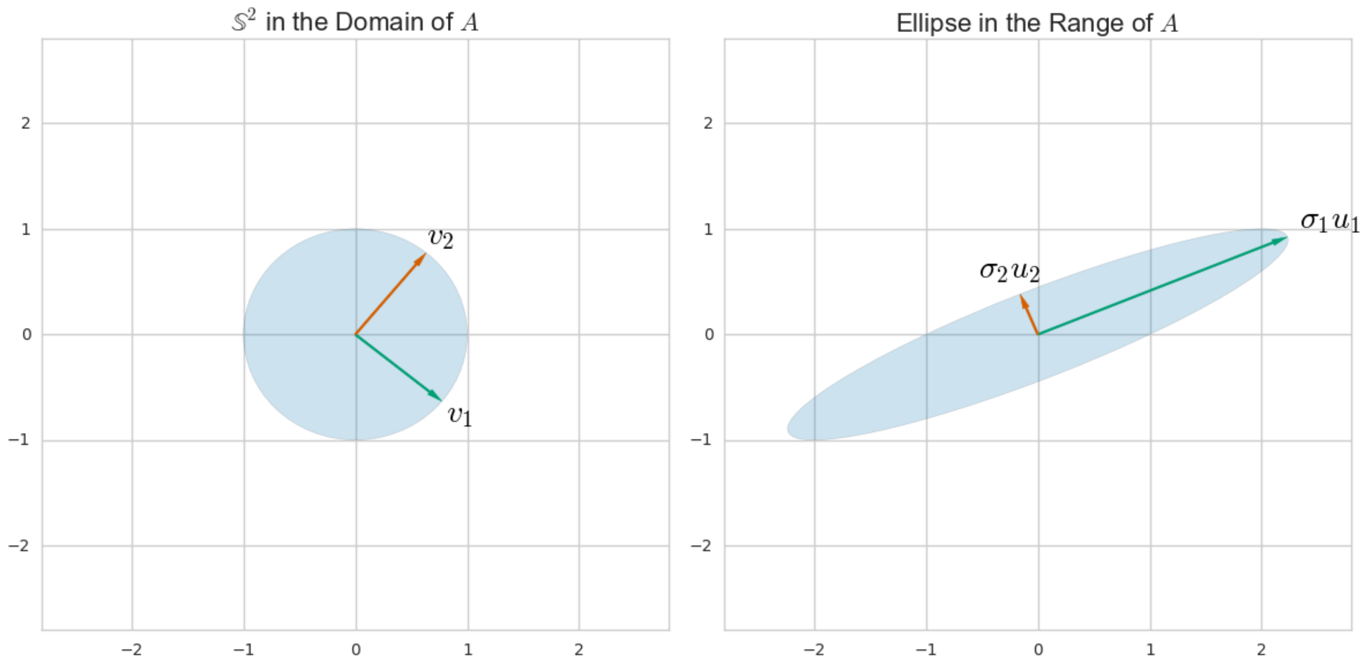
Singular Value Decomposition (SVD)

Prof. Seungchul Lee
Industrial AI Lab.

Geometry of Linear Maps

- Matrix A (or linear transformation) = rotate + stretch/compress



Geometry of Linear Maps

- An extremely important fact:

