Computational Aspects of Constrained L_1-L_2 Minimization for Compressive Sensing

Yifei Lou[1], Stanley Osher[2], and Jack Xin[3]

[1] University of Texas Dallas
[2] University of California Los Angeles
[3] University of California Irvine

Abstract. We study the computational properties of solving a constrained L_1-L_2 minimization via a difference of convex algorithm (DCA), which was proposed in our previous work [13, 19] to recover sparse signals from a under-determined linear system. We prove that the DCA converges to a stationary point of the nonconvex L_1-L_2 model. We clarify the relationship of DCA to a convex method, Bregman iteration [20] for solving a constrained L_1 minimization. Through experiments, we discover that both L_1 and L_1-L_2 obtain better recovery results from more coherent matrices, which appears unknown in theoretical analysis of exact sparse recovery. In addition, numerical studies motivate us to consider a weighted difference model L_1-αL_2 ($\alpha > 1$) to deal with ill-conditioned matrices when L_1-L_2 fails to obtain a good solution.

1 Introduction

Compressive sensing (CS) [7, 9] is about acquiring or recovering a *sparse* signal (a vector with most elements being zero) from an under-determined linear system $Au = b$, where b is the data vector and A is a $M \times N$ matrix for $M < N$. Mathematically, it amounts to solving a constrained optimization problem,

$$\min \|u\|_0 \quad \text{s.t.} \quad Au = b, \tag{1}$$

where $\|\cdot\|_0$ is the L_0 norm, which counts the number of non-zero elements. Minimizing the L_0 norm is equivalent to finding the sparsest solution. Since L_0 minimization is NP-hard [14], a popular approach is to replace L_0 by a convex L_1 norm, which often gives a satisfactory sparse solution. A major step for CS was the derivation of the restricted isometry property (RIP) [3], which guarantees to recover a sparse signal by minimizing the L_1 norm. It is also proved in [3] that Gaussian random matrices satisfy the RIP with high probability. A deterministic result in [6, 8, 11] says that exact sparse recovery via L_1 minimization is possible if

$$\|u\|_0 < \frac{1 + 1/\mu}{2}, \tag{2}$$

where μ is the mutual coherence of A, defined as

$$\mu(A) = \max_{i \neq j} \frac{|\mathbf{a}_i^T \mathbf{a}_j|}{\|\mathbf{a}_i\| \|\mathbf{a}_j\|}, \quad \text{with } A = [\mathbf{a}_1, \cdots, \mathbf{a}_N]. \tag{3}$$

The inequality (2) suggests that L_1 may not perform well for highly coherent matrices in that if $\mu \sim 1$, then $\|u\|_0$ is 1 at most.

Recently, there has been a surge of activities in deploying nonconvex penalties to promote sparsity while solving the linear system, because the convex L_1 norm may not perform well on some practical problems with coherent sensing matrices. In our previous studies [13,19], we advocate the use of a nonconvex functional L_1-L_2, as opposed to L_p for $p \in (0,1)$ in [5,12,18]. To minimize L_1-L_2, a difference of convex algorithm (DCA) [16] is applied. In [13], we conduct an extensive study comparing the sparse penalties, $L_0, L_1, L_p (0 < p < 1), L_1 - L_2$, and their numerical algorithms. Numerical experiments demonstrate that L_1-L_2 is always better than L_1 to promote sparsity, and using DCA for L_1-L_2 is better than iterative reweighted algorithms for L_p minimization [5, 12] when the sensing matrix exhibits high coherence.

The contributions of this work are four-fold. First, we prove that the DCA iterations for a constrained minimization converge to a stationary point (Section 2). Second, we clarify the relation of DCA to the Bregman iteration [20], which is designed for convex functional minimization (Section 3). Third, we analyze how coherence, sparsity and minimum separation (MS) contribute to exact recovery of a sparse vector from an under-determined system, and discover that both L_1 and L_1-L_2 get better recovery results towards high coherence (Section 4). Lastly, we demonstrate that exact recovery is highly correlated with DCA converging in a few steps, and then propose a weighted difference model, minimizing L_1-αL_2 for $\alpha > 1$, to improve the reconstruction accuracy when L_1-L_2 fails to find the exact solution (Section 5).

