Rend. Istit. Mat. Univ. Trieste
Volume 47 (2015), 107–125
DOI: 10.13137/0049-4704/11222

A geometric perspective on the Singular Value Decomposition

Giorgio Ottaviani and Raffaella Paoletti

Dedicated to Emilia Mezzetti

Abstract. *This is an introductory survey, from a geometric perspective, on the Singular Value Decomposition (SVD) for real matrices, focusing on the role of the Terracini Lemma. We extend this point of view to tensors, we define the singular space of a tensor as the space spanned by singular vector tuples and we study some of its basic properties.*

Keywords: Singular Value Decomposition, Tensor.
MS Classification 2010: 15A18, 15A69, 14N05.

1. Introduction

The *Singular Value Decomposition* (SVD) is a basic tool frequently used in Numerical Linear Algebra and in many applications, which generalizes the *Spectral Theorem* from symmetric $n \times n$ matrices to general $m \times n$ matrices. We introduce the reader to some of its beautiful properties, mainly related to the *Eckart-Young Theorem*, which has a geometric nature. The implementation of a SVD algorithm in the computer algebra software *Macaulay2* allows a friendly use in many algebro-geometric computations.

This is the content of the paper. In Section 2 we see how the best rank r approximation of a matrix can be described through its SVD; this is the celebrated Eckart-Young Theorem, that we revisit geometrically, thanks to the *Terracini Lemma*. In Section 3 we review the construction of the SVD of a matrix by means of the Spectral Theorem and we give a coordinate free version of SVD for linear maps between Euclidean vector spaces. In Section 4 we define the singular vector tuples of a tensor and we show how they are related to tensor rank; in the symmetric case, we get the eigentensors. In Section 5 we define the singular space of a tensor, which is the space containing its singular vector tuples and we conclude with a discussion of the *Euclidean Distance (ED) degree*, introduced first in [5]. We thank the referee for many useful remarks.

2. SVD and the Eckart-Young theorem

The vector space $\mathcal{M} = \mathcal{M}_{m,n}$ of $m \times n$ matrices with real entries has a natural filtration with subvarieties $\mathcal{M}_r = \{m \times n \text{ matrices of rank } \leq r\}$. We have

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \ldots \subset \mathcal{M}_{\min\{m,n\}}$$

where the last subvariety $\mathcal{M}_{\min\{m,n\}}$ coincides with the ambient space.

THEOREM 2.1 (Singular Value Decomposition). *Any real $m \times n$ matrix A has the SVD*

$$A = U\Sigma V^t$$

where U, V are orthogonal (respectively of size $m \times m$ and $n \times n$) and $\Sigma = \mathrm{Diag}(\sigma_1, \sigma_2, \ldots)$, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$. The $m \times n$ matrix Σ has zero values at entries (ij) with $i \neq j$ and sometimes it is called pseudodiagonal (we use the term diagonal only for square matrices).

The diagonal entries σ_i are called *singular values* of A and it is immediate to check that σ_i^2 are the eigenvalues of both the symmetric matrices AA^t and A^tA. We give a proof of Theorem 2.1 in §3. We recommend [12] for a nice historical survey about SVD. Decomposing $\Sigma = \mathrm{Diag}(\sigma_1, 0, 0, \cdots) + \mathrm{Diag}(0, \sigma_2, 0, \cdots) + \cdots =: \Sigma_1 + \Sigma_2 + \cdots$ we find

$$A = U\Sigma_1 V^t + U\Sigma_2 V^t + \cdots$$

and the maximum i for which $\sigma_i \neq 0$ is equal to the rank of the matrix A.

Denote by u_k, v_l the columns, respectively, of U and V in the SVD above. From the equality $A = U\Sigma V^t$ we get $AV = U\Sigma$ and considering the ith columns we get $Av_i = (AV)_i = (U\Sigma)_i = ((u_1, \cdots, u_m)\mathrm{Diag}(\sigma_1, \sigma_2, \cdots))_i = \sigma_i u_i$, while, from the transposed equality $A^t = V\Sigma^t U^t$, we get $A^t u_i = \sigma_i v_i$.

So if $1 \leq i \leq min\{m, n\}$, the columns u_i and v_i satisfy the conditions

$$Av_i = \sigma_i u_i \quad \text{and} \quad A^t u_i = \sigma_i v_i. \tag{1}$$

DEFINITION 2.2. *The pairs (u_i, v_i) in (1) are called singular vector pairs.*

More precisely, if $1 \leq i \leq min\{m, n\}$, the vectors u_i and v_i are called, respectively, left-singular and right-singular vectors for the singular value σ_i.

If the value σ_i appears only once in Σ, then the corresponding pair (u_i, v_i) is unique up to sign multiplication.

REMARK 2.3. *The right-singular vectors corresponding to zero singular values of A span the kernel of A; they are the last $n - rk(A)$ columns of V.*

The left-singular vectors corresponding to non-zero singular values of A span the image of A; they are the first $rk(A)$ columns of U.

REMARK 2.4. *The uniqueness property mentioned in Definition 2.2 shows that SVD of a general matrix is unique up to simultaneous sign change in each pair of singular vectors u_i and v_i. With an abuse of notation, it is customary to think projectively and to refer to "the" SVD of A, forgetting the sign change. See Theorem 3.4 for more about uniqueness.*

For later use, we observe that $U\Sigma_i V^t = \sigma_i u_i \cdot v_i^t$.

Let $||-||$ denote the usual l^2 norm (called also Frobenius or Hilbert-Schmidt norm) on $\mathcal{M}$, that is $\forall A \in \mathcal{M}$
$||A|| := \sqrt{tr(AA^t)} = \sqrt{\sum_{i,j} a_{ij}^2}$. Note that if $A = U\Sigma V^t$, then $||A|| = \sqrt{\sum_i \sigma_i^2}$.

The Eckart-Young Theorem uses SVD of the matrix A to find the matrices in $\mathcal{M}_r$ which minimize the distance from A.

THEOREM 2.5 (Eckart-Young, 1936). *Let $A = U\Sigma V^t$ be the SVD of a matrix A. Then*

- *$U\Sigma_1 V^t$ is the best rank 1 approximation of A, that is $||A - U\Sigma_1 V^t|| \le ||A - X||$ for every matrix X of rank 1.*
- *For any $1 \le r \le rank(A)$, $U\Sigma_1 V^t + \ldots + U\Sigma_r V^t$ is the best rank r approximation of A, that is $||A - U\Sigma_1 V^t - \ldots - U\Sigma_r V^t|| \le ||A - X||$ for every matrix X of rank $\le r$.*

Among the infinitely many rank one decompositions available for matrices, the Eckart-Young Theorem detects the one which is particularly nice in optimization problems. We will prove Theorem 2.5 in the more general formulation of Theorem 2.9.

2.1. Secant varieties and the Terracini Lemma

Secant varieties give basic geometric interpretation of rank of matrices and also of rank of tensors, as we will see in section 4.

Let $\mathcal{X} \subset \mathbb{P}V$ be an irreducible variety. The k-secant variety of $\mathcal{X}$ is defined by

$$\sigma_k(\mathcal{X}) := \overline{\bigcup_{p_1,\ldots,p_k \in \mathcal{X}} \mathbb{P}Span\{p_1, \ldots, p_k\}} \tag{2}$$

where $\mathbb{P}Span\{p_1, \ldots, p_k\}$ is the smallest projective linear space containing $p_1, \ldots, p_k$ and the overbar means Zariski closure (which is equivalent to Euclidean closure in all cases considered in this paper).

