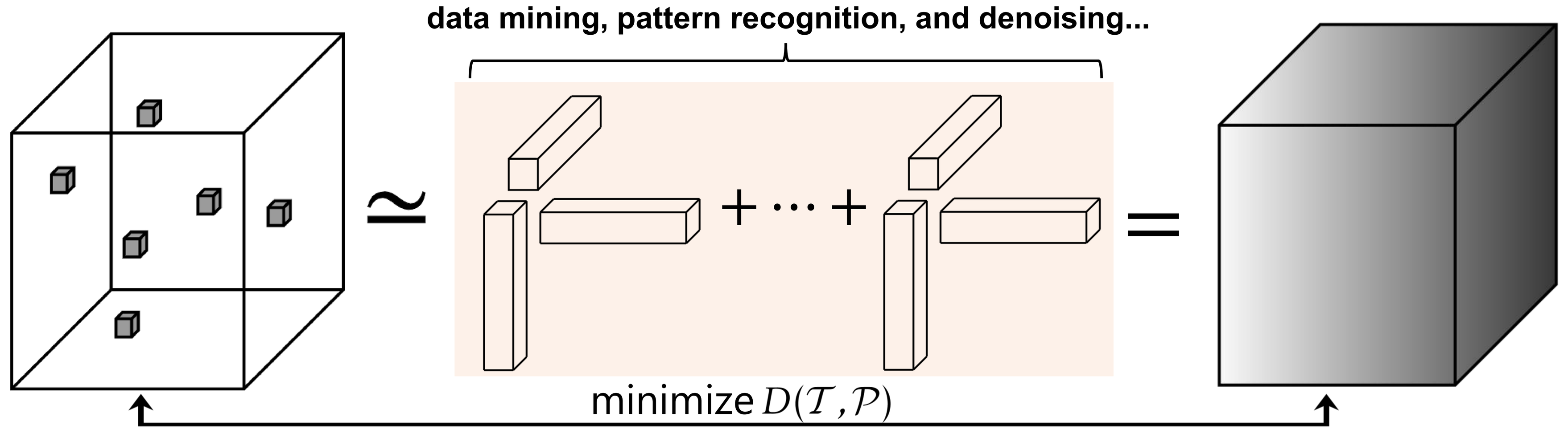
Deformed Decomposition for Non-negative Tensors

Kazu Ghalamkari[1], Petr Taborsky[1], Morten Mørup[1]

[1]Technical University of Denmark, Department of Applied Mathematics and Computer Science

Motivation & Background



Tensor decomposition is often **ill-posed** or **NP-hard**.
e.g., the rank-1 decomp. minimizing the L2 norm is NP-hard.
However, rank-1 decomp. minimizing the KL-divergence is a convex optimization.

Research Question:

What about divergences beyond KL?
Can we retain convex optimization (global optimality) under richer divergences?

When is the KL-divergence a bad fit?

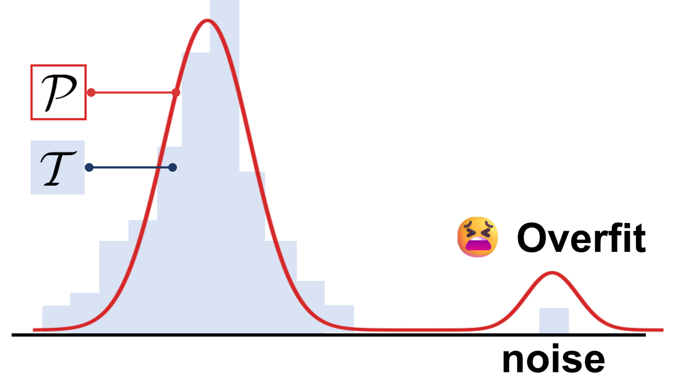
The KL-divergence can overfit to noise and outliers.