2 Constrained L_1-L_2 Minimization

Replacing L_0 in (1) by L_1-L_2, we get a constrained minimization problem,

$$\min_{u \in \mathbb{R}^N} \|u\|_1 - \|u\|_2 \quad \text{s.t.} \quad A\,u = b. \tag{4}$$

The idea of DCA [16] involves linearizing the second (nonconvex) term in the objective function at the current solution, and advancing to a new one by solving a L_1 type of subproblem, *i.e.*,

$$u_{n+1} = \arg\min \left\{ \|u\|_1 - \langle q_n, u \rangle \quad \text{s.t.} \quad Au = b \right\}, \tag{5}$$

for $q_n = \frac{u_n}{\|u_n\|_2}$. We introduce two Lagrange multipliers y, z in an augmented Lagrangian,

$$L_{\lambda,\rho}(u,v,y,z) = \lambda\|v\|_1 - \lambda\, q_n^T u + \rho y^T(u-v) + z^T(Au-b) + \frac{\rho}{2}\|u-v\|^2 + \frac{1}{2}\|Au-b\|^2.$$

Then an alternating direction of multiplier method (ADMM) [1] is applied to solve eq. (5), by alternatively updating each variable (u, v, y and z) to minimize $L_{\lambda,\rho}$. Please refer to Figure 1 for pseudo-code. Note that the update of v can

be solved efficiently by a *soft shrinkage* operator: $v = \text{shrink}(u + y, \lambda/\rho)$, where $\text{shrink}(s, \gamma) = \text{sgn}(s)\max\{|s| - \gamma, 0\}$.

Next we will prove that the DCA sequence $\{u_n\}$, defined in eq. (5), converges to a stationary point of eq. (4). The convergence proof of an unconstrained L_1-L_2 minimization can be found in [19]. For general DCA, convergence is guaranteed if the optimal value is finite and the sequence generated by DCA is bounded [16,17]. Here we can prove that the DCA sequence is bounded with probability 1, due to degree-1 homogeneity of L_1-L_2 and the fact that a nonzero signal is 1-sparse if and only if its L_1-L_2 is zero (see Lemma 1).

Lemma 1. *Suppose $u \in \mathbb{R}^N \setminus \{\mathbf{0}\}$ and $\|u\|_0 = s$, then*

$$\|u\|_1 - \|u\|_2 = 0 \quad \textit{if and only if} \quad s = 1.$$

Please refer to [19] for the proof.

Lemma 2. *The objective function $E(u) = \|u\|_1 - \|u\|_2$ is monotonically decreasing for the DCA sequence $\{u_n\}$ defined in eq. (5).*

Proof. We want to show that

$$0 \leqslant E(u_n) - E(u_{n+1}) = \|u_n\|_1 - \|u_n\|_2 - \|u_{n+1}\|_1 + \|u_{n+1}\|_2. \tag{6}$$

The first-order optimality condition for a constrained problem (5) can be formulated as

$$\frac{\partial \mathcal{L}}{\partial u} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \nu} = 0, \tag{7}$$

where $\mathcal{L}(u, \nu) = \|u\|_1 - q_n^T u + \nu^T(Au - b)$ is the Lagrangian. Since L_1 norm is not differentiable, we consider a subgradient $p_{n+1} \in \partial\|u_{n+1}\|_1$, and hence eq. (7) is equivalent to

$$p_{n+1} - q_n + A^T\nu = 0 \quad \text{and} \quad Au_{n+1} = b. \tag{8}$$

Left multiplying the first equation in (8) by $u_n - u_{n+1}$ gives

$$\begin{aligned} 0 &= \langle p_{n+1} - q_n + A^T\nu, u_n - u_{n+1}\rangle \\ &= \langle p_{n+1}, u_n\rangle + \langle q_n, u_{n+1}\rangle - \|u_{n+1}\|_1 - \|u_n\|_2, \end{aligned} \tag{9}$$

where we use $\langle p_{n+1}, u_{n+1}\rangle = \|u_{n+1}\|_1$, $\langle q_n, u_n\rangle = \|u_n\|_2$, and $Au_n = Au_{n+1} = b$. Substituting eq. (9) into eq. (6), we get

$$E(u_n) - E(u_{n+1}) = (\|u_n\|_1 - \langle p_{n+1}, u_n\rangle) + (\|u_{n+1}\|_2 - \langle q_n, u_{n+1}\rangle). \tag{10}$$

