Recovery of Sparse Signals from Noisy Measurements Using an ℓ_p-Regularized Least-Squares Algorithm

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

University of Victoria

August 25, 2011

Outline

- Compressive Sensing

Outline

- Compressive Sensing
- Use of ℓ_1 Minimization

Outline

- Compressive Sensing
- Use of ℓ_1 Minimization
- Use of ℓ_p Minimization

Outline

- Compressive Sensing
- Use of ℓ_1 Minimization
- Use of ℓ_p Minimization
- Use of ℓ_p-Regularized Least Squares

Outline

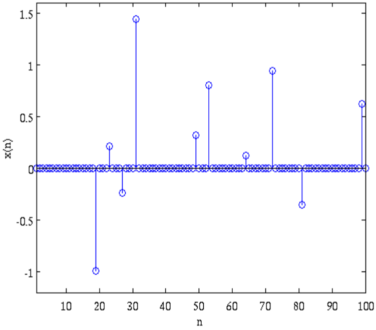
- Compressive Sensing
- Use of ℓ_1 Minimization
- Use of ℓ_p Minimization
- Use of ℓ_p-Regularized Least Squares
- Performance Evaluation

Outline

- Compressive Sensing
- Use of ℓ_1 Minimization
- Use of ℓ_p Minimization
- Use of ℓ_p-Regularized Least Squares
- Performance Evaluation
- Conclusion

- 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 sparse signal

An image with sparse gradient (Shepp-Logan phantom)

- A signal acquisition procedure in the compressive sensing framework is modelled as

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

where $\mathbf{\Phi}$ is a measurement matrix, typically with $K < M \ll N$, and $\mathbf{y}$ is measurement vector.

- A signal acquisition procedure in the compressive sensing framework is modelled as

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

where $\mathbf{\Phi}$ is a measurement matrix, typically with $K < M \ll N$, and $\mathbf{y}$ is measurement vector.

- The inverse problem of recovering signal vector $\mathbf{x}$ from $\mathbf{y}$ is an ill-posed problem.

Compressive Sensing, cont'd

- A signal acquisition procedure in the compressive sensing framework is modelled as

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

  where $\mathbf{\Phi}$ is a measurement matrix, typically with $K < M \ll N$, and $\mathbf{y}$ is measurement vector.

- The inverse problem of recovering signal vector $\mathbf{x}$ from $\mathbf{y}$ is an ill-posed problem.

- A classical approach to recover $\mathbf{x}$ from $\mathbf{y}$ is the method of least squares

$$\mathbf{x}^* = \mathbf{\Phi}^T \left(\mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1} \mathbf{y}$$

Compressive Sensing, cont'd

- Unfortunately, the method of least squares fails to recover a sparse signal.

Compressive Sensing, cont'd

- Unfortunately, the method of least squares fails to recover a sparse signal.

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

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

- Unfortunately, the method of least squares fails to recover a sparse signal.

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

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

- If signal $\mathbf{x}$ is known to be sparse, it can be estimated by solving the optimization problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & ||\mathbf{x}||_0 \\ \text{subject to} & \mathbf{\Phi x} = \mathbf{y} \end{array}$$

Use of ℓ_1 Minimization

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

Use of ℓ_1 Minimization

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

- Computationally tractable algorithms include the *basis pursuit* algorithm which solves the problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & ||\mathbf{x}||_1 \\ \text{subject to} & \mathbf{\Phi x} = \mathbf{y} \end{array}$$

where $||\mathbf{x}||_1 = \sum_{i=1}^{N} |x_i|$.

Use of ℓ_1 Minimization, cont'd

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.

