ALS Iteration / (Non-)Closedness

Wolfgang Hackbusch

Max-Planck-Institut für *Mathematik in den Naturwissenschaften*
and University of Kiel

Inselstr. 22-26, D-04103 Leipzig, Germany
wh@mis.mpg.de

https://www.mis.mpg.de/scicomp/hackbusch.en.html

Creswick, February 9, 2018

Overview

I. ALS Method for Optimisation Problems

— Formulation of the Problem
— Study of Examples
— Global Convergence for Rank-1 Approximation

II. (Non-)Closedness Questions

— r-Term Format, Rank of a Tensor
— Properties of $\mathcal{R}_r$, Numerical Instability
— Strassen's Matrix Multiplication
— Matrix-Product (TT) Format, Tensor Networks
— Nonclosedness of the Cyclic Matrix-Product Format

Minimal Subspaces

— Definition
— Tensor Spaces of Linear Mappings, Functionals
— Characterisation of Minimal Subspaces in Infinite Dimensions

Topological Tensor Spaces

— Banach spaces, Crossnorms, Projective and Injective Norm
— Final Proof

1 ALS Method for Optimisation Problems

1.1 Formulation of the Problem

Let

$$\Phi(\mathbf{u}) = \min$$

be a minimisation problem over the whole tensor space $\mathbf{u} \in \mathbf{V}$.

Approximation: Choose any format $\mathcal{F} \subset \mathbf{V}$. Solve

$$\Phi(\mathbf{u}) = \min \text{ over all } \mathbf{v} \in \mathcal{F}.$$

This is the minimisation over all parameters in the representation of $\mathbf{v} \in \mathcal{F}$.

Difficulty: While the original problem may be convex, the new problem is not.

Example: $\Phi(\mathbf{u}) = \langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle - 2 \langle \mathbf{b}, \mathbf{u} \rangle$ for the solution of $\mathbf{A}\mathbf{u} = \mathbf{b}$ with positive definite matrix $\mathbf{A}$.

Example: $\Phi(\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2$ over all $\mathbf{u} \in \mathcal{R}_1 = \mathcal{T}_{(1,\ldots,1)}$. $\mathbf{v} \in \mathbf{V}$ is arbitrary.

Ansatz:

$$\mathbf{u} = u^{(1)} \otimes u^{(2)} \otimes \ldots \otimes u^{(d)}, \quad u^{(j)} \in V_j = \mathbb{R}^{n_j}$$

Necessary condition: $\nabla\Phi(\mathbf{u}) = 0$ (multilinear system of equations).

ALS = alternating least-squares method:
1) solve $\nabla\Phi(u^{(1)} \otimes u^{(2)} \otimes \ldots \otimes u^{(d)}) = 0$ w.r.t. $u^{(1)} \Rightarrow$ solution: $\hat{u}^{(1)}$,
2) solve $\nabla\Phi(\hat{u}^{(1)} \otimes u^{(2)} \otimes \ldots \otimes u^{(d)}) = 0$ w.r.t. $u^{(2)} \Rightarrow$ solution: $\hat{u}^{(2)}$,
⋮
d) solve $\nabla\Phi(\hat{u}^{(1)} \otimes \ldots \otimes \hat{u}^{(d-1)} \otimes u^{(d)}) = 0$ w.r.t. $u^{(d)} \Rightarrow$ solution: $\hat{u}^{(d)}$
All partial steps are linear problems and easy to solve.

One ALS iteration is given by $\mathbf{u}_0 = u^{(1)} \otimes \ldots \otimes u^{(d)} \mapsto \mathbf{u}_1 = \hat{u}^{(1)} \otimes \ldots \otimes \hat{u}^{(d)}$.
This defines a ALS sequence $\{\mathbf{u}_m : m \in \mathbb{N}_0\}$.

Questions: Does $\mathbf{u}_m$ converge? To what limit? Convergence speed?

1.2 First Results

Mohlenkamp (2013, Linear Algebra Appl. **438**):

- The sequence $\{\mathbf{u}_m : m \in \mathbb{N}_0\}$ is bounded,
- $\|\mathbf{u}_m - \mathbf{u}_{m+1}\| \to 0$,
- $\sum_{m=0}^{\infty} \|\mathbf{u}_m - \mathbf{u}_{m+1}\|^2 < \infty$,
- the set S of accumulation points of $\{\mathbf{u}_m\}$ is connected and compact.

Conclusion: If S contains an isolated point $\mathbf{u}^*$, it follows that $\mathbf{u}_m \to \mathbf{u}^*$.

Note that, in general, the limit may depend on the starting value!

1.3 Study of Examples

1.3.1 Case of $d = 2$

$$\mathbf{v} := \binom{1}{0} \otimes \binom{1}{0} + 2\binom{0}{1} \otimes \binom{0}{1}, \quad \Phi(\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2 .$$

1) $\mathbf{u}^{**} = 2\binom{0}{1} \otimes \binom{0}{1}$ is the global minimiser and an attractive fixed point.
2) $\mathbf{u}^{*} = \binom{1}{0} \otimes \binom{1}{0}$ is a fixed point of the ALS iteration:

$$\Phi(\mathbf{u}^* + \delta_1 \otimes \binom{1}{0}) = \Phi(\mathbf{u}^*) + \|\delta_1\|^2 .$$

But $\Phi\left(\binom{1}{t} \otimes \binom{1}{t}\right) = \Phi(\mathbf{u}^*) - t^2\left(2 - t^2\right)$

$\Rightarrow \mathbf{u}^*$ is a saddle point and a repulsive fixed point.

