An Alternating l1 approach to the compressed sensing problem

Stéphane Chrétien [a],

[a] *Laboratoire de Mathématiques, Université de Franche-Comté and UMR CNRS 6623, 16 route de Gray, 25030 Besançon Cedex, France*

Received *****; accepted after revision +++++

Presented by £££££

Abstract

An improvement of the standard l_1 relaxation is proposed for the Compressed Sensing problem. Lagrangian duality is used in order to produce a dual approach to the original combinatorial problem of sparsest recovery. We deduce from this approach a practical alternating maximization method and provide preliminary computational experiments showing that the proposed method outperforms the l_1 relaxation. *To cite this article: S. Chrétien, C. R. Mecanique 333 (2005).*

Résumé

Une méthode d'optimization l_1-alternée pour le problème de Compressed Sensing. Une extension de la relaxation l_1 standard pour le problème de Compressed Sensing est proposée. Le problème combinatoire de la recherche d'une solution la plus parcimonieuse d'un système linéaire est approché par la dualité lagrangienne. Nous déduisons de cette approche une version implantable de type optimisation alternée de la norme l_1 tronquée adaptativement. Des simulations sont présentées où la méthode montre un comportement moyen meilleur que celui de la relaxation l_1 simple. *Pour citer cet article : S. Chrétien, C. R. Mecanique 333 (2005).*

Key words: Signal Processing ; Compressed Sensing ; Compressive Sampling ; Linear Programming ; Lagrange duality

Mots-clés : Traitement du Signal ; Compressed Sensing ; Compressive Sampling ; Programmation Linéaire ; Dualité lagrangienne ; Optimisation alternée

1. Introduction

Compressed Sensing (CS) is a very recent field of fast growing interest and whose impact on concrete applications in coding and image acquisition is already remarkable. Up to date informations on this

Email address: `stephane.chretien@univ-fcomte.fr` (Stéphane Chrétien).

new topic may be obtained from the website *http://www.dsp.ece.rice.edu/cs/*. The foundational paper is [3] where the main problem considered was the one of reconstructing a signal from a few frequency measurements. Since then, important contributions to the field have appeared; see [1] for a survey and references therein.

1.1. *The Compressed Sensing problem*

In mathematical terms, the problem can be stated as follows. Let x be a k-sparse vector in $\mathbb{R}^n$, i.e. a vector with no more than k nonzero components. The observations are simply given by

$$y = Ax \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$ and m small compared to n, and the goal is to recover x exactly from these observations. One of the main problems of the field at its beginning was the construction of observation matrices A which allowed to recover x with k as large as possible for given values of n and m, using convex optimization.

The problem of compressed sensing can be solved unambiguously if there is no sparser solution to the linear system (1) than x. Then, recovery is obtained by simply finding the sparsest solution to (1). If for any x in $\mathbb{R}^n$ we denote by $\|x\|_0$ the l_0-norm of x, i.e. the cardinal of the set of indices of nonzero components of x, the compressed sensing problem is equivalent to

$$\min_{x\in\mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad Ax = y. \tag{2}$$

We denote by $\Delta_0(y)$ the solution of problem (2) and $\Delta_0(y)$ is called a decoder. Thus, the CS problem is a combinatorial optimization problem. Moreover, the following lemma is well known.

Lemma 1.1 [4] *If A is any $m \times n$ matrix and $2k \leq m$, then the following properties are equivalent:*

i. The decoder Δ_0 satisfies $\Delta_0(Ax) = x$, for all $x \in \Sigma_k$,

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

The main problem in this approach is that obtaining $\Delta_0(y)$ is generally computationally intractable. Candes, Romberg and Tao studied the convex l_1-relaxation instead and proved that this relaxation works under stonger conditions on A than the ones given in Lemma 1.1 for the decoder $\Delta_0(y)$ to work.

