Singular Value Decomposition

We know any symmtric square matrix A can be diagonalized by orthogonal transforms via eigenvalue decomposition $A = X \Lambda X^T$.

Can we extend it to an arbitrary matrix ?

Yes. (by sacrifying some property in eigen value decomposition)

Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ be an arbitrary real matrix.

① Since $A^T A \in \mathbb{R}^{n \times n}$ is square, symmtric and spsd.

Let (λ_i, v_i), $i = 1, 2, \cdots, n$ are eigenpairs of $A^T A$ with

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0. \quad \text{(Because } A^T A \text{ is spsd)}$$

So, the eigenvalue decomposition of $A^T A$ is

$$A^T A = V \Lambda V^T, \text{ where } V = [v_1\, v_2 \cdots v_n] \text{ and } \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

② $AA^T \in \mathbb{R}^{m \times m}$ is also symmetric, square, and spsd.

Also, by a theorem of linear algebra, AA^T and $A^T A$ share the same non-zero eigenvalues,

i.e., the eigenvalues of AA^T are $\{\lambda_1, \lambda_2, \cdots, \lambda_n, \overbrace{\underset{\uparrow \lambda_{n+1}}{0}, \cdots, \underset{\uparrow \lambda_m}{0}}^{m-n \text{ 0's}}\}$.

Let (λ_i, u_i), $i = 1, \cdots, m$ be eigenpairs of AA^T. Then

$$AA^T = [\underbrace{u_1\, u_2 \cdots u_n}_{U} \mid u_{n+1} \cdots u_m] \begin{bmatrix} \Lambda & \\ & 0 \end{bmatrix} \left.\begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \\ \hline u_{n+1}^T \\ \vdots \\ u_m^T \end{bmatrix}\right\} U^T$$

$$= U \Lambda U^T$$

③ Let's construct a decomposition of A by U, V and Λ.

$$AA^T(Av_i) = A(A^T A v_i) = A(\lambda_i v_i) = \lambda_i (A v_i) \qquad i = 1, 2, \cdots, n$$

$\Rightarrow$ Av_i is an eigenvector of AA^T with eigenvalue λ_i.

$$\Rightarrow \boxed{A v_i = \sigma_i u_i, \quad i = 1, 2, \cdots, n.} \quad ①$$

Similarly,

$$A^TA(A^Tu_i) = A^T(AA^Tu_i) = A^T(\lambda_i u_i) = \lambda_i(A^Tu_i). \quad i=1,2,\cdots,n$$

$\Rightarrow$ A^Tu_i is an eigenvector of A^TA with eigenvalue λ_i

$\Rightarrow$ $A^Tu_i = \tilde{\sigma}_i v_i$ for some $\tilde{\sigma}_i \in \mathbb{R}$

Let us find $\tilde{\sigma}_i$: Multiplying v_i^T from the left gives

$$v_i^TA^Tu_i = \tilde{\sigma}_i v_i^Tv_i = \tilde{\sigma}_i .$$

Because $v_i^TA^Tu_i$ is a number,

$$\tilde{\sigma}_i = v_i^TA^Tu_i = (v_i^TA^Tu_i)^T = u_i^TAv_i \overset{\text{by ①}}{=} u_i^T(\sigma_i u_i) = \sigma_i$$

Therefore, $\tilde{\sigma}_i = \sigma_i$. Thus,

$$\Rightarrow \boxed{A^Tu_i = \sigma_i v_i, \quad i=1,2,\cdots,n.} \quad ②$$

Further,

$$\lambda_i v_i = A^TAv_i = \sigma_i A^Tu_i = \sigma_i(\sigma_i v_i) = \sigma_i^2 v_i$$

$$\Rightarrow \sigma_i^2 = \lambda_i \Rightarrow \boxed{\sigma_i = \sqrt{\lambda_i}, \quad i=1,2,\cdots,n.} \quad ③$$

Write ① into matrix form, we obtain

$$A\underbrace{[v_1\ v_2\ \cdots\ v_n]}_{V} = \underbrace{[u_1\ u_2\ \cdots\ u_n]}_{U}\underbrace{\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n\end{bmatrix}}_{\Sigma}$$

i.e.,
$$AV = U\Sigma$$

Since $V \in \mathbb{R}^{n\times n}$ is square and orthogonal, right multiplying V^T on both hands sides gives

$$A\underbrace{VV^T}_{I} = U\Sigma V^T \Rightarrow A = U\Sigma V^T.$$

A is diagonized by U and V^T. This decomposition is called singular value decomposition (SVD).

