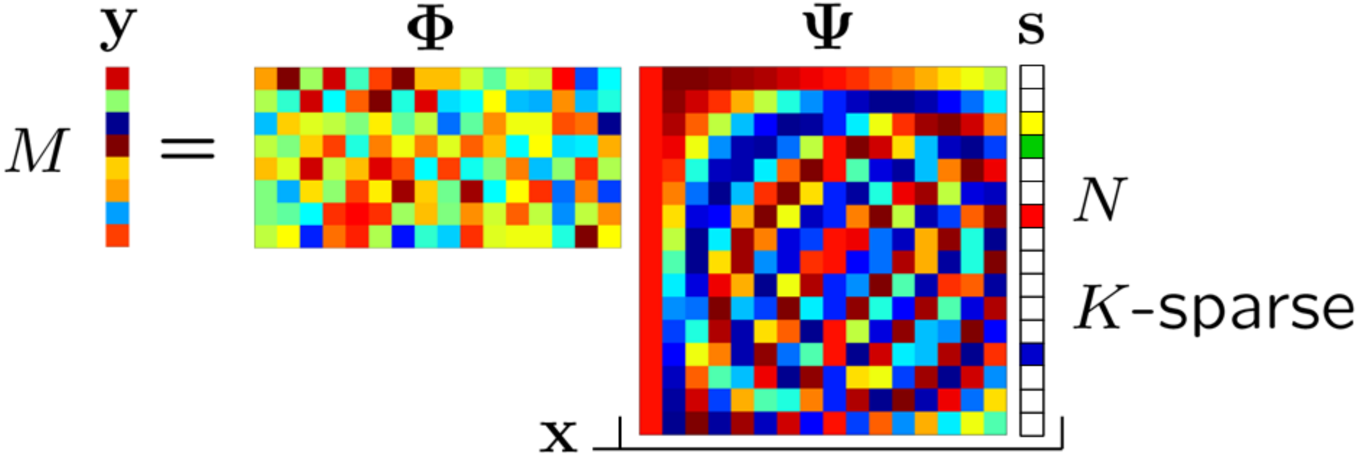
Convex approaches for group sparse signal recovery in compressed sensing

-Nikhil Rao

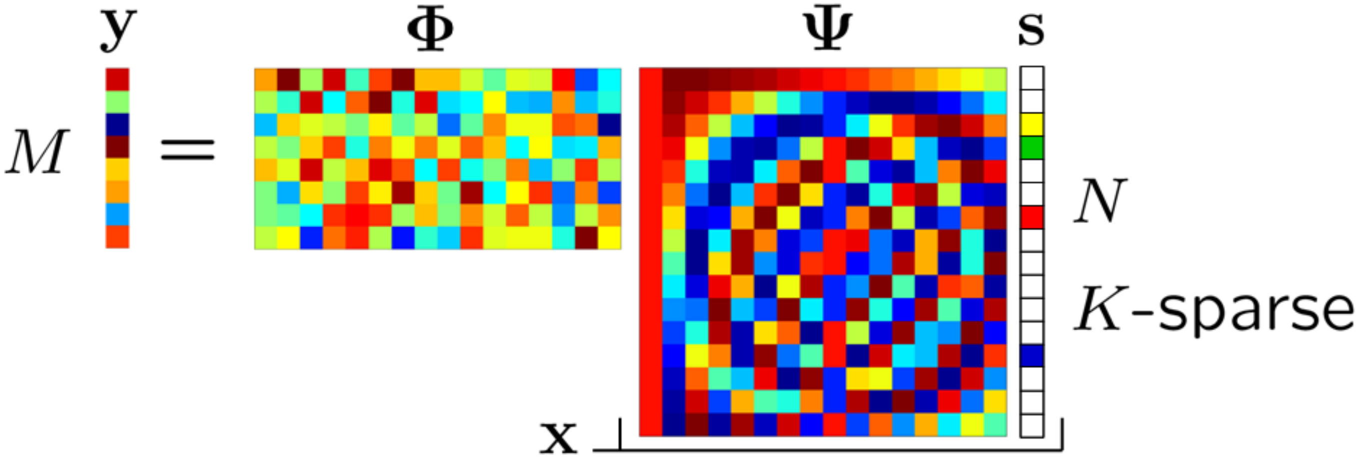
University of Wisconsin-Madison

(with Rob Nowak, Steve Wright, Nick Kingsbury, Ben Recht)

