Blind Sparse Recovery From Superimposed Non-Linear Sensor Measurements

Peter Jung
(Technische Universität Berlin)

SAMPTA 2017
Tallinn (Estonia)
July 4, 2017

Collaborators/Funding

Martin Genzel

(Research Group Applied Functional Analysis @ TU Berlin)

Outline of This Talk

1. Motivation & Measurement Model
2. The Direct Method
3. The Lifting Method
4. Extensions & Conclusion

Motivation

Environmental monitoring problems are of great importance in many practical applications.

Examples:

- Heat, Fire, and Seismic Monitoring
- Tsunami Early Warning Systems
- Structural Health Monitoring
- Medical Sensor Solutions

Motivation

Environmental monitoring problems are of great importance in many practical applications.

Examples:

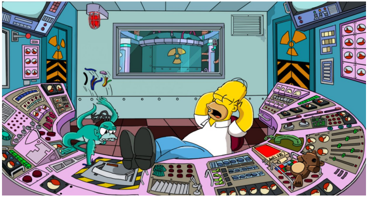
- Heat, Fire, and Seismic Monitoring
- Tsunami Early Warning Systems
- Structural Health Monitoring
- Medical Sensor Solutions
- distributed sensing & imperfect communication

Motivation

Environmental monitoring problems are of great importance in many practical applications.

Examples:

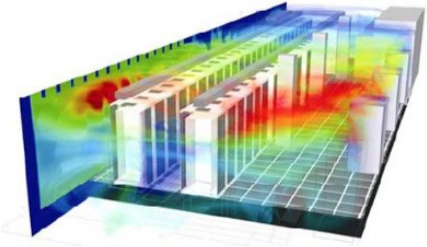
- Heat, Fire, and Seismic Monitoring
- Tsunami Early Warning Systems
- Structural Health Monitoring
- Medical Sensor Solutions
- distributed sensing & imperfect communication



Wireless sensor networks form a promising approach to many of those problems.

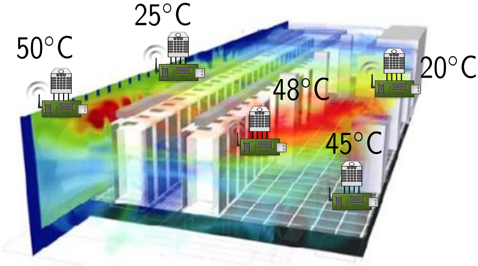
Wireless Sensor Networks

Example: Monitor the heat distribution of a server room



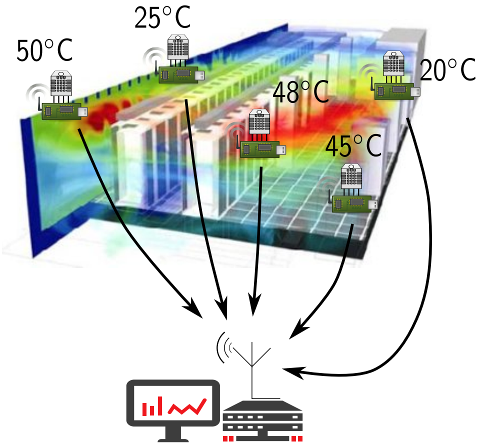
Wireless Sensor Networks

Spatially distributed sensors measure the temperature...



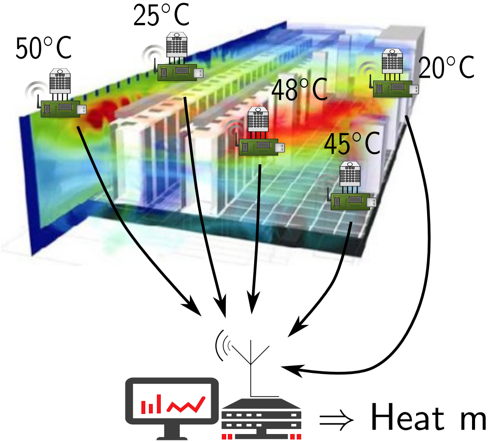
Wireless Sensor Networks

... and transmit their measurements simultaneously to a central receiver.



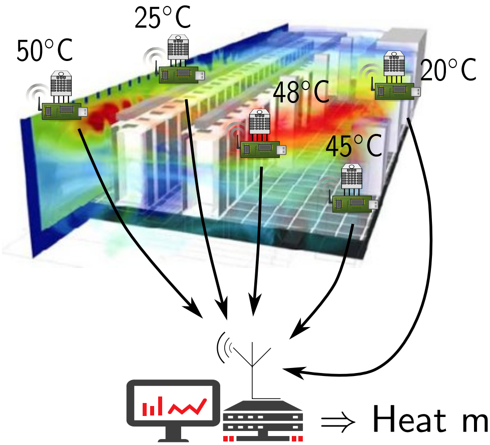
Wireless Sensor Networks

Challenge: Reconstruct the heat map!

