A Novel Sparse Bayesian Algorithm for Recovery of Spatially Correlated Sparse Vectors

Dheeraj Prasanna

SPC Lab, Department of ECE
Indian Institute of Science

January 5, 2020

Agenda

Recap

Multiple Measurements

Multiple Signal Realizations

$$\mathbf{y}_i = \mathbf{A}_i \mathbf{x}_i + \mathbf{w}_i, \; i = \{1, 2, ...L\} \quad (1)$$

where,

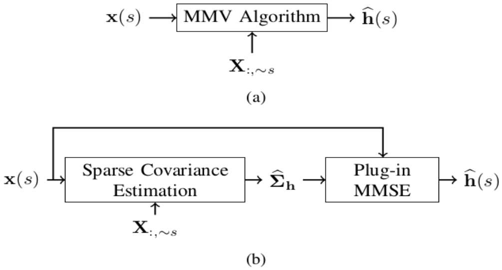
- $\mathbf{A}_i \in \mathbb{C}^{M \times N}, \forall i$ with $M \leq N$
- $\mathbf{w}_i \overset{iid}{\sim} \mathcal{CN}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}) \quad \forall i$
- $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$

$$p(\mathbf{Y}|\mathbf{X}, \sigma^2) = \prod_{l=1}^{L} \mathcal{CN}(\mathbf{y}_l; \mathbf{A}\mathbf{x}_l, \sigma_n^2 \mathbf{I}_M).$$

- Signal design: $\mathbf{x}_i \overset{iid}{\sim} \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$

Summary

- Decoupling property: Covariance Estimation and Signal Recovery
- Advantage in exploiting correlation
- Sample Covariance: Samples of $O(N)$ even with N measurements
- Exploit structure in Covariance matrix (Eg. Masked Sample covariance)



[1]

Figure. (a) Generic MMV algorithm, (b) Decomposition of MMV into Covariance Estimation and plug-in MMSE estimator

[1] Haghighatshoar et al, 2019

Prior design

Prior models

Generative model for $\mathbf{x}$ (Structure of Σ)

- No spatial correlation
  - Diagonal covariance matrix
  - Few non-zero diagonals: Sparse
- Spatial correlation (Intra-vector correlation)
  - Non-diagonal covariance matrix
  - Toeplitz, Compound symmetry, Autoregressive etc
  - $K \times K$ non-zero sub matrix: Sparse
  - Block sparse structure

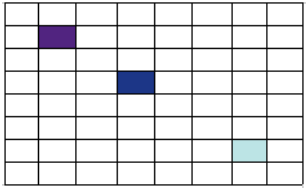


Figure. Diagonal Covariance

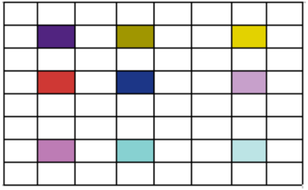


Figure. Covariance of sparse signals with spatial correlation

Correlation matrices

Correlation vs Covariance matrix

- Diagonal entries of correlation matrices are equal to 1
- Non-diagonal entries specify correlation coefficient
- Does not consider the effect of different variances of each index

Models for Correlation matrix **U**

- Compound Symmetry or Uniform correlation model
  - $U_{ij} = \begin{cases} 1 & \text{if } i = j \\ \rho & \text{if } i \neq j \end{cases}$
  - Diversity reception system for uniformly correlated Rayleigh fading channel [2]
- Auto-regressive or Exponential correlation model
  - $U_{ij} = \begin{cases} 1 & \text{if } i = j \\ \rho^{|i-j|} & \text{if } i \neq j \end{cases}$
  - mmWave Massive MIMO channel with correlated Rayleigh fading [3]

[2] Ranjan K Mallick, 2007
[3] Samimi, et.al., 2016

Building covariance matrices

Covariance model

- $p(\mathbf{x}|\boldsymbol{\gamma}, \mathbf{U}) = \mathcal{CN}(\mathbf{x}; \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$
- Hyper-parameter for variance $\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1, & \gamma_2, & \dots & \gamma_n \end{bmatrix}^T$ with $\boldsymbol{\Gamma} = diag(\boldsymbol{\gamma})$
- $\rho_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$
- $(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}})_{ij} = \sqrt{\gamma_i}\sqrt{\gamma_j}U_{ij}$

$$\Longrightarrow \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\Gamma}^{\frac{1}{2}}$$

