On the Search Algorithm for the Output Distribution that Achieves the Channel Capacity

Kenji Nakagawa, Kohei Watabe, Takuto Sabu
Department of Electrical and Electronics and Information Engineering,
Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan,
E-mail:nakagawa@nagaokaut.ac.jp

***Abstract*—We consider a search algorithm for the output distribution that achieves the channel capacity of a discrete memoryless channel. We will propose an algorithm by iterated projections of an output distribution onto affine subspaces in the set of output distributions. The problem of channel capacity has a similar geometric structure as that of smallest enclosing circle for a finite number of points in the Euclidean space. The metric in the Euclidean space is the Euclidean distance and the metric in the space of output distributions is the Kullback-Leibler divergence. We consider these two problems based on Amari's α-geometry [1]. Then, we first consider the smallest enclosing circle in the Euclidean space and develop an algorithm to find the center of the smallest enclosing circle. Based on the investigation, we will apply the obtained algorithm to the problem of channel capacity.**

I. Introduction

The channel capacity C of a discrete memoryless channel is defined as the maximum of the mutual information. C is also formulated as the solution of a $\min\max$ problem concerned with the Kullback-Leibler divergence [3], [7]. If we replace the Kullback-Leibler divergence with the Euclidean distance, a similar problem in the Euclidean space is obtained. That is the problem of smallest enclosing circle for a finite number of points. In this paper, we will investigate the problem of smallest enclosing circle in the Euclidean space geometrically, and develop an algorithm to compute the solution to the $\min\max$ problem of the Euclidean distance. Then, the resulting algorithm will be applied to the $\min\max$ problem of channel capacity to make an algorithm for calculating the output distribution that achieves the channel capacity.

The geometry of the problem of smallest enclosing circle is the Euclidean geometry and that of channel capacity is the information geometry [1]. We will use only barycentric coordinate, inner product, Pythagorean theorem, and projection onto affine subspaces, as common properties in both geometries, to develop a computation algorithm. The algorithm is called a "projection algorithm."

In this paper, first we consider the algorithm for calculating the center of the smallest enclosing circle for a finite number of points in general position in the Euclidean space. Then, similarly, in the case that the row vectors of the channel matrix are in general position, we consider the algorithm for calculating the output distribution that achieves the channel capacity. We will show that the both problems are solved by common geometric properties. Further, based on the above investigation, we consider the case that the finite number of points and the row vectors are not necessarily in general position. Then, finally we propose heuristic search algorithm and perform the proposed algorithm for randomly generated placements of row vectors. We evaluate the percentage that correct solutions are obtained by our heuristic algorithm.

A. Related Works

There are roughly two categories of calculation methods for the channel capacity, one is solving equations due to Muroga [6] and the other is a sequential calculation method due to Arimoto [2]. In Muroga [6], the input probability distribution that achieves the channel capacity is obtained by solving directly the equations derived from the Lagrange multiplier method. See also Kawabata [5]. In Arimoto's sequential approximation method [2], the channel capacity of an arbitrary channel matrix is calculated numerically by a recurrence formula.

B. Channel matrix and channel capacity

Let us consider a discrete memoryless channel $X \to Y$ with input source X and output source Y. Denote by $\{x_1, \cdots, x_m\}$ the input alphabet and $\{y_1, \cdots, y_n\}$ the output alphabet. The conditional probability P^i_j that y_j is received when x_i was transmitted is denoted by

$$P^i_j = P(Y = y_j | X = x_i),\ i = 1, \cdots, m, j = 1, \cdots, n,$$

and the row vector P^i is defined by

$$P^i = (P^i_1, \cdots, P^i_n),\ i = 1, \cdots, m. \tag{1}$$

The channel matrix Φ is defined by

$$\Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} = \begin{pmatrix} P^1_1 & \cdots & P^1_n \\ \vdots & & \vdots \\ P^m_1 & \cdots & P^m_n \end{pmatrix}. \tag{2}$$

The set $\bar{\Delta}^m$ of all input probability distributions on the input alphabet $\{x_1, \cdots, x_m\}$ is defined by $\bar{\Delta}^m = \{\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_m) | \lambda_i \geq 0, i = 1, \cdots, m, \sum_{i=1}^m \lambda_i = 1\}$. Similarly, the set $\bar{\Delta}^n$ of all output probability distributions on the output alphabet $\{y_1, \cdots, y_n\}$ is defined by $\bar{\Delta}^n = \{Q = (Q_1, \cdots, Q_n) | Q_j \geq 0, j = 1, \cdots, n, \sum_{j=1}^n Q_j = 1\}$.

