ICERM Workshop

U.S. NAVAL RESEARCH LABORATORY

Hierarchical Bayesian Approaches to Imaging and Compressive Sensing

Dr. Raghu G. Raj

Radar Division, Code 5313

raghu.raj@nrl.navy.mil

October 2017

UNCLASSIFIED

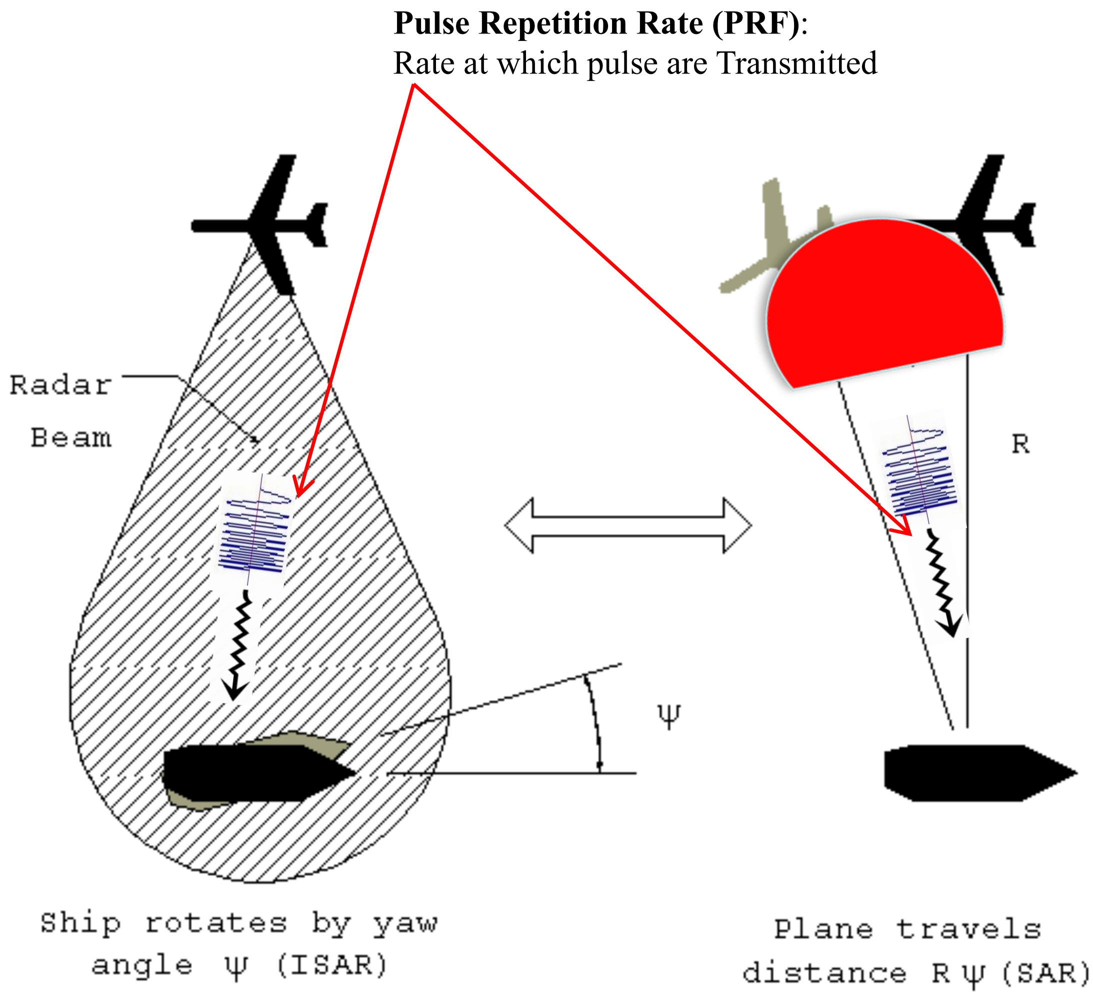
Distribution A: Approved for Public Release

Outline

- Introduction and Motivation
- System Setup and Problem Formulation
- Sparsity Inducing Priors: Radar Clutter/Prior Models
  - CG, NCG, Graphical Models
- Hierarchical Bayesian Algorithms for Imaging
- Fast Stochastic Algorithm for HB-MAP
- Discussions

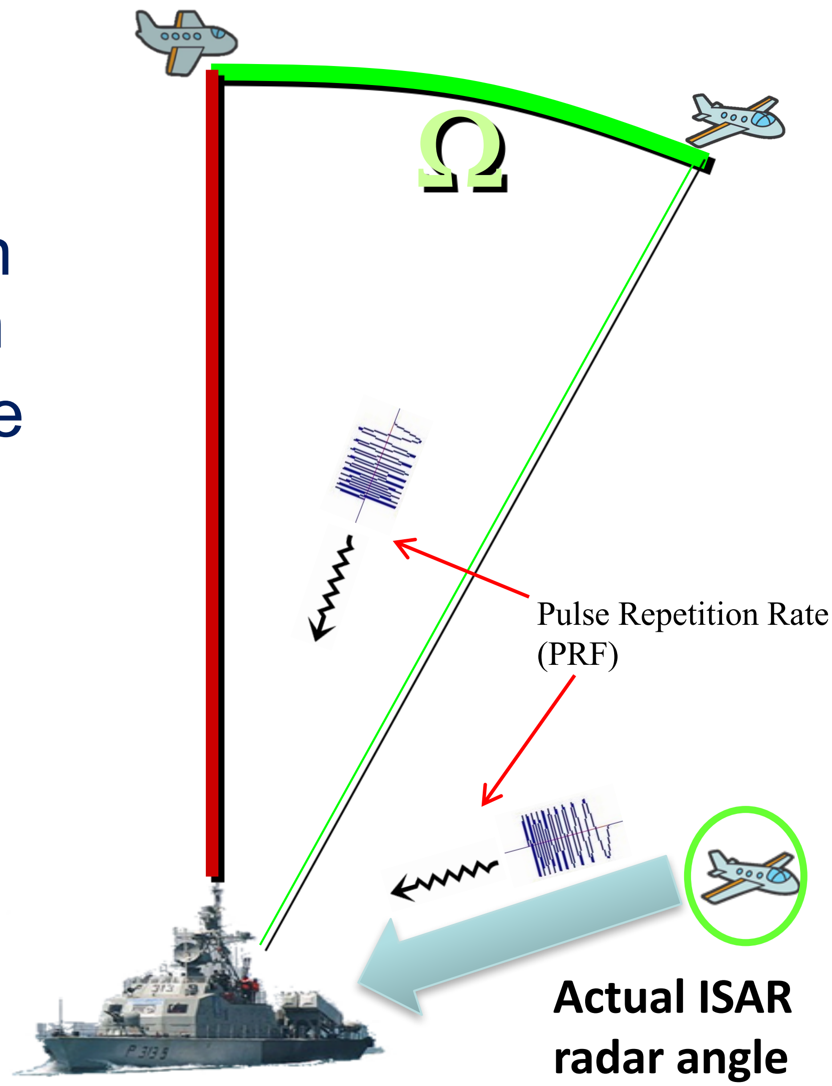
Synthetic Aperture

- Can create large effective apertures through motion
- **Synthetic Aperture Radar (SAR)** moves the platform + radar to view static targets
- **Inverse Synthetic Aperture Radar (ISAR)** uses static radar to view moving targets



ISAR: Inverse Synthetic Aperture Radar

- Motion of the target causes a Doppler shift that depends on the angular velocity of the target motion and distance from center of rotation
- Like moving a SAR radar across the sky by Ω as target is rocked by waves (faster)
- Excellent for a moving (rocking) target at *extremely long ranges*
- Allows imaging and target recognition from closer to standoff ranges and lower grazing angles



Introduction and Motivation (3)

- Radar systems sense the environment by transmitting and receiving waveforms sampled through a finite effective aperture
  - Aperture is created in two primary ways:
    1. Motion between radars and targets resulting in relative aspect changes (which manifests in terms of Doppler structure of backscattered signal)
    2. Distributed sensor structures (for example: Multi-static scenarios)

- Typically in radar systems the aperture is densely sampled; for example:
  - Large CPI (Coherent Processing Interval i.e. observation time) in ISAR imaging
  - Large aperture created due to motion of the aircraft in SAR imaging

- Nevertheless, even in such scenarios there is a need for enabling radar systems to perform robust inference/imaging when the aperture is sparse
  - We refer to such scenarios as ‘Sparse Sensing’ i.e. limited number of pulses fall on a target of interest

Motivation for Sparse Sensing Radar Imaging

- **Example#1: ISAR (Inverse Synthetic Aperture Radar) Imaging**
  - **Fundamental Problem in ISAR Imaging: Motion estimation errors due to complexities in motion dynamics**
  - **When viewed from the lens of Fourier processing and Backprojection, we need Large CPI so that we have a large enough aperture to form a high-quality image**
  - **However motion of target can be very complex and non-linear in large CPI—motion compensation (mocomp) more difficult**
  - **Solution: Imaging in Small CPI (sparse aperture) so that the target motion can be assumed to be linear i.e. simpler mocomp—however alternatives to Fourier based imaging is needed**

- **Example#2: SAR (Synthetic Aperture Radar) Imaging**
  - **Strip-map SAR modality is capable of imaging a large coverage area; however the number of pulses that interrogate any particular target of interest is likely relatively small**
  - **By enabling robust inference via 'Sparse Sensing Radar Imaging' techniques, particular targets of interest can be imaged at higher resolution than otherwise possible via backprojection techniques**

Motivation for Sparse Sensing Radar Imaging

- **Example#3: Image-While-Scan (IWS)**
  - The concept here is that a Radar is in scanning mode (i.e. rotating antenna) and updates the surface picture with each sweep of the antenna—such that it images multiple targets without the need to invest separate dwell times for each individual target
  - The difficulty here is that we have only a limited number of pulses to from which to form Doppler spectrum at each range-bin
  - This is an example where there the Sparse Sensing scenario directly applies

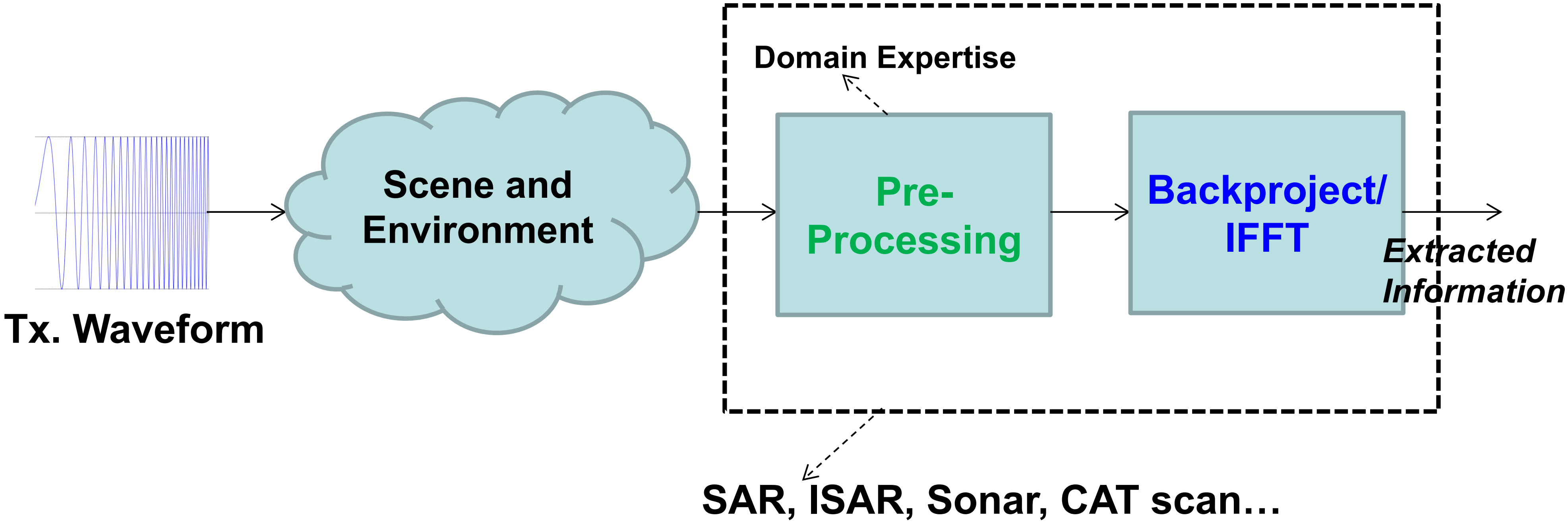
Motivation for Sparse Sensing Radar Imaging

