

## LETTER

# An Iterative Reweighted Least Squares Algorithm with Finite Series Approximation for a Sparse Signal Recovery

Kazunori URUMA<sup>†a)</sup>, *Nonmember*, Katsumi KONISHI<sup>††</sup>, *Member*, Tomohiro TAKAHASHI<sup>†</sup>, *Student Member*, and Toshihiro FURUKAWA<sup>†</sup>, *Member*

**SUMMARY** This letter deals with a sparse signal recovery problem and proposes a new algorithm based on the iterative reweighted least squares (IRLS) algorithm. We assume that the non-zero values of a sparse signal is always greater than a given constant and modify the IRLS algorithm to satisfy this assumption. Numerical results show that the proposed algorithm recovers a sparse vector efficiently.

**key words:** compressed sensing, sparse optimization, iterative reweighted least squares

## 1. Introduction

The theory of compressed sensing shows that a sparse signal can be recovered exactly from fewer measurements than traditionally believed necessary by solving the following  $\ell_0$  minimization problem [1]–[3],

$$\text{Minimize } \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{Ax} = \mathbf{y}, \quad (1)$$

where  $\|\cdot\|_0$  denotes the number of non-zero values in a vector,  $A \in R^{m \times n}$  ( $m < n$ ) is a measurement matrix, and  $\mathbf{y} \in R^m$  is an observation vector. In the case of  $\|\mathbf{a}\|_0 = K$ ,  $\mathbf{a}$  is called K-sparse. Unfortunately, the problem (1) is NP-hard in general, and therefore a lot of studies consider the following  $\ell_p$  norm minimization to relax this problem,

$$\text{Minimize } \|\mathbf{x}\|_p^p \quad \text{subject to } \mathbf{Ax} = \mathbf{y}, \quad (2)$$

where  $0 < p \leq 1$ . If some conditions are satisfied, (2) gives an optimal solution of (1) [4]. Hence various methods have been proposed such as orthogonal matching pursuit (OMP) [5], forward-backward splitting (FOBOS) [7], constraint removal (CR) [6], reweighted  $\ell_1$  minimization [8] and iterative reweighted least squares (IRLS) [9], [10]. The latest study has reported that the IRLS algorithm has the best performance in major applications [11].

In some applications, we consider signals which are not exactly sparse to be sparse, that is, we neglect relatively small value elements and obtain sparse vectors. For example, the wavelet transformed image is considered to be a K-sparse signal by selecting the top K absolute values and neglecting other ones [12]. In [13] an image colorization algorithm was proposed based on the sparse optimization, where

the signal is approximated as a sparse signal by letting the elements with smaller value than a given constant be zero. Therefore we assume in this letter that the optimal sparse signal  $\mathbf{x}_{opt}$  of the problem (1) is included in  $R_v^n$ , where  $R_v^n$  is defined by

$$R_v^n = \{\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T \in R^n : |x_i| > v \text{ or } x_i = 0\},$$

and  $v > 0$  is a given constant. The contribution of this letter is to propose a new IRLS algorithm which utilizes this assumption explicitly to provide more accurate sparse solution than other algorithms. This letter also proposes a heuristic method for unknown  $v$ . Numerical results show that the proposed algorithm recovers the sparse signal efficiently.

## 2. Main Results

### 2.1 Proposed Algorithm

This letter considers the following  $\ell_1$  norm minimization problem to find the optimal solution  $\mathbf{x}_{opt}$  of the problem (1),

$$\text{Minimize } \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{Ax} = \mathbf{y}, \mathbf{x} \in R_v^n, \quad (3)$$

where  $v > 0$  is a given constant. The IRLS algorithm [9], [10] gives the  $\ell_1$  minimization solution by alternately updating the weighted least squares solution and the weight. The update scheme of the  $t$ th iteration is follows,

$$\mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x} \in R^n \text{ s.t. } \mathbf{Ax} = \mathbf{y}} \sum_{i=1}^n w_i^{(t)} x_i^2, \quad (4)$$