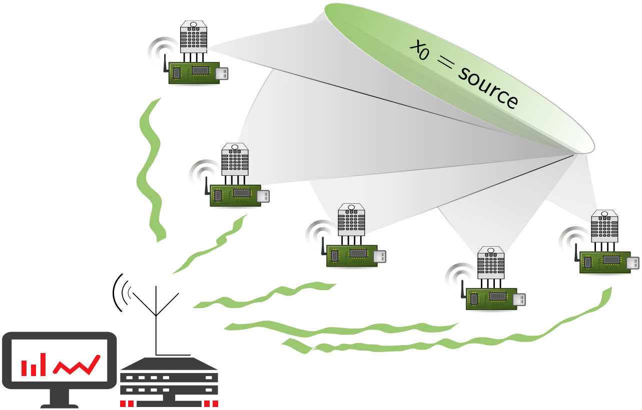


$\Rightarrow$ Heat map

The signal-of-interest is often (approximately) sparse in a known transform domain (e.g., Fourier or wavelets)



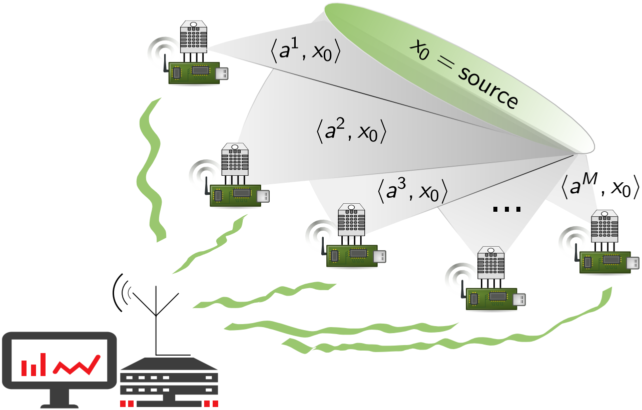
⇒ Heat map



Unknown source vector

$\boldsymbol{x}_0$

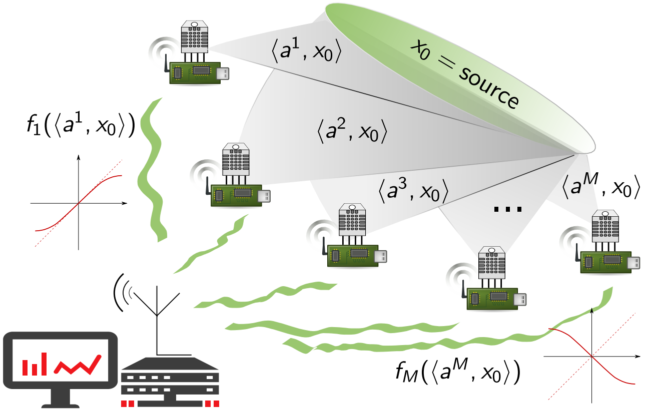
Towards a Measurement Model...



Each sensor node takes linear measurements $\rightsquigarrow$ "perspectives"
(autonomous and ad hoc)

$$\langle \boldsymbol{a}^j, \boldsymbol{x}_0 \rangle$$

Towards a Measurement Model...

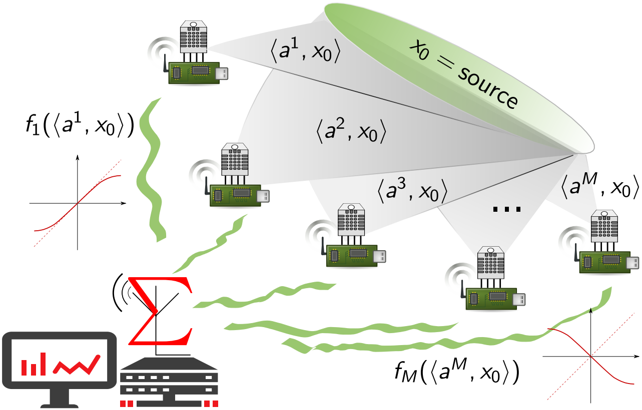


Transmission of raw data (uncoded)

Non-linear distortion due to hardware imperfections and wireless channel

$$f_j(\langle \boldsymbol{a}^j, \boldsymbol{x}_0\rangle)$$

Towards a Measurement Model...



