

Supplemental notes on the singular value decomposition
Math 129b

The Singular Value Decomposition. Let A be a $k \times n$ matrix, and let $s = \min(k, n)$. There exist a $k \times k$ orthogonal matrix U , an $n \times n$ orthogonal matrix V , and real numbers $\sigma_1 \geq \cdots \geq \sigma_s \geq 0$ such that

$$U^t A V = \Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \sigma_s \end{bmatrix}, \quad (1)$$

where the (i, i) entry of Σ is σ_i and all other entries of Σ are 0. (Note that Σ is not necessarily a square matrix; in fact, Σ is $k \times n$.)

Since $U^t = U^{-1}$ and $V^{-1} = V^t$, we also have that

$$A = U \Sigma V^t. \quad (2)$$

This method of expressing A as a product of the form orthogonal-diagonal-orthogonal is called the *singular value decomposition* of A , and the real numbers $\sigma_1, \dots, \sigma_s$ are called the *singular values* of A . The columns of the matrix V are called the *right singular vectors* of A , and the columns of U are called the *left singular vectors* of A .

Note that since 1 is equivalent to $AV = U\Sigma$, for $1 \leq i \leq s \leq n$, we have that $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$. If $k \geq n$ (i.e., if A is “tall”), this accounts for all of the columns of AV ; if $k < n$ (i.e., if A is “fat”), then we also have that $A\mathbf{v}_i = \mathbf{0}$ for $k < i \leq n$.

The main point of the singular value decomposition of A is that the SVD gives a precise description of the geometry of the linear function $T : \mathbb{R}^n \rightarrow \mathbb{R}^k$ defined by $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, in that:

1. If $k \geq n$ (i.e., if A is “tall”), then $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis for \mathbb{R}^n and $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an orthonormal basis for \mathbb{R}^k such that each \mathbf{v}_i is mapped onto the scalar multiple $\sigma_i \mathbf{u}_i$ of \mathbf{u}_i .
2. If $k < n$ (i.e., if A is “fat”), then $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis for \mathbb{R}^n and $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an orthonormal basis for \mathbb{R}^k such that for $1 \leq i \leq k$, each \mathbf{v}_i is mapped onto the scalar multiple $\sigma_i \mathbf{u}_i$ of \mathbf{u}_i , and for $k < i \leq n$, each \mathbf{v}_i is mapped to $\mathbf{0} \in \mathbb{R}^k$.

It can also be shown that \mathbf{v}_1 , which corresponds to the largest singular value σ_1 , is a unit vector in \mathbb{R}^n that has an image of largest possible norm, i.e.,

$$\max_{\mathbf{v} \in \mathbb{R}^n, \|\mathbf{v}\|=1} \|T(\mathbf{v})\| = \|T(\mathbf{v}_1)\| = \|\sigma_1 \mathbf{u}_1\| = \sigma_1. \quad (3)$$

The SVD therefore provides the answer to many min/max problems arising from the geometry of T . For more on the applications of the SVD, including some applications to statistics, see *Matrix Computations*, by Golub and Van Loan.

Proof of SVD. Let $X = A^t A$; note that X is an $n \times n$ matrix. From PS11, there exists an orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for \mathbb{R}^n such that each \mathbf{v}_i is an eigenvector of X . Let λ_i be the eigenvalue of X associated with \mathbf{v}_i . From PS11, each $\lambda_i \geq 0$, so by reordering $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ if necessary, we may assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

For $1 \leq i \leq n$, let $\sigma_i = \sqrt{\lambda_i}$, which is a real number, since $\lambda_i \geq 0$. Let r be the largest integer such that $\lambda_r > 0$; i.e., pick r so that

$$\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n.$$

Note that for $1 \leq i \leq r$, $\sigma_i = \sqrt{\lambda_i} > 0$, so we may define

$$\mathbf{u}_i = \frac{1}{\sigma_i} A \mathbf{v}_i \quad \text{for } 1 \leq i \leq r. \quad (4)$$

By PS11, since $\lambda_i = 0$ for $r + 1 \leq i \leq n$, we have that

$$\mathbf{0} = A \mathbf{v}_i \quad \text{for } r + 1 \leq i \leq n. \quad (5)$$

PS11 also implies that $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ is orthonormal. The Orthonormal Expansion Theorem therefore implies that we may expand $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_k\}$ for \mathbb{R}^k .

Now let U be the $k \times k$ matrix whose columns are $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and let V be the $n \times n$ matrix whose columns are $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. Since $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ are orthonormal bases for \mathbb{R}^k and \mathbb{R}^n , respectively, U and V are orthogonal. It therefore remains only to verify (1).

First, note that

$$AV = A[\mathbf{v}_1 \ \cdots \ \mathbf{v}_n] = [A\mathbf{v}_1 \ \cdots \ A\mathbf{v}_n] = [\sigma_1 \mathbf{u}_1 \ \cdots \ \sigma_r \mathbf{u}_r \ \mathbf{0} \ \cdots \ \mathbf{0}], \quad (6)$$

where the last equality follows from (4) and (5). Therefore,

$$\begin{aligned} U^t AV &= \begin{bmatrix} \mathbf{u}_1^t \\ \vdots \\ \mathbf{u}_k^t \end{bmatrix} [\sigma_1 \mathbf{u}_1 \ \cdots \ \sigma_r \mathbf{u}_r \ \mathbf{0} \ \cdots \ \mathbf{0}] \\ &= \begin{bmatrix} \sigma_1 \mathbf{u}_1^t \mathbf{u}_1 & \cdots & \sigma_r \mathbf{u}_1^t \mathbf{u}_r & \mathbf{u}_1^t \mathbf{0} & \cdots & \mathbf{u}_1^t \mathbf{0} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sigma_1 \mathbf{u}_k^t \mathbf{u}_1 & \cdots & \sigma_r \mathbf{u}_k^t \mathbf{u}_r & \mathbf{u}_k^t \mathbf{0} & \cdots & \mathbf{u}_k^t \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}, \end{aligned} \quad (7)$$

where the last equality holds because $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is orthonormal. (Note that $U^t AV$ will not be a square matrix in general, even though we have drawn it as a square matrix to emphasize the diagonal entries.) Then, by setting $\sigma_{r+1} = \cdots = \sigma_s = 0$ if necessary, we obtain (1). The theorem follows. \square