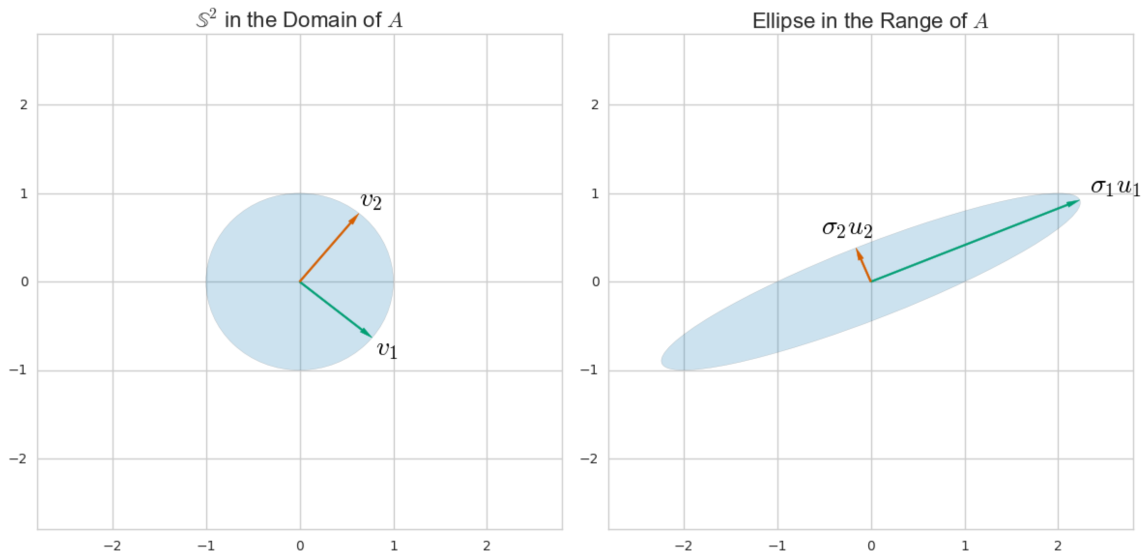
every matrix $A \in \mathbb{R}^{m\times n}$ maps the unit ball in $\mathbb{R}^n$ to an ellipsoid in $\mathbb{R}^m$

$$S = \{x \in \mathbb{R}^n \mid \; \|x\| \leq 1\}$$
$$AS = \{Ax \mid x \in \mathbf{S}\}$$

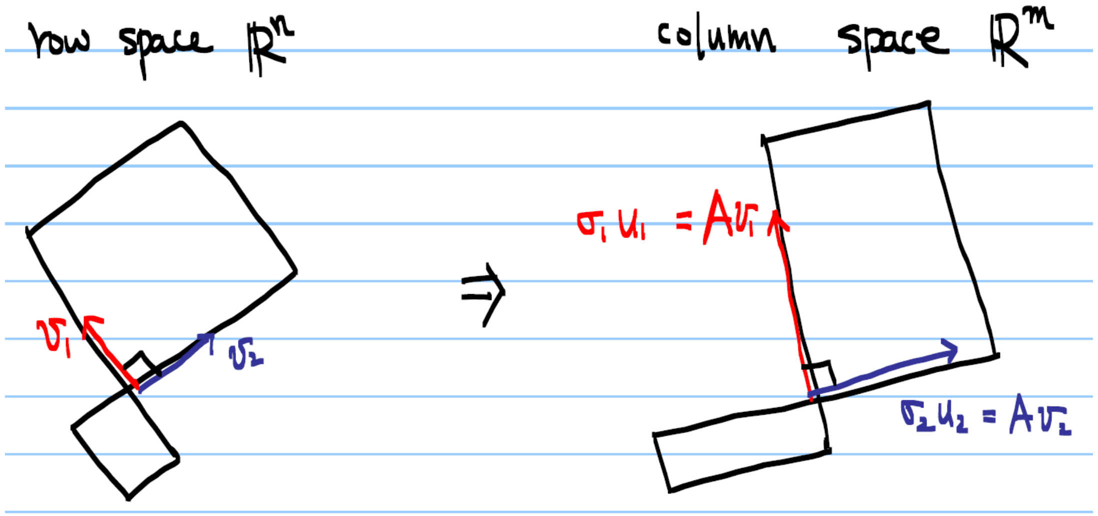
Singular Values and Singular Vectors

- the numbers $\sigma_1, \cdots, \sigma_n$ are called the singular values of A by convention, $\sigma_i > 0$
- the vectors $u_1, \cdots, u_n$ these are unit vectors along the principal semi-axes of AS
- the vectors $v_1, \cdots, v_n$ these are the preimages of the principal semi-axes, defined so that

$$Av_i = \sigma_i u_i$$



Graphical Explanation



$$
\begin{aligned}
A\begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix} &= \begin{bmatrix} Av_1 & Av_2 & \cdots & Av_r \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_r u_r \end{bmatrix} \\
&= \begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}
\end{aligned}
$$

$$\therefore \; AV = U\Sigma \quad (r \leq m, n)$$

Thin Singular Value Decomposition

$A \in \mathbb{R}^{m \times n}$, skinny and full rank (*i.e.*, $r = n$)

