PHASE TRANSITION OF JOINT-SPARSE RECOVERY FROM MULTIPLE MEASUREMENTS VIA CONVEX OPTIMIZATION

Shih-Wei Hu[*,**], *Gang-Xuan Lin*[*], *Sung-Hsien Hsieh*[*], *and Chun-Shien Lu*[*]

[*] Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC
[**] Department of Comp. Sci. & Inf. Eng., National Taiwan University, Taipei, Taiwan, ROC

ABSTRACT

In sparse signal recovery of compressive sensing, the phase transition determines the edge, which separates successful recovery and failed recovery. Moreover, the width of phase transition determines the vague region, where sparse recovery is achieved in a probabilistic manner. Earlier works on phase transition analysis in either single measurement vector (SMV) or multiple measurement vectors (MMVs) is too strict or ideal to be satisfied in real world.

Recently, phase transition analysis based on conic geometry has been found to close the gap between theoretical analysis and practical recovery result for SMV. In this paper, we explore a rigorous analysis on phase transition of MMVs. Such an extension is not intuitive at all since we need to redefine the null space and descent cone, and evaluate the statistical dimension for $\ell_{2,1}$-norm. By presenting the necessary and sufficient condition of successful recovery from MMVs, we can have a boundary on the probability that the solution of a MMVs recovery problem by convex programming is successful or not. Our theoretical analysis is verified to accurately predict the practical phase transition diagram of MMVs.

1. INTRODUCTION

Compressive sensing (CS) [1, 2, 3] of sparse signals in achieving simultaneous data acquisition and compression has been extensively studied in the past few years. Let $s \in \mathbb{R}^N$ be an original signal to be sensed, let $\Phi \in \mathbb{R}^{M\times N}$ represent a sampling matrix, and let $y \in \mathbb{R}^M$ be the measurement vector. We say that s is k-sparse in the Ψ domain if it can represented as $s = \Psi x$, where x contains k non-zero entries. At the encoder, random projection, defined as:

$$y = \Phi s = \Phi\Psi x = Ax, \tag{1}$$

is conducted on s via Φ to obtain the measurement vector y, where $A = \Phi\Psi$. Under the circumstance, y is called a single measurement vector (SMV). The measurement rate in CS is defined as $0 < \frac{M}{N} \leq 1$. At the decoder, the original signal s can be recovered to a certain extent by means of convex optimization [4, 5] or greedy algorithms [6, 7].

An efficient convex optimization algorithm to solve sparse signal recovery in CS is ℓ_1-norm minimization, defined as:

$$\text{(SL1)} \quad \min \|x\|_1 \ \text{ s.t. } \ y = Ax,$$

where $\|\cdot\|_1$ is the ℓ_1-norm.

In fact, the solution to problem (SL1) heavily depends on the parameter set composed of N, M, and k. Thus, it is interesting in both theoretical and practical aspects to explore the sufficient conditions of successful sparse recovery.

1.1. Related Work

In the literature, only few researches target the goal of estimating the phase transition of sparse signal recovery in the context of compressive sensing. Donoho and Huo [8] show that when sparsity k is less than $\frac{1}{2}\left(1+\mu^{-1}\right)$, then problem (SL1) will have a unique solution, where μ is the mutual incoherence of A. Studer *et al.* [9] consider the model $z = Ax + By$ and find the sufficient condition of unique solution depends on the mutual incoherence between A and B. The aforementioned results, however, are too ideal to be satisfied in real world.

An innovative idea is proposed in [10] in that the authors analyze the problem (SL1) by conic geometry. They find that the intersection between ℓ_1-ball and the set of solutions of $y = Ax$ can be seen as the intersection between descent cone of ℓ_1-norm of x and null space of A. Moreover, the probability of intersection can be predicted by kinematic formula [11], and the probability is dependent on a novel summary parameter, called statistical dimension. The authors show that the statistical dimension can predict the probability of successful recovery of problem (SL1) almost perfectly.

