Unconstrained Regularized ℓ_p-Norm Based Algorithm for the Reconstruction of Sparse Signals

J. K. Pant, W.-S. Lu, and A. Antoniou

University of Victoria

May 17, 2011

Outline

- Compressive Sensing

Outline

- Compressive Sensing
- Signal Recovery by Using ℓ_1 and ℓ_p Minimizations

Outline

- Compressive Sensing
- Signal Recovery by Using ℓ_1 and ℓ_p Minimizations
- Signal Recovery by Using Regularized ℓ_p Mininization

Outline

- Compressive Sensing
- Signal Recovery by Using ℓ_1 and ℓ_p Minimizations
- Signal Recovery by Using Regularized ℓ_p Mininization
- Line Search Based on Banach's Fixed-Point Theorem

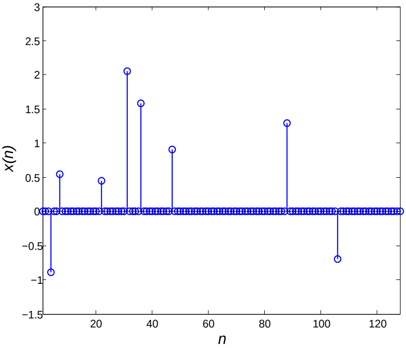
Outline

- Compressive Sensing
- Signal Recovery by Using ℓ_1 and ℓ_p Minimizations
- Signal Recovery by Using Regularized ℓ_p Mininization
- Line Search Based on Banach's Fixed-Point Theorem
- Performance Evaluation

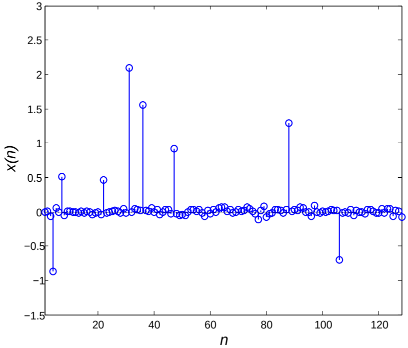
Compressive Sensing

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

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



A sparse signal

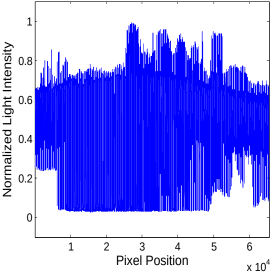


A near sparse signal

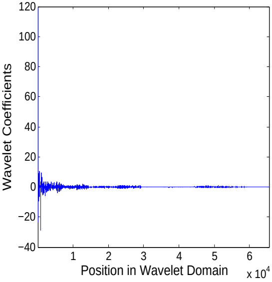
- Sparsity is a generic property of signals: A real-world signal always has a sparse or near-sparse representation with respect to an appropriate basis.

- Sparsity is a generic property of signals: A real-world signal always has a sparse or near-sparse representation with respect to an appropriate basis.

An Image



An equivalent 1-D signal



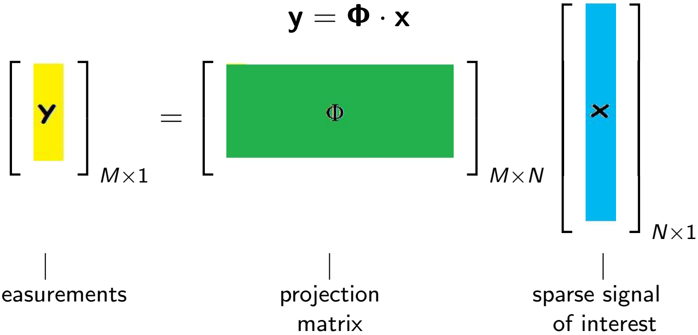
A wavelet representation of the image

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 (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 $M \times N$.
- In such a process, measurement vector $\mathbf{y}$ and signal vector $\mathbf{x}$ are interrelated by the equation

$$\mathbf{y} = \mathbf{\Phi} \cdot \mathbf{x}$$



Compressive Sensing, cont'd

- CS theory shows that these random projections contain much, sometimes all, the information content of signal $\mathbf{x}$.

