Measurement Matrix Design for Compressd Sensing: Sample Distortion Framework and Modulated Matrix

Mike E Davies, Chunli Guo

Institute for Digital Communications &
Joint Institute for Signal and Image Processing
University of Edinburgh, UK

Talk Outline

- Introduction
- Sample Distortion (SD) framework

  -definition, lower bound and convexity
- SD in Wavelet Statistical Image Model

  - bandwise sampling

  - sample allocation with tree structure
- Modulated Matrix Design

  - matrix structure, 1-D state evolution dynamics
- Two Block Matrix

  - first order phase transition, relationship with seeded matrix
- Conclusion

Stochastic CS Setting: $Y = \Phi X$

- Express signal X as a draw from a probabilistic model:

$$p_x(X) \propto \prod_{i=1}^{N} p(x_i)$$

- Appropriate $p(x)$ for compressible signal:

  -**heavy tail** -**peak at origin**

e.g. **Gaussian mixture model (GM)**

$$p(x) = \lambda N(0, \sigma_L^2) + \left(1 - \lambda\right) N(0, \sigma_S^2)$$

- **Bayesian optimal Approximation Message Passing (BAMP)**

$$\hat{X} = E(X \mid Y)$$

CS Imaging

- We focus on **natural images in wavelet domain**
- Is i.i.d Gaussian matrix optimal for CS imaging?

  No! Nature images have more properties

  e.g. exponential energy decay, tree structure….
- The matrix we want to design

  - **block diagonal**

  - tractable way to distribute samples for each block
- Our solution

  sample distortion function and sample allocation

Sample Distortion Framework

Given an i.i.d. source $X = [x_1, x_2, \dots, x_n]^T, x_i \sim p(x_i)$

Setup:

undersampling ratio: $\delta \triangleq m/n, m < n$

linear measurement encoder: $\Phi \in \mathbb{R}^{m \times n}$

nonlinear decoder: $\Delta(\Phi X)$

Then we define the ***l_2* *Sample Distortion (SD) function***

$$D_\Delta(\delta) = \inf_{\delta, \Phi} \frac{1}{n} E \| X - \Delta(\Phi_\delta X) \|_2^2$$

For **i.i.d random encoder-BAMP decoder** , SE predicts

$$D_{k+1} = E(X^2) - E\left[F^2(X + Z\sqrt{\frac{D_k}{\delta}}; \frac{D_k}{\delta}) \right]$$

$Z \sim N(0,1)$ and F is the MMSE optimal scalar denoising estimator

SD Lower Bound

- **Entropy Based Bound (EBB c.f. Shannon RD lower bound)**

Let $x_i \sim p(x_i)$, $\mathrm{var}(x_i) = 1$, $h(x_i) < \infty$ then

$$D_{\mathrm{EBB}}(\delta) \geq (1-\delta) 2^{2(h(x_i)-h_g)/(1-\delta)}$$

$h(x_i)$ -entropy of $p(x_i)$ h_g - entropy of unit Gaussian

For Gaussian source $D_{EBB}(\delta) = 1 - \delta$

- **Model Based Bound (MBB)**

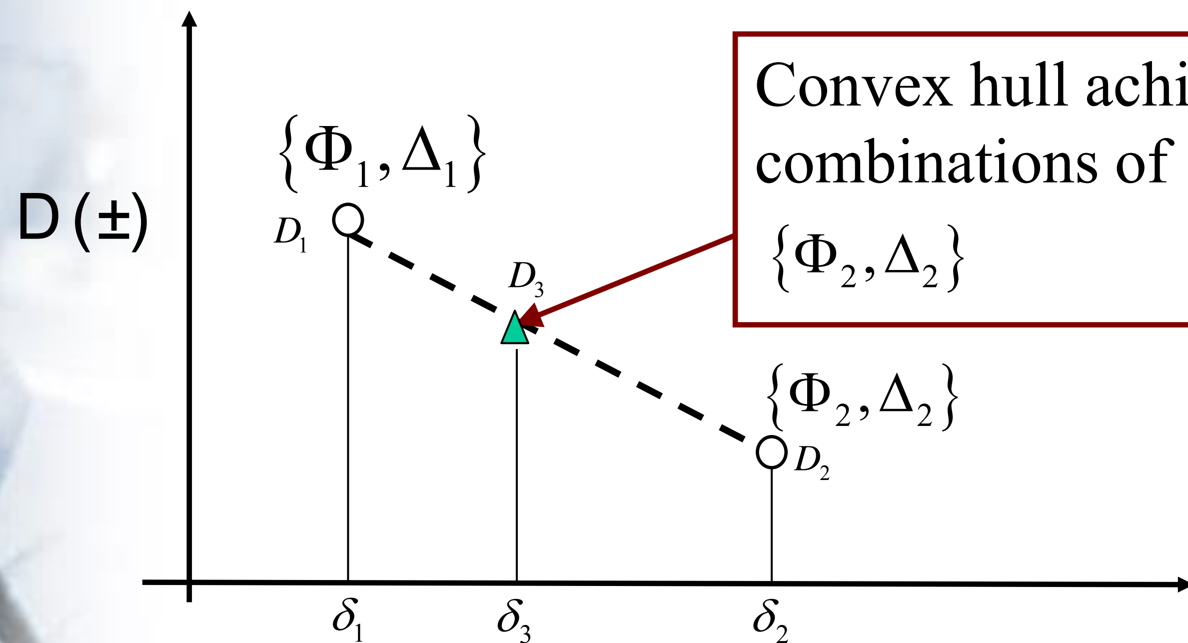
$$p(x) = \prod_{j=0}^{\infty} p(\sigma_j^2) N(0, \sigma_j^2)$$

- bounded by the combination of Gaussian lower bound
- tighter than EBB as $\delta \to 0$

Convexity of D(δ)

Theorem:

