PERFORMANCE ANALYSIS OF JOINT-SPARSE RECOVERY FROM MULTIPLE MEASUREMENT VECTORS WITH PRIOR INFORMATION VIA CONVEX OPTIMIZATION

Shih-Wei Hu, Gang-Xuan Lin, Sung-Hsien Hsieh, Wei-Jie Liang, and Chun-Shien Lu

Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC

ABSTRACT

We address the problem of compressed sensing with multiple measurement vectors associated with prior information in order to better reconstruct an original sparse signal. This problem is modeled via convex optimization with $\ell_{2,1} - \ell_{2,1}$ minimization. We establish bounds on the number of measurements required for successful recovery. Our bounds and geometrical interpretations reveal that if the prior information can decrease the statistical dimension and make it lower than that under the case without prior information, $\ell_{2,1} - \ell_{2,1}$ minimization improves the recovery performance dramatically. All our findings are further verified via simulations.

***Index Terms*—** Convex optimization, Multiple measurement vectors, Sparsity, Statistical dimension

1. INTRODUCTION

1.1. Background and Problem Definition

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. In this paper, we focus on multiple measurement vectors (MMVs) that are sensing results with respect to observed signals. MMVs gradually exhibit the applicability especially in the areas of wireless sensor networks and wearable sensors [4, 5, 6].

Let $S = [s_1, s_2, ..., s_l] \in \mathbb{R}^{n \times l}$ be the matrix of l (> 1) original signals to be sensed by a sensing matrix $\Phi \in \mathbb{R}^{m \times n} (m < n)$ and let the matrix of measurement vectors be $Y = [y_1, y_2, ..., y_l] \in \mathbb{R}^{m \times l}$, where $y_i = \Phi s_i$, $i = 1, 2, ..., l$. We also let $s_i = \Psi x_i$ and let $X_0 = [x_1, x_2, ..., x_l] \in \mathbb{R}^{n \times l}$ be k-joint sparse, where all x_i's share the common support. Given $A = \Phi\Psi$, recovery from MMVs can be efficiently solved via convex optimization as:

$$\text{(Mconvex)} \quad \min_X f(X) \;\text{ s.t. }\; Y = AX,$$

where $f(\cdot)$ denotes a convex function. We call the problem (Mconvex) succeeds if it has a unique optimal solution and is ground truth X_0. Traditionally, we usually set convex function $f(X) = \|X\|_{2,1}$ to enhance the joint-sparsity of X:

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

So far, there is very limited literature about MMVs with prior information via convex optimization. In fact, we can have some prior knowledge about the ground truth X_0 in, for example, the problem of distributed compressive video sensing (DCVS) [7]. In DCVS, we usually adopt higher/lower measurement rates to sample and transmit key/non-key frames at encoder, and then we treat these reconstructed key frames as the prior information for better recovery of the non-key frames at decoder. Mota *et al.* [8] first propose the analysis of single measurement vector (SMV) with prior information via convex optimization. They show that the performance can be improved provided good prior information can be available. In [9], we characterize when problem (ML1) succeeds and derive the phase transition of success rate inspired by the framework of conic geometry [10].

In this paper, we further extend the problem (ML1) to (ML1) plus prior information as:

$$\text{(ML1P)} \quad \min_X \|X\|_{2,1} + \lambda\|X - W\|_{2,1} \;\text{ s.t. }\; Y = AX,$$

where W is prior information associated with ground truth X_0. The goal here is to provide theoretical but practical bound of the probability of successful recovery and analyze the relationship between prior information and performance.

1.2. Contributions of This Paper

- Based on conic geometry, the phase transition of success rate in (ML1P) is derived and is consistent with the empirical results. This study indeed provides the useful insights into how to solve the problem of MMVs with prior information.

- What prior information is "good" can be concluded by our theoretical analysis. For example, instead of giving the rough conclusion such as $\|X_0 - W\|_{2,1}$ being close to 0, we clearly show how the supports of $X - W$ and the signs of $X - W$ affect the performance.

