Proceedings of the International MultiConference of Engineers and Computer Scientists 2012 Vol I,
IMECS 2012, March 14 - 16, 2012, Hong Kong

An l_p Norm Minimization Using Auxiliary Function for Compressed Sensing

Hiroshi TSUTSU, Yoshitaka MORIKAWA

Abstract— **Compressed sensing (CS) is a powerful tool for signal measurement and processing, so it has caught much attention in many fields such as medical imaging and image restoration. In CS, we assume the latent signals, such as voices and images, are sparse; i.e. they have many zero components. On this assumption, we search the sparsest signal sufficing the linear measurement equalities. Usually instead of minimizing the number of nonzero components of signal, we minimize the l_1 norm of the solution vector to satisfy the convex condition. Linear programming such as simplex method is often used for this l_1 norm minimization. This paper proposes a fast algorithm for l_p ($0<p<2$) norm minimization using auxiliary function. We derivate the algorithm and show that the smaller p is, the faster it works. We emphasize the $l_{0.95}$ norm minimization is 6.47 times faster than l_1 simplex method.**

Index Terms— **auxiliary function, compressed sensing (CS), fast algorithm, l_p norm, optimization**

I. Introduction

Compressed sensing (CS) has recently emerged as a new aspect replacing classical sampling technology [1],[2],[3]. Usually analogue band-limited signals such as images, music and voices are at first sampled with sampling rate twice larger than their bandwidth frequencies and then compressed for transmission or storage. The compression ratios are around 1 to 20, so almost all information is discarded, which is waste resource usage. CS is a technique to acquire compressed version at input stage.

In the framework of CS, we assume the signals around us are sparse, which means they are expressed as a few nonzero components in their inherent space and observe linearly independent measurements of the signal. We take advantage of the source signal sparsity to solve the sparsest solution to the underdetermined linear simultaneous equations,

$$\min_{\boldsymbol{x}\in C^N} \|\boldsymbol{x}\|_0^0 \quad \text{subject to } \boldsymbol{W}\boldsymbol{x}=\boldsymbol{y}, \tag{1}$$

where $\boldsymbol{x}\in C^N$ is the latent signal, $\boldsymbol{y}\in C^M$ is its observation, $\boldsymbol{W}\in C^{M\times N}$ ($M<N$) is the corresponding observation matrix, and $\|\boldsymbol{x}\|_0^0$ denotes the l_0 norm defined as the support size of $\boldsymbol{x}$,

$$\|\boldsymbol{x}\|_0^0 := |\operatorname{supp}\boldsymbol{x}|. \tag{2}$$

The above l_0 norm minimization is, however, NP-hard because it results in combinatory problem. Then (1) is relaxed to the following l_1 norm minimization,

$$\min_{\boldsymbol{x}\in C^N} \|\boldsymbol{x}\|_1^1 \quad \text{subject to } \boldsymbol{W}\boldsymbol{x}=\boldsymbol{y}, \tag{3}$$

where

$$\|\boldsymbol{x}\|_p := \left(\sum_{n=1}^{N} |x_n|^p\right)^{1/p}. \tag{4}$$

Even by this relaxation the solution to (3) is almost identical to that of (1). Moreover the solution to (3) is easily obtained by the linear programming like simplex method [4]. Owing to this discovery, CS has been applied to many fields, such as medical imaging and signal processing.

Chartland [5] has recently published that the l_p norm minimization

$$\min_{\boldsymbol{x}\in C^N} \|\boldsymbol{x}\|_p^p \quad \text{subject to } \boldsymbol{W}\boldsymbol{x}=\boldsymbol{y}, \tag{5}$$

is more effective than the traditional l_1 norm minimization of (3) and shown a CT reconstruction of 256×256 Shepp-Logan phantom is possible by only 9 projections.

