

# Tensorized Random Projections

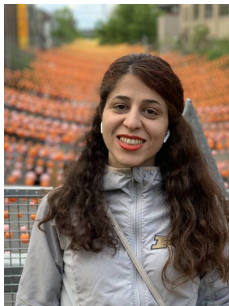
*Guillaume Rabusseau*

Assistant Professor at DIRO, UdeM  
CIFAR Canada Chair in AI at Mila

May 17, 2021

SIAM Conference on Applied Linear Algebra  
Mini-symposium *Tensor Methods: Theory and Practice*

Joint work with Beheshteh T. Rakhshan, published at AISTATS 2020.



# Motivation

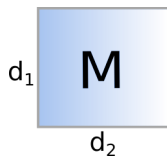
- **Random projection (RP)** and **tensor decomposition** : Two tools to deal with high-dimensional data
- But **RP** cannot scale to very high-dimensional inputs (e.g. high-order tensors)
- We use tensor decomposition to scale Gaussian RP to high-order tensors

# Outline

- 1 Introduction to Tensor Networks
- 2 Tensorized Random Projections
- 3 Experiments
- 4 Conclusion

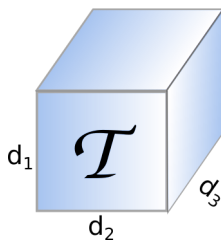
# Introduction to Tensor Networks

# Tensors



$$\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}$$

$$\mathbf{M}_{ij} \in \mathbb{R} \text{ for } i \in [d_1], j \in [d_2]$$




$$\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$$


$$(\mathcal{T}_{ijk}) \in \mathbb{R} \text{ for } i \in [d_1], j \in [d_2], k \in [d_3]$$

# Tensor Networks

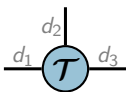
Degree of a node  $\equiv$  order of tensor



$\mathbf{v} \in \mathbb{R}^d$



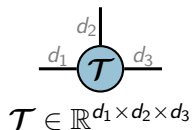
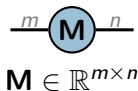
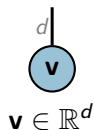
$\mathbf{M} \in \mathbb{R}^{m \times n}$



$\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$

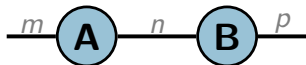
# Tensor Networks

Degree of a node  $\equiv$  order of tensor



Edge  $\equiv$  contraction

Matrix product:




$$(\mathbf{AB})_{i_1, i_2} = \sum_{k=1}^n \mathbf{A}_{i_1 k} \mathbf{B}_{k i_2}$$




# Tensor Networks

Degree of a node  $\equiv$  order of tensor



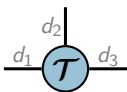
A blue circle node labeled  $\mathbf{v}$  with a vertical line extending upwards from its top, labeled  $d$ .

$$\mathbf{v} \in \mathbb{R}^d$$



A blue circle node labeled  $\mathbf{M}$  with a horizontal line extending to the left from its center, labeled  $m$ , and a horizontal line extending to the right from its center, labeled  $n$ .

$$\mathbf{M} \in \mathbb{R}^{m \times n}$$

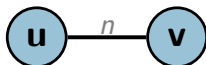


A blue circle node labeled  $\mathcal{T}$  with three lines extending from its center: a vertical line upwards labeled  $d_2$ , a horizontal line to the left labeled  $d_1$ , and a horizontal line to the right labeled  $d_3$ .

$$\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$$

Edge  $\equiv$  contraction


Inner product:




$$\mathbf{u}^\top \mathbf{v} = \sum_{k=1}^n \mathbf{u}_k \mathbf{v}_k$$

# Tensor Networks

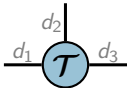
Degree of a node  $\equiv$  order of tensor



$\mathbf{v} \in \mathbb{R}^d$



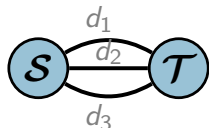
$\mathbf{M} \in \mathbb{R}^{m \times n}$



$\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$

Edge  $\equiv$  contraction

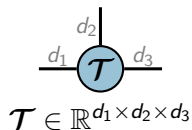
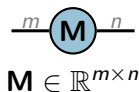
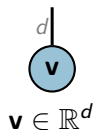
Inner product between tensors:



$$\langle \mathbf{S}, \mathbf{T} \rangle = \sum_{i_1=1}^{d_1} \sum_{i_2=1}^{d_2} \sum_{i_3=1}^{d_3} \mathbf{S}_{i_1 i_2 i_3} \mathbf{T}_{i_1 i_2 i_3}$$

