A PERSONAL INTERVIEW WITH THE SINGULAR VALUE DECOMPOSITION

MATAN GAVISH

Part 1. Theory

1. The Polar Decomposition

In what follows, $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$. The vector space $\mathbb{F}^n$ is an inner product space with the standard inner product, $\langle\cdot,\cdot\rangle$. Let us denote by $\mathcal{M}_{n\times m}(\mathbb{F})$ the set of matrices over $\mathbb{F}$ with n rows and m columns. Vectors will denote columns, that is, we will write

$$\mathbb{F}^n \ni v = \begin{pmatrix} | \\ v \\ | \end{pmatrix}.$$

The following fact is known as the (Left) Polar Decomposition: Let $m \leqslant n$. Any matrix $A \in \mathcal{M}_{n\times m}(\mathbb{F})$ may be factorized as $A = U \cdot P$ where $U \in \mathcal{M}_{n\times m}(\mathbb{F})$ has a orthonormal columns and $P \in \mathcal{M}_{m\times m}(\mathbb{F})$ is positive semi-definite.

Remark. When $\mathbb{F} = \mathbb{R}$ and $n = m$, the polar decomposition makes precise the statement, that every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is a composition of a dilation and a rotation/reflection. The proof will imply that in this case there is an orthonormal base (the eigenvectors of the matrix $\sqrt{A^*A}$) on which A acts as a rotation/reflection (U) composed with a dilation (P).

The polar decomposition follows directly from a remarkable connection between the (general) matrix A and the positive semi-definite matrix $P = \sqrt{A^*A}$.

Date: June 30, 2010.

1.1. **The Matrix $\sqrt{A^*A}$.** We first note that the matrix $A^*A \in \mathcal{M}_{m\times m}(\mathbb{F})$ is self-adjoint and positive semi-definite, since for any $v \in \mathbb{F}^m$ we have

$$0 \leqslant \|Av\|^2 = \langle Av, Av\rangle = \langle v, A^*Av\rangle = v^*A^*Av\ .$$

There are several equivalent ways to define the matrix $\sqrt{A^*A}$. To name one, note that A^*A can be diagonalized over $\mathbb{F}$, $A^*A = WDW^*$, where W is unitary and D is diagonal. In fact, since A^*A is positive semi-definite, all its eigenvalues are real and non-negative, and we can denote

$$D = \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_m^2 \end{pmatrix},$$

for some numbers $0 \leqslant \lambda_1, \ldots, \lambda_m \in \mathbb{R}$, called the *Singular Values* of A. We define

$$\sqrt{A^*A} = W \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix} W^*\ .$$

This notation is justified since $\left(\sqrt{A^*A}\right)^2 = A^*A$. Note that $\sqrt{A^*A}$ is also self-adjoint and positive semi-definite.

Why is the matrix $\sqrt{A^*A}$ is interesting? one reason is that for all $v \in \mathbb{F}^m$,

$$\|Av\|^2 = \langle Av, Av\rangle = \langle v, A^*Av\rangle = \left\langle v, \sqrt{A^*A}\cdot\sqrt{A^*A}v\right\rangle = \left\langle \sqrt{A^*A}v, \sqrt{A^*A}v\right\rangle = \left\|\sqrt{A^*A}v\right\|^2,$$

that is,

$$\|Av\| = \left\|\sqrt{A^*A}v\right\|\ . \tag{1.1}$$

Besides the geometrical interpretation that A and $\sqrt{A^*A}$ have the same effect on a vector's length, this has the useful consequence,

$$\ker A = \ker\sqrt{A^*A}\,.$$

As $m = \dim\ker A + \operatorname{rank}A = \dim\ker\sqrt{A^*A} + \operatorname{rank}\sqrt{A^*A}$, it also follows that

$$\operatorname{rank}\sqrt{A^*A} = \operatorname{rank}A\,.$$

We denote their common value by ℓ.

