Sparse Modes and a Case of Structed Sparsity (A Note)

Luo Peng
altlp@nuaa.edu.cn

http://parnec.nuaa.edu.cn

Nanjing University of Aeronautics and Astronautics

2012.03.23

Two References

1. Jun Liu, Shuiwang Ji, and Jieping Ye, "A manual of SLEP," http://www.public.asu.edu/~jye02/Software/SLEP.

2. Rao, N.S.; Nowak, R.D.; Wright, S.J.; Kingsbury, N.G., "Convex approaches to model wavelet sparsity patterns" Image Processing (ICIP), 2011 .

Dominant modeling tool

- Genomics
- Genetics
- Signal and audio processing
- Image processing
- Neuroscience (theory of sparse coding)
- Machine learning
- Data mining
- …

Outline

1. Sparse Modes

2. a Case of Structed Sparsity

1. SparseModes

The ℓ_p-norm of the vector $\mathbf{v} \in \mathbb{R}^n$ is defined as $\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}$. The ℓ_0-norm of the vector $\mathbf{v} \in \mathbb{R}^n$ is defined as $\|\mathbf{v}\|_0 = \sum_{i=1}^{n} |v_i|^0$. The Frobenius norm of the matrix $\mathrm{M} \in \mathbb{R}^{n\times m}$ is defined as

$$\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} m_{ij}^2} = \sqrt{\sum_{i=1}^{n} \|\mathbf{m}^i\|_2^2}.$$

$$\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{m} m_{ij}^2} = \sum_{i=1}^{n} \|\mathbf{m}^i\|_2$$

- Let x be the model parameter to be estimated. A commonly employed model for estimating x is

$$\min \ \text{loss}(x) + \lambda \ \text{penalty}(x) \tag{1}$$

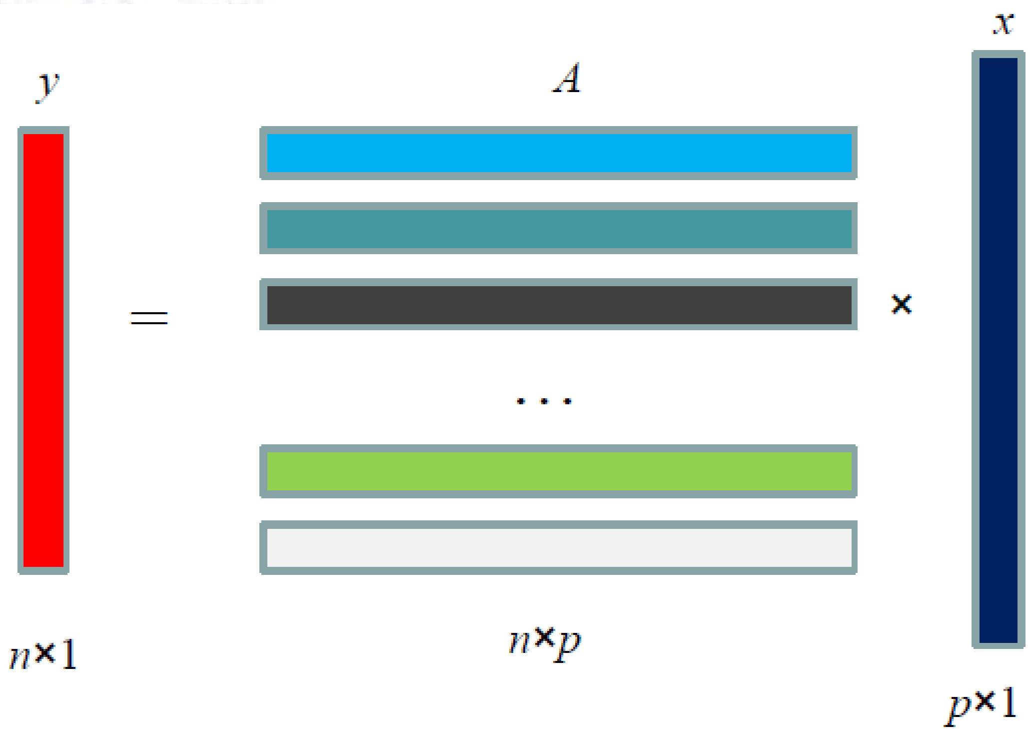
- (1) is equivalent to the following model:

$$\begin{aligned} &\min \ \text{loss}(x) \\ &\text{s.t.} \quad \text{penalty}(x) \le z \end{aligned} \tag{2}$$

loss(x) penalty(x)

- Least squares
- Logistic loss
- Hinge loss
- …

- Zero norm is the natural choice
  - The number of nonzero elements of x
  - Not a valid norm, nonconvex, NP-hard



- *x* is sparse
- *p>>n*
- *A is a measurement matrix satisfying certain conditions*

