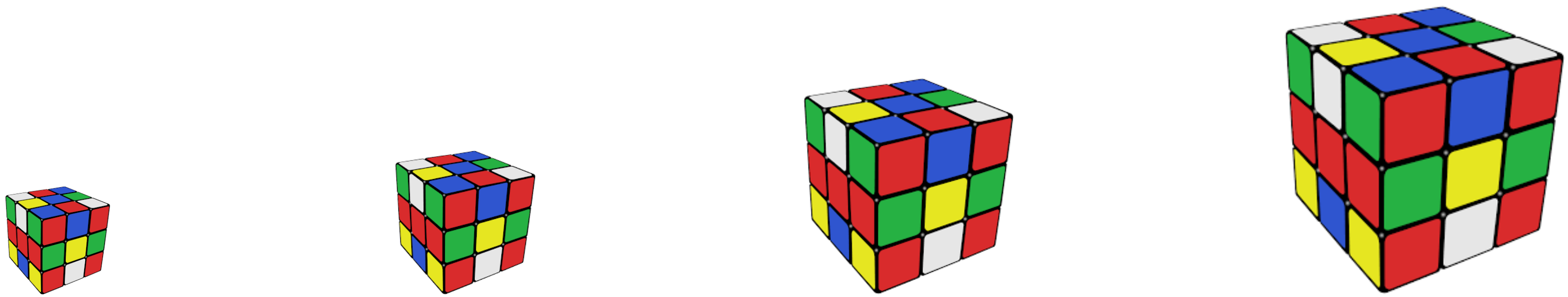
Tensors for multi-dimensional data analysis

Neriman Tokcan

March 22, 2019

LSA MATHEMATICS
UNIVERSITY OF MICHIGAN

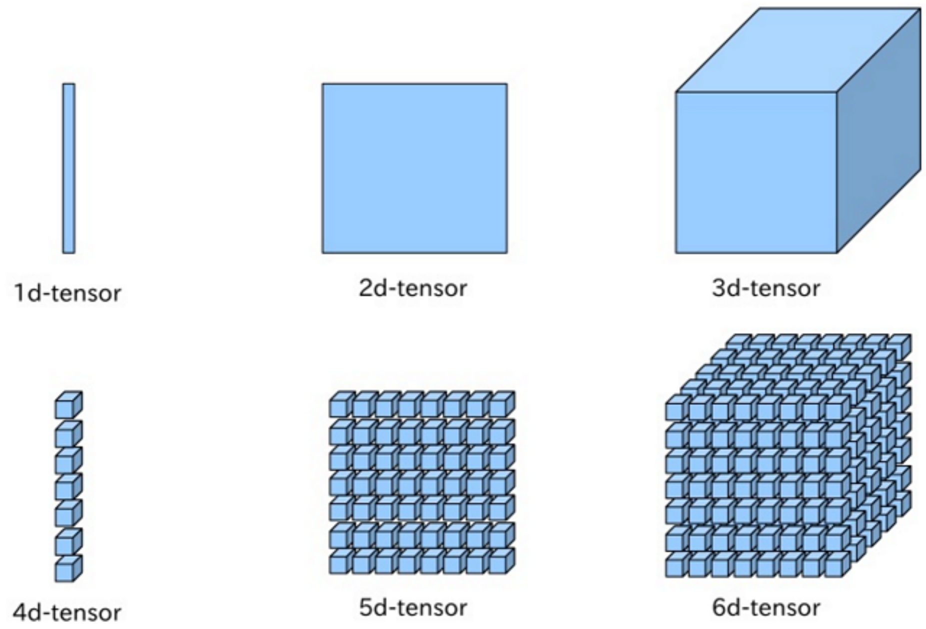
Department of
Computational Medicine and Bioinformatics
University of Michigan



Outline

- Tensors and why tensors?
- Quick review on CP-tensor decomposition
- Convex Decomposition
- Tensor Amplification
- Similarity Score for Tensors

Tensor is a multi-dimensional array



source:https://www.slideshare.net/yokotatsuya/principal-component-analysis-for-tensor-analysis-and-eeg-classification

Examples

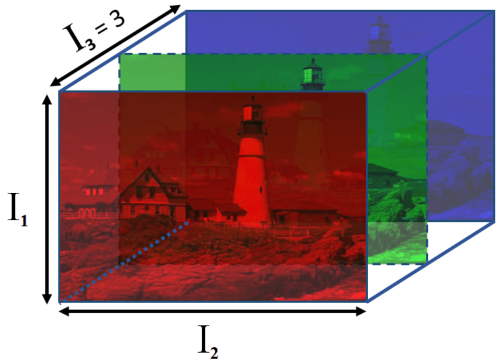


Figure: 3rd order tensor

color video is 4th-order tensor

Figure: 4th order tensor

sources:https://www.researchgate.net/publication/307091456_Quantifying_Blur_in_Color_Images_using_Higher_Order_Singular_Values, https://www.slideshare.net/BertonEarnshaw/a-brief-survey-of-tensors

Multilinear algebra

Let $V_1, V_2, \ldots, V_d$ be vector spaces. A function

$$f : V_1 \times V_2 \times \ldots V_d \to \mathbb{C}$$

is called multilinear if it is linear in each factor V_ℓ. The space of such multilinear functions is denoted $V_1^* \otimes V_2^* \otimes \ldots \otimes V_d^*$ and called the tensor product of the vector spaces $V_1^*, V_2^* \ldots, V_d^*$. Elements $T \in V_1^* \otimes V_2^* \otimes \ldots \otimes V_d^*$ are called tensors.
Let $V \cong \mathbb{R}^n$ be a Euclidean vector space with basis $e_1, e_2, \ldots, e_n$. On V we have an inner product that allows us to identify V with its dual space $V^\star$. We consider the d-fold tensor product of V:

$$V^{\otimes d} = \underbrace{V \otimes V \otimes \cdots \otimes V}_{d} \cong \mathbb{R}^{n \times n \times \cdots \times n} \cong \mathbb{R}^{n^d} \tag{1}$$

There are various ways to think of elements of $V^{\otimes d}$.The following statement is well known.