1.3. Notations

For a matrix H, we denote its transpose by H^T; its i^{th} row by h^i; its j^{th} column by h_j; and the i^{th} entry of j^{th} column by h^i_j. $\Lambda_H := \{i : \|h^i\|_2 \neq 0\}$ for a matrix H is a support set that collects the indices of nonzero rows of H. $\|\cdot\|_p$ and $\|\cdot\|_F$ denote the ℓ_p-norm and Frobenius norm, respectively. The $\ell_{p,q}$-norm of a matrix is defined as $\|X\|_{p,q} = \|(\|x^i\|_p)_{n\times 1}\|_q$. The null space of matrix $A \in \mathbb{R}^{m\times n}$ is defined as $\text{null}(A,l) = \left\{Z \in \mathbb{R}^{n\times l} : AZ = \mathbf{0}_{m\times l}\right\}$. Let $\mathbb{E}$ denote the expected value and let $\overline{\mathcal{B}} = \left\{x : \|x\|_2 \le 1,\ x \in \mathbb{R}^n\right\}$ denote closed unit ball. The dot product of two matrices is $\langle X, Y\rangle = \text{tr}\left(X^T Y\right)$.

2. CONIC GEOMETRY

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

Definition 2.1. (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.

By the definition of descent cone, the necessary and sufficient condition of the success of problem (ML1) is described and proved in our earlier work [9]. But in this paper, the main problem we are studying is not related to a norm function, so we need to modify the proof slightly to fit the problem (Mconvex) with general convex function.

Lemma 2.2. (Optimality condition for MMVs recovery with general convex function)
The matrix X_0 is the unique optimal solution to problem (Mconvex) if and only if $\mathcal{D}(f, X_0) \cap \text{null}(A,l) = \{\mathbf{0}_{n\times l}\}$.

Since linear subspace is also a cone, Lemma 2.2 connects the optimal conditions to the relation that the intersection between the descent cone at X_0 and matrix null space is singleton (*i.e.*, problem (Mconvex) succeeds).

For a random sensing matrix A, the probability of success for problem (Mconvex) can be related to the “sizes” of two cones in Lemma 2.2. Amelunxen *et al.* [10] give a way to measure the size of a cone, as described in the following.

Definition 2.3. (Statistical Dimension [10])
The statistical dimension (S.D.) $\delta(\mathcal{C})$ of a closed convex cone $\mathcal{C} \subset \mathbb{R}^n$ is defined as:

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

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

According to the definition of S.D. of a cone, Amelunxen *et al.* [10] derive the probability that two cones with a random rotation are separated as follows.

Theorem 2.4. (Approximate kinematic formula [10])
Fix a tolerance $\eta \in (0,1)$. Suppose that $\mathcal{C}_1,\ \mathcal{C}_2 \subset \mathbb{R}^N$ are closed convex cones, but one of them is not a subspace. Draw an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ uniformly at random. Then

$$\delta(\mathcal{C}_1)+\delta(\mathcal{C}_2) \le n - a_\eta\sqrt{n} \ \Rightarrow\ \mathbb{P}\{\mathcal{C}_1 \cap Q\mathcal{C}_2 = \{\boldsymbol{0}\}\} \ge 1-\eta,$$

$$\delta(\mathcal{C}_1)+\delta(\mathcal{C}_2) \ge n + a_\eta\sqrt{n} \ \Rightarrow\ \mathbb{P}\{\mathcal{C}_1 \cap Q\mathcal{C}_2 = \{\boldsymbol{0}\}\} \le \eta.$$

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

Let $\mathcal{C}_1 = \mathcal{D}\left(f, X_0\right)$ and let $Q\mathcal{C}_2 = \text{null}(A,l)$ with a random matrix $A = \Phi\Psi$ [11]. The probability of intersection given in Theorem 2.4 can be reformulated as the probability of existence of unique optimal solution by Lemma 2.2, *i.e.*,

$$\begin{aligned}&\mathbb{P}(\mathcal{C}_1 \cap Q\mathcal{C}_2 = \{\mathbf{0}\}) = \mathbb{P}(\mathcal{D}(f,X_0) \cap \mathit{null}(A,l) = \{\mathbf{0}_{n\times l}\})\\ &= \mathbb{P}(\text{(Mconvex) succeeds}).\end{aligned}$$

Since the nullity of A is $n-m$ almost surely, the dimension of $\mathcal{C}_2$ is $\delta\left(\text{null}(A,l)\right) = \dim\left(\text{null}(A,l)\right) = (n-m)l$. Then, the probability that (Mconvex) succeeds can be estimated by Theorem 2.5, which was derived in our earlier work [9].