KL-divergence

$$D_{\mathrm{KL}}(\mathcal{T},\mathcal{P}) = \sum_{ijk} \mathcal{T}_{ijk} \log \frac{\mathcal{T}_{ijk}}{\mathcal{P}_{ijk}}$$

Model, Data

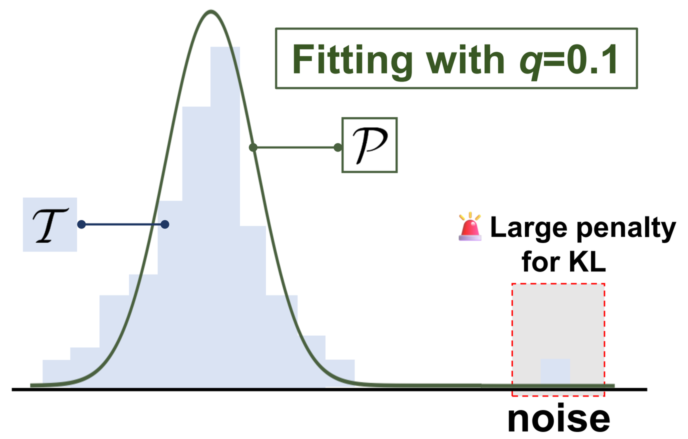
Large penalty if $\mathcal{T}_{ijk} \neq 0$ and $\mathcal{P}_{ijk} = 0$.



Tsallis *q*-divergence

$$D_q(\mathcal{T},\mathcal{P}) = \frac{1}{\sum_{ijkl} \mathcal{T}^q_{ijkl}} \frac{1}{1-q} \sum_{ijkl} \left\{ 1 - \mathcal{T}^q_{ijkl} \mathcal{P}^{1-q}_{ijkl} \right\}$$

Temperature *q*>0 controls the sensitivity to noise



Deformed algebra

χ-exponential and logarithm function

For any increasing function $\chi : \mathbb{R}_+ \to \mathbb{R}_+$

$$\mathrm{Log}_\chi[x] = \int_1^x \frac{dt}{\chi(t)}, \quad \mathrm{Exp}_\chi[x] = \mathrm{Log}_\chi^{-1}[x]$$

Deformed product

$$\mathrm{Exp}_\chi[a+b] = \mathrm{Exp}_\chi[a] \underset{\chi}{\circledast} \mathrm{Exp}_\chi[b]$$

Example 1: Tsallis deformation

$$\chi(t) = t^q \ (q>0)$$
$$\mathrm{Log}_q(x) = \frac{x^{1-q}-1}{1-q}$$
$$\mathrm{Exp}_q(x) = [1+(1-q)x]^{1/(1-q)}$$

Example 2: Kaniadakis deformation

$$\chi(t) = \frac{2t^{\kappa+1}}{t^{2\kappa}+1}$$
$$\mathrm{Log}_\kappa(x) = \frac{x^\kappa - x^{-\kappa}}{2\kappa}$$
$$\mathrm{Exp}_\kappa(x) = (\kappa x + \sqrt{1+\kappa^2 x^2})^{1/\kappa}$$

Deformed Many-Body Approximation

$$\mathcal{T}_{ijkl} = \mathrm{Exp}_\chi\left[-\psi_\chi + E_\theta(i,j,k,l)\right]$$

Free energy; Natural parameters of the deformed exponential family

$$= \mathrm{Exp}_\chi\left[-\psi_\chi + E^1_i + \dots + E^4_l + E^{12}_{ij} + \dots + E^{34}_{kl} + E^{123}_{ijk} + \dots + E^{1234}_{ijkl}\right]$$

One-body approx.

