Representative of $L_{1/2}$ Regularization among L_q $(0<q\leq 1)$ Regularizations: an Experimental Study Based on Phase Diagram

XU Zong-Ben[1] GUO Hai-Liang[1] WANG Yao[1] ZHANG Hai[1,2]

Abstract Recently, regularization methods have attracted increasing attention. L_q $(0 < q < 1)$ regularizations were proposed after L_1 regularization for better solution of sparsity problems. A natural question is which is the best choice among L_q regularizations with all q in $(0, 1)$? By taking phase diagram studies with a set of experiments implemented on signal recovery and error correction problems, we show the following: 1) As the value of q decreases, the L_q regularization generates sparser solution. 2) When $1/2 \leq q < 1$, the $L_{1/2}$ regularization always yields the best sparse solution and when $0 < q \leq 1/2$, the performance of the regularizatons takes no significant difference. Accordingly, we conclude that the $L_{1/2}$ regularization can be taken as a representative of L_q $(0 < q < 1)$ regularizations.

Key words L_q regularization, phase diagram, signal recovery, error correction

Citation Xu Zong-Ben, Guo Hai-Liang, Wang Yao, Zhang Hai. Representative of $L_{1/2}$ regularization among L_q $(0 < q \leq 1)$ regularizations: an experimental study based on phase diagram. *Acta Automatica Sinica*, 2012, **38**(7): 1225−1228

DOI 10.3724/SP.J.1004.2012.01225

Recently, considerable attention has been paid to the following sparsity problem. We are given an $n \times N$ matrix Φ which is in some sense "random", for example, a matrix with i.i.d Gaussian entries, and we are also given an n-vector $\boldsymbol{y}$ and know that $\boldsymbol{y} = \Phi\boldsymbol{x}_0$ where $\boldsymbol{x}_0 \in \mathbf{R}^N$ is an unknown sparse vector. We expect to recover $\boldsymbol{x}_0$. However, $n \ll N$, the system of equations is underdetermined and hence, it is not a properly-posed problem in linear algebra. Nevertheless, sparsity of x_0 is a very useful priority that sometimes allows unique solution. Accordingly, one naturally proposes to use the following optimization model (P_0) to obtain the sparsest solutions

$$(P_0) \qquad \min_{\boldsymbol{x}\in\mathbf{R}^N} \|\boldsymbol{x}\|_0 \ \text{ s. t. } \boldsymbol{y} = \Phi\boldsymbol{x} \tag{1}$$

where $\|\boldsymbol{x}\|_0 = |\{i : x_i \neq 0\}|$. This is of little practical use, however, since the problem (P_0) is combinatorial in feature and generally difficult to be solved as its solution requires an intractable combinatorial search[1].

To solve this problem, the subsequent (P_1) optimization problem was suggested[2], which then can be transformed into a linear programming problem:

$$(P_1) \qquad \min_{\boldsymbol{x}\in\mathbf{R}^N} \|\boldsymbol{x}\|_1$$

$$\text{s. t.} \qquad \boldsymbol{y} = \Phi\boldsymbol{x} \tag{2}$$

where $\|\boldsymbol{x}\|_1 = \sum_{i=1}^N |x_i|$. We call (P_0) as L_0 regularization and (P_1) as L_1 regularization. The use of L_1 regularization has become so widespread that it has been arguably considered as the "modern least squares"[3]. However, the solutions of the L_1 regularization are often not as sparse as those of the L_0 regularization. To find solutions more sparse than L_1 regularization is definitely imperative and required for many applications. A natural try for this purpose is to apply the L_q $(0 < q < 1)$ regularization, that is, to solve the following (P_q) model,

$$(P_q) \qquad \min_{\boldsymbol{x}\in\mathbf{R}^N} \|\boldsymbol{x}\|_q$$

$$\text{s.t.} \qquad \boldsymbol{y} = \Phi\boldsymbol{x} \tag{3}$$

or equivalently,

$$\min_{\boldsymbol{x}\in\mathbf{R}^N} \{\|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_q^q\} \tag{4}$$

where $\|\boldsymbol{x}\|_q = \sum_{i=1}^N |x_i|_q^q$, and λ is a regularization parameter. Obviously, the $L_q\,(0 < q < 1)$ model is no longer a convex optimization problem, and thus we can only get the local optimal solutions in most cases, yet it can yield solutions sparser than the L_1 regularization model[4−5].

