Necessary and Sufficient Conditions for Recovery of Sparse Signals over Finite Fields

Jin-Taek Seong, *Student Member, IEEE,* and Heung-No Lee, *Senior Member, IEEE*

***Abstract*—We consider a compressed sensing (CS) framework over finite fields. We derive sufficient and necessary conditions for recovery of sparse signals in terms of the ambient dimension of the signal space, the sparsity of the signal, the number of measurements, and the field size. We show that the sufficient condition coincides with the necessary condition if the sensing matrix is sufficiently dense while both the length of the signal and the field size grow to infinity. One of the interesting conclusions includes that unless the signal is very sparse, the sensing matrix does not have to be dense to have the upper bound coincide with the lower bound.**

***Index Terms*—L_0 norm minimization, compressed sensing, finite fields.**

I. Introduction

RECENTLY, compressed sensing (CS) theory has emerged as a new paradigm for signal acquisition in which compression and sampling of signals can be done simultaneously, introduced in the signal processing and information theory, such as Candes and Tao [1] and Donoho [2]. One of the main issues in the CS problems has been to quantify how many measurements are needed for perfect recovery of unknown signals. The most surprising and interesting discovery is that perfect recovery is possible with the number of measurements much smaller than the ambient dimension of the unknown signal as long as the signal is sufficiently sparse in a certain domain.

In general, the problems of CS have been considered mainly over the field of real and complex numbers. One of the key points in CS problems is to minimize the number of measurements while unknown signals are perfectly recovered. There are some applications that this CS problem over finite fields can be useful, including, *i*) the problems of collecting data samples from a group of correlated sources [3], [4], *ii*) group testing [5] , *iii*) the problem of sensor failure detection [6], and *iv*) minimization of file servers in order to complete download in file sharing networks [7]. For instance, in [3], Bassi *et al.* addressed the problem of the collecting spatially correlated measurements in a wireless sensor network. All sensors quantize their measured values, and map them to q-level symbols. A sink receives coded packets which are linear combinations of source packets over Galois fields.

Manuscript received April 1, 2013. The associate editor coordinating the review of this letter and approving it for publication was M. Wigger.

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (Do-Yak Research Program, No. 2013-035295, and Haek-Sim Reserach Program, No. 2012-047744).

The authors are with the Department of Information and Communications, Gwangju Institute of Science and Technology (GIST), South Korea (e-mail: heungno@gist.ac.kr).

Digital Object Identifier 10.1109/LCOMM.2013.13.130753

In this paper, we aim to investigate the core question of CS problems, but for the CS systems over finite fields where the sparse signals, the sensing matrix, and the measurements are made of the elements from a finite field. We use the ideal L_0 norm (which is equal to the Hamming weight in coding theory) minimization with a goal of providing benchmark to performance of any practical recovery routines. We first derive an upper bound on the probability of error. By using Fano's inequality [8], we derive a lower bound. We show that the upper bound and the lower bound converge with each other if the sensing matrix is sufficiently dense.

II. Compressed Sensing over Finite Fields

We describe the CS framework over a finite field of size q, $\mathbb{F}_q$: Let $\boldsymbol{x} \in \mathbb{F}_q^N$ be a sparse signal of length N with sparsity k_1 which indicates the number of nonzero entries in $\boldsymbol{x}$, $k_1 \in \{1, 2, \ldots, K\}$, where K, $2K \leq N$, denotes the maximum number of nonzero entries in $\boldsymbol{x}$. Let $\mathcal{L}$ denote the set of sparse signals, i.e., $\mathcal{L} := \bigcup_{k_1=1}^{K} \mathcal{L}_{k_1}$ where $\mathcal{L}_{k_1}$ denotes the set of signals $\boldsymbol{x}$ of length N with sparsity k_1. The size of the set $\mathcal{L}$ is given by $|\mathcal{L}| = \sum_{k_1=1}^{K} \binom{N}{k_1}(q-1)^{k_1}$ where $|\cdot|$ denotes the cardinality of the set. A sparse signal $\boldsymbol{x}$ is randomly and uniformly selected from the set $\mathcal{L}$. Let $\boldsymbol{A} \in \mathbb{F}_q^{M \times N}$ be an $M \times N$ sensing matrix with $N > M$. The measured signal $\boldsymbol{y}$ is given by