and

$$w_i^{(t+1)} = \frac{1}{|x_i^{(t+1)}|}. \quad (5)$$

As mentioned in the previous section, we assume here that the optimal solution  $\mathbf{x}_{opt}$  is included in  $R_v^n$ . To obtain  $\mathbf{x}^{(t)} \in R_v^n$  in each iteration, this letter proposes the following simple soft-thresholding scheme before updating the weight  $w_i^{(t+1)}$ ,

$$x_i^{(t+1)} \leftarrow \text{sign}(x_i^{(t+1)}) \cdot (|x_i^{(t+1)}| - v)_+, \quad (6)$$

where  $\text{sign}(\cdot)$  denotes the signum function and  $(a)_+ = \max(a, 0)$ . Then we obtain a new weight as follows from (5) and (6),

Manuscript received August 28, 2013.

Manuscript revised October 18, 2013.

<sup>†</sup>The authors are with the Department of Management Engineering, Tokyo University of Science, Tokyo, 162–8601 Japan.

<sup>††</sup>The author is with the Department of Computer Science, Kogakuin University, Tokyo, 163–8677 Japan.

a) E-mail: uru-kaz@ms.kagu.tus.ac.jp

DOI: 10.1587/transinf.E97.D.319

$$w_i^{(t+1)} = \begin{cases} \frac{1}{|x_i^{(t+1)}| - \nu} & (|x_i^{(t+1)}| > \nu) \\ \infty & (|x_i^{(t+1)}| \leq \nu) \end{cases}. \quad (7)$$

Next we focus on an implementation of (7). A simple way to implement this weight is to replace the infinity with a large constant. Experimental results show that this implementation works well only if  $\nu$  is small enough or if the initial value  $\mathbf{x}^{(0)}$  is nearly an optimal solution. In the IRLS algorithm, we usually use the least square solution  $\mathbf{x}_{LS} = A^T(AA^T)^{-1}\mathbf{y}$  as  $\mathbf{x}^{(0)}$ , and the most elements of non-sparse  $\mathbf{x}_{LS}$  are smaller than  $\nu$  since the optimal solution is sparse. This causes the most elements of  $\mathbf{x}^{(t)}$  for  $t \geq 1$  to remain to be less than  $\nu$  due to a large  $w_i^{(t)}$ , and therefore we obtain a wrong solution. In order to avoid sticking in  $[0, \nu]$  of  $x_i^{(t+1)}$ , this letter approximates the weight (7) by the strictly monotonically decreasing function for  $|x_i|$  as follows,

$$w_i^{(t+1)} = \frac{1}{\nu} \sum_{l=1}^L \left( \frac{\nu}{|x_i^{(t+1)}|} \right)^l = \frac{1}{\nu} \sum_{l=1}^L \frac{1}{\left( \frac{|x_i^{(t+1)}|}{\nu} \right)^l}. \quad (8)$$

Note that the weight (8) is exactly equal to (7) if  $L \rightarrow \infty$  because

$$\begin{aligned} \frac{1}{\nu} \sum_{l=1}^{\infty} \left( \frac{\nu}{|x_i^{(t+1)}|} \right)^l &= \frac{1 - \left( \frac{\nu}{|x_i^{(t+1)}|} \right)^{\infty}}{|x_i^{(t+1)}| - \nu} \\ &= \begin{cases} \frac{1}{|x_i^{(t+1)}| - \nu} & (|x_i^{(t+1)}| > \nu) \\ \infty & (|x_i^{(t+1)}| \leq \nu) \end{cases}. \end{aligned}$$

