Proofs of Number of Compressed Measurements Needed for Noisy Distributed Compressed Sensing

Sangjun Park and Heung-No Lee*
School of Information and Communications
Gwangju Institute of Science and Technology
Gwangju, Republic of Korea
{sjpark1@gist.ac.k and heungno@gist.ac.kr}
*Correspondence Author

Abstract—**In this paper, we consider a data collection network (DCN) system where sensors take samples and transmit them to a Fusion Center (FC). Signal correlation is modeled with signal sparseness. The number of compressed measurements which allows correct *support set* recovery at FC is investigated. This is done by finding the probability of support set recovery errors. The joint typical receiver considered by Akcakaya and Tarokh is used to avoid dependence on particular choice of recovery routines. The following interesting results have been obtained: 1) The number of compressed measurements per sensor (PSM) converges to sparsity as the number of sensors increases. 2) The detection failure probability *linearly* converges to zero as the number of sensors increases. 3) The number of sensors guaranteeing a certain level of failure probability is given the system parameters such as Gaussian noise variance, PSM, and sparsity.**

Keywords-Compressed Sensing, Joint Typicality, Distributed Source Coding, Distributed Compressed Sensing.

I. INTRODUCTION

We consider a data collection network (DCN) system in which there are one signal fusion center (FC) and many sensors reporting to it. Sensors acquire signal samples independently and transmit acquired signal samples to FC. FC then aims to reconstruct each individual signal perfectly. The important problem we aim to investigate here is how to utilize the signal correlation present in the acquired signals and reduce the traffic volume from sensors to FC. This type of questions frequently arise in wireless sensor networks where sensors operate drawing power from onboard batteries and thus saving power from unnecessary transmissions is of utmost importance. To deal with this type of problem, distributed source coding [1][2] has been studied in the past.

Signals in the DCN system are often correlated with each other because sensors are usually deployed in a restricted region and put to observe a phenomenon globally occurring in the region. Sensors can utilize signal correlations and reduce the amount of traffic. The signal reconstruction unit at FC also notices the presence of signal correlation and utilizes this information in a joint signal reconstruction. As the result, the amount of traffic each sensor has to transmit is reduced. This is the main idea of distributed source coding. Recently, Duarte *et al.* [6] coined the term *Distributed Compressed Sensing* which means that distributed source coding is achieved via compressed sensing (CS) at each sensor. CS [3], as a new signal acquisition paradigm, is suitable for sensors with limited onboard resources such as power and storage element.

In CS, signal correlation is modeled by signal *sparse*ness. A signal $\mathbf{x} \in \Re^N$ is said to be *sparse* with *sparsity* $\|\mathbf{x}\|_0 = K$, where $\|\mathbf{x}\|_0$ is the number of non-zero elements of $\mathbf{x}$. A *support set* is the collection of indices of the non-zero elements of $\mathbf{x}$. The more a signal is correlated, the smaller the sparsity K. A sparse signal $\mathbf{x}$, i.e., a correlated signal, can be compressively sampled, via a linear transformation, i.e., $\mathbf{y} = \mathbf{F}\mathbf{x}$ where $\mathbf{F}$ is $M \times N$, called the sensing matrix. Compression is said to be made when $M < N$. It is perhaps the most important and surprising fact in the CS theory that the unknown signal $\mathbf{x}$ can be found uniquely from the compressed signal $\mathbf{y}$ as long as a certain set of conditions on $\mathbf{F}$ are satisfied [5].

For the DCN system, inter-sensor correlations exist between any two acquired signals. Inter-sensor correlations can be modeled by the portion of sensors having the same *support set*. Intra-sensor correlations, in contrast, are signal correlations that exist inside a single sensor signal. Thus, the collection of signals acquired by a group of sensors contains inter- and intra-sensor correlations. A jointly sparse signal set can be defined to describe each signal in the collection. The joint signal reconstruction at FC thus should be able to exploit both the inter- and *intra-sensor* correlations and have each sensor take a less number of compressed samples transmitted to FC.

The main focus of this paper is to determine how many number of measurements per sensor (PSM) is needed for correct recovery of the support of the jointly sparse signals, as the number of sensors and the noise variance are varied. The jointly typical decoder (JT decoder) introduced in [2][4] is extended for the DCN system so that a result which does not dependent upon any particular choice of recovery algorithms can be attained. We obtain an upper bound on the detection failure probability. We explicitly prove that PSM converges to sparsity as the number of sensors increases. We prove that the detection failure probability *linearly* converges to zero as the number of sensors increases. We obtain the number of sensors required for a guaranteed detection performance, given the system parameters such as Gaussian noise variance, PSM, and sparsity.

II. RELATED WORKS

Duarte *et al* [6] introduced a new theory and algorithm for a distributed compressed sensing. In their system, each sparse signal shares the same support set. Their strategy is described as follows: 1) Each sensor independently compresses their signals, i.e., $\mathbf{y}_i = \mathbf{F}_i \mathbf{x}_i \in \Re^{M_i}$, where sub index *i* denotes the i^{th} sensors. 2) All the compressed signals are collected at the central unit via noiseless channel. 3) A decoder at the central unit tries to jointly reconstruct all the signals. To jointly reconstruct, they designed One-Step Greedy Algorithm (OSGA). Furthermore, they analyzed OSGA by using the central limit theorem. Their analysis and simulation result shows that $M \ge K+1$ is sufficient for the perfect recovery, as the number of sensors increases. We note that all the sensing matrices can be different at each sensor and they did not consider the presence of noise in their work.

Tang and Nehorai [7] worked an MMV problem which is similar to this work in the assumption that all the sensing matrices at each sensor are the same. They analyzed performance of estimating the support set that is shared by each sparse signal under a AWGN channel. They introduced a hypothesis test framework; obtained both upper and lower bounds on the probability that the support set is not correctly detected by using the Chernoff bound and Fano's inequality. Their main result is that $M \ge 2K$ is sufficient for estimating the support set in their Theorem 1. Similar to our result, they also mentioned it, i.e., $M \ge K+1$. They insisted that this result can be derived from their Theorem 3, as the following quote indicates "Actually, Malioutov *et al*. made the empirical observation that l_1-SVD technique can resolve $M-1$ sources if they are well separated. Theorem 3 still applies to this extreme case." But it is not explicitly done and in fact difficult to make use of their Theorem 3 and draw the result, $M \ge K+1$.