1.2. **Proof of the (Right) Polar Decomposition.** Since $\sqrt{A^*A}$ is self-adjoint, there exists an orthonormal basis $\psi_1, \dots, \psi_m$ of $\mathbb{F}^m$ whose elements are eigenvectors of $\sqrt{A^*A}$. Using our previous notation for the eigenvalues of $\sqrt{A^*A}$, suppose that $\sqrt{A^*A}\,\psi_i = \lambda_i \psi_i$ for $i = 1, \dots, m$. For simplicity, let us assume that $\lambda_1 \dots \lambda_\ell > 0$ (where $m \geqslant \ell = \mathrm{rank} A$) or in other words, that $\psi_{\ell+1}, \dots, \psi_m$ are basis vectors for $\ker A = \ker \sqrt{A^*A}$ (in case A has a non-trivial kernel). To factor $A = U \cdot P$, let $P = \sqrt{A^*A}$ and define $U \in \mathcal{M}_{n\times m}(\mathbb{F})$ as follows.

Consider the vectors $\frac{1}{\lambda_1}A\psi_1, \dots, \frac{1}{\lambda_\ell}A\psi_\ell$. This set is orthonormal: for $1 \leqslant i, j \leqslant \ell$, we have

$$\left\langle \frac{1}{\lambda_i}A\psi_i\,,\ \frac{1}{\lambda_j}A\psi_j \right\rangle = \frac{1}{\lambda_i\lambda_j}\langle A\psi_i, A\psi_j\rangle = \frac{1}{\lambda_i\lambda_j}\langle \psi_i, A^*A\psi_j\rangle = \frac{1}{\lambda_i\lambda_j}\langle \psi_i, \lambda_j^2\psi_j\rangle = \frac{\lambda_j}{\lambda_i}\langle \psi_i, \psi_j\rangle = \delta_{i,j}$$

(recall that $\lambda_1, \dots \lambda_m$ are reals).

Remark. In fact, we discovered that $\psi_{\ell+1}, \dots, \psi_m$ is an orthonormal basis for $\ker A$, and $\frac{1}{\lambda_1}A\psi_1, \dots, \frac{1}{\lambda_\ell}A\psi_\ell$ is an orthonormal basis for $\mathrm{Im} A$.

We now use the assumption $m \leqslant n$. that If $\ell < m$, take an orthonormal completion $\varphi_{\ell+1}, \dots, \varphi_m$ such that $\frac{1}{\lambda_1}A\psi_1, \dots, \frac{1}{\lambda_\ell}A\psi_\ell, \varphi_{\ell+1}, \dots, \varphi_m$ constitutes an orthonormal set of vectors in $\mathbb{F}^n$. This is possible since $m \leqslant n$. Finally, define

$$U = \begin{pmatrix} | & & | & | & & | \\ \frac{1}{\lambda_1}A\psi_1 & \cdots & \frac{1}{\lambda_\ell}A\psi_\ell & \varphi_{\ell+1} & \cdots & \varphi_m \\ | & & | & | & & | \end{pmatrix} \cdot \begin{pmatrix} -\ \psi_1^*\ - \\ -\ \psi_2^*\ - \\ \vdots \\ -\ \psi_m^*\ - \end{pmatrix} \in \mathcal{M}_{n\times m}(\mathbb{F})\ .$$

Let us check that the matrix U has the desired properties. First, U has orthonormal columns: the left-hand matrix in the above product is unitary and hence has orthonormal columns, as does the right-hand matrix. It is simple to verify that generally, a product of matrices with orthonormal columns again has orthonormal columns. Next, Since

$$\begin{pmatrix} -\ \psi_1^*\ - \\ -\ \psi_2^*\ - \\ \vdots \\ -\ \psi_m^*\ - \end{pmatrix} \cdot \begin{pmatrix} | \\ \psi_i \\ | \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow\ i \\ \\ \\ \\ \end{matrix}\ ,$$

we find that

$$U\psi_i = \begin{cases} \frac{1}{\lambda_i} A\psi_i & 1 \leqslant i \leqslant \ell \\ \varphi_i & \ell < i \leqslant m \end{cases}$$

so that

$$UP\psi_i = \begin{cases} \lambda_i U\psi_i & 1 \leqslant i \leqslant \ell \\ 0 \cdot U\psi_i & \ell < i \leqslant m \end{cases} = \begin{cases} A\psi_i & 1 \leqslant i \leqslant \ell \\ 0 \cdot \varphi_i & \ell < i \leqslant m \end{cases} = A\psi_i$$

where we have used the fact that $A\psi_i = 0$ for $i = \ell + 1, \ldots, m$ (these vector actually span $\ker A$). Since the identity $A = UP$ holds on the basis vectors, $\psi_1, \ldots, \psi_m$, we are done.