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{1}$$

We assume that the elements of the sensing matrix $\boldsymbol{A}$ are independent and identically distributed (i.i.d.), so that

$$\Pr\{A_{ij} = \alpha\} = \begin{cases} 1-\gamma, \text{ if } \alpha = 0 \\ \gamma/(q-1), \text{ if } \alpha \neq 0 \end{cases} \tag{2}$$

where γ denotes the sparse factor which is the probability that an element of the sensing matrix has nonzero values, and A_{ij} denotes the element of the ith row and the jth column of the sensing matrix, for $i \in \{1, 2, \ldots, M\}$ and $j \in \{1, 2, \ldots, N\}$, and α denotes a dummy variable, i.e., $\alpha \in \mathbb{F}_q$.

III. Upper and Lower Bounds for Recovery Performance

A. Probability of error for L_0 norm minimization

In this section, we aim to derive an upper and a lower bound for recovery of sparse signals in a CS framework for given parameters, i.e., N, K, M, and γ. We assume that the decoder in our scheme finds the sparsest feasible solution $\hat{\boldsymbol{x}}$ using the L_0 norm minimization as follows,

$$(\mathbf{P}_0) \quad \hat{\boldsymbol{x}} = \min \|\bar{\boldsymbol{x}}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\bar{\boldsymbol{x}} = \boldsymbol{y}, \tag{3}$$

where $\bar{\boldsymbol{x}} \in \mathcal{L}$ is a feasible solution. Let k_2 be the sparsity of $\bar{\boldsymbol{x}}$ as $k_2 := \|\bar{\boldsymbol{x}}\|_0$.

For a given $\boldsymbol{x}$, the decision $\hat{\boldsymbol{x}}$ is a function of the random matrix $\boldsymbol{A}$. Let us define two sets of matrices, $\mathcal{E}_0(\boldsymbol{x}) := \{\boldsymbol{A} : \boldsymbol{x} \neq \hat{\boldsymbol{x}}\}$ and $\mathcal{E}(\boldsymbol{x}, \bar{\boldsymbol{x}}) := \{\boldsymbol{A} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\bar{\boldsymbol{x}}\}$. Given these definitions, an error is then said to occur when a realized sensing matrix belongs to the set $\mathcal{E}_0(\boldsymbol{x})$, i.e., $\boldsymbol{A} \in \mathcal{E}_0(\boldsymbol{x})$. Note the following inclusion: $\mathcal{E}_0(\boldsymbol{x}) \subseteq \bigcup_{\bar{\boldsymbol{x}} \in \mathcal{L}, \boldsymbol{x} \neq \bar{\boldsymbol{x}}} \mathcal{E}(\boldsymbol{x}, \bar{\boldsymbol{x}})$. Let $\Pr\{\boldsymbol{x} \neq \hat{\boldsymbol{x}}\}$ be the probability of error averaged over all $\boldsymbol{x}$. We consider $\Pr\{\boldsymbol{x} \neq \hat{\boldsymbol{x}}\} = \sum_{\boldsymbol{x} \in \mathcal{L}} \Pr\{\boldsymbol{A} \in \mathcal{E}_0(\boldsymbol{x}) | \boldsymbol{x}\} \Pr\{\boldsymbol{x}\}$. Then, the probability of error is upper bounded by the inclusion as follows:

