Lecture 4

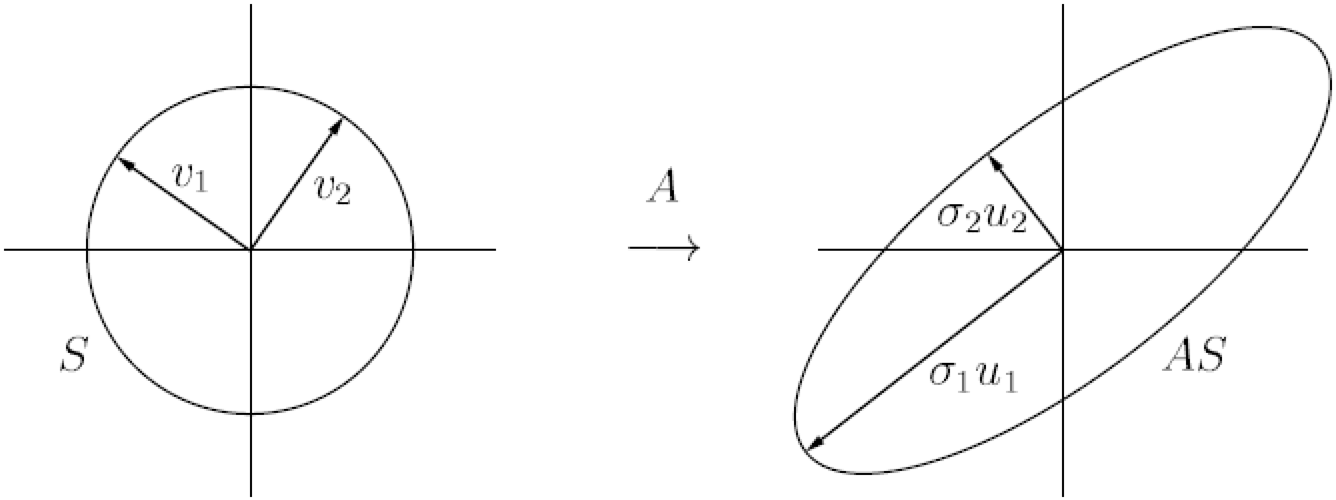
Singular Value Decomposition

NLA Reading Group Spring ’13

by Hakan Güldaş

Geometric Interpretation

The image of the unit sphere under any $m \times n$ matrix is a hyperellipse.



n singular values : the lengths of the n principal semiaxes of AS
n left singular vectors : the unit vectors oriented in the directions of the principal semiaxes
n right singular vectors : the unit vectors that are preimages of left singular vectors so that $Av_j = \sigma_j u_j$

Reduced Svd

$$Av_j = \sigma_j u_j, \qquad 1 \le j \le n$$

This collection of vector equations can be expressed as a matrix equation,

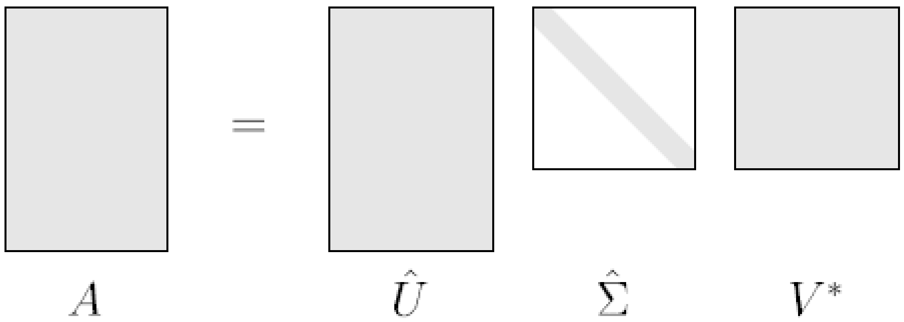
$$\left[\begin{array}{c} \\ \\ A \\ \\ \\ \end{array}\right]\left[\begin{array}{c|c|c|c} \\ v_1 & v_2 & \cdots & v_n \\ \\ \end{array}\right] = \left[\begin{array}{c|c|c|c} \\ \\ u_1 & u_2 & \cdots & u_n \\ \\ \\ \end{array}\right]\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix},$$

$\hat{U}$ is an $n\times m$ with orthonormal columns i.e. $\hat{U}\hat{U}^* = I_n$