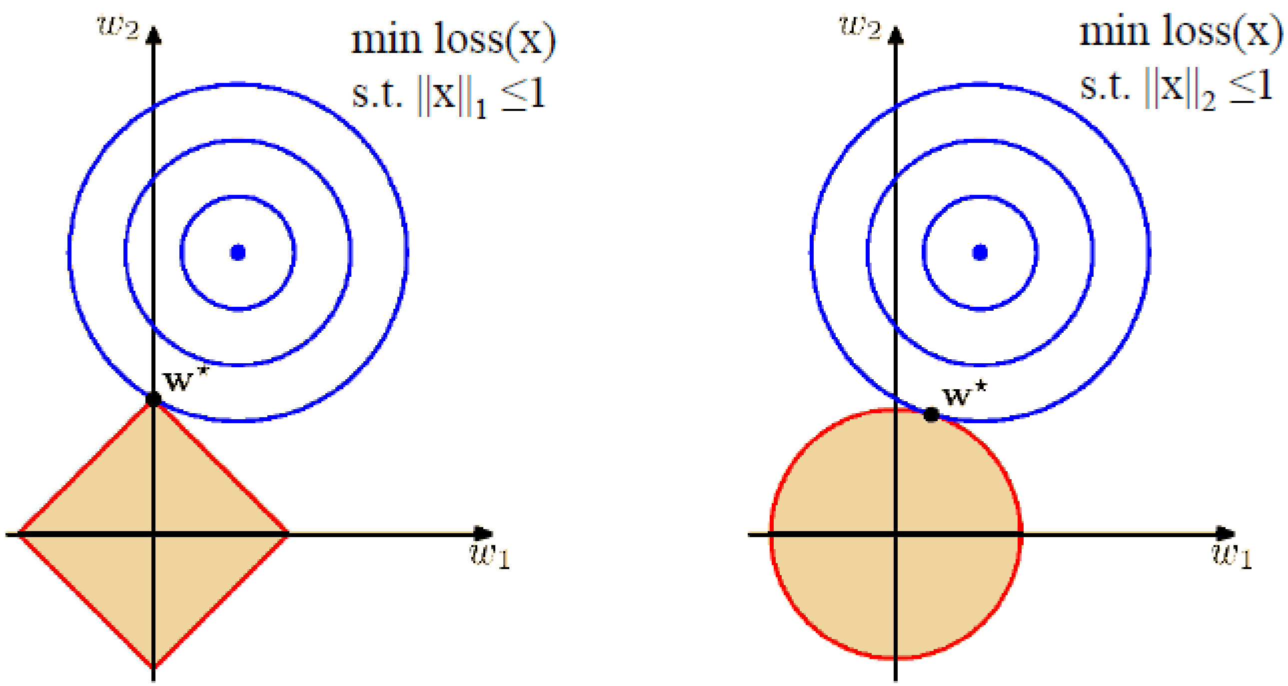
NP hard

$$P_0 \quad \min \ \|x\|_0 \quad \text{s.t.} \ Ax = y$$

$$P_1 \quad \min \ \|x\|_1 \quad \text{s.t.} \ Ax = y$$

$$\min \ \text{loss}(x) + \lambda \, \|x\|_0$$

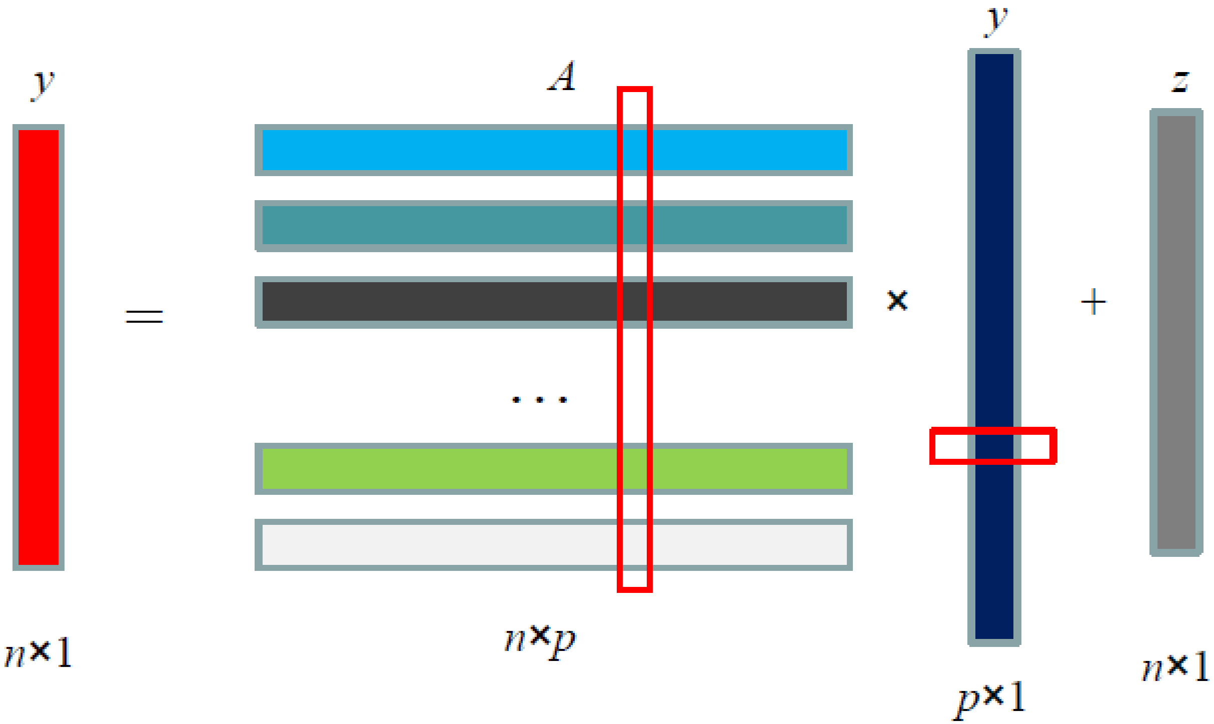
$$\min \ \text{loss}(x) + \lambda \, \|x\|_1$$



$$y = Ax + z$$ ← noise

Lasso

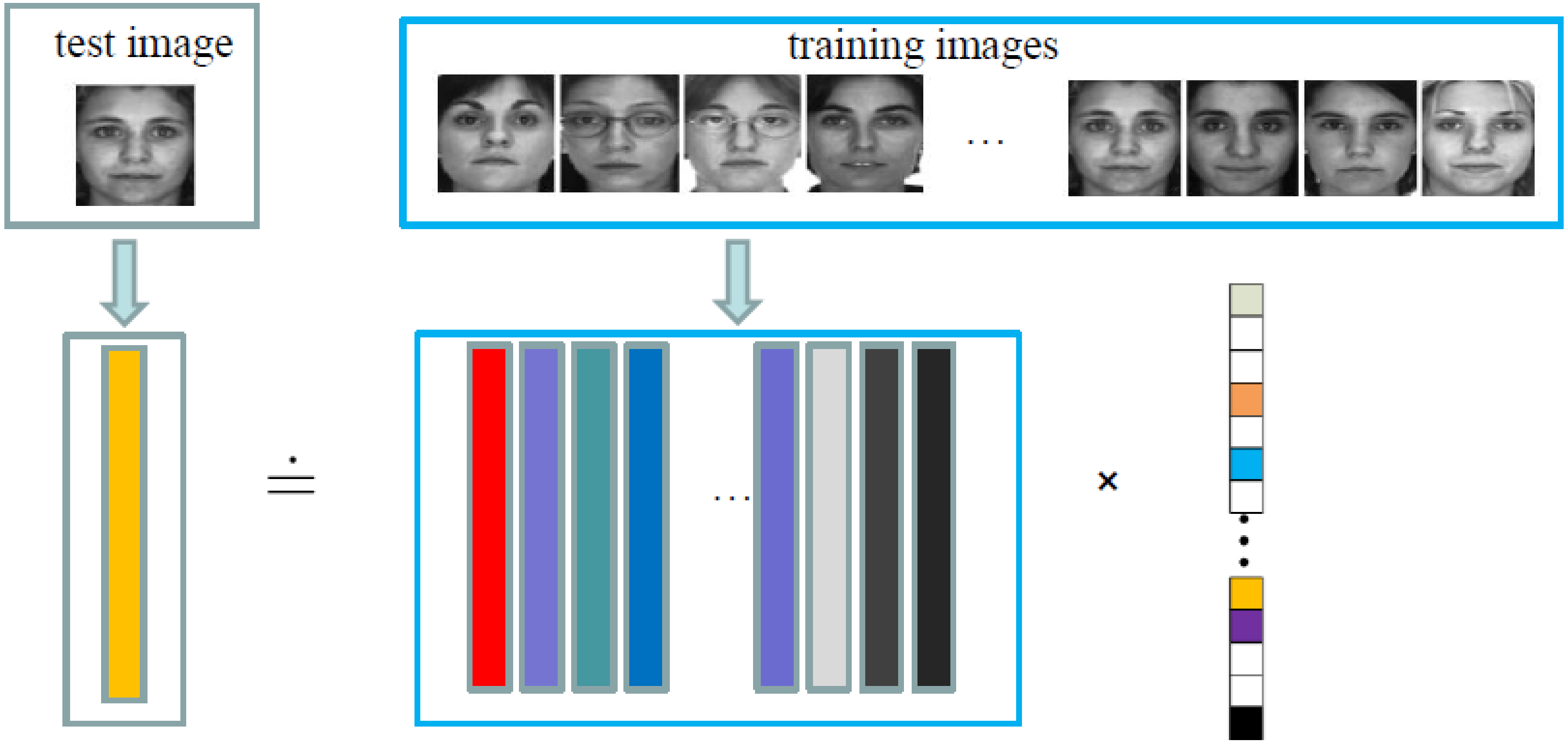
(Tibshirani, 1996)



