Robust Reconstruction Algorithms for Compressive Imaging

Rafael E. Carrillo

http://people.epfl.ch/rafael.carrillo

Signal Processing Laboratory LTS2
Department of Electrical Engineering
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

BASP Frontiers September 5 2011

Outline

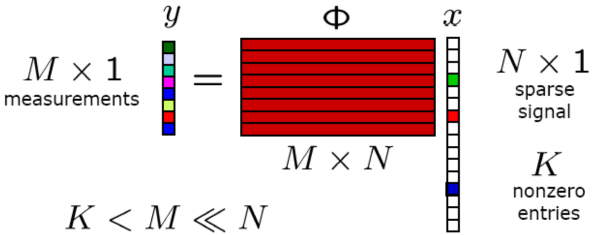
1. Introduction and Problem Set-up
2. Lorentzian Iterative Hard Thresholding
   - Lorentzian Based IHT Algorithm
   - Numerical Experiments
3. Robust Bayesian CS
   - Bayesian CS Using Generalized Cauchy Models
   - Numerical Experiments
4. Summary

Introduction

- When signals are sparse/compressible, directly acquire compressed data

Introduction

- When signals are sparse/compressible, directly acquire compressed data
- Replace samples by more general measurements



- Recover original signal from y

Introduction

- Compressive sensing systems are not immune to noise, which is always present in practical acquisition systems

Introduction

- Compressive sensing systems are not immune to noise, which is always present in practical acquisition systems
- Most CS systems treat all noise contributions as sampling noise,

$$y = \Phi(x_0 + w) + z = \Phi x_0 + r$$

Introduction

- Compressive sensing systems are not immune to noise, which is always present in practical acquisition systems
- Most CS systems treat all noise contributions as sampling noise,

$$y = \Phi(x_0 + w) + z = \Phi x_0 + r$$

- What if the measurements are corrupted by sparse or impulsive noise?

Introduction

- Compressive sensing systems are not immune to noise, which is always present in practical acquisition systems
- Most CS systems treat all noise contributions as sampling noise,

$$y = \Phi(x_0 + w) + z = \Phi x_0 + r$$

- What if the measurements are corrupted by sparse or impulsive noise?
- Traditional noise aware CS systems consider finite variance noise models

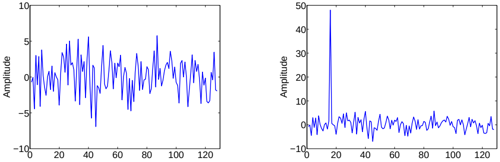
Introduction

- Compressive sensing systems are not immune to noise, which is always present in practical acquisition systems
- Most CS systems treat all noise contributions as sampling noise,

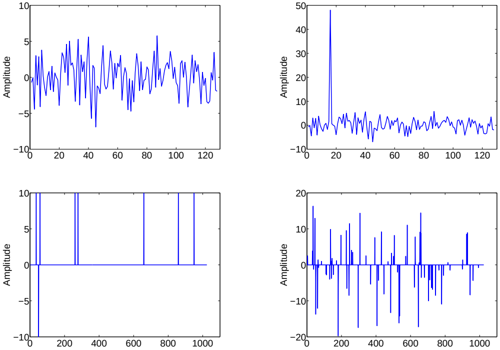
$$y = \Phi(x_0 + w) + z = \Phi x_0 + r$$

- What if the measurements are corrupted by sparse or impulsive noise?
- Traditional noise aware CS systems consider finite variance noise models
- Impulsive noise has infinite or very large variance breaking the assumptions of traditional LS-based recovery algorithms

Impulsive Noise Effects



Impulsive Noise Effects



Motivation

Develop simple robust reconstruction strategies that render faithful reconstruction of sparse signals in impulsive environments.