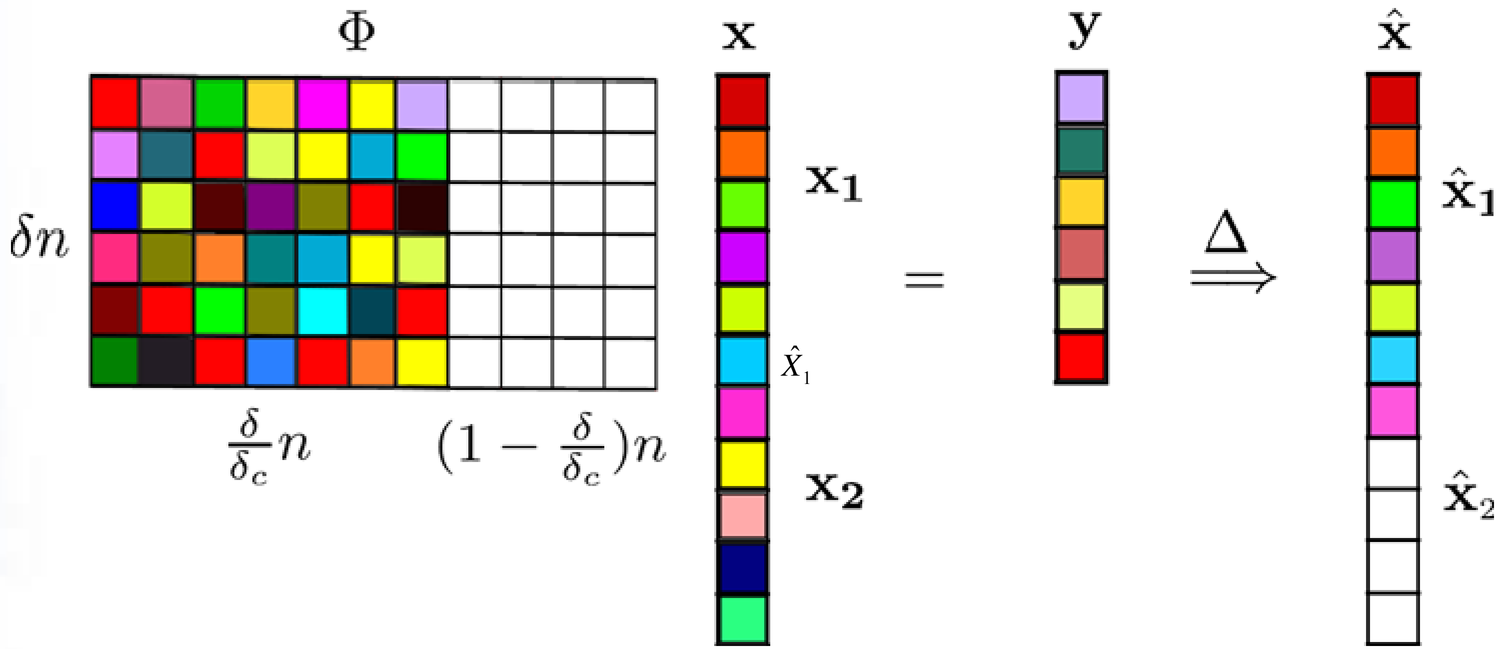
The SD function, D(δ), is convex



Convex hull achievable by combinations of $\{\Phi_1, \Delta_1\}$ and $\{\Phi_2, \Delta_2\}$

$$\delta_3 = \alpha\delta_1 + (1 - \alpha)\delta_2$$

Hybrid Zeroing Matrix

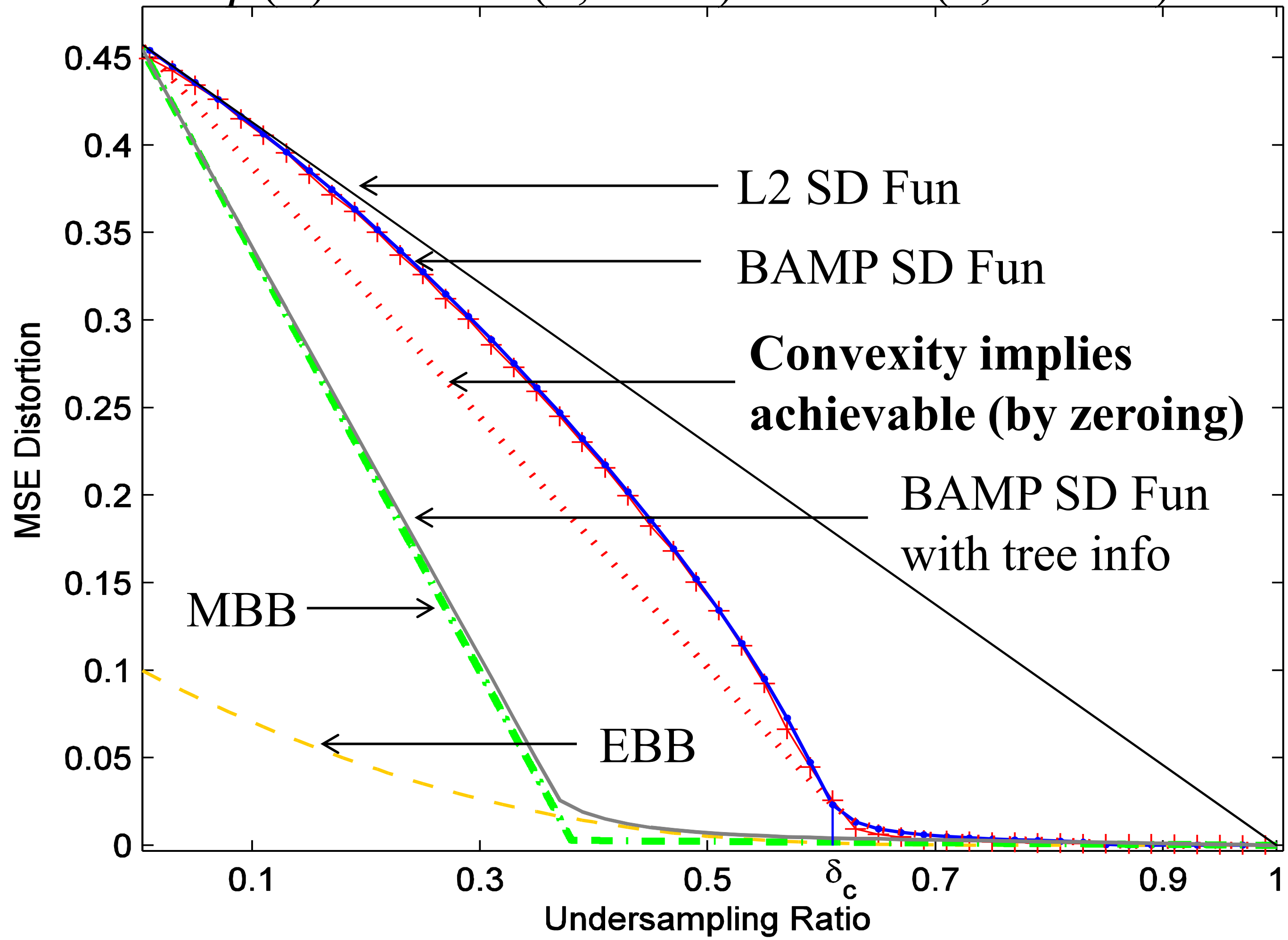


$$\Phi = [\tilde{\Phi}, 0]$$

Setting a portion of the measurement matrix as zero (convex combination of the trivial decoder and BAMP decoder) effectively convexifies the SD function.

