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

\[ \Phi(u) = \min \]

be a minimisation problem over the whole tensor space \( u \in V \).

Approximation: Choose any format \( F \subset V \). Solve

\[ \Phi(u) = \min \text{ over all } v \in F. \]

This is the minimisation over all parameters in the representation of \( v \in F \).

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

Example: \( \Phi(u) = \langle Au, u \rangle - 2 \langle b, u \rangle \) for the solution of \( Au = b \) with positive definite matrix \( A \).
Example: $\Phi(u) = \|v - u\|^2$ over all $u \in \mathcal{R}_1 = T_{(1,\ldots,1)}$. $v \in V$ is arbitrary.

Ansatz:

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

Necessary condition: $\nabla \Phi(u) = 0$ (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: $\hat{u}^{(2)}$,

\vdots

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 $u_0 = u^{(1)} \otimes \ldots \otimes u^{(d)} \mapsto u_1 = \hat{u}^{(1)} \otimes \ldots \otimes \hat{u}^{(d)}$.

This defines a ALS sequence $\{u_m : m \in \mathbb{N}_0\}$.

Questions: Does $u_m$ converge? To what limit? Convergence speed?
1.2 First Results

Mohlenkamp (2013, Linear Algebra Appl. 438):

- The sequence \( \{u_m : m \in \mathbb{N}_0\} \) is bounded,

\[
\|u_m - u_{m+1}\| \to 0,
\]

- \[
\sum_{m=0}^{\infty} \|u_m - u_{m+1}\|^2 < \infty,
\]

- the set \( S \) of accumulation points of \( \{u_m\} \) is connected and compact.

**Conclusion:** If \( S \) contains an isolated point \( u^* \), it follows that \( u_m \to u^* \).

Note that, in general, the limit may depend on the starting value!
1.3 Study of Examples

1.3.1 Case of \( d = 2 \)

\[
v := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \Phi(u) = \|v - u\|^2.
\]

1) \( u^{**} = 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \) is the global minimiser and an attractive fixed point.

2) \( u^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \) is a fixed point of the ALS iteration:

\[
\Phi(u^* + \delta_1 \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}) = \Phi(u^*) + \|\delta_1\|^2.
\]

But \( \Phi \left( \begin{pmatrix} 1 \\ t \end{pmatrix} \otimes \begin{pmatrix} 1 \\ t \end{pmatrix} \right) = \Phi(u^*) - t^2 \left( 2 - t^2 \right) \)

\( \Rightarrow u^* \) is a saddle point and a repulsive fixed point.

Conclusion: Almost all starting values lead to \( u_m \rightarrow u^{**} \).
1.3.2 Case of $d \geq 3$

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

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

Again $u^* = \otimes^3 a$ and $u^{**} = 2 \otimes^3 b$ are fixed points, $\Phi(u^{**}) < \Phi(u^*)$. But now both are local minima (attractive fixed points)!

Additional saddle point (repulsive fixed point): $u^{***} = c \otimes^3 (a + \frac{1}{2} b)$.

The sequence $\{u_m\}$ corresponding to the starting value

$$u_0 = c^{(0)} \left( a + t_1^{(0)} b \right) \otimes \left( a + t_2^{(0)} b \right) \otimes \left( a + t_3^{(0)} b \right)$$

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

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

(A) $\tau_0 > 2^{-\gamma}$, $\gamma = 5^{1/2} + 1 \Rightarrow u_m \rightarrow u^{**}$ (global minimiser),

(B) $\tau_0 < 2^{-\gamma} \Rightarrow u_m \rightarrow u^*$ (local minimiser),

(C) $\tau_0 = 2^{-\gamma} \Rightarrow u_m \rightarrow u^{***}$ (saddle point, global minimiser on the manifold $\tau = 2^{-\gamma}$).
We recall:

**Conclusion:** If the set of accumulation points of \( \{u_m\} \) contains an isolated point \( u^* \), it follows that \( u_m \rightarrow u^* \).

Wang–Chu (2014): Global convergence for almost all \( u_0 \).

