Improving Imprecise Compressive Sensing Models

Dongeun Lee and **Rafael Lima** and **Jaesik Choi**
School of Electrical and Computer Engineering
Ulsan National Institute of Science and Technology
Ulsan, 44919, Korea
{eundong, rafael_glima, jaesik}@unist.ac.kr

Abstract

Random sampling in compressive sensing (CS) enables the compression of large amounts of input signals in an efficient manner, which is useful for many applications. CS reconstructs the compressed signals exactly with overwhelming probability when incoming data can be sparsely represented with a few components. However, the theory of CS framework including random sampling has been focused on exact recovery of signal; impreciseness in signal recovery has been neglected. This can be problematic when there is uncertainty in the number of sparse components such as signal sparsity in dynamic systems that can change over time. We present a new theoretical framework that handles uncertainty in signal recovery from the perspective of recovery success and quality. We show that the signal recovery success in our model is more accurate than the success probability analysis in the CS framework. Our model is then extended to the case where the success or failure of signal recovery can be relaxed. We represent the number of components included in signal recovery with a right-tailed distribution and focus on recovery quality. Experimental results confirm the accuracy of our model in dynamic systems.

1 INTRODUCTION

Continuous flows of big data are generated by 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) (Seeger & Nickisch, 2008; Hsu et al., 2009; Lopes, 2013; Malioutov & Varshney, 2013; Zhu & Gu, 2015) is a well suited choice for resource limited devices because it enables the sensing and compression of massive data without the complexity burden imposed by conventional schemes. Recent advances in CS reduce the complexity burden even further with *random sampling*, by which CS schemes have been successfully applied to broader application areas such as sampling of spatio-temporal data (Foucart & Rauhut, 2013; Lee & Choi, 2014).

CS reconstructs the exact signals from the compressed measurements with overwhelming probability when incoming data can be sparsely represented (i.e., small numbers of components). Therefore, most CS frameworks are built based on the assumption that incoming data with sparse representation can be *exactly* recovered when an enough number of measurements is given.

Unfortunately, this assumption does not hold in practice when there is no guarantee of enough measurements for varying signal sparsity. This *uncertainty* occurs especially with many dynamic systems where the numbers of components change over time. The assumption also implies that the reconstruction would fail when input signals have more components (denser) than a predefined threshold. This prevents deriving a tight probabilistic model which takes account of the numbers of components and measurements in signal recovery. In this regard, recently introduced dynamic CS frameworks (Malioutov et al., 2010; Sejdinovic et al., 2010; Shahrasbi et al., 2011; Vaswani & Lu, 2010; Ziniel & Schniter, 2013) provide the way of reducing the number of necessary measurements exploiting temporal correlation between measurements. Nevertheless, a recovery success/quality analysis with uncertainty in signal sparsity has not been provided by existing CS frameworks yet.

This paper presents a new theoretical framework for the random sampling in CS that handles *impreciseness* in signal recovery when the number of measurements lacks for varying signal sparsity. Our framework incorporates the beta distribution to present the signal recovery success more accurately than the success probability analysis in the CS framework. Furthermore, we relax the concept of signal recovery success and present the number of components

included in the signal recovery as a varying quantity, for which we propose right-tailed distribution modeling. We believe our new framework will bridge the gap between success and failure of signal recovery in CS frameworks.

2 COMPRESSIVE SENSING AND RANDOM SAMPLING

Compressive sensing, or compressed sampling (CS), is an efficient signal processing framework which incorporates signal acquisition and compression simultaneously (Baraniuk, 2007; Candès & Wakin, 2008). CS enables a signal to be acquired with a number of samples that is far fewer than an original signal dimension and of the same order as the number of (significant) 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 $\mathbf{x} \in \mathbb{R}^N$ is compactly represented with a sparsifying basis $\boldsymbol{\Psi}$ having just a few components: $\mathbf{x} = \boldsymbol{\Psi}\mathbf{s}$, where $\mathbf{s} \in \mathbb{R}^N$ is the vector of transformed coefficients with a few significant coefficients. Here, $\boldsymbol{\Psi}$ could be a basis that makes $\mathbf{x}$ sparse in a transform domain such as the DCT, wavelet transform domains, or even the canonical basis, i.e., the identity matrix $\mathbf{I}$, if $\mathbf{x}$ is sparse itself without the help of a transform.

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

In practice, some signals may not be exactly K-sparse. Rather, they can be closely approximated with K basis vectors by ignoring many small coefficients close to zero. This type of signal is called *compressible* (Baraniuk, 2007; Foucart & Rauhut, 2013).

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

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\Psi}\mathbf{s}, \qquad (1)$$

where $\boldsymbol{\Phi}$ should have the *restricted isometry property* (RIP).[1] A conventional approach for $\boldsymbol{\Phi}$ to satisfy RIP is sampling its independent identically distributed (i.i.d.) elements from the Gaussian distribution or other sub-Gaussian distributions (e.g., Rademacher/Bernoulli distribution).

The system shown in (1) is underdetermined, as the number of equations M is smaller than the number of variables N, i.e., there are infinitely many $\mathbf{x}$'s that satisfy $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x}$. Nevertheless, this system can be solved with overwhelming probability exploiting the fact that $\mathbf{s}$ is K-sparse. Here $M = O(K \log(N/K))$ in the case of Gaussian and sub-Gaussian sensing matrices (Candès & Wakin, 2008).

2.2 RANDOM SAMPLING

Random sampling is a variant of CS which can further reduce the computational complexity to a constant time (Foucart & Rauhut, 2013; Lee & Choi, 2014). The random sampling scheme is based on the fact that it is possible to construct $\boldsymbol{\Phi}$ in (1) from a random selection of rows from the identity matrix $\mathbf{I}$, which is equivalent to the random sampling of coefficients in $\mathbf{x}$.

Note that the sparsifying basis $\boldsymbol{\Psi}$ should be *incoherent*[2] with $\mathbf{I}$, such as the DCT and wavelet transform bases, for the successful recovery of the original signal (Candès & Wakin, 2008; Foucart & Rauhut, 2013). Unless they are incoherent, the measurement vector $\mathbf{y} \in \mathbb{R}^M$ in (1) would contain zero entries. Here, the number of required measurements M is larger than in the cases of Gaussian and sub-Gaussian matrices, that is, $M = O(K \log N)$.