SD Function for Two-state GM

$$p(x) = 0.38N(0,1.198) + 0.62N(0,0.0044)$$



Note there is no first order phase transition for the BAMP SD curve thus the magic matrix is not beneficial [Barbier,Krzakala 2011]

Statistical image model

A simple **statistical multi-resolution model** [Mallat 89, Choi & Baraniuk 99] represent image with wavelets:

$$f = \sum_k u_{j_0,k}\phi_{j_0,k} + \sum_{j \geq j_0,k} w_{j,k}\varphi_{j,k}$$

model wavelet coefficients as i.i.d. GM with fixed variance per band

$$w_{j,k} \;:\; \lambda_j N(0,\sigma_{L,j}^2) + \left(1-\lambda_j\right) N(0,\sigma_{S,j}^2)$$

where $\sigma_{L,j}$ and $\sigma_{S,j}$ **decay exponentially** across scale

This model is related to the deterministic Besov signal model.

Bandwise Sampling

We proposed to (randomly) sample each band independently, e.g. [Donoho 2006, Tsaig 2007, Chang et al 2009] - makes analysis tractable.

$$\Phi = \begin{pmatrix} \Phi_0 & & & \\ & \Phi_1 & & \\ & & \ddots & \\ & & & \Phi_L \end{pmatrix}$$

Optimizing Sample Allocation

Need to balance placing a sample in one band over another

$$\min_{m_i} \sum_{i=1}^{L} \sigma_i^2 n_i D_i\left(\frac{m_i}{n_i}\right)$$

$$s.t. \sum_{i=1}^{L} m_i = m \quad \text{and} \quad 0 \le m_i \le n_i \quad i = 1, \ldots, L$$

Bandwise Sampling

Optimizing Sample Allocation

From the Lagrangian formulation, define a **distortion reduction function** for each band:

$$\eta_j(m_i) = \sigma_j^2 n_j \left(D\left((m_i+1)/n_j\right) - D(m_i)/n_j \right)$$

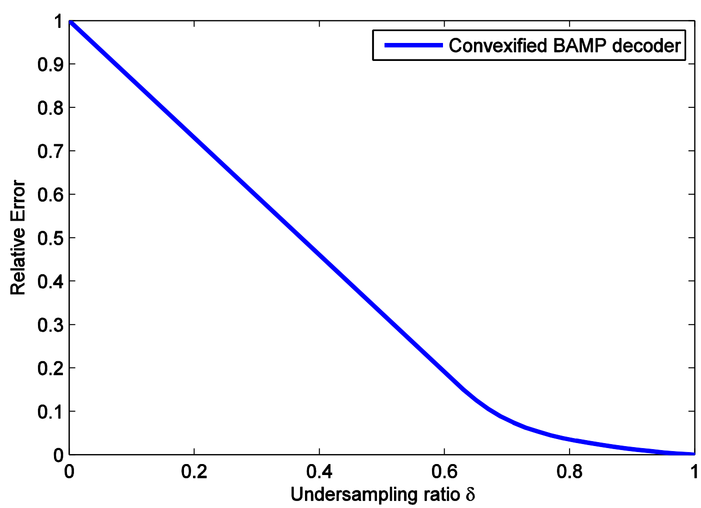
Optimal solution is a consequence of convex SD function and achieved by a greedy sample allocation strategy.

Similar idea to **reverse water filling** in Rate Distortion Theory

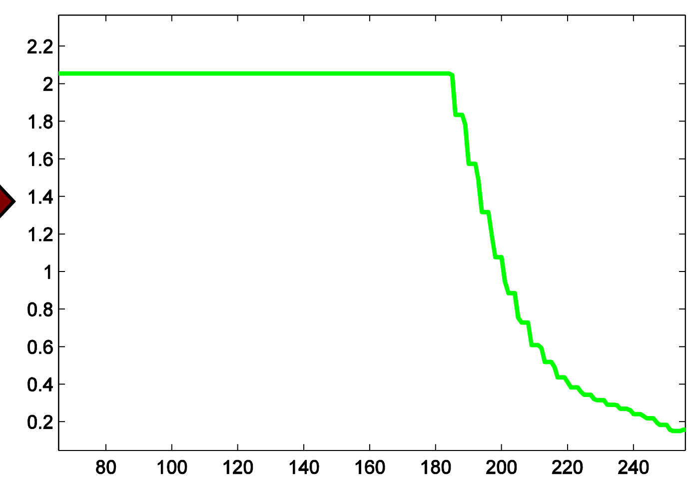
Bandwise Sampling

Convexified BAMP distortion reduction function (band 1 for cameraman image model)