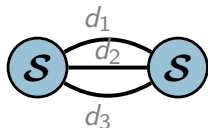
# Tensor Networks

Degree of a node  $\equiv$  order of tensor



Edge  $\equiv$  contraction

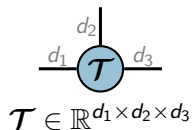
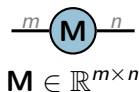
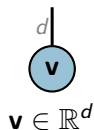
Frobenius norm of a tensor:



$$\|\mathbf{S}\|_F^2 = \sum_{i_1=1}^{d_1} \sum_{i_2=1}^{d_2} \sum_{i_3=1}^{d_3} (\mathbf{s}_{i_1 i_2 i_3})^2$$

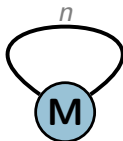
# Tensor Networks

Degree of a node  $\equiv$  order of tensor



Edge  $\equiv$  contraction

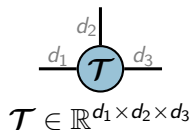
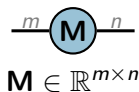
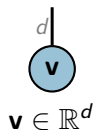
Trace of an  $n \times n$  matrix:



$$\text{Tr}(\mathbf{M}) = \sum_{i=1}^n \mathbf{M}_{ii}$$

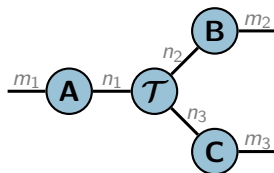
# Tensor Networks

Degree of a node  $\equiv$  order of tensor



Edge  $\equiv$  contraction

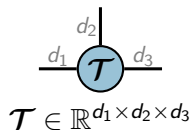
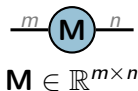
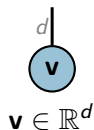
Tensor times matrices:



$$(\mathcal{T} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C})_{i_1, i_2, i_3} = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \sum_{k_3=1}^{n_3} \mathcal{T}_{k_1 k_2 k_3} \mathbf{A}_{i_1 k_1} \mathbf{B}_{i_2 k_2} \mathbf{C}_{i_3 k_3}$$

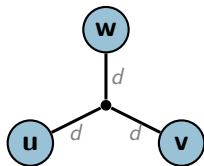
# Tensor Networks

Degree of a node  $\equiv$  order of tensor



Edge  $\equiv$  contraction

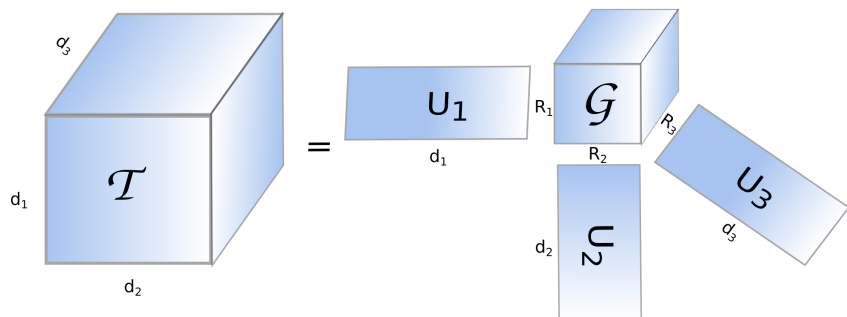
Hyperedge  $\equiv$  contraction between more than 2 indices:



$$\sum_{i=1}^n \mathbf{u}_i \mathbf{v}_i \mathbf{w}_i$$

# Tensor Decomposition Techniques

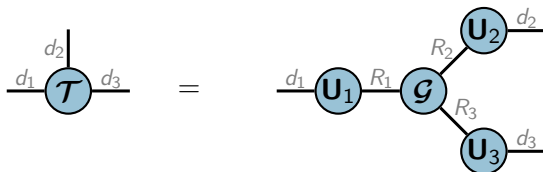
- Tucker decomposition [Tucker, 1966]:



$\Rightarrow R_1 R_2 R_3 + d_1 R_1 + d_2 R_2 + d_2 R_2$  parameters instead of  $d_1 d_2 d_3$ .

