Probabilistic Tensor Train Decomposition

Jesper Løve Hinrich and Morten Mørup

Jesper Løve
Hinrich

Morten Mørup

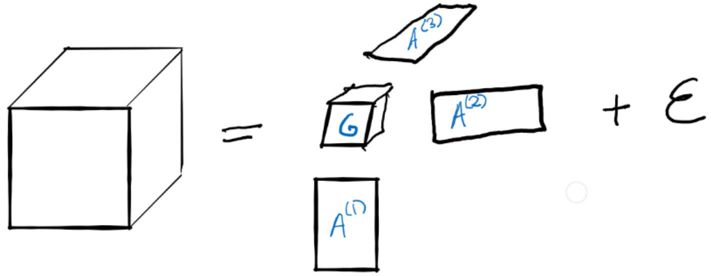
DTU Compute
Department of Applied Mathematics and Computer Science

27th European Signal Processing Conference,
A Coruña, Spain, Sept. 2-6, 2019

Tensor Decomposition

- Tucker

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$



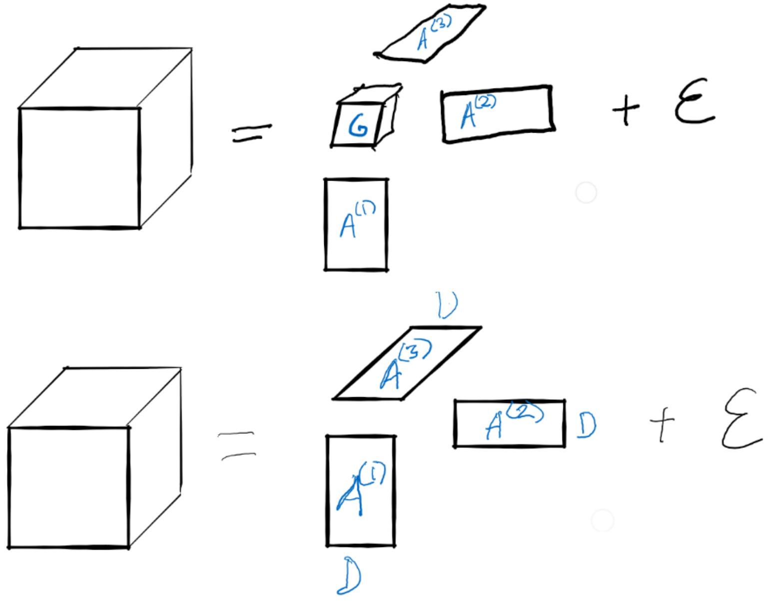
Tensor Decomposition

- Tucker

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$

- Candecomp/PARAFAC (CPD)

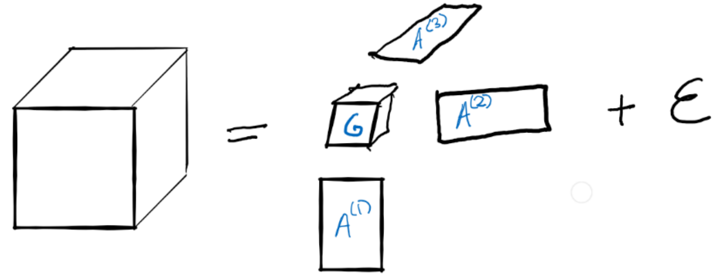
$$\boldsymbol{\mathcal{X}} = \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$



Tensor Decomposition

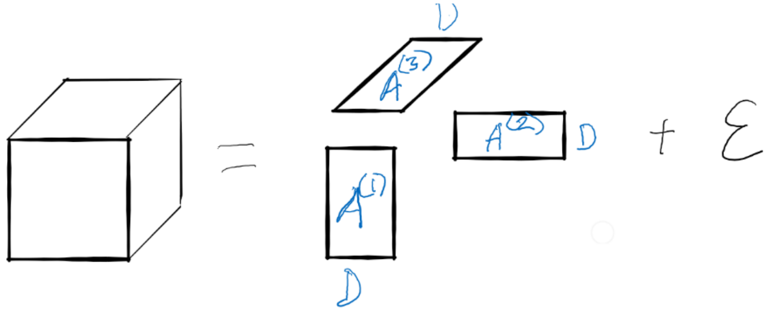
- Tucker

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$



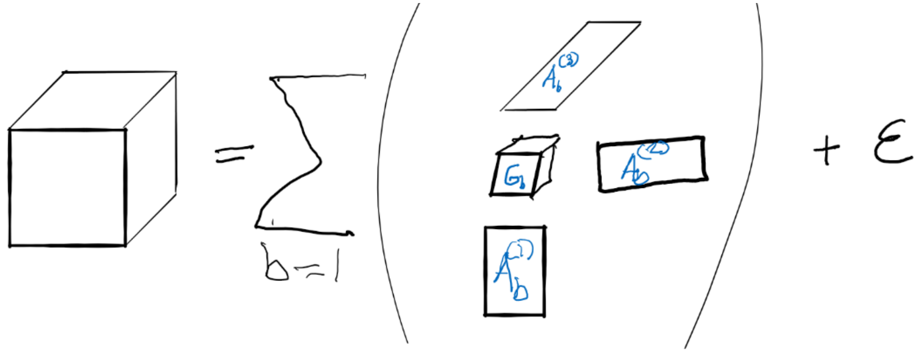
- Candecomp/PARAFAC (CPD)

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{I}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$



- Block-Term (BTD)

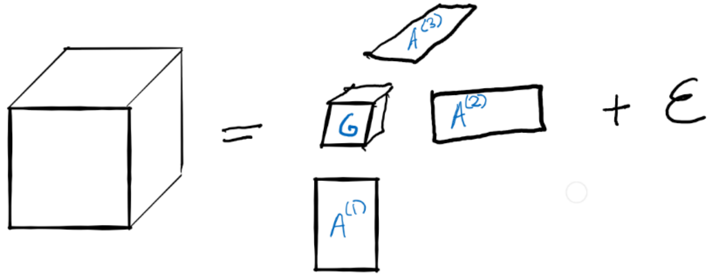
$$\boldsymbol{\mathcal{X}} = \sum_{b=1}^{\#Blocks} \boldsymbol{\mathcal{G}}_b \times_1 \mathbf{A}_b^{(1)} \times_2 \mathbf{A}_b^{(2)} \times_3 \ldots \times_N \mathbf{A}_b^{(N)}$$



Tensor Decomposition

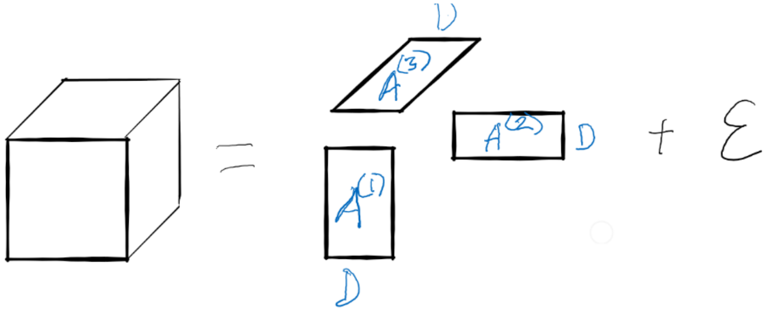
- Tucker

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$



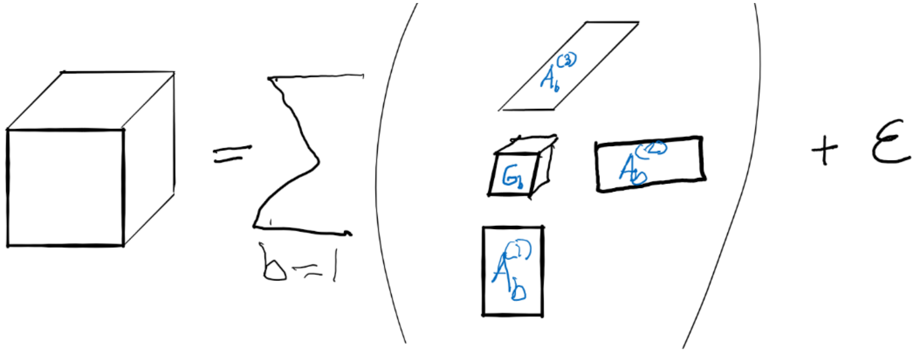
- Candecomp/PARAFAC (CPD)

