

確率分布に関する凸最適化問題の 最適解探索アルゴリズム

中川 健治, 渡部 康平, 佐武 拓斗
長岡技術科学大学

待ち行列研究部会 2016 年 7 月 16 日 (土) 東京工業大学

発表の概要

1. 通信路容量とユークリッド空間における最小包含円との関係について
2. 最小包含円の中心および通信路容量を達成する出力分布を計算するアルゴリズム
 - P^1, \dots, P^m からの等距離点
 - 内積, Pythagorasの定理, 射影
 - 探索アルゴリズム
3. P^1, \dots, P^m の配置が任意のときのアルゴリズム
 - P^1, \dots, P^m の次元を上げる方法
4. ヒューリスティックな探索アルゴリズム
 - 数値例
5. まとめと今後の課題

離散的無記憶通信路 $X \longrightarrow Y$

入力アルファベット $\{x_1, \dots, x_m\}$,

出力アルファベット $\{y_1, \dots, y_n\}$

x_i を送信したとき y_j が受信される条件付き確率

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

行ベクトル

$$P^i = (P_1^i, \dots, P_n^i), i = 1, \dots, m$$

通信路行列

$$\Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} = \begin{pmatrix} P_1^1 & \dots & P_n^1 \\ \vdots & & \vdots \\ P_1^m & \dots & P_n^m \end{pmatrix}$$

入力分布全体

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

出力分布全体

$$\bar{\Delta}^n = \{Q = (Q_1, \dots, Q_n) | Q_j \geq 0, j = 1, \dots, n, \sum_{j=1}^n Q_j = 1\}$$

入力分布 $\boldsymbol{\lambda} \in \bar{\Delta}^m$ に対する出力分布 $Q \in \bar{\Delta}^n$

$$Q = \boldsymbol{\lambda}\Phi, Q_j = \sum_{i=1}^m \lambda_i P_j^i, j = 1, \dots, n$$

$$\text{相互情報量} : I(\boldsymbol{\lambda}, \Phi) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i P_j^i \log \frac{P_j^i}{Q_j}$$

$$\text{通信路容量} : C = \max_{\boldsymbol{\lambda} \in \bar{\Delta}^m} I(\boldsymbol{\lambda}, \Phi)$$

通信路容量 : $C = \max_{\lambda \in \bar{\Delta}^m} I(\lambda, \Phi)$

$\bar{\Delta}^m$ は凸集合, $I(\lambda, \Phi)$ は λ に関して上に凸な関数

$$\lambda \in \bar{\Delta}^m \Leftrightarrow \sum_{i=1}^m \lambda_i = 1, \lambda_i \geq 0, i = 1, \dots, m$$

つまり, 上記の max 問題は不等式制約 $\lambda_i \geq 0$ がある最適化問題であり, 解くことが難しい。

$Q, Q' \in \bar{\Delta}^n$ に対して, Kullback-Leibler 情報量 (divergence) を

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

と定義する。

通信路容量 $C = \max_{\lambda \in \bar{\Delta}^m} I(\lambda, \Phi)$ に対する Kuhn-Tucker 条件 :

入力分布 $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*) \in \bar{\Delta}^m$ が通信路容量 C を達成するための必要十分条件は, ある値 C_0 が存在して,

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

を満たすことである。このとき, C_0 は C に等しい。

この Kuhn-Tucker 条件で問題が解決するように見えるが, 解法には場合分けが必要で計算量が大きい。

例題：最小化問題（最適化と変分法（丸善出版） p.72）

$$\min x_1^2 + 2x_2^2$$

制約条件：

$$g_1(\mathbf{x}) = -x_1 - x_2 + 1 \leq 0, \quad g_2(\mathbf{x}) = x_1 - 2x_2 - 1 \leq 0$$

Lagrange関数 $F(\mathbf{x}, \boldsymbol{\mu}) = x_1^2 + 2x_2^2 + \mu_1 g_1(\mathbf{x}) + \mu_2 g_2(\mathbf{x})$

Kuhn-Tucker条件

$$\partial F / \partial x_1 = 2x_1 - \mu_1 + \mu_2 = 0$$

$$\partial F / \partial x_2 = 4x_2 - \mu_1 - 2\mu_2 = 0$$

$$g_1(\mathbf{x}) \leq 0, \quad \mu_1 \geq 0, \quad \mu_1 g_1(\mathbf{x}) = 0$$

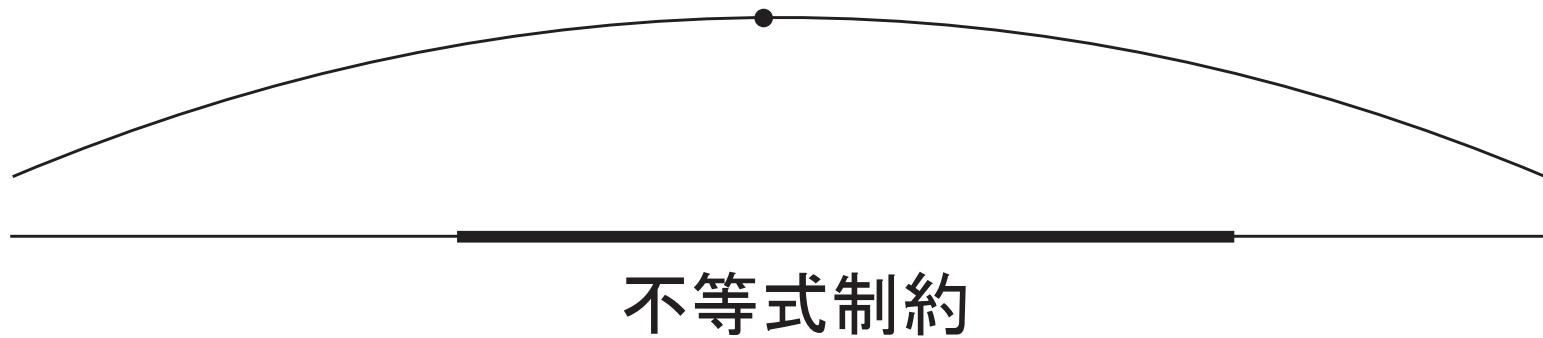
$$g_2(\mathbf{x}) \leq 0, \quad \mu_2 \geq 0, \quad \mu_2 g_2(\mathbf{x}) = 0$$

場合分け (a) $\mu_1 = 0, \mu_2 = 0$ (b) $g_1(\mathbf{x}) = 0, \mu_2 = 0$

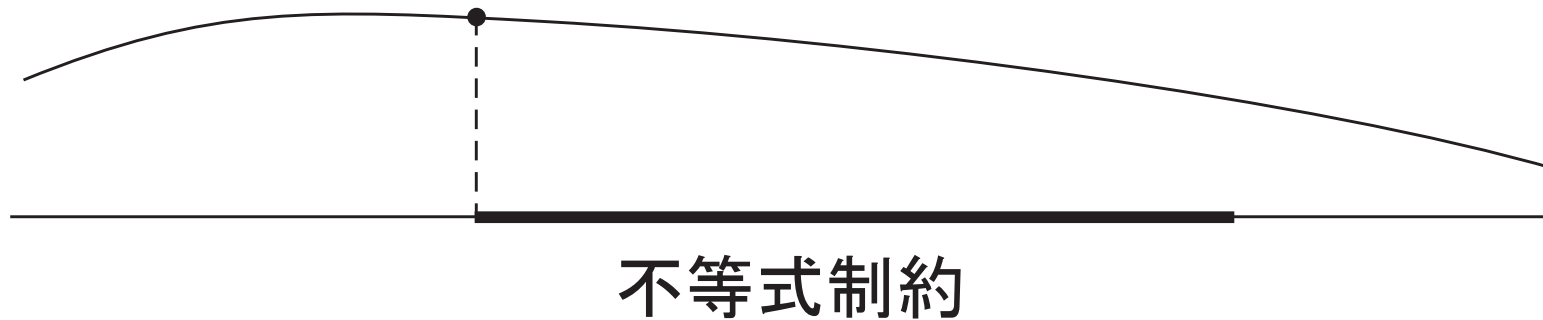
(c) $\mu_1 = 0, g_2(\mathbf{x}) = 0$ (d) $g_1(\mathbf{x}) = 0, g_2(\mathbf{x}) = 0$

