ISIT'2013 1569720207

Least Squares Superposition Codes with Bernoulli Dictionary are Still Reliable at Rates up to Capacity

Yoshinari Takeishi, Masanori Kawakita, and Jun'ichi Takeuchi
Graduate School of Information Science and Electrical Engineering, Kyushu University,
Motooka 744, Nishi-ku, Fukuoka-city, Fukuoka 819-0395, Japan

***Abstract*—For the additive white Gaussian noise channel with average power constraint, sparse superposition codes with least squares decoding were proposed by Barron and Joseph in 2010. The codewords are designed by using a dictionary which is drawn from a Gaussian distribution. The error probability is shown to be exponentially small in code length for all rates up to the capacity. This paper proves that when the dictionary is drawn from a Bernoulli distribution, the error probability is also exponentially small for all rates up to the capacity.**

I. Introduction

We analyze error probability of a certain modification of the sparse superposition codes [2], [3], [4], [11] ignoring computational complexity and prove that the error probability is exponentially small in code length.

The sparse superposition codes are codes for Additive White Gaussian Noise (AWGN) channel and designed using a matrix called dictionary. The codewords are sum of chosen column vectors from the dictionary. The codewords are real number strings of length n, and they are directly sent to the channel. Each entry of the dictionary is drawn from a Gaussian distribution so that the codeword is subject to the optimal Gaussian distribution for the input of the AWGN channel.

In [4], [11], Barron and Joseph proved that the error probability of the sparse superposition codes with least squares decoder is $O(\exp\{-nd\Delta^2\})$, where n is length of codeword, d is a positive constant and Δ is a gap between the rate and the capacity. Note that the least squares decoding is optimal but computationally intractable. This bound corresponds to the form of the optimal reliability bounds as in [10], [13], but the constant d is not confirmed to be optimal. Further, for every n, a specific upper bound of the error probability is stated in [11]. We review it as Theorem 1 in this paper.

Barron and Joseph conjectured that the codes for which each entry of dictionary is equiprobable $\pm$ 1 also achieve the capacity. In this paper we prove this conjecture. In particular, we demonstrated an upper bound on the error probability, which is comparable to the bound by Joseph and Barron.

Barron et al. proposed efficient decoding algorithms [2], [3] rather than the least squares decoder. Therefore, sparse superposition codes, as well as Polar codes [1], may be competitive to modern codes such as turbo codes [5] and Low Density Parity Check (LDPC) codes [9]. However, for implementation, Gaussian dictionary causes a problem that entries are not bounded and require arbitrary precision. This paper will help to solve the problem.

The earlier version of this material was presented as [15] (without peer review), but the provided bound was asymptotic one and the proof technique was different.

In Section II, we review the sparse superposition codes. We describe our result in Section III.

II. Sparse superposition codes

In this section, we review the sparse superposition codes, which employ the dictionary each entry of which is drawn from the Gaussian distribution.

A. Communication problem

We consider the communication for the AWGN channel. A sender would like to send a message $u \in \{0,1\}^K$ to the receiver correctly. Let u be subject to the uniform distribution, namely each message arises with probability $1/2^K$. The sender encodes the message to the codeword, then sends it to the channel. The codeword is a real number string of length n and denoted by c. We define the transmission rate R as K/n. The power of codeword c is defined by $(1/n)\sum_{i=1}^n c_i^2$. We constrain the average of the power across the 2^K codewords to be not more than P.

Let $N(\mu, \sigma^2)$ denote the Gaussian distribution with mean μ and variance σ^2. The AWGN channel adds noise ϵ to the codeword, where ϵ is a real number string of length n and each coordinate is subject to $N(0, \sigma^2)$. Thus, let Y denote the received word, then $Y = c + \epsilon$. Under the average power constraint P, it is well known that the channel capacity is $C = \frac{1}{2}\log(1+v)$, where $v = P/\sigma^2$.

The receiver estimates the massage u from the received word. Let $\hat{u}$ denote the estimate. It is desirable that $\hat{u} = u$ for the correct communication. The block error is the event $\hat{u} \neq u$. We require that the probability of the block error is small for sufficiently large n.

B. Encoding

For a sparse superposition code, we first map an input u into a coefficient vector $\beta \in \{0,1\}^N$ by a one to one function. The β is a sparse vector and the number of nonzero elements equals L. Here, we split $\beta = (\beta_1, \beta_2, \ldots, \beta_N)$ into L sections of size M ($N = LM$) and arrange with one coordinate 1 and the other coordinates 0 in each section. Then the codeword is formed as follows:

$$c = X\beta = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_N X_N,$$

where X is an $n \times N$ matrix (dictionary) and X_j is the jth column vector of X. Thus c is a superposition of L column vectors of X, with exactly one column selected from each section.

The dictionary X is made before coding, and shared by the sender and the receiver. Each element of X is independently drawn from $N(0, P/L)$. Then if we consider X as a random variable, each entry of c is independently subject to $N(0, P)$ because c is sum of L column vectors of X. For the random coding in [14], this distribution is the optimal for the input of the AWGN channel.