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{I}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$



- Block-Term (BTD)

$$\boldsymbol{\mathcal{X}} = \sum_{b=1}^{\#Blocks} \boldsymbol{\mathcal{G}}_b \times_1 \mathbf{A}_b^{(1)} \times_2 \mathbf{A}_b^{(2)} \times_3 \ldots \times_N \mathbf{A}_b^{(N)}$$

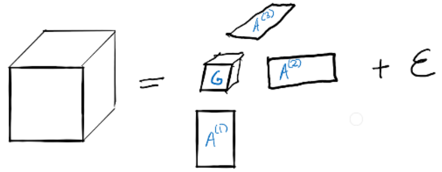


- Tensor Train (TTD)

$$\boldsymbol{\mathcal{X}} = \mathbf{U}^{(1)} \times_{[2,1]} \boldsymbol{\mathcal{U}}^{(2)} \times_{[3,1]} \boldsymbol{\mathcal{U}}^{(3)} \times_{[4,1]} \cdots \times_{[N-1,1]} \boldsymbol{\mathcal{U}}^{(N-1)} \times_{[N,1]} \mathbf{U}^{(N)}$$

Tensor Decomposition

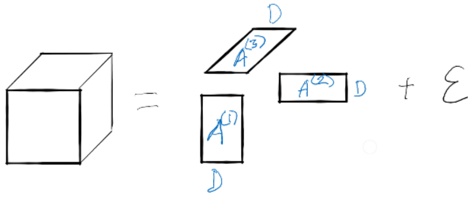
- Tucker



$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d_1} \sum_{d_2} \cdots \sum_{d_N} g_{d_1,d_2,\ldots,d_N} a^{(1)}_{i_1,d_1} a^{(2)}_{i_2,d_2} \cdots a^{(N)}_{i_N,d_N}$$

- Candecomp/PARAFAC (CPD)



$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{I}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d} a^{(1)}_{i_1,d_1} a^{(2)}_{i_2,d_2} \cdots a^{(N)}_{i_N,d_N}$$

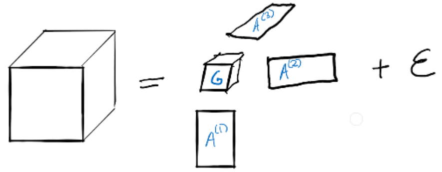
- Tensor Train (TTD)

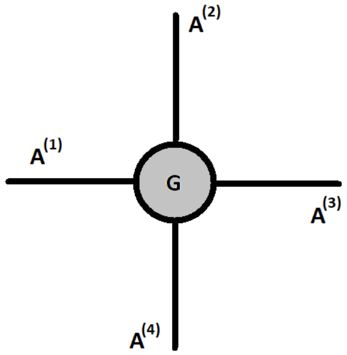
$$\boldsymbol{\mathcal{X}} = \mathbf{U}^{(1)} \times_{[2,1]} \boldsymbol{\mathcal{U}}^{(2)} \times_{[3,1]} \boldsymbol{\mathcal{U}}^{(3)} \times_{[4,1]} \cdots \times_{[N-1,1]} \boldsymbol{\mathcal{U}}^{(N-1)} \times_{[N,1]} \mathbf{U}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d_0} \sum_{d_1} \cdots \sum_{d_N} u^{(1)}_{d_0,i_1,d_1} u^{(2)}_{d_1,i_2,d_2} \cdots u^{(N)}_{d_{N-1},i_N,d_N}$$

Tensor Decomposition

- Tucker

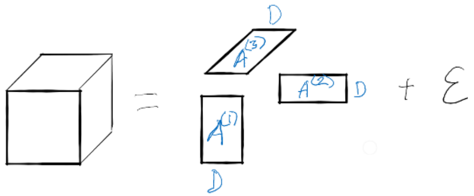


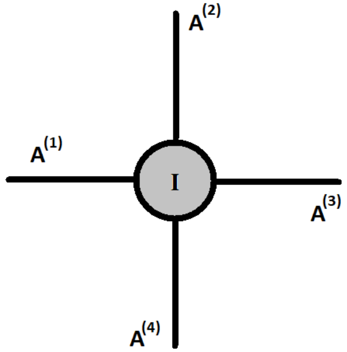


$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d_1}\sum_{d_2}\cdots\sum_{d_N} g_{d_1,d_2,\ldots,d_N} a^{(1)}_{i_1,d_1} a^{(2)}_{i_2,d_2} \cdots a^{(N)}_{i_N,d_N}$$

- Candecomp/PARAFAC (CPD)





$$\mathcal{X} = \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d} a^{(1)}_{i_1,d_1} a^{(2)}_{i_2,d_2} \cdots a^{(N)}_{i_N,d_N}$$

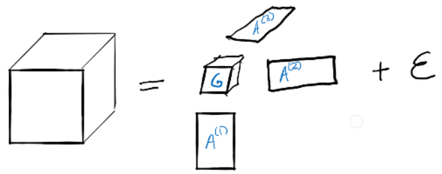
- Tensor Train (TTD)

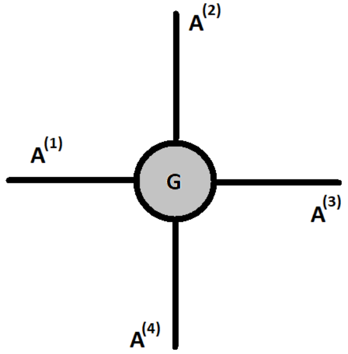
$$\mathcal{X} = \mathbf{U}^{(1)} \times_{[2,1]} \mathcal{U}^{(2)} \times_{[3,1]} \mathcal{U}^{(3)} \times_{[4,1]} \cdots \times_{[N-1,1]} \mathcal{U}^{(N-1)} \times_{[N,1]} \mathbf{U}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d_0}\sum_{d_1}\cdots\sum_{d_N} u^{(1)}_{d_0,i_1,d_1} u^{(2)}_{d_1,i_2,d_2} \cdots u^{(N)}_{d_{N-1},i_N,d_N}$$

Tensor Decomposition

- Tucker

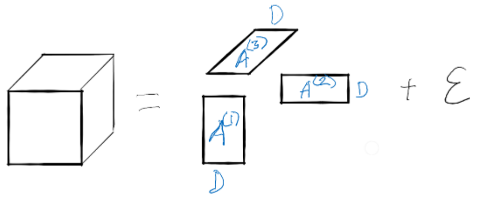


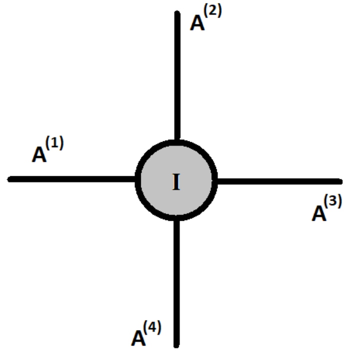


$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d_1}\sum_{d_2}\cdots\sum_{d_N} g_{d_1,d_2,\ldots,d_N} a^{(1)}_{i_1,d_1} a^{(2)}_{i_2,d_2} \cdots a^{(N)}_{i_N,d_N}$$

- Candecomp/PARAFAC (CPD)





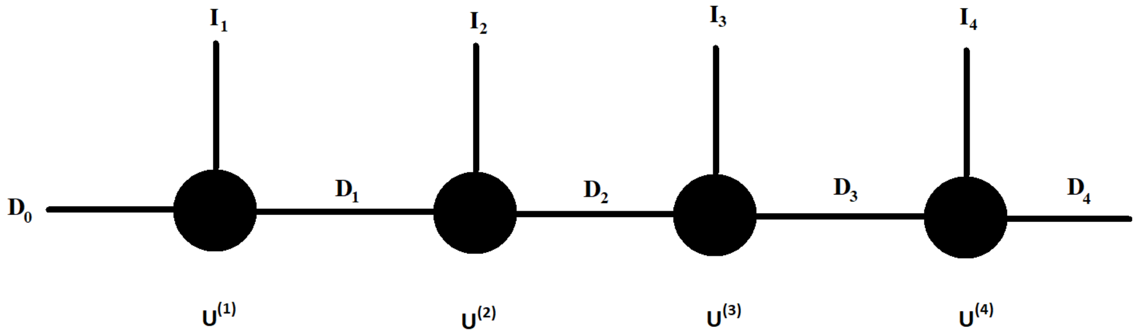
$$\mathcal{X} = \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \ldots \times_N \mathbf{A}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d} a^{(1)}_{i_1,d_1} a^{(2)}_{i_2,d_2} \cdots a^{(N)}_{i_N,d_N}$$

- Tensor Train (TTD)

$$\mathcal{X} = \mathbf{U}^{(1)} \times_{[2,1]} \boldsymbol{\mathcal{U}}^{(2)} \times_{[3,1]} \boldsymbol{\mathcal{U}}^{(3)} \times_{[4,1]} \cdots \times_{[N-1,1]} \boldsymbol{\mathcal{U}}^{(N-1)} \times_{[N,1]} \mathbf{U}^{(N)}$$

$$x_{i_1,i_2,\ldots,i_N} = \sum_{d_0}\sum_{d_1}\cdots\sum_{d_N} u^{(1)}_{d_0,i_1,d_1} u^{(2)}_{d_1,i_2,d_2} \cdots u^{(N)}_{d_{N-1},i_N,d_N}$$



How Many Parameters?

