Tensor Networks (IN2388)

Christian B. Mendl, summer term 2022 25.4.2022

1. Introduction

(see corresponding slides)

2. Mathematical framework and graphical diagrams

2.1 Linear algebra fundamentals

Vector spaces: $\mathbb{R}^n$, $\mathbb{C}^n$, $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{C}^n$

$C([0,1], \mathbb{C})$: space of continuous functions $f: [0,1] \to \mathbb{C}$

Matrices: $A \in \mathbb{C}^{m \times n}$: $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$

Matrix-vector product:

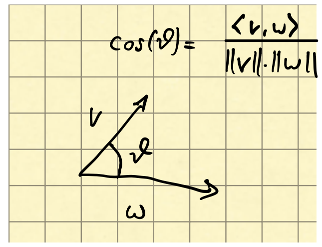
$$A \cdot v = \begin{pmatrix} a_{11} v_1 + a_{12} v_2 + \dots a_{1n} v_n \\ \vdots \\ a_{m1} v_1 + a_{m2} v_2 + \dots a_{mn} v_n \end{pmatrix} \in \mathbb{C}^m$$

A can be interpreted as linear operator from $\mathbb{C}^n \to \mathbb{C}^m$

Remark: We will formulate definitions, theorems, ... in terms of complex numbers; this includes real numbers as special case.

Inner product and norm on a vector space V:

$$\langle \cdot , \cdot \rangle : V \times V \to \mathbb{C}$$



Can be defined abstractly,

in this course:

complex conjugate

$$\langle v, w \rangle = \sum_{j=1}^{n} v_j^* w_j \quad \text{for all } v, w \in \mathbb{C}^n$$

Note: $\langle \cdot , \cdot \rangle$ is linear in its second argument,

but anti-linear in its first: $\langle \alpha \cdot v, w \rangle = \alpha^* \langle v, w \rangle$

$\langle \cdot , \cdot \rangle$ induces a norm on V via $\|v\| = \sqrt{\langle v, v \rangle}$

Cauchy-Schwarz inequality:

for all $v, w \in V$: $|\langle v, w \rangle| \leq \|v\| \cdot \|w\|$

Example on $C([0,1], \mathbb{C})$: $\langle f, g \rangle = \int_0^1 f(x)^* g(x) \, dx$

Adjoint (conjugate transpose) of a matrix:

$$A^\dagger = (A^*)^T$$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{C}^{m \times n} \rightsquigarrow A^\dagger = \begin{pmatrix} a_{11}^* & \cdots & a_{m1}^* \\ \vdots & & \vdots \\ a_{1n}^* & \cdots & a_{mn}^* \end{pmatrix} \in \mathbb{C}^{n \times m}$$

Note: $\langle v, A w \rangle = \langle A^\dagger v, w \rangle \quad \forall v \in \mathbb{C}^m, w \in \mathbb{C}^n$

A matrix $A \in \mathbb{C}^{n \times n}$ is called Hermitian (or self-adjoint)

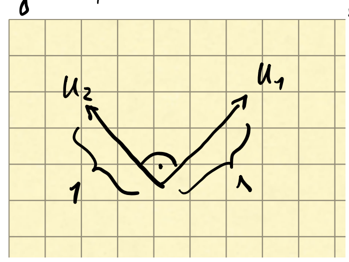
if $A^\dagger = A$

Unitary matrices: $U \in \mathbb{C}^{n \times n}$ is called unitary if $U^\dagger U = I$ (identity matrix)

i.e. $U^\dagger$ is the inverse of U

Intuition: $U = (u_1 | u_2 | \ldots | u_n)$

consists of orthonormal column vectors



U describes a change of basis which preserves the length of vectors.

Note: $U^\dagger U = I \iff U U^\dagger = I$:

U is unitary if and only if $U^\dagger$ is unitary.

<u>Normal</u> matrix: $A \in \mathbb{C}^{n\times n}$ is called normal if it commutes with its adjoint: $A^\dagger A = A A^\dagger$

In particular: every unitary and every Hermitian matrix is normal.

<u>Eigenvalues and -vectors</u>: Let $A \in \mathbb{C}^{n\times n}$. Then a non-zero vector $v \in \mathbb{C}^n$ is called an eigenvector of A with corresponding eigenvalue $\lambda \in \mathbb{C}$ if

$$A v = \lambda v$$

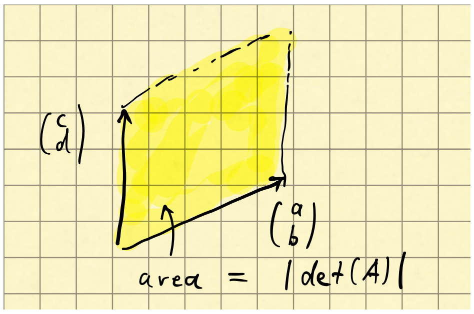
<u>Determinant</u> of a square matrix $A \in \mathbb{C}^{n\times n}$:

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{j,\sigma(j)}$$

with S_n the group of all permutations of $\{1, .. n\}$

Example for $n=2$: $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $S_2 = \{ id, \underset{1\leftrightarrow 2}{(1,2)} \}$

$$\det(A) = \underbrace{a \cdot d}_{\sigma = id} - \underbrace{b \cdot c}_{\sigma = 1\leftrightarrow 2}$$



Properties:

- A is invertible (column or row vectors are linearly independent) if and only if $\det(A) \neq 0$
- $\det(A^T) = \det(A)$
- For all $A, B \in \mathbb{C}^{n\times n}$:

$$\det(A \cdot B) = \det(A) \cdot \det(B)$$

2.5.2022

Relevance for eigenvalues:

$Av = \lambda v \iff (\lambda I - A)v = 0 \iff \lambda I - A$ is <u>not</u> invertible

$$\iff \underbrace{\det(\lambda I - A)}_{\text{characteristic polynomial of } A} = 0 \quad (\lambda \text{ regarded as variable})$$

Fundamental theorem of algebra guarantees that there exist n complex roots of the characteristic polynomial ($A \in \mathbb{C}^{n\times n}$)

Example: $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \rightsquigarrow \det(\lambda I - A) = \det\begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix} =$

$$= \lambda^2 + 1 \overset{!}{=} 0$$

$$\iff \lambda = \pm i$$

Theorem (Spectral decomposition). Any normal matrix $A \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable, i.e. there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ and eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ such that

$$A \cdot U = U \underbrace{\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}}_{\text{diagonal matrix}},$$

equivalently: $A = U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^{\dagger}$

Conversely, every matrix representable in this form is normal.