- λ as hyper-parameter for noise variance σ^2

Bayesian inference

Cost function

- Estimation of hyper-parameters $\theta = \{\boldsymbol{\gamma}, \lambda\}$
- Marginalized pdf: $p(\mathbf{y}|\theta) = \int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}, \lambda) p(\mathbf{x}|\gamma, \mathbf{U}) \, \mu d\mathbf{x}$

$$\Longrightarrow p(\mathbf{y}|\theta) = \frac{1}{(2\pi)^{M/2}} \frac{1}{|\boldsymbol{\Sigma}_{\mathbf{y}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}\right)$$

where $\boldsymbol{\Sigma}_{\mathbf{y}} = \left[\frac{\mathbf{I} - \mathbf{A}\left[\lambda \boldsymbol{\Sigma}_{\gamma}^{-1} + \mathbf{A}^H \mathbf{A}\right]\mathbf{A}^H}{\lambda}\right]^{-1} = \lambda \mathbf{I} + \mathbf{A} \boldsymbol{\Sigma}_{\gamma} \mathbf{A}^H$

- Evidence maximization of θ: Maximizing $p(\mathbf{y}|\theta)$
- Alternatively, minimizing

$$L(\gamma, \mathbf{U}, \lambda) = -log(p(\mathbf{y}|\theta)) = \frac{1}{2}\log(|\boldsymbol{\Sigma}_{\mathbf{y}}|) + \frac{1}{2}\mathbf{y}^H \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}$$

Expectation-Maximization

- Special case of Majorization Minimization (MM) algorithms
- Local maximum likelihood parameters
- Set of incomplete observed data (**y**) assumed to be generated from a specific model

Algorithm

1. Consider set of starting parameters.
2. Expectation-step: Estimate the variables using the observed data
3. Maximization-step: Use complete data ($\mathbf{x}, \mathbf{y}$) obtained in E-step to update parameters or hypothesis
4. Check whether values are converging or not

Analytic expression

$$\theta_{new} = \arg\max_{\theta} \mathbb{E}_{\mathbf{x}|\mathbf{y},\theta_{old}} [\log(p(\mathbf{x}, \mathbf{y}|\theta))]$$

E-Step

Posterior distribution

$$p(\mathbf{x}|\mathbf{y},\theta) = \mathcal{CN}(\mathbf{x};\boldsymbol{\mu}_\theta,\boldsymbol{\Sigma}_\theta)$$

where, $\boldsymbol{\Sigma}_\theta = \left(\frac{1}{\lambda}\mathbf{A}^H\mathbf{A} + \boldsymbol{\Sigma}_\gamma^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_\theta = \frac{1}{\lambda}\boldsymbol{\Sigma}_\theta\mathbf{A}^H\mathbf{y}$.

Expectation step

$$Q(\theta_{old},\theta) = \mathbb{E}_{\mathbf{x}|\mathbf{y},\theta_{old}}\left[\log(p(\mathbf{x},\mathbf{y}|\theta))\right] = \mathbb{E}_{\mathbf{x}|\mathbf{y},\theta_{old}}\left[\log(p(\mathbf{x}|\gamma,\mathbf{U}))\right] + \mathbb{E}_{\mathbf{x}|\mathbf{y},\theta_{old}}\left[\log(p(\mathbf{y}|\mathbf{x},\lambda))\right]$$

$$= Q_1(\theta_{old},\gamma,\mathbf{U}) + Q_2(\theta_{old},\lambda)$$

where,

$$Q_1(\theta_{old},\gamma,\mathbf{U}) = c - \frac{1}{2}log(|\boldsymbol{\Sigma}_\gamma|) - \frac{1}{2}Tr\left[\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_{old}\right],$$

$$Q_2(\theta_{old},\lambda) = c - \frac{M}{2}\log(\lambda) - \frac{\|\mathbf{y}\|^2}{2\lambda} + \frac{\boldsymbol{\mu}_\theta^H\mathbf{A}^H\mathbf{y}}{\lambda} - \frac{1}{2\lambda}Tr\left[\mathbf{A}^H\mathbf{A}\boldsymbol{\Sigma}_{old}\right],$$

where $\boldsymbol{\Sigma}_{old} = \boldsymbol{\Sigma}_{\theta_{old}} + \boldsymbol{\mu}_{\theta_{old}}\boldsymbol{\mu}_{\theta_{old}}^H$.

