特集論文 「発見科学」 Technical Papers

スケールタイプ制約に基づく科学的法則式の発見

Discovery of Scientific Law Equations Based on Scale-Type Constraints

鷲尾 隆 Takashi Washio
大阪大学産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University.
washio@sanken.osaka-u.ac.jp, http://www.ar.sanken.osaka-u.ac.jp/washprjp.html

元田 浩 Hiroshi Motoda
（同　　上）
motoda@sanken.osaka-u.ac.jp, http://www.ar.sanken.osaka-u.ac.jp/motoprjp.html

Keywords: scientific discovery, scientific law equation, scale-type constraint, identity constraint, dimensional analysis.

Summary

A novel and generic theory is formulated to characterize the structure of a scientific law/model equation. Based on the theory, an efficient algorithm is developed to discover scientific law/model equations governing an objective process under experimental environments, and the algorithm is implemented to the "*Smart Discovery System* (SDS)" program. SDS derives the quantitative equations reflecting the scientific first principles underlying the objective process. The power of the proposed approach comes from the use of "*scale-type constraints*" to limit the mathematically admissible relations among the measurement quantities representing the states of the objective process. These constraints well specify the admissible formulae of the scientific law/model equations, and provide a measure to efficiently reduce the search space of the equation formulae. In this paper, the theoretical foundation to discover the scientific law/model equations and the algorithm of SDS are presented, and its efficiency and practicality are demonstrated and discussed with complex working examples. Since the conventional equation discovery systems could not sufficiently guarantee the mathematical admissibility of the discovered equations, this work is expected to open up a new research field on the scientific equation discovery.

1. 科学的法則式発見研究の概観

科学的問題領域ではしばしば実験や観測によって，方程式で表される規則性を示す数値データが収集されることが多い．古来，科学者達は数値データからそのような規則性を読みとり，"科学的法則式"を発見し，更にそれらを組み合わせて複雑な現象の"科学的モデル式"を構築して来た．近年の計算機による情報処理技術の発達と共に，統計数理をはじめ人工知能や知識発見など幅広い分野において，このような科学者の行為を自動化ないしは支援する試みが展開されている．しかしながら，対象の規則性を表す方程式が科学的法則式であるための条件に関する体系的かつ詳細な研究はあまり多くない．法則式は各学問体系発展の歴史的経緯の中で発見・命名されてきており，その学問体系上の慣習によって広く知られたあくまでも経験的な関係であるという解釈が取られることも多い．その一方で，法則式の定義や成立条件を演繹的立場から明確化しようという努力も，断片的ながら多くの著名な科学者達によって行われてきた．もちろん，慣習的側面を有する"科学的法則式"という用語を，例外なく定式化することには無理があろう．しかしながら，対象解明の基本的道具である法則式が具備すべき条件を明確化することは，学問体系の基礎をより堅牢なものにするためにも意義深い．

R. Descartes は演繹法に関する考察の中で推論の明証性や問題分割，健全性や無矛盾性の規範を提唱しており [Descartes 1637]，これらは学問体系の構成要素として法則式を見いだす際にも重要である．I. Newton は自然法則を発見する際の規範として，自然界の原因以外を排除する客観性，なるべく少数の原因のみを仮定する簡潔性，広範囲での成立を要請する普遍性，すべての試行結果に反しない健全性を要求した [Newton 1686]．H.A. Simon は法則記述の簡潔性のみならず，それが簡明な原理によって発見されるべきであることも主張している [Simon 97]．また，R.P. Feynman は特に法則形式の時間や空間に関する数学的許容性の重要性を述べている [Feynman 65]．更に近年，R.E. Valdes-Perez は法則式を含む科学的知識の正しさを経験を通じて証明することは困難であり，問

題にすべきはその "尤もらしさ (*plausibility*)" であると主張している [Valdes-Perez 99]. 筆者等もこのような知見を踏まえ，法則式に要請される条件の体系化を試みており，数値データから得る方程式が "科学的" であるためには，客観性，普遍性，再現性，健全性，簡潔性，数学的許容性といった規範を経験上尤もらしく満たすことが結論されねばならないと考えている [鷲尾 98a]. しかし，実際には計算機を用いて限られた環境やデータから，これら全ての規範を満たすことが十分尤もらしいと結論づけられる方程式を得ることは困難なことが多い. 従って，過去に提案された手法の殆どが，一定の前提の下でより緩和された規範を用い，"科学的法則式及びモデル式である可能性の高い式" を発見する立場を取っている.

特に対象に関する数値データの収集条件の前提は，上記規範の確認しやすさに大きな影響を与える. この条件は以下の2種類に分類される.

(1) 対象状態の受動的観測のみが許される場合

(2) 対象状態を能動的に変更する観測が許される場合

(1) では統計的回帰分析やシステム同定理論，ニューラルネットワークなどの手法が用いられてきた. この条件下では，解析者が事前の知見から法則式形式として適切なものを採用するか，それが困難な場合でも採用する方程式形式が真の法則式を漸近的に近似可能であると仮定して，採用した式をデータに当てはめ，対象データの規則性を表す方程式やモデル式を得る [Chatterjee 86, Ljung 87, Wasserman 89]. 近年では，ニューラルネットワークの階層構造に自然法則式で経験上よく見られる形式を埋め込み，法則式の可能性を有するモデルを導く研究も進められている [Saito 97]. また，式の全体形式を最初から仮定せず各部分的形式のみを仮定し，数値データに合うようにそれらをボトムアップに組み上げる手法も研究されて来た. これには部分的依存関係を調べながら線形方程式構築を行う TETRAD[Glymour 87] や同じく多項式を用いる GMDH[Ivakhnenko 70]，微分と積の2つの演算子を繰り返し適応して多数の数量を生成し，それらの部分集合内の数量関係を線形近似する生成・テスト法に基づく LAGRANGE[Dzeroski 94] などがある. 以上の手法の何れにおいても，対象やそれに類似した系に対する観測で得られた様々なデータへの当てはめを繰り返すことで，得られる結果の客観性や再現性を経験的に確認することは可能である. また，AIC や MDL 等の簡潔性に関する規範を適用して，より妥当性の高い式を選ぶことも可能である [Akaike 68, Rissanen 78]. しかしながら，対象に関する受動的観測下では，観測可能な対象状態や観測条件に偏りを生じることが多く，得られた関係式の適用限界，即ち普遍性や，他の知見との整合性，即ち健全性に関する確認は一般に困難である. 更に以上の手法の何れもが，法則式やモデル式の形式に関して何らかの蓋然的前提を置く立場をとっており，その前提が対象を表す法則式形式として妥当であるか否かの数学的許容性に関する直接的検証は含まない. 前提の妥当性検証は，データへの当てはめ誤差に関する統計検定，即ち与えられたデータの範囲内での経験的検証に留まり，前提が正しくない場合にはデータに見かけ上よく合う経験式が得られるに過ぎない.

これに対して (2) の場合には，科学者が実験において通常行うのと同じく，対象の種々の性質を詳細に分離して観測可能であるため，各規範の確認が比較的容易である. P. W. Langley 等はこの収集条件の下で法則式発見を行う手法を提案し，それに基づく BACON というシステムを構築した [Langley 81, Langley 87]. 更にこの研究を引き継ぐ形で，FAHRENHEIT [Koehn 86]，ABACUS [Falkenhainer 86]，IDS [Nordhausen 90] など，BACON で使わ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. これらの手法では，多数の観測数量の中から任意の2数量組を取り上げ，実験データへの当てはめを通じて，その2項関係を多様な候補の中から選択する. 更に多数の2数量組の関係同定や中間変数の生成を繰り返し，ボトムアップに全体の法則式ないしモデル式を組み上げていく. 様々な条件設定による実験データを収集し関係式を導出するため，客観性，再現性はもとより，実験条件の範囲での普遍性や健全性の確認も行われる. また，ボトムアップな式の組み上げにより，データを説明可能な最も簡潔な方程式が得られる. ただし，方程式形式の妥当性検証については，基本的には (1) の場合と同じくデータへの経験的当てはまりの良さの統計検定に留まり，数学的許容性の確認を欠いている. また，種々の2項関係やそれらの組み合わせを探索するため必要とする計算量が膨大であり，僅か数個の数量で表現されるような小規模な問題でも，法則式とは見なせない解を導くことが多い. そこで，COPER のように数量の単位次元の情報を用いて数学的許容性の確認を行う手法が考えられた [Kokar 86]. ただし，この方法では物理学など数量単位が明確な分野にしか適用できない.

