Sparse Representation of White Gaussian Noise with Application to \( \ell_0 \)-Norm Decoding in Noisy Compressed Sensing

Ori Shental
Qualcomm Inc.
5775 Morehouse Drive
San Diego, CA 92121, USA
Email: oshental@qualcomm.com

Abstract—The achievable and converse regions for sparse representation of white Gaussian noise based on an overcomplete dictionary are derived in the limit of large systems. Furthermore, the marginal distribution of such sparse representations is also inferred. The results are obtained via the Replica method which stems from statistical mechanics. A direct outcome of these results is the introduction of a sharp threshold for \( \ell_0 \)-norm decoding in noisy compressed sensing, and its mean-square error for underdetermined Gaussian vector channels.

I. INTRODUCTION

White Gaussian noise (WGN) is a canonical component in innumerable systems, models and applications related to information and signal theory. In this contribution, we study sparse representations of WGN. Sparse representations are based on dictionary matrices. The columns of the dictionary are termed atoms. Often, the atom elementary signals are chosen from, a so called, overcomplete dictionary for which the number of atoms exceeds the dimension of the signal space. Such a dictionary exhibits a full-rank and fat matrix. Sparse representation of a signal defines a linear combination of only a few atoms of the dictionary. In this paper we concentrate on zero mean, unit variance, i.i.d. entries (0, 1) dictionaries (like the Gaussian or Bernoulli dictionaries) in the limit of infinite dimensions.

Although being non-compressible, WGN vector realizations do have sparse representations. For example, consider a trivial sparse representation (based on a Gaussian dictionary) with Hamming weight equal to the dimension of the WGN vector. Such a sparse representation is clearly achievable since the resultant square dictionary matrix is almost surely invertible in the large-system limit. This square matrix is generated by concatenating together the columns of the original fat dictionary matrix corresponding to the non-zero entries of the sparse representation. Evidently, denser representations are also attainable, up to the fully dense representation. On the other hand, it is clear that a sparse representation consisting of only a single atom does not necessarily exist. That is true since the probability the WGN realization is identical to one of the atom columns of the Gaussian dictionary instance has a measure zero.

This observation raises some interesting, yet so far unresolved questions:

1) Can one go below the trivial sparsity when representing WGN?
2) if so, what is the minimal Hamming weight delimiting the achievable and converse regions of WGN sparse representations?
3) and how do such representations look like in the sense of probability density function?

In this paper, we answer these fundamental questions by utilizing the Replica method born from the study of disordered systems in statistical physics. The Replica method [1], despite being mathematically non-rigorous [2], has proven itself as a powerful tool in solving open problems in information theory and communications (see, e.g., [3]), and particularly in the analysis of sparse signals [4]–[8]. One immediate consequence of our analysis of WGN sparse representation is in revealing a sharp threshold for \( \ell_0 \)-norm decoding in noisy compressed sensing [9]–[11] and its mean-square error performance.

The paper is organized as follows. The problem formulation is first introduced in Section II. Section III investigates the characteristics of sparse representation of WGN in the large-system limit, while in Section IV the latter is applied for the analysis of \( \ell_0 \)-norm decoding in noisy compressed sensing. We conclude in Section V.

II. PROBLEM FORMULATION

Let \( \mathbf{w} \in \mathbb{R}^m \) be a WGN vector of dimensions \( m \in \mathbb{N}^* \) with i.i.d. entries \( w_i \sim \mathcal{N}(0, 1), i = 1, \ldots, m \)[1]. Consider an overcomplete dictionary \( \mathbf{D} \in \mathbb{R}^{m \times n} \) with \( n = m/\alpha \in \mathbb{N}^* \) atoms and zero mean, unit variance, i.i.d. entries \( D_{ij}, j = 1, \ldots, n \). Examples for such dictionaries are the Gaussian, \( D_{ij} \sim \mathcal{N}(0, 1) \), and Bernoulli, \( D_{ij} \pm 1 \), dictionaries. The scalar \( \alpha \in (0, 1) \) is termed measurement ratio. The vector \( \mathbf{w} \) and matrix \( \mathbf{D} \) are statistically independent. The realizations of the WGN, \( \mathbf{w} \), and dictionary, \( \mathbf{D} \), are denoted by \( \omega \) and \( \mathcal{D} \), respectively.

