Journal Watch:

IEEE Transanctions on Signal Processing March 2014

Saurabh Khanna,
Signal Processing for Communication, ECE, IISc

Sparse Signal Estimation by Maximally Sparse Convex Optimization.

Authors: Ivan W. Selesnick and Ilker Bayram

- **Problem Statement:**
  - $$\arg\min_{\mathbf{x}\in\mathbb{R}^n} \{F(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \sum_{n=0}^{N-1} \lambda_n \phi_n(\mathbf{x}_n)\}$$
  - To find "non convex" penalties ϕ_n which induce sparsity more strongly than ℓ_1 norm regularization such that overall objective function $F(.)$ is convex.
- **Main Idea:**
  - Balance the +ve second derivative of fit error term against -ve second derivative of penalty terms.
  - Select $\phi_n(x)$ to be parameterized functions $\phi_n(x; a_n)$, with parameter a_n.
  - Candidates for ϕ_n:
    - logarithm penalty: $\frac{1}{a}\log(1 + a|x|)$
    - arc-tangent penalty: $\frac{2}{a\sqrt{3}}\left(\tan^{-1}\left(\frac{1+2a|x|}{\sqrt{3}}\right)\right)$

- **How to find a_n**
  - Rewrite $F(\mathbf{x})$ as

$$F(\mathbf{x}) = \left(\frac{1}{2}\mathbf{x}^T(\mathbf{H}^T\mathbf{H} - \mathbf{R})\mathbf{x} - \mathbf{y}^T\mathbf{H}\mathbf{x} + \frac{1}{2}\mathbf{y}^T\mathbf{y}\right) + \left(\frac{1}{2}\mathbf{x}^T\mathbf{R}\mathbf{x} + \sum_n \lambda_n\phi_n(\mathbf{x}_n; n)\right)$$

    where $\mathbf{R}$ is a +ve definite **diagonal** matrix
  - For ϕ_n being logarithmic penalty, the second term is convex if

$$0 < a_n < \frac{r_n}{\lambda_n} \tag{1}$$

    where $r_n = \mathbf{R}(n, n)$. So, we need to find $\mathbf{R}$ which satisfies (1).
- **How to find R**
  - Larger r_n synonymous with more non-convex ϕ_n and therefore more sparsity inducing.
  - **R** found via optimization problem:

$$\arg\max_{r_1, r_2 \ldots r_n} \sum_{n=0}^{N-1} r_n$$

such that $r_n \geq \lambda_{min}(\mathbf{H}^T\mathbf{H})$ and $\mathbf{H}^T\mathbf{H} - \mathbf{R} \geq 0$

Near Optimal Sensor Placement for Linear Inverse Problems.

Authors: Juri Raniieri and Amina Chebira and Martin Vetterli

- **Problem Statement:**
  - Linear inverse problem: Find $\mathbf{x}$ from $\mathbf{y} = \Psi\mathbf{x}$.
  - There are N measurements from different sensors, $\mathbf{y} \in \mathbb{R}^N$.
  - Which are best L meaurements out of N total meaurements? (MSE wise)
  - Combinatorial complexity.
- **Contributions:**
  - Fast greedy algorithm to pick best L measurements.
  - Near optimality of greedy algorithm shown. (optimization in terms of MSE)
- **Main Idea:** Use Frame potential(FP) used as cost function.
- **What is frame potential**
  - For $G \subset (1, 2 \dots N)$ and $|G| = L$, $FP(\Psi_G) = \sum_{i,j \in G} | < \psi_i, \psi_j > |^2$ where ψ_i and ψ_j are rows of submatrix Ψ_G.
  - FP is a measure of orthogonality of rows. (Lower FP $\Rightarrow$ tighter frame).
- **Why frame potential?**
  - MSE based cost functions suffer from local minimas.
  - It is shown that $FP(\Psi_G) \rightarrow FP_{UNTF}$ implies $MSE(\Psi_G) \rightarrow MSE_{UNTF}$.
  - FP is easy to compute.
  - FP is shown to be submodular in G and greedy algorithms are known to be optimal in optimization of submodular functions.
- **FrameSense - a greedy algorithm to pick best L out of N measurements**
  - In each iteration, remove the row that maximizes $FP(\Psi_S) = FP(\Psi) - FP(\Psi_{N \setminus S})$.
  - S = set of unwanted rows/measurements and $N = (1, 2, \dots N)$

Distributed Sparse Recursive Least-Squares Over Networks.

