# Reconstruction of Block-Sparse Signals by Using an $\ell_{2/p}$ -Regularized Least-Squares Algorithm

Jeevan K. Pant, Wu-Sheng Lu, and Andreas Antoniou

University of Victoria

May 21, 2012

Compressive Sensing

- Compressive Sensing
- Signal Recovery by Using  $\ell_1$  or  $\ell_p$  Minimization

- Compressive Sensing
- Signal Recovery by Using  $\ell_1$  or  $\ell_p$  Minimization
- Recovery of Block-Sparse Signals

- Compressive Sensing
- Signal Recovery by Using  $\ell_1$  or  $\ell_p$  Minimization
- Recovery of Block-Sparse Signals
- Block-Sparse Signal Recovery by Using  $\ell_{2/p}$  Minimization

- Compressive Sensing
- Signal Recovery by Using  $\ell_1$  or  $\ell_p$  Minimization
- Recovery of Block-Sparse Signals
- Block-Sparse Signal Recovery by Using  $\ell_{2/p}$  Minimization
- Performance Evaluation

- Compressive Sensing
- Signal Recovery by Using  $\ell_1$  or  $\ell_p$  Minimization
- Recovery of Block-Sparse Signals
- Block-Sparse Signal Recovery by Using  $\ell_{2/p}$  Minimization
- Performance Evaluation
- Conclusions

#### Compressive Sensing

■ A signal  $\mathbf{x}(n)$  of length N is K-sparse if it contains K nonzero components with  $K \ll N$ .

#### Compressive Sensing

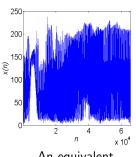
- A signal  $\mathbf{x}(n)$  of length N is K-sparse if it contains K nonzero components with  $K \ll N$ .
- $\blacksquare$  A signal is near K-sparse if it contains K significant components.

#### Compressive Sensing

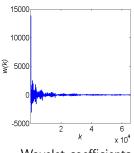
- A signal  $\mathbf{x}(n)$  of length N is K-sparse if it contains K nonzero components with  $K \ll N$ .
- $\blacksquare$  A signal is near K-sparse if it contains K significant components.
- Example: an image with near sparse wavelet coefficients:



An image of Lena



An equivalent 1-D signal



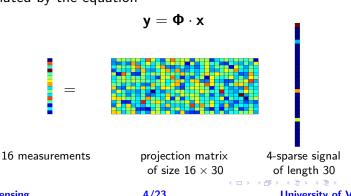
Wavelet coefficients

#### Compressive Sensing, cont'd

Compressive sensing (CS) is a data acquisition process whereby a sparse signal  $\mathbf{x}(n)$  represented by a vector  $\mathbf{x}$  of length N is determined using a small number of projections represented by a matrix  $\mathbf{\Phi}$  of dimension  $M \times N$ .

#### Compressive Sensing, cont'd

- Compressive sensing (CS) is a data acquisition process whereby a sparse signal  $\mathbf{x}(n)$  represented by a vector  $\mathbf{x}$  of length N is determined using a small number of projections represented by a matrix  $\Phi$  of dimension  $M \times N$ .
- In such a process, measurement vector **y** and signal vector **x** are interrelated by the equation



The inverse problem of recovering signal x from measurement y such that

$$egin{array}{cccc} oldsymbol{\Phi} & oldsymbol{\mathsf{x}} & oldsymbol{\mathsf{y}} & oldsymbol{\mathsf{y}} \ |_{M imes N} & |_{N imes 1} & |_{M imes 1} \end{array}$$

is an ill-posed problem.

The inverse problem of recovering signal x from measurement y such that

$$\begin{array}{c|c} \boldsymbol{\Phi} \cdot \mathbf{x} = \mathbf{y} \\ \begin{matrix} \mid & \mid \\ M \times N & N \times 1 \end{matrix}$$

is an ill-posed problem.

•  $\ell_2$  minimization often fails to yield a sparse  $\mathbf{x}$ , i.e., a signal obtained as

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{arg min}} ||\mathbf{x}||_2$$
 subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

is often not sparse.

The inverse problem of recovering signal x from measurement y such that

$$\begin{array}{c|c} \boldsymbol{\Phi} \cdot \boldsymbol{x} = \boldsymbol{y} \\ \begin{matrix} | & | \\ N \times N & N \times 1 \end{matrix}$$

is an ill-posed problem.

