Matching Pursuit Covariance Learning

Esa Ollila

EUSIPCO-2024, August 27th, 2024 at Lyon

Aalto University, Department of Information and Communications Engineering

Multiple measurements vector (MMV) model

- *Multiple measurements vector* (MMV) [DE11] model:

$$\mathbf{y}_l = \mathbf{A}\mathbf{x}_l + \mathbf{e}_l, \quad l = 1, \ldots, L,$$

  where
  - $\mathbf{A} = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_M) \in \mathbb{C}^{N \times M}$ known overcomplete matrix $(M > N)$, called the *dictionary*
  - $\mathbf{x}_l \in \mathbb{C}^M$ unobserved sparse signal vectors
  - $\mathbf{e}_l \in \mathbb{C}^N$ unobserved zero mean white noise vector, $\mathrm{cov}(\mathbf{e}_l) = \sigma^2 \mathbf{I}$

  The column vectors $\mathbf{a}_i$ of $\mathbf{A}$ are referred to as *atoms*.
- In matrix form:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E},$$

Key assumption

Signals $\mathbf{x}_l$ are jointly K-sparse: $\mathbf{X}$ is K-**rowsparse** (i.e., only K rows are different from zero).

Sparse signal recovery methods

Sparse signal recovery (SSR) methods aim at identifying the support

$$\mathcal{M} = \mathrm{supp}(\mathbf{X}) = \{i \in [\![M]\!] : x_{ij} \neq 0 \text{ for some } j \in [\![L]\!]\}$$

given only the data $\mathbf{Y}_{N\times L}$, the dictionary $\mathbf{A}_{N\times M}$, and the sparsity level K.

Popular approaches:

- Convex relaxation:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{AX}\|^2 + \lambda\ell(\mathbf{X})$$

  for some mixed matrix norm $\ell(\cdot)$ and proper choice of penalty λ.
- Approximate solution for $\|\cdot\|_0$ pseudonorm, e.g., simultaneous normalized iterative hard thresholding (**SNIHT**) [BCHJ14].
- Greedy methods: greedily add atoms to $\mathcal{M}$ that provide best fit to LS criterion as in simultaneous orthogonal matching pursuit (**SOMP**) [TGS06].
- Covariance-based (see next slide) as **M-SBL** (Sparse Bayesian Learning) [WR07],

Covariance-based sparse signal recovery

- signal: circular Gaussian random variables with uncorrelated components:

$$\mathbf{x}_l \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{\Gamma}) \quad \text{with} \quad \mathbf{\Gamma} = \mathrm{cov}(\mathbf{x}_l) = \mathrm{diag}(\boldsymbol{\gamma})$$

  where $\boldsymbol{\gamma} \in \mathbb{R}^M_{\geq 0}$ contains signal powers with only K non-zero elements, i.e., $\mathcal{M} = \mathrm{supp}(\mathbf{X}) = \mathrm{supp}(\boldsymbol{\gamma})$.
- $\mathbf{e}_l \sim \mathcal{CN}_N(\mathbf{0}, \sigma^2\mathbf{I})$ where σ^2 is the unknown noise variance.
- $\mathbf{y}_l$ i.i.d., $\mathbf{x}_l$ and $\mathbf{e}_l$ uncorrelated.

Under the above conditions, $\mathbf{y}_l \sim \mathcal{CN}_N(\mathbf{0}, \mathbf{\Sigma})$, with covariance matrix $\mathbf{\Sigma} = \mathrm{cov}(\mathbf{y}_l)$ given by

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I} = \sum_{i=1}^{M} \gamma_i \mathbf{a}_i \mathbf{a}_i^{\mathsf{H}} + \sigma^2\mathbf{I}.$$

Covariance-based support recovery: minimize the neg. log-likelihood fnc (LLF) $\ell(\boldsymbol{\gamma}, \sigma^2)$ of the data $\mathbf{Y}$ given $\mathbf{A}$ and sparsity K.

OMP formulation in [Ela10, Table 3.1]:

Initialization: Initialize $k = 0$, and set

- The initial solution $\mathbf{x}^0 = \mathbf{0}$.
- The initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0 = \mathbf{b}$.
- The initial solution support $\mathcal{S}^0 = Support\{\mathbf{x}^0\} = \emptyset$.

