Statistical Compressed Sensing of Gaussian Mixture Models

Mantri Krishna Sri Ipsit | Shaan Ul Haque
180070032 | 180070053

Statistical Compressed Sensing

- Exploiting the statistical properties of natural images
- A collection of signals can be efficiently reconstructed
- Each subpart of an image can be assumed to follow a Gaussian distribution
- Linear Decoder (in case of GMMs)

Theory - Linear Decoder is Optimal

Theorem 1. *[23] Let* $\mathbf{x} \in \mathbb{R}^N$ *be a random vector with prior pdf* $\mathcal{N}(\mathbf{0}, \Sigma)$, *and* $\Phi \in \mathbb{R}^{M \times N}$, $M \leq N$, *be a sensing matrix. From the measured signal* $\mathbf{y} = \Phi\mathbf{x} \in \mathbb{R}^M$, *the optimal decoder* Δ *that minimizes the mean square error (MSE)* $E_{\mathbf{x}}[\|\mathbf{x} - \Delta(\Phi\mathbf{x})\|_2^2] = \min_g E_{\mathbf{x}}[\|\mathbf{x} - g(\Phi\mathbf{x})\|_2^2]$, *as well as the mean absolute error (MAE)* $E_{\mathbf{x}}[\|\mathbf{x} - \Delta(\Phi\mathbf{x})\|_1] = \min_g E_{\mathbf{x}}[\|\mathbf{x} - g(\Phi\mathbf{x})\|_1]$, *where* $g : \mathbb{R}^M \to \mathbb{R}^N$, *is obtained with a linear MAP estimator,*

$$\Delta(\Phi\mathbf{x}) = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \underbrace{\Sigma\Phi^T(\Phi\Sigma\Phi^T)^{-1}}_{\Delta}(\Phi\mathbf{x}), \tag{2}$$

and the resulting error $\eta = \mathbf{x} - \Delta(\Phi\mathbf{x})$ *is Gaussian with mean zero and with covariance matrix* $\Sigma_\eta = E_{\mathbf{x}}[\eta\eta^T] = \Sigma - \Sigma\Phi^T(\Phi\Sigma\Phi^T)^{-1}\Phi\Sigma$, *whose trace yields the MSE of SCS.*

$$E_{\mathbf{x}}[\|\mathbf{x} - \Delta(\Phi\mathbf{x})\|_2^2] = Tr(\Sigma - \Sigma\Phi^T(\Phi\Sigma\Phi^T)^{-1}\Phi\Sigma). \tag{3}$$

Theory - Best-K linear approximation is as good as non-linear one

For Gaussian signals $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$, *where* $\mathbf{S} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ *whose eigenvalues* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ *decay fast, the best k-term linear approximation*

$$\mathbf{x}_k^l(m) = \begin{cases} \mathbf{x}(m) & 1 \leq m \leq k, \\ 0 & k+1 \leq m \leq N, \end{cases} \tag{8}$$

and the nonlinear approximation

$$\mathbf{x}_k^n = T_k(\mathbf{x}), \tag{9}$$

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*

$$\sigma_k^l(\{\mathbf{x}\})_X = E_{\mathbf{x}}[\|\mathbf{x} - \mathbf{x}_k^l\|_X] \quad \textit{and} \quad \sigma_k^n(\{\mathbf{x}\})_X = E_{\mathbf{x}}[\|\mathbf{x} - \mathbf{x}_k^n\|_X]. \tag{10}$$

Theory - GMM based Piecewise Linear Decoder

To decode a measured signal $\mathbf{y} = \Phi\mathbf{x}$, the GMM-based SCS decoder estimates the signal $\tilde{\mathbf{x}}$ and selects the Gaussian model $\tilde{j}$ by maximizing the log a-posteriori probability

$$(\tilde{\mathbf{x}}, \tilde{j}) = \arg\max_{\mathbf{x}, j} \log f(\mathbf{x}|\mathbf{y}, \Sigma_j). \tag{41}$$

(41) is calculated by first computing the linear MAP decoder (2) using each of the Gaussian models,

$$\tilde{\mathbf{x}}_j = \Delta_j(\Phi\mathbf{x}) = \underbrace{\Sigma_j \Phi^T (\Phi \Sigma_j \Phi^T)^{-1}}_{\Delta_j} (\Phi\mathbf{x}), \quad \forall 1 \le j \le J, \tag{42}$$

and then selecting a best model $\tilde{j}$ that maximizes the log a-posteriori probability among all the models [36]

$$\tilde{j} = \arg\max_{1 \le j \le J} -\frac{1}{2}\left(\log|\Sigma_j| + \tilde{\mathbf{x}}_j^T \Sigma_j^{-1} \tilde{\mathbf{x}}_j\right), \tag{43}$$

whose corresponding decoder $\Delta_{\tilde{j}}$ implements a piecewise linear estimate:

$$\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_{\tilde{j}} = \Delta_{\tilde{j}}(\Phi\mathbf{x}). \tag{44}$$

Alternating between E-step and M-step

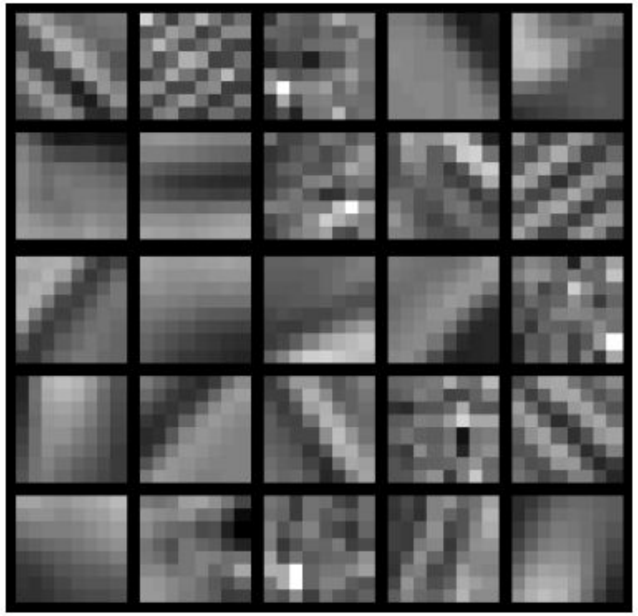
E-step: Assuming that the estimates of the Gaussian parameters are known(following the previous M-step), the E-step calculates the MAP signal estimation and model selection for all the signals, following MAP linear decoder and then maximizing the log a-posteriori.

M-step: Assuming that the Gaussian model selection j and the signal estimate x are known for all the signals (following the previous E-step), the M-step estimates (updates) the Gaussian models using the MLE method.

