

非線形連立方程式の求解による
3次元発話方向推定問題の解法
Solving 3D Speech Direction Estimation Problem
via System of Nonlinear Equations

筑波大学 人間総合科学研究群 清水 啓^{*1}
KEI SHIMIZU
DEGREE PROGRAMS IN COMPREHENSIVE HUMAN SCIENCES
UNIVERSITY OF TSUKUBA

筑波大学 数理物質系 照井 章^{*2}
AKIRA TERUI
INSTITUTE OF PURE AND APPLIED SCIENCES
UNIVERSITY OF TSUKUBA

筑波大学 図書館情報メディア系 三河 正彦^{*3}
MASAHIKO MIKAWA
FACULTY OF LIBRARY, INFORMATION AND MEDIA SCIENCE
UNIVERSITY OF TSUKUBA

Abstract

This paper proposes a method for solving the 3D speech direction estimation problem using a system of nonlinear equations for individual selection in a multi-robot system. There are methods for solving the speech direction estimation problem: machine learning and genetic algorithms. However, machine learning requires training data, and genetic algorithms may fail to solve the problem correctly due to phenomenon such as initial convergence or hitchhiking.

Therefore, we propose a method for solving the 3D speech direction estimation problem by solving a system of nonlinear equations. We use trust-region method for solving the system of nonlinear equations. Furthermore, we show result of experiments with our implementation.

1 はじめに

一般に、ある系において既知パラメータの値から推定したい未知パラメータの値を求める問題を推定問題という [4]。現実的な推定問題の一つとして、複数台のロボットが協働してタスクを行うマルチロボットシステム [16] と呼ばれる環境において、命令を行う主体が複数台のロボットから 1 台を選択する問題がある。これを個体選択と呼ぶ。個体選択の手法には主に画像処理と音声処理が用いられる。画像処理を用い

*1 〒 305-8571 茨城県つくば市天王台 1-1-1 E-mail: kei.3718.kh@gmail.com

*2 〒 305-8571 茨城県つくば市天王台 1-1-1 E-mail: terui@math.tsukuba.ac.jp

*3 〒 305-8550 茨城県つくば市春日 1-2 E-mail: mikawa@slis.tsukuba.ac.jp

る手法として、注視 [7] や指差し (ポインティング・ジェスチャー) [2] といった動作による選択手法が提案されている。しかし、人間と複数台のロボットの位置を捉えるためのセンサーの設置が必要であることや、センサーが信号を受信できる範囲が制限されていることによって、ロボットと人間の位置関係に制限があることなどの問題点がある。これらの問題に対処するために、音声処理を用いる手法が有効である。音声処理を用いる手法では、人間が複数台のロボットのうちタスクを行わせるロボットの方向を向いて話しかけることで1台を選択する。これを実現するためには、話者がどの方向を向いているかをロボット自身が認識する必要がある。この問題を発話方向推定問題と呼ぶ。

発話方向推定問題の解法として、機械学習 [1, 18] を用いた選択手法が提案されているが、大量の音声データが必要であることや、学習データと異なる環境では精度が低下するといった問題点がある。一方で、HOE (Head Orientation Estimation system) と呼ばれるモデルベースの手法も提案されている [13, 19] が、環境が変わるごとにパラメータのチューニングが必要になるという問題点がある。これらの問題点を克服するために、加藤 [20] は学習データやパラメータのチューニングの必要がない手法として、独自に構築した音声拡散モデルに対して、遺伝的アルゴリズム [12] を用いた最適化手法を提案した。しかし、遺伝的アルゴリズムでは、初期収束あるいはヒッチハイクと呼ばれる現象により問題を正しく解けない場合がある。これに対して、Kuramochi et al. [11] は Minimax 近似を用いた手法を提案した。

以上の手法は、すべて2次元における発話方向推定によるロボット選択手法である。すなわち、音源 (発話者の口) と複数台のロボットがすべて同一平面上にあるという状況を仮定している。しかし、実環境において、音源と複数台のロボットがすべて同一平面上にあるとは限らない。3次元空間における発話方向推定では、画像処理と音声処理を組み合わせた手法 [15] が主に用いられている。しかし、画像処理を用いる手法には上記のような問題点があるため、音声処理のみを用いる手法の方が望ましいと考えられる。音声処理のみによる手法 [9] では、音響に関する複数の特徴量に対して線形モデルや人工ニューラルネットワークを用いた推定手法が提案されているが、話者の位置や発話方向の制限、精度の向上、学習データの必要性が課題として挙げられる。

以上のことから、本研究では音声処理のみを用いたより精度の良い3次元発話方向推定問題の解法として、非線形連立方程式の求解による手法を提案する。本手法では、ロボットに搭載されたマイクロフォン・アレイにより計測される観測データ (ワールド座標系におけるロボットの位置と姿勢、声の到来方向、音圧レベル) の値を用いて非線形連立方程式を解く。これにより、音声処理のみを用いた発話方向推定が可能となる。また、本手法は精度や計算時間においても実用上、有効であることが示された。

2 準備

2.1 信頼領域法を用いた非線形連立方程式の解法

本節では、信頼領域法およびその応用としての非線形連立方程式の解法について概説する [21]。本節で用いる記号を以下のように定義する。

- $\boldsymbol{x}, \boldsymbol{s} \in \mathbb{R}^n$: n 次元実ベクトル
- \boldsymbol{x}^T : \boldsymbol{x} の転置ベクトル
- f : \mathbb{R}^n から \mathbb{R} への2回微分可能な関数
- $\nabla f(\boldsymbol{x})$: 点 \boldsymbol{x} における f の勾配ベクトル
- $\nabla^2 f(\boldsymbol{x})$: 点 \boldsymbol{x} における f の Hesse 行列