Theorem 2.5. (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 $Y = AX_0$. Then

$$m \ge \tfrac{\delta(\mathcal{D}(f,X_0))}{l} + \tfrac{a_\eta\sqrt{nl}}{l} \Rightarrow \mathbb{P}\left(\textit{(Mconvex) succeeds}\right) \ge 1-\eta;$$
$$m \le \tfrac{\delta(\mathcal{D}(f,X_0))}{l} - \tfrac{a_\eta\sqrt{nl}}{l} \Rightarrow \mathbb{P}\left(\textit{(Mconvex) succeeds}\right) \le \eta,$$

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

3. ESTIMATION OF S.D. IN (ML1P)

In Theorem 2.5, $\delta(\mathcal{D}(f,X_0))$ plays an important role to estimate the probability that (Mconvex) succeeds. However, calculating the exact value of S.D. of a cone is still open. In this section, we provide the bounds of S.D. of descent cone at the point X_0 associated with convex function $\zeta_W(X) = \|X\|_{2,1} + \lambda\|X-W\|_{2,1}$ in problem (ML1P), where function ζ_W is called $\ell_{2,1}$-norm with prior information.

Theorem 3.1. (Error bound of S.D. in (ML1P))
Let $\partial\zeta_W$ be subdifferential of ζ_W. Suppose $\partial\zeta_W(X)$ is nonempty and compact, and does not contain the origin. Then, we have

$$\inf_{\tau\ge 0} F(\tau) - \xi(X) \le \delta\left(\mathcal{D}\left(\zeta_W, X\right)\right) \le \inf_{\tau\ge 0} F(\tau),\text{[1]}$$

where $\xi(X) = \frac{2\|X\|_F\cdot\sup\left\{\|S\|_F : S\in\partial\zeta_W(X)\right\}}{\langle\partial\zeta_W(X),X\rangle}$,

$$F(\tau) := F(\tau,X) = \mathbb{E}\left[\mathit{dist}^2\left(G, \tau\cdot\partial\zeta_W\left(X\right)\right)\right] \ \textit{for}\ \tau \ge 0,$$

[1] The upper bound of S.D. (right inequality) follows Proposition 4.1 [10].

and $G \in \mathbb{R}^{n\times l}$ is a Gaussian random matrix.

Moreover, for k-joint sparse matrix $X_0 \in \mathbb{R}^{n\times l}$, we have

$$\inf_{\tau\geq 0} F(\tau) - \frac{2(1+\lambda)\sqrt{n}}{(1-\lambda)\sqrt{k}} \leq \delta\left(\mathcal{D}\left(\zeta_W, X_0\right)\right) \leq \inf_{\tau\geq 0} F(\tau).$$

Please refer to the full version [12] for detailed proof.

To calculate the function $F(\tau)$ in Theorem 3.1, we first compute the subdifferential of both $\ell_{2,1}$-norm and $\zeta_W(X)$.

Lemma 3.2. (Subdifferential of $\ell_{2,1}$-norm [13]) *For any $X, U \in \mathbb{R}^{n\times l}$, we have*

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

where

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

The subgradient of $\ell_{2,1}$-norm at X is calculated by row-by-row subgradient of Euclidean norm $\|\cdot\|_2$, whereas $\partial\left\|x^i\right\|_2$ consists of the gradient whenever $x^i \neq 0$, and $\partial\left\|x^i\right\|_2 = \overline{\mathcal{B}}$ if $x^i = 0$. That is, the computation of subgradient of $\ell_{2,1}$-norm at X depends on if a row of X is zero or not.

Moreover, since the subdifferential of $\zeta_W(X)$ can be calculated separately as $\partial(\|X\|_{2,1} + \lambda\|X - W\|_{2,1}) = \partial\|X\|_{2,1} + \lambda\partial\|X - W\|_{2,1}$, we calculate the subgradient of $\zeta_W(X)$ according to the indices sets of zero and nonzero rows with respect to X and $X - W$. We separate the domain of $\zeta_W(X)$ into four cases, where $E_1 = \Lambda_X \cap \Lambda_{X-W}$, $E_2 = \Lambda_X \cap \Lambda^c_{X-W}$, $E_3 = \Lambda^c_X \cap \Lambda_{X-W}$, and $E_4 = \Lambda^c_X \cap \Lambda^c_{X-W}$. Then, we have the following Lemma.