$$
\begin{aligned}
\Pr\{\boldsymbol{x} \neq \hat{\boldsymbol{x}}\} &\leq \frac{1}{|\mathcal{L}|} \sum_{\boldsymbol{x} \in \mathcal{L}} \Pr\Big\{\boldsymbol{A} \in \bigcup_{\bar{\boldsymbol{x}} \in \mathcal{L}, \boldsymbol{x} \neq \bar{\boldsymbol{x}}} \mathcal{E}(\boldsymbol{x}, \bar{\boldsymbol{x}}) \Big| \boldsymbol{x}\Big\} \\
&\overset{(a)}{\leq} \frac{1}{|\mathcal{L}|} \sum_{\boldsymbol{x} \in \mathcal{L}} \sum_{\substack{\bar{\boldsymbol{x}} \in \mathcal{L} \\ \boldsymbol{x} \neq \bar{\boldsymbol{x}}}} \Pr\Big\{\boldsymbol{A} \in \mathcal{E}(\boldsymbol{x}, \bar{\boldsymbol{x}}) \Big| \boldsymbol{x}\Big\} \\
&\overset{(b)}{=} \frac{1}{|\mathcal{L}|} \sum_{\boldsymbol{x} \in \mathcal{L}} \sum_{h=1}^{2K} \sum_{\bar{\boldsymbol{x}} \in \bar{\mathcal{L}}_h(\boldsymbol{x})} \Pr\Big\{\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\bar{\boldsymbol{x}} \Big| \boldsymbol{x}\Big\} \\
&\overset{(c)}{=} \frac{1}{|\mathcal{L}|} \sum_{\boldsymbol{x} \in \mathcal{L}} \sum_{h=1}^{2K} \Big|\bar{\mathcal{L}}_h(\boldsymbol{x})\Big| \Pr\Big\{\boldsymbol{A}\boldsymbol{d}_h = 0\Big\} \qquad (4) \\
&= \frac{1}{|\mathcal{L}|} \sum_{h=1}^{2K} \sum_{k_1=1}^{K} \sum_{\boldsymbol{x} \in \mathcal{L}_{k_1}} \Big|\bar{\mathcal{L}}_h(\boldsymbol{x})\Big| \Pr\Big\{\boldsymbol{A}\boldsymbol{d}_h = 0\Big\} \\
&= \frac{1}{|\mathcal{L}|} \sum_{h=1}^{2K} \Pr\Big\{\boldsymbol{A}\boldsymbol{d}_h = 0\Big\} \sum_{k_1=1}^{K} \binom{N}{k_1} (q-1)^{k_1} \Big|\bar{\mathcal{L}}_h(\boldsymbol{x})\Big| \\
&\overset{(d)}{=} \frac{1}{|\mathcal{L}|} \sum_{h=1}^{2K} N_h \Pr\Big\{\boldsymbol{A}\boldsymbol{d}_h = 0\Big\},
\end{aligned}
$$

where the inequality (a) is due to the union bound, and (b) is due to partition of the set $\{\bar{\boldsymbol{x}} \in \mathcal{L}\}$ with respect to the Hamming weight h, i.e., $\bar{\mathcal{L}}_h(\boldsymbol{x}) := \big\{\bar{\boldsymbol{x}} \in \mathcal{L} : \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_0 = h\big\}$, for $h = 1, 2, \ldots, 2K$. For the equality (c), we will show shortly that for each $\bar{\boldsymbol{x}} \in \bar{\mathcal{L}}_h(\boldsymbol{x})$, the probability is identically the same with each other, i.e., $\Pr\big\{\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\bar{\boldsymbol{x}} \big| \boldsymbol{x}\big\} = \Pr\big\{\boldsymbol{A}\boldsymbol{d}_h = 0\big\}$, where $\boldsymbol{d}_h := \boldsymbol{x} - \bar{\boldsymbol{x}}$ denotes a difference vector with the Hamming weight h. Before moving on, please note that $\Pr\big\{\boldsymbol{A}\boldsymbol{d}_h = 0\big\} = \prod_{i=1}^{M} \Pr\big\{A_i \boldsymbol{d}_h = 0\big\}$ where A_i denotes the ith row of $\boldsymbol{A}$ since the elements of $\boldsymbol{A}$ are i.i.d.

