Alon Rosen

Singular Values and Low Rank

Eigenvalues describe symmetric matrices. Singular values extend the same geometric picture to arbitrary rectangular matrices. We will use them to understand when linear equations are stably invertible and when a large matrix is well approximated by one of low rank.

The main points of this chapter are:

- The SVD decomposes a matrix into orthogonal input and output directions.
- The smallest singular value controls stability of least squares.
- Truncating the SVD gives the best low-rank approximation.

1 Singular values and the SVD

Let $A \in \mathbb{R}^{m \times n}$. Its *singular value decomposition* is

$$A = U\Sigma V^\top = \sum_{i=1}^{q} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top, \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_q > 0,$$

where $q = \operatorname{rank}(A)$, the vectors $\mathbf{u}_i$ are orthonormal in $\mathbb{R}^m$, and the vectors $\mathbf{v}_i$ are orthonormal in $\mathbb{R}^n$.

Fact 1. *For every i, $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ and $A^\top \mathbf{u}_i = \sigma_i \mathbf{v}_i$. Consequently,*

$$A^\top A\, \mathbf{v}_i = \sigma_i^2 \mathbf{v}_i, \qquad AA^\top \mathbf{u}_i = \sigma_i^2 \mathbf{u}_i.$$

Thus the singular values of A are the square roots of the nonzero eigenvalues of $A^\top A$.

The SVD therefore says that A takes orthogonal input directions $\mathbf{v}_i$ to orthogonal output directions $\mathbf{u}_i$, stretching the ith direction by the factor σ_i.

2 Largest and smallest singular values

For any unit vector $\mathbf{x}$, write $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i + \mathbf{x}_0$, where $\mathbf{x}_0 \in \ker A$. Then

$$\|A\mathbf{x}\|_2^2 = \sum_i \sigma_i^2 \alpha_i^2.$$

It follows immediately that

Fact 2.

$$\sigma_{\max}(A) = \max_{\|\mathbf{x}\|_2 = 1} \|A\mathbf{x}\|_2.$$

If A has full column rank, then

$$\sigma_{\min}(A) = \min_{\|\mathbf{x}\|_2 = 1} \|A\mathbf{x}\|_2.$$

Thus $\sigma_{\min}(A)$ measures the direction that A compresses the most. If it is very small, two noticeably different parameter vectors can produce very similar outputs.

3 Pseudoinverse and least squares

Assume A has full column rank. From the SVD $A = U\Sigma V^\top$, define

$$A^\dagger = V\Sigma^{-1}U^\top.$$

This agrees with $A^\dagger = (A^\top A)^{-1}A^\top$, and $A^\dagger A = I$.

Fact 3. *If A has full column rank, then*

$$\|A^\dagger\|_2 = \frac{1}{\sigma_{\min}(A)}.$$

Indeed, $A^\dagger \mathbf{u}_i = \sigma_i^{-1}\mathbf{v}_i$, so the largest stretching factor of $A^\dagger$ is $1/\sigma_{\min}(A)$. This is the quantity that appears in least-squares stability: small measurement error is amplified by at most this factor.

4 Low-rank approximation

A matrix has rank at most r if its columns lie in an r-dimensional subspace. The SVD orders the directions of a matrix from strongest to weakest, suggesting that we keep only the first r terms:

$$A_r = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top.$$

Clearly A_r has rank at most r.

For a matrix M, the *Frobenius norm* is

$$\|M\|_F^2 = \sum_{i,j} M_{ij}^2.$$

Equivalently, it is the Euclidean norm of all the entries of M viewed as one long vector. In particular, $\|A - B\|_F^2$ is the total squared entrywise error between A and B.

Fact 4 (Best rank-r approximation)**.** *Among all matrices B of rank at most r, the truncated SVD A_r minimizes $\|A - B\|_F$. Moreover,*

$$\|A - A_r\|_F^2 = \sum_{i>r} \sigma_i^2.$$

Thus the tail of the singular values measures exactly how far A is from rank r in total squared entrywise error. If the tail is small, a large matrix can be represented accurately using only a small number of singular directions.