- **From the preceding examples it is clear that there is a need to develop statistical inference techniques that can perform Radar Imaging under the constraints of Sparse-Sensing—for e.g. i) limited number of pulses sampling different targets aspects or ii) limited number of spatially distributed sensors)**
  - **Key idea: to systematically incorporate *prior (probabilistic) knowledge of the scene structure* into the inference process**

- **Importantly, an added benefit of our approach—i.e. exploiting prior knowledge of scene structure—is that the resulting methods are potentially useful even when a dense number of pulses impinges on a target interest**
  - **Especially where there is significant degradation of the received signal due to corruptions arising from environmental and other factors**
  - **Computational complexity of the inference technique is an issue however**

Overview

- **Traditional approach to scene estimation etc. is to employ the fundamental tools of matched filtering and various pre-processing steps followed by Fourier spectral analysis**



- **Strategy:**
  - ➢ **Keep the Pre-processing (range migration compensation etc.) in-tact i.e. unchanged**
  - ➢ **Replace the Backprojection/IFFT block with 'something better'**

Overview (2)

- Implicitly, Fourier based spectral analysis represents the data in Fourier bases. However wavelet based representations of signals have significant advantages over Fourier representations:
  - Radar scenes have sparse structure in wavelet representations
  - Radar scenes reveal a rich statistical structure in wavelet representations
- Sparsity-based reconstruction algorithms (such as in Compressive Sensing (CS)) try to exploit the sparse structure of such signals in order to better extract information from the measurements
  - However sparsity is only a crude measure of the probabilistic structure of radar images. Thus the challenge is to come up with novel ways of incorporating informed prior models for radar scenes into an elegant optimization framework

Outline

- Introduction and Motivation
- **System Setup and Problem Formulation**
- Sparsity Inducing Priors: Radar Clutter/Prior Models
  - CG, NCG, Graphical Models
- Hierarchical Bayesian Algorithms for Imaging
- Fast Stochastic Algorithm for HB-MAP
- Discussions

System Setup and Problem Formulation

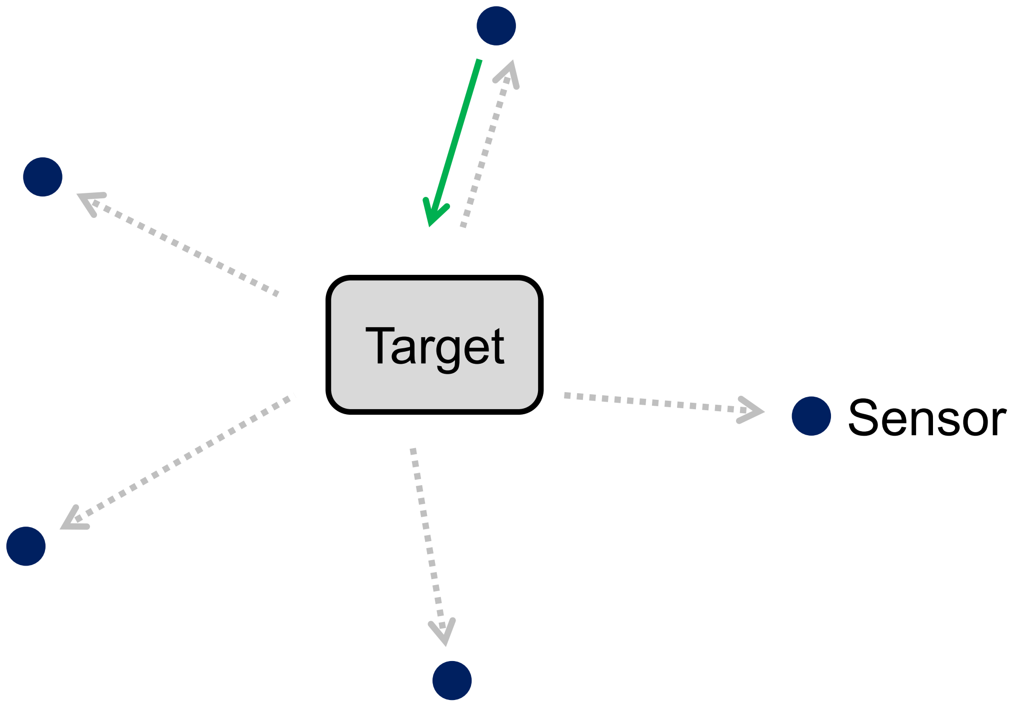
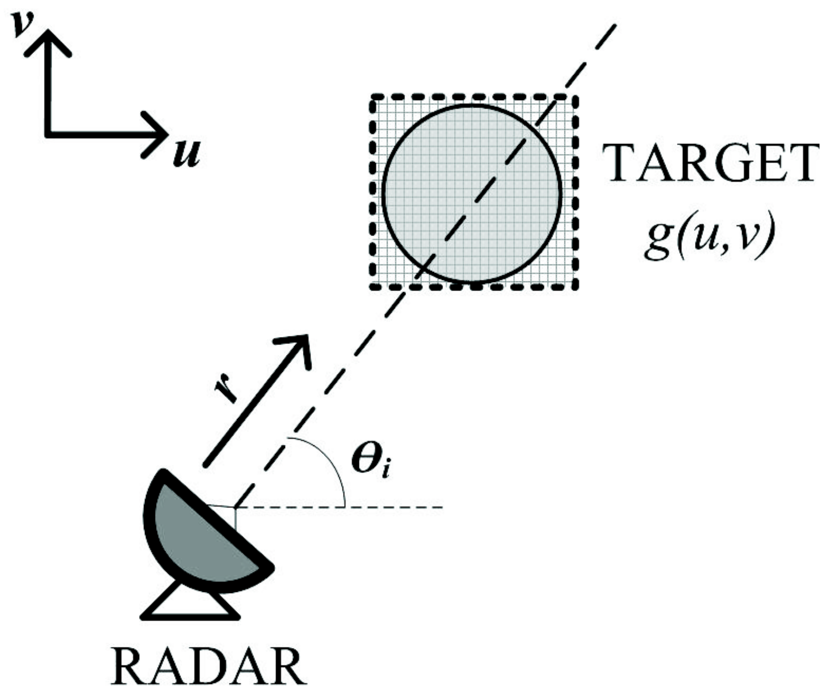


Fig. Multistatic Radar Imaging System

- Conditioned on radar pre-processing steps (mocomp etc.), the correct abstraction of the radar imaging problem (in any modality such as ISAR, SAR etc.) is embodied in the *Multistatic Radar Imaging setup*:
  - ***M*** radar sensors interrogating the scene/target at different angles
  - Target scene is assumed to be stationary

Radar System (2)



- Focus on single radar return first
- Assume discrete-time system (i.e., inherent sampling time, T_s)
- Discrete range index, r
- Aligned at a positive angle θ_i with the image axes

Radar System (3)

- Transmit waveform
  $x[k]$ for $k = 1, 2, \ldots, N$
- Power constraint
  $$\frac{1}{N}\sum_{k=1}^{N} x^2[k] \leq \rho$$
- Reflectivity Function
  $g(u, v)$ for $u = 1, 2, \ldots$ and $v = 1, 2, \ldots$
- Response from single image point
  $$y_{(u_0,v_0)}[k] = g(u_0, v_0)x[k - D(r_0)]$$

Two-way time delay

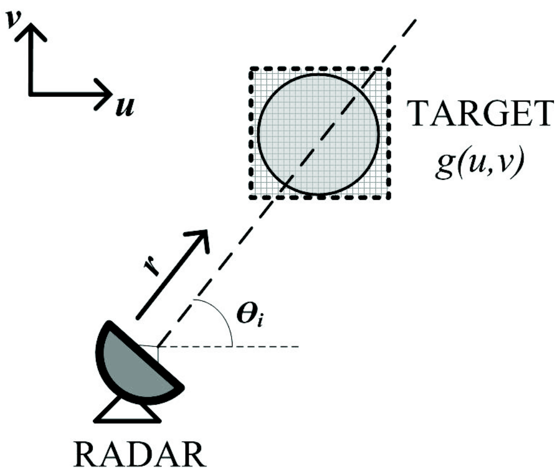


Fig. Radar Imaging System

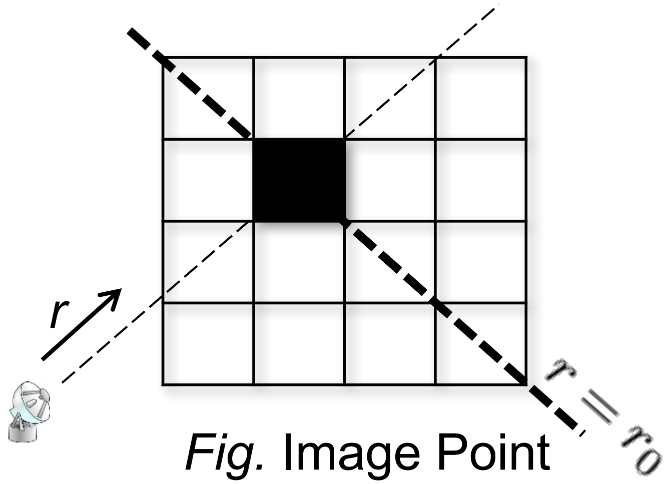


Fig. Image Point

Radar System (4)

- Response from constant range (a.k.a. *iso-range contour*)

$$y_{r=r_0}[k] = \mathcal{R}_{\theta_i}\{g(u,v)\}[r_0]x[k-(D_0+r_0)]$$

Two-way time delay from center of image

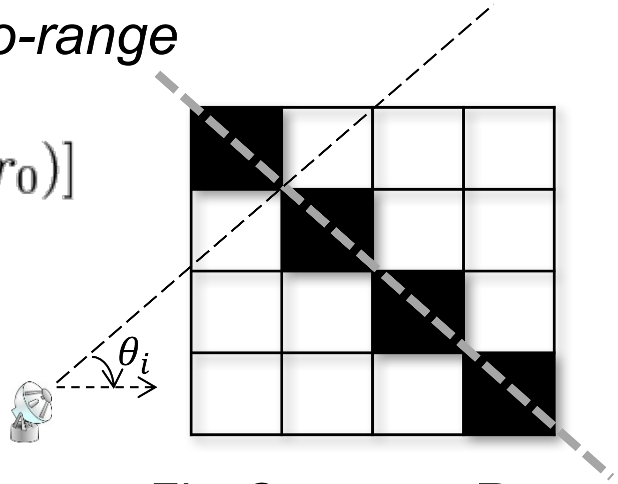


Fig. Constant- Range Cut

- Total response

$$\begin{aligned} y[k] &= \sum_r \mathcal{R}_{\theta_i}\{g(u,v)\}[r]x[k-D_0-r] \\ &= (p_{\theta_i} * x)[k-D_0] \end{aligned}$$