また別の視点として，求めたい方程式の種類の観点から，

(A) 1つの完全方程式を得る場合

(B) 区分完全方程式を得る場合

(C) 連立方程式系を得る場合

の各場合に関する手法開発が行われて来た. 完全方程式とは例えば，$f = ma$ が3数量の自由度を1減らして2次元に落とすように，数量関係の自由度を1だけ制約する方程式のことである. これに対し，$(x-r)^2+(y-r)^2=0$ は x,y,r が実数である限り $x=r, y=r$ と等価であり，3数量の自由度を2つ減らすので完全方程式ではない. 前述した (1),(2) の何れの手法も (A) の目的に用いることができる. (B) は例えば，水の物性状態方程式を発見する場合に，固体，液体，気体の物性値観測データを区別しないで与えても，自動的に融点と沸点の温度を同定し，それぞれの温度区間における物性方程式を得るような場合である. 上記 (2) のような実験環境下では，このような対

象の各モード区間を同定することが比較的容易であるため，前述のFAHRENHEITやABACUSはこの機能を備えている．また，(C)の連立方程式については，(1)の場合は統計的回帰分析や線形システム同定理論，ニューラルネット，GMDH，LAGRAGEなどが導出可能である．しかし，何れも連立方程式における数量間の依存構造について事前の知見に基づく前提を導入する必要があり，導かれた式の構造が普遍性，再現性，健全性，数学的許容性などを有する保証はない．これに対して前述のTETRADは，線形式に限定ではあるが数量の各組み合わせの依存性に関する統計的検証を積み上げる方法を用い，対象を実際に構成する科学的第一原理の依存構造を反映する可能性の高い連立方程式モデルを得る．また(2)の場合では，近年筆者等が，実験環境下において各数量の制御可能性の確認を通じて対象の数量依存関係を同定し，科学的第一原理を反映する可能性の高い連立方程式モデルを発見する手法を開発している[Washio 98b]．この方法では，実験可能な範囲で極力高い客観性や普遍性，再現性，健全性，簡潔性，数学的許容性を有する連立方程式構造が得られる．

2. 研究の目的

本研究は，以上のような一連の科学的法則式発見研究の中で，“(2)対象状態を能動的に変更する観測が許される場合”に“(A)1つの完全方程式を得る場合”を取り上げる．この分野では，BACONをはじめとする一連の法則式発見システムが提案されて来ているが，方程式形式の数学的許容性の保証や扱える問題の規模と計算量，適用可能範囲の点で多くの問題点を抱えている．そこで本研究では，従来手法が抱えるこのような問題を解消すべく，多くの数量で表される複雑な対象に関して，実験で得る数値データと数量に関する僅かな知見から，数学的に第一原理法則式として許容される1つの法則式や1つの対象モデル式を，観測ノイズ誤差にロバストでかつ多項式時間で高速に求解する手法を提案する．そして，その手法を実装した“*Smart Discovery System* (SDS)”の開発を行い，提案手法の性能評価を行うことを目的とする．

SDSが必要とする入力情報は，対象を表現するに十分な種類の数量に関する実験観測データとそれら数量各々のスケールタイプである．SDSは，法則式に要求されるスケールタイプ制約(*scale-type constraint*)及び恒等関係制約(*identity constraint*)という強くかつ一般的な数学的許容性制約を満たす範囲で，実験観測データをよく説明する方程式を探索する．そのため，単位次元制約のような対象に強く依存した知識を用いずとも十分に探索空間を絞り込んで，科学的第一原理を表す可能性が非常に高い法則式やモデル式を導出可能である．また，中間変数を多用するBACON系のアルゴリズムと異なり，SDSは各2数量組毎の関係をデータから求めた後，全ての3つ組数量の間で整合性チェックを行い，その後に多数量間の1つの方程式に組み上げる．これにより，観測ノイズや誤差に対する高いロバスト性を確保し，かつ数量の個数に関して高々3次の計算量で目的とする式を導出可能である[Washio 97]．

3. SDSが用いる制約

本研究の提案手法を説明する前に，その中で用いられる数学的制約について論じる．

3·1 スケールタイプの数学的許容性

測定過程の観点から数量の特徴付けを行う“測定論(*measurement theory*)”と呼ばれる研究領域において，S. S. Stevensは測定過程を“一定の規則（基礎的経験操作）によって対象や事象に数を割り当てること”と定義した[Stevens 46]．彼は異なる規則の下で数が割り当てられれば，異なる種類の“スケールタイプ(*scale type*)”の数量とその“測定(*measurement*)”が得られると考え，幾つかのスケールタイプを定式化した．その中で重要な3種類を表1にまとめる．この中で“間隔尺度(*interval scale*)”と“比例尺度(*ratio scale*)”は，物理学や心理学，経済学，社会学などの問題領域において主要なスケールタイプである．間隔尺度量には，例えば摂氏や華氏の単位の温度や，エネルギー，エントロピー，時刻，音程などがある．これらを測るスケールの原点は絶対的なものではなく，我々の定義によってどのようにも変更可能である．従って，間隔尺度量は2つの測定量の差（間隔）の大きさに関する情報しか持たない．数学的に許容される単位変換は一般的線形変換である．一方，比例尺度量としては，質量，絶対温度，圧力，時間間隔，周波数，金額などが挙げられる．これらの値はすべて絶対的な原点によって定められ，そこから測った2つの測定量の比はどのような単位を採用しようとも“不変(*invariant*)”である．数学的許容単位変換は相似変換である．もう1つの“絶対尺度(*absolute scale*)”はいわゆる無次元量であり，異なる測定過程を用いてもその値自体が不変であるので，単位を定義することに意味が認められない量である．例として，2つの長さの比や角度(radian)，流体力学におけるNusselt数，Reynolds数などが挙げられる．単位を持たないので許容変換は恒等変換である．ここで注意したいことは，スケールタイプは単位次元とは異なることである．スケールタイプは単に測定規則の定義を反映しているに過ぎず，数量の背後の物理的意味は表さない．これに対し，単位次元は背後の物理的意味や他の量との関係に関する情報も含んでいる．

この研究を引き継いで，R. D. Luceはもし2つの数量が絶対尺度量を媒介としない直接の依存関係を持つ場合には，その関係は各数量のスケールタイプの性質によって決まる基礎的な関数で表されると主張した[Luce 59]．

表 1 スケールタイプの種類

尺度	基礎的 経験操作	数学的 許容変換
間隔	間隔や差の 等値性の決定	一般的線形変換 $x' = kx + c$
尺度	比の等値 性の決定	相似変換 $x' = kx$
絶対	絶対的値の 等値性の決定	恒等変換 $x' = x$

例えば，x と y が両方とも比例尺度であり，y が x によって連続関数 $y = u(x)$ の形で定義されるとする．仮にその関係が対数関数，即ち $y = \log x$ であると仮定する．この時，表1に示される比例尺度 x の許容変換に抵触しないで，x にある正数 k を掛け単位を変更することができるが，これによって $u(kx) = \log k + \log x$ となり，$\log k$ 分だけ y の原点が移動してしまう．これは明らかに比例尺度である y の性質を破ってしまう．従って，x と y の間の関数関係は対数であってはならない．x と y それぞれに許容される単位変換は $x' = kx$ 及び $y' = Ky$ であるので，関係 $y = u(x)$ は $y' = u(x') \leftrightarrow Ky = u(kx)$ となる．y の単位変換係数 K は k に依存するが，x の値には依存してはならない．そこで，K を $K(k)$ と表すと，連続関数 $u(x)$ について以下の制約を得る．

$$u(kx) = K(k)u(x).$$

ここで，k, K は単位変換係数であるため，$k > 0$ 及び $K(k) > 0$ である．表2の第4列目には，比例尺度量と間隔尺度量の間の関係制約がまとめられている [Luce 59]．ルースは $x \geq 0$ かつ $u(x) \geq 0$ の下での $u(x)$ の各解を得た．我々は彼の定理を拡張し，x と $u(x)$ の負値領域までを網羅する結果を得た．それらの詳細な証明は他の文献に譲る [Washio 96]．それらを表2の最後列にまとめる．なお，単位を持たない絶対尺度量同士や絶対尺度量と他のスケールタイプ量との間にはこのような制約は存在せず，スケールタイプの観点からは任意の関係が許される．表2で間隔尺度量から比例尺度量を定義できない理由は，比例尺度量は絶対原点の情報を持つのに対し，間隔尺度量は持たず情報不足であるためである．クーロンの法則，オームの法則，ニュートンの万有引力の法則に現れる数量はすべて比例尺度であり，各法則式は表中の式1に従う．式2.1と式2.2の例は，エネルギーやエントロピーに関する法則に見られる．ある質量 m と速度 v を有する質点の全エネルギー U は，P をポテンシャルエネルギーとすると $U = mv^2/2 + P$ である．もし理想気体の温度が一定であれば，気体のエントロピー E は圧力 p について $E = -R\log p + E'$ という関係を持つ．ここで，R と E はそれぞれボルツマン定数及びエントロピーの参照基準値である．精神物理学の分野では，人間が感じる音程 s は周波数 f の対数に比例するというフェヒナーの法則が成立する．即ち $s = \alpha \log f + \beta$ である．ここでは，s は間隔尺度量であり f は比例尺度量である．式4の例としては，等速直線運動する質点に関する現時刻 t と速度 v，初期位置 x_0，現位置 x の間の関係，即ち $x = vt + x_0$ が挙げられる．

