MATRIX ANALYSIS OF COMMUNICATION AND BRAIN NETWORKS

A Dissertation

by

SHUO LI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Tie Liu
Committee Members,	Krishna Narayanan
	Le Xie
	Anxiao Jiang
Head of Department,	Miroslav M. Begovic

December 2016

Major Subject: Electrical Engineering

ABSTRACT

In this dissertation, we study two network problems using matrices as our primary analysis tools.

First, the limits of treating interference as noise are studied for the canonical two-user symmetric Gaussian interference channel. A two-step approach is proposed for finding approximately optimal input distributions in the high signal-to-noise ratio (SNR) regime. First, approximately and precisely optimal input distributions are found for the Avestimehr-Diggavi-Tse (ADT) linear deterministic model. These distributions are then translated, systematically, into Gaussian models, which we show can achieve the sum capacity to within a finite gap.

Next, the problem of clustering for brain networks based on the resting-state fMRI time-series data is studied. Our approach is based on the classical K-means algorithm, using Mahalanobis distance as the distance metric. We first consider the hypothetical case where the ground truth is available, so an optimal distance metric can be learned from it. This naturally motivates an unsupervised clustering algorithm that alternates between clustering and metric learning. The performance of the proposed algorithm is evaluated via computer simulations.

DEDICATION

To my family.

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Dr. Tie Liu, for his generous support and guidance. His patience and encouragement became the greatest support for me to finish my dissertation. The research skills and the logical thinking abilities I learned from him over the past years will be my best treasure throughout my life.

I would like to thank all my committee members, Dr. Krishna Narayanan, Dr. Le Xie and Dr. Anxiao(Andrew) Jiang, for giving me valuable suggestions and feedbacks on my dissertation.

I have been really lucky to have such wonderful lab mates, Hung D. Ly, Jinjing Jiang, Jae Won Yoo, Neeharika Marukala, Shuo Shao, Xiaopeng(Lucia) Sui, Li Xu, Paul McVay and Amir Salim. They all have been great workmates and friends. I would like to especially thank Yu-Chih(Jerry) Huang, whom has been the best workmate, guide and friend during my PhD career. I will never forget the days we worked together.

I also would like to thank Blake Bradley, Yue Deng, Kai Gao, Xiaoxi Zhang, Jian Shen, Chunyu Tian, Shuo Feng, Lizhuo Wang and all my friends in College Station, for making my PhD life an unforgettable journey.

Finally, and most importantly, I am truly thankful to my parents for their endless love and support, which helped me go through these years. I wish to thank Xinyi, for coming to my life, for her supporting and understanding.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Tie Liu [advisor], Professor Krishna Narayanan and Professor Le Xie of the Department of Electrical and Computer Engineering, and Professor Anxiao Jiang of the Department of Computer Science and Engineering.

Funding Sources

This work was made possible in part by NSF, under Grant No.1320237.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

1.1 Treating Interference as Noise for Interference Network

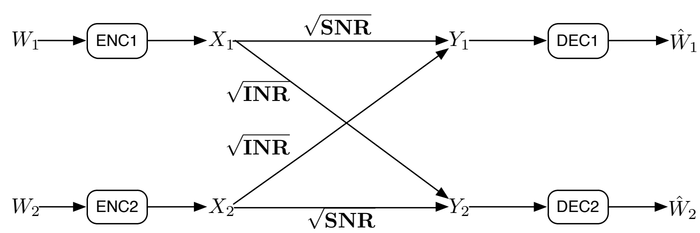


Figure 1.1: Two-user symmetric Gaussian interference channel

Interference management is one of the most critical issues in wireless communication network. Perhaps the simplest model to study interference management is the memoryless stationary two-user symmetric Gaussian interference channel, for which the channel outputs Y_1 and Y_2 at their respective users can be written as:

$$\mathsf{Y}_1 = \sqrt{\mathrm{SNR}}\mathsf{X}_1 + \sqrt{\mathrm{INR}}\mathsf{X}_2 + \mathsf{Z}_1 \quad (1.1)$$

$$\mathsf{Y}_2 = \sqrt{\mathrm{INR}}\mathsf{X}_1 + \sqrt{\mathrm{SNR}}\mathsf{X}_2 + \mathsf{Z}_2. \quad (1.2)$$

Here, X_1 and X_2 are channel inputs that are subject to a *unit* power constraint, and the noise Z_1 and Z_2 are *standard* Gaussian and are independent of the input signals. Since both the signal and the noise power in the above model are normalized to one, the quantities SNR and INR represent the signal-to-noise ratio and interference-to-noise ratio, respectively, for both users.

For the above two-user $(\mathrm{SNR}, \mathrm{INR})$ symmetric Gaussian interference channel, the following characterization on the *sum capacity* has been known since the early 1970s [1]:

$$C_{sum}(\mathrm{SNR}, \mathrm{INR}) = \lim_{k\to\infty} \sup \frac{1}{k}[I(\mathsf{X}_1^k; \mathsf{Y}_1^k) + I(\mathsf{X}_2^k; \mathsf{Y}_2^k)] \tag{1.3}$$

where for $i = 1, 2$, $\mathsf{X}_i^k := (\mathsf{X}_i(t) : t = 1, \ldots, k)$ and $\mathsf{Y}_i^k := (\mathsf{Y}_i(t) : t = 1, \ldots, k)$, and for each $k \in \mathcal{N}$, the supreme is over all possible *product* distributions on $(\mathsf{X}_1^k, \mathsf{X}_2^k)$ such that $E[(\frac{1}{k}\|\mathsf{X}_i^k\|^2)] \leq 1$ for $i = 1, 2$. Note that finding an explicit expression for the sum capacity via the above *limiting* characterization requires solving a sequence of optimization problems over input distributions. When $\mathrm{INR}(1 + \mathrm{INR})^2 \leq \frac{\mathrm{SNR}^2}{4}$, it has been recently shown [2–4] that *Gaussian* inputs with $k = 1$ can achieve the sum capacity. For the other parameter regimes, however, the problem of finding a sequence of input distributions to achieve the sum capacity remains open.

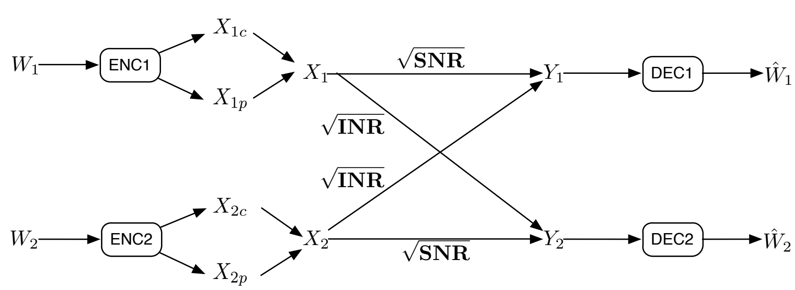


Figure 1.2: Han-Kobayashi scheme

Deviating from the limiting characterization, a different line of quests for an *explicit* characterization of the sum capacity focuses on the so-called Han-Kobayashi scheme [5]. These quests have led to a precise characterization of the sum capacity when $\mathrm{INR} \geq \mathrm{SNR}(1 + \mathrm{SNR})$ [6] and an *approximate* characterization of the sum capacity to within one bit for *all* parameter regimes [7]. The essential idea of Han-Kobayashi scheme is to split the message at each transmitter into a private message and a *common* message and the common message needs to be decodable at both the intended and the unintended receivers. By comparison, the communication scheme for achieving the limiting expression (1.3) is to simply treat the interference as adding to the noise floor at each receiver, which is very appealing from the engineering viewpoint due to its simplicity.

In their recent work [8] [9], Dytso *et al.* revisited the limiting characterization (1.3) for the two-user symmetric Gaussian interference channel and proposed *mixtures* of discrete and Gaussian distributions (with $k = 1$) for the channel inputs. Focusing on the high-SNR regime where $\mathrm{INR} = \mathrm{SNR}^{\alpha}$ for some *fixed* α and $\mathrm{SNR} \to \infty$, it was shown in [8] that the proposed input distributions can achieve the sum capacity to within $O(\log\log(\mathrm{SNR}))$ for any $\alpha > 0$.

Driven by the desire of understanding the rationale behind the proposed mixtures of discrete and Gaussian distributions and the mathematical nature of the $O(\log\log(\mathrm{SNR}))$ gap result [8], in this dissertation we consider a *two-step* approach for finding approximately optimal input distributions for the two-user symmetric Gaussian interference channel. First, we shall consider the limiting characterization for the two-user symmetric Avestimehr-Diggavi-Tse (ADT) linear deterministic interference channel [10] and find input distributions that can (approximately) achieve the sum capacity. Next, we shall translate, *systematically*, the so-found (approximately) optimal input distributions for the linear deterministic model to the Gaussian model and show that the translations can also achieve the sum capacity to within $O(1)$ for any $\alpha \neq 1$. Different from the mixtures of discrete and Gaussian distributions proposed in [8], the approximately optimal input distributions found via the ADT linear deterministic model is *purely* discrete for all parameter regimes.

We mention here that there has been a lot of success recently in using the ADT linear deterministic model [10] to find approximate characterizations of wireless network capacity. Most of these success relies on a translation from the linear deterministic model to the Gaussian model at a *scheme* level. By comparison, the proposed translation in this dissertation is at the *distribution* level and is more along the line of [11] for studying the fundamental limits of secret writing on dirty paper.

1.2 Clustering for Brain Network via Metric Learning

Brain is also a network. Some brain regions are functionally connected even they are in different places. One popular research topic recently is to analysis the functionality connectivity between brain regions with network analysis method [12] [13]. The basic unit in the brain network is voxel which represents a relatively small area of brain cells. Thanks to the Magnetic resonance imaging (MRI) technology, we can successively scan the Blood Oxygentation Level Dependent (BOLD) signal of each voxel to form a so-called functional-MRI(fMRI) time-series. Studies in [14] [15] [16] showed that when test participants were asked to relaxed and not thinking particular, the correlation between the acquired Resting-State fMRI time-series can reflect the functional connectivity among voxels.

A clustering algorithm (K-means [17], Normalized Cut [18] [19], Spectral Clustering [20] [19]) can then be chosen to group voxels with similar function together. However, since the time-series data is always noise-corrupted, the clustering result may not get close enough to the ground truth. Recently, [21] [22] proposed a large-margin supervised clustering approach to learn a Mahalanobis distance from the ground truth. The learned metric lead to a relatively better clustering result. In this dissertation, we first follow [21] [22] to discussed the performance of a supervised clustering algorithm for a synthetic fMRI time-series dataset and then proposed an iterative unsupervised clustering algorithm for the brain network since the ground truth is not always available.

1.3 Dissertation Outline

The rest of this dissertation is organized as follows. Chapter 2 presents our result on the optimality of treating interference as noise for the two user symmetric Gaussian interference channel, which was partly reported in [23]. Chapter 3 presents our discussion on the clustering for brain network with metric learning. The performance of our proposed unsupervised clustering algorithm is shown via computer simulations. Finally, in Chapter 4, we summarize our main contributions in this dissertation and discuss some possible future research directions.

2. TREATING INTERFERENCE AS NOISE FOR TWO-USER SYMMETRIC GAUSSIAN INTERFERENCE CHANNELS*

2.1 Two-user Symmetric ADT Linear Deterministic Interference Channel

2.1.1 Channel Model

For any $n, m \in \mathcal{N}$, let $q := \max(n, m)$ and S be a $q \times q$ *down-shift* matrix.

$$S = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} \tag{2.1}$$

A two-user symmetric (n, m) ADT linear deterministic interference channel is given by [10]:

$$\mathsf{Y}_1 = S^{q-n}\mathsf{X}_1 + S^{q-m}\mathsf{X}_2 \tag{2.2}$$

$$\mathsf{Y}_2 = S^{q-n}\mathsf{X}_2 + S^{q-m}\mathsf{X}_1 \tag{2.3}$$

where $\mathsf{X}_1, \mathsf{X}_2, \mathsf{Y}_1, \mathsf{Y}_2 \in \mathbb{F}_2^q$ and all matrix multiplications and additions are over $\mathbb{F}_2$. For the above ADT linear deterministic interference channel, an explicit characterization of the sum capacity of the channel is known [24] and is given by:

$$C_{sum}(n,m) = \begin{cases} 2(n-m), & \text{for } \alpha \in \left(0, \frac{1}{2}\right] \\ 2m, & \text{for } \alpha \in \left(\frac{1}{2}, \frac{2}{3}\right] \\ 2n-m, & \text{for } \alpha \in \left(\frac{2}{3}, 1\right) \\ m, & \text{for } \alpha \in (1,2) \\ 2n, & \text{for } \alpha \in [2,\infty), \end{cases} \tag{2.4}$$

where $\alpha := \frac{m}{n}$.

2.1.2 Main Results

The following result is *implicit* in [25].

Theorem 1. *For any* $n, m \in \mathcal{N}$ *such that* $\alpha \in \left(0, \frac{2}{3}\right] \cup [2,\infty)$*, there exists a product distribution on* $(\mathsf{X}_1, \mathsf{X}_2)$ *such that*

$$I(\mathsf{X}_1;\mathsf{Y}_1) + I(\mathsf{X}_2;\mathsf{Y}_2) = C_{sum}(n,m). \tag{2.5}$$

For any $n, m \in \mathcal{N}$ *such that* $\alpha \in \left(\frac{2}{3}, 1\right) \cup (1,2)$*, there exists a product distribution on* $(\mathsf{X}_1, \mathsf{X}_2)$ *such that*

$$I(\mathsf{X}_1;\mathsf{Y}_1) + I(\mathsf{X}_2;\mathsf{Y}_2) \geq C_{sum}(n,m) - 1 \tag{2.6}$$

We note here that for $\alpha \in \left(\frac{2}{3}, 1\right) \cup (1,2)$, even though the input distributions that we consider for Theorem 1 are *not* precisely optimal, translating them into the Gaussian model suffices to achieve the sum capacity to within $O(1)$.

The following result is new to the best of our knowledge.

Theorem 2. *For any* $n, m \in \mathcal{N}$ *such that* $\alpha \in \left(\frac{2}{3}, 1\right) \cup (1,2)$*, there exists a product* *distribution on* $(\mathsf{X}_1^2, \mathsf{X}_2^2)$ *such that*

$$\frac{1}{2}\left[I(\mathsf{X}_1^2;\mathsf{Y}_1^2) + I(\mathsf{X}_2^2;\mathsf{Y}_2^2)\right] = C_{sum}(n,m). \tag{2.7}$$

Whether $k \geq 2$ is *necessary* to achieve the sum capacity for $\alpha \in \left(\frac{2}{3}, 1\right) \cup (1,2)$ remains unknown.