Lemma 3.3. (Subdifferential of $\ell_{2,1}$-norm with prior information) *For any $X, U \in \mathbb{R}^{n\times l}$, we have*

$$U \in \partial\zeta_W\left(X\right) \Leftrightarrow u^i \in \partial(\|x^i\|_2 + \lambda\|x^i - w^i\|_2), 1 \leq i \leq n,$$

where

$$u^i \in \partial(\|x^i\|_2 + \lambda\|x^i - w^i\|_2) \Leftrightarrow$$
$$\begin{cases} u^i = \frac{x^i}{\|x^i\|_2} + \lambda(\frac{x^i - w^i}{\|x^i - w^i\|_2}), & \text{if } i \in E_1, \\ u^i = \frac{x^i}{\|x^i\|_2} + \lambda\beta^i, \|\beta^i\|_2 \leq 1, & \text{if } i \in E_2, \\ u^i = \alpha^i + \lambda(\frac{x^i - w^i}{\|x^i - w^i\|_2}), \|\alpha^i\|_2 \leq 1, & \text{if } i \in E_3, \\ u^i = \alpha^i + \lambda\beta^i, \|\alpha^i\|_2 \leq 1, \|\beta^i\|_2 \leq 1, & \text{if } i \in E_4. \end{cases}$$

According to Lemma 3.3, Theorem 3.1 can be rewritten as follows.

Theorem 3.4. (Statistical dimension of descent cone of $\ell_{2,1}$-norm with prior information) *With the same notations and assumptions as in Theorem 3.1, the S.D. of the descent cone of ζ_W at the point X_0 satisfies the inequality*

$$\psi_p - \frac{2(1+\lambda)\sqrt{n}}{(1-\lambda)\sqrt{k}} \leq \delta(\mathcal{D}(\zeta_W, X_0)) \leq \psi_p. \qquad (1)$$

The function ψ_p is defined as $\psi_p(\boldsymbol{E}) := \inf_{\tau\geq 0}\{R_p(\tau, \boldsymbol{E})\}$, where $\boldsymbol{E} = (|E_1|, |E_2|, |E_3|, |E_4|)$ and $R_p = T_1 + T_2 + T_3 + T_4$ with

$$T_1 = |E_1|\left(l + \tau^2 + \tau^2\lambda^2\right) + 2\tau^2\lambda\sum_{i\in E_1}\cos(\angle Ox^i_0 w^i),$$

$$T_2 = |E_2|\int_{\tau\lambda}^{\infty}(t - \tau\lambda)^2 \cdot \frac{\tau t^l e^{-\frac{t^2+\tau^2}{2}}}{(\tau t)^{l/2}} I_{l/2-1}(\tau t)dt,$$

$$T_3 = |E_3|\int_{\tau}^{\infty}(t - \tau)^2 \cdot \frac{\tau\lambda t^l e^{-\frac{t^2+\tau^2\lambda^2}{2}}}{(\tau\lambda t)^{l/2}} I_{l/2-1}(\tau\lambda t)dt,$$

$$T_4 = |E_4|\frac{2^{1-L/2}}{\Gamma(l/2)}\int_{\tau(1+\lambda)}^{\infty}(t - \tau(1+\lambda))^2 t^{l-1}e^{-t^2/2}dt,$$

where Γ is gamma function and $I_v(z) = \sum_{k=0}^{\infty}\frac{1}{\Gamma(k+1)\Gamma(v+k+1)}\left(\frac{z}{2}\right)^{2k+v}$ is modified Bessel functions of the first kind.

Please refer to the full version [12] for detailed proof.

Following Theorem 3.4, since R_p is strictly convex, the infimum value can be computed by finding the root of derivative of R_p. Moreover, if we divide the inequality in Eq. (1) by n, we can see that the error term $\frac{2(1+\lambda)}{(1-\lambda)\sqrt{nk}}$ is inversely proportional to n. That is, the error term is negligible as n is large enough. We verify Theorem 3.4 in the next section.