- Number modes, N
- For each mode:
  - Observations, I_n
  - Components, D_n

$$\texttt{Tucker}: \quad \sum_{n=1}^{N} D_n I_n + \prod_{n=1}^{N} D_n$$

$$\texttt{CP}: \quad D \sum_{n=1}^{N} I_n$$

$$\texttt{BTD}: \quad \sum_{b=1}^{\#Blocks} \left(\prod_{n=1}^{N} D_n^b + \sum_{n=1}^{N} D_n^b I_n \right), \qquad \sum_{b=1}^{\#Blocks} D_n^b = D_n$$

$$\texttt{TensorTrain}: \quad \sum_{n=1}^{N} D_{n-1} I_n D_n, \quad D_0 = D_N = 1$$

How Many Parameters?

- Number modes, N
- For each mode:
  - Observations, I_n
  - Components, D_n

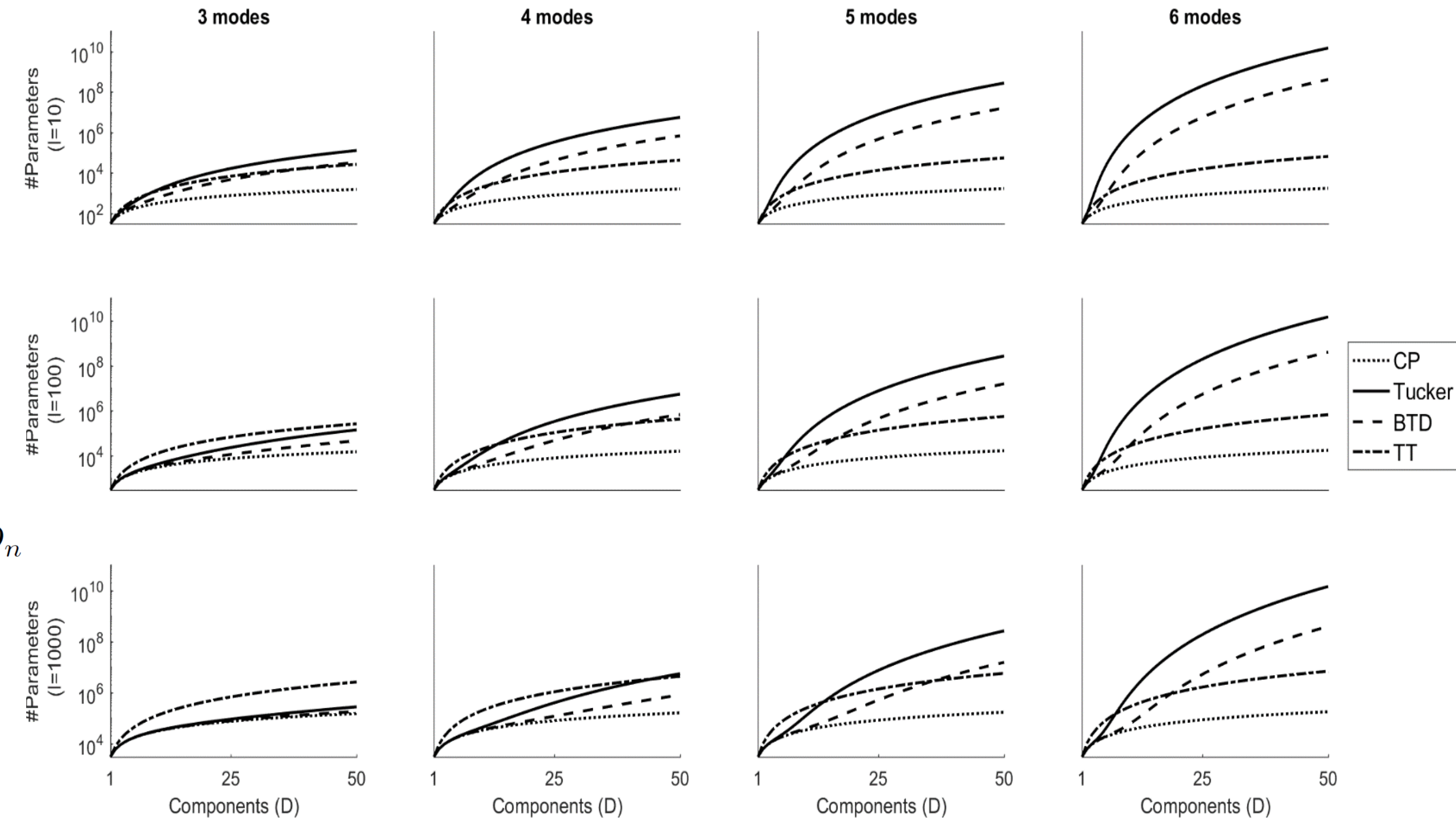
$$\text{Tucker}: \quad \sum_{n=1}^{N} D_n I_n + \prod_{n=1}^{N} D_n$$

$$\text{CP}: \quad D \sum_{n=1}^{N} I_n$$

$$\text{BTD}: \quad \sum_{b=1}^{\#Blocks} \left(\prod_{n=1}^{N} D_n^b + \sum_{n=1}^{N} D_n^b I_n \right), \qquad \sum_{b=1}^{\#Blocks} D_n^b = D_n$$

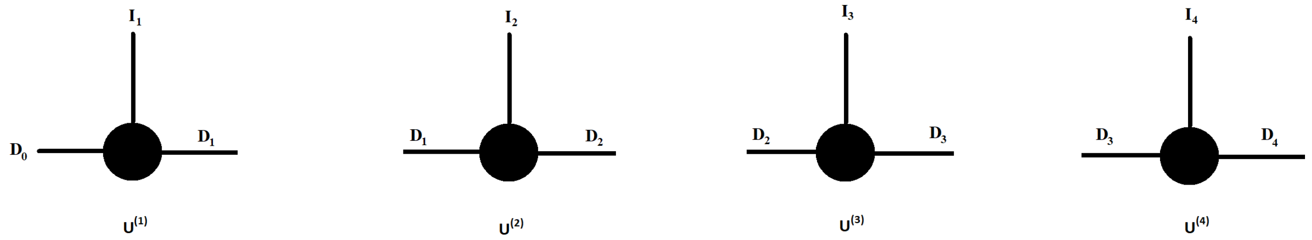
$$\text{TensorTrain}: \quad \sum_{n=1}^{N} D_{n-1} I_n D_n, \quad D_0 = D_N = 1$$

Parameter Growth in I,D, and N

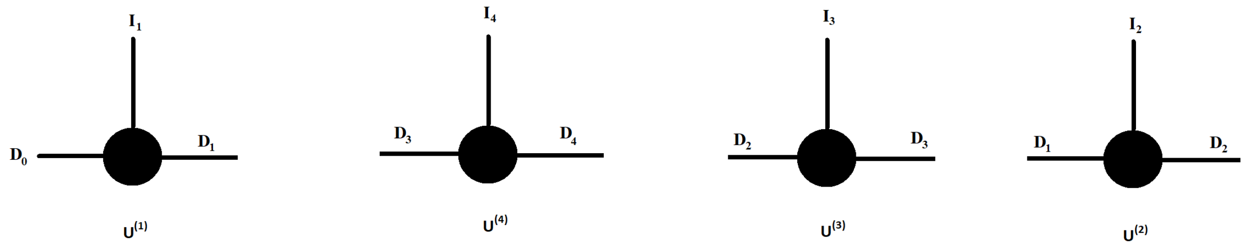


Tensor Trains are Asymmetric!

