Defining the SVD

Let $A \in \mathbb{R}^{m \times m}$, and $A^T A = V \Lambda V^T$ be an eigenvalue decomposition, with V orthogonal. Then, AV satisfies

$$(AV)^T(AV) = V^T A^T AV = V^T(V\Lambda V^T)V = \Lambda.$$

This means that the columns of AV are orthogonal, but not orthonormal: the ith column has norm $\sqrt{\lambda_i}$.

We can scale them: define $(AV)_i = \mathbf{u}_i \sigma_i$, with $\sigma_i = \sqrt{\lambda_i}$. Then the $\mathbf{u}_i$ are the columns of an orthogonal matrix U.

Singular value decomposition

This gives a variant of the eigenvalue decomposition that is well-defined for every matrix:

Singular value decomposition (SVD) (for square matrices)

Each matrix $A \in \mathbb{R}^{m \times m}$ can be decomposed as

$$A = USV^T = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_m^T \end{bmatrix}$$

$$= \mathbf{u}_1 \sigma_1 \mathbf{v}_1^T + \mathbf{u}_2 \sigma_2 \mathbf{v}_2^T + \cdots + \mathbf{u}_m \sigma_m \mathbf{v}_m^T.$$

with U, V orthogonal and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$.

In this decomposition U and V are not the inverse of each other!

We lose the ability to express matrix powers:
$A^2 = A \cdot A = USV^T USV^T \neq US^2 V^T$.

Singular value decomposition

The σ_i are called singular values and we can take them non-negative and ordered: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$.

Singular values $\neq$ eigenvalues. They are always positive and usually more 'spread apart' than the eigenvalues. (Matlab examples)

Uniqueness: singular values are unique; singular vectors $\mathbf{u}_i, \mathbf{v}_i$ are not — exactly like eigenvalues / eigenvectors.

Rectangular matrices

The same theorem holds also for a rectangular matrix, with some changes in the shape of the involved matrices.

Singular value decomposition (SVD)

Each matrix $A \in \mathbb{R}^{m \times n}$ can be decomposed as $A = USV^T$, with with U, V orthogonal and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$. $U \in \mathbb{R}^{m \times m}$, $S \in \mathbb{R}^{m \times n}$ (padded with zeros), $V \in \mathbb{R}^{n \times n}$, e.g.,

$$S = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{bmatrix}.$$

$$A = \mathbf{u}_1 \sigma_1 \mathbf{v}_1^T + \mathbf{u}_2 \sigma_2 \mathbf{v}_2^T + \cdots + \mathbf{u}_{\min(m,n)} \sigma_{\min(m,n)} \mathbf{v}_{\min(m,n)}^T,$$

Thin SVD

Note that the sum-of-rank-1 form uses only the first $\min(m, n)$ columns of U and V. This suggests a different, more compact form, the thin (or economy-sized) SVD.

For tall-thin matrices:

$$A = \begin{bmatrix} U_0 & U_c \end{bmatrix} \begin{bmatrix} S_0 \\ 0 \end{bmatrix} V^T = U_0 S_0 V^T.$$

$U_0 \in \mathbb{R}^{m \times n}, S_0 \in \mathbb{R}^{n \times n}$.

(Matlab examples, `[U, S, V] = svd(A, 0)`).

Computational costs

`[U, S, V] = svd(A, 0)` (thin) costs $O(mn^2)$ ops for $A \in \mathbb{R}^{m \times n}$ or $A \in \mathbb{R}^{n \times m}$ with $m \geq n$.

`[U, S, V] = svd(A)` (non-thin) is more expensive: it has to compute and return the large $m \times m$ factor.

Properties of the SVD: rank, image, kernel

Rank r = number of nonzero singular values:
$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0.$
We can omit row/columns after r in the product:

$$A = USV^T = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

$$= \hat{U}\hat{S}\hat{V}^T = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}$$

$$= \mathbf{u}_1\sigma_1\mathbf{v}_1^T + \mathbf{u}_2\sigma_2\mathbf{v}_2^T + \cdots + \mathbf{u}_r\sigma_r\mathbf{v}_r^T.$$

For each $\mathbf{x} \in \mathbb{R}^n$, $A\mathbf{x}$ is linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_r$ (image).
Any linear combination $\mathbf{y}$ of $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ satisfies $A\mathbf{y} = 0$ (kernel).

Exercises

(Some done in class)

1. If $A = USV^T$ is the SVD of a square invertible A, what is the SVD of A^{-1}?
2. If A is positive semidefinite, is its eigendecomposition $A = U\Lambda U^T$ also an SVD?
3. If A is symmetric but not positive semidefinite, how can we modify signs in $A = U\Lambda U^T$ to obtain an SVD?
4. Show that for a square $A = USV^T$ one has $AA^T = US^2U^T$ and $A^TA = VS^2V^T$, and that these are eigendecompositions.
5. How do the decompositions in the previous exercise change if A is rectangular? Check also with Matlab.

References: Trefethen-Bau book, Lectures 4 and 5.