[1] The symbols \{ \} and \{ \}_ij denote entries of a vector and matrix, respectively.
Let \( z(\omega, D) \in \mathbb{R}^n \) be a representation of the WGN instance, \( \omega \), via a certain overcomplete dictionary, \( D \), namely
\[
\omega = \frac{1}{\sqrt{n}} D z(\omega, D).
\]
The representation vector \( z(\omega, D) \equiv z_{\kappa}(\omega, D) \), explaining an observed \( \omega \) given the dictionary \( D \), is termed \( \kappa \)-sparse representation if at most \( k = \kappa n \in \mathbb{N}^+ \) of its entries are non-zero, i.e. \( \|z(\omega, D)\|_0/n \leq \kappa \), where \( \kappa \in (0, 1) \) is the sparsity fraction. The \( \ell_0 \) norm of a vector is defined as \( \|z\|_0 \triangleq \# \{ i \in \{1, \ldots, n\} \mid z_i \neq 0 \} \).

Hereinafter in this paper, a large-system limit is assumed: The WGN dimension and the number of dictionary atoms go to infinity, \( i.e. m, n \to \infty \) respectively, but with fixed sparsity fraction, \( \kappa \), and measurement ratio \( \alpha \).

The central question under investigation in this paper is as follows. What is the normalized Hamming weight, or sparsity fraction \( \kappa^{\star} \), of the sparsest representation of WGN based on a \( \alpha \)-measurement dictionary in the limit of large systems. Mathematically speaking, we are targeting at \( \kappa^{\star} \) such that
\[
z_{\kappa^{\star}}(\omega, D) = \arg \min \|z\|_0 \text{ subject to } \omega = \frac{1}{\sqrt{n}} D z. \tag{2}
\]

The minimal normalized Hamming weight, \( \kappa^{\star} \), as expressed in (2) is a function of the specific realizations \( \omega \) and \( D \). Owing to the self-averaging property \( [1] \), one can alternatively define \( \kappa^{\star} \) in terms of averaging over all realizations, making it amenable to evaluation. The self-averaging property, in the context of our analysis, is described in the following assumption. Note that herein the symbol \( \mathbb{E} \{ \cdot \} \) denotes expectation of the random object within the brackets with respect to (w.r.t.) the subscript random variables.

Assumption 1. The limit \( \kappa^{\star}_{\alpha} \triangleq \lim_{n \to \infty} \kappa^{\star}_{\alpha}(\omega, D) \) exists and it is equal to its average over the randomness of the WGN and dictionary, \( \lim_{n \to \infty} \mathbb{E}_{\omega, D} \{ \kappa^{\star}_{\alpha}(\omega, D) \} \), for almost all realizations of the WGN and dictionary.

Based on Assumption 1 the next section is dedicated to the explicit computation of the minimal sparsity fraction, \( \kappa^{\star} \). This minimal Hamming weight is the key for better understanding of the principal characteristics of WGN sparse representations as described in the rest of this contribution.

III. SPARSE REPRESENTATION OF WGN

A. Achievable and Converse Regions

The fundamental questions [1] and [2] brought up in the Introduction are addressed in the following claim.

Claim 2. Consider the scalars \( \alpha \in (0, 1) \) and \( \kappa^{\star}_{\alpha} \triangleq 2 Q(\xi) \in (0, 1) \), where \( \xi \geq 0 \) is determined by
\[
\alpha = \sqrt{2} \int_{\xi}^{\infty} t^2 \exp(-t^2/2) dt, \tag{3}
\]
and \( Q(\xi) \triangleq \int_{\xi}^{\infty} dt/\sqrt{2\pi} \exp(-t^2/2) \) is the Q-function.

Then, with probability 1 in the large-system limit, for a zero mean, unit variance, i.i.d. dictionary with measurement ratio \( \alpha \):

- i) (minimal Hamming weight) the sparsest WGN representation is \( \kappa^{\star}_{\alpha} \)-sparse;
- ii) (achievable region) \( \kappa \)-sparse representation of WGN exists only for \( \kappa \geq \kappa^{\star}_{\alpha} \);
- iii) (converse region) \( \kappa \)-sparse representation of WGN does not exist for \( \kappa < \kappa^{\star}_{\alpha} \).