distortion



distortion reduction

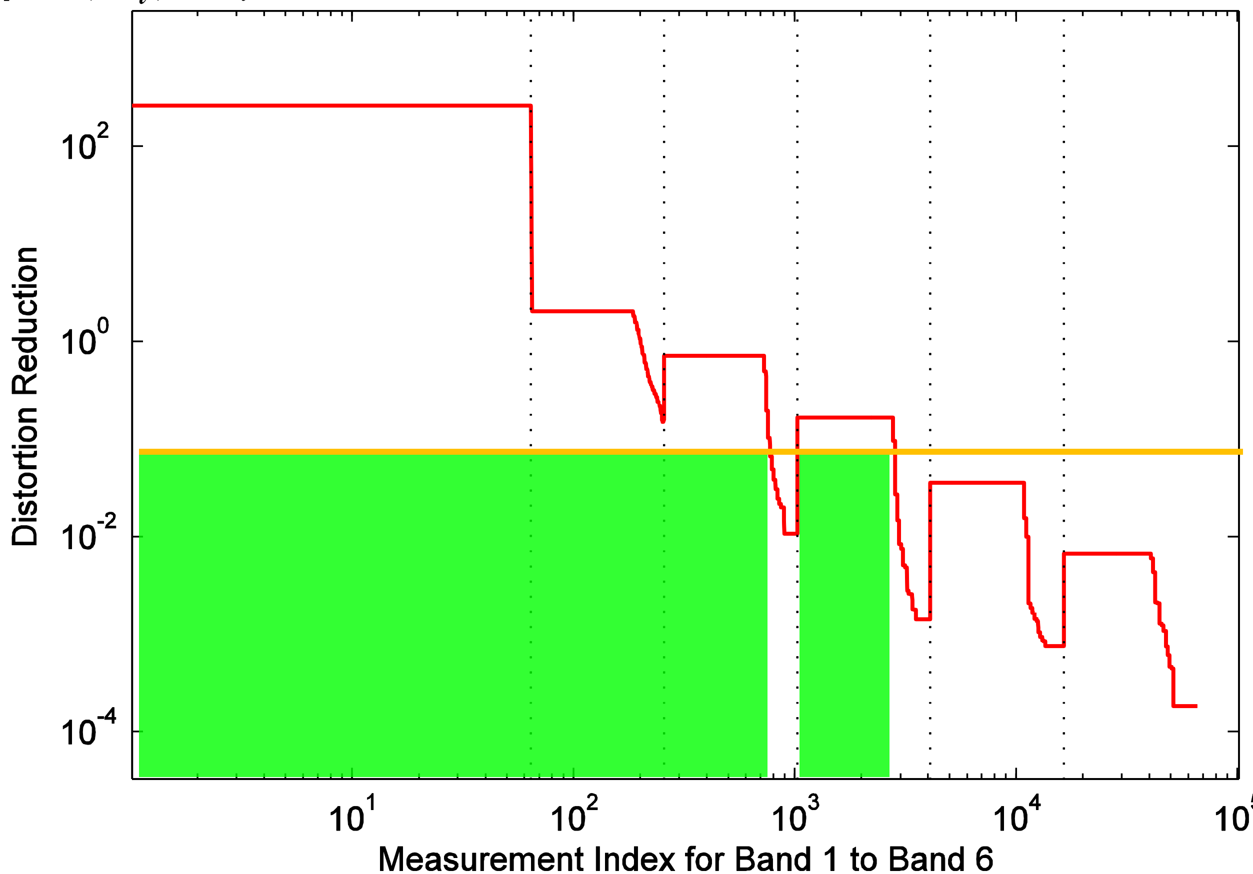


$$\eta_j(m_i) = \sigma_j^2 n_j \left(D\left((m_i+1)/n_j\right) - D(m_i)/n_j \right)$$

Bandwise Sample Allocation

We select a γ and reverse fill samples in each band until $\eta^{(i)}(m_i) \le \gamma$

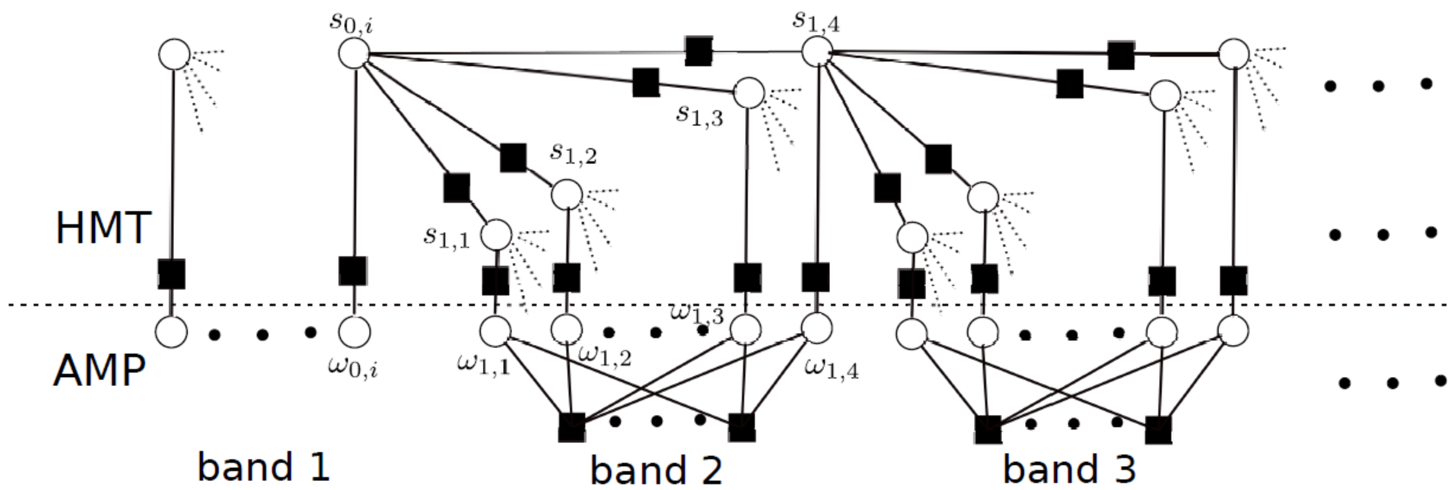
DR fun for cameraman image



γ

The optimization works for any convex SD function, including the oracle function (MBB)

Incorporating Tree Structure



$$\mathrm{P(X|Y)} \propto \sum_S p(S) \prod_{i=1}^{n} p(x_i|s_i) \prod_{j=1}^{m} p(y_i|X)$$

Turbo scheme [Som, Schinter 2012]: calculate marginal probabilities for hidden states $S_{j,k}$ and incorporate into BAMP