Lemma

There are bijections between the following sets:

1. *the set of* **tensors** $V^{\otimes d}$;
2. $(V^{\otimes d})^\star$, *the set of* **linear maps** $V^{\otimes d} \to \mathbb{R}$;
3. *the set of* $\mathbb{R}$-**multilinear maps** $V^d \to \mathbb{R}$

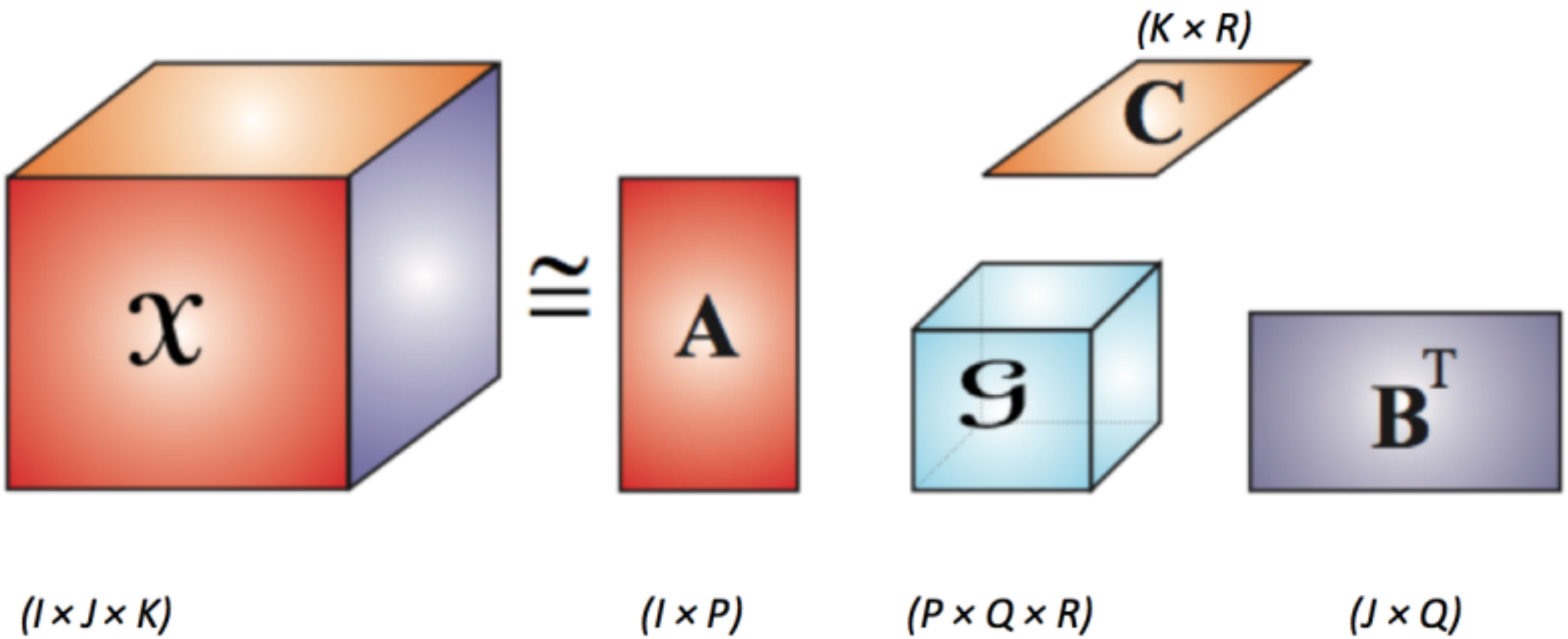
The bijection between (2) and (3) follows from the universal property of the tensor product. Since we have identified V with its dual $V^\star$ we can also identify $V^{\otimes d}$ with $(V^\star)^{\otimes d} \cong (V^{\otimes d})^\star$. We will frequently switch between these different viewpoints.

Why tensors and tensors decompositions?

- To analyze big data (As starting point express the tensor as sum of meaningful parts)
- For dimension reduction
- To exploit the structure of the data
- To reduce the computational complexity
- To deal with missing data (tensor completion)
- To deal with noisy data

Tucker decomposition

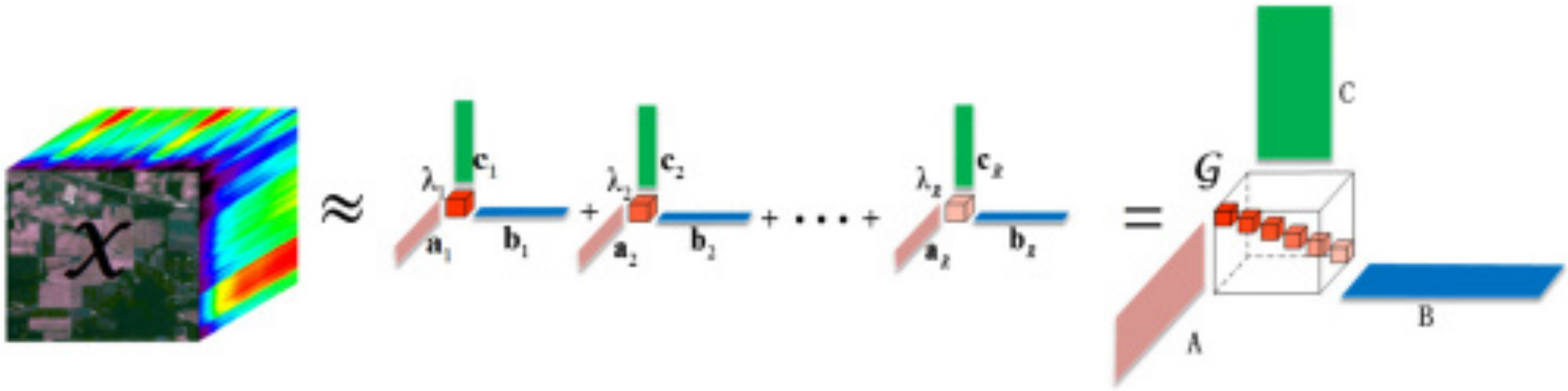
Tucker decomposition decomposes a tensor into a set of matrices and a small core tensor. Initially described as a three-mode extension of factor analysis and principal component analysis.



$$\mathcal{X} \approx \mathcal{G} \times_1 A \times_2 B \times_3 C$$

source: https://iksinc.online/2018/05/02/understanding-tensors-and-tensor-decompositions-part-3/

CP decomposition can be considered as a more restricted Tucker decomposition. In CP, the core tensor is restricted to be diagonal.



$$\mathcal{X} \approx \bar{\mathcal{X}} = \sum_{i=1}^{r} \lambda_i a_i \otimes b_i \otimes c_i = [\mathcal{G}, A, B, C]$$

Tensor rank: smallest number of rank-1 tensors that can generate X by summing up.
We want to find a rank r approximation of the tensor $\mathcal{X}$.

source: Leyuan Fang, Nanjun He, and Hui Lin, CP tensor-based compression of hyperspectral images

CP decomposition

Our goal is solving following problem:

