Learning-based Sparse Recovery for Massive Random Access

Srikrishna Bhashyam

IIT Madras

June 15, 2023

IIT Kanpur

A. P. Sabulal, S. Bhashyam, "Joint Sparse Recovery using Deep Unfolding With Application to Massive Random Access," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 5050-5054.

U. K. Sreeshma Shiv, S. Bhashyam, C. R. Srivatsa and C. R. Murthy, "Learning-Based Sparse Recovery for Joint Activity Detection and Channel Estimation in Massive Random Access Systems," in IEEE Wireless Communications Letters, vol. 11, no. 11, pp. 2295-2299, Nov. 2022

Recent Research: Overview

Research Interests

- Communication and Information Theory
- Statistical Inference

Recent work

- MIMO
- Sequential hypothesis testing
- Model-based learning for wireless communication
  - Learning-based sparse recovery for massive random access

More details at https://www.ee.iitm.ac.in/skrishna/

Multiple-Input Multi-Output (MIMO)

Multiple power constraints

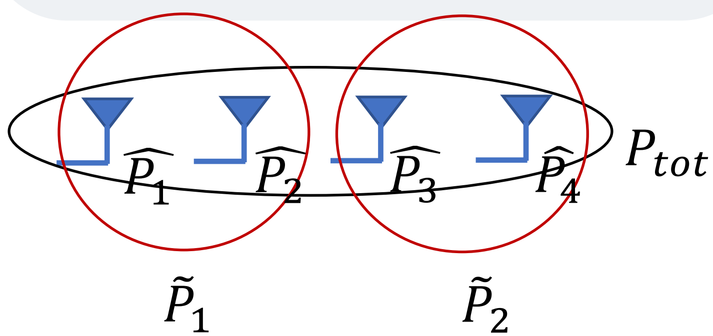
- Distributed antennas
    - Cell-free MIMO, CoMP
- Hardware constraints

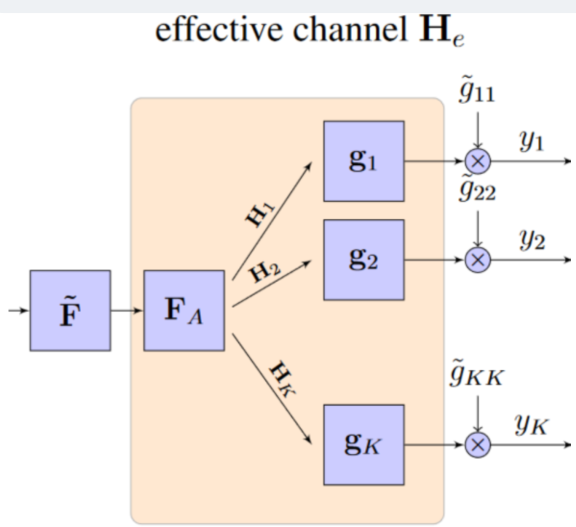
Distributed beamforming

- Limited coordination

Precoding with partial channel knowledge

- Hybrid beamforming

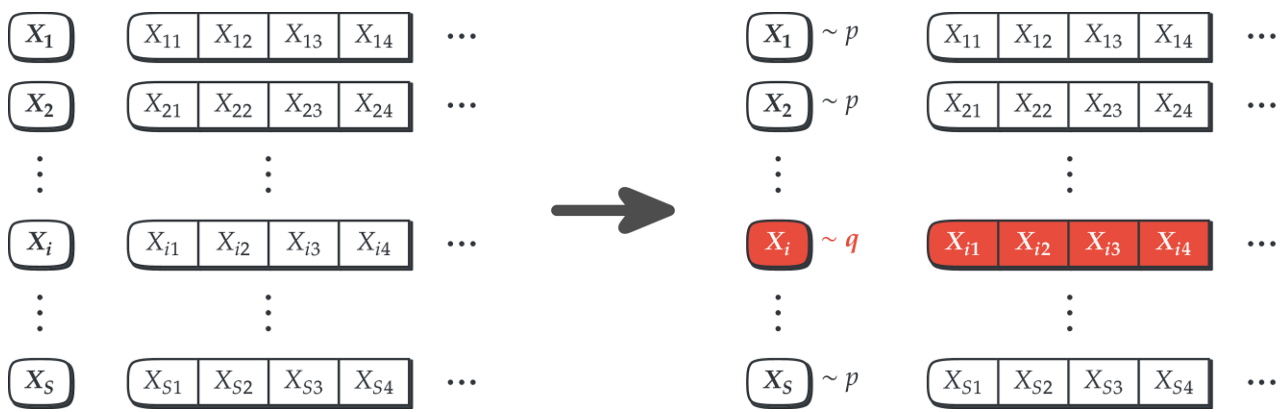




- S. S. Nair and S. Bhashyam, "Hybrid beamforming in MU-MIMO using partial interfering beam feedback," in IEEE Communications Letters, vol. 24, no. 7, pp. 1548-1552, July 2020.
- S. S. Nair and S. Bhashyam, "Robust Nonlinear Precoding in MU-MIMO using Partial Interfering Beam Feedback," *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, United Kingdom, 2023, pp. 1-6.
- R. Chaluvadi, S. S. Nair, S. Bhashyam, "Optimal Multi-antenna Transmission with Multiple Power Constraints," IEEE Transactions on Wireless Communications, vol. 18, no. 7, pp. 3382-3394, July 2019.
- V. N. Moothedath and S. Bhashyam, "Distributed Pareto Optimal Beamforming for the MISO Multi-band Multi-cell Downlink," in IEEE Transactions on Wireless Communications, vol. 19, no. 11, pp. 7196-7209, Nov. 2020.