Now, we aim to introduce the work in [4] because our JT decoder is inspired from [4]. Akcakaya and Tarokh [4] showed that $M = \mathrm{O}\left(K \log\left(N/K\right)\right)$ for a single sensor. To obtain it, they introduced JT decoder. Their JT decoder was inspired from Shannon's work. They said that "We define a decoder that characterizes events based on their typicality. We call such a decoder a "joint typicality decoder." and "Error events are defined based on atypicality, and the probability of these events are small as a consequence of the law of large numbers." They analyzed the joint typical decoder by using probabilistic approaches. They obtained the upper bound on the probability of error events. After they obtained the upper bound, they showed that it converges to zero as the number of compressed measurements increases like $\mathrm{O}\left(K \log\left(N/K\right)\right)$. Furthermore, they showed that a distortion, i.e., mean square error, is roughly bounded.

We adopt JT decoder here and we define events that JT decoder fails to detect the support set; obtained the upper bound on the probability on events. The one difficult problem is how we obtain the upper bound on the probability for the multiple sensors. In their case, they borrowed exponential inequalities from [10] and used as a tool. In our case, to obtain the upper bound, we first made the Chernoff bound; second get the new upper bound on the Chernoff bound. Finally, we minimize the further upper bound. Interestingly we found a way to factor out the number of sensors from the final form by taking the logarithm based on the natural number. Thus, the final form is well suitable for our work.

III. SYSTEM MODEL

There exist S sensors measuring signals; each compresses the signals via CS and transmitting acquired samples to FC. Let the acquired signal at each sensor be $\mathbf{x}_s \in \Re^N$ with $\|\mathbf{x}_s\|_0 = K$, where $s \in \{1, 2, \cdots, S\}$. The support set of $\mathbf{x}$ is defined as

$$\mathcal{I}\left(\mathbf{x}\right) := \operatorname{supp}\left(\mathbf{x}\right) = \left\{ i \middle| x\left(i\right) \neq 0 \right\}.$$

We assume that all the sparse signals have the same support set. Thus, $\mathcal{I} = \mathcal{I}\left(\mathbf{x}_1\right) = \cdots = \mathcal{I}\left(\mathbf{x}_S\right)$ [1]. The compressed signal at each sensor is given as

$$\mathbf{y}_s = \mathbf{F}_s \mathbf{x}_s, \tag{1}$$

where all the elements of $\mathbf{F}_s \in \Re^{M \times N}$ follow i.i.d. Gaussian $\mathcal{N} \sim \left(0, 1\right)$. We call $\mathbf{F}_s$ the s^{th} sensing matrix. All the compressed signals are transmitted to FC via an AWGN channel. Then, the decoder at FC receives

$$\mathbf{r}_s = \mathbf{y}_s + \mathbf{n}_s, \tag{2}$$

where all the elements of $\mathbf{n}_s$ follow i.i.d. Gaussian $\mathcal{N} \sim \left(0, \sigma_{\text{noise}}^2\right)$. We call $\mathbf{n}_s$ the s^{th} noise vector. We assume that all the noise vectors and all the sensing matrices are mutually independent. For simplicity, we denote $\mathbf{r} = \left[\mathbf{r}_1 \cdots \mathbf{r}_S\right]$, $\mathbf{x} = \left[\mathbf{x}_1 \cdots \mathbf{x}_S\right]$, and $\mathbf{n} = \left[\mathbf{n}_1 \cdots \mathbf{n}_S\right]$.

Similar model was considered in [6][7]. In both works, they assumed that all the signals have the same support set. In particular, when we remove all the noise vectors in our model, then ours becomes what Duarte *et al* used in [6]. In [7], they assumed that all the sensing matrices are the same. Both aims at correctly detecting the support set $\mathcal{I}$ from $\mathbf{r}$ and all the sensing matrices.

IV. JOINT TYPICAL (JT) DECODER AND EVENT

Now, we aim to introduce joint typical (JT) decoder divided into two different parts. The first part is the *Support Set Detection* part, where JT decoder estimates the support set. After getting the estimated support set, JT decoder computes all the signals using pseudo inverse operation which is the *Signal Estimation*.

Definition 1: (*Support Set Detection*) JT decoder estimates the support set by employing all the received vectors and all the sensing matrices.

$$D_1 : \left(\forall_s : \mathbf{r}_s, \forall_s : \mathbf{F}_s\right) \mapsto \mathcal{O},$$

where $\mathcal{O} \subset \{1, \cdots, N\}$ with $|\mathcal{O}| = K$.

---

1 $\mathcal{I} = \mathcal{I}\left(\mathbf{x}_1\right) = \cdots = \mathcal{I}\left(\mathbf{x}_S\right)$: $\mathcal{I}$ is always the support set in the whole paper.

Definition 2: (*Signal Estimation*) JT decoder computes all the signals by using the output of Support Set Detection.

$$D_2 : \forall_i \hat{\mathbf{x}}_i = \left(\mathbf{F}_{i,\mathcal{O}}^{\mathrm{T}} \mathbf{F}_{i,\mathcal{O}}\right)^{-1} \mathbf{F}_{i,\mathcal{O}}^{\mathrm{T}} \mathbf{r}_i ,$$

where sub-matrix $\mathbf{F}_{i,\mathcal{O}}$ is constructed by collecting the set of column vectors of $\mathbf{F}_i$ corresponding to indices of $\mathcal{O}$ which is the output of Support Set Detection.

Clearly, all the signals are reconstructed when an output from Support Set Detection is equal to the support set. Now, we introduce a $\delta-$jointly typical event.

Definition 3: ($\delta-$*Joint Typicality*) We say that an $M \times S$ matrix $\mathbf{r}$ and a set $\mathcal{J}$ with $|\mathcal{J}| = K$ are $\delta-$jointly typical if $rank\left(\mathbf{F}_{s,\mathcal{J}}\right) = K$ for all s and

$$\left| \sum_s \frac{\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{r}_s\right\|^2}{SM} - \frac{(M-K)\sigma_{\text{noise}}^2}{M} \right| < \delta, \tag{3}$$

where $\mathbf{Q}(\mathbf{A}) = \mathbf{I} - \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathrm{T}}$ and $\delta > 0$. For simplicity, we consider $\mathrm{E}(\mathbf{r}, \mathcal{J}, \delta)$ that $\mathbf{r}$ and $\mathcal{J}$ are a $\delta-$jointly typical event.

