

大規模システムに関する 科学的連立方程式モデルの発見

Discovery of Scientific Simultaneous Equation Models of Large Scale Systems

鷲尾 隆* 元田 浩*
Takashi Washio Hiroshi Motoda

* 大阪大学産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University, Ibaraki 567-0047, Japan.

19YY年MM月DD日 受理

Keywords: scientific discovery, scientific law, simultaneous equations, minimal complete subset, large scale system

Summary

A novel and generic theory is formulated to characterize the structure of simultaneous equations. Based on the theory, an algorithm is developed to discover the structure of simultaneous equations governing an objective process through experiments, and the algorithm is implemented to the “*Simultaneous Structure Finder (SSF)*” program. SSF combined with another system SDS which discovers a quantitative formula of a complete equation derives a quantitative model of simultaneous equations reflecting scientific laws underlying the objective process. The power of the proposed approach comes from the use of “*minimal complete subsets*” contained in any simultaneous equation set. They can be identified through controlled experiments on the objective process. This novel definition nicely characterizes the structure of simultaneous equations, and provides a measure to efficiently prune unnecessary search to identify simultaneous structure of an equation set. In this paper, the theoretical foundations of the structure identification and the algorithm of SSF are presented, and its efficiency and practicality are demonstrated and discussed with large scale working examples. Since the conventional equation discovery systems could not find such simultaneous equations in an efficient manner, this work is expected to provide a new research field on the scientific equation discovery.

1. はじめに

データに内在する何らかの規則性を発見ないしは同定することを目指す人工知能の研究分野には、機械学習と並んで科学的法則発見がある。科学的法則発見の研究分野の草創期研究として最も代表的なものは、Langley 等による BACON システムに関するものである [Langley et al. 1987]。この研究を引き継ぐ形で、FAHRENHEIT [Koehn and Zytkow 1986]、ABACUS [Falkenhainer and Michalski 1986]、IDS [Nordhausen and Langley 1990] など、BACON で用いられた原理に基礎を置く、多くの科学的法則発見システムが提案された。これらは、何れも連続プロ

セスを対象とした実験の数値測定データから、それらを支配する 1 本の完全方程式^{*1}を探索するものである。しかしながら BACON 方式では、本当に対象系を支配する機構を反映した科学的法則式を導出する保証はなく、実験データに見かけ上フィットするだけの単なる実験式を導出してしまう可能性が高い。また、そのアルゴリズムは計算量爆発の問題を有し、高々数個程度の数量で表される規模の対象系しか扱うことができない。更に数量の測定ノイズや誤差に非常に敏感であり、十分な精度の法則式を同定することが困難である場合が多いことが報告されている [Schaffer 1990, Huang

*1 n 個の数量の自由度を 1 だけ制約して $n-1$ 自由度とする方程式を、完全方程式という。例えば $(x-r)^2+(y-r)^2=0$ は 3 変数の自由度を 2 つ制約するので、完全方程式ではない。

and Zytkow 1996] .

このような問題点を軽減するために、前述の ABACUS の拡張版や COPER [Kokar 1986] などは、数量の有する単位次元の知識に基づき次元解析の手法を適用し、数理的にあり得ない項の探索枝刈りを効率的に行う手法を採用した。法則式に表れる数量は単なる純粋数ではなく、一定の測定過程を通じて得られたものである。そのため、数量は測定量として単位次元の情報を持ち、この単位次元に照らして数理的に許容される形式を有する方程式のみが法則式の候補となる。従って、単なる実験式ではなく科学的法則式である可能性が極めて高い方程式を発見可能である。しかしながら、これらは物理学など各数量の単位次元に関する知識が明確である問題領域にしか適用できないという、強い制限を有する。

これに対し、筆者等は新しい原理に基づく“*Smart Discovery System* (SDS)”という、定量的な科学的法則式発見システムの開発を行った [Washio and Motoda 1997]。このシステムは、各数量が有するスケールタイプの知識に基づいて、数量間に数学的に許容される定量的関係を強く絞り込むスケールタイプ制約と、測定データから得られた数量間の部分的な関係から、数学的に許容される全体関係を絞り込む恒等制約を用いる。そして、これら2つの制約を用いて非常に効率的に、法則式として数理的にあり得ない数量関係の探索の枝狩りを行う。単位次元と異なり、各数量のスケールタイプに関する知識は非常に幅広い問題領域で得られるため、SDS は実験心理学や経済学、社会学など、多くの非物理的領域に適用可能である。このような数学的制約の導入により必要とされる計算量は少なく抑えられる。そのため、SDS は1つの完全方程式に数十個程度の数量が含まれるような複雑な対象系に関し、法則式やそれに基づくモデル式である可能性が極めて高い完全方程式を得ることが可能である。更にこれらの制約に適切な統計的データ処理を組み合わせることによって、数量データに含まれる測定ノイズや誤差に対してロバストな、方程式導出が行える。

以上一連の研究成果にもかかわらず、これまでに発見対象となっている科学的法則式やモデル式は、1つの完全方程式で表されるものに限定されている。しかしながら実際的な大規模問題では、複数の機構が作用する系を対象とする場合が殆どである。このような系は多くの場合、複数の完全方程式で表されるため、上記に述べた手法では扱うことができない。過去に開発された幾つかの法則式発見システムには、この点について一定の配慮を行うものが存在する。水は固体、液体、気体のそ

れぞれの相で、全く異なった状態方程式に従う。このように対象系が異なる完全方程式によって支配される幾つかのモードを有する場合に、前述の FAHRENHEIT や ABACUS は、実験データから対象系の各モードとそれらの間の遷移条件を見だし、更に各モードを支配するそれぞれ1つの完全方程式の発見を行う手法を備えている。実際問題において、対象系が複数のモードを有する場合は多く、この手法は法則式発見の範囲拡大に貢献する。しかしながら、対象系が複数の方程式で表現されるより一般的な場合は、複数の機構が同時に作用して連立方程式系を構成するものである。特に殆どの大規模な対象系は連立方程式で表現される。科学的連立方程式モデルを発見することは、対象系で作用している複数の機構間の依存関係を把握するためにも欠くことができない [Iwasaki and Simon 1986, Murota 1987]。過去にもデータから連立微分方程式系を発見する能力を有する LAGRANGE というシステムが開発されている [Dzeroski and Todorovski 1994]。しかしながら、これは与えられたデータに対して、基本的に種々の連立方程式の生成とその当てはめ検証を繰り返す探索方式を採るため、膨大な計算量を要しかつ単なる実験式を含む多数の解候補及びそれら各々の冗長な表現形式を与えてしまうという問題を有する。

本研究では、科学的法則式やモデル式である可能性の極めて高い方程式を、効率的に導出する SDS の能力を備えつつ、大規模な対象系について実験データから、必要な数学的許容条件を満たす科学的連立方程式モデルを発見することが可能な手法の提案を目的とする。ただし、ここでは対象系やその実験条件について3つの条件を仮定する。1つは、対象系は我々が着目する状態の範囲内において、定量的かつ連続な数量のみを含む幾つかの完全方程式から成る制約不足の連立方程式で表されるという仮定である。操作実験が可能な実際の対象系は、連立方程式中に含まれる数量が制約不足であるために、種々の数量操作が可能である。実験系では制約充足や制約過剰な系は対象となり得ない。2つ目は、各方程式に表れる全ての数量を何らかの手段で観測可能なことである。3つ目は、各方程式において他の方程式の制約を満たしつつ、1つの従属な数量を除いて全ての数量を実験的に適当な値に制御する何らかの手段が存在することである。これらは、実験環境でのデータ採集を前提とする従来の BACON 系の手法に共通した仮定である。このような問題設定の下で以下を検討し、定式化並びに評価を実施した。