制約の個数が多いとき，このような場合分けは計算量が大きい。

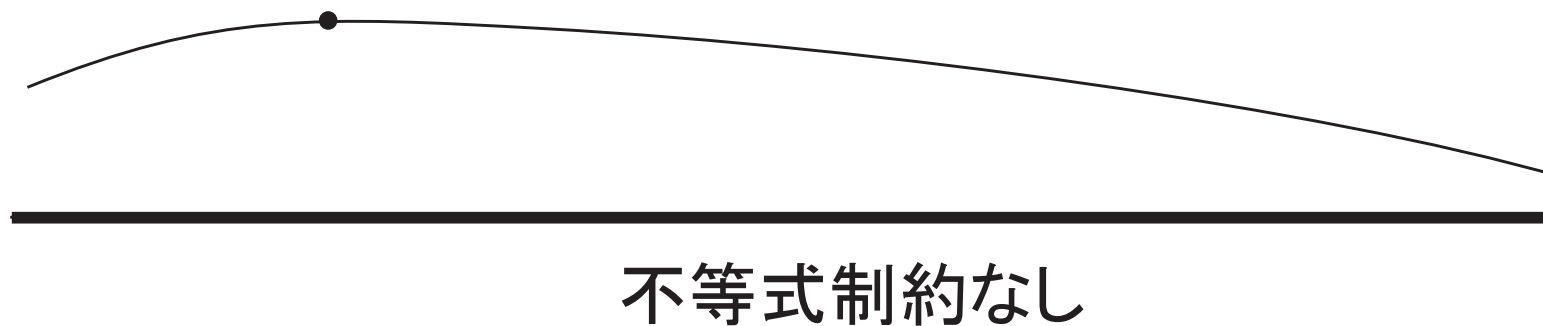
凸集合上での凸関数の最大化：



微分で(偶然)
解ける



微分で解けない



微分で解ける

本研究の目的

不等式制約のある凸最適化問題を等式制約の凸最適化問題を有限回解くことによって解く。

C は Kullback-Leibler 情報量 $D(P^i \| Q)$ によって次のように表すこともできる Csiszàr[5]。

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

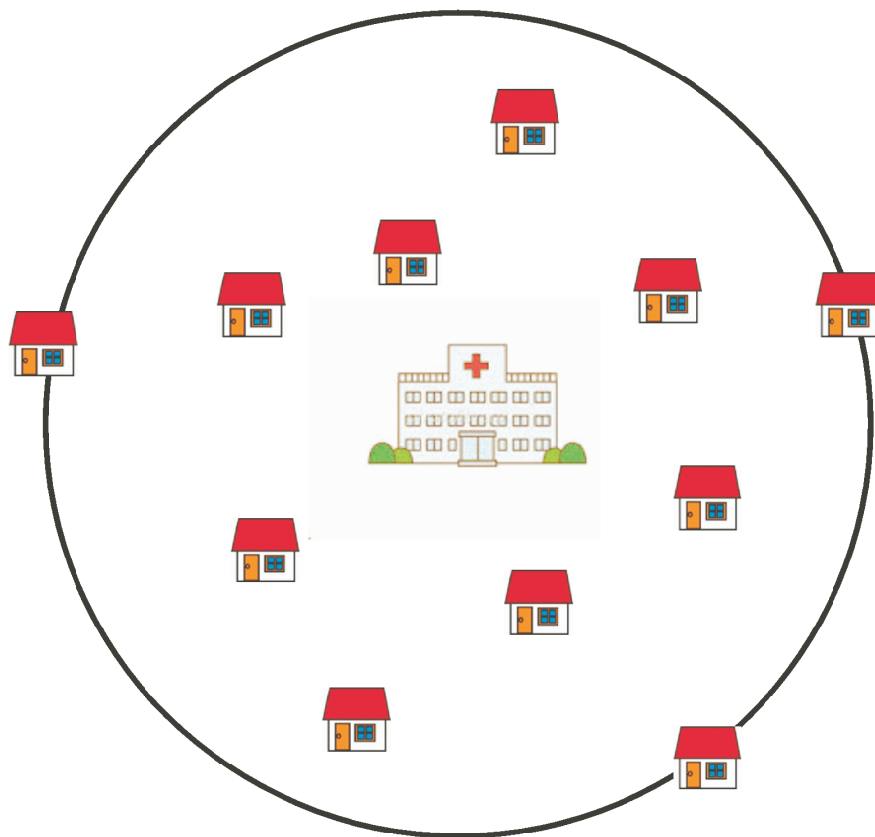
KL 情報量 D をユークリッド距離 d に置き換えると \mathbb{R}^n における類似の問題を考えることができる。

すなわち, $P^1, \dots, P^m \in \mathbb{R}^n$ に対して,

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

これは, 有限点集合 $\{P^1, \dots, P^m\}$ に対する「最小包含円」の問題である。

P^1, \dots, P^m を内部または周上に含む円のうちで半径が最小のものを最小包含円 (Smallest Enclosing Circle) といい,
 $\Gamma = \Gamma(P^1, \dots, P^m)$ と表す。



Γ は周上の点で決まる。内部の点は Γ の決定に寄与しない。
寄与しない点を除外するアルゴリズムについて考える。

最小包含円問題に関する凸最適化問題

$$P^1, \dots, P^m \in \mathbb{R}^n, \Phi = \begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix} \in \mathbb{R}^{m \times n}$$

$$\mathbf{a} = (\|P^1\|^2, \dots, \|P^m\|^2) \in \mathbb{R}^m, \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$$

$$f(\boldsymbol{\lambda}, \Phi) = \boldsymbol{\lambda}^t \mathbf{a} - \boldsymbol{\lambda} \Phi^t \Phi^t \boldsymbol{\lambda} : \mathbb{R}^m \rightarrow \mathbb{R} \text{ 上に凸}$$

(通信路容量における相互情報量に相当)

最小包含円問題は

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

と同値。

最小包含円問題に関する凸最適化問題

$$\max_{\lambda \in \bar{\Delta}^m} f(\lambda, \Phi) \quad (*)$$

に対する Kuhn-Tucker 条件 :

$\lambda^* = (\lambda_1^*, \dots, \lambda_m^*) \in \bar{\Delta}^m$ が $(*)$ を達成するための必要十分条件は, ある d_0 が存在して,

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

を満たすことである。このとき, $Q^* = \lambda^* \Phi$ が最小包含円 $\Gamma(P^1, \dots, P^m)$ の中心であり, $d^* = d_0$ がその半径に等しい。

ユークリッド空間における最小包含円問題を幾何的に考察し、
ユークリッド距離に関する min max 問題、すなわち

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

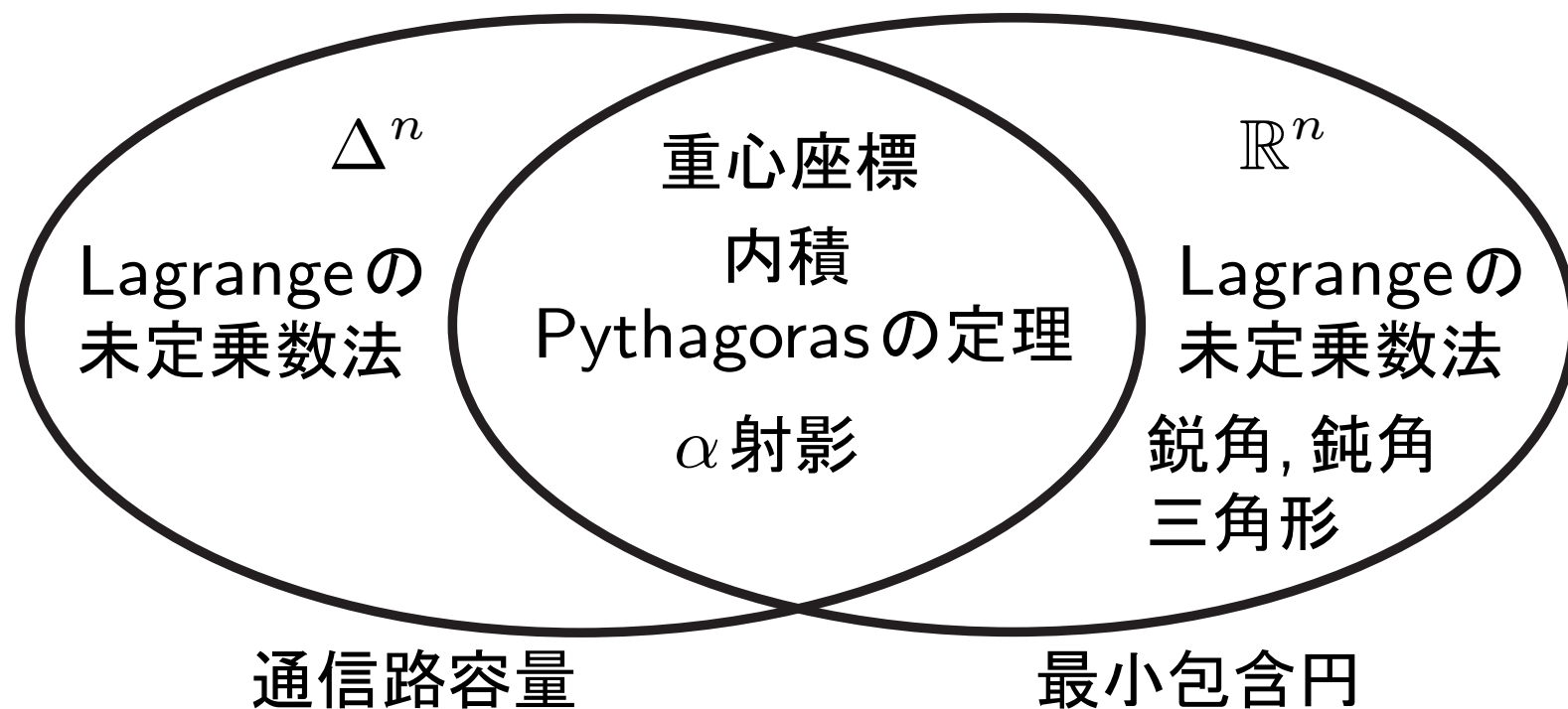
の解を計算するアルゴリズムを開発する。そして得られたアル
ゴリズムを通信路容量に関する min max 問題、すなわち

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

に適用して、通信路容量を達成する出力分布を計算するアルゴ
リズムを作る。このようなアプローチをとる理由は、ユーク
リッド幾何は我々にとってなじみがある幾何なので、新しい幾
何的なアルゴリズムを作りやすいからである。

$$D(P^i \| Q) \longleftrightarrow d(P^i, Q)$$

この置き換えは単なる形式的な置き換えではない。通信路容量の問題と最小包含円の問題は類似した幾何構造を持つ。それは Amari[1] の α 幾何である。そこで、両者に共通の性質だけを使って一方の問題でのアルゴリズムを作れば、それは自動的に他方に移植できる。



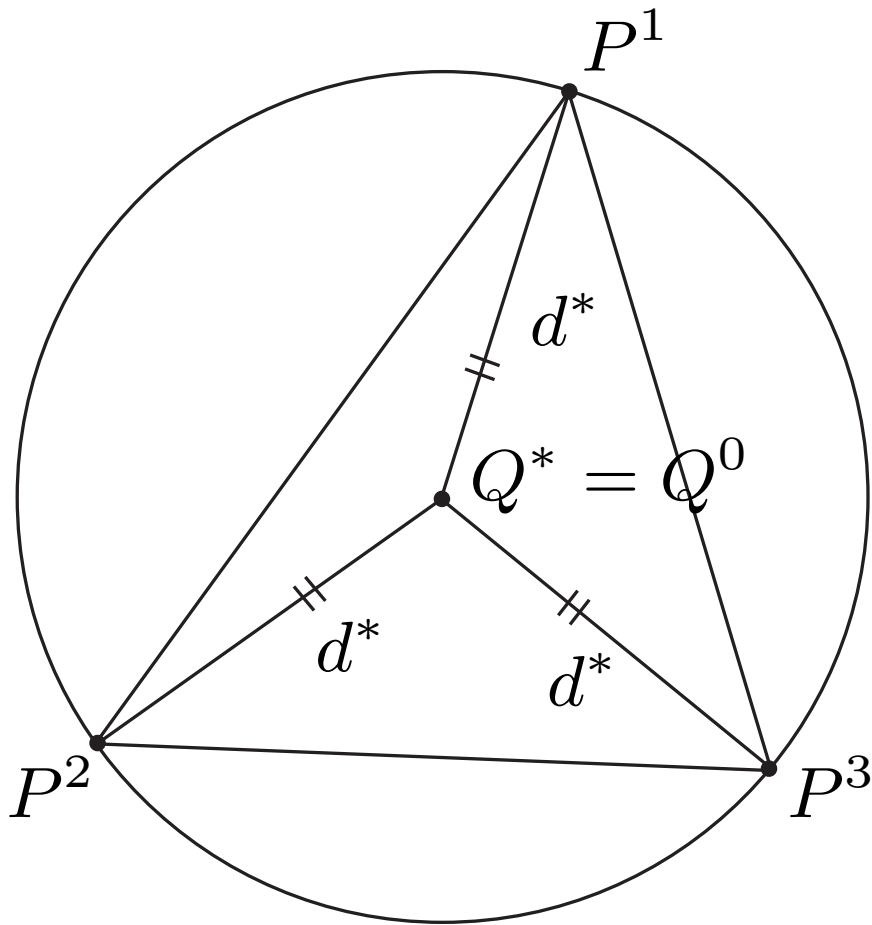
	通信路容量	最小包含円
λ	入力分布	重心座標
P^i	入力シンボル x_i に対する出力分布 $\in \Delta^n$	与えられた点 $\in \mathbb{R}^n$
Φ	$\begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix}$ 通信路行列	$\begin{pmatrix} P^1 \\ \vdots \\ P^m \end{pmatrix}$ 点を並べた行列
凸関数	$I(\lambda, \Phi)$ 相互情報量	$f(\lambda, \Phi) = \lambda^t a - \lambda \Phi^t \Phi^t \lambda$
尺度	$D(P \ Q)$ KL 情報量	$d(P, Q)$ ユークリッド距離
KT 条件	$D(P^i \ \lambda^* \Phi)$ $\begin{cases} = C_0, & \text{if } \lambda_i^* > 0 \\ \leq C_0, & \text{if } \lambda_i^* = 0 \end{cases}$	$d(P^i, \lambda^* \Phi)$ $\begin{cases} = d_0, & \text{if } \lambda_i^* > 0 \\ \leq d_0, & \text{if } \lambda_i^* = 0 \end{cases}$

