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 signal recovery at FC is investigated. This is done by studying the probability of signal recovery failure. The joint typical decoder (JT decoder) similar to the one proposed by Akcakaya and Tarokh is used to avoid dependence on particular choice of recovery routines. The following interesting results have been obtained: 1) The detection failure probability *linearly* converges to zero as the number of sensors increases. 2) The number of compressed measurements per sensor (PSM) needed for successful recovery converges to sparsity as the number of sensors increases.**

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 intends to reconstruct each individual signal perfectly. The 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 is the sparsity K. A sparse signal $\mathbf{x}$, i.e., a correlated signal, can be compressively sampled, via a linear projection system, i.e., $\mathbf{y} = \mathbf{F}\mathbf{x}$ where $\mathbf{F} \in \Re^{M \times N}$ is 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 a portion of sensors having the same *support set*. Intra-sensor correlations, in contrast, are signal correlations that exist within 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 the signals in the collection. A good 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 support of the jointly sparse signals, as the number of sensors increases. A jointly typical decoder (JT decoder) similar to the ones in [2][4] is used here 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 prove that the detection failure probability *linearly* converges to zero as the number of sensors increases, which show that PSM converges to sparsity as the number of sensors increases.

II. SYSTEM MODEL

There are S sensors are distributed in a limited region. Each sensor compressively measures the signal coming to its way in the fashion of CS and transmits a set of acquired samples to FC. Let the original signal being acquired at the s^{th} sensor be denoted as $\mathbf{x}_s \in \Re^N$ with $\|\mathbf{x}_s\|_0 = K$, $s \in \{1,2,\cdots,S\}$. The support set of sparse signal $\mathbf{x}$ is defined as

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

We assume that each sparse signal has the same support set in this paper, i.e., $\mathcal{I} = \mathcal{I}(\mathbf{x}_1) = \cdots = \mathcal{I}(\mathbf{x}_S)$. 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 distribution $\mathcal{N}(0,1)$. The matrix $\mathbf{F}_s$ denotes the sensing matrix of the s^{th} sensor. All the compressed signals are transmitted to the FC via the AWGN channel. The received signal at FC is

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

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

Our signal model encompasses the signal models used in previous works [6],[7]. In the both works, the assumption that each sparse signal has the same support set is used. On the one hand, the model in Duarte *et al.* [6] contains no observation noise and thus is equivalent to our model (2) when all the noise vectors are ignored. On the other hand, the model in Tang and Nehorai [7] is obtained when all the sensing matrices are assumed to be the same in (1).

III. JOINT TYPICAL (JT) DECODER AND EVENT

In Akcakaya and Tarokh [4], a JT decoder is used to show that the number of measurements required for reliable support set detection is $\mathrm{O}(K \log(N/K))$, a sufficient condition asymptotic to the signal length N. This JT decoder was inspired from the classic work of Shannon's channel coding theorem. The JT decoder is a fictitious decoder defined on the rare occurrence of *atypical* detection error events. As the word "atypical" indicates, the probability of occurrence of such an event is small and in fact vanishes as the signal length increases due to the law of large numbers. Our JT decoder defined in this paper is slightly different from that of Akcakaya and Tarokh. This was done with an aim to streamline our analysis in such a way so as to better focus on our goal of investigating the impact of using multiple sensors on the number of measurements.

Let us now formally introduce our joint typical (JT) decoder. It consists of two different parts: 1) the *Support Set Detection* (*SSD*) part and 2) the *Signal Estimation* (SE) part. The aim of SE is to compute all of the signal coefficients. It is well known that the task of the SE part is trivial once the support set is known. Namely, the most critical part of JT decoder is the SSD part. Motivated from this observation, we begin by defining the notion of $\delta-$ *Joint Typicality.*

Definition 1: ($\delta-$*Joint Typicality*) *We say that an* $M \times S$ *matrix* $\mathbf{r}$ *and an index set* $\mathcal{J} \subset \{1,2,\cdots,N\}$ *with* $|\mathcal{J}| = K$ *are* $\delta-$*jointly typical if* $rank(\mathbf{F}_{s,\mathcal{J}}) = K$ *for all* s *and*

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

where $\mathbf{Q}(\mathbf{F}) = \mathbf{I} - \mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}$, $\delta > 0$ *and* $(\cdot)^{\mathrm{T}}$ *represents the matrix transposition*.

Henceforth, we denote $\mathrm{E}(\mathbf{r}, \mathcal{J}, \delta)$ to mean that $\mathbf{r}$ and $\mathcal{J}$ are a $\delta-$jointly typical event.

Let $\mathrm{E}(D_{failure})$ be the detection failure event. Then,

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

There are three kinds of decoding failures incorporated in (4). The first one $\mathrm{E}(\mathbf{r}, \mathcal{I}, \delta)^C$ is to imply the case when the JT decoder makes failure not being able to declare that the correct support set is $\delta-$ *jointly typical* with the receive signal. In the second kind of events, i.e., $\mathrm{E}(\mathbf{r}, \mathcal{J} \neq \mathcal{I}, \delta)$, the JT decoder declares that an incorrect support set is $\delta-$*jointly typical* with the receive signal. The third is the case when $rank(\mathbf{F}_{s,\mathcal{J}}) < K$ for some s in which (3) cannot even be defined. The possibility that $\mathrm{E}(rank(\mathbf{F}_{s,\mathcal{J}}) < K)$ occurs is very small. Hence, it can be ignored.

