Statistical Compressed Sensing of Gaussian Mixture Models

MANTRI KRISHNA SRI IPSIT | SHAAN UL HAQUE

180070032 | *180070053*

COURSE: CS754

INSTRUCTOR: PROF. AJIT RAJWADE

Abstract - In this project, we will study and implement the concept of statistical compressed sensing that aims at efficiently sampling a collection of signals that follow a statistical distribution, and achieving accurate reconstruction on average. Signals that follow Gaussian distribution with fast decaying eigenvalues of the co-variance matrix require O(k) measurements as compared to O(klog(N/k)) number of measurements required for the convention counterpart with error bounded by constant times the best k-term approximation error. Since images do not follow Gaussian distribution as a whole we will model them using Gaussian Mixture Models and estimate the parameters of the models from the sensed signal. Lastly, we will show the superiority of SCS over conventional CS by running both algorithms over test images.

1 Introduction

1.1 Some Theory

First let us analyse the situation when the signal is distributed according to just one Gaussian whose parameters are known. Without loss of generality, we assume the signal to be zero mean. Then we have the following theorem.

Theorem: Let $\boldsymbol{x} \in \boldsymbol{R^N}$ be a random vector with pdf given as $N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Phi} \in \boldsymbol{R^{M \times N}}$ be a sensing matrix with $M \leq N$. Let $\boldsymbol{y} = \boldsymbol{\Phi x} \in \boldsymbol{R^M}$ be the sensed signal. Then the decoder Δ that minimizes the MSE(mean square error) and MAE(mean absolute error) is the linear MAP estimator, given by:

$$\Delta(\boldsymbol{\Phi x}) = argmax_{\boldsymbol{x}} p(\boldsymbol{y}|\boldsymbol{x}) = \boldsymbol{\Sigma \Phi^T (\Phi \Sigma \Phi^T)^{-1} (\Phi x)}$$

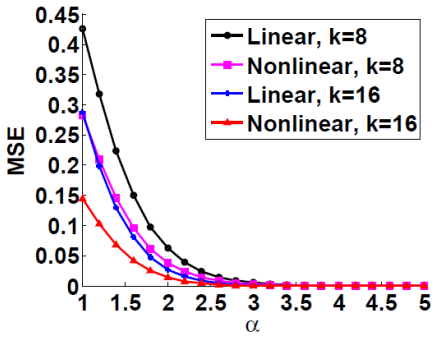
Monte Carlo simulations suggested that for signal whose eigenvalues of the co-variance matrix decreases very fast linear best-k approximator is almost as good as the best non-linear best-k approximator(like oracular estimate). The best-k linear approximator is given by:

$$\boldsymbol{x}_k^l(m) = \begin{cases} \boldsymbol{x}(m) & if\ 1 \leq m \leq k \\ 0 & if\ k+1 \leq m \leq N \end{cases}$$

and the nonlinear approximation:

$$\boldsymbol{x}_k^n = T_k(\boldsymbol{x})$$

where T_k is a thresholding operator that keeps the k coefficients of largest amplitude and setting others to zero, lead to comparable approximation errors The simulation results(taken directly from paper[1]) are shown below.



(a)

Figure 1: MSE vs α

Here eigenvalues $\lambda_i = i^{-\alpha}$. Since both errors are comparable we will evaluate the performance of our algorithm in terms of linear best-k approximator.
Further in the paper the author has shown that the linear MAP filter has error bounds comparable to the best-k linear approximator when the co-variance matrix has fast decaying eigenvalues as we have in natural image patches.

1.2 Why only O(k) measurements works well??

Without loss of generality let us assume that signal $\boldsymbol{x} \in \boldsymbol{R}^N$ is zero mean Gaussian distributed with diagonal co-variance matrix $N(\boldsymbol{0}, \boldsymbol{S})$. The diagonal elements of $\boldsymbol{S}$ decreases, i.e, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$. Since variance of the elements in the random vector $\boldsymbol{x}$ decreases probability that it lies near it's mean, i.e. zero value is higher. Thus, on an average if we extract only the first k-elements of the vector then the error with the original vector will be low on average. This is why only k measurements suffices.
In conventional CS we don't know the position of non-zero elements while in SCS we know the exact indices where the non-zero elements are present with very high probability.
Now real signal is not zero mean and not white as well. But we can always centre the signal wrt mean and whiten the signal by applying SVD to co-variance matrix. And since all these operations are reversible hence all our analysis holds for them as well.

