A Sample Distortion Analysis for Compressed Imaging

Chunli Guo, Mike E. Davies

Institute for Digital Communications
University of Edinburgh, UK

Talk Outline

- Introduction
  - Compressed Sensing: from Sparse to Compressible, from Deterministic to Stochastic
- Sample Distortion (SD) framework
  - definition, examples, lower bounds and convexity
- Multi-resolution CS
  - Wavelet Statistical Image Model
  - SD and Optimal Bandwise Sampling
  - Oracle Bounds
  - Sample Allocation with tree structure
- Natural Image Examples
- Conclusions

Sparsity and Compressed Sensing

Compressed sensing

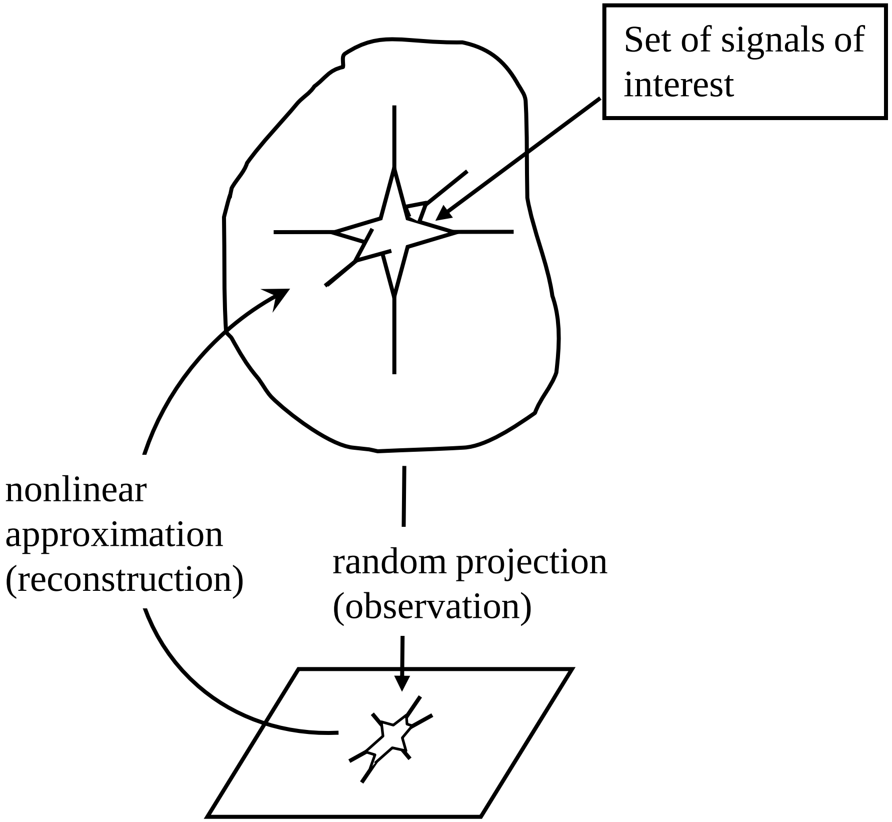
Compressed Sensing assumes a sparse/compressible set of signals

Uses random projections for observation matrices

Signal reconstruction by a nonlinear mapping.

Compressed sensing provides practical algorithms with guaranteed performance e.g. L1 min., OMP, CoSaMP, IHT.

Closely linked with theory of n-widths



The L1 solution guarantee

A (the?) popular Compressed Sensing solution is:

Basis Pursuit: $\hat{x}_1 = \underset{\{\Phi x = y\}}{\operatorname{argmin}} \|x\|_1$

CS theory asserts that when Φ has an appropriate RIP on k-sparse vectors, x_1, x_2.

$$(1-\delta)\|x_1 - x_2\| \leq \|\Phi(x_1 - x_2)\| \leq (1+\delta)\|x_1 - x_2\|$$

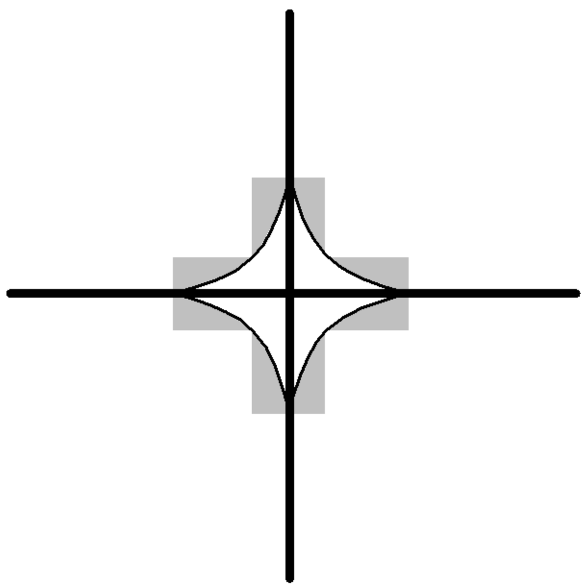
then we are guaranteed the following "instance optimality“

$$\|\hat{x}_1 - x_0\|_1 \leq C_k(\Phi) \cdot \sigma_k(x_0)_1$$

[Candes 2008, Cohen, Dahmen & DeVore 2009]

Compressible vectors (deterministic)

If x lives in either an l^p or weak wl^p ball:



$$\|x\|_{l_p} = \left(\sum_n |x|^p\right)^{1/p} \leq \|x\|_{wl_p}$$

$$\|x\|_{wl_p} := \sup_n \left\{|x_n^*| \cdot n^{1/p}\right\} \leq R$$

Instance optimality implies small approximation error:

$$\sigma_k(x)_q \leq R\left(\frac{p}{q-p}\right)^{\frac{1}{q}} k^{-\left(\frac{1}{p}-\frac{1}{q}\right)}$$

Compressible Distributions

Consider a stochastic (Bayesian) setting...
express signal as a draw from a probabilistic model:

$\longrightarrow$ Draw N samples i.i.d. from a distribution

$$p_X(x) \propto \prod_{i=1}^{N} p(x_i)$$

Question:

When does $p_X(x)$ define an approximately lower dimensional (i.e. compressible) signal model?

$\Rightarrow$ notion of compressible distributions

A Sample-Distortion framework for CS

Sample Distortion Framework

[Guo & D. 2011/2012]

What is the best we can do?

any recovery algorithm; *any* measurement matrix; *any* dimension, i.e. a sampling equivalent to Rate Distortion Theory.

Define the l_2 *Sample Distortion (SD) function* as:

$$D(\delta) := \inf_{\mathrm{n}} \inf_{\Phi} \inf_{\Delta} \frac{1}{n} \mathbb{E} \| \boldsymbol{X} - \Delta(\Phi \boldsymbol{X}) \|_2^2$$

where $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^T$ is the i.i.d. source and we define:

- sampling ratio: $\delta := m/n, m < n$
- linear measurement encoder: $\Phi \in \mathbb{R}^{m \times n}$
- nonlinear decoder: $\Delta(\Phi \boldsymbol{X})$

