Learning Dynamic Compressive Sensing Models

Dongeun Lee and Jaesik Choi

School of Electrical and Computer Engineering
Ulsan National Institute of Science and Technology (UNIST), Ulsan 689-798, Korea
eundong@unist.ac.kr, jaesik@unist.ac.kr

Abstract

Random sampling in compressive sensing (CS) enables to compress large amounts of input signals in efficient manner, which becomes useful for many applications. CS reconstructs the compressed signals exactly with overwhelming probability when incoming data can be sparsely represented with a fixed number of principal components, which is one of drawbacks of CS frameworks because the signal sparsity in many dynamic systems changes over time. We present the first CS framework that handles signals without the fixed sparsity assumption by incorporating the distribution of signal sparsity. We prove the beta distribution modeling of signal recovery success is more accurate than the success probability analysis in the CS framework, which is also confirmed by experiments. For signals that cannot be exactly represented with the sparse representation, we show the numbers of principal components included in their signal recoveries can be represented with a probability distribution. We show this distribution is skewed to the right and naturally represented by the gamma distribution.

Keywords: Compressive sensing, random sampling, dynamic signal sparsity, sparse signal recovery.

1 Introduction

Continuous flows of big data are generated from many sources nowadays. Among these, resource limited devices occupy a significant portion. For these devices, sensing and transmitting massive data are important challenges, as they are concerned with saving resources.

Compressive sensing (CS) \cite{1,2,3,4,5,6} is an ideal choice for the resource limited devices because it enables to sense and compress massive data without the complexity burden imposed by conventional schemes. Recent advances in CS reduce the complexity burden even further by random sampling, so that CS schemes are successfully applied to broader application areas \cite{7,8,9}.

More specifically, CS reconstructs the compressed signals exactly with overwhelming probability when incoming data can be sparsely represented (i.e., small numbers of principal components). In other words, the probability of recovery failure can be bounded when compressed signals has at least a predefined length. Here, the recovery failure means that CS fails to reconstruct the exact input signal.

However, existing CS frameworks are built based on a strong assumption which says that incoming data can be sparsely represented by a fixed number of principal components. This assumption does not hold in practice with many dynamic systems where the numbers of principal components change over time. The assumption also prevents deriving a tight probability distribution of principal components. Existing CS frameworks consider that the reconstruction would fail, when input signals have more principal components (denser) than the predefined threshold.
In this paper, we present a flexible CS framework that handles signals without the fixed sparsity assumption by incorporating the distribution of signal sparsity. Our framework learns the distribution of (changing) numbers of principal components of input signal in a dynamic system. By incorporating the beta distribution, our model presents the signal recovery success more accurately than the success probability analysis in the CS framework. Thus, our new model will pave the way to apply advances in CS frameworks, for signals generated from general dynamic systems. In addition, we propose the gamma distribution modeling for signals that cannot be exactly represented with the sparse representation.

The rest of this paper is organized as follows. Section 2 reviews the CS frameworks and random sampling. Section 3 provides our main contribution, a new CS framework which incorporates the distribution of signal sparsity, where we prove our beta distribution modeling is more accurate than existing CS framework. In Section 4 we show the number of principal components included in signal recovery can be represented with a right-tailed distribution. Section 5 exhibits experimental results, followed by concluding remarks in Section 6.

2 Compressive Sensing and Random Sampling

Compressive sensing or compressed sampling (CS) is an efficient signal processing technique that incorporates signal acquisition and compression simultaneously [7, 10]. If a signal can be represented by only a few components (i.e., principal components) with or without the help of a sparsifying basis, CS allows it to be efficiently acquired with a number of samples that is far fewer than the signal dimension and of the same order as the number of principal components.

2.1 Compressing While Sensing

In CS, a signal is projected onto random vectors whose cardinality is far below the dimension of the signal. Consider a signal \( x \in \mathbb{R}^N \) is compactly represented with a sparsifying basis \( \Psi \) having just a few principal components as follows:

\[
x = \Psi s,
\]

(1)

where \( s \in \mathbb{R}^N \) is the vector of transformed coefficients with a few significant coefficients.

