Message Passing Algorithm for Joint Support and Signal Recovery of Approximately Sparse Signals

Shubha Shedthikere

July 16, 2012

1. Overview of Sparse Signal Recovery
   - Problem Formulation
   - Review of existing methods
2. JSSR-MP
   - From strictly sparse to approximately sparse
   - Bayesian inference problem
   - Message Passing Algorithm
   - Joint Support and Signal Recovery using Message Passing (JSSR-MP)
   - Discussion on computational complexity
   - Results
3. Future work

Linear inverse problem

The objective of an inverse problem is to find the best model, $\mathbf{x}$, such that (at least approximately)

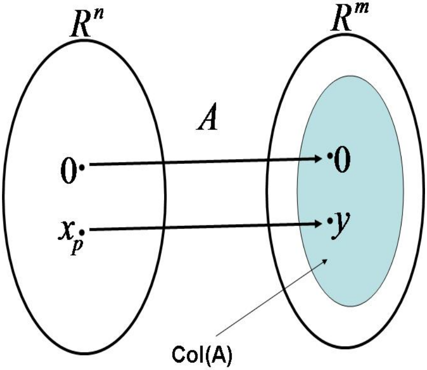
$$\mathbf{y} = A(\mathbf{x}) \tag{1}$$

where A is an operator describing the explicit relationship between the observed data, $\mathbf{y}$, and the model parameters.
If A is a linear mapping, we have, the linear inverse problem of finding $\mathbf{x}$ such that

$$\mathbf{y} = A\mathbf{x} \tag{2}$$

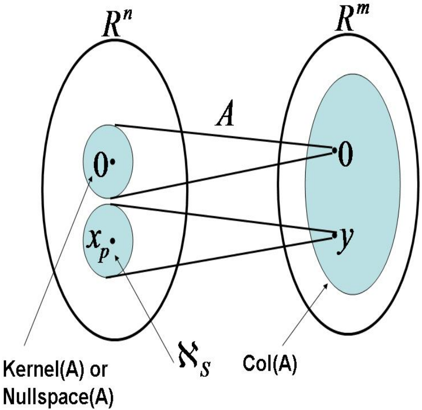
Linear inverse problem

Case 1: When A is non-singular



$\aleph_S = \{\mathbf{x} \in \mathbf{R} : A\mathbf{x} = \mathbf{y}\} = \mathbf{x}_p$

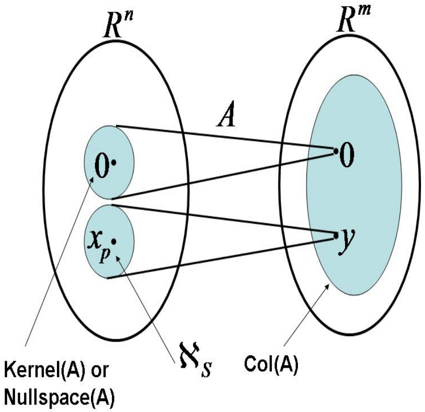
Case 2: When A is singular



$\aleph_S = \{\mathbf{x} \in \mathbf{R} : A\mathbf{x} = \mathbf{y}\} = \mathbf{x}_p + \mathcal{N}(A)$

Linear inverse problem

Consider $A \in \mathbf{R}^{m \times n}$, with $m < n$. with $\text{rank}(A) = m$. A is singular.

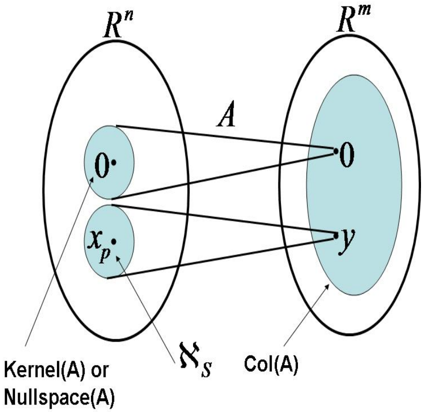


1. Is it possible to get a unique solution?

$$\aleph_S = \{\mathbf{x} \in \mathbf{R} : A\mathbf{x} = \mathbf{y}\} = \mathbf{x}_p + \mathcal{N}(A)$$

Linear inverse problem

Consider $A \in \mathbf{R}^{m \times n}$, with $m < n$. with $\mathrm{rank}(A) = m$. A is singular.

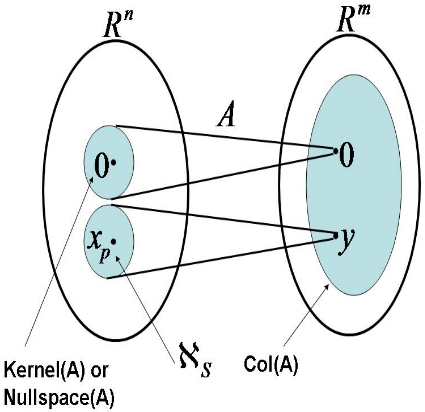


$$\aleph_S = \{\mathbf{x} \in \mathbf{R} : A\mathbf{x} = \mathbf{y}\} = \mathbf{x}_p + \mathcal{N}(A)$$

1. Is it possible to get a unique solution? Probably!

Linear inverse problem

Consider $A \in \mathbf{R}^{m \times n}$, with $m < n$. with $\text{rank}(A) = m$. A is singular.

