UNIVERSIDADE FEDERAL DO CEARÁ

# Alternating group lasso for block-term tensor decomposition with application to ECG source separation

J. H. de M. Goulart et al., 2020

# I. Introduction

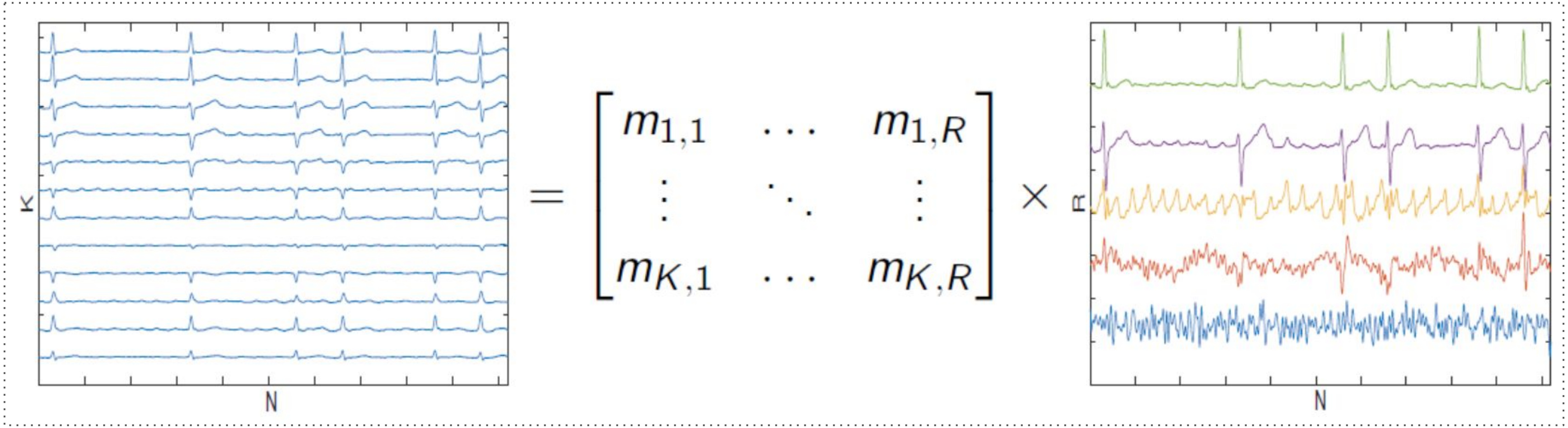
# I. Introduction

- Biomedical Application: Atrial activity **extraction** during a persistent Atrial Fibrillation (**AF**).

- Blind Source Separation **(BSS)**
    - Matrix Approach
        - Principal component analysis (PCA)
        - Independent component analysis (ICA)
    - Tensor Approach
        - **BTD**-Gauss Newton (ALS-NLS or ELS)
        - **AGL** and its Constrained Version, **CAGL**

# Matrix Approach

$$\mathbf{Y} = \mathbf{M} \times \mathbf{S}$$




$$= \begin{bmatrix} m_{1,1} & \cdots & m_{1,R} \\ \vdots & \ddots & \vdots \\ m_{K,1} & \cdots & m_{K,R} \end{bmatrix} \times$$

Figure : **Classic model for BSS.**

# Tensor Approach

$$\mathcal{Y} \approx \sum_{r=1}^{R} \mathbf{A}_r \mathbf{B}_r^\top \circ \mathbf{x}_r$$

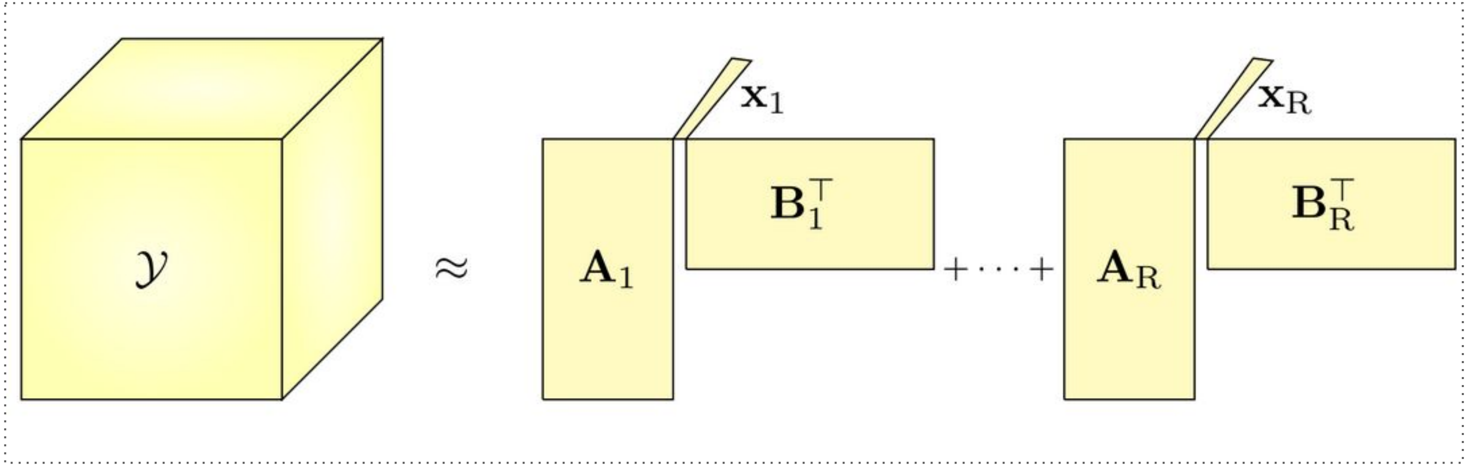



Figure : **BTD** model for **BSS**.

# II. Tensor-Based Separation of Sums of Complex Exponentials

# A. Low-rank Hankel source model

**Discrete-time signals that can be modeled as linear combination of exponentials**

- The signal $s(n)$ is represented by an all-pole model
- Mapped onto a Hankel matrix $Hs$
- Hankel matrix has rank at most $\min\{L,M\}$ [1]

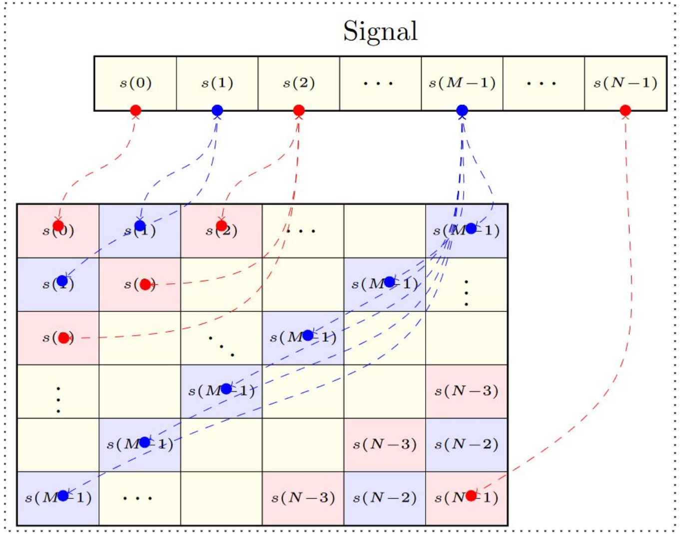
$$s(n) = \sum_{l=1}^{L} \alpha \exp(\zeta_l n), \quad n = 0, \ldots, N-1,$$

[1]D. L. Boley, et. al (1997): “Vandermonde factorization of a Hankel matrix,” in Proc. Workshop Scientific Comput.

