EXTENDED CYCLIC COORDINATE DESCENT FOR ROBUST ROW-SPARSE SIGNAL RECONSTRUCTION IN THE PRESENCE OF OUTLIERS

Huiping Huang[⋆‡], Hing Cheung So[†], Abdelhak M. Zoubir[⋆‡]

[⋆] *Signal Processing Group, Technische Universität Darmstadt*
[‡] *Graduate School of Computational Engineering, Technische Universität Darmstadt*
[†] *Department of Electrical Engineering, City University of Hong Kong*

ICASSP 2020

Content

Overview

TECHNISCHE UNIVERSITÄT DARMSTADT

Introduction
Sparse Signal Reconstruction (1)

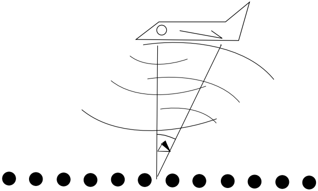
- A sparse signal can be reliably recovered from a reduced set of measurements, under certain conditions, e.g., RIP.

Introduction
Sparse Signal Reconstruction (1)

- A sparse signal can be reliably recovered from a reduced set of measurements, under certain conditions, e.g., RIP.
- A sparse signal can carry useful information, such as directions of sources.

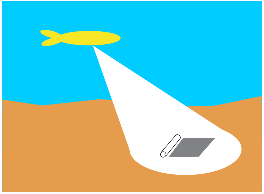
Introduction
Sparse Signal Reconstruction (1)

- A sparse signal can be reliably recovered from a reduced set of measurements, under certain conditions, e.g., RIP.
- A sparse signal can carry useful information, such as directions of sources.
- Applications



(a) Target tracking

(b) Through-the-wall radar 3D reconstruction



(c) Sonar imaging

Introduction
Sparse Signal Reconstruction (2)

Data model:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{e}(t), \quad t = 1, 2, \ldots, T$$

Introduction
Sparse Signal Reconstruction (2)

Data model:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{e}(t), \quad t = 1, 2, \ldots, T$$

- $\mathbf{y}(t) \in \mathbb{C}^N$: Measurement data
- $\mathbf{A} \in \mathbb{C}^{N \times P}$: Predictor dictionary, $N < P$
- $\mathbf{s}(t) \in \mathbb{C}^P$: Signal vector
- $\mathbf{e}(t) \in \mathbb{C}^N$: Gaussian noise

Introduction
Sparse Signal Reconstruction (3)

Data model in matrix form:

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{E}$$

where $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \cdots, \mathbf{y}(T)]$, $\mathbf{S} = [\mathbf{s}(1), \mathbf{s}(2), \cdots, \mathbf{s}(T)]$, $\mathbf{E} = [\mathbf{e}(1), \mathbf{e}(2), \cdots, \mathbf{e}(T)]$

Data model in matrix form:

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{E}$$

where $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \cdots, \mathbf{y}(T)]$, $\mathbf{S} = [\mathbf{s}(1), \mathbf{s}(2), \cdots, \mathbf{s}(T)]$,
$\mathbf{E} = [\mathbf{e}(1), \mathbf{e}(2), \cdots, \mathbf{e}(T)]$

Sparse signal reconstruction using LASSO:

$$\min_{\mathbf{S}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{A}\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_{2,1}$$

Introduction
Non-Gaussian Heavy-Tailed Noise (1)

- Data model without outliers

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{e}(t)$$

$\mathbf{e}(t) \in \mathbb{C}^N$: Gaussian noise

Introduction

Non-Gaussian Heavy-Tailed Noise (1)

- Data model without outliers

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{e}(t)$$

$\mathbf{e}(t) \in \mathbb{C}^N$: Gaussian noise

- Data model with outliers

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{q}(t) \circ \mathbf{e}(t)$$

where the symbol $\circ$ stands for the Hadamard product operator,

Introduction
Non-Gaussian Heavy-Tailed Noise (1)