Since any subgradient of L_1 norm has the property that $|p_{n+1}^{(i)}| \leqslant 1$ for all $1 \leq i \leq N$, we have $\|u_n\|_1 \geq \langle p_{n+1}, u_n\rangle$. The second term in eq. (10), $\|u_{n+1}\|_2 - \langle q_n, u_{n+1}\rangle \geqslant 0$ is due to Cauchy-Schwarz inequality. Therefore, we proved that $E(u_n) \geqslant E(u_{n+1})$. □

Lemma 3. *The DCA sequence $\{u_n\}$ is bounded with probability 1.*

Proof. It follows from Lemma 2 that $E(u_n)$ is bounded, and hence there exists a constant C so that

$$\|u_n\|_1 - \|u_n\|_2 \leq C. \tag{11}$$

Write $u_n = \|u_n\|_2 \cdot u_n/\|u_n\|_2$, or a polar decomposition into amplitude and phase. By degree-1 homogeneity, eq. (11) is:

$$\|u_n\|_2(\|\frac{u_n}{\|u_n\|_2}\|_1 - \|\frac{u_n}{\|u_n\|_2}\|_2) \leq C.$$

Suppose $\|u_n\|_2$ diverges (up to a sub-sequence, but denoted the same), then

$$\|\frac{u_n}{\|u_n\|_2}\|_1 - \|\frac{u_n}{\|u_n\|_2}\|_2 \to 0, \text{ as } n \to \infty.$$

Since $u_n/\|u_n\|_2$ is compact (on unit sphere), it converges to a limit point, denoted as u_*, on the unit sphere (up to a sub-sequence). Hence, $\|u_*\|_1 - \|u_*\|_2 = 0$, implying $\|u_*\|_0 = 1$ by Lemma 1.

On the other hand, the DCA sequence satisfies $Au_n = b$. Dividing by $\|u_n\|_2 \to \infty$, we find that $Au_* = 0$, so u_* is in Ker(A). Unless A has a zero column (or one component of u is absent in the constraint), Ker(A) does not contain a one-sparse vector, which is a contradiction.

So if A has no zero column (which happens with probability 1 for random matrices from continuous distribution), we conclude that u_n is bounded.

Theorem 1. *Any non-zero limit point u_* satisfies the first-order optimality condition, which means u_* is a stationary point.*

Proof. The objective function E is monotonically decreasing by Lemma 2, and bounded from below, so $E(u^n)$ converges and hence we have $\|u_{n+1}\|_2 - \langle q_n, u_{n+1}\rangle \to 0$, which implies that $u_n - u_{n+1} \to 0$ as $u_n \neq \mathbf{0}$.

As the sequence $\{u_n\}$ is bounded by Lemma 3, there exists (by definition) a subsequence of $\{u_n\}$ converging to a limit point u_*. The subsequence is denoted as $\{u_{n_k}\}$. The optimality condition at the n_k-th step of DCA is

$$q_{n_k-1} - A^T\nu \in \partial\|u_{n_k}\|_1 \quad \text{and} \quad Au_{n_k} = b. \tag{12}$$

Since $u_{n_k} \to u_*$ and subgradient of L_1 norm is a closed set, we have $\partial\|u_{n_k}\|_1 \subseteq \partial\|u_*\|_1$. Letting $n_k \to \infty$, we get

$$q_* - A^T\nu \in \partial\|u_*\|_1 \quad \text{and} \quad Au_* = b, \tag{13}$$

which means that u_* satisfies the first-order optimality condition and hence it is a stationary point. □