Figure 1 illustrates the outcome of Claim 2. Interestingly, for a given measurement ratio \( \alpha \), we can go below the trivial \( \alpha \)-sparse representation down to the sparsest representation which is \( \kappa^{\star}_{\alpha} \)-sparse. Note also that for the well-posed case of \( \alpha = 1 \), one gets \( \kappa^{\star}_{1} = 1 \) as expected, that is only a dense representation exists. The proof of Claim 2 is as follows.

Proof:

Define an energy or cost function
\[
\mathcal{E}_{\alpha}(\tilde{z}, \omega, D, m, n) \triangleq \frac{1}{m} \sum_{i=1}^{m} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tilde{D}_{ij} \tilde{z}_j - \omega_i \right)^2. \tag{4}
\]

Note in passing, that the vector \( \tilde{z} \in \mathbb{R}^n \) is used in the definition of the cost function (4), rather than \( z \), since the latter denotes a (sparse) representation (1) which may not necessarily exist for certain values of \( \kappa \) and instances \( \omega \) and \( D \).

The energy function (4) can be rewritten equivalently as
\[
\mathcal{E}_{\kappa}(\zeta, b, \omega, D, m, n) \triangleq \frac{1}{m} \sum_{i=1}^{m} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{n} D_{ij} b_j \zeta_j - \omega_i \right)^2, \quad \tag{5}
\]
where \( \zeta_j \triangleq b_j \zeta_j, \zeta \in \mathbb{R}^n \) and the binary variable \( b_j \in \{0, 1\} \) has a Bernoulli parameter \( \kappa = (\sum_{j=1}^{n} b_j)/n \). Based on Assumption 1 on the self-averageness in the large-system limit, one can state that almost surely
\[
\mathcal{E}_{\kappa}(\zeta, b, \omega, D, m, n) \xrightarrow{a.s.} \lim_{n \to \infty} \mathbb{E}_{\omega, D} \{ \mathcal{E}_{\kappa}(\zeta, b, \omega, D, m, n) \} \triangleq \mathcal{E}_{\kappa}(\zeta, b, \alpha). \tag{6}
\]

The minimal energy can be obtained from the limit of zero

![](image-url)
temperature $1/\beta = 0$

$$\xi_\kappa^\text{min}(\alpha) = \frac{1}{\alpha} \lim_{\beta \to \infty} \frac{\partial (\beta F(\beta, \alpha))}{\partial \beta},$$  

(7)

where $F$ is the normalized average free energy

$$-\beta F(\beta, \alpha) \triangleq \lim_{n \to \infty} \frac{1}{n} \mathbb{E}_{\mathbf{w}, \mathbf{D}} \{ \log Z(\mathbf{w}, \mathbf{D}, \beta, m, n) \},$$  

(8)

and the partition function is defined as

$$Z(\mathbf{w}, \mathbf{D}, \beta, m, n) \triangleq \sum_{\mathbf{b}} \delta \left( n - \sum_{j=1}^n b_j \right) \int_{-\infty}^\infty \prod_{j=1}^n d\zeta_j$$

$$\times \exp \left( -\beta \kappa \mathbb{E}_\kappa(\mathbf{\zeta}, \mathbf{b}, \mathbf{w}, \mathbf{D}, m, n) \right).$$  

The function $\delta(\cdot)$ denotes the Kronecker delta.

In the limit of zero temperature, $\beta \to \infty$, (7) only the vector $\mathbf{z}$ which minimizes the energy $\xi_\kappa$ will eventually contribute to the partition function $Z$ and consequently to the free energy $F$. Other solutions vanish exponentially in the summation (9).

Based on the definitions of the sparse representation (1) and energy function (4), observe that a zero minimal energy, $\xi_\kappa^\text{min} = 0$, implies the existence of $\kappa$-sparse representations, while evidently for $\xi_\kappa^\text{min} > 0$ there is no such representation.

The quenched average in (8) is too complicated to be computed directly, thus it is carried out via the Replica method [1]. This method relies on the mathematical identity

$$\mathbb{E}_{\mathbf{w}, \mathbf{D}} \{ \log Z \} \equiv \lim_{r \to 0} \frac{1}{r} \log \mathbb{E}_{\mathbf{w}, \mathbf{D}} \{ Z^r \}.  