•  $\ell_2$  minimization often fails to yield a sparse  $\mathbf{x}$ , i.e., a signal obtained as

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{arg min}} ||\mathbf{x}||_2$$
 subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

is often not sparse.

**A** sparse **x** can be recovered using  $\ell_1$  minimization as

$$\mathbf{x}^* = \underset{\mathbf{x}}{\mathsf{arg}} \; \mathsf{min} ||\mathbf{x}||_1 \quad \mathsf{subject to} \quad \; \mathbf{\Phi} \mathbf{x} = \mathbf{y}$$

- Recently,  $\ell_p$  minimization based algorithms have been shown to recover sparse signals using fewer measurements.
- In these algorithms, the signal is recovered by using the optimization problem

minimize 
$$||\mathbf{x}||_p^p = \sum_{i=1}^N |x_i|^p$$
 subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

where p < 1.

- Recently,  $\ell_p$  minimization based algorithms have been shown to recover sparse signals using fewer measurements.
- In these algorithms, the signal is recovered by using the optimization problem

minimize 
$$||\mathbf{x}||_p^p = \sum_{i=1}^N |x_i|^p$$
  
subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

where p < 1.

Note that the objective function  $||\mathbf{x}||_p^p$  in the above problem is nonconvex and nondifferentiable.

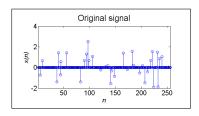
- Recently,  $\ell_p$  minimization based algorithms have been shown to recover sparse signals using fewer measurements.
- In these algorithms, the signal is recovered by using the optimization problem

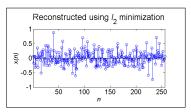
minimize 
$$||\mathbf{x}||_p^p = \sum_{i=1}^N |x_i|^p$$
  
subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

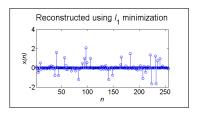
where p < 1.

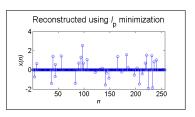
- Note that the objective function  $||\mathbf{x}||_p^p$  in the above problem is nonconvex and nondifferentiable.
- Despite this, it has been shown in the literature that if the above problem is solved with sufficient care, improved reconstruction performance can be achieved.

**Example:** N = 256, K = 35, M = 100.









Let N, d, and N/d be positive integers such that d < N and N/d < N.

- Let N, d, and N/d be positive integers such that d < N and N/d < N.
- A signal **x** of length N can be divided into N/d blocks as

$$\mathbf{x} = \begin{bmatrix} \mathbf{\tilde{x}}_1 \ \mathbf{\tilde{x}}_2 \ \cdots \ \mathbf{\tilde{x}}_{N/d} \end{bmatrix}^T$$

where

$$\tilde{\mathbf{x}}_i = \begin{bmatrix} x_{(i-1)d+1} & x_{(i-1)d+1} & \cdots & x_{(i-1)d+d} \end{bmatrix}^T$$

for i = 1, 2, ..., N/d.

- Let N, d, and N/d be positive integers such that d < N and N/d < N.
- A signal **x** of length N can be divided into N/d blocks as

$$\mathbf{x} = \begin{bmatrix} \mathbf{\tilde{x}}_1 \ \mathbf{\tilde{x}}_2 \ \cdots \ \mathbf{\tilde{x}}_{N/d} \end{bmatrix}^T$$

where

$$\widetilde{\mathbf{x}}_i = \begin{bmatrix} x_{(i-1)d+1} & x_{(i-1)d+1} & \cdots & x_{(i-1)d+d} \end{bmatrix}^T$$

for i = 1, 2, ..., N/d.

■ Signal  $\mathbf{x}$  is said to be K-block sparse if it has K nonzero blocks with  $K \ll N/d$ .

- Let N, d, and N/d be positive integers such that d < N and N/d < N.
- A signal **x** of length N can be divided into N/d blocks as

$$\mathbf{x} = \begin{bmatrix} \mathbf{\tilde{x}}_1 \ \mathbf{\tilde{x}}_2 \ \cdots \ \mathbf{\tilde{x}}_{N/d} \end{bmatrix}^T$$

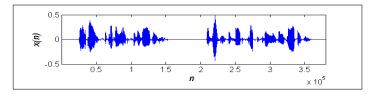
where

$$\mathbf{\tilde{x}}_i = \begin{bmatrix} x_{(i-1)d+1} & x_{(i-1)d+1} & \cdots & x_{(i-1)d+d} \end{bmatrix}^T$$
 for  $i = 1, 2, \dots, N/d$ .

