

パラメータ多様体の記述手法の提案と ウイルスダイナミクスの解析への応用

小松 瑞果^{1,a)} 谷口 隆晴^{1,2}

概要: 時系列データの背後にある系について、未知パラメータを含む状態空間モデルを用いて解析する場合、データを用いた未知パラメータの推定が問題となる。このような問題を扱うとき、モデルの同定可能性、すなわち、データから未知パラメータが一意に定まるかどうかを考慮する必要がある。特に、パラメータが一意に定まらない場合、何らかの手法によって得られたパラメータは、候補となるパラメータの集合の一部でしかないため、このような手法では、モデルとデータから定まる情報を見落とし得る。これに対し、本研究では、状態空間モデルと時系列データから定まる情報を、パラメータ多様体として一意に記述する手法を提案する。また、実際に、生体内におけるウイルスダイナミクスに関するモデルに、提案手法を適用し、既存のパラメータ推定手法では得られなかった、新たな知見が得られたことを示す。

キーワード: パラメータ多様体、構造的同定可能性解析、状態空間モデル、グレブナー基底

Description of the Parameter Manifold and Its Applications to Analysis of Virus Dynamics

Abstract: To investigate a system that lies behind observed time series data using a state space model, which is supposed to describe the system, together with unknown parameters, the problem to tackle is the estimation of the parameters. For this kind of problems, it is necessary to care about the identifiability of a model, that is, whether its unknown parameters are uniquely determined from the data. Especially if they are not uniquely determined, a parameter vector which is obtained in some way is just a part of a set of possible parameter vectors, and hence, such typical approaches result in losing information that is determined from the data and the model. Against this, we propose a method to describe the information determined uniquely from a state space model and time series data as the parameter manifold, which is a manifold that consists of possible parameters. Besides, we apply the proposed method to a model that describes the dynamics of virus in vivo, and then, show that a new insight overlooked by conventional approaches is actually obtained by our method.

Keywords: Parameter manifold, Structural identifiability analysis, State space model, Gröbner basis

1. はじめに

本研究では、観測された時系列データの背後にある系に対し、状態空間モデルを用いて解析する問題を扱い、特に、データとモデルから一意に定まる情報を陽に記述する手法

を提案する。

観測された時系列データに対し、背後にある系を未知パラメータを含む数理モデルで記述し、解析する問題は、システムズバイオロジーや制御など、様々な分野で表れる [1], [2]。この問題は、具体的には、状態空間モデル

$$\frac{dx}{dt} = f(x, u; a) \quad (1)$$

$$y = g(x; a) \quad (2)$$

について、与えられた入出力時系列データ $u_d(t), y_d(t)$ を観測モデル (2) が近似するように、未知パラメータ a の値を推定する問題に帰着される。ただし、 $x(t) \in \mathbb{R}^N$ を状態変

¹ 神戸大学大学院システム情報学研究科
Graduate School of System Informatics, Kobe University,
1-1 Rokkodai-cho, Nada-ku, Kobe, 657-8501, Japan
² JST さきがけ
JST PRESTO, K's Gobancho, 7, Gobancho, Chiyoda-ku,
Tokyo, 102-0076, Japan
^{a)} m-komatsu@stu.kobe-u.ac.jp

数, $u(t) \in \mathbb{R}^M$ を入力ベクトル, $y(t) \in \mathbb{R}$ を出力, $a \in \mathbb{R}^n$ をパラメータベクトル, N, M, n を正の整数とする. また, $u_d(t), y_d(t)$ は, それぞれ $u(t), y(t)$ と同じ大きさのベクトルであるが, すべての t について得られているとは限らない. このような問題を扱うとき, モデルの同定可能性, すなわち, 与えられたデータから, 未知パラメータが一意に定まるかどうかを考慮する必要がある [3], [4], [5]. パラメータが一意に定まる場合, 何らかの手法で推定したパラメータは, 唯一の真のパラメータとみなしてよい. 一方, そうでない場合, 推定されたパラメータは, 真のパラメータの集合の一部でしかない. つまり, 後者のような場合, 一般的なパラメータ推定では, 扱う問題に対する解を網羅的に得ることができず, 候補となる他の解を見落とす可能性がある. 実際, 後者のようなモデルは, 特に, 薬学やシステムバイオロジーの分野において用いられているため [1], [6], このような問題の解決には, 一定の意義がある.

