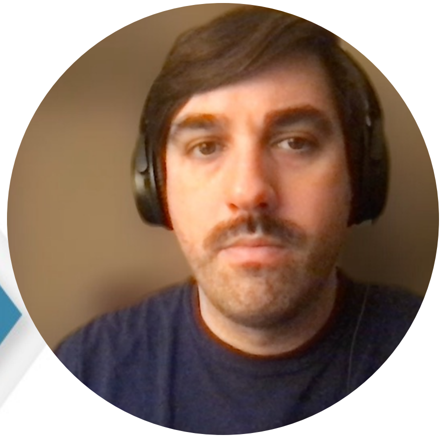
Sandia National Laboratories

Exceptional service in the national interest



THE GENERALIZED MULTILINEAR MODEL (GMLM)

Jeremy Myers

Carlos Llosa, Danny Dunlavy, Rich Lehoucq, Tian Ma

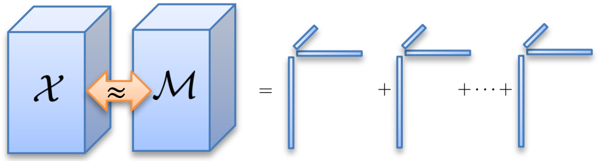
Sandia National Laboratories, USA

9th Annual Sandia Machine Learning/Deep Learning (MLDL) Workshop
Hosted from Albuquerque, NM – Aug. 11-14, 2025

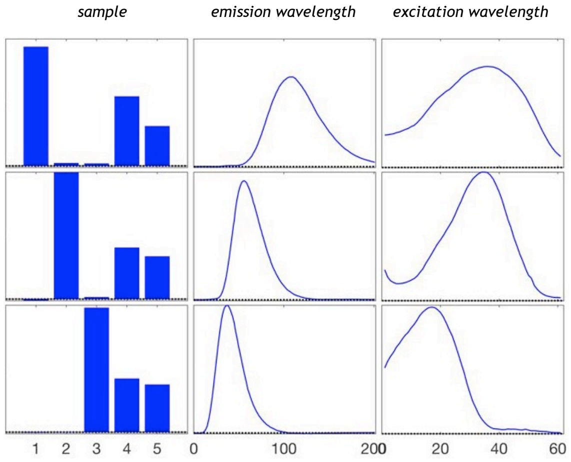
SAND2025-09251V SAND2025-07434C

OVERVIEW

Low-Rank Tensor Decompositions



Example: Amino acid analysis

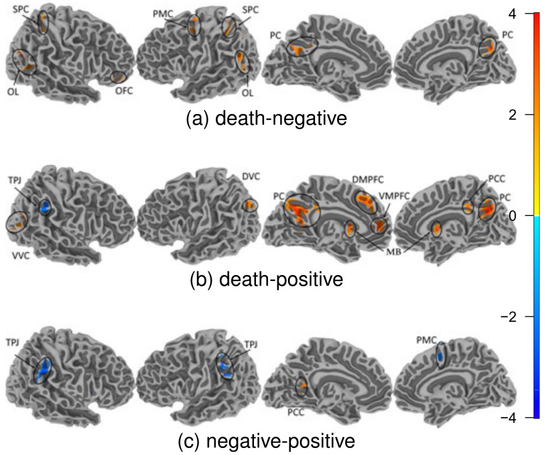


Andersen & Bro (2003)

Tensor-on-Tensor Regression Models

$$\mathcal{Y}_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + \mathcal{E}_i$$

Example: Brain activity/suicide risk

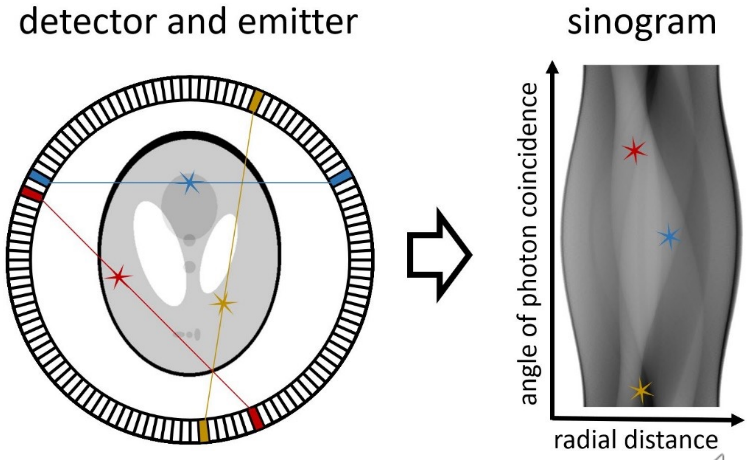


Lock (2018); Llosa & Maitra (2022)

Generalized Multilinear Models (GMLMs)

$$g\left(\mathbb{E}_{f_{\mathcal{Y}}}\left(\mathcal{Y}_i\right)\right) = \langle \mathcal{X}_i | \mathcal{B} \rangle$$