- Data model without outliers

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{e}(t)$$

$\mathbf{e}(t) \in \mathbb{C}^N$: Gaussian noise

- Data model with outliers

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{q}(t) \circ \mathbf{e}(t)$$

where the symbol $\circ$ stands for the Hadamard product operator,
$\mathbf{q}(t) \in \mathbb{R}^N$: Weight vector

$$\begin{cases} q_i(t) \gg 1 & : \text{Outlier} \\ q_i(t) = 1 & : \text{Non-outlier} \end{cases}$$

where $q_i(t)$ is the i-th entry of $\mathbf{q}(t)$

Introduction
Non-Gaussian Heavy-Tailed Noise (2)

- Data model with outliers in matrix form:

$$\mathbf{Y} = \mathbf{AS} + \mathbf{Q} \circ \mathbf{E}$$

where $\mathbf{Q} = [\mathbf{q}(1), \mathbf{q}(2), \cdots, \mathbf{q}(T)]$

Introduction
Non-Gaussian Heavy-Tailed Noise (2)

- Data model with outliers in matrix form:

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{Q} \circ \mathbf{E}$$

  where $\mathbf{Q} = [\mathbf{q}(1), \mathbf{q}(2), \cdots, \mathbf{q}(T)]$

- The problem to be addressed is as follows:

  With known **Y** and **A**, recover the row-sparse matrix **S.**

Overview

Proposed Approach
Proposed Minimization Problem (1)

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

Proposed Approach
Proposed Minimization Problem (1)

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1\|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2\|\mathbf{S}\|_{2,1}$$

- The first term: **W** is used to down-weight the the outliers while keeping the remaining observations unchanged

$$0 < W_{ij} < 1 \text{ for outliers, } W_{ij} = 1 \text{ for non-outliers}$$

Proposed Approach
Proposed Minimization Problem (1)

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

- The first term: **W** is used to down-weight the the outliers while keeping the remaining observations unchanged

$$0 < W_{ij} < 1 \text{ for outliers, } W_{ij} = 1 \text{ for non-outliers}$$

- The second term: Impose sparsity of $\mathbf{W}_{-1} = \mathbf{W} - \mathbf{1}_{N \times T}$

Proposed Approach
Proposed Minimization Problem (1)

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

- The first term: **W** is used to down-weight the the outliers while keeping the remaining observations unchanged

$$0 < W_{ij} < 1 \text{ for outliers, } W_{ij} = 1 \text{ for non-outliers}$$

- The second term: Impose sparsity of $\mathbf{W}_{-1} = \mathbf{W} - \mathbf{1}_{N \times T}$
- The third term: Impose row-sparsity of **S**

Proposed Approach
Proposed Minimization Problem (2)

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\boxed{\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}}$$

- LASSO-type problem, cyclic coordinate descent (CCD)

Proposed Approach
Proposed Minimization Problem (2)

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\boxed{\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}}$$

- LASSO-type problem, cyclic coordinate descent (CCD)
- Biconvex with respect to **W** and **S**

Proposed Approach
Remarks (1)

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{\mathbf{W},\mathbf{S}} \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

- Comparison with the one using Huber's loss function

$$\min_{\mathbf{S}} \|\rho_{\mathrm{H}}(\mathbf{Y} - \mathbf{AS})\|_F^2 + \mu\|\mathbf{S}\|_{2,1}$$

where

$$\rho_{\mathrm{H}}(x) = \begin{cases} |x|^2, & \text{for } |x| \leq c \ \text{(i.e., non-outliers)} \\ 2c|x| - c^2, & \text{for } |x| > c \ \text{(i.e., outliers)} \end{cases}$$

Proposed Approach
Remarks (2)

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

- Comparison with the original problem without outliers

$$\min_{\mathbf{S}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{AS}\|_F^2 + \lambda \|\mathbf{S}\|_{2,1}$$

Proposed Approach
Extended Cyclic Coordinate Descent Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{A}\mathbf{S})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

Proposed Approach
Extended Cyclic Coordinate Descent Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\boxed{\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\text{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}}$$

