

Section 6: Singular Value Decomposition

MATH241 Lecture Notes, notes pages Extra notes

Notation. Scalars are plain italic (a, α, z), vectors are bold ($\mathbf{u}, \mathbf{x}$), and matrices are italic capitals (A, B ; zero matrix O , identity I). The handwritten notes underline vectors once and matrices twice.

These notes are built on top of Orhan Arikan's linear algebra course notes.

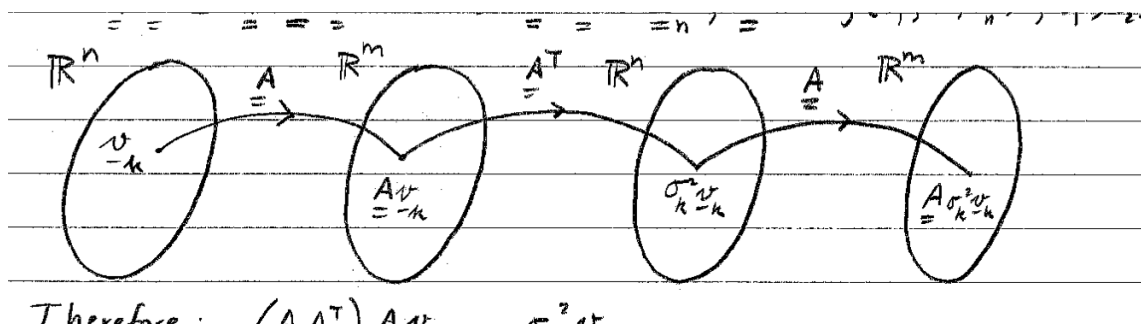
— extra note: SVD Lecture Notes, page 1 —

Supplement supplied separately; wording and calculations retained as written.

Singular Value Decomposition (SVD)

For a real $m \times n$ matrix A , the product $A^T A$ is symmetric and positive semi-definite, so it can be diagonalized by an orthogonal matrix:

$$A^T A = V \Lambda V^T \quad \text{where } V^T V = I_n, \quad \Lambda = \text{diag}(\sigma_1^2, \dots, \sigma_n^2), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$



Now look at what $A A^T$ does to the vector $A v_k$:

$$(A A^T) A v_k = A (A^T A v_k) = \sigma_k^2 A v_k.$$

Thus $A v_k$ is an eigenvector of the symmetric positive semi-definite matrix $A A^T$ with eigenvalue σ_k^2 , provided $A v_k \neq \mathbf{0}$. Write $A v_k = \alpha_k \mathbf{u}_k$ with $\|\mathbf{u}_k\| = 1$. Taking norms,

$$\|A v_k\|^2 = |\alpha_k|^2 \Rightarrow \mathbf{v}_k^T \underbrace{A^T A v_k}_{\sigma_k^2 v_k} = |\alpha_k|^2 \Rightarrow \alpha_k = \sigma_k \text{ is a solution.}$$

Remark: The product $A^T A v_k$ is replaced by $\sigma_k^2 v_k$, and then $\mathbf{v}_k^T v_k = 1$.

So $A v_k = \sigma_k \mathbf{u}_k$ for every k . Putting all n of these equations side by side,

$$A[\mathbf{v}_1 \ \dots \ \mathbf{v}_n] = [\mathbf{u}_1 \sigma_1 \ \mathbf{u}_2 \sigma_2 \ \dots \ \mathbf{u}_n \sigma_n] = [\mathbf{u}_1 \ \dots \ \mathbf{u}_n] \underbrace{\Sigma}_{\text{diag}(\sigma_1, \dots, \sigma_n)}$$

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

If $\sigma_k = 0$, then $A v_k = \mathbf{0}$ and the formula $\mathbf{u}_k = A v_k / \sigma_k$ does not define $\mathbf{u}_k$. Such columns can be completed with any orthonormal vectors, or simply dropped, as in the final form below.

For the non-zero singular values the vectors $\mathbf{u}_k$ are automatically orthonormal. Indeed, for $\sigma_k > 0$ and $\sigma_i > 0$,

$$\mathbf{u}_k^T \mathbf{u}_i = \frac{1}{\sigma_k} \mathbf{v}_k^T A^T A \frac{1}{\sigma_i} \mathbf{v}_i = \frac{1}{\sigma_k \sigma_i} \mathbf{v}_k^T \underbrace{(A^T A) \mathbf{v}_i}_{\sigma_i^2 \mathbf{v}_i}$$

$$= \frac{\sigma_i}{\sigma_k} \mathbf{v}_k^T \mathbf{v}_i = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i. \end{cases}$$

Final form. Keeping only the r non-zero singular values, the matrix is a sum of r rank-one pieces:

The SVD in its final form

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad \text{where } r \text{ is the number of non-zero singular values } \sigma_i.$$

$$A = U_r \Sigma_r V_r^T, \quad U_r = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r], \quad \Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r), \quad V_r = [\mathbf{v}_1, \dots, \mathbf{v}_r].$$

— extra note: SVD SUMMARY, page 1 —

Supplement supplied separately; wording and calculations retained as written.

