On the Derivation of RIP for Random Gaussian Matrices and Binary Sparse Signals

(*Invited Paper*)

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

***Abstract*—The number of measurements, M, sufficient for successful recovery via the L1 minimization is well known to be $M = O(K\log(N/K))$ [7] with Gaussian measurement matrices used for sensing K sparse signals of ambient dimension N. We aim to shed a light on the source of the $\log(N)$ factor, and see if the bound can be improved by considering it for simplest possible *K*-sparse signals--0/1 binary K sparse signals. Previous work exists with which it is reasonable to expect reduction in the number of measurements when the signal has smaller degrees of freedom. We derive an upper bound on the probability that any set of *K* randomly selected Gaussian column vectors are mutually independent; we use this to find an upper bound on the probability that a Gaussian sensing matrix satisfies the restricted isometry condition. Using this result, a sufficient condition for good signal recovery is found. Surprisingly, the result remains the same, i.e., $M = O(K\log(N/K))$, which may suggest the $\log(N)$ factor is generic for Gaussian measurements.**

Keywords-Compressive Sensing, Restricted Isometry Property, Binary Sparse Signal.

I. INTRODUCTION

Recently, the field of compressive sensing[1][2][3] has received considerable attention by many researchers. This compressive sensing is a new sampling method in which a sparse signal specified in a high ambient dimension can be recovered from a measurement much lower in dimension compared to the ambient dimension.

The data model in compressive sensing can be expressed as linear-system of equations. In fact, this linear model is an under-determined system. It is expressed as $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \Re^{M\times N}$ ($M < N$) , $\mathbf{x} \in \Re^{N}$ and $\mathbf{y} \in \Re^{M}$. The signal $\mathbf{x}$ in consideration is *K*-sparse signals, which means that the number of non-zero coefficients in the vector $\mathbf{x}$ is only K and $K \ll N$. Since the data model in compressive sensing is under-determined, there may exist multiple $\mathbf{x}$ s which provide the same $\mathbf{y}$. Among them, hence, a good criterion is aiming to seek the sparsest solution, i.e., one that minimizes the L0 norm of $\mathbf{x}$, $\|\mathbf{x}\|_0$, where $\|\mathbf{x}\|_0$ denotes the number of non-zero elements in $\mathbf{x}$. This is optimal but requires a combinatorial search. A suboptimal criterion, still good at promoting the sparseness in solution, is the L1 minimization.

Candes and Tao [4] obtained uniqueness conditions for the L1 minimization. They introduced the notion of Restricted Isometry Property (RIP). It is designed to provide a measure how well a given sensing matrix preserve the energy of any K sparse signals. In other words, a sensing matrix $\mathbf{A}$ satisfying RIP for a certain K implies that any set of K or less columns of $\mathbf{A}$ must be independent. Candes and Tao used it to specify the uniqueness condition. For example, the L1 minimization always recovers the sparsest solution if the RIP condition is met for any $3K$ sparse vectors.

However, checking to see if a sensing matrix satisfies the RIP is combinatorial and thus is an NP-hard problem. Rather than checking RIP for a particular matrix, therefore, they consider it for an ensemble of random matrices. Namely, how many measurements M , for a given N, are needed to have satisfaction of RIP with a high probability? The general answer to this question is $M = O(K\log(N/K))$ [7] when the matrices are made of i.i.d. Gaussian.

Meanwhile, there are a group of recent results [10][11] which utilize the idea of classical parity checking and syndrome decoding methods (error locator polynomial or annihilating filters) for Reed-Solomon codes and obtains a result that $2K$ measurements are enough to recover any K sparse signals. We make note of the fact that the number of measurements does not have the $\log(N)$ factor. There could be many factors affecting this favorable result, such as a better sensing matrix (Vandermonde frame), a good signal recovery scheme, and the finite size alphabet for signals and matrices.

In this paper, we aim to investigate the impact of the size of alphabet on the number of measurements needed for compressed sensing using Gaussian measurement matrices and recovery using the L1 minimization. To see the impact of the size of the alphabet, we assume the alphabet of the K sparse signal is binary, the simplest. We investigate how many measurements M , for a given N , are needed for the Gaussian matrix to satisfy the Restricted Isometry Property with high probability. This question is interesting as it can shed a light how the alphabet size affects the number of measurements.