Remark. If A is inverible, $\ker A = \ker \sqrt{A^*A}$ implies that $\sqrt{A^*A}$ is invertible. In this case, U is uniquely determined since $U = A \cdot P^{-1}$.

Remark. If $A \in \mathcal{M}_{n\times m}(\mathbb{F})$ and $m \geqslant n$, we can perform a right polar decomposition for $A^* \in \mathcal{M}_{m\times n}(\mathbb{F})$. We obtain a matrix $U \in \mathcal{M}_{m\times n}(\mathbb{F})$ with orthonormal columns such that $A^* = U \cdot \sqrt{AA^*}$, or equivalently, $A = \sqrt{AA^*} \cdot U^*$. Here $\sqrt{AA^*}$ is again positive semi-definite, but now U^* has orthonormal *rows*. This is the Left Polar Decomposition. If $n = m$, that is, if our matrix A is square, both decompositions are possible: there exist unitary matrices $U, U' \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that

$$A = U \cdot \sqrt{A^*A} = \sqrt{AA^*} \cdot U' .$$

It is interesting to note that t, in this case $U = U'$ holds if and only if A is normal ($AA^* = A^*A$).

2. The Singular Value Decomposition

2.1. **Roughly Speaking.** At least two different decompositions go by the name of Singular Value Decomposition (SVD):

(1) Any matrix $A \in \mathcal{M}_{n\times m}(\mathbb{F})$ may be factorized as

$$A = V \cdot D \cdot W^* ,$$

where $V \in \mathcal{M}_{n\times p}(\mathbb{F})$ and $W \in \mathcal{M}_{m\times p}(\mathbb{F})$ have orthonormal columns for $p = \min\{m, n\}$, and $D \in \mathcal{M}_{p\times p}(\mathbb{F})$ is diagonal with non-negative entries.

(2) Any matrix $A \in \mathcal{M}_{n\times m}(\mathbb{F})$ may be factorized as $A = V \cdot D \cdot W^*$, where $V \in \mathcal{M}_{n\times n}(\mathbb{F})$ and $W \in \mathcal{M}_{m\times m}(\mathbb{F})$ are unitary matrices, and $D \in \mathcal{M}_{n\times m}(\mathbb{F})$ has non-negative entries on the main diagonal and zeros elsewhere. (The main diagonal of a matrix $D \in \mathcal{M}_{n\times m}(\mathbb{F})$ is $(D_{11}, \dots, D_{nn})$.)

We will work out SVD using the first definition[1]. Suppose that such a decompisition does exist, and denote

$$V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_p \\ | & & | \end{pmatrix} \; ; \; W = \begin{pmatrix} | & & | \\ w_1 & \cdots & w_p \\ | & & | \end{pmatrix} \; ; \; D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{pmatrix}.$$

Since $W^*W = V^*V = \mathbb{I}_p$, for each $1 \leqslant i \leqslant p$, we have

$$A^*A\, w_i = WDV^*VDW^*\, w_i = WD^2W^*\, w_i = WD^2 \begin{pmatrix} - & w_1^* & - \\ & \vdots & \\ - & w_p^* & - \end{pmatrix} \begin{pmatrix} | \\ w_i \\ | \end{pmatrix} = d_i^2 w_i \tag{2.1}$$

and similarly

$$AA^*\, v_i = VDW^*WDV^*\, v_i = VD^2V^*\, v_i = VD^2 \begin{pmatrix} - & v_1^* & - \\ & \vdots & \\ - & v_p^* & - \end{pmatrix} \begin{pmatrix} | \\ v_i \\ | \end{pmatrix} = d_i^2 v_i\,. \tag{2.2}$$

