Noname manuscript No.
(will be inserted by the editor)

Point Source Super-resolution Via Non-convex L_1 Based Methods

Yifei Lou · Penghang Yin · Jack Xin

Abstract We study the super-resolution (SR) problem of recovering point sources consisting of a collection of isolated and suitably separated spikes from only the low frequency measurements. If the peak separation is above a factor in $(1, 2)$ of the Rayleigh length (physical resolution limit), L_1 minimization is guaranteed to recover such sparse signals. However, below such critical length scale, especially the Rayleigh length, the L_1 certificate no longer exists. We show several local properties (local minimum, directional stationarity, and sparsity) of the limit points of minimizing two L_1 based nonconvex penalties, the difference of L_1 and L_2 norms (L_{1-2}) and capped L_1 (CL_1), subject to the measurement constraints. In one and two dimensional numerical SR examples, the local optimal solutions from difference of convex function algorithms outperform the global L_1 solutions near or below Rayleigh length scales either in the accuracy of ground truth recovery or in finding a sparse solution satisfying the constraints more accurately.

Keywords Super-Resolution · Rayleigh Length · L_{1-2} · Capped L_1 · Difference of Convex Algorithm (DCA)

1 Introduction

Super-resolution (SR), as its name states, aims at enhancing the resolution of a sensing system, in which the resolution is limited by hardware such as lens and sensors. It is closely related to interpolation [23] in the sense of filling in information on an unknown fine grid based on what is available on the coarse grid.

Y. Lou
Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080. E-mail: yifei.lou@utdallas.edu. YL was partially supported by NSF grant DMS-1522786.

P. Yin and J. Xin
Department of Mathematics, UC Irvine, Irvine, CA 92697. PY and JX were partially supported by NSF grants DMS-1222507 and DMS-1522383.

As particularly useful in imaging applications, such as high-definition television and retina display used in Apple products, SR is often cast as an image reconstruction problem, for which some methods are directly transplanted onto SR, *e.g.*, total variation [24], non-local means [31], and sparse dictionary representation [39]. For other SR methods, please refer to two survey papers [3, 26] and references therein.

The super-resolution problem addressed in this paper is different to image zooming or magnification, but aiming to recover a real-valued signal from its low-frequency measurements. A mathematical model is expressed as

$$b_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_t e^{-i2\pi kt/N}, \qquad |k| \leq f_c, \tag{1}$$

where $x \in \mathbb{R}^N$ is a vector of interest, and $b \in \mathbb{C}^n$ is the given low frequency information with $n = 2f_c{+}1$ $(n < N)$. This is related to super-resolution in the sense that the underlying signal x is defined on a fine grid with spacing $1/N$, while we only observe the lowest n Fourier coefficients, which implies that we can only expect to recover the signal on a coarser grid with spacing $1/n$. For simplicity, we use matrix notations to rewrite eq. (1) as $b = S_n \mathcal{F} x$, where S_n is a sampling matrix by collecting the lowest n frequency coefficients, $\mathcal{F}$ is the Fourier transform, and we denote $\mathcal{F}_n = S_n \mathcal{F}$. The frequency cutoff induces a resolution limit inversely proportional to f_c; below we set $\lambda_c = 1/f_c$, which is referred to as Rayleigh length (a classical resolution limit of hardware [19]). Hence, a super-resolution factor (SRF) can be interpreted as the ratio between the spacing in the coarse and fine grids, *i.e.*, $\mathrm{SRF} = N/n \approx 0.5\lambda_c N$.

We are interested in superresolving point sources. It is particularly useful in astronomy [32], where blurred images with point sources need to be cleaned or super-resolved. Suppose x is composed of points sources, *i.e.*, $x = \sum_{t_j \in T} c_j \delta_{t_j}$, where δ_τ is a Dirac measure at τ, spikes of x are located at t_j belonging to a set T, and c_j are coefficients. Denote $K = |T|$ be the cardinality of the set T, and sparsity assumption implies that $K \ll N$. Recently, sparse recovery problem becomes popular due to rapid advances in compressive sensing (CS) [13]. The provable performance of CS methods relies on either restricted-isometry property (RIP) [4] or incoherent measurements [35, 36]. Unfortunately for SR, the sensing matrix $\mathcal{F}_n$ is highly coherent [17]. Consequently, sparse SR deserves special attention, which may lead to a better understanding of CS. For example, Demanet and Nguyen [11] discussed minimax recovery theory and error bounds by analyzing restricted isometry constant (a CS concept).

In addition to sparsity, we also assume that the point sources are separated by a critical distance, which is referred to as *minimum separation* (MS) [6] (cf. Definition 1). Theoretical results based on the analysis of L_1 certificate or interpolating trigonometric polynomials of sparse sign patterns [6] demonstrate that point sources can be exactly recovered in the noise-free case as long as any two spikes are MS distance apart (cf. Theorem 1).

Definition 1 (Minimum Separation) Let $\mathbb{T}$ be the circle obtained by identifying the endpoints on $[0, 1]$ and $\mathbb{T}^d$ the $d-$dimensional torus. For a family of points $T \in \mathbb{T}^d$, the minimum separation is defined as the closest warp-around distance between any two elements from T,

$$\mathrm{MS} := \triangle(T) := \inf_{(t,t')\in T: t\neq t'} |t - t'|, \tag{2}$$

where $|t - t'|$ is the L_∞ distance (maximum deviation in any coordinate).

Theorem 1 *[6, Corollary 1.4] Let $T = \{t_j\}$ be the support of x. If the minimum distance obeys*

$$\triangle(T) \geq 2\lambda_c N, \tag{3}$$

then x is the unique solution to L_1 minimization:

$$\min |x|_1 \quad s.t. \quad \mathcal{F}_n x = y. \tag{4}$$

If x is real-valued, then the minimum gap can be lowered to $1.87\lambda_c N$.

We want to analyze the constant in front of $\lambda_c N$ in eq. (3), referred to as minimum separation factor (MSF). Theorem 1 indicates that MSF≥ 2 guarantees the exact recovery of L_1 minimization with a recent improvement to 1.26 [18] at the cost of an additional constraint that $f_c \geq 1000$. This line of research was originated from Donoho [12], who showed that MSF> 1 is sufficient if the spikes are on the grid. Note that both aforementioned works [6, 18] are formulated in terms of off-grid spikes. Another article about off-grid spikes was [1] by Aubel *et al*, who also arrived at MSF> 1 if windowed Fourier (or short-time Fourier transform) measurements are available. Furthermore, there are two works that do not require MS. De Castro and Gamboa [10] showed that K spikes can be resolved from $2K + 1$ Fourier samples; and with additional positive assumption of point sources, Donoho *et al.* [14] showed that $2K$ noiseless measurements are sufficient to yield exact recovery of K positive spikes. In addition to these exact recovery results, errors in spike detection and noise robustness are of great interest as well. Fernandez-Granda analyzed error bounds of constrained L_1 minimization in [18], while the unconstrained version was addressed in [34] under a Gaussian noise model as well as in [2] for any sampling scheme. The robustness of spike detection was discussed in [15].

1.1 Our contributions

We investigate recovery performance of two nonconvex L_1 based penalties, the difference of L_1 and L_2 norms (L_{1-2}) and capped L_1 (CL_1). The former is recently proposed in [22, 40] as an alternative to L_1 for CS, and the latter is often used in statistics and machine learning [33, 41]. Numerical simulations show that L_1 minimization often fails when MSF< 1, in which case we demonstrate that both L_{1-2} and CL_1 outperform the classical L_1 method.

During the course of simulation study, we observe that the rank property is mostly satisfied for L_{1-2}, i.e. the L_0 norm of the reconstructed solution does not exceed n (the rank of A). We find that exact sparse recovery is almost unlikely when MSF is very small, but the reconstructed solution is still sparse with sparsity at most n. In addition, we have the following relationship: $\text{MS} \cdot K \leq N$, MSF=MS$\cdot f_c/N$, and $\text{rank}(A) = n = 2 \cdot fc + 1$. Putting them together, we get $K < 0.5n/\text{MSF}$. This inequality implies that we may reconstruct a vector sparser than the ground-truth (c.f. Figure 4).

The rest of the paper is organized as follows. Section 2 reviews numerical methods for L_1 minimization [6] and L_p $(0 < p < 1)$ minimization [20]. Section 3 describes the proposed algorithms for two non-convex functionals, L_{1-2} and CL_1, in a unified way. We analyze the theoretical aspects of the two methods in Section 4 including rank property, local minimizers, and stationary points. Experiments on both one-dimensional signals and two-dimensional images are examined in Section 5, followed by conclusions in Section 6.

