

# ITERATIVELY REWEIGHTED ALGORITHMS FOR COMPRESSIVE SENSING

Rick Chartrand

Los Alamos National Laboratory  
rickc@lanl.gov

Wotao Yin\*

Rice University  
wotao.yin@rice.edu

## ABSTRACT

The theory of compressive sensing has shown that sparse signals can be reconstructed exactly from many fewer measurements than traditionally believed necessary. In [1], it was shown empirically that using  $\ell^p$  minimization with  $p < 1$  can do so with fewer measurements than with  $p = 1$ . In this paper we consider the use of iteratively reweighted algorithms for computing local minima of the nonconvex problem. In particular, a particular regularization strategy is found to greatly improve the ability of a reweighted least-squares algorithm to recover sparse signals, with exact recovery being observed for signals that are much less sparse than required by an unregularized version (such as FOCUSS, [2]). Improvements are also observed for the reweighted- $\ell^1$  approach of [3].

**Index Terms**— Compressive sensing, signal reconstruction, nonconvex optimization, iteratively reweighted least squares,  $\ell^1$  minimization.

## 1. INTRODUCTION

Recent papers [4, 5] have introduced the concept known as *compressive sensing* (among other related terms). The basic principle is that sparse or compressible signals can be reconstructed from a surprisingly small number of linear measurements, provided that the measurements satisfy an *incoherence* property (see, e.g., [6] for an explanation of incoherence). Such measurements can then be regarded as a compression of the original signal, which can be recovered if it is sufficiently compressible. A few of the many potential applications are medical image reconstruction [7], image acquisition [8], and sensor networks [9].

The first algorithm presented in this context is known as basis pursuit [10]. Let  $\Phi$  be an  $M \times N$  measurement matrix, and  $\Phi x = b$  the vector of  $M$  measurements of an  $N$ -dimensional signal  $x$ . The reconstructed signal  $u^*$  is the minimizer of the  $\ell^1$  norm, subject to the data:

$$\min_u \|u\|_1, \quad \text{subject to } \Phi u = b. \quad (1)$$

\*The second author was supported by an internal faculty research grant from the Dean of Engineering at Rice University. He would like to thank Elaine Hale and Yin Zhang for valuable help with coding.

A remarkable result of Candès and Tao [11] is that if, for example, the rows of  $\Phi$  are randomly chosen, Gaussian distributed vectors, there is a constant  $C$  such that if the support of  $x$  has size  $K$  and  $M \geq CK \log(N/K)$ , then the solution to (1) will be exactly  $u^* = x$  with overwhelming probability. The required  $C$  depends on the desired probability of success, which in any case tends to one as  $N \rightarrow \infty$ . Donoho and Tanner [12] have computed sharp reconstruction thresholds for Gaussian measurements, so that for any choice of sparsity  $K$  and signal size  $N$ , the required number of measurements  $M$  for (1) to recover  $x$  can be determined precisely.

Variants of these results have included  $\Phi$  being a random Fourier submatrix, or having values  $\pm 1/\sqrt{N}$  with equal probability. More general matrices are considered in [6, 13]. Also,  $x$  can be sparse with respect to any basis, with  $u$  replaced with  $\Psi u$  for suitable unitary  $\Psi$ .

A family of iterative greedy algorithms [14, 15, 16] have been shown to enjoy a similar exact reconstruction property, generally with less computational complexity. However, these algorithms require more measurements for exact reconstruction than the basis pursuit method.