# A. Low-rank Hankel source model

**Hankel Matrix and Vandermonde Decomposition:**

- Since the sources can be expressed as low-rank Hankel matrices, the signal separation can be performed via BTD.




$$\mathbf{H}_s = \mathbf{V}_s \, \mathrm{Diag}(\alpha_1, \ldots, \alpha_L) \mathbf{V}_s^{\mathsf{T}}$$

$$\mathbf{V}_s \triangleq \begin{bmatrix} 1 & \ldots & 1 \\ \exp(\zeta_1) & \ldots & \exp(\zeta_L) \\ \vdots & & \vdots \\ \exp(\zeta_1(M-1)) & \ldots & \exp(\zeta_L(M-1)) \end{bmatrix} \in \mathbb{C}^{M \times L}$$

Figure : Hankel mapping.

# B. Separation of linear mixture via BTD

**Spatial diversity plays an essential role on source estimation**

- Each observed signal ($\mathbf{Y}$) is mapped onto a structured Hankel Matrix
- Where we stack each these matrices in the 3rd-mode of the tensor data *(M x M x K)*

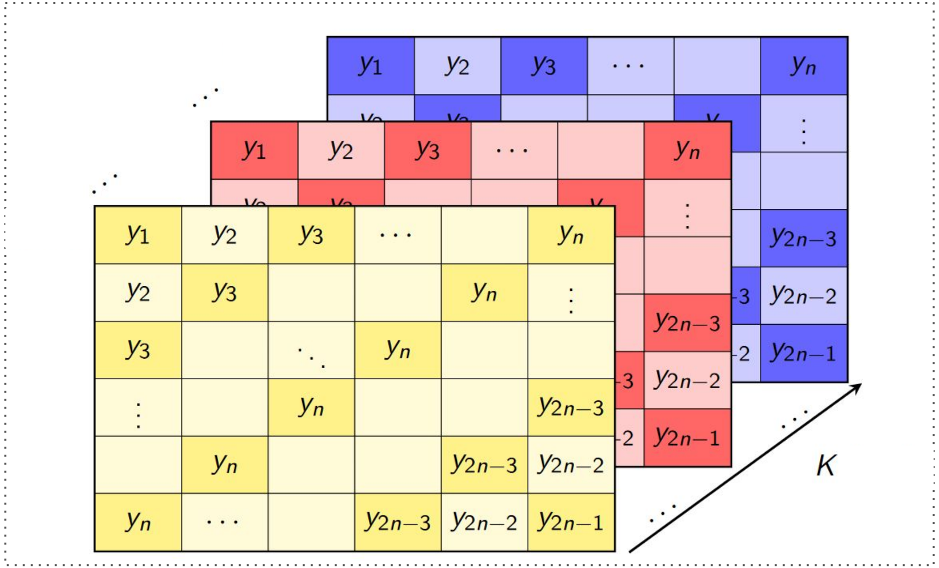



Figure : Tensor built of data Hankel matrices.

# B. Separation of linear mixture via BTD

**Source estimation and its associated Hankel matrix**

- Each Hankel Matrix has rank $Lr$
- The data tensor $\boldsymbol{y}$ consists of a sum of $R$ blocks
- Each block is given by the tensor (outer) product of a low-rank Hankel matrix and a vector

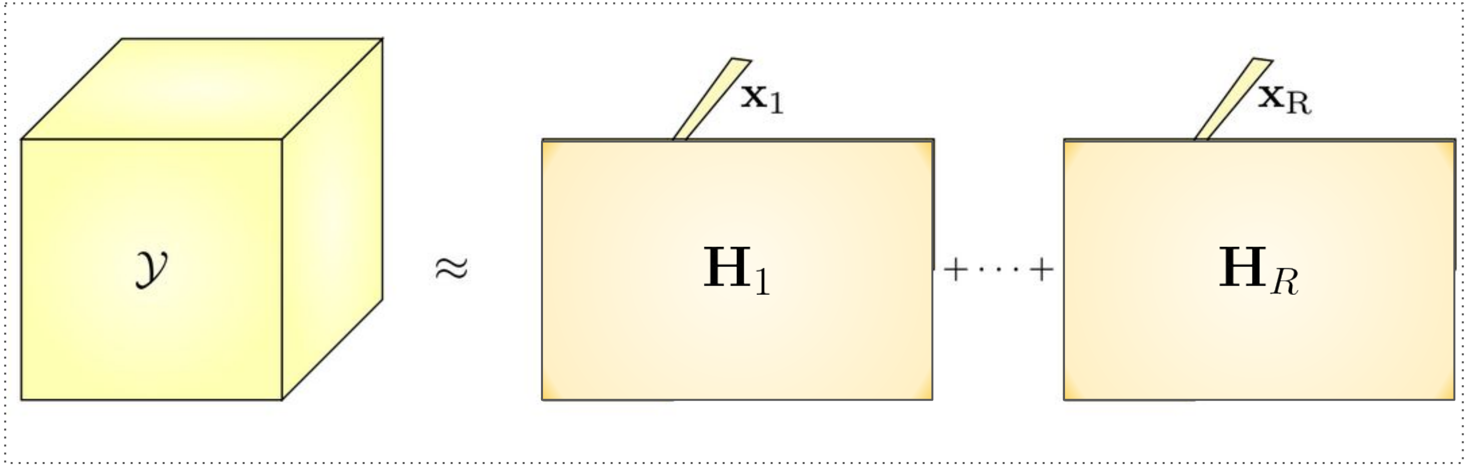



Figure : Structured-**BTD** model for **BSS**.

# C. Approximate BTD

**Existence and Uniqueness**

- Fixed structure
- Prior Knowledge of *(R, Lr)*
- Complex values and require additional constraints (as a structured matrix)

$$f(\mathbf{A}, \mathbf{B}, \mathbf{X}) \triangleq \left\| \mathcal{Y} - \sum_{r=1}^{R} \left( \mathbf{A}_r \mathbf{B}_r^{\mathsf{T}} \right) \otimes \mathbf{x}_r \right\|_F^2$$

$$\min_{(\mathbf{A},\mathbf{B},\mathbf{X}) \in \mathcal{S}} f(\mathbf{A}, \mathbf{B}, \mathbf{X})$$

# D. Shortcomings of the standard least-squares approach

**Blocks and rank remarks**

- Flexibility regarding the ranks $Lr$ of the blocks (which can be very different from one another)
- Rank Inversion and overestimation
- Set of possible ranks is restricted (it becomes more severe as the number of blocks grows)

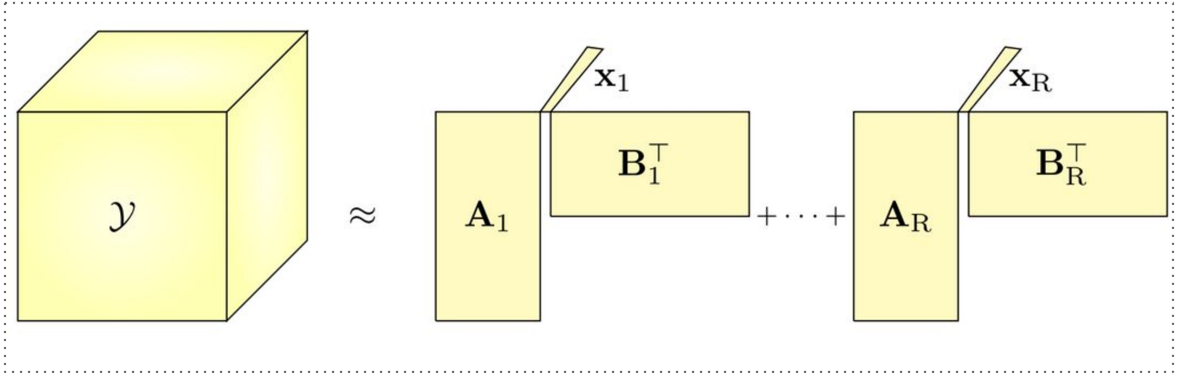



Figure : **BTD** model for **BSS**.

