Tensor networks, TT (Matrix Product States) and Hierarchical Tucker decomposition

R. Schneider (TUB Matheon)

Setting - Tensors

\[ V_\nu := \mathbb{R}^n \ , \ \mathcal{H}_d = \mathcal{H} := \bigotimes_{\nu=1}^{d} V_\nu \text{ d-fold tensor product Hilbert-s.,} \]

\[ \mathcal{H} \simeq \{ (x_1, \ldots, x_d) \mapsto U(x_1, \ldots, x_d) \in \mathbb{R} : x_i = 1, \ldots, n_i \} \text{.} \]

The function \( U \in \mathcal{H} \) will be called an order \( d \)-tensor.
For notational simplicity, we often consider \( n_i = n \). Here \( x_1, \ldots, x_d \in \{1, \ldots, n\} \) will be called variables or indices.

\[ \mathbf{k} \mapsto U(k_1, \ldots, k_d) = (U_{k_1, \ldots, k_d}) \ , \ k_i = 1 \ldots, n_i \text{.} \]

Or in index (vectorial) notation

\[ U = (U_{k_1, \ldots, k_d})_{n_i}^{n_i} \]

\[ \dim \mathcal{H} = n^d \text{ curse of dimensions!!} \]

E.g. wave function \( \psi (r_1, s_1, \ldots, r_N, s_N) \)
Contracted tensor representations

Espig & Hackbusch & Handschuh & S.

Let $V_\nu := \mathbb{R}^{r_\nu}$

$$K \simeq \{ (k_1, \ldots, k_K) \mapsto V(k_1, \ldots, k_K) : k_i = 1, \ldots, r_i \}.$$
Contracted tensor representations

Definition
Subset $\mathcal{V}^r \subseteq \mathcal{H}$: Let $d < d' \leq d + K$,

$$U \in \mathcal{V}^r \iff U = \sum_{k_1}^{r_1} \ldots \sum_{k_K}^{r_K} \bigotimes_{\nu=1}^{d'} u_\nu(x_\nu, k_\nu(1), k_\nu(2), k_\nu(3))$$

where $k_\nu(i) \in \{k_1, \ldots, k_K\}$, $i = 1, 2, 3$, and

- $u_\nu(x_\nu, k_\nu(1), k_\nu(2))$ if $\nu = 1, \ldots, d$
- $u_\nu(k_\nu(1), k_\nu(2), k_\nu(3))$ if $\nu = d + 1, \ldots, d'$

I.e. the $d'$ tensors $u_\nu(., ., .)$ depends at most on 3 variables!

We will write shortly $u_\nu(x_\nu, k_\nu) \in \mathbb{R}$, keeping in mind that $u_\nu$ depends only on at most three variables $x_\nu, k_\nu(1), k_\nu(2), k_\nu(3)$.

Example
Canonical format: $K = 1$, $d' = d$, $u_\nu(x_\nu, k)$, $k = 1, \ldots, r$. 
Contracted tensor representations

Definition
Subset $\mathcal{V}^r \subseteq \mathcal{H}$: Let $d < d' \leq d + K$,

$$U \in \mathcal{V}^r \iff U = \sum_{k_1}^{r_1} \ldots \sum_{k_K}^{r_K} \bigotimes_{\nu=1}^{d'} u_{\nu}(x_{\nu}, k_{\nu}(1), k_{\nu}(2), (k_{\nu}(3)))$$

the $d'$ tensors $u_{\nu}(..., .., ..)$ depends at most on 3 variables! (Not necessary!)

Complexity:
Let $r := \max\{r_i : i = 1, \ldots, K\}$ then storage requirement is

$$\#DOF' s \leq \max\{d' \times n \times r^2, d' \times r^3\}.$$
Contracted tensor representations

- Due to the multi-linear ansatz, the present extremely general format allow differentiation and local optimization methods! (See e.g. previous talk!)
- Redundancy has not been removed, in general redundancy is enlarged.
- Closedness would be violated in general.

