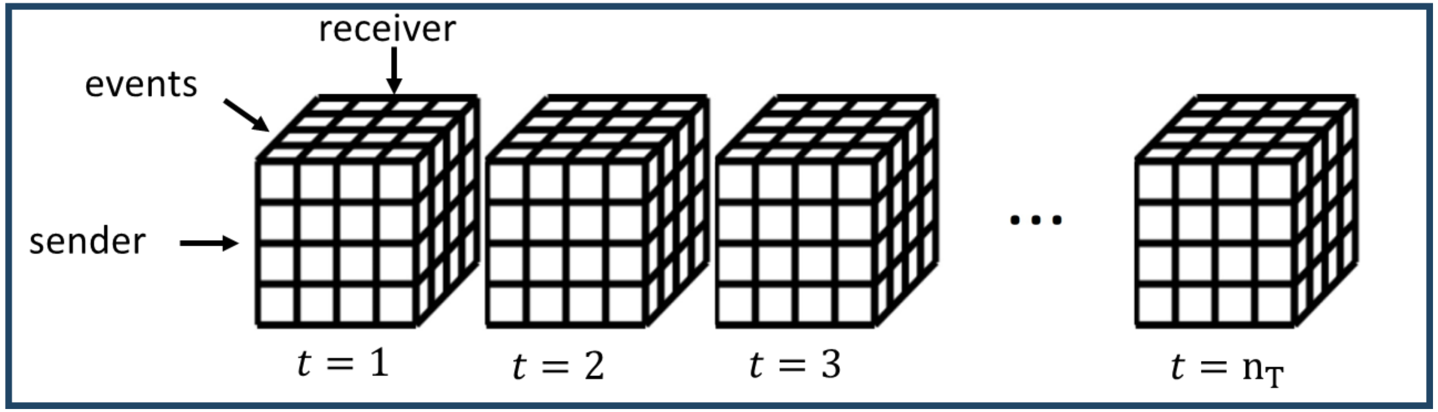
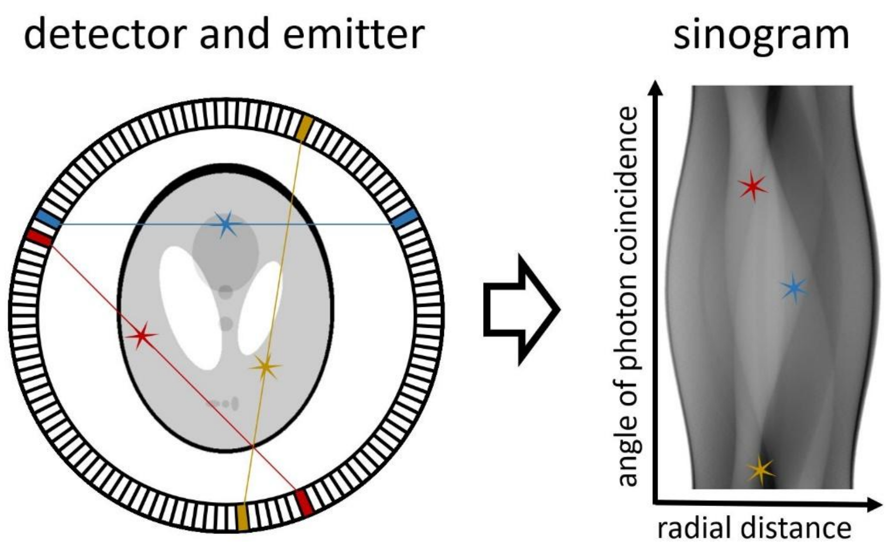
Poisson-response Tensor-on-Tensor Regression and Applications

Carlos Llosa and Daniel Dunlavy

SAND2026-24182C

Motivating problems





Temporal prediction of dyadic relationships

ICEWS database [1]

- Countries as receivers and senders
- Events such as threats or aid
- **Can we predict future relations?**

Change-point detection for multi-way relationships

Enron email database [2]

- Employees as senders and receivers
- Events such as words or topics of conversation
- **Are there changes in communication at different times?**

Positron emission tomography (PET) reconstruction [3]

- Poisson distribution commonly used for photon counts in emission tomography
- Statistical model for PET reconstruction is ill-posed without restrictions
- **Can we reformulate and improve the model using low-rank tensors?**

[1] O'Brien, *Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research*, ISR, 2010.
[2] Enron Email Dataset: https://www.cs.cmu.edu/~enron/
[3] Shepp and Vardi, *Maximum Likelihood Reconstruction for Emission Tomography,* IEEE TMI, 1982.