- (1) 数量間の制約に関する不変構造の観点から、制約不足な連立方程式を数学的に特徴付ける。

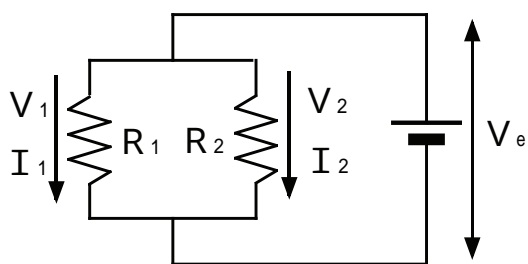


図1 並列抵抗の電子回路

- (2) 対象系を操作する実験を通じて，上記不変構造を導出するアルゴリズムを定式化する．
- (3) 従来の科学的法則式発見手法を連立方程式中の各式の分離同定に用いる原理とアルゴリズムを定式化する．
- (4) 種々の例題を通じ，提案手法の性能や適用性を評価する．

我々は(1)から(3)に至る研究によって得られたアルゴリズムと理論に基づいて，“*Simultaneous Structure Finder (SSF)*”と名付けたツールプログラムを開発した．そして(4)において，前述のように優れた特徴を有するSDSとこのSSFを組み合わせるにより，その性能と適用性を評価した．

2. 連立方程式発見の基本原理の検討

2.1 問題提起

最初に，制約不足の連立方程式に関するある簡単な例を通じて，連立方程式が有する数量間の制約構造に関する問題を提起する．この問題提起は連立方程式が有する不変構造を導く重要な糸口を与える．図1は2本の抵抗と1つの電池からなる電子回路である．これは以下の4本の式からなる連立方程式によってモデル化できる．

$$\begin{aligned} V_1 &= I_1 R_1 [1], V_2 = I_2 R_2 [2], \\ V_e &= V_1 [3], V_e = V_2 [4], \end{aligned} \quad (1)$$

ここで R_1, R_2 : 2つの抵抗値,

V_1, V_2 : 抵抗の両端電圧,

I_1, I_2 : 抵抗を流れる電流,

V_e : 電池の供給電圧

上記と異なる形式で以下のモデルを与えることも可能である．

$$\begin{aligned} I_1 R_1 &= I_2 R_2 [1], V_2 = I_2 R_2 [2], \\ V_e &= V_2 [3], V_e = V_2 [4] \end{aligned} \quad (2)$$

両モデルとも，定量的に正しい回路動作挙動を与えることができる．しかしながら，直感的に見て前者の方がより自然で理解し易いように思える．この差異は，両モデルにおける数量配置が異なることによると考えられる．連立方程式における数量の配置を体系的に扱うために，ここでは“接続行列”を用いることにする．接続行列 T の各行は連立方程式の各式に対応し，各列は各数量に対応する．もし， j -番目の数量が i -番目の方程式に現れるならば， T の (i, j) 要素である T_{ij} を1と置き，そうでない場合には0と置くことにする [Murota 1987]．以下の2つの T_1, T_2 が，それぞれ式(1)及び式(2)に関する接続行列である．

$$T_1 = \begin{bmatrix} V_e & V_1 & V_2 & I_1 & I_2 & R_1 & R_2 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

制約不足の連立方程式を対象とするため， M 行 N 列の接続行列は $M < N$ を満たす．ここで外生的に幾つかの数量を制御操作することによって，連立方程式中で $n (\leq N)$ 個の未定数量を含む同じく $n (\leq M)$ 本の方程式を得ることができれば，その n 本の部分連立方程式を解いて n 個の数量の値が決定可能となる．外生的に1つの数量を制御操作することは，その数量値が既知となって方程式の未定数量から除かれるために，上記の接続行列上で言えば対応する列を行列から除去することに相当する．従って，制約不足の連立方程式の一部が数量の外生的制御操作によって解けるようになるということは，接続行列上では列の幾つかを除去することによって， n 個の行が同じく n 個の非零列を持つようにできることに対応する．この時，これら n 個の数量の値は部分連立方程式を解くことで“決定”する．本論文では，そのような n 本の部分連立方程式を次数 n の“完全部分集合”と呼ぶことにする．電子回路の最初のモデル式(1)において，もし V_e と R_1 の値を外生的に制御操作すれば， T_1 の1行目と3, 4行目が V_1, V_2, I_1 の3つの非零列のみを含むようになる．従って，式集合 $\{[1], [3], [4]\}$ は3次の完全部分集合である．この時，3数量 V_1, V_2, I_1 の値が決定するのに対し，ここからは I_2 と R_2 の値は決まらない．一方で，

後者のモデル式 (2) において同様に V_e と R_1 の値を外生的に制御操作しても, T_2 の行の何れの組み合わせも 3 つ以上の非零列を有し, 次数 3 の完全部分集合は得られない. しかしながら, 実際の図 1 の電子回路に V_e と R_1 の外生的制御操作を施すならば, 最初のモデルによる数量の決定結果が正しいことは明かである. 実際, 前者のモデルは数量の如何なる外生的制御操作に関しても, 実際の系と同じ数量決定結果を与える. 即ち数量の制約構造に関して, 前者のモデルは実際の系と“同型”である. これに対して, 後者のモデルは幾つかの外生的制御パターンに対して実際とは異なった数量決定結果を与える. 前者のように接続行列が実際の対象系と同型な制約構造を有するモデルを, 本論文では“構造形”と呼ぶことにする.

2・2 連立方程式の制約構造の特徴付け

本節では上述の問題提起を受けて, 制約不足の連立方程式が有する制約構造に関し, より精密な数学的定式化と特徴付けを与える. まず定式化の基礎として, 幾つかの基本的定義を導入する.

【定義 2・1】(接続行列) 互いに代数独立な複数方程式の集合 $E = \{eq_i | i = 1, \dots, M\}$ が与えられ, それらが数量集合 $Q = \{q_j | j = 1, \dots, N\}$ を含むとする. ある行列 T の (i, j) 要素を T_{ij} とした時, $q_j (\in Q)$ が $eq_i (\in E)$ に現れるならば $T_{ij} = 1$, そうでなければ $T_{ij} = 0$ とする行列 T を“接続行列”という. □

【定義 2・2】(完全部分集合) T を接続行列とし, それに関して $CE (\subseteq E)$ 及び $RQ (\subseteq Q)$ とする. T から RQ に属する数量に対応する列を全て除去した時, その CE に属する式に対応する行に関する部分接続行列を $T[CE, Q - RQ]$ と表す. 更に, $T[CE, Q - RQ]$ において少なくとも 1 つ以上の要素が非零である列 (非零列) の集合を $NQ (\subseteq Q - RQ)$ とする. もし $|CE| = |NQ| = n$ であるならば, CE を次数 n の“完全部分集合”という. ここで, $|\cdot|$ は集合の濃度を表す. □

これらの定義に基づき, 完全部分集合に関する幾つかの数学的性質を導くことができる.