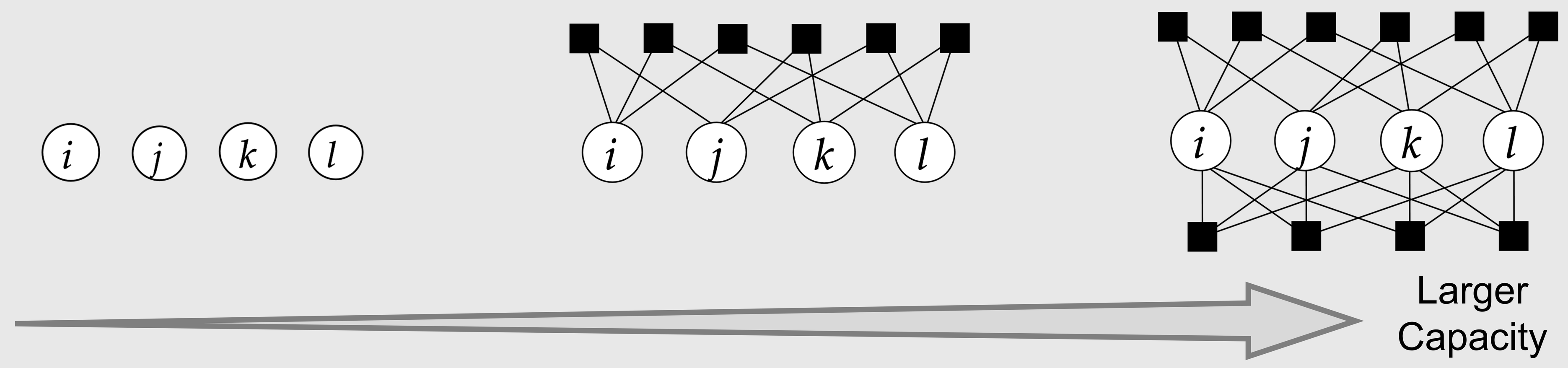
$$\mathcal{P}_{ijkl} = p^1_i \underset{\chi}{\circledast} \cdots \underset{\chi}{\circledast} p^4_l$$

Two-body approx.

$$\mathcal{P}_{ijkl} = X^{12}_{ij} \underset{\chi}{\circledast} \dots \underset{\chi}{\circledast} X^{34}_{kl}$$

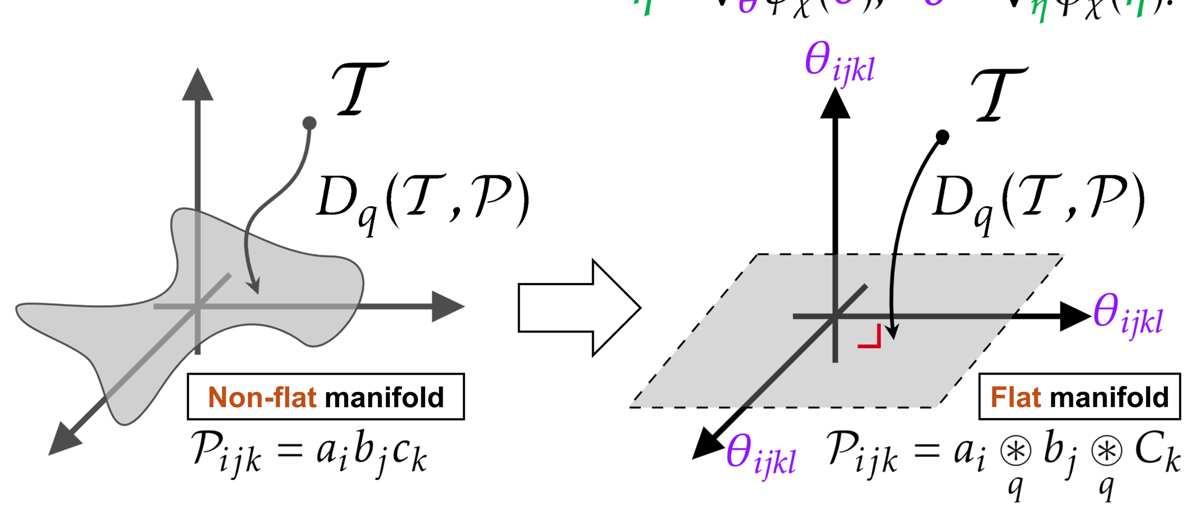
Three-body approx.

$$\mathcal{P}_{ijkl} = X^{123}_{ijk} \underset{\chi}{\circledast} \dots \underset{\chi}{\circledast} X^{234}_{jkl}$$