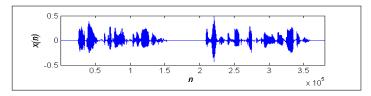
- Signal  $\mathbf{x}$  is said to be K-block sparse if it has K nonzero blocks with  $K \ll N/d$ .
- Note that the definition of K-sparse in the conventional CS is the special case of K-block sparse with d=1.

 Block-sparsity naturally arises in various signals such as speech signals, multiband signals, and some images.

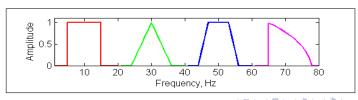
- Block-sparsity naturally arises in various signals such as speech signals, multiband signals, and some images.
- Speech signal:



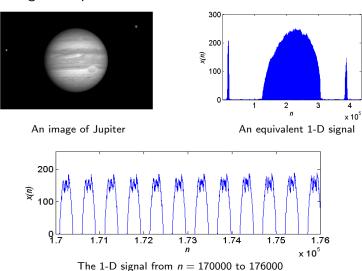
- Block-sparsity naturally arises in various signals such as speech signals, multiband signals, and some images.
- Speech signal:



Multiband spectrum:



An image of Jupiter:



■ The block sparsity of a signal can be measured using the  $\ell_{2/0}$ -pseudonorm which is given by

$$||\mathbf{x}||_{2/0} = \sum_{i=1}^{N/d} (||\mathbf{\tilde{x}}_i||_2)^0$$

■ The block sparsity of a signal can be measured using the  $\ell_{2/0}$ -pseudonorm which is given by

$$||\mathbf{x}||_{2/0} = \sum_{i=1}^{N/d} (||\mathbf{\tilde{x}}_i||_2)^0$$

■ The value of function  $||\mathbf{x}||_{2/0}$  is equal to the number of blocks of  $\mathbf{x}$  which have all-nonzero components.

■ The block sparsity of a signal can be measured using the  $\ell_{2/0}$ -pseudonorm which is given by

$$||\mathbf{x}||_{2/0} = \sum_{i=1}^{N/d} (||\mathbf{\tilde{x}}_i||_2)^0$$

- The value of function  $||\mathbf{x}||_{2/0}$  is equal to the number of blocks of  $\mathbf{x}$  which have all-nonzero components.
- A block-sparse signal can therefore be recovered by solving the optimization problem

minimize 
$$||\mathbf{x}||_{2/0}$$
 subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

■ The block sparsity of a signal can be measured using the  $\ell_{2/0}$ -pseudonorm which is given by

$$||\mathbf{x}||_{2/0} = \sum_{i=1}^{N/d} (||\mathbf{\tilde{x}}_i||_2)^0$$

- The value of function  $||\mathbf{x}||_{2/0}$  is equal to the number of blocks of  $\mathbf{x}$  which have all-nonzero components.
- A block-sparse signal can therefore be recovered by solving the optimization problem

minimize 
$$||\mathbf{x}||_{2/0}$$
 subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

Unfortunately, this problem is nonconvex with combinatorial complexity.

 A practical method for recovering a block sparse signal is to solve the problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & ||\mathbf{x}||_{2/1} \\ \text{subject to} & \mathbf{\Phi}\mathbf{x} = \mathbf{y} \end{array}$$

where

$$||\mathbf{x}||_{2/1} = \sum_{i=1}^{N/d} ||\mathbf{\tilde{x}}_i||_2$$

 A practical method for recovering a block sparse signal is to solve the problem

minimize 
$$||\mathbf{x}||_{2/1}$$
 subject to  $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$ 

where

$$||\mathbf{x}||_{2/1} = \sum_{i=1}^{N/d} ||\mathbf{\tilde{x}}_i||_2$$

■ Note that function  $||\mathbf{x}||_{2/1}$  is the  $\ell_1$  norm of the vector

$$[||\mathbf{x}_1||_2 ||\mathbf{x}_2||_2 \cdots ||\mathbf{x}_{N/d}||_2]^T$$
,

which essentially gives a measure of the inter-block sparsity of  $\mathbf{x}$ .

