Multiple Signal Realizations: Decoupling Covariance and Signal Estimation

Dheeraj Prasanna

SPC Lab, Department of ECE
Indian Institute of Science

October 12, 2019

Table of Contents

Multiple Measurements

Multiple Signal Realizations

$$\mathbf{y}_i = \mathbf{A}_i\mathbf{x}_i + \mathbf{w}_i,\ \ i = \{1, 2, ...L\} \tag{1}$$

where,

- $\mathbf{A}_i \in \mathbb{R}^{M\times N}, \forall i$ with $M \leq N$
- $\mathbf{w}_i \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}) \quad \forall i$
- $\mathbf{x}_i \overset{iid}{\sim} \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$

Examples

- $\mathbf{y}_i = \mathbf{x}_i + \mathbf{w}_i$
- Joint sparsity (MMV) setup
- Massive MIMO Uplink

$$\mathbf{Y}[n] = \sum_{l=1}^{L}\sum_{k=1}^{K}\sqrt{\mu}\mathbf{h}_{lk}\mathbf{x}_{lk}^{\top}[n] + \mathbf{N}[n]$$

Spatial Correlation

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

- 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

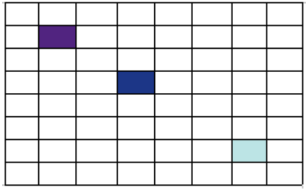


Figure. Diagonal Covariance

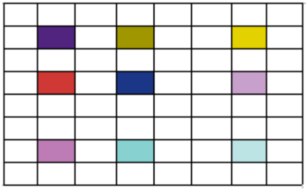


Figure. Covariance of sparse signals with spatial correlation

Spatial Correlation - II

Why spatial correlation?

1. Exploiting correlation can lead to better recovery performance
2. Existing algorithms could perform poorly when there is high intra-vector correlation

Example: Setup and Metrics

$\mathcal{S}$: Support set (Actual) of $\mathbf{X}$ $\mathcal{S}_{out}$: Support set (Recovered) of $\mathbf{X}_{out}$

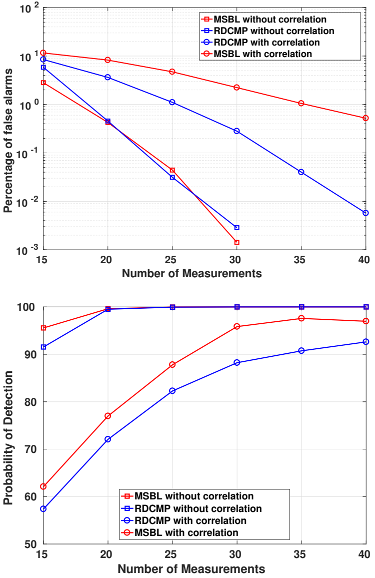
- False Alarm percentage= $\frac{|\mathcal{S}_{out}|-|\mathcal{S}\cap\mathcal{S}_{out}|}{N-|\mathcal{S}|} \times 100$
- Probability of Detection (%)= $\frac{|\mathcal{S}\cap\mathcal{S}_{out}|}{K} \times 100$
- NMSE = $\frac{||\mathbf{X}-\mathbf{X}_{out}||^2}{||\mathbf{X}||^2}$
- Algorithms: MSBL, RDCMP

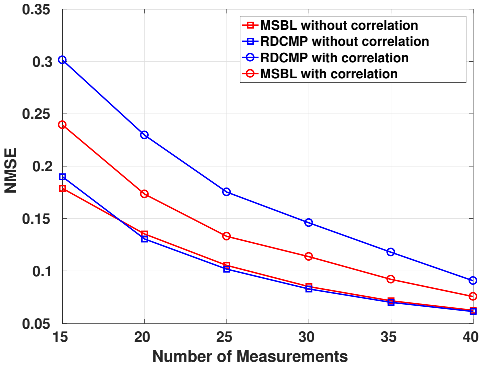
Setup:

- N = 100
- L = 200
- K = 30
- $\sigma^2 = 0.1$

Observations:

- False alarm increases for correlated signals when compared to uncorrelated signals
- Detection probability also decreases
- Support recovery methods perform better in false alarms





Observations:

- RMSE increases when correlation is present, MSBL performance is better than RDCMP
- MSBL and RDCMP perform poorly in $K > M$ regime for signals with correlation
- Performance decrease present even for M slightly greater than K

