Hyperdeterminant and Tensor Rank

Lek-Heng Lim

Workshop on Tensor Decompositions and Applications
CIRM, Luminy, France
August 29–September 2, 2005

Thanks: NSF DMS 01-01364

Collaborators

Pierre Comon
Laboratoire I3S
Université de Nice Sophia Antipolis

Vin de Silva
Department of Mathematics
Stanford University

Matrix Multiplication

Let $f : U \to V$ and $g : V \to W$ be linear maps; U, V, W vector spaces over $\mathbb{R}$ of dimensions n, m, l.

With choice of bases on U, V, W, g, f have matrix representations $A = [a_{ij}] \in \mathbb{R}^{l \times m}, B = [b_{jk}] \in \mathbb{R}^{m \times n}$.

The matrix representation of $h = g \circ f$ (ie. $h(\mathbf{x}) := g(f(\mathbf{x})))$ is then $C = [c_{ik}] \in \mathbb{R}^{l \times n}$ where

$$c_{ik} := \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Similarly for bilinear $g : V_1 \times V_2 \to \mathbb{R}$ and linear $f_1 : U_1 \to V_1, f_2 : U_2 \to V_2$ with matrix representations $A \in \mathbb{R}^{d_1 \times d_2}$, $B_1 \in \mathbb{R}^{d_1 \times s_1}$, $B_2 \in \mathbb{R}^{d_2 \times s_2}$.

The composite map h, where $h(\mathbf{x}, \mathbf{y}) := g(f_1(\mathbf{x}), f_2(\mathbf{y}))$, has matrix representation

$$C = B_2^\top A B_1 \in \mathbb{R}^{s_1 \times s_2}.$$

Multilinear Matrix Multiplication

Do the same for multilinear map $g : V_1 \times \cdots \times V_k \to \mathbb{R}$ and linear maps $f_1 : U_1 \to V_1, \dots, f_k : U_k \to V_k$; $\dim(V_i) = s_i, \dim(U_i) = d_i$.

With choice of bases on V_i's and U_i's, g is represented by $A = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ and $f_1, \dots, f_k$ by $M_1 = [m^1_{j_1 i_1}] \in \mathbb{R}^{d_1 \times s_1}, \dots, M_k = [m^k_{j_k i_k}] \in \mathbb{R}^{d_k \times s_k}$.

If we compose g by $f_1, \dots, f_k$ to get $h : U_1 \times \cdots \times U_k \to \mathbb{R}$ defined by

$$h(\mathbf{x}_1, \dots, \mathbf{x}_k) = g(f(\mathbf{x}_1), \dots, f(\mathbf{x}_k)),$$

then h is represented by $[\![c_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{s_1 \times \cdots \times s_k}$ where

$$c_{i_1 \cdots i_k} := \textstyle\sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} a_{j_1 \cdots j_k} m^1_{j_1 i_1} \cdots m^k_{j_k i_k}. \tag{1}$$

The *covariant multilinear matrix multiplication* will be written

$$A(M_1, \dots, M_k) := [\![c_{i_1 \cdots i_k}]\!] \in \mathbb{R}^{s_1 \times \cdots \times s_k}.$$

Contravariant Version

The *contravariant multilinear matrix multiplication* of $[\![a_{j_1\cdots j_k}]\!] \in \mathbb{R}^{d_1\times\cdots\times d_k}$ by matrices $L_1 = [\ell^1_{i_1 j_1}] \in \mathbb{R}^{r_1\times d_1}, \dots, L_k = [\ell^k_{i_k j_k}] \in \mathbb{R}^{r_k\times d_k}$ is defined by

$$(L_1, \dots, L_k)A = [\![b_{i_1\cdots i_k}]\!] \in \mathbb{R}^{r_1\times\cdots\times r_k},$$
$$b_{i_1\cdots i_k} := \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} \ell^1_{i_1 j_1} \cdots \ell^k_{i_k j_k} a_{j_1\cdots j_k}. \qquad (2)$$

This comes from the composition of a multilinear map $g : V_1^* \times \cdots \times V_k^* \to \mathbb{R}$ by linear maps $f_1 : V_1 \to U_1, \dots, f_k : V_k \to U_k$.

Simple relation if we disregard covariance/contravariance:

$$(L_1, \dots, L_k)A = A(L_1^\top, \dots, L_k^\top)$$
$$A(M_1, \dots, M_k) = (M_1^\top, \dots, M_k^\top)A.$$

Works over $\mathbb{C}$ too (replace $L_i^\top$ by $L_i^\dagger$).

Properties

- Let $A, B \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ and $\lambda, \mu \in \mathbb{R}$. Let $L_1 \in \mathbb{R}^{r_1 \times d_1}, \ldots, L_k \in \mathbb{R}^{r_k \times d_k}$. Then

$$(L_1, \ldots, L_k)(\lambda A + \mu B) = \lambda (L_1, \ldots, L_k) A + \mu (L_1, \ldots, L_k) B.$$

- Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$. Let $L_1 \in \mathbb{R}^{r_1 \times d_1}, \ldots, L_k \in \mathbb{R}^{r_k \times d_k}$, and $M_1 \in \mathbb{R}^{s_1 \times r_1}, \ldots, M_k \in \mathbb{R}^{s_k \times r_k}$. Then

$$(M_1, \ldots, M_k)(L_1, \ldots, L_k) A = (M_1 L_1, \ldots, M_k L_k) A$$

  where $M_i L_i \in \mathbb{R}^{s_i \times d_i}$ is simply the matrix-matrix product of M_i and L_i.

- Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ and $\lambda, \mu \in \mathbb{R}$. Let $L_1 \in \mathbb{R}^{r_1 \times d_1}, \ldots, L_j, M_j \in \mathbb{R}^{r_j \times d_j}, \ldots, L_k \in \mathbb{R}^{r_k \times d_k}$. Then

$$A(L_1, \ldots, \lambda L_j + \mu M_j, \ldots, L_k) = \lambda (L_1, \ldots, L_j, \ldots, L_k) A + \mu (L_1, \ldots, M_j, \ldots, L_k) A.$$

Aside: Relation with Kronecker Product

Forgetful map $\mathbb{R}^{d_1 \times \cdots \times d_k} \to \mathbb{R}^{d_1 \cdots d_k}$, $A \mapsto \text{vec}(A)$ ('forgets' the multilinear structure), then

$$\text{vec}((L_1, \ldots, L_k)A) = L_1 \otimes \cdots \otimes L_k \, \text{vec}(A).$$

where $L_1 \otimes \cdots \otimes L_k \in \mathbb{R}^{d_1 \cdots d_k \times d_1 \cdots d_k}$ is the Kronecker product of $L_1, \ldots, L_k$.