1.2. Our Contributions

The prior researches discussed so far are almost for the single measurement vector. In some cases (*e.g.* [12]), however, the available information may contain multiple measurements vectors (MMVs) [13, 14, 15, 16, 17]. In this paper, we theoretically analyze the phase transition behavior of sparse signal recovery from MMVs via conic geometry and verify its practicability via simulations.

2. PROBLEM DEFINITION

2.1. Notations

For the matrix H, we denote its transpose by H^T; its ith row by h^i; its jth column by h_j; and the jth entry of ith row by h^i_j.supp(h) for a vector h is a set that collects the indices of nonzero entries of h and supp(H) for a matrix H is equivalent

to $\bigcup_j \text{supp}(h_j)$. $\|\cdot\|_p$ denotes the ℓ_p-norm and the Frobenius norm of matrix H is denoted by $\|H\|_F$.

2.2. Phase Transition of Sparse Signal Recovery from Multiple Measurement Vectors

The objective of this paper is to study the behavior of phase transition of sparse signal recovery in convex optimization problem of multiple measurement vectors (MMVs) in the context of compressive sensing. Let $S = [s_1, s_2, ..., s_L] \in \mathbb{R}^{N\times L}$ be the set of L (> 1) original signals to be sensed and let the set of measurement vectors be $Y = [y_1, y_2, ..., y_L] \in \mathbb{R}^{M\times L}$, where $y_i = \Phi s_i$, $i = 1, 2, ..., L$, as in Eq. (1). We also let $s_i = \Psi x_i$ and denote $X = [x_1, x_2, ..., x_L] \in \mathbb{R}^{N\times L}$. A matrix X is called K joint sparse if $|\text{supp}(X)| = K$. For matrices, we define the $\ell_{p,q}$-norms as:

$$\|X\|_{p,q} = \|(\|x^i\|_p)_{N\times 1}\|_q.$$

In contrast to SMV, given a dictionary A, sparse recovery from MMVs can be efficiently solved via $\ell_{2,1}$-norm minimization as:

$$\text{(ML1)} \quad \min \|X\|_{2,1} \ \text{ s.t. } \ Y = AX.$$

We call the problem (ML1) succeed if it has a unique optimal solution. The sufficient condition of solving problem (ML1), similar to that of SMV, is the key to evaluate successful recovery. Nevertheless, only few researches target this problem. In [14], the authors prove that when $K < \frac{1}{2}\left(1 + \mu^{-1}\right)$, the recovered result is the unique solution to problem (ML1) where μ denotes the mutual incoherence of A. The goal of this paper is to provide theoretical but practical bound of the probability of successful recovery for problem (ML1).

3. PRELIMINARY–CONIC GEOMETRY

We briefly introduce how a convex function can be specified in terms of conic geometry to make this paper self-contained.

Definition 3.1. (Descent cone [10])
The descent cone $\mathcal{D}(f, x)$ of a function $f : \mathbb{R}^N \to \overline{\mathbb{R}}$ at a point $x \in \mathbb{R}^N$, defined as:

$$\mathcal{D}(f, x) := \bigcup_{\tau>0} \{u \in \mathbb{R}^N : f(x + \tau u) \le f(x)\},$$

is the conical hull of the perturbations that do not increase f near x.

Since cone is not a linear subspace, it doesn't have the rank to evaluate its "size" in $\mathbb{R}^{N\times L}$. For this, a new measure is required.

Definition 3.2. (Statistical Dimension [10])
The statistical dimension (S.D.) $\delta(C)$ of a closed convex cone $C \subset \mathbb{R}^N$ is defined as:

$$\delta(C) := \mathbb{E}\left[\left\|\prod(g, C)\right\|_2^2\right],$$

where $g \in \mathbb{R}^N$ is a standard normal vector, $\|\cdot\|_2$ is ℓ_2-norm, and $\prod(\cdot, C)$, denoting the Euclidean projection onto C, is defined as: $\prod(x, C) := \arg\min\{\|x - y\|_2 : y \in C\}$.