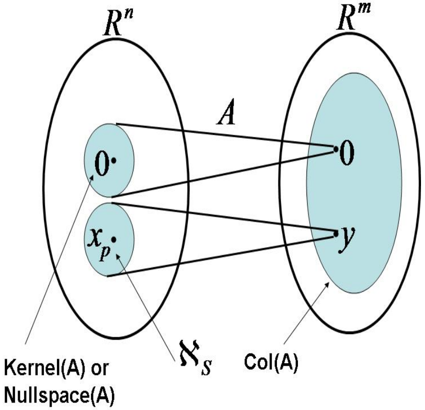


1. Is it possible to get a unique solution? Probably!
   We could put some constraints on the $\aleph_s$ such that we could get a unique solution $\mathbf{x}^*$ to the problem.
2. Is $\mathbf{x}^* = x_p$? Yes!

$$\aleph_S = \{\mathbf{x} \in \mathbf{R} : A\mathbf{x} = \mathbf{y}\} = \mathbf{x}_p + \mathcal{N}(A)$$

Linear inverse problem

Consider $A \in \mathbf{R}^{m \times n}$, with $m < n$. with rank$(A) = m$. A is singular.



$$\aleph_S = \{\mathbf{x} \in \mathbf{R} : A\mathbf{x} = \mathbf{y}\} = \mathbf{x}_p + \mathcal{N}(A)$$

1. Is it possible to get a unique solution? Probably!
   We could put some constraints on the $\aleph_s$ such that we could get a unique solution $\mathbf{x}^*$ to the problem.
2. Is $\mathbf{x}^* = x_p$? Yes!
   If we can put a constraint on the set of values $\mathbf{x}_p$ can take and if we put an appropriate constraint on $\aleph_S$.

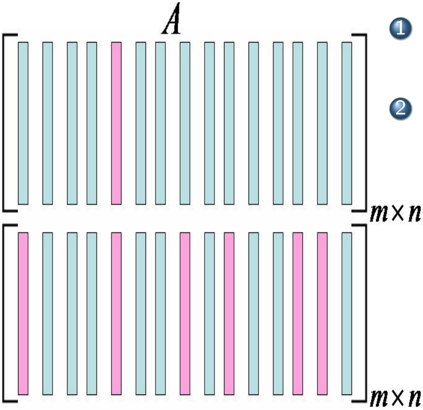
Linear inverse problem

Consider $A \in \mathbf{R}^{m \times n}$, with $m < n$, with rank$(A) = m$. Spark(A) is defined as the smallest number of linearly dependent columns. If k denotes the number of non-zero elements of $\mathbf{x}$, in other words, $k = \|\mathbf{x}\|_0$.

If $k < \frac{1}{2} Spark(A)$, then there does not exist any other vector $\mathbf{x}'\mathbf{R}^n$, such that $A\mathbf{x}' = \mathbf{y}$ and $\|\mathbf{x}'\|_0 \leq k$

Linear inverse problem

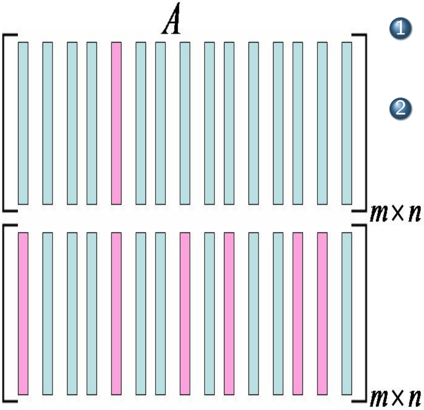
Let us see the intuition behind this. Consider $A \in \mathbf{R}^{m \times n}$, with $m < n$, with $\text{rank}(A) = m$. Let the $spark(A) = s$. $1 \leq s \leq m + 1$. Let us assume $s > 1$.



1. Case 1: k=1 Trivially the case is true. Since $Spark(A) > 1$.
2. Case 2: k= s. If $\mathbf{x}_p$ has its non-zero elements such that $\mathbf{y}$ is a linear combination of $\{\mathbf{a}_{i_1}, \mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_s}\}$, where $i_k \in \{1, \cdots, n\}$. There exists a be a vector $\mathbf{x}'$ such that it selects the column $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_s}\}$, $j_k \in \{1, \cdots, n\}$, such that $A\mathbf{x}' = y$ because the set $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_s}, \mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_s}\}$ consists of linearly dependent columns.

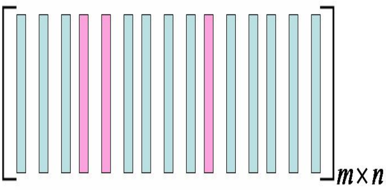
Linear inverse problem

Let us see the intuition behind this. Consider $A \in \mathbf{R}^{m\times n}$, with $m < n$, with $\text{rank}(A) = m$. Let the $spark(A) = s$. $1 \le s \le m+1$. Let us assume $s > 1$.



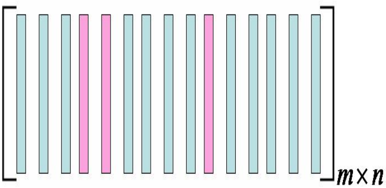
1. Case 1: k=1 Trivially the case is true. Since $Spark(A) > 1$.
2. Case 2: k= s. If $\mathbf{x}_p$ has its non-zero elements such that $\mathbf{y}$ is a linear combination of $\{\mathbf{a}_{i_1}, \mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_s}\}$, where $i_k \in \{1, \cdots, n\}$. There exists a be a vector $\mathbf{x}'$ such that it selects the column $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_s}\}$, $j_k \in \{1, \cdots, n\}$, such that $A\mathbf{x}' = y$ because the set $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_s}, \mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_s}\}$ consists of linearly dependent columns.