Sample Distortion Framework

Specific SD functions

- L_2 decoder

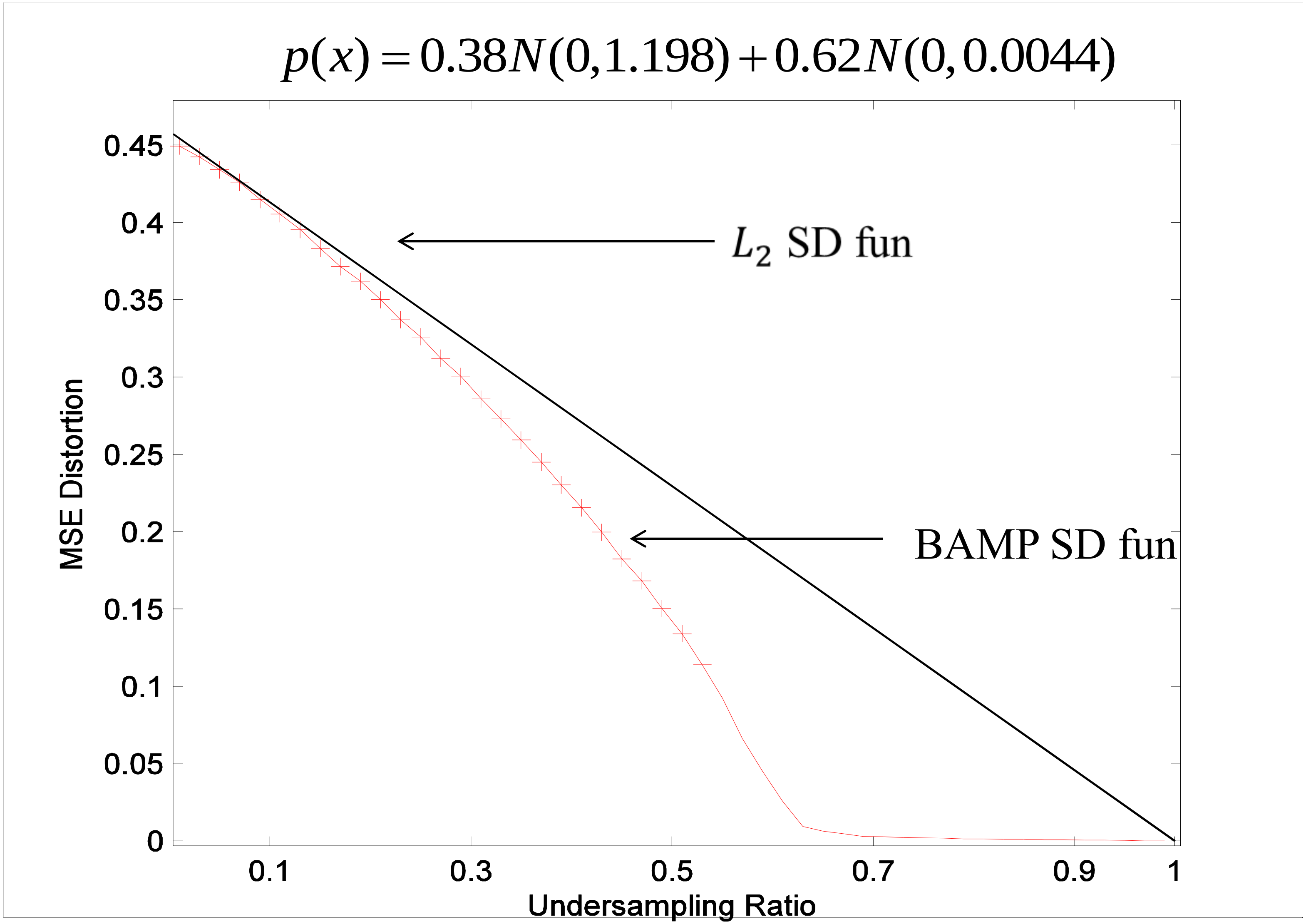
$$D(\delta) = 1 - \delta$$

- MMSE AMP with iid Gaussian encoder state evolution equations predict SD fun. a fixed point of:

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

where $F(\cdot;\tau)$ is MMSE scalar shrinkage function and
$Z \sim \mathcal{N}(0,1)$
Replica Method valid $\Longrightarrow$ MMSE AMP is Bayes optimal

SD Functions for 2-state GSM



SD Lower Bounds

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

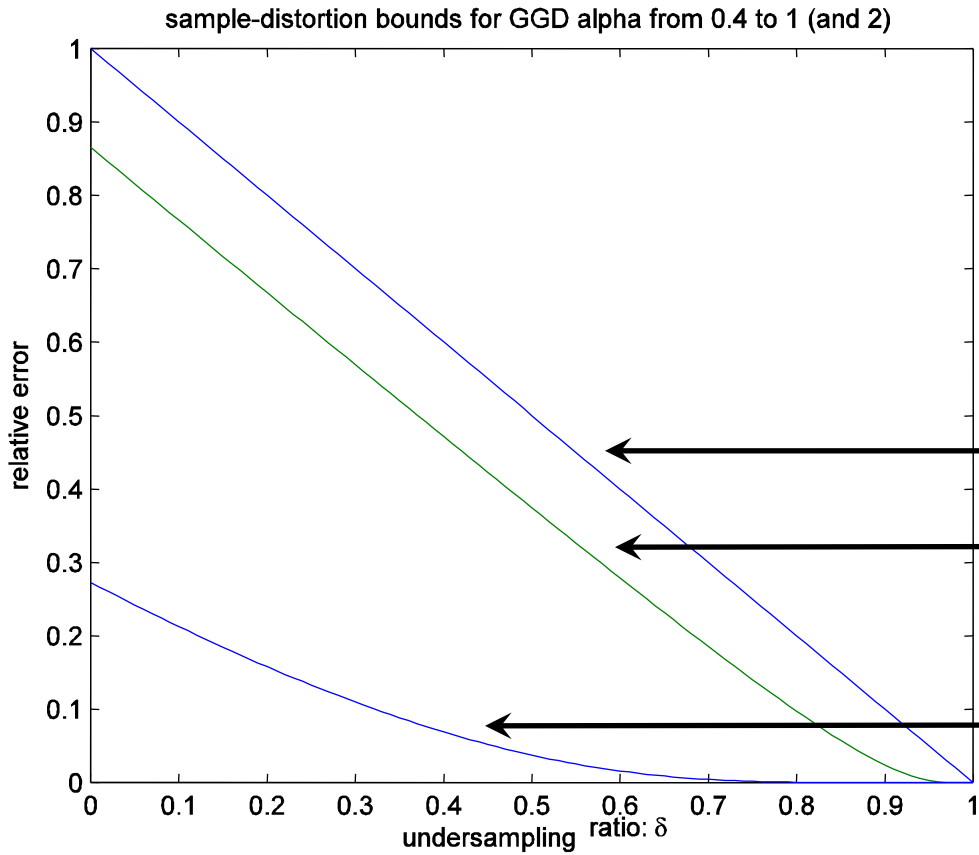
$$D_{EEB}(\delta) \geq (1-\delta)2^{2(h(x_i)-h_g)/(1-\delta)}$$

where

$h(x_i)$ - entropy of $p(x_i)$ and

h_g - entropy of Gaussian

Example: Generalized Gaussian



$$p(x) = \frac{\alpha}{2\sqrt{\beta\sigma}\,\Gamma(\frac{1}{\alpha})} \exp\left(- \left| \frac{x}{\sqrt{\beta\sigma}} \right|^{\alpha} \right)$$

where $\beta = \Gamma(1/\alpha)/\Gamma(3/\alpha)$

Wavelet coefficients of natural images are often modelled as GGD with $\alpha \approx 0.4$-1.0

