(This paper was published in *SIAM Journal on Optimization*, 22 (2012), no.3, pp. 1065-1088.)

REWEIGHTED ℓ_1-MINIMIZATION FOR SPARSE SOLUTIONS TO UNDERDETERMINED LINEAR SYSTEMS *

YUN-BIN ZHAO† AND DUAN LI‡

Abstract. Numerical experiments have indicated that the reweighted ℓ_1-minimization performs exceptionally well in locating sparse solutions of underdetermined linear systems of equations. We show that reweighted ℓ_1-methods are intrinsically associated with the minimization of the so-called merit functions for sparsity, which are essentially concave approximations to the cardinality function. Based on this observation, we further show that a family of reweighted ℓ_1-algorithms can be systematically derived from the perspective of concave optimization through the linearization technique. In order to conduct a unified convergence analysis for this family of algorithms, we introduce the concept of range space property (RSP) of a matrix, and prove that if its adjoint has this property, the reweighted ℓ_1-algorithm can find a sparse solution to the underdetermined linear system provided that the merit function for sparsity is properly chosen. In particular, some convergence conditions for the Candès-Wakin-Boyd method and the recent ℓ_p-quasi-norm-based reweighted ℓ_1-method can be obtained as special cases of the general framework.

Key words. Reweighted ℓ_1-minimization, sparse solution, underdetermined linear system, concave minimization, merit function for sparsity, compressed sensing.

AMS subject classifications. 15A06, 15A29, 65K05, 90C25, 90C26, 90C59

1. Introduction.

Given an $m \times n$ matrix A with $m \leq n$ and a nonzero vector $b \in R^m$, the linear system $Ax = b$ has infinitely many solutions when the system is underdetermined. Depending on the nature of source problems, we are often interested in finding a particular solution, and thus optimization methods come into a play through certain merit functions that measure the desired special structure of the solution.

One of the recent interests is to find the sparsest solution of an underdetermined linear system, which has found many applications in signal and image processing [16, 3, 2]. To find a sparsest solution of $Ax = b$, perhaps the ideal merit function is the cardinality of a vector, denoted by $\|x\|_0$, i.e., the number of nonzero components of x. Clearly, the set of the sparsest solutions of $Ax = b$ coincides with the set of solutions to the cardinality minimization problem

$$(P_0) \qquad \text{minimize } \{\|x\|_0 : Ax = b\},$$

which is an NP-hard discrete optimization problem [35]. The recent study in the field of compressed sensing nevertheless shows that not all cardinality minimization problems are equally hard, and there does exist a class of matrices A such that the problem (P_0) is computationally tractable. These matrices can be characterized by such concepts as the *spark* which was formally defined by Donoho and Elad [17], *restricted isometry property* (RIP) introduced by Candès and Tao [8], *mutual coherence* (MC) [34, 18, 17], and *null space property* (NSP) [17, 27, 13, 43].

*This work was partially supported by the Research Grants Council of Hong Kong under grant CUHK 414610.

†School of Mathematics, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom (`y.zhao.2@bham.ac.uk`).

‡Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, NT, Hong Kong (`dli@se.cuhk.edu.hk`).

Problem (P_0) is not easy to solve in general. From a convex analysis point of view, a natural methodology is to minimize the convex envelope of $\|x\|_0$. It is well-known that ℓ_1-norm is the convex envelope of $\|x\|_0$ over the region $\{x : \|x\|_\infty \leq 1\}$. One of the main approaches to attack (P_0) is through ℓ_1-minimization

$$(P_1) \qquad \text{minimize } \{\|x\|_1 : Ax = b\},$$

which is identical to a linear program (LP) and hence can be solved very efficiently. Using ℓ_1-norm as a merit function for sparsity can be traced back several decades in a wide range of areas from seismic traces [40], sparse-signal recovery [19, 11], sparse-model selection (LASSO algorithm) in statistics [39] to image processing [1], and continues its growth in other areas. A brief history of using ℓ_1-minimization can be found in [9]. The ℓ_1-minimization approach has been generalized to low-rank matrix recovery/matrix completion (see e.g., [4, 7]), and to the so-called matrix rank minimization as well (see e.g., [20, 21, 22, 37, 38, 44]). Thanks to the intensive study in the field of compressed sensing (see e.g., [16, 3]), both theoretical properties and numerical performances of ℓ_1-minimization have been well-established over the past few years. Various conditions (including the above-mentioned MC, RIP, NSP, and others) for the relationship

$$\arg\min\{\|x\|_0 : Ax = b\} = \{x^*\} = \arg\min\{\|x\|_1 : Ax = b\}$$

have been developed (see e.g., [18, 17, 27, 5, 6, 41, 2, 13, 43]). In terms of sparse signal recovery, this relationship implies that ℓ_1-minimization allows recovery of sparse signals from a small number of measurements [17, 16, 5, 6]. A comprehensive discussion and survey of recent results in this field can be found in [16, 3, 2, 23].

Inspired by the efficiency of ℓ_1-minimization, it is natural to ask whether there are other alternatives which can either be comparable to or even outperform ℓ_1-minimization in finding sparse solutions of linear systems. Numerical experiments indicate that the reweighted ℓ_1-minimization does outperform unweighted ℓ_1-minimization in many situations [9, 24, 31, 12, 14]. The key feature of reweighted ℓ_1-minimization is the solution of a series of weighted ℓ_1-problems

$$(P_{W^k}) \qquad \min\{\|W^k x\|_1 : \; Ax = b\},$$

where $W^k = \text{diag}(w^k)$ and $w^k = (w_1^k, ..., w_n^k)^T \in R_+^n$ is the vector of weights determined by the current iterate $x^k = (x_1^k, ..., x_n^k)^T \in R^n$. The solution to (P_{W^k}) is set to be x^{k+1}, based on which the new weight W^{k+1} is computed. Some theoretical analysis has been made for reweighted ℓ_1-algorithms since 2008, when Candès, Wakin and Boyd [9] proposed the reweighted method with

$$w_i^k = \frac{1}{|x_i^k| + \varepsilon}, \; i = 1, ..., n, \; \text{for } \varepsilon > 0. \tag{1.1}$$

We refer to this as the CWB method in this paper. Needell [36] showed that the error bounds for noisy signal recovery via the CWB method can be tighter than those of standard ℓ_1-minimization. Foucart and Lai [24] proved that under the assumption of RIP, the reweighted ℓ_1-method with weights

$$w_i^k = \frac{1}{(|x_i^k| + \varepsilon)^{1-p}}, \; i = 1, ..., n, \tag{1.2}$$

where $p \in (0,1)$ is a given parameter, can exactly recover the sparse signal. Lai and Wang [31], and Chen and Zhou [12] further prove that under RIP/NSP-type conditions, the accumulation points of the sequence generated by the reweighted ℓ_1-algorithm with weights (1.2) can converge to a stationary point of certain '$\ell_2 - \ell_p$' minimization problem that is an approximation to (P_0). Note that the objective of (P_{W^k}) is separable in x. It is worth mentioning that some nonseparable iterative reweighted methods were also proposed recently by Wipf and Nagarajan [42]. However, as pointed out in [9], the understanding of reweighted ℓ_1-minimization remains very incomplete so far. Even the convergence property of the CWB algorithm remains unclear at present.

We note that while the major study of reweighted ℓ_1-minimization is carried out recently, the reweighted least square (RLS) method has a relatively long history. RLS was proposed by Lawson [32] in 1960s, and was extended to ℓ_p-minimization later. The idea of RLS methods was also used in the algorithm for robust statistical estimation [29], and in FOCUSS methods [25] for the sparse solution of linear systems. The interplay of null space property (NSP), ℓ_1-minimization, and RLS method has been clarified recently in [14].

The main contributions of this paper are as follows. First, we provide a unified derivation of the reweighted ℓ_1-minimization, which can be viewed as the first-order method for concave programming with an objective called the merit function for sparsity that is certain approximation of $\|x\|_0$. Second, we provide a unified theoretical analysis for a large family of reweighted ℓ_1-algorithms for the sparse solution of underdetermined linear systems. Interestingly, various new reweighted ℓ_1-methods can be systematically constructed/extracted from this family. To show the generic convergence of this family of algorithms, we introduce the new concept of *range space property* (RSP) of a matrix, which is different from (but has some link to) RIP and NSP. One of the results in this paper claims that if A^T has the RSP of order K with constant $\rho > 0$ satisfying $(1+\rho)K < n$, then there exist a large number of merit functions for sparsity, associated with which reweighted ℓ_1-algorithms can generate a $\lfloor (1+\rho)K \rfloor$-sparse solution to the linear system. Based on optimization theory merely, the analysis in this paper is remarkably different from the existing RIP/NSP-based analysis. It should be stressed that the CWB method, and the algorithm with weights (1.2) are special cases of our general framework of reweighted ℓ_1-algorithms, and hence a new convergence result for these existing algorithms under RSP assumption has been established for the first time in this paper. Finally, we carry out some numerical experiments to demonstrate the performance of several new and existing reweighted ℓ_1-algorithms in locating the sparsest solution of linear systems. Our numerical results show that in many situations the NW1-NW4 algorithms proposed in this paper, the CWB method, and the algorithm with weights (1.2) do remarkably outperform the standard ℓ_1-minimization (See section 4 for details).

This paper is organized as follows. In section 2, a unified approach for deriving a large family of reweighted ℓ_1-algorithms based on concave merit functions for sparsity is proposed. In section 3, some convergence properties of this family of algorithms are proved via the range space property. Numerical results are given in section 4, and conclusions are given in the last section.

Notation: Let R^n be the n-dimensional Euclidean space, and R^n_+ and R^n_{++} the sets of nonnegative and positive vectors, respectively. For $x \in R^n$, $\|x\|_p = (\sum_{i=1}^n |x_j|^p)^{1/p}$ denotes the ℓ_p- (quasi-)norm where $p \in (0,\infty)$. Given a set $S \subseteq \{1,2,...,n\}$, the symbol $|S|$ denotes the cardinality of S, and $S_c = \{1,2,...,n\} \backslash S$ is the complement of S. For a matrix A and a vector $x \in R^n$, we use the notation A_S to denote the submatrix extracted from A with column indices in S, and x_S the subvector extracted from x with component indices in S. Let $\mathcal{T}(x) = \{i : x_i \neq 0\}$ denote the support of x. Let $\sigma(x)$ be the vector which is the non-increasing rearrangement of the absolute values of the entries of $x \in R^n$, and let $|x| = (|x_1|, ..., |x_n|)^T \in R^n$. Clearly, we have $\sigma(x) = \sigma(|x|)$, and $\sigma(x)_i$ is the ith largest component of $|x|$. In this paper, $x \in R^n$ is said to be K-sparse if x contains at most K nonzero components. Thus x is K-sparse if and only if $\sigma(x)_{K+1} = 0$. For any x and y in R^n, the inequality $x \leq y$ means $x_i \leq y_i$ for all $i = 1, ..., n$.

2. A unified framework of reweighted ℓ_1-minimization. The central idea of reweighted ℓ_1-algorithms is to define a weight based on the current iterate x^k, solve the weighted ℓ_1-minimization for this weight, and then use its solution to define a new weight. The weight is used to penalize the components which are small, in order to drive them to be as small as possible via minimizing the weighted ℓ_1-norm. In this section, we introduce a unified framework for reweighted ℓ_1-minimization. To this end, we need to specify a family of merit functions for sparsity. A function from R^n to R is said to be a *merit function for sparsity* if it is an approximation to $\|x\|_0$ in some sense. For instance, ℓ_1-norm is a convex relaxation of $\|x\|_0$ over $\|x\|_\infty \leq 1$, and $\|x\|_p$, $p \in (0, 1)$, is also a merit function for sparsity since $\|x\|_p^p \to \|x\|_0$ as $p \to 0$. Clearly, there exist a vast number of merit functions for sparsity. Minimizing such a function may drive the variable x to become sparse when a sparse solution exists.

From a computational point of view, a merit function for sparsity should admit certain desired properties such as convexity or concavity. Due to the NP-hardness of (P_0), it seems that there is no hope to approximate $\|x\|_0$ in R^n to any level of accuracy by a convex function. On the contrary, there exist various concave functions that can approximate $\|x\|_0$ to any level of accuracy. Concave merit functions appear more natural than convex ones when finding the sparsest solution of linear systems, since the 'bulged' feature of convex merit functions might prohibit locating the sparsest solution in some situations. This phenomenon was observed by Harikumar and Bresler [28]. So, throughout the remainder of this paper, we focus on concave merit functions for sparsity. For simplicity, we consider separable concave merit functions of the form

$$F(x) = \sum_{i=1}^{n} \phi_i(|x_i|),$$

where $\phi_i : R_+ \to R$ is called the kernel function. Given a separable function $F(x)$ as above and a set $S \subseteq \{1, ..., n\}$, we use $F(x_S)$ to denote the reduced separable function, i.e., $F(x_S) := \sum_{i \in S} \phi_i(|x_i|)$. To avoid the division by zero when computing the gradient of a merit function, we perturb the function by replacing $|x_i|$ by $|x_i| + \varepsilon$, where $\varepsilon > 0$ is a given parameter. This leads to the approximation problem of (P_0):

$$\min \left\{ F_\varepsilon(x) = \sum_{i=1}^{n} \phi_i(|x_i| + \varepsilon) : Ax = b \right\}. \tag{2.1}$$

For example, if all $\phi_i(t) = t^p$ with $p \in (0, 1)$, problem (2.1) is the ℓ_p-quasi-norm minimization (see e.g., [10, 15, 24]). By setting $\phi_i(t) = \log(t)$, the function $F_\varepsilon(x) = \sum_{x_i \neq 0} \log(|x_i| + \varepsilon)$ was used by Gorodnitsky and Rao [25] to design the FOCUSS algorithm for sparse signal reconstruction.