The parameters L, M, and N are selected with satisfying the following. The number of messages is 2^K and that of codewords is M^L. Thus we arrange $2^K = M^L$, equivalently, $K = L \log_2 M$. The value of M is set to be L^a and the parameter a is referred to as section size rate. Then $K = aL \log_2 L$ and $n = (aL \log_2 L)/R$.

C. Decoding

We review the least squares decoder employed in [11]. From the received word Y and knowledge of the dictionary X, the receiver estimates the message u, equivalently, estimates the corresponding β. Define a set $\mathcal{B}$ as

$$\mathcal{B} = \{\beta \in \{0,1\}^N \mid \beta_j \text{ has one 1 in each section}\}.$$

Then the least squares decoder is given by

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{B}} \|Y - X\beta\|^2$$

where $\|\cdot\|$ denotes the Euclidean norm.

Let β^* to be a true β and $\hat{\beta}$ to be a solution of the decoder. The block error occurs when $\hat{\beta} \neq \beta^*$. Further, let $mistakes$ denote the number of sections in which the position of the nonzero term in $\hat{\beta}$ is different from that in the β^*. Define the error event $\mathcal{E}_{\alpha_0} = \{mistakes \geq \alpha_0 L\}$, that the decoder makes mistakes in at least α_0 fraction of sections. A proportion of mistakes $\alpha = mistakes/L$ is called section error rate.

D. Performance

It is proved in the literature [11] that given $0 < \alpha_0 \leq 1$, the probability of the event $\mathcal{E}_{\alpha_0}$ is exponentially small in n. The following theorem provides an upper bound on the probability of the event $\mathcal{E}_{\alpha_0}$, where

$$w_v = \frac{v}{[4(1+v)^2]\sqrt{1 + (1/4)v^3/(1+v)}}$$

and $g(x) = \sqrt{1+4x^2} - 1$. It follows that

$$g(x) \geq \min\{\sqrt{2}x, x^2\} \quad \text{for all} \quad x \geq 0.$$

The definition of $a_{v,L}$ is given later as (2).

Theorem 1 (Joseph and Barron 2012): Suppose that each entry of X is independently drawn from $N(0, P/L)$. Assume $M = L^a$, where $a \geq a_{v,L}$, and rate R is less than capacity C. Then

$$\Pr[\mathcal{E}_{\alpha_0}] = e^{-nE(\alpha_0, R)}$$

with $E(\alpha_0, R) \geq h(\alpha_0, C - R) - (\log(2Rn/a))/n$, where

$$h(\alpha, \Delta) = \min\left\{\alpha w_v \Delta, \frac{1}{4} g\left(\frac{\Delta}{2\sqrt{v}}\right)\right\}$$

is evaluated at $\alpha = \alpha_0$ and $\Delta = C - R$.

Now, we need to prove that the block error probability is exponentially small. For that purpose, we use composition with an outer Reed-Solomon (RS) code [12] of rate near one.

When $C - R$ is small, Theorem 1 yields the bound of $O(\exp\{-nd\Delta^2\})$.

To prove Theorem 1, we need to evaluates the probability of the event $E_l = \{misrtakes = l\}$ for $l = 1, 2, \ldots, L$. $\Pr[E_l]$ is used for evaluating

$$\Pr[\mathcal{E}_{\alpha_0}] = \sum_{l \geq \alpha_0 L} \Pr[E_l].$$

Let v denote the signal-to-noise ratio P/σ^2, then the channel capacity of AWGN is given as $C = (1/2)\log(1+v)$. In addition, we introduce a function $C_\alpha = (1/2)\log(1+\alpha v)$ for $0 \leq \alpha \leq 1$. $C_\alpha - \alpha C$ is a nonnegative function equal to 0 when α is 0 or 1 and strictly positive in between. Thus the quantity $C_\alpha - \alpha R$ is larger than $\alpha(C - R)$, which is positive when $R < C$.

For a positive Δ and $\rho \in [-1, 1]$, we define a quantity $D(\Delta, 1-\rho^2)$ as

$$D(\Delta, 1-\rho^2) = \max_{\lambda \geq 0}\{\lambda\Delta + (1/2)\log(1 - \lambda^2(1-\rho^2))\}$$

and $D_1(\Delta, 1-\rho^2)$ as

$$D_1(\Delta, 1-\rho^2) = \max_{0 \leq \lambda \leq 1}\{\lambda\Delta + (1/2)\log(1 - \lambda^2(1-\rho^2))\}.$$

Note that these quantities are nonnegative.

The following lemma (given in [11]) evaluates $\Pr[E_l]$.