There is a filtration $\mathcal{X} = \sigma_1(\mathcal{X}) \subset \sigma_2(\mathcal{X}) \subset \ldots$

This ascending chain stabilizes when it fills the ambient space.

Example 2.6 (Examples of secant varieties in matrix spaces.). *We may identify the space $\mathcal{M}$ of $m \times n$ matrices with the tensor product $\mathbb{R}^m \otimes \mathbb{R}^n$. Hence we have natural inclusions $\mathcal{M}_r \subset \mathbb{R}^m \otimes \mathbb{R}^n$. Since $\mathcal{M}_r$ are cones, with an abuse of notation we may call with the same name the associated projective variety $\mathcal{M}_r \subset \mathbb{P}(\mathbb{R}^m \otimes \mathbb{R}^n)$. The basic equality we need is*

$$\sigma_r(\mathcal{M}_1) = \mathcal{M}_r$$

which corresponds to the fact that any rank r matrix can be written as the sum of r rank one matrices.

In this case the Zariski closure in (2) is not necessary, since the union is already closed.

The Terracini Lemma (see [9] for a proof) describes the tangent space $\mathbb{T}$ of a k-secant variety at a general point.

Lemma 2.7 (Terracini Lemma). *Let $z \in \mathbb{P}Span\,\{p_1, \ldots, p_k\}$ be general. Then*

$$\mathbb{T}_z\sigma_k(\mathcal{X}) = \mathbb{P}Span\,\{\mathbb{T}_{p_1}\mathcal{X}, \ldots, \mathbb{T}_{p_k}\mathcal{X}\}\,.$$

Example 2.8 (Tangent spaces to $\mathcal{M}_r$). *The tangent space to $\mathcal{M}_1$ at a point $u \otimes v$ is $\mathbb{R}^m \otimes v + u \otimes \mathbb{R}^n$:*

any curve $\gamma(t) = u(t) \otimes v(t)$ in $\mathcal{M}_1$ with $\gamma(0) = u \otimes v$ has derivative for $t = 0$ given by $u'(0) \otimes v + u \otimes v'(0)$ and since $u'(0), v'(0)$ are arbitrary vectors in $\mathbb{R}^m, \mathbb{R}^n$ respectively, we get the thesis.

As we have seen in Example 2.6, the variety $\mathcal{M}_r$ can be identified with the r-secant variety of $\mathcal{M}_1$, so the tangent space to $\mathcal{M}_r$ at a point $U(\Sigma_1 + \cdots + \Sigma_r)V^t$ can be described, by the Terracini Lemma, as $\mathbb{T}_{U\Sigma_1V^t}\mathcal{M}_1 + \cdots + \mathbb{T}_{U\Sigma_rV^t}\mathcal{M}_1 = \mathbb{T}_{\sigma_1u_1\otimes v_1^t}\mathcal{M}_1 + \cdots + \mathbb{T}_{\sigma_ru_r\otimes v_r^t}\mathcal{M}_1 = (\mathbb{R}^m \otimes v_1^t + u_1 \otimes \mathbb{R}^n) + \cdots + (\mathbb{R}^m \otimes v_r^t + u_r \otimes \mathbb{R}^n)$.

2.2. A geometric perspective on the Eckart-Young Theorem

Consider the variety $\mathcal{M}_r \subset \mathbb{R}^m \otimes \mathbb{R}^n$ of matrices of rank $\leq r$ and for any matrix $A \in \mathbb{R}^m \otimes \mathbb{R}^n$ let $d_A(-) = d(A, -)\colon \mathcal{M}_r \to \mathbb{R}$ be the (Euclidean) distance function from A. If $\mathrm{rk}A \geq r$ then the minimum on $\mathcal{M}_r$ of d_A is achieved on some matrices of rank r. This can be proved by applying the following Theorem 2.9 to $\mathcal{M}_{r'}$ for any $r' \leq r$. Since the variety $\mathcal{M}_r$ is singular exactly on $\mathcal{M}_{r-1}$, the minimum of d_A can be found among the critical points of d_A on the smooth part $\mathcal{M}_r \setminus \mathcal{M}_{r-1}$.

Theorem 2.9 (Eckart-Young revisited [5, Example 2.3]). *Let $A = U\Sigma V^t$ be the SVD of a matrix A and let $1 \leq r \leq rk(A)$. All the critical points of the distance function from A to the (smooth) variety $\mathcal{M}_r \setminus \mathcal{M}_{r-1}$ are given by $U(\Sigma_{i_1} + \ldots + \Sigma_{i_r})V^t$, where $\Sigma_i = \mathrm{Diag}(0, \ldots, 0, \sigma_i, 0, \ldots, 0)$, with $1 \leq i \leq$*

$rk(A)$. If the nonzero singular values of A are distinct then the number of critical points is $\binom{rk(A)}{r}$.

Note that $U\Sigma_i V^t$ are all the critical points of the distance function from A to the variety $\mathcal{M}_1$ of rank one matrices. So we have the important fact that all the critical points of the distance function from A to $\mathcal{M}_1$ allow to recover the SVD of A.

For the proof of Theorem 2.9 we need

LEMMA 2.10. *If $A_1 = u_1 \otimes v_1$, $A_2 = u_2 \otimes v_2$ are two rank one matrices, then $< A_1, A_2 >=< u_1, u_2 >< v_1, v_2 >$.*

Proof. $< A_1, A_2 >= tr(A_1 A_2^t) =$

$$tr[\left(\begin{pmatrix} u_{11} \\ \vdots \\ u_{1m} \end{pmatrix} \cdot (v_{11}, \cdots, v_{1n})\right)\left(\begin{pmatrix} v_{21} \\ \vdots \\ v_{2n} \end{pmatrix} \cdot (u_{21}, \cdots, u_{2m})\right)] =$$

$\sum_i u_{1i} \left(\sum_k v_{1k} v_{2k}\right) u_{2i} = \sum_i u_{1i} u_{2i} \sum_k v_{1k} v_{2k} =< u_1, u_2 >< v_1, v_2 >$. □

LEMMA 2.11. *Let $B \in \mathcal{M}$. If $< B, \mathbb{R}^m \otimes v >= 0$, then $< \mathrm{Row}(B), v >= 0$.*

If $< B, u \otimes \mathbb{R}^n >= 0$, then $< \mathrm{Col}(B), u >= 0$.

Proof. Let $\{e_1, \cdots, e_m\}$ be the canonical basis of $\mathbb{R}^m$; then, by hypothesis, $< B, e_k \otimes v >= 0 \ \forall k = 1, \cdots, m$.

We have $0 = tr\,[B(v^t \otimes e_k^t)] = tr\,[B(0, \cdots, 0, v, 0, \cdots, 0)] =< B^k, v >$, where B^k denotes the kth row of B, so that the space $\mathrm{Row}(B)$ is orthogonal to the vector v. In a similar way, we get $< \mathrm{Col}(B), u >= 0$. □

By using Terracini Lemma 2.7 we can prove Theorem 2.9.

Proof of Theorem 2.9. The matrix $U(\Sigma_{i_1} + \cdots + \Sigma_{i_r})V^t$ is a critical point of the distance function from A to the variety $\mathcal{M}_r$ if and only if the vector $A - (U(\Sigma_{i_1} + \cdots + \Sigma_{i_r})V^t)$ is orthogonal to the tangent space (see 2.8) $\mathbb{T}_{U(\Sigma_{i_1}+\cdots+\Sigma_{i_r})V^t}\mathcal{M}_r = (\mathbb{R}^m \otimes v_{i_1}^t + u_{i_1} \otimes \mathbb{R}^n) + \cdots + (\mathbb{R}^m \otimes v_{i_r}^t + u_{i_r} \otimes \mathbb{R}^n)$.