This paper proposes a fast searching algorithm to solve (5) by introducing auxiliary function [6]. In this method, for a given objective function $L(\boldsymbol{x})$, auxiliary variables $\boldsymbol{a}$ are introduced to lead a new objective function $L^{+}(\boldsymbol{x}, \boldsymbol{a})$ and iteratively alternating minimizations of $\boldsymbol{x}$ and $\boldsymbol{a}$ are performed like expectation-maximization (EM) algorithm [7]. If $L^{+}(\boldsymbol{x}, \boldsymbol{a})$ is quadratic in respect to $\boldsymbol{x}$ and the optimum $\boldsymbol{a}$, with $\boldsymbol{x}$ given, are uniquely predicted by analytical formulae, the algorithm becomes very efficient. This is a generalization of iterative reweighted least squares for l_1 norm minimization to l_p norm proposed by Gorodnitsky and Rao [8]. We derive the auxiliary function for l_p norm minimization and show the algorithm.

Following Chartland [5], l_p ($p<1$) norm minimization can achieve perfect reconstruction even by smaller number of observations comparing to l_1 norm minimization. So in this paper, by simulation we examine this property and l_p norm minimization with $p\approx 0.95$ can exhibit perfect reconstruction by 68 % measurements of l_1 norm minimization. Also processing times of our method are respectively 16 % and 6.5 % of those of l_1 norm simplex method when the number of measurements M is 50 and 40 for synthesized signals with sparsity $K=16$ and signal dimension $N=64$.

Manuscript was received December 4, 2011; revised January 3, 2012.

H. TSUTSU is with the Graduate School of Natural Science and Technology, Okayama University, Okayama-shi, 700-8530 Japan. He is now in the Master Course of Electronic and Information Systems Engineering. (e-mail: tsutsu@trans.cne.okayama-u.ac.jp).

Y. MORIKAWA is an emeritus professor of Okayama University. He was with the Division of Electronic and Information Systems Engineering, Graduate School of Natural Science and Technology, Okayama University, Okayama-shi, 700-8530 Japan (e-mail: morikawa@cne.okayama-u.ac.jp). He is interested in Image compression, Cryptography, and Medical Imaging.

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

In II we describe observation effectiveness of l_p ($p<1$) norm minimization, and in III we derive the algorithm using auxiliary function and describe its speed effectiveness. In IV the simulation results are exhibited. Lastly in V we conclude this paper. Without any description we use lower case italic bold-face letters for vectors and upper ones for matrices.

II. L_p MINIMIZATION IN COMPRESSED SENSING

The condition that (5) has a unique solution and exhibits perfect reconstruction for *K*-sparse signal strongly depends on the property of observation matrix $\boldsymbol{W}$. Restricted isotropy (RI) is the most important property.

Definition 1. Let $\boldsymbol{W} \in C^{M\times N}$ be an observation matrix. For $\forall \boldsymbol{x} \in \{\boldsymbol{s} \in C^N;\ |\mathrm{supp}(\boldsymbol{s})| \le S\}$, if there exists the minimum $0 \le \delta_S < 1$ satisfying the following inequalities;

$$(1-\delta_S)\,\|\,\boldsymbol{x}\,\|_2^2 \le \|\,\boldsymbol{W}\boldsymbol{x}\,\|_2^2 \le (1+\delta_S)\,\|\,\boldsymbol{x}\,\|_2^2, \tag{6}$$

then δ_S is called restricted isotropy (RI) constant of $\boldsymbol{W}$.

Candes and Tao [3] derived the perfect reconstruction condition of RI constant for l_1 norm minimization (3). After then, Chartland [5] generalized this condition to l_p norm minimization (5).