 A practical method for recovering a block sparse signal is to solve the problem

$$\begin{array}{ll} \underset{\textbf{x}}{\text{minimize}} & ||\textbf{x}||_{2/1} \\ \text{subject to} & \boldsymbol{\Phi}\textbf{x} = \textbf{y} \end{array}$$

where

$$||\mathbf{x}||_{2/1} = \sum_{i=1}^{N/d} ||\mathbf{\tilde{x}}_i||_2$$

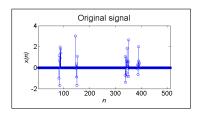
■ Note that function  $||\mathbf{x}||_{2/1}$  is the  $\ell_1$  norm of the vector

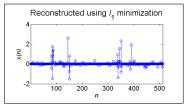
$$[||\mathbf{x}_1||_2 ||\mathbf{x}_2||_2 \cdots ||\mathbf{x}_{N/d}||_2]^T$$
,

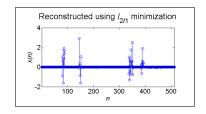
which essentially gives a measure of the inter-block sparsity of  ${\bf x}$ .

The above problem is a convex programming problem which can be solved using a semidefinite-programming or a second-order cone-programming (SOCP) solver.

Example: N = 512, d = 8, K = 5, M = 100.







# Block-Sparse Signal Recovery by Using $\ell_{2/p}$ Minimization

We propose reconstructing a block-sparse signal  ${\bf x}$  from measurement  ${\bf y}$  by solving the  $\ell_{2/p}$ -regularized least-squares problem

minimize 
$$F_{\epsilon}(\mathbf{x}) = \frac{1}{2} ||\mathbf{\Phi}\mathbf{x} - \mathbf{y}||_{2}^{2} + \lambda ||\mathbf{x}||_{2/p,\epsilon}^{p}$$
 (P)

with p < 1 for a small  $\epsilon$  where

$$||\mathbf{x}||_{2/p,\epsilon}^{p} = \sum_{i=1}^{N/d} (||\tilde{\mathbf{x}}_{i}||_{2}^{2} + \epsilon^{2})^{p/2}$$

We propose reconstructing a block-sparse signal  ${\bf x}$  from measurement  ${\bf y}$  by solving the  $\ell_{2/p}$ -regularized least-squares problem

minimize 
$$F_{\epsilon}(\mathbf{x}) = \frac{1}{2} ||\mathbf{\Phi}\mathbf{x} - \mathbf{y}||_{2}^{2} + \lambda ||\mathbf{x}||_{2/p,\epsilon}^{p}$$
 (P)

with p < 1 for a small  $\epsilon$  where

$$||\mathbf{x}||_{2/p,\epsilon}^{p} = \sum_{i=1}^{N/d} (||\tilde{\mathbf{x}}_{i}||_{2}^{2} + \epsilon^{2})^{p/2}$$

Note that

$$\begin{split} &\lim_{\epsilon \to 0} ||\mathbf{x}||_{2/p,\epsilon}^p = ||\mathbf{x}||_{2/p}^p \\ &\lim_{p \to 0} ||\mathbf{x}||_{2/p}^p = ||\mathbf{x}||_{2/0} \end{split}$$

■ Good signal reconstruction performance is expected when problem  ${\bf P}$  on slide 14 is solved with a sufficiently small  $\epsilon$ .

- Good signal reconstruction performance is expected when problem  ${\bf P}$  on slide 14 is solved with a sufficiently small  $\epsilon$ .
- However, for small  $\epsilon$  the objective function  $F_{\epsilon}(\mathbf{x})$  becomes highly nonconvex and nearly nondifferentiable.

- Good signal reconstruction performance is expected when problem  ${\bf P}$  on slide 14 is solved with a sufficiently small  $\epsilon$ .
- However, for small  $\epsilon$  the objective function  $F_{\epsilon}(\mathbf{x})$  becomes highly nonconvex and nearly nondifferentiable.
- The larger the  $\epsilon$ , the easier the optimization of  $F_{\epsilon}(\mathbf{x})$ .

