Singular Value Decomposition (SVD).

1 Motivation

We've seen that a key step solving some problems, like:

1) Finding $A^n \underline{x}$ for many n's;

2) Solving a differential equation (in the optional lecture about oscillations of the H_2O molecule),

and in many others, which we have not explored, requires having a basis of eigenvectors of A.

When A is diagonalizable,

$$A = VDV^{-1},$$

we have such a basis: it consists of columns of V.

What if A is not diagonalizable? Then it is defective (Sec. 4.5) and does not have enough eigenvectors for a basis.

How can we then proceed to solve our problems?

In Sec 4.7 – part II we saw one possible approach. Namely, let a 2×2 A has only one eigenvector, $\underline{u}$.

Then take $\underline{v} \perp \underline{u}$ (where $\underline{v}$ is not an eigenvector of A); then $\{\underline{u}, \underline{v}\}$ form an (orthogonal) basis.

We didn't study it, but, it turns out, this $\underline{v}$ is, in some sense, not an eigenvector "but the next best thing to an eigenvector", called a <u>generalized eigenvector</u>.

However, in this lecture we will <u>not</u> pursue the "generalized eigenvectors path". Instead, we will study a completely different approach – a singular value decomposition (SVD) of a matrix.

Consider $A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$, $a \neq 0$. (1)
One can easily check that it is <u>not diagonalizable</u>: the eigenvalue $\lambda = 1$ has alg. multiplicity $= 2$ and the geometric multiplicity $= 1$ (eigenvector $= \begin{pmatrix} 1 \\ 0 \end{pmatrix}$).

Let us recall that eigenvalues of a <u>diagonalizable</u> matrix tell us something about the length of $(A^k \underline{x})$ for an "initial" vector $\underline{x}$ and a sufficiently large k:

$$A^k \underline{x} \approx c \cdot \lambda_{max}^k \cdot \underline{v}, \qquad (2)$$

where the eigenvector $\underline{v}$ corresponds to the largest $|\lambda|$, denoted $|\lambda|_{max}$.

Even though our A is not diagonalizable, we can try to use that idea and look at the length of $A\underline{x}$. We'll actually look at $\|A\underline{x}\|^2$ to avoid taking the $\sqrt{\ }$.

$$\|A\underline{x}\|^2 = (A\underline{x})^T(A\underline{x}) = \underline{x}^T\underbrace{(A^TA)}_{B}\underline{x} \qquad (3)$$

We know that $\boxed{B = A^TA}$ is symmetric (Sec. 1.6) and hence diagonalizable; therefore, its eigenvectors form a basis (Sec. 4.7). Moreover, they can be chosen to form a nice, orthonormal basis;

see Property 1 of real symmetric matrices (Sec. 4.7, part II).

So, let $\{\underline{v}_1, \underline{v}_2\}$ be the orthonormal eigenvectors of B:

$$B\underline{v}_i = (\lambda_B)_i \, \underline{v}_i \quad , \quad i=1,2 \qquad (4)$$

Then, we can expand our arbitrary $\underline{x}$ as

$$\underline{x} = c_1 \underline{v}_1 + c_2 \underline{v}_2 \qquad (5)$$

and find $\|A\underline{x}\|^2$ from (3), (4), (5):

$$\|A\underline{x}\|^2 = (c_1 \underline{v}_1^T + c_2 \underline{v}_2^T)\, B\, (c_1 \underline{v}_1 + c_2 \underline{v}_2)$$

$$= (c_1 \underline{v}_1^T + c_2 \underline{v}_2^T)\left((\lambda_B)_1 c_1 \underline{v}_1 + (\lambda_B)_2 c_2 \underline{v}_2\right)$$

$$= c_1^2 (\lambda_{B_1}) \cancelto{1}{(\underline{v}_1^T \underline{v}_1)} + c_1 c_2 (\lambda_B)_2 \cancelto{0}{\underline{v}_1^T \underline{v}_2}$$

$$+ c_2 c_1 (\lambda_B)_1 \cancelto{0}{(\underline{v}_2^T \underline{v}_1)} + c_2^2 (\lambda_B)_2 \cancelto{1}{(\underline{v}_2^T \underline{v}_2)}$$

$$= c_1^2 (\lambda_B)_1 + c_2^2 (\lambda_B)_2 . \qquad (6)$$

Thus, one can find the length of $A\underline{x}$ (and, similarly, of $A^k\underline{x}$) if one knows:

- the eigenvalues λ_B of $B = A^T A$, and
- the eigenvectors of B (which determine coordinates c_1, c_2 in (6)).