Matrix Techniques

Start with $\mathbb{R}^{m\times n\times l}$ and $\mathbb{C}^{m\times n\times l}$. $l = 2$ is well understood, may be regarded as pairs of matrices $(A, B) \in (\mathbb{C}^{m\times n})^2$ or $(\mathbb{R}^{m\times n})^2$, or equivalently, as a matrix pencil $\lambda A + \mu B \in \mathbb{C}[\lambda, \mu]^{m\times n}$ or $\mathbb{R}[\lambda, \mu]^{m\times n}$.

Kronecker-Weierstrass Theory. There exist $S \in \mathrm{GL}(m), T \in \mathrm{GL}(n)$ such that (SAT, SBT) can be decomposed into block pairs of the following forms

$$\left(\begin{bmatrix}1 & & & \\ 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0\end{bmatrix}, \begin{bmatrix}0 & & & \\ 1 & 0 & & \\ & 1 & \ddots & \\ & & \ddots & 0 \\ & & & 1\end{bmatrix}\right) \in \mathbb{R}^{(p+1)\times p},$$

$$\left(\begin{bmatrix}1 & 0 & & & \\ & 1 & 0 & & \\ & & \ddots & \ddots & \\ & & & 1 & 0\end{bmatrix}, \begin{bmatrix}0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1\end{bmatrix}\right) \in \mathbb{R}^{q\times (q+1)},$$

$$\left(\begin{bmatrix}1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{bmatrix}, \begin{bmatrix}0 & & & -a_0 \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & -a_{r-2} \\ & & 1 & -a_{r-1}\end{bmatrix}\right) \in \mathbb{R}^{r\times r}.$$

Likewise for $\mathbb{C}$. Similar but simpler results obtained by Jos ten Berge for generic pairs.

Larger Sizes and Higher Orders

Want to obtain results as general as possible — for tensors of arbitrary size and order over both $\mathbb{R}$ and $\mathbb{C}$. For larger values of k or $d_1, \ldots, d_k$, techniques relying on multilinear matrix multiplications become increasingly less effective.

Inherent limitation:

$$\dim(\mathbb{R}^{d_1 \times \cdots \times d_k}) = d_1 \cdot \cdots \cdot d_k = O(d^k)$$

while

$$\dim(\mathrm{GL}(d_1) \times \cdots \times \mathrm{GL}(d_k)) = d_1^2 + \cdots + d_k^2 = O(kd^2)$$

and

$$\dim(\mathrm{O}(d_1) \times \cdots \times \mathrm{O}(d_k)) = d_1(d_1 - 1)/2 + \cdots + d_k(d_k - 1)/2 = O(kd^2).$$

The action of $\mathrm{GL}(d_1) \times \cdots \times \mathrm{GL}(d_k)$ on $\mathbb{R}^{d_1 \times \cdots \times d_k}$ has uncountably many orbits,

$$\{(L_1, \ldots, L_k)A \mid (L_1, \ldots, L_k) \in \mathrm{GL}(d_1) \times \cdots \times \mathrm{GL}(d_k)\},$$

as soon as $d_i > 2$, $k > 4$.

Multilinear Functional and its Gradient

Multilinear functional associated with $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, ie.

$$f_A : \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} \to \mathbb{R}, \qquad (3)$$
$$(\mathbf{x}_1, \ldots, \mathbf{x}_k) \mapsto \sum_{j_1=1}^{d_1} \cdots \sum_{j_k=1}^{d_k} a_{j_1 \cdots j_k} x^1_{j_1} \cdots x^k_{j_k},$$

can be written as

$$f_A(\mathbf{x}_1, \ldots, \mathbf{x}_k) = A(\mathbf{x}_1, \ldots, \mathbf{x}_k) \qquad (4)$$

where the rhs is the right multilinear multiplication by $\mathbf{x}_i = (x^i_1, \ldots, x^i_{d_i})^\top$, regarded as a $d_i \times 1$ matrix.

Gradient of f_A may be written as

$$\nabla f_A = (\nabla_{\mathbf{x}_1} f_A, \ldots, \nabla_{\mathbf{x}_k} f_A)$$

where

$$\nabla_{\mathbf{x}^i} f_A(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \left(\frac{\partial f_A}{\partial x^i_1}, \ldots, \frac{\partial f_A}{\partial x^i_{d_i}} \right) = A(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, I_{d_i}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_k).$$

I_{d_i} denotes $d_i \times d_i$ identity matrix.

Hyperdeterminant

Work in $\mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)}$ for the time being ($d_i \geq 1$). Consider

$$S := \{A \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \nabla f_A(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \mathbf{0} \text{ for some non-zero } (\mathbf{x}_1, \ldots, \mathbf{x}_k)\}.$$

Theorem (Gelfand, Kapranov, Zelevinsky, 1992). S is a hypersurface if and only if

$$d_j \leq \sum_{i \neq j} d_i$$

for all $j = 1, \ldots, k$. Let Δ be the equation of the hypersurface, ie. a multivariate polynomial in the entries of A such that

$$S = \{A \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \Delta(A) = 0\}.$$

Then Δ may be chosen to have integer coefficients.

For $\mathbb{C}^{m\times n}$, the condition becomes $m \leq n$ and $n \leq m$ — that's why matrix determinants is only defined for square matrices.

Since Δ has integer coefficients, $\Delta(A)$ is real-valued for $A \in \mathbb{R}^{(d_1+1)\times\cdots\times(d_k+1)}$.

Geometric View

Let $X = \{\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_k \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \mathbf{x}_i \in \mathbb{C}^{d_i+1}\}$ be the (smooth) manifold of decomposable tensors (X oftened called the Segre variety).