Example: PET image reconstruction

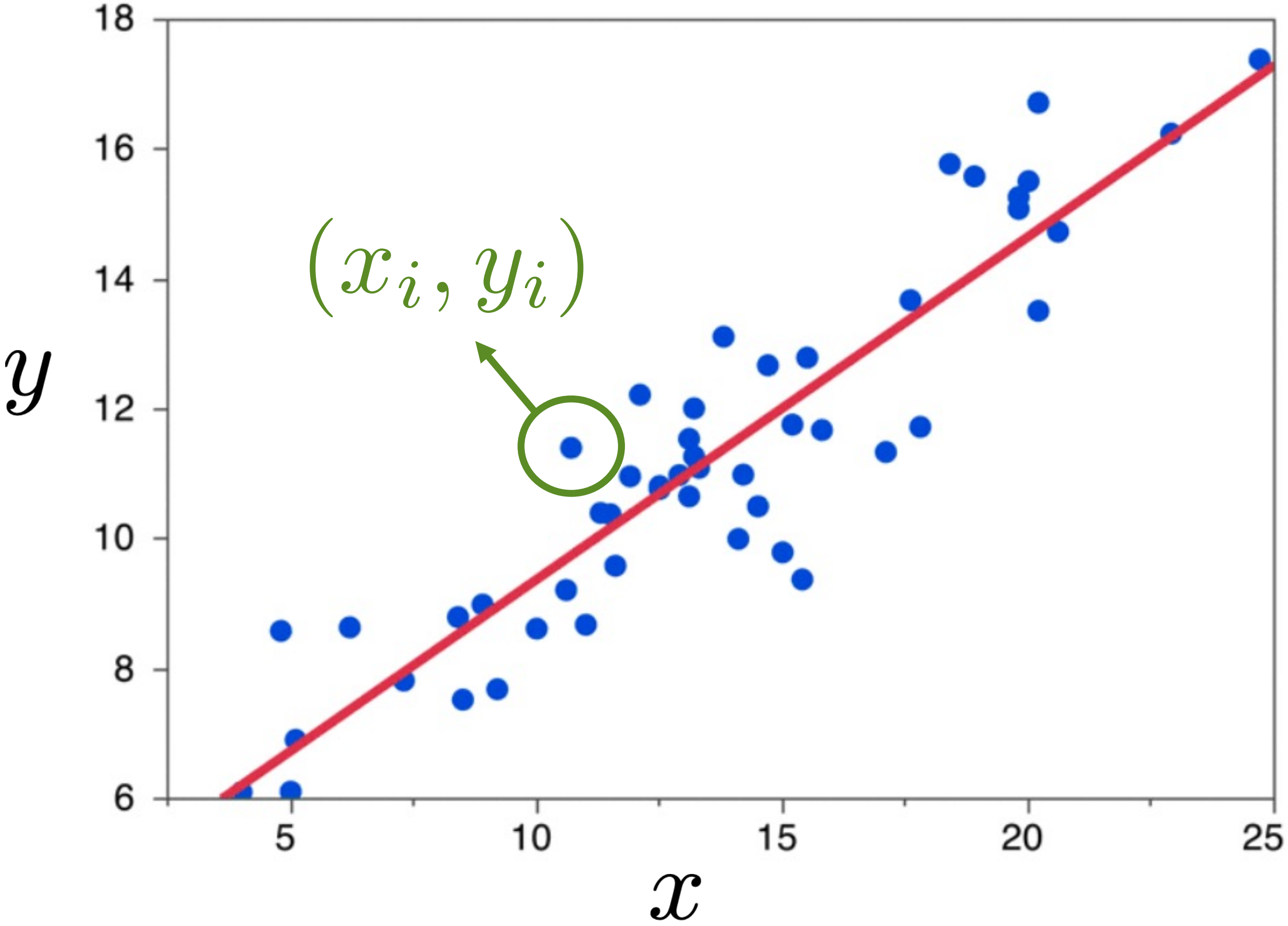


Llosa, Dunlavy, Myers, Lehoucq, Ma (2025)

OUTLINE

- Tensor-on-Tensor Regression (ToTR)
- Evolution of Generalized CP (GCP)
- Evolution of Generalized Multilinear Models (GMLMs) for ToTR
- Extending Generalized Linear Models (GLMs) to GMLMs
- Parameter Inference for GMLMs
- Example Applications

LINEAR REGRESSION



Simple linear regression:

$$y_i = \beta x_i + \epsilon_i$$

Multiple linear regression:

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \epsilon_i$$

TENSOR-ON-TENSOR REGRESSION

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

Linear regression: $$y_i = \langle \boldsymbol{x}_i | \boldsymbol{\beta} \rangle + \epsilon_i$$

Tensor regression: $$y_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + \epsilon_i$$

Multivariate linear regression: $$\boldsymbol{y}_i = \langle \boldsymbol{x}_i | B \rangle + \boldsymbol{\epsilon}_i$$

Tensor-on-Tensor regression (ToTR): $$\mathcal{Y}_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + \mathcal{E}_i$$

Lock (2018); Llosa & Maitra (2022)

TOTR: HOW MUCH DATA DO WE NEED TO FIT THE MODEL

$M_1 \times \cdots \times M_P \qquad N_1 \times \cdots \times N_Q$

Tensor-on-Tensor regression (ToTR):

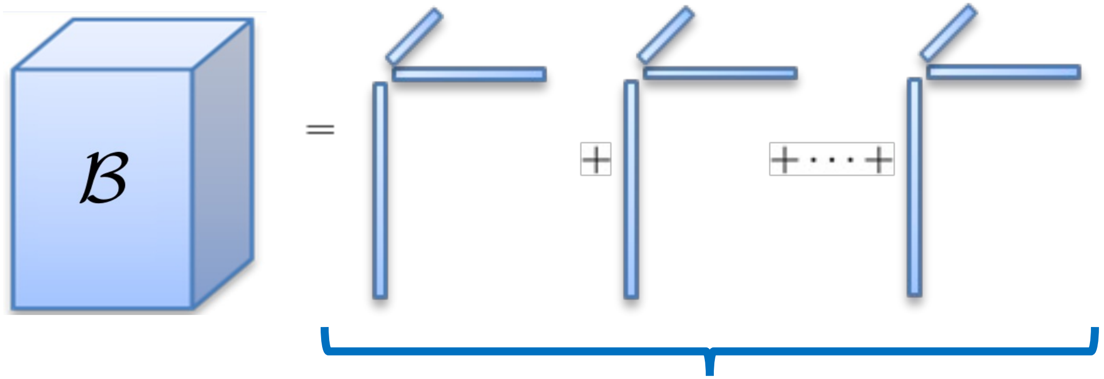
$N_1 \times \cdots \times N_Q \times M_1 \times \cdots \times M_P$

$\boxed{i = 1, 2, \ldots, n \longrightarrow \text{so many samples needed!}}$

$N_1 \times \cdots \times N_Q \times M_1 \times \cdots \times M_P$

Low-rank Canonical Polyadic (CP) Model:

Rank-1 tensor



R summands

Lock (2018); Llosa & Maitra (2022)