2.3 RECOVERY OF SIGNAL

A signal recovery algorithm takes measurements $\mathbf{y}$, a random sensing matrix $\boldsymbol{\Phi}$, and the sparsifying basis $\boldsymbol{\Psi}$. The sensing matrix $\boldsymbol{\Phi}$ and sparsifying basis $\boldsymbol{\Psi}$ are assumed to be known to a decoder. The signal recovery algorithm then recovers $\mathbf{s}$ knowing that $\mathbf{s}$ is sparse. Once we recover $\mathbf{s}$, the original signal $\mathbf{x}$ can be recovered through $\mathbf{x} = \boldsymbol{\Psi}\mathbf{s}$. The recovery algorithm reconstructs $\mathbf{s}$ by the following linear program:

$$\operatorname{argmin} \|\tilde{\mathbf{s}}\|_1 \qquad \text{subject to} \qquad \boldsymbol{\Phi}\boldsymbol{\Psi}\tilde{\mathbf{s}} = \mathbf{y}. \qquad (2)$$

The optimization problem in (2) is solved by a ℓ_1-minimization method (basis pursuit) (Boyd & Vandenberghe, 2004), greedy methods such as orthogonal matching pursuit (Pati et al., 1993), or thresholding-based methods such as iterative hard thresholding (Blumensath & Davies, 2008). Choosing a specific algorithm depends on $\boldsymbol{\Phi}$, M, N, and K: recovery success rates and speed can only be determined by numerical tests (Foucart & Rauhut, 2013).[3] In this paper, we reconstruct signals by the *basis pursuit*.

Specifically in the case of random sampling, the solution $\mathbf{s}^\star$ to (2) obeys

$$\|\mathbf{s}^\star - \mathbf{s}\|_2 \leq C_1 \cdot \|\mathbf{s} - \mathbf{s}_K\|_1 \qquad (3)$$

---

[1] The random sensing basis $\boldsymbol{\Phi}$ have RIP if $(1 - \delta)\|\mathbf{s}\|_2^2 \leq \|\boldsymbol{\Phi}\boldsymbol{\Psi}\mathbf{s}\|_2^2 \leq (1 + \delta)\|\mathbf{s}\|_2^2$ for small $\delta \geq 0$, and this condition applies to all K-sparse $\mathbf{s}$.

[2] The two bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are incoherent when the rows of $\boldsymbol{\Phi}$ cannot sparsely represent the columns of $\boldsymbol{\Psi}$ and vice versa.

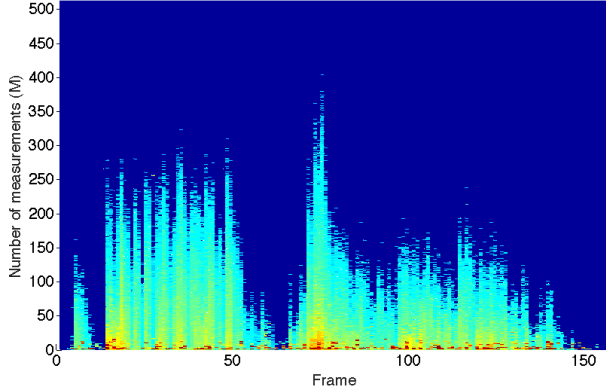[3] Note that greedy methods are not always fast.

Figure 1: Recovery error of audio data for different numbers of measurements M over time. Each frame has a signal length $N = 512$. Colors close to blue represent smaller errors, whereas colors close to red represent larger errors.

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

3 A NEW PERSPECTIVE ON RECOVERY SUCCESS

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

Most CS literature assumes a sufficient number of measurements M such that an exact recovery is almost always achievable (Candès & Wakin, 2008; Foucart & Rauhut, 2013), which is based on an assumption that sparsity K is already known or does not exceed a certain bound. However, the signal sparsity in dynamic systems may change over time and an excessive number of measurements may waste resources such as network bandwidth and storage space. For example, Figure 1 shows recovery error over time for audio data (a 7 second recording of a trumpet solo) (Ziniel et al., 2012), where varying signal sparsity incurs different recovery error with a fixed number of measurements over time. Here we cannot simply increase the number of measurements to eliminate error, as it is unreasonable in terms of compression. Therefore, we propose a new theoretical framework for the random sampling of CS and provide a new perspective on signal recovery.

3.1 COMPRESSIVE SENSING FRAMEWORK

In the random sampling of CS, the number of required measurements $M = O(K \log N)$ can be detailed as follows (Foucart & Rauhut, 2013):

$$M \geq C \cdot K \ln(N) \ln(\epsilon^{-1}) \tag{4}$$

for some constant $C > 0$, where $\epsilon \in (0, 1)$ denotes the probability of an *inexact* recovery of the K-sparse signal. In particular, the signal recovery succeeds with a probability of at least $1 - \epsilon$ if (4) holds.[4]

We can then express (4) with regard to the probability of failure ϵ, which is given by

$$\mathbf{P}(\mathbf{s}^\star \neq \mathbf{s} \mid M, N, K) := \epsilon \leq \exp\left(-\frac{M}{C \cdot \ln(N)K}\right). \tag{5}$$

Thus, the probability of failure (inexact recovery) $\mathbf{P}(\mathbf{s}^\star \neq \mathbf{s} \mid M, N, K)$ is conditional upon M, N, and K. Since we are interested in the dynamic signal sparsity K, we model K as a *random variable* with 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 k and find the upper bound of failure probability as follows:

$$\mathbf{P}(\mathbf{s}^\star \neq \mathbf{s} \mid M, N) = \int_k \mathbf{P}(\mathbf{s}^\star \neq \mathbf{s} \mid M, N, K) \cdot f_K(k)\,\mathrm{d}k$$
$$\leq \int_k \exp\left(-\frac{M}{C \cdot \ln(N)K}\right) f_K(k)\,\mathrm{d}k. \tag{6}$$

Therefore, we can state that a signal recovery succeeds with a probability of at least $1 - \int_k \exp(-M/(C \cdot \ln(N)K))f_K(k)\mathrm{d}k$, given the distribution of signal sparsity $f_K(k)$.