上述の研究背景を, 簡単な例を用いて説明する. 未知パラメータを a_1, a_2 とする状態空間モデル

$$\frac{dx}{dt} = a_1 a_2 x \quad (3)$$

$$y = x \quad (4)$$

と初期条件 $x(0) = 1$ が与えられ, 観測データが $y_d(t) = e^{5t}$ に従う場合について考える. 式 (3) を, 与えられた初期条件のもとで解くと, $x(t) = e^{a_1 a_2 t}$ であるから, データとモデルから定まる情報は $a_1 a_2 = 5$ であり, a_1, a_2 それぞれの値は一意に定まらない. 実際, この問題を, 例えば, 観測データと観測モデルの数値解の差を最小化する問題として反復法などで数値的に解くと, 得られるパラメータは, その初期解などに依存する. このような結果の依存性をなくすには, この問題の解を包括するパラメータに関する制約式 $a_1 a_2 = 5$ 自体を, データとモデルから得られればよい.

そこで, 本研究では, この制約式, 正確には, 制約式から定まる代数的多様体を, パラメータ多様体とよび, これを具体的に記述する手法を提案する. パラメータ多様体の記述には, 主に, システムズバイオロジーの分野で発展した, システム解析の一分野である, 構造的同定可能性解析 [4], [7] を部分的に応用する. 具体的には, 構造的同定可能性解析において導かれる, 入出力関係式を用いて, パラメータ多様体を記述する. また, 入出力関係式の導出にあたり, モデルが非線形であっても適用可能な, 微分代数に基づく手法を用いる. 入出力関係式の一般的な用途は, 構造的同定可能性解析, つまり, 十分滑らかでノイズのないデータが与えられたもとの, 同定可能性の判定であるが, 本手法は, これを, 実データ解析に新たに応用する. また, 実際には, 生体内ウイルスのダイナミクスに関するモデル [6] に対して, 提案手法を適用し, 従来手法では導かれなかった新たな考察が得られたことを示す.

最後に, 本稿の構成について説明する. まず, 第2節で

は, パラメータ多様体の導出方法に関する, 理論的側面について説明する. 次に, 第3節では, パラメータ多様体の具体的な導出方法を示す. そして, 第4節では, 提案手法を, 生体内におけるウイルスダイナミクスに関するモデルと, 実験データに適用し, これを用いた解析結果を示す. 最後にまとめを述べる.

2. パラメータ多様体に関する理論的側面

本節では, 以下に定義するパラメータ多様体に関する理論的側面について, 説明する:

定義1 (パラメータ多様体) 状態空間モデル (1), (2) と, 入出力時系列データ $u_d(t), y_d(t)$ が与えられたとき, モデルの出力 $y(t; a)$ が $y_d(t)$ に一致するパラメータの集合を, パラメータ多様体と定義する.

ただし, 本稿では, 観測データ $u_d(t), y_d(t)$ はモデル (1), (2) に従い, 特に, 入力については完全であること, すなわち, $u_d(t) = u(t)$ を仮定する. なお, 完全なデータとは, ノイズがなく, 十分なめらかなデータを指す. このような仮定は, 後述する構造的同定可能性解析においても, 同様に用いられる.

2.1 構造的同定可能性解析の概要

ここではパラメータ多様体の記述に関する技術的背景である, 構造的同定可能性解析について, 簡単に説明する. 構造的同定可能性解析では, モデルへの入出力データが完全である, すなわち, $u_d(t) = u(t), y_d(t) = y(t)$ と仮定し, このときのモデルの同定可能性を判定する. 構造的同定可能性の定義は, 様々であるが [4], [8], ここでは, 完全なデータ $u_d(t) = u(t), y_d(t) = y(t)$ が与えられたもとの, 状態空間モデル (1), (2) のパラメータが一意に定まるとき, モデルは構造的に同定可能であるという. モデルが構造的に同定可能であることの必要十分条件は, モデルの入出力 $u(t), y(t)$ と, パラメータ a が一対一に対応づくことであり [9], 判定においては, これを確かめればよい. 具体的な手法としては, 線形モデルについては, 例えば, [3], 非線形モデルについては, [4], [7] などを参照されたい.