Remark: The column vectors of $U = (u_1 | \dots | u_n)$ are a basis of eigenvectors of A, since $A \cdot U = U \cdot \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ is stating that $A u_j = \lambda_j u_j$ for $j = 1, \dots, n$.

Sketch of proof (for "$\Rightarrow$") : Let A be normal.

(i) Every linear subspace $V \subseteq \mathbb{C}^n$ with dimension ≥ 1 which is left invariant by A, i.e. $A v \in V$ for all $v \in V$, contains at least one eigenvector of A: namely, regarding A as linear operator on V, the corresponding characteristic polynomial must contain at least one root.

(ii) If v is an eigenvector of A with eigenvalue λ, then v is also an eigenvector of A^+ with eigenvalue λ^*:

$$0 = \| (\lambda I - A) v \|^2 = \langle (\lambda I - A) v, (\lambda I - A) v \rangle =$$

$$= \langle v, (\lambda I - A)^+ (\lambda I - A) v \rangle \overset{A \text{ normal}}{=} \langle v, (\lambda I - A)(\lambda^* I - A^+) v \rangle$$

$$|\lambda|^2 I - \lambda^* A - \lambda A^+ + A^+ A$$

$$= \| \underbrace{(\lambda^* I - A^+) v}_{\text{has to be zero}} \|^2 \quad \Leftrightarrow \quad \lambda^* v = A^+ v$$

(iii) Let λ be an eigenvalue of A and V_λ the subspace of corresponding eigenvectors : $V_\lambda = \{ v \in \mathbb{C}^n : A v = \lambda v \}$

Then A and A^+ leave $V_\lambda^\perp$ $\leftarrow$ orthogonal complement invariant, namely:

$\forall w \in V_\lambda^\perp$, $v \in V_\lambda$:

$$\langle v, A w \rangle = \langle A^+ v, w \rangle \overset{(ii)}{=} \langle \lambda^* v, w \rangle = \lambda \underbrace{\langle v, w \rangle}_{= 0 \text{ since } w \in V_\lambda^\perp} = 0$$

$\rightsquigarrow A w \in V_\lambda^\perp$, similarly $A^+ w \in V_\lambda^\perp$

(iv) Now we apply the above steps recursively to A restricted to $V_\lambda^\perp$, starting from any eigenvalue λ on $V = \mathbb{C}^n$.
The adjoint of A on $V_\lambda^\perp$ is still A^+ (restricted to $V_\lambda^\perp$) and hence A on $V_\lambda^\perp$ is likewise normal.

$\square$

Remark: for general matrices, eigenvectors might not form a basis, eigenvectors of different eigenvalues are not necessarily orthogonal to each other.

Useful criterion which guarantees the existence of a single leading eigenvalue: Perron-Frobenius theorem.

The <u>spectral radius</u> of a square matrix $A \in \mathbb{C}^{n\times n}$ is the largest absolute value of the eigenvalues of A:

$$\rho(A) := \max\{|\lambda_1|, \dots |\lambda_n|\}, \quad \text{with } \lambda_1, \dots \lambda_n \text{ the eigenvalues of } A$$

Immediate corrollary of the spectral theorem:
any Hermitian matrix is unitarily diagonalizable, and all its eigenvalues are real:

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = U^\dagger A U \quad \leftarrow \text{Hermitian}$$

A Hermitian matrix $A \in \mathbb{C}^{n\times n}$ is called <u>positive semidefinite</u> if $\langle v, Av\rangle \geq 0$ for all $v \in \mathbb{C}^n$.

Thus the eigenvalues of such a matrix are likewise non-negative, since for any eigenpair (v, λ):

$$\lambda \underbrace{\|v\|^2}_{>0} = \lambda\langle v, v\rangle = \langle v, \lambda v\rangle = \langle v, Av\rangle \geq 0.$$

Theorem (Singular value decomposition, SVD)
Let $A \in \mathbb{C}^{n\times n}$ be a square matrix. Then there exist unitary matrices $U, V \in \mathbb{C}^{n\times n}$, and a list of real numbers $\sigma_1 \dots \sigma_n$ with $\sigma_1 \geq \sigma_2 \geq \dots \sigma_n \geq 0$ called <u>singular values</u>, such that
$$A = U \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} V^+ .$$

$\checkmark \; (A^+A)^+ = A^+(A^+)^+ = A^+A$

Proof: A^+A is Hermitian and positive semidefinite, since
$\forall v \in \mathbb{C}^n$: $\langle v, A^+A v\rangle = \langle Av, Av\rangle = \|Av\|^2 \geq 0$

$\rightsquigarrow$ by the spectral theorem, A^+A is unitarily diagonalizable by some unitary matrix $V \in \mathbb{C}^{n\times n}$.

We denote the corresponding eigenvalues by $\sigma_1^2, \dots \sigma_n^2$, $\sigma_j \geq 0 \;\forall j$

w.l.o.g. $\sigma_j \geq \sigma_{j+1} \;\forall j$:
$$A^+A = V \cdot \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} V^+, \qquad V = (v_1 | \dots | v_n)$$

Denote the number of strictly positive eigenvalues by k.

For all $j = 1, \dots k$, set $u_j = \frac{1}{\sigma_j} A v_j$

The u_j's are orthonormal, since:
$$\langle u_j, u_\ell\rangle = \frac{1}{\sigma_j \sigma_\ell} \langle A v_j, A v_\ell\rangle = \frac{1}{\sigma_j \sigma_\ell} \langle v_j, \underbrace{A^+A v_\ell}_{\sigma_\ell^2 v_\ell}\rangle = \delta_{j\ell}$$

In case $k < n$, extend $(u_1 \dots u_k)$ to an orthonormal basis of $\mathbb{C}^n$.