From the SVD of A we have $A - (U(\Sigma_{i_1} + \cdots + \Sigma_{i_r})V^t) = U(\Sigma_{j_1} + \cdots + \Sigma_{j_l})V^t = \sigma_{j_1} u_{j_1} \otimes v_{j_1}^t + \cdots + \sigma_{j_l} u_{j_l} \otimes v_{j_l}^t$ where $\{j_1, \cdots, j_l\}$ is the set of indices given by the difference $\{1, \cdots, rk(A)\} \setminus \{i_1, \cdots, i_r\}$.

Let $\{e_1, \cdots, e_m\}$ be the canonical basis of $\mathbb{R}^m$. By Lemma 2.10 we get: $< \sigma_{j_h} u_{j_h} \otimes v_{j_h}^t, e_l \otimes v_{i_k}^t >= \sigma_{j_h} < u_{j_h}, e_l >< v_{j_h}, v_{i_k} >= 0$ since v_{j_h}, v_{i_k} are distinct columns of the orthogonal matrix V. So the matrices $U\Sigma_{j_h}V^t$ are orthogonal to the spaces $\mathbb{R}^m \otimes v_{i_k}^t$.

In a similar way, since U is an orthogonal matrix, the matrices $U\Sigma_{j_h}V^t$ are orthogonal to the spaces $u_{i_k} \otimes \mathbb{R}^n$. So $A - (U(\Sigma_{i_1} + \cdots + \Sigma_{i_r})V^t)$ is orthogonal to the tangent space and $U(\Sigma_{i_1} + \cdots + \Sigma_{i_r})V^t$ is a critical point.

Let now $B \in \mathcal{M}_r$ be a critical point of the distance function from A to $\mathcal{M}_r$. Then $A - B$ is orthogonal to the tangent space $\mathbb{T}_B\mathcal{M}_r$.

Let $B = U'(\Sigma'_1 + \cdots \Sigma'_r)V'^t$, $A - B = U''(\Sigma''_1 + \cdots \Sigma''_l)V''^t$ be SVD of B and $A - B$ respectively, with $\Sigma'_r \neq 0$ and $\Sigma''_l \neq 0$.

Since $A - B$ is orthogonal to $\mathbb{T}_B\mathcal{M}_r = (\mathbb{R}^m \otimes {v'_1}^t + u'_1 \otimes \mathbb{R}^n) + \cdots + (\mathbb{R}^m \otimes {v'_r}^t + u'_r \otimes \mathbb{R}^n)$, by Lemma 2.11 we get $< Col(A - B), u'_k >= 0$ and $< Row(A - B), v'_k >= 0 \quad k = 1, \cdots, r$. In particular, $Col(A - B)$ is a vector subspace of $Span\{u'_1, \cdots, u'_r\}^\perp$ and has dimension at most $m - r$ while $Row(A - B)$ is a vector subspace of $Span\{v'_1, \cdots, v'_r\}^\perp$ and has dimension at most $n - r$, so that $l \leq min\{m, n\} - r$.

From the equality $A - B = (u''_1, \ldots, u''_l, 0 \ldots, 0)\,(\Sigma''_1 + \cdots \Sigma''_l)V''^t$ we get $Col(A - B) \subset Span\{u''_1, \ldots, u''_l\}$ and equality holds by dimensional reasons.

In a similar way, $Row(A - B) = Span\{v''_1, \cdots, v''_l\}$. This implies that the orthonormal columns $u''_1, \cdots, u''_l, u'_1, \cdots, u'_r$ can be completed with orthonormal $m - l - r$ columns of $\mathbb{R}^m$ to obtain an orthogonal $m \times m$ matrix U, while the orthonormal columns $v''_1, \cdots, v''_l, v'_1, \cdots, v'_r$ can be completed with orthonormal $n - l - r$ columns of $\mathbb{R}^n$ to obtain an orthogonal $n \times n$ matrix V.

We get $A - B = U \begin{pmatrix} \Sigma'' & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^t, \quad B = U \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Sigma' & 0 \\ 0 & 0 & 0 \end{pmatrix} V^t$, where $\Sigma'' = \mathrm{Diag}(\sigma''_1, \ldots, \sigma''_l)$ and $\Sigma' = \mathrm{Diag}(\sigma'_1, \ldots, \sigma'_r)$.

So $A = (A - B) + B = U \begin{pmatrix} \Sigma'' & 0 & 0 \\ 0 & \Sigma' & 0 \\ 0 & 0 & 0 \end{pmatrix} V^t$ can easily be transformed to a SVD of A by just reordering the diagonal elements σ'_i's and σ''_i's and the critical point B is of the desired type. $\square$

The following result has the same flavour of Eckart-Young Theorem 2.9.

THEOREM 2.12 (Baaijens, Draisma [1, Theorem 3.2]). *Let $A = U\Sigma V^t$ be the SVD of a $n \times n$ matrix A. All the critical points of the distance function from A to the variety $O(n)$ of orthogonal matrices are given by the orthogonal matrices $U\mathrm{Diag}(\pm 1, \ldots, \pm 1)V^t$ and their number is 2^n.*

Actually, in [1], the result is stated in a slightly different form, which is equivalent to this one, that we have chosen to make more transparent the link with SVD. It is easy to check that, among the critical points computed in Theorem 2.12, the one with all plus signs, corresponding to the orthogonal matrix UV^t, gives the orthogonal matrix closest to A. This is called the Löwdin orthogonalization (or symmetric orthogonalization) of A.

3. SVD via the Spectral Theorem

In this section we prove Theorem 2.1 as a consequence of the Spectral Theorem. We recall

THEOREM 3.1 (Spectral Theorem). *For any symmetric real matrix B, there exists an orthogonal matrix V such that $V^{-1}BV = V^tBV$ is a diagonal matrix.*

REMARK 3.2. *Since the Euclidean inner product is positive definite, it is elementary to show that for any real $m \times n$ matrix A we have $Ker(A^tA) = Ker(A)$ and $Ker(AA^t) = Ker(A^t)$.*

Proof of Theorem 2.1 Let A be an $m \times n$ matrix with real entries. The matrix A^tA is a symmetric matrix of order n and it's positive semidefinite. By the Spectral Theorem, there exists an orthogonal matrix V (of order n) such that

$$V^{-1}(A^tA)V = V^t(A^tA)V = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where D is diagonal of order $r = \mathrm{r}k(A^tA) = \mathrm{r}k(A)$ (see Remark 3.2) and is positive definite: $D = \mathrm{Diag}(d_1, \cdots, d_r)$ with $d_1 \geq d_2 \geq \cdots d_r > 0$.
Let $v_1, \cdots, v_n$ be the columns of V; then

$$(A^tA)(v_1, \cdots, v_n) = (v_1, \cdots, v_n) \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = (d_1v_1, \cdots, d_rv_r, 0, \cdots, 0)$$