Main Iteration: Increment k by 1 and perform the following steps:

- **Sweep:** Compute the errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{k-1}\|_2^2$ for all j using the optimal choice $z_j^* = \mathbf{a}_j^T \mathbf{r}^{k-1} / \|\mathbf{a}_j\|_2^2$.
- **Update Support:** Find a minimizer, j_0 of $\epsilon(j)$: $\forall\, j \notin \mathcal{S}^{k-1}$, $\epsilon(j_0) \leq \epsilon(j)$, and update $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{j_0\}$.
- **Update Provisional Solution:** Compute $\mathbf{x}^k$, the minimizer of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ subject to $Support\{\mathbf{x}\} = \mathcal{S}^k$.
- **Update Residual:** Compute $\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k$.
- **Stopping Rule:** If $\|\mathbf{r}^k\|_2 < \epsilon_0$, stop. Otherwise, apply another iteration.

Output: The proposed solution is $\mathbf{x}^k$ obtained after k iterations.

Sweep: solve MLE when all but one signal parameter is not known.

Update provisional solution: solve the MLE in non-sparse case.

OMP formulation in [Ela10, Table 3.1]:

Initialization: Initialize $k = 0$, and set

- The initial solution $\mathbf{x}^0 = \mathbf{0}$.
- The initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0 = \mathbf{b}$.
- The initial solution support $\mathcal{S}^0 = Support\{\mathbf{x}^0\} = \emptyset$.

Main Iteration: Increment k by 1 and perform the following steps:

- **Sweep:** Compute the errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{k-1}\|_2^2$ for all j using the optimal choice $z_j^* = \mathbf{a}_j^T \mathbf{r}^{k-1} / \|\mathbf{a}_j\|_2^2$.
- **Update Support:** Find a minimizer, j_0 of $\epsilon(j)$: $\forall\, j \notin \mathcal{S}^{k-1}$, $\epsilon(j_0) \le \epsilon(j)$, and update $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{j_0\}$.
- **Update Provisional Solution:** Compute $\mathbf{x}^k$, the minimizer of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ subject to $Support\{\mathbf{x}\} = \mathcal{S}^k$.
- **Update Residual:** Compute $\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k$.
- **Stopping Rule:** If $\|\mathbf{r}^k\|_2 < \epsilon_0$, stop. Otherwise, apply another iteration.

Output: The proposed solution is $\mathbf{x}^k$ obtained after k iterations.

We can solve these steps also for our likelihood!

Maximum likelihood estimation results [1/2]: sweep

Gaussian negative LLF:

$$\ell(\boldsymbol{\gamma}, \sigma^2 \mid \mathbf{Y}, \mathbf{A}) = \operatorname{tr}\big((\underbrace{\mathbf{A}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}}_{=\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{\Sigma}}\big) + \log|\underbrace{\mathbf{A}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}}_{=\boldsymbol{\Sigma}}|$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l\mathbf{y}_l^{\mathsf{H}}$ is the sample covariance matrix (SCM).

Maximum likelihood estimation results [1/2]: sweep

Gaussian negative LLF:

$$\ell(\boldsymbol{\gamma}, \sigma^2 \mid \mathbf{Y}, \mathbf{A}) = \operatorname{tr}((\underbrace{\mathbf{A}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}}_{=\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{\Sigma}}) + \log|\underbrace{\mathbf{A}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}}_{=\boldsymbol{\Sigma}}|$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l\mathbf{y}_l^{\mathsf{H}}$ is the sample covariance matrix (SCM).