Superposition of signals at the central receiver + noise $\rightsquigarrow$ Sample

$$y = \sum_{j=1}^{M} f_j(\langle \mathbf{a}^j, \mathbf{x}_0 \rangle) + e$$

Why this setup?

- Heuristics in the linear&coherent case: Enlarging the network improves the receive SNR! But, with nonlinearities. . . ?

$$y = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}^j, \boldsymbol{x}_0 \rangle) + e$$

Towards a Measurement Model...

Why this setup?

- Heuristics in the linear&coherent case: Enlarging the network improves the receive SNR! But, with nonlinearities. . . ?
- But we shall see that similar conclusion is true for the reconstruction of $\boldsymbol{x}_0$ from non-linear measurements when M grows.

$$y = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}^j, \boldsymbol{x}_0 \rangle) + e$$

Why this setup?

- Heuristics in the linear&coherent case: Enlarging the network improves the receive SNR! But, with nonlinearities. . . ?
- But we shall see that similar conclusion is true for the reconstruction of $\boldsymbol{x}_0$ from non-linear measurements when M grows.
- Cope with phase instabilities $\rightarrow$ noncoherent case

$$y = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}^j, \boldsymbol{x}_0 \rangle) + e$$

Model Setup and Problem Statement

Superimposed non-linearly distorted measurements:

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \dots, m$$

- $\boldsymbol{x}_0 \in \mathbb{R}^n$: unknown source vector (sparse)
- $\boldsymbol{a}_i^j \in \mathbb{R}^n$: i-th measurement vector of the j-th senor

This talk: $\boldsymbol{a}_i^j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$ *i.i.d.*

- $f_j : \mathbb{R} \to \mathbb{R}$: scalar distortion (non-linear and possibly unknown)
- $\mathbb{E}[f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle)] = 0$
- e_i: independent measurement noise

Model Setup and Problem Statement

Superimposed non-linearly distorted measurements:

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \ldots, m$$

- $\boldsymbol{x}_0 \in \mathbb{R}^n$: unknown source vector (sparse)
- $\boldsymbol{a}_i^j \in \mathbb{R}^n$: i-th measurement vector of the j-th senor

This talk: $\boldsymbol{a}_i^j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$ *i.i.d.*

- $f_j : \mathbb{R} \to \mathbb{R}$: scalar distortion (non-linear and possibly unknown)
- $\mathbb{E}[f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle)] = 0$
- e_i: independent measurement noise

Model Setup and Problem Statement

Superimposed non-linearly distorted measurements:

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \ldots, m$$

- $\boldsymbol{x}_0 \in \mathbb{R}^n$: unknown source vector (sparse)
- $\boldsymbol{a}_i^j \in \mathbb{R}^n$: i-th measurement vector of the j-th senor

This talk: $\boldsymbol{a}_i^j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ *i.i.d.*

- $f_j : \mathbb{R} \to \mathbb{R}$: scalar distortion (non-linear and possibly unknown)
- $\mathbb{E}[f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle)] = 0$
- e_i: independent measurement noise

Can we recover $\boldsymbol{x}_0$ *from* $\{(\boldsymbol{a}_i^1, \ldots, \boldsymbol{a}_i^M; y_i)\}_{i=1\ldots m}$*?*

How many measurements m *do we need?*
And what is the impact of the sensor count M*?*

The Direct Method

Key Idea: Mimicking the Linear Case

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \ldots, m$$

- In the linear case ($f_j = \mathrm{Id}$), we have $y_i = \langle \sum_{j=1}^{M} \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle + e_i$.

Key Idea: Mimicking the Linear Case

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \dots, m$$

- In the linear case ($f_j = \mathrm{Id}$), we have $y_i = \langle \sum_{j=1}^{M} \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle + e_i$.
- Idea: Consider superimposed measurement vectors

$$\boldsymbol{a}_i := \sum_{j=1}^{M} \boldsymbol{a}_i^j, \quad i = 1, \dots, m,$$

and just solve the vanilla Lasso:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \sum_{i=1}^{m} (y_i - \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle)^2 \quad \text{subject to } \|\boldsymbol{x}\|_1 \leq R$$

Key Idea: Mimicking the Linear Case

$$y_i = \sum_{j=1}^M f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \dots, m$$

- In the linear case ($f_j = \mathrm{Id}$), we have $y_i = \langle \sum_{j=1}^M \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle + e_i$.
- Idea: Consider superimposed measurement vectors

$$\boldsymbol{a}_i := \sum_{j=1}^M \boldsymbol{a}_i^j, \quad i = 1, \dots, m,$$

and just solve the vanilla Lasso:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \sum_{i=1}^m (y_i - \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle)^2 \quad \text{subject to } \|\boldsymbol{x}\|_1 \leq R$$

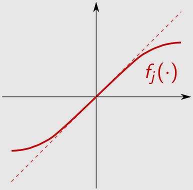
We "fool" the Lasso by fitting a linear ansatz to non-linear measurements. Can such an approach work?

