San José State University

Math 250: Mathematical Data Visualization

Singular Value Decomposition (SVD)

Dr. Guangliang Chen

Outline of the lecture:

- Existence of SVD for general matrices
- Different versions of SVD
- Computing SVD by hand and software
- Geometric interpretation
- Applications of SVD

Recall

... that symmetric matrices are (orthogonally) diagonalizable.

That is, for any symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, there exist an orthogonal matrix $\mathbf{Q} = [\mathbf{q}_1 \ldots \mathbf{q}_n]$ and a diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, both real and square, such that $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \sum_{i=1}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^T$.

Furthermore, λ_i's are the eigenvalues of $\mathbf{A}$ and $\mathbf{q}_i$'s the corresponding eigenvectors (which are orthogonal to each other and have unit norm).

Such a factorization is called the **eigendecomposition** of $\mathbf{A}$, also called the **spectral decomposition** of $\mathbf{A}$.

Existence of the SVD for general matrices

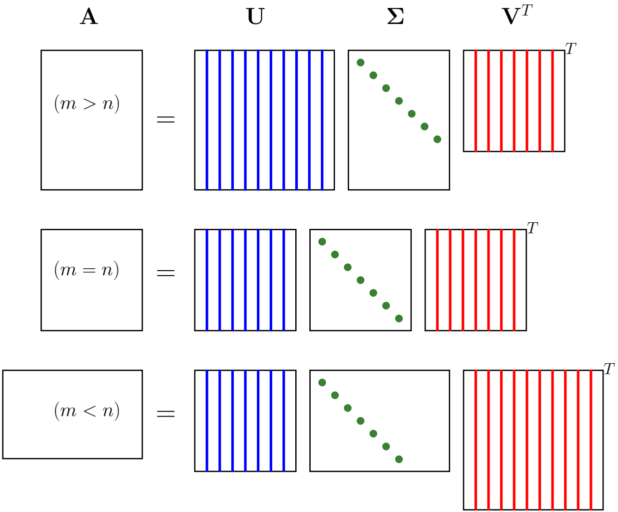
Theorem: For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, there exist two orthogonal matrices $\mathbf{U} \in \mathbb{R}^{m\times m}, \mathbf{V} \in \mathbb{R}^{n\times n}$ and a nonnegative, diagonal matrix $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ such that

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T.$$

Moreover, the number of positive diagonals of $\mathbf{\Sigma}$ equals the rank of $\mathbf{A}$.

Remark. This factorization is called the *Singular Value Decomposition (SVD)* of $\mathbf{A}$:

- The diagonals of $\mathbf{\Sigma}$ are called the singular values of $\mathbf{A}$.
- The columns of $\mathbf{U}$ are called the left singular vectors of $\mathbf{A}$.
- The columns of $\mathbf{V}$ are called the right singular vectors of $\mathbf{A}$.


$\mathbf{A}$
$\mathbf{U}$
$\mathbf{\Sigma}$
$\mathbf{V}^T$
$(m > n)$
$=$
T
$(m = n)$
$=$
T
$(m < n)$
$=$
T

Example 0.1. It can be directly verified that

$$\underbrace{\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{pmatrix} \sqrt{3} & \\ & 1 \\ & \end{pmatrix}}_{\mathbf{\Sigma}} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T}_{\mathbf{V}^T}.$$

In the above equation, $\mathbf{U}, \mathbf{V}$ are orthogonal matrices and $\mathbf{\Sigma}$ is a diagonal matrix. Therefore, the above factorization represents a singular value decomposition of $\mathbf{A}$.

Moreover, $\operatorname{rank}(\mathbf{A}) = 2$, and there are precisely 2 positive entries in the diagonal of $\mathbf{\Sigma}$.

- **Singular values**:

$$\sigma_1 = \sqrt{3}, \quad \sigma_2 = 1;$$

- **Left singular vectors**:

$$\mathbf{u}_1 = \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}$$

- **Right singular vectors**:

$$\mathbf{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Connection to symmetric matrices

From the SVD of $\mathbf{A}$ we obtain that

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \cdot \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T = \mathbf{U}\left(\mathbf{\Sigma}\mathbf{\Sigma}^T\right)\mathbf{U}^T$$

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T \cdot \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{V}\left(\mathbf{\Sigma}^T\mathbf{\Sigma}\right)\mathbf{V}^T$$

This shows that

- $\mathbf{U}$ is the eigenvectors matrix of $\mathbf{A}\mathbf{A}^T$;
- $\mathbf{V}$ is the eigenvectors matrix of $\mathbf{A}^T\mathbf{A}$;
- The nonzero eigenvalues of $\mathbf{A}\mathbf{A}^T, \mathbf{A}^T\mathbf{A}$ (which must be the same) are equal to the squared singular values of $\mathbf{A}$.