Tensor networks

Definition
A minimal tensor network is a representation in the above format, where each contraction variable $k_i$, $i = 1, \ldots, K$, appear exactly twice. In this case, they can be labeled by $i \sim (\nu, \mu)$, $1 \leq \nu \leq d'$, $1 \leq \mu < \nu$. It is often assumed, that each exterior variable $x_i$ appear at most once!
Diagram - (Graph) of a tensor network:

Diagramm:

- Each node corresponds to a factor $u_\nu$, $\nu = 1, \ldots, d'$,
- each line to a variable $x_\nu$ or a contraction variable $k_\mu$.
- Since $k_\nu$ is an edge connecting 2 tensors $u_{\mu_1}$ and $u_{\mu_2}$, one has to sum over connecting lines (edges).
Tensor networks has been introduced in quantum information theory.

**Example**

1. TT (tensor trains, Oseledets & Tyrtyshnikov) and HT (Hackbusch & Kühn) formats are tensor networks.
2. The canonical format is not (directly) a tensor network.
Tensor formats

- Hierarchical Tucker format
  (HT; Hackbusch/Kühn, Grasedyck, Kressner: Tree-tensor networks)
Tensor formats

- Hierarchical Tucker format
  (HT; Hackbusch/Kühn, Grasedyck, Kressner : Tree-tensor networks)
- Tucker format (Q: MCTDH(F))

\[
U(x_1, \ldots, x_d) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} B(k_1, \ldots, k_d) \bigotimes_{i=1}^{d} U_i(k_i, x_i)
\]
Tensor formats

▶ Hierarchical Tucker format
   (HT; Hackbusch/Kühn, Grasedyck, Kressner : Tree-tensor networks)

▶ Tucker format (Q: MCTDH(F))

▶ Tensor Train (TT-)format
   (Oseledets/Tyrtyshnikov, ≃ MPS-format of quantum physics)

\[
U(x) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{d-1}=1}^{r_{d-1}} \prod_{i=1}^{d} B_i(k_{i-1}, x_i, k_i) = B_1(x_1) \cdots B_d(x_d)
\]
Noteable special case of HT:

**TT format** (Oseledets & Tyrtyshnikov, 2009)
(matrix product states (MPS), Vidal 2003, Schöllwock et al.)

TT tensor $U$ can be written as matrix product form

$$U(x) = U_1(x_1) \cdots U_i(x_i) \cdots U_d(x_d)$$

$$= \sum_{k_1=1}^{r_1} \cdots \sum_{k_{d-1}=1}^{r_{d-1}} U_1(x_1, k_1) U_2(k_1, x_2, k_2) \cdots U_{d-1}(k_{d-2}x_{d-1}, k_{d-1}) U_d(k_{d-1}, x_d, k_d)$$

with matrices $U_i(x_i) = (u_{k_{i-1}}^{k_i}(x_i)) \in \mathbb{R}^{r_{i-1} \times r_i}$, $r_0 = r_d := 1$

Redundancy: $U(x) = U_1(x_1)G G^{-1} U_2(x_2) \cdots U_i(x_i) \cdots U_d(x_d) . $
TT approximations of Friedman data sets

\[ f_2(x_1, x_2, x_3, x_4) = \sqrt{x_1^2 + (x_2x_3 - \frac{1}{x_2x_4})^2}, \]

\[ f_3(x_1, x_2, x_3, x_4) = \tan^{-1}\left( \frac{x_2x_3 - (x_2x_4)^{-1}}{x_1} \right) \]

on 4 – D grid, \( n \) points per dim. \( \leadsto n^4 \) tensor, \( n \in \{3, \ldots, 50\} \).

**full_to_tt** (Oseledets, successive SVDs)

and **MALS** (with \( A = I \))

(Holtz & Rohwedder & S.)
Solution of $-\Delta U = b$

- Dimension $d = 4, \ldots, 128$ varying
- Gridsize $n = 10$
- Right-hand-side $b$ of rank 1
- Solution $U$ has rank 13
Hierarchical Tucker as subspace approximation

1) $D = \{1, \ldots, d\}$, tensor space $V = \bigotimes_{j \in D} V_j$
2) $T_D$ dimension partition tree, vertices $\alpha \in T_D$ are subsets $\alpha \subset T_D$, root: $\alpha = D$
3) $V_\alpha = \bigotimes_{j \in \alpha} V_j$ for $\alpha \in T_D$