The upper bound in (6) may have an analytic solution when K follows certain distributions such as the *inverse Gaussian distribution* and the *gamma distribution*.[5]

Remark Assuming $f_K(k) = \mathrm{IG}(\mu, \lambda)$, the upper bound of (6) is

$$\frac{\sqrt{\lambda}\exp(\lambda/\mu - \sqrt{2\lambda/\mu^2}\sqrt{M/(C \cdot \ln(N)) + \lambda/2})}{\sqrt{2M/(C \cdot \ln(N)) + \lambda}}, \tag{7}$$

where μ and λ are the mean and the shape parameter of the inverse Gaussian distribution, respectively.

Remark Assuming $f_K(k) = \mathrm{Gamma}(\kappa, \theta)$, the upper bound of (6) is

$$\frac{2}{\Gamma(\kappa)}\left(\frac{M}{C \cdot \ln(N)\theta}\right)^{\kappa/2} K_{-\kappa}\left(2\sqrt{\frac{M}{C \cdot \ln(N)\theta}}\right), \tag{8}$$

---

[4]See Theorem 12.20 (Foucart & Rauhut, 2013).

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

where κ and θ are the shape parameter and the scale parameter of the gamma distribution, respectively; $\Gamma(\cdot)$ is the gamma function; and $K_{-\kappa}(\cdot)$ is the modified Bessel function of the second kind.

3.2 MODELING SUCCESS AND FAILURE

Unfortunately, the probability of signal recovery failure ϵ given in (5) does not hold in practice because there is a discrepancy between the failure probabilities in the CS framework and actual random sampling, as will be further explained in Section 5.1. Thus we have to model the success or failure probability of signal recovery from a new perspective.

We can model the new pdf of *signal recovery success* using the mixture of the *Dirac delta function* and the *beta distribution*, which incorporates both stochastic and deterministic cases. We introduce $K_{\min}$ and $K_{\max}$ to denote the minimum and the maximum signal sparsities which yield stochastic probability, as opposed to a deterministic result where signal recovery always succeeds or always fails.

Definition (Recovery Success Model). *Let* $\mathbf{P}(\mathbf{s}^\star = \mathbf{s} | M, N) := \Pi$. *The pdf of Π given K is given by*[6]

$$f_{\Pi|K}(\pi \mid k) := \begin{cases} \delta(\pi - 1) & k < K_{\min} \\ \mathrm{Beta}(\alpha_K, \beta_K) & K_{\min} \le k \le K_{\max} \\ \delta(\pi) & K_{\max} < k \end{cases} \quad (9)$$

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 k as follows:

$$\begin{aligned} f_\Pi(\pi) &= \int_k f_{\Pi|K}(\pi \mid k) f_K(k) \, \mathrm{d}k \\ &= \int_0^{K_{\min}} \delta(\pi - 1) f_K(k) \, \mathrm{d}k \\ &\quad + \int_{K_{\min}}^{K_{\max}} \mathrm{Beta}(\alpha_K, \beta_K) \cdot f_K(k) \, \mathrm{d}k \\ &\quad + \int_{K_{\max}}^{\infty} \delta(\pi) f_K(k) \, \mathrm{d}k \\ &= \delta(\pi - 1) F_K(K_{\min}) + \delta(\pi)(1 - F_K(K_{\max})) \\ &\quad + \int_{K_{\min}}^{K_{\max}} \mathrm{Beta}(\alpha_K, \beta_K) \cdot f_K(k) \, \mathrm{d}k, \quad (10) \end{aligned}$$

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

The two Dirac delta function terms in (10) can be interpreted as probability masses. Since $\int_{K_{\min}}^{K_{\max}} \mathrm{Beta}(\alpha_K, \beta_K) \cdot f_K(k) \, \mathrm{d}k$ does not have an analytic solution, we compute this value numerically.
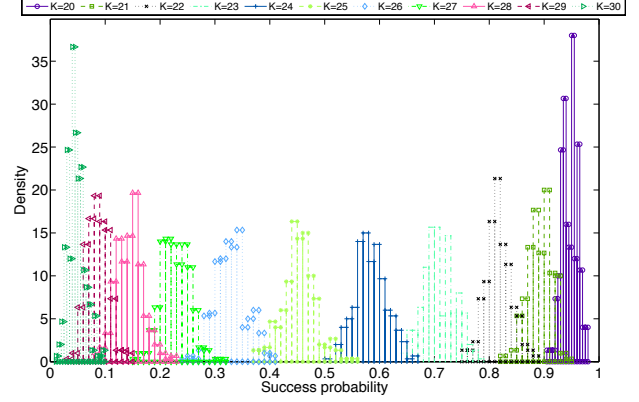

Figure 2: Histograms of success probability for various K's. Success probability distribution for each K was obtained with 300 different random signed spike vectors for $N = 512$ and $M = 100$. A single success probability for each signed spike vector was calculated with 300 experiments.

As an illustrative example, suppose that we examine the success probability by generating many different signed spike (± 1) vectors for each signal sparsity and then performing experiments for each signed spike vector.[7] Figure 2 shows histograms of success probability for various signal sparsities, where $K_{\min} = 20$ and $K_{\max} = 30$.

The success probability shown in Figure 2 naturally follows the beta distribution with its parameters α and β depending on signal sparsity, i.e., $\mathbf{P}(\mathbf{s}^\star = \mathbf{s} \mid M, N, K) \sim \mathrm{Beta}(\alpha_K, \beta_K)$.[8] The beta distribution is well known as the conjugate prior for the Bernoulli and the binomial distributions which are ideal for modeling success/failure.