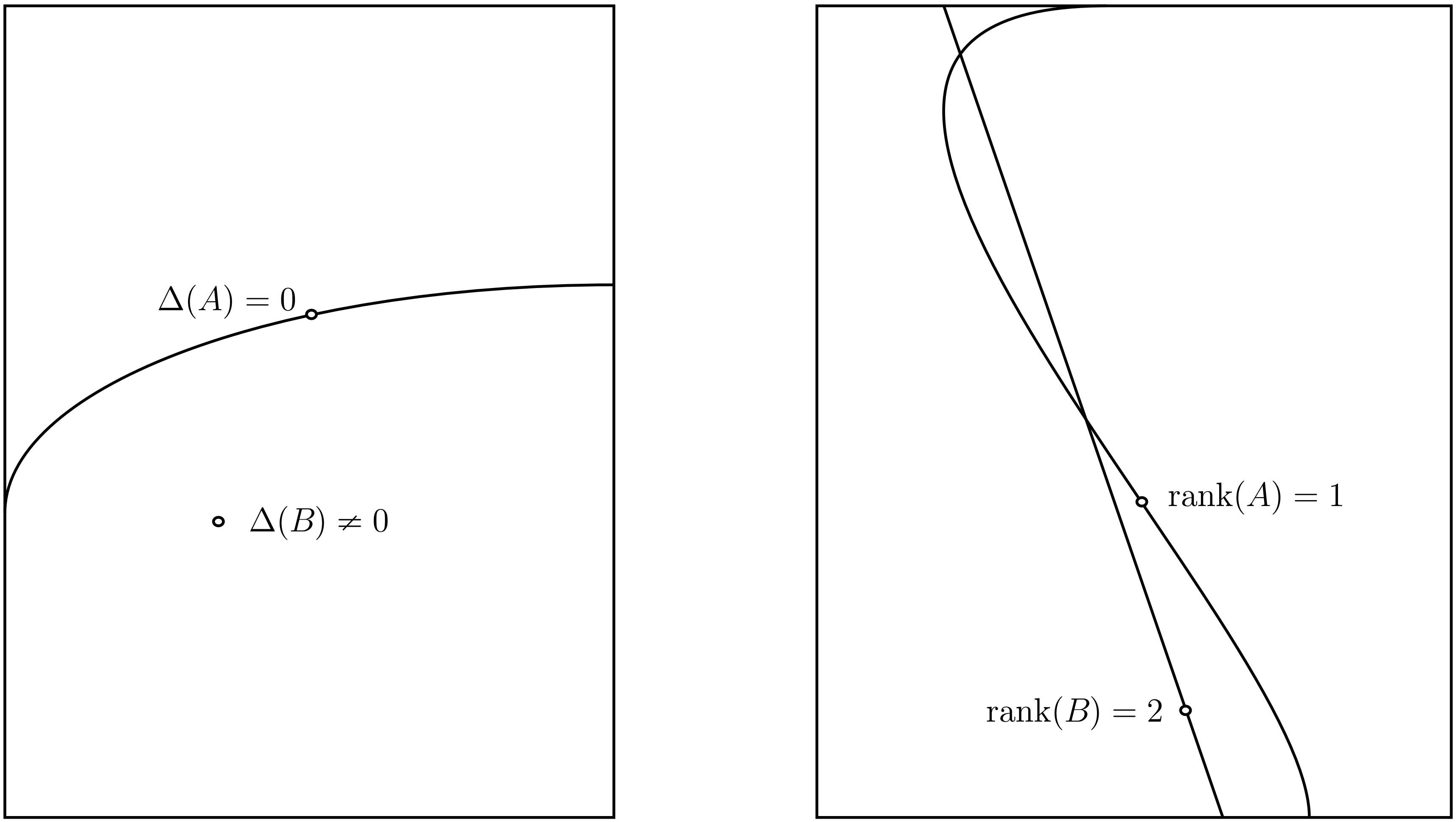
Let $A \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)}$. Then the condition $\nabla f_A(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \mathbf{0}$ for some non-zero $(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is equivalent to saying that the hyperplane orthogonal to A, ie.

$$H_A := \{B \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \langle A, B\rangle = 0\}$$

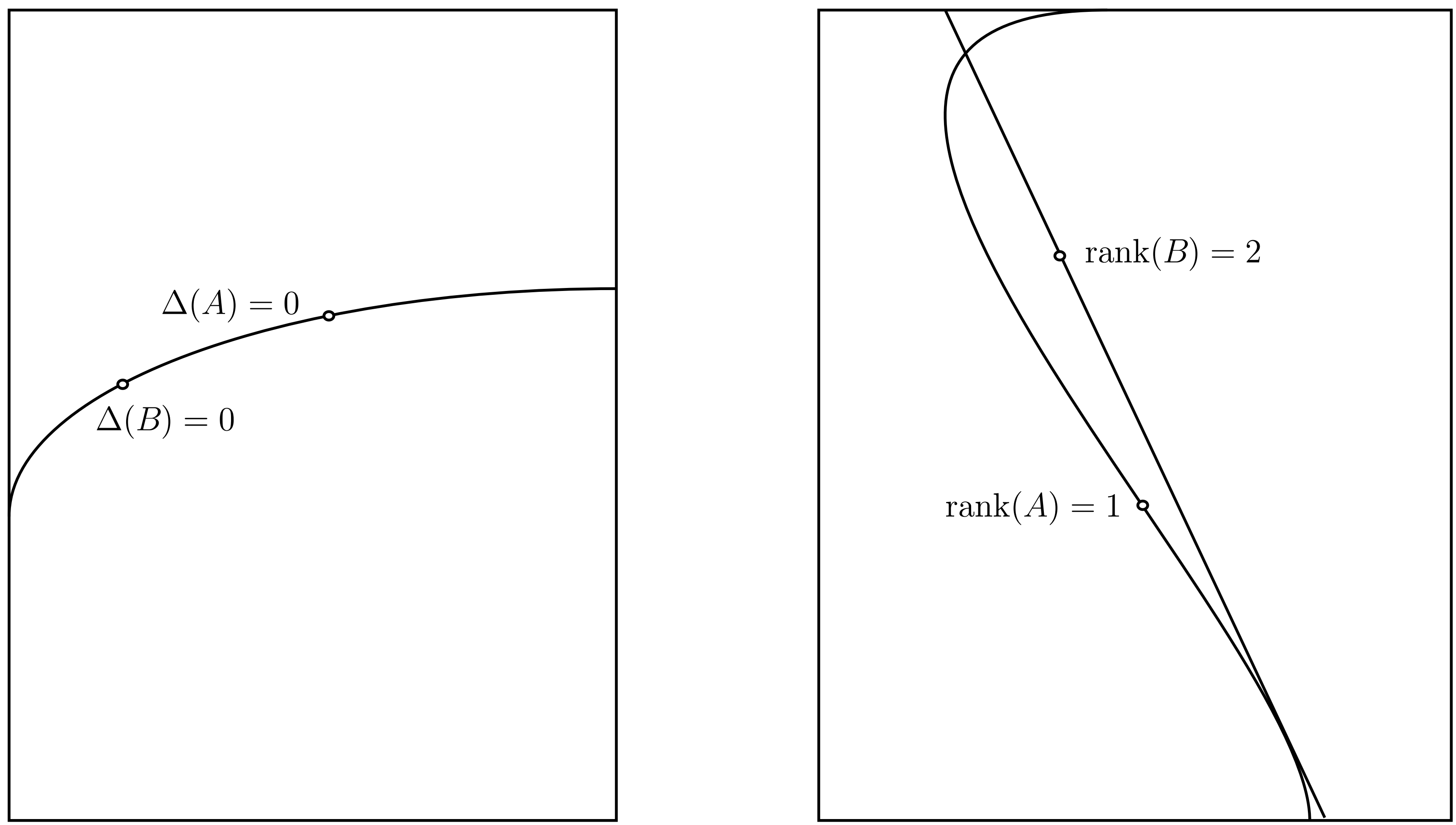
contains a tangent to X at the point $\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_k$. This may also be taken as an alternative definition of the hyperdeterminant $\Delta(A)$.

