Bayesian Compressive Sensing and Analysis

Navaneetha Krishna Madan Gopal

1005171127

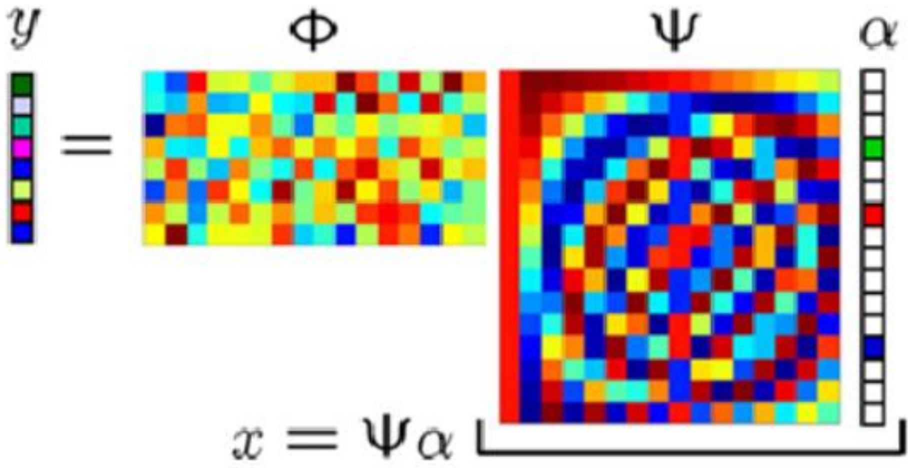
Compressive Sensing – The story

- Shannon-Nyquist’s theorem
- Data overflow problem – due to “sense then compress” philosophy
- “Compress while sensing philosophy” of CS

Math Behind Compressive Sensing

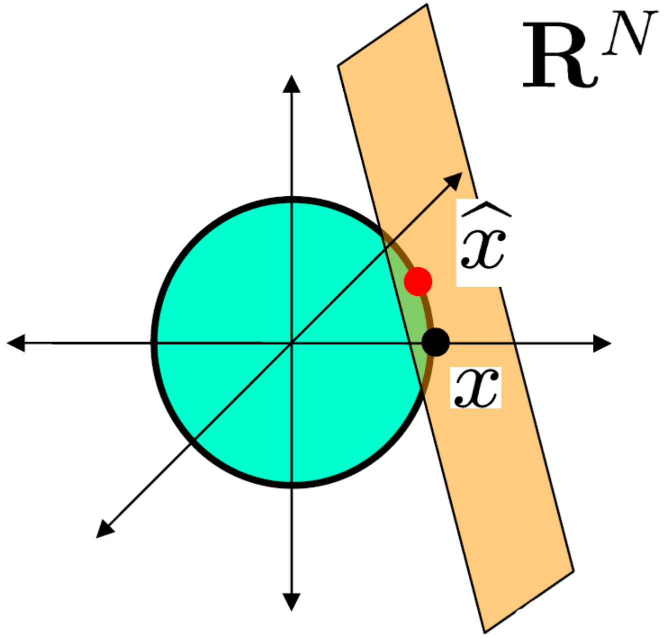
- Sparse and compressible signals
- Sensing matrix

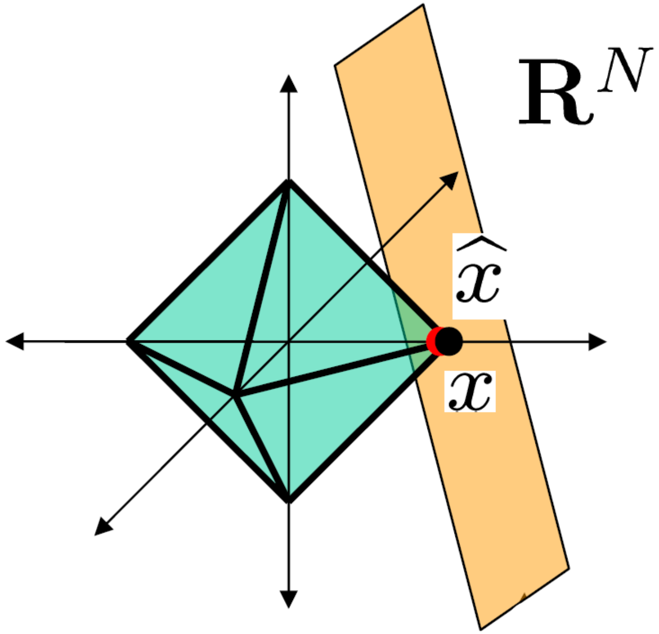
$$y = \phi x + n$$



$$min\|x\|_1 \text{ subject to } \|y - \phi x\|_2 \leq \epsilon$$

Comparison between Norms



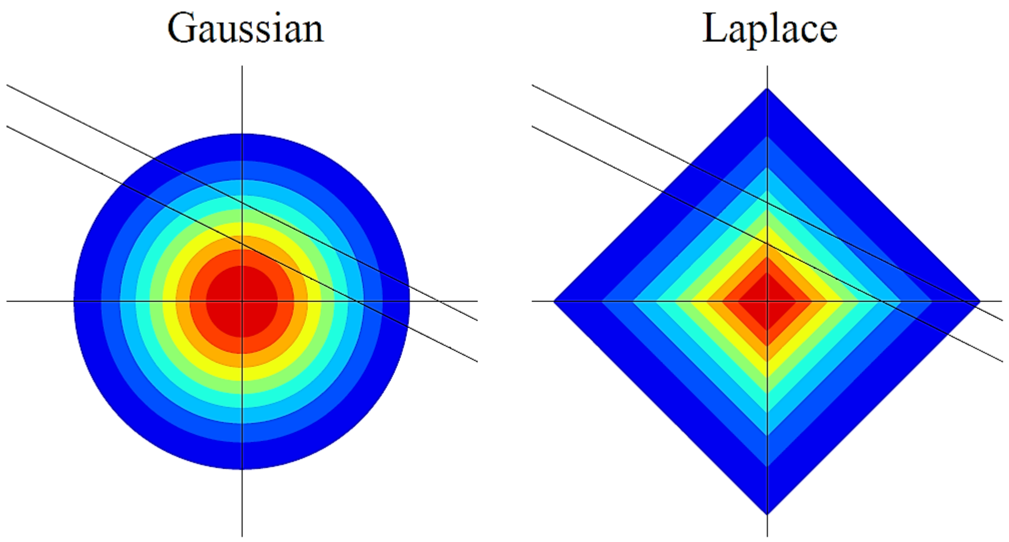


Approaches of Compressive Sensing

- Iterative relaxation
  - Use linear programming
  - Basis pursuit, gradient descent
- Greedy algorithms
  - Local optimum to find global optimum
  - Matching pursuit, orthogonal matching pursuit
- Bayesian models
  - Utilize the information that the vector is sparse
  - Relevance Vector Machine approach