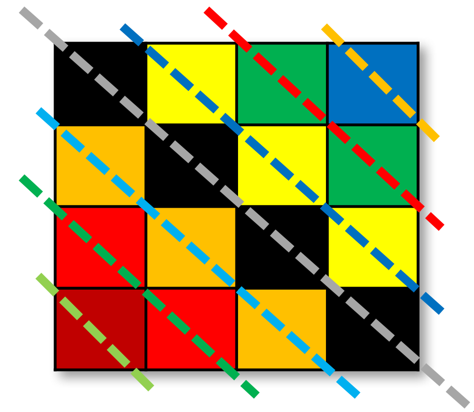


Fig. Total Response

- Discrete Radon transform

$$p_{\theta_i}[k] = \mathcal{R}_{\theta_i}\{g(u,v)\}[k]$$

Image Structure

- Define matrix version of image, $\boldsymbol{G}$

$$\mathbf{G} \in \mathbb{R}^{I_r \times I_c} : g_{i,j} = g(i,j)$$

  - I_r, I_c: number of rows and columns in the image

- Construct image on a linear basis (a.k.a. *sparse approximation*)
  - For e.g. use two-dimensional wavelets

- Implies that the image can be written as

$$\mathbf{G} = \mathbf{\Phi c}$$

  - $\mathbf{\Phi} \in \mathbb{R}^{(I_r I_c) \times D}$ is a ***dictionary of basis vectors (e.g. wavelets etc.)***
  - $\mathbf{c} \in \mathbb{R}^{D \times 1}$ is a random vector of wavelet coefficients
    - ***Distribution of*** $\mathbf{c}$ ***will be discussed later***

Monostatic Response

- Using the sparse approximation model for the image, the total response in vector form is

$$\mathbf{y} = a\mathbf{R}_{\theta_i}\mathbf{\Phi}\mathbf{c} * \mathbf{x} + \mathbf{n}$$

  - $\mathbf{y} = [y[1], y[2], \ldots, y[N+L-1]]^T$
  - a is a path loss coefficient
  - $\mathbf{R}_{\theta_i} \in \mathbb{R}^{L\times(I_r I_c)}$ is the matrix form of the Radon transform
  - $\mathbf{n} \in \mathbb{R}^{(N+L-1)\times 1}$ is an additive white Gaussian noise vector
    - $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$

- We can now extend this system to the multistatic scenario

Multistatic Response

- Multistatic system can be broken into bistatic pairs
  - Simpler to focus on bistatic case
- Interested in ***multiradar*** setup
  - Only Radar   transmits, all others receive
  - We assume RCS fluctuations are isotropic
- Using previous system model, the response at Radar

$$\mathbf{y}_i = a_{i,i}\mathbf{R}_{\theta_i}\mathbf{\Phi}\mathbf{c} * \mathbf{x} + \mathbf{n}_i$$

- Using a theorem from [Crispin59], we also have the return at Radar

$$\mathbf{y}_j = a_{i,j}\mathbf{R}_{\theta_b}\mathbf{\Phi}\mathbf{c} * \mathbf{x} + \mathbf{n}_j$$

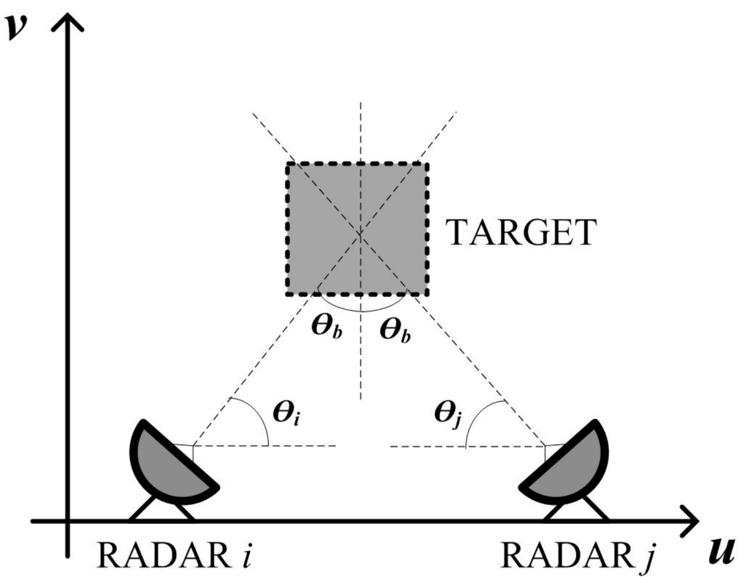


Fig. Bistatic Radar System

Multistatic Response (2)

- For simplicity, define $\boldsymbol{R_i} = a_i \boldsymbol{R_{\theta_i}}$ and form the convolution matrix $\boldsymbol{X}$:

$$\mathbf{X} = \begin{bmatrix} x[1] & 0 & & \cdots & 0 \\ x[2] & x[1] & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & x[N] & x[N-1] \\ 0 & \cdots & & 0 & x[N] \end{bmatrix}$$

- Simplified system model

$$\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{R}_i\boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{n_i}$$
$$\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{R}_j\boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{n_j}$$

- This can be done for all bistatic pairs
  - Only matters for pairs involving the transmitting radar array

Multistatic Response (3)

- All responses

$$\boldsymbol{y}_1 = \boldsymbol{X}\boldsymbol{R}_1\boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{n}_1$$
$$\vdots$$
$$\boldsymbol{y}_M = \boldsymbol{X}\boldsymbol{R}_M\boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{n}_M$$

- Gathering all *M* returns together, we form the overall response:

$$\boldsymbol{y} = \overline{\boldsymbol{X}}\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{n}$$
$$\Longrightarrow \boxed{\boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{n}}$$

  - $\overline{\boldsymbol{X}}$ is a block diagonal matrix with *M* copies of $\boldsymbol{X}$ down the diagonal
  - $\boldsymbol{R} = \left[\boldsymbol{R}_1^T, \dots, \boldsymbol{R}_M^T\right]^T$
  - $\boldsymbol{n} = \left[\boldsymbol{n}_1^T, \dots, \boldsymbol{n}_M^T\right]^T$
  - $\boldsymbol{\Psi} = \overline{\boldsymbol{X}}\boldsymbol{R}$

Problem Formulation

- **We have the following mathematical model for the radar system:**

$$\boldsymbol{y} = \boldsymbol{\Psi}\,\boldsymbol{\Phi}\,\boldsymbol{c} + \boldsymbol{v} = \boldsymbol{\Psi}\,\boldsymbol{I} + \boldsymbol{v} \qquad (1)$$

where,

$\boldsymbol{y} \in \mathbb{R}^m$ **is the Measured signal (after pre-processing)**

$\boldsymbol{\Psi} \in \mathbb{R}^{mxp}$ **is the effective Sensing matrix (i.e. Waveform)**

$\boldsymbol{\Phi} \in \mathbb{R}^{pxn}$ **is the Dictionary matrix in which the image is represented (Wavelets, Fourier, Time-Frequency Bases, etc.)**

$\boldsymbol{c} \in \mathbb{R}^n$ **is the underlying coefficients to be estimated (the resulting Radar Image Estimate is** $\mathbf{I} = \boldsymbol{\Phi c}$**)**

$\boldsymbol{v} \in \mathbb{R}^n$ **is the interference in the measurements due to both** ***clutter and noise***

In this talk we focus on novel techniques of solving for $\boldsymbol{c}$ in (1):

Contribution#1: Novel Hierarchical Bayes algorithm via Probabilistic Graphical Extension of Compound Gaussian (CG) model
#2: Fast algorithmic extensions

Outline

- Introduction and Motivation
- System Setup and Problem Formulation
- Sparsity Inducing Priors: Radar Clutter/Prior Models
  - CG, NCG, Graphical Models
- Hierarchical Bayesian Algorithms for Imaging
- Fast Stochastic Algorithm for HB-MAP
- Discussions

Sparsity Inducing Priors

- **Consider a Bayesian-MAP Formulation:**

$$c^* = \underset{c}{\operatorname{argmax}} \log P(c|y) \quad (2)$$

$$\propto \underset{c}{argmax}\{log\,P(y|c) + log\,P(c)\} \quad (3)$$

$$= \underset{c}{\operatorname{argmin}}\left\{\|y - \Psi\Phi c\|_2^2 - log\,P(c)\right\} \quad (4)$$

where, in (4) we assume that $v \sim \mathcal{N}(0, \Sigma_v)$

- **For the specific choice $P(c) \propto exp(-\lambda\|c\|_1)$ i.e. Laplacian distribution, (4) reduces to:** $c^* = \underset{c}{argmin}\left\{\|y - \Psi\Phi c\|_2^2 + \lambda\|c\|_1\right\}$ **(5)**

  which is the well known l_1 sparsity promoting optimization problem

- **Thus sparsity promoting algorithms such as (5) can be viewed a special cases of Bayesian inference algorithms wherein sparsity-inducing priors such as Laplacian distributions are incorporated**

In Search of Richer Priors

- From our experience, the Laplacian is not a rich enough prior for modeling radar and natural images. In search of richer priors we choose to *build upon* the so-called CG (Compound Gaussian) distribution
  - The Compound Gaussian (CG) is a special type of compound distribution whose structure is as follows:

$$P(c) = \int \frac{1}{\sqrt{2\pi}|\Sigma|z} exp\left(-(c/z)^H \Sigma^{-1}(c/z)\right) P(z).dz \qquad (6)$$

- Why build upon CG?
  - Many of the most well known distributions are special cases of CG, such as: Laplacian, Gamma, Student, Generalized Gaussian, Pareto, Alpha-stable, etc.
  - It is very convenient to incorporate this within an optimization framework via Hierarchical Bayesian modeling

Statistics of Coefficient Vector c : CG Prior Model (Local CG Model)

- **Let us plot the Histogram of coefficients $c \in \mathbb{R}^n$ for a typical SAR image:**

Compound Gaussian (CG) Model of c:

$$c = z \odot u \qquad \textbf{(7)}$$

where:

$u \sim N(0,\ \sigma_u)$ is a Gaussian r.v.

$z \geq 0$ is a Non-Gaussian r.v.