We use $e = (1, 1, ..., 1)^T \in R^n$ throughout to denote the vector of ones. Recall that a function $f : R^n \to R$ is said to be coercive in the region $D \subseteq R^n$ if $f(x) \to \infty$

as $\|x\| \to \infty$ and $x \in D$. We now specify a class of merit functions satisfying the following assumption.

ASSUMPTION 2.1. *For any given $\varepsilon > 0$, the merit function $F_\varepsilon : R^n \to R$ satisfies the following properties:*

(a) *$F_\varepsilon(x) = F_\varepsilon(|x|)$ for any $x \in R^n$, and $F_\varepsilon(x)$ is separable in x, and twice continuously differentiable with respect to $x \in R^n_+$.*

(b) *In R^n_+, $F_\varepsilon(x)$ is strictly increasing with respect to every component x_i and ε, and for any given $\gamma > 0$ there exists a finite number $Q(\gamma)$ such that for any $S \subseteq \{1, ..., n\}$, $g(x_S) := \inf_{\varepsilon \downarrow 0} F_\varepsilon(x_S) \geq Q(\gamma)$ provided $x_S \geq \gamma e_S$, and $g(x_S)$ is coercive in the set $\{x_S : x_S \geq \gamma e_S\}$.*

(c) *In R^n_+, the gradient satisfies that $\nabla F_\varepsilon(x) \in R^n_{++}$ for any $(x, \varepsilon) \in R^n_+ \times R_{++}$, $[\nabla F_\varepsilon(x)]_i \to \infty$ as $(x_i, \varepsilon) \to 0$, and for any given $x_i > 0$ the component $[\nabla F_\varepsilon(x)]_i$ is continuous in ε and tends to a finite positive number (dependent only on x_i) as $\varepsilon \to 0$.*

(d) *In R^n_+, the Hessian satisfies that $y^T \nabla^2 F_\varepsilon(x) y \leq -C(\varepsilon, r)\|y\|^2$ for any $y \in R^n$ and $x \in R^n_+$ with $\|x\| \leq r$, where $r > 0$ and $C(\varepsilon, r) > 0$ are constants, and $C(\varepsilon, r)$ is continuous in ε and bounded away from zero as $\varepsilon \to 0$. (Thus $F_\varepsilon(x)$ is strictly concave with respect to x.)*

There exists a large family of functions satisfying the above assumption, and we show that it is very easy to construct examples in this family. Note that any nonnegative combination of a finite number of functions satisfying Assumption 2.1 still satisfies this assumption. So the set of such functions is a convex cone, denoted by $\mathcal{M}$, i.e.,

$$\mathcal{M} = \{F_\varepsilon : \ F_\varepsilon \text{ satisfies Assumption 2.1 for } \varepsilon > 0\}.$$

Based on the merit functions in $\mathcal{M}$, we can derive a family of reweighted ℓ_1-minimization algorithms. Note that for any $F_\varepsilon \in \mathcal{M}$, problem (2.1) can be rewritten as

$$\min\{F_\varepsilon(v) : Ax = b, |x| \leq v\} = \min_{(x,v)\in\mathcal{F}} F_\varepsilon(v), \tag{2.2}$$

where $|x| = (|x_1|, ..., |x_n|)^T$ and

$$\mathcal{F} = \{(x, v) : Ax = b, |x| \leq v\}.$$

Throughout this paper, *we assume that the system $Ax = b$ has a solution, i.e., $\mathcal{F} \neq \emptyset$. This can be always guaranteed when the system is underdetermined.*

At the current point v^k, since $F_\varepsilon(v) = F_\varepsilon(v^k) + \nabla F_\varepsilon(v^k)^T(v - v^k) + o(\|v - v^k\|)$, the simplest tractable approximation to the concave minimization (2.2) is the problem of minimizing the linear approximation of $F_\varepsilon(v)$ over the same feasible set. Thus it makes sense to solve the following problem to generate the next point (x^{k+1}, v^{k+1}) :

$$\begin{aligned}(x^{k+1}, v^{k+1}) &= \arg \min_{(x,v)\in\mathcal{F}} \left\{F_\varepsilon(v^k) + \nabla F_\varepsilon(v^k)^T(v - v^k)\right\} \\ &= \arg \min_{(x,v)\in\mathcal{F}} \nabla F_\varepsilon(v^k)^T v, \end{aligned} \tag{2.3}$$

which is an LP. Note that $v^k \in R^n_+$ and $\nabla F_\varepsilon(v^k) \in R^n_{++}$ (by Assumption 2.1(c)). It is easy to see that the optimal solution (x^{k+1}, v^{k+1}) of (2.3) always satisfies

$$v^{k+1} = |x^{k+1}| \text{ for all } k \geq 0. \tag{2.4}$$

Hence it follows from (2.4) and the positiveness of $\nabla F_\varepsilon(v^k)$ that

$$\nabla F_\varepsilon(v^k)^T v^{k+1} = \nabla F_\varepsilon(|x^k|)^T |x^{k+1}| = \left\|\mathrm{diag}\left(\nabla F_\varepsilon(|x^k|)\right) x^{k+1}\right\|_1 .$$

Therefore, the iterative scheme (2.3) is nothing but

$$x^{k+1} = \arg\min\left\{\left\|\mathrm{diag}\left(\nabla F_\varepsilon(|x^k|)\right) x\right\|_1 : \ Ax = b\right\}, \tag{2.5}$$

which is the *reweighted ℓ_1-minimization* with weight $w^k = \nabla F_\varepsilon(|x^k|) \in R^n_{++}$. By Assumption 2.1(c), we have that $[\nabla F_\varepsilon(v^k)]_i \to \infty$ as $(v_i, \varepsilon) \to 0$, so for small $v_i^k (= |x_i^k|)$ and $\varepsilon > 0$, the corresponding weight $[\nabla F_\varepsilon(v^k)]_i$ is large. The iterative scheme (2.3) (equivalently, (2.5)) provides a unified approach to derive reweighted ℓ_1-algorithms for sparse solutions of linear systems, which can be described as follows.

Algorithm 2.2 (Reweighted ℓ_1-minimization)

S1. Choose $\alpha,\ \varepsilon_0 \in (0, 1)$, and let $(x^0, v^0) \in R^n \times R^n_+$ be an initial point.

S2. At the current iterate (x^k, v^k) with $\varepsilon_k > 0$, compute

$$(x^{k+1}, v^{k+1}) = \arg\min_{(x,v)\in\mathcal{F}} \left(\nabla F_{\varepsilon_k}(v^k)\right)^T v$$

(i.e., $x^{k+1} = \arg\min\{\|\mathrm{diag}(\nabla F_{\varepsilon_k}(|x^k|))x\|_1 : Ax = b\}$).

S3. Set $\varepsilon_{k+1} = \alpha\varepsilon_k$. Replace $(x^k, v^k, \varepsilon_k)$ by $(x^{k+1}, v^{k+1}, \varepsilon_{k+1})$ and repeat S2.

Note that ε is reduced by a factor after every iteration, and several stoping criteria can be used in Algorithm 2.2, for instance, $\varepsilon_k < \varepsilon^*$ or $\|v^{k+1} - v^k\| < \varepsilon^*$ (ε^* is a prescribed tolerance). We note that the general concave optimization theory is a long-lasting research topic in the field of global optimization (see for instance [30]). From a concave programming point of view, Algorithm 2.2 is not new, which is essentially a linearization method for concave minimization. However, it is the objective function (i.e., the merit function for sparsity) that makes the algorithm unique.

Algorithm 2.2 is a general framework for reweighted ℓ_1-algorithms, from which many specific algorithms, including some existing ones, can be immediately obtained by choosing various merit functions. In this paper, we focus on the merit functions in $\mathcal{M}$. A few examples of such functions together with their gradients are given as follows.

Example 2.3. (i) Note that $\left(n + \frac{\sum_{i=1}^n \log(|x_i|+\varepsilon)}{-\log\varepsilon}\right) \to \|x\|_0$ as $\varepsilon \to 0$. This motivates the merit function

$$F_\varepsilon(x) = \sum_{i=1}^n \log(|x_i| + \varepsilon)$$

in $\mathcal{M}$. The gradient of this function at $x \in R^n_+$ is given by $\nabla F_\varepsilon(x) = \left(\frac{1}{x_1+\varepsilon}, \ldots, \frac{1}{x_n+\varepsilon}\right)^T$ $\in R^n_{++}$, and the Hessian $\nabla^2 F_\varepsilon(x) = -\mathrm{diag}\left(\frac{1}{(x_1+\varepsilon)^2}, \ldots, \frac{1}{(x_n+\varepsilon)^2}\right)$. Using this merit function, (2.3) is exactly the CWB reweighted ℓ_1-method with weights given by (1.1), which has been studied by several authors [21, 22, 33, 9, 36]. A convergence result for this method can be obtained as a special case of our general framework.

(ii) Let $p \in (0, 1)$. It is easy to verify that the function

$$F_\varepsilon(x) = \sum_{i=1}^n \log\left(|x_i| + \varepsilon + (|x_i| + \varepsilon)^p\right)$$

is also in $\mathcal{M}$. At $x \in R^n_+$, the gradient is given by $[\nabla F_\varepsilon(x)]_i = \frac{p+(x_i+\varepsilon)^{1-p}}{(x_i+\varepsilon)^{1-p}[x_i+\varepsilon+(x_i+\varepsilon)^p]}$, $i = 1, ..., n$, and the Hessian $\nabla^2 F_\varepsilon(x)$ is a diagonal matrix with diagonal entries

$$-\frac{1}{[x_i+\varepsilon+(x_i+\varepsilon)^p]^2}\left(1+\frac{p(3-p)}{(x_i+\varepsilon)^{1-p}}+\frac{p}{(x_i+\varepsilon)^{2(1-p)}}\right),\ i=1,...,n.$$

This merit function yields a reweighted ℓ_1-algorithm with the weights

$$w_i = [\nabla F_\varepsilon(|x|)]_i = \frac{p+(|x_i|+\varepsilon)^{1-p}}{(|x_i|+\varepsilon)^{1-p}\,[|x_i|+\varepsilon+(|x_i|+\varepsilon)^p]},\ i=1,...,n,$$

(termed the 'NW1' algorithm) which has not been studied in the literature so far.

Example 2.4. Note that $\|x\|_p^p \to \|x\|_0$ as $p \to 0$. For a given $p \in (0,1)$, we define $F_\varepsilon(x) = \frac{1}{p}\sum_{i=1}^n (|x_i|+\varepsilon)^p$, which is in $\mathcal{M}$. At $x \in R^n_+$, the gradient of this function is given by

$$\nabla F_\varepsilon(x) = \left(\frac{1}{(x_1+\varepsilon)^{1-p}}, ..., \frac{1}{(x_n+\varepsilon)^{1-p}}\right)^T \in R^n_{++},$$

and the Hessian $\nabla^2 F_\varepsilon(x) = -\text{diag}\left(\frac{1}{(x_1+\varepsilon)^{2-p}}, ..., \frac{1}{(x_n+\varepsilon)^{2-p}}\right)$. By this merit function, (2.3) is exactly the reweighted ℓ_1-method with weights $w_i = \frac{1}{(|x_i|+\varepsilon)^{1-p}}, i = 1, ..., n$, (termed the 'Wl$_p$' algorithm) which was recently studied in [24, 31, 12]. Some new properties of this method can be extracted from our general results in this paper.

Example 2.5. Let p and $q \in (0,1)$ be given. Define

$$F_\varepsilon(x) = \frac{1}{p}\sum_{i=1}^n (|x_i|+\varepsilon+(|x_i|+\varepsilon)^q)^p,$$

which is in $\mathcal{M}$. The gradient of this function at $x \in R^n_+$ is given by

$$[\nabla F_\varepsilon(x)]_i = \frac{q+(x_i+\varepsilon)^{1-q}}{(x_i+\varepsilon)^{1-q}\,[x_i+\varepsilon+(x_i+\varepsilon)^q]^{1-p}},\ i=1,...,n,$$

and the Hessian of this function is a diagonal matrix with diagonal entries

$$-\frac{1}{[x_i+\varepsilon+(x_i+\varepsilon)^q]^{2-p}}\left((1-p)+\frac{2q(1-p)+q(1-q)}{(x_i+\varepsilon)^{1-q}}+\frac{q^2(1-p)+q(1-q)}{(x_i+\varepsilon)^{2(1-q)}}\right).$$

Based on this merit function, the associated reweighted ℓ_1-minimization (termed the 'NW2' algorithm) remains new. It is interesting to note that the parameter q above can be chosen to be greater than 1 if we restrict the range of p. For instance, we can let $q = 2$ and $p \le \frac{1}{2}$. Then the function

$$F_\varepsilon(x) = \frac{1}{p}\sum_{i=1}^n \left(|x_i|+\varepsilon+(|x_i|+\varepsilon)^2\right)^p$$

is in $\mathcal{M}$. The associated reweighted ℓ_1-minimization (termed the 'NW3' algorithm) with

$$w_i = [\nabla F_\varepsilon(|x|)]_i = \frac{1+2(|x_i|+\varepsilon)}{(|x_i|+\varepsilon+(|x_i|+\varepsilon)^2)^{1-p}}, i=1,...,n$$

is also a new algorithm. Numerical experiments show that NW2 and NW3 are quite strong in finding the sparsest solution of linear systems in many situations (see section 4 for details).

