On the extension of trace norm to tensors

Ryota Tomioka[1], Kohei Hayashi[2], Hisashi Kashima[1]

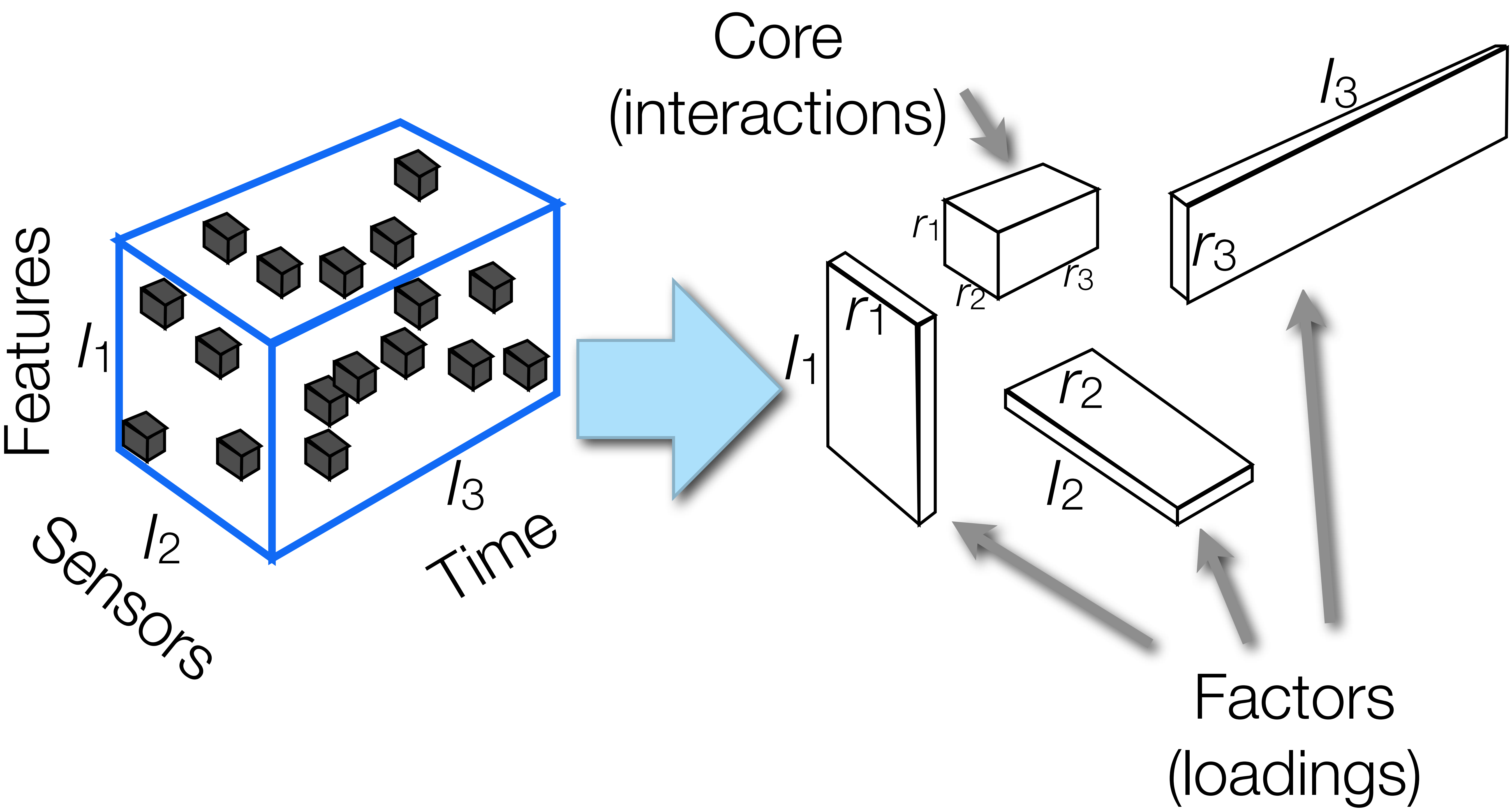
[1]The University of Tokyo
[2]Nara Institute of Science and Technology