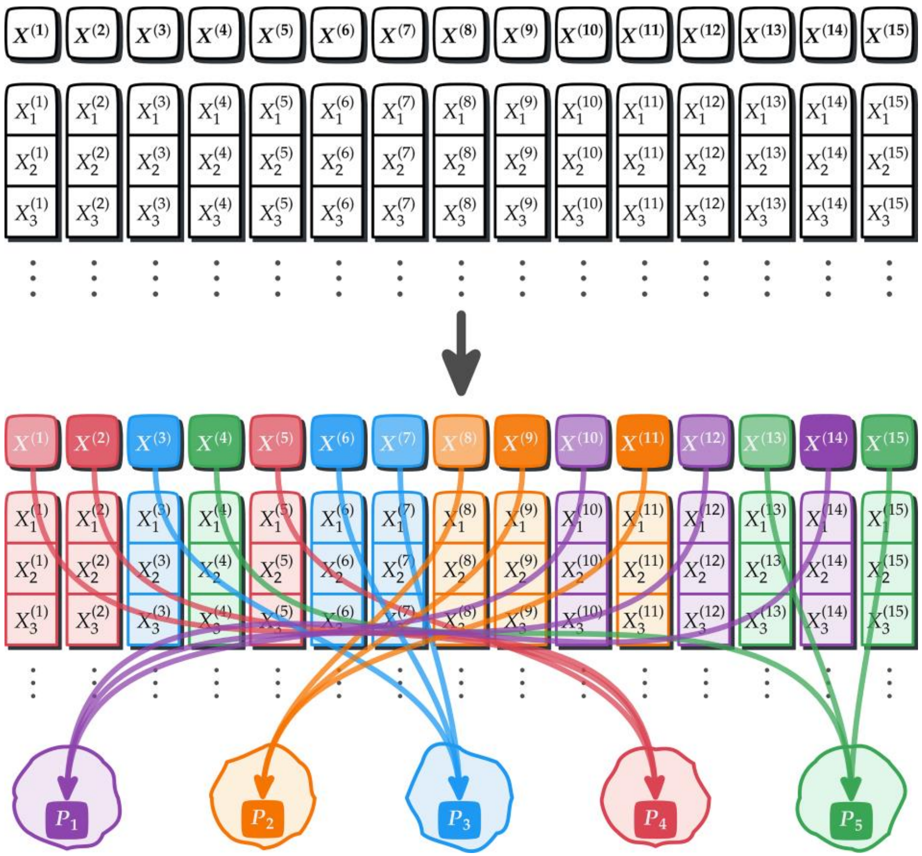
Sequential hypothesis testing in Multi-Armed Bandits

- Anomaly detection –> Generalized hypothesis testing
- Parametric setting: Vector exponential family
- Active sampling under constraints



- G. R. Prabhu, S. Bhashyam, A. Gopalan and R. Sundaresan, "Sequential Multi-Hypothesis Testing in Multi-Armed Bandit Problems: An Approach for Asymptotic Optimality," in IEEE Transactions on Information Theory, vol. 68, no. 7, pp. 4790-4817, July 2022.
- Aditya Deshmukh, Venugopal V. Veeravalli & Srikrishna Bhashyam (2021) Sequential controlled sensing for composite multihypothesis testing, Sequential Analysis, 40:2, 259-289.

Sequential hypothesis testing



- Nonparametric setting
- Anomaly detection & Clustering

- S. C. Sreenivasan and S. Bhashyam, "Sequential Nonparametric Detection of Anomalous Data Streams," in IEEE Signal Processing Letters, vol. 28, pp. 932-936, 2021.
- S. C. Sreenivasan, S. Bhashyam, Nonparametric Sequential Clustering of Data Streams with Composite Distributions, Signal Processing (2022).

Model-based learning for wireless communication:

Learning-based sparse recovery for massive random access

Model-based learning for wireless communication

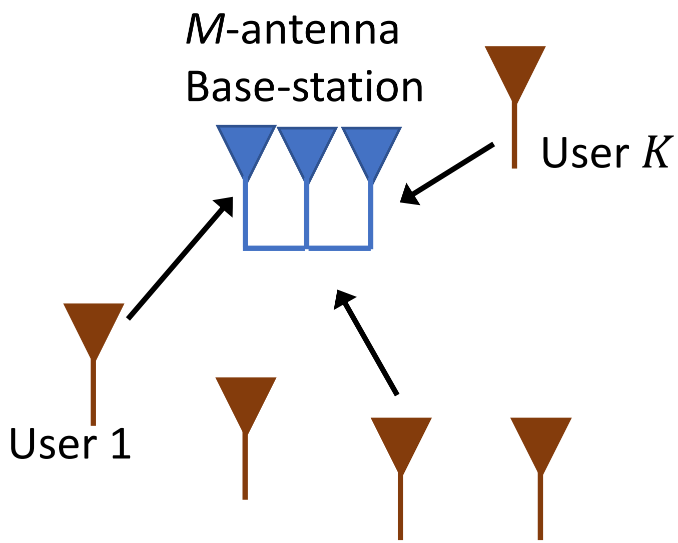
Model-based signal processing	Deep learning
Domain knowledge	Data-driven, uses large data sets
Analysis and interpretation	Not easy to interpret or analyse

Model-based learning: Hybrid approach

- Deep unfolding
- Model-aided networks

N. Shlezinger, J. Whang, Y. C. Eldar and A. G. Dimakis, "Model-Based Deep Learning: Key Approaches and Design Guidelines," 2021 IEEE Data Science and Learning Workshop (DSLW), 2021, pp. 1-6, doi: 10.1109/DSLW51110.2021.9523403.