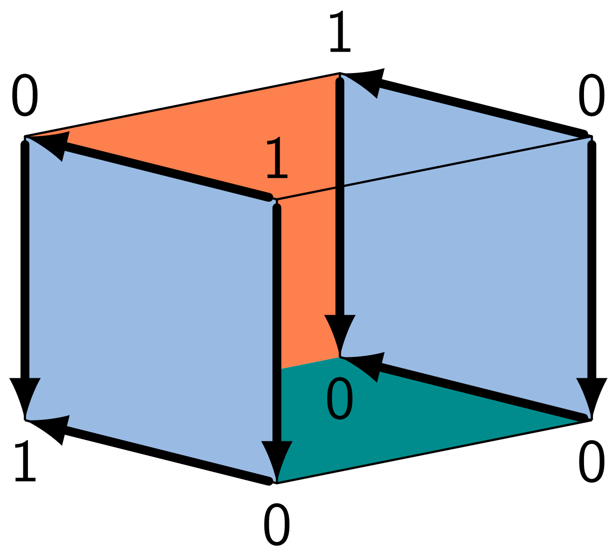
$$\text{minimize}_{\bar{\mathcal{X}}} ||\mathcal{X} - \bar{\mathcal{X}}||_F = \sum_{i,j,k} (\mathcal{X}(i,j,k) - \bar{\mathcal{X}}(i,j,k))^2$$

This is a non-convex problem with three sub-convex problems.
A common method for CP decomposition and other tensor-related optimization problems is Alternating Least Squares:

- Fix all but a subset of the unknowns
- By choosing different subsets, cycle through all the unknowns
- Repeat until converged

CP decomposition- Limitations

- CP decompositions are not always numerically stable
- Convergence is very slow
- Algorithm may not converge to a global minimum
- It is heavily dependent on the starting guess



$$T = e_2 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_1 + e_1 \otimes e_1 \otimes e_2.$$

T has rank 3, but it can be approximated by rank 2 tensors.
$T_n = n(e_1 + \frac{1}{n}e_2) \otimes (e_1 + \frac{1}{n}e_2) \otimes (e_1 + \frac{1}{n}e_2) - ne_1 \otimes e_1 \otimes e_1$, then $lim_{n \to \infty} T_n = T$.

Different decomposition and fit at each run

$$X \approx \bar{X} = \sum_{r=1}^{R} a_r \otimes b_r \otimes c_r, \quad \text{fit} = 1 - \frac{||X - \bar{X}||}{||\bar{X}||}$$

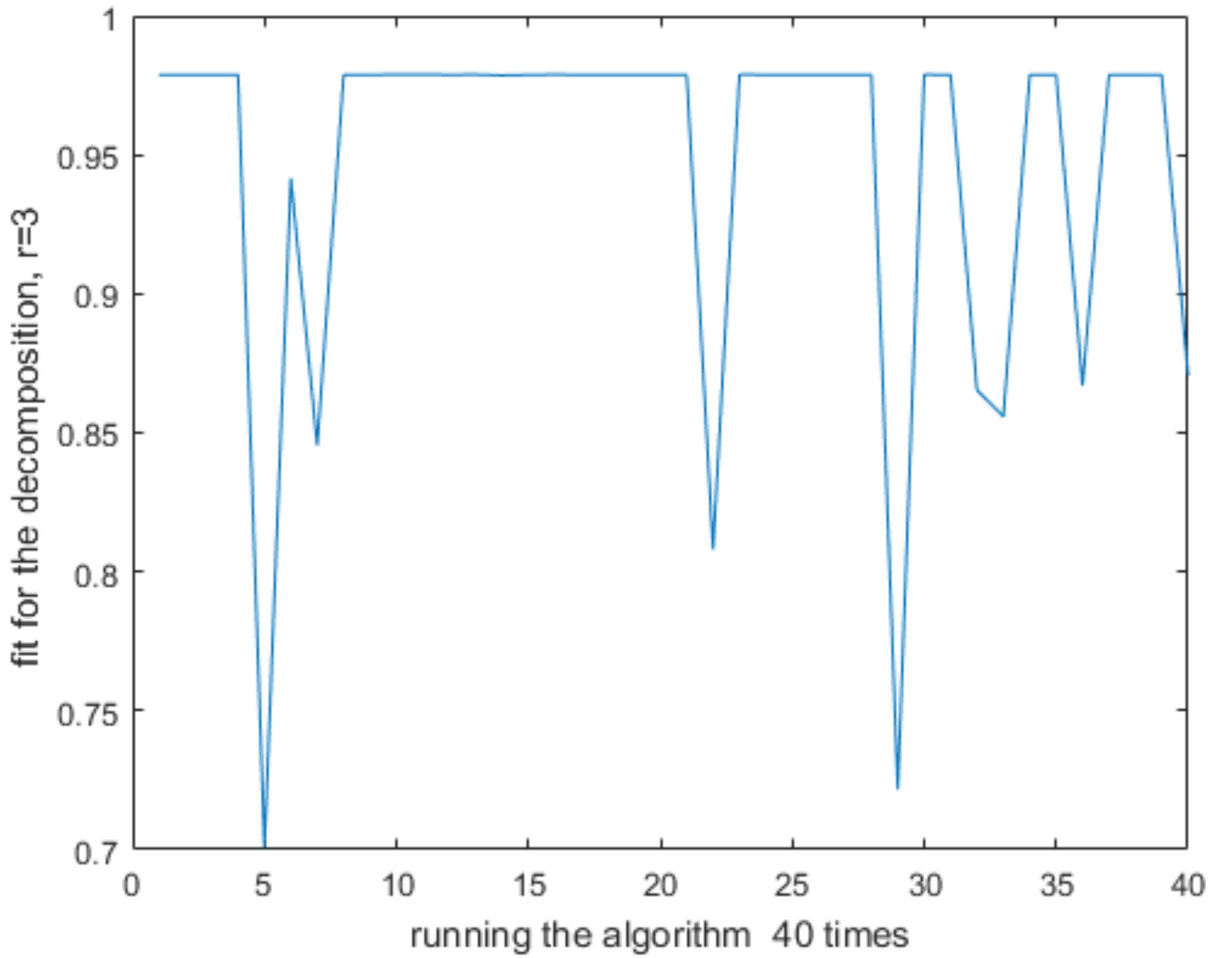
Example

```
X=tensor(rand(2,3,4))
% Generates a random tensor of size 2*3*4
Y=cp_als(X,3)
% Approximates a rank 3 decomposition of X

CP_ALS:
 Iter  1: f = 6.771001e-01 f-delta = 6.8e-01
 Iter 10: f = 9.312852e-01 f-delta = 1.5e-02
 Iter 20: f = 9.778946e-01 f-delta = 6.4e-04
 Iter 24: f = 9.787954e-01 f-delta = 9.4e-05
 Final f = 9.787954e-01
```

CP-local minimums

X is a 3rd order tensor of size $3 \times 4 \times 5$. We run the algorithm 40 times.