From Linear to Multilinear Regression

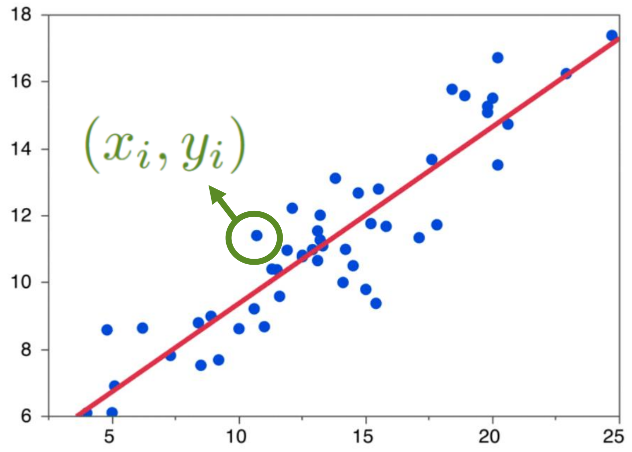
Linear regression: $$y_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i | \boldsymbol{\beta} \rangle, \sigma^2)$$

Tensor regression: $$y_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \sigma^2)$$

Multivariate linear regression: $$\boldsymbol{y}_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i | B \rangle, \Sigma)$$

Tensor-on-Tensor regression [4,5]: $$\mathcal{Y}_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$$

$m_1 \times \cdots \times m_p$ $\quad h_1 \times \cdots \times h_l$ $\quad h_1 \times \cdots \times h_l \times m_1 \times \cdots \times m_p$



Linear regression example

$$\langle S|T \rangle = \begin{cases} \text{inner product} & dim(S) = dim(T) \\ \text{contraction} & dim(S) \neq dim(T) \end{cases}$$

$$\mathcal{Y}_i \in \mathbb{R}^{10\times10\times10} \quad \mathcal{X}_i \in \mathbb{R}^{20\times20\times20}$$
$$\mathcal{B} \in \mathbb{R}^{10\times10\times10\times20\times20\times20}$$

Tensor models not tractable without using a very large sample size

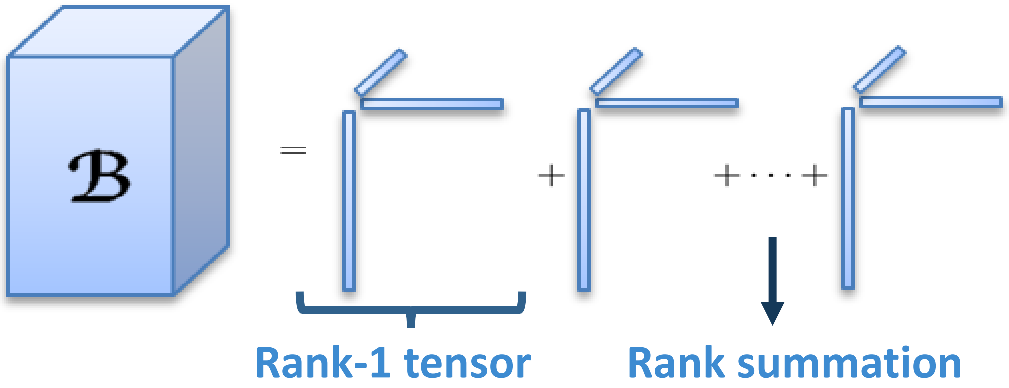
[4] Llosa and Maitra, *Reduced-Rank Tensor-on-Tensor Regression and Tensor-Variate Analysis of Variance*, IEEE TPAMI 2022.
[5] Lock, *Tensor-on-Tensor Regression*, JCGS, 2018

Tensor-on-Tensor Regression (ToTR)

$$\mathcal{Y}_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$$

Canonical polyadic (CP) ToTR [4,5] : $\mathcal{B} = [\![\boldsymbol{\lambda}; U_1, \ldots, U_l, V_1, \ldots, V_p]\!]$

3-dimensional tensor example



Each rank-1 tensor is the outer product of $p+q$ vector

Other low-rank tensor structures studied in [4]

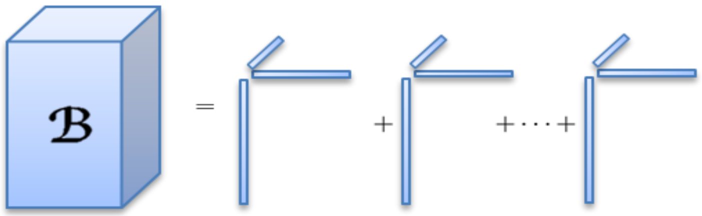
Make the model tractable with low-rank tensors!

[4] Llosa and Maitra, *Reduced-Rank Tensor-on-Tensor Regression and Tensor-Variate Analysis of Variance*, IEEE TPAMI 2022.

