IRLS Based Algorithm for Signal Recovery from Noisy Compressive Measurements

Qin Li Ji Shen Advisor: Gordon Erlebacher
Applied Math, Florida State University, Tallahassee
Email: qli@math.fsu.edu jshen@math.fsu.edu
Florida State University

Abstract

Compressive Sensing (CS) is a novel theory states that sparse signals can be recovered from only a few measurements far below that dictated by the Nyquist sampling theorem. To recover the signal, one needs to find the sparse solution to an under-determined linear system. When the measurements are contaminated by noise, an unconstrained minimization problem $(P_{0,\lambda})$ is often considered. Since it is NP hard to solve, we try to solve an approximation of $(P_{0,\lambda})$, $(P_{p,\lambda})$ with $0 < p \le 1$, using an Iterative Re-weighted Least Square (IRLS) algorithm. We discuss the convergence properties of this algorithm and provide several numerical results. Our algorithm is robust in the presence of noise and capable of recovering signals that are less sparse than possible with the best alternate approaches.

Introduction

A. Overview of Compressive Sensing

- Many signals are compressible: sparse/can be sparsely represented by a proper dictionary.
- CS theory: one can often recover compressible signals from far fewer measurements than traditional methods.
- CS applications: compressive imaging, medical imaging, geophysical data analysis, etc.

B. Statement of the Problem

Suppose $N < M$, $\mathbf{c} \in \mathrm{R}^{\mathrm{M}}$ is the signal to be recovered, $\mathbf{f} \in \mathrm{R}^{\mathrm{N}}$ is the measurement vector and $D \in \mathrm{R}^{\mathrm{M\times N}}$ is the sampling matrix, usually a random matrix. In this poster, D is a random matrix generated from $\mathcal{N}(0,1)$. To recover $\mathbf{c}$ from $\mathbf{f}$, one must solve the following,

$$(P_0) \qquad \min \| \mathbf{c} \|_0, \quad \text{subject to} \quad D\mathbf{c} = \mathbf{f}$$

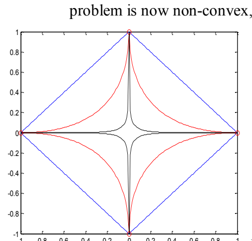
In this problem, sparseness is measured by the ℓ_0 norm of the signal. Weaker forms of measurements are ℓ_1, ℓ_p norms with $0 < p < 1$. See Fig 1. In practice, noise usually exists during the sampling process so the following unconstrained minimization is often considered,

$$(P_{0,\lambda}) \qquad \min \lambda \| \mathbf{c} \|_0 + \frac{1}{2} \| \mathbf{f} - D\mathbf{c} \|_2^2$$

The Lagrange multiplier λ controls the trade off between the sparsity and quality of fit.

C. Existent Recovery Algorithms

- Greedy Algorithms: GA iteratively solves a set of approximation problems. OMP and its improved versions have been proved to converge to a global optimal solution under some strict conditions. However, they are efficient only in recovering very sparse signals.
- BP/BPDN: These algorithms replace the ℓ_0 norm by ℓ_1 norm and solve the resulting convex problems. Algorithms include Linear Programming algorithms and iterative thresholding methods, etc. LP algorithms are robust and stable but computational burdensome. Iterative thresholding algorithms are faster but cannot recover less sparse signals.
- Non-convex Techniques: These algorithms replace the ℓ_0 norm by ℓ_p norm with $0 < p < 1$. The advantage is that the resulting problem is closer to $(P_{0,\lambda})$ than $(P_{1,\lambda})$. Intuitively, they can recover less sparse signals than the previous two. Due to the fact that the problem is now non-convex, one cannot avoid converging to local minimum.



Definition of the norms $\|\cdot\|_0, \|\cdot\|_1, \|\cdot\|_p, 0 < p < 1$

$\| \mathbf{c} \|_0 =$ # of nonzero entries in $\mathbf{c}$

$$\| \mathbf{c} \|_1 = \sum_{i=1}^{M} |c_i|$$

$$\| \mathbf{c} \|_p = \left(\sum_{i=1}^{M} |c_i|^p \right)^{\frac{1}{p}}$$

Fig 1. Unit-balls of different norms. Blue: ℓ_1 norm; red: $\ell_{0.5}$ norm; black: $\ell_{0.25}$ norm. Four red circles: ℓ_0 norm.