(10)

According to the replica trick, one first evaluates $\mathbb{E}_{\mathbf{w}, \mathbf{D}} \{ Z^r \}$ for integer $r$ and then continues analytically to $r = 0$. Applying a standard replica analysis, and particularly a replica-symmetric (RS) ansatz [1], à la Gardner and Derrida [12], [13], the minimal energy gets the form

$$\xi_\kappa^\text{min}(\alpha) = \lim_{\beta \to \infty} \frac{1 + Q + \beta(Q - q)^2}{2(1 + \beta(Q - q))^2},$$  

(11)

where the squared $\ell_2$-norm of $\mathbf{z}$ is

$$Q \triangleq \frac{1}{n} \sum_{j=1}^n b_j \zeta_j^2$$  

(12)

and

$$q = q_{ab} = \frac{1}{n} \sum_{j=1}^n b_a^j b_b^j \zeta_j^\alpha \zeta_j^\beta \quad \forall a < b$$  

(13)

is the replica-symmetric (i.e., independent of replica indices $a, b \in \mathbb{N}^\ast$) physical order parameter. The latter parameter measures the overlap between different solutions $\mathbf{z}_a$ and $\tilde{\mathbf{z}}_b$. The $Q$ and $q$ (and other saddle-point) parameters are obtained by the extremization of $\mathcal{F}_{\text{RS}}$, the replica-symmetric average free-energy density. For a finite scalar

$$x \triangleq \beta(Q - q),$$  

(14)

the minimization of $\mathcal{F}_{\text{RS}}$ w.r.t. $Q$ yields the squared norm of the optimal solution

$$Q = x = \frac{\alpha_\kappa^*}{\alpha - \alpha_\kappa^*},$$  

(15)

where the threshold

$$\alpha_\kappa^* \triangleq \left( \frac{2}{\pi} \int_0^\infty t^2 \exp (-t^2/2) dt \right)^{1/2},$$  

(16)

and $\xi$ is the unique solution of

$$\left( \frac{2}{\pi} \int_0^\infty dt \exp (-t^2/2) = \kappa. \right. \quad \left. (17) \right.$$  

Hence in the region

$$\alpha > \alpha_*^\kappa,$$  

(18)

the minimal energy boils down to

$$\xi_\kappa^\text{min}(\alpha) = \frac{1}{1 + Q} = \frac{1}{1 - \alpha/\alpha_\kappa^*},$$  

(19)

implying that there is no sparse representation in this region. Also, in this region of finite auxiliary parameter $x$ [14], since $\beta \to \infty$ we get $Q \equiv q$ meaning that there is only a single solution $\mathbf{z}$ for which the energy gets its minimal value $\xi_\kappa^\text{min}$. Instead of expressing this region in terms of inequality on the measurement ratio [18], one can describe it using the sparsity fraction of the solution $\mathbf{z}$ by requiring

$$\kappa < \kappa_*^\alpha,$$  

(20)

where the sparsity threshold

$$\kappa_*^\alpha \triangleq \left( \frac{2}{\pi} \int_0^\infty dt \exp (-t^2/2) \right)^{1/2},$$  

(21)

and this time the integral limit $\xi$ is determined by solving

$$\alpha = \left( \frac{2}{\pi} \int_0^\infty t^2 \exp (-t^2/2) dt \right)^{1/2} \quad \left. (22) \right.$$  

analogously to (16) and (17). This yields the converse region claim [15].

On the other hand for the complementary case of $\kappa \geq \kappa_*^\alpha$, one finds $x \to \infty$ [14], thus from the minimal energy expression (11) we learn that in this region $\xi_\kappa^\text{min} > 0$, establishing the achievability statement [11]. In the region, since $Q \neq q$ there is an infinite number of zero-energy solutions per $(\kappa, \alpha)$-point. Hence, the threshold $\kappa = \kappa_*^\alpha$ describes, with probability 1, the minimal possible Hamming weight of the sparsest WGN representation, as claimed in [1]. This concludes the proof. ■

B. Achieving Distribution

We now aim at question 3 from the Introduction and examine the achieving probability density function of the WGN $\kappa$-sparse representation via an $\alpha$-measurement dictionary. The results are summarized in the next claim.