Dually-flat manifold generated by convex functions $(\psi_\chi(\theta), \varphi_\chi(\eta))$

$\eta = \nabla_\theta \psi_\chi(\theta), \quad \theta = \nabla_\eta \varphi_\chi(\eta).$



Non-convex optimization / **Convex optimization**

χ-entropy: $\varphi_\chi(\eta) = \sum_{ijkl} \tilde{x}[\mathcal{P}_{ijkl}] \mathrm{Log}_\chi[\mathcal{P}_{ijkl}]$

Bregman divergence generated by φ_χ:

$$D_\chi(\mathcal{T},\mathcal{P}) = \varphi_\chi(\eta^{\mathcal{T}}) - \varphi_\chi(\eta^{\mathcal{P}}) - \nabla_\eta \varphi_\chi(\eta^{\mathcal{P}})(\eta^{\mathcal{T}} - \eta^{\mathcal{P}}) = \sum_{ijkl} \tilde{x}[\mathcal{T}_{ijkl}]\left(\mathrm{Log}_\chi[\mathcal{T}_{ijkl}] - \mathrm{Log}_\chi[\mathcal{P}_{ijkl}]\right)$$

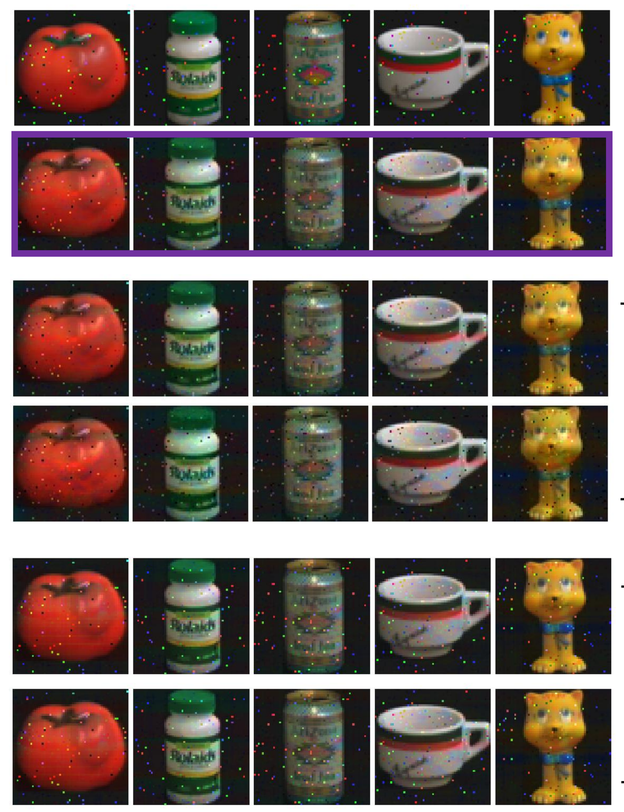
where the escort is $\tilde{x}[\mathcal{T}_{ijkl}] = \frac{x[\mathcal{T}_{ijkl}]}{\sum_{ijkl} x[\mathcal{T}_{ijkl}]}$.