Example 0.2. For the matrix $\mathbf{A}$ in the preceding example, we have

$$\mathbf{A}\mathbf{A}^T = \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{pmatrix} 3 & & \\ & 1 & \\ & & 0 \end{pmatrix}}_{\mathbf{\Sigma}\mathbf{\Sigma}^T} \cdot \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}^T}_{\mathbf{U}^T}$$

$$\mathbf{A}^T\mathbf{A} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\mathbf{V}} \cdot \underbrace{\begin{pmatrix} 3 & \\ & 1 \end{pmatrix}}_{\mathbf{\Sigma}^T\mathbf{\Sigma}} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T}_{\mathbf{V}^T}$$

How to prove the SVD theorem

Given any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the SVD can be thought of as solving a matrix equation for three unknown matrices (under constraints):

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

Suppose such solutions exist. From

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\left(\mathbf{\Sigma}^T\mathbf{\Sigma}\right)\mathbf{V}^T$$

we can find $\mathbf{V}$ and $\mathbf{\Sigma}$, which contain the eigenvectors and square roots of eigenvalues of $\mathbf{A}^T\mathbf{A}$, respectively.

After we have found both $\mathbf{V}$ and $\mathbf{\Sigma}$, rewrite the matrix equation as

$$\mathbf{A}_{m\times n}\mathbf{V}_{n\times n} = \mathbf{U}_{m\times m}\mathbf{\Sigma}_{m\times n},$$

or in columns,

$$\mathbf{A}[\mathbf{v}_1 \ldots \mathbf{v}_r\, \mathbf{v}_{r+1} \ldots \mathbf{v}_n] = [\mathbf{u}_1 \ldots \mathbf{u}_r\, \mathbf{u}_{r+1} \ldots \mathbf{u}_m] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \\ & & & \end{bmatrix}.$$

By comparing columns, we obtain

$$\mathbf{A}\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i, & 1 \le i \le r \text{ (\#nonzero singular values)} \\ \mathbf{0}, & r < i \le n \end{cases}$$

This tells us how to find the first r columns of matrix $\mathbf{U} \in \mathbb{R}^{m\times m}$:

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i \quad \text{for all } 1 \le i \le r.$$

The remaining columns of $\mathbf{U}$ will be found by completing an orthonormal basis for $\mathbb{R}^m$, starting with $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$:

$$\begin{aligned} \mathbf{u}_i^T\mathbf{x} &= 0, \quad i = 1, \dots, r \\ \|\mathbf{x}\| &= 1 \end{aligned}$$

For a rigorous proof of the SVD theorem, see notes.

Example 0.3. Find the SVD of $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$.

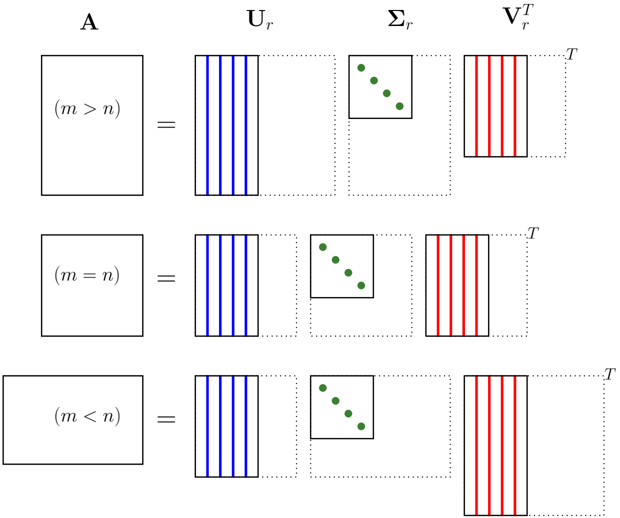
Different versions of SVD

- **Full SVD**: $\mathbf{A}_{m\times n} = \mathbf{U}_{m\times m}\mathbf{\Sigma}_{m\times n}\mathbf{V}_{n\times n}^T$
- **Compact SVD**: Suppose $\operatorname{rank}(\mathbf{A}) = r$. Define

$$\mathbf{U}_r = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{R}^{m\times r}$$
$$\mathbf{V}_r = [\mathbf{v}_1, \ldots, \mathbf{v}_r] \in \mathbb{R}^{n\times r}$$
$$\mathbf{\Sigma}_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$$

Then

$$\mathbf{A} = \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^T.$$


$\mathbf{A}$
$\mathbf{U}_r$
$\mathbf{\Sigma}_r$
$\mathbf{V}_r^T$
$(m > n)$
$=$
T
$(m = n)$
$=$
T
$(m < n)$
$=$
T

- **Rank-1 decomposition**:

$$\mathbf{A} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

  This has the interpretation that $\mathbf{A}$ is a weighted sum of rank-one matrices, as for a square, symmetric matrix.