Now, we introduce failure events that JT decoder fails to estimate the support set. The first failure event is that there exists event such that $rank\left(\mathbf{F}_{s,\mathcal{J}}\right) < K$. Clearly, JT decoder cannot estimate the support set because of (3). The second failure event is $\mathrm{E}(\mathbf{r}, \mathcal{J} \neq \mathcal{I}, \delta)$. JT decoder considers the incorrect support set as the correct support set in this event. The last event is $\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^{C}$ [2]. JT decoder is not aware of the correct support set. Hence, JT decoder fails to estimate the correct support set whenever any one of these three events occurs. Let $\mathrm{E}\left(D_{failure}\right)$ be the detection failure event. Then,

$$\begin{aligned}\mathrm{E}\left(D_{failure}\right) = \mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^{c} & \bigcup_{\forall_{\mathcal{J} \neq \mathcal{I}, |\mathcal{J}|=K}} \mathrm{E}(\mathbf{r}, \mathcal{J}, \delta) \\ & \bigcup_{\forall_s \forall_{\mathcal{J}, |\mathcal{J}|=K}} \mathrm{E}\left(rank\left(\mathbf{F}_{s,\mathcal{J}}\right) < K\right)\end{aligned} . \tag{4}$$

The last term, i.e., $\mathrm{E}\left(rank\left(\mathbf{F}_{s,\mathcal{J}}\right) < K\right)$ in (4) can be ignored because all the entries of $\mathbf{F}_s$ follow i.i.d. Gaussian $\mathcal{N} \sim (0,1)$. Therefore, the first failure event rarely occurs. Now, we address the randomness for both the remaining events, i.e., $\mathrm{E}(\mathbf{r}, \mathcal{J} \neq \mathcal{I}, \delta)$ and $\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^{C}$.

It is easy to show $\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{I}}\right)\mathbf{r}_s\right\|^2$ and $\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{r}_s\right\|^2$ become $\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{I}}\right)\mathbf{n}_s\right\|^2$ and $\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\left(\sum_{i \in \mathcal{I} \setminus \mathcal{J}} x_s(i)\mathbf{f}_s(i) + \mathbf{n}_s\right)\right\|^2$ in (3) respectively. In the System Model section, we assume that $\mathbf{n}_s$ and $\mathbf{f}_s(i)$ are random vectors. Therefore, $\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{I}}\right)\mathbf{r}_s\right\|^2$ and $\left\|\mathbf{Q}\left(\mathbf{F}_{s,\mathcal{J}}\right)\mathbf{r}_s\right\|^2$ are random variables.

V. PROBABILITES OF THE FAILURE EVENTS

We continuously aim to talk about probabilities of all the failure events. By using the union bound approach, we have

$$\begin{aligned}\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le & \Pr\left\{\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^{c}\right\} \\ & + \sum_{\forall_{\mathcal{J} \neq \mathcal{I}, |\mathcal{J}|=K}} \Pr\left\{\mathrm{E}(\mathbf{r}, \mathcal{J}, \delta)\right\}\end{aligned} . \tag{5}$$

We are interested in obtaining both probabilities, i.e., $\Pr\left\{\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^{c}\right\}$ and $\Pr\left\{\mathrm{E}(\mathbf{r}, \mathcal{J} \neq \mathcal{I}, \delta)\right\}$ respectively. Instead of obtaining exact probabilities, we get upper bounds on them. Lemma 1 and Lemma 2 provide upper bounds. The following notations become useful for representing both upper bounds:

$$p_c(T, \mathcal{I}) := \exp\left(-\frac{SM\delta'}{2}\right) \times \left(1 + \frac{M\delta'}{M-K}\right)^{S\frac{M-K}{2}} \tag{6}$$

and

$$\begin{aligned} p_i(T, \mathcal{J}) := & \exp\left(-\frac{SM}{2\sigma_{\min}^2}\left(\frac{M-K}{M}\left(\sigma_{\text{noise}}^2 - \sigma_{\min}^2\right) + \delta\right)\right) \\ & \times \left(\frac{\sigma_{\text{noise}}^2}{\sigma_{\min}^2} + \frac{M}{M-K}\frac{\delta}{\sigma_{\min}^2}\right)^{S\frac{M-K}{2}} \end{aligned} \tag{7}$$

where $T = \left\{S, M, K, \delta, \sigma_{\text{noise}}^2\right\}$ [3], $\sigma_{\min}^2 = \min_{s \in \{1,\cdots,S\}} \left(\sigma_{s,\mathcal{J}}^2\right)$, $\delta' = \delta / \sigma_{\text{noise}}^2$ and $\sigma_{s,\mathcal{J}}^2 = \sum_{i \in \mathcal{I} \setminus \mathcal{J}} x_s(i)^2 + \sigma_{\text{noise}}^2$. Here, $\mathcal{J}$ is one of the incorrect support sets.

Lemma 1: Let $\mathcal{I}$ be the correct support set and the rank of $\mathbf{F}_{s,\mathcal{I}}$ be K for all s. Then, for any $\delta > 0$, we have

$$\Pr\left\{\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^{c}\right\} \le 2 p_c(T, \mathcal{I}). \tag{8}$$

Lemma 2: Let $\mathcal{J}$ be the one of the incorrect support sets, $0 \le |\mathcal{I} \cap \mathcal{J}| < K$ and the rank of $\mathbf{F}_{s,\mathcal{J}}$ be K for all s. Then, for any $\delta > 0$, we have

$$\Pr\left\{\mathrm{E}(\mathbf{r}, \mathcal{J}, \delta)\right\} \le p_i(T, \mathcal{J}). \tag{9}$$

The detailed proofs of them are given in [11].

VI. CONVERGENCE RESULT

In the previous section, we obtained the upper bounds on the two events. In this section, we aim to examine their behavior depending on S. In other words, what is the behaviors of both the upper bounds when we increase S? This will be useful for answering these questions.

---

[2] $\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^{C}$: complement event of $\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)$.

[3] In the whole paper, T, T^* and T_1 are always defined as $T = \left\{S, M, K, \delta, \sigma_{\text{noise}}^2\right\}$, $T^* = \left\{S = 1, M, K, \delta, \sigma_{\text{noise}}^2\right\}$ and $T_1 = \left\{S+1, M, K, \delta, \sigma_{\text{noise}}^2\right\}$ respectively.

Proposition 1: Let $M > K$, $\mathcal{I}$ be the correct support set and the rank of $\mathbf{F}_{s,\mathcal{I}}$ be K for all s and $\delta > 0$. Then, $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\}$ *linearly converges to zero* with rate $p_c\left(T^*,\mathcal{I}\right)$ as S increases.

Proposition 2: Let $M > K$, $\mathcal{J}$ be one of the incorrect support sets and the rank of $\mathbf{F}_{s,\mathcal{J}}$ be K for all s, $\delta > 0$ and $\sigma^2_{\text{noise}} < \min_{s\in\{1,\cdots,S\}} \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s\left(i\right)^2$. Then, $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J},\delta\right)\right\}$ *linearly converges to zero* with rate $p_i\left(T^*,\mathcal{J}\right)$ as S increases.