In (1), \( \Psi \) could be a basis that makes \( x \) sparse in transform domain such as the DCT, wavelet transform domains, or even the canonical basis, i.e., the identity matrix \( I \), if \( x \) is sparse itself without the help of transform.

**Definition** The signal \( x \) is called \( K \)-sparse if it is a linear combination of only \( K \ll N \) basis vectors such that \( \sum_{i=1}^{K} s_{n_i} \psi_{n_i} \), where \( \{ n_1, \ldots, n_K \} \subset \{ 1, \ldots, N \} \); \( s_{n_i} \) is a coefficient in \( s \); \( \psi_{n_i} \) is a column of \( \Psi \).

In practice, some signal is not exactly \( K \)-sparse; rather it can be closely approximated with \( K \) basis vectors ignoring many small coefficients close to zeros. This type of signal is called *compressible* [7, 8].

CS projects \( x \) onto a random sensing basis \( \Phi \in \mathbb{R}^{M \times N} \) as follows (\( M < N \)):

\[
y = \Phi x = \Phi \Psi s,
\]

(2)
where $\Phi$ should have restricted isometry property (RIP)\footnote{The random sensing basis $\Phi$ have RIP if $(1 - \delta)\|s\|_2^2 \leq \|\Phi \Psi s\|_2^2 \leq (1 + \delta)\|s\|_2^2$ for small $\delta \geq 0$, and this condition applies to all $K$-sparse $s$.}. A conventional approach for $\Phi$ to satisfy RIP is sampling its independent identically distributed (i.i.d.) elements from the Gaussian or other sub-Gaussian distributions whose moment-generating function is bounded by that of the Gaussian (e.g., Rademacher/Bernoulli distribution).

The system shown in (2) is underdetermined as the number of equations $M$ is smaller than the number of variables $N$: there are infinitely many $x$’s that satisfy $y = \Phi x$. Nevertheless this system can be solved with overwhelming probability exploiting the fact that $s$ is $K$-sparse. Here $M = O(K \log(N/K))$ in the case of Gaussian and sub-Gaussian sensing matrices\footnote{The two bases $\Phi$ and $\Psi$ are incoherent when the rows of $\Phi$ cannot sparsely represent the columns of $\Psi$ and vice versa.}.

### 2.2 Random Sampling

Random sampling is a variant of CS that can further reduce the computational complexity to a constant time \cite{8, 9}. The random sampling scheme is based on the fact that it is possible to construct $\Phi$ in (2) from a random selection of rows from the identity matrix $I$, which is equivalent to the random sampling of coefficients in $x$.

Note that the sparsifying basis $\Psi$ should be incoherent\footnote{In a typical setup, the only information an encoder always has to send is $y$. The sensing matrix $\Phi$ can be explicitly sent \cite{3} or reconstructed using meta information such as the seed of pseudorandom number generator \cite{2, 11}, depending on application.} with $I$ such as the DCT and wavelet transform bases for the successful recovery of the original signal \cite{8, 10}. Unless they are incoherent, the measurement vector $y \in \mathbb{R}^M$ in (2) would contain zero entries.

The random sampling of signal in the CS setup is illustrated in Fig. 1. Here, the number of required measurements $M$ is somewhat larger than the case of Gaussian and sub-Gaussian matrices, that is, $M = O(K \log N)$.

### 2.3 Recovery of Signal

A signal recovery algorithm takes measurements $y$, a random sensing matrix $\Phi$, and the sparsifying basis $\Psi$. The sensing matrix $\Phi$ and sparsifying basis $\Psi$ are assumed to be known to a decoder.\footnote{In a typical setup, the only information an encoder always has to send is $y$. The sensing matrix $\Phi$ can be explicitly sent \cite{3} or reconstructed using meta information such as the seed of pseudorandom number generator \cite{2, 11}, depending on application.}

The signal recovery algorithm then recovers $s$ knowing that $s$ is sparse. Once we recover $s$, the original signal $x$ can be recovered through (1). The recovery algorithm reconstructs $s$ by the following linear program:

$$\argmin \|s\|_1 \quad \text{subject to} \quad \Phi \Psi \tilde{s} = y.$$  \hfill (3)
The optimization problem in (3) can be solved under categories of optimization methods, greedy methods, and thresholding-based methods [8, 12, 13, 14]. Choosing a specific algorithm depends on \( \Phi, M, N, \) and \( K \): recovery success rates and speed can only be determined by numerical tests.