and $v_{r+1}, \cdots, v_n \in Ker(A^tA) = Ker(A)$ (see Remark 3.2).
Let $\sigma_i = \sqrt{d_i},\ i = 1, \cdots, r$ and let $u_i = (1/\sigma_i)Av_i \in \mathbb{R}^m$. These vectors are orthonormal since $< u_i, u_j >= \frac{1}{\sigma_i\sigma_j} < Av_i, Av_j >= \frac{1}{\sigma_i\sigma_j} < v_i, A^tAv_j >= \frac{1}{\sigma_i\sigma_j} < v_i, d_jv_j >= \frac{\sigma_i}{\sigma_j} < v_i, v_j >= \frac{\sigma_i}{\sigma_j}\delta_{ij}$. Thus it's possible to find $m - r$ orthonormal vectors in $\mathbb{R}^m$ such that the matrix $U := (u_1, \cdots, u_r, u_{r+1}, \cdots, u_m)$ is an $m \times m$ orthogonal matrix. Define $\Sigma := \begin{pmatrix} D^{1/2} & 0 \\ 0 & 0 \end{pmatrix}$ to be an $m \times n$ matrix with $m - r$ zero rows, $D^{1/2} = \mathrm{Diag}(\sigma_1, \cdots, \sigma_r)$. Then

$$U\Sigma V^t = \left(\frac{1}{\sigma_1}Av_1, \cdots, \frac{1}{\sigma_r}Av_r, u_{r+1}, \cdots, u_m\right) \begin{pmatrix} \sigma_1 v_1^t \\ \vdots \\ \sigma_r v_r^t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = A(v_1, \cdots, v_r) \begin{pmatrix} v_1^t \\ \vdots \\ v_r^t \end{pmatrix}.$$

Since V is orthogonal we have

$$I_n = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} v_1^t \\ \vdots \\ v_n^t \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix} \begin{pmatrix} v_1^t \\ \vdots \\ v_r^t \end{pmatrix} + \begin{pmatrix} v_{r+1} & \cdots & v_n \end{pmatrix} \begin{pmatrix} v_{r+1}^t \\ \vdots \\ v_n^t \end{pmatrix}$$

Hence we get

$$U\Sigma V^t = A\left(I_n - \begin{pmatrix} v_{r+1} & \cdots & v_n \end{pmatrix}\begin{pmatrix} v_{r+1}^t \\ \vdots \\ v_n^t \end{pmatrix}\right) = A$$

since $v_{r+1}, \cdots, v_n \in Ker(A)$. □

LEMMA 3.3. • *(1) Let $\sigma_1^2 > \cdots > \sigma_k^2 > 0$ be the distinct non zero eigenvalues of A^tA and $V_i = Ker(A^tA - \sigma_i^2 I_n)$ be the corresponding eigenspaces, $V_0 = Ker(A^tA) = Ker(A)$. Then*

$$\mathbb{R}^n = \left(\oplus_{i=1}^k V_i\right) \oplus V_0$$

is an orthogonal decomposition of $\mathbb{R}^n$.

- *(2) Let $U_i = Ker(AA^t - \sigma_i^2 I_m)$, $U_0 = Ker(AA^t) = Ker(A^t)$. Then*

$$AV_i = U_i \quad \text{and} \quad A^tU_i = V_i \ \text{if } i = 1, \cdots, r.$$

- *(3)*

$$\mathbb{R}^m = \left(\oplus_{i=1}^k U_i\right) \oplus U_0$$

is an orthogonal decomposition of $\mathbb{R}^m$ and $\sigma_1^2 > \cdots > \sigma_k^2 > 0$ are the distinct non zero eigenvalues of AA^t.

- *(4) The isomorphism $\frac{1}{\sigma_i}A|_{V_i} : V_i \longrightarrow U_i$ is an isometry with inverse $\frac{1}{\sigma_i}A^t|_{U_i} : U_i \longrightarrow V_i$.*

Proof. (1) is the Spectral Theorem. In order to prove (2), $AV_i \subseteq U_i$ since $\forall w \in V_i$ one has $(AA^t)(Aw) = A(A^tA)w = \sigma_i^2 Aw$. In a similar way, $A^tU_i \subseteq V_i$. On the other hand, $\forall z \in U_i$ one has $z = \frac{1}{\sigma_i^2}(AA^t)z = A(\frac{1}{\sigma_i^2}A^tz) \in AV_i$ so that $AV_i = U_i$. In a similar way, $A^tU_i = V_i$. (3) and (4) are immediate from (2). □

Lemma 3.3 may be interpretated as the following coordinate free version of SVD, that shows precisely in which sense SVD is unique.

THEOREM 3.4 (Coordinate free version of SVD). *Let $\mathcal{V}, \mathcal{U}$ be real vector spaces of finite dimension endowed with inner products $<,>_{\mathcal{V}}$ and $<,>_{\mathcal{U}}$ and let $F\colon \mathcal{V} \to \mathcal{U}$ be a linear map with adjoint $F^t\colon \mathcal{U} \to \mathcal{V}$, defined by the property $< Fv, u >_{\mathcal{U}} = < v, F^tu >_{\mathcal{V}}$ $\forall v \in \mathcal{V}, \forall u \in \mathcal{U}$. Then there is a unique decomposition (SVD)*

$$F = \sum_{i=1}^k \sigma_i F_i$$

with $\sigma_1 > \ldots > \sigma_k > 0$, $F_i\colon \mathcal{V} \to \mathcal{U}$ linear maps such that

- $F_iF_j^t$ *and* $F_i^tF_j$ *are both zero for any* $i \neq j$,
- $F_{i|Im(F_i^t)}\colon Im(F_i^t) \to Im(F_i)$ *is an isometry with inverse* F_i^t.

Both the singular values σ_i *and the linear maps* F_i *are uniquely determined from* F.

By taking the adjoint in Theorem 3.4, $F^t = \sum_{i=1}^k \sigma_i F_i^t$ is the SVD of F^t. The first interesting consequence is that

$$FF^t = \sum_{i=1}^k \sigma_i^2 F_i F_i^t \quad \text{and} \quad F^tF = \sum_{i=1}^k \sigma_i^2 F_i^t F_i$$

are both spectral decomposition (and SVD) of the self-adjoint operators FF^t and F^tF. This shows the uniqueness in Theorem 3.4. Note that
$\mathcal{V} = \left(\oplus_{i=1}^k Im(F_i^t)\right) \bigoplus KerF$ and $\mathcal{U} = \left(\oplus_{i=1}^k Im(F_i)\right) \bigoplus KerF^t$ are both orthogonal decompositions and that $\mathrm{rk}F = \sum_{i=1}^k \mathrm{rk}F_i$.

Moreover, $F^+ = \sum_{i=1}^k \sigma_i^{-1} F_i^t$ is the Moore-Penrose inverse of F, expressing also the SVD of F^+.

Theorem 3.4 extends in a straightforward way to finite dimensional complex vector spaces $\mathcal{V}$ and $\mathcal{U}$ endowed with Hermitian inner products.

4. Basics on tensors and tensor rank

We consider tensors $A \in K^{n_1+1} \otimes \ldots \otimes K^{n_d+1}$ where $K = \mathbb{R}$ or $\mathbb{C}$. It is convenient to consider complex tensors even if one is interested only in the real case.

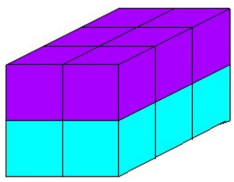

Figure 1: The visualization of a tensor in $K^3 \otimes K^2 \otimes K^2$.

Entries of A are labelled by d indices as $a_{i_1 \ldots i_d}$.

