

# The Two Stage l1 Approach to the Compressed Sensing Problem

Stéphane Chrétien

► **To cite this version:**

Stéphane Chrétien. The Two Stage l1 Approach to the Compressed Sensing Problem. Rémi Gribonval. SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations, Apr 2009, Saint Malo, France. 2009. <inria-00369534>

**HAL Id: inria-00369534**

**<https://hal.inria.fr/inria-00369534>**

Submitted on 20 Mar 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Two Stage $l_1$ Approach to the Compressed Sensing Problem

Stéphane Chrézien  
 Mathematics Department  
 University of Franche Comté  
 16 route de Gray  
 25000 Besançon, France  
 Email : stephane.chretien@univ-fcomte.fr

**Résumé**—This paper gives new results on the recovery of sparse signals using  $l_1$ -norm minimization. We introduce a two-stage  $l_1$  algorithm equivalent to the first two iterations of the alternating  $l_1$  relaxation introduced in [1].

## I. INTRODUCTION

The Compressed sensing problem is currently the focus of an extensive research activity and can be stated as follows : Given a sparse vector  $x^* \in \mathbb{R}^n$  and an observation matrix  $A \in \mathbb{R}^{m \times n}$  with  $m \ll n$ , try to recover the vector  $x$  from the small measurement vector  $y = Ax^*$ . Although the problem consists of solving an overdetermined system of linear equations, enough sparsity will allow to succeed as shown by the following lemma (where  $\Sigma_s$  will denote the set of all  $s$ -sparse vectors, i.e. vectors whose components are all zero except for at most  $s$  of them),

**Lemma I.1** [2] *If  $A$  is any  $m \times n$  matrix and  $2s \leq m$ , then the following properties are equivalent :*

*i. The decoder  $\Delta_0(y)$  given by*

$$\Delta_0(y) = \operatorname{argmin}_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad Ax = y. \quad (1)$$

*$k$  satisfies  $\Delta_0(Ax) = x$ , for all  $x \in \Sigma_s$ ,*

*ii. For any set of indices  $T$  with  $\#T = 2k$ , the matrix  $A_T$  has rank  $2s$  where  $A_T$  stands for the submatrix of  $A$  composed of the columns indexed by  $T$  only.*

### A. The $l_1$ and the Reweighted $l_1$ relaxations

The main problem with decoder  $\Delta_0$  is that the optimization problem (1) is in general NP-hard. For this reason, the now standard  $l_1$  relaxation strategy is adopted, i.e. the decoder  $\Delta_1(y)$  is obtained as

$$\Delta_1(y) = \operatorname{argmin}_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad Ax = y. \quad (2)$$

Now, solving (2) can be done in polynomial time and thus  $\Delta_1(y)$  can be efficiently computed. The second problem is to give robust conditions under which exact recovery holds. One such condition was given by Candes Romberg and Tao [3] and is now known as the Uniform Uncertainty Principle (UUP) or as the Restricted Isometry Property (RIP).

One of the main remaining challenges is to reduce the number of observations  $m$  needed to recover a given sparse

signal  $x$ . One idea is the use of  $l_p$ ,  $p < 0 < 1$  decoders  $\Delta_p(y)$ . The main draw back of the approach using  $l_p$ ,  $p < 0 < 1$  norm minimization is that the resulting decoding scheme is again NP-Hard. Another idea is to use a reweighted  $l_1$  approach as proposed in [4].

---

### Algorithm 1 Reweighted $l_1$ algorithm (Rew- $l_1$ )

---

**Input**  $u > 0$  and  $L \in \mathbb{N}_*$   
 $z_u^{(0)} = e$   
 $x_u^{(0)} \in \min_{x \in \mathbb{R}^n, Ax=y} \|x\|_1$   
 $l = 1$   
**while**  $l \leq N$  **do**  
 $z_u^{(l)} = \frac{1}{|x^{(l)}|+u}$  componentwise  
 $x_u^{(l)} \in \operatorname{argmax}_{x \in \mathbb{R}^n, Ax=y} \sum_{i=1}^n w_i^{(l-1)} |x_i|$   
 $l \leftarrow l + 1$   
**end while**  
**Output**  $z_u^{(L)}$  and  $x_u^{(L)}$ .