Choice of Prior for Data

- Let **w** be a sparse vector which is to be estimated
- **w** is given a Laplace prior as this prior is sparsity promoting



Tractable Choice of Hierarchical Prior

- Gaussian noise and the Laplacian sparse vector are not conjugate
- Leads to computational complexity
- A hierarchical gaussian prior is used

$$p(\boldsymbol{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{N} \mathcal{N}(w_i|0, \alpha_i^{-1})$$

Problem statement

- The model is defined as follows $\mathbf{t} = \phi\mathbf{w} + \mathbf{n}$
- **t** is a Kx1 vector of observations, **w** is an Nx1 sparse vector, phi is a KxN matrix and **n** is a Kx1 noise vector
- Precision of noise **n** is represented as α_0
- The pdf of t is given by

$$p(\mathbf{t}|\mathbf{w},\sigma^2) = (2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}||\mathbf{t}-\mathbf{\Phi}\mathbf{w}||^2\right\}$$

Sensing Matrix generation

- One of the advantages of Bayesian approach is that a correlated sensing matrix can also yield good estimates of the sparse signal
- Hence correlation is introduced into the system by pre-multiplying the sensing matrix by a correlation matrix ρ

Goal

- To estimate the precisions of the entries of the sparse vectors and the noise variance

Procedure

- From the theory of Relevance Vector Machine, the mean vector and covariance matrix of the signal vector as is given as:

$$\begin{aligned} \boldsymbol{\mu} &= \alpha_0 \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \boldsymbol{g}, \\ \boldsymbol{\Sigma} &= (\boldsymbol{A} + \alpha_0 \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \end{aligned}$$

Iterative method

- Calculate initial value of μ and Σ
- A parameter ϒ which measures the “well-determinedness” of the sample at a given index location us defined as

$$\gamma_i \equiv 1 - \alpha_i \Sigma_{ii}$$

- From this, the new precision is calculated as

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2}$$

Iterative method (continued)

- The noise variance is calculated as

$$(\sigma^2)^{\text{new}} = \frac{\|\mathbf{t} - \mathbf{\Phi}\boldsymbol{\mu}\|^2}{N - \sum_i \gamma_i}$$

- The covariance matrix and mean vectors are recalculated using these updates values of noise variance and vector of precisions

Optimization

- The values of ϒ for the signal components is checked after every iteration
- When they(it) go(es) below a certain value (≈ 10^(-12)), the corresponding rows and columns of Σ and the mean vector are removed
- The corresponding columns in the sensing matrix are removed
- This improves speed and accuracy

Advantages over other methods

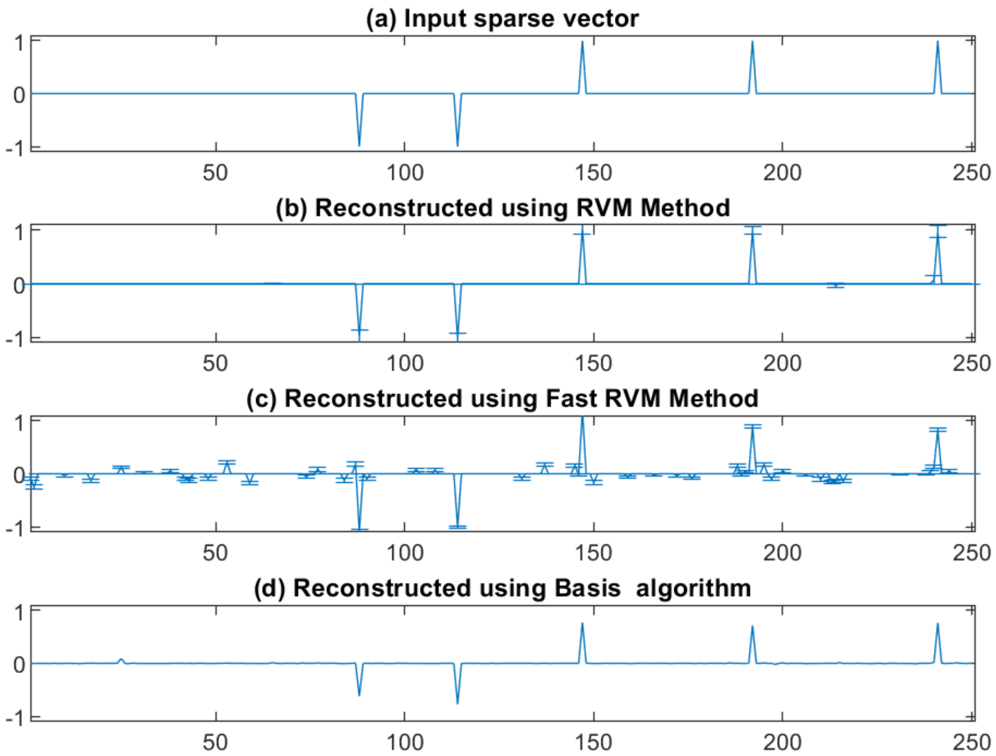
- Full posterior estimate of the sparse vector is obtained instead of just a point estimate
- Reconstruction fidelity is maintained high even in the presence of correlation

Problem Definition

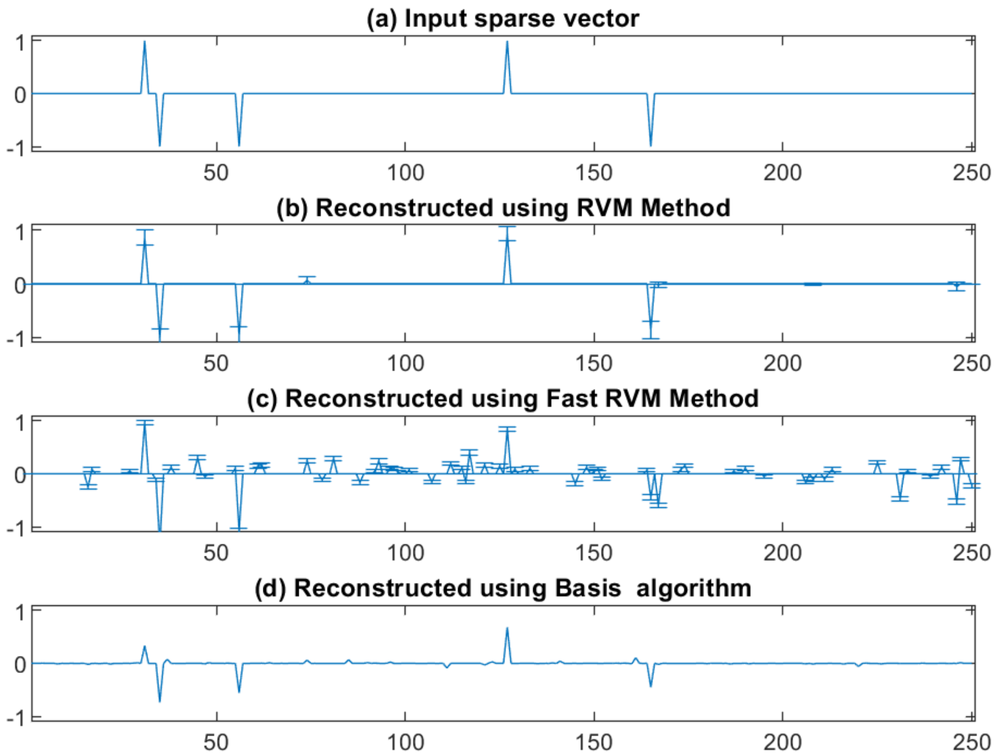
- Sparse vector of length 250 –> N = 250
- This is sampled 100 times -> K = 100
- 5 spikes in the vector -> S = 5
- The input is reconstructed using the three approaches and the following results are obtained

