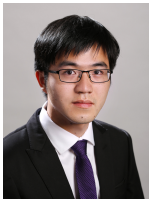


Effectiveness of Nonconvex Tensor Completion: Fast Convergence & Uncertainty Quantification



Yuxin Chen

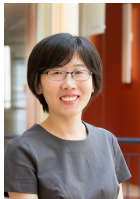
Princeton University



Changxiao Cai
Princeton



Gen Li
Tsinghua

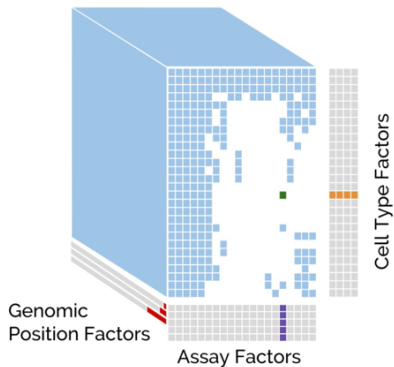


Yuejie Chi
CMU

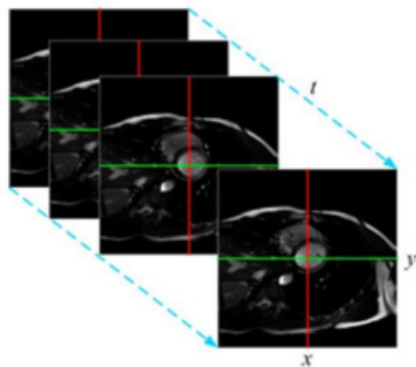


H. Vincent Poor
Princeton

Ubiquity of high-dimensional tensor data

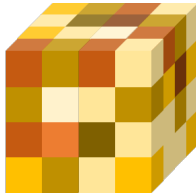


computational genomics
— *fig. credit: Schreiber et al. 19*

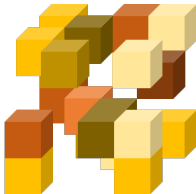


dynamic MRI
— *fig. credit: Liu et al. 17*

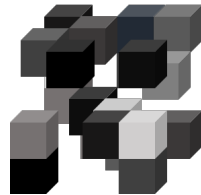
Imperfect data acquisition



a tensor of interest

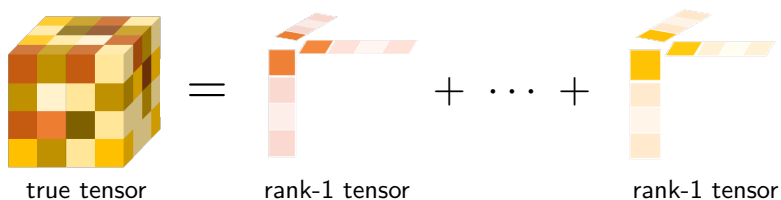


missing data



noise

Key to enabling reliable reconstruction from incomplete data
— exploiting **low CP-rank structure**



$$\mathbf{T}^* = \sum_{i=1}^r \mathbf{u}_i^* \otimes \mathbf{u}_i^* \otimes \mathbf{u}_i^*$$

Setup



T^*



T

- unknown rank- r tensor T^* :

$$T^* = \sum_{i=1}^r \mathbf{u}_i^* \otimes \mathbf{u}_i^* \otimes \mathbf{u}_i^* \in \mathbb{R}^{d \times d \times d}$$

Setup



T^*



T

- unknown rank- r tensor T^* :

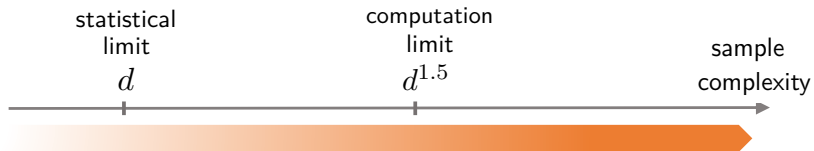
$$T^* = \sum_{i=1}^r \mathbf{u}_i^* \otimes \mathbf{u}_i^* \otimes \mathbf{u}_i^* \in \mathbb{R}^{d \times d \times d}$$

- incomplete & noisy observations over a *random* sampling set Ω :

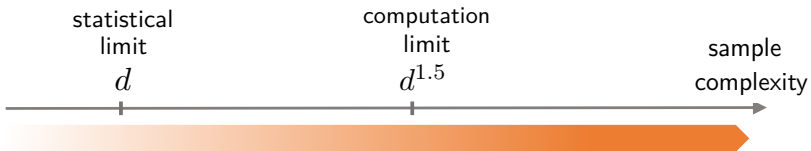
$$T_{i,j,k} = T_{i,j,k}^* + \text{noise}, \quad (i, j, k) \in \Omega$$

Goal: recover $\{\mathbf{u}_i^*\}_{i=1}^r$ and T^*

Statistical-computational gap ($r = O(1)$)