Specifically in case of the random sampling, the solution \( s^* \) to (3) obeys

\[
\|s^* - s\|_2 \leq C_1 \cdot \|s - s_K\|_1 \tag{4}
\]

for some constant \( C_1 > 0 \), where \( s_K \) is the vector \( s \) with all but the largest \( K \) components set to 0. When an original signal is exactly \( K \)-sparse, then \( s = s_K \) with \( M = O(K \log N) \) measurements, which implies the recovery is exact, i.e., \( s^* = s \).

### 3 Learning Recovery Success

The signal reconstruction in compressive sensing (CS) is closely related to a probabilistic concept. For instance, when we say an exact recovery of \( K \)-sparse signal is achievable with overwhelming probability, it implies there is also the chance of recovery not being exact.

Most existing CS literature assumes an enough number of measurements \( M \) so that an exact recovery is almost always achievable [8, 10], which is based on the assumption that the sparsity \( K \) is already known or does not exceed a certain bound. However, the signal sparsity in dynamic systems can change over time and an excessive number of measurements can waste resources such as network bandwidth and storage space. Therefore, we propose a new learning problem for the random sampling of CS and provide a new theoretical analysis on signal recovery.

#### 3.1 Compressive Sensing Framework

In the random sampling of CS, the number of required measurements \( M = O(K \log N) \) in random sampling can be detailed as follows [8]:

\[
M \geq C \cdot K \ln(N) \ln(\epsilon^{-1}) \tag{5}
\]

for some constant \( C > 0 \), where \( \epsilon \in (0, 1) \) denotes the probability of an inexact recovery of \( K \)-sparse signal. In particular, the signal recovery succeeds with a probability at least \( 1 - \epsilon \) if (5) holds.

We can then express (5) with regard to the probability of failure \( \epsilon \), which is given by

\[
P(s^* \neq s \mid M, N, K) := \epsilon \leq \exp \left( -\frac{M}{C \cdot \ln(N) K} \right). \tag{6}
\]

Thus, the probability of failure (inexact recovery) \( P(s^* \neq s \mid M, N, K) \) is conditional upon \( M, N, \) and \( K \). Since we are interested in the dynamic signal sparsity \( K \), we model \( K \) as a random variable; \( M \) and \( N \) as fixed quantities.

If we denote an arbitrary probability density function (pdf) of \( K \) as \( f_K(k) \), we can marginalize over \( f_K(k) \) and find the upper bound of failure probability as follows:

\[
P(s^* \neq s \mid M, N) = \int_k P(s^* \neq s \mid M, N, K) \cdot f_K(k) \, dk \leq \int_k \exp \left( -\frac{M}{C \cdot \ln(N) K} \right) f_K(k) \, dk. \tag{7}
\]

\( ^4 \)Note that greedy methods are not always fast.
The upper bound in (7) can have an analytic solution depending on the form of \( f_K(k) \). In particular, this is the case when \( K \) follows certain distributions such as the inverse Gaussian distribution and the gamma distribution. Therefore, we can state that a signal recovery succeeds with a probability at least 
\[
1 - \int_k \exp\left(-\frac{M}{C \cdot \ln(N) K}\right) f_K(k) \, dk,
\]
given the distribution of signal sparsity \( f_K(k) \).

### 3.2 Beta Distribution Modeling

Unfortunately, the probability of signal recovery failure \( \epsilon \) given in (6) does not hold in practice. There is a discrepancy between the failure probability in the CS framework and a failure probability in actual random sampling. Thus we have to model success or failure probability of signal recovery from new perspective.

As an example, we examine the success probability more closely by generating many different signed spike (\( \pm 1 \)) vectors for each signal sparsity and then performing experiments for each signed spike vector. Fig. 2 shows histograms of success probability for various signal sparsities.

We model the success probability shown in Fig. 2 by the beta distribution with its parameters \( \alpha \) and \( \beta \) depending on signal sparsity, i.e., \( P(s^* = s \mid M, N, K) \sim \text{Beta}(\alpha_K, \beta_K) \). Moreover in Fig. 2, we can see that the mean of success probability distribution constantly decreases as \( K \) increases; the variance of success probability distribution keeps increasing till \( K = 25 \) and then keeps decreasing.