更に以上の2数量間の許容関係を，多数量間の関係式に拡張する．複数の測定数量間の数学的な許容関係に関しては，既に単位次元解析の分野で以下の2つの有名な定理が知られている [Bridgman 22, Buckingham 14]．

[定理 3·1] (Product Theorem) 数量相互の大きさの比が絶対的に定められる条件下では，数量 $x, y, z, ...$ を相互に直接関係づける関数 f は以下の形式を有する．

$$\Pi = \rho(x, y, z, ...) = \Gamma x^a y^b z^c ....$$

ここで Π は派生量，$\Gamma, a, b, c, ...$ は定数である．

[定理 3·2] (Buckingham Π-theorem) もし $\phi(x, y, ...) = 0$ が1つの完全方程式であれば，それを以下のような形式に書き換え可能である．

$$F(\Pi_1, \Pi_2, ..., \Pi_{n-r}) = 0$$

ここで，n は ϕ の引数の数であり，r は $x, y, z, ...$ の基礎単位の数である．またすべての i について，Π_i は無次元量である．

ここで "基礎単位 (*basic unit*)" とは，長さ $[L]$，質量 $[M]$，時間 $[T]$ のように ϕ において他の次元とは独立に数量のスケールを決める次元のことである．この定理は **Product Theorem** と共に，例えば振り子の時刻 $t[T]$ とその紐の長さ $l[L]$，重力加速度 $g[LT^{-2}]$，偏角 θ[無次元] の間の関係式を求めるのに用いることができる，ここで，2つの無次元量 $\Pi_1 = t(g/l)^{1/2}$, $\Pi_2 = \theta$ を構成できるので，**Buckingham Π-theorem** より $F(\Pi_1, \Pi_2) = F(t(g/l)^{1/2}, \theta) = 0$ という形の関係式を導ける．実際には，$F(\Pi_1, \Pi_2) = \Pi_2 - C\sin\Pi_1 = 0$ である．単位次元解析では各無次元量式 $\Pi_i = \rho_i(x, y, z, ...)$ を "レジーム (*regime*) 式" と呼び，無次元量間の関係式 $F(\Pi_1, \Pi_2, ..., \Pi_{n-r}) = 0$ を "アンサンブル (*ensemble*) 式" と呼ぶ．上記の例では，$\Pi_1 = t(g/l)^{1/2}$ と $\Pi_2 = \theta$ がそれぞれレジーム式であり，式 $\Pi_2 - C\sin\Pi_1 = 0$ がアンサンブル式である．1つのレジーム式は，その内部の数量が無次元量を通じてのみ外部の数量と関係づけられる相互に分離可能な部分的機構を表す．また1つのアンサンブル式は，対象の1つのまとまった機構を表す．

上記の2つの定理は，数量相互の大きさの比が絶対的に定められる場合のみを対象としている．これは観測数量全てが比例尺度であることに相当する．そこで，筆者等は前述の各スケールタイプに関する2数量間の許容関係式に基づき，これらの定理を観測数量に間隔尺度量が含まれる場合に拡張し，以下の定理を得た．これらの証明については他文献に譲る [Washio 99]．

[定理 3·3] (Extended Product Theorem) 1つのレジーム式が比例尺度の数量の集合 R と間隔尺度の数量

表 2 2数量間のスケールタイプ制約

No.	尺度の種類 独立変数	従属(被定義)変数	制約	可能な関係
1	ratio	ratio	$u(kx)=K(k)u(x)$	$u(x)=\alpha_*\|x\|^{\beta}$
2.1	ratio	interval	$u(kx)=K(k)u(x)+C(k)$	$u(x)=\alpha_*\|x\|^{\beta}+\delta$
2.2				$u(x)=\alpha\log\|x\|+\beta_*$
3	interval	ratio	$u(kx+c)=K(k,c)u(x)$	不可能
4	interval	interval	$u(kx+c)=K(k,c)u(x)+C(k,c)$	$u(x)=\alpha_*\|x\|+\beta$

1) 表記 α_*,β_* はそれぞれ $x\geq 0$ の時 α_+,β_+，$x<0$ の時 α_-,β_- を表す．

の集合 I から構成されているとする．ただし，R も I も空集合であることが可能である．この時，$x_i\in R\cup I$ を派生量 Π に関係づける関数 ρ は，以下の2式の何れかの形式を取る．

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

$$\Pi=\sum_{x_i\in R}a_i\log|x_i|+\sum_{I_k\in P_g}a_k\log\left(\sum_{x_j\in I_k}b_{kj}|x_j|+c_k\right)+\sum_{x_\ell\in I_g}b_{g\ell}|x_\ell|+c_g$$

ここで，Π を除き各係数は定数である．また，P は I の1つの分割であり，P_g は $I-I_g$ $(I_g\subseteq I)$ の1つの分割である．

［定理 3·4］(Extended Buckingham Π-theorem) $\phi(x,y,z,...)=0$ が1つの完全方程式であり，かつその中の各数量が間隔，比例，絶対尺度量の何れかである場合には，$\phi=0$ は以下の形式に書き換え可能である．

$$F(\Pi_1,\Pi_2,...,\Pi_{n-r-s})=0$$

ここで，n は ϕ の引数の数，r と s はそれぞれ $x,y,z,...$ の数量が有する基礎単位及び基礎原点の数である．また全ての i について Π_i は無次元量である．

ここで，新たに導入した“基礎原点 (*basic origin*)”とは，摂氏温度の次元における水の融点や海抜標高の次元における基準海面のように，他とは独立に間隔尺度の測定数量を決める原点のことである．

上記の2つの定理により，数量のスケールタイプの条件から科学的第一原理を反映する法則式やモデル式に数学的に要請される形式が明らかとなった．従って，対象を表現するのに必要な一連の数量と，各々のスケールタイプ及び基礎単位や基礎原点を共有する数量組が予め分かれば，各レジーム式の形式は，実験データを用いずとも十分に絞り込める．ただし，実際には何れの数量組が1つのレジーム式を構成するかは未知なので，逆に実験データから **Extended Product Theorem** の形式を満たす数量の組み合わせを探索する．このようにして発見される数量組とその関係式は，レジーム式である可能性が高いが，必ずしも本来のレジーム式と完全に一致する保証はないため，見かけ上の関係である場合を考慮して“擬レジーム (*pseudo-regime*) 式”と呼ぶ．

3·2 恒等関係の数学的許容性

スケールタイプ制約以外にも科学的第一原理を表す法則式やモデル式の形式を限定する数学的許容性制約が存在する．筆者等は，3数量の内で2種類の数量ペアの関係が既知である時，恒等的に整合する残りの1つのペアに成立する関係が非常に限られることに基づき，“恒等関係制約 (*identity constraints*)”を導いた．例えば，3数量 $\{\Theta_h,\Theta_i,\Theta_j\}$ 間に $a(\Theta_j)\Theta_h+\Theta_i=b(\Theta_j)$ と $a(\Theta_i)\Theta_h+\Theta_j=b(\Theta_i)$ という2つの数量ペアの線形関係が知られている時には，以下の関係が恒等的に成立しなければならない．

$$\Theta_h\equiv-\frac{\Theta_i}{a(\Theta_j)}+\frac{b(\Theta_j)}{a(\Theta_j)}\equiv-\frac{\Theta_j}{a(\Theta_i)}+\frac{b(\Theta_i)}{a(\Theta_i)}$$

ここで最右の式は，如何なる Θ_i の値についても Θ_j について線形であるから，中間の式も同様でなければならない．従って，以下が成立する．

$$1/a(\Theta_j)=\alpha_1\Theta_j+\beta_1$$
$$b(\Theta_j)/a(\Theta_j)=-\alpha_2\Theta_j-\beta_2$$

これらを中間の式に代入し以下の3数量関係を得る．

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

同様にして3数量以上に関しても，その中の幾つかの数量ペアの関係が既知であれば，全体で数学的に許容される関係を絞り込むことができる．筆者等は，幾つかの数量ペアの関係が，全て線形及び全てべき乗積である場合について一般的な多数量関係を明らかにした．その結果を表3にまとめる．

恒等関係制約は幾つかの数量ペアの関係が既知であれば，数量一般に適用可能である．従って，スケールタイプ制約が適用できない絶対尺度量（無次元量）やスケールタイプが未知である数量であっても，それらの部分的な法則関係が実験データなどから明らかである場合には，法則式やモデル式の形式を絞り込むことが可能である．