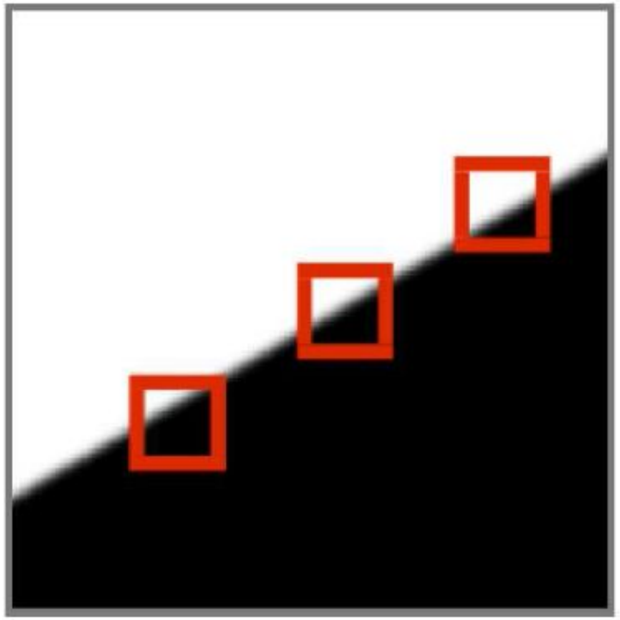
$$\tilde{\mu}_j = \frac{1}{|\mathscr{C}_j|} \sum_{i \in \mathscr{C}_j} \tilde{\mathbf{x}}_i \quad \text{and} \quad \tilde{\Sigma}_j = \frac{1}{|\mathscr{C}_j|} \sum_{i \in \mathscr{C}_j} (\tilde{\mathbf{x}}_i - \tilde{\mu}_j)(\tilde{\mathbf{x}}_i - \tilde{\mu}_j)^T .$$

MAP-EM Initialization

We randomly generate 19 angle values and construct small images for these 19 angles as shown in the fig. Then we take all the possible patches which cover both the white and black regions (i.e., near boundary) vectorize them after centring the signal, concatenate in matrix and take it's column wise covariance. This matrix is used as initialization for the covariance matrix for the 19 Gaussian models. The means are initialized as zero vectors.



Dictionary atoms learnt from Lena



A synthetic edge image. Patches (8×8) that touch the edge are used to calculate an initial covariance matrix.

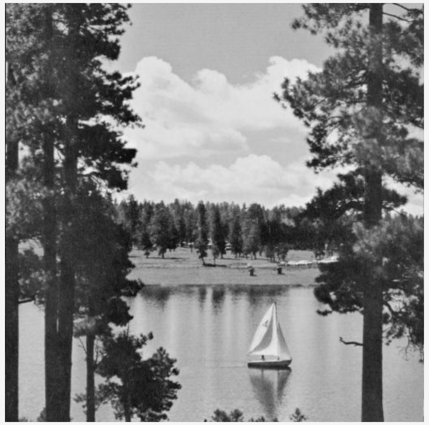
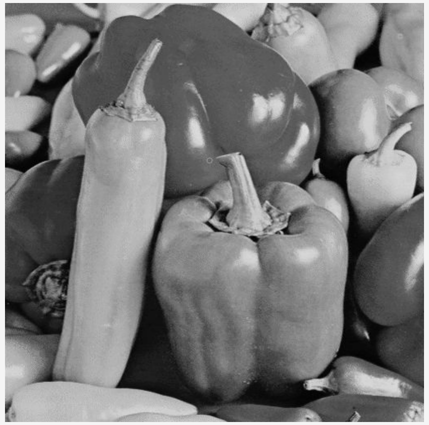
Experiments

- Lena, House, Lake and Peppers were used for comparison

- **SCS**: 8x8 overlapping patches from test images were sensed using random sensing matrices to obtain measurements of 6 different sizes for each patch followed by reconstruction using MAP-EM algorithm

- **Conventional CS**: An overcomplete dictionary was learned from 720,000 8x8 patches taken from the entire 300 images of Berkeley Segmentation dataset using K-SVD. 8x8 overlapping patches from test images were sensed using random sensing matrices to obtain measurements of 6 different sizes for each patch followed by reconstruction using ISTA