Solution: Robust Estimation Theory

- The generalized Cauchy distribution family was introduced by Rider in 1957
- The GCD PDF is

$$f(x) = a\delta(\delta^p + |x|^p)^{-2/p}, \; 0 < p \leq 2$$

with $a = \frac{p\Gamma(2/p)}{2(\Gamma(1/p))^2}$

- For $p = 2$ we have the Cauchy distribution
- For $p = 1$ we have the Meridian distribution

Solution: Robust Estimation Theory

LL_p Norm

$$\|u\|_{LL_{p},\gamma} = \sum_{i=1}^{m} \log\left\{1 + \frac{u_i^p}{\gamma^p}\right\}, \quad \gamma > 0, \; u \in \mathbb{R}^m.$$

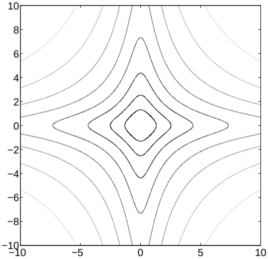
Solution: Robust Estimation Theory

LL_p Norm

$$\|u\|_{LL_{p,\gamma}} = \sum_{i=1}^{m} \log\left\{1 + \frac{u_i^p}{\gamma^p}\right\}, \quad \gamma > 0, \ u \in \mathbb{R}^m.$$

Lorentzian norm: particular case when $p = 2$

- It is an everywhere continuous function
- It is convex near the origin
- Large deviations are not heavily penalized as in the case of L_1 or L_2



Previous Work: Lorentzian Basis Pursuit

Let Φ be an $m \times n$ sensing matrix such that $\delta_{2s} < \sqrt{2} - 1$. Then for any signal $x_0 \in \mathbb{R}^n$ such that $|T_0| \leq s$, where $T_0 = supp(x_0)$, and observation noise z with $\|z\|_{LL_{2,\gamma}} \leq \epsilon$, the solution to

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ s.t. } \|y - \Phi x\|_{LL_{2,\gamma}} \leq \epsilon,$$

x^*, obeys the following bound

$$\|x^* - x_0\|_2 \leq C_s \cdot \gamma \cdot \sqrt{m(e^{\epsilon} - 1)},$$

where the constant C_s depends only on δ_{2s}.

Previous Work: Lorentzian Basis Pursuit

Let Φ be an $m \times n$ sensing matrix such that $\delta_{2s} < \sqrt{2} - 1$. Then for any signal $x_0 \in \mathbb{R}^n$ such that $|T_0| \leq s$, where $T_0 = supp(x_0)$, and observation noise z with $\|z\|_{LL_{2,\gamma}} \leq \epsilon$, the solution to

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ s.t. } \|y - \Phi x\|_{LL_{2,\gamma}} \leq \epsilon,$$

x^*, obeys the following bound

$$\|x^* - x_0\|_2 \leq C_s \cdot \gamma \cdot \sqrt{m(e^{\epsilon} - 1)},$$

where the constant C_s depends only on δ_{2s}.

Problem: Slow and complex to solve!

Outline

1 Introduction and Problem Set-up

2 Lorentzian Iterative Hard Thresholding
- Lorentzian Based IHT Algorithm
- Numerical Experiments

3 Robust Bayesian CS
- Bayesian CS Using Generalized Cauchy Models
- Numerical Experiments

4 Summary

Lorentzian Based IHT Algorithm

Ideal optimization problem:

$$\min_{x \in \mathbb{R}^n} \|y - \Phi x\|_{LL_{2},\gamma} \quad \text{s.t.} \quad \|x\|_0 \leq s$$

Lorentzian Based IHT Algorithm

Ideal optimization problem:

$$\min_{x\in\mathbb{R}^n} \|y - \Phi x\|_{LL_{2,\gamma}} \text{ s.t. } \|x\|_0 \leq s$$

Iterative algorithm:

$$x^{(t+1)} = H_s\left(x^{(t)} - \mu^{(t)} g^{(t)}\right)$$

where

$$g^{(t)} = \nabla_x \|y - \Phi x^{(t)}\|_{LL_{2,\gamma}}$$

and $H_s(a)$ is the non-linear operator that sets all but the largest (in magnitude) s elements of a to zero.