Result 1. [FT01, YLS+10] The unique minimizer of the **conditional** neg. LLF for $\gamma_i \geq 0$ (with fixed $\{\gamma_j\}_{j\neq i}$, σ^2) is:

$$\begin{aligned}\hat{\gamma}_i &= \min_{\gamma}\ \operatorname{tr}((\boldsymbol{\Sigma}_{\setminus i} + \gamma\mathbf{a}_i\mathbf{a}_i^{\mathsf{H}})^{-1}\hat{\boldsymbol{\Sigma}}) + \log|\boldsymbol{\Sigma}_{\setminus i} + \gamma\mathbf{a}_i\mathbf{a}_i^{\mathsf{H}}| \\ &= \max\left(\frac{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}_{\setminus i}^{-1}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{\setminus i})\boldsymbol{\Sigma}_{\setminus i}^{-1}\mathbf{a}_i}{(\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}_{\setminus i}^{-1}\mathbf{a}_i)^2}, 0\right),\end{aligned}$$

where $\boldsymbol{\Sigma}_{\setminus i} = \boldsymbol{\Sigma} - \gamma_i\mathbf{a}_i\mathbf{a}_i^{\mathsf{H}}$ = array CM without the contribution of the i^{th} source.

Maximum likelihood estimation results [2/2]: update provisional solution

Result 2. (non-sparse, underdetermined case) [SN95] If $M = K$ (so # of atoms = # of sources), the unrestricted minimizers of the neg. LLF $\ell(\boldsymbol{\gamma}, \sigma^2)$ are

$$\hat{\sigma}^2 = \frac{1}{N-K}\mathrm{tr}\big((\mathbf{I} - \mathbf{A}\mathbf{A}^+)\hat{\boldsymbol{\Sigma}}\big)$$
$$\hat{\gamma} = \mathrm{diag}(\mathbf{A}^+(\hat{\boldsymbol{\Sigma}} - \hat{\sigma}^2\mathbf{I})\mathbf{A}^{+\mathsf{H}})$$

provided that $\hat{\gamma}_i > 0$ for all $i = 1, \ldots, K$. (Note: $\mathbf{A}$ is $N \times K$, $\boldsymbol{\gamma}$ is $K \times 1$, $\mathbf{A}^+$ is the pseudoinverse of $\mathbf{A}$)

Remark. If $\hat{\gamma}$ in Result 2 contains negative elements, then we calculate the constrained (non-negative) solution using [Bre88, Algorithm I].

CL-OMP steps

Initialization: Initialize $k = 0$ and set

- The initial solution $\boldsymbol{\gamma}^{(0)} = \mathbf{0}_{M\times 1}$, $\sigma^{2(0)} = \mathrm{tr}(\hat{\boldsymbol{\Sigma}})/N$
- The initial solution support $\mathcal{M}^{(0)} = \mathrm{supp}(\boldsymbol{\gamma}^{(0)}) = \emptyset$
- The initial CM: $\boldsymbol{\Sigma}^{(0)} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^{(0)})\mathbf{A}^{\mathsf{H}} + \sigma^{2(0)}\mathbf{I} = \sigma^{2(0)}\mathbf{I}$

Main Iteration ($k = 0, 1, \ldots, K-1$):

1. **Sweep**: for each conditional MLE $\hat{\gamma}_i$ (cf. Result 1), compute the errors (fits):

$$\epsilon_i = \mathrm{tr}((\boldsymbol{\Sigma}^{(k)} + \hat{\gamma}_i \mathbf{a}_i \mathbf{a}_i^{\mathsf{H}})^{-1}\hat{\boldsymbol{\Sigma}}) + \log|\boldsymbol{\Sigma}^{(k)} + \hat{\gamma}_i \mathbf{a}_i \mathbf{a}_i^{\mathsf{H}}|$$

for all i. This is equivalent to computing:

$$\epsilon_i = \log(1 + \hat{\gamma}_i \mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Theta}^{(k)} \mathbf{a}_i) - \hat{\gamma}_i \mathbf{a}_i \boldsymbol{\Theta}^{(k)}\ \mathbf{a}_i + \text{constant}$$

where $\boldsymbol{\Theta}^{(0)} = (\boldsymbol{\Sigma}^{(0)})^{-1}$.