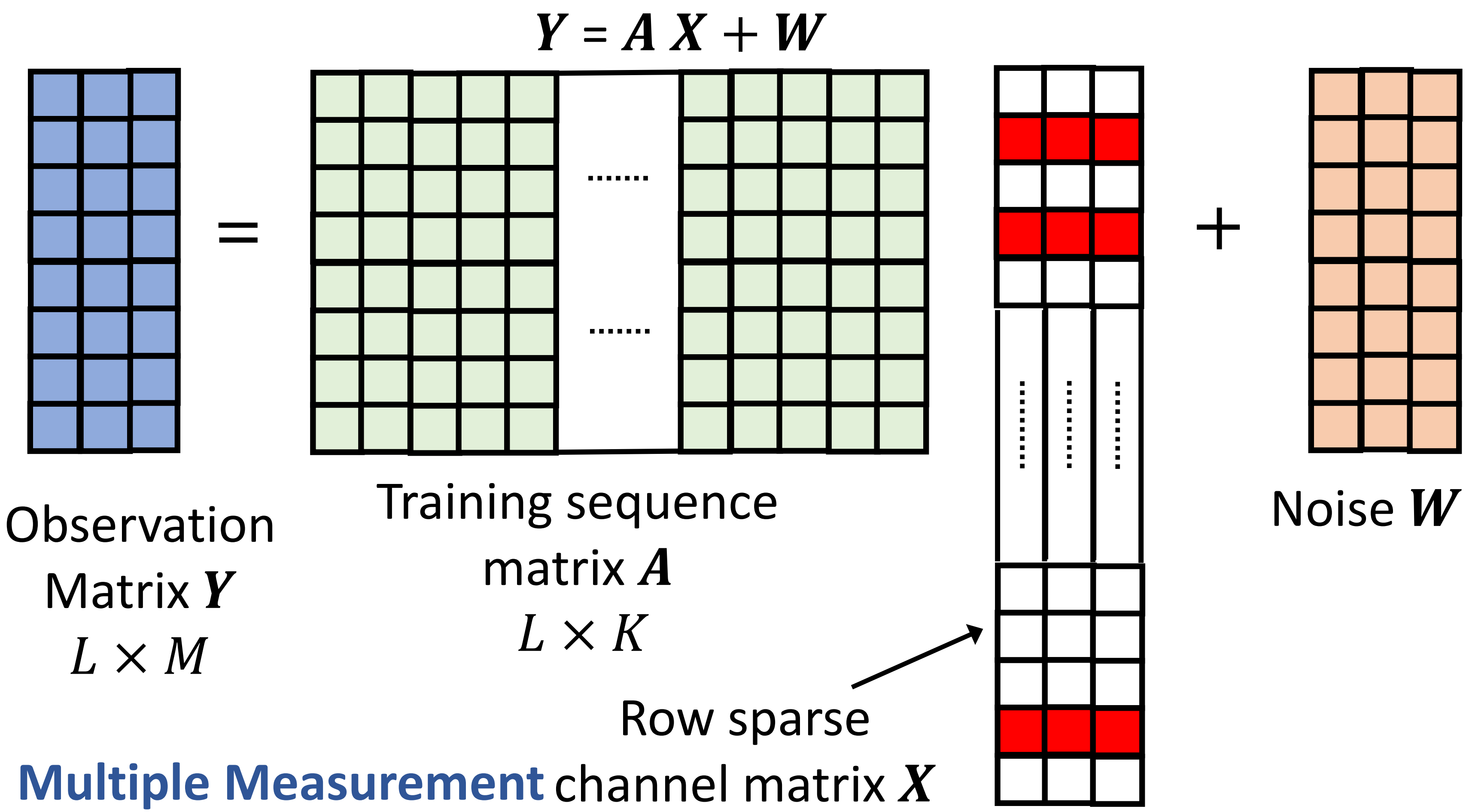
Massive Random Access



- Small fraction of users are active at any given time
- Identify the active users
- Estimate channel corresponding to the active users

- Active users send training sequences of length L

Joint Sparse Recovery: Activity detection and Channel estimation



Multiple Measurement Vector (MMV) problem

Sparse recovery

Iterative soft thresholding

ISTA

Approximate message passing

AMP
OAMP
Vector AMP

Matching Pursuit

OMP
CoSaMP

Alternating direction method of multipliers

MMV-ADM

Sparse Bayesian learning

SBL
M-SBL

Learning-based sparse recovery

Iterative soft thresholding

ISTA
LISTA
TISTA, MMV-TISTA
L-MMSE-MMV-TISTA

Approximate message passing

AMP
OAMP
Vector AMP
L-AMP
VAMP-net

Matching Pursuit

OMP
CoSaMP

Alternating direction method of multipliers

MMV-ADM
MMV-MADM
L-MMV-MADM

Sparse Bayesian learning

SBL
M-SBL
L-SBL

Our Work

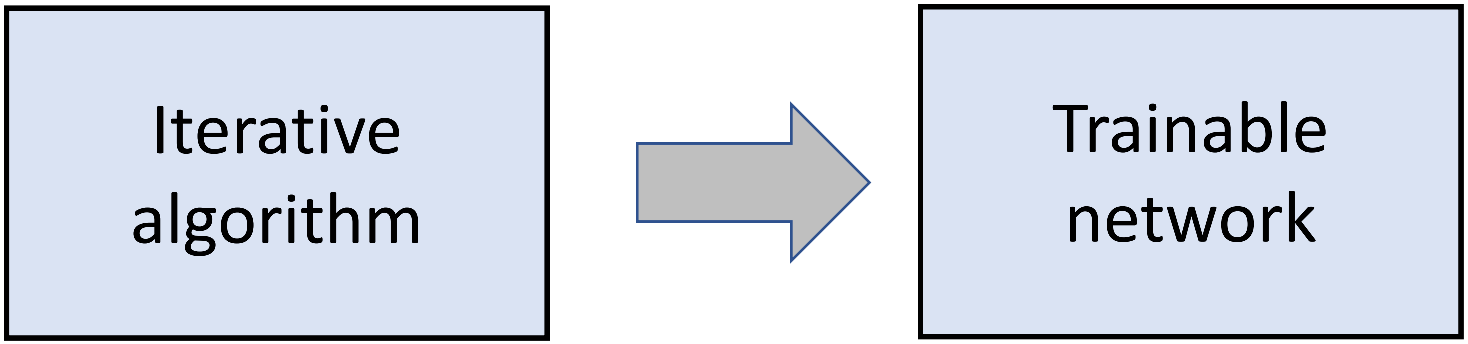
- Proposed methods
  - MMV-MADM and LMMV-MADM
    - Uses deep unfolding, modified cost
  - MMV-TISTA and learnt version
    - Replaces denoiser with a model-based neural network
- New comparisons
  - Performance-complexity trade-offs

A. P. Sabulal, S. Bhashyam, "Joint Sparse Recovery using Deep Unfolding With Application to Massive Random Access," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 5050-5054.

U. K. Sreeshma Shiv, S. Bhashyam, C. R. Srivatsa and C. R. Murthy, "Learning-Based Sparse Recovery for Joint Activity Detection and Channel Estimation in Massive Random Access Systems," in IEEE Wireless Communications Letters, vol. 11, no. 11, pp. 2295-2299, Nov. 2022

Deep Unfolding

Technique 1: Deep unfolding



- Each iteration is a layer
- Parameters in each layer untied and trained

John R. Hershey, Jonathan Le Roux, and Felix Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," CoRR, vol. abs/1409.2574, 2014.

Alexios Balatsoukas-Stimming and Christoph Studer, "Deep unfolding for communications systems: A survey and some new directions," arXiv preprint arXiv:1906.05774, 2019.

V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," IEEE Signal Processing Magazine, vol. 38, no. 2, pp. 18–44, 2021.

Proposed method: LMMV-MADM

MMV-ADM → Modified MMV-ADM → Trainable network

- MMV-ADM
  - Based on alternating direction method of multipliers
- Modification of existing algorithm to help learning
  - Back-projected error
- Unfolding: Significant reduction in training overhead
- Two learning approaches: Supervised, Unsupervised

MMV-ADM

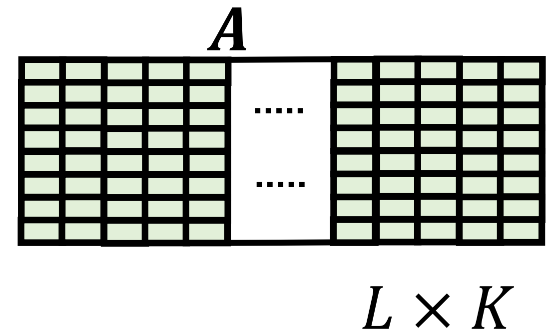
$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{2,1} + \frac{1}{2\mu}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_2$$

- Alternating direction method
- No matrix inversions ⇨ fast, scalable
- Convergence analysis feasible

H. Lu, X. Long, and J. Lv, "A fast algorithm for recovery of jointly sparse vectors based on the alternating direction methods," in Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 461–469.