- CS theory shows that these random projections contain much, sometimes all, the information content of signal $\mathbf{x}$.
- If a sufficient number of such measurements is collected, recovering signal $\mathbf{x}$ from measurements $\mathbf{y}$ is possible.

- CS theory shows that these random projections contain much, sometimes all, the information content of signal $\mathbf{x}$.
- If a sufficient number of such measurements is collected, recovering signal $\mathbf{x}$ from measurements $\mathbf{y}$ is possible.
- A condition for this to be possible is

$$M \geq c \cdot K \cdot \log(N/K)$$

where c is a small constant.

Compressive Sensing, cont'd

- CS theory shows that these random projections contain much, sometimes all, the information content of signal $\mathbf{x}$.
- If a sufficient number of such measurements is collected, recovering signal $\mathbf{x}$ from measurements $\mathbf{y}$ is possible.
- A condition for this to be possible is

$$M \geq c \cdot K \cdot \log(N/K)$$

  where c is a small constant.

- Typically,

$$K < M < N$$

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

$$\underset{M\times N}{\boldsymbol{\Phi}} \cdot \underset{N\times 1}{\mathbf{x}} = \underset{M\times 1}{\mathbf{y}}$$

is an ill-posed problem.

- The inverse problem of recovering signal vector $\mathbf{x}$ from measurement vector $\mathbf{y}$ such that

$$\underset{M\times N}{\boldsymbol{\Phi}} \cdot \underset{N\times 1}{\mathbf{x}} = \underset{M\times 1}{\mathbf{y}}$$

is an ill-posed problem.

- A classical approach for solving this problem is to find a vector $\mathbf{x}^*$ with minimum ℓ_2 norm in the translated null space of $\boldsymbol{\Phi}$ such that

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} ||\mathbf{x}||_2 \quad \text{subject to} \quad \boldsymbol{\Phi}\mathbf{x} = \mathbf{y}$$

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

$$\underset{M\times N}{\boldsymbol{\Phi}} \cdot \underset{N\times 1}{\mathbf{x}} = \underset{M\times 1}{\mathbf{y}}$$

  is an ill-posed problem.

- A classical approach for solving this problem is to find a vector $\mathbf{x}^*$ with minimum ℓ_2 norm in the translated null space of $\boldsymbol{\Phi}$ such that

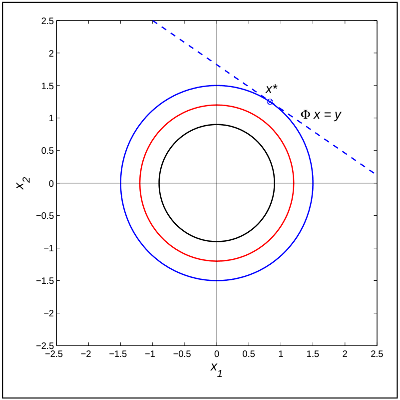
$$\mathbf{x}^* = \arg \min_{\mathbf{x}} ||\mathbf{x}||_2 \quad \text{subject to} \quad \boldsymbol{\Phi}\mathbf{x} = \mathbf{y}$$

- Unfortunately, the ℓ_2 minimization fails to recover a sparse signal.

Signal Recovery by Using ℓ_1 and ℓ_p Minimizations, cont'd

- Why ℓ_2-norm minimization fails to work?

- Why ℓ_2-norm minimization fails to work?



Contours of $||\mathbf{x}||_2 = r$

As r increases, the contour of $||\mathbf{x}||_2 = r$ grows and touches the hyperplane $\mathbf{\Phi x} = \mathbf{y}$.

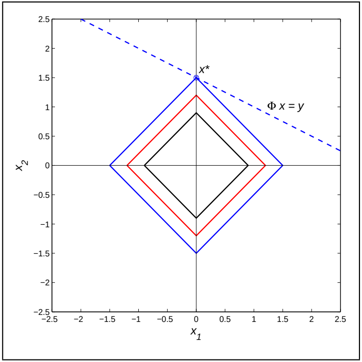
The solution $\mathbf{x}^*$ obtained is not sparse.