IRLS based Algorithm

A. IRLS-p Algorithm

Since it is NP hard to solve $(P_{0,\lambda})$, we try to solve the following instead,

$$(P_{p,\lambda}) \qquad \min \lambda \| \mathbf{c} \|_p + \frac{1}{2} \| \mathbf{f} - D\mathbf{c} \|_2^2, \quad 0 < p \le 1$$

Introduce the following functional

$$\mathcal{T}_p(\mathbf{c}, \mathbf{w}, \epsilon) := \frac{\lambda}{2} \left[\sum_i^{\mathrm{N}} c_i^2 w_i + \epsilon^2 w_i + \left(\frac{2}{p} - 1\right) w_i^{-\frac{p}{2-p}} \right] + \frac{1}{2} \| \mathbf{f} - D\mathbf{c} \|_2^2$$

where w is the weight vector and ϵ is a regularization parameter. We update c, w alternatively to ensure that the value of the functional decreases. Check Listing 1 for formal description. The parameter ϵ is initially set to $\epsilon_0 = 1$. When the sequence c^n converges, $\epsilon_{n+1} = \frac{\epsilon_n}{10}$. Otherwise, it remains constant.

Algorithm: IRLS-*p*

A. Initialize $\mathbf{c}^0$, $\epsilon_0 > 0$, calculate $w_i^0 = ((c_i^0)^2 + \epsilon_0^2)^{\frac{2-p}{p}}$
B. At iteration n
Update $\mathrm{c}^{n+1} = \mathrm{argmin}\, \mathcal{T}_p(\mathbf{c}, \mathbf{w}^n, \epsilon_n)$ by solving
$(\lambda p W^n + D^T D)\mathbf{c}^{n+1} = D^T \mathbf{f}, W^n = diag\{w_i^n\}$
Evaluate ϵ_{n+1} such that $\epsilon_{n+1} \le \epsilon_n$ (as previously described)
Update $\mathbf{w}^{n+1} = \mathrm{argmin}\, \mathcal{T}_p(\mathbf{c}^{n+1}, \mathbf{w}, \epsilon_n)$
$w_i^{n+1} = ((c_i^{n+1})^2 + \epsilon_{n+1}^2)^{-\frac{2-p}{2}}$
C. Terminate the algorithm when $\mathbf{c}^n$ converges and ϵ_n reaches its minimal

Listing 1. IRLS-p for approximating $(P_{p,\lambda})$

B. Convergence of the algorithm

Given $f^\epsilon(\mathbf{c})$ and $g^\epsilon(\mathbf{c})$ as the followings, we proved convergence properties as stated in Table 1:

$$f^\epsilon(\mathbf{c}) = \lambda \sum_i (c_j^2 + \epsilon^2)^{\frac{1}{2}} + \frac{1}{2} \| \mathbf{f} - D\mathbf{c} \|_2^2$$

$$g^\epsilon(\mathbf{c}) = \lambda \sum_i (c_j^2 + \epsilon^2)^{\frac{p}{2}} + \frac{1}{2} \| \mathbf{f} - D\mathbf{c} \|_2^2$$

These two functions are good approximations to $(P_{p,\lambda})$. As $p \to 0$, g^ϵ approximates $(P_{0,\lambda})$ better. Intuitively, we expect our algorithm with $p < 1$ to recover less sparse signals than greedy algorithms and BP methods. In addition, we have proved that when $p = 1$, $\| \mathbf{c}^\epsilon - \mathbf{c}^\star \|_2^2 \le 2AM\epsilon$, where $\mathrm{c}^\star$ minimizes $(P_{1,\lambda})$.

	$p = 1$	$0 < p < 1$
Convergence ($\lim \epsilon_n = \epsilon > 0$)	Converges to c^ϵ, a global minimizer of $f^\epsilon(\mathbf{c})$	Converges to c^ϵ, a local minimizer of $g^\epsilon(\mathbf{c})$
Advantage	Global minimum	Better chance to recover solution with less sparsity; Converges faster when p is smaller

Table 1. Convergence properties of IRLS-p. Comparison between the cases $p = 1$ and less.