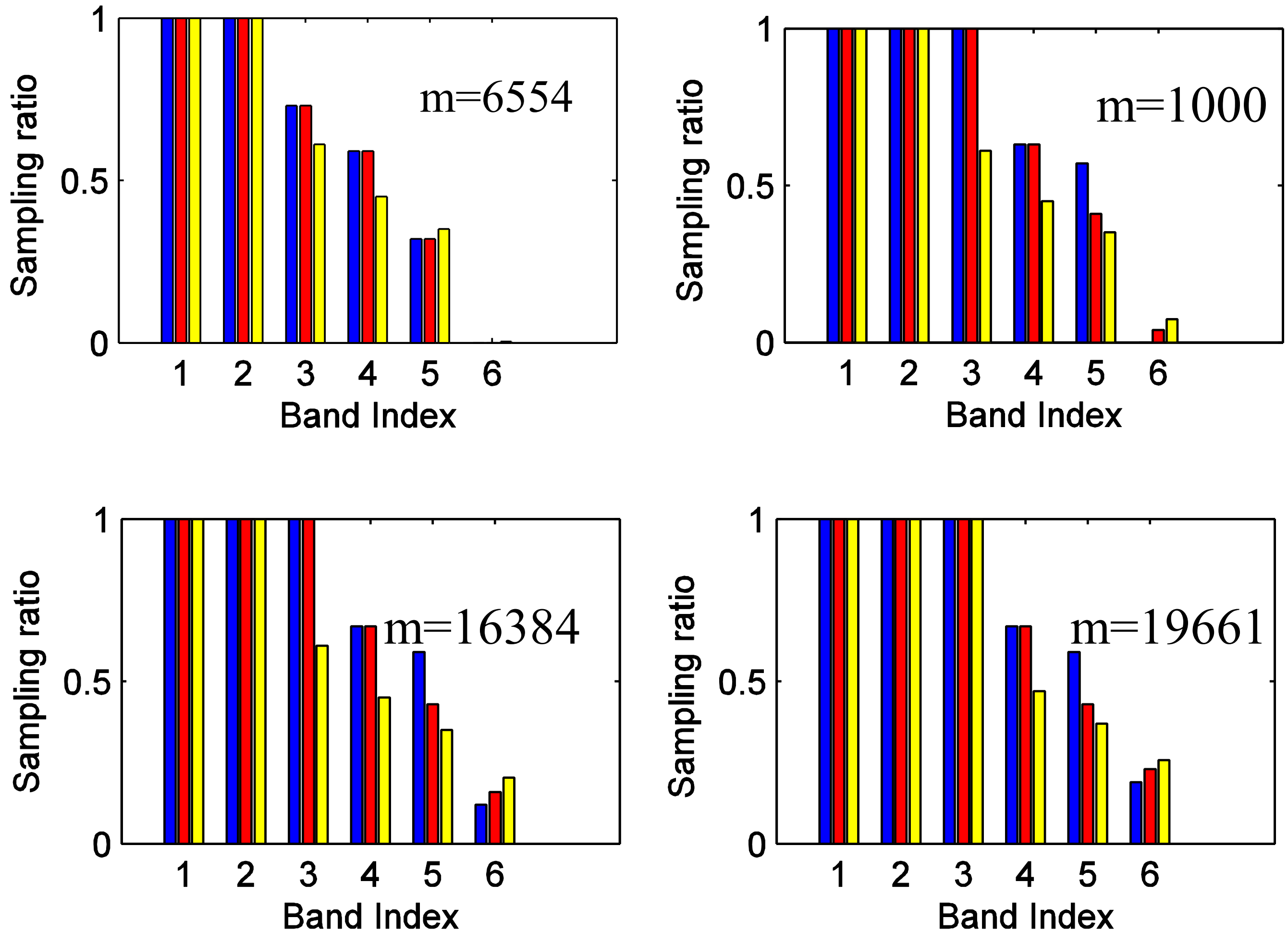
Bandwise CS Sample Allocation

Sample allocation (% of full sampling) per band for δ= 10%, 15.26%, 25% and 30%

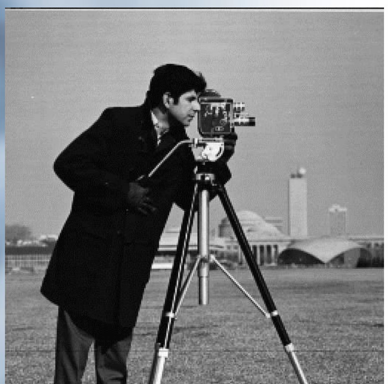
SA for cvx SD fun

Empirical best SA with tree info

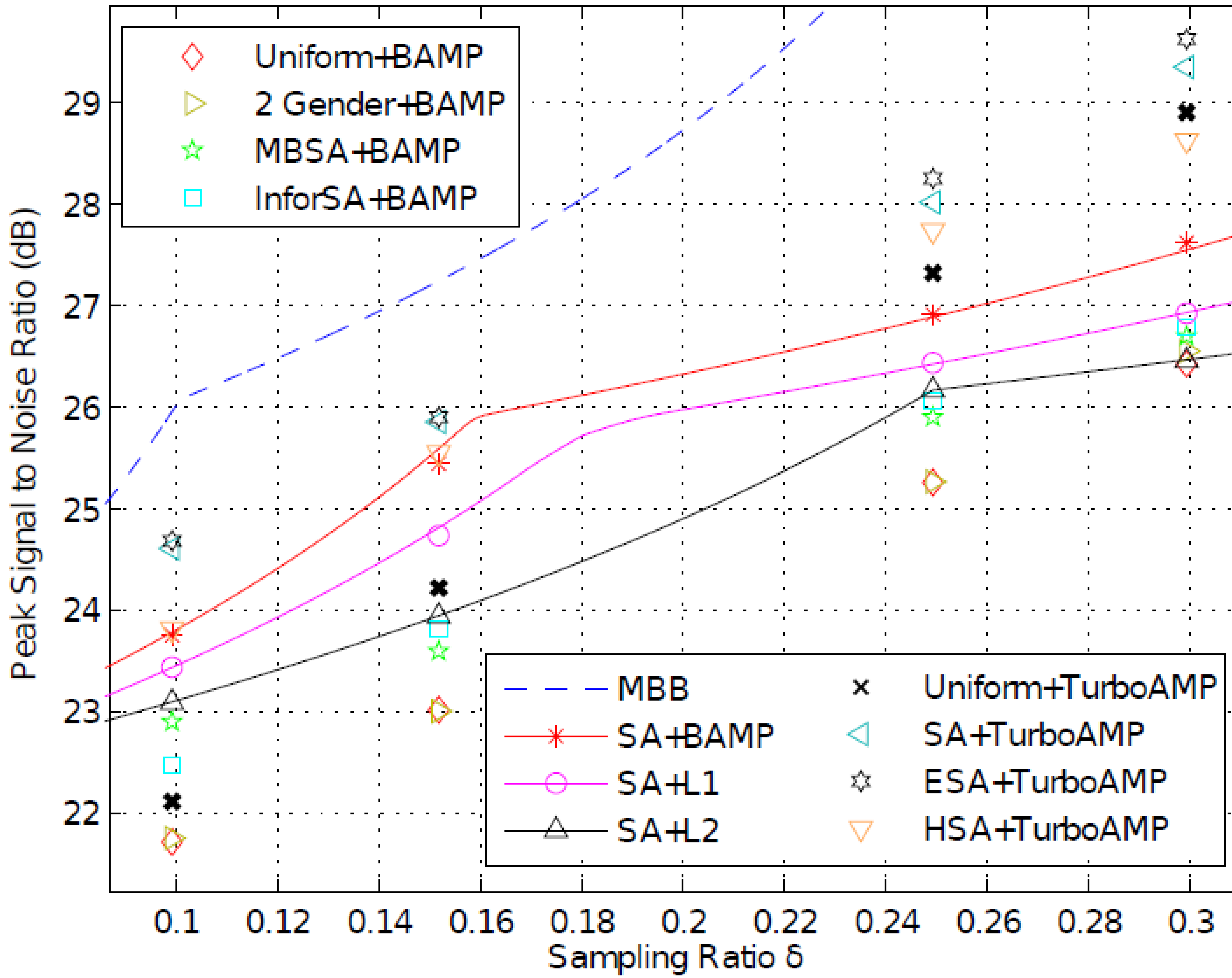
SA for SD fun with true tree info



Bandwise CS Performance



e.g. cameraman



Reconstructed Images

Image reconstructions from 10000 measurements (15%)