2 Review on L_1 and L_p minimization

To make the paper self-contained, we briefly review two numerical algorithms: L_1 minimization via semi-definite program (SDP) in [6] and L_p minimization via iteratively reweighted least square (IRLS) in [20], both of which will be examined in Section 5 as a benchmark to the proposed L_{1-2} and CL_1 methods.

2.1 L_1 via SDP

To recover the optimal solution of (4), Candés and Fernandez-Granda [6] considered a dual problem, *i.e.*,

$$max_c \ \text{Re}\langle y, c\rangle \quad \text{s.t.} \quad \|\mathcal{F}_n^* c\|_\infty \leq 1; \tag{5}$$

the constraint says that the trigonometric polynomial $\mathcal{F}_n^* c(t) = \sum_{|k| \leq f_c} c_k e^{i2\pi kt}$ has a modulus uniformly bounded by 1 over the interval $[0, 1]$. As indicated in [6, Corollary 4.1], this constraint is equivalent to the existence of a Hermitian matrix $Q \in \mathbb{C}^{n\times n}$ such that

$$\begin{bmatrix} Q & u \\ u^* & 1 \end{bmatrix} \succeq 0, \quad \sum_{i=1}^{n-j} Q_{i,i+j} = \begin{cases} 1 \ j = 0 \\ 0 \ j = 1, 2, \cdots, n-1. \end{cases} \tag{6}$$

Therefore, the dual problem is equivalent to

$$\{\hat{c}, \hat{Q}\} = \arg\max_{c,Q} \text{Re}\langle y, c\rangle \quad \text{s.t.} \quad (6), \tag{7}$$

which can be solved via SDP on the decision variations $c \in \mathcal{C}^n, Q \in \mathcal{C}^{n\times n}$, in total $(n+1)^2/2$ variables. Once the optimal dual variations $\hat{c}, \hat{Q}$ are obtained, a root-finding technique is used to retrieve a solution to the primal problem (4). In particular, the trigonometric polynomial,

$$p_{2n-2}(e^{i2\pi t}) = 1 - |\mathcal{F}_c^*(t)|^2 = 1 - \sum_{k=-2f_c}^{2f_c} u_k e^{i2\pi kt}, \quad u_k = \sum_j \hat{c}_j \bar{\hat{c}}_{j-k}, \tag{8}$$

is a real-valued and nonnegative trigonometric polynomials by construction; and $p_{2n-2}(e^{i2\pi t})$ is either equal to zero everywhere or has at most $n-1$ roots on the unit circle. Therefore, one simply locates the roots of p_{2n-2} on the unit circle in order to recover the support of the optimal solution to the L_1 minimization (4); and then amplitudes can be estimated via least-squares. The noise robustness of this algorithm was analyzed in a follow-up work [5].

2.2 L_p via IRLS

The L_p quasi-norm is often used in CS as an alternative to L_1 to approximate the L_0 norm, see [7–9, 20, 38]. As it is nonconvex for $p < 1$, L_p minimization is generally NP hard. In [20], the authors considered a smoothed L_p minimization, which is expressed as

$$\min \lambda \sum_{j=1}^{N} (|x_j|^2 + \epsilon^2)^{p/2} + \frac{1}{2}\|Ax - b\|_2^2, \tag{9}$$

for $\epsilon > 0$. Taking the gradient of (9) gives the first-order optimality condition,

$$\lambda \left[\frac{p x_j}{(|x_j|^2 + \epsilon^2)^{1-p/2}} \right]_{1 \le j \le N} + A^T(Ax - b) = 0. \tag{10}$$

Then an iterative scheme is formulated as

$$\begin{cases} x^{k+1} = \arg\min \lambda \sum\limits_{j=1}^{N} w_j^k |x_j|^2 + \frac{1}{2}\|Ax - b\|_2^2 \\ \\ w_j^{k+1} = p(|x_j^{k+1}|^2 + \epsilon^2)^{p/2-1}. \end{cases} \tag{11}$$

Each subproblem in (11) can be solved by a weighted least-square type of equation,

$$(\lambda W^k + A^T A)x^{k+1} = A^T b, \tag{12}$$

where W^k is a diagonal matrix with diagonal elements of $\{w_j^k, j = 1, \cdots, N\}$. The parameter ϵ should be discreetly chosen so as to avoid local minima. The update for ϵ in [20] is given as $\epsilon^{k+1} = \min\{\epsilon^k, c \cdot r(x^{k+1})_{K+1}\}$, where $c \in (0, 1)$ is a constant, $r(z)$ is the rearrangement of absolute value of $z \in \mathbb{R}^N$, and K is the estimated sparsity of the vector x to be constructed. This method gives better results than the classical L_1 approaches in the RIP regime and/or incoherent scenario, but it does not work so well for highly coherent CS, as observed in [21, 22, 40].

3 Nonconvex L_1 based minimization via DCA

In this section, we describe a unified approach for solving two nonconvex L_1 based minimization problems via a difference of convex algorithm (DCA) [28, 29]. The unconstrained minimization problem is formulated as follows,

$$\min F(x) := \lambda R(x) + \frac{1}{2}\|Ax - b\|_2^2, \tag{13}$$

where $R(x)$ is a regularization term, λ is a balancing parameter, and $A = \mathcal{F}_n$. We consider two regularization terms:

$$R_{L_{1-2}}(x) = \|x\|_1 - \|x\|_2 \qquad (L_{1-2}) \tag{14}$$

$$R_{CL_1}(x) = \sum_j \min\{|x_j|, \alpha\}, \qquad (CL_1) \tag{15}$$

where α in R_{CL_1} is a pre-defined parameter. A variant of CL_1 is of the form $\sum_j \min\{|x_j|/\alpha, 1\}$, referred to as a normalized capped L_1 [27]. However, the normalized CL_1 is computationally stiff in the super-resolution setting, while L_{1-2} is not, and parameter free.

The method of DCA decomposes $F(x) = G(x) - H(x)$ where both $G(x)$ and $H(x)$ are convex. By linearizing H, we obtain an iterative scheme that starts with $x^1 \neq \mathbf{0}$,

$$\begin{cases} y^k \in \partial H(x^k) \\ x^{k+1} = \arg\min_{x\in\mathbb{R}^N} G(x) - \big(H(x^k) + \langle y^k, x - x^k\rangle\big), \end{cases} \tag{16}$$

where y^k is a subgradient of $H(x)$ at x^k. The DC decomposition is

$$\begin{cases} G_{L_{1-2}}(x) = \frac{1}{2}\|A\,x - b\|_2^2 + \lambda\|x\|_1 \\ H_{L_{1-2}}(x) = \lambda\|x\|_2, \end{cases} \begin{cases} G_{CL_1}(x) = \frac{1}{2}\|A\,x - b\|_2^2 + \lambda\|x\|_1 \\ H_{CL_1}(x) = \lambda\sum_j \max(|x_j| - \alpha, 0), \end{cases} \tag{17}$$

for L_{1-2} and CL_1 respectively. Each subproblem in (16) amounts to an L_1 regularized form

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 - \langle y^k, x\rangle, \tag{18}$$

where $y^k_{L_{1-2}} = \lambda\frac{x^k}{\|x^k\|_2}$ for L_{1-2}, and $y^k_{CL_1} = \lambda\text{sign}(x^k). * \max(|x^k| - \alpha, 0)$[1] for C$L_1$. To solve (18), we consider the augmented Lagrangian

$$L_\delta(x, z, u) = \frac{1}{2}x^T(A^TA)x + \lambda\|z\|_1 - \langle y^k, x\rangle + \langle u, x - z\rangle + \frac{\delta}{2}\|x - z\|_2^2,$$

where z is an auxiliary variable; to enforce the constraint $x = z$, the Lagrange multiplier $\delta > 0$ and dual variable u are introduced. ADMM iterates between

[1] Denote $.*$ be entry-wise multiplication, and $|\cdot|$ be entry-wise absolute value (Note $\|\cdot\|_1$ is the standard L_1 norm).

minimizing L_δ with respect to x and z, and updating the dual variable u. Therefore, an iterative scheme for solving the subproblem (18) goes as follows,

$$\begin{cases} x_{l+1} = (A^T A + \lambda I_d)^{-1}(\delta(z_l + u_l) - y^k) \\ z_{l+1} = \text{shrink}(x - u, \lambda/\delta) \\ u_{l+1} = u_l + z_{l+1} - x_{l+1}, \end{cases} , \tag{19}$$

where the subscript l indexes the inner iterations. Note that the matrix inversion $(A^T A + \lambda I_d)^{-1}$ can be efficiently implemented by Fast Fourier Transforms, as A is the multiplication of a sampling matrix and Fourier matrix. The subproblem (18) is convex, and hence it is guaranteed to have an optimal solution x_* via (19), and we take it to be the solution of (18), *i.e.*, $x^{k+1} = x_*$.