表 3 恒等関係制約

2数量間の関係	一般的関係
$ax+y=b$	$\sum_{(A_i\in 2^{LQ})\&(p\subsetneq A_i\forall p\in L)} a_i \prod_{x_j\in A_i} x_j = 0$
$x^a y=b$	$\prod_{(A_i\in 2^{PQ})\&(p\subsetneq A_i\forall p\in P)} \exp(a_i \prod_{x_j\in A_i} \log x_j) = 0$

L を線形関係を有する2数量ペアの集合とし，$LQ=\cup_{p\in L}p$ とする．P を積関係を有する2数量ペアの集合とし，$PQ=\cup_{p\in P}p$ とする．

4. SDSのアルゴリズム

与えられた対象に関して，科学的第一原理を表す法則式やモデル式の諸規範を満たす完全方程式を探索するアルゴリズム構築し，それを“SDS(*Smart Discovery System*)”として実装した．そのアルゴリズムを図1に示す．以降，その原理とアルゴリズムの詳細を説明する．

間隔尺度量の集合を IQ，比例尺度量の集合を RQ，絶対尺度量の集合を AQ とする．

(1-1) IQ 内の全ての数量ペアについて，許容される線形関係式の“2変量試験 (*bi-variate test*)”を行う．そして，採択された2数量関係式をリスト IE に含め，棄却された数量ペアをリスト NIE に含める．

(1-2) IE 内の互いに関連している各3つ組の関係式について“3組試験 (*triplet test*)”を行う．採択された3つ組について，全ての極大凸集合 MCS を導き，各 MCS 内の2数量関係式を1つの多数量式に組み上げる．組み上げられた各多数量式を新たな1つの項とし，IQ から各多数量式に含まれる数量を除く代わりにこれら新たな項をつけ加える．そして，$RQ=RQ+IQ$ とする．

(2-1) RQ 内の全ての数量ペアについて，許容される積関係式の2変量試験を行う．そして，採択された2数量関係式をリスト RE に含め，棄却された数量ペアをリスト NRE に含める．

(2-2) RE 内の互いに関連している各3つ組の関係式について3つ組試験を行う．採択された3つ組について，全ての極大凸集合 MCS を導き，各 MCS 内の2数量関係式を1つの多数量式に組み上げる．組み上げられた各多数量式を新たな1つの項とし，RQ から各多数量式に含まれる数量を除く代わりにこれら新たな項をつけ加える．

(2-3) RQ 内の間隔尺度量の線形関係式を表す項と他の数量・項との全てのペアについて，許容される対数関係式の2変量試験を行う．そして，採択された各2数量関係式を新たな1つの項とし，RQ から各2数量関係式に含まれる数量を除く代わりに，これら新たな項をつけ加える．

(3) $AQ=AQ+RQ$ とする．アンサンブル式の候補集合 CE について，新たな項が生成できなくなるまで以下のステップ (3-1) と (3-2) を繰り返す．

(3-1) AQ 内の全ての数量ペアについて，CE に含まれる各関係式の2変量試験を行う．そして，採択された2数量関係式をリスト AE に含める．採択された2数量関係式を1つの項に置き換えて表すことが可能な場合には，AQ においてその2数量を1つの項に置き換える．

(3-2) AE 内の各2数量関係式に恒等関係に基づく数学的許容性制約を適用し，1つの制約式にまとめられる関係式については，その関係式を新たな1つの項で表し，AQ から各制約式に含まれる数量を除く代わりに，これら新たな項をつけ加える．ステップ (3-1) に戻る．

対象系の法則式ないしモデル式の候補は AQ 内の項を AE 内の関係式を使って展開することにより得られる．

図 1 SDSのアルゴリズムの概要

4·1 2変量試験 (*bi-variate test*)

はじめに述べたように，科学的第一原理を表す法則式やモデル式を得るには，前節で述べた数学的許容性に加えて，客観性，普遍性，再現性，健全性，簡潔性などの規範を尤もらしく満たす関係式の探索が必要とされる．これらの規範の多くについて，対象やそれに類似する系に関する広範な検証を経なければ，十分に尤もらしさを確認することは困難である．しかし，与えられた対象とその実験可能な範囲に限定して規範検証を行うように，最低限の条件緩和を受け入れるならば，これらの成立判定は比較的容易である．この条件緩和の下では，客観性に関しては対象とその実験可能範囲で観測可能なデータに限って関係式を探索すれば満たされる．また，方程式が例外なく成立することを確認する普遍性の検証は，対象データにおいて反例などの矛盾がないかを調べる健全性の検証に包含される．再現性については，対象に関して繰り返し同一実験を行い，同じ方程式を得ることを検証すればよい．また，簡潔性を満たすために，極力簡単な関係式の探索をボトムアップ的に重ねて全数量間の関係を表す法則式を得る．

以上の条件緩和の下で，はじめにSDSは実験を通じて各数量間の擬レジーム式を探索する．ただし，一度に複数の数量間の擬レジーム式を探すと，候補として複雑な式を探索し上述の簡潔性を満たせなくなるばかりか，探索上の組み合わせ爆発を起こしてしまう．そのため，はじめにステップ (1-1) において，対象を表す n 個の数量 $x_1,x_2,...,x_n$ の内，間隔尺度の数量の集合 IQ に含まれる数量の任意のペア (x_i,x_j) を取り上げて関係を求める．この時，他の数量を各々の実験で設定可能な値域内の特定の値に固定する．そして，ペアの片方 x_i の値を実験可能な全ての値域内で変化させ，所定の関数形のデータ当てはめによって両者間の一般的関係を求める．この際に用いる関数形は，**Extended Product Theorem** のレジーム式に見られる2つの間隔尺度量間の線形関係である．即ち，数学的許容性を満たす

$$b_{ij}x_i+x_j=d_{ij} \qquad (1)$$

という線形関係式である．ここで，b_{ij} は他の数量に依存しない定数である．また，ステップ (2-1) において，比例尺度量及び間隔尺度量の線形項の集合 RQ に含まれる数量の任意のペア (x_i,x_j) についても，上記と同様の2項間のデータ当てはめが行われる．この際に用いる関数形

としては，**Extended Product Theorem** のレジーム式に見られる間隔尺度量の線形項や比例尺度量の間のべき乗積関係，即ち

$$x_i{}^{a_{ij}} x_j = d_{ij} \tag{2}$$

である．ここで，a_{ij} は他の数量に依存しない定数である．更にステップ (2-3) においても，RQ 内の間隔尺度量の線形項と他の数量・項との全てのペアについて，数学的に許容される対数関係

$$a_{ij} \log x_i + x_j = d_{ij} \tag{3}$$

のデータ当てはめが行われる．ここで，x_j は間隔尺度量の線形項，x_i は他の数量や項，a_{ij} は他の数量に依存しない定数である．

これらのデータ当てはめによって得られる式が，対象データの 2 項関係を矛盾なく表しているか，即ち式の健全性成立の尤もらしさを判定するために統計検定を行う．上記の式 (1) は線形式であり，両辺対数を取った式 (2) や対数項を中間変量で置き換えた式 (3) は線形式の形式を取り，何れも $y = ax + b$ の形に書き直すことができる．これについて以下の検定を適用する．なお，全ての検定において有意水準 $\alpha = 0.95$ としている．

(i) 当てはめの回帰成分分散 $S_R = \sigma(x,y)^2/\sigma(x,x)$ と残差成分分散 $S_e = \sigma(e_{xy})$ の間の F-検定を行う．ここで，$\sigma(x,y)$ は項 x,y の共分散を，$\sigma(e_{xy})$ は y に関する x の 2 項関係式の当てはめ誤差の平方和を表す．この際，回帰式は何れも 2 係数分の自由度を持つが，回帰成分分散は 1 自由度である平均値を差し引いた偏差に関する分散なので，S_R の自由度は $\Phi_R = 2-1 = 1$ である．また，当てはめ対象データ数を k とした時，データの全分散の自由度は平均値分を差し引いた $k-1$ であり，S_e の自由度は更に Φ_R を引いた $\Phi_e = k-1-\Phi_R = k-2$ である．これら自由度で規格化した $V_R = S_R/\Phi_R$ 及び $V_e = S_e/\Phi_e$ を定義し，それらの比 $F_0 = V_R/V_e$ を得る．もしこの値が有意水準 α における自由度 (Φ_R, Φ_e) に従う F-値よりも小さければ不合格とする．

(ii) 係数 a について，その絶対値自身より当てはめ期待誤差が大きいか否かの正規検定を行う．当てはめ誤差の正規性を仮定すれば，a は正規分布 $N(a, \sigma(e_{xy})/\sigma(x,x))$ に従うことが知られるので，これらに基づいて有意水準 α における a の期待誤差 δa を得る．もし，$|a| < \delta a$ であれば $a \simeq 0$ と考えられ，不合格とする．