Linear inverse problem



1. Case 3: $k < s/2$. If $\mathbf{x}_p$ has its non-zero elements such that $\mathbf{y}$ is a linear combination of $\{\mathbf{a}_{i_1}, \mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_k}\}$, where $i_l \in \{1, \cdots, n\}$. There does not exist a vector $\mathbf{x}'$ such that it selects the column $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_k}\}$, $j_l \in \{1, \cdots, n\}$, such that $A\mathbf{x}' = y$ because the set $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_k}, \mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_k}\}$ consists of linearly independent columns.

Linear inverse problem



1. Case 3: $k < s/2$. If $\mathbf{x}_p$ has its non-zero elements such that $\mathbf{y}$ is a linear combination of $\{\mathbf{a}_{i_1}, \mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_k}\}$, where $i_l \in \{1, \cdots, n\}$. There does not exist a vector $\mathbf{x}'$ such that it selects the column $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_k}\}$, $j_l \in \{1, \cdots, n\}$, such that $A\mathbf{x}' = y$ because the set $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_k}, \mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_k}\}$ consists of linearly independent columns.

Therefore, given a matrix A, we can find out the $spark(A)$. If we restrict the vector $\mathbf{x}_p$ to be in the set $\{\mathbf{x} : \|\mathbf{x}\|_0 < \frac{1}{2} spark(A)\}$, then $\aleph_S = \{\mathbf{x} : \|\mathbf{x}\|_0 < spark(A)$ such that $A\mathbf{x} = \mathbf{y}\}$.
If $\mathbf{x}^* = \min \aleph_S$, then $\mathbf{x}^* = \mathbf{x}_p$. Thus, recovers $\mathbf{x}$ 'exactly'!

Problem Formulation

Implication of having a unique solution : We can project the sparse signals onto lower dimensional space and recover it exactly.

Problem Formulation

Implication of having a unique solution : We can project the sparse signals onto lower dimensional space and recover it exactly.

Big question

How do we recover the vector $\mathbf{x}$ from $\mathbf{y}$?
What is the computational complexity?

Sparse vector recovery

In principle, we can recover $\mathbf{x}$ exactly by solving the combinatorial optimization problem

$$\min_{\mathbf{x}\in\mathbf{R}} \|x\|_0 \text{ subject to: } A\mathbf{x} = y \tag{3}$$

Solving (3) directly is infeasible even for modest-sized signals. The algorithm would run over all subsets of $\{1, \cdots, N\}$ of cardinality less than half of the spark of A and each time check whether $A\mathbf{x}$ was in the range of the selected columns or not, and then find the solution for the submatrix A_s consisting of selected columns . It is well-known that this procedure would clearly be very computationally expensive.

Sparse vector recovery

Theory gives guarantees of exact recovery in the noiseless case. In the noisy case where

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \text{where } \mathbf{n} \text{ is the measurement noise} \tag{4}$$

we cannot expect to recover the signal exactly.
In this case, we solve the following minimization problem, which can be viewed as a regression problem.

$$\min_{\mathbf{x}\in\mathbf{R}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_0 \tag{5}$$

Sparse vector recovery

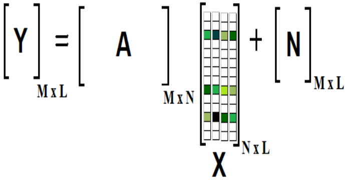
In many cases, the measured signal itself might be *approximately sparse* (signals with very few large components and large number of small components). Infact we will show that noisy measurement of strictly sparse signals can be viewed as noiseless measurements of approximately sparse signals. In this case, we solve the following minimization problem

$$\min_{\mathbf{x}\in\mathbf{R}} \sum_{i=1}^{N} \mathcal{I}_{\{|x_i|>\ \epsilon\}} \text{ subject to:} \mathbf{y} = \mathbf{A}\mathbf{x} \tag{6}$$

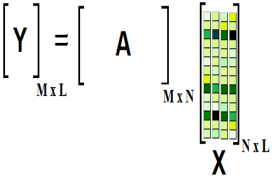
From SMV to MMV

We know that when the signal is approximately sparse or if the measurement is noisy, we cannot recover the signal exactly. In such situations the signal recovery could be improved if multiple measurement vectors(MMV) with the same sparsity profile (that is the position of the large coefficients) are available.

Noisy measurement of strictly sparse signals



Noiseless measurement of approximately sparse signals



Review of existing methods

Below we compare the recent methods of sparse recovery based on Bayesian framework namely

1. Sparse Bayesian learning (SBL)[1]
2. Compressed Sensing-Belief Propagation (CS-BP),[2]
3. Approximate Message Passing(AMP)[3]
4. Joint Support and Signal Recovery using Message Passing (JSSR-MP).[4]

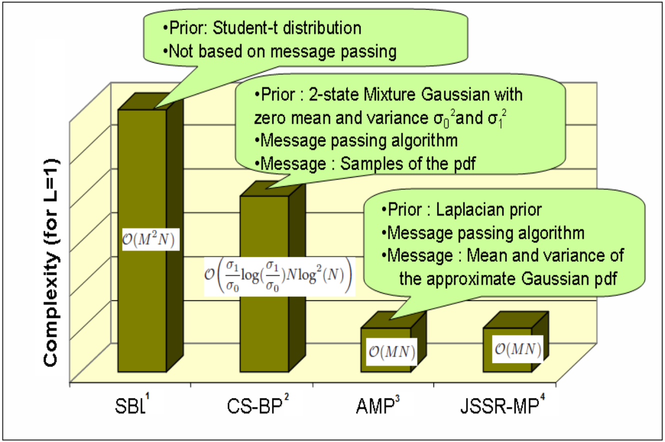
[1] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Sig. Proc.*, vol. 55, no. 7, pp. 3704-3716, Jul. 2007.