Theorem 2. Let $\boldsymbol{W} \in C^{M\times N}$ be an observation matrix. Let $\boldsymbol{x} \in C^N$ and let $K = \|\boldsymbol{x}\|_0^0$ be the size of the support of $\boldsymbol{x}$. Moreover, let $p \in (0, 1]$, and $s>1$. If the RI constant of $\boldsymbol{W}$ satisfies

$$\delta_{sK} + s^{(2-p)/p}\,\delta_{(s+1)K} < s^{(2-p)/p} - 1, \tag{7}$$

then the unique solution of (5) is exactly $\boldsymbol{x}$.

In (7), if p is very positive-small, then $s^{(2-p)/p}$ becomes very large and (7) restricts $\delta_{(s+1)K}$ only. So, this condition is less restrictive when p is small.

Fig.1 illustrates the searching performance of l_1 norm minimization (3) and l_p norm minimization (5) in the case of $M=2$ and $N=3$. In the figure, we suppose the observation space $\boldsymbol{Wx}=\boldsymbol{y}$[1] intersects at the angle θ with the x_3 axis, and let $[0,0,c]^{\mathrm{T}}$ be the intersection point. In this case, the sparsest solution is $\boldsymbol{x}_{\mathrm{sparse}}=[0,0,c]^{\mathrm{T}}$. Nevertheless, when $\theta \ll \pi/4$, the result of l_1 norm minimization is always the point $\boldsymbol{x}_{1\text{-}2}$, at which the observation space $\boldsymbol{Wx}=\boldsymbol{y}$ crosses the x_1-x_2 plane, as shown in part (a). The shown octahedron is the eqi-norm surface $\|\boldsymbol{x}\|_1=\|\boldsymbol{x}_{1\text{-}2}\|_1$. On the contrary, if we select p small enough, l_p norm minimization can find the sparsest point x_{sparse}, as shown in part (b). The shown concave-octahedron is the eqi-norm surface $\|\boldsymbol{x}\|_p=\|\boldsymbol{x}_{\mathrm{sparse}}\|_p$.

In the minimization method which will be described in the succeeding section, we initialize the algorithm by the easily obtained point satisfying the constraint equations $\boldsymbol{Wx}=\boldsymbol{y}$. This initial point is selected as l_2 norm solution to (3) with $p=2$. However, l_2 norm solution reaches more near to the point $\boldsymbol{x}_{1\text{-}2}$ in the last performance example, as θ becomes small. So, the convergent point $\hat{\boldsymbol{x}}_{\mathrm{opt}}$ is also the false point $\boldsymbol{x}_{1\text{-}2}$. Therefore, the admissible p for stable convergence must be carefully selected.

[1] Since $M=2$ and $N=3$, two observation constraints correspond to two planes in 3-dimensional space. Therefore, observation space is an intersection of those two planes, i.e. a right line.

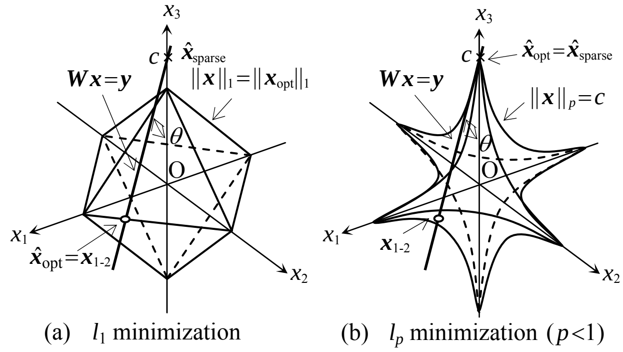

Fig. 1. Illustration for performance difference of l_1 norm minimization (3) and l_p minimization (5) in the case of $M=2$ and $N=3$. When $\boldsymbol{Wx}=\boldsymbol{y}$ intersects at $[0,0,c]^{\mathrm{T}}$ with x_3 axis, the sparsest solution is $\hat{\boldsymbol{x}}_{\mathrm{sparse}}=[0,0,c]^{\mathrm{T}}$. l_p norm minimization can find $\hat{\boldsymbol{x}}_{\mathrm{sparse}}$ but l_1 minimization fails as respectively shown in part (b) and (a).