In particular, if C is a subspace, $\delta(C) = \dim(C)$. With S.D., we can predict the probability of intersection between two cones by the following theorem.

Theorem 3.3. (Approximate kinematic formula [10])
Fix a tolerance $\eta \in (0, 1)$. Suppose that $C_1,\ C_2 \subset \mathbb{R}^N$ are closed convex cones, one of which is not a subspace. Draw an orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ uniformly at random. Then

$$\delta(C_1)+\delta(C_2) \le N-a_\eta\sqrt{N} \Rightarrow \mathbb{P}\{C_1\cap \boldsymbol{Q}C_2 = \{\boldsymbol{0}\}\} \ge 1-\eta,$$

$$\delta(C_1)+\delta(C_2) \ge N+a_\eta\sqrt{N} \Rightarrow \mathbb{P}\{C_1\cap \boldsymbol{Q}C_2 = \{\boldsymbol{0}\}\} \le \eta.$$

The quantity $a_\eta := \sqrt{8\log(4/\eta)}$.

4. MAIN RESULTS

In this section, we characterize when problem (ML1) succeeds and derive the phase transition of success rate of itself inspired by the framework of conic geometry. In order to satisfy the requirement of Theorem 3.3, A must be a Gaussian random matrix. In compressive sensing, Φ and Ψ are conventionally used to set as a Gaussian random matrix and orthonormal basis, respectively, so that $A = \Phi\Psi$ is also a Gaussian random matrix [18].

In our method, we first define the matrix null space $\text{null}(A, L)$ of matrix $A \in \mathbb{R}^{M\times N}$ as:

$$\text{null}(A, L) := \{E \in \mathbb{R}^{N\times L} : AE = \mathbf{0}_{M\times L}\}.$$

Second, the descent cone of $\ell_{2,1}$-norm $\mathcal{D}(\|\cdot\|_{2,1}, X_0)$ at the point X_0 is defined as:

$$\mathcal{D}(\|\cdot\|_{2,1}, X_0) := \bigcup_{\tau>0} \{U \in \mathbb{R}^{N\times L} : \|X_0+\tau U\|_{2,1} \le \|X_0\|_{2,1}\}.$$

The necessary and sufficient condition of the success of problem (ML1) is described in the following.

Theorem 4.1. (Optimality condition for MMVs recovery)
The matrix X_0 is the unique optimal solution to problem (ML1) if and only if $\mathcal{D}(\|\cdot\|_{2,1}, X_0) \cap \text{null}(A, L) = \{\boldsymbol{0}_{N\times L}\}$.

Proof. Assume X_0 is the unique optimal solution to problem (ML1). Given a matrix $E \in \mathcal{D}(\|\cdot\|_{2,1}, X_0) \cap \text{null}(A, L)$, we know that $X_0 + E$ is a feasible point of problem (ML1) and $\|X_0 + E\|_{2,1} \le \|X_0\|_{2,1}$, which implies that $X_0 + E$ is an optimal solution to problem (ML1). According to the uniqueness of optimal solution of problem (ML1), we have $E = 0$, and thus $\mathcal{D}(\|\cdot\|_{2,1}, X_0) \cap \text{null}(A, L) = \{\mathbf{0}_{N\times L}\}$.

Conversely, suppose $\mathcal{D}(\|\cdot\|_{2,1}, X_0) \cap \text{null}(A, L) = \{\mathbf{0}_{N\times L}\}$. Since we know that X_0 is a feasible solution of problem (ML1), for any matrix $E \in \text{null}(A, L)\setminus\{\mathbf{0}_{N\times L}\}$, $X_0 + E$ is also feasible. If $\|X_0 + E\|_{2,1} \le \|X_0\|_{2,1}$, then we have $E \in \mathcal{D}(\|\cdot\|_{2,1}, X_0) \cap \text{null}(A, L)\setminus\{\mathbf{0}_{N\times L}\} = \emptyset$, which is impossible. Therefore

$$\|X_0 + E\|_{2,1} > \|X_0\|_{2,1} \ \text{ for all } E \in \text{null}(A, L)\setminus\{\mathbf{0}_{N\times L}\},$$

which means that X_0 is the unique optimal solution to problem (ML1). $\square$

C_1 and C_2 in theorem 3.3 are replaced by $\mathcal{D}(\|\cdot\|_{2,1}, X_0)$ and null(A, L), respectively, to derive the following theorem.