[5] Lock, *Tensor-on-Tensor Regression*, JCGS, 2018

Existing: Gaussian ToTR

$$\mathcal{Y}_i \overset{ind.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$$



Loglikelihood:

$$\ell(\mathcal{B}) = \sum_{i=1}^{n} \text{vec}(\mathcal{Y}_i - \langle \mathcal{X}_i | \mathcal{B} \rangle)' \left[\bigotimes_{k=1}^{p} \Sigma_k^{-1} \right] \text{vec}(\mathcal{Y}_i - \langle \mathcal{X}_i | \mathcal{B} \rangle)$$

Least squares loss

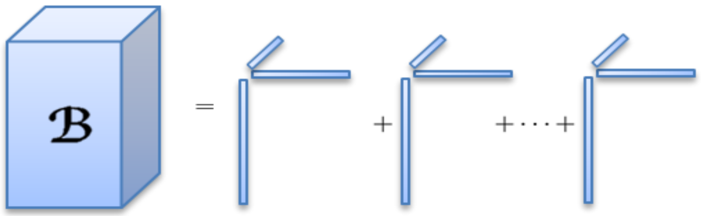
- CP decomposition constraints

New: Poisson ToTR (PToTR)

$$\mathcal{Y}_{i\boldsymbol{j}} \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}})$$

$$\Downarrow$$

$$\mathcal{Y}_i \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i | \mathcal{B} \rangle)$$



Loglikelihood:

$$\ell(\mathcal{B}) = \sum_{i=1}^{n} \sum_{\boldsymbol{j}} [\mathcal{Y}_{i\boldsymbol{j}} \log(\langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}) - \langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}]$$

Kullback-Leibler divergence loss

- CP decomposition constraints
- **Strictly positive $\mathcal{B}$ constraints**

Analogous to the existing model, harder to optimize loss function

Optimization procedure for PToTR

Model

$$\mathcal{Y}_i \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i | \mathcal{B} \rangle)$$

Loglikelihood

$$\ell(\mathcal{B}) = \sum_{i=1}^{n} \sum_{\boldsymbol{j}} [\mathcal{Y}_{i\boldsymbol{j}} \log(\langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}) - \langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}]$$

Constraints

$$\mathcal{B} = [\![\boldsymbol{\lambda}; U_1, \ldots, U_l, V_1, \ldots, V_p]\!] > 0$$

Our new multiplicative update rules extend those in the CP-alternating Poisson regression (CP-APR) algorithm [6]

Multiplicative Updates (strictly non-decreasing loglikelihood)	**Cost** (number of FLOPs)
$\boldsymbol{W} = \mathcal{X}_{(Q+1)}\left(\odot_q V_q\right) \in \mathbb{R}^{n\times R}$	$O\left(R\dim(\mathcal{X})\right)$
$\mathcal{Z} = \mathcal{Y}/\langle \mathcal{X} \vert \mathcal{B} \rangle \in \mathbb{R}^{M_1\times\cdots\times M_P\times n}$	$O\left(R\dim(\mathcal{Y}) + R\dim(\mathcal{X})\right)$
$\boldsymbol{U}^{(p)} \leftarrow \boldsymbol{U}^{(p)} * \dfrac{\mathcal{Z}_{(p)}(\boldsymbol{W} \odot \boldsymbol{U}_{-p})}{\mathbf{1}_{M_p}\mathbf{1}_n^\top \boldsymbol{W}}$	$O\left(R\dim(\mathcal{Y}) + R\dim(\mathcal{X})\right)$
$\boldsymbol{V}^{(q)} \leftarrow \boldsymbol{V}^{(q)} * \dfrac{\mathcal{X}_{(q)}\left[(\mathcal{Z}_{(P+1)}\boldsymbol{U}) \odot \boldsymbol{V}_{-q}\right]}{\mathcal{X}_{(q)}\left[(\mathbf{1}_n\mathbf{1}_R^\top) \odot \boldsymbol{V}_{-q}\right]}$	$O\left(R\dim(\mathcal{Y}) + R\dim(\mathcal{X})\right)$

When $\mathcal{X}$ or $\mathcal{Y}$ are **sparse**, these costs scale with nnz($\mathcal{X}$) ,nnz($\mathcal{Y}$) Instead of dim($\mathcal{X}$) ,dim($\mathcal{Y}$)

Initial Statistical Theory for PToTR

For a class of rank-R coefficient tensors with bounded positive entries in

$$S_R(\alpha,\beta) = \{\mathcal{B} : \mathrm{rank}_{CP}(\mathcal{B}) \leq R,\ \beta < \mathcal{B}_{n,m} \leq \alpha\}$$

We derived a minimax lower bound,

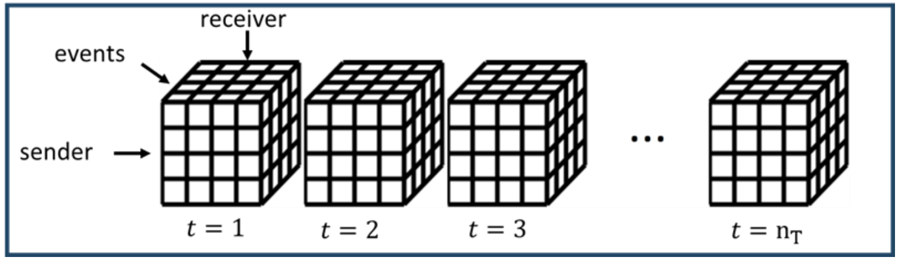
$$\inf_{\hat{\mathcal{B}}} \sup_{\mathcal{B}\in S_R(\beta,\alpha)} \mathbb{E}(||\widehat{\mathcal{B}} - \mathcal{B}||_F^2) \geq \frac{\beta \log 2}{128}\left(\frac{JR}{16} - 1\right)\frac{\xi}{\|\boldsymbol{X}\|_2^2}$$

Where J is the max ambient dimension, $\xi = \min_i ||\mathcal{X}_i||_1$ and $X = [\mathrm{vec}(\mathcal{X}_1)\ldots\mathrm{vec}(\mathcal{X}_n)]$.