For the constrained formulation,

$$\min R(x) \quad \text{s.t.} \quad Ax = b, \tag{20}$$

we apply a similar trick to the unconstrained version by considering the following iterative scheme,

$$x^{k+1} = \arg\min_x \{\|x\|_1 - \langle y^k, x\rangle \quad \text{s.t.} \quad Ax = b\}. \tag{21}$$

To solve (21), we introduce two dual variables u, v and define an augmented Lagrangian

$$L_\delta(x, z, u, v) = \|z\|_1 - \langle y^k, x\rangle + \langle u, x - z\rangle + \langle v, Ax - b\rangle + \frac{\delta}{2}\|x - z\|^2 + \frac{\delta}{2}\|Ax - y\|^2,$$

where ADMM finds a saddle point (x_*, z_*, u_*, v_*) satisfying

$$L_\delta(x_*, z_*, u, v) \leqslant L_\delta(x_*, z_*, u_*, v_*) \leqslant L_\delta(x, z, u_*, v_*) \qquad \forall x, z, u, v.$$

As a result, we take $x^{k+1} = x_*$.

4 Theoretical properties

In this section, we investigate a rank property, namely the L_0 norm of the reconstructed solution does not exceed n (the rank of A). First of all, we examine the probability of finding the exact solution with 100 random trials for L_{1-2}, CL_1 with $\alpha = 0.1$, and L_p with $p = 1/2$. The left of Figure 1 illustrates that it is unlikely to find the exact solution when MSF is small (< 0.8), which implies that multiple sparse vectors satisfying $Ax = b$ do exist. On the other hand, we plot the probability of rank property being satisfied on the right of Figure 1, by counting how many times the L_0 norm of the reconstructed solution is smaller than or equal to n. The results suggest that the rank property is true for both L_{1-2} and CL_1 when MSF> 1 or when L_1 certificate holds. More importantly in the worse case (MSF< 0.8), L_{1-2} provides a sparse solution (sparsity $\leq n$) while satisfying the constraint, which is the best one can do. It seems unlikely for L_p to have the rank property.

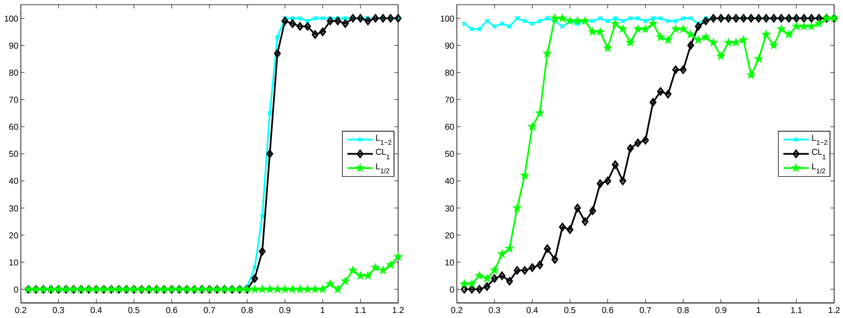

Fig. 1 Probability (%) of finding the exact solution (left) and of rank property being satisfied (right) for both L_{1-2} and CL_1 with $\alpha = 0.1$. It shows over 95% chance that the reconstructed solutions using L_{1-2} are $n-$sparse, though it does not find the exact ground-truth solution when MSF is small (< 0.8).

One deterministic result regarding the rank property is given in [40, Theorem 3.1] that there exists λ_n such that for any $\lambda > \lambda_n := \frac{\|A\|_2\|b\|}{\sqrt{n+1}-1}$, the stationary point of the unconstrained L_{1-2} minimization problem has at most n non-zero elements. In practice, we choose a much smaller λ than λ_n, which usually yields a smaller residual and better recovery. As for CL_1, we can not derive such result, as $y^k_{CL_1}$ is not bounded *a priori* and hence no upper bound of sparsity in terms of λ for CL_1. For the rest of this section, we give some theoretical analysis on the rank property that is independent of λ.

4.1 Local minimizers

It is shown in [40, Theorem 2.3-2.4] that any local minimizer of L_{1-2} has the rank property, as summarized in Theorem 2. With additional assumption, we prove the rank property for CL_1 in Theorem 3. The error bounds at high probability of a local minimizer of CL_1 from the true solution are established in [37,41] under sparse eigenvalue assumptions of the Hessian of the loss functions (similar to RIP), which is unfortunately hard to verify.

Theorem 2 *Suppose $A \in \mathbb{R}^{n\times N}$ is of full row rank. If x^* is a local minimizer of L_{1-2} in either a unconstrained (13) or constrained (20) formulation, then the sparsity of x^* is at most n.*

Theorem 3 *Suppose $A \in \mathbb{R}^{n\times N}$ is of full row rank. If x^* is a local minimizer of CL_1 in either a unconstrained (13) or constrained (20) formulation and the objective function is not flat in the neighborhood of x^*, then the sparsity of x^* is at most n.*

Proof We only provide proof for the constrained case, and the unconstrained version is almost the same. It is sufficient to show the columns of A_{Λ^*} are linearly independent. Prove by contradiction. Suppose there exists $v \in \ker(A)\setminus$

0 such that $\mathrm{supp}(d) \subseteq \Lambda^*$. Denote $\Lambda^*_{\alpha+} = \{j : |x^*_j| > \alpha\}$, $\Lambda^*_{\alpha-} = \{j : |x^*_j| < \alpha\}$, and $\Lambda^*_\alpha = \{j : |x^*_j| = \alpha\}$. For any fixed neighborhood of x^*, we scale d so that

$$\begin{cases} |x^*_j \pm d_j| > \alpha \; j \in \Lambda^*_{\alpha+} \\ |x^*_j \pm d_j| < \alpha \; j \in \Lambda^*_{\alpha-} \end{cases} \tag{22}$$

Consider two feasible vectors in $\mathbf{B}_r(x^*)$, $\hat{x} = x^* + d$ and $\breve{x} = x^* - d$. Since $\mathrm{supp}(d) \subseteq \Lambda^*$ and $d \in \ker(A)$, we have $\mathrm{supp}(\hat{x}) \subseteq \Lambda^*$, $\mathrm{supp}(\breve{x}) \subseteq \Lambda^*$, and $A\hat{x} = A\breve{x} = Ax^*$. By analyzing $R_{CL_1}(x^*)$ and $R_{CL_1}(x^* \pm d)$, we get

$$\begin{aligned} & R_{CL_1}(x^* + d) + R_{CL_1}(x^* - d) - 2R_{CL_1}(x^*) \\ = & \sum_{j \in \Lambda^*_{\alpha-}} \Big(|x^*_j + d_j| + |x^*_j - d_j| - 2|x^*_j| \Big) + \sum_{j \in \Lambda^*_0} \Big(\min(|\alpha + d_j|, |\alpha - d_j|) - \alpha \Big). \end{aligned}$$

The first term is zero for v sufficiently small, while the second term is negative if $\Lambda^*_0 \neq \emptyset$, so we have

$$R_{CL_1}(x) \geq \min\{R_{CL_1}(\hat{x}), R_{CL_1}(\breve{x})\}.$$

As long as $R_{CL_1}(x^*)$ is not flat (or constant) in $\mathbf{B}_r(x^*)$, the above inequality is strict, which contradicts with the assumption that x^* is a local minimizer in $\mathbf{B}_r(x^*)$.

Remarks: It is possible that objective function for CL_1 is not constant. For example, if the set $\{j : -\alpha < x_j < 0\}$ has different cardinality to the set $\{j : 0 < x_j < \alpha\}$, then $R_{CL_1}(\hat{x}) \neq R_{CL_1}(\breve{x})$. Or if $\Lambda^*_0 \neq \emptyset$, then $\sum_{j \in \Lambda^*_0} \Big(\min(|\alpha + d_j|, |\alpha - d_j|) - \alpha \Big) < 0$. In addition, the rank property of CL_1 depends on α. If α is small, then the set $\Lambda^*_{\alpha-}$ may be empty, and hence rank property does not hold. Another interpretation is that if α is too small, the problem is a small perturbation of the least squares problem where sparsity is absent. If α is too large, the CL_1 is no longer a good approximation of the L_0 norm. Empirically, we find that an adaptive update of α during iterations works better than a fixed value, one advantage of which is no need to tune this parameter. The analysis of adaptive α is beyond the scope of this paper.