---

The main intuition behind this reweighted  $l_1$  relaxation is the following. The greater the component  $x_i$  becomes, the smaller weight it should receive since it can be considered that this component should not be set to zero.

The main drawback of the reweighted  $l_1$  approach is that an unknown parameter is to be tuned whose order of magnitude is hard to know ahead of time.

### B. The Alternating $l_1$ algorithm

Another approach was proposed in [1] and uses Lagrange duality. Let us write down problem (1), to which  $\Delta_0$  is the solution map, as the following equivalent problem

$$\max_{z \in \{0,1\}^n, x \in \mathbb{R}^n} e^t z \quad \text{s.t.} \quad z_i x_i = 0, \quad i = 1, \dots, n, \quad Ax = y$$

where  $e$  denotes the vector of all ones. Here since the sum of the  $z_i$ 's is maximized, the variable  $z$  plays the role of an indicator function for the event that  $x_i = 0$ . This problem is clearly nonconvex due to the quadratic equality constraints  $z_i x_i = 0$ ,  $i = 1, \dots, n$ . However, these constraints can be merged into the unique constraint  $\|D(z)x\|_1 = 0$ , leading to the following equivalent problem

$$\max_{z \in \{0,1\}^n, x \in \mathbb{R}^n} e^t z \quad \text{s.t.} \quad \|D(z)x\|_1 = 0, \quad Ax = y. \quad (3)$$

The Alternating  $l_1$  algorithm consists of a suboptimal alternating minimization procedure to approximate the dual function at  $u$ . The algorithm is as follows.

---

**Algorithm 2** Alternating  $l_1$  algorithm (Alt- $l_1$ )

---

**Input**  $u > 0$  and  $L \in \mathbb{N}_*$

$$z_u^{(0)} = e$$

$$x_u^{(0)} \in \max_{x \in \mathbb{R}^n, Ax=y} \mathcal{L}(x, z^{(0)}, u)$$

$$l = 1$$

**while**  $l \leq N$  **do**

$$z_u^{(l)} \in \operatorname{argmax}_{z \in \{0,1\}^n} \mathcal{L}(x_u^{(l)}, z, u)$$

$$x_u^{(l)} \in \operatorname{argmax}_{x \in \mathbb{R}^n, Ax=y} \mathcal{L}(x, z_u^{(l)}, u)$$

$$l \leftarrow l + 1$$

**end while**

**Output**  $z_u^{(L)}$  and  $x_u^{(L)}$ .

---

Notice that, similarly to the reweighted  $l_1$  algorithm, the Alternating  $l_1$  method also requires the tuning of an unknown parameter  $u$ . However, the main motivation for this proposal is that this parameter  $u$  has a clear meaning : it is a dual variable which, in the case where the dual function  $\theta(u)$  is well approximated by the sequence  $\mathcal{L}(x^{(l)}, z^{(l)}, u)$ , can be efficiently optimized without additional prior information, due to the convexity of the dual function.

## II. THE TWO STAGE $l_1$ METHOD

The main remark about the alternating  $l_1$  method is the following (see [1]) : for a given dual variable  $u$ , the alternating  $l_1$  algorithm can be seen a sequence  $(x_u^{(l)})_{l \in \mathbb{N}}$  of truncated  $l_1$ -norm minimizers of the type

$$x_u^l = \operatorname{argmin}_{x \in \mathbb{R}^n} \|x_{T_u^l}\|_1 \quad \text{s.t.} \quad Ax = y. \quad (4)$$

where  $T_u^l$  is the set of indices for which  $|x_i^{l-1}| < \frac{1}{u}$ . Therefore, the Alternating  $l_1$  algorithm can be seen as an iterative thresholding scheme with threshold value equal to  $\frac{1}{u}$ . Now assume for instance that a fraction  $\rho m$  of the non zero components is well identified by the plain  $l_1$  step with solution  $x^{(l)}$ . Then, the practitioner might ask if the appropriate value for  $u$  is the one which imposes an  $l_1$  penalty on the index set corresponding to the  $n - \rho m$  smallest components of  $x^{(l)}$ . Moreover, the large scale simulation experiments which have been performed on the plain  $l_1$  relaxation seemed to agree on the fact that the breakdown point occurs near  $\frac{m}{4}$ . Thus, a practitioner could be tempted to wonder whether  $\rho = \frac{1}{4}$  is a sensible value. Motivated by the previous practical considerations, the two stage  $l_1$  algorithm is defined as follows (the parameter  $u$  is now replaced by the parameter  $\rho = \frac{1}{u}$ ).