The output distribution $Q \in \bar{\Delta}^n$ corresponding to the input distribution $\boldsymbol{\lambda} \in \bar{\Delta}^m$ is denoted by $Q = \boldsymbol{\lambda}\Phi$, i.e., $Q_j = \sum_{i=1}^m \lambda_i P^i_j,\ j = 1, \cdots, n$, and the mutual information $I(\boldsymbol{\lambda}, \Phi)$ is defined by $I(\boldsymbol{\lambda}, \Phi) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i P^i_j \log(P^i_j / Q_j)$. Then, the channel capacity C is defined by

$$C = \max_{\boldsymbol{\lambda} \in \bar{\Delta}^m} I(\boldsymbol{\lambda}, \Phi). \tag{3}$$

The channel capacity C is also formulated by

$$C = \min_{Q \in \bar{\Delta}^n} \max_{1 \le i \le m} D(P^i \| Q), \tag{4}$$

where $D(P^i \| Q)$ is the Kullback-Leibler divergence (see [3]).

For some channel matrix, the input distribution $\boldsymbol{\lambda}$ that achieves (3) is not unique, but the output distribution Q that achieves (4) is unique for any channel matrix [3]. By virtue of the uniqueness, it is easy to consider the method of calculating the channel capacity C based on (4) using geometric properties of the Kullback-Leibler divergence.

On the other hand, in order to prove that the resulting output distribution actually achieves C, we will use the convex optimization (3) rather than the geometrical consideration by (4). Concerning (3), the following theorem holds [4].

Theorem: (Kuhn-Tucker condition for the problem of channel capacity) A necessary and sufficient condition for an input distribution $\boldsymbol{\lambda}^* = (\lambda_1^*, \cdots, \lambda_m^*) \in \bar{\Delta}^m$ to achieve the channel capacity C is that there exists a value C_0 with

$$D(P^i \| \boldsymbol{\lambda}^* \Phi) \begin{cases} = C_0, & \text{for } i \text{ with } \lambda_i^* > 0, \\ \le C_0, & \text{for } i \text{ with } \lambda_i^* = 0. \end{cases} \tag{5}$$

Then, C_0 is equal to C.

C. Smallest enclosing circle

Now, replacing $D(P^i \| Q)$ in (4) with the Euclidean distance, we can consider a similar problem in $\mathbb{R}^n$. That is, for $P^1, \cdots, P^m \in \mathbb{R}^n$, let us consider

$$\min_{Q \in \mathbb{R}^n} \max_{1 \le i \le m} d(P^i, Q), \tag{6}$$

where $d(P^i, Q)$ denotes the Euclidean distance between the points P^i and Q in $\mathbb{R}^n$. This is the problem of smallest enclosing circle for the points $P^1, \cdots, P^m$. The purpose of this paper is to study the problem of smallest enclosing circle geometrically and obtain a search algorithm for the optimal solution. Then, through the similarity of (4) and (6), we will apply the resulting algorithm to obtain a search algorithm for the output distribution that achieves the channel capacity.

(4) and (6) not only resemble formally, but have common geometric structures from the view point of Amari's α-geometry [1]. Therefore, if we can develop an algorithm and prove its validity by using only the common properties to both geometries, then an algorithm obtained in one problem can be applied to the other problem almost automatically.

II. Problem of smallest enclosing circle in the Euclidean space

Consider a finite number of points $P^1, \cdots, P^m$ in the n dimensional Euclidean space $\mathbb{R}^n$. The smallest sphere in $\mathbb{R}^n$ that includes these points in its inside or on the boundary is called the *smallest enclosing circle*, and is represented by $\Gamma(P^1, \cdots, P^m)$. The smallest enclosing circle $\Gamma(P^1, \cdots, P^m)$ is formulated by (6). The $Q = Q^*$ that achieves (6) is the center of $\Gamma(P^1, \cdots, P^m)$ and $d^* = \max_{1 \le i \le m} d(P^i, Q^*)$ is its radius.

A. Barycentric coordinate

Let O be the origin of $\mathbb{R}^n$. For the sake of simplicity, we write P^i instead of $\overrightarrow{OP^i}$, $\lambda_1 P^1 + \lambda_2 P^2$ instead of $\lambda_1 \overrightarrow{OP^1} + \lambda_2 \overrightarrow{OP^2}$, $P^2 - P^1$ instead of $\overrightarrow{P^1 P^2}$, and so on. We say m points $P^1, \cdots, P^m \in \mathbb{R}^n$ are *in general position* if the vectors $P^2 - P^1, \cdots, P^m - P^1$ are linearly independent.