There are many choices for q when the (P_q) model is adopted. A nature and also crucial question is: which q is the best among L_q $(0 < q < 1)$ regularizations? In this paper, our aim is to provide an affirmative answer to this question through an experimental study with phase diagram. We will comparatively apply the L_q $(0 < q \leq 1)$ regularizations, according to the phase diagram requirement, to several typical sparsity problems: compressive sensing and error correction, and then, we will conclude from the resultant phase diagrams that $L_{1/2}$ regularization can be taken as a representative of L_q $(0 < q < 1)$ regularizations. This study offers a solid evidence to support the speciality and importance of $L_{1/2}$ regularization.

1 Experimental methods and test problems

1.1 Experimental methods

For an underdetermined system of linear equations $\boldsymbol{y} = \Phi\boldsymbol{x}$, when the model (P_0) has a unique sparse solution (it is then also the unique solution of (P_q)) and the solution can be obtained from the L_q $(0 < q \leq 1)$ regularization procedure, we say that the L_q and L_0 regularizations are equivalent, or briefly, of L_0/L_q equivalence. When a vector is not only a solution of (P_0) but also a solution of (L_q) problem, it is said to be a point of L_0/L_q equivalence.

In the context of L_0/L_1 equivalence, Donoho[6−7] introduced the notion of phase diagram to illustrate how sparsity (number of nonzeros in $\boldsymbol{x}$/number of rows in Φ) and indeterminacy (number of rows in Φ/number of columns in Φ) affect the success of L_1 regularization. Using the technique of high-dimensional geometry analysis, Donoho[6] provided a necessary and sufficient condition on a particular random matrix Φ of size $n \times N$ such that every $\boldsymbol{x} \in \chi^N(k)$ is a point of L_1/L_0 equivalence, where $\chi^N(k) = \{\boldsymbol{x} \in \mathbf{R}^N : \|\boldsymbol{x}\|_0 \leq k\}$. The performance exhibits two phases (success/failure) in a diagram, as shown in Fig. 1[7]. Each point on the plot of the figure corresponds to a statistical model for certain values of n, N, and k. The abscissa runs from 0 to 1, and gives values for $\delta = n/N$. The ordinate is $\rho = k/n$, measuring

Manuscript received March 16, 2011; accepted June 22, 2011
Supported by National Basic Research Program of China (973 Program) (2007CB311002) and National Natural Science Foundation of China (60975036)
Recommended by Associate Editor LIU Yi-Jun
1. Institute of Information and System Science, Xi′an Jiaotong University, Xi′an 710049, P. R. China 2. Department of Mathematics, Northwest University, Xi′an 710069, P. R. China

the level of sparsity in the model. Above the plotted phase transition curve, the L_1 method fails to find the sparsest solution; below the curve, the solution of (P_1) is the precise solution of (P_0).

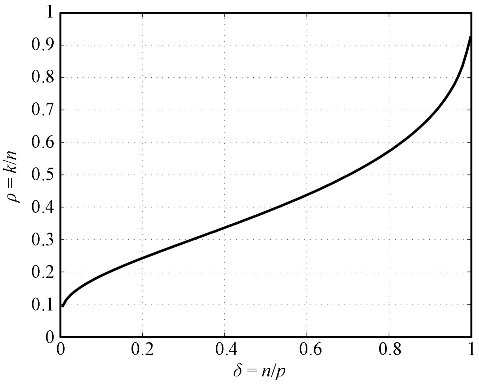

Fig. 1 Theoretical phase transition diagram: the theoretical threshold at which equivalence of the solutions to the L_1 and L_0 optimization problems breaks down (Along the x-axis the level of underdeterminedness decreases, and along the y-axis the level of sparsity of the underlying model increases.)

Donoho et al.[8] conducted a series of simulation experiments for the problem of variable selection when the number of variables exceeds the number of observations. They have defined a problem suite $S\{k, n, N\}$ as a collection of problems with sparse solutions, and each problem has an $n \times N$ model matrix Φ and a k-sparse N-vector of coefficients $\boldsymbol{x}$. We will follow the method of Donoho et al.[8] in this paper. For each k, n, N combination we run an algorithm in question multiple times, and measure its success according to a quantitative criterion. We choose the relative root square error (RRSE) RRSE $= \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2/\|\boldsymbol{x}\|_2$ as the quantization criterion, and the results are then compared across models with different problem sizes.

