k}^T \end{bmatrix}$$
  $$= \lambda_1 \mathbf{u_1}\mathbf{u_1}^T + \lambda_2 \mathbf{u_2}\mathbf{u_2}^T + \ldots + \lambda_k \mathbf{u_k}\mathbf{u_k}^T$$

Approximation

- $\mathbf{C} = \begin{bmatrix} 5 & 1 \\ 1 & 0.35 \end{bmatrix}$ $\quad \lambda_1 = 5.2059 \quad \lambda_2 = 0.1441$

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



Non-symmetric matrix

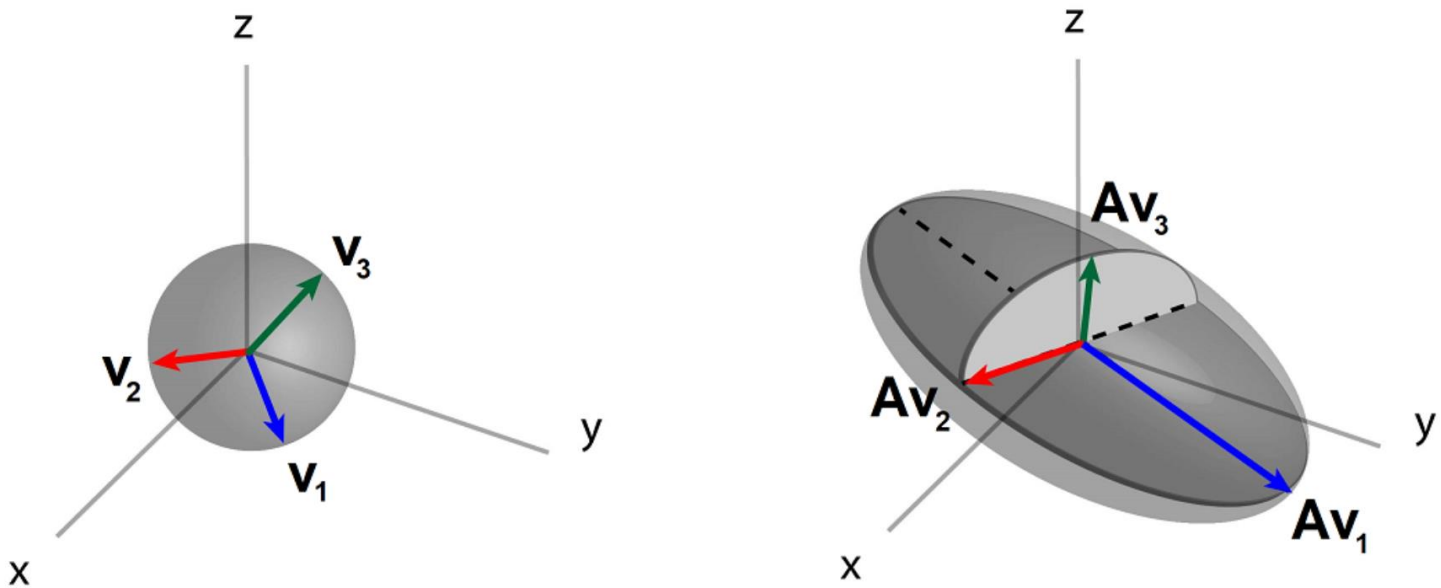
- no real eigenvalues $\begin{bmatrix} 3 & -1 \\ 1 & 2 \end{bmatrix}$

- The eigenvectors are not linearly independent $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$

- the eigenvectors are linearly independent, but they are not orthogonal $\begin{bmatrix} 3 & 2 \\ 0 & 2 \end{bmatrix}$

Singular Values

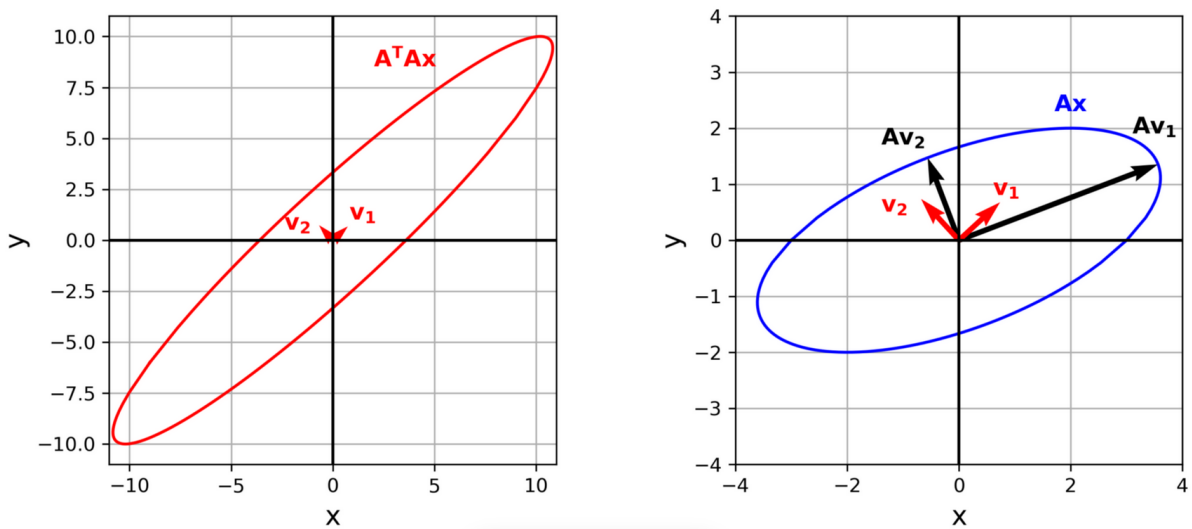
- For non-symmetric matrix, gram matrix is symmetric.