Lemma 2 (Joseph and Barron 2012): Suppose that each entry of X is independently drawn from $N(0, P/L)$. Let a positive integer $l \leq L$ be given and let $\alpha = l/L$. Then, $\Pr[E_l]$ is bounded by the minimum for t_α in the interval $[0, C_\alpha - \alpha R]$ of $err_{Gauss}(\alpha)$, where

$$\begin{aligned} err_{Gauss}(\alpha) = {}& {}_L C_{\alpha L} \exp\{-nD_1(\Delta_\alpha, 1-\rho_\alpha^2)\} \\ & + \exp\{-nD(t_\alpha, \alpha^2 v/(1+\alpha^2 v))\} \end{aligned} \tag{1}$$

with $\Delta_\alpha = C_\alpha - \alpha R - t_\alpha$ and $1 - \rho_\alpha^2 = \alpha(1-\alpha)v/(1+\alpha v)$.

Define $a_{v,L}$ by

$$a_{v,L} = \max_{\alpha \in \{\frac{1}{L}, \frac{2}{L}, \ldots, 1-\frac{1}{L}\}} \frac{R \log {}_L C_{L\alpha}}{D_1(C_\alpha - \alpha C, 1-\rho_\alpha^2) L \log L}. \tag{2}$$

To confirm (1) is exponentially small, it is sufficient that the section size rate a is larger than $a_{v,L}$ (see Lemma 5 in [11]).

III. Performance with Bernoulli dictionary

In this section, we discuss the performance of the sparse superposition codes with Bernoulli dictionary. We draw each entry of the dictionary as the following independent random variable:

$$X_{ij} = \begin{cases} -\sqrt{P/L} \text{ (with probability } 1/2) \\ \sqrt{P/L} \text{ (otherwise)} \end{cases}$$

A. Main result

In this setting, we have the following theorem:

Theorem 3: Suppose that each entry of X is independent equiprobable $\pm\sqrt{P/L}$. Assume $M = L^a$, where $a \geq a_{v,L}$, and rate R is less than capacity C. Let $L \geq 1000$ and $\alpha_0 L \geq 50$, then

$$\Pr[\mathcal{E}_{\alpha_0}] = e^{-nE(\alpha_0,R)}$$

with

$$E(\alpha_0, R) \geq h(\alpha_0, C-R) - (\log(2Rn/a))/n - \bar{\iota}(L)$$

where $\bar{\iota}(L) = \max\{\bar{\iota}_1, \bar{\iota}_2\}$, which are defined in Lemma 4.

This theorem is the correspondent to Theorem 1 in this case, though the lower bound of $E(\alpha_0, R)$ in Theorem 3 is less than that in Theorem 1. The difference $\bar{\iota}$ can be small as n gets large, which is discussed in detail below.

In order to prove this theorem, we prove the following lemma, which is the correspondent to Lemma 2 in this case.

Lemma 4: Suppose that each entry of X is independent equiprobable $\pm\sqrt{P/L}$. Assume $M = L^a$, where $a \geq a_{v,L}$, and rate R is less than capacity C. Let $L \geq 1000$ and $\alpha_0 L \geq 50$, and $\alpha = l/L$. Then for all l such that $\alpha_0 \leq \alpha \leq 1$, $\Pr[E_l]$ is bounded by the minimum for t_α in the interval $[0, C_\alpha - \alpha R]$ of $err_{Ber}(\alpha)$, where

$$\begin{aligned} err_{Ber}(\alpha) = {}& {}_L C_{\alpha L} \exp\{-n(D_1(\Delta_\alpha, 1-\rho_\alpha^2) - \bar{\iota}_1)\} \\ & + \exp\{-n(D(t_\alpha, \alpha^2 v/(1+\alpha^2 v)) - \bar{\iota}_2)\} \end{aligned}$$

with $\Delta_\alpha = C_\alpha - \alpha R - t_\alpha$ and $1-\rho_\alpha^2 = \alpha(1-\alpha)v/(1+\alpha v)$. The variables $\bar{\iota}_1$ and $\bar{\iota}_2$ are defined as

$$\begin{aligned} \bar{\iota}_1 &= \ln((1+\bar{\iota}_3)(1+\max\{\bar{\iota}_4, \bar{\iota}_5\})) \\ \bar{\iota}_2 &= f_\zeta(L) + \frac{6\sqrt{\ln L}}{\sqrt{L}} \\ 1+\bar{\iota}_3 &= \max_{\alpha_0 L \leq l \leq L} \left(e^{f_\zeta(l)}\right) \left(1 + \frac{1+v}{\sqrt{L}}\right) \\ &\quad \times \exp\left\{\frac{3(1+v)}{\sqrt{\alpha_0}} \frac{\sqrt{\ln L}}{\sqrt{L}}\right\} \\ 1+\bar{\iota}_4 &= \max_{\alpha_0 L \leq l \leq L-\sqrt{L}} \left(e^{f_\zeta(l)+f_\zeta(L-l)}\right) \left(1 + \frac{7\sqrt{\ln L}}{\sqrt{L}}\right) \\ &\quad \times \exp\left\{\frac{10\alpha_1^2\sqrt{\ln L}}{L^{1/4}}\right\} \\ 1+\bar{\iota}_5 &= \max_{L-\sqrt{L} \leq l \leq L-1} \frac{e^{f_\zeta(l)}(1+1/\sqrt{L})}{\sqrt{1-2/\sqrt{L}}} \exp\left\{\frac{3\ln L}{\sqrt{L}}\right\} \end{aligned}$$

where $\alpha_1 = \sqrt{1+1/(4\sqrt{\alpha_0})}$, and $f_\zeta(l)$ is defined in Lemma 5.