4. VERIFICATION

The verification was conducted using the CVX package [14]. Based on Theorem 3.4, it's clear to see that S.D. is highly related to ψ_p, which is dominated by $\boldsymbol{E}$ and $\sum_{i\in E_1}\cos(\angle Ox^i w^i)$ called the cosine term. Hence, our simulations are divided into three categories: (1) Examine how prior information, controlled by $|E_2|$, improve the performance, (2) Verify how prior information with correct supports but imprecise values, controlled by $|E_1|$ and cosine term, affect the performance, and (3) Examine how prior information with wrong supports, controlled by $|E_3|$, affect the performance (Please refer to our full version [12] for the 3rd category.).

4.1. Parameter Setting

The signal dimension was fixed at $n = 100$ and sparsity was set to $k = 16$. Since there are no changes with performance when the length of a measurement vector m is larger than $\frac{n}{2}$ in all simulations, m was set to range from 1 to $\frac{n}{2}$ to focus on the phase transition of performance. In our simulations, we construct a signal matrix $X_0 \in \mathbb{R}^{n\times l}$ with k nonzero rows and generate prior information W with k_W nonzero rows to satisfy $w^i = x^i, \forall i \in \Lambda_W \subset \Lambda_X$.

4.2. Prior Information Controlled by $|E_2|$

The following procedure (Step 1 $\sim$ 3) was repeated 100 times for each set of parameters, composed of l and k_W.

Step 1 Draw a standard normal matrix $A \in \mathbb{R}^{m\times n}$ and generate $Y = AX_0$.

Step 2 Solve problem (ML1P) by CVX to obtain an optimal solution X^*.

Step 3 Declare success if $\|X^* - X_0\|_F \le 10^{-5}$.

In Fig. 1, the theoretical curve (in black), indicating $\frac{\delta(\mathcal{D}(\zeta_W, X_0))}{l}$ derived in Theorem 2.5, is located at the vague region (of separating success and failure) of practical recovery results (in blue). We can observe that the theoretical results (in black) and the practical results (in blue) in Fig. 1(b) are more close to the origin than those in Fig. 1(a) because more correct supports (*i.e.*, larger k_W) are available. Similar results can also be observed in Figs. 1(c) and (d) when l becomes larger. In addition, they show that both the theoretical and practical results will be more close to the origin than those in Figs. 1(a) and (b) due to a larger l is used.

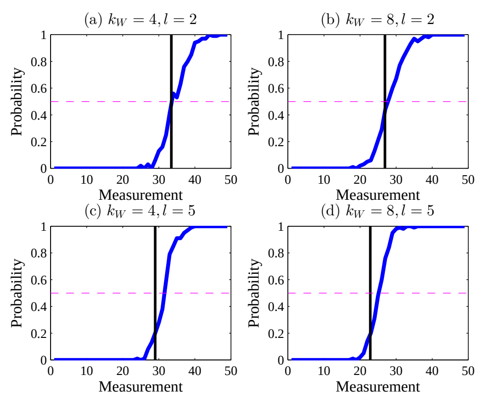

Fig. 1. The empirical probability that problem (ML1P) recovers a sparse signal matrix with the help of prior information W: (a) $k_W = 4$ and $l = 2$; (b) $k_W = 8$ and $l = 4$; (c) $k_W = 4$ and $l = 5$; (d) $k_W = 8$ and $l = 5$.

4.3. Prior Information with Correct Supports but Imprecise Values

We discuss how much influence of cosine term on S.D. and performance. This is equivalent to exploring the similarity between X_0 and W. The parameters were $l = 5$ and $k_W = 8$. We repeat the procedure (Step 1 $\sim$ 3) 100 times for four types of prior information, described as follows.

Type 1. $w^i \sim N(\mathbf{0}, I_{5\times5}), \ \forall i \in \Lambda_W$.

Type 2. $w^i = \text{sign}(x^i), \ \forall i \in \Lambda_W$.

Type 3. $w^i = (\mu + 3\sigma)\cdot \text{sign}(x^i), \ \forall i \in \Lambda_W$, where μ and σ are mean and standard deviation of x^i, respectively.