- estimation error scales with the **effective low-rank dimension** $\boldsymbol{JR}$, not the ambient dimensions.
- The design matters through the spectral norm $\|\boldsymbol{X}\|_2^2$
- When $I = 1$ this reduces to the PCP minimax case of [1]

[1] Lopez, Prasadan, Llosa-Vite, Lehoucq, Dunlavy, Near-Efficient and Non-Asymptotic Multiway Inference, *Under review.*

Application I: Autoregressive Model for the ICEWS Database



$\mathcal{Y}_t(i_1, i_2, i_3)$ = # times action i_3 was taken by country i_1 on country i_2 at week t.

$$\mathcal{Y}_t \overset{indep.}{\sim} \text{Poisson}(\langle \mathcal{Y}_{t-1} | \mathcal{B} \rangle)$$

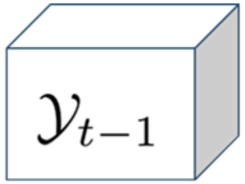
$\mathcal{B}(i_1, i_2, i_3, j_1, j_2, j_3)$ = Effect that the **previous** i_3 event from i_1 towards i_2 has on **current** j_3 event from j_1 towards j_2.

Other longitudinal effects by choosing $\mathcal{X}_t$ in:

$$\mathcal{Y}_t \overset{indep.}{\sim} \text{Poisson}(\langle \mathcal{X}_t | \mathcal{B} \rangle)$$

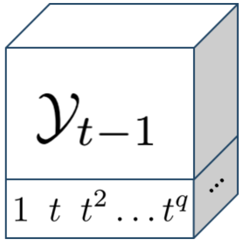
- **AR(1) with no trend (as before):**

$\mathcal{X}_t =$



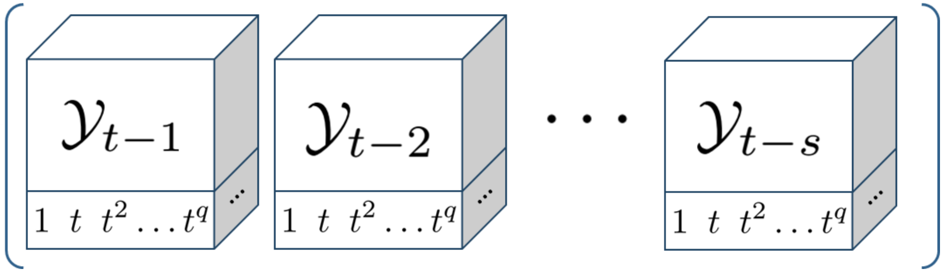
- **AR(1) with q-th order polynomial trend:**