2.1.3 Proof of Theorem 1

Fix $n, m \in \mathcal{N}$ such that $\alpha = \frac{m}{n} \neq 1$. Let $\mathsf{U}_1, \mathsf{U}_2$ be two i.i.d random vectors of length r with entries drawn independently and uniformly from $\mathbb{F}_2$. For $i = 1, 2$, let

$$\mathsf{X}_i = E\mathsf{U}_i \tag{2.8}$$

for some $E \in \mathbb{F}_2^{q,r}$. For the above symmetric choice of $(\mathsf{X}_1, \mathsf{X}_2)$, by defining $A = S^{q-n}$ and $B = S^{q-m}$, we have

$$\begin{aligned} & I(\mathsf{X}_2;\mathsf{Y}_2) \\ & \quad = I(\mathsf{X}_1;\mathsf{Y}_1) \\ & \quad = H(\mathsf{Y}_1) - H(\mathsf{Y}_1|\mathsf{X}_1) \\ & \quad = H(S^{q-n}E\mathsf{U}_1 + S^{q-m}E\mathsf{U}_2) - H(S^{q-m}E\mathsf{U}_2) \\ & \quad = rank([S^{q-n}E \; S^{q-m}E]) - rank(S^{q-m}E) \\ & \quad = rank([AE \; BE]) - rank(BE) \end{aligned} \tag{2.9}$$

where both ranks are evaluated in $\mathbb{F}_2$.

We now present our choice of E based on the following five different regimes of α.

Case 1: $\alpha \in (0, \frac{1}{2}]$ **(very weak interference channel).** Let $F_1 \in \mathbb{F}_2^{m,n}$ and $F_2 \in \mathbb{F}_2^{n-2m,n}$ such that all rows from F_1, F_2 are linearly independent. We can choose

$$E = \begin{bmatrix} F_1 \\ F_2 \\ 0_{m,n} \end{bmatrix}. \tag{2.10}$$

Then

$$\begin{aligned} & rank([AE\ BE]) \\ &= rank(\begin{bmatrix} F_1 & 0_{m,n} \\ F_2 & 0_{n-2m,n} \\ 0_{m,n} & F_1 \end{bmatrix}) \\ &= n, \end{aligned} \tag{2.11}$$

and

$$rank(BE) = rank(F_1) = m, \tag{2.12}$$

which leads to

$$\text{(2.9)} = n - m = \frac{1}{2}C_{sum}. \tag{2.13}$$

Case 2: $\alpha \in (\frac{1}{2}, \frac{2}{3})$ **(weak interference regime).** Let $F_1 \in \mathbb{F}_2^{n-m,n}$ and $F_2 \in \mathbb{F}_2^{2m-n,n}$

such that $[F_1^T, F_2^T]^T$ is full-ranked. We can choose

$$E = \begin{bmatrix} F_1 \\ 0_{n-m,n} \\ F_2 \end{bmatrix}. \tag{2.14}$$

Then

$$\begin{aligned} & rank([AE\ BE]) \\ & = rank(\begin{bmatrix} F_1 & 0_{n-m,n} \\ 0_{n-m,n} & F_1 \\ F_2 & 0_{2m-n,n} \end{bmatrix}) \\ & = n, \end{aligned} \tag{2.15}$$

and

$$rank(BE) = rank(F_1) = n - m, \tag{2.16}$$

which leads to

$$\text{(2.9)} = n - (n-m) = m = \frac{1}{2}C_{sum}. \tag{2.17}$$

Case 3: $\alpha \in \left(\frac{2}{3}, 1\right)$ **(moderate interference regime).** For this regime, we shall consider *sub-regimes* $\alpha \in \left[\frac{2l}{2l+1}, \frac{2l+2}{2l+3}\right)$ for $l \in \mathcal{N}$. For any *fixed* $l \in \mathcal{N}$, let $F_1, F_3, \ldots, F_{2l-1} \in \mathbb{F}_2^{\lfloor \frac{2l+1}{2}m \rfloor - ln, n}$, $F_2, F_4, \ldots, F_{2l-2} \in \mathbb{F}_2^{(l+1)n - \lfloor \frac{2l+3}{2}m \rfloor, n}$, and $F_{2l}, F_{2l+1} \in \mathbb{F}_2^{n-m,n}$ such that all rows from $F_1, F_2, \ldots, F_{2l+1}$ are linearly independent.

We shall choose

$$E = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_{2l-2} \\ F_{2l-1} \\ F_{2l} \\ F_{2l-1} \\ F_{2l-2} \\ \vdots \\ F_3 \\ F_2 \\ F_1 \\ 0_{n-m,n} \\ F_{2l+1} \end{bmatrix}. \tag{2.18}$$

Then

$$rank([AE\ BE]) = rank(\begin{bmatrix} F_1 & 0_{\lfloor \frac{2l+1}{2}m \rfloor - ln, n} \\ F_2 & 0_{(l+1)n - \lfloor \frac{2l+3}{2}m \rfloor, n} \\ F_3 & F_1 \\ \vdots & \\ F_{2l-2} & F_{2l-4} \\ F_{2l-1} & F_{2l-3} \\ \hline F_{2l} & \left\{ \begin{array}{l} F_{2l-2} \\ F_{2l-1} \end{array} \right. \\ \hline \left. \begin{array}{l} F_{2l-1} \\ F_{2l-2} \end{array} \right\} & F_{2l} \\ \hline F_{2l-3} & F_{2l-1} \\ \vdots & \vdots \\ F_3 & F_5 \\ F_2 & F_4 \\ F_1 & F_3 \\ 0_{(l+1)n - \lfloor \frac{2l+3}{2}m \rfloor, n} & F_2 \\ 0_{\lfloor \frac{2l+1}{2}m \rfloor - ln, n} & F_1 \\ F_{2l+1} & 0_{n-m,n} \end{bmatrix}) \qquad (2.19)$$

$$
(2.19) = rank(\begin{bmatrix} F_1 & 0 \\ F_2 & 0 \\ F_3 & 0 \\ \vdots & \\ F_{2l-2} & 0 \\ F_{2l-1} & 0 \\ F_{2l} & 0 \\ 0 & F_{2l} \\ 0 & F_{2l-1} \\ \vdots & \vdots \\ 0 & F_5 \\ 0 & F_4 \\ 0 & F_3 \\ 0 & F_2 \\ 0 & F_1 \\ F_{2l+1} & 0 \end{bmatrix}) = \begin{cases} n, & m \text{ is even} \\ n-1, & m \text{ is odd} \end{cases} \tag{2.20}
$$

Note that the solid lines here indicate that the submatrix between them have same numbers of rows.

Remark 1. *(Explanation on $n-1$) Although most of the sub matrices have the floor operation on their size, the number of rows of $[F_{odd}^T, F_{even}^T]^T$ is fixed. For example, when m is odd, F_1 has $\lfloor \frac{2l+1}{2} m \rfloor - ln$ rows which will loss $\frac{1}{2}$, F_2 has $(l+1)n - \lfloor \frac{2l+3}{2} m \rfloor$ rows which will gain $\frac{1}{2}$, so that $[F_1^T, F_2^T]^T$ has $n-m$ rows together. After counting the number of sub matrices, only $[F_{2l-1}\ 0]$ and $[0\ F_{2l-1}]$ has "half bit loss" which combine a 1 bit loss in total.*

Then

$$rank(BE) = rank\left(\begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_{2l-2} \\ F_{2l-1} \\ F_{2l} \\ F_{2l-1} \\ F_{2l-2} \\ \vdots \\ F_3 \\ F_2 \\ F_1 \end{bmatrix}\right) = \begin{cases} \frac{1}{2}m, m \text{ is even} \\ \frac{1}{2}m - \frac{1}{2}, m \text{ is odd} \end{cases} \tag{2.21}$$

which leads to

$$(2.9) = \begin{cases} n - \frac{1}{2}m & = \frac{1}{2}C_{sum} \quad , m \text{ is even} \\ n - \frac{1}{2}m - \frac{1}{2} & = \frac{1}{2}C_{sum} - \frac{1}{2}, m \text{ is odd} \end{cases} \geq \frac{1}{2}C_{sum} - \frac{1}{2} \tag{2.22}$$

Case 4: $\alpha \in (1, 2)$ **(strong interference regime).** For this regime, we shall consider *sub-regimes* $\alpha \in \left(\frac{2l+2}{2l+1}, \frac{2l}{2l-1}\right]$ for $l \in \mathcal{N}$. For any ***fixed*** $l \in \mathcal{N}$, let $F_1, F_2, \ldots F_{l-1} \in \mathbb{F}_2^{m-n,m}$, $F_l, F_{l+1} \in \mathbb{F}_2^{ln-\lceil \frac{2l-1}{2}m \rceil, m}$, and $F_{l+2} \in \mathbb{F}_2^{\lceil \frac{2l+1}{2}m \rceil - (l+1)n, m}$ such that all rows from $F_1, F_2, \ldots, F_{l+2}$ are linearly independent.

We shall choose

$$E = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_l \\ F_{l+1} \\ F_{l+2} \\ F_l \\ F_{l-1} \\ \vdots \\ F_2 \\ F_1 \end{bmatrix}. \tag{2.23}$$

Then

$$rank([AE\ BE]) = rank(\begin{bmatrix} 0_{m-n,m} & F_1 \\ F_1 & F_2 \\ \vdots & \vdots \\ F_{l-2} & F_{l-1} \\ \hline \left.\begin{matrix} F_{l-1} \\ F_l \end{matrix}\right\} & \left\{\begin{matrix} F_l \\ F_{l+1} \\ F_{l+2} \end{matrix}\right. \\ \hline F_{l+1} & F_l \\ \hline \left.\begin{matrix} F_{l+2} \\ F_l \end{matrix}\right\} & F_{l-1} \\ \hline F_{l-1} & F_{l-2} \\ \vdots & \vdots \\ F_3 & F_2 \\ F_2 & F_1 \\ F_1 & 0_{m-n,m} \end{bmatrix}) \tag{2.24}$$

$$
(2.24) = rank\left(\begin{bmatrix}
0 & F_1 \\
0 & F_2 \\
\vdots & \vdots \\
0 & F_{l-1} \\
\hline
0 & \left\{\begin{matrix} F_l \\ F_{l+1} \\ F_{l+2} \end{matrix}\right. \\
\hline
F_{l+1} & 0 \\
\hline
\left.\begin{matrix} F_{l+2} \\ F_l \end{matrix}\right\} & 0 \\
\hline
F_{l-1} & 0 \\
\vdots & \vdots \\
F_3 & 0 \\
F_2 & 0 \\
F_1 & 0
\end{bmatrix}\right) = \begin{cases} m, & m \text{ is even} \\ m-1, & m \text{ is odd} \end{cases} \tag{2.25}
$$

Note that the solid lines here indicates that the submatrix between them have same numbers of rows. Similar to the previous regime, this time, only $[0 \; F_l]$ and $[F_{l+1} \; 0]$ have 'half bit loss' which combines to 1 in total.

and

$$rank(BE) = rank(\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_l \\ F_{l+1} \\ F_{l+2} \\ F_l \\ F_{l-1} \\ \vdots \\ F_2 \\ F_1 \end{bmatrix}) = \begin{cases} \frac{1}{2}m, m \text{ is even} \\ \frac{1}{2}m - \frac{1}{2}, m \text{ is odd} \end{cases} \tag{2.26}$$

which leads to

$$\text{(2.9)} = \begin{cases} \frac{1}{2}m & = \frac{1}{2}C_{sum} \quad , m \text{ is even} \\ \frac{1}{2}m - \frac{1}{2} & = \frac{1}{2}C_{sum} - \frac{1}{2}, m \text{ is odd} \end{cases} \geq \frac{1}{2}C_{sum} - \frac{1}{2}, \tag{2.27}$$

Case 5: $\alpha \in [2, \infty)$ **(Very strong interference regime)**. Let $F_1 \in \mathbb{F}_2^{n,n}$ be full-ranked, we can choose

$$E = \begin{bmatrix} F_1 \\ 0_{m-n,n} \end{bmatrix}. \tag{2.28}$$

Then

$$rank([AE\ BE]) = rank(\begin{bmatrix} 0_{n,m} & F_1 \\ 0_{m-2n,n} & 0_{m-2n,n} \\ F_1 & 0_{n,n} \end{bmatrix}) \\ = 2n, \tag{2.29}$$

and

$$rank(BE) = rank(F_1) = n, \tag{2.30}$$

which leads to

$$(2.9) = 2n - n = n = \frac{1}{2}C_{sum}. \tag{2.31}$$

This complete the proof of Theorem 1.

2.1.4 Proof of Theorem 2

For notational simplicity, we shall consider the channel inputs X_1^2 and X_2^2 (over two channel uses) as length-$2q$ vectors in $\mathbb{F}_2$ (where the first q components correspond to the first channel use and the last q components correspond to the second channel use which are separated by dashed line). As before, let $\mathsf{U}_1, \mathsf{U}_2$ be two i.i.d random vectors of length r with entries drawn independently and uniformly from $\mathbb{F}_2$. For $i = 1, 2$, let

$$\mathsf{X}_i = E\mathsf{U}_i \tag{2.32}$$

for some $E \in \mathbb{F}_2^{2q,r}$.

For $\alpha \in \left(\frac{2}{3}, 1\right)$, we shall consider sub-regimes $\alpha \in \left(\frac{l}{l+1}, \frac{l+1}{l+2}\right]$ for $l \in \mathcal{N} - \{1\}$. Our choice of E depends on the *parity* of l. When l is even, let $F_1, F_3, F_7 \in \mathbb{F}_2^{(l+1)m-ln,2n-m}$, $F_2, F_6 \in \mathbb{F}_2^{(l+1)n-(l+2)m,2n-m}$, and $F_4, F_5, G_i, H_i \in \mathbb{F}_2^{n-m,2n-m}$ where $i = 1, \ldots, \frac{l}{2} - 1$ such that all rows from $G_1, \ldots, G_{\frac{l}{2}-1}, H_1, \ldots, H_{\frac{l}{2}-1}, F_1, \ldots, F_7$ are linearly independent. We shall choose

$$E = \begin{bmatrix} G_1 \\ 0_{n-m,2n-m} \\ \vdots \\ G_{\frac{l-2}{2}} \\ 0_{n-m,2n-m} \\ F_1 \\ F_2 \\ \boxed{F_3} \\ 0_{n-m,2n-m} \\ F_4 \\ ----- \\ H_1 \\ 0_{n-m,2n-m} \\ \vdots \\ H_{\frac{l-2}{2}} \\ 0_{n-m,2n-m} \\ F_5 \\ 0_{n-m,2n-m} \\ \boxed{F_3} \\ F_6 \\ F_7 \end{bmatrix}. \tag{2.33}$$

Note that the block F_3 (highlighted with surrounding boxes) is used in both channel uses.