# III. Alternating Group Lasso Algorithm for BTD

# A. Problem formulation

**Constrained Alternating Group Lasso (CAGL) Approach**

- Non-fixed structure minimizing $\mathrm{F}(\mathbf{A}, \mathbf{B}, \mathbf{X})$ ensuring the Hankel structure
- Penalization term $(\gamma)$ and $\mathrm{g}(\mathbf{A}, \mathbf{B}, \mathbf{X})$ limiting the multilinear ranks and number of blocks
- Allows simultaneous estimation of $(R, Lr)$ and model factors

$$F(\mathbf{A}, \mathbf{B}, \mathbf{X}) \triangleq f(\mathbf{A}, \mathbf{B}, \mathbf{X}) + \gamma\, g(\mathbf{A}, \mathbf{B}, \mathbf{X})$$

$$g(\mathbf{A}, \mathbf{B}, \mathbf{X}) \triangleq \|\mathbf{A}\|_{2,1} + \|\mathbf{B}\|_{2,1} + \|\mathbf{X}\|_{2,1}$$

$$\min_{(\mathbf{A},\mathbf{B},\mathbf{X}) \in \mathcal{S}} \quad F(\mathbf{A}, \mathbf{B}, \mathbf{X})$$

# B. Algorithm for unconstrained blocks

## AGL Algorithm

- Solve Group Lasso Subproblems, where the factor $\mathbf{A}$ is estimated depending on $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$

**Inputs:** Data tensor $\mathcal{Y}$, penalty parameter $\gamma$, proximal term weight $\tau$, initial point $(\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \mathbf{X}^{(0)})$
**Outputs:** Approximate BTD factors $(\mathbf{A}, \mathbf{B}, \mathbf{X})$

1: $t \leftarrow 1$
2: **while** stopping criteria not met **do**
3: Solve group lasso subproblem (12) to obtain $\mathbf{A}^{(t)}$ from $\mathbf{A}^{(t-1)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$
4: Solve group lasso subproblem in $\mathbf{B}$ analogous to (12) to obtain $\mathbf{B}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$
5: **for** $r = 1, \ldots, R$ **do**
6: $L_r^{(t)} \leftarrow \operatorname{rank}\left(\mathbf{A}_r^{(t)} (\mathbf{B}^{(t)})_r^{\mathsf{T}}\right)$
7: $(\mathbf{A}_r^{(t)}, \mathbf{B}_r^{(t)}) \leftarrow \texttt{slra}(\mathbf{A}_r^{(t)} (\mathbf{B}_r^{(t)})^{\mathsf{T}}, L_r^{(t)})$
8: $(\mathbf{A}_r^{(t)}, \mathbf{B}_r^{(t)}) \leftarrow ([\mathbf{A}_r^{(t)}\ \mathbf{0}_{I \times L - L_r^{(t)}}], [\mathbf{B}_r^{(t)}\ \mathbf{0}_{I \times L - L_r^{(t)}}])$
9: Solve group lasso subproblem in $\mathbf{X}$ analogous to (12) to obtain $\mathbf{X}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t)}$ and $\mathbf{X}^{(t-1)}$
10: $t \leftarrow t + 1$

Table I: Pseudocode for the unconstrained AGL algorithm (lines 5–8 must be omitted).

# C. Handling linear constraints in $\mathbf{H}_r$

**CAGL Algorithm**

- Structured low-rank approximation (SRLA)

| | |
|---|---|
| **Inputs:** | Data tensor $\mathcal{Y}$, penalty parameter $\gamma$, proximal term weight $\tau$, initial point $(\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \mathbf{X}^{(0)})$ |
| **Outputs:** | Approximate BTD factors $(\mathbf{A}, \mathbf{B}, \mathbf{X})$ |

$t \leftarrow 1$
**while** stopping criteria not met **do**
  Solve group lasso subproblem (12) to obtain $\mathbf{A}^{(t)}$ from $\mathbf{A}^{(t-1)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$
  Solve group lasso subproblem in $\mathbf{B}$ analogous to (12) to obtain $\mathbf{B}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$
  **for** $r = 1, \ldots, R$ **do**
    $L_r^{(t)} \leftarrow \operatorname{rank}\left(\mathbf{A}_r^{(t)}(\mathbf{B}^{(t)})_r^{\mathsf{T}}\right)$
    $(\mathbf{A}_r^{(t)}, \mathbf{B}_r^{(t)}) \leftarrow \mathtt{slra}(\mathbf{A}_r^{(t)}(\mathbf{B}_r^{(t)})^{\mathsf{T}}, L_r^{(t)})$
    $(\mathbf{A}_r^{(t)}, \mathbf{B}_r^{(t)}) \leftarrow ([\mathbf{A}_r^{(t)}\ \mathbf{0}_{I\times L-L_r^{(t)}}], [\mathbf{B}_r^{(t)}\ \mathbf{0}_{I\times L-L_r^{(t)}}])$
  Solve group lasso subproblem in $\mathbf{X}$ analogous to (12) to obtain $\mathbf{X}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t)}$ and $\mathbf{X}^{(t-1)}$
  $t \leftarrow t + 1$

$$\hat{\mathbf{H}}_r \approx \hat{\mathbf{A}}_r \hat{\mathbf{B}}_r^{\mathsf{T}} \in \mathcal{U}$$

Table I: Pseudocode for constrained AGL algorithm.

# IV. Numerical Evaluation on Random Block Term Decomposition Models

# A. Unconstrained BTD

**Experiment Setup**

- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{X}$ and $\mathbf{N}$ in an i.i.d. fashion from the standard normal distribution.
- Algebraic method with rank estimates $L_1 = L_2 = L_3 = L$
- AGL is run from this initial solution and using $\gamma = \gamma 0$
- Apply a γ-sweeping procedure inspired by solution-path techniques
- Overestimation and true parameters
- SNR of 20 dB

$$\mathcal{Y}_0 = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{x}_r$$

$$\mathcal{Y} = \mathcal{Y}_0 + \sigma_{\mathcal{N}} \mathcal{N}$$

$$\mathrm{NMSE}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{X}}) \triangleq \frac{1}{R} \sum_{r=1}^{R} \frac{\|(\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{x}_r - (\hat{\mathbf{A}}_r \hat{\mathbf{B}}_r^\top) \otimes \hat{\mathbf{x}}_r\|_F^2}{\|(\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{x}_r\|_F^2}$$