[定理 2・1] (対称性定理) 接続行列 T においてある列集合 RQ ($|RQ| = m$) を除いた時, 次数 n の完全部分集合 CE が得られるものとする. 更に $T[CE, Q]$ を CE 内の方程式と Q 内の数量に関する部分接続行列とする. ここで $T[CE, Q]$ の非零列の集合を CQ とする. この時, CQ において濃度が m である任意の部分集合 RQ_i (ただし $i = 1, \dots, \binom{n+m}{m}$) を除いても, 同様に次数 n の完全部分集合 CE が得られる. □

証明. 定義 2.2 の NQ は部分接続行列 $T[CE, Q - RQ]$ の非零列集合であるので, $CQ = RQ + NQ$ である. 次数 n の完全部分集合の定義から $|RQ| = m, |NQ| = n$ であるので, $|CQ| = m + n$ である. $|RQ_i| = m$ ($i = 1, \dots, \binom{n+m}{m}$) であるような任意の列集合 $RQ_i (\subseteq CQ)$ の除去について, その残りの列集合 $NQ_i = CQ - RQ_i$ は濃度 $(m + n) - m$, 即ち $|NQ_i| = n$ を有する. 従って, 任意の RQ_i の除去について, CE は次数 n の完全部分集合である. ■

[定理 2・2] (不変性定理) 方程式の全体集合 U_E 上の変換 $f: U_E \rightarrow U_E$ を考える. CE が T における次数 n の完全部分集合であるとする. もし, 変換 $f(CE)$ が方程式の数と非零列構造を保存する, 即ち $|f(CE)| = |CE|$ 及び CQ_f を $T[f(CE)]$ の非零列集合であるとした場合 $CQ = CQ_f$ であるならば, $f(CE)$ も次数 n の完全部分集合である. □

証明. $CQ = CQ_f$ であるため, RQ を $T[CE]$ 及び $T[f(CE)]$ の何れにおいて削除しても, 同一の非零列集合 NQ が得られる. CE が次数 n の完全部分集合である時 $|CE| = n$ であるので, もし $|CQ| = |CQ_f| = m + n$ ならば RQ を $|RQ| = m$ と選ぶことによって, $|NQ|$ を n とすることができる. そのようなある RQ の下では, $|CE| = |f(CE)| = |NQ| = n$ であり, $f(CE)$ は次数 n の完全部分集合の条件を満たす. ■

注意 2・1 “対称性定理”は, 各完全部分集合に含まれる数量のうち必要数 m 個を外生的に制御する限り, 実験上の制御数量の選び方に依らずに, 即ち探索アルゴリズムとは独立に, その同一の完全部分集合を同定可能であることを示している. 完全部分集合のこの性質に基づけば, 効率的で完全かつ健全な探索アルゴリズムを得ることが可能である. □

注意 2・2 先に 1 つの方程式においては直接に制御できない従属数量が 1 つ存在することを許すという仮定を置いたが, “対称性定理”はこの条件の下でも実験を通じて各完全部分集合と定量的に正しい各方程式形を正しく導出する理論的保証を与える. もし, ある完全部分集合に 1 つの従属数量が存在するならば, 他の制御可能な数量を適切に制御することで, その完全部分集合を制約しかつ従属数量を仮想的に制御したのと同様な状態を実現することができる. □

以下の注は連立方程式で表される定量的モデルの発見を考える際に非常に重要である.

注意 2・3 ある対象系のモデルを考える時“不変性定理”が示す変換 f の内容を, 更に方程式間の代入や定量的算術操作に限ることで, 数量間の定量的関係と依存構造の両面で等価な様々な連立方程式系を得るこ

とができる． □

前述の電子回路の例では，連立方程式 (1) において V_e の値が外生的に指定される，即ち T_1 の第 1 列が除去されれば， T_1 の第 3 及び第 4 行が 2 つの非零列のみを含むことになる．従って，連立方程式 (1) の方程式集合 $\{V_e = V_1[3], V_e = V_2[4]\}$ は次数 2 の完全部分集合であることがわかる．これらの方程式は，線形変換によって定量的等価性を保ったまま以下の形式に変換可能である．

$$V_e = 2V_1 - V_2 [3], V_e = -V_1 + 2V_2 [4] \quad (3)$$

この新たな 3 番目のモデルについては，接続行列は以下ようになる．

$$T_3 = \begin{matrix} & V_e & V_1 & V_2 & I_1 & I_2 & R_1 & R_2 \\ \begin{matrix} 0 \\ 0 \\ 1 \\ 1 \end{matrix} & \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

T_1 の場合と同様に V_e に対応する第 1 列を除去することによって， T_3 の第 3 及び第 4 行が 2 つの非零列のみを含むことになり，これらの 2 行からなる 2 次の完全部分集合は，この新たなモデルにおいても保持されていることがわかる．

変換 f に関して等価な連立方程式を多数枚挙するような探索アルゴリズムを採用すれば，探索の組み合わせ爆発に直面してしまう．これを避けるために，以下で定義される特定の形式を有する連立方程式のみを同定することを考える．

【定義 2・3】(完全部分集合の正準形) 次数 n の完全部分集合 CE について，その接続行列 $T[CE, Q]$ 内の全ての非零列集合 CQ の要素が 1 である表現形式を， CE の“正準形”という． □

完全部分集合の正準形の一例が連立方程式 (3) である．正準形以外の形式の連立方程式で定量的及び構造的に等価な表現は，変換 f によって全て導出可能であるため，連立方程式モデルの発見という目的には正準形のみを求めれば十分である．

連立方程式で表される系に関して，各完全部分集合は数量の値を決定するまとまった基礎的機構を表すと考えられる．しかしながら，多くの場合に幾つかの完全部分集合は互いに独立ではない．例えば，以下の 4 つの完全部分集合が連立方程式 (1) には存在する．

$$\begin{aligned} & \{[3], [4]\} \quad (n = 2), \\ & \{[1], [3], [4]\} \quad (n = 3), \end{aligned} \quad (4)$$

$$\{[2], [3], [4]\} \quad (n = 3),$$

$$\{[1], [2], [3], [4]\} \quad (n = 4)$$

[] 内の番号は各方程式を， n は各完全部分集合の次数を表す．これから分かるように，完全部分集合は相互に多くの重なりを持っており，より高次の完全部分集合はそれより低次の完全部分集合と重複した機構を表している場合がある．以下は完全部分集合間の相互関係に関する数学的特徴に関する定理である．

[定理 2・3](束定理) 方程式集合 E から成る連立方程式で表される対象系モデルについて，その完全部分集合の集合 $L = \{CE_i | \forall CE_i \subseteq E\}$ は束をなす．即ち， $\forall CE_i, CE_j \in L$ について， $CE_i \cup CE_j \in L$ かつ $CE_i \cap CE_j \in L$ である． □

証明. 2 つの完全部分集合 $CE_i, CE_j \subseteq E$ について，各々の方程式に含まれる数量の集合をそれぞれ CQ_i, CQ_j とする．

(i) $CE_i \cap CE_j = \phi$ の場合には， CQ_i と CQ_j に含まれる RQ_i と RQ_j 内の数量の値をそれぞれ統制することで， CE_i と CE_j 内の他の数量の値は各々独立に決定される．従って，もし $CQ_i \cup CQ_j$ 内の適当な数量の組み合わせ $RQ_i \cup RQ_j$ について外生的に値を指定すれば， $CQ_i \cup CQ_j$ 内の全数量の値が $CE_i \cup CE_j$ を通じて決定される．よって $CE_i \cup CE_j \in L$ ， $CE_i \cap CE_j = \phi \in L$ は自明である．

