Robust Recovery of Sparse Nonnegative Weights from Mixtures of Positive-Semidefinite matrices

Peter Jung

joint work with A. Fengler, F. Jaensch, S. Haghighatshoar and G. Caire

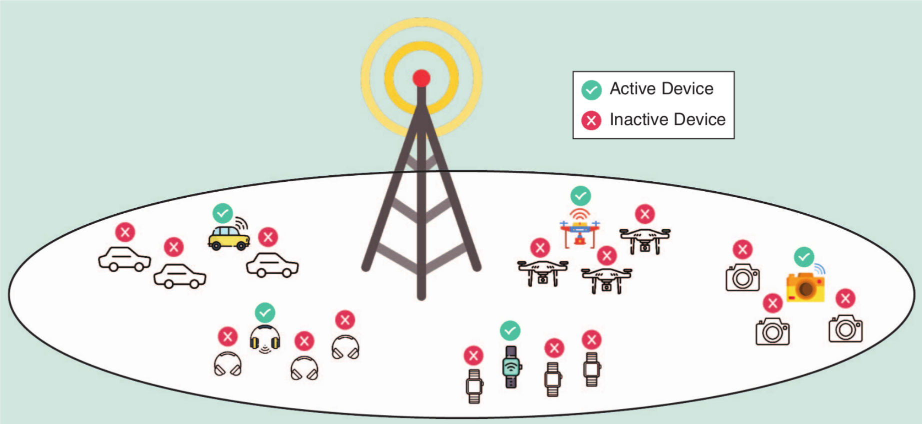
Technical University of Berlin, Germany

Berlin

July 8th, 2019

Outline

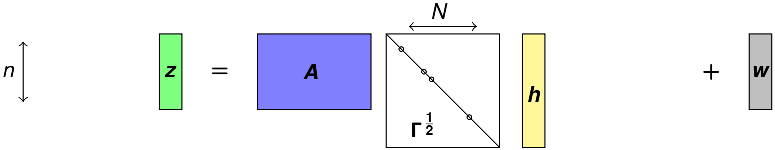
Introduction & Motivation



- Scenario where many devices/sensors send sporadic data to a central “city-wide” collector.
- Collector is a massive MIMO base station with a large number M of antennas.
- 1) activity detection; 2) large-scale fading coefficient 3) grant-free unsourced random access.

Introduction & Motivation

This simplified linear model is relevant:



$$z = A\Gamma^{\frac{1}{2}} h + w$$

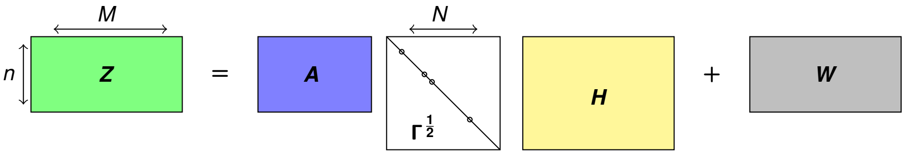
where

- $A = (a_1 | \dots | a_N) \in \mathbb{C}^{n \times N}$ sequence matrix (known), $n = 50 \dots 200$
- $h \in \mathbb{C}^N$ channel, and $w \in \mathbb{C}^n$ noise (unknown)
- $\Gamma = \text{diag}(\gamma)$ reflects **large scale fading** $\gamma \in \mathbb{R}_+^N$ (known/unknown)
- $\text{supp}(\gamma)$ is **activity pattern or data bits** (unknown)

Recent research focus is on:

- Bayesian approaches, like treat γ as iid. with “sparse” prior, unstable

This simplified linear model is relevant:



$$Z = A\Gamma^{\frac{1}{2}}H + W$$

where

- $A = (a_1|\dots|a_N) \in \mathbb{C}^{n\times N}$ sequence matrix (known), $n = 50\dots200$
- $H \in \mathbb{C}^{N\times M}$ channel, and $W \in \mathbb{C}^{n\times M}$ noise (unknown)
- $\Gamma = \text{diag}(\gamma)$ reflects **large scale fading** $\gamma \in \mathbb{R}_+^N$ (known/unknown)
- $\text{supp}(\gamma)$ is **activity pattern or data bits** (unknown)

Recent research focus is on:

- Bayesian approaches, like treat γ as iid. with "sparse" prior, unstable
- Non-Bayesian: treat γ as deterministic unknown nonnegative and sparse, recover γ

- **Raw measurement model**

$$Z = A\Gamma^{\frac{1}{2}} H + W$$

where unknown channel H and noise W are independent with iid entries

- for single antenna ($M = 1$)

$$z = A\Gamma^{\frac{1}{2}} h + w =: Ax + w \in \mathbb{C}^n \quad \text{with}$$

rule of thumb for s-sparse x: $\boldsymbol{n \gtrsim s \log(N/s)}$ for iid. subgaussian A

- **Raw measurement model**

$$Z = A\Gamma^{\frac{1}{2}} H + W$$

where unknown channel H and noise W are independent with iid entries

- for single antenna ($M = 1$)

$$z = A\Gamma^{\frac{1}{2}} h + w =: Ax + w \in \mathbb{C}^n \quad \text{with}$$

rule of thumb for s-sparse x: $\boldsymbol{n \gtrsim s \log(N/s)}$ for iid. subgaussian A