For example, let us take $h = 1$. Then, it is easy to show $\Pr\big\{A_{i1}\beta_1 = 0\big\} = \Pr\big\{A_{i1} = 0\big\}$ for any $\beta_1 \in \mathbb{F}_q \backslash \{0\}$ since it follows the property of multiplication over finite fields. Thus, $\Pr\big\{\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\bar{\boldsymbol{x}} \big| \boldsymbol{x}\big\} = \Pr\big\{\boldsymbol{A}\boldsymbol{d}_1 = 0\big\}$ for each $\bar{\boldsymbol{x}} \in \bar{\mathcal{L}}_1(\boldsymbol{x})$ regardless of position of the nonzero entry in $\boldsymbol{d}_1$. For $h = 2$ and two nonzero elements $\beta_1, \beta_2 \in \mathbb{F}_q \backslash \{0\}$, the following holds: $\Pr\big\{A_{i1}\beta_1 + A_{i2}\beta_2 = 0\big\} = \sum_{\alpha \in \mathbb{F}_q} \Pr\big\{A_{i1}\beta_1 = \alpha, A_{i2}\beta_2 = -\alpha\big\} = \sum_{\alpha \in \mathbb{F}_q} \Pr\big\{A_{i1} = \alpha\beta_1^{-1}\big\} \Pr\big\{A_{i2} = -\alpha\beta_2^{-1}\big\}$. It is trivial to show $A_{i1} = A_{i2} = 0$ for $\alpha = 0$. A little tricky is the case for any $\alpha \in \mathbb{F}_q \backslash \{0\}$. But note that both $\alpha\beta_1^{-1}$ and $-\alpha\beta_2^{-1}$ are nonzero, thus from the probability distribution (2), $\sum_{\alpha \in \mathbb{F}_q \backslash \{0\}} \Pr\big\{A_{i1} = \alpha\beta_1^{-1}\big\} \Pr\big\{A_{i2} = -\alpha\beta_2^{-1}\big\} = \sum_{\alpha \in \mathbb{F}_q \backslash \{0\}} \Pr\big\{A_{i1} = \alpha\big\} \Pr\big\{A_{i2} = -\alpha\big\}$. Thus, $\Pr\big\{A_{i1}\beta_1 + A_{i2}\beta_2 = 0\big\} = \Pr\big\{A_{i1} + A_{i2} = 0\big\}$. For $3 \leq h \leq 2K$, we can show $\Pr\big\{\sum_{j=1}^{h} A_{ij}\beta_j = 0\big\} = \Pr\big\{\sum_{j=1}^{h} A_{ij} = 0\big\}$ using a recursion, i.e., $\Pr\big\{\sum_{j=1}^{h} A_{ij}\beta_j = 0\big\} = \sum_{\alpha \in \mathbb{F}_q} \Pr\big\{\sum_{j=1}^{h-1} A_{ij}\beta_j = \alpha\big\} \Pr\big\{A_{ih}\beta_h = -\alpha\big\}$.

The last equality (d) of (4) is due to the collection of difference vectors with the same Hamming weight, where N_h denotes the total number of difference vectors with $\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_0 = h$, i.e., $N_h = \sum_{k_1=1}^{K} \binom{N}{k_1} (q-1)^{k_1} |\bar{\mathcal{L}}_h(\boldsymbol{x})|$, which will be exactly figured out in Section III.B.

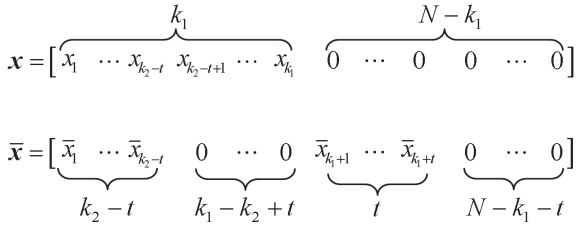


Fig. 1. One example for $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}$.

B. Upper bound

In this subsection, we aim to complete the derivation on the upper bound given in (4). Let P_h be denoted as $P_h := \Pr\big\{A_i \boldsymbol{d}_h = 0\big\} = \Pr\big\{\sum_{j=1}^{h} A_{ij} = 0\big\}$. Given the distribution (2) and noting $P_0 = 1$, P_h can be rewritten in a recursive form,

$$
\begin{aligned}
P_h &= \Pr\Big\{\sum_{j=1}^{h-1} A_{ij} = 0\Big\} \Pr\Big\{A_{ih} = 0\Big\} \\
&\quad + \sum_{\alpha \in \mathbb{F}_q \backslash \{0\}} \Pr\Big\{\sum_{j=1}^{h-1} A_{ij} = \alpha\Big\} \Pr\Big\{A_{ih} = -\alpha\Big\} \qquad (5) \\
&= P_{h-1}(1-\gamma) + (1 - P_{h-1}) \frac{\gamma}{q-1}.
\end{aligned}
$$