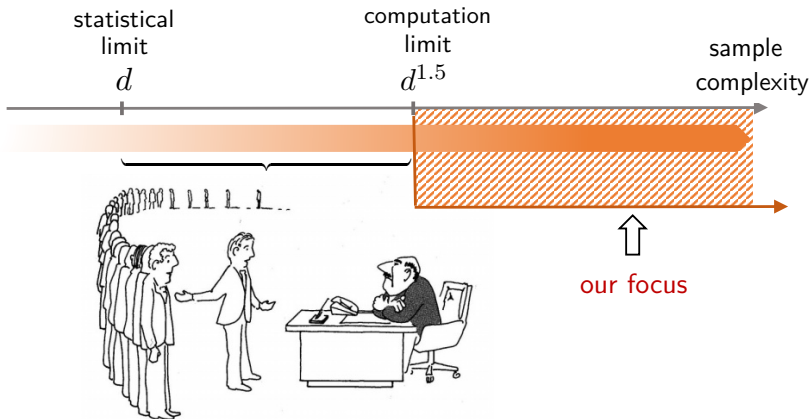


Statistical-computational gap ($r = O(1)$)



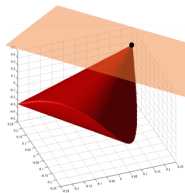
"I can't find an efficient algorithm, but neither can all these people."

Statistical-computational gap ($r = O(1)$)

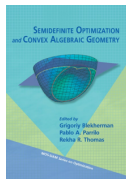


"I can't find an efficient algorithm, but neither can all these people."

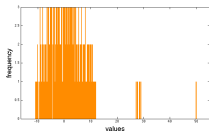
Prior art



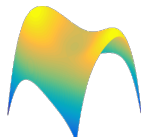
convex relaxation



sum-of-squares hierarchy



spectral methods



nonconvex optimization

- Gandy, Recht, Yamada '11
- Liu, Musialski, Wonka, Ye '12
- Kressner, Steinlechner, Vandereycken '13
- Xu, Hao, Yin, Su '13
- Romera-Paredes, Pontil '13
- Jain, Oh '14
- Huang, Mu, Goldfarb, Wright '15
- Barak, Moitra '16
- Zhang, Aeron '16
- Yuan, Zhang '16
- Montanari, Sun '16
- Kasai, Mishra '16
- Potechin, Steurer '17
- Dong, Yuan, Zhang '17
- Xia, Yuan '19
- Zhang '19
- ...

Prior art ($r = O(1)$)

| | algorithm | sample size | comput. cost | recovery type (noiseless) |
|---------------------|----------------------------------|-------------|-------------------|---------------------------|
| Yuan, Zhang '16 | tensor nuclear norm | d | NP-hard | exact |
| Xia, Yuan '17 | spectral method + GD on manifold | $d^{3/2}$ | slow | exact |
| Montanari, Sun '18 | spectral method | $d^{3/2}$ | d^3 | inexact |
| Barak, Moitra '16 | sum-of-squares | $d^{3/2}$ | slow (d^{15}) | exact |
| Potechin et al. '17 | sum-of-squares | $d^{3/2}$ | slow (d^{10}) | exact |

Prior art ($r = O(1)$)

| | algorithm | sample size | comput. cost | recovery type (noiseless) |
|---------------------|----------------------------------|-------------|-------------------|---------------------------|
| Yuan, Zhang '16 | tensor nuclear norm | d | NP-hard | exact |
| Xia, Yuan '17 | spectral method + GD on manifold | $d^{3/2}$ | slow | exact |
| Montanari, Sun '18 | spectral method | $d^{3/2}$ | d^3 | inexact |
| Barak, Moitra '16 | sum-of-squares | $d^{3/2}$ | slow (d^{15}) | exact |
| Potechin et al. '17 | sum-of-squares | $d^{3/2}$ | slow (d^{10}) | exact |

| | algorithm | ℓ_2 error (noisy) | ℓ_∞ error (noisy) |
|----------------------|---------------------------------------|------------------------|-----------------------------|
| Xia, Yuan, Zhang '17 | spectral method + tensor power method | suboptimal | n/a |
| Barak, Moitra '16 | sum-of-squares | suboptimal | n/a |

*Can we design an algorithm that is simultaneously
sample-efficient, computationally fast, & minimax-optimal?*

A nonconvex least squares formulation

$$\underset{\mathbf{U}=[\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{d \times r}}{\text{minimize}} \quad f(\mathbf{U}) := \sum_{(i,j,k) \in \Omega} \underbrace{\left\{ \left(\sum_{s=1}^r \mathbf{u}_s^{\otimes 3} \right)_{i,j,k} - T_{i,j,k} \right\}^2}_{\text{squared loss over observed entries}}$$