For example, the expression in coordinates of a $3 \times 2 \times 2$ tensor A as in Figure 1 is, with obvious notations,

$$\begin{aligned} A = {} & a_{000}x_0y_0z_0 + a_{001}x_0y_0z_1 + a_{010}x_0y_1z_0 + a_{011}x_0y_1z_1 + \\ & a_{100}x_1y_0z_0 + a_{101}x_1y_0z_1 + a_{110}x_1y_1z_0 + a_{111}x_1y_1z_1 + \\ & a_{200}x_2y_0z_0 + a_{201}x_2y_0z_1 + a_{210}x_2y_1z_0 + a_{211}x_2y_1z_1. \end{aligned}$$

Definition 4.1. *A tensor A is decomposable if there exist $x^i \in K^{n_i+1}$, for $i = 1, \ldots, d$, such that $a_{i_1 \ldots i_d} = x^1_{i_1} x^2_{i_2} \ldots x^d_{i_d}$. In equivalent way, $A = x^1 \otimes \ldots \otimes x^d$.*

Define the rank of a tensor A as the minimal number of decomposable summands expressing A, that is

$$\mathrm{rk}(A) := \min\{r | A = \sum_{i=1}^{r} A_i, A_i \text{ are decomposable}\}$$

For matrices, this coincides with usual rank. For a (nonzero) tensor, decomposable $\Longleftrightarrow$ rank one.

Any expression $A = \sum_{i=1}^{r} A_i$ with A_i decomposable is called a *tensor decomposition.*

As for matrices, the space of tensors of format $(n_1+1) \times \ldots \times (n_d+1)$ has a natural filtration with subvarieties $\mathcal{T}_r = \{A \in K^{n_1+1} \otimes \ldots \otimes K^{n_d+1} |$ rank $(A) \leq r\}$. We have

$$\mathcal{T}_1 \subset \mathcal{T}_2 \subset \ldots$$

Corrado Segre in XIX century understood this filtration in terms of projective geometry, since $\mathcal{T}_i$ are cones.

The decomposable (or rank one) tensors give the "Segre variety"

$$\mathcal{T}_1 \simeq \mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_d} \subset \mathbb{P}(K^{n_1+1} \otimes \ldots \otimes K^{n_d+1})$$

The variety $\mathcal{T}_k$ is again the k-secant variety of $\mathcal{T}_1$, like in the case of matrices.

For $K = \mathbb{R}$, the Euclidean inner product on each space $\mathbb{R}^{n_i+1}$ induces the inner product on the tensor product $\mathbb{R}^{n_1+1} \otimes \ldots \otimes \mathbb{R}^{n_d+1}$ (compare with Lemma 2.10). With respect to this product we have the equality $||x_1 \otimes \ldots \otimes x_d||^2 = \prod_{i=1}^{d} ||x_i||^2$. A *best rank r approximation* of a real tensor A is a tensor in $\mathcal{T}_r$ which minimizes the l^2-distance function from A. We will discuss mainly the best rank one approximations of A, considering the critical points $T \in \mathcal{T}_1$ for the l^2-distance function from A to the variety $\mathcal{T}_1$ of rank 1 tensors, trying to extend what we did in §2. The condition that T is a critical point is again that the tangent space at T is orthogonal to the tensor $A - T$.

Theorem 4.2 (Lim, variational principle [10]). *The critical points $x_1 \otimes \ldots \otimes x_d \in \mathcal{T}_1$ of the distance function from $A \in \mathbb{R}^{n_1+1} \otimes \ldots \otimes \mathbb{R}^{n_d+1}$ to the variety $\mathcal{T}_1$ of rank 1 tensors are given by d-tuples $(x_1, \ldots, x_d) \in \mathbb{R}^{n_1+1} \times \ldots \times \mathbb{R}^{n_d+1}$ such that*

$$A \cdot (x_1 \otimes \ldots \widehat{x_i} \ldots \otimes x_d) = \lambda x_i \quad \forall i = 1, \ldots, d \tag{3}$$

where $\lambda \in \mathbb{R}$, the dot means contraction and the notation $\widehat{x_i}$ means that x_i has been removed.

Note that the left-hand side of (3) is an element in the dual space of $\mathbb{R}^{n_i+1}$, so in order for (3) to be meaningful it is necessary to have a metric identifying $\mathbb{R}^{n_i+1}$ with its dual. We may normalize the factors x_i of the tensor product $x_1\otimes\ldots\otimes x_d$ in such a way that $||x_i||^2$ does not depend on i. Note that from (3) we get $A\cdot(x_1\otimes\ldots\otimes x_d)=\lambda||x_i||^2$. Here, $(x_1,\ldots,x_d)$ is called a *singular vector d-tuple* (defined independently by Qi in [11]) and λ is called a *singular value*. Allowing complex solutions to (3), λ may be complex.

EXAMPLE 4.3. *We may compute all singular vector triples for the following tensor in* $\mathbb{R}^3\otimes\mathbb{R}^3\otimes\mathbb{R}^2$

$$\begin{aligned} f =& 6x_0y_0z_0 && +2x_1y_0z_0+6x_2y_0z_0 \\ &-2014x_0y_1z_0 && +121x_1y_1z_0-11x_2y_1z_0 \\ &+48x_0y_2z_0 && -13x_1y_2z_0-40x_2y_2z_0 \\ &-31x_0y_0z_1 && +93x_1y_0z_1+97x_2y_0z_1 \\ &+63x_0y_1z_1 && +41x_1y_1z_1-94x_2y_1z_1 \\ &-3x_0y_2z_1 && +47x_1y_2z_1+4x_2y_2z_1 \end{aligned}$$

We find 15 *singular vector triples,* 9 *of them are real,* 6 *of them make* 3 *conjugate pairs.*

The minimum distance is 184.038 *and the best rank one approximation is given by the singular vector triple*
$(x_0-.0595538x_1+.00358519x_2)(y_0-289.637y_1+6.98717y_2)(6.95378z_0-.2079687z_1)$. *Tensor decomposition of f can be computed from the Kronecker normal form and gives f as sum of three decomposable summands, that is*

$$\begin{aligned} f =& (.450492x_0-1.43768x_1-1.40925x_2)(-.923877y_0-.986098y_1-.646584y_2)(.809777z_0+68.2814z_1)+ \\ &(-.582772x_0+.548689x_1+1.93447x_2)(.148851y_0-3.43755y_1-1.07165y_2)(18.6866z_0+28.1003z_1)+ \\ &(1.06175x_0-.0802873x_1-.0580488x_2)(-.0125305y_0+3.22958y_1-.0575754y_2)(-598.154z_0+10.8017z_1) \end{aligned}$$

Note that the best rank one approximation is unrelated to the three summands of minimal tensor decomposition, in contrast with the Eckart-Young Theorem for matrices.

THEOREM 4.4 (Lim, variational principle in symmetric case [10]). *The critical points of the distance function from $A\in\mathrm{Sym}^d\mathbb{R}^{n+1}$ to the variety $\mathcal{T}_1$ of rank* 1 *tensors are given by d-tuples $x^d\in\mathrm{Sym}^d\mathbb{R}^{n+1}$ such that*

$$A\cdot(x^{d-1})=\lambda x. \tag{4}$$

The tensor x in (4) is called a *eigenvector*, the corresponding power x^d is a *eigentensor* , λ is called a *eigenvalue.*

4.1. Dual varieties and hyperdeterminant

If $\mathcal{X}\subset\mathbb{P}V$ then

$$\mathcal{X}^*:=\overline{\{H\in\mathbb{P}V^*|\exists\text{ smooth point }p\in\mathcal{X}\text{ s.t. }\mathbb{T}_p\mathcal{X}\subset H\}}$$

is called the *dual variety* of $\mathcal{X}$ (see [7, Chapter 1]). So $\mathcal{X}^*$ consists of hyperplanes tangent at some smooth point of $\mathcal{X}$.

