The convex geometry of blind deconvolution

and matrix completion

Dominik Stöger

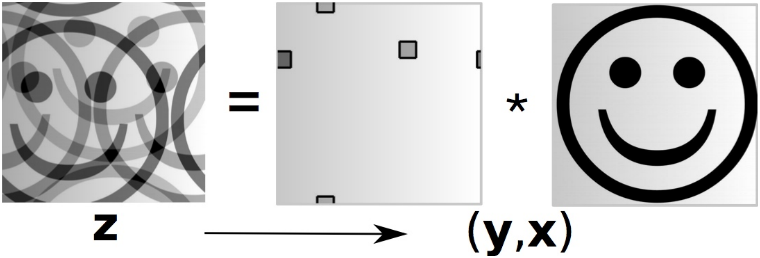
USC

February 20, 2020

Joint work with Felix Krahmer (TUM)

Blind deconvolution in imaging

- *Blind deconvolution ubiquituous in many applications:*
  - Imaging: $\boldsymbol{x}$ signal, $\boldsymbol{w}$ blur



- *(Circular) convolution of $\boldsymbol{w}, \boldsymbol{x} \in \mathbb{C}^L$:*

$$(\boldsymbol{w} * \boldsymbol{x})_k := \sum_{\ell=1}^{L} \boldsymbol{w}_k \boldsymbol{x}_{(\ell-k)} \mod L.$$

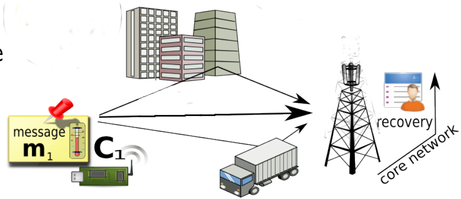
Blind deconvolution in wireless communications

- *Task:* deliver message $\boldsymbol{m} \in \mathbb{C}^N$ via unknown channel.
  Proposed approach: introduce redundancy before transmission.

- *Linear encoding:*
  $\boldsymbol{x} = \boldsymbol{C}\boldsymbol{m}$ with $\boldsymbol{C} \in \mathbb{C}^{L \times N}$
  the signal x is transmitted
- *Channel model:*
  only most direct paths are active
  $\boldsymbol{w} = \boldsymbol{B}\boldsymbol{h}$, where $\boldsymbol{B} \in \mathbb{C}^{L \times K}$
- *Received signal:* $\boldsymbol{e}$ noise

$$\boldsymbol{y} = \boldsymbol{w} * \boldsymbol{x} + \boldsymbol{e} \in \mathbb{C}^L$$

- Introduced by Ahmed, Recht, Romberg (IEEE IT '14)



Goal: recover $\boldsymbol{m}$ from $\boldsymbol{y}$

Lifting

- *Observation:* $\boldsymbol{w} * \boldsymbol{x} = \boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\boldsymbol{m}$ is bilinear in $\boldsymbol{h}$ and $\boldsymbol{m}$
  $\Rightarrow$ There is a unique linear map $\mathcal{A} : \mathbb{C}^{K \times N} \to \mathbb{C}^L$ such that

$$\boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\boldsymbol{m} = \mathcal{A}\left(\boldsymbol{h}\boldsymbol{m}^*\right)$$

  for arbitrary $\boldsymbol{h}$ and $\boldsymbol{m}$

Lifting

- *Observation:* $\boldsymbol{w} * \boldsymbol{x} = \boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\boldsymbol{m}$ is bilinear in $\boldsymbol{h}$ and $\boldsymbol{m}$
  $\Rightarrow$ There is a unique linear map $\mathcal{A} : \mathbb{C}^{K \times N} \to \mathbb{C}^{L}$ such that

$$\boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\boldsymbol{m} = \mathcal{A}(\boldsymbol{h}\boldsymbol{m}^*)$$

  for arbitrary $\boldsymbol{h}$ and $\boldsymbol{m}$

- Thus, the rank 1 matrix $\boldsymbol{X} = \boldsymbol{h}\boldsymbol{m}^*$ satisfies