Uschmajew (2015):

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

Łojasiewicz (1965, Ensembles semi-analytiques): If \( \Phi \) 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 \( \Phi(u) = \|v - u\|^2 \) with

\[
v = \otimes^3 a + \lambda (a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a),
\]

\[
a \perp b, \quad \|a\| = \|b\| = 1.
\]

Depending on the value of \( \lambda \) it is shown that the convergence can be

- sublinear (\( \lambda = 1/2 \)),

- linear (\( \lambda < 1/2 \)).

For \( v = \otimes^3 a + 2 \otimes^3 b \), \( u_m \to \otimes^3 a \) 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

\[
v = \sum_{i=1}^{r} \bigotimes_{j=1}^{d} v_{i}^{(j)}, \quad v_{i}^{(j)} \in V_j,
\]

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

\[
\text{rank}(v) := \min\{ m : v \in \mathcal{R}_m \},
\]

we may write

\[
\mathcal{R}_r := \{ v \in V : \text{rank}(v) \leq r \}.
\]

The maximal rank of \( V \) is

\[
\mu := \sup\{ \text{rank}(v) : v \in V \}.
\]

\( \mu < \infty \) 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($v$) is \textit{NP hard} (cf. Håstad 1990).

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

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

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

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

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

$$v = a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a \in \mathcal{R}_3 \setminus \mathcal{R}_2$$

$$v = (b + na) \otimes \left(a + \frac{1}{n}b\right) \otimes a + a \otimes a \otimes (b - na) - \frac{1}{n}b \otimes b \otimes a. \quad \text{for } v_n \in \mathcal{R}_2$$

- \textit{Border rank}: $\text{rank}(v) := \min\{r \in \mathbb{N}_0 : v \in \text{closure}(\mathcal{R}_r)\}$. 

Numerical Instability

In the previous example, the terms of $v_n$ grow like $O(n)$, while the result is of size $O(1)$.

This implies *numerical cancellation*: $\log_2 n$ binary digits of $v_n$ are lost.

We say that the sequence $\{v_n\}$ is unstable.

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

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

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

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

This optimisation problem need not be solvable.

The set of $v^* \in \mathcal{V}$ 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 $2 \times 2$ block matrices can be multiplied as follows:

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

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

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

$m_2 = (a_3 + a_4)b_1$,

$m_3 = a_1(b_2 - b_4)$,

$m_4 = a_4(b_3 - b_1)$,

$m_5 = (a_1 + a_2)b_4$,

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

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

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

\[c_\nu = \sum_{\mu, \lambda=1}^{4} v_{\nu \mu \lambda} a_\mu b_\lambda \quad (1 \leq \nu \leq 4).\]

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

\[v = \sum_{i=1}^{r} \bigotimes_{j=1}^{3} v_i^{(j)} \in \bigotimes_{j=1}^{3} \mathbb{K}^{4}.
\]

The insertion into \(c_\nu = \sum_{\mu, \lambda=1}^{4} v_{\nu \mu \lambda} a_\mu b_\lambda\) yields

\[c_\nu = \sum_{i=1}^{r} \sum_{\mu, \lambda=1}^{4} v_i^{(1)}[\nu] v_i^{(2)}[\mu] v_i^{(3)}[\lambda] a_\mu b_\lambda
\]

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

requiring \(r\) multiplications.

Strassen 1969: \(\text{rank}(v) \leq 7\), Winograd 1971: \(\text{rank}(v) = 7\),
Landsberg 2012: \(\text{rank}(v) = 7\).
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 $v[i_1, \ldots, i_d]$ of $v \in V = \bigotimes_{j=1}^d \mathbb{K}^{n_j}$ is expressed by

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

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 $\text{rank}_{\{1, \ldots, j\}}(v)$.
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** $C(d,(r_j))$ is used in physics:

$$v[i_1 i_2 \cdots i_d] = \text{trace}\{V^{(1)}[i_1] \cdot V^{(2)}[i_2] \cdots \cdot V^{(d-1)}[i_{d-1}] \cdot V^{(d)}[i_d]\}$$

$$= \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} V_{k_dk_1}^{(1)}[i_1] \cdot V_{k_1k_2}^{(2)}[i_2] \cdots \cdot V^{(d-1)}[i_{d-1}] \cdot V_{k_{d-1}k_d}^{(d)}[i_d].$$