Modified MMV-ADM

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{2,1} + \frac{1}{2\mu}\left\|\boldsymbol{A}^{\dagger}\boldsymbol{Y} - \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{X}\right\|_2$$



- Backprojected LS error instead of LS error
  - $\boldsymbol{A}^{\dagger}\boldsymbol{Y} - \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{X}$ instead of $\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}$
  - $\boldsymbol{A}^{\dagger} = \boldsymbol{A}^T[\boldsymbol{A}\boldsymbol{A}^T]^{-1}$

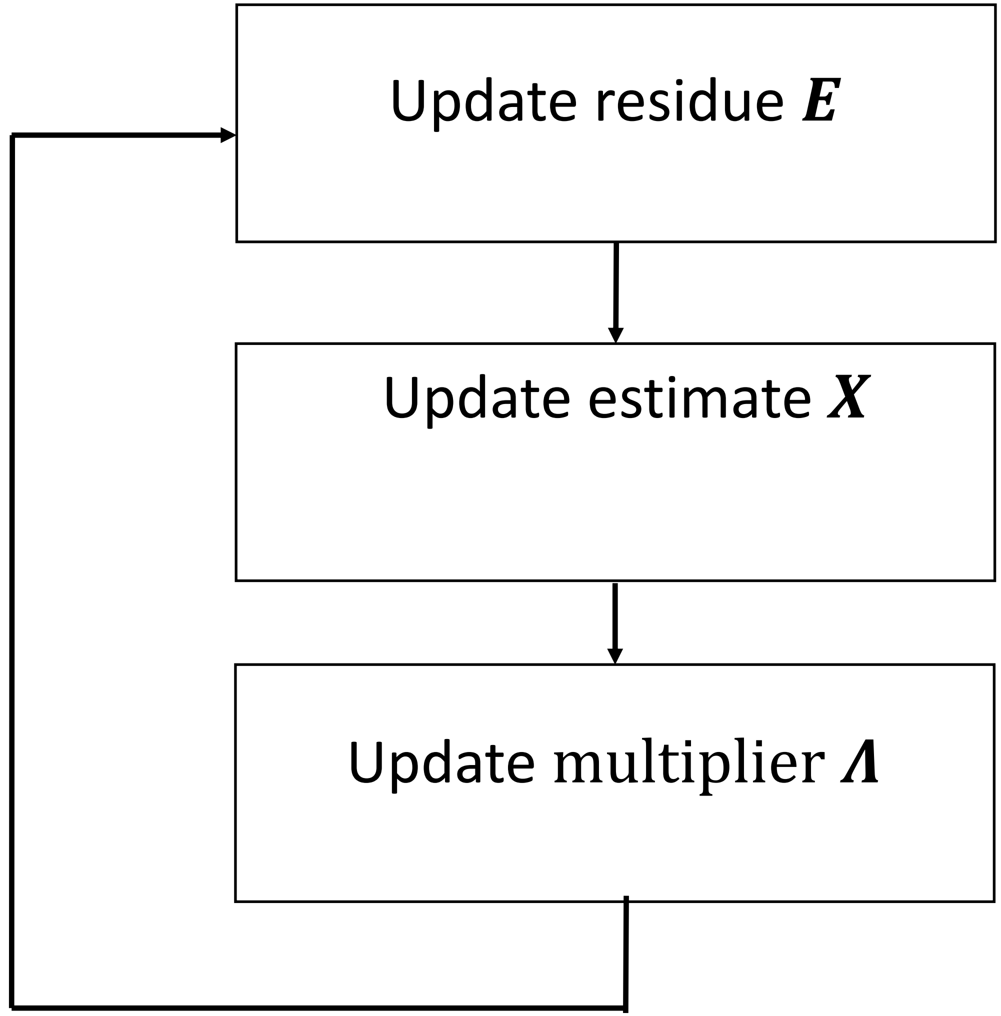
- Modified algorithm also fast, scalable
- Unfolding results in a easily trainable network

T. Tirer and R. Giryes, “Back-projection based fidelity term for ill-posed linear inverse problems,” IEEE Transactions on Image Processing, vol. 29, pp. 6164-6179, 2020.

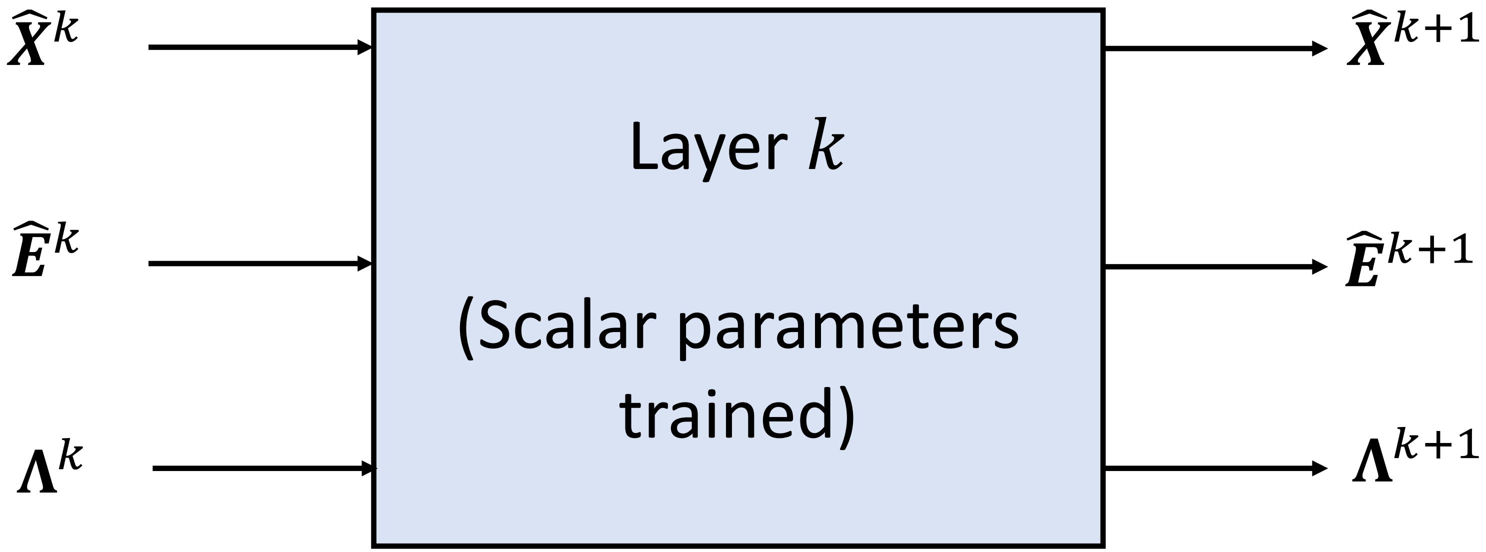
Modified MMV-ADM

- Augmented Lagrangian
  $L(\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{\Lambda})$
- $\boldsymbol{E} = \boldsymbol{A}^{\dagger}\boldsymbol{Y} - \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{X}$
- Initialize $\boldsymbol{X}, \boldsymbol{\Lambda}$

