SINGULAR VALUE DECOMPOSITION of A MATRIX & ITS BASIC PROPERTIES

Thm [SVD]. Let $A \in \mathbb{C}^{m\times n}$, $r = \text{rank}(A) \leq \min\{m, n\}$. There exists the following matrix decomposition of A:

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

where $U \in \mathbb{C}^{m\times m}$, $V \in \mathbb{C}^{n\times n}$ are UNITARY MATRICES (ORTHOGONAL IF $A \in \mathbb{R}^{m\times n}$), and Σ is a DIAGONAL MATRIX such that:

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix} \in \mathbb{R}^{m\times n},$$

where $\sigma_i \in \mathbb{R}$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, are called SINGULAR VALUES of A.

The COLUMNS of U and V

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{bmatrix} \text{ and } V = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

are called LEFT & RIGHT SINGULAR VECTORS.

This is the full SVD.

When $r < \min\{m, n\}$, the "economy-sized" or reduced SVD typically suffices:

$$A = U_1 \Sigma_1 V_1^*,$$

$U_1 \in \overset{\mathbb{C}}{(\mathbb{R})}^{m\times r}$, $\quad U_1 = U(:, 1{:}r) = [u_1\ u_2 \cdots u_r]$

$\Sigma_1 \in \mathbb{R}^{r\times r}$, $\quad \Sigma_1 = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$

$V_1 \in \mathbb{R}^{n\times r}$, $\quad V_1 = V(:, 1{:}r) = [v_1\ v_2 \cdots v_r]$

U_1 and V_1 are "orthonormal" matrices (the term "orthogonal" is reserved for SQUARE matrices).

__Remark__: It is possible to compute the SVD with the help of two SPECTRAL decompositions:

$$A^*A = V\Sigma^*\underbrace{U^*U}_{I}\Sigma V^* = V\Sigma^*\Sigma V^* = V\Sigma^2 V^*$$

$\rightarrow$ we get V

$$AA^* = U\Sigma \underbrace{V^*V}_{I}\Sigma^* U^* = U\Sigma^2 U^*$$

$\rightarrow$ we get U

For numerical reasons, it's better to compute the SVD using a spectral decomposition of the Hermitian matrix

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}$$ "augmented matrix"

In fact, there exist algorithms more efficient and more numerically stable than this, e.g., the Golub–Kahan algorithm (see §8.6 of "Matrix Computations" of Golub & Van Loan (1996)).

__SOME BASIC PROPERTIES of THE SVD__:

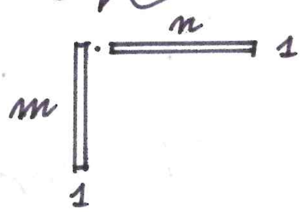
(a) $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ AND $\mathbf{u}_i^* A = \sigma_i \mathbf{v}_i^*$

(b) Spectral norm: $\|A\|_2 = \sigma_1$

(c) Frobenius norm: $\|A\|_F = \sqrt{\sum_i \sigma_i^2}$

(d) "Rank-revealing" SVD: the matrix A is a LINEAR COMBINATION of r rank-1 matrices, i.e.,

$$A = \sum_{i=1}^{r} \sigma_i \underbrace{\mathbf{u}_i \mathbf{v}_i^*}$$



(e) $\sigma_i = \sqrt{\lambda_i(AA^*)} = \sqrt{\lambda_i(A^*A)}$ (already seen above)

(h) $\operatorname{im}(A) = \operatorname{span}\{u_1, \dots, u_r\} = \operatorname{span}(U_1)$

$\operatorname{Ker}(A) = \operatorname{span}\{v_{r+1}, \dots, v_n\} = \operatorname{span}(V_1^{\perp})$

(g) The singular values of a matrix are unique.

If $\sigma_1 > \dots > \sigma_r > 0$ (STRICTLY), its singular vectors ($u_1, \dots, u_r$ AND $v_1, \dots, v_r$) are unique UP TO A FACTOR ± 1.

If SOME σ_i are EQUAL, then the SVD is no longer unique.

<u>Proofs:</u>

(a) $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ and $u_i^* A = \sigma_i v_i^*$ for $1 \leq i \leq \min\{m, n\}$.

Proof: Comparing the columns of $AV = U\Sigma$ and the rows of $U^*A = \Sigma V^*$.

(b) $$\|A\|_2 = \sqrt{\lambda_{max}(A^*A)} \underset{A = U\Sigma V^*}{=} \sqrt{\lambda_{max}(\underbrace{\Sigma^2}_{\text{DIAGONAL}})} = \sigma_1$$

(c) $$\|A\|_F := \sqrt{\sum_i \sum_j a_{ij}^2} = \sqrt{Tr(A^*A)} = \sqrt{Tr(\Sigma^*\Sigma)} = \|\Sigma\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

(h) Let $x \in \mathbb{C}^n$ (any vector $\in \mathbb{C}^n$). Since V is UNITARY, we have $x = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$, and then, because of (a),

$\mathbf{v}_i$ ← THE R.S.V.'s of A

$$Ax = \sum_{i=1}^{n} \alpha_i A\mathbf{v}_i = \sum_{i=1}^{r} (\alpha_i \sigma_i) \mathbf{u}_i \in \operatorname{span}\{u_1, \dots, u_r\}.$$

So $\operatorname{im}(A) = \{Ax : x \in \mathbb{C}^n\} = \operatorname{span}\{u_1, \dots, u_r\}$.

($\dim(\operatorname{im}(A)) = r$ by def. of rank).

From (a), we have that $A\boldsymbol{v}_i = 0$ if $i > r$.

The "nullity + rank" theorem tells us that

$$\dim \operatorname{Ker}(A) = n - r,$$

therefore $\operatorname{Ker}(A) = \operatorname{span}\{v_{r+1}, \dots, v_n\}$.

$\blacksquare$

A very important result (in numerical linear algebra): the <u>BEST</u> (possibly LOW) rank-k approximation of a matrix A <u>exists</u> and it can be computed via the SVD!

<u>Thm.</u> (<u>Eckart-Young</u>, <u>Schmidt-Mirsky theorem</u>)

Let $k < r = \operatorname{rank}(A)$ (i.e., k is the "low rank"), and let the TRUNCATED rank-revealing SVD be:

$$A_k := \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^*.$$

Then, the solution to the (non-convex) optimization problems

$$\min_{\operatorname{rank}(B)=k} \|A - B\|_{2,F}$$

is PRECISELY given by $B = A_k$. (!!!)

The <u>MINIMIZERS</u> are $\sigma_{k+1}(A)$ for the problem with $\|\cdot\|_2$,

and $\sqrt{\sum_{i=k+1}^{r} \sigma_i^2(A)}$ " " " " $\|\cdot\|_F$.

Moreover, the minimizer is UNIQUE if and only if $\sigma_k > \sigma_{k+1}$.

NB: STRICTLY!