Sparse Bayesian Learning: Block-Sparsity

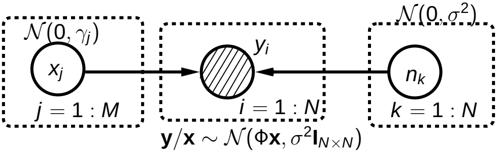
Ranjitha Prasad, C. R. Murthy, B. D. Rao

July 6, 2013

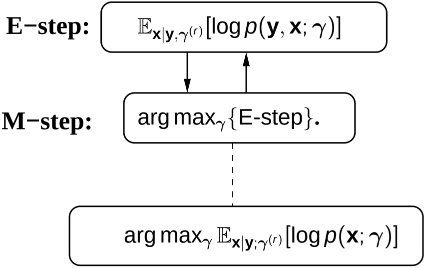
Outline

1. SBL Preliminaries
2. Block Sparse Signals
   - Definition
   - Parallel Clustered-SBL
   - Block Sparse SBL: Correlated Case
   - Nested EM Algorithm

Sparse Bayesian Learning



- Problem: $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$
- Bayesian: Prior pdf on sparse vector $\mathbf{x}$
- Sparse Bayesian Learning: $\mathbf{x} \sim \mathcal{N}(0, \Gamma)$, $\Gamma = \text{diag}(\gamma)$. $|\gamma|_0 = K$
- Based on the iterative Expectation Maximization framework



- E-step: Compute $p(\mathbf{x}|\mathbf{y};\gamma^{(r)})$
- M-step: Maximization
- Exact inference

What is a Block-sparse signal?

$\mathbf{x}_{11}$	$\mathbf{x}_{12}$	$\ldots$	$\mathbf{x}_{1M}$	$\ldots$	$\mathbf{x}_{B1}$	$\mathbf{x}_{B2}$	$\ldots$	$\mathbf{x}_{BM}$

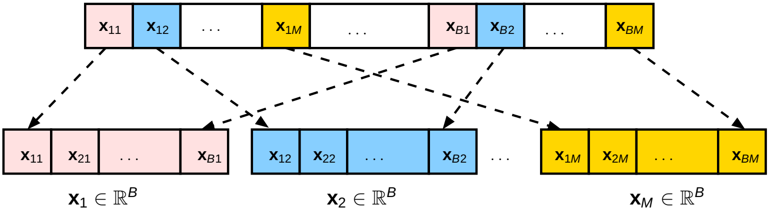
$\mathbf{x}_1 \in \mathcal{R}^M$

$\mathbf{x}_1 \sim \mathcal{N}(0, \gamma_1 \mathbf{B})$

$\mathbf{x}_B \in \mathcal{R}^M$

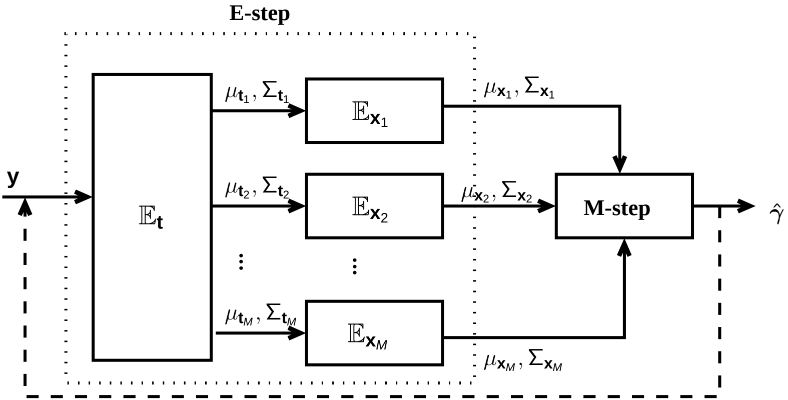
$\mathbf{x}_B \sim \mathcal{N}(0, \gamma_B \mathbf{B})$

- $\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}$
- Every block of the same length
- Block-sparsity manifested through $\gamma = [\gamma_1, \ldots, \gamma_B]$
- We consider the two cases, (i) $\mathbf{B} = \mathbf{I}$ (ii) $\mathbf{B}$ derived from an AR model