Results - SCS

ratio = 0.2

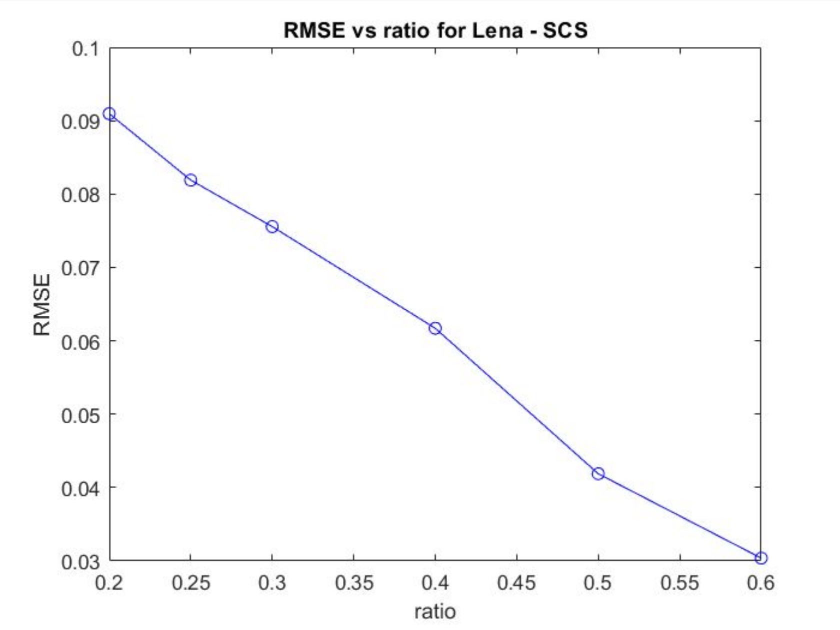
ratio = 0.25

ratio = 0.3

ratio = 0.4

ratio = 0.5

ratio = 0.6



Results - SCS

ratio = 0.2

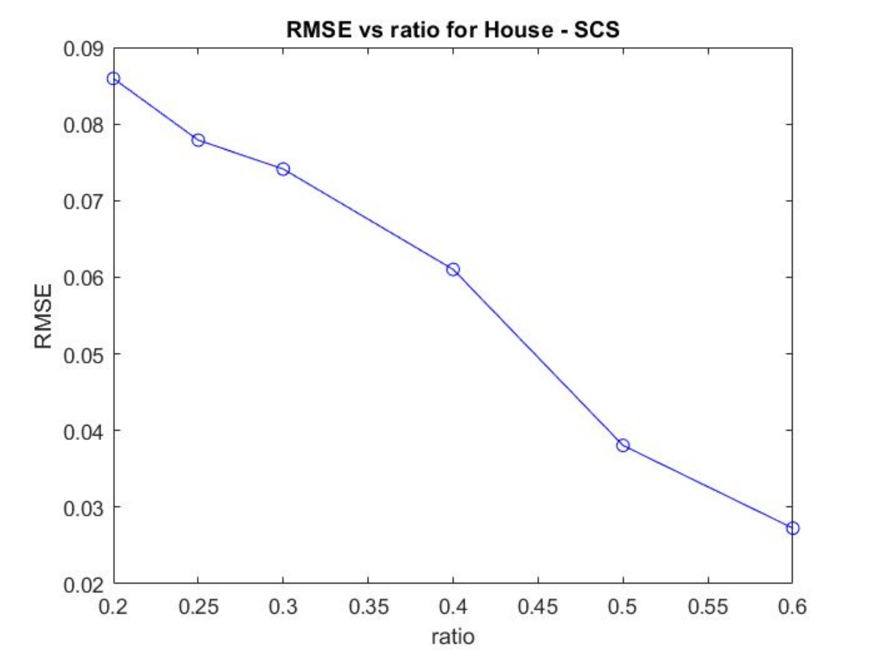
ratio = 0.25

ratio = 0.3

ratio = 0.4

ratio = 0.5

ratio = 0.6



Results - SCS

ratio = 0.2

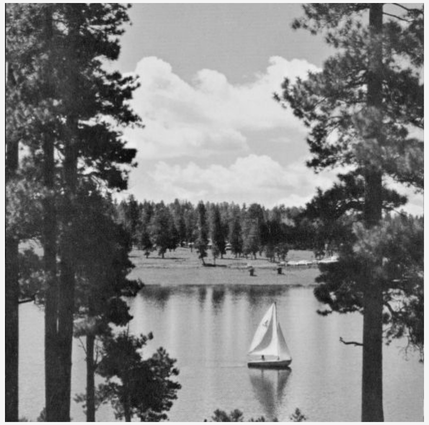
ratio = 0.25

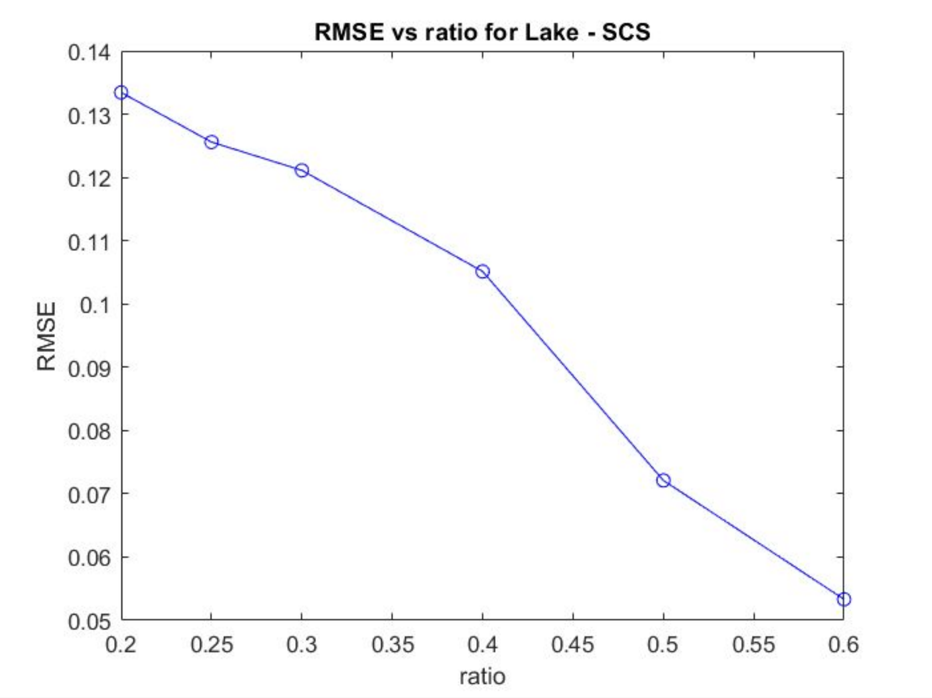