# A. Unconstrained BTD

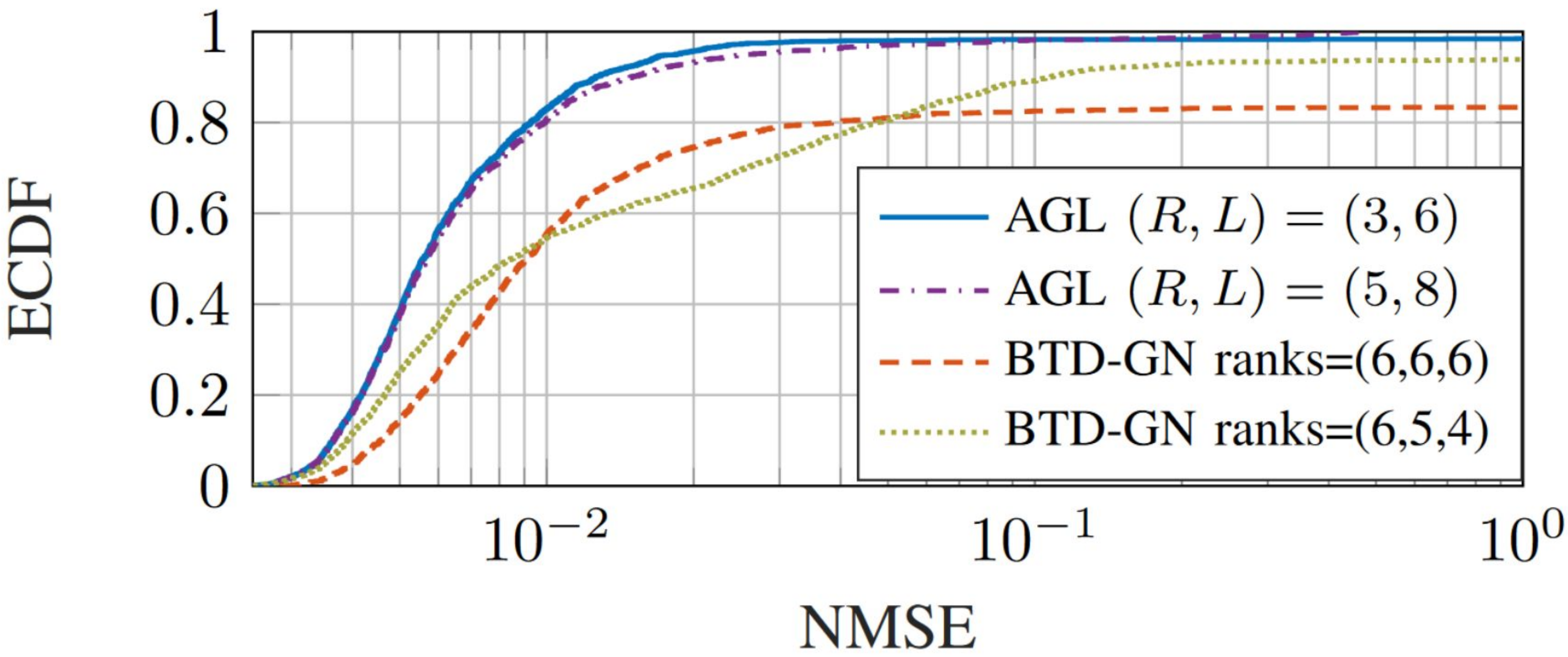



Figure 1: Empirical CDFs of NMSE over estimated blocks by AGL and BTD-GN for 500 realizations.

# A. Unconstrained BTD

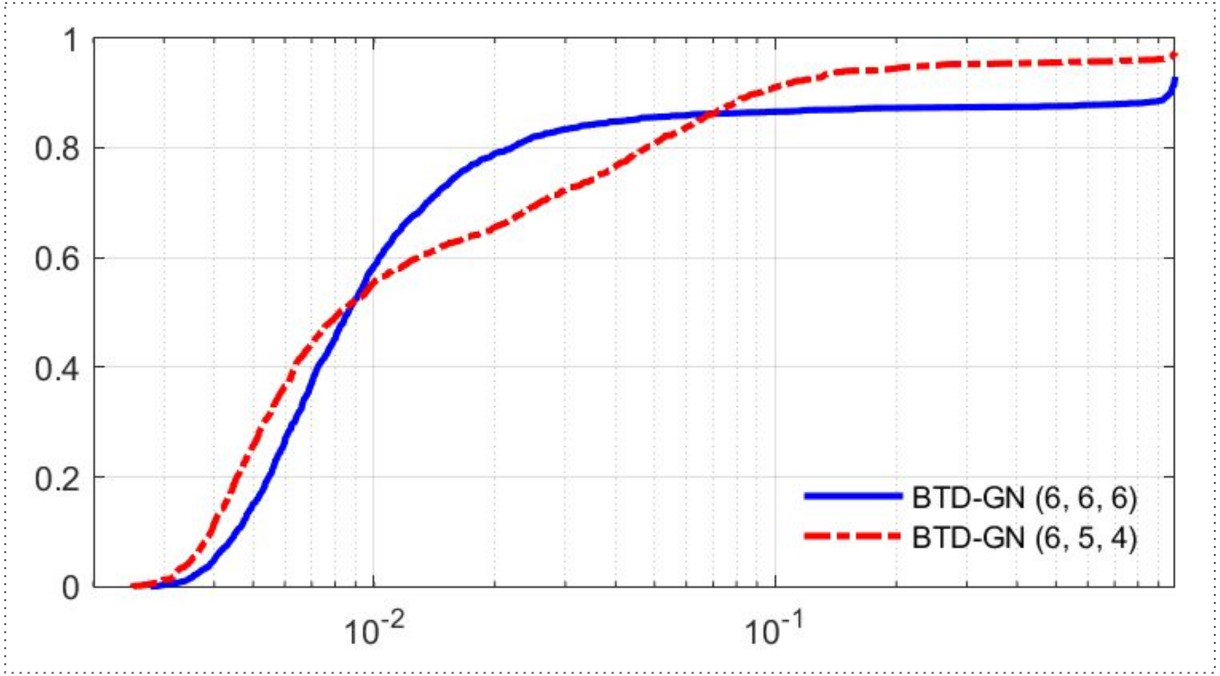



Figure: Empirical CDFs of NMSE over estimated blocks by BTD-GN for 500 realizations.

# A. Unconstrained BTD

| | AGL (3, 6) | AGL (5, 8) | BTD-GN (6, 5, 4) | BTD-GN (6, 6, 6) |
|---|---|---|---|---|
| ECDF < 0.01 | 83% | 80% | 58.8% | 58.8% |
| True model detection | 97.6% | 87.6% | - | - |
| Rank Inversion | - | - | 51% | - |

# A. Unconstrained BTD

## Experiment Setup

- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{X}$ and $\mathbf{N}$ in an i.i.d. fashion from the standard normal distribution.
- Constrained to employ an arbitrarily initial solution (random)
- AGL is run from this initial solution and using $\gamma = \gamma 0$
- Same γ-sweeping procedure
- For BTD-GN and AGL, 30 and 12 initializations taken, respectively
- SNR of 10 dB

## Constrained Scenario

- All algorithms are followed by Cadzow's algorithm to enforce the low-rank Hankel constrain

# A. Unconstrained BTD

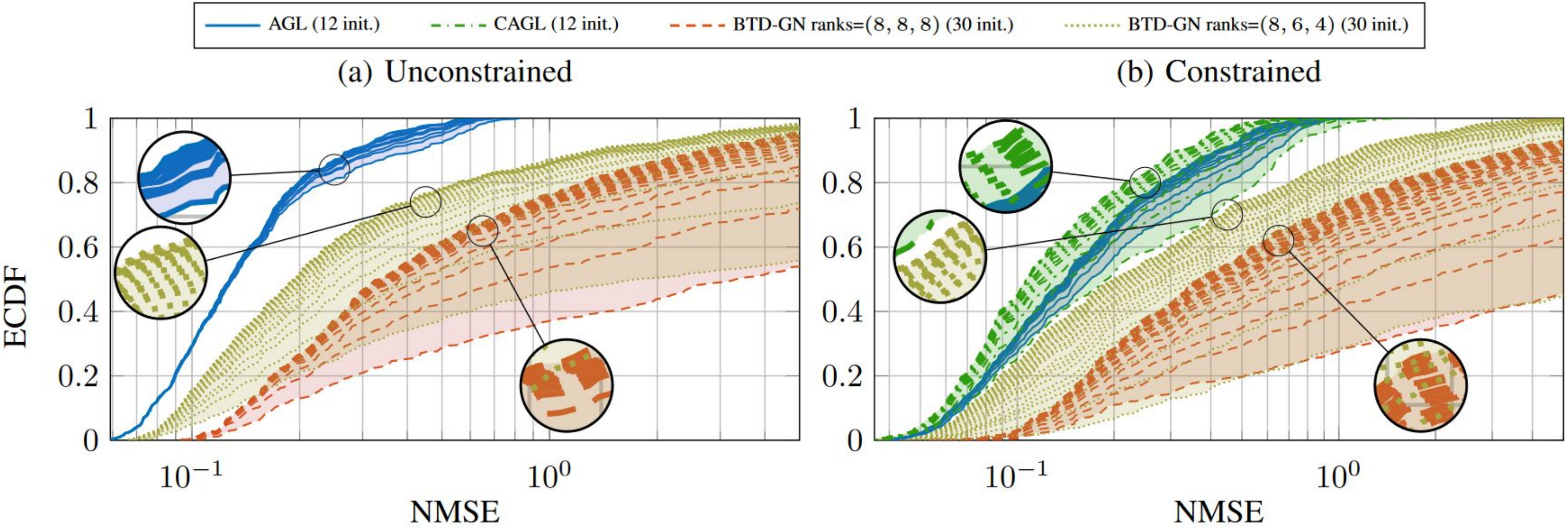



Figure 2: Empirical CDFs of NMSE over estimated blocks by AGL and BTD-GN for 500 realizations, with 12 and 30 initializations, respectively. The best solution among is chosen.