Claim 3. The marginal probability density function of the $j$’th (non-zero) entry, $\zeta_j$, of the (minimal $\ell_2$-norm) $\kappa$-sparse

2The RS ansatz is locally stable and conjectured to be globally stable also.
representation of WGN, $z_\kappa$, is given in the large-system limit by

$$p(\zeta_j) = \begin{cases} 
0 & \text{if } |\zeta_j| < \epsilon \\
\sqrt{\frac{\alpha_\kappa^2 (\alpha_\kappa^2 - \alpha)}{2\pi \alpha \kappa^2}} \exp\left(-\frac{\zeta_j^2 \alpha_\kappa^2 (\alpha_\kappa^2 - \alpha)}{2\alpha}\right) & \text{otherwise}
\end{cases}$$

where $\alpha_\kappa^\ast$ is the achievability threshold for given sparsity fraction $\kappa$.

Proof:
The proof of Claim 3 is also based on replica analysis and is deferred to another publication.

Figure 2 gives some examples of densities for different ($\kappa, \alpha$)-points. A few remarks on Claim 3 are in place. Note that the derived probability density function is only the marginal one, $p(\zeta_j)$. What is the joint distribution of $\zeta$ is an interesting, yet challenging open question. In the large-system limit the achieving probability density is not a function of the WGN vector and dictionary realizations. There is an infinite number of sparse representations per ($\kappa, \alpha$)-point in the achievable region. Furthermore, the stated achieving distribution corresponds to the representation with the minimal $\ell_2$ norm. Observe that, remarkably, as $\alpha \to \alpha_\kappa^\ast$, the measure zero gap in the probability densities increases to infinity, and the non-vanishing part of the distribution becomes more uniform.

According to Claim 3 a given measurement ratio $\alpha$ determines an optimal "compression" rate $\kappa_\alpha^\ast$. For this rate the sparsest representation consists of $\kappa_\alpha^\ast n$ entries of infinite value. The locations of these infinite spikes, distributed uniformly over the $n$ entry indices, are dictated by the specific WGN instance. The rest of the $(1 - \kappa_\alpha^\ast) n$ entries are set to zero. Thus the number of non-zero entries describing the noise instance is reduced w.r.t. its original length, i.e. $k < m$. However since their locations are noise-realization dependent, an $n > m$-length representation is still required. Hence this is only a so called compressed sparse representation rather than a real compression, which, as is well known, does not exist for WGN.

C. Experimental Results

The theoretical minimal Hamming weight (Claim 2) is corroborated by computer simulations. First, WGN vector, $\omega$, and Gaussian dictionary, $D$, are being generated. Then the sparsest representation per realization is inferred using the iteratively reweighted least-squares (IRLS) method [14]. The IRLS method is a tractable approximation of $\ell_0$-norm minimization [2]. The sparsity fraction of the simulated sparsest representation is then averaged over sufficiently large ensemble of realizations. This procedure is being repeated for several number of atoms, $n$, varying from 40 up to 200. Figure 3 displays the averaged simulated minimal sparsity fraction versus the reciprocal of the number of dictionary atoms, $1/n$, for an example case of $\alpha = 0.5$. A quadratic fitting is then applied so to extrapolate the minimal Hamming weight for infinite $n$ (i.e., the crossing point with the vertical axis at $1/n = 0$). Figure 3 presents the extrapolated minimal sparsity fraction, $\kappa_\alpha^\ast$, as a function of the measurement ratio, $\alpha$, range. Note that a fairly good agreement is obtained between the simulation-based $\kappa_\alpha^\ast$ and the theoretical curve $\kappa_\alpha^\ast$. The discrepancy between theory and simulations may be explained by first extrapolation errors and second by the fact that IRLS method is only an approximation to $\ell_0$-norm optimization due to the tendency of the former to converge to local, rather than global minimum.