ratio = 0.3

ratio = 0.4

ratio = 0.5

ratio = 0.6





Results - SCS


ratio = 0.2

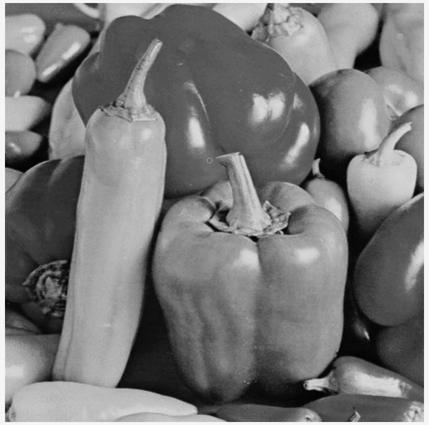

ratio = 0.25

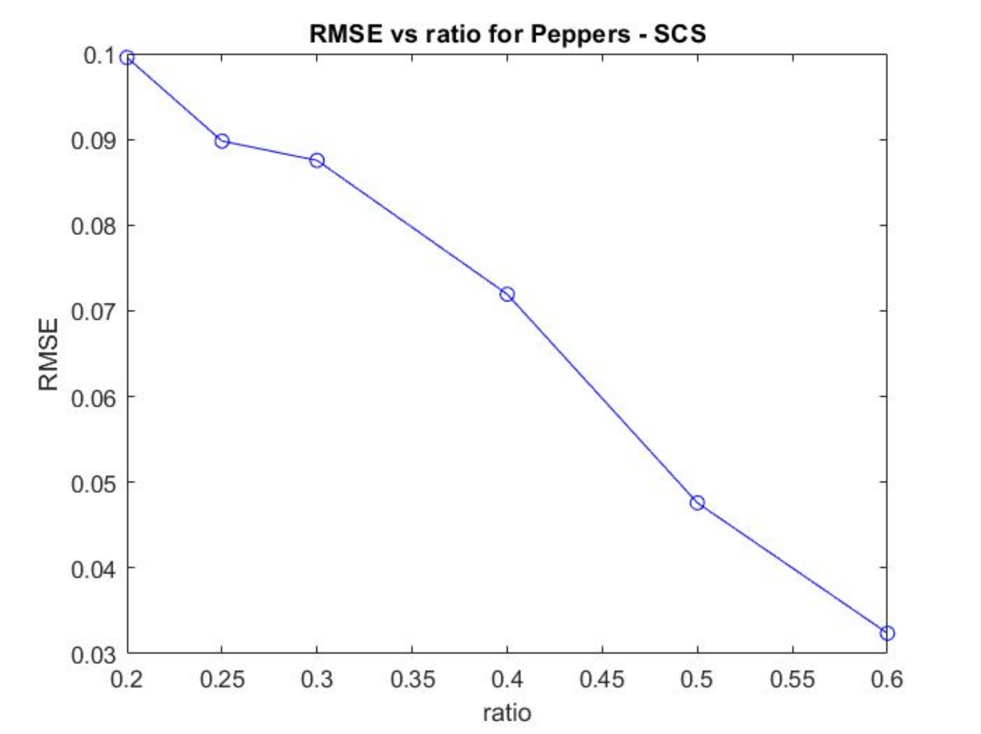

ratio = 0.3


ratio = 0.4


ratio = 0.5

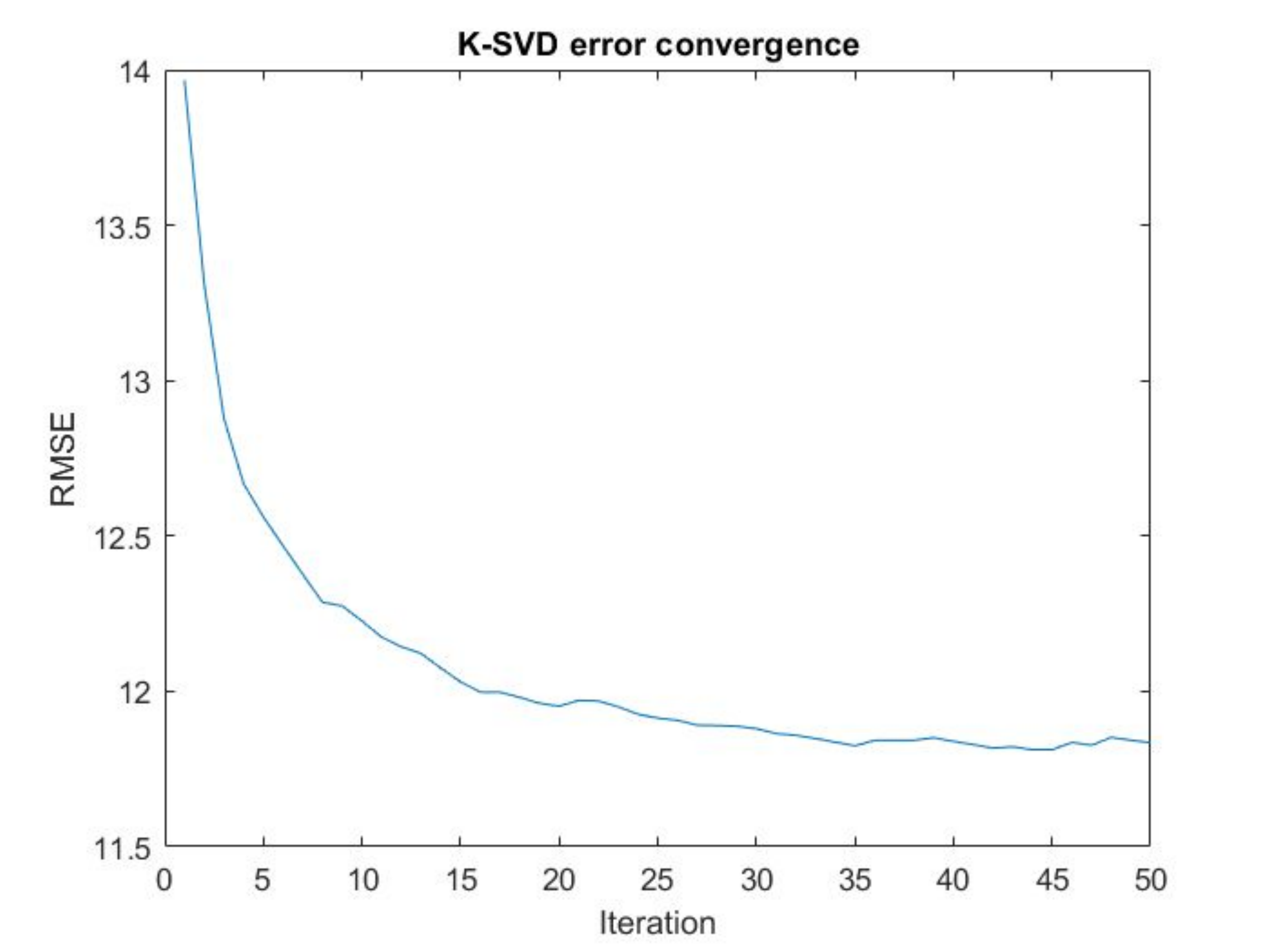

ratio = 0.6





K-SVD

- Overcomplete dictionary of size 64x225
- 720,000 64x1 training samples