Calculate Wavelet Coefficients of the SAR Image and plot its Histogram (shown in Blue)

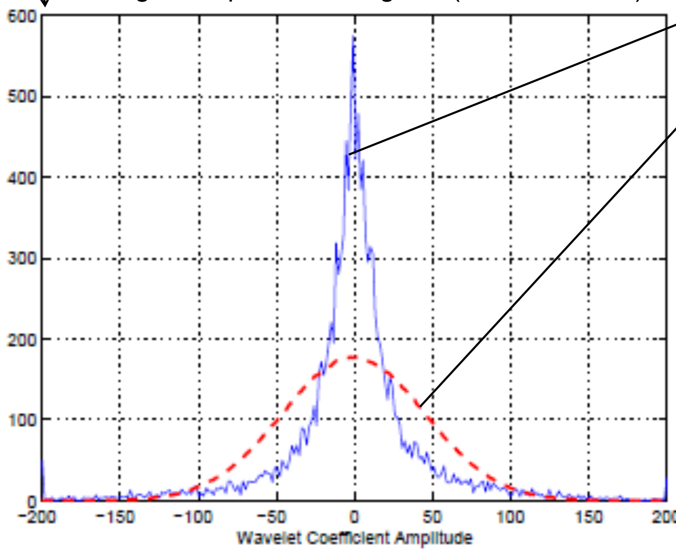


Figure: Variance-matched Gaussian. (ΔH/H) = 0.0549

Histogram of Wavelet Coefficients c

Best Gaussian Match

1) Estimate vector z (given c in this case)
2) Normalize vector c: $u(i) = c(i) / z(i)$ (for each vector component i)
3) Plot Histogram of u and its Best Gaussian Match (shown in Red Font)

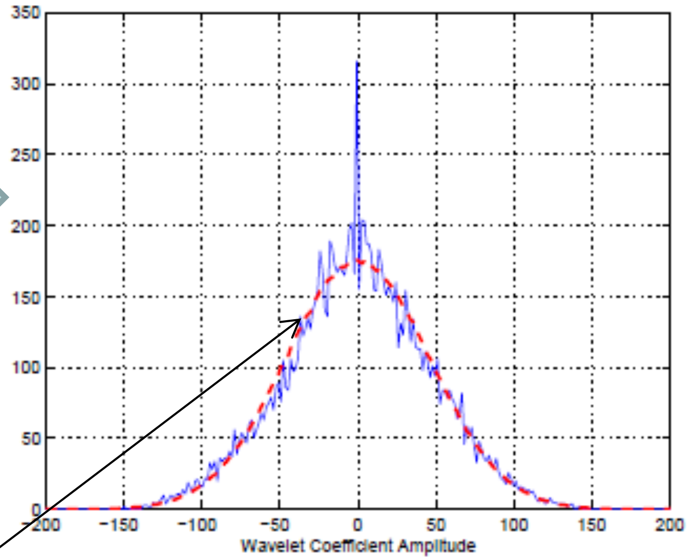


Figure: Variance-matched CG (Compound Gaussian). (ΔH/H) = 0.0023

Local Image Statistics

- Consider the local image statistics of a **SAR image**

$$\boldsymbol{c_i} \sim z_i \boldsymbol{u_i}$$

- The vector of the *i*th neighborhood coefficients can be written as
  - z_i is a non-negative, random scalar
  - z_i has pmf $p_z(z)$
  - $\boldsymbol{u_i} \sim \mathcal{N}(0, \Sigma_u)$

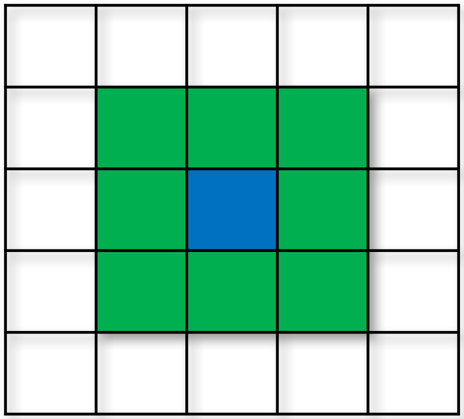


Fig. Neighborhood of Wavelet Coefficients

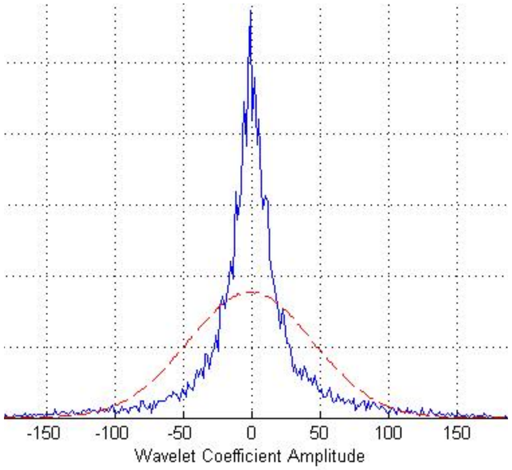


Fig. Histogram of Wavelet Coefficients **of a SAR Image**

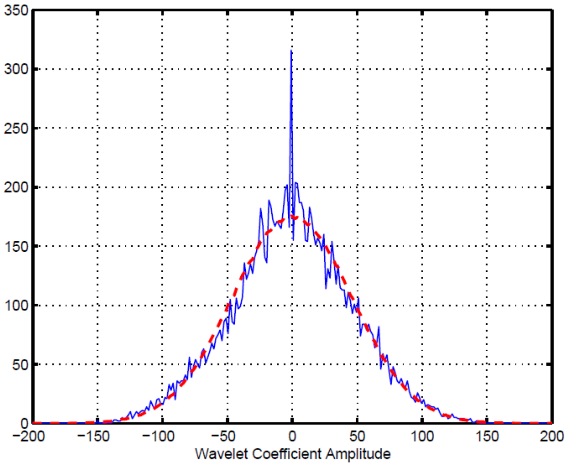


Fig. Histogram of corresponding *z-Normalized* Wavelet Coefficients

Generalizations of CG

- The CG distribution can be generalized in different directions:
  1. Analytical Generalization of CG:
     – Non-linear CG (NCG) distribution [Raj et. al. 2012]
  2. Graphical Extensions of CG [Wainwright et. al. 2001]
     – This will be used in our global image formation algorithm…

Source Statistics: Global CG Model

- <u>The need for a Global CG Model for Imaging Applications:</u>
  - Local CG model such as shown in the previous slide do NOT capture (non-linear) spatial correlation information of wavelet coefficients
  - The critical step is to estimate the non-Gaussian z- field:
    - For imaging applications, unlike the previous slide, the wavelet coefficients are completely inaccessible. Local CG model does NOT allow z-field to be estimated in such cases.
    - A Global CG model models the non-linear statistical interactions of coefficients across scale and space (via a Probabilistic Graphical Model). Exploiting this information (~nonlinear covariance model) allows us to estimate the z-field.

Global Compound Gaussian (CG) Model:

We model $\boldsymbol{c}$ as a random vector that can be decomposed into the following (Hadamard) product form:

$$\boldsymbol{c} = \boldsymbol{z} \odot \boldsymbol{u}$$

such that:

1) $u \sim \mathcal{N}(0, P_u)$, $z = h(x)$, and x follows a *Multi-scale Gaussian Tree structure*
2) u and z are independent random variables
3) $\mathbb{E}[z^2] = 1$
4) h is a non-linearity (which ultimately controls the sparse structure of c)

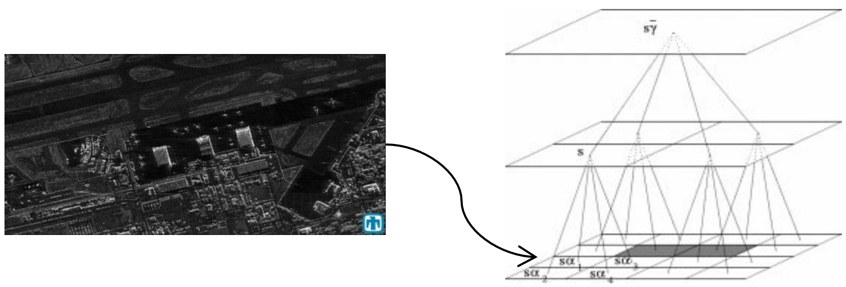


Figure: *Quad-tree structure that captures the non-Gaussian interaction between wavelet coefficients across scales.*

Graphical Extensions of CG:
Tree-based CG model

- **The CG models we have considered thus far is a local-patch based model i.e. it captures the statistics of local neighborhoods of an image under *i.i.d.* sampling**
  - **Can be employed in local-patch based image reconstruction**
- **However such models do not model the global structure of images**
  - **Hence are unsuitable for Global image estimation**
- **For global reconstruction strategies for images, we need a global statistical model for the image that is locally consistent with CG**
  - **This can be accomplished via Probabilistic Graphical Model Extensions of the CG distribution**

Quick Overview of Probabilistic Graphical Models

- **(Undirected) Graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined by a set of nodes $\mathcal{V} = \{1, \cdots, n\}$, and a set of edges $\mathcal{E} \subset \binom{\mathcal{V}}{2}$**
- **Graphical Model: Random vector defined on a graph; nodes represent random variables, edges reveal conditional dependencies**
- **Graph structure defines factorization of joint probability distribution**

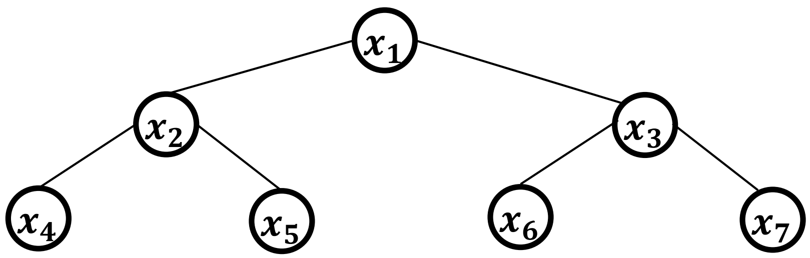


Figure: Tree – acyclic graph with n nodes and (n-1) edges (here, n=7)

$$P(x) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2)P(x_5|x_2)P(x_6|x_3)P(x_7|x_3)$$

- **Given a graphical model very efficient algorithms (such as Message Passing) exist for making statistical inferences on such graphs**

Graphical CG Model

- **The specific model that we use for modeling Global CG:**

$$\boldsymbol{c} = \boldsymbol{z} \bullet \boldsymbol{u}$$

where, $\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{P_u})$

$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x})$

$\boldsymbol{u}$ **and** $\boldsymbol{z}$ **are independent random variables**

$\mathbb{E}[\boldsymbol{z}^2] = \mathbf{1}$ **(i.e.** ***variance of*** $\boldsymbol{c}$ ***is controlled by*** $\boldsymbol{u}$**:** $\boldsymbol{u}(\boldsymbol{s}) = \mathbf{2}^{-\boldsymbol{\gamma}}\mathcal{N}(\mathbf{0}, \boldsymbol{I})$

and where, $\boldsymbol{x} \sim$ **Multi-scale Gaussian Tree structure**

$\boldsymbol{x}(\boldsymbol{s}) = \boldsymbol{A}(\boldsymbol{s})\ \boldsymbol{x}(\boldsymbol{par}(\boldsymbol{s})) + \boldsymbol{B}(\boldsymbol{s})\boldsymbol{w}(\boldsymbol{s}) \qquad \boldsymbol{P_x}(\boldsymbol{s}) = \boldsymbol{A}(\boldsymbol{s})\boldsymbol{P_x}(\boldsymbol{par}(\boldsymbol{s}))\boldsymbol{A}(\boldsymbol{s})^{\boldsymbol{T}} + \boldsymbol{Q}(\boldsymbol{s})$

- **In our simulations we instantiated the Graphical CG model as follows:**

1. We employed the following non-linearity in our simulations: $\boldsymbol{h}(\boldsymbol{x}) = \sqrt{\boldsymbol{exp}}(\boldsymbol{x}/\boldsymbol{\alpha})$

   where, α controls the sparsity-level in the generated signal: smaller the α, sparser the signal

2. We set $A \equiv \boldsymbol{\mu}$ and $B \equiv \sqrt{1-\mu^2} \Rightarrow P_x(s) = I_d\ \forall s.$

   Given this the entire covariance matrix corresponding to the Gaussian process $\boldsymbol{P_x}(\boldsymbol{s}, \boldsymbol{t}), \forall \boldsymbol{s}, \boldsymbol{t}$ can be calculated by a simple set of recursive equations