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}) + \boldsymbol{e}$$

- Finding $\boldsymbol{X}$ is a low rank matrix recovery problem

Lifting

- *Observation:* $\boldsymbol{w} * \boldsymbol{x} = \boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\boldsymbol{m}$ is bilinear in $\boldsymbol{h}$ and $\boldsymbol{m}$
  $\Rightarrow$ There is a unique linear map $\mathcal{A} : \mathbb{C}^{K \times N} \to \mathbb{C}^{L}$ such that
  $$\boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\boldsymbol{m} = \mathcal{A}(\boldsymbol{h}\boldsymbol{m}^*)$$
  for arbitrary $\boldsymbol{h}$ and $\boldsymbol{m}$
- Thus, the rank 1 matrix $\boldsymbol{X} = \boldsymbol{h}\boldsymbol{m}^*$ satisfies
  $$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}) + \boldsymbol{e}$$
- Finding $\boldsymbol{X}$ is a low rank matrix recovery problem
- Ideally find
  $$\operatorname{argmin} \ \operatorname{rank} \boldsymbol{X} \quad \text{subject to } \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2 \leq \eta$$
- Such problems are NP-hard in general
  $\rightarrow$ try convex relaxation

A convex approach

SDP relaxation (Ahmed, Recht, Romberg '14)

Solve the semidefinite program (SDP)

$$\widetilde{\boldsymbol{X}} = \text{argmin} \, \|\boldsymbol{X}\|_* \qquad \text{subject to } \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2 \leq \eta. \tag{SDP}$$

A convex approach

SDP relaxation (Ahmed, Recht, Romberg '14)

Solve the semidefinite program (SDP)

$$\widetilde{\boldsymbol{X}} = \operatorname{argmin} \|\boldsymbol{X}\|_* \qquad \text{subject to } \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2 \leq \eta. \tag{SDP}$$

The *nuclear norm* $\|\boldsymbol{X}\|_* := \sum_{j=1}^{\operatorname{rank}(\boldsymbol{X})} \sigma_j(\boldsymbol{X})$ is the sum of all singular values.

Model assumptions

- $\boldsymbol{y} = \boldsymbol{Bh} * \boldsymbol{Cm} + \boldsymbol{e}$
- Adversarial noise: $\|\boldsymbol{e}\|_2 \leq \eta$
- $\boldsymbol{C} \in \mathbb{C}^{L \times N}$ has i.i.d. standard Gaussian entries
- $\boldsymbol{B} \in \mathbb{C}^{L \times K}$ satisfies $\boldsymbol{B}^* \boldsymbol{B} = \boldsymbol{Id}$ and is such that $\boldsymbol{FB}$ (for $\boldsymbol{F}$ the DFT) has rows of equal norm.

Recovery guarantees

Theorem (Ahmed, Recht, Romberg '14)

Assume

$$\frac{L}{\log^3 L} \geq C\left(K + N\mu_{\boldsymbol{h}}^2\right).$$

Then with high probability every minimizer $\widetilde{\boldsymbol{X}}$ *of (SDP) satisfies*

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{h}\boldsymbol{m}^*\|_F \lesssim \sqrt{K+N}\,\eta.$$

- $\mu_{\boldsymbol{h}}$ coherence parameter (typically small, see next slide)

Recovery guarantees

Theorem (Ahmed, Recht, Romberg '14)

Assume

$$\frac{L}{\log^3 L} \geq C\left(K + N\mu_{\boldsymbol{h}}^2\right).$$

Then with high probability every minimizer $\widetilde{\boldsymbol{X}}$ *of (SDP) satisfies*

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{h}\boldsymbol{m}^*\|_F \lesssim \sqrt{K+N}\ \eta.$$

- $\mu_{\boldsymbol{h}}$ coherence parameter (typically small, see next slide)
- Consequences:
    - No noise, i.e., $\eta = 0$:
      $\rightarrow$ Exact recovery with a near optimal-amount of measurements

Recovery guarantees

Theorem (Ahmed, Recht, Romberg '14)