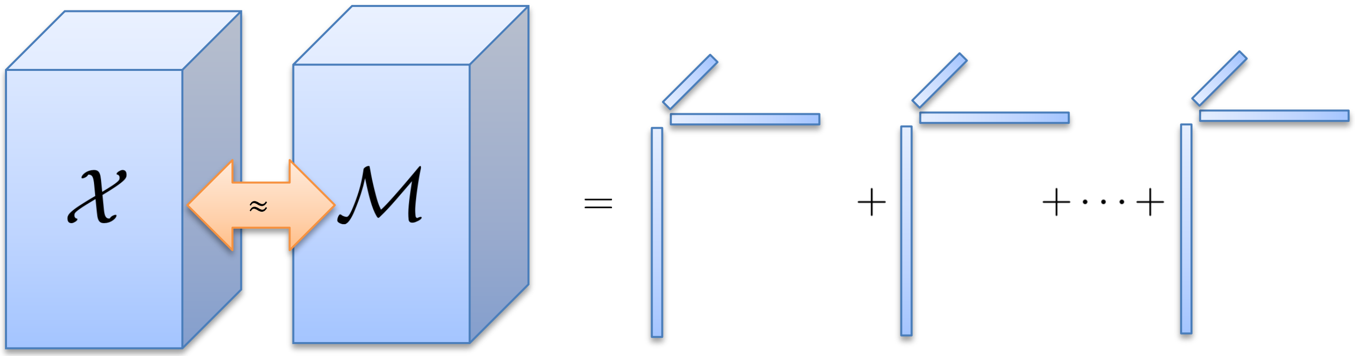
$N_1 \times R \qquad M_1 \times R$

$$\mathcal{B} = [\![V_{N_1}, \ldots, V_{N_Q}, U_{M_1}, \ldots, U_{M_P}]\!]$$

Fewer parameters $\longrightarrow$ fewer samples needed

$$\prod_p M_p \prod_q N_q \to R\left(\sum_p M_p + \sum_q N_q\right)$$

GENERALIZING CP DECOMPOSITIONS FOR D-WAY TENSORS



$$= [\![A^{(1)}, A^{(2)}, \ldots, A^{(d)}]\!] = [\![\{A^{(k)}\}_1^d]\!]$$

Multi-index for *d*-way tensors: $\boldsymbol{j} = (j_1, j_2, \ldots, j_d) \to x_{\boldsymbol{j}} = \mathcal{X}(j_1, j_2, \ldots, j_d)$

CP-ALS, CP-OPT

Carroll & Chang (1970); Harshman (1970); Acar, Dunlavy & Kolda (2011)

$$\min_{\{A^{(k)}\}_1^d} \sum_{\boldsymbol{j}} (x_{\boldsymbol{j}} - m_{\boldsymbol{j}})^2$$

$$x_{\boldsymbol{j}} \overset{indep.}{\sim} N(m_{\boldsymbol{j}}, \sigma^2)$$

CP-APR, CP-POPT

Chi & Kolda (2012); Ranadive & Baskaran (2021)

$$\min_{\{A^{(k)}\}_1^d} \sum_{\boldsymbol{j}} m_{\boldsymbol{j}} - x_{\boldsymbol{j}} \log(m_{\boldsymbol{j}})$$

$$x_{\boldsymbol{j}} \overset{indep.}{\sim} Poisson(m_i)$$

GCP

Hong, Kolda & Duersch (2020); Kolda & Hong (2020)

$$\min_{\{A^{(k)}\}_1^d} \sum_{\boldsymbol{j}} f(x_{\boldsymbol{j}}, m_{\boldsymbol{j}})$$

$$x_{\boldsymbol{j}} \overset{indep.}{\sim} P_f(m_i)$$

GENERALIZING TENSOR-ON-TENSOR REGRESSION

Tensor-on-Tensor regression (ToTR):

$$\mathcal{Y}_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + \mathcal{E}_i$$

$M_1 \times \cdots \times M_P$

$N_1 \times \cdots \times N_Q$

$N_1 \times \cdots \times N_Q \times M_1 \times \cdots \times M_P$

ToTR

Lock (2018); Llosa & Maitra (2022)

$$\mathcal{Y}_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + \mathcal{E}_i$$

$$\mathcal{Y}_i \sim N\left(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_P\right)$$

CP, Tucker, Tensor Train, Tensor Network, …

Poisson ToTR (PToTR)

Llosa & Dunlavy (2025)

$$\mathcal{Y}_i \sim Poisson\left(\langle \mathcal{X}_i | \mathcal{B} \rangle\right)$$

CP

GMLM

Llosa, Dunlavy, Myers, Lehoucq, Ma (2025)

$$g\left(\mathbb{E}_{f_\mathcal{Y}}\left(\mathcal{Y}_i\right)\right) = \langle \mathcal{X}_i | \mathcal{B} \rangle$$

CP

THE GENERALIZED MULTILINEAR MODEL (GMLM)

Same as in ToTR

$$g\left(\mathbb{E}_{f_{\mathcal{Y}}}\left(\mathcal{Y}_i\right)\right) = \langle \mathcal{X}_i | \mathcal{B} \rangle$$

Link Function

- $g : \mathbb{R}^{\times_{p=1}^{P} M_p} \to \mathbb{R}^{\times_{p=1}^{P} M_p}$
- Applied entry-wise
- e.g., identity, logit, log, inverse,…

Random component

- e.g., normal, binomial, Poisson, …

GENERALIZED LINEAR MODELS (GLMS)

$$y_i \sim P_{f_Y}(\psi_i), \quad \mathbb{E}(y_i) = \mu_i, \quad \mu_i = g^{-1}(\eta_i), \quad \eta_i = \boldsymbol{b}^\top \boldsymbol{x}_i$$

- y: response (independent)
- $\boldsymbol{x}$: covariate
- μ: mean response
- g: link function, relating random component and linear predictor
- $\boldsymbol{b}$: model parameters
- n samples $\rightarrow X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]\,; \boldsymbol{y} = [y_1, \ldots, y_n]$

$$\boxed{\mathrm{GLM}(X, \boldsymbol{y})}$$

Nelder & Wedderburn (1972)

