Open Access

Non-Convex Compressed Sensing from Noisy Measurements

A. Majumdar* and R. K. Ward

Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada

Abstract: This paper proposes solution to the following non-convex optimization problem:

$\min \| x \|_p$ subject to $\| y - Ax \|_q \leq \varepsilon$

Such an optimization problem arises in a rapidly advancing branch of signal processing called 'Compressed Sensing' (CS). The problem of CS is to reconstruct a k-sparse vector x_{nX1}, from noisy measurements $y = Ax + \eta$, where A_{mXn} (m<n) is the measurement matrix and η_{mX1} is additive noise.

In general the optimization methods developed for CS minimizes a sparsity promoting l_1-norm (p=1) for Gaussian noise (q=2). This is restrictive for two reasons: i) theoretically it has been shown that, with positive fractional norms (0<p<1), the sparse vector x can be reconstructed by fewer measurements than required by l_1-norm; and ii) Noises other than Gaussian require the norm of the misfit (q) to be something other than 2. To address these two issues an Iterative Reweighted Least Squares based algorithm is proposed here to solve the aforesaid optimization problem.

Keywords: Compressed Sensing, Non-convex optimization, Non-Gaussian noise.

1. INTRODUCTION

Finding sparse solution to a system of under-determined equation is of considerable interest to the rapidly growing branch of signal processing called Compressed Sensing (CS). It has widespread applications in blind source separation, source coding, imaging, channel estimation etc. In a typical CS setting, the problem is to reconstruct a k-sparse signal 'x' of length n, (a vector of k non-zeroes and n-k zeroes). This signal is not sampled directly but is measured *via* a measurement matrix A (size mXn, m<n); and a measurement vector 'y' of length m is obtained.

$$y_{m\times 1} = A_{m\times n} x_{n\times 1} \quad (1)$$

In any practical situation, the measured signal is corrupted by noise, so a more practical model for the measurement process is,

$$y = Ax + \eta, \text{ where } \eta \text{ is the noise} \quad (2)$$

The problem of CS is to reconstruct x from noisy under-sampled measurements y.

If the measurement matrix satisfies certain properties, it is possible to reconstruct the k-sparse signal by solving the following problem [1],

$$\min \| x \|_0 \text{ subject to } \| y - Ax \|_2 \leq \varepsilon \quad (3)$$

In (3) it is assumed that the noise is Gaussian, and therefore the l_2-norm for the misfit is used.

The k-sparse signal can be recovered by solving (3) from only $m = 2k + 1$ measurements [2]. However, solving (3) is known to be an NP hard problem. To achieve a practical solution, a convex surrogate of the NP hard problem is solved instead

$$\min \| x \|_1 \text{ subject to } \| y - Ax \|_2 \leq \varepsilon \quad (4)$$

It has been proved that under certain conditions, the solutions (3) and (4) are equivalent [3]. This is good because, it allows solution of (4) *via* well known quadratic programming methods instead of solving an NP hard problem (3). But, sparse signal recovery *via* solving the l_1-norm minimization problem requires more measurements $m = Ck \log \frac{n}{k}$ [1].

In practical situations, the cost of acquiring the measurements is proportional to the number of measurements (m); therefore the challenge is to reconstruct the signal with the least number of measurements possible. Theoretically the least number of measurements required for reconstruction is $m = 2k + 1$, however this will require solution of an NP hard problem (3) – which is not practical. It is shown in [1], that by minimizing a fractional norm p (0<p<1) it is possible reconstruct the k-sparse signal from $m = C_1 k + pC_2 k \log \frac{n}{k}$ measurements. One can directly see that when the value of p is small, the contribution due to the second factor ($pC_2 k \log \frac{n}{k}$) is negligible and the number of required measurements is only proportional to the sparsity of the vector ($C_1 k$). Based on this theoretical result, several previous studies proposed methods for solving the signal

*Address correspondence to this author at the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada; E-mail: angshulm@ece.ubc.ca

reconstruction problem for the noiseless case *via* the following optimization:

$$\min \| x \|_p \text{ subject to } y = Ax, 0 < p \le 1 \quad (5)$$

Generally gradient descent [2] or iterative least squares [1, 4 and 5] is used to solve (5).