# A. Unconstrained BTD

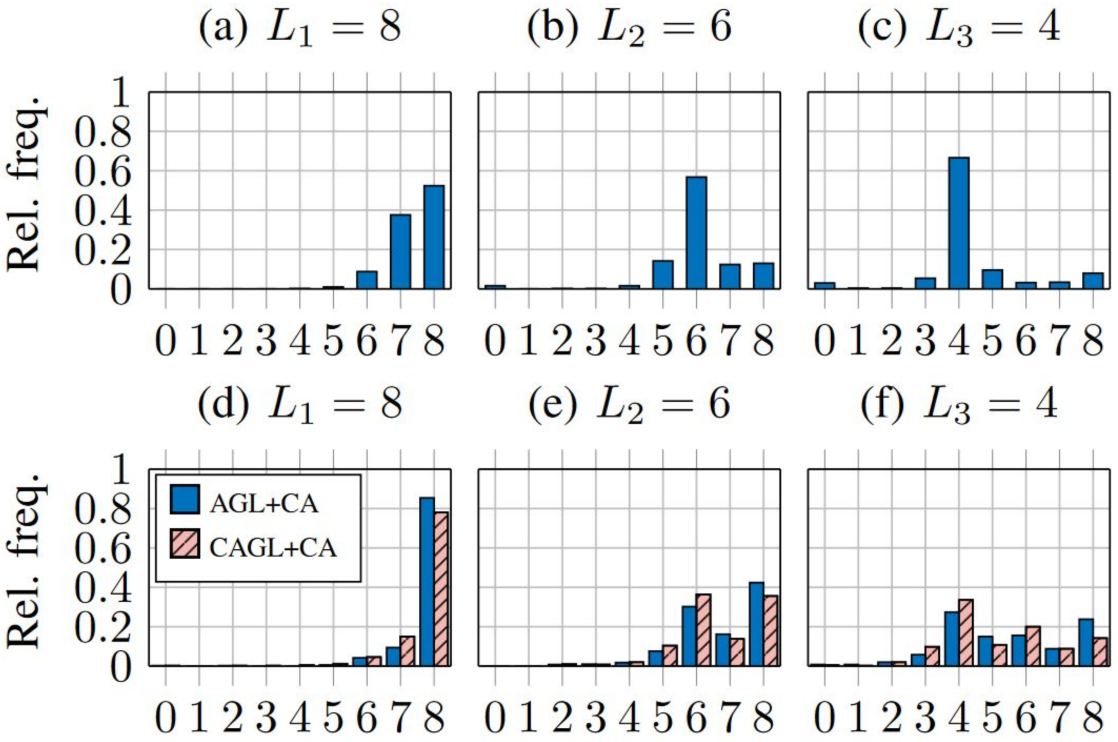


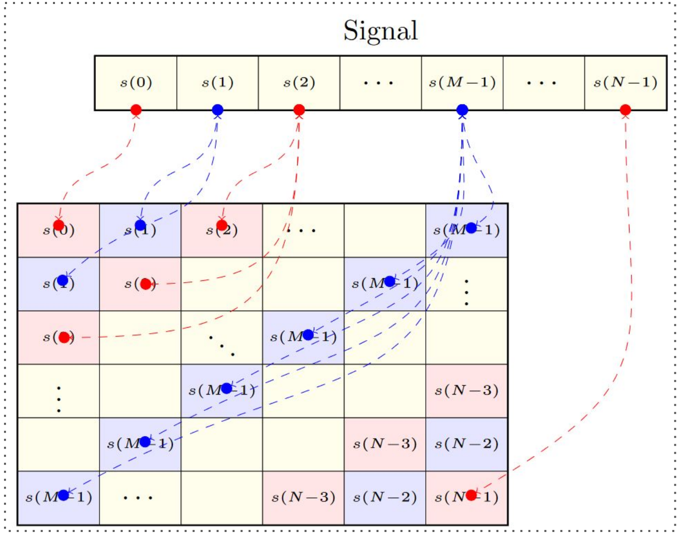

Figure 3: Proportion of block ranks estimated by AGL and CAGL with SNR = 10dB.

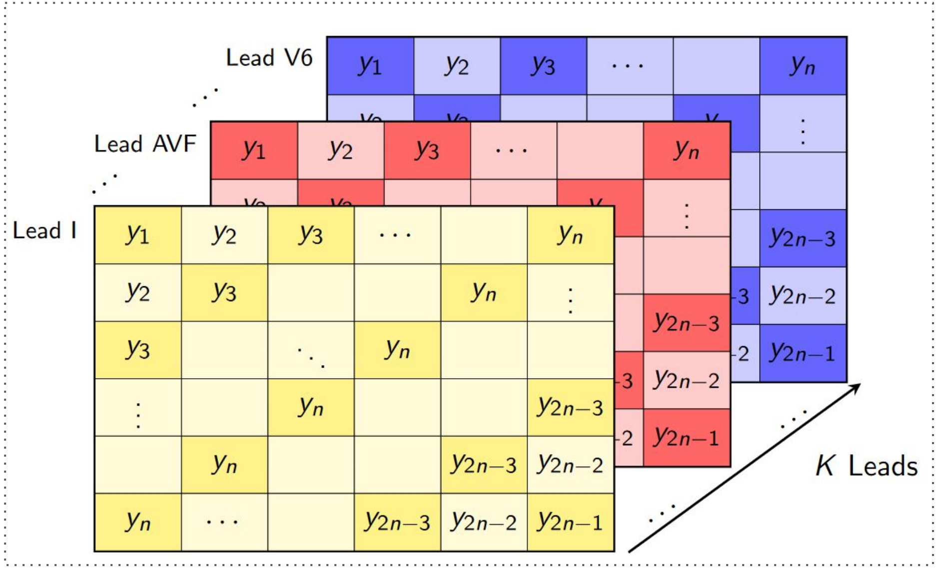
# V. Experimental Results with ECG Data

# A. Tensor representation of ECG signals

**Experimental Setup**

- Each ECG lead mapped onto a Hankel matrix. Stack in the data tensor 3rd-mode.



# B. Semi-synthetic AF data

**Stridh Model**

- Sum real ventricular activity and a signal that mimics the sawtooth pattern (**AF**)
- Random spatial signature and Additive white Gaussian noise (AWGN)

$$\mathbf{Y} = \mathbf{V} + \alpha\, \mathbf{x}\mathbf{s}^{\mathsf{T}} + \mathbf{N} \in \mathbb{R}^{12 \times N}$$

$$s(n) = -\textstyle\sum_{p=1}^{P} a_p(n) \sin\left(p\,\theta(n)\right)$$

| Model | $P$ | $a$ | $\Delta a$ | $f_a$ | $F_s$ | $f_0$ | $\Delta f$ | $F_f$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 5 | 150 | 50 | 0.08 | 1000 | 6 | 0.2 | 0.10 |
| 2 | 3 | 60 | 18 | 0.50 | 1000 | 8 | 0.3 | 0.23 |

Table III: Parameters of the synthetic AA signal model.