"Model Mismatch = Noise"

Definition (The Mismatch Parameters)

For each node $j = 1, \ldots, M$, we define the scaling parameter

$$\mu_j := \mathbb{E}[f_j(g) \cdot g], \quad g \sim \mathcal{N}(0,1),$$



and the mean scaling parameter

$$\mu = \frac{1}{M} \sum_{j=1}^{M} \mu_j \; .$$

The deviation parameter is given by

$$\sigma^2 := \frac{1}{M} \sum_{j=1}^{M} \| f_j(g) - \mu g \|_{\psi_2}^2, \quad g \sim \mathcal{N}(0,1).$$

- μ_j measures how well "aligns" f_j to Id in expectation.
- If $\|\boldsymbol{x}_0\|_2 = 1$, we have $g := \langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle \sim \mathcal{N}(0,1)$.

This concept originates from Plan and Vershynin (2016) and some extensions by G. (2017).

Recovery via the Direct Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ be s-sparse, $\|\boldsymbol{x}_0\|_2 = 1$, and $e_i \sim \mathcal{N}(0, \nu^2)$. For every $\delta \in (0, 1]$, the following holds true with probability at least $1 - 5\exp(-C\delta^2 m)$: If

$$m \gtrsim \delta^{-2} s \log(\tfrac{2n}{s}),$$

then any minimizer $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ of the Lasso with $R = \|\mu \boldsymbol{x}_0\|_1$ satisfies

$$\|\hat{\boldsymbol{x}} - \mu \boldsymbol{x}_0\|_2 \leq (\sigma^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

$$\mu = \tfrac{1}{M} \textstyle\sum_{j=1}^{M} \mu_j, \qquad \sigma^2 := \tfrac{1}{M} \textstyle\sum_{j=1}^{M} \|f_j(g) - \mu g\|_{\psi_2}^2$$

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} \sum_{i=1}^{m} (y_i - \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle)^2 \quad \text{subject to } \|\boldsymbol{x}\|_1 \leq R$$

Recovery via the Direct Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ be s-sparse, $\|\boldsymbol{x}_0\|_2 = 1$, and $e_i \sim \mathcal{N}(0, \nu^2)$. For every $\delta \in (0, 1]$, the following holds true with probability at least $1 - 5\exp(-C\delta^2 m)$: If

$$m \gtrsim \delta^{-2} s \log(\tfrac{2n}{s}),$$

then any minimizer $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ of the Lasso with $R = \|\mu \boldsymbol{x}_0\|_1$ satisfies

$$\|\hat{\boldsymbol{x}} - \mu \boldsymbol{x}_0\|_2 \leq (\sigma^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

Recovery becomes feasible if $m \gtrsim s \log(\frac{2n}{s})$!

Recovery via the Direct Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ be s-sparse, $\|\boldsymbol{x}_0\|_2 = 1$, and $e_i \sim \mathcal{N}(0, \nu^2)$. For every $\delta \in (0, 1]$, the following holds true with probability at least $1 - 5\exp(-C\delta^2 m)$: If

$$m \gtrsim \delta^{-2} s \log(\tfrac{2n}{s}),$$

then any minimizer $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ of the Lasso with $R = \|\mu \boldsymbol{x}_0\|_1$ satisfies

$$\|\hat{\boldsymbol{x}} - \mu \boldsymbol{x}_0\|_2 \leq (\sigma^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

Recovery becomes feasible if $m \gtrsim s \log(\frac{2n}{s})$!

- The non-linearities f_j and the node count M affect the error bound only in terms of the rescaling factor μ and the variance σ^2.

Recovery via the Direct Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ be s-sparse, $\|\boldsymbol{x}_0\|_2 = 1$, and $e_i \sim \mathcal{N}(0, \nu^2)$. For every $\delta \in (0,1]$, the following holds true with probability at least $1 - 5\exp(-C\delta^2 m)$: If

$$m \gtrsim \delta^{-2} s \log(\tfrac{2n}{s}),$$

then any minimizer $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ of the Lasso with $R = \|\mu \boldsymbol{x}_0\|_1$ satisfies

$$\|\hat{\boldsymbol{x}} - \mu \boldsymbol{x}_0\|_2 \leq (\sigma^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

Recovery becomes feasible if $m \gtrsim s \log(\frac{2n}{s})$!