(ii) $CE_i \cap CE_j \neq \phi$ かつ $CE_i \not\subseteq CE_j, CE_j \not\subseteq CE_i$ の場合には，以下の 2 条件が成立する．

$$|CQ_i - CQ_i \cap CQ_j| > |CE_i - CE_i \cap CE_j|$$

$$|CQ_j - CQ_i \cap CQ_j| > |CE_j - CE_i \cap CE_j|$$

もし前者の条件が成立しなければ，完全部分集合 CE_j によって CQ_j 内の数量の値が決定された場合， $CQ_i - CQ_i \cap CQ_j$ 内の数量は制約充足ないしは制約過剰になってしまう．このことは CE_i が独立にその内部の数量の値を決定し得る完全部分集合であることと矛盾する．同様の議論は後者の条件にも当てはまる．この事は，方程式集合 $CE_i \cap CE_j$ は他の完全部分集合とは独立に $CQ_i \cap CQ_j$ 内の数量の値を決定可能であることを意味する．従って， $CE_i \cap CE_j \in L$ である．更に，適当な数量を $CQ_i \cap CQ_j$ 及び $CQ_i - CQ_i \cap CQ_j, CQ_j - CQ_i \cap CQ_j$ から選んで外生的に値を指定すると， $CQ_i \cup CQ_j$ 内の全ての数量の値が決定される．従って， $CE_i \cup CE_j \in L$ である．

(iii) $CE_i \subseteq CE_j$ または $CE_j \subseteq CE_i$ の場合は， $CE_i \cap CE_j \in L$ かつ $CE_i \cup CE_j \in L$ は自明で

ある .

[定理 2・4](モジュラ束定理) 方程式集合 E から成る連立方程式で表される対象系モデルについて, モデルの完全部分集合の集合 $L = \{CE_i | \forall CE_i \subseteq E\}$ は, 完全部分集合の回数に関してモジュラ束をなす. 即ち, $n(CE)$ を完全部分集合 CE の回数とした時, $n(CE_i \cup CE_j) = n(CE_i) + n(CE_j) - n(CE_i \cap CE_j)$ である. □

証明. 定義より完全部分集合の回数はその濃度に等しい. $|CE_i \cup CE_j| = |CE_i| + |CE_j| - |CE_i \cap CE_j|$ の関係から, 定理に示される回数間関係は明かである. ■

完全部分集合間のモジュラ束構造に基づいて, 各完全部分集合に固有の独立成分とその回数を以下のように定義する.

【定義 2・4】(完全部分集合の独立成分) 完全部分集合 CE_i の独立成分 DE_i を以下のように定義する.

$$DE_i = CE_i - \bigcup_{\substack{\forall CE_j | CE_j \subseteq CE_i \\ \text{and } CE_j \in L}} CE_j$$

CE_i に含まれる他の完全部分集合には属さない CE_i 中の固有の数量集合 DQ_i も以下のように定義する.

$$DQ_i = CQ_i - \bigcup_{\substack{\forall CQ_j | CQ_j \subseteq CQ_i \text{ where} \\ CQ_j \text{ is a set of quantities} \\ \text{appearing in } CE_j \in L}} CQ_j$$

また, DE_i の回数 δn_i 及び自由度 δm_i を以下のように定義する.

$$\delta n_i = |DE_i|, \quad \delta m_i = |DQ_i| - |DE_i|$$

□

注意 2・4 ある連立方程式系における回数間の依存構造の下で, 完全部分集合の“独立成分”は回数の値の決定に関し他の完全部分集合とは独立した機構を表す. 独立成分 DE_i 内に現れる回数の値は, 他の方程式制約に抵触することなしに $(|CQ_i| - \delta n_i)$ の自由度を持って変更可能である. □

方程式 (4) の例では, 3 つの独立成分が導かれる.

$$\begin{aligned} DE_1 &= \{[3], [4]\} - \phi = \{[3], [4]\}, \\ \delta n_1 &= 2 - 0 = 2, \\ DE_2 &= \{[1], [3], [4]\} - \{[3], [4]\} = \{[1]\}, (5) \\ \delta n_2 &= 3 - 2 = 1, \\ DE_3 &= \{[2], [3], [4]\} - \{[3], [4]\} = \{[2]\}, \\ \delta n_3 &= 3 - 2 = 1 \end{aligned}$$

モジュラ束の集合包含関係は単調性を持つため, 小さい完全部分集合から *greedy* に探索しても, 余り多くの計算量を必要とせず全完全部分集合を導くこと

ができる.

各独立成分 DE_i は完全部分集合 CE_i の部分集合であるため, DE_i の非零構造も同様に不変性定理に従う. 従って, CE_i の正準形の中で DE_i に対応する部分によって, 同じく DE_i の正準形を定めることができる. この考えに基づいて, 対象系を表す連立方程式系の正準形を以下のように定義する.

【定義 2・5】(連立方程式の正準形) DE_i をその完全部分集合 CE_i の正準形に対応する部分によって表す時, 連立方程式系の“正準形”を各独立成分の和集合 $\bigcup_{i=1}^b DE_i$ によって定める. ここで b は連立方程式系に含まれる DE_i の全数である. □

電子回路の例における連立方程式系の正準形は, 式 (5) に基づき

$$T_4 = \begin{matrix} & V_c & V_1 & V_2 & I_1 & I_2 & R_1 & R_2 \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

となる.

もし実験を通じて対象系における回数間の実際の依存関係を反映するように, 連立方程式を導出するならばそれは“構造形”となる. 更にそれを正準形として導くことを考え, 以下の定義を導入する.

【定義 2・6】(構造正準形) もし連立方程式の正準形が“構造形”となるように導出されるならば, それを“構造正準形”と呼ぶ. □

例えば, 構造形 T_1 から得られた正準形の接続行列 T_4 は, 構造正準形である.

3. SSF のアルゴリズムと実装

回数の測定と制御可能性に関する前述の仮定の下で, 小さい完全部分集合から *greedy* に探索する図 2 のアルゴリズムを開発し, SSF に実装した. SSF には, 対象系をモデル化するために必要な回数のリストとそれらの測定値を入力する. 本アルゴリズムでは (S2) において, 少ない濃度を有する回数集合の制御から始めて, 完全に値が被制御になる回数が生じないかどうかを実験 (*experiment*) によって調べる. この実験では図 3 に示されるアルゴリズムによって, 被制御回数が決定される. この (S2-3), (S2-4) において, 各相関係数 r が十分有意 (*significant*) な相関を示すか否かは, 統計的 t -検定により判定される. 今, データ数 d から得られた相関係数 r が無相関を示すという帰無仮説の下で,

$$t = \frac{r\sqrt{d-2}}{\sqrt{1-r^2}},$$

は、自由度 $d-2$ の t -分布に従うことが知られている。これから逆に、有意水準 $\alpha (= 0.05)$ の下での自由度 $d-2$ の t -分布の両側検定区間の片幅 $t(d-2, \alpha)$ から、対応する

$$r(d-2, \alpha) = \frac{t(d-2, \alpha)}{\sqrt{d-2 + \{t(d-2, \alpha)\}^2}},$$

を求め、

$$|r| \geq r(d-2, \alpha)$$

であれば2つの数量は有意な相関を有すると見なし、そうでなければ無相関であると見なす。このようにしてもし被制御数量が見つければ、図2の(S3)において制御数量集合とそれに対応する被制御数量集合の和集合を CQ_i とする。 CQ_i は新たに見つかった完全部分集合 CE_i が含む数量の集合である。そして定義2・4に従って、その $DE_i, DQ_i, \delta n_i, \delta m_i$ が求められ記録される。一度、独立成分集合が見つければ、それまでに見つかった各 DQ_i から δm_i 以内の個数分選ばれた数量と、何れの DQ_i にもまだ属さないものの中から選ばれた幾つかの数量が制御に用いられ、新たな完全部分集合が探索される。制御数量以外の全ての数量が被制御になるまで探索は続けられる。このアルゴリズムの計算量は非多項式オーダーであるが、 DQ_i による上述の探索制約によって、必要計算量は効率的に減少する。 DQ_i による探索制約を用いても、完全部分集合相互の束構造が有する単調性のため、完全部分集合の探索漏れはあり得ない。