Assume

$$\frac{L}{\log^3 L} \geq C\left(K + N\mu_{\boldsymbol{h}}^2\right).$$

Then with high probability every minimizer $\widetilde{\boldsymbol{X}}$ of (SDP) satisfies

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{h}\boldsymbol{m}^*\|_F \lesssim \sqrt{K+N}\,\eta.$$

- $\mu_{\boldsymbol{h}}$ coherence parameter (typically small, see next slide)
- Consequences:
  - No noise, i.e., $\eta = 0$:
    $\rightarrow$ Exact recovery with a near optimal-amount of measurements
  - Noisy scenario, i.e., $\eta > 0$:
    $\rightarrow$ dimension factor $\sqrt{K+N}$ appears in the noise

The coherence term $\mu_{\boldsymbol{h}}$ ...

... is given as

$$\mu_{\boldsymbol{h}} = \max(\mu_1, \mu_2)$$

The coherence term $\mu_{\boldsymbol{h}}$ ...

... is given as

$$\mu_{\boldsymbol{h}} = \max(\mu_1, \mu_2)$$

1. $\mu_1 = \frac{\sqrt{L}\|\boldsymbol{F}\boldsymbol{B}\boldsymbol{h}\|_\infty}{\|\boldsymbol{h}\|}$, where $\boldsymbol{F}$ is the discrete Fourier transform.

The coherence term $\mu_{\boldsymbol{h}}$ ...

... is given as

$$\mu_{\boldsymbol{h}} = \max(\mu_1, \mu_2)$$

1. $\mu_1 = \frac{\sqrt{L}\|\boldsymbol{F}\boldsymbol{B}\boldsymbol{h}\|_\infty}{\|\boldsymbol{h}\|}$, where $\boldsymbol{F}$ is the discrete Fourier transform.
   - *Intuition:* If mass is concentrated in only few entries of $\boldsymbol{FBh}$, it is more likely to get lost in the pointwise multiplication with $\boldsymbol{FCm}$.
   - *Range:* typically between 1 (optimal) and $\sqrt{K}$ (yields quadratic dependence on dimensions).
   - *Numerical evidence by Ahmed et al.:* μ_1 large $\Rightarrow$ more measurements needed.

The coherence term $\mu_{\boldsymbol{h}}$ ...

... is given as

$$\mu_{\boldsymbol{h}} = \max(\mu_1, \mu_2)$$

1. $\mu_1 = \frac{\sqrt{L}\|\boldsymbol{F}\boldsymbol{B}\boldsymbol{h}\|_\infty}{\|\boldsymbol{h}\|}$, where $\boldsymbol{F}$ is the discrete Fourier transform.
   - *Intuition:* If mass is concentrated in only few entries of $\boldsymbol{F}\boldsymbol{B}\boldsymbol{h}$, it is more likely to get lost in the pointwise multiplication with $\boldsymbol{F}\boldsymbol{C}\boldsymbol{m}$.
   - *Range:* typically between 1 (optimal) and $\sqrt{K}$ (yields quadratic dependence on dimensions).
   - *Numerical evidence by Ahmed et al.:* μ_1 large $\Rightarrow$ more measurements needed.
2. μ_2: technical term needed for golfing scheme.

Hope: Proof artifact, not necessary.

Noise robustness in low-rank matrix recovery

- Gaussian measurement matrices (implies RIP) ✓
- phase retrieval ✓
- *blind deconvolution* (this presentation) ?
- *matrix completion* (later in the talk) ?
- Robust PCA ?
- many more... ?

Noise robustness in low-rank matrix recovery

- Gaussian measurement matrices (implies RIP) ✓
- phase retrieval ✓
- *blind deconvolution* (this presentation) ?
- *matrix completion* (later in the talk) ?
- Robust PCA ?
- many more... ?

$\Longrightarrow$ Despite the popularity of convex relaxations for low-rank matrix recovery in the literature, their noise robustness is not well-understood.

What is the problem?

- Proof technique for these models:
  - Golfing scheme originally developed by D. Gross.
  - *Idea:* Show existence of (approximate) dual certificate w.h.p.