1.3 Model selection with GMM

At the heart of the GMM-based SCS decoder is the model selection. As correct Gaussian models are selected, the SCS performance bounds described in paper apply. Selection of the model is a simple task. We first subtract mean from the signal and then apply MAP linear estimator or linear decoder to compute the estimated signal. The estimated signal is computed for each Gaussian model and then that estimate is chosen which maximizes the log a-posteriori among all models. Let J be the total number of Gaussians, them mathematically:

$$\tilde{\boldsymbol{x}}_j = \Delta_j(\boldsymbol{\Phi x}) = \boldsymbol{\Sigma}_j \boldsymbol{\Phi}^T (\boldsymbol{\Phi} \boldsymbol{\Sigma}_j \boldsymbol{\Phi}^T)^{-1} (\boldsymbol{\Phi x}) \; \forall \; 1 \leq j \leq J$$

Then selecting that model which maximizes a-posteriori, given as:

$$\tilde{j} = argmax_{1 \leq j \leq J} - \frac{1}{2}(log(\boldsymbol{\Sigma}_j) + \boldsymbol{x}^T \boldsymbol{\Sigma}_j \boldsymbol{x})$$

Further the paper has shown using simulations and mathematical arguments that for covariance matrices with fast decaying eigenvalues the above estimation of model gives correct model almost always.

1.4 SCS with GMM – Algorithm

In real sensing, the Gaussian parameters are unknown to us and we need to estimate them from the sensed signal then proceed with reconstruction. We will use the famous MAP-EM algorithm from Machine Learning to estimate the parameters as well as the signal or image. The overview of the algorithm is as follows:

- E-step - Assuming that the estimates of the Gaussian parameters $\{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}_{1\leq j\leq J}$ are known (following the previous M-step), the E-step calculates the MAP signal estimation and does model selection for all the signals.
- M-step - Assuming that the Gaussian model selection $\tilde{j}$ and the signal estimate $\tilde{\boldsymbol{x}}$ are known for all the signals (following the previous E-step), the M-step estimates (updates) the Gaussian models $\{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}_{1\leq j\leq J}$. The updates are done based on the Maximum Likelihood Estimate. Let $\boldsymbol{C}_j$ be the set of signals which belong to jth model.

$$\tilde{\boldsymbol{\mu}}_j = \frac{1}{|C_j|}\sum_{i in C_j} \tilde{\boldsymbol{x}}_i \text{ and } \tilde{\boldsymbol{\Sigma}}_j = \frac{1}{|C_j|}\sum_{i in C_j} (\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{\mu}}_j)(\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{\mu}}_j)^T$$

  It might happen that some of the co-variance matrix are not invertible, hence we add $\epsilon \boldsymbol{I}$ to all the matrix to have numerical stability.

The biggest problem with MAP-EM algorithm is that this is highly non-convex problem hence initialization plays a big role in settling to the global minima and not getting stuck at local minimas. Thus for initialization we applied the trick as mentioned in the paper[2]. We also carried out blind compressed sensing to compare the performance of SCS with conventional CS. The algorithm for blind CS was same as given the paper[3] The total number of Gaussian models were 19. For all cases we used randomly generated Gaussian sensing matrices.

2 Analysis of Experimental Results

The dictionary for conventional CS is learned with K-SVD from 720, 000 image patches, extracted from the entire standard Berkeley segmentation database containing 300 natural images. Instead of 3 images we took four standard images Lena, House, Pepper and Lake all 512*512 pixel images.

(a) Lena. **(b)** House. **(c)** Lake. **(d)** Peppers.

Figure 2: Standard Test Images.

To reduce time taken by algorithm we resized the images to 128*128. We extracted $\approx$ 14500 image patches from the test images and carried out reconstruction by averaging out the overlapping patches. Averaging helped to reduce block artifacts significantly. SCS was found to be superior then both blind CS and conventional CS almost always. The RMSE was significantly lower for higher measurement ratios. Some reconstructed images for the case of 50% measurement ratio are shown below.

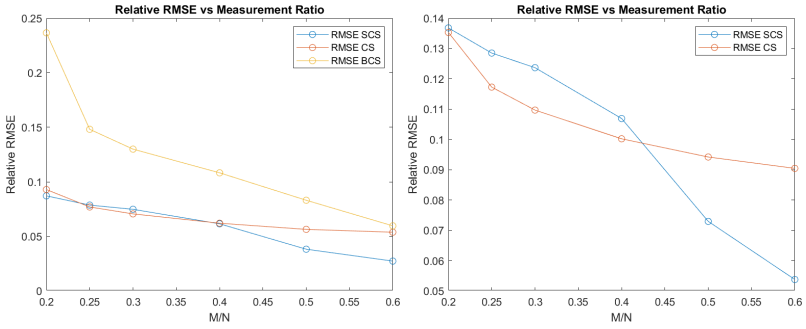
(a) SCS reconstruction. **(b)** KSVD CS reconstruction.

Figure 3: Measurement Ratio 0.5.