2.2 入出力関係式を用いたパラメータ多様体の導出

構造的同定可能性解析の手法のうち, 微分代数に基づく手法 [4], [7] では, モデルの入出力 $u(t), y(t)$ と, パラメータ a の関係を, 入出力関係式を用いることで解析する. なお, 入出力関係式は, 専ら, 構造的同定可能性解析に用いられる. 微分代数とは, 微分演算を演算に含む代数であり, 微分方程式を, グレブナー基底のようなイデアルに対する代数的操作により解析できる [10]. これは, 非線形モデルに対しても適用可能である. そこで, 本研究では, 微分代数的操作により得られる入出力関係式を用いて, パラメータ多様体を記述する. このように, 本研究では, 代数

に基づく手法を用いるため、以下、状態空間モデルの右辺 $f(x, u; a), g(x; a)$ は、状態変数 $x(t)$ に関する多項式で表せると仮定する。この仮定が満たされない場合は、モデルの右辺を多項式近似することで、提案手法が適用可能となる。

状態空間モデル (1), (2) に対する入出力関係式とは、モデルの入出力 $u(t), y(t)$ とパラメータ a の関係式であり、言い換えると、モデル (1), (2) から生成された、微分代数イデアルの元のうち、状態変数 $x(t)$ を用いずに表される式である。具体的には、次のように表される：

$$\sum_k c_k(a) f_k(y, \dots, \frac{d^N y}{dt^N}, u, \dots, \frac{d^{N-1} u}{dt^{N-1}}) = 0. \quad (5)$$

ここで、 $c_k(a)$ はパラメータの有理関数、 f_k は、 $u(t), y(t)$ に関する微分多項式である。なお、本稿における微分とは、時間に関する微分 $\frac{d}{dt}$ を表す。入出力関係式 (5) に関して、次の定理が成り立つ：

定理 1 (入出力関係式を用いたパラメータ多様体の導出)
入出力データ $u_d(t) = u(t), y_d(t)$ が得られていると仮定する。出力データが完全： $y_d(t) = y(t)$ であれば、各 k に対し、入出力関係式 (5) にデータとその微分値を与えることで、 $c_k(a)$ の値が一意に得られる。これを v_k とすると、連立代数方程式 $c_k(a) = v_k$ は、パラメータ空間 \mathbb{R}^n における多様体を定める。これは、データ $u_d(t) = u(t), y_d(t) = y(t)$ に対する、モデル (1), (2) のパラメータ多様体である。この定理が示すように、パラメータ多様体を得るには、モデルに対する入出力関係式を計算し、完全なデータを与えれば良い。入出力関係式 (5) は、モデルを表す式から状態変数を消去すれば得られる。入出力関係式の導出にあたり、[7] において示されている、以下の定理を用いる。

定理 2 $x(t) \in \mathbb{R}^N$ を状態変数、 $u(t) \in \mathbb{R}^M$ を入力ベクトル、 $y(t) \in \mathbb{R}$ を出力、 $a \in \mathbb{R}^n$ をパラメータベクトル、 N, M, n を正の整数とする状態変数モデル (1), (2) について、 $\mathbb{Q}(a)[x, y, u, dx/dt, dy/dt, du/dt, \dots, d^{N-1}x/dt^{N-1}, d^{N-1}y/dt^{N-1}, d^{N-1}u/dt^{N-1}, d^N x/dt^N, d^N y/dt^N]$ 上のイデアル I を次のように定める：

$$I = \langle \frac{dx}{dt} - f(x, u; a), \dots, \frac{d^N x}{dt^N} - \frac{d^{N-1}}{dt^{N-1}} f(x, u; a), \\ y - g(x; a), \dots, \frac{d^N y}{dt^N} - \frac{d^N}{dt^N} g(x; a) \rangle. \quad (6)$$

このとき、 $J = I \cap \mathbb{Q}(a)[y, u, dy/dt, du/dt, \dots, d^{N-1}y/dt^{N-1}, d^{N-1}u/dt^{N-1}, d^N y/dt^N]$ は零イデアルとはならず、入出力関係式 (5) を含む。

注意 1 ここでは、出力 y が 1 つであると仮定したが、複数出力がある場合についても、同様の定理を示すことができる [7]。簡単のため、以下、1 出力の場合のみを扱う。

グレブナー基底は、イデアルを特徴づける多項式の集合であり、以下のように定義される：

定義 2 (グレブナー基底) ある単項式順序が与えられた環 R 上のイデアルを I とする。任意の多項式 f について、

$f \in I$ であることと、 f の G による、与えられた順序に従う割り算のあまりが 0 であることが等価であるとき、 G を I のグレブナー基底という。なお、単項式順序とは、多項式における各単項式間の順序である。

この定義は、イデアルの元である、どの多項式も、グレブナー基底の元である多項式の線形結合で表されることと、それ以外はイデアルに含まれないことを意味している。また、単項式順序をうまく選ぶと、消去定理と呼ばれる定理により、入出力関係式はグレブナー基底の要素のうち、 $x, dx/dt, \dots, d^{N-1}x/dt^{N-1}$ が含まれない元として得られること、また、それで尽くされることが保証されている。