GENERALIZED LINEAR MODELS (GLMS)

$$y_i \sim P_{f_Y}(\psi_i), \quad \mathbb{E}(y_i) = \mu_i, \quad \mu_i = g^{-1}(\eta_i), \quad \eta_i = \boldsymbol{b}^\top \boldsymbol{x}_i$$

Loglikelihood:

$$\ell(\boldsymbol{b}) = \sum_{i=1}^{n} \log f_Y(\psi_i)$$

Gradient:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{b}} \ell(\boldsymbol{b}) &= \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{b}} \log f_Y(\psi_i) \\ &= \sum_{i=1}^{n} \underbrace{\left[\frac{\partial}{\partial \boldsymbol{b}} \eta_i\right]}_{\boldsymbol{x}_i} \underbrace{\left[\frac{\partial}{\partial \eta_i} \mu_i\right] \left[\frac{\partial}{\partial \mu_i} \psi_i\right] \left[\frac{\partial}{\partial \psi_i} \log f_Y(\psi_i)\right]}_{\text{weight } i} \end{aligned}$$

Nelder & Wedderburn (1972)

THE GENERALIZED MULTILINEAR MODEL (GMLM)

Same as in ToTR

$$g\left(\mathbb{E}_{f_{\mathcal{Y}}}\left(\mathcal{Y}_i\right)\right) = \langle \mathcal{X}_i | \mathcal{B} \rangle$$

Link Function

- $g : \mathbb{R}^{\times_{p=1}^{P} M_p} \to \mathbb{R}^{\times_{p=1}^{P} M_p}$
- Applied entry-wise
- e.g., identity, logit, log, inverse,…

Random component

- Exponential dispersion family:

$$\log f_{\mathcal{Y}}(\psi) = \sum_{\boldsymbol{j}} \left(\frac{\mathcal{Y}_{\boldsymbol{j}}\psi_{\boldsymbol{j}} - b(\psi_{\boldsymbol{j}})}{a(\delta_{\boldsymbol{j}})} + C \right)$$

- e.g., normal, binomial, Poisson, …

Extends the Generalized Linear Model (GLM) to tensors

Nelder & Wedderburn (1972)

FROM GLM TO GMLM

$$\mathcal{Y}_i \sim P_{f_{\mathcal{Y}}}(\psi_i), \quad \mathbb{E}(\mathcal{Y}_i) = \mu_i, \quad \mu_i = g^{-1}(\eta_i), \quad \eta_i = \langle \mathcal{X}_i | \mathcal{B} \rangle$$

where $\mathcal{Y}_i, \psi_i, \mu_i, \eta_i$ are all tensors of size $M_1 \times \cdots \times M_P$

Low-rank constraint: $\mathcal{B} = [\![V_1, \ldots, V_{N_Q}, U_1, \ldots, U_{M_P}]\!]$

Loglikelihood: $\ell(\boldsymbol{\theta}) = \sum_{i,\boldsymbol{j}} \log f_{\mathcal{Y}}(\psi_{i,\boldsymbol{j}})$

Gradient: $\frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta}) = \left[\left(\frac{\partial}{\partial \operatorname{vec}(V_1)} \ell(\boldsymbol{\theta}) \right)^\top \ldots \left(\frac{\partial}{\partial \operatorname{vec}(U_P)} \ell(\boldsymbol{\theta}) \right)^\top \right]^\top$

where $\boldsymbol{\theta} = [\ \operatorname{vec}(V_1)^\top \ldots\ \operatorname{vec}(V_Q)^\top\ \operatorname{vec}(U_1)^\top \ldots\ \operatorname{vec}(U_P)^\top]^\top$

GMLM: MAXIMUM LIKELIHOOD ESTIMATION

Stack all $\mathcal{Y}_i, \mathcal{X}_i$ into $\mathcal{Y} \in \mathbb{R}^{(\times_{p=1}^{P} M_p)\times n}, \mathcal{X} \in \mathbb{R}^{(\times_{q=1}^{Q} N_q)\times n}$. Use $\mathrm{GLM}(X, \boldsymbol{y})$ as follows:

- **Row-based inference** for U_p, where $W = \mathcal{X}_{(Q+1)}(\odot_q V_q)$:

Khatri-Rao Product

$$U_p[j_p, :] \leftarrow \mathrm{GLM}\left(X = \mathrm{reshape}\left[\left(\bigodot_{k\neq p} U_k\right) \odot W\right],\ \boldsymbol{y} = (Y_{(p)}[j_p, :])^\top\right)$$

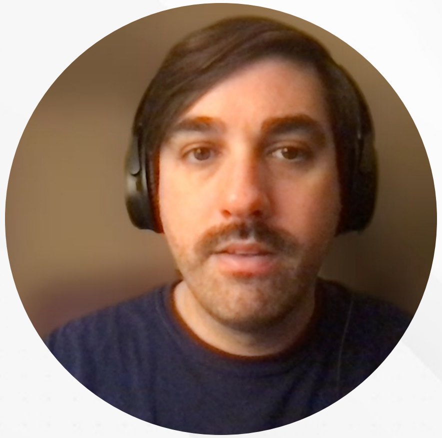
- **Factor-based inference** for V_q, where $W_q = \mathcal{X}_{(Q+1,q)}(\odot_{k\neq q} V_k)$:

$$\mathrm{vec}(V_q) \leftarrow \mathrm{GLM}\left(X = \mathrm{reshape}\left[\left(\bigodot_{p} U_p\right) \odot W_q\right],\ \boldsymbol{y} = \mathrm{vec}(\mathcal{Y})\right)$$

1 outer iteration

NOTE: Leverage iteratively reweighted least squares (IRLS) to solve each GLM.