Numerical Results

A. Exact Recovery from Noise-Free Measurements

We conducted numerical experiments to evaluate algorithms performance to recover the original signal from noise-free measurements. We want to test whether perfect recovery is possible from a low number of samples. To this end, we set $\lambda = 10^{-8}$. Let the length of $\mathbf{c}$ be 512 and the number of samplings be 256 and generate a random signal $\mathbf{c}$ with sparsity 128, i.e., 128 entries are non-zero. The value for p is set to 0.01, 0.5, 0.75, 1. For 20 different cases, the signals are almost perfectly reconstructed. However, for larger p, more iterations are required for convergence and the error is larger.

Remark: This signal is not very sparse. GA cannot recover it and BP only reach the accuracy of IRLS-1.

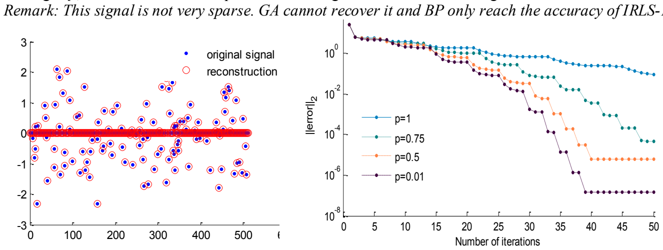


Fig 2. Left: Recovery of an 128-sparse signal. Right: Decreasing of $\| \mathbf{c}^{\text{recovered}} - \mathbf{c}^{\text{exact}} \|_2^2$ for different p.

B. Recovery from Noisy Measurements

We try to illustrate the stable approximation when $\mathbf{f}$ is noisy. Let $\mathbf{f} = D\mathbf{c} + \mathbf{n}$, $\mathbf{n}$ is Gaussian noise. The result of one experiment is illustrated in Fig 3. The length of the measurement vector is 128, and is contaminated with noise of variance 0.03. If we still require $\mathbf{f} = D\mathbf{c}$, (let λ be small enough), the recovered signal is mildly affected by the noise. Setting λ properly leads to better recovery.

We then fixed the number of measurements to be 128, let the sparsity of the signal range from 2 to 50 (where exact recovery is possible for the noise-free case), and evaluate the SNR of the recovered signal for the output of $\lambda = 10^{-8}$ and $\lambda = 0.02$ respectively (noise variance remains 0.03 for each experiment). The enhancement of SNR is illustrated in Fig 4. We also calculated the noise reduction, which is not shown here. The noise reduction is about 80% for very sparse signals and almost linearly decreases to 20% for denser signals. For all experiments, we use IRLS-0.5.

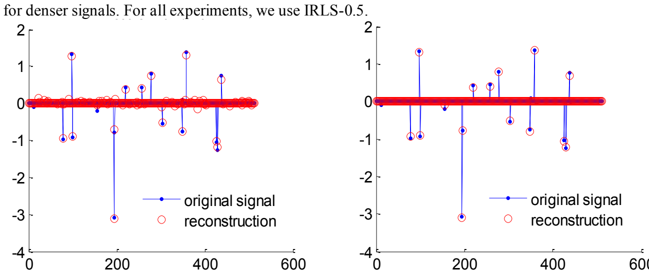


Fig 3. Left: Recovered assuming no noise exists. Right: Recovered with λ chosen properly.

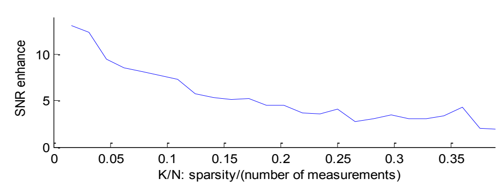


Fig 4. Horizontal axis: Ratio between sparsity and number of measurements. Vertical axis: SNR enhancement. For a fixed sparsity, 20 cases are evaluated and averaged.

Discussion

We applied an IRLS-based algorithm to the CS problem. It outperforms other algorithms in recovering less sparse signals. It is also robust in the presence of noise. Our algorithm is slower than GA and iterative thresholding algorithms. To the best of our knowledge, we are the first to provide convergence analysis for IRLS based algorithm solving the unconstrained minimization in CS. A related work of Daubechies discusses a similar algorithm that solves the constrained minimization (P_0), which is not suitable to recover signals from noisy samplings.

Acknowledgement

We thank the National Science Foundation for support under grant NSF-0417850