Hierarchical Tucker Decomposition

Notation

$\mathcal{I} := \bigtimes_{\mu=1}^{d} I_\mu$ with $I_\mu := \{1,\ldots,n_\mu\}$ for $\mu \in \{1,\ldots,d\}$. For $t \subseteq \{1,\ldots,d\}$, define $t' := \{1,\ldots,d\}\setminus t$ and $\mathcal{I}_t := \bigtimes_{\mu\in t} I_\mu$, $\mathcal{I}_{t'} := \bigtimes_{\mu\in t'} I_\mu$. $A \in \mathbb{R}^{\mathcal{I}}$ denotes a d dimensional tensor where $\otimes$ denotes the outer product and $\otimes_{\mathcal{K}}$ the Kronecker product. The standard scalar product is denoted by $\langle\cdot,\cdot\rangle$, and the corresponding induced norm by $\|\cdot\|$, i.e., the Frobenius norm $\|A\| := \sqrt{\langle A, A\rangle}$.

Definition 1 (t-matricization)**.** *Let $A \in \mathbb{R}^{\mathcal{I}}$ and $t \subseteq \{1,\ldots,d\}$ then $A^{(t)} := \mathcal{M}_t(A)$ where $\mathcal{M}_t(A) : \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}_t\times\mathcal{I}_{t'}}$ and $\mathcal{M}_t(A)_{(i_\mu)_{\mu\in t},(i_\mu)_{\mu\in t'}} := A(i_1,\ldots,i_d)$. We also define $\mathcal{M}_\emptyset := \mathcal{M}_{\{1,\ldots,d\}} := A$.*

Definition 2 (t-multiplication)**.** *Let $A \in \mathbb{R}^{\mathcal{I}}$ and $U_t \in \mathbb{R}^{\mathcal{J}_t\times\mathcal{I}_t}$ then $(U_t \circ_t A)^{(t)} := U_t A^{(t)} \in \mathbb{R}^{\mathcal{J}_t\times\mathcal{I}_{t'}}$. Also for $v = 1,\ldots,d$ and $U_v \in \mathbb{R}^{\mathcal{J}_v\times\mathcal{I}_v}$ we define $(U_1,\ldots,U_d)\circ A := U_1\circ_1\ldots U_d\circ_d A \in \mathbb{R}^{\mathcal{J}_1\times\cdots\times\mathcal{J}_d}$.*

Definition 3. *A* dimension tree *$\mathcal{T}_{\mathcal{I}}$ for dimension $d \in \mathbb{N}$ is a tree such that:*

(i) The root node is $\{1,\ldots,d\}$ and the depth is $\lceil \log_2(d)\rceil =: p$.

(ii) Each non-leaf node has exactly two disjoint children whose union is the parent and all leaves are singletons and lie at level $p-1$ or p.

We denote $\mathcal{T}_{\mathcal{I}}^{l} := \{t \in \mathcal{T}_{\mathcal{I}} \mid level(t) = l\}, \mathcal{L}(\mathcal{T}_{\mathcal{I}})$ for the leaves and $\mathcal{I}(\mathcal{T}_{\mathcal{I}})$ for the interior nodes.

Dimension trees are almost complete binary trees, except that on the penultimate level leaves may appear. For ease of presentation, we restrict to complete binary trees, i.e. all leaves appear on level p. From now on $\mathcal{T}_{\mathcal{I}}$ will always denote a complete binary tree if not mentioned otherwise.

