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

Kenji Nakagawa, *Member, IEEE*, Kohei Watabe, *Member, IEEE*, and Takuto Sabu

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. 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.

Index Terms—Channel capacity, discrete memoryless channel, smallest enclosing circle, information geometry, projection algorithm.

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 [4], [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 reason for taking such an approach is because the Euclidean geometry is familiar to us, so it may be easier to make new geometric algorithms.

As mentioned above, the problem of channel capacity is described by an optimization problem concerned with the output probability distributions. The geometry on the set of output probability distributions is the information geometry [1]. The Euclidean geometry and the information geometry can be considered in a unified manner from the viewpoint of α-geometry by Amari [1]. Therefore, the solution algorithm for the problem of smallest enclosing circle can be applied to the problem of channel capacity through this geometric similarity. For that purpose, it is necessary to use only the applicable properties to the problem of channel capacity. If it is achieved, then an algorithm obtained in the smallest enclosing circle can be transplanted almost automatically to the channel capacity. In this paper, we will actually 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 algorithms and perform the proposed algorithm for randomly generated placements of points and 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 [6] and the other is a sequential calculation method due to Arimoto [2]. In [6], the input probability distribution that achieves the channel capacity is obtained by solving directly the equations derived from the Lagrange multiplier method. In this case, we should take care of the inequality condition that the probability takes a non-negative value. Simply solving the equations by ignoring this inequality condition may yield a solution of a negative probability. Of course this is not a correct solution. The inequality condition makes the problem difficult, however, we can say this difficulty makes the problem attractive. In [6],

if a solution includes a negative probability, one of the input symbols should be removed and solving equations should be continued repeatedly until all the probabilities become non-negative. Therefore, if all the possibilities are exhausted, the total number of equations to be solved is exponential.

Our calculation method of the channel capacity C proposed in this paper belongs to the same category as [6]. For the comparison of our method with Muroga's, we will evaluate the percentage of the correctly obtained values of the channel capacity by our heuristic algorithm and further, compare the computational complexity of two methods.

In Arimoto's sequential approximation method [2], the channel capacity of an arbitrary channel matrix is calculated numerically by a recurrence formula. This method does not yield a negative probability, but corresponding to it, if some input probability becomes 0, an exceptional treatment is required in the calculation of a reverse channel. Both in [2] and [6], it is important to determine which input symbol has a positive probability. That is a main subject of this paper.

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_j^i that y_j is received when x_i was transmitted is denoted by

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

and the row vector P^i is defined by

$$P^i = (P_1^i, \cdots, P_n^i), \quad i = 1, \cdots, m. \tag{1}$$

The channel matrix Φ is defined by

$$\Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} = \begin{pmatrix} P_1^1 & \cdots & P_n^1 \\ \vdots & & \vdots \\ P_1^m & \cdots & P_n^m \end{pmatrix}. \tag{2}$$

The set $\bar{\Delta}^m$ of all input probability distributions on the input alphabet $\{x_1, \cdots, x_m\}$ is defined by

$$\bar{\Delta}^m = \{\lambda = (\lambda_1, \cdots, \lambda_m) | \lambda_i \geq 0, i = 1, \cdots, m, \sum_{i=1}^m \lambda_i = 1\}.$$

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 $\lambda \in \bar{\Delta}^m$ is denoted by $Q = \lambda\Phi$, i.e., $Q_j = \sum_{i=1}^m \lambda_i P_j^i$, $j = 1, \cdots, n$, and the mutual information $I(\lambda, \Phi)$ is defined by

$$I(\lambda, \Phi) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i P_j^i \log \frac{P_j^i}{Q_j}. \tag{3}$$

Then, the channel capacity C is defined by

$$C = \max_{\lambda \in \bar{\Delta}^m} I(\lambda, \Phi). \tag{4}$$

For two output distributions $Q = (Q_1, \cdots, Q_n)$, $Q' = (Q_1', \cdots, Q_n') \in \bar{\Delta}^n$, the Kullback-Leibler divergence $D(Q \| Q')$ is defined by

$$D(Q \| Q') = \sum_{j=1}^n Q_j \log \frac{Q_j}{Q_j'}, \tag{5}$$

see [4]. The channel capacity C is also formulated by the Kullback-Leibler divergence as follows [4]:

$$C = \min_{Q \in \bar{\Delta}^n} \max_{1 \leq i \leq m} D(P^i \| Q). \tag{6}$$

For some channel matrix, the input distribution λ that achieves (4) is not unique, but the output distribution Q that achieves (6) is unique for any channel matrix [4]. By virtue of the uniqueness, it is easier to consider the method of calculating the channel capacity C based on (6) 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 (4) rather than the geometrical consideration by (6). Concerning (4), the following Kuhn-Tucker condition holds [5].

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

$$D(P^i \| \lambda^* \Phi) \begin{cases} = C_0, & \text{for } i \text{ with } \lambda_i^* > 0, \\ \leq C_0, & \text{for } i \text{ with } \lambda_i^* = 0. \end{cases} \tag{7}$$

Then, C_0 is equal to C.

C. Smallest Enclosing Circle

Now, replacing $D(P^i \| Q)$ in (6) 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

$$\min_{Q \in \mathbb{R}^n} \max_{1 \leq i \leq m} d(P^i, Q), \tag{8}$$

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 set of 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 (6) and (8), we will apply the resulting algorithm to obtain a search algorithm for the output distribution that achieves the channel capacity.

Then, first, let us consider the problem of smallest enclosing circle in the Euclidean space.
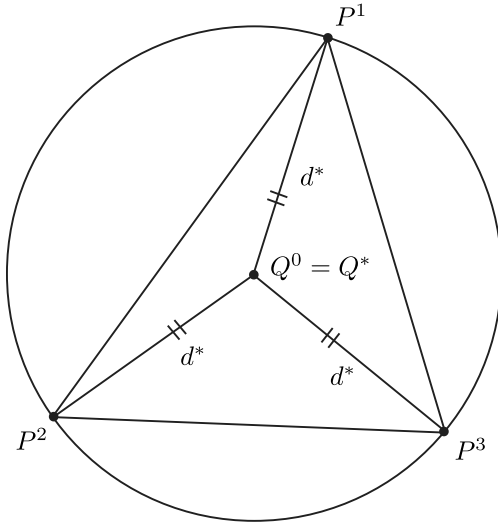
Fig. 1.   Smallest enclosing circle for acute triangle $\triangle P^1 P^2 P^3$.
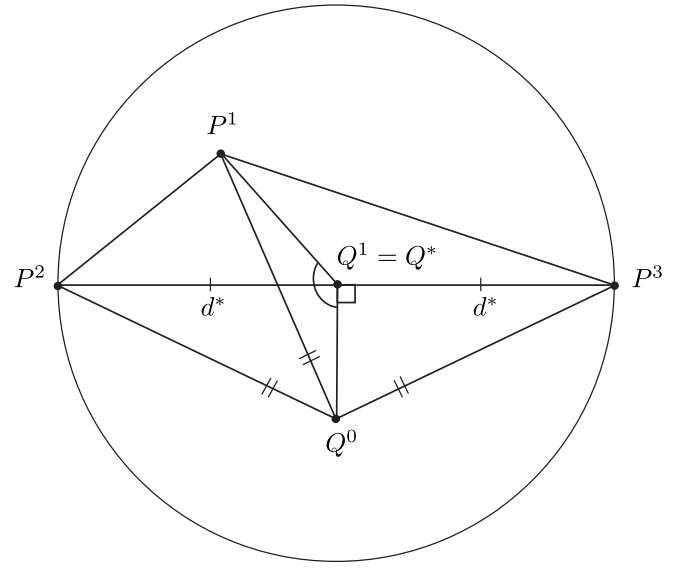
Fig. 2.   Smallest enclosing circle for obtuse triangle $\triangle P^1 P^2 P^3$.

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 (8). The $Q = Q^*$ that achieves (8) is the center of $\Gamma(P^1, \cdots, P^m)$ and $d^* = \max_{1 \leq i \leq m} d(P^i, Q^*)$ is its radius.

A. Equidistant Point and Projection

As a simplest example, let us consider the center Q^* and the radius d^* of the smallest enclosing circle $\Gamma = \Gamma(P^1, P^2, P^3)$ for three points P^1, P^2, P^3 in $\mathbb{R}^2$. We will investigate separately in cases that $\triangle P^1 P^2 P^3$ is an acute triangle and obtuse triangle.

(I) Case that $\triangle P^1 P^2 P^3$ is an acute triangle (see Fig.1).

In this case, the circumcenter of $\triangle P^1 P^2 P^3$, i.e., the equally distant point Q^0 from P^1, P^2, P^3 is the center of the smallest enclosing circle Γ and $d(P^1, Q^0)(= d(P^2, Q^0) = d(P^3, Q^0))$ is its radius d^*.

From this example, we find it valid to consider the equidistant point Q^0 from the given points $P^1, \cdots, P^m$.

(II) Case that $\triangle P^1 P^2 P^3$ is an obtuse triangle (see Fig.2).

In this case, assuming $\angle P^1$ is an obtuse angle, we see that the center Q^* of the smallest enclosing circle Γ is the midpoint of the side $P^2 P^3$, and the radius d^* is equal to $d(P^2, Q^*)$ $(= d(P^3, Q^*))$. The equidistant point Q^0 from P^1, P^2, P^3 exists, however, Q^0 is not the center of the smallest enclosing circle because Q^0 is outside of $\triangle P^1 P^2 P^3$. Then, defining Q^1 as the nearest point from Q^0 on the straight line $L(P^2, P^3)$ passing through P^2, P^3, we write it as $Q^1 = \pi(Q^0|L(P^2, P^3))$. Q^1 is called the projection of Q^0 onto $L(P^2, P^3)$. Q^1 is the midpoint of the side $P^2 P^3$, therefore, we have $Q^* = Q^1$.

From this example, we find it valid to consider the projection of the equidistant point Q^0 onto some set if Q^0 is outside of $\triangle P^1 P^2 P^3$. In this paper, we call it the *projection algorithm* that calculates the center Q^* of the smallest enclosing circle (or the output distribution Q^* achieving the channel capacity) by the projection onto a straight line passing through two points or an affine subspace spanned by plural points.

B. Barycentric Coordinate

Let O be the origin of $\mathbb{R}^n$. Henceforth, for the sake of simplicity, we write P^i instead of $\overrightarrow{OP^i}$, $\lambda_1 P^1 + \lambda_2 P^2$ instead of $\lambda_1 \overrightarrow{OP^1} + \lambda_2 \overrightarrow{OP^2}$, and $P^2 - P^1$ instead of $\overrightarrow{P^1 P^2}$, and so on. Depending on the case, we consider P^i as a point in $\mathbb{R}^n$, or as a vector $\overrightarrow{OP^i}$.

We say that 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, or

$$\text{rank} \begin{pmatrix} P^2 - P^1 \\ \vdots \\ P^m - P^1 \end{pmatrix} = m - 1. \qquad (9)$$

Let $L_0 = L(P^1, \cdots, P^m)$ denote the affine subspace spanned by $P^1, \cdots, P^m$, i.e., the minimum affine subspace including $P^1, \cdots, P^m$, then we have $\dim L_0 = m - 1$ under the condition (9).

We will use the barycentric coordinate to represent the position of a point in L_0. Consider m points $P^1, \cdots, P^m \in \mathbb{R}^n$ in general position. 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

$$\begin{cases} Q = \lambda_1 P^1 + \cdots + \lambda_m P^m, & (10) \\ \lambda_1 + \cdots + \lambda_m = 1. & (11) \end{cases}$$

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.

C. Analysis for Smallest Enclosing Circle

The smallest enclosing circle $\Gamma(P^1, \cdots, P^m)$ for $P^1, \cdots, P^m \in \mathbb{R}^n$ is obtained by solving the min max problem (8). In (8), the problem is expressed by the distance $d(P^i, Q)$, so it is easily understood geometrically. Therefore, it is possible to develop a solution algorithm by geometric considerations. In fact, in this paper, we will develop new algorithms based on (8). However, in order to prove that the algorithm is correct, the double optimization problem as (8) is difficult to apply. Then, a convex optimization problem (convex programming) as a simple optimization problem equivalent to (8) is given as follows.

The coordinates of $P^1, \cdots, P^m$ are defined by

$$P^i = (P_1^i, \cdots, P_n^i) \in \mathbb{R}^n, \quad i = 1, \cdots, m, \qquad (12)$$

and a matrix Φ with row vectors P^i is defined by

$$\Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} = \begin{pmatrix} P_1^1 & \cdots & P_n^1 \\ \vdots & & \vdots \\ P_1^m & \cdots & P_n^m \end{pmatrix} \in \mathbb{R}^{m \times n}. \qquad (13)$$

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

$$f(\boldsymbol{\lambda}, \Phi) = \boldsymbol{\lambda}\,^t\boldsymbol{a} - \boldsymbol{\lambda}\Phi\,^t\Phi\,^t\boldsymbol{\lambda}, \qquad (14)$$

where, ^{t} denotes the transposition of vector or matrix. $f(\boldsymbol{\lambda}, \Phi)$ is a differentiable and convex upward function of $\boldsymbol{\lambda}$. Let us define $\bar{\Delta}^m$ by

$$\bar{\Delta}^m = \{\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_m) | \lambda_i \geq 0, i = 1, \cdots, m, \sum_{i=1}^m \lambda_i = 1\},$$

then the convex optimization problem

$$\max_{\boldsymbol{\lambda} \in \bar{\Delta}^m} f(\boldsymbol{\lambda}, \Phi) \qquad (15)$$

is equivalent to the problem of smallest enclosing circle [8]. For $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ that achieves (15), $Q^* \equiv \boldsymbol{\lambda}^*\Phi$ 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 [8].

For (15), the following Kuhn-Tucker condition holds [5], and [8].

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 (15) is that there exists d_0 with

$$d(P^i, \boldsymbol{\lambda}^*\Phi) \begin{cases} = d_0, & \text{for } i \text{ with } \lambda_i^* > 0, \\ \leq d_0, & \text{for } i \text{ with } \lambda_i^* = 0. \end{cases} \qquad (16)$$

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.

D. 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. Then, defining

$$\Psi = \begin{pmatrix} P^2 - P^1 \\ \vdots \\ P^m - P^1 \end{pmatrix} \in \mathbb{R}^{(m-1) \times n}, \qquad (17)$$

we have from (9)

$$\text{rank } \Psi = m - 1. \qquad (18)$$

We will calculate the barycentric coordinate $\boldsymbol{\lambda}^0$ of the equidistant point $Q^0 \in L_0$.

1) Calculation of $\boldsymbol{\lambda}^0$: The coordinate of P^i is defined by (12) and the $m \times n$ matrix Φ is defined by (13). Define the coordinate of Q^0 by

$$Q^0 = (Q_1^0, \cdots, Q_n^0) \in \mathbb{R}^n, \qquad (19)$$

and further define the following:

$$\hat{P}^i = (1, P^i)$$
$$= (1, P_1^i, \cdots, P_n^i) \in \mathbb{R}^{n+1}, \quad i = 1, \cdots, m, \qquad (20)$$
$$\hat{Q}^0 = (1, Q^0) = (1, Q_1^0, \cdots, Q_n^0) \in \mathbb{R}^{n+1}, \qquad (21)$$

$$\hat{\Phi} = \begin{pmatrix} \hat{P}^1 \\ \vdots \\ \hat{P}^m \end{pmatrix}$$
$$= \begin{pmatrix} 1 & P_1^1 & \cdots & P_n^1 \\ \vdots & \vdots & & \vdots \\ 1 & P_1^m & \cdots & P_n^m \end{pmatrix} \in \mathbb{R}^{m \times (n+1)}, \qquad (22)$$

$$J = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{(m-1) \times m}, \qquad (23)$$

$$\hat{\boldsymbol{a}} = (\|\hat{P}^1\|^2, \cdots, \|\hat{P}^m\|^2) \in \mathbb{R}^m, \qquad (24)$$
$$\mathbf{1} = (1, \cdots, 1) \in \mathbb{R}^m. \qquad (25)$$

Because $P^1, \cdots, P^m \in \mathbb{R}^n$ are in general position, we see rank $\hat{\Phi} = m$. In fact, if $\sum_{i=1}^m c_i \hat{P}^i = O$, then $\sum_{i=2}^m c_i (P^i - P^1) = O$, hence from (18) we have $c_i = 0$, $i = 1, \cdots, m$.