When l is odd, let $F_1, F_3, F_5, F_9 \in \mathbb{F}_2^{(l+1)m-ln,2n-m}$, $F_2, F_4, F_8 \in \mathbb{F}_2^{(l+1)n-(l+2)m,2n-m}$, and $F_6, F_7, G_i, H_i \in \mathbb{F}_2^{n-m,2n-m}$ where $i = 1, \ldots, \frac{l-3}{2}$ such that all rows from $G_1, \ldots, G_{\frac{l-3}{2}}$,

$H_1, \ldots, H_{\frac{l-3}{2}}, F_1, \ldots, F_9$ are linearly independent. We shall choose

$$E = \begin{bmatrix} G_1 \\ 0_{n-m,2n-m} \\ \vdots \\ G_{\frac{l-3}{2}} \\ 0_{n-m,2n-m} \\ F_1 \\ F_2 \\ \boxed{F_3} \\ \boxed{F_4} \\ F_5 \\ 0_{n-m,2n-m} \\ F_6 \\ ----- \\ H_1 \\ 0_{n-m,2n-m} \\ \vdots \\ H_{1\frac{l-3}{2}} \\ 0_{n-m,2n-m} \\ F_7 \\ 0_{n-m,2n-m} \\ \boxed{F_3} \\ \boxed{F_4} \\ 0_{4m-3n,2n-m} \\ F_8 \\ F_9 \end{bmatrix}. \tag{2.34}$$

For $\alpha \in (1,2)$, we shall consider the sub-regimes $\alpha \in \left[\frac{l+2}{l+1}, \frac{l+1}{l}\right)$ for $l \in \mathcal{N}$. Our choice of E again depends on the parity of l. When l odd, let $F_1, F_3, F_4 \in \mathbb{F}_2^{(l+1)n-lm,m}$, $F_2, F_5 \in \mathbb{F}_2^{(l+1)m-(l+2)n,m}$, and $G_i, H_i \in \mathbb{F}_2^{m-n,m}$ such that all rows from $G_1, \ldots, \ldots, G_{\frac{l-1}{2}}, H_1, \ldots,$

$H_{\frac{l-1}{2}}, F_1, \ldots, F_5$ are linearly independent. We shall choose

$$E = \begin{bmatrix} G_1 \\ 0_{m-n,m} \\ \vdots \\ G_{\frac{l-1}{2}} \\ 0_{m-n,m} \\ \boxed{F_1} \\ F_2 \\ F_3 \\ 0_{m-n,m} \\ ---- \\ H_1 \\ 0_{m-n,m} \\ \vdots \\ H_{\frac{l-1}{2}} \\ 0_{m-n,m} \\ F_4 \\ F_5 \\ \boxed{F_1} \\ 0_{m-n,m} \end{bmatrix}. \tag{2.35}$$

When l is even, let $F_1, F_4, F_5, F_7 \in \mathbb{F}_2^{(l+1)n-lm,m}$, $F_2, F_3, F_6 \in \mathbb{F}_2^{(l+1)m-(l+2)n,m}$, and $G_i, H_i \in \mathbb{F}_2^{m-n,m}$ where $i = 1, \ldots, \frac{l}{2} - 1$ such that all rows from $G_1, \ldots, G_{\frac{l}{2}-1}, H_1, \ldots,$

$H_{\frac{l}{2}-1}, F_1, \ldots, F_7$ are linearly independent. We shall choose

$$E = \begin{bmatrix} G_1 \\ 0_{m-n,m} \\ \vdots \\ G_{\frac{l}{2}-1} \\ 0_{m-n,m} \\ F_1 \\ \boxed{F_2} \\ 0_{(l+1)n-lm} \\ F_3 \\ F_4 \\ 0_{m-n,m} \\ ----- \\ H_1 \\ 0_{m-n,m} \\ \vdots \\ H_{\frac{l-1}{2}-1} \\ 0_{m-n,m} \\ F_5 \\ F_6 \\ 0_{(l+1)n-lm} \\ \boxed{F_2} \\ F_7 \\ 0_{m-n,m} \end{bmatrix}. \tag{2.36}$$

With the above choices of E, we have

$$\frac{1}{2}\left[I(\mathsf{X}_1^2;\mathsf{Y}_1^2) + I(\mathsf{X}_2^2;\mathsf{Y}_2^2)\right] = C_{sum}(n,m)$$

for any $\alpha \in \left(\frac{2}{3}, 1\right) \cup (1,2)$. The proof is omitted for brevity.

2.2 Two-user Symmetric Gaussian Interference Channel

2.2.1 Main Result

Theorem 3. *For the two-user symmetric* $(\mathrm{SNR}, \mathrm{INR})$ *Gaussian interference channel with* $\mathrm{INR} = \mathrm{SNR}^{\alpha}$ *and* $\mathrm{SNR} \to \infty$*, there exists a product distribution on* $(\mathsf{X}_1, \mathsf{X}_2)$ *with purely discrete marginals such that*

$$I(\mathsf{X}_1; \mathsf{Y}_1) + I(\mathsf{X}_2; \mathsf{Y}_2) \geq C_{sum}(\mathrm{SNR}, \mathrm{SNR}^{\alpha}) - O(1) \tag{2.37}$$

for any $\alpha > 0$ *such that* $\alpha \neq 1$*.*

2.2.2 Proof of Theorem 3

To translate the (approximately) optimal input distributions proposed for the ADT linear deterministic model to the Gaussian model, we shall view the inputs to the linear deterministic model as *binary representations* of the inputs to the Gaussian model. More specifically, each distinct block F_j translates into two i.i.d. random variables $\mathsf{F}_{1,j}$ and $\mathsf{F}_{2,j}$ that are uniform over a PAM constellation. The size of the PAM constellation and the power level of $\mathsf{F}_{1,j}$ and $\mathsf{F}_{2,j}$ are determined by the number of rows of F_j and the relative position of F_j in E, respectively.

To make the correspondence between the linear deterministic model and the Gaussian model more explicit, let us define

$$n := \frac{1}{2}\log \mathrm{SNR} \tag{2.38}$$

$$m := \frac{1}{2}\log \mathrm{INR} = \frac{1}{2}\log \mathrm{SNR}^{\alpha} = n\alpha. \tag{2.39}$$

(We therefore have $n \to \infty$ and hence $m \to \infty$ for any fixed $\alpha > 0$ in the limit as

$\mathrm{SNR} \to \infty$.) With the above definition of m, n, we can rewrite the upper bound of the sum capacity in [7]

$$C_{sum}(\mathrm{SNR}, \mathrm{INR}) \leq \begin{cases} \log(\frac{1+2^{2n}}{1+2^{2m}}), & \text{for } \alpha \in \left(0, \frac{1}{2}\right] \\ \log(1 + 2^{2m} + \frac{2^{2n}}{1+2^{2m}}), & \text{for } \alpha \in \left(\frac{1}{2}, \frac{2}{3}\right] \\ \frac{1}{2}\log\left(\frac{(1+2^{2n})(1+2^{2m}+2^{2n})}{1+2^{2m}}\right), & \text{for } \alpha \in \left(\frac{2}{3}, 1\right) \\ \frac{1}{2}\log\left(1 + 2^{2m} + 2^{2n}\right), & \text{for } \alpha \in (1, 2) \\ \log\left(1 + 2^{2n}\right), & \text{for } \alpha \in [2, \infty) \end{cases}, \tag{2.40}$$

where $\alpha := \frac{m}{n}$.

We first introduce the following proposition from [9], with the original proof from [26].

Proposition 1. *Let X_D be a discrete random variable with minimum distance $d_{min(X_D)} > 0$. Let Z be a zero-mean unit-variance random variable independent of X_D. Then*

$$I(X_D; Y) \geq H(X_D) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(1 + \frac{12}{d^2_{min(X_D)}}\right) \tag{2.41}$$

Proof. See Appendix A. □

We now present the proof of Theorem 3 based on the following five different regimes of α.

Case 1: $\alpha \in (0, \frac{1}{2}]$. In this regime, $F_1 \in \mathbb{F}_2^{m,n}$ translates into two i.i.d random variables $F_{1,1}$ and $F_{2,1}$ that are uniform over PAM(2^m) and $F_2 \in \mathbb{F}_2^{n-2m,n}$ translates into two i.i.d random variables $F_{1,2}$ and $F_{2,2}$ that are uniform over PAM(2^{n-2m}).

Then the choice

$$E = \begin{bmatrix} F_1 \\ F_2 \\ 0_{m,n} \end{bmatrix}. \tag{2.42}$$

translates into the following input distributions for the Gaussian model

$$\begin{aligned} \mathsf{X}_i &= 2^{-n}(2^{m+(n-2m)}\mathsf{F}_{i,1} + 2^m \mathsf{F}_{i,2}) \\ &= 2^{-n}(2^{n-m}\mathsf{F}_{i,1} + 2^m \mathsf{F}_{i,2}), \end{aligned} \tag{2.43}$$

for $i = 1, 2$ where all four random variables $\mathsf{F}_{1,1}, \mathsf{F}_{2,1}, \mathsf{F}_{1,2}, \mathsf{F}_{2,2}$ are assumed to be independent of each other. Note that the terms m and $n-2m$ represent the numbers of rows of the blocks $0_{m,n}$ and F_2 in E respectively, and the term 2^{-n} is chosen such that $E[\mathsf{X}_i^2] \leq 1$ for $i = 1, 2$.

The mutual information between the channel inputs and outputs can be calculated as follows

$$\begin{aligned} I(\mathsf{X}_2; \mathsf{Y}_2) &= I(\mathsf{X}_1; \mathsf{Y}_1) \\ &= I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1) - I(2^m\mathsf{X}_2; 2^m\mathsf{X}_2 + \mathsf{Z}_1). \end{aligned} \tag{2.44}$$

To bound from below the mutual information $I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1)$, note that $2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 = 2^{n-m}\mathsf{F}_{1,1} + 2^m\mathsf{F}_{1,2} + \mathsf{F}_{2,1} + 2^{2m-n}\mathsf{F}_{2,2}$.

For $\alpha \in (0, \frac{1}{2}]$, the last term $2^{2m-n}\mathsf{F}_{2,2}$ on the righthand side satisfies $E[(2^{2m-n}\mathsf{F}_{2,2})^2] \to 0$ in the limit as $n \to \infty$. Thus it is below the "noise floor", and we rewrite

$$2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 = 2^n\mathsf{X}_A + 2^{2m-n}\mathsf{F}_{2,2}, \tag{2.45}$$

where

$$\mathsf{X}_A = 2^{n-m}\mathsf{F}_{1,1} + 2^m\mathsf{F}_{1,2} + \mathsf{F}_{2,1}. \tag{2.46}$$

By the mutual independence among X_A, $\mathsf{F}_{2,2}$ and Z_1, we have

$$\begin{aligned} &I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1) \\ &= h(\mathsf{X}_A + 2^{m-n}\mathsf{F}_{2,2} + \mathsf{Z}_1) - h(\mathsf{Z}_1) \\ &\geq h(\mathsf{X}_A + \mathsf{Z}_1) - h(\mathsf{Z}_1) \\ &= I(\mathsf{X}_A; \mathsf{X}_A + \mathsf{Z}_1). \end{aligned} \tag{2.47}$$

To further bound from below the mutual information $I(\mathsf{X}_A; \mathsf{X}_A + \mathsf{Z}_1)$, we shall use the following simple fact on X_A:

Fact 1. X_A *is uniform over a discrete constellation of size* $N(\mathsf{X}_A) = 2^n$ *with minimum distance* $d_{min}(\mathsf{X}_A) = 1$.

Combine the Fact 1 and Proposition 1 we have

$$\begin{aligned} I(\mathsf{X}_A;\mathsf{X}_A+\mathsf{Z}_1) &\geq H(X_A) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(1+\frac{12}{d^2_{min(X_A)}}\right) \\ &= n\log 2 - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log 13 \end{aligned} \tag{2.48}$$

To bound from above the mutual information $I(2^m\mathsf{X}_2; 2^m\mathsf{X}_2+\mathsf{Z}_1)$, note that

$$2^m\mathsf{X}_2 = \mathsf{F}_{2,1} + 2^{2m-n}\mathsf{F}_{2,2}, \tag{2.49}$$

where the last term $2^{2m-n}\mathsf{F}_{2,2}$ on the right-hand side is below the noise floor as before. We thus have

$$\begin{aligned} & I(2^m\mathsf{X}_2; 2^m\mathsf{X}_2+\mathsf{Z}_1) \\ &= h(2^m\mathsf{X}_2+\mathsf{Z}_1) - h(\mathsf{Z}_1) \\ &= h(2^m\mathsf{X}_2+\mathsf{Z}_1) - h(2^{2m-n}\mathsf{F}_{2,2}+\mathsf{Z}_1) + [h(2^{2m-n}\mathsf{F}_{2,2}+\mathsf{Z}_1) - h(\mathsf{Z}_1)] \\ &= I(F_{2,1}; 2^m\mathsf{X}_2+\mathsf{Z}_1) + I(2^{2m-n}\mathsf{F}_{2,2}; 2^{2m-n}\mathsf{F}_{2,2}+\mathsf{Z}_1) \\ &\leq H(F_{2,1}) + \frac{1}{2}\log[1+E[(2^{2m-n}\mathsf{F}_{2,2})^2]] \\ &\leq m\log 2 + \frac{1}{2}\log 2. \end{aligned} \tag{2.50}$$

Collecting results from above we have

$$\begin{aligned} I(\mathsf{X}_1;\mathsf{Y}_1)+I(\mathsf{X}_2;\mathsf{Y}_2) &\geq 2(n-m)\log 2-\log\left(\frac{2\pi e}{12}\right)-\log 13-\log 2 \\ &= 2(n-m)-c_1, \end{aligned} \tag{2.51}$$

where c_1 is a constant.

Case 2: $\alpha \in (\frac{1}{2}, \frac{2}{3}]$. In this regime, $F_1 \in \mathbb{F}_2^{n-m,n}$ translates into two i.i.d random variables $F_{1,1}$ and $F_{2,1}$ that are uniform over PAM(2^{n-m}) and $F_2 \in \mathbb{F}_2^{2m-n,n}$ translates into two i.i.d random variables $F_{1,2}$ and $F_{2,2}$ that are uniform over PAM(2^{2m-n}) . Then the choice

$$E=\begin{bmatrix} F_1 \\ 0_{n-m,n} \\ F_2 \end{bmatrix} \tag{2.52}$$

translates into the following input distributions for the Gaussian model

$$\begin{aligned} \mathsf{X}_i &= 2^{-n}(2^{(n-m)+(2m-n)}\mathsf{F}_{i,1}+\mathsf{F}_{i,2}) \\ &= 2^{-n}(2^m\mathsf{F}_{i,1}+\mathsf{F}_{i,2}), \end{aligned} \tag{2.53}$$

for $i=1,2$ where all four random variables $\mathsf{F}_{1,1},\mathsf{F}_{2,1},\mathsf{F}_{1,2},\mathsf{F}_{2,2}$ are assumed to be independent of each other. Note that the terms $n-m$ and $2m-n$ represent the numbers of rows of the blocks $0_{n-m,n}$ and F_2 in E respectively, and the term 2^{-n} is chosen such that $E[\mathsf{X}_i^2] \leq 1$ for $i=1,2$.

As (2.44) in Case 1, to bound from below the mutual information $I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2;\ 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1)$, note that

$$2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 = 2^m\mathsf{F}_{1,1} + 2^{2m-n}\mathsf{F}_{2,1} + \mathsf{F}_{1,2} + 2^{m-n}\mathsf{F}_{2,2}. \tag{2.54}$$

For $\alpha \in (\frac{1}{2}, \frac{2}{3}]$, the last term $2^{m-n}\mathsf{F}_{2,2}$ on the righthand side satisfies $E[(2^{m-n}\mathsf{F}_{2,2})^2] \to 0$ in the limit as $n \to \infty$. Thus it is below the "noise floor", and we rewrite

$$2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 = 2^n\mathsf{X}_B + 2^{m-n}\mathsf{F}_{2,2}, \tag{2.55}$$

where

$$\mathsf{X}_B = 2^m\mathsf{F}_{1,1} + 2^{2m-n}\mathsf{F}_{2,1} + \mathsf{F}_{1,2}. \tag{2.56}$$

By the mutual independence among X_B, $\mathsf{F}_{2,2}$ and Z_1, as (2.47) in Case 1, we can also claim that

$$I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1) \geq I(\mathsf{X}_B; \mathsf{X}_B + \mathsf{Z}_1) \tag{2.57}$$

and a simple fact on X_B:

Fact 2. X_B *is uniform over a discrete constellation of size* $N(\mathsf{X}_B) = 2^n$ *with minimum distance* $d_{min}(\mathsf{X}_B) = 1$.