V is an $n\times n$ unitary matrix

Σ is diagonal with nonnegative diagonal entries $\sigma_1, \ldots, \sigma_n$

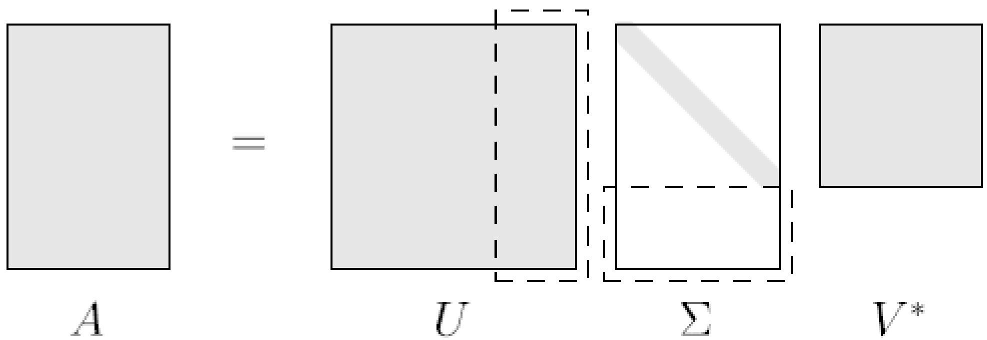
$$A = \hat{U}\hat{\Sigma}V^*$$



Full SVD

The idea is as follows. The columns of $\hat{U}$ are n orthonormal vectors in the m-dimensional space $\mathbb{C}^m$. Unless $m = n$, they do not form a basis of $\mathbb{C}^m$, nor is $\hat{U}$ a unitary matrix. However, by adjoining an additional $m - n$ orthonormal columns, $\hat{U}$ can be extended to a unitary matrix. Let us do this in an arbitrary fashion, and call the result U.

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



Now U and V are both unitary matrices and Σ is $m\times n$ diagonal matrix.
If A is rank deficient and of rank r, then exactly r diagonal entries of Σ are nonzero.

Formal Definition

singular value decomposition (SVD) of A is a factorization

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

where

$$\begin{aligned} U &\in \mathbb{C}^{m\times m} && \text{is unitary,} \\ V &\in \mathbb{C}^{n\times n} && \text{is unitary,} \\ \Sigma &\in \mathbb{R}^{m\times n} && \text{is diagonal.} \end{aligned}$$

In addition, it is assumed that the diagonal entries σ_j of Σ are nonnegative and in nonincreasing order; that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, where $p = \min(m, n)$.

unitary map V^* preserves the sphere

diagonal matrix Σ stretches the sphere into a hyperellipse

unitary map U rotates or reflects the hyperellipse

Existence and Uniqueness

Theorem 4.1. *Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition (4.4). Furthermore, the singular values $\{\sigma_j\}$ are uniquely determined, and, if A is square and the σ_j are distinct, the left and right singular vectors $\{u_j\}$ and $\{v_j\}$ are uniquely determined up to complex signs (i.e., complex scalar factors of absolute value 1).*

Set $\sigma_1 = \|A\|_2$. By a compactness argument, there must be vectors $v_1 \in \mathbb{C}^n$ and $u_1 \in \mathbb{C}^m$ with $\|v_1\|_2 = \|u_1\|_2 = 1$ and $Av_1 = \sigma_1 u_1$.

$$U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix}$$

$$\left\| \begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix} \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 \geq \sigma_1^2 + w^*w = (\sigma_1^2 + w^*w)^{1/2} \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2,$$

implying $\|S\|_2 \geq (\sigma_1^2 + w^*w)^{1/2}$. Since U_1 and V_1 are unitary, we know that $\|S\|_2 = \|A\|_2 = \sigma_1$, so this implies $w = 0$.

By the induction hypothesis, B has an SVD $B = U_2\Sigma_2V_2^*$

$$A = U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^* V_1^*$$

Existence and Uniqueness cont.

If the semiaxis lengths are distinct, then the the semiaxes themselves are determined by geometry, up to sign changes.

note that σ_1 is uniquely determined by the condition that it is equal to $\|A\|_2$

another linearly independent vector w with $\|w\|_2 = 1$ and $\|Aw\|_2 = \sigma_1$

$$v_2 = \frac{w - (v_1^* w)v_1}{\|w - (v_1^* w)v_1\|_2}$$

Since $\|A\|_2 = \sigma_1$, $\|Av_2\|_2 \leq \sigma_1$; but this must be an equality, for otherwise, since $w = v_1 c + v_2 s$ for some constants c and s with $|c|^2 + |s|^2 = 1$, we would have $\|Aw\|_2 < \sigma_1$. This vector v_2 is a second right singular vector of A corresponding to the singular value σ_1; it will lead to the appearance of a vector y (equal to the last $n-1$ components of $V_1^* v_2$) with $\|y\|_2 = 1$ and $\|By\|_2 = \sigma_1$