Lorentzian Based IHT Algorithm Formulation

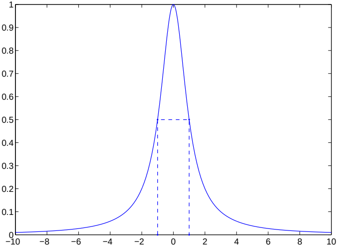
The gradient can be expressed as

$$g^{(t)} = -\Phi^T W_{(t)}(y - \Phi x^{(t)})$$

where

$$W_{(t),ii} = \frac{\gamma^2}{\gamma^2 + e_i^2}$$

with $e^{(t)} = y - \Phi x^{(t)}$



Lorentzian Based IHT Algorithm Formulation

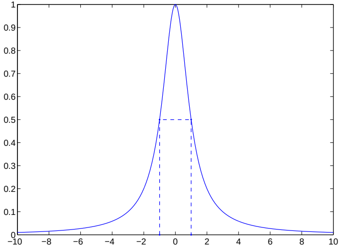
The gradient can be expressed as

$$g^{(t)} = -\Phi^T W_{(t)}(y - \Phi x^{(t)})$$

where

$$W_{(t),ii} = \frac{\gamma^2}{\gamma^2 + e_i^2}$$

with $e^{(t)} = y - \Phi x^{(t)}$



LIHT Algorithm

$$x^{(t+1)} = H_s\left(x^{(t)} + \mu^{(t)}\Phi^T W_{(t)}(y - \Phi x^{(t)})\right)$$

Computational complexity per iteration: $O(mn)$

Stability of LIHT Algorithm

Theorem

Let $x \in \mathbb{R}^n$ be a s-sparse signal. Suppose $\Phi \in \mathbb{R}^{m \times n}$ meets the RIP of order 3s. Then if $\|z\|_{LL_{2,\gamma}} \leq \epsilon$ and $\delta_{3s} < 1/2$, the reconstruction error of the LIHT algorithm at iteration t is bounded by

$$\|x - x^{(t)}\|_2 \leq \alpha^t \|x\|_2 + \beta\gamma\sqrt{m(e^{\epsilon} - 1)},$$

where

$$\alpha = 2\delta_{3s} \quad \text{and} \quad \beta = \sqrt{1 + \delta_{2s}} \left(\frac{1 - \alpha^t}{1 - \alpha} \right).$$

Parameter Tuning

- $\hat{\gamma} = (Q^y_{(0.875)} - Q^y_{(0.125)})/2$
- Considers a measurement vector with 25% of the samples corrupted by outliers and 75% well behaved

Parameter Tuning

- $\hat{\gamma} = (Q^{y}_{(0.875)} - Q^{y}_{(0.125)})/2$
- Considers a measurement vector with 25% of the samples corrupted by outliers and 75% well behaved
- Define $S = supp(x^{(t)})$ and assume support does not change. Step size:

$$\mu^{(t)} = \min_{\mu} \left\| W_{(t)}^{1/2} \left[y - \Phi_S(x_S^{(t)} + \mu g_S^{(t)}) \right] \right\|_2^2$$
$$= \frac{\|g_S^{(t)}\|_2^2}{\|W_{(t)}^{1/2} \Phi_S g_S^{(t)}\|_2^2}$$

- It guarantees that

$$\|y - \Phi_S x_S^{(t+1)}\|_{LL_2,\gamma} \leq \|y - \Phi_S x_S^{(t)}\|_{LL_2,\gamma}$$

Outline

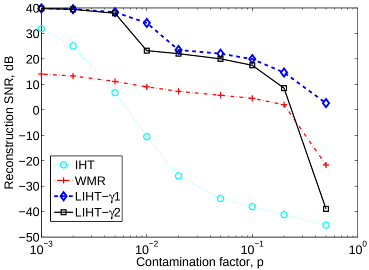
1. Introduction and Problem Set-up
2. Lorentzian Iterative Hard Thresholding
   - Lorentzian Based IHT Algorithm
   - Numerical Experiments
3. Robust Bayesian CS
   - Bayesian CS Using Generalized Cauchy Models
   - Numerical Experiments
4. Summary