For m points $P^1, \cdots, P^m \in \mathbb{R}^n$ in general position, let $L_0 = L(P^1, \cdots, P^m)$ denote the affine subspace spanned by $P^1, \cdots, P^m$. The *barycentric coordinate* of a point $Q \in L_0$ about $P^1, \cdots, P^m$ is defined as the m-tuple of real numbers $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_m)$ with

$$\begin{cases} Q = \lambda_1 P^1 + \cdots + \lambda_m P^m, & (7) \\ \lambda_1 + \cdots + \lambda_m = 1. & (8) \end{cases}$$

The barycentric coordinate $\boldsymbol{\lambda}$ in the problem of smallest enclosing circle corresponds to the input probability $\boldsymbol{\lambda}$ in the problem of channel capacity.

B. Analysis for smallest enclosing circle

A convex optimization problem as a simple optimization problem equivalent to (6) is given as follows.

The coordinates of $P^1, \cdots, P^m$ are denoted by $P^i = (P_1^i, \cdots, P_n^i) \in \mathbb{R}^n$, $i = 1, \cdots, m$ and a matrix Φ with row vectors P^i is defined by

$$\Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} = \begin{pmatrix} P_1^1 & \cdots & P_n^1 \\ \vdots & & \vdots \\ P_1^m & \cdots & P_n^m \end{pmatrix} \in \mathbb{R}^{m \times n}. \tag{9}$$

A vector $\boldsymbol{a}$ is defined by $\boldsymbol{a} = (\|P^1\|^2, \cdots, \|P^m\|^2)$, where $\|P^i\|^2 = \sum_{j=1}^n (P_j^i)^2$ is the squared norm of the vector P^i, $i = 1, \cdots, m$. Then, a function $f(\boldsymbol{\lambda}, \Phi)$ of $\boldsymbol{\lambda} \in \mathbb{R}^m$ associated with Φ is defined by $f(\boldsymbol{\lambda}, \Phi) = \boldsymbol{\lambda}\, {}^t\boldsymbol{a} - \boldsymbol{\lambda} \Phi\, {}^t\Phi\, {}^t\boldsymbol{\lambda}$, where ^{t} denotes the transposition of vector or matrix. $f(\boldsymbol{\lambda}, \Phi)$ is a differentiable and convex upward function of $\boldsymbol{\lambda}$. The convex optimization problem

$$\max_{\boldsymbol{\lambda} \in \bar{\Delta}^m} f(\boldsymbol{\lambda}, \Phi) \tag{10}$$

is equivalent to the problem of smallest enclosing circle [9]. For $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ that achieves (10), $Q^* \equiv \boldsymbol{\lambda}^* \Phi$ is the center of the smallest enclosing circle $\Gamma(P^1, \cdots, P^m)$ and $d^* \equiv \sqrt{f(\boldsymbol{\lambda}^*, \Phi)}$ is its radius [9].

For (10), the following theorem holds [4],[9].

Theorem: (Kuhn-Tucker condition for the problem of smallest enclosing circle) A necessary and sufficient condition for $\boldsymbol{\lambda}^* = (\lambda_1^*, \cdots, \lambda_m^*) \in \bar{\Delta}^m$ to achieve (10) is that there exists d_0 with

$$d(P^i, \boldsymbol{\lambda}^* \Phi) \begin{cases} = d_0, & \text{for } i \text{ with } \lambda_i^* > 0, \\ \le d_0, & \text{for } i \text{ with } \lambda_i^* = 0. \end{cases} \tag{11}$$

Then, $Q^* = \boldsymbol{\lambda}^* \Phi$ is the center of the smallest enclosing circle $\Gamma(P^1, \cdots, P^m)$ and $d^* = d_0$ is its radius.

C. Equidistant point from $P^1, \cdots, P^m$ and its barycentric coordinate

For $P^1, \cdots, P^m \in \mathbb{R}^n$, the *equidistant point* from $P^1, \cdots, P^m$ is a point $Q^0 \in L_0 = L(P^1, \cdots, P^m)$ that satisfies $d(P^1, Q^0) = \cdots = d(P^m, Q^0)$.

Now, we assume in this chapter that $P^1 \cdots, P^m$ are in general position.

D. Inner product, Pythagorean theorem and projection in $\mathbb{R}^n$

We will use the inner product, Pythagorean theorem and projection in $\mathbb{R}^n$, which are important to determine the solution of the problem of smallest enclosing circle by the projection algorithm (see [8] for details).