(a) SCS reconstruction. **(b)** Blind CS reconstruction.

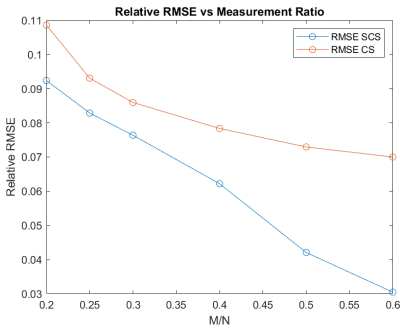
Figure 4: Measurement Ratio 0.5.

We didn't run blind cs for all images as it took 1 hour to run on single image (for all ratios) and RMSE was worse then KSVD method. The rmse plots for all the four images are shown in the next page. We can see the superiority of SCS over other CS algorithms. For some cases KSVD was better perhaps a better selection of the number of Gaussian model would have done better but nevertheless SCS proved to be better at most always especially at high measurement ratios.
The reason for significantly lower RMSE at high measurement ratio is because of SCS demands only O(k) measurements as compared to O(klog(N/k)) measurements in conventional CS thus giving almost exact reconstruction.

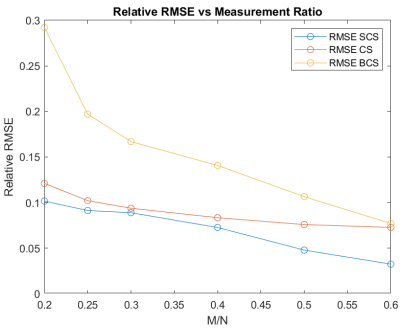


(a) Lena.

(b) House.



(c) Lake.



(d) Peppers.

Figure 5: RMSE vs measurement ratio(M/N) for all images.

The main reason for SCS to produce better reconstruction is it's better ability to capture pixels near geometric boundaries. Some of the image patches extracted from Lena image near object boundaries are shown below. The first column is ground truth, second is SCS while third is CS. Measurement ratio was 0.4.

Figure 6: Patch 1

Figure 7: Patch 2

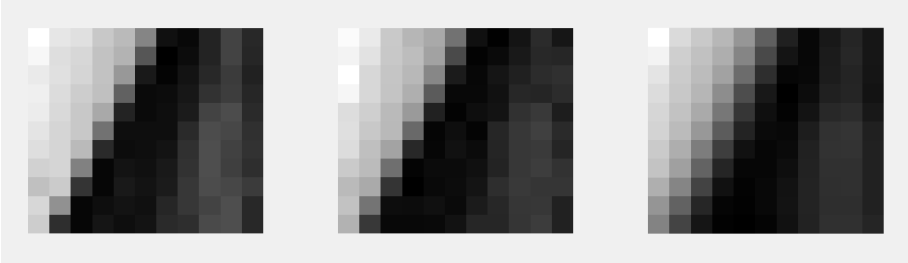


Figure 8: Patch 3

Figure 9: Patch 4

3 Conclusion

From our observations we can conclude that SCS is a better technique for image or signal reconstruction when the signal's distribution or atleast form is known to us. Unlike, KSVD no training data was required and reconstruction speed was also almost same.
The only major impediment the SCS is the initialization of the MAP-EM algorithm if signal's distribution parameters are also to be estimated. Because of the non-convexity of the problem, we can get stuck at local minima thus initialization needs to proper or atleast best results should be taken across large number of initializations.

4 Usage of Code

- There are a total of three set of codes SCS, conventional CS(KSVD) and blind CS. The main algorithm is the SCS code while others are for comparison purpose. SCS code file has a main.m code which reconstructs the image while other code files are functions. SCS code takes about 40 mins to run for a single image for a total of 6 M/N values (measurement ratio).

- Training data for KSVD can be downloaded from The Berkeley Segmentation Dataset and Benchmark

- To perform KSVD, we took the ksvd and omp toolboxes from Ron Rubinstein's toolboxes. To be able to use these toolboxes, a C language compiler is required (e.g. MinGW or Microsoft Visual C++). The procedure to properly use these toolboxes is given on the website mentioned above.

5 References

1. G. Yu and G. Sapiro, "Statistical Compressed Sensing of Gaussian Mixture Models," in IEEE Transactions on Signal Processing, vol. 59, no. 12, pp. 5842-5858, Dec. 2011.
2. G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. Submitted, arxiv.org/abs/1006.3056, June, 2010.
3. F. Pourkamali-Anaraki, S. Becker and S. M. Hughes, "Efficient dictionary learning via very sparse random projections," 2015 International Conference on Sampling Theory and Applications (SampTA), 2015, pp. 478-482.
4. S.M. Kay. Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory. Prentice Hall, 1998.
5. M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. on Signal Proc., 54(11):4311, 2006