3.3 MODELING ACCURACY

Here, we present the main theoretical contribution: the recovery success model defined in (9) is tighter than the lower bound of that in the existing CS framework explained in Section 3.1, when the number of measurements is not enough. We show the failure probability in the CS framework (5) is incapable of reflecting the actual failure probability of signal recovery. Not only can't the inequality $\mathbf{P}(\mathbf{s}^\star \ne \mathbf{s} \mid M, N, K) \le \exp(-M/(C \cdot \ln(N)K))$ 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.

---

[6]$\mathrm{Beta}(\alpha_K, \beta_K)$ here is used to denote the pdf of the beta distribution.

[7]Detailed settings are explained in Section 5.1.

[8]If more than 300 experiments had been performed, each success probability distribution would have been more sharply peaked due to a smaller variance.

Lemma 1 (Slackness of Recovery Success Probability). *There exists $K_0 > 0$ such that for all $K > K_0$, the lower bound of recovery success probability in the CS framework (Section 3.1) is greater than the value of a power-law-decay function.*

Proof. We need to show the following inequality

$$1 - \exp\left(-\frac{M}{C \cdot \ln(N)K}\right) > K^{-\alpha} \qquad (11)$$

holds if $K > K_0$ for some $K_0 > 0$, where $\alpha > 0$. Adding, subtracting, and taking the power K on both sides yields

$$(1 - K^{-\alpha})^K > \exp\left(-\frac{M}{C \cdot \ln(N)}\right). \qquad (12)$$

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

$$1 - K^{1-\alpha} > \exp\left(-\frac{M}{C \cdot \ln(N)}\right) \qquad (13)$$

holds if $K > K_0$ for some $K_0 > 0$.

If we assume $\alpha > 1$, then adding, subtracting, and taking the power $1/(1 - \alpha)$ on both sides of (13) yields

$$\left(1 - \exp\left(-\frac{M}{C \cdot \ln(N)}\right)\right)^{1/(1-\alpha)} < K. \qquad (14)$$

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. $\square$

Corollary 2. *In the CS framework (Section 3.1), there is always a chance of succeeding at signal recovery however large K is.*

Proof. The power-law-decay function $K^{-\alpha}$ in (11) *slowly* converges to zero as $K \to \infty$: its value is noticeably greater than zero even with a large K. As the lower bound of recovery success probability is greater than the value of the power-law-decay function for all $K > K_0$, there is always a chance of recovery success however large K is. $\square$

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 signal recovery with a number of measurements M less than K. We now show that our recovery success model provides more accurate success probability by the following theorem.

Theorem 3. *The recovery success model in (9) is tighter than the lower bound of recovery success probability given by the CS framework (Section 3.1) with a limited number of measurements.*

Proof. In contrast to Corollary 2, our recovery success model can yield $\mathbf{P}(\mathbf{s}^\star = \mathbf{s} \mid M, N, K) = 0$ with a bounded $K_{\max}$. In particular, we can let the mean of $\mathrm{Beta}(\alpha_K, \beta_K)$, $\alpha_K/(\alpha_K + \beta_K)$, converge to zero with $\alpha_{K_{\max}} \to 0$.

Similarly, we show this mean converges to one ($\mathbf{P}(\mathbf{s}^\star = \mathbf{s} \mid M, N, K) = 1$) with $K_{\min}$ 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.

We can let $\alpha_K/(\alpha_K + \beta_K)$ converge to one with $\beta_{K_{\min}} \to 0$. In contrast, $1 - \exp(-M/(C \cdot \ln(N)K)) \to 1$ if, and only if, $K \to 0$. Since $0 < K_{\min} < K_{\max} < \infty$, we can argue that our recovery success model can provide tighter recovery success probability. $\square$

3.4 PARAMETER LEARNING IN DYNAMIC SYSTEMS

When the signal sparsity K changes in dynamic systems, it does not change in an abrupt manner; rather, it tends to smoothly change over time (Vaswani & Lu, 2010; Ziniel & Schniter, 2013). One simple way to model this correlation between K's is to utilize the *Markov model* (Ziniel & Schniter, 2013). In this setup, each K makes up a state and each state is associated with the recovery success probability. This can be best modeled by the hidden Markov model, where each state K generates success/failure according to the emission probability.

In our scenario, signal recovery success is observed in an environment where the signal sparsity varies over time. We want to estimate parameters of the hidden Markov model, especially the emission probabilities. Since our recovery success model employs the beta distribution as conjugate distributions (prior and posterior), we can learn its parameters α_K and β_K for each state K.

Specifically, the decoder can observe signal recovery success/failure and corresponding signal sparsity K at each decoding step. Then using these emission and state sequences, it can sequentially update the parameters α_K and β_K for each state K (Durbin et al., 1998), by simply incrementing the value of α_K by 1 for each success or the value of β_K by 1 for each failure.

In order to prevent over-fitting with insufficient observations, it is preferable to have hyperparameters of the prior beta distribution set according to K's. In Figure 2, we can clearly see the trend of α_K and β_K for different K's: α_K decreases, whereas β_K increases as K grows.[9] Therefore, these hyperparameters can act as the effective numbers of observations of recovery success/failure, reflecting our prior belief on signal recovery success for different K's.

---

[9]Also, see the proof of Theorem 3.

4 FURTHER ANALYSIS ON RECOVERY QUALITY

When a signal of interest is not exactly K-sparse but compressible, as discussed in Section 2.1, the signal recovery in Section 2.3 can be treated from a different perspective (Lee & Choi, 2015). In particular, the inequality (3) is considered 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. We get progressively better results as we compute more measurements M, since $M = O(K \log N)$ (Candès & Wakin, 2008). Therefore, $\mathbf{\Psi s}^\star \in \mathbb{R}^N$ also makes progress on its quality as M increases.[10]

From this viewpoint, the success or failure of signal recovery no longer exists. Rather, we can view the number of components included in the signal recovery as a *varying quantity*. Specifically, if a signal recovery is about to fail with a given K, then K can be lowered to make the recovery eventually succeed. Here the number of *included* 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, (3) can be utilized to infer varying K's over different recoveries and signals. Here our assumption is that the upper bound in (3) is *tight* such that we solve the following optimization problem:

$$\max K \qquad \text{subject to} \qquad \|\mathbf{s}^\star - \mathbf{s}\|_2 \leq C_1 \cdot \|\mathbf{s} - \mathbf{s}_K\|_1. \tag{15}$$

In (15), C_1 has to be determined, where the maximum signal sparsity $K_{\max}$ introduced in Section 3.2 plays a key role to set the upper limit on how large K can be, since $K > K_{\max}$ is not reasonable.