$\mathcal{X}_t =$



- **AR(s) with q-th order polynomial trend:**

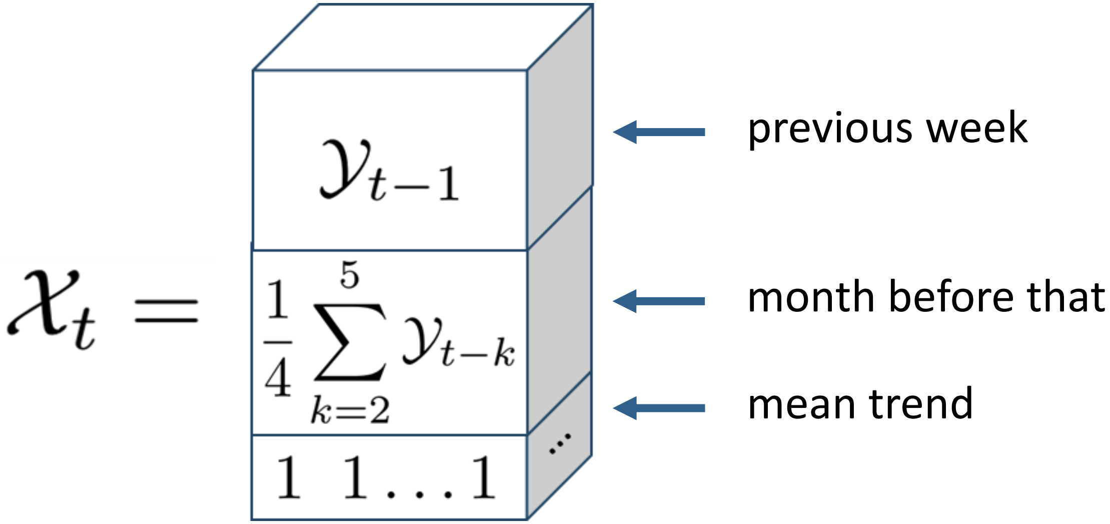
$\mathcal{X}_t =$



Application I: Autoregressive Model for the ICEWS Database

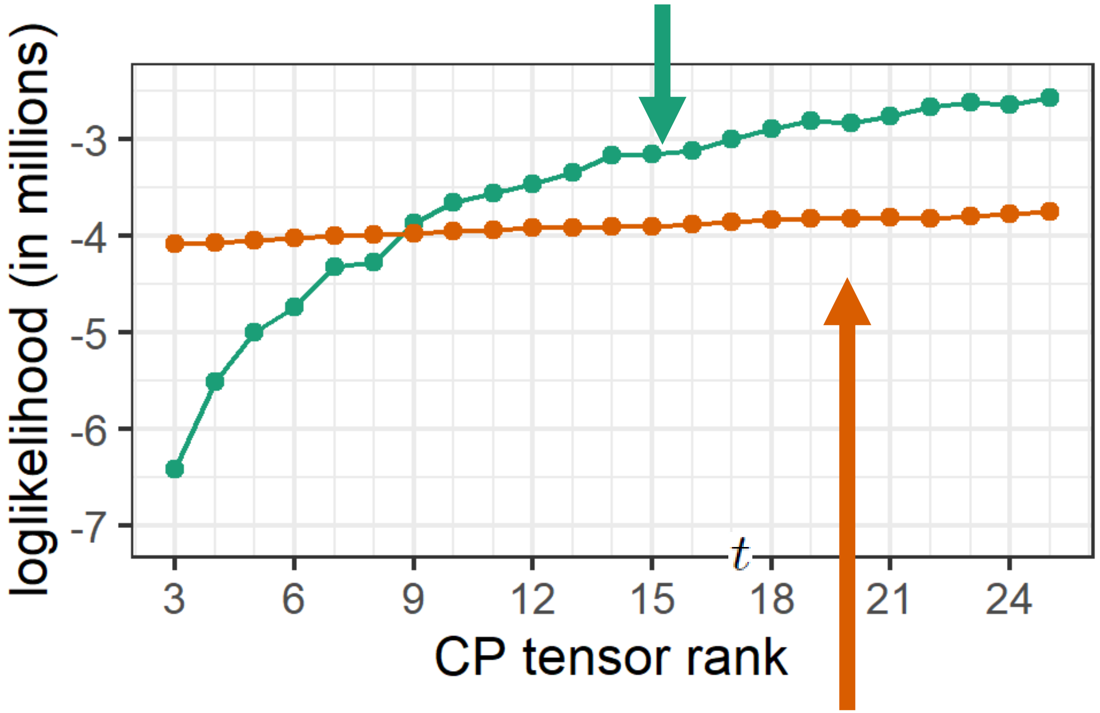
Data selected as in [7]:

- Weekly data from 2004 to mid-2014
- 25 countries
- 4 quad classes (type of events)



Poisson ToTR :

$$\mathcal{Y}_i \overset{indep.}{\sim} \text{Poisson}(\langle \mathcal{X}_i | \mathcal{B} \rangle)$$



Gaussian ToTR:

$$\mathcal{Y}_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$$

Our Poisson model fits the count data better!

[7] Hoff, *Multilinear tensor regression for longitudinal relational data*, Ann. Appl. Stat., 2015

positron emission tomography Image reconstruction

- **Element-wise regression** (ML-EM) [1]:

$$y_{i_1 i_2} \sim \text{Poisson}\left(\langle K_{i_1 i_2} | B \rangle\right)$$

- Ill-posed without constraints [2]

- **NEW: Poisson-response Tensor-on-tensor regression** (PToTR):

$$Y_{i_1 i_2} \sim \text{Poisson}\left(\langle K_{i_1 i_2} | \mathcal{B} \rangle\right)$$

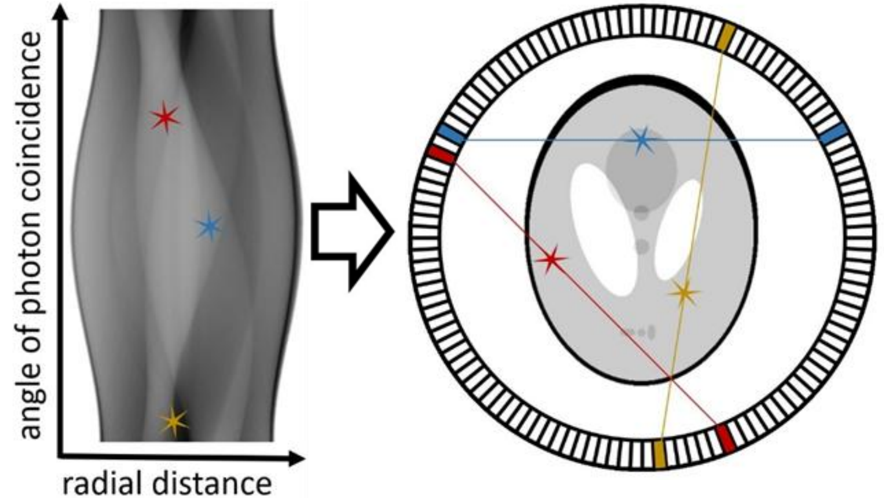
- Recovers full image and is well-posed due to use of low-rank model for $\mathcal{B}$