EXAMPLE 2.6. Let $p \in (0,\infty)$ be fixed. Note that $\varphi_\varepsilon(x) = \sum_{i=1}^n \frac{(|x_i|+\varepsilon)^p-\varepsilon^p}{(|x_i|+\varepsilon)^p} \to \|x\|_0$ as $\varepsilon \to 0$ (in particular, for $p=1$, $\varphi_\varepsilon(x) = \sum_{i=1}^n \frac{|x_i|}{|x_i|+\varepsilon} \to \|x\|_0$ as $\varepsilon \to 0$). This motivates the merit function $g_\varepsilon(x) = -\sum_{i=1}^n \frac{1}{(|x_i|+\varepsilon)^p}$ which satisfies almost all conditions of Assumption 2.1 except for the coercivity. However, we can make this function coercive as follows: Adding a log function to $g_\varepsilon(x)$ yields the merit function (in $\mathcal{M}$)

$$F_\varepsilon(x) = \sum_{i=1}^n \left[\log(|x_i|+\varepsilon) - \frac{1}{(|x_i|+\varepsilon)^p}\right], \tag{2.6}$$

for which the gradient at $x \in R^n_+$ is given by $\nabla F_\varepsilon(x) = \left(\frac{1+(x_1+\varepsilon)^p}{(x_1+\varepsilon)^{p+1}}, ..., \frac{1+(x_n+\varepsilon)^p}{(x_n+\varepsilon)^{p+1}}\right)$, and the Hessian $\nabla^2 F_\varepsilon(v) = -\text{diag}\left(\frac{p+1+(v_1+\varepsilon)^p}{(v_1+\varepsilon)^{p+2}}, ..., \frac{p+1+(v_n+\varepsilon)^p}{(v_n+\varepsilon)^{p+2}}\right)$. The associated reweighted ℓ_1-minimization (termed the 'NW4' algorithm) uses the weights $w_i = \frac{1+(|x_i|+\varepsilon)^p}{(|x_i|+\varepsilon)^{p+1}}$, $i=1,...,n$. Another way to make g_ε coercive is to add the ℓ_1-norm, leading to the merit function

$$F_\varepsilon(x) = \sum_{i=1}^n \left(|x_i| - \frac{1}{(|x_i|+\varepsilon)^p}\right),$$

which is in $\mathcal{M}$ and yields a new reweighted ℓ_1-method with weights $w_i = [\nabla F_\varepsilon(|x|)]_i = \frac{1+(|x_i|+\varepsilon)^{p+1}}{(|x_i|+\varepsilon)^{p+1}}$, $i=1,...,n$.

Note that when $p \to 0$, the function in Example 2.4 reduces to that of Example 2.3 (i), and the function in Example 2.5 reduces to that of Example 2.3(ii). Thus the CWB method can be viewed as an extreme case of the Wl_p method as $p \to 0$, and it can be also viewed as an extreme case of (2.6) as $p \to 0$.

Remark. Examples 2.3-2.6 show that it is not difficult to construct merit functions for sparsity in $\mathcal{M}$, and the well-known CWB and Wl_p methods fall into this family. To construct a function in $\mathcal{M}$, we may start with a choice of the kernel function $\phi_i(t)$, which should be twice continuously differentiable, strictly concave and increasing in R^n_+. Every merit function in $\mathcal{M}$ yields a reweighted ℓ_1-minimization algorithm for the sparse solution of linear systems. The family of such algorithms is large. Because $\mathcal{M}$ is a convex cone, any positive combination of functions in $\mathcal{M}$ is in this family. Thus, the combination can be used to generate new merit functions from known ones. For instance, a simple combination of Example 2.3(i) and Example 2.4 yields the function

$$F_\varepsilon(v) = \left(\sum_{i=1}^n \log(v_i+\varepsilon)\right) + \left(\sum_{i=1}^n (v_i+\varepsilon)^p\right)$$

in $\mathcal{M}$, where $p \in (0,1)$. Its kernel function is $\phi_i(t) = \log(t) + t^p$ $(i=1,...,n)$. In many cases, applying the log operation to a nonnegative convex function a finite number of times, we may reverse the convexity to concavity (see Zhao et al [45]). For instance, if $\phi: R_+ \to R_+$ is a twice differentiable, strictly increasing and convex function obeying $\phi''(t)(1+\phi(t)) < (\phi'(t))^2$, then $\log(1+\phi(t))$ is concave, so is $\log(1+\log(1+\phi(t)))$. Since the log operation maintains the coercivity and monotonicity of the original function, this strategy can be used to construct a concave merit function for sparsity from a convex function.

3. Convergence analysis of reweighted ℓ_1-minimization. From the above section, we see that there exist infinitely many reweighted algorithms. Therefore it is necessary to study these algorithms in a unified approach, in order to identify their common properties. The remainder of this paper is devoted to this task. We carry out a unified convergence analysis for Algorithm 2.2 based on the merit functions for sparsity in $\mathcal{M}$. We note that Algorithm 2.2 can start at any initial point (x^0, v^0) with $v^0 = |x^0|$ where x^0 is not necessarily a solution to $Ax = b$. After the first step, the algorithm will generate an iterate (x^1, v^1) satisfying $v^1 = |x^1|$ and $Ax^1 = b$, from which all subsequent iterates satisfy $Ax = b$ and (2.4), i.e., $v^k = |x^k|$ for all $k \geq 1$.

The first result below shows that if the algorithm generates at the kth step a new point $v^{k+1} \neq v^k$, then the value of the merit function for sparsity strictly decreases.

LEMMA 3.1. *Suppose that F_ε satisfies Assumption 2.1, i.e., $F_\varepsilon \in \mathcal{M}$. For a given $(\widetilde{x}, \widetilde{v}) \in \mathcal{F}$ with $\widetilde{v} = |\widetilde{x}|$ and a parameter $\varepsilon > 0$, let*

$$(x^+, v^+) = arg \min_{(x,v)\in\mathcal{F}} \nabla F_\varepsilon(\widetilde{v})^T v. \tag{3.1}$$

If $v^+ \neq \widetilde{v}$, then $F_\varepsilon(v^+) < F_\varepsilon(\widetilde{v})$.

Proof. Note that $(\widetilde{x}, \widetilde{v})$ is feasible to the problem (3.1). Thus the minimizer v^+ satisfies that $\nabla F_\varepsilon(\widetilde{v})^T v^+ \leq \nabla F_\varepsilon(\widetilde{v})^T \widetilde{v}$. When $v^+ \neq \widetilde{v}$, we have only two cases.

Case 1. $\nabla F_\varepsilon(\widetilde{v})^T (v^+ - \widetilde{v}) < 0$. In this case, by concavity of F_ε, we have $F_\varepsilon(v^+) \leq F_\varepsilon(\widetilde{v}) + \nabla F_\varepsilon(\widetilde{v})^T (v^+ - \widetilde{v}) < F_\varepsilon(\widetilde{v})$.

Case 2. $\nabla F_\varepsilon(\widetilde{v})^T (v^+ - \widetilde{v}) = 0$. Let $r > 0$ be a constant such that $\|\widetilde{v}\| \leq r$. Then for any sufficiently small $t > 0$, we have

$$\begin{aligned}
&F_\varepsilon(\widetilde{v} + t(v^+ - \widetilde{v})) \\
&= F_\varepsilon(\widetilde{v}) + t\nabla F_\varepsilon(\widetilde{v})^T (v^+ - \widetilde{v}) + \frac{1}{2}t^2 (v^+ - \widetilde{v})^T \nabla^2 F_\varepsilon(\widetilde{v})(v^+ - \widetilde{v}) + o(t^2) \\
&= F_\varepsilon(\widetilde{v}) + \frac{1}{2}t^2 (v^+ - \widetilde{v})^T \nabla^2 F_\varepsilon(\widetilde{v})(v^+ - \widetilde{v}) + o(t^2) \\
&\leq F_\varepsilon(\widetilde{v}) - \frac{t^2}{2} C(\varepsilon, r)\|v^+ - \widetilde{v}\|^2 + o(t^2) \\
&< F_\varepsilon(\widetilde{v}).
\end{aligned} \tag{3.2}$$

The first inequality above follows from the strict concavity of F_ε (Assumption 2.1(d)). The concavity of F_ε also implies that, for any sufficiently small $t > 0$, we have

$$F_\varepsilon(\widetilde{v} + t(v^+ - \widetilde{v})) \geq tF_\varepsilon(v^+) + (1 - t)F_\varepsilon(\widetilde{v}) = t(F_\varepsilon(v^+) - F_\varepsilon(\widetilde{v})) + F_\varepsilon(\widetilde{v}),$$

which, together with (3.2), implies that $F_\varepsilon(v^+) - F_\varepsilon(\widetilde{v}) < 0$. ∎

By the structure of Algorithm 2.2, we have the following corollary, showing that the merit function strictly decreases after every iteration.

COROLLARY 3.2. *Let $F_\varepsilon \in \mathcal{M}$ and $\{(x^k, v^k)\}$ be generated by Algorithm 2.2. Then the sequence $\{F_{\varepsilon_k}(v^k)\}$ is strictly decreasing in the sense that $F_{\varepsilon_{k+1}}(v^{k+1}) < F_{\varepsilon_k}(v^k)$ for all $k \geq 1$.*

Proof. First, we note that $F_{\varepsilon_k}(v^{k+1}) \leq F_{\varepsilon_k}(v^k)$. In fact, it holds trivially if $v^{k+1} = v^k$; otherwise, it holds strictly by Lemma 3.1. Since F_ε is strictly increasing in ε (by Assumption 2.1(b)), we have $F_{\varepsilon_{k+1}}(v^{k+1}) < F_{\varepsilon_k}(v^{k+1}) \leq F_{\varepsilon_k}(x^k)$. ∎

LEMMA 3.3. *Let $F_\varepsilon \in \mathcal{M}$ and $\{(x^k, v^k)\}$ be the sequence generated by Algorithm 2.2. If there is a subsequence, denoted by $\{(x^{k_j}, v^{k_j}) : j = 1, 2, ...,\}$, such that $|x^{k_j}| \geq \gamma e$ for all j where $\gamma > 0$ is a constant, then there exists a finite constant $C^* > 0$*

such that $\sum_{j=1}^{\infty}\left\|v^{k_j+1}-v^{k_j}\right\|_2^2 \leq C^*$. *In particular, we have* $\left\|v^{k_j+1}-v^{k_j}\right\| \to 0$ *as* $j \to \infty$.

