Computing the Best Rank$-(r_1, r_2, r_3)$ Approximation of a Tensor

Lars Eldén
Department of Mathematics
Linköping University
Sweden
Joint work with Berkant Savas

Harrachov 2007

Best rank$-k$ approximation of a matrix

Assume $X_k^T X_k = I$ and $Y_k^T Y_k = I$

$$\min_{X_k, Y_k, S_k} \|A - X_k S_k Y_k^T\|_F =: \min_{X_k, Y_k, S_k} \|A - (X_k, Y_k) \cdot S_k\|_F$$

(Almost) equivalent problem:

$$\max_{X_k, Y_k} \|X_k^T A Y_k\|_F = \max_{X_k, Y_k} \|A \cdot (X_k, Y_k)\|_F$$

Solution by SVD: $X_k S_k Y_k^T = U_k \Sigma_k V_k^T = (U_k, V_k) \cdot \Sigma_k$

Eckart-Young property

Sketch of "proof":

Determine u_1 and v_1 $(k = 1)$

Put u_1 and v_1 in orthogonal matrices $(u_1 \; \widehat{U})$ and $(v_1 \; \widehat{V})$

$$(u_1 \; \widehat{U})^T A (v_1 \; \widehat{v}) = \begin{pmatrix} \sigma_1 & 0 \\ 0 & B \end{pmatrix}$$

Optimality	$\Longrightarrow$	zeros	$\Longrightarrow$	deflation: continue with B

Orthogonality of vectors comes automatically

Number of degrees of freedom in U_k and V_k is equal to the number of zeros produced.

Best rank$-(k, k, k)$ approximation of a tensor

Assume $X^T X = Y^T Y = Z^T Z = I_k$

$$\min_{X,Y,Z,\mathcal{S}} \|\mathcal{A} - (X,Y,Z)\cdot\mathcal{S}\|_F \qquad \Longleftrightarrow \qquad \max_{X,Y,Z} \|\mathcal{A}\cdot(X,Y,Z)\|_F$$

Why is this problem much more complicated?

Not enough degrees of freedom in X, Y, Z to zero many $(O(k^3) + O(kn^2))$ elements in $\mathcal{A}$

$$\Downarrow$$

Deflation is impossible in general

$$\Downarrow$$

Orthogonality constraints must be enforced

Talk outline

- Some basic tensor concepts (For simplicity: only tensors of order 3)
- Best rank-(r_1, r_2, r_3) approximation problem
- Optimization on the Grassmann manifold
- Newton-Grassmann for solving the best rank-(r_1, r_2, r_3) approximation problem
- Numerical examples
- Ongoing work

“Old and New” Research Area

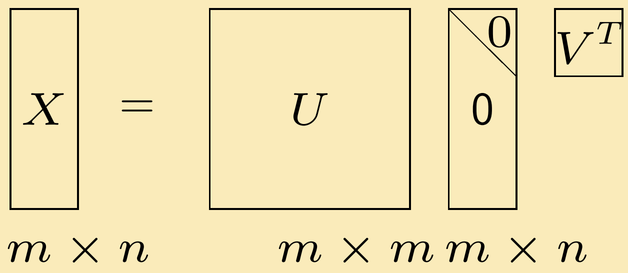
- Tensor methods have been used since the 1960’s in psychometrics and chemometrics! Only recently in numerical community.

- Available mathematical theory deals very little with computational aspects. Many fundamental mathematical problems are open!

- Applications in signal processing and various areas of data mining.

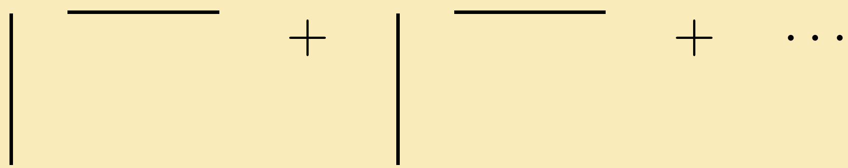
Two aspects of SVD

Singular Value Decomposition: $\mathbb{R}^{m\times n} X = U\Sigma V^T$



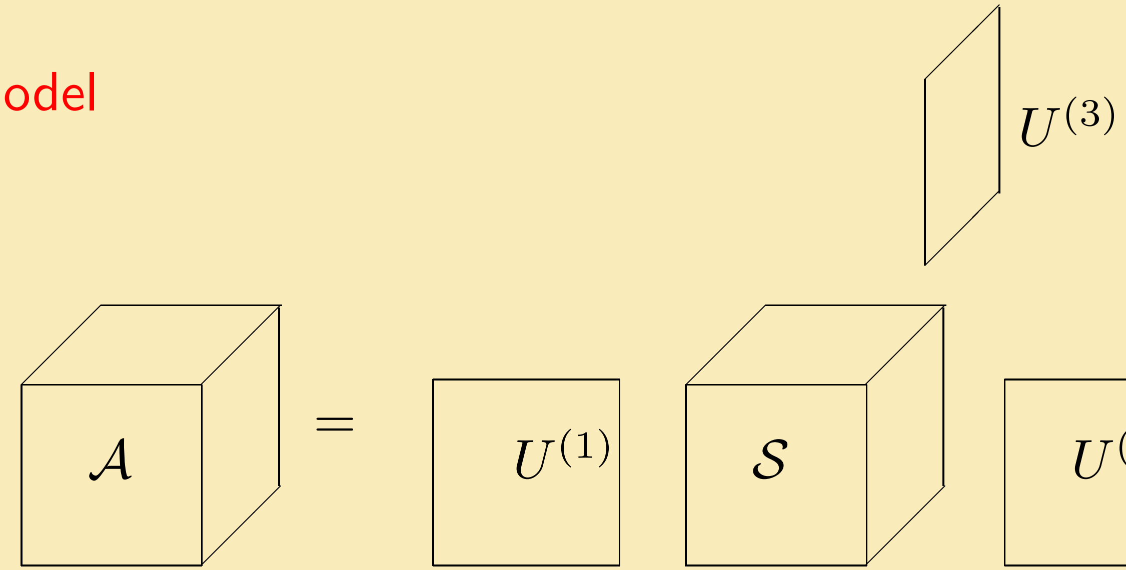
Singular value expansion: sum of rank-1 matrices:

