# TOML2020-21 Low Rank Approximations

Jorge Garcia Vidal

May 2021

## 1 Introduction

In this section we want to investigate the best low-rank approximation for an arbitrary matrix $A$. We need first to define a measure of distance between matrices.

## 2 Induced matrix norms

One way of defining a matrix norm, is using vector norms. From this we can define:

- $||A||_2 = max\frac{||Ax||_2}{||x||_2} = max_{||x||_2=1}||Ax||_2$.
- $||A||_1 = max\frac{||Ax||_1}{||x||_1} = max_{||x||_1=1}||Ax||_1$.
- $||A||_\infty = max\frac{||Ax||_\infty}{||x||_\infty} = max_{||x||_\infty=1}||Ax||_\infty$.

## 3 Schatten norms: Matrix norms that can be expressed in terms of singular values

Other possible matrix norms are the Schatten norms, which are defined in terms of the singular values, and which in some cases can be expressed in terms of the trace operator:

### 3.1 Spectral norm

$$||A||_2 = \sigma_1$$

. This norm coincides with the induced $p = 2$ norm:

$$||A||_2 = max\frac{||Ax||_2}{||x||_2}.$$

### 3.2 Frobenius norm

$$||A||_F = \sqrt{\sum_k \sigma_k^2} = \sqrt{trace(A^tA)} = \sqrt{trace(AA^t)} = \sqrt{\sum_k \sum_i |a_{i,k}|^2}.$$

### 3.3 Nuclear norm

For the semidefinite matrix $A^tA$ we can define an square root as a matrix $B$ for which $B^2 = A^tA$. We define the nuclear (or Ky-Fan) norm as:

$$||A||_N = \sum_k \sigma_k = trace(\sqrt{A^tA}).$$

## 4 Eckart-Young approximation

Let $A$ be a matrix with $n$ columns and $m$ rows with $m \geq n$.

Suppose that $A = U\Sigma V^t$ is the SVD, with $U$ and $V$ are orthonormal matrices, and $\Sigma$ is an $m \times n$ diagonal matrix with entries $(\sigma_1, \sigma_2, \cdots, \sigma_n)$ such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

The Eckart-Young results says that the best rank $k$ approximation to $A$ in the spectral norm, is given by:

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^t.$$

This result can be extended to the Frobenius and the Nuclear norms.

We can prove it for the special cases $n = m = 2$ and $k = 1$. For the general case, the proof follows the same reasoning.

Let $A = \sigma_1 u_1 v_1^t + \sigma_2 u_1 v_1^t$, and $A_1 = \sigma_1 u_1 v_1^t$. We can easily see that $||A - A_1|| = \sigma_2$.

Let $B$ an arbitrary rank-1 $2 \times 2$ general matrix. We can thus express $B$ as: $B = \rho_1 x_1 y_1^t$. Let $w$ be an element of $Ker(B)$ of length 1, which in our case would be an orthonormal vector to $y$. We can express $w$ in terms of $v_1$ and $v_2$ as $w = \gamma_1 v_1 + \gamma_2 v_2$, with $\gamma_1^2 + \gamma_2^2 = 1$, with $Bw = 0$.

Using the definition of the spectral norm (which coincides with the induced norm with $p = 2$) we have:

$$||A - B||_2^2 \geq ||(A - B)w||_2^2 = ||Aw||_2^2 = \sigma_1^2\gamma_1^2 + \sigma_1^2\gamma_2^2 \geq \sigma_2^2.$$

meaning: $||A - B||_2^2 \geq ||A - A_1||_2^2$.