3. パラメータ多様体の記述手法

3.1 パラメータ多様体の記述の流れ

第 2 節に述べた理論的説明に基づき、ここでは、観測された入出力データ $u_d(t) = u(t), y_d(t)$ と状態空間モデル (1), (2) から、パラメータ多様体を記述する手順を示す。

- (1) 観測された出力データが完全でない場合、モデルとデータを用いて、完全な出力データを生成する。なお、観測された出力データが完全である場合、それを用いる。
- (2) 定理 2 に従い、イデアル I の生成系を定める。
- (3) イデアル I に対するグレブナー基底を導出する。この元のうち、状態変数 $x(t)$ とその微分が含まれていないものとして、入出力関係式が得られる。
- (4) 入出力関係式について、データを代入し、係数 $c_k(a)$ の値 v_k を求める。これにより、 $c_k(a) = v_k$ を満たす a の集合として、パラメータ多様体が求まる。

3.2 手順 (1)

定理 1 が示すように、完全な入出力データが得られている場合、これを用いて入出力関係式を導けばよいが、実問題において、特に、出力に相当する完全なデータが得られていることは稀である。そこで、本研究では、観測された出力データ $y_d(t)$ が完全でない場合、これと、完全な入力データ $u_d(t) = u(t)$ 、及び、モデル (1), (2) から完全な出力データ $y(t)$ を生成する前処理を行う。これにより、提案するパラメータ多様体の記述手法を実用可能にする。具体的には、何らかのパラメータ推定手法により得た、あるパラメータ a^* を用いて、状態空間モデルを数値計算し、出力 $y(t)$ を完全な出力データとみなす。なお、 a^* が、パラメータ多様体上の点として推定可能であれば、 $y(t; a^*)$ は完全な出力データに一致する。モデルの解析解が得られない場合、モデルを数値計算することになり、このとき、状態変数の初期条件を与えるか、これを未知パラメータとして推定する必要がある。本稿では、状態変数の初期条件は与えられているものと仮定する。なお、この仮定は、構造的同定可能性解析においても一般的である。

3.3 手順 (2), (3)

定理 2 に従い、イデアル I の生成系を定める。ここで、技術的には、微分演算が作用した変数を新たな変数として扱うことで、微分を含まない一般的な代数系として、グレブナー基底を計算する。グレブナー基底は、Maple, DAISY, Singular などのソフトウェアパッケージを用いて計算することができる。本稿では、Singular を用い、実行例などを後述する。なお、これらの手順は、モデルが非線形であっても適用可能である。

3.4 手順 (4)

手順 (3) で導出した入出力関係式に対し、完全な入力データと、手順 (1) で生成した完全な出力データ $y(t)$ とを与え、係数 $c_k(a)$ の値を推定する。以下、入出力関係式の定数項 C を除いた項の総数を K とし、具体的な推定方法を説明する。任意の K 時点におけるデータとその微分値、すなわち、 $y(k), \dots, \frac{d^N y}{dt^N} \Big|_{t=k}, u(k), \dots, \frac{d^{N-1} u}{dt^{N-1}} \Big|_{t=k}, (k = 1, \dots, K)$ を、それぞれ代入した入出力関係式は、次のような連立方程式をなす：

$$\begin{pmatrix} f_1(t=1) & \cdots & f_K(t=1) \\ \vdots & \ddots & \vdots \\ f_1(t=K) & \cdots & f_K(t=K) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_K \end{pmatrix} = \begin{pmatrix} C \\ \vdots \\ C \end{pmatrix}. \quad (7)$$

ここで、 $f_k(t=l)$ は、 l 時点におけるデータとその微分値が代入された k 番目の単項式の値である。式 (7) の左辺の行列の逆行列が存在すれば係数の値 $v_k, (k = 1, \dots, K)$ が求まり、 $c_k(a) = v_k, (k = 1, \dots, K)$ なる制約式の集合として、パラメータ多様体が記述できる。

4. ウイルスダイナミクスの解析への応用

4.1 扱うモデルとデータ

第 3 節に示した提案手法を、生体内における C 型肝炎ウイルス (以下、HCV) の、インターフェロン- α -2b (以下、IFN) 治療開始後 2 週間以内におけるダイナミクスを表すモデルに適用する。以下、扱うモデルとデータについて簡単に説明する。詳細については、[6] を参照されたい。