Lemma 1. *(i) On each level ℓ of $\mathcal{T}_{\mathcal{I}}$, the nodes form disjoint subsets of $\{1,\ldots,d\}$ and $|\mathcal{T}_{\mathcal{I}}^{\ell}| = 2^\ell$.*

(ii) $|\mathcal{T}_{\mathcal{I}}| = 2d-1, \quad |\mathcal{L}(\mathcal{T}_{\mathcal{I}})| = d, \quad |\mathcal{I}(\mathcal{T}_{\mathcal{I}})| = d-1.$

Definition 4. *The* hierarchical rank *of a tensor $A \in \mathbb{R}^{\mathcal{I}}$ is the tuple $(k_t)_{t\in\mathcal{T}_{\mathcal{I}}}$ with $k_t := \mathrm{rank}(A^{(t)})$. We also define $\mathcal{H}$-Tucker$((k_t)_{t\in\mathcal{T}_{\mathcal{I}}}) := \{A \in \mathbb{R}^{\mathcal{I}} \mid \mathrm{rank}(A^{(t)}) \le k_t \ \forall t \in \mathcal{T}_{\mathcal{I}}\}$.*

Definition 5. *Let $(k_t)_{t\in\mathcal{T}_{\mathcal{I}}}$ be a given family of non-negative integers. For each node $t \in \mathcal{T}_{\mathcal{I}}$, a matrix $U_t \in \mathbb{R}^{\mathcal{I}_t\times k_t}$ is called a* t-frame, *and the collection $(U_t)_{t\in\mathcal{T}_{\mathcal{I}}}$ is called a* frame tree.

A frame tree is called nested *if for all $t \in \mathcal{I}(\mathcal{T}_{\mathcal{I}})$ with $\mathrm{succ}(t) = \{t_1, t_2\}$, the following holds:*

$$\mathrm{span}\{(U_t)_{:,i} \mid 1 \le i \le k_t\} \subseteq \mathrm{span}\{(U_{t_1})_{:,j} \otimes (U_{t_2})_{:,\ell} \mid 1 \le j \le k_{t_1},\ 1 \le \ell \le k_{t_2}\}.$$

Then there exists a transfer tensor *$B_t \in \mathbb{R}^{k_t\times k_{t_1}\times k_{t_2}}$ such that:*

$$(U_t)_{:,i} = \sum_{j=1}^{k_{t_1}}\sum_{\ell=1}^{k_{t_2}} (B_t)_{i,j,\ell}\cdot (U_{t_1})_{:,j} \otimes (U_{t_2})_{:,\ell}, \quad \forall i = 1,\ldots,k_t.$$

Definition 6. *Let $A \in \mathcal{H}$-Tucker$((k_t)_{t\in\mathcal{T}_{\mathcal{I}}})$ and $(U_t)_{t\in\mathcal{T}_{\mathcal{I}}}$ be a nested frame tree such that* $\mathrm{image}(A^{(t)}) = \mathrm{image}(U_t)$ *for all $t \in \mathcal{T}_{\mathcal{I}}$ and $U_{\{1,\ldots,d\}} := A$. Then $\big((B_t)_{t\in\mathcal{I}(\mathcal{T}_{\mathcal{I}})},\ (U_t)_{t\in\mathcal{L}(\mathcal{T}_{\mathcal{I}})}\big)$ is called a* hierarchical Tucker representation *of A and $(k_t)_{t\in\mathcal{T}_{\mathcal{I}}}$ is called the* hierarchical representation rank *of A.*

The representation of a tensor $A \in \mathcal{H}$-Tucker$((k_t)_{t\in\mathcal{T}_{\mathcal{I}}})$ in the hierarchical tucker format with an orthogonal frame tree and minimal ranks k_t is unique up to orthogonal transformations of the t-frames.

Lemma 2. *Let $A \in \mathcal{H}$-Tucker$((k_t)_{t\in\mathcal{T}_{\mathcal{I}}})$ be given in the HT representation $\big((B_t)_{t\in\mathcal{I}(\mathcal{T}_{\mathcal{I}})},\ (U_t)_{t\in\mathcal{L}(\mathcal{T}_{\mathcal{I}})}\big)$, then: Storage$\big((B_t)_{t\in\mathcal{I}(\mathcal{T}_{\mathcal{I}})},\ (U_t)_{t\in\mathcal{L}(\mathcal{T}_{\mathcal{I}})}\big) \le (d-1)k^3 + k\sum_{\mu=1}^{d} n_\mu, \quad k := \max_{t\in\mathcal{T}_{\mathcal{I}}} k_t$.*

The storage behaves linearly in the dimension d provided that k is uniformly bounded.