Proof. Note that $v^k = |x^k|$ for all $k \geq 1$. By Assumption 2.1(b), it follows from $v^{k_j} = |x^{k_j}| \geq \gamma e$ that

$$F_{\varepsilon_{k_j}}(v^{k_j}) \geq F_{\varepsilon_{k_j}}(\gamma e) \geq \inf_{\varepsilon\downarrow 0} F_\varepsilon(\gamma e) \geq Q(\gamma),$$

and hence the sequence $\{F_{\varepsilon_{k_j}}(v^{k_j})\}$ is bounded below. By Corollary 3.2, the sequence $\{F_{\varepsilon_k}(v^k)\}$ is strictly decreasing, and thus $F_{\varepsilon_{k_j}}(v^{k_j}) \leq F_{\varepsilon_0}(v^0)$ for any j. This, together with the coercivity of $g = \inf_{\varepsilon\downarrow 0} F_\varepsilon$ (Assumption 2.1 (b)), implies that the sequence $\{v^{k_j}\}$ must be bounded. In fact, if $\{v^{k_j}\}$ is unbounded, then there exists a subsequence, denoted still by $\{v^{k_j}\}$, such that $\|v^{k_j}\| \to \infty$. Since $v^{k_j} \geq \gamma e$, it follows from Assumption 2.1 (b) that $F_{\varepsilon_{k_j}}(v^{k_j}) \geq \inf_{\varepsilon\downarrow 0} F_\varepsilon(v^{k_j}) = g(v^{k_j}) \to \infty$ as $\|v^{k_j}\| \to \infty$, which contradicts the inequality $F_{\varepsilon_{k_j}}(v^{k_j}) \leq F_{\varepsilon_0}(v^0)$ for all j. Thus, the sequence $\{v^{k_j}\}$ must be bounded. So there is a positive constant $\gamma' > \gamma$ such that $\gamma e \leq v^{k_j} \leq \gamma' e$ for all j. For any $\varepsilon > 0$, since $F_\varepsilon(v)$ is separable in $v \in R^n_+$ and $[\nabla^2 F_\varepsilon(v)]_{ii}$ is negative (Assumption 2.1 (d)), it implies that for every i, $[\nabla F_\varepsilon(v)]_i$ is decreasing with respect to v_i. Therefore,

$$\nabla F_{\varepsilon_{k_j}}(\gamma' e) \leq \nabla F_{\varepsilon_{k_j}}(v^{k_j}) \leq \nabla F_{\varepsilon_{k_j}}(\gamma e). \tag{3.3}$$

Assumption 2.1 (c) implies that for every i, the components $[\nabla F_\varepsilon(\gamma' e)]_i$ and $[\nabla F_\varepsilon(\gamma e)]_i$ are both positive, and bounded from above and away from zero over the region $\varepsilon \in (0, \varepsilon_0]$. This, together with (3.3), implies that there exist two constants β_1 and β_2 $(0 < \beta_1 < \beta_2)$ such that

$$\beta_1 e \leq \nabla F_{\varepsilon_{k_j}}(v^{k_j}) \leq \beta_2 e \quad \text{for all } j. \tag{3.4}$$

By optimality, we have $\nabla F_{\varepsilon_{k_j}}(v^{k_j})^T v^{k_j+1} \leq \nabla F_{\varepsilon_{k_j}}(v^{k_j})^T v^{k_j}$ for any j. This, together with (3.4) and $\gamma e \leq v^{k_j} \leq \gamma' e$, implies that the sequence $\{v^{k_j+1}\}$ is bounded.

Since $\{F_{\varepsilon_{k_j}}(v^{k_j})\}$ is decreasing and bounded below, we have $F_{\varepsilon_{k_j}}(v^{k_j}) \to F^*$ as $j \to \infty$, where F^* is a constant. Let $r > 0$ be a constant such that

$$\max\{\max_{j\geq 1}\|v^{k_j}\|, \max_{j\geq 1}\|v^{k_j+1}\|\} \leq r,$$

which is finite since $\{v^{k_j}\}$ and $\{v^{k_j+1}\}$ are bounded. Note that $\varepsilon_{k_{(j+1)}} \leq \cdots \leq \varepsilon_{(k_j+1)} < \varepsilon_{k_j}$. By the decreasing property of the sequence $\{F_{\varepsilon_k}(v^k)\}$ and Assumption 2.1 (b) and (d), we have

$$\begin{aligned}
&F_{\varepsilon_{k_{(j+1)}}}(v^{k_{(j+1)}}) \leq F_{\varepsilon_{(k_j+1)}}(v^{k_j+1}) < F_{\varepsilon_{k_j}}(v^{k_j+1})\\
&= F_{\varepsilon_{k_j}}(v^{k_j}) + \nabla F_{\varepsilon_{k_j}}(v^{k_j})^T(v^{k_j+1}-v^{k_j}) + \frac{1}{2}(v^{k_j+1}-v^{k_j})^T\nabla^2 F_{\varepsilon_{k_j}}(\hat{v})(v^{k_j+1}-v^{k_j})\\
&\leq F_{\varepsilon_{k_j}}(v^{k_j}) + \nabla F_{\varepsilon_{k_j}}(v^{k_j})^T(v^{k_j+1}-v^{k_j}) - \frac{1}{2}C(\varepsilon_{k_j}, r)\|v^{k_j+1}-v^{k_j}\|_2^2,
\end{aligned}$$

where $\hat{v}$ is some point on the line segment between v^{k_j+1} and v^{k_j}, and hence $\|\hat{v}\| \leq r$. The last inequality above follows from Assumption 2.1 (d). Since $\nabla F_{\varepsilon_{k_j}}(v^{k_j})^T(v^{k_j+1} - v^{k_j}) \leq 0$, it follows from the above inequality that

$$\frac{1}{2}C(\varepsilon_{k_j}, r)\|v^{k_j+1}-v^{k_j}\|_2^2 \leq F_{\varepsilon_{k_j}}(v^{k_j}) - F_{\varepsilon_{k_{(j+1)}}}(v^{k_{(j+1)}}).$$

By Assumption 2.1 (d), $C(\varepsilon_{k_j}, r) > 0$ is bounded away from zero as $j \to \infty$. Therefore there exists a constant $\gamma^* > 0$ such that $C(\varepsilon_{k_j}, r) \geq \gamma^* > 0$ for all j. Thus the above inequality implies that

$$\lim_{j\to\infty} \|v^{k_j+1} - v^{k_j}\| = 0, \text{ and } \sum_{j=0}^{\infty} \|v^{k_j+1} - v^{k_j}\|_2^2 \leq \frac{F_{\varepsilon_0}(v^0) - F^*}{\gamma^*} =: C^*.$$

The proof is complete. ☐

From Examples 2.3 and 2.6, we see that a merit function in $\mathcal{M}$ is not necessarily bounded below as $|x_i| \to 0$. However, when a merit function is bounded below (such as Examples 2.4 and 2.5), we have the next result which claims that the result of Lemma 3.3 holds for the whole sequence generated by the algorithm, instead of a subsequence.

COROLLARY 3.4. *Let $F_\varepsilon \in \mathcal{M}$ and $F_\varepsilon(x)$ be bounded below in the region $(x, \varepsilon) \in R_+^n \times R_+$, and let $g(x) = \inf_{\varepsilon\downarrow 0} F_\varepsilon(x)$ be coercive in R_+^n. Let $\{(x^k, v^k)\}$ be generated by Algorithm 2.2. Then $\{(x^k, v^k)\}$ is bounded and there exists a finite constant C^* such that $\sum_{k=1}^{\infty} \|v^{k+1} - v^k\| \leq C^*$. In particular, $\|v^{k+1} - v^k\| \to 0$ as $k \to \infty$.*

Proof. By the assumption and Corollary 3.2, there exists a constant F^* such that $F^* \leq F_{\varepsilon_k}(v^k) \leq F_{\varepsilon_0}(v^0)$ for all k which, together with the coercivity of $g(x)$, implies that $\{v^k\}$ is bounded. Applying the rest proof of Lemma 3.3 to $\{v^k\}$ instead of $\{v^{k_j}\}$, we can show that $\|v^{k+1} - v^k\| \to 0$ and $\sum_{k=1}^{\infty} \|v^{k+1} - v^k\| \leq C^*$ for some C^*. ☐

To further investigate the properties of Algorithm 2.2, we need some condition on A. We introduce the *range space property* of A^T, based on which we can establish some convergence results for Algorithm 2.2.

DEFINITION 3.5. *(Range Space Property (RSP)) Let A be an $m \times n$ matrix with $m \leq n$. Then A^T is said to satisfy the range space property of order K with a constant $\rho > 0$ if*

$$\|\xi_{S_c}\|_1 \leq \rho \|\xi_S\|_1$$

for all sets $S \subseteq \{1, ..., n\}$ with $|S| \geq K$, and for all $\xi \in \mathcal{R}(A^T)$, the range space of A^T.

Clearly, RSP can be equivalently stated in some other ways. First, note that the inequality $|S| \geq K$ in the above definition can be replaced by the equality $|S| = K$. In fact, it is easy to see that if $\|\xi_{S_c}\|_1 \leq \rho\|\xi_S\|_1$ holds for any $S \subseteq \{1, ..., n\}$ with $|S| = K$, then it holds for all S with $|S| > K$. Second, it is evident that the RSP can be restated as follows: *A^T is said to satisfy the range space property of order K with constant $\rho > 0$ if $\|\xi_S\|_1 \leq \rho\|\xi_{S_c}\|_1$ for any set $S \subseteq \{1, ..., n\}$ with $|S| \leq n - K$, and for all $\xi \in \mathcal{R}(A^T)$.* It is interesting to understand the relationship between this property and the restricted isometry property (RIP) and null space property (NSP) of A which have been widely used in the compressed sensing literature. Recall that A has the RIP of order k if there exists a constant $\delta \in (0, 1)$ such that $(1-\delta)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1+\delta)\|z\|_2^2$ for any k-sparse vector $z \in R^n$, and that A has the NSP of order k if there exists a constant $\tau \in (0, 1)$ such that $\|\eta_S\|_1 \leq \tau\|\eta_{S_c}\|_1$ for all $S \subseteq \{1, ..., n\}$ with $|S| \leq k$, and any $\eta \in \mathcal{N}(A)$, the null space of A. In this paper, we do not make use of RIP and NSP in our analysis. The following proposition and remark shed some light on the relationship between RIP, NSP, and RSP.

PROPOSITION 3.6. *Let $m < n$, and let $A \in R^{m\times n}$ and $M \in R^{(n-m)\times n}$ be full-rank matrices satisfying $AM^T = 0$. Then the following hold.*

(i) M has the NSP of order k with constant $\tau \in (0, 1)$ if and only if A^T has the RSP of order $(n - k)$ with the same constant $\rho = \tau$.

(ii) If M has the RIP of order k with constant $\delta \in (0,1)$, then A^T has the RSP of order $\left(n - \lfloor \frac{\varrho k}{1+\varrho} \rfloor\right)$ with the constant $\rho = \left(\frac{\lfloor \frac{\varrho k}{1+\varrho} \rfloor}{k - \lfloor \frac{\varrho k}{1+\varrho} \rfloor}\right)^{1/2} \left(\frac{1}{\varrho}\right) < 1$ where $\varrho = (1-\delta)/(1+\delta)$.

Proof. (i) Since A and M are full-rank matrices and $AM^T = 0$ (and thus $MA^T = 0$), the columns of A^T comprise the basis of the null space of M. Thus $\mathcal{R}(A^T) = \mathcal{N}(M)$. Suppose that M has the NSP of order k with constant $\tau \in (0,1)$. Then $\|\eta_L\|_1 \leq \tau \|\eta_{L_c}\|_1$ for any $\eta \in \mathcal{N}(M)$ and for any set $L \subseteq \{1, ..., n\}$ with $|L| \leq k$. By setting $S = L_c$, this is equivalent to saying that $\|\eta_{S_c}\|_1 \leq \tau \|\eta_S\|_1$ for any $\eta \in \mathcal{R}(A^T)$ and for any set $S \subseteq \{1, ..., n\}$ with $|S| \geq n - k$. Thus A has the RSP of order $(n-k)$ with constant $\rho = \tau$. Clearly, the converse can be shown in a similar way.

(ii) Suppose that M has the RIP of order k with constant $\delta \in (0,1)$. Denote by $\varrho = (1-\delta)/(1+\delta)$. Let $\nu = \lfloor \frac{\varrho k}{1+\varrho} \rfloor$, which is smaller than k. It was shown in [13] that M has the NSP of order ν with constant $\tau = \sqrt{\frac{\nu}{k-\nu}} \left(\frac{1}{\varrho}\right)$, which is less than 1 by the definition of ν. By (i), this in turn implies that A^T has the RSP of order $n - \nu = n - \lfloor \frac{\varrho k}{1+\varrho} \rfloor$ with constant $\rho = \tau$. □

From Proposition 3.6, if $M \in R^{(n-m)\times n}$ has the NSP (or more restrictively, the RIP), we may construct a matrix satisfying the range space property by simply choosing $A = [v_1, ..., v_m]^T$ where $v_1, ..., v_m \in R^n$ are the basis of the null space of M.

Remark. One might be also interested in the direct relationship between the NSP of A and RSP of A^T, which remains not quite clear at this stage. However, such a direct relationship might exist as shown by the following observation. Note that for any $\xi \in \mathcal{R}(A^T)$ and $\eta \in \mathcal{N}(A)$, $\xi^T \eta = 0$ implies that for any index set $S \subset \{1, ..., n\}$ we have $(\xi_S)^T \eta_S + (\xi_{S_c})^T \eta_{S_c} = 0$, and hence

$$\|\xi_S\|_2 \|\eta_S\|_2 |\cos(\theta)| = \|\xi_{S_c}\|_2 \|\eta_{S_c}\|_2 |\cos(\theta')|, \tag{3.5}$$

where θ is the angle between ξ_S and η_S, and θ' between ξ_{S_c} and η_{S_c}. Assume that there exists a k such that for any $S \subset \{1, ..., n\}$ with $|S| \leq k$, it holds that $|\cos(\theta')| \geq \gamma > 0$ (where γ is a constant) for any $\eta \in \mathcal{N}(A)$ and $\xi \in \mathcal{R}(A^T)$ with $\eta_{S_c} \neq 0$ and $\xi_{S_c} \neq 0$. Under this assumption, we see that when $\eta_{S_c} \neq 0$ and $\xi_{S_c} \neq 0$, all other terms in (3.5) are nonzero. In this case, (3.5) can be written as

$$\|\xi_S\|_2 / \|\xi_{S_c}\|_2 = (\|\eta_{S_c}\|_2 / \|\eta_S\|_2) \left(|\cos(\theta')| / |\cos(\theta)|\right).$$

By the equivalence of $\|\cdot\|_1$ and $\|\cdot\|_2$, it is not difficult to see that there exists a constant $\vartheta > 0$ (dependent on n and k) such that

$$\|\xi_S\|_1 / \|\xi_{S_c}\|_1 \geq \vartheta \left(\|\eta_{S_c}\|_1 / \|\eta_S\|_1\right) \left(|\cos(\theta')| / |\cos(\theta)|\right).$$

Now, if A has the NSP of order k, then there exists a constant $\tau \in (0,1)$ such that $\|\eta_{S_c}\|_1 / \|\eta_S\|_1 \geq 1/\tau$, and thus $\|\xi_S\|_1 / \|\xi_{S_c}\|_1 \geq (\vartheta/\tau)(|\cos(\theta')|/|\cos(\theta)|)$, i.e.,

$$\|\xi_{S_c}\|_1 \leq \left[\frac{\tau}{\vartheta}(|\cos(\theta)|/|\cos(\theta')|)\right] \|\xi_S\|_1 \leq \frac{\tau}{\gamma\vartheta} \|\xi_S\|_1,$$

which holds for all $S \subset \{1, ..., n\}$ with $|S| \leq k$. In particular, it holds for all subset S with $|S| = 1$. This implies that A^T has the RSP of order 1, and thus A^T has the RSP of any order $k \geq 1$, which is the strongest RSP. Note that this conclusion is restrictive, since it is drawn from the strong assumption on the angle θ' as above. So we believe that this observation has not given a full picture of the true relationship between RSP and NSP yet.