信頼領域法では、無制約最適化問題の目的関数 f に対して、点 \mathbf{x}_k ($k = 1, 2, \dots$) における f の 2 次近似を最小化する、次の制約付き最適化問題（以下、部分問題と呼ぶ）を解くことで探索ステップを求める。

$$\begin{aligned} & \text{minimize} && q_k(\mathbf{s}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T B_k \mathbf{s} \\ & \text{subject to} && \|\mathbf{s}\| \leq \Delta_k \end{aligned} \quad (1)$$

ここで、 \mathbf{x}_k は k 回目の反復で得られる近似解であり、 B_k は $\nabla^2 f(\mathbf{x}_k)$ 、または $\nabla^2 f(\mathbf{x}_k)$ を近似する対称行列である。Newton 法 [21, Section 4.1] や準 Newton 法 [21, Section 4.2] で用いる直線探索では、探索方向が降下方向であることを保証するために $\nabla^2 f(\mathbf{x}_k)$ や B_k が正定値である必要があるが、信頼領域法では $\nabla^2 f(\mathbf{x}_k)$ や B_k が正定値でない場合でも部分問題 (1) は大域最適解をもつ [21, Section 4.3]。また、 $\Delta_k > 0$ を信頼半径といい、 $\|\mathbf{s}\| \leq \Delta_k$ の表す領域を信頼領域という。

2.1.1 近似解の更新

近似解を \mathbf{x}_k から \mathbf{x}_{k+1} に更新する方法について述べる。信頼領域法では、 k 回目の反復で部分問題 (1) を解いて \mathbf{s}_k を求めた後、目的関数 f の値の減少量と f の 2 次近似の値の減少量の比、すなわち

$$\rho_k := \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{s}_k)}{q_k(\mathbf{0}) - q_k(\mathbf{s}_k)} = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{s}_k)}{-\nabla f(\mathbf{x}_k)^T \mathbf{s}_k - \frac{1}{2} \mathbf{s}_k^T B_k \mathbf{s}_k} \quad (2)$$

と与えられた閾値 ξ ($0 < \xi < 1$) を用いて以下のように近似解 \mathbf{x}_{k+1} を求める。

- (i) f の値がよく減少している、すなわち $\rho_k \geq \xi$ ならば、 $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ として近似解を更新。
- (ii) f の値があまり減少していない、または増加している、すなわち $\rho_k < \xi$ ならば、 $\mathbf{x}_{k+1} = \mathbf{x}_k$ として近似解を維持。

2.1.2 信頼半径の更新

信頼半径を Δ_k から Δ_{k+1} に更新する方法について述べる。信頼領域法では、 k 回目の反復で部分問題 (1) を解いて \mathbf{s}_k を求めた後、式 (2) で表される ρ_k と与えられた閾値 $\xi_1, \xi_2, \eta_1, \eta_2$ ($0 < \xi_1 \leq \xi_2 < 1, 0 < \eta_1 < 1 < \eta_2$) を用いて以下のように信頼半径 Δ_{k+1} を求める。

- (i) f の値が十分減少している、すなわち $\rho_k \geq \xi_2$ ならば、 $\Delta_{k+1} = \eta_2 \Delta_k$ として信頼半径を拡大。
- (ii) f の値が十分減少していない、すなわち $\rho_k < \xi_1$ ならば、 $\Delta_{k+1} = \eta_1 \Delta_k$ として信頼半径を縮小。
- (iii) $\xi_1 \leq \rho_k < \xi_2$ ならば、 $\Delta_{k+1} = \Delta_k$ として信頼半径を維持。

以上のことをまとめて、 $B_k = \nabla^2 f(\mathbf{x}_k)$ とした場合の信頼領域法のアルゴリズムをアルゴリズム 2.1 に示す。

2.1.3 非線形連立方程式の解法

非線形連立方程式

$$f_1(\mathbf{x}) = f_2(\mathbf{x}) = \dots = f_m(\mathbf{x}) = 0$$

アルゴリズム 2.1 信頼領域法 [21, Section 4.3]

入力 : \mathbb{R}^n から \mathbb{R} への 2 回微分可能な関数 f , 初期点 $\mathbf{x} \in \mathbb{R}^n$, 初期信頼半径 $\Delta > 0$,

パラメータ $\xi_1, \xi_2, \eta_1, \eta_2$ ($0 < \xi_1 \leq \xi_2 < 1, 0 < \eta_1 < 1 < \eta_2$), 許容誤差 ε

出力 : f の停留点

```

1: while  $\|\nabla f(\mathbf{x})\| > \varepsilon$  do
2:    $B := \nabla^2 f(\mathbf{x})$ ;
3:    $\mathbf{s} :=$  部分問題 (1) の解;
4:    $\rho := (f(\mathbf{x}) - f(\mathbf{x} + \mathbf{s})) / (-\nabla f(\mathbf{x})^T \mathbf{s} - \frac{1}{2} \mathbf{s}^T B \mathbf{s})$ ;
5:   if  $\rho < \xi_1$  then
6:      $\Delta := \eta_1 \Delta$ ; ▷ 信頼半径を縮小
7:   else if  $\xi_1 \leq \rho < \xi_2$  then
8:      $\mathbf{x} := \mathbf{x} + \mathbf{s}$ ; ▷ 近似解を更新
9:   else
10:     $\mathbf{x} := \mathbf{x} + \mathbf{s}$ ;  $\Delta = \eta_2 \Delta$ ; ▷ 近似解を更新し, 信頼半径を拡大
11:   end if
12: end while
13: return  $\mathbf{x}$ ;

```

の数値解法は以下の非線形最小二乗問題 [21, Section 5.2] の解法として定式化される。

$$\text{minimize } \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2 \quad (3)$$

ここで, $\mathbf{r}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))^T$ であり, 一般に残差ベクトルと呼ばれる。