A nonconvex least squares formulation

$$\underset{\mathbf{U}=[\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{d \times r}}{\text{minimize}} \quad f(\mathbf{U}) := \underbrace{\sum_{(i,j,k) \in \Omega} \left\{ \left(\sum_{s=1}^r \mathbf{u}_s^{\otimes 3} \right)_{i,j,k} - T_{i,j,k} \right\}^2}_{\text{squared loss over observed entries}}$$

- **pros:** statistically efficient *if we can find global solutions*

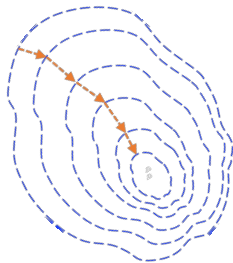
A nonconvex least squares formulation

$$\underset{\mathbf{U}=[\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{d \times r}}{\text{minimize}} \quad f(\mathbf{U}) := \underbrace{\sum_{(i,j,k) \in \Omega} \left\{ \left(\sum_{s=1}^r \mathbf{u}_s^{\otimes 3} \right)_{i,j,k} - T_{i,j,k} \right\}^2}_{\text{squared loss over observed entries}}$$

- **pros:** statistically efficient *if we can find global solutions*
- **cons:** highly nonconvex \rightarrow computationally challenging

Gradient descent (GD) with random initialization?

$$\underset{U=[\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{d \times r}}{\text{minimize}} \quad f(U) := \sum_{(i,j,k) \in \Omega} \left\{ \left(\sum_{s=1}^r \mathbf{u}_s^{\otimes 3} \right)_{i,j,k} - T_{i,j,k} \right\}^2$$



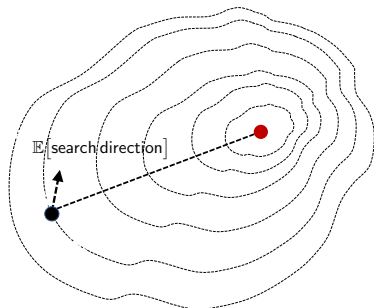
- **initialize** U^0 randomly
- **gradient descent:** for $t = 0, 1, \dots$,

$$U^{t+1} = U^t - \eta_t \nabla f(U^t)$$

— *succeeds for phase retrieval (Chen et al. '18)*

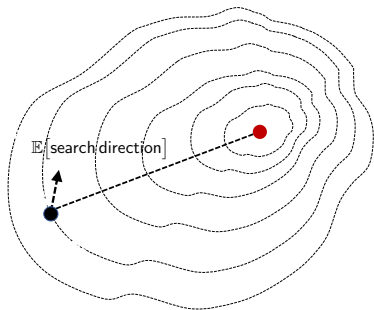
A negative conjecture

Randomly initialized GD does NOT work unless sample size $> d^2$



A negative conjecture

Randomly initialized GD does NOT work unless sample size $> d^2$

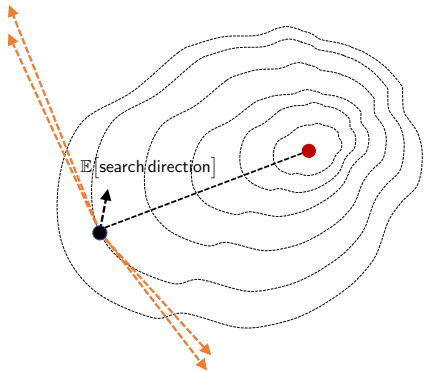


When sample size $\asymp d^{1.5}$:

- $\mathbb{E}[\text{search direction}]$ is desirable

A negative conjecture

Randomly initialized GD does NOT work unless sample size $> d^2$

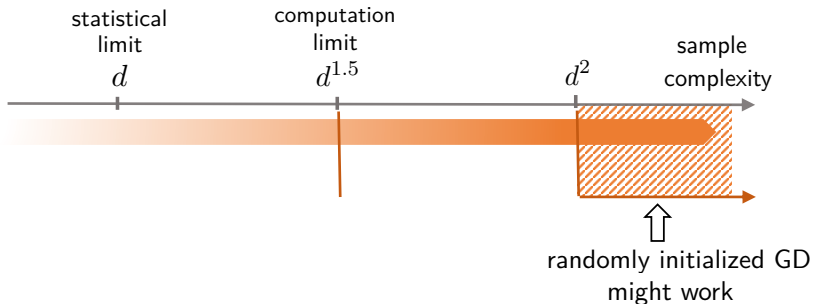


When sample size $\asymp d^{1.5}$:

- $\mathbb{E}[\text{search direction}]$ is desirable
- **issue:** variance $\gtrsim \sqrt{d} \text{mean}^2$