4 scalar parameters to choose: μ, β, γ, τ



Unfolded network



- One iteration of ADM algorithm is one layer

Training the network

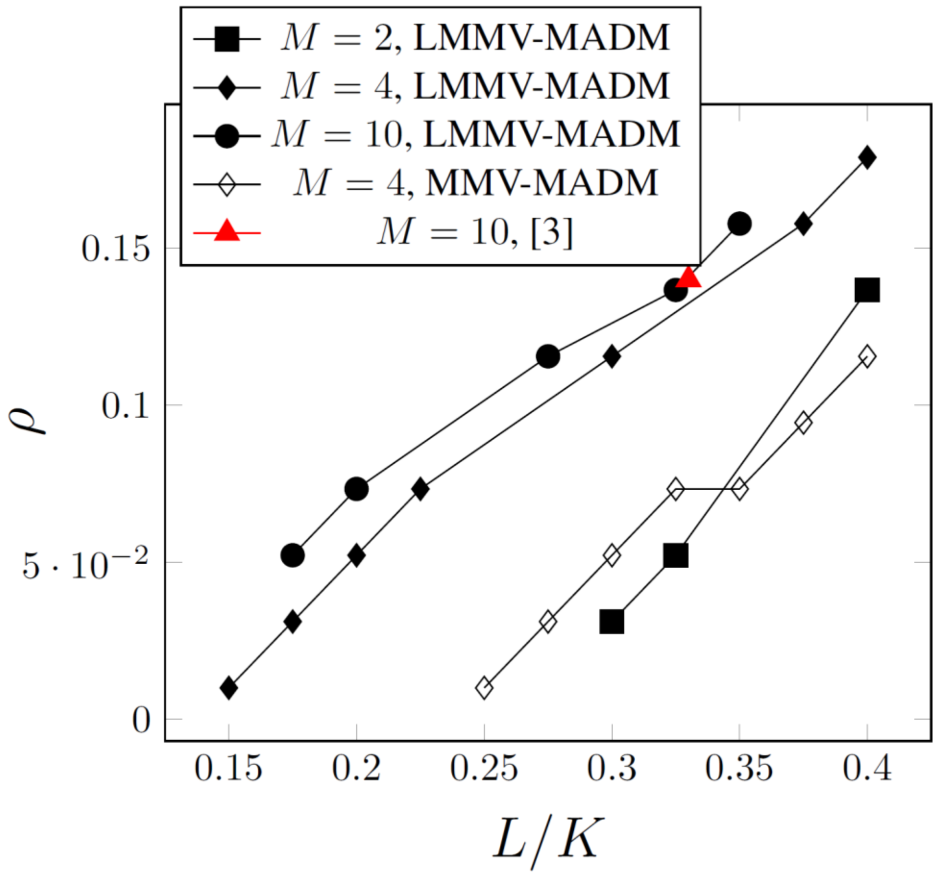
Supervised

- True $\boldsymbol{X}, \boldsymbol{Y}$ pairs available
- Generated using a channel model for training
- Layers trained sequentially
- MSE between layer output $\widehat{\boldsymbol{X}}^{k+1}$ and true $\boldsymbol{X}$ used as loss function for training

Unsupervised

- True $\boldsymbol{X}, \boldsymbol{Y}$ pairs not needed
- Loss function for training
  - $\lambda \left\| \widehat{\boldsymbol{X}}^{k+1} \right\|_{2,1/p}^{1/p} + \left\| \boldsymbol{Y} - \boldsymbol{A}\widehat{\boldsymbol{X}}^{k+1} \right\|_F^2$

Performance: Phase transition

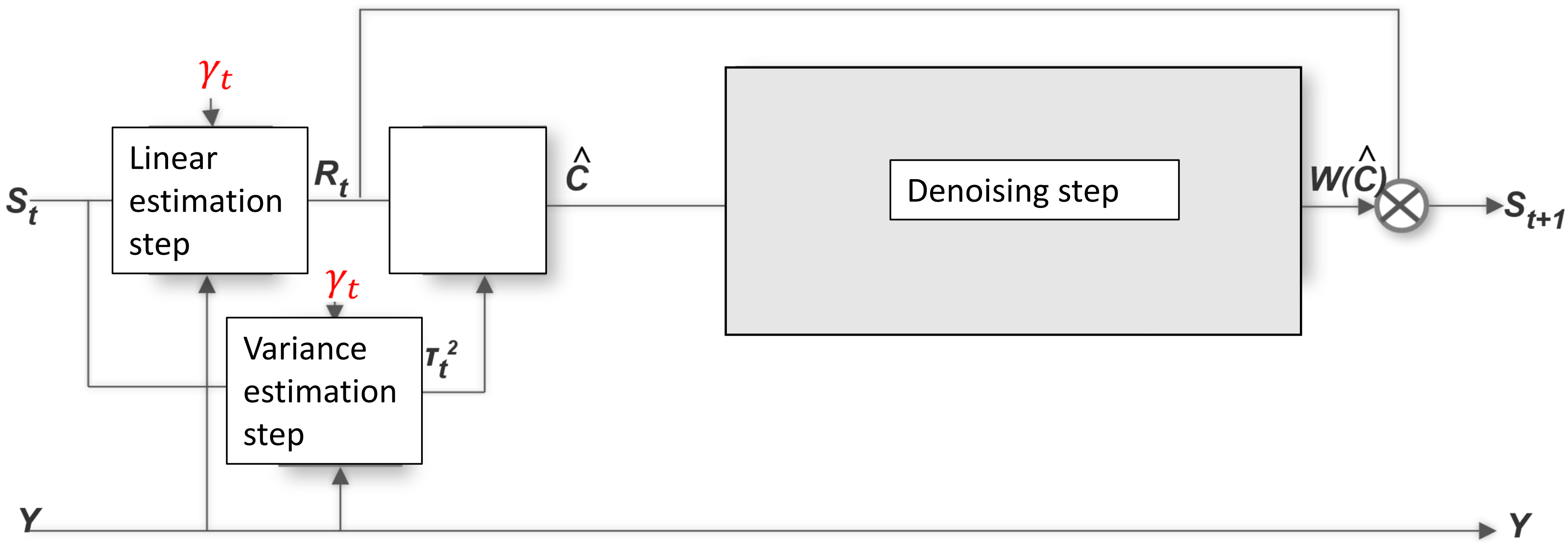


- $K = 500$, 20 layer network, MMV-MADM with 40 iterations
- Minimum L/K for a given activity probability ρ
- Training and test SNR at 30 dB
- Training at ρ = 0.2
- Success if NMSE < -20 dB

[3] T. Jiang, Y. Shi, J. Zhang, and K. B. Letaief, “Joint activity detection and channel estimation for IoT networks: Phase transition and computation-estimation tradeoff,” IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6212–6225, Aug 2019.