非線形最小二乗問題 (3) は無制約最適化問題であるため, 信頼領域法を用いて解くことができる。また, 変数 $\mathbf{x} = (x_1, \dots, x_n)^T$ に境界制約がある問題に対しては, 制約付き最適化における 1 次の必要条件 (KKT 条件) [21, 定理 6.1.4] とアフィン変換 [5] を用いることで, 信頼領域法を適用可能となる。さらに, Reflection technique を取り入れた Trust-region Interior Reflective method (TIR 法) [5] という手法が提案されており, 非線形連立方程式を効率的に解くことができる。また, 大規模な問題に対して, 信頼領域法の部分問題を解く際の計算コストを抑えるように TIR 法を改良した Subspace Trust-region Interior Reflective method (STIR 法) [5] という手法も提案されており, Python の SciPy ライブラリ [17] に含まれるモジュール `scipy.optimize` の関数 `least_squares` として実装されている。本研究では, 非線形連立方程式の求解に `least_squares` を用いる。

2.2 Minimax 近似

本節では, 発話方向推定問題の解法に用いる Minimax 近似の定義や性質, および Minimax 多項式の生成アルゴリズムである Remez 第 2 アルゴリズムについて述べる [3]。まず, 本節で用いる記号を以下のように定義する。

- $\mathbb{R}[x]$: \mathbb{R} 上の多項式環
- $\mathcal{P}_k \subset \mathbb{R}[x]$: k 次以下の 1 変数多項式の集合
- f : 閉区間 $[a, b] \subset \mathbb{R}$ ($a < b$) から \mathbb{R} への連続関数

次に、多項式近似の最大誤差と Minimax 近似およびその性質を復習する。

定義 1 (多項式近似の最大誤差 [3, Section 10.11])

閉区間 $[a, b]$ 上で、連続関数 f を多項式 $P(x) \in \mathcal{P}_k$ で近似したとき、

$$\max_{a \leq x \leq b} |f(x) - P(x)|$$

を f と P の最大誤差という。

定義 2 (Minimax 近似 [3, Section 10.11])

連続関数 f に対して、多項式 $P(x) \in \mathcal{P}_k$ で、 f との最大誤差を最小化する P を f の Minimax 多項式という。また、 f を Minimax 多項式を用いて近似することを f の Minimax 近似という。

定義 2 より、任意の $Q \in \mathcal{P}_k$ に対して、

$$P \in \mathcal{P}_k \text{ が } f \text{ の Minimax 近似} \Leftrightarrow \max_{a \leq x \leq b} |f(x) - P(x)| \leq \max_{a \leq x \leq b} |f(x) - Q(x)| \quad (4)$$

が成り立つことがわかる。すなわち、Minimax 多項式は誤差の点で最良の近似多項式といえる。

Minimax 多項式の性質に関する定理を以下に述べる。

定理 3 (Minimax 多項式の存在性と一意性 [3, Section 10.11])

任意の連続関数 f に対して、 f の Minimax 多項式 $P(x) \in \mathcal{P}_k$ は一意に存在する。

定理 4 (Minimax 近似の等振動性 [3, Section 10.11])

$P(x) \in \mathcal{P}_k$ を連続関数 f の Minimax 多項式とする。このとき、閉区間 $[a, b]$ 上に少なくとも $k+2$ 個の点 $x_0 < x_1 < \cdots < x_\ell$ ($\ell \geq k+1$) が存在して、

$$f(x_i) - P(x_i) = (-1)^i \max_{a \leq x \leq b} |f(x) - P(x)| \quad (i = 1, 2, \dots, \ell) \quad (5)$$

が成り立つ。

式 (5) が成り立つとき、 f と P の誤差関数 $f - P$ は $[a, b]$ において等振動するという。

定理 5 ([3, Section 10.11])

連続関数 f と多項式 $Q(x) \in \mathcal{P}_k$ に対して、 $f(x) - Q(x)$ の閉区間 $[a, b]$ 上の $k+2$ 個の点 $x_0 < x_1 < \cdots < x_{k+1}$ で極値¹⁾をとるとし、

$$f(x_i) - Q(x_i) = (-1)^i \lambda_i \quad (i = 0, 1, \dots, k+1)$$

が成り立つとする。ここで、 λ_i は正の実数とする。このとき、任意の $S(x) \in \mathcal{P}_k$ に対して、

$$\max_{a \leq x \leq b} |f(x) - S(x)| \geq \min_i \lambda_i \quad (6)$$

が成り立つ。

式 (6) において、 $S(x)$ を f の Minimax 多項式 $P(x)$ で置き換え、式 (4) と合わせて次の不等式を得る。

$$\min_i \lambda_i \leq \max_{a \leq x \leq b} |f(x) - P(x)| \leq \max_i \lambda_i.$$

よって、 $\min_i \lambda_i = \max_i \lambda_i$ を満たす多項式を求めれば、それが f の Minimax 多項式である。

¹⁾閉区間 $[a, b]$ の端点を含む。

アルゴリズム 2.2 Remez 第 2 アルゴリズム [3, Section 10.11]

入力 : 連続関数 f , 閉区間 $[a, b]$, 次数 k , 許容誤差 ε

出力 : f の Minimax 多項式 $P(x) \in \mathcal{P}_n$

```

1:  $E_{\max} := \infty; E_{\min} := 0;$ 
2:  $I := [x_0, x_1, \dots, x_{k+1}];$  ▷  $k + 2$  個の初期補間点
3: while  $E_{\max} - E_{\min} > \varepsilon$  do
4:    $[a_0, a_1, \dots, a_k, E] := \left( f(x_i) - \sum_{j=0}^k a_j(x_i)^j = (-1)^i E \quad (i = 0, 1, \dots, k + 1) \right)$  の解;
5:    $P(x) := \sum_{j=0}^k a_j x^j;$ 
6:    $G(x) := f(x) - P(x);$ 
7:    $[x_0, x_1, \dots, x_{k+1}] := G(x)$  の  $k + 2$  個の極値点;
8:    $I := [x_0, x_1, \dots, x_{k+1}];$ 
9:    $E_{\max} := \max_i |G(x_i)|; E_{\min} := \min_i |G(x_i)|;$ 
10: end while
11: return  $P;$ 

```

以上のことから, Remez 第 2 アルゴリズム [3, Section 10.11] と呼ばれる, 連続関数に対する Minimax 多項式の生成アルゴリズムが得られる. 離散点の集合に対する Minimax 多項式の生成アルゴリズムもあり, Remez 第 1 アルゴリズムまたは Remez exchange アルゴリズム [3, Section 10.11] と呼ばれる. Remez 第 2 アルゴリズムをアルゴリズム 2.2 に示す.