This weight enables  $x_i^{(t+1)}$  to take a larger value than  $\nu$  even if  $x_i^{(t)} \leq \nu$ , and therefore we can avoid the value 0 of undesirable elements in a solution. This letter proposes the following computable expression of (8),

$$w_i^{(t+1)} = \frac{1}{\nu} \sum_{l=1}^L \frac{1}{\left[ \frac{|x_i^{(t+1)}|}{\nu} + \epsilon^{1/l} \right]^l}, \quad (9)$$

where  $L \geq 1$  and  $\epsilon > 0$  are given constants. Because the value of  $\epsilon$  affects the performance of IRLS, this letter applies the update scheme of  $\epsilon$  proposed in [10], where  $\epsilon$  is updated as  $\epsilon \leftarrow \epsilon/10$  when the relative error  $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_2 / \|\mathbf{x}^{(t)}\|_2$  is less than  $\sqrt{\epsilon}/100$ . Finally, this letter proposes a new IRLS algorithm as shown in Algorithm 1.

## 2.2 Termination Property of the Algorithm

This subsection gives an important property of the algorithm, which guarantees the termination of the algorithm. Because it is enough to prove  $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\| \rightarrow 0$  as  $t \rightarrow \infty$  for any  $\epsilon > 0$ , we deal  $\epsilon$  as a given constant in this subsection.

For convenience, we define a vector  $\mathbf{w}_{(l)}^{(t)} = [w_{(l)1}^{(t)}, \dots, w_{(l)n}^{(t)}] \in \mathbf{R}^n$ , where  $w_{(l)i}^{(t)}$  is by

$$w_{(l)i}^{(t)} = \frac{1}{\nu \left[ \frac{|x_i^{(t)}|}{\nu} + \epsilon^{1/l} \right]^l}. \quad (10)$$

### Algorithm 1 Proposed algorithm

**Input:**  $\mathbf{A}, \mathbf{y}, \nu > 0, L \geq 1$

set  $t = 0$

set  $\epsilon = 1$

set  $w_i^{(0)} = 1$  for  $i = 1, 2, \dots, n$

**repeat**

**repeat**

calculate  $\mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x} \in \mathbf{R}^n \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{y}} \sum_{i=1}^n w_i^{(t)} x_i^2$

update  $w_i^{(t+1)} = \frac{1}{\nu} \sum_{l=1}^L \frac{1}{\left[ \frac{|x_i^{(t+1)}|}{\nu} + \epsilon^{1/l} \right]^l}$

$t \leftarrow t + 1$

**until**  $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_2 / \|\mathbf{x}^{(t)}\|_2 < \sqrt{\epsilon}/100$

$\epsilon \leftarrow \epsilon/10$

**until**  $\epsilon = 10^{-9}$

**Output:**  $\mathbf{x}^{(t)}$

Then  $w_i^{(t+1)}$  in (9) can be rewritten as  $w_i^{(t+1)} = \sum_{l=1}^L w_{(l)i}^{(t+1)}$ . Let us define the function  $\mathcal{J}$  for  $0 < w_{(l)i}^{(t)} \leq \frac{1}{\nu\epsilon}$  as follows,

$$\begin{aligned} \mathcal{J}(\mathbf{x}^{(t)}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}) \\ = \sum_{l=1}^L \left[ \sum_{i=1}^n w_{(l)i}^{(t)} (x_i^{(t)})^2 + \nu^2 \int_{w_{(l)i}^{(t)}}^{\frac{1}{\nu\epsilon}} ((\nu u)^{-\frac{1}{l}} - \epsilon^{\frac{1}{l}})^2 du \right], \end{aligned} \quad (11)$$

where  $\epsilon, \nu, L$  are given constants, and  $t$  is the number of iterations. The function  $\mathcal{J}$  is a convex function of each variable (in fact,  $\partial^2 \mathcal{J} / \partial w_{(l)i}^{(t)2} \geq 0$ ), and  $w_{(l)i}^{(t)}$  in (10) satisfies  $\partial \mathcal{J} / \partial w_{(l)i}^{(t)} = 0$ . This implies that (10) minimizes  $\mathcal{J}$ . Then Algorithm 1 minimizes the objective function  $\mathcal{J}$  in each iteration, and it holds that

$$\mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x} \in \mathbf{R}^n \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{y}} \mathcal{J}(\mathbf{x}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}), \quad (12)$$

