Weighted Superimposed Codes and Constrained Integer Compressed Sensing

Wei Dai and Olgica Milenkovic
Dept. of Electrical and Computer Engineering
University of Illinois, Urbana-Champaign

Abstract

We introduce a new family of codes, termed weighted superimposed codes (WSCs). This family generalizes the class of Euclidean superimposed codes (ESCs), used in multiuser identification systems. WSCs allow for discriminating all bounded, integer-valued linear combinations of real-valued codewords that satisfy prescribed norm and non-negativity constraints. By design, WSCs are inherently noise tolerant. Therefore, these codes can be seen as special instances of robust compressed sensing schemes. The main results of the paper are lower and upper bounds on the largest achievable code rates of several classes of WSCs. These bounds suggest that with the codeword and weighting vector constraints at hand, one can improve the code rates achievable by standard compressive sensing.

I. Introduction

Superimposed codes (SCs) and designs were introduced by Kautz and Singleton [1], for the purpose of studying database retrieval and group testing problems. In their original formulation, superimposed designs were defined as arrays of binary codewords with the property that bitwise OR functions of all sufficiently small collections of codewords are distinguishable. Superimposed designs can therefore be viewed as binary "parity-check" matrices for which syndromes represent bitwise OR, rather than XOR, functions of selected sets of columns.

The notion of binary superimposed codes was further generalized by prescribing a distance constraint on the OR evaluations of subsets of columns, and by extending the fields in which the codeword symbols lie [2]. In the latter case, Ericson and Györfi introduced Euclidean superimposed codes (ESCs), for which the symbol field is $\mathbb{R}$, for which the OR function is replaced by real addition, and for which all sums of less than K codewords are required to have pairwise Euclidean distance at least d. The best known upper bound on the size of Euclidean superimposed codes was derived by Füredi and Ruszinko [3], who used a combination of sphere packing arguments and probabilistic concentration formulas to prove their result.

On the other hand, compressed sensing (CS) is a new sampling method usually applied to K-*sparse signals*, i.e. signals embedded in an N-dimensional space that can be represented by only $K \ll N$ significant coefficients [4]–[6]. Alternatively, when the signal is projected onto a properly chosen basis of the transform space, its accurate representation relies only on a small number of coefficients. Encoding of a K-sparse discrete-time signal $\mathbf{x}$ of dimension N is accomplished by computing a measurement vector $\mathbf{y}$ that consists of $m \ll N$ linear projections, i.e. $\mathbf{y} = \Phi\mathbf{x}$. Here, Φ represents an $m \times N$ matrix, usually over the field of real numbers. Consequently, the measured vector represents a linear combination of columns of the matrix Φ, with weights prescribed by the nonzero entries of the vector $\mathbf{x}$. Although the reconstruction of the signal $\mathbf{x} \in \mathbb{R}^N$ from the (possibly noisy) projections is an ill-posed problem, the prior knowledge of signal sparsity allows for accurate recovery of $\mathbf{x}$.

*This work is supported by the NSF Grant CCF 0644427, the NSF Career Award, and the DARPA Young Faculty Award of the second author. Parts of the results were presented at the CCIS'2008 ant ITW'2008 conferences.

The connection between error-correcting coding theory and compressed sensing was investigated by Candès and Tao in [7], and remarked upon in [8]. In the former work, the authors studied random codes over the real numbers, the noisy observations of which can be decoded using linear programming techniques. As with the case of compressed sensing, the performance guarantees of this coding scheme are probabilistic, and the K-sparse signal is assumed to lie in $\mathbb{R}^N$.

We propose to study a new class of codes, termed *weighted superimposed codes* (WSCs), which provide a link between SCs and CS matrices. As with the case of the former two entities, WSCs are defined over the field of real numbers. But unlike ESCs, for which the sparse signal $\mathbf{x}$ consists of zeros and ones only, and unlike CS, for which $\mathbf{x}$ is assumed to belong to $\mathbb{R}^N$, in WSCs the vector $\mathbf{x}$ is drawn from B^N, where B denotes a *bounded, symmetric set of integers*. The motivation for studying WSCs comes from the fact that in many applications, the alphabet of the sparse signal can be modeled as a finite set of integers.

Codewords from the family of WSCs can be designed to obey prescribed norm and non-negativity constraints. The restriction of the weighting coefficients to a bounded set of integers ensures reconstruction robustness in the presence of noise - i.e., all weighted sums of at most K codewords can be chosen at "minimum distance" d from each other. This minimum distance property provides deterministic performance guarantees, which CS techniques usually lack. Another benefit of the input alphabet restriction is the potential to reduce the decoding complexity compared to that of CS reconstruction techniques. This research problem was addressed by the authors in [9]–[11], but is beyond the scope of this paper.

The central problem of this paper is to characterize the rate region for which a WSC with certain parameters exists. The main results of this work include generalized sphere packing upper bounds and random coding lower bounds on the rates of several WSC families. The upper and lower bounds differ only by a constant, and therefore imply that the superposition constraints are ensured whenever $m = O\left(K \log N / \log K\right)$. In the language of CS theory, this result suggests that the number of required signal measurements is less than the standard $O\left(K \log\left(N/K\right)\right)$, required for discriminating real-valued linear combinations of codewords. This reduction in the required number of measurements (codelength) can be seen as a result of restricting the input alphabet of the sparse signal.

The paper is organized as follows. Section III introduces the relevant terminology and definitions. Section IV contains the main results of the paper – upper and lower bounds on the size of WSCs. The proofs of the rate bounds are presented in Sections V and VIII. Concluding remarks are given in Section IX.

II. Motivating Applications

We describe next two applications - one arising in wireless communication, the other in bioengineering - motivating the study of WSCs.

The adder channel and signature codes: One common application of ESCs is for signaling over multi-access channels. For a given set of $k \leq K$ active users in the channel, the input to the receiver $\mathbf{y}$ equals the sum of the signals (signatures) $\mathbf{x}_{i_j}$, $j = 1, \ldots, k$, of the k active users, i.e. $\mathbf{y} = \sum_{j=1}^{k} \mathbf{x}_{i_j}$. The signatures are only used for identification purposes, and in order to minimize energy consumption, all users are assigned unit energy [2], [3]. Now, consider the case that in addition to identifying their presence, active users also have to convey some limited information to the receiver by adapting their transmission power. The received signal can in this case be represented by a weighted sum of the signatures of active users, i.e., $\mathbf{y} = \sum_{j=1}^{k} \sqrt{p_{i_j}} \mathbf{x}_{i_j}$. The codebook used in this scheme represents a special form of WSC, termed Weighted Euclidean Superimposed Codes (WESCs); these codes are formally defined in Section III.

Compressive sensing microarrays: A microarray is a bioengineering device used for measuring the level of certain molecules, such as RNA (ribonucleic acid) sequences, representing the joint expression profile of thousands of genes.

A microarray consist of thousands of microscopic spots of DNA sequences, called probes. The complementary DNA (cDNA) sequences of RNA molecules being measured are labeled with fluorescent tags, and such units are termed targets. If a target sequence has a significant homology with a probe sequence on the microarray, the target cDNA and probe DNA molecules will bind or "hybridize" so as to form a stable structure. As a result, upon exposure to laser light of the appropriate wavelength, the microarray spots with large hybridization activity will be illuminated. The specific illumination pattern and intensities of microarray spots can be used to infer the concentration of RNA molecules. In traditional microarray design, each spot of probes is a unique identifier of only one target molecule. In our recent work [12], [13], we proposed the concept of *compressive sensing microarrays* (CSM), for which each probe has the potential to hybridize with several different targets. It uses the observation that, although the number of potential target RNA types is large, not all of them are expected to be present in a significant concentration at all observed times.

Mathematically, a microarray is represented by a measurement matrix, with an entry in the i^{th} row and the j^{th} column corresponding to the hybridization probability between the i^{th} probe and the j^{th} target. In this case, all the entries in the measurement matrix are nonnegative real numbers, and all the columns of the measurement matrix are expected to have l_1-norms equal to one. In microarray experiments, the input vector $\mathbf{x}$ has entries that correspond to integer multiples of the smallest detectable concentration of target cDNA molecules. Since the number of different target cDNA types in a typical test sample is small compared to the number of all potential types, one can assume that the vector $\mathbf{x}$ is sparse. Furthermore, the number of RNA molecules in a cell at any point in time is upper bounded due to energy constraints, and due to intracellular space limitations. Hence, the integer-valued entries of $\mathbf{x}$ are assumed to have bounded magnitudes and to be relatively small compared to the number of different RNA types. With the above considerations, the measurement matrix of a CSM can be described by nonnegative l_1-WSCs, formally defined in Section III.

III. Definitions and Terminology

Throughout the paper, we use the following notation and definitions.

A code $\mathcal{C}$ is a finite set of N codewords (vectors) $\mathbf{v}_i \in \mathbb{R}^{m\times 1}$, $i = 1, 2, \cdots, N$. The code $\mathcal{C}$ is specified by its *codeword matrix (codebook)* $\mathbf{C} \in \mathbb{R}^{m\times N}$, obtained by arranging the codewords in columns of the matrix.

For two given positive integers, t and K, let

$$B_t = [-t, t] = \{-t, -t+1, \cdots, t-1, t\} \subset \mathbb{Z}$$

be a symmetric, bounded set of integers, and let

$$\mathcal{B}_K = \left\{\mathbf{b} \in B_t^N : \ \|\mathbf{b}\|_0 \leq K\right\}$$

denote the l_0 ball of radius K, with $\|\mathbf{b}\|_0$ representing the number of nonzero components in the vector $\mathbf{b}$ (i.e., the support size of the vector). We formally define WESCs as follows.

Definition 1: A code $\mathcal{C}$ is said to be a WESC with parameters (N, m, K, d, η, B_t) for some $d \in (0, \eta)$, if

1) $\mathbf{C} \in \mathbb{R}^{m\times N}$,
2) $\|\mathbf{v}_i\|_2 = \eta$, for all $i = 1, \cdots, N$, and,
3) if the following minimum distance property holds:

$$d_E(\mathcal{C}, K, B_t) := \min_{\mathbf{b}_1 \neq \mathbf{b}_2} \|\mathbf{C}\mathbf{b}_1 - \mathbf{C}\mathbf{b}_2\|_2 \geq d$$

for all $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}_K$.

Henceforth, we focus our attention on WESCs with $\eta = 1$, and denote the set of parameters of interest by (N, m, K, d, B_t).

The definition above can be extended to hold for other normed spaces.

Definition 2: A code $\mathcal{C}$ is said to be an l_p-WSC with parameters (N, m, K, d, B_t) if

1) $\mathbf{C} \in \mathbb{R}^{m \times N}$,
2) $\|\mathbf{v}_i\|_{l_p} = 1$, for all $i = 1, \cdots, N$, and,
3) if the following minimum distance property holds:

$$d_p(\mathcal{C}, K, B_t) := \min_{\mathbf{b}_1 \neq \mathbf{b}_2} \|\mathbf{C}\mathbf{b}_1 - \mathbf{C}\mathbf{b}_2\|_{l_p} \geq d$$

for all $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}_K$.