It is reasonable to conjecture that the number of measurements can be reduced when the size of the alphabet is small. Bruckstein et. al said that the number of measurements can be reduced by adding the non-negative constraint on the K sparse signal since the constraint leads to reduce the size of the feasible set of solutions. They obtained an improved uniqueness condition in [9].

The rest of this paper is organized as follows. In II, the background related to compressive sensing is described. A system model is given in III. We describe and prove our analytical results in IV. Conclusions and future work are described in V.

II. BACKGROUND

A. The L0 minimization and the L1 minimization

The well known L0 and L1 minimization are defined as follows:

$$(L_0) \quad \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad s.t. \quad \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

and

$$(L_1) \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad s.t. \quad \mathbf{y} = \mathbf{A}\mathbf{x} \tag{2}$$

where $\mathbf{y} \in \Re^M$, $\mathbf{A} \in \Re^{M \times N}$ ($M < N$), $\mathbf{x} \in \Re^N$, $\|\mathbf{x}\|_0 = K$ and $\|\mathbf{x}\|_1 = \sum_{i=1}^{N} |x_i|$.

Both algorithms give the sparsest solution. However, the L0 minimization is NP hard. Instead of it, we use the L1 minimization.

B. The sufficient condition on the uniqueness of L1 minimization

Candes and Tao introduced the RIP constant in [4]. They used the RIP constant for guaranteeing the uniqueness of the L1 minimization. Definition of the RIP constant is as follows:

Definition 1: Let $\mathbf{A}$ be an $M \times N$ matrix. Let $K \ll N$ be an positive integer. Suppose that there exists a smallest constant $\delta_K \in (0,1)$ such that $\mathbf{A}$ satisfies the following condition for every K sparse signal $\mathbf{x}$,

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le (1+\delta_K)\|\mathbf{x}\|_2^2. \tag{3}$$

Then, the matrix $\mathbf{A}$ is said to satisfy the $K-$ restricted Isometry property with the RIP constant of order K, δ_K.

It is well known that the L0 minimization has the unique K sparse solution if $\delta_{2K} \in (0,1)$ [4]; the L1 minimization finds the unique K sparse solution if $\delta_{2K} < \sqrt{2}-1$ [5]. Furthermore, any K sparse signal is recovered by the L1 minimization if $\delta_{2K} < 2\left(3-\sqrt{2}\right)/7 \approx 0.4531$ [6], a slight improvement from the result of Candes and Tao [5].

III. SYSTEM MODEL

A. The binary K sparse signal and notations.

we define the binary K sparse signal as follows:

Definition 2: Let $\mathbf{x}$ be the binary K sparse signal. Then, $x_i = 1$ for $\forall i \in \operatorname{supp}(\mathbf{x})$, $x_i = 0$ for $\forall i \notin \operatorname{supp}(\mathbf{x})$.

The major features of the binary K sparse signal are that its squared L2 norm and L1 norm always have the value K and coefficients are either one or zero.

We assume that the measurement matrix $\mathbf{A}$ is randomly generated and its i,j^{th} element follows i.i.d. Gaussian distribution with zero-mean and variance $1/M$. The compressed measurement $\mathbf{y}$ is obtained by the linear projection, i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}$. Since $\mathbf{x}$ is the binary K sparse signal, $y_i = a_{i,\mathcal{I}[1]} + \cdots + a_{i,\mathcal{I}[K]}$ for $\forall i$, where $\mathcal{I} := \operatorname{supp}(\mathbf{x})$.

IV. ANALYSIS

Now, we aim to obtain an probabilistic upper bound on event $\left\{\left|\|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\right| \ge \delta\|\mathbf{x}\|_2^2\right\}$ for the given parameters. Carefully seeing it, we notice that $\|\mathbf{A}\mathbf{x}\|_2^2$ is a chi-square random variable. Hence, we first analyze this random variable. We note that we use $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ as the expectation and variance operation respectively.

Lemma 1 and Lemma 2, which provide the mean and variance of $\|\mathbf{A}\mathbf{x}\|_2^2$ respectively, are to used to drive the upper bound on the probability of event $\left\{\left|\|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\right| \ge \delta\|\mathbf{x}\|_2^2\right\}$ in Lemma 3. The reason why we have interest in the mean and variance of $\|\mathbf{A}\mathbf{x}\|_2^2$ is that we utilize them to obtain the Chernoff bound [8] in Lemma 3.