In the proof of Lemma 4, we need to bound ${}_k C_l (1/2)^l$ nearly by $N(k \mid l/2, l/4)$ uniformly for all $k \in \{0, 1, \ldots, l\}$. For that purpose, we proved the following lemma via Stirling's formula and Taylor expansion.

Lemma 5: For any integer l and any $\zeta \in (0, 1/2)$, it follows that

$$\max_{k \in \{0,1,\ldots,l\}} \frac{{}_l C_k (1/2)^l}{N(k \mid l/2, l/4)} \leq \exp\{f_\zeta(l)\}$$

where

$$\begin{aligned} f_\zeta(l) = \max\Big\{ & \left(\frac{3}{16}c_\zeta^2 + \frac{1}{12}\right)\frac{1}{l}, -\frac{4\zeta^4}{3}l + \log\frac{l}{2} + \frac{1}{12l}, \\ & \frac{1}{2}\log\frac{\pi l}{2} - \left(\log 2 - \frac{1}{2}\right) l \Big\} \end{aligned}$$

and $c_\zeta = 1/(1+2\zeta)^2 + 1/(1-2\zeta)^2$.

B. Outline of the proof of Lemma 4

For a discrete random variable Z', let $P_{Z'}$ denote the probability mass function of Z'.

The proof proceeds along the lines of the proof of Lemma 4 in [11]. We evaluate the probability of the event E_l. The random variables are the dictionary $X = (X_1, X_2, \ldots, X_N)$ and the noise ϵ.

For $\beta \in \mathcal{B}$, let $S(\beta) = \{j \mid \beta_j = 1\}$ denote the set of indices for which β is nonzero. Further, let $\mathcal{A} = \{S(\beta) \mid \beta \in \mathcal{B}\}$ denote the set of allowed subset of terms. Let β^* denote β which is sent, and let $S^* = S(\beta^*)$. Furthermore, let $X_S = \sum_{j \in S} X_j$. For the occurrence of E_l, there must be an $S \in \mathcal{A}$ which differs from S^* in an amount l and which has $\|Y - X_S\|^2 \leq \|Y - X_{S^*}\|^2$. Here we define $T(S)$ as

$$T(S) = \frac{1}{2}\left[\frac{|\,Y - X_S\,|^2}{\sigma^2} - \frac{|\,Y - X_{S^*}\,|^2}{\sigma^2}\right]$$

where for a vector a of length n, $|a|^2$ denote $(1/n)\sum_{i=1}^n a_i^2$. Then $\|Y - X_S\|^2 \leq \|Y - X_{S^*}\|^2$ and $T(S) \leq 0$ are equivalent.

The subsets S and S^* have an intersection $S_1 = S \cap S^*$ of size $L - l$ and a deference $S_2 = S - S_1$ of size l.

We decompose $T(S) = \widetilde{T}(S) + T^*$, where

$$\widetilde{T}(S) = \frac{1}{2}\left[\frac{|\,Y - X_S\,|^2}{\sigma^2} - \frac{|\,Y - (1-\alpha)X_{S^*}\,|^2}{\sigma^2 + \alpha^2 P}\right]$$

and

$$T^* = \frac{1}{2}\left[\frac{|\,Y - (1-\alpha)X_{S^*}\,|^2}{\sigma^2 + \alpha^2 P} - \frac{|\,Y - X_{S^*}\,|^2}{\sigma^2}\right].$$

For a positive $\tilde{t} = t_\alpha$, let $\tilde{E}_l$ denotes a event that there is an $S \in \mathcal{A}$ which differs from S^* in an amount l and $\widetilde{T}(S) \leq \tilde{t}$. Similarly, for a negative $t^* = -t_\alpha$, let E_l^* denotes a corresponding event that $T^* \leq t^*$.

First, we evaluate $\Pr[\tilde{E}_l]$. We further decompose $\widetilde{T}(S) = \widetilde{T}_1(S_1) + \widetilde{T}_2(S)$, where

$$\widetilde{T}_1(S_1) = \frac{1}{2}\left[\frac{|\,Y - X_{S_1}\,|^2}{\sigma^2 + \alpha P} - \frac{|\,Y - (1-\alpha)X_{S^*}\,|^2}{\sigma^2 + \alpha^2 P}\right]$$

and

$$\widetilde{T}_2(S) = \frac{1}{2}\left[\frac{|\,Y - X_S\,|^2}{\sigma^2} - \frac{|\,Y - X_{S_1}\,|^2}{\sigma^2 + \alpha P}\right].$$

In [11], the following inequality is provided ((21) in p. 2547);

$$\Pr[\tilde{E}_l] \le \sum_{S_1} \mathbb{E}_{Y,X_{S^*}} e^{-n\lambda(\widetilde{I}_1(S_1)-\tilde{\imath})} \left(\sum_{S_2} \mathbb{E}_{X_{S_2}} e^{-n\widetilde{I}_2(S)} \right)^{\lambda}.$$

We can prove

$$\mathbb{E}_{X_{S_2}} e^{-n\widetilde{I}_2(S)} \quad \le \quad (1+\bar{\iota}_3)^n e^{-nC_\alpha}.$$

Here, we omit the proof, but it can be done similarly as the proof of (7) described later.