2010/12/10

NIPS2010 Workshop: Tensors, Kernels, and Machine Learning

Convex low-rank *tensor* completion

Tucker decomposition



Conventional formulation (nonconvex)

$$\underset{\mathcal{C},\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3}{\text{minimize}} \|\Omega \circ (\mathcal{Y} - \mathcal{C} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3)\|_F^2 + \text{regularization.}$$

observation — mode-k product

$$\underset{\mathcal{X}}{\text{minimize}} \|\Omega \circ (\mathcal{Y} - \mathcal{X})\|_F^2 \quad \text{s.t.} \quad \text{rank}(\mathcal{X}) \leq (r_1, r_2, r_3).$$

- Alternate minimization
- Have to fix the rank beforehand

Our approach

Matrix

Estimation of *low-rank matrix* (hard) → Trace norm minimization (tractable)
[Fazel, Hindi, Boyd 01]

Generalization ↓

Tensor

Estimation of *low-rank tensor* (hard)
Rank defined in the sense of Tucker decomposition
→ Extended trace norm minimization (tractable)

Trace norm regularization (for matrices)

$$\boldsymbol{X} \in \mathbb{R}^{R \times C}, \quad m = \min(R, C)$$

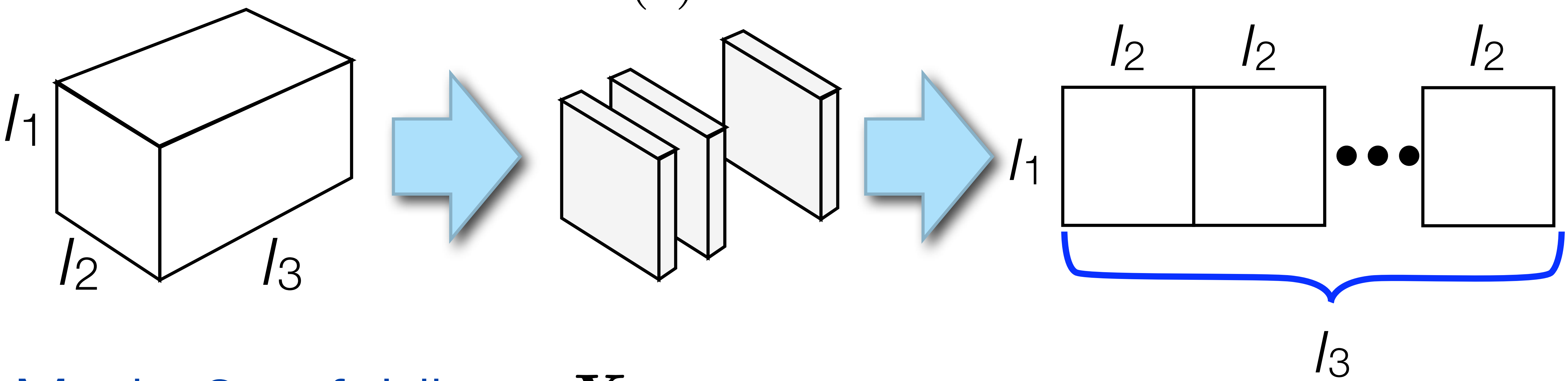
$$\|\boldsymbol{X}\|_{\mathrm{tr}} = \sum_{j=1}^{m} \sigma_j(\boldsymbol{X})$$