This motivates us to further study the eigenpairs of B: $\{(\lambda_B)_i, \underline{v}_i\}$.

2 SVD of matrix A

Properties of $((\lambda_B)_i, \underline{v}_i)$:

$\boxed{1}$ $\underline{v}_i \perp \underline{v}_j$ for $i \neq j$, and one can choose $\|\underline{v}_i\| = 1$.
Thus, the basis $\{\underline{v}_1, \underline{v}_2\}$ is orthonormal, and matrix

$$\mathbb{V} \equiv [\underline{v}_1, \underline{v}_2] \text{ is orthogonal.} \qquad (7)$$

This is nothing but Property 1 of a real symmetric matrix B (p. 21-15) and its Corollary (p. 21-16).

$\boxed{2}$ All $(\lambda_B)_i \geq 0$. (8)

<u>Proof</u>: We know from Property 2 of a symmetric matrix (p. 21-16) that all $(\lambda_B)_i$ are real. However, (8) is a stronger statement.

To prove it, consider

$$\underset{\substack{\uparrow \\ \text{length} \geq 0}}{0 \leq} \| A\, \underline{v}_i \|^2 = (A\,\underline{v}_i)^T (A\,\underline{v}_i) = \underline{v}_i^T \underbrace{(A^T A)}_{B} \underline{v}_i = (\lambda_B)_i \underbrace{\|\underline{v}_i\|^2}_{=1} \quad (9)$$

$$\Rightarrow (\lambda_B)_i \geq 0.$$

For future convenience, we define

$$\text{sigma} \longrightarrow \sigma_i \equiv \sqrt{(\lambda_B)_i} \geq 0 \quad (10)$$

Note that it follows from (9) & (10) that

$$\sigma_i = \| A\, \underline{v}_i \|. \quad (11)$$

$\boxed{3}$ $A\,\underline{v}_i \perp A\,\underline{v}_j$ for $i \neq j$ (12)

The proof is very similar to (9):

we need to compute

$$(A\underline{v}_i)^T(A\underline{v}_j) = \underline{v}_i^T B \underline{v}_j = (\lambda_B)_j \underbrace{\cancel{\underline{v}_i^T \underline{v}_j}}_{=0 \text{ By } \boxed{1}} = 0,$$

whence (12) follows.

Motivated by Property $\boxed{3}$, we introduce <u>unit</u> vectors

$$\underline{u}_i = \frac{1}{\sigma_i} A \underline{v}_i . \qquad (13)$$

As follows from above:

$$\underline{u}_i \perp \underline{u}_j \ (i \neq j); \quad \|\underline{u}_i\| = 1. \qquad (14)$$

Therefore, matrix $U \equiv [\underline{u}_1, \underline{u}_2]$ is also orthogonal.

<u>The SVD Thm.</u> (for 2×2 matrices).

Let A be 2×2 and V, U and $\sigma_{1,2}$ be as defined in (7), (14), and (11). Then

$$\boxed{A = U \underbrace{\Sigma}_{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}} \underbrace{V^T}_{V^{-1}, \text{ since } V \text{ is orthogonal}}} \qquad (15).$$

Proof: We need to show that:

$$A\underline{x} = U \Sigma V^T \underline{x} \quad \text{for any } \underline{x} \tag{16}$$

(same idea as for the spectral decomposition of a symmetric matrix, p. 21-17).

