The Convex Geometry of Blind Deconvolution

Dominik Stöger
Technische Universität München
Department of Mathematics

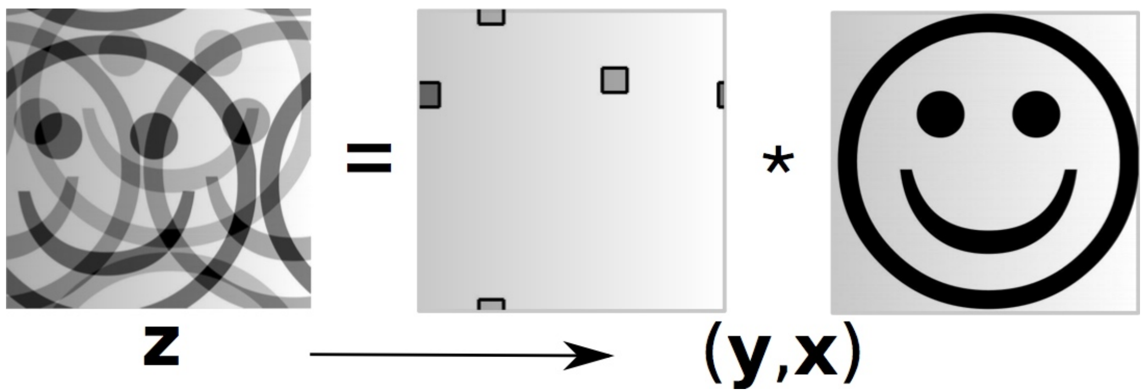
July 12, 2019

Joint work Felix Krahmer (TUM),

Funded by the DFG in the context of SPP 1798 CoSIP

Blind deconvolution in imaging

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



- *(Circular) convolution of $\boldsymbol{w}, \boldsymbol{x} \in \mathbb{C}^L$:* $(\boldsymbol{w} * \boldsymbol{x})_k := \sum_{\ell=1}^{L} \boldsymbol{w}_k \boldsymbol{x}_{(\ell-k) \mod L}$.

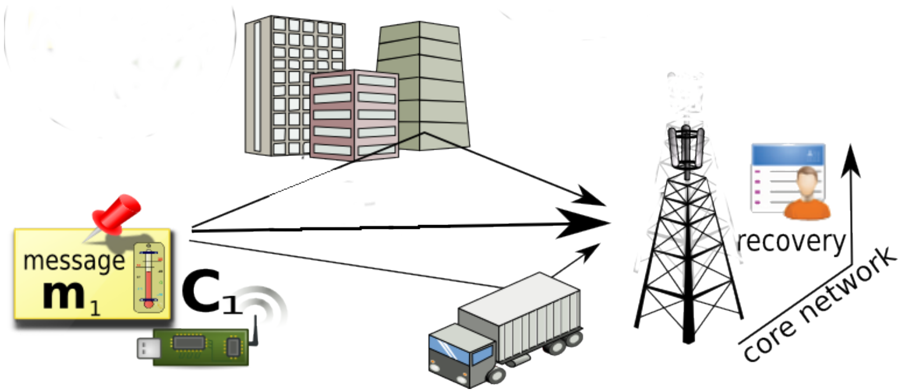
Blind deconvolution in wireless communications

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

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

Lifting

- *Observation:* $\boldsymbol{w} * \boldsymbol{x} = \boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\overline{\boldsymbol{m}}$ is bilinear in $\boldsymbol{h}$ and $\overline{\boldsymbol{m}}$
  $\Rightarrow$ There is a unique linear map $\mathcal{A} : \mathbb{C}^{K \times N} \to \mathbb{C}^{L}$ such that
  $$\boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\overline{\boldsymbol{m}} = \mathcal{A}(\boldsymbol{h}\boldsymbol{m}^*)$$
  for arbitrary $\boldsymbol{h}$ and $\boldsymbol{m}$
- Thus, the rank 1 matrix $\boldsymbol{X}_0 = \boldsymbol{h}\boldsymbol{m}^*$ satisfies
  $$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0) + \boldsymbol{e}$$
- Finding $\boldsymbol{X}_0$ is a **low rank matrix recovery problem**

Lifting

- *Observation:* $\boldsymbol{w} * \boldsymbol{x} = \boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\overline{\boldsymbol{m}}$ is bilinear in $\boldsymbol{h}$ and $\overline{\boldsymbol{m}}$
  $\Rightarrow$ There is a unique linear map $\mathcal{A} : \mathbb{C}^{K \times N} \to \mathbb{C}^L$ such that
  $$\boldsymbol{B}\boldsymbol{h} * \boldsymbol{C}\overline{\boldsymbol{m}} = \mathcal{A}(\boldsymbol{h}\boldsymbol{m}^*)$$
  for arbitrary $\boldsymbol{h}$ and $\boldsymbol{m}$
- Thus, the rank 1 matrix $\boldsymbol{X}_0 = \boldsymbol{h}\boldsymbol{m}^*$ satisfies
  $$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0) + \boldsymbol{e}$$
- Finding $\boldsymbol{X}_0$ 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)

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

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

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:

- $\boldsymbol{y} = \boldsymbol{Bh} * \boldsymbol{C\bar{m}} + \boldsymbol{e}$
- **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)

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)
- **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)
- **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
    Does not explain empirical success of (SDP)

Noise robustness in low-rank matrix recovery

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

Noise robustness in low-rank matrix recovery

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

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:**

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

D. Gross

What is the problem?