---

\( ^5 \) Since \( K \geq 0 \), probability distributions supported on semi-infinite intervals, i.e., \((0, \infty)\), are rational choices.

\( ^6 \) Detailed settings are explained in Section 5.
Fitting distributions in Fig. 2 with beta distributions using the maximum-likelihood estimation, we can obtain their parameters $\alpha$’s and $\beta$’s. In order to relate these parameters to the signal sparsity, we then use the least-squares curve fitting for the parameters.

In particular, we consider the quantity $M/(\ln(N)K)$ in (6) because the signal sparsity $K$ should be considered together with the number of measurements $M$ and the signal length $N$. Thus we model both $\alpha$ and $\beta$ as functions of this quantity, which is given by

\[
\alpha(M, N, K) = a_\alpha \ln \left( \frac{M}{\ln(N)K} \right) + b_\alpha, \tag{8}
\]

\[
\beta(M, N, K) = a_\beta \frac{M}{\ln(N)K} + b_\beta, \tag{9}
\]

where $a_\alpha > 0$, $b_\alpha > 0$; $a_\beta < 0$, $b_\beta > 0$ are model parameters for $\alpha(\cdot)$ and $\beta(\cdot)$ respectively.

The resulting model functions are plotted in Fig. 3 with $\alpha$’s and $\beta$’s obtained by the maximum-likelihood estimation for comparison. Fig. 4 also presents (sample) means and variances of the success probability distributions shown in Fig. 2, along with means and variances calculated with the model functions in (8) and (9). We can see that the beta distribution modeling with parameters (8) and (9) effectively follows the actual success probability distributions.

Incorporating extreme cases where signal recovery always succeeds or always fails (e.g., $K < 20$ or $K > 30$ in Fig. 2), we introduce $K_{\text{min}}$ and $K_{\text{max}}$ to denote the minimum and the maximum signal sparsities respectively that yield stochastic probability, not a deterministic result which can be represented as a success or a failure. Then we can model the new pdf of signal recovery success using the mixture of the Dirac delta function and the beta distribution.
Figure 4: Mean and variance of success probability distributions in Fig. 2 estimated with the beta distribution modeling $\alpha(\cdot)$ and $\beta(\cdot)$.

**Definition** Let $P(s^* = s \mid M, N) := \Pi$. The pdf of $\Pi$ given $K$ is given by

$$f_{\Pi \mid K}(\pi \mid k) := \begin{cases} 
\delta(\pi - 1) & k < K_{\min} \\
\text{Beta}(\alpha_K, \beta_K) & K_{\min} \leq k \leq K_{\max} \\
\delta(\pi) & K_{\max} < k
\end{cases}.$$  

Combining this definition with an arbitrary pdf $f_K(k)$ of the dynamic signal sparsity $K$, we can find the success probability distribution marginalized over $f_K(k)$ as follows:

$$f_{\Pi}(\pi) = \int f_{\Pi \mid K}(\pi \mid k)f_K(k) \, dk$$

$$= \int_{0}^{K_{\min}} \delta(\pi - 1)f_K(k) \, dk$$

$$+ \int_{K_{\min}}^{K_{\max}} \text{Beta}(\alpha_K, \beta_K) \cdot f_K(k) \, dk$$

$$+ \int_{K_{\max}}^{\infty} \delta(\pi)f_K(k) \, dk$$

$$= \delta(\pi - 1)F_K(K_{\min}) + \delta(\pi)(1 - F_K(K_{\max}))$$

$$+ \int_{K_{\min}}^{K_{\max}} \text{Beta}(\alpha_K, \beta_K) \cdot f_K(k) \, dk,$$

where $F_K(\cdot)$ is the cumulative distribution function (CDF) of $K$.

$^7$Beta($\alpha_K, \beta_K$) here is used to denote the pdf of the beta distribution.
The two Dirac delta function terms in (11) can be interpreted as probability masses. Since
\[ \int_{K_{\min}}^{K_{\max}} \text{Beta}(\alpha_K, \beta_K) \cdot f_K(k) \, dk \]
does not have the analytic solution, we compute the values numerically.