Projective duality:

$$X^* = S.$$



$\Delta(A) = 0$
$\Delta(B) \neq 0$
$\mathrm{rank}(A) = 1$
$\mathrm{rank}(B) = 2$



$\Delta(A) = 0$
$\Delta(B) = 0$
$\mathrm{rank}(B) = 2$
$\mathrm{rank}(A) = 1$

Minor Inaccuracy

Should really be working in projective spaces $P(\mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)}) = \mathbb{P}^{(d_1+1)\cdots(d_k+1)-1}$. This is the set of equivalence classes

$$[A] := \{\lambda A \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \lambda \in \mathbb{C}^\times\}.$$

Thing to note is that the for any $A \in \mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)}$ and $\lambda \in \mathbb{C}^\times$,

$$\operatorname{rank}_\otimes(\lambda A) = \operatorname{rank}_\otimes(A).$$

So outer-product rank is well-defined in $P(\mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)})$, ie. given $[A] \in P(\mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)})$ define

$$\operatorname{rank}_\otimes([A]) = \operatorname{rank}_\otimes(A)$$

for any $A \in [A]$.

Examples

A. Cayley, "On the theory of linear transformation," *Cambridge Math. J.*, **4** (1845), pp. 193–209.

Hyperdeterminant of $A = [\![a_{ijk}]\!] \in \mathbb{R}^{2\times 2\times 2}$ is

$$\Delta(A) = \frac{1}{4}\left[\det\left(\begin{bmatrix} a_{000} & a_{010} \\ a_{001} & a_{011} \end{bmatrix} + \begin{bmatrix} a_{100} & a_{110} \\ a_{101} & a_{111} \end{bmatrix}\right) - \det\left(\begin{bmatrix} a_{000} & a_{010} \\ a_{001} & a_{011} \end{bmatrix} - \begin{bmatrix} a_{100} & a_{110} \\ a_{101} & a_{111} \end{bmatrix}\right)\right]^2 - 4\det\begin{bmatrix} a_{000} & a_{010} \\ a_{001} & a_{011} \end{bmatrix}\det\begin{bmatrix} a_{100} & a_{110} \\ a_{101} & a_{111} \end{bmatrix}.$$

A result that parallels the matrix case is the following: the system of bilinear equations

$$\begin{aligned}
a_{000}x_0y_0 + a_{010}x_0y_1 + a_{100}x_1y_0 + a_{110}x_1y_1 &= 0,\\
a_{001}x_0y_0 + a_{011}x_0y_1 + a_{101}x_1y_0 + a_{111}x_1y_1 &= 0,\\
a_{000}x_0z_0 + a_{001}x_0z_1 + a_{100}x_1z_0 + a_{101}x_1z_1 &= 0,\\
a_{010}x_0z_0 + a_{011}x_0z_1 + a_{110}x_1z_0 + a_{111}x_1z_1 &= 0,\\
a_{000}y_0z_0 + a_{001}y_0z_1 + a_{010}y_1z_0 + a_{011}y_1z_1 &= 0,\\
a_{100}y_0z_0 + a_{101}y_0z_1 + a_{110}y_1z_0 + a_{111}y_1z_1 &= 0,
\end{aligned}$$

has a non-trivial solution iff $\Delta(A) = 0$.

Examples

Hyperdeterminant of $A = [\![a_{ijk}]\!] \in \mathbb{R}^{2\times2\times3}$ is

$$\Delta(A) = \det \begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{100} & a_{101} & a_{102} \\ a_{010} & a_{011} & a_{012} \end{bmatrix} \det \begin{bmatrix} a_{100} & a_{101} & a_{102} \\ a_{010} & a_{011} & a_{012} \\ a_{110} & a_{111} & a_{112} \end{bmatrix}$$
$$- \det \begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{100} & a_{101} & a_{102} \\ a_{110} & a_{111} & a_{112} \end{bmatrix} \det \begin{bmatrix} a_{000} & a_{001} & a_{002} \\ a_{010} & a_{011} & a_{012} \\ a_{110} & a_{111} & a_{112} \end{bmatrix}$$

Again, the following is true:

$$\begin{aligned}
a_{000}x_0y_0 + a_{010}x_0y_1 + a_{100}x_1y_0 + a_{110}x_1y_1 &= 0, \\
a_{001}x_0y_0 + a_{011}x_0y_1 + a_{101}x_1y_0 + a_{111}x_1y_1 &= 0, \\
a_{002}x_0y_0 + a_{012}x_0y_1 + a_{102}x_1y_0 + a_{112}x_1y_1 &= 0, \\
a_{000}x_0z_0 + a_{001}x_0z_1 + a_{002}x_0z_2 + a_{100}x_1z_0 + a_{101}x_1z_1 + a_{102}x_1z_2 &= 0, \\
a_{010}x_0z_0 + a_{011}x_0z_1 + a_{012}x_0z_2 + a_{110}x_1z_0 + a_{111}x_1z_1 + a_{112}x_1z_2 &= 0, \\
a_{000}y_0z_0 + a_{001}y_0z_1 + a_{002}y_0z_2 + a_{010}y_1z_0 + a_{011}y_1z_1 + a_{012}y_1z_2 &= 0, \\
a_{100}y_0z_0 + a_{101}y_0z_1 + a_{102}y_0z_2 + a_{110}y_1z_0 + a_{111}y_1z_1 + a_{112}y_1z_2 &= 0,
\end{aligned}$$

has a non-trivial solution iff $\Delta(A) = 0$.