Note that specializing $p = 2$ reproduces the definition of a WESC.

Motivated by the practical applications described in the previous section, we also define the class of *nonnegative* l_p-WSC.

Definition 3: A code $\mathcal{C}$ is said to be a nonnegative l_p-WSC with parameters (N, m, K, d, B_t) if it is an l_p-WSC such that all entries of $\mathbf{C}$ are nonnegative.

Given the parameters m, K, d and B_t, let $N(m, K, d, B_t)$ denote the maximum size of a WSC,

$$N(m, K, d, B_t) := \max\{N : \mathcal{C}(N, m, K, d, B_t) \neq \phi\}.$$

The *asymptotic code exponent* is defined as

$$R(K, d, B_t) := \limsup_{m \to \infty} \frac{\log N(m, K, d, B_t)}{m}.$$

We are interested in quantifying the asymptotic code exponent of WSCs, and in particular, WESCs and nonnegative WSCs with $p = 1$. Results pertaining to these classes of codes are summarized in the next section.

IV. On the Cardinality of WSC Families

The central problem of this paper is to determine the existence of a superimposed code with certain parameters. In [2], [3], it was shown that for ESCs, for which the codeword alphabet B_t is replaced by the asymmetric set $\{0, 1\}$, one has

$$\frac{\log K}{4K}(1 + o_d(1)) \leq R(K, d, \{1\}) \leq \frac{\log K}{2K}(1 + o_d(1)),$$

where $o_d(1)$ converges to zero as $K \to \infty$.

The main result of the paper is the upper and lower bounds on the asymptotic code exponents of several WSC families. For WESCs, introducing weighting coefficients larger than one does not change the asymptotic order of the code exponent.

Theorem 1: Let t be a fixed parameter. For sufficiently large K, the asymptotic code exponent of WESCs can be bounded as

$$\frac{\log K}{4K}(1 + o(1)) \leq R(K, d, B_t) \leq \frac{\log K}{2K}(1 + o_{t,d}(1)) \tag{1}$$

where $o(1) \to 0$ and $o_{t,d}(1) \to 0$ as $K \to \infty$. The exact expressions of the $o(1)$ and $o_{t,d}(1)$ terms are given in Equations (19) and (7), respectively.

Remark 1: The derivations leading to the expressions in Theorem 1 show that one can also bound the code exponent in a non-asymptotic regime. Unfortunately, those expressions are too complicated for practical use.

Nevertheless, this observation implies that the results pertaining to WESC are applicable for the same parameter regions as those arising in the context of CS theory.

Remark 2: The parameter t can also be allowed to increase with K. For WESCs, the value of t does not affect the lower bound on the asymptotic code exponent, while the upper bound is valid as long as $t = o(K)$.

For clarity of exposition, the proof of the lower bound is postponed to Section VI, while the proof of the upper bound, along with the proofs of the upper bounds for other WSC families, are presented in Section V. We briefly sketch the main steps of the proofs in the discussion that follows.

The proof of the upper bound is based on the sphere packing argument. The classical sphere packing argument is valid for all WSC families discussed in this paper. The leading term of the resulting upper bound is $(\log K)/K$. This result can be improved when restricting one's attention to the Euclidean norm. The key idea is to show that most points of the form $\mathbf{Cb}$ lie in a ball of radius significantly smaller than the one derived by the classic sphere packing argument. The leading term of the upper bound can in this case be improved from $(\log K)/K$ to $(\log K)/(2K)$.

The lower bound in Theorem 1 is proved by random coding arguments. We first randomly generate a family of WESCs from the Gaussian ensemble, with the code rates satisfying

$$\lim_{(m,N)\to\infty} \frac{\log N}{m} < \frac{\log K}{4K}\left(1 + o(1)\right).$$

Then we prove that these randomly generated codebooks satisfy

$$d_E(\mathcal{C}, K) \geq d$$

with high probability. This fact implies that the asymptotic code exponent

$$\begin{aligned} R(K, d, B_t) &= \limsup_{m\to\infty} \frac{\log N(m, K, d, B_t)}{m} \\ &\geq \frac{\log K}{4K}\left(1 + o(1)\right). \end{aligned}$$

We also analyze two more classes of WSCs: the class of general l_1-WSCs and the family of nonnegative l_1-WSCs. The characterization of the asymptotic code rates of these codes is given in Theorems 2 and 3, respectively.

Theorem 2: For a fixed value of the parameter t and sufficiently large K, the asymptotic code exponent of l_1-WSCs is bounded as

$$\frac{\log K}{4K}\left(1 + o(1)\right) \leq R(K, d, B_t) \leq \frac{\log K}{K}\left(1 + o_{t,d}(1)\right), \tag{2}$$

where the expressions for $o(1)$ and $o_{t,d}(1)$ are given in Equations (31) and (6), respectively.

Proof: The lower bound is proved in Section VII, while the upper bound is proved in Section V. ■

Theorem 3: For a fixed value of the parameter t and sufficiently large K, the asymptotic code exponent of *nonnegative* l_1-WSCs is bounded as

$$\frac{\log K}{4K}\left(1 + o_t(1)\right) \leq R(K, d, B_t) \leq \frac{\log K}{K}\left(1 + o_{t,d}(1)\right), \tag{3}$$

where the expressions for $o_t(1)$ and $o_{t,d}(1)$ are given by Equations (40) and (6), respectively.

Proof: The lower and upper bounds are proved in Sections VIII and V, respectively. ■

Remark 3: The upper bounds in Equations (2) and (3) also hold if one allows t to grow with K, so that $t = o(K)$. The lower bound in (2) for general l_1-WSCs does not depend on the value of t. However, the lower bound (3) for nonnegative l_1-WSCs requires that $t = o\left(K^{1/3}\right)$ (see Equation (40) for details). This difference in the convergence regime of the two l_1-WSCs is a consequence of the use of different proof techniques. For the proof of the rate regime of general l_1-WSCs, Gaussian codebooks were used. On the other hand, for nonnegative l_1-WSCs, the analysis is complicated by the fact that one has to analyze linear combinations of nonnegative random variables. To overcome this difficulty, we used the Central Limit Theorem and Berry-Essen type of distribution approximations [14]. The obtained results depend on the value of t.

Remark 4: The upper bound for WESCs is roughly one half of the corresponding bound for l_1-WSCs. This improvement in the code exponent of WESCs rests on the fact that the l_2-norm of a vector can be expressed as an inner product, i.e. $\|\mathbf{v}\|_2^2 = \mathbf{v}^\dagger \mathbf{v}$ (in other words, l_2 is a Hilbert space). Other normed spaces considered in the paper lack this property, and at the present, we are not able to improve the upper bounds for l_p-WSCs with $p \neq 2$.

V. Proof of the Upper Bounds by Sphere packing Arguments

It is straightforward to apply the sphere packing argument to upper bound the code exponents of WSCs. Regard an l_p-WSC with arbitrary $p \in \mathbb{Z}^+$. The superposition $\mathbf{Cb}$ satisfies

$$\|\mathbf{Cb}\|_p \leq \sum_{j=1}^{\|\mathbf{b}\|_0} \left\|\mathbf{v}_{i_j} b_{i_j}\right\|_p \leq Kt$$

for all $\mathbf{b}$ such that $\|\mathbf{b}\|_0 \leq K$, where the b_{i_j}s, $1 \leq j \leq \|\mathbf{b}\|_0 \leq K$, denote the nonzero entries of $\mathbf{b}$. Note that the l_p distance of any two superpositions is required to be at least d. The size of the l_p-WSC codebook, N, satisfies the sphere packing bound

$$\sum_{k=1}^{K} \binom{N}{k} (2t)^k \leq \left(\frac{tK + \frac{d}{2}}{\frac{d}{2}} \right)^m . \tag{4}$$

A simple algebraic manipulation of the above equation shows

$$\sum_{k=1}^{K} \binom{N}{k} (2t)^k \geq \binom{N}{K} (2t)^K \geq \left(\frac{N-K}{K} \right)^K (2t)^K ,$$

so that one has

$$\begin{aligned} \frac{\log N}{m} &\leq \frac{1}{K} \log\left(1 + \frac{2tK}{d}\right) - \frac{\log(2t)}{m} - \frac{\log\left(\frac{1}{K} - \frac{1}{N}\right)}{m} \\ &= \frac{\log K}{K} + \frac{1}{K}\log\left(\frac{2t}{d} + \frac{1}{K}\right) \\ &\quad - \frac{\log(2t)}{m} - \frac{\log\left(\frac{1}{K} - \frac{1}{N}\right)}{m}. \end{aligned}$$

The asymptotic code exponent is therefore upper bounded by

$$\frac{\log K}{K} \left(1 + o_{t,d}(1)\right), \tag{5}$$

where

$$o_{t,d}(1) = \frac{\log\left(\frac{2t}{d} + \frac{1}{K}\right)}{\log K} \stackrel{K \to \infty}{\longrightarrow} 0 \tag{6}$$

if $t = o(K)$.

This sphere packing bound can be significantly improved when considering the Euclidean norm. The result is an upper bound with the leading term $(\log K) / (2K)$. The proof is a generalization of the ideas used by Füredi and

Ruszinko in [3]: most points of the form $\mathbf{Cb}$ lie in a ball with radius smaller than $\sqrt{\frac{K}{3}}\,(t+1)$, and therefore the right hand side of the classic sphere packing bound (5) can be reduced by a factor of two.