[2] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressing sensing via belief propagation," *IEEE Trans. Sig. Proc.*

[3] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *PNAS*, Sept. 2009.

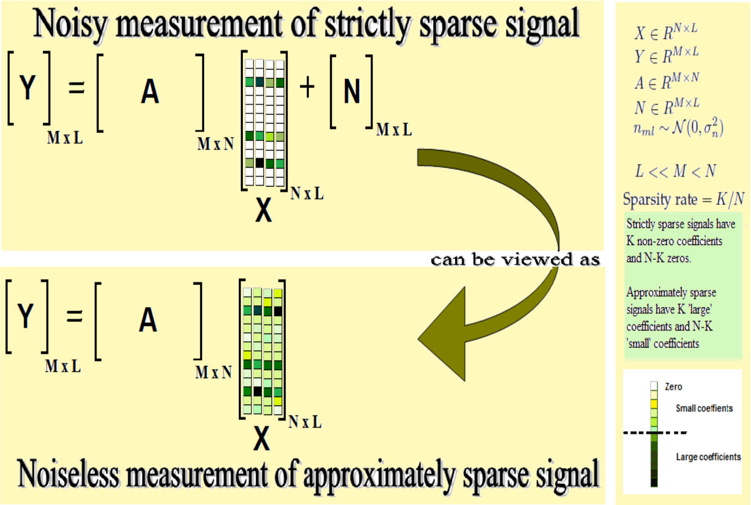
[4] JSSR-MP : S. Shedthikere and A. Chockalingam, "Message Passing Algorithm for Joint Support and Signal Recovery of Approximately Sparse Signals," ICASSP, May 2011.

Review of existing methods



Note : the above picture is not drawn to scale in term of orders. It is for representation purpose only.

From strictly sparse to approximately sparse



From strictly sparse to approximately sparse

Let $\mathbf{x}$ be a strictly sparse signal and $\mathbf{x}' = \mathbf{x} + \mathbf{e}$ be an approximately sparse signal. A noiseless measurement of $\mathbf{x}'$ of the form $\mathbf{y} = \mathbf{A}\mathbf{x}'$ can be viewed as equivalent to the noisy measurement $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, if $\mathbf{n}' = \mathbf{A}\mathbf{e}$ has the same statistical characteristics as $\mathbf{n}$. The following lemma gives the the statistical characteristics of $\mathbf{e}$ for $\mathbf{n}$ and $\mathbf{n}'$ to have the same distribution.

Lemma

If $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M < N$, with i.i.d entries from $\mathcal{N}(0, \sigma_a^2)$, $\mathbf{n} \in \mathbb{R}^M$ with i.i.d entries from $\mathcal{N}(0, \sigma_n^2)$ and $\mathbf{e}$ whose entries are i.i.d and distributed from $\mathcal{N}(0, \sigma_e^2)$, then $\mathbf{n}' = \mathbf{A}\mathbf{e}$ has the same distribution as that of $\mathbf{n}$, where $\sigma_e^2 = \frac{\sigma_n^2}{N\sigma_a^2}$.

1. Overview of Sparse Signal Recovery
   - Problem Formulation
   - Review of existing methods

2. JSSR-MP
   - From strictly sparse to approximately sparse
   - Bayesian inference problem
   - Message Passing Algorithm
   - Joint Support and Signal Recovery using Message Passing (JSSR-MP)
   - Discussion on computational complexity
   - Results

3. Future work

Bayesian inference problem

The recovery of approximately sparse signals can be viewed as Bayesian inference problem in which, the key idea is to determine that signal that is consistent with the observed measurements and that best matches our signal model.

A re-look at the minimization problem

$$\min_{\mathbf{x}\in\mathbf{R}} \underbrace{\sum_{i=1}^{N} \mathcal{I}_{\{|x_i|>\ \epsilon\}}}_{\text{sparsity promoting prior}} \quad \text{subject to: } \underbrace{\mathbf{y} = \mathbf{A}\mathbf{x}}_{\text{likelihoods}} \tag{7}$$

The above minimization problem can be cast as:

$$\widehat{\mathbf{x}} = \arg\max_{\mathbf{x}} \; p(\mathbf{x})p(\mathbf{y}|\mathbf{x}) \tag{8}$$

$$= \arg\max_{\mathbf{x}} \; p(\mathbf{x}) \prod_{i=1}^{M} \delta_{\{y_u = \mathbf{a}_u . \mathbf{x}\}} \tag{9}$$

Choice of prior: 2-state Gaussian mixture

We choose the 2-state Gaussian mixture to model the approximately signal. The hyperparameter Q_i represents whether the signal component is large or small

$$Q_i = \begin{cases} 1 & \text{represents a large component} \\ 0 & \text{represents a small component} \end{cases}$$

$p(x_i|Q_i)$ is Gaussian and is given as

$$p(x_i|Q_i = 1) = \begin{cases} \mathcal{N}(x_i; 0, \sigma_L^2) & \text{if } Q_i = 1 \\ \mathcal{N}(x_i; 0, \sigma_S^2) & \text{if } Q_i = 0 \end{cases}$$