Lastly, in some cases observation matrix $\boldsymbol{W} \in C^{M\times N}$ must be designed. Relevant matrices are obtained as follows.
1. Gaussian Matrices: The entries of $\boldsymbol{W}$ are chosen as independent-identically distributed Gaussian random variables with expectation 0 and variance $1/M$.
2. Bernoulli Matrices: The entries of Bernoulli matrices are independent realizations of $\pm 1/\sqrt{M}$ Bernoulli random variables.

III. AUXILIARY FUNCTION METHOD

For a given objective function, auxiliary function (AF) method [6] iteratively decreases an appropriately designed objective function. AF method does not always apply to any optimization problem, but it produces an effective optimization algorithm if we design auxiliary functions satisfying a condition. In the following we outline a discipline of AF method and derive a condition satisfied by the auxiliary function.

For a objective function $L(\boldsymbol{s})$, if the relation

$$L(\boldsymbol{s}) = \min_{\boldsymbol{a}} L^{+}(\boldsymbol{s}, \boldsymbol{a}), \tag{8}$$

is true, we call $L^{+}(\boldsymbol{s}, \boldsymbol{a})$ the auxiliary function of $L(\boldsymbol{s})$ and $\boldsymbol{a}$ the corresponding auxiliary variables. The next theorem yields the discipline for AF method.

Theorem 3. The objective function $L(\boldsymbol{s})$ monotonically decreases by the iteratively alternating minimization of the auxiliary function $L^{+}(\boldsymbol{s},\boldsymbol{a})$, satisfying (8), in respect to $\boldsymbol{a}$ followed by that in respect to $\boldsymbol{s}$.

When $L^{+}(\boldsymbol{s}, \boldsymbol{a})$ is quadratic in $\boldsymbol{s}$ and the optimum $\boldsymbol{a}$ is analytically computable, the above iterative optimization becomes very efficient.

A. Auxiliary function for $L(x)=|x|^p$

In order to design the auxiliary function for l_p minimization of CS, let us consider the case of one variable objective

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

function $L(x)=|x|^p$ $(0<p<2)$. We at first find quadratic functions $L^+(x, a)=x^2/(2a)+b(a)$ not smaller than $L(x)=|x|^p$ and tangent to $L(x)$ as shown in Fig. 2. Such quadratic functions are given as

$$L^+(x,a)=\frac{x^2}{2a}+\left(1-\frac{p}{2}\right)(pa)^{p/(2-p)},\quad 0<a,\ 0<p<2. \tag{9}$$

To certify that (9) satisfies the condition given by (8), take the derivative of $L^+(x,a)$ with respect to a.

$$\begin{aligned}\frac{\partial L^+(x,a)}{\partial a}&=-\frac{x^2}{2a^2}+\frac{p^2}{2}(pa)^{2(p-1)/(2-p)}\\&=-\frac{x^2-(pa)^{2/(2-p)}}{2a^2}.\end{aligned} \tag{10}$$

Thus, $L^+(x, a)$ takes the minimum value with respect to a at $a=|x|^{2-p}/p$. The minimum value is

$$L^+\left(x,\frac{x^{2-p}}{p}\right)=\frac{p}{2}|x|^p+\left(1-\frac{p}{2}\right)|x|^p=|x|^p. \tag{11}$$

Therefore, we see $L^+(x,a)$ given by (9) is the auxiliary function of $L(x)=|x|^p$.

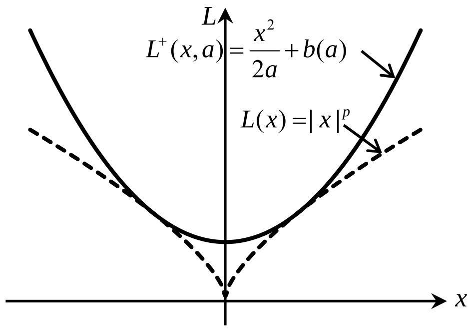

Fig. 2. Illustration for the auxiliary function $L^+(x,a)$ of objective function $L(x)=|x|^p$. The auxiliary variable a is determined so that the quadratic auxiliary function $L^+(x,a)$ is tangent to the original objective function $L(x)$.