To proceed, we assign to every $\mathbf{b} \in \mathcal{B}_K$ the probability

$$\frac{1}{|\mathcal{B}_K|} = \frac{1}{\sum_{k=1}^{K}\binom{N}{k}(2t+1)^k}.$$

For a given codeword matrix $\mathbf{C}$, define a random variable

$$\xi = \|\mathbf{Cb}\|_2 .$$

We shall upper bound the probability $\Pr\{\xi \geq \lambda\mu\}$, for arbitrary $\lambda, \mu \in \mathbb{R}^+$, via Markov's inequality

$$\Pr(\xi \geq \lambda\mu) \leq \frac{\mathrm{E}[\xi]}{\lambda\mu} \leq \frac{\sqrt{\mathrm{E}[\xi^2]}}{\lambda\mu}.$$

We calculate $\mathrm{E}\left[\xi^2\right]$ as follows. For a given vector $\mathbf{b}$, let $I \subset [1, N]$ be its support set - i.e., the set of indices for which the entries of $\mathbf{b}$ are nonzero. Let $\mathbf{b}_I$ be the vector composed of the nonzero entries of $\mathbf{b}$. Furthermore, define

$$B_{t,k} = (B_t \backslash \{0\})^k .$$

Then,

$$E\left[\xi^2\right] = \frac{1}{|\mathcal{B}_K|}\sum_{k=1}^{K}\sum_{|I|=k}\sum_{\mathbf{b}_I \in B_{t,k}}\left\|\sum_{j=1}^{k} b_{i_j}\mathbf{v}_{i_j}\right\|_2^2,$$

where $i_j \in I$, $j = 1, \cdots, k$. Note that

$$\begin{aligned}
&\sum_{|I|=k}\sum_{\mathbf{b}_I \in B_{t,k}}\left\|\sum_{j=1}^{k} b_{i_j}\mathbf{v}_{i_j}\right\|_2^2 \\
&= \sum_{|I|=k}\sum_{\mathbf{b}_I \in B_{t,k}}\left(\sum_{j=1}^{k} b_{i_j}^2 + \sum_{1\leq l\neq j\leq k} b_{i_j}b_{i_l}\mathbf{v}_{i_j}^{\dagger}\mathbf{v}_{i_l}\right) \\
&= \underbrace{\sum_{|I|=k}\sum_{\mathbf{b}_I \in B_{t,k}}\sum_{j=1}^{k} b_{i_j}^2}_{(*)} + \underbrace{\sum_{|I|=k}\sum_{\mathbf{b}_I \in B_{t,k}}\sum_{1\leq l\neq j\leq k} b_{i_j}b_{i_l}\mathbf{v}_{i_j}^{\dagger}\mathbf{v}_{i_l}}_{(**)} .
\end{aligned}$$

It is straightforward to evaluate the two sums in the above expression in closed form:

$$\begin{aligned}
(*) &= \binom{N}{k}\sum_{\mathbf{b}_I \in B_{t,k}}\sum_{j=1}^{k} b_{i_j}^2 = \binom{N}{k} k \sum_{\mathbf{b}_I \in B_{t,k}} b_{i_1}^2 \\
&= \binom{N}{k} k\,(2t)^{k-1}\sum_{b_{i_1} \in B_{t,1}} b_{i_1}^2 \\
&= \binom{N}{k} k\,(2t)^{k-1}\,\frac{t\,(t+1)\,(2t+1)}{3};
\end{aligned}$$

and

$$
\begin{aligned}
(**) &= \binom{N}{k} \sum_{1\le l\neq j\le k} \sum_{\mathbf{b}_I \in B_{t,k}} b_{i_l} b_{i_j} \mathbf{v}_{i_l}^{\dagger} \mathbf{v}_{i_j} \\
&= \binom{N}{k} (2t)^{k-2} \sum_{1\le i\neq j\le k} \sum_{b_{i_l}, b_{i_j} \in B_{t,1}} b_{i_l} b_{i_j} \mathbf{v}_{i_l}^{\dagger} \mathbf{v}_{i_j} \\
&= 0,
\end{aligned}
$$

where the last equality follows from the observation that

$$
\begin{aligned}
&\sum_{b_{i_l}\in B_{t,1}, b_{i_j} \in B_{t,1}} b_{i_l} b_{i_j} \mathbf{v}_{i_l}^{\dagger} \mathbf{v}_{i_j} \\
&= \sum_{b_{i_l}>0, b_{i_j} \in B_{t,1}} b_{i_l} b_{i_j} \mathbf{v}_{i_l}^{\dagger} \mathbf{v}_{i_j} \\
&\quad + \sum_{b_{i_l}>0, b_{i_j} \in B_{t,1}} (-b_{i_l}) b_{i_j} \mathbf{v}_{i_l}^{\dagger} \mathbf{v}_{i_j} \\
&= 0.
\end{aligned}
$$

Consequently, one has

$$
\begin{aligned}
&\sum_{|I|=k} \sum_{\mathbf{b}_I \in B_{t,k}} \left\| \sum_{j=1}^{k} b_{i_j} \mathbf{v}_{i_j} \right\|_2^2 \\
&= \binom{N}{k} \frac{(2t)^k \; k \; (t+1)(2t+1)}{6},
\end{aligned}
$$

so that

$$
E\left[\xi^2\right] = \frac{\sum_{k=1}^{K} \binom{N}{k} \frac{(2t)^k \, k \, (t+1)(2t+1)}{6}}{\sum_{k=1}^{K} \binom{N}{k} (2t)^k}.
$$

Next, substitute $\mathrm{E}\left[\xi^2\right]$ into Markov's inequality, with

$$
\mu = \sqrt{\mathrm{E}\left[\xi^2\right]},
$$

so that for any $\lambda > 1$, it holds that

$$
\Pr\left(\xi \ge \lambda\mu\right) \le \frac{1}{\lambda}.
$$

This result implies that at least a $(1-1/\lambda)$-fraction of all possible $\mathbf{Cb}$ vectors lie within an m-dimensional ball of radius $\lambda\mu$ around the origin. As a result, one obtains a sphere packing bound of the form

$$
\left(1 - \frac{1}{\lambda}\right) |\mathcal{B}_K| \le \left(\frac{\lambda\mu + \frac{d}{2}}{\frac{d}{2}} \right)^m .
$$

Note that

$$
\mu^2 = E\left[\xi^2\right] \le \frac{K}{3} (t+1)^2 ,
$$

and that

$$
|\mathcal{B}_K| \ge \left(\frac{N-K}{K} \right)^K (2t)^K .
$$

Consequently, one has

$$\left(1-\frac{1}{\lambda}\right)\left(\frac{N-K}{K}\right)^{K}(2t)^{K}\leq\left(1+\frac{\lambda\sqrt{k}\,(t+1)}{d}\right)^{m},$$

or, equivalently,

$$\frac{\log N}{m}\leq\frac{\log K}{2K}+\frac{1}{K}\log\left(\frac{\lambda\,(t+1)}{d}+\frac{1}{\sqrt{K}}\right)$$
$$-\frac{\log\left(1-\frac{1}{\lambda}\right)}{mK}-\frac{1}{m}\log\left(\frac{1}{K}-\frac{1}{N}\right)-\frac{\log\left(2t\right)}{m}.$$

Without loss of generality, we choose $\lambda=2$. The asymptotic code exponent is therefore upper bounded by

$$\frac{\log K}{2K}\left(1+o_{t,d}\left(1\right)\right),$$

where

$$o_{t,d}\left(1\right)=\frac{2}{\log K}\log\left(\frac{2\left(t+1\right)}{d}+\frac{1}{\sqrt{K}}\right)\stackrel{K\to\infty}{\longrightarrow}0 \tag{7}$$

if $t=o\left(K\right)$. This proves the upper bound of Theorem 1.

VI. Proof of the Lower Bound for WESCs

Similarly as for the case of compressive sensing matrix design, we show that standard Gaussian random matrices, with appropriate scaling, can be used as codebooks of WESCs. Let $\mathbf{H}\in\mathbb{R}^{m\times N}$ be a standard Gaussian random matrix, and let $\mathbf{h}_j$ denote the j^{th} column of $\mathbf{H}$. Let $\mathbf{v}_j=\mathbf{h}_j/\left\|\mathbf{h}_j\right\|_2$ and $\mathbf{C}=\left[\mathbf{v}_1\cdots\mathbf{v}_N\right]$. Then $\mathbf{C}$ is a codebook with unit l_2-norm codewords. Now choose a $\delta>0$ such that $d\left(1+\delta\right)<1$. Let

$$E_1=\bigcup_{j=1}^{N}\left\{\mathbf{H}:\ \frac{1}{\sqrt{m}}\left\|\mathbf{h}_j\right\|_2\in\left(1-\delta,1+\delta\right)\right\} \tag{8}$$

be the event that the normalized l_2-norms of all the columns of $\mathbf{H}$ concentrate around one. Let

$$E_2=\bigcup_{\mathcal{B}_K\ni\mathbf{b}_1\neq\mathbf{b}_2\in\mathcal{B}_K}\left\{\mathbf{H}:\ \left\|\mathbf{C}\left(\mathbf{b}_1-\mathbf{b}_2\right)\right\|_2\geq d\right\}. \tag{9}$$

In other words, E_2 denotes the event that any two different superpositions of codewords lie at Euclidean distance at least d from each other. In the following, we show that for any

$$R<\frac{\log K}{4K}\left(1+o\left(1\right)\right),$$

for which $o\left(1\right)$ is given by Equation (19), if

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}\leq R, \tag{10}$$

then

$$\lim_{(m,N)\to\infty}\Pr\left(E_2\right)=1. \tag{11}$$

This will establish the lower bound of Theorem 1.

Note that

$$\Pr\left(E_2\right)\geq\Pr\left(E_2\bigcap E_1\right)=\Pr\left(E_1\right)-\Pr\left(E_1\bigcap E_2^c\right).$$

According to Theorem 4, stated and proved in the next subsection, one has

$$\lim_{(m,N)\to\infty} \Pr\left(E_1\right) = 1.$$

Thus, the desired relation (11) holds if

$$\lim_{(m,N)\to\infty} \Pr\left(E_1 \bigcap E_2^c\right) = 0.$$

Observe that

$$\mathbf{C}\left(\mathbf{b}_1 - \mathbf{b}_2\right) = \frac{1}{\sqrt{m}}\mathbf{H}\mathbf{b}',$$

where

$$\mathbf{b}' := \mathbf{\Lambda}_{\mathbf{H}}\left(\mathbf{b}_1 - \mathbf{b}_2\right), \tag{12}$$

and

$$\mathbf{\Lambda}_{\mathbf{H}} = \begin{bmatrix} \sqrt{m}/\left\|\mathbf{h}_1\right\|_2 & & \\ & \ddots & \\ & & \sqrt{m}/\left\|\mathbf{h}_N\right\|_2 \end{bmatrix}. \tag{13}$$

By Theorem 19 in Section VI-B, in the asymptotic domain of (10),

$$\begin{aligned} &\Pr\left(E_1 \bigcap \left\{\mathbf{H}:\ \frac{1}{\sqrt{m}}\mathbf{H}\left(\left(1+\delta\right)\mathbf{b}'\right) \leq d\left(1+\delta\right)\right\}\right) \\ &= \Pr\left(E_1 \bigcap \left\{\mathbf{H}:\ \frac{1}{\sqrt{m}}\mathbf{H}\mathbf{b}' \leq d\right\}\right) \\ &= \Pr\left(E_1 \bigcap \left\{\mathbf{H}:\ \mathbf{C}\left(\mathbf{b}_1 - \mathbf{b}_2\right) \leq d\right\}\right) \\ &\to 0. \end{aligned}$$

This establishes the lower bound of Theorem 1.

A. *Column Norms of* $\mathbf{H}$

In this subsection, we quantify the rate regime in which the Euclidean norms of all columns of $\mathbf{H}$, when properly normalized, are concentrated around the value one with high probability.

Theorem 4: Let $\mathbf{H} \in \mathbb{R}^{m\times N}$ be a standard Gaussian random matrix, and $\mathbf{h}_j$ be the j^{th} column of $\mathbf{H}$.

1) For a given $\delta \in (0,1)$,

$$\Pr\left(\left|\frac{1}{m}\left\|\mathbf{h}_j\right\|_2^2 - 1\right| > \delta\right) \leq 2\exp\left(-\frac{m}{4}\delta^2\right)$$

for all $1 \leq j \leq N$.

