Joint Sparse Recovery using Deep Unfolding With Application to Massive Random Access

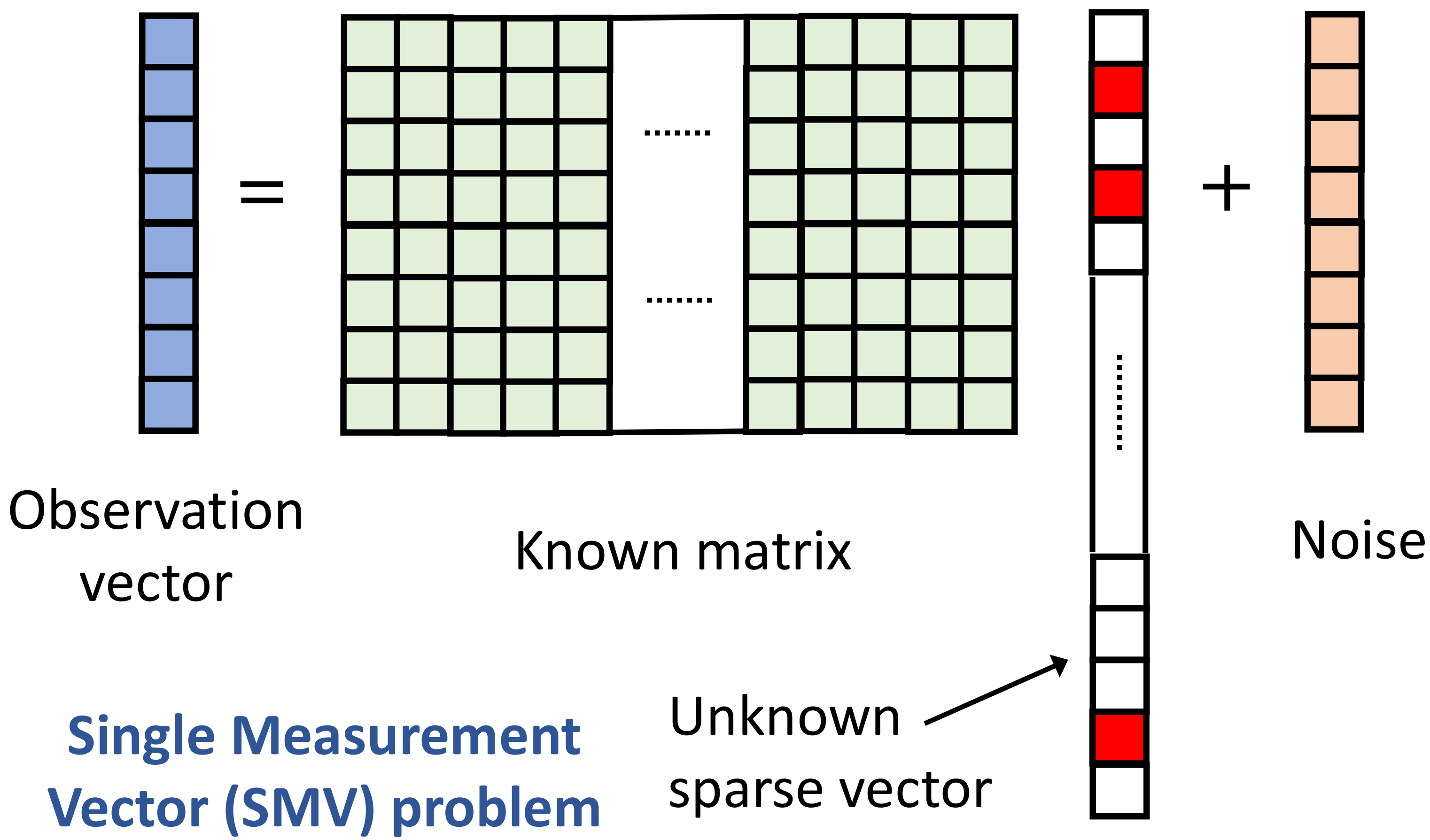
Anand P. Sabulal and Srikrishna Bhashyam