- **Proof technique for these models:**

  – *Idea:* Show existence of (approximate) dual certificate w.h.p.
  – *Golfing scheme* originally developed by D. Gross.

D. Gross

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

Are the dimension factors necessary?

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

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

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 $\tau_0 > 0$ *such that for all* $\tau \le \tau_0$ *there exists an adversarial noise vector* $\boldsymbol{e} \in \mathbb{C}^L$ *with* $\|\boldsymbol{e}\|_2 \le \tau$ *that admits an alternative solution* $\widetilde{\boldsymbol{X}}$ *with the following properties.*

- $\widetilde{\boldsymbol{X}}$ *is feasible, i.e.,* $\|\mathcal{A}\left(\widetilde{\boldsymbol{X}}\right) - \boldsymbol{y}\|_2 = \tau$
- $\widetilde{\boldsymbol{X}}$ *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.,*

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

What does this mean?

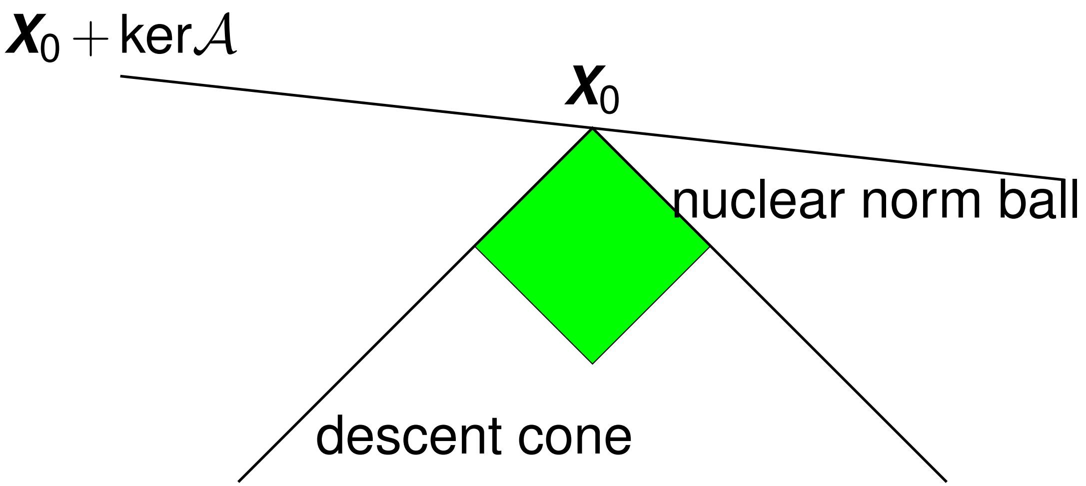
- Assume $K = N$ and $L \approx CK$ up to log-factors

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

  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)!

- Analogous result can be shown for matrix completion.

Ideas of the analysis I



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

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

Ideas of the analysis II

- **Minimum conic singular value**:

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

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

- **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, $\lambda_{\min}(\mathcal{A}, \mathcal{K}_*(\boldsymbol{hm}^*)) \asymp 1$ w.h.p., whenever $L \gtrsim K + N$)

Ideas of the analysis III

Lemma (Krahmer, DS '19)

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

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

- 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

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

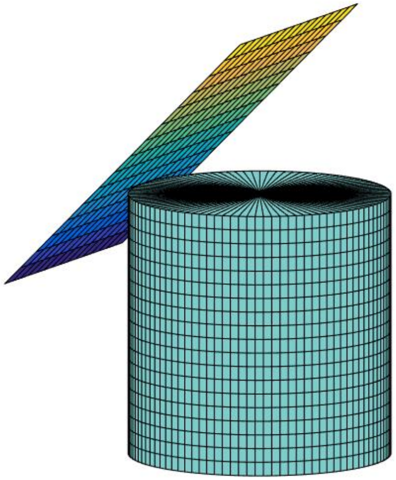
Then with high probability the following statement holds for all $\boldsymbol{h} \in S^{K-1}$ with $\mu_{\boldsymbol{h}} \leq \mu$, all $\boldsymbol{m} \in S^{N-1}$, all $\tau > 0$, and all $\boldsymbol{e} \in \mathbb{C}^L$ with $\|\boldsymbol{e}\|_2 \leq \tau$:
Any minimizer $\widetilde{\boldsymbol{X}}$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

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

$\rightarrow$ **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 $\mathcal{K}_*(\boldsymbol{hm}^*) = \mathcal{K}_1 \cup \mathcal{K}_2$, where
  - $\mathcal{K}_1$ contains all elements in $\mathcal{K}_*(\boldsymbol{hm}^*)$, which are near-orthogonal to $\boldsymbol{hm}^*$
  - $\mathcal{K}_2 := \mathcal{K}_*(\boldsymbol{hm}^*) \setminus \mathcal{K}_1$

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.

Outlook and open questions

- What can we say about the actual minimizer in the scenario of small noise?
- Stability of matrix completion?

Thank you for your attention!