Conclusion: Almost all starting values lead to $\mathbf{u}_m \to \mathbf{u}^{**}$.

1.3.2 Case of $d \geq 3$

For $a \perp b$ with $\|a\| = \|b\| = 1$ consider $\Phi(\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2$ with

$$\mathbf{v} = \otimes^3 a + 2 \otimes^3 b.$$

Again $\mathbf{u}^* = \otimes^3 a$ and $\mathbf{u}^{**} = 2 \otimes^3 b$ are fixed points, $\Phi(\mathbf{u}^{**}) < \Phi(\mathbf{u}^*)$.
But now both are local minima (attractive fixed points)!
Additional saddle point (repulsive fixed point): $\mathbf{u}^{***} = c \otimes^3 (a + \frac{1}{2} b)$.

The sequence $\{\mathbf{u}_m\}$ corresponding to the starting value

$$\mathbf{u}_0 = c^{(0)} \left(a + t_1^{(0)} b \right) \otimes \left(a + t_2^{(0)} b \right) \otimes \left(a + t_3^{(0)} b \right)$$

is completely defined by $t_2^{(0)}$ and $t_3^{(0)}$. The characteristic value is

$$\tau_m := \left| t_2^{(m)} \right|^\alpha \left| t_3^{(m)} \right|^\beta \quad \text{with} \quad \alpha = 5^{1/2} - 1, \ \beta = 2.$$

(A) $\tau_0 > 2^{-\gamma}$, $\gamma = 5^{1/2} + 1 \Rightarrow \mathbf{u}_m \to \mathbf{u}^{**}$ (global minimiser),
(B) $\tau_0 < 2^{-\gamma} \Rightarrow \mathbf{u}_m \to \mathbf{u}^*$ (local minimiser),
(C) $\tau_0 = 2^{-\gamma} \Rightarrow \mathbf{u}_m \to \mathbf{u}^{***}$ (saddle point, global minimiser on the manifold $\tau = 2^{-\gamma}$).

We recall:

Conclusion: If the set of accumulation points of $\{\mathbf{u}_m\}$ contains an isolated point $\mathbf{u}^*$, it follows that $\mathbf{u}_m \to \mathbf{u}^*$.

Wang–Chu (2014): Global convergence for almost all $\mathbf{u}_0$.

Uschmajew (2015):

Analysis based on the Łojasiewicz inequality yields:
All sequences $\mathbf{u}_m$ converge to some $\mathbf{u}^*$ with $\nabla\Phi(\mathbf{u}^*) = 0$.

Łojasiewicz (1965, Ensembles semi-analytiques): If Φ is analytic,

$$\exists\theta \in (0, 1/2] \quad |\Phi(x) - \Phi(x_*)|^{1-\theta} \leq \|\nabla\Phi(x)\|$$

in some neighbourhood of x_*.

Convergence speed?

The proof by the Łojasiewicz inequality is not constructive.

Espig–Khachatryan (2015): Study of sequences for $\Phi(\mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|^2$ with

$$\begin{aligned} \mathbf{v} &= \otimes^3 a + \lambda \left(a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a \right), \\ & a \perp b, \qquad \|a\| = \|b\| = 1. \end{aligned}$$

Depending on the value of λ it is shown that the convergence can be

- sublinear ($\lambda = 1/2$),
- linear ($\lambda < 1/2$).

For $\mathbf{v} = \otimes^3 a + 2 \otimes^3 b$, $\mathbf{u}_m \to \otimes^3 a$ or $2 \otimes^3 b$, we have

- superlinear convergence (of order $2 + 5^{1/2} > 1$)

Study of the general case: Gong–Mohlenkamp–Young 2017

2 (Non-)Closedness Questions

2.1 r-Term Format, Rank of a Tensor

$\mathbb{K}$: underlying field ($\mathbb{R}$ or $\mathbb{C}$). V_j vector spaces over $\mathbb{K}$. Any algebraic tensor has the form $\mathbf{v} = \sum_{i=1}^{r} \bigotimes_{j=1}^{d} v_i^{(j)}$, $v_i^{(j)} \in V_j$, for some $r \in \mathbb{N}_0$. Fixing r, we obtain the set

$$\mathcal{R}_r := \left\{ \sum_{i=1}^{r} \bigotimes_{j=1}^{d} v_i^{(j)} : v_i^{(j)} \in V_j \right\}$$

of tensors with *representation rank* r. Using the rank

$$\mathrm{rank}(\mathbf{v}) := \min\{m : \mathbf{v} \in \mathcal{R}_m\},$$

we may write $\mathcal{R}_r := \{\mathbf{v} \in \mathbf{V} : \mathrm{rank}(\mathbf{v}) \leq r\}$.