General Sample Allocation

Test images from the Berkeley dataset for the GSA profile

Average statistics for db2 wavelet coefficients of 200 images

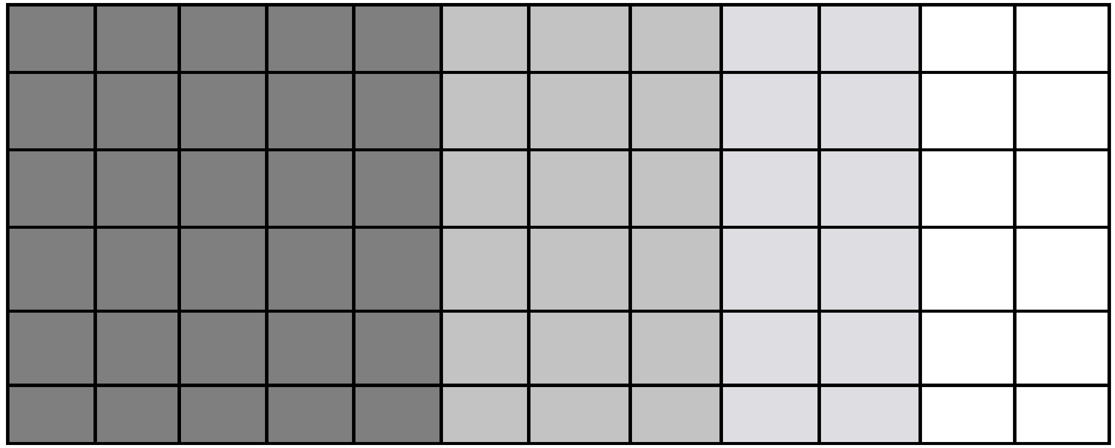
subband	b_1	b_2	b_3	b_4	b_5
λ	0.5108	0.4374	0.4076	0.3616	0.3137
σ_L^2	3.6910	0.7506	0.1595	0.0385	0.0081
σ_S^2	0.4596	0.0490	0.0075	0.0015	0.0003

General Sample Allocation

Reconstruction comparision for sampling ratio 0.2

Image	GSA	InforSA	MBSA	Uniform	2 Gender	SA+TurboAMP
car	**25.56**	24.11	25.29	22.92	22.98	25.92
plane	**28.28**	27.32	28.13	26.19	26.25	28.52
eagle	**28.66**	27.84	28.59	26.31	26.44	28.95
sculpture	**23.81**	22.89	23.54	22.05	22.61	24.58
surfer	**25.37**	24.00	25.13	22.81	22.95	25.65
tourists	**24.15**	22.93	23.75	22.08	22.37	24.53
building	**24.84**	23.59	24.66	22.48	22.55	25.37
castle	**23.65**	22.76	23.41	21.02	21.42	23.96
man	**30.32**	29.33	30.08	28.05	28.49	30.80
fish	**27.26**	27.57	26.76	24.62	24.83	27.76
average	**26.10**	25.23	25.93	23.85	24.09	26.60

Modulated Matrix Structure



The modulated matrix is a product of the homogeneous Gaussian matrix G and the rescaling matrix R

$$R=\begin{pmatrix} \sqrt{J_1}I_N & 0 & \cdots & 0 \\ 0 & \sqrt{J_2}I_{N_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{J_{Lc}}I_{N_{Lc}} \end{pmatrix} \qquad \Phi_M = GR$$

Each block is a Gaussian matrix with zero mean and J_i / N variance

1-D State Evolution Dynamics

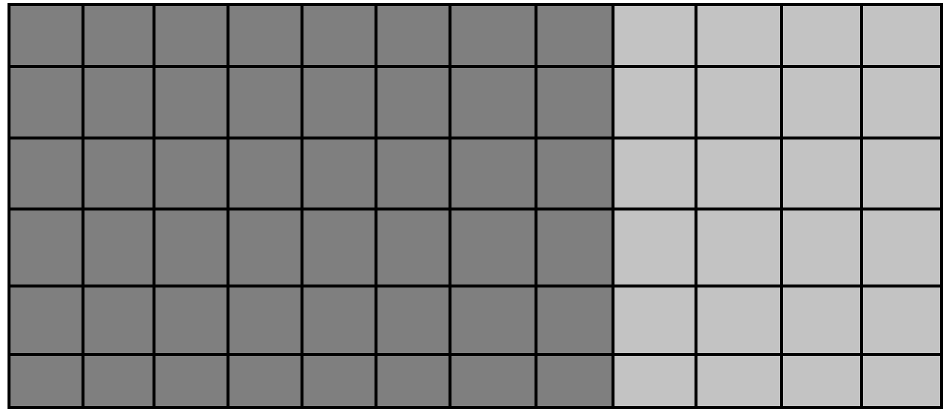
For modulated matrix, a 1-D SE equation is derived to track the performance based on the seeded matrix analysis [Krzakala 13]

$$\hat{\tau}^{(t+1)} = \frac{\sum_k J_k \gamma_k S\left(\hat{\tau}^t / J_k\right)}{\delta}$$

When the SE equation converges, the distortion is predicted as

$$\bar{E} = \frac{1}{L_c} \sum_k S(\frac{\hat{\tau}^*}{J_k})$$

Two Block Matrix



rescaling matrix $\hat{R} = \begin{pmatrix} I_1 & 0 \\ 0 & \sqrt{J_2} I_2 \end{pmatrix}$

1-D SE equation $\hat{\tau}^{(t+1)} = \frac{1}{\alpha} M(\hat{\tau}^t) = \frac{1}{\alpha}\left[\gamma_1 S\left(\hat{\tau}^t\right) + (1-\gamma_1) J_2 S\left(\frac{\hat{\tau}^t}{J_2}\right)\right]$

Distortion equation $\bar{E} = \gamma_1 S\left(\hat{\tau}^*\right) + (1-\gamma_1) S\left(\frac{\hat{\tau}^*}{J_2}\right)$