The barycentric coordinate $\boldsymbol{\lambda}^0$ of Q^0 about $P^1, \cdots, P^m$ satisfies

$$\begin{cases} Q^0 = \boldsymbol{\lambda}^0 \Phi, & (26) \\ \boldsymbol{\lambda}^{0\,t}\mathbf{1} = 1. & (27) \end{cases}$$

Together (26) and (27) is written as

$$\hat{Q}^0 = \boldsymbol{\lambda}^0 \hat{\Phi}. \qquad (28)$$

The equidistant point Q^0 from $P^1, \cdots, P^m$ satisfies

$$d(P^i, Q^0) = d(P^1, Q^0), \quad i = 2, \cdots, m, \qquad (29)$$

hence we have from (29)

$$2(P^i - P^1)^t Q^0 = \|P^i\|^2 - \|P^1\|^2, \ i = 2, \cdots, m, \quad (30)$$

and from (20), (21), (30),

$$2(\hat{P}^i - \hat{P}^1)^t \hat{Q}^0 = \|\hat{P}^i\|^2 - \|\hat{P}^1\|^2, \ i = 2, \cdots, m. \quad (31)$$

Since (31) is written as

$$2J\hat{\Phi}^t \hat{Q}^0 = J^t \hat{a}, \quad (32)$$

we have from (28)

$$2J\hat{\Phi}^t \hat{\Phi}^t \lambda^0 = J^t \hat{a}. \quad (33)$$

Defining $M \equiv \hat{\Phi}^t \hat{\Phi} \in \mathbb{R}^{m \times m}$, we have from (33)

$$J(2M^t \lambda^0 - {}^t \hat{a}) = \mathbf{0}. \quad (34)$$

Because $\text{rank } M = \text{rank } \hat{\Phi} = m$, M is non-singular. Since $\text{Ker } J = \{\tau {}^t \mathbf{1} | \tau \in \mathbb{R}\}$, from (34) there exists $\tau \in \mathbb{R}$ with

$$\lambda^0 = \frac{1}{2}(\hat{a} + \tau \mathbf{1})M^{-1}. \quad (35)$$

So, from (27)

$$1 = \lambda^0 {}^t \mathbf{1} = \frac{1}{2}(\hat{a} + \tau \mathbf{1})M^{-1 \, t} \mathbf{1}, \quad (36)$$

and from (36), we have

$$\tau = \frac{2 - \hat{a}M^{-1 \, t} \mathbf{1}}{\mathbf{1}M^{-1 \, t} \mathbf{1}}. \quad (37)$$

Substituting (37) into (35), we finally have

$$\lambda^0 = \frac{1}{2}\left(\hat{a} + \frac{2 - \hat{a}M^{-1 \, t} \mathbf{1}}{\mathbf{1}M^{-1 \, t} \mathbf{1}}\mathbf{1}\right)M^{-1}. \quad (38)$$

This is the barycentric coordinate of the equidistant point Q^0.

E. Inner Product, Pythagorean Theorem and Projection in $\mathbb{R}^n$

We will describe 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.

1) Inner Product: For three points $Q^k = (Q_1^k, \cdots, Q_n^k) \in \mathbb{R}^n$, $k = 1, 2, 3$, let us define the inner product $(Q^1 - Q^2, Q^3 - Q^2)$ by

$$(Q^1 - Q^2, Q^3 - Q^2) = \sum_{j=1}^{n}(Q_j^1 - Q_j^2)(Q_j^3 - Q_j^2). \quad (39)$$

This is the inner product of two vectors $Q^1 - Q^2 (= \overrightarrow{Q^2 Q^1})$ and $Q^3 - Q^2 (= \overrightarrow{Q^2 Q^3})$.

We have the following lemmas.

Lemma 1: For points P^i ($i = 1, \cdots, m$), Q, R in $\mathbb{R}^n$, consider the inner products $\sigma_i = (P^i - Q, R - Q)$, $i = 1, \cdots, m$. If $\sum_{i=1}^{m} \lambda_i = 1$, then we have

$$\sum_{i=1}^{m} \lambda_i \sigma_i = \left(\sum_{i=1}^{m} \lambda_i P^i - Q, R - Q\right).$$

Proof: By a simple calculation. □

Lemma 2: For any P, $Q \in \mathbb{R}^n$, $P \neq Q$, we have

$$(P - Q, P - Q) > 0.$$

Proof: By a simple calculation. □

Lemma 3: For any P, Q, $R \in \mathbb{R}^n$, we have

$$2(P - Q, R - Q) = d^2(P, Q) + d^2(Q, R) - d^2(P, R).$$

Proof: By a simple calculation. □

2) Pythagorean Theorem: For three points P, Q, R in $\mathbb{R}^n$, the following Pythagorean theorem and its inequality versions hold.

Theorem 1 (Pythagorean): For P, Q, $R \in \mathbb{R}^n$, we have

$$(P - Q, R - Q) \gtreqqless 0 \iff d^2(P, Q) + d^2(Q, R)$$
$$\gtreqqless d^2(P, R). \quad (40)$$

Proof: By Lemma 3. □

3) 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.

Lemma 4: Let L be an affine subspace of $\mathbb{R}^n$. For any $Q' \in \mathbb{R}^n$, the projection $Q'' = \pi(Q'|L)$ exists and is unique. Moreover, $Q'' = \pi(Q'|L)$ is equivalent to that $(P - Q'', Q' - Q'') = 0$ holds for any $P \in L$.

Proof: see [9]. □

4) Projection of Equidistant Point: For $P^1, \cdots, P^m \in \mathbb{R}^n$ in general position, let $L_0 = L(P^1, \cdots, P^m)$ be the affine subspace spanned by $P^1, \cdots, P^m$ and $Q^0 \in L_0$ be the equidistant point from $P^1, \cdots, P^m$. We see from (38) that Q^0 uniquely exists. Further, 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. Let $Q^1 = \pi(Q^0|L_1)$ be the projection of Q^0 onto L_1. Further, let $Q^k = \pi(Q^{k-1}|L_k)$, $k = 1, \cdots, m - 2$. We see that Q^k is the equidistant point from $P^{k+1}, \cdots, P^m$.

We have the following lemmas.

Lemma 5: $Q^k = \pi(Q^0|L_k)$, $k = 0, 1, \cdots, m - 2$.

Proof: It is trivial for $k = 0, 1$ by definition. Then, we prove for $k = 2$. From Lemma 4 and Theorem 1, we have $d^2(Q, Q^0) = d^2(Q, Q^1) + d^2(Q^1, Q^0)$ for any $Q \in L_2 \subset L_1$. Therefore, with respect to $Q \in L_2$ minimizing $d(Q, Q^0)$ is equivalent to minimizing $d(Q, Q^1)$. Because the projections $\pi(Q^0|L_2)$ and $\pi(Q^1|L_2)$ are unique by Lemma 4, so we have $\pi(Q^0|L_2) = \pi(Q^1|L_2)$. For $k \geq 3$, we can prove it by mathematical induction. □

Lemma 6: $(Q^i - Q^k, Q^0 - Q^k) = (Q^i - Q^k, Q^i - Q^k)$, $i = 0, 1, \cdots, k, k = 0, 1, \cdots, m - 2$.

Proof: By calculation, we have $(Q^i - Q^k, Q^0 - Q^k) = -(Q^k - Q^i, Q^0 - Q^i) + (Q^i - Q^k, Q^i - Q^k)$. Since $Q^k \in L_k \subset L_i$, we have $(Q^k - Q^i, Q^0 - Q^i) = 0$ by Lemma 4. □

F. Search for Q^* by Projection Algorithm

As we discussed in section II-A, it is valid to consider the equidistant point Q^0 and the projection of Q^0 onto some affine subspace. We are calling this a 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.

1) 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 equidistant point from $P^1, \cdots, P^4$ is $Q^0 = (1.93, 4.59, 2.85)$, and its barycentric coordinate $\boldsymbol{\lambda}^0$ is

$$\boldsymbol{\lambda}^0 = (-0.84, 0.04, 1.11, 0.69), \qquad (41)$$

which is calculated by (38). 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., $Q^1 = \pi(Q^0|L_1)$. We have $Q^1 = (0.48, 1.45, -0.16)$ and its barycentric coordinate $\boldsymbol{\lambda}^1$ is

$$\boldsymbol{\lambda}^1 = (0, -0.31, 0.63, 0.68). \qquad (42)$$

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., $Q^2 = \pi(Q^1|L_2)$. We have $Q^2 = (-2, 1, 1.5)$, and its barycentric coordinate $\boldsymbol{\lambda}^2$ is

$$\boldsymbol{\lambda}^2 = (0, 0, 0.5, 0.5). \qquad (43)$$

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 (16), 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 (41), (42), and all the components are non-negative at the last projection like (43). We will calculate below the smallest enclosing circle by the projection algorithm in situation 1.

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. That is,

$$\boldsymbol{\lambda}^0 = (\lambda_1^0, \lambda_2^0, \cdots, \lambda_m^0) = (-, +, \cdots, +),$$
$$\boldsymbol{\lambda}^1 = (\lambda_1^1, \lambda_2^1, \lambda_3^1, \cdots, \lambda_m^1) = (0, -, +, \cdots, +),$$
$$\boldsymbol{\lambda}^2 = (\lambda_1^2, \lambda_2^2, \lambda_3^2, \lambda_4^2, \cdots, \lambda_m^2) = (0, 0, -, +, \cdots, +),$$
$$\vdots$$
$$\boldsymbol{\lambda}^{K-1} = (\lambda_1^{K-1}, \cdots, \lambda_{K-1}^{K-1}, \lambda_K^{K-1}, \cdots, \lambda_m^{K-1})$$
$$= (0, \cdots, 0, -, +, \cdots, +),$$
$$\boldsymbol{\lambda}^K = (\lambda_1^K, \cdots, \lambda_K^K, \lambda_{K+1}^K, \cdots, \lambda_m^K)$$
$$= (0, \cdots, 0, +, \cdots, +).$$

If $K = 0$, we assume $\boldsymbol{\lambda}^0 = (+, +, \cdots, +)$. Summarizing above, we have, for $k = 0, 1, \cdots, K - 1$,

$$\lambda_i^k \begin{cases} = 0, i = 1, \cdots, k, & (44a) \\ < 0, i = k + 1, & (44b) \\ > 0, i = k + 2, \cdots, m, & (44c) \end{cases}$$

and for $k = K$,

$$\lambda_i^K \begin{cases} = 0, i = 1, \cdots, K, & (45a) \\ > 0, i = K + 1, \cdots, m. & (45b) \end{cases}$$

(The end of Assumption of situation 1)

Now, let us consider the inner products $\sigma_i \equiv (P^i - Q^K, Q^0 - Q^K)$, $i = 1, \cdots, m$. We have the following lemma.

Lemma 7: $\sigma_i \begin{cases} < 0, i = 1, \cdots, K, \\ = 0, i = K + 1, \cdots, m. \end{cases}$

Proof: By Lemma 5, we have $Q^K = \pi(Q^0|L_K)$, $L_K = L(P^{K+1}, \cdots, P^m)$. Since $P^i \in L_K$ for $i = K + 1, \cdots, m$, so by Lemma 4 we have $P^i - Q^K \perp Q^0 - Q^K$, thus

$$\sigma_i = 0, i = K + 1, \cdots, m. \qquad (46)$$

Next, we will prove $\sigma_i < 0$, $i = 1, \cdots, K$ by mathematical induction in the order of $i = K, K - 1, \cdots, 1$.

(I) Prove $\sigma_K < 0$:

From (44a), (46), we have

$$\lambda_K^{K-1} \sigma_K = \sum_{i=1}^m \lambda_i^{K-1} \sigma_i$$
$$= \left(\sum_{i=1}^m \lambda_i^{K-1} P^i - Q^K, Q^0 - Q^K \right) \text{ (by Lemma 1)}$$
$$= (Q^{K-1} - Q^K, Q^0 - Q^K)$$
$$= (Q^{K-1} - Q^K, Q^{K-1} - Q^K) \text{ (by Lemma 6)}$$
$$> 0. \text{ (by Lemma 2)} \qquad (47)$$

Since $\lambda_K^{K-1} < 0$ from (44b), we have $\sigma_K < 0$.

(II) Assuming $\sigma_K < 0, \sigma_{K-1} < 0, \cdots, \sigma_{k+1} < 0$, prove $\sigma_k < 0$:

From (44a), (46), we have

$$\sum_{i=k}^K \lambda_i^{k-1} \sigma_i = \sum_{i=1}^m \lambda_i^{k-1} \sigma_i$$
$$= \left(\sum_{i=1}^m \lambda_i^{k-1} P^i - Q^K, Q^0 - Q^K \right)$$
$$= (Q^{k-1} - Q^K, Q^0 - Q^K)$$
$$= (Q^{k-1} - Q^K, Q^{k-1} - Q^K)$$
$$> 0, \qquad (48)$$

i.e., $\lambda_k^{k-1} \sigma_k + \sum_{i=k+1}^K \lambda_i^{k-1} \sigma_i > 0$. By the induction hypothesis, $\sigma_i < 0$, $i = k + 1, \cdots, K$, and by (44b), $\lambda_k^{k-1} < 0$, by (44c), $\lambda_i^{k-1} > 0$, $i = k + 1, \cdots, K$, therefore we obtain $\sigma_k < 0$.

(III) From the above (I),(II), we have $\sigma_i < 0$, $i = 1, \cdots, K$. $\square$

Lemma 8: $d(P^i, Q^K) < d(P^{K+1}, Q^K) = \cdots = d(P^m, Q^K)$, $i = 1, \cdots, K$.

Proof: By Lemma 7, $\sigma_i = 0$, $i = K + 1, \cdots, m$, thus by Theorem 1, we have $d(P^{K+1}, Q^K) = \cdots = d(P^m, Q^K)$.

Also by Lemma 7, $\sigma_i < 0, i = 1, \cdots, K$, thus for $i = 1, \cdots, K$, we have by Theorem 1,

$$d^2(P^i, Q^K) < d^2(P^i, Q^0) - d^2(Q^K, Q^0)$$
$$= d^2(P^{K+1}, Q^0) - d^2(Q^K, Q^0)$$
$$= d^2(P^{K+1}, Q^K).$$

$\square$

Theorem 2: The center of the smallest enclosing circle Γ is $Q^* = Q^K$, and its radius is $d^* = d(P^{K+1}, Q^K)$.