過去に開発されてきた1本の完全方程式を発見する科学的法則式発見システムは、何れも上記SSFが導く構造正準形の情報を直接に受け取ることはできない。連立方程式系の定量的形式を得るためには、問題を各方程式を個別に求める副問題に分割せねばならない。そこで、連立方程式系の構造正準形の情報に基づいて全体の定量的形式を求める問題をそのような小さな副問題に分解するアルゴリズムも開発し、SSFに実装した。SDSを含め従来の科学的法則式発見システムは、対象系を表す数量のうちある2個以外の値を固定してその2数量の間の関係を実験で同定する“2数量テスト”と呼ばれる方法を積み上げ、最終的に1本の完全方程式全体の形式を得る。ところで、定義2・4に示されるように、各完全部分集合の独立成分に含まれる数量の値は δn_i 次元の分、相互に制約されている。従ってもし次元 δn_i が1以上であれば、独立成分内の制約によって構造正準形に示される方程式に現れる数量間

- (S1) Let $Q = \{q_k | k = 1, \dots, N\}$ be a set of quantities that appear in the model of an objective process. Set $X = \{x_k | x_k = q_k, \text{ for all directly controllable } q_k \in Q\}$, $CQ = \phi$, $DQ = \phi$, $N = \phi$, $M = \phi$, $h = 1$ and $i = 1$.
- (S2) Choose $C_j \subset DQ_j \in DQ$ for some DQ_j and also $C_x \subseteq X$, and take their union $C_{hi} = \dots \cup C_j \cup \dots \cup C_x$, while maintaining $|C_j| \leq \delta m_j$ and $|C_{hi}| = h$. Control all $x_k \in C_{hi}, k = 1, \dots, |C_{hi}|$ in experiment. Let a set of all quantities whose values are determined in the experiment be $D_{hi} (\subseteq Q - C_{hi})$.
- (S3) Set $CQ_{hi} = C_{hi} + D_{hi}$,
 $DQ_{hi} = CQ_{hi} - \cup_{\substack{CQ_{h'i'} \subset CQ_{hi} \\ CQ_{h'i'} \in CQ}} CQ_{h'i'}$,
 $\delta n_{hi} = |D_{hi}| - \sum_{\substack{\forall \delta n_{h'i'} \in N \\ \text{where } CQ_{h'i'} \subset CQ_{hi} \\ \text{and } CQ_{h'i'} \in CQ}} \delta n_{h'i'}$,
and $\delta m_{hi} = |DQ_{hi}| - \delta n_{hi}$.
If $\delta n_{hi} > 0$, then add CQ_{hi} to the list CQ , DQ_{hi} to the list DQ , δn_{hi} to the list N , δm_{hi} to the list M and $X = X - DQ_{hi}$.
- (S4) If all quantities are determined, i.e., $D_{hi} = Q - C_{hi}$, then go to (S5), else if any more C_{hi} where $|C_{hi}| = h$ does not exist, $h = h + 1, i = 1$ and go to (S2), else $i = i + 1$ and go to (S2).
- (S5) $DE = \{DE_i | \forall CQ_i \in CQ, DE_i \text{ is a set of } \delta n_i \text{ equations in which only all quantities in } CQ_i \text{ appear}\}$. The contents of the lists DE, DQ, N and M represent the independent components of equations, those of quantities, their orders and their degree of freedom respectively.

図2 構造正準形の探索アルゴリズム

- (S2-1) Let $R_{hi} = Q - C_{hi}$ and $p = 1$.
- (S2-2) For every $x_k \in C_{hi}$, its value is set to a random value $V_p(x_k)$ respectively.
- (S2-3) For $q = 1$ to n {
Try to exogenously control every $x_r \in R_{hi}$ to a random value $V_{cpq}(x_r)$ respectively, and let the actually observed value of x_r be $V_{opq}(x_r)$.}
- (S2-4) For every pair of $x_s, x_t \in R_{hi}$, calculate the correlation coefficient $r_{oc}(x_s, x_t)$ from $V_{opq}(x_s), V_{cpq}(x_t), q = 1, \dots, n$. Let $D_{hip} (\subseteq R_{hi})$ be a set of x_s not showing significant correlations with any x_t , and calculate the mean value $\bar{V}_{op}(x_s) = \frac{1}{n} \sum_{q=1}^n V_{opq}(x_s)$ for such x_s . Let $p = p + 1$. If $p \leq m$, then go to (S2-2).
- (S2-5) Let $D'_{hi} = \cap_{p=1}^m D_{hip}$ where the quantities have been judged not to show any significant correlation over the m tests. Calculate the correlation coefficient $r(x_s, x_k)$ between $\bar{V}_{op}(x_s)$ of $x_s \in D'_{hi}$ and $V_p(x_k)$ of $x_k \in C_{hi}$ over $p = 1, \dots, m$. Let a set of x_s showing significant correlation with any $x_k \in D_{hi}$.

図3 被制御数量決定アルゴリズム

で、任意の2数量に関する2数量テストを行うことは困難に見える。しかしながら、実際には独立成分 DE_i 内の各方程式について、互いに他の $(\delta n_i - 1)$ 個の方程式を用いた数量の代入消去を行うことによって、各

- (S1) Let DE , DQ and N be the lists obtained in the algorithm of Fig.2. Set $DQ' = \phi$.
- (S2) For $i = 1$ to $|DE|$ {
 For $j = 1$ to $|DE|$ where $j \neq i$ {
 If $DE_i \supset DE_j$ where $DE_i, DE_j \in DE$ {
 $DQ'_i = DQ'_i - NQ_j$,
 where NQ_j is arbitrary, and
 $NQ_j \subset DQ_j \in DQ$ and $|NQ_j| = N_j$.
 Add DQ'_i to the list DQ' .}}}
- (S3) $DQ = DQ'$ and $DQ' = \phi$.
 For $i = 1$ to $|DE|$ {
 For $j = 1$ to N_i {
 $DQ'_{ij} = DQ_i - DQ_{ij}$,
 where DQ_{ij} is arbitrary, but
 $\forall DQ'_h \in DQ'$, $DQ_{ij} \neq DQ_i - DQ'_h$,
 $DQ_{ij} \subset DQ_i \in DQ$ and $|DQ_{ij}| = N_i - 1$.
 Add DQ'_{ij} to the list DQ' .}}}
- (S4) Every $DQ_{ij} \in DQ'$ shows the list of quantities contained in a transformed equation. $DE' = \{DE_h | \forall DQ'_h \in DQ', DE_h \text{ is an equation in which only all quantities in } DQ'_h \text{ appear}\}$. DE' represents the simultaneous equation where the number of quantities is minimized in each equation.