2) If $m, N \to \infty$ simultaneously, so that

$$\lim_{(m,N)\to\infty} \frac{1}{m}\log N < \frac{\delta^2}{4},$$

then it holds that

$$\lim_{(m,N)\to\infty} \Pr\left(\bigcup_{j=1}^{N}\left\{\left|\frac{1}{m}\left\|\mathbf{h}_j\right\|_2^2 - 1\right| > \delta\right\}\right) = 0.$$

Proof: The first part of this theorem is proved by invoking large deviations techniques. Note that $\|\mathbf{h}_j\|_2^2 = \sum_{i=1}^{m} |H_{i,j}|^2$ is χ^2 distributed. We have

$$\begin{aligned}
&\Pr\left\{\frac{1}{m}\sum_{i=1}^{m}|H_{i,j}|^2 > 1+\delta\right\}\\
&\overset{(a)}{\leq} \exp\left\{-m\left(\alpha(1+\delta) - \log \mathrm{E}\left[e^{\alpha|H_{i,j}|^2}\right]\right)\right\}\\
&= \exp\left\{-m\left(\alpha(1+\delta) + \frac{1}{2}\log(1-2\alpha)\right)\right\}\\
&\overset{(b)}{=} \exp\left\{-\frac{m}{2}\left(\delta - \log(1+\delta)\right)\right\},
\end{aligned} \tag{14}$$

and

$$\begin{aligned}
&\Pr\left\{\frac{1}{m}\sum_{i=1}^{m}|H_{i,j}|^2 < 1-\delta\right\}\\
&\overset{(c)}{\leq} \exp\left\{m\left(\alpha(1-\delta) + \log \mathrm{E}\left[e^{-\alpha|H_{i,j}|^2}\right]\right)\right\}\\
&= \exp\left\{m\left(\alpha(1-\delta) - \frac{1}{2}\log(1+2\alpha)\right)\right\}\\
&\overset{(d)}{=} \exp\left\{-\frac{m}{2}\left(-\log(1-\delta) - \delta\right)\right\},
\end{aligned} \tag{15}$$

where (a) and (c) are obtained by applying Chernoff's inequality [15] and hold for arbitrary $\alpha > 0$, and (b) and (d) are obtained by specializing α to $\frac{1}{2}\frac{\delta}{1+\delta}$ and $\frac{1}{2}\frac{\delta}{1-\delta}$, respectively. By observing that

$$-\log(1-\delta) - \delta > \delta - \log(1+\delta) > 0,$$

we arrive at

$$\begin{aligned}
&\Pr\left\{\left|\frac{1}{m}\|\mathbf{h}_j\|_2^2 - 1\right| > \delta\right\}\\
&\leq 2\exp\left\{-\frac{m}{2}\left(\delta - \log(1+\delta)\right)\right\}\\
&\leq 2\exp\left(-\frac{m}{4}\delta^2\right).
\end{aligned}$$

The second part of the claimed result is proved by applying the union bound, i.e.

$$\begin{aligned}
&\Pr\left\{\bigcup_{j=1}^{N}\left\{\left|\frac{1}{m}\|\mathbf{h}_j\|^2 - 1\right| > \delta\right\}\right\}\\
&\leq 2N\exp\left\{-\frac{m}{4}\delta^2\right\}\\
&= \exp\left\{-m\left(\frac{\delta^2}{4} - \left(\frac{1}{m}\log N + \frac{\log 2}{m}\right)\right)\right\}.
\end{aligned}$$

This completes the proof of Theorem 4. ■

B. The Distance Between Two Different Superpositions

This section is devoted to identifying the rate regime in which any pair of different superpositions is at sufficiently large Euclidean distance. The main result is presented in Theorem 6 at the end of this subsection. Since the proof of this theorem is rather technical, and since it involves complicated notation, we first prove a simplified version of the result, stated in Theorem 5.

Theorem 5: Let $\mathbf{H} \in \mathbb{R}^{m\times N}$ be a standard Gaussian random matrix, and let $\delta \in (0,1)$ be fixed. For sufficiently large K, if

$$\lim_{(m,N)\to\infty} \frac{\log N}{m} < \frac{\log K}{4K}\left(1+o_\delta\left(1\right)\right)$$

where the exact expression for $o_\delta\left(1\right)$ given by Equation (18), then

$$\lim_{(m,N)\to\infty} \Pr\left(\frac{1}{m}\left\|\mathbf{H}\mathbf{b}\right\|_2^2 \le \delta^2\right) = 0, \tag{16}$$

for all $\mathbf{b} \in B_{2t}^N$ such that $\left\|\mathbf{b}\right\|_0 \le 2K$.

Proof: By the union bound, we have

$$\begin{aligned} &\Pr\left(\bigcup_{\left\|\mathbf{b}\right\|_0\le 2K}\left\{\mathbf{H}:\ \frac{1}{m}\left\|\mathbf{H}\mathbf{b}\right\|_2^2\le\delta^2\right\}\right) \\ &= \Pr\left(\bigcup_{k=1}^{2K}\bigcup_{\left\|\mathbf{b}\right\|_0=k}\left\{\mathbf{H}:\ \frac{1}{m}\left\|\mathbf{H}\mathbf{b}\right\|_2^2\le\delta^2\right\}\right) \\ &\le \sum_{k=1}^{2K}\binom{N}{k}(4t)^k \Pr\left(\frac{1}{m}\left\|\mathbf{H}\mathbf{b}\right\|_2^2\le\delta^2,\ \left\|\mathbf{b}\right\|_0=k\right). \end{aligned} \tag{17}$$

We shall upper bound the probability

$$\Pr\left(\frac{1}{m}\left\|\mathbf{H}\mathbf{b}\right\|_2^2\le\delta^2,\ \left\|\mathbf{b}\right\|_0=k\right)$$

for each $k=1,\cdots,2K$. From Chernoff's inequality, for all $\alpha>0$, it holds that

$$\begin{aligned} &\Pr\left(\frac{1}{m}\left\|\mathbf{H}\mathbf{b}\right\|_2^2\le\delta^2,\ \left\|\mathbf{b}\right\|_0=k\right) \\ &\le \exp\left\{m\left(\alpha\delta^2+\log \mathrm{E}\left[e^{-\alpha(\mathbf{H}_{i,\cdot}\mathbf{b})^2}\right]\right)\right\}, \end{aligned}$$

where $\mathbf{H}_{i,\cdot}$ is the i^{th} row of the $\mathbf{H}$ matrix. Furthermore,

$$\begin{aligned} &\mathrm{E}\left[e^{-\alpha(\mathbf{H}_{i,\cdot}\mathbf{b})^2}\right] \\ &= \mathrm{E}\left[\exp\left\{-\alpha\left\|\mathbf{b}\right\|_2^2\left(\mathbf{H}_{i,\cdot}\left(\mathbf{b}/\left\|\mathbf{b}\right\|_2\right)\right)^2\right\}\right] \\ &\overset{(a)}{\le} \mathrm{E}\left[\exp\left\{-\alpha k\left(\mathbf{H}_{i,\cdot}\left(\mathbf{b}/\left\|\mathbf{b}\right\|_2\right)\right)^2\right\}\right] \\ &\overset{(b)}{\le} -\frac{1}{2}\log\left(1+2\alpha k\right), \end{aligned}$$

where (a) follows from the fact that $\left\|\mathbf{b}\right\|_2^2 \ge k$ for all $\mathbf{b}\in B_{2t}^N$ such that $\left\|\mathbf{b}\right\|_0 = k$, and where (b) holds because $\mathbf{H}_{i,\cdot}\left(\mathbf{b}/\left\|\mathbf{b}\right\|_2\right)$ is a standard Gaussian random variable. Let

$$\alpha = \frac{1}{2k}\frac{k-\delta^2}{\delta^2} = \frac{1}{2\delta^2}-\frac{1}{2k}.$$

Then

$$\begin{aligned}&\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_2^2\leq\delta^2,\ \left\|\mathbf{b}\right\|_0=k\right)\\&\leq\exp\left\{m\left(\left(\frac{1}{2}-\frac{\delta^2}{2k}\right)-\frac{1}{2}\log\frac{k}{\delta^2}\right)\right\}\\&=\exp\left\{-\frac{m}{2}\left(\log k-\log\delta^2+\frac{\delta^2}{k}-1\right)\right\}.\end{aligned}$$

Substituting the above expression into the union bound gives

$$\begin{aligned}&\Pr\left(\bigcup_{\left\|\mathbf{b}\right\|_0\leq 2K}\left\{\mathbf{H}:\ \frac{1}{m}\left\|\mathbf{Hb}\right\|_2^2\leq\delta^2\right\}\right)\\&\leq\sum_{k=1}^{2K}\exp\left\{-\frac{m}{2}\left(\log k-\log\delta^2+\frac{\delta^2}{k}-1\right.\right.\\&\qquad\left.\left.-\frac{2k}{m}\log N-\frac{2k}{m}\log\left(4t\right)\right)\right\}\\&\leq\sum_{k=1}^{2K}\exp\left\{-mk\left(\frac{\log k}{2k}+\frac{\delta^2/k-\log\delta^2-1}{2k}\right.\right.\\&\qquad\left.\left.-\frac{1}{m}\log N-\frac{1}{m}\log\left(4t\right)\right)\right\}.\end{aligned}$$

Now, let K be sufficiently large so that

$$\begin{aligned}&\frac{\log 2K}{4K}+\frac{\delta^2/\left(2K\right)-\log\delta^2-1}{4K}\\&=\min_{1\leq k\leq 2K}\left(\frac{\log k}{2k}+\frac{\delta^2/k-\log\delta^2-1}{2k}\right).\end{aligned}$$

If

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}<\frac{\log K}{4K}\left(1+o_\delta\left(1\right)\right),$$

where

$$o_\delta\left(1\right)=\frac{\log 2+\delta^2/\left(2K\right)-\log\delta^2-1}{\log K},\tag{18}$$

then

$$\lim_{(m,N)\to\infty}\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_2^2\leq\delta^2\right)=0$$

for all $\mathbf{b}\in B_{2t}^N$ such that $\left\|\mathbf{b}\right\|_0\leq 2K$. This completes the proof of the claimed result. ■

Based on Theorem 5, the asymptotic region in which $\Pr\left(E_2^c\bigcap E_1\right)\to 0$ is characterized in below.

Theorem 6: Let $\mathbf{H}\in\mathbb{R}^{m\times N}$ be a standard Gaussian random matrix, and for a given $\mathbf{H}$, let $\mathbf{\Lambda_H}$ be as defined in (13). For a given $d\in(0,1)$, choose a $\delta>0$ such that $d\left(1+\delta\right)<1$. Define the set E_1 as in (8). For sufficiently large K, if

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}<\frac{\log K}{4K}\left(1+o\left(1\right)\right)$$

where

$$o\left(1\right)=\frac{\log 2-1}{\log K},\tag{19}$$

then

$$\lim_{(m,N)\to\infty} \Pr\left(\frac{1}{m}\left\|\mathbf{H}\mathbf{\Lambda}_{\mathbf{H}}\left(\mathbf{b}_1-\mathbf{b}_2\right)\right\|_2^2 \le d^2,\ E_1\right)=0, \tag{20}$$

for all pairs of $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}_K$ such that $\mathbf{b}_1 \neq \mathbf{b}_2$.