In the other direction, it was shown in [1] that a nonconvex variant of basis pursuit will produce exact reconstruction with fewer measurements. Specifically, the  $\ell^1$  norm is replaced with the  $\ell^p$  norm, where  $0 < p < 1$  (in which case  $\|\cdot\|_p$  isn't actually a norm, though  $d(x, y) = \|x - y\|_p^p$  is a metric):

$$\min_u \|u\|_p^p, \quad \text{subject to } \Phi u = b. \quad (2)$$

That fewer measurements are required for exact reconstruction than when  $p = 1$  was demonstrated by numerical experiments in [1], with random and nonrandom Fourier measurements. A theorem was also proven in terms of the restricted isometry constants of  $\Phi$ , generalizing a result of [17] to show that a condition sufficient for (2) to recover  $x$  exactly is weaker for smaller  $p$ . More recently, for the case of random, Gaussian measurements, the above condition of Candès and Tao has been shown (using different methods) [18] to generalize to

$$M \geq C_1(p)K + pC_2(p)K \log(N/K), \quad (3)$$

where  $C_1, C_2$  are determined explicitly, and are bounded in  $p$ . Thus, the dependence of the sufficient number of mea-

measurements  $M$  on the signal size  $N$  decreases as  $p \rightarrow 0$ . The constants are not sharp, however.

## 2. ALGORITHMS FOR NONCONVEX COMPRESSIVE SENSING

Early papers considering iteratively reweighted least squares (IRLS) approaches include [19, 20]. IRLS for solving (2) has mostly been considered for  $p \geq 1$ . The case of  $p < 1$  was studied by Rao and Kreutz-Delgado [2]. The approach is to replace the  $\ell^p$  objective function in (2) by a weighted  $\ell^2$  norm,

$$\min_u \sum_{i=1}^N w_i u_i^2, \quad \text{subject to } \Phi u = b, \quad (4)$$

where the weights are computed from the previous iterate  $u^{(n-1)}$ , so that the objective in (4) is a first-order approximation to the  $\ell^p$  objective:  $w_i = |u_i^{(n-1)}|^{p-2}$ . The solution of (4) can be given explicitly, giving the next iterate  $u^{(n)}$ :

$$u^{(n)} = Q_n \Phi^T (\Phi Q_n \Phi^T)^{-1} b, \quad (5)$$

where  $Q_n$  is the diagonal matrix with entries  $1/w_i = |u_i^{(n-1)}|^{2-p}$ . This comes from solving the Euler-Lagrange equation of (4), using the constraint to solve for the Lagrange multipliers, then substituting this value into the solution. The result is a fixed-point iteration for solving the Euler-Lagrange equation of (2).

In this paper, we are considering  $0 \leq p \leq 1$ . The case of  $p = 0$  corresponds to an objective of  $\sum \log(|u_i|^2)$ . Given that  $p - 2$  will be negative, the weights  $w_i$  are undefined whenever  $u_i^{(n-1)} = 0$ . A common approach for dealing with this issue is to regularize the optimization problem, by incorporating a small  $\epsilon > 0$ . For example, below we consider the damping approach:

$$w_i = ((u_i^{(n-1)})^2 + \epsilon)^{p/2-1}. \quad (6)$$

As observed in [2], the iteration (5) depends only on  $1/w_i$ , which if defined directly as  $|u_i^{(n-1)}|^{2-p}$  is well defined for any value of  $u_i^{(n-1)}$ . In principle, no regularization is needed. For many systems, however, the matrix being inverted in (5) is too ill-conditioned to allow accurate computation of  $u^{(n)}$ .

In Section 3, we perform experiments that show that the strategy of using a relatively large  $\epsilon$  in (6), then repeating the process of decreasing  $\epsilon$  after convergence and repeating the iteration (5), dramatically improves the ability of IRLS to recover sparse signals. Exact recovery becomes possible with many fewer measurements, or with signals that are much less sparse. This  $\epsilon$ -regularization strategy was used effectively with a projected gradient algorithm in [1].