and that

$$\begin{aligned} [\mathbf{w}_{(1)}^{(t+1)}, \dots, \mathbf{w}_{(l)}^{(t+1)}, \dots, \mathbf{w}_{(L)}^{(t+1)}] \\ = \arg \min_{[\mathbf{w}_{(1)}, \dots, \mathbf{w}_{(l)}, \dots, \mathbf{w}_{(L)}]} \mathcal{J}(\mathbf{x}^{(t+1)}, \mathbf{w}_{(1)}, \dots, \mathbf{w}_{(l)}, \dots, \mathbf{w}_{(L)}). \end{aligned} \quad (13)$$

We also have the following monotonicity property which holds for all  $t \geq 0$ ,

$$\begin{aligned} \mathcal{J}(\mathbf{x}^{(t)}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}) \\ \geq \mathcal{J}(\mathbf{x}^{(t+1)}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}) \\ \geq \mathcal{J}(\mathbf{x}^{(t+1)}, \mathbf{w}_{(1)}^{(t+1)}, \mathbf{w}_{(2)}^{(t+1)}, \dots, \mathbf{w}_{(L)}^{(t+1)}). \end{aligned} \quad (14)$$

We define a constant  $\alpha > 0$  as follows

$$\mathcal{J}(\mathbf{x}^{(1)}, \mathbf{w}_{(1)}^{(0)}, \mathbf{w}_{(2)}^{(0)}, \dots, \mathbf{w}_{(L)}^{(0)}) = \alpha. \quad (15)$$

Then we obtain the following inequalities,

$$\begin{aligned} \alpha &\geq \mathcal{J}(\mathbf{x}^{(t)}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}) \\ &\geq w_{(1)i}^{(t)} x_i^{(t)2} \\ &\geq w_{(1)i}^{(t)} (w_{(1)i}^{(t)})^{-1} - \nu\epsilon^2 \\ &\geq w_{(1)i}^{(t)-1} - 2\nu\epsilon, \end{aligned}$$

and hence it holds that  $w_{(1)i}^{(t)} \geq (\alpha + 2\nu\epsilon)^{-1}$ . Thus we have

$$\begin{aligned}
& \mathcal{J}(\mathbf{x}^{(t)}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}) \\
& \quad - \mathcal{J}(\mathbf{x}^{(t+1)}, \mathbf{w}_{(1)}^{(t+1)}, \mathbf{w}_{(2)}^{(t+1)}, \dots, \mathbf{w}_{(L)}^{(t+1)}) \\
& \geq \mathcal{J}(\mathbf{x}^{(t)}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}) \\
& \quad - \mathcal{J}(\mathbf{x}^{(t+1)}, \mathbf{w}_{(1)}^{(t)}, \mathbf{w}_{(2)}^{(t)}, \dots, \mathbf{w}_{(L)}^{(t)}) \\
& = \sum_{l=1}^L \sum_{i=1}^n w_{(l)i}^{(t)} [x_i^{(t)2} - x_i^{(t+1)2}] \\
& = \sum_{l=1}^L \sum_{i=1}^n w_{(l)i}^{(t)} [x_i^{(t)} - x_i^{(t+1)}]^2 \\
& \quad - 2 \sum_{l=1}^L \sum_{i=1}^n w_{(l)i}^{(t)} x_i^{(t+1)} [x_i^{(t+1)} - x_i^{(t)}] \\
& = \sum_{l=1}^L \sum_{i=1}^n w_{(l)i}^{(t)} [x_i^{(t)} - x_i^{(t+1)}]^2 \\
& \geq (\alpha + 2\nu\epsilon)^{-1} \sum_{i=1}^n [x_i^{(t)} - x_i^{(t+1)}]^2 \\
& = (\alpha + 2\nu\epsilon)^{-1} \|\mathbf{x}^{(t)} - \mathbf{x}^{(t+1)}\|^2, \tag{16}
\end{aligned}$$

where the second equality uses the fact that

$$\sum_{k=1}^n \sum_{i=1}^N w_{(k)i}^{(t)} x_i^{(t+1)} [(x_i^{(t+1)}) - (x_i^{(t)})] = 0$$