Then, $\Pr[\tilde{E}_l]$ is evaluated as

$$\Pr[\tilde{E}_l] \le \sum_{S_1} \mathbb{E}_{Y,X_{S^*}} e^{-n\lambda(\widetilde{I}_1(S_1)-\tilde{\imath})} \left(\sum_{S_2} (1+\bar{\iota}_3)^n e^{-nC_\alpha} \right)^{\lambda}.$$

Noting that the number of the sum for S_2 is less than $M^l = e^{nR\alpha}$, it follows that

$$\Pr[\tilde{E}_l] \le (1+\bar{\iota}_3)^n \sum_{S_1} \mathbb{E}_{Y,X_{S^*}} e^{-n\lambda\widetilde{I}_1(S_1)} e^{-n\lambda\Delta_\alpha}, \qquad (3)$$

where $\Delta_\alpha = C_\alpha - \alpha R - t_\alpha$.

Now we evaluate $\mathbb{E}_{Y,X_{S^*}} e^{-n\lambda\widetilde{I}_1(S_1)}$. Let $\tilde{W}_i (i = 1,2,\ldots,L)$ be random variables independently equiprobable ± 1. Let $W_1 = \sum_{i=1}^{l} \tilde{W}_i/\sqrt{l}$, $W_2 = \sum_{i=l+1}^{L} \tilde{W}_i/\sqrt{L-l}$ and $\epsilon' = \epsilon/\sigma$. We define Z and $\tilde{Z}$ as $Z = (\sigma\epsilon' + \alpha(\sqrt{\alpha P}W_1 + \sqrt{(1-\alpha)P}W_2))/\sqrt{\sigma^2+\alpha^2 P}$ and $\tilde{Z} = (\sigma\epsilon' + \sqrt{\alpha P}W_1)/\sqrt{\sigma^2+\alpha P}$, respectively. Then it follows that

$$\mathbb{E}_{Y,X_{S^*}} e^{-n\lambda\widetilde{I}_1(S_1)} = \left[\mathbb{E}_{Z_2,\tilde{Z}_2} e^{(\lambda/2)(Z_2^2-\tilde{Z}_2^2)} \right]^n.$$

We want to bound $\mathbb{E}_{Z_2,\tilde{Z}_2} e^{(\lambda/2)(Z_2^2-\tilde{Z}_2^2)}$ nearly by the same expression in case the dictionary is drawn from the Gaussian distribution.

Here, according to the assumption $\alpha_0 L \le l \le L$, the central limit theorem about W_2 does not always work. Thus, we will make case argument for (i) $l \le L - \sqrt{L}$ and (ii) $l > L - \sqrt{L}$.

First, we consider the case (i) that the central limit theorem about W_2 does work. Define sets $\mathcal{X}_2 = \{h_2(k - l/2) \mid k = 0,1,\ldots,l\}$ and $\mathcal{X}_3 = \{h_3(k' - l'/2) \mid k' = 0,1,\ldots,l'\}$, where $h_2 = 2/\sqrt{l}$ and $h_3 = 2/\sqrt{l'}$ with $l' = L - l$. Then, for $w_3 \in \mathcal{X}_3$, we have by Lemma 5

$$P_{W_1}(w_1) \le \exp\{f_\zeta(l')\} h_2 N(w_1 \mid 0,1).$$

The expression $z_2^2 - \tilde{z}_2^2$ can be represented by the quadratic form about $\mathbf{x} = (w_1, w_2, \epsilon')$ as $z_2^2 - \tilde{z}_2^2 = \mathbf{x}^T \tilde{B} \mathbf{x}$, where $\tilde{B}$ is a 3×3 matrix defined as

$$\tilde{B} = \begin{pmatrix} \frac{\alpha v}{1+\alpha v} - \frac{\alpha^3 v}{1+\alpha^2 v} & -\frac{\alpha^2\sqrt{\alpha(1-\alpha)}v}{1+\alpha^2 v} & \frac{\sqrt{\alpha v}}{1+\alpha v} - \frac{\alpha\sqrt{\alpha v}}{1+\alpha^2 v} \\ -\frac{\alpha^2\sqrt{\alpha(1-\alpha)}v}{1+\alpha^2 v} & \frac{\alpha^2(1-\alpha)v}{1+\alpha^2 v} & -\frac{\alpha\sqrt{(1-\alpha)v}}{1+\alpha^2 v} \\ \frac{\sqrt{\alpha v}}{1+\alpha v} - \frac{\alpha\sqrt{\alpha v}}{1+\alpha^2 v} & -\frac{\alpha\sqrt{(1-\alpha)v}}{1+\alpha^2 v} & \frac{1}{1+\alpha v} - \frac{1}{1+\alpha^2 v} \end{pmatrix}.$$