Q^*	C を達成する出力分布	最小包含円の中心
内積	(P, Q) 通常の内積	(η, θ) 双対座標による内積
ピタゴラス	$D(P\ Q) + D(Q\ R)$ $= D(P\ R)$	$d^2(P, Q) + d^2(Q, R)$ $= d^2(P, R)$
射影	$\min_{P \in L} D(P\ Q)$	$\min_{P \in L} d(P, Q)$

ユークリッド空間における最小包含円問題

簡単な例： \mathbb{R}^2 における3点 P^1, P^2, P^3

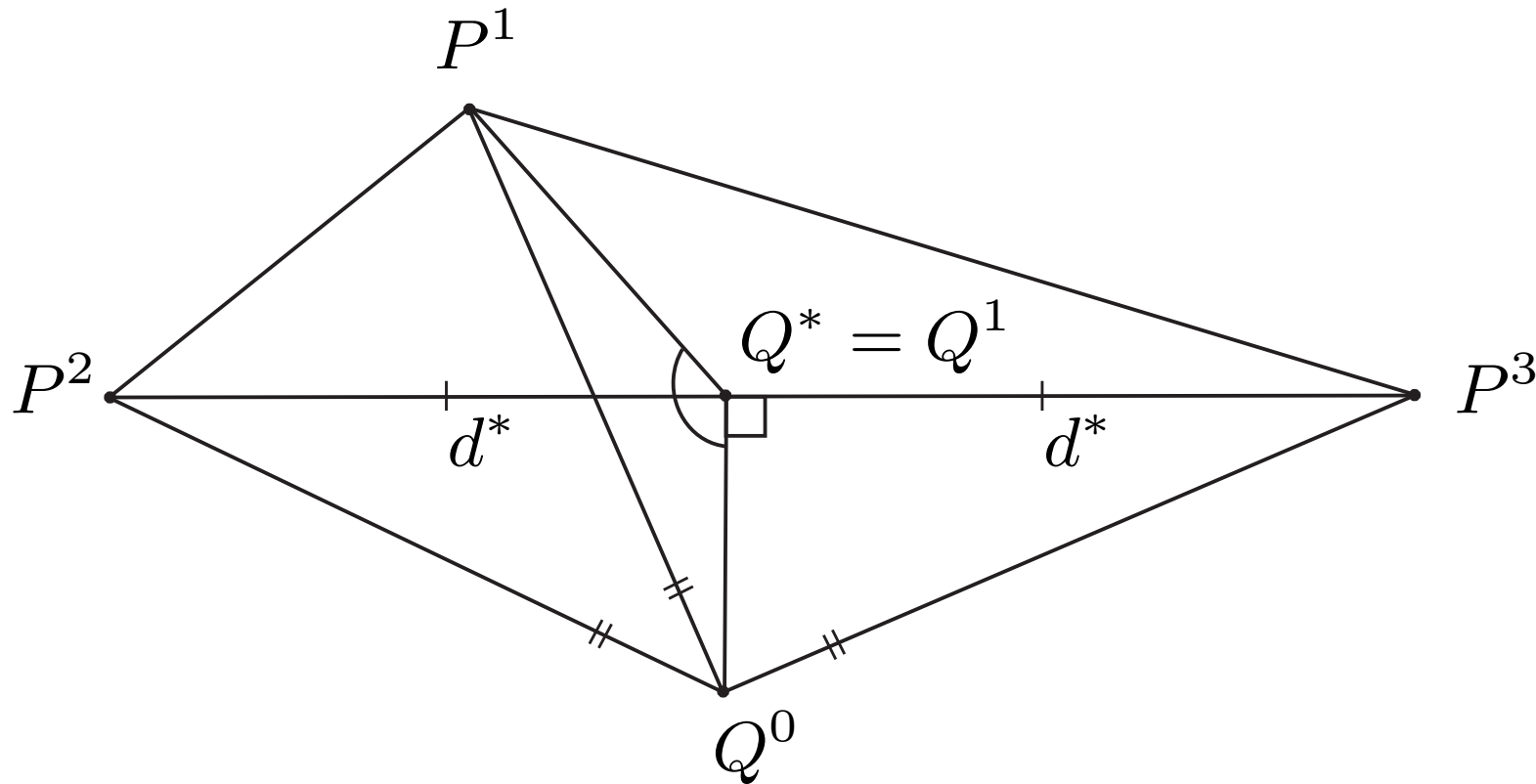
(I) $\triangle P^1 P^2 P^3$ が鋭角三角形の場合



最小包含円 Γ は $\triangle P^1 P^2 P^3$ の外接円である。 $\triangle P^1 P^2 P^3$ の外心（すなわち等距離点） Q^0 が Γ の中心 Q^* であり、 $d^* = d(P^1, Q^*)$ である。

この例から、与えられた点 P^1, \dots, P^m からの等距離点 Q^0 を考えることが有効であるといえる。

(II) $\triangle P^1 P^2 P^3$ が鈍角三角形の場合



最小包含円 Γ の中心 Q^* は辺 $P^2 P^3$ の中点 Q^1 である。等距離点 Q^0 は Γ の中心ではない。 Q^1 は Q^0 を直線 $P^2 P^3$ へ射影した点である。

この例から、等距離点 Q^0 が $\triangle P^1 P^2 P^3$ の外部にあるとき、射影を考えることが有効であるといえる。

定理

$P^1, \dots, P^m \in \mathbb{R}^n$: 一般の位置にある

Q^0 : P^1, \dots, P^m からの等距離点

V : P^1, \dots, P^m に対する convex hull

このとき,

$$\min_{Q \in V} d(Q, Q^0)$$

を達成する点 $Q = Q^*$ が最小包含円の中心である。

したがって、原理的には1回の最小化で最適点を得られるのだが、不等式制約条件のため計算ができない（計算量が多い）。

以下において，不等式制約がある最適化問題を等式制約の最適化問題を有限回解いて最適解を得る方法を考える。
等式制約だと微分で解けるので解が式で得られるから。

$L \subset \mathbb{R}^n$: アフライン部分空間（直線，平面， \dots ）
に対して

$$\min_{Q \in L} d(Q, Q^0)$$

は解ける。

P^1, \dots, P^m が一般の位置にある場合の最小包含円とその重心座標

記号の簡単化

$$P = \overrightarrow{OP}, \lambda_1 P^1 + \lambda_2 P^2 = \lambda_1 \overrightarrow{OP^1} + \lambda_2 \overrightarrow{OP^2},$$
$$P^2 - P^1 = \overrightarrow{P^1 P^2}$$

さらに、場合によって、

$$P \in \mathbb{R}^n \text{ (点)}$$

$$= (P_1, \dots, P_n) \text{ (座標)}$$

$$= \overrightarrow{OP} \text{ (ベクトル)}$$

$$= (P_1, \dots, P_n) \text{ (ベクトルの成分)}$$

 P^1, \dots, P^m が一般に位置にある, とは, ベクトルの集合 $\{P^2 - P^1, \dots, P^m - P^1\}$ が一次独立であること。

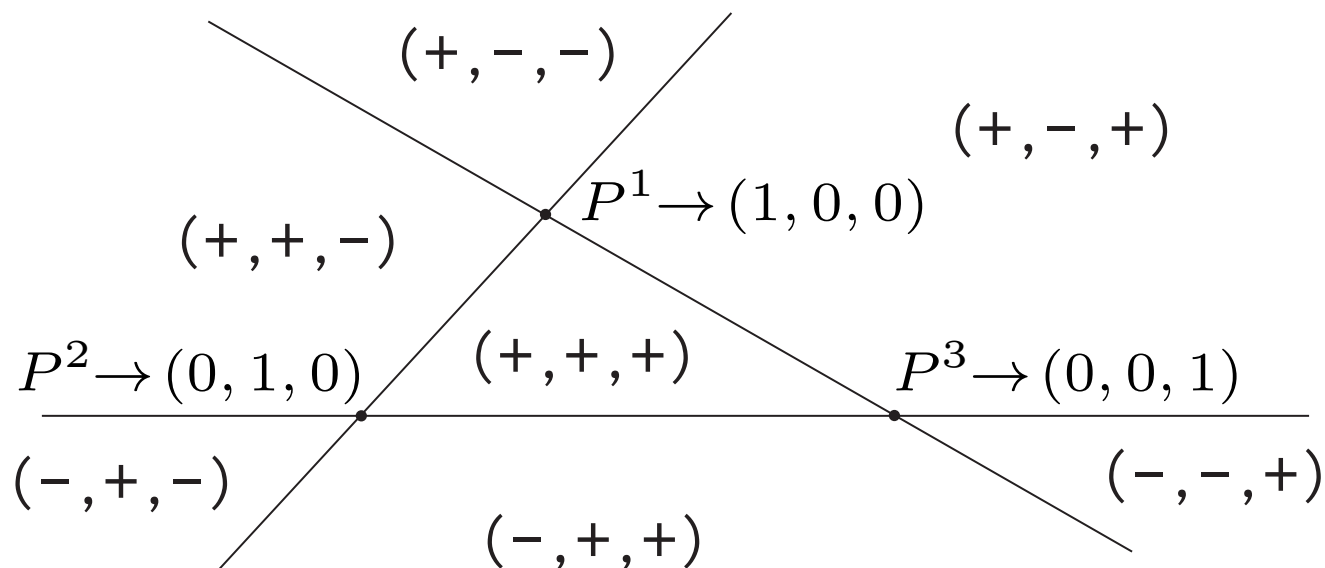
重心座標

$P^1, \dots, P^m \in \mathbb{R}^n$ が一般の位置にあるとする。

$L_0 = L(P^1, \dots, P^m) \subset \mathbb{R}^n$ を P^1, \dots, P^m で張られるアファイン部分空間とする。

$$Q = \lambda_1 P^1 + \dots + \lambda_m P^m, \quad \sum_{i=1}^m \lambda_i = 1$$

を満たす実数の組 $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$ を点 Q の P^1, \dots, P^m に関する重心座標という。