$$X = \sum_{i=1}^{n} \sigma_i u_i v_i^T =$$



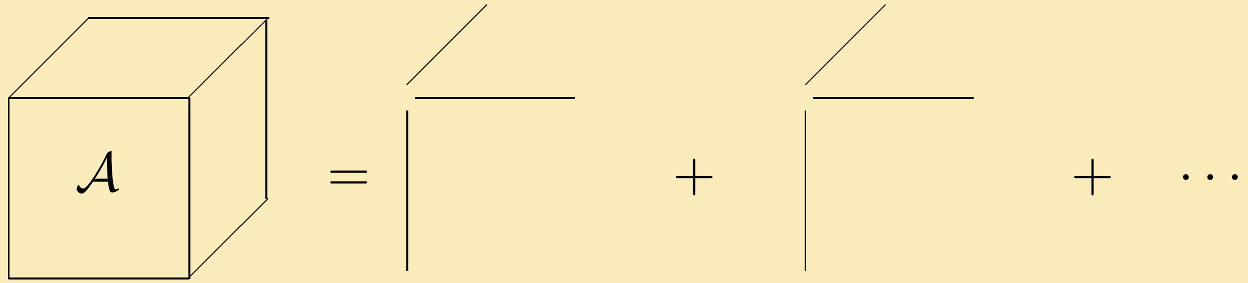
Two approaches to tensor decomposition

Tucker Model



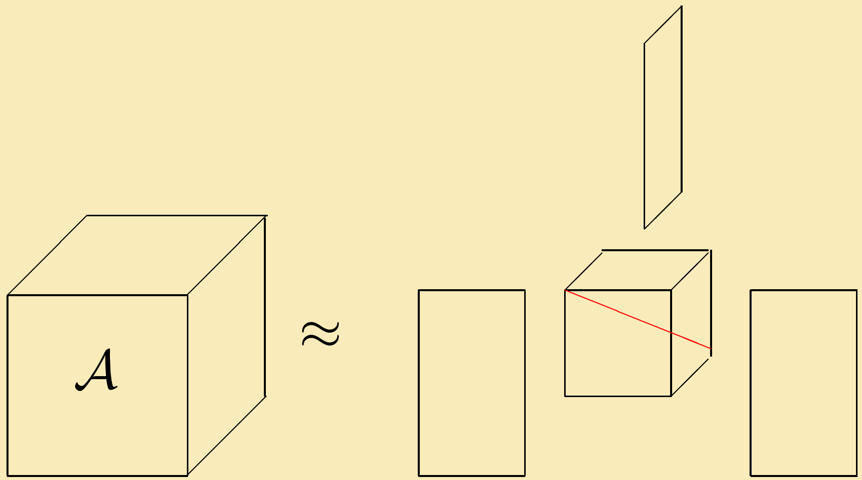
- Tucker 1966, numerous papers in psychometrics and chemometrics
- De Lathauwer, De Moor, Vandewalle, SIMAX 2000: notation, theory.

Expansion in rank-1 terms



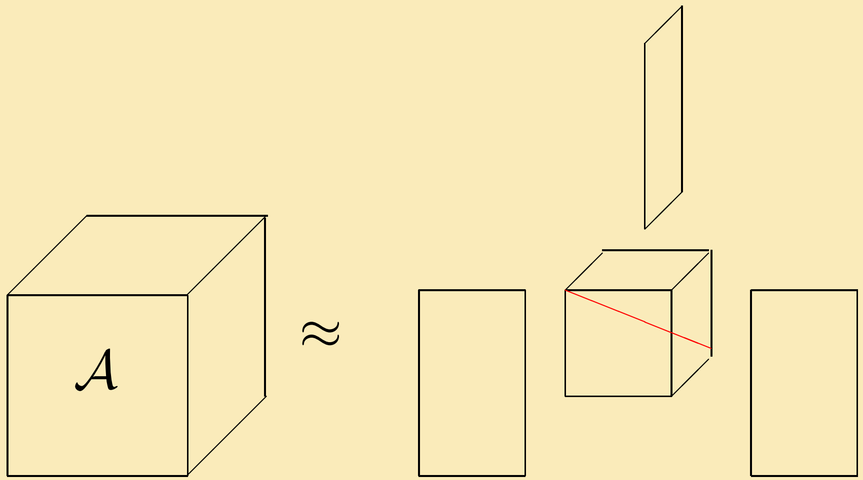
- Parafac/Candecomp/Kruskal: Harshman, Caroll, Chang 1970
- Numerous papers in psychometrics and chemometrics
- Kolda, SIMAX 2001, Zhang, Golub, SIMAX 2001, De Silva and Lim 2006

Parafac/... model: low rank approximation



The core tensor is zero except along the superdiagonal.

Parafac/... model: low rank approximation



The core tensor is zero except along the superdiagonal.

Why is it difficult to obtain this?

Because we do not have enough degrees of freedom to zero the tensor elements: $O(k^2)$ and $O(k^3)$

The Parafac approximation problem may be ill-posed![1]

Theorem 1. *There are tensors $\mathcal{A}$ for which the problem*

$$min_{x_i,y_i,z_i} \|\mathcal{A} - x_1 \otimes y_1 \otimes z_1 - x_2 \otimes y_2 \otimes z_2\|_F$$

does not have a solution. The set of tensors for which the approximation problem does not have a solution has positive volume.

The problem is illposed! (in exact arithmetic)

A well-posed problem (in floating point) near to an ill-posed one is ill-conditioned: $\Longrightarrow$ unstable computations.

Still: There are applications (e.g. in chemistry) where the Parafac model corresponds closely to the process that generates the tensor data.

[1] See De Silva and Lim (2006), Bini (1986)

Mode$-I$ multiplication of a tensor by a matrix[2]

Contravariant multiplication

$$\mathbb{R}^{n\times n\times n} \ni \mathcal{B} = (W)_{\{1\}} \cdot \mathcal{A}, \qquad \mathcal{B}(i,j,k) = \sum_{\nu=1}^{n} w_{i\nu} a_{\nu jk}.$$

All column vectors in the 3-tensor are multiplied by the matrix W.