# Tensor Decomposition Techniques

- Tucker decomposition [Tucker, 1966]:

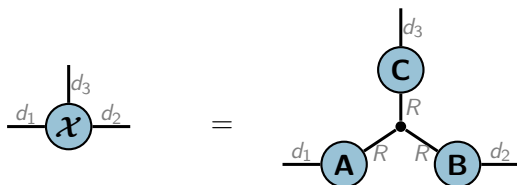
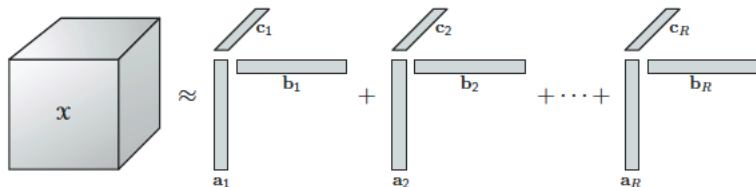


$\Rightarrow R_1 R_2 R_3 + d_1 R_1 + d_2 R_2 + d_3 R_3$  parameters instead of  $d_1 d_2 d_3$ .



# Tensor Decomposition Techniques

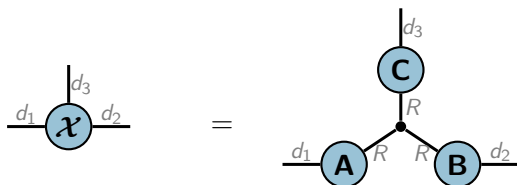
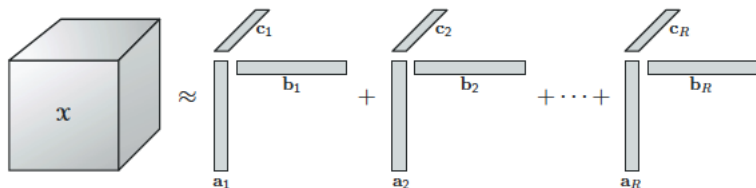
- CP decomposition [Hitchcock, 1927]<sup>1</sup>:



<sup>1</sup>fig. from [Kolda and Bader, *Tensor decompositions and applications*, 2009].

# Tensor Decomposition Techniques

- CP decomposition [Hitchcock, 1927]<sup>1</sup>:

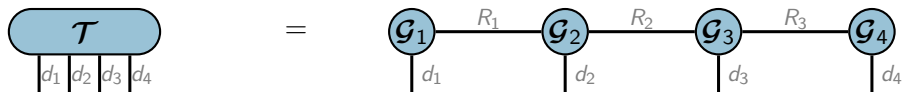


$\Rightarrow R(d_1 + d_2 + d_3)$  parameters instead of  $d_1 d_2 d_3$ .

<sup>1</sup>fig. from [Kolda and Bader, *Tensor decompositions and applications*, 2009].

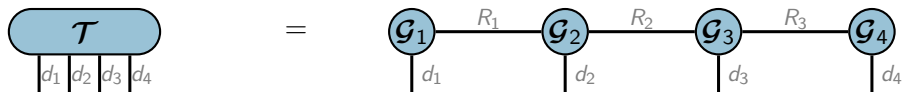
# Tensor Decomposition Techniques

- Tensor Train decomposition [Oseledets, 2011]:



# Tensor Decomposition Techniques

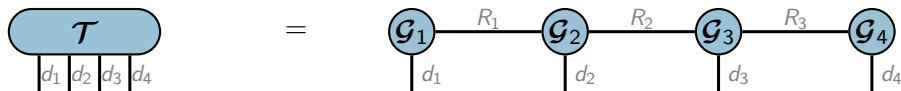
- Tensor Train decomposition [Oseledets, 2011]:



$\Rightarrow d_1 R_1 + R_1 d_2 R_2 + R_2 d_2 R_3 + R_3 d_4$  parameters instead of  $d_1 d_2 d_3 d_4$ .

