Sandia National Laboratories

Exceptional service in the national interest

William & Mary Graduate Arts & Sciences

Cyclic GCP-CPAPR Hybrid

Feb. 26, 2022

Jeremy M. Myers[1, 2]

Daniel M. Dunlavy[1]

[1] Sandia National Laboratories, Albuquerque, NM & Livermore, CA

[2] Department of Computer Science, College of William and Mary, Williamsburg, VA

SIAM Conference on Parallel Processing for Scientific Computing (PP22)

MS71 Parallel Algorithms for Tensor Computations and their Applications

SAND2022-1347C

Low-Rank Canonical Polyadic (CP) Poisson Tensor Decomposition

William & Mary
Graduate Arts & Sciences

Task

- Fit low-rank CP tensor model to Poisson-distributed nonnegative integer data.
- Nonlinear, non-convex optimization problem
- Approach: Use local method for maximum likelihood estimation from many initial guesses ("*multi-start*").
  - Current local methods converge to maximum likelihood estimator (MLE) only a fraction of solves.
  - Previous work: Examine trade-offs between several state-of-the-art local methods.[†]

Our Contributions

- Leverage trade-offs between multiple methods CP Poisson tensor decomposition in a hybrid fashion.
- Preliminary result: hybrid approach can minimize approximation error & reduce computational cost on synthetic data.

[†] Jeremy M. Myers and Daniel M. Dunlavy. *Using computation effectively for scalable Poisson tensor factorization: Comparing methods beyond computational efficiency.* In 2021 IEEE High Performance Extreme Computing Conference, HPEC 2021, Waltham, MA, USA, September 20-24, 2021, pages 1–7. IEEE, 2021.

Low-Rank CP Poisson Tensor Decomposition

- Let $\mathcal{X}$ be a d-way tensor of size $n_1 \times \cdots \times n_d$ of Poisson-distributed non-negative integers.
- A low-rank CP Poisson tensor decomposition can be computed by estimating the parameters $\mathcal{M}_{\mathbf{i}}$ that minimizes the negative log-likelihood function (NLL):

$$\min_{\mathcal{M}} f(\mathcal{X}, \mathcal{M}) = \sum_{\mathbf{i}} \mathcal{M}_{\mathbf{i}} - \mathcal{X}_{\mathbf{i}} \log(\mathcal{M}_{\mathbf{i}}),$$

  where $\mathbf{i}$ is a tuple over the tensor entries (multi-index), $\mathcal{M}$ is a rank-R CP tensor model, and $A_k, k \in \{1, \ldots, d\}$ defined as:

$$\mathcal{M} = \sum_{r=1}^{R} \lambda_j A_1(:, r) \circ \cdots \circ A_d(:, r).$$

- The maximum likelihood estimator, $\widehat{M}^*$, estimates the global optimizer.

- Applications
  - network analysis
  - term-document analysis
  - email analysis
  - link prediction
  - geospatial analysis
  - web page analysis

William & Mary
Graduate Arts & Sciences

Generalized CP (GCP)

- General loss function framework.
- All-at-once, gradient descent
- Variant to consider: *GCP with Adam optimization (GCP-Adam).*
  - stochastic gradient descent
  - linear convergence
  - scalable: uses sampling for objective function estimation and gradient computations
  - lower fraction of multi-starts converge to MLE

CP Alternating Poisson Regression (CPAPR)

- Specialized framework for Poisson loss with identity link.
- Alternating, block-coordinate descent
- Variant to consider: *Multiplicative Updates (MU).*
  - fixed-point iteration
  - sublinear convergence
  - performant: rich in dense matrix operations
  - higher fraction of multi-starts converge to MLE

Goal: *a hybrid method that leverages these advantages*

Cyclic GCP-CPAPR Hybrid

William & Mary
Graduate Arts & Sciences

Inspired by Simulated Annealing[†]

- Model solution space as thermodynamic system & move to a state with the lowest possible energy/temperature.

```
While (not converged)
    "Heat" the system to rise above local minima via stochastic search.
    "Cool" the system toward global minimum via deterministic search.
```

- Heating & cooling steps often follow a *strategy*---some parameterization of stochastic and deterministic search.