P^1, \dots, P^m からの等距離点 Q^0

$$d(P^1, Q^0) = \dots = d(P^m, Q^0),$$

$$Q^0 \in L_0 = L(P^1, \dots, P^m)$$

を満たす点 Q^0 を P^1, \dots, P^m からの等距離点という。

【定理】 P^1, \dots, P^m が一般の位置にあるとき, P^1, \dots, P^m からの等距離点 $Q^0 \in L_0$ が唯一つ存在し, その重心座標 λ^0 は

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

$$M \equiv \hat{\Phi} {}^t \hat{\Phi}, \hat{\mathbf{a}} = (\|\hat{P}^1\|^2, \dots, \|\hat{P}^m\|^2), \mathbf{1} = (1, \dots, 1) \in \mathbb{R}^m$$

となる (詳細略)。

\mathbb{R}^n における内積, Pythagorasの定理, 射影

内積 $Q^1, Q^2, Q^3 \in \mathbb{R}^n$ に対して, 内積を

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

と定義する。(これは, ベクトル $\overrightarrow{Q^2Q^1}$ と $\overrightarrow{Q^2Q^3}$ の内積である。)

補題1 $P^1, \dots, P^m, Q, R \in \mathbb{R}^n$ に対して $\sigma_i = (P^i - Q, R - Q)$, $i = 1, \dots, m$ とおく。 $\sum_{i=1}^m \lambda_i = 1$ のとき

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

補題3 $P, Q, R \in \mathbb{R}^n$ に対して

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

定理 (Pythagoras)

$$(P - Q, R - Q) \begin{matrix} \geq \\ \leq \end{matrix} 0 \iff d^2(P, Q) + d^2(Q, R) \begin{matrix} \geq \\ \leq \end{matrix} d^2(P, R)$$

射影

$Q^0 \in \mathbb{R}^n$, $L \subset \mathbb{R}^n$ に対して

$$\min_{Q \in L} d(Q, Q^0)$$

を達成する点 $Q = Q^1$ を Q^0 の L への射影といい,

$$Q^1 = \pi(Q^0 | L)$$

と表す。以下では, 部分集合 L として \mathbb{R}^n のアファイン部分空間だけを考える。

等距離点からの射影

$P^1, \dots, P^m \in \mathbb{R}^n$: 一般の位置にある

$L_0 = L(P^1, \dots, P^m)$: P^1, \dots, P^m で張られるアファイン
部分空間

$Q^0 \in L_0$: P^1, \dots, P^m からの等距離点

$L_k = L(P^{k+1}, \dots, P^m)$: P^{k+1}, \dots, P^m で張られるアファ
イン部分空間 $k = 0, \dots, m-2$

$$L_0 \supset L_1 \supset \dots \supset L_k \supset \dots$$

は次元が1ずつ小さくなるアファイン部分空間の減少列である。

Q^0 の L_1 への射影を $Q^1 = \pi(Q^0|L_1)$ とする。以下, 同様に
して, $Q^k = \pi(Q^{k-1}|L_k)$, $k = 1, \dots, m-2$ とする。

射影アルゴリズムによる Q^* の探索

例題1 $P^1 = (-10, 1, -3), P^2 = (-9, -2, 8),$
 $P^3 = (-8, 10, -5), P^4 = (4, -8, 8) \in \mathbb{R}^3$: 一般の位置にある
 P^1, \dots, P^4 からの等距離点 Q^0 の重心座標 λ^0 は

$$\lambda^0 = (-0.84, 0.04, 1.11, 0.69)$$

となる。 $\lambda_1^0 = -0.84 < 0$ に対する P^1 を除外して Q^0 を
 $L_1 = L(P^2, P^3, P^4)$ へ射影する : $Q^1 = \pi(Q^0|L_1)$

以下同様に, 負の重心座標成分を除外して射影する :

$$\lambda^1 = (0, -0.31, 0.63, 0.68)$$

$$Q^2 = \pi(Q^1|L_2), L_2 = L(P^3, P^4)$$

$$\lambda^2 = (0, 0, 0.5, 0.5) \Leftarrow \text{すべての成分が0以上}$$

このとき, $d(P^1, Q^2) = 9.18, d(P^2, Q^2) = 10.01, d(P^3, Q^2) =$
 $d(P^4, Q^2) = 12.62$ となり, Kuhn-Tucker 条件より Q^2 が最小
包含円の中心であることがわかる。

この例題のように毎回の射影で負の重心座標成分がちょうど1個あり、最後にすべてが非負になる場合を「状況1」という。すなわち、

$$\lambda^0 = (-, +, \dots, +)$$

$$\lambda^1 = (0, -, +, \dots, +)$$

$$\vdots$$

$$\lambda^{K-1} = (0, \dots, 0, -, +, \dots, +)$$

$$\lambda^K = (0, \dots, 0, +, \dots, +)$$

このとき、次の定理が成り立つ。

【定理】 状況1のとき、 P^1, \dots, P^m に対する最小包含円の中心は $Q^* = Q^K$ であり半径は $d^* = d(P^{K+1}, Q^K)$ である。

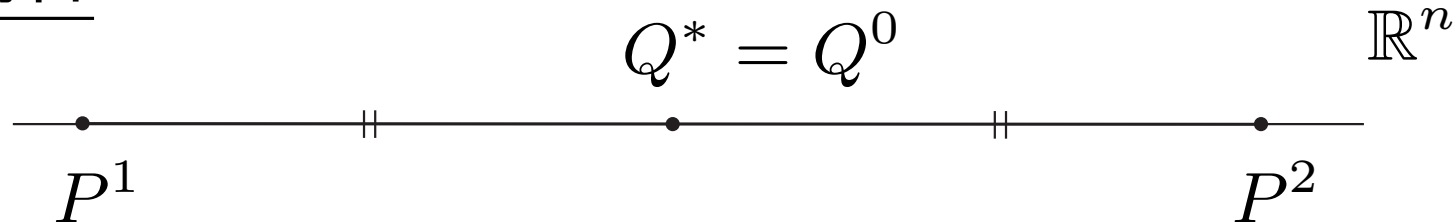
証明 内積, Pythagorasの定理, 射影だけで証明できる。

状況2 [$m = 2, 3, 4$, n は任意]

\mathbb{R}^n において一般の位置にある m 個の点 P^1, \dots, P^m に対する最小包含円 $\Gamma = \Gamma(P^1, \dots, P^m)$ の中心 Q^* と半径 d^* を射影アルゴリズムによって計算する。

目標はすべての m に対して計算するアルゴリズムを見い出すことであるが、現時点では $m = 2, 3, 4$ の場合だけ解決できている。

$m = 2$ の場合



【定理】 最小包含円の中心は $Q^* = Q^0$ であり、その半径は $d(P^1, Q^0)$ である。

$m = 3$ の場合（この場合だけ証明を詳細に紹介する）

$P^1, P^2, P^3 \in \mathbb{R}^n$: 一般の位置にある

$Q^0 \in L_0$: P^1, P^2, P^3 からの等距離点

$\lambda^0 = (\lambda_1^0, \lambda_2^0, \lambda_3^0)$: Q^0 の重心座標

λ^0 の各成分の符号に基づいて、一般性を失うことなく、以下の場合分けをすることができる。

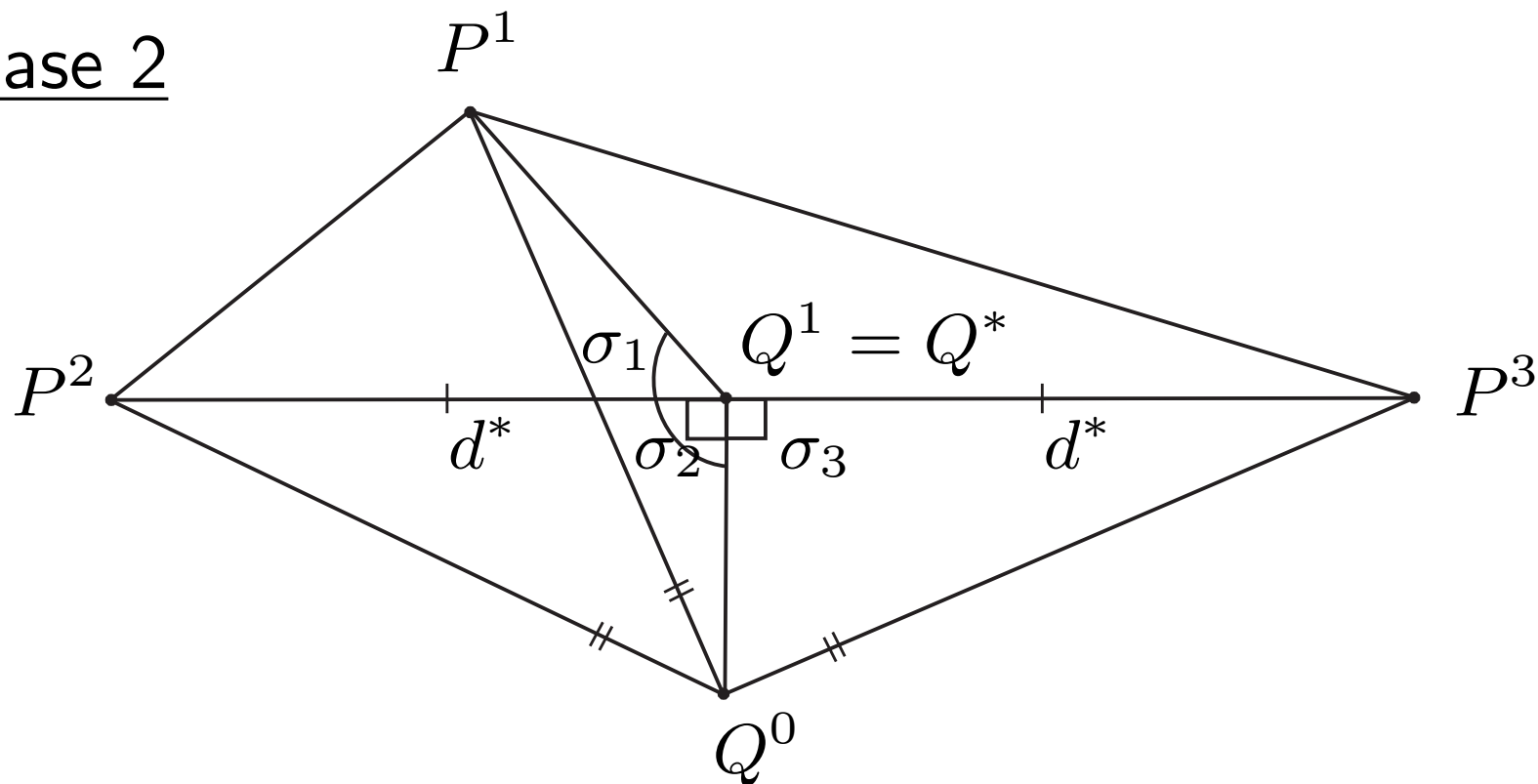
Case 1: $\lambda_1^0 \geq 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0$ （鋭角三角形）