For more examples, see:

I.M. Gelfand, M.M. Kapranov, and A.V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Birkhäuser Publishing, Boston, MA, 1994.

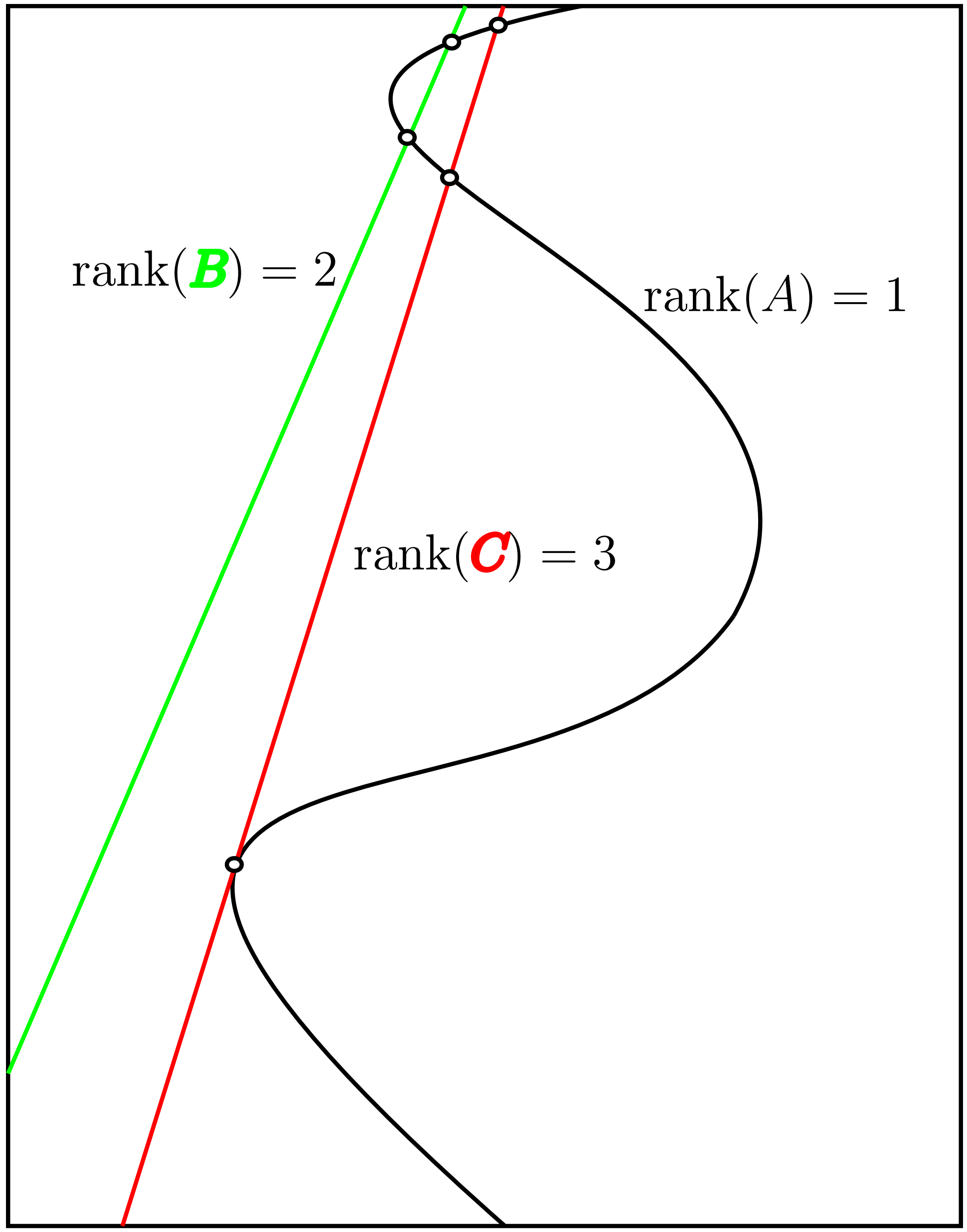
Another Explanation for Degeneracy

Degeneracy here means: a sequence of tensors B_n of rank $\leq r$ converging to a tensor A of rank $r+1$, ie. A can be approximated arbitrarily well by tensors of lower-rank. In particular, A has no best rank-r approximations.

Question: Why do degeneracy occur in PARAFAC?