Cyclic GCP-CPAPR Hybrid Approach

For $l = 1, \ldots, L$
    Perform heating step via GCP according to some strategy.
    Perform cooling step via CPAPR according to some strategy.

- Possibly update strategy for each value of l (i.e., for each *cycle*).

[†] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. *Optimization by simulated annealing*. SCIENCE, 220 (4598): 671–680, 1983.

Numerical Experiments

- **Synthetic tensor $\mathcal{X}$:** 1000 x 1000 x 1000, R = 20, 0.01% dense, 10% of nonzeros are noisy
- **Out-of-sample validation set**
  - Run GCP-Adam & CPAPR-MU separately to convergence with very high precision & very large number of epochs (GCP) or iterations (CPAPR).
  - Repeat for N = 10,000 random starting points for each method.
  - Set MLE $\widehat{M}^* \coloneqq$ CP Poisson tensor model among all 20,000 approximations with lowest NLL value.
- **Cyclic GCP-CPAPR Hybrid experiment**
  - Fix W = 100, a *work budget* for all experiments.
  - Repeat for n = 100 random starting points.

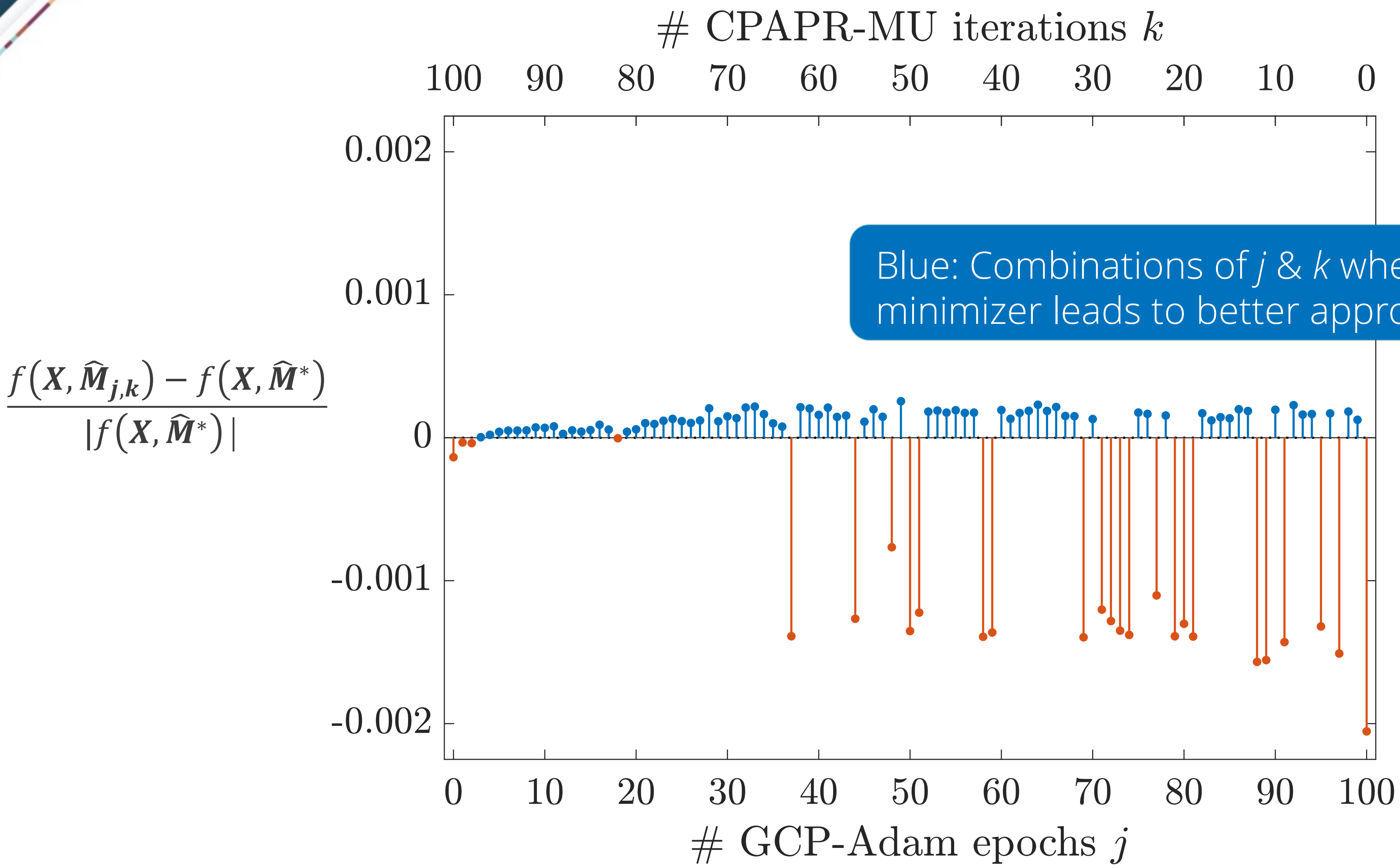
```
for j = 0, …, W,
    k = W - j
    run GCP-Adam starting with random M̂ for maximum j epochs -> M̂_1
    run CPAPR-MU starting with M̂_1 for maximum k iterations -> M̂_2
    Set M̂_{j,k} = M̂_2 as the current estimator
```