Linear sum of singular-values

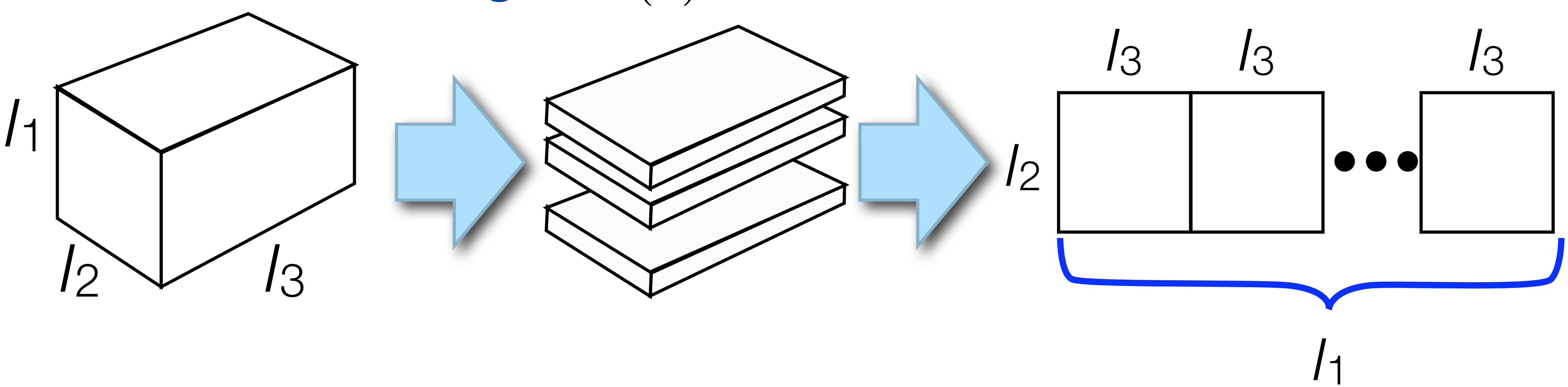
- Roughly speaking, L1 regularization on the singular-values.
- Stronger regularization --> more zero singular-values --> low rank.
- Not obvious for tensors (no singular-values for tensors)

Mode-k unfolding (matricization)