1) Projection: For a point $Q' \in \mathbb{R}^n$ and a subset $L \subset \mathbb{R}^n$, the point $Q = Q''$ that achieves $\min_{Q\in L} d(Q, Q')$ is called the *projection* of Q' onto L, and is denoted by by $Q'' = \pi(Q'|L)$. In this paper, we consider only affine subspaces of $\mathbb{R}^n$ as L. The projection onto an affine subspace is easily calculated because there is no inequality constraint.

2) Projection of equidistant point: We define $L_k = L(P^{k+1}, \cdots, P^m)$, $k = 0, \cdots, m-2$ as the affine subspace spanned by $P^{k+1}, \cdots, P^m$. $L_0 \supset L_1 \supset \cdots \supset L_k \supset \cdots$ is a decreasing sequence of affine subspaces whose dimensions are decreasing by 1. For the equidistant point Q^0 from P^1, $\cdots, P^m$, define $Q^k = \pi(Q^{k-1}|L_k)$, $k = 1, \cdots, m-2$. We see that Q^k is the equidistant point from $P^{k+1}, \cdots, P^m$.

E. Search for Q^* by projection algorithm

To calculate Q^*, a series of iterated projections of Q^0 onto affine subspaces is called *projection algorithm*.

For given points $P^1, \cdots, P^m$, let Q^* be the center of the smallest enclosing circle $\Gamma = \Gamma(P^1, \cdots, P^m)$ and d^* be its radius. In this chapter, we are assuming that $P^1, \cdots, P^m$ are in general position, and then we will show that Q^*, d^* are calculated by the projection algorithm in some situations defined below.

F. Situation 1 [*There is just one negative component of barycentric coordinate at every projection.*]

First, let us consider the following example.

Example 1: Consider four points $P^1 = (-10, 1, -3)$, $P^2 = (-9, -2, 8)$, $P^3 = (-8, 10, -5)$, $P^4 = (4, -8, 8)$ given in $\mathbb{R}^3$. $P^1, \cdots, P^4$ are in general position. The barycentric coordinate $\boldsymbol{\lambda}^0$ of the equidistant point Q^0 from $P^1, \cdots, P^4$ is $\boldsymbol{\lambda}^0 = (-0.84, 0.04, 1.11, 0.69)$. Since $\lambda_1^0 = -0.84 < 0$, we remove P^1 and calculate the projection of Q^0 onto $L_1 = L(P^2, P^3, P^4)$, i.e., $Q^1 = \pi(Q^0|L_1)$. The barycentric coordinate $\boldsymbol{\lambda}^1$ of Q^1 is $\boldsymbol{\lambda}^1 = (0, -0.31, 0.63, 0.68)$. Since $\lambda_2^1 = -0.31 < 0$, we remove P^2 and calculate the projection Q^2 of Q^1 onto $L_2 = L(P^3, P^4)$, i.e., $Q^2 = \pi(Q^1|L_2)$. The barycentric coordinate $\boldsymbol{\lambda}^2$ of Q^2 is $\boldsymbol{\lambda}^2 = (0, 0, 0.5, 0.5)$. Then, we have $d(P^1, Q^2) = 9.18$, $d(P^2, Q^2) = 10.01$, $d(P^3, Q^2) = d(P^4, Q^2) = 12.62$, so by the Kuhn-Tucker condition (11), we see that Q^2 is the center of the smallest enclosing circle, i.e., $Q^* = Q^2$, and thus, $d^* = d(P^3, Q^2)$.

In this section, we represent "situation 1" as the case that there is just one negative component of barycentric coordinate at every projection like $\boldsymbol{\lambda}^0$ and $\boldsymbol{\lambda}^1$ above, and all the components are non-negative at the last projection like $\boldsymbol{\lambda}^2$.

Assumption of situation 1 Assume $P^1, \cdots, P^m \in \mathbb{R}^n$ are in general position, and let $L_k = L(P^{k+1}, \cdots, P^m)$, $k = 0, 1, \cdots, m-2$ be the affine subspace spanned by P^{k+1}, $\cdots, P^m$. Let $Q^0 \in L_0$ be the equidistant point from $P^1, \cdots, P^m$, and define $Q^k = \pi(Q^{k-1}|L_k)$, $k = 1, \cdots, m-2$. The barycentric coordinate of Q^k about $P^1, \cdots, P^m$ is denoted by $\boldsymbol{\lambda}^k = (\lambda_1^k, \cdots, \lambda_m^k)$. Let $K = 0, 1, \cdots, m-2$. We assume that for $k = 0, 1, \cdots, K-1$, there is just one negative component of $\boldsymbol{\lambda}^k$, and for $k = K$, all the components of $\boldsymbol{\lambda}^K$ are non-negative.