By construction, $A v_j = \sigma_j u_j \quad \forall j = 1 \dots n$,
which can be summarized as
$$A \cdot V = U \cdot \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}$$

Multiplying from the right by V^+ yields the SVD of A. $\square$

Note that the SVD exists for any matrix A. 9.5.2022

Remark: The SVD can be generalized to non-square matrices as follows:

Let $A \in \mathbb{C}^{m \times n}$, set $k = \min(m,n)$. Then there exist linear isometries $U \in \mathbb{C}^{m \times k}$ and $V \in \mathbb{C}^{n \times k}$, and real non-negative singular values $\sigma_1, \dots \sigma_k$ with $\sigma_1 \geq \sigma_2 \geq \dots \sigma_k \geq 0$ such that

$$A = U \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} V^\dagger .$$

Note: dimensions are compatible:

$$U \in \mathbb{C}^{m \times k}, \quad \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} \in \mathbb{C}^{k \times k}, \quad V^\dagger \in \mathbb{C}^{k \times n}$$

Isometries are generalizations of unitary matrices:

$U \in \mathbb{C}^{m \times n}$ with $m \geq n$ is called an <u>isometry</u> if

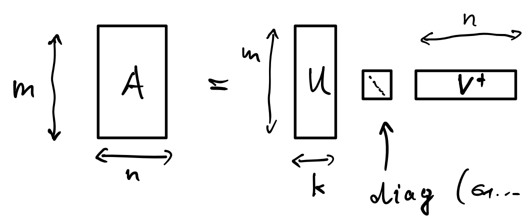
$$U^\dagger U = I_n .$$

The columns of $U = (u_1 | \dots | u_n)$ are orthonormal ($\langle u_j, u_\ell \rangle = \delta_{j\ell}$), but if $m > n$ they cannot form a basis of $\mathbb{C}^m$; instead, they span a proper subspace.

The <u>rank</u> of $A \in \mathbb{C}^{m \times n}$ is the dimension of the vector space generated by the column vectors of A.

rank(A) is equal to the number of non-zero singular values.

SVD becomes a low-rank factorization if many singular values are zero (and can hence be omitted):



diag$(\sigma_1 \dots \sigma_k)$

only keep the non-zero singular values

<u>Matrix functions</u>: Let $f: \mathbb{C} \to \mathbb{C}$ be an arbitrary function and $A \in \mathbb{C}^{n \times n}$ be normal, with spectral decomposition

$$A = U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^\dagger.$$

Then we define

$$f(A) = U \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix} U^\dagger.$$

Equivalently, if f admits a power series expansion

$$f(x) = \sum_{j=0}^{\infty} q_j \, x^j \quad \text{for all } x \in \mathbb{C}$$

can be finite as well, i.e. $q_{j+1} = 0, \ q_{j+2} = 0 \dots$

then $$f(A) = \sum_{j=0}^{\infty} q_j \, A^j$$

j-fold product of A with itself

Exercise: why equivalent?

Hint: insert spectral decomposition

Examples:

- $e^A \equiv \exp(A) = \sum_{j=0}^{\infty} \frac{1}{j!} A^j$

  Note: $\frac{d}{dt} e^{tA} = A \cdot e^{tA}$ for $t \in \mathbb{R}$

  $e^{A+B} \underset{\text{in general}}{\neq} e^A \cdot e^B$

- $\sqrt{A} \in \mathbb{C}^{n \times n}$ satisfies $\sqrt{A} \cdot \sqrt{A} = A$ (as expected)

<u>Kronecker product</u> of two matrices $A \in \mathbb{C}^{m \times n}$, $B \in \mathbb{C}^{p \times q}$:

$$A \otimes B = \begin{pmatrix} a_{11} B & \cdots & a_{1n} B \\ a_{21} B & & \vdots \\ \vdots & & \\ a_{m1} B & \cdots & a_{mn} B \end{pmatrix} \in \mathbb{C}^{mp \times nq}$$

Similarly: Kronecker product of two vectors:

$v \in \mathbb{C}^n$, $w \in \mathbb{C}^q$:

$$v \otimes w = \begin{pmatrix} v_1 w \\ \vdots \\ v_n w \end{pmatrix} \in \mathbb{C}^{nq}$$

Properties:

- For $A \in \mathbb{C}^{m \times n}$, $C \in \mathbb{C}^{n \times o}$, $B \in \mathbb{C}^{p \times q}$, $D \in \mathbb{C}^{q \times r}$:

  $$(A \otimes B) \cdot (C \otimes D) = (A \cdot C) \otimes (B \cdot D)$$

  usual matrix-matrix product

  Similarly for matrix-vector:

  $$(A \otimes B) \cdot (v \otimes w) = (A v) \otimes (B \cdot w)$$

- $I_m \otimes I_n = I_{m \cdot n}$

  $m \times m$ identity

- $\mathrm{tr}[A \otimes B] = \mathrm{tr}[A] \cdot \mathrm{tr}[B]$

  matrix trace: sum of diagonal entries

Vectorization of matrices:

Entries of $A \in \mathbb{C}^{m \times n}$ interpreted as vector of length $m \cdot n$:

$$\mathrm{vec}(A) \in \mathbb{C}^{m \cdot n}$$

Two options:

- row-major ordering: row by row, e.g.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ \vdots \\ a_{33} \end{pmatrix}$$

  used in C/C++,
  default in Python / NumPy ($\rightarrow$ this lecture),
  Mathematica, ...

- colum-major ordering:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{12} \\ \vdots \\ a_{33} \end{pmatrix}$$

  used by Fortran, Matlab, Julia, ...

2.2 Tensors and graphical tensor diagrams

General tensor $T \in \mathbb{C}^{n_1 \times n_2 \cdots \times n_d}$:

generalization of vectors and matrices to multi-dimensional arrays.

The number of dimensions of a tensor T is called the <u>order</u> or <u>degree</u> of T.

(An alternative name is "rank", but might lead to confusion with rank of a matrix.)

Python / NumPy : a.ndim for "a" an array
= len(a.shape)
a.shape = $(n_1, \dots, n_d)$

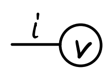
Examples :
- a vector is a tensor of degree 1
- a matrix is a tensor of degree 2

Notation : $T = (t_{i_1 \dots i_d})_{i_1 \dots i_d} \in \mathbb{C}^{n_1 \times \dots \times n_d}$

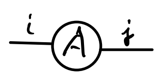
<u>Graphical representation:</u>

Draw a leg for each dimension (or index) :

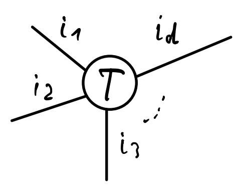
vector $v \in \mathbb{C}^n$:



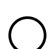
matrix $A \in \mathbb{C}^{m \times n}$:



general tensor T:



tensor of degree 0:



single (complex) number

identity map :

Contraction: (generalization of matrix-vector or matrix-matrix multiplication):