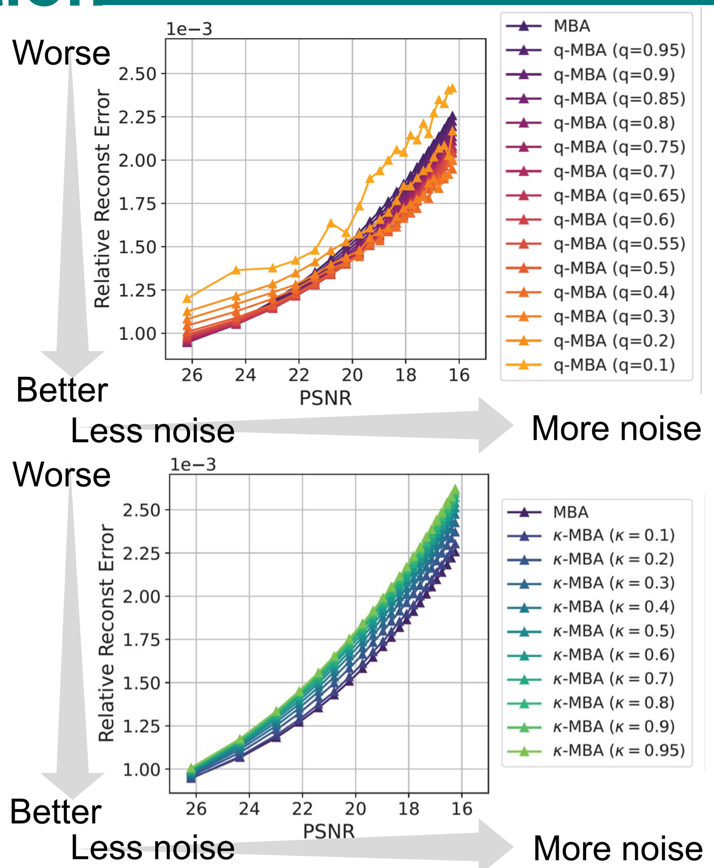
Examples of *χ*-divergence

$\chi(t) = t^q \quad (q>0)$

$$D_q(\mathcal{T},\mathcal{P}) = \frac{1}{\sum_{ijkl} \mathcal{T}^q_{ijkl}} \frac{1}{1-q} \sum_{ijkl} \left\{1 - \mathcal{T}^q \mathcal{P}^{1-q}\right\}$$

Experiment: Denoising by factorization





Deformed Low-rank Approximation

$$\mathcal{T}_{ijk} \simeq \mathcal{P}_{ijk} = \sum_{r=1}^{R} A_{ir} \underset{\chi}{\circledast} B_{jr} \underset{\chi}{\circledast} C_{kr}$$

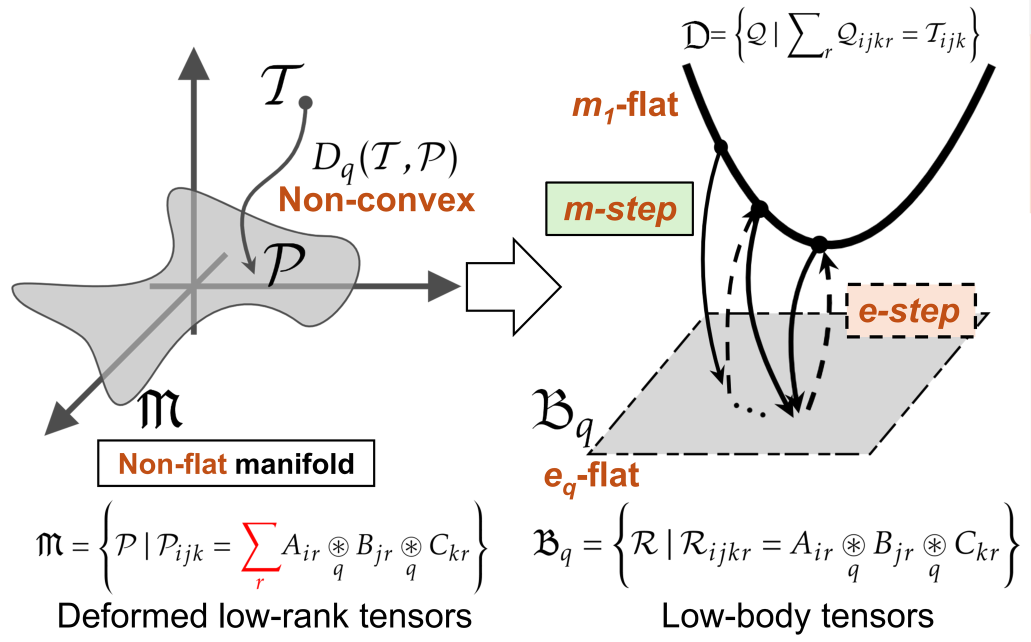
No longer convex, Is there a benefit?
Yes, Implicit regularization!