Case 2: $\lambda_1^0 < 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0$ （鈍角三角形）

Case 1 この場合は易しい。Kuhn-Tucker 条件より

【定理】 最小包含円の中心は $Q^* = Q^0$ であり、その半径は $d(P^1, Q^0)$ である。

Case 2



$$L_1 = L(P^2, P^3), Q^1 = \pi(Q^0 | L_1)$$

【定理】 最小包含円の中心は $Q^* = Q^1$ であり, その半径は $d(P^2, Q^1)$ である。

証明 内積 $\sigma_i = (P^i - Q^1, Q^0 - Q^1)$, $i = 1, 2, 3$ を考える。
直交射影だから $\sigma_2 = \sigma_3 = 0$ である。 $\sigma_1 < 0$ を示す。

$$\begin{aligned}
\lambda_1^0 \sigma_1 &= \sum_{i=1}^3 \lambda_i^0 \sigma_i \\
&= \sum_{i=1}^3 \lambda_i^0 (P^i - Q^1, Q^0 - Q^1) \\
&= \left(\sum_{i=1}^3 \lambda_i^0 P^i - Q^1, Q^0 - Q^1 \right) \\
&= (Q^0 - Q^1, Q^0 - Q^1) > 0
\end{aligned}$$

Case 2の仮定 $\lambda_1^0 < 0$ より $\sigma_1 < 0$ となる。したがって、Pythagorasの定理とその不等式版（余弦定理）より

$$d(P^1, Q^1) < d(P^2, Q^1) = d(P^3, Q^1)$$

となる。Kuhn-Tucker条件から定理を得る。

$m = 4$ の場合

$P^1, P^2, P^3, P^4 \in \mathbb{R}^n$: 一般の位置にある

$Q^0 \in L_0$: P^1, P^2, P^3, P^4 からの等距離点

$\lambda^0 = (\lambda_1^0, \lambda_2^0, \lambda_3^0, \lambda_4^0)$: Q^0 の重心座標

λ^0 の各成分の符号に基づいて, 一般性を失うことなく, 以下の場合分けをすることができる。

Case 1: $\lambda_1^0 \geq 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$

Case 2: $\lambda_1^0 < 0, \lambda_2^0 \geq 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$

Case 3: $\lambda_1^0 < 0, \lambda_2^0 < 0, \lambda_3^0 \geq 0, \lambda_4^0 \geq 0$

各 Case において, 射影アルゴリズムで Q^* が得られ, $m = 3$ の場合と同じ証明法で証明できる。

Caseによる上記の分類をまとめると， 重心座標の符号と (Q^*, d^*) は下記のようなになる。

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

Case 2 $(- + + +)$

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

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

Case 3 $(- - + +)$

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

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

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

ここで，

$$Q^{1(1)} = \pi(Q^0 | L(P^2, P^3, P^4)), \quad Q^{2\dagger} = \pi(Q^0 | L(P^2, P^4))$$

通信路容量の問題

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

KL情報量 $D(P^i \| Q)$ の幾何的性質, すなわち, $\alpha = \pm 1$ として α 幾何的性質に基づいて上記の最小値を達成する出力分布 $Q^* \in \bar{\Delta}^n$ を探索する。

特に, ユークリッド幾何と共通な性質である重心座標, 内積, α -divergence, Pythagorasの定理, α -射影だけを使う。

全て正の確率からなる出力分布全体

$$\Delta^n = \{Q = (Q_1, \dots, Q_n) | Q_j > 0, j = 1, \dots, n, \sum_{j=1}^n Q_j = 1\}$$

Δ^n 上の幾何構造

双対座標系

η 座標: $\eta = (\eta_2, \dots, \eta_n)$, $\eta_j = Q_j$, $j = 2, \dots, n$,

θ 座標: $\theta = (\theta_2, \dots, \theta_n)$, $\theta_j = \log(Q_j/Q_1)$, $j = 2, \dots, n$

内積

$Q^1, Q^2, Q^3 \in \Delta^n$ に対して

$\eta^1 = (\eta_j^1)_{j=2, \dots, n}$, $\eta^2 = (\eta_j^2)_{j=2, \dots, n}$: Q^1, Q^2 の η 座標

$\theta^2 = (\theta_j^2)_{j=2, \dots, n}$, $\theta^3 = (\theta_j^3)_{j=2, \dots, n}$: Q^2, Q^3 の θ 座標

このとき, 内積が次のように定義される。

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

アファイン部分空間

$P^1, \dots, P^m \in \Delta^n$ で張られるアファイン部分空間

$$L(P^1, \dots, P^m) = \left\{ \sum_{i=1}^m \lambda_i P^i \mid \sum_{i=1}^m \lambda_i = 1 \right\} \cap \Delta^n$$

と定義される。

KL 情報量による射影

$Q^0 \in \Delta^n, L \subset \Delta^n$ に対して,

$$\min_{Q \in L} D(Q \| Q^0)$$

を達成する $Q = Q^1$ を Q^0 の L への射影といい,
 $Q^1 = \pi(Q^0 | L)$ と表す。

以下では, L としてアファイン部分空間だけを考える。

補題 19 P^i ($i = 1, \dots, m$), $Q, R \in \Delta^n$ に対して,
 $\sigma_i = (P^i - Q, R \ominus Q)$, $i = 1, \dots, m$ とする。 $\sum_{i=1}^m \lambda_i = 1$ の
 とき,

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

補題 20 任意の $P, Q, R \in \Delta^n$ に対して

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

定理 (Pythagoras)

$P, Q, R \in \Delta^n$ に対して

$$(P - Q, R \ominus Q) \begin{matrix} \geq \\ \leq \end{matrix} 0 \iff D(P \| Q) + D(Q \| R) \begin{matrix} \geq \\ \leq \end{matrix} D(P \| R)$$

通信路容量を達成する出力分布について

点 $P^1, \dots, P^m \in \bar{\Delta}^n$ が一般の位置にある, とは, ベクトルの集合 $\{P^2 - P^1, \dots, P^m - P^1\}$ が一次独立であること。

等距離点 (等KL情報量分布)

$$D(P^1 \| Q^0) = \dots = D(P^m \| Q^0),$$

$$Q^0 \in L_0 = L(P^1, \dots, P^m)$$

を満たす点 Q^0 を P^1, \dots, P^m からの等距離点という。

補題22 P^1, \dots, P^m が一般の位置にあるとき, P^1, \dots, P^m からの等距離点 Q^0 が唯一つ存在する。

$L_k = L(P^{k+1}, \dots, P^m)$, $k = 0, 1, \dots, m-2$ とおく。

$L_0 \supset L_1 \supset \dots \supset L_k \supset \dots$ は次元が1ずつ小さくなる Δ^n のアファイン部分空間の減少列である。

等距離点 Q^0 を $L_0, L_1, \dots, L_k, \dots$ に次々と射影して,
 $Q^k = \pi(Q^{k-1} | L_k)$, $k = 1, \dots, m-2$ とする。

$\lambda^k : Q^k$ の P^1, \dots, P^m に関する重心座標

状況1の仮定 $k = 0, 1, \dots, K-1$ のとき λ^k の成分のうち負のものがちょうど1個あり, $k = K$ のとき λ^K の成分はすべて非負であるとする。

【定理】 状況1のとき, 通信路容量を達成する出力分布は $Q^* = Q^K$ であり, 通信路容量は $C = D(P^{K+1} \| Q^K)$ である。

証明 最小包含円の場合と同様に証明できる。(なぜ同様に証明できるかというと, 最小包含円問題と通信路容量問題が α 幾何に共通な性質だけを使って証明できる問題だからである。)

状況2 [入力アルファベット数 $m = 2, 3, 4$, 出力アルファベット数 n は任意]