IDEA: Decoding via **sparse covariance matching**

- Each column z of Z is treated as realization of random vector with covariance

$$\begin{aligned}\Sigma_\gamma := \mathbb{E} zz^* &= A\Gamma^{\frac{1}{2}} \mathbb{E}(hh^*) \Gamma^{\frac{1}{2}} A^* + \mathbb{E}(ww^*) \\ &= A\Gamma A^* + \sigma^2 \mathrm{Id} =: \mathcal{A}(\gamma)\end{aligned}$$

- **Raw measurement model**

$$Z = A\Gamma^{\frac{1}{2}} H + W$$

where unknown channel H and noise W are independent with iid entries

- for single antenna ($M = 1$)

$$z = A\Gamma^{\frac{1}{2}} h + w =: Ax + w \in \mathbb{C}^n \quad \text{with}$$

rule of thumb for s-sparse x: $\boldsymbol{n \gtrsim s \log(N/s)}$ for iid. subgaussian A

IDEA: Decoding via **sparse covariance matching**

- Each column z of Z is treated as realization of random vector with covariance

$$\begin{aligned}\Sigma_\gamma &:= \mathbb{E}zz^* = A\Gamma^{\frac{1}{2}} \mathbb{E}(hh^*)\Gamma^{\frac{1}{2}} A^* + \mathbb{E}(ww^*) \\ &= A\Gamma A^* + \sigma^2 \mathrm{Id} =: \mathcal{A}(\gamma)\end{aligned}$$

- **Decode from:** empirical covariance $\hat{\Sigma}_\gamma := \frac{1}{M} ZZ^* = \Sigma_\gamma + E$ where E depends on γ
- “Compressed sensing problem” (informal version)

$$\text{find } \gamma \geq 0 \quad \text{such that} \quad \hat{\Sigma}_\gamma \approx \Sigma_\gamma = \mathcal{A}(\gamma)$$

- **Raw measurement model**

$$Z = A\Gamma^{\frac{1}{2}}H + W$$

where unknown channel H and noise W are independent with iid entries

- for single antenna ($M = 1$)

$$z = A\Gamma^{\frac{1}{2}}h + w =: Ax + w \in \mathbb{C}^n \quad \text{with}$$

rule of thumb for s-sparse x: $\boldsymbol{n \gtrsim s \log(N/s)}$ for iid. subgaussian A

IDEA: Decoding via **sparse covariance matching**

- Each column z of Z is treated as realization of random vector with covariance

$$\begin{aligned}\Sigma_\gamma &:= \mathbb{E}zz^* = A\Gamma^{\frac{1}{2}}\mathbb{E}(hh^*)\Gamma^{\frac{1}{2}}A^* + \mathbb{E}(ww^*)\\ &= A\Gamma A^* + \sigma^2\mathrm{Id} =: \mathcal{A}(\gamma)\end{aligned}$$

- **Decode from:** empirical covariance $\hat{\Sigma}_\gamma := \frac{1}{M}ZZ^* = \Sigma_\gamma + E$ where E depends on γ
- “Compressed sensing problem” (informal version)

$$\text{find } \gamma \geq 0 \quad \text{such that} \quad \hat{\Sigma}_\gamma \approx \Sigma_\gamma = \mathcal{A}(\gamma)$$

☺ This work: $\boldsymbol{s \lesssim n^2 \log^2(N/n^2)}$ for iid. subgaussian A and $M \simeq s$

Recovery of Nonnegative-Sparse Vectors from Matrix Observation

Background, BPDN Recovery via Nullspace Property

$\mathcal{A}:\mathbb{R}^N \mapsto \mathbb{C}^{n\times n}$ satisfies ℓ^q-**robust nullspace property** (ℓ^q-NSP, $q \geq 1$) of order s wrt $\|\cdot\|$ with parameters $\rho \in (0,1)$ and $\tau > 0$ if

$$\|v_S\|_q \leq \frac{\rho}{s^{1-1/q}} \|v_{S^c}\|_1 + \tau \|\mathcal{A}(v)\|$$

holds for all $v \in \mathbb{R}^N$ and $S \subset [N]$ with $|S| \leq s$

$\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{C}^{n\times n}$ satisfies ℓ^q-**robust nullspace property** (ℓ^q-NSP, $q \geq 1$) of order s wrt $\|\cdot\|$ with parameters $\rho \in (0,1)$ and $\tau > 0$ if

$$\|v_S\|_q \leq \frac{\rho}{s^{1-1/q}} \|v_{S^c}\|_1 + \tau \|\mathcal{A}(v)\|$$

holds for all $v \in \mathbb{R}^N$ and $S \subset [N]$ with $|S| \leq s$

Well-known results for BPDN [Foucart, Rauhut 2014]

$$x^\sharp = \arg\min \|z\|_1 \quad \text{s.t.} \quad \|\mathcal{A}(z) - Y\| \leq \epsilon \tag{BPDN}$$

If $\mathcal{A}$ has ℓ^q–NSP of order s with parameters (ρ, τ) wrt $\|\cdot\|$:

$$\left\|x^\sharp - x\right\|_p \lesssim \frac{C(\rho)}{s^{1-1/p}} \sigma_s(x)_1 + \frac{D(\rho)\tau}{s^{1/q-1/p}} \epsilon$$

holds for all $Y = \mathcal{A}(x) + E$ with $\|E\| \leq \epsilon$ and $1 \leq p \leq q$.