Define $\tilde{A} = I - \lambda\tilde{B}$, where I is a unit matrix of order 3. Then, the expectation $\mathbb{E}_{Z_2,\tilde{Z}_2} \exp\left[\frac{\lambda}{2}(Z_2^2 - \tilde{Z}_2^2)\right]$ is bounded by

$$\frac{e^{f_\zeta(l)+f_\zeta(l')}}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} h_2 h_3 \sum_{x_1 \in \mathcal{X}_2} \sum_{x_2 \in \mathcal{X}_3} e^{-\mathbf{x}^T \tilde{A}\mathbf{x}/2} dx_3. \qquad (4)$$

Now, we are to replace the summation about x_1 and x_2 by the integral. The integral $(2\pi)^{-3/2} \int_{\Re^3} e^{-\mathbf{x}^T\tilde{A}\mathbf{x}/2} d\mathbf{x}$ equals $\mathbb{E}_{Z_2,\tilde{Z}_2} \exp\left[\frac{\lambda}{2}(Z_2^2 - \tilde{Z}_2^2)\right]$ in the Gaussian dictionary case, which is $1/\sqrt{1-\lambda^2(1-\rho_1^2)}$, where $1-\rho_1^2 = \alpha(1-\alpha)v/(1+\alpha v)$ [11]. Recalling $0 \le \lambda \le 1$, this integral is finite, so $\tilde{A}$ is a strictly positive definite symmetric matrix.

The symmetric matrix $\tilde{A}$ can be diagonalized as $\Gamma = U^T \tilde{A} U$ with an orthogonal matrix U and a diagonal matrix Γ. The diagonal entries of Γ are γ_1, γ_2 and γ_3, where they are eigenvalues of $\tilde{A}$ with $0 < \gamma_1 \le \gamma_2 \le \gamma_3$. Note that $\gamma_3 \le \sqrt{12}$. Define $\tilde{A}^{1/2}$ as $U\Gamma^{1/2}U^T$, where $\Gamma^{1/2}$ is a diagonal matrix of which the diagonal entries are $\sqrt{\gamma_1}$, $\sqrt{\gamma_2}$ and $\sqrt{\gamma_3}$. Here, for fixed x_3, define $\mathbf{y}_{k,k'} = \tilde{A}^{1/2}\mathbf{x}_{k,k'}$, where $\mathbf{x}_{k,k'} = (h_2(k - l/2), h_3(k' - l'/2), x_3)$ for $k = 0,1,\ldots,l$ and $k' = 0,1,\ldots,l'$. Then, we have

$$\sum_{x_1 \in \mathcal{X}_2} \sum_{x_2 \in \mathcal{X}_3} \exp\left\{\frac{-\mathbf{x}^T\tilde{A}\mathbf{x}}{2}\right\} = \sum_{k=0}^{l} \sum_{k'=0}^{l'} \exp\{f(\mathbf{y}_{k,k'})\}.$$

where $f(\mathbf{y}) = -\mathbf{y}^T\mathbf{y}/2$. Let $B_{k,k'}(x_3)$ denote a region $\{\mathbf{x} = (x_1,x_2,x_3)^T \in \Re^3 \mid x_1 \in [h_2(k-l/2), h_2(k+1-l/2)], x_2 \in [h_3(k'-l'/2), h_3(k'+1-l'/2)]\}$ and $A_{k,k'}(x_3)$ denote a region $\{\mathbf{y} \mid \mathbf{y} = \tilde{A}^{1/2}\mathbf{x}, \mathbf{x} \in B_{k,k'}(x_3)\}$. For $\mathbf{y} \in A_{k,k'}(x_3)$, according to Taylor expansion of $f(\mathbf{y})$ at $\mathbf{y} = \mathbf{y}_{k,k'}$, there exists $t \in [0,1]$ such that

$$f(\mathbf{y}) = f(\mathbf{y}_{k,k'}) - (\mathbf{y} - \mathbf{y}_{k,k'})^T \mathbf{y}^{(t)}$$

where $\mathbf{y}^{(t)} = t\mathbf{y}_{k,k'} + (1-t)\mathbf{y}$. Here, $\|\mathbf{y} - \mathbf{y}_{k,k'}\|$ is bounded by

$$\|\tilde{A}^{1/2}(\mathbf{x} - \mathbf{x}_{k,k'})\| \le \sqrt{\gamma_3}\|\mathbf{x} - \mathbf{x}_{k,k'}\| \le \sqrt{\gamma_3} h_4$$