4) $U_\alpha \subset V_\alpha$ subspaces of dimension $r_\alpha$ with the characteristic nesting

\[ U_\alpha \subset U_{\alpha_1} \otimes U_{\alpha_2} \quad (\alpha_1, \alpha_2 \text{ sons of } \alpha) \]

5) $v \in U_D$ (w.l.o.g. $U_D = \text{span} \ (v)$).
Subspace approximation: formulation with bases

\[ U_\alpha = \text{span} \{b^{(\alpha)}_i : 1 \leq i \leq r_\alpha \} \]

\[ b^{(\alpha)}_\ell = \sum_{i=1}^{r_{\alpha_1}} \sum_{j=1}^{r_{\alpha_2}} c_{ij}^{(\alpha,\ell)} b^{(\alpha_1)}_i \otimes b^{(\alpha_2)}_j \quad (\alpha_1, \alpha_2 \text{ sons of } \alpha \in T_D). \]

Coefficients \( c_{ij}^{(\alpha,\ell)} \) form the matrices \( C^{(\alpha,\ell)} \).

Final representation of \( \mathbf{v} \) is

\[ \mathbf{v} = \sum_{i=1}^{r_D} c_i^{(D)} b_i^{(D)} , \quad (\text{usually with } r_D = 1). \]
Orthonormal basis and recursive description

We can choose \( \{ \mathbf{b}_i^{(\alpha)} : 1 \leq i \leq r_\alpha \} \) orthonormal basis for \( \alpha \in T_D \).

This is equivalent to
1) At the leaves \( (\alpha = \{ j \}, j \in D) \): \( \{ \mathbf{b}_i^{(j)} : 1 \leq i \leq r_j \} \) is chosen orthonormal
2) The matrices \( \{ C^{(\alpha, \ell)} : 1 \leq \ell \leq r_\alpha \} \) are orthonormal (w.r.t. the Frobenius scalar product).
Hierarchical Tucker tensors

- Canonical decomposition
- Subspace approach (Hackbusch/Kühn, 2009)

(Example: \( d = 5, \mathbf{U}_j \in \mathbb{R}^{n \times k_j}, \mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}} \))
Hierarchical Tucker tensors

- Canonical decomposition: not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)

(Example: $d = 5, U_j \in \mathbb{R}^{n \times k_j}, B_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}}$)
Hierarchical Tucker tensors

- Canonical decomposition not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)

(Example: $d = 5$, $U_j \in \mathbb{R}^{n \times k_j}$, $B_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}}$)
Hierarchical Tucker tensors

- Canonical decomposition: not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)

(Example: \(d = 5\), \(U_j \in \mathbb{R}^{n \times k_j}\), \(B_t \in \mathbb{R}^{k_1 \times k_{t1} \times k_{t2}}\))
Hierarchical Tucker tensors

- Canonical decomposition: not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)

(Example: \(d = 5\), \(\mathbf{U}_j \in \mathbb{R}^{n \times k_j}\), \(\mathbf{B}_t \in \mathbb{R}^{k_t \times k_{t1} \times k_{t2}}\))
Hierarchical Tucker tensors

- Canonical decomposition not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)

(Example: $d = 5$, $U_i \in \mathbb{R}^{n \times k_i}$, $B_t \in \mathbb{R}^{k_t \times k_{t_1} \times k_{t_2}}$)
Hierarchical Tucker tensors

- Canonical decomposition: not closed, no embedded manifold!
- Subspace approach (Hackbusch/Kühn, 2009)

\[
\begin{align*}
\mathbf{B}_{\{1,2,3,4,5\}} & \quad \mathbf{U}_4 \quad \mathbf{U}_5 \\
\mathbf{B}_{\{1,2,3\}} & \quad \mathbf{U}_3 \\
\mathbf{B}_{\{1,2\}} & \quad \mathbf{U}_1 \quad \mathbf{U}_2 \\
\mathbf{U}_{\{1,2\}} & \quad \mathbf{U}_{\{1,2,3\}}
\end{align*}
\]

(Example: \( d = 5, \mathbf{U}_j \in \mathbb{R}^{n \times k_j}, \mathbf{B}_t \in \mathbb{R}^{k_1 \times k_{i_1} \times k_{i_2}} \))
HOSVD bases (Hackbusch)