Proof: The proof is analogous to that of Theorem 4 with minor changes. Let $\mathbf{b}' = \mathbf{\Lambda}_{\mathbf{H}}\left(\mathbf{b}_1-\mathbf{b}_2\right)$. On the set E_1, since

$$\frac{1}{\sqrt{m}}\left\|\mathbf{h}_j\right\|_2 \le 1+\delta$$

for all $1 \le j \le N$, the nonzero entries of $(1+\delta)\,\mathbf{b}'$ satisfy

$$\left|(1+\delta)\,b_i'\right| \ge 1.$$

Replace $\mathbf{b}$ in Theorem 5 with $(1+\delta)\,\mathbf{b}'$. All the arguments in the proof of Theorem 5 are still valid, except that the higher order term is changed to

$$\begin{aligned}
& o_{d(1+\delta)}(1) \\
&= \frac{\log 2 + d^2(1+\delta)^2/(2K) - 1 - \log\left(d^2(1+\delta)^2\right)}{\log K} \\
&\ge \frac{\log 2 - 1}{\log K}.
\end{aligned}$$

This completes the proof of the theorem. ■

VII. Proof of the Lower Bound for l_1-WSCs

The proof is similar to that of the lower bound for WESCs. Let $\mathbf{A} \in \mathbb{R}^{m\times N}$ be a standard Gaussian random matrix, and let $\mathbf{H}$ be the matrix with entries

$$H_{i,j} = \frac{\sqrt{2\pi}}{2}A_{i,j},\ 1 \le i \le m,\ 1 \le j \le N.$$

Once more, let $\mathbf{h}_j$ be the j^{th} column of $\mathbf{H}$. Let $\mathbf{v}_j = \mathbf{h}_j / \left\|\mathbf{h}_j\right\|_1$ and $\mathbf{C} = [\mathbf{v}_1 \cdots \mathbf{v}_N]$. Then $\mathbf{C}$ is a codebook with unit l_1-norm codewords. Now choose a $\delta > 0$ such that $d(1+\delta) < 1$. Let

$$E_1 = \bigcup_{j=1}^{N}\left\{\mathbf{H}:\ \frac{1}{m}\left\|\mathbf{h}_j\right\|_1 \in (1-\delta, 1+\delta)\right\}, \tag{21}$$

and

$$E_2 = \bigcup_{\mathcal{B}_K \ni \mathbf{b}_1 \neq \mathbf{b}_2 \in \mathcal{B}_K}\left\{\mathbf{H}:\ \left\|\mathbf{C}\left(\mathbf{b}_1-\mathbf{b}_2\right)\right\|_1 \ge d\right\}. \tag{22}$$

We consider the asymptotic regime where

$$\lim_{(m,N)\to\infty}\frac{\log N}{m} \le R,$$

$$R < \frac{\log K}{4K}\left(1+o(1)\right),$$

and $o_d(1)$ is given in Equation (31). Theorem 7 in Section VII-A suggests that

$$\lim_{(m,N)\to\infty}\Pr\left(E_1\right) = 1,$$

while Theorem 8 in Section VII-B shows that

$$\lim_{(m,N)\to\infty} \Pr\left(E_1 \bigcap E_2^c\right) = 0.$$

Therefore,

$$\lim_{(m,N)\to\infty} \Pr\left(E_2\right) \geq \lim_{(m,N)\to\infty} \Pr\left(E_1\right) - \Pr\left(E_1 \bigcap E_2^c\right) = 1.$$

This result implies the lower bound of Theorem 2.

A. *Column Norms of* $\mathbf{H}$

The following theorem quantifies the rate regime in which the l_1-norms of all columns of $\mathbf{H}$, with proper normalization, are concentrated around one with high probability.

Theorem 7: Let $\mathbf{A} \in \mathbb{R}^{m\times N}$ be a standard Gaussian random matrix. Let $\mathbf{H}$ be the matrix with entries

$$H_{i,j} = \frac{\sqrt{2\pi}}{2} A_{i,j},\ 1 \leq i \leq m,\ 1 \leq j \leq N.$$

Let $\mathbf{h}_j$ be the j^{th} column of $\mathbf{H}$.

1) For a given $\delta \in (0,1)$,

$$\Pr\left(\left|\frac{1}{m}\left\|\mathbf{h}_j\right\|_1 - 1\right| > \delta\right) \leq c_1 e^{-mc_2\delta^2}$$

for some positive constant c_1 and c_2;

2) Let $m, N \to \infty$ simultaneously, with

$$\lim_{(m,N)\to\infty} \frac{1}{m} \log N < c_2\delta^2.$$

The it holds that

$$\lim_{(m,N)\to\infty} \Pr\left(\bigcup_{j=1}^{N}\left\{\left|\frac{1}{m}\left\|\mathbf{h}_j\right\|_1 - 1\right| > \delta\right\}\right) = 0.$$

Proof:

1) Since $A_{i,j}$ is a standard Gaussian random variable, $|A_{i,j}|$ is a Subgaussian distributed random variable, and $\mathrm{E}\left[|A_{i,j}|\right] = \frac{2}{\sqrt{2\pi}}$. According to Proposition 1 in Appendix A, $|A_{i,j}| - \frac{2}{\sqrt{2\pi}}$ is a Subgaussian random variable with zero mean. A direct application of Theorem 12 stated in Appendix A gives

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{m}\left\|\mathbf{h}_j\right\|_1 - 1\right| > \delta\right) \\
&= \Pr\left(\left|\sum_{i=1}^{m}\left(|A_{i,j}| - \frac{2}{\sqrt{2\pi}}\right)\right| > \frac{2m\delta}{\sqrt{2\pi}}\right) \\
&\leq c_1 \exp\left(-c_2 m\delta^2\right),
\end{aligned}$$

which proves claim 1).

2) This part is proved by using the union bound: first, note that

$$\begin{aligned}&\Pr\left(\bigcup_{j=1}^{N}\left\{\left|\frac{1}{m}\left\|\mathbf{h}_j\right\|_1-1\right|>\delta\right\}\right)\\&\leq\exp\left(-mc_2\delta^2+\log c_1+\log N\right)\\&=\exp\left\{-m\left(c_2\delta^2-\frac{1}{m}\log c_1-\frac{1}{m}\log N\right)\right\}.\end{aligned}$$

If

$$\lim_{(m,N)\to\infty}\frac{1}{m}\log N<c_2\delta^2,$$

then one has

$$\lim_{(m,N)\to\infty}\Pr\left(\bigcup_{j=1}^{N}\left\{\left|\frac{1}{m}\left\|\mathbf{h}_j\right\|_1-1\right|>\delta\right\}\right)=0.$$

This completes the proof of claim 2).

■

B. The Distance Between Two Different Superpositions

Similarly to the analysis performed for WESCs, we start with a proof of a simplified version of the result needed in order to simplify tedious notation. We then explain how to establish the proof of Theorem 9 by modifying some of the steps of the simplified theorem.

Theorem 8: Let $\mathbf{A}\in\mathbb{R}^{m\times N}$ be a standard Gaussian random matrix. Let $\mathbf{H}$ be the matrix with entries

$$H_{i,j}=\frac{\sqrt{2\pi}}{2}A_{i,j},\ 1\leq i\leq m,\ 1\leq j\leq N.$$

Let $\delta\in(0,1)$ be given. For sufficiently large K, if

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}<\frac{\log K}{4K}\left(1+o_\delta\left(1\right)\right),$$

where

$$o_\delta\left(1\right)=\frac{2}{\log K}\left(\log\frac{\pi}{2\delta}-1\right),\tag{23}$$

then

$$\lim_{(m,N)\to\infty}\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_1\leq\delta\right)=0\tag{24}$$

for all $\mathbf{b}\in B_{2t}^N$ such that $\left\|\mathbf{b}\right\|_0\leq 2K$.

Proof: The proof starts by using the union bound, as

$$\begin{aligned}&\Pr\left(\bigcup_{\left\|\mathbf{b}\right\|_0\leq 2K}\left\{\mathbf{H}:\ \frac{1}{m}\left\|\mathbf{Hb}\right\|_1\leq\delta\right\}\right)\\&\leq\sum_{k=1}^{2K}\binom{N}{k}\left(4t\right)^k\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_1\leq\delta,\ \left\|\mathbf{b}\right\|_0=k\right).\end{aligned}\tag{25}$$

To estimate the above upper bound, we have to upper bound the probability

$$\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_1\leq\delta,\ \left\|\mathbf{b}\right\|_0=k\right),$$

for each $k = 1, \cdots, 2K$. Let us derive next an expression for such an upper bound that holds for arbitrary values of $k \geq 1$.

Note that

$$\begin{aligned}
&\mathrm{E}\left[e^{-\alpha\left|\sum_{j=1}^{k} b_j A_{i,j}\right|}\right] = \int_0^\infty \frac{2}{\sqrt{2\pi}\,\|\mathbf{b}\|_2} e^{-\frac{x^2}{2\|\mathbf{b}\|_2^2}} e^{-\alpha x}\cdot dx \\
&\overset{(a)}{=} \int_0^\infty \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{-\alpha\|\mathbf{b}\|_2 x}\cdot dx \\
&= e^{\frac{\alpha^2\|\mathbf{b}\|_2^2}{2}} \int_0^\infty \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{\left(x+\alpha\|\mathbf{b}\|_2\right)^2}{2}\right)\cdot dx \\
&\overset{(b)}{=} e^{\frac{\alpha^2\|\mathbf{b}\|_2^2}{2}} \int_{\alpha\|\mathbf{b}\|_2}^\infty \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\cdot dx \\
&\leq e^{\frac{\alpha^2\|\mathbf{b}\|_2^2}{2}} \int_{\alpha\|\mathbf{b}\|_2}^\infty \frac{x}{\alpha\|\mathbf{b}\|_2}\cdot\frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\cdot dx \\
&= \frac{1}{\alpha\|\mathbf{b}\|_2}\cdot\frac{2}{\sqrt{2\pi}} e^{\frac{\alpha^2\|\mathbf{b}\|_2^2}{2}} e^{-\frac{\alpha^2\|\mathbf{b}\|_2^2}{2}} \\
&\overset{(c)}{\leq} \frac{1}{\alpha}\frac{2}{\sqrt{2\pi k}},
\end{aligned} \tag{26}$$

where (a) and (b) follow from the change of variables $x' = x/\|\mathbf{b}\|_2$ and $x' = x + \alpha\|\mathbf{b}\|_2$, respectively. Inequality (c) holds based on the assumption that $\|\mathbf{b}\|_2 \geq k$. As a result,

$$\begin{aligned}
&\Pr\left(\frac{1}{m}\sum_{i=1}^{m}\left|\sum_j b_j H_{i,j}\right| \leq \delta\right) \\
&= \Pr\left(\frac{1}{m}\sum_{i=1}^{m}\left|\sum_{j=1}^{k} b_j A_{i,j}\right| \leq \frac{2}{\sqrt{2\pi}}\delta\right) \\
&\leq \exp\left\{m\left(\alpha\frac{2\delta}{\sqrt{2\pi}} + \log\mathrm{E}\left[e^{-\alpha\left|\sum_j b_j H_j\right|}\right]\right)\right\} \\
&\leq \exp\left\{m\left(\alpha\frac{2\delta}{\sqrt{2\pi}} + \log\left(\frac{2}{\sqrt{2\pi k}}\frac{1}{\alpha}\right)\right)\right\} \\
&= \exp\left\{m\left(\alpha\frac{2\delta}{\sqrt{2\pi}} - \log\left(\alpha\sqrt{\frac{\pi k}{2}}\right)\right)\right\} \\
&= \exp\left\{m\left(1 - \log\left(\sqrt{k}\frac{\pi}{2\delta}\right)\right)\right\},
\end{aligned} \tag{27}$$

where the last equality is obtained by specializing $\alpha = \sqrt{2\,\pi}/2\,\delta$.