Use of ℓ_1 Minimization, cont'd

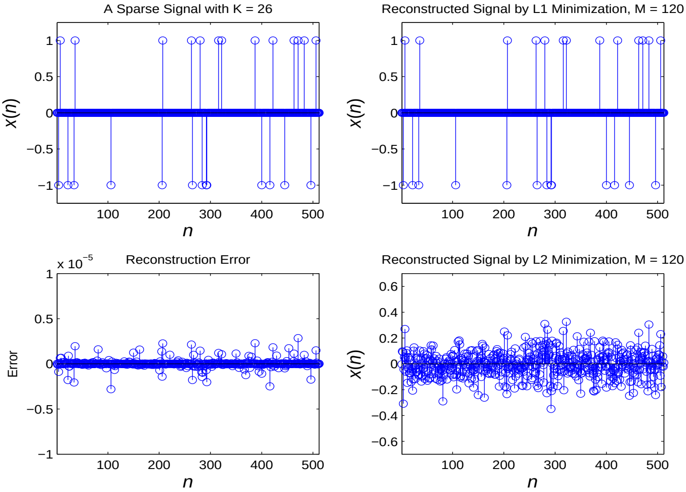
Theorem

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.

Use of ℓ_1 Minimization, cont'd

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



Use of ℓ_p Minimization

- Chartrand's ℓ_p-minimization based iteratively reweighted algorithm which solves the problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & ||\mathbf{x}||_p^p \\ \text{subject to} \quad & \mathbf{\Phi x} = \mathbf{y} \end{aligned}$$

where $||\mathbf{x}||_p^p = \sum_{i=1}^{N} |x_i|^p$ with $p < 1$ yields improved performance.

Use of ℓ_p Minimization

- Chartrand's ℓ_p-minimization based iteratively reweighted algorithm which solves the problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & ||\mathbf{x}||_p^p \\ \text{subject to} & \mathbf{\Phi x} = \mathbf{y} \end{array}$$

  where $||\mathbf{x}||_p^p = \sum_{i=1}^N |x_i|^p$ with $p < 1$ yields improved performance.

- Mohimani et al.'s smoothed ℓ_0-norm minimization algorithm solves the problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \sum_{i=1}^{N} \left[1 - \exp\left(-x_i^2/2\sigma^2\right)\right] \\ \text{subject to} & \mathbf{\Phi x} = \mathbf{y} \end{array}$$

  with $\sigma > 0$ using a sequential steepest-descent algorithm.

Use of ℓ_p Minimization, cont'd

- The unconstrained regularized ℓ_p norm minimization algorithm estimates signal $\mathbf{x}$ as

$$\mathbf{x}^* = \mathbf{x}_s + \mathbf{V}_n \boldsymbol{\xi}^*$$

where $\mathbf{x}_s$ is the least-squares solution of $\mathbf{\Phi x} = \mathbf{y}$, the columns of $\mathbf{V}_n$ constitue orthonormal basis of null space of $\mathbf{\Phi}$, and $\boldsymbol{\xi}^*$ is obtained as

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}}{\text{arg minimize}} \sum_{i=1}^{N} \left[\left(x_{si} + \mathbf{v}_i^T \boldsymbol{\xi} \right)^2 + \epsilon^2 \right]^{p/2-1}$$

Use of ℓ_p Minimization, cont'd

- The unconstrained regularized ℓ_p norm minimization algorithm estimates signal $\mathbf{x}$ as

$$\mathbf{x}^* = \mathbf{x}_s + \mathbf{V}_n \boldsymbol{\xi}^*$$

where $\mathbf{x}_s$ is the least-squares solution of $\mathbf{\Phi x} = \mathbf{y}$, the columns of $\mathbf{V}_n$ constitue orthonormal basis of null space of $\mathbf{\Phi}$, and $\boldsymbol{\xi}^*$ is obtained as

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}}{\text{arg minimize}} \quad \sum_{i=1}^{N} \left[\left(x_{si} + \mathbf{v}_i^T \boldsymbol{\xi} \right)^2 + \epsilon^2 \right]^{p/2-1}$$

- This algorithm finds a vector $\boldsymbol{\xi}^*$ that would give the sparsest estimate $\mathbf{x}^*$.

Proposed ℓ_p-Regularized Least-Squares Algorithm

- The proposed ℓ_p regularized least-squares algorithm is an extension of the unconstrained regularized ℓ_p algorithm for noisy measurement.

Proposed ℓ_p-Regularized Least-Squares Algorithm

- The proposed ℓ_p regularized least-squares algorithm is an extension of the unconstrained regularized ℓ_p algorithm for noisy measurement.