B. Auxiliary function for $L(\boldsymbol{x})=\|\boldsymbol{x}\|_p^p$

In the multivariate case, the objective function is l_p norm of $\boldsymbol{x}$, i.e. $L(\boldsymbol{x})=\|\boldsymbol{x}\|_p^p=\sum_n|x_n|^p$. To the variable x_n $(n=1,\cdots,N)$ we can respectively introduce an auxiliary variable a_n to write down its auxiliary function as

$$L^+(\boldsymbol{x},\boldsymbol{a})=\frac{1}{2}\sum\nolimits_n\frac{x_n^2}{a_n}+\sum\nolimits_n\left(1-\frac{p}{2}\right)(pa_n)^{p/(2-p)},\quad 0<a_n,\ 0<p<2. \tag{12}$$

More simply, we have in vector notation,

$$L^+(\boldsymbol{x},\boldsymbol{a})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\mathrm{diag}^{-1}(\boldsymbol{a})\boldsymbol{x}+\mathbf{1}^{\mathrm{T}}\gamma(\boldsymbol{a}), \tag{13}$$

where diag($\boldsymbol{a}$) is the $N\times N$ diagonal matrix whose diagonal components are $[a_1,a_2,\cdots,a_N]$, $\cdot^{-1}$ shows the matrix inversion, and

$$\gamma(\boldsymbol{a})=\left(1-\frac{p}{2}\right)[(pa_1)^{p/(2-p)},\cdots,(pa_N)^{p/(2-p)}]^{\mathrm{T}},$$

and $$\mathbf{1}=[1,1,\cdots,1]^{\mathrm{T}}.$$

In (12), when $\boldsymbol{a}$ is constant vector, $L^+(\boldsymbol{x},\boldsymbol{a})$ is quadratic for $\boldsymbol{x}$ and thus the optimum $\boldsymbol{x}$ is uniquely determined. On the other hand, when $\boldsymbol{x}$ is constant the first term decreases and the second one increases as a_n increases, and thus the optimum $\boldsymbol{a}$ are uniquely determined depending on initial values of $\boldsymbol{x}$. Therefore, iterative minimization uniquely determines the optimum $\boldsymbol{x}$ near to the given initial vector $\boldsymbol{x}$.

The constraint l_p norm minimization, Eq.(5), can be re-written using auxiliary function as

$$\min_{\boldsymbol{x}\in R^n,\boldsymbol{a}\in R^{+n}}\left\{\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\mathrm{diag}^{-1}(\boldsymbol{a})\boldsymbol{x}+\mathbf{1}^{\mathrm{T}}\gamma(\boldsymbol{a})\right\}\quad \text{subject to } \boldsymbol{W}\boldsymbol{x}=\boldsymbol{y}. \tag{14}$$

Using the Lagrange multiplies, the above problem can be translated to

$$\min_{\boldsymbol{x}\in R^n,\boldsymbol{a}\in R^{+n},\boldsymbol{\lambda}\in R^n} J,$$
$$J=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\mathrm{diag}^{-1}(\boldsymbol{a})\boldsymbol{x}+\mathbf{1}^{\mathrm{T}}\gamma(\boldsymbol{a})+\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}), \tag{15}$$

where $\boldsymbol{\lambda}$ is the $M\times 1$ Lagrange multiplier vector. When $1\le p<2$, the above l_p norm minimization problem is convex but when $0\le p<1$, it is not. Therefore the local optimum solution is obtained. By putting the partial derivatives of J by $\boldsymbol{x}$, $\boldsymbol{a}$ and $\boldsymbol{\lambda}$ to zeros, each of which respectively leads to