Mode-1 unfolding $\boldsymbol{X}_{(1)}$

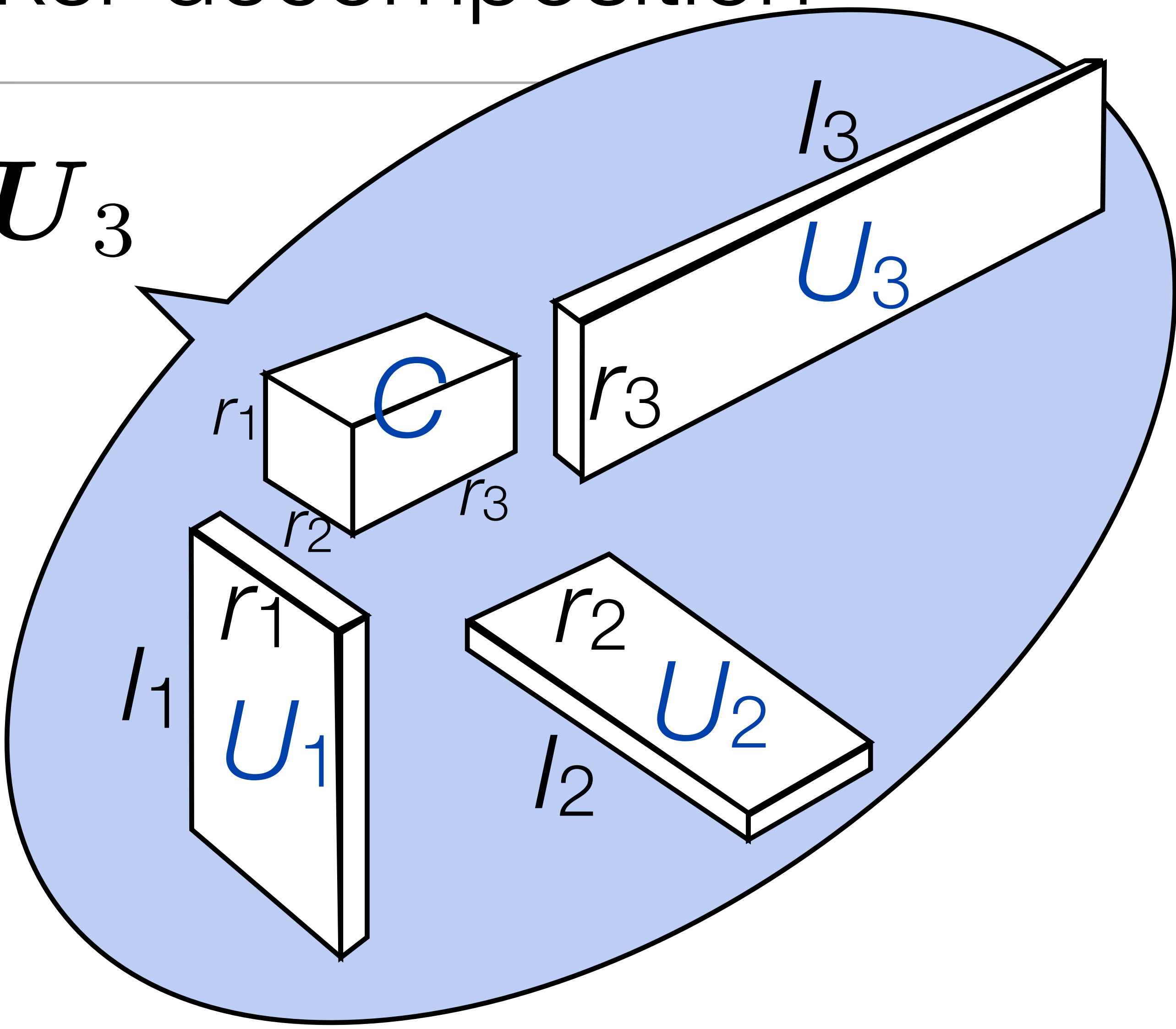


Mode-2 unfolding $\boldsymbol{X}_{(2)}$



Elementary facts about Tucker decomposition

$$\mathcal{X} = \mathcal{C} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3$$



Mode-1 unfolding

$$\boldsymbol{X}_{(1)} = \boldsymbol{U}_1 \boldsymbol{C}_{(1)} (\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)^\top$$

rank ≦ r_1

Mode-2 unfolding

$$\boldsymbol{X}_{(2)} = \boldsymbol{U}_2 \boldsymbol{C}_{(2)} (\boldsymbol{U}_1 \otimes \boldsymbol{U}_3)^\top$$