1.2. *The l_1 relaxation*

The main problem in using the decoder $\Delta_0(y)$ for given observations y is that the optimization problem (2) is NP-hard and cannot reasonably be expected to be solved in polynomial time. In order to overcome this difficulty, the original decoder $\Delta_0(y)$ has to be replaced by simpler ones in terms of computational complexity. Assuming that A is given, two methods have been studied for solving the compressed sensing problem. The first one is the orthognal matching pursuit (OMP) and the second one is the l_1-relaxation which consists of replacing the l_0-norm by the convex l_1-norm and thus can be expressed as

$$\min_{x\in\mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad Ax = y. \tag{3}$$

In the following, we will denote by $\Delta_1(y)$ the solution of the l_1-relaxation (3). Both methods are not comparable since OMP is a greedy algorithm with cheaper complexity than the l_1-relaxation whereas the l_1 approach usually enjoys better performances in terms of recovery at the price of a computational complexity equivalent to the one of linear programming. Indeed, (3) is equivalent to

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} z_i \quad \text{s.t.} \quad -z \le x \le z, \quad \text{and} \quad Ax = y \tag{4}$$

which is nothing but a linear program. The main subsequent problem induced by this choice of relaxation is to obtain simple to identify sufficient conditions on A for the relaxation to be exact, i.e. to produce the sparsest solution to the underdetermined system (1). Such a nice condition was given by Candes Romberg and Tao [3] and is now known as the Uniform Uncertainty Principle and also as the Restricted Isometry Property in certain recent papers.

The goal of our paper is to present a new method for solving the CS problem generalizing the original l_1-relaxation of [3] and with much better performance in pratice as measured by success rate of recovery versus original sparsity k.

2. The Alternating l_1 method

One important problem addressed in subsequent works and still of great interest now is the one of increasing the value of k for which every or most k-sparse signals can be reconstructed exactly for a given pair (n, m). We now present a generalization of the l_1 relaxation which we call the Alternating l_1 relaxation, with better experimental performances than the standard l_1 relaxation.

2.1. *Description of the method*

2.1.1. *The Lagrangian dual*

Recall that the problem of exact reconstruction of sparse signals can be solved using Δ_0 and Lemma 1.1. Let us start by writing down problem (2), to which Δ_0 is the solution map, as the following equivalent problem

$$\max_{z \in \{0,1\}^n, x \in \mathbb{R}^n} e^t z \quad \text{s.t.} \quad z_i x_i = 0, \quad i = 1, \ldots, n, \quad Ax = y \tag{5}$$

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 $z_i x_i = 0, \quad i = 1, \ldots, n$. However, these constraints can be merged into the unique constraint $\|D(z)x\|_1 = 0$, leading to the following equivalent problem

$$\max_{z \in \{0,1\}^n, x \in \mathbb{R}^n} e^t z \quad \text{s.t.} \quad \|D(z)x\|_1 = 0, \quad Ax = y. \tag{6}$$

Lagrangian duality is a very convenient framework for building convex relaxations to hard nonconvex optimization problems as demonstrated in [7]. In this framework, some constraints are kept implicit whereas the others are explicitely incorporated into the Lagrange function. Optimizing the lagrange function in the primal variables, e.g. x and z in (6), gives a dual function of the Lagrange multipliers. The main justification of the Lagrangian approach is that optimizing the dual function is a convex problem and can thus be solved efficiently most of the time. In the case of a maximization-type initial problem, the optimal value of the dual problem is an upper bound to the optimal value of the original problem. This property is called weak duality. Moreover, this optimal upper bound can be shown to be very sharp in some important instances like the Max-Cut problem [5]. Deciding the appropriate combination of implicit and explicit constraints can be quite tricky and there are few general rules. Choosing to make the hard constraints explicit often gives tighter bounds in general at the price of computationally harder convex relaxations. The main example for such phenomenon is the integer programming problem in n variables for which keeping the combinatorial constraints implicit leads to an easy linear programming problem in $\mathbb{R}^n$ whereas making them explicit gives a semidefinite programming (SDP) problem in $\mathbb{S}_n$, the cone of positive semidefinite real matrices of order n; see [7] for an in depth treatement.