We now study the properties of reweighted ℓ_1-algorithms under the RSP assumption. The first result below shows that if the RSP is satisfied, Algorithm 2.2 with $F_\varepsilon \in \mathcal{M}$ generates a sparse solution from any initial point in the sense that at least one component of x^k tends to zero.

THEOREM 3.7. *Let $A \in R^{m\times n}$ with $m \leq n$, and assume that A^T has the RSP of order K with constant $\rho > 0$ satisfying $1+\rho < \frac{n}{K}$. Let $F_\varepsilon \in \mathcal{M}$ and $\{(x^k, v^k)\}$ be generated by Algorithm 2.2. Then*

$$[\sigma(x^k)]_n = \min_{1\leq i\leq n} |x_i^k| \to 0 \text{ as } k \to \infty. \tag{3.6}$$

Proof. We now assume the contrary that (3.6) does not hold. Then there exist a constant $\gamma > 0$ and a subsequence, denoted by $\{(x^{k_j}, v^{k_j})\}_{j=1}^{\infty}$, such that $[\sigma(x^{k_j})]_n \geq \gamma > 0$ for all j. Thus $v^{k_j} (= |x^{k_j}|) \geq \gamma e$ for all j. By Lemma 3.3, we have $\|v^{k_j+1} - v^{k_j}\| \to 0$ as $j \to \infty$. So there exists a j' such that for all $j \geq j'$ the vector v^{k_j+1} is positive, i.e., $v^{k_j+1} \in R^n_{++}$. Let $j \geq j'$ and consider the k_j-th step of the algorithm. Note that (x^{k_j+1}, v^{k_j+1}) is an optimal solution to the LP $\min\{\nabla F_{\varepsilon_{k_j}}(v^{k_j})^T v : Ax = b, |x| \leq v\}$, which can be written as

$$\min\left\{\nabla F_{\varepsilon_{k_j}}(v^{k_j})^T v : \ Ax = b, \ x_i \leq v_i, \ -x_i \leq v_i, \ i = 1, ..., n\right\}.$$

By the optimality condition, there exist $\alpha^{k_j}, \beta^{k_j} \in R^n_+$ and $\lambda^{k_j} \in R^m$ such that

$$A^T\lambda^{k_j} - \alpha^{k_j} + \beta^{k_j} = 0, \tag{3.7}$$

$$\nabla F_{\varepsilon_{k_j}}(v^{k_j}) - \alpha^{k_j} - \beta^{k_j} = 0, \tag{3.8}$$

$$-x_i^{k_j+1} - v_i^{k_j+1} \leq 0, \ \alpha_i^{k_j}\left(-x_i^{k_j+1} - v_i^{k_j+1}\right) = 0, \ \alpha_i^{k_j} \geq 0, \ i = 1, ..., n, \tag{3.9}$$

$$x_i^{k_j+1} - v_i^{k_j+1} \leq 0, \ \beta_i^{k_j}\left(x_i^{k_j+1} - v_i^{k_j+1}\right) = 0, \ \beta_i^{k_j} \geq 0, \ i = 1, ..., n, \tag{3.10}$$

$$Ax^{k_j+1} = b.$$

Since $v^{k_j+1} \in R^n_{++}$, for every i one of the inequalities $x_i^{k_j+1} - v_i^{k_j+1} \leq 0$ and $-x_i^{k_j+1} - v_i^{k_j+1} \leq 0$ holds strictly. Thus by the complementarity conditions (3.9) and (3.10), we see that for every i, either $\alpha_i^{k_j}$ or $\beta_i^{k_j}$ must be zero. On the other hand, since $v^{k_j} \in R^n_{++}$, $\varepsilon_{k_j} > 0$ and $F_{\varepsilon_{k_j}} \in \mathcal{M}$, it implies that $\nabla F_{\varepsilon_{k_j}}(v^{k_j}) \in R^n_{++}$ (Assumption 2.1(c)). Thus it follows from (3.8) that for every i, $\alpha_i^{k_j}$ and $\beta_i^{k_j}$ cannot vanish at the same time. So we conclude that for every i, one and only one of $\alpha_i^{k_j}$ and $\beta_i^{k_j}$ is equal to zero, and hence $\alpha_i^{k_j} + \beta_i^{k_j} = |\alpha_i^{k_j} - \beta_i^{k_j}|$ which, together with (3.7) and (3.8), implies that

$$\nabla F_{\varepsilon_{k_j}}(v^{k_j}) = \alpha^{k_j} + \beta^{k_j} = |\alpha^{k_j} - \beta^{k_j}| = |A^T\lambda^{k_j}| = |\xi^{k_j}|,$$

where $\xi^{k_j} = A^T\lambda^{k_j} \in \mathcal{R}(A^T)$. Thus $\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j})) = \sigma(\xi^{k_j})$. Since A^T has the RSP of order K with $\rho > 0$, we have

$$\left\|\xi_{S_c}^{k_j}\right\|_1 \leq \rho \left\|\xi_S^{k_j}\right\|_1, \tag{3.11}$$

for any $S \subseteq \{1, ..., n\}$ with $|S| \geq K$. In particular, we consider the index set S which is the set of indices i such that $|\xi_i^{k_j}| = \left[\sigma(\xi^{k_j})\right]_j$ where $n-K+1 \leq j \leq n$, i.e., S denotes the set of indices corresponding to the K smallest components of $|\xi^{k_j}|$. Thus,

$$\left\|\xi_{S_c}^{k_j}\right\|_1 = \sum_{i=1}^{n-K} \left[\sigma(\xi^{k_j})\right]_i = \sum_{i=1}^{n-K} \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_i,$$

and

$$\left\|\xi_S^{k_j}\right\|_1 = \sum_{i=n-K+1}^{n} [\sigma(\xi^{k_j})]_i = \sum_{i=n-K+1}^{n} \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_i.$$

Therefore, by (3.11), we have

$$\sum_{i=1}^{n-K} \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_i \leq \rho \sum_{i=n-K+1}^{n} \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_i. \tag{3.12}$$

However, we have

$$\begin{aligned}\sum_{i=1}^{n-K} \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_i &\geq (n-K) \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_{n-K+1} \\ &> \rho K \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_{n-K+1}, \end{aligned} \tag{3.13}$$

where the second inequality follows from $n-K > \rho K$ and $\left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_{n-K+1} > 0$ (since $\nabla F_{\varepsilon_{k_j}}(v^{k_j}) \in R^n_{++}$). The right-hand side of (3.13) is greater than or equal to

$$\rho \sum_{i=n-K+1}^{n} \left[\sigma(\nabla F_{\varepsilon_{k_j}}(v^{k_j}))\right]_i,$$

contradicting with (3.12). Thus (3.6) holds. □

The above result holds for the whole family of merit functions for sparsity in $\mathcal{M}$. Roughly speaking, Algorithm 2.2 guarantees to generate a sparse solution if the RSP is satisfied. However, this result does not confirm how sparse the solution is. So Theorem 3.7 can be called a weak convergence theorem, based on which some stronger convergence result can be further proved. To this end, we need to establish the next technical result. Recall that for a given vector x, $\mathcal{T}(x) = \{i : x_i \neq 0\}$ denotes the support of x.

LEMMA 3.8. *Let $A \in R^{m\times n}$ with $m \leq n$, and assume that A^T has the RSP of order K with constant $\rho > 0$ satisfying $1 + \rho < \frac{n}{K}$. Let $F_\varepsilon \in \mathcal{M}$ and $\{(x^k, v^k)\}$ be generated by Algorithm 2.2 such that $\|v^{k+1} - v^k\| \to 0$ as $k \to \infty$. If there exists a positive constant $\mu > 0$ such that $|I_\mu(x^k)| \geq K$ for all sufficiently large k, where $I_\mu(x^k) = \{i : |x_i^k| \geq \mu\}$, then there exists a k' such that $\|x^{k'}\|_0 < n$ and $\mathcal{T}(x^k) \subseteq \mathcal{T}(x^{k'})$ for all $k \geq k'$.*

Proof. At the k-th step, since (x^{k+1}, v^{k+1}) is an optimal solution to the LP problem $\min\{\nabla F_{\varepsilon_k}(v^k)^T v : Ax = b, |x| \leq v\}$, by the optimality condition there exist

$\alpha^k, \beta^k \in R^n_+$ and $\lambda^k \in R^m$ such that

$$A^T\lambda^k - \alpha^k + \beta^k = 0, \tag{3.14}$$

$$\nabla F_{\varepsilon_k}(v^k) - \alpha^k - \beta^k = 0, \tag{3.15}$$

$$-x_i^{k+1} - v_i^{k+1} \leq 0,\ \alpha_i^k\left(-x_i^{k+1} - v_i^{k+1}\right) = 0,\ \alpha_i^k \geq 0,\ i = 1, ..., n, \tag{3.16}$$

$$x_i^{k+1} - v_i^{k+1} \leq 0,\ \beta_i^k\left(x_i^{k+1)} - v_i^{k+1}\right) = 0,\ \beta_i^k \geq 0,\ i = 1, ..., n, \tag{3.17}$$

$$Ax^{k+1} = b. \tag{3.18}$$

Since $F_\varepsilon(v)$ is separable in v. Each component $[\nabla F_{\varepsilon_k}(v^k)]_i$ is dependent on v_i^k only. Since the second order derivative $[\nabla^2 F_{\varepsilon_k}(v^k)]_{ii}$ is negative (by Assumption 2.1(d)), $[\nabla F_{\varepsilon_k}(v^k)]_i$ is decreasing with respect to v_i^k. This implies that for all $i \in I_\mu(x^k)$ we have

$$[\nabla F_{\varepsilon_k}(v^k)]_i \leq [\nabla F_{\varepsilon_k}(\mu e)]_i. \tag{3.19}$$

Since $\|v^{k+1} - v^k\| \to 0$ as $k \to \infty$, for all sufficiently large k we have $v_i^{k+1} \neq 0$ for every $i \in I_\mu(x^k)$. Thus, for every $i \in I_\mu(x^k)$, (3.16) and (3.17) imply that either $\alpha_i^k = 0$ or $\beta_i^k = 0$. By Assumption 2.1(c), we have $\nabla F_{\varepsilon_k}(v^k) \in R^n_{++}$ which, together with (3.15), implies that for every i we have either $\alpha_i^k \neq 0$ or $\beta_i^k \neq 0$. So we conclude that for every $i \in I_\mu(x^k)$, α_i^k and β_i^k must be strictly complementary, i.e., one and only one of them is zero. Thus

$$|\alpha_i^k - \beta_i^k| = \alpha_i^k + \beta_i^k = \left[\nabla F_{\varepsilon_k}(v^k)\right]_i \quad \text{for every } i \in I_\mu(x^k). \tag{3.20}$$

By (3.19) and (3.20), for all sufficiently large k we have

$$\sum_{i \in I_\mu(x^k)} |\alpha_i^k - \beta_i^k| = \sum_{i \in I_\mu(x^k)} \alpha_i^k + \beta_i^k = \sum_{i \in I_\mu(x^k)} \left[\nabla F_{\varepsilon_k}(v^k)\right]_i \leq \sum_{i \in I_\mu(x^k)} \left[\nabla F_{\varepsilon_k}(\mu e)\right]_i. \tag{3.21}$$

