Learning Hidden Markov Sparse Models

Lin Li and Anna Scaglione

Department of Electrical and Computer Engineering, University of California, Davis, CA 95616

Email: {llli, ascaglione}@ucdavis.edu

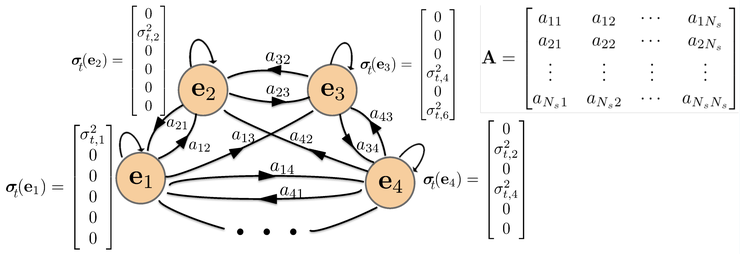
Hidden Markov Sparse Models

The signal observed from multiple receivers:

$$\boldsymbol{y}_t = \boldsymbol{D}\boldsymbol{x}_t + \boldsymbol{M}\boldsymbol{s}_t + \boldsymbol{w}_t \quad (1)$$

$$\boldsymbol{s}_t = \boldsymbol{A}\boldsymbol{s}_{t-1} + \boldsymbol{v}_{t-1} \quad (2)$$

1) $\boldsymbol{D} = [\boldsymbol{d}_1, \cdots, \boldsymbol{d}_K]$ is an over-complete dictionary;
2) $\boldsymbol{s}_t \in [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_{N_s}]$ is a set of Markov states;
3) $\boldsymbol{x}_t|_{\boldsymbol{s}_t} \sim \mathcal{N}(0, \mathrm{diag}\,(\boldsymbol{\sigma}_t))$, where $\boldsymbol{\sigma}_t = \boldsymbol{\sigma}(\boldsymbol{s}_t)$ is **sparse**.

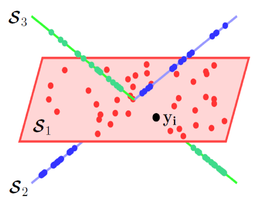


In summary:

- Sources are **on-off** determined by the hidden Markov chain
- Each state corresponds to a set of active or inactive sources.
- Given only the observations, find $\boldsymbol{D}$, $\boldsymbol{M}$, $\boldsymbol{s}_t$ and $\boldsymbol{A}$.

Applications

- **SUBSPACE CLUSTERING:** (data mining, motion segmentation)

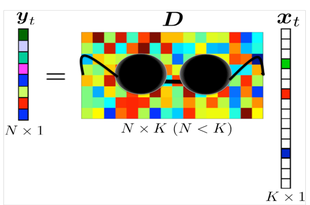


$S_i = \{\boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{U}_i\boldsymbol{x} + \boldsymbol{\mu}_i\},\ i = 1, \cdots, I.$

- $\boldsymbol{\mu}_i = \boldsymbol{M}\boldsymbol{s}_i$
- $\boldsymbol{U}_i$: a subset of columns of $\boldsymbol{D}$

The **goal** is to find (1) the number I; (2) the affine subspace S_i with unknown centroid $\boldsymbol{\mu}_i$, unknown basis $\boldsymbol{U}_i$ and unknown dimensions d_i.

- **UNDER-DETERMINED BLIND SOURCE SEPARATION:**



$\boldsymbol{y}_t = \boldsymbol{D}\boldsymbol{x}_t + \boldsymbol{w}_i$

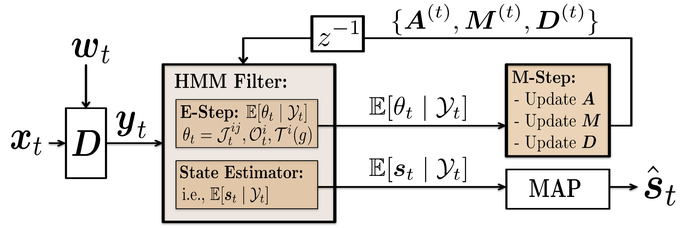
- $\boldsymbol{D}\boldsymbol{\mu}(\boldsymbol{s}_t) = \boldsymbol{M}\boldsymbol{s}_t$

The **goal** is to recover $\boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{s}_t), \mathrm{diag}(\boldsymbol{\sigma}(\boldsymbol{s}_t)))$.