Definition 7. *For $t \in \mathcal{T}_\mathcal{I}$ and t not the root, let U_t be an orthonormal t-frame. Define $\pi_t(A) := U_t U_t^T \circ_t A$, implying $\pi_t(A)^{(t)} = U_t U_t^\top A^{(t)}$. For $t = \{1, \ldots, d\}$, set $\pi_t(A) := A$.*

Lemma 3. *Let π_t, π_s be orthogonal projections as in Definition 7. Then the following holds:*

(i) $\|A + \pi_t \pi_s A\|^2 \leq \|A + \pi_t A\|^2 + \|A + \pi_s A\|^2$.

(ii) $\left\|A + \prod_{t \in \mathcal{T}_\mathcal{I}} \pi_t A\right\|^2 \leq \sum_{t \in \mathcal{T}_\mathcal{I}} \|A + \pi_t A\|^2$.

Theorem 1. *Let A^{best} denote the best approx. of A in $\mathcal{H}$-Tucker$((k_t)_{t \in \mathcal{T}_\mathcal{I}})$ and let $\psi_{t,k}(A) \in \mathbb{R}^{\mathcal{I}_t \times k}$ denote the matrix whose columns are the left singular vectors of $A^{(t)}$ corresponding to the k largest singular values $\sigma_{t,i}$ of $A^{(t)}$. Let π_t be based on $U_t := \psi_{t,k_t}(A)$. For any order of projections π_t:*

$$\left\| A - \prod_{t \in \mathcal{T}_\mathcal{I}} \pi_t A \right\| \leq \left(\sum_{t \in \mathcal{T}_\mathcal{I}} \sum_{i > k_t} \sigma_{t,i}^2 \right)^{1/2} \leq \sqrt{2d-2} \|A - A^{best}\|.$$

Definition 8 (root-to-leaves truncation)**.** *Let π_t be based on $U_t := \psi_{t,k_t}(A)$. Then $A_\mathcal{H} := \prod_{t \in \mathcal{T}_\mathcal{I}^p} \pi_t \cdots \prod_{t \in \mathcal{T}_\mathcal{I}^1} \pi_t A$ is called the root-to-leaves truncation and $A_\mathcal{H} \in \mathcal{H}$-Tucker$((k_t)_{t \in \mathcal{T}_\mathcal{I}})$.*

Definition 9 (leaves-to-root truncation)**.** *Let $A_{\tilde{\mathcal{H}},p} := A$, and for $t \in \mathcal{T}_\mathcal{I}^\ell$, let π_t be based on $U_t := \psi_{t,k_t}(A_{\tilde{\mathcal{H}},\ell+1})$, where $A_{\tilde{\mathcal{H}},\ell} := \prod_{t \in \mathcal{T}_\mathcal{I}^\ell} \pi_t A_{\tilde{\mathcal{H}},\ell+1}$. Then $A_{\tilde{\mathcal{H}}} := A_{\tilde{\mathcal{H}},1} = \prod_{t \in \mathcal{T}_\mathcal{I}^1} \pi_t \cdots \prod_{t \in \mathcal{T}_\mathcal{I}^p} \pi_t A$ is called the* leaves-to-root truncation *and $A_{\tilde{\mathcal{H}}}$ belongs to $\mathcal{H}$-Tucker$((k_t)_{t \in \mathcal{T}_\mathcal{I}})$.*