σ_i, $i=1,\cdots,n$ are called singular values.

u_i, v_i, $i=1,\cdots,n$ are called singular vectors.

(σ_i, u_i, v_i), $i=1,\cdots,n$ are called singular triplets.

When $A \in \mathbb{R}^{m\times n}$ with $m \le n$, AA^T will has a smaller size than A^TA.

Thus, a similar anlalysis gives

$$\begin{cases} Av_i = \sigma_i u_i & i=1,\cdots,m. \\ A^T u_i = \sigma_i v_i & i=1,\cdots,m. \\ \sigma_i = \sqrt{\lambda_i} & i=1,\cdots,m. \end{cases}$$

Thus $$A^T \underbrace{[u_1\ u_2 \cdots u_m]}_{U} = \underbrace{[v_1\ v_2 \cdots v_m]}_{V} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{bmatrix}}_{\Sigma}$$

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

Since $U \in \mathbb{R}^{m\times m}$ is square, $UU^T = I$.

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

$$\Rightarrow A = U\Sigma V^T. \quad \text{(This is the SVD).}$$

In summary:

Singular Value Decomposition (SVD):

For any matrix $A \in \mathbb{R}^{m\times n}$, there exists SVD:

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

where $U = [u_1\ u_2 \cdots u_p] \in \mathbb{R}^{m\times p}$ with $p = \min\{m,n\}$ satisfying $U^TU = I$,

$V = [v_1\ v_2 \cdots v_p] \in \mathbb{R}^{n\times p}$ satisfying $V^TV = I$,

$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{bmatrix} \in \mathbb{R}^{p\times p}$ satisfying $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_p \ge 0$.

Furthermore, σ_i^2 are eigenvalues of AA^T and A^TA, and u_i and v_i are the corresponding eigenvectors.

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

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = U \Sigma V^T$$

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

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = U \Sigma V^T$$

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

$$= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = U \Sigma V^T$$

Basic Properties of SVD:

① From the derivation

$$\begin{cases} A v_i = \sigma_i u_i \\ A^T u_i = \sigma_i v_i \end{cases} \quad i = 1, 2, \cdots, p, \quad p = \min\{m, n\}.$$

② $\operatorname{rank}(A) = r$ if and only if

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \text{ and } \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_p = 0.$$

That is: rank(A) is the number of non-zero singular values.

③ We can rewrite the SVD as

$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T$$

If $\operatorname{rank}(A) = r$, then the SVD is

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

$$= [u_1 \ u_2 \ \cdots \ u_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} [v_1 \ v_2 \ v_r]^T$$

This is called compact SVD of A.