- Tensor train split into carts (factors)

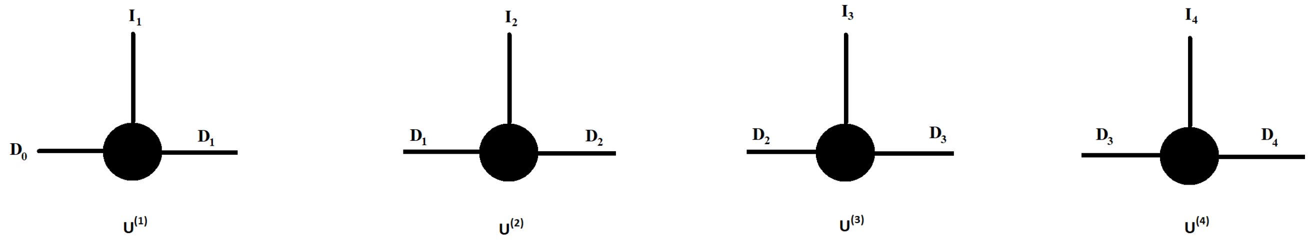


- Permuting the factors results in a different TT decomposition

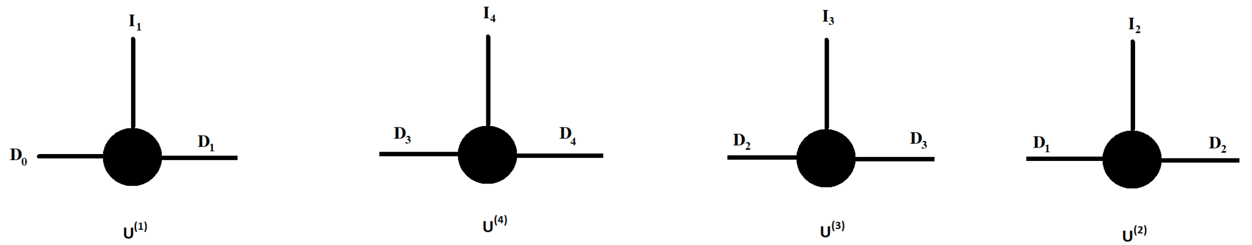


Tensor Trains are Asymmetric!

- Tensor train split into carts (factors)



- Permuting the factors results in a different TT decomposition



- Not an issue in Tucker and CP

Determining the Tensor Train by

Maximum Likelihood Based (Oseledets [1])

- No way to validate model or mode order
- Doesn’t remove noise
- Doesn’t account for uncertainty

- Fast implementation (nested SVD)
- Lots of known theoretical properties
- Widely applied

[1] Oseledets IV. Tensor-train decomposition. SIAM Journal on Scientific Computing. 2011 Sep 22;33(5):2295-317.

[2] Hinrich JL., Mørup M. Probabilistic Tensor Train Decomposition. 27th European Signal Processing Conference (EUSIPCO) 2019.

Determining the Tensor Train by

Maximum Likelihood Based (Oseledets [1])

- No way to validate model or mode order
- Doesn't remove noise
- Doesn't account for uncertainty

- Fast implementation (nested SVD)
- Lots of known theoretical properties
- Widely applied

Bayesian Inference (Hinrich & Mørup [2])

- Validate model and mode order via log-evidence
- Penalizes model complexity (simpler is better)
- Removes unstructured noise
- Uncertainty characterized though probability distributions

- Slower implementation (iterative)
- Theoretical properties less known (currently)

[1] Oseledets IV. Tensor-train decomposition. SIAM Journal on Scientific Computing. 2011 Sep 22;33(5):2295-317.

[2] Hinrich JL., Mørup M. Probabilistic Tensor Train Decomposition. 27th European Signal Processing Conference (EUSIPCO) 2019.

From ML to Bayes

Maximum Likelihood TT

$$
\begin{aligned}
\boldsymbol{\mathcal{X}} =& \mathbf{U}^{(1)} \times_{[2,1]} \boldsymbol{\mathcal{U}}^{(2)} \times_{[3,1]} \boldsymbol{\mathcal{U}}^{(3)} \times_{[4,1]} \cdots \\
& \times_{[N-1,1]} \boldsymbol{\mathcal{U}}^{(N-1)} \times_{[N,1]} \mathbf{U}^{(N)}
\end{aligned}
$$

- Found through nested SVDs

- Each unfolded train cart is orthogonal, i.e.

$$
{\mathbf{U}_{(1,2)}^{(n)}}^{\top} \mathbf{U}_{(1,2)}^{(n)} = \mathbf{I}, \quad n = 1, 2, \ldots, N
$$

From ML to Bayes

Maximum Likelihood TT

$$\mathcal{X} = \mathbf{U}^{(1)} \times_{[2,1]} \mathcal{U}^{(2)} \times_{[3,1]} \mathcal{U}^{(3)} \times_{[4,1]} \cdots \times_{[N-1,1]} \mathcal{U}^{(N-1)} \times_{[N,1]} \mathbf{U}^{(N)}$$

- Found through nested SVDs
- Each unfolded train cart is orthogonal, i.e.

$$\mathbf{U}_{(1,2)}^{(n)\top} \mathbf{U}_{(1,2)}^{(n)} = \mathbf{I}, \quad n = 1, 2, \ldots, N$$

Probabilistic TT

$$\mathcal{X} \approx \mathcal{M} = \mathcal{U}^{(1)} \times_{[2,1]} \mathcal{U}^{(2)} \times_{[3,1]} \mathcal{U}^{(3)} \times_{[4,1]} \cdots \times_{[M-1,1]} \mathcal{U}^{(M-1)} \times_{[M,1]} (\mathbf{S}\mathbf{V}^\top)$$

- Note $\mathcal{U}^{(M)} \equiv \mathbf{S}\mathbf{V}^\top$ and $\mathbf{U}_{(1,2)}^{(m)\top} \mathbf{U}_{(1,2)}^{(m)} = \mathbf{I}$.

From ML to Bayes

Maximum Likelihood TT

$$\mathcal{X} = \mathbf{U}^{(1)} \times_{[2,1]} \mathcal{U}^{(2)} \times_{[3,1]} \mathcal{U}^{(3)} \times_{[4,1]} \cdots \times_{[N-1,1]} \mathcal{U}^{(N-1)} \times_{[N,1]} \mathbf{U}^{(N)}$$

- Found through nested SVDs
- Each unfolded train cart is orthogonal, i.e.

$$\mathbf{U}_{(1,2)}^{(n)\top} \mathbf{U}_{(1,2)}^{(n)} = \mathbf{I}, \quad n = 1, 2, \ldots, N$$

Probabilistic TT

$$\mathcal{X} \approx \mathcal{M} = \mathcal{U}^{(1)} \times_{[2,1]} \mathcal{U}^{(2)} \times_{[3,1]} \mathcal{U}^{(3)} \times_{[4,1]} \cdots \times_{[M-1,1]} \mathcal{U}^{(M-1)} \times_{[M,1]} (\mathbf{S}\mathbf{V}^\top)$$