rank ≦ r_2

Mode-3 unfolding

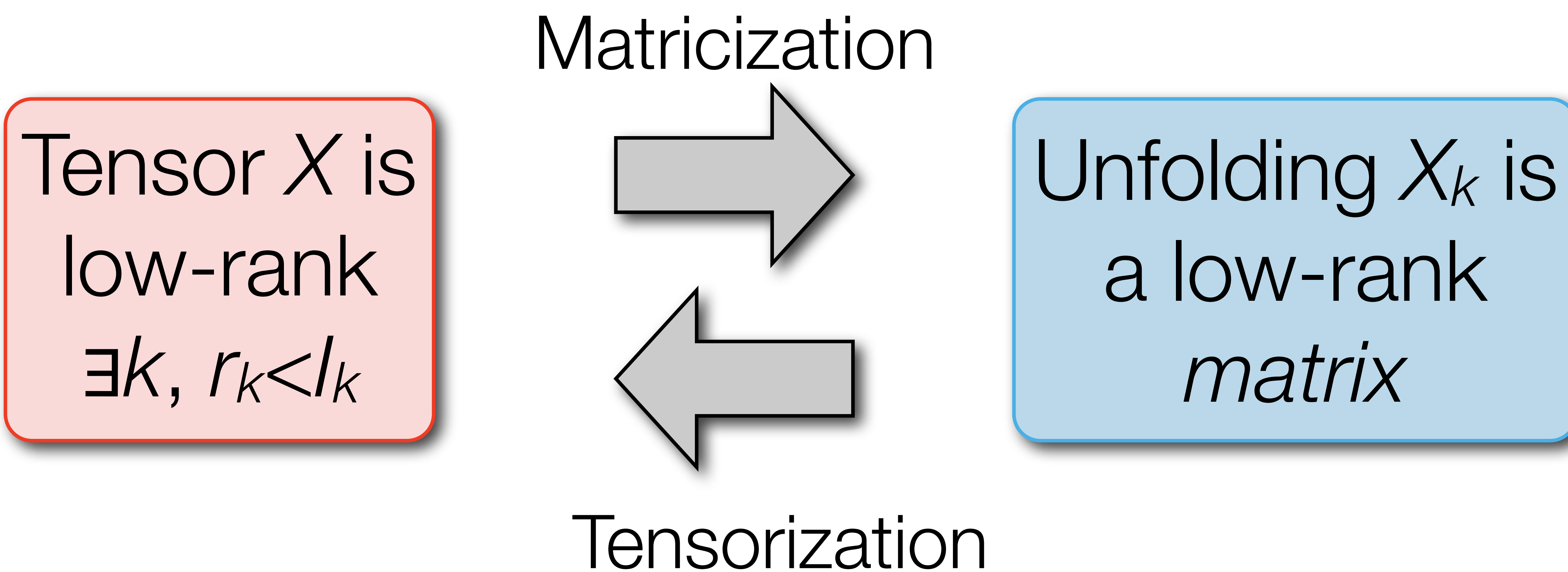
$$\boldsymbol{X}_{(3)} = \boldsymbol{U}_3 \boldsymbol{C}_{(3)} (\boldsymbol{U}_2 \otimes \boldsymbol{U}_1)^\top$$

rank ≦ r_3

The rank of $X_{(k)}$ is no more than the rank of $C_{(k)}$

What it means

- We can use the trace norm of an unfolding of a tensor *X* to learn low-rank *X.*



Approach 1: As a matrix

- Pick a mode *k*, and hope that the tensor to be learned is low rank in mode *k*.

$$\underset{\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_K}}{\text{minimize}} \quad \frac{1}{2\lambda} \|\Omega \circ (\mathcal{Y} - \mathcal{X})\|_F^2 + \|\boldsymbol{X}_{(k)}\|_*,$$

Pro: Basically a matrix problem
→ Theoretical guarantee (Candes & Recht 09)

Con: Have to be lucky to pick the right mode.

Approach 2: Constrained optimization

- Constrain so that each unfolding of *X* is simultaneously low rank.

$$\underset{\mathcal{X}\in\mathbb{R}^{I_1\times\cdots\times I_K}}{\text{minimize}} \quad \frac{1}{2\lambda}\|\Omega\circ(\mathcal{Y}-\mathcal{X})\|_F^2+\sum_{k=1}^{K}\gamma_k\|\boldsymbol{X}_{(k)}\|_*.$$

Pro: Jointly regularize every mode

Con: Strong constraint

γ_k: tuning parameter usually set to 1.

See also Marco Signoretto et al.,10

Approach 3: Mixture of low-rank tensors

- Each mixture component Z_k is regularized to be low-rank only in mode-k.