検定 (i) は残差のパワーの意味で方程式がデータに示される 2 項間の関係を十分に説明しているかどうかを検定する．(ii) は 2 項を直接関係づける係数 a が意味のある値を示しているかどうかを検定する．いずれの検定に不合格でも，当てはめに用いられた関数形は棄却される．更に，再現性も加えた確認及びノイズや誤差の影響を排除するため，このような各項ペアに関する実験データの採取とデータ当てはめが，他の数量を適当な別の値に設定することにより得られる異なる条件下で $m = 10$ 回繰り返される．そして，以下の検定が行われる．

(iii) 各当てはめにおいて，2 項関係式が (1),(2) の検定によって採択されることを確認する．(1),(2) の検定は有意水準 $\alpha = 0.95$ で行われるため，実際には $m = 10$ 回の試行のうち 9 回以下しか採択されなければ不合格とする．

(iv) 係数 a について，その値の m 回の試行における分散 $\sigma(a,a)$ を，各試行における期待正規分布 $N(a, \sigma(e_{xy})/\sigma(x,x))$ の分散の m 回にわたる平均 $av(\sigma(e_{xy})/\sigma(x,x))$ で割って得た比 χ_0^2 は，自由度 $\Phi = m-1$ の χ^2-分布に従う．そこで有意水準 α について，χ_0^2 の値がその χ^2-分布の上下限より外に位置すれば不合格とする．

(iii) により，2 項関係が他の数量値やノイズの影響を受けずに安定し成立することを確認する．更に (iv) により，2 項関係そのもののみならず，その係数も一定の誤差範囲に常に安定して収まる関係が成立することを確認する．これらによって，当該 2 項間の関係が各試行において他の数量の値に依存しないで健全性と再現性を持つことが確認される．これらの検定を通過した関係式の各係数は，m 回の試行の平均値 $\bar{a}$ を期待値とする．

このようにして確認された科学的第一原理に基づく各ペア (x_i, x_j) に関する関係は，所定の有意水準 α の下で法則式であるための各種条件に関して尤もらしい式であると考えることができる．そこで，各式はそれらの関数形，係数及び誤差の情報と共に，リスト IE や RE にまとめられる．以上のような実験試行とデータ探索処理を “2 変量試験 (*bi-variate test*)” と呼ぶ．この試験では，数量の個数 n に対して $O(n^2)$ の実験や計算処理量で各 2 項関係式のデータ当てはめを行うことができる．また，2 変量試験を m 回繰り返すことを考慮すると，$O(mn^2)$ の実験や計算処理量となる．

4·2 3 つ組試験 (*triplet test*)

次にステップ (1-2) や (2-2) において，それぞれリスト IE ないし RE 内の数量や項の 2 項関係式について相互の整合性を確認し，整合な関係式同士をボトムアップに組み上げ，健全性と簡潔性を満たすことが尤もらしい擬レジーム式の全体または部分式を求める．そのために，数量や項の任意の 3 つ組について，その間に確認されている 2 項関係式の間で，互いの関係が矛盾しないものの集合を得る．まず，ステップ (1-2) においては，IQ 内の間隔尺度量のある 3 つ組 $\{x_i, x_j, x_k\}$ について，以下の関係が IE 内に得られているとする．

$$\bar{b}_{ij} x_i + x_j = d_{ij}, \qquad \bar{b}_{jk} x_j + x_k = d_{jk},$$
$$\bar{b}_{ki} x_k + x_i = d_{ki}$$

最初の式の x_j を 2 番目の式の x_j に代入することで，以下を得る．

$$-\bar{b}_{ij}\bar{b}_{jk} x_i + x_k = d_{jk} - \bar{b}_{jk} d_{ij}$$

従って，3番目の式と比較して3式が互いに無矛盾であるための以下の条件を得る．

$$1 = -\bar{b}_{ij}\bar{b}_{jk}\bar{b}_{ki}$$

従って，3つ組の2項関係式のお互いの整合性を確認するには，この式の成否を判定すればよい．同様にステップ(2-2)においても，RQ内の間隔尺度量の線形項または比例尺度量のある3つ組$\{x_i, x_j, x_k\}$について，以下の関係がRE内に得られているとする．

$$x_i{}^{\bar{a}_{ij}} x_j = d_{ij}, \quad x_j{}^{\bar{a}_{jk}} x_k = d_{jk}, \quad x_k{}^{\bar{a}_{ki}} x_i = d_{ki}$$

最初の式のx_jを2番目の式のx_jに代入することで以下を得る．

$$x_i{}^{-\bar{a}_{ij}\bar{a}_{jk}} x_k = d_{ij}{}^{-\bar{a}_{jk}} d_{jk}$$

従って，3番目の式と比較して3式が互いに無矛盾であるための条件として，ステップ(1-2)の場合と似た関係を得る．

$$1 = -\bar{a}_{ij}\bar{a}_{jk}\bar{a}_{ki}$$

しかしながら，データ当てはめ時のノイズや誤差により，たとえ整合な関係式同士であっても厳密にこれらの関係が成立するとは限らない．そこで，以下の統計的検定により整合性の判定を行う．

(v) 整合性判定条件式の両辺の値の差をErrとした時，各係数の期待誤差は正規分布するためErrも正規分布する．そこで，各係数の期待誤差から計算される両辺の期待標準偏差をExpとしたとき，正規分布$N(0, Exp^2)$に関する有意水準αの下での上下限値よりErrの絶対値が大きければ不合格とする．

ちなみに，ステップ(1-2)の場合，

$$\begin{aligned} Err_{ijk} &= 1 + \bar{b}_{ij}\bar{b}_{jk}\bar{b}_{ki}, \\ Exp_{ijk} &= \{(\overline{db}_{ij}\bar{b}_{jk}\bar{b}_{ki})^2 \\ &\quad + (\bar{b}_{ij}\overline{db}_{jk}\bar{b}_{ki})^2 + (\bar{b}_{ij}\bar{b}_{jk}\overline{db}_{ki})^2\}^{1/2} \end{aligned}$$

となる．ここで，各$\overline{db}$は，前節で述べた係数bの期待正規分布$N(a, \sigma(e_{xy})/\sigma(x,x))$から計算される$av(\sigma(e_{xy})/\sigma(x,x))$の平方根である．ステップ(2-2)でも同様に計算される．

それぞれIEないしはREに含まれる関係式の内，全ての3つ組の数量に関する組み合わせについて以上の検定が行われる．そして，任意の3つ組について整合な関係式が存在する数量集合の内，極大のものを全て導き出す．このような集合を“極大凸集合(MCS: *maximal convex set*)”と呼ぶ．あるMCSは，内部の関係式が全て整合な数量集合の内で，これ以上の包含集合が存在しない集合を表す．よって1つのMCSは整合性を満たす極大の関係である擬レジーム式に対応する．ただしREに含まれる関係式で何れのMCSにも属さないものは，その2数量のみに関するMCSとする．現実には，観測ノイズや誤差，各種の検定誤差要因により，本来はMCSを構成するはずである関係式の一部がREから欠落してしまっている可能性がある．このような場合には，大半の関係式が共通だが一部のみが異なるMCSが2つ以上存在することになる．そこで，一旦全てのMCSを求めた後，以下の条件を緩めた検定によりMCSを再構成する．

(vi) ある1つ以上の関係式を共有するMCSの集合を$S = \{M_1, M_2, ...\}$とする．もし$p \leq p_{th}$ならば，これらのMCSを全て結合，即ち$M_S = \cup_{M_i \in S} M_i$として，1つの新たな$MCS$に置き換える．ここで$p$は，$M_S$が本来の$MCS$であると仮定した際に欠落していたと見なされる関係式の個数である．

なお，pは以下の式で計算される．

$$p = f(M_s) + \Sigma_{A \in 2^S}(-1)^{|A|} f(\cap_{M_i \in A} M_i)$$

ここで，$f(M) = |M|(|M|-1)/2$は集合M内の任意の要素ペアの数であり，また2^SはSのべき集合，$|A|$は集合Aの濃度である．p_{th}は判定閾値であり経験的に$p_{th} = 3$とした．ただし，$|S|$が3より大きい場合には常にpが3を超えるため，本検定によるMCSの再構成は$|S| \leq 3$の条件下でのみ行う．以上のようにして全てのMCSが発見されれば，各々においてその内部の全ての関係式を**Extended Product Theorem**に示される式の形式に沿って組み合わせ，擬レジーム式を求める．この時，組み合わせる関係式同士において重複する$\bar{a}$や$\bar{b}$などの係数については，それらの平均値を代表値$\hat{a}$や$\hat{b}$とする．また通常，第一原理に基づく法則式の多くの係数が整数値を取ることが多いという経験的事実から，以下の検定により係数値を修正する．