- The non-linearities f_j and the node count M affect the error bound only in terms of the rescaling factor μ and the variance σ^2.
- Even in the nonlinear case: Enlarging the sensor network indeed helps to improve the SNR!

Is That the End of the Story?

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{m}(y_i - \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle)^2 \quad \text{subject to } \|\boldsymbol{x}\|_1 \leq R$$

$$m \gtrsim \delta^{-2} s \log(\tfrac{2n}{s}) \quad \Rightarrow \quad \|\hat{\boldsymbol{x}} - \mu \boldsymbol{x}_0\|_2 \leq (\sigma^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta$$

- ☺ Low sample complexity
- ☺ Fast & Simple
- ☺ No knowledge of the network configuration required

☹ Fails to work if

$$\mu = \tfrac{1}{M}\sum_{j=1}^{M} \mu_j \approx 0$$

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{m} (y_i - \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle)^2 \quad \text{subject to } \|\boldsymbol{x}\|_1 \leq R$$

$$m \gtrsim \delta^{-2} s \log(\tfrac{2n}{s}) \quad \Rightarrow \quad \|\hat{\boldsymbol{x}} - \mu \boldsymbol{x}_0\|_2 \leq (\sigma^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta$$

☺ Low sample complexity

☺ Fast & Simple

☺ No knowledge of the network configuration required

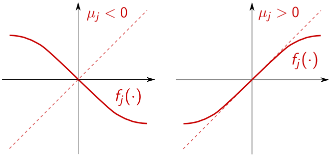
Due to non-coherent transmission, the non-linearities often take the form

$$f_j(v) = h_j \cdot f(v)$$

with $\operatorname{sign}(h_j)$ unknown.

☹ Fails to work if

$$\mu = \tfrac{1}{M} \textstyle\sum_{j=1}^{M} \mu_j \approx 0$$



The Lifting Method

Key Idea: Treating Each Sensor Individually

Model: $y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \ldots, m$

$$\min_{\substack{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^M \\ \in \mathbb{R}^n}} \sum_{i=1}^{m} \Big(y_i - \sum_{j=1}^{M} \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle \Big)^2 \quad \text{subject to } \big\| [\boldsymbol{x}^1 \ldots \boldsymbol{x}^M] \big\|_{1,2} \leq R$$

- Every node j is fitted individually by $\boldsymbol{x}^j \mapsto \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle$.
- The $\ell^{1,2}$-norm

$$\big\| [\boldsymbol{x}^1 \ldots \boldsymbol{x}^M] \big\|_{1,2} := \sum_{k=1}^{n} \Big(\sum_{j=1}^{M} ([\boldsymbol{x}^j]_k)^2 \Big)^{1/2}$$

"couples" all vectors $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M$, but does not enforce equal signs.

Key Idea: Treating Each Sensor Individually

Model: $y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \dots, m$

$$\min_{\substack{\boldsymbol{x}^1,\dots,\boldsymbol{x}^M \\ \in \mathbb{R}^n}} \sum_{i=1}^{m} \Big(y_i - \sum_{j=1}^{M} \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle \Big)^2 \quad \text{subject to } \big\| [\boldsymbol{x}^1 \dots \boldsymbol{x}^M] \big\|_{1,2} \leq R$$

- Every node j is fitted individually by $\boldsymbol{x}^j \mapsto \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle$.
- The $\ell^{1,2}$-norm

$$\big\| [\boldsymbol{x}^1 \dots \boldsymbol{x}^M] \big\|_{1,2} := \sum_{k=1}^{n} \Big(\sum_{j=1}^{M} ([\boldsymbol{x}^j]_k)^2 \Big)^{1/2}$$

"couples" all vectors $\boldsymbol{x}^1, \dots, \boldsymbol{x}^M$, but does not enforce equal signs.