The upper bound in (27) is useful only when it is less than one, or equivalently,

$$\log\left(\sqrt{k}\frac{\pi}{2\delta}\right) > 1. \tag{28}$$

For any $\delta \in (0,1)$, if $k \geq 4$, inequality (28) holds. Thus, for any $k \geq 4$,

$$\begin{aligned}
&\binom{N}{k}(4t)^k \Pr\left(\frac{1}{m}\|\mathbf{Hb}\|_1 \leq \delta,\ \|\mathbf{b}\|_0 = k\right) \\
&\leq \exp\left\{-mk\left(\frac{\log k}{2k}\left(1+o_\delta(1)\right) - \frac{\log(4t)}{m} - \frac{\log N}{m}\right)\right\} \\
&\to 0,
\end{aligned} \tag{29}$$

as $(m,N)\to\infty$ with

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}<\frac{\log k}{2k}\left(1+o_{\delta}\left(1\right)\right),$$

where

$$o_{\delta}\left(1\right)=\frac{2}{\log k}\left(\log\frac{\pi}{2\delta}-1\right).$$

Another upper bound is needed for $k=1,2,3$. For a fixed k taking one of these values,

$$\begin{aligned}&\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_1\leq\delta\right)\\&=\Pr\left(\frac{1}{m}\sum_i\left|\sum_j A_{i,j}b_j\right|<\frac{2}{\sqrt{2\pi}}\delta\right)\\&=\Pr\left(\sum_i\left(\left|\sum_j A_{i,j}b_j\right|-\frac{2\left\|\mathbf{b}\right\|_2}{\sqrt{2\pi}}\right)<\frac{2m}{\sqrt{2\pi}}\left(\delta-\left\|\mathbf{b}\right\|_2\right)\right).\end{aligned}$$

It is straightforward to verify that $\sum_j A_{i,j}b_j$ is Gaussian and that

$$\mathrm{E}\left[\left|\sum_j A_{i,j}b_j\right|\right]=\frac{2\left\|\mathbf{b}\right\|_2}{\sqrt{2\pi}}.$$

Thus

$$\sum_i\left(\left|\sum_j A_{i,j}b_j\right|-\frac{2\left\|\mathbf{b}\right\|_2}{\sqrt{2\pi}}\right)$$

is a sum of independent zero-mean subgaussian random variables. Furthermore, $\left\|\mathbf{b}\right\|_2\in\left[\sqrt{k},\sqrt{2kt}\right]$ and therefore, $\delta-\left\|\mathbf{b}\right\|_2<0$. Hence, we can apply Theorem 12 of Appendix A: as a result, there exist positive constants $c_{3,k}$ and $c_{4,k}$ such that

$$\begin{aligned}&\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_1\leq\delta\right)\\&\leq c_{3,k}\exp\left(-c_{4,k}m\left(\delta-\left\|\mathbf{b}\right\|_2\right)^2\right)\\&\leq c_{3,k}\exp\left(-c_{4,k}m\left(\sqrt{k}-\delta\right)^2\right).\end{aligned}$$

Note that the values of $c_{3,k}$ and $c_{4,k}$ depend on k. Consequently,

$$\begin{aligned}&\binom{N}{k}\left(4t\right)^k\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_1\leq\delta,\ \left\|\mathbf{b}\right\|_0=k\right)\\&\leq c_{3,k}\exp\left\{-mk\left(c_{4,k}\left(1-\frac{\delta}{\sqrt{k}}\right)^2\right.\right.\\&\qquad\qquad\left.\left.-\frac{\log\left(4t\right)}{m}-\frac{\log N}{m}\right)\right\}\\&\to 0\end{aligned}\tag{30}$$

as $m,N\to\infty$ with

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}<c_{4,k}\left(1-\frac{\delta}{\sqrt{k}}\right)^2.$$

Finally, substitute the upper bounds of (29) and (30) into the union bound of Equation (25). If K is large enough so that

$$\frac{\log K}{4K}\left(1+o_{\delta}\left(1\right)\right)<c_{4,k}\left(1-\frac{\delta}{\sqrt{k}}\right)^{2}\text{ for all }k=1,2,3,$$

and if

$$\frac{\log K}{4K}\left(1+o_{\delta}\left(1\right)\right)\leq\min_{4\leq k\leq 2K}\frac{\log k}{2k}\left(1+o_{\delta}\left(1\right)\right),$$

where $o_{\delta}\left(1\right)$ is as given in (23), then the desired result (24) holds. ■

Based on Theorem 8, we are ready to characterize the asymptotic region in which $\Pr\left(E_2^c\bigcap E_1\right)\rightarrow 0$.

Theorem 9: Define $\mathbf{A}$ and $\mathbf{H}$ as in Theorem 9. For a given $\mathbf{H}$, define the diagonal matrix

$$\mathbf{\Lambda_H}=\begin{bmatrix} m/\left\|\mathbf{h}_1\right\|_1 & & 0 \\ & \ddots & \\ 0 & & m/\left\|\mathbf{h}_N\right\|_1 \end{bmatrix}.$$

For a given $d\in(0,1)$, choose a $\delta>0$ such that $d\left(1+\delta\right)<1$. Define the set E_1 as in (21). For sufficiently large K, if

$$\lim_{(m,N)\rightarrow\infty}\frac{\log N}{m}<\frac{\log K}{4K}\left(1+o\left(1\right)\right),$$

where

$$o\left(1\right)=\frac{2}{\log K}\left(\log\pi-1-\log 2\right),\tag{31}$$

then

$$\lim_{(m,N)\rightarrow\infty}\Pr\left(\frac{1}{m}\left\|\mathbf{H\Lambda_H}\left(\mathbf{b}_1-\mathbf{b}_2\right)\right\|_1\leq d,\ E_1\right)=0$$

for all pairs of $\mathbf{b}_1,\mathbf{b}_2\in\mathcal{B}_K$ such that $\mathbf{b}_1\neq\mathbf{b}_2$.

Proof: Let $\mathbf{b}'=\mathbf{\Lambda_H}\left(\mathbf{b}_1-\mathbf{b}_2\right)$. On the set E_1, since

$$\frac{1}{\sqrt{m}}\left\|\mathbf{h}_j\right\|_1\leq 1+\delta,$$

for all $1\leq j\leq N$, all the nonzero entries of $\left(1+\delta\right)\mathbf{b}'$ satisfy

$$\left|\left(1+\delta\right)b_i'\right|\geq 1.$$

Replace $\mathbf{b}$ in Theorem 8 with $\left(1+\delta\right)\mathbf{b}'$. All arguments used in the proof of Theorem 8 are still valid, except that now, the higher order term (23) in the asymptotic expression reads as

$$\begin{aligned}&o_{d(1+\delta)}\left(1\right)\\&=\frac{2}{\log K}\left(\log\pi-1-\log 2-\log\left(d\left(1+\delta\right)\right)\right)\\&\geq\frac{2}{\log K}\left(\log\pi-1-\log 2\right).\end{aligned}$$

This completes the proof. ■

VIII. Proof of the Lower Bound for Nonnegative l_1-WSCs

The proof follows along the same lines as the one described for l_1-WSCs. However, there is a serious technical difficulty associated with the analysis of nonnegative l_1-WSCs. Let $\mathbf{A}\in\mathbb{R}^{m\times N}$ be a standard Gaussian random matrix. For general l_1-WSCs, we let

$$H_{i,j} = \frac{\sqrt{2\pi}}{2} A_{i,j},\ 1 \le i \le m,\ 1 \le j \le N,$$

and therefore,

$$\sum_{j=1}^{N} H_{i,j} b_j$$

is a Gaussian random variable, whose parameters are easy to determine. However, for nonnegative l_1-WSCs, one has to set

$$H_{i,j} = \frac{\sqrt{2\pi}}{2} \left| A_{i,j} \right|,\ 1 \le i \le m,\ 1 \le j \le N. \tag{32}$$

Since the random variables $H_{i,j}$s are not Gaussian, but rather one-sided Gaussian,

$$\sum_{j=1}^{N} H_{i,j} b_j$$

is not Gaussian distributed, and it is complicated to exactly characterize its properties.

Nevertheless, we can still define E_1 and E_2 as in Equations (21) and (22). The results of Theorem 7 are still valid under the non-negativity assumption: the norms of all $\mathbf{H}$ columns concentrate around one in the asymptotic regime described in Theorem 7. The key step in the proof of the lower bound is to identify the asymptotic region in which any two different superpositions are sufficiently separated in terms of the l_1-distance. We therefore use an approach similar to the one we invoked twice before: we first prove a simplified version of the claim, and then proceed with proving the needed result by introducing some auxiliary variables and notation.

Theorem 10: Let $\mathbf{A} \in \mathbb{R}^{m \times N}$ be a standard Gaussian random matrix. Let $\mathbf{H}$ be the matrix with entries

$$H_{i,j} = \frac{\sqrt{2\pi}}{2} \left| A_{i,j} \right|,\ 1 \le i \le m,\ 1 \le j \le N.$$

Let $\delta \in (0,1)$ be given. For a given sufficiently large K, if

$$\lim_{(m,N)\to\infty} \frac{\log N}{m} < \frac{\log K}{4K} \left(1 + o_t\left(1\right)\right)$$

where $o_t\left(1\right)$ is given in (39), then

$$\lim_{(m,N)\to\infty} \Pr\left(\frac{1}{m} \left\| \mathbf{H}\mathbf{b} \right\|_1 \le \delta \right) = 0 \tag{33}$$

for all $\mathbf{b} \in B_{2t}^N$ and $\left\| \mathbf{b} \right\|_0 \le 2K$.