IIT Madras

ICASSP 2020

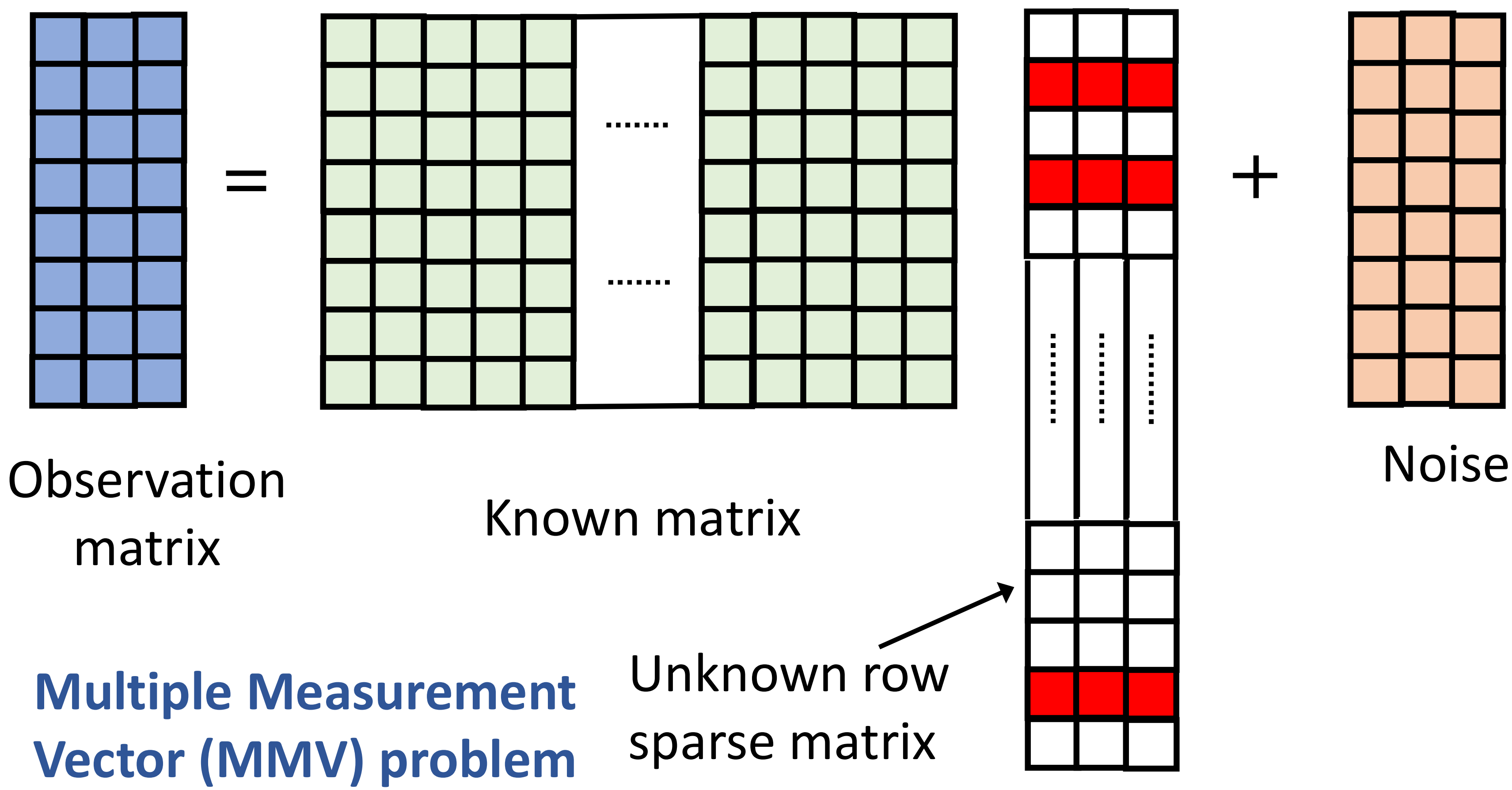
May 4-8, 2020

Introduction: Sparse recovery



Single Measurement Vector (SMV) problem

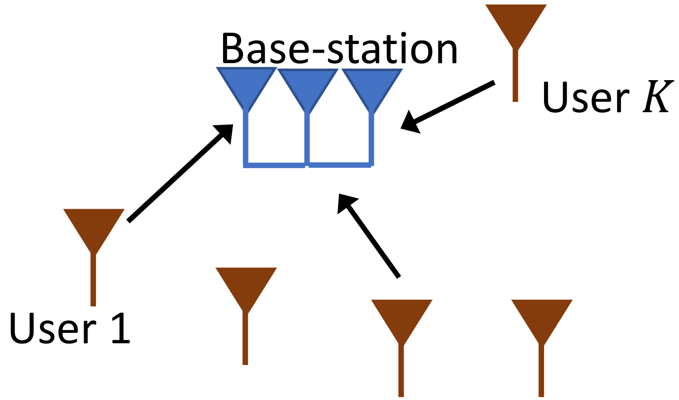
Introduction: Joint sparse recovery

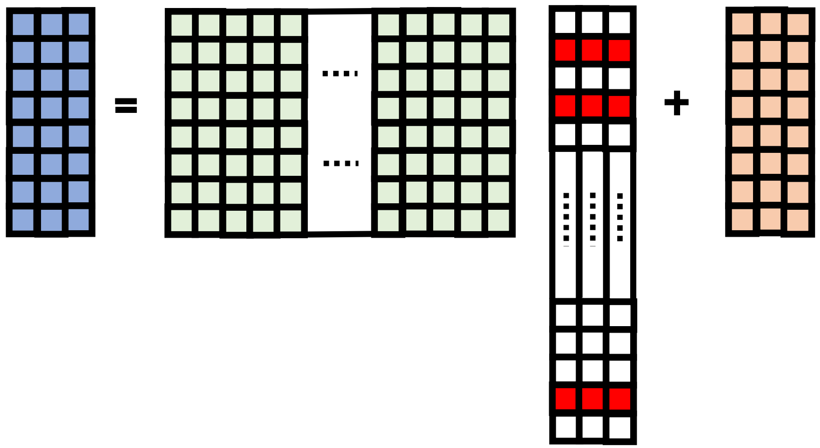


Multiple Measurement Vector (MMV) problem

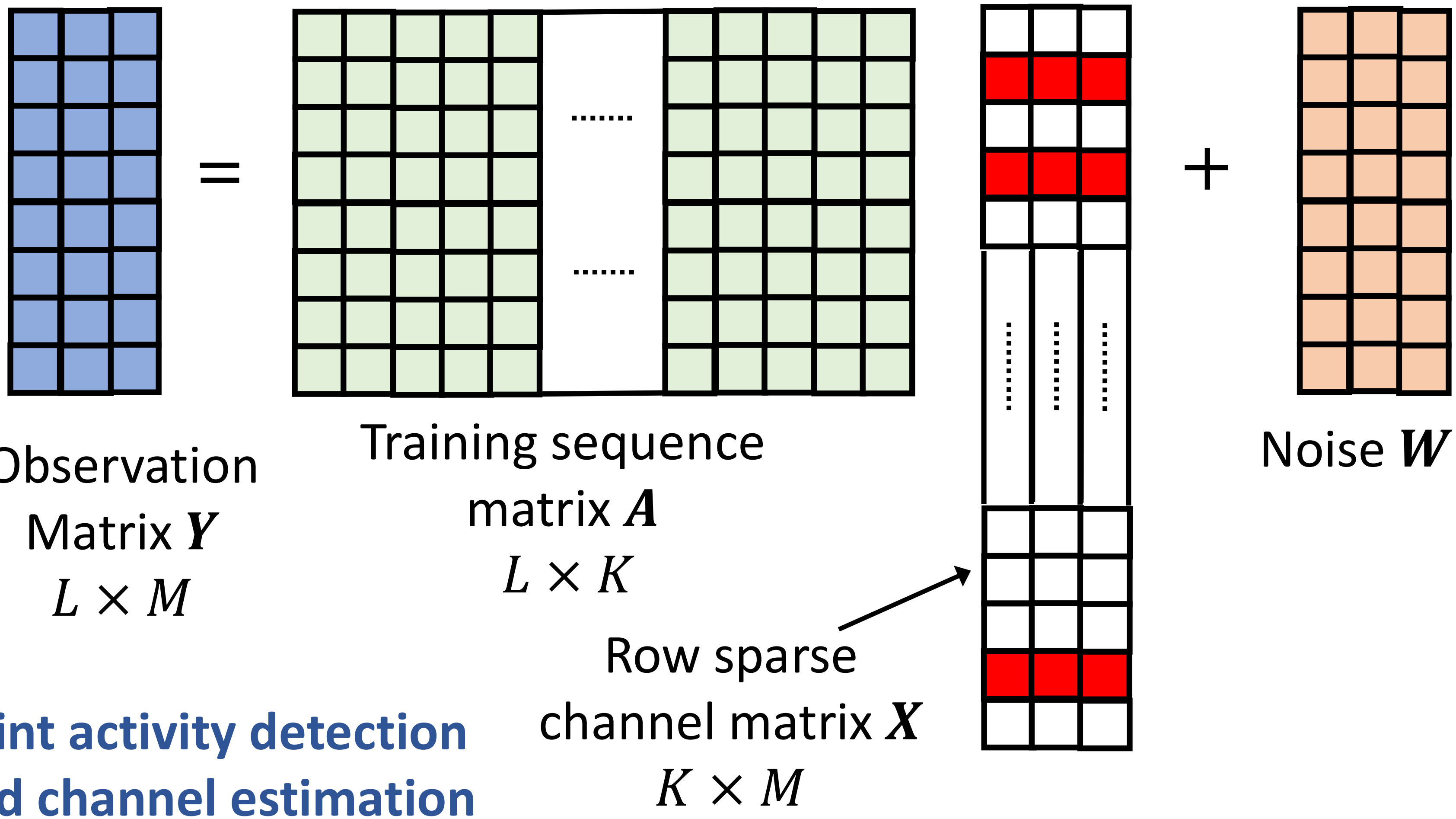