- A sparse signal, say $\mathbf{x}^*$, can be obtained by finding a vector with minimum ℓ_1 norm in the translated null space of $\mathbf{\Phi}$, i.e., using

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{subject to} \quad \mathbf{\Phi x} = \mathbf{y}$$

- A sparse signal, say $\mathbf{x}^*$, can be obtained by finding a vector with minimum ℓ_1 norm in the translated null space of $\mathbf{\Phi}$, i.e., using

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{subject to} \quad \mathbf{\Phi}\mathbf{x} = \mathbf{y}$$



Contours for $||\mathbf{x}||_1 = c$

As c increases, the contour of $||\mathbf{x}||_1 = c$ grows and touches the hyperplane $\mathbf{\Phi}\mathbf{x} = \mathbf{y}$, yielding a sparse solution

$$\mathbf{x}^* = \begin{bmatrix} 0 \\ c \end{bmatrix}$$

Theorem

If $\mathbf{\Phi} = \{\phi_{ij}\}$ where ϕ_{ij} are independent and identically distributed random variables with zero-mean and variance $1/N$ and $M \geq cK \log(N/K)$, the solution of the ℓ_1-minimization problem would recover exactly a K-sparse signal with high probability.

Theorem

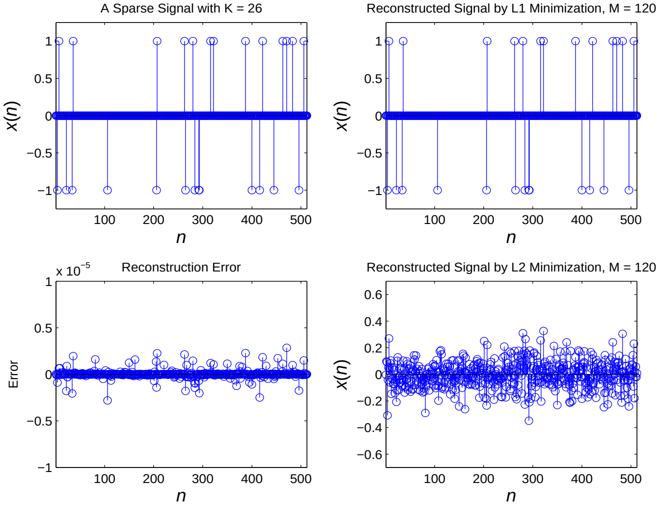
If $\mathbf{\Phi} = \{\phi_{ij}\}$ where ϕ_{ij} are independent and identically distributed random variables with zero-mean and variance $1/N$ and $M \geq cK \log(N/K)$, the solution of the ℓ_1-minimization problem would recover exactly a K-sparse signal with high probability.

- For real-valued data $\{\mathbf{\Phi}, \mathbf{y}\}$, the ℓ_1-minimization problem is a linear programming problem.

Signal Recovery by Using ℓ_1 and ℓ_p Minimizations, cont'd

- Example: $N = 512$, $M = 120$, $K = 26$

- Example: $N = 512$, $M = 120$, $K = 26$



- The sparsity of a signal can be measured by using its ℓ_0 pseudonorm

$$||\mathbf{x}||_0 = \sum_{i=1}^{N} |x_i|^0$$

Signal Recovery by Using ℓ_1 and ℓ_p Minimizations, cont'd

- The sparsity of a signal can be measured by using its ℓ_0 pseudonorm

$$||\mathbf{x}||_0 = \sum_{i=1}^{N} |x_i|^0$$

- Hence the sparsest solution of $\mathbf{\Phi x} = \mathbf{y}$ can be obtained by finding the vector $\mathbf{x}^*$ with the smallest value of the ℓ_0 pseudonorm in the translated null space of $\mathbf{\Phi}$, i.e.,

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} ||\mathbf{x}||_0 \quad \text{subject to } \mathbf{\Phi x} = \mathbf{y}$$

Signal Recovery by Using ℓ_1 and ℓ_p Minimizations, cont'd

- The sparsity of a signal can be measured by using its ℓ_0 pseudonorm

$$||\mathbf{x}||_0 = \sum_{i=1}^{N} |x_i|^0$$

- Hence the sparsest solution of $\mathbf{\Phi x} = \mathbf{y}$ can be obtained by finding the vector $\mathbf{x}^*$ with the smallest value of the ℓ_0 pseudonorm in the translated null space of $\mathbf{\Phi}$, i.e.,

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} ||\mathbf{x}||_0 \quad \text{subject to } \mathbf{\Phi x} = \mathbf{y}$$

- Unfortunately, the above ℓ_0-pseudonorm minimization problem is nonconvex with combinatorial complexity.