Theorem 1: (See [8].) In the situation 1, the center of the smallest enclosing circle Γ is $Q^* = Q^K$, and its radius is $d^* = d(P^{K+1}, Q^K)$.

G. Situation 2 [$m = 2, 3, 4$ *and* n *is arbitrary.*]

We will calculate Q^* and d^* by the projection algorithm in the case that $m = 2, 3, 4$ and n is arbitrary. Of course our goal is to find an algorithm to calculate Q^* and d^* for every m, but at present it is possible to solve only for $m = 2, 3, 4$.

Theorem 2: (See [8].) In the situation 2, the center Q^* and the radius d^* of the smallest enclosing circle Γ is obtained by the projection algorithm.

III. Problem of Channel Capacity

In this chapter, we will consider the problem of channel capacity geometrically based on the results in chapter II, to exploit a projection algorithm of searching for the output distribution that achieves the channel capacity.

A. Information geometry

The underlying geometry of the problem of channel capacity is the information geometry [1], not the Euclidean geometry. Amari [1] investigated α-geometry for real α, which is a family of geometric structures. The Euclidean geometry corresponds to $\alpha = 0$ and the geometry of Δ^n corresponds to $\alpha = \pm 1$, so, they can be regarded as a special case of α-geometry. In α-geometry, α-divergence, inner product, Pythagorean theorem, α-projection can be used. In the proof of theorems for the problem of smallest enclosing circle, we only used the Euclidean distance, inner product, Pythagorean theorem and projection among the properties of the Euclidean geometry. Thus, the resulting theorems or algorithms are expected to apply easily to the problem of channel capacity. In fact, we show it in the following.

B. Geometric structure on Δ^n

Let Δ^n be the set of probability distributions with positive components on the output alphabet $\{y_1, \cdots, y_n\}$, i.e., $\Delta^n = \{Q = (Q_1, \cdots, Q_n) | Q_j > 0, j = 1, \cdots, n, \sum_{j=1}^n Q_j = 1\}$. Geometric structures, such as dual coordinate systems, geodesic, inner product, Pythagorean theorem and projection by Kullback-Leibler divergence are introduced on Δ^n (see [1], [8]).

1) Projection by Kullback-Leibler divergence: For $Q' \in \Delta^n$ and a subset $L \subset \Delta^n$, the $Q = Q''$ that achieves $\min_{Q\in L} D(Q\|Q')$ is called the *projection* of Q' onto L, and denoted by $Q'' = \pi(Q'|L)$. In this paper, we consider only affine subspaces as L. The projection onto an affine subspace is easily calculated because there is no inequality constraint.

C. Equidistant point and projection of equidistant point

We will calculate the channel capacity C and the output distribution Q^* that achieves C, in case that the row vectors $P^1, \cdots, P^m$ of the channel matrix Φ in (2) are in general position. For $P^1, \cdots, P^m \in \Delta^n$, let $L_0 = L(P^1, \cdots, P^m) \subset \Delta^n$ be the affine subspace spanned by $P^1, \cdots, P^m$ (see [8]). $Q^0 \in L_0 = L(P^1, \cdots, P^m)$ with $D(P^i\|Q^0) = D(P^1\|Q^0)$, $i = 2, \cdots, m$ is called the *equidistant point* from $P^1, \cdots, P^m$. The existence and uniqueness of Q^0 is guaranteed (see [8]).

TABLE I
CORRESPONDENCE OF SYMBOLS BETWEEN SMALLEST ENCLOSING CIRCLE AND CHANNEL CAPACITY

	smallest enclosing circle	channel capacity
$\boldsymbol{\lambda}$	barycentric coordinate	input distribution
$P^1, \cdots, P^m$	given points $\in \mathbb{R}^n$	output distributions $\in \Delta^n$ for input symbols $x_1, \cdots, x_m$
Φ	$\begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix}$ matrix of given points	$\begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix}$ channel matrix
convex function	$f(\boldsymbol{\lambda}, \Phi) = \boldsymbol{\lambda}\,{}^t\boldsymbol{a} - \boldsymbol{\lambda}\Phi\,{}^t\Phi\,{}^t\boldsymbol{\lambda}$ $\boldsymbol{a} = (\|P^1\|^2, \cdots, \|P^m\|^2)$	$I(\boldsymbol{\lambda}, \Phi)$ mutual information
metric	$d(P, Q)$ Euclidean distance	$D(P\|Q)$ Kullback-Leibler divergence
Kuhn-Tucker condition	$d(P^i, \boldsymbol{\lambda}^*\Phi)$ $\begin{cases} = d_0, \text{if } \lambda_i^* > 0 \\ \le d_0, \text{if } \lambda_i^* = 0 \end{cases}$	$D(P^i\|\boldsymbol{\lambda}^*\Phi)$ $\begin{cases} = C_0, \text{if } \lambda_i^* > 0 \\ \le C_0, \text{if } \lambda_i^* = 0 \end{cases}$
Q^*	center of smallest enclosing circle	capacity achieving output distribution
inner product	$\sum_{j=1}^{n}(Q_j^1 - Q_j^2)(Q_j^3 - Q_j^2)$	$\sum_{j=2}^{n}(\eta_j^1 - \eta_j^2)(\theta_j^3 - \theta_j^2)$
Pythagorean theorem	$d^2(P, Q) + d^2(Q, R) = d^2(P, R)$	$D(P\|Q) + D(Q\|R) = D(P\|R)$
projection	$\min_{P \in L} d(P, Q)$	$\min_{P \in L} D(P\|Q)$