EXAMPLE APPLICATIONS



PARAMETER INFERENCE VALIDATION: SIMULATED DATA

$$g\left(\mathbb{E}_{f_{\mathcal{Y}}}\left(\mathcal{Y}_i\right)\right) = \langle \mathcal{X}_i | \mathcal{B} \rangle$$

Gaussian distribution, identity link:

- 4 different camelid images
- 150 noisy samples of each

- $\mathcal{Y}_i$: 87 x 106 (*pixel height* x *width*)
- $\mathcal{X}_i$: 4 x 3 (*camelid* x *RGB channel*)
- $\mathcal{B}$: 4 x 3 x 87 x 106

Can recover original images from noisy samples using GMLM with sufficiently high rank of parameter tensor.

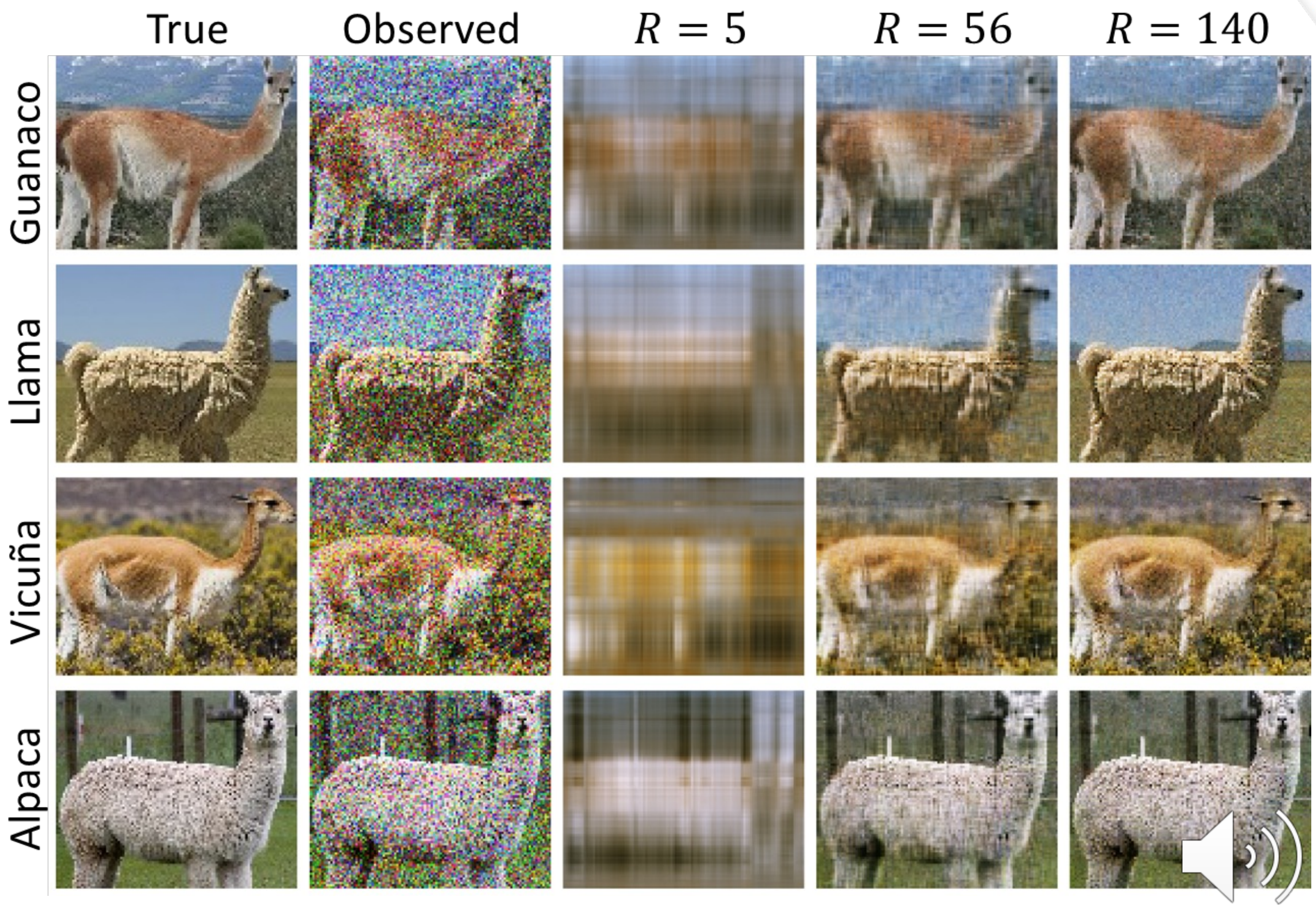


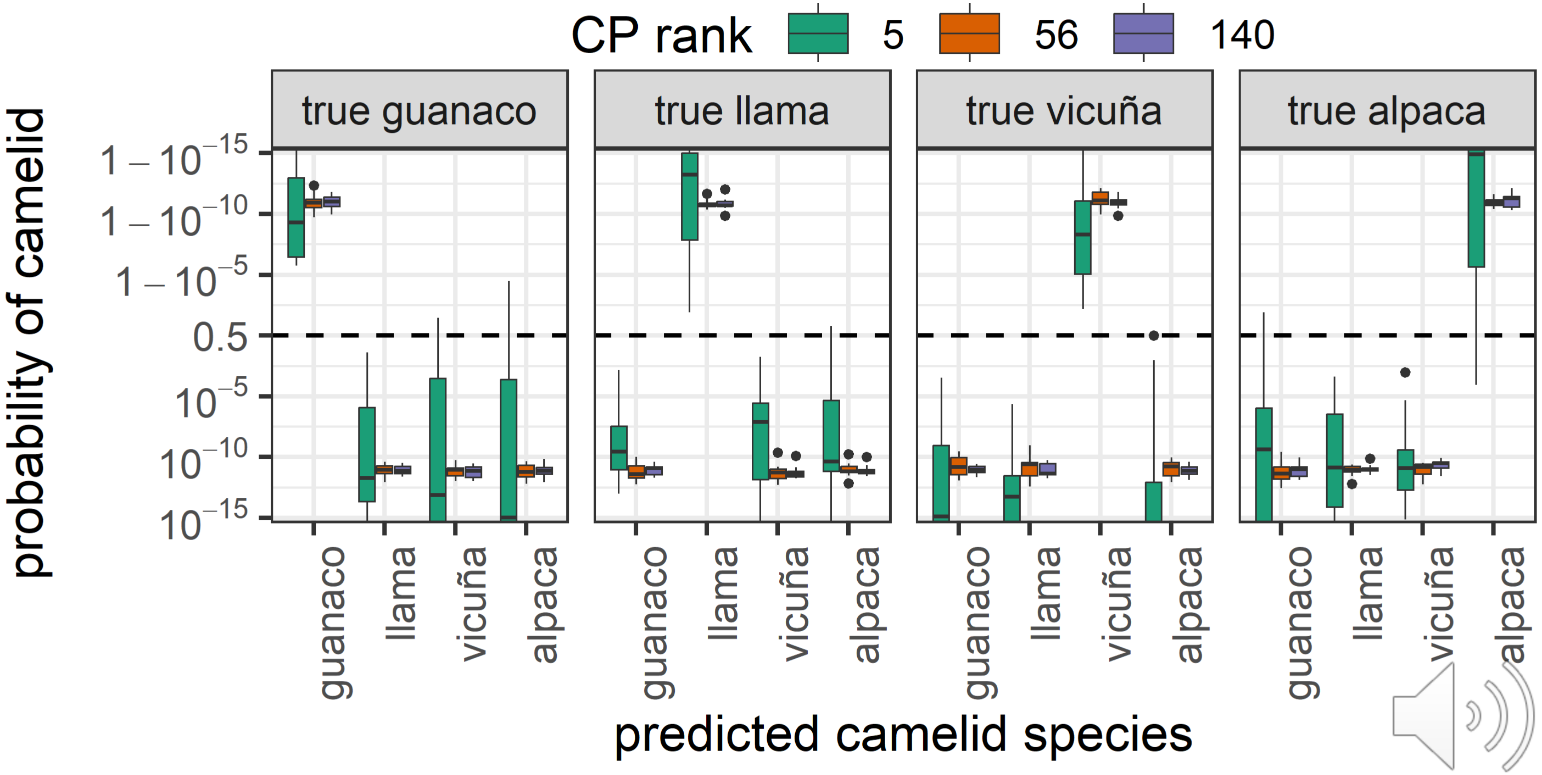
IMAGE CLASSIFICATION (PREDICTION): SIMULATED DATA

$$g\left(\mathbb{E}_{f_{\mathcal{X}}}\left(\mathcal{X}_i\right)\right) = \langle \mathcal{Y}_i | \mathcal{B} \rangle$$

Bernoulli distribution, logit link:

- 4 different camelid images
- 150 noisy samples of each

- $\mathcal{Y}_i$: 87 x 106 (*pixel height* x *width*)