SD Lower Bounds

Model Based Bound (MBB)

for distributions with a finite/infinite Gaussian scale mixture form

$$p(x) = \int_0^\infty \mathcal{N}(x{:}\, 0, \tau)\, p(\tau) d\tau$$

we have the following `oracle' based bound

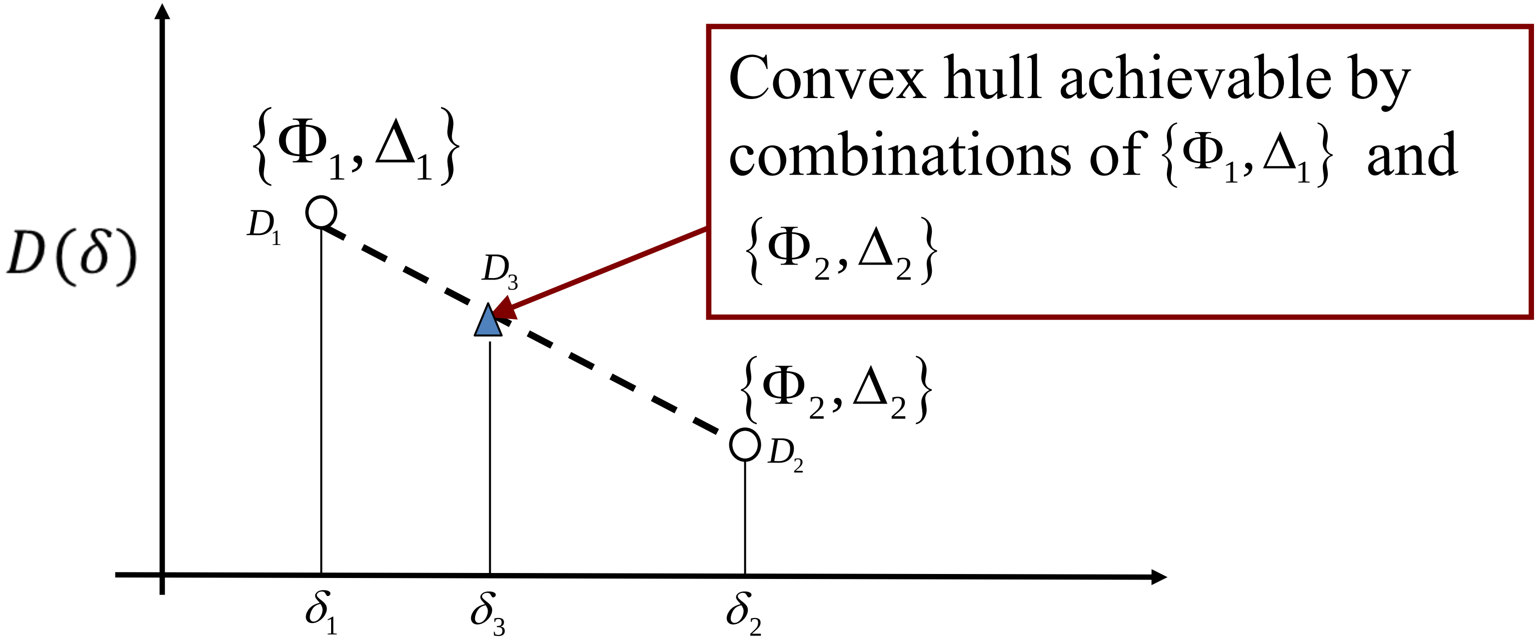
$$D_{MBB}(\delta) = \int_0^c \tau p(\tau) d\tau$$

where $\delta = \int_c^\infty p(\tau) d\tau$

Convexity of D(δ)

Theorem:

The SD function, D(δ), is convex



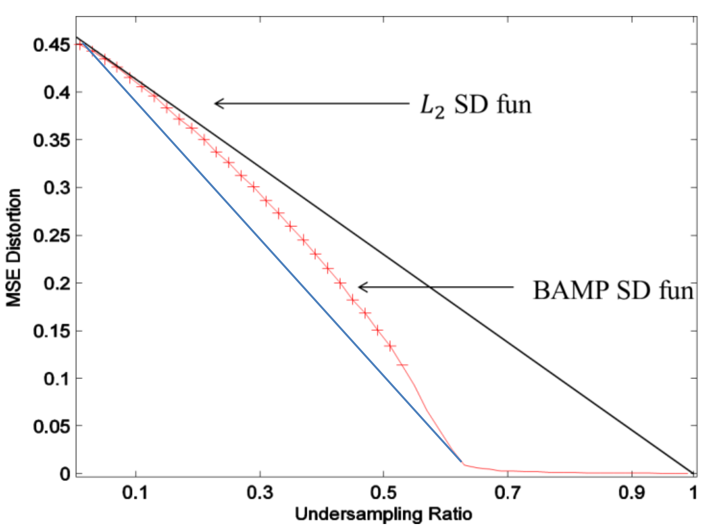
$$\delta_3 = \alpha\delta_1 + (1-\alpha)\delta_2$$

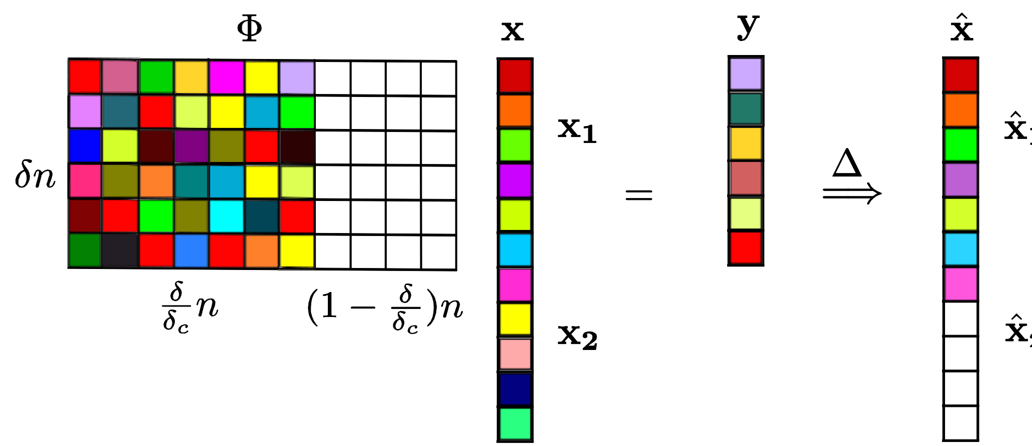
Gaussian encoders are not optimal!

Folk theorem - Gaussian encoders are optimal.

False!

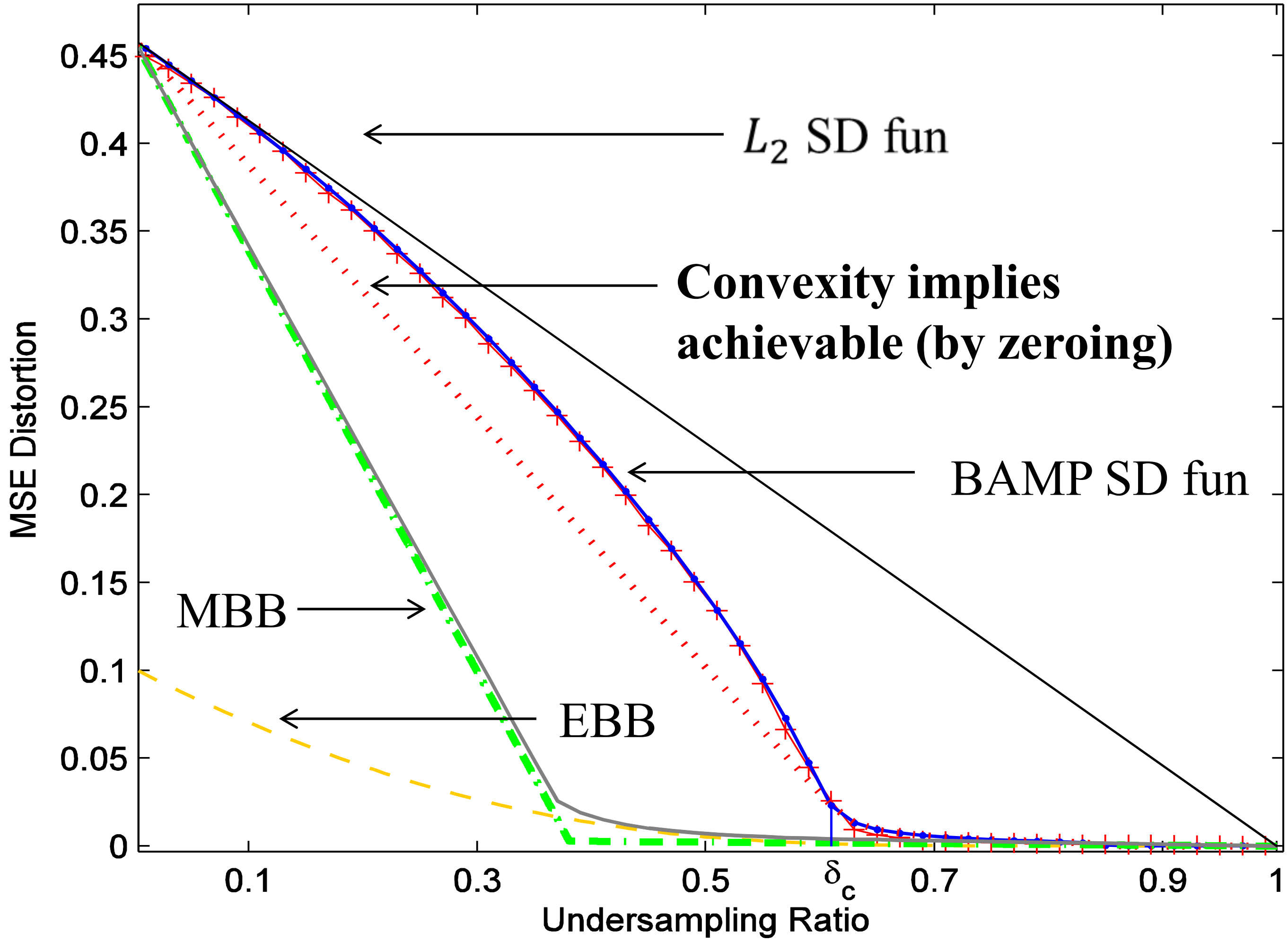
If Gaussian-specific SD function is not convex we can do better