$$\frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_1$$

Application: Face Recognition

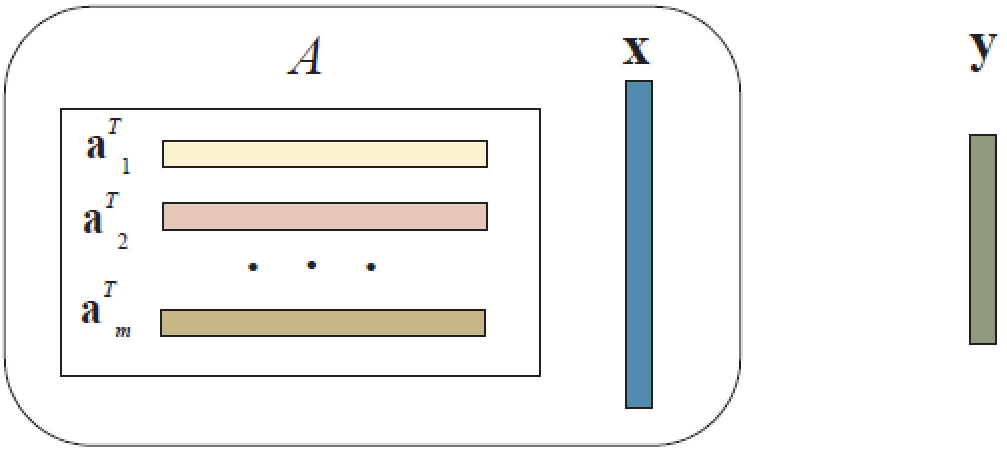
(Wright et al. 2009)



Use the computed sparse coefficients for classification

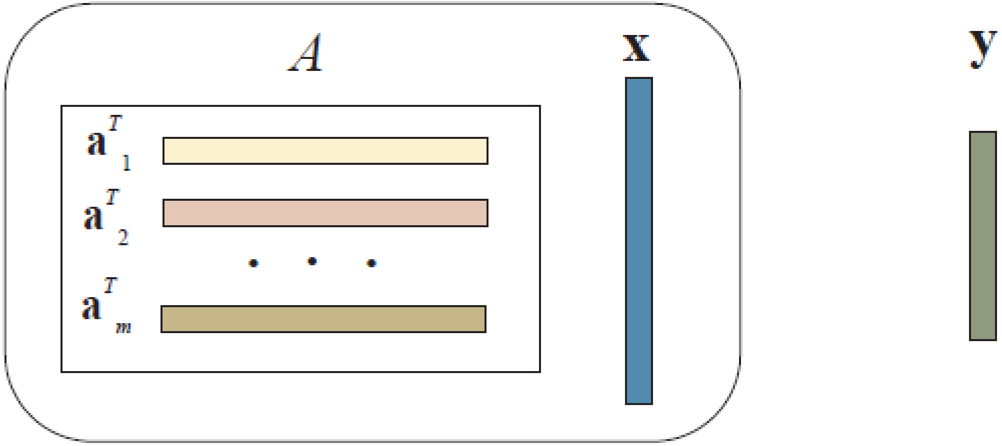
LeastR

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x}-\mathbf{y}\|_2^2 + \frac{\rho}{2}\|\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$



LeastC

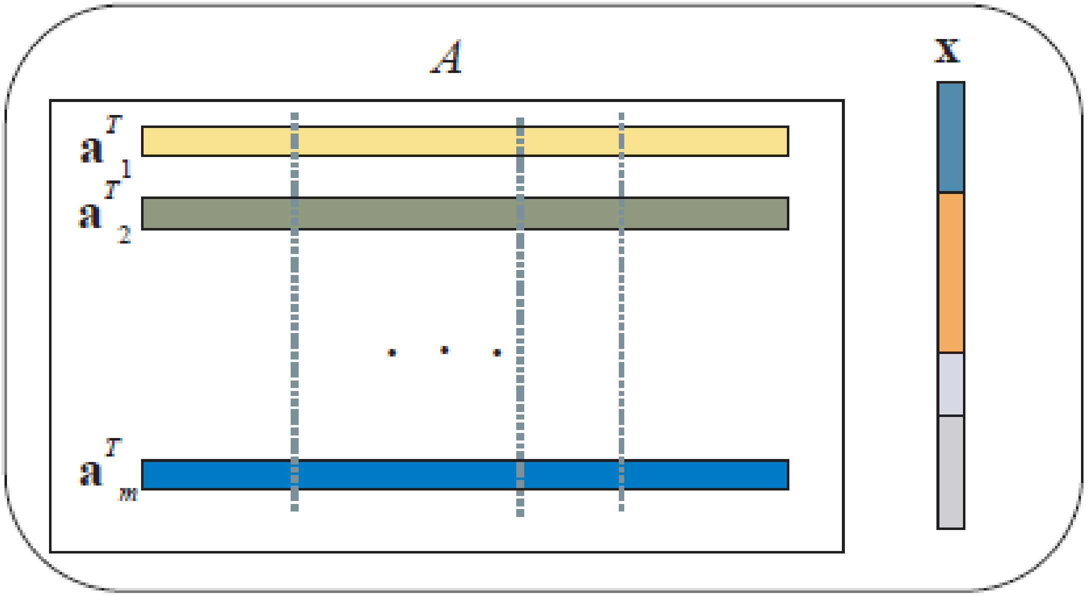
$$\min_{\mathbf{x}} \quad \frac{1}{2}\|A\mathbf{x}-\mathbf{y}\|_2^2 + \frac{\rho}{2}\|\mathbf{x}\|_2^2$$
$$\text{subject to} \quad \|\mathbf{x}\|_1 \le z,$$