D. Gross

What is the problem?

- Proof technique for these models:
  - Golfing scheme originally developed by D. Gross.
  - *Idea:* Show existence of (approximate) dual certificate w.h.p.

D. Gross

- Works well in the noiseless case, where $\boldsymbol{X}_0$ is expected to be the minimizer
- Problem: In the noisy setting we do not know the actual minimizer

Are the dimension factors necessary?

Recall: We are interested in the scenario $L \ll KN$ and we optimize

$$\widetilde{\boldsymbol{X}} = \operatorname{argmin} \|\boldsymbol{X}\|_* \qquad \text{subject to } \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2 \leq \eta. \tag{SDP}$$

Are the dimension factors necessary?

Recall: We are interested in the scenario $L \ll KN$ and we optimize

$$\widetilde{\boldsymbol{X}} = \operatorname{argmin} \|\boldsymbol{X}\|_* \qquad \text{subject to } \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2 \le \eta. \qquad \text{(SDP)}$$

Theorem (Krahmer, DS '19)

There exists an admissible $\boldsymbol{B}$ such that:
With high probability there is $\eta_0 > 0$ such that for all $\eta \le \eta_0$ there exists an adversarial noise vector $\boldsymbol{e} \in \mathbb{C}^L$ with $\|\boldsymbol{e}\|_2 \le \eta$ that admits an alternative solution $\widetilde{\boldsymbol{X}}$ with the following properties.

- $\widetilde{\boldsymbol{X}}$ *is feasible, i.e.,* $\|\mathcal{A}\left(\widetilde{\boldsymbol{X}}\right) - \boldsymbol{y}\|_2 = \eta$
- $\widetilde{\boldsymbol{X}}$ *is preferred to* $\boldsymbol{hm}^*$ *by (SDP) i.e.,* $\|\widetilde{\boldsymbol{X}}\|_* \le \|\boldsymbol{hm}^*\|_*$, *but*
- $\widetilde{\boldsymbol{X}}$ *is far from the true solution in Frobenius norm, i.e.,*

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{hm}^*\|_F \ge \frac{\eta}{C_3}\sqrt{\frac{KN}{L}}.$$

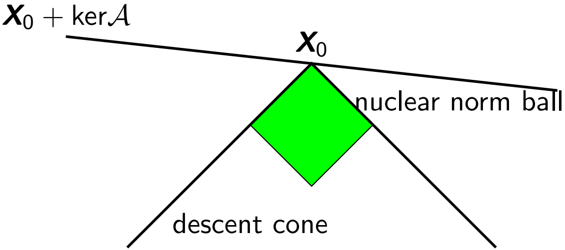
What does this mean?

- Assume $K = N$ and $L \approx CK$ up to log-factors

$$\Rightarrow \quad \|\widetilde{\boldsymbol{X}} - \boldsymbol{hm}^*\|_F \gtrsim \eta \sqrt{\frac{KN}{L}} \approx \eta \sqrt{K+N}.$$

  up to log-factors
  $\rightarrow$ The factor $\sqrt{K+N}$ is not a pure proof artifact.
- *Caution:* $\widetilde{\boldsymbol{X}}$ might not be the minimizer of (SDP)!

Ideas of the analysis I



- *Crucial geometric object:* Descent cone for $\boldsymbol{X}_0 \in \mathbb{C}^{K\times N}$

$$\mathcal{K}_* \left(\boldsymbol{X}_0\right) = \left\{ \boldsymbol{Z} \in \mathbb{C}^{K\times N} : \|\boldsymbol{X}_0 + \varepsilon \boldsymbol{Z}\|_* \le \|\boldsymbol{X}_0\|_* \text{ for some small } \varepsilon > 0 \right\}$$

Ideas of the analysis II

- Minimum conic singular value:

$$\lambda_{\min}\left(\mathcal{A}, \mathcal{K}_*\left(\boldsymbol{X}_0\right)\right) := \min_{\boldsymbol{Z} \in \mathcal{K}_*(\boldsymbol{X}_0)} \frac{\left\|\mathcal{A}\left(\boldsymbol{Z}\right)\right\|_2}{\|\boldsymbol{Z}\|_F}$$