図4 数量個数最小化アルゴリズム

方程式に含まれる数量の個数を減らすことが可能である。これにより DE_i 内の各方程式を、個別にあたかも1つの完全方程式であるように扱って数量の制御実験を行い、各々の定量的形式を求めることが可能である。代入消去によるこの数量個数の削減は、2数量テストに基づく方程式発見を可能にする上で必須である。また、各方程式の形式を同定する上で、その方程式に含まれる数量の個数が必要計算量に大きな影響を与えるため、計算量削減の面からも利点を有する。更に、同定対象とする1本の方程式内に含まれる数量の個数が少ない方が、定量的な測定データ処理における観測ノイズやフィッティング数値誤差に対する計算ロバスト性を向上できる利点もある。SSFにおける個別方程式同定問題への分割アルゴリズムは、不変性定理に示される許容変換に基づく式の変形と数量の代入消去によって、各方程式に現れる数量の個数を最小化した形式を出力する。そして一度、科学的法則式発見システムによって各方程式の定量的形式が同定されれば、それらは不変性定理に示される許容変換に従う限り、人間にとって理解しやすく目的に応じた形式に如何様にも変形可能である。

図4のアルゴリズムは、図2のアルゴリズムで得られた構造正準形を各方程式が最小個数の数量を含むように変換するものである。このアルゴリズムは、図2で得られた完全部分集合のリスト及びそれぞれの次数の情報を用いる。アルゴリズム内の(S2)において、各方程式に現れる数量は他の完全部分集合に含まれる方

式によって消去される。そして(S3)において、もし完全部分集合の次数が1を越えていれば、同一の完全部分集合内の他の方程式によって更に各方程式に含まれる数量の消去が行われる。現状のSSFでは、(S2)や(S3)において消去される数量は単に名前の辞書順に選ばれる。しかしこの選択には、測定データのノイズやフィッティング誤差を小さくする数量の組み合わせを残すなど、より良い基準の採用も将来考えられる。

4. SDS の概要

SDSはSSFから与えられる1本の完全方程式中の数量のリスト情報とそれらに関する実際の実験データ以外に、それら数量各々のスケールタイプの情報を必要とする[Washio and Motoda 1997]。測定数量のスケールタイプとは、その測定過程において対象に割り当てられる数値を定める規則を表すものである。定量的な数量に関する代表的なスケールタイプとしては、間隔尺度、比例尺度、絶対尺度がある。間隔尺度の典型例は摂氏温度や人間の音程感覚などであり、これらの基礎原点には人為的な意味しかない。比例尺度の例は質量や絶対温度などであり、これらの基礎原点は物理的に意味があり人為的な変更は許されない。絶対尺度はラジアン角のように測定単位を持たない無次元量であり、原点のみならずその測定目盛間隔すら人為的な変更は許されない。

SDSは“拡張 Buckingham II-theorem”と“拡張 Product Theorem”という2つの重要な定理に基づいて、方程式形式の探索を行う。前者は間隔尺度と比例尺度、絶対尺度の数量から成る物理的に意味のある完全方程式 $\phi(x_1, x_2, x_3, \dots, x_n) = 0$ は、全て“ensemble方程式”と呼ばれる $F(\Pi_1, \Pi_2, \dots, \Pi_{n-w}) = 0$ の形式に変換可能であることを述べている。ここで n は ϕ が含む数量の数、 w はこれら数量が有する基礎単位や基礎原点の自由度の総数である。また、全ての Π_i は絶対尺度量である。後者の定理は、間隔尺度と比例尺度の数量によって上記の各 Π_i が以下のような“regime”と呼ばれる数式で表現されることを述べている。

$$\Pi = \left(\prod_{x_i \in R} |x_i|^{a_i} \right) \left(\prod_{I_k \subseteq I} \left(\sum_{x_j \in I_k} b_{kj} |x_j| + c_k \right)^{a_k} \right)$$

$$\Pi = \sum_{x_i \in R} a_i \log |x_i|$$

$$+ \sum_{I_k \subseteq I} a_k \log \left(\sum_{x_j \in I_k} b_{kj} |x_j| + c_k \right)$$

$$+ \sum_{x_\ell \in I_g \subseteq I} b_{g\ell} |x_\ell| + c_g$$

ここで、 I と R はそれぞれ元の ϕ が含む間隔尺度及び比例尺度の数量の集合である。これらの式で、 Π を除いて全ての係数は定数であり、かつ $I_k \cap I_g = \phi$ である。SDS は、はじめ2数量テストを繰り返して拡張 *Product Theorem* に示される *regime* 式に当てはまる数量間の関係を探る。本定理に示されるスケールタイプによる制約関係は、第一原理法則式であるための数学的必要条件であるため、この方法で発見される数量間の関係は対象系を支配する第一原理を表す可能性が高い。

全ての *regime* 式が見つかった後、SDS は更に“恒等制約”という数学的制約を用いて *ensemble* 方程式の探索を行う。恒等制約は、例えば $\Theta_h, \Theta_i, \Theta_j$ という3数量の内2つのペアに、 $a(\Theta_j)\Theta_h + \Theta_i = b(\Theta_j)$ と $a(\Theta_i)\Theta_h + \Theta_j = b(\Theta_i)$ という線形関係が知られている時、3数量全体に数学的許容される関係を導くものである。この場合には簡単な証明によって、

$$\Theta_h + \alpha_1 \Theta_i \Theta_j + \beta_1 \Theta_i + \alpha_2 \Theta_j + \beta_2 = 0$$

という2重線形関係が導かれる。この原理は複数の数量間の様々な関係について一般化可能である。

SDS では、2数量テストと以上に述べた数学的制約を用いて全数量間の各2数量関係の同定を試みる。そして、“3つ組テスト”と呼ばれる2数量関係の間の整合性チェックを行い、データのノイズやフィッティング誤差によって得られる間違っただけの関係は排除する。更に正しいものについて3数量以上の多数量関係へのマージを逐次行い、最後に全体の数量関係として完全方程式を同定する。

少ない必要計算量、ノイズや誤差に対する高いロバスト性、大規模問題への高い適用性、幅広い分野への適用性など、SDS は優れた性能を持つことが確認されている [Washio and Motoda 1997]。このような性能は、従来にはない新たな数学的制約やテストアルゴリズムの導入によって得られたものである。

5. SSF と SDS の組み合わせ評価

SSF プログラムを MATLAB と呼ばれる数値処理シェル環境を利用して構築した [Matlab 1992]。対象系の制御実験を通じ SSF によって発見された連立方程式系の構造情報は方程式1本分の情報ずつ SDS に送られ、SDS はその方程式内に含まれる数量のスケールタイプ情報に基づいて対象系の測定実験を行う。この

過程が各方程式に関して繰り返され、最後に連立方程式系全体の定量的形式が同定される。なお本研究では、シミュレーションでの実験環境による対象系同定を試みた。

はじめに SSF 単独の性能を結果の妥当性と必要計算量の観点から評価した。そのために、かなり大規模な対象系を含む幾つかの評価例題を準備した。更に、SSF に SDS を組み合わせた場合の性能を同じ例題で評価した。取り上げた例題は以下のものである。