SD Functions for 2-state GSM

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



Multi-resolution Compressive Imaging

Statistical image model

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

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

with $w_{j,k}$ drawn from i.i.d. GMM with fixed variance per band

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

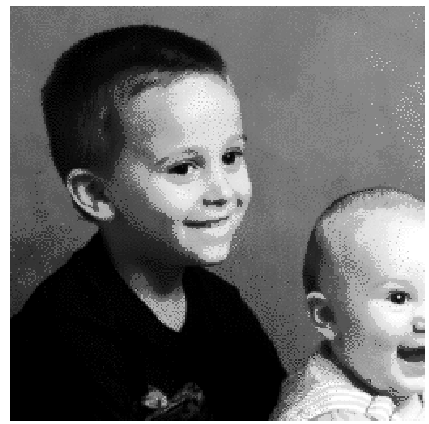
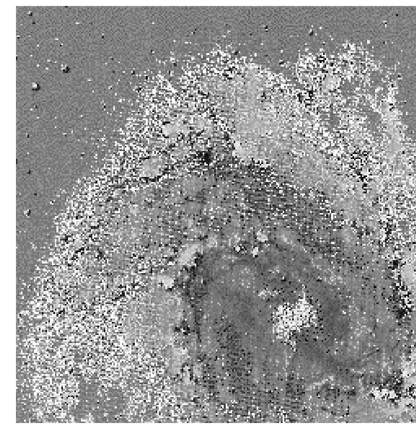
or GGD

$$p\left(w_{j,k}\right) = \frac{\alpha}{2\sqrt{\beta\sigma_j}\Gamma(\frac{1}{\alpha})} \exp\left(-\left|\frac{w_{j,k}}{\sqrt{\beta\sigma_j}}\right|^{\alpha}\right)$$

Where variances **decay exponentially** across scale

Test Images



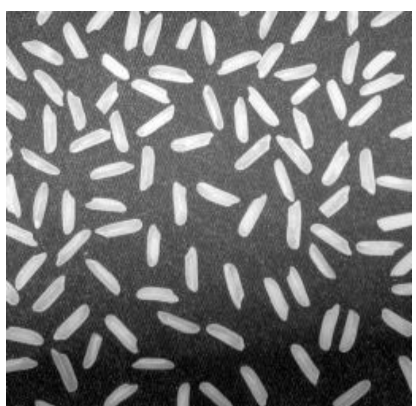


Image model example

TABLE I
STATISTICS FOR DB2 WAVELET COEFFICIENTS OF CAMERAMAN

subband		b_0	b_1	b_2	b_3	b_4	b_5
GGD	α	2	0.7	0.4	0.3	0.3	0.4
	σ^2	261.4383	2.0822	0.4559	0.0902	0.0167	0.0033
GMD	λ	1	0.4155	0.5309	0.4842	0.3664	0.2792
	σ_L^2	261.4383	4.4215	0.8542	0.1856	0.0453	0.0115
	σ_S^2		0.3331	0.0038	0.0004	0.0002	0.0001

cameraman

Image model example

cameraman

GGD representation with Haar wavelets...