In both propositions, $T^* = \left\{S = 1, M, K, \delta, \sigma^2_{\text{noise}}\right\}$. The detailed proofs of them are given in [11].

Let us consider the assumptions in Proposition 1 and Proposition 2. From inspection of the upper bounds, we note $M > K$, seen in (6) and (7). Next, $\sigma^2_{\text{noise}} < \min_{s\in\{1,\cdots,N\}} \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s\left(i\right)^2$ appears only in Proposition 2. It is reasonable. If this condition is not satisfied, JT decoder cannot distinguish between the noise and the signal. It does not appear in Proposition 1 because JT decoder observes noise components when $\mathbf{F}_{s,\mathcal{I}}$ is used.

Furthermore, we have proved that $0 < p_c\left(T^*,\mathcal{I}\right) < 1$ and $0 < p_i\left(T^*,\mathcal{J}\right) < 1$ whenever $M > K$ and $\delta > 0$. Therefore, $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\}$ and $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J},\delta\right)\right\}$ linearly converge to zero.

VII. THEOREMS AND DISCUSSIONS

Theorem 1: Let $M > K$, $\mathcal{I}$ be the correct support set, $\mathcal{J} \subset \{1,\cdots,N\}$ with $|\mathcal{J}| = K$ and $\mathcal{J} \neq \mathcal{I}$, all the ranks of $\mathbf{F}_{s,\mathcal{J}}$ and $\mathbf{F}_{s,\mathcal{I}}$ be K for all s, $\delta > 0$ and $\sigma^2_{\text{noise}} < \min_{s\in\{1,\cdots,S\}} \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s\left(i\right)^2$. Then, $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\}$ *linearly converges to zero* with a rate, $\max\left(p_c\left(T^*,\mathcal{I}\right), p_i\left(T^*,\mathcal{J}^*\right)\right)$[4], as S increases.

Proof : Let us remind the upper bound on $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\}$, i.e., (5). The upper bound can be bounded by using both (6) and (7):

$$\begin{aligned}\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} &\leq 2p_c\left(T,\mathcal{I}\right) + \sum_{\forall \mathcal{J}\neq\mathcal{I}, |\mathcal{J}|=K} p_i\left(T,\mathcal{J}\right) \\ &\leq 2p_c\left(T,\mathcal{I}\right) + \binom{N}{K} \max_{\forall \mathcal{J}\neq\mathcal{I}, |\mathcal{J}|=K} p_i\left(T,\mathcal{J}\right) \\ &= 2p_c\left(T,\mathcal{I}\right) + \binom{N}{K} p_i\left(T,\mathcal{J}^*\right) =: L\left(S\right)\end{aligned} \quad (10)$$

The term $\binom{N}{K}$ appears because all the sparse signals have the same support set. We now know that $p_c\left(T,\mathcal{J}\right)$ and $p_i\left(T,\mathcal{J}\right)$ converge to zero with given rates as S increases. Therefore, the right term in (10) must converge to zero as S increases. Now, we investigate the convergence rate of $L\left(S\right)$.

$$\begin{aligned}\lim_{S\to\infty} \frac{\left|L\left(S+1\right)\right|}{\left|L\left(S\right)\right|} &= \lim_{S\to\infty} \frac{L\left(S+1\right)}{L\left(S\right)} \\ &= \lim_{S\to\infty} \frac{2p_c\left(T_1,\mathcal{I}\right) + \binom{N}{K} p_i\left(T_1,\mathcal{J}^*\right)}{2p_c\left(T,\mathcal{I}\right) + \binom{N}{K} p_i\left(T,\mathcal{J}^*\right)}\end{aligned} \quad (11)$$

where $T_1 = \left\{S+1, M, K, \delta, \sigma^2_{\text{noise}}\right\}$. Let us consider the one of cases. 1) $p_c\left(T^*,\mathcal{I}\right) > p_i\left(T^*,\mathcal{J}^*\right)$: We divide the last term in (11) by $p_c\left(T,\mathcal{I}\right)$, then we have

$$\lim_{S\to\infty} \frac{2\dfrac{p_c\left(T_1,\mathcal{I}\right)}{p_c\left(T,\mathcal{I}\right)} + \binom{N}{K} \dfrac{p_i\left(T_1,\mathcal{J}^*\right)}{p_c\left(T,\mathcal{I}\right)}}{2\dfrac{p_c\left(T,\mathcal{I}\right)}{p_c\left(T,\mathcal{I}\right)} + \binom{N}{K} \dfrac{p_i\left(T,\mathcal{J}^*\right)}{p_c\left(T,\mathcal{I}\right)}} = p_c\left(T^*,\mathcal{I}\right), \quad (12)$$

where $\lim_{S\to\infty} \frac{p_i\left(T_1,\mathcal{J}^*\right)}{p_c\left(T,\mathcal{I}\right)} = p_i\left(T^*,\mathcal{J}\right) \lim_{S\to\infty} \frac{p_i\left(T,\mathcal{J}^*\right)}{p_c\left(T,\mathcal{I}\right)} = 0$ and $\lim_{S\to\infty} \frac{p_c\left(T_1,\mathcal{I}\right)}{p_c\left(T,\mathcal{I}\right)} = p_c\left(T^*,\mathcal{I}\right)$. Similar to this first case, we can prove the last term in (11) goes to zero for cases, where $p_c\left(T^*,\mathcal{I}\right) < p_i\left(T^*,\mathcal{J}^*\right)$ and $p_c\left(T^*,\mathcal{I}\right) = p_i\left(T^*,\mathcal{J}^*\right)$.

Therefore, the convergence rate of $L\left(S\right)$ is the maximum value between $p_c\left(T^*,\mathcal{I}\right)$ and $p_i\left(T^*,\mathcal{J}^*\right)$. Q.E.D.

Similar result was reported. Duarte *et al.* [6] proved and demonstrated that M converges to $K+1$. Limitation of their work is that they did not consider the presence of noise. Tang and Nehorai [7] proved $M \geq 2K$ for correct support set recovery from compressed signals obtained over an AWGN channel. They mentioned that M converges to $K+1$ when they discussed Theorem 3 of [7]. But, it is not explicitly done. From Theorem 3 of [7], it is difficult to draw $M \geq K+1$. Davies and Eldar [8] designed a practical algorithm to recover K sparse signals from the MMV model, $\mathbf{r} = \mathbf{F}\mathbf{x} + \mathbf{n}$, but without considering noises. Their empirical results also showed that only $K+1$ measurements per sensor are enough for good recovery as well.

