Hierarchically Sparse MIMO Channel Estimation using L-SBL

Unnikrishnan N

Indian Institute of Science

unnikrishnann@iisc.ac.in

November 19, 2019

Overview

Hierarchical Sparse Model

Model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$$
where $\mathbf{y}, \mathbf{z} \in \mathbb{R}^M$, $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{A} \in \mathbb{R}^{M \times N}$

Problem Statement

given $\mathbf{y}$, estimate $\mathbf{x} \ni ||\mathbf{x}||_0 \leq K$ and $K < M \ll N$

Hierarchical Model - Definition

Let $\mathbf{s} = \{s_1, s_2, \ldots, s_l\}$ be an l-tuple of natural numbers and consider an l-level block vector $\tilde{x} \in \mathbb{C}^{N_1 \cdot N_2 \cdots N_l}$ with $N_i \geq s_i, i \in [l]$. $\tilde{x}$ is **s**-hierarchically sparse (s-Hi-sparse) if it has the property of hierarchical s-sparsity defined inductively as follows:

- For $l = 1, \tilde{x}$ is s-Hi-sparse if atmost s_1 of its N_1 elements are non-zero.
- For $l > 1$, $\tilde{x}$ is s-Hi-sparse if it consists of N_1 blocks and atmost s_1 of these are non-zero with each non-zero block being $s_2, \ldots, s_l$-Hi-sparse.

Hierarchical Sparsity Illustration

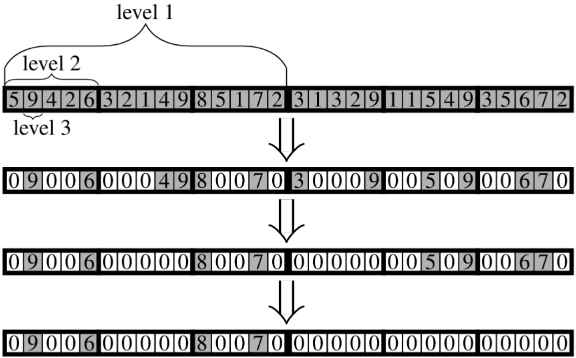

Figure: Illustration of the sequence of actions of Thresholding operator $T_{(1,2,2)}()$ on a three-level block vector in $C^{2\cdot 3\cdot 5}$(2 blocks of 3 blocks of 5 elements each).

Thresholding Based Sparse Recovery Algorithms

Motivation for Iterative Thresholding Methods

A weakness of the OMP algorithm is that, once an incorrect index has been selected in a target support S^n ,it remains in all the subsequent target supports $S^{n'}$ for $n' \geq n$. Hence, if an incorrect index has been selected, s iterations of the orthogonal matching pursuit are not enough to recover a vector with sparsity s.

Iterative Thresholding Algorithms

- IHT(Iterative Hard Thresholding)
- HTP(Hard Thresholding Pursuit)

Thresholding Based Sparse Recovery Algorithms

Motivation for Iterative Methods

From classical iteration techniques
for $y = \mathbf{A}x$ and $\mathbf{A}$ is square, then

$$x = (\mathbf{I} - \mathbf{A})x + y$$
$$x^{n+1} = (\mathbf{I} - \mathbf{A})x^n + y$$

If $\mathbf{A}$ is fat then modify $y = \mathbf{A}x$ problem to $\mathbf{A}^*y = \mathbf{A}^*\mathbf{A}x$.

$$x^{n+1} = (\mathbf{I} - \mathbf{A}^*\mathbf{A})x^n + \mathbf{A}^*y$$
$$x^{n+1} = x^n + \mathbf{A}^*(y - \mathbf{A}x^n)$$

for convergence $||\mathbf{I} - \mathbf{A}||_2 < 1$ in the first case and $||\mathbf{I} - \mathbf{A}^*\mathbf{A}||_2 < 1$ in the second case. So problem modified to

$$x^{n+1} = x^n + \mu\mathbf{A}^*(y - \mathbf{A}x^n)$$

Algorithm -IHT

1. *input* : measurement matrix $\mathbf{A}$, measurement vector y, sparsity k
2. *initialization* : k-sparse vector $\mathbf{x}^0$, typically $x^0 = \mathbf{0}$
3. *iteration* : repeat until a stopping criteria is met at $n = \bar{n}$:

   $$\mathbf{x}^{n+1} = H_k(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))$$

   where H_k is hard thresholding operator that keeps k largest absolute entries and sets the other ones to zero
4. *output* : the k-sparse vector $\mathbf{x}^* = \mathbf{x}^{\bar{n}}$