Let $Q_h := P_h - q^{-1}$. Then, the following equality can be obtained

$$
Q_h = Q_{h-1} \left(1 - \frac{\gamma}{1 - q^{-1}}\right). \qquad (6)
$$

Solving the recursion, a closed form expression for P_h is obtained

$$
P_h = q^{-1} + \left(1 - q^{-1}\right) \left(1 - \frac{\gamma}{1 - q^{-1}}\right)^h. \qquad (7)
$$

Next step is to compute N_h. For this, we use a combinatorial approach which is to enumerate all difference vectors into mutually exclusive groups with the same Hamming weight. Please see Figure 1 for counting N_h. Let us consider $\boldsymbol{x}$ in which the first k_1 elements are nonzero and the rest of the $N - k_1$ elements are zero, i.e., $\boldsymbol{x} = [x_1 x_2 \cdots x_{k_1} 0 \cdots 0]$ where x_j denotes the jth element of $\boldsymbol{x}$. Let the first and second index set denote $\{1, 2, \ldots, k_1\}$ and $\{k_1+1, k_1+2, \ldots, N\}$ respectively. Suppose that a candidate signal $\bar{\boldsymbol{x}}$ has k_2 nonzero entries in total. Among them, $t \in \{0, 1, \cdots, k_2\}$ nonzero elements are placed in the second index set of $\bar{\boldsymbol{x}}$. The rest $k_2 - t$ nonzero elements of $\bar{\boldsymbol{x}}$ are in the first index set as shown in Figure 1, where $\bar{x}_j$ denotes the jth element of $\bar{\boldsymbol{x}}$.

We enumerate all feasible signals $\bar{\boldsymbol{x}}$ with sparsity k_2 corresponding to the same Hamming weight h. It is to be noted that

TABLE I
THE NUMBER OF DIFFERENCE VECTORS $N_{h,k_1,k_2,t}$.

$N_{h,k_1,k_2,t}$	$t=0$	$t=1$	$t=2$	$t=3$
$h=1$	6	0	0	0
$h=2$	12	63	0	0
$h=3$	8	252	0	0
$h=4$	0	252	567	0
$h=5$	0	0	1134	0
$h=6$	0	0	0	945

for a given t, the Hamming weight of the difference vector is in the following range, i.e., $k_1 - k_2 + 2t \le h \le k_1 + t$. Given k_1, k_2, and t, the number of difference vectors with the Hamming weight h for $q > 2$ can be computed by

$$N_{h,k_1,k_2,t} = \binom{N-k_1}{t}(q-1)^t\binom{k_1}{k_2-t} \times \binom{k_2-t}{h-2t-k_1+k_2}(q-2)^{h-2t-k_1+k_2}, \tag{8}$$

where the first term $\binom{N-k_1}{t}(q-1)^t$ indicates the number of sequences of the length $N-k_1$ with t nonzero entries in the second set, and the second term $\binom{k_1}{k_2-t}\binom{k_2-t}{h-2t-k_1+k_2}(q-2)^{h-2t-k_1+k_2}$ indicates the number of sequences of the length k_1 having $h-2t-k_1+k_2$ nonzero entries. For $q=2$, the second term is $\binom{k_1}{k_2-t}$ by only considering binary sequences.

Example: Let us consider one example of counting $N_{h,k_1,k_2,t}$ where $N=10$, $k_1=3$, $k_2=3$ and $q=4$. There are $\binom{10}{3}(4-1)^3$ signal vectors with sparsity 3. We assume that the first 3 elements of $\boldsymbol{x}$ are nonzero, i.e., $\boldsymbol{x}=[1110000000]$, a feasible signal $\bar{\boldsymbol{x}}$ has t nonzero entries in the second set, i.e., for $t=2$, $\bar{\boldsymbol{x}}=[1001100000]$. In this example, the maximum t is 3, the Hamming weight of the difference vectors ranges from 1 to 6. Table 1 shows the number of difference vectors, $N_{h,k_1,k_2,t}$, with respect to t and h. **End of Example**

So far, we have found $N_{h,k_1,k_2,t}$ for given k_1, k_2, and t. Since we aim to find N_h, we take summation with respect to k_1, k_2, and t,

$$N_h = \sum_{k_1=1}^{K}\binom{N}{k_1}(q-1)^{k_1}\sum_{k_2=1}^{k_1}\sum_{t=0}^{k_2}N_{h,k_1,k_2,t}. \tag{9}$$

Substituting (7) and (9) into (4), we complete the derivation on the upper bound.