Theorem 4.2. (Phase transitions in MMVs recovery)
Fix a tolerance $\eta \in (0,1)$. Let $X_0 \in \mathbb{R}^{N\times L}$ be a fixed matrix. Suppose $A \in \mathbb{R}^{M\times N}$ has independent standard normal entries, and let $Y = AX_0$. Then

$$M \geq \frac{\delta(\mathcal{D}(\|\cdot\|_{2,1}, X_0))}{L} + \frac{a_\eta\sqrt{NL}}{L} \Rightarrow \mathbb{P}\left((ML1) \text{ succeeds}\right) \geq 1-\eta;$$
$$M \leq \frac{\delta(\mathcal{D}(\|\cdot\|_{2,1}, X_0))}{L} - \frac{a_\eta\sqrt{NL}}{L} \Rightarrow \mathbb{P}\left((ML1) \text{ succeeds}\right) \leq \eta,$$

where the quantity $a_\eta := \sqrt{8\log(4/\eta)}$.

5. CALCULATING THE STATISTICAL DIMENSION

In Theorem 4.2, the calculation of $\delta(\mathcal{D}(\|\cdot\|_{2,1}, X_0))$ is the key to realize theoretical results but is still an open problem. To compute the S.D. of descent cone of $\ell_{2,1}$-norm in our method, we follow Amelunxen *et al.*'s theorem [10] below that provides the upper bound of general S.D. of a cone.

Theorem 5.1. (Statistical dimension of a descent cone [10])
Let $f : \mathbb{R}^N \to \bar{\mathbb{R}}$ be a proper convex function and let $x \in \mathbb{R}^N$. Assume that the subdifferential $\partial f(x)$ is nonempty, compact, and does not contain the origin. Define the function

$$F(\tau) := F(\tau, \partial f(x)) = \mathbb{E}\left[\text{dist}^2\left(g, \tau\cdot\partial f(x)\right)\right] \text{ for } \tau \geq 0$$

where $g \sim$ NORMAL$(0, I)$. We have the upper bound

$$\delta(\mathcal{D}(f, x)) \leq \inf_{\tau\geq 0} F(\tau). \tag{2}$$

Furthermore, the function F is strictly convex, continuous at $\tau = 0$, and differentiable for $\tau \geq 0$. It achieves its minimum at a unique point.

A theorem stating the lower bound of S.D. is described as follows.

Theorem 5.2. (Error bound for descent cone recipe [10])
Let $f : \mathbb{R}^N \to \bar{\mathbb{R}}$ be a proper convex function, and fix a nonzero point x. Then

$$\inf_{\tau\geq 0} F(\tau) - \frac{2\sup\{\|s\|_2 : s \in \partial f(x)\}}{f(\frac{x}{\|x\|_2})} \leq \delta(\mathcal{D}(f, x)),$$

where the function F is defined in Theorem 5.1.

However, the above norm f is, in fact, an $\ell_{2,1}$-norm in our method. So we need the subdifferential of $\ell_{2,1}$-norm, as mentioned in [19].

Lemma 5.3. (Subdifferential of $\|\cdot\|_{2,1}$ [19])
(a) For any $x^i, u^i \in \mathbb{R}^L$, we have

$$u^i \in \partial\|x^i\|_2 \Leftrightarrow \begin{cases} u^i = x^i/\|x^i\|_2 & \text{if } x^i \neq 0, \\ \|u^i\|_2 \leq 1 & \text{if } x^i = 0. \end{cases} \tag{3}$$

(b) We have $\partial\|X\|_{2,1} \neq \emptyset$ if and only if

$$U \in \partial\|X\|_{2,1} \Leftrightarrow u^i \in \partial\|x^i\|_2,\ 1 \leq i \leq N,$$

where $U = \left[\left(u^1\right)^T, \left(u^2\right)^T, ..., \left(u^N\right)^T\right]^T \in \mathbb{R}^{N\times L}$.