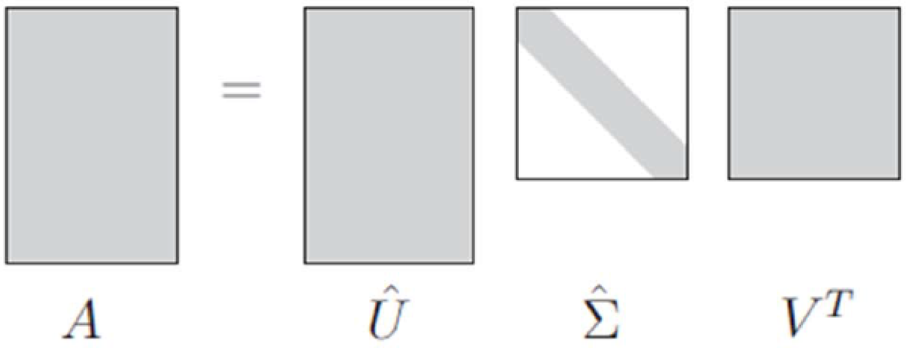
$$Av_i = \sigma_i u_i \quad \text{for } 1 \leq i \leq n$$

$$\hat{U} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

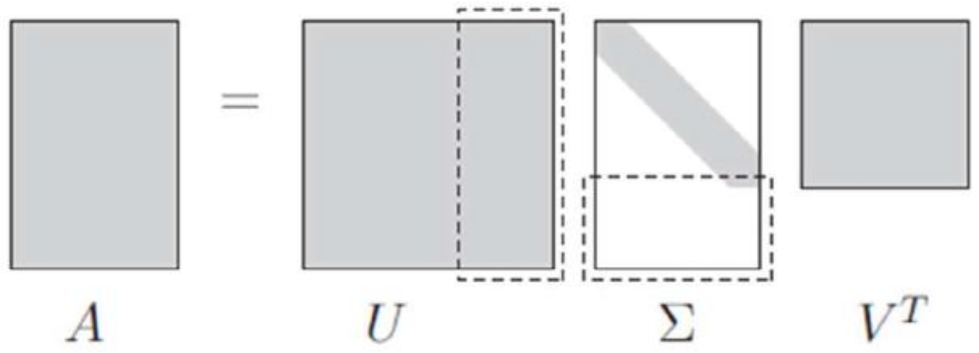
$$A = \hat{U}\hat{\Sigma}V^T$$



Full Singular Value Decomposition

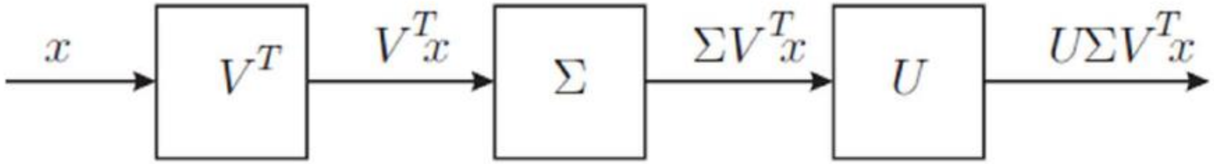
- We can add extra orthonormal columns to U
- We also add extra rows of zeros to Σ

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



Interpretation of SVD

- The SVD decomposes the linear map into
  - rotate by V^T
  - diagonal scaling by σ_i
  - rotate by U



- Note that, unlike the eigen-decomposition, input and output directions are different

