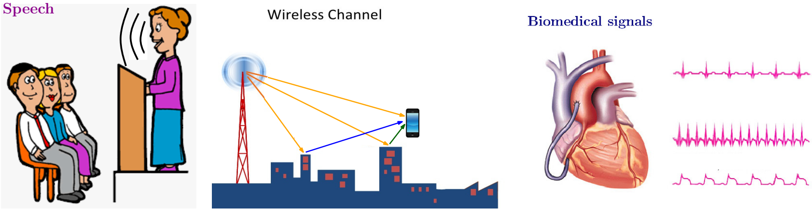
Online Sparse Bayesian Learning using Stochastic Approximation

Geethu Joseph

January 23, 2016

Motivation

- Natural signals are sparse and have significant structure

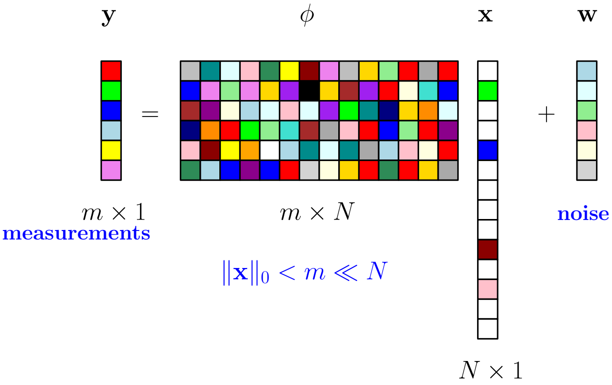


- Exploiting structure may improve recovery performance

Online Computation

- Complete input is not known in advance
- Input arrives incrementally, one piece at a time
- Improvements in
  - Memory
  - Computational complexity
  - Latency
- Disadvantage: Slightly poorer performance than offline computation

Sparse Signal Recovery Problem

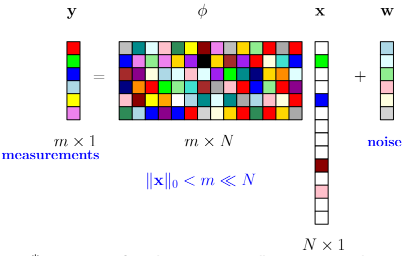


- **Goal**: Recover sparse $\boldsymbol{x}$ from $\boldsymbol{y}$

Sparse Bayesian Learning*

- Impose a fictitious sparsity inducing prior on x

$$
\begin{aligned}
x &\sim \mathcal{N}(0, \Gamma) \\
\Gamma &= \text{diag}\{\gamma_1, \gamma_2, \ldots, \gamma_N\} \\
y|x &\sim \mathcal{N}(\phi x, \sigma^2 I)
\end{aligned}
$$



E-step: $Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{(r)}) = E_{\mathbf{x}|\mathbf{y};\boldsymbol{\Gamma}^{(r)}} \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\Gamma})$

Iterate

M Step: $\boldsymbol{\Gamma}^{(r+1)} = \arg\max_{\boldsymbol{\Gamma}} Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{(r)})$

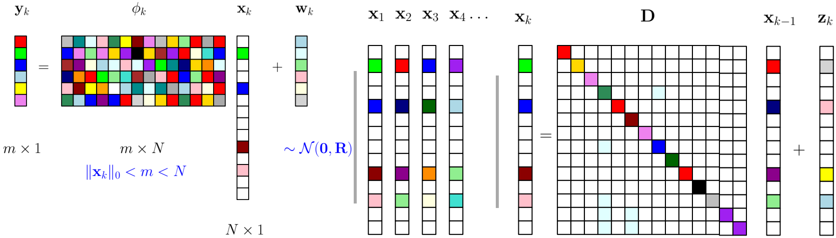
$\hat{\boldsymbol{\Gamma}}$

$\hat{\mathbf{x}} = E(\mathbf{x}|\mathbf{y}; \hat{\boldsymbol{\Gamma}})$

*D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," TSP 2007

System Model

- Multiple measurement model



- Temporally correlated sparse vectors share same support
- **Goal**: Online recovery the sparse vectors with a maximum time lag of Δ between measurement and estimation

