The Complexity of Sparse Tensor PCA

Davin Choo ¹  Tommaso D’Orsi ²

¹National University of Singapore

²ETH Zürich
Motivating example

Figure: $x = \left[ \frac{1}{\sqrt{5}}, -\frac{1}{\sqrt{5}}, 0, \ldots, 0, -\frac{1}{\sqrt{5}}, 0, \ldots, 0, -\frac{1}{\sqrt{5}}, 0, \frac{1}{\sqrt{5}}, 0, \ldots, 0 \right]^{\top}$.

Supp($x$) = \{1, 2, 9, 14, 16\}. Noise $W \in \mathbb{R}^{20 \times 20}$ has i.i.d. $N(0, 1)$ entries.
Motivating example

\[ Y = W + \lambda xx^T \]

\[ Y = W + 16xx^T \]

\[ Y = W + 32xx^T \]

**Figure:** Observation \( Y = W + \lambda xx^T \) under different signal strengths \( \lambda \in \{1, 16, 32\} \). Colours are rescaled to emphasize relative values.
Sparse tensor PCA model (Simplified)

Observe: Tensor $Y = W + \lambda x^\otimes p$, for $p \geq 2$
- $W$ is order $p$ tensor with i.i.d. $N(0,1)$ entries
- Signal $x$ is flat, $k$-sparse, unit length

Approximate recovery: Find $\hat{x}$ such that $|\langle x, \hat{x} \rangle| \gg 1 - o(1)$
Observe: Tensor $Y = W + \lambda x \otimes^p$, for $p \geq 2$

- $W$ is order $p$ tensor with i.i.d. $N(0, 1)$ entries
- Signal $x$ is flat, $k$-sparse, unit length

Approximate recovery: Find $\hat{x}$ such that $|\langle x, \hat{x} \rangle| \gg 1 - o(1)$

Extensions (briefly discussed later)

- Approximately flat signals
- Multiple spikes
- General tensor spikes

Remark: See paper for references (no citations for cleaner slides)
Sparse (Wigner) PCA: $p = 2, \ k \leq n$

- Observe tensor $Y = W + \lambda xx^\top \in \mathbb{R}^{n \times n}$
- $x$ is $k$-sparse ($|\text{supp}(x)| = k$) and unit length ($\|x\|_2 = 1$)
Sparse (Wigner) PCA: $p = 2, k \leq n$

- Observe tensor $\mathbf{Y} = \mathbf{W} + \lambda \mathbf{x} \mathbf{x}^\top \in \mathbb{R}^{n \times n}$
- $\mathbf{x}$ is $k$-sparse ($|\text{supp}(\mathbf{x})| = k$) and unit length ($\|\mathbf{x}\|_2 = 1$)
Sparse (Wigner) PCA: $p = 2, k \leq n$

- Observe tensor $\mathbf{Y} = \mathbf{W} + \lambda \mathbf{x} \mathbf{x}^T \in \mathbb{R}^{n \times n}$
- $\mathbf{x}$ is $k$-sparse ($|\text{supp}(\mathbf{x})| = k$) and unit length ($\|\mathbf{x}\|_2 = 1$)
Sparse (Wigner) PCA: \( p = 2, \ k \leq n \)

- Observe tensor \( Y = W + \lambda xx^\top \in \mathbb{R}^{n \times n} \)
- \( x \) is \( k \)-sparse (\( |\text{supp}(x)| = k \)) and unit length (\( \|x\|_2 = 1 \))
Sparse (Wigner) PCA: \( p = 2, \ k \leq n \)

- Observe tensor \( Y = W + \lambda xx^T \in \mathbb{R}^{n \times n} \)
- \( x \) is \( k \)-sparse (\( |\text{supp}(x)| = k \)) and unit length (\( \|x\|_2 = 1 \))

- \( \sqrt{k \log \frac{n}{k}} \quad k \sqrt{\log \frac{n}{k^2}} \quad k \sqrt{\log n} \quad \sqrt{n} \)

- Efficient algorithms when \( \sqrt{k} \ll \lambda \ll \min\{k, \sqrt{n}\} \)?
Sparse (Wigner) PCA: $p = 2, \ k \leq n$

- Observe tensor $\mathbf{Y} = \mathbf{W} + \lambda \mathbf{x} \mathbf{x}^\top \in \mathbb{R}^{n \times n}$
- $\mathbf{x}$ is $k$-sparse ($|\text{supp}(\mathbf{x})| = k$) and unit length ($\|\mathbf{x}\|_2 = 1$)

Efficient algorithms when $\sqrt{k} \ll \lambda \ll \min\{k, \sqrt{n}\}$?