Post Processing: $\boldsymbol{D}^{(t)}$, $\hat{\boldsymbol{s}}_t$ → MMSE → $\hat{\boldsymbol{x}}_t$

Proposed Learning Algorithm

PROPOSED RECURSIVE ARCHITECTURE: (EM)



$$\mathcal{O}_t^i = \sum_{\ell=1}^{t} \langle \boldsymbol{s}_{\ell-1}, \boldsymbol{e}_i \rangle;\ \mathcal{J}_t^{ij} = \sum_{\ell=1}^{t} \langle \boldsymbol{s}_\ell, \boldsymbol{e}_j \rangle \langle \boldsymbol{s}_{\ell-1}, \boldsymbol{e}_i \rangle;\ \mathcal{T}^i(g) = \sum_{\ell=1}^{t} \langle \boldsymbol{s}_\ell, \boldsymbol{e}_i \rangle g(\boldsymbol{y}_\ell)$$

HMM FILTERING: (Change of Measure)

Radon-Nikodym Derivative: $\Lambda_t := \prod_{\ell=1}^{t} \lambda_\ell$

$$\lambda_\ell = \det R_\ell^{-\frac{1}{2}} \exp\left[\boldsymbol{y}_\ell^T \boldsymbol{y}_\ell / 2 - (\boldsymbol{y}_\ell - \boldsymbol{M}\boldsymbol{s}_\ell)^T R_\ell^{-1} (\boldsymbol{y}_\ell - \boldsymbol{M}\boldsymbol{s}_\ell)/2\right]$$

$$R_\ell = \boldsymbol{D}\,\mathrm{diag}\,(\boldsymbol{\sigma}(\boldsymbol{s}_\ell))\,\boldsymbol{D}^T + \sigma_n^2 \boldsymbol{I}$$

Conditional Bayes' Theorem:

$$\mathbb{E}\{\boldsymbol{s}_t \mid \mathcal{Y}_t\} = \frac{\tilde{\mathbb{E}}\{\Lambda_t \boldsymbol{s}_t \mid \mathcal{Y}_t\} := \mathcal{F}(\boldsymbol{s}_t)}{\tilde{\mathbb{E}}\{\Lambda_t \mid \mathcal{Y}_t\}} \quad \text{(prob. dist. of the state)}$$

where $\mathcal{F}(\boldsymbol{s}_t) = \mathrm{diag}\,(\lambda_t(\boldsymbol{e}_1), \cdots, \lambda_t(\boldsymbol{e}_{N_s}))\,\boldsymbol{A}^T \mathcal{F}(\boldsymbol{s}_{t-1})$ (**recursive!**).

PARAMETER ESTIMATION: ($\boldsymbol{A}, \boldsymbol{M}$ and $\boldsymbol{D}$)

- Closed-Form Solutions for $\boldsymbol{A}, \boldsymbol{M}$:

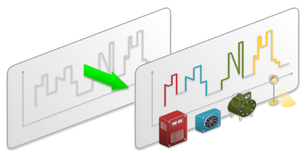
$$\boldsymbol{A}_{ij}^{(t)} = \mathbb{E}[\mathcal{O}_t^i \mid \mathcal{Y}_t]^{-1} \mathbb{E}[\mathcal{J}_t^{ij} \mid \mathcal{Y}_t]\ ,\quad \boldsymbol{M}_{ri}^{(t)} = \mathbb{E}[\mathcal{O}_t^i \mid \mathcal{Y}_t]^{-1} \mathbb{E}[\mathcal{T}^i(\boldsymbol{y}_\ell(r)) \mid \mathcal{Y}_t].$$

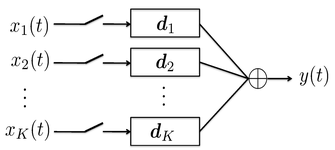
- Dictionary Learning: $\boldsymbol{D}' = \boldsymbol{D}\,\mathrm{diag}(\sigma_t)^{1/2}$

$$\min_{\boldsymbol{D}} C(\boldsymbol{D}) := \mathbb{E}\left[\sum_{\ell=1}^{m} (\boldsymbol{y}_\ell - \boldsymbol{M}\boldsymbol{s}_\ell)^T R_\ell^{-1} (\boldsymbol{y} - \boldsymbol{M}\boldsymbol{s}_\ell) + \log\det R_\ell \,\middle|\, \mathcal{Y}_t\right]$$

Generalization to Structured Dictionary

Applications: Non-Intrusive Load Monitoring, Hybrid Systems, etc.