(1) 2個の並列抵抗と電池から成る回路

これは図1に既に示した例題である。これは連立方程式(1)に示されるように4本の方程式と7つの数量で表される

(2) 環状流路壁における熱伝達

十分に大きな熱容量を有する水平な分厚い固体壁に、垂直に2つの環状の流路穴が十分に離れて空いている系を考える。そこには上部から気体流れ込み、それぞれの環状流路を通過する間に固体流路壁から熱を奪われ液体に凝縮する現象が起きている。これら2つの環状流路壁における熱伝達現象は、以下のような17数量、8方程式で表される [Kalagnanam et al. 1994]。

$$\omega = 0.9423 \left(\frac{v s k^3}{L \mu} \right)^{1/4}$$

$$\Delta T_1 = T_f - T_{w1}$$

$$\Delta T_2 = T_f - T_{w2}$$

$$h_1 = \Delta T_1^{-1/4} \omega \quad (6)$$

$$h_2 = \Delta T_2^{-1/4} \omega$$

$$\dot{H}_1 = 2\pi r L h_1 \Delta T_1$$

$$\dot{H}_2 = 2\pi r L h_2 \Delta T_2$$

$$\dot{H} = \dot{H}_1 + \dot{H}_2$$

ここで、 v, s, k, μ は、それぞれ流体の単位体積当たりの潜熱、密度、熱伝導率、粘性であり、 L, r はそれぞれ流路長及び流路半径、 T_f は流体温度、 T_{w1}, T_{w2} は2つの環状流路の壁温、 \dot{H} は流体から壁面への単位時間当たりの全熱伝達量である。

(3) 光量増加率測定電子回路

これは図5に示される一定時間内の受光量の増加率を測定する回路である。キャパシタに平行して設けられている抵抗とスイッチは、この回路動作を一定時間でリセットするためのものである。この系のモデルは22数量を含む14本の方程式によって表される。

(4) 原子力プラントの炉心

炉心の核分裂動特性、燃料体の除熱、炉心冷却材の熱及び質量収支を含むプロセスである。60個の数量を

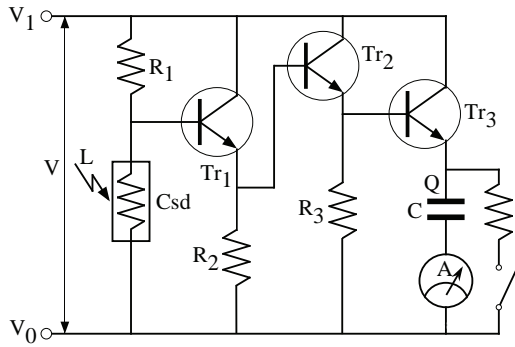


図5 光量増加率測定電子回路

む 26 本の式で表される大規模な系である。

表 1 は上記各問題の規模と、それに対する計算量、ノイズに対するロバスト性をまとめたものである。 T_{scf} は SSF において構造正準形を導出するのに必要な計算量、 T_{min} は各方程式に現れる数量の個数を最小化するのに要した計算量である。また T_{tl} と T_{av} は、それぞれ SDS によって全方程式形式を求めるのに要した計算量とその各方程式当たりの計算量である。 T_{scf} は m, n 、即ち問題の規模に強い依存性を示している。これは SSF のアルゴリズムが問題の規模に対して非多項式オーダーの計算量を必要とすることによる。また T_{scf} は各対象系モデルに含まれる数量の個数と方程式の個数の差、即ち $n - m$ にもある程度依存する。これは、 $n - m$ が大きいと対象とする連立方程式系に高い数量制御の自由度が存在し、探索空間がやはり指数的に大きくなるためである。これらに対し、 T_{min} は問題の規模について僅かな依存性しか持たず、その値自体も無視できる程に小さい。これは方程式に現れる数量の個数最小化を行う図 4 のアルゴリズムが $O(n^2)$ の複雑さしかもたないことを裏付けている。SDS に必要とされる全計算量 T_{tl} は SSF の T_{scf} に比べれば穏やかな変化を示す。SDS は各方程式を独立に扱うので、 m については比例する計算量しか必要としないのは当然である。しかし、SDS の計算量は各方程式に含まれる数量の個数の平均値 av には比較的強い依存性を示す。この傾向は T_{av} を見るとより明確である。過去の我々の研究において、SDS の必要計算量は 1 本の方程式に含まれる式の本数の約 2 乗のオーダーになることが明らかとなっている [Washio and Motoda 1997]。 av と T_{av} の関係はほぼこれに従っており、全体の T_{tl} の計算量は $O(m av^2)$ になることがわかる。要約すると大規模問題に関しては、問題の規模について計算量が指数オー

表 1 計算量とロバスト性の評価

Ex.	m	n	av	T_{scf}	T_{min}	T_{tl}	T_{av}	NL
(1)	4	7	2.5	3	0.00	206	52	35
(2)	8	17	3.9	1035	0.05	725	91	29
(3)	14	22	2.6	1201	0.05	773	55	31
(4)	26	60	4.0	42395	0.11	3315	128	26

m: number of equation, n: number of quantities, av: average number of quantities/equation, T_{scf} : CPU time (sec) to derive structural canonical form, T_{min} : CPU time to derive minimum quantities form, T_{tl} : CPU time to derive all equations by SDS, T_{av} : average CPU time per equation by SDS, NL: limitation of % noise level of SDS.

ダーで増加する SSF がネックとなる。しかしながら、表 1 に示されるように、 DE_i による探索空間の絞り込みにより、それでもなおかなり大規模な工学的問題に対応可能であることがわかる。

表 1 の右端の列は、測定実験データに測定ノイズを加えた場合に対する SSF と SDS の組み合わせのロバスト性の目安を表している。ここでは数量の測定値の絶対値に対して標準偏差の 2 倍が、あるパーセンテージとなる強さのガウスノイズを加えた。NL は 10 回の試行中 8 回まで、定量的に正しい解が得られた最大のパーセンテージを表している。これを見ると、各数量の測定値の絶対値に対して 25-35% のノイズを加えた場合でも、妥当な答えが得られることがわかる。先述の通り、SSF は幾つかの数量の制御の下で他の数量が被制御になるか否かを、統計 t -検定を用いて判定するので、測定の標本数がある程度得られればノイズが多くても高い信頼性を有する答えを導くことができる。また、SDS も基本的には 2 数量テストという極めて単純な数値データフィッティングを繰り返すだけなので、ノイズに関して極めてロバストであることが知られている [Washio and Motoda 1997]。また、対象問題が大規模であっても SSF はその一部分である完全部分集合の導出のみを対象とし、SDS も 2 数量間のテストのみを繰り返すだけなので、対象規模の増大の割にロバスト性は失われないことがわかる。このように、両者の組み合わせは実用上十分な耐ノイズ性を示す。

最後にこれらの例に対する結果の妥当性を検証した。例題 (1) では、SSF は T_4 に示される構造正準形を導出し、更に以下のように各方程式について含まれる数量の個数を最小化した組み合わせを SDS に出力した。

$$\{V_e, R_1, I_1\}, \{V_e, R_2, I_2\}, \{V_e, V_1\}, \{V_e, V_2\}$$

その結果、SDS は次のような解を導いた。

$$V_e = I_1 R_1 [1], V_e = I_2 R_2 [2],$$

$$V_e = V_1 [3] \text{ and } V_e = V_2 [4]$$

この解は定量的のみならず構造に関する不変性の面でも連立方程式 (1) と等価である。(2) の例では, SSF は以下のような構造正準形を導いた.

$$\begin{aligned} & \{\omega, v, s, k, L, \mu\}, \\ & \{\Delta T_1, T_f, T_{w1}\}, \\ & \{\Delta T_2, T_f, T_{w2}\}, \\ & \{h_1, \Delta T_1, \omega\}, \\ & \{h_2, \Delta T_2, \omega\}, \\ & \{\dot{H}_1, \Delta T_1, r, L, h_1, \omega\}, \\ & \{\dot{H}_2, \Delta T_2, r, L, h_2, \omega\}, \\ & \{\dot{H}, \dot{H}_1, \dot{H}_2\} \end{aligned}$$

そして, 最後の 2 つの方程式を表す集合に対し, 5 番目と 6 番目の方程式を表す集合の ω の代入消去によって, SSF は連立方程式 (6) と同じ数量配置を導出した. その結果を受けて, SDS は同連立方程式と同一の解を得た. 同様に, 他の例においてもほぼ元の連立方程式系と同じ形をした等価な解が得られた.

