Interpolating Convex and Non-Convex Tensor Decompositions via the Subspace Norm

Qinqing Zheng* and Ryota Tomioka†,‡ *University of Chicago †TTI Chicago ‡Microsoft Research Cambridge

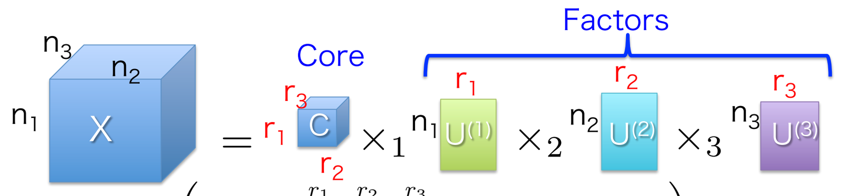
NIPS 2015, Montréal, Canada

Our results

1. We show that a rank-one tensor corrupted by independent Gaussian noise can be reliably recovered when the signal to noise ratio is larger than $n^{K/4}$ where n is the dimension and K is the order of the tensor. This naturally includes the matrix case ($K=2$) and confirms the conjecture of RM14.
2. Our estimator is a simple mode-wise unfolding. Our careful analysis shows an interesting two-phase behavior.
3. We use the estimated top H left singular vectors to define a collections of subspaces and define a new tensor norm based on it. The denoising bound for the proposed subspace norm that scales much milder w.r.t. the dimension when H is sufficiently small than n.
4. Experimental results confirm that the denoising performance of the proposed subspace norm is nearly optimal ($\sqrt{n}$).

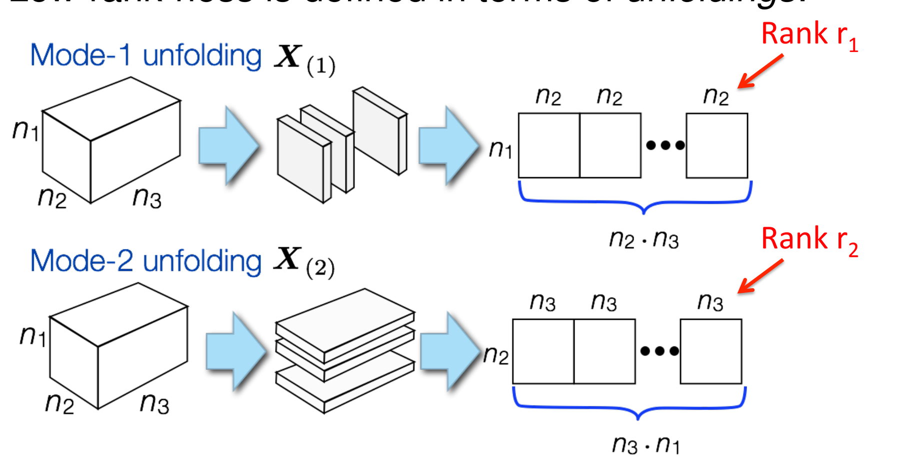
Tensor decomposition

Tucker decomposition / HOSVD [Tucker 66; de Lathauwer+00]



$$\left(X_{ijk}=\sum_{a=1}^{r_1}\sum_{b=1}^{r_2}\sum_{c=1}^{r_3}C_{abc}U^{(1)}_{ia}U^{(2)}_{jb}U^{(3)}_{kc}\right)$$

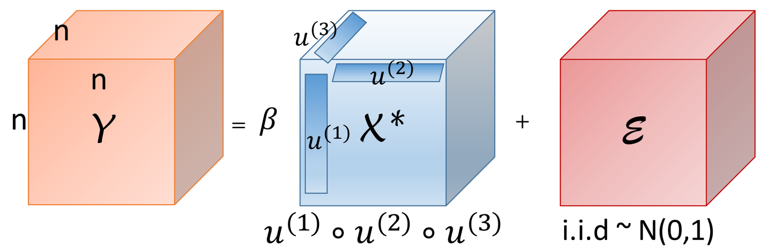
Low-rank-ness is defined in terms of *unfoldings*:



- mode-k rank $r_k := \mathrm{rank}(\boldsymbol{X}_{(k)})$.
- Higher-order generalization of singular-value decomposition.

Tensor denoising: previous results

Rank-one K-th order tensor corrupted by standard Gaussian noise

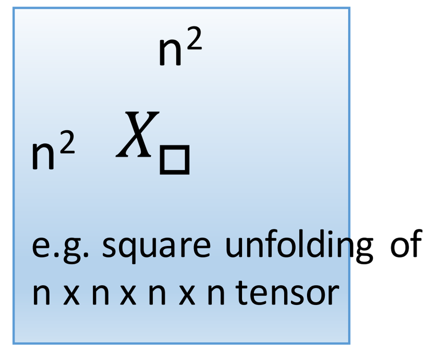


How large does the signal-to-noise ratio β have to be in order to recover $\mathcal{X}^*$?