Compressed Sensing



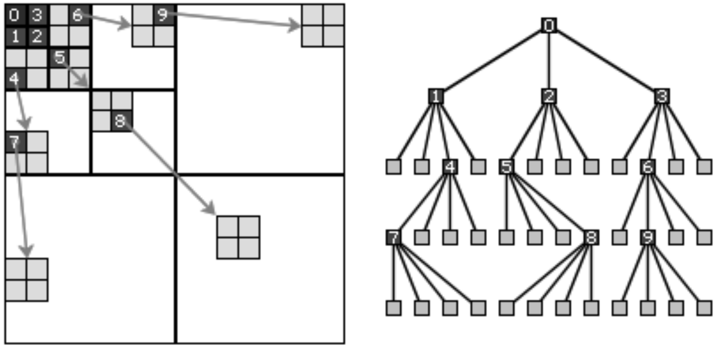
Compressed Sensing

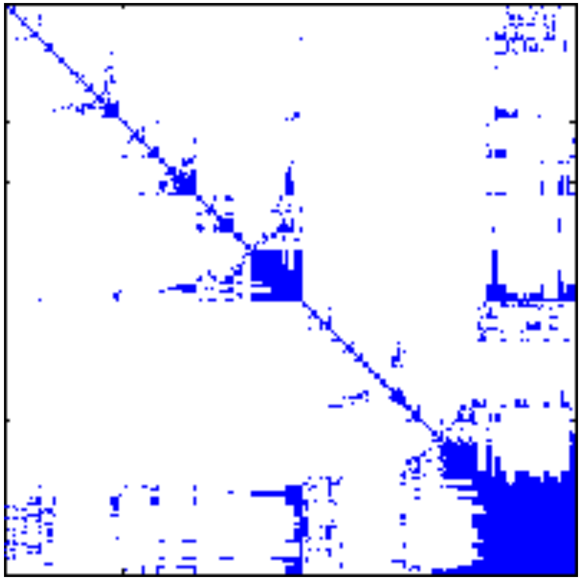


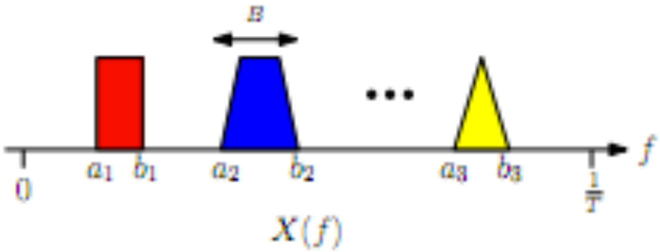
- GOAL : given y, recover s (equivalently x)
- Greedy (OMP) and convex (LASSO) methods to recover s
- M ~ O(K log(N))

Structured Sparsity

- In several cases, signals exhibit structure within sparsity







The variables can be seen to lie in a union of subspaces/groups

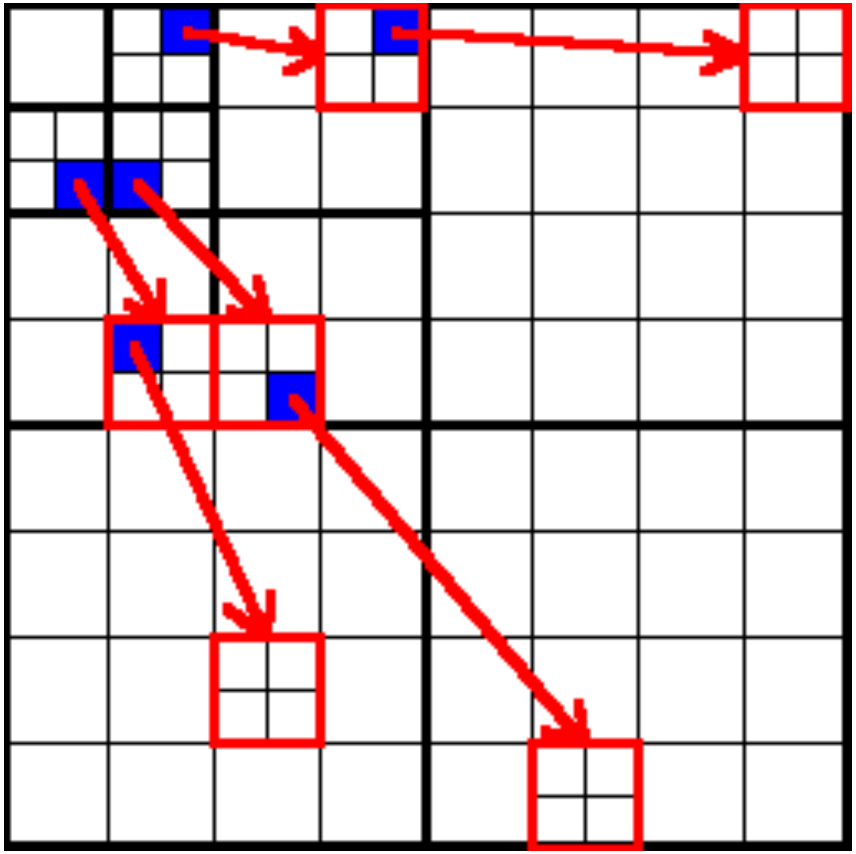
Structured Sparsity

- We can leverage this additional known structure to design better recovery algorithms