Background, BPDN Recovery via Nullspace Property

$\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{C}^{n \times n}$ satisfies ℓ^q-**robust nullspace property** (ℓ^q-NSP, $q \geq 1$) of order s wrt $\|\cdot\|$ with parameters $\rho \in (0,1)$ and $\tau > 0$ if

$$\|v_S\|_q \leq \frac{\rho}{s^{1-1/q}} \|v_{S^c}\|_1 + \tau \|\mathcal{A}(v)\|$$

holds for all $v \in \mathbb{R}^N$ and $S \subset [N]$ with $|S| \leq s$

Well-known results for BPDN [Foucart, Rauhut 2014]

$$x^\sharp = \arg\min \|z\|_1 \quad \text{s.t.} \quad \|\mathcal{A}(z) - Y\| \leq \epsilon \qquad \text{(BPDN)}$$

If $\mathcal{A}$ has ℓ^q–NSP of order s with parameters (ρ, τ) wrt $\|\cdot\|$:

$$\left\|x^\sharp - x\right\|_p \lesssim \frac{C(\rho)}{s^{1-1/p}} \sigma_s(x)_1 + \frac{D(\rho)\tau}{s^{1/q-1/p}} \epsilon$$

holds for all $Y = \mathcal{A}(x) + E$ with $\|E\| \leq \epsilon$ and $1 \leq p \leq q$.

- ☹ Problematic, if noise depends on x (Poisson noise, empirical covariance matching problems)
- ▶ Quotient property [Wojtaszczyk,2010] . . . or algorithmic approaches. . .

$\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{C}^{n\times n}$ satisfies ℓ^q-**robust nullspace property** (ℓ^q-NSP, $q \geq 1$) of order s wrt $\|\cdot\|$ with parameters $\rho \in (0, 1)$ and $\tau > 0$ if

$$\|v_S\|_q \leq \frac{\rho}{s^{1-1/q}} \|v_{S^c}\|_1 + \tau \|\mathcal{A}(v)\|$$

holds for all $v \in \mathbb{R}^N$ and $S \subset [N]$ with $|S| \leq s$

Well-known results for BPDN [Foucart, Rauhut 2014]

$$x^\sharp = \arg\min \|z\|_1 \quad \text{s.t.} \quad \|\mathcal{A}(z) - Y\| \leq \epsilon \qquad \text{(BPDN)}$$

If $\mathcal{A}$ has ℓ^q–NSP of order s with parameters (ρ, τ) wrt $\|\cdot\|$:

$$\left\|x^\sharp - x\right\|_p \lesssim \frac{C(\rho)}{s^{1-1/p}} \sigma_s(x)_1 + \frac{D(\rho)\tau}{s^{1/q-1/p}} \epsilon$$

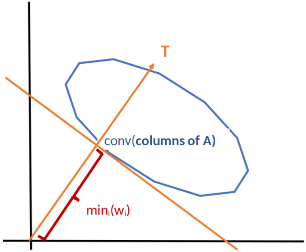
holds for all $Y = \mathcal{A}(x) + E$ with $\|E\| \leq \epsilon$ and $1 \leq p \leq q$.

- ☹ Problematic, if noise depends on x (Poisson noise, empirical covariance matching problems)
- ▶ Quotient property [Wojtaszczyk,2010] . . . or algorithmic approaches. . .
- ☺ **Non-negativity & “biased” measurement models** [Donoho92,Bruckstein04,. . .]

- "biased measurements" means

- “biased measurements” means
- ***“...origin does not belong to convex hull of the columns of* A*”*** [Donoho92,Bruckstein04,...
- Equivalent to $\mathcal{M}^+$-criterion

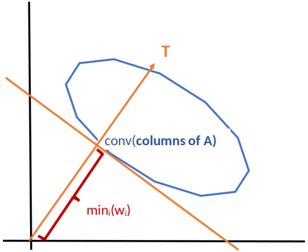
$$\exists T \text{ s.t. } w = \mathcal{A}^*(T) > 0$$



M^+-criterion makes ℓ_1-regularization superfluous

- "biased measurements" means
- ***"...origin does not belong to convex hull of the columns of* A*"*** [Donoho92,Bruckstein04,...
- Equivalent to $\mathcal{M}^+$-criterion

  $\exists T$ s.t. $w = \mathcal{A}^*(T) > 0$



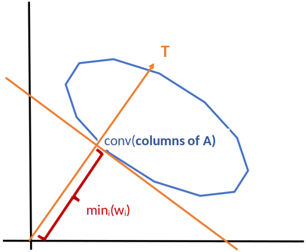
- Imagine a noiseless case with $w = 1 \simeq (1,1,1,1\dots)$:

$$\|x\|_1 \overset{x\geq 0}{=} \langle 1, x\rangle = \langle w, x\rangle = \langle \mathcal{A}^*(T), x\rangle = \langle T, \mathcal{A}(x)\rangle = \langle T, Y\rangle = \text{const},$$

no reason to minimize the ℓ_1-norm!!