# B. Semi-synthetic AF data

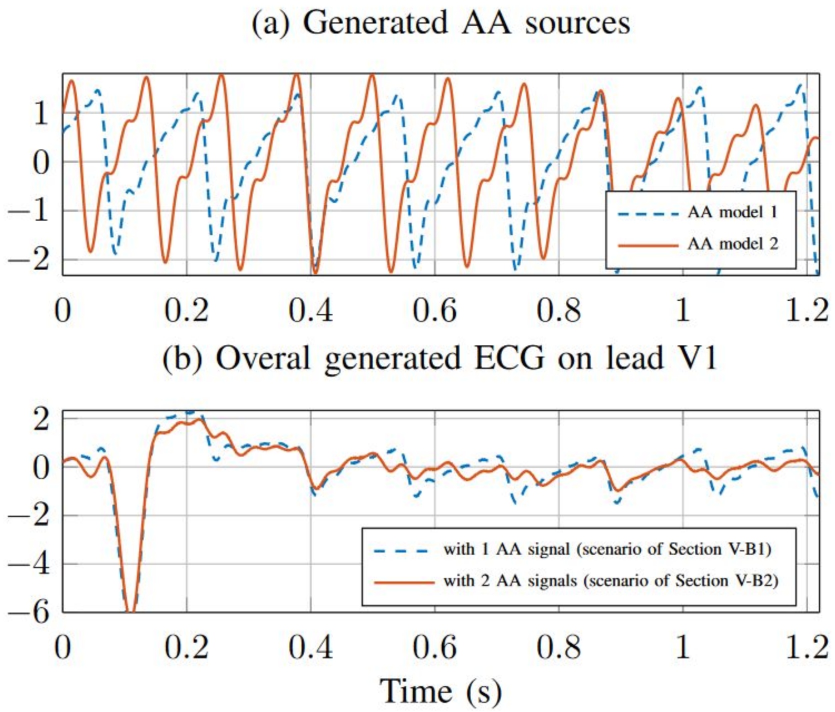



Figure 5: Examples of generated semi-synthetic models

# B. Semi-synthetic AF data

## BTD-GN vs CAGL

- **BTD-GN**: Produces good results with a proper combination of $R$, $L$ and initial point.
- **CAGL**: Only requires choosing a reasonable range for $\gamma$, and behaves much more robustly with regard to initialization.

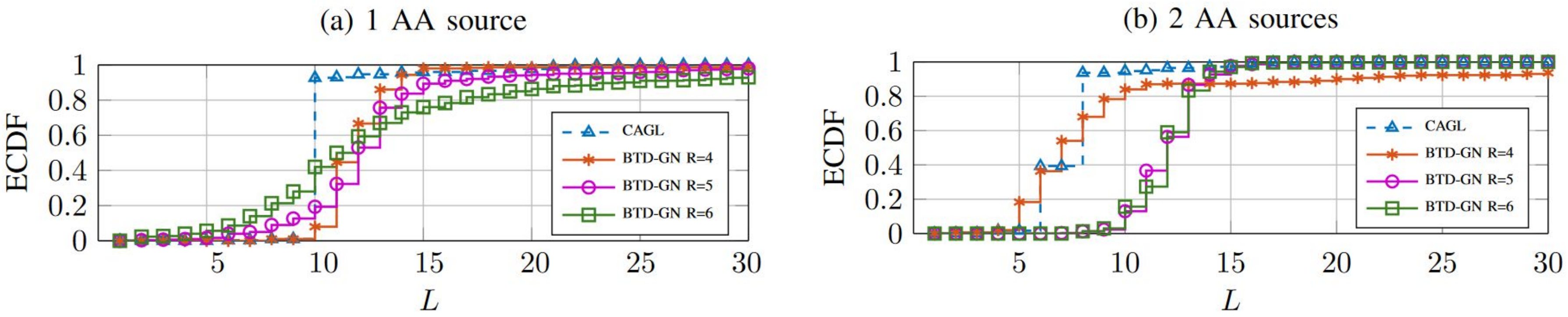



Figure 6: Empirical distribution of rank chosen by **CAGL** for the **AA** source and of rank $L$ yielding the best **AA** extraction for **BTD-GN** with different numbers of blocks $R$.

# B. Semi-synthetic AF data

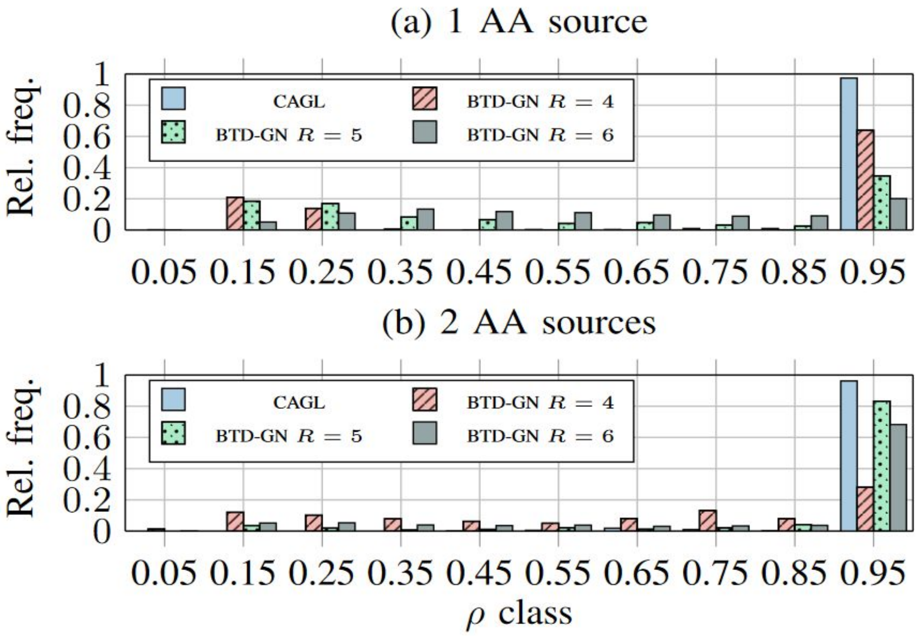



Figure 7: Histogram of computed correlation coefficient ρ (in absolute value) between true and estimated AA sources.

# C. Real **AF** data

## 12-lead ECG recordings

- Cardiology Department of Princess Grace Hospital
- **Very short ECG** segments: 1.08 to 1.40 seconds (Largest TQ interval)
- Zero-phase forward-backward Type-II Chebyshev bandpass filter (cutoff: 0.5 and 40 Hz)

TABLE IV: Block ranks of the ECG sources extracted by CAGL and characteristics of the potential AA sources.

| Patient | Non-AA source ranks | AA source | SC (%) | DF (Hz) | $\hat{\kappa}$ | AA source rank |
|---|---|---|---|---|---|---|
| 1 | 3, 9, 10, 16, 18 | 1 | 80.23 | 6.44 | 163.69 | 8 |
| 2 | 28, 32 | 1 | 59.36 | 6.20 | 116.77 | 29 |
| | | 2 | 82.05 | 6.20 | 163.77 | 12 |
| 3 | 21, 27, 29 | 1 | 66.13 | 6.20 | 132.29 | 20 |
| 4 | 8, 16, 20 | 1 | 74.51 | 5.96 | 196.16 | 10 |
| 5 | 32, 38, 39 | 1 | 91.70 | 5.72 | 348.42 | 10 |
| | | 2 | 78.63 | 5.01 | 166.95 | 21 |

Table IV: Block ranks of the **ECG** sources extracted by **CAGL** and characteristics of the potential **AA** sources.