In Euclidean setting, duality may be understood in terms of orthogonality. Considering the affine cone of a projective variety $\mathcal{X}$, the dual variety consists of the cone of all vectors which are orthogonal to some tangent space to $\mathcal{X}$.

Let $m \leq n$. The dual variety of $m \times n$ matrices of rank $\leq r$ is given by $m \times n$ matrices of rank $\leq m - r$ ([7, Prop. 4.11]). In particular, the dual of the Segre variety of matrices of rank 1 is the determinant hypersurface.

Let $n_1 \leq \ldots \leq n_d$. The dual variety of tensors of format $(n_1 + 1) \times \ldots \times (n_d + 1)$ is by definition the *hyperdeterminant* hypersurface, whenever $n_d \leq \sum_{i=1}^{d-1} n_i$. Its equation is called the hyperdeterminant. Actually, this defines the hyperdeterminant up to scalar multiple, but it can be normalized asking that the coefficient of its leading monomial is 1.

5. Basic properties of singular vector tuples and of eigentensors. The singular space of a tensor.

5.1. Counting the singular tuples

In this subsection we expose the results of [6] about the number of singular tuples (see Theorem 4.2) of a general tensor.

Theorem 5.1 ([6]). *The number of (complex) singular d-tuples of a general tensor $t \in \mathbb{P}(\mathbb{R}^{n_1+1} \otimes \ldots \otimes \mathbb{R}^{n_d+1})$ is equal to the coefficient of $\prod_{i=1}^{d} t_i^{n_i}$ in the polynomial*

$$\prod_{i=1}^{d} \frac{\hat{t}_i^{\,n_i+1} - t_i^{n_i+1}}{\hat{t}_i - t_i}$$

where $\hat{t}_i = \sum_{j \neq i} t_j$.

Amazingly, for $d = 2$ this formula gives the expected value $\min(n_1 + 1, n_2 + 1)$.

For the proof, in [6] the d-tuples of singular vectors were expressed as zero loci of sections of a suitable vector bundle on the Segre variety $\mathcal{T}_1$.

Precisely, let $\mathcal{T}_1 = \mathbb{P}(\mathbb{C}^{n_1+1}) \times \ldots \times \mathbb{P}(\mathbb{C}^{n_d+1})$ and let $\pi_i \colon \mathcal{T}_1 \to \mathbb{P}(\mathbb{C}^{n_i+1})$ be the projection on the i-th factor. Let $\mathcal{O}(\underbrace{1, \ldots, 1}_{d})$ be the very ample line bundle which gives the Segre embedding and let Q be the quotient bundle.

Then the bundle is $\oplus_{i=1}^{d} (\pi_i^* Q) \otimes \mathcal{O}(1, \ldots, 1, \underset{\substack{\uparrow \\ i}}{0}, 1, \ldots, 1)$.

The top Chern class of this bundle gives the formula in Theorem 5.1.

In the format $(\underbrace{2, \ldots, 2}_{d})$ the number of singular d-tuples is $d!$.

The following table lists the number of singular triples in the format (d_1, d_2, d_3)

d_1, d_2, d_3	$c(d_1, d_2, d_3)$	
$2, 2, 2$	6	
$2, 2, n$	8	$n \geq 3$
$2, 3, 3$	15	
$2, 3, n$	18	$n \geq 4$
$2, n, n$	$n(2n-1)$	
$3, 3, 3$	37	
$3, 3, 4$	55	
$3, 3, n$	61	$n \geq 5$
$3, 4, 4$	104	
$3, 4, 5$	138	
$3, 4, n$	148	$n \geq 6$

The number of singular d-tuples of a general tensor $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_d+1}$, when $n_1, \ldots, n_{d-1}$ are fixed and n_d increases, stabilizes for $n_d \geq \sum_{i=1}^{d-1} n_i$, as it can be shown from Theorem 5.1.

For example, for a tensor of size $2 \times 2 \times n$, there are 6 singular vector triples for $n = 2$ and 8 singular vector triples for $n \geq 3$.

The format with $n_d = \sum_{i=1}^{d-1} n_i$ is the *boundary format*, well known in hyperdeterminant theory [7]. It generalizes the square case for matrices.

The symmetric counterpart of Theorem 5.1 is the following

THEOREM 5.2 (Cartwright-Sturmfels [2]). *The number of (complex) eigentensors of a general tensor $t \in \mathbb{P}(\mathrm{Sym}^d \mathbb{R}^{n+1})$ is equal to*

$$\frac{(d-1)^{n+1} - 1}{d-2}.$$

5.2. The singular space of a tensor

We start informally to study the singular triples of a 3-mode tensor A, later we will generalize to any tensor. The singular triples $x \otimes y \otimes z$ of A satisfy (see Theorem 4.2) the equations

$$\sum_{i_0, i_1} A_{i_0 i_1 k} x_{i_0} y_{i_1} = \lambda z_k \quad \forall k$$

hence, by eliminating λ, the equations (for every $k < s$)

$$\sum_{i_0, i_1} \left(A_{i_0 i_1 k} x_{i_0} y_{i_1} z_s - A_{i_0 i_1 s} x_{i_0} y_{i_1} z_k \right) = 0$$

which are linear equations in the Segre embedding space. These equations can be permuted on x, y, z and give

$$\begin{cases} \sum_{i_0,i_1} (A_{i_0 i_1 k} x_{i_0} y_{i_1} z_s - A_{i_0 i_1 s} x_{i_0} y_{i_1} z_k) = 0 & \text{for } 0 \le k < s \le n_3 \\ \sum_{i_0,i_2} (A_{i_0 k i_2} x_{i_0} y_s z_{i_2} - A_{i_0 s i_2} x_{i_0} y_k z_{i_2}) = 0 & \text{for } 0 \le k < s \le n_2 \\ \sum_{i_1,i_2} (A_{k i_1 i_2} x_s y_{i_1} z_{i_2} - A_{s i_1 i_2} x_k y_{i_1} z_{i_2}) = 0 & \text{for } 0 \le k < s \le n_1 \end{cases} \tag{5}$$

These equations define the singular space of A, which is the linear span of all the singular vector triples of A.

The tensor A belongs to the singular space of A, as it is trivially shown by the following identity (and its permutations)

$$\sum_{i_0,i_1} (A_{i_0 i_1 k} A_{i_0 i_1 s} - A_{i_0 i_1 s} A_{i_0 i_1 k}) = 0.$$

In the symmetric case, the eigentensors x^d of a symmetric tensor $A \in \mathrm{Sym}^d \mathbb{C}^{n+1}$ are defined by the linear dependency of the two rows of the $2 \times (n+1)$ matrix

$$\begin{pmatrix} \nabla A(x^{d-1}) \\ x \end{pmatrix}$$

Taking the 2×2 minors we get the following

DEFINITION 5.3. *If $A \in \mathrm{Sym}^d \mathbb{C}^{n+1}$ is a symmetric tensor, then the singular space is given by the following $\binom{n+1}{2}$ linear equations in the unknowns x^d*

$$\frac{\partial A(x^{d-1})}{\partial x_j} x_i - \frac{\partial A(x^{d-1})}{\partial x_i} x_j = 0$$

It follows from the definition that the singular space of A is spanned by all the eigentensors x^d of A.

PROPOSITION 5.4. *The symmetric tensor $A \in \mathrm{Sym}^d \mathbb{C}^{n+1}$ belongs to the singular space of A. The dimension of the singular space is $\binom{n+d}{d} - \binom{n+1}{2}$. The eigentensors are independent for a general A (and then make a basis of the singular space) just in the cases* $\mathrm{Sym}^d \mathbb{C}^2$, $\mathrm{Sym}^2 \mathbb{C}^{n+1}$, $\mathrm{Sym}^3 \mathbb{C}^3$.