Notice that we restrict  $\rho$  to lie in  $(0, \frac{1}{2})$ . The reason should be obvious since, due to Lemma I.1, even decoder  $\Delta_0(y)$  is unable to identify more than  $\frac{m}{2}$ -sparse vectors. Another remark is that the procedure could be continued for more than 2 steps but simulation experiments of the Alternating  $l_1$  method seem to confirm that in most cases two steps suffice to converge.

---

**Algorithm 3** Two stage  $l_1$  algorithm (2Stage- $l_1$ )

---

**Input**  $\rho \in (0, \frac{1}{2})$

*Step 0* :  $x^{(0)} \in \operatorname{argmax}_{x \in \mathbb{R}^n, Ax=y} \|x\|_1$  and  $T = \text{index set of the } \rho m \text{ largest components of } x^{(0)}$

*Step 1* :  $x^{(1)} \in \operatorname{argmax}_{x \in \mathbb{R}^n, Ax=y} \|x_{T^c}\|_1$

**Output**  $x_\rho^{(1)}$ .

---

## III. MAIN RESULTS

At Step 1 of the method, a subset  $T$  is selected with cardinal  $\rho m$  and optimization is then performed with objective function  $\|x_{T^c}\|_1$ . In this section, we will adopt the following notations :  $S$  will denote the support of  $x^*$ ,  $T$  will denote the index set of the  $\rho m$  largest components of  $x^{(0)}$  as defined in the two-stage  $l_1$  algorithm.  $T_g^c$  will be an abbreviation of  $(T^c)_g$ , the "good" subset of  $T^c$  or, in mathematical terms, the subset of indices of  $S$  which also belong to  $T^c$ . On the other hand,  $T_b^c$  will denote the complement of  $T_g^c$  in  $T^c$ .

**Lemma III.1** *Assume the cardinal of  $T_g^c$  is less than  $\gamma/2$  and that  $A$  satisfies  $RIP(\delta, \gamma s)$ . Let  $h^{(1)} = x^{(1)} - x^*$ . Then, there exists a positive number  $C^*$  depending on  $x^*$  such that  $\|h_T^{(1)}\|_1 \leq C^* \|x_{T_b^c}^*\|_1$ . Moreover, if  $\|h_T^{(1)}\|_1 = 0$ , then  $\|x_{T_b^c}^*\|_1 = 0$ .*

**Proof.** Let  $N(h_T)$  denote the optimal value of the problem

$$\min_{h_{T_b^c} \in \mathbb{R}^{n-s}} \|x_{T_b^c}^* + h_{T_b^c}\|_1 \quad (5)$$

subject to

$$A_{T_b^c}(x_{T_b^c}^* + h_{T_b^c}) + A_{T_g^c}x_g^{(1)} = y - A_T(x_T^* + h_T).$$

Assume that  $x_{T_b^c}^* = 0$ . Then,  $N(h_T)$  plays the role of a norm for  $h_T$  although it does not satisfy the triangle inequality. In particular,  $N(h_T)$  is nonnegative, convex and  $N(h_T) = 0$  implies that  $h_T = 0$ .

Nonnegativity and convexity are straightforward. Assume that  $N(h_T) = 0$ , i.e. the solution  $\tilde{h}$  of (5) is null. This implies that  $A_{T_g^c}x_{T_g^c}^{(1)} = y - A_Tx_T^* = A_{T_b^c}x_{T_b^c}^*$ , which implies that  $x_{T_g^c}^{(1)} - x_{T_b^c}^*$  is in the kernel of  $A_{T_g^c}$ . Using the fact that  $T_g^c$  has cardinal less than  $\gamma s/2$  and the  $RIP(\delta, \gamma s)$  assumption, we conclude that  $x_{T_g^c}^{(1)} = x_{T_b^c}^*$ . In order to finish the proof of the lemma, it remains to recall that  $N(h_T)$  is convex and that, by Theorem 1.1 in [5],  $\|h^{(1)}\|_1$  (and thus  $\|h_T^{(1)}\|_1$ ) is bounded from above by  $C \inf_{\#U \leq \gamma/2} \|x_{U^c}\|_1$  in order to obtain existence of a sufficiently small positive constant  $C^*$  depending on  $x^*$  such that  $N(h_T) \geq C^* \|h_T\|_1$  for all  $h_T$  in the ball  $B(0, C \|x^*\|_1)$ . The desired result then follows.