Applying convergence properties of general DCA studied in [29, 30] for CL_1, we know the limit point, x^*, is a local minimizer if no component of x^* is equal to $\pm\alpha$. We numerically calculate the probability of the computed solution not taking values $\pm\alpha$, which implies local minimizers. For this purpose, we test 100 random sparse (ground-truth) vectors from Gaussian distribution and 25 random choices of α from $[0, 1]$ by uniform distribution, and compute how many times that the computed solution does not take values $\pm\alpha$. Finally we plot the probability of the limit points being local minimizers in Figure 2, which is almost for sure ($\sim 99.6\%$) at each MSF. The probabilities of having exact recovery and rank property are also provided, which validates that local minimizers do not imply the rank property, as indicated by Theorem 3. Compared with Figure 1, rank property is more likely to occur for CL_1 when α is chosen randomly instead of a fixed value. This phenomenon again suggests that an adaptive α may be better.

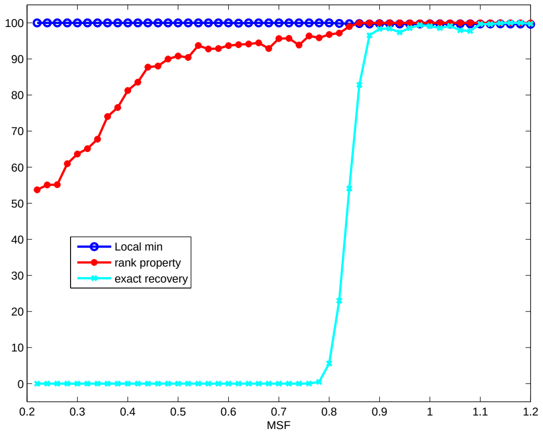

Fig. 2 Probability (%) of the computed solution of CL_1 with no elements equal to $\pm\alpha$. The results are averaged over 100 random signals and 25 random choices of α drawn from uniform distribution $[0, 1]$ at each MSF.

4.2 Second-order optimality condition

By analyzing the second-order optimality condition, we show that either the stationary point x^* has at most n non-zero elements or there exists a vector in any neighborhood of x^* that has a smaller objective function. We will need the following technical lemma.

Lemma 1 *If* $\lambda < \min\{\frac{\|A^Tb\|_2}{\sqrt{N}+\|A\|^2}, \frac{\|A^Tb\|_2}{\sqrt{N}+1}\}$, *then any first-order stationary point* $x^* \in \mathbb{R}^N$ *of* L_{1-2} *unconstrained problem (13) satisfies* $\|x^*\|_2 > \lambda$.

Proof First, we show that x^* can not be zero. Suppose $x^* = 0$, then by the optimality condition,

$$\lambda(p^* - q^*) - A^Tb = 0, \tag{23}$$

where $p^* \in \partial\|x^*\|_1$ is the subgradient of $\|x\|_1$ at x^*, and $q^* \in \partial\|x^*\|_2$. It is easy to see that when $x^* = 0$, $\|p^*\|_\infty \leq 1$ and $\|q^*\|_2 \leq 1$. By (23), we have

$$\|A^Tb\|_2 = \lambda\|p^* - q^*\|_2 \leq \lambda(\sqrt{N} + 1),$$

or $\lambda \geq \frac{\|A^Tb\|_2}{\sqrt{N}+1}$, which is a contradiction.

Therefore $x^* \neq 0$, and

$$\lambda(p^* - \frac{x^*}{\|x^*\|_2}) + A^T(Ax^* - b) = 0. \tag{24}$$

It follows from (24) that

$$\begin{aligned}\|A\|^2\|x^*\|_2 =& \|A^TA\|\|x^*\|_2 \geq \|A^TAx^*\|_2 = \|-\lambda(p^* - \frac{x^*}{\|x^*\|_2}) + A^Tb\|_2 \\ \geq& \|A^Tb\|_2 - \lambda\|p^* - \frac{x^*}{\|x^*\|_2}\|_2.\end{aligned} \tag{25}$$

Let Λ^* be the support of x^*, then $p_i^* = \operatorname{sign}(x)$ for $i \in \Lambda^*$, and $|p_i^*| \leq 1$ otherwise. So we have $|p^* - \frac{x^*}{\|x^*\|_2}|_i < 1$ for $i \in \Lambda^*$, and $|p^* - \frac{x^*}{\|x^*\|_2}|_i \leq 1$ otherwise. Using the assumption $\lambda \leq \frac{\|A^T b\|_2}{\sqrt{N}+\|A\|^2}$, from (25) it follows that

$$\|A\|^2 \|x^*\|_2 > \|A^T b\|_2 - \lambda\sqrt{N} \geq \lambda \|A\|^2,$$

and thus $\|x^*\|_2 > \lambda$.

Theorem 4 *Suppose $\lambda < \min\{\frac{\|A^T b\|_2}{\sqrt{N}+\|A\|^2}, \frac{\|A^T b\|_2}{\sqrt{N}+1}\}$. Let x^* be any limit point of the DCA L_{1-2} minimizing sequence. Then we have either $\|x^*\|_0 \leq n$ (rank property) or there exists $d \in \mathbb{R}^N$ such that $F(x^*+d) < F(x^*)$ and x^*+d is sparser than x^*.*

Proof Taking the difference of objective function values at x^*+d and x^*, we get

$$\begin{aligned}
&\frac{1}{\lambda}\Big(F(x^*+d) - F(x^*)\Big) \\
&= \left(\sum_{j\in\Lambda^*} \langle \operatorname{sign}(x_j^*), d_j\rangle + \sum_{j\in\Lambda_c^*} |d_j| + \langle \frac{1}{\lambda}A^T(Ax-b) - \frac{x^*}{\|x^*\|_2}, d\rangle\right) \\
&\quad + \frac{1}{2} d^T \left(\frac{1}{\lambda}A^T A - \frac{1}{\|x^*\|_2} + \frac{x^* {x^*}^T}{\|x^*\|_2^3}\right) d + O(\|d\|_2^3). \qquad (26)
\end{aligned}$$

Note that x^* is a column vector in (26), and hence $x^* {x^*}^T$ is a (rank-one) matrix. Since x^* is the limit point of DCA sequence, it satisfies the first-order optimality condition (24). Therefore, the first term in (26) is equal to $\sum_{j\in\Lambda_c^*} |d_j| - \langle p_j^*, d_j\rangle$, and is nonnegative, since $p_j^* \in [-1,1]$ for $j \in \Lambda_c^*$.

As for the Hessian matrix in (26), denoted as H, we have

$$H := \frac{1}{\lambda}A^T A - \frac{1}{\|x^*\|_2} + \frac{x^* {x^*}^T}{\|x^*\|_2^3} = \frac{1}{\|x^*\|_2}\mathcal{F}^T\left(\frac{\|x^*\|_2}{\lambda} S_n^T S_n - I_d + yy^T\right)\mathcal{F},$$

where $y = \mathcal{F}x^*/\|x^*\|_2$ and $A = S_n\mathcal{F}$ (S_n is a sampling matrix and $\mathcal{F}$ is the Fourier matrix). As $S_n^T S_n$ are a diagonal matrix taking values of either 1 or 0, the matrix $D := \frac{\|x^*\|_2}{\lambda} S_n^T S_n - I_d$ is also diagonal, the elements of which are $\beta := \frac{\|x^*\|_2}{\lambda} - 1$ with multiplicity n, and -1 with multiplicity $N-n$. By Lemma 1, we have $\beta > 0$.

We want to analyze the eigenvalues of H, which is equivalent to analyzing a diagonal matrix D with rank-one perturbation yy^T. Suppose u is an eigenvector of $D + yy^T$ with corresponding eigenvalue γ, then we have

$$(u^T y) \cdot \begin{bmatrix} y_n \\ y_{N-n} \end{bmatrix} + \begin{bmatrix} (\beta-\gamma)I_n & 0 \\ 0 & (-1-\gamma)I_{N-n} \end{bmatrix} \begin{bmatrix} u_n \\ u_{N-n} \end{bmatrix} = 0, \qquad (27)$$

where $y = [y_n, y_{N-n}]^T$ and $u = [u_n, u_{N-n}]^T$. So the eigenvalues of $D + yy^T$ are β with multiplicity $n-1$, -1 with multiplicity $N-n-1$, and other two, denoted as γ_1, γ_2, satisfying $\frac{\|y_n\|^2}{\gamma-\beta} + \frac{\|y_{N-n}\|^2}{\gamma+1} = 1$, or $\gamma^2 - \beta\gamma - (\beta+1)\|y_n\|^2 = 0$, where we use $\|y_n\|_2^2 + \|y_{N-n}\|_2^2 = 1$. It follows from the quadratic formula that these eigenvalues satisfy $-1 < \gamma_1 < 0 < \beta < \gamma_2$ and $\gamma_1 + \gamma_2 = \beta$.