- Cases addressed before…
  - Disjoint groups [Yuan and Lin ‘06, Zhang et. al. ‘09]
  - Tree structures [Jenatton et. al. ‘09, Baraniuk et. al. ‘10, La and Do 06, Zhang et. al. ‘09]
  - Arbitrary overlapping groups [Baraniuk et. al. ’10]

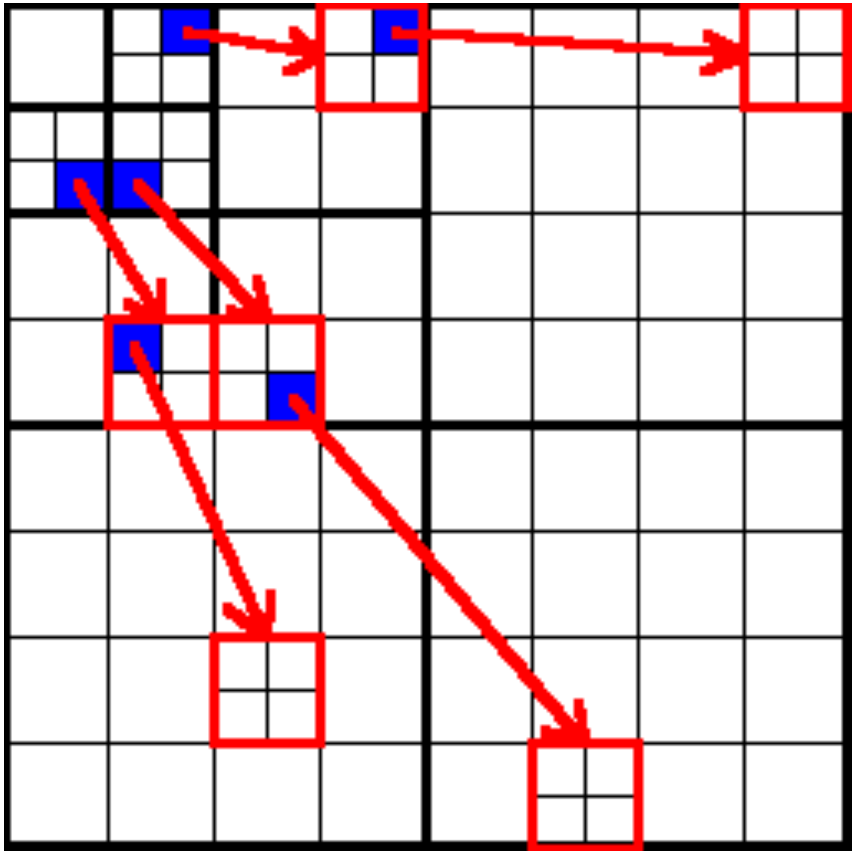
We look to apply these (and other) results to denoise images in a Compressed Sensing framework

- DWT coefficients can be modeled along a tree – in which parent-child pairs exhibit similar properties



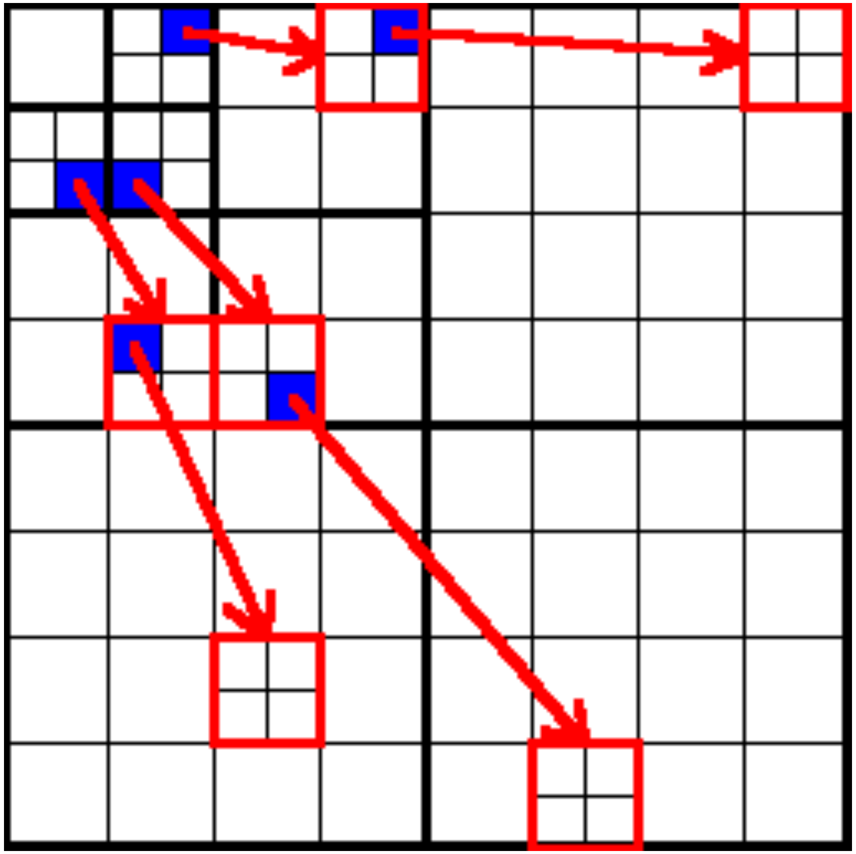
- DWT coefficients can be modeled along a tree – in which parent-child pairs exhibit similar properties