# Tensor Decomposition Techniques

- Tensor Train decomposition [Oseledets, 2011]:

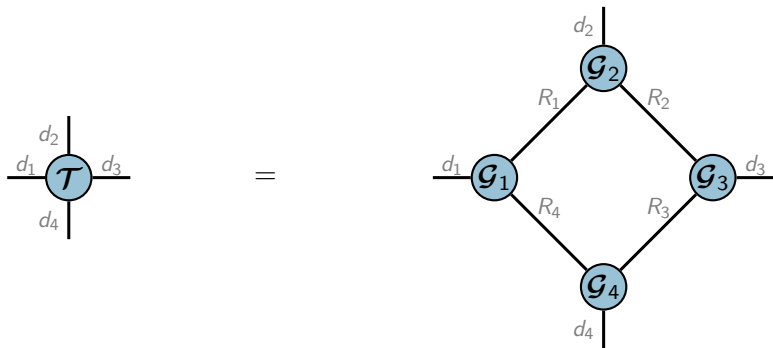


$\Rightarrow d_1 R_1 + R_1 d_2 R_2 + R_2 d_2 R_3 + R_3 d_3$  parameters instead of  $d_1 d_2 d_3 d_4$ .

- If the ranks are all the same ( $R_1 = R_2 = \dots = R$ ), can represent a vector of size  $2^n$  with  $\mathcal{O}(nR^2)$  parameters!

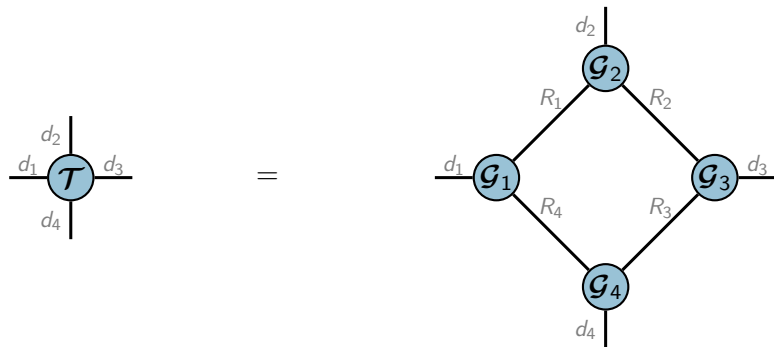
# Tensor Decomposition Techniques

- Tensor Ring decomposition [Zhao et al., 2016]:



# Tensor Decomposition Techniques

- Tensor Ring decomposition [Zhao et al., 2016]:



$\Rightarrow R_4 d_1 R_1 + R_1 d_2 R_2 + R_2 d_3 R_3 + R_3 d_4 R_4$  parameters instead of  $d_1 d_2 d_3 d_4$ .

# Tensorized Random Projections



# Random Projections (RP)

- Goal: find a low-dimensional projection  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$  ( $k \ll d$ ) that preserves distances (with high proba.).

## Random Projections (RP)

- Goal: find a low-dimensional projection  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$  ( $k \ll d$ ) that preserves distances (with high proba.).
- Johnson-Lindenstrauss Transform (or Gaussian RP).

$$f : \mathbf{x} \mapsto \frac{1}{\sqrt{k}} \mathbf{M} \mathbf{x} \quad \text{where } \mathbf{M}_{ij} \sim_{iid} \mathcal{N}(0, 1) \text{ for each } i, j$$

## Random Projections (RP)

- Goal: find a low-dimensional projection  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$  ( $k \ll d$ ) that preserves distances (with high proba.).
- Johnson-Lindenstrauss Transform (or Gaussian RP).

$$f : \mathbf{x} \mapsto \frac{1}{\sqrt{k}} \mathbf{M} \mathbf{x} \quad \text{where } \mathbf{M}_{ij} \sim_{iid} \mathcal{N}(0, 1) \text{ for each } i, j$$

### Theorem (JL, 1984)

Let  $\varepsilon > 0$  and  $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathbb{R}^d$ .

If  $k \gtrsim \varepsilon^{-2} \log m$ , then, with high proba.,  $\|f(\mathbf{x}_i)\| = (1 \pm \varepsilon) \|\mathbf{x}_i\|$  for all  $i = 1, \dots, m$ .

- Applications: sketched linear regression, randomized SVD, pre-processing step in ML pipeline, ...

# Random Projections (RP)

- Goal: find a low-dimensional projection  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$  ( $k \ll d$ ) that preserves distances (with high proba.).
- Johnson-Lindenstrauss Transform (or Gaussian RP).

$$f : \mathbf{x} \mapsto \frac{1}{\sqrt{k}} \mathbf{M} \mathbf{x} \quad \text{where } \mathbf{M}_{ij} \sim_{iid} \mathcal{N}(0, 1) \text{ for each } i, j$$

## Theorem (JL, 1984)

Let  $\varepsilon > 0$  and  $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathbb{R}^d$ .

