

EE16B - Fall'16 - Lecture 8B Notes¹

Murat Arcak

20 October 2016

¹ Licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).

Singular Value Decomposition (SVD)

SVD separates a rank- r matrix $A \in \mathbb{R}^{m \times n}$ into a sum of r rank-1 matrices of the form $\vec{u}\vec{v}^T$ (column times row). Specifically, we can find:

- 1) orthonormal vectors $\vec{u}_1, \dots, \vec{u}_r \in \mathbb{R}^m$,
- 2) orthonormal vectors $\vec{v}_1, \dots, \vec{v}_r \in \mathbb{R}^n$,
- 3) real, positive numbers $\sigma_1, \dots, \sigma_r$ such that

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \dots + \sigma_r \vec{u}_r \vec{v}_r^T. \quad (1)$$

The numbers $\sigma_1, \dots, \sigma_r$ are called *singular values* and, by convention, we order them from the largest to smallest:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0.$$

In its original form A has mn entries to be stored. In the SVD form each of the r terms is the product of a column of m entries with a row of n entries; therefore we need $r(m+n)$ numbers to store. This is an advantage when r is small relative to m and n , that is $r(m+n) \ll mn$.

In a typical application the exact rank r may not be particularly small, but we may find that the first few singular values, say $\sigma_1, \dots, \sigma_{\hat{r}}$, are much bigger than the rest, $\sigma_{\hat{r}+1}, \dots, \sigma_r$. Then it is reasonable to discard the small singular values and approximate A as

$$A \approx \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \dots + \sigma_{\hat{r}} \vec{u}_{\hat{r}} \vec{v}_{\hat{r}}^T \quad (2)$$

which has rank $= \hat{r}$, thus $\hat{r}(m+n) \ll mn$ numbers to store.

Example (Netflix): Suppose we have a $m \times n$ matrix that contains the ratings of m viewers for n movies. A truncated SVD as suggested above not only saves memory; it also gives insight into the preferences of each viewer. For example we can interpret each rank-1 matrix $\sigma_i \vec{u}_i \vec{v}_i^T$ to be due to a particular attribute, e.g., comedy, action, sci-fi, or romance content. Then σ_i determines how strongly the ratings depend on the i th attribute, the entries of \vec{v}_i^T score each movie with respect to this attribute, and the entries of \vec{u}_i evaluate how much each viewer cares about this particular attribute. Then truncating the SVD as in (2) amounts to identifying a few key attributes that underlie the ratings. This is useful, for example, in making movie recommendations as you will see in a homework problem.