3 Study A: DCA v.s. Bregman Iteration

We want to point out that DCA is related to Bregman iteration, which was derived from Bregman divergence [2], defined as

$$D_J^p(u, v) := J(u) - J(v) - \langle p, u - v\rangle, \tag{14}$$

where $J(\cdot)$ is a convex functional, and $p \in \partial J(v)$ is a subgradient of J at the point v. For constrained L_1 minimization,

$$\min_{u\in\mathbb{R}^N} \|u\|_1 \quad \text{s.t.} \quad Au = b, \tag{15}$$

Bregman iteration incorporates the Bregman divergence into an unconstrained formulation, and advances to a new solution u_{n+1} based on a Taylor expansion of $J(u) = \|u\|_1$ at current step, *i.e.*, $u_{n+1} = \arg\min_u \lambda D_J^p(u, u_n) + \frac{1}{2}\|Au - b\|_2^2$. The optimality condition is

$$\lambda(p_{n+1} - p_n) + A^T(Au_{n+1} - b) = 0. \tag{16}$$

Summing from 0 to $n+1$, we have $\lambda p_{n+1} + \sum_{k=0}^{n+1} A^T(Au_k - b) = 0$, or

$$\begin{cases} \lambda p_{n+1} + A^T(Au_{n+1} - v_n) = 0 \\ v_n = \sum_{k=1}^{n}(b - Au_k), \end{cases} \tag{17}$$

for $p_0 = u_0 = 0$, which is equivalent to

$$\begin{cases} u_{n+1} = \arg\min \lambda\|u\|_1 + \frac{1}{2}\|Au - v_n\|_2^2 \\ v_{n+1} = v_n + b - Au_{n+1}. \end{cases} \tag{18}$$

We consider the same idea for L_1-L_2. In particular, we get an optimality condition by lagging the second term,

$$\lambda(p_{n+1} - p_n) - \lambda(q_n - q_{n-1}) + A^T(Au_{n+1} - b) = 0, \tag{19}$$

where p and q be subgradients of $\|u\|_1$ and $\|u\|_2$ respectively. Summing from 0 to $n+1$ and letting $p_1 = q_0 = z_1 = 0$, we obtain

$$\begin{cases} \lambda p_{n+1} - \lambda q_n + A^T(Au_{n+1} - z_n) = 0 \\ z_{n+1} = z_n + (b - Au_{n+1}), \end{cases} \tag{20}$$

for $z_n = \sum_{k=1}^{n} b - Au_k$. The first equation in (20) is equivalent to $u_{n+1} = \arg\min \lambda\|u\|_1 - \lambda\langle q_n, u\rangle + \frac{1}{2}\|Au - z_n\|_2^2$, which can be solved via ADMM.

DCA+ADMM (Alg.1) and Bregman+ADMM (Alg.2) are summarized in Fig. 1, which shows that their difference lies in the update of z and q (compare boxed lines in Figure 1). For DCA, z is updated MaxOuter iterations and then q is updated, while Bregman iteration updates z and q simultaneously. We plot relative errors of each inner solution to the ground-truth versus computational time in Fig. 2, which illustrates that DCA is more computationally efficient than Bregman iteration.

The update of z is to account for the constraint $Au = b$, which is enforced by DCA at every inner iteration. This constraint also plays an important role in proving DCA's convergence (see Theorem 1). As for Bregman iteration, Osher *et. al.* [15] proved that Bregman iteration (18) converges if the regularization function is convex, while convergence analysis for nonconvex formulation (20) is subject to future investigation.

Define MaxOuter, MaxInner and initialize $u \neq 0, v = y = z = 0$

Alg.1: DCA + ADMM

for 1 **to** MaxOuter **do**
  for 1 **to** MaxInner **do**
    $u = (A^TA + \rho I)^{-1}(A^T(b - z) + \rho(y + v) + \lambda q)$
    $v = shrink(u + y, \lambda/\rho)$
    $y = y + (u - v)$
    $z = z + Au - b$
  end for
  $q = u/\|u\|_2$
end for

Alg.2: Bregman + ADMM

for 1 **to** MaxOuter **do**
  for 1 **to** MaxInner **do**
    $u = (A^TA + \rho I)^{-1}(A^T(b - z) + \rho(y + v) + \lambda q)$
    $v = shrink(u + y, \lambda/\rho)$
    $y = y + (u - v)$
  end for
  $z = z + Au - b$
  $q = u/\|u\|_2$
end for

Fig. 1. Pseudo-codes for DCA+ADMM (left) and Bregman+ADMM (right).