Nuclear decomposition

Definition

For a given tensor 3rd order tensor A, we want to find a decomposition

$$A = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i \tag{2}$$

such that

$$\sum_{i=1}^{r} ||u_i \otimes v_i \otimes w_i|| \tag{3}$$

is minimal.

The smallest value in (3) is called the nuclear norm and it is denoted by $||A||_*$. The decomposition in (2) is called nuclear decomposition.

Spectral norm

The dual of the nuclear norm is the spectral norm $||.||_\sigma$. The spectral norm $\|T\|_\sigma$ of a tensor $T \in V = R \otimes G \otimes B$ is defined by

$$\|T\|_\sigma := \max\{|T \cdot (x \otimes y \otimes z)| \mid \|x\| = \|y\| = \|z\| = 1\}.$$

We have

$$\|T\|_\sigma := \lim_{d \to \infty} \|T\|_{\overline{\sigma},d}$$

where

$$\|T\|_{\overline{\sigma},d} := \left(\int_{S^{p-1} \times S^{q-1} \times S^{r-1}} |T \cdot (x \otimes y \otimes z)|^d \, d\mu \right)^{1/d}.$$

Suppose that $d = 2e$ is even.

$$\int_{S^{p-1} \times S^{q-1} \times S^{r-1}} |T \cdot (x \otimes y \otimes z)|^d \, d\mu =$$

$$= T^{\otimes d} \cdot \int_{S^{p-1} \times S^{q-1} \times S^{r-1}} (x \otimes y \otimes z)^{\otimes d} \, d\mu.$$

Up to permuting the tensor factors, we have the following equality

$$\int_{S^{p-1}\times S^{q-1}\times S^{r-1}} (x\otimes y\otimes z)^{\otimes d}\, d\mu = \left(\int_{S^{p-1}} x^{\otimes d}\, d\mu\right)\otimes\left(\int_{S^{q-1}} y^{\otimes d}\, d\mu\right)\otimes\left(\int_{S^{r-1}} z^{\otimes d}\, d\mu\right).$$

How to calculate $E = \left(\int_{S^{p-1}} x^{\otimes d}\, d\mu\right)$?

This integral is over the (p-1) unit sphere with the normalized Haar measure. Because the Haar measure is $\mathrm{O}(V)$-invariant (??), the resultant tensor is also $\mathrm{O}(V)$-invariant.

The orthogonal group is defined as

$$\mathrm{O}(V) = \mathrm{O}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid AA^\dagger = I_n\}$$

The group $\mathrm{O}(V)$ acts on the n-fold tensor product space as follows. If $T = \sum_{j=1}^{r} v_{j,1} \otimes v_{j,2} \otimes \cdots \otimes v_{j,d} \in V^{\otimes d}$ is a tensor, then $A \in \mathrm{O}(V)$ acts by

$$A \cdot T = \sum_{j=1}^{r} Av_{j,1} \otimes Av_{j,2} \otimes \cdots \otimes Av_{j,d}. \tag{4}$$

The subspace of $\mathrm{O}(V)$-invariant tensors is

$$(V^{\otimes d})^{\mathrm{O}(V)} = \{T \in V^{\otimes d} \mid A \cdot T = T \text{ for all } A \in \mathrm{O}(V)\} \tag{5}$$

Theorem

The space $(V^{\otimes d})^{\mathrm{O}(V)} = 0$ *if* d *is odd. If* $d = 2e$*, the space* $(V^{\otimes d})^{\mathrm{O}(V)}$ *of* $\mathrm{O}(V)$*-invariant tensors is spanned by all multilinear maps of the form*

$$(v_1, v_2, \ldots, v_d) \mapsto \prod_{k=1}^{e} \langle v_{i_k}, v_{j_k} \rangle \tag{6}$$

where $\{i_1, i_2, \ldots, i_e, j_1, j_2, \ldots, j_e\} = \{1, 2, \ldots, d\}$.

The invariant tensor in (6) will be represented by its Brauer diagram.

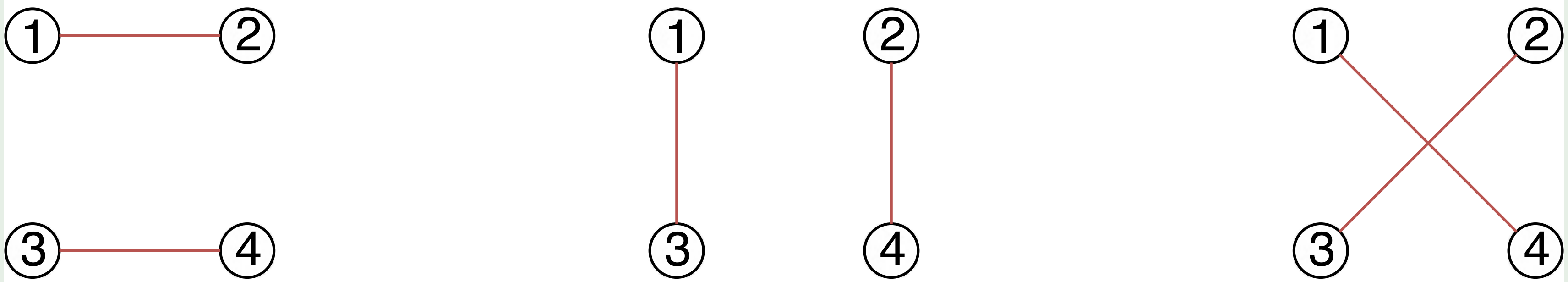
A labeled Brauer diagram of size $d = 2e$ is a perfect matching graph with vertices labeled $1, 2, \ldots, 2e$ edges between i_k and j_k for $k = 1, 2, \ldots, e$. There is a unique linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes d} \to \mathbb{R}$ such that

$$\begin{aligned}\mathcal{L}_{\mathbf{D}}(v_1 \otimes v_2 \otimes \cdots \otimes v_d) = \mathcal{M}_{\mathbf{D}}(v_1, v_2, \ldots, v_d) = \\ = (v_{i_1} \cdot v_{j_1})(v_{i_2} \cdot v_{j_2}) \cdots (v_{i_e} \cdot v_{j_e}). \quad (7)\end{aligned}$$

There is a unique tensor $\mathcal{T}_{\mathbf{D}} \in V^{\otimes d}$ such that $\mathcal{L}_{\mathbf{D}}(A) = \mathcal{T}_{\mathbf{D}} \cdot A$ for all tensors $A \in \mathrm{O}(V)$.
We also will represent this invariant tensor by $\partial^{i_1, j_1} \partial^{i_2, j_2} \ldots \partial^{i_e, j_e}$.