Estimated shape parameters for each level

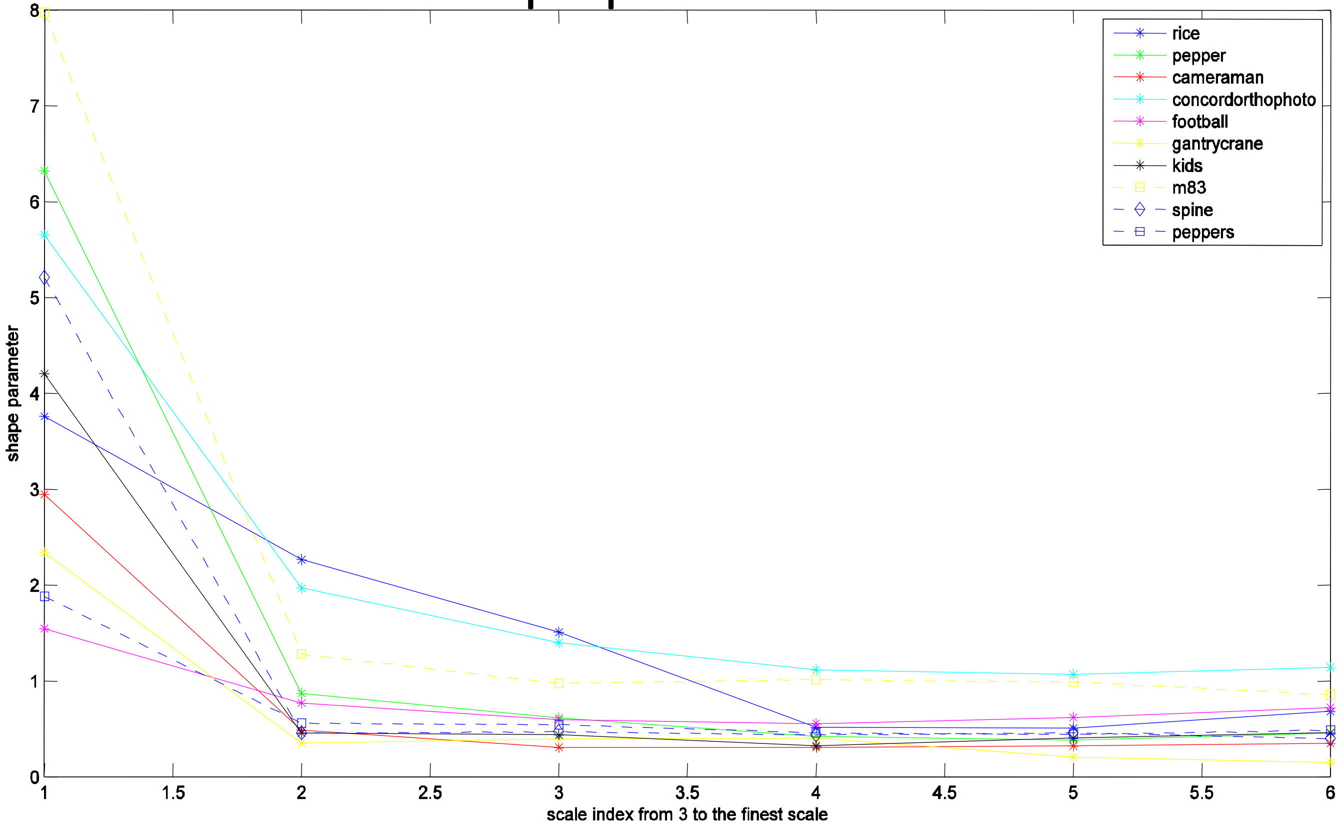
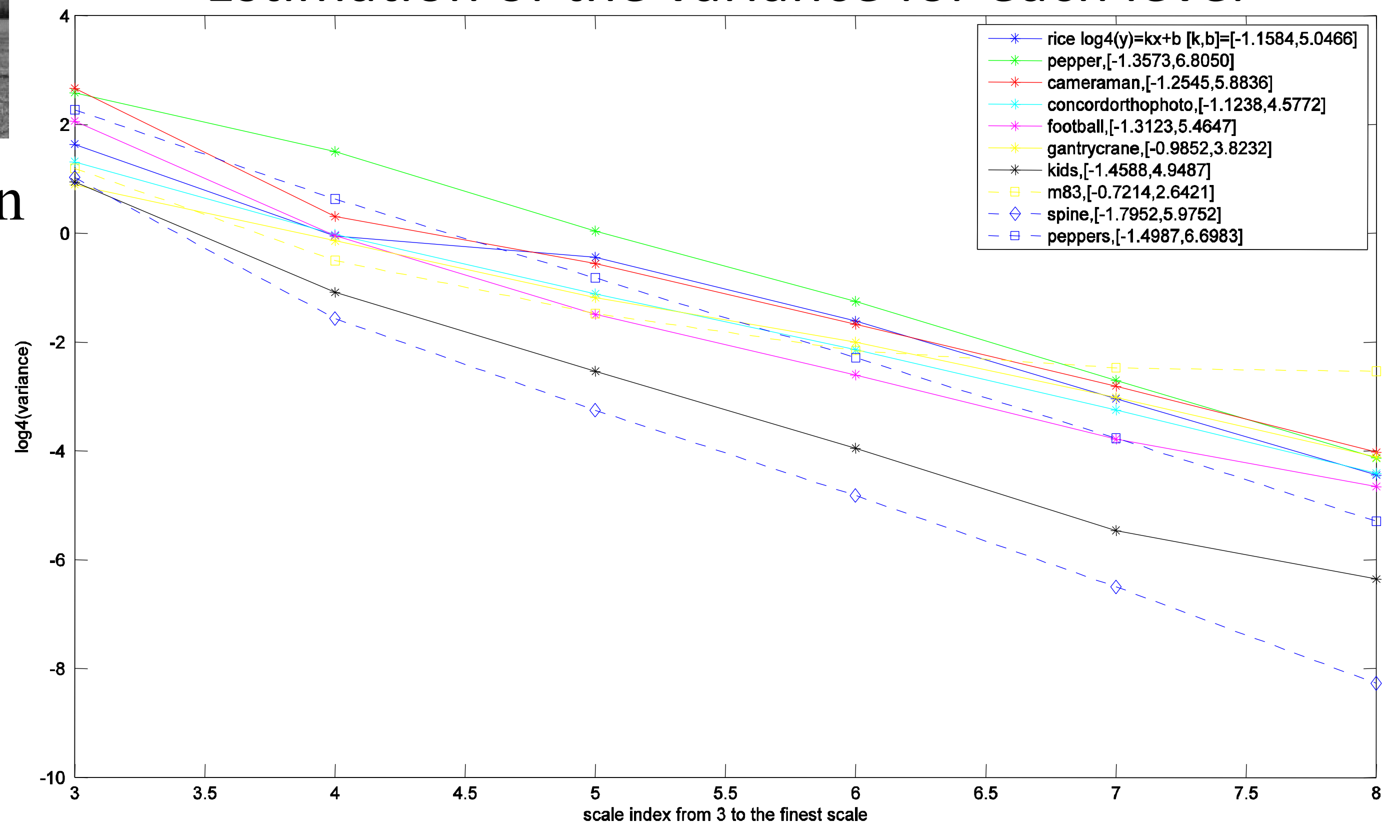


Image model example

e.g. cameraman

GGD wavelet representation

Estimation of the variance for each level



Bandwise Compressive Imaging

Bandwise sampling

Plan: (randomly) sample each wavelet band separately .

$$\begin{bmatrix} y_0 \\ \vdots \\ y_k \\ \vdots \end{bmatrix} = \begin{pmatrix} \Phi_0 & & & \\ & \ddots & & \\ & & \Phi_k & \\ & & & \ddots \end{pmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_k \\ \vdots \end{bmatrix}$$

Why?

- Many researchers have proposed bandwise sampling schemes, e.g. [Donoho 2006, Tsaig 2007, Chang et al 2009]
- Makes analysis tractable (consider problem of bandwise sample allocation)
- Linked to the near-optimal sampling for n-widths of function spaces [Kashin, Maiorov].

Optimal Bandwise sample allocation

Need to balance placing a sample in one band over another.

- Can be formulated as parallel CS problem
- Convex SD function reverse water filling solution similar to Rate Distortion theory.

Define a distortion reduction function for each band:

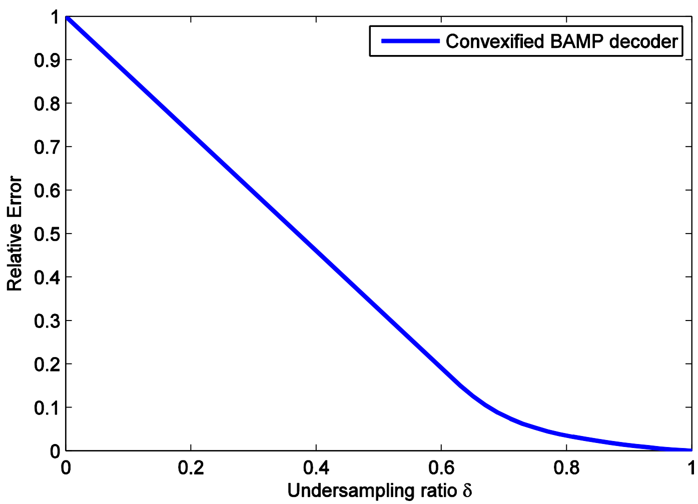
$$\eta^{(i)}(m_i) := \sigma_i^2 n_i\Big(D\big((m_i+1)/n_i\big) - D(m_i/n_i)\Big)$$

Optimal solution when $0 \leq \eta^{(i)}(m_i) \leq \lambda$ for all i and some λ.