SVD (Singular Value Decomposition)

Definition. A real $m \times n$ matrix A of rank r can be written in the “reduced” form

Reduced SVD

$$A_{m \times n} = U \Sigma V^T, \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \sigma_3 & \vdots \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix}_{r \times r}.$$

The numbers σ_i are the *singular values* of A , ordered as $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_r > 0$. The columns of U and V are orthonormal bases of the column space and the row space:

$$U_{m \times r} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r], \quad \text{where } \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\} \text{ is an ON basis for } \text{Col}(A),$$

$$V_{n \times r} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_r], \quad \text{where } \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\} \text{ is an ON basis for } \text{Row}(A).$$

Scaling Effect of Σ

The matrix of singular values acts as a pure scaling: it stretches or shrinks each coordinate separately. To see this, start from $A\mathbf{x} = \mathbf{y}$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, and multiply by U^T :

$$\begin{aligned} A_{m \times n} \underbrace{\mathbf{x}}_{\in \mathbb{R}^n} &= \underbrace{\mathbf{y}}_{\in \mathbb{R}^m} \xrightarrow{\times U^T} U^T A \mathbf{x} = U^T \mathbf{y} \\ &\Rightarrow \underbrace{U^T U}_I \Sigma V^T \mathbf{x} = U^T \mathbf{y} \\ &\Rightarrow \Sigma \underbrace{V^T \mathbf{x}}_{\tilde{\mathbf{x}}} = \underbrace{U^T \mathbf{y}}_{\tilde{\mathbf{y}}} \Rightarrow \Sigma \tilde{\mathbf{x}} = \tilde{\mathbf{y}}. \end{aligned}$$

In the rotated coordinates $\tilde{\mathbf{x}} = V^T \mathbf{x}$ and $\tilde{\mathbf{y}} = U^T \mathbf{y}$ the map is diagonal:

Scaling effect, componentwise

$$\sigma_i \tilde{x}_i = \tilde{y}_i.$$

So the directions $\mathbf{v}_i$ in the input space are sent to the directions $\mathbf{u}_i$ in the output space, and a singular value $\sigma_i > 1$ stretches along that direction while $\sigma_i < 1$ shrinks it.

The Matrix as a Sum of Rank-One Pieces

Multiplying out the three factors,

$$A = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r] \begin{bmatrix} \sigma_1 & \cdots & 0 \\ & \sigma_2 & \\ 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Each term $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is a rank-one matrix, and together they add up to a matrix of rank r . Keeping only the terms with the largest σ_i gives the best low-rank approximation of A .

— extra note: SVD SUMMARY, page 2 —

Supplement supplied separately; wording and calculations retained as written.

Derivation of the Singular Vectors

Let $W = A^T A$. Then $W \in \mathbb{R}^{n \times n}$ and $W = W^T$.

1) The vectors $\mathbf{v}_i$ (row space)

Start from $A = U \Sigma V^T$, which gives $A^T = V \Sigma^T U^T$. Multiplying,

$$\begin{aligned} A &= U \Sigma V^T \\ A^T &= V \Sigma^T U^T \end{aligned} \Rightarrow A^T A = V \underbrace{\Sigma^T}_{=\Sigma} \underbrace{U^T U}_I \Sigma V^T.$$

Remark: The middle product is I because the $\mathbf{u}_i$'s are orthonormal.

$$A^T A = V \Sigma^2 V^T.$$

Multiplying on the right by V and using $V^T V = I$ (the $\mathbf{v}_i$'s are orthonormal),

$$\Rightarrow A^T A V = V \Sigma^2 \underbrace{V^T V}_I = V \Sigma^2.$$

Reading this column by column,

$$(A^T A) \mathbf{v}_i = \mathbf{v}_i \lambda_i, \quad \lambda_i = \sigma_i^2.$$

So the eigenvalues of $A^T A$ are the squares of the singular values, and the columns $\mathbf{v}_i$ of V are the corresponding eigenvectors of $A^T A$. We order them by eigenvalue: if $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n$, then $\mathbf{v}_1$ is the eigenvector for λ_1 , and so on.

2) The vectors $\mathbf{u}_i$ (column space)

From $A = U \Sigma V^T$ we get $AV = U \Sigma$, and column by column $A \mathbf{v}_i = \mathbf{u}_i \sigma_i$. Hence

The left singular vectors

$$\mathbf{u}_i = \frac{A \mathbf{v}_i}{\sigma_i},$$

and these form an orthonormal basis for $\text{Col}(A)$. Each $\mathbf{u}_i$ must be listed in the same order as its partner $\mathbf{v}_i$, so that the pairs $(\mathbf{u}_i, \mathbf{v}_i)$ share the singular value σ_i .