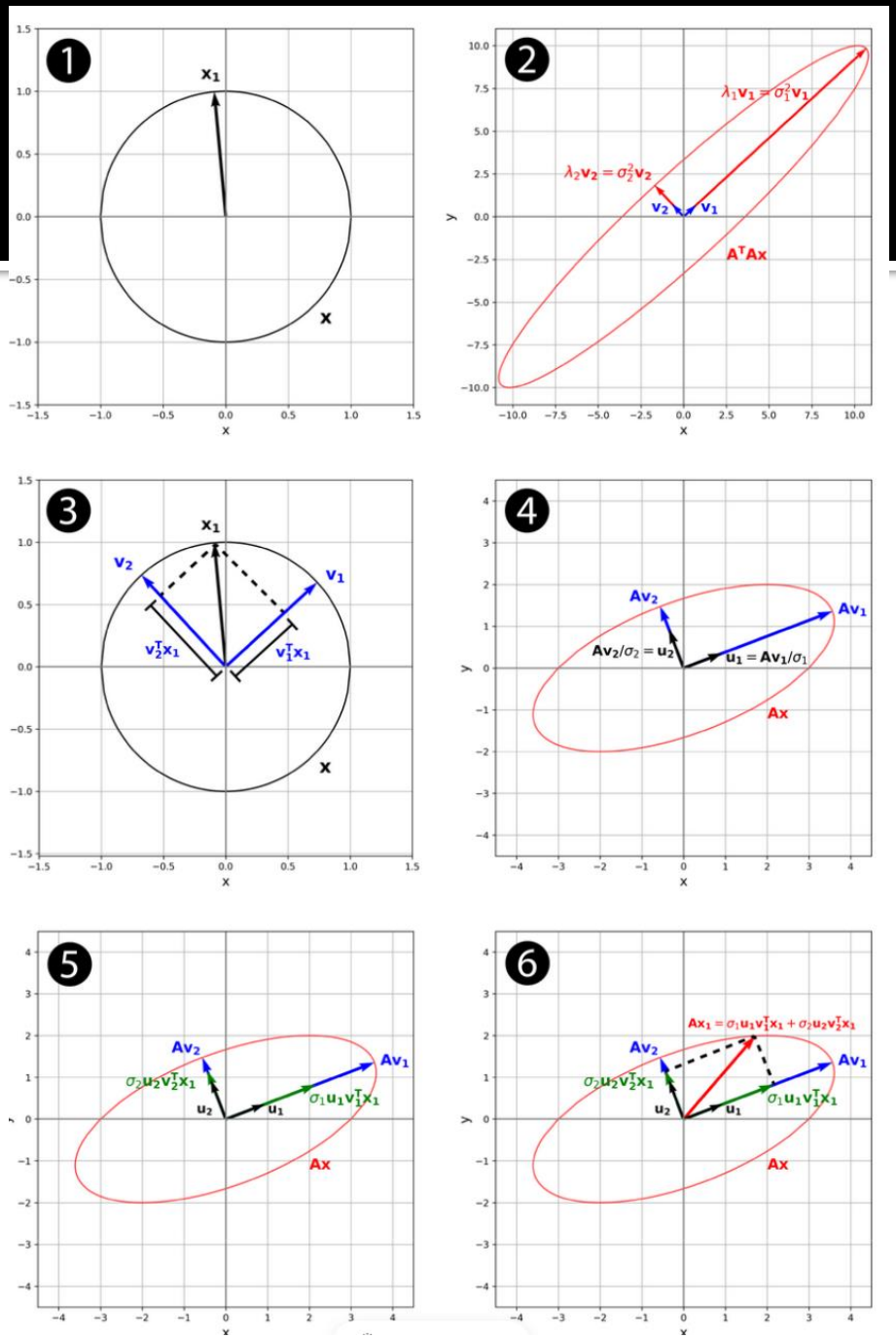
Singular Values

- Now let **A** be an $m \times n$ matrix. We showed that **A^T A** is a symmetric matrix, so it has n real eigenvalues and n linear 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 **v1**, **v2**, … **vn** and we assume they are normalized.