Method	Required β	Drawback
Overlapped / Latent nuclear norm (ordinary unfolding)	$O(n^{(K-1)/2})$	disappointing recovery guarantee
Square unfolding	$O(n^{K/4})$	only applies to even order tensor
Recursive unfolding	$O(n^{\lceil K/2\rceil/2})$	
Ideal	$\boldsymbol{O(\sqrt{nK\log(K)})}$	

Conjecture (RM 14): $O(n^{K/4})$ **bound also applies to odd order tensor.**

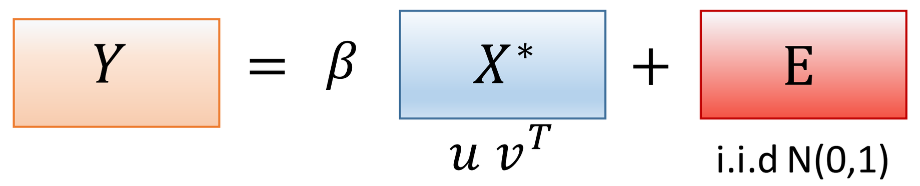
Square Unfolding for even order tensor



- For $n\times m$ matrix filled with i.i.d. standard Gaussian entries, the spectral norm scales as $O(\sqrt{n}+\sqrt{m})$.
- $\beta \geq O(n^{K/4})$ is sufficient because $\mathbb{E}\|\mathcal{E}_{\square}\|_{\mathrm{op}} = O(n^{K/4})$.
- The limitation comes because the best we can do with an odd order tensor is unfolding it as $O(n^{\lfloor K/2\rfloor}\times n^{\lceil K/2\rceil})$.

Tensor denoising via ordinary unfolding

Consider the perturbed matrix:



Key observation:

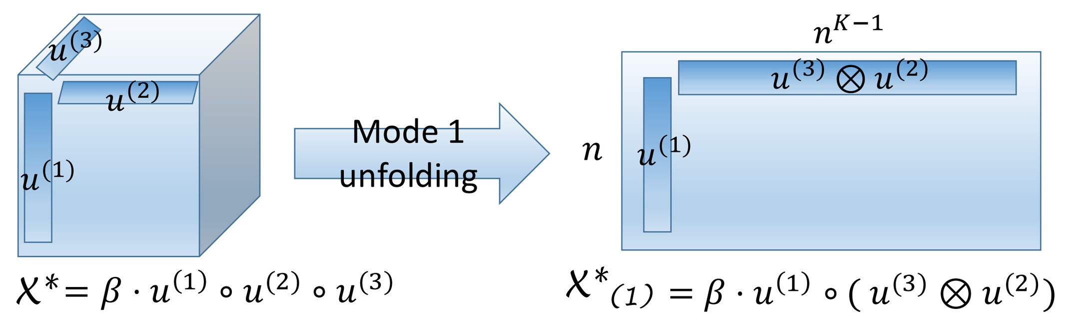
$$\boldsymbol{Y}\boldsymbol{Y}^\top = (\beta^2\boldsymbol{u}\boldsymbol{u}^\top + m\sigma^2\boldsymbol{I}) + \underbrace{(\boldsymbol{E}\boldsymbol{E}^\top - m\boldsymbol{I})}_{\text{Wishart noise}} + \beta\underbrace{(\boldsymbol{u}\boldsymbol{v}^\top\boldsymbol{E}^\top + \boldsymbol{E}\boldsymbol{v}\boldsymbol{u}^\top)}_{\text{Gaussian noise}}.$$

Theory of two phase behavior:

Let $\hat{\boldsymbol{u}}$ be the leading left singular vector of $\boldsymbol{Y}$. There exists a constant C such that with high probability if $m/n \geq C$