Let us expand

$$\underline{x} = c_1 \underline{v}_1 + c_2 \underline{v}_2 \equiv V \underline{c}$$

which is possible since $\{\underline{v}_1, \underline{v}_2\}$ is a basis.

lhs of (16):

$$A\underline{x} = AV\underline{c} = [A\underline{v}_1, A\underline{v}_2]\,\underline{c}$$

direct calculation

$$= \left[\frac{A\underline{v}_1}{\sigma_1}, \frac{A\underline{v}_2}{\sigma_2}\right]\begin{pmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{pmatrix}\underline{c}$$

$$= U\Sigma\underline{c} = U\Sigma V^{-1}\underline{x}$$

V is orthogonal

$$= U\Sigma V^T \underline{x}.$$

q.e.d.

3 Interpretation of SVD (15)

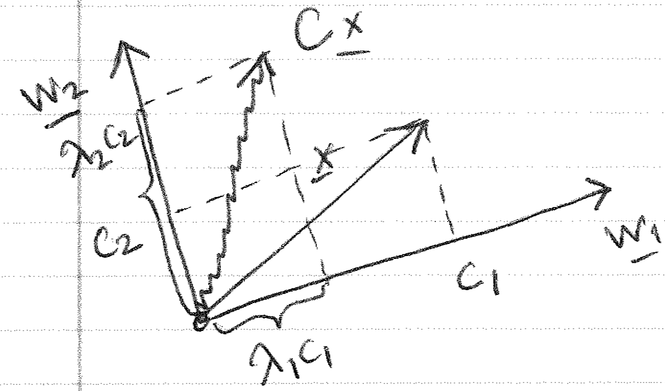
Let us first recall the interpretation of multiplication by a symmetric matrix (pp. 21-18, 21-19).

- Let C be real symmetric, with eigenpairs $(\lambda_i, \underline{w}_i)$, where $\underbrace{\underline{w}_1, \underline{w}_2}$ form an orthonormal basis.

$W = [\underline{w}_1, \underline{w}_2]$.

Then

$$C\underline{x} = W \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \underbrace{W^{-1}\underline{x}}$$



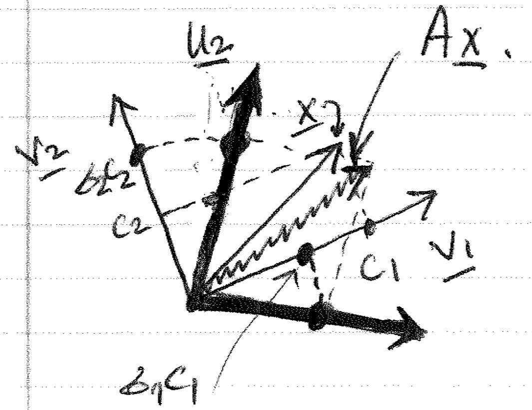
Project on $\underline{w}_1, \underline{w}_2$; get coord's c_1, c_2

stretch coord's; get $\lambda_1 c_1, \lambda_2 c_2$.

Use $\lambda_1 c_1, \lambda_2 c_2$ as <u>new</u> coord's in the same basis $\underline{w}_1, \underline{w}_2$

This is the only thing that will change in SVD.

- Now follow analogous steps for SVD.



Note that since $\{\underline{v}_1, \underline{v}_2\}$ and $\{\underline{u}_1, \underline{u}_2\}$ are two orthonormal bases, one of them is obtained by a <u>rotation</u> (or reflection) of the other.

Then:

$$A\underline{x} = U \, \underbrace{\Sigma \, \underbrace{V^{-1}\underline{x}}_{\text{project on } \underline{v}_1, \underline{v}_2;\ \text{get coord's } c_1, c_2}}_{\text{stretch coord's; get } \sigma_1 c_1,\ \sigma_2 c_2}$$

use $\sigma_1 c_1$, $\sigma_2 c_2$ as new coord's in the other basis, $\{\underline{u}_1, \underline{u}_2\}$.

4 Generalizations

If A is not square but is rectangular, the SVD will still work.

For example:

$$\underset{2\times 3}{A} = \underset{2\times 2}{U} \cdot \underset{2\times 3}{\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}} \underset{3\times 3}{V^T}$$

or

$$\underset{3\times 2}{A} = \underset{3\times 3}{U} \underset{3\times 2}{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix}} \underset{2\times 2}{V^T}.$$