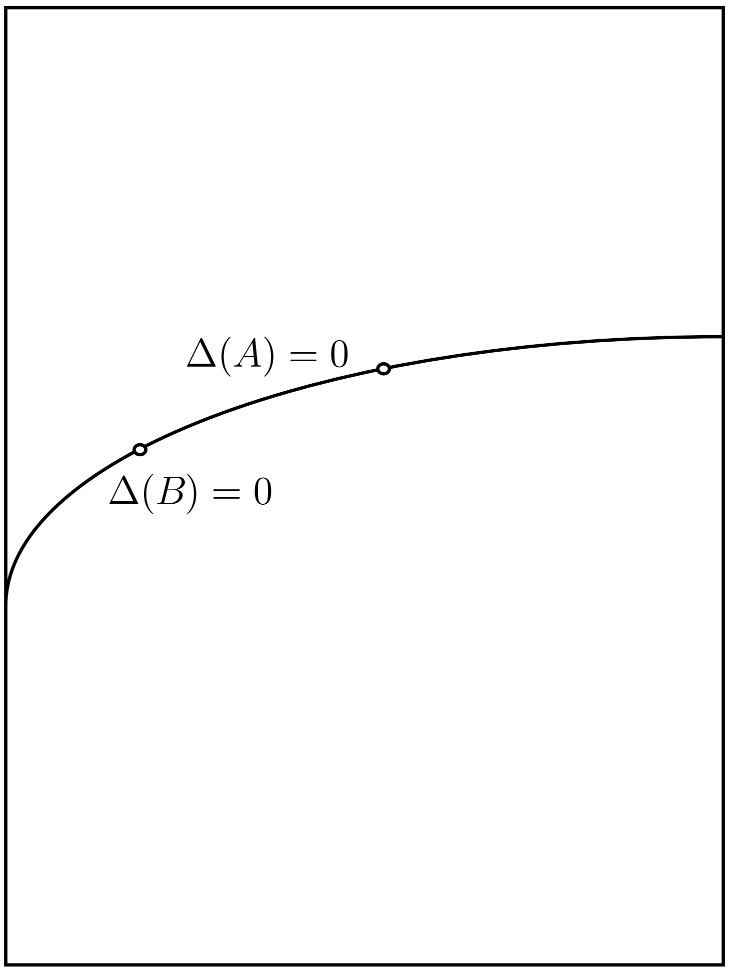
One Explanation: Alwin's talk

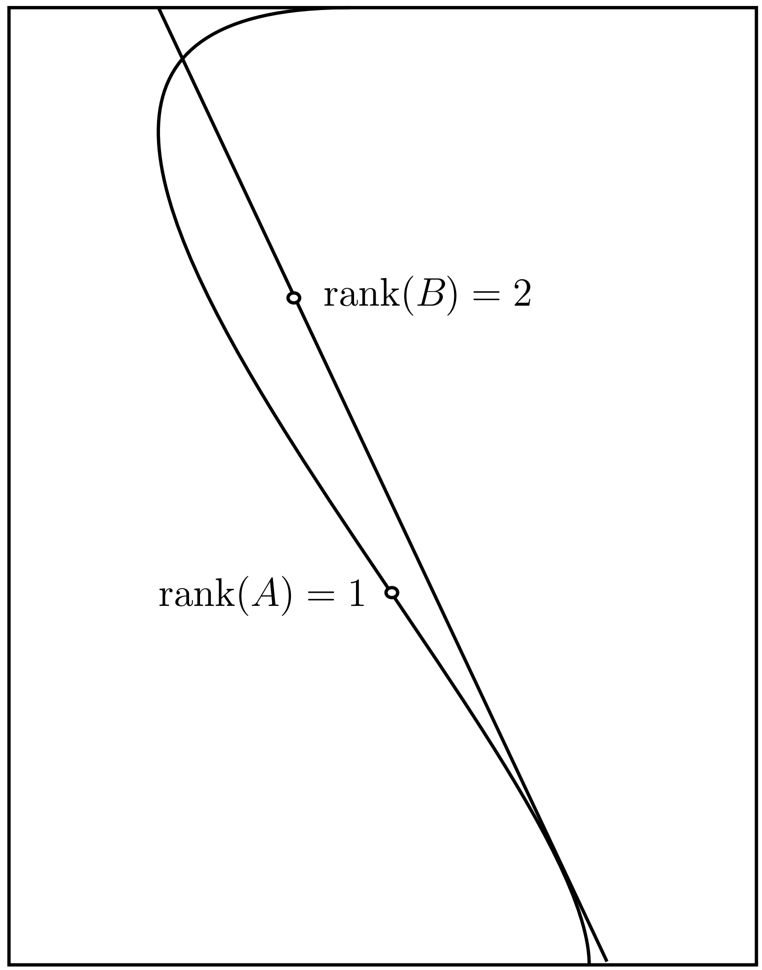
Alternative explanation: Degeneracy happens when a sequence of r-secants to X converges to a $(r+1)$-secant. Furthermore, this $(r+1)$-secant is always tangent to X (since X is a smooth manifold). This doesn't happen for order-2 tensors because the geometry of $X = \{A \in \mathbb{C}^{m\times n} \mid \text{rank}(A) \leq 1\}$ prevents this.