The maximal rank of $\mathbf{V}$ is

$$\mu := \sup\{\mathrm{rank}(\mathbf{v}) : \mathbf{v} \in \mathbf{V}\}.$$

$\mu < \infty$ holds for finite-dimensional V_j and is equal to $\min\{m : \mathcal{R}_{m+1} = \mathcal{R}_m\}$.

Properties of $\mathcal{R}_r$:

- In general, the determination of rank($\mathbf{v}$) is *NP hard* (cf. Håstad 1990).

- In general, the *maximal rank* is not explicitly known. For equal dimensions $\dim(V_j) = n$:

$$\frac{n^{d-1}}{d} \leq r_{\max} \leq \frac{d}{2(d-1)} n^{d-1} + O(n^{d-2}).$$

- For *random tensors* there may be more than one tensor rank with positive probability. These ranks are called *typical*.

- Real tensors may have different rank depending on the underlying fields $\mathbb{R}$ or $\mathbb{C}$.

- In general, $\mathcal{R}_r$ is *not closed*. Example: a, b linearly independent and

$$\begin{aligned} \mathbf{v} &= a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a \in \mathcal{R}_3 \backslash \mathcal{R}_2 \\ \mathbf{v} &= \underbrace{(b + na) \otimes \left(a + \tfrac{1}{n} b\right) \otimes a + a \otimes a \otimes (b - na)}_{\mathbf{v}_n \in \mathcal{R}_2} - \tfrac{1}{n} b \otimes b \otimes a. \end{aligned}$$

- *border rank*: $\underline{\text{rank}}(\mathbf{v}) := \min\{r \in \mathbb{N}_0 : \mathbf{v} \in closure(\mathcal{R}_r)\}$.

Numerical Instability

In the previous example, the terms of $\mathbf{v}_n$ grow like $O(n)$, while the result is of size $O(1)$.

This implies *numerical cancellation*: $\log_2 n$ binary digits of $\mathbf{v}_n$ are lost.

We say that the sequence $\{\mathbf{v}_n\}$ is unstable.

Proposition: Suppose $\dim(V_j) < \infty$ and $\mathbf{v} \in \mathbf{V} = \bigotimes_{j=1}^{d} V_j$.
A stable sequence $\mathbf{v}_n \in \mathcal{R}_r$ with $\lim \mathbf{v}_n = \mathbf{v}$ exists if and only if $\mathbf{v} \in \mathcal{R}_r$.

Conclusion: If $\mathbf{v} = \lim \mathbf{v}_n \notin \mathcal{R}_r$, the sequence $\mathbf{v}_n \in \mathcal{R}_r$ is unstable.

Best approximation problem: Let $\mathbf{v}^* \in \mathbf{V}$. Try to find $\mathbf{v} \in \mathcal{R}_r$ with

$$\|\mathbf{v}^* - \mathbf{v}\| = \inf\{\|\mathbf{v}^* - \mathbf{w}\| : \mathbf{w} \in \mathcal{R}_r\}.$$

This optimisation problem need **not be solvable.**

The set of $\mathbf{v}^* \in \mathbf{V}$ with inf $\neq$ min has a positive measure if $\mathbb{K} = \mathbb{R}$ (De Silva–Lim 2008), but measure zero if $\mathbb{K} = \mathbb{C}$ (Qi–Michałek–Lim, 2017).

3 Strassen's Matrix Multiplication

Standard matrix-matrix multiplication costs $2n^3$ operations.
Strassen 1969: $4.7n^{\log_2 7} = 4.7n^{2.8074}$

Two 2×2 block matrices can be multiplied as follows:

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}, \qquad a_i, b_i, c_i \text{ submatrices with}$$

$$c_1 = m_1 + m_4 - m_5 + m_7, c_2 = m_2 + m_4, c_3 = m_3 + m_5, c_4 = m_1 + m_3 - m_2 + m_6$$

$$m_1 = (a_1 + a_4)(b_1 + b_4),$$

$$m_2 = (a_3 + a_4)b_1,$$

$$m_3 = a_1(b_2 - b_4),$$

$$m_4 = a_4(b_3 - b_1),$$

$$m_5 = (a_1 + a_2)b_4,$$

$$m_6 = (a_3 - a_1)(b_1 + b_2),$$

$$m_7 = (a_2 - a_4)(b_3 + b_4).$$

Tensor of the matrix-matrix multiplication $\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}$:

$$c_\nu = \sum_{\mu,\lambda=1}^{4} \mathbf{v}_{\nu\mu\lambda}\, a_\mu\, b_\lambda \qquad (1 \le \nu \le 4).$$

For instance for $\nu = 1$, the identity $c_1 = a_1 b_1 + a_2 b_3$ shows that $\mathbf{v}_{111} = \mathbf{v}_{123} = 1$, and $\mathbf{v}_{1\mu\lambda} = 0$ otherwise. Assume a representation of $\mathbf{v}$ by r terms:

$$\mathbf{v} = \sum_{i=1}^{r} \bigotimes\nolimits_{j=1}^{3} v_i^{(j)} \in \bigotimes\nolimits_{j=1}^{3} \mathbb{K}^4.$$

The insertion into $c_\nu = \sum_{\mu,\lambda=1}^{4} \mathbf{v}_{\nu\mu\lambda}\, a_\mu\, b_\lambda$ yields