1) Projection of equidistant point: Now, we further define $L_k = L(P^{k+1}, \cdots, P^m)$, $k = 0, 1, \cdots, m-2$. $L_0 \supset L_1 \supset \cdots \supset L_k \supset \cdots$ is a decreasing sequence of affine subspaces whose dimensions are decreasing by 1. Let us define $Q^k = \pi(Q^{k-1}|L_k)$, $k = 1, \cdots, m-2$. We see that Q^k is the equidistant point from $P^{k+1}, \cdots, P^m$.

D. Search for Q^* by projection algorithm

For a given channel matrix Φ in (2), let C be the channel capacity and Q^* be the output distribution that achieves C. In this section, similarly to chapter II, we will show that Q^* and C are obtained by the projection algorithm, under the assumption that the row vectors of the channel matrix Φ are in general position and in the situations 1 and 2, which are defined as in the case of smallest enclosing circle.

Theorem 3: (See [8].) In the situation 1, we have $Q^* = Q^K$ and $C = D(P^{K+1}\|Q^K)$.

Theorem 4: (See [8].) In the situation 2, Q^* and C are obtained by the projection algorithm.

In chapters II and III, we used common symbols both in smallest enclosing circle and channel capacity. Then, we will show in TABLE I the correspondence of symbols between them.

IV. SEARCH FOR OPTIMAL SOLUTION FOR ARBITRARY PLACEMENT OF POINTS

In the previous chapters, we assumed that the given points $P^1, \cdots, P^m \in \mathbb{R}^n$ or $\in \Delta^n$ are in general position. In this chapter, we assume no constraint on the placement of the points $P^1, \cdots, P^m$.

We will consider a method of giving a little deformation to the placement of given points $P^1, \cdots, P^m$ so that the points of the deformed placement are in general position, then apply the projection algorithm.

A. Method of lifting dimension of point in $\mathbb{R}^n$

For m points $P^1, \cdots, P^m$ in $\mathbb{R}^n$, define the points $\widetilde{P}^1, \cdots, \widetilde{P}^m \in \mathbb{R}^{n+m}$ by lifting the dimension as follows:

$$\begin{aligned} P^1 = (P_1^1, \cdots, P_n^1) &\to \widetilde{P}^1 = (P_1^1, \cdots, P_n^1, \varepsilon, 0, \cdots, 0), \\ P^2 = (P_1^2, \cdots, P_n^2) &\to \widetilde{P}^2 = (P_1^2, \cdots, P_n^2, 0, \varepsilon, \cdots, 0), \\ &\vdots \\ P^m = (P_1^m, \cdots, P_n^m) &\to \widetilde{P}^m = (P_1^m, \cdots, P_n^m, 0, \cdots, 0, \varepsilon), \end{aligned}$$

where $\varepsilon \in \mathbb{R}$, $\varepsilon \neq 0$, and $|\varepsilon|$ is sufficiently small. Then, $\widetilde{P}^1, \cdots, \widetilde{P}^m$ are in general position.

Suppose we had the center $\widetilde{Q}^*$ of the smallest enclosing circle for $\widetilde{P}^1, \cdots, \widetilde{P}^m$ by the projection algorithm in chapter II, then for small ε, $\widetilde{Q}^*$ is close to the true Q^*, which is the center of the smallest enclosing circle for $P^1, \cdots, P^m$. Denote by $\widetilde{\boldsymbol{\lambda}}^*$ the barycentric coordinate of $\widetilde{Q}^*$ about $\widetilde{P}^1, \cdots, \widetilde{P}^m$, and $\widetilde{\boldsymbol{\lambda}}^*\big|_{\varepsilon=0}$ by substituting $\varepsilon = 0$ in $\widetilde{\boldsymbol{\lambda}}^*$. Then $Q^* = \widetilde{\boldsymbol{\lambda}}^*\big|_{\varepsilon=0}\Phi$ is expected to be the center of the original smallest enclosing circle. Examples for this method are given in [8] (see Example 2 and 3 in [8]).