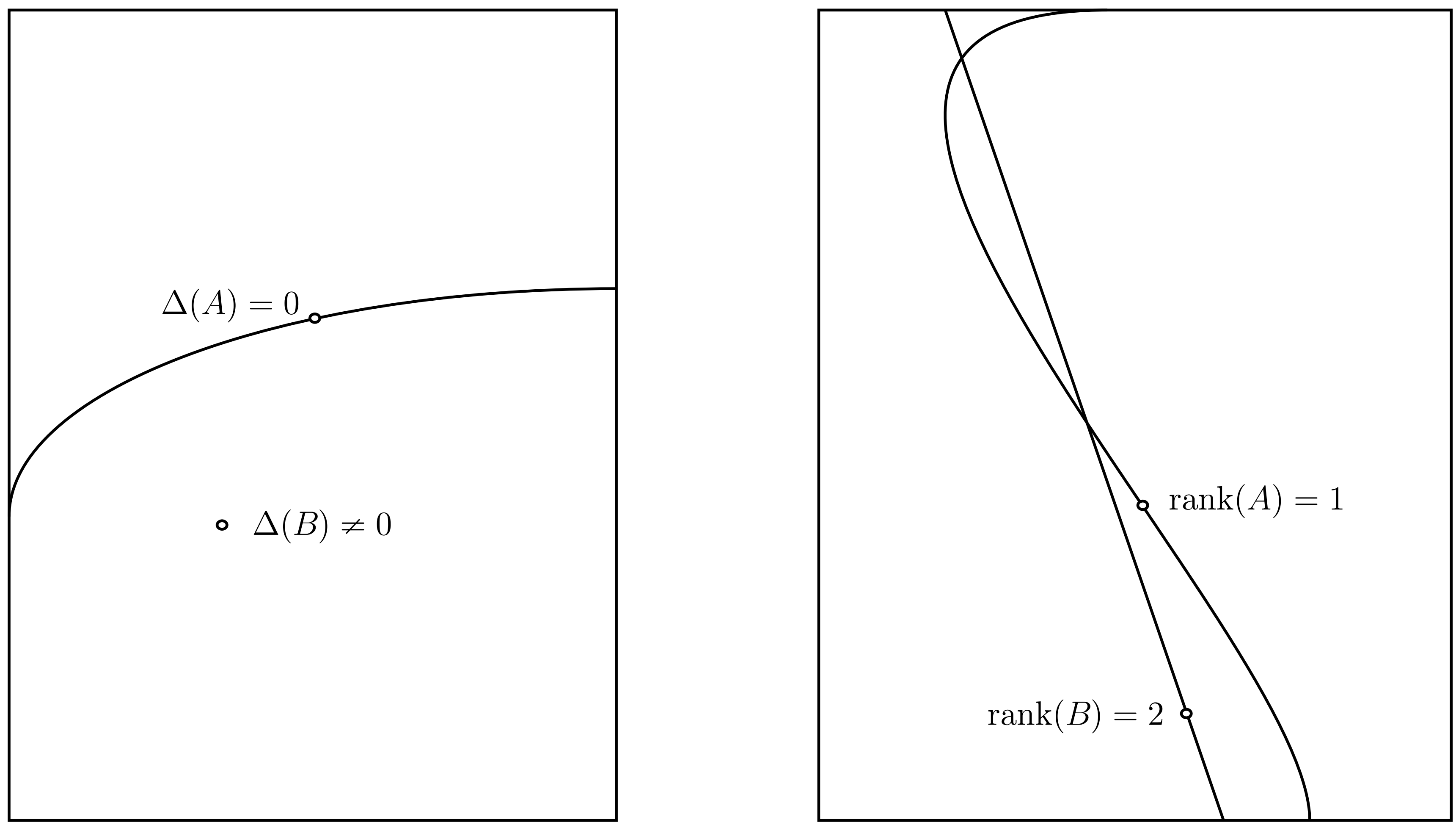


rank($\boldsymbol{B}$) = 2
rank(A) = 1
rank($\boldsymbol{C}$) = 3

Relation between the two explanations







$\Delta(A) = 0$
$\Delta(B) \neq 0$
$\text{rank}(A) = 1$
$\text{rank}(B) = 2$

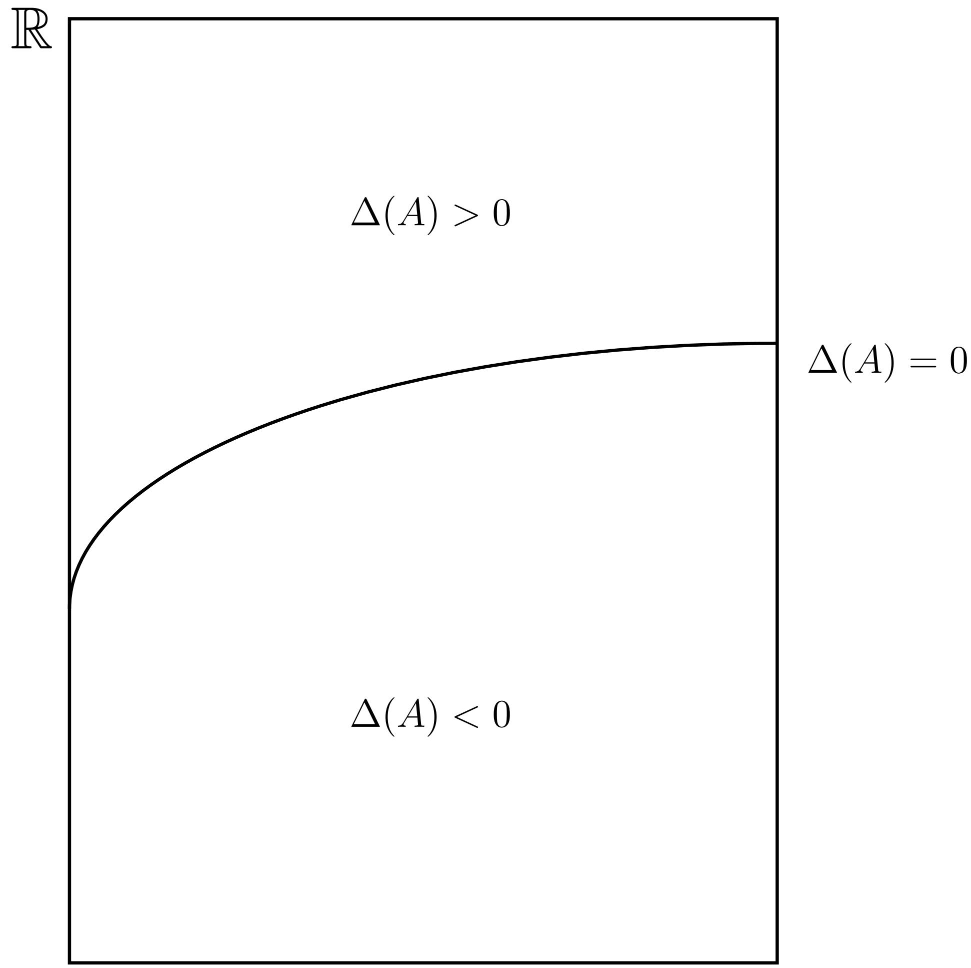
Typical Rank(s) for Real Tensors

Conjecture. For any $d_1, \ldots, d_k \geq 1$ satisfying

$$d_j \leq \sum_{i \neq j} d_i$$

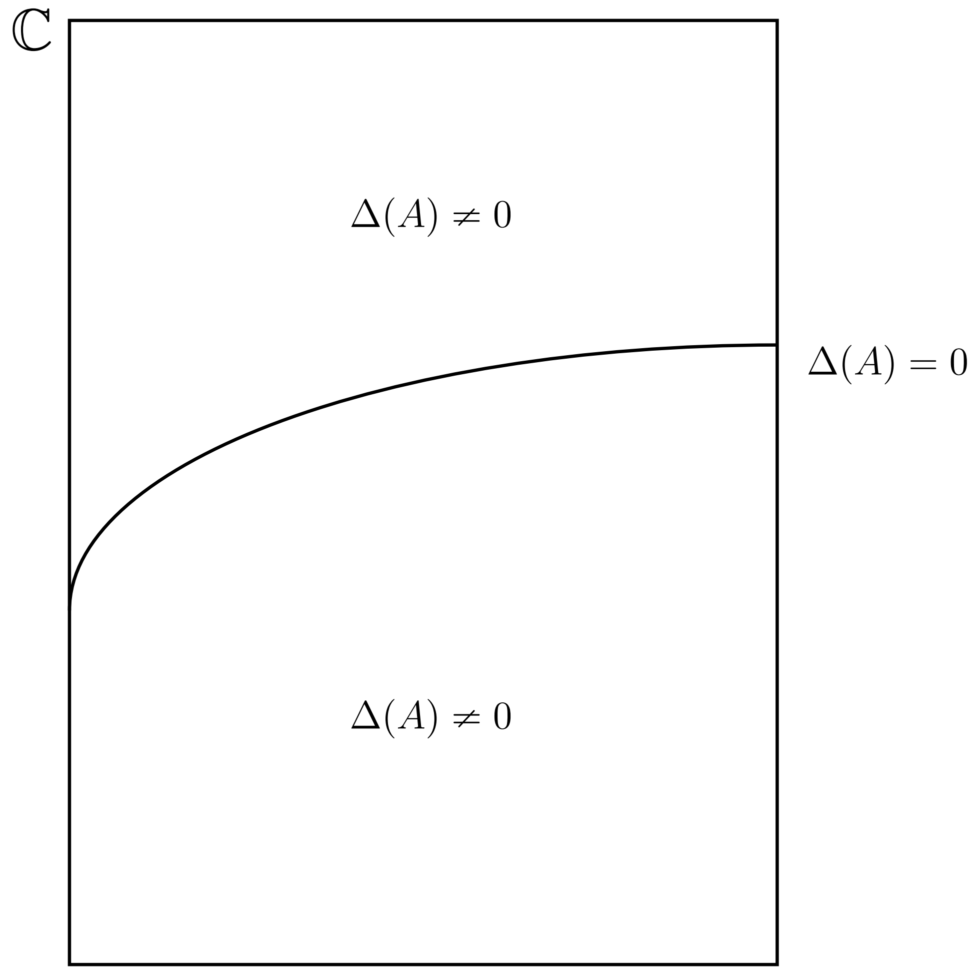
for all j, the outer-product rank is constant on the sets $\{A \in \mathbb{R}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \Delta(A) < 0\}$ and $\{A \in \mathbb{R}^{(d_1+1)\times\cdots\times(d_k+1)} \mid \Delta(A) > 0\}$ (but may take different values).

Corollary. If the conjecture is true, then there exist at most two typical ranks for $\mathbb{R}^{(d_1+1)\times\cdots\times(d_k+1)}$.


$\mathbb{R}$
$\Delta(A) > 0$
$\Delta(A) = 0$
$\Delta(A) < 0$

Generic rank for Complex Tensors

Theorem. For any $d_1, \ldots, d_k \geq 1$, a (unique) generic rank always exist for $\mathbb{C}^{(d_1+1)\times\cdots\times(d_k+1)}$.



Extending hyperdeterminant

Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, $d_i \geq r = \text{rank}_{\otimes}(A)$. Let Δ denote the hyperdeterminant in $\mathbb{R}^{r \times \cdots \times r}$.

Easy to show: There exist $Q_i \in \mathbb{R}^{r \times d_i}$ with orthonormal columns, $i = 1, \ldots, k$, such that

$$(Q_1, \ldots, Q_k)A \in \mathbb{R}^{r \times \cdots \times r}.$$

Lemma. If A can be approximated arbitrarily closely by tensors of rank $\leq r$ in $\mathbb{R}^{d_1 \times \cdots \times d_k}$, then $(Q_1, \ldots, Q_k)A$ can be approximated arbitrarily closely by tensors of rank $\leq r$ in $\mathbb{R}^{r_1 \times \cdots \times r_k}$.

Define r-hyperdeterminant of A by

$$\Delta_{k,r}(A) := \Delta((Q_1, \ldots, Q_k)A).$$

Whether $\Delta_{k,r}(A) = 0$ or not is independent of choice of $Q_1, \ldots, Q_k$.

Conditioning

J.W. Demmel, "On condition numbers and the distance to the nearest ill-posed problem," *Numer. Math.*, **51** (1987), no. 3, pp. 251–289.

J.W. Demmel, "The geometry of ill-conditioning," *J. Complexity*, **3** (1987), no. 2, pp. 201–229.

J.W. Demmel, "The probability that a numerical analysis problem is difficult," *Math. Comp.*, **50** (1988), no. 182, pp. 449–480.

An ill-posed problem is one that lacks either existence or uniqueness or stability (of a solution).