**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≠tree are in general not closed.

**Site-independent format** $C_{ind}(d, r)$: $V^{(j)}[i] = V[i]$ and $r_j = r$ for all $j$. 
4.1 Example for $d = 3$, $V = \bigotimes^3 \mathbb{K}^{2 \times 2}$, $r_1 = r_2 = r_3 = 2$ by Harris–Michałek–Sertöz 2018

Let

$$m := \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} E^{(1)}_{k_d,k_1} \otimes E^{(2)}_{k_1k_2} \otimes \cdots \otimes E^{(d-1)}_{k_{d-2}k_{d-1}} \otimes E^{(d)}_{k_{d-1},k_d} \in \bigotimes_{j=1}^d \mathbb{K}^{r_j-1 \times r_j}.$$ 

$E^{(j)}_{pq}$ is the matrix with entries $E^{(j)}_{pq}[k,\ell]=\delta_{pk}\delta_{q\ell}$.

$\{E^{(j)}_{pq} : 1 \leq p \leq r_{j-1}, 1 \leq q \leq r_j\}$ is the canonical basis of $\mathbb{K}^{r_j-1 \times r_j}$.

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

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

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 $1 \leq j \leq d := 3$. 

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

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

where $m = \sum_{k_1=1}^2 \sum_{k_2=1}^2 \sum_{k_3=1}^2 E_{k_3k_1} \otimes E_{k_1k_2} \otimes E_{k_2k_3} \in V$.

Multilinearity yields $v(t) = v_0 + t \cdot v_1 + t^2 \cdot v_2 + t^3 \cdot v_3$ with

$$v_0 = (\otimes^3\psi)(m), \quad v_1 = [\psi \otimes \psi \otimes id + \psi \otimes id \otimes \psi + id \otimes \psi \otimes \psi](m),$$
$$v_2 = [id \otimes id \otimes \psi + id \otimes \psi \otimes id + \psi \otimes id \otimes id](m), \quad v_3 = m.$$

Note that $\psi(E_{ij}) \cdot \psi(E_{k\ell}) = 0$. Since $v_0$ and $v_1$ involve three or two $\psi$ applications, $v_0 = v_1 = 0$ follows.

Evaluation of $v_2$ yields

$$v_2 = E_{21} \otimes E_{11} \otimes E_{12} + E_{22} \otimes E_{21} \otimes E_{12} + E_{11} \otimes E_{12} \otimes E_{21}$$
$$+ E_{21} \otimes E_{12} \otimes E_{22} + E_{12} \otimes E_{21} \otimes E_{11} + E_{12} \otimes E_{22} \otimes E_{21}.$$

$v_0 = v_1 = 0$ allows us to form the limit $v_2 = \lim_{t \to 0} t^{-2}v(t)$. The Lemma states that $t^{-2}v(t) \in C_{\text{ind}}(3, 2)$ for $t > 0$. 
The non-closedness of $C_{\text{ind}}(3, 2)$ will follow from $v_2 \notin C_{\text{ind}}(3, 2)$.

For an indirect proof assume $v_2 \in C_{\text{ind}}(3, 2)$. The Lemma implies that there is some $\phi \in L(K^{2 \times 2}, K^{2 \times 2})$ with $v_2 = (\otimes^3 \phi)(m)$.

It is easy to check that the range of the matricisation $M_1((\otimes^3 \phi)(m)) = \phi M_1(m) (\otimes^2 \phi)^T$ is $K^{2 \times 2}$.

Therefore the map $\phi$ must be surjective.

Since $\phi \in L(K^{2 \times 2}, K^{2 \times 2})$, surjectivity implies injectivity.

Hence $\phi : K^{2 \times 2} \to K^{2 \times 2}$ is a vector space isomorphism and $\otimes^3 \phi : V \to V$ a tensor space isomorphisms. $v_2 = (\otimes^3 \phi)(m) \Rightarrow \text{rank}(v_2) = \text{rank}(m)$.

The representation of $v_2$ yields $\text{rank}(v_2) \leq 6$.

On the other hand, $\text{rank}(m) = 7$ holds for the Strassen tensor $m$.