In particular, we can generate a compressible signal $\mathbf{s}_i \in S$ such that $\|\mathbf{s}_i\|_1 = C_{\ell_1}$ and $\|\mathbf{s}_i\|_2 = C_{\ell_2}$ for all i, where S is the set containing many different signals; $C_{\ell_1} > 0$ and $C_{\ell_2} > 0$ are constants. For each $\mathbf{s}_i$, we have a set $S_i^\star$ which contains many different recoveries $\mathbf{s}_{ij}^\star$. Then C_1 can be found as follows:

$$C_1 = \frac{\min \|\mathbf{s}_{ij}^\star - \mathbf{s}_i\|_2}{\|\mathbf{s}_i - \mathbf{s}_i^{K_{\max}}\|_1}, \tag{16}$$

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

Varying K's obtained through (15) can be represented by a pdf, which has been empirically shown to follow the *gamma distribution* (Lee & Choi, 2015). We are interested

in the shape of this pdf, which is shown by the following proposition.

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

Proof. Since $\|\mathbf{s}_i\|_1 = C_{\ell_1}$ and $\|\mathbf{s}_i\|_2 = C_{\ell_2}$ for all i, we can conceive the same sequence $\{s_n\}$ of elements (absolute values) in $\mathbf{s}_i$ for all i. Then we have

$$\|\mathbf{s}_i - \mathbf{s}_i^K\|_1 = \sum_{n=1}^{N-K} s_n. \tag{17}$$

Without loss of generality, we consider the partial sum $\sum_{n=1}^{N-K} s_n$ in (17) to be an *arithmetic series* which can be represented by a quadratic function in terms of K. We also assume the inequality constraint in (15) is the equality constraint such that $\|\mathbf{s}^\star - \mathbf{s}\|_2 = C_1 \cdot \|\mathbf{s} - \mathbf{s}_K\|_1$.

If we take the (partial) inverse function of the quadratic function, we have $K \sim K_{\max} - \sqrt{\|\mathbf{s}^\star - \mathbf{s}\|_2 - (\min \|\mathbf{s}_{ij}^\star - \mathbf{s}_i\|_2)}$. Assuming the distribution of $\|\mathbf{s}^\star - \mathbf{s}\|_2$ is *symmetric* (zero skewness), this asymptotic relation says $\|\mathbf{s}^\star - \mathbf{s}\|_2$ will be *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 *geometric series*, where $K \sim N - \log(\|\mathbf{s}^\star - \mathbf{s}\|_2)$. In this case, the pdf of K is skewed to the right as well. □

4.1 ERROR ANALYSIS IN DYNAMIC SYSTEMS

Since the success or failure of signal recovery does not exist in this framework, we instead investigate the amount of error occurring during the recovery procedure in an *expected value* sense. In particular, the best K-term approximation $\|\mathbf{s} - \mathbf{s}_K\|_1$ in (3) is known to be bounded as follows (Baraniuk et al., 2010):

$$\|\mathbf{s} - \mathbf{s}_K\|_1 \leq \frac{2G}{K}, \tag{18}$$

where the constant G can be learned by the power-law decay such that each magnitude of components in $\mathbf{s}$, sorted in decreasing order, is upper bounded by G/i^2. ($i = 1, \ldots, N$ is the sorted index.)

Then we can analyze the ℓ_2 error E of signal recovery assuming $f_K(k) = \text{Gamma}(\kappa, \theta)$, which is given by

$$E = \int_k C_1 \cdot \frac{2G}{k} f_K(k) \, \mathrm{d}k = \frac{2C_1 G}{\theta} \text{B}(\kappa - 1, 1), \tag{19}$$

where $\text{B}(\cdot, \cdot)$ is the beta function (Lee & Choi, 2015). Here the pdf $f_K(k)$ is employed to represent varying K's.[11]

---

[10] The error bound follows (3) as well if $\mathbf{\Psi}$ is an orthogonal matrix, which is usually the case.

[11] Note that this pdf is different from the one introduced in Section 3.

In this framework, there is no longer such an indicator as the timely varying signal sparsity K in Section 3, because signals are compressible and their coefficients are already populated with small, but non-zero, coefficients. Thus, we may assume the same gamma distribution over time, whose parameters κ and θ can then be estimated.

In order to prevent over-fitting to insufficient observations, we introduce the conjugate prior for the gamma distribution. It is known that the conjugate prior of the gamma distribution has the following form (Miller, 1980; Fink, 1997).

$$\mathbf{P}(\kappa, \theta \mid p, q, r, s) = \frac{1}{Z} \cdot \frac{p^{\kappa-1} \exp(-q/\theta)}{\Gamma(\kappa)^r \theta^{s\kappa}}, \qquad (20)$$

where $p, q, r, s > 0$ are hyperparameters that are sequentially updated with $p' = pk$, $q' = q + k$, $r' = r + 1$, and $s' = s + 1$;[12] and the normalizing constant Z is

$$Z = \int_0^\infty \frac{p^{\kappa-1}\Gamma(s\kappa+1)}{\Gamma(\kappa)^r q^{s\kappa+1}} \, \mathrm{d}\kappa. \qquad (21)$$

Using (19) and (20), we can marginalize over κ and θ to estimate error $\widehat{E}$ as follows:

$$\begin{aligned}
\widehat{E} &= \int_\kappa \int_\theta E \cdot \mathbf{P}(\kappa, \theta \mid p, q, r, s) \, \mathrm{d}\theta \, \mathrm{d}\kappa \\
&= \frac{2C_1 G}{Z} \int_0^\infty \frac{p^{\kappa-1}}{(\kappa-1)\Gamma(\kappa)^r} \int_0^\infty \frac{\exp(-q/\theta)}{\theta^{s\kappa+1}} \, \mathrm{d}\theta \, \mathrm{d}\kappa \\
&= \frac{2C_1 G}{Z} \int_0^\infty \frac{p^{\kappa-1}\Gamma(s\kappa)}{(\kappa-1)\Gamma(\kappa)^r q^{s\kappa}} \, \mathrm{d}\kappa, \qquad (22)
\end{aligned}$$

which can be computed numerically.