Massive random access

- Large number of users (K) associated with a cell
- Small fraction of users are active at any given time (sparsity)
- Active users send training sequences of length L
- Receiver has multiple antennas (M)



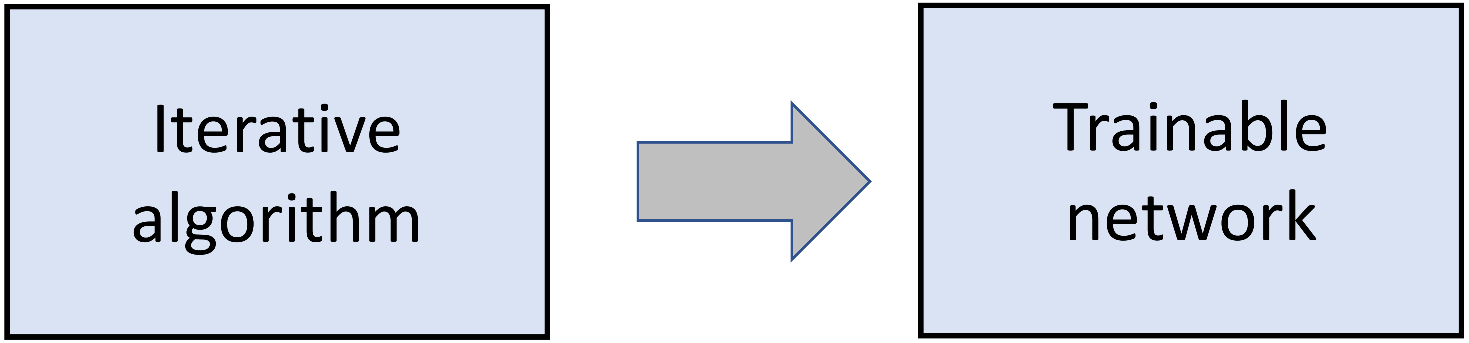


Massive random access: MMV problem



Joint activity detection and channel estimation