IV. PROBABILITES OF THE FAILURE EVENTS

Now, we aim to discuss the probability of the detection failure event. It is hard, we note, to obtain the exact detection failure probability due to dependency structure present amongst different error events. Hence, we turn to the union bound approach with which we have an upper bound on the probability

$$\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\} \le \Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\} + \sum_{\forall \mathcal{J}\neq\mathcal{I},|\mathcal{J}|=K} \Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J},\delta\right)\right\}. \quad (5)$$

Obtaining exact expression for the probabilities such as $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{I},\delta\right)^c\right\}$ and $\Pr\left\{\mathrm{E}\left(\mathbf{r},\mathcal{J},\delta\right)\right\}$ in (5), is again difficult. Further upper bounding on both kinds of probabilities is made using the Chernoff bounds. In fact, it is exciting to find out that these upper bounds eventually came out tight enough to investigate the behavior of the detection failure probability and provided answers to the research questions poised in this paper.

The two upper bounds are given below as Lemma 1 and Lemma 2 respectively. The following notations become useful for representing both upper bounds:

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

and

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

where $\mathcal{T}=\left\{S,M,K,\delta,\sigma_{\text{noise}}^2\right\}$, $\sigma_{\min}^2 = \min_{s\in\{1,\cdots,S\}}\left(\sigma_{s,\mathcal{J}}^2\right)$, $\delta^{'}=\delta/\sigma_{\text{noise}}^2$ and $\sigma_{s,\mathcal{J}}^2 = \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s\left(i\right)^2+\sigma_{\text{noise}}^2$.

Lemma 1: *Let an index set $\mathcal{I}$ 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*

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

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

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

The detailed proofs for both lemmas are given in [10].

V. Results And Discusson

In the next section, we will discuss asymptotic behavior of both $p_c\left(\mathcal{T},\mathcal{I}\right)$ and $p_i\left(\mathcal{T},\mathcal{J}\right)$ with respect to S in two propositions and summarize our main result in Theorem 1.

Proposition 1: *Let $M>K$, an index set $\mathcal{I}$ 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 $p_c\left(\mathcal{T}^*,\mathcal{I}\right)$ as S increases.*

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

In both propositions, we use $\mathcal{T}^* := \left\{S=1,M,K,\delta,\sigma_{\text{noise}}^2\right\}$. The detailed proofs for both propositions are given in [10].

Now, we use both Proposition 1 and Proposition 2 to make Theorem 1. Clearly, each term in the right hand side of the upper bound in (5) can be further upper bounded by $2p_c\left(\mathcal{T},\mathcal{I}\right)$ and $\sum_{\forall\mathcal{J}\neq\mathcal{I},|\mathcal{J}|=K} p_i\left(\mathcal{T},\mathcal{J}\right)$ respectively using Lemma 1 and 2. Owing to Propositions 1 and 2, then, the upper bound converges *linearly* to zero as S increases. Hence, the detection failure probability converges *linearly* to zero as S increases. This result is established as Theorem 1.

Theorem 1: *Let $M>K$, $\mathcal{I}$ be the correct support set, $\mathcal{J}\subset\{1,\cdots,N\}$ with $\left|\mathcal{J}\right|=K$ 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 $\sigma_{\text{noise}}^2 < \min_{s\in\{1,\cdots,S\}} \sum_{i\in\mathcal{I}\setminus\mathcal{J}} x_s\left(i\right)^2$. Then, $\Pr\left\{\mathrm{E}\left(D_{failure}\right)\right\}$ linearly converges to zero with a rate, $\max\left(p_c\left(\mathcal{T}^*,\mathcal{I}\right),p_i\left(\mathcal{T}^*,\mathcal{J}^*\right)\right)$, as S increases.*

The detailed proof of the theorem is given in [10].

Theorem 1 says that as long as the PSM *M* is greater than *K*, then by increasing the number of sensors the detection failure probability can be made to converge to zero. It is noted that for a fixed number of sensors the condition given here, *viz.* $M>K$, may not be sufficient for convergence. Another line worthy of notice is that there is a condition on the minimal signal value included among the set of sufficient conditions.

There are results similar to ours reported in the literature. Duarte *et al.* [6] proved and demonstrated that M converges to $K+1$. Difference to our work is that they did not consider the presence of noise. Tang and Nehorai [7] proved $M\ge 2K$ for correct support set recovery from compressed signals obtained over an AWGN channel. It is mentioned in [7] that they can show that M converges to $K+1$ using Theorem 3 in their paper. From what is written in Theorem 3 of [7], however, it is difficult to draw $M\ge K+1$, at least not at first hand. Davies and Eldar [8] designed a practical algorithm to recover K sparse signals from the MMV model, i.e., $\mathbf{y}=\mathbf{Fx}$, without considering noises. Their simulation results 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 a benchmark independently usable of any practical recovery algorithms. For example, given the systems parameters, the behavior of detection failure probability of the DCN system can be studied immediately.

VI. CONCLUSIONS

The main focus of this paper was to investigate how many PSM is needed for almost perfect support set recovery, as the number of sensors increases, in DCN systems. For this objective, we obtained a series of upper bounds on the detection failure probability. Using them, we proved that the upper bound *linearly* converges to zero as S increases in Theorem 1 and showed that PSM converges to sparsity as the number of sensors increases.

Proofs for Theorem, propositions and lemmas are relegated to the technical report in [10].

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. 2012-0005656)

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] S.J. Park and Heung-No Lee, "INFONET Technical Report: Proofs for Number of Compressed Measurements Needed for Noisy Distributed Compressed Sensing," http://infonet.gist.ac.kr/twiki/pub/Main/Paper/proofs_for_noisydistributed_compressed_sensing.pdf, Feb. 2012.