- An effective signal recovery strategy is to solve the ℓ_p-minimization problem

$$\begin{array}{llll} \underset{\mathbf{x}}{\text{minimize}} & ||\mathbf{x}||_p^p & \text{with} & 0 < p < 1 \\ \text{subject to} & & \boldsymbol{\Phi}\mathbf{x} = \mathbf{y} & \end{array}$$

where $||\mathbf{x}||_p^p = \sum_{i=1}^{N} |x_i|^p$.

Signal Recovery by Using ℓ_1 and ℓ_p Minimizations, cont'd

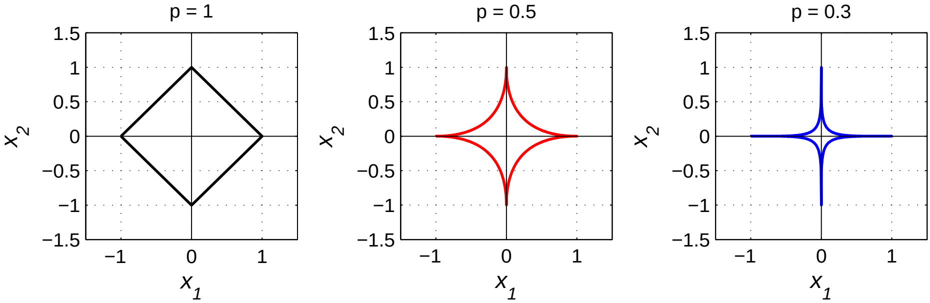
- An effective signal recovery strategy is to solve the ℓ_p-minimization problem

$$\begin{array}{lll} \underset{\mathbf{x}}{\text{minimize}} & ||\mathbf{x}||_p^p & \text{with} \quad 0 < p < 1 \\ \text{subject to} & & \mathbf{\Phi x} = \mathbf{y} \end{array}$$

where $||\mathbf{x}||_p^p = \sum_{i=1}^{N} |x_i|^p$.

- This ℓ_p-norm minimization problem is nonconvex.

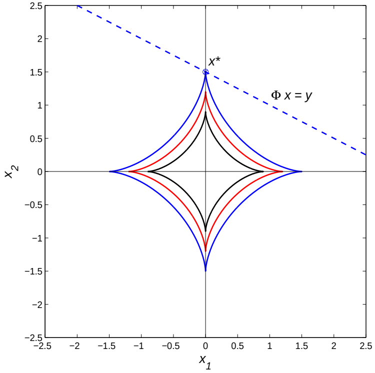
- Contours of $||\mathbf{x}||_p = 1$ with $p < 1$



$||\mathbf{x}||_1 = 1$ $\qquad$ $||\mathbf{x}||_{0.5} = 1$ $\qquad$ $||\mathbf{x}||_{0.3} = 1$

- Why ℓ_p minimization with $p < 1$?

- Why ℓ_p minimization with $p < 1$?



Contours of $||\mathbf{x}||_p^p = c$ with $p < 1$

As c increases, the contour $||\mathbf{x}||_p^p = c$ grows and touches the hyperplane $\mathbf{\Phi x} = \mathbf{y}$, yielding a sparse solution

$$\mathbf{x}^* = \begin{bmatrix} 0 \\ c \end{bmatrix}.$$

The possibility that the contour will touch the hyperplane at another point is eliminated.

- We propose to minimize a regularized ℓ_p norm

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

where $\mathbf{x}$ lies in the null space of $\boldsymbol{\Phi}$ translated by the ℓ_2-norm solution vector, say $\mathbf{x}_s$, of $\boldsymbol{\Phi}\mathbf{x} = \mathbf{y}$, namely,

$$\mathbf{x} = \mathbf{x}_s + \mathbf{V}_r \boldsymbol{\xi}$$

where $\mathbf{V}_r$ is an orthonormal basis of the null space of $\boldsymbol{\Phi}$.