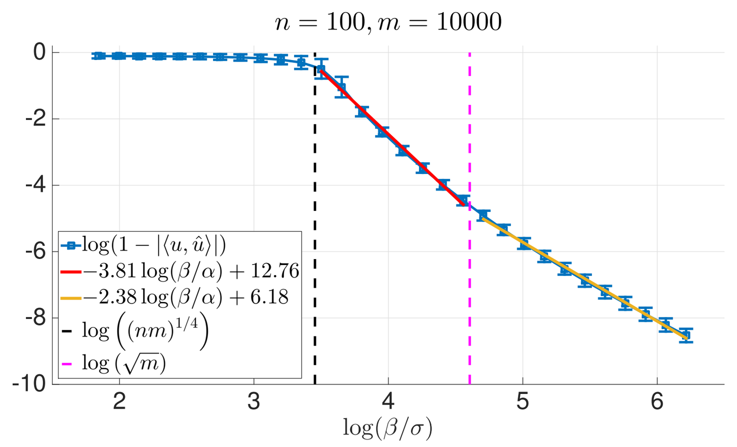
$$|\langle \hat{\boldsymbol{u}}, \boldsymbol{u}\rangle| \geq \begin{cases} 1-\dfrac{Cnm}{\beta^4} & \text{if } \sqrt{m} > \beta \geq (Cnm)^{\frac{1}{4}} \quad \textit{(Wishart term dominates)} \\ 1-\dfrac{Cn}{\beta^2} & \text{if } \beta \geq \sqrt{m} \quad \textit{(Gaussian term dominates)} \end{cases}$$

Apply to ordinarily unfolded tensor: $n\times n^{K-1}$ **matrix**



$O(n^{k/4})$ **recovery guarantee! Apply to both even and odd order tensors.**

Synthetic Experiments: ordinary unfolding



Recover the left singular vector of a 100x10000 rank one matrix. The distance between $\hat{u}$ and u first decreases as $1/\beta^4$ (red line) and then $1/\beta^2$ (yellow line).

The subspace norm

Definition: associated with $\boldsymbol{S}^{(k)}$

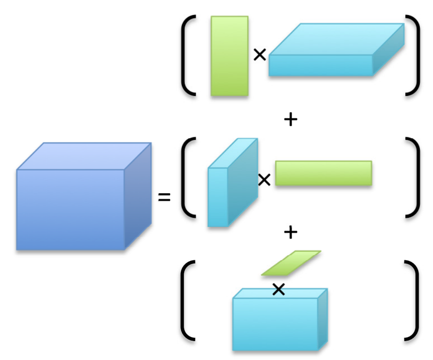
$$\|\mathcal{X}\|_{\text{subspace}} = \inf_{\mathcal{X}=\mathrm{fold}(\sum_{k=1}^K \boldsymbol{M}^{(k)}\boldsymbol{S}^{(k)})} \sum_{k=1}^K \|\boldsymbol{M}^{(k)}\|_{\mathrm{tr}}$$

dual

$$\|\mathcal{X}\|_{\text{subspace}^*} = \max_k \|\boldsymbol{X}_{(k)}\boldsymbol{S}^{(k)}\|_{\mathrm{op}}$$

Intuition

Models a low-rank tensor as a mixture of tensors that each are low-rank in a specific mode. The low-rank subspaces in each mode are pre-fixed to be $\boldsymbol{S}^{(k)}$'s.



- Reduce to the latent norm (T+10; TS13) if $\boldsymbol{S}^{(k)} = \boldsymbol{I}$:

$$\|\mathcal{X}\|_{\text{latent}} = \inf_{\mathcal{X}=\mathrm{fold}(\sum_{k=1}^K \boldsymbol{M}^{(k)})} \sum_{k=1}^K \|\boldsymbol{M}^{(k)}\|_{\mathrm{tr}}$$

- Scaling of the dual norm depends of the size of $\boldsymbol{S}^{(k)}$ ($n^{K-1}\times T$):

$$\mathbb{E}\|\mathcal{E}\|_{\text{subspace}^*} = O(\sqrt{T}+\sqrt{n})$$

Approach: Ordinary unfolding + Subspace norm

1. For each mode k, unfold observed tensor and compute the top H left singular vectors. Concatenate them to obtain a $n\times H$ matrix $\widehat{\boldsymbol{P}}^{(k)}$.
2. Construct $\boldsymbol{S}^{(k)}$ as $\boldsymbol{S}^{(k)} = \widehat{\boldsymbol{P}}^{(1)} \otimes\cdots\otimes \widehat{\boldsymbol{P}}^{(k-1)} \otimes \widehat{\boldsymbol{P}}^{(k+1)} \otimes\cdots\otimes \widehat{\boldsymbol{P}}^{(K)}$.
3. Solve the subspace norm (associated with $\{\boldsymbol{S}^{(k)}\}_{k=1}^K$) regularized problem

$$\min_{\mathcal{X}} \ \frac{1}{2}\|\mathcal{Y}-\mathcal{X}\|_F^2 + \lambda\|\mathcal{X}\|_{\text{subspace}}$$

- The parameter H controls the blend from the non-convex estimator (e.g. HOSVD) to mode-wise nuclear norm.