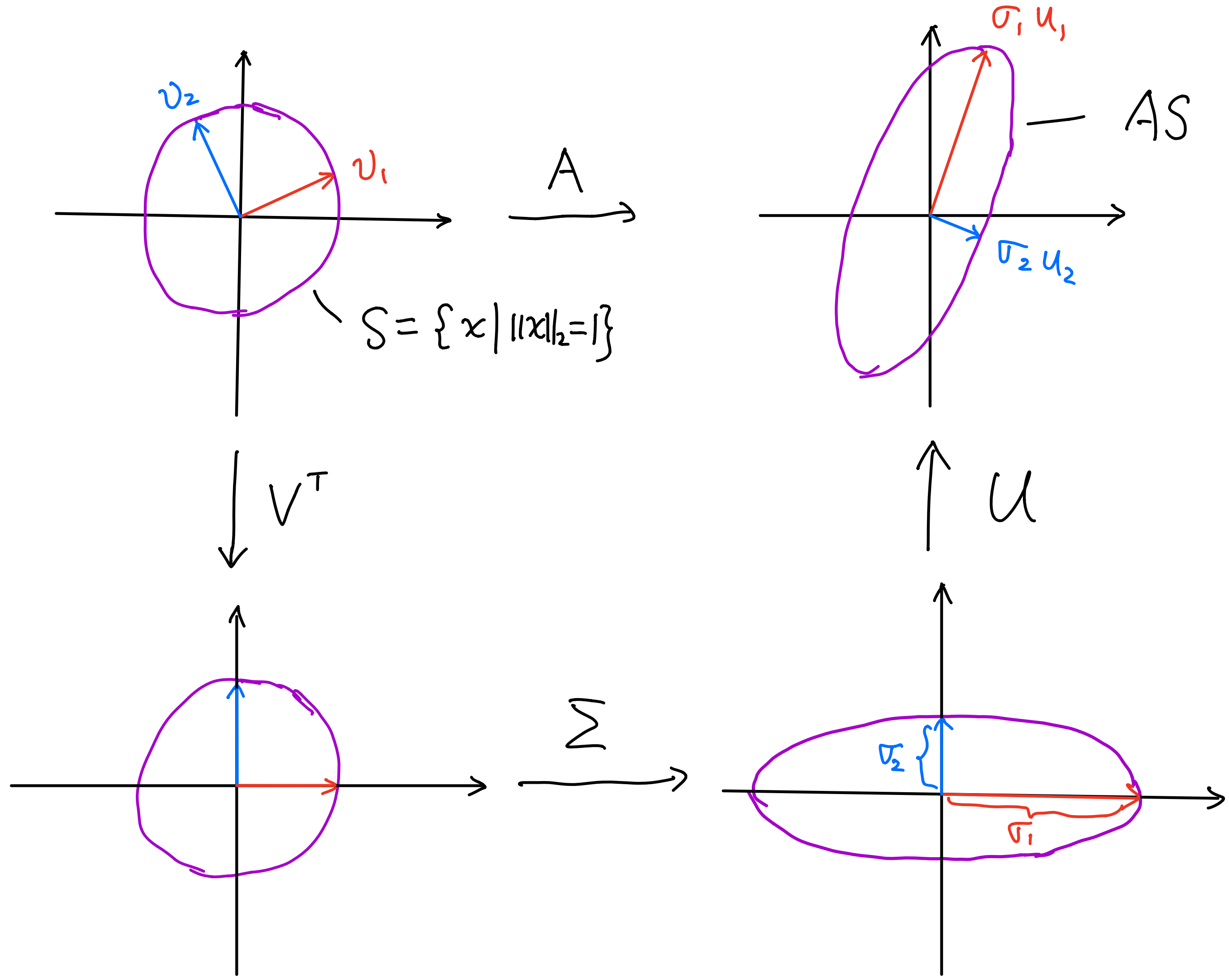
④ $\operatorname{Ran}(A) = \{Ax \mid x \in \mathbb{R}^n\} = \{\sum_{i=1}^{r} \sigma_i u_i v_i^T x \mid x \in \mathbb{R}^n\} = \{\sum_{i=1}^{r} (\sigma_i v_i^T x) u_i \mid x \in \mathbb{R}^n\}$

$= \operatorname{span}\{u_1, u_2, \cdots, u_r\}$

Similarly, $\operatorname{Ran}(A^T) = \operatorname{span}\{v_1, v_2, \cdots, v_r\}$

Geometry of SVD

Recall $A v_i = \sigma_i u_i$, $i = 1, 2, \cdots, n$.



(Application of a unitary matrix means rotation and/or reflection)

(Application of a diagonal matrix means scaling the axises)

Let $S = \{x \mid \|x\|_2 = 1\}$ be the 2-norm unit ball in $\mathbb{R}^n$

Then the image AS is an ellipsoid in $\mathbb{R}^m$

— u_i, $i = 1, \cdots, n$, are directions of axises of AS, and σ_i is the length of the corresponding axis.

SVD is a fundamental tool for matrix analysis and computation.