Proof: It follows from (45a), (45b), Lemma 8 and the Kuhn-Tucker condition (16). $\square$

2) Situation 2 [$m = 2, 3, 4$ and n Is Arbitrary.]: We will calculate by the projection algorithm the center Q^* and the radius d^* of the smallest enclosing circle $\Gamma = \Gamma(P^1, \cdots, P^m)$ 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 $m = 2, 3, 4$.

Case of $m = 2$

Consider the smallest enclosing circle $\Gamma = \Gamma(P^1, P^2)$ for two different points P^1, P^2 in $\mathbb{R}^n$. Let Q^0 be the midpoint of the line segment $P^1 P^2$, then we have

Theorem 3: The center of Γ is $Q^* = Q^0$ and the radius is $d^* = d(P^1, Q^0)$.

Case of $m = 3$

Consider the smallest enclosing circle $\Gamma = \Gamma(P^1, P^2, P^3)$ for three points P^1, P^2, P^3 in general position in $\mathbb{R}^n$. Let $L_0 = L(P^1, P^2, P^3)$ be the affine subspace spanned by P^1, P^2, P^3, and $Q^0 \in L_0$ be the equidistant point from P^1, P^2, P^3. The barycentric coordinate of Q^0 about P^1, P^2, P^3 is denoted by $\lambda^0 = (\lambda_1^0, \lambda_2^0, \lambda_3^0)$.

Based on the signs of the components of λ^0, we can make the following classification without loss of generality.

Case 3-1: $\lambda_1^0 \geq 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0$
Case 3-2: $\lambda_1^0 < 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0$

Since Q^0 is the equidistant point from P^1, P^2, P^3, there must be at least two positive components in the barycentric coordinate of Q^0. Therefore, all the cases are exhausted by Case 3-1 and Case 3-2. For each case, we will determine the center Q^* and the radius d^* of the smallest enclosing circle $\Gamma = \Gamma(P^1, P^2, P^3)$.

Case 3-1

In this case, $\triangle P^1 P^2 P^3$ is an acute triangle. We have the following theorem by the Kuhn-Tucker condition (16).

Theorem 4: The center of Γ is $Q^* = Q^0$ and the radius is $d^* = d(P^1, Q^0)$.

Case 3-2

In this case, $\triangle P^1 P^2 P^3$ is an obtuse triangle. Let $L_1 = L(P^2, P^3)$ be the line connecting the two points P^2, P^3. By the assumption $\lambda_1^0 < 0$, the equidistant point Q^0 is in the opposite side of P^1 with respect to L_1. Let $Q^1 = \pi(Q^0|L_1)$ be the projection of Q^0 onto L_1, then Q^1 is the midpoint of $P^2 P^3$. Denoting by $\lambda^1 = (\lambda_1^1, \lambda_2^1, \lambda_3^1)$ the barycentric coordinate of Q^1 about P^1, P^2, P^3, we have $\lambda_1^1 = 0, \lambda_2^1 = \lambda_3^1 = 1/2$, hence this case is the situation 1. So, we have

Theorem 5: The center of Γ is $Q^* = Q^1$ and the radius is $d^* = d(P^2, Q^1)$.

Case of $m = 4$

Consider the smallest enclosing circle $\Gamma = \Gamma(P^1, P^2, P^3, P^4)$ for four points P^1, P^2, P^3, P^4 in general position in $\mathbb{R}^n$. Let $L_0 = L(P^1, \cdots, P^4)$ be the affine subspace spanned by $P^1, \cdots, P^4$ and $Q^0 \in L_0$ be the equidistant point from $P^1, \cdots, P^4$. The barycentric coordinate of Q^0 about $P^1, \cdots, P^4$ is denoted by $\lambda^0 = (\lambda_1^0, \cdots, \lambda_4^0)$. Without loss of generality, we have the following exhaustive classification:

Case 4-1: $\lambda_1^0 \geq 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$
Case 4-2: $\lambda_1^0 < 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$
Case 4-3: $\lambda_1^0 < 0, \lambda_2^0 < 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$

Case 4-1

By the Kuhn-Tucker condition (16), we have

Theorem 6: The center of Γ is $Q^* = Q^0$ and the radius is $d^* = d(P^1, Q^0)$.

Case 4-2

Because $\lambda_1^0 < 0$, we consider the projection $Q^1 = \pi(Q^0|L_1)$ of Q^0 onto $L_1 = L(P^2, P^3, P^4)$. Let $\lambda^1 = (\lambda_1^1, \cdots, \lambda_4^1)$ be the barycentric coordinate of Q^1 about $P^1, \cdots, P^4$. Since $P^1 \notin L_1$, we have $\lambda_1^1 = 0$. Based on the signs of the components of λ^1, we have the following classification without loss of generality.

Case 4-2-1: $\lambda_1^1 = 0, \lambda_2^1 \geq 0, \lambda_3^1 \geq 0, \lambda_4^1 \geq 0$
Case 4-2-2: $\lambda_1^1 = 0, \lambda_2^1 < 0, \lambda_3^1 \geq 0, \lambda_4^1 \geq 0$

Case 4-2-1

This case is the situation 1, so we have

Theorem 7: The center of Γ is $Q^* = Q^1$ and the radius is $d^* = d(P^2, Q^1)$.

Case 4-2-2

Let $Q^2 = \pi(Q^1|L_2)$ with $L_2 = L(P^3, P^4)$, then Q^2 is the midpoint of $P^3 P^4$. Denoting by $\lambda^2 = (\lambda_1^2, \cdots, \lambda_4^2)$ the barycentric coordinate of Q^2 about $P^1, \cdots, P^4$, we have $\lambda_1^2 = \lambda_2^2 = 0, \lambda_3^2 = \lambda_4^2 = 1/2$, hence also this case is the situation 1. So, we have

Theorem 8: The center of Γ is $Q^* = Q^2$ and the radius is $d^* = d(P^3, Q^2)$.

This completes Case 4-2, then let us consider Case 4-3.

Case 4-3

For the equidistant point $Q^0 \in L_0$, let us define $Q^{1(1)} = \pi(Q^0|L(P^2, P^3, P^4))$ and $Q^{1(2)} = \pi(Q^0|L(P^1, P^3, P^4))$. Then, denote by $\lambda^{1(1)} = (\lambda_1^{1(1)}, \cdots, \lambda_4^{1(1)})$ and $\lambda^{1(2)} = (\lambda_1^{1(2)}, \cdots, \lambda_4^{1(2)})$ the barycentric coordinates of $Q^{1(1)}$ and $Q^{1(2)}$ about $P^1, \cdots, P^4$, respectively. Because $P^1 \notin L(P^2, P^3, P^4)$, $P^2 \notin L(P^1, P^3, P^4)$, we have

$$\lambda_1^{1(1)} = 0, \quad \lambda_2^{1(2)} = 0. \tag{49}$$

Then, consider the inner products

$$\sigma_i^{1(1)} = (P^i - Q^{1(1)}, Q^0 - Q^{1(1)}), \ i = 1, \cdots, 4, \tag{50}$$
$$\sigma_i^{1(2)} = (P^i - Q^{1(2)}, Q^0 - Q^{1(2)}), \ i = 1, \cdots, 4. \tag{51}$$

Lemma 9: $\sigma_i^{1(1)} \begin{cases} < 0, \ i = 1, \\ = 0, \ i = 2, 3, 4, \end{cases}$

$\sigma_i^{1(2)} \begin{cases} < 0, \ i = 2, \\ = 0, \ i = 1, 3, 4. \end{cases}$

Proof: From $P^i - Q^{1(1)} \perp Q^0 - Q^{1(1)}$, $i = 2, 3, 4$, we have $\sigma_i^{1(1)} = 0$, $i = 2, 3, 4$. Thus,

$$\lambda_1^0 \sigma_1^{1(1)} = \sum_{i=1}^{4} \lambda_i^0 \sigma_i^{1(1)}$$
$$= (Q^0 - Q^{1(1)}, Q^0 - Q^{1(1)}) > 0 \qquad (52)$$

holds, and by the assumption $\lambda_1^0 < 0$ we obtain $\sigma_1^{1(1)} < 0$. We can prove for $\sigma_i^{1(2)}$ similarly. $\qquad \square$

Lemma 10: $\lambda_1^{1(2)} \sigma_1^{1(1)} + \lambda_2^{1(1)} \sigma_2^{1(2)} > 0$.

Proof: From Lemma 9,

$$\lambda_1^{1(2)} \sigma_1^{1(1)} = \sum_{i=1}^{4} \lambda_i^{1(2)} \sigma_i^{1(1)}$$
$$= (Q^{1(2)} - Q^{1(1)}, Q^0 - Q^{1(1)}), \qquad (53)$$
$$\lambda_2^{1(1)} \sigma_2^{1(2)} = \sum_{i=1}^{4} \lambda_i^{1(1)} \sigma_i^{1(2)}$$
$$= (Q^{1(1)} - Q^{1(2)}, Q^0 - Q^{1(2)}), \qquad (54)$$

so, by (53)+(54) we obtain $\lambda_1^{1(2)} \sigma_1^{1(1)} + \lambda_2^{1(1)} \sigma_2^{1(2)} = (Q^{1(1)} - Q^{1(2)}, Q^{1(1)} - Q^{1(2)}) > 0$. $\qquad \square$

Now from Lemmas 9, 10, we see that $\lambda_1^{1(2)} < 0$ or $\lambda_2^{1(1)} < 0$ holds. So far, the points P^1 and P^2 have been treated exactly equally, hence without loss of generality, we assume

$$\lambda_1^{1(2)} < 0, \qquad (55)$$

and proceed to the next step. Here, we reconfirm the signs of the components of $\boldsymbol{\lambda}^{1(2)}$ as follows:

$$\lambda_1^{1(2)} < 0, \lambda_2^{1(2)} = 0, \lambda_3^{1(2)} \geq 0, \lambda_4^{1(2)} \geq 0. \qquad (56)$$

We make a classification based on the sings of components of $\boldsymbol{\lambda}^{1(1)}$. So far, the points P^3 and P^4 have been treated exactly equally, so without loss of generality, we have the following classification:

Case 4-3-1: $\lambda_1^{1(1)} = 0, \lambda_2^{1(1)} \geq 0, \lambda_3^{1(1)} \geq 0, \lambda_4^{1(1)} \geq 0$
Case 4-3-2: $\lambda_1^{1(1)} = 0, \lambda_2^{1(1)} < 0, \lambda_3^{1(1)} \geq 0, \lambda_4^{1(1)} \geq 0$
Case 4-3-3: $\lambda_1^{1(1)} = 0, \lambda_2^{1(1)} \geq 0, \lambda_3^{1(1)} < 0, \lambda_4^{1(1)} \geq 0$

Case 4-3-1

From Lemma 9, we obtain $d(P^1, Q^{1(1)}) < d(P^2, Q^{1(1)}) = \cdots = d(P^4, Q^{1(1)})$ in a similar way to the proof of Lemma 8. Therefore, by the Kuhn-Tucker condition (16), we have

Theorem 9: The center of Γ is $Q^* = Q^{1(1)}$ and the radius is $d^* = d(P^2, Q^{1(1)})$.

Case 4-3-2

Define $Q^2 \equiv \pi(Q^{1(1)}|L(P^3, P^4)) = \pi(Q^0|L(P^3, P^4))$ and consider the inner products $\sigma_i^2 = (P^i - Q^2, Q^0 - Q^2)$, $i = 1, \cdots, 4$. We have

Lemma 11: $\sigma_i^2 \begin{cases} < 0, i = 1, 2, \\ = 0, i = 3, 4. \end{cases}$

Proof: From $P^i - Q^2 \perp Q^0 - Q^2$, $i = 3, 4$, we have $\sigma_i^2 = 0$, $i = 3, 4$. Next, we will prove $\sigma_1^2 < 0$, $\sigma_2^2 < 0$. Note that

$Q^2 = \pi(Q^{1(2)}|L(P^3, P^4)) = \pi(Q^0|L(P^3, P^4))$. From (49),

$$\lambda_1^{1(2)} \sigma_1^2 = \sum_{i=1}^{4} \lambda_i^{1(2)} \sigma_i^2$$
$$= (Q^{1(2)} - Q^2, Q^0 - Q^2)$$
$$= (Q^{1(2)} - Q^2, Q^{1(2)} - Q^2)$$
$$> 0,$$

and by the assumption $\lambda_1^{1(2)} < 0$, we obtain $\sigma_1^2 < 0$. Similarly, by considering $\sum_{i=1}^{4} \lambda_i^{1(1)} \sigma_i^2$, we obtain $\sigma_2^2 < 0$. $\qquad \square$

Lemma 12: $d(P^i, Q^2) < d(P^3, Q^2) = d(P^4, Q^2)$, $i = 1, 2$.

Proof: Similar to the proof of Lemma 8. $\qquad \square$

Theorem 10: The center of Γ is $Q^* = Q^2$ and the radius is $d^* = d(P^3, Q^2)$.

Proof: By Lemma 12 and the Kuhn-Tucker condition (16). $\qquad \square$

Case 4-3-3

Put $Q^{1(3)} = \pi(Q^0|L(P^1, P^2, P^4))$, and consider the inner products $\sigma_i^{1(3)} = (P^i - Q^{1(3)}, Q^0 - Q^{1(3)})$, $i = 1, \cdots, 4$. We have

Lemma 13: $\sigma_i^{1(3)} \begin{cases} > 0, i = 3, \\ = 0, i = 1, 2, 4. \end{cases}$

Proof: From $P^i - Q^{1(3)} \perp Q^0 - Q^{1(3)}$, $i = 1, 2, 4$, we have $\sigma_i^{1(3)} = 0$, $i = 1, 2, 4$. Thus, $\lambda_3^0 \sigma_3^{1(3)} = \sum_{i=1}^{4} \lambda_i^0 \sigma_i^{1(3)} = (Q^0 - Q^{1(3)}, Q^0 - Q^{1(3)}) > 0$, so by the assumption $\lambda_3^0 > 0$, we obtain $\sigma_3^{1(3)} > 0$. $\qquad \square$

Lemma 14: $\lambda_1^{1(3)} < 0$.

Proof: By Lemma 13, we have

$$\lambda_3^{1(1)} \sigma_3^{1(3)} = \sum_{i=1}^{4} \lambda_i^{1(1)} \sigma_i^{1(3)}$$
$$= (Q^{1(1)} - Q^{1(3)}, Q^0 - Q^{1(3)}). \qquad (57)$$

Further, by Lemma 9,

$$\lambda_1^{1(3)} \sigma_1^{1(1)} = \sum_{i=1}^{4} \lambda_i^{1(3)} \sigma_i^{1(1)}$$
$$= (Q^{1(3)} - Q^{1(1)}, Q^0 - Q^{1(1)}). \qquad (58)$$

By (57)+(58), we have

$$\lambda_3^{1(1)} \sigma_3^{1(3)} + \lambda_1^{1(3)} \sigma_1^{1(1)} = (Q^{1(1)} - Q^{1(3)}, Q^{1(1)} - Q^{1(3)})$$
$$> 0. \qquad (59)$$

By the assumption of Case 4-3-3, $\lambda_3^{1(1)} < 0$ holds, by Lemma 13, $\sigma_3^{1(3)} > 0$ holds, and by Lemma 9, $\sigma_1^{1(1)} < 0$ holds, therefore, by (59), we obtain $\lambda_1^{1(3)} < 0$. $\qquad \square$

Here, we reconfirm the signs of the components of $\boldsymbol{\lambda}^{1(1)}$ and $\boldsymbol{\lambda}^{1(3)}$ as follows:

$$\lambda_1^{1(1)} = 0, \lambda_2^{1(1)} \geq 0, \lambda_3^{1(1)} < 0, \lambda_4^{1(1)} \geq 0, \qquad (60)$$
$$\lambda_1^{1(3)} < 0, \lambda_2^{1(3)} \geq 0, \lambda_3^{1(3)} = 0, \lambda_4^{1(3)} \geq 0. \qquad (61)$$

If we exchange the points P^2 and P^3, then the second coordinate and the third coordinate are exchanged in (60) and (61), hence as a result, the signs of the components become the same combination as in Case 4-3-2 and (56). In the proof of

Case 4-3-2, the barycentric coordinate $\boldsymbol{\lambda}^0$ of Q^0 is not used, only $\boldsymbol{\lambda}^{1(1)}$ and $\boldsymbol{\lambda}^{1(3)}$ are used in the proof, therefore, exchanging P^2 and P^3 in the proof of Case 4-3-2 gives the proof of Case 4-3-3. Hence, in this Case 4-3-3, we have the following theorem.