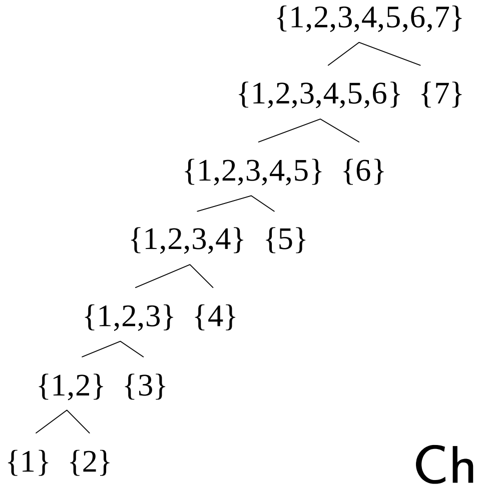
$$\begin{aligned} c_\nu &= \sum_{i=1}^{r} \sum_{\mu,\lambda=1}^{4} v_i^{(1)}[\nu]\, v_i^{(2)}[\mu]\, v_i^{(3)}[\lambda]\, a_\mu\, b_\lambda \\ &= \sum_{i=1}^{r} v_i^{(1)}[\nu] \left(\sum_{\mu=1}^{4} v_i^{(2)}[\mu]\, a_\mu \right) \left(\sum_{\lambda=1}^{4} v_i^{(3)}[\lambda]\, b_\lambda \right), \end{aligned}$$

requiring r multiplications.

Strassen 1969: rank$(\mathbf{v}) \le 7$, Winograd 1971: rank$(\mathbf{v}) = 7$,
Landsberg 2012: $\underline{\text{rank}}(\mathbf{v}) = 7$.

4 Matrix-Product (TT) Format, Tensor Networks

The hierarchical tensor format is based on a binary tree. A particular binary tree is



Choosing $U_j := V_j$ for the subspaces at the leaves $j = 1, \ldots, d$, one obtains the TT format (Oseledets–Tyrtyshnikov 2005). It coincides with the description of the matrix product states (Vidal 2003, Verstraete–Cirac 2006) used in physics:

Each component $\mathbf{v}[i_1, \ldots, i_d]$ of $\mathbf{v} \in \mathbf{V} = \bigotimes_{j=1}^{d} \mathbb{K}^{n_j}$ is expressed by

$$\mathbf{v}[i_1 i_2 \cdots i_d] = V^{(1)}[i_1] \cdot V^{(2)}[i_2] \cdot \ldots \cdot V^{(d-1)}[i_{d-1}] \cdot V^{(d)}[i_d] \in \mathbb{K},$$

where $V^{(j)}[i]$ are matrices of size $r_{j-1} \times r_j$ with $r_0 = r_d = 1$. The minimal size of r_j is $\mathrm{rank}_{\{1,\ldots,j\}}(\mathbf{v})$.

To avoid the special roles of the vectors $V^{(1)}[i_1], V^{(d)}[i_d]$ and to describe periodic situations, the Cyclic Matrix-Product format $\mathcal{C}(d,(r_j))$ is used in physics:

$$\begin{aligned}\mathbf{v}[i_1 i_2 \cdots i_d] &= \mathsf{trace}\{V^{(1)}[i_1]\cdot V^{(2)}[i_2]\cdot\ldots\cdot V^{(d-1)}[i_{d-1}]\cdot V^{(d)}[i_d]\} \\ &= \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d} V_{k_d k_1}{}^{(1)}[i_1]\cdot V^{(2)}_{k_1 k_2}[i_2]\cdot\ldots\cdot V^{(d-1)}[i_{d-1}]\cdot V_{k_{d-1}k_d}{}^{(d)}[i_d].\end{aligned}$$

Tensor Network: tensor representations based on general graphs which are in general not a tree. Here the graph is a cycle with d vertices.

THEOREM (Landsberg–Qi–Ye 2012) Formats based on a graph$\neq$tree are in general not closed.

Site-independent format $\mathcal{C}_{ind}(d,r)$: $V^{(j)}[i] = V[i]$ and $r_j = r$ for all j.

4.1 Example for $d = 3$, $\mathbf{V} = \otimes^3 \mathbb{K}^{2\times 2}$, $r_1 = r_2 = r_3 = 2$ by Harris–Michałek–Sertöz 2018

Let

$$\mathbf{m} := \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} E^{(1)}_{k_d,k_1} \otimes E^{(2)}_{k_1 k_2} \otimes \ldots \otimes E^{(d-1)}_{k_{d-2}k_{d-1}} \otimes E^{(d)}_{k_{d-1},k_d} \in \bigotimes_{j=1}^{d} \mathbb{K}^{r_{j-1}\times r_j}.$$

$E^{(j)}_{pq}$ is the matrix with entries $E^{(j)}_{pq}[k,\ell] = \delta_{pk}\delta_{q\ell}$.
$\{E^{(j)}_{pq} : 1 \le p \le r_{j-1}, 1 \le q \le r_j\}$ is the canonical basis of $\mathbb{K}^{r_{j-1}\times r_j}$.