★ Many matrix properties can be revealed by SVD.

- $\text{rank}(A) = $ # of non-zero singular values

- $\mathrm{Ran}(A) = \mathrm{span}\{u_1, u_2, \cdots, u_r\}$
- $\mathrm{Ker}(A) = \mathrm{Span}\{v_1, v_2, \cdots, v_r\}^{\perp} = \mathrm{span}\{v_{r+1}, v_{r+2}, \cdots, v_n\}$.
- $\|A\|_F = (\sigma_1^2 + \cdots + \sigma_p^2)^{1/2}$ and $\|A\|_2 = \sigma_1$

proof. ① $\|A\|_F^2 = \sum_{i,j} a_{ij}^2 = \mathrm{trace}(A^TA) = \sum_{i=1}^{p} \lambda_i$ (λ_i are eigenvalues of A^TA)

$= \sum_{i=1}^{p} \sigma_i^2$

② $\|A\|_2^2 = \max_{\|x\|_2=1} \|Ax\|_2^2 = \max_{\|x\|_2=1} \|U\Sigma V^Tx\|_2^2 = \max_{\|x\|_2=1} \|\Sigma V^Tx\|_2^2$

Assume $m \geq n$. Then $\|V^Tx\|_2 = \|x\|_2$. So,

$$\|A\|_2^2 = \max_{\|x\|_2=1} \|\Sigma V^Tx\|_2^2 = \max_{\|y\|_2=1} \|\Sigma y\|_2^2 = \sigma_1^2.$$

The case $m \leq n$ can be done similarly. ☒

☆ Best low-rank approximation

Theorem: (Eckart–Young–Mirsky theorem):

Let $A \in \mathbb{R}^{m\times n}$ with SVD $A = U\Sigma V^T$. Define

$$A_k \equiv [u_1\ u_2 \cdots u_k] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_k \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{bmatrix} = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

Obviously, $\mathrm{rank}(A_k) = k$. Then, A_k is the best rank-k approximation of A under 2-norm or Frobenius norm, i.e.,

$$A_k = \arg\min_{\mathrm{rank}(B)\leq k} \|A - B\|_F \quad \text{and} \quad A_k = \arg\min_{\mathrm{rank}(B)\leq k} \|A - B\|_2$$

Furthermore,

$$\|A - A_k\|_F = (\sigma_{k+1}^2 + \cdots + \sigma_n^2)^{1/2} \text{ and } \|A - A_k\|_2 = \sigma_{k+1}.$$

proof. Step 1: Show that $\min_{\mathrm{rank}(B)\leq k} \|A - B\|_2 = \sigma_{k+1} = \|A - A_k\|_2$

For any $B \in \mathbb{R}^{m\times n}$ satisfying $\mathrm{rank}(B) \leq k$.

Consider $\{Bv_1, Bv_2, \cdots, Bv_{k+1}\} \subset \mathrm{Ran}(B)$

Since $\mathrm{rank}(B) \leq k$, the range of B is at most k-dimensional.

$\Longrightarrow$ $Bv_1, Bv_2, \cdots, Bv_{k+1}$ are linearly dependent.

$\Rightarrow \exists (x_1, \cdots, x_{k+1}) \neq 0$ s.t. $\sum_{i=1}^{k+1} x_i B v_i = 0$

Let $v = \sum_{i=1}^{k+1} x_i v_i \Rightarrow Bv = \sum_{i=1}^{k+1} x_i B v_i = 0$

$$\Rightarrow \|A-B\|_2 \geq \frac{\|(A-B)v\|_2}{\|v\|_2} = \frac{\|Av\|_2}{\|v\|_2}$$

$$= \frac{\|\sum_{i=1}^{k+1} x_i A v_i\|_2}{\|\sum_{i=1}^{k+1} x_i v_i\|_2} = \frac{\|\sum_{i=1}^{k+1} x_i \sigma_i u_i\|_2}{\|\sum_{i=1}^{k+1} x_i v_i\|_2} = \left(\frac{\sum_{i=1}^{k+1} x_i^2 \sigma_i^2}{\sum_{i=1}^{k+1} x_i^2}\right)^{1/2} \geq \sigma_{k+1}$$