***q*-deformation induces implicit regularization**

$\mathrm{rank}(\mathcal{P}) \leq D \quad (q \to 0)$ — Tensor order

$$x \underset{q}{\circledast} y = (x^{1-q} + y^{1-q} - 1)^{1/(1-q)} = \begin{cases} xy & (q \to 1) \\ x + y - 1 & (q \to 0) \end{cases}$$

em-based optimization



***e*-step: e_1-projection onto m_1-flat manifold**

$$\mathcal{Q}^{t+1} \leftarrow \underset{\mathcal{Q}^t \in \mathfrak{D}}{\arg\min}\, D_q(\mathcal{Q}^t, \mathcal{R}^t)$$

The optimal update is $\mathcal{Q}^*_{ijkr} = \frac{\mathcal{T}_{ijk}\mathcal{R}_{ijkr}}{\sum_r \mathcal{R}_{ijkr}}$.

***m*-step: m_q-projection onto e_q-flat manifold**

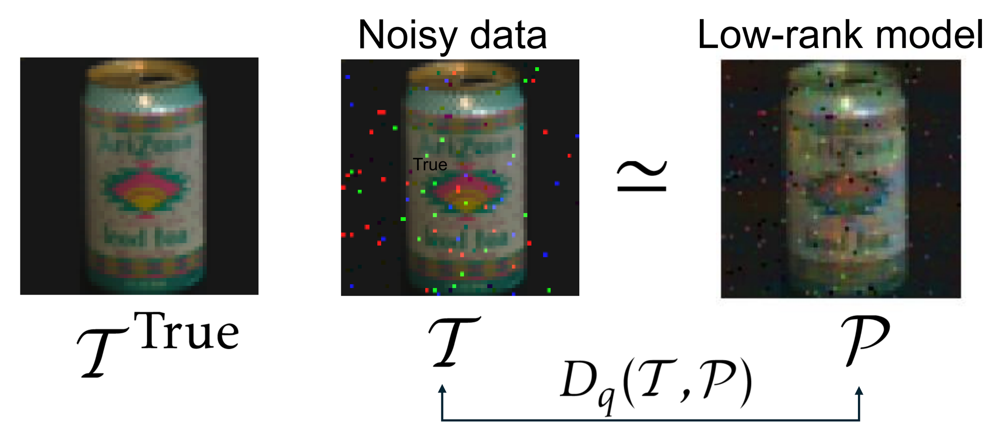
$$\mathcal{R}^{t+1} \leftarrow \underset{\mathcal{R}^t \in \mathcal{B}_q}{\arg\min}\, D_q(\mathcal{Q}^{t+1}, \mathcal{R}^t)$$

***q*-deformed many-body approximation**

$$\mathcal{Q}_{ijkr} \simeq \mathcal{R}_{ijkr} = A_{ir} \underset{q}{\circledast} B_{jr} \underset{q}{\circledast} C_{kr}$$