1) Assume \( r_D = 1 \). Then

\[
C^{(D,1)} = \sum_{\alpha} := \text{diag}\{\sigma_1^{(D)}, \ldots\},
\]

where \( \sigma_i^{(D)} \) are the singular values of \( M_{\alpha_1}(v) \)

2) For non-leaf vertices \( \alpha \in T_D, \alpha \neq D \), we have

\[
\sum_{\ell=1}^{r_{\alpha}} (\sigma_{\ell}^{(\alpha)})^2 \ C(\alpha,\ell) \ C(\alpha,\ell)^H = \sum_{\alpha_1}^2,
\]

\[
\sum_{\ell=1}^{r_{\alpha}} (\sigma_{\ell}^{(\alpha)})^2 \ C(\alpha,\ell)^T \ C(\alpha,\ell) = \sum_{\alpha_2}^2,
\]

where \( \alpha_1, \alpha_2 \) are the first and second son of \( \alpha \in T_D \) and \( \Sigma_{\alpha_i} \) the diagonal of the singular values of \( M_{\alpha_i}(v) \).
Historical remarks – hierarchical Tucker approx.

Physics: (I do not have suff. complete knowledge)
The ideas are well established in many body quantum physics, quantum optics and quantum information theory

1. DMRG: S. White (91)
2. MPS: Roemmer & Ostlund (94), Vidal (03), Verstraete et al.
3. (tree) tensor networks: Verstrate, Cirac, Schollwöck, Wolf, Eisert .... (?)
4. MERA, PEPS, ..... 
5. Chemistry: Multilevel MCTDH $\simeq$ HT : Meyer et al. (2000)

Mathematics: Completely new in tensor product approximation

1. Khoromskij et al. (2006), remark about multi level Tucker
5. Oseledets & Tyrtishnikov (2009): TT and sequential SVD
Successive SVD for e.g. TT tensors
- Vidal (2003), Oseledets (2009), Grasedyck (2009)

1. Matricisation - unfolding

\[ F(x_1, \ldots, x_d) \approx F_{x_1}^{x_2, \ldots, x_d}. \]

low rank approximation up to an error \( \epsilon_1 \), e.g. by SVD.

\[ F_{x_1}^{x_2, \ldots, x_d} = \sum_{k_1 = 0}^{r_1} u_{x_1}^{k_1} v_{x_2, \ldots, x_d}^{k_1}, \]

\[ U_1(x_1, k_1) := u_{x_1}^{k_1}, \quad k_1 = 1, \ldots, r_1. \]

2. Decompose \( V(k_1, x_2, \ldots, x_d) \) via matricisation up to an accuracy \( \epsilon_2 \),

\[ V_{k_1, x_2}^{x_3, \ldots, x_d} = \sum_{k_2}^{r_2} u_{k_1, x_2}^{k_2} v_{x_3, \ldots, x_d}^{k_2}, \]

\[ U_2(k_1, x_2, k_2) := u_{k_1, x_2}^{k_2}. \]

3. repeat with \( V(k_2, x_3, \ldots, x_d) \) until one ends with

\[ \left[ v_{k_{d-1}, x_d} \right] \mapsto U_d(k_{d-1}, x_d). \]
Example

The function

\[ U(x_1, \ldots, x_d) = x_1 + \ldots + x_d \in \mathcal{H} \]

in canonical representation

\[ U(x_1, \ldots, x_d) = x_1 \cdot 1 \cdot \ldots \cdot 1 + 1 \cdot x_2 \cdot 1 \cdot \ldots \cdot 1 + \ldots + 1 \cdot \ldots \cdot 1 \cdot x_d \]

In Tucker format, let

\[ \mathbf{u}_{i,1} = 1, \mathbf{u}_{i,2} = x_i \] be the basis of \( U_i \) (non-orthogonal), \( r_i = 2 \).

\( U \) in TT or MPS representation

\[ U(x) = x_1 + \ldots + x_d = (x_1 \ 1) \begin{pmatrix} 1 & 0 \\ x_2 & 1 \end{pmatrix} \ldots \begin{pmatrix} 1 & 0 \\ x_{d-1} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ x_d \end{pmatrix}. \]

(non-orthogonal), \( r_1 = \ldots = r_{d-1} = 2 \)
Successive SVD for e.g. TT tensors

The above algorithm can be extended to HT tensors!