Covariant multiplication

$$\mathbb{R}^{n\times n\times n} \ni \mathcal{B} = \mathcal{A} \cdot (W)_{\{1\}}, \qquad \mathcal{B}(i,j,k) = \sum_{\nu=1}^{n} a_{\nu jk} w_{\nu i}.$$

[2]Lim's notation

Matrix-tensor multiplication performed in all modes in the same expression:

$$(X, Y, Z) \cdot \mathcal{A} = \mathcal{A} \cdot (X^T, Y^T, Z^T)$$

Standard matrix multiplication of three matrices:

$$XAY^T = (X, Y) \cdot A$$

Inner product, orthogonality and norm

Inner product (contraction: $\mathbb{R}^{n\times n\times n} \rightarrow \mathbb{R}$)

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i,j,k} a_{ijk} b_{ijk}$$

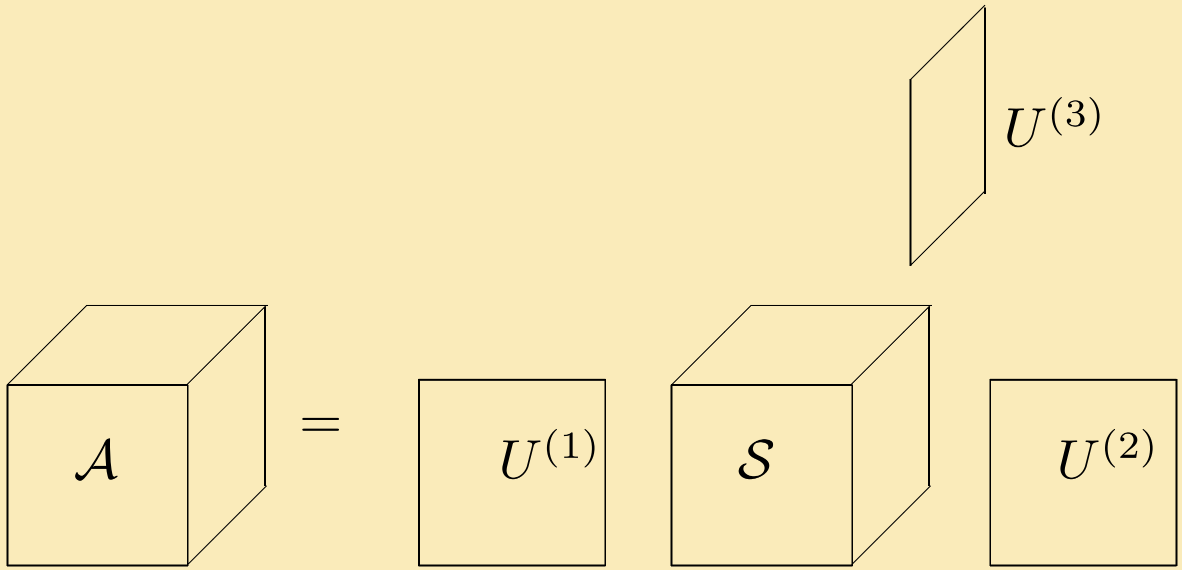
The Frobenius norm of a tensor is

$$\|\mathcal{A}\| = \langle \mathcal{A}, \mathcal{A} \rangle^{1/2}$$

Matrix case

$$\langle A, B \rangle = \mathrm{tr}(A^T B)$$

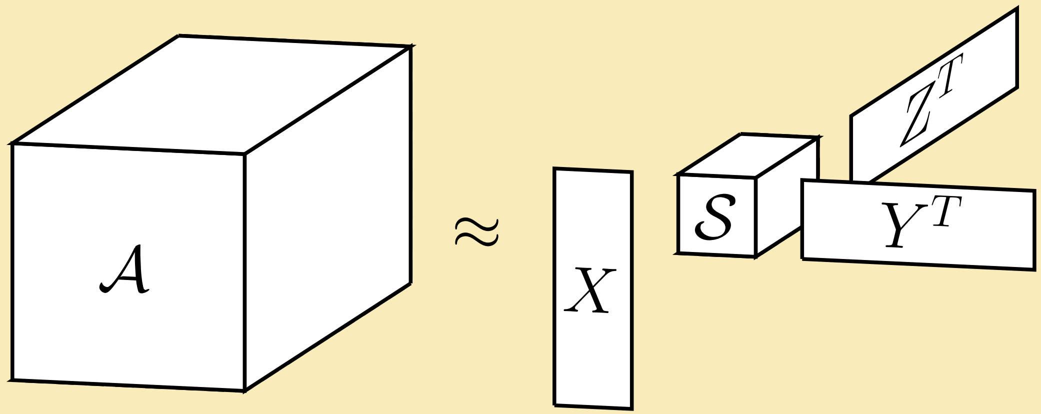
Tensor SVD (HOSVD)[3]: $\mathcal{A} = (U^{(1)}, U^{(2)}, U^{(3)}) \cdot \mathcal{S}$



The "mass" of $\mathcal{S}$ is concentrated around the $(1, 1, 1)$ corner.

Not optimal: does not solve $\min_{\mathrm{rank}(\mathcal{B})=(r_1,r_2,r_3)} \|\mathcal{A} - \mathcal{B}\|$

[3] De Lathauwer et al (2000)



Best rank$-(r_1, r_2, r_3)$ approximation:

$$\min_{X,Y,Z,\mathcal{S}} \|\mathcal{A} - (X, Y, Z) \cdot \mathcal{S}\|, \qquad X^T X = I, \quad Y^T Y = I, \quad Z^T Z = I$$

The problem is over-parameterized!

Best approximation: $\min_{\text{rank}(\mathcal{B})=(r_1,r_2,r_3)} \|\mathcal{A}-\mathcal{B}\|$

Equivalent to

$$\max_{X,Y,Z} \Phi(X,Y,Z) = \frac{1}{2}\|\mathcal{A}\cdot(X,Y,Z)\|^2 = \frac{1}{2}\sum_{j,k,l} A_{jkl}^2,$$

$$A_{jkl} = \sum_{\lambda,\mu,\nu} a_{\lambda\mu\nu} x_{\lambda j} y_{\mu k} z_{\nu l},$$

subject to

$$X^T X = I_{r_1}, \qquad Y^T Y = I_{r_2}, \qquad Z^T Z = I_{r_3}$$

Grassmann Optimization

The Frobenius norm is invariant under orthogonal transformations:

$$\Phi(X, Y, Z) = \Phi(XU, YV, ZW) = \frac{1}{2}\|\mathcal{A} \cdot (XU, YV, ZW)\|^2$$

for orthogonal $U \in \mathbb{R}^{r_1 \times r_1}$, $V \in \mathbb{R}^{r_2 \times r_2}$, and $W \in \mathbb{R}^{r_3 \times r_3}$.

