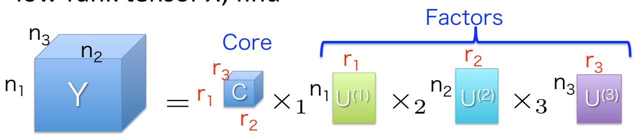
Statistical Performance of Convex Tensor Decomposition

Ryota Tomioka[1], Taiji Suzuki[1], Kohei Hayashi[2], Hisashi Kashima[1,3]

[1]The University of Tokyo, [2]Nara Institute of Science & Technology, [3]PRESTO, JST

Tucker decomposition [Tucker 66]

- Problem: Given a partially observed approximately low-rank tensor *X*, find



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

- Applications: chemo-/psycho-metrics, signal processing, computer vision, neuroscience
- Estimation: alternate minimization (non-convex)

Convex Tensor Estimation

Matrix

Estimation of *low-rank* matrix (hard) → Convex relaxation → Schatten 1-norm minimization (tractable) [Fazel, Hindi, Boyd 01]

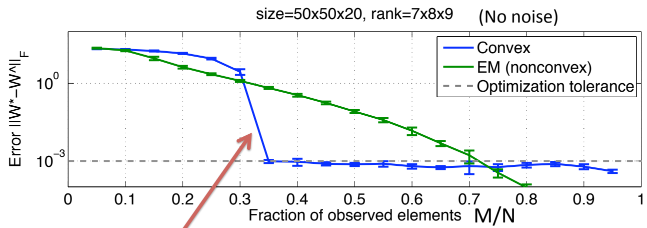
Generalize

Tensor

Estimation of *low-rank tensor* (hard) → Convex relaxation → Overlapped Schatten 1-norm minimization [Liu+09, Signoretto+10, Tomioka+10, Gandy+11]

Motivation: Phase-transition in Convex Tensor Estimation

Tensor completion result [Tomioka et al. 2010]



Goal: Explain this number of samples M from the size of the tensor [n1, n2, n3] and the Tucker rank [r1, r2, r3]

Model: Convex Tensor Estimation

Observation model $\mathcal{W}^*$ true tensor rank-(r_1,…,r_K)

$$y_i = \langle \mathcal{X}_i, \mathcal{W}^* \rangle + \epsilon_i \quad (i = 1, \ldots, M)$$

Gaussian noise $N(0, \sigma^2)$

Optimization

Empirical error, Regularization

$$\hat{\mathcal{W}} = \mathop{\mathrm{argmin}}_{\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}} \left(\frac{1}{2M} \|\boldsymbol{y} - \mathfrak{X}(\mathcal{W})\|_2^2 + \lambda_M \|\mathcal{W}\|_{S_1} \right)$$

$$(N = \textstyle\prod_{k=1}^{K} n_k)$$

Observation model $\mathfrak{X} : \mathbb{R}^N \to \mathbb{R}^M$

$\mathfrak{X}(\mathcal{W}) = (\langle \mathcal{X}_1, \mathcal{W} \rangle, \ldots, \langle \mathcal{X}_M, \mathcal{W} \rangle)^\top$

Reg. Const.

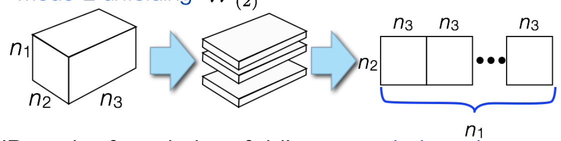
Overlapped Schatten 1-norm for Tensors

$$\left\|\mathcal{W}\right\|_{S_1} := \frac{1}{K} \sum_{k=1}^{K} \|\boldsymbol{W}_{(k)}\|_{S_1}$$

Schatten 1-norm for the mode-k unfolding

Example of unfolding (matricization)

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



NB: rank of mode-k unfolding = mode-k rank r_k

Previous work

Authors	Observation model	Assumption	Target
Recht, Fazel, Parrilo 2007	$y_i = \langle X_i, W \rangle$ $(i = 1, \ldots, M)$	Restricted Isometry	Matrix
Candès & Recht 2009	$Y_{ij} = W_{ij}$ $((i,j) \in \Omega)$	Incoherence	Matrix
Negahban & Wainwright 2011	$y_i = \langle X_i, W \rangle + \epsilon_i$ $(i = 1, \ldots, M)$	Restited Strong Convexity	Matrix
This work	$y_i = \langle X_i, W \rangle + \epsilon_i$ $(i = 1, \ldots, M)$	Restited Strong Convexity	Tensor