B. Lifting dimension of channel matrix

In this section, for the row vectors $P^1, \cdots, P^m$ of the channel matrix Φ in (2), we will define $\widetilde{P}^1, \cdots, \widetilde{P}^m$ by lifting the dimension so that they are in general position.

Now, define $\widetilde{P}^i \in \Delta^{n+2m}$ by

$$\widetilde{P}^i = \Big(\frac{P_1^i}{2m+1}, \cdots, \frac{P_n^i}{2m+1}, \overbrace{\frac{1}{2m+1}, \cdots, \frac{1}{2m+1}}^{2i-2}, \underset{2i-1\text{ th}}{\frac{1+\epsilon}{2m+1}}, \underset{2i\text{ th}}{\frac{1-\epsilon}{2m+1}}, \overbrace{\frac{1}{2m+1}, \cdots, \frac{1}{2m+1}}^{2m-2i}\Big), \quad (12)$$

where $|\varepsilon|$ is sufficiently small and $\varepsilon \neq 0$. Then, $\widetilde{P}^1, \cdots, \widetilde{P}^m$ are in general position.

Example 2: Consider the channel matrix

$$\Phi = \begin{pmatrix} P^1 \\ P^2 \\ P^3 \\ P^4 \end{pmatrix} = \begin{pmatrix} 2/5 & 2/5 & 1/5 \\ 1/3 & 1/3 & 1/3 \\ 4/5 & 1/10 & 1/10 \\ 1/10 & 4/5 & 1/10 \end{pmatrix}, \quad (13)$$

where the rows are not in general position. By lifting the dimension, we have

$$\begin{aligned} \widetilde{P}^1 &= (2/5, 2/5, 1/5, 1+\varepsilon, 1-\varepsilon, 1, 1, 1, 1, 1, 1)/9, \\ \widetilde{P}^2 &= (1/3, 1/3, 1/3, 1, 1, 1+\varepsilon, 1-\varepsilon, 1, 1, 1, 1)/9, \\ \widetilde{P}^3 &= (4/5, 1/10, 1/10, 1, 1, 1, 1, 1+\varepsilon, 1-\varepsilon, 1, 1)/9, \\ \widetilde{P}^4 &= (1/10, 4/5, 1/10, 1, 1, 1, 1, 1, 1, 1+\varepsilon, 1-\varepsilon)/9. \end{aligned}$$

Setting $\varepsilon = 0.05$, we have the barycentric coordinate $\widetilde{\boldsymbol{\lambda}}^0$ of the equidistant point $\widetilde{Q}^0$ from $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3, \widetilde{P}^4$ as $\widetilde{\boldsymbol{\lambda}}^0 = (-19.00, 7.98, 6.01, 6.01)$. Since $\widetilde{\lambda}_1^0 = -19.00 < 0$, we remove $\widetilde{P}^1$ and consider the projection $\widetilde{Q}^1 = \pi(\widetilde{Q}^0 | L(\widetilde{P}^2, \widetilde{P}^3, \widetilde{P}^4))$. We have the barycentric coordinate of $\widetilde{Q}^1$ as $\widetilde{\boldsymbol{\lambda}}^1 = (0, -0.14, 0.57, 0.57)$. Since $\widetilde{\boldsymbol{\lambda}}_2^1 = -0.14 < 0$, we further remove $\widetilde{P}^2$ and consider the projection $\widetilde{Q}^2 = \pi(\widetilde{Q}^1 | L(\widetilde{P}^3, \widetilde{P}^4))$. We have the barycentric coordinate of $\widetilde{Q}^2$ as $\widetilde{\boldsymbol{\lambda}}^2 = (0, 0, 1/2, 1/2)$. Thus, this case is the situation 1, so we have $\widetilde{\boldsymbol{\lambda}}^* = \widetilde{\boldsymbol{\lambda}}^2$. Therefore, $Q^* = \widetilde{\boldsymbol{\lambda}}^* \big|_{\varepsilon=0} \Phi = (9/20, 9/20, 1/10)$, and the channel capacity is $C = D(P^3 \| Q^*) = 0.447$ [bit/symbol]. In this case, the capacity achieving Q^* and the capacity C of Φ are correctly obtained.