In this work we will consider the problem of sparse signal reconstruction under noisy conditions. Moreover our method will consider any additive noise and will not be restricted to the Gaussian noise model. Therefore, we propose to solve the following optimization problem,

$$\min \| x \|_p \text{ subject to } \| y - Ax \|_q \le \varepsilon \quad (6)$$

where p and q can take any positive value.

We follow an Iterative Reweighted Least Squares (IRLS) methodology for solving (6). The rest of the paper will consist of several sections. Section 2 discusses the main algorithm. The experimental results are described in section 3. The conclusions are discussed in section 4.

2. THE OPTIMIZATION PROBLEM

As mentioned earlier, the signal model is given by (2)

$$y = Ax + \eta$$

The problem is to obtain x by solving the optimization problem

$$\hat{x} = \min \| x \|_p^p \text{ subject to } \| y - Ax \|_q^q \le \varepsilon \quad (7)$$

The form in (7) is equivalent to (6) since the minimizer for $\| . \|_m$ is the same as of $\| . \|_m^m$. The first term $\| x \|_p^p$ is the modeling term and the second term $\| y - Ax \|_q^q$ is the misfit.

The above constrained form can be converted to the following unconstrained form

$$\hat{x} = \min \frac{1}{p} \| x \|_p^p + \frac{\lambda}{q} \| y - Ax \|_q^q \quad (8)$$

The constant λ is related to ε. But the relationship between λ and ε can not be found analytically. In many situations the term ε is not even known beforehand; consequently trying to find λ is meaningless. In this paper, we propose a regularization method which will not require specification of λ and ε.

2.1. Convergence

Critics of non-convex optimization are apprehensive stating that the solution to such problems does not reach a global minima. The statement is true in general but not for our problem; since this problem is quasi-convex. And for quasi-convex problems the local minima and the global minima are the same.

The misfit term $\| y - Ax \|_q^q$ is convex, since the value of 'q' is always an integer. We will show that the modeling term $\| x \|_p^p$ is quasi-convex.

A function $f : R^n \to R'$ is said to be quasi-convex if it satisfies the following relationship for $\beta \in (0,1)$

$$f(\beta x + (1-\beta) y) \le \max[f(x), f(y)]$$

The above relationship is satisfied by the modeling term.

$$\| \beta x + (1-\beta) y \|_p^p = \sum_i (\beta x_i + (1-\beta) y_i)^p$$

$$\text{let } z_i = \max[x_i, y_i]$$

$$\therefore \; \| \beta x + (1-\beta) y \|_p^p \le \sum_i (\beta z_i + (1-\beta) z_i)^p$$

$$\le \sum_i (z_i)^p = \max[\| x \|_p^p, \| y \|_p^p]$$

As our optimization problem is quasi-convex it is guaranteed to reach a global minima.

2.2. Modeling Term

In the IRLS method the modeling term (l_p-norm) is approximated by a weighted l_2-norm.

$$\frac{1}{p} \| x \|_p^p \approx \frac{1}{2} \| W_m x \|_2^2 \quad (9)$$

The weight matrix is updated at each iteration. The weight matrix at iteration t is given by

$$W_m(t) = diag(\frac{2}{p} | x(t-1) |^{(p/2-1)}) \quad (10)$$

When the solution converges (assuming that it does), we have $x(t-1) = x(t)$. Therefore,

$$\frac{1}{2} \| W_m x(t) \|_2^2 = \frac{1}{2} \frac{2}{p} \sum_i x_i(t)^{p/2-1} x_i(t)$$

$$i.e \; \frac{1}{p} \sum_i x_i(t)^p = \frac{1}{p} \| x(t) \|_p^p$$

The choice behind the particular weight matrix is now clear. It is based on the idea that, when the solution reaches convergence, the weighted l_2-norm behaves as a near perfect approximation of the original l_p-norm.