Theoretical results

Subspace: the Kronecker structure

Let the $\boldsymbol{X}^*_{(k)} = \boldsymbol{P}^{(k)}\boldsymbol{\Lambda}^{(k)}\boldsymbol{Q}^{(k)}$ be the SVD of $\boldsymbol{X}^*_{(k)}$. For all k,

i) $\boldsymbol{U}^{(k)} \in \mathrm{Span}(\boldsymbol{P}^{(k)})$,

ii) $\boldsymbol{Q}^{(k)} \in \mathrm{Span}\left(\boldsymbol{P}^{(1)}\otimes\cdots\otimes\boldsymbol{P}^{(k-1)}\otimes\boldsymbol{P}^{(k+1)}\otimes\cdots\otimes\boldsymbol{P}^{(K)}\right)$.

Recovery Bound

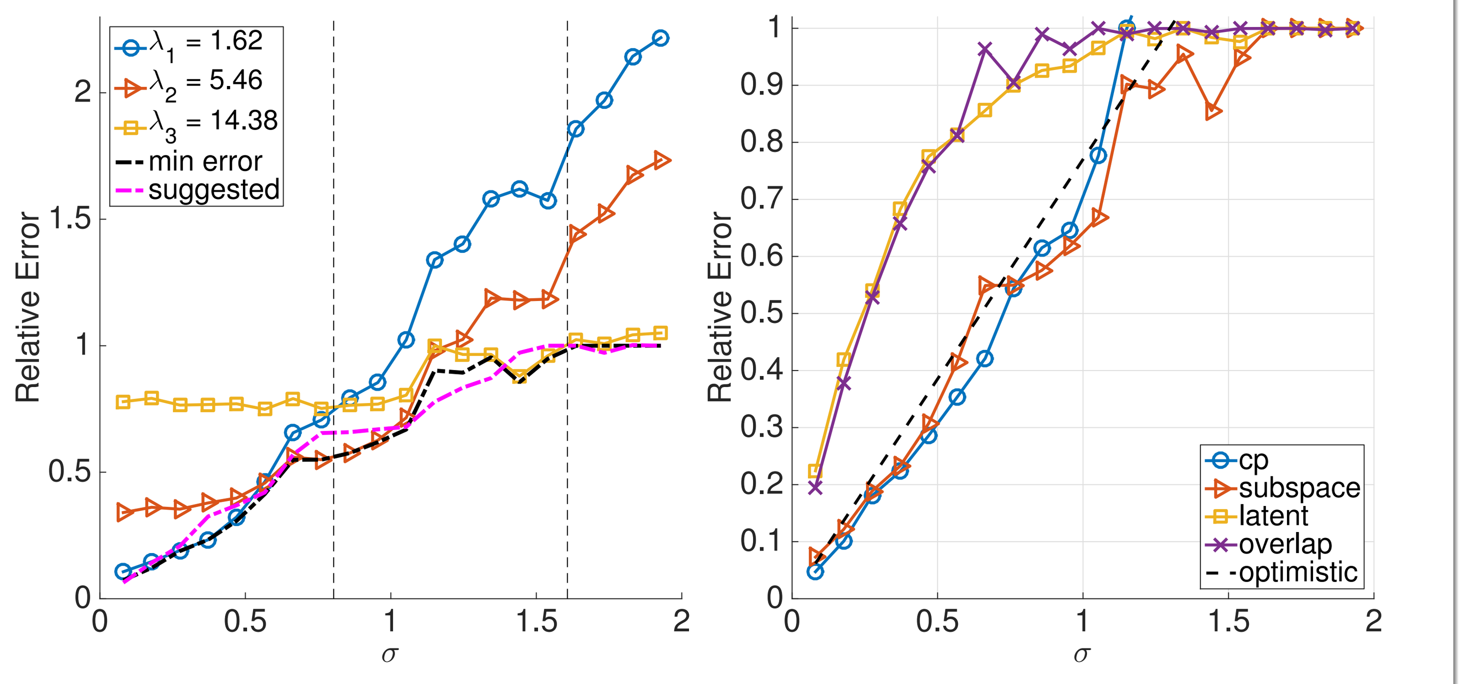
Under mild conditions, with high probability

$$\|\hat{\mathcal{X}}-\mathcal{X}^*\|_F \leq \underbrace{\inf_{\mathcal{X}_p=\sum_{k=1}^K \mathrm{fold}_k\left(\boldsymbol{M}_p^{(k)}\boldsymbol{S}^{(k)}\right)} \|\mathcal{X}_p-\mathcal{X}^*\|_F}_{\text{approximation error}} + \underbrace{O\left(\left(\sqrt{n}+\sqrt{H^{K-1}}\right)\sqrt{\sum_{k=1}^K r_k}\right)}_{\text{estimation error}}$$

- Trade-off: $H\uparrow$, approximation error $\downarrow$ but estimation error $\uparrow$. If $\mathcal{X}^* \in \mathrm{Span}(\{\boldsymbol{S}^{(k)}\}_1^K)$, approximation error is zero. Reduce to latent trace norm if $H=n$.
- Open questions: how to choose H and how to analyze the trade-off.