Model-based Neural Network for Denoising

Trainable ISTA (TISTA)



- Already uses deep unfolding
- Denoising step based on an approximate model

D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," IEEE Transactions on Signal Processing, vol. 67, no. 12, pp. 3113–3125, June 2019.

Technique 2: Model-based neural network

$$\boldsymbol{y}_t = \boldsymbol{h}_t + \boldsymbol{z}_t, \qquad\qquad \boldsymbol{Y} = [\boldsymbol{y}_1\ \boldsymbol{y}_2\ \dots\ \boldsymbol{y}_T]$$

- $\boldsymbol{h}_t$: Conditionally Gaussian $(\mathbf{0}, \boldsymbol{C}_\delta)$ given parameters δ
- $\boldsymbol{z}_t$: Gaussian $(\mathbf{0}, \sigma^2 \boldsymbol{I})$ noise
- $\delta \sim \boldsymbol{p}(\delta)$
- MMSE estimate of $\boldsymbol{h}_t = \widehat{\boldsymbol{W}}(\widehat{\boldsymbol{C}})\boldsymbol{y}_t$

$$\widehat{\boldsymbol{C}} = \frac{\mathbf{1}}{\boldsymbol{\sigma}^2} \sum_{t=1}^{T} \boldsymbol{y}_t {\boldsymbol{y}_t}^{\boldsymbol{H}}$$

$$\widehat{\boldsymbol{W}}(\widehat{\boldsymbol{C}}) = \frac{\int \exp\left(\operatorname{tr}\left(\boldsymbol{W}_\delta \widehat{\boldsymbol{C}}\right) + T \log|\boldsymbol{I} - \boldsymbol{W}_\delta|\right) \boldsymbol{W}_\delta \boldsymbol{p}(\delta) d\delta}{\int \exp\left(\operatorname{tr}\left(\boldsymbol{W}_\delta \widehat{\boldsymbol{C}}\right) + T \log|\boldsymbol{I} - \boldsymbol{W}_\delta|\right) \boldsymbol{p}(\delta) d\delta}$$

$$\boldsymbol{W}_\delta = \boldsymbol{C}_\delta \left(\boldsymbol{C}_\delta + \sigma^2 \boldsymbol{I}\right)^{-1}$$

D. Neumann, T. Wiese and W. Utschick, "Learning the MMSE Channel Estimator," in IEEE Transactions on Signal Processing, vol. 66, no. 11, pp. 2905-2917, 1 June1, 2018.

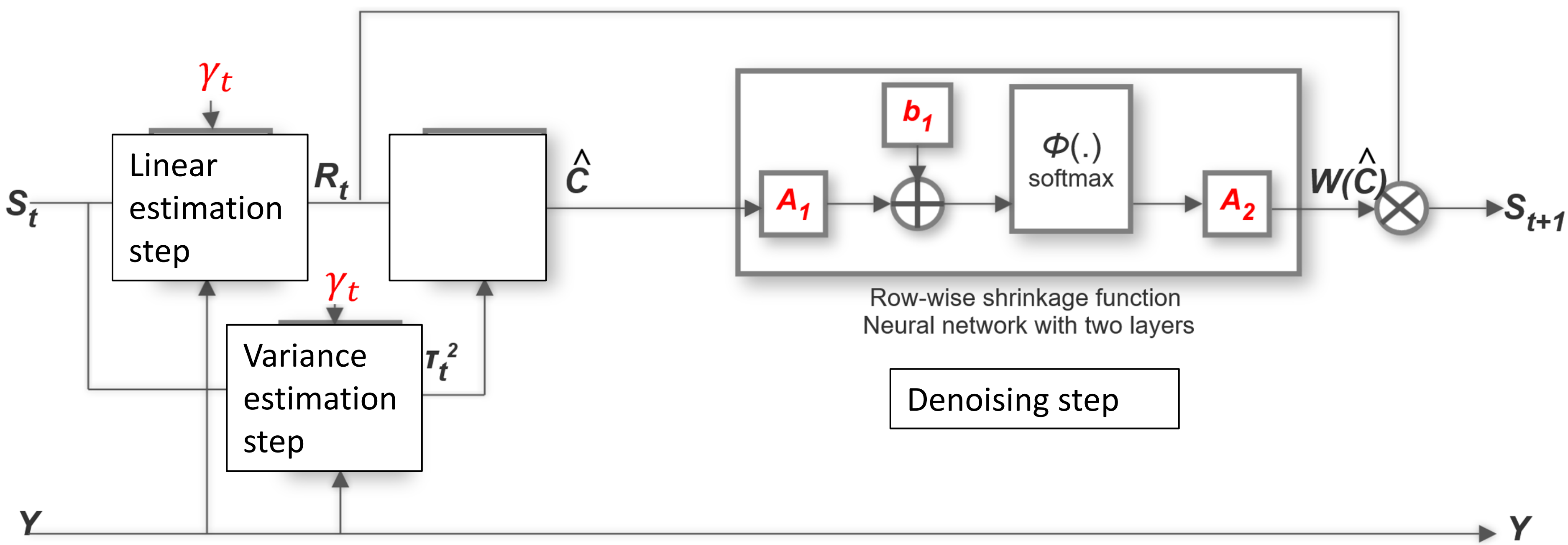
Model-based neural network

$$\widehat{\boldsymbol{W}}(\widehat{\boldsymbol{C}}) = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \exp\left(\operatorname{tr}\left(\boldsymbol{W}_{\boldsymbol{\delta}_i}\widehat{\boldsymbol{C}}\right) + b_i\right)\boldsymbol{W}_{\boldsymbol{\delta}_i}\, p_i}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \exp\left(\operatorname{tr}\left(\boldsymbol{W}_{\boldsymbol{\delta}_i}\widehat{\boldsymbol{C}}\right) + b_i\right) p_i}$$

$$\operatorname{vec}\left(\widehat{\boldsymbol{W}}(\widehat{\boldsymbol{C}})\right) = \boldsymbol{A}\,\frac{\exp\left(\operatorname{tr}\left(\boldsymbol{A}^T \operatorname{vec}(\widehat{\boldsymbol{C}})\right) + \boldsymbol{b}\right)}{\boldsymbol{1}^T \exp\left(\operatorname{tr}(\boldsymbol{A}^T \operatorname{vec}(\widehat{\boldsymbol{C}})) + \boldsymbol{b}\right)}$$

- MMSE estimator of $\boldsymbol{h}_t$: a two-stage neural network with linear layers and soft-max activation function
- Use a trained network for the denoising step
  - Parameters learnt from training data
  - Reduces modelling approximation error