LEMMA. Let $\mathbf{V} = \bigotimes_{j=1}^{d} V_j$. The set $\mathcal{C}(d,(r_j))$ consists of all

$$\mathbf{v} = \Phi(\mathbf{m}) \qquad \text{with} \quad \Phi = \bigotimes_{j=1}^{d} \phi^{(j)} \quad \text{and} \quad \phi^{(j)} \in L(\mathbb{K}^{r_{j-1}\times r_j}, V_j).$$

In our case, we have $\phi^{(j)} \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$.
We first consider the site-independent case $V^{(j)}[i] = V[i]$ for all $1 \le j \le d := 3$.

Define $\psi \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$ by $\psi(E_{12}) = E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $\psi(E_{pq}) = 0$ for $(p,q) \neq (1,2)$. Together with the identity $id \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$, define

$$\mathbf{v}(t) = \left(\otimes^3(\psi + t \cdot id)\right)(\mathbf{m}) \qquad \text{for } t \in \mathbb{R},$$

where $\mathbf{m} = \sum_{k_1=1}^{2} \sum_{k_2=1}^{2} \sum_{k_3=1}^{2} E_{k_3 k_1} \otimes E_{k_1 k_2} \otimes E_{k_2 k_3} \in \mathbf{V}$.
Multilinearity yields $\mathbf{v}(t) = \mathbf{v}_0 + t \cdot \mathbf{v}_1 + t^2 \cdot \mathbf{v}_2 + t^3 \cdot \mathbf{v}_3$ with

$$\mathbf{v}_0 = (\otimes^3 \psi)(\mathbf{m}), \qquad \mathbf{v}_1 = [\psi \otimes \psi \otimes id + \psi \otimes id \otimes \psi + id \otimes \psi \otimes \psi](\mathbf{m}),$$
$$\mathbf{v}_2 = [id \otimes id \otimes \psi + id \otimes \psi \otimes id + \psi \otimes id \otimes id](\mathbf{m}), \qquad \mathbf{v}_3 = \mathbf{m}.$$

Note that $\psi(E_{ij}) \cdot \psi(E_{k\ell}) = 0$. Since $\mathbf{v}_0$ and $\mathbf{v}_1$ involve three or two ψ applications, $\mathbf{v}_0 = \mathbf{v}_1 = 0$ follows.
Evaluation of $\mathbf{v}_2$ yields

$$\begin{aligned} \mathbf{v}_2 = {} & E_{21} \otimes E_{11} \otimes E_{12} + E_{22} \otimes E_{21} \otimes E_{12} + E_{11} \otimes E_{12} \otimes E_{21} \\ & + E_{21} \otimes E_{12} \otimes E_{22} + E_{12} \otimes E_{21} \otimes E_{11} + E_{12} \otimes E_{22} \otimes E_{21}. \end{aligned}$$

$\mathbf{v}_0 = \mathbf{v}_1 = 0$ allows us to form the limit $\mathbf{v}_2 = \lim_{t \to 0} t^{-2} \mathbf{v}(t)$. The Lemma states that $t^{-2}\mathbf{v}(t) \in \mathcal{C}_{\text{ind}}(3,2)$ for $t > 0$.

The non-closedness of $\mathcal{C}_{\mathsf{ind}}(3,2)$ will follow from $\mathbf{v}_2 \notin \mathcal{C}_{\mathsf{ind}}(3,2)$.

For an indirect proof assume $\mathbf{v}_2 \in \mathcal{C}_{\mathsf{ind}}(3,2)$. The Lemma implies that there is some $\phi \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$ with $\mathbf{v}_2 = (\otimes^3 \phi)(\mathbf{m})$.
It is easy to check that the range of the matricisation $\mathcal{M}_1((\otimes^3\phi)(\mathbf{m})) = \phi\mathcal{M}_1(\mathbf{m})(\otimes^2\phi)^{\mathsf{T}}$ is $\mathbb{K}^{2\times 2}$.
Therefore the map ϕ must be surjective.

Since $\phi \in L(\mathbb{K}^{2\times 2}, \mathbb{K}^{2\times 2})$, surjectivity implies injectivity.
Hence $\phi : \mathbb{K}^{2\times 2} \to \mathbb{K}^{2\times 2}$ is a vector space isomorphism and $\otimes^3\phi : \mathbf{V} \to \mathbf{V}$ a tensor space isomorphisms. $\mathbf{v}_2 = (\otimes^3\phi)(\mathbf{m}) \Rightarrow \mathsf{rank}(\mathbf{v}_2) = \mathsf{rank}(\mathbf{m})$.

The representation of $\mathbf{v}_2$ yields $\mathsf{rank}(\mathbf{v}_2) \le 6$.
On the other hand, $\mathsf{rank}(\mathbf{m}) = 7$ holds for the Strassen tensor $\mathbf{m}$.
This contradiction proves that $\mathbf{v}_2 \notin \mathcal{C}_{\mathsf{ind}}(3,2)$.

Similarly $\mathbf{v}_2 \notin \mathcal{C}(3,(2,2,2))$ follows (no site-independence).