Unlikely due to Sum-of-Squares (SoS) lower bound
Tensor PCA: $p \geq 2$, $k = n$

- Observe $Y = W + \lambda x^{\otimes p} \in \mathbb{R}^{n \otimes p}$
- Signal $x$ is unit length ($\|x\|_2 = 1$)
- Remark: Computing $\max_{\|x\|=1} \langle Y, x^{\otimes p} \rangle$ is NP-hard for $p \geq 3$
Tensor PCA: $p \geq 2$, $k = n$

- Observe $Y = W + \lambda x^\otimes p \in \mathbb{R}^{n \times p}$
- Signal $x$ is unit length ($\|x\|_2 = 1$)
- Remark: Computing $\max_{\|x\|=1} \langle Y, x^\otimes p \rangle$ is NP-hard for $p \geq 3$
Tensor PCA: \( p \geq 2, \ k = n \)

- Observe \( Y = W + \lambda x \otimes^p \in \mathbb{R}^{n \otimes^p} \)
- Signal \( x \) is unit length (\( \| x \|_2 = 1 \))
- Remark: Computing \( \max_{\| x \| = 1} \langle Y, x \otimes^p \rangle \) is NP-hard for \( p \geq 3 \)
Tensor PCA: $p \geq 2, \ k = n$

- Observe $Y = W + \lambda x^{\otimes p} \in \mathbb{R}^{n \otimes p}$
- Signal $x$ is unit length ($\|x\|_2 = 1$)
- Remark: Computing $\max_{\|x\|=1} \langle Y, x^{\otimes p} \rangle$ is NP-hard for $p \geq 3$

- Efficient algorithms when $\sqrt{n} \ll \lambda \ll n^{p/4}$?
Tensor PCA: $p \geq 2$, $k = n$

- Observe $Y = W + \lambda x \otimes p \in \mathbb{R}^{n \times p}$
- Signal $x$ is unit length ($\|x\|_2 = 1$)
- Remark: Computing $\max_{\|x\|=1} \langle Y, x \otimes p \rangle$ is NP-hard for $p \geq 3$

Information-theoretically impossible

Exhaustive search works

SoS algorithm

Tensor unfolding

Efficient algorithms when $\sqrt{n} \ll \lambda \ll n^{p/4}$?

Unlikely due to Sum-of-Squares (SoS) lower bound
Setup

**Sparse tensor PCA (single spike)**

Observe tensor \( Y = W + \lambda x \otimes p \)

- \( W \) is order \( p \) tensor with i.i.d. \( N(0, 1) \) entries
- Signal \( x \) is flat, \( k \)-sparse, unit length

Goal: Find \( \hat{x} \) such that \( |\langle x, \hat{x} \rangle| \gg 1 - o(1) \)
Sparse tensor PCA (single spike)

Observe tensor $Y = W + \lambda x \otimes p$

- $W$ is order $p$ tensor with i.i.d. $N(0, 1)$ entries
- Signal $x$ is flat, $k$-sparse, unit length

Goal: Find $\hat{x}$ such that $|\langle x, \hat{x} \rangle| \gg 1 - o(1)$

Simplifying assumptions

- We will ignore some technical preprocessing steps
- We will briefly discuss how we handle some extensions at the end such as the case where there are multiple planted signals
A parametric recovery algorithm

Observe tensor \( Y = W + \lambda x \otimes^p \); \( x \) is \( k \)-sparse, flat and unit length

**Recovery algorithm**

Let \( 1 \leq t \leq k \) be a computational parameter. Suppose

\[
\lambda \gtrsim \sqrt{t \left( \frac{k}{t} \right)^p \log n}.
\]

Then, there exists an algorithm that runs in \( O(pn^{p+t}) \) time and, with probability 0.99, outputs the support of \( x \).
A parametric recovery algorithm

Observe tensor $Y = W + \lambda x \otimes^p$; $x$ is $k$-sparse, flat and unit length

Recovery algorithm
Let $1 \leq t \leq k$ be a computational parameter. Suppose

$$\lambda \gtrsim \sqrt{t \left( \frac{k}{t} \right)^p \log n}.$$ 

Then, there exists an algorithm that runs in $O(pn^{p+t})$ time and, with probability 0.99, outputs the support of $x$.