It must be noted that (2) is a nonconvex optimization problem when  $p < 1$ , and all of the algorithms considered here are only designed to produce local minima. However, the fact

that in practice we are able to recover signals exactly, combined with theoretical results [1, 18] that give circumstances in which (2) has a unique, global minimizer that is exactly  $u^* = x$ , strongly suggests that the computed local minimizers are actually global, at least under a broad set of circumstances. A possible explanation for this in the context of the  $\epsilon$ -regularization strategy is that adding a relatively large  $\epsilon$  in the weights results in undesirable local minima being “filled in.” Since it is known that the sparsity of  $x$  and the incoherence of  $\Phi$  combine to make  $x$  the global minimum of (2), it is also plausible that the basin containing  $x$  will be deeper, and less likely to be filled in than other, local minima. With a large  $\epsilon$ , we observe our algorithm to converge to a reasonably nearby point; once  $u$  is in the correct basin, decreasing  $\epsilon$  allows the basin to deepen, and  $u$  approaches  $x$  more closely, converging to  $x$  as  $\epsilon \rightarrow 0$ . We hope to turn these notions into a proof of convergence of the  $\epsilon$ -regularized IRLS algorithm to the global minimum of (2).

As partial justification, we show that  $x$  will be a fixed point of the procedure described above.

**Theorem 2.1.** *Let  $x \in \mathbb{R}^N$  be a vector of sparsity  $\|x\|_0 = K$ . Let  $\Phi$  be an  $M \times N$  matrix with the property that every collection of  $2K$  columns of  $\Phi$  is linearly independent. Let  $\epsilon_j \in (0, 1)$  be a sequence converging to zero. For each  $j$ , let  $u^{*,j}$  be the unique solution of the strictly convex optimization problem (4), where the weights  $w_i$  are given by  $w_i = (x_i^2 + \epsilon_j)^{p/2-1}$ ,  $0 \leq p < 2$ , and  $b = \Phi x$ . Then  $u^{*,j} \rightarrow x$ .*

The property assumed of  $\Phi$  is called the unique representation property by Gorodnitsky and Rao [21], who observe that it implies that  $x$  is the unique solution of  $\Phi u = b$  having sparsity  $\|u\|_0 \leq K$ . Among many other examples, this property will hold with probability 1 for a random, Gaussian matrix  $\Phi$  provided  $M \geq 2K$ .

*Proof.* First, the sequence  $(u^{*,j})$  is bounded. In fact, a known property of IRLS is boundedness independent of the weights; see [22].

Next, fix  $k$  such that  $x_k = 0$ . Then

$$\sum_{i=1}^N w_i (u_i^{*,j})^2 \geq w_k (u_k^{*,j})^2 = (u_k^{*,j})^2 / \epsilon_j^{1-p/2}, \quad (7)$$

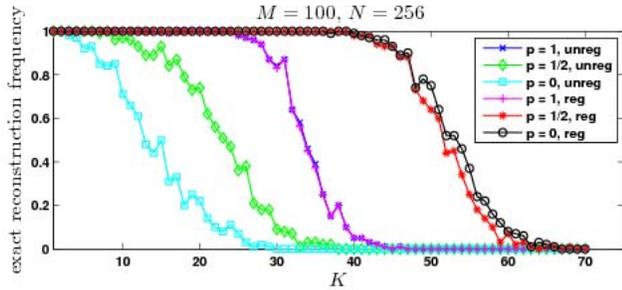
while by the optimality of  $u^{*,j}$ ,

$$\sum_{i=1}^N w_i (u_i^{*,j})^2 \leq \sum_{i=1}^N w_i x_i^2 \leq \|x\|_p^p. \quad (8)$$

Combining, we obtain

$$(u_k^{*,j})^2 \leq \epsilon_j^{1-p/2} \|x\|_p^p \rightarrow 0. \quad (9)$$

So we see  $(u^{*,j})$  tends to zero off the support of  $x$ . Then any limit point  $u_0$  of  $(u^{*,j})$  will be a solution of  $\Phi u_0 = b$  having at most  $K$  nonzero components. By the remarks preceding the proof, it must be that  $u_0 = x$ . Therefore  $u^{*,j} \rightarrow x$ .  $\square$



**Fig. 1.** Plots of recovery frequency as a function of  $K$ . Regularized IRLS has a much higher recovery rate than unregularized IRLS, except when  $p = 1$  when they are almost identical. Regularized IRLS recovers the greatest range of signals when  $p$  is small, while unregularized IRLS performs less well for small  $p$  than when  $p = 1$ .