アルゴリズム 2.2 の正当性や停止性については [8] を参照せよ. アルゴリズム 2.2 では, まず $k + 2$ 個の初期補間点 $x_0 < x_1 < \dots < x_{k+1}$ を与える必要があるが, その選び方に依らず反復は収束する [3, Section 10.11].

次に, Minimax 近似における次数や閉区間と最大誤差の関係を以下に示す.

定理 6 ([3, Section 4.2])

閉区間 $[a, b]$ において, $P \in \mathcal{P}_k$ を連続関数 f の Minimax 多項式とする. このとき, k が大きく, $b - a$ が小さいほど f と P の最大誤差 $\max_{a \leq x \leq b} |f(x) - P(x)|$ は小さくなる.

アルゴリズム 2.2 の while 文内では, まず

$$f(x_i) - \sum_{j=0}^k a_j(x_i)^j = (-1)^i E \quad (i = 0, 1, \dots, k + 1) \quad (7)$$

を解く. 式 (7) は $k + 2$ 個の変数 a_0, a_1, \dots, a_k, E をもつ $k + 2$ 本の方程式からなる連立 1 次方程式である. 次に, 誤差関数 $f(x) - P(x)$ の極値点の計算が必要となるが, 微分可能でない場合でも Brent 法を用いることで極値点の近似値を求められる [3, Section 10.11]. よって, アルゴリズム 2.2 では入力する連続関数に微分可能性を必要としない. また, 連続関数 f の定義域が閉区間でない場合は, 定義域を含む閉区間上でアルゴリズム 2.2 を用いる.

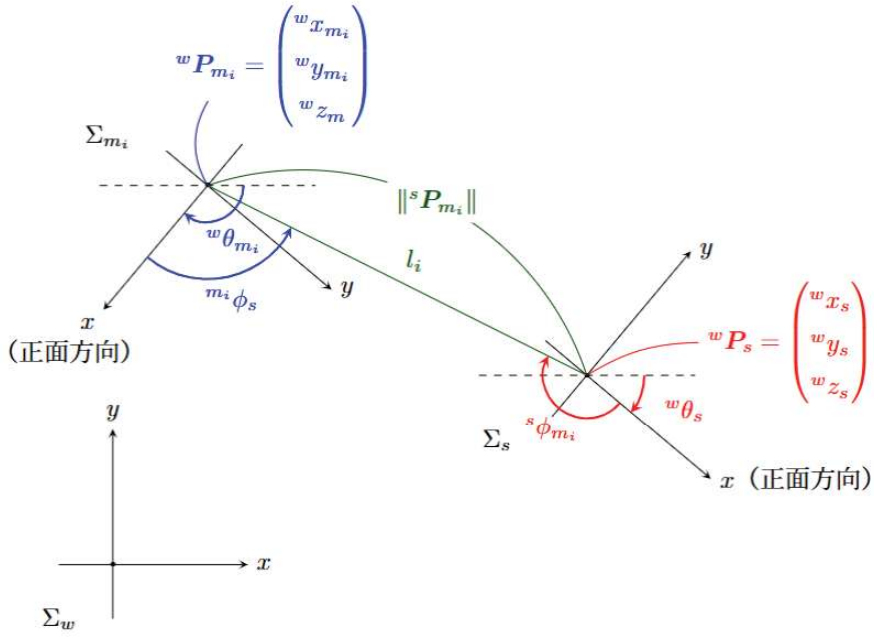


図 1: 話者の真上から見た図

3 座標系の定義とパラメータの分類

3.1 座標系の定義

基準となる座標系や話者, i 番目のロボットの位置を原点とする座標系の定義を図 1, 2 に示す. なお, 本論文では n 台のロボットに 1 から n までの番号をつけ, Robot- i ($i = 1, \dots, n$) と表す. 本研究では, 以下の条件でロボットが配置された状況を仮定する.

- n 台のロボットはすべて同一平面 ($z = w z_m$) 上にあるとする. よって, ワールド座標系における Robot- i の位置は, $w \mathbf{P}_{m_i} = (w x_{m_i}, w y_{m_i}, w z_m)^T$ とする.
- ロボットの姿勢は水平に固定されているものとする. すなわち, $w \omega_{m_i} = 0$ とする.

座標系の定義から, 座標系 Σ_s におけるロボットの位置 $s \mathbf{P}_{m_i} = (s x_{m_i}, s y_{m_i}, s z_{m_i})^T$ は変換行列 $w \mathbf{R}_s$ を用いて次式で表される.

$$s \mathbf{P}_{m_i} = w \mathbf{R}_s^{-1} (w \mathbf{P}_{m_i} - w \mathbf{P}_s). \quad (8)$$

ここで, $w \mathbf{R}_s$ は y 軸および z 軸を中心とする回転行列の積である. 式 (8) より, $s \mathbf{P}_{m_i} = (s x_{m_i}, s y_{m_i}, s z_{m_i})^T$ は次式で表される.