If  $k \gtrsim \varepsilon^{-2} \log m$ , then, with high proba.,  $\|f(\mathbf{x}_i)\| = (1 \pm \varepsilon) \|\mathbf{x}_i\|$  for all  $i = 1, \dots, m$ .

- Applications: sketched linear regression, randomized SVD, pre-processing step in ML pipeline, ...
- **Problem:** if  $\mathbf{x} \in \mathbb{R}^{d^N}$  is a high-order tensor represented in CP/TT format, the Gaussian RP has  $d^N k$  parameters...

# Objective

- We want to find a RP map  $f : \mathbb{R}^{d^N} \rightarrow \mathbb{R}^k$  such that:
  - ▶ the number of parameters is linear in  $N$
  - ▶ computing  $f(\mathbf{x})$  is efficient when  $\mathbf{x}$  is in the CP or TT format
  - ▶  $f$  preserves distances with high probability.

# Objective

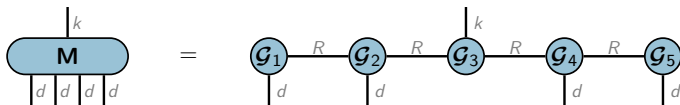
- We want to find a RP map  $f : \mathbb{R}^{d^N} \rightarrow \mathbb{R}^k$  such that:
  - ▶ the number of parameters is linear in  $N$
  - ▶ computing  $f(\mathbf{x})$  is efficient when  $\mathbf{x}$  is in the CP or TT format
  - ▶  $f$  preserves distances with high probability.
- Two important properties that a RP must satisfy:
  - ▶  $\mathbb{E}[\|f(\mathbf{x})\|^2] = \|\mathbf{x}\|^2$  for all  $\mathbf{x}$
  - ▶  $\lim_{k \rightarrow \infty} \mathbb{V}[\|f(\mathbf{x})\|^2] = 0$

# Objective

- We want to find a RP map  $f : \mathbb{R}^{d^N} \rightarrow \mathbb{R}^k$  such that:
    - ▶ the number of parameters is linear in  $N$
    - ▶ computing  $f(\mathbf{x})$  is efficient when  $\mathbf{x}$  is in the CP or TT format
    - ▶  $f$  preserves distances with high probability.
  - Two important properties that a RP must satisfy:
    - ▶  $\mathbb{E}[\|f(\mathbf{x})\|^2] = \|\mathbf{x}\|^2$  for all  $\mathbf{x}$
    - ▶  $\lim_{k \rightarrow \infty} \mathbb{V}[\|f(\mathbf{x})\|^2] = 0$
- ↪ the rate at which  $\mathbb{V}[\|f(\mathbf{x})\|^2]$  converges to 0 captures the **quality** of a RP.

# Tensor Train RP: First Attempt

- We build a Gaussian RP  $f : \mathbf{x} \mapsto \frac{1}{Z} \mathbf{M} \mathbf{x}$  where  $\mathbf{M} \in \mathbb{R}^{k \times d^N}$  is represented using the TT format:

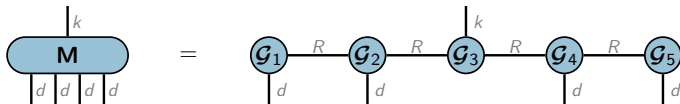


where the entries of each core tensor  $\mathcal{G}_n$  are drawn iid from  $\mathcal{N}(0, 1)$ .



# Tensor Train RP: First Attempt

- We build a Gaussian RP  $f : \mathbf{x} \mapsto \frac{1}{Z} \mathbf{M} \mathbf{x}$  where  $\mathbf{M} \in \mathbb{R}^{k \times d^N}$  is represented using the TT format:



where the entries of each core tensor  $\mathcal{G}_n$  are drawn iid from  $\mathcal{N}(0, 1)$ .