4.2 Example for $\mathbf{V} = \bigotimes^d \mathbb{C}^2$, $r_j = 2$

Smallest (nontrivial) dimension: $V_j = \mathbb{C}^2$,

tensor space $\mathbf{V} = \bigotimes^d \mathbb{C}^2$

Site-independent cyclic format $\mathcal{C}_{\mathsf{ind}}(d, 2)$, i.e., $r_j = 2$

Result:

$d = 3$: $\mathcal{C}_{\mathsf{ind}}(3, 2)$ is closed (cf. Harris–Michałek–Sertöz 2018)

$d > 3$: $\mathcal{C}_{\mathsf{ind}}(d, 2)$ is not closed (cf. Seynnaeve 2018)

For $\mathbb{K} = \mathbb{R}$, $d \geq 3$, $\mathcal{C}_{\mathsf{ind}}(d, 2)$ is not closed (cf. Seynnaeve 2018)

5 Minimal Subspaces

5.1 Tensor Subspace Format

Set of tensors of multilinear rank $\leq \mathbf{r} = (r_1, \ldots, r_d) \in \mathbb{N}^d$ is

$$\mathcal{T}_{\mathbf{r}} := \bigcup_{\dim(U_j) \leq r_j} \bigotimes_{j=1}^{d} U_j.$$

Question: Is $\mathcal{T}_{\mathbf{r}}$ closed?

In the finite-dimensional case, $\dim V_j < \infty$, compactness arguments show that $\mathcal{T}_{\mathbf{r}}$ is closed.

What happens in the case of infinite-dimensional Banach spaces $\mathbf{V} = \bigotimes_{j=1}^{d} V_j$?

5.2 Minimal Subspaces

Let $\mathbf{v} \in \mathbf{V} = \bigotimes_{j=1}^{d} V_j$ — possibly $\dim V_j = \infty$ — be an *algebraic* tensor. The minimal subspaces $U_j^{\min}(\mathbf{v})$ are defined by

$$\mathbf{v} \in \bigotimes_{j=1}^{d} U_j^{\min}(\mathbf{v}), \quad \text{and}$$

$$\text{if } \mathbf{v} \in \bigotimes_{j=1}^{d} U_j \ (U_j \text{ subspace of } V_j), \text{ then } U_j^{\min}(\mathbf{v}) \subset U_j.$$

REMARK: (a) $\dim U_j^{\min}(\mathbf{v}) \leq \operatorname{rank}(\mathbf{v}) < \infty$.

(b) $\left(\bigotimes_{j=1}^{d} U_j'\right) \cap \left(\bigotimes_{j=1}^{d} U_j''\right) = \bigotimes_{j=1}^{d} \left(U_j' \cap U_j''\right)$.

Conclusion: $U_j^{\min}(\mathbf{v})$ is the subspace of minimal dimension in

$$\mathbf{v} \in U_j \otimes \mathbf{V}_{[j]} \quad \text{with} \quad \mathbf{V}_{[j]} := \bigotimes_{k \neq j} V_j.$$

5.2.1 Matricisation

The j-th *matricisation* $\mathcal{M}_j : \mathbf{V} = \bigotimes_{k=1}^{d} \mathbb{K}^{n_k} \to \mathbb{K}^{n_j \times n_{[j]}}$ defined by

$$\mathbf{v} \mapsto M_j := \mathcal{M}_j(\mathbf{v}) \in \mathbb{K}^{n_j \times n_{[j]}} \qquad \text{with } n_{[j]} := \prod_{k \neq j} n_k,$$

$$M_j[i_j, \mathbf{i}_{[j]}] := \mathbf{v}[i_1, \ldots, i_d], \qquad \mathbf{i}_{[j]} := (i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d).$$

REMARK. If $V_j = \mathbb{K}^{n_j}$, then $U_j^{\min}(\mathbf{v}) = range(\mathcal{M}_j(\mathbf{v})) = range(M_j)$.

Consequences:

$$\text{rank}_j(\mathbf{v}) := \text{rank}(\mathcal{M}_j(\mathbf{v})) \qquad \text{for } 1 \leq j \leq d,$$

$$\mathcal{T}_\mathbf{r} = \{\mathbf{v} \in \mathbf{V} : \text{rank}_j(\mathbf{v}) \leq r_j \text{ for } 1 \leq j \leq d\}.$$

Generalisation for infinite-dimensional Hilbert spaces is possible ($n_j = \infty$), but not for general Banach spaces.

6 Tensor Spaces of Linear Mappings

Let V_j, W_j be vector spaces defining $\mathbf{V} := \bigotimes_{j=1}^{d} V_j$ and $\mathbf{W} := \bigotimes_{j=1}^{d} W_j$. Then the sets of linear maps

$$L_j := L(V_j, W_j)$$

are again vector spaces. They define the tensor space

$$\mathbf{L} := \bigotimes_{j=1}^{d} L_j.$$

$\mathbf{L}$ can be embedded into $L(\mathbf{V}, \mathbf{W})$: $\mathbf{A} = \bigotimes_{j=1}^{d} A^{(j)} \in \mathbf{L}$ is the linear map defined by

$$\mathbf{A} \bigotimes_{j=1}^{d} v^{(j)} := \bigotimes_{j=1}^{d} \left(A^{(j)} v^{(j)} \right).$$

