

# 1 28th of September 2018 — F. Poloni

We are left with the task of reaching a (sort of) “eigenvalue decomposition” when the target matrix is not symmetric.

**Definition 1.1** (Schur decomposition). *Let  $A \in M(n, \mathbb{R})$ ,  $\exists U \in M(n, \mathbb{R})$  orthogonal matrix and  $T \in M(n, \mathbb{R})$  triangular matrix such that  $A = UTU^T$  and this is called **Schur decomposition**.*

**Definition 1.2** (Singular value decomposition). *Let  $A \in M(n, \mathbb{R})$ ,  $\exists U, V \in M(n, \mathbb{R})$  orthogonal matrices and  $\Sigma \in \text{Diag}(n, \mathbb{R})$  such that  $A = U\Sigma V^T$  and this is called **Singular Value Decomposition**.*

$$A = \begin{pmatrix} U^1 & U^2 & \dots & U^n \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \cdot \begin{pmatrix} V^{1T} \\ V^{2T} \\ \vdots \\ V^{nT} \end{pmatrix} = \sum_{i=1}^n U^i \sigma_i V^{iT}$$

Where  $\sigma_i$  are called **singular values** and they are sorted such that  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$ .



## Something on Matlab ...

In Matlab the SVD decomposition is obtained through the command `svd(A)`, which return value is made of the three matrices  $U, \Sigma, V$ .

As an example, `[U, S, V] = svd(A)`. Notice that, if `svd(A)` is assigned to one variable, then such variable is an array of singular values.

**Definition 1.3** (Thin (fat) SVD). *Let  $A \in M(m, n, \mathbb{R})$ , there exist  $U \in M(m, n, \mathbb{R})$ ,  $V \in M(n, \mathbb{R})$  and  $\Sigma(m, n, \mathbb{R})$  such that  $A = \sum_{i=1}^{\min(m,n)} U^i \sigma_i V^{iT}$  has a **thin** (or **fat**) **SVD factorization**, where  $\Sigma$  has the following shape (case  $m = 3, n = 5$ )*

$$\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$$

We are not going into details of algorithms for computing SVD, but we would like to add a consideration about the computational complexity of such an algorithm. For thin SVD, it is  $O(m \cdot n^2)$ .

**Definition 1.4** (Rank). *Let  $A \in M(n, \mathbb{R})$  we call the **rank** of  $A$  the number of non-zero singular values.*

*Equivalently, the **rank** is the size of the column space.*

**Property 1.1.** A matrix  $A \in M(n, \mathbb{R})$  has rank  $r$  iff all its eigenvalues starting from the  $r + 1$ -th are 0, formally iff  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > \sigma_{k+1} = \dots = \sigma_n = 0$ .

Thanks to Property 1.1, we can somehow talk about an “even thinner” SVD, where all the 0s in the bottom right part of the matrix  $\Sigma$ , cancel out the latter columns of  $U$  and the latter rows of  $V$  (aka columns of  $V^T$ ). A pictorial representation of the shape of  $\Sigma$  can be found below.

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ \hline & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \\ \hline & & & & & & & \mathbf{0} \end{pmatrix}$$

This factorization represents  $A$  as  $\sum_{i=1}^r U_i \sigma_i V_i$ .

An attentive reader may notice that  $Ax = \sum_{i=1}^r U_i \sigma_i V_i x$ , where the last three terms are dimensionally a scalar. It goes without saying that the image of  $A$  is the span of  $U_1, U_2, \dots, U_r$ , hence  $rk(A) = r$ .

Moreover,  $\ker(A) = \text{span}(V_{r+1}, V_{r+2}, \dots, V_n)$ , since  $V$  is orthogonal (proof: plugging in  $x = V_j$ , where  $j > r$ ).

**Definition 1.5** (Matrix norm). Let  $A \in M(m, n, \mathbb{R})$ . We define the **matrix norm** of  $A$  as

$$\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{z=1} \|Az\|$$

Where the norm may be any of the ones defined in ?? second equality is introduced in order to work in a compact set, the one of normalized vectors  $z$ .

Notice that  $\|Ax\| \leq \|A\| \cdot \|x\|$ .