- ☺  $\mathcal{O}(R^2 N d + R^2 k)$  parameters instead of  $d^N k$ .
- ☺ Efficient computation of  $\mathbf{M} \mathbf{x}$  when  $\mathbf{x}$  is in the CP/TT format.
- ☺ We have  $\mathbb{E}[\|f(\mathbf{x})\|^2] = \|\mathbf{x}\|^2$ .
- ☹ We can show that  $\lim_{k \rightarrow \infty} \mathbb{V}[\|f(\mathbf{x})\|^2] > 0 \dots$

# Tensor Train Random Projection (TT-RP)

- Tensor Train RP:

$$f_{TT(R)} : \mathbf{x} \mapsto \frac{1}{\sqrt{kR^N}} \mathbf{M}\mathbf{x}$$

where each row of  $\mathbf{M} = \begin{pmatrix} -\mathbf{m}_1^\top - \\ -\mathbf{m}_2^\top - \\ \vdots \\ -\mathbf{m}_k^\top - \end{pmatrix} \in \mathbb{R}^{k \times d^N}$  is in the TT format:

$$\mathbf{m}_i = \mathcal{G}_1^i \overset{R}{-} \mathcal{G}_2^i \overset{R}{-} \mathcal{G}_3^i \overset{R}{-} \mathcal{G}_4^i \quad \text{for each } i = 1, \dots, k$$

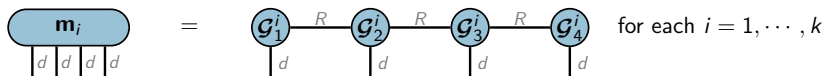
and the entries of each core tensor  $\mathcal{G}_n^i$  are drawn iid from  $\mathcal{N}(0, 1)$ .

# Tensor Train Random Projection (TT-RP)

- Tensor Train RP:

$$f_{TT(R)} : \mathbf{x} \mapsto \frac{1}{\sqrt{kR^N}} \mathbf{M}\mathbf{x}$$

where each row of  $\mathbf{M} = \begin{pmatrix} -\mathbf{m}_1^T - \\ -\mathbf{m}_2^T - \\ \vdots \\ -\mathbf{m}_k^T - \end{pmatrix} \in \mathbb{R}^{k \times d^N}$  is in the TT format:



and the entries of each core tensor  $\mathcal{G}_n^i$  are drawn iid from  $\mathcal{N}(0, 1)$ .

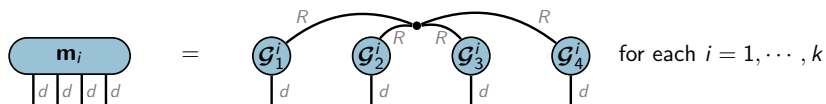
- ▶  $\mathcal{O}(kNdR^2)$  parameters instead of  $d^N k$ .
- ▶ Efficient computation of  $\mathbf{M}\mathbf{x}$  when  $\mathbf{x}$  is in the CP/TT format.
- ▶ We have  $\mathbb{E}[\|f_{TT(R)}(\mathbf{x})\|^2] = \|\mathbf{x}\|^2$ .
- ▶ We have  $\lim_{k \rightarrow \infty} \mathbb{V}[\|f_{TT(R)}(\mathbf{x})\|^2] = 0$ .

# CP Random projection (CP-RP)

- CP Random Projection:

$$f_{\text{CP}(R)} : \mathbf{x} \mapsto \frac{1}{\sqrt{kR^N}} \mathbf{M}\mathbf{x}$$

where each row of  $\mathbf{M} = \begin{pmatrix} -\mathbf{m}_1^T \\ -\mathbf{m}_2^T \\ \vdots \\ -\mathbf{m}_k^T \end{pmatrix} \in \mathbb{R}^{k \times d^N}$  is in the CP format:



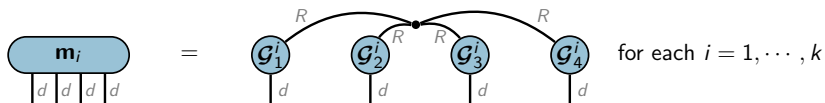
and the entries of each core tensor  $\mathcal{G}_n^i$  are drawn iid from  $\mathcal{N}(0, 1)$ .

# CP Random projection (CP-RP)

- CP Random Projection:

$$f_{CP(R)} : \mathbf{x} \mapsto \frac{1}{\sqrt{kR^N}} \mathbf{M}\mathbf{x}$$

where each row of  $\mathbf{M} = \begin{pmatrix} -\mathbf{m}_1^\top \\ -\mathbf{m}_2^\top \\ \vdots \\ -\mathbf{m}_k^\top \end{pmatrix} \in \mathbb{R}^{k \times d^N}$  is in the CP format:



and the entries of each core tensor  $\mathcal{G}_n^i$  are drawn iid from  $\mathcal{N}(0, 1)$ .