Optimality condition

Gamma update

$$\boldsymbol{\gamma}^* = \arg\max_{\boldsymbol{\gamma}} Q_1(\theta_{old}, \boldsymbol{\gamma})$$

Optimality condition

$$\begin{aligned}
0 =& -\frac{1}{2}\frac{\partial\left(\log|\boldsymbol{\Gamma}| + \log|\mathbf{U}| + Tr\left[\left(\boldsymbol{\Gamma}^{-\frac{1}{2}}\mathbf{P}\boldsymbol{\Gamma}^{-\frac{1}{2}}\right)\boldsymbol{\Sigma}_{old}\right]\right)}{\partial\gamma_i} \\
=& \frac{1}{\gamma_i} - Tr\left[\left(\frac{1}{2\gamma_i^{\frac{3}{2}}}\mathbf{J}_{ii}\mathbf{P}\boldsymbol{\Gamma}^{-\frac{1}{2}} + \frac{1}{2\gamma_i^{\frac{3}{2}}}\boldsymbol{\Gamma}^{-\frac{1}{2}}\mathbf{P}\mathbf{J}_{ii}\right)\boldsymbol{\Sigma}_{old}\right] \text{ where } \mathbf{J}_{ii} \text{ is single entry matrix} \\
=& 2\sqrt{\gamma_i} - 2\,Tr\left[\mathbf{J}_{ii}\mathbf{P}\boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{old}\right] \\
=& 2\sqrt{\gamma_i} - 2\left(\mathbf{P}\boldsymbol{\Gamma}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{old}\right)_{ii}.
\end{aligned}$$

$$\Rightarrow \sqrt{\gamma_i} = \sum_{k=1}^{N}\frac{1}{\sqrt{\gamma_k}}P_{ik}(\Sigma_{old})_{ki}\ \forall i \in [N].$$

Fixed point iterations

- Optimality condition has variables dependent on each other
- Let $c_i = 1/\sqrt{\gamma_i}$

$$\frac{1}{\mathbf{c}} = \mathbf{K}\mathbf{c},$$

$$\Rightarrow \mathbf{c} = \mathbf{K}^{-1}\frac{1}{\mathbf{c}}$$

where $\frac{1}{\mathbf{c}}$ is element wise inverse of $\mathbf{c}$ and $\mathbf{K} = \mathbf{P} \odot \boldsymbol{\Sigma}_{old}$.

- One iteration leads to reduction in cost function for each EM iteration
- $\mathbf{c}_{new} = \mathbf{K}^{-1}\frac{1}{\mathbf{c}_{old}}$ and update $\gamma_i = \left(\frac{1}{c_i}\right)^2, \qquad \forall i \in [N].$

Noise variance update

$$\lambda^* = \arg\max_{\lambda} Q_2(\theta_{old}, \lambda).$$

Optimality condition

$$0 = \frac{\partial Q_2}{\partial \lambda}$$

$$= \frac{\partial \left(c - \frac{M}{2}\log(\lambda) - \frac{\mathbf{y}^H\mathbf{y}}{2\lambda} + \frac{\boldsymbol{\mu}_{old}^H \mathbf{A}^H \mathbf{y}}{\lambda} - \frac{Tr\left(\mathbf{A}^H\mathbf{A}\boldsymbol{\Sigma}_{old}\right)}{2\lambda} \right)}{\partial \lambda}$$

$$= \frac{-1}{2\lambda^*} M + \frac{1}{2{\lambda^*}^2} \left[\mathbf{y}^H\mathbf{y} - 2\boldsymbol{\mu}_{old}^H \mathbf{A}^H\mathbf{y} + Tr\left(\mathbf{A}^H\mathbf{A}\boldsymbol{\Sigma}_{old}\right)\right]$$

$$\Rightarrow \lambda^* = \frac{1}{M}\left[\mathbf{y}^H\mathbf{y} - 2\mu_{old}^H\mathbf{A}^H\mathbf{y} + Tr\left[\mathbf{A}^H\mathbf{A}\boldsymbol{\Sigma}_{old}\right]\right]$$