$$\boldsymbol{x}=\mathrm{diag}(\boldsymbol{a})\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\lambda}, \tag{16a}$$

$$\boldsymbol{a}=\frac{|\boldsymbol{x}|^{2-p}}{p}, \tag{16b}$$

and $$\boldsymbol{W}\boldsymbol{x}=\boldsymbol{y}. \tag{16c}$$

It is noted that $|\cdot|^{2-p}$ is a component-wise operator. Substituting (16a) to (16c) and rearranging the order, we lastly have

$$[\boldsymbol{W}\mathrm{diag}(\boldsymbol{a})\boldsymbol{W}^{\mathrm{T}}]\boldsymbol{\lambda}=\boldsymbol{y}, \tag{17a}$$

$$\boldsymbol{x}=\mathrm{diag}(\boldsymbol{a})\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\lambda}, \tag{17b}$$

and $$\boldsymbol{a}=\frac{1}{p}|\boldsymbol{x}|^{2-p}. \tag{17c}$$

The minimization is performed iteratively as follows. At first, we initialize $\boldsymbol{a}=\mathbf{1}$, which corresponds to the l_2 norm minimization in (5) with $p=2$. Then we solve $\boldsymbol{\lambda}$ to the simultaneous equations (17a) and substitute it for (17b) to get the corresponding vector $\boldsymbol{x}$. From this $\boldsymbol{x}$, we update coefficient vector $\boldsymbol{a}$ by (17c). This cycle is repeated until the root mean square differences of $\boldsymbol{x}$'s between succeeding stages becomes less than ε.

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

In (13) $\mathrm{diag}^{-1}(\boldsymbol{a})$ describes the weighs of $\boldsymbol{x}$'s components, so if the n-th component of $\boldsymbol{x}$ in some iteration stage were very small compared to other components, the evaluation weigh, $\mathrm{diag}^{-1}(a_n)=p/|x_n|^{2-p}$, to this component will come to front and x_n will be much smaller in the succeeding iterations. In this way, the sparse positions (zero components) are explored. The smaller p, the greater the relative difference of weights is. Therefore we can conclude the convergence of the above iteration algorithm will be faster as p becomes smaller.

IV. Simulations

We executed simulations of the above-mentioned l_p norm reconstruction. The simulations are done for various values of p and compared them to the l_1 simplex method. We supposed the dimension of latent signals is $N=128$, and observed them by M linear observations. The latent sparse signals $\boldsymbol{x}$ are randomly generated 1000,000 times with sparsity K=32. The measurement matrices $\boldsymbol{W}$ are generated by arrangement of i.i.d. Gaussian random numbers with zero mean and unit variance. The convergence decision constant was chosen $\varepsilon=10^{-3}$. We measured the success rates of perfect reconstructions as we change M from 64 to 128.

Fig. 3 shows the relation of success rate of perfect reconstruction to the number of measurements. This shows the

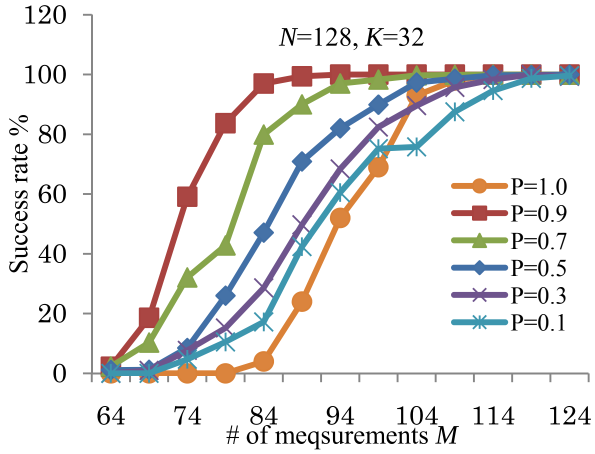

Fig.3. Success rate v.s. the number of measurements. Success rate of perfect reconstruction increases as the number of measurements increases. The curve of $p=0.9$ seems to have good performance ; perfect reconstruction occurs even with the least number of measurement.