- For noisy measurements, the signal acquisition procedure is modelled as

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{w}$$

where $\mathbf{w}$ is the measurement noise.

Proposed ℓ_p-Regularized Least-Squares Algorithm

- The proposed ℓ_p regularized least-squares algorithm is an extension of the unconstrained regularized ℓ_p algorithm for noisy measurement.

- For noisy measurements, the signal acquisition procedure is modelled as

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{w}$$

  where $\mathbf{w}$ is the measurement noise.

- In such case, the equality condition $\mathbf{\Phi x} = \mathbf{y}$ should be relaxed to

$$||\mathbf{\Phi x} - \mathbf{y}||_2^2 \leq \delta$$

  where δ is a small positive scalar.

- Consider the optimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad ||\mathbf{x}||_{p,\epsilon}^{p} \quad \text{subject to} \quad ||\mathbf{\Phi x} - \mathbf{y}||_2^2 \leq \delta$$

where

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

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- Consider the optimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad ||\mathbf{x}||_{p,\epsilon}^{p} \quad \text{subject to} \quad ||\mathbf{\Phi x} - \mathbf{y}||_2^2 \leq \delta$$

  where

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

- An unconstrained formulation of the above problem is given by

$$\underset{\mathbf{x}}{\text{minimize}} \quad \tfrac{1}{2}\,||\mathbf{\Phi x} - \mathbf{y}||_2^2 + \lambda\,||\mathbf{x}||_{p,\epsilon}^{p}$$

  where λ is a regularization parameter.

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- Consider the optimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad ||\mathbf{x}||_{p,\epsilon}^p \quad \text{subject to} \quad ||\mathbf{\Phi x} - \mathbf{y}||_2^2 \leq \delta$$

  where

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

- An unconstrained formulation of the above problem is given by

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\,||\mathbf{\Phi x} - \mathbf{y}||_2^2 + \lambda\,||\mathbf{x}||_{p,\epsilon}^p$$

  where λ is a regularization parameter.

- We solve the above optimization problem in the null space of $\mathbf{\Phi}$ and its complement space.

- Let $\mathbf{\Phi} = \mathbf{U}\,[\mathbf{\Sigma}\ \mathbf{0}]\,\mathbf{V}^T$ be the singular-value decomposition of $\mathbf{\Phi}$.
  - $\mathbf{\Sigma}$ is a diagonal matrix whose diagonal elements $\sigma_1, \sigma_2, \ldots, \sigma_M$ are the singular values of $\mathbf{\Phi}$.
  - The columns of $\mathbf{U}$ and $\mathbf{V}$ are, respectively, the left and right singular vectors of $\mathbf{\Phi}$.
  - $\mathbf{V} = [\mathbf{V}_r\ \mathbf{V}_n]$ where $\mathbf{V}_r$ consists of the first M columns and $\mathbf{V}_n$ consists the remaining $N - M$ columns of $\mathbf{V}$.
  - The columns of $\mathbf{V}_n$ and $\mathbf{V}_r$ form orthogonal basis vectors for the null space of $\mathbf{\Phi}$ and its complement space, respectively.

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- Let $\mathbf{\Phi} = \mathbf{U}\,[\mathbf{\Sigma}\ \mathbf{0}]\,\mathbf{V}^T$ be the singular-value decomposition of $\mathbf{\Phi}$.
  - $\mathbf{\Sigma}$ is a diagonal matrix whose diagonal elements $\sigma_1, \sigma_2, \ldots, \sigma_M$ are the singular values of $\mathbf{\Phi}$.
  - The columns of $\mathbf{U}$ and $\mathbf{V}$ are, respectively, the left and right singular vectors of $\mathbf{\Phi}$.
  - $\mathbf{V} = [\mathbf{V}_r\ \mathbf{V}_n]$ where $\mathbf{V}_r$ consists of the first M columns and $\mathbf{V}_n$ consists the remaining $N - M$ columns of $\mathbf{V}$.
  - The columns of $\mathbf{V}_n$ and $\mathbf{V}_r$ form orthogonal basis vectors for the null space of $\mathbf{\Phi}$ and its complement space, respectively.