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Note that as $\epsilon \longrightarrow 0$, we have

$$\left(x_i^2 + \epsilon^2\right)^{p/2} \approx |x_i|^p$$

Therefore,

$$||\mathbf{x}||_{p,\epsilon}^p\Big|_{\epsilon \longrightarrow 0} \approx ||\mathbf{x}||_p^p$$

i.e., the regularized ℓ_p norm closely approximates the ℓ_p norm.

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Note that as $\epsilon \longrightarrow 0$, we have

$$\left(x_i^2 + \epsilon^2\right)^{p/2} \approx |x_i|^p$$

Therefore,

$$||\mathbf{x}||_{p,\epsilon}^p\Big|_{\epsilon \longrightarrow 0} \approx ||\mathbf{x}||_p^p$$

i.e., the regularized ℓ_p norm closely approximates the ℓ_p norm.

- The reconstruction involves solving the optimization problem

$$\textbf{(P1)} \quad \underset{\boldsymbol{\xi}}{\text{minimize}} \sum_{i=1}^{n} \left\{ [x_s(i) + \mathbf{v}_i^T \boldsymbol{\xi}]^2 + \epsilon^2 \right\}^{p/2}$$

for a small value of ϵ.

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Optimization overview:

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Optimization overview:
  - Obtain an ℓ_2-norm solution $\mathbf{x}$, set $\boldsymbol{\xi} = \mathbf{0}$, and select an initial value of ϵ to satisfy the inequality

$$\epsilon \geq \sqrt{1-p} \cdot \underset{1 \leq i \leq N}{\text{maximum}} |x_{si}|$$

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Optimization overview:
  - Obtain an ℓ_2-norm solution $\mathbf{x}$, set $\boldsymbol{\xi} = \mathbf{0}$, and select an initial value of ϵ to satisfy the inequality

$$\epsilon \geq \sqrt{1-p} \cdot \underset{1 \leq i \leq N}{\text{maximum}} |x_{si}|$$

  - Using $\boldsymbol{\xi}$ as an initializer, solve the optimization problem **P1** using a quasi-Newton algorithm such as Broyden-Fletcher-Goldfarb-Shanno algorithm. Set the resulting solution to $\boldsymbol{\xi}$.

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Optimization overview:
  - Obtain an ℓ_2-norm solution $\mathbf{x}$, set $\boldsymbol{\xi} = \mathbf{0}$, and select an initial value of ϵ to satisfy the inequality

$$\epsilon \geq \sqrt{1-p} \cdot \underset{1 \leq i \leq N}{\text{maximum}} |x_{si}|$$

  - Using $\boldsymbol{\xi}$ as an initializer, solve the optimization problem **P1** using a quasi-Newton algorithm such as Broyden-Fletcher-Goldfarb-Shanno algorithm. Set the resulting solution to $\boldsymbol{\xi}$.
  - Reduce the value of ϵ, use $\boldsymbol{\xi}$ as an initializer, and solve problem **P1** again using the same quasi-Newton algorithm.

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Optimization overview:
  - Obtain an ℓ_2-norm solution $\mathbf{x}$, set $\boldsymbol{\xi} = \mathbf{0}$, and select an initial value of ϵ to satisfy the inequality

$$\epsilon \geq \sqrt{1-p} \cdot \underset{1 \leq i \leq N}{\text{maximum}} |x_{si}|$$

  - Using $\boldsymbol{\xi}$ as an initializer, solve the optimization problem **P1** using a quasi-Newton algorithm such as Broyden-Fletcher-Goldfarb-Shanno algorithm. Set the resulting solution to $\boldsymbol{\xi}$.
  - Reduce the value of ϵ, use $\boldsymbol{\xi}$ as an initializer, and solve problem **P1** again using the same quasi-Newton algorithm.
  - Repeat this procedure until problem **P1** is solved for a sufficiently small value of ϵ.