CL-OMP steps

1. **Update support**: Find a minimizer, i_k of ϵ_i:

$$\epsilon_{i_k} \leq \epsilon_i, \quad \forall i \notin \mathcal{M}^{(k)},$$

and update the support $\mathcal{M}^{(k+1)} = \mathcal{M}^{(k)} \cup \{i_k\}$

2. **Update provisional solution**: minimize the negative LLF s.t. $\mathrm{supp}(\boldsymbol{\gamma}) = \mathcal{M}^{(k+1)}$, i.e. solve the underdetermined (non-sparse) case:

$$(\hat{\mathbf{g}}, \hat{\sigma}^2) = \arg\min_{\mathbf{g}, \sigma^2} \ell(\mathbf{g}, \sigma^2 \mid \mathbf{Y}, \mathbf{A}_{\mathcal{M}^{(k+1)}}),$$

where ℓ is the neg. LLF. (Note: solution via Result 2 or [Bre88, Algorithm I])

3. **Update CM**:

$$\boldsymbol{\Sigma}^{(k+1)} = \mathbf{A}_{\mathcal{M}^{(k+1)}} \operatorname{diag}(\hat{\mathbf{g}}) \mathbf{A}^{\mathsf{H}}_{\mathcal{M}^{(k+1)}} + \hat{\sigma}^2 \mathbf{I}.$$

Algorithm 1: CL-OMP: **C**ovariance **L**earning **O**rthogonal **M**atching **P**ursuit algorithm

Input : $\hat{\mathbf{\Sigma}}$, $\mathbf{A}$, K

Initialize: $\mathbf{\Sigma} = [\mathrm{tr}(\hat{\mathbf{\Sigma}})/p]\mathbf{I}$, $\mathcal{M} = \emptyset$

for $k = 1, \ldots, K$ **do**

$\boldsymbol{\gamma} = (\gamma_i)_{M\times 1}$, $\gamma_i \leftarrow \max\left(\frac{\mathbf{a}_i^{\mathsf{H}}\mathbf{\Sigma}^{-1}(\hat{\mathbf{\Sigma}}-\mathbf{\Sigma})\mathbf{\Sigma}^{-1}\mathbf{a}_i}{(\mathbf{a}_i^{\mathsf{H}}\mathbf{\Sigma}^{-1}\mathbf{a}_i)^2}, 0\right)$ `// Result 1`

$\boldsymbol{\epsilon} = (\epsilon_i) \leftarrow \left(\log(1+\gamma_i\mathbf{a}_i^{\mathsf{H}}\mathbf{\Sigma}^{-1}\mathbf{a}_i) - \gamma_i\mathbf{a}_i^{\mathsf{H}}\mathbf{\Sigma}^{-1}\mathbf{a}_i\right)_{M\times 1}$ `// errors (value of neg. LLF at solution)`

$\mathcal{M} \leftarrow \mathcal{M} \cup \{i_k\}$ with $i_k \leftarrow \arg\min_{i\notin\mathcal{M}} \epsilon_i$ `// choose source with smallest error`

$\sigma^2 \leftarrow \frac{1}{N-k}\mathrm{tr}\big((\mathbf{I} - \mathbf{A}_{\mathcal{M}}\mathbf{A}_{\mathcal{M}}^{+})\hat{\mathbf{\Sigma}}\big)$ `// by Result 2`

$\boldsymbol{\gamma}_{\mathcal{M}} \leftarrow \max\left(\mathrm{diag}\left(\mathbf{A}_{\mathcal{M}}^{+}(\hat{\mathbf{\Sigma}} - \sigma^2\mathbf{I})\mathbf{A}_{\mathcal{M}}^{+\mathsf{H}}\right), 0\right)$ `// by Result 2`

$\boldsymbol{\gamma}_{\mathcal{M}^{\complement}} \leftarrow \mathbf{0}$

$\mathbf{\Sigma} \leftarrow \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$

Output : $\mathcal{M}$, $\boldsymbol{\gamma}$, σ^2

Simulation set-up

Methods:

- **SOMP** [TGS06, Algorithm 3.1]
- **SNIHT** [BCHJ14, Algorithm 1].
- **M-SBL** (joint maximization of σ^2 and $\boldsymbol{\gamma}$, with σ^2 update [WR07, Eq. (21)] in M-step)

Set-up:# of atoms $M = 256$, sparsity level $K = 4$.