Type 4. $w^i = x^i, \ \forall i \in \Lambda_W \subset \Lambda_X$.

The results are shown in Fig. 2 and are summarized as follows: (1) As shown in Fig. 2 (a), Type 1 makes the cosine term $\cos(\angle Ox^i w^i)$ unpredictable but is expected to be the highest one among the four types and cause the worst performance. (2) In Fig. 2 (b), W only has correct signs, so it cannot ensure if $\cos(\angle Ox^i w^i)$ is greater than or less than 0. However, correct direction still improves the performance. (3) In Fig. 2 (c), W has correct signs with the original signal and satisfies $|x^i_j| < |w^i_j|$ for $i \in \Lambda_W$ and $1 \le j \le l$ with probability as high as 99%. These make the cosine term less than 0 and lead to better performance. (4) Since Type 4 carries the best prior information, Fig. 2 (d) exhibits the upper bound of performance.

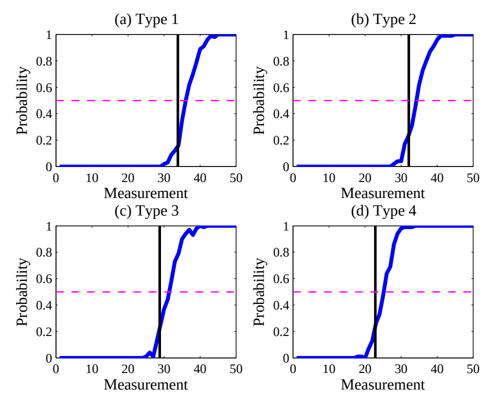

Fig. 2. The empirical probability that problem (ML1P) identifies a sparse matrix with l measurement vectors under prior information W: (a) Type 1; (b) Type 2; (c) Type 3; (d) Type 4.

5. CONCLUSION

In view of the fact that the phase transition analysis in joint-sparse signal recovery with prior information of compressive sensing is relatively unexplored, we have presented a new phase transition analysis based on conic geometry to figure out the effect of prior information for MMVs in this paper. Our studies indeed provide useful insights into the critical problem of selecting prior information to guarantee improvement of signal recovery in the context of compressive sensing.

6. ACKNOWLEDGMENT

This work was supported by Ministry of Science and Technology, Taiwan, ROC, under grants MOST 104-2221-E-001-019-MY3 and 104-2221-E-001-030-MY3.

7. 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] Weiwei Li, Ting Jiang, and Ning Wang, "Compressed sensing based on the characteristic correlation of ecg in hybrid wireless sensor network," *International Journal of Distributed Sensor Networks*, vol. 501, pp. 325103, 2015.

[5] Zhilin Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *Biomedical Engineering, IEEE Transactions on*, vol. 62, no. 8, pp. 1902–1910, Aug 2015.

[6] Ling Xiao, Renfa Li, Juan Luo, and Mengqin Duan, "Activity recognition via distributed random projection and joint sparse representation in body sensor networks," in *Advances in Wireless Sensor Networks*, Limin Sun, Huadong Ma, and Feng Hong, Eds., vol. 418 of *Communications in Computer and Information Science*, pp. 51–60. Springer Berlin Heidelberg, 2014.

[7] Li-Wei Kang and Chun-Shien Lu, "Distributed compressive video sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2009 IEEE International Conference on*, April 2009, pp. 1169–1172.

[8] João F. C. Mota, Nikos Deligiannis, and Miguel R. D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *CoRR*, vol. abs/1408.5250, 2014.

[9] Shih-Wei Hu, Gang-Xuan Lin, Sung-Hsien Hsieh, and Chun-Shien Lu, "Phase transition of joint-sparse recovery from multiple measurements via convex optimization," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, April 2015, pp. 3576–3580.

[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, no. 3, pp. 224–294, 2014.

[11] 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.

[12] S.-W. Hu, G.-X. Lin, S.-H. Hsieh, W.-J. Liang, and C.-S. Lu, "Performance analysis of joint-sparse recovery from multiple measurement vectors with prior information via convex optimization," *http://arxiv.org/abs/1509.06655*, 2015.

[13] 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.

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