Theorem 1 (*Upper bound*). For any sensing matrix with the distribution (2), an upper bound on probability of error for the $\mathbf{P}_0$ problem is given by

$$\Pr\{\boldsymbol{x}\neq\hat{\boldsymbol{x}}\} \le \frac{1}{|\mathcal{L}|}\sum_{h=1}^{2K}\sum_{k_1=1}^{K}\binom{N}{k_1}(q-1)^{k_1} \times \sum_{k_2=1}^{k_1}\sum_{t=0}^{k_2}N_{h,k_1,k_2,t}P_h^M. \tag{10}$$

This result is general. For sparse sensing matrices, one may use the distribution given in (2) and obtain P_h from (7).

For dense sensing matrices, let $\gamma = 1-q^{-1}$ in (2); then, $\Pr\{A_{ij}=\alpha\} = q^{-1}$ for any $\alpha\in\mathbb{F}_q$ and the matrix becomes uniform random. In this special case, $P_h = q^{-1}$. Note there is no dependency on h. Thus, the upper bound (10) can be simplified as

$$\begin{aligned}\Pr\{\boldsymbol{x}\neq\hat{\boldsymbol{x}}\} &\le \frac{1}{|\mathcal{L}|}\sum_{h=1}^{2K}N_h q^{-M}\\ &\overset{(a)}{=}(|\mathcal{L}|-1)q^{-M}\\ &< Kq^{-M}\binom{N}{K}(q-1)^K\\ &\le 2^{\log_2 K - M\log_2 q + NH_b(K/N) + K\log_2(q-1)},\end{aligned} \tag{11}$$

where $H_b(\cdot)$ denotes the binary entropy function. The equality (a) originates from the fact that $\sum_{h=1}^{2K}N_h = (|\mathcal{L}|-1)|\mathcal{L}|$, which is the total number of sequences except for the original vector $\boldsymbol{x}$. Consequently, from the condition that the exponent of the R.H.S. of (11) remains negative so that the probability of error goes to 0 as $N\to\infty$, we derive the following sufficient condition on M,

$$M \ge \frac{\log_2 K + NH_b(K/N) + K\log_2(q-1)}{\log_2 q}. \tag{12}$$

Corollary 2 (*Sufficient condition on M*). Let $\gamma = 1-q^{-1}$. If (12) is satisfied, then $\Pr\{\boldsymbol{x}\neq\hat{\boldsymbol{x}}\}\to 0$ as $N\to\infty$.

C. Lower bound

Next, we aim to derive the lower bound on the probability of error for the $\mathbf{P}_0$ problem. For this, we use the Markov chain relation, a decision $\hat{\boldsymbol{x}}$ is made given $\boldsymbol{A}$ and $\boldsymbol{y}$, i.e., $\boldsymbol{x}\to(\boldsymbol{A},\boldsymbol{y})\to\hat{\boldsymbol{x}}$, a standard approach in information theory. Then, by the Fano's inequality, the probability of error is lower bounded as follows,

$$\Pr\{\boldsymbol{x}\neq\hat{\boldsymbol{x}}\} \ge \frac{H(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{A})-1}{\log_q|\mathcal{L}|} = \frac{H(\boldsymbol{x})-H(\boldsymbol{y}|\boldsymbol{A})-1}{\log_q|\mathcal{L}|}, \tag{13}$$