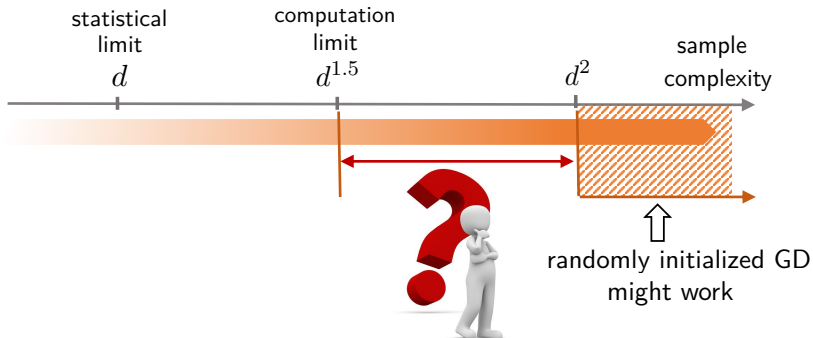
A negative conjecture

Randomly initialized GD does NOT work unless sample size $> d^2$

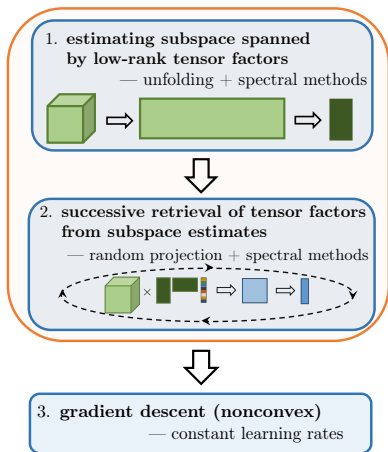


A negative conjecture

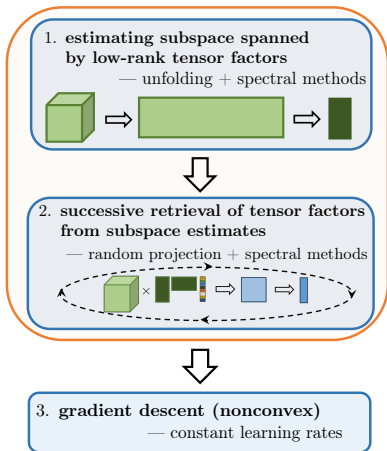
Randomly initialized GD does NOT work unless sample size $> d^2$



Our proposal: a two-stage nonconvex algorithm



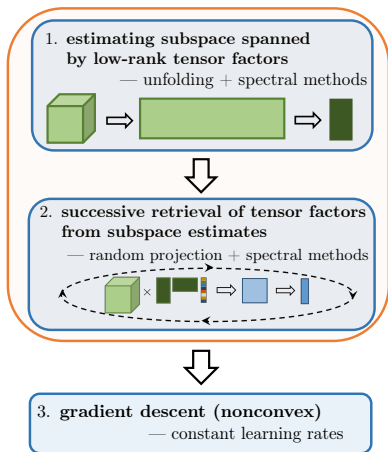
Our proposal: a two-stage nonconvex algorithm



1. initialization: U^0

- estimate $\text{span}\{\mathbf{u}_i^*\}$ via spectral method
- disentangle individual factors $\{\mathbf{u}_i^*\}$ from subspace estimate

Our proposal: a two-stage nonconvex algorithm



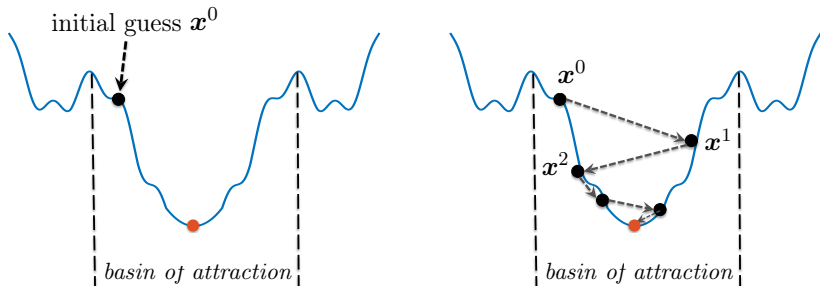
1. initialization: U^0

- estimate **span** $\{\mathbf{u}_i^*\}$ via spectral method
- disentangle individual factors $\{\mathbf{u}_i^*\}$ from subspace estimate

2. gradient descent: for $t = 0, 1, \dots$

$$U^{t+1} = U^t - \eta \nabla f(U^t)$$

Rationale of two-stage approach



1. initialize within a local basin sufficiently close to global min
(restricted) strongly convex
2. iterative refinement

A bit more details about initialization

Step 1.1: estimate $\text{span}\{\mathbf{u}_i^*\}_{1 \leq i \leq r} \longrightarrow \mathbf{U}_{\text{sub}}$

- matricization: $\mathbf{A} = \text{unfold}(\mathbf{T})$
- estimate rank- r subspace of $\mathcal{P}_{\text{off-diag}}(\mathbf{A}\mathbf{A}^\top)$ (diagonal deletion)