According to Theorem 5.1, Theorem 5.2, and Lemma 5.3, we can estimate the S.D. of descent cone of $\ell_{2,1}$-norm.

Theorem 5.4. (Statistical dimension of descent cone of $\ell_{2,1}$-norm)
Let X be a matrix in $\mathbb{R}^{N\times L}$ with K nonzero rows. Then the normalized S.D. of the descent cone of $\ell_{2,1}$-norm at the point X satisfies the inequality

$$\psi\left(\frac{K}{N}\right) - \frac{2}{\sqrt{KN}} \leq \frac{\delta(\mathcal{D}(\|\cdot\|_{2,1}, X))}{N} \leq \psi\left(\frac{K}{N}\right). \tag{4}$$

The function ψ is defined as

$$\psi(\rho) := \inf_{\tau\geq 0}\{R(\tau)\}, \tag{5}$$

$$R(\tau) = \rho(L+\tau^2) + (1-\rho)\frac{2^{1-L/2}}{\Gamma(L/2)}T(\tau), \tag{6}$$

where $T(\tau) = \int_\tau^\infty (S-\tau)^2 S^{L-1} e^{-S^2/2} dS$ and $\Gamma(L/2)$ is gamma function.

Proof. Without loss of generality, we may assume that the sparse matrix $X \in \mathbb{R}^{N\times L}$ takes the form of

$$X = \left[\left(x^1\right)^T, \ldots, \left(x^K\right)^T, \mathbf{0}_{L\times(N-K)}\right]^T$$

where the first K rows of X are nonzero.

Since $\|\cdot\|_{2,1}$ is a proper convex function, by replacing the function f as $\|\cdot\|_{2,1}$, Eq. (2) becomes

$$\delta(\mathcal{D}(\|\cdot\|_{2,1}, X)) \leq \inf_{\tau\geq 0} \mathbb{E}[\text{dist}^2(G, \tau\cdot\partial\|X\|_{2,1})], \tag{7}$$

where dist$(\mathcal{S}, \mathcal{T}) = \|\mathcal{S} - \mathcal{T}\|_F$. The matrix $G \in \mathbb{R}^{N\times L}$ has independent standard normal entries, which can be written as $G = \left[\left(g^1\right)^T, \ldots, \left(g^K\right)^T, \left(g^{K+1}\right)^T, \ldots, \left(g^N\right)^T\right]^T = \left[G_1^T, G_2^T\right]^T$, where $G_1 \in \mathbb{R}^{K\times L}$ and $G_2 \in \mathbb{R}^{(N-K)\times L}$. By Eq. (3), the distance from matrix G to $\tau\cdot\partial\|X\|_{2,1}$ is

$$\begin{aligned}&\text{dist}^2(G, \tau\cdot\partial\|X\|_{2,1}) \\ &= \sum_{i=1}^K\sum_{j=1}^L \left(g_j^i - \tau\cdot\frac{x_j^i}{\|x^i\|_2}\right)^2 + \sum_{i=K+1}^N \left(\|g^i\|_2 - \tau\right)_+^2,\end{aligned}$$

where $(a)_+ = \max(a, 0)$. Since the entries of G are independent standard normal, $\|g^i\|_2 \sim \chi_L$ follow the chi distribution with L degrees of freedom. After taking the expectation, we have

$$\begin{aligned}&\mathbb{E}\left(\text{dist}^2(G, \tau\cdot\partial\|X\|_{2,1})\right) \\ &= K(L+\tau^2) + (N-K)\frac{2^{1-L/2}}{\Gamma(L/2)}\int_\tau^\infty (S-\tau)^2 S^{L-1} e^{-S^2/2} dS \\ &= K(L+\tau^2) + (N-K)\frac{2^{1-L/2}}{\Gamma(L/2)}T(\tau).\end{aligned}$$