- Better methods (compared to lasso) exist:
    - TOMP (La and Do '06)
    - Message passing (Schniter '10)
    - HMT based reweighting (Duarte et. al. '08)



- DWT coefficients can be modeled along a tree – in which parent-child pairs exhibit similar properties

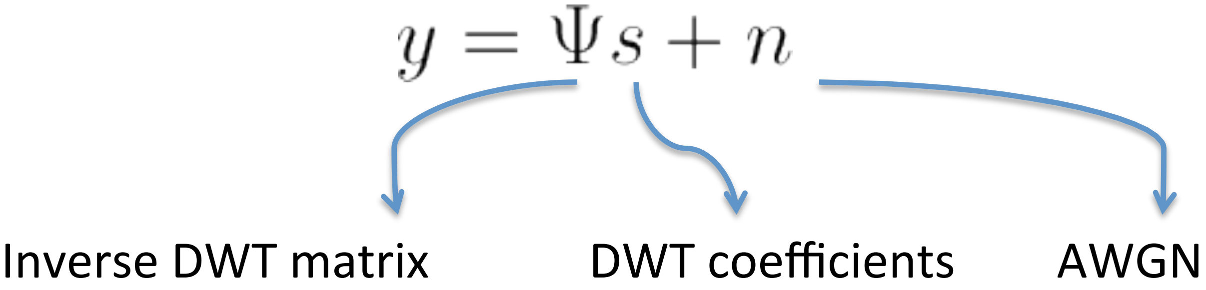
- Better methods (compared to lasso) exist:
  - TOMP (La and Do '06)
  - Message passing (Schniter '10)
  - HMT based reweighting (Duarte et. al. '08)



These methods take into account the inherent tree structure of the coefficients

But…

- Standard denoising:



$$y = \Psi s + n$$

Inverse DWT matrix

DWT coefficients

AWGN

But…

- Standard denoising:

$$y = \underbrace{\Psi}_{\text{Inverse DWT matrix}} \underbrace{s}_{\text{DWT coefficients}} + \underbrace{n}_{\text{AWGN}}$$

- In a compressed sensing framework:

$$y = \underbrace{\Phi}_{\substack{\text{iid Gaussian (or any other)} \\ \text{measurement matrix}}} \Psi s + n$$

Wavelet Based Image Restoration

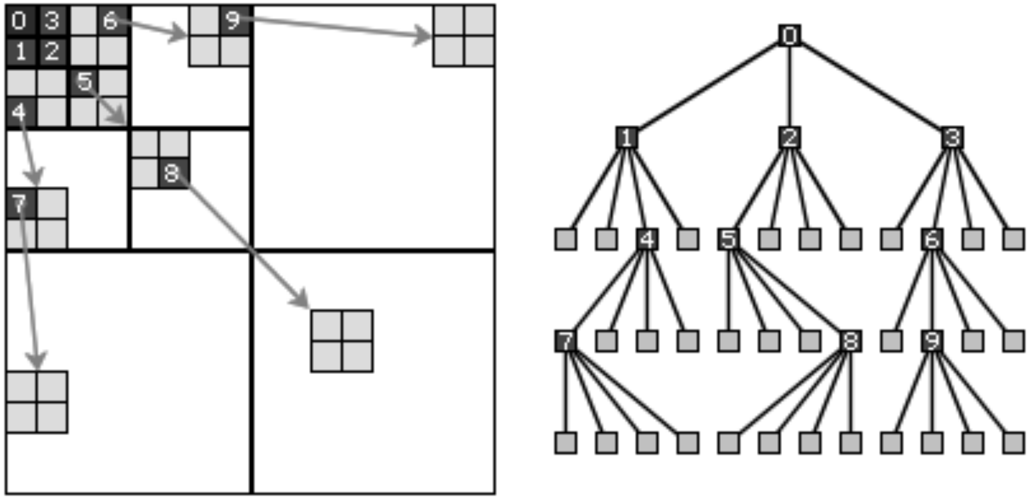
- The measurement matrix linearly (and randomly) mixes the DWT coefficients, and hence we can no longer use existing approaches

$$y_i = \sum_j \Phi_{ij} [\Psi s]_j$$

- GOAL : given compressive measurements, can we recover the sparse coefficients, while taking advantage of the inherent tree structure and maintain convexity?