- "biased measurements" means
- ***"...origin does not belong to convex hull of the columns of* A*"*** [Donoho92,Bruckstein04,...
- Equivalent to $\mathcal{M}^+$-criterion

  $\exists T$ s.t. $w = \mathcal{A}^*(T) > 0$



- Imagine a noiseless case with $w = 1 \simeq (1, 1, 1, 1 \dots)$:

$$\|x\|_1 \overset{x \geq 0}{=} \langle 1, x \rangle = \langle w, x \rangle = \langle \mathcal{A}^*(T), x \rangle = \langle T, \mathcal{A}(x) \rangle = \langle T, Y \rangle = \text{const},$$

  no reason to minimize the ℓ_1-norm!!
- For general $w > 0$ this will depend on

$$\kappa := \kappa(\mathcal{A}^*(T)) = \frac{\max w_i}{\min w_i}$$

- combining with nullspace property [Kabanava et al., 2016] and [Kueng and PJ, 2018]

Non-neg. Recovery Guarantee for Generic Norm

Non-neg. Recovery Guarantee for Generic Norm

Theorem ([Jaensch, PJ 2019], [Kueng, PJ 2018])

Let $\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{C}^{n\times n}$ *be a linear map with*

1. ℓ^q*–NSP of order* s *wrt norm* $\|\cdot\|$ *and parameter* $\rho \in [0,1)$ *and* $\tau > 0$ *and*
2. *fulfills* M^+*-criterion for* $T \in \mathbb{C}^{n\times n}$ *yielding* $\kappa = \kappa(\mathcal{A}^*(T))$

If $\rho\kappa < 1$ *then for any* $x \geq 0$ *and* E *the solution*

$$x^\sharp = \arg\min_{z\geq 0} \|\mathcal{A}(z) - Y\|. \quad (1)$$

for $Y = \mathcal{A}(x) + E$ *obeys*

$$\|x^\sharp - x\|_p \leq \frac{C'\sigma_s(x)_1}{s^{1-\frac{1}{p}}} + \frac{D'\left(\tau + \frac{\|\mathcal{A}^*(T)\|_\infty^{-1}\|T\|^\circ}{s^{1-\frac{1}{q}}}\right)}{s^{\frac{1}{q}-\frac{1}{p}}}\|E\|$$

for all $1 \leq p \leq q$*, where* $C' := 2\frac{(1+\kappa\rho)^2}{1-\kappa\rho}\kappa$ *and* $D' := 2\frac{3+\kappa\rho}{1-\kappa\rho}\kappa$.

- Almost same guarantees as for optimally tuned BPDN, i.e. with instantaneous noise bound

Non-neg. Recovery Guarantee for Generic Norm

Theorem ([Jaensch, PJ 2019], [Kueng, PJ 2018])

Let $\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{C}^{n\times n}$ *be a linear map with*

1. ℓ^q*–NSP of order s wrt norm* $\|\cdot\|$ *and parameter* $\rho \in [0,1)$ *and* $\tau > 0$ *and*
2. *fulfills* M^+*-criterion for* $T \in \mathbb{C}^{n\times n}$ *yielding* $\kappa = \kappa(\mathcal{A}^*(T))$

If $\rho\kappa < 1$ *then for any* $x \geq 0$ *and E the solution*

$$x^\sharp = \arg\min_{z\geq 0} \|\mathcal{A}(z) - Y\|. \tag{1}$$

for $Y = \mathcal{A}(x) + E$ *obeys*

$$\|x^\sharp - x\|_p \leq \frac{C'\sigma_s(x)_1}{s^{1-\frac{1}{p}}} + \frac{D'\left(\tau + \frac{\|\mathcal{A}^*(T)\|_\infty^{-1}\|T\|^\circ}{s^{1-\frac{1}{q}}}\right)}{s^{\frac{1}{q}-\frac{1}{p}}}\|E\|$$

for all $1 \leq p \leq q$*, where* $C' := 2\frac{(1+\kappa\rho)^2}{1-\kappa\rho}\kappa$ *and* $D' := 2\frac{3+\kappa\rho}{1-\kappa\rho}\kappa$.

- Almost same guarantees as for optimally tuned BPDN, i.e. with instantaneous noise bound
- Strength of “self-tuning” depends on dimension-scaling of $\|\mathcal{A}^*(T)\|_\infty^{-1}\|T\|^\circ$.