Another empirical study is devoted to affirming the derived marginal achieving probability (Claim 3). To this end, we first generate a Gaussian dictionary, $D$, and a sparse representation, $z$. The latter is being generated according to the stochastic profile provided by Claim 3 (i.e., the corresponding probability densities as depicted, for instance, by the blue Gaussian tail in Figure 2). The non-zero entry locations in
the sparse representation are chosen uniformly at random. Thus, the vector \( w = Dz / \sqrt{n} \) is being calculated. Repeating this process for a large ensemble of realizations, we build a histogram for the vector \( w \). Figure 4 displays the values of a standard Gaussian distribution, \( \mathcal{N}(0,1) \), in the horizontal axis and the values of the obtained histogram in its vertical axis. For the case the histogram would be a perfect Gaussian, then the \( \times \)-markers should fall exactly on the reference solid line. One may observe that the experimentally generated noise, \( w \), is approximately Gaussian, as the \( \times \)-marks fall in the vicinity of the straight line. Note that this mismatch is very well expected, mainly due to the inaccurate implicit assumption, taken in this simulation study, about the non-zero entries of the sparse representation being i.i.d. and taken only based on the marginal probabilities, rather than the joint one (which is unfortunately unknown). Furthermore, this empirical test does not say anything whether the generated (approximately) Gaussian noise vector, \( w \), is white or not. In the next section, we explore some consequences of what we have learned so far on sparse representations of WGN on the analysis of \( \ell_0 \)-norm optimization in noisy compressed sensing (CS, [9]–[11]).

IV. NOISY COMPRESSED SENSING

Consider a \( \kappa_x \in \{0,1\} \)-sparse data vector, \( x_\kappa \in \mathbb{R}^n \), with a finite second moment. Suppose a noiseless zero mean, unit variance and i.i.d. (e.g., Gaussian) linear transformation, with ratio \( \alpha \), of the data, \( y = Dx_\kappa \), is observed. We are interested in perfectly reconstructing the data from the underdetermined measurements. According to CS literature (e.g., [15]) given the \( \ell_0 \)-norm decoder

\[
\hat{x} = \arg \min ||x||_0 \quad \text{subject to} \quad y = Dx_\kappa \quad (24)
\]

then a prefect recovery, \( \hat{x} = x_\kappa \), can be achieved for almost any \( x_\kappa \), with probability 1, if \( \alpha > \kappa_x \). This is termed weak, or typical noiseless \( \ell_0 \)-norm decodable region\(^3\). Exact reconstruction is impossible with overwhelming probability for \( \alpha \leq \kappa_x \), which is the strong converse region.

Unfortunately, the \( \ell_0 \)-norm decoder is prohibitively complex exhibiting an NP-complete optimization problem since it requires combinatorial enumeration of the \( \binom{n}{\kappa_x} \) possible sparse vectors. One of the existing wonders of CS is that the \( \ell_0 \) norm in the optimization problem [21] could be replaced by an \( \ell_1 \) norm (\( ||x||_1 \equiv \sum_i |x_i| \)). This replacement turns [21] into a tractable optimization problem with polynomial complexity, but miraculously it still generates perfect reconstruction, \( \hat{x} = x_\kappa \), with probability 1. However, the feasibility of the \( \ell_1 \)-norm decoder emerges at the cost of more required linear measurement (i.e., larger \( \alpha \) per given sparsity \( \kappa_x \)). Nevertheless, the \( \ell_0 \)-norm decoder is of major theoretical importance as it bounds the performance of practical reconstruction methods.

Moving to the more realistic case of noisy compressed sensing, consider an overloaded Gaussian vector channel

\[
y = \sqrt{\frac{\text{snr}}{m}} Dx_\kappa + \omega, \quad (25)
\]

where \( \text{snr} \) is the signal-to-noise ratio (SNR) gain of the channel. Examples of such Gaussian vector channel include, to name a few, CDMA and MIMO communication systems. For CDMA (code-division multiple-access) channel the Bernoulli dictionary is mapped onto the spreading matrix, while the number of measurements, \( m \), and atoms, \( n \), correspond to the processing gain and number of active users, respectively. Similarly, for MIMO (multiple-input multiple-output) system in Gaussian fading the number of measurements, \( m \), and atoms, \( n \), are translated to the number of receiving and transmitting antennas, respectively.