Now we discuss eigenvectors and diagonalization. Each eigenvector for β has the form of $[u_n, 0]$ with $u_n^T y_n = 0$. Denote U_n be a matrix with each column being one of the eigenvectors corresponding to β. We further assume U_n is orthonormal after Gram-Schmidt orthogonalization. Similarly, each eigenvectors for -1 has the form of $[0, u_{N-n}]$ with $u_{N-n}^T y_{N-n} = 0$, and we denote U_{N-n} be an orthonormal matrix composed of all the corresponding eigenvectors. Therefore, an orthonormal matrix, denoted as U, that diagonalizes H can be expressed as

$$U = \begin{bmatrix} U_n & 0 & \dfrac{y_n}{(\gamma_1 - \beta)\alpha_1} & \dfrac{y_n}{(\gamma_2 - \beta)\alpha_2} \\ & & & \\ 0 & U_{N-n} & \dfrac{y_{N-n}}{(\gamma_1 + 1)\alpha_1} & \dfrac{y_{N-n}}{(\gamma_2 + 1)\alpha_2} \end{bmatrix}, \tag{28}$$

where α_1, α_2 are normalizing factors.

For any $d \in \mathbb{R}^N$, we can decompose $d = d^k + d^r$, where $d^k \in \ker(A)$ and $d^r \in \text{range}(A^T)$, and hence $\mathcal{F}d^k, \mathcal{F}d^r$ have the forms of $[0, g_{N-n}]^T, [g_n, 0]^T$ respectively. Denote $s_1 := \langle y_n, g_n \rangle,\ s_2 := \langle y_{N-n}, g_{N-n} \rangle$. We can prove that s_1, s_2 are real numbers and $s_1 + s_2 = \langle d, x^* \rangle$. Then after tedious calculation, (26) reduces to

$$\begin{aligned} F(x^* + d) - F(x^*) = \lambda \sum_{j \in \Lambda_c^*} \Big(|d_j| - \langle p_j^*, d_j \rangle \Big) & \\ + \beta \|U_n^T g_n\|^2 - \|U_{N-n}^T g_{N-n}\|^2 + P_1 s_1^2 + P_2 s_1 s_2 + N_0 s_2^2 & \end{aligned} \tag{29}$$

where $P_1, P_2 > 0$ and $N_0 < 0$ are constant with respect to d.

If $\|x^*\|_0 > n$, then the columns of A_{Λ^*} are linearly dependent, and hence there exists $d \in \ker(A) \setminus \{0\}$ such that $d \in S_{\Lambda^*}$, where $S_{\Lambda^*} = \{x : \text{supp}(x) \in \Lambda^*\}$. As $d \in \ker(A)$, $\mathcal{F}d = [0, g_{N-n}]$, *i.e.*, $g_n = 0$ and $s_1 = 0$. Since $g_{N-n} \neq 0$, we have $F(x^* + d) - F(x^*) < 0$, *i.e.*, $x^* + d$ has a smaller objective function than x^*. In addition, we can scale d to cancel one non-zero element of x^*, and hence $x^* + d$ is sparser than x^*.

4.3 Stationary points

It is shown in [21, 40] that any limit point of the DCA sequence converges to a stationary point; and in Theorem 5, we give a tighter result, which states that the limit point is d-stationary rather than stationary. These stationarity concepts are related as the set of local minimizers belongs to the set of d-stationary points, which belongs to the set of stationary points. As we often observe that limit points of DCA are sparse (see Figure 1), it is likely that any d-stationary point may have the rank property, which will be left to a future work. We first give the definition of d-stationary points [16], and then prove the DCA sequence converges to a d-stationary point in Theorem 5.

Definition 2 (D-stationary) We say that a vector $\hat{x} \in X$ is a d(irectional) stationary point of the minimization of a function $F(x)$, or in short, d-stationary, if

$$F'(\hat{x}; x - \hat{x}) \geq 0, \quad \forall x \in X,$$

where the directional derivative is defined as one-sided derivative

$$F'(x; d) := \lim_{\tau \searrow 0} \frac{F(x + \tau d) - F(x)}{\tau}. \tag{30}$$

For example, directional derivatives of L_1 and L_2 norms at x^* are

$$\| \cdot \|_1'(x^*; d) = \sum_{j \in \Lambda^*} \langle \text{sign}({}^* x_j), d_j \rangle + \sum_{j \notin \Lambda^*} |d_j|, \tag{31}$$

and

$$\| \cdot \|_2'(x^*; d) = \begin{cases} \langle \frac{x^*}{\|x^*\|_2}, d \rangle & \text{if } x^* \neq 0 \\ \|d\|_2 & \text{if } x^* = 0. \end{cases} \tag{32}$$

As a result, the directional derivative of $F(x)$ for L_{1-2} can be expressed as

$$F'(x^*; d) = \lambda \sum_{j \in \Lambda^*} \langle \text{sign}(x_j^*), d_j \rangle + \lambda \sum_{j \in \Lambda_c^*} |d_j| - \lambda \langle \frac{x^*}{\|x^*\|_2}, d \rangle + \langle A^T (Ax^* - b), d \rangle, \tag{33}$$

for $x^* \neq 0$.

Theorem 5 *Let $\{x^k\}$ be the sequence of iterates generated by DCA* (16)*, or DCA sequence in short, for L_{1-2}, then any limit point x^* of $\{x^n\}$ is a d-stationary point of $F(x)$, defined in* (13) *and $R(x) = R_{L_{1-2}}(x)$.*

Proof In [40], the DCA sequence $\{x^k\}$ was shown to converge to a stationary point x^*. We now prove that all the iterates (except for the first one) and the limit point x^* are non-zero. We assume that the initial point is $x^0 = 0$. Then it follows from (16) that

$$F(x^1) = \lambda(||x^1||_1 - ||x^1||_2) + \frac{1}{2}||Ax^1 - b||^2 \tag{34}$$

$$\leq \lambda ||x^1||_1 + \frac{1}{2}\|Ax^1 - b\|^2 \leq \frac{1}{2}\|b\|^2 = F(0). \tag{35}$$

Strict inequality holds if zero is not global minimum of L_1 problem (which is generically true). Therefore, nonzero property is maintained during descending iterations of DCA, *i.e.*, $x^k \neq 0, \ \forall k$. In addition, $x^* \neq 0$ as $F(x^*) < F(0)$.

As x^* satisfies the first-order optimality condition (24), we can simplify the directional derivative of $F'(x; d)$, given in (33),

$$F'(x; d) = \lambda \sum_{j \in \Lambda^*} \langle \text{sign}(x_j^*), d_j \rangle + \lambda \sum_{j \in \Lambda_c^*} |d_j| - \langle p^*, d \rangle \tag{36}$$

$$= \lambda \sum_{j \in \Lambda_c^*} \Big(|d_j| - \langle p_j^*, d_j \rangle \Big), \tag{37}$$

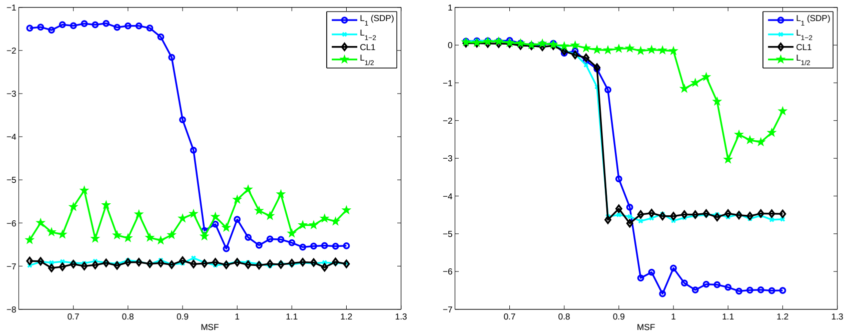

Fig. 3 Error analysis of residuals (left) and relative reconstruction errors (right) on log 10 scale. The underlying signal in this case is with N=1000 and MS=20.

where $p^* \in \partial \|x^*\|_1$. As $p_j^* \in [-1, 1]$ for $j \in \Lambda_c^*$, then $|d_j| - \langle p_j^*, d_j \rangle \geq 0$, and hence $F'(x^*; d) \geq 0 \ \forall d$, which means $x^* \neq 0$ is a d-stationary point.