Condition number in a nutshell: The condition number of a well-posed problem measures the distance of that problem to the manifold of ill-posed problems. The larger the condition number, the closer the problem is to an ill-posed one.

Example. The manifold of ill-posed $n \times n$ linear system is $\mathcal{S} = \{A \in \mathbb{R}^{n\times n} \mid \det(A) = 0\}$. For a non-singular B, the (normalized) condition number

$$\frac{1}{\|B^{-1}\|}$$

is the distance of B to $\mathcal{S}$. Note that the value of $\det(B)$ does not enter into the picture — $\det(B)$ can be arbitrarily small for well-conditioned B.

Using Hyperdeterminants

The set $\{A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \mid \Delta_{k,r}(A) = 0\}$ characterizes the ill-posed best rank-r approximation problems in $\mathbb{R}^{d_1 \times \cdots \times d_k}$.

The (normalized) condition number of the problem of finding the best rank-r approximation to B is given by the reciprocal of the distance of B to $\{A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \mid \Delta_{k,r}(A) = 0\}$.

Note again that the exact value of $\Delta_{k,r}(A)$ is unimportant (only whether it is 0).

Example: Order-3, Rank-2

Easy Fact. Let $l, m, n \geq 2$. Let $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{R}^l$, $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{R}^m$, $\mathbf{x}_3, \mathbf{y}_3 \in \mathbb{R}^n$ and define

$$A := \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{y}_3 + \mathbf{x}_1 \otimes \mathbf{y}_2 \otimes \mathbf{x}_3 + \mathbf{y}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3 \in \mathbb{R}^{l \times m \times n}.$$

Then $\operatorname{rank}_{\otimes}(A) = 3$ if and only if $\mathbf{x}_i, \mathbf{y}_i$ are linearly independent, $i = 1, 2, 3$.

Theorem (de Silva and L., 2004). Let $l, m, n \geq 2$. Let $A \in \mathbb{R}^{l \times m \times n}$ with $\operatorname{rank}_{\otimes}(A) \geq 3$. The following are equivalent:

1. A is the limit of a sequence $B_n \in \mathbb{R}^{l \times m \times n}$ with $\operatorname{rank}_{\otimes}(B_n) \leq 2$,
2. $\Delta_{3,2}(A) = 0$,
3. $A = \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{y}_3 + \mathbf{x}_1 \otimes \mathbf{y}_2 \otimes \mathbf{x}_3 + \mathbf{y}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$.