Authors: Zhaoting Liu, Ying Liu and Chunguang Li

- **Problem Statement:**
  - Distributed online learning of sparse vector
  - Measurement model at time instant i.

$$d_{n,i} = \mathbf{u}_{n,i}\mathbf{w} + \eta_{n,i}$$

  - $\mathbf{w} \in \mathbb{R}^M$ is sparse, $d_{n,i}$ is measurement taken at node n $\eta_{n,i} \sim \mathcal{N}(0, \sigma_n^2)$.
- **Contributions:** Distributed RLS algorithm for learning $\mathbf{w}$.
- **Main ideas:**
  - Each node n minimizes local cost function

$$\psi_{n,i} = \arg\max_{\mathbf{w}} \sum_{l \in N_n} \sum_{j=1}^{i} \mu^{i-j} \log p(d_{l,i}/\mathbf{w}) - \gamma J(\mathbf{w}) \qquad (2)$$

  - $J(\mathbf{w})$ is sparsity inducing penalty term.
  - Local cost function (2) simplifies to:

$$\psi_{n,i} = \arg\min_{\mathbf{w}} \sum_{l \in N_n} \frac{(\mathbf{d}_{l,i} - \mathbf{U}_{l,i}\mathbf{w})^T \Lambda_i (\mathbf{d}_{l,i} - \mathbf{U}_{l,i}\mathbf{w})}{2\sigma_l^2} + \gamma J(\mathbf{w})$$

    where $\mathbf{d}_{l,i} = (d_{n,1}, \ldots d_{n,i})$ and $\mathbf{U}_{l,i} = \mathrm{col}(\mathbf{u}_{n,1}, \ldots \mathbf{u}_{n,i})$ and $\Lambda_i = \mathrm{diag}(\mu_{i-1}, \mu_{i-2} \ldots 1)$
  - As i increases, dimensions of $\mathbf{d}_{l,i}, \mathbf{U}_{l,i}, \Lambda_i$ increase. So RLS type algorithm is needed!

- **Main ideas (contd):**
- Combined measurement model for i time instances at node n:

$$\mathbf{d}_{n,i} = \mathbf{U}_{n,i}\mathbf{w} + \boldsymbol{\xi}_{n,i} \tag{3}$$

- Trick: Decompose noise vector $\boldsymbol{\xi}_{n,i}$ into two parts:

$$\boldsymbol{\xi}_{n,i} = \alpha_n \mathbf{U}_{n,i}\boldsymbol{\mu}_{n,i} + \Lambda_i^{\frac{1}{2}}\boldsymbol{\nu}_{n,i} \tag{4}$$

where $\nu_{n,i} \sim \mathcal{N}(0, \sigma_n^2\mathbf{I} - \alpha_n^2\Lambda_i^{\frac{1}{2}}\mathbf{U}_{n,i}\mathbf{U}_{n,i}^T\Lambda_i^{\frac{1}{2}})$ and $\mu_{n,i} = \mathcal{N}(0, \mathbf{I})$.

- Using (3) and (4), we can write

$$\mathbf{z}_{n,i} = \mathbf{w} + \alpha_n\boldsymbol{\mu}_{n,i} \tag{5}$$

$$\mathbf{d}_{n,i} = \mathbf{U}_{n,i}\mathbf{z}_{n,i} + \Lambda_i^{\frac{1}{2}}\boldsymbol{\nu}_{n,i} \tag{6}$$

- Using (5) and (6), we can now formulate EM algorithm !
- E-step:

$$Q(\mathbf{w}, \hat{\mathbf{w}}_{n,i-1}) = E_z[\log p(\mathbf{d}_{l,i}, \mathbf{z}_{l,i}/\mathbf{w})|\mathbf{d}_{l,i}, \hat{\mathbf{w}}_{n,i-1}].$$

- modified M-step:

$$\arg\max_{\mathbf{w}} \sum_{l\in N_n} Q_l(\mathbf{w}, \hat{\mathbf{w}}_{n,i-1}) - \gamma J(\mathbf{w}).$$

Decentralized Dynamic Optimization Through the Alternating Direction Method of Multipliers

Authors: Qing Ling and Alejandro Ribeiro

- **Problem Statement:**
  - Dynamic optimization problem with separable cost function @ time instant k:
$$\min_{\mathbf{x}} \sum_{i=1}^{L} f_i^k(\mathbf{x})$$
    where $\mathbf{x} \in \mathbb{R}^p$.
  - Can be formulated as a decentralized consensus optimization problem, given by
$$\min_{\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_L} \sum_{i=1}^{n} f_i^k(\mathbf{x}_i) \qquad \text{such that } \mathbf{x}_i = \mathbf{x}_j \forall i, j \in (1, 2 \ldots L)$$
  - Need a decentralized algorithm whose iterations consider same time scale as the evolution of the functions $f_i^k(.)$.
- **Main Results:**
  - Decentralized ADMM suggested for above problem (run single ADMM iteration in each time instant).
  - In steady state, bounds on tracking error provided.
  - Steady state tracking error and decay of primal gap before reaching steady state depends on:
    1. Condition number of underlying graph
    2. Condition number of objective function of primal problem
    3. $\max_k ||\mathbf{x}_k^* - \mathbf{x}_{k-1}^*||_2$
    4. $\max_k ||\nabla f_k(\mathbf{x}_k^*) - \nabla f_{k-1}(\mathbf{x}_{k-1}^*)||_2$

Interesting papers...

- ℓ_q Sparsity penalized Linear Regression With Cyclic Descent
- Estimation for Linear Model With Uncertain Covariance Matrices
- On Kronecker and Linearly Structured Covariance Matrix Estimation
- Detection of Spatially Correlated Time Series From a Network of Sensor Arrays
- A Factor Graph Approach to Joint OFDM Channel Estimation and Decoding in Impulsive Noise Environments
- Joint Power and Antenna Selection Optimization in Large Cloud Radio Access Networks

Thank You !!!