Since the above inequality holds for any B satisfying $\operatorname{rank}(B) \leq k$,

$$\min_{\operatorname{rank}(B)\leq k} \|A-B\|_2 \geq \sigma_{k+1}$$

On the other hand,

$$\min_{\operatorname{rank}(B)\leq k} \|A-B\|_2 \leq \|A-A_k\|_2 = \|\sum_{i=k+1}^{n} \sigma_i u_i v_i^T\|_2 = \sigma_{k+1}$$

Altogether, $\min_{\operatorname{rank}(B)\leq k} \|A-B\|_2 = \sigma_{k+1} = \|A-A_k\|_2$.

Step 2: Show that $\min_{\operatorname{rank}(B)\leq k} \|A-B\|_F = \left(\sum_{i=k+1}^{n} \sigma_i^2\right)^{1/2} = \|A-A_k\|_F$

We have shown in Step 1 that:

$$\forall B: \operatorname{rank}(B) \leq k, \qquad \sigma_{k+1}(A) = \sigma_1(A-A_k) \leq \sigma_1(A-B).$$

We further show that

$$\sigma_{k+i}(A) \leq \sigma_i(A-B) \text{ for } i \geq 2 \text{ and } \operatorname{rank}(B) \leq k. \qquad (1)$$

To this end, let $B \in \mathbb{R}^{m\times n}$ satisfying $\operatorname{rank}(B) \leq k$.

From step 1, the best rank-$(i-1)$ approximation C to $A-B$ satisfies

$$\sigma_1(A-B-C) = \|A-B-C\|_2 = \sigma_i(A-B) \qquad (2)$$

Since $\operatorname{rank}(B+C) \leq \operatorname{rank}(B) + \operatorname{rank}(C) \leq k+(i-1)$

$$\|A-(B+C)\|_2 \geq \min_{\operatorname{rank}(D)\leq k+i-1} \|A-D\|_2 \underset{\uparrow}{=} \sigma_{k+i}(A) \qquad (3)$$

from step 1 again

Here $\sigma_{k+i}(A) = 0$ if $k+i > n$.

(2) & (3) together implies (1).

With (1),

$$\|A-B\|_F^2 = \sum_{i=1}^{n} (\sigma_i(A-B))^2 \geq \sum_{i=1}^{n} (\sigma_{i+k}(A))^2 = \sum_{i=k+1}^{n} \sigma_i^2 = \|A-A_k\|_F^2$$

which implies $\min_{\operatorname{rank}(B)\leq k} \|A-B\|_F^2 \geq \|A-A_k\|_F^2$.

7

Since $\operatorname{rank}(A_k) \le k$, $\quad \|A-A_k\|_F^2 \ge \min_{\operatorname{rank}(B)\le k} \|A-B\|_F^2$ $\}$

$$\Longrightarrow \min_{\operatorname{rank}(B)\le k} \|A-B\|_F = \|A-A_k\|_F = \left(\sum_{i=k+1}^{n} \sigma_i^2\right)^{1/2}. \qquad \boxtimes$$

Example:

① $A = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$,

The SVD is $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & \\ & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = 3e_1e_1^T + 2e_2(-e_2)^T$

So a best rank 1 approximation is $A_1 = 3e_1e_1^T = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$

$\|A-A_1\|_F = 2$ and $\|A-A_1\|_2 = 2$.

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

The SVD is $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 3 & \\ & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = 3e_2e_1^T + 2e_1(-e_2)^T$

So a best rank 1 approximation is $A_1 = 3e_2e_1^T = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix}$

$\|A-A_1\|_F = 2$ and $\|A-A_1\|_2 = 2$.

Computation of SVD

Given $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$ with SVD $A = U\Sigma V^T$
it is seen from the previous chapter that

$$A^TA = V\Sigma^2V^T \quad \text{and} \quad AA^T = U\Sigma U^T$$

Therefore, computation of SVD of A can be converted to eigenvalue/eigenvector problems of A^TA or AA^T.