When the coefficients in x become zeroes (which is expected since x is sparse) the corresponding diagonal elements of the weight matrix approach infinity. To avoid such a situation, the weight matrix is perturbed slightly, so that

$$W_m(t) = diag(\frac{2}{p} | x(t-1) + \delta(t) |^{(p/2-1)}) \quad (11)$$

The perturbation δ binds the elements of the weight matrix. At each iteration, the perturbation δ is reduced so that when the solution converges, the weighted l_2-norm is a good approximation of the desired l_p-norm. The idea of perturbing the IRLS was proposed in [1]. It showed that the perturbed method was significantly better than the unperturbed ones such as [5].

Other methods (based on thresholding) to bind the values of the weight matrix have been proposed in [6] (the problem in [6] was to solve the problem of Total Variation minimization; it is related to ours but not exactly the same). However such methods are in general not convergent, while ours is. The convergence of perturbed IRLS have been proved in [7].

2.3. Misfit

The l_q-norm misfit is manipulated in the same way as the modeling term. It is approximated by a weighted l_2-norm.

$$\frac{1}{q}\| y-Ax\|_q^q \approx \frac{1}{2}\| W_f(y-Ax)\|_2^2 \tag{12}$$

At each iteration, the actual misfit is approximated in the way shown in (12). The weight matrix is updated at each iteration. It has the form,

$$W_f(t)=diag(\frac{2}{q}|y-Ax(t-1)|^{(q/2-1)}) \tag{13}$$

It can be easily verified, that this form of the weight matrix leads to perfect approximation of the desired l_q-norm when the solution converges.

The elements of the weight matrix needs to be bounded. Therefore, the weight matrix is slightly perturbed by a factor δ. The perturbation is reduced at each iteration, so that it is negligible when the solution converges. The perturbed weight matrix takes the form,

$$W_f(t)=diag(\frac{2}{q}|y-Ax(t-1)+\delta(t)|^{(q/2-1)}) \tag{14}$$

Such reweighted schemes for approximating the l_q-norm misfit have been employed earlier [6]. The earlier work employed a thresholding scheme to bound the weight matrix. Our method of perturbing the misfit iteratively showed faster convergence. There are other advantages of using the perturbation method compared to thresholding which we will explain later.

2.4. General Algorithm

The sole purpose of this work is to solve the problem (8). With the approximations made in sections 2.1 and 2.2, (8) can be represented (approximately) in the following form

$$\hat{x}=\min\frac{1}{2}\|W_m x\|_2^2+\frac{\lambda}{2}\|W_f(y-Ax)\|_2^2 \tag{15}$$

Alternatively (15) can be expressed in the form

$$\hat{x}=\min\frac{1}{2}\left\|\begin{pmatrix}W_f & 0\\ 0 & W_m\end{pmatrix}\left(\begin{pmatrix}A\\ \lambda I\end{pmatrix}x-\begin{pmatrix}y\\ 0\end{pmatrix}\right)\right\|_2^2 \tag{16}$$

The closed form solution to (16) is

$$\hat{x}=(\lambda W_m^T W_m + A^T W_f^T W_f A)^{-1}A^T W_f^T W_f y \tag{17}$$

It is possible to apply (17) iteratively by updating the weight matrices till the solution converges. In fact this is the solution proposed in [6] for the problem of Total Variation minimization. Applying (17) directly for our purpose has one major drawback – choosing the regularization parameter λ.

We set forth to solve (7). There is a relation between the amount of misfit ε in (7) and the regularization parameter λ of (8). In general there is no analytical relation between these two. To make things worse, in many practical situations even the misfit ε is not known. Studies that develop algorithms dependent on regularization parameter mostly guess/tune its value. In this work, we propose a solution that does not require such parameter tuning.

With a change of variable, $u=W_m^{-1}x$ (this is just a rescaling, the sparsity of original variable x and the transformed one u is the same), the closed form solution becomes,

$$\hat{u}=(\lambda I+W_m^T A^T W_f^T W_f A W_m)^{-1}W_m^T A^T W_f^T W_f y \tag{18}$$

The usual way to solve it is by conjugate gradient method. (18) can be solved very efficiently by making $\lambda=0$ and control the number of iterations of the conjugate gradient (CG) algorithm for regularization [8].

The number of CG iterations is to be controlled based on Global Cross Validation (GCV). Following [8], the following approximate GCV criterion is used to control the number of iterations for the CG algorithm.

$$GCV(t)=\frac{\|Ax(t)-y\|_2^2}{(N-t)^2} \tag{19}$$

where N is the length of the vector x. The iterations (t) are to be stopped when the GCV criterion reaches a minimum.

