Convex Tensor Decomposition via Structured Schatten Norm Regularization

Ryota Tomioka
tomioka@ttic.edu
TTI Chicago

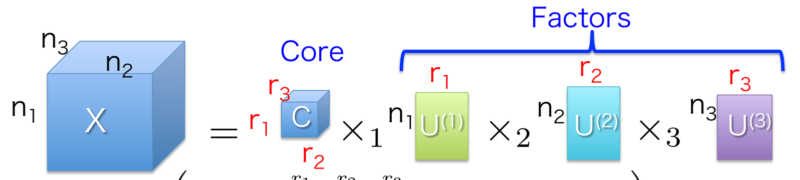
Taiji Suzuki
s-taiji@is.titech.ac.jp
Tokyo Institute of Technology

Summary

- We propose and analyze a convex-optimization-based algorithm for recovering a low-rank tensor (multi-way array) from noisy measurements.
- Our result shows that while the bound for the previously analyzed overlapped Schatten 1-norm depends on the average (square root) of the mode-k ranks, the bound for the new latent Schatten 1-norm depends on the minimal mode-k norm.
- Thus it performs almost as well as knowing the mode with the lowest rank.
- The bound agrees well with simulation.

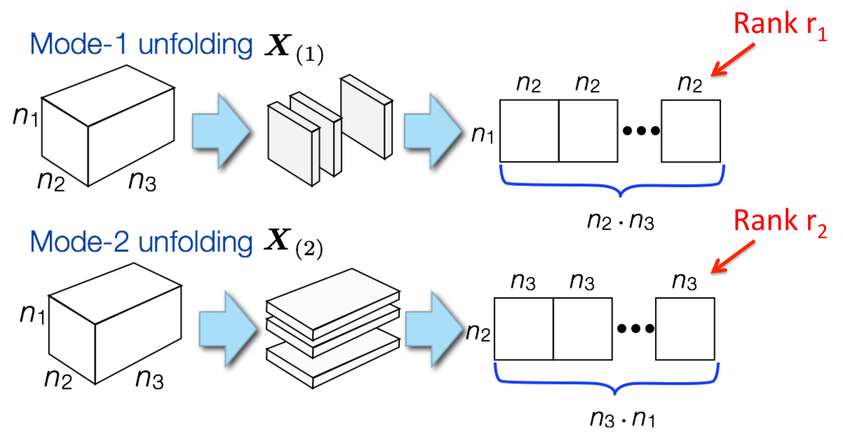
Tensor Decomposition

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.
- Used in neuroimaging, collaborative filtering, etc.

Schatten Norm (for matrices)

- Schatten p-norm:

$$\|\boldsymbol{X}\|_{S_p} = \left(\sum_{j=1}^{r} \sigma_j^p(\boldsymbol{X})\right)^{1/p},$$

where

$$\boldsymbol{X} = \boldsymbol{U}\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}\boldsymbol{V}^\top \qquad \text{(singular-value decomposition.)}$$

- $p = 1$: Nuclear norm / trace norm.
- $p \to \infty$: Spectral norm.

Overlapped Schatten Norms

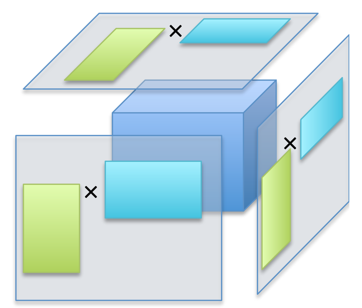
- Overlapped $S_1/1$-norm

$$|||\mathcal{W}|||_{\underline{S_1/1}} = \sum\nolimits_{k=1}^{K} \|\boldsymbol{W}_{(k)}\|_{S_1}.$$

- Studied by Gandy+11; Liu+09; Signoretto+10; T+10 for low-rank tensor recovery.

Intuition

Models a low-rank tensor as a simultaneously low rank object.