Turning back to problem (6) and, for the purpose of keeping the computational complexity low, choosing to keep the constraints $Ax = y$ and $z \in \{0,1\}^n$ implicit, the Lagrangian function is given by

$$\mathcal{L}(x, z, u) = e^t z - u \|D(z)x\|_1 \tag{7}$$

where $D(z)$ is the diagonal matrix with diagonal vector equal to z. The dual function (with values in $\mathbb{R} \cup +\infty$) is defined by

$$\theta(u) = \max_{z \in \{0,1\}^n,\, x \in \mathbb{R}^n,\, Ax=y} \mathcal{L}(x, z, u) \tag{8}$$

and the dual problem is

$$\inf_{u \in \mathbb{R}} \theta(u). \tag{9}$$

Notice that if $u < 0$ then $\theta(u) = +\infty$. The main problem with the dual problem (9) is that the solutions to (8) are as difficult to obtain as the solution of the original problem (6) because of the nonconvexity of the Lagrangian function $\mathcal{L}$.

2.1.2. *A practical alternative*

Due to the difficulty of computing the dual function θ, the interest of the Lagrangian dual seems at first to be of pure theoretical nature only. In this section, we propose a simple but suboptimal alternating minimization approach.

Recovering the l_1 relaxation. When we fix $u > 0$ and we restrict z to the value $z = e$, solving the problem

$$\theta(u) = \max_{z=e,\, x \in \mathbb{R}^n,\, Ax=y} \mathcal{L}(x, z, u) \tag{10}$$

gives exactly the solution $\Delta_1(y)$ of the l_1 relaxation.

The Alternating l_1 relaxation. From the Lagrangian duality theory above, it may be supected that a better relaxation than the plain l_1 relaxation can be obtained by trying to optimize the Lagrangian even in a suboptimal manner.

Algorithm 1 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 $l \le N$ **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 $x_u^{(L)}$.

The following lemma says that at each step, the next support is first estimated by thresholding.

Lemma 2.1 *For all x in $\mathbb{R}^n$, any solution z of*

$$\max_{z\in[0,1]^n} \mathcal{L}(x,z,u) \tag{11}$$

satisfies that $z_i = 1$ if $|x_i| < \frac{1}{u}$, 0 if $|x_i| > \frac{1}{u}$ and $z_i \in [0,1]$ otherwise.

Proof. Problem (11) is clearly separable and the solution can be easily computed coordinatewise. □

Remark 1 This lemma proves in particular that the solution set of the convex problem (11) always contains a binary vector and the binary constraints $z \in \{0,1\}^n$ in the computation of $z_u^{(l)}$ can be relaxed to $z \in [0,1]^n$. As a consequence, we may conclude that each step of the algorithm is computable in polynomial time.

2.2. *Open problems*

Leaving aside the Lagrangian dual problem for the moment, a fully rigorous analysis of the rudimentary Alternating l_1 algorithm seems quite challenging already. However, we have the two following basic properties:

- Taking the suboptimal choice $z_u^{(l)} = e$ at each step l gives the standard l_1 relaxation.
- Using Lemma 2.1, the computation of $x_u^{(l)}$ becomes

$$x_u^{(l)} \in \operatorname{argmax}_{x\in\mathbb{R}^n,\ Ax=y} \sum_{i \text{ s.t.} (z_u^{(l)})_i=1} |x_i|, \tag{12}$$

  and thus, the number of components of x taken into account in the l_1 objective funtion will hopefully be lower than n.

Based on this, if one is allowed to expect that minimizing the l_1-norm over a smaller set of components increases the number of detected zero components, the Alternating l_1 approach should improve over the plain l_1. It is an open problem however to write a precise quantitative result supporting this intuition. Another important question would be to know when the alternating procedure does provide a solution to the optimization problem in the very definition (8) of θ in the case where the number of iterations L goes to $+\infty$, and when an ϵ-approximate solution can be extracted from this sequence within polynomial time. Based on such results, one could safely try and generalize the approach by associating a Lagrange multiplier to each constraint $|x_i z_i| = 0$ and attack the resulting Lagrangian dual problem using modern non-smooth optimization algorithms such as bundle methods [6].

One further remark that needs to be made is that, contrarily to the SDP approach for combinatorial problems, we chose to keep the combinatorial constraints implicit. Thus, our approach still lets possible the option of incorporating these constraints into the Lagrange function and obtain a richer dual function with maybe better solutions than the present rudimentary alternating l_1 approach. Further investigations should be made in this direction in order to decide wether the increase in computational complexity induced by such a generalization is worthwhile.