Results – Conventional CS


ratio = 0.2

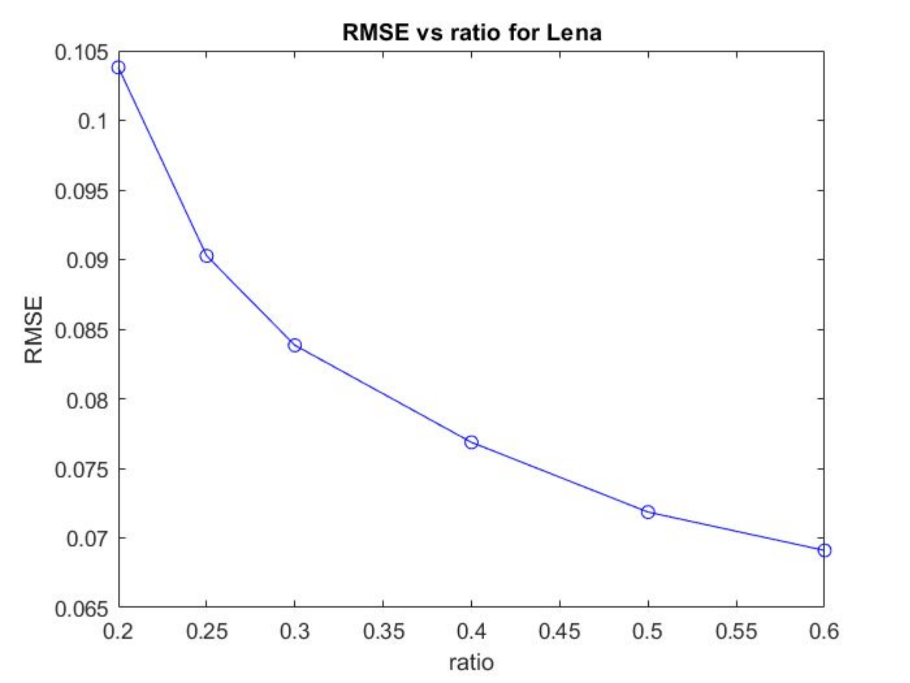

ratio = 0.25


ratio = 0.3


ratio = 0.4


ratio = 0.5


ratio = 0.6



Results – Conventional CS


ratio = 0.2

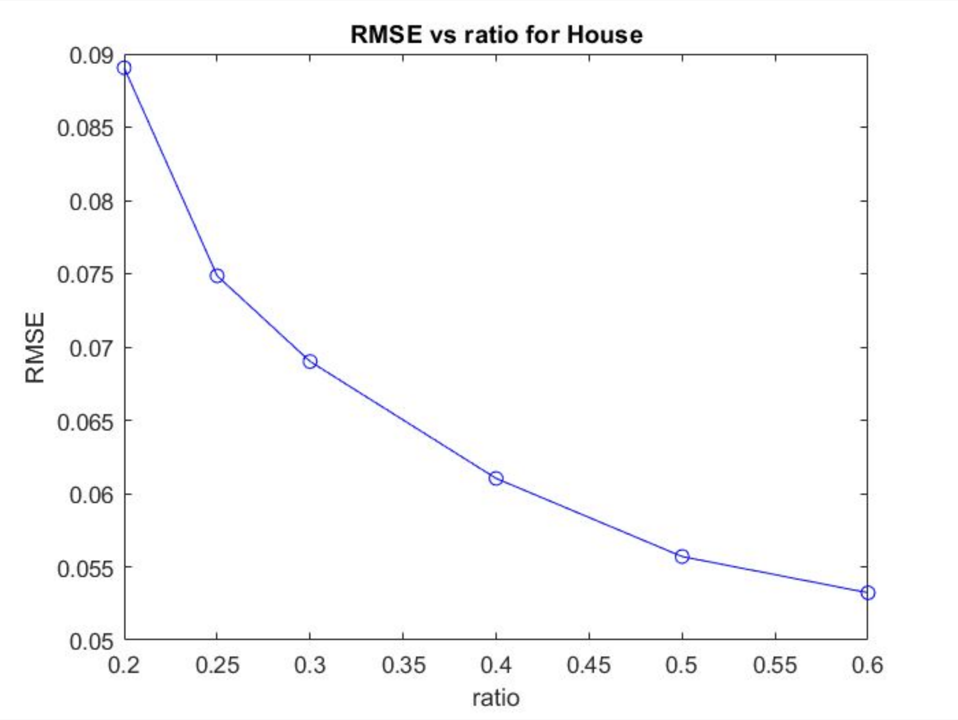

ratio = 0.25


ratio = 0.3


ratio = 0.4


ratio = 0.5


ratio = 0.6



Results – Conventional CS


ratio = 0.2

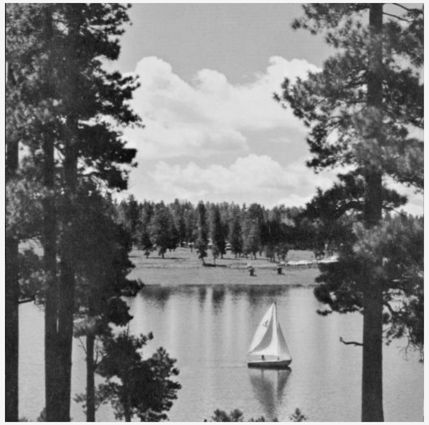