Restricted strong convexity (RSC)

(cf. Negahban & Wainwright 11)

- Assume that there is a positive constant $\kappa(\mathfrak{X})$ such that for all tensors $\Delta \in C$

$$\frac{1}{M}\|\mathfrak{X}(\Delta)\|_2^2 \geq \kappa(\mathfrak{X})\|\!|\Delta|\!\|_F^2$$

(The set C needs to be defined carefully)

Note:

- If $C=R^N$, $\kappa(X)=\min \mathrm{eig}(X^TX)$ $(X \in R^{M\times N})$
- When M<N, restriction is necessary.
- The smaller C, the weaker the assumption.

Lemma 1: A key inequality

$$\mathcal{W}, \mathcal{X} \in \mathbb{R}^{n_1\times\cdots\times n_K}$$

$$\langle \mathcal{W}, \mathcal{X}\rangle \leq \|\!|\mathcal{W}|\!\|_{S_1}\|\!|\mathcal{X}|\!\|_{\mathrm{mean}}$$

where

$$\|\!|\mathcal{W}|\!\|_{S_1} := \frac{1}{K}\sum_{k=1}^K \|\boldsymbol{W}_{(k)}\|_{S_1} \qquad \|\!|\mathcal{X}|\!\|_{\mathrm{mean}} := \frac{1}{K}\sum_{k=1}^K \|\boldsymbol{X}_{(k)}\|_{S_\infty}$$

K=2: norm duality (tight)
K>2: not tight

$$\|\boldsymbol{X}\|_{S_1} := \sum_{j=1}^m \sigma_j(\boldsymbol{X})$$

$$\|\boldsymbol{X}\|_{S_\infty} := \max_{j\in\{1,\dots,m\}} \sigma_j(\boldsymbol{X})$$

Theorem 1 (deterministic)

- Solution of the opt. problem $\hat{\mathcal{W}}$
- Reg const λ_M satisfies

$$\lambda_M \geq 2\|\!|\mathfrak{X}^*(\boldsymbol{\epsilon})|\!\|_{\mathrm{mean}}/M$$

where $\mathfrak{X}^*(\epsilon) = \sum_{i=1}^M \epsilon_i \mathcal{X}_i$ (noise design correlation)

$$\|\!|\mathcal{X}|\!\|_{\mathrm{mean}} := \frac{1}{K}\sum_{k=1}^K \|\boldsymbol{X}_{(k)}\|_{S_\infty}$$

- Under the RSC assumption

$$\|\!|\hat{\mathcal{W}} - \mathcal{W}^*|\!\|_F \leq \frac{32\lambda_M}{\kappa(\mathfrak{X})}\frac{1}{K}\sum_{k=1}^K \sqrt{r_k}$$

(cf. Negahban & Wainwright 11)

Two special cases

- Noisy tensor decomposition (M=N)
  - RSC: trivial. $\kappa(\mathfrak{X}) = 1/M$
  - bound on the noise-design correlation term

$$\mathbb{E}\|\!|\mathfrak{X}^*(\boldsymbol{\epsilon})|\!\|_{\mathrm{mean}} \leq \frac{\sigma}{K}\sum_{k=1}^K\left(\sqrt{n_k} + \sqrt{N/n_k}\right) \quad \text{(Lemma 3)}$$

- Random Gauss design
  - RSC: more difficult (Lemma 5)
  - bound on the noise-design correlation term

$$\mathbb{E}\|\!|\mathfrak{X}^*(\boldsymbol{\epsilon})|\!\|_{\mathrm{mean}} \leq \frac{\sigma\sqrt{M}}{K}\sum_{k=1}^K\left(\sqrt{n_k} + \sqrt{N/n_k}\right) \quad \text{(Lemma 4)}$$

Theorem 2 (noisy tensor decomp.)