まず、生体内における HCV の、IFN 治療下におけるダイナミクスを表すモデルは次のように表される：

$$\frac{dx_1}{dt} = a_1 - a_2 x_1 - a_3 x_1 x_3, \quad (8)$$

$$\frac{dx_2}{dt} = a_3 x_1 x_3 - a_4 x_2, \quad (9)$$

$$\frac{dx_3}{dt} = (1 - a_5) a_6 x_2 - a_7 x_3, \quad (10)$$

$$y = x_3. \quad (11)$$

x_1, x_2, x_3 は時間 t に依存する変数、 y は観測されるモデルからの出力、 a_1, \dots, a_7 は時不変パラメータである。 x_1 は非感染細胞の数、 x_2 は感染肝細胞の数、 x_3 は血漿 1ml

中の HCV の RNA コピー数を表す。非感染細胞の生産率は a_1 、死亡率は a_2 、感染肝細胞の死亡率は a_4 であり、単位はいずれも [1/日] である。なお、非感染細胞は、感染率 a_3 [ml/(コピー × 日)] で感染肝細胞となる。感染肝細胞による HCV の生産率は a_6 である。IFN の薬効は、感染肝細胞による HCV 生産率を低下させることである。ここで、阻害率を $a_5, (0 \leq a_5 \leq 1)$ とすると、IFN 治療下における HCV 生産率は $(1 - a_5) a_6$ [コピー/(ml × 日)] と表される。また、HCV の除去率は a_7 [1/日] である。なお、式 (11) が表すように、実験においては、HCV 量が観測された。

ここで、治療開始後 2 週間以内は、非感染細胞数 x_1 が変化せず、 $x_1(0)$ は、 $a_5 = 0$ としたモデル (8)–(11) の定常状態の値 $\frac{a_4 a_7}{a_3 a_6}$ をとると仮定する： $x_1(0) = \frac{a_4 a_7}{a_3 a_6} = x_1(t)$ 。なお、 $a_5 = 0$ という仮定は、 $t < 0$ において、IFN による感染肝細胞への影響がないことを表す。治療開始後 2 週間以内におけるモデルは、次のように表される：

$$\frac{dx_2}{dt} = \frac{a_4 a_7}{a_6} x_3 - a_4 x_2, \quad (12)$$

$$\frac{dx_3}{dt} = (1 - a_5) a_6 x_2 - a_7 x_3, \quad (13)$$

$$y = x_3. \quad (14)$$

以下の解析では、モデル (12)–(14) と、観測データとしては、[6] の図 1A の黒丸で表されている、ウイルス量に関する完全でない時系列データを用いる。これは、IFN の投与量が異なる、3 人の被験者 (以下、1-H, 2-D, 3-D) からそれぞれ得られたデータである。ここで、モデル (12), (13) には入力を与えられていないため、 $u(t) = u_d(t) = 0$ である。また、IFN の投与後、HCV 量の減衰が始まるまでにかかる時間と、HCV 量の初期値 $x_3(0)$ が、被験者ごとに異なることが、[6] において明かにされている。具体的には、被験者 1-H, 2-D, 3-D に対して、それぞれ、遅延時間が 10[時間], 7[時間], 5[時間]、 $x_3(0)$ が 4.1×10^6 [コピー/ml], 1.0×10^6 [コピー/ml], 0.4×10^6 [コピー/ml] と解析されており、本稿でも、この解析結果を用いる。なお、微分方程式 (12)–(14) は線形モデルであるが、提案手法は一般の非線形モデルについても、適用可能であることに注意されたい。

4.2 パラメータ多様体の記述

以下、第 3 節に示した手順に従い、第 4.1 節で示したモデルとデータに対する、パラメータ多様体を記述する。

まず、手順 (1) であるが、ここでは、観測データと観測モデルの数値解の差を最小化する最小二乗問題を Levenberg-Marquardt アルゴリズムで解くことで、パラメータを推定し、図 1 に示すデータを生成した。このうち、HCV 量の減衰開始後のデータを表す観測モデル (14) の数値解を、完全なデータ $y(t)$ として扱う。

次に、手順 (2) として、モデル (12)–(14) に対するイデ

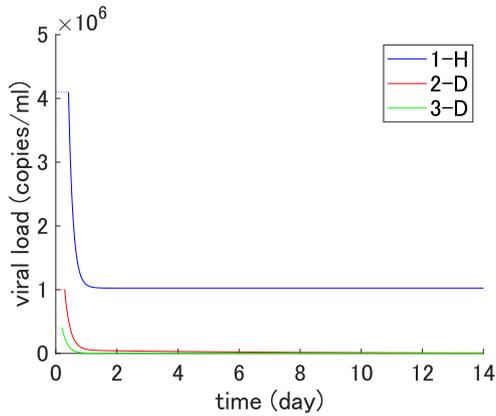


図 1 被験者 1-H, 2-D, 3-D それぞれに対して生成された、完全なデータ。各被験者らに対する、IFN の投与量は、それぞれ、 5×10^6 , 10×10^6 , 15×10^6 (IU) である。また、実線部分を、完全なデータ $y(t)$ として扱う。