Maximize Φ over equivalence classes

$$[X] = \{XU \mid U \text{ orthogonal}\}.$$

Product of Grassmann manifolds: $\mathrm{Gr}^3 = \mathrm{Gr}(J, r_1) \times \mathrm{Gr}(K, r_2) \times \mathrm{Gr}(L, r_3)$

$$\max_{(X,Y,Z)\in\mathrm{Gr}^3} \Phi(X, Y, Z) = \max_{(X,Y,Z)\in\mathrm{Gr}^3} \frac{1}{2}\langle \mathcal{A} \cdot (X, Y, Z), \mathcal{A} \cdot (X, Y, Z)\rangle$$

Newton's Method on one Grassmann Manifold

Taylor expansion + linear algebra on tangent space[4] at X

$$G(X(t)) \approx G(X(0)) + \langle \Delta, \nabla G \rangle + \frac{1}{2}\langle \Delta, H(\Delta) \rangle,$$

Grassmann gradient:

$$\nabla G = \Pi_X G_x, \qquad (G_x)_{jk} = \frac{\partial G}{\partial x_{jk}}, \qquad \Pi_X = I - XX^T$$

The Newton equation for determining Δ:

$$\Pi_X \langle \mathcal{G}_{xx}, \Delta \rangle_{1:2} - \Delta \langle X, G_x \rangle_1 = -\nabla G, \qquad (\mathcal{G}_{xx})_{jklm} = \frac{\partial^2 G}{\partial X_{jk}\, \partial X_{lm}}.$$

[4]Tangent space at X: all matrices Z satisfying $Z^T X = 0$.

Newton-Grassmann Algorithm on Gr^3

Here: local coordinates

Given tensor $\mathcal{A}$ and starting points $(X_0, Y_0, Z_0) \in \mathrm{Gr}^3$

repeat

compute the Grassmann gradient $\nabla\widehat{\Phi}$
compute the Grassmann Hessian $\widehat{\mathcal{H}}$
matricize $\widehat{\mathcal{H}}$ and vectorize $\nabla\widehat{\Phi}$
solve $D = (D_x, D_y, D_z)$ from the Newton equation
take a geodesic step along the direction D, giving new iterates (X,Y,Z)

until $\|\nabla\widehat{\Phi}\|/\Phi <$ TOL

Implementation using TensorToolbox and object-oriented Grassmann classes in Matlab

Newton's method on Gr^3

Differentiate $\Phi(X, Y, Z)$ along a geodesic curve $(X(t), Y(t), Z(t))$ in the direction $(\Delta_x, \Delta_y, \Delta_z)$:

$$\frac{\partial x_{st}}{\partial t} = (\Delta_x)_{st},$$

and

$$\left(\frac{dX(t)}{dt}, \frac{dY(t)}{dt}, \frac{dZ(t)}{dt}\right) = (\Delta_x\,,\ \Delta_y\,,\ \Delta_z),$$

Since $\mathcal{A} \cdot (X, Y, Z)$ is linear in X, Y, Z separately:

$$\frac{d(\mathcal{A} \cdot (X, Y, Z))}{dt} = \mathcal{A} \cdot (\Delta_x, Y, Z) + \mathcal{A} \cdot (X, \Delta_y, Z) + \mathcal{A} \cdot (X, Y, \Delta_z).$$

First Derivative

$$\frac{d\Phi}{dt} = \frac{1}{2}\frac{d}{dt}\langle \mathcal{A}\cdot(X,Y,Z), \mathcal{A}\cdot(X,Y,Z)\rangle = \langle \mathcal{A}\cdot(\Delta_x,Y,Z), \mathcal{A}\cdot(X,Y,Z)\rangle$$
$$+ \langle \mathcal{A}\cdot(X,\Delta_y,Z), \mathcal{A}\cdot(X,Y,Z)\rangle + \langle \mathcal{A}\cdot(X,Y,\Delta_z), \mathcal{A}\cdot(X,Y,Z)\rangle.$$

We want to write $\langle \mathcal{A}\cdot(\Delta_x,Y,Z), \mathcal{A}\cdot(X,Y,Z)\rangle$ in the form $\langle \Delta_x, \Phi_x\rangle$

Define the tensor $\mathcal{F} = \mathcal{A}\cdot(X,Y,Z)$ and write

$$\langle \mathcal{A}\cdot(\Delta_x,Y,Z), \mathcal{F}\rangle =: \langle \mathcal{K}_x(\Delta_x), \mathcal{F}\rangle = \langle \Delta_x, \mathcal{K}_x^*\mathcal{F}\rangle,$$

Linear operator:

$$\Delta_x \longmapsto \mathcal{K}_x(\Delta_x) = \mathcal{A}\cdot(\Delta_x,Y,Z)$$

Adjoint Operator

Linear operator:

$$\Delta_x \longmapsto \mathcal{K}_x(\Delta_x) = \mathcal{A} \cdot (\Delta_x, Y, Z)$$

with adjoint

$$\langle \mathcal{K}_x(\Delta_x), \mathcal{F} \rangle = \langle \Delta_x, \mathcal{K}_x^* \mathcal{F} \rangle = \langle \Delta_x, \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{F} \rangle_{-1} \rangle$$

where the partial contraction is defined

$$\langle \mathcal{B}, \mathcal{C} \rangle_{-1}(i_1, i_2) = \sum_{\mu,\nu} b_{i_1\mu\nu} c_{i_2\mu\nu}$$

Grassmann Gradient

X-part: multiply by $\Pi_x = I - XX^T$

$$\begin{aligned}\Pi_X \Phi_x &= \Pi_X \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{F} \rangle_{-1} \\ &= \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{A} \cdot (X, Y, Z) \rangle_{-1} - XX^T \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{F} \rangle_{-1} \\ &= \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{A} \cdot (I, Y, Z) \rangle_{-1} X - X \langle \mathcal{F}, \mathcal{F} \rangle_{-1},\end{aligned}$$

Complete gradient (recall $\mathcal{F} = \mathcal{A} \cdot (X, Y, Z)$):

$$\begin{pmatrix} \Pi_X \Phi_x \\ \Pi_Y \Phi_y \\ \Pi_Y \Phi_z \end{pmatrix} = \begin{pmatrix} \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{A} \cdot (I, Y, Z) \rangle_{-1} X - X \langle \mathcal{F}, \mathcal{F} \rangle_{-1} \\ \langle \mathcal{A} \cdot (X, I, Z), \mathcal{A} \cdot (X, I, Z) \rangle_{-2} Y - Y \langle \mathcal{F}, \mathcal{F} \rangle_{-2} \\ \langle \mathcal{A} \cdot (X, Y, I), \mathcal{A} \cdot (X, Y, I) \rangle_{-3} Z - Z \langle \mathcal{F}, \mathcal{F} \rangle_{-3} \end{pmatrix}.$$

Second Derivative

$$\frac{d^2\Phi}{dt^2} = \langle \mathcal{A} \cdot (\Delta_x, Y, Z), \mathcal{A} \cdot (\Delta_x, Y, Z) \rangle + \langle \mathcal{A} \cdot (\Delta_x, \Delta_y, Z), \mathcal{A} \cdot (X, Y, Z) \rangle$$
$$+ \langle \mathcal{A} \cdot (\Delta_x, Y, Z), \mathcal{A} \cdot (X, \Delta_y, Z) \rangle + \langle \mathcal{A} \cdot (\Delta_x, Y, \Delta_z), \mathcal{A} \cdot (X, Y, Z) \rangle$$
$$+ \langle \mathcal{A} \cdot (\Delta_x, Y, Z), \mathcal{A} \cdot (X, Y, \Delta_z) \rangle + \cdots ,$$

plus 10 analogous terms.