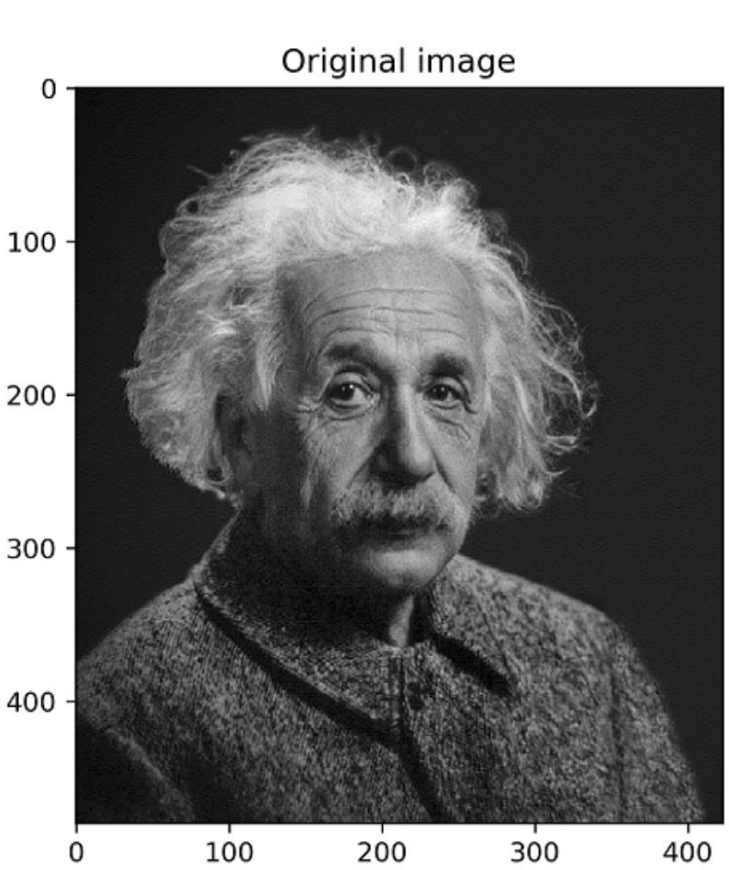
- Back to page 4

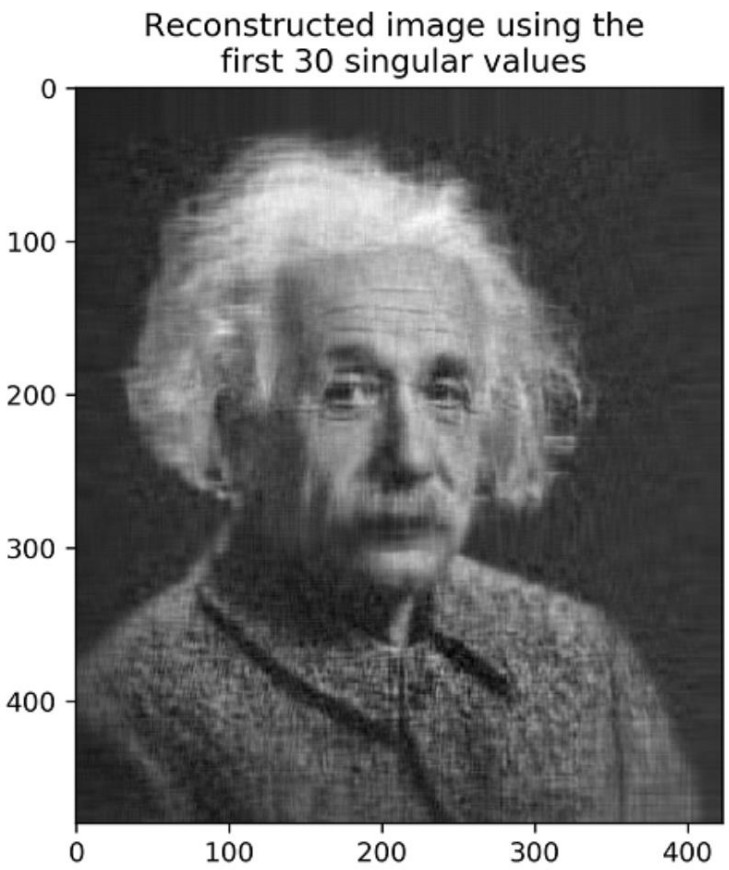


SVD

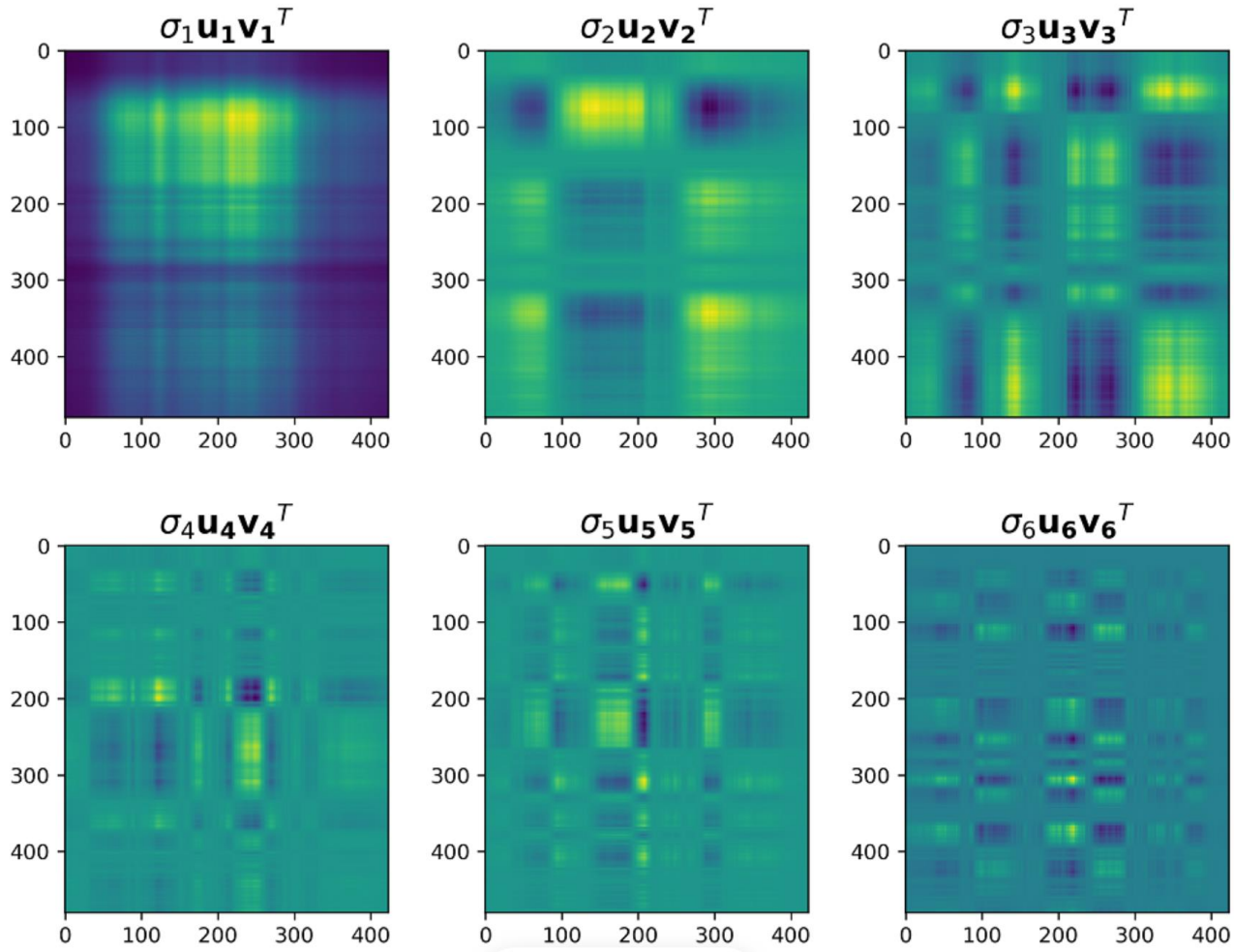

1
x_1
x
2
$\lambda_1 v_1 = \sigma_1^2 v_1$
$\lambda_2 v_2 = \sigma_2^2 v_2$
v_2
v_1
$A^T Ax$
3
x_1
v_2
v_1
$v_2^T x_1$
$v_1^T x_1$
x
4
Av_2
Av_1
$Av_2/\sigma_2 = u_2$
$u_1 = Av_1/\sigma_1$
Ax
5
Av_2
Av_1
$\sigma_2 u_2 v_2^T x_1$
$\sigma_1 u_1 v_1^T x_1$
u_2
u_1
Ax
6
$Ax_1 = \sigma_1 u_1 v_1^T x_1 + \sigma_2 u_2 v_2^T x_1$
Av_2
Av_1
$\sigma_2 u_2 v_2^T x_1$
$\sigma_1 u_1 v_1^T x_1$
u_2
u_1
Ax