To prove that  $N(h_T) = 0$  if  $h_T = 0$  is a bit harder. Thus, assume that  $h_T = 0$ . Then, the solution  $\tilde{h}$  of (5) is just the solution of

$$\min_{h_{T_b^c} \in \mathbb{R}^{n-s}} \|h_{T_b^c}\|_1 \quad A_{T_b^c}h_{T_b^c} = y - A_T - x_T^* - A_{T_g^c}x_g^{(1)}.$$

Now since  $y = Ax^*$ , we obtain that  $y - A_T x_T^* - A_{T_g^c} x_g^{(1)} = A_{T_g^c} (x_{T_g^c}^* - x_g^{(1)})$  and thus, the right hand side term is nothing but the image of a  $\gamma s/2$ -sparse vector. Now, recalling that we assumed  $RIP(\delta, \gamma s)$ , Theorem 1.1 in [5] implies that  $\tilde{h}$  must be the sparsest solution of the system  $A_{T_g^c} h_{T_g^c} = y - A_T x_T^* - A_{T_g^c} x_g^{(1)}$  from which we deduce that  $\tilde{h}$  is  $\gamma s/2$ -sparse. Therefore the vector  $(h_{T_g^c}, x_{T_g^c}^{(1)})$  is  $\gamma s$  which solves  $A_{T_g^c} x_{T_g^c} = y - A_T x_T^*$ . On the other hand,  $x_{T_g^c}^*$  also solves  $A_{T_g^c} x_{T_g^c} = y - A_T x_T^*$  and its support is included in the support of  $(h_{T_g^c}, x_{T_g^c}^{(1)})$ . Therefore,  $(h_{T_g^c}, x_{T_g^c}^{(1)}) - x_{T_g^c}^*$  is a  $\gamma s/2$  sparse vector which lies in the kernel of  $A$ . Using again the fact that  $RIP(\delta, \gamma s)$  holds, we conclude that  $(h_{T_g^c}, x_{T_g^c}^{(1)}) - x_{T_g^c}^* = 0$ . Thus,  $h_{T_g^c} = 0$  and  $x_{T_g^c}^{(1)} = x_{T_g^c}^*$ .  $\square$

Using this lemma, we deduce the following theorem.

**Theorem III.1** Assume that  $RIP(\delta, \gamma s)$  holds and that an index set  $T_g$  of cardinal greater than or equal to  $(1 - \gamma/2)s$  has been recovered at Step 0 after thresholding, then  $x^{(1)}$  satisfies

$$\|x^{(1)} - x^*\|_1 \leq C^{**} \|x_{T_g^c}^*\|_1.$$

for some constant  $C^{**}$  depending on  $x^*$ .

**Proof.** The vector  $x^{(1)}$  satisfies

$$\|x_{T_g^c}^{(1)}\|_1 \leq \|x_{T_g^c}^*\|_1. \quad (6)$$

Let us write  $h^{(1)} = x^{(1)} - x^*$ . Using (6), a now standard decomposition gives

$$\|x_{T_g^c}^*\|_1 - \|h_{T_g^c}\|_1 + \|h_{T_g^c}\|_1 - \|x_{T_g^c}^*\|_1 \leq \|x_{T_g^c}^*\|_1 + \|x_{T_g^c}^*\|_1.$$