- Noiseless scenario, i.e., $\eta = 0$:
  Exact recovery $\Longleftrightarrow \lambda_{\min}\left(\mathcal{A}, \mathcal{K}_*\left(\boldsymbol{X}_0\right)\right) > 0$

- Noisy scenario: Conic singular value controls stability [Chandrasekaran et al. '12]:

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{X}_0\|_F \leq \frac{2\eta}{\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(\boldsymbol{X}_0))}$$

  (As $\mathcal{A}$ is Gaussian, $\lambda_{\min}\left(\mathcal{A}, \mathcal{K}_*\left(\boldsymbol{X}_0\right)\right) \asymp 1$ w.h.p., whenever $m \gtrsim rn$ and $\boldsymbol{X}_0$ has rank r)

Ideas of the analysis III

Lemma (Krahmer, DS '19)

There exists $\boldsymbol{B} \in \mathbb{C}^{L \times K}$ *satisfying* $\boldsymbol{B}^* \boldsymbol{B} = Id_K$ *and* $\mu_{\max}^2 = 1$, *whose corresponding measurement operator* $\mathcal{A}$ *satisfies the following:*
Let $\boldsymbol{m} \in \mathbb{C}^N \setminus \{0\}$ *and let* $\boldsymbol{h} \in \mathbb{C}^K \setminus \{0\}$ *be incoherent. Then with high probability it holds that*

$$\lambda_{\min}\left(\mathcal{A}, \mathcal{K}_*\left(\boldsymbol{h}\boldsymbol{m}^*\right)\right) \leq C_3 \sqrt{\frac{L}{KN}}.$$

- Lemma can be used to prove the previous theorem.
- (Analogous result holds for matrix completion.)

All hope is lost?!

Recovery for high noise levels

Theorem (Krahmer, DS '19)

Let $\alpha > 0$. Assume that

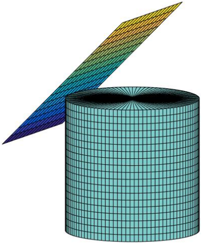
$$L \geq C_1 \frac{\mu^2}{\alpha^2} (K+N) \, \log^2 L.$$

Then with high probability the following statement holds for all $\boldsymbol{h} \in S^{K-1}$ with $\mu_{\boldsymbol{h}} \leq \mu$, all $\boldsymbol{m} \in S^{N-1}$, all $\eta > 0$, and all $\boldsymbol{e} \in \mathbb{C}^L$ with $\|\boldsymbol{e}\|_2 \leq \eta$: Any minimizer $\widetilde{\boldsymbol{X}}$ of (SDP) satisfies

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{h}\boldsymbol{m}^*\|_F \leq \frac{C_3 \mu^{2/3} \log^{2/3} L}{\alpha^{2/3}} \max\{\eta; \alpha\}.$$

$\rightarrow$ Near-optimal recovery guarantees for high noise-levels.

Proof sketch I



- Descent cone local approximation to descent set near $\boldsymbol{hm}^*$.
- *Geometric Intuition:* Close to $\ker \mathcal{A}$, the descent set is not pointy.

- Consider the partition $\mathcal{K}_* (\boldsymbol{X}_0) = \mathcal{K}_1 \cup \mathcal{K}_2$, where
  - $\mathcal{K}_1$ contains all elements in $\mathcal{K}_* (\boldsymbol{X}_0)$, which are near-orthogonal to $\boldsymbol{X}_0$
  - $\mathcal{K}_2 := \mathcal{K}_* (\boldsymbol{X}_0) \setminus \mathcal{K}_1$

Proof sketch II

Geometric intution: No large error can occur in directions belonging to $\mathcal{K}_1$ due to the curved nature of the nuclear norm ball

- $\lambda_{\min}(\mathcal{A}, \mathcal{K}_2)$ can be bounded from below using *Mendelson's small-ball method*
- $\rightarrow$ No large error can occur in these directions

S. Mendelson

Combining these two ideas yields the result.