Combine the Fact 2 and Proposition 1 we have

$$I(\mathsf{X}_B;\mathsf{X}_B+\mathsf{Z}_1) \geq H(\mathsf{X}_B) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(1+\frac{12}{d^2_{min(X_B)}}\right)$$
$$= n\log 2 - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log 13 \quad (2.58)$$

To bound from above the mutual information $I(2^m\mathsf{X}_2; 2^m\mathsf{X}_2+\mathsf{Z}_1)$, note that

$$2^m\mathsf{X}_2 = 2^{2m-n}\mathsf{F}_{2,1} + 2^{m-n}\mathsf{F}_{2,2}, \quad (2.59)$$

where the last term $2^{m-n}\mathsf{F}_{2,2}$ on the right-hand side is below the noise floor as before. We thus have

$$I(2^m\mathsf{X}_2; 2^m\mathsf{X}_2+\mathsf{Z}_1)$$
$$\leq H(2^{2m-n}\mathsf{F}_{2,1}) + \frac{1}{2}\log[1+E[(2^{m-n}\mathsf{F}_{2,2})^2]]$$
$$\leq (n-m)\log 2 + \frac{1}{2}\log 2. \quad (2.60)$$

Collecting results from above we have

$$I(\mathsf{X}_1;\mathsf{Y}_1) + I(\mathsf{X}_2;\mathsf{Y}_2) \geq 2m\log 2 - \log\left(\frac{2\pi e}{12}\right) - \log 13 - \log 2$$
$$= 2m - c_2, \quad (2.61)$$

where c_2 is a constant.

Case 3: $\alpha \in (\frac{2}{3}, 1)$. In this regime, consider *sub-regimes* $\alpha \in \left[\frac{2l}{2l+1}, \frac{2l+2}{2l+3}\right)$ for

$l = 1, 2, 3, \ldots$, $F_1, F_3, \ldots, F_{2l-1} \in \mathbb{F}_2^{\lfloor \frac{2l+1}{2} m \rfloor - ln, n}$ translates into $2l$ i.i.d random variables $\mathsf{F}_{1,1}, \mathsf{F}_{1,3}, \ldots, \mathsf{F}_{1,2l-1}$ and $\mathsf{F}_{2,1}, \mathsf{F}_{2,3}, \ldots, \mathsf{F}_{2,2l-1}$ that are uniform over PAM$(2^{\frac{2l+1}{2} m - ln - 1})$, $F_2, F_4, \ldots, F_{2l-2} \in \mathbb{F}_2^{(l+1)n - \lfloor \frac{2l+3}{2} m \rfloor, n}$ translates into $2(l-1)$ i.i.d random variables $\mathsf{F}_{1,2}$, $\mathsf{F}_{1,4}, \ldots, \mathsf{F}_{1,2l-2}$ and $\mathsf{F}_{2,2}, \mathsf{F}_{2,4}, \ldots, \mathsf{F}_{2,2l-2}$ that are uniform over PAM$(2^{(l+1)n - \frac{2l+3}{2} m - 1})$ and $F_{2l} \in \mathbb{F}_2^{n-m, n}$ translates into 2 i.i.d random variables $\mathsf{F}_{1,2l}$ and $\mathsf{F}_{2,2l}$ that are uniform over PAM(2^{n-m-1}) and $F_{2l+1} \in \mathbb{F}_2^{n-m,n}$ translates into 2 i.i.d random variables $\mathsf{F}_{1,2l+1}$ and $\mathsf{F}_{2,2l+1}$ that are uniform over PAM(2^{n-m}). The -1 is chosen to avoid the carryover between variables. Then the choice

$$E = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_{2l-2} \\ F_{2l-1} \\ F_{2l} \\ F_{2l-1} \\ F_{2l-2} \\ \vdots \\ F_3 \\ F_2 \\ F_1 \\ 0_{n-m,n} \\ F_{2l+1} \end{bmatrix}. \tag{2.62}$$

translates into the following input distributions for the Gaussian model

$$\begin{aligned}
\mathsf{X}_i = 2^{-n}\Bigg[&\sum_{j=1}^{l-1} \left[2^{(l-j+2)n-\frac{2l-2j+3}{2}m}\mathsf{F}_{i,2j-1} + 2^{jm-(j-1)n}\mathsf{F}_{i,2j}\right] \\
&+ 2^{2n-\frac{3}{2}m}\mathsf{F}_{i,2l-1} + 2^{n-\frac{1}{2}m}\mathsf{F}_{i,2l} + 2^{(l+1)n-(l+1)m}\mathsf{F}_{i,2l-1} \\
&+ \sum_{k=l-1}^{1} \left[2^{\frac{2l-2k-1}{2}m-(l-k-1)n}\mathsf{F}_{i,2k} + 2^{(k+1)n-(k+1)m}\mathsf{F}_{i,2k-1}\right] + \mathsf{F}_{i,2l+1}\Bigg], \qquad (2.63)
\end{aligned}$$

for $i = 1, 2$ where all random variables are assumed to be independent of each other. $2^{-(n-n\epsilon_n)}$ is chosen such that $E[\mathsf{X}_i^2] \leq 1$ for $i = 1, 2$.

As (2.44) in Case 1, to bound from below the mutual information $I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1)$, note that

$$\begin{aligned}
&2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 \\
&= 2^{-n}\Bigg[2^{(l+2)n-\frac{2l+1}{2}m+n\epsilon_n}\mathsf{F}_{1,1} + 2^{m+n+n\epsilon_n}\mathsf{F}_{1,2} \\
&+ \sum_{j=2}^{l-1}\left[2^{(l-j+3)n-\frac{2l-2j+3}{2}m}(\mathsf{F}_{1,2j-1} + \mathsf{F}_{2,2j-3}) + 2^{jm-(j-2)n}(\mathsf{F}_{1,2j} + \mathsf{F}_{2,2j-2})\right] \\
&+ 2^{3n-\frac{3}{2}m}(\mathsf{F}_{1,2l-1} + \mathsf{F}_{2,2l-3}) \\
&+ 2^{2n-\frac{1}{2}m}(\mathsf{F}_{1,2l} + 2^{\frac{2l+1}{2}m-ln}\mathsf{F}_{2,2l-2} + \mathsf{F}_{2,2l-1}) \\
&+ 2^{n+\frac{1}{2}m}(\mathsf{F}_{2,2l} + 2^{(l+1)n-\frac{2l+3}{2}m}\mathsf{F}_{1,2l-1} + \mathsf{F}_{1,2l-2}) \\
&+ 2^{(l+1)n-lm}(\mathsf{F}_{2,2l-1} + \mathsf{F}_{1,2l-3}) \\
&+ \sum_{k=l-1}^{2}\left[2^{\frac{2l-2k+1}{2}m-(l-k-1)n}(\mathsf{F}_{2,2k} + \mathsf{F}_{1,2k-2}) + 2^{(k+1)n-km}(\mathsf{F}_{2,2k-1} + \mathsf{F}_{1,2k-3})\right] \\
&+ 2^{\frac{2l-1}{2}m-(l-2)n}\mathsf{F}_{2,2} + 2^{2n-m}\mathsf{F}_{2,1} + 2^n\mathsf{F}_{1,2l+1} + 2^m\mathsf{F}_{2,2l+1}\Bigg].
\end{aligned} \tag{2.64}$$

For $\alpha \in (\frac{2}{3}, 1)$, the last term $2^{m-n}\mathsf{F}_{2,2l+1}$ on the righthand side satisfies $E[(2^{m-n}\mathsf{F}_{2,2l+1})^2] \to 0$ in the limit as $n \to \infty$. Thus it is below the "noise floor", and we rewrite

$$2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 = \mathsf{X}_C + 2^{m-n}\mathsf{F}_{2,2l+1}. \tag{2.65}$$

By the mutual independence among X_C, $\mathsf{F}_{2,2}$ and Z_1,as (2.47) in Case 1, we can also claim that

$$I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1) \geq I(\mathsf{X}_C; \mathsf{X}_C + \mathsf{Z}_1) \tag{2.66}$$

and a fact on X_C:

Fact 3. X_C *is uniform over a discrete constellation of size* $N(\mathsf{X}_C) = 2^{n-4l}$ *with minimum distance* $d_{min}(\mathsf{X}_C) = 1$.

Proof. First, the minimum distance is determined by the last part of X_C which is $\mathsf{F}_{1,2l+1}$ with $d_{min}(\mathsf{F}_{1,2l+1}) = 1$. Then, we proceed to prove X_C is uniform by contradiction. First, thank to the -1 we add, the only thing could cause us trouble is when two random variables align (say for example, $\mathsf{F}_{1,2j-1} + \mathsf{F}_{2,2j-3}$). To show that the align can not happen, suppose there are two different realizations x_C with $\mathsf{F}_{i,j} = f_{i,j}$ and x'_C with $\mathsf{F}_{i,j} = f'_{i,j}$ where not all $f_{i,j} = f'_{i,j}$ such that $x_C = x'_C$. By inspecting the last parts of (2.64), one observes that for $x_C = x'_C$ we must have $f_{1,2l+1} = f'_{1,2l+1}$, $f_{2,1} = f'_{2,1}$ and $f_{2,2} = f'_{2,2}$ since they are corresponding to the disjoint intervals. Now, given this observation, inspecting the top terms of (2.64) provides $f_{1,1} = f'_{1,1}$, and $f_{1,2} = f'_{1,2}$. Keeping doing this will enforce $f_{i,j} = f'_{i,j}$ for all i, j participating in x_C and x'_C which contradicts to the assumption. Simple calculation will then provide the cardinality result.

□

Combine the Fact 3 and Proposition 1 we have

$$
\begin{aligned}
& I(\mathsf{X}_C; \mathsf{X}_C + \mathsf{Z}_1) \\
& \geq H(\mathsf{X}_C) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(1 + \frac{12}{d^2_{min(X_C)}}\right) \\
& = (n-4l)\log 2 - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log 13
\end{aligned} \tag{2.67}
$$

To bound from above the mutual information $I(2^m\mathsf{X}_2; 2^m\mathsf{X}_2 + \mathsf{Z}_1)$, note that

$$
\begin{aligned}
& 2^m\mathsf{X}_2 \\
& = 2^{-n}\Bigg[\sum_{j=1}^{l-1}\left[2^{(l-j+2)n - \frac{2l-2j+2}{2}m}\mathsf{F}_{2,2j-1} + 2^{(j+1)m-(j-1)n}\mathsf{F}_{2,2j}\right] \\
& + 2^{2n-\frac{1}{2}m}\mathsf{F}_{2,2l-1} + 2^{n+\frac{1}{2}m}\mathsf{F}_{2,2l} + 2^{(l+1)n-lm}\mathsf{F}_{2,2l-1} \\
& + \sum_{k=l-1}^{1}\left[2^{\frac{2l-2k+1}{2}m-(l-k-1)n}\mathsf{F}_{2,2k} + 2^{(k+1)n-km}\mathsf{F}_{2,2k-1}\right] + 2^m\mathsf{F}_{2,2l+1}\Bigg],
\end{aligned} \tag{2.68}
$$

where the last term $2^{m-n}\mathsf{F}_{2,2l+1}$ on the right-hand side is below the noise floor as before. We thus have

$$
\begin{aligned}
& I(2^m\mathsf{X}_2; 2^m\mathsf{X}_2 + \mathsf{Z}_1) \\
& \leq H(2^m\mathsf{X}_2 - 2^{m-n}\mathsf{F}_{2,2l+1}) + \frac{1}{2}\log[1 + E[(2^{m-n}\mathsf{F}_{2,2l+1})^2]] \\
& \leq (\frac{1}{2}m - 2l)\log 2 + \frac{1}{2}\log 2.
\end{aligned} \tag{2.69}
$$

Collecting results from above we have

$$
\begin{aligned}
& I(\mathsf{X}_1;\mathsf{Y}_1)+I(\mathsf{X}_2;\mathsf{Y}_2) \\
& \geq 2(n-\frac{1}{2}m)\log 2 - 4l\log 2 - \log\left(\frac{2\pi e}{12}\right) - \log 13 - \log 2 \\
& = 2n - m - c_3(l),
\end{aligned} \tag{2.70}
$$

where $c_3(l)$ is a function of l. Note that for a fixed α, $c_3(l)$ will become a constant.

Case 4: $\alpha \in (1,2]$. In this regime, consider *sub-regimes* $\alpha \in \left(\frac{2l+2}{2l+1}, \frac{2l}{2l-1}\right]$ for $l = 1,2,3,\ldots,$ $F_1, F_2, \ldots F_{l-1} \in \mathbb{F}_2^{m-n,m}$ translates into $2(l-1)$ i.i.d random variables $\mathsf{F}_{1,1}, \mathsf{F}_{1,2}, \ldots, \mathsf{F}_{1,l-1}$ and $\mathsf{F}_{2,1}, \mathsf{F}_{2,2}, \ldots, \mathsf{F}_{2,l-1}$ that are uniform over PAM(2^{m-n-1}), $F_l, F_{l+1} \in \mathbb{F}_2^{ln-\lceil\frac{2l-1}{2}m\rceil,m}$ translates into 4 i.i.d random variables $\mathsf{F}_{1,l}, \mathsf{F}_{1,l+1}$ and $\mathsf{F}_{2,l}, \mathsf{F}_{2,l+1}$ that are uniform over PAM$(2^{ln-\frac{2l-1}{2}m-1})$ and $F_{l+2} \in \mathbb{F}_2^{\lceil\frac{2l+1}{2}m\rceil-(l+1)n,m}$ translates into 2 i.i.d random variables $\mathsf{F}_{1,l+2}, \mathsf{F}_{1,l+2}$ and $\mathsf{F}_{2,l+2}, \mathsf{F}_{2,l+2}$ that are uniform over PAM$(2^{\frac{2l+1}{2}m-(l+1)n-1})$.

Then the choice

$$E = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_l \\ F_{l+1} \\ F_{l+2} \\ F_l \\ F_{l-1} \\ \vdots \\ F_2 \\ F_1 \\ 0_{m-n,m} \end{bmatrix}. \tag{2.71}$$

translates into the following input distributions for the Gaussian model

$$\begin{aligned} \mathsf{X}_i = 2^{-m} \Bigg[& \sum_{j=1}^{l-1} 2^{jn-(j-1)m} \mathsf{F}_{i,j} + 2^{\frac{3}{2}m-n} \mathsf{F}_{i,l} \\ & + 2^{(l+1)m-(l+1)n} \mathsf{F}_{i,l+1} + 2^{\frac{1}{2}m} \mathsf{F}_{i,l+2} + \sum_{k=l}^{1} 2^{km-kn} \mathsf{F}_{i,k} \Bigg], \end{aligned} \tag{2.72}$$

for $i = 1, 2$, where all random variables are assumed to be independent of each other. Note that the term $m - n$, $ln - \frac{2l-1}{2}m$ and $\frac{2l+1}{2}m - (l+1)n$ represent the numbers of rows of the blocks $\{F_1, \ldots, F_{l-1}\}$, F_l, F_{l+1} and F_{l+2} in E respectively. 2^{-m} is chosen such that $E[\mathsf{X}_i^2] \leq 1$ for $i = 1, 2$.