Proof. To check that A belongs to the singular space, consider dual variables $y_j = \frac{\partial}{\partial x_j}$. Then we have $\left(\frac{\partial A}{\partial y_j} y_i - \frac{\partial A}{\partial y_i} y_j\right) \cdot A(x) = \frac{\partial A}{\partial y_j} \cdot \frac{\partial A}{\partial x_i} - \frac{\partial A}{\partial y_i} \cdot \frac{\partial A}{\partial x_j}$, which vanishes by symmetry. To compute the dimension of the singular space, first recall that symmetric tensors in $\mathrm{Sym}^d \mathbb{C}^{n+1}$ correspond to homogeneous polynomials of degree d in $n+1$ variables. We have to show that for a general polynomial A, the $\binom{n+1}{2}$ polynomials $\frac{\partial A}{\partial x_j} x_i - \frac{\partial A}{\partial x_i} x_j$ for $i < j$ are independent. This is easily checked for the Fermat polynomial $A = \sum_{i=0}^n x_i^d$ for $d \ge 3$ and for the polynomial $A = \sum_{p<q} x_p x_q$ for $d = 2$. The case listed are the ones where the inequality

$$\frac{(d-1)^{n+1}-1}{d-2} \geq \binom{n+d}{d} - \binom{n+1}{2}$$

is an equality (for $d = 2$ the left-hand side reads as $\sum_{i=0}^{n}(d-1)^i = n+1$). $\square$

Denote by e_j the canonical basis in any vector space $\mathbb{C}^n$.

PROPOSITION 5.5. *Let $n_1 \leq \ldots \leq n_d$. If $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_d+1}$ is a tensor, then the singular space of A is given by the following $\sum_{i=1}^{d} \binom{n_i+1}{2}$ linear equations. The $\binom{n_i+1}{2}$ equations of the i-th group ($i = 1, \ldots, d$) for $x^1 \otimes \ldots \otimes x^d$ are*

$$A(x^1, x^2, \ldots, \underset{\substack{\uparrow \\ i}}{e_p}, \ldots, x^d)(x^i)_q - A(x^1, x^2, \ldots, \underset{\substack{\uparrow \\ i}}{e_q}, \ldots, x^d)(x^i)_p = 0$$

for $0 \leq p < q \leq n_i$. The tensor A belongs to this linear space, which we call again the singular space of A.

Proof. Let $A = \sum_{i_1,\ldots,i_d} A_{i_1,\ldots,i_d} x^1_{i_1} \otimes \ldots \otimes x^d_{i_d}$. Then we have for the first group of equations

$$\sum_{i_2,\ldots,i_d} (A_{k,i_2,\ldots,i_d} A_{s,i_2,\ldots,i_d} - A_{s,i_2,\ldots,i_d} A_{k,i_2,\ldots,i_d}) = 0 \quad \text{for } 0 \leq k < s \leq n_1$$

and the same argument works for the other groups of equations. $\square$

We state, with a sketch of the proof, the following generalization of the dimensional part of Prop. 5.4.

PROPOSITION 5.6. *Let $n_1 \leq \ldots \leq n_d$ and $N = \prod_{i=1}^{d-1}(n_i + 1)$. If $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_d+1}$ is a tensor, the dimension of the singular space of A is*

$$\begin{cases} \prod_{i=1}^{d}(n_i+1) - \sum_{i=1}^{d} \binom{n_i+1}{2} & \text{for } n_d + 1 \leq N \\ \binom{N+1}{2} - \sum_{i=1}^{d-1} \binom{n_i+1}{2} & \text{for } n_d + 1 \geq N. \end{cases}$$

The singular d-tuples are independent (and then make a basis of this space) just in cases $d = 2$, $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$, $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^n$ for $n \geq 4$.

Proof. Note that if $n_d + 1 \geq N$, for any tensor $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_d+1}$, there is a subspace $L \subset \mathbb{C}^{n_d+1}$ of dimension N such that $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_{d-1}+1} \otimes L$, hence all singular d-tuples of A lie in $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_{d-1}+1} \otimes L$. Note that for $n_d + 1 = \prod_{i=1}^{d-1}(n_i + 1)$, then the singular space has dimension $N =$

$\prod_{i=1}^d (n_i+1) - \sum_{i=1}^d \binom{n_i+1}{2}$. It can be shown that posing $k(i_1,\ldots,i_{d-1}) = \sum_{j=1}^{d-1}\left[\left(\prod_{s\le j-1}(n_s+1)\right) i_j\right]$, then the tensor

$$A = \sum_{i_1=0}^{n_1} \ldots \sum_{i_{d-1}=0}^{n_{d-1}} k(i_1,\ldots,i_{d-1}) e^1_{i_1} \ldots e^{d-1}_{i_{d-1}} e^d_{k(i_1,\ldots,i_{d-1})}$$

is general in the sense that the $\sum_{i=1}^d \binom{n_i+1}{2}$ corresponding equations are independent (note that $k(i_1,\ldots,i_{d-1})$ covers all integers between 0 and $N-1$). For $n_d+1 \ge N$ the dimension stabilizes to $N^2 - \sum_{i=1}^{d-1}\binom{n_i+1}{2} - \binom{N+1}{2} = \binom{N+1}{2} - \sum_{i=1}^{d-1}\binom{n_i+1}{2}$. □

Remark 5.7. *For a general $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_d+1}$, the $\sum_{i=1}^d \binom{n_i+1}{2}$ linear equations of the singular space of A are independent if $n_d + 1 \le N + 1$.*

Remark 5.8. *In the case of symmetric matrices, the singular space of A consists of all matrices commuting with A. If A is regular, this space is spanned by the powers of A. If A is any matrix (not necessarily symmetric), the singular space of A consists of all matrices with the same singular vector pairs as A. These properties seem not to generalize to arbitrary tensors. Indeed the tensors in the singular space of a tensor A may have singular vectors different from those of A, even in the symmetric case. This is apparent for binary forms. The polynomials g having the same eigentensors as f, satisfy the equation $g_x y - g_y x = \lambda(f_x y - f_y x)$ for some λ, which in degree d even has (in general) the solutions $g = \mu_1 f + \mu_2(x^2+y^2)^{d/2}$ with $\mu_1, \mu_2 \in \mathbb{C}$, while for degree d odd has (in general) the solutions $g = \mu_1 f$. In both cases, these solutions are strictly contained in the singular space of f.*

In any case, a positive result which follows from Prop. 5.4, 5.5, 5.6 is the following

Corollary 5.9. *(i) Let $n_1 \le \ldots \le n_d$, $N = \prod_{i=1}^{d-1}(n_i+1)$ and $M = \min(N, n_d+1)$. A general tensor $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_d+1}$ has a tensor decomposition given by $NM - \sum_{i=1}^{d-1}\binom{n_i+1}{2} - \binom{N+1}{2}$ singular vector d-tuples.*

(ii) A general symmetric tensor $A \in \mathrm{Sym}^d\mathbb{C}^{n+1}$ has a symmetric tensor decomposition given by $\binom{n+d}{d} - \binom{n+1}{2}$ eigentensors.

The decomposition in (i) is not minimal unless $d = 2$, when it is given by the SVD.

The decomposition in (ii) is not minimal unless $d = 2$, when it is the spectral decomposition, as sum of $(n+1)$ (squared) eigenvectors.