Table of Contents

Decoupling: $l_{2,1}$ Regularized LS

- Multiple measurement vector setup: $\mathbf{x}(s) = \mathbf{\Psi}(s)\mathbf{h}(s) + \mathbf{z}(s), \forall s \in [T]$
- MMV variant of l_1-LS

$$\widehat{\mathbf{C}} = \underset{\mathbf{C}\in\mathbb{C}^{G\times T}}{\arg\min} \left[\frac{1}{2} \sum_{s\in[T]} \|\mathbf{x}(s) - \mathbf{\Psi}(s)\mathbf{A}\mathbf{c}(s)\|_2^2 + \varrho\sqrt{T}\|\mathbf{C}\|_{2,1} \right]$$

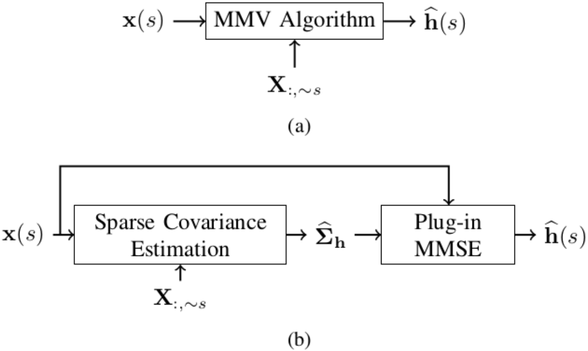
- Convex cost function

$$\widehat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}\in\mathbb{R}_+^G}{\arg\min} \left[g(\boldsymbol{\gamma}) = \frac{1}{T} \sum_{s\in[T]} \mathbf{x}(s)^\dagger \left(\mathbf{\Psi}(s)\mathbf{A}\mathbf{\Gamma}\mathbf{A}^\dagger\mathbf{\Psi}(s)^\dagger + \varrho\mathbf{I}_m \right)^{-1} \mathbf{x}(s) + \mathrm{tr}(\mathbf{\Gamma}) \right]$$

- Theorem 1: $\widehat{\gamma}_i = \frac{\|\hat{\mathbf{C}}_{i,:}\|_2}{\sqrt{T}}$ for $i \in [G]$
- Theorem 2: If $\widehat{\mathbf{h}}(s) = \mathbf{A}\widehat{\mathbf{c}}(s)$, where $\widehat{\mathbf{c}}(s) = \widehat{\mathbf{C}}_{:,s}$, then it is also given by

$$\widehat{\mathbf{h}}(s) = \mathbf{A}\widehat{\mathbf{\Gamma}}\mathbf{A}^\dagger\mathbf{\Psi}(s)^\dagger \left(\mathbf{\Psi}(s)\mathbf{A}\widehat{\mathbf{\Gamma}}\mathbf{A}^\dagger\mathbf{\Psi}(s)^\dagger + \varrho\mathbf{I}_m \right)^{-1} \mathbf{x}(s)$$

Estimation Model



[1]

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

[1] Haghighatshoar et al, 2019

Table of Contents

Sample Complexity Bounds

Setup: Observations $\left\{ \left(A_t, A_t^T x_t \right) \right\}_{t=1}^n$, where $A_t \in \mathbb{R}^{d \times m}$ are orthonormal basis for m-dim subspace drawn uniformly at random. Consider $\Phi_t = A_t A_t^T$

- Observed covariance

$$\hat{\Sigma}_1 \triangleq \frac{d^2}{nm^2} \sum_{t=1}^{n} \left(\Phi_t x_t \right) \left(\Phi_t x_t \right)^T$$

- Unbiased estimate for covariance considered

$$\hat{\Sigma} \triangleq \frac{m \left((d+2)(d-1)\hat{\Sigma}_1 - (d-m)\,\mathrm{tr}\left(\hat{\Sigma}_1 \right) I_d \right)}{d(dm+d-2)}$$

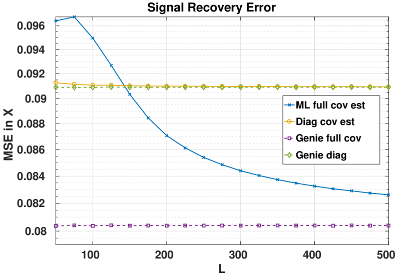
- **Corollary 1 (Gaussian Upper Bounds)**: Let $x_1, \ldots, x_n \sim \mathcal{N}(0, \Sigma)$ and construct $\hat{\Sigma}$. Then for any $\delta \in (0, 1)$, there exist universal constants $\kappa_1, \kappa_2 > 0$ such that, with probability at least $1 - \delta$, the

$$|\hat{\Sigma} - \Sigma\|_2 \leq \kappa_2 \|\Sigma\|_2 \left(\sqrt{\frac{d^3 \log^2(nd/\delta)}{nm^2}} + \frac{d^3 \log^2(nd/\delta)}{nm^2} + \sqrt{\frac{d \log(1/\delta)}{n}} \right)$$

The bound holds when $d \geq 2$ and $n \geq d \log(1/\delta)$.