First term:

$$\begin{aligned} \langle \mathcal{A} \cdot (\Delta_x, Y, Z), \mathcal{A} \cdot (\Delta_x, Y, Z) \rangle &= \langle \Delta_x, \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{A} \cdot (\Delta_x, Y, Z) \rangle_{-1} \rangle \\ &= \langle \Delta_x, \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{A} \cdot (I, Y, Z) \rangle_{-1} \Delta_x \rangle. \end{aligned}$$

"xx" Part of Grassmann Hessian

Sylvester operator:

$$\mathcal{H}_{xx}(\Delta_x) = \Pi_X \langle \mathcal{A} \cdot (I, Y, Z), \mathcal{A} \cdot (I, Y, Z) \rangle_{-1} \Delta_x - \Delta_x \langle \mathcal{F}, \mathcal{F} \rangle_{-1},$$

“xy” Part of Grassmann Hessian

Second term:

$$\langle \mathcal{A} \cdot (\Delta_x, \Delta_y, Z), \mathcal{A} \cdot (X, Y, Z) \rangle = \langle \Delta_x, \langle \mathcal{A} \cdot (I, \Delta_y, Z), \mathcal{A} \cdot (X, Y, Z) \rangle_{-1} \rangle$$
$$= \langle \Delta_x, \langle \mathcal{F}^1_{xy}, \Delta_y \rangle_{2,4;1:2} \rangle.$$

where $\mathcal{F}^1_{xy}$ is the $4-$tensor

$$\mathcal{F}^1_{xy} = \langle \mathcal{A} \cdot (I, I, Z), \mathcal{A} \cdot (X, Y, Z) \rangle_{-(1,2)} = \langle \mathcal{A} \cdot (I, I, Z), \mathcal{A} \cdot (X, Y, Z) \rangle_3,$$

and

$$(\langle \mathcal{B}, \Delta \rangle_{2,4;1:2})_{ik} = \sum_{\mu\nu} b_{i\mu k\nu} \delta_{\mu\nu}$$

Grassmann Hessian

$$\mathcal{H}(\Delta) = \begin{pmatrix} \mathcal{H}_{xx}(\Delta_x) + \mathcal{H}_{xy}(\Delta_y) + \mathcal{H}_{xz}(\Delta_z) \\ \mathcal{H}_{yx}(\Delta_x) + \mathcal{H}_{yy}(\Delta_y) + \mathcal{H}_{yz}(\Delta_z) \\ \mathcal{H}_{zx}(\Delta_x) + \mathcal{H}_{zy}(\Delta_y) + \mathcal{H}_{zz}(\Delta_z) \end{pmatrix},$$

"Diagonal part":

$$\mathcal{H}_{xx}(\Delta_x) = \Pi_x \langle \mathcal{B}_x, \mathcal{B}_x \rangle_{-1} \Delta_x - \Delta_x \langle \mathcal{F}, \mathcal{F} \rangle_{-1}, \qquad \mathcal{B}_x = \mathcal{A} \cdot (I, Y, Z),$$
$$\mathcal{H}_{yy}(\Delta_y) = \Pi_y \langle \mathcal{B}_y, \mathcal{B}_y \rangle_{-2} \Delta_y - \Delta_y \langle \mathcal{F}, \mathcal{F} \rangle_{-2}, \qquad \mathcal{B}_y = \mathcal{A} \cdot (X, I, Z),$$
$$\mathcal{H}_{zz}(\Delta_z) = \Pi_z \langle \mathcal{B}_z, \mathcal{B}_z \rangle_{-3} \Delta_z - \Delta_z \langle \mathcal{F}, \mathcal{F} \rangle_{-3}, \qquad \mathcal{B}_z = \mathcal{A} \cdot (X, Y, I).$$

Grassmann Hessian, “upper triangular part”,

$$\mathcal{H}_{xy}(\Delta_y) = \Pi_x \left(\langle \langle \mathcal{C}_{xy}, \mathcal{F} \rangle_{-(1,2)}, \Delta_y \rangle_{2,4;1:2} + \langle \langle \mathcal{B}_x, \mathcal{B}_y \rangle_{-(1,2)}, \Delta_y \rangle_{4,2;1:2} \right),$$

$$\mathcal{H}_{xz}(\Delta_z) = \Pi_x \left(\langle \langle \mathcal{C}_{xz}, \mathcal{F} \rangle_{-(1,3)}, \Delta_z \rangle_{2,4;1:2} + \langle \langle \mathcal{B}_x, \mathcal{B}_z \rangle_{-(1,3)}, \Delta_z \rangle_{4,2;1:2} \right),$$

$$\mathcal{H}_{yz}(\Delta_z) = \Pi_y \left(\langle \langle \mathcal{C}_{yz}, \mathcal{F} \rangle_{-(2,3)}, \Delta_z \rangle_{2,4;1:2} + \langle \langle \mathcal{B}_y, \mathcal{B}_z \rangle_{-(2,3)}, \Delta_z \rangle_{4,2;1:2} \right),$$

where we have also introduced $\mathcal{C}_{xy} = \mathcal{A} \cdot (I, I, Z)$, $\mathcal{C}_{xz} = \mathcal{A} \cdot (I, Y, I)$ and $\mathcal{C}_{yz} = \mathcal{A} \cdot (X, I, I)$.