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Line Search Based on Banach's Fixed-Point Theorem:
  - The $(k+1)$th iterate is computed as

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \alpha \mathbf{d}_k$$

Signal Recovery by Using Regularized ℓ_p Minimization, cont'd

- Line Search Based on Banach's Fixed-Point Theorem:
  - The $(k+1)$th iterate is computed as

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \alpha \mathbf{d}_k$$

  - According to Banach's fixed-point theorem, the step size α can be computed using a finite number of iterations of

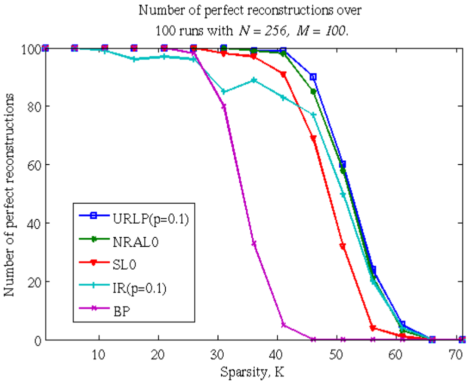
$$\alpha_{l+1} = -\frac{\sum_{i=1}^{N} x_i \cdot v_i \cdot \gamma_i(\alpha_l, \epsilon)^{p/2-1}}{\sum_{i=1}^{N} v_i^2 \cdot \gamma_i(\alpha_l, \epsilon)^{p/2-1}}$$

where

$$\gamma_i(\alpha_l, \epsilon) = (x_i + \alpha v_i)^2 + \epsilon^2, \quad x_i = x_{si} + \mathbf{v}_i^T \boldsymbol{\xi}_k, \quad v_i = \mathbf{v}_i^T \mathbf{d}_k$$

Performance Evaluation

Number of perfectly recovered instances versus sparsity K by various algorithms with $N = 256$ and $M = 100$ over 100 runs.



URLP: Proposed

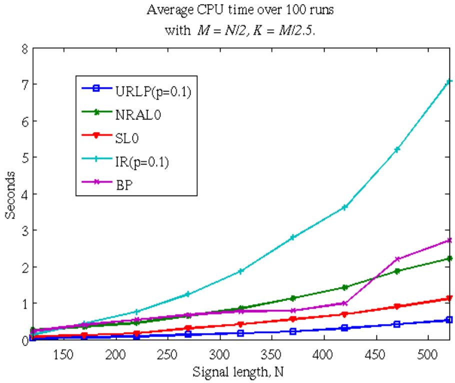
NRAL0: Null space re-weighted approximate ℓ_0 (Pant, Lu, and Antoniou, 2010)

SL0: Smoothed ℓ_0-norm minimization (Mohimani et. al., 2009)

IR: Iterative re-weighting (Chartrand and Yin, 2008)

Performance Evaluation, cont'd

Average CPU time versus signal length for various algorithms with $M = N/2$ and $K = M/2.5$.



URLP: Proposed

NRAL0: Null space re-weighted approximate ℓ_0 (Pant, Lu, and Antoniou, 2010)

SL0: Smoothed ℓ_0-norm minimization (Mohimani et. al., 2009)

IR: Iterative re-weighting (Chartrand and Yin, 2008)

Conclusions

- Compressive sensing is an effective technique for sampling sparse signals.

Conclusions

- Compressive sensing is an effective technique for sampling sparse signals.
- ℓ_1 minimization works in general for the reconstruction of sparse signals.

Conclusions

- Compressive sensing is an effective technique for sampling sparse signals.
- ℓ_1 minimization works in general for the reconstruction of sparse signals.
- ℓ_p minimization with $p < 1$ can improve the recovery performance for signals that are less sparse.

Conclusions

- Compressive sensing is an effective technique for sampling sparse signals.
- ℓ_1 minimization works in general for the reconstruction of sparse signals.
- ℓ_p minimization with $p < 1$ can improve the recovery performance for signals that are less sparse.
- Regularized ℓ_p minimization offers improved signal rconstruction performance.

Conclusions

- Compressive sensing is an effective technique for sampling sparse signals.
- ℓ_1 minimization works in general for the reconstruction of sparse signals.
- ℓ_p minimization with $p < 1$ can improve the recovery performance for signals that are less sparse.
- Regularized ℓ_p minimization offers improved signal rconstruction performance.
- A line search method based on Banach's fixed-point theorem offers improved complexity.

Thank you for your attention.

This presentation can be downloaded from:

http://www.ece.uvic.ca/∼andreas/RLectures/ISCAS2011-Jeevan-Web.pdf