Deep unfolding

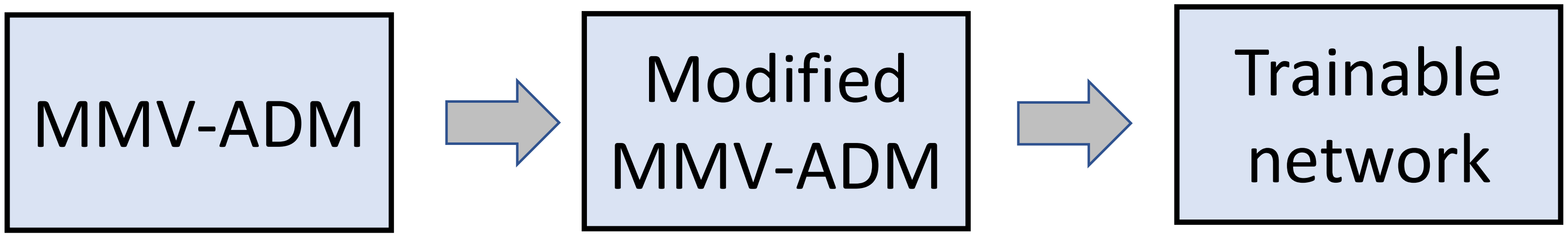


- Each iteration is a layer
- Parameters in each layer untied and trained
- SMV problem: ISTA → TISTA, AMP → L-AMP

M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," IEEE Transactions on Signal Processing, vol. 65, no. 16, pp. 4293–4308, 2017.

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.