Linear operator: Fourth order tensor $\langle \mathcal{C}_{xy}, \mathcal{F} \rangle_{-(1,2)}$ acting on matrix giving matrix:

$$\langle \langle \mathcal{C}_{xy}, \mathcal{F} \rangle_{-(1,2)}, \Delta_y \rangle_{2,4;1:2}$$

Local Coordinates

Hessian is singular in Euclidean space, but non-singular on the tangent space

$(\Delta_x, \Delta_y, \Delta_z)$ to be determined, live on the tangent space:

$$\Delta_x^T X = 0, \qquad \Delta_y^T Y = 0, \qquad \Delta_z^T X = 0$$

$X_\perp$ determined so that $(X, X_\perp)$ is a (square) orthogonal matrix

$$\begin{aligned} \Delta_x &= X_\perp D_x, & D_x &\in \mathbb{R}^{(J-r_1)\times r_1}, \\ \Delta_y &= Y_\perp D_y, & D_y &\in \mathbb{R}^{(K-r_2)\times r_2}, \\ \Delta_z &= Z_\perp D_z, & D_x &\in \mathbb{R}^{(L-r_3)\times r_3}; \end{aligned}$$

Grassmann Hessian in local coordinates

$$\widehat{\mathcal{H}}(D) = \begin{pmatrix} \widehat{\mathcal{H}}_{xx}(D_x) + \widehat{\mathcal{H}}_{xy}(D_y) + \widehat{\mathcal{H}}_{xz}(D_z) \\ \widehat{\mathcal{H}}_{yx}(D_x) + \widehat{\mathcal{H}}_{yy}(D_y) + \widehat{\mathcal{H}}_{yz}(D_z) \\ \widehat{\mathcal{H}}_{zx}(D_x) + \widehat{\mathcal{H}}_{zy}(D_y) + \widehat{\mathcal{H}}_{zz}(D_z) \end{pmatrix}$$

where the diagonal operators are

$$\begin{aligned} \widehat{\mathcal{H}}_{xx}(D_x) &= \langle \widehat{\mathcal{B}}_x, \widehat{\mathcal{B}}_x \rangle_{-1} D_x - D_x \langle \mathcal{F}, \mathcal{F} \rangle_{-1}, & \widehat{\mathcal{B}}_x &= \mathcal{A} \cdot (X_\perp, Y, Z), \\ \widehat{\mathcal{H}}_{yy}(D_y) &= \langle \widehat{\mathcal{B}}_y, \widehat{\mathcal{B}}_y \rangle_{-2} D_y - D_y \langle \mathcal{F}, \mathcal{F} \rangle_{-2}, & \widehat{\mathcal{B}}_y &= \mathcal{A} \cdot (X, Y_\perp, Z), \\ \widehat{\mathcal{H}}_{zz}(D_z) &= \langle \widehat{\mathcal{B}}_z, \widehat{\mathcal{B}}_z \rangle_{-3} D_z - D_z \langle \mathcal{F}, \mathcal{F} \rangle_{-3}, & \widehat{\mathcal{B}}_z &= \mathcal{A} \cdot (X, Y, Z_\perp). \end{aligned}$$

Grassmann Hessian, “upper triangular” operators

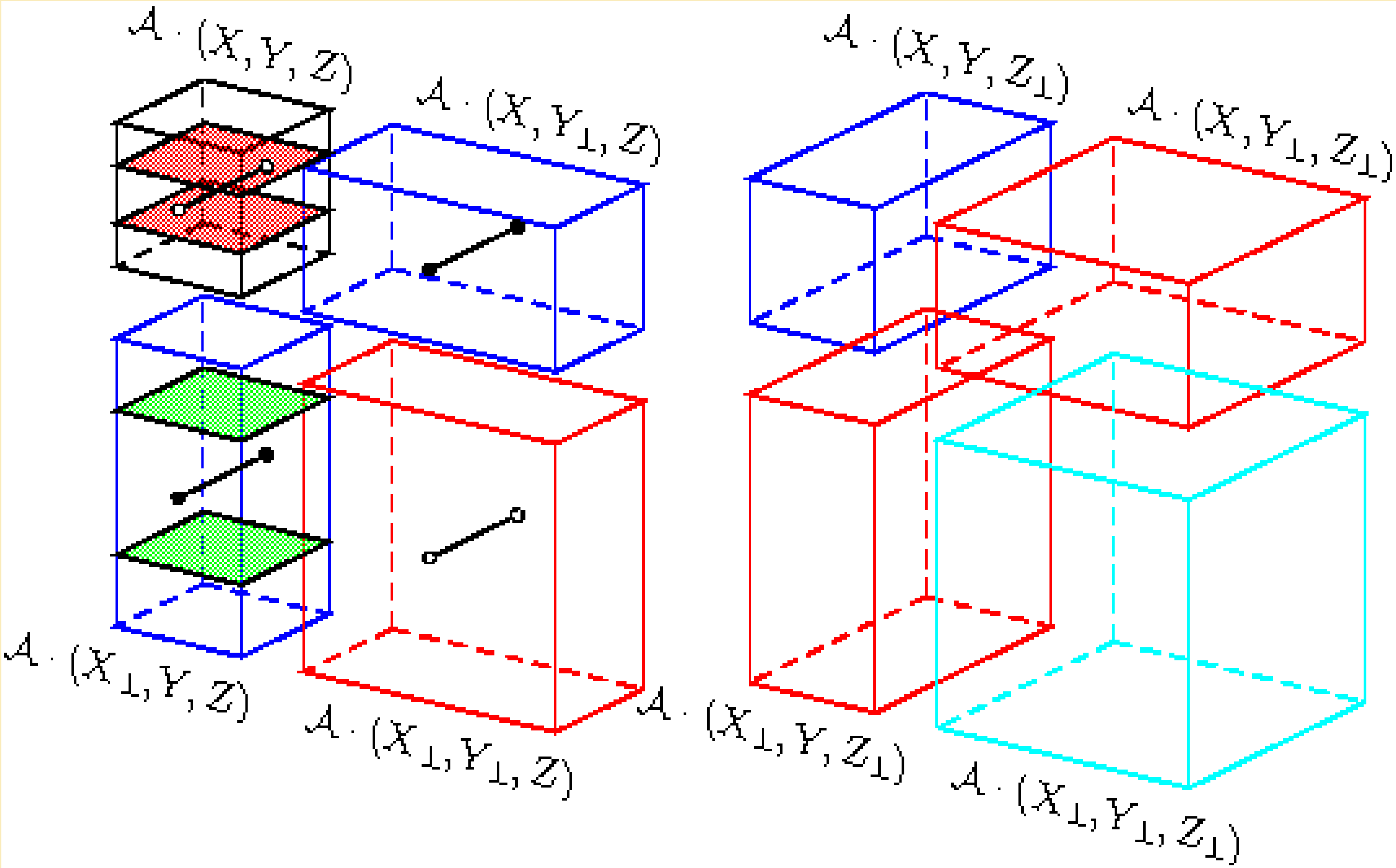
$$\widehat{\mathcal{H}}_{xy}(D_y) = \Big(\langle \langle \widehat{\mathcal{C}}_{xy}, \mathcal{F} \rangle_{-(1,2)} \, , D_y \rangle_{2,4;1:2} + \langle \langle \widehat{\mathcal{B}}_x, \widehat{\mathcal{B}}_y \rangle_{-(1,2)}, D_y \rangle_{4,2;1:2} \Big) \, ,$$