Should such a decomposition exist, then, the column w_i of W must be an eigenvector of A^*A with eigenvalue d_i^2. Using notation from the previous section, if $m \leqslant n$ (so $p = m$) we must have simply

$$W = \begin{pmatrix} | & & | \\ \psi_1 & \cdots & \psi_m \\ | & & | \end{pmatrix} \; ; \; D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}.$$

[1]Matlab, for example, uses the second one

2.2. **SVD from the Polar Decomposition.** Formally, the SVD is an easy consequence of the polar decomposition. Assume that $m \leqslant n$ (so $p = m$) and $A \in \mathcal{M}_{n\times m}(\mathbb{F})$. We use the notations defined above, and employ the right polar decomposition:

$$\begin{aligned}
A &= U\cdot\sqrt{A^*A} = UW\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_m\end{pmatrix}W^* = \\
&= \begin{pmatrix} | & & | & | & & | \\ \frac{1}{\lambda_1}A\psi_1 & \cdots & \frac{1}{\lambda_\ell}A\psi_\ell & \varphi_{\ell+1} & \cdots & \varphi_m \\ | & & | & | & & | \end{pmatrix}\cdot WW\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_m\end{pmatrix}W^* = \\
&= \underbrace{\begin{pmatrix} | & & | & | & & | \\ \frac{1}{\lambda_1}A\psi_1 & \cdots & \frac{1}{\lambda_\ell}A\psi_\ell & \varphi_{\ell+1} & \cdots & \varphi_m \\ | & & | & | & & | \end{pmatrix}}_{V}\cdot\underbrace{\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_m\end{pmatrix}}_{D}\cdot\underbrace{\begin{pmatrix} - & \psi_1^* & - \\ & \vdots & \\ - & \psi_m^* & - \end{pmatrix}}_{W^*};
\end{aligned}$$

Define, then

$$\begin{aligned}
D &= \begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_m\end{pmatrix}\\
V &= \begin{pmatrix} | & & | & | & & | \\ \frac{1}{\lambda_1}A\psi_1 & \cdots & \frac{1}{\lambda_\ell}A\psi_\ell & \varphi_{\ell+1} & \cdots & \varphi_m \\ | & & | & | & & | \end{pmatrix}\\
W &= \begin{pmatrix} | & & | \\ \psi_1 & \cdots & \psi_m \\ | & & | \end{pmatrix}
\end{aligned}$$

and we have, by definition, an SVD decomposition of A. Manifestly, (2.1) holds. Let us convince ourselves that (2.2) also holds, namely, that the column v_i of V is an eigenvectors of AA^*, which corresponds to the eigenvalue λ_i^2. Indeed, for $1 \leqslant i \leqslant \ell$ we have

$$AA^*v_i = AA^*\frac{1}{\lambda_i}A\psi_i = \frac{1}{\lambda_i}A\left(A^*A\psi_i\right) = \frac{1}{\lambda_i}A\left(\lambda_i^2\psi_i\right) = \lambda_i^2\left(\frac{1}{\lambda_i}A\psi_i\right).$$

For the case where $\ell+1 \leqslant i \leqslant m$, recall that by our construction $\varphi_{\ell+1}, \ldots, \varphi_m \in (\mathrm{Im} A)^{\perp}$. We always have $(\mathrm{Im} A)^{\perp} \subset \ker AA^*$: indeed, fix $v \in (\mathrm{Im} A)^{\perp}$, then for any $u \in \mathbb{F}^n$,

$$0 = \langle AA^* u, v \rangle = \langle u, AA^* v \rangle$$

- which implies $AA^* v = 0$. Now, if $\ell+1 \leqslant i \leqslant m$ then $\lambda_i = 0$ and $v_i = \varphi_i \in \ker AA^*$, so that

$$AA^* v_i = 0 = 0 \cdot \varphi_i = \lambda_i^2 \cdot \varphi_i \,.$$

Thus, in either case, the column v_i of V is an eigenvectors of AA^*that corresponds to the eigenvalue λ_i^2.