4 Study B: Sparsity v.s. Coherence

Theoretically, the success of L_1 depends on the RIP or incoherence condition. Unfortunately, RIP is difficult to verify for a given matrix, while incoherence is not strong enough to account for exact recovery. Therefore, we are interested in non-RIP conditions to evaluate the performance of L_1 or L_1-L_2, which will contribute a better characterization of sparse solutions.

We consider a family of randomly oversampled partial discrete cosine transform (DCT) matrices of the form

$$A = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \in \mathbb{R}^{M\times N} \quad \text{with} \quad \mathbf{a}_j = \frac{1}{\sqrt{M}} \cos(\frac{2\pi \mathbf{w} j}{F}),\ j = 1, \cdots, N, \quad (21)$$

where $\mathbf{w}$ is a random vector uniformly distributed in $[0, 1]^M$. This matrix arises in spectral estimation [10], if the cosine function in (21) is replaced by complex exponential. The coherence of this type of matrices is controlled by F in the sense that larger F corresponds to larger coherence.

We then generate random sparse vectors as ground-truth, denoted as u_g, whose sparsity (L_0 norm) is S with nonzero elements being at least R distance apart, referred to as *minimum separation*. Let $b = Au_g$, and u_* is a reconstructed solution, from L_1 minimization using Bregman iteration (18) or L_1-L_2 minimization using DCA (5). We consider the algorithm successful, if the relative error of u_* to the ground truth u_g is less than .001, *i.e.*, $\frac{\|u_* - u_g\|}{\|u_g\|} < .001$.

We analyze whether success rates (based on 100 random realizations) are related to coherence (F), sparsity (S), and minimum separation (R) using randomly oversampled DCT matrices of size 100×2000. We include the discussion of MS here, due to the work of [4], which suggests that sparse spikes need to be further apart for more coherent matrices. However, we observe that MS seems to play a minor role in sparse recovery when it is above $2F$, a theoretical lower bound [4], as indicated by the first plot in Fig. 3 showing that success rates as

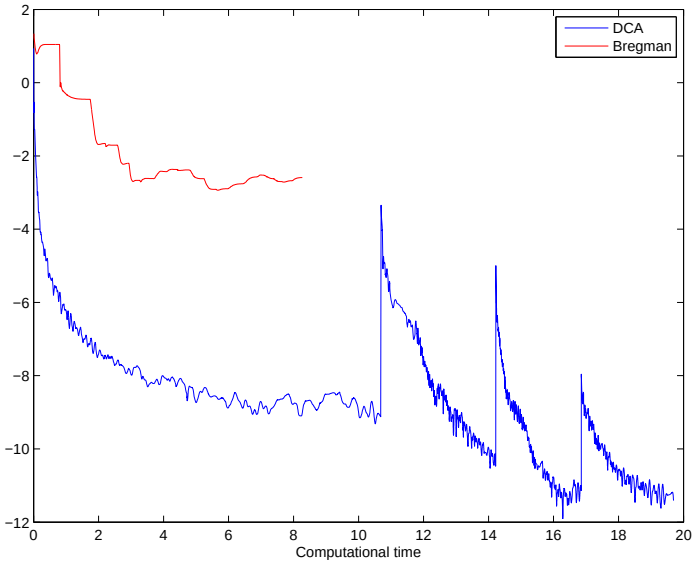

Fig. 2. Comparison between DCA and Bregman in terms of relative errors to the ground-truth solution versus computational time (sec.), which shows that DCA is more effiient than Bregman iteration. The stopping condition for outer iterations is $\|x_n - x_{n-1}\|/\|x_{n-1}\| < 1e-4$, and Bregman stops eariler than DCA after about 8 seconds.

a function of S are almost identical for different R. The second plot in Fig. 3 is success rates as a function of F for different S while $R = 50$, which suggests that sparsity is the most important factor that contributes to exact sparse recovery, compared to MS and coherence.