Theorem 11: Put $Q^{2\dagger} = \pi(Q^0|L(P^2, P^4))$. Then the center of Γ is $Q^* = Q^{2\dagger}$ and the radius is $d^* = d(P^2, Q^{2\dagger})$.

III. PROBLEM OF CHANNEL CAPACITY

In chapter II, we obtained some theorems on the problem of smallest enclosing circle using Euclidean geometry. In particular, by using the distance, inner product, Pythagorean theorem and projection in the Euclidean space, we obtained a method of searching for the center Q^* of the smallest enclosing circle by the projection algorithm. The problem of channel capacity has a similar geometric structure to that of smallest enclosing circle based on the similarity of (6) and (8).

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], rather than the Euclidean geometry. A difference between the Euclidean geometry and the information geometry is that the Euclidean geometry uses one coordinate system but the information geometry uses two mutually dual coordinate systems. 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 structure is introduced on Δ^n as follows [1].

1) Dual Coordinate Systems: Two coordinate systems, in other words, two ways to specify $Q = (Q_1, \cdots, Q_n) \in \Delta^n$

are given on Δ^n.

η coordinate:

$$\boldsymbol{\eta} = (\eta_2, \cdots, \eta_n), \; \eta_j = Q_j, \; j = 2, \cdots, n,$$

θ coordinate:

$$\boldsymbol{\theta} = (\theta_2, \cdots, \theta_n), \; \theta_j = \log \frac{Q_j}{Q_1}, \; j = 2, \cdots, n.$$

The η coordinate system and the θ coordinate system are mutually dual coordinate systems [1].

2) Geodesic: A straight line with respect to the η coordinate is called an η geodesic. Let $\boldsymbol{\eta}^1, \boldsymbol{\eta}^2$ be the η coordinates of $Q^1, Q^2 \in \Delta^n$, respectively, then the η geodesic passing through Q^1, Q^2 is defined by

$$\boldsymbol{\eta}(t) = (1 - t)\boldsymbol{\eta}^1 + t\boldsymbol{\eta}^2 \in \Delta^n, \; t \in \mathbb{R}. \tag{62}$$

Further, a straight line with respect to the θ coordinate is called a θ geodesic. Let $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2$ be the θ coordinates of $Q^1, Q^2 \in \Delta^n$, respectively, then the θ geodesic passing through Q^1, Q^2 is defined by

$$\boldsymbol{\theta}(t) = (1 - t)\boldsymbol{\theta}^1 + t\boldsymbol{\theta}^2 \in \Delta^n, \; t \in \mathbb{R}. \tag{63}$$

C. Inner Product, Pythagorean Theorem and Projection in Δ^n

1) Inner Product: Consider three points $Q^1, Q^2, Q^3 \in \Delta^n$. The inner product $(Q^1 - Q^2, Q^3 \ominus Q^2)$ is defined as follows. Let $\boldsymbol{\eta}^1 = (\eta_2^1, \cdots, \eta_n^1), \boldsymbol{\eta}^2 = (\eta_2^2, \cdots, \eta_n^2)$ be the η coordinates of Q^1, Q^2, respectively, and $\boldsymbol{\theta}^2 = (\theta_2^2, \cdots, \theta_n^2), \boldsymbol{\theta}^3 = (\theta_2^3, \cdots, \theta_n^3)$ be the θ coordinates of Q^2, Q^3, respectively. Then the *inner product* is defined by

$$(Q^1 - Q^2, Q^3 \ominus Q^2) \equiv \sum_{j=2}^n (\eta_j^1 - \eta_j^2)(\theta_j^3 - \theta_j^2). \tag{64}$$

This is the inner product (in the usual sense) of the two tangent vectors $d\boldsymbol{\eta}(t)/dt|_{t=0} = \boldsymbol{\eta}^1 - \boldsymbol{\eta}^2$ and $d\boldsymbol{\theta}(t)/dt|_{t=0} = \boldsymbol{\theta}^3 - \boldsymbol{\theta}^2$ at Q^2 for two geodesics $\boldsymbol{\eta}(t) = (1 - t)\boldsymbol{\eta}^2 + t\boldsymbol{\eta}^1$ and $\boldsymbol{\theta}(t) = (1 - t)\boldsymbol{\theta}^2 + t\boldsymbol{\theta}^3$ passing through Q^2 [1].

For $Q^1, Q^2, Q^3 \in \Delta^n$, the η geodesic $\boldsymbol{\eta}(t) = (1-t)\boldsymbol{\eta}^2 + t\boldsymbol{\eta}^1$ and the θ geodesic $\boldsymbol{\theta}(t) = (1 - t)\boldsymbol{\theta}^2 + t\boldsymbol{\theta}^3$ are said to be *orthogonal* at Q^2 if [1]

$$(Q^1 - Q^2, Q^3 \ominus Q^2) = 0. \tag{65}$$

We have the following lemmas.

Lemma 15: For $P^i \; (i = 1, \cdots, m), Q, R \in \Delta^n$, consider the inner products $\sigma_i = (P^i - Q, R \ominus Q), i = 1, \cdots, m$. If $\sum_{i=1}^m \lambda_i = 1$, then

$$\sum_{i=1}^m \lambda_i \sigma_i = \left(\sum_{i=1}^m \lambda_i P^i - Q, R \ominus Q\right). \tag{66}$$

Proof: By the definition of inner product and a simple calculation. □

Lemma 16: For any $P, Q, R \in \Delta^n$, we have

$$(P - Q, R \ominus Q) = -(Q - P, R \ominus P) + (P - Q, P \ominus Q).$$
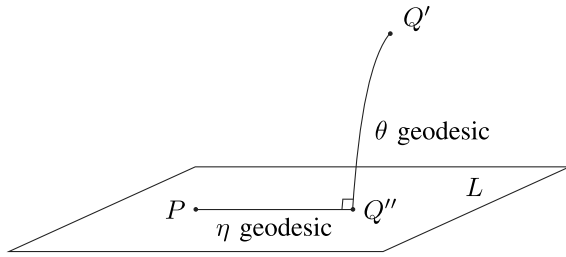
Proof: By a simple calculation. □

Fig. 3.   Projection $Q'' = \pi(Q'|L)$ of Q' onto affine subspace L.

Lemma 17: For any $P, Q, R \in \Delta^n$, we have

$$(P - Q, R \ominus Q) = D(P\|Q) + D(Q\|R) - D(P\|R).$$

Proof: By a simple calculation (see also [1]).   □

2) Pythagorean Theorem: For three points P, Q, R in Δ^n, the following Pythagorean theorem and its inequality version hold.

Theorem 12: (Pythagorean) For $P, Q, R \in \Delta^n$, we have

$$(P - Q, R \ominus Q) \gtreqqless 0 \iff D(P\|Q) + D(Q\|R)$$
$$\gtreqqless D(P\|R). \tag{67}$$

Proof: By Lemma 17 (see also [1]).   □

3) Projection by Kullback-Leibler Divergence: For $P^1, \cdots, P^m \in \Delta^n$, the affine subspace $L(P^1, \cdots, P^m) \subset \Delta^n$ spanned by $P^1, \cdots, P^m$ is defined by

$$L(P^1, \cdots, P^m) = \left\{ \sum_{i=1}^m \lambda_i P^i \ \middle|\ \sum_{i=1}^m \lambda_i = 1 \right\} \cap \Delta^n. \tag{68}$$

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.

Lemma 18: Let L be an affine subspace in Δ^n. Then, for any $Q' \in \Delta^n$, the projection $Q'' = \pi(Q'|L)$ exists and is unique. Moreover, $Q'' = \pi(Q'|L)$ is equivalent to that $(P - Q'', Q' \ominus Q'') = 0$ holds for any $P \in L$ (see Fig.3).

Proof: see [1].   □

D. 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. In chapter II, we investigated the problem of smallest enclosing circle based on the Euclidean geometry in case that the points $P^1, \cdots, P^m$ are in general position. In this chapter, we will consider the problem of channel capacity based on the information geometry.

1) Equidistant Point From $P^1, \cdots, P^m$: For a channel matrix Φ in (2), a matrix $\Psi \in \mathbb{R}^{(m-1) \times n}$ is defined by

$$\Psi = \begin{pmatrix} P^2 - P^1 \\ \vdots \\ P^m - P^1 \end{pmatrix}$$
$$= \begin{pmatrix} P_1^2 - P_1^1 & \cdots & P_n^2 - P_n^1 \\ \vdots & & \vdots \\ P_1^m - P_1^1 & \cdots & P_n^m - P_n^1 \end{pmatrix}. \tag{69}$$

Then, $P^1, \cdots, P^m$ are said to be *in general position*, if the vectors $P^2 - P^1, \cdots, P^m - P^1$ are linearly independent, or

$$\text{rank } \Psi = m - 1. \tag{70}$$

Now, we assume in this chapter that $P^1, \cdots, P^m$ are in general position. Then, denote by $\Psi' \in \mathbb{R}^{(m-1) \times (n-1)}$ the matrix that is made by removing the first column of Ψ, i.e.,

$$\Psi' = \begin{pmatrix} P_2^2 - P_2^1 & \cdots & P_n^2 - P_n^1 \\ \vdots & & \vdots \\ P_2^m - P_2^1 & \cdots & P_n^m - P_n^1 \end{pmatrix}. \tag{71}$$

We see rank $\Psi' = m - 1$. In fact, let $P^{i\prime} = (P_2^i, \cdots, P_n^i)$, $i = 1, \cdots, m$, and suppose $\sum_{i=2}^m c_i (P^{i\prime} - P^{1\prime}) = O$. Noticing $\sum_{j=1}^n P_j^i = 1$, $i = 1, \cdots, m$, we have $\sum_{i=2}^m c_i (P^i - P^1) = O$, therefore, from (70), $c_i = 0$, $i = 2, \cdots, m$.

Now, similarly to the problem of smallest enclosing circle, we consider an output distribution that has the equal Kullback-Leibler divergence from $P^1, \cdots, P^m$, i.e., we consider $Q \in \Delta^n$ that satisfies

$$D(P^i\|Q) = D(P^1\|Q), \quad i = 2, \cdots, m. \tag{72}$$

Let $\boldsymbol{\theta} = (\theta_2, \cdots, \theta_n)$ be the θ coordinate of Q, and $H(P^i) = -\sum_{j=1}^n P_j^i \log P_j^i$ be the entropy of P^i, $i = 1, \cdots, m$. Then by a simple calculation, we have from (72),

$$D(P^i\|Q) - D(P^1\|Q)$$
$$= -\sum_{j=2}^n (P_j^i - P_j^1) \theta_j - H(P^i) + H(P^1)$$
$$= 0, \quad i = 2, \cdots, m. \tag{73}$$

Putting $\boldsymbol{b} = (-H(P^2) + H(P^1), \cdots, -H(P^m) + H(P^1)) \in \mathbb{R}^{m-1}$, we can rewrite (73) as

$$\Psi'^t \boldsymbol{\theta} = {}^t\boldsymbol{b}. \tag{74}$$

Because rank $\Psi' = m - 1$, the equation (74) has a solution $\boldsymbol{\theta}$, but it is not necessarily unique. From a solution $\boldsymbol{\theta} = (\theta_2, \cdots, \theta_n)$ of (74), we make $Q = (Q_1, \cdots, Q_n) \in \Delta^n$ by

$$\begin{cases} Q_1 = \left(1 + \sum_{j=2}^n \exp \theta_j \right)^{-1}, \\ Q_j = Q_1 \exp \theta_j, \quad j = 2, \cdots, n. \end{cases} \tag{75}$$

Then the θ coordinate of Q becomes $\boldsymbol{\theta}$.

Lemma 19: Let $L_0 = L(P^1, \cdots, P^m)$ be the affine subspace spanned by $P^1, \cdots, P^m$ (see (68)). Make $Q \in \Delta^n$

by (75) from a solution $\boldsymbol{\theta}$ of the equation (74) and put $Q^0 = \pi(Q|L_0)$. Then, Q^0 is the unique point in L_0 with

$$D(P^i \| Q^0) = D(P^1 \| Q^0), \; i = 2, \cdots, m. \quad (76)$$

Proof: For Q, $Q' \in \Delta^n$ with (74), (75), let $Q^0 = \pi(Q|L_0)$, $Q^{0'} = \pi(Q'|L_0)$. We will prove $Q^0 = Q^{0'}$. Put $Q^0 = (Q_1^0, \cdots, Q_n^0)$, $Q^{0'} = (Q_1^{0'}, \cdots, Q_n^{0'})$. Both Q and Q' satisfy the equation (72), so from Theorem 12 and Lemma 18, we have

$$\sum_{j=1}^{n} P_j^i \log \frac{P_j^i}{Q_j^0} = \sum_{j=1}^{n} P_j^1 \log \frac{P_j^1}{Q_j^0}, \; i = 2, \cdots, m, \quad (77)$$

$$\sum_{j=1}^{n} P_j^i \log \frac{P_j^i}{Q_j^{0'}} = \sum_{j=1}^{n} P_j^1 \log \frac{P_j^1}{Q_j^{0'}}, \; i = 2, \cdots, m. \quad (78)$$

Subtracting (77) from (78),

$$\sum_{j=1}^{n} P_j^i \log \frac{Q_j^0}{Q_j^{0'}} = \sum_{j=1}^{n} P_j^1 \log \frac{Q_j^0}{Q_j^{0'}}, \; i = 2, \cdots, m. \quad (79)$$

From (79), we can write

$$\sum_{j=1}^{n} P_j^i \log \frac{Q_j^0}{Q_j^{0'}} = \kappa \; (\text{constant}), \; i = 1, \cdots, m. \quad (80)$$

Let $\boldsymbol{\lambda}^0 = (\lambda_1^0, \cdots, \lambda_m^0)$, $\boldsymbol{\lambda}^{0'} = (\lambda_1^{0'}, \cdots, \lambda_m^{0'})$ be the barycentric coordinates of Q^0, $Q^{0'} \in L_0$ about $P^1, \cdots, P^m$, respectively. Then, from (80),

$$\kappa = \sum_{i=1}^{m} \lambda_i^0 \sum_{j=1}^{n} P_j^i \log \frac{Q_j^0}{Q_j^{0'}} = D(Q^0 \| Q^{0'}), \quad (81)$$

$$\kappa = \sum_{i=1}^{m} \lambda_i^{0'} \sum_{j=1}^{n} P_j^i \log \frac{Q_j^0}{Q_j^{0'}} = -D(Q^{0'} \| Q^0). \quad (82)$$

Subtracting (82) from (81),

$$0 = D(Q^0 \| Q^{0'}) + D(Q^{0'} \| Q^0), \quad (83)$$

thus, we obtain $Q^0 = Q^{0'}$. □

$Q^0 \in L_0 = L(P^1, \cdots, P^m)$ with (76) is called the *equidistant point* from $P^1, \cdots, P^m$. The existence and uniqueness of Q^0 is guaranteed by Lemma 19. It might be better to call Q^0 an equi-divergence distribution, but, by analogy with smallest enclosing circle, we call it an equidistant point, too.