(vii) 各係数の値$\hat{a}$や$\hat{b}$について，その期待誤差を$\hat{da}$ないしは$\hat{db}$としたとき，各々の値が最寄りの整数値に誤差の範囲で近いか否かを正規分布検定する．

もし，近いと判定された場合には，$\hat{a}$や$\hat{b}$を最寄りの整数値で置き換える．そして，これらの式の集合でIEやREの内容を置き換える．

このように法則式探索の早い段階において，**Extended Product Theorem**の数学的許容性制約を用いて，多数の数量を少数の擬レジーム式に基づく中間項$\{\Pi_i | i = 1, ..., n-r-s\}$にまとめることにより，法則式探索全体の計算量を節減することができる．以上のような計算処理を“3つ組試行(*triplet test*)”と呼ぶ．この部分は$O(n^3)$の処理量となる．

4·3 アンサンブル式の探索

次にステップ(3-1)において，REに含まれる各擬レジーム式に現れるΠや元々無次元量であった絶対尺度量同士の関係の探索を行う．まず，REの各擬レジーム式のΠに元々の絶対尺度量を加えた数量の集合をAQとおく．そして，AQ内の任意の項ペアに関して，関係式候補集合CEに含まれる関係式のデータ当てはめを行う．AQ内の数量は**Extended Product Theorem**には

従わないため，如何なる関係式でも候補になり得る．ただし，現状のSDSでは第一原理を表す法則式によく見られる以下のべき乗積式及び線形式のみの2項関係の探索を行う．候補式をより増やせば，一層探索能力は向上する．

$$\Pi_i^a \Pi_j = b$$
$$a\Pi_i + \Pi_j = b$$

そして，前述の2変量試験の検定(1)~(4)が適用される．この内，(1),(2), (3)の検定に合格した関係式のみが集合AEに入れられる．(4)の検定については擬レジーム式の場合とは異なり，検定に不合格であっても関係式を棄却せず，AEの中でその係数が他の数量に依存性を持つ事実を記録するに留める．以上により全ての2項関係の検定が終了した後，係数が他の数量に依存しないことが確認された関係式同士のみについて，擬レジーム式に関する場合と同様に整合な極大凸集合MCSを求める．そして，得られた各MCSについて，べき乗積式同士の場合は

$$\Theta_i = \prod_{x_j \in MCS_i} {x_j}^{a_j}$$

の形式に，また線形式同士の場合は

$$\Theta_i = \sum_{x_j \in MCS_i} a_j x_j$$

の形式にマージを行う．ここでΘは数量がマージされることで構成された中間項である．そして，AQにおいて統合された数量を除き，新たに得られた中間項Θを含める．この数量の統合化操作はそれ以上の統合化ができなくなるまで反復される．

その後ステップ(3-2)において，前述の表3に示した恒等関係に関する数学的許容性制約に従う関係をAQの各数量の3つ組の関係について調べる．即ち，AQ内の任意のペア項間の関係について，前述の擬レジーム式に関する2変量試験のデータ当てはめと検定(1)~(4)が適用される．そして，関係式に現れる係数が他の数量と依存関係を有し，かつ他と線形式の関係を持つ数量は集合Lへ，べき乗積式の関係を持つ数量は集合Pに含められる．これに基づいて，それぞれ表3に示された恒等関係に関し，数学的に許容される式を導出する．この時，もし1つの係数を除いて係数が数量に依存しない関係式が得られれば，依存する唯一の係数を新たな中間項ΘとしてAQに加え，AQから当該関係式に含まれる数量を除く．また，全ての係数が数量に依存しない定数であれば，当該式がアンサンブル式となる．これ以外であれば，当該関係式は棄却される．アンサンブル式が見つからない場合には，本副節の最初に戻って処理を繰り返す．

以上のSDSのアルゴリズムでは，前述の通り2変量試験の繰り返し数m，数量の個数nに関して$O(mn^2)$の複雑さを有し，3つ組試験が$O(n^3)$の複雑さを有する．

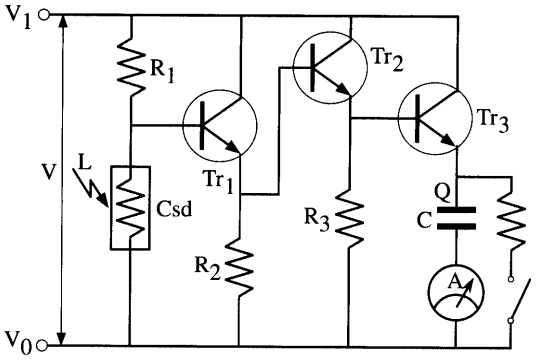

図2 光度計の電子回路

従って，高々3次の計算量で求解することが可能であるが，実際には実験試行を含む2変量試験部分の処理が大きなウエートを占めるため，おおよそ2次オーダーの処理量となる．

5. 複雑な対象への適用

本章では，SDSの手法を複雑な対象を含む2つの例題に適用した結果について説明する．何れの場合もSDSは実験環境シミュレータが有する対象式を知らずに，シミュレータを実験操作して種々の数量組のデータを取得する．シミュレータは各数量の絶対値に対して4%の標準偏差を有する正規性白色雑音をデータに混入するようになっている．1回の2変量試験では$k=10$個の(x_i, x_j)のデータが当てはめに用いられた．未知係数が2個の線形式当てはめに対して，$k=10$であれば最小二乗フィッティング上の問題は生じない．また，kを少なくすることで全体で必要となる実験回数をなるべく低く抑えた．

1つ目の例は，図2に示される3石トランジスタからなる一定時間内の光量増加率を測定する回路である．この系は以下の18数量からなる式で表される．

$$\left(\frac{R_3 h_{fe_2}}{R_3 h_{fe_2}+h_{ie_2}} \frac{R_2 h_{fe_1}}{R_2 h_{fe_1}+h_{ie_1}} \frac{rL^2}{rL^2+R_1}\right)(V_1-V_0) - \frac{Q}{C} - \frac{K h_{ie_3} X}{B h_{fe_3}} = 0$$

ここで，L, rは光量と光素子Csdの感度，X, K, Bは表示電流計針の位置，針バネの定数，磁石磁場の強さを表す．また，h_{ie_i}, h_{fe_i}は，それぞれi番のトランジスタのベース入力インピーダンスと電流増幅率を表す．他の記号は電子回路の標準的定義に従う．ここで，電位は間隔尺度量及び電流増幅率は無次元量であるため，$RQ = \{L, r, R_1, R_2, R_3, h_{ie_1}, h_{ie_2}, h_{ie_3}, Q, C, X, K, B\}$が比例尺度量の集合，$IQ = \{V_0, V_1\}$が間隔尺度量の集合，$AQ = \{h_{fe_1}, h_{fe_2}, h_{fe_3}\}$が絶対尺度量の集合となる．

はじめにステップ(1-1)で，実験データからV_1とV_2の関係の2変量試験が行われ$\Pi_1 = V_1 - V_0$が得られた．次にステップ(1-2)において，IQが3数量以上含まないので，そのままIQ内のV_0とV_1はΠ_1とまとめられ，

$RQ = RQ + IQ$ とされた．更にステップ (2-1) において，RQ 内の数量や中間項の間の2変量試験が行われた．例えば，2つの比例尺度 Q と X の関係 $Q^aX = d$ については，$m = 10$ 回の検定 (1) に関する F-値及び検定 (2) に関する係数 a の絶対値 $|a|$ とその不確定誤差 da がそれぞれ

```
 1:  F=25.93  |a|= 0.6682  da=0.0100
 2:  F=1.986  |a|= 0.6339  da=0.0346
 3:  F=0.748  |a|= 0.4840  da=0.1086
 4:  F=27.08  |a|= 0.6789  da=0.0100
 5:  F=1.421  |a|= 0.5833  da=0.0640
 6:  F=0.405  |a|= 0.3902  da=0.1539
 7:  F=0.860  |a|= 0.2351  da=0.6268
 8:  F=37.09  |a|= 0.7655  da=0.0100
 9:  F=1.843  |a|= 0.6226  da=0.0424
10:  F=6.324  |a|=-0.0494  da=0.0557
```