- Note $\mathcal{U}^{(M)} \equiv \mathbf{S}\mathbf{V}^\top$ and $\mathbf{U}_{(1,2)}^{(m)\top} \mathbf{U}_{(1,2)}^{(m)} = \mathbf{I}$.
- Found by inferring the generative model,

$$\begin{aligned}
\tau &\sim \mathcal{G}(\alpha_\tau, \beta_\tau), \qquad \lambda \sim \mathcal{G}(\alpha_\lambda, \beta_\lambda), \\
s_{dd} &\sim \mathcal{TN}_{[0,\infty]}\left(0, \lambda^{-1}\right), \quad d = 1, \ldots, D_{M-1} \\
\mathbf{V} &\sim v\mathcal{MF}\left(\mathbf{0}^{I_M \times D_{M-1}}\right), \\
\mathbf{U}_{(1,2)}^{(m)} &\sim v\mathcal{MF}\left(\mathbf{0}^{D_{m-1} I_m \times D_m}\right), m = 1, \ldots, M-1 \\
\mathcal{X}|\boldsymbol{\theta} &\sim \mathcal{N}_{I_1 \times I_2 \times \cdots \times I_M}\left(\mathcal{M}, \mathbf{I}_{I_1}\tau^{-1}, \mathbf{I}_{I_2}, \ldots, \mathbf{I}_{I_M}\right)
\end{aligned}$$

where $\boldsymbol{\theta} = \{\tau, \lambda, \mathbf{S}, \mathbf{V}, \mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \ldots, \mathcal{U}^{(M-1)}\}$

$$\alpha_\tau = \beta_\tau = \alpha_\lambda = \beta_\lambda = 10^{-6}$$

Probabilistic Tensor Train

$$\mathcal{X} \approx \mathcal{M} = \mathcal{U}^{(1)} \times_{[2,1]} \mathcal{U}^{(2)} \times_{[3,1]} \mathcal{U}^{(3)} \times_{[4,1]} \cdots \times_{[M-1,1]} \mathcal{U}^{(M-1)} \times_{[M,1]} (\mathbf{S}\mathbf{V}^\top)$$

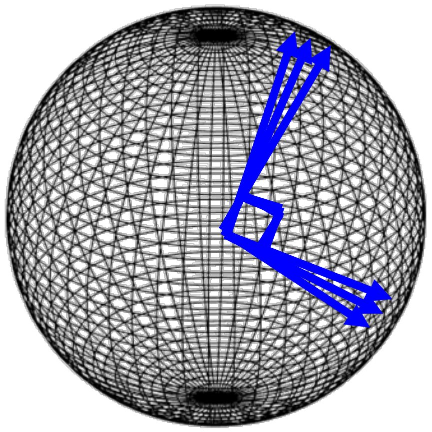
- Note $\mathcal{U}^{(M)} \equiv \mathbf{S}\mathbf{V}^\top$ and $\mathbf{U}_{(1,2)}^{(m)\top} \mathbf{U}_{(1,2)}^{(m)} = \mathbf{I}$.

- Generative model,

$$\begin{aligned}
\tau &\sim \mathcal{G}(\alpha_\tau, \beta_\tau), \qquad \lambda \sim \mathcal{G}(\alpha_\lambda, \beta_\lambda), \\
s_{dd} &\sim \mathcal{TN}_{[0,\infty]}\left(0, \lambda^{-1}\right), \quad d = 1, \ldots, D_{M-1} \\
\mathbf{V} &\sim v\mathcal{MF}\left(\mathbf{0}^{I_M \times D_{M-1}}\right), \\
\mathbf{U}_{(1,2)}^{(m)} &\sim v\mathcal{MF}\left(\mathbf{0}^{D_{m-1} I_m \times D_m}\right), m = 1, \ldots, M-1 \\
\mathcal{X}|\boldsymbol{\theta} &\sim \mathcal{N}_{I_1 \times I_2 \times \cdots \times I_M}\left(\mathcal{M}, \mathbf{I}_{I_1}\tau^{-1}, \mathbf{I}_{I_2}, \ldots, \mathbf{I}_{I_M}\right)
\end{aligned}$$

- Parameters: $\boldsymbol{\theta} = \{\tau, \lambda, \mathbf{S}, \mathbf{V}, \mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \ldots, \mathcal{U}^{(M-1)}\}$
- Broad priors: $\alpha_\tau = \beta_\tau = \alpha_\lambda = \beta_\lambda = 10^{-6}$

von Miss-Fisher Matrix (vMF) distribution



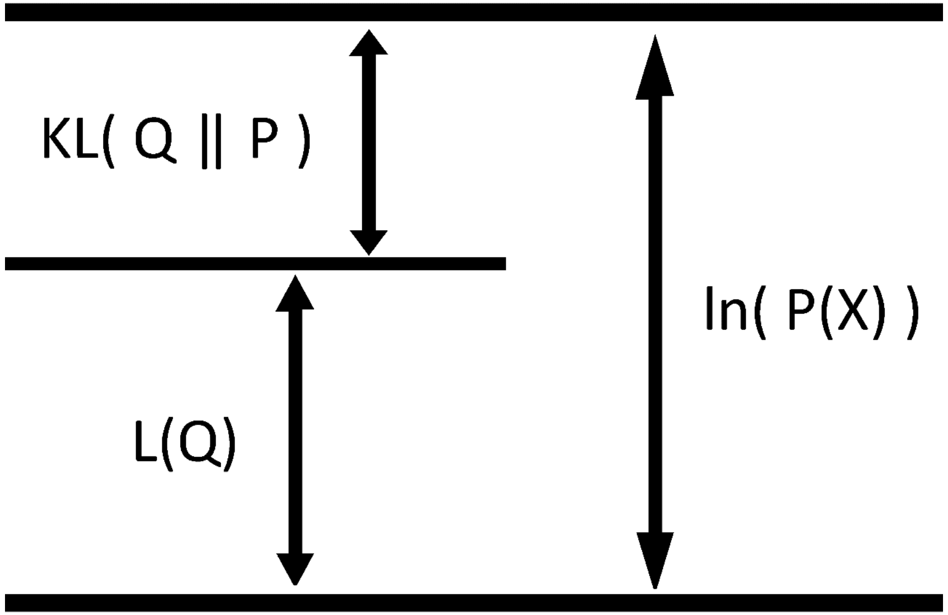
Model Inference by Variational Bayesian optimization

- The exact posterior is typically intractable

$$P(\boldsymbol{\theta}|\mathbf{X}) = \frac{P(\mathbf{X}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{X})}$$

- Approximation by Variational Bayesian (VB)
  - True distribution: $P(\boldsymbol{\theta}|\boldsymbol{X})$
  - Variational distribution: $Q(\boldsymbol{\theta})$

- Minimize KL-divergence



$$\log P(\boldsymbol{X}) = \log \int P(\boldsymbol{X}|\theta)P(\theta)d\theta \geq \int \log[P(\boldsymbol{X}|\theta)P(\theta)]Q(\theta)\, d\theta - \int \log[Q(\theta)]Q(\theta)\, d\theta$$

Alternating optimization of the ELBO

$$\tilde{F}^{(m)} = \langle \tau \rangle \left(\texttt{contract} \left(\mathcal{X}, \{ \left\langle \mathcal{U}^{(n)} \right\rangle \}_{n \neq m} \right) \right)_{(1,2)}$$
$$Q(\mathbf{U}^{(m)}_{(1,2)}) \sim v\mathcal{MF}(\tilde{F}^{(m)})$$

$$\tilde{\alpha}_\lambda = \alpha_\lambda + 0.5 D_{M-1} \quad , \quad \tilde{\beta}_\lambda = \beta_\lambda + 0.5\mathrm{trace}(\langle \mathbf{S}\mathbf{S}^\top \rangle),$$
$$Q(\lambda) \sim \mathcal{G}(\tilde{\alpha}_\lambda, \tilde{\beta}_\lambda)$$

$$\tilde{F}^{(M)^\top} = \langle \tau \rangle \ \langle \mathbf{S} \rangle \ \texttt{contract} \left(\mathcal{X}, \{ \left\langle \mathcal{U}^{(n)} \right\rangle \}_{n \neq M} \right)$$
$$Q(\mathbf{V}) \sim v\mathcal{MF}(\tilde{F}^{(M)})$$

$$\tilde{\alpha}_\tau = \alpha_\tau + 0.5 \prod_{m=1}^{M} I_m$$
$$\tilde{\beta}_\tau = \beta_\tau + 0.5 \Big[\mathrm{trace} \left(\mathcal{X} \times \mathcal{X} \right) + \mathrm{trace} \left(\langle \mathbf{S}\mathbf{S}^\top \rangle \right)$$
$$- 2\mathrm{trace} \left(\texttt{contract} \left(\mathcal{X}, \{\{ \left\langle \mathcal{U}^{(n)} \right\rangle \}_{n=1,\ldots,M} \right) \right) \Big]$$
$$Q(\tau) \sim \mathcal{G} \left(\tilde{\alpha}_\tau, \tilde{\beta}_\tau \right)$$

$$\mathbf{W}^{D_{M-1} \times I_M} = \texttt{contract} \left(\mathcal{X}, \{ \left\langle \mathcal{U}^{(n)} \right\rangle \}_{n \neq M} \right)$$
$$\sigma^2_{s_{dd}} = \left(\langle \tau \rangle + \langle \lambda \rangle \right)^{-1}, \quad \mu_{s_{dd}} = \sigma^2_{s_{dd}} \ \langle \mathbf{W}_d \rangle \ \langle \mathbf{v}_d \rangle \ \langle \tau \rangle$$
$$Q(\mathbf{s}_{dd}) \sim \mathcal{TN}_{[0,\infty]} \left(\mu_{s_{dd}}, \sigma^2_{s_{dd}} \right)$$