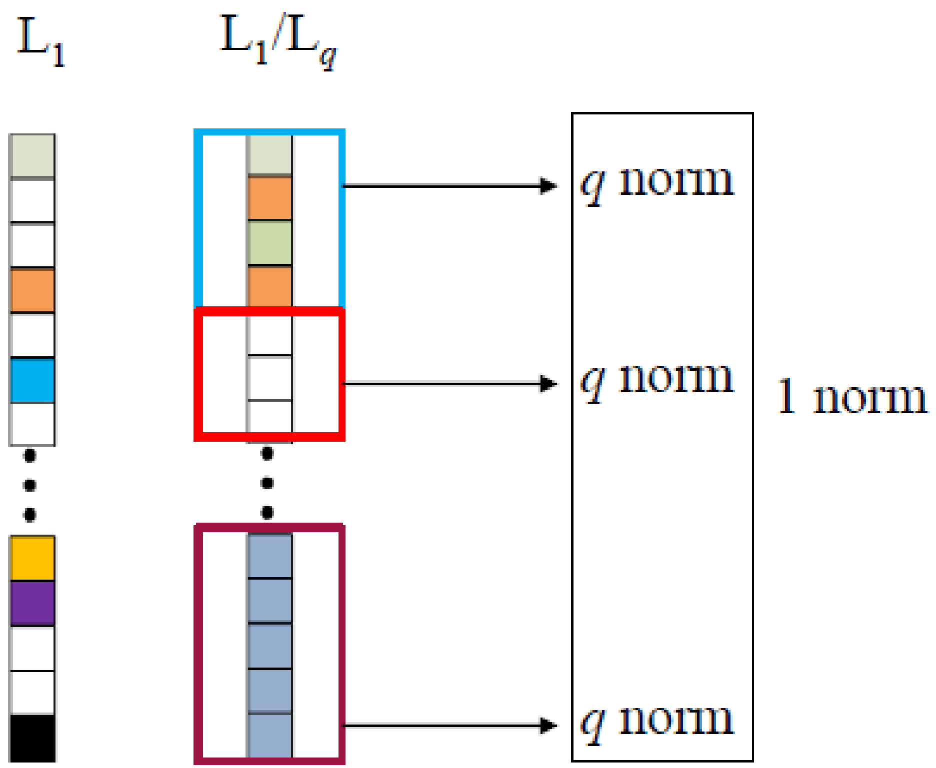


ℓ_1/ℓ_q-Norm Regularized Problems

glLeastR

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x}-\mathbf{y}\|_2^2 + \lambda \sum_{i=1}^{k} w_i^g \|\mathbf{x}_{G_i}\|_q$$

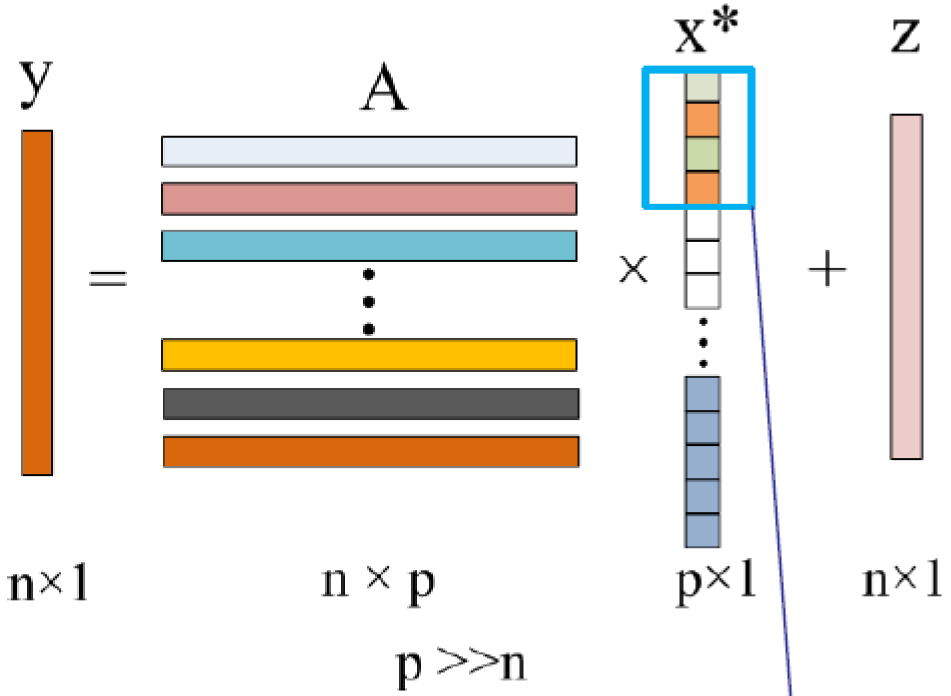




$$\|X\|_{q,1} = \sum_i \|X_{G_i}\|_q$$

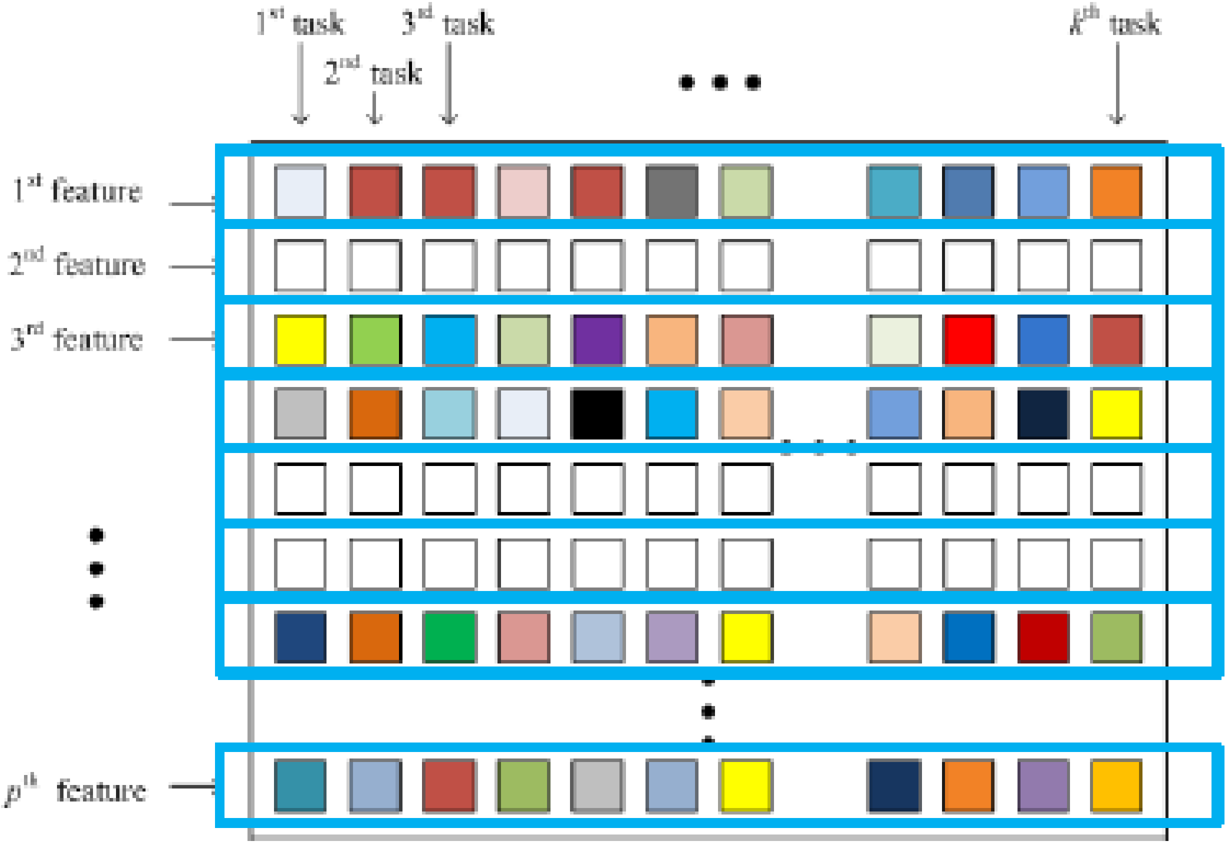
Group Lasso

(Yuan and Lin, 2006)