Example

if $e = 2$ (and $d = 4$), then the Brauer diagrams are



We will give the corresponding maps and tensor of the third diagram $\mathbf{D_3}$.

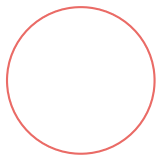
$$\mathcal{L}_{\mathbf{D_3}}(v_1 \otimes v_2 \otimes v_3 \otimes v_4) = \mathcal{M}_{\mathbf{D_3}}(v_1, v_2, v_3, v_4) = (v_1 \cdot v_4)(v_2 \cdot v_3)$$

If $e_1, \ldots, e_n$ is a basis of V, then we have

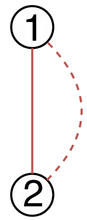
$$\mathcal{T}_{\mathbf{D_3}} = \sum_{i=1}^{n} \sum_{j=1}^{n} e_i \otimes e_j \otimes e_j \otimes e_i$$

Brauer diagrams-cycles

We can view $\partial^{1,2}$ as an element in $V \otimes V$, but also as a linear map $V \otimes V \to \mathbb{R}$. If we apply $\partial^{1,2} \in (V \otimes V)^\star$ to $\partial^{1,2} \in V \otimes V$, then we get $\sum_{i=1}^{n} \langle e_i, e_i \rangle = n$. In diagrams, we can write this as:



=



= n

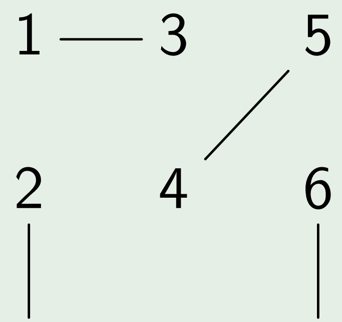
Partial Brauer Diagrams

A **partial Brauer diagram** of size d is a graph with d vertices labeled $1, 2, \ldots, d$ whose edges form a partial matching.
To a partial Brauer diagram $\mathbf{D}$ with e edges, we can associate an $\mathrm{O}(V)$-invariant multi-linear map $\mathcal{M}_{\mathbf{D}} : V^d \to V^{\otimes(d-2e)}$ and a linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes d} \to V^{\otimes d-2e}$.

Example

For the diagram:

$\mathbf{D} =$



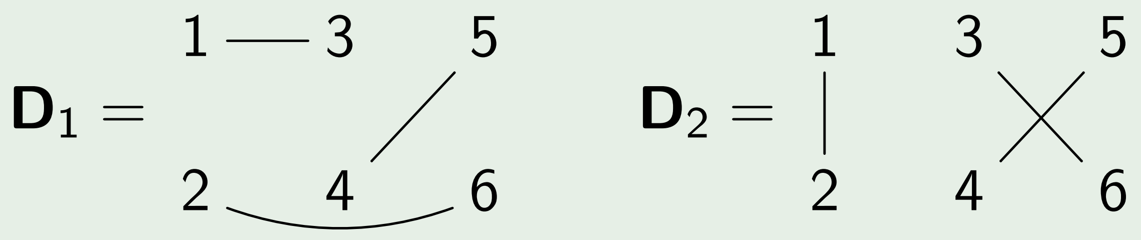
we have

$$\mathcal{M}_{\mathbf{D}}(v_1, v_2, \ldots, v_6) = (v_1 \cdot v_3)(v_4 \cdot v_5) v_2 \otimes v_6 \in V^{\otimes 2}$$

for $v_1, v_2, \ldots, v_6 \in V$.

Example

We compute the inner product $\mathcal{T}_{\mathbf{D}_1}$ and $\mathcal{T}_{\mathbf{D}_2}$ where $\mathbf{D}_1$ and $\mathbf{D}_2$ are the diagrams below:



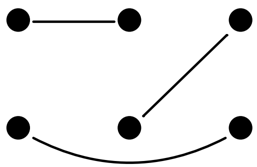
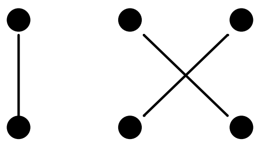
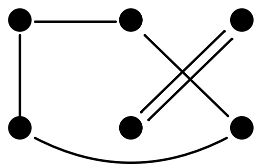
(8)

$$\mathcal{T}_{\mathbf{D}_1} \cdot \mathcal{T}_{\mathbf{D}_2} = \left(\sum_{i,j,k} e_i \otimes e_j \otimes e_i \otimes e_k \otimes e_k \otimes e_j \right) \cdot$$

$$\left(\sum_{p,q,r} e_p \otimes e_p \otimes e_q \otimes e_r \otimes e_r \otimes e_q \right) =$$

$$\sum_{i,j,k,p,q,r} (e_i \cdot e_p)(e_j \cdot e_p)(e_i \cdot e_q)(e_k \cdot e_r)(e_k \cdot e_r)(e_j \cdot e_q) = n^2. \quad (9)$$

We can visualize this computation as follows

 $\cdot$  $=$  $= n^2, \qquad (10)$

We overlay the diagrams. The indices of the edges in a cycle must all be the same. Since there are two cycles, namely i, p, j, q and k, r we essentially sum over two indices and get n^2.

General rule for inner product

The dot product of two tensors $\mathcal{T}_{\mathbf{D}_1}, \mathcal{T}_{\mathbf{D}_2} \in V^{\otimes d}$ can be computed as follows. We overlay the two diagrams $\mathbf{D}_1$ and $\mathbf{D}_2$ so that the (labeled) nodes coincide. Then $\mathcal{T}_{\mathbf{D}_1} \cdot \mathcal{T}_{\mathbf{D}_2} = n^k$ where k is the number of cycles (including 2-cyles).

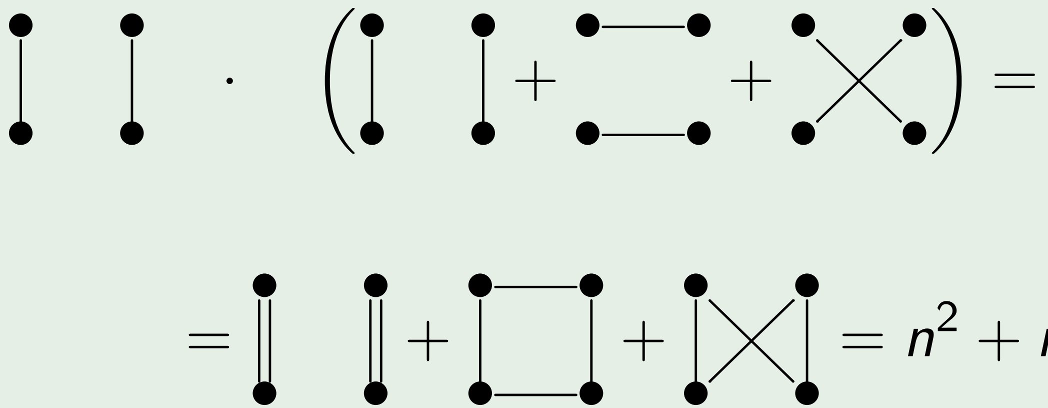
Proposition

Suppose that $T_d = \mathcal{T}_{\mathbf{D}}$ where $\mathbf{D}$ is Brauer diagram on $d = 2e$ vertices, and $\mathcal{S}_d = \sum_{\mathbf{E}} \mathcal{T}_{\mathbf{E}}$, where the sum is over all Brauer diagrams $\mathbf{E}$ on d vertices. Then we have $T_d \cdot \mathcal{S}_d = n(n+2)\cdots(n+d-2)$.

