SINGULAR VALUE DECOMPOSITION

MANUEL W. LUETHI AND MEIKE AKVELD

Abstract. We motivate the proof of the singular value decompostion for a linear map between two finite dimensional $\mathbb{R}$ vector spaces and argue that the proof follows with elementary trickery from the right formulation of the question.

During his PhD studies at ETH, the first named author got into a discussion with an undergraduate student, who admitted that the technicality of the proof of the singular value decomposition blurred its meaning to him. Convinced, that this is by no means necessary, the first named author insisted on providing him with a geometric argument. From this the following proof was reconstructed, and afterwards incorporated in the course taught by the second named author.

We first motivate the statement. Assume that $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear of rank $r = \operatorname{rank} T$, i.e. r denotes the dimension of the image of T. Let $(w_1, \dots, w_r)$ be any ordered basis of $\operatorname{im} T$. By definition there are $v_1, \dots, v_r \in \mathbb{R}^n$ such that $Tv_i = w_i$ and thus $(v_1, \dots, v_r)$ is a linearly independent set. One easily checks that $\ker T \cap \langle v_1, \dots, v_r \rangle = \{0\}$, and hence we can find an extension $(v_1, \dots, v_n)$ to a basis of $\mathbb{R}^n$ such that the matrix representing T with respect to this basis and any extension of the w_i to a basis of $\mathbb{R}^m$ is of the form

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

where I_r denotes the $r \times r$ identity matrix. If we normalize the v_i with respect to the Euclidean metric on $\mathbb{R}^n$, then we would have to replace the identity matrix in the top left corner by a diagonal matrix, whose diagonal entries are the original length of the vectors v_i.

The singular value decomposition is a refined version of this statement. Its proof is a bit more intricate, as we can certainly start with an orthonormal basis of $\operatorname{im} T$ and we can of course extend it to an orthonormal basis of $\mathbb{R}^m$, but there is no reason why the v_i should be orthogonal. However the singular value decomposition states that there exists an orthonormal basis of $\mathbb{R}^m$ such that the corresponding v_i can be chosen orthonormally.

There are two main ingredients:

- We want to somehow single out a designated basis of $\operatorname{im} T$ for which the v_i could be orthogonal, but ex ante it is not clear, where this would come from. Let T^* denote the adjoint to T. Then TT^* is self-adjoint non-negative definite and hence $\mathbb{R}^m$ has an orthonormal basis consisting of eigenvectors of TT^*. As $\operatorname{im} TT^* = \operatorname{im} T$, there is a very special basis of $\operatorname{im} T$, namely the eigenvectors of TT^* for positive eigenvalues.
- The map T defines a bijection between $(\ker T)^\perp$ and $\operatorname{im} T$, hence given $w \in \operatorname{im} T$, there is a unique $v \in (\ker T)^\perp$ such that $Tv = w$. Note that $\operatorname{im} T^* \subseteq (\ker T)^\perp$, and as the restriction of TT^* is diagonalizable, so is its inverse. Hence one expects that for the unique $v_i \in (\ker T)^\perp$ satisfying $Tv_i = w_i$ for some eigenvector $w_i \in \operatorname{im} T$

Date: January 12, 2019.

of TT^*, one also has that T^*w_i is a multiple of v_i. It will however turn out that in the proof it is much easier to define the v_i using T^* directly.

We now set out to give a rigorous proof of the following

Theorem 1 (Singular Value Decomposition). *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ linear and $r = \operatorname{rank} T \geq 1$. Then there exist $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and orthonormal bases $(v_i)_{i=1}^n$ and $(w_j)_{j=1}^m$ of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, so that $Tv_i = \sigma_i w_i$ whenever $1 \leq i \leq r$ and $v_i \in \ker T$ otherwise. Moreover, the* singular values *$\sigma_1, \cdots, \sigma_r$ are uniquely determined by T.*

Proof. As TT^* is non-negative definite self-adjoint, there exists an orthonormal basis $(w_j)_{j=1}^m$ of $\mathbb{R}^m$ consisting of eigenvectors of TT^*. For all $1 \leq j \leq m$ let $\lambda_j \in \mathbb{R}$ be the eigenvalue of TT^* corresponding to the eigenvector w_j. After permutation of the elements of the basis, we can assume without loss of generality that $\lambda_1 \geq \cdots \geq \lambda_\rho > 0 = \lambda_{\rho+1} = \cdots = \lambda_m$, where $\rho = \operatorname{rank} TT^*$. In particular, $(w_j)_{j=1}^\rho$ is an orthonormal basis of $\operatorname{im} TT^*$.

We show now that $\operatorname{im} TT^* = \operatorname{im} T$. The inclusion $\operatorname{im} TT^* \subseteq \operatorname{im} T$ is immediate, and using the assumption of finite dimensionality, it suffices to show that $\operatorname{rank} TT^* = \operatorname{rank} T$. First we show that $\ker TT^* = \ker T^*$. Again, one inclusion is clear, as $T^*w = 0 \implies TT^*w = 0$. For the opposite inclusion, assume that $w \in \ker TT^*$. Then

$$0 = \langle TT^*w, w\rangle = \langle T^*w, T^*w\rangle \implies T^*w = 0$$

and hence follows $\ker TT^* \subseteq \ker T^*$. Hence the dimension formula implies that

$$\begin{aligned}\operatorname{rank} TT^* &= m - \dim(\ker TT^*) = m - \dim(\ker T^*) \\ &= \operatorname{rank} T^* = \operatorname{rank} T\end{aligned}$$

as desired. It follows, that $\rho = r$ and that $(w_j)_{j=1}^r$ is an orthonormal basis of $\operatorname{im} T$.

For $1 \leq i \leq r$ set $\tilde{v}_i = T^*w_i$, then for all $1 \leq i, j \leq r$ holds

$$\lambda_i \delta_{ij} = \langle TT^*w_i, w_j\rangle = \langle \tilde{v}_i, \tilde{v}_j\rangle$$

and thus $(\tilde{v}_j)_{j=1}^r$ form an orthogonal family of non-zero vectors. In particular, they are linearly independent. Let $v \in \ker T$, then we get for $1 \leq i \leq r$

$$0 = \langle Tv, w_i\rangle = \langle v, \tilde{v}_i\rangle$$

and thus $(\ker T) \perp \langle \tilde{v}_1, \ldots, \tilde{v}_r\rangle$, so that using $n = \dim(\ker T) + r$ there exists an extension $(\tilde{v}_i)_{i=1}^n$ to an orthogonal basis of $\mathbb{R}^n$. Let $v_i = \frac{1}{\|\tilde{v}_i\|}\tilde{v}_i$ for all $1 \leq i \leq n$, then it follows that

$$T\Big(\sum_{i=1}^n \alpha_i v_i\Big) = \sum_{i=1}^n \alpha_i T v_i = \sum_{i=1}^r \alpha_i \sqrt{\lambda_i} w_i$$

and setting $\sigma_i = \sqrt{\lambda_i}$ the existence follows.

For the uniqueness, we note that for any set of singular values and corresponding normal basis $(v_1, \ldots, v_n)$, we have

$$\langle T^*Tv_i, v_j\rangle = \sigma_i^2 \delta_{ij}$$

and thus v_i is an eigenvector of T^*T for eigenvalue σ_i^2. Hence by positivity and ordering, σ_i is uniquely determined by T. ☐

Email address: `manuel.luethi@math.ethz.ch`