$$\min \frac{1}{2}\|Ax - y\|_2^2 + \lambda \sum_{i=1}^{g} d_i \|x_{G_i}\|_2$$

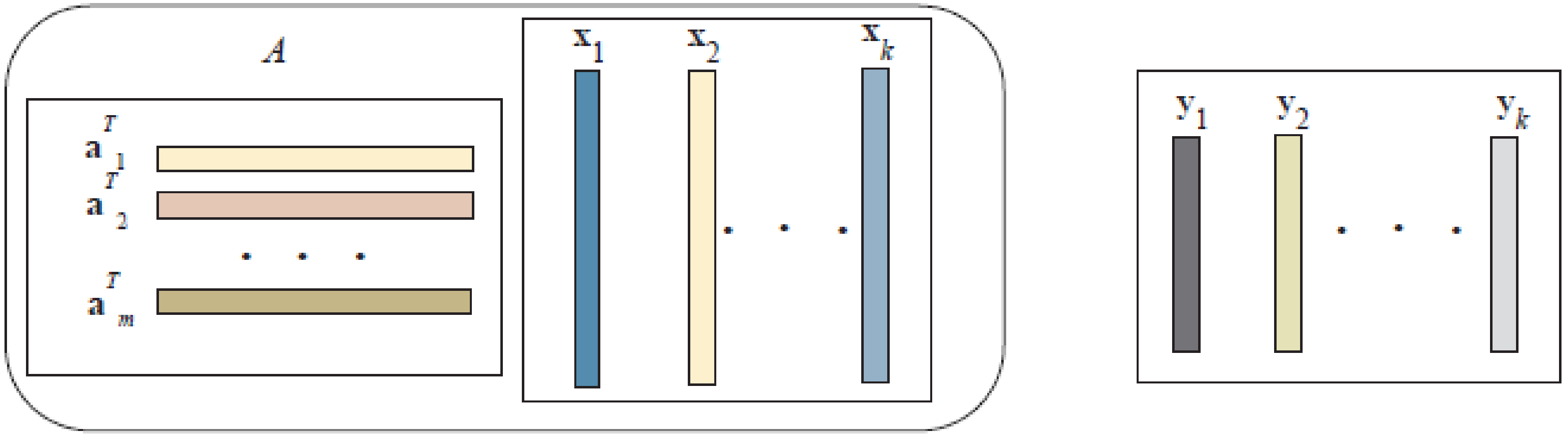
L_1/L_q



1st task
2nd task
3rd task
kth task
1st feature
2nd feature
3rd feature
pth feature

mcLeastR

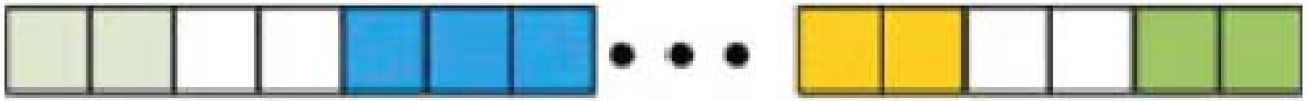
$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_{\ell_1/\ell_q}$$



where $A \in \mathbb{R}^{m \times n}$, $\mathbf{y} \in \mathbb{R}^{m \times k}$, and $\mathbf{x} \in \mathbb{R}^{n \times k}$.

Fused Lasso

$$\min_{\mathbf{x}} \quad \frac{1}{2}\|A\mathbf{x}-\mathbf{y}\|_2^2 + \lambda_1\|\mathbf{x}\|_1 + \lambda_2\sum_{i=1}^{p-1}|x_i - x_{i+1}|$$



Sparse Group Lasso

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x}-\mathbf{y}\|_2^2+\lambda_1\|\mathbf{x}\|_1+\lambda_2\sum_{i=1}^{g} w_i^g\|\mathbf{x}_{G_i}\|_2$$

Tree Structured Group Lasso

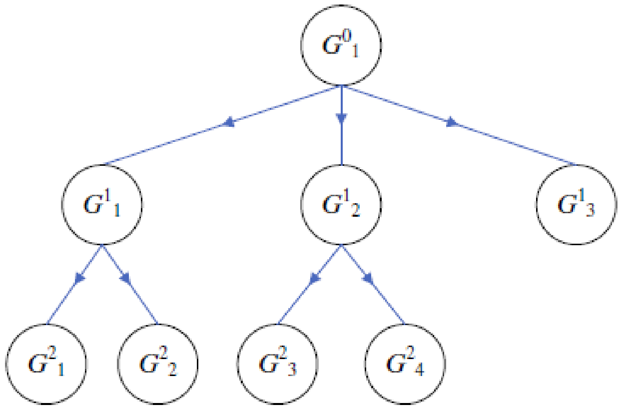


Figure 11: A sample index tree for illustration. Root: $G_1^0 = \{1,2,3,4,5,6,7,8\}$. Depth 1: $G_1^1 = \{1,2\}$, $G_2^1 = \{3,4,5,6\}$, $G_3^1 = \{7,8\}$. Depth 2: $G_1^2 = \{1\}$, $G_2^2 = \{2\}$, $G_3^2 = \{3,4\}$, $G_4^2 = \{5,6\}$.

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \sum_{i=0}^{d} \sum_{j=1}^{n_i} w_j^i \|\mathbf{x}_{G_j^i}\|$$

tree_mtLeastR

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \sum_{i=0}^{d} \sum_{j=1}^{n_i} w_j^i \|\mathbf{x}_{G_j^i}\|$$

Overlapping Group Lasso

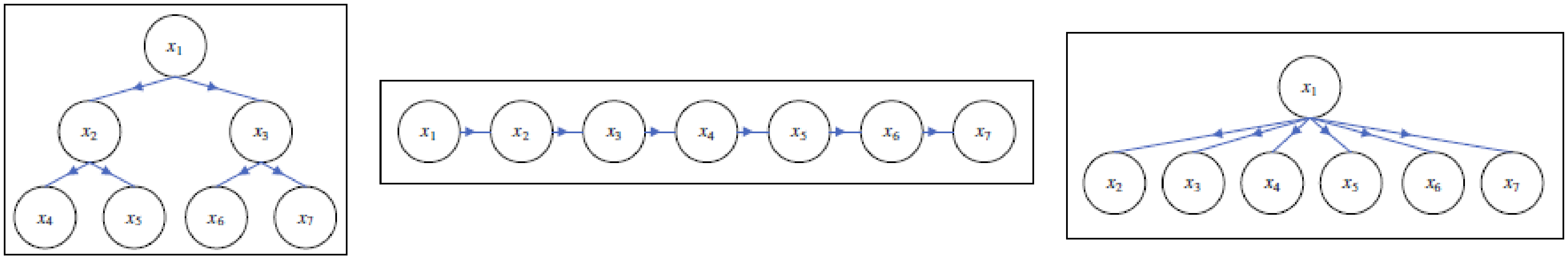
$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_1\|\mathbf{x}\|_1 + \lambda_2 \sum_{i=}^{g} w_i^g \|\mathbf{x}_{G_i}\|$$

the groups G_i may overlap.