Example

Example

For $e = 2$, we get



$$= n^2 + n + n = n(n+2). \qquad (11)$$

Back to integrals

Proposition

$$\int_{S^{n-1}} v^{\otimes 2e}\, d\mu = \frac{1}{n(n+2)\cdots(n+2e-2)}\mathcal{S}_d$$

Let $U = \int_{S^{n-1}} v^{\otimes 2e}\, d\mu$. Since U is $\mathrm{O}(V)$ and Σ_{2d}-invariant, it is a linear combination (coefficients are same) of Brauer diagrams. So we have $U = C\mathcal{S}_{2e}$ where C is some constant. Let $\mathbf{D}$ be some Brauer diagram on d vertices. The value of C is obtained from

$$1 = \int_{S^{n-1}} d\mu = \int_{S^{n-1}} (\mathcal{T}_{\mathbf{D}} \cdot v^{\otimes 2e}) d\mu = \mathcal{T}_{\mathbf{D}} \cdot \int_{S^{n-1}} v^{\otimes 2e}\, d\mu =$$
$$C(\mathcal{T}_{\mathbf{D}} \cdot \mathcal{S}_d) = Cn(n+2)\cdots(n+2e-2).$$

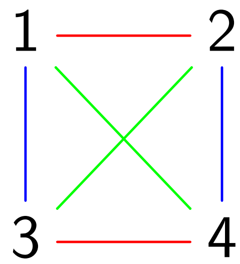
We now consider 3 euclidean $\mathbb{R}$-vector spaces R, G and B of dimension p, q and r respectively. The tensor product space $V = R \otimes G \otimes B$ is a representation of $H := \mathrm{O}(R) \times \mathrm{O}(G) \times \mathrm{O}(B)$. We are interested in H-invariant tensors in V. We have

$$(V^{\otimes d})^H \cong (R^{\otimes d})^{\mathrm{O}(R)} \otimes (G^{\otimes d})^{\mathrm{O}(G)} \otimes (B^{\otimes d})^{\mathrm{O}(B)}. \tag{12}$$

We see that $(V^{\otimes d})^H$ is spanned by diagrams with vertices $1, 2, 3, \ldots, d$ and red, green and blue edges such that for each of the colors we have a perfect matching. This means that each vertex has exactly one red, one green and one blue edge.

Definition

A **colored Brauer diagram** of size $d = 2e$ is a graph with d vertices labeled $1, 2, \ldots, d$ and e red, e green and e blue edges such that for each color, the edges of that color form a perfect matching.



(13)

corresponds to the linear map $\mathcal{L}_{\mathbf{D}} : V^{\otimes 4} = (R \otimes G \otimes B)^{\otimes 4} \to \mathbb{R}$ defined by

$$\begin{aligned}(a_1 \otimes b_1 \otimes c_1) \otimes (a_2 \otimes b_2 \otimes c_2) \otimes (a_3 \otimes b_3 \otimes c_3) \otimes (a_4 \otimes b_4 \otimes c_4) \mapsto \\ (a_1 \cdot a_2)(a_3 \cdot a_4)(b_1 \cdot b_4)(b_2 \cdot b_3)(c_1 \cdot c_3)(c_2 \cdot c_4). \quad (14)\end{aligned}$$

Approximation of spectral norm d=2

Suppose that $d = 2$. Then we have

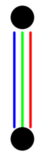
$$\int_{S^{p-1}} x \otimes x \, d\mu = \tfrac{1}{p}\mathcal{S}_{R,2}, \quad \int_{S^{q-1}} y \otimes y \, d\mu = \tfrac{1}{q}\mathcal{S}_{G,2},$$

$$\text{and } \int_{S^{r-1}} z \otimes z \, d\mu = \tfrac{1}{r}\mathcal{S}_{B,2}.$$

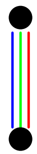
Therefore, we get

$$\int_{S^{p-1}\times S^{q-1}\times S^{r-1}} (x \otimes y \otimes z)^{\otimes 2} \, d\mu = \frac{1}{pqr}\mathcal{S}_{R,2} \otimes \mathcal{S}_{G,2} \otimes \mathcal{S}_{B,2}.$$

In diagrams, we get

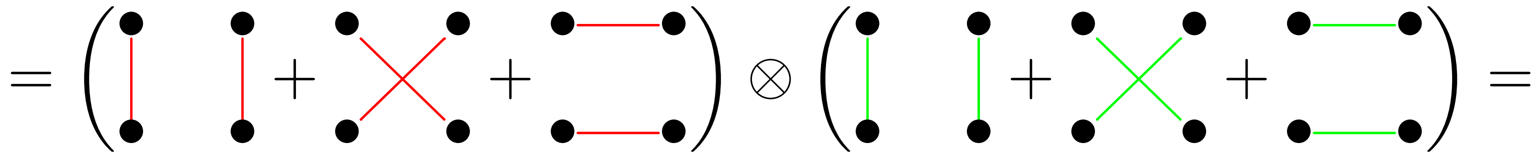
$$\int_{S^{p-1}\times S^{q-1}\times S^{r-1}} (x \otimes y \otimes z)^{\otimes 2} \, d\mu = \frac{1}{pqr}$$ 

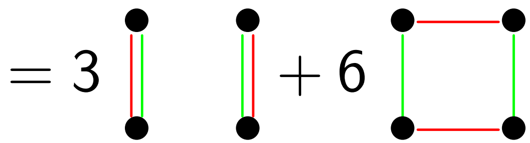
So we have

$$\|T\|_{\sigma,2}^2 = (T \otimes T) \cdot \left(\frac{1}{pqr} \right.$$  $$\left. \right) = \frac{1}{pqr} T \cdot T = \frac{1}{pqr}\|T\|^2.$$