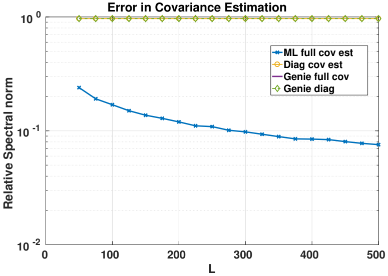
Setup:

- $\mathbf{y}_i = \mathbf{x}_i + \mathbf{w}_i \in \mathbb{R}^N$ $\forall i = 1, 2, \ldots, L$
- N = 50
- $\sigma^2 = 0.1$
- Compound symmetric covariance matrix



Observations:

- Error in Covariance estimation is lower when covariance is exploited for all L
- Error in signal recovery does not match Error in Covariance



Discussion - I

Summary

- Decoupling property: Covariance Estimation and Signal Recovery
- Advantage in exploiting correlation
- Sample Covariance: Samples of $O(N)$ even with N measurements

Robustness vs Over-fitting trade-off characterization

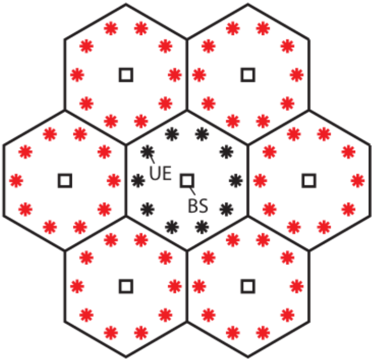
- Adaptive Setup
- Metric for covariance estimation output
- $\tilde{\Sigma} = f(\widehat{\Sigma})$ as estimate for covariance.
  - Chen, et al. proposed Masked Sample Covariance
  - Exploit structure in Covariance matrix

Table of Contents

Massive MIMO System

- L cells
- Base Station with M antenna
- K users with single antenna
- Uplink channel
  - Block flat-fading channel
  - Coherence time T_c
  - Coherence Bandwidth B_c

$$\mathbf{Y}[n] = \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{\mu}\mathbf{h}_{lk}\mathbf{x}_{lk}^{\top}[n] + \mathbf{N}[n]$$

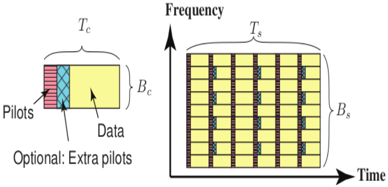


[2]

[2] Bjornson, et al, 2016

Variations in Channel Statistics

- Rayleigh fading: $\mathbf{h}_{lk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{lk})$
  Constant for $\tau_c = T_c B_c$ channel uses
- Channel Covariance matrix $\mathbf{R}_{lk}$
  - Independent Rayleigh fading: $\mathbf{R}_{lk} = \beta_{lk}\mathbf{I}_M$
  - In practice, Correlated Rayleigh fading
- $\mathbf{R}_{lk}$ constant for τ_s channel blocks
- Two orders of magnitude slower variations



(a) Coherence block (b) Grid with fixed statistics [3]

Pilot sequences

- Pilot associated with UE k in cell j: $\phi_{jk} \in \mathbb{C}^K$ with $\|\phi_{jk}\|^2 = 1$
- Reused across cell
- Total pilot power ρ^{tr} per UE

[3] Bjornson, et al, 2016

Channel Estimation

LMMSE estimation of $\mathbf{h}_{jk}$

$$\hat{\mathbf{h}}_{lk} = \mathbf{R}_{lk}\mathbf{Q}_k^{-1}\mathbf{y}_{lk}^p$$

with

$$\mathbf{y}_{lk}^p = \mathbf{h}_{lk} + \sum_{i=1, i\neq j}^{L} \mathbf{h}_{ik} + \frac{1}{\sqrt{\rho^{\mathrm{tr}}}}\mathbf{N}_j^p \phi_{jk}^\star$$

and $\mathbf{Q}_k = \mathbb{E}\left\{\mathbf{y}_{lk}^p \left(\mathbf{y}_{lk}^p\right)^{\mathrm{H}}\right\}$ given by

$$\mathbf{Q}_k = \sum_{i=1}^{L} \mathbf{R}_{ik} + \frac{1}{\rho^{\mathrm{tr}}}\mathbf{I}_M$$

Spectral efficiency (Linear receive combining

- Channel capacity of UE k in cell l can be lower bounded by SE

$$\mathrm{SE}_{lk} = \left(1 - \frac{K}{\tau_c}\right)\mathbb{E}\left\{\log_2\left(1 + \gamma_{lk}\right)\right\} \quad [\mathrm{bit/s/Hz}]$$