Algorithm

Algorithm 1: Correlated EM

Input: $\mathbf{A}, \mathbf{Y}, \mathbf{U}$

Initialize: $k \leftarrow 0, \boldsymbol{\gamma} \leftarrow \mathbf{1}, \lambda \leftarrow 0.1$

while $k \leq k_{max}$ *and* $\boldsymbol{\mu}_\theta \neq \boldsymbol{\mu}_{\theta_{old}}$ **do**

$$\boldsymbol{\Sigma}_{\theta_{old}} = \left(\frac{1}{\lambda}\mathbf{A}^H\mathbf{A} + \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\right)^{-1}$$

$$\boldsymbol{\mu}_{\theta_{old}} = \frac{1}{\lambda}\boldsymbol{\Sigma}_{\theta_{old}}\mathbf{A}^H\mathbf{Y}$$

$$\boldsymbol{\Sigma}_{old} = \boldsymbol{\Sigma}_{\theta_{old}} + \boldsymbol{\mu}_{\theta_{old}}\boldsymbol{\mu}_{\theta_{old}}^H$$

$$c_i = \sqrt{\frac{1}{\gamma_i}}\ \forall i$$

$$\mathbf{c} \leftarrow \left(\mathbf{U}^{-1} \odot \boldsymbol{\Sigma}_{old}\right)^{-1}\frac{1}{\mathbf{c}}$$

$$\gamma_i = \left(\frac{1}{c_i}\right)^2$$

$$\lambda^* = \frac{1}{M}\left[\frac{1}{L}\|\mathbf{Y}\|_F^2 - \frac{2}{L}\boldsymbol{\mu}_{old}^H\mathbf{A}^H\mathbf{Y} + Tr\left[\mathbf{A}^H\mathbf{A}\boldsymbol{\Sigma}_{old}\right]\right]$$

$$k \leftarrow k + 1$$

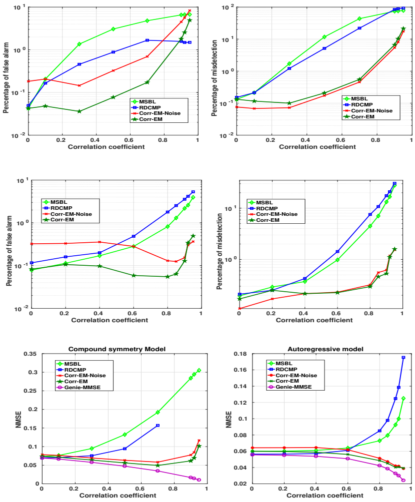
end

Output: $\mathbf{X} = \boldsymbol{\mu}_\theta, \boldsymbol{\gamma}$, Support obtained by non-zero entries of $\boldsymbol{\gamma}$

Results

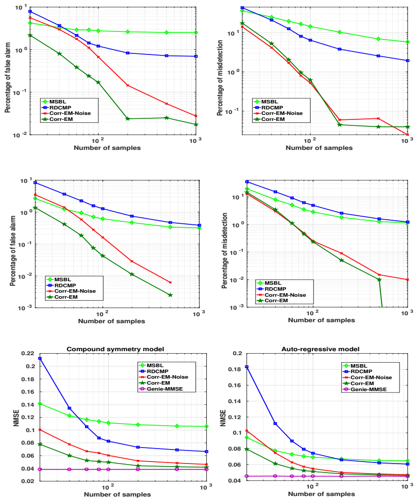
Effect of correlation

- $N = 100$
- $K = 20$
- $M = 20$
- $L = 100$
- Row1 : Support recovery for uniform model
- Row2 : Support recovery for exponential model
- Row3 : NMSE recovery for both models



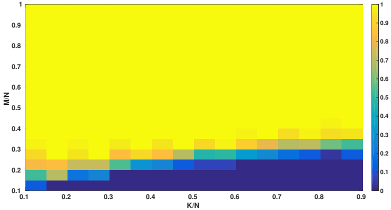
Effect of number of samples

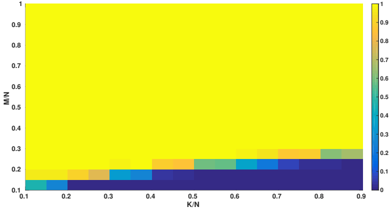
- $N = 100$
- $K = 20$
- $M = 20$
- Row1 : Support recovery for uniform model $\rho = 0.5$
- Row2 : Support recovery for exponential model $\rho = 0.75$
- Row3 : NMSE recovery for both models



Phase transition

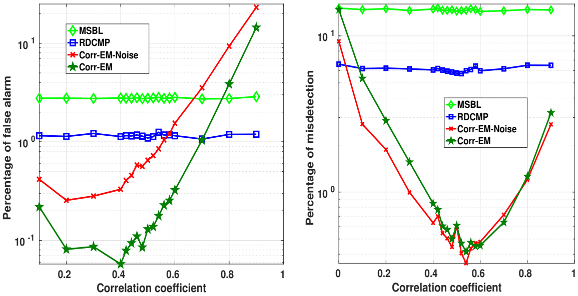
- $N = 100$
- $L = 100$
- Figure 1: Uniform model $\rho = 0.5$
- Figure 2 : Exponential model $\rho = 0.75$