Key Idea: Treating Each Sensor Individually

Model: $y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \ldots, m$

$$\min_{\substack{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^M \\ \in \mathbb{R}^n}} \sum_{i=1}^{m} \Big(y_i - \sum_{j=1}^{M} \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle \Big)^2 \quad \text{subject to } \big\| [\boldsymbol{x}^1 \ldots \boldsymbol{x}^M] \big\|_{1,2} \leq R$$

- Every node j is fitted individually by $\boldsymbol{x}^j \mapsto \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle$.
- The $\ell^{1,2}$-norm

$$\big\| [\boldsymbol{x}^1 \ldots \boldsymbol{x}^M] \big\|_{1,2} := \sum_{k=1}^{n} \Big(\sum_{j=1}^{M} ([\boldsymbol{x}^j]_k)^2 \Big)^{1/2}$$

"couples" all vectors $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M$, but does not enforce equal signs.

Key Idea: Treating Each Sensor Individually

Model: $y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \ldots, m$

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times M}} \sum_{i=1}^{m} \Big(y_i - \langle \boldsymbol{A}_i, \boldsymbol{X} \rangle_{\mathrm{HS}} \Big)^2 \quad \text{subject to } \big\| \boldsymbol{X} \big\|_{1,2} \leq R$$

- Every node j is fitted individually by $\boldsymbol{x}^j \mapsto \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle$.
- The $\ell^{1,2}$-norm

$$\big\| [\boldsymbol{x}^1 \ldots \boldsymbol{x}^M] \big\|_{1,2} := \sum_{k=1}^{n} \Big(\sum_{j=1}^{M} ([\boldsymbol{x}^j]_k)^2 \Big)^{1/2}$$

  "couples" all vectors $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M$, but does not enforce equal signs.
- We are actually working in the matrix space $\mathbb{R}^{n \times M}$:

$$\boldsymbol{A}_i := [\boldsymbol{a}_i^1 \ldots \boldsymbol{a}_i^M], \qquad \boldsymbol{X} := [\boldsymbol{x}^1 \ldots \boldsymbol{x}^M]$$

  $\rightsquigarrow$ Higher computational burden

Recovery via the Lifting Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ *be s-sparse,* $\|\boldsymbol{x}_0\|_2 = 1$*, and* $e_i \sim \mathcal{N}(0, \nu^2)$*. For every* $\delta \in (0, 1]$*, the following holds true with probability at least* $1 - 5\exp(-C\delta^2 m)$*: If*

$$m \gtrsim \delta^{-2} s \max\{M, \log(\tfrac{2n}{s})\},$$

then any minimizer $[\hat{\boldsymbol{x}}^1 \dots \hat{\boldsymbol{x}}^M] \in \mathbb{R}^{n \times M}$ *of the Group-Lasso with* $R = \|[\mu_1 \boldsymbol{x}_0 \dots \mu_M \boldsymbol{x}_0]\|_{1,2}$ *satisfies*

$$\Big(\tfrac{1}{M}\textstyle\sum_{j=1}^{M} \|\hat{\boldsymbol{x}}^j - \mu_j \boldsymbol{x}_0\|_2^2\Big)^{1/2} \leq (\tilde{\sigma}^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

$$\mu_j := \mathbb{E}[f_j(g) \cdot g], \qquad \tilde{\sigma}^2 := \tfrac{1}{M}\textstyle\sum_{j=1}^{M} \|f_j(g) - \mu_j g\|_{\psi_2}^2$$

$$\min_{\substack{\boldsymbol{x}^1, \dots, \boldsymbol{x}^M \\ \in \mathbb{R}^n}} \sum_{i=1}^{m} \Big(y_i - \sum_{j=1}^{M} \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j \rangle\Big)^2 \quad \text{subject to } \big\|[\boldsymbol{x}^1 \dots \boldsymbol{x}^M]\big\|_{1,2} \leq R$$

Recovery via the Lifting Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ *be s-sparse,* $\|\boldsymbol{x}_0\|_2 = 1$*, and* $e_i \sim \mathcal{N}(0, \nu^2)$*. For every* $\delta \in (0,1]$*, the following holds true with probability at least* $1 - 5\exp(-C\delta^2 m)$*: If*

$$m \gtrsim \delta^{-2} s \max\{M, \log(\tfrac{2n}{s})\},$$

then any minimizer $[\hat{\boldsymbol{x}}^1 \dots \hat{\boldsymbol{x}}^M] \in \mathbb{R}^{n \times M}$ *of the Group-Lasso with* $R = \|[\mu_1 \boldsymbol{x}_0 \dots \mu_M \boldsymbol{x}_0]\|_{1,2}$ *satisfies*

$$\Big(\tfrac{1}{M}\sum_{j=1}^M \|\hat{\boldsymbol{x}}^j - \mu_j \boldsymbol{x}_0\|_2^2\Big)^{1/2} \le (\tilde{\sigma}^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