As (2.44) in Case 1, to bound from below the mutual information $I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1)$, note that

$$
\begin{aligned}
&2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 \\
&= 2^{-m}\Bigg[2^{m+n}\mathsf{F}_{2,1} + \sum_{j=2}^{l-1} 2^{jn-(j-2)m}(\mathsf{F}_{2,j} + \mathsf{F}_{1,j-1}) \\
&+ 2^{\frac{3}{2}m}(\mathsf{F}_{2,l+2} + 2^{\frac{2l+1}{2}m-(l+1)n}\mathsf{F}_{2,l+1} + 2^{m-n}\mathsf{F}_{2,l} + 2^{ln-\frac{2l-1}{2}m}\mathsf{F}_{1,l-1} + \mathsf{F}_{1,l}) \\
&+ 2^{(l+1)m-ln}(\mathsf{F}_{2,l} + \mathsf{F}_{1,l+1}) + 2^{lm-(l-1)n}(\mathsf{F}_{2,l-1} + \mathsf{F}_{1,l} + 2^{ln-\frac{2l-1}{2}m}\mathsf{F}_{1,l+2}) \\
&+ \sum_{k=l-2}^{1} 2^{(k+1)m-kn}(\mathsf{F}_{2,k} + \mathsf{F}_{1,k+1}) + 2^m\mathsf{F}_{1,1}\Bigg].
\end{aligned} \tag{2.73}
$$

For $\alpha \in (1, 2]$, there is no term below the "noise floor". Also, there is a fact that

Fact 4. $2^n\mathsf{X}_1 + 2^m\mathsf{X}_2$ *is uniform over a discrete constellation of size* $N(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2) = 2^{m-(2l+4)}$*with minimum distance* $d_{min}(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2) = 2$.

Proof. Similar to the proof of Fact 3. □

Combine the Fact 4 and Proposition 1 we have

$$
\begin{aligned}
&I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1) \\
&\geq H(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(1 + \frac{12}{d^2_{min(2^n\mathsf{X}_1+2^m\mathsf{X}_2)}}\right) \\
&= (m - (2l+4))\log 2 - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log 4
\end{aligned} \tag{2.74}
$$

To bound from above the mutual information $I(2^m \mathsf{X}_2; 2^m \mathsf{X}_2 + \mathsf{Z}_1)$, note that

$$2^m \mathsf{X}_2 = 2^{-m} \Bigg[\sum_{j=1}^{l-1} 2^{jn-(j-2)m} \mathsf{F}_{2,j} + 2^{\frac{5}{2}m-n} \mathsf{F}_{2,l} + 2^{(l+2)m-(l+1)n} \mathsf{F}_{2,l+1} + 2^{\frac{3}{2}m} \mathsf{F}_{2,l+2} + \sum_{k=l}^{1} 2^{(k+1)m-kn} \mathsf{F}_{2,k} \Bigg] \tag{2.75}$$

is uniform over PAM$(2^{\frac{1}{2}m-(l+2)})$ with the minimum distance $d_{min}(2^m \mathsf{X}_2) > 1$, so that

$$I(2^m \mathsf{X}_2; 2^m \mathsf{X}_2 + \mathsf{Z}_1) \leq (\frac{1}{2}m - (l+2)) \log 2.$$

Collecting results from above we have

$$\begin{aligned} I(\mathsf{X}_1; \mathsf{Y}_1) + I(\mathsf{X}_2; \mathsf{Y}_2) &\geq m \log 2 - (2l+4) \log 2 - \log \left(\frac{2\pi e}{12} \right) - \log 4 \\ &= m - c_4(l), \end{aligned} \tag{2.76}$$

where $c_4(l)$ is a function of l. Note that for a fixed α, $c_4(l)$ will become a constant.

Case 5: $\alpha \in [2, \infty)$. In this regime, $F_1 \in \mathbb{F}_2^{n,n}$ translates into 2 i.i.d random variables $\mathsf{F}_{1,1}$and $\mathsf{F}_{2,1}$ that are uniform over PAM(2^n). Then the choice

$$E = \begin{bmatrix} F_1 \\ 0_{m-n,n} \end{bmatrix} \tag{2.77}$$

translates into the following input distributions for the Gaussian model

$$\mathsf{X}_i = 2^{-m}2^{m-n}\mathsf{F}_{i,1} = 2^{-n}\mathsf{F}_{i,1}, \tag{2.78}$$

for $i = 1, 2$ where $\mathsf{F}_{1,1}, \mathsf{F}_{2,1}$ are assumed to be independent of each other. Note that the term 2^{-m} is chosen such that $E[\mathsf{X}_i^2] \leq 1$ for $i = 1, 2$.

As (2.44) in Case 1, to bound from below the mutual information $I(2^n\mathsf{X}_1+2^m\mathsf{X}_2; 2^n\mathsf{X}_1+2^m\mathsf{X}_2 + \mathsf{Z}_1)$, note that

$$2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 = \mathsf{F}_{1,1} + 2^{m-n}\mathsf{F}_{2,1} \tag{2.79}$$

is uniform over PAM(2^{2n}) with the minimum distance $d_{min}(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2) = 1$. By Proposition 1 we have

$$I(2^n\mathsf{X}_1 + 2^m\mathsf{X}_2; 2^n\mathsf{X}_1 + 2^m\mathsf{X}_2 + \mathsf{Z}_1) \geq 2n\log 2 - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log 13. \tag{2.80}$$

Also, note that

$$2^m\mathsf{X}_2 = 2^{-m}2^{2m-n}\mathsf{F}_{2,1} = 2^{m-n}\mathsf{F}_{2,1} \tag{2.81}$$

is uniform over PAM(2^n) with the minimum distance $d_{min}(2^m \mathsf{X}_2) > 1$, so that

$$I(2^m \mathsf{X}_2; 2^m \mathsf{X}_2 + \mathsf{Z}_1) \leq n \log 2. \tag{2.82}$$

Collecting results from above we have

$$\begin{aligned} I(\mathsf{X}_1;\mathsf{Y}_1) + I(\mathsf{X}_2;\mathsf{Y}_2) &\geq 2n \log 2 - \log\left(\frac{2\pi e}{12}\right) - \log 13 \\ &= 2n - c_5, \end{aligned} \tag{2.83}$$

where c_5 is a constant.

Collecting (2.51) (2.61) (2.70) (2.76) and (2.83) above we have the following bound,

$$I(\mathsf{X}_1;\mathsf{Y}_1) + I(\mathsf{X}_2;\mathsf{Y}_2) \geq \begin{cases} 2(n-m) - c_1, & \text{for } \alpha \in \left(0, \frac{1}{2}\right] \\ 2m - c_2, & \text{for } \alpha \in \left(\frac{1}{2}, \frac{2}{3}\right] \\ 2n - m - c_3(l), & \text{for } \alpha \in \left(\frac{2}{3}, 1\right) \\ m - c_4(l), & \text{for } \alpha \in (1, 2) \\ 2n - c_5, & \text{for } \alpha \in [2, \infty) \end{cases} \tag{2.84}$$

Recall that from (2.40)

$$C_{sum}(\mathrm{SNR},\mathrm{INR}) \leq \begin{cases} \log(\frac{1+2^{2n}}{1+2^{2m}}), & \text{for } \alpha \in \left(0,\frac{1}{2}\right] \\ \log(1+2^{2m}+\frac{2^{2n}}{1+2^{2m}}), & \text{for } \alpha \in \left(\frac{1}{2},\frac{2}{3}\right] \\ \frac{1}{2}\log\left(\frac{(1+2^{2n})(1+2^{2m}+2^{2n})}{1+2^{2m}}\right), & \text{for } \alpha \in \left(\frac{2}{3},1\right) \\ \frac{1}{2}\log\left(1+2^{2m}+2^{2n}\right), & \text{for } \alpha \in (1,2) \\ \log\left(1+2^{2n}\right), & \text{for } \alpha \in [2,\infty) \end{cases}$$

$$= \begin{cases} 2(n-m)+O(1), & \text{for } \alpha \in \left(0,\frac{1}{2}\right] \\ 2m+O(1), & \text{for } \alpha \in \left(\frac{1}{2},\frac{2}{3}\right] \\ 2n-m+O(1), & \text{for } \alpha \in \left(\frac{2}{3},1\right) \\ m+O(1), & \text{for } \alpha \in (1,2) \\ 2n+O(1), & \text{for } \alpha \in [2,\infty) \end{cases} \tag{2.85}$$

Combine (2.84) and (2.85) will complete the proof.

3. CLUSTERING FOR BRAIN NETWORK VIA METRIC LEARNING

3.1 K-means Clustering Algorithm

K-means algorithm was first introduced in [17] with the main idea of finding K centroids of clusters. This algorithm is shown as follows,

1. Randomly pick K nodes as the initial centroids of K clusters
2. Assign each nodes to the cluster with nearest centroid
3. When all nodes are assigned, recalculate the locations of K centroids
4. Repeat Step 2 and 3 until the centroids do not change.

Although using K-means to find a optimal solution is NP-hard even in 2 dimension space [27] [28], it is still a widely commonly used clustering algorithm due to its simplicity. In this dissertation, as in [21] [22] , we consider the matrix representation of K-means algorithm. Let $X = [x_1, \dots, x_N]^\top \in \mathcal{R}^{N \times T}$ be the set of fMRI time-series data of N voxels in $\in \mathcal{R}^T$. In K-means, looking for K clusters is equivalent to looking for

- An assignment matrix $Y \in \{0,1\}^{N \times K}$ with the following properties,

  1.
$$Y_{i,k} = \begin{cases} 1, \text{if } v_i \text{ is in cluster } k \\ 0, \text{otherwise.} \end{cases} \tag{3.1}$$

  2. $rank(Y)$=K: because there are K clusters.
  3. $Y\mathbf{1}_K = \mathbf{1}_N$, where $\mathbf{1}_K$ is a all one vector with length K : because one voxel is assigned to one and only one cluster.

- A centroid matrix $Z = [z_1, z_2, \dots, z_K]^\top \in \mathcal{R}^T$ where $z_k \in \mathcal{R}^T$ is the centroid of cluster k.

Then, K-means is to minimize to following distortion function

$$\min_{Y \in \{0,1\}^{N \times K}, Y\mathbf{1}_K = \mathbf{1}_T, rank(Y)=K, Z \in \mathcal{R}^{K \times p}} \sum_{i=1}^{N} \sum_{k=1}^{K} Y_{i,k} ||x_i - c_k||^2 \tag{3.2}$$

where $||A||^2$ is the Euclidean norm of A. By considering Frobenius norm $||A||_F^2 = \text{Tr}(AA^\top)$, we can rewrite 3.2 in matrix form

$$\min_{Y \in \{0,1\}^{N \times K}, Y\mathbf{1}_K = \mathbf{1}_T, rank(Y)=K, Z \in \mathcal{R}^{K \times p}} ||X - YZ||_F^2 \tag{3.3}$$

For a given assignment matrix Y, the centroid matrix Z can be solved in closed form as $Z^* = (Y^\top Y)^{-1} Y^\top X = Y^\dagger X$ where $Y^\dagger$ is the Moore Penrose Pseudoinverse of Y. By defining $C = YY^\dagger \in \mathcal{R}^{N \times N}$ as the clustering matrix, (3.3) now becomes

$$\begin{gathered} \min_{C \in \mathcal{C}_\mathcal{A}} tr(XX^\top (I - C)) \\ \mathcal{C}_\mathcal{A} = \{YY^\dagger : Y \in \{0,1\}^{T \times K}, Y\mathbf{1}_K = \mathbf{1}_T, rank(Y) = K\} \end{gathered} \tag{3.4}$$

As shown in [21] [22], the optimization problem above can be relaxed to the following form

$$\begin{gathered} \max_{C \in \mathcal{C}_\mathcal{R}} tr(XX^\top C) \\ \mathcal{C}_\mathcal{R} = \{C : C \in \mathcal{R}^{T \times T}, C^2 = C, tr(C) = K\} \end{gathered} \tag{3.5}$$

with the optimal solution $\tilde{C}$ being the orthogonal projector onto the K leading eigenvectors of $XX^\top$. Since $\tilde{C} \in \mathcal{C}_\mathcal{R}$ is not always in $\mathcal{C}_\mathcal{A}$, one way to find the hard assignment C^* heuristically is using K-means over the k leading eigenvectors of $XX^\top$. In this dissertation, we use the multiclass spectral clustering method proposed by [19] and also in [13] with a faster converge rate to a nearly global-optimal solution.

3.2 Metric Learning

We consider training a Mahalanobis distance [29] $d_M(\mathbf{x}, \mathbf{y})$ which is parameterized by a symmetric positive semidefinite(PSD) matrix $M \in \mathcal{R}^{T\times T}$,

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^\top \mathbf{M}(\mathbf{x}-\mathbf{y})} \tag{3.6}$$

Equivalently, we can rewrite (3.5) as

$$\begin{gathered}\max_{C\in\mathcal{C}_\mathcal{R}} tr(XMX^\top C)\\ \mathcal{C}_\mathcal{R} = \{C : C \in \mathcal{R}^{T\times T}, C^2 = C, tr(C) = K\}\end{gathered} \tag{3.7}$$

Note that when $M = I$, (3.7) becomes (3.5). The solution of (3.7) is the orthogonal projector onto the K leading eigenvectors of $XMX^\top$. Then, as in [22], this Mahalanobis distance matrix M can be learned as the minimizer of the following formulation,

$$\min_{M\succeq 0, tr(M)=1} [\max_{C\in\mathcal{C}_\mathcal{R}} (\triangle(C, \bar{C}) + \gamma(M, C))], \tag{3.8}$$

where

- γ is an indicator function

$$\gamma(M, C) = \begin{cases} 0, & if\ C \in \text{argmax}_{C \in \mathcal{C}_\mathcal{R}} tr(XMX^\top C) \\ -\infty, & \text{otherwise} \end{cases} \tag{3.9}$$

- $\bar{C}$ is the ground truth

- $\triangle(C, \bar{C}) = ||C - \bar{C}||^2$ measures the difference between C and $\bar{C}$.

(3.8) has a closed form solution

$$M = sX^\dagger \bar{C}(X^\dagger)^\top, s > 0 \tag{3.10}$$

3.3 Unsupervised Clustering

In practice, the ground truth information is not always available especially for the brain network. We propose an iterative unsupervised clustering algorithm as follows,

First, (3.7) can be rewritten with (3.10) as follows,

$$\max_{C \in \mathcal{C}_\mathcal{R}} tr((XX^\dagger)\bar{C}(XX^\dagger)^\top C)$$
$$\mathcal{C}_\mathcal{R} = \{C : C \in \mathcal{R}^{T \times T}, C^2 = C, tr(C) = K\}, \tag{3.11}$$

Then, the proposed iterative algorithm is,

INPUT: Data set X:

1. Set $M = I$.

2. Solve (3.7) to get a clustering matrix C_p.

3. Solve (3.11) with $\bar{C}$ replaced by C_p. Calculate the accuracy.

4. Redo step 3 until improvement is less than a threshold.

3.4 Simulation Results

In this simulation, we use the same set of synthetic fMRI data as in [13]. As shown in Fig 3.1, There are 10 simulated data set and each has $N = 489$ voxels that are separated into six regions.

