1 Singular Value Decomposition, Continued (4.1, 4.2)

(Continued from previous notes.)

1.1 Relationship to Norm and Condition Number

Recall that we defined the matrix 2-norm as

$$||A|| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{||A\mathbf{x}||_2}{||\mathbf{x}||} = \sigma_1,$$

where σ_1 is the largest singular value. Note that this definition also makes sense for $A \in \mathbb{R}^{n\times m}$.

Theorem 1.1

$$||A||_2 = \sigma_1.$$

Since A and A^T have the same singular values, we have the following corollary.

Corollary 1.1

$$||A||_2 = ||A^T||_2.$$

Since A is nonsingular, A has full rank, i.e., rank n. A has n strictly positive singualr values, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$. Now,

$$A^{-1}Av_i = A^{-1}(\sigma_i u - i) \implies v_i = \sigma_i A^{-1}u_i \implies A^{-1}u_i = \frac{1}{\sigma_i}v_i,$$

so in particular we can map each σ like so:

A	A^{-1}
$v_1 \xrightarrow[\sigma_1]{} u_1$	$u_1 \xrightarrow[\sigma_1^{-1}]{} v_1$
$v_2 \xrightarrow[\sigma_2]{} u_2$	$u_2 \xrightarrow[\sigma_2^{-1}]{} v_2$
$v_3 \xrightarrow[\sigma_3]{} u_3$	$u_3 \xrightarrow[\sigma_3^{-1}]{} v_3$
$\vdots$	$\vdots$
$v_n \xrightarrow[\sigma_n]{} u_n$	$u_n \xrightarrow[\sigma_n^{-1}]{} v_n$

This tells us that the singular values of A^{-1} must be

$$\frac{1}{\sigma_n} \geq \frac{1}{\sigma_{n-1}} \geq \ldots \geq \frac{1}{\sigma_2} \geq \frac{1}{\sigma_1} > 0$$

such that

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_n} \end{bmatrix}.$$

And, in particular,

$$||A^{-1}||_2 = \frac{1}{\sigma_n} \qquad ||A||_2 = \sigma_1.$$

Theorem 1.2

Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix with singular values $\sigma_1 \geq \ldots \geq \sigma_n > 0$. Then,

$$\kappa_2(A) = \frac{\sigma_1}{\sigma_n}.$$

1.2 More on SVD

Remember that there are two types of SVD:

- Full SVD.
$$A = U\Sigma V^T,$$
where A is $n \times m$, U is $n \times n$, Σ is $n \times m$, and V^T is $m \times m$. Here, $\text{rank}(A) = r \leq m$ and $n \geq m$.
- Reduced SVD
$$A = \hat{U}\hat{\Sigma}\hat{V}^T,$$
where A is $n \times m$, $\hat{U}$ is $n \times r$, $\hat{\Sigma}$ is $r \times r$, and $\hat{V}^T$ is $r \times m$.

In any case, we now know that

$$||A||_2 = \sigma_1 \qquad \kappa_2(A) = \frac{\sigma_1}{\sigma_n}.$$

1.3 Rank-1 Decomposition

Theorem 1.3

Let $A \in \mathbb{R}^{n\times m}$ be a nonzero matrix with rank r. Let $\sigma_1, \ldots, \sigma_r$ be the singular values of A, with associated right and left singular vectors $v_1, \ldots, v_r$ and $u_1, \ldots, u_r$, respectively. Then,

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^T,$$

where $u_j \in \mathbb{R}^n$, $v_j \in \mathbb{R}^m$, and $u_j v_j^T \in \mathbb{R}^{n\times m}$.

To see why this theorem works,

$$\begin{aligned}
A &= \hat{U}\hat{\Sigma}\hat{V}^T \\
&= \underbrace{\begin{bmatrix} u_1 & v_2 & \ldots & u_r \end{bmatrix}}_{\hat{U}} \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_r \end{bmatrix}}_{\hat{\Sigma}} \underbrace{\begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \end{bmatrix}}_{\hat{V}} \\
&= \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \ldots & \sigma_r u_r \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \end{bmatrix} \\
&= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_r u_r v_r^T \\
&= \sum_{i=1}^{r} \sigma_i u_i v_i^T.
\end{aligned}$$

This is called the **rank-1 decomposition** because A is written as a sum of rank-1 matrices ($u_i v_i^T$ is a rank-1 matrix.)

1.3.1 Low Rank Approximation

We know that

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_r u_r v_r^T.$$

We also know that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$. So, we can choose some $k \leq r$ and define

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Then, A_k is called the rank-k approximation of A (it's of type "low-rank" when $k < r$).

Essentially, we cut-off parts of the sum belonging to small singular values, producing an approximation (A_k) to the original matrix (A). It should, then, be noted that $\text{rank}(A_k) = k$, with each $u_i v_i^T$ having rank 1.

Theorem 1.4

$$||A - A_k||_2 = \sigma_{k+1}.$$

Proof. We know that $A = \hat{U}\hat{\Sigma}\hat{V}^T$ and $A_k = \hat{U}\hat{\Sigma}_k\hat{V}^T$. So,

$$\hat{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \qquad \hat{\Sigma}_k = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & 0 \end{bmatrix}.$$

From this,

$$\begin{aligned} A - A_k &= \hat{U}(\hat{\Sigma} - \hat{\Sigma}_k)\hat{V}^T \\ &= \hat{U} \begin{bmatrix} 0 & & & & & & \\ & 0 & & & & & \\ & & \ddots & & & & \\ & & & 0 & & & \\ & & & & \sigma_{k+1} & & \\ & & & & & \ddots & \\ & & & & & & \sigma_r \end{bmatrix} \hat{V}^T, \end{aligned}$$

as desired. $\square$

In addition, A_k is the matrix of rank k that is closest to A (in the 2-norm). In other words, $\min ||A - B||_2$ with minimum over all matrix B of rank k is A_k.