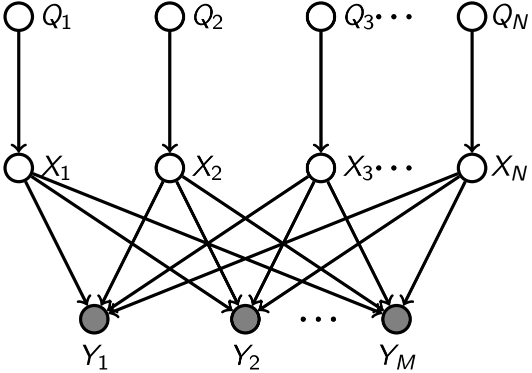
Thus, $p(x_i)$ is given as

$$p(x_i) \ = \ p(Q_i = 1)\mathcal{N}(x_i; 0, \sigma_L^2) + p(Q_i = 0)\mathcal{N}(x_i; 0, \sigma_S^2), \qquad (10)$$

giving rise to a 2-state mixture Gaussian prior.

Graphical representation

The joint probability distribution
$p(\mathbf{x},\mathbf{y},\mathbf{q};\theta) = \prod_{i=1}^{N} p(q_i;\zeta_i) \prod_{i=1}^{N} p(x_i|q_i;\theta_i) \prod_{u=1}^{M} p(y_u|\mathbf{x})$
where, $\theta = \{\zeta_1, \zeta_2, \cdots, \zeta_N, \sigma_L^2, \sigma_S^2\}$
We want to find the Maximum aposteriori (MAP) estimate of $\mathbf{x}$



Signal and Support Recovery

The MAP estimate of $\mathbf{x}$ can be obtained by marginalizing out the hyperparameters from the complete posterior distribution $p(\mathbf{x}, \mathbf{q}|\mathbf{y})$, as

$$\widehat{\mathbf{x}}_{MAP} = \arg\max_{\mathbf{x}} \sum_{\mathbf{q}\in\{0,1\}^N} p(\mathbf{x}|\mathbf{y}, \mathbf{q})p(\mathbf{q}|\mathbf{y}), \tag{11}$$

where $\mathbf{q} = [q_1 \ q_2 \cdots q_N]^T \in \{0, 1\}^N$ denotes the sparsity profile of $\mathbf{x}$, and the posterior distribution of the hyperparameters $p(\mathbf{q}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{q})p(\mathbf{q})$ is obtained by marginalizing out $\mathbf{x}$ as

$$p(\mathbf{q}|\mathbf{y}) \propto \left(\int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{q})d\mathbf{x} \right) p(\mathbf{q}). \tag{12}$$

The support is recovered from the posterior distribution of Q_i's as

$$\widehat{S} = \{i : p(Q_i = 1|\mathbf{y}) \geq p(Q_i = 0|\mathbf{y})\}. \tag{13}$$

1. Overview of Sparse Signal Recovery
   - Problem Formulation
   - Review of existing methods

2. JSSR-MP
   - From strictly sparse to approximately sparse
   - Bayesian inference problem
   - **Message Passing Algorithm**
   - Joint Support and Signal Recovery using Message Passing (JSSR-MP)
   - Discussion on computational complexity
   - Results

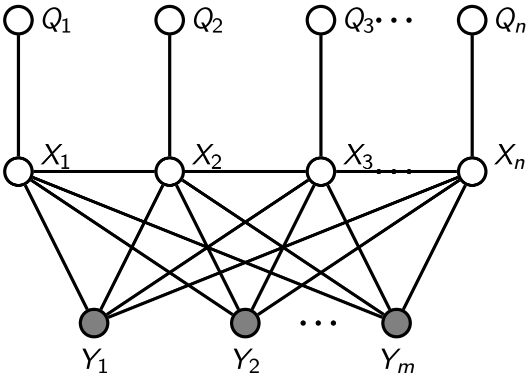
3. Future work

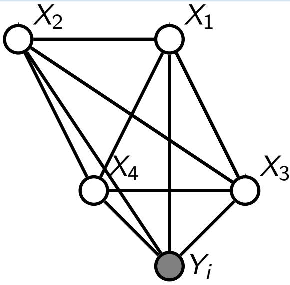
From directed graphs to factor graphs

We know that the sum-product algorithm is a very efficient way to calculate the marginal probabilities at all the nodes. But sum product algorithm give exact marginals only on directed trees.