Although the JT decoder is not a practical decoder as an OSGA developed in [6], it has benefit as a performance analysis tool. It provides benchmark independent of recovery algorithms. For example, given the systems parameters, the detection failure probability of the network system can be found immediately.

When we see (10), (6) and (7), it is easy to see that when $M \geq K+1$, the expressions go to zero as S increases. We find the convergence rate and show that $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\}$ *linearly*

[4] $p_i\left(T,\mathcal{J}^*\right) := \max_{\forall \mathcal{J}\neq\mathcal{I}, |\mathcal{J}|=K} p_i\left(T,\mathcal{J}\right)$ throughout the entire part of this paper.

converges to zero with the rate given in Theorem 1 as S increases.

Proposition 3: Let all the parameters except for S, which are N, M, K, σ^2_{noise}, $\delta > 0$, and $\rho \in (0,1)$ be given. Then, the minimum number of sensors such that $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le \rho$ is

$$S_{\min} := \left\lceil \frac{\log \rho}{\log\left(p_{up}\left(T^*\right)\right)} \right\rceil, \tag{13}$$

if $p_{up}\left(T^*\right) \in (0,1)$, where

$$p_{up}\left(T^*\right) = 2p_c\left(T^*, \mathcal{I}\right) + \binom{N}{K} p_i\left(T^*, \mathcal{J}^*\right). \tag{14}$$

Proof: From (10), we derive

$$\begin{aligned} \Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} &\le 2p_c\left(T, \mathcal{I}\right) + \binom{N}{K} p_i\left(T, \mathcal{J}^*\right) \\ &= 2p_c\left(T^*, \mathcal{I}\right)^S + \binom{N}{K} p_i\left(T^*, \mathcal{J}^*\right)^S \\ &\le \left[2p_c\left(T^*, \mathcal{I}\right) + \binom{N}{K} p_i\left(T^*, \mathcal{J}^*\right)\right]^S \\ &= p_{up}\left(T^*\right)^S \end{aligned}. \tag{15}$$

By taking logarithm on both sides in (15), we get

$$\log\left(\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\}\right) \le S \log\left(p_{up}\left(T^*\right)\right) \tag{16}$$

Now we aim to find S such that $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le \rho$. By letting the right hand side term in (16) be less than $\log(\rho)$, we have

$$S \log\left(p_{up}\left(T^*\right)\right) \le \log(\rho). \tag{17}$$

Eventually, we derive (13) from (17) , which is valid when $p_{up}\left(T^*\right) \in (0,1)$.

Q.E.D.

Proposition 3 gives us the number of sensors sufficient for $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le \rho$ when all the other system parameters except S are given and fixed. This is useful result. For example, suppose $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le 0.3$ for a single sensor DCN. Now, we aim to find the number of sensors S which guarantees $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le 0.01$. From (13), we have $\left\lceil \log(0.01)/\log(0.3) \right\rceil = 4$. It implies $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le 0.01$ when $S \ge 4$.

VIII. Conclusions and Future Works

The main focus of this paper was to investigate how many number of measurements per sensor (PSM) is needed for almost perfect support set recovery, as the number of sensors and the noise variance are varied. For this objective, we obtained an upper bound on the recovery failure probability. Using this upper bound, we explicitly proved that PSM converges to sparsity as the number of sensors increases, Theorem 1. We also proved that the upper bound *linearly* converges to zero as S increases in Theorem 1. Finally, we provided a result (Proposition 3) which is useful to determine the sufficient number of sensors which guarantees a certain level of detection success given all the parameters except for S are given and fixed.

Due to space provided is limited; proofs for propositions and lemmas are relegated to the technical report in [11].

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (Do-Yak Research Program, No. 2011-0016496)

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding for correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471– 480, July, 1973.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory, Second Edition*, Wiley, New York, 2006

[3] D. Donoho, "Compressive sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289 – 1306, 2006

[4] M. Akcakaya and V. Tarokh, "Shannon-Theoretic Limits on Noisy Compressive Sampling," *IEEE Trans. Inform. Thoery*, vol. 56, no. 1 , Jan 2010

[5] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197 – 2202, March 2003

[6] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin and R. G. Baraniuk, "Distributed Compressed Sensing of Jointly Sparse Signals," *Signals, Systems and Computers*, 2005. *Conference Record of the Thirty-Ninth Asilomar Conference*, pp. 1537 – 1541, Oct 28 – Nov 1, 2005.

[7] G. Tang and A. Nehorai, "Performance Analysis for Sparse Support Recovery," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, March. 2010.

[8] M. E. Davies and Y. C. Eldar, "Rank awareness for joint sparse recovery," arXiv:1004.4529, 2010.

[9] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, pp. 21 – 30, March. 2008.

[10] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann. Statist.*, vol. 28, no. 5, pp. 1303-1339, Oct. 2000.

[11] S.J. Park and Heung-No Lee, "Proofs for Number of Compressed Measurements Needed for Noisy Distributed Compressed Sensing," http://infonet.gist.ac.kr/twiki/pub/Main/Paper/proofs_for_noisy_distributed_compressed_sensing.pdf.

APPENDIX

In Appendix, we always note that $\delta > 0$, $M > K$, $\mathcal{J}$ is one of the incorrect support set and $\mathcal{I}$ is the support set.

A. Proofs of Lemma 1 and 2

Let us consider $\mathbf{Q}(\mathbf{F}_{s,\mathcal{I}})\mathbf{r}_s$ and $\mathbf{Q}(\mathbf{F}_{s,\mathcal{J}})\mathbf{r}_s$ respectively. They can be expressed as

$$\begin{aligned}\mathbf{Q}(\mathbf{F}_{s,\mathcal{I}})\mathbf{r}_s &= \mathbf{F}_s\mathbf{x}_s + \mathbf{n}_s - \left(\mathbf{F}_{s,\mathcal{I}}\left(\mathbf{F}_{s,\mathcal{I}}^{\mathrm{T}}\mathbf{F}_{s,\mathcal{I}}\right)^{-1}\mathbf{F}_{s,\mathcal{I}}^{\mathrm{T}}\right)(\mathbf{F}_s\mathbf{x}_s + \mathbf{n}_s) \\ &= \mathbf{Q}(\mathbf{F}_{s,\mathcal{I}})\mathbf{n}_s\end{aligned},$$

and

$$\begin{aligned}\mathbf{Q}(\mathbf{F}_{s,\mathcal{J}})\mathbf{r}_s &= \mathbf{F}_s\mathbf{x}_s + \mathbf{n}_s - \left(\mathbf{F}_{s,\mathcal{J}}\left(\mathbf{F}_{s,\mathcal{J}}^{\mathrm{T}}\mathbf{F}_{s,\mathcal{J}}\right)^{-1}\mathbf{F}_{s,\mathcal{J}}^{\mathrm{T}}\right)(\mathbf{F}_s\mathbf{x}_s + \mathbf{n}_s) \\ &= \mathbf{Q}(\mathbf{F}_{s,\mathcal{J}})\left(\sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s(i)\mathbf{f}_s(i) + \mathbf{n}_s\right)\end{aligned},$$

where $x_s(i)$ is the i^{th} coefficient of the s^{th} signal and $\mathbf{f}_s(i)$ is the i^{th} column vector of the s^{th} sensing matrix.