**Property 1.2.** Let  $A$  and  $B \in M(n, m, \mathbb{R})$  and let  $x \in \mathbb{R}^n$ , the following holds, for any norm defined in ??:

- $\|A\| \geq 0$  (and the equality holds iff  $A = 0$ );
- $\|\alpha A\| = |\alpha| \|A\|, \forall \alpha \in \mathbb{R}$ ;
- $\|A + B\| \leq \|A\| + \|B\|$ ;
- $\|AB\| \leq \|A\| \|B\|$ ;

- $\|Ax\| \leq \|A\| \|x\|$ .

**Fact 1.3.** Let  $A \in (n, m, \mathbb{R})$  and let  $U \in M(m, n, \mathbb{R})$  orthogonal, in the case of 2-norm  $\|A\|_2 = \|AU\|_2 = \|UA\|_2$ .

*Proof.*  $\|UA\|_2 = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|UAx\|_2}{\|x\|_2} \stackrel{(1)}{=} \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2$ , where  $\stackrel{(1)}{=}$  follows from a property of vector norms.

$\|AU\|_2 = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|AUx\|_2}{\|x\|_2} \stackrel{(2)}{=} \max_{y \in \mathbb{R}^n, y \neq 0} \frac{\|Ay\|_2}{\|y\|_2} = \|A\|_2$ , where  $\stackrel{(2)}{=}$  follows from the substitution  $y = Ux$ .  $\square$

**Definition 1.6** (Frobenius norm). Let  $A \in M(n, m, \mathbb{R})$ , we term **Frobenius norm** of  $A$

$$\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m (A)_{ij}^2}.$$

Notice that all the properties enumerated in Property 1.2 hold for the Frobenius norm as well.

**Fact 1.4.** Let  $A \in M(n, m, \mathbb{R})$  and let  $A = U\Sigma V^T$  be its singular value decomposition. The following hold:

1.  $\|A\|_2 = \|\Sigma\|_2 = \sigma_1$
2.  $\|A\|_F = \|\Sigma\|_F = \sqrt{\sum_{i=1}^{\min\{n,m\}} \sigma_i^2}$

*Proof.* 1. The first equality follows from Proposition 1.3, while the second is proved as follows:

$$\begin{aligned} \|\Sigma\|_2 &= \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|\Sigma x\|_2}{\|x\|_2} = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\left\| \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & & & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_2}{\left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_2} = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\left\| \begin{pmatrix} \sigma_1 x_1 \\ \sigma_2 x_2 \\ \vdots \\ \sigma_n x_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|_2}{\left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_2} \\ &= \frac{\sqrt{(\sigma_1 x_1)^2 + (\sigma_2 x_2)^2 + \cdots + (\sigma_n x_n)^2}}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}} \leq \frac{\sqrt{(\sigma_1 x_1)^2 + (\sigma_1 x_2)^2 + \cdots + (\sigma_1 x_n)^2}}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}} \\ &= \sqrt{\sigma_1^2} \cdot \frac{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}} = \sigma_1 \end{aligned} \tag{1.1}$$

The equality is achieved if we pick  $x = e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ .

2. The proof of this assertion is similar to the other and it is left to the reader. □

**Theorem 1.5** (Eckart-Young). *Let  $A \in M(n, m, \mathbb{R})$  and let  $A = U\Sigma V^T$  be its singular value decomposition.*

*The solution of  $\min_{rk(X) \leq k} \|A\| - X$  is given by the truncated SVD:*

$$X = \begin{pmatrix} U^1 & U^2 & \dots & U^k \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_k \end{pmatrix} \cdot \begin{pmatrix} V^1 \\ V^2 \\ \vdots \\ V^k \end{pmatrix}$$

Where the norm is both  $\|\cdot\|_2$  and  $\|\cdot\|_F$ .

**Fact 1.6.** *Let  $A \in M(n, \mathbb{R})$  and let  $A$  be invertible. The following holds:  $\|A^{-1}\| = \frac{1}{\sigma_n}$*

*Proof.* Since  $A$  is invertible, none of the  $\sigma_i$  is 0, hence the smaller (namely  $\sigma_n$ ) is not 0.

$$A^{-1} = (U\Sigma V^T)^{-1} \stackrel{(1)}{=} V^{T-1}\Sigma^{-1}U^{-1} = V \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix} U^T$$

Where  $\stackrel{(1)}{=}$  follows from the orthogonality of  $V$  and  $U$ .

Notice that this is *almost* an SVD, because the values on the diagonal are not sorted in a decreasing order.

Plugging in the norm, we have:

$$\|A^{-1}\| = \left\| V \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix} U^T \right\| = \left\| \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix} \right\| = \frac{1}{\sigma_n}$$

□