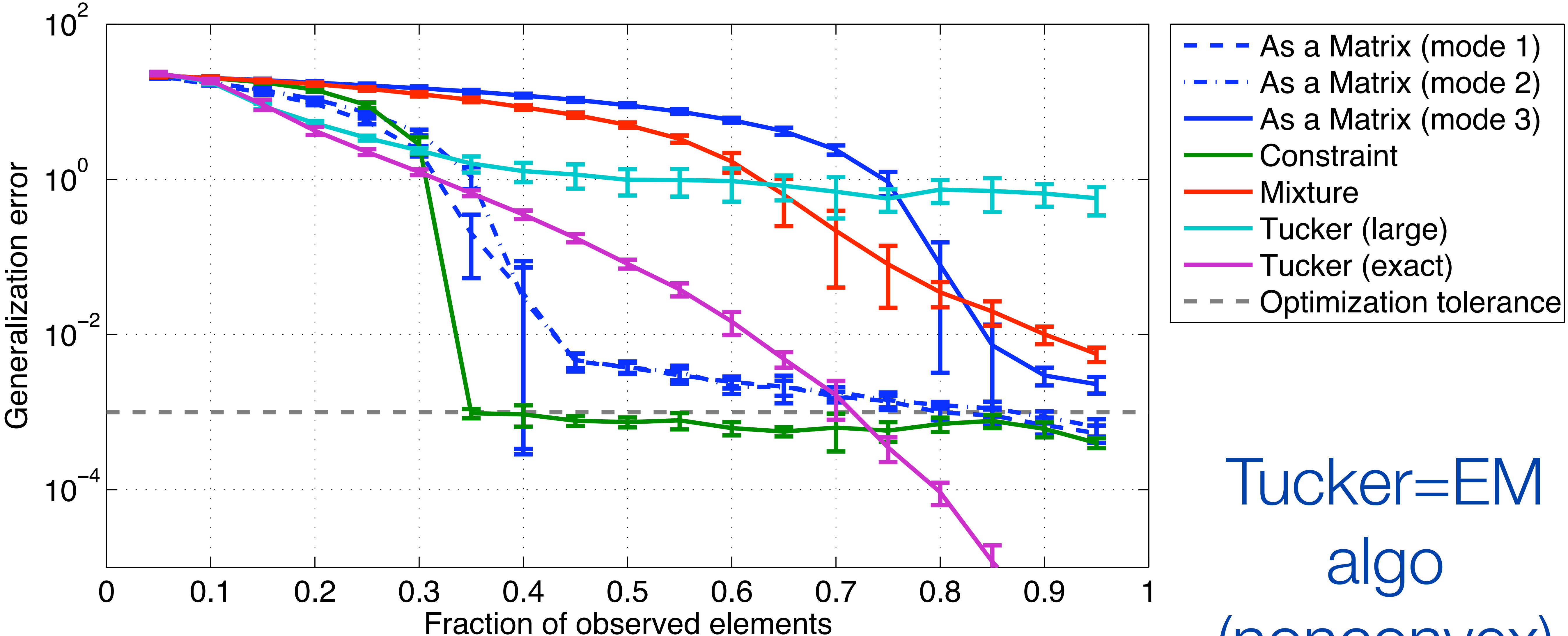
$$\underset{\mathcal{Z}_1,\ldots,\mathcal{Z}_K}{\text{minimize}} \quad \frac{1}{2\lambda}\left\|\Omega \circ \left(\mathcal{Y} - \sum\nolimits_{k=1}^{K} \mathcal{Z}_k\right)\right\|_F^2 + \sum_{k=1}^{K} \gamma_k \|\boldsymbol{Z}_{k(k)}\|_*,$$