(observe that  $A(\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}) = 0$  and that  $\mathbf{x}^{(t+1)}$  is the least squares solution with  $\mathbf{w}^{(t)}$ ). If we now sum these inequalities over  $t \geq 1$ , we have that

$$\begin{aligned}
\alpha & \geq \alpha - \mathcal{J}(\mathbf{x}^{(\infty)}, \mathbf{w}_{(1)}^{(\infty)}, \mathbf{w}_{(2)}^{(\infty)}, \dots, \mathbf{w}_{(L)}^{(\infty)}) \\
& \geq \mathcal{J}(\mathbf{x}^{(1)}, \mathbf{w}_{(1)}^{(1)}, \mathbf{w}_{(2)}^{(1)}, \dots, \mathbf{w}_{(L)}^{(1)}) \\
& \quad - \mathcal{J}(\mathbf{x}^{(\infty)}, \mathbf{w}_{(1)}^{(\infty)}, \mathbf{w}_{(2)}^{(\infty)}, \dots, \mathbf{w}_{(L)}^{(\infty)}) \\
& \geq (\alpha + 2\nu\epsilon)^{-1} \sum_{t=1}^{\infty} \|\mathbf{x}^{(t)} - \mathbf{x}^{(t+1)}\|^2,
\end{aligned}$$

and therefore it holds that

$$\alpha(\alpha + 2\nu\epsilon) \geq \sum_{t=1}^{\infty} \|\mathbf{x}^{(t)} - \mathbf{x}^{(t+1)}\|^2.$$

In particular we have

$$\lim_{t \rightarrow \infty} (\mathbf{x}^{(t)} - \mathbf{x}^{(t+1)}) = 0. \tag{17}$$

This implies that Algorithm 1 always terminates. This property guarantees only the termination condition of Algorithm 1 and not the convergence to the optimal solution. As well as the original IRLS algorithm, the proposed algorithm is not guaranteed to provide exact solutions, however, numerical results show that it works well to obtain sparse solutions [10].

### 2.3 Case of Unknown $\nu$

Although the performance of the algorithm depends on the value of  $\nu$ , it is not always known. Since the proposed weight is approximated by series, the algorithm has robust to the value of  $\nu$  which is larger than a true value. It is trivial that  $0 < \nu < \max(|x_{opt_i}|)$ , where  $\max(|x_i|)$  denotes the largest absolute element in  $\mathbf{x}$ , and therefore this paper proposes a heuristic where  $\nu$  is given as  $\nu = \max(|x_i^{(0)}|)$  at the first iteration and then is reduced gradually as follows,

$$\nu^{(0)} = \max(|x_i^{(0)}|),$$

and

$$\nu^{(t+1)} = \max\left(\epsilon, \min\left(\nu^{(t)}, \eta^t \max(|x_i^{(t+1)}|)\right)\right), \tag{18}$$

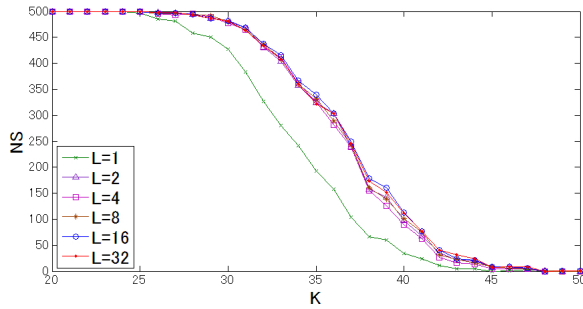
where  $\eta < 1$ . The value of  $\nu$  gradually decreases at each iteration, and  $\nu^{(t)} \rightarrow \epsilon$  as  $t \rightarrow \infty$ . Empirical results show that  $\eta = 0.995$  and  $\epsilon = 10^{-8}$  achieve the best performance, and this value is used in all experiments of the next section.