where $\mathbf{x} = \tilde{A}^{-1/2}\mathbf{y}$ and $h_4 = \sqrt{h_2^2 + h_3^2}$ with $\tilde{A}^{-1/2} = (\tilde{A}^{1/2})^{-1}$. Similarly, $\|\mathbf{y} - \tilde{\mathbf{y}}\|$ is also bounded by $\sqrt{\gamma_3}h_4$. Thus, it follows that $\|\tilde{\mathbf{y}}\| \le \|\mathbf{y}\| + \|\mathbf{y} - \tilde{\mathbf{y}}\| \le \|\mathbf{y}\| + \sqrt{\gamma_3}h_4$. Then, we have

$$\begin{aligned} f(\mathbf{y}) &\ge f(\mathbf{y}_{k,k'}) - \|\mathbf{y} - \mathbf{y}_{k,k'}\| \times \|\tilde{\mathbf{y}}\| \\ &\ge f(\mathbf{y}_{k,k'}) - a(\mathbf{y}) \end{aligned}$$

where $a(\mathbf{y}) = \sqrt{\gamma_3}h_4(\|\mathbf{y}\| + \sqrt{\gamma_3}h_4)$. Thus, it follows that

$$\exp\{f(\mathbf{y}_{k,k'})\} \le \min_{\mathbf{y} \in A_{k,k'}(x_3)} \exp\{f(\mathbf{y}) + a(\mathbf{y})\}.$$

Let $S(A_{k,k'}(x_3)) = \sqrt{\gamma_1\gamma_2\gamma_3}h_2h_3$ denote the area of $A_{k,k'}(x_3)$, we have

$$S(A_{k,k'}(x_3)) \min_{\mathbf{y} \in A_{k,k'}(x_3)} e^{f(\mathbf{y})+a(\mathbf{y})} \le \int_{A_{k,k'}(x_3)} e^{f(\mathbf{y})+a(\mathbf{y})} d\mathbf{y}.$$

Finally, we have

$$\int_{-\infty}^{\infty} h_2h_3 \sum_{x_1,x_2} e^{-\mathbf{x}^T\tilde{A}\mathbf{x}/2} dx_3 \le \frac{1}{\sqrt{\gamma_1\gamma_2\gamma_3}} \int_{\Re^3} e^{f(\mathbf{y})+a(\mathbf{y})} d\mathbf{y}. \qquad (5)$$

Here, divide the region of the integral as $\|\mathbf{y}\| \le B_3$ and $\|\mathbf{y}\| \ge B_3$. In the region $\|\mathbf{y}\| \le B_3$, $a(\mathbf{y})$ can be bounded by $\bar{a} = \sqrt{\gamma_3}h_4(B_3 + \sqrt{\gamma_3}h_4)$. Thus, we have

$$\int_{\|\mathbf{y}\| \le B_3} e^{f(\mathbf{y})+a(\mathbf{y})} d\mathbf{y} \le e^{\bar{a}} \int_{\Re^3} e^{f(\mathbf{y})} d\mathbf{y} = e^{\bar{a}}(2\pi)^{3/2}. \qquad (6)$$

Now consider the region $\|\mathbf{y}\| \geq B_3$. In this range, $\|y\|$ is bounded by $\|y\|^2/B$. Then, $a(\mathbf{y})$ is bounded by

$$\frac{\sqrt{\gamma_3}h_4}{B_3}\mathbf{y}^T\mathbf{y} + \gamma_3 h_4^2$$

Thus, $\exp\{f(\mathbf{y}) + a(\mathbf{y})\}$ is bounded by $\exp\{-\xi\mathbf{y}^T\mathbf{y}/2 + \gamma_3 h_4^2\}$, where $\xi = 1 - 2\sqrt{\gamma_3}h_4/B_3$. Here, consider the integral

$$I(B_3) = \int_{\|\mathbf{y}\|\geq B_3} \exp\left\{-\frac{\xi}{2}\mathbf{y}^T\mathbf{y}\right\} d\mathbf{y}.$$

Now that $B_3 > 2\sqrt{\gamma_3}h_4$, we have $\xi > 0$, then this integral is finite. By some manipulation, we have

$$I(B_3) \leq \frac{4\pi(B_3^2\xi + 1)}{B_3\xi^2} \exp\left\{-\frac{\xi B_3^2}{2}\right\}.$$

Finally, we have

$$\int_{\|\mathbf{y}\|\geq B_3} e^{f(\mathbf{y})+a(\mathbf{y})} d\mathbf{y} \leq \frac{4\pi(B_3^2\xi + 1)}{B_3\xi^2} \exp\left\{-\frac{\xi B_3^2}{2} + \gamma_3 h_4^2\right\}$$

Together with (5) and (6), the expectation $\mathbb{E}_{Z,\tilde{Z}} \exp\left[\frac{\lambda}{2}(Z^2 - \tilde{Z}^2)\right]$ is bounded by

$$\frac{(2\pi)^{3/2}}{\sqrt{\gamma_1\gamma_2\gamma_3}}\left(e^{\bar{a}} + \frac{B_3^2\xi + 1}{B_3\xi^2} \exp\left\{-\frac{\xi B_3^2}{2} + \gamma_3 h_4^2\right\}\right).$$