# C. Real **AF** data

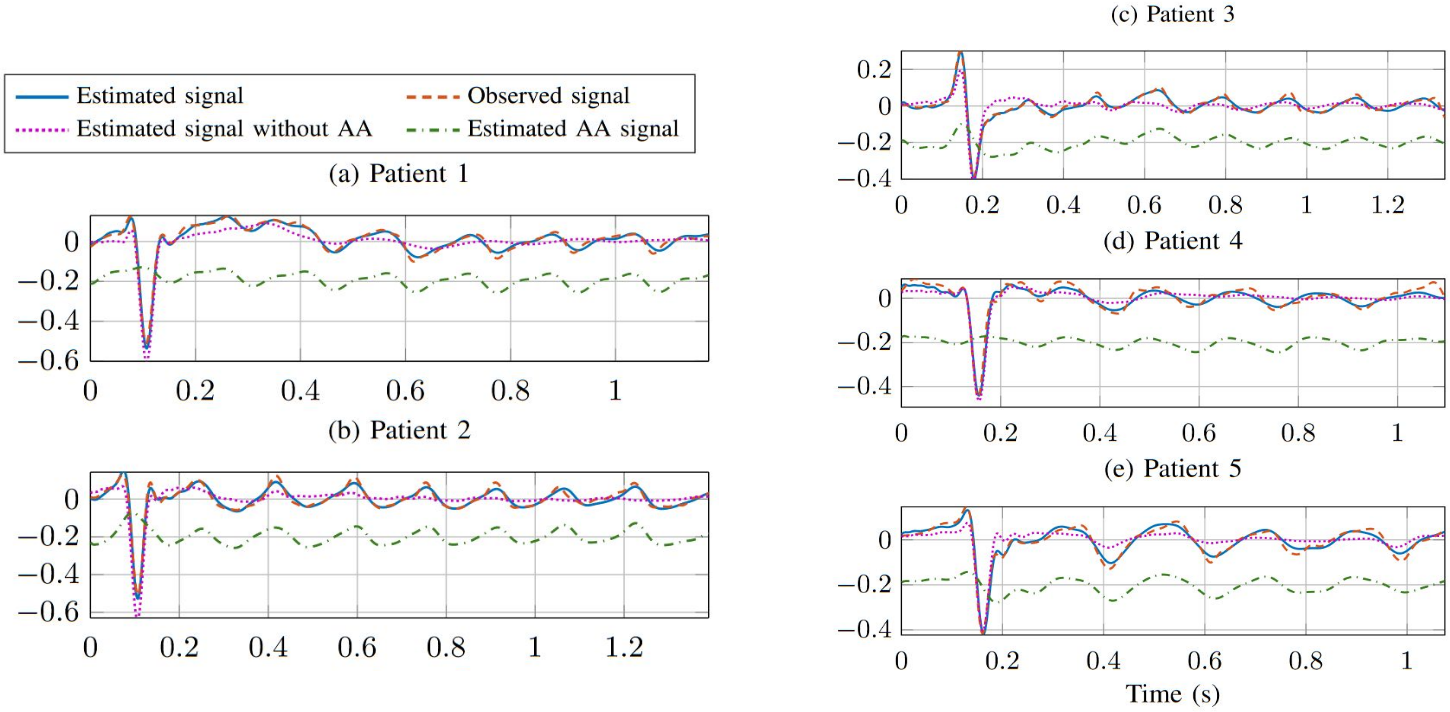



Figure 8: Results produced by CAGL with real-world ECG data: observed and estimated signals at lead V1. Estimated **AA** signals are vertically shifted for ease of visualization.

# C. Real **AF** data

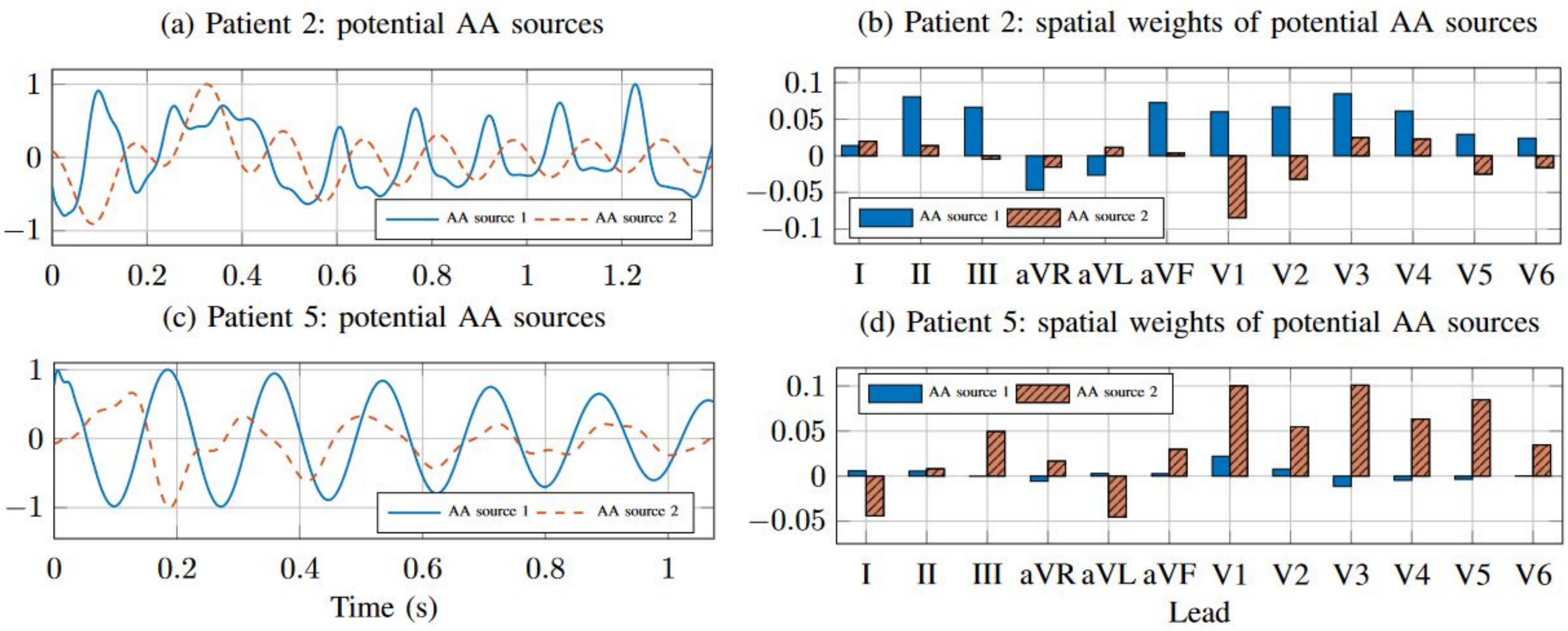



Figure 9: Results produced by CAGL with ECG data from Patients 2 and 5.

# VI. Conclusion

# VI. Conclusion

## Contributions

- Allows simultaneous estimation of model parameters and factors
- The resulting subproblems can be solved by existing group lasso method
- Experimental results show that AGL approach and its constrained version is much more robust with respect to initialization than the conventional least-squares approach

## Clinical Impact

- Extract AA by using a structured low-rank approximation method
- Without the need of choosing structural parameters *a priori*

## Further Work

- How to automatically tune $\gamma$ in practice (CAGL)
- Physiological interpretation of characteristics and spatial signatures diversity
- Multiple atrial sources analysis and enlarge database of AF patients

# Supplementary Material

# Algorithm for unconstrained blocks

## AGL e CAGL

- Solve Group Lasso Subproblems, where $\mathscr{W}_{\mathbf{A}}^{(t)}(\mathbf{A})$ is a linear map depending on $\mathbf{B}^{(t\text{-}1)}$ and $\mathbf{X}^{(t\text{-}1)}$

$$\min_{\mathbf{A}\in\mathbb{C}^{I\times LR}} \frac{1}{2}\|\mathcal{Y}-\mathscr{W}_{\mathbf{A}}^{(t)}(\mathbf{A})\|_F^2+\gamma\,\|\mathbf{A}\|_{2,1}+\tfrac{\tau}{2}\|\mathbf{A}-\hat{\mathbf{A}}^{(t-1)}\|_F^2$$

**Inputs:** Data tensor $\mathcal{Y}$, penalty parameter $\gamma$, proximal term weight $\tau$, initial point $(\mathbf{A}^{(0)},\mathbf{B}^{(0)},\mathbf{X}^{(0)})$

**Outputs:** Approximate BTD factors $(\mathbf{A},\mathbf{B},\mathbf{X})$

1: $t \leftarrow 1$

2: **while** stopping criteria not met **do**

3: Solve group lasso subproblem (12) to obtain $\mathbf{A}^{(t)}$ from $\mathbf{A}^{(t-1)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$