5 Experimental Results

We numerically demonstrate that the proposed L_{1-2} and CL_1 via DCA can recover signals beyond the Rayleigh length. Two existing methods, L_1 via SDP [6] and $L_{1/2}$ via IRLS [20], serve as benchmarks. For 2D image super-resolution, it is computationally expensive to solve the L_1 minimization via SDP, so we adopt the ADMM approach instead.

5.1 One-dimensional Signal Super-resolution

We consider a sparse signal (the ground-truth) x_g of 1000-dimensional with MS = 20. We vary f_c from 31 to 60, thus MSF:= $\Delta(T) \cdot f_c/N$:=MS$\cdot f_c/N$=0.62 : 0.02 : 1.2. Denoted x^* as the reconstructed signal using any of the methods: SDP, constrained and unconstrained L_{1-2}. In Figure 3, we plot the residual ($\|Au^* - b\|/\|b\|$) and relative reconstruction errors ($\|x^* - x_g\|/\|x_g\|$) on log 10 scale. As MSF decreases towards and passes 1, L_{1-2} with DCA maintains fidelity (constraints) much better, even for the unconstrained L_{1-2}. More importantly, we observe smaller relative errors of L_{1-2} than SDP for MSF< 1 when unique sparse solution is not guaranteed. The L_{1-2} approaches pick one among a solution pool, while errors in SDP's root findings tend to violate the fidelity $Ax = b$. For MSF> 1 where the L_1 certificate holds, L_{1-2} seems not as good as SDP in terms of reconstruction errors, which is due to stopping conditions; on the other hand, relative errors on the order of $1e - 5$ are accurate enough.

In Figures 4-5, we examine one particular ground-truth vector and choose f_c to have MSF=0.4 and 0.8 respectively, when all the methods fail to recover the ground-truth. Figure 4 shows that the reconstructed solutions are sparser

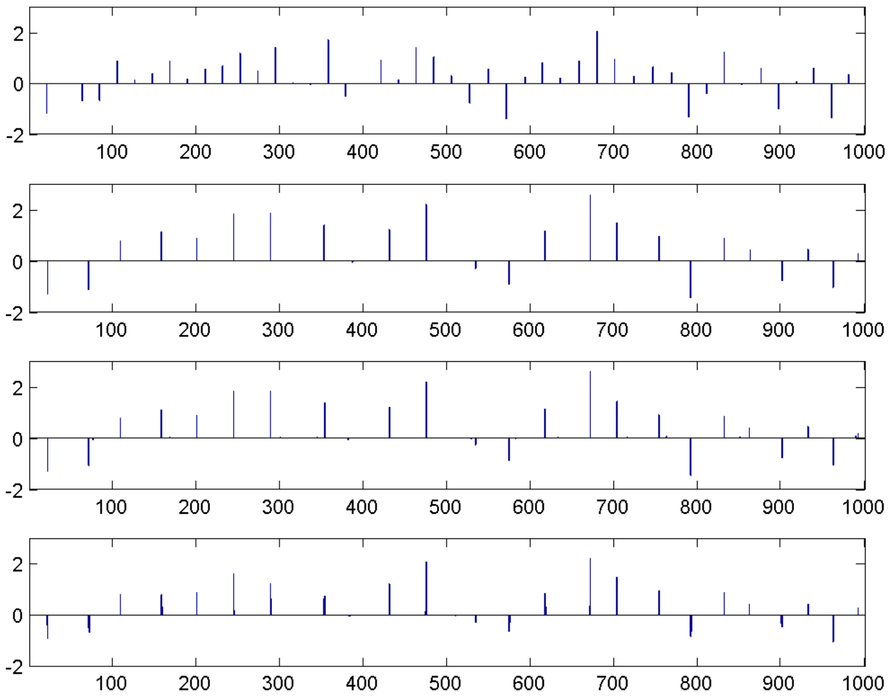

Fig. 4 Reconstruction comparison for MSF=0.4, from top to bottom: ground-truth, L_1 via SDP (residual $\sim 10^{-1.6}$), L_{1-2} (residual $\sim 10^{-7.7}$), and CL_1 (residual $\sim 10^{-5.0}$). All the reconstruction methods yield overly sparser vectors compared to the ground-truth.

than the ground-truth when MSF=0.4, which is consistent with our intuition as discussed in Section 4. We see in Figure 5 that the solution of CL_1 is not sparse when $\alpha = 0.1$ (the result becomes sparse if we increase α to 0.25). For results in Figure 5 (MSF=0.8), we observe "peak splitting", *i.e.*, all the methods miss one single peak in the ground-truth, and instead recover two nearby peaks, marked in blue squares. In addition, a peak shift, circled in green, is probably attributed to both peak splitting and peak merging, as there is a tiny peak on the left of the green in the ground-truth vector. Since there is no certificate guarantee in this regime, there are acceptable solutions if the tolerance on residual is satisfied. The large residual of SDP is clearly due to shifted peak locations, and there are very few small peaks in SDP. In contrast, there are some small peaks appearing in L_{1-2}, which may be the cost to satisfy the constraint. No matter how peaks split or merge, the reconstructed solutions of L_{1-2} are sparse.

We now look at success rates in two tests with fixed MS and fixed f_c respectively. In the first case, we consider 100 random realizations of the same setting as discussed above to get Figure 3, and success rates of three methods can be computed. An incident (or a reconstructed signal x^*) is labeled as "successful" if $\|x^* - x_g\|/\|x_g\| < 1.5e-3$ and $\|Au^* - b\|/\|b\| < 5e-4$. In the

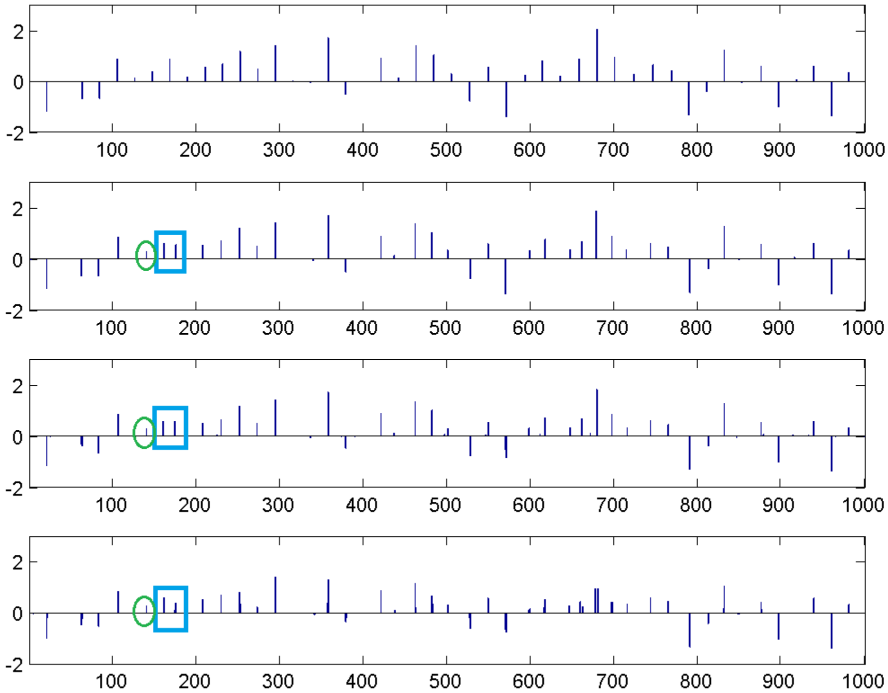

Fig. 5 Reconstruction comparison for MSF=0.8, from top to bottom: ground-truth, L_1 via SDP (residual $\sim 10^{-1.3}$), L_{1-2} (residual $\sim 10^{-7.7}$), and CL_1 (residual $\sim 10^{-4.9}$).

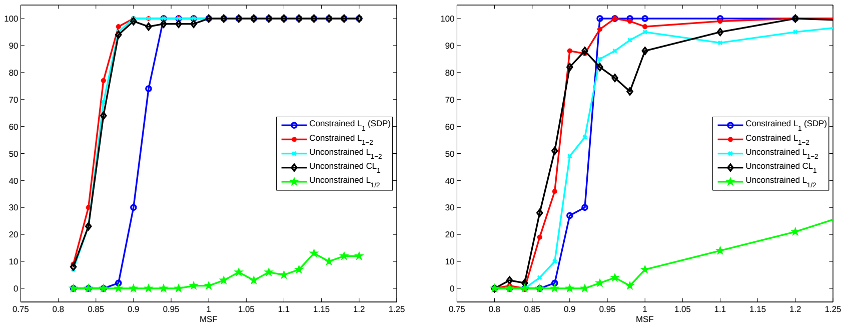

Fig. 6 Success rates (%) of fixed MS= 20 (left) and fixed $f_c = 20$ (right) when $N = 1000$.

second case, we fix $f_c = 20$, generate a sparse signal with MS = MSF*N/f_c, and repeat 100 random realizations to compute success rates. Both plots in Figure. 6 show big advantages of L_{1-2} and CL_1 over SDP when MSF< 1.