The right-hand side of the above is bounded. In fact, by Assumption 2.1(c), $[\nabla F_\varepsilon(\mu e)]_i$ is continuous in ε and there exists constant $\gamma_i^* > 0$ such that $[\nabla F_\varepsilon(\mu e)]_i \to \gamma_i^*$ as $\varepsilon \to 0$. Thus there exists a constant W^* such that $\sum_{i=1}^n [\nabla F_\varepsilon(\mu e)]_i \leq W^*$ for all $\varepsilon \in (0, \varepsilon^0]$. Note that $\varepsilon_k \in (0, \varepsilon^0]$. It follows from (3.21) that for all sufficiently large k we have

$$\sum_{i \in I_\mu(x^k)} |\alpha_i^k - \beta_i^k| \leq \sum_{i \in I_\mu(x^k)} [\nabla F_{\varepsilon_k}(\mu e)]_i \leq \sum_{i=1}^n [\nabla F_{\varepsilon_k}(\mu e)]_i \leq W^*. \tag{3.22}$$

We now prove that $\|x^k\|_0 = |\mathcal{T}(x^k)| < n$ for all sufficiently large k. From Assumption 2.1(c), $[\nabla F_\varepsilon(v)]_i \to \infty$ as $(v_i, \varepsilon) \to 0$. Thus there exists a small constant $0 < \epsilon^* < \mu$ such that

$$[\nabla F_\varepsilon(v)]_i > \rho W^* \text{ for any } |v_i| + \varepsilon \leq \epsilon^*. \tag{3.23}$$

By Theorem 3.7, we have that $[\sigma(x^k)]_n \to 0$ as $k \to \infty$. Thus there exists a sufficiently large number k' such that $[\sigma(x^k)]_n + \varepsilon_k < \epsilon^*$ for all $k \geq k'$. Let i_0 be the index such that $v_{i_0}^{k'} = |x_{i_0}^{k'}| = [\sigma(x^{k'})]_n$. Since $\epsilon^* < \mu$, we see that $i_0 \notin I_\mu(x^{k'})$. Consider the next point $(x^{k'+1}, v^{k'+1})$, which satisfies the optimality condition (3.14)-(3.18) where $(\lambda^k, v^k, x^{k+1}, v^{k+1}, \alpha^k, \beta^k, \varepsilon_k)$ is replaced by $(\lambda^{k'}, v^{k'}, x^{k'+1}, v^{k'+1}, \alpha^{k'}, \beta^{k'}, \varepsilon_{k'})$. We now prove that $x_{i_0}^{k'+1} = 0$. In fact, if it is not, by (3.15)-(3.17), one and only one of $\alpha_{i_0}^{k'}$ and $\beta_{i_0}^{k'}$ is zero. This, together with (3.15) and (3.23), implies that

$$|\alpha_{i_0}^{k'} - \beta_{i_0}^{k'}| = \alpha_{i_0}^{k'} + \beta_{i_0}^{k'} = \left[\nabla F_{\varepsilon_{k'}}(v^{k'})\right]_{i_0} > \rho W^*,$$

and thus

$$\sum_{i \notin I_\mu(x^{k'})} |\alpha_i^{k'} - \beta_i^{k'}| \geq |\alpha_{i_0}^{k'} - \beta_{i_0}^{k'}| > \rho W^*. \tag{3.24}$$

However, since $|I_\mu(x^{k'})| \geq K$ and A^T has the RSP of order K with constant $\rho > 0$, by (3.14) and (3.22), we have

$$\begin{aligned} \sum_{i \notin I_\mu(x^{k'})} |\alpha_i^{k'} - \beta_i^{k'}| &= \sum_{i \notin I_\mu(x^{k'})} |A^T \lambda^{k'}|_i \\ &\leq \rho \sum_{i \in I_\mu(x^{k'})} |A^T \lambda^{k'}|_i = \rho \sum_{i \in I_\mu(x^{k'})} |\alpha_i^{k'} - \beta_i^{k'}| \\ &\leq \rho W^*. \end{aligned} \tag{3.25}$$

This contradicts (3.24). So we conclude that $x_{i_0}^{k'+1} = 0$ and thus $|\mathcal{T}(x^{k'+1})| < n$. This also indicates that $[\sigma(x^{k'+1})]_n = v_{i_0}^{k'+1} = 0$. Replacing $x^{k'}$ by $x^{k'+1}$, considering the point $(x^{k'+2}, v^{k'+2})$ and repeating the same proof above, we can show that $x_{i_0}^{k'+2} = 0$. Thus, by induction, we conclude that for all $k > k'$ the iterates will keep this component being zero. This proof can be applied to any other component $x_i^{k'} = 0$ from which we can show that $x_i^k = 0$ for all $k > k'$. Thus there exists a k' such that $\mathcal{T}(x^k) \subseteq \mathcal{T}(x^{k'})$ for all $k \geq k'$. ∎

The requirement that $\|v^{k+1} - v^k\| \to 0$ as $k \to \infty$ used in Lemma 3.8 is mild, and it can hold trivially when the merit function is suitably chosen. For instance, by Corollary 3.4, the merit functions in Examples 2.4 and 2.5 can ensure this condition.

In what follows, we prove the next main result in this section. Let $I_\mu(x)$ be defined as in Lemma 3.8. Since $b \neq 0$, there exists a small number $\mu_0 > 0$ such that for any given $\mu \in (0, \mu_0)$ the set $I_\mu(x) \neq \emptyset$ for any solution of the system $Ax = b$. Clearly, we have $I_\mu(x) \subseteq \mathcal{T}(x)$ (and thus $|I_\mu(x)| \leq |\mathcal{T}(x)|$) for any $x \in R^n$. The next result is stronger than Theorem 3.7.

THEOREM 3.9. *Let $A \in R^{m\times n}$ with $m \leq n$. Assume that A^T has the RSP of order K with constant $\rho > 0$ satisfying $1 + \rho < \frac{n}{K}$. Let $F_\varepsilon \in \mathcal{M}$ and the sequence $\{(x^k, v^k)\}$ be generated by Algorithm 2.2. If $\|v^{k+1} - v^k\| \to 0$ as $k \to \infty$, then there is a subsequence $\{x^{k_j}\}$ that converges to a $\lfloor (1+\rho)K \rfloor$-sparse solution of $Ax = b$ in the sense that $[\sigma(x^{k_j})]_{\lfloor (1+\rho)K+1 \rfloor} \to 0$ as $j \to \infty$.*

Proof. Consider the sequence $\{(x^k, v^k)\}$ generated by Algorithm 2.2. Under the condition of the theorem, we prove that it has a subsequence $\{x^{k_j}\}$ convergent to a $\lfloor (1+\rho)K \rfloor$-sparse solution in the sense that $[\sigma(x^{k_j})]_{\lfloor (1+\rho)K+1 \rfloor} \to 0$ as $j \to \infty$. We prove this by contradiction. Assume the contrary that there is no such subsequence. Then $[\sigma(x^k)]_{\lfloor (1+\rho)K+1 \rfloor}$ must be bounded away from zero, i.e., there exists a number $\mu^* > 0$ such that $[\sigma(x^k)]_{\lfloor (1+\rho)K+1 \rfloor} \geq \mu^*$ for all sufficiently large k. In other words

$$|I_{\mu^*}(x^k)| \geq \lfloor (1+\rho)K + 1 \rfloor > (1+\rho)K$$

for all large k. By Lemma 3.8, there exists a k' such that

$$\|x^k\|_0 = |\mathcal{T}(x^k)| < n,\ \mathcal{T}(x^k) \subseteq \mathcal{T}(x^{k'}) \text{ for all } k \geq k'. \tag{3.26}$$

Note that $I_{\mu^*}(x^k) \subseteq \mathcal{T}(x^k)$. So

$$n > |\mathcal{T}(x^k)| \geq |I_{\mu^*}(x^k)| > (1+\rho)K > K \text{ for any } k \geq k'. \tag{3.27}$$

If $|\mathcal{T}(x^k)| > |I_{\mu^*}(x^k)|$ at the k-step ($k \geq k'$), we can prove that the algorithm will continue to reduce the value of $|\mathcal{T}(x^k)|$ until for some $k'' > k'$ we have $|\mathcal{T}(x^k)| = |I_{\mu^*}(x^k)|$ for all $k \geq k''$. Since $F_\varepsilon(v)$ is separable in v, it can be represented as $F_\varepsilon(v) = \sum_{i=1}^n \phi_i(\varepsilon, v_i)$ where ϕ_i's are some kernel functions. From (3.26), we see that if $x_i^{k'} = 0$ then $x_i^k = 0$ for all $k \geq k'$. Thus for all $k \geq k'$ the problem (2.5), i.e., $\min\left\{\nabla F_{\varepsilon_k}(v^k)^T v : Ax = b, |x| \leq v\right\}$, is exactly equivalent to the reduced problem

$$\min\left\{[\nabla F_{\varepsilon_k}(v^k)]_S^T v_S :\ A_S x_S = b,\ |x_S| \leq v_S\right\}, \tag{3.28}$$

where $S = \mathcal{T}(x^{k'})$. In other words, for all $k \geq k'$ the solution (x^{k+1}, v^{k+1}) to (2.5) can be partition into $x^{k+1} = (x_S^{k+1}, 0), v^{k+1} = (v_S^{k+1}, 0)$ where (x_S^{k+1}, v_S^{k+1}) is the solution to the reduced problem (3.28). The merit function for sparsity associated with (3.28) is given by

$$F_\varepsilon(v_S) := \sum_{i\in S} \phi_i(\varepsilon, v_i),$$

which still satisfies the Assumption 2.1 in space $R^{|S|}$, where $|S| = |\mathcal{T}(x^{k'})| > K$ by (3.27). (The reduced function above is obtained from $F_\varepsilon(v)$ by simply dropping the components $\phi_i(\varepsilon, v_i)$ with $i \in S_c = \{1, ..., n\}\backslash S$, and retaining the ones with indices in S only.) We now show that the reduced matrix A_S^T has the same RSP as that of A^T. Indeed, let $\eta \in \mathcal{R}(A_S^T)$. Then there exists a λ such that $\eta = A_S^T\lambda$. Setting $\eta' = A_{S_c}^T\lambda$ and rearranging the components of (η, η') if necessary, we have $(\eta, \eta') \in \mathcal{R}(A^T)$. For any $L \subseteq S$ with $|L| \geq K$, by the RSP of A^T, we have

$$\|(\eta_{L_c}, \eta')\|_1 \leq \rho\|\eta_L\|,$$

where $L_c = S\backslash L$. Thus $\|\eta_{L_c}\|_1 \leq \rho\|\eta_L\|$, which implies that A_S^T satisfies the RSP of order K with the same constant ρ. Similarly, it is evident that removing rows from the matrix A, the transpose of the resulting submatrix still satisfies the RSP with the same constant ρ as that of A^T. Note that $A_S \in R^{m\times|S|}$. If $m > |S|$, the rows of A_S are linearly dependent, and hence some equations of $A_S x_S = b$ are redundant, and can be removed from the system without any change to the solution of (3.28).

Therefore without loss of generality, the size of A_S can be assume to be $m \times |S|$ with $m \leq |S|$, and A_S^T has the RSP of order K with the same constant as that of A^T, as shown above. By (3.27), we have that $1 + \rho < \frac{|S|}{K}$ where $|S| = |\mathcal{T}(x^{k'})|$. Therefore, applying Lemma 3.8 to the $m \times |S|$ matrix A_S and the reduced merit function for sparsity $F_\varepsilon(v_S) = \sum_{i\in S} \phi_i(\varepsilon, v_i)$ where $v_S \in R^{|S|}$, we conclude that there exists a $k'' > k'$ such that $|\mathcal{T}(x_S^{k''})| < |S|,\ \mathcal{T}(x_S^k) \subseteq \mathcal{T}(x_S^{k''})$ for all $k \geq k''$. Notice that $x^{k''}$ and x^k are partitioned, respectively, into $(x_S^{k''}, 0)$ and $(x_S^k, 0)$ for all $k \geq k''$. This is equivalent to

$$|\mathcal{T}(x^{k''})| < |S| = |\mathcal{T}(x^{k'})|,\ \mathcal{T}(x^k) \subseteq \mathcal{T}(x^{k''}) \text{ for all } k \geq k''.$$

If $|\mathcal{T}(x^{k''})|$ remains larger than $|I_{\mu^*}(x^{k''})|$, which is larger than $\lfloor(1+\rho)K\rfloor$ by (3.27), then replace $x^{k'}$ by $x^{k''}$ and repeat the same proof above, we can conclude that there exists $k''' > k''$ such that $|\mathcal{T}(x^{k'''})|$ is strictly smaller than $|\mathcal{T}(x^{k''})|$. Therefore, by induction, there must exist an integer number, denoted still by k''', such that

$$\mathcal{T}(x^k) = I_{\mu^*}(x^k),\ \mathcal{T}(x^k) \subseteq \mathcal{T}(x^{k'''}) \text{ for all } k \geq k'''.$$

Let $S = I_{\mu^*}(x^k)$, which is larger than $(1+\rho)K$ by (3.27). The above relation implies that for all $k \geq k'''$ the vector x^k is $|S|$-sparse vector where $|S| = |I_{\mu^*}(x^k)| > (1+\rho)K$, and all nonzero components of x^k are bounded below by $\mu^* > 0$. All the rest iterations are equivalent to solving the reduced minimization problem (3.28) with $S = I_{\mu^*}(x^k) = \mathcal{T}(x^k)$. Note that A_S is a submatrix of A, so A_S satisfies the RSP with the same order and constant. Thus applying to the reduced merit function $F_\varepsilon(x_S)$, Theorem 3.7 implies that $[\sigma(x_S^k)]_{|S|} \to 0$ as $k \to \infty$, i.e., the smallest component of $x_{|S|}^k$ tends to zero, which contradicts with $x_{|S|}^k \geq \mu^* e > 0$. This contradiction shows that there must exist a subsequence $\{x^{k_j}\}$ convergent to a $\lfloor(1+\rho)K\rfloor$-sparse solution in the sense that $[\sigma(x^{k_j})]_{\lfloor(1+\rho)K+1\rfloor} \to 0$ as $j \to \infty$. □

An immediate result is given as follows.