When all the elements are observed (M=N) and the regularization const. satisfies

$$\lambda_M \geq c_0\sigma\sum_{k=1}^K\left(\sqrt{n_k} + \sqrt{N/n_k}\right)/(KN)$$

Then

$$\frac{\|\!|\hat{\mathcal{W}} - \mathcal{W}^*|\!\|_F^2}{N} \leq O_p\left(\sigma^2 \|\boldsymbol{n}^{-1}\|_{1/2}\|\boldsymbol{r}\|_{1/2}\right)$$

Normalized rank

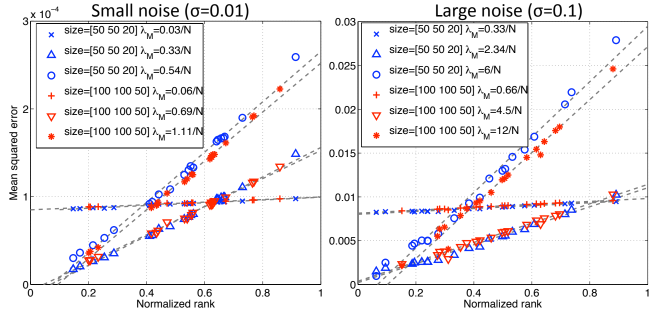
where

$$\|\boldsymbol{n}^{-1}\|_{1/2} := \left(\frac{1}{K}\sum_{k=1}^K\sqrt{1/n_k}\right)^2, \quad \|\boldsymbol{r}\|_{1/2} := \left(\frac{1}{K}\sum_{k=1}^K\sqrt{r_k}\right)^2$$

If n_k=n and r_k=r, normalized rank = r/n

Simulation: Noisy tensor decomposition

Mean squared error $\frac{\|\!|\hat{\mathcal{W}} - \mathcal{W}^*|\!\|_F^2}{N}$



linear relation between MSE and normalized rank!

Theorem 3: random Gauss design

Assume elements of X_i are drown iid from standard normal distribution. Moreover

$$\lambda_M \geq c_0\sigma\sum_{k=1}^K\left(\sqrt{n_k} + \sqrt{N/n_k}\right)/(K\sqrt{M}) \quad \text{and}$$

$$\frac{\#\text{samples }(M)}{\#\text{variables }(N)} \geq c_1\|\boldsymbol{n}^{-1}\|_{1/2}\|\boldsymbol{r}\|_{1/2} \approx \frac{r}{n}$$

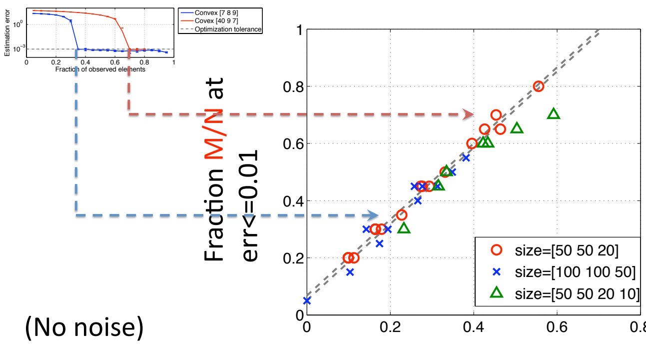
Normalized rank

Convergence!

$$\frac{\|\!|\hat{\mathcal{W}} - \mathcal{W}^*|\!\|_F^2}{N} \leq O_p\left(\frac{\sigma^2\|\boldsymbol{n}^{-1}\|_{1/2}\|\boldsymbol{r}\|_{1/2}}{M}\right)$$

$$\|\boldsymbol{n}^{-1}\|_{1/2} := \left(\frac{1}{K}\sum_{k=1}^K\sqrt{1/n_k}\right)^2, \quad \|\boldsymbol{r}\|_{1/2} := \left(\frac{1}{K}\sum_{k=1}^K\sqrt{r_k}\right)^2$$

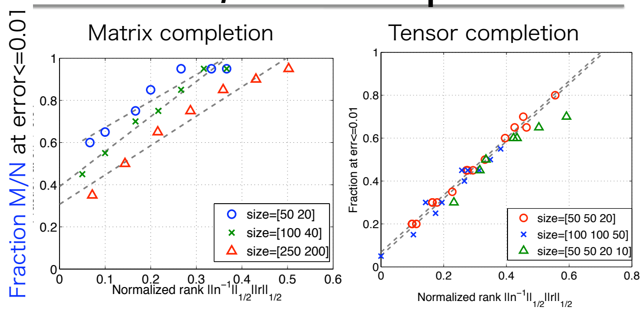
Simulation: Tensor Completion



(No noise)

Normalized rank $\|n^{-1}\|_{1/2}\|r\|_{1/2}$

Matrix / tensor completion



Tensor completion *easier* than matrix completion!?