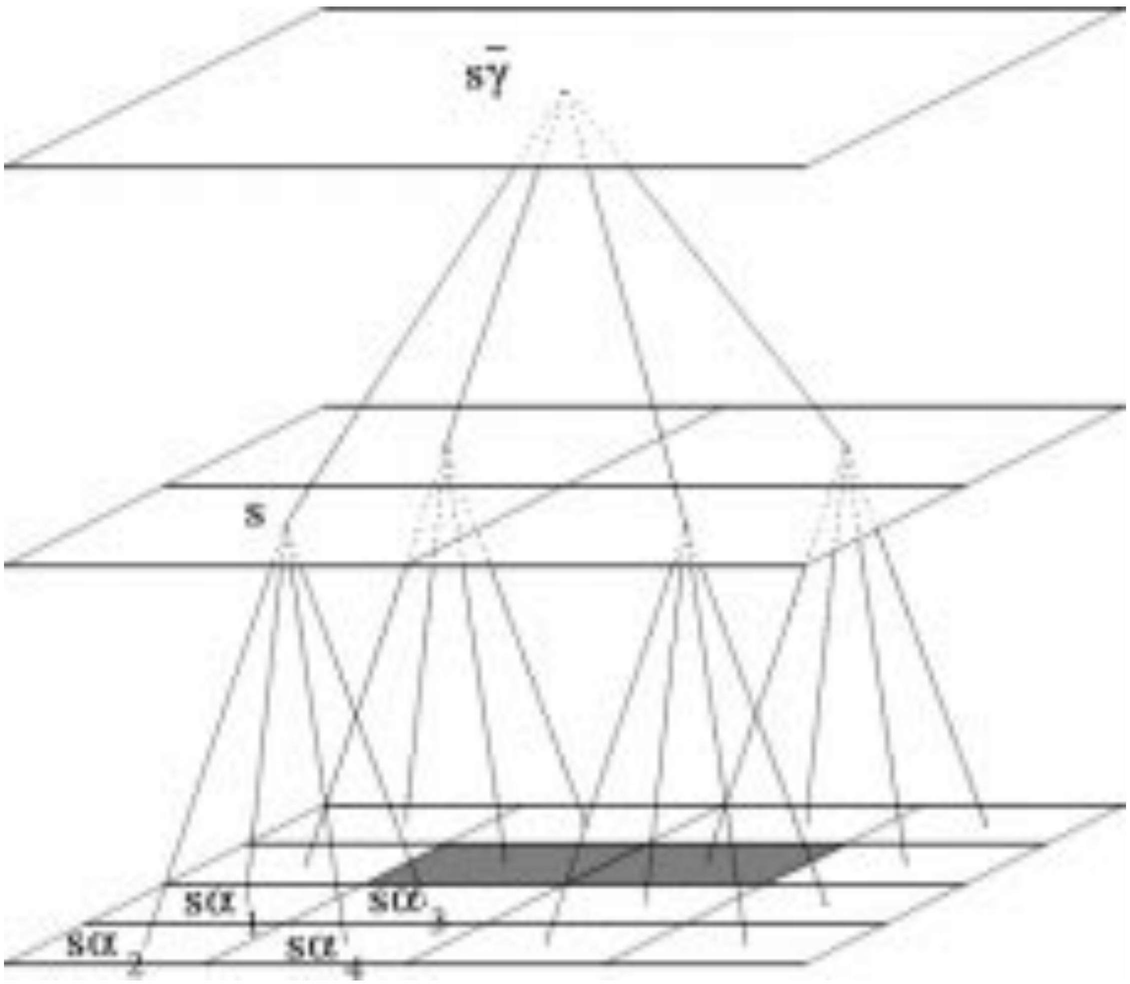
3. **Thus 3 parameters are associated with the Graphical CG model:**

   $\boldsymbol{\alpha}$ (controls sparsity)

   $\boldsymbol{\mu}$ (controls inter-scale dependency structure)

   $\boldsymbol{\gamma}$ (controls distribution of variance (power) across scales)… (currently in our simulations we set $\boldsymbol{\gamma}$=0)

Mapping Image Space to Graphical Models



(Figure taken from [Wainwright et. al. 2001])

Outline

- Introduction and Motivation
- System Setup and Problem Formulation
- Sparsity Inducing Priors: Radar Clutter/Prior Models
  - CG, NCG, Graphical Models
- Hierarchical Bayesian Algorithms for Imaging
- Fast Stochastic Algorithm for HB-MAP
- Discussions

Hierarchical Bayesian Algorithms for Imaging [1-2]

- **We focus here on the Global image reconstruction problem: i.e. estimate c given measurements y**

$$y = \Psi\Phi c + v = \tilde{\Psi} c + v \tag{8}$$

  - **We employ the Tree-based CG Graphical model as our Global Prior on the unknown coefficients *c***
    - The CG model subsumes Laplacian and other distributions as special cases

- **Due to the presence of the integral structure in the CG pdf expression, it is difficult to optimize with respect to the complete Graphical CG prior model**
  - **Therefore we approach this via a Hierarchical Bayes-MAP perspective wherein we first estimate the optimum z- field (i.e. Type-II estimation) followed by estimation of the optimum u-field (i.e. Type-I estimation)**

[1] R.G. Raj, "A Hierarchical Bayesian-MAP Approach to Inverse Problems in Imaging," *Inverse Problems*, vol. 32, no. 7, July 2016.
[2] R.G. Raj, "Hierarchical Bayesian-MAP Methodology for Solving Inverse Problems," U.S. Patent 9,613,439, April 04, 2017.

U.S. NAVAL RESEARCH LABORATORY

Novel Hierarchical Bayesian-MAP Algorithm for Imaging

- **Our high-level pseudo-code is as follows:**
  1. **Initialize parameters α and μ**
  2. **Perform Type-II estimation to obtain $\hat{z}$, the MAP estimate of z**
  3. **Given $\hat{z}$ perform approximate EM algorithm to estimate the optimum parameter μ**
  4. **Iterate between Steps 2-3 until convergence**
  5. **Perform Type-I estimation to obtain $\hat{u}$, the optimum estimate of u**
  6. **The optimum estimate of the image is $\hat{I} = \Phi\hat{c}$, where $\hat{c} = \hat{z}.\hat{u}$**

- ***In our work we derive novel equations and algorithms for Type-I and Type-II estimation for solving (8)***
  - **For Type-II estimation we employ a Steepest Descent approach to calculate the z- field. We do this in a couple of ways:**
    1) **Newton method: Computationally intensive but Best results**
    2) **Gradient method: Sub-Optimal but computationally more efficient**
  - **For Type-I estimation: It turns out that sparsity of the *z*-field *must* be explicitly exploited**

Type-II Estimation

- Since we are modeling z as $\mathrm{h}(x)$ (where h is a nonlinearity and x follows a Gaussian Tree- structured distribution), it suffices to estimate the x field given the measurements y

- We can easily show that the optimum x can be determined as follows:

$$x^* = \underset{x}{argmin}\{y^T(M_x)^{-1}y + logdet(M_x) + x^T(P_x)^{-1}x\} \qquad \textbf{(9)}$$
$$= \underset{x}{argmin}\{f(x)\}$$

where, $M_x = A_x P_u (A_x)^T + \Sigma_v$

$$A_x = \widetilde{\Psi}\Lambda_x$$
$$\Lambda_x = diag(h(x))$$

Type-II Estimation (2)

- This leads to the following Steepest-Descent algorithm for calculating the optimum x- field:

1. Initialize x:

   First set $x^0 = h^{-1}(\widetilde{\Psi}^T y)$

   Thereafter we perform Kalman smoothing of x^0 w.r.t. the Gaussian Tree structured model with parameters A and B

   Initialize $n = 0$

2. Calculate the descent direction d^n either by Gradient descent or Newton Descent methods; where:

   $d^n = -\nabla f(x^n)$ for Gradient Descent

   $d^n = -[\nabla^2 f(x^n)]^{-1} \nabla f(x^n)$ for Newton Descent

3. Update the x-field:

   $$x^{n+1} = x^n + \lambda d^n$$

   where λ is chosen by a line search

4. Repeat Steps (1)-(2) until convergence

Type-II Estimation (3)

- Gradient Equation:

$$\nabla f(x) = 2(P_x)^{-1}x + G_x\left(vec\left((M_x)^{-1}\right) - [(M_x)^{-1}y \otimes (M_x)^{-1}y]\right) \quad \textbf{(10)}$$

where,

$$G_x = \frac{\partial vec(H_x)}{\partial x}[(P_u H_x) \otimes I_n + I_n \otimes (P_u H_x)]\left(\left(\widetilde{\Psi}\right)^T \otimes \left(\widetilde{\Psi}\right)^T\right)$$

- Hessian Equation:

$$\nabla^2 f(x) = 2(P_x)^{-1} + L(x) + Q(x) \quad \textbf{(11)}$$

where,

$$L(x) = \frac{\partial vec(G_x)}{\partial x}\left(vec\left((M_x)^{-1}\right) \otimes I_n\right) - G_x[(M_x)^{-1} \otimes (M_x)^{-1}](G_x)^T$$

Type-II Estimation (4)

- Hessian Equation (contd.):

$$Q(x) = -\frac{\partial vec(G_x)}{\partial x}\left((M_x)^{-1}y \otimes (M_x)^{-1}y \otimes I_n\right)$$
$$+G_x\left[\left(\left((M_x)^{-1}y\right)^T \otimes (M_x)^{-1}\right)\left(\left((M_x)^{-1}y\right)^T \otimes I_m\right.\right.$$
$$\left.\left.+\left(\left((M_x)^{-1}y\right)^T \otimes I_m\right)K_{mm}\right)\right](G_x)^T$$

$$\frac{\partial vec(G_x)}{\partial x} = E_x\left(\left(\widetilde{\Psi}\right)^T \otimes \left(\widetilde{\Psi}\right)^T \otimes I_n\right)$$

$$E_x = \begin{bmatrix} \nabla^2 H_x\{(P_u H_x) \otimes I_{n^2} + I_n \otimes (P_u H_x) \otimes I_n\} \\ + \\ \nabla H_x\left[I_n \otimes \left(P_u \widetilde{K}_{nn}\right) + (I_n \otimes (P_u)^T)\left(\widetilde{K}_{nn} \otimes I_n\right)\right]\left(I_{n^2} \otimes \nabla {H_x}^T\right) \end{bmatrix}$$

Type-II Estimation (5)

- Comments on Type-II estimation
  - In implementing the Gradient and Hessian equations one must pay **special attention to the sparsity structure of the matrices involved** so that the computations can be performed efficiently and with low storage
    - Brute force implementation is **infeasible**—storage for intermediate matrices on the order to 1 to 10 million TB for a 64x64 image
  - We employed the following non-linearity in our simulations

$$h(x) = \sqrt{exp}(x/\alpha)$$

where, α controls the sparsity-level in the generated signal: The smaller the α, the sparser the resulting signal

This choice of non-linearity gives us good control over the sparsity levels in a tractable manner