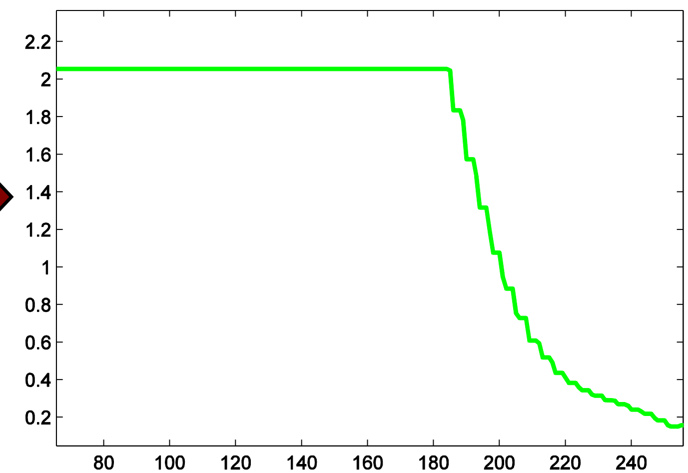
Bandwise Sampling

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

distortion



distortion reduction

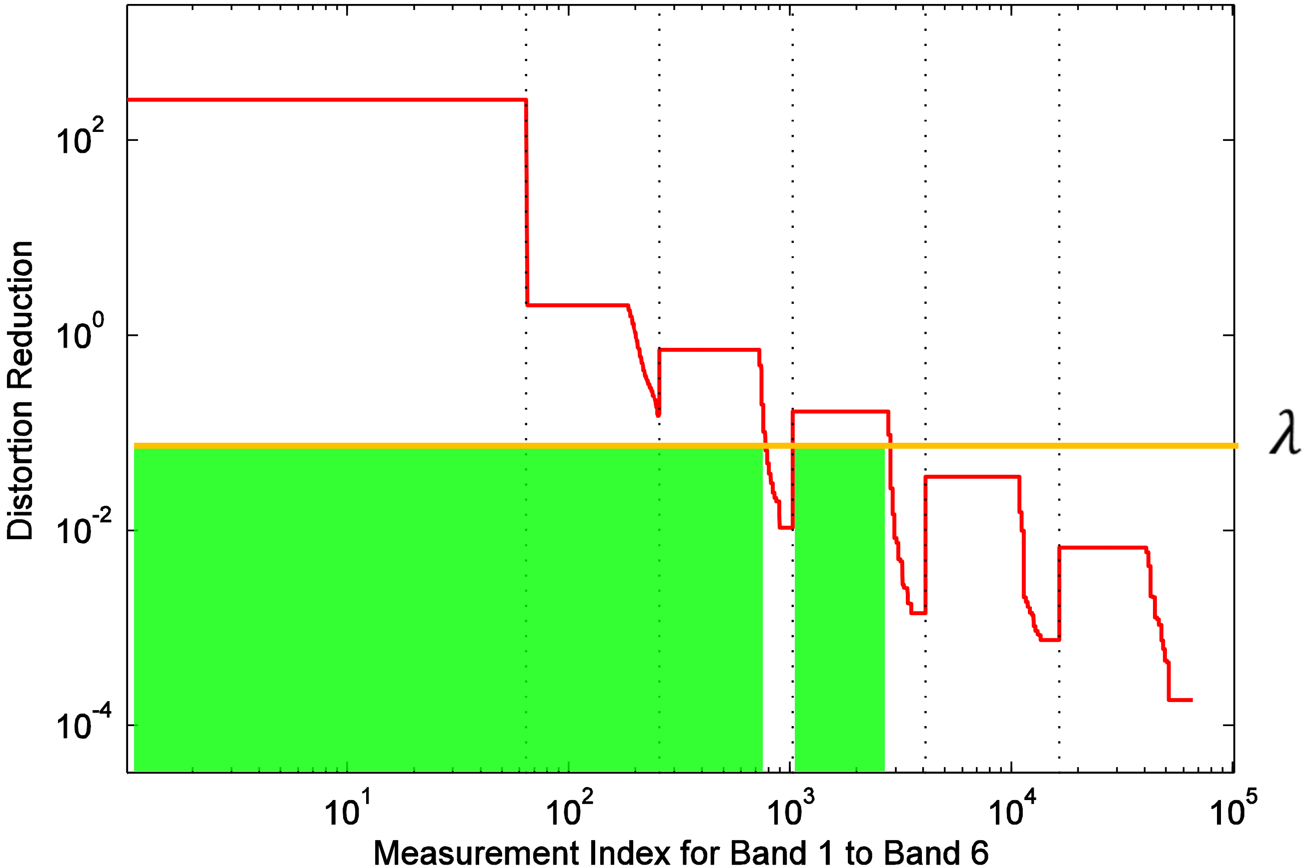


$$\eta_j(m_i) = \sigma_j^2 n_j\big(D(m_i/n_j) - D((m_i+1)/n_j)\big)$$

Bandwise Sample Allocation

We select a λ and reverse fill samples in each band until $\eta^{(i)}(m_i) \leq \lambda$

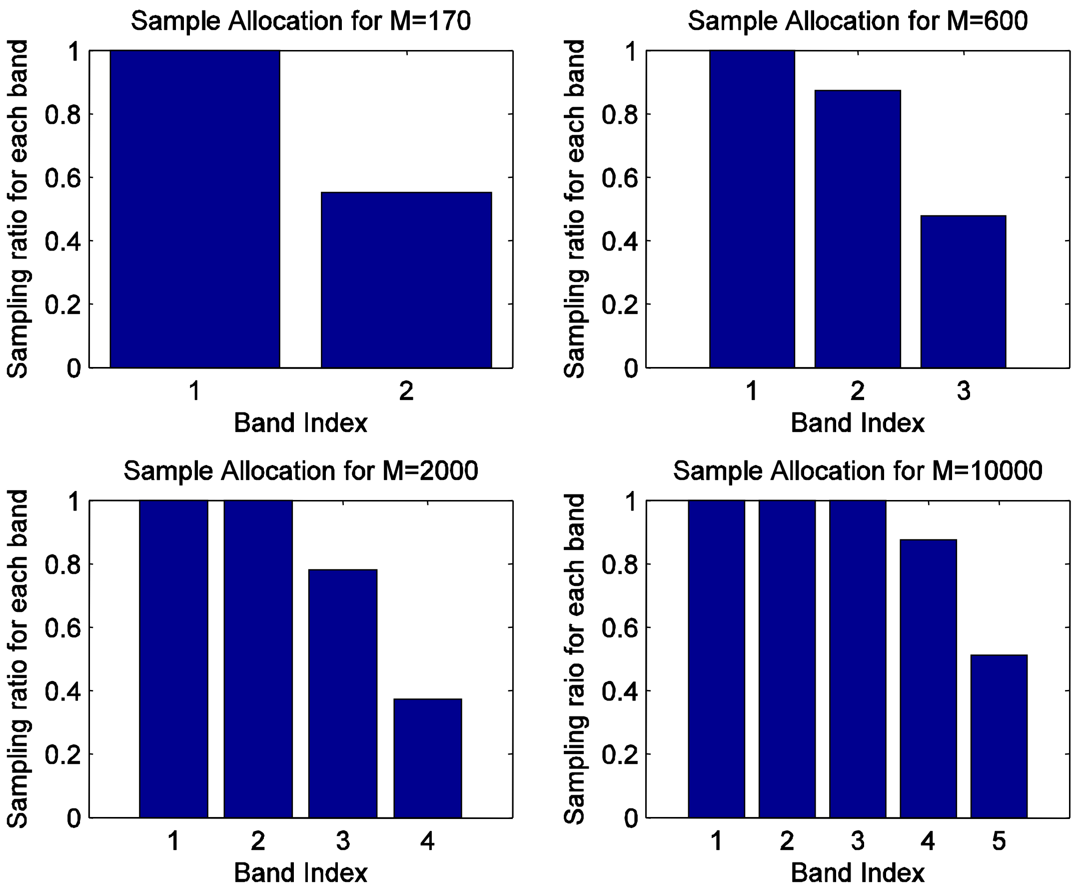
DR fun for cameraman image



The optimization works for any convex SD function, including L_2 SD function and lower bounds (EBB, MMB)