Then Eq. (7) becomes

$$\frac{\delta(\mathcal{D}(\|\cdot\|_{2,1}, X))}{N} \leq \psi\left(\frac{K}{N}\right), \tag{8}$$

where

$$\psi\left(\frac{K}{N}\right) := \inf_{\tau \geq 0} \left\{ \frac{K}{N}(L+\tau^2) + \left(1 - \frac{K}{N}\right) \frac{2^{1-L/2}}{\Gamma(L/2)} T(\tau) \right\}.$$

On the other hand, we aim to approximate the lower bound of S.D. of descent cone of $\ell_{2,1}$-norm . By the fact that the descent cone of a point is affine invariant provided the sign of such point is preserved, then we have

$$\delta(\mathcal{D}(\|\cdot\|_{2,1}, X)) = \delta(\mathcal{D}(\|\cdot\|_{2,1}, \hat{X})),$$

where $\hat{x}^i = \frac{x^i}{\|x^i\|_2}$ for $i = 1, 2, ..., K$, and $\hat{x}^j = \mathbf{0}$ for $j = K+1, ..., N$.

By Theorem 5.2 and Eq. (2), we can see that the gap between the upper and lower bounds of $\delta(\mathcal{D}(\|\cdot\|_{2,1}, \hat{X}))$ is

$$\text{gap} = \frac{2 \sup \left\{ \|U\|_F : U \in \partial \left\| \hat{X} \right\|_{2,1} \right\}}{\left\| \hat{X} / \|\hat{X}\|_F \right\|_{2,1}}. \tag{9}$$

By Eq. (3), we have $\|U\|_F \leq \sqrt{N}$ for every subgradient $U \in \partial \|\hat{X}\|_{2,1}$. In addition, $\left\| \hat{X} / \|\hat{X}\|_F \right\|_{2,1} = \sqrt{K}$. Hence, the gap in Eq. (9) can be derived as:

$$\text{gap} = \frac{2\sqrt{N}}{\sqrt{K}}$$

and we complete the proof. □

Note that the upper bound of normalize S.D. in Eq. (8) is exactly the one in Theorem 5.1, and the infimum in Eq. (5) is achieved for the unique positive τ that solves the stationary equation $R'(\tau) = 0$.

6. VERIFICATIONS

We verify our theoretical analysis about phase transition in compressed sensing via $\ell_{2,1}$-norm minimization by simulations, which were conducted using the CVX package [20]. In our simulations, the signal dimension was fixed at $N = 100$ and both M and K ranged from 1 to $N-1$. The number L of measurement vectors was selected as $L = 2$, 5, and 10. We repeated the following verification procedure 100 times for each set of parameters, composed of M, K, and L.

1. Construct a matrix $X_0 \in \mathbb{R}^{N \times L}$ with K nonzero rows. The locations of nonzero rows were selected at random, where all the entries are nonzero and are either $+1$ or -1 with equal probability.
2. Draw a standard normal matrix $A \in \mathbb{R}^{M \times N}$ and form $Y = AX_0$.
3. Solve problem (ML1) by CVX to obtain an optimal point X^*.
4. Declare success if $\|X^* - X_0\| \leq 10^{-5}$.