Pro: Each Z_k takes care of each mode

Con: Sum is not low-rank

Numerical experiment

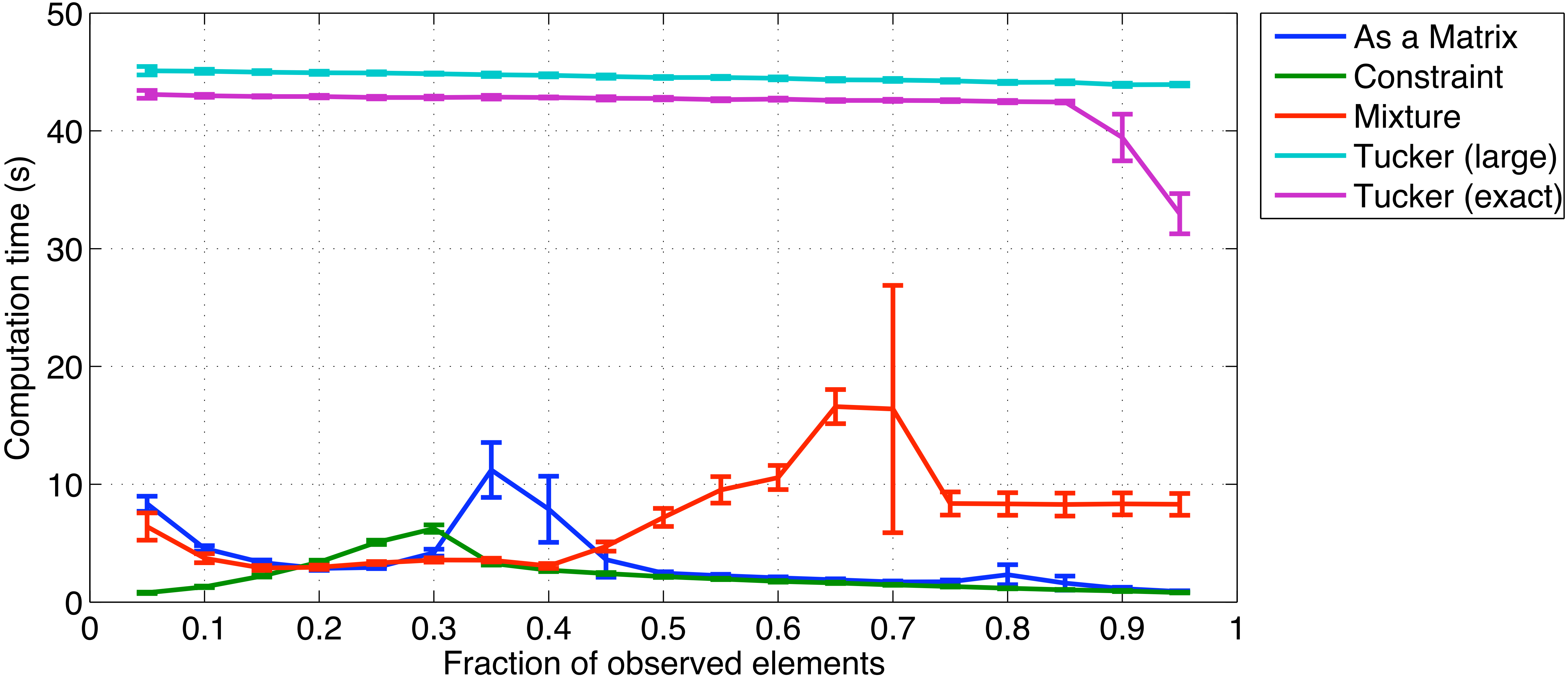
- True tensor: Size 50x50x20, rank 7x8x9. No noise (λ=0).
- Random train/test split.



Tucker=EM algo (nonconvex)