- $\mathbf{x}_1, \ldots, \mathbf{x}_M$ have the same support: group-sparse, $\mathbf{x}_i \sim \mathcal{N}(0, \Gamma)$
-

$$\mathbf{y} = \sum_{i=1}^{M} \mathbf{t}_i \quad \text{where} \quad \mathbf{t}_i = \Phi_i \mathbf{x}_i + \mathbf{n}_i, \quad 1 \leq i \leq M,$$



- E-step: decomposed into $\mathbb{E}_{\mathbf{t}}$ and $\mathbb{E}_{\mathbf{x}}$
- M-step combines and computes an estimate of γ
- Exact inference

$$\text{E-step}: Q\left(\gamma|\gamma^{(r)}\right) = \mathbb{E}_{\mathbf{t},\mathbf{x}|\mathbf{y};\gamma^{(r)}}[\log p(\mathbf{y},\mathbf{t},\mathbf{x};\gamma)]$$

$$\text{M-step}: \gamma^{(r+1)} = \arg\max_{\boldsymbol{\gamma}:\gamma_i\in\mathbb{R}_+} Q\left(\gamma|\gamma^{(r)}\right)$$

$$p\left(\mathbf{t},\mathbf{x}|\mathbf{y};\gamma^{(r)}\right) = p\left(\mathbf{x}|\mathbf{t},\mathbf{y};\gamma^{(r)}\right)p\left(\mathbf{t}|\mathbf{y};\gamma^{(r)}\right)$$
$$= p\left(\mathbf{x}|\mathbf{t};\gamma^{(r)}\right)p\left(\mathbf{t}|\mathbf{y};\gamma^{(r)}\right)$$

E-step can be rewritten as

$$\text{E-step}: Q\left(\gamma|\gamma^{(r)}\right) = \underbrace{\mathbb{E}_{\mathbf{t}|\mathbf{y};\gamma^{(r)}}}_{\mathbb{E}_{\mathbf{t}}}\underbrace{\mathbb{E}_{\mathbf{x}|\mathbf{t};\gamma^{(r)}}}_{\mathbb{E}_{\mathbf{x}}}[\log p(\mathbf{y},\mathbf{t},\mathbf{x};\gamma)]$$

E-step

- Compute $p(\mathbf{t}|\mathbf{y};\gamma^{(r)})$: $p(\mathbf{t}_m|\mathbf{x}_m) = \mathcal{N}(\Phi_m \mathbf{x}_m, \beta_m \sigma^2)$
- $p(\mathbf{x}) = \prod_{m=1}^{M} p(\mathbf{x}_m)$, $p(\mathbf{x}) = \prod_{m=1}^{M} p(\mathbf{x}_m)$
- Given $\mathbf{H} = \underbrace{[\mathbf{I}_N, \ldots, \mathbf{I}_N]}_{M \text{ times}}$ and $\mathbf{y} = \mathbf{Ht}$, compute

$$p(\mathbf{t}|\mathbf{y};\gamma^{(r)}) = \mathcal{N}(\mu_t, \Sigma_t),$$

$$\mu_t = (\mathbf{R} + \Phi_B \Gamma_B \Phi_B^T)\mathbf{H}^T(\mathbf{H}(\mathbf{R} + \Phi_B \Gamma_B \Phi_B^T)\mathbf{H}^T)^{-1}\mathbf{y}$$

$$\Sigma_t = (\mathbf{R} + \Phi_B \Gamma_B \Phi_B^T) -$$

$$(\mathbf{R} + \Phi_B \Gamma_B \Phi_B^T)\mathbf{H}^T(\mathbf{H}(\mathbf{R} + \Phi_B \Gamma_B \Phi_B^T)\mathbf{H}^T)^{-1}\mathbf{H}(\mathbf{R} + \Phi_B \Gamma_B \Phi_B^T).$$

where Φ_B: block diagonal, $\Phi_1 \ldots \Phi_M$ along the diagonal, and $\Gamma_B = \mathbf{I}_B \otimes \Gamma^{(r)}$