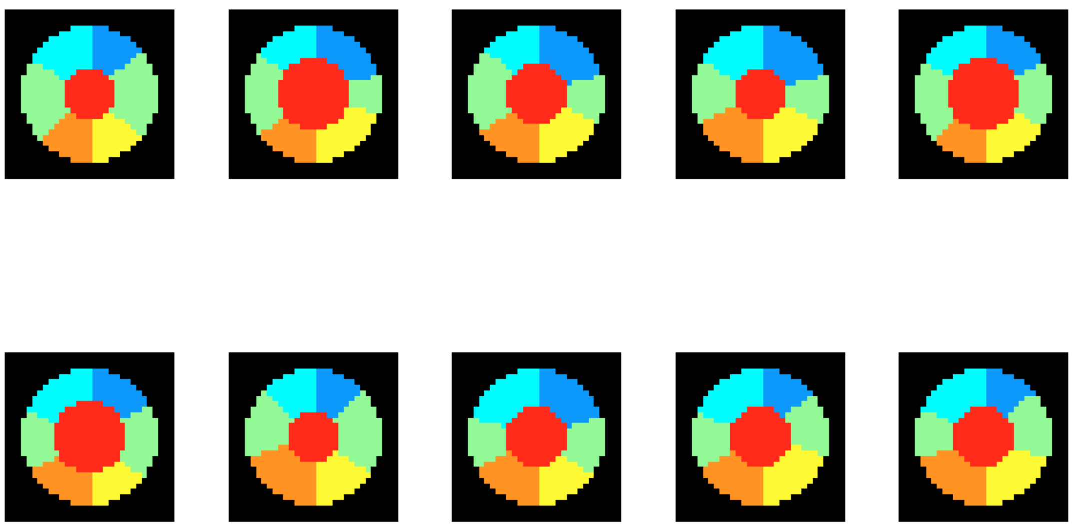


Figure 3.1: Synthetic data template

There are $K = 6$ synthetic time-series data $s_k(t),\ k \in [1 : K]$ which are generated using real resting-state fMRI data from six regions of the brain [13]. For the voxel in the kth cluster, the time-series is generated by adding a Gaussian noise to $s_k(t)$ as follows

$$x_i^k(t) = s_k(t) + \alpha n(t), i \in [1 : N],\ k \in [1 : K] \tag{3.12}$$

where

- $x_i^k(t)$ is the time series for voxel i which is in region k.
- $n(t)$ is standard Gaussian noise.
- α is to control the SNR.

In order to understand the nature of the learned metric M, we consider using a special noise pattern (α_1, α_2). As shown in Figure 3.2, the first half of time series is added with noise at level α_1 and the second half is with noise at level α_2.

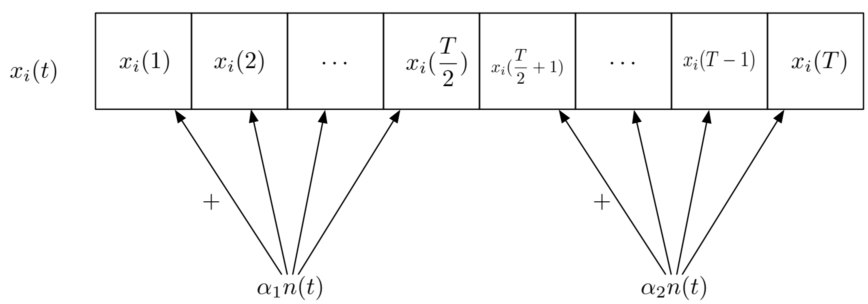


Figure 3.2: Special noise pattern (α_1, α_2)

To evaluate the performance of the clustering results, we use the Dice's coefficient

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \tag{3.13}$$

to measure the overlap between the clustering result C and the ground truth $\bar{C}$.

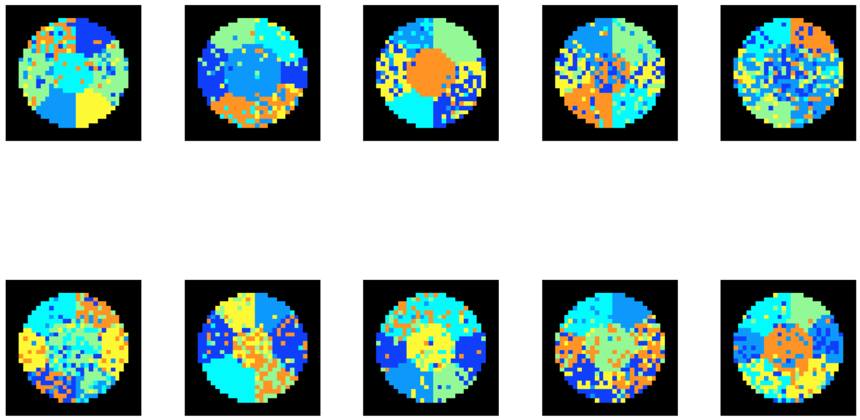
We now show the clustering result of supervised clustering and unsupervised clustering with the choice of $\alpha \in \{0.2, 0.4, (0.1, 0.3)\}$ as examples shown in Figure 3.3, Figure 3.4, Figure 3.5 while the clustering accuracy is shown in Figure 3.7.

Remark 2.

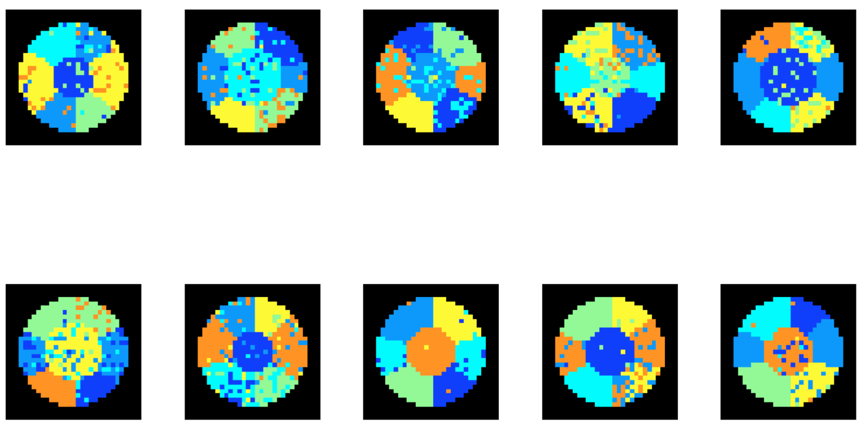
- *Note that the color assignments in the clustering result are not the same as in template, but it will not affect the clustering accuracy.*
- *In Figure 3.7, the abbreviations are as follows,*
  - *NL: Without learning*
  - *Ni,* $i \in [1, 2, 3, 4]$*: Learning from previous result*
  - *GT: Learn from ground truth*

3.5 Observations

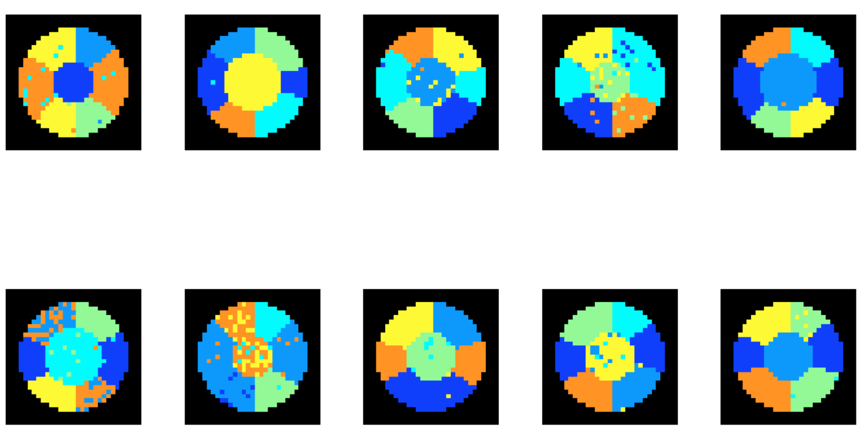
By investigating the simulation results, we have the following key observations. First, when the ground truth is available, implementing the learned Mahalanobis distance can greatly improve the clustering results no matter how large the noise is. For example, in Figure 3.4a, the noise is large enough to mess up the original clustering result, but Figure 3.4c shows a almost perfect clustering result after learning. Second, when the noise pattern is (α_1, α_2), the learned metric M acts as a feature selection matrix to follow the data with better quality. For example, in Figure 3.6a, the top-left part of M has larger coefficients (brighter) than the bottom-right part (darker). It looks like that M 'selects' the part with smaller noise which, as expected, leads to a good clustering result shown in Figure 3.5c. Finally, for the unsupervised algorithm, as in Figure 3.7, there are always improvements from NL to N1. The level of improvements varies with different noise level. For example, in Figure 3.3b, we can see visible improvements when noise is small ($\alpha = 0.2$) and barely see any improvements when noise is large ($\alpha = 0.4$) in Figure 3.4b. However, this algorithm does not converge after iterations which can be observed from L2 to L4 in Figure 3.7.



(a) Without learning

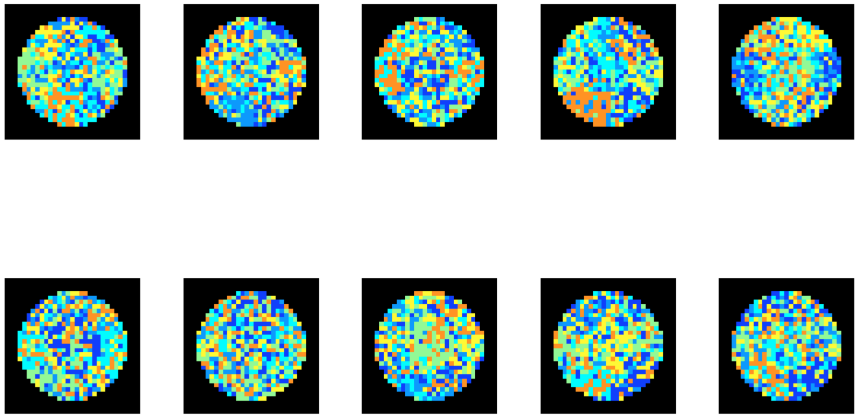


(b) Learning from previous result

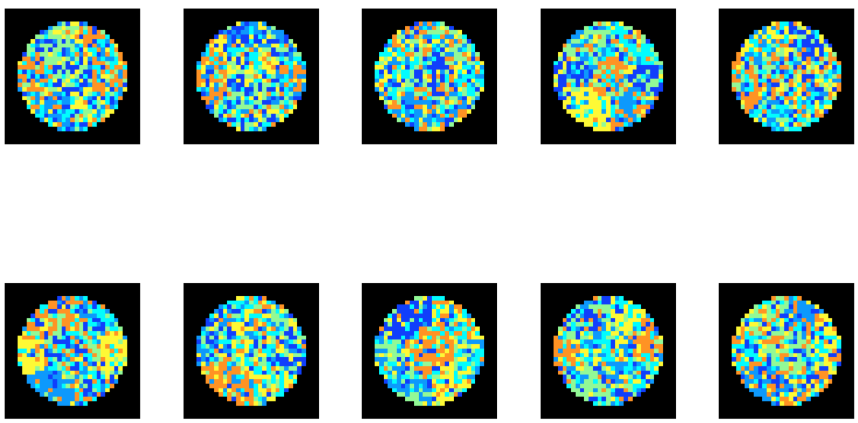


(c) Learning from ground truth

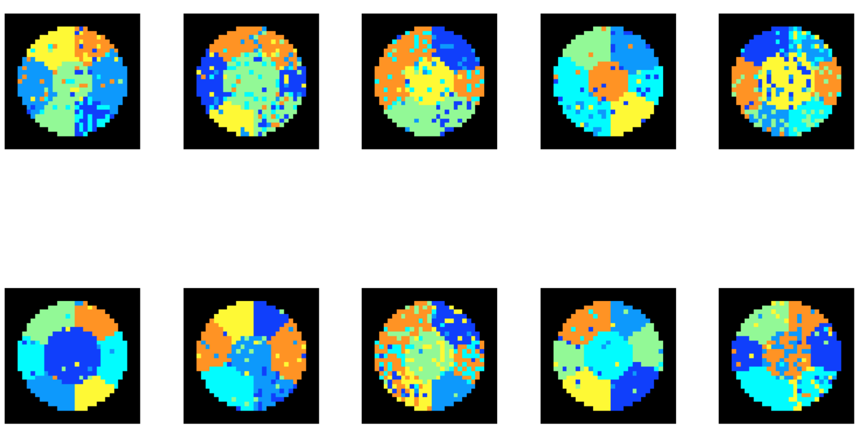
Figure 3.3: Clustering results at noise level 0.2



(a) Without learning

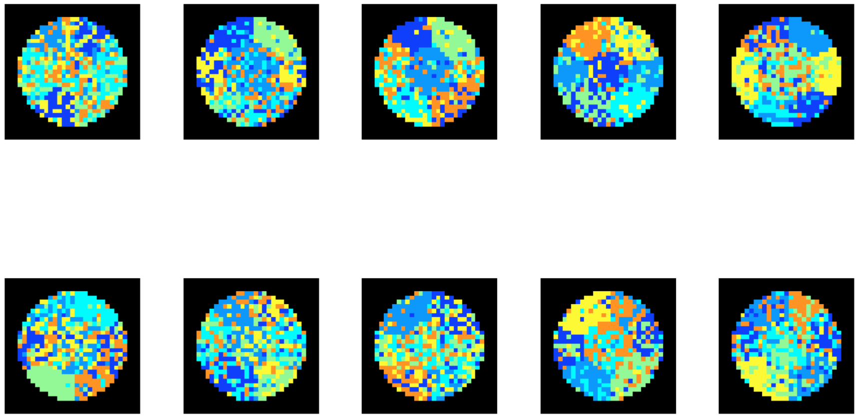


(b) Learning from previous result

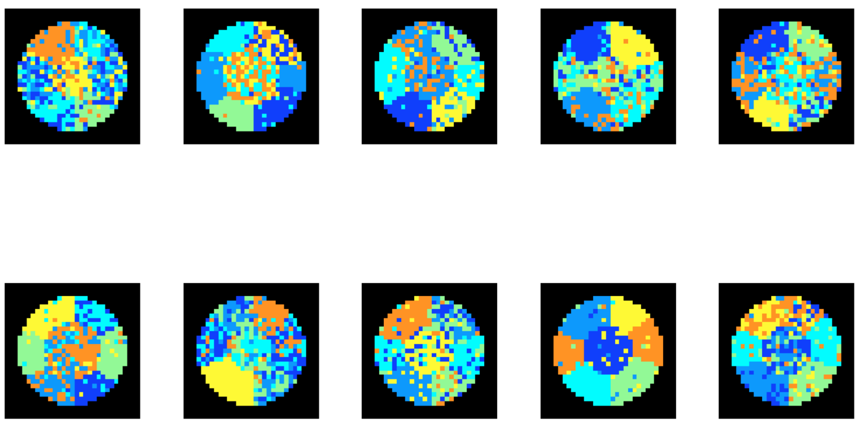


(c) Learning from Ground Truth

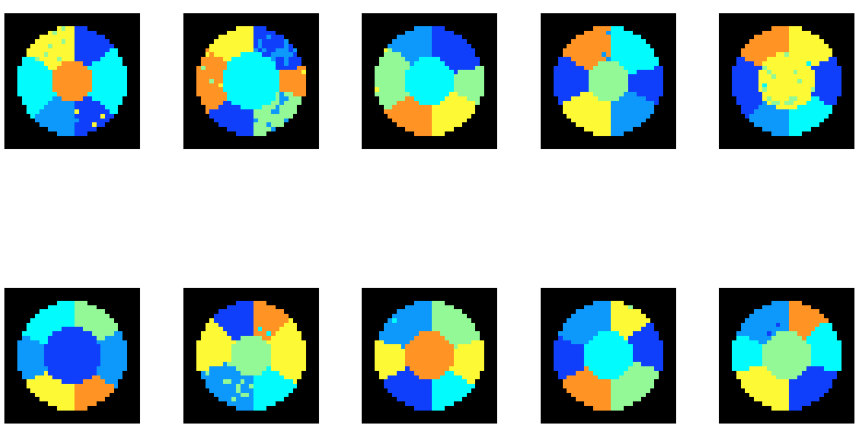
Figure 3.4: Clustering results at noise level 0.4