$$\widehat{\mathcal{H}}_{xz}(D_z) = \Big(\langle \langle \widehat{\mathcal{C}}_{xz}, \mathcal{F} \rangle_{-(1,3)}, D_z \rangle_{2,4;1:2} + \langle \langle \widehat{\mathcal{B}}_x, \widehat{\mathcal{B}}_z \rangle_{-(1,3)}, D_z \rangle_{4,2;1:2} \Big) \, ,$$

$$\widehat{\mathcal{H}}_{yz}(D_z) = \Big(\langle \langle \widehat{\mathcal{C}}_{yz}, \mathcal{F} \rangle_{-(2,3)} \, , D_z \rangle_{2,4;1:2} + \langle \langle \widehat{\mathcal{B}}_y, \widehat{\mathcal{B}}_z \rangle_{-(2,3)}, D_z \rangle_{4,2;1:2} \Big) \, ,$$

where $\widehat{\mathcal{C}}_{xy} = \mathcal{A} \cdot (X_\perp, Y_\perp, Z)$, $\widehat{\mathcal{C}}_{xz} = \mathcal{A} \cdot (X_\perp, Y, Z_\perp)$ and $\widehat{\mathcal{C}}_{yz} = \mathcal{A} \cdot (X, Y_\perp, Z_\perp)$.

Illustration of Hessian



Numerical Examples. Test 1

Simulate a "signal tensor" with low rank and normally distributed noise.

Two $20 \times 20 \times 20$ tensors:

$$\mathcal{A}_1 = \mathcal{B}_1 + \rho\mathcal{E}_1, \qquad \mathrm{rank}(\mathcal{B}_1) = (10, 10, 10)$$
$$\mathcal{A}_2 = \mathcal{B}_2 + \rho\mathcal{E}_2, \qquad \mathrm{rank}(cB_2) = (15, 15, 15)$$

$\mathcal{E}_i$ are noise tensors, and $\rho = 0.1$

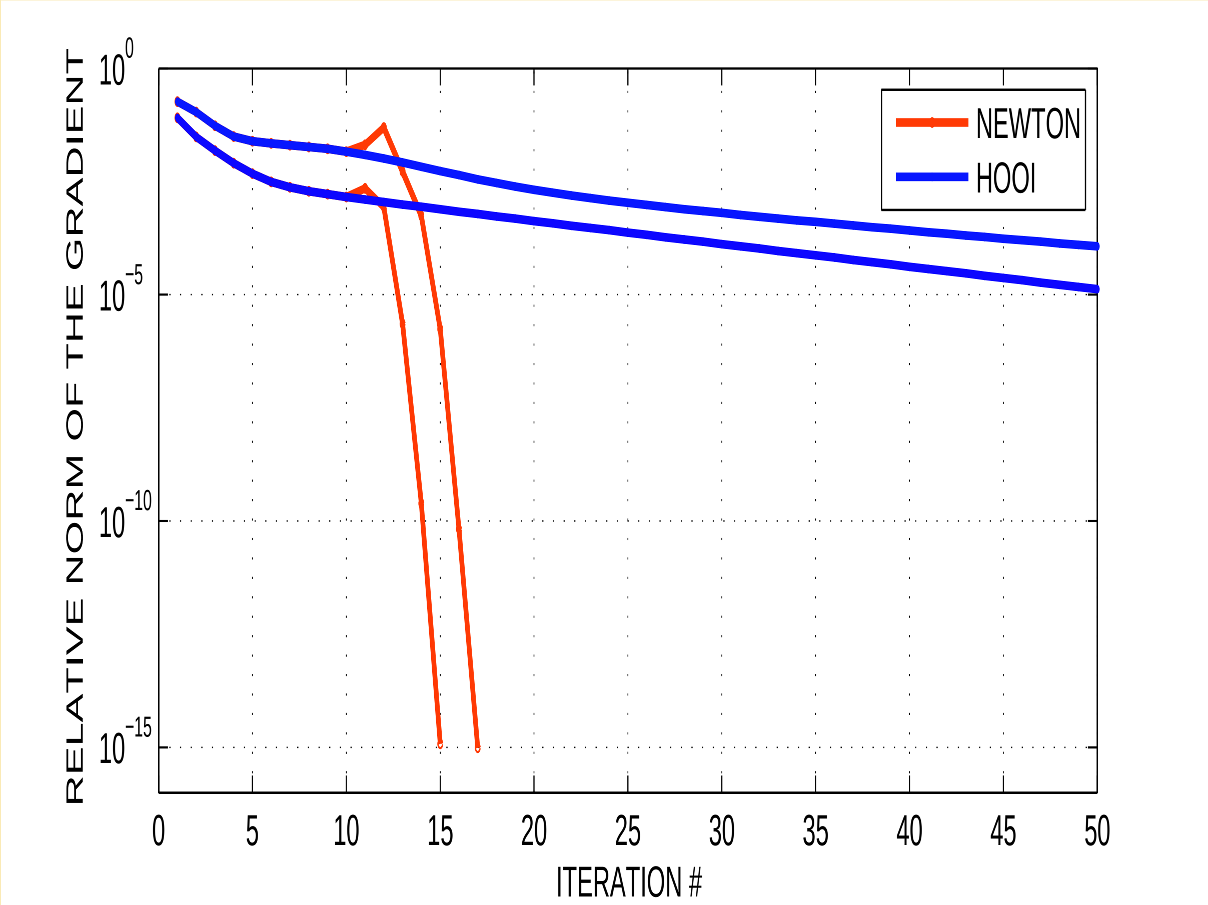
A rank$-(5, 5, 5)$ approximation was computed