4: Solve group lasso subproblem in $\mathbf{B}$ analogous to (12) to obtain $\mathbf{B}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t-1)}$ and $\mathbf{X}^{(t-1)}$

5: **for** $r=1,\dots,R$ **do**

6: $L_r^{(t)} \leftarrow \operatorname{rank}\left(\mathbf{A}_r^{(t)}(\mathbf{B}^{(t)})_r^{\mathsf{T}}\right)$

7: $(\mathbf{A}_r^{(t)},\mathbf{B}_r^{(t)}) \leftarrow \mathtt{slra}(\mathbf{A}_r^{(t)}(\mathbf{B}_r^{(t)})^{\mathsf{T}}, L_r^{(t)})$

8: $(\mathbf{A}_r^{(t)},\mathbf{B}_r^{(t)}) \leftarrow ([\mathbf{A}_r^{(t)}\ \mathbf{0}_{I\times L-L_r^{(t)}}],[\mathbf{B}_r^{(t)}\ \mathbf{0}_{I\times L-L_r^{(t)}}])$

9: Solve group lasso subproblem in $\mathbf{X}$ analogous to (12) to obtain $\mathbf{X}^{(t)}$ from $\mathbf{A}^{(t)}$, $\mathbf{B}^{(t)}$ and $\mathbf{X}^{(t-1)}$

10: $t \leftarrow t+1$

Table I: Pseudocode for the unconstrained AGL algorithm (lines 5–8 must be omitted).

# Constrained BTD: Further Results

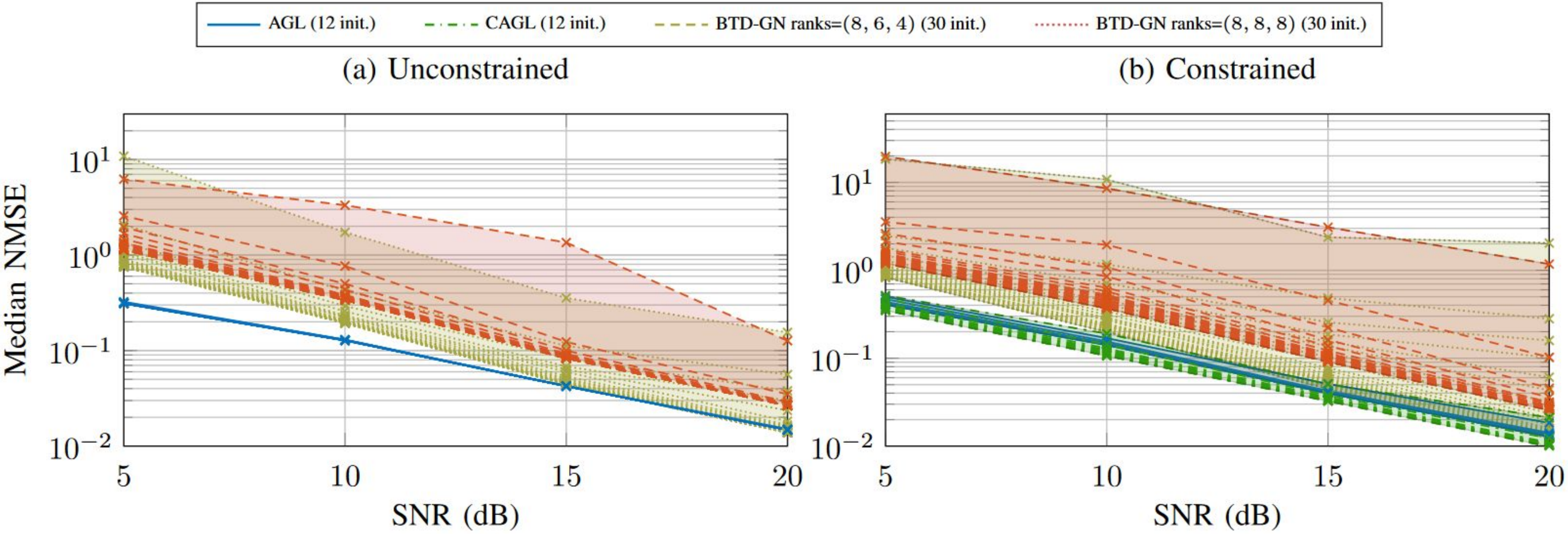



Figure 4: Median NMSE over estimated blocks attained by AGL and BTD-GN for 500 realizations of a noisy random BTD model in the unconstrained and (Hankel-)constrained scenarios.

# Unconstrained BTD

## Cadzow's Algorithm (CA)

- Performs alternating projections onto the Hankel subspace with rank bounded by a prescribed value.
- Approximations provided by CA are not (locally) optimal, but they are often satisfying in practice.

| | |
|---|---|
| **Inputs:** | Data matrix $\mathbf{Y}$, target rank $L$, tolerance $\epsilon_{\text{CA}}$ and maximum number of iterations $T_{\text{CA}}$ |
| **Outputs:** | Low-rank Hankel approximation $\mathbf{H}$ of $\mathbf{Y}$ |

```
1: Initialization: t ← 1,  H^(0) ← Y
2: for t = 1, 2, ... do
3:     H^(t) ← P_H(H^(t−1))
4:     Compute the SVD: H^(t) = UΣV^T
5:     Truncate the computed SVD at rank L: H^(t) ← U_L Σ_L V_L^T
6:     if ||H^(t) − H^(t−1)||_F < ε_CA ||H^(t−1)||_F or t = T_CA then
7:         break for loop and output H = H^(t)
```

1: **Initialization**: $t \leftarrow 1, \quad \mathbf{H}^{(0)} \leftarrow \mathbf{Y}$
2: **for** $t = 1, 2, \ldots$ **do**
3: $\quad \mathbf{H}^{(t)} \leftarrow \mathcal{P}_{\mathcal{H}}(\mathbf{H}^{(t-1)})$
4: $\quad$ Compute the SVD: $\mathbf{H}^{(t)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}}$
5: $\quad$ Truncate the computed SVD at rank $L$: $\mathbf{H}^{(t)} \leftarrow \mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_L^{\mathsf{T}}$
6: $\quad$ **if** $\|\mathbf{H}^{(t)} - \mathbf{H}^{(t-1)}\|_F < \epsilon_{\text{CA}} \|\mathbf{H}^{(t-1)}\|_F$ or $t = T_{\text{CA}}$ **then**
7: $\quad\quad$ **break for loop and output** $\mathbf{H} = \mathbf{H}^{(t)}$

Table II: Pseudocode for the Cadzow's algorithm.

# Stridh Model

**Details**

- Sum real ventricular activity and a signal that mimics the sawtooth pattern (**AF**)
- Random spatial signature and Additive white Gaussian noise (AWGN)

$$\mathbf{Y} = \mathbf{V} + \alpha\, \mathbf{x}\mathbf{s}^{\mathsf{T}} + \mathbf{N} \in \mathbb{R}^{12\times N}$$

$$a_p(n) = \frac{2}{p\pi}\left[a + \Delta a \sin\left(2\pi \frac{f_a}{F_s} n\right)\right]$$

$$s(n) = -\sum_{p=1}^{P} a_p(n) \sin\left(p\,\theta(n)\right)$$

$$\theta(n) = 2\pi \frac{f_0}{F_s} n + \left(\frac{\Delta f}{F_f}\right) \sin\left(2\pi \frac{F_f}{F_s} n\right)$$

| Model | $P$ | $a$ | $\Delta a$ | $f_a$ | $F_s$ | $f_0$ | $\Delta f$ | $F_f$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 5 | 150 | 50 | 0.08 | 1000 | 6 | 0.2 | 0.10 |
| 2 | 3 | 60 | 18 | 0.50 | 1000 | 8 | 0.3 | 0.23 |

Table III: Parameters of the synthetic AA signal model.