We examine the scability of the algorithms for $N = 1000, 2000, 4000$, while keeping SRF fixed, specifically $N/f_c = 50$. Roughly speaking, all the algorithms are scalable to some extent, as illustrated in Figure 7. For SDP, the smaller N is, the smaller MSF is observed for exact recovery. As for L_{1-2} and

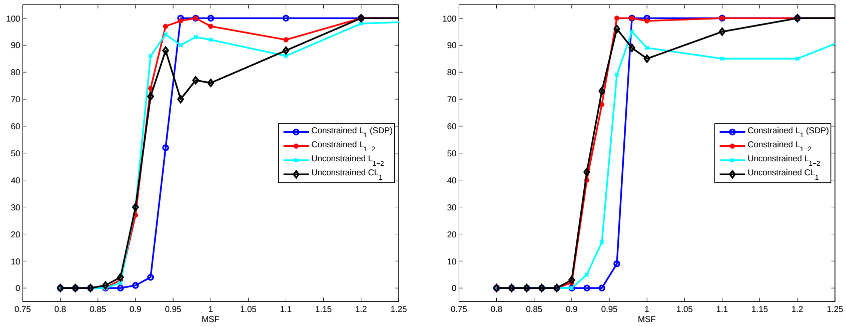

Fig. 7 Scability of the algorithms: N = 2000 (top) and 4000 (bottom) when $f_c = 20$. $N = 1000$ is plotted on the right of Figure 6.

CL_1, the success rates diminish while N increases, which attributes to the curse of dimension: as N is large, the iterative DCA scheme does not converge (for both inner and outer loops) or it takes too long to converge so we have to stop earlier to obtain less accurate results within a reasonable amount of time.

5.2 Two-dimensional Image Super-resolution

We present the super-resolution results of 2D images. There are two types of minimum separation. Definition 1 [6] uses L_∞ norm to measure the distance, while another definition is called *Rayleigh regularity* (RR) [12, 25], as given below.

Definition 3 (Rayleigh regularity) Fix N, n, and set $\lambda = 1/f_c = 2/(n-1)$. We say that the set of points $\mathcal{T} \subset \{0, 1/N, \cdots, 1-1/N\} \subset \mathbb{T}^d$ is Rayleigh regular with parameters (d, r) and write $\mathcal{T} \in \mathcal{R}_d(t, r; N, n)$ if it may be partitioned as $\mathcal{T} = \mathcal{T}_1 \cup \cdots \cup \mathcal{T}_r$ where the $\mathcal{T}_i$'s are disjoint, and each obeys a minimum separation constraint, that is, for all square subsets $\mathcal{D} \subset \mathcal{T}^d$ of side length $t\lambda_c/2$, $|\mathcal{T}_i \cap \mathcal{D}| \leq 1$.

Images with these two definitions are illustrated in Figure 8, where Def. 3 (RR) produces more spikes than Def. 1, thus more challenging for image super-resolution. The theoretical guarantee is studied in [6] for MS and in [25] for RR with additional assumption of positive sources (the spikes have positive values). In both papers, MSF is theoretically proven to be capped at 2.38, while we observe empirically that an exact recovery via L_1 minimization occurs at MSF $= 1.7$. Here is the problem setting: image is of size 100×100, $MS = 10$, and $f_c = 4 : 20$, thus yielding MSF$= 0.4 : 0.1 : 2$. We examine three regularization terms: traditional L_1, L_{1-2}, and CL_1, all of which are formulated in a constrained model. The success rates for these two cases are present in Figure 8. When MSF is below 1.7, the success rates of MS are much

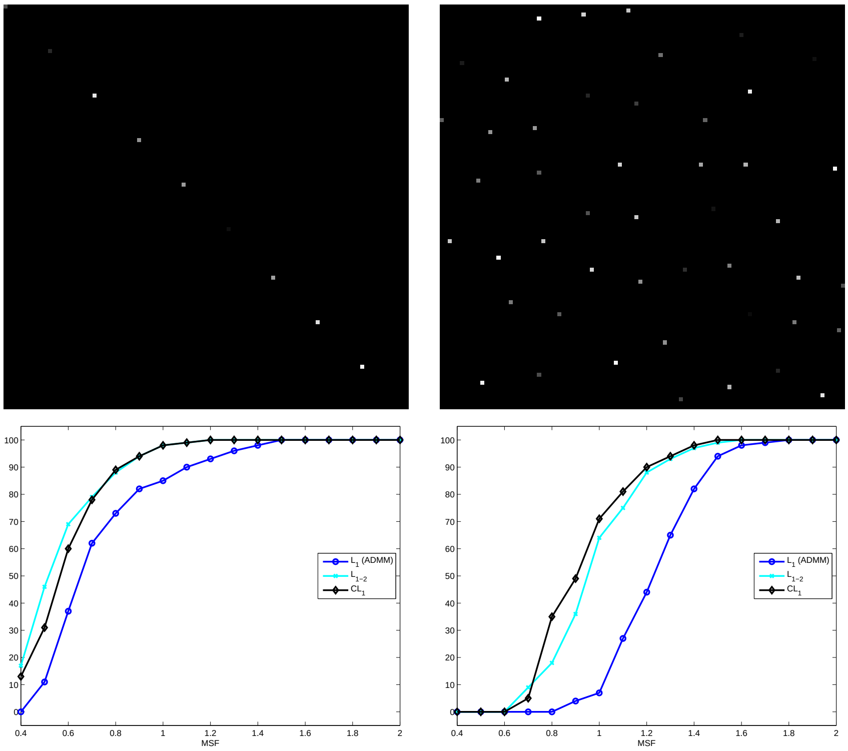

Fig. 8 Image examples shown on the top row: positive spikes of size 100 × 100 and MS=10, which is defined differently in [6] (left) and [25] (right). The corresponding success rates (%) of L_1, L_{1-2}, and CL_1, all in a constrained formulation, are plotted on the bottom. An exact recovery via L_1 minimization occurs at MSF = 1.7 for both MS definitions.

higher than that of RR. Both plots illustrate that L_{1-2} and CL_1 ($\alpha = 0.1$) have advantages over L_1 (solved by ADMM). Note that CL_1 is better than L_{1-2} on the right plot of Figure 8 in the 2d examples where sources only take positive values.

Finally we show an image super-resolution example to illustrate the visual difference. We look at a particular point source image similar to the upper right plot of Figure 8, which reminds one of the stars in a clear night sky. The image is of size 100 × 100 with MS=10 based on RR definition, and we only take 15 × 15 ($f_c = 7$) measurements from low-frequency data, thus yielding MSF=0.7. If using the inverse FFT after the zero-padded frequency data, the reconstruction looks very blurry as shown on the upper left of Figure 9. All the L_1 variants (L_1, capped L_1, and L_{1-2}) result in much sparser and clearer reconstructions. To have a better visual comparison, we plot the error maps of the reconstructed result to the ground-truth for the L_1 methods. All of them can find the spikes' location relatively well, L_1 also picks up some neighboring pixels, and L_{1-2} has the smallest error.