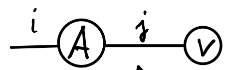
summation over a shared index

Graphical form: connect corresponding legs

Examples: $A \in \mathbb{C}^{m \times n}, \quad v \in \mathbb{C}^n$

$$(A \cdot v)_i = \sum_{j=1}^{n} a_{ij} v_j , \quad i = 1, \dots, m$$

$A \cdot v \hat{=}$



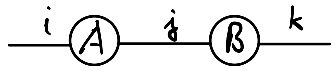
summation over connected leg

matrix-matrix product:

$B \in \mathbb{C}^{n \times p}$:

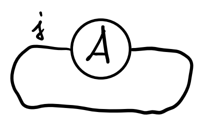
$$(A \cdot B)_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} \quad \forall\, i = 1, \dots m, \quad k = 1, \dots p$$

$A \cdot B \hat{=}$



For $A \in \mathbb{C}^{n \times n}$:

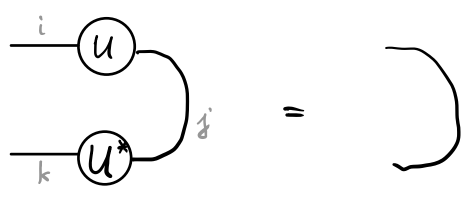
$$\mathrm{tr}[A] = \sum_{j=1}^{n} a_{jj} \hat{=}$$



trace

Can express the unitary property of $U \in \mathbb{C}^{n \times n}$ as

$U U^{\dagger} = I$:



identity

$(U \cdot U^{\dagger})_{ik} =$

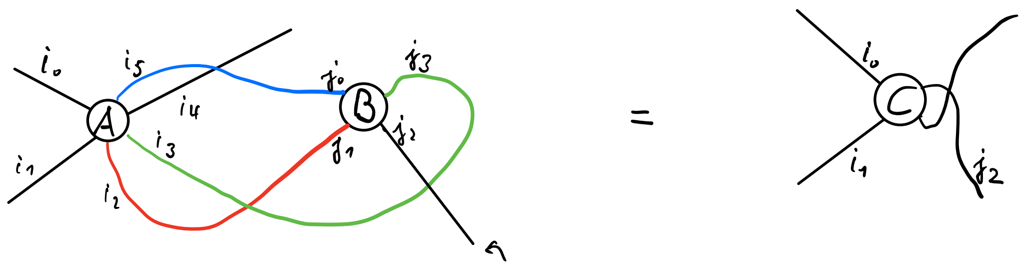
$= \sum_j u_{ij} u^*_{kj}$

transposition taken into account by how legs are connected

16. 5. 2022

General contraction of two tensors:

Example:



"open" (uncontracted) leg

A has degree 6   B has degree 4

$$\underbrace{c_{i_0, i_1, i_4,}}_{\text{open indices of } A}\underbrace{{}_{j_2}}_{\text{open indices of } B} = \sum_{k,\ell,m} a_{i_0 i_1 k \ell i_4 m}\, b_{m k j_2 \ell}$$

NumPy:

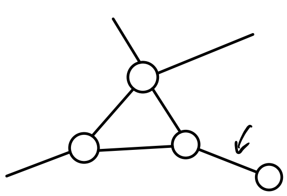
`c = np.tensordot(a, b, axes=([2,3,5], [1,3,0]))`

`c = np.einsum("ijklmn,nkol->ijmo", a, b)`

$$c_{ijmo} = \sum_{k,\ell,n} a_{ijk\ell mn}\, b_{nko\ell}$$

$$c = \text{np.einsum}(a, (0,1,2,3,4,5), b, (\cancel{6}^{\,5}, \cancel{7}^{\,2}, 8, \cancel{9}^{\,3}), (0,1,4,8))$$

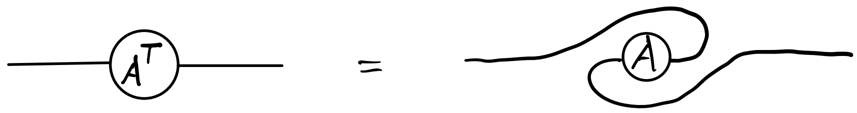
General tensor network:



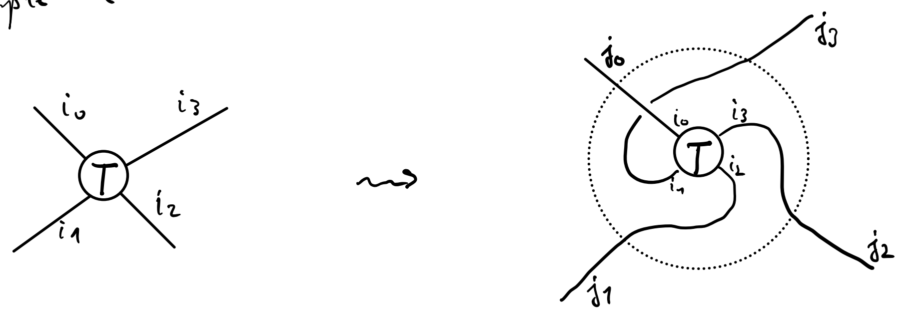
each connected edge corresponds to a summation over shared inde

Permutation of dimensions

Generalization of matrix transposition:



Example (counter-clockwise ordering of indices):



$$\text{np.transpose}\,(t,\ \overbrace{(0,\ 2,\ 3,\ 1)}^{\text{perm}})$$

$$\begin{array}{cccc} i_0 & i_2 & i_3 & i_1 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ j_0 & j_1 & j_2 & j_3 \end{array}$$

Note: Reverse permutation?

invperm = np.argsort (perm) (0, 3, 1, 2)

np.transpose (np.transpose (t, perm), np.argsort (perm)) = t

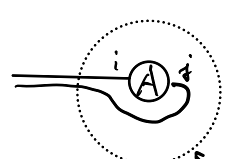
Vectorization

Graphical equivalent of vectorization:

combine tensor legs

For a matrix $A \in \mathbb{C}^{m \times n}$: $\text{vec}(A) \in \mathbb{C}^{m \cdot n}$

single "virtual" index (i,j) }

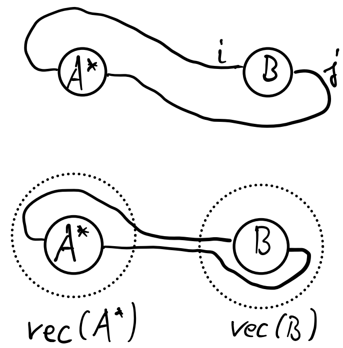


tensor with a single virtual leg

Remark: row- vs. column major not specified in diagrams

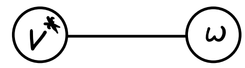
Example: $\mathrm{tr}[A^\dagger B]$ can be interpreted as dot product of two vectors:

$$\mathrm{tr}[A^\dagger B] = \sum_{i,j} a_{ij}^* b_{ij} \quad \hat{=}$$



Compare with inner product of $v, w \in \mathbb{C}^n$:

$$\langle v, w \rangle = \sum_{j=1}^{n} v_j^* w_j$$



In the following:

$$\langle A, B \rangle := \mathrm{tr}[A^\dagger B] = \langle \mathrm{vec}(A), \mathrm{vec}(B) \rangle$$