- $\mathcal{X}_i$: 4 x 1 (*camelid* x *RGB channel*)
- $\mathcal{B}$: 87 x 106 x 4 x 1



Can accurately predict image classifications from noisy samples using GMLM with sufficiently high rank of parameter tensor.

POSITRON EMISSION TOMOGRAPHY RECONSTRUCTION

Poisson distribution, identity link:

- Poisson distribution for number of photon coincidences in sinogram
- Radon basis $\mathcal{K}$ is a 4D tensor with combined dimensions of Y and B
- Element-wise formulation (ML-EM) [1]:

$$y_{i_1 i_2} \overset{ind.}{\sim} \text{Poisson}(\langle B, K_{i_1 i_2} \rangle)$$

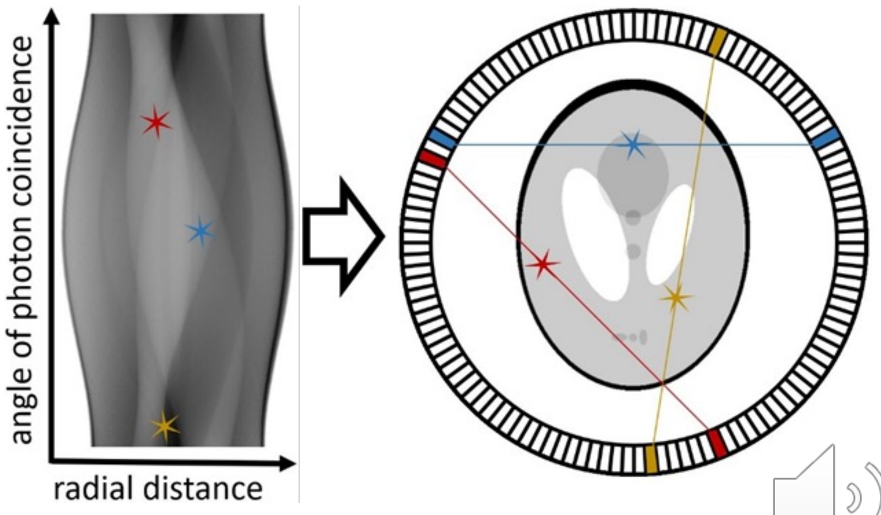
- Ill-posed without additional restrictions [2]
- Here $K_{i_1 i_2}$ is one matrix slice of $\mathcal{K}$
- **New: 4D PET model (volume across time):**

$$Y_{i_1, i_2} \overset{ind.}{\sim} \text{Poisson}(\langle K_{i_1, i_2} | \mathcal{B} \rangle)$$

$$Y \sim \text{Poisson}(\langle \overbrace{B|\mathcal{K}}^{\text{Radon transform of } B} \rangle)$$