Extreme values of $t$

- $t = 1 \Rightarrow$ Polynomial running time, $\lambda \gtrsim \sqrt{k^p \log n}$
- $t = k \Rightarrow$ Exponential running time, $\lambda \gtrsim \sqrt{k \log n}$
A parametric recovery algorithm

Observe tensor $Y = W + \lambda x \otimes^p$; $x$ is $k$-sparse, flat and unit length

**Recovery algorithm**

Let $1 \leq t \leq k$ be a computational parameter. Suppose

$$\lambda \gtrsim \sqrt{t \left( \frac{k}{t} \right)^p \log n}.$$ 

Then, there exists an algorithm that runs in $O(pn^{p+t})$ time and, with probability 0.99, outputs the support of $x$.

**Extreme values of $t$**

- $t = 1 \Rightarrow$ Polynomial running time, $\lambda \gtrsim \sqrt{k^p \log n}$
- $t = k \Rightarrow$ Exponential running time, $\lambda \gtrsim \sqrt{k \log n}$

Remark: To obtain $\hat{x}$, run an existing tensor PCA algorithm on the smaller subtensor defined by the support of $x$. 
Limited brute force

- Define $U_t = \left\{ u \in \left\{ -\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}} \right\} : |\text{supp}(u)| = t \right\}$
Limited brute force

- Define \( U_t = \left\{ u \in \left\{ -\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}} \right\} : |\text{supp}(u)| = t \right\} \)
- For any \( u \in U_t \),
  - Since \( Y = W + \lambda x^\otimes p \), \( \langle Y, u^\otimes p \rangle = \langle W, u^\otimes p \rangle + \lambda \langle u, x \rangle^p \)
Limited brute force

- Define $U_t = \left\{ u \in \left\{-\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}} \right\} : |\text{supp}(u)| = t \right\}$
- For any $u \in U_t$,
  - Since $Y = W + \lambda x \otimes^p$, $\langle Y, u \otimes^p \rangle = \langle W, u \otimes^p \rangle + \lambda \langle u, x \rangle^p$
  - Since $W$ is Gaussian tensor, $|\langle W, u \otimes^p \rangle| \leq O(\sqrt{t \log n})$
Limited brute force

- Define \( U_t = \{ u \in \{-\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}}\} : |\text{supp}(u)| = t \} \)
- For any \( u \in U_t \),
  - Since \( Y = W + \lambda x \otimes^p \), \( \langle Y, u \otimes^p \rangle = \langle W, u \otimes^p \rangle + \lambda \langle u, x \rangle^p \)
  - Since \( W \) is Gaussian tensor, \(|\langle W, u \otimes^p \rangle| \leq O(\sqrt{t \log n}) \)
  - If \(|\text{supp}(u) \cap \text{supp}(x)| = t \), then
    \[
    \langle Y, u \otimes^p \rangle \geq \lambda \cdot \left(\frac{t}{k}\right)^{\frac{p}{2}} - O(\sqrt{t \log n})
    \]
Limited brute force

- Define $U_t = \left\{ u \in \left\{ -\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}} \right\} : |\text{supp}(u)| = t \right\}$

- For any $u \in U_t$,
  - Since $Y = W + \lambda x \otimes p$, $\langle Y, u \otimes p \rangle = \langle W, u \otimes p \rangle + \lambda \langle u, x \rangle^p$
  - Since $W$ is Gaussian tensor, $|\langle W, u \otimes p \rangle| \leq O(\sqrt{t \log n})$
  - If $|\text{supp}(u) \cap \text{supp}(x)| = t$, then
    \[
    \langle Y, u \otimes p \rangle \geq \lambda \cdot \left( \frac{t}{k} \right)^{\frac{p}{2}} - O(\sqrt{t \log n})
    \]
  - If $|\text{supp}(u) \cap \text{supp}(x)| < (1 - \epsilon) \cdot t$, then
    \[
    \langle Y, u \otimes p \rangle \leq \lambda \cdot (1 - \epsilon)^p \cdot \left( \frac{t}{k} \right)^{\frac{p}{2}} + O(\sqrt{t \log n})
    \]
**Limited brute force**

- Define $U_t = \left\{ u \in \left\{-\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}}\right\} : |\text{supp}(u)| = t \right\}$