Radon transform of B

$$\underset{\text{Observed sinogram containing counts}}{Y} \sim \text{Poisson}\big(\overbrace{\langle \underset{\text{Image to reconstruct}}{B} | \mathcal{K} \rangle}^{\text{Radon transform of } B}\big)$$

Observed sinogram containing counts

Image to reconstruct

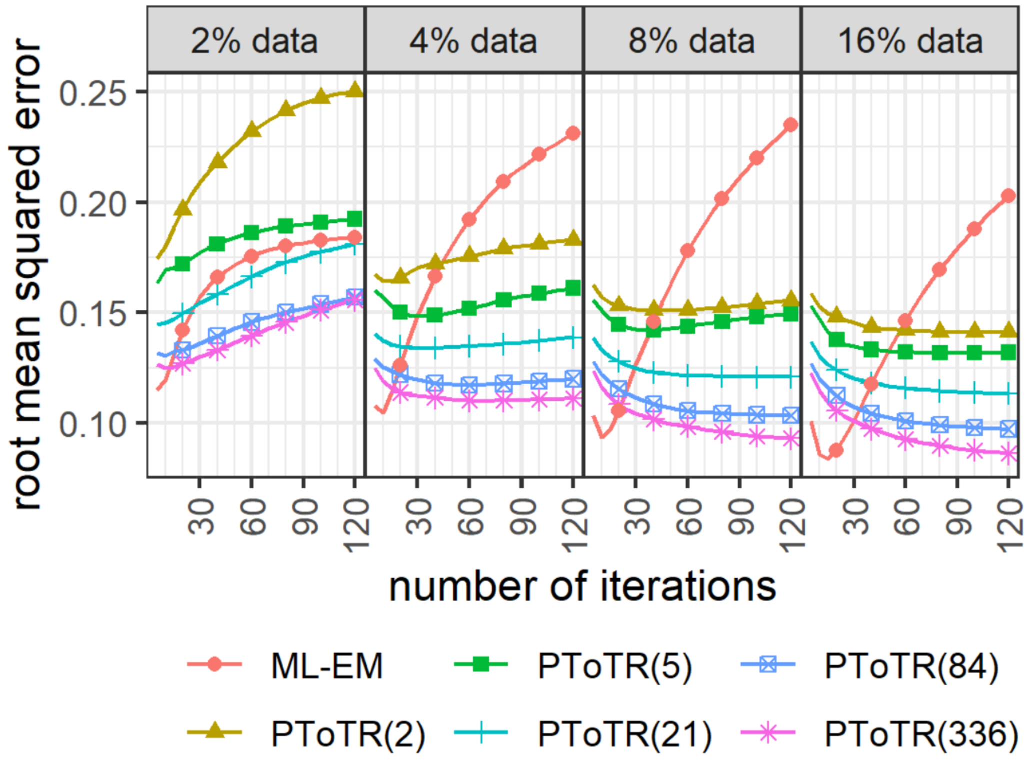


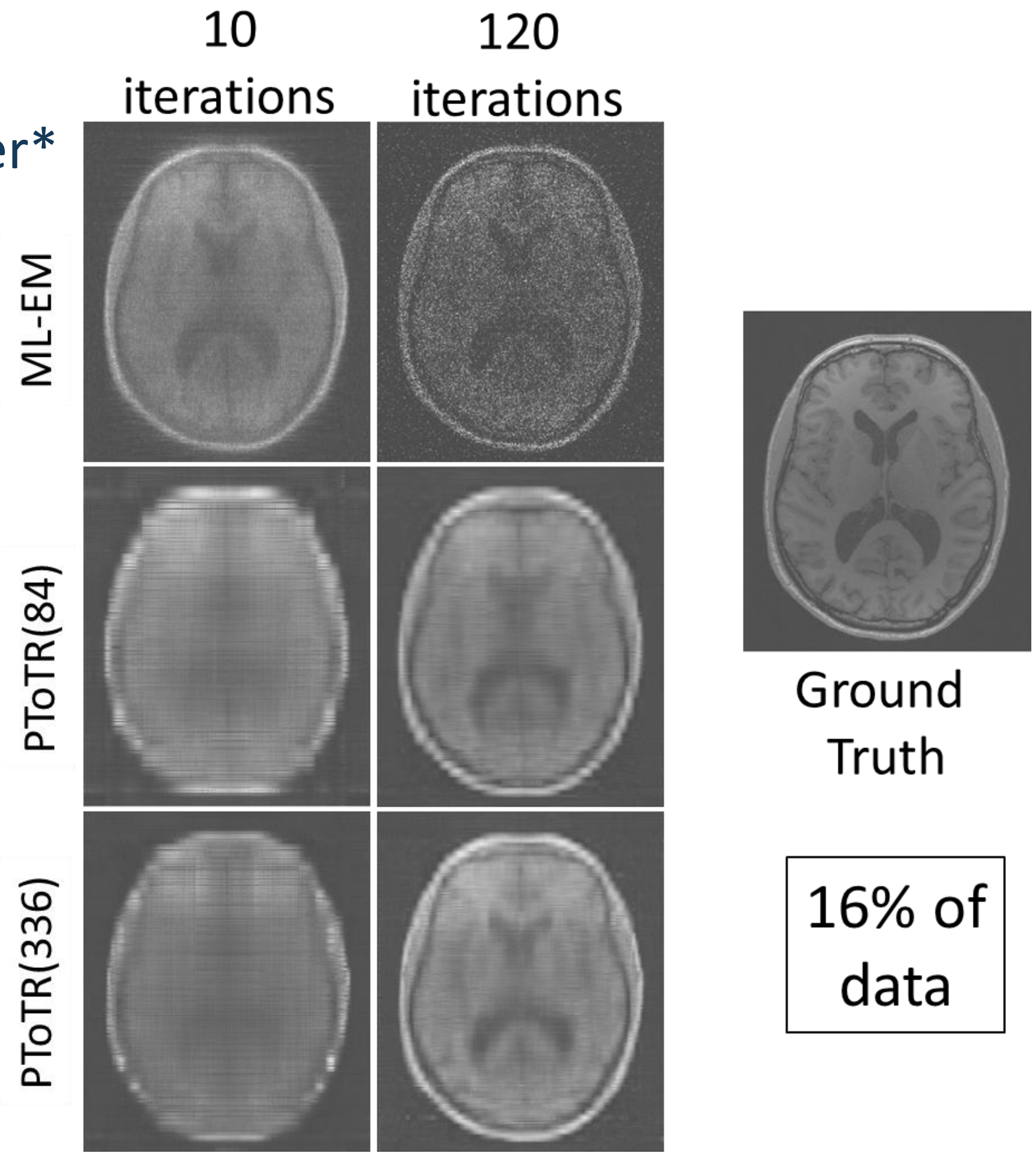
[1] Shepp & Vardi (1982); [2] Snyder, Miller, Thomas & Politte (1987)

PET Image reconstruction

- four MRI measurements on same subject, scanner*
- 256 x 256 matrix slices → 256 x 1024 sinograms
- Parameters: ML-EM (no low-rank): **~ 63 million**

  PToTR: **~ 63 thousand** (rank-84)





*Hawco, et al. (2022)

PToTR and Change-Point Detection