- Good signal reconstruction performance is expected when problem  ${\bf P}$  on slide 14 is solved with a sufficiently small  $\epsilon$ .
- However, for small  $\epsilon$  the objective function  $F_{\epsilon}(\mathbf{x})$  becomes highly nonconvex and nearly nondifferentiable.
- The larger the  $\epsilon$ , the easier the optimization of  $F_{\epsilon}(\mathbf{x})$ .
- Therefore, we propose to solve problem P on slide 14 by using the following sequential optimization procedure:
  - Choose a sufficiently large value of  $\epsilon$  and solve problem  $\mathbf{P}$  using Fletcher-Reeves' conjugate-gradient (CG) algorithm. Set the solution to  $\mathbf{x}$ .
  - Reduce the value of  $\epsilon$ , use **x** as an initializer, and solve problem **P** again.
  - Repeat this procedure until problem **P** is solved for a sufficiently small value of  $\epsilon$ . Output the final solution and stop.

■ In the kth iteration of Fletcher-Reeves' CG algorithm, iterate  $\mathbf{x}_k$  is updated as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

where

$$\begin{array}{rcl} \mathbf{d}_k & = & -\mathbf{g}_k + \beta_{k-1} \mathbf{d}_{k-1} \\ \beta_{k-1} & = & \frac{||\mathbf{g}_k||_2^2}{||\mathbf{g}_{k-1}||_2^2} \end{array}$$

■ In the kth iteration of Fletcher-Reeves' CG algorithm, iterate  $\mathbf{x}_k$  is updated as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

where

$$\mathbf{d}_{k} = -\mathbf{g}_{k} + \beta_{k-1}\mathbf{d}_{k-1}$$
$$\beta_{k-1} = \frac{||\mathbf{g}_{k}||_{2}^{2}}{||\mathbf{g}_{k-1}||_{2}^{2}}$$

Given  $\mathbf{x}_k$  and  $\mathbf{d}_k$ , the step size  $\alpha_k$  is obtained by solving the optimization problem

minimize 
$$f(\alpha) = F_{\epsilon}(\mathbf{x}_k + \alpha \mathbf{d}_k)$$

■ By setting the first derivative of  $f(\alpha)$  to zero, we get

$$\alpha = G(\alpha)$$

where

$$G(\alpha) = -\frac{\mathbf{d}_{k}^{T} \mathbf{\Phi}^{T} (\mathbf{\Phi} \mathbf{x}_{k} - \mathbf{y}) + \lambda \cdot p \cdot \sum_{i=1}^{N/d} \gamma_{i} \cdot (\tilde{\mathbf{x}}_{ki}^{T} \tilde{\mathbf{d}}_{ki})}{||\mathbf{\Phi} \mathbf{d}_{k}||_{2}^{2} + \lambda \cdot p \cdot \sum_{i=1}^{N/d} \gamma_{i} \cdot (\tilde{\mathbf{d}}_{ki}^{T} \tilde{\mathbf{d}}_{ki})}$$

$$\gamma_{i} = (||\tilde{\mathbf{x}}_{i} + \alpha \tilde{\mathbf{d}}_{i}||_{2}^{2} + \epsilon^{2})^{p/2-1}$$

■ By setting the first derivative of  $f(\alpha)$  to zero, we get

$$\alpha = G(\alpha)$$

where

$$G(\alpha) = -\frac{\mathbf{d}_{k}^{T} \mathbf{\Phi}^{T} (\mathbf{\Phi} \mathbf{x}_{k} - \mathbf{y}) + \lambda \cdot p \cdot \sum_{i=1}^{N/d} \gamma_{i} \cdot (\tilde{\mathbf{x}}_{ki}^{T} \tilde{\mathbf{d}}_{ki})}{||\mathbf{\Phi} \mathbf{d}_{k}||_{2}^{2} + \lambda \cdot p \cdot \sum_{i=1}^{N/d} \gamma_{i} \cdot (\tilde{\mathbf{d}}_{ki}^{T} \tilde{\mathbf{d}}_{ki})}$$

$$\gamma_{i} = (||\tilde{\mathbf{x}}_{i} + \alpha \tilde{\mathbf{d}}_{i}||_{2}^{2} + \epsilon^{2})^{p/2-1}$$

In the above equations,  $\tilde{\mathbf{x}}_{ki}$  and  $\tilde{\mathbf{d}}_{ki}$  are the *i*th blocks of vectors  $\mathbf{x}_k$  and  $\mathbf{d}_k$ , respectively.