Ordered Tree–Nonnegative Max-Heap

$$T^t = (V^t, E^t) V^t = \{x_1, x_2, \dots, x_p\}$$

$$P = \{\mathbf{x} \geq 0, x_i \geq x_j \ \forall (x_i, x_j) \in E^t\}$$

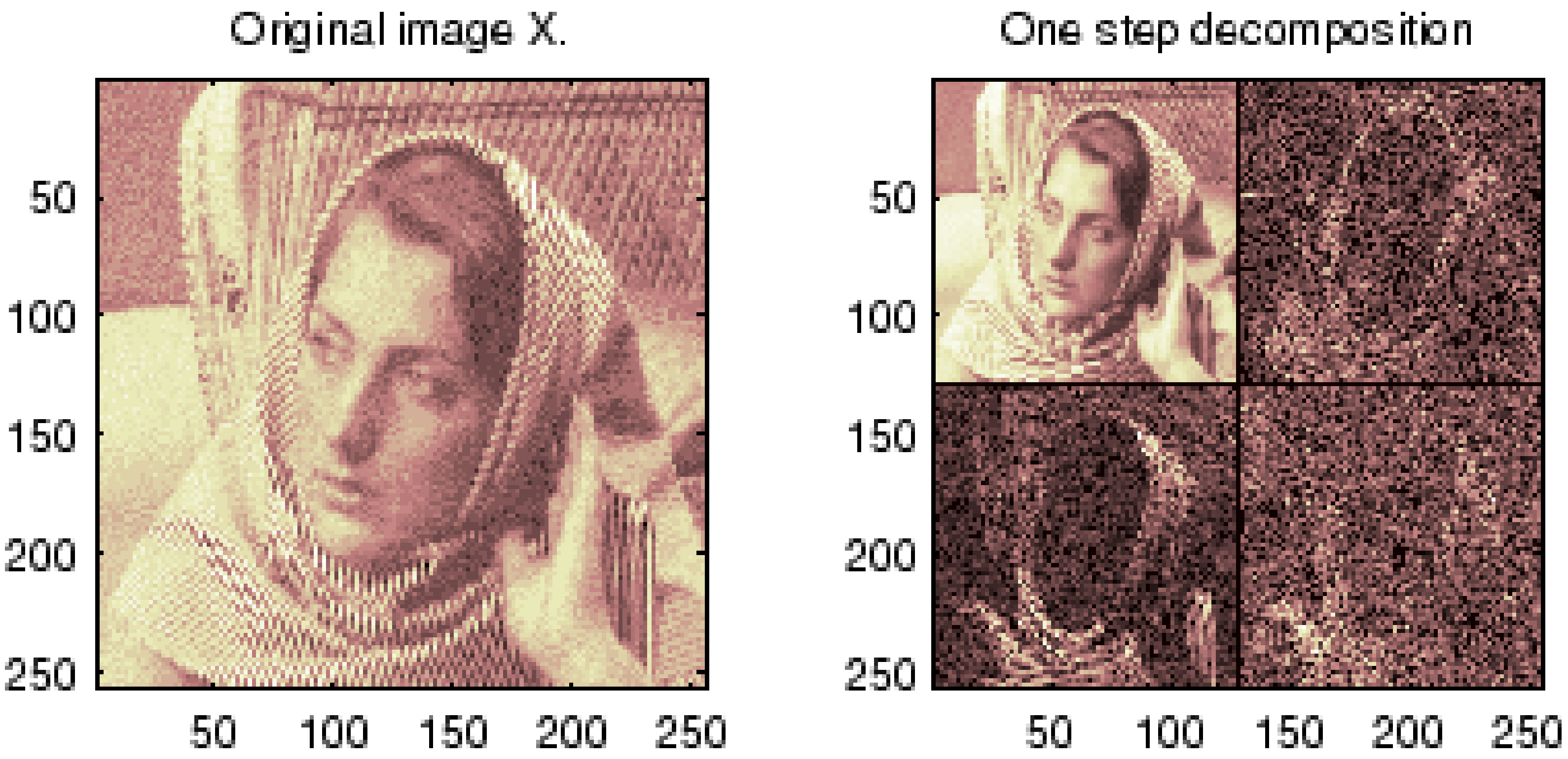


(a) (b) (c)

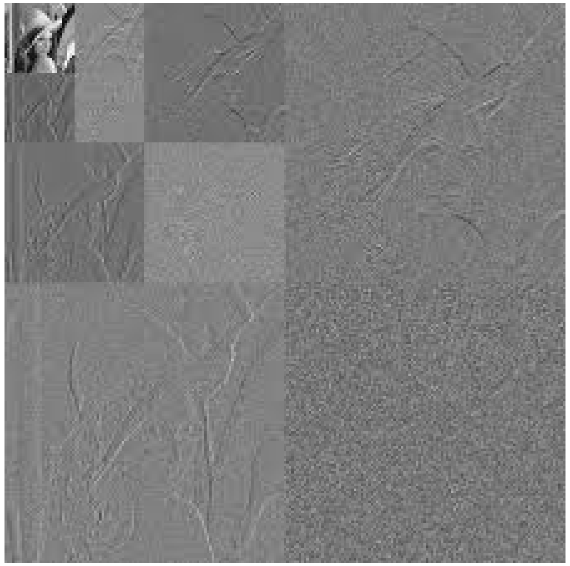
$$\min_{\mathbf{x} \in P} \frac{1}{2} \|A\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

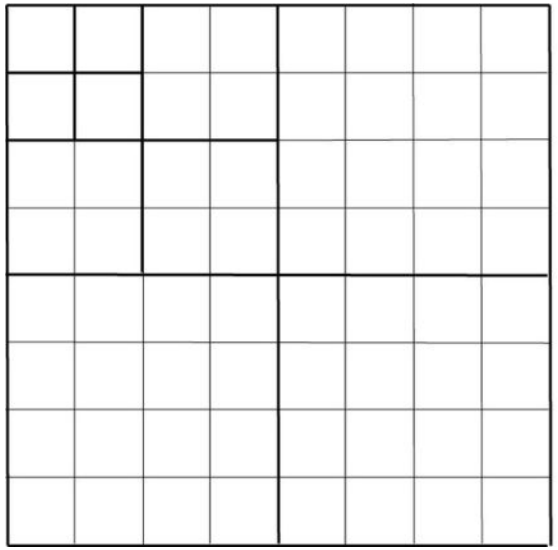
2. An Approach to Model Wavelet Sparsity Patterns