Algorithm 1 Root-to-Leaves Truncation

Require: Tensor $A \in \mathbb{R}^\mathcal{I}$, tree $\mathcal{T}_\mathcal{I}$, ranks $(k_t)_{t \in \mathcal{T}_\mathcal{I}}$
for all leaves $t \in \mathcal{L}(\mathcal{T}_\mathcal{I})$ **do**
  Compute SVD of $A^{(t)}$; store dominant k_t left singular vectors in columns of U_t.
end for
for $\ell = p-1, \ldots, 0$ **do**
  for all $t \in \mathcal{I}(\mathcal{T}_\mathcal{I})$ on level ℓ **do**
    Compute SVD of $A^{(t)}$; store dominant k_t left singular vectors in U_t.
    Let U_{t_1}, U_{t_2} denote the frames of t's children. Compute the transfer tensor:
    $(B_t)_{i,j,\nu} := \langle (U_t)_{:,i}, (U_{t_1})_{:,j} \otimes (U_{t_2})_{:,\nu} \rangle$
  end for
end for
Compute root transfer tensor:
$(B_{\{1,\ldots,d\}})_{1,j,\nu} := \langle A, (U_{t_1})_{:,j} \otimes (U_{t_2})_{:,\nu} \rangle$
return $\mathcal{H}$-Tucker representation $\big((B_t)_{t \in \mathcal{I}(\mathcal{T}_\mathcal{I})},\ (U_t)_{t \in \mathcal{L}(\mathcal{T}_\mathcal{I})}\big)$ for $A_\mathcal{H}$

Algorithm 2 Leaves-to-Root Truncation

Require: Tensor $A \in \mathbb{R}^\mathcal{I}$, tree $\mathcal{T}_\mathcal{I}$, ranks $(k_t)_{t \in \mathcal{T}_\mathcal{I}}$
for all leaves $t \in \mathcal{L}(\mathcal{T}_\mathcal{I})$ **do**
  Compute SVD of $A^{(t)}$; store dominant k_t left singular vectors in columns of U_t.
end for
Compute core $C_p := (U_1^\top, \ldots, U_d^\top) \circ A$
for $\ell = p-1, \ldots, 0$ **do**
  Initialize $C_\ell := C_{\ell+1}$
  for all $t \in \mathcal{I}(\mathcal{T}_\mathcal{I})$ on level ℓ **do**
    Compute SVD of $(C_{\ell+1})^{(t)}$; store dominant k_t left singular vectors in U_t.
    Let U_{t_1}, U_{t_2} denote the frames of t's children. Compute the transfer tensor:
    $(B_t)_{i,j,\nu} := \langle (U_t)_{:,i}, (U_{t_1})_{:,j} \otimes (U_{t_2})_{:,\nu} \rangle$
    Update the core $C_\ell := U_t^\top \circ_t C_\ell$
  end for
end for
return $\mathcal{H}$-Tucker representation $\big((B_t)_{t \in \mathcal{I}(\mathcal{T}_\mathcal{I})},\ (U_t)_{t \in \mathcal{L}(\mathcal{T}_\mathcal{I})}\big)$ for $A_{\tilde{\mathcal{H}}}$

Lemma 4 (Complexity of Algorithm 1)**.** *Let $N := \prod_{\mu=1}^d n_\mu$. The complexity behaves like $\mathcal{O}(N^{3/2})$.*

Lemma 5 (Complexity of Algorithm 2)**.** *The complexity behaves like $\mathcal{O}\left(N \cdot \sum_{\mu=1}^d n_\mu + dk^{d+2}\right)$.*

Theorem 2. *Let A^{best} denote the best approximation of A in $\mathcal{H}$-Tucker$((k_t)_{t \in \mathcal{T}_\mathcal{I}})$ then:*

$$\|A - A_{\tilde{\mathcal{H}}}\| \leq (2 + \sqrt{2})\sqrt{d} \|A - A^{best}\|.$$

References

[1] L. Grasedyck, *Hierarchical Singular Value Decomposition of Tensors*, SIAM Journal on Matrix Analysis and Applications, vol. 31, no. 4, pp. 2029–2054, 2010.

[2] D. Kressner and C. Tobler, *htucker—A MATLAB Toolbox for Tensors in Hierarchical Tucker Format*, ACM Transactions on Mathematical Software, vol. 40, no. 3, Article 22, pp. 1–22, 2014.