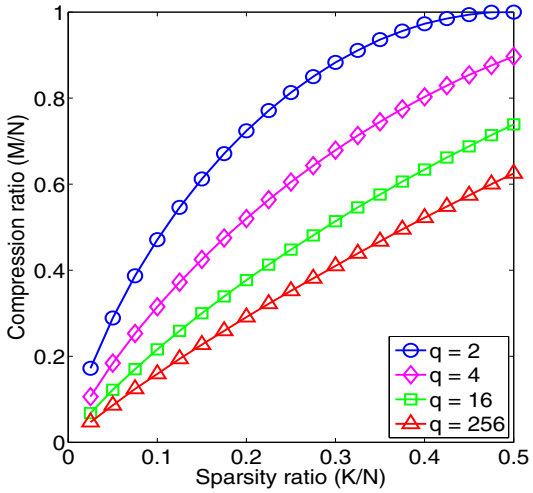
where $H(\cdot)$ denotes the entropy. According to the definition of conditional entropy, $H(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{A}) = H(\boldsymbol{x}) - I(\boldsymbol{x};\boldsymbol{y},\boldsymbol{A})$ where $I(\cdot)$ denotes the mutual information. Assuming that $\boldsymbol{A}$ is independent of $\boldsymbol{x}$, we have $I(\boldsymbol{x};\boldsymbol{y},\boldsymbol{A}) = I(\boldsymbol{x};\boldsymbol{y}|\boldsymbol{A})$. We use the following $I(\boldsymbol{x};\boldsymbol{y}|\boldsymbol{A}) = I(\boldsymbol{y};\boldsymbol{x}|\boldsymbol{A}) = H(\boldsymbol{y}|\boldsymbol{A}) - H(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{A})$. Since $\boldsymbol{y}$ is a function of $\boldsymbol{A}$ and $\boldsymbol{x}$, then $H(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{A}) = 0$, so that $H(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{A}) = H(\boldsymbol{x}) - H(\boldsymbol{y}|\boldsymbol{A})$. Since $H(\boldsymbol{y}|\boldsymbol{A}) \le H(\boldsymbol{y}) \le MH(y_1) \le M\log_q q = M$ and $\boldsymbol{x}$ is randomly and uniformly chosen from the set $\mathcal{L}$, we obtain the lower bound,

$$\Pr\{\boldsymbol{x}\neq\hat{\boldsymbol{x}}\} \ge 1 - \frac{M+1}{\log_q|\mathcal{L}|}. \tag{14}$$

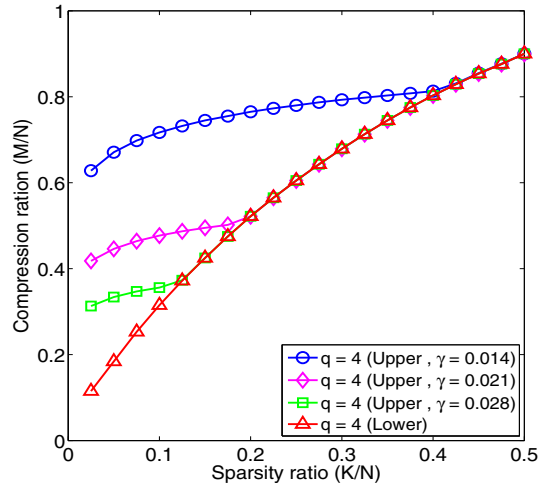
If the number of measurements is smaller than $\log_q|\mathcal{L}|-1$ in the R.H.S. of (14), the probability of error of the CS system is strictly away from and greater than or equal to the positive number in the R.H.S. of (14). This means that the negated condition, $M > \log_q|\mathcal{L}|-1$, is a necessary condition for an unboundedly arbitrary probability of error. Note the following inequlities, $\log_q|\mathcal{L}| > \log_q\binom{N}{K} + \log_q(q-1)^K \ge \log_q\frac{2^{NH_b(K/N)}}{N+1} + \log_q(q-1)^K$.

Theorem 3 (*Necessary condition on M*). For an arbitrarily small probability of error, the following

$$M > \frac{NH_b(K/N) + K\log_2(q-1) - \log_2(N+1)}{\log_2 q}, \tag{15}$$



(a)



(b)

Fig. 2. (a) Lower bounds for $N = 1000$ (note that if N is sufficiently large and $\gamma = 1 - q^{-1}$, the upper and lower bounds coincide with each other). (b) Upper and lower bounds with different sparse factors for $N = 1000$ and $q = 4$. In the region above the upper bound, the probability of error is less than 10^{-2}, while in the region below the lower bound, the probability of error is greater than 10^{-2}.

is a necessary condition.