Applications: Dimensionality reduction

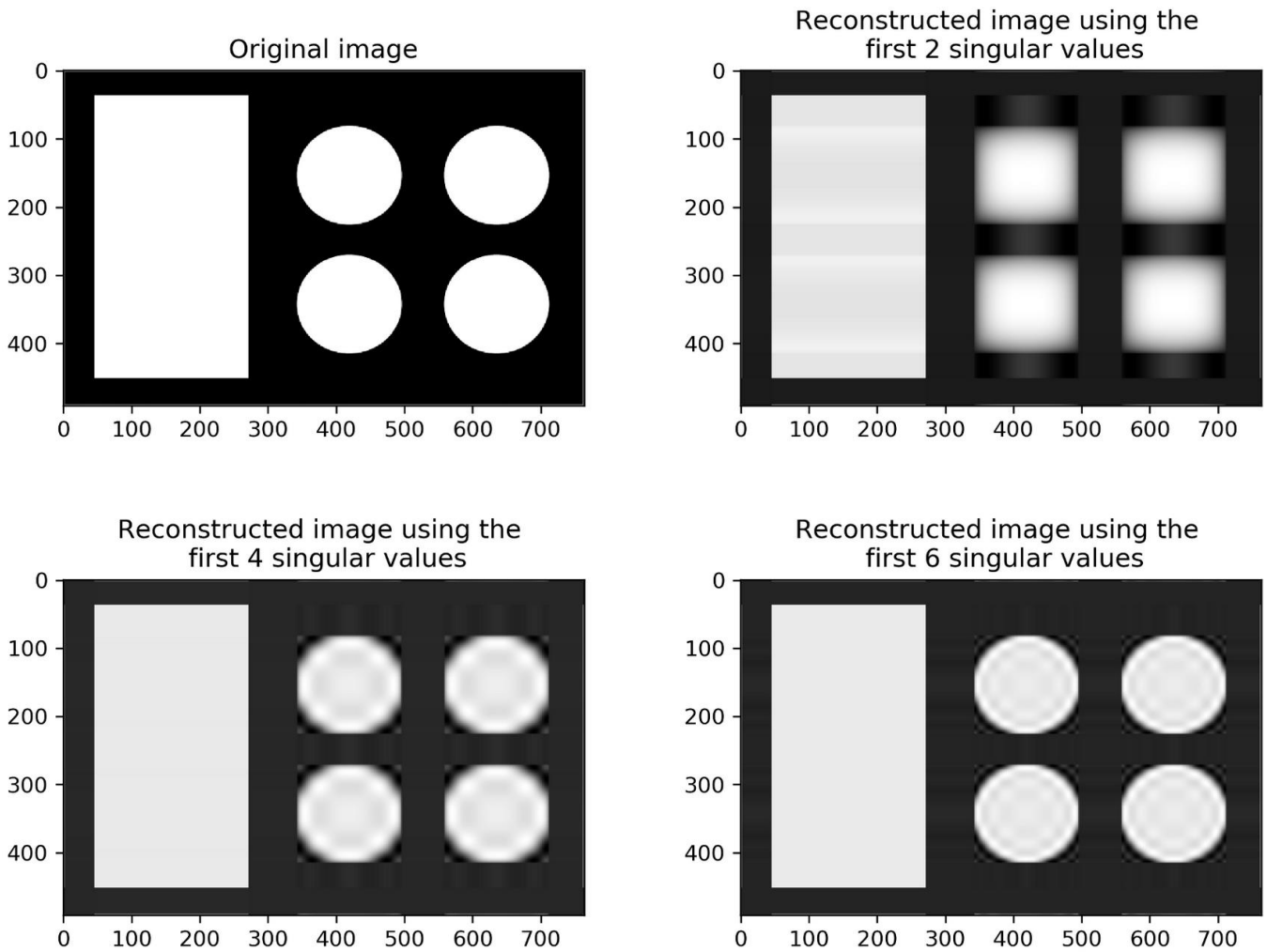




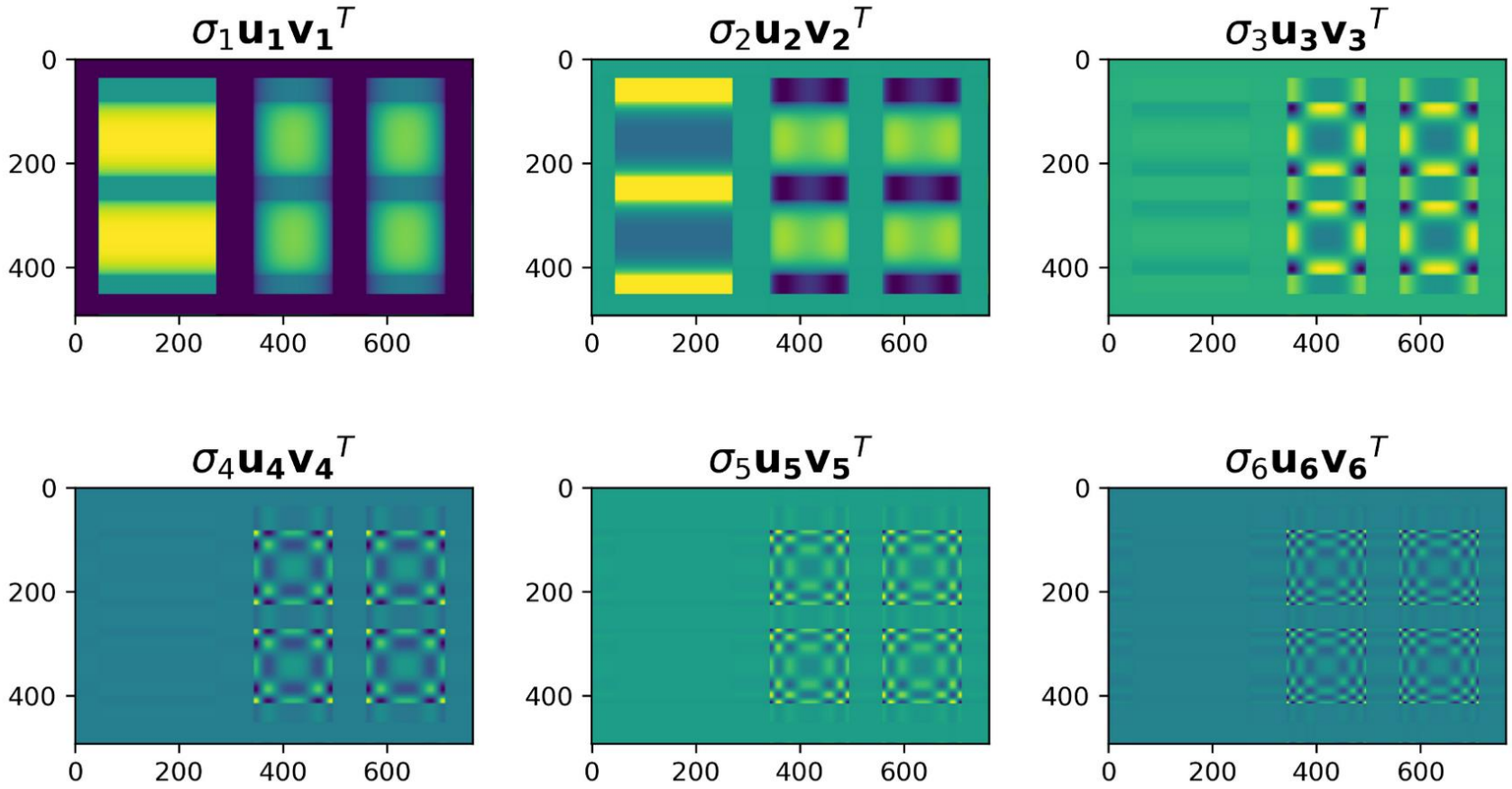
Applications: Dimensionality reduction



Applications: Dimensionality reduction



Applications: Dimensionality reduction



PCA

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

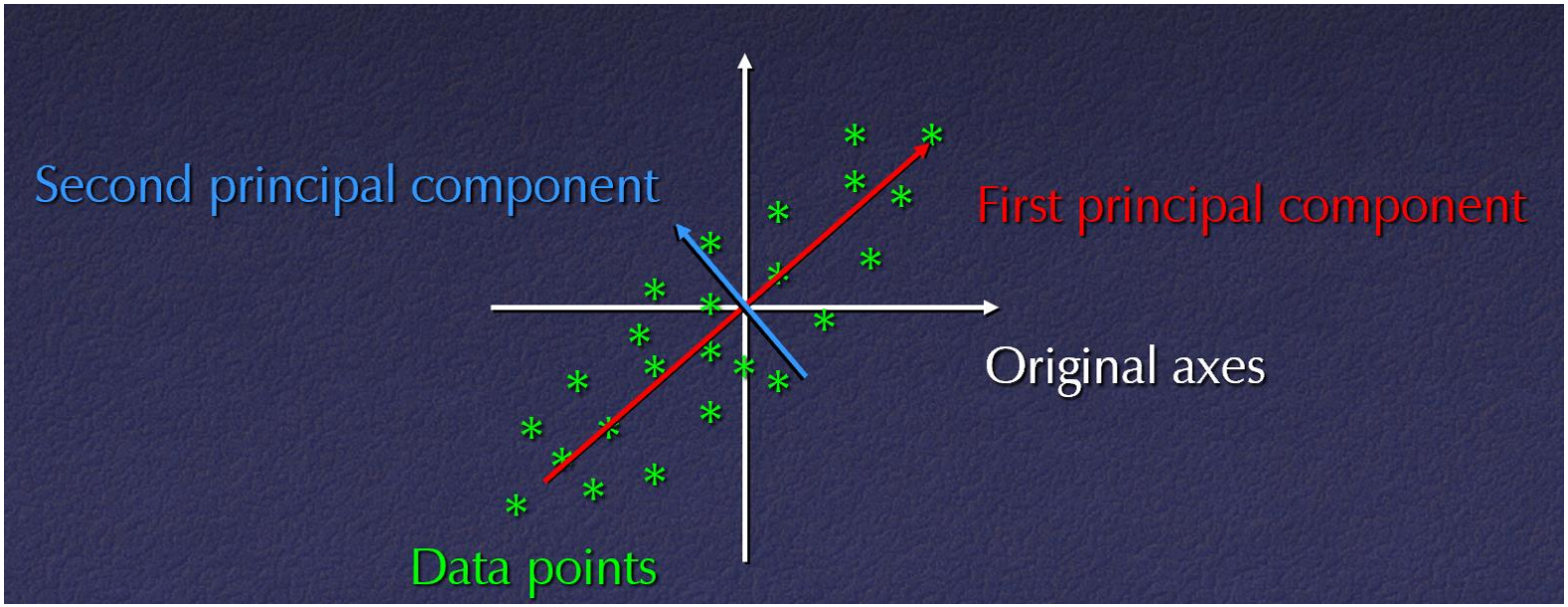
  The central ideas of PCA are **orthonormal coordinate** *systems, the distinction between* **variance and covariance**, *and the possibility of choosing an orthonormal basis to* **eliminate covariance**. *Technically, PCA may be performed either by* **eigenvector analysis** *of the covariance matrix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