### 3.3 Modeling Accuracy

Here, we present the main theoretical contribution: the recovery success probability modeled by our beta distribution is tighter than the lower bound of that in the existing CS framework. We show the failure probability in the CS framework is incapable of reflecting the actual failure probability of signal recovery. Not only the inequality
\[ P(s^* \neq s \mid M, N, K) \leq \exp(-M/(C \cdot \ln(N)K)) \]
cannot provide tight probability of failure, but the inequality itself is inaccurate.

This inaccuracy results from the slowly decaying lower bound of success probability, that is, \(1 - \exp(-M/(C \cdot \ln(N)K))\). In fact, we can show this lower bound decays slower than a power-law decay by the following lemma.

**Lemma 1** There exists \(K_0 > 0\) such that for all \(K > K_0\), the lower bound of recovery success probability is greater than the value of a power-law-decay function.

**Proof** We need to show the following inequality
\[ 1 - \exp \left( -\frac{M}{C \cdot \ln(N)K} \right) > K^{-\alpha} \]
holds if \(K > K_0\) for some \(K_0 > 0\), where \(\alpha > 0\). Adding, subtracting, and taking the power \(K\) in both sides yields
\[ (1 - K^{-\alpha})^K > \exp \left( -\frac{M}{C \cdot \ln(N)} \right). \]

We now use the *binomial approximation* on the left-hand side: \((1 - K^{-\alpha})^K \geq 1 - K \cdot K^{-\alpha}\). Thus we instead prove the following inequality
\[ 1 - K^{1-\alpha} > \exp \left( -\frac{M}{C \cdot \ln(N)} \right), \]
holds if \(K > K_0\) for some \(K_0 > 0\).

If we assume \(\alpha > 1\), then adding, subtracting, and taking the power \(1/(1 - \alpha)\) in both sides of (14) yields
\[ \left( 1 - \exp \left( -\frac{M}{C \cdot \ln(N)} \right) \right)^{1/(1-\alpha)} < K. \]
Setting \(K_0 = (1 - \exp(-M/(C \cdot \ln(N))))^{1/(1-\alpha)}\), we can argue that for all \(K > K_0\), the lower bound of recovery success probability is greater than the value of a power-law-decay function.

**Corollary 2** In the CS framework, there is always a chance of succeeding in signal recovery however large \(K\) is.

**Proof** The power-law-decay function \(K^{-\alpha}\) in (12) slowly converges to zero as \(K \to \infty\): its value is noticeably greater than zero even with large \(K\). As the lower bound of recovery success probability is greater than the value of the power-law-decay function for all \(K > K_0\), we can say there is always a chance of recovery success however large \(K\) is.
We now show that our beta distribution modeling provides more accurate success probability by the following theorem.

**Theorem 3** The recovery success probability modeled by the beta distribution is tighter than the lower bound of that given by the CS framework.

**Proof** The claim of the CS framework in Corollary 2 is in fact implausible because it says we can even set $K > M$ and there is still a chance of success. We cannot expect the signal recovery with the number of measurements $M$ less than $K$.

On the contrary, our beta distribution modeling can yield $P(s^* = s \mid M, N, K) = 0$ with a bounded $K_{\text{max}}$. In particular, we show that the mean of $\text{Beta}(\alpha_K, \beta_K)$, $\alpha_K/(\alpha_K + \beta_K)$, converges to zero with a bounded $K_{\text{max}}$. If we state $\alpha_K/(\alpha_K + \beta_K) = 0$ using (8) and (9), we obtain

$$a_\alpha \ln \left( \frac{M}{\ln(N)K} \right) + b_\alpha = 0. \tag{16}$$

Representing (16) with respect to $K$, we have

$$K_{\text{max}} = \frac{M}{\ln(N)} \exp \left( \frac{b_\alpha}{a_\alpha} \right) < \infty. \tag{17}$$

Similarly, we show this mean converges to one ($P(s^* = s \mid M, N, K) = 1$) with $K_{\text{min}}$ which is not so close to zero, whereas the lower bound of the recovery success probability given by the CS framework converges to one only if $K$ is very close to zero.