Furthermore, in the limit case, we prove that $M > K$ is necessary and sufficient for successful recovery by solving $\mathbf{P}_0$ problem. To do this, from both Corollary 2 and Theorem 3, by dividing both sides of the inequalities by N, and we have $\frac{\log_q K}{N} \to 0$ in (12) and $\frac{\log_2(N+1)}{N} \to 0$ in (15) as $N \to \infty$ while the ratios M/N and K/N are fixed. In addition, when the field size goes to infinity, $q \to \infty$, then $\frac{H_b(K/N)}{\log_2 q} \to 0$ and $\frac{\log_2(q-1)}{\log_2 q} \to 1$. Thus, for both the necessary and the sufficient condition, we come to the following condition, $M > K$.

Corollary 4 (*Coincidence*). For fixed ratios M/N and K/N, as $N \to \infty$ and $q \to \infty$, the necessary, and the sufficient condition, for successful recovery of the K sparse signals over finite fields $\mathbb{F}_q$ is $M > K$.

IV. Numerical Results and Discussion

Figure 2 shows the compression ratio (=*M*/*N*) versus the sparsity ratio (=*K*/*N*) for recovery of a K sparse signal of length $N = 1000$ at the probability of error of 10^{-2}. We consider the following size finite fields: $q = 2$, 4, 16, and 256. Fixing K, we find the smallest integer M satisfying the upper (10) and the lower bound (14) at 10^{-2}. One interesting feature of Figure 2(a) is that for the lower bound, the compression ratio required for recovery of unknown sparse signals dramatically decreases as the field size grows. This result means that less number of measurements is needed for a larger finite field. In addition, the upper bound for uniform random sensing matrix is nearly identical with the lower bound.

In Figure 2(b), with respect to different sparse factor γ for a fixed field size, i.e., $q = 4$, we obtain the compression ratio which satisfies the upper bound at 10^{-2}. It can be observed that a higher value of sparse factor γ is required for recovery of very sparse signals. The aim of Figure 2(b) is to show that as the sparse factor of the sensing matrix increases, the upper bound approaches the lower bound even in the region of small sparsity ratios. Namely, if the sensing matrix is sufficiently dense, the upper bound coincides with the lower bound over finite fields. It is easy to see that if the signal and the sensing matrix are both sparse, the chance of making a zero measurement gets high; then the number of measurements needs to increase so as to compensate for missed sensing opportunities.

V. Conclusion

In this work, we considered a CS framework over finite fields. We derived the sufficient and necessary conditions for recovery of sparse signals. We showed that the both conditions are tight, and they coincide when the sparse factor of the sensing matrix is sufficiently large. We found that for recovery of ultra-sparse signals, the sensing matrix is required to be dense. One interesting result is that when the sensing matrix is sufficiently large and dense, and the field size is large, the number of measurements needed for perfect recovery is only $M > K$.

References

[1] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] F. Bassi, C. Liu, L. Iwaza, and M. Kieffer, "Compressive linear network coding for efficient data collection in wireless sensor networks," in *Proc. 2012 Euro. Sig. Proc. Conf.*, pp. 714–718.

[4] E. Bourtsoulatze, N. Thomos, and P. Frossard, "Correlation-aware reconstruction of network coded sources," in *Proc. 2012 Int. Sym. Network Coding*, pp. 91–96.

[5] M. Cheraghchi, A. Hormati, A. Karbasi, and M. Vetterli, "Group testing with probabilistic tests: theory, design and application," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 7057–7067, Oct. 2011.

[6] T. Tosic, N. Thomos, and P. Frossard, "Distributed sensor failure detection in sensor networks," *Signal Process.*, vol. 93, no. 2, pp. 399–410, Feb. 2013.

[7] C. Huimin, "Distributed file sharing: network coding meets compressed sensing," *2006 Intl. Conf. Commun. Net. in China*.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.