Results for Numerical Experiments

- Recall: Problem is non-convex, so we use multi-start to estimate MLE (global optimizer).
- $\hat{P}_A(\epsilon)$: estimates probability from our numerical experiments that method A converges to solution with NLL value in radius-ϵ ball of the MLE.

ϵ	$\hat{\boldsymbol{P}}_{\boldsymbol{GCP-Adam}}$	$\hat{\boldsymbol{P}}_{\boldsymbol{CPAPR-MU}}$	$\hat{\boldsymbol{P}}_{\boldsymbol{hybrid}}$	Best hybrid pair *(j,k)*
10^{-1}	1.00	1.00	1.00	all
10^{-2}	0.27	0.69	0.65	(0,100)
10^{-3}	0	0.05	0.16	(1,99)
10^{-4}	0	< 0.01	0.13	(4,96)
10^{-5}	0	0	0.03	(8,92)
10^{-6}	0	0	0.01	(8,92)

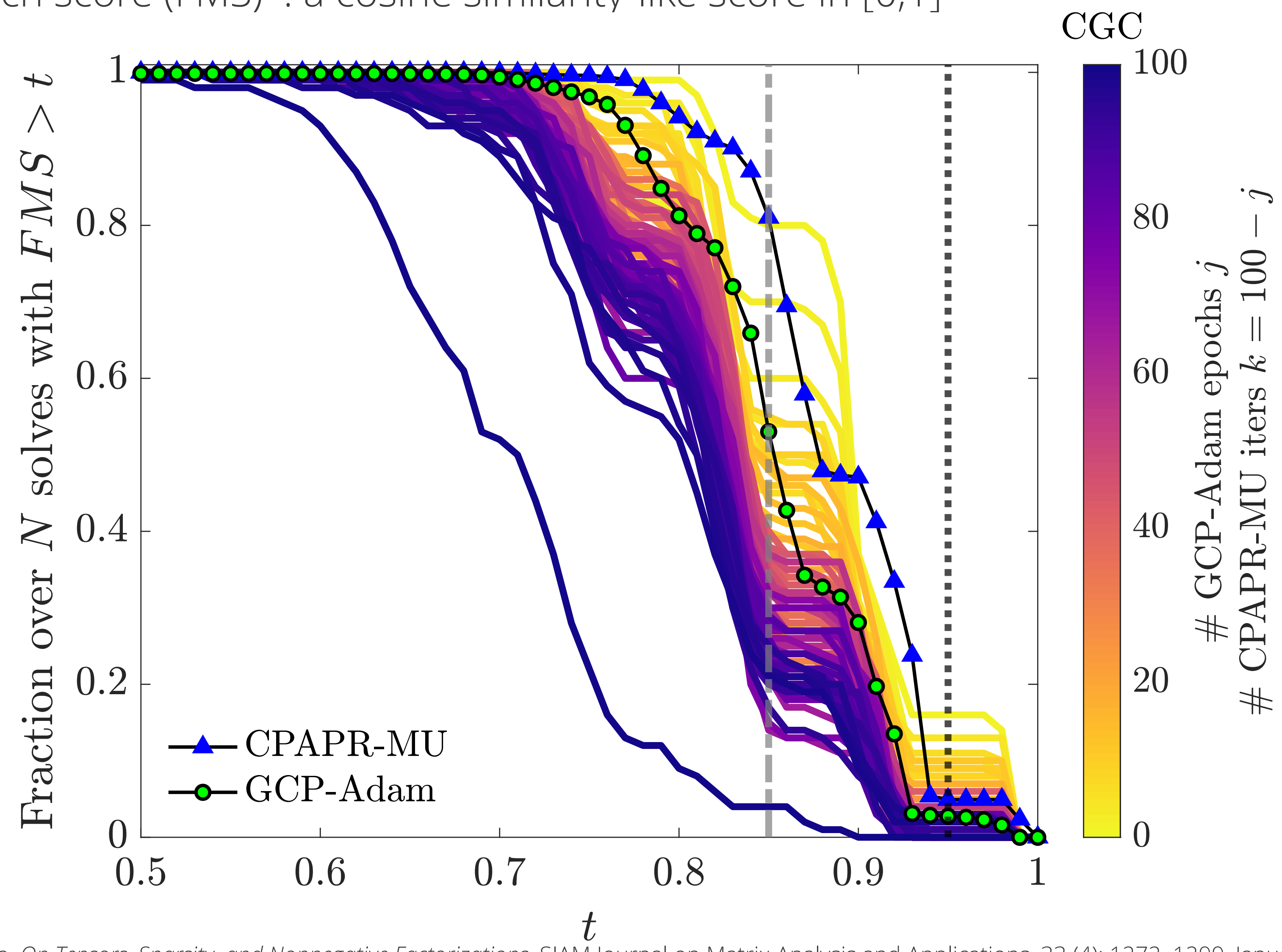
Results for Numerical Experiments



Blue: Combinations of j & k where hybrid minimizer leads to better approximate MLE