Proof: Similarly as for the corresponding proof for general l_1-WSCs, we need a tight upper bound on the moment generation function of the random variable

$$\left| \sum_{j=1}^{k} b_j \left| A_{i,j} \right| \right|.$$

For this purpose, we resort to the use of the Central Limit Theorem. We first approximate the distribution of $\sum_{j=1}^{k} b_j \left| A_{i,j} \right|$ by a Gaussian distribution. Then, we uniformly upper bound the approximation error according to the Berry-Esseen Theorem (see [14] and Appendix B for an overview of this theory). Based on this approximation, we obtain an upper bound on the moment generating function, with leading term $\left(\log k\right)/\sqrt{k}$ (see Equation (38) for details).

To simplify the notation, for a $\mathbf{b} \in B_{2t}^N$ with $\|\mathbf{b}_0\| = k$, let

$$Y_{\mathbf{b},k} = \sum_{j=1}^{N} \frac{\sqrt{2\pi}}{2} |A_j| \, b_j,$$

where A_js are standard Gaussian random variables. Then,

$$\begin{aligned} &\Pr\left(\frac{1}{m} \|\mathbf{Hb}\|_1 \leq \delta, \ \|\mathbf{b}\|_0 = k\right) \\ &\leq \exp\left\{m\left(\alpha\delta + \log \mathrm{E}\left[e^{-\alpha|Y_{\mathbf{b},k}|}\right]\right)\right\}, \end{aligned}$$

where the inequality holds for all $\alpha > 0$. Now, we fix α and upper bound the moment generating function as follows. Note that

$$\begin{aligned} &\mathrm{E}\left[e^{-\alpha|Y_{\mathbf{b},k}|}\right] \\ &= \mathrm{E}\left[e^{-\alpha|Y_{\mathbf{b},k}|}, \ |Y_{\mathbf{b},k}| \geq \frac{1}{\alpha} \log \sqrt{k}\right] \qquad (34) \\ &\quad + \mathrm{E}\left[e^{-\alpha|Y_{\mathbf{b},k}|}, \ |Y_{\mathbf{b},k}| < \frac{1}{\alpha} \log \sqrt{k}\right]. \qquad (35) \end{aligned}$$

The first term (34) is upper bounded by

$$\begin{aligned} &\mathrm{E}\left[e^{-\alpha\frac{1}{\alpha}\log\sqrt{k}}, \ |Y_{\mathbf{b},k}| \geq \frac{1}{\alpha} \log \sqrt{k}\right] \\ &\leq \mathrm{E}\left[\frac{1}{\sqrt{k}}, \ |Y_{\mathbf{b},k}| \geq \frac{1}{\alpha} \log \sqrt{k}\right] \\ &\leq \frac{1}{\sqrt{k}} \Pr\left(|Y_{\mathbf{b},k}| \geq \frac{1}{\alpha} \log \sqrt{k}\right) \\ &\leq \frac{1}{\sqrt{k}}. \qquad (36) \end{aligned}$$

In order to upper bound the second term in Equation (35), we apply Lemma 2 from the Appendix, proved using the Central Limit Theorem and the Berry-Esseen result:

$$\begin{aligned} &\mathrm{E}\left[1, \ |Y_{\mathbf{b},k}| < \frac{1}{\alpha} \log \sqrt{k}\right] \\ &= \Pr\left(|Y_{\mathbf{b},k}| < \frac{1}{\alpha} \log \sqrt{k}\right) \\ &= \Pr\left(\left|\sum_{j=1}^{k} b_j |A_j|\right| < \frac{2}{\sqrt{2\pi}} \frac{1}{\alpha} \log \sqrt{k}\right) \\ &\leq \frac{2}{\sqrt{2\pi}} \frac{1}{\alpha\pi} \frac{\log \sqrt{k}}{\sqrt{k}} + 12 \frac{t^3}{\sqrt{k}} \mathrm{E}\left[|A|^3\right] \\ &= \frac{1}{\sqrt{k}} \frac{1}{\sqrt{2\pi}} \left(\frac{\log k}{\alpha\pi} + 48t^3\right). \qquad (37) \end{aligned}$$

Combining the upper bounds in (36) and (37) shows that

$$\begin{aligned} \mathrm{E}\left[e^{-\alpha|Y_{\mathbf{b},k}|}\right] &\leq \frac{1}{\sqrt{k}} \left(1 + \frac{1}{\sqrt{2\pi}} \left(\frac{\log k}{\alpha\pi} + 48t^3\right)\right) \\ &\leq \frac{1}{\sqrt{k}} \left(1 + \frac{\log k}{4\alpha} + 24t^3\right). \qquad (38) \end{aligned}$$

Next, set $\alpha = 1/\delta$. Then

$$\begin{aligned}&\Pr\left(\frac{1}{m}\left\|\mathbf{Hb}\right\|_1 \leq \delta,\ \left\|\mathbf{b}\right\|_0 = k\right)\\&\leq \exp\left\{-m\left(\frac{1}{2}\log k\right)\left(1+o_{t,k}\left(1\right)\right)\right\},\end{aligned}$$

where

$$o_{t,k}\left(1\right) = -\frac{2+2\log\left(1+\frac{\log k}{4}+24t^3\right)}{\log k}.$$

Now we choose a $k_0 \in \mathbb{Z}^+$ such that for all $k \geq k_0$,

$$\frac{\log k}{2k}\left(1+o_{t,k}\left(1\right)\right) > 0.$$

It is straightforward to verify that k_0 is well defined. Consider the case when $1 \leq k \leq k_0$. It can be verified that

$$\sum_{j=1}^{k} b_j H_{i,j} = \frac{\sqrt{2\pi}}{2}\sum_{j=1}^{k} b_j \left|A_{i,j}\right|$$

is Subgaussian and that

$$\mathrm{E}\left[\left|\sum_{j=1}^{k} b_j H_{i,j}\right|\right] \geq 1$$

for all $\mathbf{b} \in B_{2t}^N$ such that $\left\|\mathbf{b}\right\|_0 = k$. By applying the large deviations result for Subgaussian random variables, as stated in Theorem 12, and the union bound, it can be proved that there exists a $c_k > 0$ such that

$$\begin{aligned}&\Pr\left(\bigcup_{\left\|\mathbf{b}\right\|_0 = k}\left\{\mathbf{H}:\ \frac{1}{m}\left\|\mathbf{Hb}\right\|_1 \leq \delta\right\}\right)\\&\leq \exp\left\{-mk\left(c_k - \frac{\log\left(4t\right)}{m} - \frac{\log N}{m}\right)\right\}\\&\to 0.\end{aligned}$$

The above result holds whenever $m, N \to \infty$ simultaneously, with

$$\lim_{(m,N)\to\infty}\frac{\log N}{m} < c_k.$$

Finally, let K be sufficiently large so that

$$\frac{\log K}{4K}\left(1+o_{t,2K}\left(1\right)\right) \leq \min_{k_0 \leq k \leq 2K}\frac{\log k}{2k}\left(1+o_{t,k}\left(1\right)\right),$$

and

$$\frac{\log K}{4K}\left(1+o_{t,2K}\left(1\right)\right) \leq \min_{1 \leq k \leq k_0} c_k.$$

Then

$$\Pr\left(\bigcup_{\|\mathbf{b}\|_0\le 2K}\left\{\mathbf{H}:\ \frac{1}{m}\|\mathbf{H}\mathbf{b}\|_1\le\delta\right\}\right)$$
$$\le\sum_{k=1}^{2K}\binom{N}{k}(4t)^k\Pr\left(\frac{1}{m}\|\mathbf{H}\mathbf{b}\|_1\le\delta,\ \|\mathbf{b}\|_0=k\right)$$
$$\le\sum_{k=1}^{k_0}\exp\left\{-mk\left(c_k-\frac{\log(4t)}{m}-\frac{\log N}{m}\right)\right\}$$
$$+\sum_{k=k_0+1}^{2K}\exp\left\{-mk\left(\frac{\log k}{2k}\left(1+o_{t,k}(1)\right)\right.\right.$$
$$\left.\left.-\frac{\log(4t)}{m}-\frac{\log N}{m}\right)\right\}$$
$$\to 0,$$

as $m,N\to\infty$ with

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}<\frac{\log K}{4K}\left(1+o_t(1)\right),$$

where

$$o_t(1)=-\frac{2+2\log\left(1+\frac{\log 2K}{4}+24t^3\right)}{\log(2K)}. \tag{39}$$

■

Based on Theorem 10, we can characterize the rate region in which any two distinct superpositions are sufficiently separated in the l_1 space.

Theorem 11: Define $\mathbf{A}$ and $\mathbf{H}$ as in Theorem 10. For a given $\mathbf{H}$, define the diagonal matrix

$$\mathbf{\Lambda_H}=\begin{bmatrix} m/\|\mathbf{h}_1\|_1 & & 0\\ & \ddots & \\ 0 & & m/\|\mathbf{h}_N\|_1\end{bmatrix}.$$

Also, for $d\in(0,1)$, choose a $\delta\in\left(0,\frac{1}{2}\right)$ such that $d(1+\delta)<1$. Define the set E_1 as in (21). Provided that K is sufficiently large, if

$$\lim_{(m,N)\to\infty}\frac{\log N}{m}<\frac{\log K}{4K}\left(1+o_t(1)\right)$$

where

$$o_t(1)=-\frac{2+2\log\left(1+\frac{\log 2K}{4}+648t^3\right)}{\log(2K)}, \tag{40}$$

then it holds

$$\lim_{(m,N)\to\infty}\Pr\left(\frac{1}{m}\left\|\mathbf{H}\mathbf{\Lambda_H}\left(\mathbf{b}_1-\mathbf{b}_2\right)\right\|_1\le d,\ E_1\right)=0,$$

for all pairs of $\mathbf{b}_1,\mathbf{b}_2\in\mathcal{B}_K$ such that $\mathbf{b}_1\ne\mathbf{b}_2$.

Proof: The proof is very similar to that of Theorem 10. The only difference is the following. Let $\mathbf{b}'=\mathbf{\Lambda_H}\left(\mathbf{b}_1-\mathbf{b}_2\right)$. Since

$$\frac{1}{2}\le 1-\delta\le\frac{1}{m}\|\mathbf{h}_j\|_1\le 1+\delta\le\frac{3}{2},$$

all the nonzero entries of $(1+\delta)\,\mathbf{b}'$ on the set E_1 satisfy the following inequality

$$1 \leq \left|(1+\delta)\,b_i'\right| \leq 3t.$$

As a result, we have the higher order term $o_t(1)$ as given in Equation (40). ■

IX. Conclusions

We introduced a new family of codes over the reals, termed weighted superimposed codes. Weighted superimposed codes can be applied to all problems in which one seeks to robustly distinguish between bounded integer valued linear combinations of codewords that obey predefined norm and sign constraints. As such, they can be seen as a special instant of compressed sensing schemes in which the sparse sensing vectors contain entries from a symmetric, bounded set of integers. We characterized the achievable rate regions of three classes of weighted superimposed codes, for which the codewords obey l_2, l_1, and non-negativity constraints.