6. 関連研究と議論

SSF が出力する各方程式の数量配置, 即ち構造正準形から構造情報を保存したままで各方程式が含む数量の個数を最小化した配置は, 多くの場合, 我々が理解し易い連立方程式モデルである. これは方程式間で共有する数量をなるべく少なく整理して, 各方程式が表す過程を明確にする方が理解し易いためであると思われる.

本論文の最初に述べたように, 従来の科学的法則式発見システムは必要計算量が非常に多く, かつ高々 2, 3 本の完全方程式を導出し得ただけであった. SSF 及びそれが持つ原理はこのような限界を超えるのみならず, これまでの殆ど全ての科学的法則式発見システムとの組み合わせが可能であり, 一般的なツール及び原理として従来システムの性能向上に用いることが可能である. 更にこの基本原理は, 連続系に限らず連立的な制約を有する限り離散的な対象系にも適用可能であり, 非常に広い一般性を有する.

本論文が示した連立方程式系の完全部分集合に関する基本理論は, 因果的順序づけの理論 [Iwasaki and Simon 1986] の拡張であると捉えることができる. 完全部分集合は, 完結部分集合 (“*self-contained subset*”) の全ての候補を内包している. 一度, 幾つかの数量の

値が外生的に指定されると, 完全部分集合の一部が完結部分集合となる. 本研究が提案した構造形もまた構造方程式 [Iwasaki and Simon 1986] の拡張である. 本理論は構造方程式のより精密な定義と特徴づけを与えている.

7. おわりに

本研究は数量間の依存性に関する不変構造の観点から, 制約不足の連立方程式系を数学的に特徴づけた. そして, 実験を通じてその構造を同定するアルゴリズムを提案した. 更に, これらの原理やアルゴリズムを従来の科学的法則式発見システムに適用し, 連立方程式系を同定するための一般的枠組みを確立した. このアルゴリズムを SSF という一般的プログラムツールにまとめ, それを SDS と組み合わせることで優れた性能を発揮することを確認した.

参考文献

- [Langley et al. 1987] P.W. Langley, H.A. Simon, G. Bradshaw and J.M. Zytkow, Scientific Discovery; Computational Explorations of the Creative Process, MIT Press, Cambridge, Massachusetts, 1987.
- [Koehn and Zytkow 1986] B. Koehn and J.M. Zytkow, Experimenting and theorizing in theory formation., Proceedings of the International Symposium on Methodologies for Intelligent Systems, pp.296-307, ACM SIGART Press, 1986.
- [Falkenhainer and Michalski 1986] B.C. Falkenhainer and R.S. Michalski, Integrating Quantitative and Qualitative Discovery: The ABACUS System, Machine Learning, pp.367-401, Kluwer Academic Publishers, BOSTON, 1986.
- [Nordhausen and Langley 1990] B. Nordhausen and P.W. Langley, An Integrated Approach to Empirical Discovery, Computational Models of Scientific Discovery and Theory Formation, Morgan Kaufman Publishers, San Mateo, California, 1990.
- [Schaffer 1990] C. Schaffer, A Proven Domain-Independent Scientific Function-Finding Algorithm, Proceedings Eighth National Conference on Artificial Intelligence, pp.828-833, AAAI Press/The MIT Press, 1990.
- [Huang and Zytkow 1996] K.M. Huang and J.M. Zytkow, Robotic discovery: the dilemmas of empirical equations, Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery, pp.217-224, Tokyo, 1996.
- [Kokar 1986] M.M. Kokar, Determining Arguments of Invariant Functional Descriptions, Machine Learning, pp.403-422, Kluwer Academic Publishers, BOSTON, 1986.
- [Washio and Motoda 1997] T. Washio and H. Motoda, Discovering Admissible Models of Complex Systems Based on Scale-Types and Identity Constraints, Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, pp.810-817, Nagoya, 1997.
- [Iwasaki and Simon 1986] Y. Iwasaki and H.A. Simon,

Causality in Device Behavior, Artificial Intelligence, pp.3-32, Elsevier Science Publishers B.V., North-Holland, 1986.

[Murota 1987] K. Murota, Systems Analysis by Graphs and Matroids - Structural Solvability and Controllability, Algorithms and Combinatorics, Vol.3, Springer-Verlag, Berlin, 1987.

[Dzeroski and Todorovski 1994] S. Dzeroski and L. Todorovski, Discovering Dynamics: From Inductive Logic Programming to Machine Discovery, Journal of Intelligent Information Systems, Vol.3, pp.1-20, Kluwer Academic Publishers, Boston, 1994.

[Kalagnanam et al. 1994] J. Kalagnanam and M. Henrion and E. Subrahmanian, The Scope of Dimensional Analysis in Qualitative Reasoning, Computational Intelligence, Vol.10, No.2, pp.117-133, 1994.

[Matlab 1992] MATLAB Reference Guide, The Math Works, Inc., 1992.

〔担当編集委員：××，査読者：××〕

— (略) —

著者紹介

鷲尾 隆 (正会員)

1960年生。1983年東北大学工学部原子核工学科卒業。1988年東北大学大学院原子核工学専攻博士課程修了。工学博士。1988年から1990年にかけてマセチューセツ工科大学原子炉研究所客員研究員。1990年(株)三菱総合研究所入社。1996年退社。現在、大阪大学産業科学研究所助教授(知能システム科学研究部門)原子カシステムの異常診断手法に関する研究、定性推論に関する研究を経て、現在は人工知能の基礎研究、特に科学的知識発見、データマイニング

などの研究に従事。著書に“*Expert Systems Applications within the Nuclear Industry*”, American Nuclear Society, 『知能工学概論』: 第2章エージェント(共著, 廣田 薫 編, 昭晃堂)など。人工知能学会, 計測自動制御学会, 日本ファジイ学会, 情報処理学会, AAAI, 各会員。

washio@sanken.osaka-u.ac.jp

元田 洸 (正会員)

1943年生。1965年東京大学工学部原子力工学科卒業。1967年東京大学大学院原子力工学専攻修士課程修了。同年(株)日立製作所入社。同社中央研究所, 原子力研究所, エネルギー研究所, 基礎研究所を経て1995年退社。現在、大阪大学産業科学研究所教授(知能システム科学研究部門)原子カシステムの設計, 運用, 制御に関する研究, 診断型エキスパート・システムの研究を経て、現在は

人工知能の基礎研究、特に機械学習、知識獲得、知識発見などの研究に従事。工学博士。認知科学会, 人工知能学会, 情報処理学会, 日本ソフトウェア科学会, AAAI, IEEE Computer Society, 各会員。

motoda@sanken.osaka-u.ac.jp