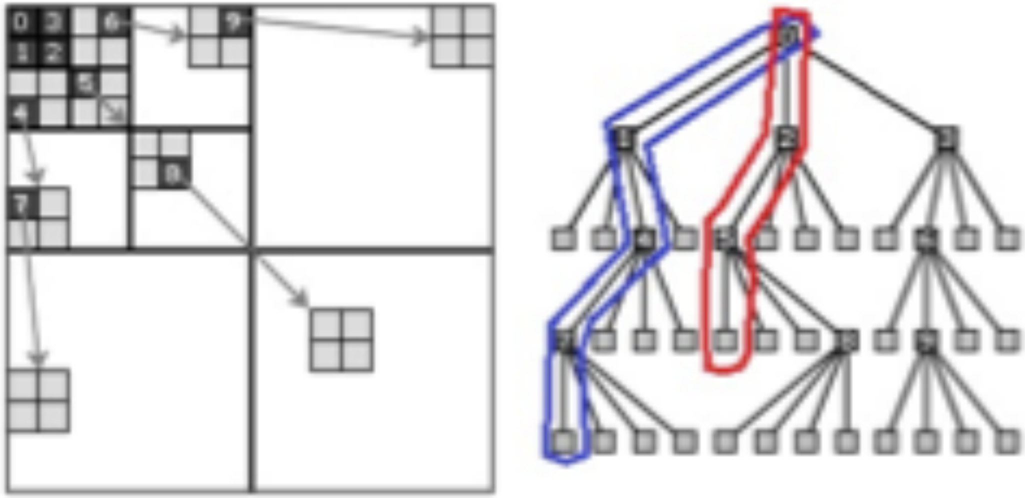
Wavelet Based Image Restoration

- We consider the groups to be paths along the tree from the root to leaf



Wavelet Based Image Restoration

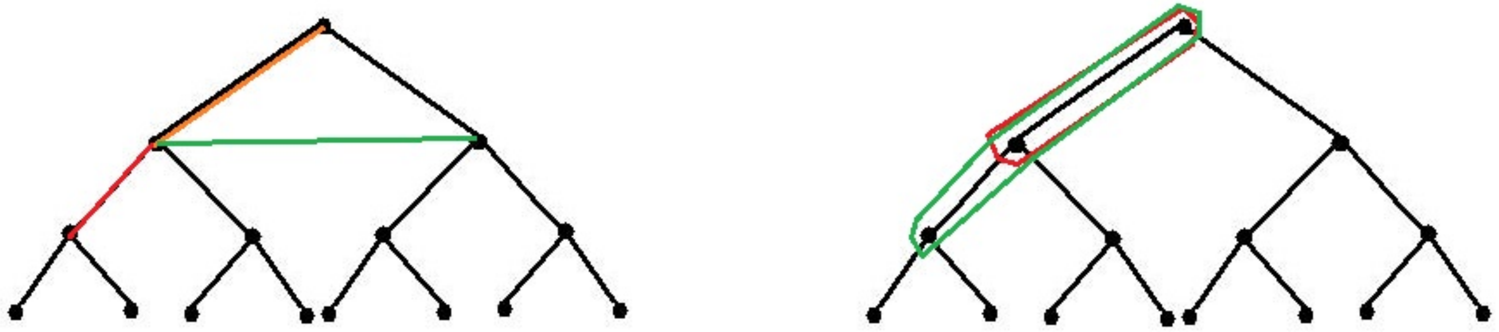
- We consider the groups to be paths along the tree from the root to leaf



- The groups overlap, and hence, to recover a union of groups, we use the method by Jacob et. al.
- Resulting coefficient structure same as Jenatton et. al 09

Other Grouping strategies

- Parent-child pairs, in scale dependencies



- Hierarchical groups along paths of the tree
- Better groupings ➔ better reconstruction

Group Lasso with Overlap (Jacob et. al. 09)