Matrix Completion

Low-rank matrix completion

$$\boldsymbol{X}_0 = \begin{pmatrix} 1 & ? & ? & ? & ? \\ ? & ? & 3 & 7 & ? \\ ? & 2 & ? & ? & ? \\ ? & ? & 4 & ? & 1 \\ ? & 2 & ? & ? & ? \end{pmatrix}$$

Can one complete the matrix $\boldsymbol{X}_0$?

Our model

- $\boldsymbol{X}_0 \in \mathbb{R}^{n\times n}$ rank-r matrix
- *Measurement operator* $\mathcal{A}: \mathbb{R}^{n\times n} \to \mathbb{R}^m$:

$$\mathcal{A}(\boldsymbol{X})(i) := \frac{n}{\sqrt{m}} \boldsymbol{X}_{a_i,b_i},$$

  where (a_i, b_i) chosen uniformly at random for each $i \in [m]$ (i.i.d.)
- *Observation vector:*

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0) + \boldsymbol{e},$$

  $\boldsymbol{e}$ noise with $\|\boldsymbol{e}\|_2 \leq \eta$

Recovery guarantees

Recovery algorithm: We optimize

$$\widetilde{\boldsymbol{X}} = \operatorname{argmin} \|\boldsymbol{X}\|_* \qquad \text{subject to } \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2 \le \eta. \tag{SDP}$$

Recovery guarantees

Recovery algorithm: We optimize

$$\widetilde{\boldsymbol{X}} = \operatorname{argmin} \|\boldsymbol{X}\|_* \qquad \text{subject to } \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2 \leq \eta. \tag{SDP}$$

Theorem (Candes, Plan '10)

Let $\boldsymbol{X}_0 \in \mathbb{R}^{n \times n}$ be an incoherent rank-r matrix. Assume that

$$m \geq Cn \, poly \log(n).$$

Then with high probability every minimizer $\widetilde{\boldsymbol{X}}$ of (SDP) satisfies

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{X}_0\|_F \lesssim \sqrt{n}\, \eta.$$

Our result

Theorem (Krahmer, DS '19)

Assume that $\boldsymbol{X}_0 \in \mathbb{R}^{n\times n} \setminus \{0\}$ is a rank-r matrix. Then w.h.p. there is $\eta_0 > 0$ such that for all $\eta \le \eta_0$ there exists an adversarial noise vector $\boldsymbol{e} \in \mathbb{R}^m$ with $\|\boldsymbol{e}\|_2 \le \eta$ that admits an alternative solution $\widetilde{\boldsymbol{X}} \in \mathbb{R}^{n\times n}$ with the following properties.

- *$\widetilde{\boldsymbol{X}}$ is feasible, i.e., $\left\|\mathcal{A}\left(\widetilde{\boldsymbol{X}}\right) - \boldsymbol{y}\right\|_2 = \eta$ for $\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0) + \boldsymbol{e}$ the noisy measurement vector*
- *$\widetilde{\boldsymbol{X}}$ is preferred to $\boldsymbol{X}_0$ by (SDP), i.e., $\|\widetilde{\boldsymbol{X}}\|_* \le \|\boldsymbol{X}_0\|_*$, but*
- *$\widetilde{\boldsymbol{X}}$ is far from the true solution in Frobenius norm, i.e.,*

$$\|\widetilde{\boldsymbol{X}} - \boldsymbol{X}_0\|_F \ge \frac{\eta}{C} n \sqrt{\frac{r}{m}}.$$

$\Longrightarrow$ If $m \asymp n\text{polylog}\,(n_1)$, then $n\sqrt{\frac{r}{m}} \asymp \sqrt{\frac{n}{\text{poly}\log n}}$

Key lemma

Analogous to blind deconvolution, the result is a consequence of the following key lemma.