Experimental Set-up

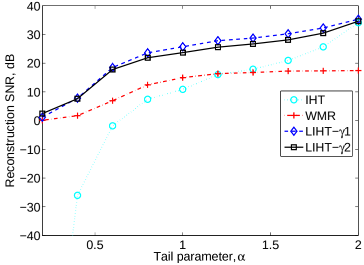
- Sparse signals with $n = 1024$, $s = 8$ and $P_{av} = 0.78$ in a Hadamard basis
- Nonzero coefficients drawn from a Rademacher distribution. Position randomly chosen
- Number of random measurements set to $m = 128$. Gaussian measurement matrices
- Contaminated Gaussian and α–stable noise models
- 1000 repetitions of each experiment averaged

Robustness Against Impulsive Noise

Cont. Gaussian, $\sigma^2 = 0.01$



α–stable noise, $\sigma = 0.1$



Performance varying the number of samples

α–stable noise, $\sigma = 0.1$

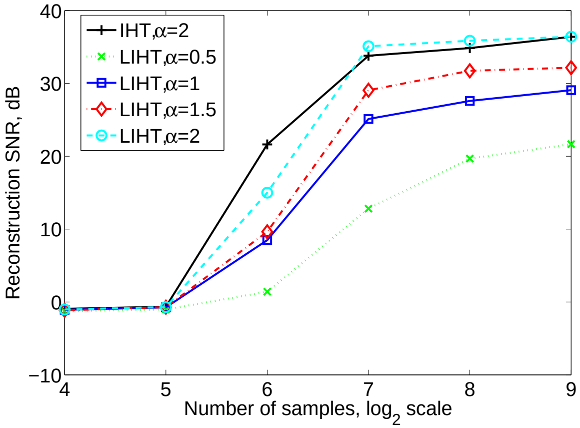


Image Reconstruction Example (I)

256×256 Lena image

Cauchy corrupted measurements

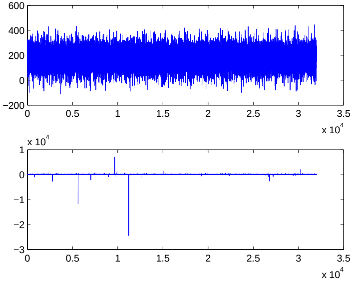
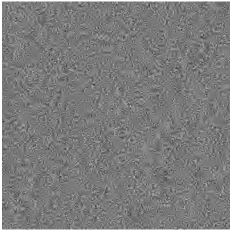


Image Reconstruction Example (II)

LS-IHT,
R-SNR=-10.7 dB



LS-IHT with clipping,
R-SNR=6.2 dB

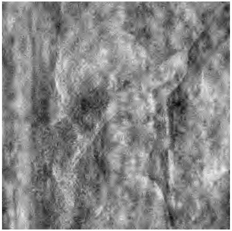


Image Reconstruction Example (III)

LS-IHT noiseless,
R-SNR=23.9 dB

LIHT,
R-SNR=20.5 dB

Outline

1. Introduction and Problem Set-up
2. Lorentzian Iterative Hard Thresholding
   - Lorentzian Based IHT Algorithm
   - Numerical Experiments
3. Robust Bayesian CS
   - Bayesian CS Using Generalized Cauchy Models
   - Numerical Experiments
4. Summary

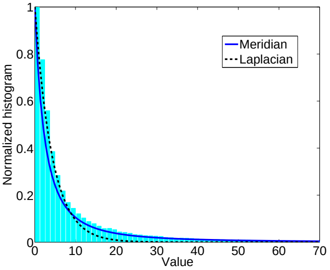
Sparse and Compressible Models

- Compressible signals are modeled in an L_p ball

$$\|x\|_p \leq R$$

- Order statistics obey power law decay

$$x_{(i)} \leq R \cdot i^{-1/p}$$



- Algebraic-tailed distributions are more suitable models than exponential-tailed distributions

Bayesian Formulation

Bayesian modeling: All unknowns treated as stochastic quantities with assigned distributions.

Bayesian Formulation

Bayesian modeling: All unknowns treated as stochastic quantities with assigned distributions.