$$\mu_j := \mathbb{E}[f_j(g) \cdot g], \qquad \tilde{\sigma}^2 := \tfrac{1}{M}\sum_{j=1}^M \|f_j(g) - \mu_j g\|_{\psi_2}^2$$

$$\min_{\substack{\boldsymbol{x}^1,\dots,\boldsymbol{x}^M \\ \in \mathbb{R}^n}} \sum_{i=1}^m \Big(y_i - \sum_{j=1}^M \langle \boldsymbol{a}_i^j, \boldsymbol{x}^j\rangle\Big)^2 \quad \text{subject to } \big\|[\boldsymbol{x}^1 \dots \boldsymbol{x}^M]\big\|_{1,2} \le R$$

Recovery via the Lifting Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ be s-sparse, $\|\boldsymbol{x}_0\|_2 = 1$, and $e_i \sim \mathcal{N}(0, \nu^2)$. For every $\delta \in (0, 1]$, the following holds true with probability at least $1 - 5\exp(-C\delta^2 m)$: If

$$m \gtrsim \delta^{-2} s \max\{M, \log(\tfrac{2n}{s})\},$$

then any minimizer $[\hat{\boldsymbol{x}}^1 \dots \hat{\boldsymbol{x}}^M] \in \mathbb{R}^{n \times M}$ of the Group-Lasso with $R = \|[\mu_1 \boldsymbol{x}_0 \dots \mu_M \boldsymbol{x}_0]\|_{1,2}$ satisfies

$$\left(\tfrac{1}{M} \textstyle\sum_{j=1}^{M} \|\hat{\boldsymbol{x}}^j - \mu_j \boldsymbol{x}_0\|_2^2\right)^{1/2} \leq (\tilde{\sigma}^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

Recovery becomes feasible if $m \gtrsim s \max\{M, \log(\frac{2n}{s})\}$!

- Every $\hat{\boldsymbol{x}}^j$ approximates rescaled $\boldsymbol{x}_0$ even if $\mu = \frac{1}{M}\sum_{j=1}^{M} \mu_j \approx 0$, ...

Recovery via the Lifting Method

Theorem (Genzel & J., 2017)

Let $\boldsymbol{x}_0$ be s-sparse, $\|\boldsymbol{x}_0\|_2 = 1$, and $e_i \sim \mathcal{N}(0, \nu^2)$. For every $\delta \in (0, 1]$, the following holds true with probability at least $1 - 5\exp(-C\delta^2 m)$: If

$$m \gtrsim \delta^{-2} s \max\{M, \log(\tfrac{2n}{s})\},$$

then any minimizer $[\hat{\boldsymbol{x}}^1 \ldots \hat{\boldsymbol{x}}^M] \in \mathbb{R}^{n \times M}$ of the Group-Lasso with $R = \|[\mu_1 \boldsymbol{x}_0 \ldots \mu_M \boldsymbol{x}_0]\|_{1,2}$ satisfies

$$\left(\tfrac{1}{M} \sum_{j=1}^{M} \|\hat{\boldsymbol{x}}^j - \mu_j \boldsymbol{x}_0\|_2^2\right)^{1/2} \leq (\tilde{\sigma}^2 + \tfrac{\nu^2}{M})^{1/2} \cdot \delta \ .$$

Recovery becomes feasible if $m \gtrsim s \max\{M, \log(\frac{2n}{s})\}$!

- Every $\hat{\boldsymbol{x}}^j$ approximates rescaled $\boldsymbol{x}_0$ even if $\mu = \frac{1}{M} \sum_{j=1}^{M} \mu_j \approx 0$, ...
- ... but the sample complexity now grows linearly in M.
- rule for sensor count: $M \sim log(2n/s)$.

Direct vs. Lifting

Direct Method

- ☺ Low sample complexity
- ☺ Fast & Simple
- ☺ No knowledge of the network configuration required
- ☹ Cannot deal with $\mu \approx 0$

Lifting Method

- ☺ Can deal with $\mu \approx 0$ ($\rightsquigarrow$ non-coherent transmission)
- ☺ Allows to learn about the network configuration ("bilinear" problem)
- ☹ Higher sample complexity for large networks $O(s \cdot M)$
- ☹ Slower

Direct vs. Lifting

Direct Method

- ☺ Low sample complexity
- ☺ Fast & Simple
- ☺ No knowledge of the network configuration required