This paper: Deep unfolding for joint sparse recovery



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

MMV-ADM

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

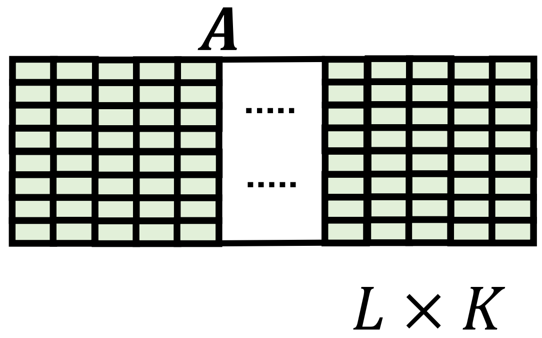
- Alternating direction method
- No matrix inversions $\Rightarrow$ fast, scalable
- Convergence analysis feasible

- Unfolding does not result in a easily trainable network

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

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

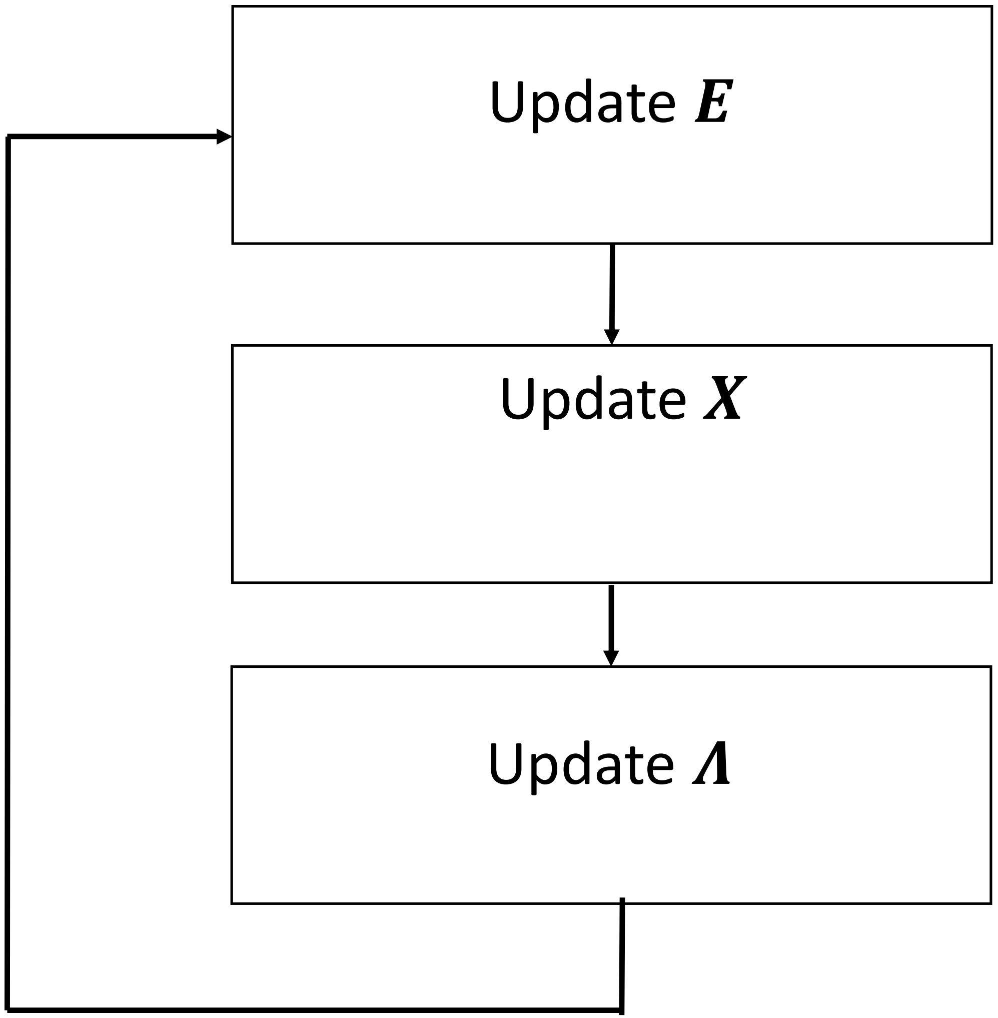


- Backprojected LS error instead of LS error
  - $\boldsymbol{A}^{\dagger}\boldsymbol{Y} - \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{X}$ instead of $\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}$
  - $\boldsymbol{A}^{\dagger} = \boldsymbol{A}^T[\boldsymbol{A}\boldsymbol{A}^T]^{-1}$