From directed graphs to factor graphs

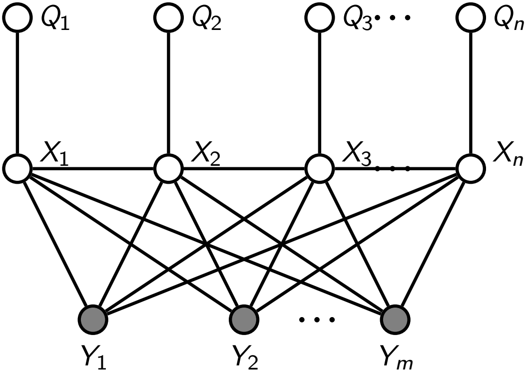
Step 1: Moralize the graph



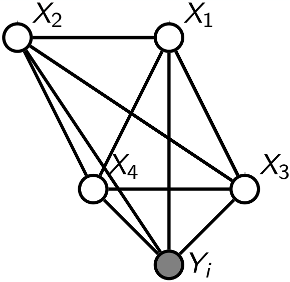


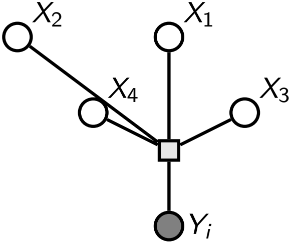
From directed graphs to factor graphs

Step 1: Moralize the graph

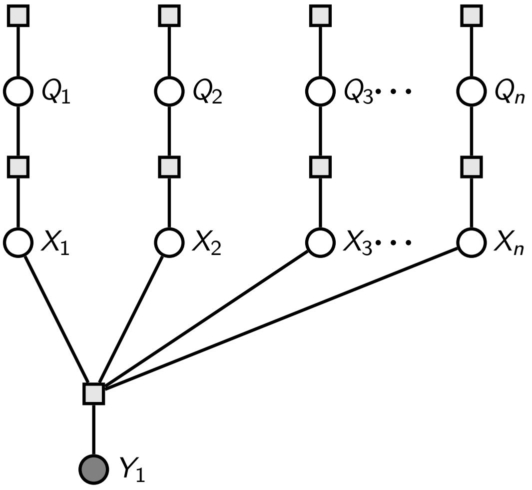


Step 2: Add factor nodes



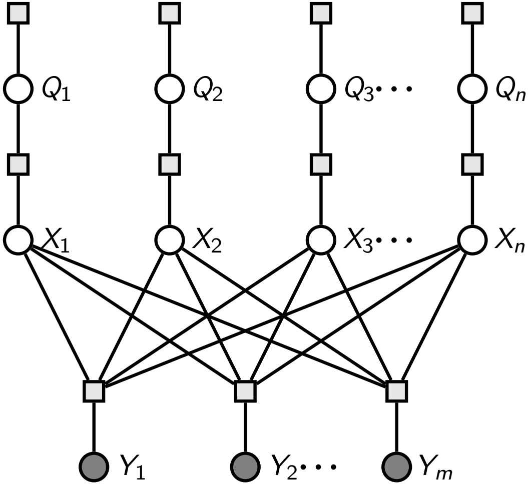


From directed graphs to factor graphs



Looks like the problem is solved to certain extent

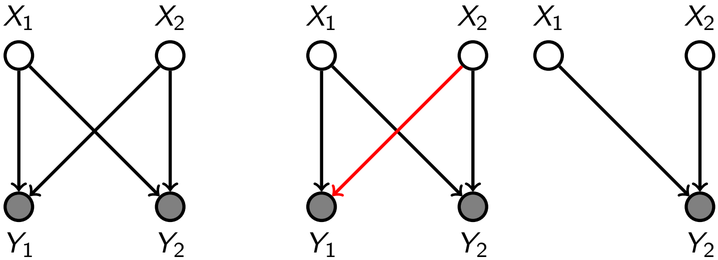
From directed graphs to factor graphs



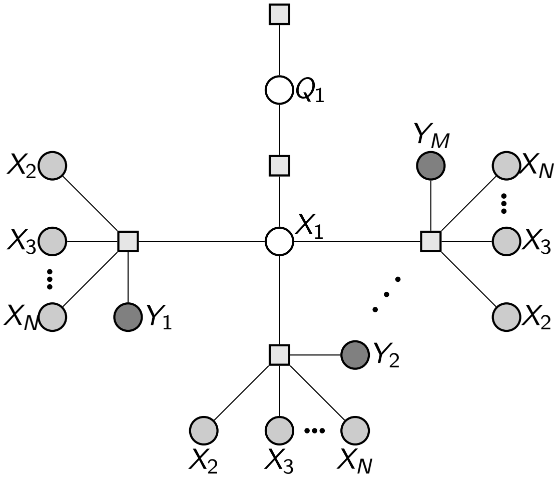
Not really!!

Cycle-free approximations

The root cause of the cycle shown below is both Y_1 and Y_2 have same parents. So having observed Y_1 and Y_2, X_1 and X_2 become dependent and so the path is not block.



Consider the path $X_1 Y_2 X_2 Y_1 X_1$. If X_2 is observed, then knowing X_2, would block the path.



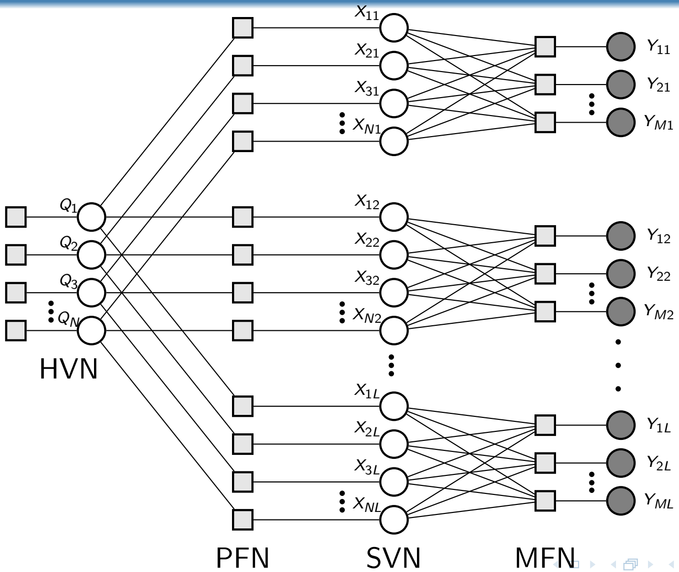
During each iteration we could use the posterior distribution of the previous iteration as the 'soft' observation and consider using the sum-product algorithm on this approximated cycle-free graph.

