Matrix vs Tensor Denoising Methods under Block Sparse Perturbations

Animashree Anandkumar, Prateek Jain, Yang Shi, U.N. Niranjan
EECS, UC Irvine; Microsoft Research, India; EECS, UC Irvine; ICS, UC Irvine

Highlights

- Robust tensor CP decomposition: separate a tensor into low-rank and sparse components.
- Proposed method: alternating projections between power method and hard thresholding.
- Global convergence guarantees: under incoherence and bounded sparsity assumption.
- Improvements: can handle more gross corruptions than matrix-based methods.

Introduction

Figure 1: Robust tensor decomposition problem: decomposing a given tensor into low-rank and sparse components.

Problem: Let $T, L, S \in \mathbb{R}^{n \times n \times n}$. Wlog, $\sigma^*_i > 0$. Given $T$, find $L, S$ such that:

$$T = L + S, \quad L = \frac{1}{n^2} \sigma^*_i u_i^{(1)} \otimes u_i^{(2)} \otimes u_i^{(3)} \quad \|S\|_0 \leq d$$

such that $L$ is a CP-rank $r$ orthogonal tensor, i.e., $(u_i, v_i) = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and 0, else.

- $(L)$: $L$ is $\mu$-incoherent, i.e., $\|u_i\| \leq \frac{\sqrt{d}}{\mu}$.
- $(S)$: Random block sparsity pattern with $B$ blocks of size $d$ and overlap fraction $\eta$, i.e.,

$$\psi = \frac{B}{r} \otimes \psi_i \otimes \psi_i, \quad \|\psi_i\|_0 \leq d, \quad \psi_i(j) = 0 \quad \text{or} \quad 1$$

$\eta = O((\mu r)^{1/3}, B = O(min(r^{1/3}, r^{1/2}, \eta^{-1/4}))$.

Under this model, the support tensor $\Psi$ which encodes sparsity pattern, has rank $B$.

Tensor vs Matrix Method

- Under random block sparsity,

$$\frac{d_{\text{on}}}{d_{\text{matrix}}} = \frac{\Omega(n^{1/3} \eta^{1/3})}{\eta^{1/2}} \cdot O(1)$$

- Thus, we can handle more gross corruptions than matrix methods.

Alternating Projection Algorithm

1. for Stage $l = 1$ to $r$
2. repeat
3. $L^{(l+1)} = P(T - S^{(l)})$ (use power method and gradient ascent to compute the eigenvectors).
4. $S^{(l+1)} = \mathcal{H}_i(T - L^{(l+1)})$ (hard thresholding of grossly corrupted entries).
5. until Convergence
6. end for

Gradient Ascent Algorithm

1. **Power method**\cite{1} to land in spectral ball of sparse tensor: $v_i^{(l)} \leftarrow T_i(u_i^{(l)} \otimes v_i^{(l)}) / \|T_i(u_i^{(l)} \otimes v_i^{(l)})\|_2$.
2. **Gradient ascent** iterations to compute tensor eigenvectors: $v_i^{(l+1)} \leftarrow v_i^{(l)} + \frac{1}{\lambda_i} E_i(u_i^{(l)} \otimes v_i^{(l)})$.
3. Deflation to obtain all leading components: $T_i \leftarrow T_i - \lambda_i u_i \otimes u_i \otimes u_i$.

Theorem (Convergence to Global Optimum)

Let $L^*, S^*$ satisfy (L) and (S). The outputs $L$ (and its parameters $\hat{u}_i$ and $\hat{\lambda}_i$) and $S$ of the alternating projection algorithm satisfy w.h.p.:

$$\|\hat{u}_i - u_i\|_\infty \leq \frac{\delta}{\mu^2 r^{1/2}} \sigma^*_i, \quad \|\hat{\lambda}_i - \lambda_i\| \leq \delta, \quad \forall i \in [n]$$

$$\|L - L^*\|_F \leq \delta, \quad \|S - S^*\|_\infty \leq \delta/n^{1/2}, \quad \text{and sup} \sup S \subseteq \sup S^*.$$

Error Plots

- Figure 2: (a),(b) Error with deterministic sparsity. (c),(d) Error with block sparsity.

Time Plots

- Figure 3: (a),(b) Running time with deterministic sparsity. (c),(d) Running time with block sparsity.

Figure 4: Foreground filtering in the Curtain video dataset. (a): Original image frame. (b): Foreground filtered using tensor method; time taken is 5.1s. (c): Foreground filtered using matrix method; time taken is 5.7s.

Proof Outline

- The inductive proof for alternating projections is along the similar lines of \cite{2}.
- Assuming $T = L^* + S^* L^* = u \otimes u \otimes u$
- Update $L_{t+1} = T - S_t = L^* + S^* - S_t = L^* + E_t$
- Consider tensor fixed point equation $L_{t+1}^* = \mathcal{P}_L L_{t+1}$ and apply perturbation arguments from \cite{1}.
- Goal is to prove $\|L_{t+1} - L^*\|_\infty \leq \epsilon \|L_{t} - L^*\|_\infty$.
- Prove $\|L_{t+1} - L^*\|_\infty \leq \epsilon \|E_t\|_\infty$ using inductive assumption $\|E_t\|_\infty \leq \epsilon \|L_{t} - L^*\|_\infty$ and then prove $\|E_{t+1}\|_\infty \leq \epsilon \|E_{t+1}\|_\infty$. Piecing these together we obtain the full proof.
- The challenge in the tensor case is to prove the validity of the tensor fixed point equation.

References