Non-neg. Recovery Guarantee for Generic Norm

- For non-negative mixture of positive-semidefinite matrices $A_i \succ 0$:

$$\mathcal{A}(x) = \sum_{i=1}^{N} x_i A_i \quad x \geq 0$$

we have $\mathcal{A}^*(\text{Id}) = (\text{Tr}A_i)_{i=1}^{N} > 0 \Rightarrow$ M^+ is trivially fulfilled, and

$$\min_{z \geq 0} \|\mathcal{A}(z) - Y\|$$

is very useful for robust sparse covariance matching problems

- For non-negative mixture of positive-semidefinite matrices $A_i \succ 0$:

$$\mathcal{A}(x) = \sum_{i=1}^{N} x_i A_i \quad x \geq 0$$

we have $\mathcal{A}^*(\mathrm{Id}) = (\mathrm{Tr} A_i)_{i=1}^N > 0 \Rightarrow M^+$ is trivially fulfilled, and

$$\min_{z \geq 0} \|\mathcal{A}(z) - Y\|$$

is very useful for robust sparse covariance matching problems

- **NSP &** $M^+ \Rightarrow$ *"for all s-sparse $x \geq 0$ and all $z \geq 0$ it holds"*

$$\|z - x\|_2 \lesssim \kappa \left(\tau + \frac{\|T\|^\circ}{\|\mathcal{A}^*(T)\|_\infty \sqrt{s}} \right) \|\mathcal{A}(z - x)\|$$

Rankone Subgaussian Random Model

- Sparse non-negative mixture of $\{a_i a_i^*\}_{i=1}^N$:

$$\mathcal{A}(x) = \sum_{i=1}^{N} x_i a_i a_i^* \quad x \geq 0$$

$\mathcal{A}$ is also called as (columnwise) self-**Khatri-Rao product** of $(a_1|\dots|a_N)$.

Random model for the sequences

- $\{a_i\}_{i=1}^N \subset \mathbb{C}^n$ independent with iid. circular-symmetric **subgaussian entries** (real/imag. independent, zero mean, variance $\frac{1}{2}$, ψ_2–norm at most ψ_2)
- $\mathcal{A}^*(T) = (\langle a_i, Ta_i\rangle)_{i=1}^N \overset{T=\text{Id}}{=} (\|a_i\|_2^2)_{i=1}^N \Rightarrow$ M^+-criterion holds whp for $\kappa \approx 1\dots$

Rankone Subgaussian Random Model

- Sparse non-negative mixture of $\{a_i a_i^*\}_{i=1}^N$:

$$\mathcal{A}(x) = \sum_{i=1}^{N} x_i a_i a_i^* \quad x \geq 0$$

$\mathcal{A}$ is also called as (columnwise) self-**Khatri-Rao product** of $(a_1|\ldots|a_N)$.

Random model for the sequences

- $\{a_i\}_{i=1}^N \subset \mathbb{C}^n$ independent with iid. circular-symmetric **subgaussian entries** (real/imag. independent, zero mean, variance $\frac{1}{2}$, ψ_2–norm at most ψ_2)
- $\mathcal{A}^*(T) = (\langle a_i, Ta_i\rangle)_{i=1}^N \overset{T=\text{Id}}{=} (\|a_i\|_2^2)_{i=1}^N \Rightarrow$ M^+-criterion holds whp for $\kappa \approx 1\ldots$

Lemma (M^+-criterion holds whp for $\kappa \approx 1$)

For $T \sim Id$ and $\eta \in (0,1)$ it holds that $\kappa(\mathcal{A}^*(T)) \leq \frac{1+\eta}{1-\eta}$ with probability at least

$$1 - N\exp(-cn \cdot \min\{\frac{\eta^2}{\psi_2^4}, \frac{\eta}{\psi_2^2}\}).$$

Beyond independent components, "convex concentration property" [Adamczak, 2015]

Steps to proof Nullspace Properties

Nullspace Property...

- From our assumption: $\mathcal{A}$ has "independent columns" $\text{vec}(a_i a_i^*) \in \mathbb{C}^{n^2}$
- $a_i = (a_{i,1}, \dots, a_{i,n}) \in \mathbb{C}^n$ is subgaussian
- $a_i a_i^* \in \mathbb{C}^{n \times n}$ is subexponential
- $\mathbb{E}\|a_i a_i\|_F^2 = \mathbb{E}\|a_i\|_2^4 = n^2$

Nullspace Property...

- From our assumption: $\mathcal{A}$ has "independent columns" $\text{vec}(a_i a_i^*) \in \mathbb{C}^{n^2}$
- $a_i = (a_{i,1}, \ldots, a_{i,n}) \in \mathbb{C}^n$ is subgaussian
- $a_i a_i^* \in \mathbb{C}^{n \times n}$ is subexponential
- $\mathbb{E}\|a_i a_i\|_F^2 = \mathbb{E}\|a_i\|_2^4 = n^2$
- well–known: (after normalization) sufficient RIP implies NSP [Foucart, Rauhut 2014]

  Thus, if $\delta_{2s}(\Phi) \leq \delta < \frac{4}{\sqrt{41}}$ then Φ has ℓ^2–NSP of order s wrt $\|\cdot\|_2$ with parameters

$$\rho \leq \frac{\delta}{\sqrt{1-\delta^2} - \delta/4} \quad \text{and} \quad \tau \leq \frac{\sqrt{1+\delta}}{\sqrt{1-\delta^2} - \delta/4}.$$

- RIP for independent heavy-tailed columns...