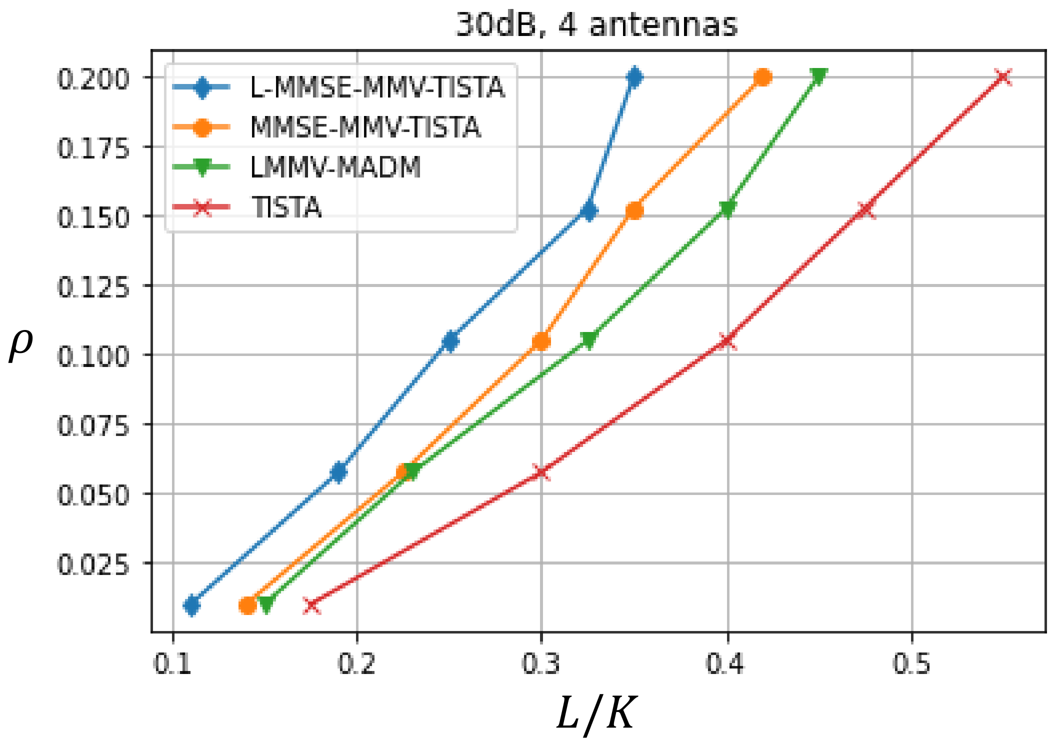
Trainable ISTA (TISTA) and modification



D. Ito, S. Takabe, and T. Wadayama, “Trainable ISTA for sparse signal recovery,” IEEE Transactions on Signal Processing, vol. 67, no. 12, pp. 3113–3125, June 2019.
D. Neumann, T. Wiese and W. Utschick, "Learning the MMSE Channel Estimator," in IEEE Transactions on Signal Processing, vol. 66, no. 11, pp. 2905-2917, 1 June1, 2018.

Simulation Results

Performance

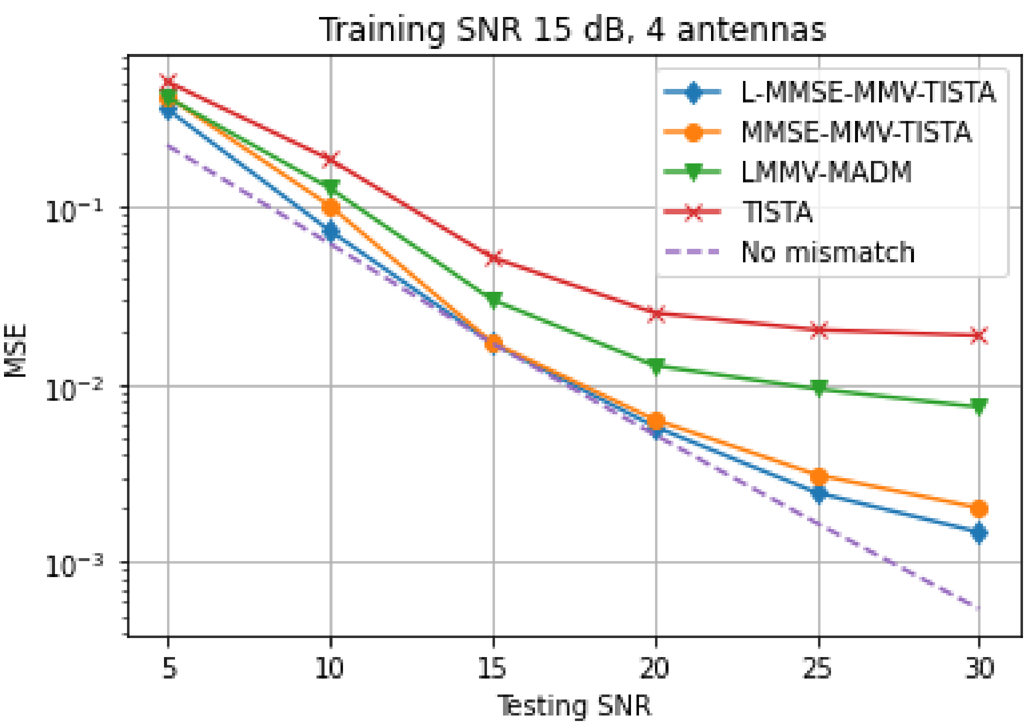


- $K = 500$ users
- 12 layer network
- Minimum L/K for a given activity probability ρ
- Training and test SNR at 30 dB
- Success if NMSE < -20 dB
- Correlated channel

Unfolded network needs fewer iterations
Learnt denoiser gives better performance

U. K. Sreeshma Shiv, S. Bhashyam, C. R. Srivatsa and C. R. Murthy, "Learning-Based Sparse Recovery for Joint Activity Detection and Channel Estimation in Massive Random Access Systems," in IEEE Wireless Communications Letters, vol. 11, no. 11, pp. 2295-2299, Nov. 2022