NOTE: Our approach in general is not limited by the specific choice of $h(x)$

Type-I Estimation

- Here we calculate the optimum u- field given the z- field estimated by the Type II procedure. We can easily show that the corresponding Bayesian MAP formulation approximately leads to the following equation to be solved for u:

$$\Lambda_{I_\tau(z)}\left(\left(\widetilde{\Psi}\right)^T\Lambda_1\widetilde{\Psi}+\Lambda_2\right)\Lambda_z u=\Lambda_{I(z)}\widetilde{\Psi}\Lambda_1 y \qquad \textbf{(12)}$$

where,

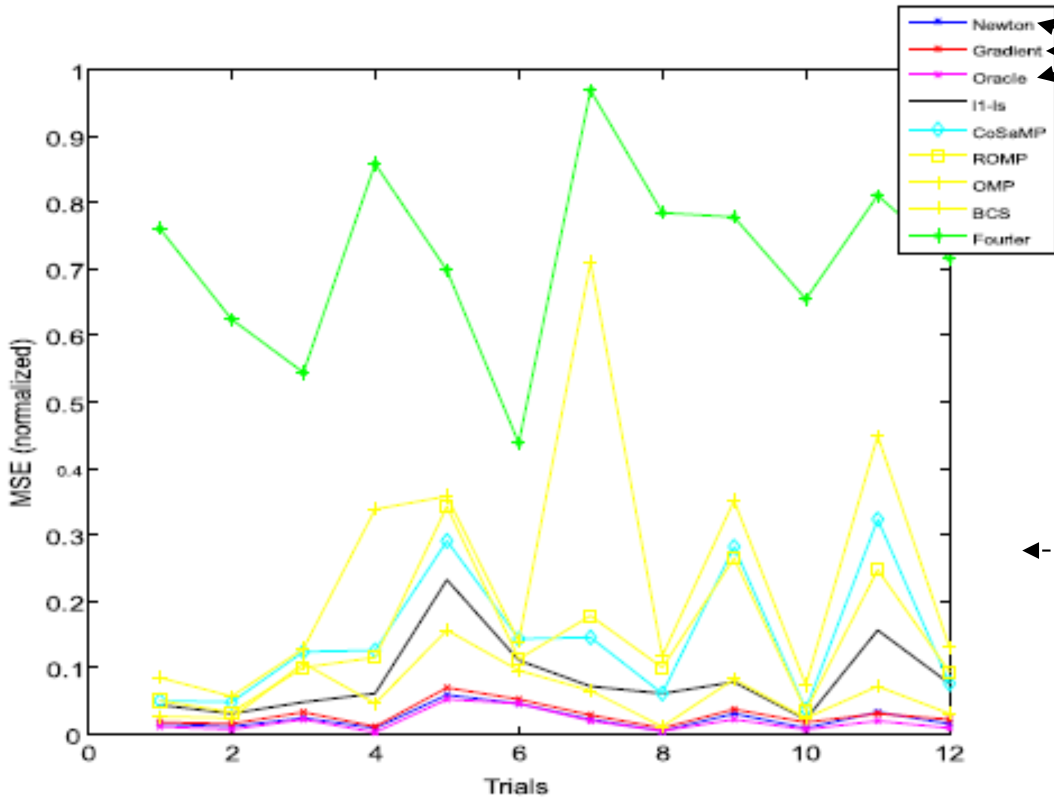
$$\Lambda_z = diag(z)$$

$$\Lambda_{I_\tau(z)} = diag(I_\tau(z)), \text{ where } I_\tau(z) = \begin{matrix} z & if\ z \geq \tau \\ 0 & else \end{matrix}$$

- The above is analogous to a weighted l_1- shrinkage procedure where the optimum weights are furnished by the z- field estimated by the Type II procedure.

HB-MAP w.r.t. Oracle: Imaging

Our Algorithm



Ψ: Sampling Operator

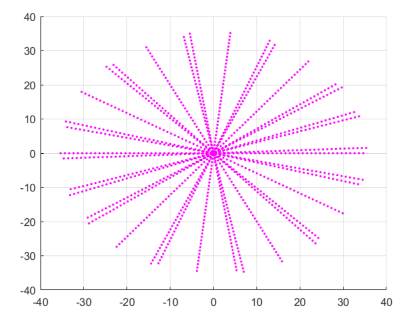
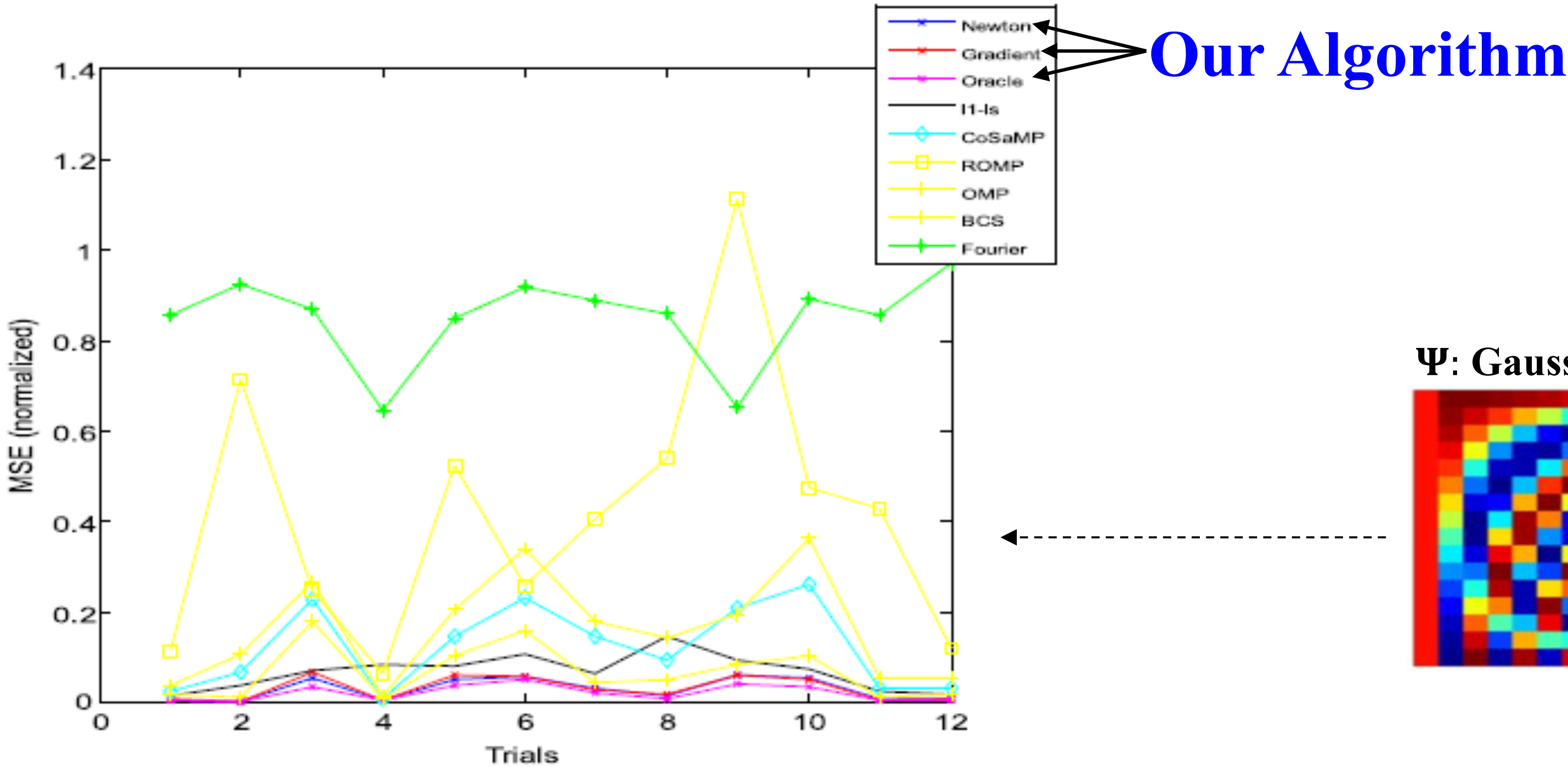


Figure. Performance for the case where 16×16 images are reconstructed based on measurements from a Radon transform at 15 different uniformly-spaced angles. The simulated images have sparsity level $\alpha = 0.2$. Higher levels of sparsity are more typical of natural images. (i) Our algorithm (Newton, gradient descent, and oracle). (ii) Sparsity based convex opt. algorithms (l1-ls, CoSaMP, ROMP and BCS). (iii) Standard Fourier-based reconstruction. Our HB-MAP algorithm consistently shows the best performance in terms of MSE across all trials. Newton HB-MAP outperforms gradient HB-MAP. The oracle HB-MAP plot is the result of the Type-I estimation for the case when we know the true z-field and thus gives the best performance as expected.

HB-MAP w.r.t. Oracle: Compressive Sensing



Ψ: Gaussian Noise

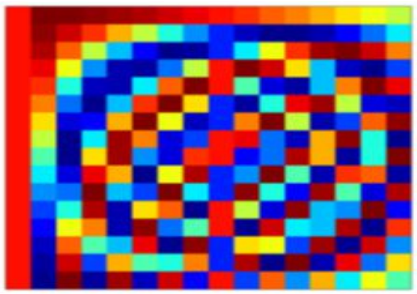


Figure. Performance for the case where 16×16 images are reconstructed based on measurements from a Random sensing matrix (with number of rows *equivalent to* a channel consisting of eight different uniformly-spaced angles). The simulated images have sparsity level $\alpha = 0.2$. Higher levels of sparsity are more typical of natural images. (i) Our algorithm (Newton, gradient descent, and oracle). (ii) Sparsity based convex opt. algorithms (l1-ls, CoSaMP, ROMP and BCS). (iii) Standard Fourier-based reconstruction. Our HB-MAP algorithm consistently shows the best performance in terms of MSE across all trials. Newton HB-MAP outperforms gradient HB-MAP. The oracle HB-MAP plot is the result of the Type-I estimation for the case when we know the true z-field and thus gives the best performance as expected.

Performance on Empirical Data: 15 Angles

Tomographic sampling operator: Radon transform at a Sparse number of angles

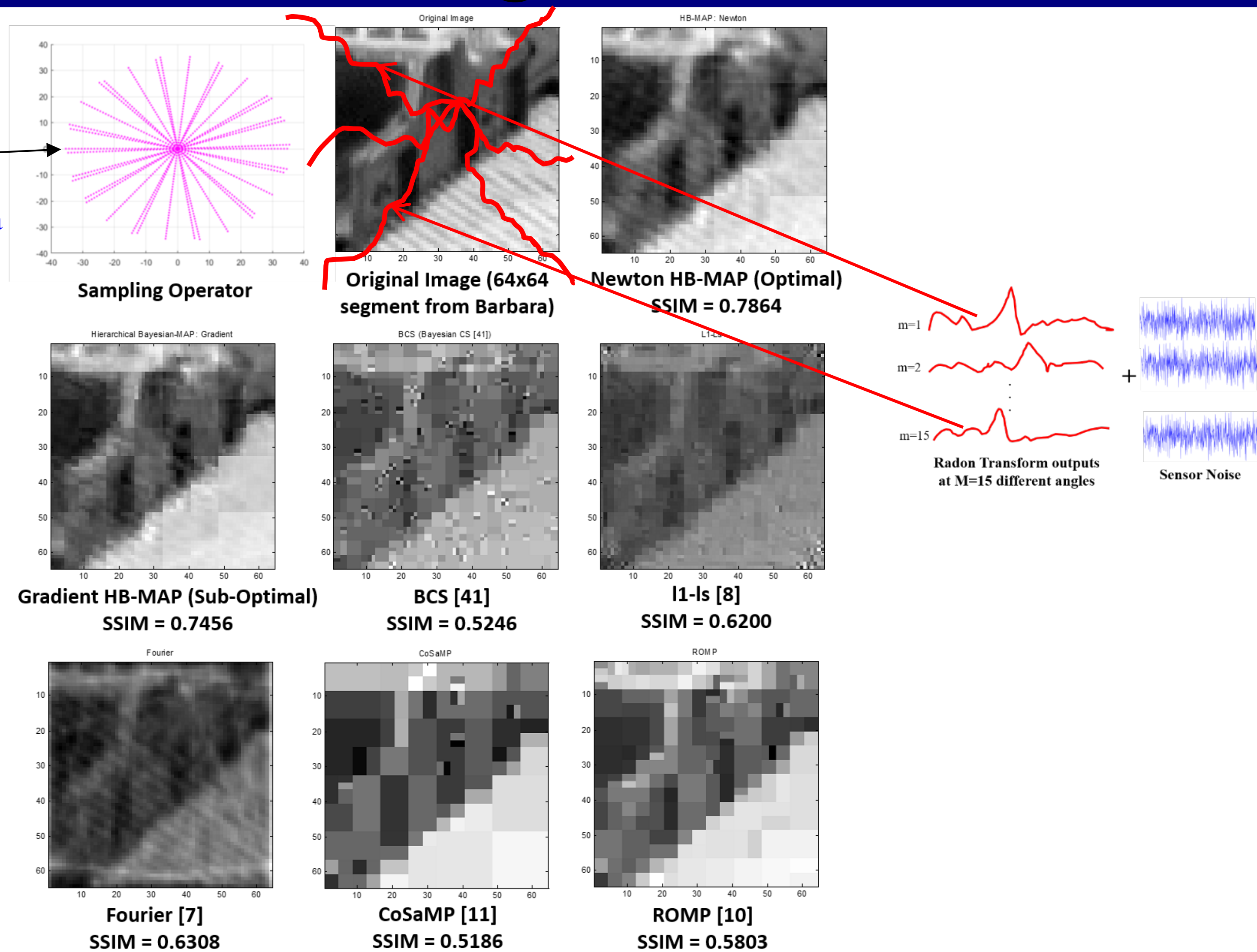
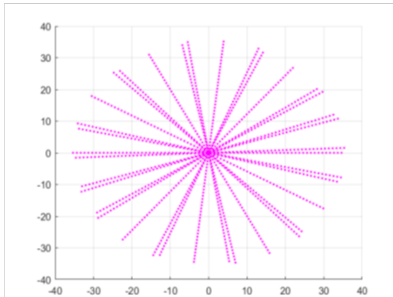
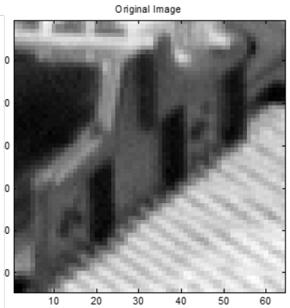