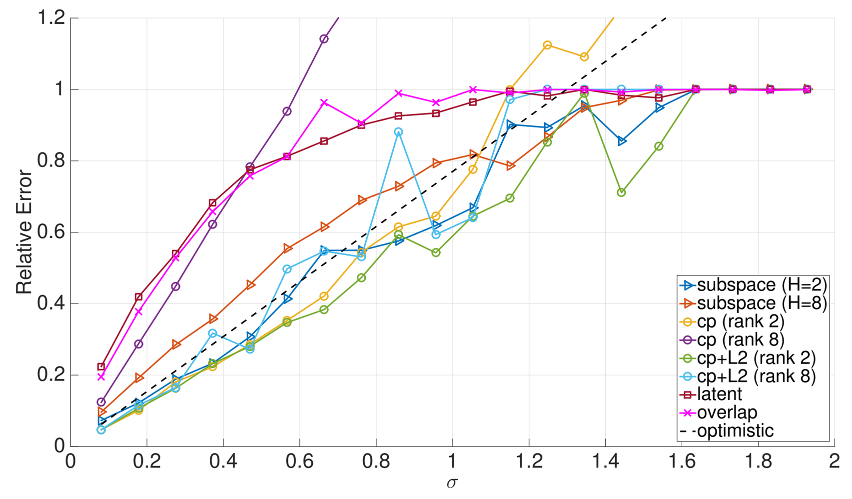
Synthetic Experiments: subspace norm

- Random CP tensor of size $20\times30\times40$ and rank 2.
- Added gaussian noise with standard deviation σ.



(L) Subspace norm with 3 representative values of λ. Magenta: theoretical motivated choice $\lambda = \sigma(\max_k(\sqrt{n_k}+\sqrt{H^{K-1}})+\sqrt{2\log K})$.

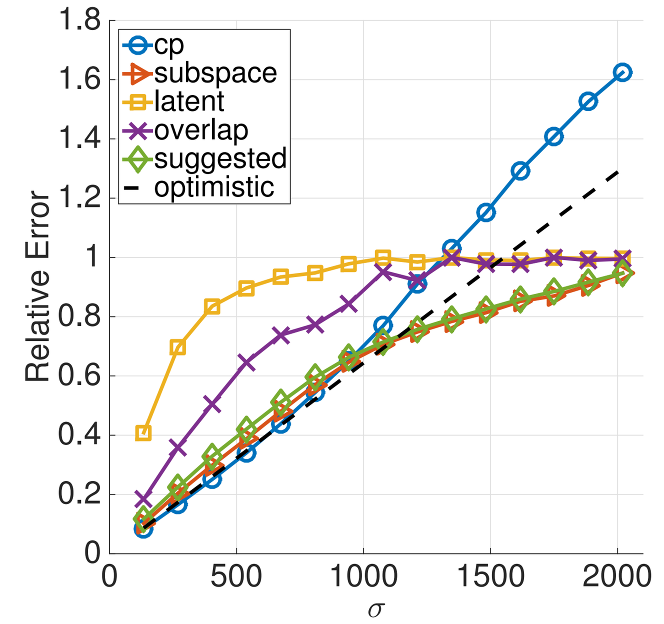
(R) The best results (across λ's or initializations) of different methods. The rank is correctly specified for CP and subspace norm.



(B) The rank is over-specified. ℓ_2 regularization is added to CP.

Subspace norm significantly outperforms overlap and latent norms, and performs almost equally with the CP.

Amino acids data



- Specturm of excitation wavelength and emission of five laboratory made samples, measured by fluorescence. See Bro 97 for details.
- $n_1 = 5$ (sample), $n_2 = 61$ (excitation wavelength), $n_3 = 201$ (emission).
- Samples contain different amounts of 3 types of acids, so that the data suits rank-3 CP model. The true rank is used for both CP and Subspace approach.

Contacts: Qinqing Zheng (qinqing@cs.uchicago.edu) Ryota Tomioka (tomioka@ttic.edu),