In Fig. 4, we examine the success rates of using L_1 and L_1-L_2 as a function of R, while S is fixed to be 25. The two plots illustrate that the success rates of $F = 10, 15$ are higher than that of $F = 5$, which implies that more coherent matrices yield better recovery rates for both L_1 and L_1-L_2. This phenomenon appears new in L_1 sparse recovery, which is worthy of future study.

5 Study C: Exact Recovery v.s. DCA Convergence

We find that if DCA converges in a few iterations (say 3-5), the reconstructed solution coincides with the ground-truth solution with high probability. It follows from Theorem 1 that DCA sequence always converges to a stationary point. We consider a stopping condition of DCA to be $\frac{\|u_{n+1}-u_n\|_2}{\|u_n\|} < .001$. In this section, we say DCA converges if the number of iterations is less than 10. Table 1 is the confusion matrix or joint occurrence of whether DCA converges and whether the algorithm finds the exact solution, which illustrates exact recovery is highly correlated with DCA converging in a few steps.

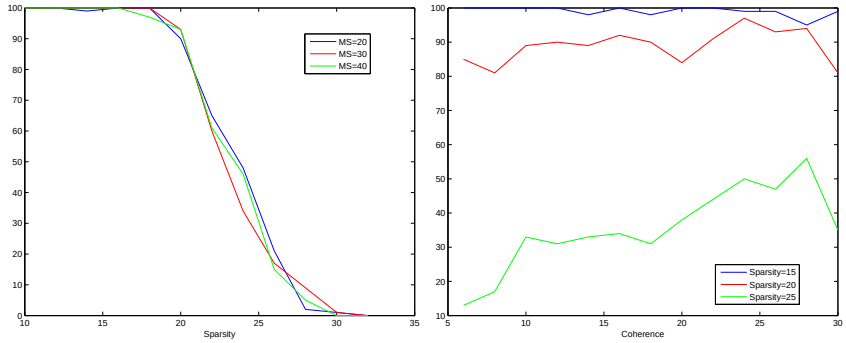

Fig. 3. Success rates of recovering sparse signals from randomly oversampled partial DCT matrices of size 100×2000 as a function of S for different R while $F = 10$ (left) and as a function of F for different S while $R = 50$ (right). Both plots suggest that sparsity is the most important factor in sparse recovery, compared to MS and coherence.

Table 1. Confusion matrix of whether DCA converges and whether the algorithm finds the exact solution.

	converge	not converge
exact	8104	17
not exact	445	5732

We examine the relative errors between two consecutive DCA iterations for both converging and non-converging cases. Fig. 5 shows that the relative errors monotonically decreases, and DCA stops in 3-5 steps. On the other hand, if the relative errors are oscillatory, it often implies that the algorithm does not converge (within 10 iterations).

We found that one reason that DCA does not converge is that L_1-L_2 of the exact solution is larger than that of some DCA iterates u_n, and hence the algorithm jumps among these local mimima. This observation suggests that L_1-L_2 is unable to promote sparsity in some degenerate cases. As a remedy, we propose a weighted difference model L_1-αL_2 with more weight on the nonconvex term ($\alpha > 1$). We find that the weighted difference model sometimes improves the recovery rate, though there is no convergence proof for $\alpha > 1$, as the objective function is not bounded from below. To illustrate this phenomenon numerically, we consider to apply the DCA for L_1-$2L_2$, if the DCA for L_1-L_2 does not converge within 10 iterations; otherwise, we use the solution of L_1-L_2 to be the one in lieu of L_1-$2L_2$. The success rates for both $\alpha = 1$ and $\alpha = 2$ are then recorded in Tables 2, which report at least 10% improvement for most testing cases.

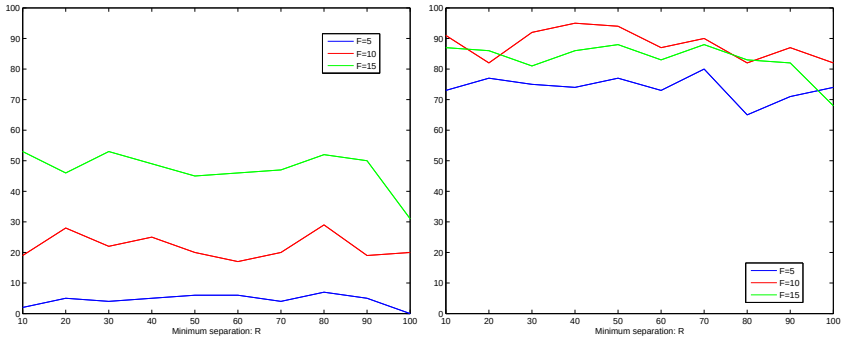

Fig. 4. Success rates of using L_1 (left) and L_1-L_2 (right) as a function of R while fixing $S = 20$ for different F showing that larger coherence yields better recovery rates.