Remark 1: In [6], the solution $X_1, \cdots, X_m$ of [6, eq. (11)] is equivalent to our equidistant point. The existence of $X_1, \cdots, X_m$ is assumed in [6], but we proved the existence and the uniqueness of the equidistant point by Lemma 19.

2) 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. For the equidistant point Q^0 from $P^1, \cdots, P^m$, let $Q^1 = \pi(Q^0|L_1)$ be the projection of Q^0 onto L_1. Further, we define $Q^k = \pi(Q^{k-1}|L_k)$, $k = 1, \cdots, m - 2$. We see that Q^k is the equidistant point from $P^{k+1}, \cdots, P^m$.

Lemma 20: $Q^k = \pi(Q^0|L_k)$, $k = 0, 1, \cdots, m - 2$.

Proof: It is trivial for $k = 0, 1$ by definition. Next for $k = 2$, putting $Q^{2'} = \pi(Q^0|L_2)$, we will prove $Q^{2'} = Q^2 \equiv \pi(Q^1|L_2)$. Since $Q^1 = \pi(Q^0|L_1)$, for any $Q \in L_1$ we have by Theorem 12 and Lemma 18

$$D(Q \| Q^0) = D(Q \| Q^1) + D(Q^1 \| Q^0). \quad (84)$$

Therefore, by (84), with respect to $Q \in L_2(\subset L_1)$ minimizing $D(Q \| Q^0)$ and minimizing $D(Q \| Q^1)$ are equivalent. Because the projections $\pi(Q^1|L_2)$ and $\pi(Q^0|L_2)$ are unique by Lemma 18, we obtain $Q^{2'} = Q^2$. For $k \geq 3$, it is proved by mathematical induction. □

Lemma 21: $(Q^i - Q^k, Q^0 \ominus Q^k) = (Q^i - Q^k, Q^i \ominus Q^k)$, $i = 0, 1, \cdots, k, k = 0, 1, \cdots, m - 2$.

Proof: By Lemma 16, we have

$$(Q^i - Q^k, Q^0 \ominus Q^k)$$
$$= -(Q^k - Q^i, Q^0 \ominus Q^i) + (Q^i - Q^k, Q^i \ominus Q^k). \quad (85)$$

By Lemma 20, $Q^i = \pi(Q^0|L_i)$, and by $k \geq i$, $Q^k \in L_i$, so we obtain $(Q^k - Q^i, Q^0 \ominus Q^i) = 0$ by Lemma 18. □

Lemma 22: For any $P, Q \in \Delta^n$, $P \neq Q$, we have $(P - Q, P \ominus Q) > 0$.

Proof: By Lemma 17, we obtain $(P - Q, P \ominus Q) = D(P \| Q) + D(Q \| P) > 0$. □

E. 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. We will show that Q^* and C are obtained by the projection algorithm, under the assumption (70) that the row vectors of the channel matrix Φ are in general position and in the situations 1 and 2. The obtained Lemmas and Theorems are almost the same form as in the case of smallest enclosing circle, thus for the sake of easy reading, we put them in the Appendix.

Remark 2: We investigated the same problem in the previous paper [7], and considered the projection algorithm by the Kullback-Leibler divergence. However, there was an error in the proof of [7, Th. 3].

In chapters II and III, we used common symbols both in smallest enclosing circle and channel capacity. Then, we 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, i.e., (18) or (70). Under these assumptions, there exists the equidistant point Q^0 from $P^1, \cdots, P^m$, hence by the projection algorithm, we could calculate the center of the smallest enclosing circle Q^*, or the capacity achieving output distribution Q^*. However, we cannot expect that arbitrarily given points $P^1, \cdots, P^m$ are in general position, especially if m is large. For example, four points in $\mathbb{R}^2$ are not in general position. If $P^1, \cdots, P^m$ are not in general position, the projection algorithms in the previous chapters cannot be used.

TABLE I
CORRESPONDENCE OF SYMBOLS BETWEEN SMALLEST ENCLOSING CIR-
CLE 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	$d(P^i, \boldsymbol{\lambda}^*\Phi)$ $\begin{cases} = d_0, \text{if } \lambda_i^* > 0 \\ \leq d_0, \text{if } \lambda_i^* = 0 \end{cases}$	$D(P^i\|\boldsymbol{\lambda}^*\Phi)$ $\begin{cases} = C_0, \text{if } \lambda_i^* > 0 \\ \leq C_0, \text{if } \lambda_i^* = 0 \end{cases}$
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	$\min_{P \in L} d(P, Q)$	$\min_{P \in L} D(P\|Q)$

In this chapter, we assume

$$\operatorname{rank} \Psi \leq m - 1 \ \text{ for } \Psi = \begin{pmatrix} P^2 - P^1 \\ \vdots \\ P^m - P^1 \end{pmatrix}, \qquad (86)$$

which implies that there is no constraint on the placement of the points $P^1, \cdots, P^m$. Because the projection algorithms in chapters II, III cannot be used in its present form if rank $\Psi < m - 1$, we must consider some way to avoid this difficulty. In this chapter, 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. According to the method which will be proposed in this chapter, it is not necessary to check the rank of Ψ in advance. Furthermore, we do not need to know about geometric conditions, such as, a point is contained in the convex hull of several other points, or what the dimension of the subspace spanned by the whole points is, and so on. However, because we make a little deformation to the original problem, there might be some possibility that the obtained result is different from the true solution of the original problem.

Let us consider an example of Fig.4. The three points P^1, P^2, P^3 are on a straight line in $\mathbb{R}^2$. In this case, how can we determine the smallest enclosing circle $\Gamma(P^1, P^2, P^3)$ without knowing the positional relationship such as P^1 lies
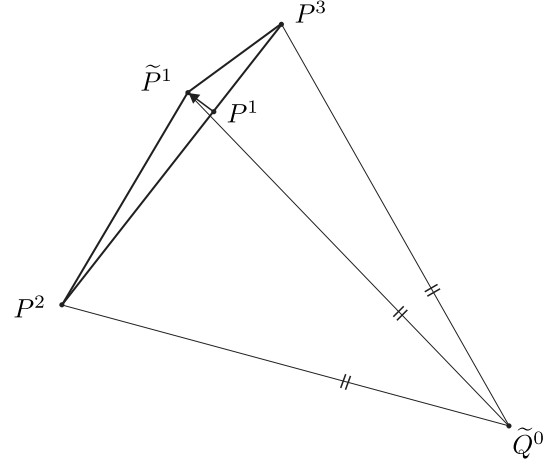

Fig. 4.   Getting equidistant point by shifting a point.

between P^2 and P^3? A method that we conceive immediately is to select two points from the three points and check all the combinations $\Gamma(P^1, P^2)$, $\Gamma(P^1, P^3)$, $\Gamma(P^2, P^3)$. Among them, the one which includes all the points with the minimum radius is the smallest enclosing circle. However, if the number of points becomes large the computational complexity becomes large, thus, such a combinatorial method cannot be applied to general cases.

Now, let us shift P^1 slightly to have $\widetilde{P}^1$ (see Fig. 4), then there exists the equidistant point $\widetilde{Q}^0$ from $\widetilde{P}^1, P^2, P^3$, so, we can calculate the smallest enclosing circle $\Gamma(\widetilde{P}^1, P^2, P^3)$ by the projection algorithm in the previous chapters. If the amount of shift is small, then we can expect $\Gamma(\widetilde{P}^1, P^2, P^3) = \Gamma(P^1, P^2, P^3)$.

Based on the above fundamental idea, we will give the following algorithm.

A. Method of Lifting Dimension of Point in $\mathbb{R}^n$

For m points $P^1, \cdots, P^m$ in $\mathbb{R}^n$ with (86), define the points $\widetilde{P}^1, \cdots, \widetilde{P}^m \in \mathbb{R}^{n+m}$ by lifting the dimension as follows:

$$P^1 = (P_1^1, \cdots, P_n^1) \to \widetilde{P}^1 = (P_1^1, \cdots, P_n^1, \varepsilon, 0, \cdots, 0),$$
$$P^2 = (P_1^2, \cdots, P_n^2) \to \widetilde{P}^2 = (P_1^2, \cdots, P_n^2, 0, \varepsilon, \cdots, 0),$$
$$\vdots \qquad\qquad\qquad\qquad \vdots$$
$$P^m = (P_1^m, \cdots, P_n^m) \to \widetilde{P}^m = (P_1^m, \cdots, P_n^m, 0, \cdots, 0, \varepsilon),$$
$$(87)$$

where $\varepsilon \in \mathbb{R}$, $\varepsilon \neq 0$, and $|\varepsilon|$ is sufficiently small. The above correspondence $P^i \to \widetilde{P}^i$ can be written as follows. Defining the ith fundamental vector $\boldsymbol{e}_i \in \mathbb{R}^m$ by
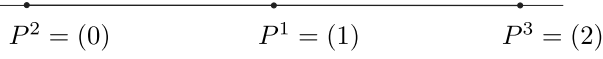
$$\boldsymbol{e}_i = (0, \cdots, 0, \overset{\overset{i\,\text{th}}{\vee}}{1}, 0, \cdots, 0), \quad i = 1, \cdots, m, \qquad (88)$$

we have the correspondence

$$\mathbb{R}^n \ni P^i \to \widetilde{P}^i = (P^i, \varepsilon \boldsymbol{e}_i) \in \mathbb{R}^{n+m}, \quad i = 1, \cdots, m. \quad (89)$$

Then let

$$\widetilde{\Phi} = \begin{pmatrix} \widetilde{P}^1 \\ \vdots \\ \widetilde{P}^m \end{pmatrix} \in \mathbb{R}^{m \times (n+m)}. \qquad (90)$$

$$P^2 = (0) \qquad P^1 = (1) \qquad P^3 = (2)$$

Fig. 5. Three points $P^1 = (1)$, $P^2 = (0)$, $P^3 = (2)$ in $\mathbb{R}$.

We have rank $\widetilde{\Phi} = m$ for $\varepsilon \neq 0$, hence, by the same argument as (II-D), we see that there exists the equidistant point $\widetilde{Q}^0 \in L(\widetilde{P}^1, \cdots, \widetilde{P}^m) \subset \mathbb{R}^{n+m}$ from $\widetilde{P}^1, \cdots, \widetilde{P}^m$. Denote by $\widetilde{\lambda}^0 = (\widetilde{\lambda}_1^0, \cdots, \widetilde{\lambda}_m^0)$ the barycentric coordinate of $\widetilde{Q}^0$ about $\widetilde{P}^1, \cdots, \widetilde{P}^m$. Define $\widetilde{M} \equiv \widetilde{\Phi}^t \widetilde{\Phi} \in \mathbb{R}^{m \times m}$, then since rank $\widetilde{M} = $ rank $\widetilde{\Phi} = m$, $\widetilde{M}$ is non-singular. Putting $\widetilde{a} = (\|\widetilde{P}^1\|^2, \cdots, \|\widetilde{P}^m\|^2) \in \mathbb{R}^m$, we have

$$\widetilde{\lambda}^0 = \frac{1}{2}\left(\widetilde{a} + \frac{2 - \widetilde{a}\widetilde{M}^{-1t}\mathbf{1}}{\mathbf{1}\widetilde{M}^{-1t}\mathbf{1}}\mathbf{1}\right)\widetilde{M}^{-1} \qquad (91)$$

in a similar way as (38).

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 $\widetilde{\lambda}^*$ the barycentric coordinate of $\widetilde{Q}^*$ about $\widetilde{P}^1, \cdots, \widetilde{P}^m$, and $\widetilde{\lambda}^*|_{\varepsilon=0}$ by substituting $\varepsilon = 0$ in $\widetilde{\lambda}^*$. Then $Q^* = \widetilde{\lambda}^*|_{\varepsilon=0}\Phi$ is expected to be the center of the original smallest enclosing circle.

Example 2: Let us consider three points $P^1 = (1)$, $P^2 = (0)$, $P^3 = (2)$ in $\mathbb{R}$ (see Fig.5).

Lifting the dimension, we have $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3 \in \mathbb{R}^4$ by

$$P^1 = (1) \rightarrow \widetilde{P}^1 = (1, \varepsilon, 0, 0),$$
$$P^2 = (0) \rightarrow \widetilde{P}^2 = (0, 0, \varepsilon, 0),$$
$$P^3 = (2) \rightarrow \widetilde{P}^3 = (2, 0, 0, \varepsilon).$$

From (91), we have the barycentric coordinate $\widetilde{\lambda}^0 = (\widetilde{\lambda}_1^0, \widetilde{\lambda}_2^0, \widetilde{\lambda}_3^0)$ about $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3$ of the equidistant point $\widetilde{Q}^0$ from $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3$ as

$$\widetilde{\lambda}^0 = \left(\frac{-2 + 2\varepsilon^2}{6\varepsilon^2}, \frac{1 + 2\varepsilon^2}{6\varepsilon^2}, \frac{1 + 2\varepsilon^2}{6\varepsilon^2}\right). \qquad (92)$$

For sufficiently small $|\varepsilon|$ with $\varepsilon \neq 0$, $\widetilde{\lambda}_1^0 = (-2 + 2\varepsilon^2)/(6\varepsilon^2) < 0$, thus we remove $\widetilde{P}^1$ and project $\widetilde{Q}^0$ onto $L(\widetilde{P}^2, \widetilde{P}^3)$. The barycentric coordinate $\widetilde{\lambda}^1 = (\widetilde{\lambda}_1^1, \widetilde{\lambda}_2^1, \widetilde{\lambda}_3^1)$ of $\widetilde{Q}^1 = \pi(\widetilde{Q}^0|L(\widetilde{P}^2, \widetilde{P}^3))$ about $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3$ is $\widetilde{\lambda}^1 = (0, 1/2, 1/2)$. Because $\widetilde{\lambda}_i^1 \geq 0$, $i = 1, 2, 3$, this case is the situation 1, so we have $\widetilde{\lambda}^1 = \widetilde{\lambda}^*$ by Theorem 2. Then, substituting $\varepsilon = 0$, we obtain $\widetilde{\lambda}^*|_{\varepsilon=0} = (0, 1/2, 1/2)$ and

$$Q^* = \widetilde{\lambda}^*|_{\varepsilon=0}\Phi$$
$$= \left(0, \frac{1}{2}, \frac{1}{2}\right)\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$
$$= 1.$$

Therefore, the center Q^* of $\Gamma(P^1, P^2, P^3)$ is obtained correctly.