Vectorization of arbitrary tensors works analogously:

$$T \in \mathbb{C}^{n_1 \times \dots \times n_d}$$

NumPy: np.reshape(t, -1) ← returned vector points to same data

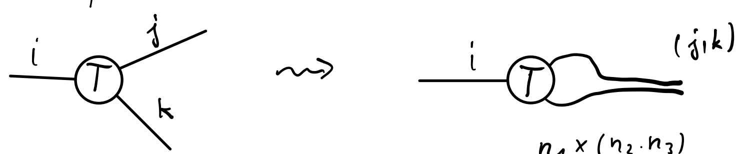
t.flatten() ← copies data

Matricization

Often it is useful to interpret a tensor as matrix

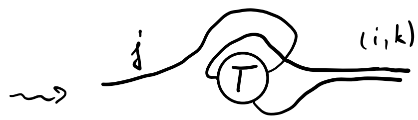
⇝ combine tensor legs into two "virtual" index tuples

Example: $T \in \mathbb{C}^{n_1 \times n_2 \times n_3}$



np.reshape(t, $(n_1, n_2 * n_3)$)

$\in \mathbb{C}^{n_1 \times (n_2 \cdot n_3)}$



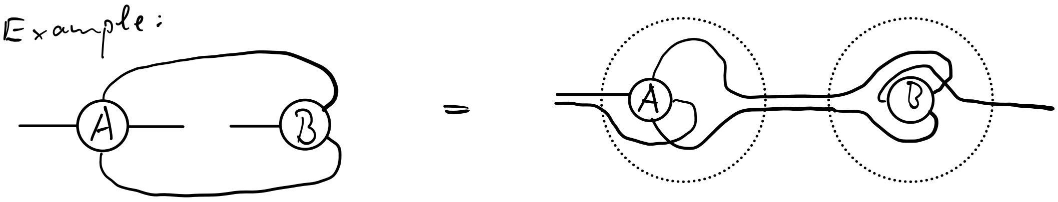
$\in \mathbb{C}^{n_2 \times (n_1 \cdot n_3)}$

np.reshape(np.transpose(t, (1,0,2)), $(n_2, n_1 * n_3)$)

$i \leftrightarrow j$

Contraction of two arbitrary tensors can thus be interpreted as matrix - matrix multiplication:

Example:



tensors after permuting dimensions and matricization

Outer and Kronecker product

Recall for $v \in \mathbb{C}^m$, $w \in \mathbb{C}^n$:

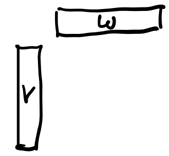
Kronecker product $$v \otimes w = \begin{pmatrix} v_1 w_1 \\ v_1 w_2 \\ \vdots \\ v_1 w_n \\ \vdots \\ v_m w_n \end{pmatrix} \in \mathbb{C}^{m \cdot n}$$

np.kron(v, w)

Can be interpreted as vectorization of

<u>outer product</u> $v \circ w$:

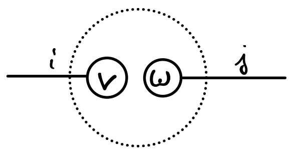
$$v \circ w = (v_i w_j)_{i=1..m,\, j=1,..n} = \begin{pmatrix} v_1 w_1 & v_1 w_2 & \cdots & v_1 w_n \\ v_2 w_1 & & & \vdots \\ \vdots & & & \\ v_m w_1 & \cdots & & v_m w_n \end{pmatrix}$$



$\in \mathbb{C}^{m \times n}$

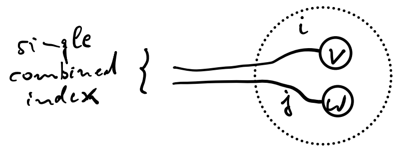
np.outer(v, w)

$v \circ w \;\hat{=}$



23.5.2022

Thus $v \otimes w \;\hat{=}$



Leads to alternative representation of SVD :

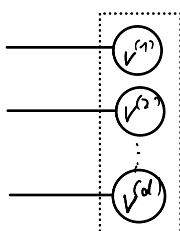
$$A = \sum_j \sigma_j \, u_j \circ v_j^*$$

$A = U \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} V^\dagger$

$U = (u_1 \mid u_2 \mid \cdots \mid u_n)$

In general: <u>outer product</u> of vectors $v^{(1)}, \ldots v^{(d)}$ defined as

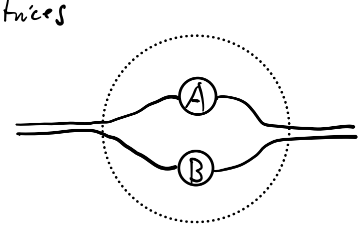
$v^{(1)} \circ v^{(2)} \circ \cdots \circ v^{(d)} \;\hat{=}$



(tensor of degree d)

Kronecker product of matrices

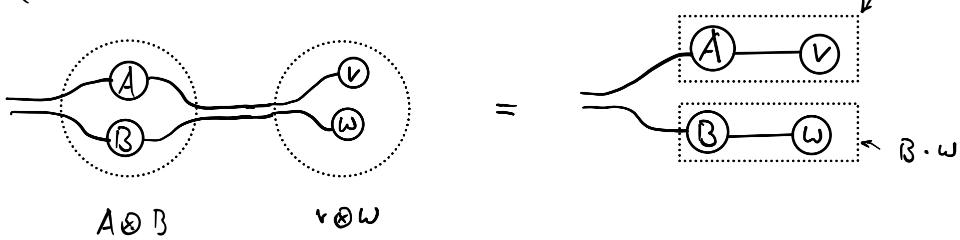
$A \otimes B \;\hat{=}$



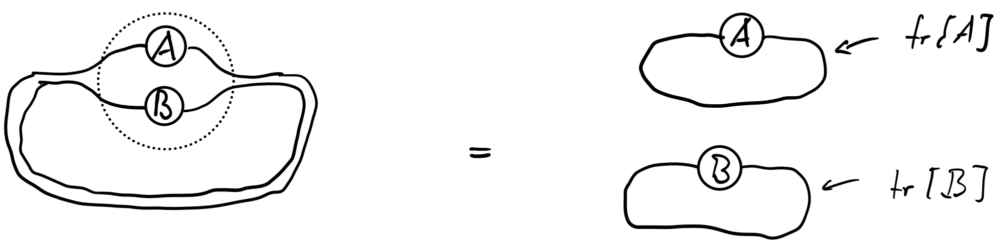
np.kron(A, B)