$$s \mathbf{P}_{m_i} = \begin{pmatrix} (w x_{m_i} - w x_s) \cos(w \theta_s) \cos(w \omega_s) + (w y_{m_i} - w y_s) \sin(w \theta_s) \cos(w \omega_s) + (w z_m - w z_s) \sin(w \omega_s) \\ -(w x_{m_i} - w x_s) \sin(w \theta_s) + (w y_{m_i} - w y_s) \cos(w \theta_s) \\ -(w x_{m_i} - w x_s) \cos(w \theta_s) \sin(w \omega_s) - (w y_{m_i} - w y_s) \sin(w \theta_s) \sin(w \omega_s) + (w z_m - w z_s) \cos(w \omega_s) \end{pmatrix}. \quad (9)$$

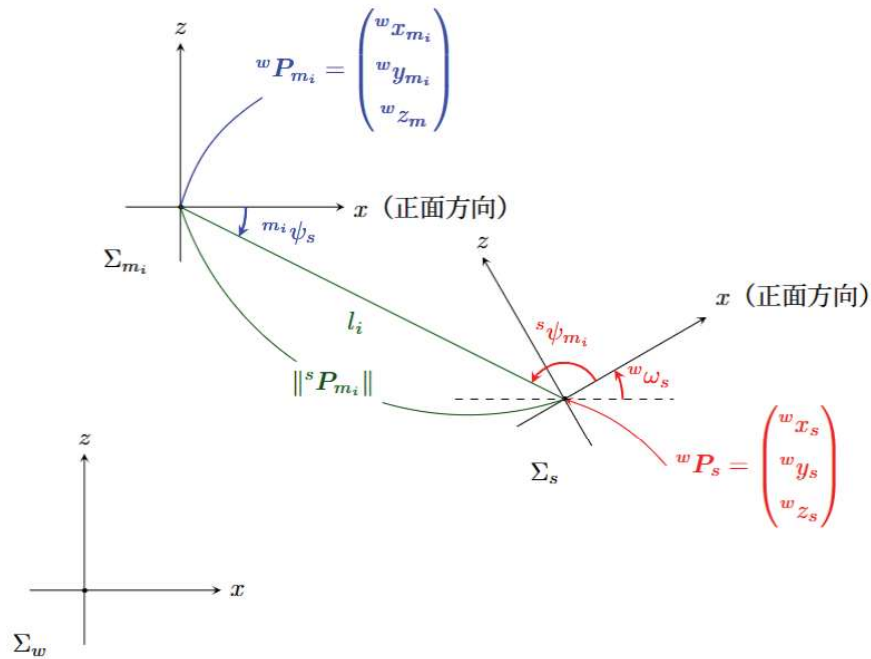


図 2: 話者の真横から見た図

3.2 パラメータの分類

第 3.1 節で述べたパラメータを既知パラメータと未知パラメータに分類する。本研究で扱うロボットはマイクロフォン・アレイを搭載しており、水平方向における声の到来方向 (Direction of Arrival, DoA) を計測できる。また、Ultra Wide Band (UWB) [10] による Simultaneous Localization and Mapping (SLAM) [6] という技術により、ロボット自身の位置と姿勢を推定することができる。以上のことから、既知パラメータと未知パラメータは以下の通りである。

既知パラメータ

- $m_i \phi_s$: マイクロフォン・アレイにより推定可能。
- $w P_{m_i}, w \theta_{m_i}$: UWB による SLAM により推定可能。

未知パラメータ

- $w P_s, w \theta_s, w \omega_s, s \phi_{m_i}, s \psi_{m_i}, m_i \psi_s$

本問題において推定したい発話方向は、ワールド座標系に対して話者が向いている方向である。よって、推定したい未知パラメータは $w \theta_s, w \omega_s$ である。

4 音声拡散モデル

ロボットに搭載されたマイクロフォン・アレイで計測される音圧レベルの理論値を求めるための数学モデルを音声拡散モデル [20] といい、次式で表される。

$$L_{P_i} = L_{P_0} - 20 \log_{10} \| {}^s \mathbf{P}_{m_i} \| - a \left(1 - \frac{1 + \cos \theta_i}{2} \right) \quad (10)$$

$$\Leftrightarrow 2L_{P_i} - 2L_{P_0} + 40 \log_{10} \| {}^s \mathbf{P}_{m_i} \| + a(1 - \cos \theta_i) = 0. \quad (11)$$

式 (10) の右辺第 2 項を距離減衰項、右辺第 3 項を角度減衰項という。本研究では角度減衰の大きさを表す数学モデルとして、次式で表される 3 次元カージオイド曲面を用いる。

$$\begin{cases} D = -a \left(1 - \frac{1 + \cos \theta}{2} \right) \\ x = D \cos \theta \\ y = D \sin \theta \cos \phi \\ z = D \sin \theta \sin \phi \end{cases} \quad (-\pi < \theta \leq \pi, 0 \leq \phi < \pi) \quad (12)$$

ここで、 D は角度 θ , ϕ における角度減衰の大きさを表す。

式 (12) で表される 3 次元カージオイド曲面の極座標系を図 3 に示す。ただし、3 次元カージオイド曲面は座標系 Σ_s における曲面であることに注意する。原点 O (話者の位置) を端点とする半直線 OP 上に Robot- i があるので、3 次元カージオイド曲面における角度は θ_i , ϕ_i と表す。 θ_i は x 軸正方向と動径 OP のなす角である。また、点 P を yz 平面に射影した点を P' とすると、 ϕ_i は y 軸正の向きと動径 OP' とのなす角である。以上のことから、座標系 Σ_s におけるロボットの位置 ${}^s \mathbf{P}_{m_i} = ({}^s x_{m_i}, {}^s y_{m_i}, {}^s z_{m_i})^T$ は次式で表される。

$${}^s \mathbf{P}_{m_i} = \begin{pmatrix} {}^s x_{m_i} \\ {}^s y_{m_i} \\ {}^s z_{m_i} \end{pmatrix} = \begin{pmatrix} \| {}^s \mathbf{P}_{m_i} \| \cos \theta_i \\ \| {}^s \mathbf{P}_{m_i} \| \sin \theta_i \cos \phi_i \\ \| {}^s \mathbf{P}_{m_i} \| \sin \theta_i \sin \phi_i \end{pmatrix} \quad (-\pi < \theta_i \leq \pi, 0 \leq \phi_i < \pi). \quad (13)$$