After we decompose $\mathbf{Q}(\mathbf{F}_{s,\mathcal{I}})$ and $\mathbf{Q}(\mathbf{F}_{s,\mathcal{J}})$, then we get $\mathbf{U}_{s,\mathcal{I}}\sum_s\mathbf{U}_{s,\mathcal{I}}^{\mathrm{T}}$ and $\mathbf{V}_{s,\mathcal{J}}\sum_s\mathbf{V}_{s,\mathcal{J}}^{\mathrm{T}}$, where $\sum_s \in \Re^{M\times M}$ is a diagonal matrix with $M-K$ diagonal entries equal to one and the remaining entries equal to zero. $\mathbf{U}_{s,\mathcal{I}}$ is a unitary matrix depending on $\mathbf{f}_s(i)$ for $\forall i \in \mathcal{I}$. It is obvious that $\mathbf{U}_{s,\mathcal{I}}$ is independent of $\mathbf{n}_s$. Similarly, $\mathbf{V}_{s,\mathcal{J}}$ is also unitary matrix depending on $\mathbf{f}_s(i)$ for $\forall i \in \mathcal{J}$. However, they are not only independent of $\mathbf{n}_s$ and but also of the columns $\mathbf{f}_s(i)$ for $\forall i \in \mathcal{I}\setminus\mathcal{J}$. Now we compute $\left\|\mathbf{Q}(\mathbf{F}_{s,\mathcal{I}})\mathbf{r}_s\right\|^2$ and $\left\|\mathbf{Q}(\mathbf{F}_{s,\mathcal{J}})\mathbf{r}_s\right\|^2$. They are

$$\left\|\mathbf{Q}(\mathbf{F}_{s,\mathcal{I}})\mathbf{r}_s\right\|^2 = \left\|\Sigma_s\mathbf{n}_s\right\|^2, \tag{18}$$

and

$$\left\|\mathbf{Q}(\mathbf{F}_{s,\mathcal{J}})\mathbf{r}_s\right\|^2 = \left\|\Sigma_s(\mathbf{t}_s + \mathbf{n}_s)\right\|^2, \tag{19}$$

where $\mathbf{t}_s = \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s(i)\mathbf{f}_s(i)$ whose entries follows i.i.d. Gaussian $\mathcal{N}\left(0, \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s(i)^2\right)$.

For simplicity, we assume that the first $M-K$ diagonal entries of $\sum_s$ are one and remaining entries are zero. Then, both (18) and (19) can be rewritten as

$$\left\|\mathbf{Q}(\mathbf{F}_{s,\mathcal{I}})\mathbf{r}_s\right\|^2 = \sum_{i=1}^{M-K} n_s(i)^2, \tag{20}$$

and

$$\left\|\mathbf{Q}(\mathbf{F}_{s,\mathcal{J}})\mathbf{r}_s\right\|^2 = \sum_{i=1}^{M-K} c_s(i)^2 \tag{21}$$

where $n_s(i)$, which follows i.i.d. Gaussian $\mathcal{N}(0, \sigma_{\text{noise}}^2)$, is the i^{th} element of the s^{th} noise vector and $c_s(i) = t_s(i) + n_s(i)$ which follows i.i.d. Gaussian $\mathcal{N}\left(0, \sigma_{s,\mathcal{J}}^2 := \sigma_{\text{noise}}^2 + \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s(i)^2\right)$ owing to $\mathbf{t}_s$ and $\mathbf{n}_s$ are mutually independent. Since $n_s(i)$ and $c_s(i)$ are Gaussian, both (18) and (19) are chi-square random variables.

For simplicity, we let $Z_{s,1} := \sum_{i=1}^{M-K} n_s(i)^2/\sigma_{\text{noise}}^2$ and $Z_{s,2} := \sum_{i=1}^{M-K} c_s(i)^2$. Now, we consider $\Pr\left\{\mathrm{E}(\mathbf{r},\mathcal{I},\delta)^c\right\}$ and $\Pr\left\{\mathrm{E}(\mathbf{r},\mathcal{J}\neq\mathcal{I},\delta)\right\}$. From (3), we have

$$\begin{aligned}\Pr\left\{\mathrm{E}(\mathbf{r},\mathcal{I},\delta)^c\right\} &= \Pr\left\{\sum_s Z_{s,1} \geq S(M-K) + \frac{SM\delta}{\sigma_{\text{noise}}^2}\right\} \\ &\quad + \Pr\left\{\sum_s Z_{s,1} \leq S(M-K) - \frac{SM\delta}{\sigma_{\text{noise}}^2}\right\}\end{aligned}, \tag{22}$$

and

$$\Pr\left\{\mathrm{E}(\mathbf{r},\mathcal{J}\neq\mathcal{I},\delta)\right\} \leq \Pr\left\{\sum_s Z_{s,2} \leq SQ\right\}. \tag{23}$$

where $Q = (M-K)\sigma_{\text{noise}}^2 + M\delta$, $Z_{s,1}$ and $Z_{s,2}$ are the chi-square random variables of $M-K$ degrees of freedom respectively, their means are $M-K$ and $(M-K)\sigma_{s,\mathcal{J}}^2$, their variances are $2(M-K)$ and $2(M-K)\sigma_{s,\mathcal{J}}^2$ respectively. After defining $Z_1 := \sum_s Z_{s,1}$, then (22) is rewritten as

$$\begin{aligned}\Pr\left\{\mathrm{E}(\mathbf{r},\mathcal{I},\delta)^c\right\} &= \Pr\left\{Z_1 \geq S(M-K) + \frac{SM\delta}{\sigma_{\text{noise}}^2}\right\} \\ &\quad + \Pr\left\{Z_1 \leq S(M-K) - \frac{SM\delta}{\sigma_{\text{noise}}^2}\right\}\end{aligned}, \tag{24}$$