A bit more details about initialization

Step 1.2: retrieve tensor factors from subspace estimate

- generate a random vector \mathbf{g} from U_{sub}
- compute leading eigenvector of $\mathbf{T} \otimes \mathbf{g} = \sum_i \langle \mathbf{u}_i^*, \mathbf{g} \rangle \mathbf{u}_i^* \mathbf{u}_i^{*\top} + \text{noise}$
- repeat ...

find the \mathbf{u}_i^* most aligned with \mathbf{g}

Assumptions

$$\mathbf{T}^* = \sum_{i=1}^r \mathbf{u}_i^* \otimes \mathbf{u}_i^* \otimes \mathbf{u}_i^* \in \mathbb{R}^{d \times d \times d}$$

- **random sampling**: each entry is observed independently with prob. p

Assumptions

$$\mathbf{T}^* = \sum_{i=1}^r \mathbf{u}_i^* \otimes \mathbf{u}_i^* \otimes \mathbf{u}_i^* \in \mathbb{R}^{d \times d \times d}$$

- **random sampling**: each entry is observed independently with prob. p
- **random noise**: independent zero-mean sub-Gaussian noise with variance $O(\sigma^2)$

Assumptions

$$\mathbf{T}^* = \sum_{i=1}^r \mathbf{u}_i^* \otimes \mathbf{u}_i^* \otimes \mathbf{u}_i^* \in \mathbb{R}^{d \times d \times d}$$

- **random sampling:** each entry is observed independently with prob. p
- **random noise:** independent zero-mean sub-Gaussian noise with variance $O(\sigma^2)$
- **ground truth:** low-rank ($r = O(1)$), well-conditioned, incoherent ($\{\mathbf{u}_i^*\}$ are de-localized and not aligned)

l_2 and l_∞ theoretical guarantees

Theorem 1 (Cai, Li, Poor, Chen '19)

There exists some constant $\rho < 1$ and some permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{r \times r}$ s.t. with high prob., the t -th iterate satisfies

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{U}^*\|_F$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{T}^*\|_F$$

l_2 and l_∞ theoretical guarantees

Theorem 1 (Cai, Li, Poor, Chen '19)

There exists some constant $\rho < 1$ and some permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{r \times r}$ s.t. with high prob., the t -th iterate satisfies

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{U}^*\|_F$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{T}^*\|_F$$

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{U}^*\|_\infty$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{T}^*\|_\infty$$

provided that sample size $\gtrsim d^{1.5} \text{poly} \log(d)$

ℓ_2 and ℓ_∞ theoretical guarantees

Theorem 1 (Cai, Li, Poor, Chen '19)

There exists some constant $\rho < 1$ and some permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{r \times r}$ s.t. with high prob., the t -th iterate satisfies

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{U}^*\|_F$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{T}^*\|_F$$

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{U}^*\|_\infty$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{T}^*\|_\infty$$

provided that sample size $\gtrsim d^{1.5} \text{poly} \log(d)$

- linear/geometric convergence \longrightarrow linear-time algorithm

l_2 and l_∞ theoretical guarantees

Theorem 1 (Cai, Li, Poor, Chen '19)

There exists some constant $\rho < 1$ and some permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{r \times r}$ s.t. with high prob., the t -th iterate satisfies

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{U}^*\|_F$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{T}^*\|_F$$

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{U}^*\|_\infty$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{T}^*\|_\infty$$

provided that *sample size* $\gtrsim d^{1.5} \text{poly} \log(d)$

- near-optimal sample complexity (among poly-time algorithms)

l_2 and l_∞ theoretical guarantees

Theorem 1 (Cai, Li, Poor, Chen '19)

There exists some constant $\rho < 1$ and some permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{r \times r}$ s.t. with high prob., the t -th iterate satisfies

$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{U}^*\|_F$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_F \lesssim (\rho^t + \sigma \sqrt{d/p}) \|\mathbf{T}^*\|_F$$

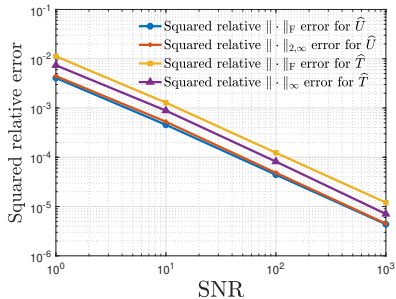
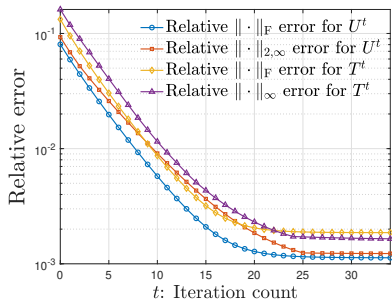
$$\|\mathbf{U}^t \mathbf{\Pi} - \mathbf{U}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{U}^*\|_\infty$$

$$\|\mathbf{T}^t - \mathbf{T}^*\|_\infty \lesssim (\rho^t + \sigma \sqrt{d \log d/p}) \|\mathbf{T}^*\|_\infty$$

provided that sample size $\gtrsim d^{1.5} \text{poly} \log(d)$

- near-optimal statistical accuracy (both l_2 and l_∞)