式 (9), (13) の x 座標より、次式を得る。

$$\begin{aligned} ({}^w x_{m_i} - {}^w x_s) \cos({}^w \theta_s) \cos({}^w \omega_s) + ({}^w y_{m_i} - {}^w y_s) \sin({}^w \theta_s) \cos({}^w \omega_s) \\ + ({}^w z_{m_i} - {}^w z_s) \sin({}^w \omega_s) - \| {}^s \mathbf{P}_{m_i} \| \cos \theta_i = 0. \end{aligned} \quad (14)$$

5 3次元発話方向推定問題の解法

本章では、3 次元空間における発話方向推定問題の解法、すなわち ${}^w \theta_s$ と ${}^w \omega_s$ を求める方法について述べる。解法は以下の 2 ステップに分けられる。

ステップ 1. 話者の位置 ${}^w \mathbf{P}_s = ({}^w x_s, {}^w y_s, {}^w z_s)^T$ を推定する。

ステップ 2. 発話方向 ${}^w \theta_s$, ${}^w \omega_s$ を推定する。

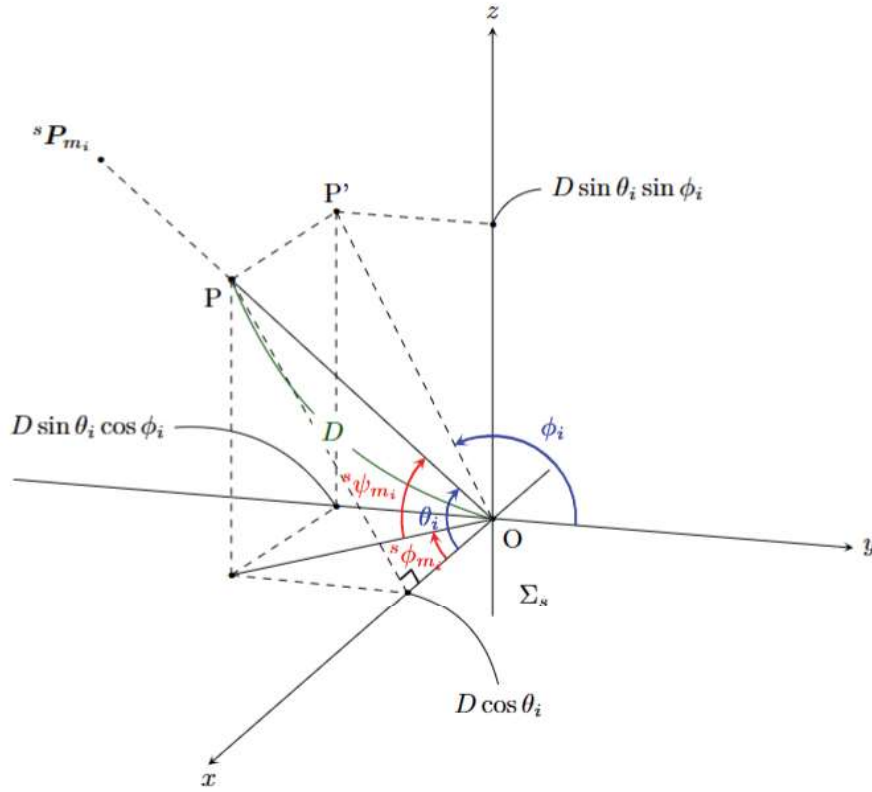


図 3: 式 (12) の極座標系

5.1 話者の位置推定

座標系 Σ_w における話者の位置 ${}^w \mathbf{P}_s = ({}^w x_s, {}^w y_s, {}^w z_s)^T$ を求める. 直線 l_i の方程式は $t_i \in \mathbb{R}$ を用いて次式で表される.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = {}^w \mathbf{P}_{m_i} + t_i \begin{pmatrix} 1 \\ \tan({}^w \theta_{m_i} + {}^{m_i} \phi_s) \\ \tan({}^{m_i} \psi_s) \end{pmatrix} = \begin{pmatrix} {}^w x_{m_i} \\ {}^w y_{m_i} \\ {}^w z_m \end{pmatrix} + t_i \begin{pmatrix} 1 \\ \tan({}^w \theta_{m_i} + {}^{m_i} \phi_s) \\ \tan({}^{m_i} \psi_s) \end{pmatrix}.$$

直線 l_1, \dots, l_n の交点を求めれば, それが話者の位置 ${}^w \mathbf{P}_s$ である. 実際は計測誤差の影響により, 一般に l_1, \dots, l_n は 1 点 ${}^w \mathbf{P}_s$ で交わらないため最小二乗解を求める.

5.2 発話方向推定

解法 1 : 式 (11), (14) を合わせて得られる以下の非線形連立方程式を解く.

$$\begin{aligned} & ({}^w x_{m_i} - {}^w x_s) \cos({}^w \theta_s) \cos({}^w \omega_s) + ({}^w y_{m_i} - {}^w y_s) \sin({}^w \theta_s) \cos({}^w \omega_s) \\ & + ({}^w z_m - {}^w z_s) \sin({}^w \omega_s) - \|{}^s \mathbf{P}_{m_i}\| \cos \theta_i = 0, \end{aligned} \quad (15)$$

$$2L_{P_i} - 2L_{P_0} + 40 \log_{10} \|{}^s \mathbf{P}_{m_i}\| + a(1 - \cos \theta_i) = 0.$$

式 (15) は変数 $n + 5$ 個, 方程式 $2n$ 本からなる非線形連立方程式である. よって, 解が有限個に定まるための必要条件は $2n \geq n + 5$, すなわち $n \geq 5$ である.