Given groups G = {g_1, g_2, g_3, … g_k}

A = ΦΨ

$$\hat{x} = arg\min_{x} \frac{1}{2}||y - Ax||^2 + \lambda\Omega(x)$$

$$\Omega(x) = \inf_{\sum v_g = x} \sum_{g \in G} ||v_g||$$

v_g = vector with support in g

The norm will look to recover a signal whose support is a union of groups

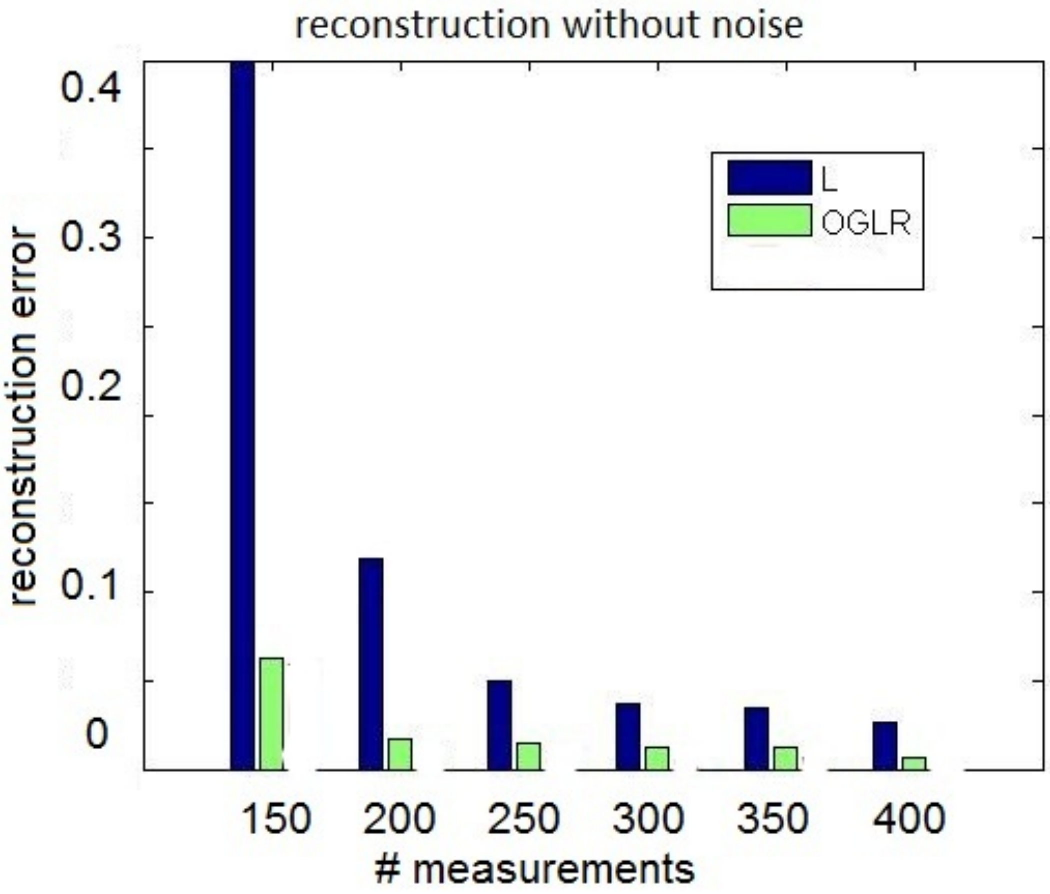
Denoising

- Cameraman image corrupted by WGN of variance 0.3
- Image resized to 64X64 , vectorized (length 4096)
- # of random samples collected = 800