Numerical experiments

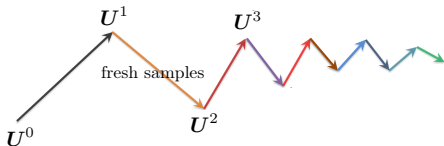


$$d = 100, r = 4, p = 0.1$$

No need of sample splitting

Sample-splitting (fresh samples at each iteration)?

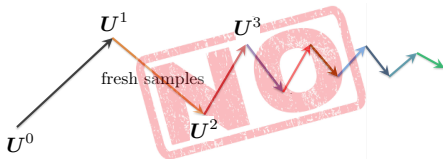
— helps analysis but waste data



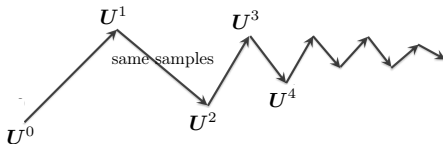
No need of sample splitting

Sample-splitting (fresh samples at each iteration)?

— helps analysis but waste data



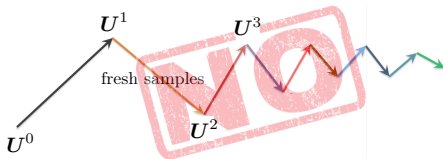
Our results: reusing all samples in all iterations



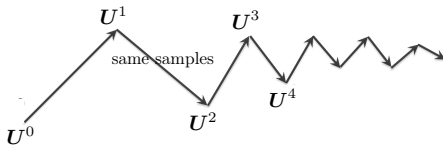
No need of sample splitting

Sample-splitting (fresh samples at each iteration)?

— helps analysis but waste data



Our results: reusing all samples in all iterations



How to deal with complicated statistical dependency across iterations?

Key proof idea: leave-one-out decoupling

Leave out a small amount of randomness and re-run the algorithm

Key proof idea: leave-one-out decoupling

Leave out a small amount of randomness and re-run the algorithm

- El Karoui, Bean, Bickel, Lim, Yu '13
- El Karoui '15
- Javanmard, Montanari '15
- Zhong, Boumal '17
- Lei, Bickel, El Karoui '17
- Sur, Chen, Candès '17
- Abbe, Fan, Wang, Zhong '17
- Chen, Fan, Ma, Wang '17
- Ma, Wang, Chi, Chen '17
- Chen, Chi, Fan, Ma '18
- Ding, Chen '18
- Dong, Shi '18
- Sur, Candès '18
- Chen, Liu, Li '19
- Chen, Fan, Ma, Yan '19
- Pananjady, Wainwright '19
- Ling '20
- Chen, Fan, Ma, Yan '20
- Agarwal, Kakade, Yang '20
- Abbe, Fan, Wang '20
- Li, Wei, Chi, Gu, Chen '20

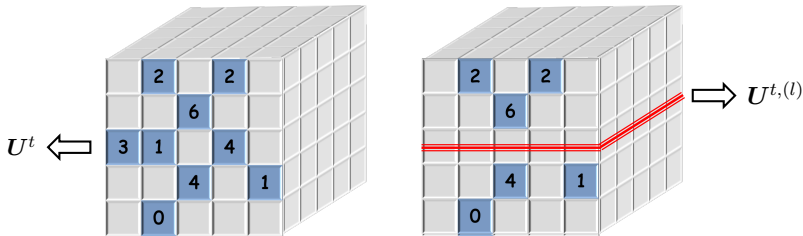
Foundations and Trends[®] in Machine Learning
**Spectral Methods for Data
Science: A Statistical Perspective**

Suggested Citation: Yuxin Chen, Yuejie Chi, Jianqing Fan and Cong Ma (2020), "Spectral Methods for Data Science: A Statistical Perspective", Foundations and Trends[®] in

| | | |
|----------|---|------------|
| 4 | Fine-grained analysis: ℓ_∞ and $\ell_{2,\infty}$ perturbation theory | 126 |
| 4.1 | Leave-one-out-analysis: An illustrative example | 127 |

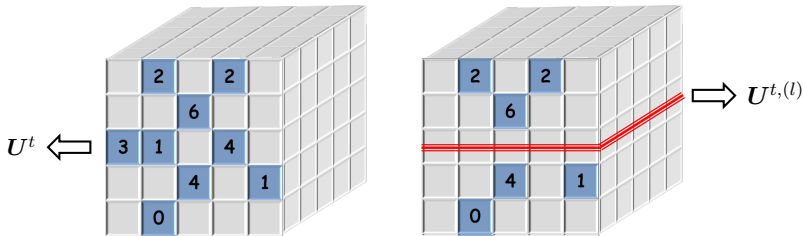
Key proof idea: leave-one-out decoupling