ratio = 0.25

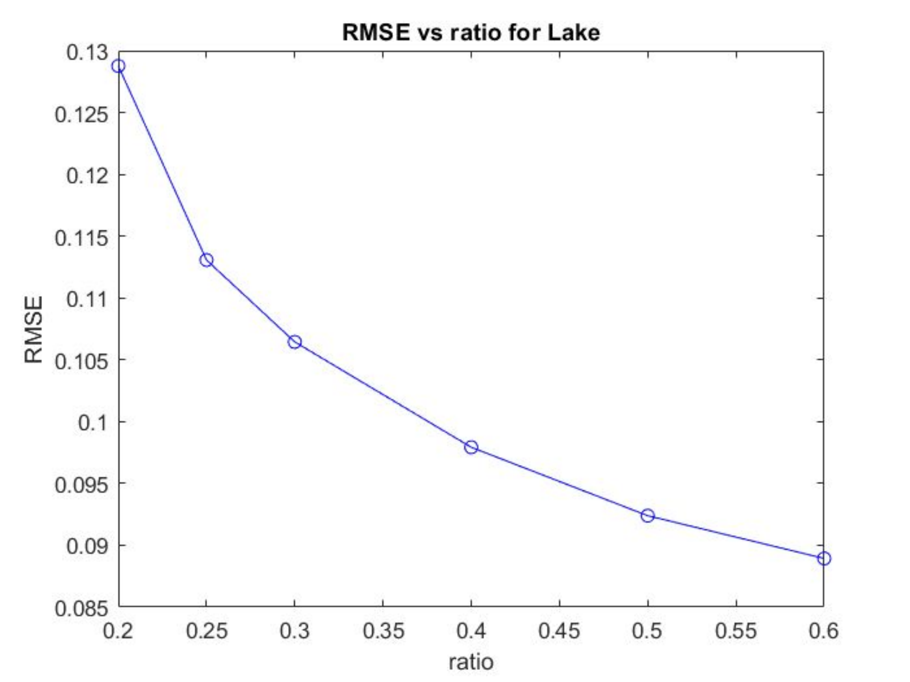

ratio = 0.3


ratio = 0.4


ratio = 0.5


ratio = 0.6





Results – Conventional CS

ratio = 0.2

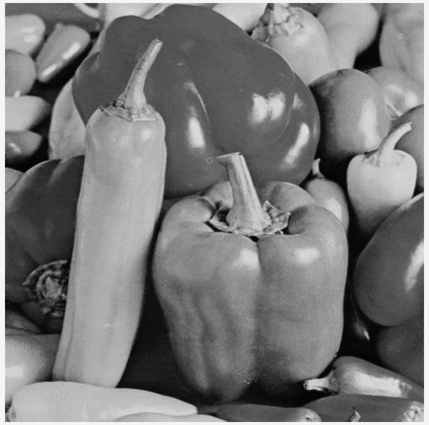
ratio = 0.25

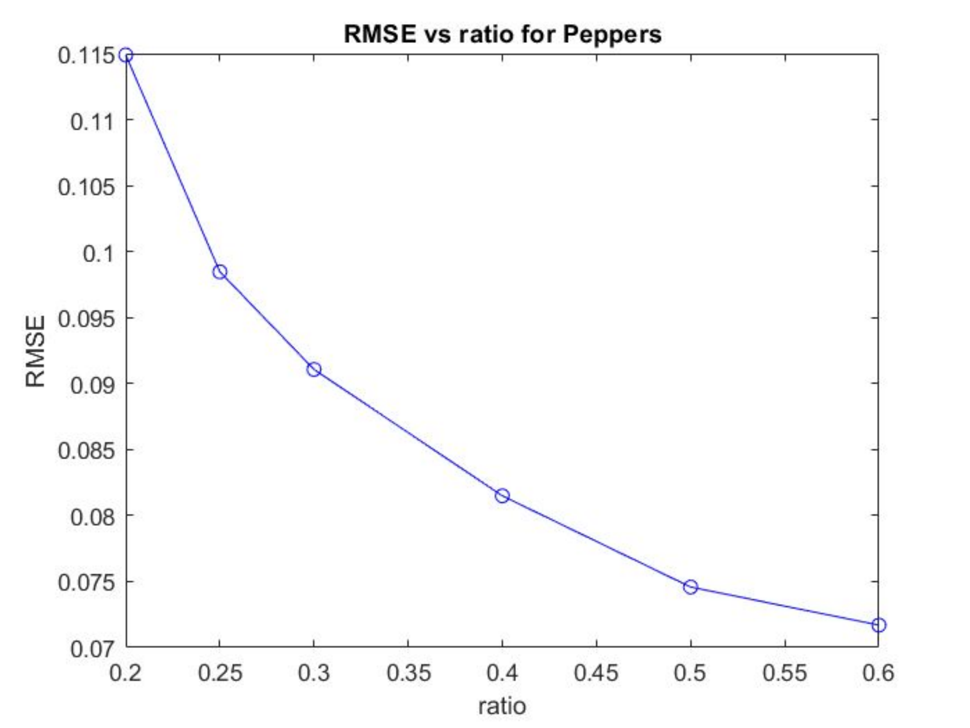
ratio = 0.3

ratio = 0.4

ratio = 0.5

ratio = 0.6





Results - Blind CS

ratio = 0.2

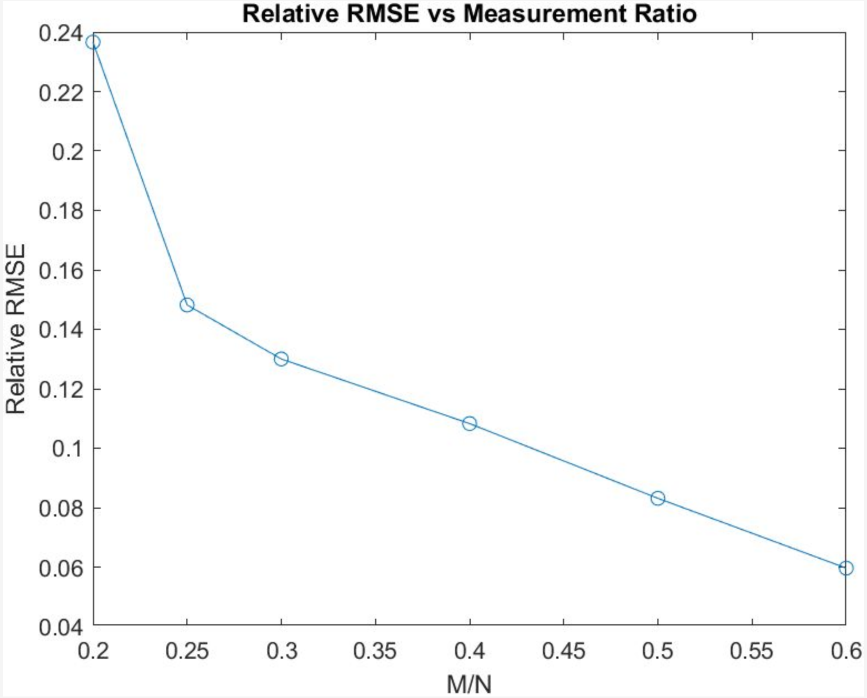