Results

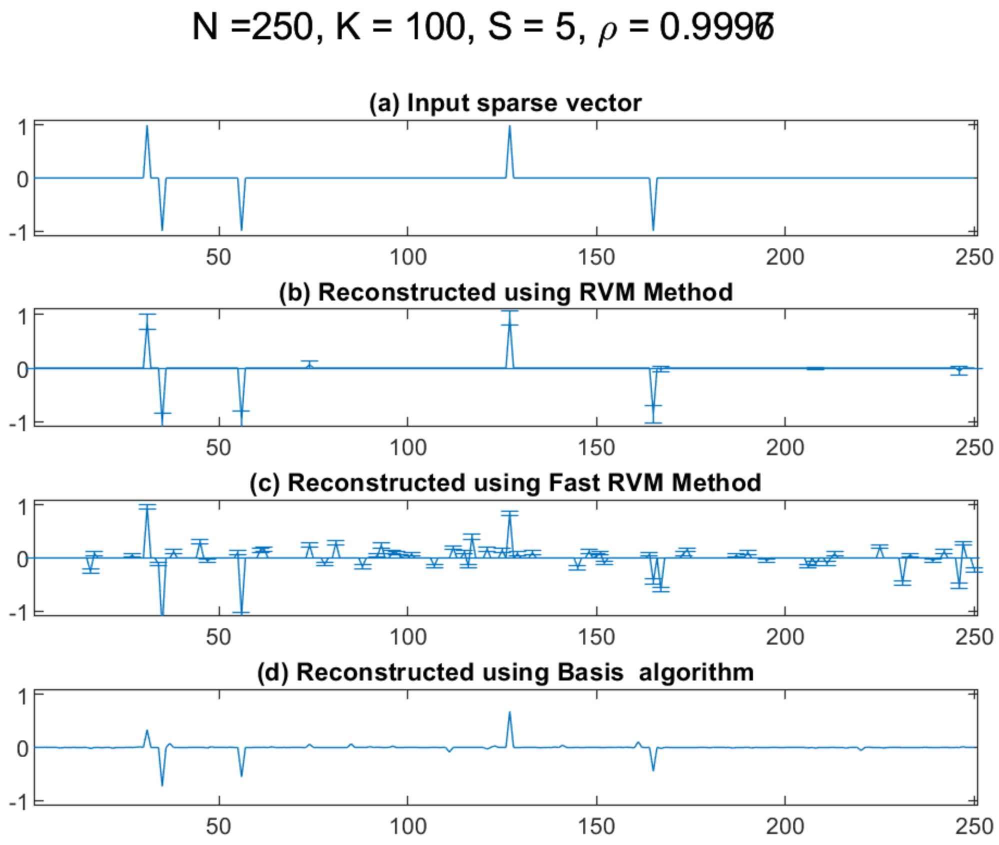
N =250, K = 100, S = 5, ρ = 0.9995

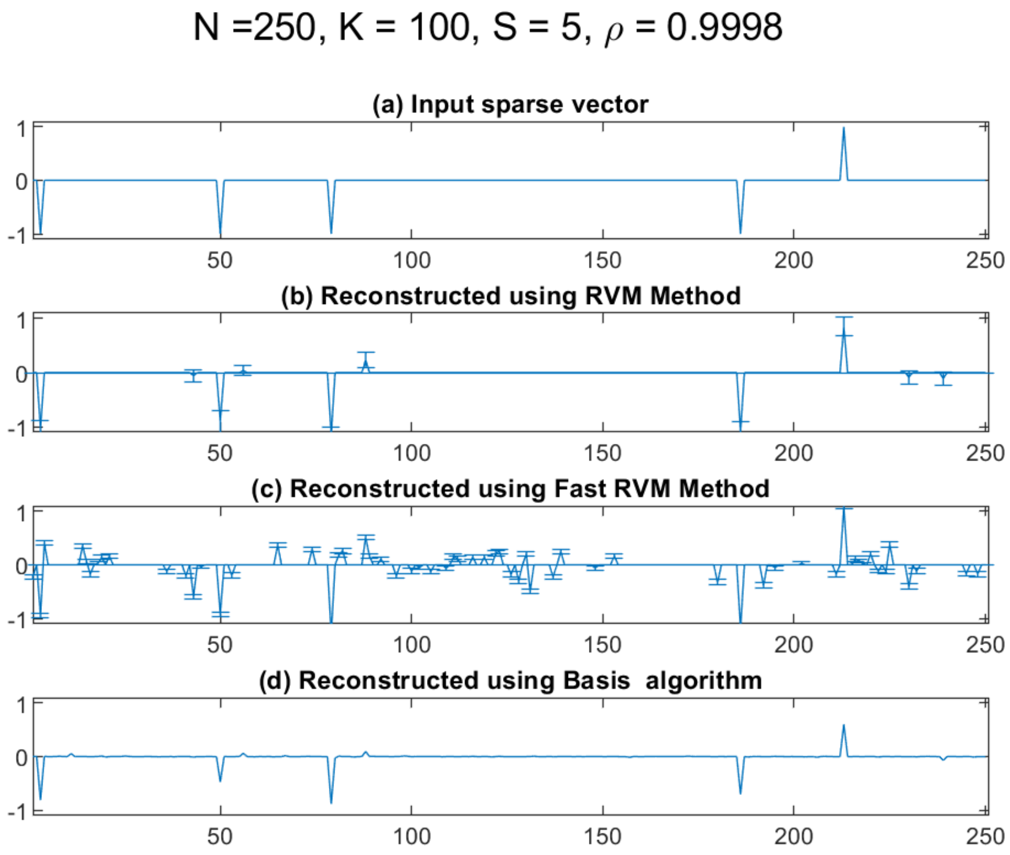


N =250, K = 100, S = 5, ρ = 0.9996

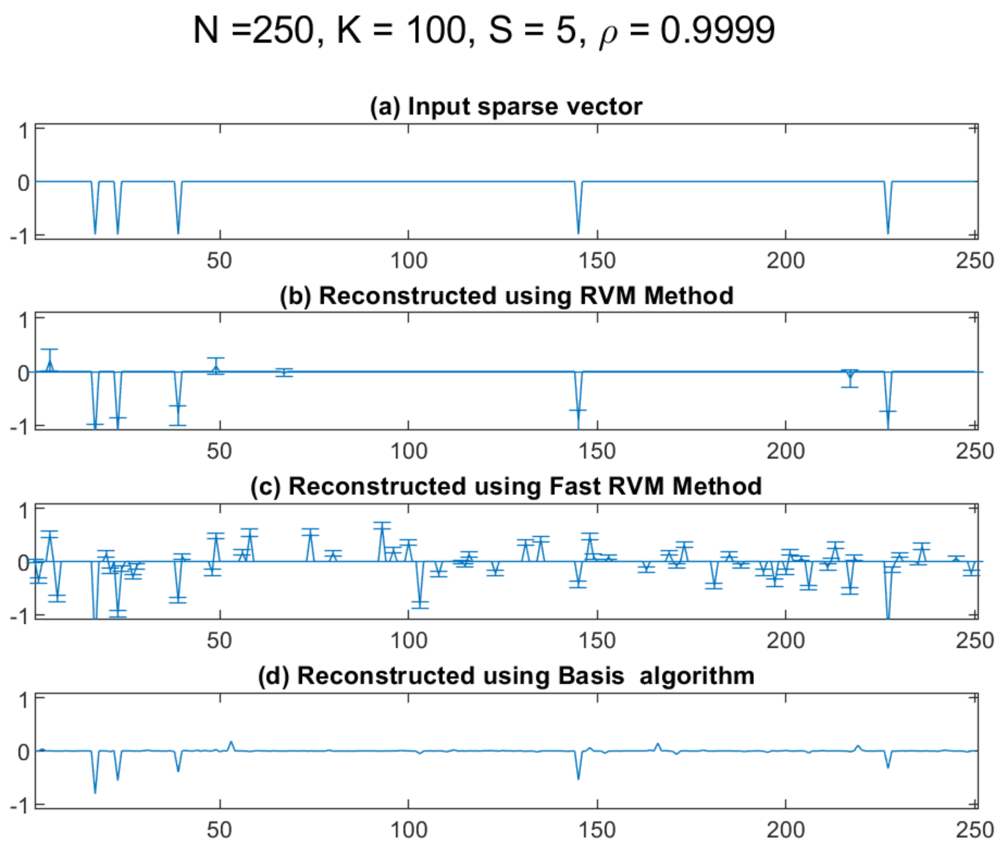


Results(continued)