Results for Numerical Experiments

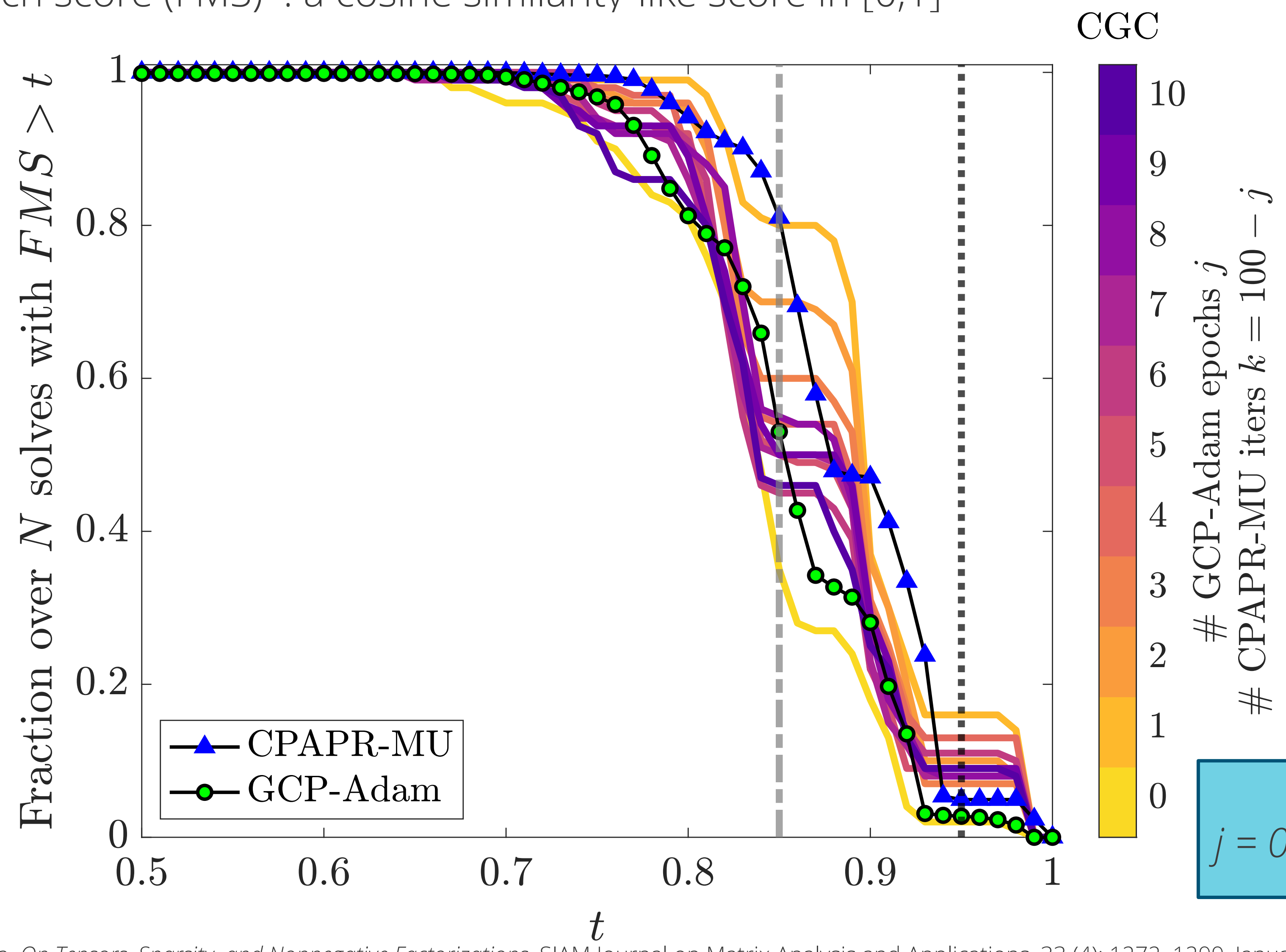
Factor match score (FMS) †: a cosine similarity-like score in [0,1]



† Eric C. Chi and Tamara G. Kolda. *On Tensors, Sparsity, and Nonnegative Factorizations*. SIAM Journal on Matrix Analysis and Applications, 33 (4): 1272–1299, January 2012. (Appendix E)

Results for Numerical Experiments

Factor match score (FMS) †: a cosine similarity-like score in [0,1]



Focus:
j = 0, ..., 10 only

† Eric C. Chi and Tamara G. Kolda. *On Tensors, Sparsity, and Nonnegative Factorizations*. SIAM Journal on Matrix Analysis and Applications, 33 (4): 1272–1299, January 2012. (Appendix E)

Conclusions and Future Work

Preliminary Conclusions regarding GCP-CPAPR Hybrid

- Can lead to better approximate MLEs (than using either method separately)
- Can be more computationally efficient (by using fewer multi-starts)

Ideas for Future Work

- Extend idea with $L > 1$ cycles
- Adaptive updates to strategies with $L > 1$ cycles
- Compare to black-box methods

Thank you!

Questions?

Jeremy Myers: jmmyers@cs.wm.edu, jermyer@sandia.gov
Danny Dunlavy: dmdunla@sandia.gov