Let \( z_{\kappa_x} \) be a \( \kappa_x \)-sparsest WGN representation. Therefore the overloaded Gaussian vector channel can be reformulated as

\[
y = \sqrt{\frac{\text{snr}}{m}} Dx_\kappa + \omega = \sqrt{\frac{1}{m}} D(\sqrt{\text{snr}}x_\kappa + \sqrt{\omega}z_{\kappa_x}) \triangleq \sqrt{\frac{1}{m}} Dx^*, \quad (26)
\]

where \( x^* \triangleq (\sqrt{\text{snr}}x_\kappa + \sqrt{\omega}z_{\kappa_x}) \) is the sparsest explanation.
of the observations $y$ given the dictionary/channel $\mathbf{D}$. Thus interestingly based on Claim 2, the noisy channel with $\kappa_x$-sparse data input vector may be mapped into an equivalent noiseless channel with a denser $(\kappa_x + \kappa_x - \kappa_x \kappa_x)$-sparse input vector as shown in Figure 5. The subtraction of $\kappa_x \kappa_x$ accounts for the partial overlap between the $\kappa_x$ finite non-zero entries of the input vector and the $\kappa_x$ infinite entries of WGN sparsest representation determined by nature. Although the infinite-value entries of $\mathbf{z}$, it is noteworthy that the sparsest vector $\mathbf{x}^*$, which explains the observations $y$ and is generated via the $l_0$-norm optimization problem (24), may consist of only finite non-zero entries. An extreme example to this observation occurs when a sparse (but not the sparsest) representation of the WGN realization is formed by $(\kappa_x + \kappa_x - \kappa_x \kappa_x) n$ non-zero entries which are finite (since this representation no longer resides on the border of the achievable region) and fully covers the indices of the $\kappa_x$ finite non-zero entries of the data input. Still in either case the minimal sparsity of the total $\mathbf{x}^*$ remains unchanged. The above argument is summarized in the following corollary leaning on Claim 2 and 3.

**Corollary 4.** Given a Gaussian vector channel (25) with measurement ratio $\alpha \in (0, 1)$ and arbitrary SNR $> 0$, then in the large-system limit an $l_0$-norm decoder (24) results, with probability 1, in $(\kappa_x + \kappa_x - \kappa_x \kappa_x)$-sparse vector with support $\Omega_0$, where $\kappa_x$ is the channel input sparsity fraction and $\kappa_x^*$ is the minimal normalized Hamming weight per given measurement ratio $\alpha$. The support, $\Omega$, of the data input vector, $\mathbf{x}$, maintains $\Omega \subseteq \Omega_0$.

The following claim states the noisy counterpart to the $l_0$-decodable and converse regions as described at the beginning of this section for the noiseless case.

**Claim 5.** Given a Gaussian vector channel (25) with measurement ratio $\alpha \in (0, 1)$ and arbitrary SNR $> 0$, then in the large-system limit an $l_0$-norm decoder (24) results, with probability 1, in average mean-square error (MSE) per unknown

$$\frac{1}{n} \| \hat{x} - x \|^2_2 = \frac{\kappa_x + \kappa_x^* - \kappa_x \kappa_x^*}{\text{SNR}},$$

(27)

as long as

$$\kappa_x \leq \frac{\alpha - \kappa_x^*}{1 - \kappa_x^*}$$

(28)

for almost any $x$. Otherwise $l_0$-reconstruction is impossible with overwhelming probability.

Figure 6 draws the sharp threshold (28) and displays the $l_0$-decodable region in $(\kappa_x, \alpha)$ plane for the noisy case. Also marked is the classical threshold for the noiseless case. As may be expected, the existence of an ambient noise polluting the observations increases the threshold on the number of linear measurements required for optimal recovery. However, note that the noisy threshold itself is insensitive to the SNR level. Note also that in the absence of any other information (e.g.,

4The $\sqrt{\lambda}$ scaling of $\mathbf{z}$ in (26) is due to the fact that the Gaussian vector channel is normalized by $\sqrt{\lambda}$, so as to maintain unity power per atom, while the definition $\| \mathbf{z} \|^2_2$ of the sparse representation uses normalization by $\sqrt{\lambda}$.

prior on $x$), reconstruction with MSE which is proportional to the SNR, as stated in Claim 5, is the best one can hope for. The derived MSE (27) is also comparable to the MSE of an oracle decoder, $\kappa_x / \text{SNR}$, magically knowing the support $\Omega$ of the original data input.