where Z_1 is also a chi-square random variable of $S(M-K)$ degrees of freedom since it is sum of independent chi-square random variables. Now we apply the Chernoff bound to obtain an upper bound on (24). We have

$$\Pr\left\{\mathrm{E}(\mathbf{r},\mathcal{I},\delta)^c\right\} \leq \frac{\mathbb{E}\left[\exp(\nu_1 Z_1)\right]}{\exp(\nu_1 S\lambda_1)} + \frac{\mathbb{E}\left[\exp(\nu_2 Z_1)\right]}{\exp(\nu_2 S\lambda_2)}, \tag{25}$$

where $\mathbb{E}\left[\exp(\nu_j Z_1)\right] = (1-2\nu_j)^{-\frac{S(M-K)}{2}}$, $\lambda_i = M-K+(-1)^{i-1}\frac{M\delta}{\sigma_{\text{noise}}^2}$. To optimize (25) with respect to ν_1 and ν_2 , we find them such that $\frac{d}{d\nu_i}\frac{\mathbb{E}\left[\exp(\nu_i Z_1)\right]}{\exp(\nu_i S\lambda_i)} = 0$ for $\forall i \in \{1,2\}$. It is easy to find ν_1 and ν_2 . They are

$$\nu_1 = \frac{1}{2}\left(1 - \frac{M-K}{\lambda_1}\right) > 0, \tag{26}$$

and

$$\nu_2 = \frac{1}{2}\left(1 - \frac{M-K}{\lambda_2}\right) < 0 . \tag{27}$$

Using (26) and (27), we finally obtain the upper bound on $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\}$. That is

$$\begin{aligned}\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\} &\le \exp\left(-\frac{SM\delta'}{2}\right)\left(1+\frac{M\delta'}{M-K}\right)^{S\frac{M-K}{2}} \\ &\quad + \exp\left(\frac{SM\delta'}{2}\right)\left(1-\frac{M\delta'}{M-K}\right)^{S\frac{M-K}{2}} \\ &\le 2\exp\left(-\frac{SM\delta'}{2}\right)\left(1+\frac{M\delta'}{M-K}\right)^{S\frac{M-K}{2}} \\ &=: p_c\left(T,\mathcal{I}\right)\end{aligned} . \tag{28}$$

where $T = \left\{S, M, K, \delta, \sigma_{\text{noise}}^2\right\}$, $\delta' = \delta/\sigma_{\text{noise}}^2$ and $p_c\left(T,\mathcal{I}\right)$ appears in (6). Last inequality appears due to

$$\exp\left(\tfrac{-SM\delta'}{2}\right)\left(1+\tfrac{M\delta'}{M-K}\right)^{\frac{S(M-K)}{2}} \ge \exp\left(\tfrac{SM\delta'}{2}\right)\left(1-\tfrac{M\delta'}{M-K}\right)^{\frac{S(M-K)}{2}} .$$

It completes the proof of Lemma 1. Now, we continue to prove Lemma 2. By applying the Chernoff bound to (23), we have

$$\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J}\neq\mathcal{I},\delta\right)\right\} \le \frac{\mathbb{E}\left[\exp\left(\nu_3 \sum_s Z_{s,2}\right)\right]}{\exp\left(\nu_3 S\lambda_3\right)}, \tag{29}$$

where $\lambda_3 = \left(M-K\right)\sigma_{\text{noise}}^2 + M\delta$ and $\mathbb{E}\left[\exp\left(\nu_3 Z_{s,2}\right)\right] = \left(1-2\nu_3\sigma_{s,\mathcal{J}}^2\right)^{-\frac{M-K}{2}}$. For simplicity, we take the further upper bound on (29). That is

$$\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J}\neq\mathcal{I},\delta\right)\right\} \le \frac{\max_{s\in\{1,\cdots,S\}}\left(\mathbb{E}\left[\exp\left(\nu_3 Z_{s,2}\right)\right]\right)^S}{\exp\left(\nu_3 S\lambda_3\right)}. \tag{30}$$

Similar to both (26) and (27), we aim to find ν_3 such that $\frac{d}{d\nu_3}\frac{\max_{s\in\{1,\cdots,S\}}\left(\mathbb{E}\left[\exp\left(\nu_3 Z_{s,2}\right)\right]\right)^S}{\exp\left(\nu_3 S\lambda_3\right)} = 0$. It is

$$\nu_3 = \frac{1}{2\sigma_{\min}^2}\left(1 - \frac{\left(M-K\right)\sigma_{\min}^2}{\lambda_3}\right) < 0, \tag{31}$$

thus, by using (31), (30) becomes

$$\begin{aligned}\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J}\neq\mathcal{I},\delta\right)\right\} &\le \left(\frac{\sigma_{\text{noise}}^2}{\sigma_{\min}^2} + \frac{M}{M-K}\frac{\delta}{\sigma_{\min}^2}\right)^{S\frac{M-K}{2}} \\ &\quad \times \exp\left(-\frac{SM}{2\sigma_{\min}^2}\left(\frac{M-K}{M}\left(\sigma_{\text{noise}}^2 - \sigma_{\min}^2\right) + \delta\right)\right) \\ &=: p_i\left(T,\mathcal{J}\right)\end{aligned}, \tag{32}$$

where $T = \left\{S, M, K, \delta, \sigma_{\text{noise}}^2\right\}$, $\sigma_{\min}^2 = \min_{s\in\{1,\cdots,S\}}\left(\sigma_{s,\mathcal{J}}^2\right)$, $\mathcal{J}$ is one of the incorrect support set and $p_i\left(T,\mathcal{J}\right)$ appears in (7). It completes the proof of Lemma 2.

B. Proof of Proposition 1

First, we aim to show that $p_c\left(T,\mathcal{I}\right)$ converges to zero as S increases. From (6), we get

$$\begin{aligned}p_c\left(T,\mathcal{I}\right) &= p_c\left(T^*,\mathcal{I}\right)^S \\ &= \left[\exp\left(-\frac{M\delta'}{2}\right)\times\left(1+\frac{M\delta'}{M-K}\right)^{\frac{M-K}{2}}\right]^S,\end{aligned}$$

where $T = \left\{S, M, K, \delta, \sigma_{\text{noise}}^2\right\}$, $T^* = \left\{S=1, M, K, \delta, \sigma_{\text{noise}}^2\right\}$ and $\delta' = \delta/\sigma_{\text{noise}}^2$. If we prove that $0 \le p_c\left(T^*,\mathcal{I}\right) < 1$, then, $p_c\left(T,\mathcal{I}\right)$ can converges to zero as S increases. Obviously, $0 < p_c\left(T^*,\mathcal{I}\right)$ is always true. Now, we aim to prove $p_c\left(T^*,\mathcal{I}\right) < 1$. By taking logarithm operator to both sides on $p_c\left(T^*,\mathcal{I}\right) < 1$, then we get

$$\log\left(1+t\right) < t, \tag{33}$$

where $t = \frac{M\delta'}{M-K}$. It is obvious that (33) is true whenever $t > 0$. Therefore, $p_c\left(T^*,\mathcal{I}\right) < 1$ is always true for any positive t. It implies that $p_c\left(T,\mathcal{I}\right)$ must converge to zero as S increases. Finally, $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\}$ also converges to zero.