Figure. Here performance of our Newton- and Gradient- HB-MAP and other algorithms is shown for the case where the original image is a 64x64 image segment for which a Radon transform is performed at **15 uniformly spaced angles** plus 60dB of Measurement Noise at each Channel
The HB-MAP algorithm has the best Visual performance and in terms of SSIM.

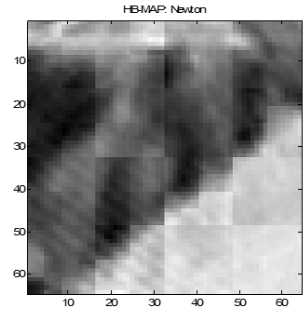
Performance on Empirical Data: 10 Angles



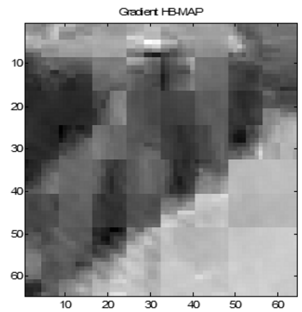
Sampling Operator



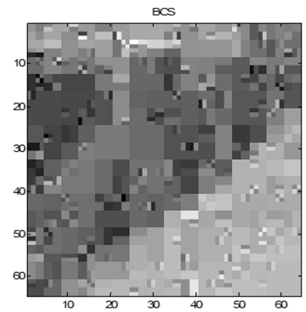
Original Image (64x64 segment from Barbara)



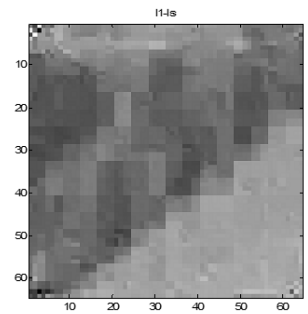
Newton HB-MAP (Optimal)
SSIM = 0.7031



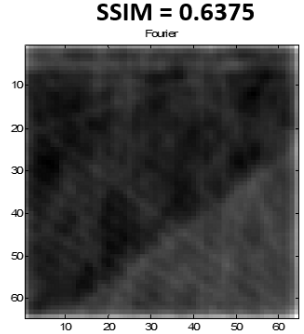
Gradient HB-MAP (Sub-Optimal)
SSIM = 0.6375



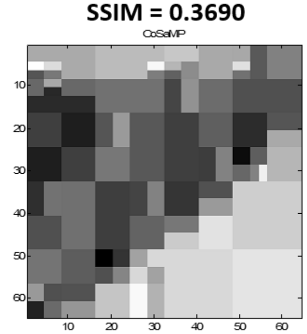
BCS [41]
SSIM = 0.3690



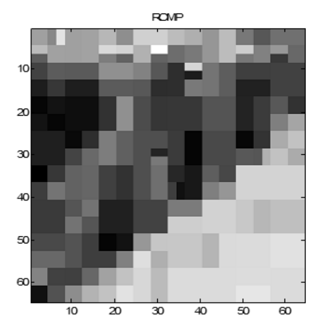
l1-ls [8]
SSIM = 0.5559



Fourier [7]
SSIM = 0.3284



CoSaMP [11]
SSIM = 0.4445

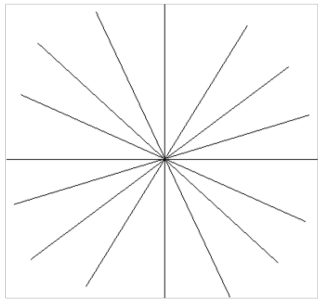


ROMP [10]
SSIM = 0.5104

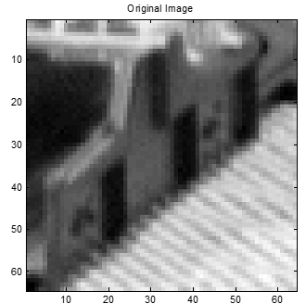
Figure. Here performance of our Newton- and Gradient- HB-MAP and other algorithms is shown for the case where the original image is a 64x64 image segment for which a Radon transform is performed at **10 uniformly spaced angles** plus 60dB of Measurement Noise at each Channel

The HB-MAP algorithm has the best Visual performance and in terms of SSIM.

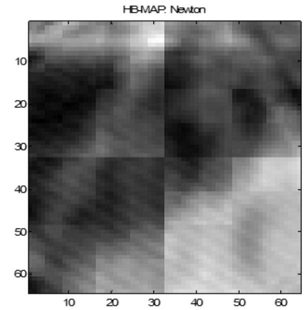
Performance on Empirical Data: 6 Angles



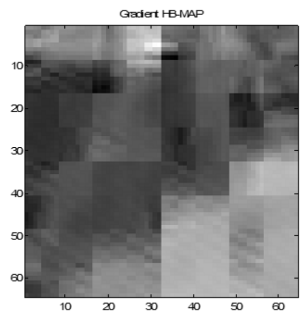
Sampling Operator



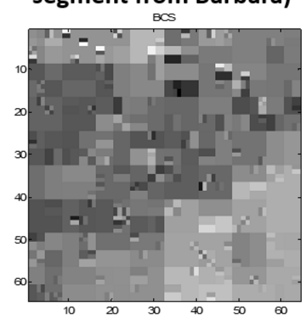
Original Image (64x64 segment from Barbara)



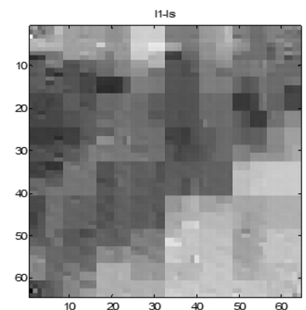
Newton HB-MAP (Optimal)
SSIM = 0.5637



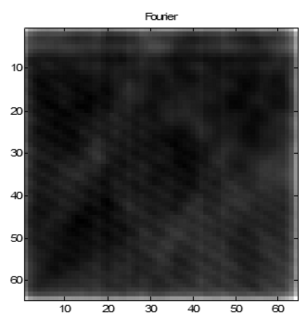
Gradient HB-MAP (Sub-Optimal)
SSIM = 0.4883



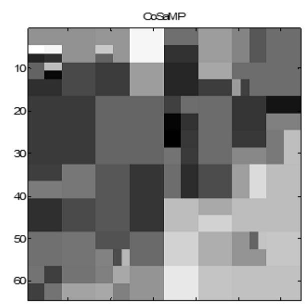
BCS [41]
SSIM = 0.2204



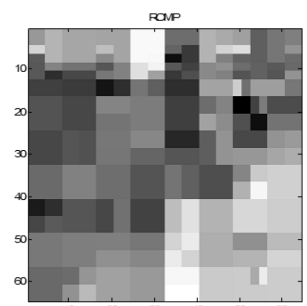
l1-ls [8]
SSIM = 0.4173



Fourier [7]
SSIM = 0.1992



CoSaMP [11]
SSIM = 0.3070



ROMP [10]
SSIM = 0.2661

Figure. Here performance of our Newton- and Gradient- HB-MAP and other algorithms is shown for the case where the original image is a 64x64 image segment for which a Radon transform is performed at **6 uniformly spaced angles** plus 60dB of Measurement Noise at each Channel
The HB-MAP algorithm has the best Visual performance and in terms of SSIM [31].

Outline

- Introduction and Motivation
- System Setup and Problem Formulation
- Sparsity Inducing Priors: Radar Clutter/Prior Models
  - CG, NCG, Graphical Models
- Hierarchical Bayesian Algorithms for Imaging
- **Fast Stochastic Algorithm for HB-MAP**
- Discussions

Fast Stochastic Algorithm for Imaging [1]

- Key problem: To alleviate the computational complexity of the Gradient HB-MAP (gHBMAP) and Newton HB-MAP (nHBMAP) algorithms
  - **Observation:** The Type-II objective is the computational bottleneck
    - Thus Type-II is the main focus of our work [1]

- Consider again the Type-II estimation objective function to be maximized:

$$f(x) = f_1(x) + f_2(x) + f_3(x) \tag{13}$$

where:

$$f_1(x) = y^T B(x)^{-1} y$$
$$f_2(x) = \log \det \mathrm{B(x)}$$
$$f_3(x) = x^T \Sigma_x^{-1} x$$

Such that: $B(x) = \sigma_u^2 X^T H^2(x) X + \sigma_\varepsilon^2 I$

$H(x) = diag(h(x))$; $X = \Psi\Phi$

[1] J. McKay, R.G. Raj, and V. Monga "Fast Stochastic Hierarchical Bayesian MAP for Tomographic Imaging," Accepted into the Proceedings of *IEEE Asilomar Conf. on Signals, Systems and Computers*, 2017.

Fast Stochastic Algorithm for SSRI (2)

- The key technical issues revolve around ameliorating the numerical tractability of f_1 and f_2

- We can obtain the following numerical simplification of f_2:

$$f_2(x) = \log \det \mathrm{B(x)}$$
$$\geq det(\sigma_u^2\, X^T H^2(x) X) + det(\sigma_\varepsilon^2 I) \quad \text{(Minkowski inequality)}$$
$$\geq \sum_{i=1}^{n} \frac{x_i}{\alpha} + K \quad \text{(where K is independent of } x\text{)}$$

$\Longrightarrow$ Thus we can maximize $\widetilde{f_2}(x) = \sum_{i=1}^{n} \frac{x_i}{\alpha}$ instead

- On the other hand, $f_1(x) = y^T B(x)^{-1} y$ reveals no readily actionable (convex) approximation
  - *Our approach is to avoid gradient calculations altogether by invoking a stochastic approximation (SA) approach*

Brief overview of SA

- The particular SA approach that we will be considering is Simultaneous Perturbation SA (SPSA):
  - The key idea in SPSA (and related approaches such as Finite Difference SA (FDSA)) is to start with finite difference approximations (for e.g. of the gradient) and to incorporate probabilistic theories to reduce computations [1-2]

- For SPSA the idea is to modify the update step of a gradient descent algorithm for gradient descent algorithm as follows:

$$x_{k+1} = x_k - \alpha_k \tilde{g}(x) \quad \textbf{(14)}$$

where $\tilde{g}(x)$ is a gradient approximation:

$$\tilde{g}(x) = \frac{1}{c_k}\left(f(x + c_k\Delta_k) - f(x - c_k\Delta_k)\right) \oslash \Delta_k$$

such that: Δ_k ~ symmetrix $\pm$ Bernoulli distribution

$\oslash$ ~ pointwise division operator

$c_k > 0$ is a small value that decreases to zero as $k \rightarrow \infty$

[1] J.C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," IEEE transactions on automatic control, vol. 37, no. 3, pp. 332–341, 1992.