Property

The norm is bounded by the average (square root) mode-k ranks

$$|||\mathcal{W}|||_{\underline{S_1/1}} \leq \sum_{k=1}^{K} \sqrt{r_k}\, |||\mathcal{W}|||_F.$$

- More general: Overlapped S_p/q-norm

$$|||\mathcal{W}|||_{\underline{S_p/q}} = \left(\sum\nolimits_{k=1}^{K} \|\boldsymbol{W}_{(k)}\|_{S_p}^q\right)^{1/q}.$$

Bound for the Overlapped $S_1/1$-Norm [T+11]

Assume $\mathcal{W}^*$ has rank $(r_1, \ldots, r_K)$. Given observation

$$\mathcal{Y} = \mathcal{W}^* + \mathcal{E},$$

($\mathcal{E}$ is Gaussian $\mathcal{N}(0, \sigma^2)$ noise) any minimizer $\hat{\mathcal{W}}$of the convex problem

$$\underset{\mathcal{W}}{\text{minimize}} \quad \frac{1}{2}|||\mathcal{Y} - \mathcal{W}|||_F^2 + \lambda |||\mathcal{W}|||_{\underline{S_1/1}}$$

satisfies the bound

$$\frac{1}{N}|||\hat{\mathcal{W}} - \mathcal{W}^*|||_F^2 \leq c_1'\sigma^2 \left(\frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{n_k}}\right)^2 \left(\frac{1}{K}\sum_{k=1}^{K}\sqrt{r_k}\right)^2 = O\left(\frac{\frac{1}{K}\sum_{k=1}^{K} r_k}{n}\right). \qquad (1)$$

with high probability if

$$\lambda = \frac{c_0'\sigma}{K^2}\sum_{k=1}^{K}\left(\sqrt{n_k} + \sqrt{N/n_k}\right),$$

where $N = \prod_{k=1}^{K} n_k$.

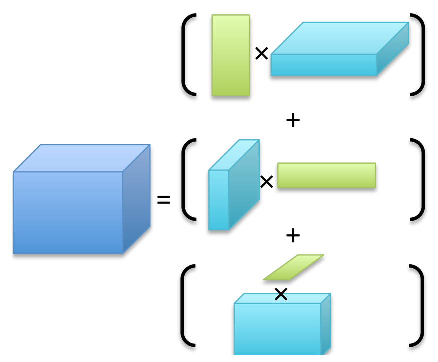
Latent Schatten Norms

- Latent $S_1/1$-norm

$$|||\mathcal{W}|||_{\overline{S_1/1}} = \inf_{(\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(K)})=\mathcal{W}} \sum_{k=1}^{K} \|\boldsymbol{W}^{(k)}_{(k)}\|_{S_1}.$$

Intuition

Models a low-rank tensor as a mixture of tensors that each are low-rank in a specific mode.



Property

The norm is bounded by the minimal mode-k rank

$$|||\mathcal{W}|||_{\overline{S_1/1}} \leq \left(\min_k \sqrt{r_k}\right) |||\mathcal{W}|||_F,$$

- More general: Latent S_p/q-norm

$$|||\mathcal{W}|||_{\overline{S_p/q}} = \inf_{(\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(K)})=\mathcal{W}} \left(\sum\nolimits_{k=1}^{K} \|\boldsymbol{W}^{(k)}_{(k)}\|_{S_p}^q\right)^{1/q}.$$

Bound for the Latent $S_1/1$-Norm (NEW)

Assume $\mathcal{W}^*$ has rank $(r_1, \ldots, r_K)$. Given observation

$$\mathcal{Y} = \mathcal{W} + \mathcal{E} = \sum_{k=1}^{K} \mathcal{W}^{*(k)} + \mathcal{E},$$

which satisfies $\|\boldsymbol{W}^{*(l)}_{(k)}\|_{S_\infty} \leq \alpha \quad (\forall l \neq k, k, l = 1, \ldots, K)$, any minimizer $\hat{\mathcal{W}}$ of the convex problem

$$\underset{\mathcal{W}}{\text{minimize}} \quad \frac{1}{2}|||\mathcal{Y} - \mathcal{W}|||_F^2 + \lambda |||\mathcal{W}|||_{\overline{S_1/1}}$$
$$\text{s.t.} \quad \mathcal{W} = \sum\nolimits_{k=1}^{K} \mathcal{W}^{(k)}, \ \|\boldsymbol{W}^{(l)}_{(k)}\|_{S_\infty} \leq \alpha, \quad \forall l \neq k$$