Poisson-tensor change-point detection

$\mathcal{Y}_t$ changes in mean at time τ_1

$$\mathcal{Y}_t \sim \left\{ \begin{array}{ll} \text{Poisson}(\mathcal{M}_1) & t = 1, \ldots, \tau_1 \\ \text{Poisson}(\mathcal{M}_2) & t = \tau_1 + 1, \ldots, n_T \end{array} \right\}$$

- Mean before change-point: $\mathcal{M}_1$
- Mean after change-point: $\mathcal{M}_2$
- Change-point location: τ_1

Equivalent PTANOVA formulation

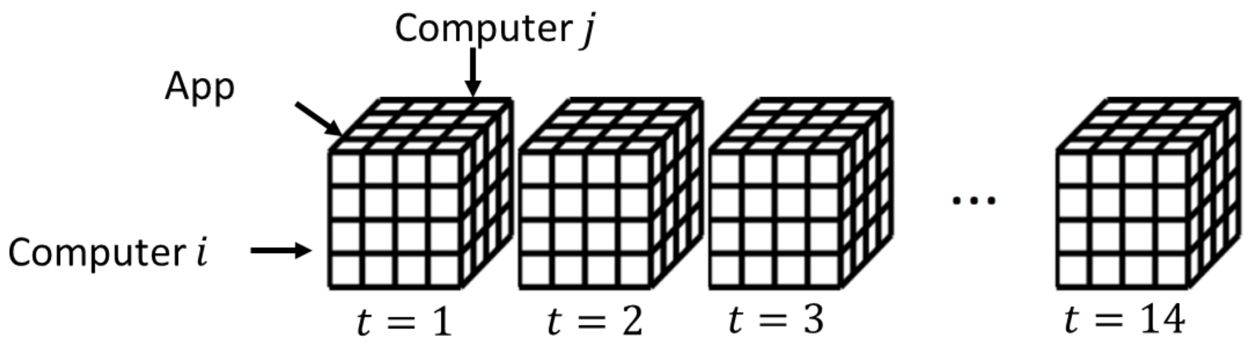
$$\mathcal{Y}_t \sim \text{Poisson}(\langle \boldsymbol{x}_t | \mathcal{B} \rangle)$$

$$\boldsymbol{x}_t = \left\{ \begin{array}{ll} (1,0)' & t = 1, \ldots, \tau_1 \\ (0,1)' & t = \tau_1 + 1, \ldots, n_T \end{array} \right\}$$