⇝ can graphically "prove" several properties of the Kronecker product:

- $(A \otimes B) \cdot (v \otimes w) = (Av) \otimes (Bw)$:

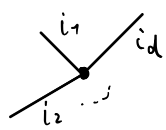


- Similarly $(A \otimes B) \cdot (C \otimes D) = (A \cdot C) \otimes (B \cdot D)$
- $\mathrm{tr}[A \otimes B] = \mathrm{tr}[A] \cdot \mathrm{tr}[B]$



Delta tensor (also denoted COPY tensor)

Can be defined for arbitrary degree d



$$\delta_{i_1 \dots i_d} = \begin{cases} 1 & \text{if } i_1 = i_2 = \dots i_d \\ 0 & \text{otherwise} \end{cases}$$

For $d = 2$: identity matrix

3. Canonical tensor formats and low rank approximation

3.1 Low-rank approximation

Norms

- Spectral norm: for $A \in \mathbb{C}^{m \times n}$:

$$\|A\|_2 := \max \left\{ \|A \cdot v\| : v \in \mathbb{C}^n \text{ with } \|v\| = 1 \right\}$$

standard Euclidean norm

Can derive that:

$$\|A\|_2 = \sigma_1(A) \quad \text{(largest singular value)}$$

- Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \|\mathrm{vec}(A)\|$$

since $\sum_{i,j} |a_{ij}|^2 = \mathrm{tr}[A^+ A] = \sum_j \sigma_j^2(A)$

$$\|A\|_F = \sqrt{\sum_j \sigma_j^2(A)}$$

Corresponding inner product:

$$\langle A, B \rangle = \langle \mathrm{vec}(A), \mathrm{vec}(B) \rangle$$

Can be generalized to tensors of arbitrary degree:

for $T \in \mathbb{C}^{n_1 \times \cdots \times n_d}$

$$\|T\|_F := \|\mathrm{vec}(T)\| = \sqrt{\sum_{i_1 \dots i_d} |t_{i_1 \dots i_d}|^2}$$

$$\langle T_1, T_2 \rangle = \langle \mathrm{vec}(T_1), \mathrm{vec}(T_2) \rangle$$

Basic property: $\|\cdot\|_2$ and $\|\cdot\|_F$ for matrices are <u>unitarily invariant</u>:

for any $A \in \mathbb{C}^{m\times n}$ and unitary $U \in \mathbb{C}^{m\times m}$, $V \in \mathbb{C}^{n\times n}$:

$$\|UAV\|_2 = \|A\|_2, \quad \|UAV\|_F = \|A\|_F$$

(such unitary transformations leave singular values invariant)

Von Neumann's trace inequality

Theorem: Let $A, B \in \mathbb{C}^{m\times n}$, w.l.o.g. $m \geq n$, with singular values denoted

$$\sigma_1(A) \geq \sigma_2(A) \geq \dots \geq \sigma_n(A) \geq 0$$
$$\sigma_1(B) \geq \sigma_2(B) \geq \dots \geq \sigma_n(B) \geq 0,$$

respectively.

Then $$|\langle A, B\rangle| \leq \sum_{j=1}^{n} \sigma_j(A)\cdot \sigma_j(B)$$

Optimal low-rank approximation by truncated SVD

$A \in \mathbb{C}^{m\times n}$, w.l.o.g $m \geq n$, SVD: $A = USV^{+}$ with

isometry $U = (u_1 | \dots | u_n) \in \mathbb{C}^{m\times n}$, $S = \mathrm{diag}(\sigma_1 \dots \sigma_n)$

unitary $V = (v_1 | \dots | v_n) \in \mathbb{C}^{n\times n}$

For $k \leq n$, define

$$T_k(A) = \underbrace{(u_1 | \dots | u_k)}_{\tilde{U}} \cdot \underbrace{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix}}_{\tilde{S}} \cdot \underbrace{(v_1 | \dots | v_k)^{+}}_{\tilde{V}^{+}} =$$

$$= U \cdot \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \cdot V^{+}$$

(see Exercise 2.2 (b))

$\rightarrow$ for any unitary invariant matrix norm $\|\cdot\|$:

$$\|A - T_k(A)\| = \| USV^+ - U \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_k & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix} V^+ \| =$$

$$= \| U \cdot \left(S - \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_k & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix} \right) V^+ \|$$

$$= \| \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \sigma_{k+1} & \\ & & & & \ddots \; \sigma_n \end{pmatrix} \|$$

In particular: $\|A - T_k(A)\|_2 = \sigma_{k+1}$, $\|A - T_k(A)\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_n^2}$

Theorem (Eckart-Young-Mirsky): Let $A \in \mathbb{C}^{m \times n}$.
Then for any unitarily invariant matrix norm $\|\cdot\|$,
$$\|A - T_k(A)\| = \min \{ \|A - B\| : B \in \mathbb{C}^{m\times n} \text{ has rank at most } k. \}$$

Proof for $\|\cdot\|_F$:

For any $B \in \mathbb{C}^{m \times n}$:

$$\|A - B\|_F^2 = \langle A - B, A - B \rangle = \langle A, A \rangle - \langle A, B \rangle - \underbrace{\langle B, A \rangle}_{\langle A, B\rangle^*} + \langle B, B \rangle =$$

$$= \|A\|_F^2 - 2 \operatorname{Re} \langle A, B \rangle + \|B\|_F^2$$

$$\geq \|A\|_F^2 - 2 |\langle A, B \rangle| + \|B\|_F^2$$

$$\underset{\uparrow \text{trace inequality}}{\geq} \|A\|_F^2 - 2 \sum_j \sigma_j(A) \cdot \sigma_j(B) + \|B\|_F^2 = \sum_j (\sigma_j(A) - \sigma_j(B))^2$$