Fig. 1 Perfect data that is generated for patients 1-H, 2-D and 3-D. For each patient, 5×10^6 , 10×10^6 , 15×10^6 (IU) of IFN was administrated respectively. We use the data that is shown in solid lines as perfect data $y(t)$.

アルを定める。ここで、 $N = 2$ であることに注意し、環 $\mathbb{Q}(a_4, a_5, a_6, a_7)[x_2, x_3, y, dx_2/dt, dx_3/dt, dy/dt, d^2x_2/dt^2, d^2x_3/dt^2, d^2y/dt^2]$ に対するイデアルを次のように定める：

$$I = \left\langle \frac{dx_2}{dt} - \frac{a_4 a_7}{a_6} x_3 + a_4 x_2, \frac{dx_3}{dt} - (1 - a_5) a_6 x_2 + a_7 x_3, \frac{d^2 x_2}{dt^2} - \frac{a_4 a_7}{a_6} \frac{dx_3}{dt} + a_4 \frac{dx_2}{dt}, \frac{d^2 x_3}{dt^2} - (1 - a_5) a_6 \frac{dx_2}{dt}, a_7 \frac{dy}{dt}, y - x_3, \frac{dy}{dt} - \frac{dx_3}{dt}, \frac{d^2 y}{dt^2} - \frac{d^2 x_3}{dt^2} \right\rangle. (15)$$

グレブナー基底の計算に向け、以下の変数を導入する：

$$\begin{pmatrix} X_1 & X_3 & X_5 \\ X_2 & X_4 & X_6 \\ Y_1 & Y_2 & Y_3 \end{pmatrix} := \begin{pmatrix} x_2 & \frac{dx_2}{dt} & \frac{d^2 x_2}{dt^2} \\ x_3 & \frac{dx_3}{dt} & \frac{d^2 x_3}{dt^2} \\ y & \frac{dy}{dt} & \frac{d^2 y}{dt^2} \end{pmatrix}.$$

これらの変数を用いて、イデアル I に対するグレブナー基底求める。Singular のプログラムは、次の通りである：

```
ring r=(0, a4, a5, a6, a7), (X6, X5, X4, X3, X2, X1, Y3, Y2, Y1), lp;
ideal I=X3-(a4*a7*X2)/a6+a4*X1, X4-(1-a5)*a6*X1+a7*X2, X5-(a4*a7*X4)/a6+a4*X3, X6-(1-a5)*a6*X3+a7*X4, Y1-X2, Y2-X4, Y3-X6;
stdhilb(I);
```

なお、lp は単項式順序として辞書式順序を与えることを表す。これにより、次に示すグレブナー基底を得る。

```
_ [1]=Y3+(a4+a7)*Y2+(a4*a5*a7)*Y1
_ [2]=(a5*a6-a6)*X1+Y2+(a7)*Y1
_ [3]=X2-Y1
_ [4]=(a6)*X3+(a4*a6)*X1+(-a4*a7)*Y1
_ [5]=X4-Y2
_ [6]=(a6)*X5+(-a4*a7)*X4+(a4*a6)*X3
_ [7]=X6-Y3
```

このうち、状態変数 X_1, \dots, X_6 を含まない 1 番目の元から、次に示す入出力関係式が得られる：

$$(a_4 a_5 a_7) y + (a_4 + a_7) \frac{dy}{dt} + \frac{d^2 y}{dt^2} = 0. \quad (16)$$

ここで、 $c_1(a) = a_4 a_5 a_7, c_2(a) = a_4 + a_7, C = -\frac{d^2 y}{dt^2}$ とおき、 $c_1(a), c_2(a)$ の値 v_1, v_2 を推定する。異なる任意の 2 時点 $t = t_1, t_2$ におけるデータとその微分値を、入出力関係式 (16) に与えると、次の連立方程式が得られる：

$$\begin{pmatrix} y(t_1) & \frac{dy}{dt} \Big|_{t=t_1} \\ y(t_2) & \frac{dy}{dt} \Big|_{t=t_2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = - \begin{pmatrix} \frac{d^2 y}{dt^2} \Big|_{t=t_1} \\ \frac{d^2 y}{dt^2} \Big|_{t=t_2} \end{pmatrix}.$$