- ☹ Cannot deal with $\mu \approx 0$

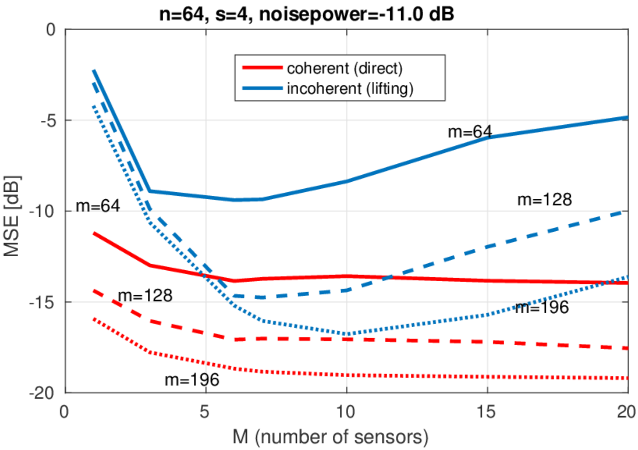
Lifting Method

- ☺ Can deal with $\mu \approx 0$ ($\leadsto$ non-coherent transmission)
- ☺ Allows to learn about the network configuration ("bilinear" problem)

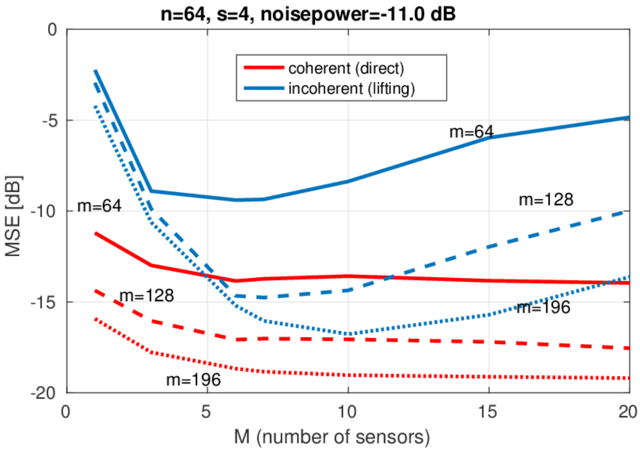
- ☹ Higher sample complexity for large networks $O(s \cdot M)$
- ☹ Slower

$\Rightarrow$ *The direct approach should be preferred unless $\mu \approx 0$.*

Direct vs. Lifting



Direct vs. Lifting



Can we have the best of both worlds?

Some Available Extensions

- *The Hybrid Method*

  Allows to incorporate prior knowledge of the sensor configurations

$$\boldsymbol{b}_i^k := \sum_{j=1}^{M} c_{j,k} \boldsymbol{a}_i^j, \quad k = 1, \ldots, M'$$

- *Sub-Gaussian Measurements*
- *Imperfect Tuning*

  Recovery if the Lasso parameter R was not perfectly chosen
  $\rightsquigarrow$ The error rate may drop down to $O(m^{-1/4})$

- *Stability and Robustness*

  Different types of noise and (compressible) source vectors

The proofs are based on a more general framework for convex recovery from non-linear observations.

Let's conclude...

What to Take Home...?

- Some problems in wireless sensor networks can be modeled by superimposed non-linearly distorted measurements

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \ldots, m.$$

- Recovery is often feasible with a very few measurements ...
- ... without knowing the exact model configuration $(f_1, \ldots, f_M)$.

What to Take Home...?

- Some problems in wireless sensor networks can be modeled by superimposed non-linearly distorted measurements

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \quad i = 1, \dots, m.$$

- Recovery is often feasible with a very few measurements ...
- ... without knowing the exact model configuration $(f_1, \dots, f_M)$.

But there is still a lot of work to do!

- *Structured measurements?*

  Sub-Gaussians do not always reflect real-world applications.

- *Non-convex recovery?*

  Breaking the complexity barrier: $O(s \cdot M) \rightsquigarrow O(s + M)$

  related to bilinear inverse problems with sparsity priors!

THANK YOU FOR YOUR ATTENTION!

References

M. Genzel and P. Jung
Recovering Structured Data From Superimposed Non-Linear Measurements.
In preparation, 2017.

M. Genzel and P. Jung
Blind Sparse Recovery From Superimposed Non-Linear Sensor Measurements.
Proceedings of SampTA 2017.

M. Genzel
A Unified Framework for Convex Recovery From Non-Linear Observations.
In preparation, 2017.

M. Genzel
High-Dimensional Estimation of Structured Signals From Non-Linear Observations With General Convex Loss Functions.
IEEE Trans. Inf. Theory 63.3 (2017), 1601–1619.

Low- or High-Quality Sensors?