If $\operatorname{rank}(B) \leq k$, at most k singular values of B can be non-zero; in this case

$$\sum_j (\sigma_j(A) - \sigma_j(B))^2 \text{ is minimized by } \begin{array}{ll} \sigma_j(B) = \sigma_j(A), & j = 1, \dots, k \\ \sigma_j(B) = 0, & j = k+1, \dots \end{array}$$

Inserting $B = T_k(A)$ attains this lower of $\|A - B\|_F^2$ $\square$

3.2 Canonical Polyadic (CP) decomposition

$T \in \mathbb{C}^{n_1 \times \dots \times n_d}$, CP decomposition of T:

$$T = \sum_{j=1}^{r} u_j^{(1)} \circ \dots \circ u_j^{(d)} =: [\![U^{(1)}, \dots, U^{(d)}]\!]$$

with $U^{(\ell)} = \left(u_1^{(\ell)} \,|\, \dots \,|\, u_r^{(\ell)} \right) \in \mathbb{C}^{n_\ell \times r}$ (not neccessarily isometrics)

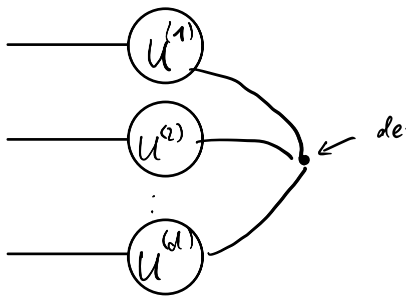
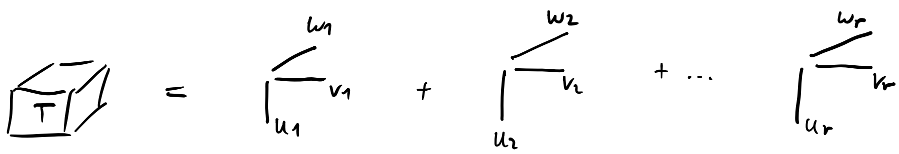


Illustration for $d = 3$, notation $T = \sum_{j=1}^{r} u_j \circ v_j \circ w_j$



30.5.2022

<u>Tensor rank of T:</u> minimal possible r

Remark:

For $d = 2$ (matrix case), the rank is a lower semi-continuous function

i.e. a converging sequence of matrizes of rank r always converges to a matrix of rank at most r

$\{A^j\}_{j=1,\dots} \in \mathbb{C}^{m \times n}$

$\lim_{j \to \infty} \| A^j - A \| = 0$

Does not hold for $d \geq 3$:

Counterexample by Silva and Lim:

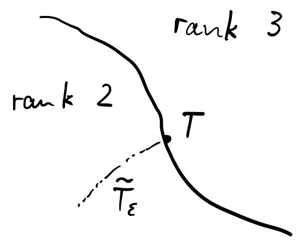
$T = e_1 \circ e_1 \circ e_2 + e_1 \circ e_2 \circ e_1 + e_2 \circ e_1 \circ e_1$, $\quad e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

has rank 3

$\tilde{T}_\varepsilon := \frac{1}{\varepsilon} (e_1 + \varepsilon e_2) \circ (e_1 + \varepsilon e_2) \circ (e_1 + \varepsilon e_2) - \frac{1}{\varepsilon} e_1 \circ e_1 \circ e_1$

has rank 2 (2 summands)

$\tilde{T}_\varepsilon \xrightarrow[\varepsilon \to 0]{} T$



"border rank problem"

<u>Alternating least squares (ALS)</u> algorithm commonly used to find CP decomposition

for $d = 3$:

Goal: given T, $\min_{U,V,W} \| T - [\![U, V, W]\!] \|^2$

Idea: optimize U, V, W one-by-one (leads to least-squares problem)

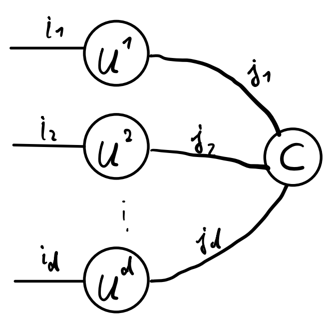
iterate until convergence

3.3 Tucker decomposition

$T = (t_{i_1 \dots i_d}) \in \mathbb{C}^{n_1 \times \dots \times n_d}$

$$t_{i_1 \dots i_d} = \sum_{j_1=1}^{k_1} \dots \sum_{j_d=1}^{k_d} c_{j_1 \dots j_d} \, u^1_{i_1, j_1} \circ \dots \circ u^d_{i_d, j_d}$$

$T =$



$U^\ell = (u^\ell_{ij}) \in \mathbb{C}^{n_\ell \times k_\ell}$

"core tensor"

$C \in \mathbb{C}^{k_1 \times \dots \times k_d}$

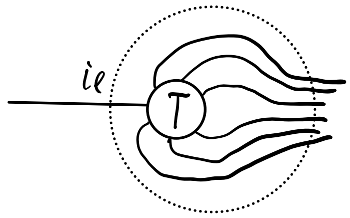
$k_\ell \ll n_\ell$

U^ℓ can be chosen to be isometries

Definition $\underline{\ell\text{-mode matricization}}$ $T^{(\ell)}$ of a tensor $T \in \mathbb{C}^{n_1 \times \dots \times n_d}$:

partitioning of dimensions into ℓ-th dimension and the remaining dimensions:

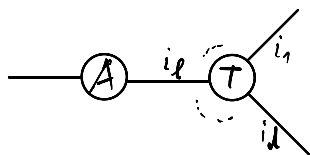
Thus $T^{(\ell)} \in \mathbb{C}^{n_\ell \times (n_1 \cdots n_{\ell-1} n_{\ell+1} \cdots n_d)}$

$T^{(\ell)} =$ 

$\underline{\text{Multilinear rank}}$ of T: $(r_1, \dots, r_d)$ with $r_\ell := \operatorname{rank}(T^{(\ell)})$

For Tucker format tensor: $r_\ell \le k_\ell \quad \forall \ell = 1, \dots d$

For $\ell \in \{1, \dots, d\}$ and a matrix $A \in \mathbb{C}^{m \times n_\ell}$,
the ℓ-mode matrix product $A \cdot_\ell T$ is the
multiplication of A with the ℓ-th dimension of T:

$A \cdot_\ell T =$  $\in \mathbb{C}^{n_1 \times \dots \times n_{\ell-1} \times m \times n_{\ell+1} \times \dots n_d}$

(ordering of dimensions same as for T)