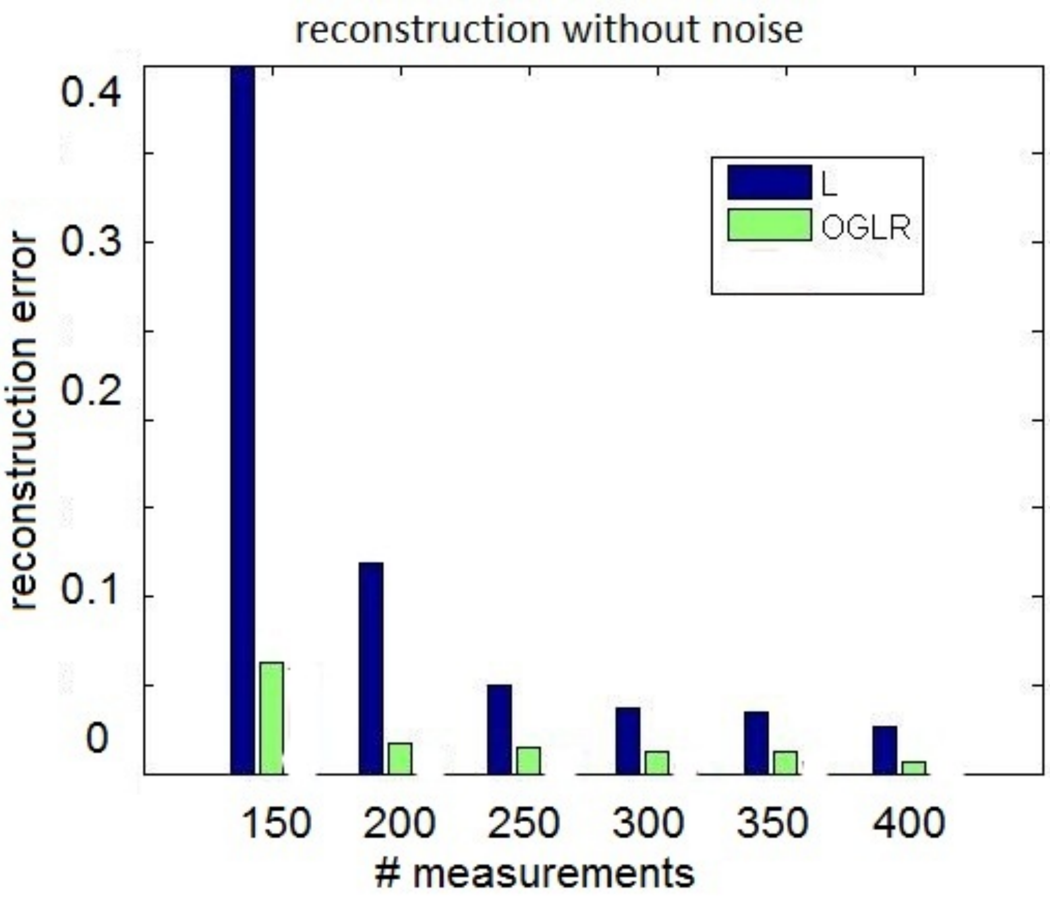
PSNR = 22.75

PSNR = 25.08

gaussian measurements



gaussian measurements



Can we find the number of measurements needed to exactly recover a signal using a convex program?

Exact reconstruction of structured sparse signals (with Ben Recht)

Consider a k- group sparse signal x, made up of M (arbitrary) groups, with max. group size = B

Theorem (NR, Recht, Nowak): The number of *iid* Gaussian measurements needed to exactly recover x is lower bounded by:
#m > 2k(log(M-k) + B) + k *

*Preprint : http://arxiv.org/abs/1106.4355

Exact reconstruction of structured sparse signals (with Ben Recht)

Consider a k- group sparse signal x, made up of M (arbitrary) groups, with max. group size = B

Theorem (NR, Recht, Nowak): The number of *iid* Gaussian measurements needed to exactly recover x is lower bounded by:
#m > 2k(log(M-k) + B) + k *

#m is independent of the nature of groups

Arbitrary groupings entail tougher recovery algorithms

*Preprint : http://arxiv.org/abs/1106.4355

Exact Reconstruction of Structured Sparse Signals

Atomic norm

$$||x||_{\mathcal{A}} = \inf \left\{ \sum_{a \in \mathcal{A}} c_a : x = \sum_{a \in \mathcal{A}} c_a a \quad c_a \geq 0 \quad \forall a \in \mathcal{A} \right\}$$

Exact Reconstruction of Structured Sparse Signals

Atomic norm

$$||x||_{\mathcal{A}} = \inf \left\{ \sum_{a \in \mathcal{A}} c_a : x = \sum_{a \in \mathcal{A}} c_a a \quad c_a \geq 0 \quad \forall a \in \mathcal{A} \right\}$$

$$\forall G \in \mathcal{G}, \text{ let } A_G = \{a^G \in \mathbb{R}^p : ||(a^G)_G||_2 = 1, (a^G)_{G^c} = 0\}$$

$$\mathcal{A} = \{A_G\}_{G \in \mathcal{G}}$$

$$||x||_{\mathcal{A}} = \Omega^{\mathcal{G}}_{overlap}(x)$$

Overlap Group lasso norm (Jacob et al 09)

Exact Reconstruction of Structured Sparse Signals

Atomic norm $$||x||_{\mathcal{A}} = \inf \left\{ \sum_{a \in \mathcal{A}} c_a : x = \sum_{a \in \mathcal{A}} c_a a \;\; c_a \geq 0 \;\; \forall a \in \mathcal{A} \right\}$$

$$\forall G \in \mathcal{G}, \text{ let } A_G = \{a^G \in \mathbb{R}^p : ||(a^G)_G||_2 = 1, (a^G)_{G^c} = 0\}$$

$$\mathcal{A} = \{A_G\}_{G \in \mathcal{G}}$$

$$||x||_{\mathcal{A}} = \Omega^{\mathcal{G}}_{overlap}(x)$$ Overlap Group lasso norm (Jacob et al 09)