Lemma (Krahmer, DS '19)

Let $\boldsymbol{X}_0 \in \mathbb{R}^{n\times n} \setminus \{0\}$ be an incoherent rank-r matrix. Then with high probability it holds that

$$\lambda_{\min}\left(\mathcal{A}, \mathcal{K}_\star\left(\boldsymbol{X}_0\right)\right) \le C_3 \frac{1}{n}\sqrt{\frac{m}{r}}.$$

Proof sketch I

Our goal is to construct $\boldsymbol{Z} \in \mathbb{R}^{n \times n}$ such that

$$\frac{\| \mathcal{A}(\boldsymbol{Z}) \|_2}{\| \boldsymbol{Z} \|_F} \lesssim \frac{1}{n} \sqrt{\frac{m}{r}}$$

In the following, without loss of generality: $\| \boldsymbol{X}_0 \|_F = 1$

Proof sketch I

Our goal is to construct $\boldsymbol{Z} \in \mathbb{R}^{n \times n}$ such that

$$\frac{\|\mathcal{A}(\boldsymbol{Z})\|_2}{\|\boldsymbol{Z}\|_F} \lesssim \frac{1}{n}\sqrt{\frac{m}{r}}$$

In the following, without loss of generality: $\|\boldsymbol{X}_0\|_F = 1$

- Step 1: Characterize the descent cone of the nuclear norm:

$$\overline{\mathcal{K}_\star(\boldsymbol{X}_0)} = \left\{ \boldsymbol{Z} \in \mathbb{R}^{n \times n} : -\langle \boldsymbol{U}_0 \boldsymbol{V}_0^T, \boldsymbol{Z} \rangle_F \geq \|\mathcal{P}_{\mathrm{T}^\perp}(\boldsymbol{Z})\|_* \right\}$$

  - T denotes tangent space of $\boldsymbol{X}_0 = \boldsymbol{U}_0 \Sigma_0 \boldsymbol{V}_0^T$ (truncated SVD)

$$\mathrm{T} := \left\{ \boldsymbol{U}_0 \boldsymbol{V}^T + \boldsymbol{U}\boldsymbol{V}_0^T : \boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{n \times r} \right\}$$

  - $\mathrm{T}^\perp$ orthogonal complement of T
  - $\mathcal{P}_{\mathrm{T}^\perp}$ orthogonal projection onto $\mathrm{T}^\perp$

Proof sketch II

- Step 2: With high probability there is $\boldsymbol{W} \in \mathbb{R}^{n \times n}$ such that $\|\boldsymbol{W}\|_F \approx 1$,

$$\mathcal{A}(\boldsymbol{W}) = 0$$

  and,

$$\|\mathcal{P}_{\mathsf{T}^\perp} \boldsymbol{W}\|_* \lesssim \frac{1}{n}\sqrt{\frac{m}{r}}$$

- Step 3: Define

$$\boldsymbol{Z} := \boldsymbol{W} - \lambda \boldsymbol{X}_0$$

  for $\lambda \approx \frac{1}{n}\sqrt{\frac{m}{r}}$

Proof sketch III

- Step 5: Check $\boldsymbol{Z} = \boldsymbol{W} - \lambda \boldsymbol{X}_0 \in \mathcal{K}_\star(\boldsymbol{X}_0)$
- Step 6: Show that

$$\|\mathcal{A}(\boldsymbol{Z})\|_2 \lesssim \frac{1}{n}\sqrt{\frac{m}{r}}\|\boldsymbol{Z}\|_F$$

$\square$

Proof sketch III

- Step 5: Check $\boldsymbol{Z} = \boldsymbol{W} - \lambda \boldsymbol{X}_0 \in \mathcal{K}_\star(\boldsymbol{X}_0)$
- Step 6: Show that

$$\|\mathcal{A}(\boldsymbol{Z})\|_2 \lesssim \frac{1}{n}\sqrt{\frac{m}{r}}\|\boldsymbol{Z}\|_F$$

$\square$

- Geometric consequence of the proof:
  The angle between $\ker \mathcal{A}$ and $\mathcal{K}_\star(\boldsymbol{X}_0)$ is rather small.

Outlook and open questions

- What can we say about the actual minimizer in the scenario of small noise?
- Stability of matrix completion?
- What if the ground truth is approximately low-rank?
- Random noise?!

Thank you for your attention!