5 EXPERIMENTAL RESULTS

5.1 RECOVERY SUCCESS

In Section 3, we discussed the discrepancy between the failure probabilities in the CS framework and actual random sampling. In order to show this discrepancy, we *artificially* generated signed spikes ± 1 at random locations in proportion to desired sparsities and *densified* these spikes using $\boldsymbol{\Psi}$[13] to perform the random sampling.

For each signal sparsity K, the actual failure probability can be calculated for different recovery experiments. To this end, we adopted a standard optimization method (*basis pursuit*) to solve the optimization problem in (2) (Chen et al., 1998). Specifically, the primal-dual algorithm based on the interior point method was employed to solve (2) (Boyd & Vandenberghe, 2004).

---

[12]Here, p', q', r', and s' are updated posterior hyperparameters; and k is a single observation.

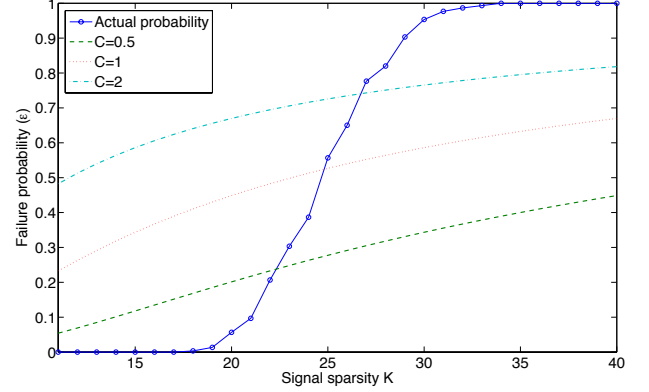[13]We used DCT as the sparsifying basis $\boldsymbol{\Psi}$ throughout experiments.

Figure 3: Comparison between actual failure probability and failure probabilities given in (5) with varying C's. Actual failure probability of each signal sparsity K was obtained with 300 experiments for $N = 512$ and $M = 100$.

Figure 3 shows that the 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 (5) cannot model the actual failure probability of signal recovery, regardless of the value chosen for constant C. This result confirms Lemma 1 and Corollary 2.

Moreover, in Section 3.2 we modeled the new pdf of signal recovery success $f_\Pi(\pi)$ in (10). We compared this new pdf with the upper bound of failure in (6), given a dynamic signal sparsity K. Specifically, we employed the inverse Gaussian distribution such that $f_K(k) = \mathrm{IG}(30, 200)$. Figure 4 exhibits the efficacy of our recovery success model, where the lower bounds of success probability given in the CS framework fail to capture actual success probability in random sampling case. This result confirms Theorem 3.

Note that our recovery success model provides the baseline of recovery success for any CS frameworks that are specifically designed to handle varying signal sparsity. For instance, Figure 5 shows histograms of success probability for various signal sparsities using Modified-CS (Vaswani & Lu, 2010).[14] Compared with Figure 2, the success probability shown in Figure 5 also follows the beta distribution; but success probability is higher than that of basis pursuit for a given sparsity K ($K_{\min} = 21$ and $K_{\max} = 31$), thanks to the ability of Modified-CS to handle dynamic signal sparsity. The recovery success model in (9) is still effective here as a theoretical framework, or the recovery success model using the basis pursuit may promise a minimum guarantee for the recovery success of other CS frameworks.

We also employed real-world environmental data sets ob-

---

[14]Results were obtained with two frames where the second frame has one more spike than the first frame so that Modified-CS could exploit smoothly varying signal sparsity. Histograms in Figure 5 are the success probabilities of the second frame.
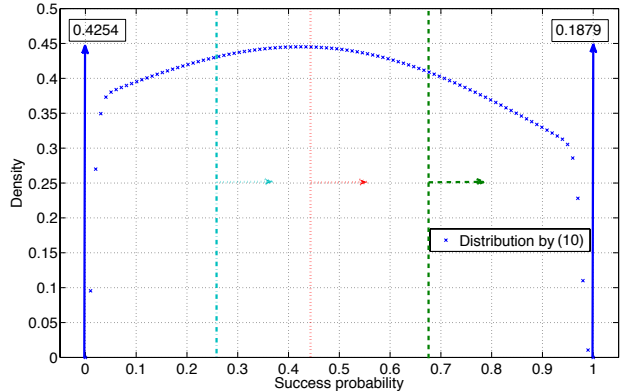
Figure 4: Comparison between our new success probability distribution in (10) and the lower bounds of success probability obtained by (6) with varying C's. The inverse Gaussian distribution was used for $f_K(k)$. Two probability masses are shown by vertical arrows, where solid boxes atop the arrows denote their probabilities. Three vertical dashed/dotted lines represent the lower bounds by (6): $C = 0.5$ at 0.6781; $C = 1$ at 0.4450; and $C = 2$ at 0.2596. Here, $K_{\min} = 20$ and $K_{\max} = 30$.
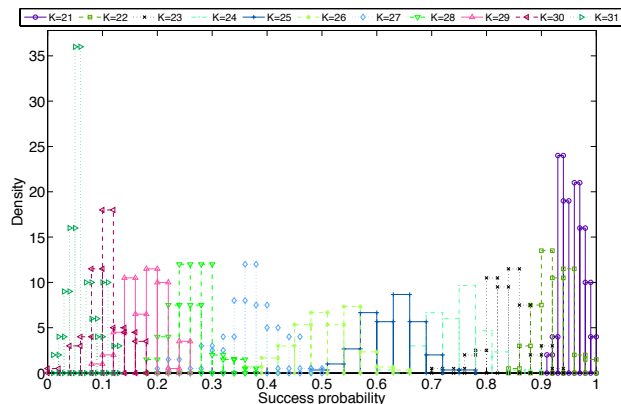

Figure 5: Histograms of success probability for various K's using Modified-CS (Vaswani & Lu, 2010). Success probability distribution for each K was obtained with 100 different random signed spike vectors for $N = 512$ and $M = 100$. A single success probability for each signed spike vector was calculated with 100 experiments. Compared with the results with basis pursuit in Figure 2, success probability is higher for a given sparsity K. Note also that 100 experiments resulted in a larger variance for each K.

tained from wireless sensor network deployments (Quer et al., 2012): humidity and temperature. In addition, audio data shown in Figure 1 was used for comparison as well. Random numbers representing the dynamic signal sparsity K were drawn from the inverse Gaussian distribution ($f_K(k) = \mathrm{IG}(30, 200)$) and we used this K to randomly choose components sorted in decreasing order; other

components were set to zero. Figure 6 shows the success probability of signal recovery follows the shape of Figure 4.