Appendix

A. Subgaussian Random Variables

Definition 4 (The Subgaussian and Subexponential distributions): A random variable X is said to be Subgaussian if there exist positive constants c_1 and c_2 such that

$$\Pr\left(|X| > x\right) \leq c_1 e^{-c_2 x^2} \quad \forall x > 0.$$

It is Subexponential if there exist positive constants c_1 and c_2 such that

$$\Pr\left(|X| > x\right) \leq c_1 e^{-c_2 x} \quad \forall x > 0.$$

Lemma 1 (Moment Generating Function): Let X be a zero-mean random variable. Then, the following two statements are equivalent.

1) X is Subgaussian.
2) $\exists c$ such that $\mathrm{E}\left[e^{\alpha X}\right] \leq e^{c\alpha^2}$, $\forall \alpha \geq 0$.

Theorem 12: Let $X_1, \cdots, X_n$ be independent Subgaussian random variables with zero mean. For any given $a_1, \cdots, a_n \in \mathbb{R}$, $\sum_k a_k X_k$ is a Subgaussian random variable. Furthermore, there exist positive constants c_1 and c_2 such that

$$\Pr\left(\left|\sum_k a_k X_k\right| > x\right) \leq c_1 e^{-c_2 x^2/\|\mathbf{a}\|_2^2}, \quad \forall x > 0,$$

where $\|\mathbf{a}\|_2^2 = \sum_k a_k^2$.

Proof: See [16, Lecture 5, Theorem 5 and Corollary 6]. ■

We prove next a result that asserts that translating a Subgaussian random variable produces another Subgaussian random variable.

Proposition 1: Let X be a Subgaussian random variable. For any given $a \in \mathbb{R}$, $Y = X + a$ is a Subgaussian random variable as well.

Proof: It can be verified that for any $y \in \mathbb{R}$,

$$(y-a)^2 \leq \frac{1}{2}y^2 - a^2,$$

and

$$(y+a)^2 \leq \frac{1}{2}y^2 - a^2.$$

Now for $y > |a|$,

$$\begin{aligned}\Pr\left(|Y| > y\right) &= \Pr\left(X + a > y\right) + \Pr\left(X + a < -y\right) \\ &\leq \Pr\left(X > y - a\right) + \Pr\left(X < -y - a\right). \end{aligned} \tag{41}$$

When $a > 0$,

$$\begin{aligned}(41) &\leq \Pr\left(|X| > y - a\right) \\ &\leq c_1 e^{-c_2(y-a)^2} \\ &\leq c_1 c^{c_2 a^2} e^{-c_2 y^2/2}. \end{aligned} \tag{42}$$

When $a \leq 0$,

$$\begin{aligned}(41) &\leq \Pr\left(|X| > y + a\right) \\ &\leq c_1 e^{-c_2(y+a)^2} \\ &\leq c_1 c^{c_2 a^2} e^{-c_2 y^2/2}. \end{aligned} \tag{43}$$

Combining Equations (42) and (43), one can show that

$$\Pr\left(|Y| > y\right) \leq c_1 c^{c_2 a^2} e^{-c_2 y^2/2},\ \forall y > |a|.$$

On the other hand,

$$\Pr\left(|Y| \leq y\right) \leq 1 \leq e^{c_2 a^2/2} e^{-c_2 y^2/2},\ \forall y \leq |a|.$$

Let $c_3 = \max\left(c_1 e^{c_2 a^2}, e^{c_2 a^2/2}\right)$ and $c_4 = c_2/2$. Then

$$\Pr\left(|Y| > y\right) \leq c_3 e^{-c_4 y^2}.$$

This proves the claimed result. ∎

B. The Berry-Esseen Theorem and Its Consequence

The Central Limit Theorem (CLT) states that under certain conditions, an appropriately normalized sum of independent random variables converges weakly to the standard Gaussian distribution. The Berry-Esseen theorem quantifies the rate at which this convergence takes place.

Theorem 13 (The Berry-Esseen Theorem): Let $X_1, X_2, \ldots, X_k$ be independent random variables such that $\mathrm{E}\left[X_i\right] = 0$, $\mathrm{E}\left[X_i^2\right] = \sigma_i^2$, $\mathrm{E}\left[\left|X_i^3\right|\right] = \rho_i$. Also, let

$$s_k^2 = \sigma_1^2 + \cdots + \sigma_k^2,$$

and

$$r_k = \rho_1 + \cdots + \rho_k.$$

Denote by F_k the cumulative distribution function of the normalized sum $\left(X_1 + \cdots + X_k\right)/s_k$, and by $\mathscr{N}$ the standard Gaussian distribution. Then for all x and k,

$$\left|F_k\left(x\right) - \mathscr{N}\left(x\right)\right| \leq 6\frac{r_k}{s_k^3}.$$

The Berry-Esseen theorem is used in the proof of the lower bound for the achievable rate region of nonnegative l_1-WSCs. In the proof, one need to identify a tight bound on the probability of a weighted sum of nonnegative random variables. The probability of this sum lying in a given interval can be estimated by the Berry-Esseen, as summarized in the following lemma.

Lemma 2: Assume that $\mathbf{b} \in B_t^k$ is such that $\|\mathbf{b}\|_0 = k$, let $X_1, X_2, \cdots, X_k$ be independent standard Gaussian random variables. For a given positive constant $c > 0$, one has

$$\Pr\left(\left|\sum_{j=1}^{k} b_j \left|X_j\right|\right| < c\log\sqrt{k}\right) \leq \frac{c}{\pi}\frac{\log\sqrt{k}}{\sqrt{k}} + 12\rho\frac{t^3}{\sqrt{k}},$$

where $\rho := \mathrm{E}\left[\left|X\right|^3\right]$.

Proof: This lemma is proved by applying the Berry-Essen theorem. Note that the $b_j\left|X_j\right|$'s are independent random variables. Their sum $\sum_{j=1}^{k} b_j\left|X_j\right|$ can be approximated by a Gaussian random variable with properly chosen mean and variance, according to the Central Limit Theorem. In the proof, we first use the Gaussian approximation to estimate the probability

$$\Pr\left(\left|\sum_{j=1}^{k} b_j \left|X_j\right|\right| < c\log\sqrt{k}\right).$$

Then we subsequently employ the Berry-Essen theorem to upper bound the approximation error.

To simplify notation, let

$$Y_{\mathbf{b},k} = \sum_{j=1}^{k} b_j\left|X_j\right|,$$

and let $\mathcal{N}(x)$ denote, as before, the standard Gaussian distribution. Then,

$$\begin{aligned}
&\Pr\left(\left|Y_{\mathbf{b},k}\right| < c\log\sqrt{k}\right) \\
&\leq \Pr\left(\frac{Y_{\mathbf{b},k}}{\sqrt{\sum_j b_j^2}} \in \left(-\frac{c\log\sqrt{k}}{\sqrt{\sum_j b_j^2}}, \frac{c\log\sqrt{k}}{\sqrt{\sum_j b_j^2}}\right)\right) \\
&\leq \Pr\left(\frac{Y_{\mathbf{b},k}}{\sqrt{\sum_j b_j^2}} \in \left(-\frac{c\log\sqrt{k}}{\sqrt{k}}, \frac{c\log\sqrt{k}}{\sqrt{k}}\right)\right) \\
&\leq \Pr\left(\frac{Y_{\mathbf{b},k}}{\|\mathbf{b}\|_2} \leq \frac{c\log\sqrt{k}}{\sqrt{k}}\right) - \Pr\left(\frac{Y_{\mathbf{b},k}}{\|\mathbf{b}\|_2} \leq \frac{c\log\sqrt{k}}{\sqrt{k}}\right),
\end{aligned}$$

where in the second inequality we used the fact that $b_j \geq 1$, so that $\sum_{j=1}^{k} b_j^2 \geq k$.

According to Theorem 13, for all $x \in \mathbb{R}$ and all k,

$$\begin{aligned}
&\left|\Pr\left(\frac{Y_{\mathbf{b},k}}{\|\mathbf{b}\|_2} \leq x\right) - \mathcal{N}(x)\right| \\
&\leq 6\rho\,\frac{\sum_{j=1}^{k}\left|b_j\right|^3}{\left(\sum_{j=1}^{k}\left|b_j\right|^2\right)^{3/2}} \\
&\leq \frac{6k\rho t^3}{k^{3/2}} = \frac{6\rho t^3}{\sqrt{k}},
\end{aligned}$$

since $\sum_{j=1}^{k}\left|b_j\right|^3 \leq k\,t^3$, and $\sum_{j=1}^{k}\left|b_j\right|^2 \geq k$.

Thus,

$$\begin{aligned}
&\Pr\left(|Y_{\mathbf{b},k}| < c\log\sqrt{k}\right)\\
&\leq \mathcal{N}\left(\frac{c\log\sqrt{k}}{\sqrt{k}}\right) + \frac{6\rho t^3}{\sqrt{k}}\\
&\quad - \mathcal{N}\left(-\frac{c\log\sqrt{k}}{\sqrt{k}}\right) + \frac{6\rho t^3}{\sqrt{k}}\\
&\leq \frac{2}{2\pi}\frac{c\log\sqrt{k}}{\sqrt{k}} + \frac{12\rho t^3}{\sqrt{k}},
\end{aligned}$$

which completes the proof of the claimed result. ■

REFERENCES

[1] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Trans. Inform. Theory*, vol. 10, no. 4, pp. 363–377, 1964.

[2] T. Ericson and L. Györfi, "Superimposed codes in $\mathbf{R}^n$," *IEEE Trans. Inform. Theory*, vol. 34, no. 4, pp. 877–880, 1988.

[3] Z. Füredi and M. Ruszinkó, "An improved upper bound of the rate of Euclidean superimposed codes," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 799–802, 1999.

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

[5] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[6] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

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

[8] G. Cormode and S. Muthukrishnan, "What's hot and what's not: Tracking most frequent items dynamically," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[9] W. Dai and O. Milenkovic, "Weighted euclidean superimposed codes for integer compressed sensing," in *IEEE Information Theory Workshop (ITW)*, 2008, submitted.

[10] ——, "Constrained compressed sensing via superimposed coding," in *Information Theory and Applications Workshop*, San Diego, CA, invited talk, Jan. 2008.

[11] ——, "Sparse weighted euclidean superimposed coding for integer compressed sensing," in *Conference on Infomation Sciences and Systems (CISS)*, 2008, submitted.

[12] M. Sheikh, O. Milenkovic, and R. Baraniuk, "Designing compressive sensing DNA microarrays," *Proceedings of the IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), St. Thomas, U.S. Virgin Islands*, Dec. 2007.

[13] W. Dai, M. Sheikh, O. Milenkovic, and R. Baraniuk, "Probe designs for compressed sensing microarrays," in *IEEE International Conference on Bioinformatics and Biomedicine, Philadelphia, PA*, submitted, 2008.

[14] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume 2*, 2nd ed. Wiley, 1971.

[15]

[16] R. Vershynin, *Non-asymptotic Random Matrix Theory (Lecture Notes)*, 2007.