$\underline{\text{Higher-order SVD algorithm (HOSVD)}}$

Goal: Find a Tucker format approximation of a given tensor T, using prescribed dimensions $k_1, \dots k_d$ (with $k_\ell \le r_\ell \ \forall \ell$)

1. Compute SVD of ℓ-mode matricizations:

$$T^{(\ell)} = U^\ell \cdot S^\ell \cdot (V^\ell)^\dagger, \qquad S^\ell = \operatorname{diag}(\sigma^\ell)$$

$(\sigma_1^\ell, \sigma_2^\ell, \dots)$

for $\ell = 1, \dots, d$

2. Truncate : $\tilde{U}^\ell := U^\ell_{:,1:k_\ell}$ for $\ell = 1, \dots, d$

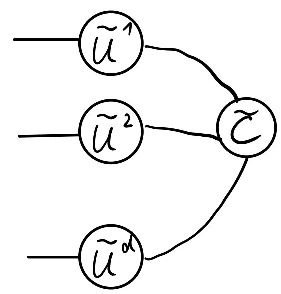
3. Form the core tensor $\tilde{C}$:

Initialize $\tilde{C} \leftarrow T$

For $\ell = 1, \dots, d$:

$$\tilde{C} \leftarrow (\tilde{U}^\ell)^+ \cdot_\ell \tilde{C}$$

return $(\tilde{U}^1, \dots, \tilde{U}^d, \tilde{C}, \sigma^1, \dots, \sigma^d)$

Then: $\quad T \approx$ 

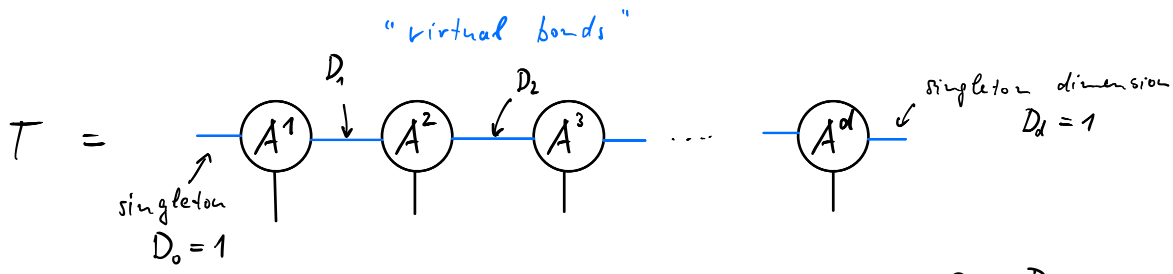
Note: $\tilde{U}^\ell \cdot (\tilde{U}^\ell)^+$ defines an projection onto the subspace spanned by the columns of $\tilde{U}^\ell$.

Theorem : The Tucker format tensor $\tilde{T}$ resulting from HOSVD with dimensions $(k_1, \dots, k_d)$ satisfies the quasi-optimal condition

$$\| T - \tilde{T} \| \leq \sqrt{d} \cdot \| T - T_{best} \| ,$$

with T_{best} the best approximation of T with multilinear rank $(k_1, \dots, k_d)$.

3.4 Tensor train (TT) / matrix product state (MPS) decomposition



TT cores / MPS "matrices" : $A^\ell \in \mathbb{C}^{n_\ell \times D_{\ell-1} \times D_\ell}$

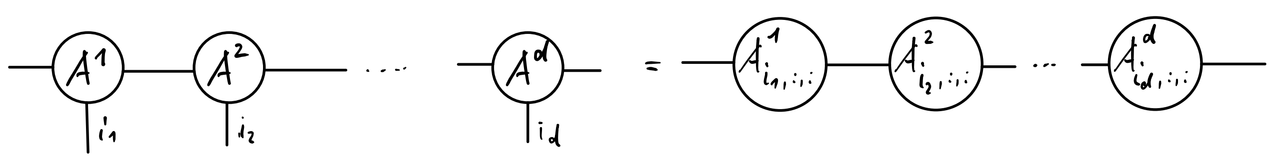
$\rightsquigarrow \quad T \in \mathbb{C}^{n_1 \times \dots \times n_d}$

Smallest possible tuple $(D_0, \dots D_d)$ is called <u>TT rank</u> of T

Notes : only polynomially many entries with increasing d
in A^ℓ tensors if virtual bonds are bounded : $D_\ell \leq D_{max}$

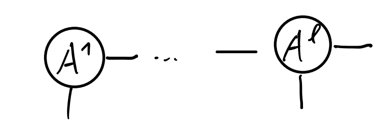
Compare with full tensor $T \in \mathbb{C}^{\overbrace{n \times \dots \times n}^{}_{d}}$: n^d entries

Each single entry $t_{i_1 \dots i_d}$ of a MPS tensor originates
from a sequence of matrix products , hence "matrix product state"



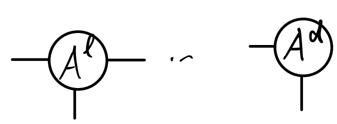
Properties :

Set $T_{\leq \ell}$ =

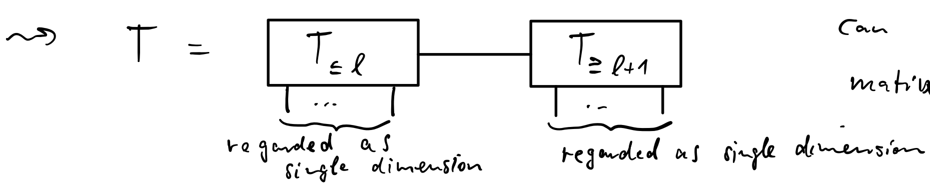


$\in \mathbb{C}^{n_1 \times \dots \times n_\ell \times D_\ell}$

$T_{\geq \ell}$ =



$\in \mathbb{C}^{D_{\ell-1} \times n_\ell \times \dots \times n_d}$



can be interpreted as matrix - matrix product

Interpreted as matices, $\operatorname{rank}(T_{\leq \ell}) \leq D_\ell$

$\operatorname{rank}(T_{\geq \ell+1}) \leq D_\ell$

and thus also $\operatorname{rank}(T_{\leq \ell} \cdot T_{\geq \ell+1}) \leq D_\ell$

$\rightsquigarrow$ matricization of MPS T as $\mathbb{C}^{(n_1 \cdots n_\ell) \times (n_{\ell+1} \cdots n_d)}$ matrix

has rank at most D_ℓ.