The following recipes are employed to study an algorithm for L_q $(0 < q \leq 1)$ regularization models:

1) Generate a prototype model $\boldsymbol{y} = \Phi\boldsymbol{x}$, which has a k-sparse solution, where $k < N$.

2) Run an algorithm of L_q regularization to obtain a reconstructed solution $\hat{\boldsymbol{x}}$.

3) Evaluate performance to test if $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2/\|\boldsymbol{x}\|_2 \leq \gamma$, where γ is a tolerance bound set in advance.

After getting all the RRSEs of the problem suite $S\{k, n, N\}$, a phase diagram can be drew for the tasted algorithm. With such methodology, we will present the experiments in the next section. However, we need to first introduce two typical test problems.

1.2 Test problems

1.2.1 Signal recovery

Assume that $\boldsymbol{x}$ is a signal with k nozero spikes, and that Ψ is a unit matrix, i.e., the canonical basis is used to denote the signal. We attempt to reconstruct $\boldsymbol{x}$ through random sampling Φ by (P_1) and (P_q) models and compare the performance of L_q regularizations when q varies from 0 to 1.

Many researchers have suggested methods to solve the L_q $(0 < q \leq 1)$ regularization problems. Since L_1 regularization is convex and the others are not, there are many exclusive efficient methods for L_1 regularization, while very few for L_q $(0 < q < 1)$ regularizations. We briefly introduce the methods applied in the experiments below.

For the L_1 regularization problem, we will use the basis pursuit method suggested by Donoho et al.[2], which is based on a linear program solver.

For the L_q regularization problems, we will apply the reweighted L_1 method proposed by Xu et al.[9], which transforms the L_q problems into a series of convex weighted L_1 regularization problems, to which the existing L_1 regularization algorithms can be efficiently applied[10−11].

1.2.2 Error correction

Suppose that a response variable $\boldsymbol{y}$ is dependent on independent variables $A_1, \cdots, A_p$, and that we are in a classically designed experiment, with $p < n$. If the dependence is linear, we then have $\boldsymbol{y} = \sum_j x_j A_j + \boldsymbol{e}$, where the error $\boldsymbol{e}$ is a "wild" variable containing occasionally very large outliers. We assume that the outlier generators $\boldsymbol{e}$ have most entries of 0, with k being the number of nozero entries, but we know neither which entries are affected nor how they are affected. We would like to recover the information $\boldsymbol{x}$ exactly from the corrupted N-dimensional vector $\boldsymbol{y}$.

To decode, Candes et al.[12] proposed to use the subsequent (D_1) model to solve the error correction problem:

$$(D_1) \quad \min_{\tilde{\boldsymbol{x}} \in \mathbf{R}^p} \|\boldsymbol{y} - A\tilde{\boldsymbol{x}}\|_1 \tag{5}$$

which can also be recast as an linear programming (LP) problem. They showed that if A is chosen suitably, the solution of (D_1) model can correctly retrieve the information $\boldsymbol{x}$ without error provided that $\boldsymbol{e}$ is sparse enough.

Also, Chartrand[13] suggested to solve the error correction problem by L_q $(0 < q < 1)$ minimization:

$$(D_q) \quad \min_{\tilde{\boldsymbol{x}} \in \mathbf{R}^p} \|\boldsymbol{y} - A\tilde{\boldsymbol{x}}\|_q \tag{6}$$

For the L_1 minimization, we use the corresponding Matlab program in L_1-magic software, which is available on Candes' web page[14]. For L_q $(0 < q < 1)$ minimization, we slightly adjust the reweighted L_1 algorithm proposed in [9] to solve the (D_q) model.

2 Experiment results

2.1 Signal recovery

To apply the phase diagram methodology, we fix the length of signal, $N = 512$, and then build a prototype model for a certain level of underdeterminedness $\delta \equiv n/N$ for $\delta \in [0, 1]$, and a sparsity level $\rho \equiv k/n$ for $\rho \in [0, 1]$. Then, a problem suite is constructed through varying sampling number n and sparsity k. The performance of the regularization algorithms is then evaluated over this grid in a systematic way. For each k, n, N combination, the procedure of study is as follows:

Algorithm 1.

Step 1. Generate $A_{n\times N}$ with $A_{ij} \in \mathrm{N}(0, 1)$, and create $\boldsymbol{y} = A\boldsymbol{x}$ where $\boldsymbol{x}$ has k nonzeros drawn from $\mathrm{N}(0, 1)$.