Now, we examine the convergence rate of $p_c\left(T,\mathcal{I}\right)$. It is

$$\lim_{S\to\infty}\frac{p_c\left(T_1,\mathcal{I}\right)}{p_c\left(T,\mathcal{I}\right)} = \lim_{S\to\infty}\frac{p_c\left(T^*,\mathcal{I}\right)p_c\left(T,\mathcal{I}\right)}{p_c\left(T,\mathcal{I}\right)} = p_c\left(T^*,\mathcal{I}\right), \tag{34}$$

where $T = \left\{S, M, K, \delta, \sigma_{\text{noise}}^2\right\}$, $T^* = \left\{S=1, M, K, \delta, \sigma_{\text{noise}}^2\right\}$ and $T_1 = \left\{S+1, M, K, \delta, \sigma_{\text{noise}}^2\right\}$. Furthermore, $0 < p_c\left(T^*,\mathcal{I}\right) < 1$. Therefore, $p_c\left(T,\mathcal{I}\right)$ linearly converges to zero with rate $p_c\left(T^*,\mathcal{I}\right)$. It finally implies that $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\}$ also linearly converges to zero with rate $p_c\left(T^*,\mathcal{I}\right)$. It completes of the proof of Proposition 1.

C. Proof of Proposition 2

First, we aim to show that $p_i\left(T,\mathcal{J}\right)$ converges to zero as S increases. From (7), we get

$$p_i\left(T,\mathcal{J}\right)=p_i\left(T^*,\mathcal{J}\right)^S$$
$$=\left[\begin{array}{l}\exp\left(-\frac{M}{2\sigma_{\min}^2}\left(\frac{M-K}{M}\left(\sigma_{\text{noise}}^2-\sigma_{\min}^2\right)+\delta\right)\right)\\ \times\left(\frac{\sigma_{\text{noise}}^2}{\sigma_{\min}^2}+\frac{M}{M-K}\frac{\delta}{\sigma_{\min}^2}\right)^{\frac{M-K}{2}}\end{array}\right]^S,$$

where $T=\left\{S,M,K,\delta,\sigma_{\text{noise}}^2\right\}$, $T^*=\left\{S=1,M,K,\delta,\sigma_{\text{noise}}^2\right\}$, $\sigma_{\min}^2=\min_{s\in\{1,\cdots,S\}}\left(\sigma_{s,\mathcal{J}}^2\right)$, and $\sigma_{s,\mathcal{J}}^2=\sum_{i\in\mathcal{I}\setminus\mathcal{J}}x_s\left(i\right)^2+\sigma_{\text{noise}}^2$. If we prove that $0<p_i\left(T^*,\mathcal{J}\right)<1$, then, $p_i\left(T^*,\mathcal{J}\right)$ can converges to zero as S increases. Obviously, $0<p_i\left(T^*,\mathcal{J}\right)$ is always true. Now, we aim to prove $p_i\left(T^*,\mathcal{J}\right)<1$. By taking logarithm operator to both sides on $p_i\left(T^*,\mathcal{J}\right)<1$, then we get

$$\log\left(t\right)<t-1, \tag{35}$$

where $t=\frac{\sigma_{\text{noise}}^2}{\sigma_{\min}^2}+\frac{M}{M-K}\frac{\delta}{\sigma_{\min}^2}$. When either $t>1$ or $0<t<1$, (35) is true. Regions on t can be translated to regions on δ. They are

$$\delta>\frac{\left(M-K\right)\left(\sigma_{\min}^2-\sigma_{\text{noise}}^2\right)}{M}>0,$$

owing to $\sigma_{\min}^2=\min_s\left(\sum_{i\in\mathcal{I}\setminus\mathcal{J}}x_s\left(i\right)^2+\sigma_{\text{noise}}^2\right)$ and

$$0<\delta<\frac{\left(M-K\right)\left(\sigma_{\min}^2-\sigma_{\text{noise}}^2\right)}{M}.$$

By combining two regions, finally we have

$$\left(\delta>0\right)\cap\left(\delta\neq\frac{\left(M-K\right)\left(\sigma_{\min}^2-\sigma_{\text{noise}}^2\right)}{M}\right). \tag{36}$$

We note that we only represent $\delta>0$ in Proposition 2 because the probability that $\delta\neq\frac{\left(M-K\right)\left(\sigma_{\min}^2-\sigma_{\text{noise}}^2\right)}{M}$ occurs is very rare. Thus, we can say that $0<p_i\left(T^*,\mathcal{J}\right)<1$ is true for $\delta>0$. It implies that $p_i\left(T^*,\mathcal{J}\right)$ can converge to zero as S increases. $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J}\neq\mathcal{I},\delta\right)\right\}$ finally must converge to zero as S increases as well.

Now, we examine the convergence rate of $p_i\left(T^*,\mathcal{J}\right)$. It is

$$\lim_{S\to\infty}\frac{p_i\left(T_1,\mathcal{J}\right)}{p_i\left(T,\mathcal{J}\right)}=\lim_{S\to\infty}\frac{p_i\left(T^*,\mathcal{J}\right)p_i\left(T,\mathcal{J}\right)}{p_i\left(T,\mathcal{J}\right)}=p_i\left(T^*,\mathcal{J}\right), \tag{37}$$

where $T=\left\{S,M,K,\delta,\sigma_{\text{noise}}^2\right\}$, $T^*=\left\{S=1,M,K,\delta,\sigma_{\text{noise}}^2\right\}$ and $T_1=\left\{S+1,M,K,\delta,\sigma_{\text{noise}}^2\right\}$. Furthermore, $0<p_i\left(T^*,\mathcal{J}\right)<1$. Therefore, $p_i\left(T,\mathcal{J}\right)$ linearly converges to zero with rate $p_i\left(T^*,\mathcal{J}\right)$. It finally implies that $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J}\neq\mathcal{I},\delta\right)\right\}$ also linearly converges to zero with rate $p_i\left(T^*,\mathcal{J}\right)$. It completes of the proof of Proposition 2.