Lemma 1: Let $\mathbf{A}$ be a Gaussian matrix, each element of which follows i.i.d. Gaussian distribution with zero-mean and variance $1/M$. Let $\mathbf{x}$ be the binary K sparse signal. Then, $\mathbb{E}\left[\|\mathbf{A}\mathbf{x}\|_2^2\right] = \|\mathbf{x}\|_2^2$.

Proof: We write $y_i = \sum_{j=1}^{K} a_{i,j}$. Hence, $y_i^2 = \left(\sum_{j=1}^{K} a_{i,j}\right)^2$. Now, we take the expectation operator with respect to the distribution of $a_{i,j}$. We have $\mathbb{E}\left[y_i^2\right] = K/M$. Since the squared L2 norm of the vector $\mathbf{y}$ is $\sum_{i=1}^{M} y_i^2$, $\mathbb{E}\left[\|\mathbf{y}\|_2^2\right] = \sum_{i=1}^{M} \mathbb{E}\left[y_i^2\right] = M\left(\frac{K}{M}\right) = K$. We know $\|\mathbf{x}\|_2^2$ is K due to Definition 3. Hence, $\mathbb{E}\left[\|\mathbf{A}\mathbf{x}\|_2^2\right] = \|\mathbf{x}\|_2^2$. Q.E.D.

Using Lemma 1, we can rewrite $\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2$ as $\left|\|\mathbf{Ax}\|_2^2-\mathbb{E}\left[\|\mathbf{Ax}\|_2^2\right]\right|\geq\delta\|\mathbf{x}\|_2^2$. Next, we provide Lemma 2 for the variance of $\|\mathbf{Ax}\|_2^2$.

Lemma 2: Let $\mathbf{A}$ be a Gaussian matrix, each element of which follows i.i.d. Gaussian distribution with zero-mean and $1/M$ variance. Let $\mathbf{x}$ be the binary K sparse signal. Then, $\mathbb{V}\left[\|\mathbf{Ax}\|_2^2\right]=2K^2/M$.

Proof: From the definition of the variance, we have $\mathbb{V}\left[\|\mathbf{Ax}\|_2^2\right]=\mathbb{E}\left[\left(y_1^2+\cdots+y_M^2\right)^2\right]-\left[\mathbb{E}\left(y_1^2+\cdots+y_M^2\right)\right]^2$. Due to Lemma 1, we know $\left[\mathbb{E}\left(y_1^2+\cdots+y_M^2\right)\right]^2$ is K^2. Hence, we only consider $\mathbb{E}\left[\left(y_1^2+\cdots+y_M^2\right)^2\right]$ which can be expanded as $2\left(\mathbb{E}\left[y_1^2y_2^2\right]+\mathbb{E}\left[y_1^2y_3^2\right]+\cdots+\mathbb{E}\left[y_{M-1}^2y_M^2\right]\right)+\sum_i^M\mathbb{E}\left[y_i^4\right]$. y_i is a Gaussian random variable with zero mean and variance K/M. Hence, $\mathbb{E}\left[\mathbf{y}_1^2\right]=K/M$. $\mathbb{E}\left[\mathbf{y}_i^4\right]$ is just the fourth moment generating function of the random variable distributed as $\mathcal{N}\left(0,K/M\right)$. Therefore, $\mathbb{E}\left[\mathbf{y}_i^4\right]=3K^2/M^2$. We drive

$$\begin{aligned}\mathbb{V}\left[\|\mathbf{y}\|_2^2\right]&=\mathbb{E}\left[\left(\mathbf{y}_1^2+\cdots+\mathbf{y}_M^2\right)^2\right]-\mathbb{E}\left[\left(\mathbf{y}_1^2+\cdots+\mathbf{y}_M^2\right)\right]^2\\&=\sum_i^M\mathbb{E}\left[\mathbf{y}_i^4\right]+2\left(\mathbb{E}\left[\mathbf{y}_1^2\mathbf{y}_2^2\right]+\cdots\right)-K^2\\&=M\times\frac{3K^2}{M^2}+2\binom{M}{2}\frac{K^2}{M^2}-K^2\\&=\frac{3K^2}{M}+2\frac{M\left(M-1\right)}{2}\frac{K^2}{M^2}-K^2\\&=\frac{2K^2}{M}\end{aligned}.$$

Q.E.D.