であり，幾つかの試行で $F < 5.317$ 及び $da > |a|$ の棄却条件に該当するため，2項関係式 $Q^aX = d$ の存在可能性はデータから否定された．ちなみに，係数 a の一定性を検定する検定 (4) では $\chi^2 = 39.54$ が得られ，これも $\chi^2 > 16.92$ の棄却条件に該当した．このような各項ペアに関する2変量試験の結果，2項関係の集合 RE が

$$
\begin{aligned}
RE = \{&L^{(1.999\pm0.010)}r = b_1, L^{(-1.999\pm0.010)}R_1 = b_2,\\
&r^{(-1.000\pm0.010)}R_1 = b_3, R_2^{(-1.000\pm0.010)}h_{ie_1} = b_4,\\
&R_3^{(-1.000\pm0.010)}h_{ie_2} = b_5, Q^{(-1.000\pm0.010)}C = b_6,\\
&h_{ie_3}^{(1.000\pm0.010)}X = b_7, h_{ie_3}^{(1.000\pm0.010)}K = b_8,\\
&h_{ie_3}^{(-1.000\pm0.010)}B = b_9, X^{(1.000\pm0.010)}K = b_{10},\\
&X^{(-0.999\pm0.010)}B = b_{11}, K^{(-1.000\pm0.010)}B = b_{12}\}
\end{aligned}
$$

と得られた．

更にステップ (2-2) において，3つ組試験により互いに整合な最大凸集合が求められ，それに従って以下の疑レジーム式の集合 RQ が導出された．

$$
\begin{aligned}
RQ = \{&\Pi_1 = V_1 - V_0, \Pi_2 = R_1 r^{-1.0} L^{-2.0},\\
&\Pi_3 = h_{ie_1} R_2^{-1.0}, \Pi_4 = h_{ie_2} R_3^{-1.0},\\
&\Pi_5 = h_{ie_3} XKB^{-1.0}, \Pi_6 = QC^{-1.0}\}
\end{aligned}
$$

次のステップ (2-3) では，この RQ 内で間隔尺度量の線形関係式を表す Π_1 と他の項との全ペアについて対数関係式を探索するが，何れのペアについても該当する関係は発見されなかった．また $AQ = AQ + RQ$ として，ステップ (3-1) において上記の Π と AQ 内の元々の無次元量の間で2変量試験を行い，各試験で採択された2項関係に関する極大凸集合 MCS に関する項のマージを行った．その結果，以下の5つの中間数量に集約され，$AQ = \{\Theta_1, \Theta_2, \Theta_3, \Theta_5\}$ となった．

$$
\begin{aligned}
\Theta_1 &= \Pi_2 h_{fe_1} = R_1 r^{-1.0} L^{-2.0} h_{fe_1}\\
\Theta_2 &= \Pi_3 h_{fe_2} = h_{ie_1} R_2^{-1.0} h_{fe_2}\\
\Theta_3 &= \Pi_4 h_{fe_3} = h_{ie_2} R_3^{-1.0} h_{fe_3}\\
\Theta_4 &= \Pi_5 + \Pi_6 = h_{ie_3} XKB^{-1.0} + QC^{-1.0}\\
\Theta_5 &= \Pi_1 \Theta_4^{-1.0}\\
&= (V_1 - V_0)(h_{ie_3} XKB^{-1.0} + QC^{-1.0})^{-1.0}
\end{aligned}
$$

最後にステップ (3-2) において，データから $\{\Theta_1, \Theta_5\}$, $\{\Theta_2, \Theta_5\}$, $\{\Theta_3, \Theta_5\}$ の3つのペアに線形関係が成立することが判明し，これに表3の最初の恒等関係制約を適用し以下を得た．

$$
\begin{aligned}
&\Theta_1\Theta_2\Theta_3 + \Theta_1\Theta_2 + \Theta_2\Theta_3 + \Theta_1\Theta_3 +\\
&\Theta_1 + \Theta_2 + \Theta_3 + \Theta_5 + 1 = 0
\end{aligned}
$$

先に求まった一連の式を本式に代入すれば，対象回路を表す第一原理モデル式 (4) が得られる．

更に物理法則以外の分野についても適用を試みた．前述のフェヒナーの法則について適応した結果，比例尺度量である音の周波数 f と間隔尺度量である人間の音程 I に関して，以下の2種類の関係式候補が得られた．

$$I = \alpha f^{\beta} + \gamma \text{ または } I = \alpha \log f + \beta$$

この内，後者がフェヒナーの法則に合致する．また別の例として，部屋の容積 R とその中で我々が居る位置の照度 L_p，同じくそこから見た窓の立体角 W_p に関して，我々が感じる部屋の主観的開放感 S_p が以下の式で関係づけられるとされている [乾 72].

$$S_p = cRL_p^{0.3} W_p^{0.3},$$

ここで，S_p はマグニチュード推定法と呼ばれる心理学的測定によって被験者から得られ，比例尺度のスケールタイプを有する．その他の物理的数量は R, L_p が比例尺度量，W_p が絶対尺度量である．これについて本アルゴリズムを適用したところ上記の式が正確に求められた．

6. 議論及び関連研究

SDS の主な特徴は，対象を表す数量の個数に対する少ない計算量，対象の複雑さや規模の大きさに対するスケーラビリティ，観測ノイズや誤差に対するロバスト性，そして幅広い適用性である．表4に種々の対象に関する本アルゴリズムの計算量評価結果をまとめる．理想気体状態方程式の導出に関する結果を基準にした場合の相対的な CPU 計算時間が示されている．比較のために，本研究で提唱した数学的許容制約を用いずに，種々の数式形式の組み合わせについてしらみ潰し探索を行う従来型の ABACUS の計算量も掲載した [Falkenhainer 86]．これらの例では，本アルゴリズムの計算量はおおむね対象とする法則式やモデル式に含まれる数量個数の2乗に比例する．この結果は，先の理論的考察の結果と一致する．これに対し，従来型のものは全体として組み合わせ爆発的な探索を行うため，数量の個数が多いと SDS に比して非常に多くの計算時間を要する．ただし，対象方程式の形式に依存して探索範囲が大きく変化するため厳密な計算量の見積は困難である．実際，数量個数が同じ8個である運動量保存則と運動エネルギー保存則の例でもかなり計算時間の違いを生じる．しかしながら，定性的には概ね数量個数に関する指数乗に比例する探索が必要となる

表 4 計算量とノイズロバスト性に関する評価

例	n	TC(S)	TC(A)	NL(S)
Fechner	2	0.34	-	±45%
開放感	4	1.06	-	±40%
理想気体	4	1.00	1.00	±40%
運動量保存	8	6.14	22.7	±35%
Coulomb	5	1.63	24.7	±35%
Stoke's	5	1.59	16.3	±35%
運動エネルギー	8	6.19	285.	±30%
電子回路	18	21.9	-	±20%

n: 属性量の数, TC(S): CPU 計算時間, TC(A): 比較対象 ABACUS の CPU 計算時間, NL(S): ノイズ許容限界.

はずであり，電子回路の例であれば相対的 CPU 計算時間にして数千倍から数億倍の計算時間が必要になると推定され，従来型の手法では，実規模問題について実際的な時間内に解を得ることは困難と思われる．

また表の右端列には，各数量の観測に人工的に混入したガウスノイズに対する本アルゴリズムのロバスト性が示されている．各々の事例について 10 回の評価を行い，その内 8 回以上正しい式が得られるための最大のノイズ振幅標準偏差の数量絶対値に対する相対値を示している．従来型の ABACUS は，僅かな実験ノイズに対しても極めて脆弱であることが報告されているのに対し，本アルゴリズムがかなり複雑な法則式に関しても非常にロバストな結果を与えることが示されている．これは本アルゴリズムが 2 変量試験という 2 数量ずつの単純な関係のみを実験で求めることに加えて，3 つ組試験と法則式の数学的許容性制約を用いることで，見かけ上は当てはまるが制約に合致しない虚偽の関係式を効果的に排除し得るためと考えられる．

以上のように SDS は，10 個を超える数量からなる比較的規模の大きい問題に関しても，法則式やそれによって構成されるモデル式を実際的な条件で発見可能である．10 数量を超える規模の対象をまとめて一度にモデル化することは，通常，科学者や技術者にとっても困難な場合が多く，本研究のような体系的手法を採用するメリットは大きいと考えられる．また，先の主観的開放感の例のように，各数量の単位が明確でない場合でもスケールタイプは明らかであることが殆どであり，本アルゴリズムが幅広い適用性を有することがわかる．単位次元を用いる COPER などは，このような場合には適用困難である [Kokar 86].