M-step

$$\gamma^{(r+1)} = \underset{\gamma:\gamma_i \in \mathbb{R}_+}{\arg\max}\, \mathbb{E}_{\mathbf{t},\mathbf{x}|\mathbf{y};\gamma^{(r)}}[\log p(\mathbf{t},\mathbf{x};\gamma)]$$

$$= \underset{\gamma:\gamma_i \in \mathbb{R}_+}{\arg\max} \left(c' - \mathbb{E}_{\mathbf{t}|\mathbf{y};\gamma^{(r)}} \mathbb{E}_{\mathbf{x}|\mathbf{t};\gamma^{(r)}} \left[\frac{\mathbf{x}^T \Gamma_B \mathbf{x}}{2} + \frac{1}{2} \log |\Gamma_B| \right] \right)$$

$$= \underset{\gamma:\gamma_i \in \mathbb{R}_+}{\arg\min} \left(c' + \frac{M}{2} \log |\Gamma| + \frac{1}{2} \sum_{m=1}^{M} \mathbb{E}_{\mathbf{t}|\mathbf{y};\gamma^{(r)}} \mathbb{E}_{\mathbf{x}|\mathbf{t};\gamma^{(r)}} \left[\mathrm{Tr} \left(\Gamma^{-1} (\mathbf{x}_m \mathbf{x}_m^T) \right) \right] \right)$$

Hence,

$$\gamma^{(r+1)} = \frac{1}{M} \left(\sum_{m=1}^{M} \mathrm{diag} \left(\Sigma_{\mathbf{x}_m} + \frac{\Sigma_{\mathbf{x}_m} \Phi_m^T [\Sigma_{\mathbf{t}_m} + \mu_{\mathbf{t}_m} \mu_{\mathbf{t}_m}^T] \Phi_m \Sigma_{\mathbf{x}_m}}{\beta_m^2 \sigma^4} \right) \right)$$

Kalman Clustered-SBL

- Model the intra-block correlation using a first order AR model:

$$\mathbf{x}_m = \rho \mathbf{x}_{m-1} + \mathbf{u}_m, \quad m = 1, \ldots, M.$$

$\mathbf{u}_m(i) \sim \mathcal{CN}(0, (1-\rho^2)\gamma(i))$, ρ: AR coefficient, $\rho \in \mathbb{R}$ and $0 \leq \rho \leq 1$.

- Assume that $\mathbf{B}_1 = \ldots = \mathbf{B}_B = \mathbf{B}$: $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{M-1} \\ \rho & 1 & \rho & \cdots & \rho^{M-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{M-1} & \rho^{M-2} & \rho^{M-3} & \cdots & 1 \end{pmatrix}$$

- State space model representation:

$$\mathbf{t}_m = \Phi_m \mathbf{x}_m + \mathbf{n}_m,$$
$$\mathbf{x}_m = \rho \mathbf{x}_{m-1} + \mathbf{u}_m, \quad m = 1, \ldots, M.$$

- Vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$ have a common sparsity profile given by $\gamma = \text{diag}(\Gamma)$
- Joint pdf of $\mathbf{t}$ and the temporally correlated vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M$ is given by

$$p(\mathbf{t}, \mathbf{x}_1, \ldots, \mathbf{x}_M; \gamma) = \prod_{m=1}^{M} p(\mathbf{t}_m | \mathbf{x}_m) p(\mathbf{x}_m | \mathbf{x}_{m-1}; \gamma),$$

where $p(\mathbf{x}_1 | \mathbf{x}_0; \gamma) \triangleq p(\mathbf{x}_1; \gamma)$.

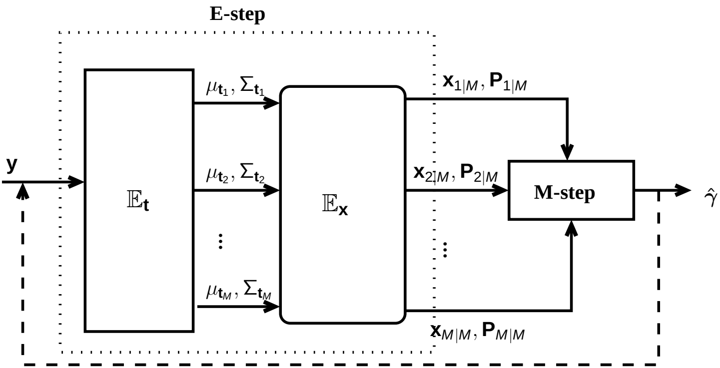
- The KC-SBL based update equations are given by

$$\text{E-step}: \mathbb{E}_{\mathbf{t}|\mathbf{y};\gamma^{(r)}}\left[\mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_M|\mathbf{t};\gamma^{(r)}}[\log p(\mathbf{y},\mathbf{t},\mathbf{x}_1,\ldots,\mathbf{x}_M;\gamma)]\right]$$