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

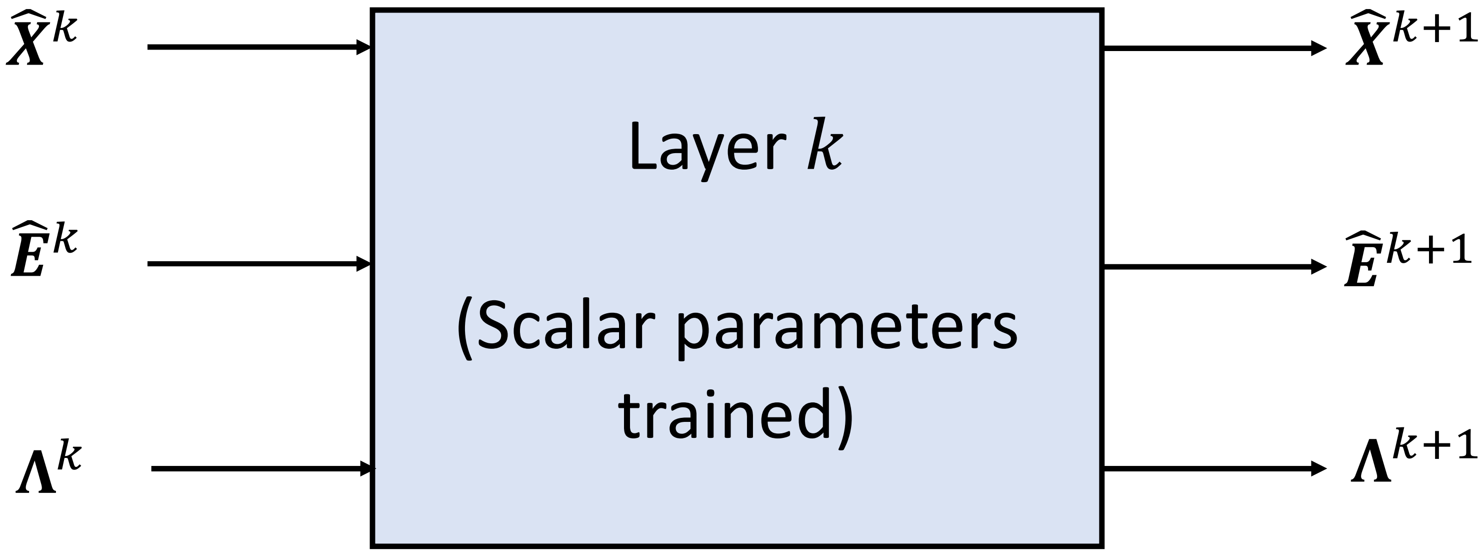
T. Tirer and R. Giryes, “Back-projection based fidelity term for ill-posed linear inverse problems,” CoRR, vol. abs/1906.06794, 2019.