- Dictionary $\mathbf{A}$: Gaussian random $N \times M$ measurement matrix, unit-norm columns.
- Support $\mathcal{M} = \mathrm{supp}(\mathbf{X})$: randomly chosen from $\{1, \dots, M\}$ for each trial.
- Noise $\mathbf{e}_i$: white circular Gaussian with variance σ^2
- Sparse signal: $[\mathbf{x}_i]_j \sim \mathcal{CN}(0, \gamma_j)$, for $j \in \mathcal{M}$.
- SNR: frst source $10 \log_{10} \gamma_1 / \sigma^2$ while others have 1 dB, 2 dB, and 4 dB lower SNR.

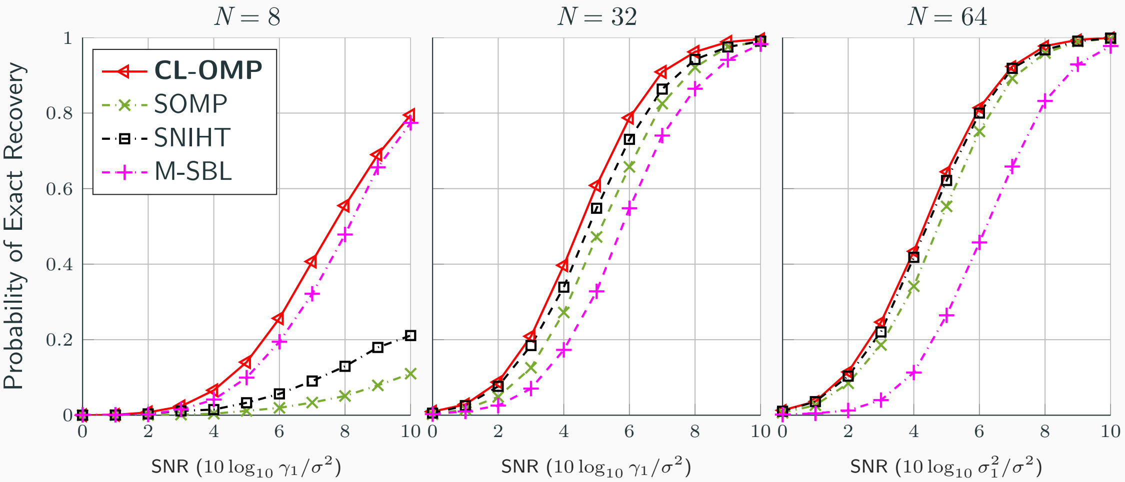
Simulation set-up

- (Empirical) *probability of exact recovery*,

$$\text{PER} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{I}\big(\hat{\mathcal{M}}^{(t)} = \mathcal{M}^{(t)}\big),$$

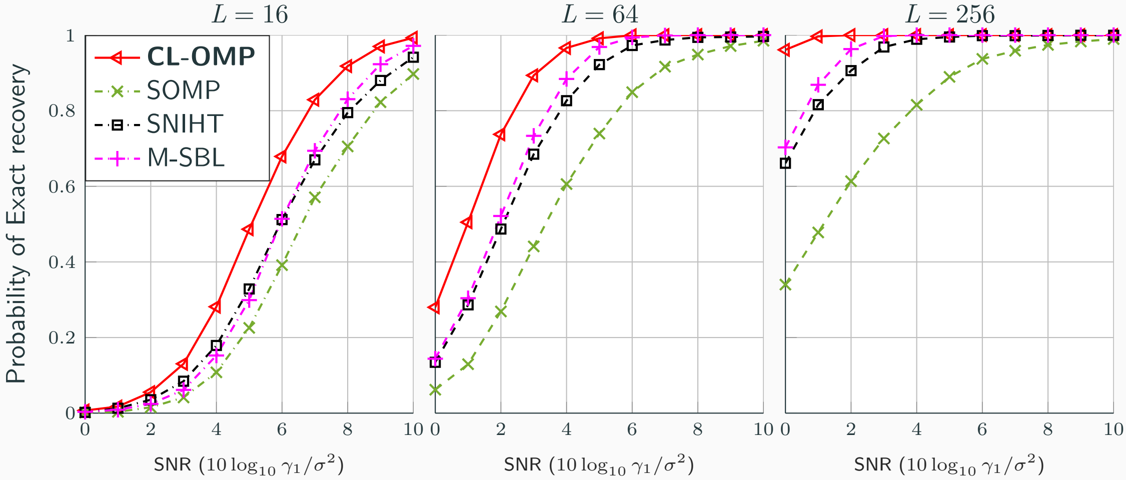
  where $\mathrm{I}(\cdot)$ denotes the indicator function, and $\hat{\mathcal{M}}^{(t)}$ denotes the estimate of the true signal support $\mathcal{M}^{(t)}$ for t^{th} MC trial
- # of MC trials $T = 2000$.