- For any $u \in U_t$,
  - Since $Y = W + \lambda x \otimes^p$, $\langle Y, u \otimes^p \rangle = \langle W, u \otimes^p \rangle + \lambda \langle u, x \rangle^p$
  - Since $W$ is Gaussian tensor, $|\langle W, u \otimes^p \rangle| \leq O(\sqrt{t \log n})$
  - If $|\text{supp}(u) \cap \text{supp}(x)| = t$, then
    $$\langle Y, u \otimes^p \rangle \geq \lambda \cdot \left( \frac{t}{k} \right)^{\frac{p}{2}} - O(\sqrt{t \log n})$$
  - If $|\text{supp}(u) \cap \text{supp}(x)| < (1 - \epsilon) \cdot t$, then
    $$\langle Y, u \otimes^p \rangle \leq \lambda \cdot (1 - \epsilon)^p \cdot \left( \frac{t}{k} \right)^{\frac{p}{2}} + O(\sqrt{t \log n})$$

- For $v_* = \arg\max_{u \in U_t} \langle Y, u \otimes^p \rangle$,
  $$|\text{supp}(v_*) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t$$
Limited brute force

- Define $U_t = \left\{ u \in \left\{ -\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}} \right\} : |\text{supp}(u)| = t \right\}$

- For $v_* = \arg\max_{u \in U_t} \langle Y, u^p \rangle$,
  $$|\text{supp}(v_*) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t$$
Limited brute force

- Define $U_t = \left\{ u \in \left\{-\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}}\right\} : |\text{supp}(u)| = t \right\}$
- For $v_* = \arg\max_{u \in U_t} \langle Y, u^\otimes p \rangle$,
  $$|\text{supp}(v_*) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t$$
- Define $\alpha \in \mathbb{R}^n$ with entries $\alpha_\ell := \langle Y, v_*^{p-1} \otimes e_\ell \rangle$, $\forall \ell \in [n]$
Limited brute force

- Define $U_t = \left\{ u \in \{-\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}}\} : |\text{supp}(u)| = t \right\}$
- For $v_* = \arg\max_{u \in U_t} \langle Y, u^\otimes p \rangle$,
  $$|\text{supp}(v_*) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t$$
- Define $\alpha \in \mathbb{R}^n$ with entries $\alpha_\ell := \langle Y, v_*^{\otimes p-1} \otimes e_\ell \rangle$, $\forall \ell \in [n]$
- For all $\ell \in \text{supp}(x)$,
  $$\alpha_\ell = \lambda \cdot x_\ell \cdot \langle x, v_* \rangle^{p-1} + \langle W, v_*^{\otimes p-1} \otimes e_\ell \rangle \gtrsim \sqrt{\log n}$$
Limited brute force

- Define $U_t = \left\{ u \in \left\{ -\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}} \right\} : |\text{supp}(u)| = t \right\}$
- For $v_* = \arg\max_{u \in U_t} \langle Y, u \otimes^p \rangle$,
  $$|\text{supp}(v_*) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t$$
- Define $\alpha \in \mathbb{R}^n$ with entries $\alpha_\ell := \langle Y, v_* \otimes^{p-1} \otimes e_\ell \rangle$, $\forall \ell \in [n]$
- For all $\ell \in \text{supp}(x)$,
  $$\alpha_\ell = \lambda \cdot x_\ell \cdot \langle x, v_* \rangle^{p-1} + \langle W, v_* \otimes^{p-1} \otimes e_\ell \rangle \gtrsim \sqrt{\log n}$$
- For all $\ell \notin \text{supp}(x)$ (i.e. $x_\ell = 0$),
  $$\alpha_\ell = \langle W, v_* \otimes^{p-1} \otimes e_\ell \rangle \lesssim \sqrt{\log n}$$
Limited brute force

- Define $U_t = \left\{ u \in \left\{-\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}}\right\} : |\text{supp}(u)| = t \right\}$
- For $v_* = \arg\max_{u \in U_t} \langle Y, u^\otimes p \rangle$, 
  $|\text{supp}(v_*) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t$
- Define $\alpha \in \mathbb{R}^n$ with entries $\alpha_\ell := \langle Y, v_*^{\otimes p-1} \otimes e_\ell \rangle$, $\forall \ell \in [n]$ 
- For all $\ell \in \text{supp}(x)$, 
  $\alpha_\ell = \lambda \cdot x_\ell \cdot \langle x, v_* \rangle^{p-1} + \langle W, v_*^{\otimes p-1} \otimes e_\ell \rangle \gtrsim \sqrt{\log n}$ 
- For all $\ell \not\in \text{supp}(x)$ (i.e. $x_\ell = 0$), 
  $\alpha_\ell = \langle W, v_*^{\otimes p-1} \otimes e_\ell \rangle \lesssim \sqrt{\log n}$
- Using $v_*$, the vector $\alpha$ acts as indicator of the support of $x$
Limited brute force