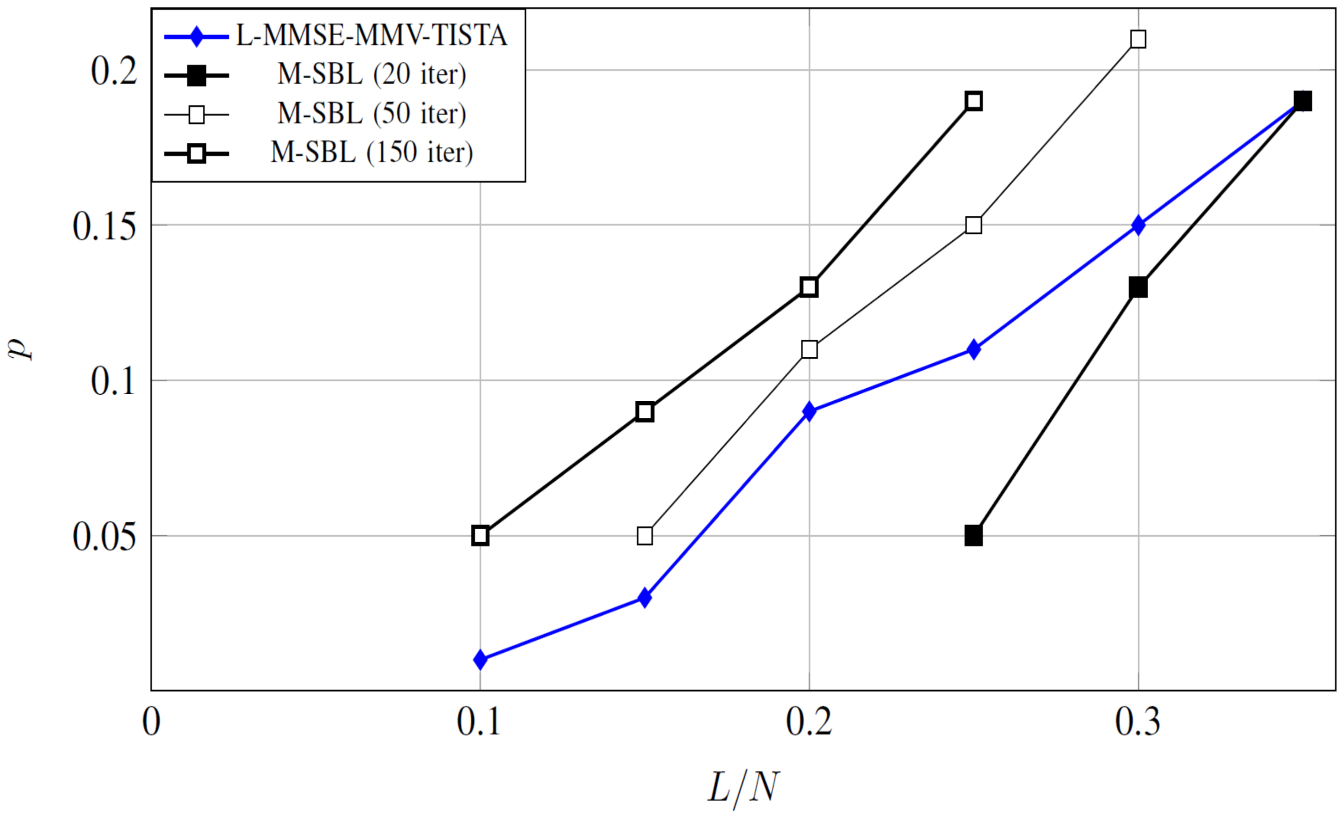
Robustness



- Robustness study
- $K = 500$, 12 layer network
- $\rho = 0.1$
- Training SNR at 15 dB
- L = 200, M = 4
- Correlated channel

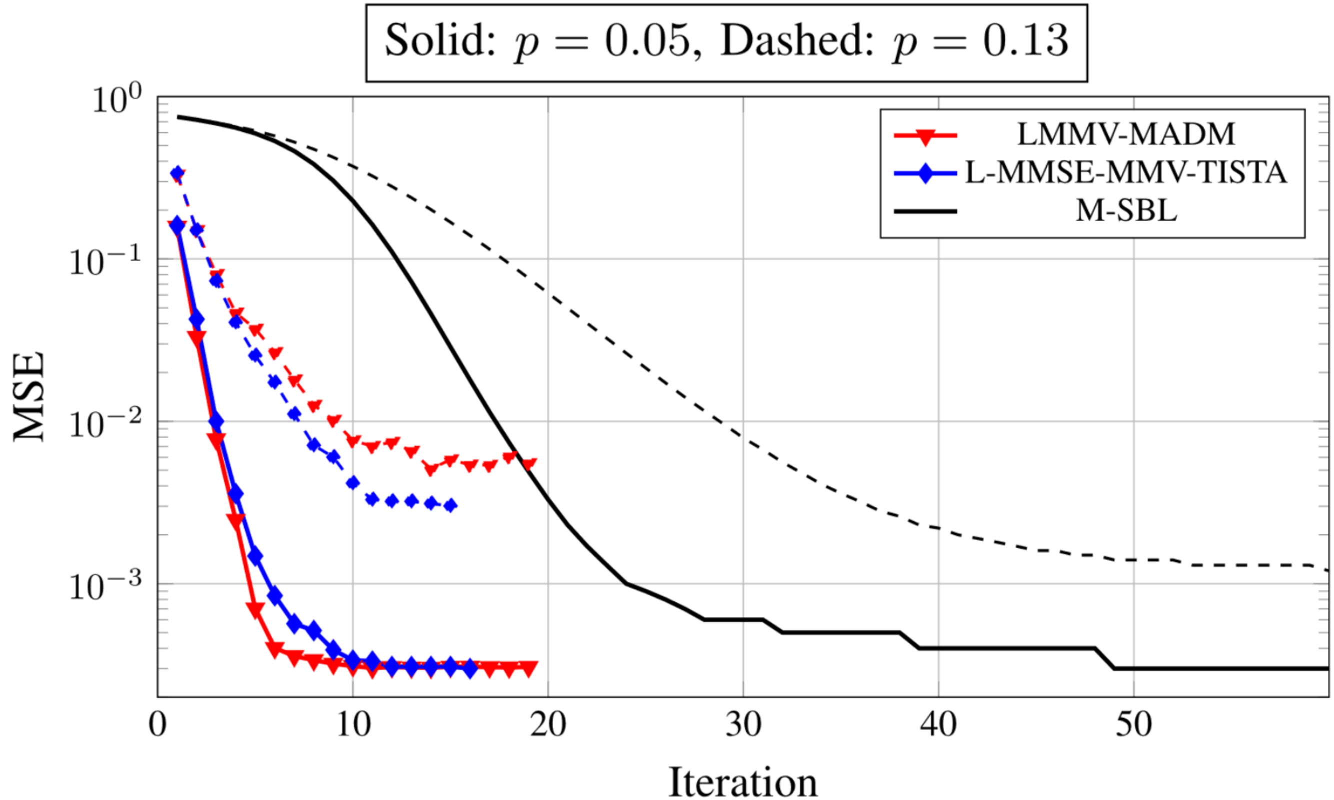
Robust to differences in training and testing SNR

Comparison with M-SBL



- 500 users, 30 dB, 4 antennas, correlated channel
- M-SBL can perform better with higher complexity

Comparison with M-SBL



- 500 users, 30 dB, 10 antennas, L/N = 0.3
- Complexity advantage for smaller p

Summary

- New learning-based sparse recovery methods
  - Back-projected error
  - Deep unfolding
  - Model-based neural network
  - Both supervised and unsupervised training
- Massive random access
  - Reduction in pilot overhead
- Ongoing work
  - Probability of error threshold
  - Large scale fading effects and estimation for MMV-TISTA

Thank you