Step 2. Run a regularization method, and get the reconstructed solution $\hat{\boldsymbol{x}}$.

Step 3. Evaluate the "success/failure" property: if the relative root square error (RRSE) is smaller than 10^{-5}, the recovery is considered success, or failure otherwise.

Step 4. Repeat Steps $1 \sim 3$ for 50 times, and evaluate the frequency of success/failure recovery.

After getting the "success/failure frequency" of all the defined problem suite $S\{k, n, N\}$, we can plot it on the phase plane (δ, ρ), where $\delta = n/p$ and $\rho = k/n$ (Fig. 2). The contours indicate the success rate, where light gray (above the belt curve) means the success rate of this combination of $\{k, n, N\}$ is 0 %, and dark gray (below the belt curve) means the success rate of this combination of $\{k, n, N\}$ is 100 %. We can also find the belt area with other color in Fig. 2, which means that the success rate is between 0 % and 100 %. Fig. 2 shows the recovery performances of $L_{0.1}$, $L_{0.3}$, $L_{0.5}$, $L_{0.7}$, $L_{0.9}$, and $L_{1.0}$ separately, in which the thin curve is the theoretical L_1/L_0 equivalent curve, proved by Donoho[6−7]. After comparing the phase diagrams in Fig. 2, we can obtain the following conclusions:

1) The phase transition phenomenon occurs very clearly, with the belt area displaying a phase transition curve for any L_q regularizations.

2) The L_q phase transition curves are almost all above the theoretical L_1/L_0 equivalent curve, showing that L_q regularizations have stronger sparsity promoting ability than L_1 regularization.

3) The smaller q, the better the performance, but when $q \in (0, 0.5)$, the difference is invisible.

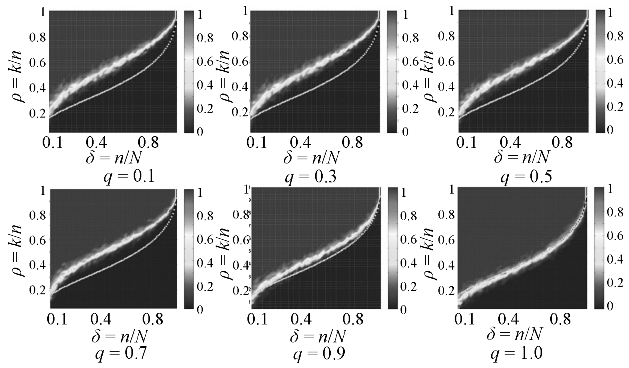

Fig. 2 Phase diagrams of L_q regularizations (q = 0.1, 0.3, 0.5, 0.7, 0.9, 1.0) when applied to signal recovery, where abscissa $\delta = n/N$ and ordinate $\rho = k/n$

To quantitatively compare the differences of behaviors of the L_q regularizations for different $q \in (0, 1)$, we have gotten more phase diagrams for different values of q and calculated the success percentage in all the suite for a regularization, that is, the ratio of the dark gray region in the whole region of the phase plane. The interpolated success percentage curve is depicted in Fig. 3, where the horizontal axis is the value of q, and the vertical axis is the percentage of successive restoration. From Fig. 3, we can observe the following:

1) The ratio changes very slowly between $q \in (0, 0.5)$.

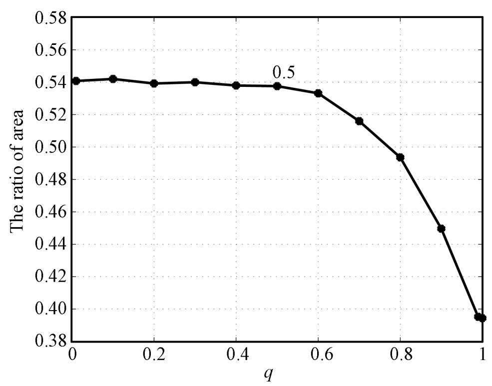

Fig. 3 Success recovery percentage of L_q regularizations when applied to signal recovery

2) The ratio changes rapidly when q increases from 0.5 to 1.

This reveals that $L_{1/2}$ regularization is significantly better than L_1 regularization, while it takes no significant difference from other L_q regularizations when $q \in (0, 0.5)$. Thus, in the sense of bringing the benefit of exact recovery, $L_{1/2}$ can be regarded as the best.