【定理】 状況2において通信路容量 C を達成する出力分布 Q^* は射影アルゴリズムによって探索できる。

証明 ユークリッド空間における最小包含円の中心 Q^* の探索と同様にできる。

任意の点配置に対する最適点の探索

いままで、与えられた点 P^1, \dots, P^m が \mathbb{R}^n または Δ^n において一般の位置にあることを仮定した。

しかし、点の個数が多いと、この仮定は成り立たない。そこで、これから必ずしも一般の位置にあるとは限らない任意の点配置について考える。

P^1, \dots, P^m が一般の位置にない（すなわち、 $\{P^2 - P^1, \dots, P^m - P^1\}$ が一次従属）

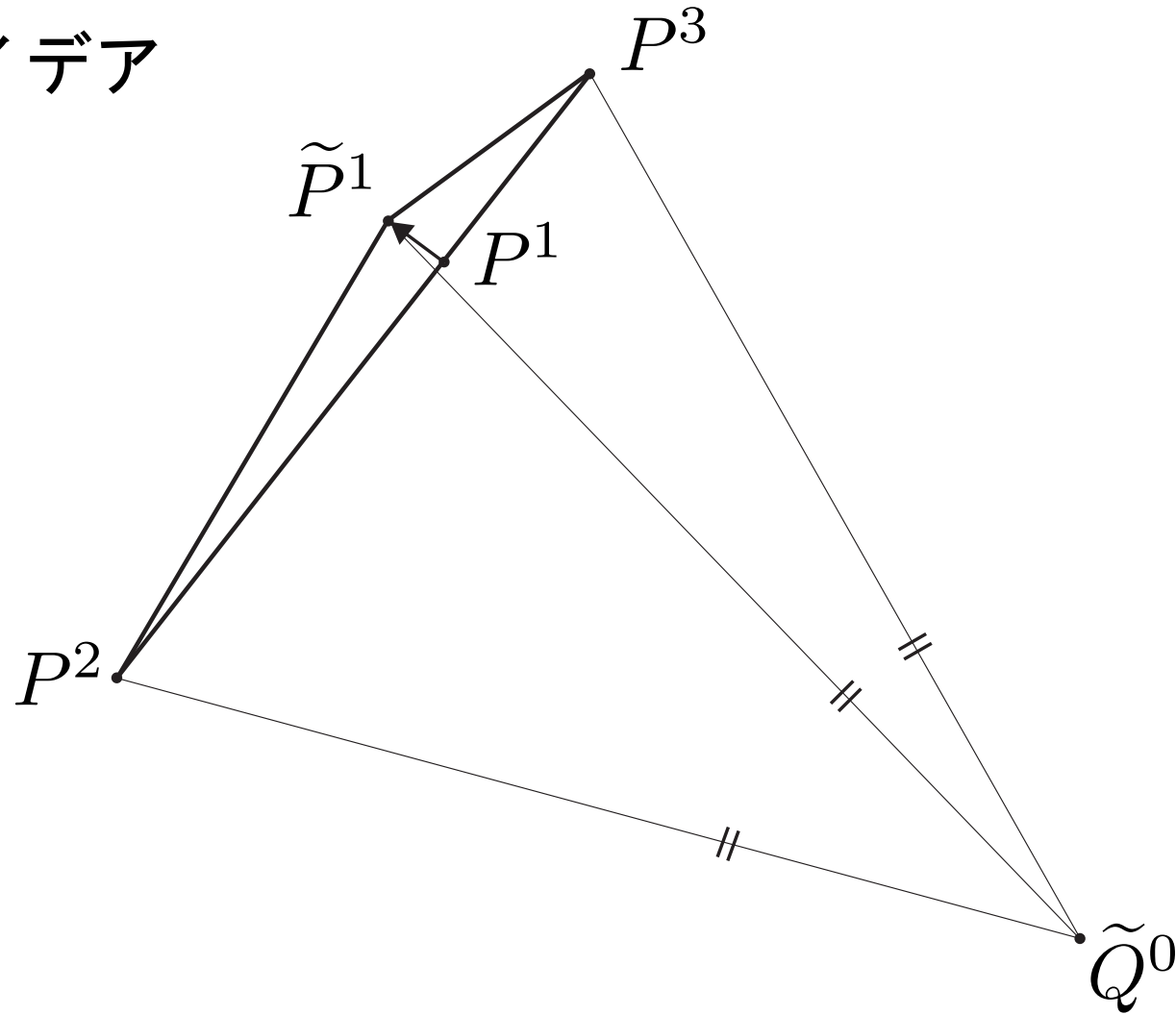
⇓

等距離点が存在しない

⇓

点 P^1, \dots, P^m の次元を上げて $\tilde{P}^1, \dots, \tilde{P}^m$ とし、 $\tilde{P}^1, \dots, \tilde{P}^m$ が一般の位置にあるようにする。

基本的なアイデア



P^1, P^2, P^3 が一直線上にあるとき等距離点は存在しない。
そこで, P^1 を少しずらして \tilde{P}^1 とすると, \tilde{P}^1, P^2, P^3 は
一般の位置にあり, 等距離点 \tilde{Q}^0 が存在する。

必ずしも一般の位置にあるとは限らない m 個の点
 $P^1, \dots, P^m \in \mathbb{R}^n$ の次元を上げる方法

$$P^1(P_1^1, \dots, P_n^1) \rightarrow \tilde{P}^1(P_1^1, \dots, P_n^1, \varepsilon, 0, \dots, 0)$$

$$P^2(P_1^2, \dots, P_n^2) \rightarrow \tilde{P}^2(P_1^2, \dots, P_n^2, 0, \varepsilon, \dots, 0)$$

$$\vdots \qquad \qquad \qquad \vdots$$

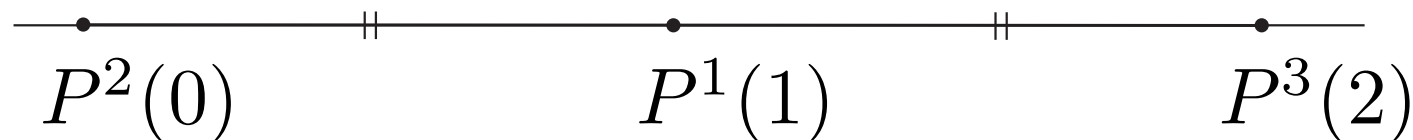
$$P^m(P_1^m, \dots, P_n^m) \rightarrow \tilde{P}^m(P_1^m, \dots, P_n^m, 0, \dots, 0, \varepsilon)$$

ただし, $\varepsilon \in \mathbb{R}, \varepsilon \neq 0$ で $|\varepsilon|$ は十分小さいとする。そこで,

$$\tilde{\Phi} = \begin{pmatrix} \tilde{P}^1 \\ \vdots \\ \tilde{P}^m \end{pmatrix} \in \mathbb{R}^{m \times (n+m)} \text{ とおくと } \text{rank } \tilde{\Phi} = m$$

したがって $\tilde{P}^1, \dots, \tilde{P}^m$ は一般の位置にあり, $\tilde{P}^1, \dots, \tilde{P}^m$ からの等距離点 \tilde{Q}^0 が存在する。

例題2



各点の次元を上げる

$$P^1(1) \rightarrow \tilde{P}^1(1, \varepsilon, 0, 0)$$

$$P^2(0) \rightarrow \tilde{P}^2(0, 0, \varepsilon, 0)$$

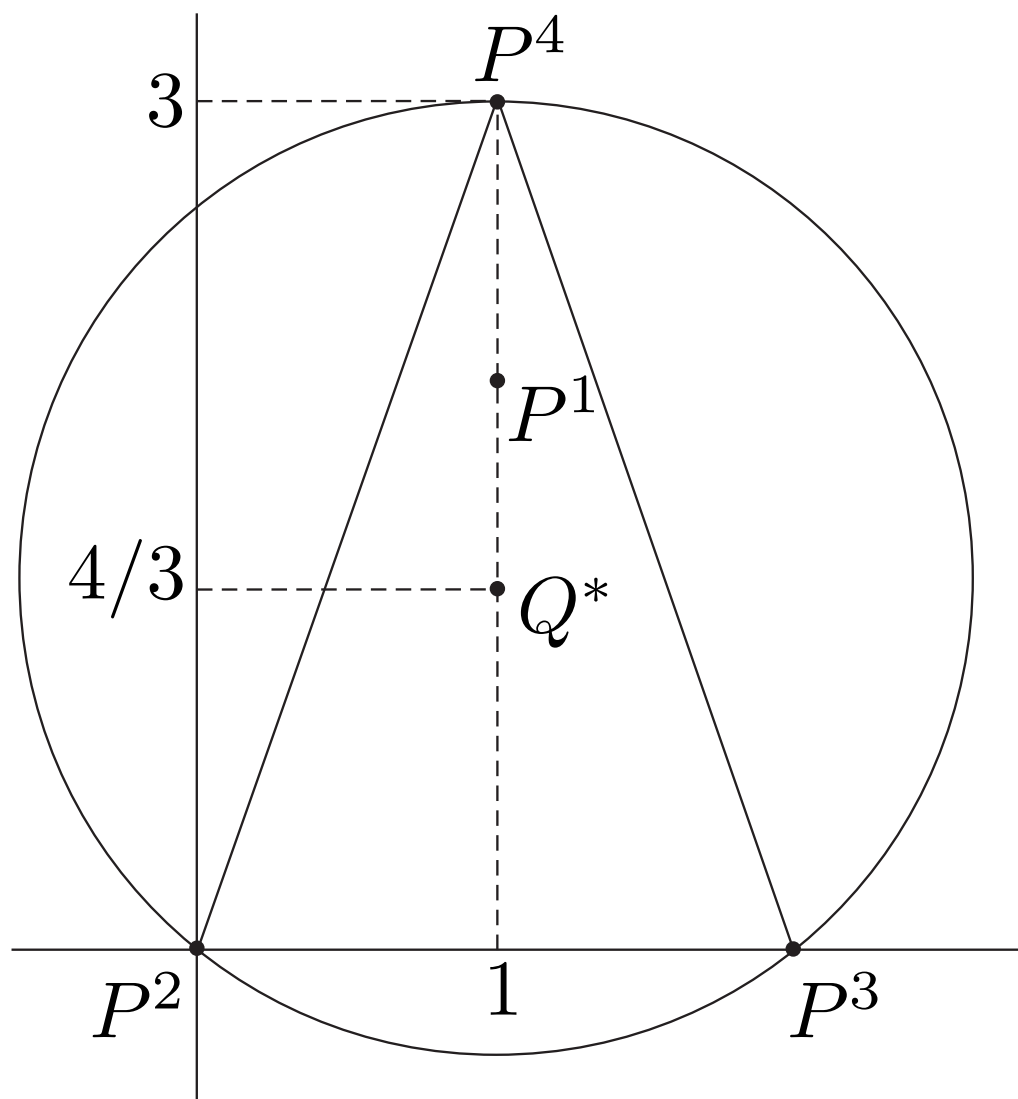
$$P^3(2) \rightarrow \tilde{P}^3(2, 0, 0, \varepsilon)$$

$\tilde{P}^1, \tilde{P}^2, \tilde{P}^3$ からの等距離点 \tilde{Q}^0 の重心座標 $\tilde{\lambda}^0$

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

$|\varepsilon|$ が十分小さいとき $\tilde{\lambda}_1^0 = (-2 + 2\varepsilon^2)/(6\varepsilon^2) < 0$ となるので、 \tilde{P}^1 を除外する。そして、射影 $\tilde{Q}^1 = \pi(\tilde{Q}^0 | L(\tilde{P}^2, \tilde{P}^3))$ の重心座標 $\tilde{\lambda}^1 = (0, 1/2, 1/2)$ を得る。 $Q^* = \tilde{\lambda}^*|_{\varepsilon=0} \Phi = 1$ となり、正しく最小包含円の中心 Q^* が得られる。

例題3 \mathbb{R}^2 の4点 $P^1(1, 2), P^2(0, 0), P^3(2, 0), P^4(1, 3)$



次元を上げる

$$P^1(1, 2) \rightarrow \tilde{P}^1(1, 2, \varepsilon, 0, 0, 0)$$

$$P^2(0, 0) \rightarrow \tilde{P}^2(0, 0, 0, \varepsilon, 0, 0)$$

$$P^3(2, 0) \rightarrow \tilde{P}^3(2, 0, 0, 0, \varepsilon, 0)$$

$$P^4(1, 3) \rightarrow \tilde{P}^4(1, 3, 0, 0, 0, \varepsilon)$$

等距離点 \tilde{Q}_0 の重心座標 $\tilde{\lambda}^0$

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

$\tilde{\lambda}_1^0 < 0$ より \tilde{P}^1 を除外して $\tilde{Q}^1 = \pi(\tilde{Q}^0 | L(\tilde{P}^2, \tilde{P}^3, \tilde{P}^4))$ とする。 \tilde{Q}^1 の重心座標 $\tilde{\lambda}^1$