Now, we know the mean and the variance of the random variable $\|\mathbf{Ax}\|_2^2$. In fact, the mean and the variance are easily obtained when we consider the random variable $\|\mathbf{Ax}\|_2^2$. Clearly, $y_i=a_{i,1}+\cdots+a_{i,K}$, hence y_i becomes the Gaussian random variable with zero mean and variance K/M. Therefore, y_i^2 becomes the chi-square random variable with 1 degree of freedom, K/M mean and variance $2K^2/M^2$. Therefore, $\|\mathbf{Ax}\|_2^2$, which is $\sum_{i=1}^M y_i^2$, becomes the chi-square random variable with M degree of freedom, K mean and variance $2K^2/M$. It proves Lemma 1 and Lemma 2 again.

We now have all information about the random variable $\|\mathbf{Ax}\|_2^2$. Hence, we can exactly compute the probability of the event $\left\{\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2\right\}$. In this paper, we aim to obtain an upper bound on the probability. Lemma 3 provides it.

Lemma 3: Let $\mathbf{A}$ be a Gaussian matrix, each element of which follows i.i.d. Gaussian distribution with zero-mean and variance $1/M$. Let $\mathbf{x}$ be a binary K sparse signal and $\delta\in\left(0,1\right)$. Then,

$$P\left(\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2\right)\leq2\exp\left(-\frac{M\delta}{2}\right)\left(\delta+1\right)^{\frac{M}{2}}. \quad (4)$$

Proof: From the Chernoff bound[8] it is well known that $P\left(\|\mathbf{Ax}\|_2^2\geq K\left(1+\delta\right)\right)\leq\mathbb{E}\left[\exp\left(t\left(\|\mathbf{Ax}\|_2^2-K\left(1+\delta\right)\right)\right)\right]$ for $t>0$, where $\mathbb{E}\left[\exp\left(t\|\mathbf{Ax}\|_2^2\right)\right]=\left(1-2tK/M\right)^{-M/2}$ since $\mathbb{E}\left[\exp\left(ty_i^2\right)\right]$ is $\left(1-2tK/M\right)^{-1/2}$ for $\forall i$. Hence, we have

$$P\left(\|\mathbf{Ax}\|_2^2\geq K\left(1+\delta\right)\right)\leq f\left(t\right):=\frac{\left(1-2tK/M\right)^{-M/2}}{\exp\left(tK\left(1+\delta\right)\right)}. \quad (5)$$

Now, we aim to find the optimal t such that $\partial f\left(t\right)/\partial t=0$, to make the bound (5) be tight. The optimal t is $\frac{M}{2K}\left(\frac{\delta}{1+\delta}\right)$. By using it in (5), we first get the upper bound on the probability of event $\|\mathbf{Ax}\|_2^2\geq K\left(1+\delta\right)$. That is

$$P\left(\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\geq\delta\|\mathbf{x}\|_2^2\right)\leq\exp\left(-\frac{M\delta}{2}\right)\left(\delta+1\right)^{\frac{M}{2}}. \quad (6)$$

Similarly, we also consider the upper bound probability of the event $\|\mathbf{Ax}\|_2^2\leq K\left(1-\delta\right)$. The upper bound probability is bounded by $\mathbb{E}\left[\exp\left(t\left(\|\mathbf{Ax}\|_2^2-K\left(1-\delta\right)\right)\right)\right]$ for $t<0$. That is

$$P\left(\|\mathbf{Ax}\|_2^2\geq K\left(1-\delta\right)\right)\leq g\left(t\right):=\frac{\left(1-2tK/M\right)^{-M/2}}{\exp\left(tK\left(1-\delta\right)\right)}. \quad (7)$$

We again find the optimal t such that $\partial f\left(t\right)/\partial t=0$. The optimal t is $-\frac{M}{2K}\left(\frac{\delta}{1-\delta}\right)$. By using it in (7), we get

$$P\left(\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\leq-\delta\|\mathbf{x}\|_2^2\right)\leq\exp\left(\frac{M\delta}{2}\right)\times\left(1-\delta\right)^{\frac{M}{2}}. \quad (8)$$

In addition, it is easy to show that

$$\exp\left(-\frac{M\delta}{2}\right)\left(\delta+1\right)^{\frac{M}{2}}\geq\exp\left(\frac{M\delta}{2}\right)\times\left(1-\delta\right)^{\frac{M}{2}}.$$

Finally, the upper bound on the probability of the event $\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2$ is obtained as follows:

$$P\left(\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2\right)=P\left(\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\geq\delta\|\mathbf{x}\|_2^2\right)$$
$$+P\left(\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\leq-\delta\|\mathbf{x}\|_2^2\right)$$
$$\leq\exp\left(-\frac{M\delta}{2}\right)(\delta+1)^{\frac{M}{2}} \quad (9)$$
$$+\exp\left(\frac{M\delta}{2}\right)\times(1-\delta)^{\frac{M}{2}}$$
$$\leq 2\exp\left(-\frac{M\delta}{2}\right)(\delta+1)^{\frac{M}{2}}$$

It completes the proof of Lemma.