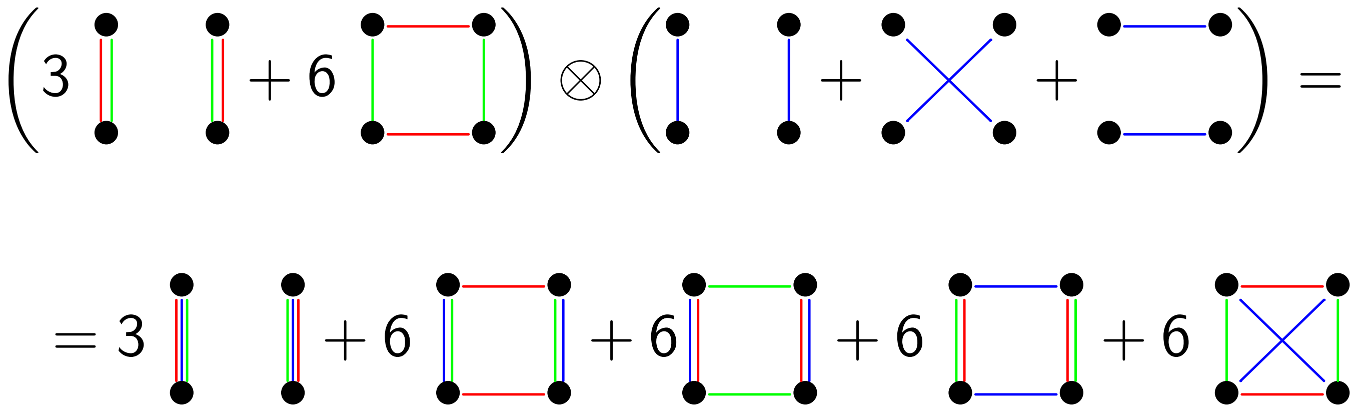
$$\int_{S^{p-1}\times S^{q-1}\times S^{r-1}} (x\otimes y\otimes z)^{\otimes 4}\, d\mu = \frac{1}{p(p+2)q(q+2)r(r+2)} \mathcal{S}_{R,4}\otimes\mathcal{S}_{G,4}\otimes\mathcal{S}_{B,4}.$$

$$\mathcal{S}_{R,4}\otimes\mathcal{S}_{G,4} =$$



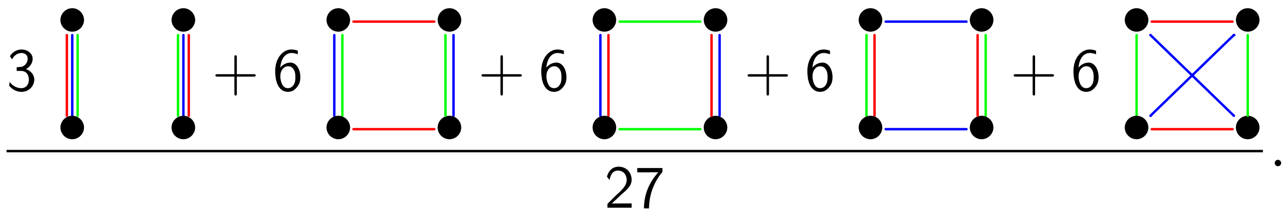


$$\mathcal{S}_{R,4} \otimes \mathcal{S}_{G,4} \otimes \mathcal{S}_{B,4} =$$



(15)

$$\|T\|_{\sigma,4}^4 =$$



This is a spectral-like norm and dual of this norm is a nuclear-like norm

Suppose that the singular values of A are $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r \geq$, then the spectral norm of A is λ_1. The singular values of $\Phi(A) = AA'A$ are $\lambda_1^3, \ldots, \lambda_r^3$. So Φ amplifies the largest singular value relative to the other singular values. We define $\bar{\Phi}(A) = \Phi(A)/||\Phi(A)||$. One can find the main component in the singular value decomposition by iterating Φ. For tensors, we define

$$\phi(T) = \int_{||v||=||w||=||z||=1} T(v, w, z)^3 v \otimes w \otimes z. \tag{16}$$

This operation amplifies the low rank structure of the tensor T. As $n \to \infty$, $\overline{\Phi}(T)^n$ typically converges to a rank 1 tensor rapidly. It is realted to spectral norm approximation, beacuse of the relation $||T||_{\sigma,4} = \langle \Phi(T), T \rangle$.

function DECOMPOSE(T, r, n)