Algorithm -HTP

1. *input* : measurement matrix $\mathbf{A}$, measurement vector $\mathbf{y}$, sparsity k
2. *initialization* : k-sparse vector $\mathbf{x}^0$, typically $\mathbf{x}^0 = \mathbf{0}$
3. *iteration* : repeat until a stopping criteria is met at $n = \bar{n}$:

$$S^{n+1} = L_k(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))$$
$$\mathbf{x}^{n+1} = \underset{\mathbf{z}\in\mathbb{R}^N}{\operatorname{argmin}} \left\{\|\mathbf{y} - \mathbf{A}z\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^{n+1}\right\}$$

   where L_k is the index operator that returns set of first k absolute largest entries of a vector and S^n is the index set optained in n^{th} iteration

4. *output* : the k-sparse vector $\mathbf{x}^* = \mathbf{x}^{\bar{n}}$

Iterative Sparse Recovery Algorithms - Assuming Hierarchical Structure

Modified Iterative Algorithms

- HiIHT(Hierarchical Iterative Hard Thresholding)
- HiHTP(Hierarchical Hard Thresholding Pursuit)

Algorithm-HiIHT and HiHTP

1. *input* : $\mathbf{A}$, $\mathbf{y}$, $\mathbf{s}$-Hi sparse vector. where $\mathbf{s} = (s_1, s_2, \ldots s_l)$
2. *initialization* : $\mathbf{s}$-Hi sparse vector $\mathbf{x}^0$, typically $\mathbf{x}^0 = \mathbf{0}$
3. *iteration* : repeat until a stopping criteria is met at $n = \bar{n}$:

   $\mathbf{x}_{temp} = \mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)$

   $S^{n+1} = T_{\mathbf{s}}(\mathbf{x}^n + A^*(\mathbf{y} - A\mathbf{x}^n))$

   if **HiIHT**

   $\mathbf{x}^{n+1} = \mathbf{0} \in \mathbb{R}^N$

   $\mathbf{x}^{n+1}_{S^{n+1}} = \mathbf{x}_{temp,S^{n+1}}$

   else if **HiHTP**

   $\mathbf{x}^{n+1} = \underset{\mathbf{z} \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^{n+1} \right\}$

   where $T_{\mathbf{s}}$ is Thresholding operator for s-Hi sparse vector
4. *output* : the $\mathbf{s}$-Hi sparse vector $\mathbf{x}^* = \mathbf{x}^{\bar{n}}$

Algorithm - Thresholding Operator

1. *input* : $\tilde{\mathbf{x}} \in \mathbb{C}^{N_1 \cdot N_2 \cdots N_l}$, $\mathbf{s} = (s_1, s_2, \ldots, s_l)$, $l \geq 2$.
2. *initialization* : $\mathbf{z} = \mathbf{x}$
3. For each of the $N_1 \cdot N_2 \cdots N_{l-1}$ blocks at level $l-1$ of $\mathbf{z}$, identify the s_l (out of a total N_l) largest-modulus elements and set the remaining elements equal to zero.
4. *iteration* : stops when $k = 0$, where $k = l - 2$
   while($k \geq 1$)
   For each of the $N_1 \cdot N_2 \cdots N_k$ blocks at level k of $\mathbf{z}$, identify the s_{k+1} (out of a total N_{k+1}) blocks with the largest Euclidean norm and set the elements of the remaining blocks equal to zero.
   $k = k - 1$
   end while
5. *output* : the $\mathbf{s}$-Hi sparse vector $\mathbf{z}$

Sequence of Operation of Thresholding Operator

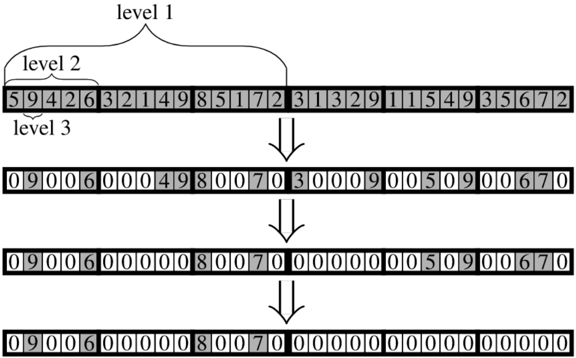

Figure: Illustration of the sequence of actions of Thresholding operator $T_{(1,2,2)}()$ on a three-level block vector in $C^{2\cdot3\cdot5}$(2 blocks of 3 blocks of 5 elements each).