左辺の行列が逆行列をもつとき、 v_1, v_2 が次のように求まる：

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = - \begin{pmatrix} y(t_1) & \frac{dy}{dt} \Big|_{t=t_1} \\ y(t_2) & \frac{dy}{dt} \Big|_{t=t_2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{d^2 y}{dt^2} \Big|_{t=t_1} \\ \frac{d^2 y}{dt^2} \Big|_{t=t_2} \end{pmatrix}. \quad (17)$$

以下、実際に、手順 (1) において生成したデータ $y(t)$ を用いて、各被験者に対する v_1, v_2 を求める。まず、乱数を用いて、被験者 1-H, 2-D, 3-D に対する t_1, t_2 を、それぞれ、 $(t_1, t_2) = (0.8183, 0.8631), (1.9529, 12.2390), (8.7774, 1.5303)$ [日] とした。次に、 t_1, t_2 における、データとその微分値 $y, \frac{dy}{dt}, \frac{d^2 y}{dt^2}$ の値を求める。ここで、生成したデータは数値解であるため、微分値を近似的に求める必要がある。今回は、中心差分により、例えば被験者 2-D について、 $y(t_1) \simeq 4.1136 \times 10^4, \frac{dy}{dt}(t_1) \simeq -0.6708 \times 10^4, \frac{d^2 y}{dt^2}(t_1) \simeq 1.5174 \times 10^4, y(t_2) \simeq 0.8616 \times 10^4, \frac{dy}{dt}(t_2) \simeq -0.1308 \times 10^4, \frac{d^2 y}{dt^2}(t_2) \simeq 0.0029 \times 10^4$ を得た。また、これらを式 (17) に代入すれば、 $(v_1, v_2) = (-4.9069, -32.3517)$ が得られる。被験者 1-H, 3-D についても同様の手法で、それぞれ、 $(v_1, v_2) = (-0.0007, -47.6106), (-15.2059, -38.5831)$ を得た。最後に、手順 (3) で求めた入出力関係式 (16) の係数 $c_1(a), c_2(a)$ より、 $a_4 a_5 a_7 = v_1, a_4 + a_7 = v_2$ を得る。ここで、 a_7 を消去すれば、

$$v_1 = a_4 a_5 (v_2 - a_4). \quad (18)$$

が得られ、これに、手順 (4) で求めた v_1, v_2 を、それぞれ与えると、パラメータ多様体得られる。このようにして得られた、被験者 2-D, 3-D に対するパラメータ多様体を、図 2 に示す。

4.3 パラメータ多様体を用いた解析

ここでは、第 4.2 節で得たパラメータ多様体を用いて、2 つの解析を行い、提案手法の有効性を確認する。

まず、被験者 1-H に対する v_1 の値が、他の被験者と比較して、極めて小さいことに着目し、式 (18) において、 v_1 を 0 とすると、被験者 1-H に対するパラメータ多様体とし

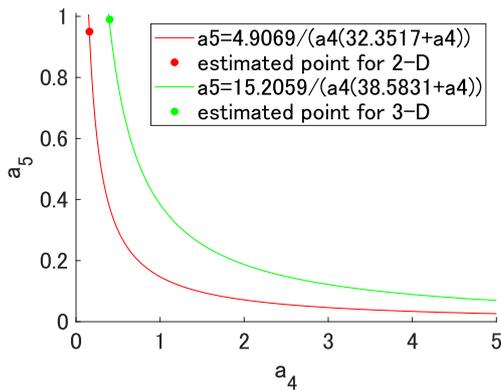


図 2 被験者 2-D, 3-D に対するデータ $y_d(t)$ とモデル (12)–(14) から得られた, パラメータ多様体. ここでは, 範囲を $0 \leq a_4 \leq 5, 0 \leq a_5 \leq 1$ に限定して図示した. なお, 図中の点は, それぞれ, 手順 (1) において推定されたパラメータを表す.

Fig. 2 The parameter manifold for patients 2-D and 3-D obtained from the model (12)–(14) and the corresponding data $y_d(t)$. They are drawn under the limitation such that $0 \leq a_4 \leq 5, 0 \leq a_5 \leq 1$. The dots in the figure represent the estimated parameters in the procedure (1).