5.2 RECOVERY QUALITY

When a signal is compressible and not exactly K-sparse, this signal is basically *dense*. In Section 4, we regarded the number of components included in the signal recovery as a varying quantity. We are interested in the general shape of this quantity in distribution. In order to verify Proposition 1, we performed experiments using real data sets as well as artificially generated random signed spikes.

We first provide results with real-world data sets to verify Proposition 1. Figure 7 displays the histograms of K, the number of 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 all distributions are skewed to the right. Furthermore, the distributions follow the gamma distribution, which is also natural since the gamma distribution has positive skewness, i.e., right tailed.

In addition, 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 signal recovery always fails when $K > 30$ in our case, as shown in Figure 3. Figure 8 displays the histogram of K and the gamma distribution fitting, where we can again see that Proposition 1 holds.

Furthermore, we analyze the ℓ_2 error E of signal recovery assuming $f_K(k) = \mathrm{Gamma}(\kappa, \theta)$ using (19). In order to show its efficacy, we compared the solutions of (19) with real data sets. For humidity data, $E = 94.3533$ while the average ℓ_2 norm of data is 564.8585; for temperature data, $E = 75.5441$ while the average ℓ_2 norm of data is 627.8038; and for audio data, $E = 5.0979$ while the average ℓ_2 norm of data is 1.5866. Apart from the case of audio data, (19) provides useful estimators for the upper bound of amount of error during recovery. It should be noted that this bound is rather loose due to a large constant G in (18), which could be improved with a less conservative G.

6 CONCLUSION

We have presented a new theoretical CS framework in random sampling which handles uncertainty in signal recovery from a new perspective. The success probability of signal recovery in random sampling was investigated when signal sparsity can vary with an insufficient number of measurements. The success probability analysis in the existing CS framework was shown to be incapable of reflecting actual
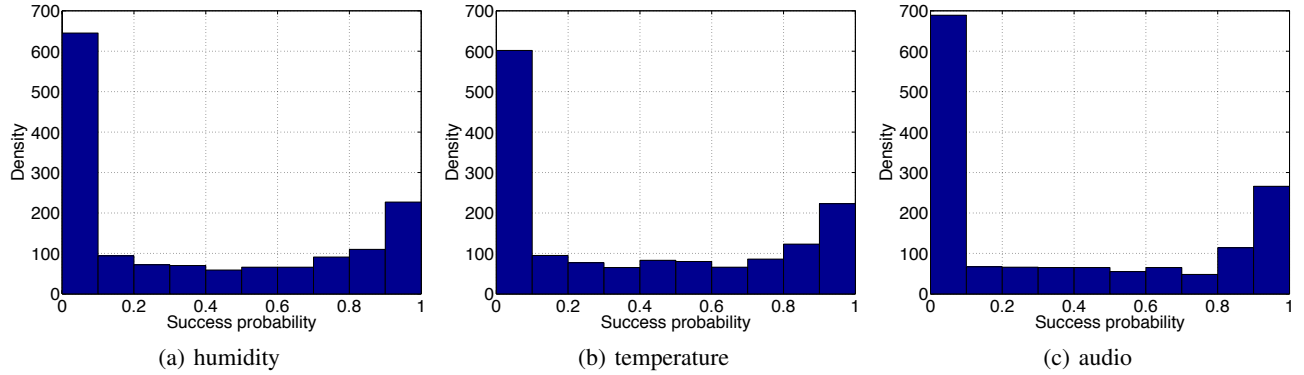
(a) humidity

(b) temperature

(c) audio

Figure 6: Histograms of success probability for (a) humidity data, (b) temperature data, and (c) audio data. Histogram was obtained with 1,500 random number generations (the inverse Gaussian distribution) to choose different signals and 100 different experiments for each signal with $N = 512$ and $M = 100$. All histograms closely follow the shape of Figure 4.
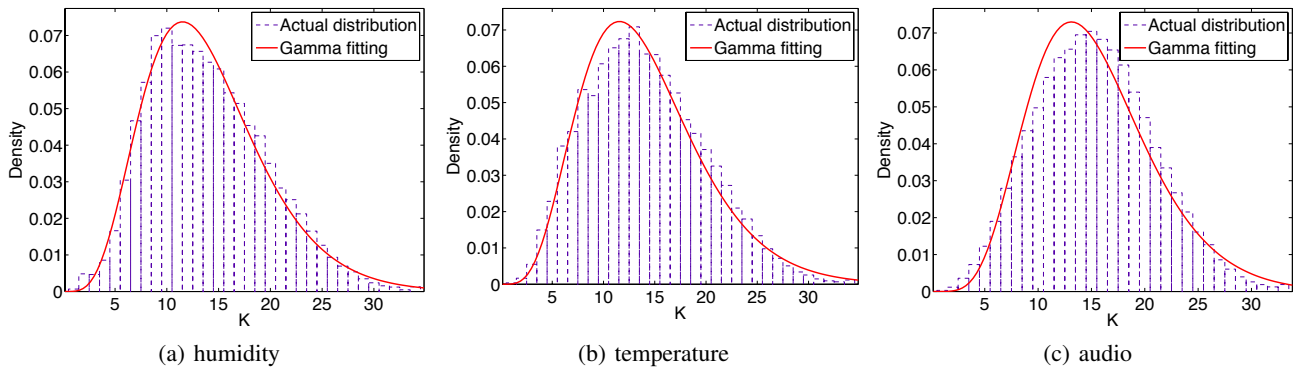