SDS の弱点は，発見される法則式やモデル式のクラスにある程度限界があることである．1 つの限界は，レジーム式やアンサンブル式が "*read-once formulae*" でなければならないことである [Bshouty 94]．これは各数量が式の中で高々1 カ所にしか現れないように整理できる数式のことである．例えばブラックの比熱法則と呼ばれる式は，M_1, M_2 を同じ物質でできた 2 つの物体それぞれの質量，T_1, T_2 をそれぞれの初期温度，T_f を両方を接合した後の平衡温度とすると

$$\frac{M_2}{M_1} = \frac{T_f - T_1}{T_f - T_2}$$

という形式をとる．この式を変形して T_f が 1 カ所にのみ現れるようにすると，必ず他の何れかの数量が 2 カ所以上に現れてしまうので *read-once formula* ではない．2 つ目の限界は，恒等関係制約で用いる 2 項関係集合 CE の内容が線形やべき乗積関係のみであるため，数量間の関係が加算や減算，乗算，除算，指数，対数といった初等的な演算子や関数関係で表される場合にしか適用できないことである．3 つ目に，図 1 のステップ (3-1) におけるアンサンブル式探索の 2 変量試験における 2 項関係式は，単純な 2 項演算子で表される関係に限定され，もっと複雑な一般的 2 項演算子 $F(x,y)$ に関する扱いが困難なことである．殆どの法則式やモデル式はこれらの限界には抵触せず SDS は問題なく動作するが，このような限界は将来は取り除かれるべきである．3 つ目の限界については，D. Bshouty 等が複雑な $F(x,y)$ を初等的 2 項演算子で関連づけられた 2 つの 1 項演算子 $g(x), h(y)$ と他の 1 項演算子 $f(\cdot)$ に，関係の不変性から例えば $f(g(x)+h(y))$ のように分解する方法を提案している [Bshouty 94]．法則発見の場合には，事前には $F(x,y)$ の形式が未知なのでこの方法は適用困難であるが，このような不変性原理に基づいてより一般的な関係を扱えるように拡張することは可能であると考えられる．2 つ目の限界に関しては，より広範な 2 項関係に関して恒等関係制約の考察を進め，CE の内容をより豊富なものにして行くことで解決可能である．1 つ目の限界に関しても，何らかの不変性や恒等性に関する原理を導入することで，緩和可能であると予想し研究を進めている．

7. 結　　言

SDS はスケールタイプ制約や恒等関係制約，2 変量試験や 3 つ組試験などに基づいて，新しい法則発見システムの枠組みを確立した．この枠組みは，少ない計算量，高いロバスト性やスケーラビリティ，広い適用性を持つことが確認された．殆どの科学的法則発見が大量の実験や観測の積み上げによってなされたことは間違いない．しかし，科学者達はデータだけに依存して法則発見を行って来たのではなく，光速度不変性やダイナミクスの時間対称性，状態連続性などの様々な数学的許容性に関する知見を駆使して来た．スケールタイプ制約や恒等関係制約は，幅広い適用性を有するそのような知見の例である．今後は，更に大規模なシステムへの適用や非物理領域での全く新しい法則の発見などを目指す予定である．

◇ 参 考 文 献 ◇

[Akaike 68] Akaike, H.: *On the use of a linear model for the identification of feedback systems*, Annals of institute for

Statistical Mathematics, Vol.20, pp.425–439 (1968).
[Bridgman 22] Bridgman, P.W.: *Dimensional Analysis*, Yale University Press, New Haven, CT (1922).
[Bshouty 94] Bshouty, D. and Bshouty, N.H.: *On Learning Arithmetic Read-Once Formulas with Exponentiation*, Proc. of the Seventh Annual ACM Conference on Computational Learning Theory, pp.311–317 (1994).
[Buckingham 14] Buckingham, E.: *On physically similar systems; Illustrations of the use of dimensional equations*, Physical Review, Vol.IV, No.4, pp.345–376 (1914).
[Chatterjee 86] Chatterjee, S. and Hadi, A.S.: *Influential Observations, High Leverge Points, and Outliers in Linear Regression*, Statistical Science, pp.379–416 (1986).
[Descartes 1637] Descartes, R.: *Discours de la Methode* (1637), 落合太郎 訳: 方法序説, 岩波書店.
[Dzeroski 94] Dzeroski, A.: *Discovering Dynamics: From Inductive Logic Programing to Machine Discovery*, Journal of Intelligent Information Systems, Vol.3, pp.1–20 (1994).
[Falkenhainer 86] Falkenhainer, B.C. and Michalski, R.S.: *Integrating Quantitative and Qualitative Discovery: The ABACUS System*, Machine Learning, pp.367–401, Boston, Kluwer Academic Publishers (1986).
[Feynman 65] Feynman, R. P.: *The Character of Physical Law*, Charles E. Tuttle Co. Inc. (1965).
[Glymour 87] Glymour, C. et al.: *Discovering Causal Structure*, Academic Press (1987).
[Ivakhnenko 70] Ivakhnenko, A.G.:*Heuristic Self-Organization Problems of Engineering Cybernetics*, Automatica, Vol.6, pp.207–219 (1970).
[乾 72] 乾正雄, 宮田紀元, 渡辺圭子: 開放感に関する研究, 日本建築学会論文報告集, No.193, pp.51–57 (1972).
[Koehn 86] Koehn, B. and Zytkow, J.M.: *Experimeting and theorizing in theory formation*, Proc. of the International Symposium on Methodologies for Intelligent Systems, pp.296–307, ACM SIGART Press (1986).
[Kokar 86] Kokar, M.M.: *Determining Arguments of Invariant Functional Descriptions*, Machine Learning, pp.403–422, Kluwer Academic Publishers (1986).
[Langley 81] Langley, P.W.: *Data-driven discovery of physical laws*, Cognitive Science, Vol.5, pp.31-54 (1981).
[Langley 87] Langley, P.W., Simon, H.A., Bradshaw, G. and Zytkow, J.M.: *Scientific Discovery; Computational Explorations of the Creative Process*, Cambridge, Massachusetts: MIT Press (1987).
[Ljung 87] Ljung, L.: *System Identification Theory for the User*, PTR Prentice Hall, Englewood Cliffs, New Jersey (1987).
[Luce 59] Luce, R.D.: *On the Possible Psychological Laws*, The Psychological Review, Vol.66, No.2, pp.81–95 (1959).
[Newton 1686] Newton, I.: *Principia*, Vol.II, The System of the World (1686), Translated into English by A. Motte (1729), London, England, University of California Press, Ltd. (Copyright 1962).
[Nordhausen 90] Nordhausen, B. and Langlay, P.W.: *An Integrated Approach to Empirical Discovery*, Computational Models of Scientific Discovery and Theory Formation, Morgan Kaufman Publishers, Inc. (1990).
[Rissanen 78] Rissanen, J.: *Modeling By Shortest Data Description*, Automatica, Vol.14, pp.465–471 (1978).
[Saito 97] Saito, K. and Nakano, R.: *Law Discovery Using Neural Networks*, Proc. of IJCAI-97: Fifteenth International Joint Conference on Artificial Intelligence, Vol.2, pp.1078–1083 (1997).
[Simon 97] Simon, H. A. et al.: *Scientific discovery and simplicity of method*, Artificial Intelligence, Vol.91, pp.177-181 (1997).
[Stevens 46] Stevens, S.S.: *On the Theory of Scales of Measurement*, Science, Vol.103, No.2684, pp.677–680 (1946).
[Valdes-Perez 99] Valdes-Perez, R. E.: *Principles of human-computer collaboration for knowledge discovery in science*, Artificial Intelligence, Vol.107, pp.335–346 (1999).
[Washio 96] Washio, T. and Motoda, H.: *Scale-Based Reasoning on Possible Law Equations*, Working Papers of QR'96: Tenth Int. Workshop on Qualitative Reasoning, Stanford Sierra Camp, Fallen Leaf Lake, California, USA, pp.255–264 (1996).
[Washio 97] Washio, T. and Motoda, H.: *Discovering Admissible Models of Complex Systems Based on Scale-Types and Identity Constraints*, Proc. of Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97), Vol.2, pp.810–817 (1997).
[鷲尾 98a] 鷲尾隆, 元田浩: 属性変量の尺度認知に基づく構成的法則発見手法, 認知科学, Vol.5, No.2, pp.80–94 (1998).
[Washio 98b] Washio, T. and Motoda, H.: *Discovering Admissible Simultaneous Equations of Large Scale Systems*, Proc. of AAAI'98: Fifteenth National Conference on Artificial Intelligence, pp.189-196 (1998).
[Washio 99] Washio, T. and Motoda, H.: *Extension of Dimensional Analysis for Scale-types and its Application to Discovery of Admissible Models of Complex Processes*, Working Papers of Similarity Workshop'99: The Second International Workshop on Similarity Methods, University Stuttgart, Stuttgart, Germany, pp.129–147 (1999).
[Wasserman 89] Wasserman, P.D.: *Neural Computing: Theory and Practice*, New York: Van Nostrand Reinhold (1989).

〔担当委員：山口高平〕

1999年10月30日 受理

著者紹介

鷲尾 隆(正会員)は，前掲(Vol.15, No.1, p.186)参照.

元田 浩(正会員)は，前掲(Vol.15, No.1, p.186)参照.