- Define $U_t = \left\{ u \in \left\{ -\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}} \right\} : |\text{supp}(u)| = t \right\}$
- For $v_\ast = \arg\max_{u \in U_t} \langle Y, u \otimes p \rangle$,
  \[ |\text{supp}(v_\ast) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t \]
- Using $v_\ast$, the vector $\alpha$ acts as indicator of the support of $x$

...
Limited brute force

• Define $U_t = \left\{ u \in \left\{-\frac{1}{\sqrt{t}}, 0, \frac{1}{\sqrt{t}}\right\} : |\text{supp}(u)| = t \right\}$

• For $v_\ast = \arg \max_{u \in U_t} \langle Y, u \otimes p \rangle$,

$$|\text{supp}(v_\ast) \cap \text{supp}(x)| \geq (1 - \epsilon) \cdot t$$

• Using $v_\ast$, the vector $\alpha$ acts as indicator of the support of $x$

• Relationship with known algorithms
  • When $t = 1$, this recovers the idea of diagonal thresholding (Pick out largest coordinate, one at a time)
  • When $t = k$, this is literally brute force (MLE)
Algorithmic extensions

Multiple spikes

• \( \mathbf{Y} = \mathbf{W} + \sum_{q=1}^{r} \lambda_q \mathbf{x}^{(q)} \otimes \mathbf{p} \)

• Disjoint support assumption: \( \text{supp}(x_i) \cap \text{supp}(x_j) = \emptyset \)
Algorithmic extensions

Multiple spikes

- \( \mathbf{Y} = \mathbf{W} + \sum_{q=1}^{r} \lambda_q \mathbf{x}^{(q)} \otimes \mathbf{p} \)
- Disjoint support assumption: \( \text{supp}(x_i) \cap \text{supp}(x_j) = \emptyset \)

Approximately flat \( k \)-sparse unit length signals

Fix a constant \( A \geq 1 \). For \( k \)-sparse signal \( x \),

\[
\frac{1}{A\sqrt{k}} \leq |x_\ell| \leq \frac{A}{\sqrt{k}} \quad \text{for } \ell \in \text{supp}(x)
\]
Algorithmic extensions

Multiple spikes

- \( \mathbf{Y} = \mathbf{W} + \sum_{q=1}^{r} \lambda_q \mathbf{x}^{\otimes p}_{(q)} \)

- Disjoint support assumption: \( \text{supp}(x_i) \cap \text{supp}(x_j) = \emptyset \)

Approximately flat \( k \)-sparse unit length signals

Fix a constant \( A \geq 1 \). For \( k \)-sparse signal \( x \),

\[
\frac{1}{A\sqrt{k}} \leq |x_\ell| \leq \frac{A}{\sqrt{k}} \quad \text{for} \ \ell \in \text{supp}(x)
\]

General tensor spike

Instead of just \( x^{\otimes p} \), we can allow the tensor signal to be \( x^{(1)} \otimes \ldots \otimes x^{(p)} \) involving \( 1 \leq \ell \leq p \) distinct \( k \)-sparse vectors
Lower bound summary

• Given a recovery algorithm, one can distinguish whether $\lambda = 0$
Lower bound summary

- Given a recovery algorithm, one can distinguish whether $\lambda = 0$
- Computational lower bound
  - Premise: Low-deg polynomials capture efficient functions
  - Likelihood ratio test, restricted to low-degree polynomials
  - Special cases of our bounds match known sparse PCA and tensor PCA low-degree bounds
Lower bound summary

- Given a recovery algorithm, one can distinguish whether $\lambda = 0$
- Computational lower bound
  - Premise: Low-deg polynomials capture efficient functions
  - Likelihood ratio test, restricted to low-degree polynomials
  - Special cases of our bounds match known sparse PCA and tensor PCA low-degree bounds
- Information-theoretic lower bound
  - Fano’s inequality on $\epsilon$-packing of flat $k$-sparse unit vectors $U_k$
  - Our bound is equivalent (up to constants) with recent works that study phase transition for weak recovery
Key contributions

1. A parametric multi-spike recovery algorithm for sparse tensor PCA that trades off running time with signal strength requirements
   - Given exponential time, our algorithm can recover the signal at the best known information-theoretic threshold
   - If we insist on polynomial time, our algorithm recovers the signal at the best known computational threshold

2. A computational lower bound based on low-degree polynomials and the low-degree likelihood method

3. An information-theoretic lower bound for approximate recovery