RIP for heavy-tailed column-independent matrices

Theorem ([Adamczak et al. 2011])

Let $X_1, \dots, X_N \in \mathbb{R}^m$ be independent ψ_1-random vectors with $\mathbb{E}\{\|X_i\|_2^2\} = m$ and let $\psi = \max_{i \le N} \|X_i\|_{\psi_1}$. Let $\theta' \in (0,1)$, $K, K' \geq 1$ and set $\xi = \psi K + K'$. Then for $\mathcal{A} := (X_1|\dots|X_N)$

$$\delta_s\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq C\xi^2 \sqrt{\frac{s}{m}} \log\left(\frac{eN}{s\sqrt{\frac{s}{m}}}\right) + \theta' \tag{2}$$

holds with probability larger then

$$1 - \exp(-cK\sqrt{s}\log(\frac{eN}{s\sqrt{\frac{s}{m}}})) \tag{3}$$

$$- \mathbb{P}(\max_{i \le N} \|X_i\|_2 \geq K'\sqrt{m}) - \mathbb{P}(\max_{i \le N} \left| \frac{\|X_i\|_2^2}{m} - 1 \right| \geq \theta'), \tag{4}$$

where $C, c > 0$ are universal constants.

Adamczak, Litvak, Pajor and Tomczak-Jaegermann, “Restricted Isometry Property of Matrices with Independent Columns and Neighborly Polytopes by Random Sampling”

RIP for heavy-tailed column-independent matrices

Theorem ([Adamczak et al. 2011])

Let $X_1, \ldots, X_N \in \mathbb{R}^m$ be independent ψ_1-random vectors with $\mathbb{E}\{\|X_i\|_2^2\} = m$ and let $\psi = \max_{i \leq N} \|X_i\|_{\psi_1}$. Let $\theta' \in (0,1)$, $K, K' \geq 1$ and set $\xi = \psi K + K'$. Then for $\mathcal{A} := (X_1|\ldots|X_N)$

$$\delta_s\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq C\xi^2\sqrt{\frac{s}{m}}\log\left(\frac{eN}{s\sqrt{\frac{s}{m}}}\right) + \theta' \tag{2}$$

holds with probability larger then

$$1 - \exp(-cK\sqrt{s}\log(\frac{eN}{s\sqrt{\frac{s}{m}}})) \tag{3}$$

$$- \mathbb{P}(\max_{i \leq N}\|X_i\|_2 \geq K'\sqrt{m}) - \mathbb{P}(\max_{i \leq N}\left|\frac{\|X_i\|_2^2}{m} - 1\right| \geq \theta'), \tag{4}$$

where $C, c > 0$ are universal constants.

Adamczak, Litvak, Pajor and Tomczak-Jaegermann, "Restricted Isometry Property of Matrices with Independent Columns and Neighborly Polytopes by Random Sampling"

- ☹ $\mathcal{A}$ can not have optimal RIP since $\mathbb{E}a_i a_i^* = \mathrm{Id}$ (which was essential for $M^+ \ldots$)
- ☹ Equiv., $\|a_i a_i^*\|_{\psi_1} = \sup_{\|Z\|_F \leq 1} \|\langle a_i, Z a_i \rangle\|_{\psi_1} = \mathcal{O}(\sqrt{n})$ depends on dimension
- ☺ but optimal NSP is still feasible

- ☹ $\mathcal{A}$ can not have optimal RIP since $\mathbb{E}a_i a_i^* = \mathrm{Id}$ (which was essential for $M^+ \ldots$)
- ☹ Equiv., $\|a_i a_i^*\|_{\psi_1} = \sup_{\|Z\|_F \leq 1} \|\langle a_i, Z a_i \rangle\|_{\psi_1} = \mathcal{O}(\sqrt{n})$ depends on dimension
- ☺ but optimal NSP is still feasible
- ▶ two options to get RIP statements here. . .

1. **centering** (of independent interest. . .) and/or
2. considering only **off-diagonal part**, better for proving NSP in our case. . .

1. **Centering**

Theorem ([Fengler, PJ 2019])

Let $A = (a_1|\dots|a_N) \in \mathbb{R}^{n\times N}$ with iid ψ_2-entries a_{ij} with $\mathbb{E}a_{ij} = 0$, $\mathbb{E}a_{ij}^2 = 1$ and $\|a_{ij}\|_{\psi_2} \leq \psi_2$.
Let $\mathcal{A} = (X_1|\dots|X_N) \in \mathbb{R}^{n^2\times N}$ be the centered **self-Khatri-Rao product** *of A, i.e., with iid. columns*

$$X_i \simeq vec(a_i a_i^* - \mathbb{E}a_i a_i^*)$$

Then $\delta_s\left(\frac{\mathcal{A}}{n}\right) < \delta$ for any $\delta > 0$ with probability $\geq 1 - C\exp(-c_\delta n/\psi_2^2)$ as long as

$$s \lesssim n^2/\log^2(N/n^2)$$

- similar statement if a_i is drawn uniformly from sphere of radius $\sqrt{n}$ (no iid. components anymore...)