### 3. Experimental Result

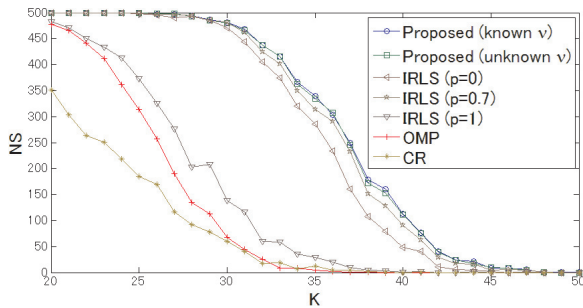
This section presents numerical examples to show the efficiency of the proposed algorithm comparing with IRLS, orthogonal matching pursuit (OMP) [5], and constraint removal (CR) [6]. We select entries of a measurement matrix  $A \in \mathbf{R}^{100 \times 256}$  from Gaussian distribution of mean 0 and standard deviation  $\frac{1}{100}$ , then scale the columns to have unit 2-norm. A K-sparse vector  $\mathbf{x}_{opt} \in \mathbf{R}^{256}$  is generated by taking the top K absolute values of a vector where the value of each element is chosen from Gaussian distribution of mean 0 and standard deviation  $\sigma$ , and an observation signal vector  $\mathbf{y} \in \mathbf{R}^{100}$  is constructed according to the equation  $A\mathbf{x}_{opt} = \mathbf{y}$ . In IRLS, we use  $p \in \{0, 0.7, 0.8, 1\}$  of the  $\ell_p$  norm minimization, where  $p = 0.7$  and  $0.8$  are the best values for  $\sigma = 1$  and  $100$ , respectively, obtained by examining of  $p \in [0, 1]$  with each  $0.05$ . Since CR requires a priori information about the sparsity, it is applied with the value of  $K$ . All parameters of IRLS, OMP and CR are selected to achieve the best performance. In all experiments, if the solution  $\mathbf{x}$  satisfies  $\|\mathbf{x} - \mathbf{x}_{opt}\|_2 / \|\mathbf{x}_{opt}\|_2 < 10^{-3}$ , we regard it as successful recovery. Each result shows the number of successful recoveries (NS) of 500 trials.

First we examine Algorithm 1 to investigate the effects of  $L$  in (9), where the value of  $\nu$  is given. Figure 1 shows the NS with  $L \in \{1, 2, 4, 8, 16, 32\}$ . As can be seen, the algorithm achieves the best performance with  $L = 16$ , and therefore we use  $L = 16$  in the rest examples.

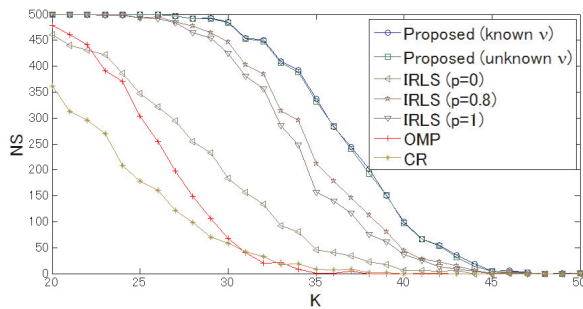
Next we compare the proposed algorithm with OMP [5], CR [6] and IRLS [10] by applying them to the problems of K-sparse vector generated with  $\sigma = 1$  and  $100$ . We set  $\nu$  as the value of  $K$ th element of  $0.95\mathbf{x}_{opt}$ . The proposed algorithm is applied in both cases that the value of  $\nu$  is known and not known. Figures 2–4 show the results. As can be seen, the proposed algorithm with known  $\nu$  has the



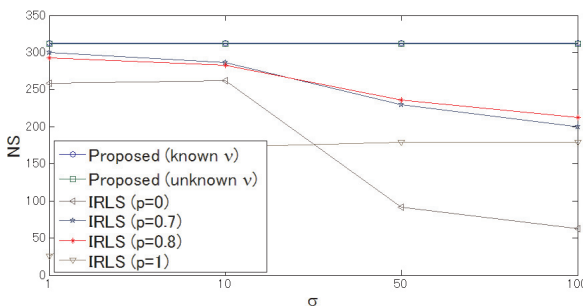
**Fig. 1** The number of successful recoveries vs. the sparsity  $K$  of the proposed algorithm with  $L = 1, 2, 3, 8, 16$  and  $32$ .