Sampling model:

$$y = \Phi x + z.$$

Maximum a posteriori (MAP) estimate:

$$\max_{x \in \mathbb{R}^n} p(x|y,\Gamma) = \max_{x \in \mathbb{R}^n} p(y|x,\Gamma) p(x|\Gamma)$$

- $p(y|x,\Gamma)$: likelihood of the samples
- $p(x|\Gamma)$: prior of x

Bayesian CS Using Generalized Cauchy Models

- Model for sparse and compressible signals: algebraic tailed priors

$$p(x|\delta, p) = (a\delta)^n \prod_{i=1}^{n} (\delta^p + |x_i|^p)^{-2/p}$$

Bayesian CS Using Generalized Cauchy Models

- Model for sparse and compressible signals: algebraic tailed priors

$$p(x|\delta, p) = (a\delta)^n \prod_{i=1}^{n} (\delta^p + |x_i|^p)^{-2/p}$$

- Assume independent GCD distribution for the noise, with tail parameter q. Likelihood function:

$$p(y|x, \sigma, q) = (a\sigma)^m \prod_{i=1}^{m} (\sigma^q + |y_i - \theta_i|^q)^{-2/q}$$

with location vector $\theta = \Phi x$

MAP Estimate

The MAP estimate assuming a Cauchy distribution for the noise ($q = 2$), σ, δ and p known, is:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \|y - \Phi x\|_{LL_2,\sigma} + 2\|x\|_{LL_p,\delta}$$

MAP Estimate

The MAP estimate assuming a Cauchy distribution for the noise ($q = 2$), σ, δ and p known, is:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \|y - \Phi x\|_{LL_2,\sigma} + 2\|x\|_{LL_p,\delta}$$

- Easy to formulate in a RWLS framework

MAP Estimate

The MAP estimate assuming a Cauchy distribution for the noise ($q = 2$), σ, δ and p known, is:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \|y - \Phi x\|_{LL_2,\sigma} + 2\|x\|_{LL_p,\delta}$$

- Easy to formulate in a RWLS framework
- Parameter estimation: noise scale parameter, $\hat{\sigma} = 0.5(Q^y_{(0.875)} - Q^y_{(0.125)})$

MAP Estimate

The MAP estimate assuming a Cauchy distribution for the noise ($q = 2$), σ, δ and p known, is:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \|y - \Phi x\|_{LL_2,\sigma} + 2\|x\|_{LL_p,\delta}$$

- Easy to formulate in a RWLS framework
- Parameter estimation: noise scale parameter, $\hat{\sigma} = 0.5(Q^y_{(0.875)} - Q^y_{(0.125)})$
- Prior parameter estimation: EM approach
  - Prior tail parameter, $\hat{p}_t \equiv$ML estimate
  - Prior scale parameter, $\hat{\delta}_t = 0.5(Q^x_{(0.75)} - Q^x_{(0.25)})$

Fixed Point Formulation

Update equation:

$$x^{(t+1)} = [\Phi^T\Phi + 2W]^{-1}\Phi^T Hy$$
$$= W^{-1}\Phi^T[H\Phi W^{-1}\Phi^T + 2I]^{-1}Hy.$$

where

$$W_{ii} = \frac{1}{(\delta^p + |x_i|^p)|x_i|},$$

$$H_{ii} = \frac{\sigma^2}{\sigma^2 + e_i^2}, \quad e^{(t)} = y - \Phi x^{(t)}.$$

GCBCS Algorithm

Require σ, δ_{min}, γ and J.
Initialize $t = 0$ and $\hat{x}_0 = \Phi^T(\Phi\Phi^T + I)^{-1}Hy$.
While $\|\hat{x}_t - \hat{x}_{t-1}\|_2 > \gamma$ or $t < J$

1. Update $\hat{\delta}_t$, $\hat{p}_t$ W and H.
2. Compute $\hat{x}_{t+1}$.
3. $t \leftarrow t + 1$

Return $\hat{x}$

GCBCS Algorithm

Require σ, δ_{min}, γ and J.
Initialize $t = 0$ and $\hat{x}_0 = \Phi^T(\Phi\Phi^T + I)^{-1}Hy$.
While $\|\hat{x}_t - \hat{x}_{t-1}\|_2 > \gamma$ or $t < J$

1. Update $\hat{\delta}_t$, $\hat{p}_t$ W and H.
2. Compute $\hat{x}_{t+1}$.
3. $t \leftarrow t + 1$

Return $\hat{x}$

- Computational complexity per iteration: $O(n^3)$
- Bottleneck: matrix inversion