(a) Without learning

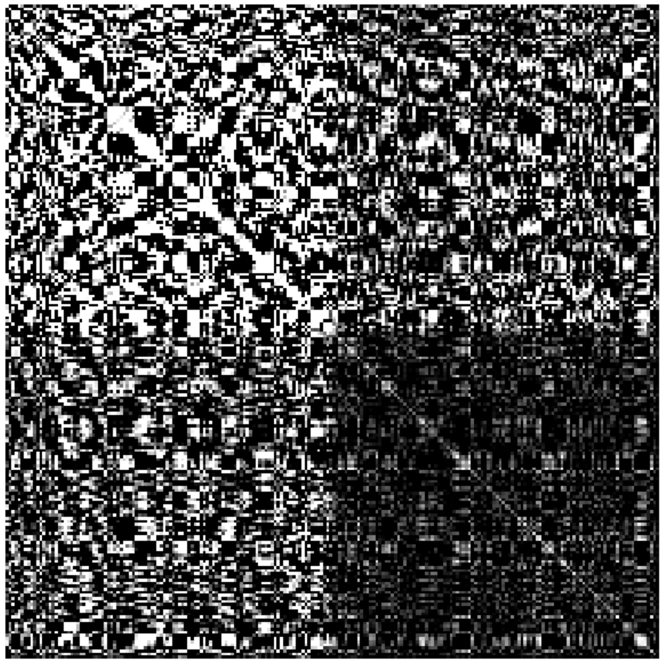


(b) Learning from previous result

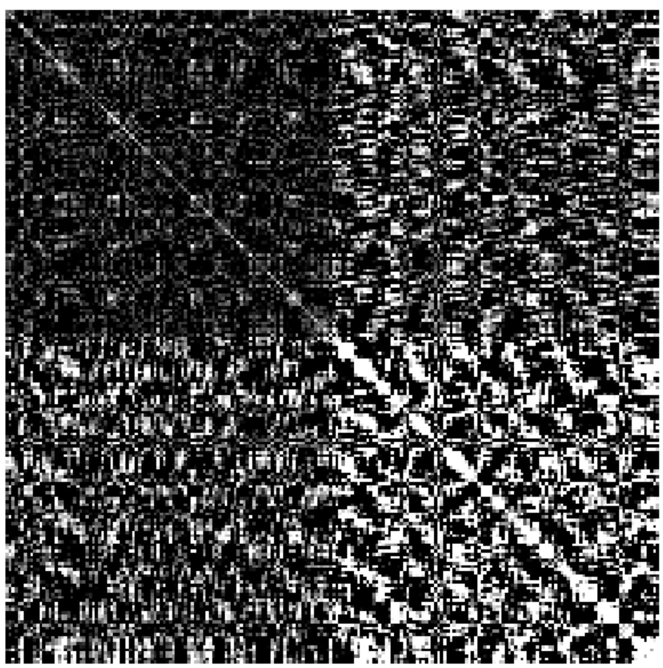


(c) Learning from ground truth

Figure 3.5: Clustering results at noise level (0.1, 0.3)



(a) Learned M at noise level (0.1, 0.3)



(b) Learned M at noise level (0.4, 0.2)

Figure 3.6: Property of M

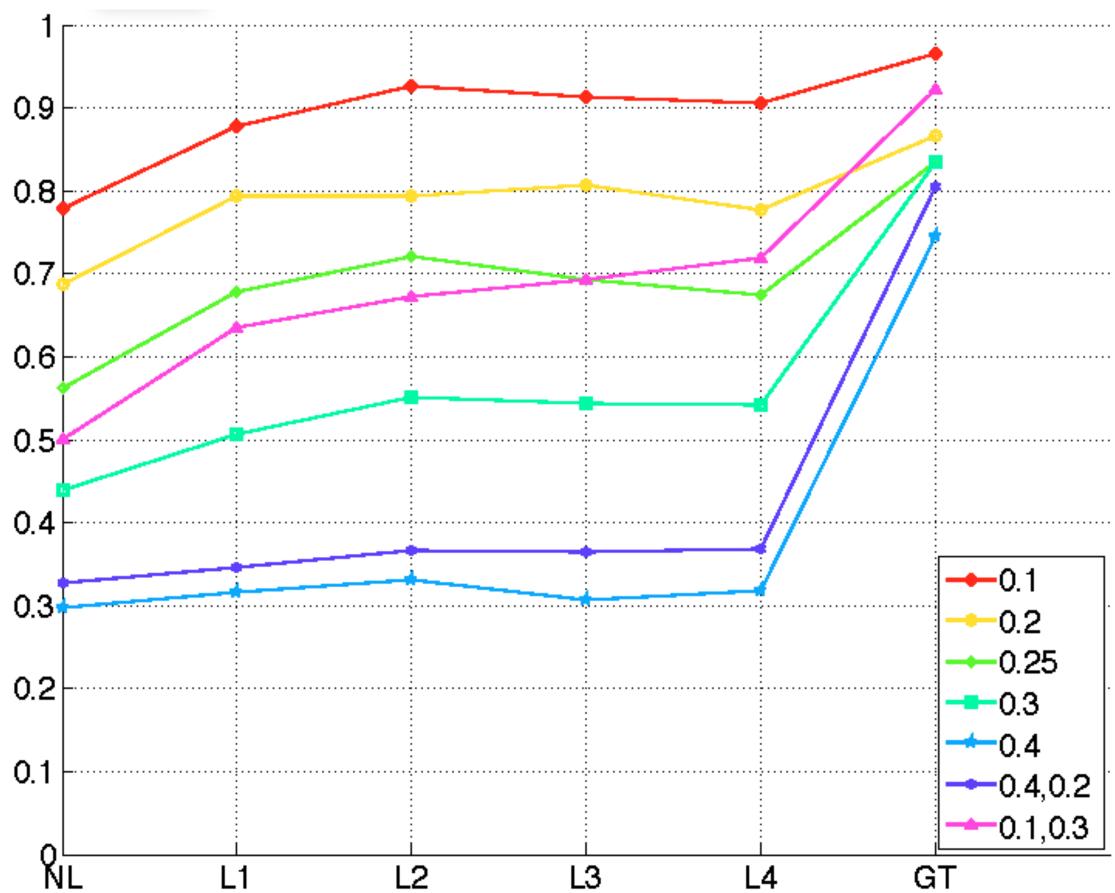


Figure 3.7: Clustering accuracy

4. CONCLUSIONS

In this dissertation, we have studied the following two network problems.

4.1 Treating Interference as Noise for Communication Network

First, we have studied the two-user Gaussian symmetric interference channel. Unlike abundant existing research, we have tackled the problem by finding (approximately) optimal input distributions for the conventional scheme which uses random codebooks at the transmitters and typical decoding with treating interference as noise at the receivers. In order to do so, we first looked into the corresponding linear deterministic model and obtained (approximately) optimal input distributions. This result has then been leveraged to propose approximately optimal distributions for the original Gaussian problem. Our result is of practical importance as it is based on purely discrete input distributions at the transmitters and treating interference as noise at receivers, both are of practical interest. Our result can also be extended to a more general network scenario.

4.2 Clustering for Brain Network via Metric Learning

Second, we have studied the clustering for brain network with metric learning based on synthetic resting-state fMRI time-series. We first introduced the K-means clustering algorithm in matrix form. Then a metric learning method has been implemented into the clustering algorithm by learning a Mahalanobis distance matrix from ground truth. Since the ground truth is not accessible for the brain network, we have proposed an iterative unsupervised clustering algorithm. Property of the learned metric M has also been studied by introducing a dataset with special noise pattern. Computer simulation results have shown the benefit of using metric learning and also shown the nature of M as a feature selection matrix to use the better part of data.

In the future, the following research directions are considered. First, by deeper investigating the nature of the learned matrix M, the unsupervised clustering method could be improved by implementing proper constraints. Then, real resting-state fMRI data can be tested in our algorithm. Finally, group clustering method is worth investigating.

REFERENCES

[1] R. Ahlswede, “Multi-way communication channels,” pp. 23–52, Sept. 1971.

[2] V. S. Annapureddy and V. V. Veeravalli, “Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region,” vol. 55, pp. 3032–3050, July 2009.

[3] X. Shang, G. Kramer, and B. Chen, “A new outer bound and the noisy interference sum-rate capacity for the Gaussian interference channels,” vol. 55, pp. 689–699, Feb. 2009.

[4] A. S. Motahari and A. K. Khandani, “Capacity bounds for the Gaussian interference channel,” vol. 55, pp. 620–643, Feb. 2009.

[5] T. S. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” pp. 49–60, Jan. 1981.

[6] M. H. M. Costa and A. El Gamma, “The capacity region of the discrete and memoryless interference channel with strong interference,” pp. 710–711, Sept. 1987.

[7] R. H. Etkin, D. N. C. Tse, and H. Wang, “Gaussian interference channel capacity to within one bit,” vol. 54, pp. 5534–5562, Dec. 2008.

[8] A. Dytso, N. Devroye, and D. Tuninetti, “On Gaussian interference channels with mixed Gaussian and discrete inputs,” in *Proc. ISIT*, pp. 261–265, June 2014.

[9] A. Dytso, D. Tuninetti, and N. Devroye, “Interference as noise: Friend or foe?,” *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3561–3596, 2016.

[10] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: A deterministic approach,” vol. 57, pp. 1872–1905, Apr. 2011.

[11] M. El-Halabi, T. Liu, C. N. Georghiades, and S. Shamai, “Secret writing on dirty paper: A deterministic view,” vol. 58, pp. 3419–3429, June 2012.

[12] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, “A whole brain fmri atlas generated via spatially constrained spectral clustering,” *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.

[13] X. Shen, F. Tokoglu, X. Papademetris, and R. T. Constable, “Groupwise whole-brain parcellation from resting-state fmri data for network node identification,” *Neuroimage*, vol. 82, pp. 403–415, 2013.

[14] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, “Functional connectivity in the motor cortex of resting human brain using echo-planar mri,” *Magnetic resonance in medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[15] B. B. Biswal, J. V. Kylen, and J. S. Hyde, “Simultaneous assessment of flow and bold signals in resting-state functional connectivity maps,” *NMR in Biomedicine*, vol. 10, no. 45, pp. 165–170, 1997.

[16] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, “Functional connectivity in the resting brain: a network analysis of the default mode hypothesis,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 1, pp. 253–258, 2003.

[17] S. P. Lloyd, “Least square quantization in pcm,” *Technical report*, 1957.

[18] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[19] S. X. Yu and J. Shi, “Multiclass spectral clustering,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 313–319, IEEE, 2003.

[20] W. E. Donath and A. J. Hoffman, “Lower bounds for the partitioning of graphs,” *IBM Journal of Research and Development*, vol. 17, no. 5, pp. 420–425, 1973.

[21] R. Lajugie, F. R. Bach, and S. Arlot, "Large-margin metric learning for constrained partitioning problems.," in *ICML*, pp. 297–305, 2014.

[22] M. T. Law, Y. Yu, M. Cord, and E. P. Xing, "Closed-form training of mahalanobis distance for supervised clustering," in *Computer Vision and Pattern Recognition, 2016, Proceedings. 29th IEEE Conference on*, IEEE, 2016.

[23] S. Li, Y.-C. Huang, T. Liu, and H. D. Pfister, "On the limits of treating interference as noise for two-user symmetric gaussian interference channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1711–1715, IEEE, 2015.

[24] G. Bresler and D. N. C. Tse, "The two-user gaussian interference channel: A deterministic view," *Eur. Trans. Telecommunications*, vol. 19, pp. 333–354, Apr. 2008.

[25] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian k-unser interference channel," vol. 56, pp. 3297–3303, July 2010.

[26] L. H. Ozarow and A. D. Wyner, "On the capacity of the gaussian channel with a finite number of input levels," *IEEE transactions on information theory*, vol. 36, no. 6, pp. 1426–1428, 1990.

[27] A. Vattani, "The hardness of k-means clustering in the plane," *Manuscript, accessible at http://cseweb. ucsd. edu/avattani/papers/kmeans_hardness. pdf*, vol. 617, 2009.

[28] M. Mahajan, P. Nimbhorkar, and K. Varadarajan, "The planar k-means problem is np-hard," in *International Workshop on Algorithms and Computation*, pp. 274–285, Springer, 2009.

[29] P. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Science of India 12*, 1936.