This contradiction proves that $v_2 \notin C_{\text{ind}}(3, 2)$.

Similarly $v_2 \notin C(3, (2, 2, 2))$ follows (no site-independence).
4.2 Example for $V = \otimes^d \mathbb{C}^2, r_j = 2$

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

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

Site-independent cyclic format $C_{\text{ind}}(d, 2)$, i.e., $r_j = 2$

Result:

$d = 3 : C_{\text{ind}}(3, 2)$ is closed (cf. Harris–Michałek–Sertöz 2018)

$d > 3 : C_{\text{ind}}(d, 2)$ is not closed (cf. Seynnaeve 2018)

For $K = \mathbb{R}, d \geq 3, C_{\text{ind}}(d, 2)$ is not closed (cf. Seynnaeve 2018)
5 Minimal Subspaces

5.1 Tensor Subspace Format

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

$$T_r := \bigcup_{\dim(U_j) \leq r_j} \bigotimes_{j=1}^{d} U_j.$$ 

Question: Is $T_r$ closed?

In the finite-dimensional case, $\dim V_j < \infty$, compactness arguments show that $T_r$ is closed.

What happens in the case of infinite-dimensional Banach spaces $V = \bigotimes_{j=1}^{d} V_j$?
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 \( U_j^{\text{min}}(\mathbf{v}) \) are defined by

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

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

REMARK: (a) \( \dim U_j^{\text{min}}(\mathbf{v}) \leq \text{rank}(\mathbf{v}) < \infty \).
(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: \( U_j^{\text{min}}(\mathbf{v}) \) 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 $\mathcal{M}_j : V = \bigotimes_{k=1}^{d} \mathbb{K}^{n_k} \to \mathbb{K}^{n_j \times n[j]}$ defined by

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

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

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

Consequences:

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

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

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 $V := \bigotimes_{j=1}^{d} V_j$ and $W := \bigotimes_{j=1}^{d} W_j$. Then the sets of linear maps

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

are again vector spaces. They define the tensor space

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

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

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

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

\[
V[j] := V_1 \otimes V_2 \otimes \ldots \otimes V_{j-1} \otimes V_{j+1} \ldots \otimes V_d;
\]
in particular

\[
K \otimes K \otimes \ldots \otimes K \otimes V_j \otimes K \otimes \ldots \otimes K \simeq V_j.
\]

Let functionals \( \varphi_k : V_k \to K \) be given for all \( k \neq j \). Then

\[
\varphi[j] := \varphi_1 \otimes \ldots \otimes \varphi_{j-1} \otimes id \otimes \varphi_{j+1} \otimes \ldots \otimes \varphi_d \in L(V, V_j)
\]
maps \( V \) into \( V_j \).

We identify \( \bigotimes_{k \neq j} \varphi_k \in V'_j \) with \( \varphi[j] \in L(V, V_j) \).
6.1.1 Minimal Subspaces

\[ U_j^{\text{min}}(v) := \left\{ \varphi(v) : \varphi \in a \bigotimes_{k \neq j} V'_k \right\} \]

\[ = \left\{ \varphi(v) : \varphi \in \left( a \bigotimes_{k \neq j} V'_k \right)' \right\}. \]

\( V'_k \): algebraic dual space of \( V_k \).

In the finite-dimensional case, this statement is equivalent to \( U_j^{\text{min}}(v) = \text{range}(M_j(v)) \).

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

\[ \text{rank}_j(v) := \dim(U_j^{\text{min}}(v)). \]
Under rather general assumptions on the norms of $V_j$ and $V$ we shall prove that

$$v_n \rightarrow v \quad \Rightarrow \quad \dim(U_{j\min}(v)) \leq \liminf_{n \rightarrow \infty} \dim(U_{j\min}(v_n)).$$

**Conclusion:**

(1) $\mathcal{T}_r$ is weakly closed.

(2) If $V$ is a reflexive Banach space,

$$\inf_{u \in \mathcal{T}_r} \|v - u\| = \|v - u_{\text{best}}\|
$$

has a solution $u_{\text{best}} \in \mathcal{T}_r$.