$\boldsymbol{y}_t = \sum_{k=1}^{K} H(\boldsymbol{d}_k)\boldsymbol{x}_{t,k} + \boldsymbol{w}_t = \boldsymbol{D}\boldsymbol{x}_t + \boldsymbol{w}_i$

- $\boldsymbol{y}_t = [y(t), \cdots, y(t - N - L_s)]^T$
- $\boldsymbol{D} = [H(\boldsymbol{d}_1), \cdots, H(\boldsymbol{d}_K)]$, $H(\boldsymbol{d}_i)$ convolution matrix
- $\boldsymbol{x}_t = [\boldsymbol{x}_{t,1}^T, \cdots, \boldsymbol{x}_{t,K}^T]^T$, $\boldsymbol{x}_{t,k} = [x_k(t), \cdots, x_k(t - L_s - 1)]^T$

  $\boldsymbol{x}_{t,k} \sim \mathcal{N}(\boldsymbol{\mu}_k, \sigma_k^2 \boldsymbol{I})$ if the kth source is active.

Numerical Results

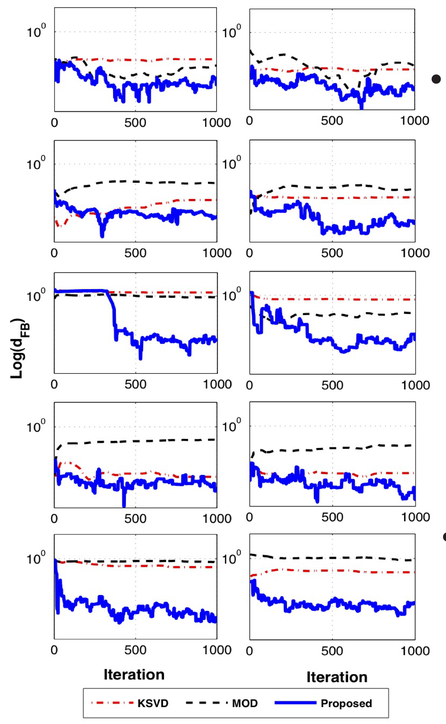
- **Parameters** :

$t = 1000$ source blocks
$N = 6$ receivers
$K = 10$ sources
$S = 2$ (sparsity)
$\max_{1 \le i \ne j \le K} |d_i^H d_j| = 0.6014$ (mutual coherence)

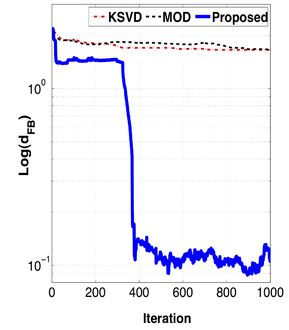
- **Dictionary Learning Techniques** :

1) Proposed (proposed recursive method)
2) MOD (method of optimal direction)
3) KSVD (clustering-based method)

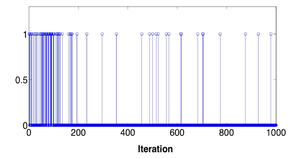
- **Numerical Resuls** : $d_{\mathrm{FB}}(\boldsymbol{d}_k, \boldsymbol{d}_k') = \cos^{-1}\left[\frac{|\boldsymbol{d}_k^T \boldsymbol{d}_k'|}{\|\boldsymbol{d}_k\| \|\boldsymbol{d}_k'\|}\right]$



- $d_{\mathrm{FB}}(\boldsymbol{D}, \boldsymbol{D}') = \sum_{k=1}^{K} d_{\mathrm{FB}}(\boldsymbol{d}_k, \boldsymbol{d}_k')$



- Err. in State Est.:



Related Works

[1] R.J. Elliott et al., "Hidden Markov Models: Estimation and Control," Springer-Verlag, 1994.

[2] M. Aharon et al., "On the Uniqueness of Overcomplete Dictionaries, and a Practical Way to Retrieve Them," *Linear Algebra and Its Applications*, 2006, vol. 416, pp. 48-67.

[3] J. H. Manton,"Optimization algorithms exploiting unitary constraints," *IEEE Trans. on Sig. Proc.*, 2002, vol. 50, pp. 635-650.

[4] D. Donoho, "Compressed Sensing", *Information Thoery, IEEE Trans on*, vol. 52, pp. 1289-1306, 2006.