perfect reconstruction is occurred even by the smaller number of measurements when $p=0.9$ than when $p=1.0$. The more precise simulations exhibited the least measurement number M=69 (50% success rate) occurs when $p=0.95$. Comparing with $M=94$ of $p=1.00$, about 68 % of measurements will ensure perfect reconstruction in the $l_{0.95}$ norm minimization.

Table 1 shows the comparison of processing time and success rate of perfect reconstruction between the proposed algorithm of $l_{0.95}$ norm minimization and simplex method of l_1 norm minimization. The comparison was done for measurements of 50 random signals of length $N=64$. We compared the cases of (K,M)=(16,50), (16,40) and (8,15), and the experiments were performed 50 times with different random signals and obtained the average processing time of perfect reconstruction. The programs were written in C language (compiler: gcc) and executed on CPU Core i7 (2.8GHz) with 4.00GB main memory.

From Table 1, it is concluded in our method the perfect reconstruction ratio 100 % but in l_1 simplex method the rates are not full mark. Therefore $l_{0.95}$ norm minimization is robust

Table 1 Comparison of average processing time between $l_{0.95}$ norm minimization and simplex method of l_1 norm minimization (N=64 and 50 times repetition).

Condition K	Condition M	Comparison item	$l_{0.95}$ minimization	l_1 simplex method
16	50	time (msec)	26	163
		PR rate (%)	100	96
16	40	time (msec)	35	538
		PR rate (%)	100	88
8	15	time (msec)	38	18
		PR rate (%)	100	86

to the number of measurements. In addition, processing times of our method are respectively 16 % and 6.5 % of those of l_1 simplex method when M is 50 and 40. But when both K and M are very small, l_1 norm simplex method is a little faster.

V. Conclusions

This paper proposes a fast l_p norm minimization method for compressed sensing. This method is based on auxiliary function method and the optimization is performed in iterative reweighted least squares. We showed the unique solution is obtained depending on initial values by our method. For very small p, l_2 norm solution may not be a suitable initial vector. Simulation showed $l_{0.95}$ norm minimization method is almost always faster than l_1 norm simplex method.

Future work will be to explain the reason why the robustness decreases for small K like 0.1. Also, we want to apply our algorithm to X-ray CT reconstruction.

References

[1] H.Tanaka, "Mathematics of compressed sensing (in Japanese)," *IEICE Fundamentals Review*, vol.4, no.1, pp.39-47, 2010.
[2] D.L.Donho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol.52, no.4, pp.1289-1306, 2006.
[3] E.J.Candes and T.Tao, "The restricted isometry property and its implications for compressed sensing," *C.R.Acad.Sci.I, Math.*, vol.346, nos.9-10, pp.589-592, 2008.
[4] E.J.Candes and T.Tao, "Decoding by liner programing," *IEEE Trans. Inf. Theory*, vol.51, no.12, pp.4203-4215, 2005.
[5] R.Chartrand and V.Stqneva, "Restricted isometry properties and nonconvex compressed sensing," *Inverse Problems*, vol.24, no.035020, pp.1-14, 2008.
[6] H.Kameoka, Y.Kamamoto, N.Harada and K. Moritani, "A linear predictive coding algorithm minimizing the Golomb-Rice code length of the residual signal (in Japanese)," *IEICE Trans. Fundamentals*, vol.J91-A, no.11, pp.1017-1025, 2008.
[7] R.Hogg, J.McKean and A.Craig, " Introduction to Mathematical Statistics," Pearson Prentice Hall, 2005.
[8] Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. ," IEEE Trans. Signal Process., vol.45, no.3, pp.600–616, 1997.