Tensor Train and the contract(X,U) operator

- Contracting the data **X** with the train carts **U**, except the m'th cart.

$$\mathcal{X}^{I_1 \times \ldots \times I_M}$$

$$\{\mathcal{U}_i \in \mathbb{R}^{D_{i-1} \times I_i \times D_i}\}_{i=1,\ldots,M}$$

$$D_0 = D_M = 1$$

$$\mathcal{C} = \texttt{contract}\left(\mathcal{X}, \{\mathcal{U}_i\}_{i \neq m}\right)$$
$$\mathcal{C} \in \mathbb{R}^{D_{m-1} \times I_m \times D_m}$$

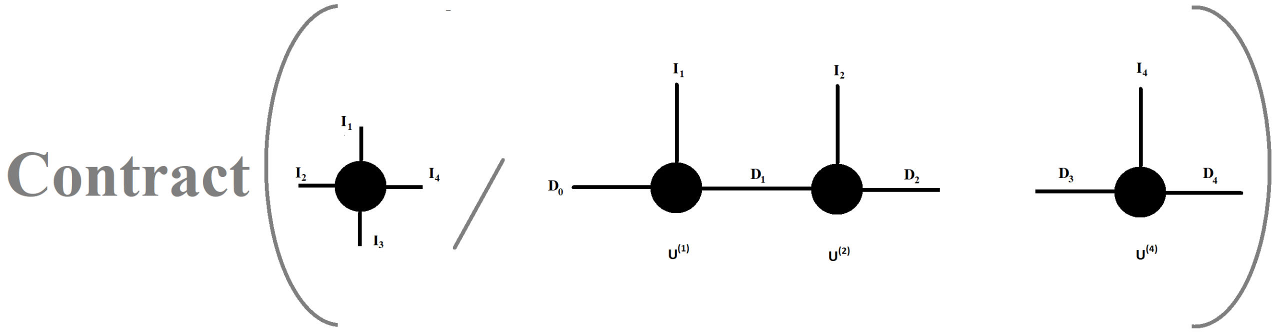
Tensor Train and the contract(X,U) operator

- Contracting the data **X** with the train carts **U**, except the m'th cart.

$$\mathcal{X}^{I_1 \times \ldots \times I_M}$$

$$\{\mathcal{U}_i \in \mathbb{R}^{D_{i-1} \times I_i \times D_i}\}_{i=1,\ldots,M}$$

$$D_0 = D_M = 1$$



$$\mathcal{C} = \mathtt{contract}\left(\mathcal{X}, \{\mathcal{U}_i\}_{i \neq m}\right)$$
$$\mathcal{C} \in \mathbb{R}^{D_{m-1} \times I_m \times D_m}$$

Tensor Train and the contract(X,U) operator

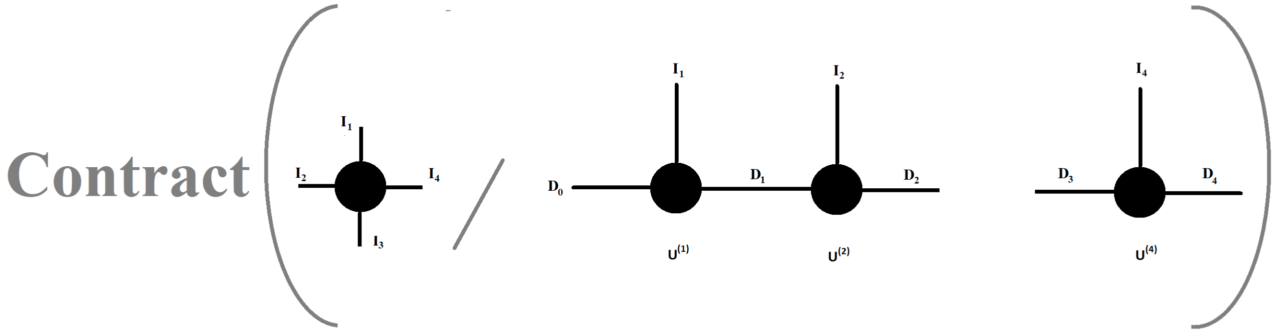
- Contracting the data **X** with the train carts **U**, except the m'th cart.

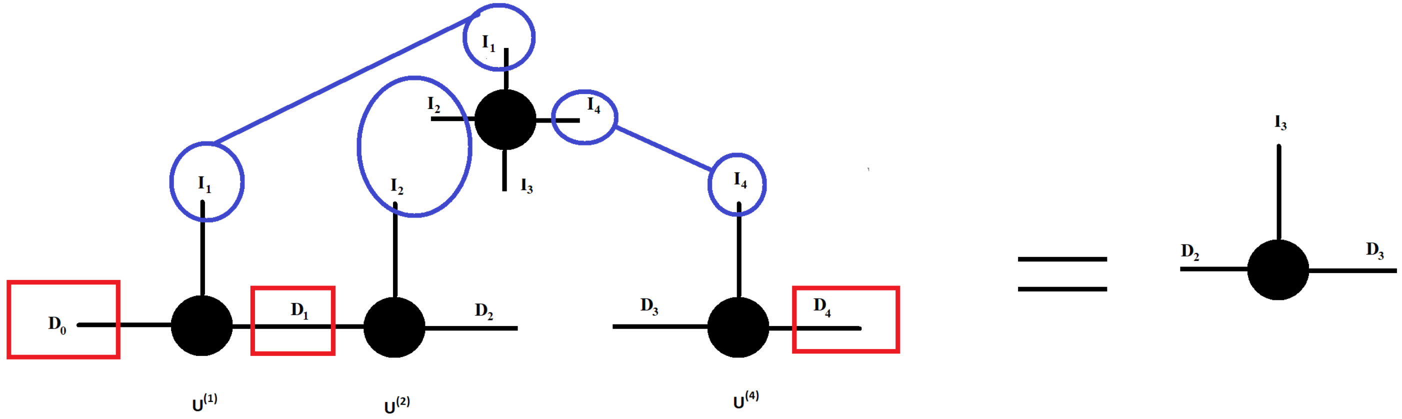
$$\mathcal{X}^{I_1 \times \ldots \times I_M}$$

$$\{\mathcal{U}_i \in \mathbb{R}^{D_{i-1} \times I_i \times D_i}\}_{i=1,\ldots,M}$$

$$D_0 = D_M = 1$$

$$\mathcal{C} = \texttt{contract}\left(\mathcal{X}, \{\mathcal{U}_i\}_{i \neq m}\right)$$
$$\mathcal{C} \in \mathbb{R}^{D_{m-1} \times I_m \times D_m}$$