Alternating direction method



- 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 be chosen

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

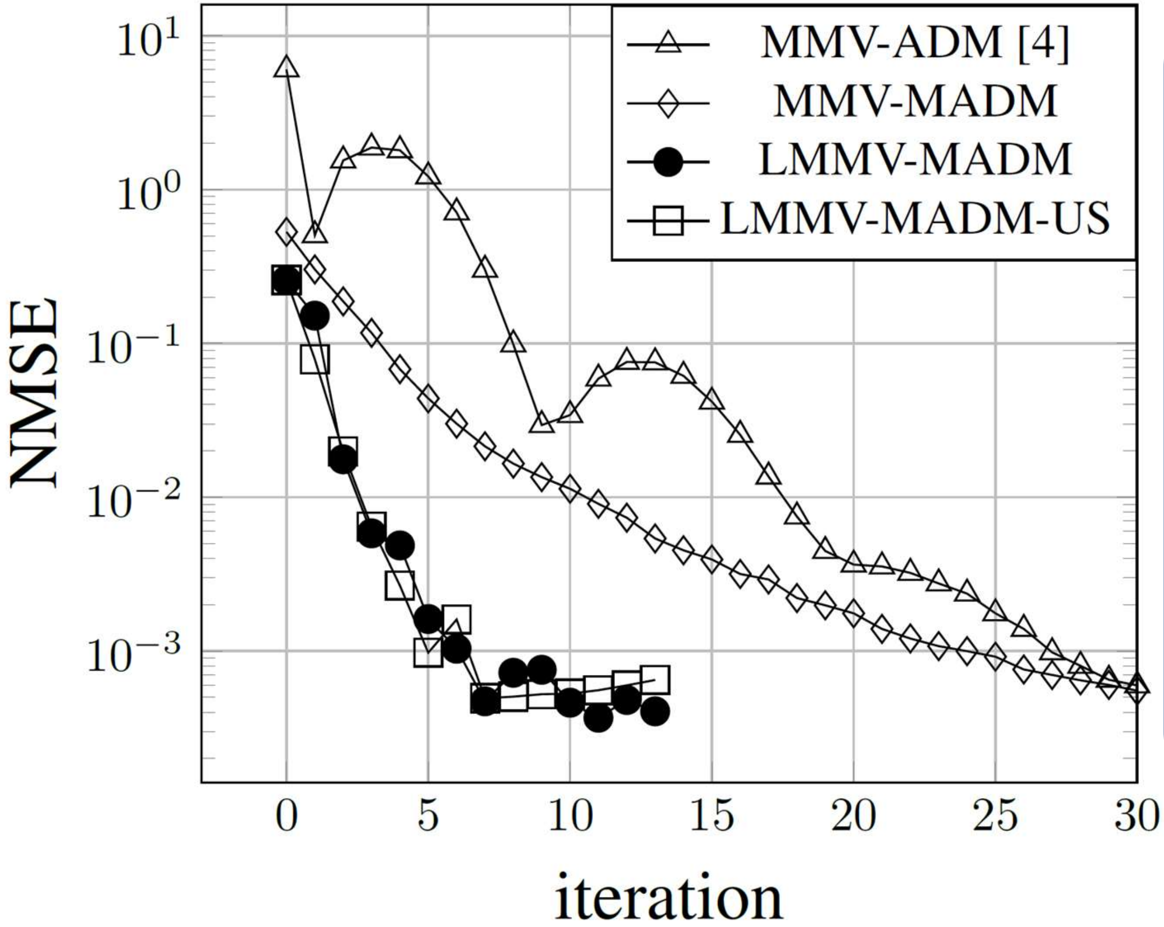
Training the network: Unsupervised

- True $\boldsymbol{X}, \boldsymbol{Y}$ pairs not needed

- Layers trained sequentially
- Loss function for training
  - $\lambda\left\|\widehat{\boldsymbol{X}}^{k+1}\right\|_{2,1/p}^{1/p} + \left\|\boldsymbol{Y} - \boldsymbol{A}\widehat{\boldsymbol{X}}^{k+1}\right\|_F^2$

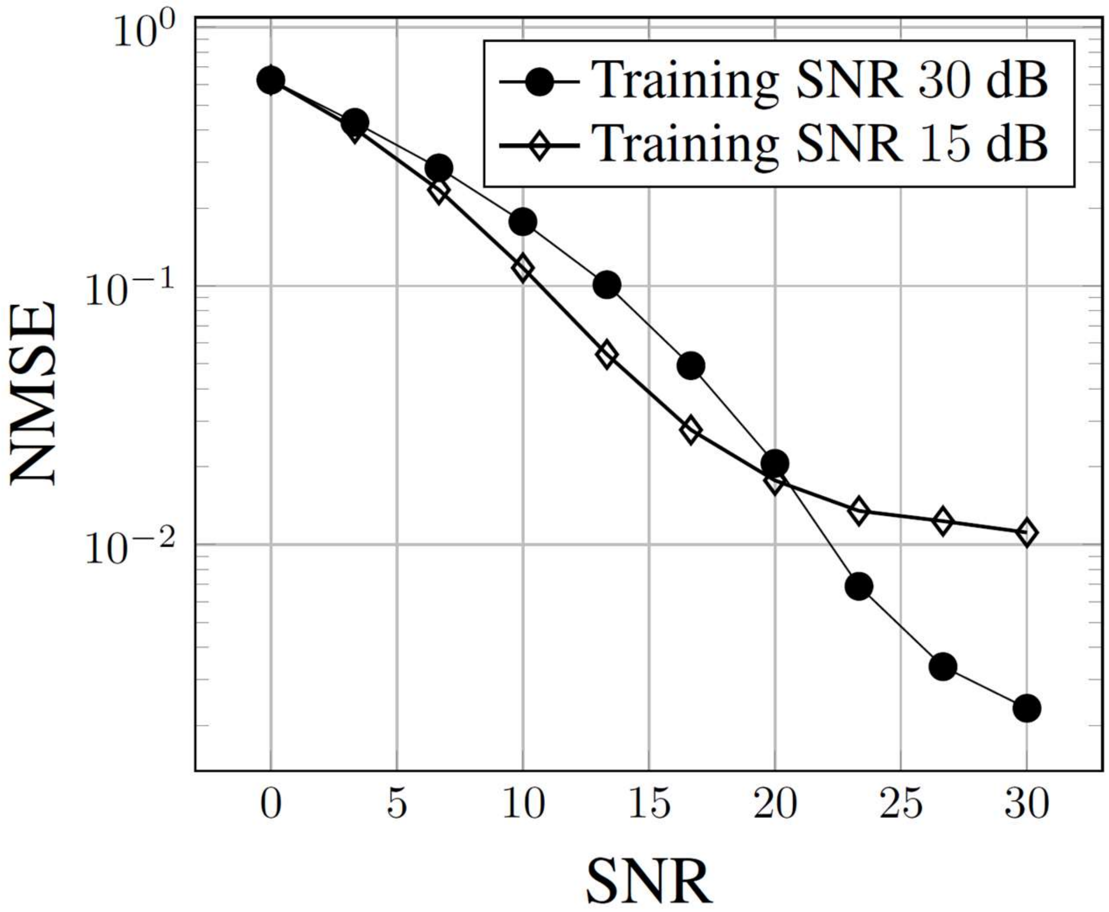
- Unsupervised method can be used after initialization with the supervised method