Observed sinogram containing counts

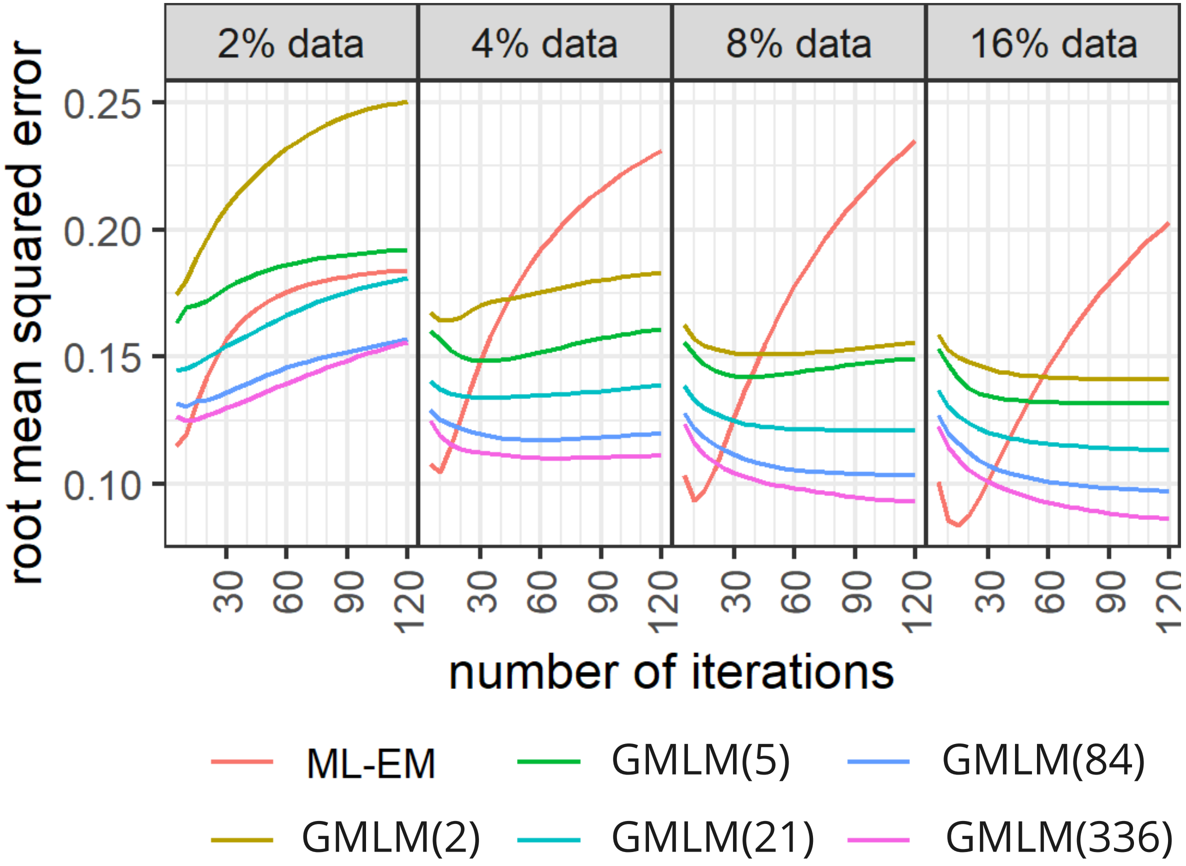
Image to reconstruct

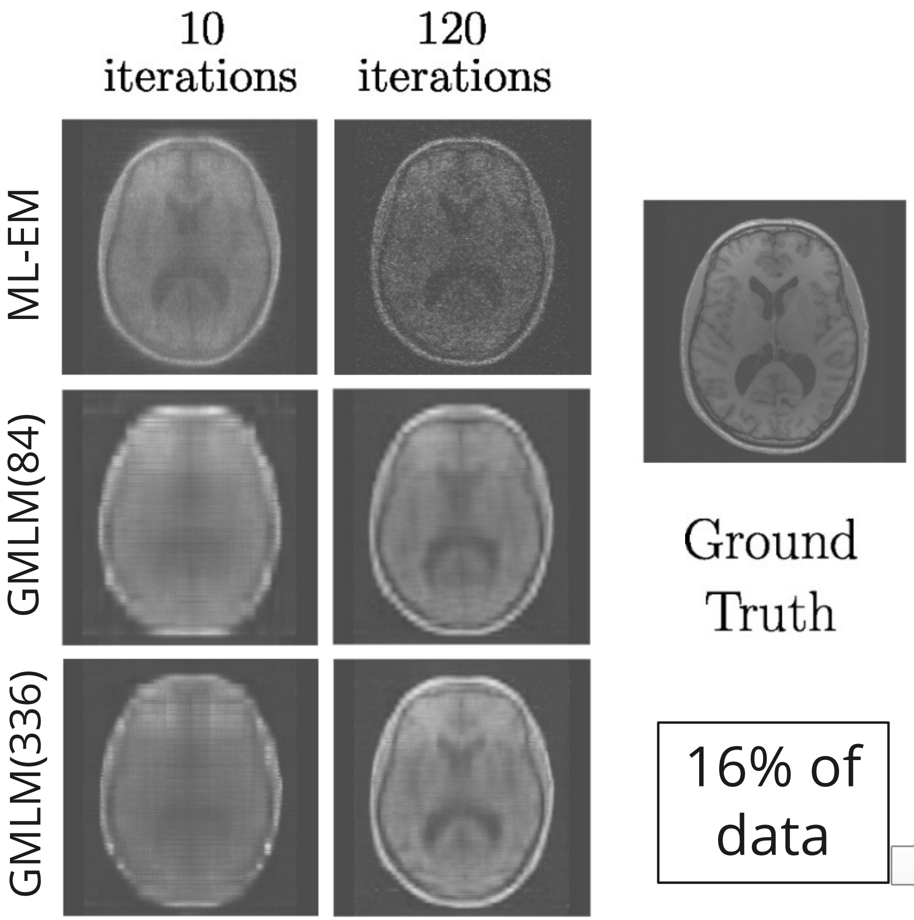


[1] Shepp & Vardi (1982); [2] Snyder, et al. (1987)

POSITRON EMISSION TOMOGRAPHY RECONSTRUCTION

- 4-way tensor data: four MRI measurements on the same subject and scanner *Hawco, et al. (2022)*
- 256 x 256 matrix slices → 256 x 1024 sinograms
- Parameters: ML-EM (no low-rank): ~ 63 million; GMLM: ~ 63 thousand (rank-84)





16% of data

TENSOR-VARIATE ANALYSIS OF VARIANCE (TANOVA)

TANOVA is a special case of ToTR ($\mathcal{X}_i$ are indicator tensors) that generalizes ANOVA