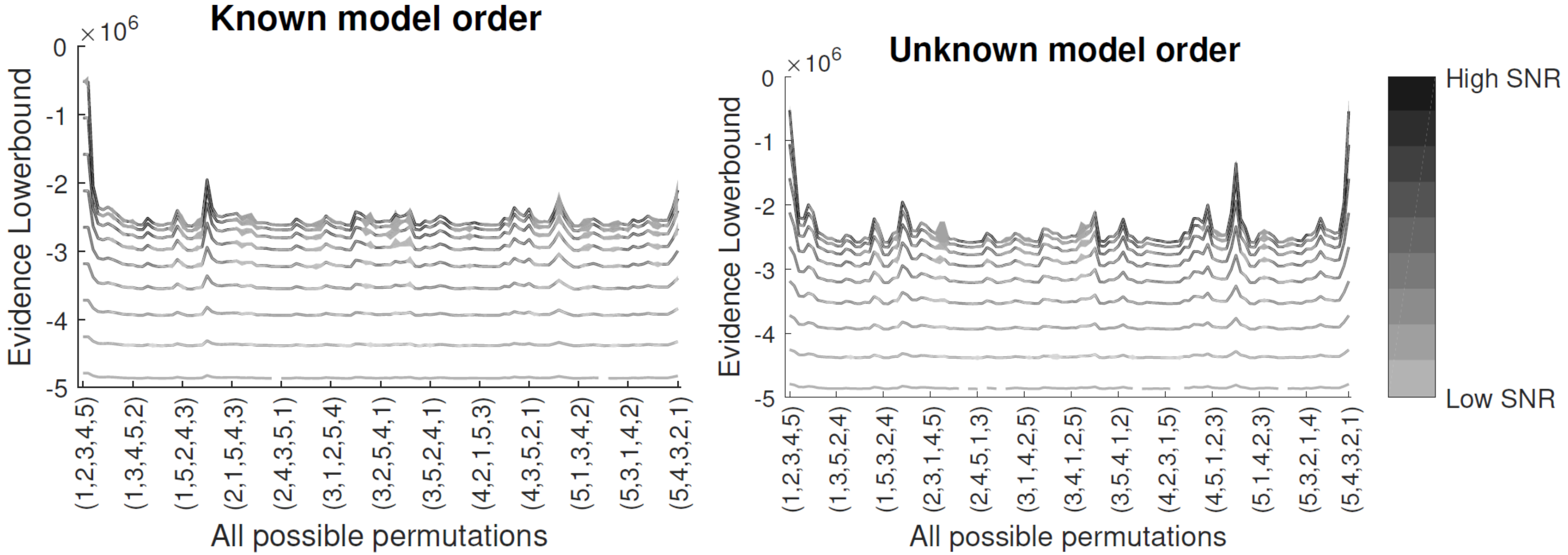




Simulated Experiment: PTTD

- Data generation (ground truth)
  - N=5 mode tensor
  - I=(20, 19, 18, 17, 16)
  - D=(1, 6, 5, 4, 3, 1)
  - True mode order (1, 2, 3, 4, 5)
- Two scenarios
  - Known D
  - Unknown D, but known max(D).
    $D_{est} = (1, D_{max}, D_{max}, D_{max}, D_{max}, 1)$
- For each scenario
  - Testing all 5!=120 mode order permutations
  - Varying signal-to-noise (SNR) ratio (-10 to 10 dB)
  - 10 Repeats

Simulated Experiment: PTTD



(a) (b)

Fig. 1: Performance of PTTD when the model order is known (a) and unknown (b). Higher ELBO indicates a better model.

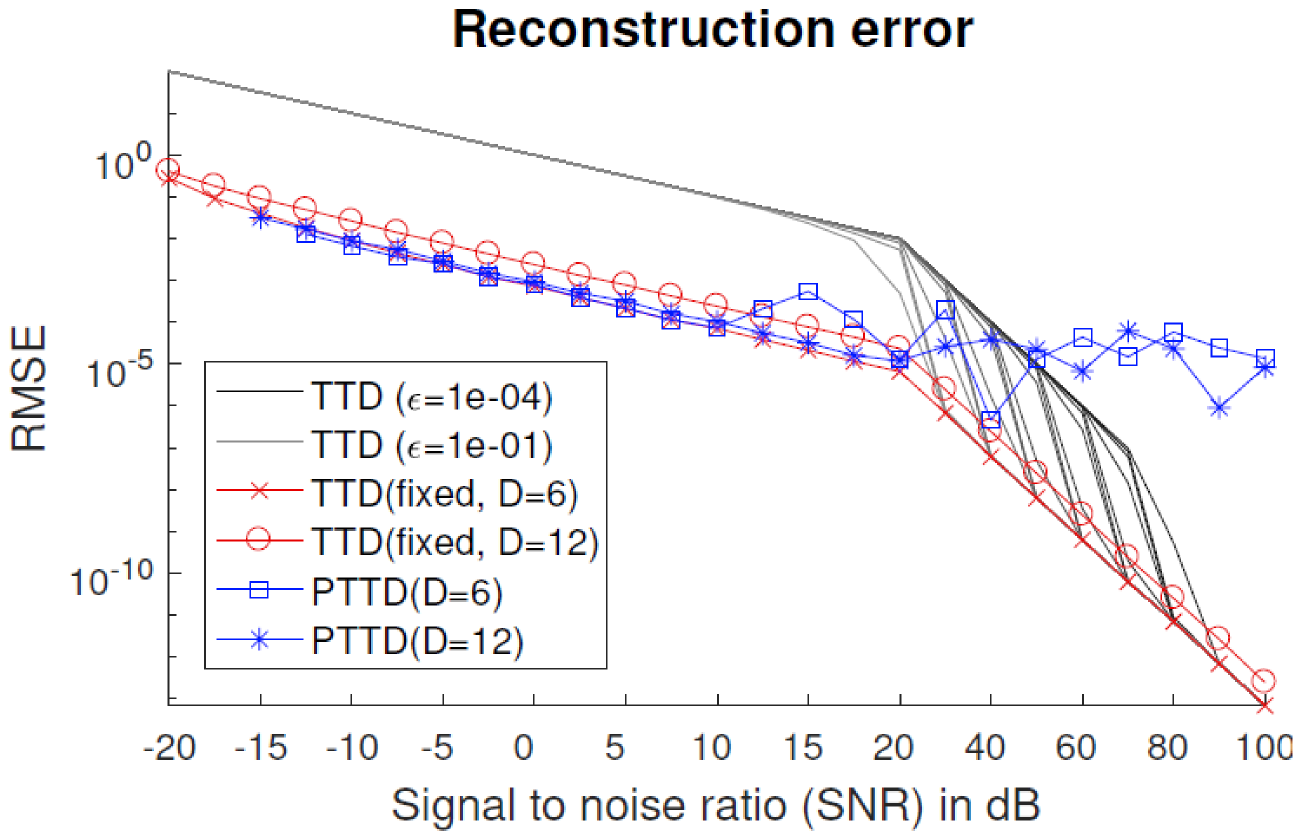
- Highest ELBO for the true mode order.
- High noise (low SNR) reduces identifiability.
- Permuting results in different performance.
- Special case for true (1,2,3,4,5) and reverse (5,4,3,2,1) mode order.

Simulated Experiment: TTD vs PTTD

- Data simulated as before.
  (only mode order (1,2,3,4,5))

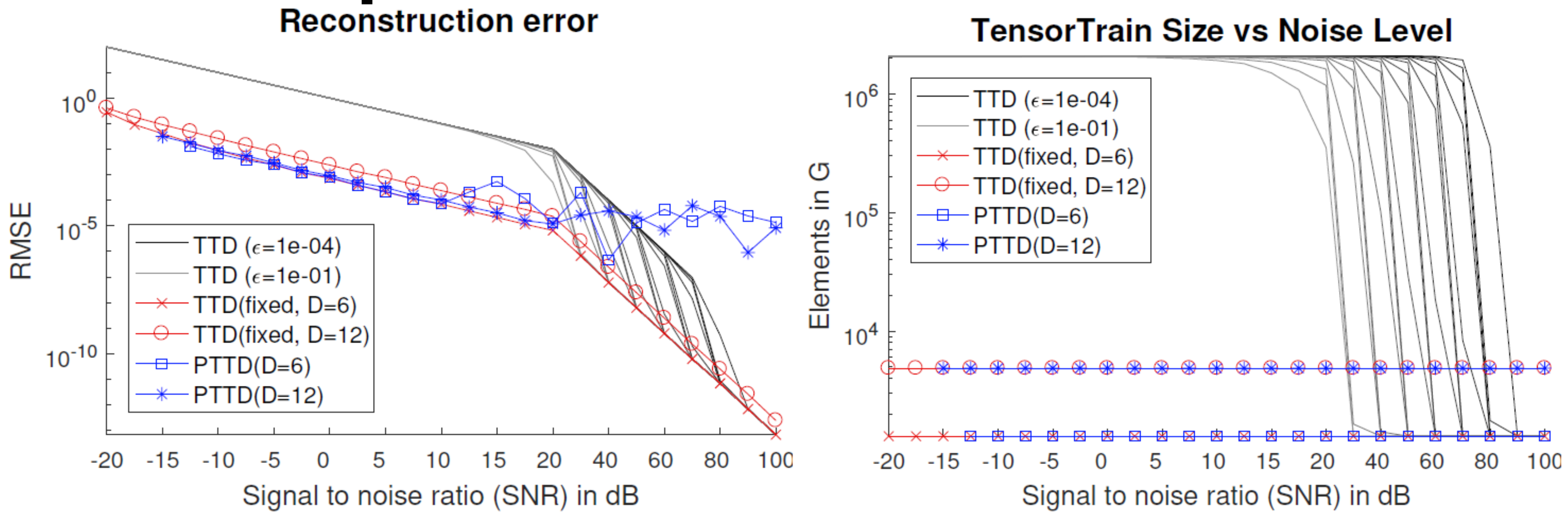
- Models
  - TTD $\boldsymbol{\varepsilon}$ (Oseledets)
    $\varepsilon = [10^{-4}, \ldots, 10^{-1}]$
  - PTTD and TTD(fixed)
    D=6 : (1,6,5,4,3,1)
    D=12 : (1,12,10,8,6,1)