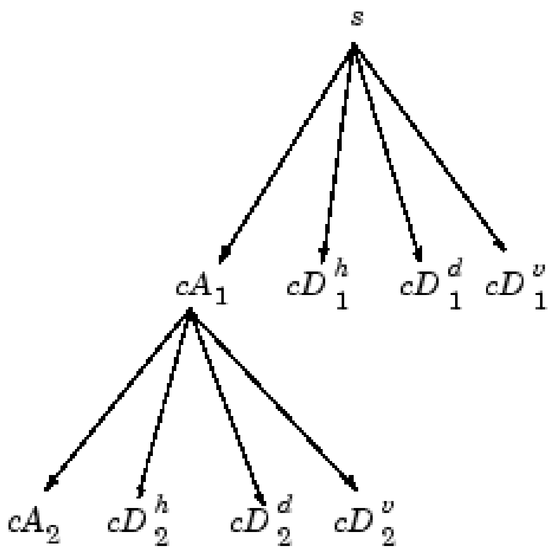
Structured Sparsity



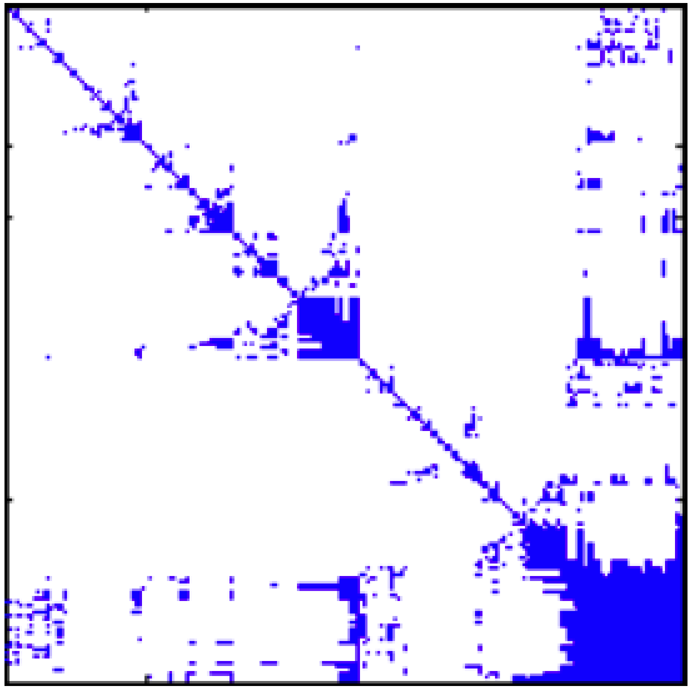
Structured Sparsity







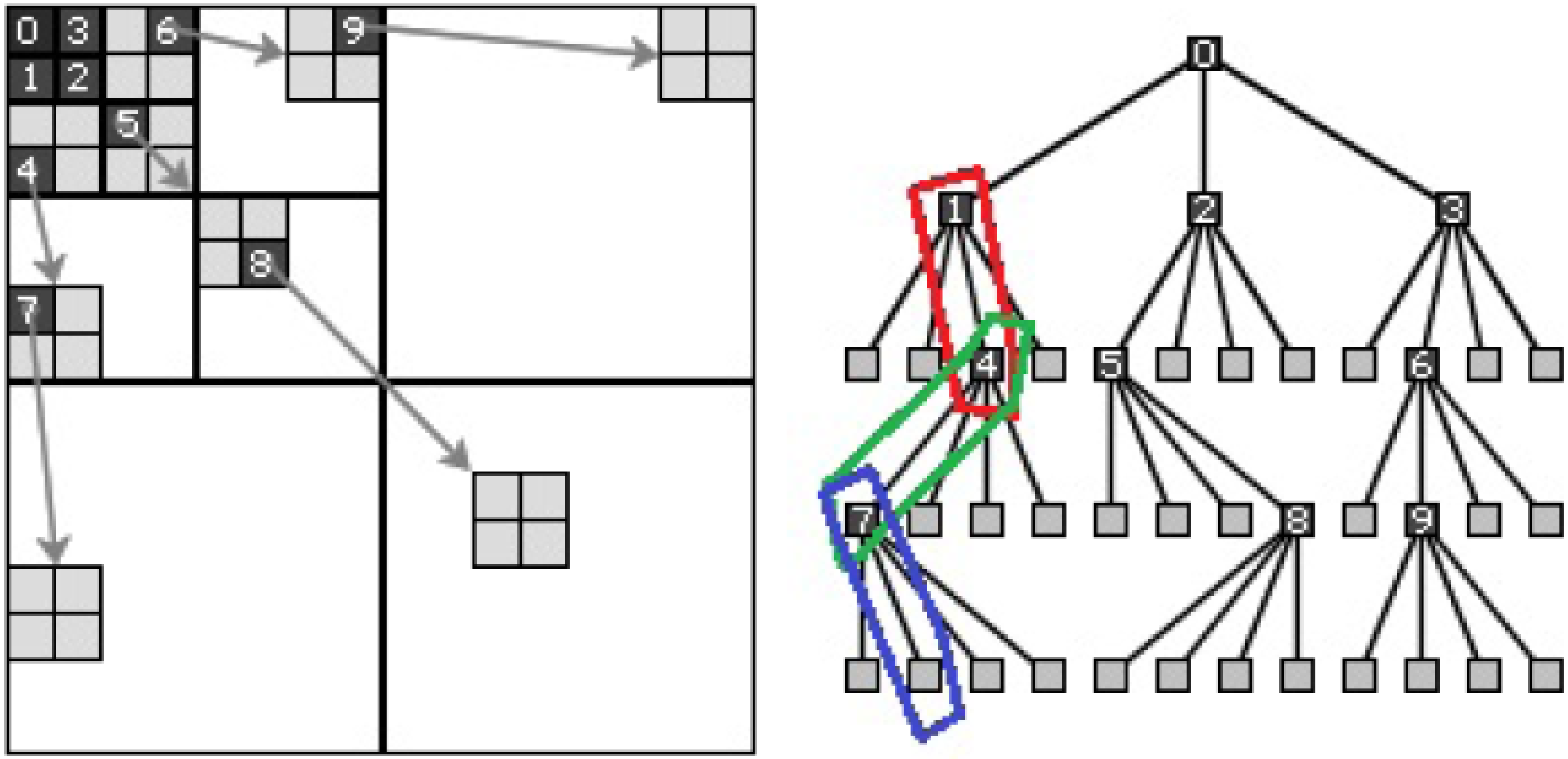
Structured Sparsity



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]