The GCV criterion is very cheap to compute, it only requires dividing the current residual by a scalar. How the GCV acts as a regularizer can be intuitively understood. The numerator (misfit) reduces at each CG iteration so that the solution is a better fit for the observed data. But at every iteration the denominator in (19) reduces as well, so that the reduction in misfit is penalized. For a theoretical understanding, the reader is asked to read [9].

Our discussion behind the algorithm is complete. The complete algorithm is expressed tersely in the following pseudo code.

Initialization – set δ(0) = 0 and $\hat{x}(0)=\min\|y-Ax\|_2^2$ solved by regularized CG.

At iteration t – continue the following steps till convergence (i.e. either δ is less than 10^{-6} or the number of iterations has reached maximum limit)

1. Find the current modeling and misfit weight matrices from equations (11) and (14).

   a. $W_m(t)=diag(\frac{2}{p}|x(t-1)+\delta(t)|^{(p/2-1)})$

   b. $W_f(t)=diag(\frac{2}{q}|y-Ax(t-1)+\delta(t)|^{(q/2-1)})$

2. Form a new matrix as required by (18), $L = W_f A W_m$.

3. Scale y as required by (18), $t = W_d y$.

4. Solve $\hat{u}(t) = \min \| t - Lu \|_2^2$ *via* regularized CG.

5. Find x by rescaling u, $x(t) = W_m u(t)$.

6. Reduce δ by a factor of 10 if $||y\text{-}Ax||_q$ has reduced.

This algorithm has several desirable features. First, it is applicable to explicit matrices as well as fast operators. Second, it is non-parametric – it does not require specifying parameters like λ or thresholding hyper-parameters as in [6]. Third, the algorithm is guaranteed to converge (may be to a local minima owing to its con-convexity), since it follows IRLS methodology whose convergence has been proved in [7].

The computational cost of the algorithm is dominated at each step by step 4, the least squares solution. We are employing a regularized version of the Conjugate Gradient method to solve this problem. At most 'k' CG iterations are run. For each CG iteration the computational cost is $O(n^2)$ for explicit matrices or $O(n\log n)$ for fast operators. Therefore the dominating cost at each step is $O(kn^2)$ or $O(n\log n)$ as the case may be.

3. EXPERIMENTAL EVALUATION

The experiments are performed on synthetic data. A sparse signal of length 150 with only 25 non-zero coefficients is generated at random. It is measured by a measurement matrix A of size 100X150, which, in our case, is an i.i.d. Gaussian matrix with its columns normalized to unity. The measurement vector (length 100) is corrupted by additive noise. In all the experiments the value of the sparsity promoting norm ($||.||_p$) was fixed at 0.6 as the best results were obtained at this value. Simulations were carried out 10,000 times for all the experimental configurations. The results shown are the averaged over all simulations.

It was theoretically shown in [2] that the results from the following two optimizations are nearly the same for Ga ussian noise.

$$\min \| x \|_p \text{ subject to } y = Ax \quad (20a)$$

$$\min \| x \|_p \text{ subject to } \| y - Ax \|_2 \le \varepsilon \quad (20b)$$

This implies that solution of the sparse reconstruction problem by the proposed method (20b) will be the same as the solution proposed in [1]. In the first experiment we will show that, practically such is not the case. Reconstruction errors from our proposed method is significantly less compared to [1]. The normalized mean squared errors between the original (noiseless) signal and the reconstrucred one for different values of noise variance are tabulated in Table **1**.

Results from Table **1** validate that the optimization method that incorporates noisyness of the data (20b) yields better values compared to the optimization which treats the data to be noiseless (20a). Results from (20b) show better performance when the data becomes progressively more noisy.

Table 1. NMSE for Gaussian Noise

Optimization Constraint	σ^2 value for Gaussian Noise			
	0.05	**0.1**	**0.25**	**0.5**
y=Ax	0.027	0.030	0.037	0.054
$\|\|y\text{-}Ax\|\|_2<\varepsilon$	0.019	0.020	0.026	0.039

In the following experiment we will add non-Gaussian noise to the data. In such cases the $||.||_2$ for the misfit is not optimal. Tables **2** and 3 show the results for Poisson noise and Impulse noise respectively. The best results for Poisson noise are obtained when the norm of the misfit is 3 and for Impulse noise the corresponding value is 6.