Error analysis

► Quasi best approximation

\[ \| F(x) - U(x_1) \ldots U_d(x_d) \|_2 \leq \left( \sum_{i=1}^{d-1} \epsilon_i^2 \right)^{1/2} \leq \sqrt{d-1} \inf_{U \in \mathcal{T} \leq r} \| F(x) - U(x) \|_2. \]

► Exact recovery: if \( U \in \mathcal{T} \leq r \), then is will be recovered exactly! (up to rounding errors)
The minimal numbers $r_i$ for a TT or HT representation of a tensor $U$ is well defined.

1. The optimal ranks $r_i, r = (r_1, \ldots, r_{d-1})$, of the TT decomposition are equal to the ranks of the following matrices, $i = 1, \ldots, d-1$

   $$r_i = \text{rank of } A_i = U_{x_1, \ldots, x_i}^{x_{i+1}, \ldots, x_d}.$$

2. If $\nu \subset D$ then $\mu := D \setminus \mu$ the the mode $\nu$-rank $r_{\nu}$ is given by

   $$r_{\nu} = \text{rank of } A_i = U_{x_{\mu}}^{x_{\nu}}.$$

(If $\nu = \{1, 2, 7, 8\}$ and $\mu = \{3, 4, 5, 6\}$ then $x_{\nu} = (x_1, x_2, x_7, x_8)$. )
A tree $T_D$ is characterized by the property, if one remove one edge yields two separate trees.
Observation: Let $A_i$ with ranks $\text{rank} A_i \leq r$. If
\[ \lim_{i \to \infty} \| A_i - A \|_2 = 0 \] then $\text{rank} A \leq r$: ⇒ closedness of Tucker and HT tensor in $T \leq r$ (Falco & Hackbsuch).

\[
T \leq r = \bigcup_{s \leq r} T_s \subset \mathcal{H} \text{ is closed!}
\]
due to Hackbusch & Falco

Landsberg & Ye: If a tensor network has not a tree structure, the set of all tensor of this form need not to be closed!
Landsbergs result - tensor networks

Let \( \mu \) are the vertices and \( s_\mu \in S_\mu \) are incident edges of a vertex \( \mu \), and \( V_\mu \) the vector space attached to it. (it can be empty!)

A vertex \( \mu \) is called super critical if

\[
\dim V_\mu \geq \prod_{s(\mu) \in S(\mu)} r_s(\mu),
\]

and critical if \( \dim V_\mu = \prod_{s(\mu) \in S(\mu)} r_s(\mu) \)

**Lemma**

*Any supercritical vertex can be reduced to a critical one by Tucker decomposition.*

**Theorem (Landsberg-Ye)**

*If a (proper) loop a tensor network contains only (super)-critical vertices, then the network is not (Zariski)-closed.*

I.e. if the ranks are too small! For \( \bigotimes_{i=1}^d \mathbb{C}^2 \), i.e. \( n = 2 \), Landsbergs result has no consequences.
Summary
For Tucker and HT redundancy can be removed (see next talk)

Table: Comparison

<table>
<thead>
<tr>
<th></th>
<th>canonical</th>
<th>Tucker</th>
<th>HT</th>
</tr>
</thead>
<tbody>
<tr>
<td>complexity</td>
<td>$O(n_{dr})$</td>
<td>$O(r^d + n_{dr})$</td>
<td>$O(n_{dr} + d_{r^3})$</td>
</tr>
<tr>
<td></td>
<td>++</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>rank</td>
<td>no</td>
<td>defined $r_T$</td>
<td>defined $r_{HT}, r_T \leq r_{HT}$</td>
</tr>
<tr>
<td></td>
<td>$r_c \geq$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>closedness</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>essential redundancy</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>recovery</td>
<td>??</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>quasi best approx.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>best approx.</td>
<td>no</td>
<td>exist</td>
<td>exist</td>
</tr>
<tr>
<td></td>
<td></td>
<td>but NP hard</td>
<td>but NP hard</td>
</tr>
</tbody>
</table>
Thank you for your attention.

References:
UNDER CONSTRUCTION,
D. Kressner and C. Tobler *htucker A Matlab toolbox for tensor in hierarchical Tucker format* Tech Report, SAM ETH Zürich (2011)

I do not have suff. complete knowledge about physics literatur