2.2 Error correction

Inspired by the phase transition experiment for signal recovery, we fix $n = 512$, and then build a prototype model for a fixed $\gamma = p/n$, $\gamma \in [0, 1]$, and a fixed $\epsilon = k/p$ for $\epsilon \in [0, 1]$. Almost exactly the same with the experiment of signal recovery, for each k, p, n combination, the procedure of study is given as follows.

Algorithm 2.

Step 1. Generate $A = (A_{ij}) \in \mathbf{R}^{n\times p}$ with $A_{ij} \sim \mathrm{N}(0, 1)$ and p-dimensional vector $\boldsymbol{x_0}$ with $x_i \sim \mathrm{N}(0, 1)$, and then create $\boldsymbol{y} = A\boldsymbol{x_0} + \boldsymbol{e}$ where $\boldsymbol{e}$ has k nonzeros drawn from $\mathrm{N}(0, 1)$.

Step 2. Run a regularization method, and get the solution $\hat{\boldsymbol{x}}$.

Step 3. Evaluate the "success/failure" property: if the relative root square error (RRSE) is smaller than 10^{-5}, the recovery is considered success, or failure otherwise.

Step 4. Repeat Steps 1 ∼ 3 for 50 times, and evaluate the "success/failure".

After getting the "success/failure" of all the defined problem suite $S\{k, p, n\}$, we plot it on the plane (γ, ϵ), where $\gamma = p/n$ and $\epsilon = k/p$. The contours indicate the success/failure rate, with light gray (above the belt curve) meaning the success of this combination of $\{k, n, p\}$ is 0 %, and dark gray (below the belt curve) means the success rate of this combination of $\{k, n, p\}$ is 100 %. In the figure, we can also find the belt area with other color, which means that the success rate is between 0 % and 100 %. Following Donoho and Tanner[15], we display the result in different ordinate systems with variables $\rho = k/(n - p)$ and $\delta = (n - p)/n$. Fig. 4 then shows the performances of $L_{0.1}$, $L_{0.3}$, $L_{0.5}$, $L_{0.7}$, $L_{0.9}$, $L_{1.0}$ regularization algorithms, in which the thin curve is the theoretical L_1/L_0 equivalent curve. After comparing the phase diagrams in Fig. 4, we find that all the phenomena observed in signal recovery application occur again in this error correction application.

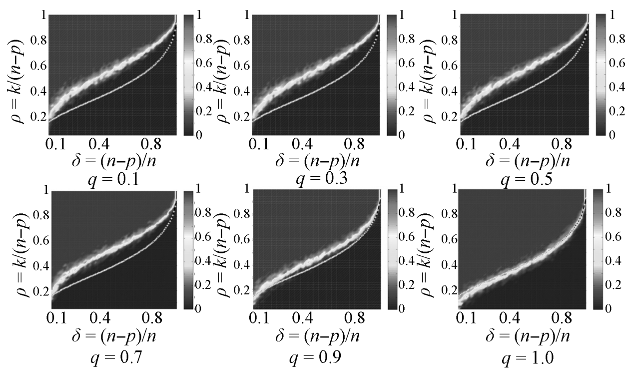

Fig. 4 Phase diagrams of L_q regularizations (q = 0.1, 0.3, 0.5, 0.7, 0.9, 1.0) when applied to error correction, where abscissa $\delta = (n - p)/n$ and ordinate $\rho = k/(n - p)$

To quantitatively compare the differences of behaviors of the L_q minimizations for different $q \in (0, 1)$, we have also gotten more phase diagrams for different values of q and calculated the success percentage in all the suite for a regularization algorithm, which is defined as the ratio of the dark gray region in the whole phase plane. The interpolated success rate curve is given in Fig. 5. These experiments for error correction reveal again that $L_{1/2}$ regularization is significantly better than L_1 regularization, while it takes no significant difference from other L_q regularization when $q \in (0, 0.5)$. Thus, in the sense of getting the benefit of exact recovery, $L_{1/2}$ is the best.