解法 2 : 式 (15) において, $T := \sin({}^w\theta_s)$, $U := \cos({}^w\theta_s)$, $X := \sin({}^w\omega_s)$, $Y := \cos({}^w\omega_s)$, $Q_i := \cos\theta_i$ とおくと以下の非線形連立方程式が得られる.

$$\begin{aligned} &({}^w x_{m_i} - {}^w x_s)UY + ({}^w y_{m_i} - {}^w y_s)TY + ({}^w z_m - {}^w z_s)X - \|{}^s \mathbf{P}_{m_i}\|Q_i = 0, \\ &2L_{P_i} - 2L_{P_0} + 40 \log_{10} \|{}^s \mathbf{P}_{m_i}\| + a(1 - Q_i) = 0, \\ &T^2 + U^2 - 1 = 0, \\ &X^2 + Y^2 - 1 = 0. \end{aligned} \tag{16}$$

式 (16) は変数 $n + 7$ 個, 方程式 $2n + 2$ 本からなる非線形連立方程式である. よって, 解が有限個に定まるための必要条件は $2n + 2 \geq n + 7$, すなわち $n \geq 5$ である.

解法 3 : 式 (16) において, $\delta_i := 40 \log_{10} \|{}^s \mathbf{P}_{m_i}\|$ とおくと以下の連立代数方程式が得られる.

$$\begin{aligned} &\{({}^w x_{m_i} - {}^w x_s)UY + ({}^w y_{m_i} - {}^w y_s)TY + ({}^w z_m - {}^w z_s)X\}^2 - \|{}^s \mathbf{P}_{m_i}\|^2 Q_i^2 = 0, \\ &2L_{P_i} - 2L_{P_0} + \delta_i + a(1 - Q_i) = 0, \\ &T^2 + U^2 - 1 = 0, \\ &X^2 + Y^2 - 1 = 0. \end{aligned} \tag{17}$$

式 (17) は変数 $2n + 7$ 個, 方程式 $2n + 2$ 本からなる連立代数方程式である. よって, 変数の個数が方程式の本数より多いので解が有限個に定まらない. そこで, $\delta_i := 40 \log_{10} \|{}^s \mathbf{P}_{m_i}\|$ を変形して, 変数の個数を減らす. δ_i は ${}^w z_s$ のみを変数にもつ連続関数であるから, 式 (17) において δ_i を k 次 Minimax 多項式

$$d_{i,0} + d_{i,1} {}^w z_s + d_{i,2} {}^w z_s^2 + \cdots + d_{i,k} {}^w z_s^k$$

で置き換えることにより, 次式を得る.

$$\begin{aligned} &\{({}^w x_{m_i} - {}^w x_s)UY + ({}^w y_{m_i} - {}^w y_s)TY + ({}^w z_m - {}^w z_s)X\}^2 - \|{}^s \mathbf{P}_{m_i}\|^2 Q_i^2 = 0, \\ &2L_{P_i} - 2L_{P_0} + d_{i,0} + d_{i,1} {}^w z_s + d_{i,2} {}^w z_s^2 + \cdots + d_{i,k} {}^w z_s^k + a(1 - Q_i) = 0, \\ &T^2 + U^2 - 1 = 0, \\ &X^2 + Y^2 - 1 = 0. \end{aligned} \tag{18}$$

式 (18) は変数 $n + 7$ 個, 方程式 $2n + 2$ 本からなる連立代数方程式である. よって, 解が有限個に定まるための必要条件は $2n + 2 \geq n + 7$, すなわち $n \geq 5$ である.

6 実験

本章では, 各解法における連立方程式の解法の精度や誤差, 計算量を評価するための実験について述べる.

6.1 実験手順

実験の手順は以下の通りとする.

1. ロボットの台数は $n = 5$ とし, 未知パラメータの値を無作為に与える.
2. 手順 1 で与えた未知パラメータの値から既知パラメータの値を求める.
3. 手順 2 で得られた既知パラメータの値を用いて未知パラメータの値を求める.
4. 手順 3 の結果を用いて精度評価や誤差評価, 計算量評価を行う.

6.1.1 連立方程式の解法

連立方程式 (15), (16), (18) の解法には, 信頼領域法 [21, Section 4.3] に基づく制約付き最適化の手法である, Python の SciPy ライブラリ [17] に含まれるモジュール `scipy.optimize` の関数 `least_squares` を用いる. ただし, ある初期値に対して `least_squares` を用いて連立方程式の近似解を求めても, 解になっていないとは限らない. そこで, 以下の手順で連立方程式の解を求める.

1. 1 個の初期値を無作為に与え, `least_squares` で式 (15) または式 (16) の近似解を求める.
2. 近似解を式 (15) または式 (16) の左辺に代入し, 与えられた閾値を下回れば終了する.
3. 手順 2 が満たされなければ, 手順 1 に戻る. 手順 1, 2 を 100 回繰り返しても解が得られなければ終了する.

6.2 計算環境

本研究における計算環境は以下の通りである.

- CPU: Intel(R) Xeon(R) Silver 4210 CPU @ 2.20GHz
- RAM: 251GiB
- OS: Ubuntu 20.04.6 LTS
- Python: version 3.8.10
- Risa/Asir: version 20230315 [14]

6.3 実験結果

第 5.2 節で述べた各解法それぞれに対して, 第 6.1 節の手順による実験を行った結果を表 1 に示す. 実験に用いたサンプル数は 10000 である.

第 6.1.1 節で述べたように, ある初期値に対して得られた近似解を連立方程式の左辺に代入し, 与えられた閾値を下回るかを確認する. 今回の実験ではその閾値を 10^{-8} とする. また, 得られた解の平均相対誤差が 10^{-5} を下回った場合に正しい解が得られたとする.

また, 解法 3 では Minimax 近似の次数は 5 次, 許容誤差は 10^{-5} とし, 7 個の初期補間点には 6 次 Chebyshev 多項式 $T_6(x)$ の極値点を用いた.