**Fig. 2** The number of successful recoveries vs. the sparsity  $K$  of  $x_{opt}$  for  $\sigma = 1$ .



**Fig. 3** The number of successful recoveries vs. the sparsity  $K$  of  $x_{opt}$  for  $\sigma = 100$ .



**Fig. 4** The number of successful recoveries vs. the standard deviation  $\sigma$  for  $K = 35$ .

best performance, and the update scheme (18) of  $v$  provide the almost same performance as the case of known  $v$ . We can also see that the performance of the proposed algorithm is robust for the value of  $\sigma$  although that of IRLS with the best  $p$  is worse when  $\sigma$  is greater, that is, the ratio of  $\|x_{opt}\|_2$  to  $\|A\|_F$  is greater for  $p = 0, 0.7, 0.8$ .

#### 4. Conclusion

This letter deals with the sparse signal recovery problem under the assumption that the non-zero values of sparse signal are greater than a known constant  $v$ . This letter proposes a new algorithm by modifying the IRLS algorithm using this assumption explicitly and gives the proof of termination. Numerical results show that the proposed algorithm recovers sparse signals better than other algorithms and works well even when  $v$  is unknown.

#### References

- [1] K. Hayashi, M. Nagahara, and T. Tanaka, "A user's guide to compressed sensing for communications systems," IEICE Trans. Commun., vol.E96-B, no.3, pp.685–712, March 2013.
- [2] E.J. Candes, J. Romberg, and T. Tao, "Stable signal recovery for incomplete and inaccurate measurements," Commun. Pure Appl. Math., vol.59, no.8, pp.1207–1223, Aug. 2006.
- [3] D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, pp.1289–1306, April 2006.
- [4] E.J. Candes and T. Tao, "Decoding by linear programming," IEEE Trans. Inf. Theory, vol.51, no.12, pp.4203–4215, Dec. 2005.
- [5] G.M. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," J. Constructive Approximation, vol.13, no.1, pp.57–98, 1997.
- [6] A. Sahin and S. Ozen, "Constraint removal for sparse signal recovery," Signal Process., vol.92, no.4, pp.1172–1175, 2012.
- [7] J. Duchi and Y. Singer, "Efficient online and batch learning using forward backward splitting," J. Machine Learning Research, vol.10, pp.2899–2934, Dec. 2009.
- [8] E.J. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," J. Fourier Anal. Appl., vol.14, pp.877–905, 2008.
- [9] I. Daubechies, R. DeVore, M. Fornasier, and C. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," Comm. Pure Appl. Math., vol.63, no.1, pp.1–38, 2010.
- [10] R. Chartrand and W. Yin, "An iteratively reweighted algorithms for compressive sensing," Proc. IEEE Int. Conf. Acoust. Speech Signal Process., pp.3869–3872, April 2008.
- [11] Q. Lyu, Z. Lin, Y. She, and C. Zhang, "A comparison of typical  $\ell_p$  minimization algorithm," Neurocomputing, Jan. 2013. (Preprint submitted).
- [12] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, R. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," Proc. European Conference on Computer Vision, pp.155–168, 2009.
- [13] K. Uruma, K. Konishi, T. Takahashi, and T. Furukawa, "Mixed  $\ell_0/\ell_1$  norm minimization approach to image colorization," IEICE Trans. Inf. & Syst., vol.E95-D, no.8, pp.2150–2151, Aug. 2012.