2.3. *Monte Carlo experiments*

Comparison between the success rate of ℓ_1 and Alternating ℓ_1 is shown in Figure 1. Optimization of the Lagrange multiplier u was performed using coarse dichotomic search and we finally used $u = 3$ uniformly over all experiments in order to reduce the simulation's complexity, keeping however in mind that experiment-wise optimization of u should give better results, even though probably not much better. We also chose the value $L = 4$ iterations in the Alternating l_1 in order to make fair comparison with Boyd, Candes and Wakin's recent proposal [2] called the Reweighted l_1 relaxation. Our proposal outperformed both the plain l_1 and the Reweighted l_1 relaxations for the given data sizes. Moreover, we also observed that the optimal value of $z^{(l)}$ was most of the time obtained after the first step of the alternating procedure and thus, we could as well have taken $L = 1$ most of the time, dividing in this way by four the computational effort when compared to Candes, Wakin and Boyd's Reweighted l_1 relaxation. Further experiments are currently being performed in order to explore the practical efficiency of the Alternating l_1 approach in greater details.

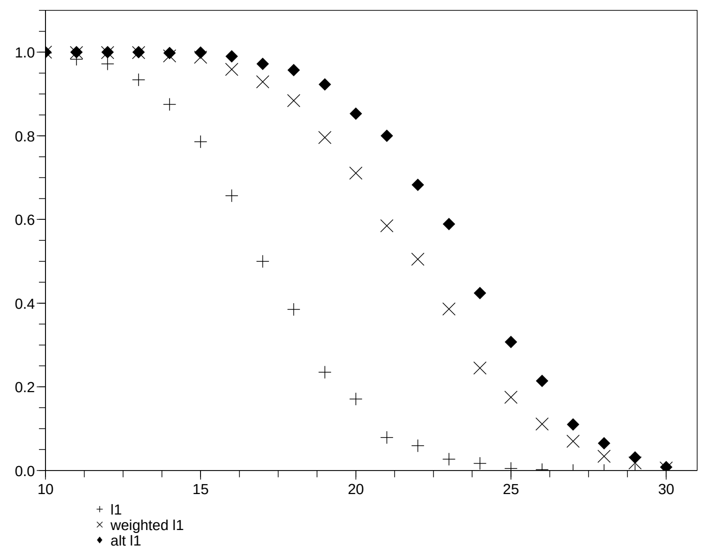

Figure 1. Rate of success over 1000 Monte Carlo experiments in recovering the support of the signal vs. signal sparsity k for $n = 128$, $m = 50$, $L = 4$, $u = 3$. A and nonnul components of x were drawn from the gaussian $\mathcal{N}(0, 1)$ distribution. In Boyd, Candes and Wakin's new Reweighted l_1 relaxation we chose $\epsilon = .1$, the best value found in [2].

Acknowledgements. The author would like to thank E. Candès for his great hospitality during a stay at Caltech and for extremely fruitful discussions together with Y. Plan concerning the proposed method.

References

[1] E. Candes, *Compressive sampling*, International Congress of Mathematics, vol. 3, 2006, pp. 1433–1452.

[2] E. Candes, Wakin M., and Boyd S., *Enhancing sparcity by reweighted l_1 minimization*, preprint (2007).

[3] E. Candes, J. Romberg, and Tao T., *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Information Theory **2** (2006), no. 52, 489–509.

[4] A. Cohen, W. Dahmen, and DeVore R., *Compressed sensing and best k-term approximation*, Preprint (2006).

[5] M. Goemans and D. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. Assoc. Comput. Mach. **42** (1995), no. 6, 1115–1145.

[6] J.-B. Hiriart Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms ii: Advanced theory and bundle methods*, Grundlehren der Mathematischen Wissenschaften, vol. 306, Springer- Verlag, 1993.

[7] C. Lemaréchal and F. Oustry, *Sdp relaxations in combinatorial optimization from a lagrangian viewpoint*, Advances in convex analysis and global optimization (Pythagorion, 2000), Nonconvex Optim. Appl., vol. 54, 2001, pp. 119–134.