ratio = 0.25

ratio = 0.3

ratio = 0.4

ratio = 0.5

ratio = 0.6



Results - Blind CS

ratio = 0.2

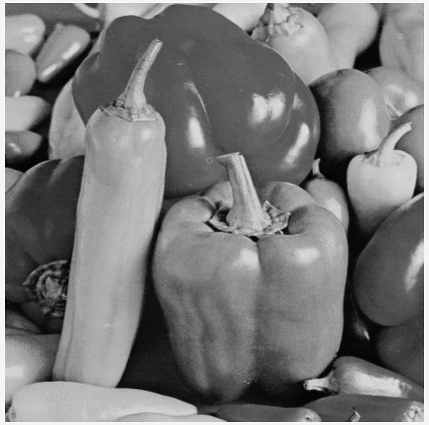
ratio = 0.25

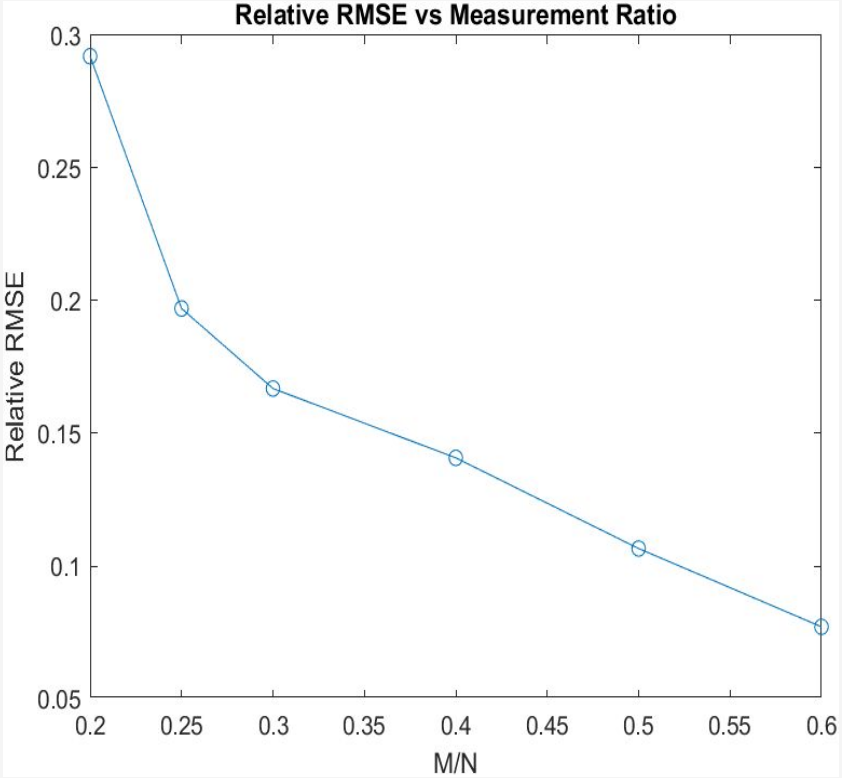
ratio = 0.3

ratio = 0.4

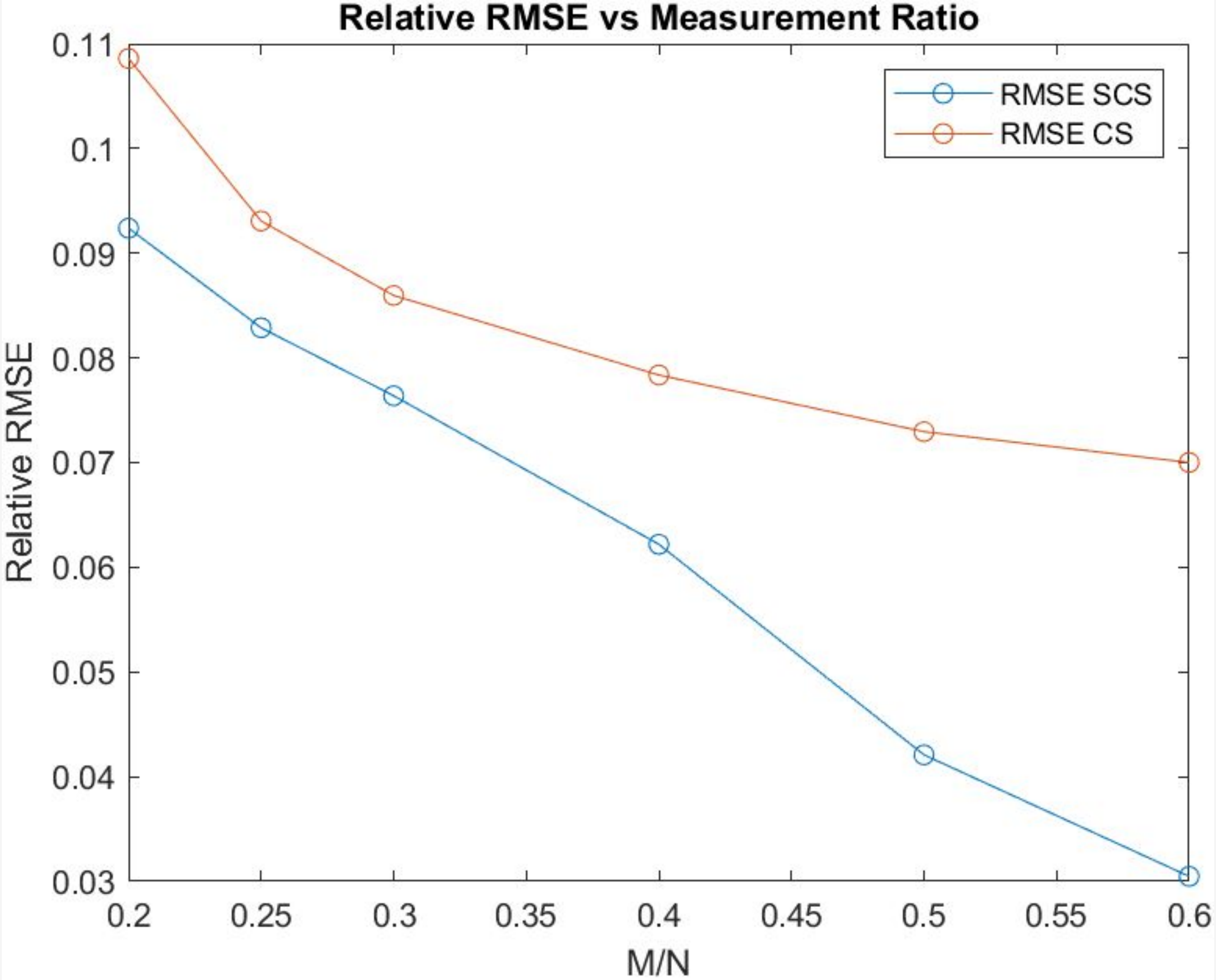
ratio = 0.5

ratio = 0.6

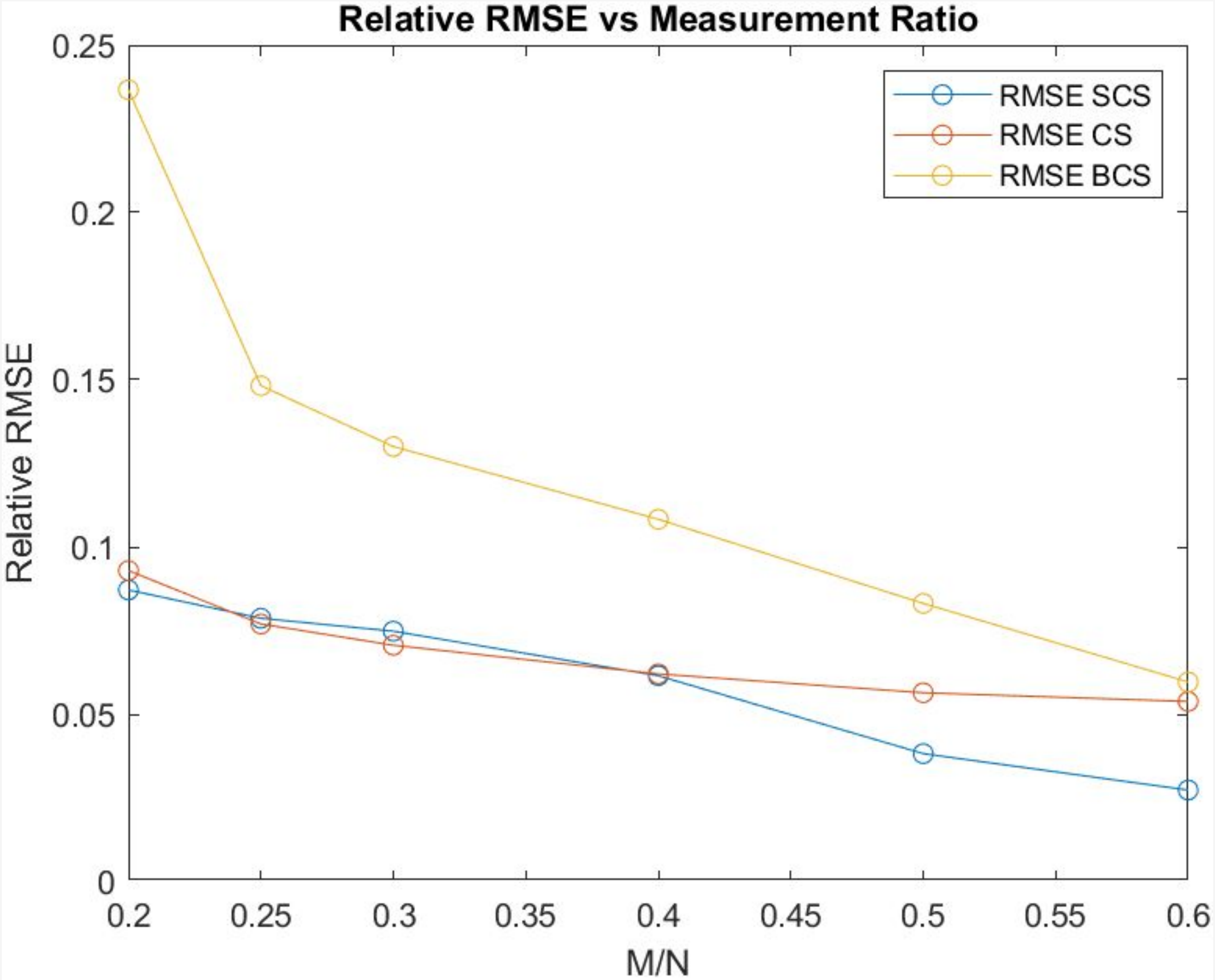




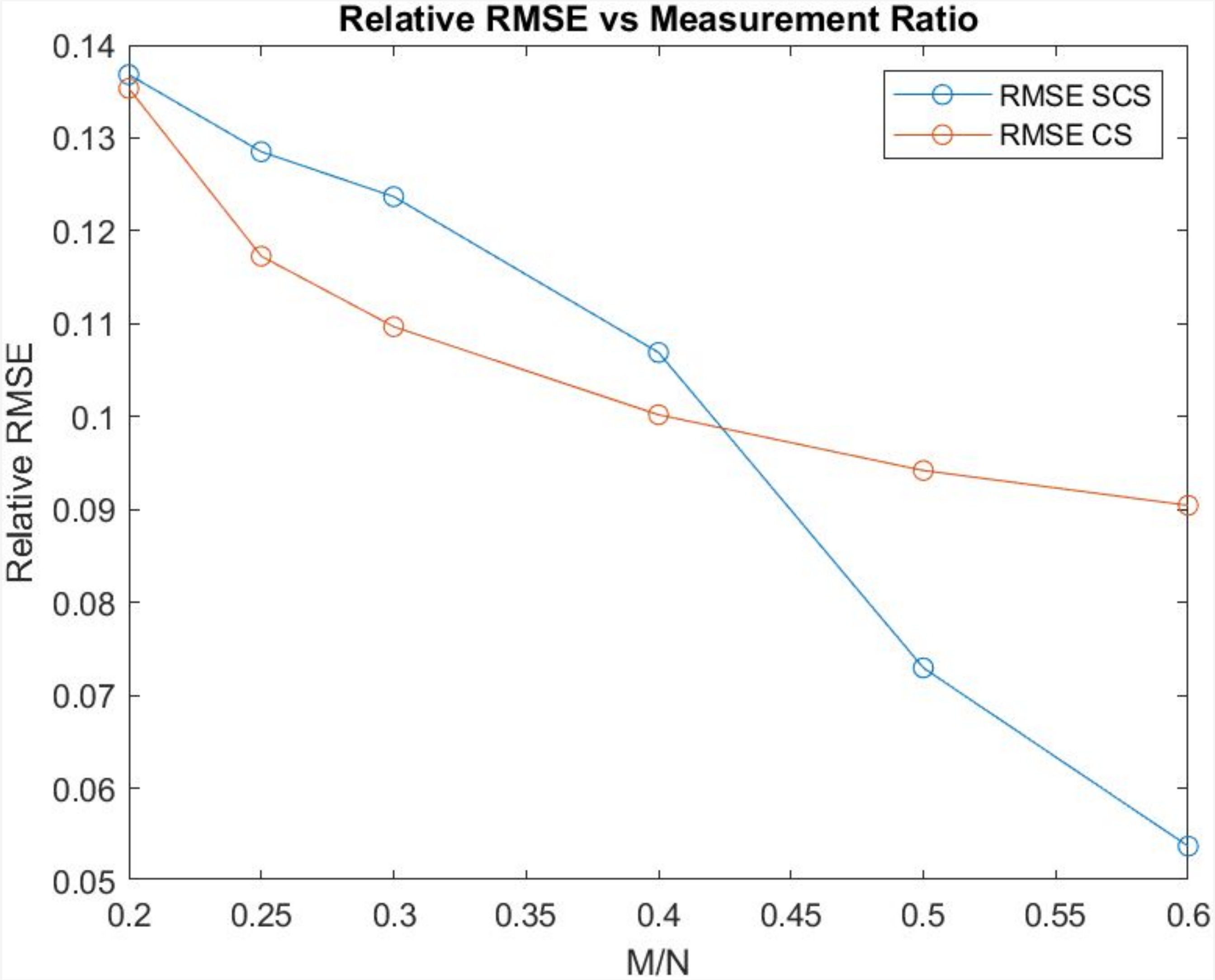
