Adaptive Compressive Imaging Using Sparse Hierarchical Learned Dictionaries

#### Jarvis Haupt

University of Minnesota Department of Electrical and Computer Engineering





- Motivation -

New Agile Sensing Platforms

# A Host of New Agile Imaging Sensors





#### CMOS Separable Transform Image Sensor (Georgia Tech)



# – Overview of This Talk –

Fusing Adaptive Sensing and Structured Sparsity

– Background –

Sparse Inference and Adaptive Sensing

#### A Model for Sparsity



Objects of interest are vectors  $x \in \mathbb{R}^n$ 

Signal Support:  $S \triangleq \{i : x_i \neq 0\}$ 



### A Sparse Inference Task



Noisy Linear Observation Model:

$$y = \Phi x + w \qquad \left\{ \begin{array}{l} \Phi \in \mathbb{R}^{m \times n} \\ w \sim \mathcal{N}(0, I_{m \times m}) \end{array} \right.$$

Support Recovery

Goal: Obtain an (accurate) estimate  $\widehat{S} = \widehat{S}(y, \Phi)$  of true support S

#### A Sparse Inference Task



Noisy Linear Observation Model:

$$y = \Phi x + w \qquad \left\{ \begin{array}{l} \Phi \in \mathbb{R}^{m \times n} \\ w \sim \mathcal{N}(0, I_{m \times m}) \end{array} \right.$$

Assume

- "Sensing energy"  $\|\Phi\|_F^2$  fixed:  $\|\Phi\|_F^2 = R$
- $|x_i| \ge \mu$  for all  $i \in S$

What conditions are necessary/sufficient for *exact* support recovery? (eg., such that  $P(S \neq \hat{S}) \rightarrow 0$  as  $n \rightarrow \infty$ )

### Exact Support Recovery?



"Point sampling" y = x + w(Sensing energy R = n)

Necessary & Sufficient for Exact Support Recovery:

$$\mu \geq \text{const.} \sqrt{\left(\frac{n}{R}\right)\log n}$$

"Uncompressed" Sensing (Donoho & Jin 2004; JH, Castro, & Nowak 2010)

"Compressed" Sensing (Genovese, Jin, & Wasserman 2009; Aeron, Saligrama & Zhao, 2010)

# **Conditions for Exact Support Recovery**



"Uncompressed" Sensing (Donoho & Jin 2004; JH, Castro, & Nowak 2010) "Compressed" Sensing (Genovese, Jin, & Wasserman 2009; Aeron, Saligrama & Zhao, 2010)

# Question: Can we do better by exploiting structure, or adaptivity, or both?

# **Conditions for Exact Support Recovery**



Necessity: (Castro 2012) Sufficiency (uncompressed): (Malloy & Nowak, 2010; Malloy & Nowak, 2011) Sufficiency (compressed): (JH, Baraniuk, Castro, & Nowak 2012, Malloy & Nowak 2013)

$$y = \Phi x + w \qquad \begin{cases} \Phi \in \mathbb{R}^{m \times n} \\ w \sim \mathcal{N}(0, I_{m \times m}) \end{cases} \quad \|\Phi\|_F^2 = R \end{cases}$$

# – Beyond Simple Sparsity –

The Role of Structure

#### **Our Focus: Tree Sparsity**



Characteristics of tree structure:

- Elements of x in one-to-one correspondence with nodes of  $\mathcal{T}$
- Nonzeros of tree-sparse vector form rooted connected subtree of  ${\mathcal T}$

Question: Does tree structure help in support recovery?

# **Conditions for Exact Support Recovery**



<u>Detection</u> of simple trail (uncompressed sensing)

Signal Detection Problem: (Arias-Castro, Candes, Helgason, & Zeitouni 2008)

# **Conditions for Exact Support Recovery**

The intersection of adaptivity and (tree) structure...

	Non-structured	Structured	
Non-adaptive	$\mu \ge \operatorname{const.} \sqrt{\left(\frac{n}{R}\right) \log n}$	$\mu \geq \operatorname{const.} \sqrt{\left(\frac{n}{R}\right)}$	Non-adaptive
Adaptive	$\mu \ge \operatorname{const.} \sqrt{\left(\frac{n}{R}\right)\log k}$	$\mu \ge \operatorname{const.} \sqrt{\left(\frac{k}{R}\right) \log k}$ (S)	Adaptive
	Non-structured	Structured	



Akshay Soni University of Minnesota

(A. Soni & JH, 2011) http://arxiv.org/pdf/1111.6923.pdf

Recent related work:

Adaptivity and Structure in finding activated blocks in a matrix (Balakrishnan, Kolar, Rinaldo, and Singh 2012)

#### Adaptive Tree Sensing: An Example



If the hypothesis test is correct at each step, then

$$m = dk + 1 = O(k)$$

Adaptive Wavelet "Tree Sensing" in the Literature:

- Non-Fourier encoded MRI (Panych & Jolesz, 1994)
- Compressive Imaging (Deutsch, Averbuch, & Dekel, 2009)

(none analyzed the case of *noisy* measurements...)