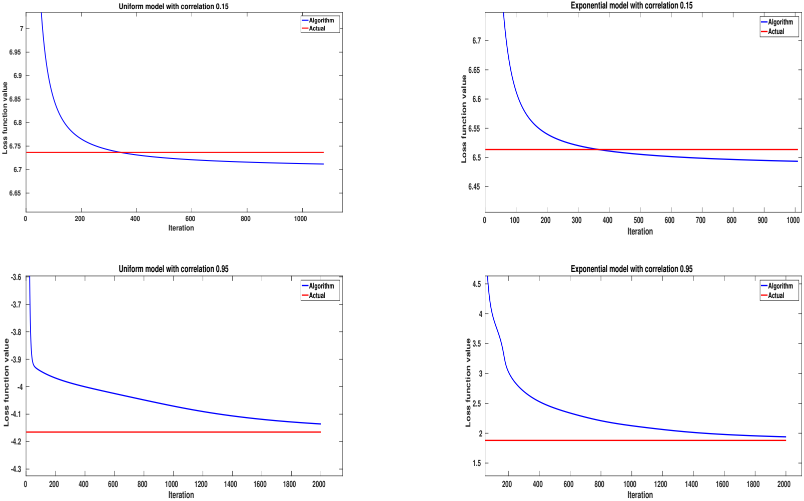


Error in correlation coefficient

- $N = 100$
- $K = 20$
- $M = 20$
- $L = 100$
- Uniform model
- Actual $\rho = 0.5$



Does the EM converge?



Challenges

Theory behind the algorithm?

Open Questions

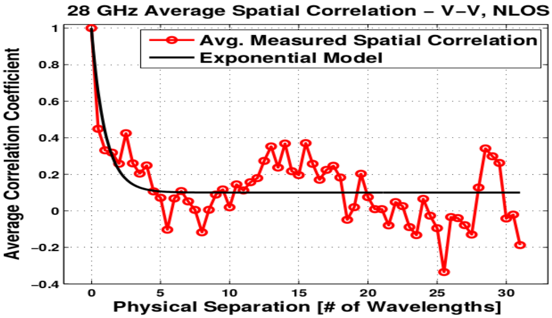
- Why only single iteration of fixed point is useful?
- Does the EM converge to a stationary point? What are the properties? How do you show convergence?
- What happens when exact correlation coefficient is not known?
- Can we learn the correlation coefficients from the available data?
- How useful is the algorithm for practical scenarios?
- What is the computational complexity? Can we make it faster?

Learning correlation coefficients

- Works only for applications with known **U**
- Parametric form for **U**
- Can we learn the parameters from sample covariance of $\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{y}$?
- Simplify the optimality conditions for **U** or its parameters as hyper-parameters

Where can this be applied?

- Correlated Rayleigh fading in Massive MIMO channels
- Can models with correlation counter fundamental limit created due to pilot contamination? [4]



[5]

[4] Sanguinetti, et.al., 2019
[5] Samimi, et.al., 2016

References

- M. Azizyan, A. Krishnamurthy, and A. Singh, “Extreme compressive sampling for covariance estimation,” *IEEE Transactions on Information Theory*, vol. 64, no. 12, pp. 7613–7635, Dec 2018.
- R. Y. Chen, A. Gittens, and J. A. Tropp, “The masked sample covariance estimator: an analysis using matrix concentration inequalities,” *Information and Inference: A Journal of the IMA*, vol. 1, no. 1, pp. 2–20, 2012.
- S. Haghighatshoar and G. Caire, “Multiple measurement vectors problem: A decoupling property and its applications,” 2018.
- L. Sanguinetti, E. Björnson, and J. Hoydis, “Towards massive mimo 2.0: Understanding spatial correlation, interference suppression, and pilot contamination,” *arXiv preprint arXiv:1904.03406*, 2019.
- R. K. Mallik, “The uniform correlation matrix and its application to diversity,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1619–1625, May 2007.
- M. K. Samimi, S. Sun, and T. S. Rappaport, “Mimo channel modeling and capacity analysis for 5g millimeter-wave wireless systems,” in *2016 10th European Conference on Antennas and Propagation (EuCAP)*, April 2016, pp. 1–5.
- L. Sanguinetti, E. Björnson, and J. Hoydis, “Towards massive mimo 2.0: Understanding spatial correlation, interference suppression, and pilot contamination,” 2019.