If we state $\alpha_K/(\alpha_K + \beta_K) = 1$ using (8) and (9), we have

$$a_\beta \frac{M}{\ln(N)K} + b_\beta = 0. \tag{18}$$

Representing (18) with respect to $K$, we have

$$K_{\text{min}} = -\frac{a_\beta M}{b_\beta \ln(N)}. \tag{19}$$

Plugging $K_{\text{min}}$ into $1 - \exp(-M/(C \cdot \ln(N)K))$ shows that

$$1 - \exp \left( \frac{b_\beta}{C \cdot a_\beta} \right) > 0, \tag{20}$$

as $C > 0$, $a_\beta < 0$, and $b_\beta > 0$: that is, $1 - \exp(-M/(C \cdot \ln(N)K)) \to 1$ if and only if $K \to 0$.

Since $0 < K_{\text{min}} < K_{\text{max}} < \infty$, we can argue that the beta distribution modeling can provide tighter recovery success probability. □

### 4 Learning Recovery Quality

When a signal in question is not exactly $K$-sparse, but compressible as discussed in Section 2.1, the recovery of signal in Section 2.3 should be treated from different perspective. In particular, the inequality (4) is rendered differently.

If an original signal is compressible, then the quality of a recovered signal is proportional to that of the $K$ most significant pieces of information (principal components). We get progressively better
results as we compute more measurements $M$, since $M = O(K \log N)$ \[10\]. Therefore, $\Psi s^* \in \mathbb{R}^N$ also makes progress on its quality as $M$ increases. (The error bound follows \[4\] as well if $\Psi$ is an orthogonal matrix, which is usually the case.)

In this framework, the success or failure of signal recovery no longer exists. Rather, we can view the number of principal components included in the signal recovery as a \textit{varying quantity} \[15\]. Specifically, if a signal recovery is about to fail with a given $K$, then $K$ can be lowered to make the recovery succeed eventually. Here the number of included principal components $K$ varies for different recoveries and signals, as analogous to the success probability in Section \[3.2\] that can be calculated with different recoveries and varies for different signals.

In this regard, \[4\] can be utilized to infer varying $K$’s over different recoveries and signals. Here our assumption is that the upper bound in \[4\] is \textit{tight} such that we solve the following optimization problem:

\[
\max K \quad \text{subject to} \quad \|s^* - s\|_2 \leq C_1 \cdot \|s - s_K\|_1. \tag{21}
\]

In \[21\], $C_1$ has to be determined, where the maximum signal sparsity $K_{\text{max}}$ introduced in Section \[3.2\] plays a key role to set the upper limit on how large $K$ can be, since $K > K_{\text{max}}$ is not reasonable.

In particular, we can generate a compressible signal $s_i \in S$ such that $\|s_i\|_1 = C_{\ell_1}$ for all $i$, where $S$ is the set containing many different signals; $C_{\ell_1} > 0$ is some constant. For each $s_i$, we have a set $S_{ij}^*$ that contains many different recoveries $s_{ij}^*$. Then $C_1$ can be found as follows:

\[
C_1 = \min \|s_{ij}^* - s_i\|_2 \|s_i - s_i^{K_{\text{max}}}\|_1, \tag{22}
\]

where $s_i^{K_{\text{max}}}$ denotes the compressible signal $s_i$ with all but the largest $K_{\text{max}}$ components set to 0.

Varying $K$’s obtained through \[21\] can be represented by a pdf. We are interested in the shape of this pdf, which is revealed by the following proposition.

**Proposition 1** The pdf of $K$, the number of principal components included in the signal recovery of a compressible signal, is skewed to the right, i.e., right tailed.

**Proof** Since $\|s_i\|_1 = C_{\ell_1}$ for all $i$, we can conceive the same sequence $\{s_n\}$ of elements (absolute values) in $s_i$ for all $i$. Then we have

\[
\|s_i - s_i^{K_{\text{max}}}\|_1 = \sum_{n=1}^{N-K} s_n. \tag{23}
\]

Without loss of generality, we consider the partial sum $\sum_{n=1}^{N-K} s_n$ in \[23\] to be an \textit{arithmetic series} that can be represented by a quadratic function in terms of $K$. We also assume the inequality constraint in \[21\] is the equality constraint such that $\|s^* - s\|_2 = C_1 \cdot \|s - s_K\|_1$.