Note that $1/\sqrt{\gamma_1\gamma_2\gamma_3}$ equals $\mathbb{E}_{Z,\tilde{Z}} \exp\left[\frac{\lambda}{2}(Z^2 - \tilde{Z}^2)\right]$ in the Gaussian dictionary case, which is $1/\sqrt{1-\lambda^2(1-\rho_\alpha^2)}$. Since $B_3 = \sqrt{\ln L}$, we can show

$$\mathbb{E}_{Z_2,\tilde{Z}_2} \exp\left[\frac{\lambda}{2}(Z_2^2 - \tilde{Z}_2^2)\right] \leq \frac{1 + \bar{\iota}_4}{\sqrt{1-\lambda^2(1-\rho_1^2)}}. \tag{7}$$

For the case (ii) $l \geq L - \sqrt{L}$, the central limit theorem for W_2 does not work for some l, but we take advantage of the fact that the influence of W_2 is small when $L - l$ is small. In fact, it is not difficult to show that

$$\mathbb{E}_{Z,\tilde{Z}} \exp\left[\frac{\lambda}{2}(Z^2 - \tilde{Z}^2)\right] \leq \frac{1 + \bar{\iota}_5}{\sqrt{1-\lambda^2(1-\rho_\alpha^2)}}. \tag{8}$$

Thus $\Pr[\tilde{E}_l]$ is evaluated as

$$\Pr[\tilde{E}_l] = {}_L C_{\alpha L} \exp\{-n(D_1(\Delta_\alpha, 1-\rho_\alpha^2) - \bar{\iota}_1)\}. \tag{9}$$

By the similar discussion with the evaluation of $\Pr[\tilde{E}_l]$, $\Pr[E_l^*]$ is evaluated as

$$\Pr[E_l^*] \leq \exp\{-n(D(t_\alpha - t, 1-\rho^2) - \bar{\iota}_2)\}. \tag{10}$$

Thus (9) and (10) yield the bound to be obtained.

C. Proof of Theorem 3

We evaluate $\Pr[\mathcal{E}_\alpha]$. If $R < C$, it follows that

$$\begin{aligned}\Pr[\mathcal{E}_\alpha] = \sum_{l=\alpha_0 L}^{L} \Pr[E_l] &\leq \sum_{l=\alpha_0 L}^{L} err_{Ber}(\alpha) \\ &\leq \sum_{l=\alpha_0 L}^{L} err_{Gauss}(\alpha)\exp\{n\bar{\iota}\}\end{aligned}$$

from Lemma 2 and Lemma 4. If $a \geq a_{v,L}$, it follows that

$$\Pr[\mathcal{E}_\alpha] \leq \exp\{-n(h(\alpha_0, C-R) - (\log(2Rn/a))/n - \bar{\iota})\}$$

from Theorem 1. Thus, we have Theorem 3.

Acknowledgment

The authors thank Professor Andrew R. Barron for his valuable comments. This research was partially supported by JSPS KAKENHI Grant Number 24500018 and FUJIFILM Corporation.

References

[1] E. Arikan, "Channel polarization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051-3073, Jul. 2009.

[2] A. R. Barron and A. Joseph, "Sparse superposition codes are fast and reliable at rates approaching capacity with Gaussian noise," Jun. 2011. Available: http://www.stat.yale.edu/~arb4/publications.html

[3] A. R. Barron and A. Joseph, "Towards fast reliable communication at rates near capacity with Gaussian noise," *Proc. IEEE. Int. Symp. Inf. Theory*, Austin, Texas, Jun. 13-18, 2010, pp. 315-319.

[4] A. R. Barron and A. Joseph, "Least squares superposition coding of moderate dictionary size, reliable at rates up to channel capacity," *Proc. Int. Symp. Inf. Theory*, Austin, Texas, June 13-18, 2010, pp. 275-279.

[5] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding: turbo codes," in *Proc. Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 1064-1070.

[6] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2-14, Jan. 1972.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 2006.

[8] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1957.

[9] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol.IT-8, pp. 21-28, Jan. 1962.

[10] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[11] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541-2557, May 2012.

[12] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. SIAM*, vol. 8, pp. 300-304, Jun. 1960.

[13] Y. Polyanskiy, H. V. Poor, and S. Verdu, " Channel coding rate in the finite block length regime," *IEEE Trans. Inf. Theory,* vol. 56, no. 5, pp. 2307-2359, May 2010.

[14] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.,* vol.27, pp. 379-423, 1948.

[15] Y. Takeishi, M. Kawakita, and J. Takeuchi, "Sparse superposition codes with Bernoulli dictionary," *Technical Report of IEICE*, Vol. 112, No. 215, IT2012-38, pp. 41-46, Sep. 2012.