1. Overview of Sparse Signal Recovery
   - Problem Formulation
   - Review of existing methods

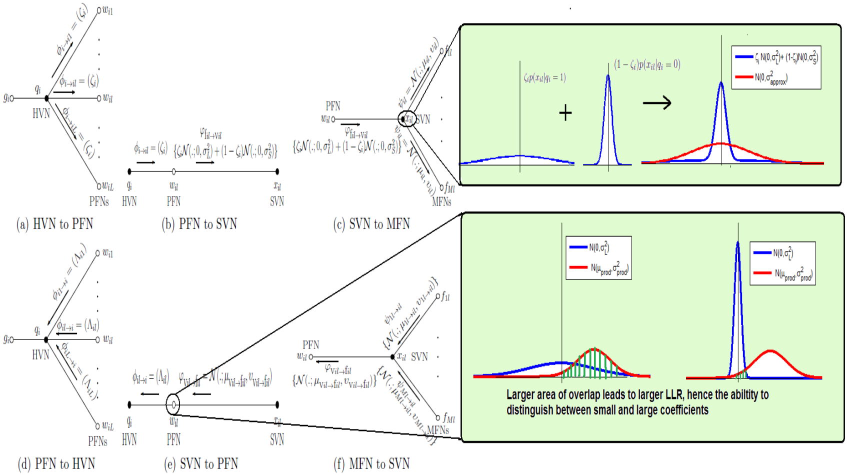
2. JSSR-MP
   - From strictly sparse to approximately sparse
   - Bayesian inference problem
   - Message Passing Algorithm
   - Joint Support and Signal Recovery using Message Passing (JSSR-MP)
   - Discussion on computational complexity
   - Results

3. Future work

Factor Graph



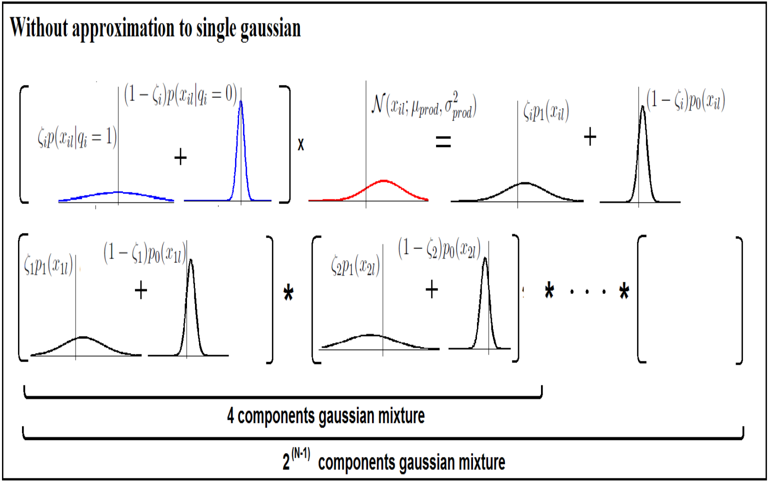
JSSR-MP for MMV



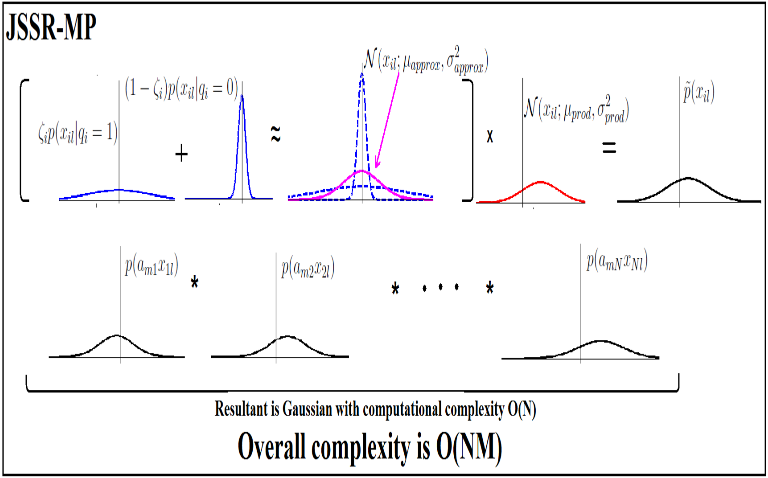
(a) HVN to PFN (b) PFN to SVN (c) SVN to MFN

(d) PFN to HVN (e) SVN to PFN (f) MFN to SVN

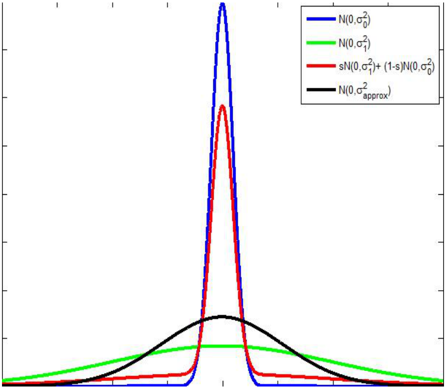
Complexity without Gaussian approximation



Computation complexity of JSSR-MP



Effect of the approximation on the sparsity of the recovered vector



Observe that if we were minimizing the vector $\mathbf{x}$ in the energy sense or $\|\cdot\|_2$ then using Gaussian would be appropriate. But we see that using this approximation at the MFN and use of the mixture to classify it as large and small components works well in our case.