  Note that $-\mathbf{u}_i, -\mathbf{v}_i$ are also corresponding singular vectors to σ_i:

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^{r} \sigma_i (-\mathbf{u}_i)(-\mathbf{v}_i)^T.$$

  This shows that the SVD of a matrix is not unique.

- **Truncated SVD**: For any integer $1 \leq K \leq r$, let $\sigma_1, \ldots, \sigma_K$ represent the largest K singular values of $\mathbf{A}$ with corresponding left and right singular vectors $(\mathbf{u}_i, \mathbf{v}_i), 1 \leq i \leq K$. We define the K-term truncated SVD of $\mathbf{A}$ as

$$\mathbf{A} \approx \underbrace{\sum_{i=1}^{K} \sigma_i \mathbf{u}_i \mathbf{v}_i^T}_{\mathbf{A}_K}$$

Note that $\mathbf{A}_K$ has a rank of K and it can be regarded as a low-rank approximation to $\mathbf{A}$ (if K is small).

Geometric interpretation of SVD

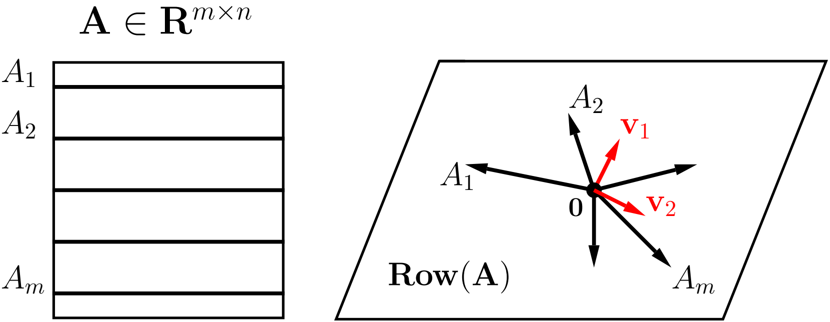
Given any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with rank r, let its compact SVD be

$$\mathbf{A} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T.$$

We rewrite it in the following way:

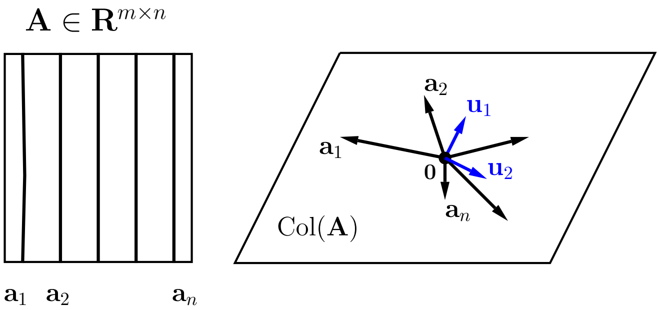
$$\begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} = \mathbf{A} = \underbrace{(\mathbf{U}_r \mathbf{\Sigma}_r)}_{\text{coefficients}} \cdot \underbrace{\mathbf{V}_r^T}_{\text{basis}} = \begin{bmatrix} * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}.$$

This shows that the rows of $\mathbf{V}_r^T$ (columns of $\mathbf{V}_r$) provide an orthonormal basis for the row space of $\mathbf{A}$.



Similarly, the columns of $\mathbf{U}_r$ provide an orthonormal basis for the column space of $\mathbf{A}$:

$$\mathbf{A} = \underbrace{\mathbf{U}_r}_{\text{basis}} \cdot \underbrace{\left(\mathbf{\Sigma}_r \mathbf{V}_r^T\right)}_{\text{coefficients}} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_r\end{bmatrix} \cdot \left(\mathbf{\Sigma}_r \mathbf{V}_r^T\right).$$



Example 0.4. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

By direct calculation, we obtain the compact SVD of $\mathbf{A}$ as follows:

$$\mathbf{U}_2 = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 3 & \\ & 1 \end{pmatrix}, \quad \mathbf{V}_2^T = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

Therefore, $\{\mathbf{v}_1^T, \mathbf{v}_2^T\}$ forms an orthonormal basis for the row space of $\mathbf{A}$, and the spanning coefficients for the row vectors of $\mathbf{A}$ are along the rows of the following matrix

$$\mathbf{U}_2\mathbf{\Sigma}_2 = \begin{pmatrix} \sqrt{6} & 0 \\ \frac{3}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Similarly, $\{\mathbf{u}_1, \mathbf{u}_2\}$ forms an orthonormal basis for $\mathrm{Col}(\mathbf{A})$ and the spanning coefficients for the columns of $\mathbf{A}$ are along the columns of

$$\mathbf{\Sigma}_2\mathbf{V}_2^T = \begin{pmatrix} 3 & \\ & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix} = \begin{pmatrix} \frac{3}{\sqrt{6}} & \frac{3}{\sqrt{6}} & \sqrt{6} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

MATLAB commands for computing matrix SVD

1. Full SVD

svd – Singular Value Decomposition.