 $y_{(5)} = x_7 + w_{(5)}$ 

 $Q = \{\emptyset\}$ 

 $\hat{S} = \{1, 5\}$ 

# **Orthogonal Dictionaries and Tree Sparsity**

Consider signals  $z \in \mathbb{R}^p$  that are sparse in a known dictionary  $D \in \mathbb{R}^{p \times n}$ . That is, z = Dx, where

- $x \in \mathbb{R}^n$  is k-sparse,
- D satisfies  $D^T D = I_{n \times n}$ , and
- columns of D are  $d_j, j = 1, 2, \ldots, n$

We are interested in the case where x is *tree-sparse*...

Collect (noisy) observations of z by projecting onto (scaled) columns of D. Suppose, for example, that the j-th measurement is obtained by projecting onto columnn  $d_i$ , then

$$y_{(j)} = \beta d_i^T z + w_{(j)}$$

where  $w_{(j)} \sim \mathcal{N}(0, 1)$ .

Nonnegative scaling factor (equivalently, could consider non-unit noise variance)

## Support Recovery via Adaptive Tree Sensing

#### Theorem (A. Soni & JH, 2011)

Let  $\mathcal{T}_{n,d}$  be a balanced, rooted connected tree of degree d with n nodes. Suppose that  $z \in \mathbb{R}^p$  can be expressed as z = Dx, where D is a known dictionary with orthonormal columns and x is k-sparse. Further, suppose the support of x corresponds to a rooted connected subtree of  $\mathcal{T}_{n,d}$ . Observations of z are of the form of projections of z onto columns of D.

Let the index corresponding to the root of  $\mathcal{T}_{n,d}$  be known, and apply the top-down tree sensing procedure with threshold  $\tau$  and scaling parameter  $\beta$ . For any  $c_1 > 0$  and  $c_2 \in (0, 1)$ , there exists a constant  $c_3 > 0$  such that if

$$u = \min_{i \in \mathcal{S}} |x_i| \ge \sqrt{c_3 \beta^{-2} \log k}$$

and  $\tau = c_2 \mu \beta$ , the tree sensing procedure collects m = dk+1 measurements, and produces a support estimate  $\widehat{S}$  that equals S with probability at least  $1 - k^{-c_1}$ .

Choose  $\beta = \sqrt{\frac{R}{(d+1)k}}$ , then the theorem guarantees exact support recovery (whp) when

$$\mu \ge \sqrt{c_3(d+1)\left(\frac{k}{R}\right)\log k}$$

# **Conditions for Exact Support Recovery**

The intersection of adaptivity and (tree) structure...

	Non-structured	Structured	
Non-adaptive	$\mu \geq \operatorname{const.} \sqrt{\left(\frac{n}{R}\right) \log n}$	$\mu \geq  ext{const.} \sqrt{\left(rac{n}{R} ight)}$ (*conjecture for support recovery)	Non-adaptive
Adaptive	$\mu \geq \operatorname{const.} \sqrt{\left(\frac{n}{R}\right)\log k}$	$\mu \ge \operatorname{const.} \sqrt{\left(\frac{k}{R}\right) \log k}$ (S)	Adaptive
	Non-structured	Structured	



(A. Soni & JH, 2011) http://arxiv.org/pdf/1111.6923.pdf

Akshay Soni University of Minnesota

– LASeR –

Learning Adaptive Sensing Representations

#### Beyond Wavelet Trees: Learned Representations

Given training data  $Z \in \mathbb{R}^{p \times q}$ , want to *learn* a dictionary D so that

 $Z \approx DX, \ D \in \mathbb{R}^{p \times n}, \ X \in \mathbb{R}^{n \times q},$ 

and each column of  $X, x_i \in \mathbb{R}^n$ , is tree-sparse in  $\mathcal{T}_{n,d}$ .