If we take the (partial) inverse function of the quadratic function, we have $K \sim K_{\text{max}} - \sqrt{\|s^* - s\|_2 - (\min \|s_{ij}^* - s_i\|_2)}$. Assuming the distribution of $\|s^* - s\|_2$ is \textit{symmetric} (zero skewness), this asymptotic relation says $\|s^* - s\|_2$ will be \textit{compressed} as it becomes large, which in turn makes the pdf of $K$ right tailed.

A similar claim can be made if we consider the partial sum $\sum_{n=1}^{N-K} s_n$ to be a \textit{geometric series}, where $K \sim N - \log(\|s^* - s\|_2)$. In this case, the pdf of $K$ is skewed to the right as well. \qed
Figure 5: Comparison between actual failure probability and failure probabilities given in (6) with varying $C$’s. Actual failure probability of each signal sparsity $K$ was obtained with 300 experiments for $N = 512$ and $M = 100$.

5 Experimental Results

5.1 Recovery Success

In Section 3, we discussed the discrepancy between the failure probability in the CS framework and the failure probability in actual random sampling. In order to show this discrepancy, we artificially generated signed spikes $\pm 1$ at random locations in proportion to desired sparsities and densified these spikes using $\Psi^8$ as in (1) to perform the random sampling.

For each signal sparsity $K$, actual failure probability can be calculated for different recovery experiments. To this end, we adopted an optimization method to solve the optimization problem in (3) that is also called basis pursuit [16]. Specifically, the primal-dual algorithm based on the interior point method was employed to solve (3) [12].

Fig. 5 shows that actual failure probability of signal recovery with varying signal sparsity does not follow the failure probability given in the CS framework. The failure probability in (6) cannot model actual failure probability of signal recovery, regardless of various choices of a constant $C$. This result confirms Lemma 4 and Corollary 2.

Moreover in Section 3.2, we modeled the new pdf of signal recovery success $f_\Pi(\pi)$ in (11). We compare this new pdf with the upper bound of failure in (7), given a dynamic signal sparsity $K$. Specifically, we employed the inverse Gaussian distribution such that $f_K(k) = IG(30, 200)$.

8 We used DCT as the sparsifying basis $\Psi$ throughout experiments.
Figure 6: Comparison between our new success probability distribution in (11) and the lower bounds of success probability obtained by (7) with varying $C$’s. The inverse Gaussian distribution was used for $f_K(k)$. Two probability masses are shown in vertical arrows, where solid boxes atop the arrows denote their probabilities. Three vertical dashed/dotted lines represent the lower bounds by (7): $C = 0.5$ at 0.6781; $C = 1$ at 0.4450; $C = 2$ at 0.2596. Here, $K_{\text{min}} = 20$ and $K_{\text{max}} = 30$.

Fig. 6 exhibits the efficacy of our beta distribution modeling, where the lower bounds of success probability given in the CS framework fail to capture actual success probability in random sampling. This result confirms Theorem 3.

We also employed an environmental data set obtained from a wireless sensor network deployment for comparison [17], where relative humidity data (%) was used here. Random numbers representing the dynamic signal sparsity $K$ were drawn from the inverse Gaussian distribution and we used this $K$ to randomly choose principal components in humidity data; other components were set to zeros. Fig. 7 displays the success probability of signal recovery follows the shape of Fig. 6. Specifically, two probability masses of success and failure are evident in Fig. 7 although their values are different. This is attributed to different $K_{\text{min}}$ and $K_{\text{max}}$ for the humidity data.

5.2 Recovery Quality

When a signal is compressible and not exactly $K$-sparse, the dynamic signal sparsity $K$ cannot be easily defined. In this scenario, every signal is basically dense, although it can be approximated with $K$ basis vectors.

In Section 4 we regarded the number of principal components included in the signal recovery as varying quantity. And we are interested in the general shape of this quantity in distribution.
To this end, random signed spikes were artificially generated in different magnitudes at random locations and densified to perform random sampling. In particular, we considered an arithmetic sequence of length 50, \((2, 4, 6, \ldots, 98, 100)\), whose elements were placed at random locations in each vector. These signals are dense enough to be used for experiments because a signal recovery always fails when \(K > 30\) in our case, as shown in Fig. 5.