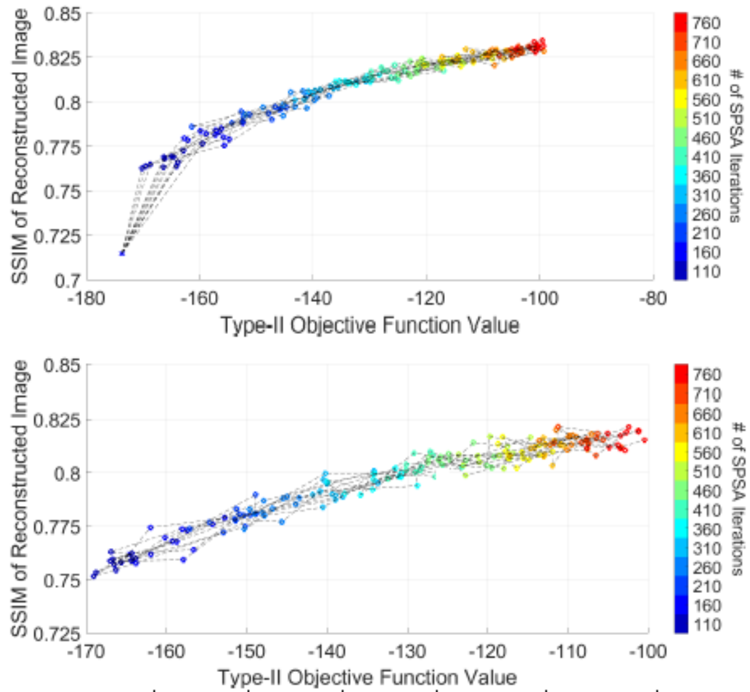
[2] J.C. Spall, "An overview of the simultaneous perturbation method for efficient optimization," Johns Hopkins APL technical digest, vol. 19, no. 4, pp. 482–492, 1998.

fsHBMAP Algorithm

- Our resulting algorithm, of incorporating SA into Type-II estimation, is called **Fast Stochastic HB-MAP (fsHBMAP)**:
  1. **Initialize parameters**
  2. **Perform SPSA based Type-II estimation to obtain $\hat{\boldsymbol{z}}$, the MAP estimate of $\boldsymbol{z}$**
  3. **Perform Type-I estimation to obtain $\widehat{\boldsymbol{u}}$, the optimum estimate of $\boldsymbol{u}$**
  4. **The optimum estimate of the image is $\hat{\boldsymbol{I}} = \boldsymbol{\Phi}\hat{\boldsymbol{c}}$, where $\hat{\boldsymbol{c}} = \hat{\boldsymbol{z}}.\widehat{\boldsymbol{u}}$**

- As before:
  - Hierarchical Bayesian structure is encapsulated within the above CG based framework
  - The CG prior in our framework reduces to Laplacian (i.e. l_1) and other priors (such as spike-and-slab, Generalized Gaussian etc.) with suitable choice of non-linearity in our model
  - Our framework is *not* limited by the choice of dictionary or sensing matrix

Illustration of Computational Speedup



Iteration	260	360	460	560	660	760
Time (s)	4.72	6.52	8.42	10.55	13.26	13.85

Figure. We took random initial values of x for a 16x16 image snippet of the Barbara image—sampled by the radon transform at 18 angles—and performed SPSA iterations in 10 trials.

- In every case incorporating SPSA iterations was not only able to decrease the Type-II objective function but also in a way that improves overall image quality (as measured by SSIM)
- Top Panel: Wavelet dictionary
  Bottom Panel: DCT dictionary
  Mean fsHBMAP completion times (seconds) are given for wavelets

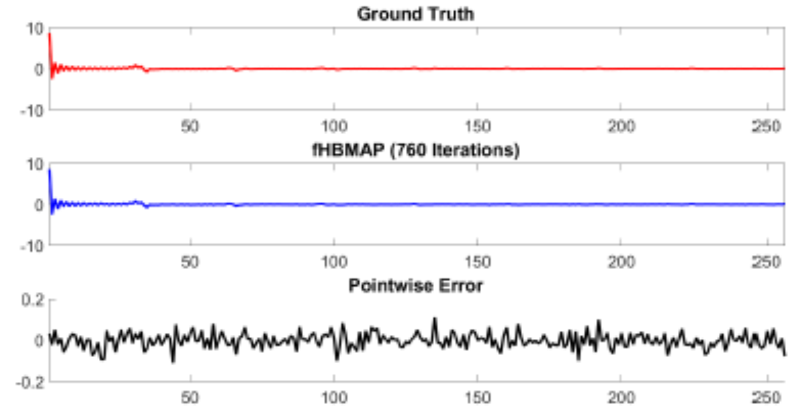
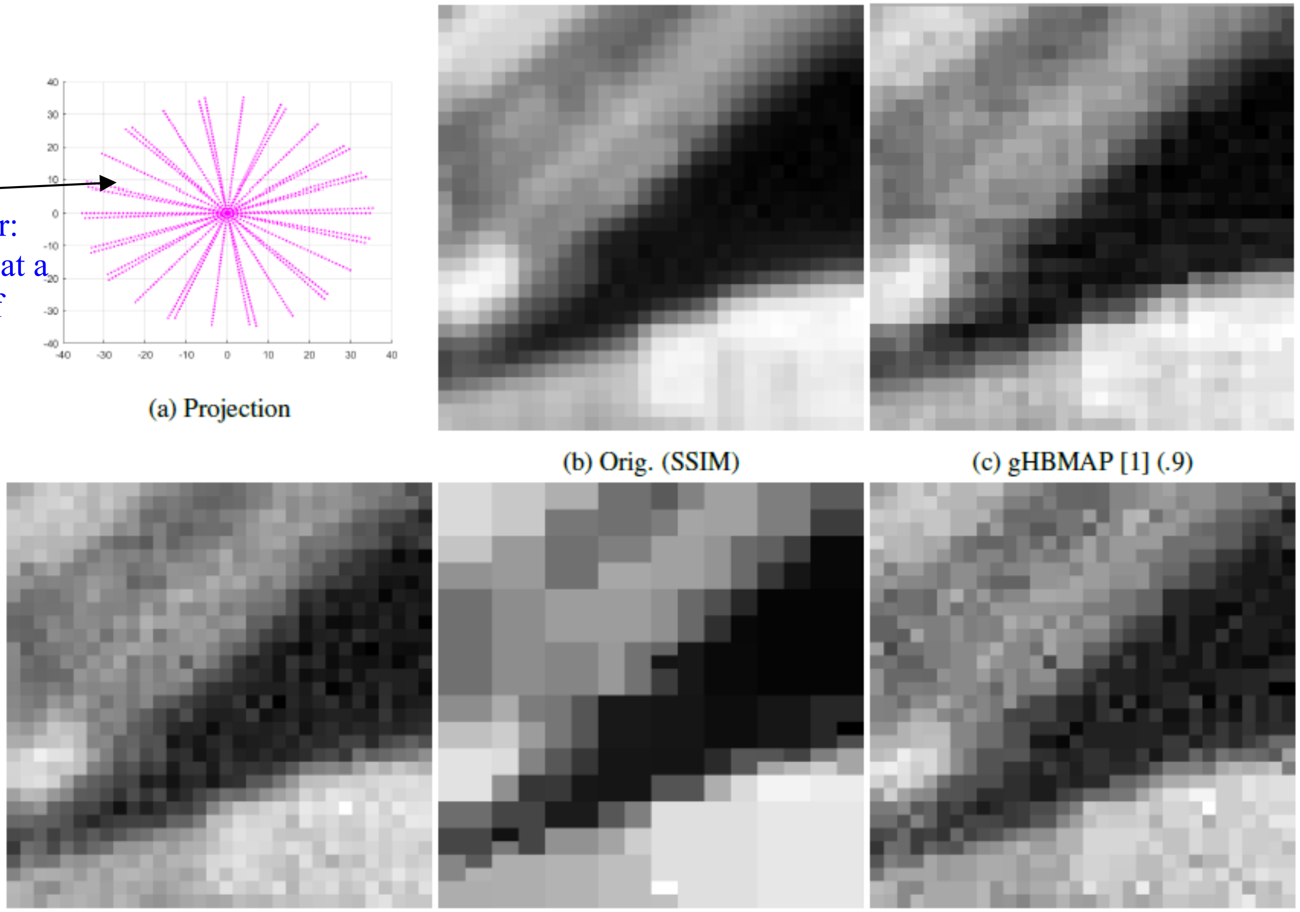


Figure. Comparison of wavelet coefficients of action (top) & fsHBMAP (middle) with point-wise error given (bottom).

fsHBMAP is much more amenable to being applied to large scale imaging problems because–unlike gHBMAP or nHBMAP–no explicit gradient or Kronecker calculations are involved.

fsHBMAP Performance

Tomographic sampling operator: Radon transform at a Sparse number of angles



(a) Projection

(b) Orig. (SSIM)

(c) gHBMAP [1] (.9)

(d) fsHBMAP(.78)

(e) CoSaMP [13] (.74)

(f) ICR [4] (.73)

	(c) gHBMAP	(d) fsHBMAP	(e) CoSaMP	(f) ICR
Time (seconds)	480	54	3	61

Traditional Backprojection based imaging performs poorly for a sparse aperture: SSIM ~ 0.6

Reconstructions of 32×32 Barbara photo snippet after projected, sampled in the pattern of 4a. Noise was $\sigma_\varepsilon^2 = .1$. Newton HB-MAP had an SSIM of .94 at a cost of 1.1×10^4 seconds (not shown for space).

Ongoing Investigations

- Detailed analysis of performance in Compressive Sensing applications
- Detailed analysis of different choices of non-linear models (i.e. prior distributions) matched to different application domains (such as Radar, Sonar)
- Extension to Deterministic Compressive Sensing matrices and Complex sensing matrices
  - Applications to Radar Imaging [1-2] and Sonar Imaging

[1] R.G. Raj, R.W. Jansen, M.A. Sletten, "A Sparsity based approach to Velocity SAR Imaging," *IEEE Radar Conference 2016*

[2] S. Samadi, M. Çetin, and M. A. Masnadi-Shirazi, “Sparse representation-based synthetic aperture radar imaging,” *IET Radar, Sonar Navigat.*, vol. 5, no. 2, pp. 182–193, Feb. 2011

Outline

- Introduction and Motivation
- System Setup and Problem Formulation
- Sparsity Inducing Priors: Radar Clutter/Prior Models
  - CG, NCG, Graphical Models
- Hierarchical Bayesian Algorithms for Imaging
- Fast Stochastic Algorithm for HB-MAP
- Discussions

Discussions

- **Thus far we have:**
  - We have shown how state-of-the-art imaging results can be achieved via systematically exploiting hierarchical Bayesian prior models
  - We have explored novel methods for Image formation
  - We have developed novel Bayesian models for (radar) signal processing applications

- **Some of the Future Work stemming from this includes:**
  - Systematically applying these computational tools in different application domains such as SAR, ISAR, Sonar imaging.
  - Rigorously investigating theoretical bounds for our Bayesian image formation algorithm (HB-MAP)
  - Application of novel waveform designs in conjunction with our optimum reconstruction algorithm
  - Extension to more general graphical structures and dictionaries

Thank You