For each $1 \leq l \leq d$, generate leave-one-out auxiliary iterates $\{U^{t,(l)}\}$ by replacing l^{th} slice with true values



Key proof idea: leave-one-out decoupling

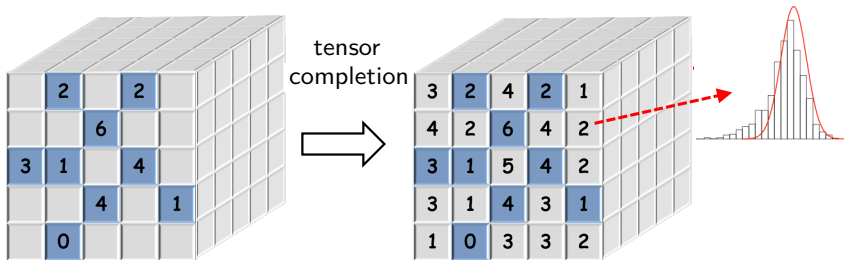
For each $1 \leq l \leq d$, generate leave-one-out auxiliary iterates $\{U^{t,(l)}\}$ by replacing l^{th} slice with true values



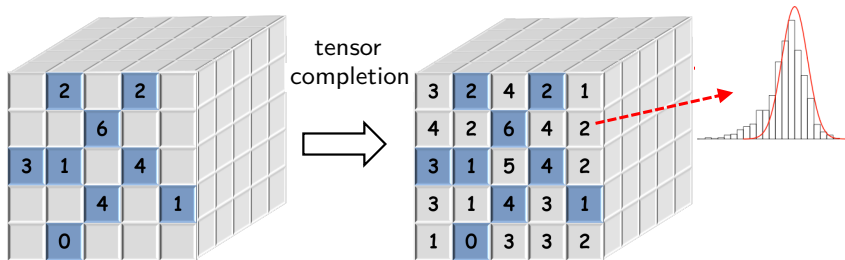
- exploit partial statistical independence
- exploit leave-one-out stability
- enable optimal l_∞ error control

Inference and uncertainty quantification

One step further: uncertainty quantification?



One step further: uncertainty quantification?



How to assess uncertainty, or “confidence”, of nonconvex estimates due to imperfect data acquisition?

- noise
- missing data

Challenges

$$\underset{U=[\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{d \times r}}{\text{minimize}} \quad f(U) := \underbrace{\sum_{(i,j,k) \in \Omega} \left\{ \left(\sum_{s=1}^r \mathbf{u}_s^{\otimes 3} \right)_{i,j,k} - T_{i,j,k} \right\}^2}_{\text{squared loss}}$$

- How to pin down distributions of nonconvex solutions?

Challenges

$$\underset{U=[\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{d \times r}}{\text{minimize}} \quad f(U) := \underbrace{\sum_{(i,j,k) \in \Omega} \left\{ \left(\sum_{s=1}^r \mathbf{u}_s^{\otimes 3} \right)_{i,j,k} - T_{i,j,k} \right\}^2}_{\text{squared loss}}$$

- How to pin down distributions of nonconvex solutions?
- How to adapt to unknown noise distributions and heteroscedasticity (i.e. location-varying noise variance)?

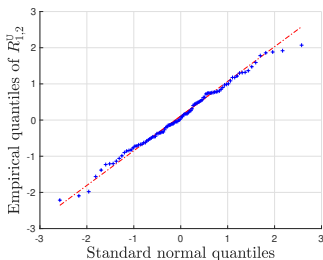
Challenges

$$\underset{U=[\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{d \times r}}{\text{minimize}} \quad f(U) := \underbrace{\sum_{(i,j,k) \in \Omega} \left\{ \left(\sum_{s=1}^r \mathbf{u}_s^{\otimes 3} \right)_{i,j,k} - T_{i,j,k} \right\}^2}_{\text{squared loss}}$$

- How to pin down distributions of nonconvex solutions?
- How to adapt to unknown noise distributions and heteroscedasticity (i.e. location-varying noise variance)?
- Existing estimation guarantees are highly insufficient
→ Overly wide confidence intervals

Distributional theory

- random sampling
- independent Gaussian noise
- ground truth: low-rank, incoherent, well-conditioned



Theorem 2 (Cai, Poor, Chen '20)