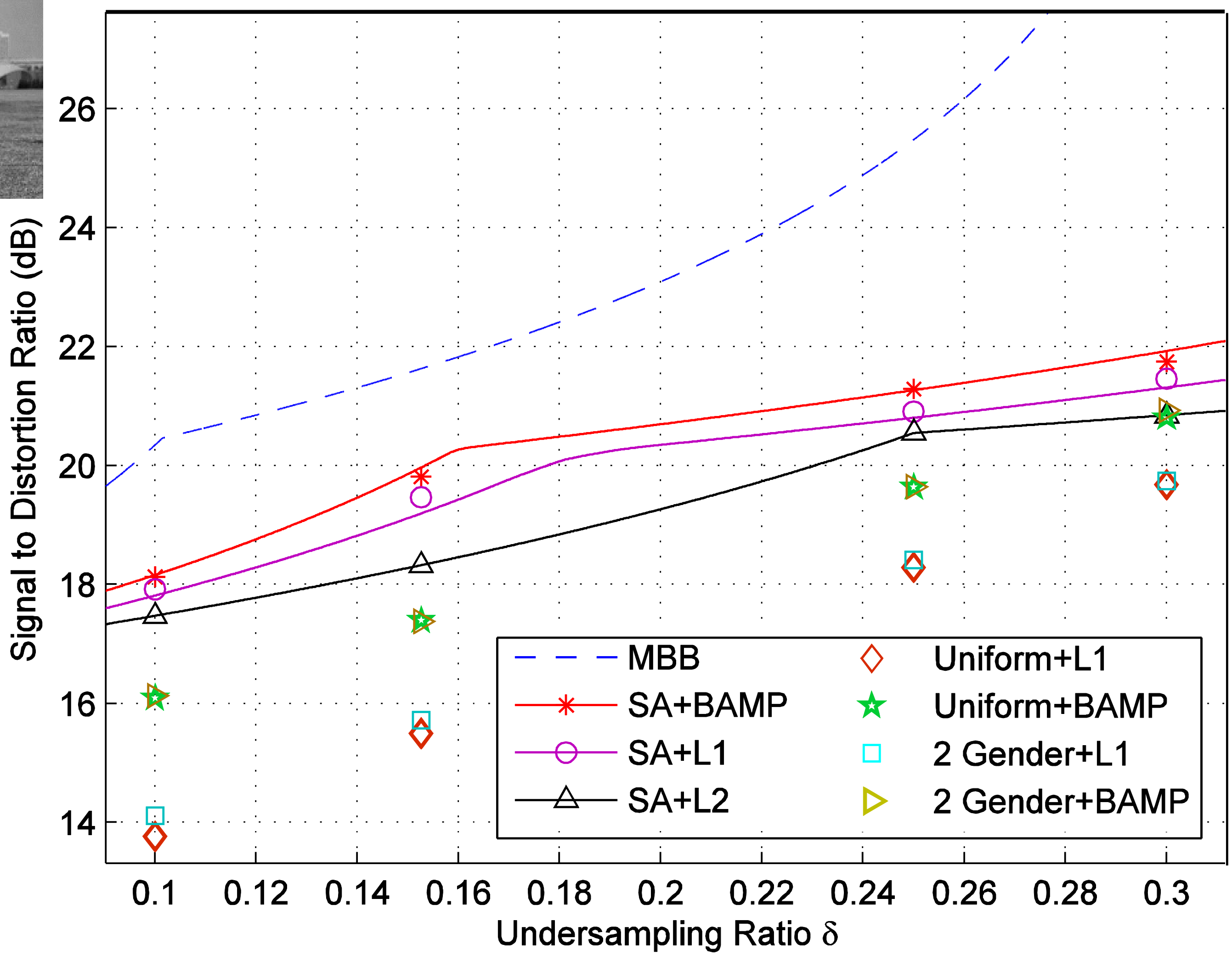
Bandwise CS sample allocation

Sample allocation (% of full sampling) per band for m = 170, 600, 2000 and 10000 measurements. There are typically no more than 2-3 partially sampled bands