1. **Off-diagonal part** $X_i \simeq P(a_i a_i^*)$ where $P : \mathbb{C}^{n\times n} \to \mathbb{R}^{2n(n-1)}$

Theorem ([Jaensch, PJ 2019])

Let $\mathcal{A}(x) = \sum_{i=1}^N x_i a_i a_i^*$ *with* $a_i \in \mathbb{C}^n$ *independent circular-symmetric and* $\max_{i\in[N]} \|a_i\|_{\psi_2} \leq \psi_2$*. Let* $m = 2n(n-1)$ *and assume*

$$s \lesssim m \log^{-2}(N/m)$$

Then, with probability $\geq 1 - \exp(-cn/\psi_2^2)$ *(i) the matrix* $\Phi = \frac{1}{\sqrt{m}} P \circ \mathcal{A} \in \mathbb{R}^{m\times N}$ *has RIP and (ii)* $\mathcal{A}$ *has the* ℓ^2*-NSP of order s w.r.t. to* $\|\cdot\|_F$ *with parameters* ρ *and* $\tau/\sqrt{m}$ *(depending on the RIP constant* δ_{2s}*).*

1. **Off-diagonal part** $X_i \simeq P(a_i a_i^*)$ where $P : \mathbb{C}^{n\times n} \to \mathbb{R}^{2n(n-1)}$

Theorem ([Jaensch, PJ 2019])

Let $\mathcal{A}(x) = \sum_{i=1}^N x_i a_i a_i^*$ with $a_i \in \mathbb{C}^n$ independent circular-symmetric and $\max_{i\in[N]} \|a_i\|_{\psi_2} \leq \psi_2$. Let $m = 2n(n-1)$ and assume

$$s \lesssim m \log^{-2}(N/m)$$

Then, with probability $\geq 1 - \exp(-cn/\psi_2^2)$ (i) the matrix $\Phi = \frac{1}{\sqrt{m}} P \circ \mathcal{A} \in \mathbb{R}^{m\times N}$ has RIP and (ii) $\mathcal{A}$ has the ℓ^2-NSP of order s w.r.t. to $\|\cdot\|_F$ with parameters ρ and $\tau/\sqrt{m}$ (depending on the RIP constant δ_{2s}).

- (ii) is obvious since RIP for Φ implies NSP for $\mathcal{A}$:

$$\begin{aligned}\|x_S\|_2 &\leq \frac{\rho}{\sqrt{s}}\|x_{S^c}\|_1 + \tau\|\Phi x\|_2 = \frac{\rho}{\sqrt{s}}\|x_{S^c}\|_1 + \frac{\tau}{\sqrt{m}}\|P(\mathcal{A}(x))\|_2 . \\ &\leq \frac{\rho}{\sqrt{s}}\|x_{S^c}\|_1 + \frac{\tau}{\sqrt{m}}\|\mathcal{A}(x)\|_F\end{aligned}$$

$X_i \simeq P(a_i a_i^*) \in \mathbb{R}^{2n(n-1)}$, concentration of the polynomial

$$\|X_i\|_2^2 \simeq \sum_{k \neq l}^{n} (|a_{i,k}||a_{i,l}|)^2 = \sum_{(k,l)\in\mathcal{I}} w_k^2 w_l^2 =: f(w) \tag{6}$$

for $w_k = [\mathrm{Re}(a_{i,k}), \mathrm{Im}(a_{i,k})]$ and $\mathcal{I} = \{(k,l) \in [2n] \times [2n] : k \neq l, k \neq n+l, l \neq n+k\}$

RIP for Off-Diagonal Part

$X_i \simeq P(a_i a_i^*) \in \mathbb{R}^{2n(n-1)}$, concentration of the polynomial

$$\|X_i\|_2^2 \simeq \sum_{k\neq l}^{n} (|a_{i,k}||a_{i,l}|)^2 = \sum_{(k,l)\in\mathcal{I}} w_k^2 w_l^2 =: f(w) \tag{6}$$

for $w_k = [\mathrm{Re}(a_{i,k}), \mathrm{Im}(a_{i,k})]$ and $\mathcal{I} = \{(k,l) \in [2n] \times [2n] : k \neq l, k \neq n+l, l \neq n+k\}$

Lemma

Let $w = (w_1, \ldots, w_{2N})$ be RV with independent $\|w_i\|_{\psi_2} \leq \psi_2$. The polynomial $f(w)$ in (6) fulfills

$$\mathbb{P}\left(\frac{1}{m}|f(w) - \mathbb{E}f(w)| \geq t\right) \leq 2\exp(-c\sqrt{t} \cdot n/\psi_2^2)$$

where $m = 2n(n-1)$.