Outline

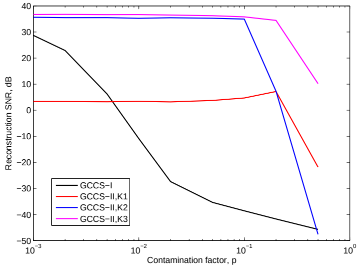
1. Introduction and Problem Set-up
2. Lorentzian Iterative Hard Thresholding
   - Lorentzian Based IHT Algorithm
   - Numerical Experiments
3. Robust Bayesian CS
   - Bayesian CS Using Generalized Cauchy Models
   - Numerical Experiments
4. Summary

Experimental Set-up

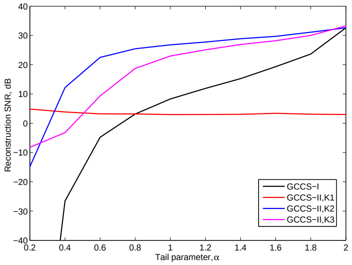
- Sparse signals with $n = 1024$, $s = 8$ and $P_{av} = 0.78$ in a Hadamard basis
- Nonzero coefficients drawn from a Rademacher distribution. Position randomly chosen
- Number of random measurements set to $m = 128$. Gaussian measurement matrices
- Contaminated Gaussian and α–stable noise models
- 1000 repetitions of each experiment averaged

Robustness Against Impulsive Noise

Cont. Gaussian, $\sigma^2 = 0.01$



α–stable noise, $\sigma = 0.1$



Performance varying the number of samples

α–stable noise, $\sigma = 0.1$

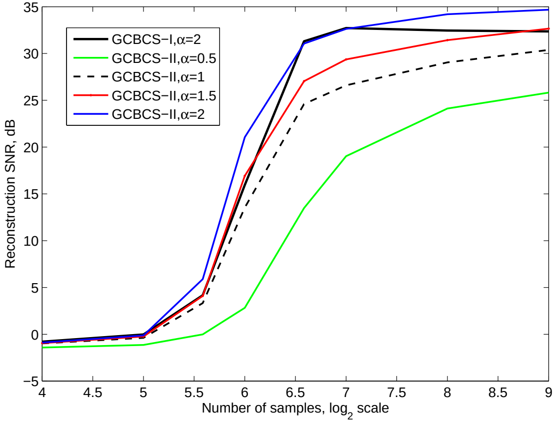


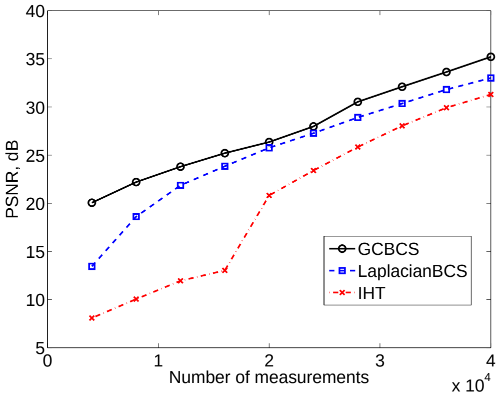
Image Reconstruction Example (I)

256×256 Lena image

- Left column: Laplacian BCS
- Right column: GCBCS
- Top: $m = 8000$.
  L: $PSNR = 18.61$ dB,
  GC: $PSNR = 23.81$ dB
- Middle: $m = 20000$.
  L: $PSNR = 25.56$ dB,
  GC: $PSNR = 26.36$ dB
- Bottom: $m = 32000$.
  L: $PSNR = 30.36$ dB,
  GC: $PSNR = 32.10$ dB

Image Reconstruction Example (II)

Average results for 10 different images, $n = 256 \times 256 = 65536$



Summary

- Robust reconstruction algorithms are presented based on the GC models
- Performance and properties of the algorithms are investigated in heavy and light tail environments
- LIHT and GCBCS outperform LS-CS methods in heavy–tailed noise while providing comparable performance in light–tailed environments
- Future work:
  1. Development of non i.i.d. algebraic models for sparse and compressible signals
  2. Inclusion of prior information like support knowledge or signal model