Comparison (Lena) – SCS vs CS



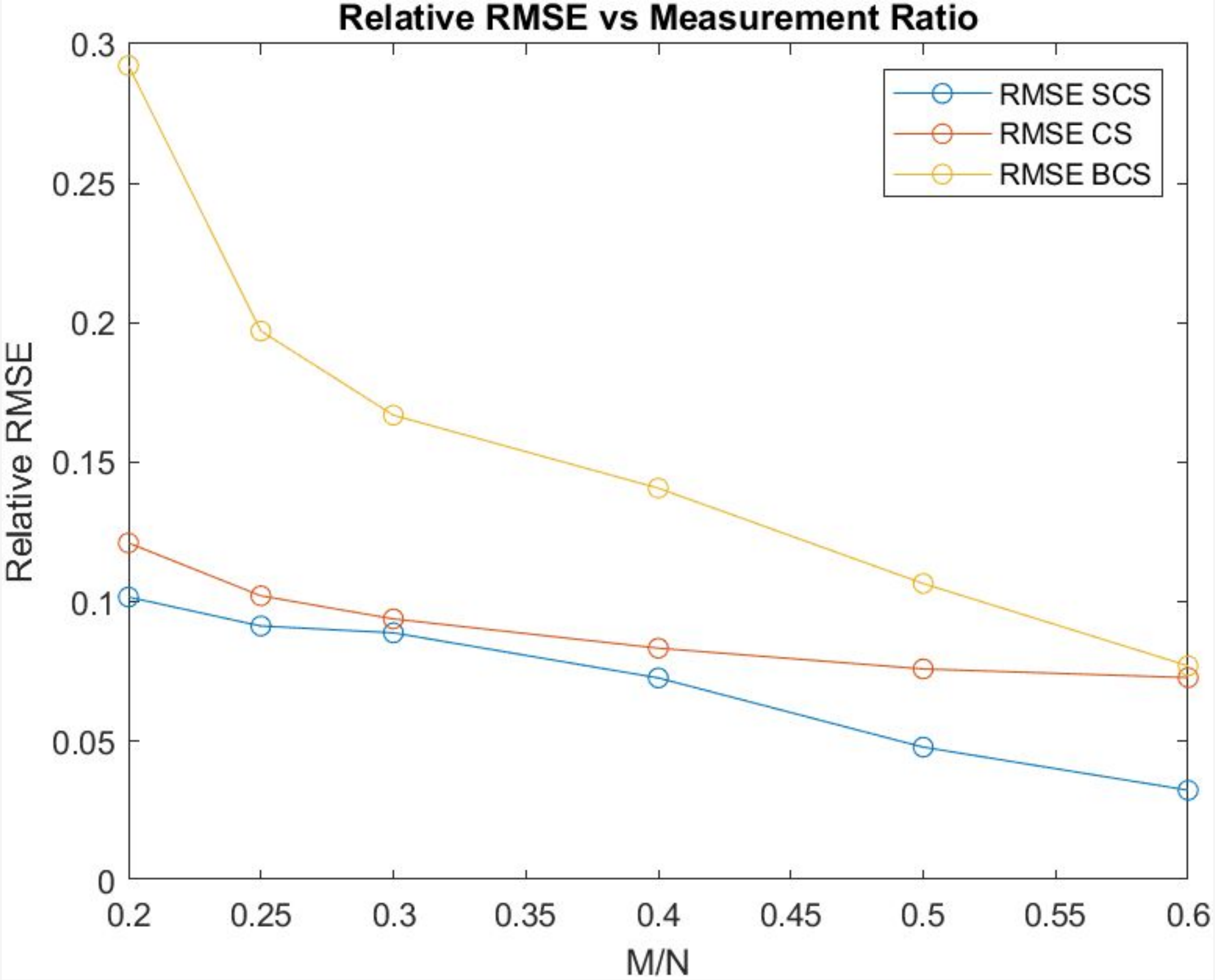
Comparison (House) – SCS vs CS vs BCS



Comparison (Lake) – SCS vs CS



Comparison (Peppers) – SCS vs CS vs BCS



Comparison (Lena)

First Column - True image patch

Second Column - SCS reconstruction

Third Column - CS reconstruction

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