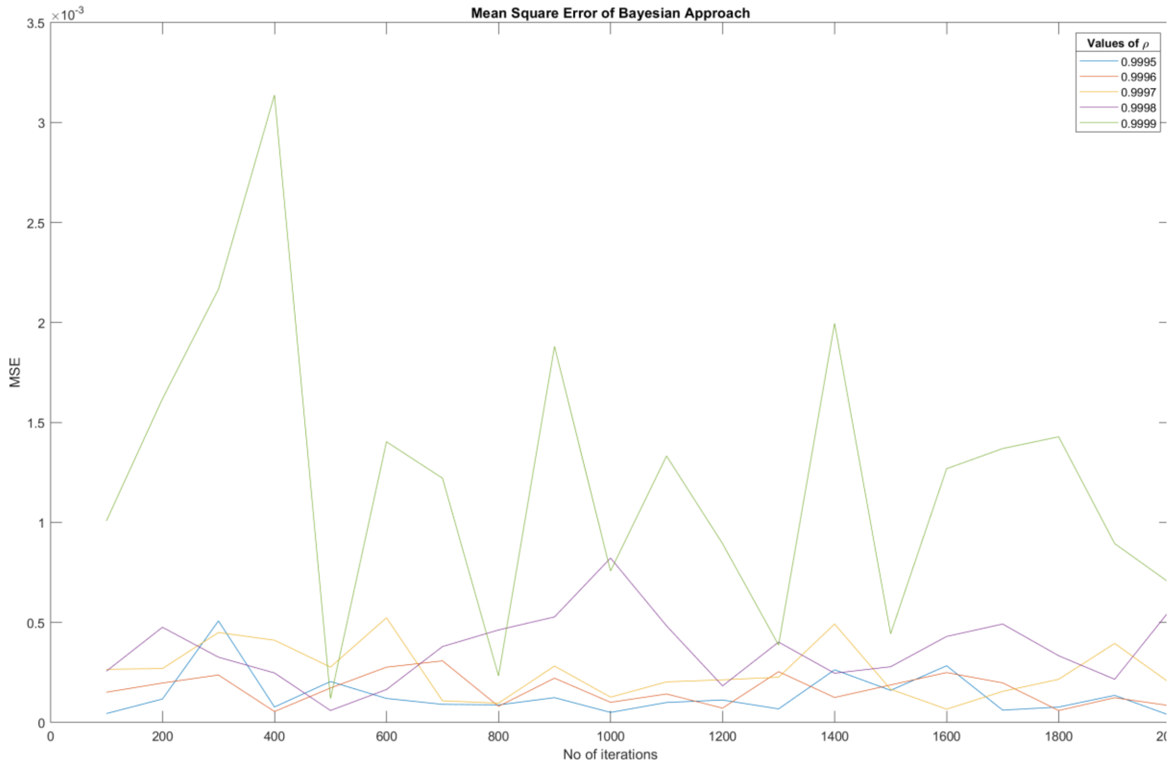


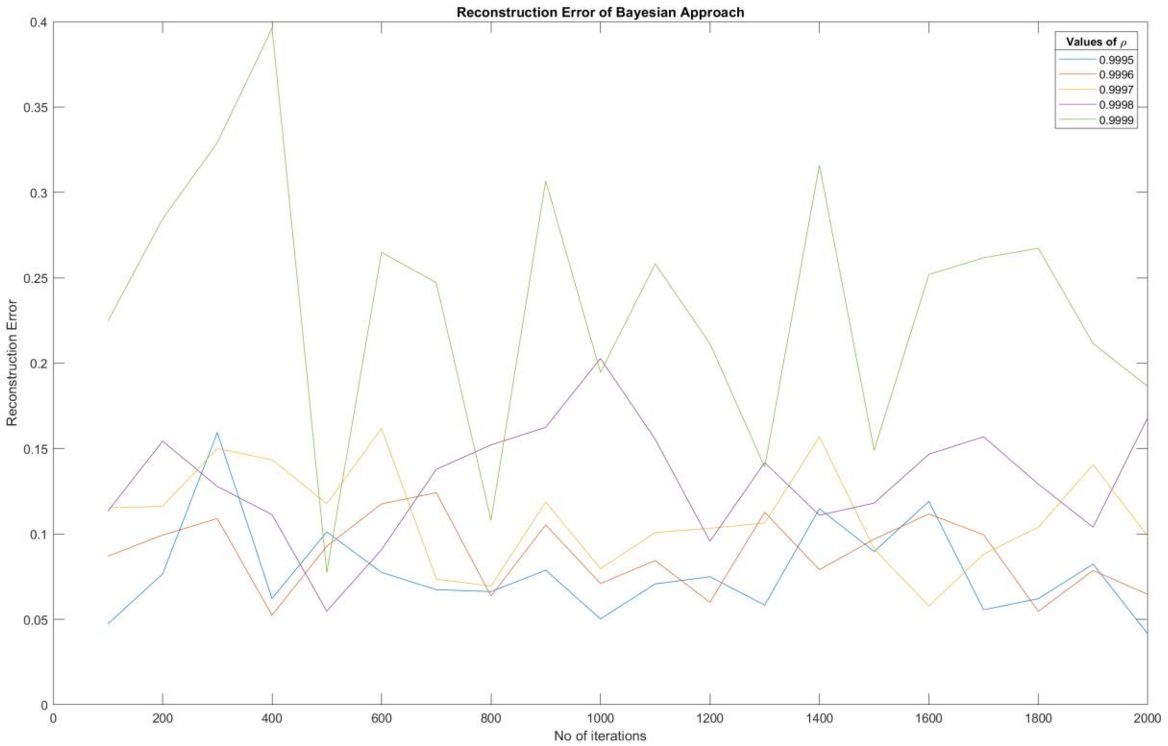


Results(continued)