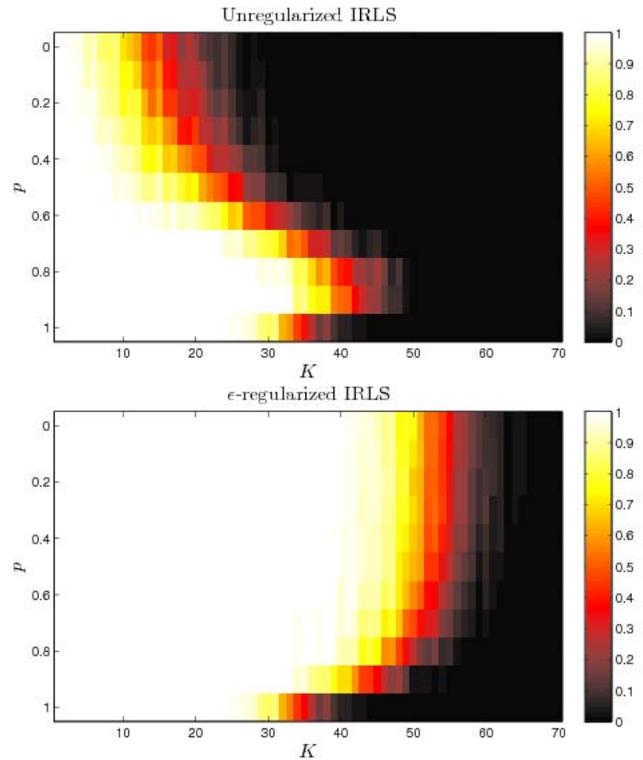
Candès, Wakin, and Boyd have proposed an iteratively reweighted  $\ell^1$  minimization algorithm corresponding to the  $p = 0$  case above [3]. Below we compare our  $\epsilon$ -regularized IRLS results to a similar,  $\epsilon$ -regularized iteratively reweighted  $\ell^1$  minimization algorithm for a few values of  $p$ .

### 3. NUMERICAL EXPERIMENTS

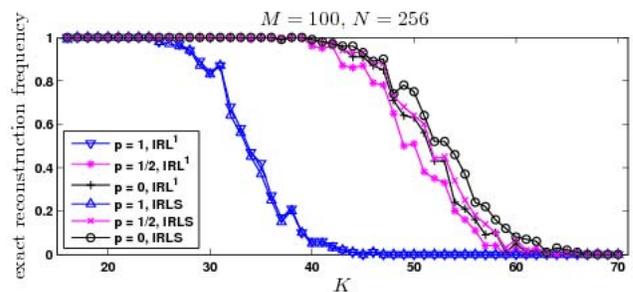
For our experiments, for each of 100 trials we randomly select entries of a  $100 \times 256$  matrix  $\Phi$  from a mean-zero Gaussian distribution, then scale the columns to have unit 2-norm. For each value of  $K$ , we randomly choose the support of  $x$ , then choose the components from a Gaussian distribution of mean 0 and standard deviation 2. The same  $\Phi$  and  $x$  would then be used for each algorithm and choice of  $p$ . For  $\epsilon$ -regularized IRLS,  $\epsilon$  is initialized to 1 and  $u^{(0)}$  initialized to the minimum 2-norm solution of  $\Phi u = b$ . The iteration (5) with weights defined by (6) is run until the change in relative 2-norm from the previous iterate is less than  $\sqrt{\epsilon}/100$ , at which point  $\epsilon$  is reduced by a factor of 10, and the iteration repeated beginning with the previous solution. Decreasing  $\epsilon$  too soon results in much poorer recovery performance. This process is continued through a minimum  $\epsilon$  of  $10^{-8}$ , below which the matrix being inverted in (5) can become ill-conditioned. For this reason, the “unregularized” IRLS is implemented as  $\epsilon$ -regularized IRLS, but with a single  $\epsilon$  of  $10^{-8}$ .