Q.E.D.

Lemma 3 provides the upper bound on the probability of the event that a certain K column vectors of the matrix $\mathbf{A}$ does not become independent.

Our goal is to make the matrix $\mathbf{A}$ satisfies $K-$restricted isometry property with the RIP constant δ. Hence, we need to make sure that all possible binary K sparse signals satisfy the RIP for a given matrix $\mathbf{A}$. Due to all the possible cases, the number of measurements can increase. Theorem 1 provides the bound on the number of measurements.

Theorem 1: Let $\mathbf{A}$ be a Gaussian matrix, each element of which follows i.i.d. Gaussian distribution with zero-mean and variance $1/M$. Let $\mathbf{x}$ be the binary K sparse signal and $\delta\in(0,1)$. Then, the matrix $\mathbf{A}$ satisfies $K-$ restricted isometry property with the RIP constant δ if

$$\frac{2K\log(eN/K)}{\delta-\log(\delta+1)}<M, \quad (10)$$

as $K\to\infty$ or $N\to\infty$.

Proof: From Lemma 3, we know the upper bound probability of the event $\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2$ for the one binary K sparse signal. Since the number of binary K sparse signals is $\binom{N}{K}$, the probability of event $\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2$ for any binary K sparse signals is bounded by

$$\leq\binom{N}{K}2\exp\left(-\frac{M\delta}{2}\right)(\delta+1)^{\frac{M}{2}}$$
$$\leq 2(eN/K)^K\exp\left(-\frac{M\delta}{2}\right)(\delta+1)^{\frac{M}{2}} \quad (11)$$
$$\leq 2\exp\left(-\frac{M}{2}(\delta-\log(\delta+1))+K\log(eN/K)\right)$$

From (11), we notice that $P\left(\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2\right)$ goes to zero exponentially fast as a function of M if $\frac{2K\log(eN/K)}{\delta-\log(\delta+1)}<M$. Hence, we now can say that the matrix $\mathbf{A}$ satisfies $K-$restricted isometry property with the δ if sufficiently large $M=\mathrm{O}(K\log(N/K))$.

Q.E.D.

We remind that the L0 minimization gives the unique K sparse solution if $\delta_{2K}\in(0,1)$. Using this fact, we can say that if $6.7K\log(eN/K)<M$, the L0 minimization uniquely gives the K sparse signal with high probability.

Before we connect Theorem 1 with the L1 uniqueness, we introduce a Corollary. That is the matrix $\mathbf{A}$ satisfies (3) with high probability.

Corollary 1: Let $\mathbf{A}$ be a Gaussian matrix, each element of which follows i.i.d. Gaussian distribution with zero-mean and $1/M$ variance. Let $\mathbf{x}$ be the binary K sparse signal and $\delta\in(0,1)$. Then, the matrix $\mathbf{A}$ satisfies $K-$ restricted isometry property with the RIP constant δ with $1-\varepsilon$ probability if

$$\frac{2K\log(eN/K)}{\delta-\log(\delta+1)}+\frac{1.3863-2\log\varepsilon}{\delta-\log(\delta+1)}\leq M. \quad (12)$$

Proof: It is trivial to obtain from (11). Thus, we omit it.