0
3
6
9
1
2
5
4
8
7
0
1
2
3
4
5
6
7
8
9

Dblurring based on wavelet transform:

$$y = Lx + w$$

$$\widehat{\theta} \quad := \quad \arg\min_{\theta} \left\{ \frac{1}{2} \|y - A\theta\|_2^2 + \lambda_l \|\theta\|_1 \right\}$$

$$A := LW$$

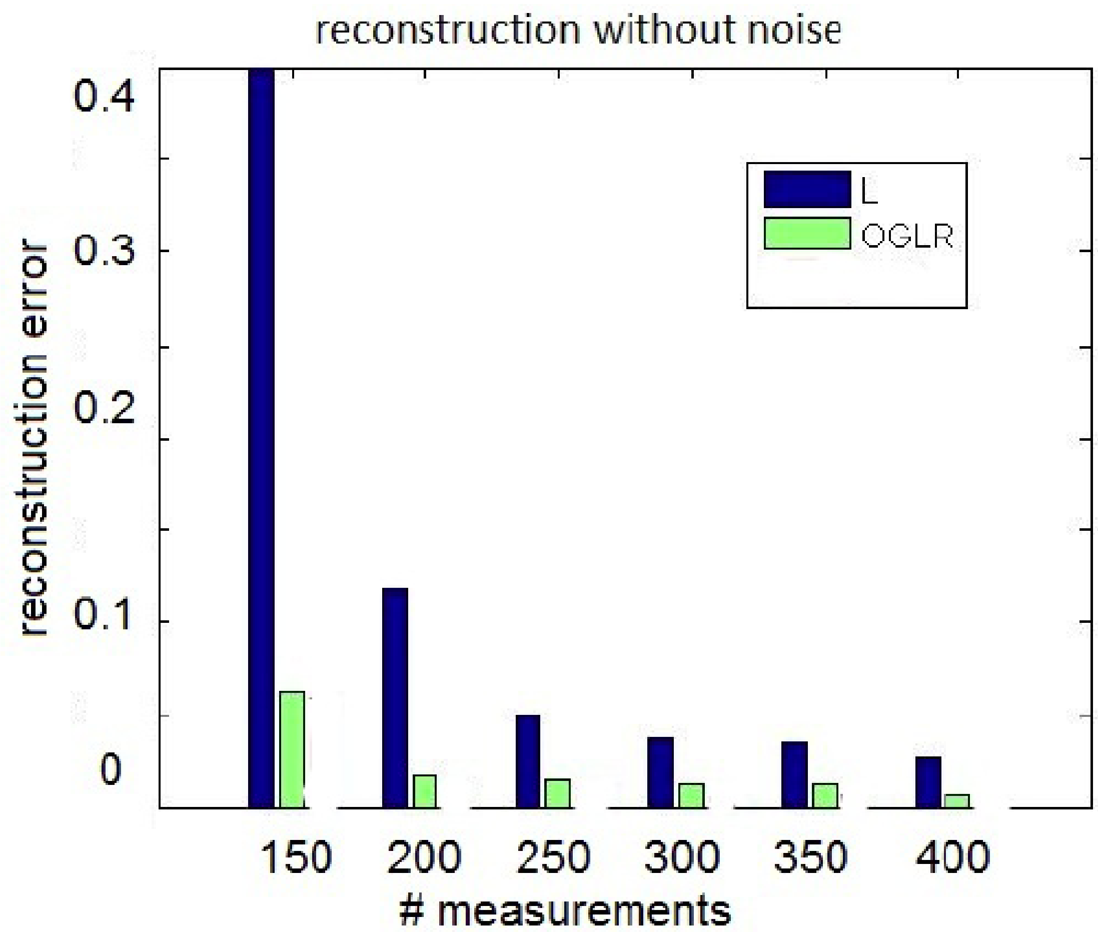
$$\widehat{\theta}_{OGLR} := \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2} \|y - \tilde{A}\tilde{\theta}\|_2^2 + \lambda_{rep} \sum_{\tilde{g} \in \tilde{\mathcal{G}}} ||\tilde{\theta}_{\tilde{g}}||_2 \right\}$$

Dblurring based on wavelet transform:

PSNR = 22.75

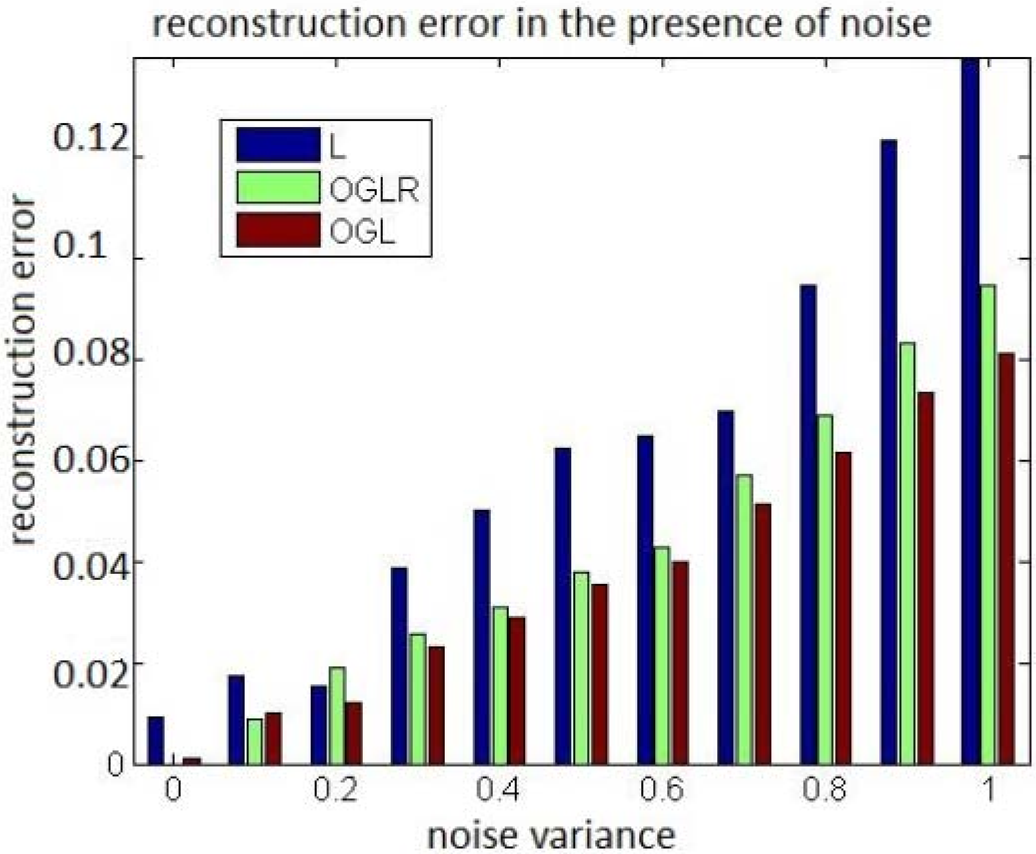
PSNR = 25.08

Dblurring based on wavelet transform:



Dblurring based on wavelet transform:

$$\widehat{\theta}_{OGL} := \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2}\|y - A\theta\|_2^2 + \lambda_{ogl} \sum_{\tilde{g}\in\tilde{\mathcal{G}}} ||\tilde{\theta}_{\tilde{g}}||_2. + \frac{1}{2}\tau^2 \sum_{i=1}^{n} \sum_{j\in J_i} (\theta_i - \theta_i^{(j)})^2 \right\}$$



Thanks