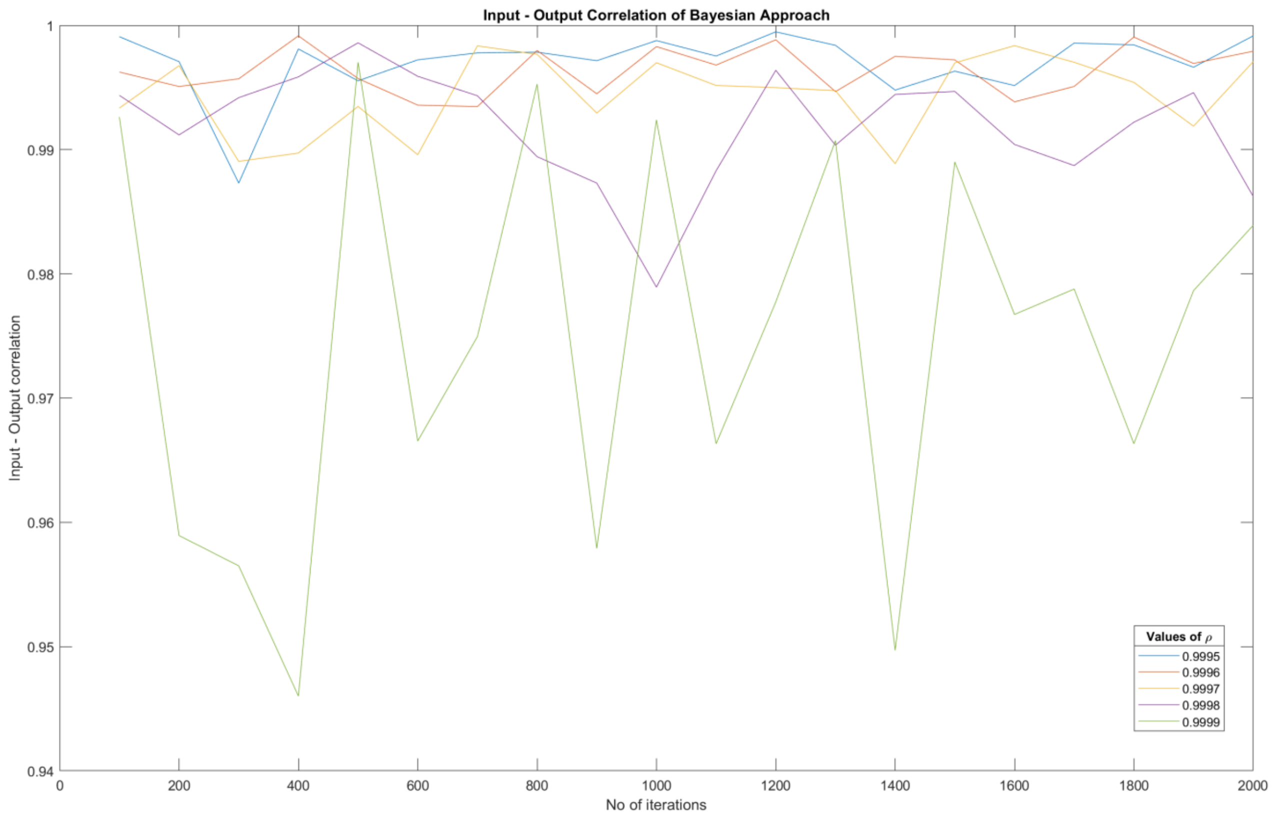


Performance parameters

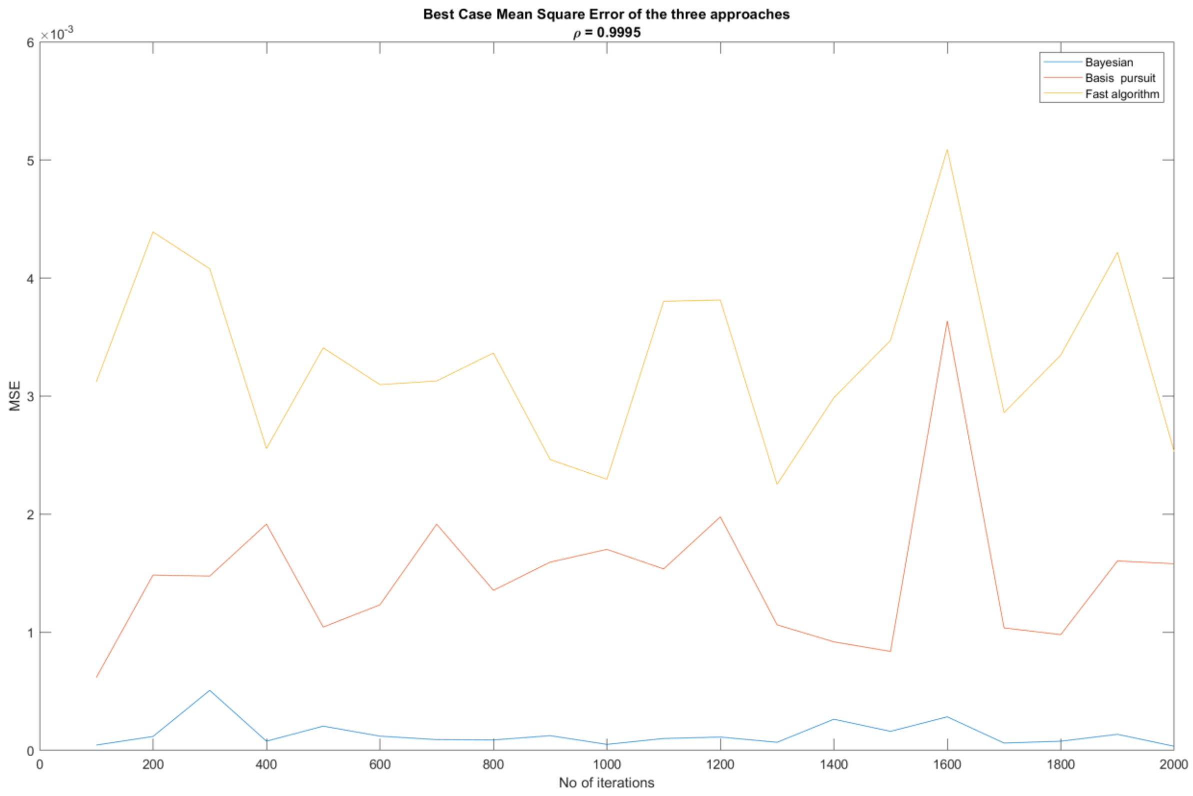