We thus obtain

$$\|h_{T_g^c}\|_1 \leq \|h_{T_g^c}\|_1 + 2\|x_{T_g^c}^*\|_1. \quad (7)$$

However, since  $RIP(\delta, \gamma s)$  holds,  $NSP(C, \gamma/2s)$  holds too, with  $C < 1$ . Therefore, we obtain that

$$\|h_{T_g^c}\|_1 \leq C(\|h_{T_g^c}\|_1 + \|h_T\|_1). \quad (8)$$

Combining (7) and (8), we obtain

$$\|h_{T_g^c}\|_1 \leq \frac{C}{1-C} \|h_T\|_1 + \frac{2}{1-C} \|x_{T_g^c}^*\|_1. \quad (9)$$

As a consequence, we obtain that

$$\begin{aligned} \|h\|_1 &\leq C(2\|x_{T_g^c}^*\|_1 + C'\|h_T\|_1) + \|h_T\|_1 + 2\|x_{T_g^c}^*\|_1 \\ &\quad + C\|h_T\|_1 + \|h_T\|_1 \\ &= (1 + C + CC')\|h_T\|_1 + 2(1 + C)\|x_{T_g^c}^*\|_1. \end{aligned}$$

which, using Lemma III.1, implies

$$\|h\|_1 \leq ((1 + C + CC')C^* + 2(1 + C))\|x_{T_g^c}^*\|_1. \quad (10)$$

which is the desired bound.  $\square$

The following corollary is a straightforward consequence of the previous theorem.

**Corollary III.1** Assume that the assumptions of Theorem III.1 are satisfied. Then, exact reconstruction is obtained if  $x_{T_g^c}^* = 0$ , i.e.  $x^*$  is  $s$ -sparse.

## IV. MONTE CARLO EXPERIMENTS

The following Monte Carlo experiments show that the performance of the two-stage  $l_1$  algorithm which drops the penalty over the index set of the  $m/4$  largest components of the solution of plain  $l_1$  are almost as good as the performance of the reweighted  $l_1$  with the best parameter which is usually unknown in practice. A Python program is available at <http://stephane.g.chretien.googlepages.com/alternatingl1> and can be used to perform these experiments and other involving the Alternating  $l_1$  algorithm.

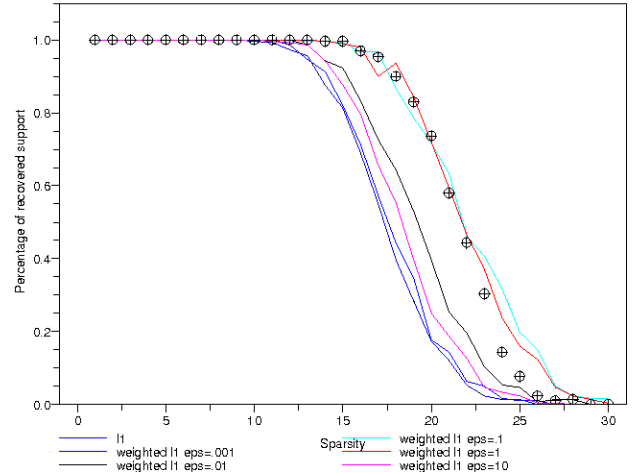


FIG. 1. Rate of success over 300 Monte Carlo experiments in recovering the support of the signal vs. signal sparsity  $k$  for  $n = 128$ ,  $m = 50$ ,  $L = 4$ ,  $u = 3$ .  $A$  and nonnull components of  $x$  were drawn from the gaussian  $\mathcal{N}(0, 1)$  distribution. The results for the two-stage  $l_1$  method are represented by the “+ in a circle” sign.

## RÉFÉRENCES

- [1] S. Chretien, “An alternating  $l_1$  for the compressed sensing problem,” Preprint, available at <http://arxiv.org/abs/0809.0660>, 2008.
- [2] A. Cohen, W. Dahmen, and D. R., “Compressed sensing and best  $k$ -term approximation,” Preprint, 2006.
- [3] E. Candes, J. Romberg, and T. T., “Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information,” *IEEE Information Theory*, vol. 2, no. 52, pp. 489–509, 2006.
- [4] E. Candes, W. M., and B. S., “Enhancing sparsity by reweighted  $l_1$  minimization,” *Journal of Fourier Analysis and Applications*, 2008.
- [5] E. Candès, “The restricted isometry property and its implications for compressed sensing.” *Compte Rendus de l’Academie des Sciences, Paris, Serie I*, vol. 346, pp. 589–592, 2008.