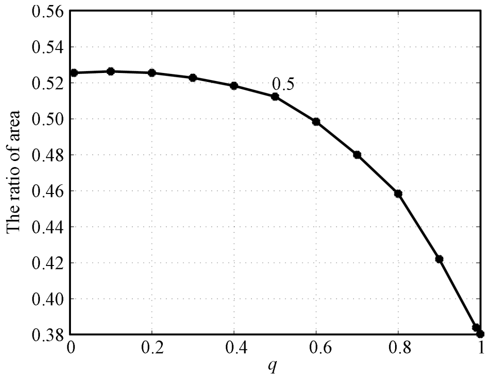

Fig. 5 Success recovery percentage of L_q regularizations when applied to error correction

3 Conclusion

With the phase diagram tool, we have conducted an experimental study on performance comparison between L_1 regularization and L_q $(0 < q < 1)$ regularizations for sparse signal recovery and error correction. The comparisons show that when $0 < q < 1$, the L_q regularizations can always yield sparser solutions than L_1 regularization, and that the smaller, the better the performance of L_q regularizations. Nevertheless, when $0 < q \leq 1/2$, the performance of L_q regularizations has no significant difference. This suggests that among the L_q regularizations with $0 < q \leq 1$, $L_{1/2}$ can be taken as a representative.

The study of this paper reveals the extreme importance and special role of $L_{1/2}$ regularization. It particularly leads to a guess or an expectation that the $L_{1/2}$ regularization might be more powerfully applied to sparsity problems. In a very recent research, a fast efficient iterative half thresholding algorithm was suggested for implementation of $L_{1/2}$ regularization.

References

1 Natarajan B K. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 1995, **24**(2): 227−234

2 Chen S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 1998, **20**(1): 33−61

3 Candes E, Wakin M, Boyd S. Enhancing sparsity by reweighted L_1 minimization. *Journal of Fourier Analysis and Applications*, 2008, **14**(5): 877−905

4 Chartrand R. Exact reconstructions of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 2007, **14**(10): 707−710

5 Chartrand R, Staneva V. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 2008, **24**(3): 1−14

6 Donoho D L. Neighborly polytopes and the sparse solution of underdetermined linear equations [Online], available: http: // www-stat.stanford.edu/ ∼donoho/ Reports/ 2005/ NPaSSULE-01-28-05.pdf, November 8, 2011

7 Donoho D L. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete and Computational Geometry*, 2006, **35**(4): 617−652

8 Donoho D L, Stodden V. Breakdown point of model selection when the number of variables exceeds the number of observations. In: Proceedings of the International Joint Conference on Neural Networks. Vancouver, USA: IEEE, 2006. 1916−1921

9 Xu Z B, Zhang H, Wang Y, Chang X Y, Yong L. $L_{1/2}$ regularization. *Science in China Series F: Information Sciences*, 2010, 53(6): 1159−1169

10 Chen X, Xu F M, Ye Y. Lower bound theory of nonzero entries in solutions of L_2-L_p minimization. *SIAM Journal on Scientific Computing*, 2010, **32**(5): 2832−2852

11 Chartrand R, Yin W. Iteratively reweighted algorithms for compressive sensing. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Las Vegas, USA: IEEE, 2008. 3869−3872

12 Candes E, Tao T. Decoding by linear programming. *IEEE Transactions on Information Theory*, 2005, **51**(12): 4203−4215

13 Chartrand R. Nonconvex compressed sensing and error correction. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Honolulu, USA: IEEE, 2007. 889−892

14 Candes E, Caltech J R. L_1-MAGIC: recovery of sparse signals via convex programming [Online], available: http: //users.ece.gatech.edu/~justin/l1magic/downloads/l1magic.pdf, June 30, 2011

15 Donoho D L, Tanner J. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A*, 2009, **367**(1906): 4273−4293

XU Zong-Ben Professor at the Institute for Information and System Sciences, Faculty of Science, Xi′an Jiaotong University. His research interest covers machine learning, visual computation and image processing, and sparse modeling and optimization. E-mail: zbxu@mail.xjtu.edu.cn

GUO Hai-Liang Master student at the Institute for Information and System Sciences, Faculty of Science, Xi′an Jiaotong University. His research interest covers compressive sensing, sparse modeling of signals and images. Corresponding author of this paper. E-mail: nianlongguo@gmail.com

WANG Yao Ph. D. candidate at the Institute for Information and System Sciences, Faculty of Science, Xi′an Jiaotong University. His research interest covers gradient boosting and its application to variable selection, compressive sensing, and sparse modeling of signals and images.
E-mail: yao.s.wang@gmail.com

ZHANG Hai Associate professor in the Department of Mathematics, Northwest University and Ph. D. candidate at the Institute for Information and System Sciences, Faculty of Science, Xi′an Jiaotong University. His research interest covers stability of learning algorithms, learning theory, sparsity and regularization. E-mail: zhanghaiwu@tom.com