5.3. The Euclidean Distance Degree and its duality property

The construction of critical points of the distance from a point p, can be generalized to any affine (real) algebraic variety $\mathcal{X}$.

Following [5], we call Euclidean Distance Degree (shortly *ED degree*) the number of critical points of $d_p = d(p, -)\colon \mathcal{X} \to \mathbb{R}$, allowing complex solutions. As before, the number of critical points does not depend on p, provided p is generic. For a elementary introduction, see the nice survey [13].

Theorem 2.9 says that the ED degree of the variety $\mathcal{M}_r$ defined in §2 is $\binom{\min\{m,n\}}{r}$, while Theorem 2.12 says that the ED degree of the variety $O(n)$ is 2^n. The values computed in Theorem 5.1 give the ED degree of the Segre variety $\mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_d}$, while the Cartwright-Sturmfels formula in Theorem 5.2 gives the ED degree of the Veronese variety $v_d(\mathbb{P}^n)$.

Theorem 5.10 ([5, Theorem 5.2, Corollary 8.3]). *Let p be a tensor. There is a canonical bijection between*

- *critical points of the distance from p to rank ≤ 1 tensors*
- *critical points of the distance from p to hyperdeterminant hypersurface.*

Correspondence is $x \mapsto p - x$

In particular, from the 15 critical points for the distance from the $3 \times 3 \times 2$ tensor f defined in Example 4.3 to the variety of rank one matrices, we may recover the 15 critical points for the distance from f to the hyperdeterminant hypersurface. It follows that $\mathrm{Det}(f - p_i) = 0$ for the 15 critical points p_i.

The following result generalizes Theorem 5.10 to any projective variety $\mathcal{X}$.

Theorem 5.11 ([5, Theorem 5.2]). *Let $\mathcal{X} \subset \mathbb{P}^n$ be a projective variety, $p \in \mathbb{P}^n$. There is a canonical bijection between*

- *critical points of the distance from p to $\mathcal{X}$*
- *critical points of the distance from p to the dual variety $\mathcal{X}^*$.*

Correspondence is $x \mapsto p - x$. In particular $\mathrm{EDdegree}(\mathcal{X}) = \mathrm{EDdegree}(\mathcal{X}^*)$

5.4. Higher order SVD

In [3], L. De Lathauwer, B. De Moor, and J. Vandewalle proposed a higher order generalization of SVD. This paper has been quite influential and we sketch this contruction for completeness (in the complex field).

Theorem 5.12 (HOSVD, De Lathauwer, De Moor, Vandewalle, [3]). *A tensor $A \in \mathbb{C}^{n_1+1} \otimes \ldots \otimes \mathbb{C}^{n_d+1}$ can be multiplied in the i-th mode by unitary matrices $U_i \in U(n_i + 1)$ in such a way that the resulting tensor S has the following properties:*

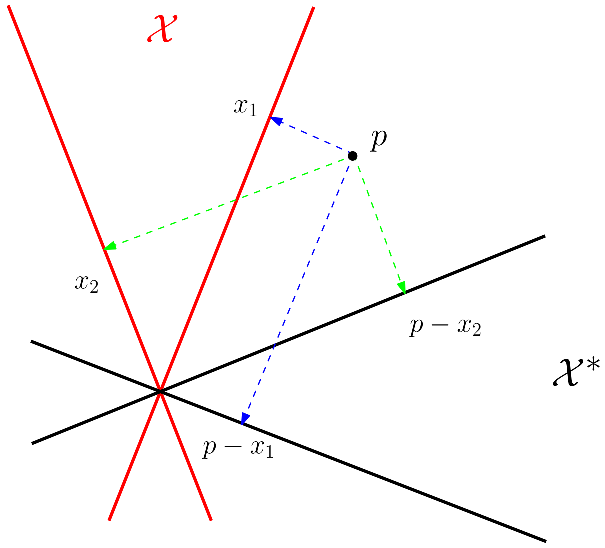

Figure 2: The bijection between critical points on $\mathcal{X}$ and critical points on $\mathcal{X}^*$.

1. *(all-orthogonality) For any $i = 1, \ldots, d$ and $\alpha = 0, \ldots, n_i$ denote by S^i_α the slice in $\mathbb{C}^{n_1+1} \otimes \ldots \widehat{\mathbb{C}^{n_i+1}} \ldots \otimes \mathbb{C}^{n_d+1}$ obtained by fixing the i-index equal to α. Then for $0 \leq \alpha < \beta \leq n_i$ we have $\overline{S^i_\alpha} \cdot S^i_\beta = 0$, that is any two parallel slices are orthogonal according to Hermitian product.*

2. *(ordering) for the Hermitian norm, for all $i = 1, \ldots, d$*

$$\left\|S^i_0\right\| \geq \left\|S^i_1\right\| \geq \ldots \geq \left\|S^i_{n_i}\right\|$$

$\left\|S^i_j\right\|$ are the i-mode singular values and the columns of U_i are the i-mode singular vectors. For $d = 2$, $\left\|S^i_j\right\|$ do not depend on i and we get the classical SVD. This notion has an efficient algorithm computing it. We do not pursue it further because the link with the critical points of the distance is weak, although it can be employed by suitable iterating methods.

References

[1] J. Baaijens and J. Draisma, *Euclidean distance degrees of real algebraic groups*, Linear Algebra Appl. **467** (2015), 174–187.

[2] D. Cartwright and B. Sturmfels, *The number of eigenvectors of a tensor*, Linear Algebra Appl. **438** (2013), 942–952.

[3] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl. **21** (2000), 1253–1278.

[4] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-*1 *and rank-*$(r_1, r_2, \ldots, r_d)$ *approximation of higher order tensors*, SIAM J. Matrix Anal. Appl. **21** (2000), 1324–1342.

[5] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. Thomas, *The euclidean distance degree of an algebraic variety*, to appear in Found. Comput. Math.

[6] S. Friedland and G. Ottaviani, *The number of singular vector tuples and uniqueness of best rank one approximation of tensors*, Found. Comput. Math. **14** (2014), 1209–1242.

[7] I.M. Gelfand, M. Kapranov, and A. Zelevinsky, *Discriminants, resultants and multidimensional determinants*, Birkhäuser, Boston, 1994.

[8] D. Grayson and M. Stillman, *Macaulay2, a software system for research in algebraic geometry*, available at `www.math.uiuc.edu/Macaulay2/`.

[9] J.M. Landsberg, *Tensors: Geometry and applications*, Graduate Studies in Mathematics, no. 128, Birkhäuser, Providence, 2012.

[10] L.-H. Lim, *Singular values and eigenvalues of tensors: a variational approach*, Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP '05, vol. 1, 2005, pp. 129–132.

[11] L. Qi, *Eigenvalues and invariants of tensors*, J. Math. Anal. Appl. **325** (2007), 1363–1377.

[12] G.W. Stewart, *On the early history of the singular value decomposition*, SIAM Review **35** (1993), no. 4, 551–566.

[13] R. Thomas, *Euclidean distance degree*, SIAM News **47** (October 2014), no. 8.

Authors' addresses:

G. Ottaviani
Dipartimento di Matematica e Informatica "U. Dini"
Università di Firenze
viale Morgagni 67/A
50134 Firenze (Italy)
E-mail: `ottavian@math.unifi.it`

R. Paoletti
Dipartimento di Matematica e Informatica "U. Dini"
Università di Firenze
viale Morgagni 67/A
50134 Firenze (Italy)
E-mail: `raffy@math.unifi.it`

Received January 7, 2015
Revised March 20, 2015