- Vector $\mathbf{x}$ is expressed as

$$\mathbf{x} = \mathbf{V}_r\phi + \mathbf{V}_n\xi$$

where ϕ and ξ are vectors of lengths M and $N - M$, respectively.

- Using the SVD, we recast the optimization problem for the ℓ_p-RLS algorithm as

$$\underset{\phi,\boldsymbol{\xi}}{\text{minimize}} \quad F_{p,\epsilon}(\phi, \boldsymbol{\xi})$$

where

$$F_{p,\epsilon}(\phi, \boldsymbol{\xi}) = \frac{1}{2}\sum_{i=1}^{M} (\sigma_i \phi_i - \tilde{y}_i)^2 + \lambda \, ||\mathbf{x}||_{p,\epsilon}^{p}$$

$\tilde{y}_i$ is the ith component of vector $\tilde{\mathbf{y}} = \mathbf{U}^T\mathbf{y}$, and $\mathbf{x} = \mathbf{V}_r\phi + \mathbf{V}_n\boldsymbol{\xi}$.

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- In the kth iteration of the ℓ_p-RLS algorithm, signal $\mathbf{x}^{(k)}$ is updated as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha \mathbf{d}_v^{(k)}$$

where

$$\mathbf{d}_v^{(k)} = \mathbf{V}_r \mathbf{d}_r^{(k)} + \mathbf{V}_n \mathbf{d}_n^{(k)}$$

and $\alpha > 0$.

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- In the kth iteration of the ℓ_p-RLS algorithm, signal $\mathbf{x}^{(k)}$ is updated as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha \mathbf{d}_v^{(k)}$$

where

$$\mathbf{d}_v^{(k)} = \mathbf{V}_r \mathbf{d}_r^{(k)} + \mathbf{V}_n \mathbf{d}_n^{(k)}$$

and $\alpha > 0$.

- Vectors $\mathbf{d}_r^{(k)}$ and $\mathbf{d}_n^{(k)}$ are of lengths M and $N - M$, respectively.

- In the kth iteration of the ℓ_p-RLS algorithm, signal $\mathbf{x}^{(k)}$ is updated as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha \mathbf{d}_v^{(k)}$$

where

$$\mathbf{d}_v^{(k)} = \mathbf{V}_r \mathbf{d}_r^{(k)} + \mathbf{V}_n \mathbf{d}_n^{(k)}$$

and $\alpha > 0$.

- Vectors $\mathbf{d}_r^{(k)}$ and $\mathbf{d}_n^{(k)}$ are of lengths M and $N - M$, respectively.
- Each component of vectors $\mathbf{d}_r^{(k)}$ and $\mathbf{d}_n^{(k)}$ is efficiently computed using *the first step* of a fixed-point iteration.

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

$$\alpha = -\frac{\mathbf{d}_r^{(k)^T} \mathbf{\Sigma} \left(\mathbf{\Sigma} \phi - \tilde{y} \right) + \lambda \cdot p \cdot \mathbf{x}^{(k)^T} \boldsymbol{\zeta}_v}{||\mathbf{\Sigma} \mathbf{d}_r^{(k)}||_2^2 + \lambda \cdot p \cdot \mathbf{d}_v^{(k)^T} \boldsymbol{\zeta}_v}$$

where

$$\boldsymbol{\zeta}_v = [\zeta_{v1} \ \zeta_{v2} \ \cdots \ \zeta_{vN}]^T$$

with

$$\zeta_{vj} = \left[\left(x_j^{(k)} + \alpha d_{vj}^{(k)} \right)^2 + \epsilon^2 \right]^{p/2-1} d_{vj}^{(k)}$$

for $j = 1, 2, \ldots, N$ where $d_{vj}^{(k)}$ is the jth component of the descent direction $\mathbf{d}_v^{(k)}$.

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- Optimization overview:
  - First, set ϵ to a large value, say, ϵ_1, typically $0.5 \leq \epsilon_1 \leq 1$, and initialize $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ to zero vectors.