**Proof:**

Combining the classical result from CS theory on the $l_0$-norm decodable region for noiseless CS with Corollary 4 immediately answers the question of what are the decodable and converse regions of $l_0$-norm decoder for additive WGN compressed sensing. As depicted in Figure 7, the $l_0$-norm decoder can successfully operate as long as the sum of the minimal “physical” sparsity originated from the WGN vector realization (on which we have no control), $\kappa_x^*$, and the net sparsity of the channel input data, $\kappa_x - \kappa_x^* \kappa_x$, is less than the noiseless reconstruction threshold. Thus the operability of the $l_0$-norm decoder demands

$$\kappa_x^* + \kappa_x - \kappa_x^* \kappa_x \leq \alpha,$$

(29)

which establishes the threshold (28).

Now based on the $l_0$-norm recovery, the support $\Omega_0$ of the sparsest representation, $\mathbf{x}^*$, of the observation vector, $\mathbf{y}$, is
known. Thus eliminating atoms (columns) of the dictionary corresponding to indices which are not in $\Omega_0$, the problem becomes well-posed, and one could optimally reconstruct $x$ with the Least-Squares (LS) method

$$
\hat{x}_{LS} = \begin{cases} 
\sqrt{\frac{m}{\text{snr}}} (D^T_{\Omega_0} D_{\Omega_0})^{-1} D^T_{\Omega_0} y & \text{on } \Omega_0 \\
0 & \text{elsewhere}
\end{cases}
$$

(30)

The MSE of this LS solution results in (27).

V. Conclusion

A sharp threshold for the achievability of sparse representation of WGN is introduced via Replica method. The marginal distribution of such sparse representations is derived, showing that the sparsest representation is composed of infinite-value entries. Based on this WGN analysis, we have also established sharp threshold for $\ell_0$-norm decoding in noisy compressed sensing and its corresponding MSE.

Bear in mind that for any orthonormal basis matrix $\Psi$ (e.g., DFT matrix) and a Gaussian dictionary $D$, the matrix $D \Psi$ will be also a Gaussian dictionary, thus the discussed results apply for these case too. Extension of this analysis to other dictionaries is called for. Also, it may be of a major interest to search for applications, insights and consequences of the WGN sparse representation analysis to various fields, like data hiding, cryptography and watermarking.

References

[1] V. S. Dotsenko, *Introduction to the Replica Theory of Disordered Statistical Systems*. Cambridge, UK: Cambridge University Press, 2001.

[2] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models*. Berlin, Germany: Springer, 2003.

[3] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. Inf. Theory*, vol. 48, pp. 2888–2910, Nov. 2002.

[4] Y. Kabashima, T. Wadayama, and T. Tanaka, “A typical reconstruction limit for compressed sensing based on $L_p$-norm minimization,” *J. Stat. Mech.*, p. L09003, 2009.

[5] S. Rangan, A. K. Fletcher, and V. K. Goyal, “Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing,” 2009. [Online]. Available: arXiv:0906.3234v2

[6] D. Guo, D. Baron, and S. Shamai, “A single-letter characterization of optimal noisy compressed sensing,” in *Proc. 47th Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, Sep. 2009.

[7] Y. Kabashima, T. Wadayama, and T. Tanaka, “Statistical mechanical analysis of a typical reconstruction limit of compressed sensing,” in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 1533–1537.

[8] K. Takeda and Y. Kabashima, “Statistical mechanical analysis of compressed sensing using a correlated compression matrix,” in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 1538–1542.

[9] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[10] D. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[11] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[12] E. Gardner, “The space of interactions in neural network models,” *J. Phys. A*, vol. 21, pp. 257–270, 1988.

[13] E. Gardner and B. Derrida, “Optimal stages properties of neural network models,” *J. Phys. A*, vol. 21, pp. 271–284, 1988.

[14] R. Chartrand, “Iteratively reweighted algorithms for compressive sensing,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, USA, Mar. 2008, pp. 3869–3872.

[15] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, “Distributed compressed sensing,” Rice University, Tech. Rep., 2005.