て $a_4 a_5 (47.6106 + a_4) = 0$ が得られる. ここで, 第 4.1 節で述べたように, $a_4 \geq 0$ であるため, $47.6106 + a_4 \neq 0$ である. また, $0 \leq a_5 \leq 1$ であるから,

$$a_4 a_5 = 0 \Leftrightarrow \begin{cases} 0 \leq a_5 \leq 1, & (a_4 = 0), \\ a_5 = 0, & (a_4 > 0) \end{cases} \quad (19)$$

を得る. ここから, 被験者 1-H に対するパラメータ多様体 (19) と, 図 2 に示した被験者 2-D, 3-D に対するパラメータ多様体が, 構造的に異なることが分かる. また, [6] において, 被験者 1-H と, 被験者 2-D, 3-D に対して観測されたウイルスダイナミクスが, 大きく異なることが確認されており, IFN の投与量の違いによるものだと考察された. この考察は, 提案手法による解析結果に一致し, これにより, パラメータ多様体の幾何学的構造に着目することで, ダイナミクスの変化を捉えられうることが示唆された.

続いて, 被験者 1-H に対するパラメータ多様体 (19) について, モデリングの観点から考察する. $a_4 = 0$ は, 感染肝細胞が, 観測期間内に死亡しないことを表す. 一方で, $a_5 = 0$ は, IFN が機能しないことを表す. [6] において, 被験者 1-H については, 感染肝細胞の死亡率が極めて低い可能性があることが考察されているが, 一方で, 式 (19) からは, IFN が機能していない可能性もあることが分かる. 実際, [6] では, パラメータ推定によって得られた $(a_4, a_5) = (0, 0.75)$ について考察しているため, 前者の考察が自然に得られるが, このような考察は, パラメータ推定の詳細設定に依存する. これに対し, 提案手法は, パラメータ多様体の記述により, パラメータ推定では見落とされていた, 系のダイナミクスに関する新たな知見を得ることができた.

5. まとめ

本研究では, 系を表す状態空間モデルと, 時系列データから定まる情報を, パラメータ多様体として, 一意に記述する手法を提案した. パラメータ多様体の記述には, 入出力関係式を応用した. 入出力関係式の導出にあたり, 得られている入出力データが完全であることを仮定した, 構造的同定可能性解析の手法を応用した. ただし, 提案手法は, パラメータ多様体の記述の前処理として, 観測されたデータから完全な出力データを生成するため, 観測された出力データが完全でないような実問題にも, 適用可能である. 実際, 本稿では, 生体内ウイルスダイナミクス [6] について, 提案手法を用いて解析し, 既存手法では得られなかった知見を得た. 今後の課題としては, 状態変数の初期値が得られていない場合についても, 適用可能となるように手法を改善することが挙げられ, 具体的には, 状態変数の初期値を考慮した構造的同定可能性解析の技術を導入することを検討している.

謝辞 本研究は, JST さきがけ (JPMJPR16EC) に支援を受けている.

参考文献

- [1] Peters, S. A.: Physiologically-based pharmacokinetic(PBPK) modeling and simulations: principles, methods, and applications in the pharmaceutical industry, Wiley (2011).
- [2] Nelles, O.: Nonlinear system identification : From classical approaches to neural networks and fuzzy models, Springer (2001).
- [3] Meshkat, N., Sullivant, S.: Identifiable reparametrizations of linear compartment models, J. Symb. Comput., Vol. 63, pp. 46–67 (2014).
- [4] Ljung, L., Glad, T.: On global identifiability for arbitrary model parametrizations, Automatica, Vol. 30, pp. 265–276 (1994).
- [5] Bellman, R. and Astrom, K., J.: On structural Identifiability, Math Biosci., Vol. 7, pp. 7:329–339 (1970).
- [6] Neumann, A., U., Lam, N., P., Dahari, H., Gretch, D., R., Wiley, T. E., Layden, T., J. and Perelson, A., S.: Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon- α therapy, Science, Vol. 282, pp. 103–107 (1998).
- [7] Meshkat, N., Rosen, Z. and Sullivant, S.: Algebraic tools for the analysis of state space models, Adv. Stud. Pure Math., Vol. 77, pp. 171–205 (2018).
- [8] Xia, X. and Moog, C. H.: Identifiability of nonlinear systems with application to HIV/AIDS models, IEEE Transactions on Automatic Control, Vol. 48, pp. 330–336 (2003).
- [9] Miao, H., Xia, X., Perelson, A. S. and Wu, H.: On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics, SIAM Rev Soc Ind Appl Math., Vol. 53, pp. 3–39 (2011).
- [10] Ritt, J., F.: Differential Algebra, American Mathematical Society Colloquium Publications, Vol. XXXIII, (1950).