- ▶  $\mathcal{O}(kNdR)$  parameters instead of  $d^N k$ .
- ▶ Efficient computation of  $\mathbf{M}\mathbf{x}$  when  $\mathbf{x}$  is in the CP/TT format.
- ▶ We have  $\mathbb{E}[\|f_{CP(R)}\|^2] = \|\mathbf{x}\|^2$ .
- ▶ We have  $\lim_{k \rightarrow \infty} \mathbb{V}[\|f_{CP(R)}(\mathbf{x})\|^2] = 0$ .

# Main Result

## Theorem

Let  $\mathbf{x} \in \mathbb{R}^{d^N}$  and  $R \in \mathbb{N}$ .

The RP maps  $f_{TT(R)}$  and  $f_{CP(R)}$  satisfy the following properties:

- $\mathbb{E}[\|f_{CP(R)}(\mathbf{x})\|^2] = \mathbb{E}[\|f_{TT(R)}(\mathbf{x})\|^2] = \|\mathbf{x}\|^2$
- $\mathbb{V}[\|f_{TT(R)}(\mathbf{x})\|^2] \leq \frac{1}{k} \left(3 \left(1 + \frac{2}{R}\right)^{N-1} - 1\right) \|\mathbf{x}\|^4$
- $\mathbb{V}[\|f_{CP(R)}(\mathbf{x})\|^2] \leq \frac{1}{k} \left(3^{N-1} \left(1 + \frac{2}{R}\right) - 1\right) \|\mathbf{x}\|^4$

# Main Result

## Theorem

Let  $\mathbf{x} \in \mathbb{R}^{d^N}$  and  $R \in \mathbb{N}$ .

The RP maps  $f_{TT(R)}$  and  $f_{CP(R)}$  satisfy the following properties:

- $\mathbb{E}[\|f_{CP(R)}(\mathbf{x})\|^2] = \mathbb{E}[\|f_{TT(R)}(\mathbf{x})\|^2] = \|\mathbf{x}\|^2$
- $\mathbb{V}[\|f_{TT(R)}(\mathbf{x})\|^2] \leq \frac{1}{k} \left( 3 \left( 1 + \frac{2}{R} \right)^{N-1} - 1 \right) \|\mathbf{x}\|^4$
- $\mathbb{V}[\|f_{CP(R)}(\mathbf{x})\|^2] \leq \frac{1}{k} \left( 3^{N-1} \left( 1 + \frac{2}{R} \right) - 1 \right) \|\mathbf{x}\|^4$

↪ The bounds on the variances are substantially different...

# Main Result

## Theorem

Let  $\mathbf{x} \in \mathbb{R}^{d^N}$  and  $R \in \mathbb{N}$ .

The RP maps  $f_{TT(R)}$  and  $f_{CP(R)}$  satisfy the following properties:

- $\mathbb{E}[\|f_{CP(R)}(\mathbf{x})\|^2] = \mathbb{E}[\|f_{TT(R)}(\mathbf{x})\|^2] = \|\mathbf{x}\|^2$
- $\mathbb{V}[\|f_{TT(R)}(\mathbf{x})\|^2] \leq \frac{1}{k} \left( 3 \left( 1 + \frac{2}{R} \right)^{N-1} - 1 \right) \|\mathbf{x}\|^4$
- $\mathbb{V}[\|f_{CP(R)}(\mathbf{x})\|^2] \leq \frac{1}{k} \left( 3^{N-1} \left( 1 + \frac{2}{R} \right) - 1 \right) \|\mathbf{x}\|^4$

↪ The bounds on the variances are substantially different...

- More details in [arxiv:2003.05101](https://arxiv.org/abs/2003.05101).



## Comparison between $f_{CP}$ and $f_{TT}$

	$f_{CP(R)} : \mathbb{R}^{d^N} \rightarrow \mathbb{R}^k$	$f_{TT(R)} : \mathbb{R}^{d^N} \rightarrow \mathbb{R}^k$
Number of parameters	$\mathcal{O}(kNdR)$	$\mathcal{O}(kNdR^2)$
Computing $f(\mathbf{x})$ $\mathbf{x}$ in CP with rank $\tilde{R}$	$\mathcal{O}(kNd \max(r, \tilde{R}^2))$	$\mathcal{O}(kNd \max(r, \tilde{R}^3))$
Computing $f(\mathbf{x})$ $\mathbf{x}$ in TT with rank $\tilde{R}$	$\mathcal{O}(kNd \max(r, \tilde{R}^3))$	$\mathcal{O}(kNd \max(r, \tilde{R}^3))$
With proba $\geq 1 - \delta$ , $\mathbb{P}(\ f(\mathbf{x})\ ^2 = (1 \pm \epsilon)\ \mathbf{x}\ ^2)$ as soon as $k \gtrsim \dots$	$\frac{3^{N-1}(1+2/R)}{\epsilon^2} \log^{2N} \left(\frac{1}{\delta}\right)$	$\frac{(1+2/R)^N}{\epsilon^2} \log^{2N} \left(\frac{1}{\delta}\right)$