Example 3: Let us consider four points $P^1 = (1, 2)$, $P^2 = (0, 0)$, $P^3 = (2, 0)$, $P^4 = (1, 3)$ in $\mathbb{R}^2$, which are not in general position (see Fig.6). By lifting the dimension, we have

$$P^1 = (1, 2) \rightarrow \widetilde{P}^1 = (1, 2, \varepsilon, 0, 0, 0),$$
$$P^2 = (0, 0) \rightarrow \widetilde{P}^2 = (0, 0, 0, \varepsilon, 0, 0),$$
$$P^3 = (2, 0) \rightarrow \widetilde{P}^3 = (2, 0, 0, 0, \varepsilon, 0),$$
$$P^4 = (1, 3) \rightarrow \widetilde{P}^4 = (1, 3, 0, 0, 0, \varepsilon).$$

Let $\widetilde{Q}^0$ be the equidistant point from $\widetilde{P}^1, \cdots, \widetilde{P}^4$, and $\widetilde{\lambda}^0 = (\widetilde{\lambda}_1^0, \cdots, \widetilde{\lambda}_4^0)$ be the barycentric coordinate of $\widetilde{Q}^0$ about $\widetilde{P}^1, \cdots, \widetilde{P}^4$. From (91),

$$\widetilde{\lambda}^0 = \left(\frac{-42 + 7\varepsilon^2 + 2\varepsilon^4}{54\varepsilon^2 + 8\varepsilon^4}, \frac{7 + 15\varepsilon^2 + 2\varepsilon^4}{54\varepsilon^2 + 8\varepsilon^4}, \frac{7 + 15\varepsilon^2 + 2\varepsilon^4}{54\varepsilon^2 + 8\varepsilon^4}, \frac{28 + 17\varepsilon^2 + 2\varepsilon^4}{54\varepsilon^2 + 8\varepsilon^4}\right). \qquad (93)$$

For sufficiently small $|\varepsilon|$ with $\varepsilon \neq 0$, $\widetilde{\lambda}_1^0 < 0$, thus we remove $\widetilde{P}^1$ and project $\widetilde{Q}^0$ onto $L(\widetilde{P}^2, \widetilde{P}^3, \widetilde{P}^4)$. The barycentric coordinate $\widetilde{\lambda}^1 = (\widetilde{\lambda}_1^1, \cdots, \widetilde{\lambda}_4^1)$ about $\widetilde{P}^1, \cdots, \widetilde{P}^4$ of $\widetilde{Q}^1 = \pi(\widetilde{Q}^0|L(\widetilde{P}^2, \widetilde{P}^3, \widetilde{P}^4))$ is

$$\widetilde{\lambda}^1 = \left(0, \frac{5 + \varepsilon^2}{18 + 3\varepsilon^2}, \frac{5 + \varepsilon^2}{18 + 3\varepsilon^2}, \frac{8 + \varepsilon^2}{18 + 3\varepsilon^2}\right). \qquad (94)$$

Because $\widetilde{\lambda}_i^1 \geq 0$, $i = 1, \cdots 4$, this case is the situation 1, so we have $\widetilde{\lambda}^1 = \widetilde{\lambda}^*$ by Theorem 2. Substituting $\varepsilon = 0$, we have $\widetilde{\lambda}^*|_{\varepsilon=0} = (0, 5/18, 5/18, 8/18)$. The center Q^* of $\Gamma(P^1, \cdots, P^4)$ is

$$Q^* = \widetilde{\lambda}^*|_{\varepsilon=0}\Phi \qquad (95)$$
$$= \left(0, \frac{5}{18}, \frac{5}{18}, \frac{8}{18}\right)\begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 2 & 0 \\ 1 & 3 \end{pmatrix} \qquad (96)$$
$$= \left(1, \frac{4}{3}\right). \qquad (97)$$

Hence, also in this case the center Q^* of $\Gamma(P^1, \cdots, P^4)$ is correctly obtained.

B. Lifting Dimension of Channel Matrix

In the previous section, for the points $P^1, \cdots, P^m$ in $\mathbb{R}^n$, by lifting the dimension of these points, we had $\widetilde{P}^1, \cdots, \widetilde{P}^m \in \mathbb{R}^{n+m}$ which are in general position. 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.

In the case of smallest enclosing circle, we added εe_i to P^i for lifting the dimension in (89). εe_i is in the vicinity of the origin of $\mathbb{R}^n$. In the problem of channel capacity, we will add a constant multiple of a distribution in the vicinity of the uniform distribution to a constant multiple of the row vector P^i of the channel matrix. Since the sum of the probabilities is 1, if some component is increased then another component must be decreased. By considering so, we will use a vector $((1 + \varepsilon)/2m, (1 - \varepsilon)/2m, 1/2m, \cdots, 1/2m)$ etc., which is in the vicinity of the uniform distribution
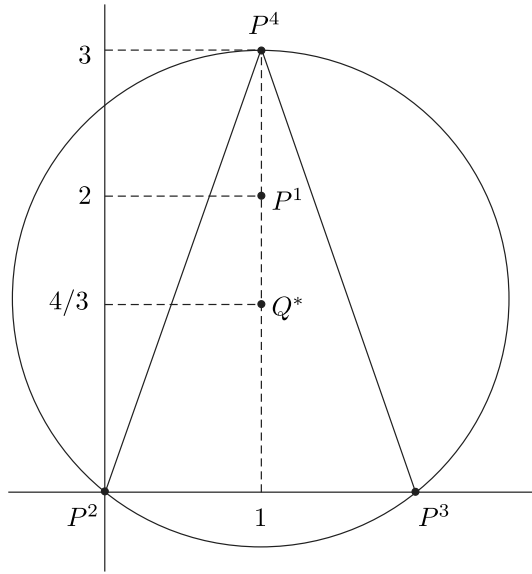
Fig. 6. Four points $P^1 = (1,2)$, $P^2 = (0,0)$, $P^3 = (2,0)$, $P^4 = (1,3)$ in $\mathbb{R}^2$.

$(1/2m, 1/2m, \cdots, 1/2m) \in \Delta^{2m}$. Now, define $\widetilde{P}^i \in \Delta^{n+2m}$ by

$$
\widetilde{P}^i = \Big(\frac{P_1^i}{2m+1}, \cdots, \frac{P_n^i}{2m+1}, \overbrace{\frac{1}{2m+1}, \cdots, \frac{1}{2m+1}}^{2i-2},
$$

$$
\overset{2i-1\,\text{th}}{\overset{\vee}{\frac{1+\epsilon}{2m+1}}}, \overset{2i\,\text{th}}{\overset{\vee}{\frac{1-\epsilon}{2m+1}}}, \overbrace{\frac{1}{2m+1}, \cdots, \frac{1}{2m+1}}^{2m-2i} \Big), \quad (98)
$$

where $|\varepsilon|$ is sufficiently small and $\varepsilon \neq 0$. Let

$$
\widetilde{\Phi} = \begin{pmatrix} \widetilde{P}^1 \\ \vdots \\ \widetilde{P}^m \end{pmatrix}, \quad (99)
$$

then rank $\widetilde{\Phi} = m$ for $\varepsilon \neq 0$.

Define an $m \times 2m$ matrix Ξ by

$$
\Xi = \frac{1}{2m}
$$

$$
\times \begin{pmatrix}
1+\epsilon & 1-\epsilon & 1 & 1 & \cdots & 1 & 1 \\
1 & 1 & 1+\epsilon & 1-\epsilon & \cdots & 1 & 1 \\
\vdots & \vdots & & \ddots & \ddots & \vdots & \vdots \\
1 & 1 & 1 & 1 & \cdots & 1+\epsilon & 1-\epsilon
\end{pmatrix},
$$

and denote by Ξ^i the i th row vector of Ξ, where $\Xi^i \in \Delta^{2m}$, $i = 1, \cdots, m$. (98) can be written as

$$
\widetilde{P}^i = \Big(\frac{1}{2m+1} P^i, \frac{2m}{2m+1} \Xi^i \Big), \quad i = 1, \cdots, m. \quad (100)
$$

1) Dummy Output Alphabet: The channel defined by (98), (99), (100) is modeled as follows. That is, let the input alphabet $\{x_1, \cdots, x_m\}$ be the same as the original one and then to the original output alphabet $\{y_1, \cdots, y_n\}$ dummy alphabet $\{y_1', \cdots, y_{2m}'\}$ is newly added. Let us define random variables

Y, Y' taking values on $\{y_1, \cdots, y_n\}$, $\{y_1', \cdots, y_{2m}'\}$, respectively, by

$$
P(Y = y_j | X = x_i) = P_j^i, \quad i = 1, \cdots, m, \, j = 1, \cdots, n,
$$
$$
P(Y' = y_{j'}' | X = x_i) = \Xi_{j'}^i, \quad i = 1, \cdots, m, \, j' = 1, \cdots, 2m,
$$

where $\Xi_{j'}^i$ denote the j'th element of the row vector Ξ^i. Further, define a random variable $\widetilde{Y}$ as the mixture of Y and Y' with weight $(1/(2m+1), 2m/(2m+1))$. Then $X \to \widetilde{Y}$ is a model of the channel defined by (98), (99), (100). The mutual information $I(\lambda, \Xi)$ between X and Y' is close to 0. Dummy alphabet is used for lifting the dimension of the row vectors of the channel matrix to make them be in general position.

2) Equation of Barycentric Coordinate: Now, for $Q = (Q_1, \cdots, Q_n) \in \Delta^n$, $T = (T_1, \cdots, T_{2m}) \in \Delta^{2m}$, let $\widetilde{Q} \in \Delta^{n+2m}$ be

$$
\widetilde{Q} = \Big(\frac{1}{2m+1} Q, \frac{2m}{2m+1} T \Big). \quad (101)
$$

We will calculate the equidistant point $\widetilde{Q} \in L(\widetilde{P}^1 \cdots, \widetilde{P}^m) \subset \Delta^{n+2m}$ from $\widetilde{P}^1 \cdots, \widetilde{P}^m$, i.e., $\widetilde{Q}$ satisfies

$$
D(\widetilde{P}^i || \widetilde{Q}) = D(\widetilde{P}^1 || \widetilde{Q}), \quad i = 2, \cdots, m. \quad (102)
$$

By calculation,

$$
D(\widetilde{P}^i || \widetilde{Q}) = \frac{-1}{2m+1} \Big\{ \sum_{j=1}^n P_j^i \log Q_j + \epsilon \log \frac{T_{2i-1}}{T_{2i}} + H(P^i)
$$
$$
+ \sum_{j'=1}^{2m} \log(2m T_{j'}) + \bar{H}(\epsilon) \Big\}, \quad (103)
$$

where we set $H(P^i) = -\sum_{j=1}^n P_j^i \log P_j^i$, $\bar{H}(\epsilon) = -(1+\epsilon)\log(1+\epsilon) - (1-\epsilon)\log(1-\epsilon)$. By (102), (103), we have

$$
\sum_{j=1}^n \Big(P_j^i - P_j^1 \Big) \log Q_j + \epsilon \log \frac{T_{2i-1}}{T_{2i}} - \epsilon \log \frac{T_1}{T_2}
$$
$$
= -H(P^i) + H(P^1), \quad i = 2, \cdots, m. \quad (104)
$$

Next, let $\widetilde{\lambda} = (\widetilde{\lambda}_1, \cdots, \widetilde{\lambda}_m)$ be the barycentric coordinate of $\widetilde{Q}$ about $\widetilde{P}^1, \cdots, \widetilde{P}^m$, i.e., $\widetilde{Q} = \widetilde{\lambda} \widetilde{\Phi}$ from (99). By (100), (101), $Q = \widetilde{\lambda} \Phi$, $T = \widetilde{\lambda} \Xi$, thus substituting these into (104), we have

$$
\sum_{j=1}^n \Big(P_j^i - P_j^1 \Big) \log \sum_{i'=1}^m \widetilde{\lambda}_{i'} P_j^{i'} + \epsilon \log \frac{1 + \varepsilon \widetilde{\lambda}_i}{1 - \varepsilon \widetilde{\lambda}_i} - \epsilon \log \frac{1 + \varepsilon \widetilde{\lambda}_1}{1 - \varepsilon \widetilde{\lambda}_1}
$$
$$
= -H(P^i) + H(P^1), \quad i = 2, \cdots, m. \quad (105)
$$

For m unknowns $\widetilde{\lambda}_1, \cdots, \widetilde{\lambda}_m$, there are $m-1$ equations (105) and one equation $\sum_{i=1}^m \widetilde{\lambda}_i = 1$, then we have the solution $\widetilde{\lambda}$. The existence and the uniqueness of the solution $\widetilde{\lambda}$ is guaranteed by Lemma 19.

Example 4: Let us consider a channel matrix

$$
\Phi = \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = \begin{pmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \\ 0.8 & 0.2 \end{pmatrix}, \quad (106)
$$

where the rows are not in general position (see Fig.7). By lifting the dimension of Φ, we have the channel matrix
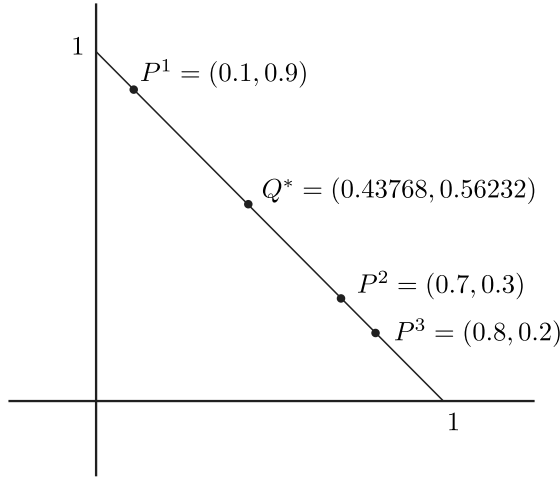
Fig. 7. Three points $P^1 = (0.1, 0.9)$, $P^2 = (0.7, 0.3)$, $P^3 = (0.8, 0.2)$ in Δ^2.
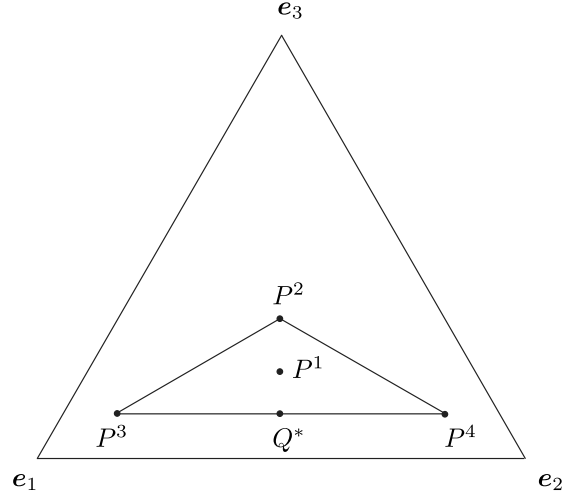

Fig. 8. Four points $P^1 = (2/5, 2/5, 1/5)$, $P^2 = (1/3, 1/3, 1/3)$, $P^3 = (4/5, 1/10, 1/10)$, $P^4 = (1/10, 4/5, 1/10)$ in Δ^3.