- Step 1: Update W_{ij} by fixing $\mathbf{S}$ and W_{nt} ($n = 1, \ldots, N$ and $n \neq i$, $t = 1, \ldots, T$ and $t \neq j$)

Proposed Approach
Extended Cyclic Coordinate Descent Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{A}\mathbf{S})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

- Step 1: Update W_{ij} by fixing $\mathbf{S}$ and W_{nt} ($n = 1, \ldots, N$ and $n \neq i$, $t = 1, \ldots, T$ and $t \neq j$)
- Step 2: Update S_{ij} by fixing $\mathbf{W}$ and S_{pt} ($p = 1, \ldots, P$ and $p \neq i$, $t = 1, \ldots, T$ and $t \neq j$)

Proposed Approach
Extended Cyclic Coordinate Descent Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

- Step 1: Update W_{ij} by fixing $\mathbf{S}$ and W_{nt} ($n = 1, \ldots, N$ and $n \neq i$, $t = 1, \ldots, T$ and $t \neq j$)
- Step 2: Update S_{ij} by fixing $\mathbf{W}$ and S_{pt} ($p = 1, \ldots, P$ and $p \neq i$, $t = 1, \ldots, T$ and $t \neq j$)
- Step 3: Cyclically repeating Steps 1 and 2 until convergence

Proposed Approach
Extended Cyclic Coordinate Descent Algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

$$\min_{\mathbf{W},\mathbf{S}} \quad \frac{1}{2}\|\mathbf{W} \circ (\mathbf{Y} - \mathbf{AS})\|_F^2 + \lambda_1 \|\mathrm{vec}(\mathbf{W}_{-1})\|_1 + \lambda_2 \|\mathbf{S}\|_{2,1}$$

- Step 1: Update W_{ij} by fixing **S** and W_{nt} ($n = 1, \ldots, N$ and $n \neq i$, $t = 1, \ldots, T$ and $t \neq j$)
- Step 2: Update S_{ij} by fixing **W** and S_{pt} ($p = 1, \ldots, P$ and $p \neq i$, $t = 1, \ldots, T$ and $t \neq j$)
- Step 3: Cyclically repeating Steps 1 and 2 until convergence

Detailed derivations of the whole algorithm can be found in our paper.

Overview

Application to Source Localization
Relative Frequency (RF) Comparison

Application to Source Localization
Relative Frequency (RF) Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

Setups:

- ULA of $N = 8$ sensors
- $K = 3$ signals with directions: $\{-20^\circ, 0^\circ, 15^\circ\}$
- Over-complete dictionary $\mathbf{A}$ contains $P = 101$ columns, covering from -50° to 50°
- $T = 64$ snapshots
- $\text{SNR} = 0$ dB

Application to Source Localization
Relative Frequency (RF) Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

Setups:

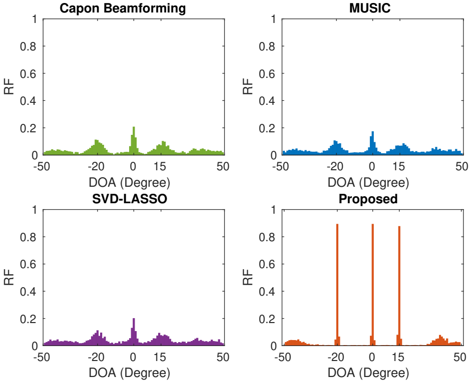
- ULA of $N = 8$ sensors
- $K = 3$ signals with directions: $\{-20°, 0°, 15°\}$
- Over-complete dictionary $\mathbf{A}$ contains $P = 101$ columns, covering from $-50°$ to $50°$
- $T = 64$ snapshots
- $\text{SNR} = 0$ dB
- $\text{RF}(\theta) = \frac{\hat{\kappa}(\theta)}{\text{MC}}$

Application to Source Localization
Relative Frequency (RF) Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

Setups:

- ULA of $N = 8$ sensors
- $K = 3$ signals with directions: $\{-20^\circ, 0^\circ, 15^\circ\}$
- Over-complete dictionary $\mathbf{A}$ contains $P = 101$ columns, covering from -50° to 50°
- $T = 64$ snapshots
- $\text{SNR} = 0$ dB
- $\text{RF}(\theta) = \frac{\hat{\kappa}(\theta)}{\text{MC}}$



Application to Source Localization
Probability of Exact Recovery (PER) Comparison

Application to Source Localization
Probability of Exact Recovery (PER) Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

Setups:

- ULA of $N = 5$ sensors
- $K = 2$ signals with directions: $\{-15^\circ, 15^\circ\}$
- Over-complete dictionary $\mathbf{A}$ contains $P = 101$ columns, covering from -50° to 50°
- $T = 64$ snapshots
- SNR varies from -10 dB to 10 dB

Application to Source Localization
Probability of Exact Recovery (PER) Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

Setups:

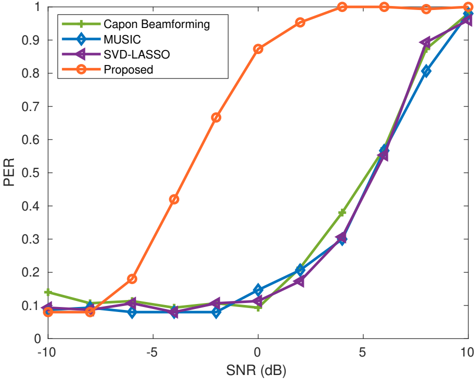
- ULA of $N = 5$ sensors
- $K = 2$ signals with directions: $\{-15^\circ, 15^\circ\}$
- Over-complete dictionary $\mathbf{A}$ contains $P = 101$ columns, covering from -50° to 50°
- $T = 64$ snapshots
- SNR varies from -10 dB to 10 dB
- $\text{PER} = \frac{1}{K} \sum_{k=1}^{K} \text{RF}(\theta_k)$

Application to Source Localization
Probability of Exact Recovery (PER) Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

Setups:

- ULA of $N = 5$ sensors
- $K = 2$ signals with directions: $\{-15^\circ, 15^\circ\}$
- Over-complete dictionary $\mathbf{A}$ contains $P = 101$ columns, covering from -50° to 50°
- $T = 64$ snapshots
- SNR varies from -10 dB to 10 dB
- $\text{PER} = \frac{1}{K}\sum_{k=1}^{K} \text{RF}(\theta_k)$



Overview

TECHNISCHE UNIVERSITÄT DARMSTADT

Summary and Outlook

Summary:

- This paper studied the problem of row-sparse signal recovery for complex-valued data contaminated with outliers.
- Using a weight matrix, we formulated the problem in the form of LASSO-type.
- We extended CCD algorithm to solve the resulting problem.
- Simulation results showed that the proposed algorithm is robust against outliers even in the low SNR regime.

[1] C. Steffens, M. Pesavento, and M.E. Pfetsch, “A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem,” *IEEE TSP* 2018.

Summary and Outlook

Summary:

- This paper studied the problem of row-sparse signal recovery for complex-valued data contaminated with outliers.
- Using a weight matrix, we formulated the problem in the form of LASSO-type.
- We extended CCD algorithm to solve the resulting problem.
- Simulation results showed that the proposed algorithm is robust against outliers even in the low SNR regime.

Outlook:

- The sparsity structure of the weight matrix will be further studied.
- Some state-of-the-art techniques, such as SPARROW[1], will be studied in the framework of row-sparse signal reconstruction.

[1] C. Steffens, M. Pesavento, and M.E. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," *IEEE TSP* 2018.

TECHNISCHE UNIVERSITÄT DARMSTADT

Thank you for your attention!

SIGNAL PROCESSING GROUP

Department of Electrical Engineering

香港城市大學
City University of Hong Kong

Acknowledgment:
This work of Huiping Huang is supported by the ‘Excellence Initiative’ of the German Federal and State Governments and the Graduate School of Computational Engineering at Technische Universität Darmstadt.