Performance



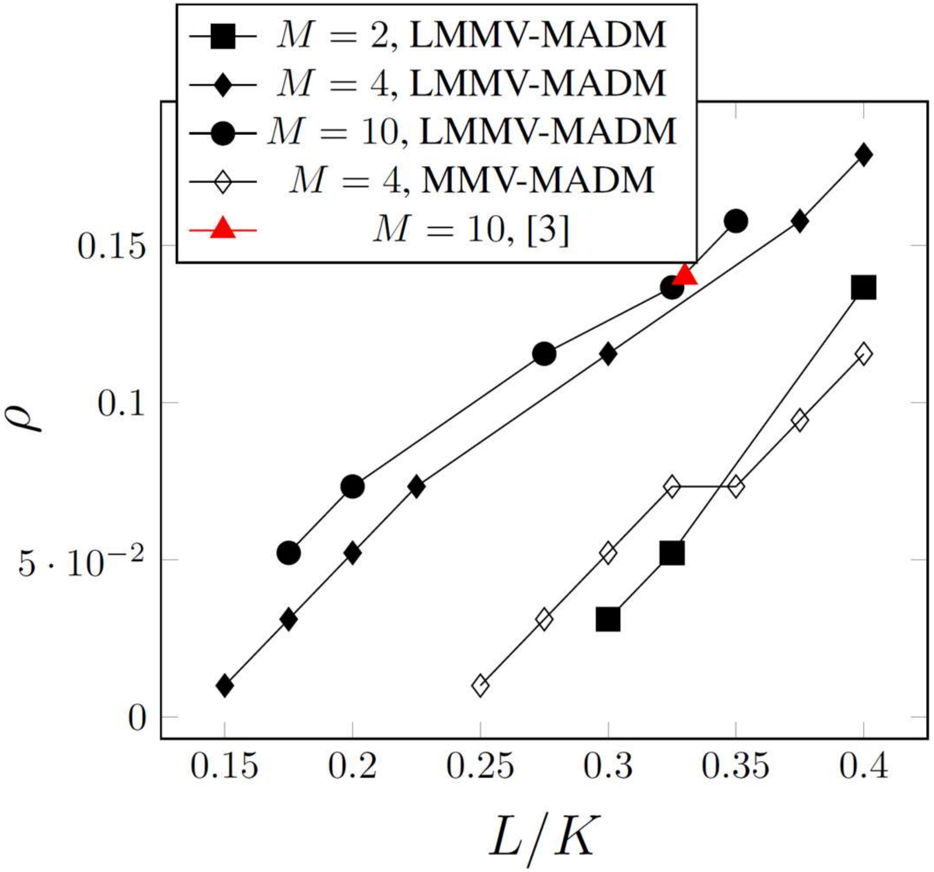
- $K = 2000, L = 500, \rho = 0.05, M = 10$
- Use of backprojected error results in smooth convergence
- Unfolded network converges faster
- Unsupervised training also works well

Training SNR



- $K = 2000, L = 300,$ $\rho = 0.05, M = 10$
- Training at $\rho = 0.07$
- Network trained at higher SNR preferable

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.

Summary

- Modified MMV-ADM algorithm
  - Backprojected error
- Significant improvement with deep unfolding and training
  - Both supervised and unsupervised training
- No matrix inversion steps
- Massive random access
  - Reduction in pilot overhead
  - Reduction in number of iterations

https://www.ee.iitm.ac.in/~skrishna/