$\widetilde{\Phi}$ of (99) by

$$P^1 = (0.1, 0.9) \to \widetilde{P}^1 = (0.1, 0.9, 1+\varepsilon, 1-\varepsilon, 1, 1, 1, 1)/7,$$
$$P^2 = (0.7, 0.3) \to \widetilde{P}^2 = (0.7, 0.3, 1, 1, 1+\varepsilon, 1-\varepsilon, 1, 1)/7,$$
$$P^3 = (0.8, 0.2) \to \widetilde{P}^3 = (0.8, 0.2, 1, 1, 1, 1, 1+\varepsilon, 1-\varepsilon)/7.$$

Let $\widetilde{\lambda}^0$ be the barycentric coordinate about $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3$ of the equidistant point $\widetilde{Q}^0$ from $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3$. Solving the equation (105) with $\varepsilon = 0.05$, we have

$$\widetilde{\lambda}^0 = (2.39, -12.96, 11.57). \tag{107}$$

Since $\widetilde{\lambda}^0_2 = -12.96 < 0$, we remove $\widetilde{P}^2$ and project $\widetilde{Q}^0$ onto $L(\widetilde{P}^1, \widetilde{P}^3)$ to have $\widetilde{Q}^1 = \pi(\widetilde{Q}^0|L(\widetilde{P}^1, \widetilde{P}^3))$. Let $\widetilde{\lambda}^1$ be the barycentric coordinate of $\widetilde{Q}^1$ about $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3$. By calculation $\widetilde{\lambda}^1 = (0.5175, 0, 0.4825)$, so this case is the situation 1 in the Appendix. Therefore, we have $\widetilde{\lambda}^* = \widetilde{\lambda}^1$ by Theorem 13 and $\widetilde{\lambda}^*|_{\varepsilon=0} = (0.5176, 0, 0.4824)$. Thus,

$$Q^* = \widetilde{\lambda}^*|_{\varepsilon=0} \Phi \tag{108}$$

$$= (0.5176, 0, 0.4824) \begin{pmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \\ 0.8 & 0.2 \end{pmatrix} \tag{109}$$

$$= (0.43768, 0.56232), \tag{110}$$

(see Fig.7). The channel capacity is $C = D(P^1 \| Q^*) = 0.398$ [bit/symbol]. So, the capacity achieving Q^* and the capacity C of Φ are correctly obtained.

Example 5: Consider the channel matrix

$$\Phi = \begin{pmatrix} P^1 \\ P^2 \\ P^3 \\ P^4 \end{pmatrix} = \begin{pmatrix} 2/5 & 2/5 & 1/5 \\ 1/3 & 1/3 & 1/3 \\ 4/5 & 1/10 & 1/10 \\ 1/10 & 4/5 & 1/10 \end{pmatrix}, \tag{111}$$

where the rows are not in general position (see Fig.8). By lifting the dimension of Φ, we have the channel matrix

$\widetilde{\Phi}$ of (99) by

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

Solving the equation (105) with $\varepsilon = 0.05$, we have the barycentric coordinate $\widetilde{\lambda}^0$ of the equidistant point $\widetilde{Q}^0$ from $\widetilde{P}^1, \widetilde{P}^2, \widetilde{P}^3, \widetilde{P}^4$ as

$$\widetilde{\lambda}^0 = (-19.00, 7.98, 6.01, 6.01). \tag{112}$$

Since $\widetilde{\lambda}^0_1 = -19.00 < 0$, 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

$$\widetilde{\lambda}^1 = (0, -0.14, 0.57, 0.57). \tag{113}$$

Since $\widetilde{\lambda}^1_2 = -0.14 < 0$, 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

$$\widetilde{\lambda}^2 = (0, 0, 1/2, 1/2). \tag{114}$$

Thus, this case is the situation 1 of the Appendix, so we have $\widetilde{\lambda}^* = \widetilde{\lambda}^2$ from Theorem 13. $\widetilde{\lambda}^2$ in (114) does not depend on $\varepsilon \in \mathbb{R}$, thus we have $\widetilde{\lambda}^*|_{\varepsilon=0} = \widetilde{\lambda}^2 = (0, 0, 1/2, 1/2)$. Therefore,

$$Q^* = \widetilde{\lambda}^*|_{\varepsilon=0} \Phi \tag{115}$$

$$= (0, 0, 1/2, 1/2) \begin{pmatrix} 2/5 & 2/5 & 1/5 \\ 1/3 & 1/3 & 1/3 \\ 4/5 & 1/10 & 1/10 \\ 1/10 & 4/5 & 1/10 \end{pmatrix} \tag{116}$$

$$= (9/20, 9/20, 1/10), \tag{117}$$

TABLE II

SIMULATION RESULTS OF HEURISTIC ALGORITHM FOR
SMALLEST ENCLOSING CIRCLE (HS)

m \ n	2	3	10	20
3	100%	100%	100%	100%
4	97.6%	99.7%	100%	100%
5	96.3%	95.0%	100%	100%
8	94.4%	85.3%	99.5%	100%
10	92.9%	82.5%	98.5%	100%

and the channel capacity is $C = D(P^3 \| Q^*) = 0.447$ [bit/symbol]. Also in this case, the capacity achieving Q^* and the capacity C of Φ are correctly obtained.

V. HEURISTIC ALGORITHM OF CALCULATING SMALLEST ENCLOSING CIRCLE AND CHANNEL CAPACITY

In this chapter, we will propose heuristic projection algorithms for the calculation of the smallest enclosing circle and the channel capacity with arbitrary m, n and arbitrary placement of points, based on the algorithms in chapter II, III and the dimension lifting in chapter IV. In chapters II, III, we gave algorithms to calculate the smallest enclosing circle and the channel capacity under the limited situations, i.e., situation 1 and situation 2. The methods which will be proposed in this chapter can be applied to any number of points and any placement of points, but the obtained results are not always correct. In the following, we describe heuristic algorithms and apply them to many concrete problems generated by random numbers, then show the percentage of getting correct solutions.

A. Heuristic Algorithm for Smallest Enclosing Circle (HS)

1) For m points $P^1, \cdots, P^m$ in $\mathbb{R}^n$, lift the dimension by (87) or (89) to obtain $\widetilde{P}^1, \cdots, \widetilde{P}^m$.
2) Calculate by (91) the barycentric coordinate $\widetilde{\lambda}^0 = (\widetilde{\lambda}_1^0, \cdots, \widetilde{\lambda}_m^0)$ 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 are shown in TABLE II. Simulation method is as follows: Generate n uniform random integers from -1000 to 1000 and let them be the coordinate of one point in $\mathbb{R}^n$. Repeat it m times to obtain m points of $\mathbb{R}^n$, and let them $P^1, \cdots, P^m$. Calculate the center Q^* of the smallest enclosing circle by the heuristic algorithm HS. The above m time generation of points and the calculation of Q^* are defined to be one set. Then 10000 sets are executed for each pair of m, n in TABLE II to show the percentage that the correct Q^* is obtained. Hence, in total, 20 million sets
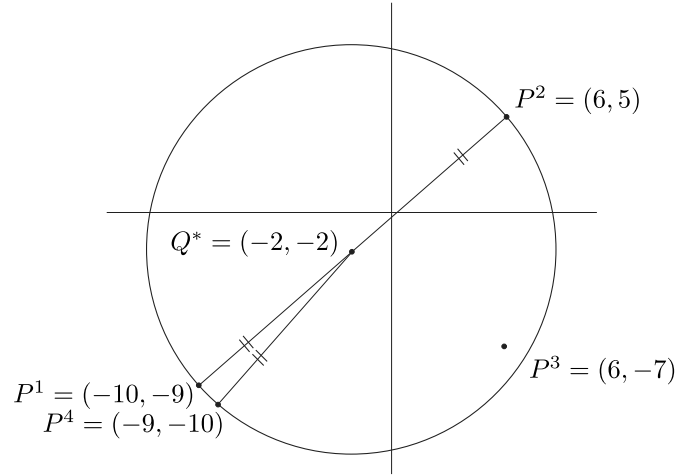

Fig. 9.   Example of placement of points for which our algorithm HS fails.

are executed in the whole TABLE II. The correctness of the obtained Q^* is checked by the Kuhn-Tucker condition (16).

In every case of TABLE II, we have succeeded in obtaining the correct solutions at fairly high percentages. From these results, we can consider that many of actual placements of points are the situation 1. For a fixed value of m, the percentage of success increases as n increases. This is because the placements of points are easier to become in general position for a larger dimension n. Further, for a fixed value of n, the percentage of success decreases as m increases. This is because there exist so many points in a low dimensional space that our HS becomes difficult to succeed.

1) Example of Placement of Points for Which our Algorithm Fails: We will show an example of the placement of points for which the correct smallest enclosing circle cannot be obtained by our heuristic algorithm HS (see Fig.9). Consider 4 points $P^1 = (-10, -9)$, $P^2 = (6, 5)$, $P^3 = (6, -7)$, $P^4 = (-9, -10)$ in $\mathbb{R}^2$. The center Q^* of $\Gamma = \Gamma(P^1, \cdots, P^4)$ is $Q^* = (-2, -2)$, which is the midpoint of $P^1 P^2$. Thus, P^1 is necessary for determining Γ. However, if we apply our HS to this placement of points, we have

$$\lambda^0 = \left(\frac{-64800 + 22536\varepsilon^2 + 357\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6}, \right.$$

$$\frac{6480 + 43350\varepsilon^2 + 523\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6},$$

$$\frac{-10800 - 13278\varepsilon^2 + 319\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6},$$

$$\left. \frac{69120 + 17880\varepsilon^2 + 343\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6} \right).$$

Because λ_1^0 is negative and the smallest among $\lambda_1^0, \cdots, \lambda_4^0$ for sufficiently small $|\varepsilon|$ with $\varepsilon \neq 0$, our HS removes P^1. Hence, in this case our HS fails.

B. Heuristic Algorithm for Channel capacity (HC)

1) For m probability distributions $P^1, \cdots, P^m$ in Δ^n, lift the dimension by (98) or (100) to obtain $\widetilde{P}^1, \cdots, \widetilde{P}^m$.

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%

TABLE IV
RATIO N_2/N_1 OF COMPUTATIONAL COMPLEXITY

m \ n	2	3	10	20
3	0.250	0.250	0.250	0.250
4	0.200	0.182	0.182	0.182
5	0.150	0.120	0.115	0.115
8	0.071	0.039	0.024	0.024
10	0.048	0.021	0.008	0.008

2) Calculate by (105) the barycentric coordinate $\widetilde{\lambda}^0 = (\widetilde{\lambda}_1^0, \cdots, \widetilde{\lambda}_m^0)$ 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 are shown in TABLE III. Simulation method is as follows: Generate n uniform random numbers U_j in $(0, 1)$, $j = 1, \cdots, n$ and put $S = \sum_{j=1}^n U_j$. Then define $P = (U_1/S, \cdots, U_n/S) \in \Delta^n$. Repeat it m times to obtain m probability distributions $P^1, \cdots, P^m \in \Delta^n$. Calculate the capacity achieving Q^* by the heuristic algorithm HC. The above m time generation of probability distributions and the calculation of Q^* are defined to be one set. Then 10000 sets are executed for each pair of m, n in TABLE III to show the percentage that the correct Q^* is obtained. The correctness of the obtained Q^* is checked by the Kuhn-Tucker condition (7). The failures of HC in TABLE III include computational errors by the function FindRoot in Mathematica.

In every case of TABLE III, the percentage of success is more than 99%. But, in comparison with TABLE II of the smallest enclosing circle, we do not understand well the reason why TABLE III has higher rate of success, although the geometric structures of the two problems are similar.

C. Comparison of Computational Complexity

For given m points $P^1, \cdots, P^m \in \mathbb{R}^n$ or $\in \Delta^n$, the simplest way to calculate the center of the smallest enclosing circle or the channel capacity achieving distribution is to examine all the possible combinations of the points [6]. This is a brute force method. For example, consider the case of 4 points in $\mathbb{R}^2$. Because a circle in $\mathbb{R}^2$ is determined by two or three points, the total number of all the possible combinations of the points in the brute force method is $_4C_2 + _4C_3 = 10$. In a similar way, the total number N_1 of all the possible combinations of m points in $\mathbb{R}^n$ or Δ^n is

$$N_1 = \sum_{\ell=0}^{\min(m, n+1)} {}_mC_\ell. \tag{118}$$

If we consider the computational complexity for one combination as 1, then the computational complexity of the brute force method is N_1.

On the other hand, in our proposed heuristic algorithms, the number of points is decreasing by one at every time, hence the computational complexity N_2 is

$$N_2 = m - 2. \tag{119}$$

We will show in TABLE IV the ratio N_2/N_1 for the case of TABLE II and III. N_2 is very smaller than N_1, however, note that the proposed method does not necessarily yield the correct solutions.

VI. CONCLUSION

The contributions of this paper are as follows:

- We showed that the problem of channel capacity has a similar geometric structure to the problem of smallest enclosing circle in $\mathbb{R}^n$. Then, based on this geometric similarity, we developed a projection algorithm for calculating the channel capacity.
- We obtained theorems of searching for the capacity achieving output distribution by projection, and showed that the barycentric coordinate plays an important role.
- For the channel matrix Φ which does not have the equidistant point, we newly defined a channel matrix $\widetilde{\Phi}$ which is close to Φ and has the equidistant point. $\widetilde{\Phi}$ is made by lifting the dimension of the row vectors of Φ. Then, we proposed a heuristic algorithm to calculate the capacity achieving output distribution. Sometimes in some problem, the difficulty can be relaxed by lifting dimension. Our method is one example of this lifting dimension.

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

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

APPENDIX

A. Search for Q^* by Projection Algorithm

For a given channel matrix Φ in (2), the channel capacity C and the capacity achieving distribution Q^* are obtained by the projection algorithm, under the assumption (70) that the row vectors of the channel matrix Φ are in general position and in the situations 1 and 2 below.

1) Situation 1 [There Is Just One Negative Component of Barycentric Coordinate at Every Projection.]: We consider the following assumption similar to section II-F1.

Assumption of Situation 1: Assume $P^1, \cdots, P^m \in \Delta^n$ 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. That is, for $k = 0, 1, \cdots, K-1$,

$$\lambda_i^k \begin{cases} = 0, i = 1, \cdots, k, & (120a) \\ < 0, i = k+1, & (120b) \\ > 0, i = k+2, \cdots, m, & (120c) \end{cases}$$

and for $k = K$,

$$\lambda_i^K \begin{cases} = 0, i = 1, \cdots, K, & (121a) \\ > 0, i = K+1, \cdots, m. & (121b) \end{cases}$$

If $K = 0$, we assume $\lambda_i^0 \geq 0$, $i = 1, \cdots, m$. (The end of Assumption of situation 1)

Here, we consider the inner products $\sigma_i = (P^i - Q^K, Q^0 \ominus Q^K)$, $i = 1, \cdots, m$.

Lemma 23: $\sigma_i \begin{cases} < 0, i = 1, \cdots, K, \\ = 0, i = K+1, \cdots, m. \end{cases}$

Proof: By Lemma 20, $Q^K = \pi(Q^0|L_K)$, and since $P^{K+1} \cdots, P^m \in L_K$, we have, by Lemma 18,

$$\sigma_i = 0, \quad i = K+1, \cdots, m. \tag{122}$$

Next, we will prove $\sigma_i < 0$, $i = 1, \cdots, K$ by mathematical induction in the order of $i = K, K-1, \cdots, 1$.
(I) Prove $\sigma_K < 0$:

By (120a), (122), we have

$$\lambda_K^{K-1}\sigma_K = \sum_{i=1}^m \lambda_i^{K-1}\sigma_i$$
$$= (Q^{K-1} - Q^K, Q^0 \ominus Q^K)$$
$$= (Q^{K-1} - Q^K, Q^{K-1} \ominus Q^K) \quad \text{(by Lemma 21)}$$
$$> 0 \quad \text{(by Lemma 22)}. \tag{123}$$

By (120b), $\lambda_K^{K-1} < 0$, thus we obtain $\sigma_K < 0$ by (123).
(II) Assuming $\sigma_K < 0$, $\sigma_{K-1} < 0, \cdots, \sigma_{k+1} < 0$, prove $\sigma_k < 0$:

By (120a), (122), we have

$$\lambda_k^{k-1}\sigma_k + \sum_{i=k+1}^K \lambda_i^{k-1}\sigma_i = \sum_{i=1}^m \lambda_i^{k-1}\sigma_i$$
$$= (Q^{k-1} - Q^K, Q^0 \ominus Q^K)$$
$$= (Q^{k-1} - Q^K, Q^{k-1} \ominus Q^K)$$
$$> 0. \tag{124}$$

By the induction hypothesis, $\sigma_i < 0$, $i = k+1, \cdots, K$, and by (120b), $\lambda_k^{k-1} < 0$, by (120c), $\lambda_i^{k-1} > 0$, $i = k+1, \cdots, K$, thus by (124), we obtain $\sigma_k < 0$.
(III) By the above (I), (II), we have $\sigma_i < 0$, $i = 1, \cdots, K$. $\square$

Lemma 24: $D(P^i\|Q^K) < D(P^{K+1}\|Q^K) = \cdots = D(P^m\|Q^K)$, $i = 1, \cdots, K$.