Table 2. Success rates (%) of L_1-L_2 and L_1-$2L_2$.

$F = 10, R = 50, S =$	18	21	24	27	30	Total
L_1-L_2	97	78	34	10	0	219
L_1-$2L_2$	98	81	40	17	4	240
improve	1.03	3.85	17.65	70.00	inf	9.59
$F = 10, S = 25, R =$	20	30	40	50	60	Total
L_1-L_2	28	33	25	36	26	216
L_1-$2L_2$	37	37	34	41	32	266
improve	32.14	12.12	36.01	13.89	23.08	23.15

6 Study D: Constrained V.S. Unconstrained

We now compare the constrained L_1-L_2 minimization (4) with the unconstrained version, *i.e.*,

$$\min_{u \in \mathbb{R}^N} \lambda(\|u\|_1 - \|u\|_2) + \frac{1}{2}\|Au - b\|_2^2, \tag{22}$$

which is studied thoroughly in [19]. The constrained formulation is a parameter-free model, while two auxiliary variables are introduced in the ADMM algorithm. In all experiments, we choose $\lambda = 2, \rho = 10$. For the unconstrained formulation, a small λ is chosen to enforce $Au = b$ implicitly. Here we choose $\lambda = 10^{-5}$ in (22). The comparison between constrained and unconstrained formulations for both L_1 and L_1-L_2 is given in Fig. 6, which shows that the two optimization problems yield similar performance when λ is small.

7 Conclusion and Future Work

This paper studied the computational aspects of a constrained L_1-L_2 minimization, as an alternative to the conventional L_1 approach, to recover sparse

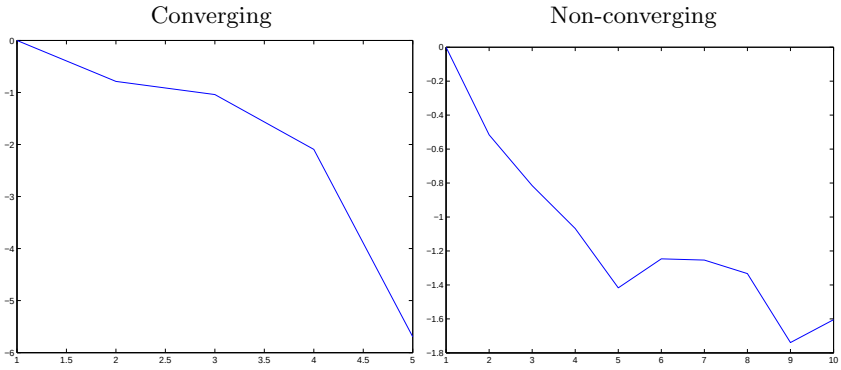

Fig. 5. Relative errors (in logarithmic scales) betweent two consecutive DCA iterations for converging (left) and non-converging (right) cases. If DCA converges in a few steps (3-5), it gives an exact recovery with high probability.

signals from an under-determined system. The DCA is applied to solve this nonconvex model with guaranteed convergence to a stationary point. The relation of DCA for a nonconvex model with a convex method, Bregman iteration, was also presented. Numerical experiments demonstrated that more coherent matrices give better recovery results and we proposed a weighted difference model to improve the reconstruction results, when L_1-L_2 is not sharp enough to promote sparse solution.

As for future directions, we will analyze the convergence of Bregman iteration applied to a nonconvex model, *i.e.*, eq. (20). Furthermore, we will devote ourselves to the understanding of the peculiar phenomenon of larger coherence giving better results. As for the weighted difference model, we will further explore the possibilities of stabilizing the resulting algorithm, and adaptively choosing the weighting parameter α with respect to the matrix A in eq. (1).