解法	平均計算時間 [s]	解が出た回数 [回]	正しい解が出た回数 [回]
1	0.2754	9976	9919
2	0.1847	9958	9894
3	5.5050	9502	2513

表 1: 解法の比較

解法 3 の精度が悪化した原因として, Minimax 近似を用いたことによって, 正しい解が連立代数方程式 (18) の解になっていないことが予想される. このことを確かめるために, Minimax 多項式を含む方程式 (式 (18) の第 2 式) の左辺に正しい解を代入したときの値と 0 (右辺の値) との誤差を評価したところ, 解法 1, 2 よりも誤差が大きいことがわかった.

7 まとめと今後の課題

本研究では、3次元発話方向推定問題の解法として、信頼領域法を用いた非線形連立方程式の求解による手法を提案した。また、3種類の連立方程式の解法の精度や誤差、計算量を評価するための実験を行った。解法1、解法2では精度、誤差、計算量すべてにおいて実用上有効といえる結果が得られたが、解法3では精度、誤差、計算量すべてにおいて解法1および解法2に及ばない結果となった。最後に、今後の課題を以下に挙げる。

- 解法1、解法2で正しい解が得られない場合の原因の解明と対処法の提案
- 解法3の精度向上と計算時間の短縮
- 既知パラメータの計測誤差を考慮した場合の精度評価と誤差評価
- 声の指向特性モデルの改良

謝 辞

本研究は、京都大学の国際共同利用・共同研究拠点である数理科学研究所の支援を受けています。また、本研究は公益財団法人 JKA 機械振興補助事業 振興事業補助 研究補助・個別研究「2025年度数式処理によるロバストかつ効率的な計算に基づくマニピュレータの動作計画手法の開発補助事業」および NTU-UGA-UT Trilateral Center の補助を受けています。

参考文献

- [1] K. Ahuja, A. Kong, M. Goel and C. Harrison. Direction-of-voice (dov) estimation for intuitive speech interaction with smart devices ecosystems. In UIST. pp. 1121–1131, 2020.
- [2] J. Alonso-Mora, S. H. Lohaus, P. Leemann, R. Siegwart and P. Beardsley. Gesture based human-multi-robot swarm interaction and its application to an interactive display. In 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 5948–5953, 2015.
- [3] H. M. Antia. Numerical Methods for Scientists and Engineers, Third Edition. Hindustan Book Agency, 2012.
- [4] A. van den Bos. Parameter Estimation for Scientists and Engineers. John Wiley & Sons, 2007.
- [5] M. A. Branch, T. F. Coleman, and Y. Li. A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. SIAM Journal on Scientific Computing. vol. 21, No. 1, pp. 1–23, 1999.
- [6] C. Cadena, L. Carlone, et al. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. IEEE Transactions on Robotics. vol. 32, No. 6, pp. 1309–1332, 2016. <https://doi.org/10.1109/TRO.2016.2624754>.
- [7] A. Couture-Beil, R. T. Vaughan, and G. Mori. Selecting and commanding individual robots in a multi-robot system. In 2010 Canadian Conference on Computer and Robot Vision. IEEE, pp. 159–166, 2010.

- [8] N. Daili and A. Guesmia. Remez Algorithm Applied to the Best Uniform Polynomial Approximations. *General Mathematics Notes*. vol. 17, No. 1, pp. 16–31, 2013.
- [9] R. C. Felsheim et al. Head Orientation Estimation from Multiple Microphone Arrays. *Proceedings of the 28th European Signal Processing Conference (EUSIPCO)*. pp. 491–495, 2020.
- [10] J. Haapola et al. An Ultra Wideband Survey: Global Regulations and Impulse Radio Research Based on Standards. *IEEE Communications Surveys & Tutorials*. vol. 19, No. 2, pp. 874–890, 2016. <https://doi.org/10.1109/COMST.2016.2634593>.
- [11] K. Kuramochi, A. Terui, and M. Mikawa. Solving Estimation Problems Using Minimax Polynomials and Gröbner Bases. *Proceedings of the 10th International Symposium on Symbolic Computation in Software Science (SCSS 2024), Work-in-progress workshop*. pp. 57–62, 2024.
- [12] M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1998.
- [13] K. Müller, B. Cakmak, P. Didier, S. Doclo, J. Østergaard and T. Wolff. Head orientation estimation with distributed microphones using speech radiation patterns. *Asilomar Conference on Signals, Systems, and Computers (ACSSC 2023)*. IEEE, pp. 1404–1409, 2023.
- [14] M. Noro and T. Takeshima. Risa/Asir — A Computer Algebra System. *Papers from the International Symposium on Symbolic and Algebraic Computation*. pp. 387–396, 1992.
- [15] C. Segura and J. Hernando. 3D Joint Speaker Position and Orientation Tracking with Particle Filters. *Sensors*. vol. 14, pp. 2259–2279, 2014.
- [16] J. C. Tejada et al. A Review of Multi-Robot Systems and Soft Robotics: Challenges and Opportunities. *Sensors*. vol. 25, No. 5, 1353p, 2025.
- [17] P. Virtanen, R. Gommers, T. E. Oliphant, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. 17, pp. 261–272, 2020.
- [18] J. J. Yang, G. Banerjee, V. Gupta, M. S. Lam and J. A. Landay. Soundr: Head position and orientation prediction using a microphone array. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI’20*. pp. 1–12, 2020.
- [19] Q. Yang and Y. Zheng. Model-based head orientation estimation for smart devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* vol. 5, No. 3, 2021.
- [20] 加藤虎之介. 発話方向推定に基づくマルチロボットののための個体選択手法. 筑波大学人間総合科学学術院人間総合科学研究群修士論文, 74p, 2024.
- [21] 成島康史, 中山舜民, 矢部博. 非線形最適化法 — 数理的基礎と Python による実装. オーム社, 376p, 2025.