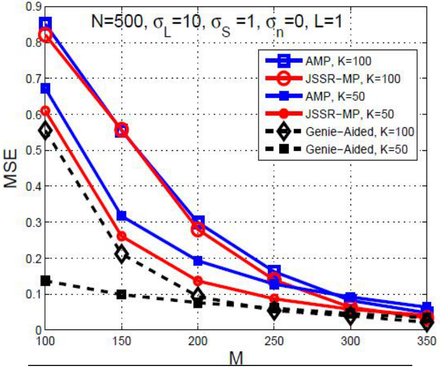
1. Overview of Sparse Signal Recovery
   - Problem Formulation
   - Review of existing methods

2. JSSR-MP
   - From strictly sparse to approximately sparse
   - Bayesian inference problem
   - Message Passing Algorithm
   - Joint Support and Signal Recovery using Message Passing (JSSR-MP)
   - Discussion on computational complexity
   - Results

3. Future work

Performance of JSSR-MP for SMV

MSE versus M performance of signal recovery of approximately sparse signals with noiseless measurements for SMV(L=1)

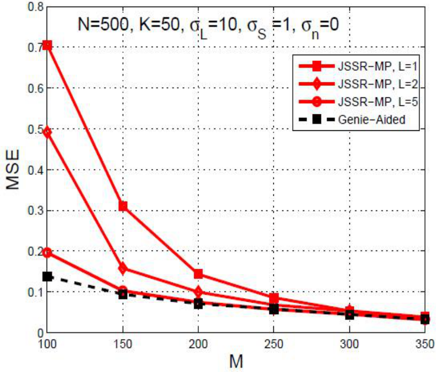


- As K/N increases performance degrades
- JSSR-MP outperforms CoSaMP[1]
- JSSR-MP performance almost same as AMP
- Genie-aided system has perfect knowledge of support

[1] D. Needell, J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Jl. of Appl. Comput. Harmon. Anal*, vol. 26, pp. 301-321, 2009.

Performance of JSSR-MP MMV

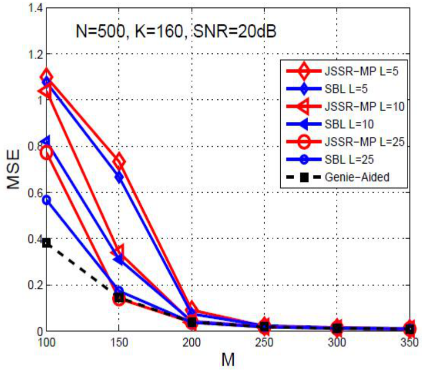
MSE versus M performance of signal recovery of approximately sparse signals with noiseless measurements for MMV($L > 1$)



- JSSR-MP shows significant improvement in performance as L increases
- Performance close to genie-aided system is achieved for small values of L (as shown for $L = 5$)

Performance of JSSR-MP for noisy measurements

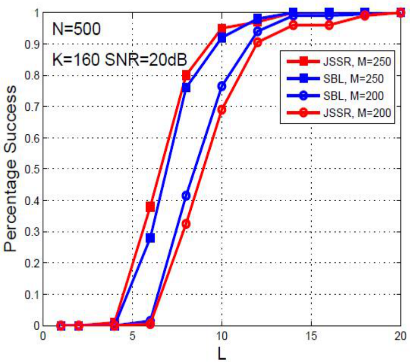
MSE versus M performance of JSSR-MP algorithm for recovery of strictly sparse signals with noisy measurements for MMV



- Perfomance of JSSR-MP is comparable to that of M-SBL and this is achieved by JSSR-MP($\mathcal{O}(MNL)$) at one order less complexity than M-SBL($\mathcal{O}(M^2N)$)

Support recovery

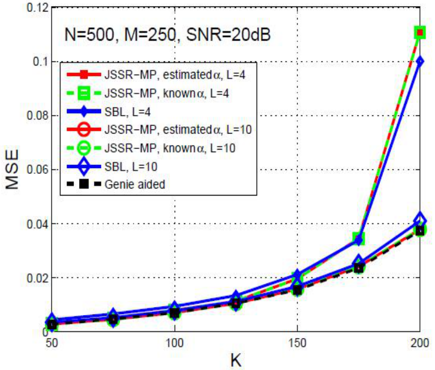
% success of support recovery versus L for strictly sparse signals with noisy measurements for MMV



•Perfomance of JSSR-MP is comparable to that of M-SBL in terms of support recovery

Performance of JSSR-MP with increasing sparsity rate

MSE versus K performance of JSSR-MP algorithm for recovery of strictly sparse signals with noisy measurements for MMV



- MSE performance degrades as K increases.
- Performance close to genie aided system can be achieved by increasing L as K increases.

Future work

1. Recovery of signals which are correlated in magnitude.

Future work

1. Recovery of signals which are correlated in magnitude.
2. Recovery of signals which are correlated in support.

Future work

1. Recovery of signals which are correlated in magnitude.
2. Recovery of signals which are correlated in support.
3. Application of JSSR-MP to sparse channel estimation.

Future work

1. Recovery of signals which are correlated in magnitude.
2. Recovery of signals which are correlated in support.
3. Application of JSSR-MP to sparse channel estimation.

Refernces

- M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing of approximately sparse signals," *Proc. IEEE ISIT 2008*, Toronto, Jul. 2008.
- D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressing sensing via belief propagation," *IEEE Trans. Sig. Proc.*, Jan. 2010.
- S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346-2356, June 2008.
- M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Jl. of Mach. Learn. Res.*, vol. 1, pp. 211-244, Sept. 2001.
- O. Shental, D. Bickson, P. H. Siegel, J. K. Wolf, and D. Dolev, "Gaussian belief propagation solver for systems of linear equations," *Proc. IEEE ISIT'2008*, Toronto, Jul. 2008.