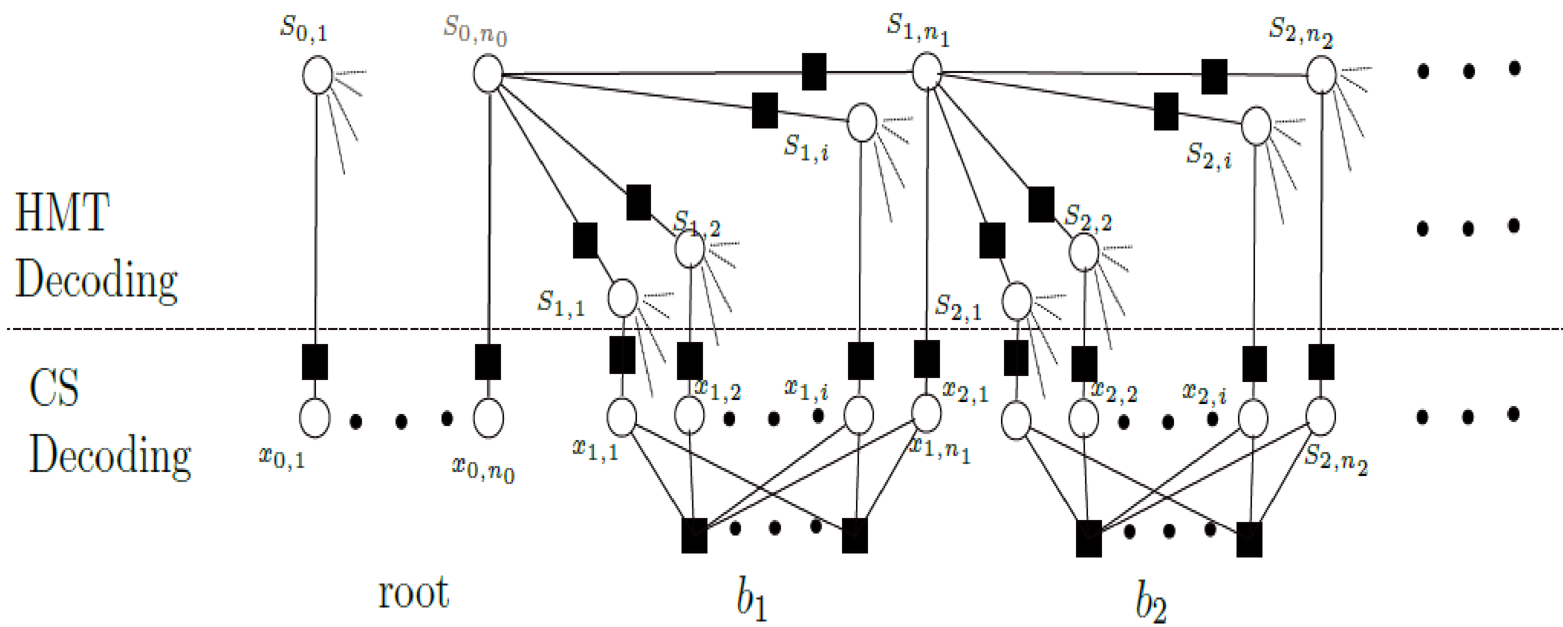
Bandwise CS Performance

e.g. cameraman



adding Tree Structure

Incorporating Tree Structure

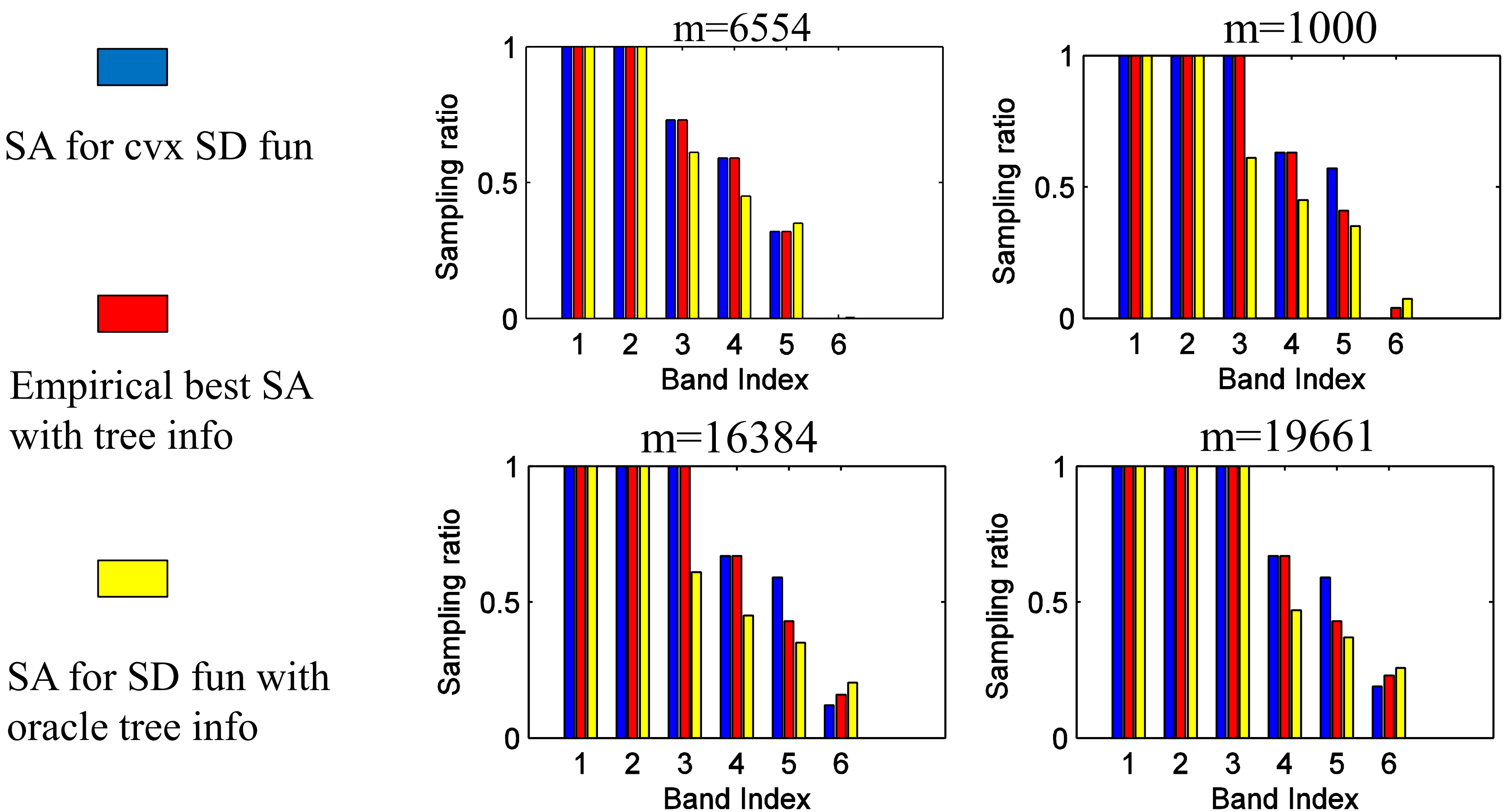


We can add tree-based priors on coefficients and decode using Turbo AMP scheme [Som, Schniter 2012]:

This calculates marginal probabilities for hidden states and incorporate into MMSE AMP

Bandwise CS Sample Allocation

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



Bandwise CS Performance

e.g. cameraman

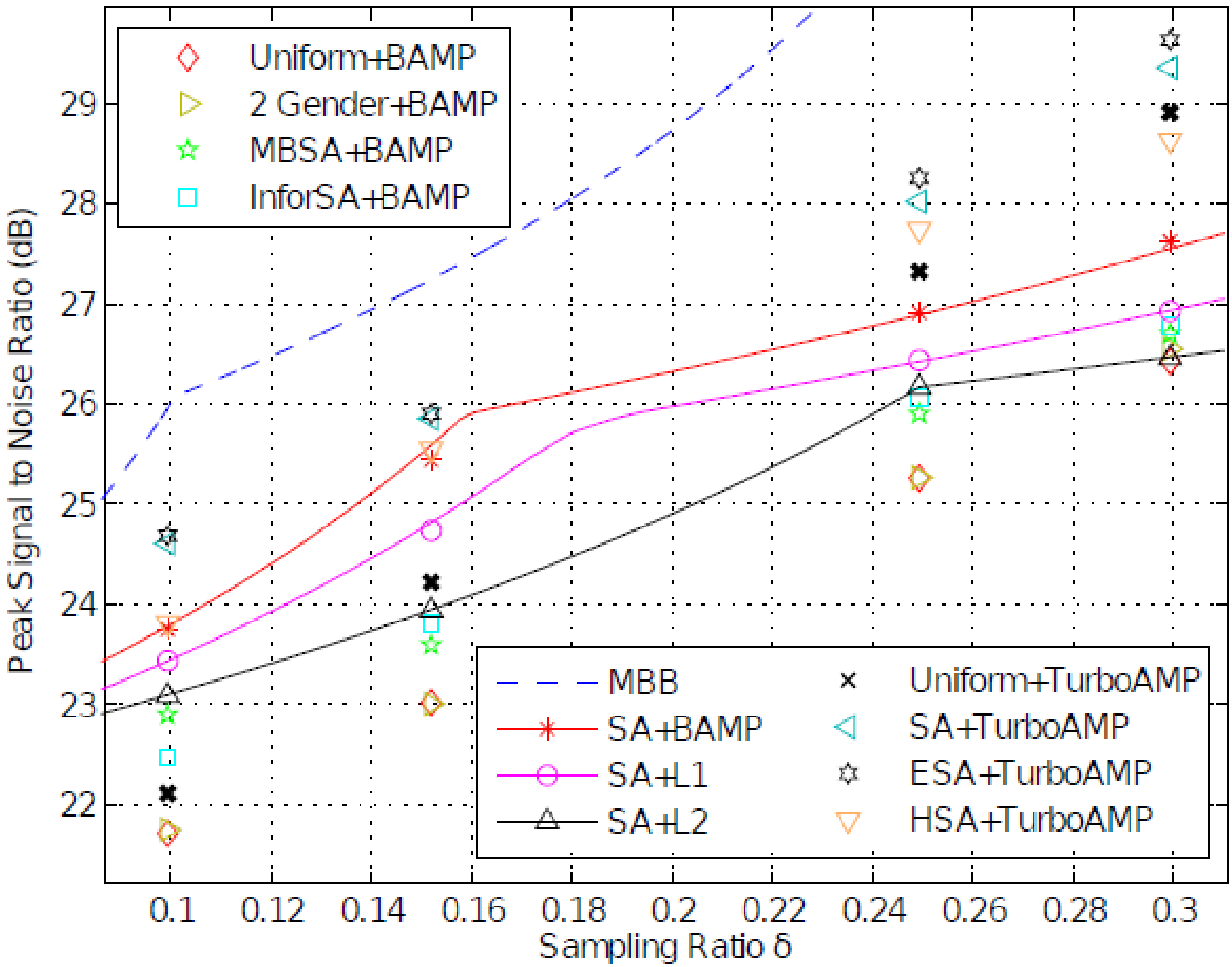


Image reconstructions from 10000 measurements (15%)

SA for General Image Statistics

Open questions

- How to derive sample allocations for more sophisticated models? – analysis representations, tree structured model, etc.

- How to allocate samples within constrained sampling schemes (e.g. partial Fourier)?

References

- R. Gribonval, V. Cevher, and M. E. Davies. *Compressible Distributions for High-dimensional Statistics*. Preprint, 2011, available at arXiv:1102.1249v2.
- M. E. Davies and C. Guo, *Sample-Distortion Functions for Compressed Sensing*, 49th Allerton Conf. on Communication, Control, and Computing, 2011.
- C. Guo and M. E. Davies , *Sample-Distortion for Compressed Imaging*, on IEEE TSP, 2013.

Thank You