COROLLARY 3.10. *Under the same condition of Theorem 3.7. Let $\{(x^k, v^k)\}$ be the sequence generated by Algorithm 2.2, and let $\{(x^{k_j}, v^{k_j})\}$ be the subsequence such that $[\sigma(x^{k_j})]_{\lfloor(1+\rho)K+1\rfloor} \to 0$ as $j \to \infty$. Then the following hold.*

(i) For any given integer number $t \geq 1$, the subsequence $\{x^{k_j+t}\}$ converges also to $\lfloor(1+\rho)K\rfloor$-sparse solution in the sense that $[\sigma(x^{k_j+t})]_{\lfloor(1+\rho)K+1\rfloor} \to 0$ as $j \to \infty$.

(ii) If $v^k \to v^*$, then any accumulation point of $\{x^k\}$ is a $\lfloor(1+\rho)K\rfloor$-sparse solution of $Ax = b$.

(iii) If the $\lfloor(1+\rho)K\rfloor$-sparse solution of $Ax = b$ is unique, then the subsequence $\{x^{k_j}\}$ converges to the sparsest solution of $Ax = b$. In particular, if $x^k \to x^*$ or $v^k \to v^*$, then x^k converges to the sparsest solution.

Proof. Clearly, we have

$$\begin{aligned}\left|[\sigma(x^{k_j+t})]_{\lfloor(1+\rho)K+1\rfloor} - [\sigma(x^{k_j})]_{\lfloor(1+\rho)K+1\rfloor}\right| &\leq \left\|\sigma(x^{k_j+t}) - \sigma(x^{k_j})\right\| \\ &= \left\|\sigma(v^{k_j+t}) - \sigma(v^{k_j})\right\| \leq \left\|v^{k_j+t} - v^{k_j}\right\|.\end{aligned}$$

Note that $\left\|v^{k_j+t} - v^{k_j}\right\| \leq \sum_{i=1}^t \left\|v^{k_j+i} - v^{k_j+(i-1)}\right\| \to 0$ as $j \to \infty$, which follows from $\|v^{k+1} - v^k\| \to 0$ as $k \to \infty$. Combining the two relations above leads to the result (i). The results (ii) and (iii) are evident. □

As shown by Corollary 3.4, the merit function in $\mathcal{M}$ can be chosen to ensure that the sequence $\{v^k\}$ is bounded and $\|v^{k+1} - v^k\| \to 0$. So the requirement $\|v^{k+1} - v^k\| \to 0$ used in Theorem 3.9 and Corollary 3.10 can be removed when the merit functions are suitably chosen (e.g., Examples 2.4 and 2.5). We summarize this result as follows.

THEOREM 3.11. *Assume that A^T has the RSP of order K with constant $\rho > 0$ satisfying that $1+\rho < \frac{n}{K}$. Let $F_\varepsilon \in \mathcal{M}$ and $F_\varepsilon(v)$ be bounded below in $(x,v) \in R^n_+ \times R_+$, and $g(x) = \inf_{\varepsilon\downarrow 0} F_\varepsilon(x)$ be coercive in R^n_+. Let $\{(x^k, v^k)\}$ be generated by Algorithm 2.2. Then there is a subsequence $\{x^{k_j}\}$ that converges to a $\lfloor(1+\rho)K\rfloor$-sparse solution of $Ax = b$ in the sense that $[\sigma(x^{k_j})]_{\lfloor(1+\rho)K+1\rfloor} \to 0$ as $j \to \infty$.*

Remark. Except for Corollary 3.4 and Theorem 3.11, all other results established in this section can be viewed as the common properties shared among the reweighted ℓ_1-algorithms based on the merit function in $\mathcal{M}$. Theorem 3.7 claims that under RSP, any reweighted ℓ_1-minimization algorithm associated with a merit function in $\mathcal{M}$ can find a sparse solution of the underdetermined linear system. If the sequence generated by the algorithm satisfies $\|v^{k+1} - v^k\| \to 0$, then Theorem 3.9 further claims that, under RSP of order K with $(1+\rho)K < n$, the algorithm can find at least a $\lfloor (1+\rho)K \rfloor$-sparse solution. Since the CWB method falls into the framework of Algorithm 2.2 and it is based on a merit function in $\mathcal{M}$ (see Example 2.3), a convergence result for the CWB method can be extracted from Theorem 3.7 and 3.9 and their corollaries immediately. The statement of this special result is omitted here. Theorem 3.11 that is stronger than Theorem 3.9 has further identified a subclass of merit functions in $\mathcal{M}$, including Wl_p, NW2, and NW3 methods, which can ensure that the generated sequence is bounded and satisfies $\|v^{k+1} - v^k\| \to 0$. Thus, the convergence results for Wl_p, NW2, and NW3 methods can be immediately obtained from Theorem 3.11 as well.

4. Numerical Experiments. As seen in section 3, $\mathcal{M}$ is a large family of merit functions for sparsity, based on which various reweighted ℓ_1-methods can be constructed. It is interesting to compare these algorithms through numerical experiments. Since it is impossible to test all algorithms of this family, we single out a few of them, and compare their performances in finding sparse solution of underdetermined linear systems. Let us first list a few of these specific methods as follows.

(a) Candès-Wakin-Boyd (**CWB**) method

$$x^{k+1} = \arg\ \min \left\{ \sum_{i=1}^{n} \left(\frac{1}{|x_i^k| + \varepsilon_k} \right) |x_i| : Ax = b \right\}.$$

(b) $\mathbf{Wl}_p$ method

$$x^{k+1} = \arg\min \left\{ \sum_{i=1}^{n} \left(\frac{1}{(|x_i^k| + \varepsilon_k)^{1-p}} \right) |x_i| : Ax = b \right\}, \quad p \in (0,1).$$

(c) **NW1** algorithm derived from Example 2.3(ii)

$$x^{k+1} = \arg\min \left\{ \sum_{i=1}^{n} \left(\frac{p + (|x_i^k| + \varepsilon_k)^{1-p}}{(|x_i^k| + \varepsilon_k)^{1-p} \left[|x_i^k| + \varepsilon_k + (|x_i^k| + \varepsilon_k)^p \right]} \right) |x_i| : Ax = b \right\},$$

where $p \in (0,1)$.

(d) **NW2** algorithm derived from Example 2.5

$$x^{k+1} = \arg\min \left\{ \sum_{i=1}^{n} \left[\frac{q + (|x_i^k| + \varepsilon_k)^{1-q}}{(|x_i^k| + \varepsilon_k)^{1-q} \left[|x_i^k| + \varepsilon_k + (|x_i^k| + \varepsilon_k)^q \right]^{1-p}} \right] |x_i| : Ax = b \right\},$$

where $p, q \in (0,1)$.

(e) **NW3** algorithm derived from Example 2.5

$$x^{k+1} = \arg\min \left\{ \sum_{i=1}^{n} \left(\frac{1 + 2(|x_i^k| + \varepsilon_k)}{(|x_i^k| + \varepsilon_k + (|x_i^k| + \varepsilon_k)^2)^{1-p}} \right) |x_i| : Ax = b \right\},$$

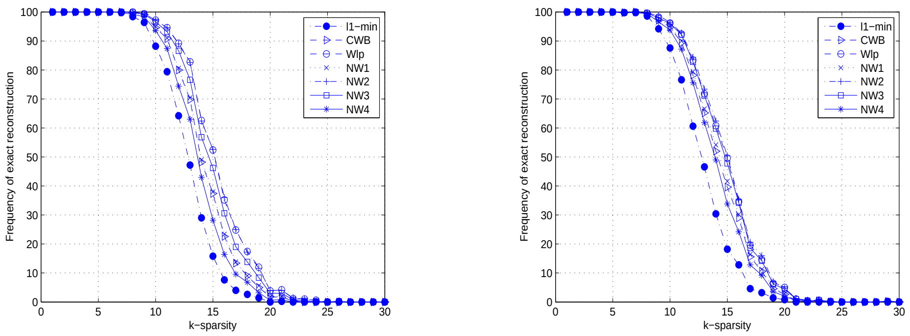
where $p \in (0, 1/2]$.

(f) **NW4** algorithm based on Example 2.6

$$x^{k+1} = \arg\min \left\{ \sum_{i=1}^{n} \left(\frac{1 + (|x_i^k| + \varepsilon_k)^p}{(|x_i^k| + \varepsilon_k)^{1+p}} \right) |x_i| : Ax = b \right\}, \quad p \in (0, \infty).$$

To compare these methods, we randomly generate the pair (A, x), where $A \in R^{50\times 250}$ and x is a k-sparse vector in R^{250} with $k = 1, 2, ..., 30$. Throughout this section, all random pairs (A, x) are generated based on the following assumption: The entries of A and x on its support are i.i.d Gaussian random variables with zero mean and unit variances. Once (A, x) is generated, we set $b = Ax$ and test the algorithms on this system. For every given sparsity k, 500 pairs of (A, x) were generated, and we compare the success probability of all the above-mentioned algorithms in locating k-sparse solutions. For all tested instances of $Ax = b$, every reweighted algorithm was executed only 4 iterations, and the same parameters $\alpha = 0.5, \varepsilon_0 = 0.01$ and the initial point $x^0 = e \in R^{250}$ were used in Algorithm 2.2. Note that $x^0 = e$ implies that the first iteration of the algorithm is actually the ℓ_1-minimization. Given a k-sparse solution x of $Ax = b$, the algorithm claims to be successful in finding (or exact reconstruction of) the k-sparse solution x if the found solution x^k satisfies that $\|x^k\|_{\tilde{0}} \le k$ and $\|x^k - x\| \le 10^{-5}$ where $\|x^k\|_{\tilde{0}}$ is defined, in our experiments, to be the number of components of x satisfying $|x_i^k| \ge 10^{-5}$. Clearly, the main computational cost in Algorithm 2.2 is solving weighted ℓ_1-minimization problems. To solve these problems, we use CVX, a package for specifying and solving convex programs [26].

Since NW2 has two parameters (p, q), we set $q = p$ for this algorithm in all tests for simplicity. Our experiments show that no matter what value of $p \in (0, 1)$ is taken (for NW3, p is restricted in $(0, 1/2]$), all six reweighted ℓ_1-algorithms defined above remarkably outperform ℓ_1-minimization in recovering the desired sparse solutions. The new algorithms NW2 and NW3 proposed in this paper and the existing Wl$_p$ method are particularly strong. The results for $p = 0.3$ and 0.5 are summarized in Figure 4.1, in which the success probability of ℓ_1-minimization is also included.



(a) $p = 0.5$ (b) $p = 0.3$

FIG. 4.1. *Comparison of success rates of finding the k-sparse solution x of $b = Ax$, where $A \in R^{50\times 250}$ and $x \in R^{250}$. For each k-sparsity, 500 attempts were made. All six reweighted ℓ_1-methods outperform ℓ_1-minimization.*

The success probability of these algorithms are different. For example, for the sparsity $k = 15$, Figure 4.1 (a) shows that the success probability of ℓ_1-minimization to find the desired sparse solution is about 17%, NW4 is about 28%, NW1 and CWB are about 38%, NW3 is about 46%, and NW2 and Wl$_p$ are about 52%. A similar result can be seen from Figure 4.1 (b). Figure 4.1 demonstrates that NW2 and Wl$_p$ methods perform particularly well in finding the sparse solutions of linear systems. Thus it is interesting to single out these two algorithms, and further compare their performances (also with the ℓ_1- and CWB method). The experiments show that when

p is relatively small (e.g. $p \leq 0.2$), NW2 outperforms Wl_p method. For $0.3 \leq p \leq 0.6$, NW2 and Wl_p are quite comparable, and in many cases their numerical performances are almost identical. When p is relatively large (e.g. $p \geq 0.7$), Wl_p method can outperform NW2 in many situations. The results for $p = 0.01, 0.1, 0.3$ and 0.7 were summarized in Figure 4.2.