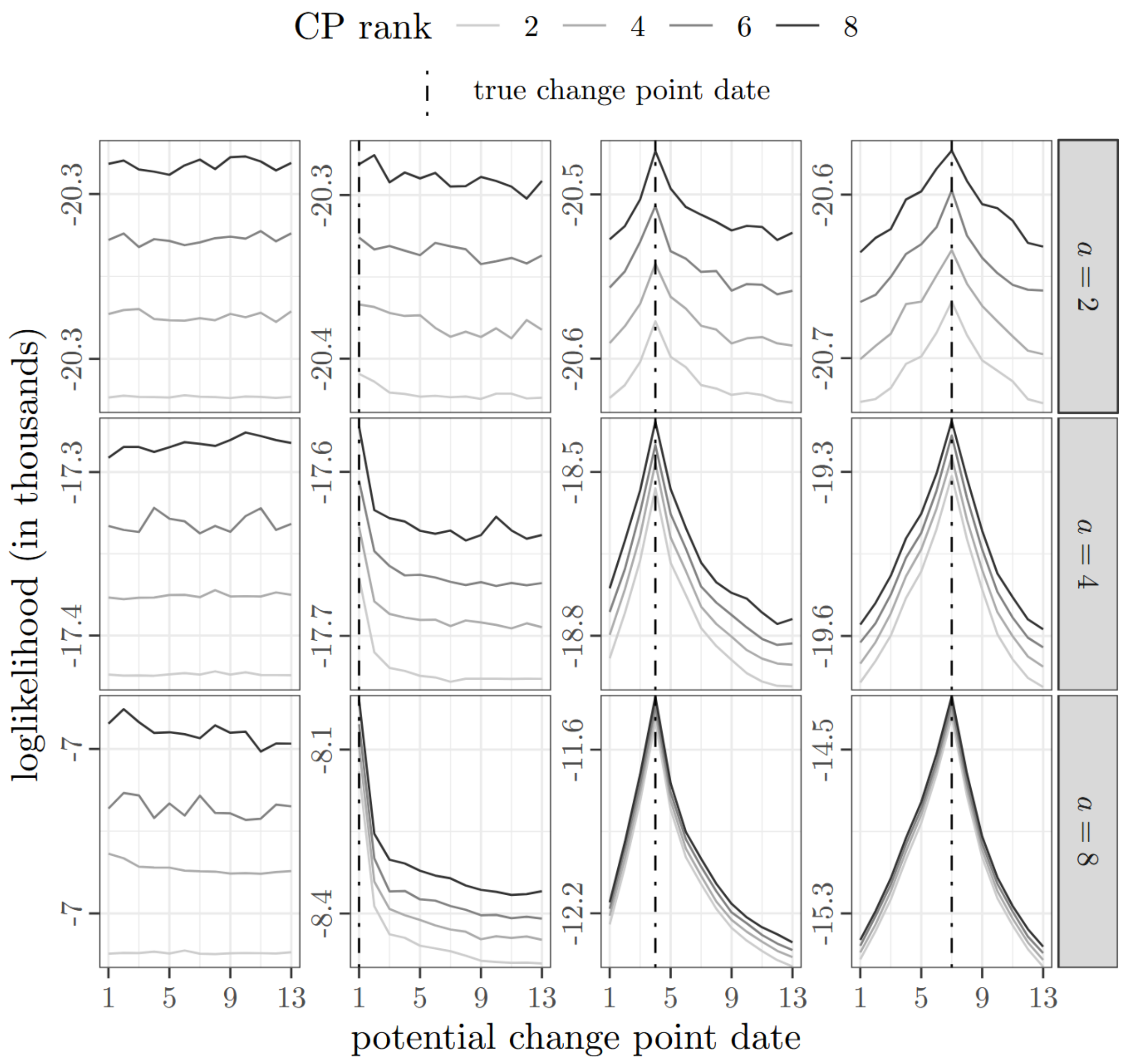
- Mean before change-point: $\langle (1,0)' | \mathcal{B} \rangle$
- Mean after change-point: $\langle (0,1)' | \mathcal{B} \rangle$
- Change-point location: τ_1

Originally solved using PToTR → Now we can perform parameter inference via GToTR

Change-point detection: Simulated Data



- 10 computers communicating via 15 apps over 14 time-steps: 14 count tensors of size 10 x 10 x 15.
- Event occurs at time τ that changes communication pattern:
  - All communications generated independently from Poisson(λ), $\lambda = 5$.
  - One app usage patterns changes to Poisson($\lambda \times a$) after time τ, where $a = 1, 2, 3, 5, 20, 100$.



Conclusions and path forward

- Model: Linear regression → Tensor regression → Poisson tensor regression
  - Low-rank tensors and constrained optimization

- Methods:
  - Tensor autoregressive models: temporal prediction
  - 2D PET → 4D PET: stable estimation
  - ANOVA → TANOVA → PTANOVA: change-point detection for tensor count data

- Where we are going
  - Generalized Tensor-on-Tensor Regression
  - Other low-rank tensor formats (Tucker, TT, others)
  - Uncertainty quantification
  - Error bounds and applications

Thank You!

For follow-up questions reach out to:

- Carlos Llosa cjllosa@sandia.gov,
- Danny Dunlavy dmdunla@sandia.gov