$E = T$

$s = 0$

while $s < r$ **do**

$s \leftarrow s + 1$

$U \leftarrow \bar{\Phi}(E)$

Approximate U with $v_s = a_s \otimes b_s \otimes c_s$

Choose $\lambda_1, \ldots, \lambda_s$ with $\|E\|$ minimal, where

$E = T - (\lambda_1 v_1 + \cdots + \lambda_s v_s)$

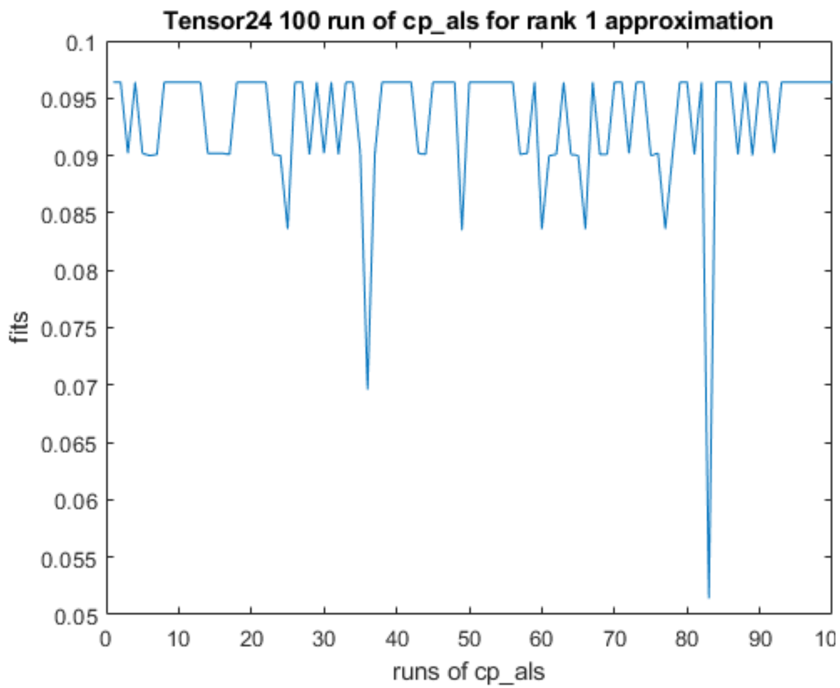
end while

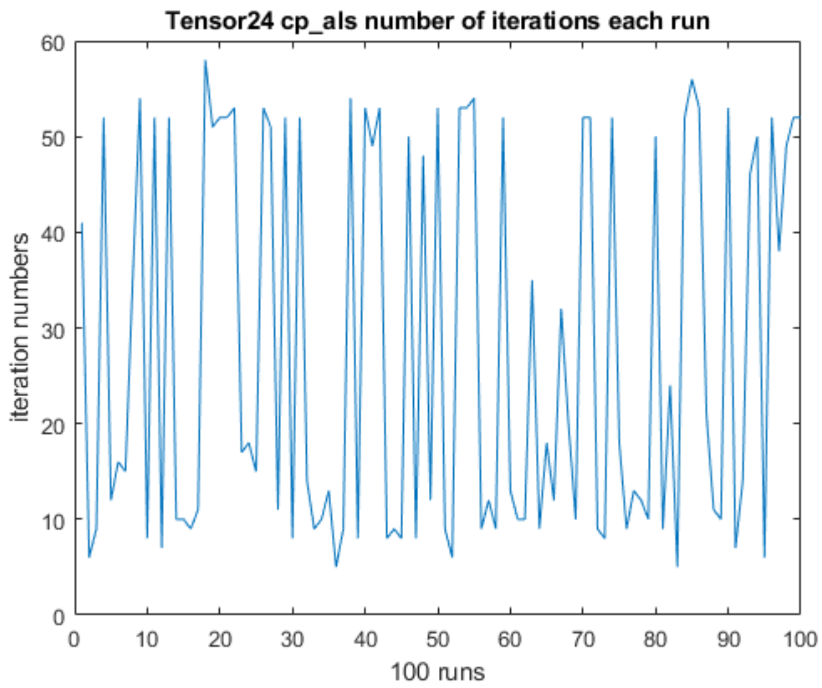
return decomposition $T = \lambda_1 v_1 + \cdots + \lambda_s v_s + E$

end function

Test- rank 1 approximation

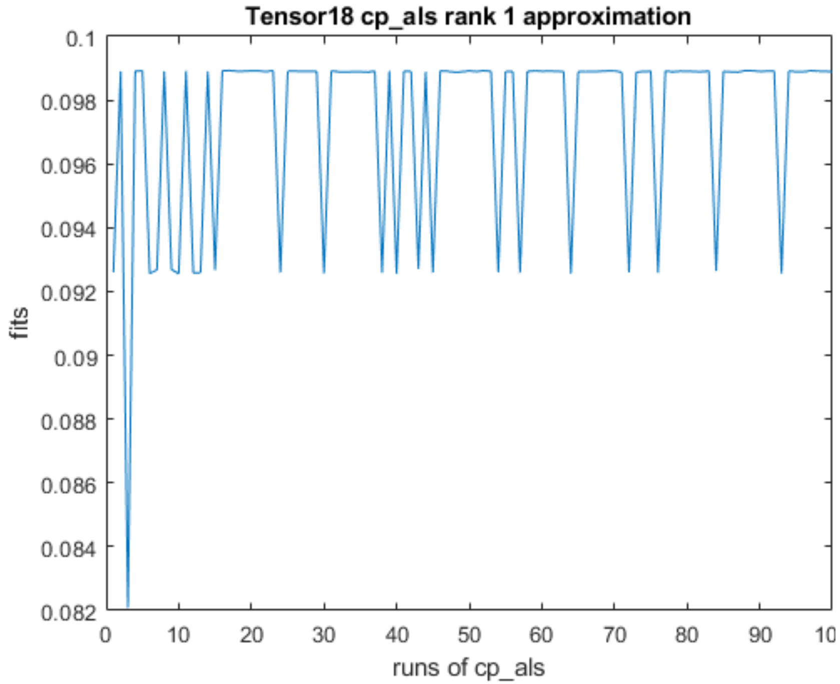
T is a $10 \times 10 \times 10$ tensor of rank 10, where components are random unit vectors independently drawn from the uniform distribution on the unit sphere.

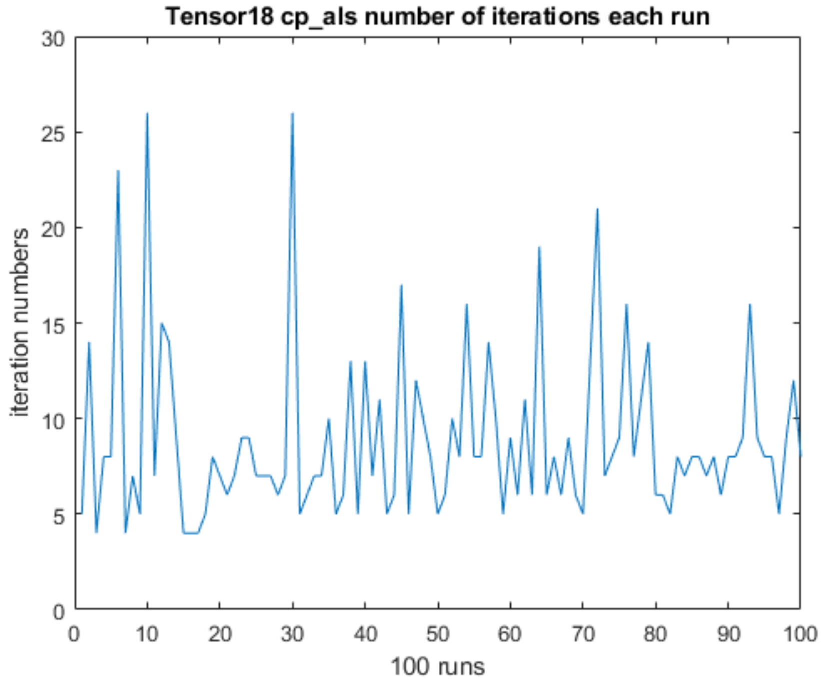




cp fits: min=0.0073, mean=0.0806, max=0.0964
cp iteration numbers: min=5, mean=28, max=58
amplification: fit=0.0964, iteration number=10

Test 2-more regular cp_als





cp fits: min=0.0821, mean=0.0974, max=0.0989
cp iteration numbers: min=4, mean=9, max=26
amplification: fit=0.0989, iteration number=2

Test on tensor decomposition/ denoising

We started with a random 50* 50 *50 tensor of rank 400, $T = \sum_{i=1}^{400} a_i \otimes b_i \otimes c_i$ where components are random unit vectors, independently drawn from the uniform distribution on the unit sphere. We then added a random tensor E with $||E|| = 20$. We created a noisy tensor $T_n = T + E$, and used three methods for denoising. Each method gives a denoised tensor Td and we measured the mean squared error $||T - T_d||^2$.

	Mean squared error	Signal/Noise ratio (dB)
No Denoising	400	0
CP-ALS	388	0.13
HOSVD	340	0.70
Amplification	193	3.17