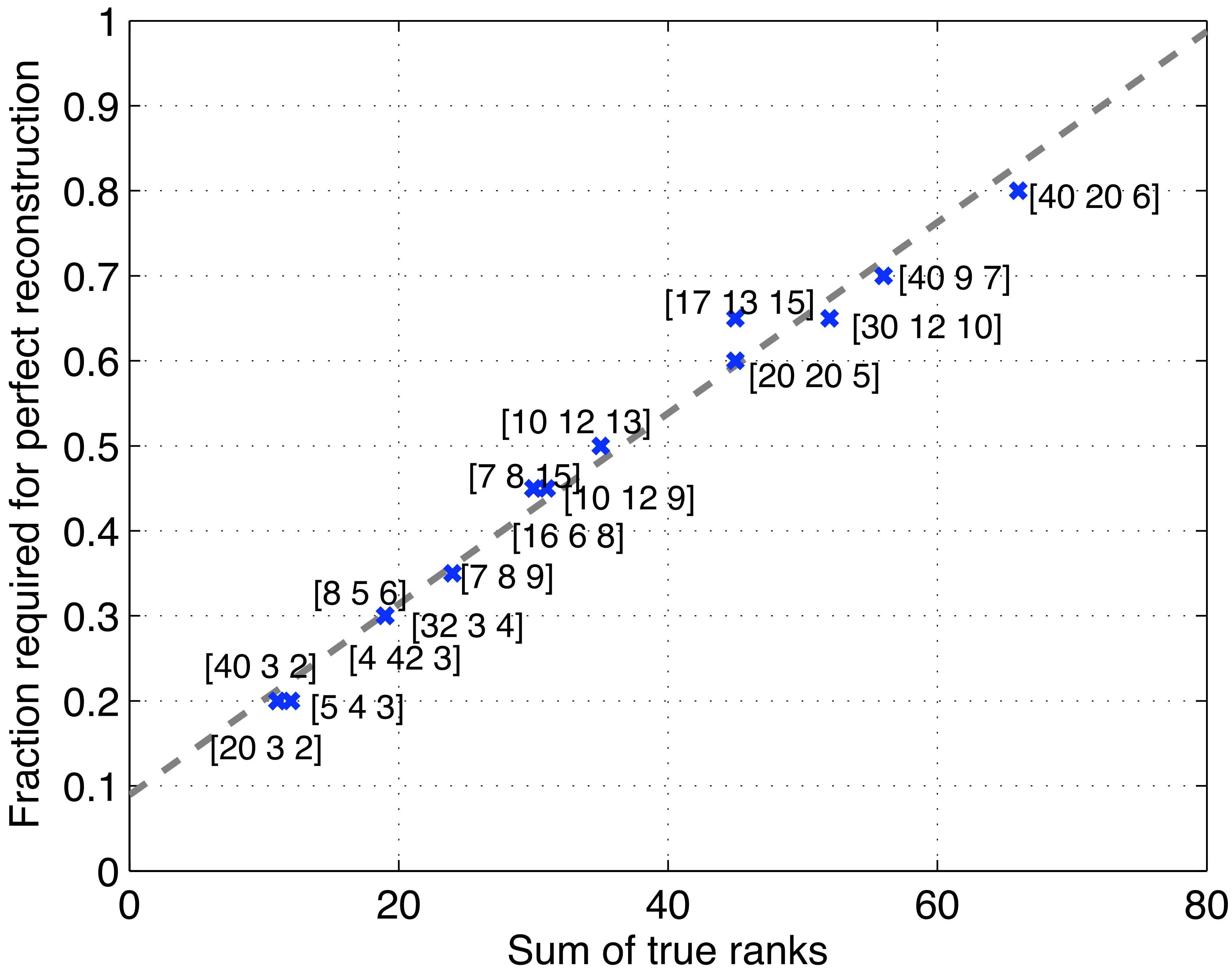
Computation time

- Convex formulation is also fast



Phase transition behaviour

- Sum of true ranks= $\min(r_1, r_2 r_3) + \min(r_2, r_3 r_1) + \min(r_3, r_1 r_2)$



Summary

- Low-rank tensor completion can be computed in a convex optimization problem using the trace norm of the unfoldings.
  - No need to specify the rank beforehand.
- Convex formulation is more accurate and faster than conventional EM-based Tucker decomposition.
- Curious "phase transition" found → compressive-sensing-type analysis is an on-going work.
- Technical report: Arxiv:1010.0789 (including optimization)
- Code:
  - http://www.ibis.t.u-tokyo.ac.jp/RyotaTomioka/Softwares/Tensor

Acknowledgment

- This work was supported in part by MEXT KAKENHI 22700138, 22700289, and NTT Communication Science Laboratories.