## Comparison between $f_{CP}$ and $f_{TT}$

	$f_{CP(R)} : \mathbb{R}^{d^N} \rightarrow \mathbb{R}^k$	$f_{TT(R)} : \mathbb{R}^{d^N} \rightarrow \mathbb{R}^k$
Number of parameters	$\mathcal{O}(kNdR)$	$\mathcal{O}(kNdR^2)$
Computing $f(\mathbf{x})$ $\mathbf{x}$ in CP with rank $\tilde{R}$	$\mathcal{O}(kNd \max(r, \tilde{R}^2))$	$\mathcal{O}(kNd \max(r, \tilde{R}^3))$
Computing $f(\mathbf{x})$ $\mathbf{x}$ in TT with rank $\tilde{R}$	$\mathcal{O}(kNd \max(r, \tilde{R}^3))$	$\mathcal{O}(kNd \max(r, \tilde{R}^3))$
With proba $\geq 1 - \delta$ , $\mathbb{P}(\ f(\mathbf{x})\ ^2 = (1 \pm \varepsilon) \ \mathbf{x}\ ^2)$ as soon as $k \gtrsim \dots$	$\frac{3^{N-1}(1+2/R)}{\varepsilon^2} \log^{2N} \left( \frac{1}{\delta} \right)$	$\frac{(1+2/R)^N}{\varepsilon^2} \log^{2N} \left( \frac{1}{\delta} \right)$

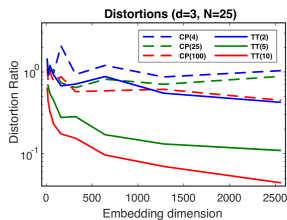
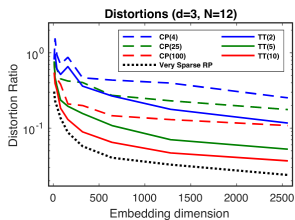
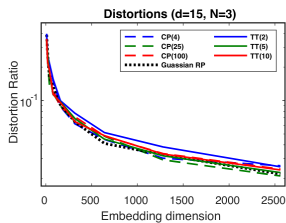
- Lower bounds on  $k$  suggest that  $f_{TT}$  is a better RP than  $f_{CP}$ .
- Classical Gaussian RP needs  $k \gtrsim \frac{1}{\varepsilon^2} \log \left( \frac{1}{\delta} \right)$ .
- Comparisons with other approaches: see [arxiv:2003.05101](https://arxiv.org/abs/2003.05101).

# Experiments

# Experiment Setup

- Compare  $f_{TT}$ ,  $f_{CP}$  and classical Gaussian RP to project  $d^N$  dimensional vectors
  - ▶ small order:  $d = 15, N = 3$
  - ▶ medium order:  $d = 3, N = 12$
  - ▶ higher order:  $d = 3, N = 25$
- Input  $\mathbf{x}$  is a random unit-norm TT vector with rank  $\tilde{R} = 10$ .
- Metric: distortion ratio  $\frac{\|f(\mathbf{x})\|^2}{\|\mathbf{x}\|^2} - 1$
- Report averages over 100 trials

# Experiment Results



# Conclusion

# Discussion

- We proposed an efficient way to tensorize classical Gaussian RP
- Theory and experiments suggest that TT is better suited than CP for very high dimensional RP

# Discussion

- We proposed an efficient way to tensorize classical Gaussian RP
- Theory and experiments suggest that TT is better suited than CP for very high dimensional RP
- Future work:
  - ▶ Leverage results to design efficient linear regression and SVD algorithms
  - ▶ Beyond classical tensor decomposition: other TN structures better suited for RP?
  - ▶ Study of statistical properties of TT vectors with random Gaussian cores



Thank you! Questions?