Acknowledgments. We would like to thank the anonymous referee for pointing out a general convergence property of DCA and a suggestion to reorganize our previous proof of convergence. Stanley Osher was partially supported by ONR grants N000141210838, N000141410444, NSF DMS-1118971, and the Keck Foundation. Jack Xin was partially supported by NSF grant DMS-1211179.

References

1. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (Jan 2011)

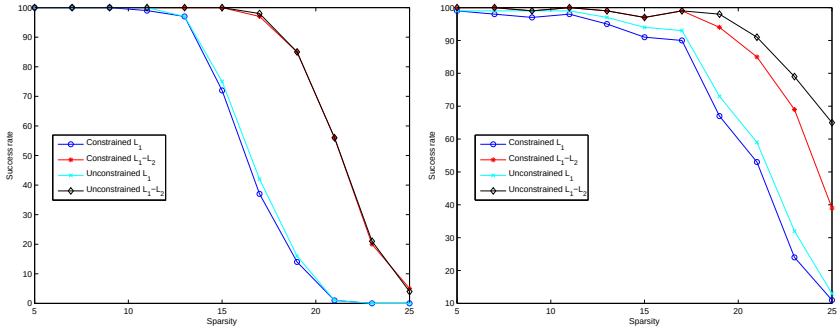

Fig. 6. The comparison between constrained and unconstrained formulations for $F = 5$ (left) and $F = 20$ (right), which shows that they give similar performance with the model parameter in unconstrained minimization problem properly chosen.

2. Bregman, L.: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. USSR Comp. Math. Math. Phys. (7), 200–217 (1967)
3. Candès, E.J., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. 59, 1207–1223 (2006)
4. Candès, E., Fernandez-Granda, C.: Super-resolution from noisy data. J. Fourier Anal. Appl. 19(6), 1229–1254 (2013)
5. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP). pp. 3869–3872 (2008)
6. Donoho, D., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. Proc. Nat. Acad. Scien. 100, 2197–2202 (2003)
7. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory 52(4), 1289–1306 (2006)
8. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. Information Theory, IEEE Transactions on 47(7), 2845–2862 (2001)
9. E. J. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Process. Mag. 25(2), 21–30 (2008)
10. Fannjiang, A., Liao, W.: Coherence pattern-guided compressive sensing with unresolved grids. SIAM J. Imaging Sci. 5(1), 179–202 (Feb 2012)
11. Gribonval, R., Nielsen, M.: Sparse representations in unions of bases. IEEE Trans. Inf. Theory 49(12), 3320–3325 (2003)
12. Lai, M.J., Xu, Y., Yin, W.: Improved iteratively reweighted least squares for unconstrained smoothed lq minimization. SIAM J. Numer. Anal. 5(2), 927–957 (2013)
13. Lou, Y., Yin, P., He, Q., Xin, J.: Computing sparse representation in a highly coherent dictionary based on difference of l1 and l2. J. Sci. Comput., to appear (2014)
14. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. pp. 227–234 (1995)
15. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterated regularization method for total variation-based image restoration. Multiscale Model. Simul (4), 460–489 (2005)

16. Pham-Dinh, T., Le-Thi, H.A.: Convex analysis approach to d.c. programming: Theory, algorithms and applications. Acta Math. Vietnam. 22(1), 289–355 (1997)
17. Pham-Dinh, T., Le-Thi, H.A.: A d.c. optimization algorithm for solving the trust-region subproblem. SIAM J. Optim. 8(2), 476–505 (1998)
18. Xu, Z., Chang, X., Xu, F., Zhang, H.: $l_{1/2}$ regularization: A thresholding representation theory and a fast solver. IEEE Trans. Neural Netw. 23, 1013–1027 (2012)
19. Yin, P., Lou, Y., He, Q., Xin, J.: Minimization of $l_1 - l_2$ for compressed sensing. SIAM J. Sci. Comput., to appear (2014)
20. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for l1 minimization with applications to compressed sensing. SIAM J. Imaging Sci. 1, 143–168 (2008)