Table 2. NMSE for Poisson Noise

Optimization Constraint	NMSE
y=Ax	0.124
$\|\|y\text{-}Ax\|\|_2<\varepsilon$	0.107
$\|\|y\text{-}Ax\|\|_3<\varepsilon$	0.057

Table 3. NMSE for Impulse Noise

Optimization Constraint	Fraction of coefficients affected			
	0.01	**0.05**	**0.1**	**0.25**
y=Ax	0.050	0.052	0.052	0.055
$\|\|y\text{-}Ax\|\|_2<\varepsilon$	0.048	0.049	0.050	0.053
$\|\|y\text{-}Ax\|\|_6<\varepsilon$	0.036	0.039	0.040	0.042

Tables **2** and **3** show that the ability to vary the norm of the misfit helps in better denoising. By varying the norm according to the type of noise, it is possible to obtain significantly better results. For Poisson noise the NMSE improved by around 40% and for Impulse noise the improvement is around 20%.

4. CONCLUSION

Compressed Sensing addreses the problem of solving an under-determined system of linear equations where the solution is known to be sparse. Most of the recovery algorithms in this area propose to solve this problem by convex optimization methods. This is a major shortcoming since it is known that better solutions can be achieved *via* non-convex optimization. Moreover such solutions almost always assume that the noise is Gaussian – which may not be the case. The few studies related to non-convex optimization are confined to the noiseless case. This work addresses these shortcomings and s a non-convex optimization method that can handle any type of additive noise.

Our algorithm is simple to implement. We have made our code publicly available from [10]. The algorithm is non-parametric and does not require the knowledge of the noise variance or the amount of model misfit. Experimental results indicate that significant improvements can be achieved by our proposed method over previous non-convex methods.

5. ACKNOWLEDGEMENT

The authors want to thank Igor Carron for sharing the code for the following optimization problem.

$$\min \| x \|_p \text{ subject to } y = Ax$$

which in turn is based upon IRLS algorithm in [1].

REFERENCES

[1] R. Chartrand, and W. Yin, "Iteratively reweighted algorithms for compressive sensing," *ICASSP 2008*. 2008, pp. 3869-3872.

[2] R. Saab. and O. Yilmaz, “Sparse recovery by non-convex optimization - instance optimality”, *Appl. Comput. Harmonic Anal*., in press.

[3] D. L. Donoho, “For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution”, *Commun. Pure Appl. Math.*, Vol. 59, pp. 797-829, 2006.

[4] S. Foucart and M. J. Lai, “Sparsest solutions of underdetermined linear systems *via* l_q-minimization for 0 0<q<1”, *Appl. Comput. Harmonic Anal.*, in press.

[5] B. D. Rao and K. Kreutz-Delgado, “An affine scaling methodology for best basis selection”, *IEEE Trans Signal Process*, Vol. 47, no. 1, pp. 187-200, 1999.

[6] B. Wohlberg and P. Rodríguez, “An Iteratively Reweighted Norm Algorithm for Minimization of Total Variation Functionals”, *IEEE Signal Process Lett.*, Vol. 14, no. 12, pp. 948-951, 2007.

[7] I. Daubechies, R. DeVore, M. Fornasier, and C. Sinan Güntürk, Iteratively re-weighted least squares minimization for sparse recovery, 2008, in press.

[8] D. Tradz, T. Ulrych, and M. Sacchi, “Latest views of the sparse Radon transform”, *Geophysics*, Vol. 68, no. 1, pp. 386-399, 2003.

[9] E. Haber, Numerical strategies for the solution of inverse problems: Ph.D. thesis, University of British Columbia, 1997.

[10] http://www.mathworks.com/matlabcentral/fileexchange/24814-non-convex-compressed-sensing-for-non-gaussian-noise

Received: October 19, 2009 Revised: November 03, 2009 Accepted: November 03, 2009

© Majumdar and Ward; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.