**Performance of 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

**Convex Tensor Estimation**

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

- **Tensor**
  - Estimation of low-rank tensor (hard)
  - Schatten 1-norm minimization ([Liu09, Signoretto+10, Tomioka+10, Gandy+11])

**Motivation: Phase-transition in Convex Tensor Estimation**

- Tensor completion result [Tomioka et al. 2010]

**Previous work**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Observation model</th>
<th>Assumption</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recht, Fazel, Parrilo 2007</td>
<td>( y_i = \langle X_i, W \rangle ) ((i = 1, \ldots, M))</td>
<td>Restricted Isometry</td>
<td>Matrix</td>
</tr>
<tr>
<td>Candès &amp; Recht 2009</td>
<td>( Y_{ij} = W_{ij} ) ((i, j) \in \Omega)</td>
<td>Incoherence</td>
<td>Matrix</td>
</tr>
<tr>
<td>Negahban &amp; Wainwright 2011</td>
<td>( y_i = \langle X_i, W \rangle + \epsilon_i ) ((i = 1, \ldots, M))</td>
<td>Restricted Strong Convexity</td>
<td>Matrix</td>
</tr>
<tr>
<td>This work</td>
<td>( y_i = \langle X_i, W \rangle + \epsilon_i ) ((i = 1, \ldots, M))</td>
<td>Restricted Strong Convexity</td>
<td>Tensor</td>
</tr>
</tbody>
</table>

**Goal:**

- Explain this number of samples \( M \) from the size of the tensor \([n_1, n_2, n_3]\) and the Tucker rank \([r_1, r_2, r_3]\)
Restricted strong convexity (RSC)

(cf. Negahban & Wainwright 11)

- Assume that there is a positive constant $\kappa(x)$ such that for all tensors $x \in C$

$$\frac{1}{M} \| x(\Delta) \|^2 \geq \kappa(x) \| \Delta \|^2$$

(The set $C$ needs to be defined carefully)

Note:
- If $C = R^N$, $\kappa(x) = \min\text{eig}([X^T X])$ ($X \in R^{p \times m}$)
- When $M < N$, restriction is necessary.
- The smaller $C$, the weaker the assumption.

Two special cases

- Noisy tensor decomposition ($M = N$)
  - RSC: trivial, $\kappa(x) = 1/M$
  - bound on the noise-design correlation term $E \| [X'(x)]_\text{mean} \| \leq \frac{1}{K} \sum_{k=1}^K (\sqrt{n_k} + \sqrt{N/n_k})$ (Lemma 3)

- Random Gauss design
  - RSC: more difficult (Lemma 5)
  - bound on the noise-design correlation term $E \| [X'(x)]_\text{mean} \| \leq \frac{\sqrt{M}}{K} \sum_{k=1}^K (\sqrt{n_k} + \sqrt{N/n_k})$ (Lemma 4)

Theorem 2 (noisy tensor decompp.)

When all the elements are observed ($M = N$) and the regularization const. satisfies $\lambda_M \geq \epsilon_0 \sigma \sum_{k=1}^K \left( \sqrt{n_k} + \sqrt{N/n_k} \right) / (KN)$

Then

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

where

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

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

Theorem 1 (deterministic)

- Solution of the opt. problem $\hat{W}$
- Reg const $\lambda_M$ satisfies $\lambda_M \geq 2 \| \hat{X}^\star(\epsilon) \|_\text{mean} / M$

where $\hat{X}^\star(\epsilon) = \sum_{i=1}^M t_i x_i$ (noise design correlation)

$$\| \hat{X}^\star(\epsilon) \|_\text{mean} := \frac{1}{K} \sum_{k=1}^K \| x_k \|_{S_m}$$

(inside the RSC assumptions)

$$\| \hat{W} - W^\star \|_F \leq \frac{32 \lambda_M}{\kappa(x)} \frac{1}{K} \sum_{k=1}^K \sqrt{r_k}$$

(c.f. Negahban & Wainwright 11)

Simulation: Noisy tensor decomposition

Mean squared error $\| \hat{W} - W^\star \|_F^2$

Linear relation between MSE and normalized rank!

Simulation: Tensor Completion

Matrix / tensor completion

Matrix 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]

Choosing the set C

\[
\Delta_{(k)} = \Delta'_{(k)} + \Delta''_{(k)}
\]

mode-k unfolding of the residual

Component spanned by the truth

Orthogonal to the truth

\[
\text{Lemma 2. Let } \hat{W} \text{ be the solution of the minimization problem (2) with } \lambda_M \geq 2\|\Sigma(x)\|_{\max}/M,
\text{ and let } \Delta := \hat{W} - W^*, \text{ where } W^* \text{ is the true low-rank tensor. Let } \Delta_{(k)} = \Delta'_{(k)} + \Delta''_{(k)} \text{ be the decomposition defined in Equation (4). Then for all } k = 1, \ldots, K \text{ we have the following inequalities:}
\]

1. \( \text{rank}(\Delta_{(k)}) \leq K \)
2. \( \sum_{k=1}^{K} \| \Delta_{(k)} \|_F \leq \sum_{k=1}^{K} \| \Delta_{(k)} \|_F 

\]

References


Lemma 5 (RSC for random Gaussian)

Let \( X : \mathbb{R}^{n_1 \times \cdots \times n_K} \rightarrow \mathbb{R}^M \)
be a random Gaussian design. Then

\[
\| \hat{X}(\Delta) \|_F \geq \frac{1}{4} \left\| \Delta \right\|_F - \frac{1}{K} \sum_{k=1}^{K} \left( \sqrt{\frac{n_k}{M} + \frac{n_k \lambda_{\max}}{M}} \right) \| \Delta \|_{S_k},
\]

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

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