satisfies the bound

$$\frac{1}{N}|||\hat{\mathcal{W}} - \mathcal{W}^*|||_F^2 \leq c_1 K F \sigma^2 \frac{\min_k r_k}{\min_k n_k} = O\left(\frac{\min_k r_k}{n}\right), \qquad (2)$$

where $F = \left(\left(1 + \sqrt{\frac{n_1 n_K}{N}}\right) + \left(\sqrt{\log K} + \frac{\alpha(K-1)}{c_0\sigma}\right)\sqrt{\frac{n_K}{N}}\right)^2$, with high probability, if

$$\lambda = c_0\sigma(\sqrt{N/n_K} + \sqrt{n_1} + \sqrt{\log K}) + \alpha(K-1),$$

where we assume $n_1 \geq \cdots \geq n_K$ (without loss of generality).

Duality

Overlapped norms

$$|||\mathcal{W}|||_{\underline{S_1/1}} = \sum_{k=1}^{K} \|\boldsymbol{W}_{(k)}\|_{S_1}$$

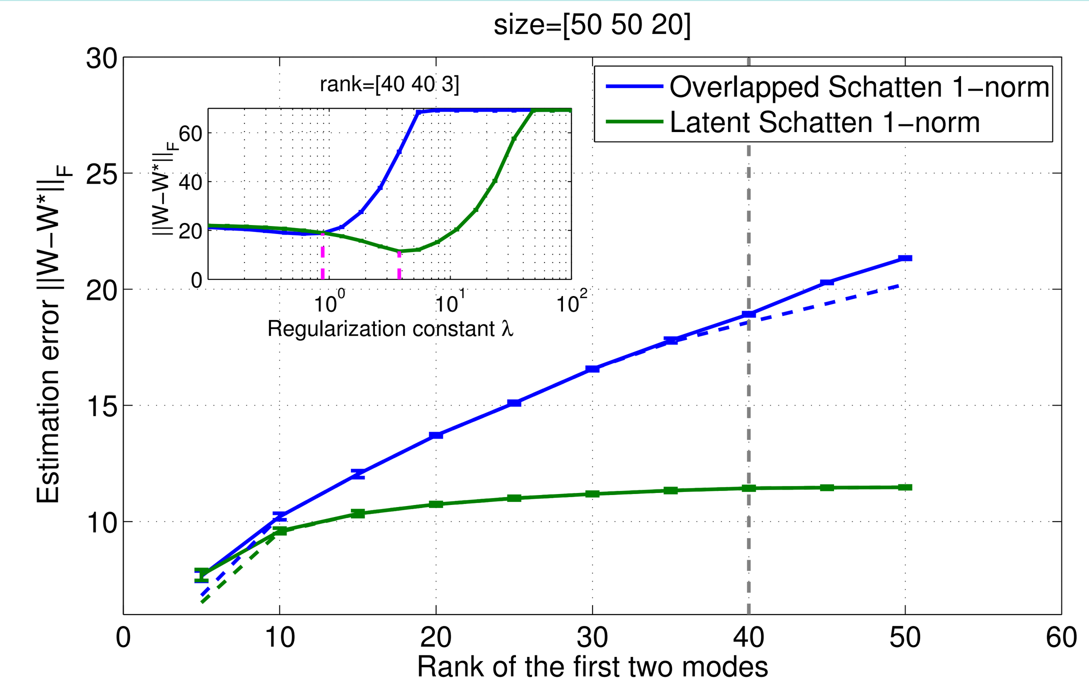
$$|||\mathcal{W}|||_{\underline{S_\infty/\infty}} = \max_k \|\boldsymbol{W}_{(k)}\|_{S_\infty}$$

Latent norms

$$|||\mathcal{W}|||_{\overline{S_\infty/\infty}} = \inf_{(\mathcal{W}^{(k)})_k} \max_k \|\boldsymbol{W}_{(k)}\|_{S_\infty} \quad \text{s.t.} \quad \mathcal{W}^{(1)} + \cdots + \mathcal{W}^{(K)} = \mathcal{W}$$