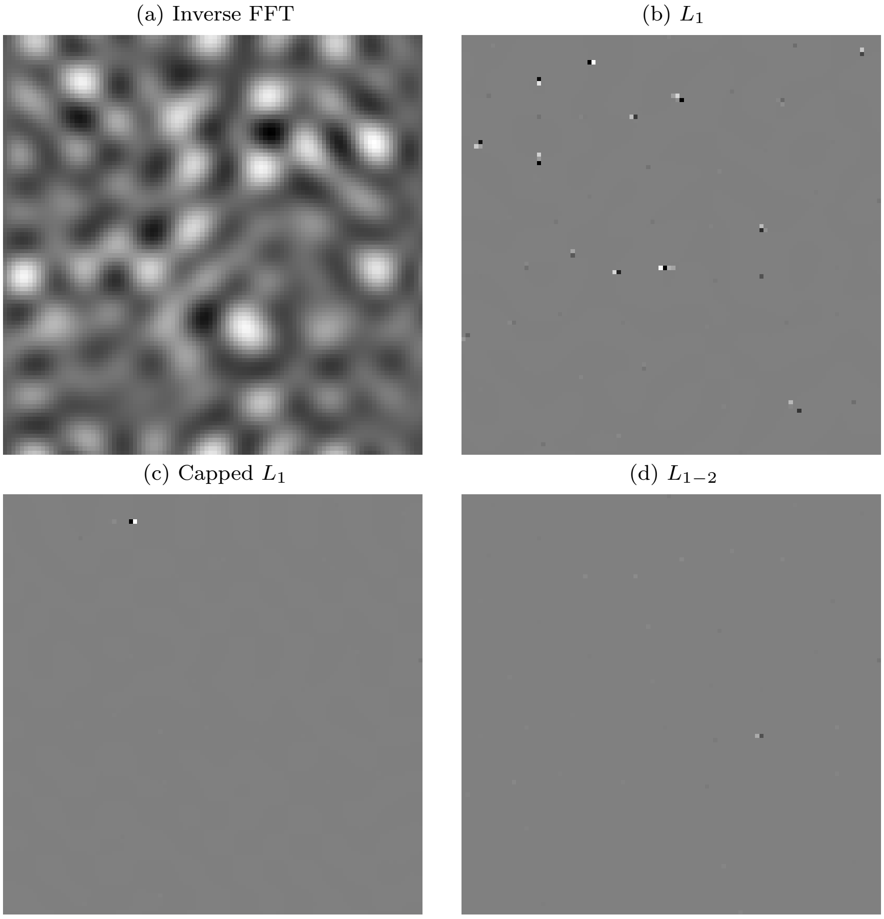

Fig. 9 A particular example at MSF = 0.7 using the RR definition (the top right plot of Figure 8). (a) A reconstruction via a direct inverse FFT, and (b)-(d) are the difference of the reconstructed solutions using L_1, capped L_1, and L_{1-2} to the ground-truth image, with intensity window [-0.1, 0.1]. The root-means-errors for these results are 0.004 (L_1), 0.001 (capped L_1), and 0.0005 (L_{1-2}).

6 Conclusions

We presented local properties (local minimum, d-stationary points, and sparsity upper bound) of computed solutions minimizing L_{1-2} and capped-L_1 penalties for super-resolution of point sources. At a high probability, the limit point of DCA-capped-L_1 algorithm is a local minimum and is sparse if the intrinsic parameter of the capped-L_1 is suitably chosen. The limit point of DCA-L_{1-2} algorithm is a directional stationary point, which is observed numerically to have sparsity upper bounded by the rank of the sensing matrix at a high probability. In numerical experiments in one and two dimensions, the two non-convex penalties produced better solutions either in the relative accuracy of ground truth recovery or seeking a sparse solution while maintaining the measurement constraints when peak distance of the sparse solutions is below the Rayleigh length (the classical barrier).

Acknowledgment

The authors would like to thank C. Fernandez–Granda at New York University, V. Morgenshtern at Stanford University, J–S. Pang at University of Southern California, and M. Yan at Michigan State University for helpful discussions.

References

1. Aubel, C., Stotz, D., Bölcskei, H.: A theory of super-resolution from short-time fourier transform measurements. Tech. rep., arXiv preprint arXiv:1509.01047 (2015)
2. Azais, J.M., De-Castro, Y., Gamboa, F.: Spike detection from inaccurate samplings. Appl. Comput. Harmon. Anal. **38**(2), 177–195 (2015)
3. Borman, S., Stevenson, R.L.: Super-resolution from image sequences-a review. In: Midwest Symposium on Circuits and Systems, pp. 374–378 (1998)
4. Candes, E., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. **59**, 1207–1223 (2006)
5. Candès, E.J., Fernandez-Granda, C.: Super-resolution from noisy data. J. Fourier Anal. Appl. **19**(6), 1229–1254 (2013)
6. Candès, E.J., Fernandez-Granda, C.: Towards a mathematical theory of super-resolution. Comm. Pure Appl. Math. **67**(6), 906–956 (2014)
7. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process. Lett **10**(14), 707–710 (2007)
8. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 3869–3872 (2008)
9. Chen, X., Xu, F., Ye, Y.: Lower bound theory of nonzero entries in solutions of \ell_2-\ell_p minimization. SIAM Journal on Scientific Computing **32**(5), 2832–2852 (2010)
10. De-Castro, Y., Gamboa, F.: Exact reconstruction using beurling minimal extrapolation. J. Math. Anal. Appl. **395**(1), 336–354 (2012)
11. Demanet, L., Nguyen, N.: The recoverability limit for superresolution via sparsity. Tech. rep., arXiv preprint arXiv:1502.01385 (2015)
12. Donoho, D.L.: Superresolution via sparsity constraints. SIAM J. Math. Anal. **23**(5), 1309–1331 (1992)
13. Donoho, D.L.: Compressed sensing. IEEE Trans. on Inform. Theory **52**(4) (2006)
14. Donoho, D.L., Johnstone, I.M., Hoch, J.C., Stern, A.S.: Maximum entropy and the nearly black object. J. Roy. Sat. Soc. B Met. pp. 41–81 (1992)
15. Duval, V., Peyré, G.: Exact support recovery for sparse spikes deconvolution. Found. Comput. Math. pp. 1–41 (2015)
16. Facchinei, F., Pang, J.S.: Finite-dimensional variational inequalities and complementarity problems. Springer Science & Business Media (2007)
17. Fannjiang, A., Liao, W.: Coherence pattern-guided compressive sensing with unresolved grids. SIAM J. Imaging Sci. **5**(1), 179–202 (2012)
18. Fernandez-Granda, C.: Super-resolution of point sources via convex programming. Tech. rep., arXiv preprint arXiv:1507.07034 (2015)
19. Goodman, J.W.: Introduction to Fourier optics. Roberts and Company Publishers (2005)
20. 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)

21. Lou, Y., Osher, S., Xin, J.: Computational aspects of constrained L1-L2 minimization for compressive sensing. In: Model. Comput. & Optim. in Inf. Syst. & Manage. Sci., pp. 169–180. Springer (2015)
22. 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., online: Oct 2014, DOI 10.1007/s10915-014-9930-1 (2014)
23. Mallat, S., Yu, G.: Super-resolution with sparse mixing estimators. IEEE Trans. Image Process. **19**(11), 2889–2900 (2010)
24. Marquina, A., Osher, S.: Image super-resolution by TV-regularization and Bregman iteration. J. Sci. Computing **37**(3), 367–382 (2008)
25. Morgenshtern, V.I., Candès, E.J.: Super-resolution of positive sources: the discrete setup. Tech. rep., arXiv preprint arXiv:1504.00717 (2015)
26. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: a technical overview. IEEE Signal Process. Mag. **20**(3), 21–36 (2003)
27. Peleg, D., Meir, R.: A bilinear formulation for vector sparsity optimization. Signal Process. **88**(2), 375–389 (2008)
28. Pham-Dinh, T., Le-Thi, H.A.: Convex analysis approach to d.c. programming: Theory, algorithms and applications. Acta Mathematica Vietnamica **22**(1), 289–355 (1997)
29. 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)
30. Pham-Dinh, T., Le-Thi, H.A.: The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems. Annals of Operations Research **133**(1-4), 23–46 (2005)
31. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the non-local-means to super-resolution reconstruction. IEEE Trans. Image Process. **18**(1), 36–51 (2009)
32. Shahram, M., Milanfar, P.: Statistical and information-theoretic analysis of resolution in imaging. IEEE trans. Inf. Theory **8**(52), 3411–3437 (2006)
33. Shen, X., Pan, W., Zhu, Y.: Likelihood-based selection and sharp parameter estimation. J. Am. Statist. Assoc. **107**(497), 223–232 (2012)
34. Tang, G., Bhaskar, B.N., Recht, B.: Near minimax line spectral estimation. IEEE Trans. Inf. Theory **61**(1), 499–512 (2015)
35. Tropp, J.: Greed is good: Algorithmic results for sparse approximation. IEEE Trans. Inform. Theory **50**, 2231–2242 (2004)
36. Tropp, J., Gilbert, A.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inform. Theory **53**(12), 4655–4666 (2007)
37. Wang, Z., Liu, H., Zhang, T.: Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. Annals of statistics **42**(6), 2164 (2014)
38. Xu, Z., Chang, X., Xu, F., Zhang, H.: $L_{1/2}$ Regularization: A Thresholding Representation Theory and a Fast Solver. IEEE Trans. on Neural Networks **23**, 1013–1027 (2012)
39. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Proc. **19**(11), 2861–2873 (2010)
40. Yin, P., Lou, Y., He, Q., Xin, J.: Minimization of $l_1 - l_2$ for compressed sensing. SIAM J. Sci. Comput. **37**, A536–A563 (2015)
41. Zhang, T.: Multi-stage convex relaxation for learning with sparse regularization. In: Adv. Neural Inf. Process. Syst., pp. 1929–1936 (2009)