PCA and SVD

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

SVD and PCA

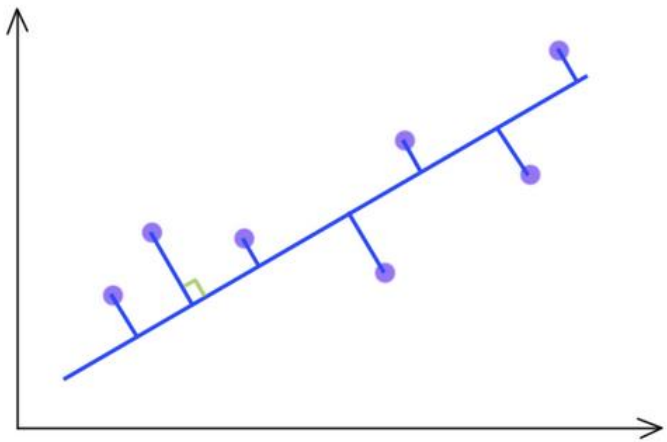
- Principal Components Analysis (PCA): approximating a high-dimensional data set with a lower-dimensional subspace

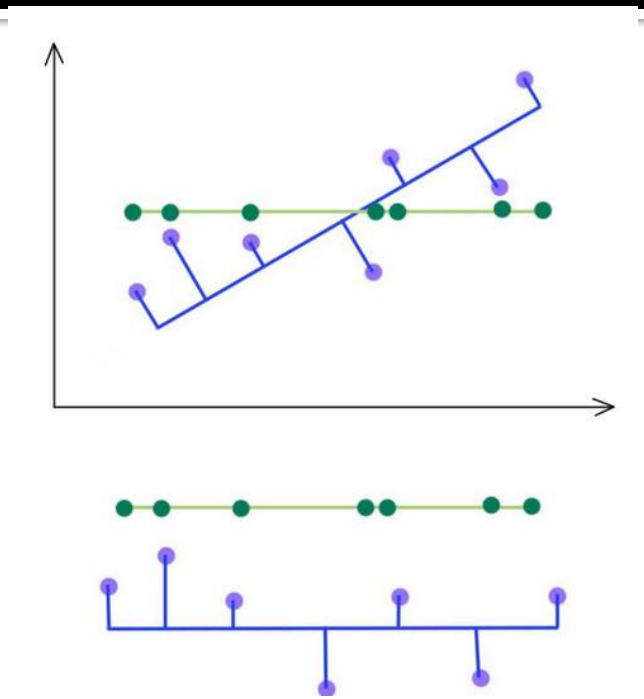


SVD and PCA

- Data matrix with points as rows, take SVD
  - Subtract out mean ("whitening")
- Columns of $\mathbf{V}_k$ are principal components
- Value of $\sum_i$ gives importance of each component

SVD and PCA





$$
\begin{array}{ccccc}
\mathbf{A} & & \mathbf{U} & \mathbf{S} & \mathbf{V}^{\mathrm{T}} \\
\begin{pmatrix} x_{11} & & x_{1n} \\ & \ddots & \\ x_{m1} & & x_{mn} \end{pmatrix} & \rightarrow & \begin{pmatrix} u_{11} & u_{kl} & u_{r1} \\ & \ddots & \ddots \\ u_{1m} & u_{km} & u_{rm} \end{pmatrix} & \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_k & \\ & & & \ddots \\ 0 & & & \sigma_r \end{pmatrix} & \begin{pmatrix} v_{11} & v_{1k} & v_{1n} \\ & \ddots & \ddots \\ v_{r1} & v_{rk} & v_{rn} \end{pmatrix} \\
m \times n & & & &
\end{array}
$$

$$
\rightarrow \begin{pmatrix} u_{11} & u_{kl} \\ & \ddots \\ u_{1m} & u_{km} \end{pmatrix} \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{pmatrix} \begin{pmatrix} v_{11} & v_{1n} \\ & \ddots \\ v_{kl} & v_{kn} \end{pmatrix}
$$