Pose this as an optimization:

$$\{D, X\} = \arg \min_{D \in \mathbb{R}^{p \times n}, D^T D = I_{n \times n}, \{x_i\}} \sum_{i=1}^{q} \|z_i - Dx_i\|_2^2 + \lambda \Omega(x_i)$$

The regularization term is  $\Omega(x_i) = \sum_{g \in \mathcal{G}} \omega_g ||(x_i)_g||$ , where

- $\mathcal{G}$  denotes a set of (overlapping) groups of indices for x,
- $(x_i)_g$  is  $x_i$  restricted to the indices in the group  $g \in \mathcal{G}$ ,
- $\omega_g$  are non-negative weights, and
- the norm can be, eg.,  $\ell_2$  or  $\ell_\infty$

Solve by *alternating minimization* over *D* and *X* (Jenatton, Mairal, Obozinski, & Bach, 2010) Sparse Modeling Software (SPAMS): http://spams-devel.gforge.inria.fr/

# Hierarchical Overlapping Groups

Example: Binary Tree, 15 nodes, 4 levels...



Number of groups same as number of nodes (but varying sizes)

– LASeR –

An Illustrative Example

# Learning Adaptive Sensing Representations



Learn representation for 163 images from Psychological Image Collection at Stirling (PICS) http://pics.psych.stir.ac.uk/

#### Example images $(128 \times 128)$



#### Learned Orthogonal Tree-Basis Elements



(First four levels of 7 total)



### **Qualitative Results**



#### **Qualitative Results**



#### **Quantitative Results**



——: LASeR

- $\Box$  : PCA
  - o: CS Lasso
  - $\diamondsuit$ : CS Tree Lasso
  - o: Wavelet Sensing



original image

$$SNR = 10\log_{10}\left(\frac{\|\mathbf{x}\|_2^2}{\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2}\right)$$

 $\sigma = 1$ 

– LASeR –

#### Imaging via "Patch-wise" Sensing

# "Patch-wise" Sensing Experiment

Motivated by EO Imaging Application (Thanks: Bob Muise @ Lockheed Martin)

Training Data:

3 Sample images from the Columbus Large Image Format (CLIF) 2007 Dataset Each image is 1024x1024



Randomly extracted 3000 32x32 patches (at random locations)... and vectorized them into length 1024 vectors

Applied PCA and LASeR (7-level 127 node binary tree) to this training data

#### **Compare: PCA Basis Elements**



In the tree-sensing context, can view PCA sensing approach in terms of a tree of degree 1

# Learned Orthogonal Tree-Basis Elements



# Example: Approximation by "Patch-wise" Sensing

Test Image (another image from CLIF database)



Sense & reconstruct non-overlapping 32x32 patches... ...comparing LASeR, PCA, Wavelets...

Sampling rate: 12.5%



LASeRrSNR = 16.5 dB

 $\begin{array}{c} \mathsf{PCA} \\ \mathsf{rSNR} = 17.6 \ \mathrm{dB} \end{array}$ 

Sampling rate: 12.5%



LASeRrSNR = 16.5 dB

Sampling rate: 12.5%



 $\begin{array}{c} \mathsf{PCA} \\ \mathsf{rSNR} = 17.6 \ \mathrm{dB} \end{array}$ 

Sampling rate: 12.5%



LASeRrSNR = 16.5 dB

2D Haar Wavelet rSNR = 13.5 dB

#### Approximation Results – Adaptive Sampling Rate

Average sampling rate: 7.2%



LASeR rSNR = 13.9 dB  $\begin{array}{c} \mathsf{PCA} \\ \mathsf{rSNR} = 15.0 \ \mathrm{dB} \end{array}$ 

### Approximation Results – Adaptive Sampling Rate

Average sampling rate: 7.2%



LASeR rSNR = 13.9 dB 2D Haar Wavelet rSNR = 11.9 dB

### Approximation: Zoomed In

A Closer Look... (Average sampling rate: 7.2%)



Original

LASeR

PCA

2D Haar

#### Approximation: Zoomed In

A Closer Look... (Average sampling rate: 7.2%)



Original

LASeR

PCA

2D Haar

# Sampling Rate Adapts to Block "Complexity"

Sampling rate per block (Average sampling rate: 7.2%)



# Sampling Rate Adapts to Block "Complexity"

Sampling rate per block (Average sampling rate: 7.2%)



# Summary: Adaptivity + Structure

#### In Theory (Support Recovery)



Conclusions:

Polynomial reduction in SNR required for exact support recovery (for fixed "sensing energy")

# Summary: Adaptivity + Structure

In Practice (Learned Representations and Patch-wise Sensing)



Original

LASeR

PCA

Conclusions:

PCA works very well on "small" patch sizes (shared, elemental structure)!

### Summary: Adaptivity + Structure



#### Conclusions:

Potential benefit for learned representations depend on patch size, data "regularity", noise Can we sense with non-orthogonal trees? (Relations to CS experimental design?)

Thank You! www.ece.umn.edu/~jdhaupt jdhaupt@umn.edu