SBL: Correlated MMV

- Impose a sparsity inducing prior on

$$
\begin{aligned}
\boldsymbol{x}_k &\sim \mathcal{N}(0, \boldsymbol{\Gamma}) \\
\boldsymbol{\Gamma} &= \text{diag}\{\gamma\}
\end{aligned}
$$

Problem: At time k, given $\boldsymbol{y}_{1:k}$, estimate $\boldsymbol{x}_{k-\Delta}$

Offline SBL Approach†

- Iterating the E and M steps yield the MLE of γ

$$\begin{aligned}\textbf{E-step: } \mathcal{Q}(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{(r)}) &= \mathbb{E}_{\boldsymbol{x}_{1:K}|\boldsymbol{y}_{1:K};\boldsymbol{\Gamma}^{(r)}}\left\{\log\left[p\left(\boldsymbol{y}_{1:K},\boldsymbol{x}_{1:K}\right)\right]\right\} \\ \textbf{M-step: } \boldsymbol{\Gamma}^{(r+1)} &= \arg\max_{\boldsymbol{\Gamma}} \mathcal{Q}(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{(r)})\end{aligned}$$

- **M-step:** Computes $\boldsymbol{\Gamma}^{(r+1)}$ as a closed form function of the state statistics:
  - Mean: $\hat{\boldsymbol{x}}_{t|K} = \mathbb{E}\left\{\boldsymbol{x}_t|\boldsymbol{y}_{1:K}\right\}$
  - Autocovariance: $\boldsymbol{P}_{t|K} = \text{cov}\left\{\boldsymbol{x}_t, \boldsymbol{x}_t|\boldsymbol{y}_{1:K}\right\}$
  - Cross-covariance: $\boldsymbol{P}_{t,t-1|K} = \text{cov}\left\{\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:K}\right\}$, $t \in [K]$
- **E-step**: Computes state statistics using $\boldsymbol{\Gamma}^{(r)}$

†R. Prasad, et al., "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning" TSP 2014

Online Version

Approximation:

$$\begin{aligned} \hat{\boldsymbol{x}}_{t|K} &\rightarrow \hat{\boldsymbol{x}}_{t|t+\Delta} \\ \boldsymbol{P}_{t|K} &\rightarrow \boldsymbol{P}_{t|t+\Delta} \\ \boldsymbol{P}_{t,t-1|K} &\rightarrow \boldsymbol{P}_{t,t-1|t+\Delta} \end{aligned}$$

Online Recursions:

$$\gamma_k = \gamma_{k-1} + \frac{1}{k}\text{Diag}\left\{\frac{1}{(1-\rho_i^2)}\boldsymbol{T}_{k|k+\Delta} - \boldsymbol{\Gamma}_{k-1}\right\}$$

$\boldsymbol{T}_{t|k}$: function of $\hat{\boldsymbol{x}}_{t|K}$, $\boldsymbol{P}_{t|K}$, and $\boldsymbol{P}_{t,t-1|K}$

Implementation of Algorithm

- New state space model:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{z}_k$$
$$\boldsymbol{y}_{k+\Delta} = \tilde{\boldsymbol{A}}_k \boldsymbol{x}_k + \tilde{\boldsymbol{w}}_{k+\Delta}.$$

- Modified parameters:
  - $\tilde{\boldsymbol{A}}_k = \boldsymbol{A}_{k+\Delta} \boldsymbol{D}^{\Delta}$
  - $\tilde{\boldsymbol{w}}_k \sim \mathcal{N}\left(0, \tilde{\boldsymbol{R}}_k\right)$
  - $\tilde{\boldsymbol{R}}_k = \boldsymbol{A}_{k+\Delta}\left(\boldsymbol{I} - \boldsymbol{D}^{2\Delta}\right)\boldsymbol{\Gamma}\boldsymbol{A}_{k+\Delta}^{\mathsf{T}} + \boldsymbol{R}_k$