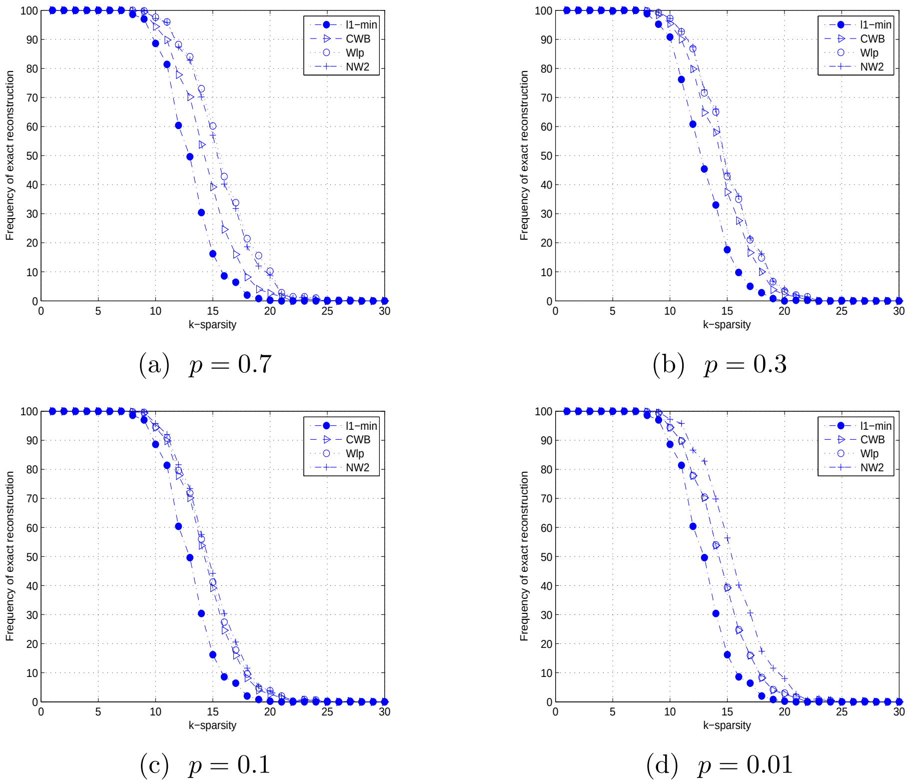


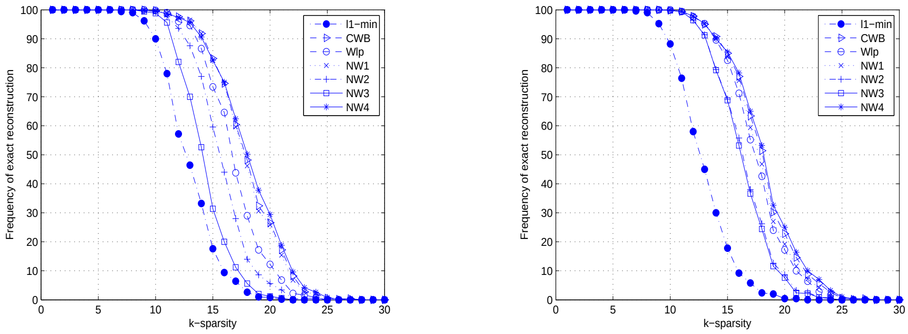
FIG. 4.2. *Comparison of NW2, Wl_p, CWB, and ℓ_1-minimization through the success frequency of finding the k-sparse solution of $b = Ax$, where $A \in R^{50\times 250}$. 500 attempts were made for each k-sparsity ($k = 1, ..., 30$), and different values of p were tested.*

The above numerical experiments were carried out by using the updating rule $\varepsilon_{k+1} = \alpha\varepsilon_k$ where $\alpha = 0.5$. These experiments have demonstrated that all tested reweighted ℓ_1-algorithms (NW1-NW4, CWB, and Wl_p) do outperform the standard ℓ_1-minimization in many situations. We also observed that in the aforementioned testing environment, NW2, NW3 and Wl_p perform better than CWB, NW1, and NW4 in many situations, and NW2 and Wl_p methods are quite comparable to each other.

However, these numerical results cannot imply that the overall performance of NW2, NW3, and Wl_p is always better than CWB, NW1, and NW4. It is interesting to test algorithms using a different parameter updating rule. Candès, Wakin and Boyd [9] proposed the following rule:

$$\varepsilon_k = \max\left\{[\sigma(x^k)]_{i_0}, 10^{-3}\right\}, \tag{4.1}$$

where i_0 denotes the nearest integer to $m/[4\log(n/m)]$. Let us replace the updating scheme $\varepsilon_{k+1} = \varepsilon_k/2$ in Algorithm 2.2 by (4.1), and redo experiments. The results for $p = 0.5$ and 0.3 were summarized in Figure 4.3, from which we see that all tested reweighted algorithms still remarkably outperform the standard ℓ_1-minimization, but this time CWB, NW1, and NW4 perform better than NW2, NW3, and Wl_p, and these three methods (CWB, NW1, and NW4) are quite comparable and the recovery by these three methods is robust with respect to the choice of the parameter p.



(i) $p = 0.5$, ε_k is updated by (4.1) (ii) $p = 0.3$, ε_k is updated by (4.1)

FIG. 4.3. *Comparison of success rates of reweighted algorithms with (4.1) in finding the k-sparse solution of $b = Ax$, where $A \in R^{50\times 250}$. For each k-sparsity, 500 attempts were made.*

In summary, all the tested reweighted methods can outperform the standard ℓ_1-method in finding sparse solutions of linear systems. From Figures 4.1 and 4.3, the numerical performance of a reweighted algorithm may depend on the updating rule of the parameter ε_k, and we observe that CWB, NW1, and NW4 perform best when using (4.1), compared with the remaining algorithms using either the rule (4.1) or $\varepsilon_{k+1} = \varepsilon_k/2$.

5. Conclusions. Via a merit function for sparsity which is certain concave approximation of the cardinality function, the concave minimization plays an important role in locating sparse solutions of underdetermined linear systems of equations. Through a linearization technique, minimizing concave merit functions for sparsity yields a unified approach for reweighted ℓ_1-minimization algorithms. This unified approach not only makes it easy to construct various new specific reweighted ℓ_1-algorithms for the sparse solution of linear systems, but also enables us to develop a new and unified convergence theory for a large family of such algorithms. The analysis in this paper is based on the so-called range space property, which is different from the existing RIP/NSP-based analysis. As special cases of our general framework, a convergence result for the well-known ℓ_p-quasi-norm-based reweighted algorithm and Candès-Wakin-Boyd method can be obtained, respectively, from Theorems 3.11 and 3.9 in this paper. Moreover, several specific reweighted ℓ_1-algorithms have been constructed, and their efficiency of finding sparse solutions of linear systems has been demonstrated by numerical experiments. Although the simulation shows that reweighted ℓ_1-algorithms outperform the standard ℓ_1-method in many situations, a rigorous mathematical proof for this phenomena has not been carried out so far. This remains an open question in this field. What we have actually proved in this paper is that, under suitable conditions, a large family of reweighted ℓ_1-algorithms can generate a solution with a certain level of sparsity to the linear system.

REFERENCES

[1] P. Blomgren and T.F. Chan, Color TV: total variation methods for restoration of vector-valued images, *IEEE Trans. Image Process.*, 7 (1998), pp. 304-309.
[2] A. Bruckstein, D. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, 51 (2009), pp. 34-81.
[3] E. Candès, Compressive sampling, *International Congress of Mathematicians*, Vol. III, 2006, pp.1433-1452.
[4] E. Candès and B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.*, 9 (2009), pp. 717-772.
[5] E. Candès, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly impcomplete frequency information, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 489-509.
[6] E. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.*, 59 (2006), pp. 1207-1223.
[7] E. Candès and T. Tao, The power of convex relaxation: near-optimal matrix completion, *IEEE Trans. Inform. Theory*, 56 (2009), pp. 2053-2080.
[8] E. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), pp. 4203-4215.
[9] E. Candès, M. Wakin and S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, 14 (2008), pp. 877-905.
[10] R. Chartrand, Exact reconstruction of sparse signals via nonconex minimization, *IEEE Signal Process. Lett.*, 14 (2007), pp. 707-710.
[11] S. Chen, D. Donoho and M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, 20(1998), pp. 33-61.
[12] X. Chen and W. Zhou, Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization, Technical Report, HK Polytech. Univ., 2010.
[13] A. Cohen, W. Dahmen and R. DeVore, Compressed sensing and best k-term approximation, *J. Amer. Math. Soc.,* 22 (2009), pp. 211-231.
[14] I. Daubechies, R. DeVore, M. Fornasier and C.S. Güntürk, Iteratively reweighted least squares minimization for sparse recovery, *Commun. Pure. Appl. Math.*, 63 (2010), pp. 1-38.
[15] M. Davies and R. Gribonval, Restricted isometry constants where ℓ_p sparse recovery can fail for $0 < q \leq 1$, *IEEE Trans. Inform. Theory*, 55 (2009), pp. 2203-2214.
[16] D. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 1289-1306.
[17] D. Donoho and M. Elad, Optimality sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization, *Proc. Natl. Acad. Sci.*, 100 (2003), pp. 2197-2202.
[18] D. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inform. Theory*, 47 (1999), pp. 2845-2862.
[19] D. Donoho and P. Stark, Uncertainty priciples and signal recovery, *SIAM J. Appl. Math.*, 49(1989), pp. 906-931.
[20] M. Fazel, *Matrix rank minimization with applications*, PhD Thesis, Stanford Univ., 2002.
[21] M. Fazel, H. Hindi and S. Boyd, Rank minimization and applicaitons in system theory, in *Proc. Am. Control Conf.*, Vol. 4, 2004, pp. 3273-3278.
[22] M. Fazel, H. Hindi and S. Boyd, Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices, in *Proc. Am. Control Conf.*, June 2003.
[23] M. Fornasier and H. Rauhut, *Compressive sensing*, in the *Handbook of Mathematical Methods in Imaging*, Otmar Scherzer (ed.), Vol.1, Springer, 2011.
[24] S. Foucart and M. Lai, Sparsest solutions of undertermined linear systems via ℓ_p-minimization for $0 < q \leq 1$, *Appl. Comput. Harmon. Anal.*, 26 (2009), pp. 395-407.
[25] I. Gorodnitsky and B. Rao, Sparse signal reconstruction from limited data using FOCUSS: A reweighted minimum norm algorithm, *IEEE Trans. Signal Process.*, 45 (1997), pp. 600-616.
[26] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 1.21, April 2011.
[27] R. Gribonval and M. Nielsen, Sparse decompositions in unions of bases, *IEEE Trans. Inform. Theory*, 49 (2003), pp. 3320-3325.
[28] G. Harikumar and Y. Bresler, A new algorithm for computing sparse solutions to linear inverse problems, in *Proc. Int. Conf. Acoustics, Speech, Singnal Processing (ICASSP)*, May 1996.
[29] P. Holland and R. Welsch, Robust regression using iteratively reweighted least-squares, *Commun. Stat. Theor. Methods*, A6 (1977), pp. 813-827.
[30] R. Horst, P.M. Pardalos and N. V. Thoai, *Introduction to Global Optimization*, Kluwer Aca-

demic publishers, 2rd Edition, 2000.
[31] M. Lai and J. Wang, An unconstrainted ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems, *SIAM J. Optim.*, 21 (2010), pp. 82-101.
[32] C. Lawson, Contributions to the theory of linear least maximum approximation, PhD thesis, University of Califonia, Los Angeles, 1961.
[33] M. Lobo, M. Fazel and S. Boyd, Portifolio optimization with linear and fixed transaction costs, *Ann. Oper. Res.*, 152(2007), pp. 341-365.
[34] S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.*, 41 (1993), pp. 3397-3415.
[35] B. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Comput.*, 24 (1995), pp. 227-234.
[36] D. Needell, Noisy signal recovery via iterative reweighted ℓ_1-minimization, In *Proceedings of the 43rd Asilomar conference on Signals, Systems and Computers*, Asilomar'09, 2009, pp. 113-117.
[37] B. Recht, M. Fazel and P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Rev.,* 52 (2010), pp. 471-501.
[38] B. Recht, W. Xu and B. Hassibi, Null space conditions and thresholds for rank minimization, *Math. Program. Ser. B*, 127 (2011), pp. 175-211.
[39] R. Tibshirani, Regression shrignkage and selection via the lasso, *J. Royal Statist. Soc.*, B 58 (1996), pp. 267-288.
[40] H. Taylor, S. Banks and J. McCoy, Deconvolution with the ℓ_1 norm, *Geophysics*, 44 (1979), 39-52.
[41] J. Tropp and D. Needell, CoSaMP: Iterative signal recovery from imcomplete and inaccurate samples, *Appl. Comput. Harmon. Anal.,* 26 (2008), pp. 301-321.
[42] D. Wipf and S. Nagarajan, Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions, Preprint, 2010.
[43] Y. Zhang, Theory of compressive sensing via ℓ_1-mimimization: A Non-RIP analysis and extensions, Technical Report, Rice Univ., 2008.
[44] Y.B. Zhao, An approximation Theory of matrix rank minimization and its application to quadratic equations, *Linear Algebra Appl.*, 437 (2012), pp. 77-93.
[45] Y.B. Zhao, F.C. Fang and D. Li, Constructing generalized mean functions using convex functions with regularity conditions, *SIAM J. Optim.*, 17 (2006), pp. 37-51.