APPENDIX A

PROOF OF PROPOSITION 1

Proof. We first introduce a random variable U which is uniformly distributed on $[-\frac{d_{min}}{2}, \frac{d_{min}}{2}]$ and independent of X_D and Z. Let $Y = X_D + Z$ and $\hat{X} = X_D + U$, then $\hat{X} \rightarrow X_D \rightarrow Y$ form a Markov chain. From data-processing inequality we have

$$I(X_D; Y) \geq I(\hat{X}; Y) = h(\hat{X}) - h(\hat{X}|Y) \tag{A.1}$$

Also,

$$h(\hat{X}) = H(X_D) + \log(d_{min}) \tag{A.2}$$

Then, for a given y

$$\begin{aligned} h(\hat{X}|Y = y) &= -\int p(\hat{x}|y) \log(p(\hat{x}|y))d\hat{x} \\ &\leq -\int p(\hat{x}|y) \log(q_y(\hat{x}))d\hat{x} \end{aligned} \tag{A.3}$$

for any $q_y(\hat{x})$. Let us pick

$$q_y(\hat{x}) = \frac{1}{\sqrt{2\pi}s} exp\left\{-\frac{(\hat{x} - ky)^2}{2s^2}\right\} \tag{A.4}$$

Then we have

$$h(\hat{X}|Y=y) \leq (\log e)\left\{\frac{1}{2}\ln 2\pi s^2 + \frac{1}{2s^2}E[(\hat{X}-ky)^2|Y=y]\right\} \tag{A.5}$$

and

$$h(\hat{X}|Y) \leq (\log e)\left\{\frac{1}{2}\ln 2\pi s^2 + \frac{1}{2s^2}E[(\hat{X}-kY)^2]\right\} \tag{A.6}$$

Furthermore, since X_D, U and Z are mutually independent and U, Z are zero-mean, we have

$$\begin{aligned}
E(\hat{X}-kY)^2 &= E((1-k)X_D + U - kZ)^2 \\
&= (1-k)^2 EX_D^2 + EU^2 + k^2 EZ^2 \\
&= (1-k)^2 EX_D^2 + \frac{d_{min}^2}{12} + k^2 \\
&= EX_D^2 + \frac{d_{min}^2}{12} - 2kEX_D^2 + k^2(EX_D^2+1)
\end{aligned} \tag{A.7}$$

This expression is minimized for $k = \frac{EX_D^2}{EX_D^2+1}$, so that

$$h(\hat{X}|Y) \leq (\log e)\left\{\frac{1}{2}\ln 2\pi s^2 + \frac{1}{2s^2}\left(\frac{d_{min}^2}{12} + \frac{EX_D^2}{EX_D^2+1}\right)\right\} \tag{A.8}$$

and this expression is minimized for $s^2 = \frac{d_{min}^2}{12} + \frac{EX_D^2}{EX_D^2+1}$, so that

$$h(\hat{X}|Y) \leq \frac{1}{2}\log\left[2\pi e\left(\frac{d_{min}^2}{12} + \frac{EX_D^2}{EX_D^2+1}\right)\right] \leq \frac{1}{2}\log\left[2\pi e\left(\frac{d_{min}^2}{12} + 1\right)\right] \tag{A.9}$$

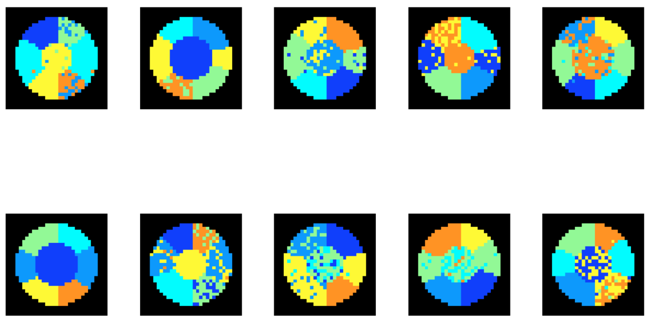
Collecting the results above, we have the bound

$$
\begin{aligned}
I(X_D;Y) &\geq h(\hat{X}) - h(\hat{X}|Y) \\
&= H(X_D) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \frac{1}{2}\log\left(1 + \frac{12}{d^2_{min(X_D)}}\right)
\end{aligned} \tag{A.10}
$$

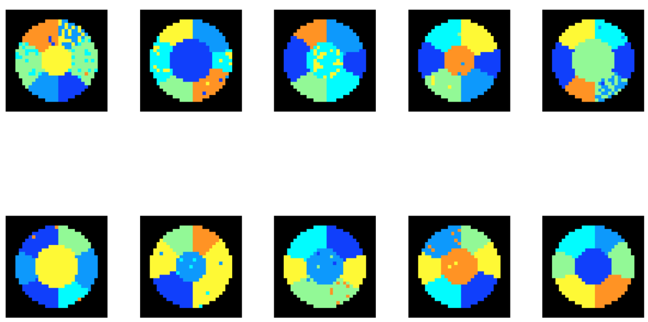
□

APPENDIX B

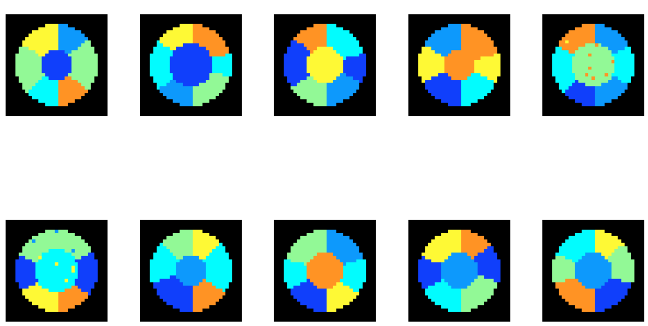
REST SIMULATION RESULTS



(a) Without learning

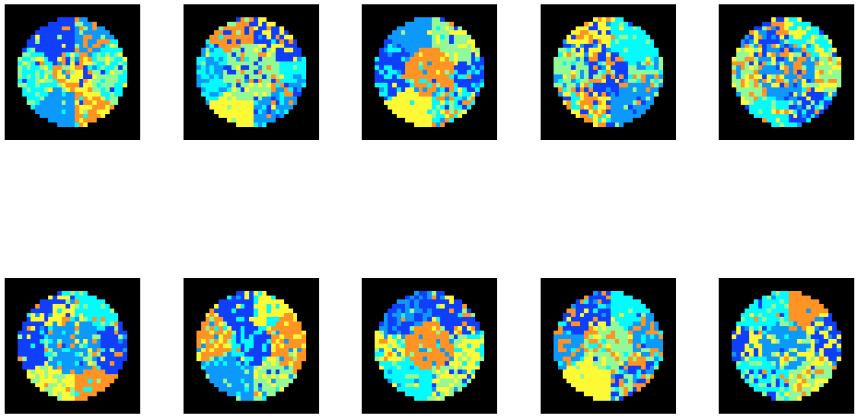


(b) Learning from previous result

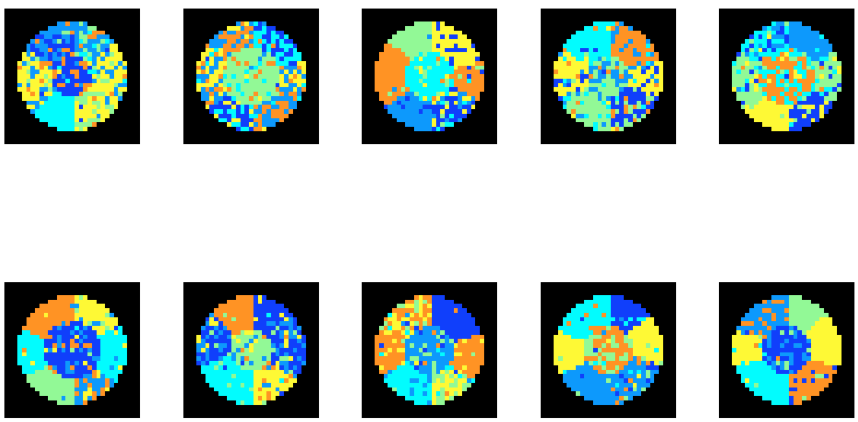


(c) Learning from ground truth

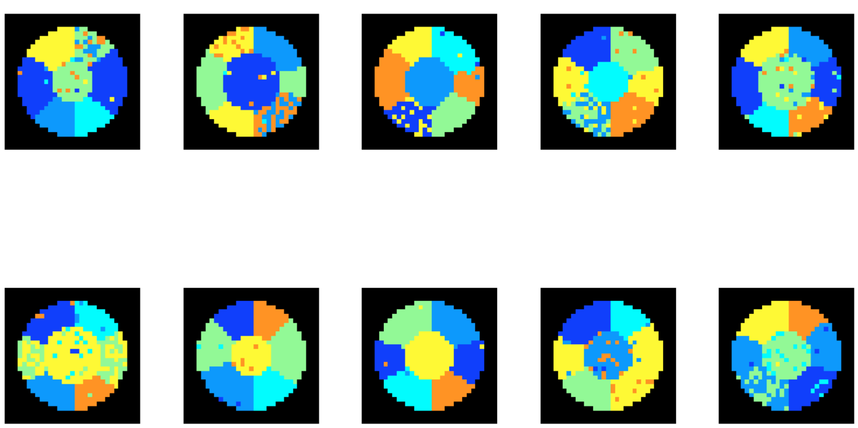
Figure B.1: Clustering results at noise level 0.1



(a) Without learning

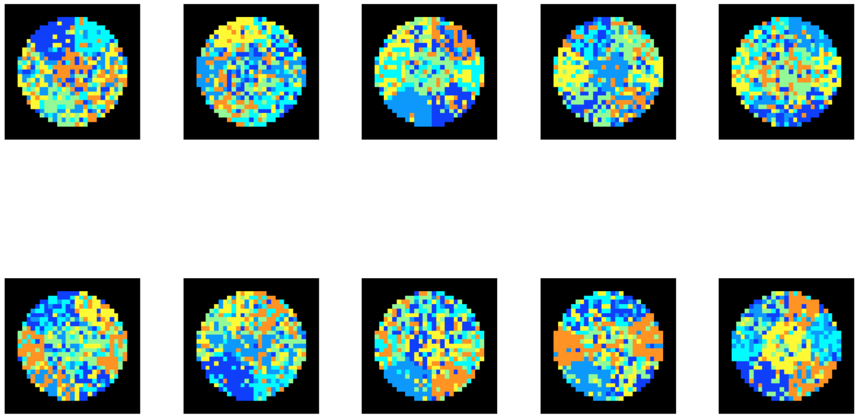


(b) Learning from previous result

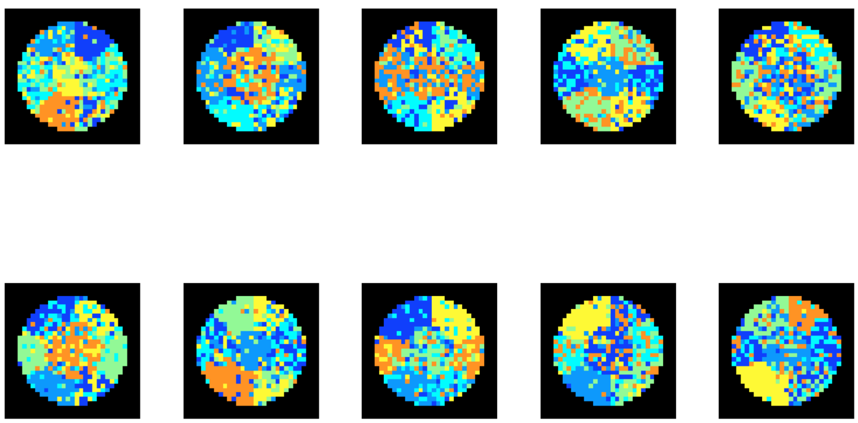


(c) Learning from ground truth

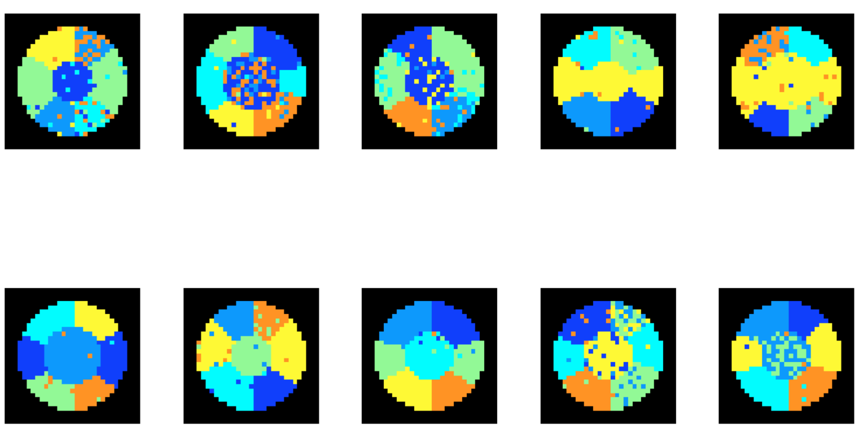
Figure B.2: Clustering results at noise level 0.25



(a) Without learning

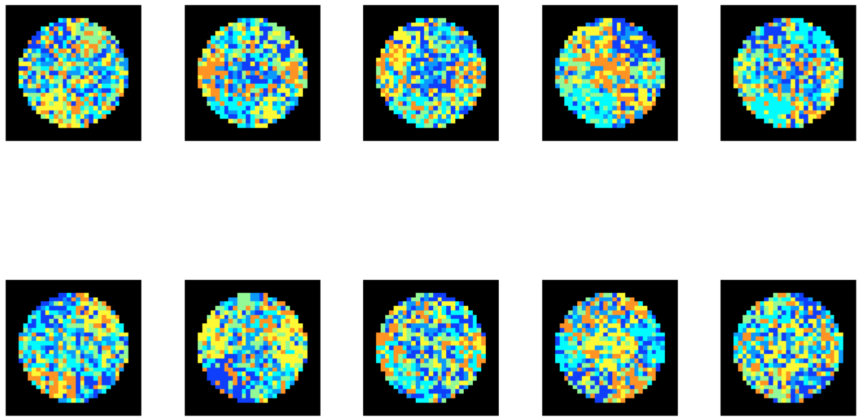


(b) Learning from previous result

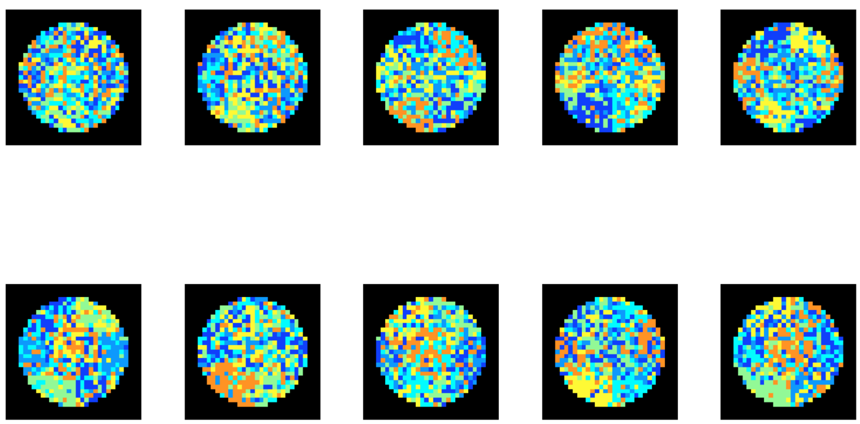


(c) Learning from ground truth

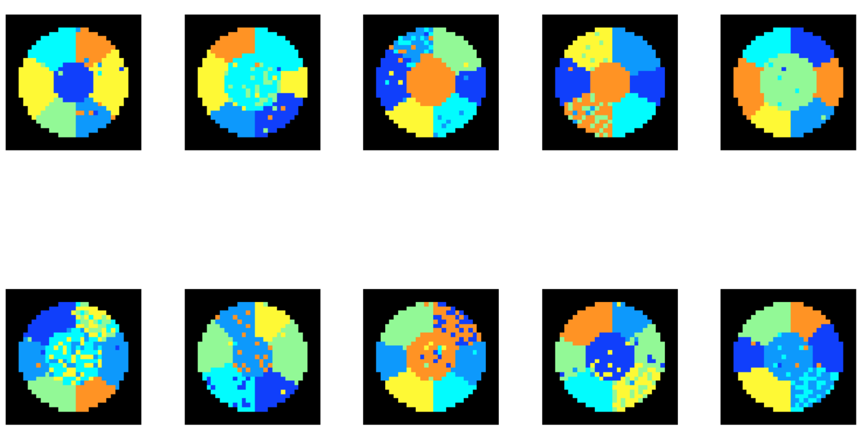
Figure B.3: Clustering results at noise level 0.3



(a) Without learning



(b) Learning from previous result



(c) Learning from ground truth

Figure B.4: Clustering results at noise level (0.4,0.2)