Initial approximation: random tensor

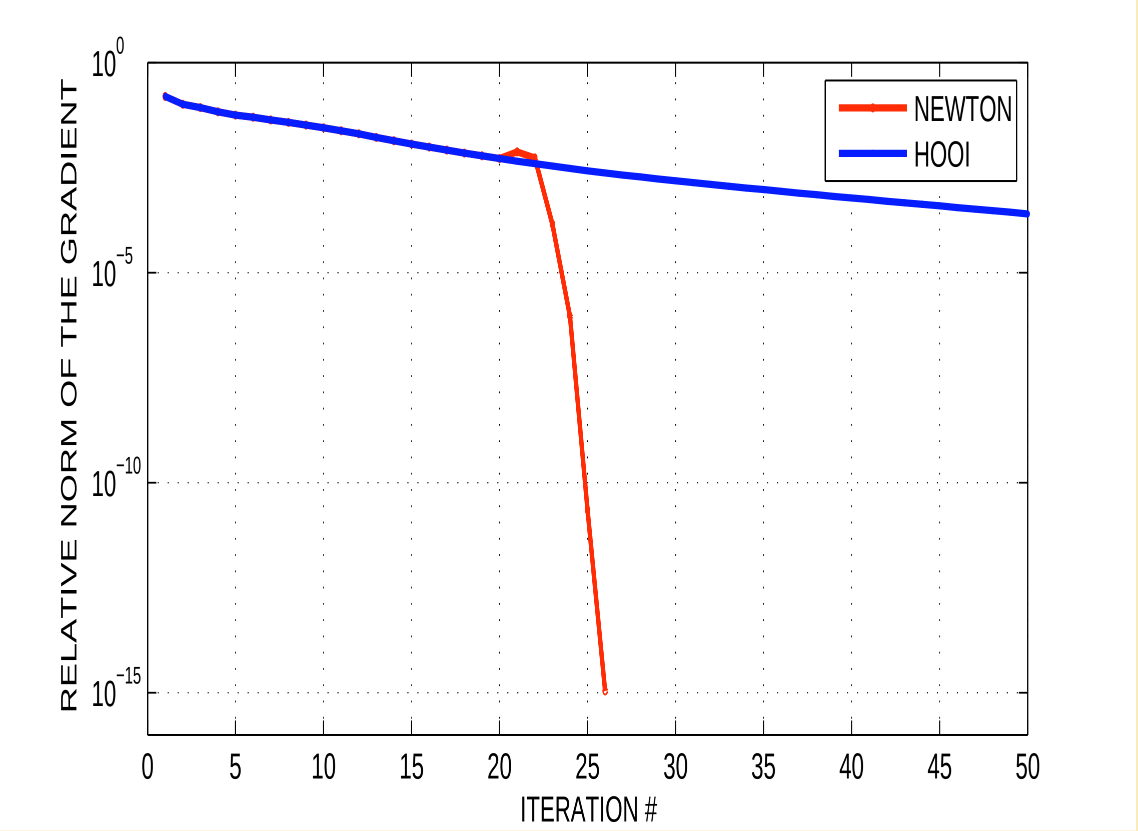
10 HOOI iterations were performed before the Newton method was started.

HOOI: "Alternating least squares" approach (De Lathauwer)

Convergence history for Test 1



Test 2: Random $20 \times 20 \times 20$ Tensor



Initialization: HOSVD and 20 HOOI iterations

Ongoing work

Matrix case:

$$\min_{\text{rank}(B)=k} \|A - B\|_F = \|A - U_1 \Sigma_1 V_1^T\|_F$$

Put

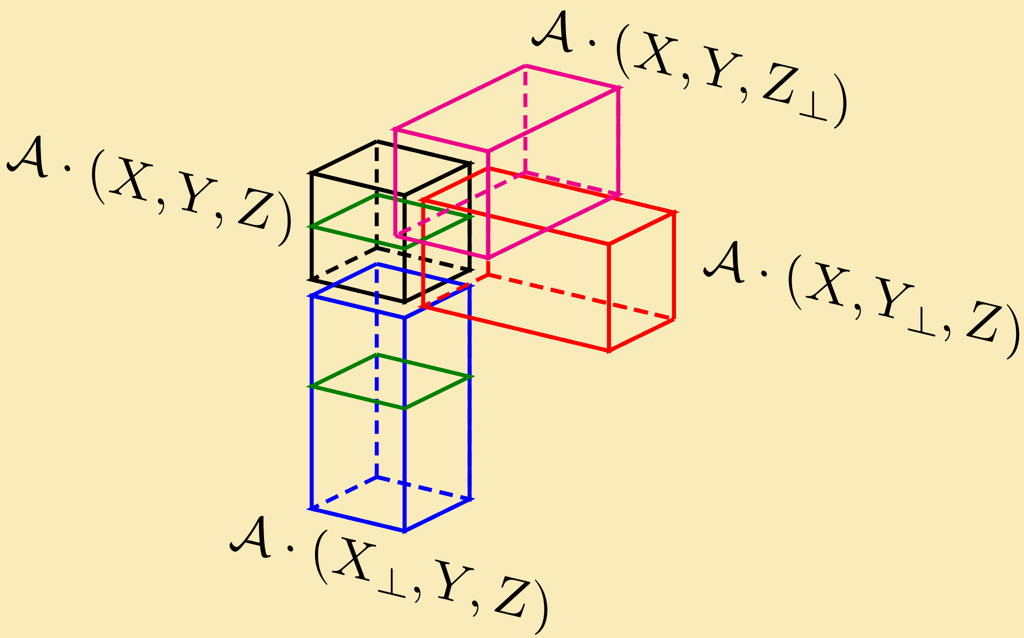
$$\widetilde{U} = \begin{pmatrix} U_1 & U_\perp \end{pmatrix}, \qquad \widetilde{V} = \begin{pmatrix} V_1 & V_\perp \end{pmatrix}$$

Then

$$\widetilde{U}^T A \widetilde{V} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & C \end{pmatrix}$$

How much can be generalized to tensors?

Tensor "SVD"?



All slices orthogonal: $\langle \mathcal{A} \cdot (X, Y, Z), \mathcal{A} \cdot (X_\perp, Y, Z) \rangle_{-1} = 0.$

Conclusions

- To exhibit structure: matricize as late as possible
- Tensor framework without extensive index wrestling
- Partial contractions play the role of adjoints
- Newton-Grassmann $\Longrightarrow$ unconstrained optimization. Quadratic convergence
- Generalization to higher order tensors is straightforward
- Present work: investigation of theoretical properties and implementation of other methods (Quasi-Newton: Savas & Lim, trust-region: Ishteva (Louvain))