6.1 Functionals

REMARK: $\dim(V_j) = 1 \Rightarrow \mathbf{V} = \bigotimes_{k=1}^{d} V_k$ isomorphic to

$$\mathbf{V}_{[j]} := V_1 \otimes V_2 \otimes \ldots \otimes V_{j-1} \otimes V_{j+1} \ldots \otimes V_d;$$

in particular

$$\mathbb{K} \otimes \mathbb{K} \otimes \ldots \otimes \mathbb{K} \otimes V_j \otimes \mathbb{K} \otimes \ldots \otimes \mathbb{K} \simeq V_j.$$

Let functionals $\varphi_k : V_k \to \mathbb{K}$ be given for all $k \neq j$. Then

$$\boldsymbol{\varphi}^{[j]} := \varphi_1 \otimes \ldots \otimes \varphi_{j-1} \otimes id \otimes \varphi_{j+1} \otimes \ldots \otimes \varphi_d \in L(\mathbf{V}, V_j)$$

maps $\mathbf{V}$ into V_j.

We identify $\bigotimes_{k \neq j} \varphi_k \in \mathbf{V}'_{[j]}$ with $\boldsymbol{\varphi}^{[j]} \in L(\mathbf{V}, V_j)$.

6.1.1 Minimal Subspaces

$$U_j^{\mathrm{min}}(\mathbf{v}) := \left\{ \varphi(\mathbf{v}) : \ \varphi \in {}_a\bigotimes_{k \neq j} V_k' \right\}$$
$$= \left\{ \varphi(\mathbf{v}) : \ \varphi \in \left({}_a\bigotimes_{k \neq j} V_k \right)' \right\}.$$

V_k': algebraic dual space of V_k.

In the finite-dimensional case, this statement is equivalent to $U_j^{\mathrm{min}}(\mathbf{v}) = range(\mathcal{M}_j(\mathbf{v}))$.

In the infinite-dimensional case, the definition of $\mathrm{rank}_j(\mathbf{v})$ can be extended by

$$\mathrm{rank}_j(\mathbf{v}) := \dim(U_j^{\mathrm{min}}(\mathbf{v})).$$

Under rather general assumptions on the norms of V_j and $\mathbf{V}$ we shall prove that

$$\mathbf{v}_n \rightharpoonup \mathbf{v} \qquad \Rightarrow \qquad \dim(U_j^{\min}(\mathbf{v})) \leq \liminf_{n\to\infty} \dim(U_j^{\min}(\mathbf{v}_n)).$$

Conclusion:

(1) $\mathcal{T}_{\mathbf{r}}$ is weakly closed.

(2) If $\mathbf{V}$ is a reflexive Banach space,

$$\inf_{\mathbf{u}\in\mathcal{T}_{\mathbf{r}}} \|\mathbf{v}-\mathbf{u}\| = \|\mathbf{v}-\mathbf{u}_{\mathsf{best}}\|$$

has a solution $\mathbf{u}_{\mathsf{best}} \in \mathcal{T}_{\mathbf{r}}$.

Why weak convergence?

There is a sequence $\mathbf{u}_n \in \mathcal{T}_{\mathbf{r}}$ with $\|\mathbf{v}-\mathbf{u}_n\| \to \inf_{\mathbf{u}\in\mathcal{T}_{\mathbf{r}}} \|\mathbf{v}-\mathbf{u}\|$.

In the reflexive case, there is subsequence such that $\mathbf{u}_n \rightharpoonup \mathbf{u}_{\mathsf{best}} \in \mathbf{V}$.

$\dim(U_j^{\min}(\mathbf{u}_n)) \leq r_j \Rightarrow \dim(U_j^{\min}(\mathbf{u}_{\mathsf{best}})) \leq r_j \Rightarrow \mathbf{u}_{\mathsf{best}} \in \mathcal{T}_{\mathbf{r}}$.

7 Topological Tensor Spaces

7.1 Case of Banach Spaces

V_j $(1 \le j \le d)$: normed space with $\|\cdot\|_j$, possibly a Banach space (i.e., complete).

$\mathbf{V}_{\mathsf{alg}} := {}_a\bigotimes_{j=1}^d V_j$ is the algebraic tensor space.

$\|\cdot\|$ chosen norm on $\mathbf{V}_{\mathsf{alg}}$.

Completion of $\mathbf{V}_{\mathsf{alg}}$ w.r.t. $\|\cdot\|$ yields the topological tensor space (Banach tensor space)

$$\mathbf{V} := \mathbf{V}_{\mathsf{top}} := {}_{\|\cdot\|}\bigotimes_{j=1}^d V_j \,.$$

REMARKS: (1) $\mathbf{V}_{\mathsf{top}}$ depends on the choice of $\|\cdot\|$
(2) $\|\cdot\|$ is not fixed by the norms $\|\cdot\|_j$.

7.2 Crossnorms

A necessary condition for reasonable topological tensor spaces is the continuity of the tensor product, i.e.,

$$\left\| \bigotimes_{j=1}^{d} v^{(j)} \right\| \le C \prod_{j=1}^{d} \left\| v^{(j)} \right\|_j$$

for some $C < \infty$ and all $v^{(j)} \in V_j$.

DEFINITION: $\|\cdot\|$ is called a crossnorm if

$$\left\| \bigotimes_{j=1}^{d} v^{(j)} \right\| = \prod_{j=1}^{d} \left\| v^{(j)} \right\|_j .$$

REMARK: There are different crossnorms $\|\cdot\|$ for the same $\|\cdot\|_j$!