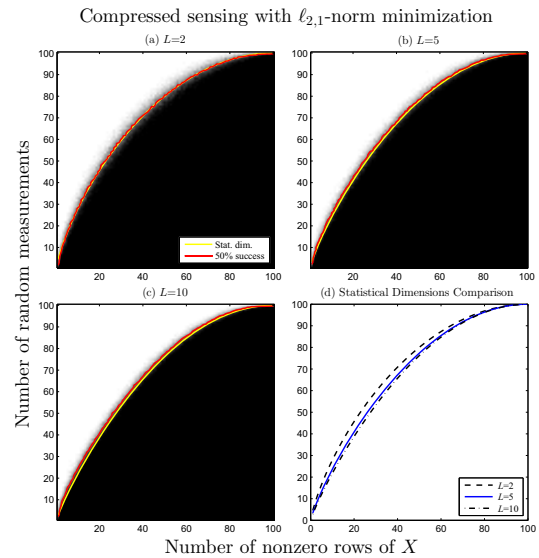

Fig. 1. The empirical probability that problem (ML1) identifies a sparse matrix (a) $X \in \mathbb{R}^{100\times 2}$, (b) $X \in \mathbb{R}^{100\times 5}$, (c) $X \in \mathbb{R}^{100\times 10}$, given random linear measurements $Y = AX$, (d) Comparison of statistical dimensions under $L = 2$ (dash line), $L = 5$ (solid line), and $L = 10$ (dash-dot line).

In Theorem 5.4, we show that $\delta(\mathcal{D}(\|\cdot\|_{2,1}), X_0)$ depends on N, K, and L. In Figs. 1(a)-(c), we can observe that the theoretical curve (in yellow) is extremely close to success rate of 50% (in red), which implies that the derived phase transition of sparse recovery from problem (ML1) meets practical results. Furthermore, we can also observe from Fig. 1(d) that the larger L is, the smaller $\delta(\mathcal{D}(\|\cdot\|_{2,1}), X_0)$ and the width of phase transition are. Such a phenomenon is reasonable because more measurement vectors will be helpful in recovery of sparse signals.

In real applications, N and L are known/determined in advance, K is unknown, and M should be properly determined to save measurement rates. According to Theorem 5.4, with K fixed temporarily, we can determine the lower bound of M, but with M fixed temporarily, we can determine the upper bound of K. Thus, we can dynamically adjust the unknown parameters based on real scenarios.

7. CONCLUSION

Since the phase transition analysis in sparse signal recovery of compressive sensing is too strict or ideal to be satisfied in real world, we present a new phase transition analysis based on conic geometry to close the gap between theoretical analysis and practical recovery result for MMVs.

8. ACKNOWLEDGMENT

This work was supported by Ministry of Science and Technology, Taiwan, under grants MOST 102-2221-E-001-002-MY2 and 102-2221-E-001-022-MY2.

9. REFERENCES

[1] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[2] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[3] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[5] A. Milzarek and M. Ulbrich, "A semismooth newton method with multidimensional filter globalization for l1-optimization," *SIAM J. Optim.*, vol. 24, no. 1, pp. 298–333, 2014.

[6] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, pp. 93–100, Dec 2010.

[7] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[8] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.

[9] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *Information Theory, IEEE Transactions on*, vol. 58, no. 5, pp. 3115–3130, May 2012.

[10] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: phase transitions in convex programs with random data," *Information and Inference*, vol. 3, pp. 224–294, 2014.

[11] R. Schneider and W. Weil, *Stochastic and Integral Geometry*, Springer series in statistics: Probability and its applications. Springer, 2008.

[12] T. H. Chien, W. J. Liang, and C. S. Lu, "A practical subspace multiple measurement vectors algorithm for cooperative spectrum sensing," in *Proc. IEEE Globecom, Austin, TX, USA,*, December 2014.

[13] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477–2488, July 2005.

[14] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4634–4643, Dec 2006.

[15] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3704–3716, July 2007.

[16] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5302–5316, Nov 2009.

[17] Y. Jin and B. D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," in *Acoustics, Speech and Signal Processing, IEEE International Conference on*, March 2008, pp. 3921–3924.

[18] T. T. Do, G. Lu, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 139–154, Jan 2012.

[19] M. Haltmeier, "Block-sparse analysis regularization of ill-posed problems via $\ell_{2,1}$-minimization," in *Methods and Models in Automation and Robotics (MMAR), International Conference on*, Aug 2013, pp. 520–523.

[20] M. G. Grant and S. P. Boyd, *The CVX user's guide, release 2.0(beta)*, cvxr.com/cvx/doc, 2013.