Fig. 8 displays the histogram of \(K\), the number of principal components included in each signal recovery, which was obtained using the method explained in Section 4. We can identify that Proposition 1 actually holds here, as this distribution is skewed to the right. Furthermore, we empirically found that the distribution followed the gamma distribution, which is also natural since the gamma distribution has the positive skewness, i.e., right tailed.

We also provide a result with a real-world signal to verify Proposition 1 and the gamma distribution fitting. Fig. 9 displays the histogram of \(K\) and its gamma distribution fitting, where we can see that \(K\) follows the gamma distribution as well.

6 Conclusion

We have presented the first CS framework that handles signals without the fixed sparsity assumption. The success probability of signal recovery in random sampling was investigated when the signal sparsity can vary. The success probability analysis in the existing CS framework was shown to be incapable of reflecting actual success probability by both theoretical analysis and experiments.
Figure 8: Distribution of $K$ fitted with a gamma distribution $\text{Gamma}(242.81, 0.09)$, using the maximum likelihood estimation. Histogram was obtained with 300 different signals and 300 different experiments for each signal, for $N = 512$ and $M = 100$.

On the contrary, our beta distribution modeling could closely reflect the actual success probability. We also considered signals that cannot be exactly represented with the sparse representation, where we viewed the number of principal components included in the signal recovery as varying quantity. This quantity was shown to follow a right-tailed distribution such as the gamma distribution by both theoretical analysis and experiments. We plan to investigate how to define and model the dynamic signal sparsity for these signals.

References

[1] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, “Compressive wireless sensing,” in Proc. Int’l Conf. Inf. Process. Sens. Netw. (IPSN ’06), 2006, pp. 134–142.

[2] C. Luo, F. Wu, J. Sun, and C. W. Chen, “Compressive data gathering for large-scale wireless sensor networks,” in Proc. MobiCom ’09, 2009, pp. 145–156.

[3] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, “On the interplay between routing and signal representation for compressive sensing in wireless sensor networks,” in Proc. Inf. Theory Appl. Workshop (ITA ’09), 2009, pp. 206–215.

[4] S. Ji and L. Carin, “Bayesian compressive sensing and projection optimization,” in Proc. ICML ’07, 2007, pp. 377–384.
Figure 9: Distribution of $K$ fitted with a gamma distribution $\Gamma(11.78, 2.04)$ for relative humidity (%) data. Histogram was obtained with 1,000 experiments over the same signal for $N = 512$ and $M = 100$.

[5] M. Lopes, “Estimating unknown sparsity in compressed sensing,” in Proc. ICML ’13, 2013, pp. 217–225.

[6] D. Malioutov and K. Varshney, “Exact rule learning via boolean compressed sensing,” in Proc. ICML ’13, 2013, pp. 765–773.

[7] R. G. Baraniuk, “Compressive sensing [lecture notes],” IEEE Signal Process. Mag., vol. 24, no. 4, pp. 118–121, Jul. 2007.

[8] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing. Springer, 2013.

[9] D. Lee and J. Choi, “Low complexity sensing for big spatio-temporal data,” in Proc. Int’l Conf. Big Data (BigData ’14), 2014, pp. 323–328.

[10] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” IEEE Signal Process. Mag., vol. 25, no. 2, pp. 21–30, Mar. 2008.

[11] M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, “Universal distributed sensing via random projections,” in Proc. Int’l Conf. Inf. Process. Sens. Netw. (IPSN ’06), 2006, pp. 177–185.

[12] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.
[13] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in Proc. Asilomar Conf. Signals, Syst. Comput. (ACSSC ’93). IEEE, 1993, pp. 40–44.

[14] T. Blumensath and M. E. Davies, “Iterative thresholding for sparse approximations,” J. Fourier Anal. Appl., vol. 14, no. 5-6, pp. 629–654, 2008.

[15] D. Lee and J. Choi, “Learning compressive sensing models for big spatio-temporal data,” Proc. SDM ’15, Apr./May 2015, to be published.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM J. Sci. Comput., vol. 20, no. 1, pp. 33–61, 1998.

[17] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, “Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework,” IEEE Trans. Wireless Commun., vol. 11, no. 10, pp. 3447–3461, Oct. 2012.