PCA Proof

- 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)

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

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

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

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

Low-Rank Approximation

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

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

When using the induced 2-norm, the best **rank-$\boldsymbol{k}$** approximation is given by:

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

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots \geq 0$$

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

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

Low-Rank Approximation

- Example $\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$, the best rank-1 approximation is

- Solution

$$\mathbf{X}_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = \sqrt{3} \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

In this problem, the approximation error under either norm (spectral or Frobenius) is the same: $\|\mathbf{X} - \mathbf{X}_1\| = \sigma_2 = 1$.

Matrix norms

The Euclidean norm of an orthogonal matrix is equal to 1

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

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

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

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

Where we used the fact that $\|\boldsymbol{U}\|_2 = 1$, $\|\boldsymbol{V}\|_2 = 1$ and $\boldsymbol{\Sigma}$ is diagonal

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

σ_{max} is the largest singular value

Norm for the inverse of a matrix

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

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

$$\|\boldsymbol{A}^{-1}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \left\|(\boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\boldsymbol{T}})^{-1}\boldsymbol{x}\right\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \left\|\boldsymbol{V}\,\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{x}\right\|_2$$

Since $\|\boldsymbol{U}\|_2 = 1$, $\|\boldsymbol{V}\|_2 = 1$ and $\boldsymbol{\Sigma}$ is diagonal then

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

σ_{min} is the smallest singular value

Norm of the pseudo-inverse matrix

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

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

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, $\|\boldsymbol{A}^{+}\|_2$ is the same as $\|\boldsymbol{A}^{-1}\|_2$.

Zero matrix: If $\boldsymbol{A}$ is a zero matrix, then $\boldsymbol{A}^{+}$ is also the zero matrix, and $\|\boldsymbol{A}^{+}\|_2 = 0$

SVD Summary

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

- In reduced form: $\boldsymbol{A} = \boldsymbol{U_R}\boldsymbol{\Sigma_R}\boldsymbol{V_R}^{\boldsymbol{T}}$, 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 $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$, denoted the right singular vectors.

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

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

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

Matrix Norm

- Example
- Solution

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

$$\|\mathbf{C}\|_1 = \max_{\|\mathbf{x}\|_1=1} \|\mathbf{C}\mathbf{x}\|_1$$

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

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

$$\|\mathbf{C}\|_1 = \max(4, 5)$$

$$\|\mathbf{C}\|_1 = 5$$

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

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

$$\|\mathbf{C}\|_\infty = 5$$

$$\|\mathbf{C}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{C}\mathbf{x}\|_2$$

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

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

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

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

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

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

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

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

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

Matrix Norm

- Example

Show that $\|A\|_{\mathrm{F}} = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$, where $\sigma_1, \ldots, \sigma_r$ are the singular values of A. Then show that $\sigma_{\max}(A) \le \|A\|_{\mathrm{F}} \le \sqrt{r}\,\sigma_{\max}(A)$.

- Solution

$\|U\Sigma V^T\|_F$

We start with the full SVD of $A = U\Sigma V^{\mathsf{T}}$. By the previous problem,

$$\|A\|_{\mathrm{F}} = \|\Sigma V^{\mathsf{T}}\|_{\mathrm{F}} = \|\Sigma\|_{\mathrm{F}} = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}.$$

Since $\sigma_2^2, \ldots, \sigma_r^2 \ge 0$, we have $\sigma_1 \le \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} = \|A\|_{\mathrm{F}}$. Next, since $\sigma_2, \ldots, \sigma_r \le \sigma_1$, we have $\|A\|_{\mathrm{F}} = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \le \sqrt{\sigma_1^2 + \cdots + \sigma_1^2} = \sqrt{r}\sigma_1$.

$(\Sigma V^T)^T(\Sigma V^T)$

چرا؟ $\|U\Sigma V^T\|_F^2 = tr(U\Sigma V^T)^T(U\Sigma V^T) = tr(V\Sigma^T \underbrace{U^T U}_{I} \Sigma V^T) = tr(V\Sigma^T\Sigma V^T) = \|\Sigma V^T\|_F^2$

چرا؟ $\|\Sigma V^T\|_F^2 = tr((\Sigma V^T)^T(\Sigma V^T)) = tr((\Sigma V^T)(V\Sigma^T))$

$= tr(\Sigma \underbrace{V^T V}_{I} \Sigma^T)$

از این خاصیت استفاده شد: $tr(AB) = tr(BA)$

$= tr(\Sigma\Sigma^T) = tr(\Sigma^T\Sigma)$

$= \|\Sigma\|_F$

Matrix norm and low-rank approximation

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

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

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

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^2 = \mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^T$$

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

Application

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