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- Optimization overview:
  - First, set ϵ to a large value, say, ϵ_1, typically $0.5 \leq \epsilon_1 \leq 1$, and initialize $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ to zero vectors.
  - Solve the optimization problem by i) computing descent directions $\mathbf{d}_v$ and $\mathbf{d}_r$, ii) computing the step size α; and iii) updating solution $\mathbf{x}$ and coefficient vector $\boldsymbol{\phi}$.

- Optimization overview:
  - First, set ϵ to a large value, say, ϵ_1, typically $0.5 \leq \epsilon_1 \leq 1$, and initialize $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ to zero vectors.
  - Solve the optimization problem by i) computing descent directions $\mathbf{d}_v$ and $\mathbf{d}_r$, ii) computing the step size α; and iii) updating solution $\mathbf{x}$ and coefficient vector $\boldsymbol{\phi}$.
  - Reduce ϵ to a smaller value and again solve the optimization problem.

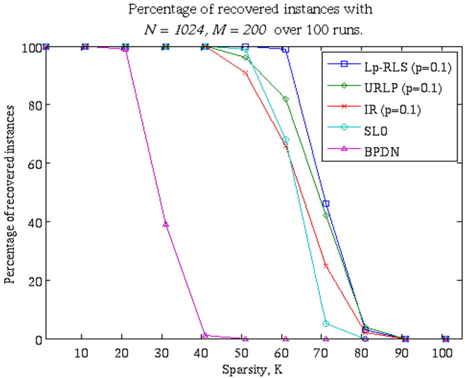
- Optimization overview:
  - First, set ϵ to a large value, say, ϵ_1, typically $0.5 \le \epsilon_1 \le 1$, and initialize $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ to zero vectors.
  - Solve the optimization problem by i) computing descent directions $\mathbf{d}_v$ and $\mathbf{d}_r$, ii) computing the step size α; and iii) updating solution $\mathbf{x}$ and coefficient vector $\boldsymbol{\phi}$.
  - Reduce ϵ to a smaller value and again solve the optimization problem.
  - Repeat this procedure until a sufficiently small target value, say, ϵ_J is reached.

Proposed ℓ_p-Regularized Least-Squares Algorithm, cont'd

- Optimization overview:
  - First, set ϵ to a large value, say, ϵ_1, typically $0.5 \leq \epsilon_1 \leq 1$, and initialize $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ to zero vectors.
  - Solve the optimization problem by i) computing descent directions $\mathbf{d}_v$ and $\mathbf{d}_r$, ii) computing the step size α; and iii) updating solution $\mathbf{x}$ and coefficient vector $\boldsymbol{\phi}$.
  - Reduce ϵ to a smaller value and again solve the optimization problem.
  - Repeat this procedure until a sufficiently small target value, say, ϵ_J is reached.
  - Output $\mathbf{x}$ as the solution.

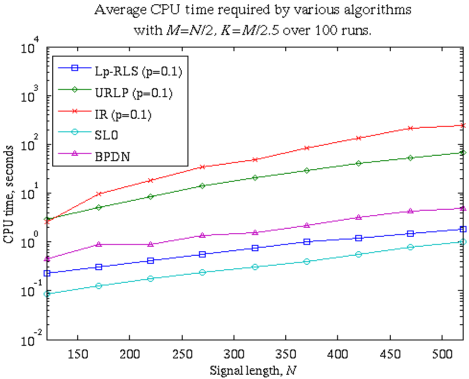
Performance Evaluation

Number of recovered instances versus sparsity K by various algorithms with $N = 1024$ and $M = 200$ over 100 runs.



ℓ_p-RLS: Proposed
URLP: Unconstrained regularized ℓ_p (Pant, Lu, and Antoniou, 2011)
SL0: Smoothed ℓ_0-norm minimization (Mohimani et al., 2009)
IR: Iterative re-weighting (Chartrand and Yin, 2008)

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



ℓ_p-RLS: Proposed
URLP: Unconstrained regularized ℓ_p (Pant, Lu, and Antoniou, 2011)
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.
- Proposed ℓ_p-regularized least-squares offers improved signal reconstruction from noisy measurements.

Thank you for your attention.