SVD: Matrix factorization

- for any matrix A

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

- for symmetric and positive definite matrix A

$$A = S\Lambda S^{-1} = S\Lambda S^T \quad (S : \text{eigenvectors})$$

PCA and SVD

- Any real symmetric and positive definite matrix B has a eigen decomposition

$$B = S\Lambda S^T$$

- A real matrix $(m \times n)$ A, where $m > n$, has the decomposition

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

- From A (skinny and full rank) we can construct two positive-definite symmetric matrices, AA^T and $A^T A$

$$\begin{aligned} AA^T &= U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T \quad (m \times m,\ n \text{ eigenvalues and } m-n \text{ zeros},\ U \text{ eigenvectors}) \\ A^T A &= V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T \quad (n \times n,\ n \text{ eigenvalues},\ V \text{ eigenvectors}) \end{aligned}$$

PCA and SVD

- PCA by SVD

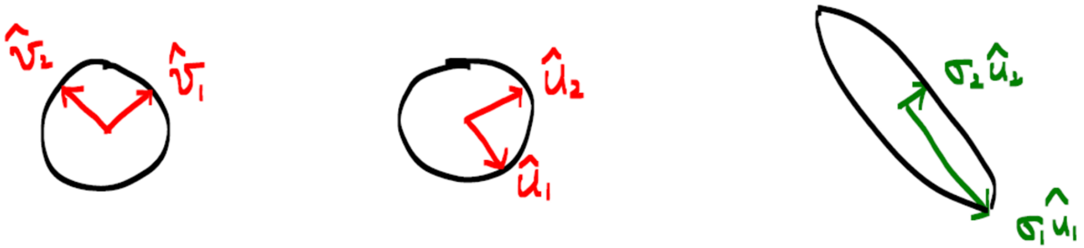
$$B = A^T A = V\Sigma^T \Sigma V^T = V\Lambda V^T$$

  - V is eigenvectors of $B = A^T A$
  - $\sigma_i^2 = \lambda_i$

- Note that in PCA,

$$VSV^T \quad \left(\text{where } S = \frac{1}{m} X^T X = \left(\frac{X}{\sqrt{m}}\right)^T \frac{X}{\sqrt{m}} = A^T A\right)$$

Low Rank Approximation: Dimension Reduction



$$
\begin{aligned}
x &= c_1 v_1 + c_2 v_2 \\
&= \langle x, v_1 \rangle v_1 + \langle x, v_2 \rangle v_2 \\
&= (v_1^T x) v_1 + (v_2^T x) v_2
\end{aligned}
$$

$$
\begin{aligned}
Ax &= c_1 A v_1 + c_2 A v_2 \\
&= c_1 \sigma_1 u_1 + c_2 \sigma_2 u_2 \\
&= u_1 \sigma_1 v_1^T x + u_2 \sigma_2 v_2^T x \\
&= \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} x \\
&= U \Sigma V^T x \\
&\approx u_1 \sigma_1 v_1^T x \quad (\text{if } \sigma_1 \gg \sigma_2)
\end{aligned}
$$

$$A = U_r \Sigma_r V_r^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

$$\tilde{A} = U_k \Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T \qquad (k \leq r)$$

Example: Image Approximation

- Approximation of A

$$\tilde{A} = U_k \Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T \qquad (k \le r)$$

Original image

Approximated image w/ rank = 20

Proper Orthogonal Modes (POM)

- Principal Components and Proper Orthogonal Modes (POM)
  - the principal component analysis seems to suggest that we are simply expanding our solution in another orthonormal basis, one which can always diagonalize the underlying system.

$$f(x,t) \approx \sum_{i=1}^{k} c_i(t)\phi_i(x)$$

- Here are some of the more common expansion bases used in practice

$$\begin{aligned}
\phi_i(x) &= (x - x_0)^i && \text{Taylor expansion} \\
\phi_i(x) &= e^{ix} && \text{Fourier transform} \\
\phi_i(x) &= \psi_{a,b}(x) && \text{Wavelet transform} \\
\phi_i(x) &= \phi_{\lambda_i}(x) && \text{Eigenfunction expansion}
\end{aligned}$$

Eigenface

Seeing through a Disguise with SVD

Dimension Reduction = Model Reduction

- Practical use of first principles models for physical phenomena described by partial differential equations, advanced control and signal processing algorithms
- Facilitate system identification, data compression, and knowledge extraction