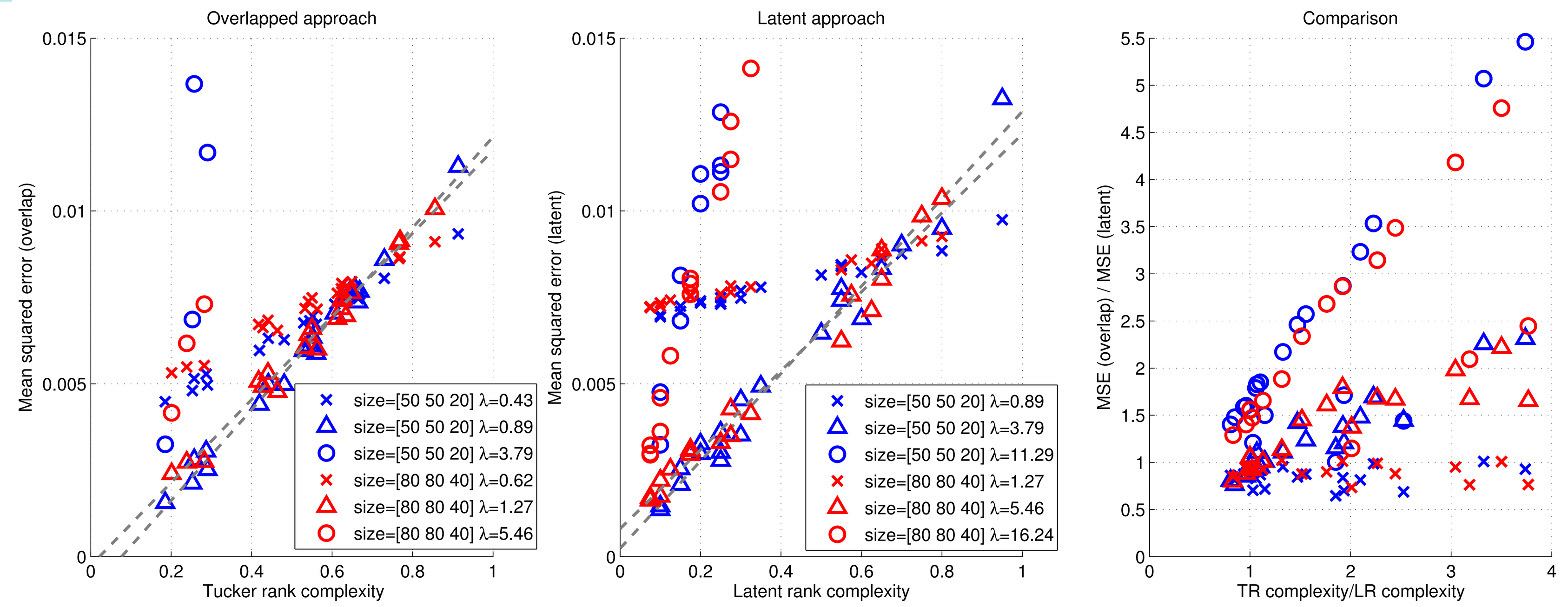
$$|||\mathcal{W}|||_{\overline{S_1/1}} = \inf_{(\mathcal{W}^{(k)})_k} \sum_{k=1}^{K} \|\boldsymbol{W}^{(k)}_{(k)}\|_{S_1} \quad \text{s.t.} \quad \mathcal{W}^{(1)} + \cdots + \mathcal{W}^{(K)} = \mathcal{W}$$

Experiment



- Goal: recover a synthetic rank $(r, r, 3)$ tensor of size $50 \times 50 \times 20$.
- Error measured by $|||\mathcal{W}^* - \hat{\mathcal{W}}|||_F$.

Numerical Comparison



- Randomly generated problems with various ranks and sizes.
- Tucker rank complexity = (1); Latent rank complexity = (2)
- Latent Schatten 1-norm performs better in most cases.

Future work

- Tensor completion.
- Optimality? – "statistical dimension" (Amelunxen, Lotz, McCoy, & Tropp, 2013) could be used as in Mu, Huang, Wright, & Goldfarb, 2013.
- Other norms and better relaxation/computation trade-off.