■ Step size  $\alpha_k$  is determined by using the recursive relation

$$\alpha_{l+1} = G(\alpha_l)$$
 for  $l = 1, 2, ...$ 

■ Step size  $\alpha_k$  is determined by using the recursive relation

$$\alpha_{l+1} = G(\alpha_l)$$
 for  $l = 1, 2, \dots$ 

According to Banach's fixed-point theorem, if  $|dG(\alpha)/d\alpha| < 1$  then function  $G(\alpha)$  is a contraction mapping, i.e.,

$$|G(\alpha_1) - G(\alpha_2)| \le \eta |\alpha_1 - \alpha_2|$$

with  $\eta < 1$  and, as a consequence, the above recursion converges to a solution.

■ Step size  $\alpha_k$  is determined by using the recursive relation

$$\alpha_{I+1} = G(\alpha_I)$$
 for  $I = 1, 2, \dots$ 

According to Banach's fixed-point theorem, if  $|dG(\alpha)/d\alpha| < 1$  then function  $G(\alpha)$  is a contraction mapping, i.e.,

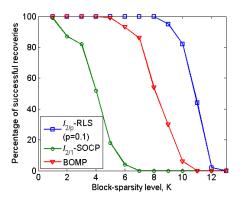
$$|G(\alpha_1) - G(\alpha_2)| \le \eta |\alpha_1 - \alpha_2|$$

with  $\eta < 1$  and, as a consequence, the above recursion converges to a solution.

Extensive experimental results have shown that function  $G(\alpha)$  for function  $f(\alpha)$  is, in practice, a contraction mapping.

#### Performance Evaluation

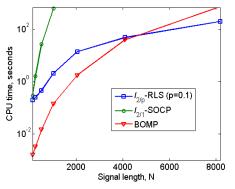
Number of perfectly recovered instances with N = 512, M = 100, and d = 8 over 100 runs.



 $\ell_{2/p}\text{-RLS:}$  Proposed  $\ell_{2/p}\text{-Regularized}$  Least-Squares  $\ell_{2/1}\text{-SOCP:}$   $\ell_{2/1}$  Second-Order Cone-Programming (Eldar and Mishali, 2009) BOMP: Block Orthogonal Matching Pursuit (Eldar et. al., 2010)

#### Performance Evaluation, cont'd

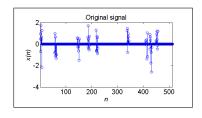
Average CPU time with M = N/2, K = M/2.5d, and d = 8 over 100 runs.

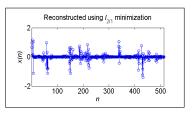


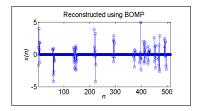
 $\ell_{2/p}\text{-RLS:}$  Proposed  $\ell_{2/p}\text{-Regularized}$  Least-Squares  $\ell_{2/1}\text{-SOCP:}$   $\ell_{2/1}$  Second-Order Cone-Programming (Eldar and Mishali, 2009) BOMP: Block Orthogonal Matching Pursuit (Eldar et. al., 2010)

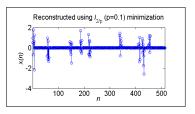
#### Performance Evaluation, cont'd

Example: N = 512, d = 8, K = 9, M = 100.









Compressive sensing is an effective sampling technique for sparse signals.

- Compressive sensing is an effective sampling technique for sparse signals.
- $\ell_1$ -minimization and  $\ell_p$ -minimization with p<1 work well for the reconstruction of sparse signals.

- Compressive sensing is an effective sampling technique for sparse signals.
- $\ell_1$ -minimization and  $\ell_p$ -minimization with p < 1 work well for the reconstruction of sparse signals.
- $\ \ \, = \ell_{2/1} \mbox{-minimization}$  offers improved reconstruction performance for block-sparse signals.

- Compressive sensing is an effective sampling technique for sparse signals.
- $\ell_1$ -minimization and  $\ell_p$ -minimization with p < 1 work well for the reconstruction of sparse signals.
- =  $\ell_{2/1}$ -minimization offers improved reconstruction performance for block-sparse signals.
- The proposed  $\ell_{2/p}$ -regularized least-squares algorithm offers improved reconstruction performance for block-sparse signals relative to the  $\ell_{2/1}$ -SOCP and BOMP algorithms.

#### Thank you for your attention.

This presentation can be downloaded from:

 $http://www.ece.uvic.ca/{\sim} and reas/RLectures/ISCAS2012\text{-}Jeevan\text{-}Pres.pdf$