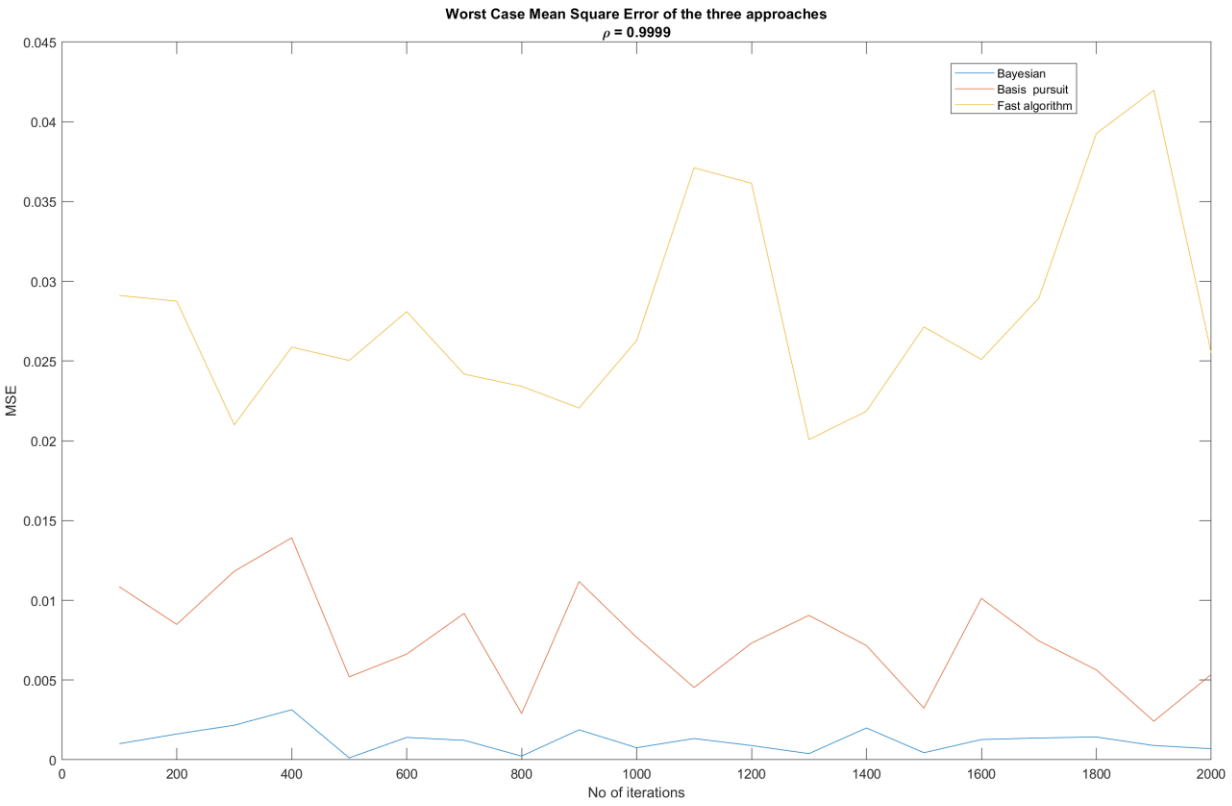


Performance Parameters(Continued)