The group lasso with overlap is equivalent to the minimization of an atomic norm

Proof Outline

The GAUSSIAN WIDTH of $S \subset \mathcal{S}^{p-1}$ $w(S) = E[\sup_{z \in S} g^t z]$

Proof Outline

The GAUSSIAN WIDTH of $S \subset \mathcal{S}^{p-1}$ $w(S) = E[\sup_{z \in S} g^t z]$

From an extension of Gordon's minimum restricted singular value theorem,

Let T(x) be the spherical part of the tangent cone at x w.r.t the atomic norm. If A is a random normalized gaussian map, then x is the unique optimum of the atomic norm minimization provided $M \geq w(T) + \mathcal{O}(1)$

Proof Outline

The GAUSSIAN WIDTH of $S \subset \mathcal{S}^{p-1}$ $\quad w(S) = E[\sup_{z \in S} g^t z]$

From an extension of Gordon's minimum restricted singular value theorem,

Let T(x) be the spherical part of the tangent cone at x w.r.t the atomic norm. If A is a random normalized gaussian map, then x is the unique optimum of the atomic norm minimization provided $M \geq w(T) + \mathcal{O}(1)$

For any cone C, from the properties of tangent cones and Jensen's inequality

$$w(\underbrace{C \cap \mathbb{S}^{p-1})}_{T}^2 \leq \mathbb{E}_g[dist(g, C^*)^2]$$

Dual cone of C (the Normal cone in our case)

Proof Outline

The GAUSSIAN WIDTH of $S \subset \mathcal{S}^{p-1}$ $\quad w(S) = E[\sup_{z \in S} g^t z]$

From an extension of Gordon's minimum restricted singular value theorem,

Let T(x) be the spherical part of the tangent cone at x w.r.t the atomic norm. If A is a random normalized gaussian map, then x is the unique optimum of the atomic norm minimization provided $M \geq w(T) + \mathcal{O}(1)$

For any cone C, from the properties of tangent cones and Jensen's inequality

$$w(\underbrace{C \cap \mathbb{S}^{p-1}}_{T})^2 \leq \mathbb{E}_g[dist(g, C^*)^2]$$

Dual cone of C (the Normal cone in our case)

So the problem reduces to finding the expected distance between the normal cone and a random gaussian vector

Exact reconstruction

- 5000 measurements considered. Image size = 128 X 128

(a) Peppers

(b) CoSaMP (MSE = 22.8)

(c) Model-Based (MSE = 11.1)

(d) OG Lasso (MSE = 7.8)

Summary

- DWT coefficients can be modeled into groups
- Can use convex methods to recover the image
- Superior to lasso in both denoising and deconvolution applications
- Accuracy depends on "proper" grouping
- Derived a bound on the number of measurements needed for exact recovery of structured sparse signals
- Bound is non asymptotic, and applies to ANY grouping

FUTURE WORK

- "optimal" groupings of DWT coefficients
- Extend to noisy setting

References

- La C. and Do M.N. “Tree based orthogonal matching pursuit algorithm for signal reconstruction”. IEEE International Conference on Image Processing, Atlanta, GA., pages 1277 – 1280, Oct 2006.
- Huang J., Zhang T., and Metaxas D. “Learning with structured sparsity”. Technical report, arXiv:0903.3002. Preprint available at http://arxiv.org/pdf/0903.3002v2, May 2009.
- Jacob L., Obozinski G., and Vert J.P. “Group lasso with overlap and graph lasso”. Proceedings of the 26th International Conference on machine Learning, 2009.
- Schniter P. “Turbo reconstruction of structured sparse signals”. Proc. Conference on Information Sciences and Systems, Mar 2010.
- Jenatton R., Audibert J.Y., and Bach F. “Structured variable selection with sparsity inducing norms”. Tech- nical report, arXiv:0904.3523. Preprint available at http://arxiv.org/pdf/0904.3523v3, Sep 2009.

References

- Baraniuk R.G., Cevher V., M.F. Duarte, and Hegde C. “Model-based compressive sensing”. IEEE Transactions on Information Theory, 2010.
- Wright S.J., Nowak R.D., and Figueiredo M.A.T. “Sparse reconstruction by separable approximation”. Transactions on Signal Processing, 57:2479–2493, 2009.
- Yuan M. and Lin Y. “Model selection and estimation in regression with grouped variables”. Journal of the royal statistical society. Series B, 68:49–67, 2006
- DuarteM.F.,WakinM.B.,and BaraniukR.G.”Wavelet-domain compressive signal reconstruction using a hidden markov tree model.” Proc. International Conference on Acoustics, Speech, and Signal Processing, pages 5137–5140, Mar 2008.