Implementation: Kalman filter for new system

Convergence Analysis: Special Case

- Assumptions:
  1. The sparse vectors are uncorrelated: $\boldsymbol{D} = \mathbf{0}$
  2. All measurement matrices are identical: $\boldsymbol{A}_k = \boldsymbol{A}, \forall k$
  3. All measurement noise statistics are identical: $\boldsymbol{R}_k = \boldsymbol{R}$, $\forall k$
- Simplified Algorithm:

$$\gamma_k = \gamma_{k-1} + \frac{1}{k}\text{Diag}\left\{\boldsymbol{P}(\gamma_k) + \hat{\boldsymbol{x}}(\gamma_k)\hat{\boldsymbol{x}}(\gamma_k)^{\mathsf{T}} - \boldsymbol{\Gamma}_k\right\}$$

$$\boldsymbol{P}(\gamma) = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{T}}\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{T}} + \boldsymbol{R}\right)^{-1}\boldsymbol{A}\boldsymbol{\Gamma}$$

$$\hat{\boldsymbol{x}}(\gamma) = \boldsymbol{P}(\gamma)\boldsymbol{A}^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{y}_k$$

Stochastic Approximation Framework

Algorithm as stochastic approximation recursion:

$$\gamma_k = \gamma_{k-1} + \frac{1}{k}\boldsymbol{f}(\gamma_{k-1}) + \frac{1}{k}\boldsymbol{e}_k$$

- Mean field function:

$$\boldsymbol{f}(\gamma) = \mathbb{E}\left\{\text{Diag}\left\{\boldsymbol{P}(\gamma) + \hat{\boldsymbol{x}}(\gamma)\hat{\boldsymbol{x}}(\gamma)^{\mathsf{T}} - \boldsymbol{\Gamma}\right\}\right\}$$

- Martingale difference sequence:

$$\boldsymbol{e}_k = \text{Diag}\left\{\boldsymbol{P}(\gamma_k) + \hat{\boldsymbol{x}}(\gamma_k)\hat{\boldsymbol{x}}(\gamma_k)^{\mathsf{T}} - \boldsymbol{\Gamma}_k\right\} - \boldsymbol{f}(\gamma_{k-1})$$

Convergence Results

Algorithm converges to a closed set $\mathbb{G}$ if there exists a nonnegative $\mathcal{C}^1$ function V such that

(i) $\boldsymbol{f}$ is continuous and bounded on all compact sets of its domain

(ii) $\langle \nabla_{\boldsymbol{\gamma}} V(\boldsymbol{\gamma}), \boldsymbol{f}(\boldsymbol{\gamma}) \rangle < 0$, $\forall \boldsymbol{\gamma} \notin \mathbb{G}$

(iii) $V(\mathbb{G})$ is a finite set

(iv) $\boldsymbol{\gamma}_k$ remains in a compact subset of $\mathbb{R}^N$

(v) $\lim_{k\to\infty} \sum_{t=1}^{k} \frac{1}{t} \boldsymbol{e}_t$ exists

Our Choices

- Closed Set:

$$\mathbb{G} = \{\gamma \in \mathbb{O} : \boldsymbol{f}(\gamma) = \mathbf{0}\}$$
$$=$$

- Potential function:

$$V(\gamma) = \mathrm{Tr}\left\{\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{T}} + \boldsymbol{R}\right)^{-1}\left(\boldsymbol{A}\boldsymbol{\Gamma}_{\mathrm{opt}}\boldsymbol{A}^{\mathsf{T}} + \boldsymbol{R}\right)\right\} - \log\left|\left(\boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{T}} + \boldsymbol{R}\right)^{-1}\left(\boldsymbol{A}\boldsymbol{\Gamma}_{\mathrm{opt}}\boldsymbol{A}^{\mathsf{T}} + \boldsymbol{R}\right)\right|$$

Main Theorem

Assumption: The nonzero entries of $\boldsymbol{x}$ are orthogonal

Result: With probability one,

$$\gamma_k \to \{\mathbf{0}\} \cup \left\{\gamma \in \mathbb{R}_+^N : \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^\mathsf{T} = \boldsymbol{A}\mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\right\}\boldsymbol{A}^\mathsf{T}\right\}$$

Corollary

Assumption:

- The nonzero entries of $\boldsymbol{x}$ are orthogonal
- Rank $\{\boldsymbol{A} \odot \boldsymbol{A}\} = N$

Result: With probability one,

$$\gamma_k[i] \to \left\{\mathbf{0}, \mathbb{E}\boldsymbol{x}[i]^2\right\}$$

Summary

- Proposed a stochastic approximation recursion for online recovery of temporally correlated sparse vectors
- Temporal correlation is modeled using first order AR process
- Algorithm is implemented using Kalman filter