Results are shown in Figures 1 and 2. We see that  $\epsilon$ -regularized IRLS is able to recover signals with many more nonzero components when  $p = 0$ , in comparison with unregularized IRLS, or with  $\epsilon$ -regularized IRLS with  $p = 1$ . For  $p = 1$ , the two algorithms perform almost identically, but then unregularized IRLS decays as  $p$  decreases, while  $\epsilon$ -regularized IRLS improves.

For comparison with the proposed  $IR\ell^1$  approach of Candès, Wakin, and Boyd, we implemented the above procedure, except at each iteration the weighted least squares step (4) was



**Fig. 2.** Same data as in Figure 1, but with every  $p$  shown. Unregularized IRLS is best at  $p = 0.9$ , then decays quickly for smaller  $p$ , Regularized IRLS improves as  $p$  gets smaller, and recovers many more signals than unregularized IRLS.



**Fig. 3.** Comparison of  $IR\ell^1$  with IRLS, both  $\epsilon$ -regularized. The recovery rate is near 1 for the same values of  $K$ , while IRLS has a slightly higher recovery rate for larger  $K$ , except for  $p = 1$  where the two are essentially the same.

replaced with solving the following weighted- $\ell^1$  problem:

$$\min_u \sum_{i=1}^N w_i |u_i|, \quad \text{subject to } \Phi u = b, \quad (10)$$

where the weights are given by

$$w_i = (|u_i^{(n-1)}| + \epsilon_j)^{p-1}. \quad (11)$$

We used a Mosek-based, interior-point linear programming solver to solve (10). The results are somewhat better than those in [3]; see Figure 3. The recovery rate remains near 1 for as many signals as  $\epsilon$ -regularized IRLS. For larger  $K$ , the recovery rate of  $\text{IR}^{\ell^1}$  is just slightly lower than that of IRLS, except for  $p = 1$  where the difference is negligible. Since IRLS is computationally simpler, our results suggest that IRLS is the better approach.

#### 4. REFERENCES

- [1] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.
- [2] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 187–200, 1999.
- [3] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization." Preprint.
- [4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.
- [5] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, 2006.
- [6] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, 2007.
- [7] M. Lustig, J. M. Santos, J.-H. Lee, D. L. Donoho, and J. M. Pauly, "Application of compressed sensing for rapid MR imaging," in *SPARS*, (Rennes, France), 2005.
- [8] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Computational Imaging IV - Proceedings of SPIE-IS and T Electronic Imaging*, vol. 6065, 2006.
- [9] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *39th Asilomar Conference on Signals, Systems and Computers*, pp. 1537–1541, 2005.
- [10] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.
- [11] E. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, 2006.
- [12] D. L. Donoho and J. Tanner, "Thresholds for the recovery of sparse solutions via L1 minimization," in *40th Annual Conference on Information Sciences and Systems*, pp. 202–206, 2006.
- [13] R. A. DeVore, "Deterministic construction of compressed sensing matrices." Preprint.
- [14] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit." Preprint.
- [15] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from incoherent projections," in *SPARS*, (Rennes, France), 2005.
- [16] C. La and M. N. Do, "Signal reconstruction using sparse tree representations," in *Wavelets XI*, vol. 5914, SPIE, 2005.
- [17] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, 2005.
- [18] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing." Preprint.
- [19] C. L. Lawson, *Contributions to the theory of linear least maximum approximations*. PhD thesis, UCLA, 1961.
- [20] A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, vol. 16, pp. 147–185, 1974.
- [21] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a reweighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, pp. 600–616, 1997.
- [22] G. W. Stewart, "On scaled projections and pseudoinverses," *Linear Algebra Appl.*, vol. 112, pp. 189–194, 1989.