**Why weak convergence?**

There is a sequence $u_n \in \mathcal{T}_r$ with $\|v - u_n\| \rightarrow \inf_{u \in \mathcal{T}_r} \|v - u\|$. In the reflexive case, there is subsequence such that $u_n \rightarrow u_{\text{best}} \in V$.

$$\dim(U_{j\min}(u_n)) \leq r_j \Rightarrow \dim(U_{j\min}(u_{\text{best}})) \leq r_j \Rightarrow u_{\text{best}} \in \mathcal{T}_r.$$
7 Topological Tensor Spaces

7.1 Case of Banach Spaces

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

\( V_{\text{alg}} := a \bigotimes_{j=1}^{d} V_j \) is the algebraic tensor space.

\( \| \cdot \| \) chosen norm on \( V_{\text{alg}} \).

Completion of \( V_{\text{alg}} \) w.r.t. \( \| \cdot \| \) yields the topological tensor space (Banach tensor space)

\[ V := V_{\text{top}} := \| \cdot \| \bigotimes_{j=1}^{d} V_j. \]

REMARKS: (1) \( V_{\text{top}} \) 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.,

\[
\left\| \bigotimes_{j=1}^{d} v(j) \right\| \leq C \prod_{j=1}^{d} \left\| v(j) \right\| _j
\]

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

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

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

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

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

$$
\| \cdot \| \lesssim \| \cdot \|_{\wedge}.
$$
7.4 Duals and Injective Norm \(\|\cdot\|_V\)

The dual space \(V^*_j\) is the space of the \textit{continuous} and linear functions on \(V_j\). We now require:

also the tensor product \(\otimes : \times_{j=1}^d V^*_j \to a \otimes_{j=1}^d V^*_j\) is continuous, i.e.,

\[
\left\| \otimes_{j=1}^d \varphi_j \right\|_*^* \leq C \prod_{j=1}^d \|\varphi_j\|_*' \quad \text{for all } \varphi_j \in V^*_j.
\]

**•** For \(v \in a \otimes_{j=1}^d V_j\) define \(\|\cdot\|_\forall(V_1,\ldots,V_d)\) by

\[
\|v\|_\forall(V_1,\ldots,V_d) := \|v\|_\forall := \sup_{0 \neq \varphi_j \in V^*_j, 1 \leq j \leq d} \frac{|(\varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_d)(v)|}{\prod_{j=1}^d \|\varphi_j\|_*'}.
\]

**•** \(\|\cdot\|_\forall\) is a crossnorm.

**•** \(\|\cdot\|_\forall\) is the \textit{weakest norm} with the continuity condition from above.
We recall $U_j^{\text{min}}(v) := \{ \varphi(v) : \varphi \in a \otimes_k \neq j V'_k \}$. Hahn–Banach theorem yields

$$U_j^{\text{min}}(v) = \{ \varphi(v) : \varphi \in a \otimes_k \neq j V^*_k \}.$$ 

$\varphi = \otimes_k \neq j \varphi^{(k)} \in a \otimes_k \neq j V'_k$ induces the map $\varphi[j] \in \mathcal{L}(V_j)$.

(1) If $\| \cdot \| \gtrsim \| \cdot \|_V$ then $\varphi \in a \otimes_k \neq j V^*_k$ implies that $\varphi[j] \in \mathcal{L}(V_j, V_j)$ is continuous.

(2) Weak convergence $v_n \rightharpoonup v$ implies $\varphi[j](v_n) \rightharpoonup \varphi[j](v)$ in $V_j$.

Proof. For any $\varphi(j) \in V^*_j$ we have $\varphi(j)(\varphi[j](v_n)) = (\otimes_k \varphi^{(k)})(v_n)$.

Since $\Phi := \otimes_k \varphi^{(k)} \in V^*$, $v_n \rightharpoonup v$ yields $\Phi(v_n) \rightharpoonup \Phi(v) = \varphi(j)(\varphi[j](v))$.

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

Hence $v_n \rightharpoonup v \quad \Rightarrow \quad \dim(U_j^{\text{min}}(v)) \leq \lim \inf_{n \to \infty} \dim(U_j^{\text{min}}(v_n)).$