Conclusion

- Convex tensor decomposition --- now with performance guarantee
- Normalized rank predicts empirical scaling behavior well

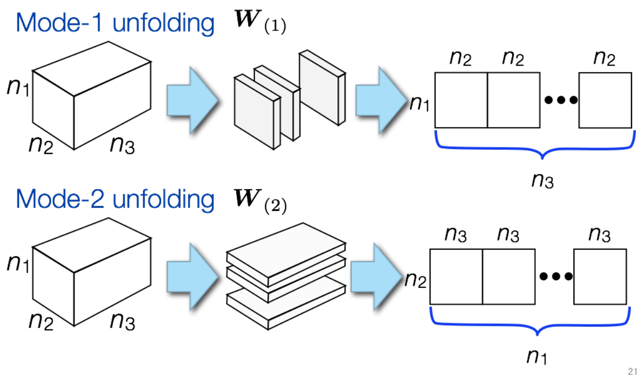
Issues

- Why matrix completion more difficult than tensor completion?
- Worst case analysis-> average case analysis
- Analyze tensor completion more carefully
  - Incoherence [Candes & Recht 09]
  - Spikiness [Negahban et al. 10]

References

- Candes & Recht (2009) Exact matrix completion via convex optimization. Found. Comput. Math., 9(6):717–772.
- Gandy, Recht, & Yamada (2011) Tensor completion and low-n-rank tensor recovery via convex optimization. Inverse Problems, 27:025010.
- Kolda & Bader (2009) Tensor decompositions and applications. SIAM Review, 51(3):455–500.
- Liu, Musialski, Wonka, & Ye. (2009) Tensor completion for estimating missing values in visual data. In Prof. ICCV.
- Negahban & Wainwright (2011) Estimation of (near) low-rank matrices with noise and high-dimensional scaling. Ann. Statist., 39(2).
- Recht, Fazel, & Parrilo (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review, 52(3):471–501.
- Signoretto, de Lathauwer, & Suykens (2010) Nuclear norms for tensors and their use for convex multilinear estimation. Tech Report 10-186, K.U.Leuven.
- Tomioka, Hayashi, & Kashima (2010) On the extension of trace norm to tensors. In NIPS2010 Workshop: Tensors, Kernels and Machine Learning.
- Tomioka, Hayashi, & Kashima (2011) Estimation of low-rank tensors via convex optimization. Technical report, arXiv:1010.0789, 2011.
- Tucker (1966) Some mathematical notes on three-mode factor analysis. Psychometrika, 31(3): 279–311.

Mode-k unfolding (matricization)



Choosing the set C

- We only need the residual Δ to be in C

$$\Delta_{(k)} = \Delta'_k + \Delta''_k$$

mode-k unfolding of the residual = Component spanned by the truth + Orthogonal to the truth

Lemma 2. *Let $\hat{\mathcal{W}}$ be the solution of the minimization problem (7) with $\lambda_M \geq 2\|\mathfrak{X}^*(\boldsymbol{\epsilon})\|_{\mathrm{mean}}/M$, and let $\Delta := \hat{\mathcal{W}} - \mathcal{W}^*$, where $\mathcal{W}^*$ is the true low-rank tensor. Let $\boldsymbol{\Delta}_{(k)} = \boldsymbol{\Delta}'_k + \boldsymbol{\Delta}''_k$ be the decomposition defined in Equation (4). Then for all $k = 1, \ldots, K$ we have the following inequalities:*

1. $\mathrm{rank}(\boldsymbol{\Delta}'_k) \leq 2r_k$.
2. $\sum_{k=1}^{K} \|\boldsymbol{\Delta}''_k\|_{S_1} \leq 3\sum_{k=1}^{K} \|\boldsymbol{\Delta}'_k\|_{S_1}$.

Lemma 5 (RSC for random Gaussian)

Let $\mathfrak{X} : \mathbb{R}^{n_1 \times \cdots \times n_K} \to \mathbb{R}^M$

be a random Gaussisan design. Then

$$\frac{\|\mathfrak{X}(\Delta)\|_2}{\sqrt{M}} \geq \frac{1}{4}|||\Delta|||_F - \frac{1}{K}\sum_{k=1}^{K}\left(\sqrt{\frac{n_k}{M}} + \sqrt{\frac{\bar{n}_{\setminus k}}{M}}\right)|||\Delta|||_{S_1},$$

with probability at least $1 - 2\exp(-N/32)$

Proof: analogous to that of Prop 1 in Negahban & Wainwright 2011 (use Lemma 1)