Covariance Estimation - I

Estimation of $\mathbf{Q}_k$

- Estimation using pilot signal $\mathbf{y}_{jk}^p$ over $N_Q \leq \tau_s$ coherence blocks

$$\left[\hat{\mathbf{Q}}_{jk}^{(\text{sample})}\right] = \frac{1}{N_Q}\sum_{n=1}^{N_Q}\left[\mathbf{y}_{jk}^p[n]\left(\mathbf{y}_{jk}^p[n]\right)^{\mathrm{H}}\right]$$

- $$\frac{1}{N_Q}\sum_{n=1}^{N_Q}\left[\mathbf{y}_{jk}^p[n]\left(\mathbf{y}_{jk}^p[n]\right)^{\mathrm{H}}\right]_{m,m} \xrightarrow{\text{a.s.}} \left[\mathbf{Q}_{jk}\right]_{m,m}$$

  From Law of Large numbers, Standard deviation decays as $\frac{1}{\sqrt{N_Q}}$

- Regularization using a convex combination

$$\hat{\mathbf{Q}}_{jk}(\eta) = \eta\hat{\mathbf{Q}}_{jk}^{(\text{sample})} + (1-\eta)\hat{\mathbf{Q}}_{jk}^{(\text{diagonal})}$$

Covariance Estimation - II

Estimation of $\mathbf{R}_{lk}$

R direct: Specific training phase for $\mathbf{R}_{jk}$

- Every UE uses set of unique orthogonal pilots
- Implemented using pattern of $N_R KL$ extra pilots, N_R for each UE
- Sample covariance estimate and regularization same as above

Via Q: Two stage estimation procedure

- Each UE associated with unique orthogonal pilots, with total N_R pilots
- All UE's that cause pilot contamination will send the pilot
- BS estimates sample covariance $\mathbf{Q}_{jk,-k}^{(sample)}$ of $\mathbf{Q}_{jk} - \mathbf{h}_{jk}\mathbf{h}_{jk}^H$

$$\hat{\mathbf{R}}_{jjk}^{(\text{sample})} = \hat{\mathbf{Q}}_{jk}^{(\text{sample})} - \hat{\mathbf{Q}}_{jk,-k}^{(\text{sample})}$$

- Regularization

$$\hat{\mathbf{R}}_{jjk}(\mu) = \mu \hat{\mathbf{R}}_{jjk}^{(\text{sample})} + (1-\mu)\hat{\mathbf{R}}_{jjk}^{(\text{diagonal})}$$

Spectral Efficiency: Imperfect Covariance

- Lemma 2. The channel capacity of *UEk* in cell *j* is lower bounded by

$$\underline{\mathrm{SE}}_{jk} = \left(1 - \frac{K}{\tau_c} - \alpha\right) \log_2\left(1 + \underline{\gamma}_{jk}\right) \quad [\mathrm{bit/s/Hz}]$$

with $\alpha = \frac{N_R K L}{\tau_s}$ accounting for the extra pilots used for covariance matrix estimation and

$$\underline{\gamma}_{jk} = \frac{\left|\mathbb{E}\left\{\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\right\}\right|^2}{\sum_{l=1}^{L}\sum_{i=1}^{K}\mathbb{E}\left\{\left|\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jli}\right|^2\right\} - \left|\mathbb{E}\left\{\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\right\}\right|^2 + \frac{1}{\rho}\mathbb{E}\left\{\left\|\mathbf{v}_{jk}\right\|^2\right\}}$$

where the expectations are with respect to channel realizations.

- Closed form expressions for MRC type schemes, where $\mathbf{v}_{jk} = \mathbf{W}_{jk}\mathbf{y}_{jk}^{p}$

$$\mathbf{W}_{jk} = \begin{cases} \mathbf{R}_{jk}\mathbf{Q}_k^{-1} & \text{MMSE estimator} \\ \hat{\mathbf{R}}_{jk}(\mu)\left(\hat{\mathbf{Q}}_k(\eta)\right)^{-1} & \text{Approximate MMSE estimator} \\ \mathbf{I}_M & \text{LS estimator.} \end{cases}$$

Discussion - II

Summary

- Imperfect Channel Covariance Estimation
- Aim to maximize Spectral efficiency
- LMMSE vs element wise LMMSE estimation (Kocharlakota, et al.)

Future Scope

- Exploit structure (Eg. Toeplitz)
- mmWave Massive MIMO
- Modified pilot transmission scheme

References

- E. Björnson, L. Sanguinetti, and M. Debbah, “Massive mimo with imperfect channel covariance information,” in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 974–978.
- 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.
- A. K. Kocharlakota, K. Upadhya, and S. A. Vorobyov, “On achievable rates for massive mimo system with imperfect channel covariance information,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 4504–4508.
- 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.