MIMO Channel Model

Consider an uplink channel of a single cell with a BS equipped with $M \gg 1$ antenna elements serving multiple single antenna UEs. Array manifold for a ULA is

$$\mathbf{a}(\theta) \triangleq [1, e^{-j2\pi\theta} \ldots, e^{-j2\pi(M-1)\theta}]^T$$

where $\theta = d \sin\phi$, d is normalized spatial separation and $\phi \in [-\pi/2, \pi/2]$
Consider a steering dictionary (sampled version)

$$\mathbf{A}_\theta \triangleq [\mathbf{a}(0), \mathbf{a}(1/M) \ldots, \mathbf{a}((M-1)/M)] = \mathbf{F}_{M,M} \in \mathbb{C}^{M \times M}$$

Consider transmissions are performed via wideband OFDM signals with $N \gg 1$ subcarriers centered at the baseband frequencies $\{2\pi k/N\}_{k=0}^{N-1}$, with $T_s > 0$ OFDM symbol duration. The delay manifold

$$\mathbf{b}(\tau) \triangleq [1, e^{-j2\pi\tau/T_s}, \ldots, e^{-j2\pi(N-1)\tau/T_s}]^T$$

Consider sampled version of delay as $\{kT_s/N\}_{k=0}^{N-1}$

$$\mathbf{A}_\tau \triangleq [\mathbf{b}(0), \mathbf{b}(T_s/N) \ldots, \mathbf{b}((D-1)T_s/N)] = \mathbf{F}_{N,D} \in \mathbb{C}^{N \times D}$$

where $D = \alpha N, \alpha < 1$ corresponds to maximum delay spread

MIMO Channel Model

Channel of an arbitrary UE is a superposition of a small number L of impinging wavefronts characterized by delay,angle pairs $\{\tau_p, \theta_p\}_{p=0}^{L-1}$ with $\tau_p \in [0, \alpha T_s]$ and $\theta_p \in [0, 1]$. The channel transfer matrix is given by

$$\mathbf{H} = \sum_{p=0}^{L-1} \rho_p \mathbf{b}\left(\tau_p\right) \mathbf{a}^H\left(\theta_p\right)$$

where $\rho_p \in \mathbb{C}$ is the complex gain of the p-th path. Here L is known and is independent of system parameters M and N. Based on sampled version of array manifold and delay manifold the above equation can be written as

$$\mathbf{H} = \mathbf{A}_\tau \mathbf{X} \mathbf{A}_\theta^H$$

$$\mathbf{X} \triangleq \sum_{p=0}^{L-1} \rho_p \mathbf{e}_{k_p, D} \mathbf{e}_{l_p, M}^T \in \mathbb{C}^{D \times M}$$

$$(\tau_p, \theta_p) = (k_p T_s / N, l_p / M)$$

where $\mathbf{e}_{n,N} \in \mathbb{C}^N$ denoting the canonical basis with nth element is 1.

Multiuser Channel Estimation problem

BS partitions the uplink UEs to groups of U UEs. Each group is assigned exclusive set of pilot subcarriers. Consider a single group with $V \leq U$ active UEs and $\mathcal{N}_p \subset [N]$ subcarriers, $\mathcal{M}_p \subset [M]$ antennas at BS. Observation at the BS during pilot transmission is ($\mathbf{c}_u$ is pilot signature)

$$\mathbf{Y} = \sum_{u=0}^{U-1} \mathrm{diag}\left(\mathbf{c}_u\right) \mathbf{P}_{\mathcal{N}_p} \mathbf{H}_u \mathbf{P}_{\mathcal{M}_p}^T + \mathbf{Z} \in \mathbb{C}^{N_p \times M_p}$$

where $c_u \in \mathbb{C}^{N_p}$, $H_u \in \mathbb{C}^{N \times M}$, $\mathbf{P}_{\mathcal{N}_p} \in I_{N,N_p}$ and $\mathbf{P}_{\mathcal{M}_p} \in I_{M,M_p}$ can be written in terms of delay angle combination as

$$\mathbf{Y} = \bar{\mathbf{A}}_\tau \bar{\mathbf{X}} \bar{\mathbf{A}}_\theta^H + \mathbf{Z} \in \mathbb{C}^{N_p \times M_p}$$

$$\bar{\mathbf{A}}_\tau \triangleq \frac{1}{\sqrt{N_p}} \left[\mathrm{diag}\left(\mathbf{c}_0\right) \mathbf{P}_{\mathcal{N}_p} \mathbf{A}_\tau, \ldots, \mathrm{diag}\left(\mathbf{c}_{U-1}\right) \mathbf{P}_{\mathcal{N}_p} \mathbf{A}_\tau\right]$$

$$\bar{\mathbf{A}}_\theta \triangleq \frac{1}{\sqrt{M_p}} \mathbf{P}_{\mathcal{M}_p} \mathbf{A}_\theta$$

$$\bar{\mathbf{X}} \triangleq \left[\mathbf{X}_0^T, \mathbf{X}_1^T, \ldots, \mathbf{X}_{U-1}^T\right]^T$$

Problem Formulation to CS Framework

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z} \in \mathbb{C}^{N_p M_p}$$
$$\mathbf{y} = vec(\mathbf{Y})$$
$$\mathbf{A} \triangleq \bar{\mathbf{A}}_\theta^* \otimes \bar{\mathbf{A}}_\tau$$
$$\mathbf{x} \triangleq \text{vec}(\bar{\mathbf{X}})$$

How hierarchy comes

$\mathbf{x} \in \mathbb{C}^{M \cdot U \cdot D}$
out of M angles at max LV number of angles are present and $LV \leq M$ (M being large)
At max at every UD, V active UEs are present.
At max at every D, L delays can be present.