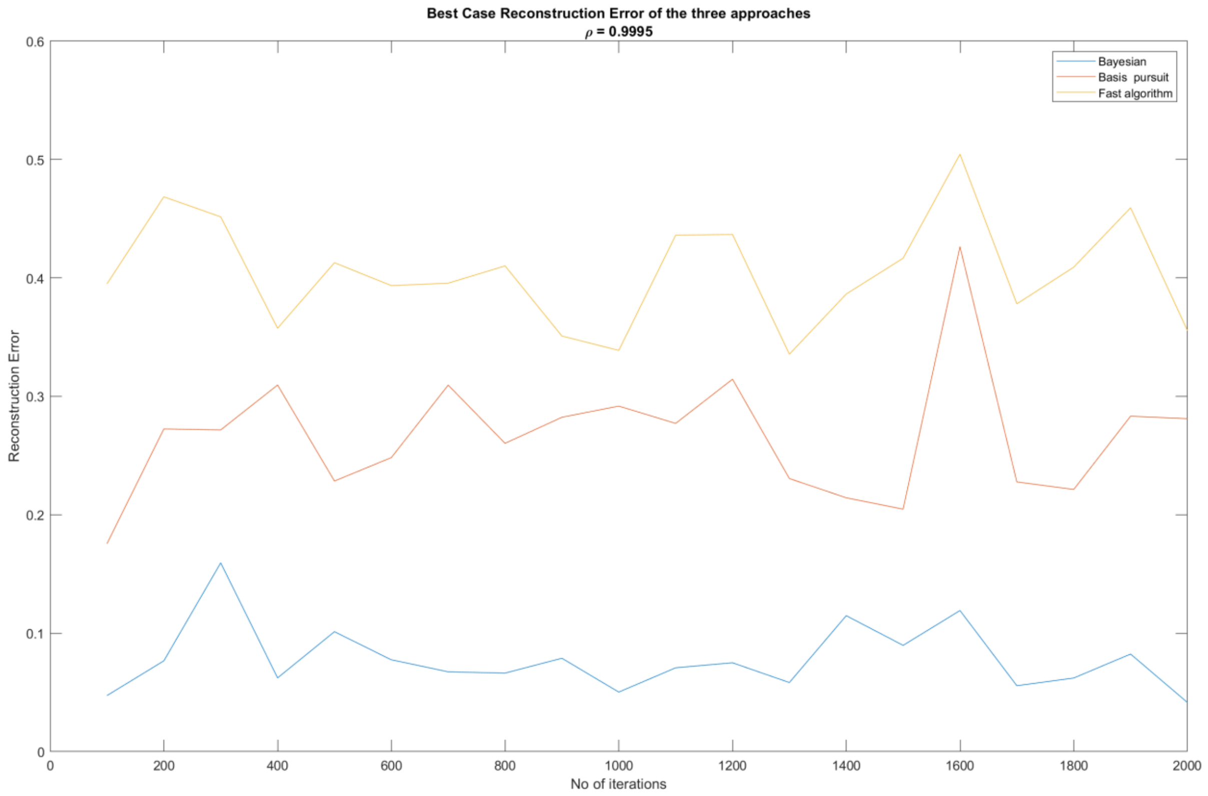


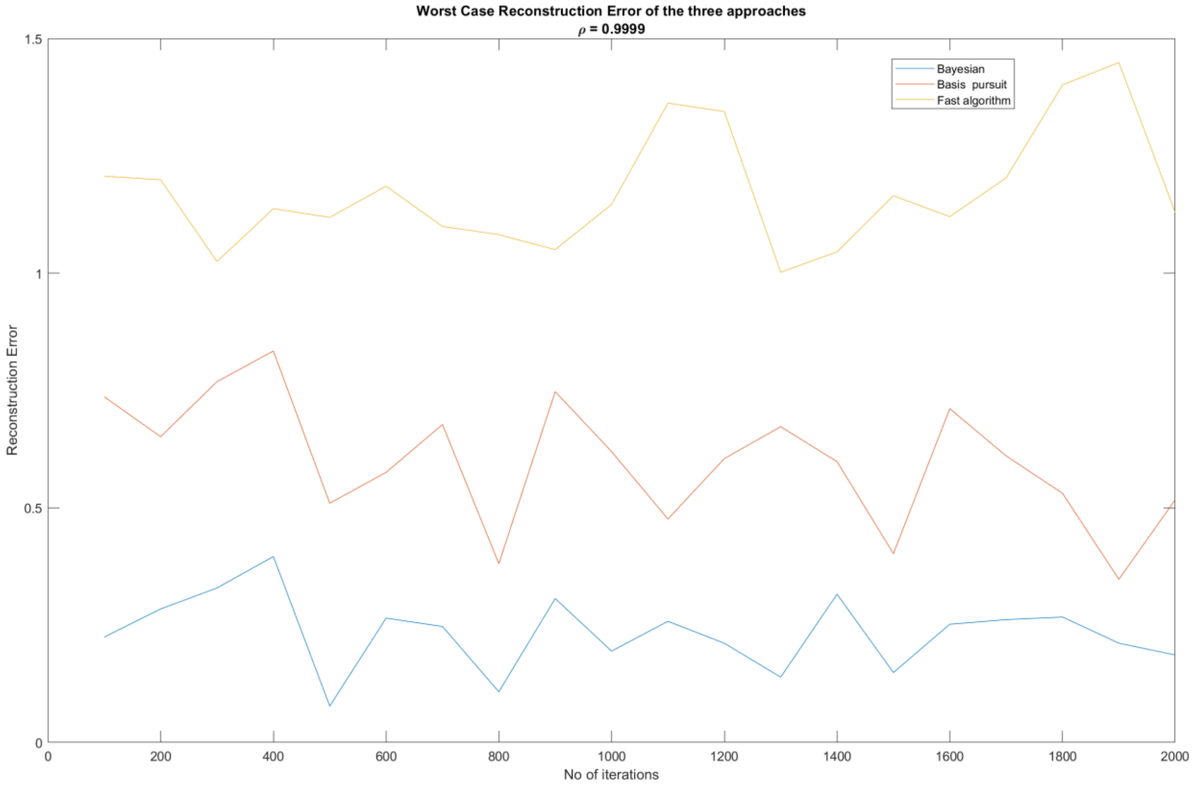
Best Case – Worst Case Comparisons



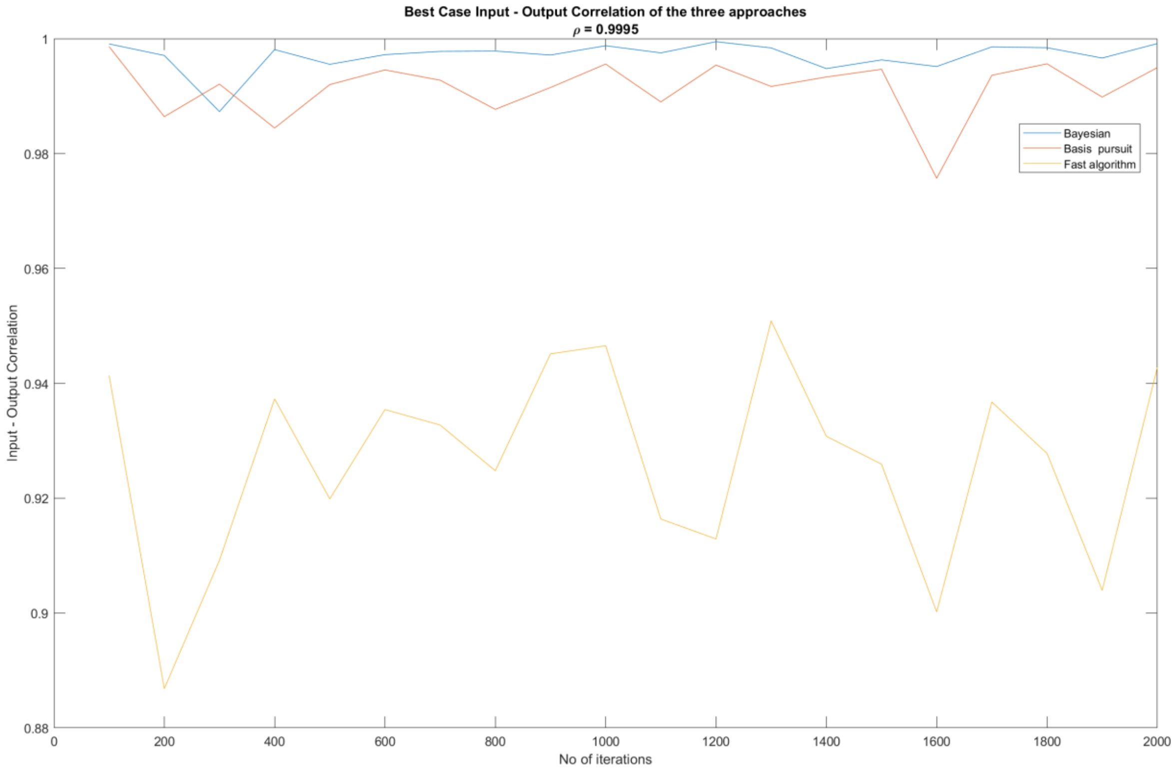


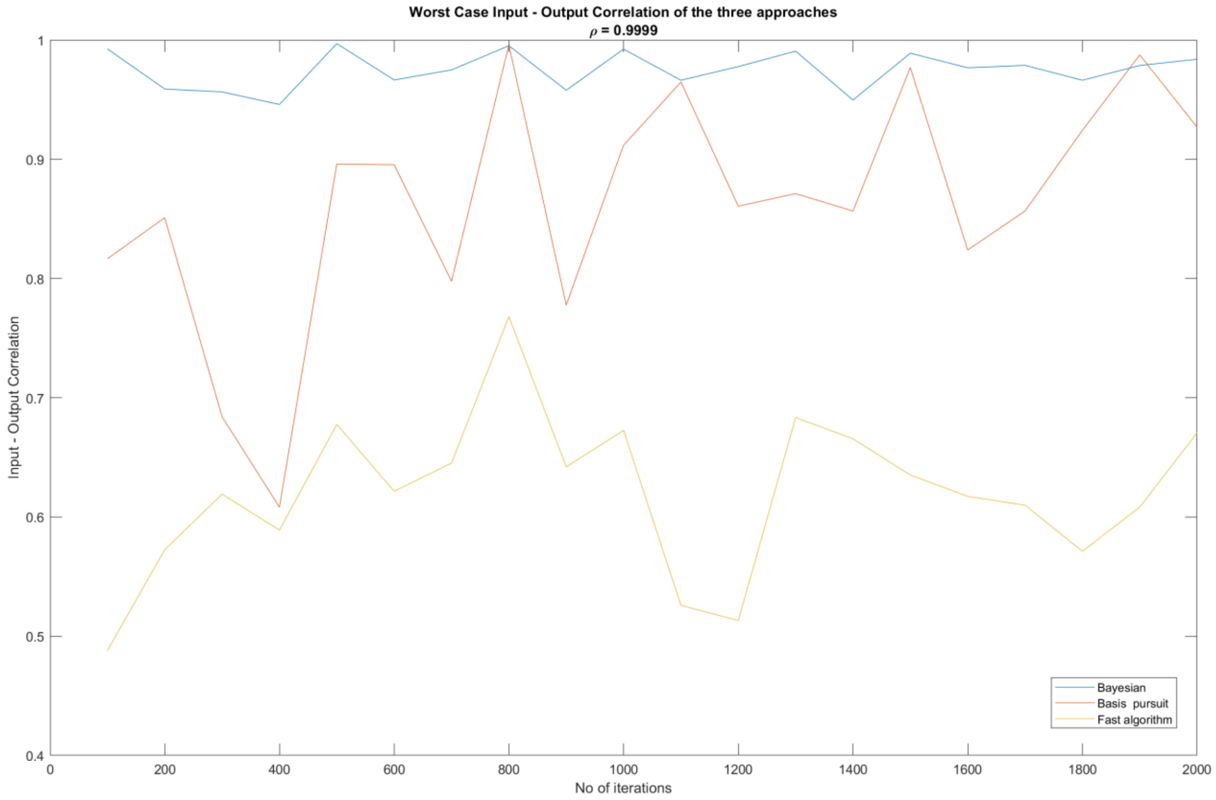
Best Case – Worst Case Comparisons





Best Case – Worst Case Comparisons





UFOs – Unexplainable Funky Observations

- The number of iterations to reach steady state is a random number with a very large variance
- This improves as the correlation is reduced
- The number of iterations required to converge to 20% more data in the final mean vector increases asymptotically with the initial value of the noninformative prior of the sparse vector

References

- Ji, Shihao, Ya Xue and Lawrence Carin. “Bayesian Compressive Sensing.” IEEE Transactions on Signal Processing 56 (2008): 2346-2356.
- MatthiasW. Seeger, “Bayesian Inference and Optimal Design for the Sparse Linear Model”, Journal of Machine Learning Research 9 (2008) 759-813, 2008
- F.Salahdine, N.Kaabouch, and H. El Ghazi, "Bayesian Compressive Sensing with Circulant Matrix for Spectrum Sensing in Cognitive Radio Networks," IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference, 2016

References (Continued)

- Compressive sensing - Microsoft research
- Justin Romberg – Lecture series, Tsinghua University

Thank you! Questions?

Goal

- To compute the posterior densities to estimate the noise variance and sparse vector

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) = \frac{p(\mathbf{t} | \mathbf{w}, \boldsymbol{\alpha}, \sigma^2) p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{t})}$$

where

$$p(\mathbf{t}) = \int p(\mathbf{t} | \mathbf{w}, \boldsymbol{\alpha}, \sigma^2) p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2) \, d\mathbf{w} \, d\boldsymbol{\alpha} \, d\sigma^2$$

Tractability

- The integration is not tractable, hence the posterior is expressed as
- Hence the posterior is expressed as

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) = p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2 | \mathbf{t})$$

- The first term is expressed as

$$\begin{aligned} p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) &= \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)}, \\ &= (2\pi)^{-(N+1)/2} |\boldsymbol{\Sigma}|^{-1/2} \ \exp\left\{ -\frac{1}{2} (\mathbf{w} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu}) \right\} \end{aligned}$$

Tractability (Continued)

- The second term in the posterior in approximated to be the mode of the parameters
- This has been proven experimentally to be accurate