Study 1: $M = 256$, $K = 4$, $L = 16$, while N varies.



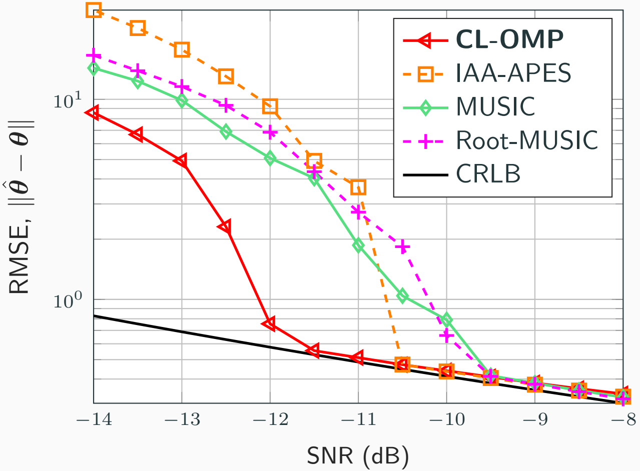
M-SBL not consistent; SOMP and SNIHT perform poorly in small N.

Study 2: $M = 256$, $K = 4$, $N = 16$, while L varies



Other methods perform poorly in small L and low SNR.

How about DOA estimation?



$K = 4$ Gaussian sources; $L = 125$, $N = 20$, $M = 1801$ $(\Delta\theta = 0.1^o)$

$\boldsymbol{\theta} = (-30.1^o, -20.02^o, -10.02^o, 3.02^o)$. MC trials $= 2000$.

Conclusions

- We proposed a covariance learning orthogonal matching pursuit (**CL-OMP**) algorithm.
- CL-OMP outperformed traditional SSR methods
- Especially, when N or L is small, or in low SNR, the SNIHT and SOMP performed very poorly compared to CL-OMP
- As DOA estimation method, CL-OMP outperformed MUSIC, Root-MUSIC, IAA-APES.

MATLAB and python codes available at github:

`https://github.com/esollila/CovLearn`

References i

- J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jin, *Greedy algorithms for joint sparse recovery*, IEEE Trans. Signal Process. **62** (2014), no. 7, 1694 – 1704.
- Yoram Bresler, *Maximum likelihood estimation of a linearly structured covariance with application to antenna array processing*, Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling, IEEE, 1988, pp. 172–175.
- M. F. Duarte and Y. C. Eldar, *Structured compressed sensing: From theory to applications*, IEEE Trans. Signal Process. **59** (2011), no. 9, 4053–4085.
- M. Elad, *Sparse and redundant representations*, Springer, New York, 2010.
- Anita Faul and Michael Tipping, *Analysis of sparse bayesian learning*, Advances in neural information processing systems **14** (2001).

References ii

- Petre Stoica and Arye Nehorai, *On the concentrated stochastic likelihood function in array signal processing*, Circuits, Systems and Signal Processing **14** (1995), no. 5, 669–674.
- J. A. Tropp, A. C. Gilbert, and M. J. Strauss, *Algorithms for simultaneous sparse approximation. Part I: greedy pursuit*, Signal Processing **86** (2006), 572–588.
- David P Wipf and Bhaskar D Rao, *An empirical Bayesian strategy for solving the simultaneous sparse approximation problem*, IEEE Transactions on Signal Processing **55** (2007), no. 7, 3704–3716.
- Tarik Yardibi, Jian Li, Petre Stoica, Ming Xue, and Arthur B Baggeroer, *Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares*, IEEE Trans. Aerosp. Electron. Syst. **46** (2010), no. 1, 425–443.