[U,S,V] = svd(X) produces a diagonal matrix S, of the same dimension as X and with nonnegative diagonal elements in decreasing order, and orthogonal matrices U and V so that X = U*S*V^T.

s = svd(X) returns a vector containing the singular values.

2. Truncated SVD

svds – Find a few singular values and vectors.

S = svds(A,K) computes the K largest singular values of A.

[U,S,V] = svds(A,K) computes the singular vectors as well. If A is M-by-N and K singular values are computed, then U is M-by-K with orthonormal columns, S is K-by-K diagonal, and V is N-by-K with orthonormal columns.

In many applications, a truncated SVD is enough, and it is much easier to compute than the full SVD.

3. SVD Sketch

[U,S,V] = svdsketch(A) returns the singular value decomposition (SVD) of a low-rank matrix sketch of A. The matrix sketch only reflects the most important features of A (up to a tolerance), which enables faster calculation of the SVD of large matrices compared to using SVDS.

[U,S,V] = svdsketch(A, tol) specifies a tolerance for the sketch of A such that norm(U*S*V'-A,'fro')/norm(A,'fro') <= tol.

Power method for numerical computing of SVD

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix whose SVD is to be computed: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Consider $\mathbf{C} = \mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n \times n}$. We have

$$\begin{aligned}
\mathbf{C} &= \mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})\mathbf{V}^T = \sum \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T \\
\mathbf{C}^2 &= \mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})^2\mathbf{V}^T = \sum \sigma_i^4 \mathbf{v}_i \mathbf{v}_i^T \\
&\vdots \\
\mathbf{C}^k &= \mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})^k\mathbf{V}^T = \sum \sigma_i^{2k} \mathbf{v}_i \mathbf{v}_i^T
\end{aligned}$$

If $\sigma_1 > \sigma_2$, then the first term dominates, so

$$\mathbf{C}^k \to \sigma_1^{2k} \mathbf{v}_1 \mathbf{v}_1^T, \quad \text{as } k \to \infty.$$

Note that $\mathbf{v}_1\mathbf{v}_1^T$ is a rank-1 matrix, with columns being multiples of $\mathbf{v}_1$.

This means that a close estimate of $\mathbf{v}_1$ can be computed by simply taking the first column of $\mathbf{C}^k$ (for some large k) and normalizing it to a unit vector.

This method works but can be very costly due to the matrix power part, which has a complexity of $\mathcal{O}(n^3)$.

A better approach. Instead of computing $\mathbf{C}^k$, we select a random vector $\mathbf{x} \in \mathbb{R}^n$ and compute $\mathbf{C}^k\mathbf{x}$ through a sequence of matrix-vector multiplications (which are very efficient especially when one dimension of $\mathbf{A}$ is small, or $\mathbf{A}$ is sparse):

$$\mathbf{C}^k\mathbf{x} = \mathbf{A}^T\mathbf{A}\cdots\mathbf{A}^T\mathbf{A}\mathbf{x}$$

Write $\mathbf{x} = \sum c_i\mathbf{v}_i$ (since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form an orthonormal basis for $\mathbb{R}^n$). Then

$$\mathbf{C}^k\mathbf{x} \approx (\sigma_1^{2k}\mathbf{v}_1\mathbf{v}_1^T)\left(\sum c_i\mathbf{v}_i\right) = \sigma_1^{2k}c_1\mathbf{v}_1.$$

Normalizing the vector $\mathbf{C}^k\mathbf{x}$ for some large k then yields $\mathbf{v}_1$, the first right singular vector of $\mathbf{A}$.

Applications of SVD

The matrix SVD has lots of applications such as

- **Orthogonal best-fit plane**
- **Dimension reduction**
- Image compression[1]
- Recommender systems (matrix completion)[2]

We will cover the first two applications later in the course.

[1] https://www.mathworks.com/help/matlab/math/image-compression-with-low-rank-svd.html

[2] https://engineering.purdue.edu/ChanGroup/ECE695Notes/Lecture_SVT.pdf