Proof: By Lemma 23 and Theorem 12. $\square$

Theorem 13: The output distribution that achieves the channel capacity is $Q^* = Q^K$, and the channel capacity is $C = D(P^{K+1}\|Q^K)$.

Proof: By (121a), (121b), Lemma 24 and the Kuhn-Tucker condition (7). $\square$

We see that the above Lemmas 23, 24, Theorem 13 and their proofs are very similar to Lemmas 7, 8, Theorem 2 and their proofs in the problem of smallest enclosing circle in chapter II. This is because that the problems of smallest enclosing circle and channel capacity can be solved using only common properties of α-geometry.

2) Situation 2 [$m = 2, 3, 4$ and n Is Arbitrary.]: We will calculate by the projection algorithm the capacity achieving output distribution Q^* and the channel capacity C, under the assumption that the row vectors $P^1, \cdots, P^m$ of the channel matrix Φ in (2) are in general position in Δ^n.

Similarly to chapter II, our goal is to find an algorithm to calculate Q^* and C for every m, but at present it is possible to solve only for $m = 2, 3, 4$.

As can be seen from the results of the previous section, the proofs of lemmas and theorems in the case of smallest enclosing circle are almost the same as those of channel capacity. Because those are similar also in this section, the proofs of the following lemmas and theorems will be described simply.

Case of $m = 2$

Consider the channel capacity of a channel matrix Φ with two different row vectors $P^1, P^2 \in \Delta^n$. Let $L_0 = L(P^1, P^2)$, and $Q^0 \in L_0$ be the equidistant point from P^1, P^2, i.e.,

$$D(P^1\|Q^0) = D(P^2\|Q^0), \ Q^0 \in L_0. \tag{125}$$

Let $\boldsymbol{\lambda}^0 = (\lambda_1^0, \lambda_2^0)$ be the barycentric coordinate of Q^0 about P^1, P^2. Because Q^0 is between P^1 and P^2 in L_0, we have $\lambda_1^0 > 0, \lambda_2^0 > 0$. Thus, by (125) and the Kuhn-Tucker condition (7), we have

Theorem 14: The output distribution that achieves the channel capacity is $Q^* = Q^0$, and the channel capacity is $C = D(P^1 \| Q^0)$.

Case of $m = 3$

Suppose the channel matrix consists of three row vectors $P^1, P^2, P^3 \in \Delta^n$ in general position. Let $L_0 = L(P^1, P^2, P^3)$ and let $Q^0 \in L_0$ be the equidistant point from P^1, P^2, P^3. Let $\boldsymbol{\lambda}^0 = (\lambda_1^0, \lambda_2^0, \lambda_3^0)$ be the barycentric coordinate of Q^0 about P^1, P^2, P^3.

Based on the signs of the components of $\boldsymbol{\lambda}^0$, we have the following classification without loss of generality.

Case 3-1: $\lambda_1^0 \geq 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0$
Case 3-2: $\lambda_1^0 < 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0$

As in the problem of smallest enclosing circle, all the cases are exhausted by Case 3-1 and Case 3-2.

Case 3-1

By the Kuhn-Tucker condition (7), we have

Theorem 15: The output distribution that achieves the channel capacity is $Q^* = Q^0$ and the channel capacity is $C = D(P^1 \| Q^0)$.

Case 3-2

Similarly to the Case 3-2 in chapter II, this case is the situation 1, hence we have

Theorem 16: The output distribution that achieves the channel capacity is $Q^* = Q^1$ and the channel capacity is $C = D(P^2 \| Q^1)$.

Case of $m = 4$

Suppose the channel matrix Φ consists of four row vectors $P^1, P^2, P^3, P^4 \in \Delta^n$ in general position. Let $L_0 = L(P^1, \cdots, P^4)$ and $Q^0 \in L_0$ be the equidistant point from $P^1, \cdots, P^4$. Let $\boldsymbol{\lambda}^0 = (\lambda_1^0, \cdots, \lambda_4^0)$ be the barycentric coordinate of Q^0 about $P^1, \cdots, P^4$.

Without loss of generality, we have the following classification:

Case 4-1: $\lambda_1^0 \geq 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$
Case 4-2: $\lambda_1^0 < 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$
Case 4-3: $\lambda_1^0 < 0, \lambda_2^0 < 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$

Case 4-1

By the Kuhn-Tucker condition (7), we have

Theorem 17: The output distribution that achieves the channel capacity is $Q^* = Q^0$ and the channel capacity is $C = D(P^1 \| Q^0)$.

Case 4-2

Let $L_1 = L(P^2, P^3, P^4)$ and $Q^1 = \pi(Q^0 | L_1)$, and denote by $\boldsymbol{\lambda}^1 = (\lambda_1^1, \cdots, \lambda_4^1)$ the barycentric coordinate of Q^1 about $P^1, \cdots, P^4$. Since $P^1 \notin L_1$, we have $\lambda_1^1 = 0$.

Based on the signs of the components of $\boldsymbol{\lambda}^1$, we have the following classification without loss of generality.

Case 4-2-1: $\lambda_1^1 = 0, \lambda_2^1 \geq 0, \lambda_3^1 \geq 0, \lambda_4^1 \geq 0$
Case 4-2-2: $\lambda_1^1 = 0, \lambda_2^1 < 0, \lambda_3^1 \geq 0, \lambda_4^1 \geq 0$
Consider the inner products $\sigma_i = (P^i - Q^1, Q^0 \ominus Q^1)$, $i = 1, \cdots, 4$.

Case 4-2-1

This case is the situation 1, hence we have

Theorem 18: The output distribution that achieves the channel capacity is $Q^* = Q^1$ and the channel capacity is $C = D(P^2 \| Q^1)$.

Case 4-2-2

Let us define $Q^2 = \pi(Q^1 | L_2)$ with $L_2 = L(P^3, P^4)$, then Q^2 is between P^3 and P^4 in the line L_2. Denoting by $\boldsymbol{\lambda}^2 = (\lambda_1^2, \cdots, \lambda_4^2)$ the barycentric coordinate of Q^2 about $P^1, \cdots, P^4$, we have $\lambda_1^2 = \lambda_2^2 = 0, \lambda_3^2 > 0, \lambda_4^2 > 0$, hence this is the situation 1. So, we have

Theorem 19: Let $Q^2 = \pi(Q^1 | L(P^3, P^4))$. The output distribution that achieves the channel capacity is $Q^* = Q^2$ and the channel capacity is $C = D(P^3 \| Q^2)$.

This completes Case 4-2, then let us consider Case 4-3.

Case 4-3

Here, let us define $Q^{1(1)} = \pi(Q^0 | L(P^2, P^3, P^4))$ and $Q^{1(2)} = \pi(Q^0 | L(P^1, P^3, P^4))$. Then denote by $\boldsymbol{\lambda}^{1(1)} = (\lambda_1^{1(1)}, \cdots, \lambda_4^{1(1)})$ and $\boldsymbol{\lambda}^{1(2)} = (\lambda_1^{1(2)}, \cdots, \lambda_4^{1(2)})$ the barycentric coordinates of $Q^{1(1)}$ and $Q^{1(2)}$ about $P^1, \cdots, P^4$, respectively. Consider the inner products

$$\sigma_i^{1(1)} = (P^i - Q^{1(1)}, Q^0 \ominus Q^{1(1)}), \ i = 1, \cdots, 4, \ (126)$$
$$\sigma_i^{1(2)} = (P^i - Q^{1(2)}, Q^0 \ominus Q^{1(2)}), \ i = 1, \cdots, 4. \ (127)$$

Lemma 25: $\sigma_i^{1(1)} \begin{cases} < 0, \ i = 1, \\ = 0, \ i = 2, 3, 4, \end{cases}$

$\sigma_i^{1(2)} \begin{cases} < 0, \ i = 2, \\ = 0, \ i = 1, 3, 4. \end{cases}$

Lemma 26: $\lambda_1^{1(2)} \sigma_1^{1(1)} + \lambda_2^{1(1)} \sigma_2^{1(2)} > 0$.

The above Lemmas 25, 26 are proved in the same way as Lemmas 9, 10.

Now from Lemmas 25, 26, we see that $\lambda_1^{1(2)} < 0$ or $\lambda_2^{1(1)} < 0$ holds. So far, the points P^1 and P^2 have been treated exactly equally, hence without loss of generality, we assume $\lambda_1^{1(2)} < 0$. Based on the signs of the components of $\boldsymbol{\lambda}^{1(1)}$, we have the following classification:

Case 4-3-1: $\lambda_1^{1(1)} = 0, \lambda_2^{1(1)} \geq 0, \lambda_3^{1(1)} \geq 0, \lambda_4^{1(1)} \geq 0$
Case 4-3-2: $\lambda_1^{1(1)} = 0, \lambda_2^{1(1)} < 0, \lambda_3^{1(1)} \geq 0, \lambda_4^{1(1)} \geq 0$
Case 4-3-3: $\lambda_1^{1(1)} = 0, \lambda_2^{1(1)} \geq 0, \lambda_3^{1(1)} < 0, \lambda_4^{1(1)} \geq 0$

Case 4-3-1

Theorem 20: The output distribution that achieves the channel capacity is $Q^* = Q^{1(1)}$ and the channel capacity is $C = D(P^2 \| Q^{1(1)})$.

Proof: Similar to the proof of Theorem 9. $\square$

Case 4-3-2

Define $Q^2 \equiv \pi(Q^{1(1)} | L(P^3, P^4)) (= \pi(Q^0 | L(P^3, P^4)))$, and consider the inner products $\sigma_i^2 = (P^i - Q^2, Q^0 \ominus Q^2)$, $i = 1, \cdots, 4$.

Lemma 27: $\sigma_i^2 \begin{cases} < 0, \ i = 1, 2, \\ = 0, \ i = 3, 4. \end{cases}$

Lemma 28: $D(P^i \| Q^2) < D(P^3 \| Q^2) = D(P^4 \| Q^2)$, $i = 1, 2$.

Theorem 21: The output distribution that achieves the channel capacity is $Q^* = Q^2$ and the channel capacity is $C = D(P^3 \| Q^2)$.

The above Lemmas 27, 28 and Theorem 21 are proved in the same way as Lemmas 11, 12 and Theorem 10.

Case 4-3-3

Let us define $Q^{1(3)} = \pi(Q^0|L(P^1, P^2, P^4))$, and denote by $\lambda^{1(3)} = (\lambda_1^{1(3)}, \cdots, \lambda_4^{1(3)})$ the barycentric coordinate of $Q^{1(3)}$ about $P^1, \cdots, P^4$. Consider the inner products $\sigma_i^{1(3)} = (P^i - Q^{1(3)}, Q^0 \ominus Q^{1(3)})$, $i = 1, \cdots, 4$.

Lemma 29: $\sigma_i^{1(3)} \begin{cases} > 0, & i = 3, \\ = 0, & i = 1, 2, 4. \end{cases}$

Lemma 30: $\lambda_1^{1(3)} < 0$.

Theorem 22: Let $Q^{2\dagger} \equiv \pi(Q^0|L(P^2, P^4))$. The output distribution that achieves the channel capacity is $Q^* = Q^{2\dagger}$ and the channel capacity is $C = D(P^2 \| Q^{2\dagger})$.

The above Lemmas $29, 30$ and Theorem 22 are proved in the same way as Lemmas $13, 14$ and Theorem 11.

Summarizing the classification of case $m = 4$ in chapters II and III, we have the signs of barycentric coordinates and the pair (Q^*, d^*) as follows:

Case 4-1 $(+ + + +)$ $(Q^0, d(P^1, Q^0))$

| Case 4-2 $(- + + +)$ |
| Case 4-2-1 $(0 + + +)$ $(Q^1, d(P^2, Q^1))$ |
| Case 4-2-1 $(0 - + +)$ $(Q^2, d(P^3, Q^2))$ |

| Case 4-3 $(- - + +)$ |
| Case 4-3-1 $(0 + + +)$ $(Q^{1(1)}, d(P^2, Q^{1(1)}))$ |
| Case 4-3-2 $(0 - + +)$ $(Q^2, d(P^3, Q^2))$ |
| Case 4-3-3 $(0 + - +)$ $(Q^{2\dagger}, d(P^2, Q^{2\dagger}))$ |

REFERENCES

[1] S. I. Amari, *Differential-Geometrical Methods in Statistics* (Lecture Notes in Statistics). Springer-Verlag, 1985.

[2] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[3] D. Cheng, X. Hu, and C. Martin, "On the smallest enclosing balls," *Commun. Inf. Syst.*, vol. 6, no. 2, pp. 137–160, 2006.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic, 1982.

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

[6] S. Muroga, "On the capacity of a discrete channel. I mathematical expression of capacity of a channel which is disturbed by noise in its every one symbol and expressible in one state diagram," *J. Phys. Soc. Jpn.*, vol. 8, pp. 484–494, 1953.

[7] K. Nakagawa and F. Kanaya, "A new geometric capacity characterization of a discrete memoryless channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 318–321, Mar. 1988.

[8] S. Schönherr, "Quadratic programming in geometric optimization: Theory, implementation, and applications," Ph.D. dissertation, Swiss Federal Inst. Technol., Zürich, Switzerland. [Online]. Available: http://www.inf.ethz.ch/personal/emo/DoctThesisFiles/schoenherr02.pdf

[9] G. Strang, *Linear Algebra and Its Applications*. New York, NY, USA: Academic, 1976.

Kenji Nakagawa (M'85) received the B.S., M.S. and D.S. degrees from Tokyo Institute of Technology, Tokyo, Japan, 1980, 1982 and 1986, respectively. In 1985, he joined NTT (Nippon Telegraph and Telephone Corp.). Since 1992, he has been an associate professor of Nagaoka University of Technology, and from 2012 he is a professor. His research interests include information theory, performance evaluation of network, and queueing theory. He is a member of IEEE, IEICE, Operations Research Society of Japan, Mathematical Society of Japan and Japanese Society of Engineering Education.

Kohei Watabe (S'11–M'14) received his B.E. and M.E. degrees in Engineering from Tokyo Metropolitan University, Tokyo, Japan, in 2009 and 2011, respectively. He also received the Ph.D. degree from Osaka University, Japan, in 2014. He was a JSPS research fellow (DC2) from April 2012 to March 2014. He has been an Assistant Professor of Department of Electrical Engineering, Nagaoka University of Technology since April 2014. He is a member of the IEEE and the IEICE.

Takuto Sabu graduated from National Institute of Technology, Toyama college in 2013. He received the B.E. and M.E. degrees from Nagaoka University of Technology, Niigata, Japan, in 2015 and 2017, respectively. The theme of his master thesis is the calculation algorithm of the channel capacity.