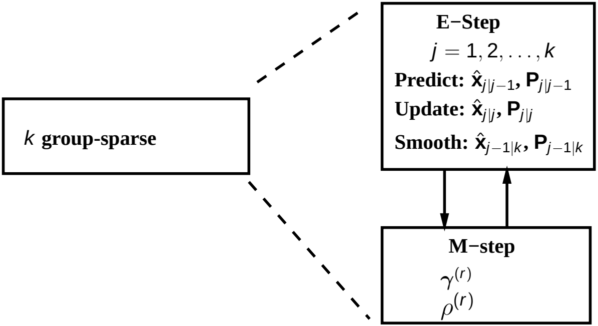
- ML estimate of γ given by

$$\gamma^{(r+1)} = \arg\max_{\gamma:\gamma_i\in\mathbb{R}_+} \mathbb{E}_{\mathbf{t}|\mathbf{y};\gamma^{(r)}}\left[\mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_M|\mathbf{t};\gamma^{(r)}}\left(c - \frac{\mathbf{x}_1^T\Gamma^{-1}\mathbf{x}_1}{2} - \frac{M}{2}\log|\Gamma| - \sum_{m=2}^{M}\frac{(\mathbf{x}_m-\rho\mathbf{x}_{m-1}))^T\Gamma^{-1}(\mathbf{x}_m-\rho\mathbf{x}_{m-1})}{2(1-\rho^2)}\right)\right]$$

- Complexity scales with M



- E-step: decomposed into $\mathbb{E}_t$ and $\mathbb{E}_x$
- $\mathbf{x}_1, \ldots \mathbf{x}_M$ are correlated $\rightarrow \mathbb{E}_x$ comprises of a recursive Kalman filter
- M-step: aggregates outputs of the $\mathbb{E}_x$ to compute an estimate of γ



- E-step: Recursive Kalman filter
- Compared to regular KF, E-step requires $\mathbf{P}_{j,j-1|k}$
- M-step aggregates outputs of each recursive step: Batch algorithm.

- The nested E and the M steps:

$$\text{E-step}: Q\left(\gamma|\gamma^{(r+\frac{k}{K})},\gamma^{(r)}\right) =$$

$$\mathbb{E}_{\mathbf{t}|\mathbf{y};\gamma^{(r)}}\left[\mathbb{E}_{\mathbf{x}_1,\ldots,\mathbf{x}_M|\mathbf{t};\gamma^{(r+\frac{k}{K})}}[\log p(\mathbf{y},\mathbf{t},\mathbf{x}_1,\ldots,\mathbf{x}_M;\gamma)]\right]$$

$$\text{M-step}: \gamma^{(r+\frac{k+1}{K})} = \underset{\gamma:\gamma_i\in\mathbb{R}_+}{\arg\max}\, Q\left(\gamma|\gamma^{(r+\frac{k}{K})},\gamma^{(r)}\right)$$

- The inner EM iteration is initialized by $\gamma^{(r+\frac{0}{K})} = \gamma^{(r)}$.
- At the end of the iterations of the inner EM loop, we set $\gamma^{(r+1)} = \gamma^{(r+\frac{K}{K})}$, effectively updating the posterior distribution of $\mathbf{t}$ in the outer EM loop.

Preliminary Result

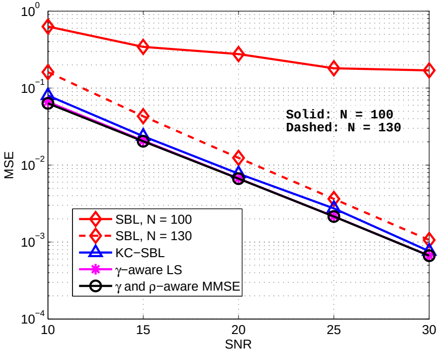


Figure: $M = 8$, $B = 32$, 5 non-zero blocks and $\rho = 0.9$.