Design Criteria

Reduced Model

1. Chosen a few antennas $M_p < M$ for reception
2. Chosen a few subcarries for each group $N_p < N$
3. Chosen random subset of antennas $\mathcal{M}_p \subset [M]$
4. Chosen random subset of subcarriers $\mathcal{N}_p \subset [N]$

First two points are for complexity reduction and third and fourth points are for satisfying CS(compressed sensing) recovery conditions.

DNN(Deep Neural Network) based Solution

Modifications

1. Measurement matrix A generation for lower mutual coherence.
2. Reformulated the problem from complex to real for training
3. Hierarchical sparse vector generations
4. Used L-SBL (with 11 layers) to learn the recovery from measurements.

L-SBL

1. Its a unfolded version of SBL(sparse bayesion learning) algorithm.

Simulation Results

Measurement matrix **A** as random 30×50 and no structure in sparsity.

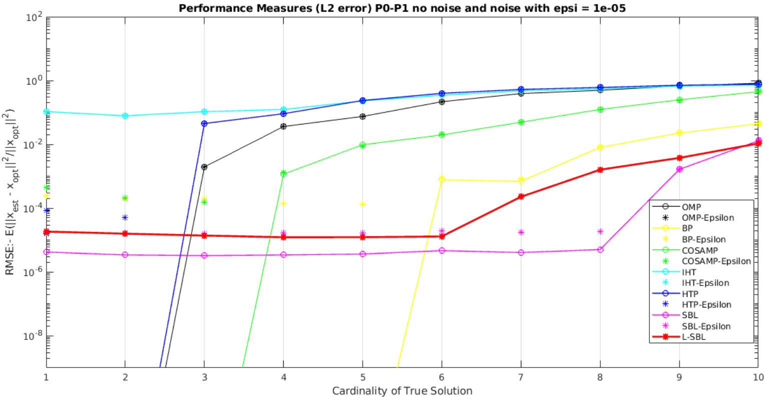

Figure: RMSE of sparse recovery using multiple algorithms with random matrix (30x50)

Simulation Results

Measurement matrix **A** as random 30×50 and (1,5,5) Hi-sparse structure is used.

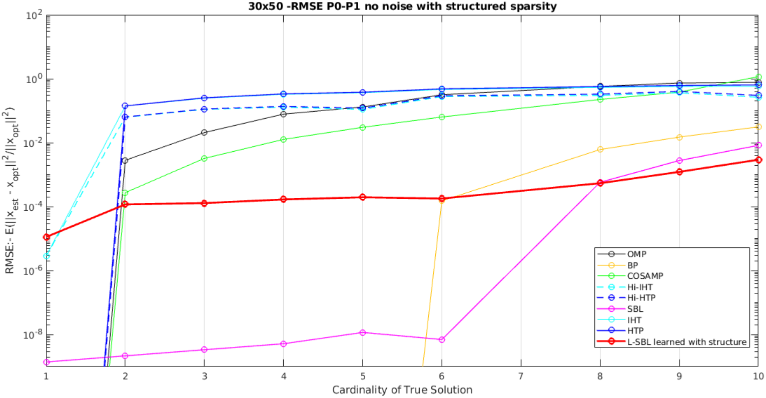

Figure: RMSE of sparse recovery using multiple algorithms with random matrix

Conclusion And Future Scope

Conclusion

1. Various iterative thresholding algorithms were discussed.
2. Hierarchical sparse recovery based on HiIHT/HiHTP were discussed.
3. s-Hi sparse vector recovery using L-SBL is tried out for lower dimensions.
4. Recovery performance of L-SBL is better than *HiIHT/HiHTP*.

Future Scope

1. Extend to higher dimensions

References

Elad, Michael,"Sparse and Redundant Representations - From Theory to Applications in Signal and Image Processing", 2010

S. Foucart and H. Rauhut,A "Mathematical Introduction to CompressiveSensing", Birkhuser, 2013.

G. Wunder, S. Stefanatos, A. Flinth, I. Roth and G. Caire, "Low-Overhead Hierarchically-Sparse Channel Estimation for Multiuser Wideband Massive MIMO," in IEEE Transactions on Wireless Communications, vol. 18, no. 4, pp. 2186-2199, April 2019.

Thank You