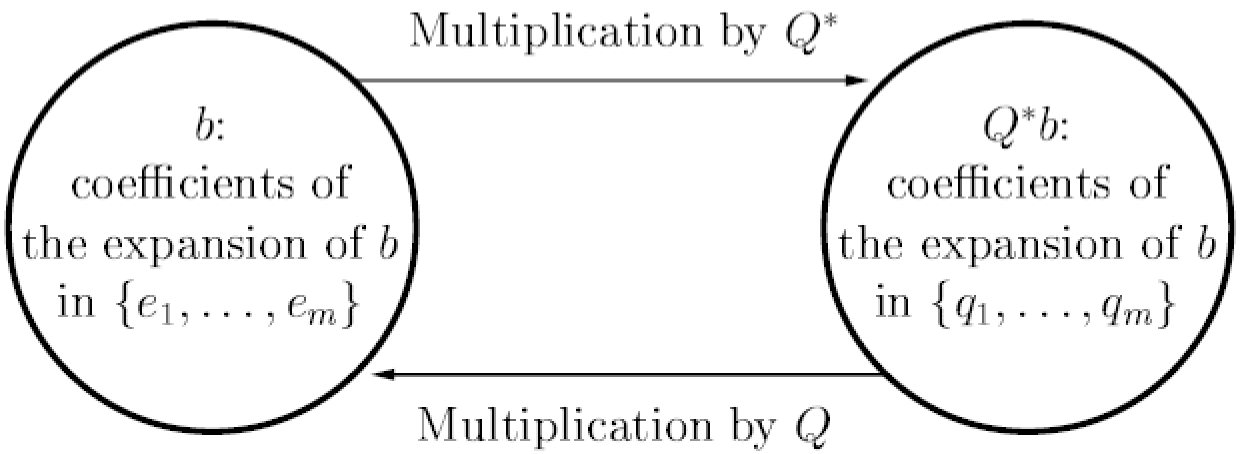
Lecture 5

More on the SVD

NLA Reading Group Spring ’13

by Hakan Güldaş

A Change of Basis



$$b \in \mathbb{C}^m \qquad x \in \mathbb{C}^n \qquad b' = U^*b, \qquad x' = V^*x$$

$$b = Ax \iff U^*b = U^*Ax = U^*U\Sigma V^*x \iff b' = \Sigma x'$$

Whenever $b = Ax$, we have $b' = \Sigma x'$. Thus A reduces to the diagonal matrix Σ when the range is expressed in the basis of columns of U and the domain is expressed in the basis of columns of V.

SVD vs. Eigenvalue Decomposition

$A \in \mathbb{C}^{m\times m}$, the *eigenvalue decomposition* of A is

$$A = X\Lambda X^{-1},$$

if we define, for $b, x \in \mathbb{C}^m$ satisfying $b = Ax$,

$$b' = X^{-1}b, \qquad x' = X^{-1}x,$$

- SVD uses two different bases, whereas EVD uses one.
- In SVD bases are orthonormal, in EVD bases are not necessarily orthonormal.
- Not all matrices have an EVD but every matrix has an SVD.

Matrix Properties via the SVD

The power of the SVD becomes apparent as we begin to catalogue its connections with other fundamental topics of linear algebra. For the following theorems, assume that A has dimensions $m \times n$. Let p be the minimum of m and n, let $r \leq p$ denote the number of nonzero singular values of A, and let $\langle x, y, \ldots, z \rangle$ denote the space spanned by the vectors $x, y, \ldots, z$.

Theorem 5.1. *The rank of A is r, the number of nonzero singular values.*

Theorem 5.2. $\mathrm{range}(A) = \langle u_1, \ldots, u_r \rangle$ *and* $\mathrm{null}(A) = \langle v_{r+1}, \ldots, v_n \rangle$

Theorem 5.3. $\|A\|_2 = \sigma_1$ *and* $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$.

Theorem 5.4. *The nonzero singular values of A are the square roots of the nonzero eigenvalues of A^*A or AA^*. (These matrices have the same nonzero eigenvalues.)*

Theorem 5.5. *If $A = A^*$, then the singular values of A are the absolute values of the eigenvalues of A.*

Theorem 5.6. *For $A \in \mathbb{C}^{m \times m}$,* $|\det(A)| = \prod_{i=1}^{m} \sigma_i$.

Low Rank Approximation

Theorem 5.7. *A is the sum of r rank-one matrices:*

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*. \tag{5.3}$$

Formula (5.3), however, represents a decomposition into rank-one matrices with a deeper property: *the νth partial sum captures as much of the energy of A as possible.*

Theorem 5.8. *For any ν with $0 \le \nu \le r$, define*

$$A_\nu = \sum_{j=1}^{\nu} \sigma_j u_j v_j^*; \tag{5.4}$$

if $\nu = p = \min\{m, n\}$, define $\sigma_{\nu+1} = 0$. Then

$$\|A - A_\nu\|_2 \;=\; \inf_{\substack{B \in \mathbb{C}^{m\times n} \\ \operatorname{rank}(B) \le \nu}} \|A - B\|_2 \;=\; \sigma_{\nu+1}.$$

Low Rank Approximation cont.

Theorem 5.8 has a geometric interpretation. What is the best approximation of a hyperellipsoid by a line segment? Take the line segment to be the longest axis. What is the best approximation by a two-dimensional ellipsoid? Take the ellipsoid spanned by the longest and the second-longest axis. Continuing in this fashion, at each step we improve the approximation by adding into our approximation the largest axis of the hyperellipsoid not yet included. After r steps, we have captured all of A. This idea has ramifications in areas as disparate as image compression (see Exercise 9.3) and functional analysis.

Theorem 5.9. *For any ν with $0 \le \nu \le r$, the matrix A_ν of (5.4) also satisfies*

$$\|A - A_\nu\|_F = \inf_{\substack{B \in \mathbb{C}^{m \times n} \\ \operatorname{rank}(B) \le \nu}} \|A - B\|_F = \sqrt{\sigma_{\nu+1}^2 + \cdots + \sigma_r^2}.$$