Finally, we connect both Corollary 1 and Theorem 1 with the uniqueness condition. As mentioned in the Background, the L1 minimization exactly recover the K sparse signal if $\delta_{2K}<0.4531$. Hence, by letting $\delta\leq\delta_{2K}<0.4531$ in (12), then the L1 minimization can exactly recover the binary K sparse signal with $1-\varepsilon$ probability. A similar story is described in [7]

V. Conclusion

In this paper, we have obtained a sufficient condition on the number of measurements when the non-zero elements of the K-sparse signal are taken from the binary set {0,1}. We have obtained in Lemmas 1 and 2 that the energy of the measurements $\|\mathbf{Ax}\|_2^2$ is the chi-square random variable with M degrees of freedom, with mean K and variance $2K^2/M$. Using this result, we obtained in Lemma 3 the upper bound on the probability of a large deviation event, i.e. violating the RIP condition $\left\{\left|\|\mathbf{Ax}\|_2^2-\|\mathbf{x}\|_2^2\right|\geq\delta\|\mathbf{x}\|_2^2\right\}$, for a fixed binary K sparse signal; the result is $2e^{-M\delta/2}(\delta+1)^{M/2}$. Then, using the union bound, we have obtained an upper bound on the probability that a Gaussian matrix $\mathbf{A}$ with N columns fail to satisfy the RIP condition for each and every K sparse signal, which is given by $2e^{-M\delta/2}(\delta+1)^{M/2}\binom{N}{K}$ since there are $\binom{N}{K}$ number of distinct binary K sparse signals of length N. This gave the sufficient condition given in Theorem 1, which is still $O(K\log(N/K))$.

Our main result suggests that the size of the alphabet of the signal does not have a large impact in determining the order in the number of measurements. However, we notice that the $\log(N)$ factor appears as the result of the union bound taken in (11). So that, the $\log(N)$ factor maybe is reduced or removed if we obtain the tightest bound such the Gallager's random coding bound[12].

On a separate note, we may suppose using Vandermonde measurement matrices, instead of the Gaussian matrices, and apply the procedure given here. As pointed out in the introduction, the use of Vandermonde frames is one of the important factors in reducing the number of measurements. The reason is that they guarantee that any set of M or less column vectors of an $M \times N$ Vandermonde frame is linearly independent. Putting into the words of this paper, the probability of a large deviation event $\left\{ \left| \|\mathbf{Ax}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \geq \delta \|\mathbf{x}\|_2^2 \right\}$ is exactly zero as long as $M \geq 2K$; thus, the union bound is still zero. Hence, the probability that the matrix satisfies (3) is exactly 1 as long as $M \geq 2K$.

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. L. Donoho, "Compressed Sensing," IEEE Transaction on Information Theory, vol. 52, no. 4, pp. 1289-1306, 2006.

[2] Emmaneul Candes, "Compressive Sampling," Proceedings of the International Congress of Mathematicians, pp. 1433-1452, 2006.

[3] Richard Baraniuk, "Compressive Sensing," IEEE Transaction on Signal Processing Magazine, vol. 24, no. 4, pp. 118-121, 2007.

[4] Emmaneul Candes and Terence Tao, "Decoding by Linear Programming," IEEE Transaction on Information Theory, vol. 51, no. 12, pp. 4203-4215, 2005.

[5] Emmaneul Candes, "The Restricted Isometry Property and Its Implications for Compressed Sensing," Comptes Rendus Mathematique, pp. 589-592, 2008.

[6] Simon Foucart and Ming-Jun Lai, "Sparsest Solutions of Underdetermined Linear Systems via Lq-minimization for 0< q<=1," Applied and Computational Harmonic Analysis, vol. 26, no. 3 pp. 395-407, 2009.

[7] Richard Baraniuk, Mark Davenport, Ronald DeVore and Michael Wakin, "A Simple Proof of the Restricted Isometry Property for Random Matrices", Constructive Approximation, vol 28, pp. 253-263, 2008.

[8] John G. Proakis and Masoud Salehi, "Digital Communications", 5th edition, McGRAW. HILL, pp. 58 ~ 62, 2008.

[9] A. M. Bruckstein, Michael Elad and Michael Zibulevsky, "On the Uniqueness of Nonnegative Sparse Solutions to Underdetermined Systems of Equations", IEEE Transaction on Information Theory, vol. 54, no. 11, pp. 4813 – 4820, Nov, 2008.

[10] Mehmet Akcakaya and Vahid Tarokh, "A Frame Construction and a Universal Distrotion Bound for Sparse Represenations", IEEE Transaction on Signal Proceesing, vol. 56, no. 6, pp. 2443 – 2450, Jun, 2008.

[11] Ali Hormati and Martin Vetteli, "Annihilating filter-based decoding in the compressed sesning framework", Proceedings of SPIE, 2007.

[12] Robert G. Gallager, "A Simple Derivation of the Coding Theorem and Some Application", IEEE Transaction Information Theory, Jan. 1965.