- Follows from [Götze, Sambala & Sinulis, 2019], [Adamczak & Wolf, 2015]

Theorem ([Jaensch, PJ 2019])

Let $\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{C}^{n \times n}$ *be defined as*

$$\mathcal{A}(x) = \sum_{i=1}^{N} x_i a_i a_i^*$$

where $\{a_i\}_{i=1}^N \subset \mathbb{C}^n$ *are independent with zero-mean, independent circular-symmetric* ψ_2*-entries of unit variance. If* $s \lesssim n^2 / \log^2(N/n^2)$ *the following holds with probability* $\geq 1 - \exp(-cn/\psi_2^2)$*. For all* $x \geq 0$ *and all* E *the solution*

$$x^\sharp = \arg\min_{z \geq 0} \|\mathcal{A}(z) - Y\|_F \quad \text{for} \quad Y = \mathcal{A}(x) + E \tag{NNLS}$$

obeys for $1 \leq p \leq 2$ *the following bound:*

$$\|x^\sharp - x\|_p \leq \frac{c_1 \sigma_s(x)_1}{s^{1-\frac{1}{p}}} + \frac{c_2(c_3 + \sqrt{\frac{n}{s}})}{s^{\frac{1}{2}-\frac{1}{p}}} \frac{\|E\|_F}{n}$$

Recovery Guarantee

Theorem ([Jaensch, PJ 2019])

Let $\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{C}^{n\times n}$ *be defined as*

$$\mathcal{A}(x) = \sum_{i=1}^{N} x_i a_i a_i^*$$

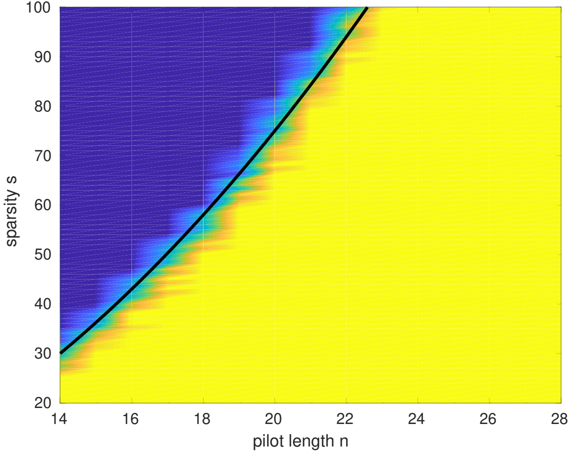
where $\{a_i\}_{i=1}^N \subset \mathbb{C}^n$ *are independent with zero-mean, independent circular-symmetric* ψ_2*-entries of unit variance. If* $s \lesssim n^2/\log^2(N/n^2)$ *the following holds with probability* $\geq 1 - \exp(-cn/\psi_2^2)$*. For all* $x \geq 0$ *and all* E *the solution*

$$x^\sharp = \arg\min_{z\geq 0} \|\mathcal{A}(z) - Y\|_F \quad \text{for} \quad Y = \mathcal{A}(x) + E \qquad \text{(NNLS)}$$

obeys for $1 \leq p \leq 2$ *the following bound:*

$$\|x^\sharp - x\|_p \leq \frac{c_1\sigma_s(x)_1}{s^{1-\frac{1}{p}}} + \frac{c_2(c_3 + \sqrt{\frac{n}{s}})}{s^{\frac{1}{2}-\frac{1}{p}}}\frac{\|E\|_F}{n}$$

- one can show $\mathbb{E}\|E\|_F^2 = \mathbb{E}\|\hat{\Sigma}_\gamma - \Sigma_\gamma\|_F^2 \leq n\|\gamma\|_1/M$ where $M = \#\text{antennas}$
- this yields $\frac{\|\gamma-\gamma^\sharp\|_1}{\|\gamma\|_1} \lesssim (\sqrt{n} + \sqrt{s})/\sqrt{M} \quad \Rightarrow M \sim s$



Phase transition for NNLS in the noiseless case ($N = 2000$). The function $s \approx n^2/4 - n - 5$ is overlayed in black.

Conclusion

- massive antennas support activity $s \simeq n^2$ (instead of $s \simeq n$) with random pilots of length n
- sparse covariance matching problem formulated as tuning-free convex program
- guarantees depend on M^+-criterion and nullspace properties
- both hold whp for centered subgaussian iid. pilots
- RIP for column-independent subexp. matrices with particular $\otimes$-structure

Thanks for Your Attention

- Fengler and PJ. *"On the Restricted Isometry Property of Centered Self Khatri-Rao Products"*, arXiv:1905.09245
- Jaensch and PJ. *"Robust Recovery of Sparse Nonnegative Weights from Mixtures of Positive-Semidefinite Matrices"*, soon on arXiv
- Haghighatshoar, PJ and Caire, *"Improved Scaling Law for Activity Detection in Massive MIMO Systems"*, ISIT2018

based on

- Kueng and PJ. *"Robust nonnegative sparse recovery and the nullspace property of 0/1 measurements"*, 2018
- Adamczak, Litvak, Pajor, and Tomczak-Jaegermann *"Restricted Isometry Property of Matrices with Independent Columns and Neighborly Polytopes by Random Sampling"*, 2011
- Götze, Sambale, and Sinulis. *"Concentration inequalities for polynomials in α-sub-exponential random variables"* arXiv 1903.05964
- Adamczak and Wolff. *"Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order"*, 2019

Peter Jung // TU Berlin, Germany [peter.jung@tu-berlin.de]