V. Heuristic algorithm of calculating channel capacity

In this chapter, we will propose heuristic projection algorithm for the calculation of the channel capacity with arbitrary m, n and arbitrary placement of row vectors, based on the algorithms in chapter II, III and the dimension lifting in chapter IV. The method which will be proposed in this chapter can be applied to any number of row vectors and any placement of row vectors, but the obtained results are not always correct. In the following, we describe the heuristic algorithm and apply them to many concrete problems generated by random numbers, then show the percentage of getting correct solutions.

A. Heuristic algorithm for Channel capacity (HC)

1) For m probability distributions $P^1, \cdots, P^m$ of Φ, lift the dimension by (12) to obtain $\widetilde{P}^1, \cdots, \widetilde{P}^m$.
2) Calculate the barycentric coordinate $\widetilde{\boldsymbol{\lambda}}^0 = (\widetilde{\lambda}_1^0, \cdots, \widetilde{\lambda}_m^0)$ of the equidistant point from $\widetilde{P}^1, \cdots, \widetilde{P}^m$.
3) If $\widetilde{\lambda}_i^0 \geq 0$, $i = 1, \cdots, m$, then end the algorithm and output $\widetilde{Q}^0$.
4) If some of $\widetilde{\lambda}_i^0$ are negative, consider the smallest one, i.e., the negative one with maximum absolute value, say, it is supposed to be $\widetilde{\lambda}_1^0$. Then remove $\widetilde{P}^1$ and leave $\widetilde{P}^2, \ldots, \widetilde{P}^m$.
5) For $m-1$ points $\widetilde{P}^2, \ldots, \widetilde{P}^m$, repeat the algorithm from 2).

The simulation results of the above algorithm for randomly generated 10000 channel matices are shown in TABLE II. In every case of TABLE II, the percentage of success is more than 99%.

A heuristic algorithm for smallest enclosing circle is also given in [8].

TABLE II
Simulation results of heuristic algorithm for channel capacity (HC)

m \ n	2	3	10	20
3	99.9%	99.9%	99.9%	99.9%
4	99.9%	99.9%	99.9%	99.9%
5	99.9%	99.8%	99.9%	99.9%
8	99.9%	99.8%	99.9%	99.9%
10	99.8%	99.5%	99.8%	99.9%

VI. Conclusion

Since the Euclidean geometry is familiar to us, it is easy to develop an algorithm and to prove the correctness of the obtained algorithm. A proposition in the Euclidean geometry can be proved in many ways because there are many tools that we can use. However, if we consider the corresponding proposition in the information geometry, all the proofs in the Euclidean geometry are not necessarily applicable to the proposition. We found that there is one natural proof in the Euclidean geometry that is directly applicable to the corresponding proposition of the information geometry. Such a proof uses only common properties of both Euclidean and information geometries. We also found that the natural proof is the simplest proof. This is a very interesting result. If we considered only the problem of channel capacity from the beginning, I think we could not obtain the projection algorithm developed in this paper. It is a result of a successful link between the Euclidean and information geometries.

VII. Future works

(1) Make an algorithm to calculate Q^* for arbitrary number m of points in general position.
(2) Make a projection algorithm for the rate distortion function and the capacity constraint function.
(3) Transplant the Arimoto algorithm to the problem of smallest enclosing circle.

References

[1] S.Amari, *Differential-Geometrical Methods in Statistics,* Lecture Notes in Statistics, Springer-Verlag, 1985.
[2] Arimoto, S., "An algorithm for computing the capacity of arbitrary discrete memoryless channels," IEEE Trans. Inform. Theory, vol.IT-18, pp.14-20, 1972.
[3] Csiszàr, I. and Körner, J., *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Orlando, 1982.
[4] Gallager, G., *Information Theory and Reliable Communication*, Wiley, New York, 1968.
[5] Kawabata, T., "A Revisit to The Muroga Method of Computing Channel Capacity," in Proceedings of ISITA2010, pp.801-806, 2010.
[6] Muroga, S., "On the capacity of a discrete channel. I," J. Phys. Soc. Japan, vol.8, pp.484-494, 1953.
[7] Nakagawa, K. and Kanaya, F., "A New Geometric Capacity Characterization of a Discrete Memoryless Channel," IEEE Trans. Inform. Theory, vol.34, no.2, pp.318-321, 1988.
[8] Nakagawa, K., Watabe, K. and Sabu, T., "On the Search Algorithm for the Output Distribution that Achieves the Channel Capacity," arXiv:1601.01394v2.
[9] Schönherr, S., "Quadratic Programming in Geometric Optimization: Theory, Implementation, and Applications," Dissertation of Swiss Federal Institute of Technology, http://www.inf.ethz.ch/personal/emo/DoctThesisFiles/schoenherr02.pdf