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

$\tilde{\lambda}_i^1 \geq 0, i = 1, \dots, 4$ なので, Kuhn-Tucker条件より $\tilde{\lambda}^1 = \tilde{\lambda}^*$

ここで $\varepsilon = 0$ として, $\tilde{\lambda}^*|_{\varepsilon=0} = (0, 5/18, 5/18, 8/18)$

$$Q^* = \tilde{\lambda}^*|_{\varepsilon=0} \Phi = (0, 5/18, 5/18, 8/18) \Phi = (1, 4/3)$$

となり正しく最小包含円 $\Gamma(P^1, \dots, P^4)$ の中心 Q^* が得られる。

通信路行列の次元を上げる

通信路行列の行ベクトル P^1, \dots, P^m に対してそれらの次元を上げた行ベクトル $\tilde{P}^1, \dots, \tilde{P}^m$ を以下のように定義する。

$$\tilde{P}^i = \left(\frac{P_1^i}{2m+1}, \dots, \frac{P_n^i}{2m+1}, \overbrace{\frac{1}{2m+1}, \dots, \frac{1}{2m+1}}^{2i-2 \text{ 個}}, \right. \\ \left. \frac{\overset{2i-1 \text{ 番目}}{\vee} 1 + \epsilon}{2m+1}, \frac{\overset{2i \text{ 番目}}{\vee} 1 - \epsilon}{2m+1}, \overbrace{\frac{1}{2m+1}, \dots, \frac{1}{2m+1}}^{2m-2i \text{ 個}} \right)$$

ここで, ϵ は $|\epsilon|$ が十分小さい実数で $\epsilon \neq 0$ とする。

次元を上げた通信路行列は

$$\tilde{\Phi} = \begin{pmatrix} \tilde{P}^1 \\ \vdots \\ \tilde{P}^m \end{pmatrix}$$

となる。 $\varepsilon \neq 0$ なので, $\text{rank } \tilde{\Phi} = m$ である。したがって, $\tilde{P}^1, \dots, \tilde{P}^m$ からの等距離点 \tilde{Q}^0 が存在する。

この通信路モデルは, 入力アルファベット $\{x_1, \dots, x_m\}$ は元のままとし, 出力アルファベット $\{y_1, \dots, y_n\}$ に新たにダミーのアルファベット $\{y'_1, \dots, y'_{2m}\}$ を追加したものである。入力アルファベットとダミーアルファベットとの相互情報量はほぼ0である。

$\tilde{P}^1, \dots, \tilde{P}^m$ からの等距離点 \tilde{Q}^0 の重心座標
 $\tilde{\lambda} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_m)$ の方程式：

$$\sum_{j=1}^n (P_j^i - P_j^1) \log \sum_{i'=1}^m \tilde{\lambda}_{i'} P_j^{i'} + \epsilon \log \frac{1 + \epsilon \tilde{\lambda}_i}{1 - \epsilon \tilde{\lambda}_i}$$

$$-\epsilon \log \frac{1 + \epsilon \tilde{\lambda}_1}{1 - \epsilon \tilde{\lambda}_1} = -H(P^i) + H(P^1), \quad i = 2, \dots, m$$

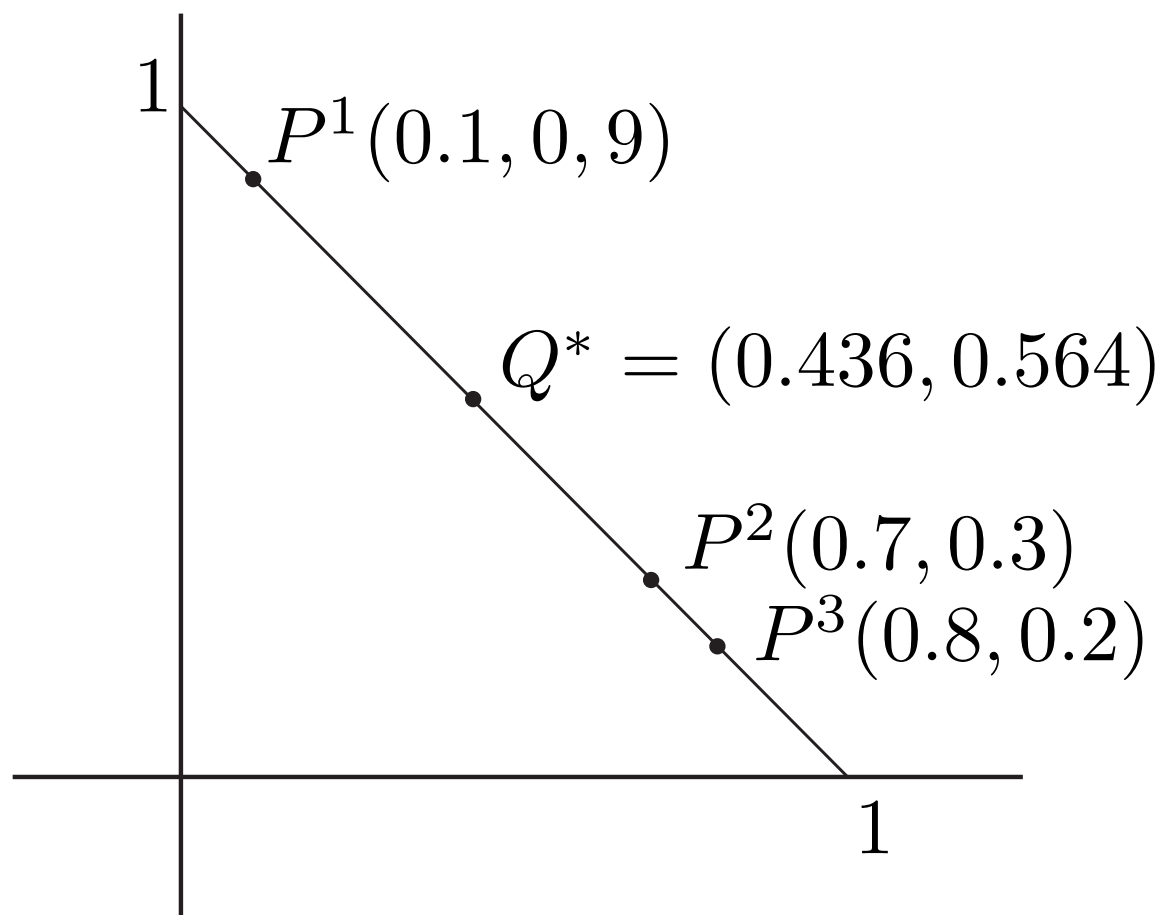
$H(P^i)$ は P^i のエントロピー

解 $\tilde{\lambda}$ の存在と一意性は補題24で証明されている。

例題 4

$$\Phi = \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix}$$

$$= \begin{pmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \\ 0.8 & 0.2 \end{pmatrix}$$



各 P^i の次元を上げる :

$$P^1(0.1, 0.9) \rightarrow \tilde{P}^1(0.1, 0.9, 1 + \varepsilon, 1 - \varepsilon, 1, 1, 1, 1)/7$$

$$P^2(0.7, 0.3) \rightarrow \tilde{P}^2(0.7, 0.3, 1, 1, 1 + \varepsilon, 1 - \varepsilon, 1, 1)/7$$

$$P^3(0.8, 0.2) \rightarrow \tilde{P}^3(0.8, 0.2, 1, 1, 1, 1, 1 + \varepsilon, 1 - \varepsilon)/7$$

$\varepsilon = 0.05$ として, 等距離点 \tilde{Q}^0 の重心座標 $\tilde{\lambda}^0$

$$\tilde{\lambda}^0 = (2.39, -12.96, 11.57)$$

$\tilde{\lambda}_2^0 = -12.96 < 0$ より \tilde{P}^2 を除外して $\tilde{Q}^1 = \pi(\tilde{Q}^0 | L(\tilde{P}^1, \tilde{P}^3))$

\tilde{Q}^1 の重心座標 $\tilde{\lambda}^1$

$$\tilde{\lambda}^* = \tilde{\lambda}^1 = (0.52, 0, 0.48)$$

となる。そして,

$$Q^* = \tilde{\lambda}^* \Big|_{\varepsilon=0} \Phi = (0.52, 0, 0.48) \Phi = (0.436, 0.564)$$

となり, 正しい解 Q^* を得る。通信路容量は

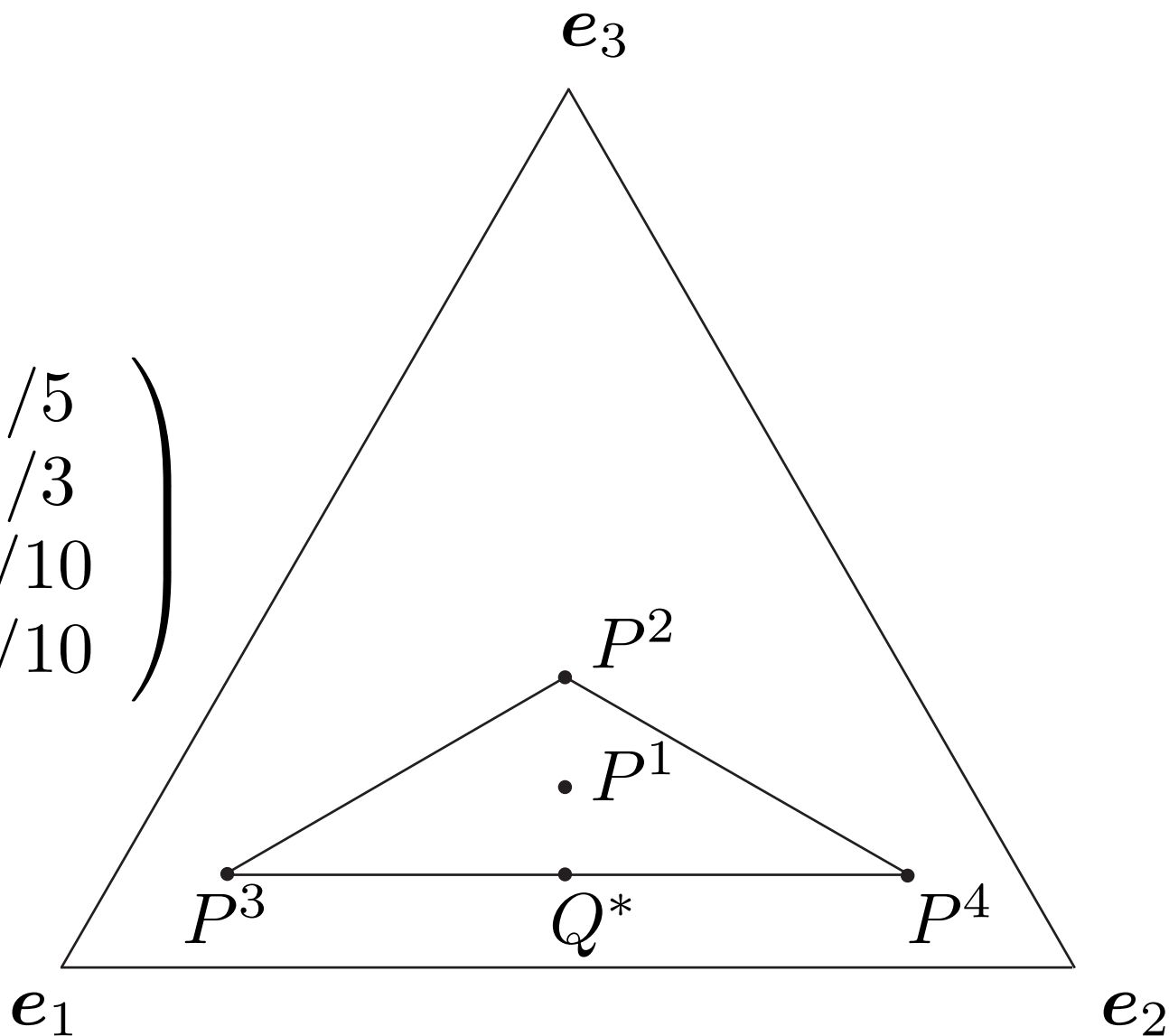
$$C = D(P^1 \| Q^*) = 0.398[\text{bit/symbol}]$$