7.3 Projective Norm $\|\cdot\|_\wedge$

The *strongest possible norm* is the projective norm (Schatten, Grothendieck), defined by

$$\begin{aligned} \|\mathbf{v}\|_{\wedge(V_1,\ldots,V_d)} &:= \|\mathbf{v}\|_\wedge \\ &:= \inf\left\{ \sum_{i=1}^{m} \prod_{j=1}^{d} \left\| v_i^{(j)} \right\|_j : \mathbf{v} = \sum_{i=1}^{m} \bigotimes_{j=1}^{d} v_i^{(j)} \right\} \end{aligned}$$

for $\mathbf{v} \in {}_a\bigotimes_{j=1}^{d} V_j$.

- $\|\cdot\|_\wedge$ is crossnorm.

- Any norm $\|\cdot\|$ satisfying the continuity requirement satisfies

$$\|\cdot\| \lesssim \|\cdot\|_\wedge .$$

7.4 Duals and Injective Norm $\|\cdot\|_\vee$

The dual space V_j^* is the space of the continuous and linear functions on V_j.
We now require:
also the tensor product $\bigotimes : \times_{j=1}^d V_j^* \to {}_a\bigotimes_{j=1}^d V_j^*$ is continuous, i.e.,

$$\left\|\bigotimes_{j=1}^d \varphi_j\right\|^* \leq C \prod_{j=1}^d \left\|\varphi_j\right\|_j^* \text{ for all } \varphi_j \in V_j^*.$$

- For $\mathbf{v} \in {}_a\bigotimes_{j=1}^d V_j$ define $\|\cdot\|_{\vee(V_1,\ldots V_d)}$ by

$$\|\mathbf{v}\|_{\vee(V_1,\ldots,V_d)} := \|\mathbf{v}\|_\vee := \sup_{\substack{0\neq\varphi_j\in V_j^* \\ 1\leq j\leq d}} \frac{\left|\left(\varphi_1 \otimes \varphi_2 \otimes \ldots \otimes \varphi_d\right)(\mathbf{v})\right|}{\prod_{j=1}^d \|\varphi_j\|_j^*}.$$

- $\|\cdot\|_\vee$ is a crossnorm.

- $\|\cdot\|_\vee$ is the weakest norm with the continuity condition from above.

7.5 Minimal Subspaces, Final Part

We recall $U_j^{\mathsf{min}}(\mathbf{v}) := \left\{\varphi(\mathbf{v}) : \ \varphi \in {}_a\bigotimes_{k\neq j} V_k'\right\}$. Hahn–Banach theorem yields

$$U_j^{\mathsf{min}}(\mathbf{v}) = \left\{\varphi(\mathbf{v}) : \ \varphi \in {}_a\bigotimes\nolimits_{k\neq j} V_k^*\right\}.$$

$\varphi = \bigotimes_{k\neq j} \varphi^{(k)} \in {}_a\bigotimes_{k\neq j} V_k'$ induces the map $\boldsymbol{\varphi}^{[j]} \in L(\mathbf{V}, V_j)$.

(1) If $\|\cdot\| \gtrsim \|\cdot\|_\vee$ then $\varphi \in {}_a\bigotimes_{k\neq j} V_k^*$ implies that $\boldsymbol{\varphi}^{[j]} \in \mathcal{L}(\mathbf{V}, V_j)$ is continuous.

(2) Weak convergence $\mathbf{v}_n \rightharpoonup \mathbf{v}$ implies $\boldsymbol{\varphi}_{[j]}(\mathbf{v}_n) \rightharpoonup \boldsymbol{\varphi}^{[j]}(\mathbf{v})$ in V_j.

Proof. For any $\varphi^{(j)} \in V_j^*$ we have $\varphi^{(j)}(\boldsymbol{\varphi}^{[j]}(\mathbf{v}_n)) = \left(\bigotimes_k \varphi^{(k)}\right)(\mathbf{v}_n)$.
Since $\Phi := \bigotimes_k \varphi^{(k)} \in \mathbf{V}^*$, $\mathbf{v}_n \rightharpoonup \mathbf{v}$ yields $\Phi(\mathbf{v}_n) \to \Phi(\mathbf{v}) = \varphi^{(j)}(\boldsymbol{\varphi}^{[j]}(\mathbf{v}))$.

(3) Let the sequences $(\mathbf{v}_n^{(i)})_{n\in\mathbb{N}}$ for $1 \le i \le N$ converge weakly to linearly independent limits $\mathbf{v}^{(i)} \in \mathbf{V}$ (i.e., $\mathbf{v}_n^{(i)} \rightharpoonup \mathbf{v}^{(i)}$). Then there is an n_0 such that for all $n \ge n_0$, the N-tuples $(\mathbf{v}_n^{(i)} : 1 \le i \le N)$ are linearly independent.

Hence $\quad \mathbf{v}_n \rightharpoonup \mathbf{v} \qquad \Rightarrow \qquad \dim(U_j^{\mathsf{min}}(\mathbf{v})) \le \liminf_{n\to\infty} \dim(U_j^{\mathsf{min}}(\mathbf{v}_n))$.