Power Iteration:

$$
\begin{aligned}
&\text{Choose } v^{(0)} \text{ s.t. } \|v^{(0)}\|_2 = 1\\
&\text{for } k = 1, 2, \dots\\
&\qquad y^{(k)} = A v^{(k-1)}\\
&\qquad u^{(k)} = y^{(k)} / \|y^{(k)}\|_2\\
&\qquad z^{(k)} = A^T u^{(k)}\\
&\qquad v^{(k)} = z^{(k)} / \|z^{(k)}\|_2\\
&\qquad \sigma^{(k)} = \|Av^{(k)}\|_2\\
&\text{end}
\end{aligned}
$$

We will have
$v^{(k)} = A^TA v^{(k-1)} / \|A^TAv^{(k-1)}\|_2$
and
$u^{(k)} = AA^Tu^{(k-1)} / \|AA^Tu^{(k-1)}\|_2$
So $v^{(k)} \to v_1$ and $u^{(k)} \to u_1$

Assume $y^{(k)} \approx v_1$
Then $\sigma^{(k)} \approx \|Av_1\|_2 = \|\sigma_1 u_1\|_2 = \sigma_1$
So $\sigma^{(k)} \to \sigma_1$

Variants of power iteration can be applied.

QR algorithm:

1. Compute $A^TA = V(\Sigma^2)V^T$ by QR algorithm.
2. To recover U stably:

   Compute QR decomposition of $AV = QR$.

   Then $U = Q$.

We don't need to form A^TA explicitly in the QR algorithm for

A^TA :

Phase I: Reduce A to bidiagonal.

$$P_1 A = \begin{bmatrix} \times & \times & \times & \cdots & \times \\ 0 & \times & \times & \cdots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \times & \times & \cdots & \times \end{bmatrix}$$

P_1 — Householder that transform $A(1:n, 1)$ to $c e_1$.

$$P_1 A O_1^T = \begin{bmatrix} \times & \times & 0 & \cdots & 0 \\ 0 & \times & \times & \cdots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \times & \times & \cdots & \times \end{bmatrix}$$

O_1 — Householder that transform the row vector $((P_1 A)(1, 2:n))^T$ to $c e_1$.

The first column will not be modified.

After $n-1$ steps,

$$B_0 = P_{n-1} \cdots P_2 P_1 A O_1^T O_2^T \cdots O_{n-1}^T = \begin{bmatrix} \times & \times & & & \\ & \ddots & \ddots & & \\ & & \ddots & \times & \\ & & & \times & \\ & & & & \end{bmatrix}$$

and $T_0 = B_0^T B_0 = (O_{n-1} \cdots O_2 O_1) A^T A (O_1^T O_2^T \cdots O_{n-1}^T)$ is tridiagonal.

Phase 2: Compute eigenvalue/eigenvectors of $T_0 = B_0^T B_0$ by QR algorithm

There are also smart implementations that works on the bi-diagonal matrix B_0. The details are omitted.

- Some other methods converts SVD to eigenvalue decomposition of some other matrices.

For example, let (σ_i, u_i, v_i) are singular triplets of A.

Then,

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} A v_i \\ A^T u_i \end{bmatrix} = \begin{bmatrix} \sigma_i u_i \\ \sigma_i v_i \end{bmatrix} = \sigma_i \begin{bmatrix} u_i \\ v_i \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u_i \\ -v_i \end{bmatrix} = \begin{bmatrix} -Av_i \\ A^T u_i \end{bmatrix} = \begin{bmatrix} -\sigma_i u_i \\ \sigma_i v_i \end{bmatrix} = -\sigma_i \begin{bmatrix} u_i \\ -v_i \end{bmatrix}$$

Therefore, $\pm\sigma_i$ are eigenvalues of $\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$ and the corresponding eigenvectors are $\begin{pmatrix} u_i \\ \pm v_i \end{pmatrix}$.

To compute SVD (i.e., σ_i, u_i, v_i) of A, we only compute eigenvalue decomposition of the symmetric matrix

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix},$$

which can be done by power iterations or QR algorithms.