となる。

例題5 通信路行列

$$\Phi = \begin{pmatrix} P^1 \\ P^2 \\ P^3 \\ P^4 \end{pmatrix}$$

$$= \begin{pmatrix} 2/5 & 2/5 & 1/5 \\ 1/3 & 1/3 & 1/3 \\ 4/5 & 1/10 & 1/10 \\ 1/10 & 4/5 & 1/10 \end{pmatrix}$$



各点の次元を上げる

$$P^1 \rightarrow \tilde{P}^1(2/5, 2/5, 1/5, 1 + \varepsilon, 1 - \varepsilon, 1, 1, 1, 1, 1, 1)/9$$

$$P^2 \rightarrow \tilde{P}^2(1/3, 1/3, 1/3, 1, 1, 1 + \varepsilon, 1 - \varepsilon, 1, 1, 1, 1)/9$$

$$P^3 \rightarrow \tilde{P}^3(4/5, 1/10, 1/10, 1, 1, 1, 1, 1 + \varepsilon, 1 - \varepsilon, 1, 1)/9$$

$$P^4 \rightarrow \tilde{P}^4(1/10, 4/5, 1/10, 1, 1, 1, 1, 1, 1 + \varepsilon, 1 - \varepsilon)/9$$

$\varepsilon = 0.05$ として $\tilde{P}^1, \dots, \tilde{P}^4$ からの等距離点 \tilde{Q}^0 の重心座標 $\tilde{\lambda}^0$ は

$$\tilde{\lambda}^0 = (-19.00, 7.98, 6.01, 6.01)$$

負の重心座標に対する点を除外して射影することを繰り返して

$$\tilde{\lambda}^1 = (0, -0.14, 0.57, 0.57)$$

$$\tilde{\lambda}^2 = (0, 0, 1/2, 1/2) = \tilde{\lambda}^*$$

となる。そして、

$$Q^* = \tilde{\lambda}^*|_{\varepsilon=0} \Phi = (0, 0, 1/2, 1/2) \Phi = (9/20, 9/20, 1/10)$$

を得る。通信路容量は $C = D(P^3 \| Q^*) = 0.447[\text{bit/symbol}]$

最小包含円を計算するヒューリスティックな アルゴリズム

任意の m, n および任意の点配置に対して適用できるヒューリスティックな射影アルゴリズムを提案する。(ただし, 常に正しい解が得られるとは限らない。)

最小包含円を計算するヒューリスティックなアルゴリズム

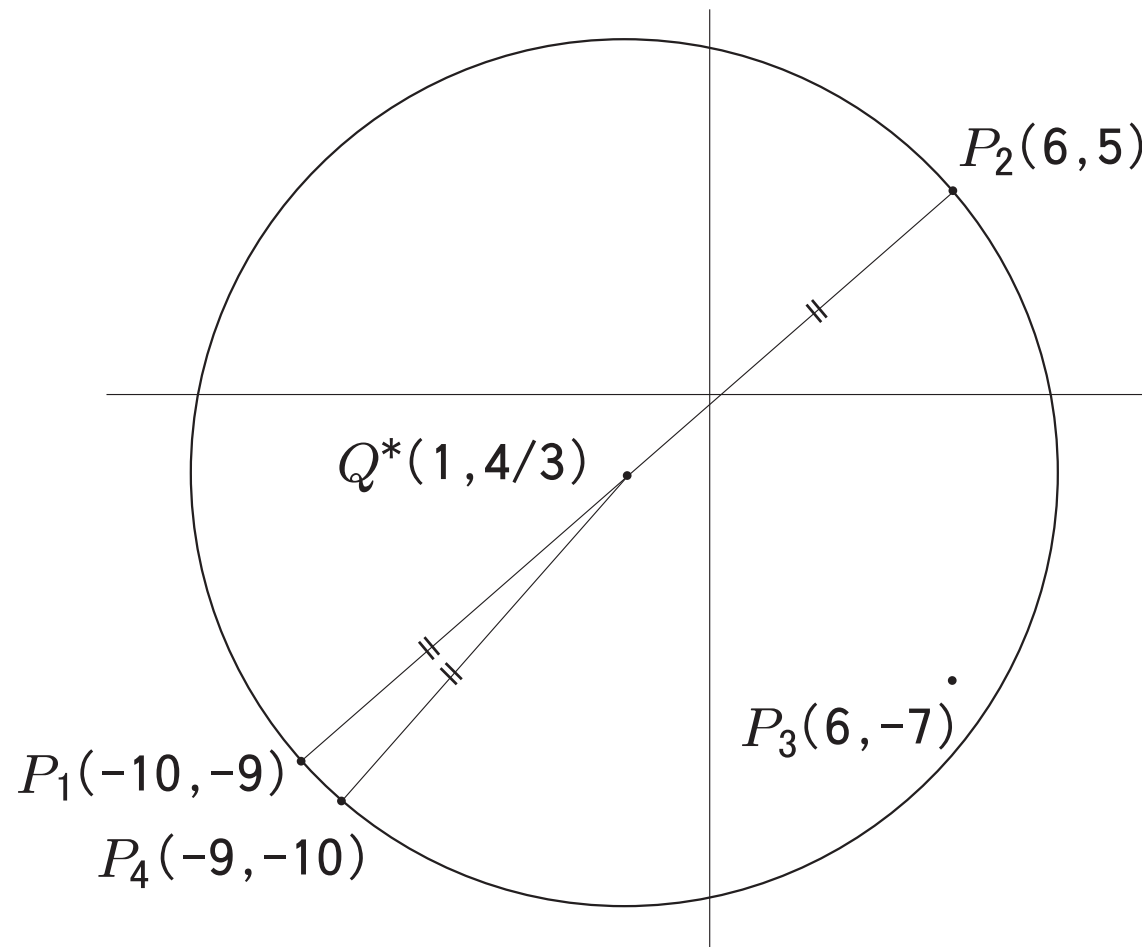
- (HA1) 点 P^1, \dots, P^m に対して, それらの次元を上げた点を $\tilde{P}^1, \dots, \tilde{P}^m$ とする。
- (HA2) $\tilde{P}^1, \dots, \tilde{P}^m$ からの等距離点 \tilde{Q}^0 の重心座標 $\tilde{\lambda}^0 = (\tilde{\lambda}_i^0)_{i=1, \dots, m}$ を計算する。
- (HA3) $\tilde{\lambda}_i^0 \geq 0, i = 1, \dots, m$ ならばアルゴリズムは終了し \tilde{Q}^0 を出力する。
- (HA4) $\tilde{\lambda}_i^0, i = 1, \dots, m$ に中に負のものがあれば, 最小のものに対応する点, 例えば \tilde{P}^1 とする, を除外して $\tilde{P}^2, \dots, \tilde{P}^m$ を残す。
- (HA5) $m - 1$ 個の点 $\tilde{P}^2, \dots, \tilde{P}^m$ に対して (HA1) から繰り返す。

シミュレーション結果

m, n のそれぞれの組に対してランダムに発生させた10000通りの最小包含円問題のうちで正しく Q^* が得られた割合を示す。

$m \backslash n$	2	3	10	20
3	100%	100%	100%	100%
4	98%	99%	100%	100%
5	96%	95%	100%	100%
8	94%	85%	99%	100%
10	93%	83%	98%	100%

ヒューリスティックアルゴリズムではうまくいかない例



$\lambda^0 =$

$$\left(\frac{-64800 + 22536\varepsilon^2 + 357\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6}, \frac{6480 + 43350\varepsilon^2 + 523\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6}, \right. \\ \left. \frac{-10800 - 13278\varepsilon^2 + 319\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6}, \frac{69120 + 17880\varepsilon^2 + 343\varepsilon^4 + \varepsilon^6}{70488\varepsilon^2 + 1542\varepsilon^4 + 4\varepsilon^6} \right)$$

計算量の比較

総当たり法の計算量

例えば, \mathbb{R}^2 における4点の場合, \mathbb{R}^2 において円は2点または3点で決定されるので, 総当たり法では ${}_4C_2 + {}_4C_3 = 10$ 通り。
一般に, \mathbb{R}^n における m 個の点の場合の計算量は

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

提案法の計算量

提案のヒューリスティックアルゴリズムでは毎回必ず点が1つずつ減っていくので, 計算量は

$$N_2 = m - 2$$

である。

総当たり法の計算量 N_1 に対する提案法の計算量 N_2 の比
 N_2/N_1

$m \backslash n$	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

今後の課題

1. 本論文では点の個数が $m = 2, 3, 4$ の場合にのみ最適点 Q^* を計算するアルゴリズムを得たが、これを任意の m に対して拡張する。
2. 本論文で利用した最小包含円問題と通信路容量問題の幾何学的な類似性を、レートひずみ関数の最適点を計算するアルゴリズムを開発する。
3. 通信路容量に関する有本アルゴリズムを、逆に最小包含円問題に移植して、最適点を探索するアルゴリズムを得る。

参考文献

Nakagawa, K., Watabe, K. and Sabu, T., “On the Search Algorithm for the Output Distribution that Achieves the Channel Capacity,” arXiv:1601.01394v2.