- Recovery of noiseless data
  $RMSE(\boldsymbol{X_{truth}} - \boldsymbol{X_{recon}})$



(a) $\mathcal{X}_{\text{truth}} - \mathcal{X}_{\text{recon}}$

Simulated Experiment: TTD vs PTTD



(a) $\boldsymbol{\mathcal{X}}_{\text{truth}} - \boldsymbol{\mathcal{X}}_{\text{recon}}$ (b) $|\boldsymbol{\mathcal{G}}_{est}|$

- TTD- ε is highly sensitive to noise!
- Sufficient TT-size is much smaller for PTTD than TTD-ε, i.e. $|G_{PTTD}| \ll |G_{TTD-\varepsilon}|$.
- Both PTTD and TTD(fixed) with high noise.

Amino Acid Dataset

- 3-mode flourescence spectroscopy data
  - 5 samples
  - 201 emission levels (250-450nm)
  - 51 excitation levels (250-300nm)
  - Total elements 51255

- Original mode order
  - (samples, emission, excitation)
  - (1, 2, 3)

- Fit PTTD with $\boldsymbol{D} = (1, D_1, D_2, 1)$
  - for all mode orders (3!=6 permutaions)
  - Varying $D_1 = [1,2,..,30]$ and $D_2 = [1,2,..,30]$

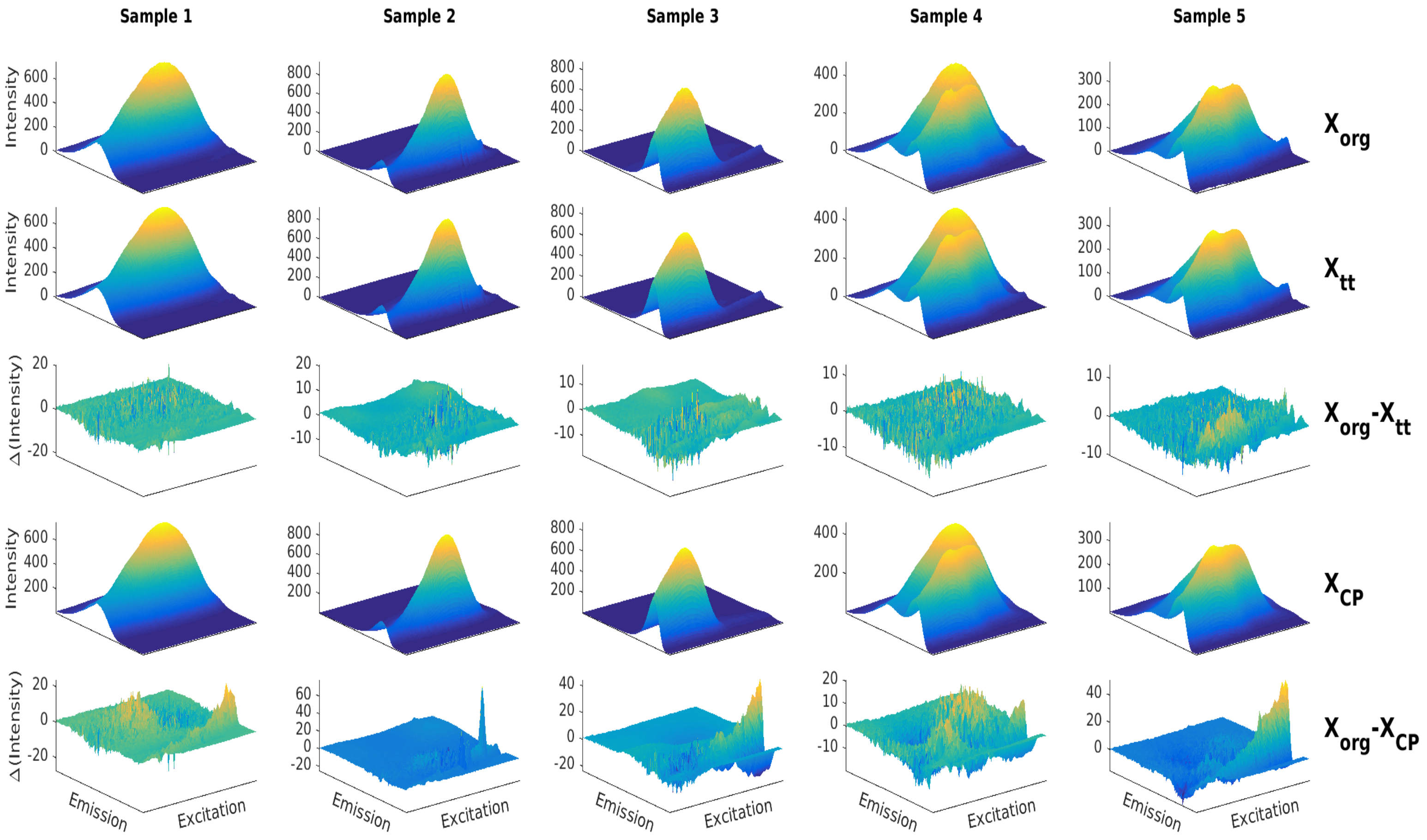
- Model with highest evidence lowerbound (ELBO)
  - Mode order (2,3,1)
  - (emission, excitation, samples)

- High compression in all modes

Mode Order	ELBO	$\mathrm{D}_{\mathrm{ELBO}}$	$\lvert\mathcal{G}_{ELBO}\rvert$
(1,2,3)	0.774	(4,3)	806
(1,3,2)	0.882	(4,5)	1330
(2,1,3)	0.969	(7,4)	1671
(2,3,1)	1.000	(6,5)	1536
(3,1,2)	0.967	(10,4)	1434
(3,2,1)	0.804	(4,5)	1274

Max. ELBO for each mode order.

Comparing PTTD and the 3 component CPD



- CPD removes unstructured and structured (Rayleight scattering) noise.
- PTTD only removes unstructured noise.
- NB. TTD-ε describes the data using more parameters than data elements.

End of the Line

Conclusions

- Prob. Tensor Train
  - Quantify mode (N) and model (D) order.
  - Remove unstructured noise.
  - Incorperate model uncertainty.
  - High compression.

- Facilitates model comparison using the evidence lowerbound
  - Maximum Likelihood TT always give lower RMSE when increasing model complexity.

- Probabilistic TT code available at **https://github.com/JesperLH/prob-tt**

End of the Line

Conclusions

- Prob. Tensor Train
  - Quantify mode (N) and model (D) order.
  - Remove unstructured noise.
  - Incorperate model uncertainty.
  - High compression.

- Facilitates model comparison using the evidence lowerbound
  - Maximum Likelihood TT always give lower RMSE when increasing model complexity.

- Probabilistic TT code available at **https://github.com/JesperLH/prob-tt**

Future work and ideas

- Efficient $\mathrm{contract}(\mathbf{X},\mathbf{U})$
- Input TTD(**X**) to PTTD
  - Allows evaluation of likely TTDs via PTTD
- Non-parametric
  - Automatically infer **D**
- Stochastic updates rules
- PTTD on PTTD operations (inspired by Oselets)
  - e.g. +, -, *, / PTTDs

Contact

jehi@dtu.dk

http://people.compute.dtu.dk/jehi/