General Model	Definition	Special Case (Indicator $\mathcal{X}_i$)
Linear regression (LR)	$y_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i \vert \boldsymbol{\beta} \rangle, \sigma^2)$	**ANOVA**
Multivariate LR	$\boldsymbol{y}_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i \vert B \rangle, \Sigma)$	**MANOVA**
Tensor regression	$y_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i \vert \mathcal{B} \rangle, \sigma^2)$	**Factorial designs**
Tensor-on-tensor regression	$\mathcal{Y}_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i \vert \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$	**TANOVA**

$\mathcal{Y}_i \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i \vert \mathcal{B} \rangle)$ → **Poisson TANOVA**

Llosa & Maitra (2022)

PTANOVA 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

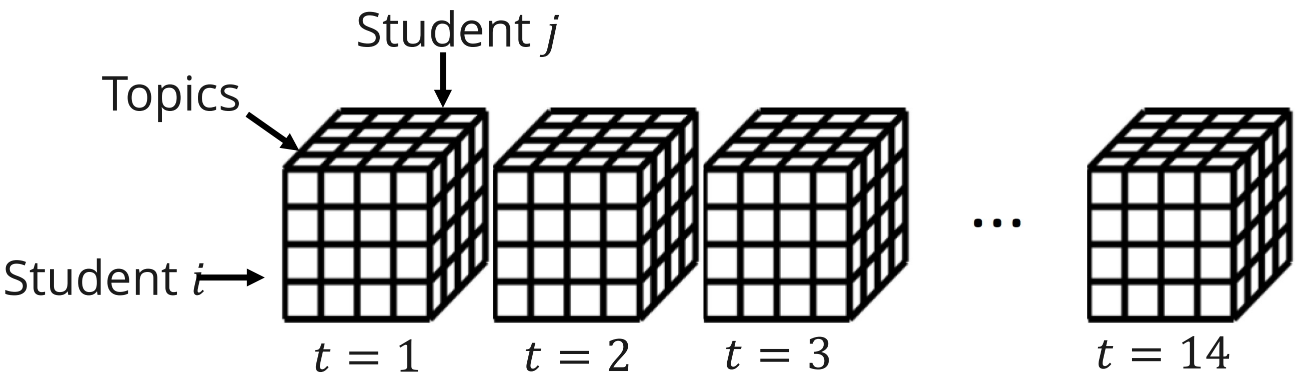
$$\mathcal{Y}_t \sim \text{Poisson}(\langle \boldsymbol{x}_t | \mathcal{B} \rangle)$$

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

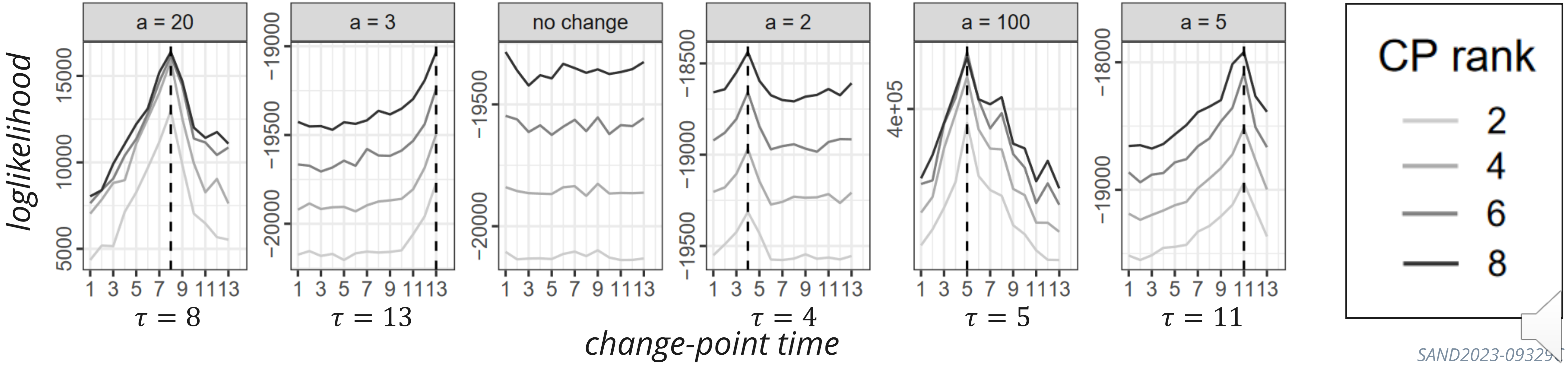
- Mean before change-point: $\langle (1,0)' | \mathcal{B} \rangle$
- Mean after change-point: $\langle (0,1)' | \mathcal{B} \rangle$
- Change-point location: τ_1

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

CHANGE-POINT DETECTION: SIMULATED DATA



- 10 Students talking about 15 topics over 14 time-steps: 14 count tensors of size 10 x 10 x 15.
- Event occurs at time τ that changes communication pattern:
  - All conversation generated independently from Poisson(λ), $\lambda = 5$.
  - One topic changes to Poisson($\lambda \times a$) after time τ, where $a = 1, 2, 3, 5, 20, 100$.



SUMMARY AND NEXT STEPS

- **GMLM: Generalized Multilinear Model for Tensor-on-Tensor Regression**
  - *Inference efficiency:* leverages low-rank decomposition of parameter tensor
  - *Well-posed model:* low-rank parameter tensor avoids need for regularization
  - *Reasonable computation:* leverages IRLS for alternating parameter inference subproblems

- **Future Directions**
  - *GMLM vs. GCP:* GMLM reduces to tensor decomposition with $\mathcal{X}_i = 1$
  - *Analysis of variance:* GMLM is equivalent to ANOVA with indicator $\mathcal{X}_i$
  - *Accelerating computation:* randomized IRLS
  - *Uncertainty quantification:* Fisher Information and Cramér-Rao lower bounds on variance
  - *Applications:* temporal prediction, change-point detection, imaging reconstruction

GMLM: GENERALIZED MULTILINEAR MODEL

Jeremy Myers
Jeremy.moulton.myers@gmail.com