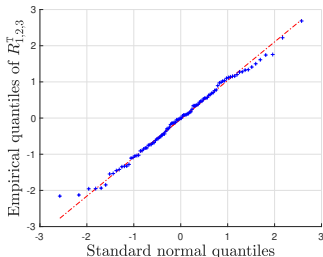
Consider any (i, j, k) s.t. the corresponding “SNR” is not exceedingly small. Then with high prob.,

$$\hat{T}_{i,j,k} - T_{i,j,k}^* \sim \mathcal{N}(0, \text{Cramér-Rao}) + \text{negligible term}$$

— asymptotically optimal

Distributional theory

- random sampling
- independent Gaussian noise
- ground truth: low-rank, incoherent, well-conditioned

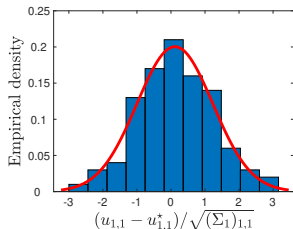


Theorem 2 (Cai, Poor, Chen '20)

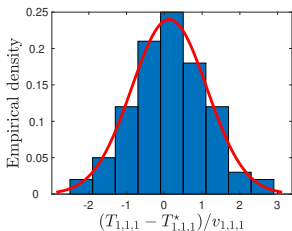
Consider any (i, j, k) s.t. the corresponding “SNR” is not exceedingly small. Then with high prob.,

$$\hat{T}_{i,j,k} - T_{i,j,k}^* \sim \mathcal{N}(0, \text{Cramér-Rao}) + \text{negligible term}$$

— asymptotically optimal

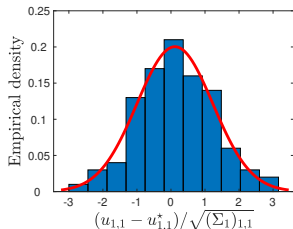


tensor factor entry

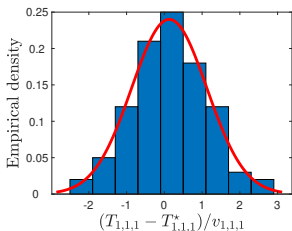


tensor entry

- **approximate Gaussianity:** estimation error of our nonconvex approach is zero-mean Gaussian

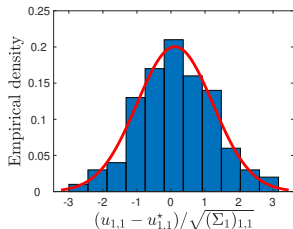


tensor factor entry

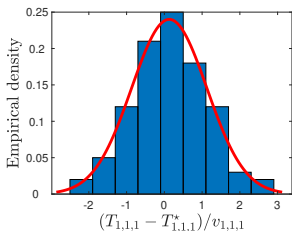


tensor entry

- **approximate Gaussianity:** estimation error of our nonconvex approach is zero-mean Gaussian
- **confidence intervals:** error (co)-variance can be accurately estimated, leading to valid CI construction



tensor factor entry



tensor entry

- **approximate Gaussianity:** estimation error of our nonconvex approach is zero-mean Gaussian
- **confidence intervals:** error (co)-variance can be accurately estimated, leading to valid CI construction
- **adaptivity:** our procedure is data-driven, and adaptive to unknown and heteroscedastic noise levels

Back to estimation: ℓ_2 optimality

Distributional theory in turn allows us to track estimation accuracy

Theorem 3 (Cai, Poor, Chen '20)

Suppose noise is i.i.d. $\mathcal{N}(0, \sigma^2)$. Then one has

$$\|\widehat{\mathbf{T}} - \mathbf{T}^*\|_{\mathbb{F}}^2 = \underbrace{\frac{(6 + o(1))\sigma^2 r d}{p}}_{\text{Cramér-Rao lower bound}}$$

Back to estimation: ℓ_2 optimality

Distributional theory in turn allows us to track estimation accuracy

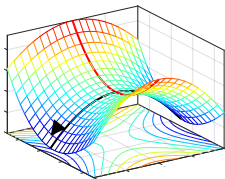
Theorem 3 (Cai, Poor, Chen '20)

Suppose noise is i.i.d. $\mathcal{N}(0, \sigma^2)$. Then one has

$$\|\widehat{\mathbf{T}} - \mathbf{T}^*\|_{\mathbb{F}}^2 = \underbrace{\frac{(6 + o(1))\sigma^2 r d}{p}}_{\text{Cramér-Rao lower bound}}$$

- precise characterization of estimation accuracy
- achieves full statistical efficiency (including **pre-constant**)

Summary



nonconvex
optimization

optimal estimation guarantees

linear-time algorithm

minimal sample size

fine-grained uncertainty quantification

“Nonconvex Low-Rank Tensor Completion from Noisy Data” C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, 2021+

“Subspace Estimation from Unbalanced and Incomplete Data Matrices: $\ell_{2,\infty}$ Statistical Guarantees,” C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *Annals of Statistics*, 2021

“Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality,” C. Cai, H. V. Poor, Y. Chen, *ICML 2020*