Zeroing matrix is a special case where $J_2 = 0$

Two Block Matrix vs. Seeded Matrix

The seeded matrix with 4 sub-matrices takes the form

$$\Phi_s = \begin{pmatrix} G_1 & \sqrt{J_2}G_2 \\ \sqrt{J_1}G_3 & G_4 \end{pmatrix}$$

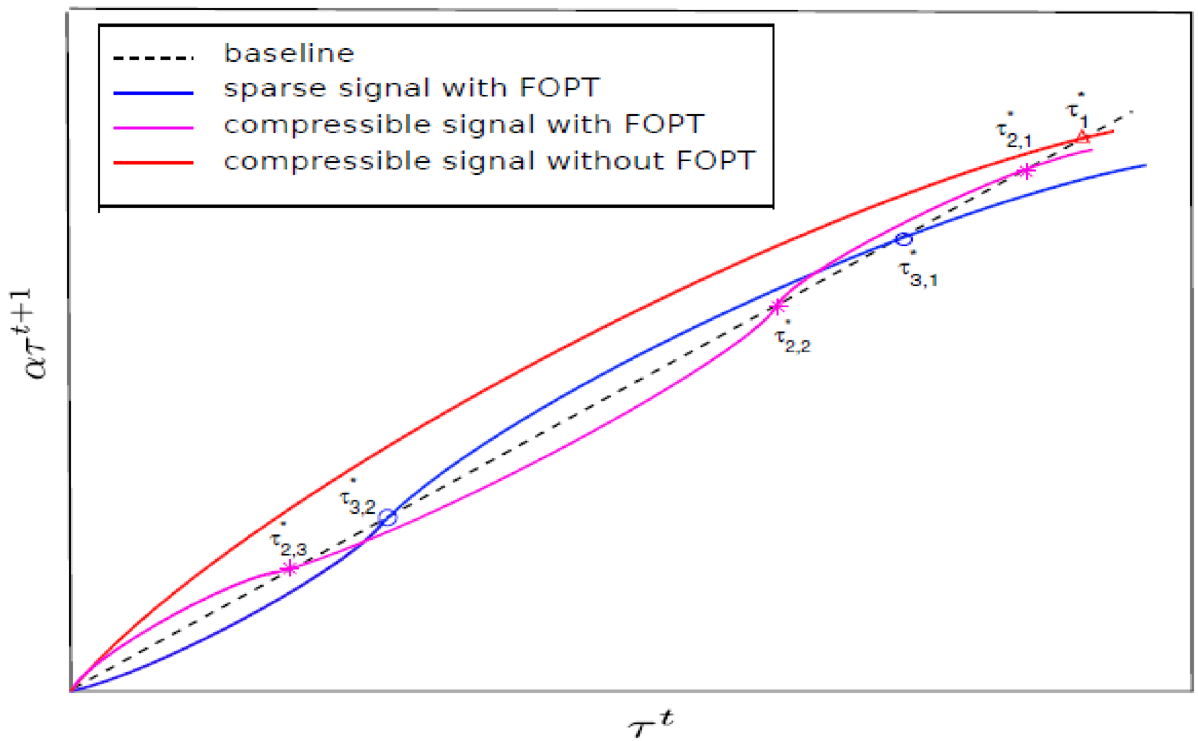
If we set $J_1 = 1/J_2$ the two block matrix is the rescaled seeded matrix

$$\hat{\Phi}_M \overset{J_1=1/J_2}{=} \begin{pmatrix} G_1 & \sqrt{J_2}G_2 \\ G_3 & \sqrt{J_2}G_4 \end{pmatrix} = \begin{pmatrix} I_1 & 0 \\ 0 & \sqrt{J_2}I_4 \end{pmatrix} \Phi_s$$

The two block matrix has a relatively simple 1-D SE dynamics, which makes the analytical optimization possible.

First Order Phase Transition (FOPT)

A discontinuous drop of the MSE at a particular δ in the SD context



Necessary and sufficient condition for signals without FOPT:

for any $\tau^* > 0$

$$\frac{f(\tau^*)}{\tau^*} < \eta(\tau^*)$$

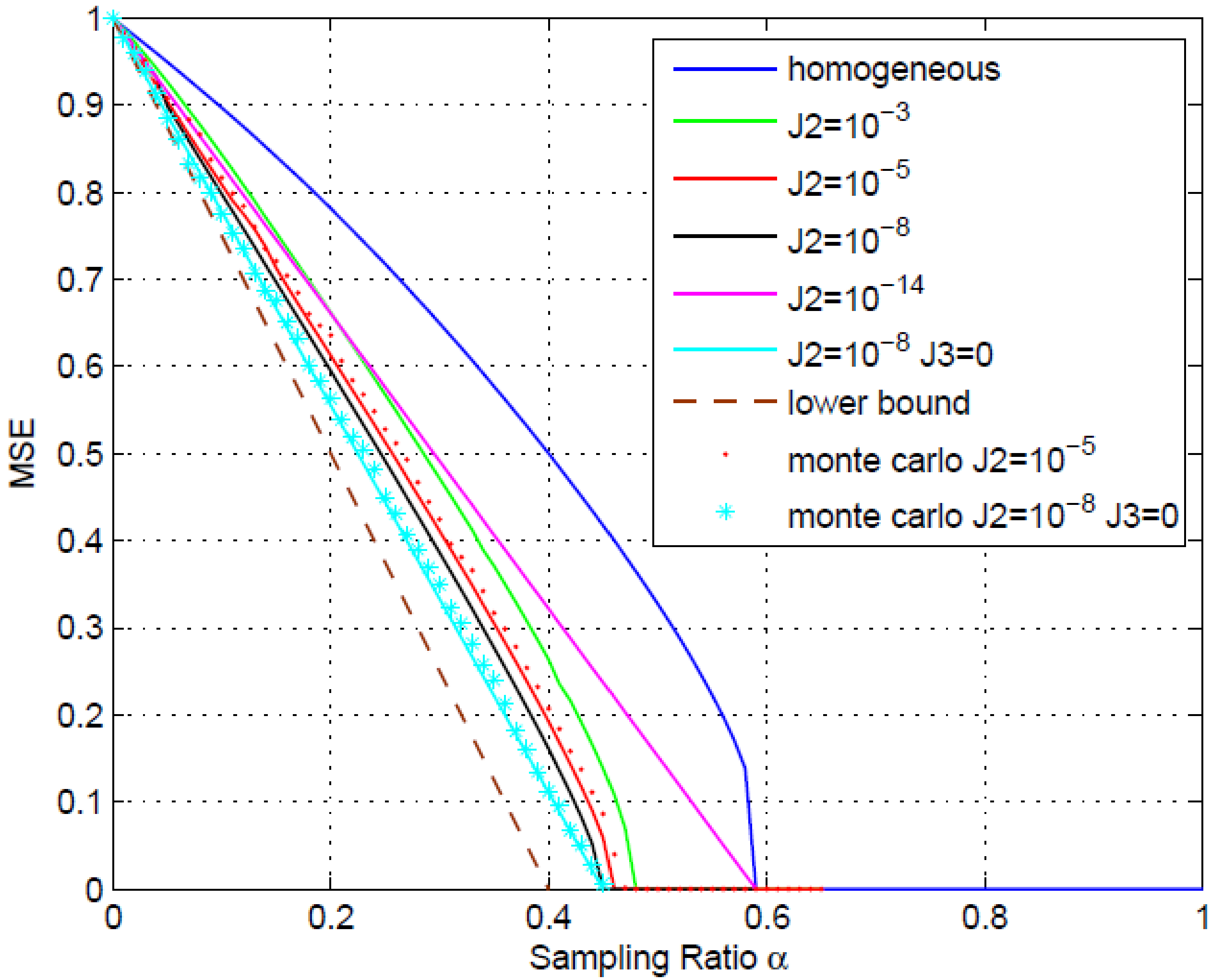
Where $\eta(\tau) \triangleq \dfrac{df(\tau)}{d\tau}$ and SE equation takes the form $\delta\tau^{(t+1)} = f(\tau^t)$