(a) humidity

(b) temperature

(c) audio

Figure 7: Distributions of K fitted with gamma distributions for (a) humidity data ($\mathrm{Gamma}(5.69, 2.45)$), (b) temperature data ($\mathrm{Gamma}(5.56, 2.54)$), and (c) audio data ($\mathrm{Gamma}(6.92, 2.21)$). Histograms were obtained with 34 different signals and 1,000 different experiments for each signal (a and b); with 153 different signals and 500 different experiments for each signal (c), with $N = 512$ and $M = 100$.
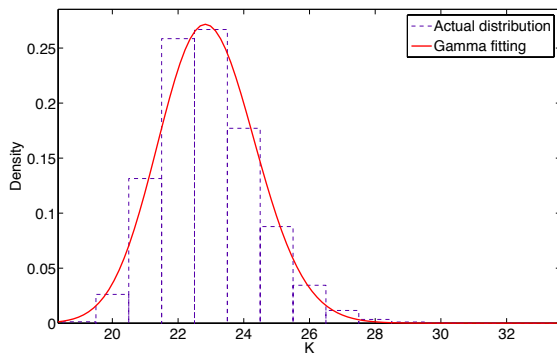

Figure 8: Distribution of K fitted with a gamma distribution $\mathrm{Gamma}(242.81, 0.09)$, using the maximum likelihood estimation. Histogram was obtained with 300 different signals and 300 different experiments for each signal, with $N = 512$ and $M = 100$.

success probability by both theoretical analysis and experiments. On the contrary, our recovery success model could closely reflect actual success probability.

We also considered signals which cannot be exactly represented with sparse representations, where we could alternatively view the number of components included in the signal recovery as a varying quantity. This quantity was shown by both theoretical analysis and experiments to follow a right-tailed distribution such as the gamma distribution. We provided error analysis for these signals.

Acknowledgements

References

Baraniuk, Richard G. Compressive sensing [lecture notes]. *IEEE Signal Process. Mag.*, 24(4):118–121, Jul. 2007.

Baraniuk, Richard G., Cevher, Volkan, Duarte, Marco F., and Hegde, Chinmay. Model-based compressive sensing. *IEEE Trans. Inf. Theory*, 56(4):1982–2001, Apr. 2010.

Blumensath, Thomas and Davies, Mike E. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14(5-6):629–654, Dec. 2008.

Boyd, Stephen and Vandenberghe, Lieven. *Convex Optimization*. Cambridge University Press, 2004.

Candès, Emmanuel J. and Wakin, Michael B. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, Mar. 2008.

Chen, Scott S., Donoho, David L., and Saunders, Michael A. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, Jan. 1998.

Durbin, Richard, Eddy, Sean R, Krogh, Anders, and Mitchison, Graeme. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

Fink, Daniel. A compendium of conjugate priors, 1997.

Foucart, Simon and Rauhut, Holger. *A Mathematical Introduction to Compressive Sensing*. Springer, 2013.

Hsu, Daniel, Kakade, Sham, Langford, John, and Zhang, Tong. Multi-label prediction via compressed sensing. In *Proc. NIPS*, pp. 772–780, 2009.

Lee, Dongeun and Choi, Jaesik. Low complexity sensing for big spatio-temporal data. In *Proc. IEEE BigData*, pp. 323–328, 2014.

Lee, Dongeun and Choi, Jaesik. Learning compressive sensing models for big spatio-temporal data. In *Proc. SDM*, pp. 667–675, 2015.

Lopes, Miles. Estimating unknown sparsity in compressed sensing. In *Proc. ICML*, pp. 217–225, 2013.

Malioutov, Dmitry and Varshney, Kush. Exact rule learning via boolean compressed sensing. In *Proc. ICML*, pp. 765–773, 2013.

Malioutov, Dmitry M., Sanghavi, Sujay R., and Willsky, Alan S. Sequential compressed sensing. *IEEE J. Sel. Top. Signal Process.*, 4(2):435–444, Apr. 2010.

Miller, Robert B. Bayesian analysis of the two-parameter gamma distribution. *Technometrics*, 22(1):65–69, Feb. 1980.

Pati, Yagyensh C., Rezaiifar, Ramin, and Krishnaprasad, P. S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. ACSSC*, pp. 40–44, 1993.

Quer, Giorgio, Masiero, Riccardo, Pillonetto, Gianluigi, Rossi, Michele, and Zorzi, Michele. Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework. *IEEE Trans. Wireless Commun.*, 11(10):3447–3461, Oct. 2012.

Seeger, Matthias W. and Nickisch, Hannes. Compressed sensing and bayesian experimental design. In *Proc. ICML*, pp. 912–919, 2008.

Sejdinovic, Dino, Andrieu, Christophe, and Piechocki, Robert. Bayesian sequential compressed sensing in sparse dynamical systems. In *Proc. Allerton*, pp. 1730–1736, 2010.

Shahrasbi, Behzad, Talari, Ali, and Rahnavard, Nazanin. TC-CSBP: Compressive sensing for time-correlated data based on belief propagation. In *Proc. CISS*, pp. 1–6, 2011.

Vaswani, Namrata and Lu, Wei. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE Trans. Signal Process.*, 58(9):4595–4607, Sep. 2010.

Zhu, Rongda and Gu, Quanquan. Towards a lower sample complexity for robust one-bit compressed sensing. In *Proc. ICML*, pp. 739–747, 2015.

Ziniel, Justin and Schniter, Philip. Dynamic compressive sensing of time-varying signals via approximate message passing. *IEEE Trans. Signal Process.*, 61(21): 5270–5284, Nov. 2013.

Ziniel, Justin, Rangan, Sundeep, and Schniter, Philip. A generalized framework for learning and recovery of structured sparse signals. In *Proc. IEEE Statistical Signal Processing Workshop*, pp. 325–328, 2012.