- Convergence guarantee
- Optimization of each step is convex

Experiment: Noisy image reconstruction



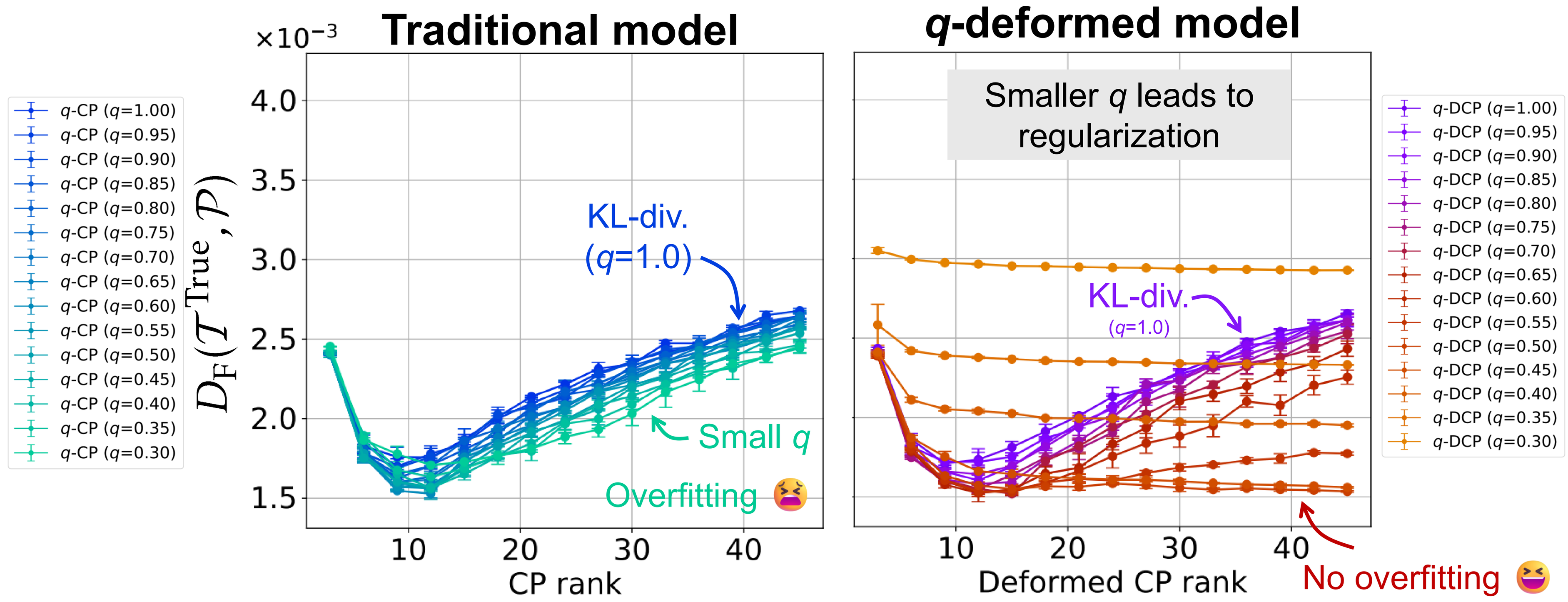
① Traditional model

$$\mathcal{T}_{ijk} \simeq \mathcal{P}_{ijk} = \sum_{r=1}^{R} A_{ir} B_{jr} C_{kr}$$

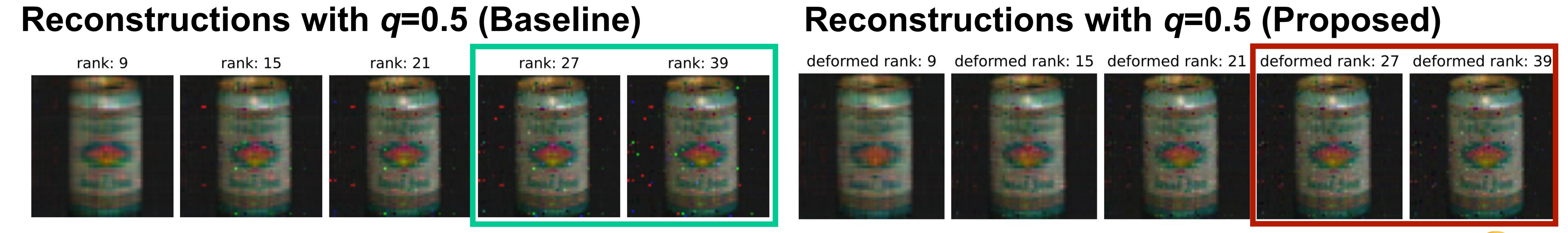
② *q*-deformed model

$$\mathcal{T}_{ijk} \simeq \mathcal{P}_{ijk} = \sum_{r=1}^{R} A_{ir} \underset{q}{\circledast} B_{jr} \underset{q}{\circledast} C_{kr}$$

Deformed rank



Reconstructions with *q*=0.5 (Baseline) / **Reconstructions with *q*=0.5 (Proposed)**



Overfits to noise / No noise

The 29th International Conference on Artificial Intelligence and Statistics (AISTATS 2026), Tangier, Morocco, 2-5 May, 2026