Two Block Matrix Effect on FOPT

- For signals with FOPT, the spurious fixed points of the SE equation will be removed so that perfect reconstruction is achievable.
- For signals without FOPT, the dynamics of the two block matrix keep this property

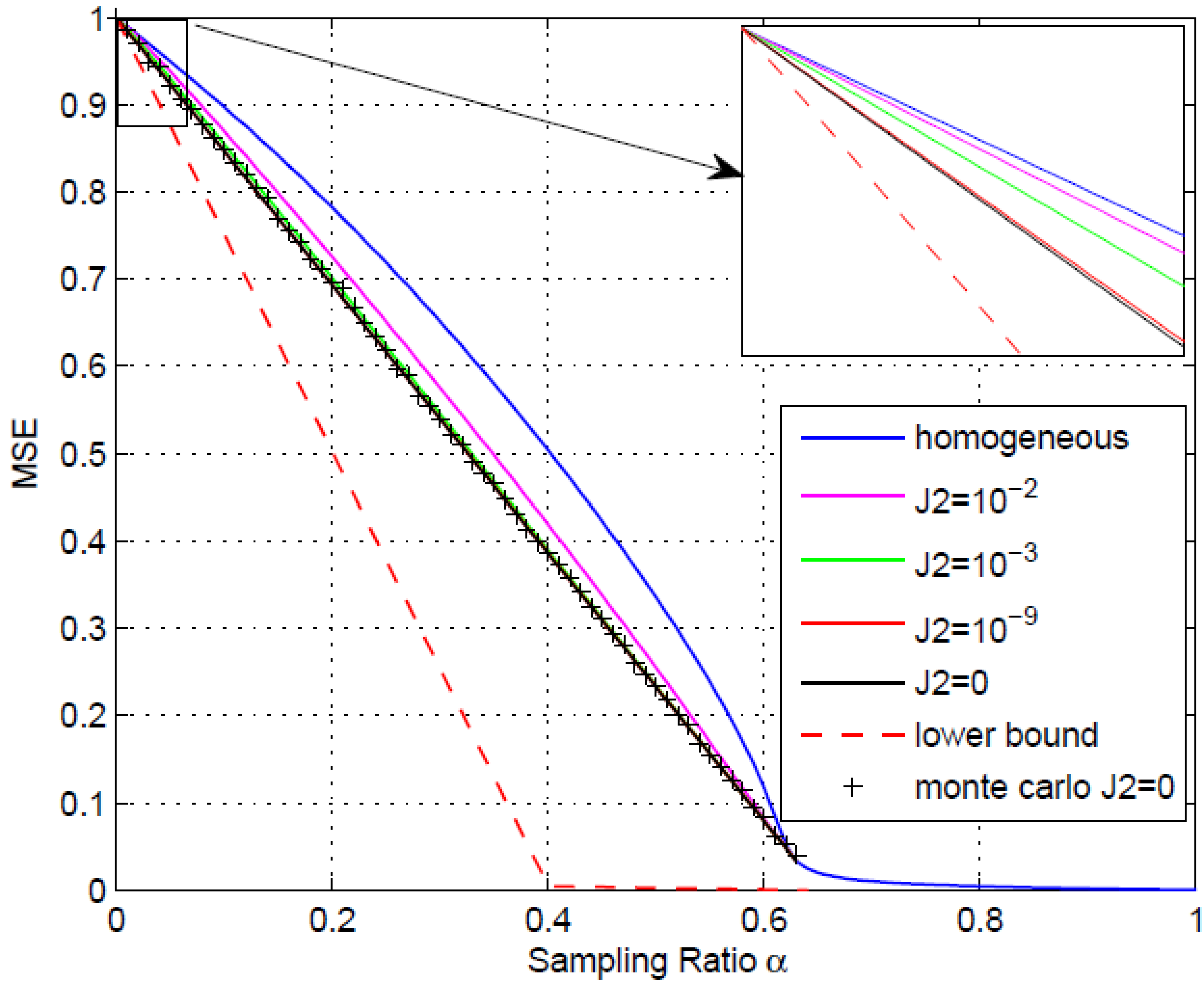
Theorem:
If the SE equation for signals with the homogeneous Gaussian matrix $S(\tau)$ satisfies the no FOPT condition, then the SE equation for using the two block matrix $M(\tau)$ also satisfies the no FOPT condition.

Two Block Matrix for Sparse Signal



The perfect reconstruction ratio is moved from 0.59 to 0.45 by the two block matrix with $J_2 = 10^{-8}$ and can be further convexified by a three block matrix.

Two Block Matrix for Compressible Signal



Empirically we observed that zeroing matrix is optimal for compressible signals without FOPT

Conclusion

- We have introduced a SD framework to characterize a signal's "compressibility" in a stochastic setting.
- We used SD functions to derive a natural discretization for CS imaging & it gives accurate estimation of CS performance
- Modulated matrix is introduced as an extension of the seeded matrix with a simple 1-D SE dynamics
- First order phase transition is analyzed from the SE function perspective and necessary and sufficient condition for signals without FOPT is provided
- Two-block matrix is studied as a special case.

Thank You