# Scale-Based Reasoning on Possible Law Equations

**Takashi Washio**
Research Center for Safety Science, Mitsubishi Research Institute, Inc.
2-3-6, Otemachi, Chiyoda-ku, Tokyo, 100, Japan
e-mail: washio@mri.co.jp

**Hiroshi Motoda**
Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567, Japan
e-mail: motoda@isir2.sanken.osaka-u.ac.jp

## Abstract

The objective of this paper is to propose a novel approach to automatically reason formulae of laws and their solutions. Our approach takes an intermediate position between the deductive approaches such as dimension-based and symmetry-based reasoning and the empirical approaches such as BACON. It does not require a priori insights into the objective system but data through measurement, and hence it can be applied to various domains like BACON and is not limited to physics. In spite of its data-driven feature, the solutions of the formulae obtained by our approach are ensured to be sound similarly to the dimension-based approach. The basic idea is the combined use of deductive "*scale-based reasoning*" and data-driven reasoning. Especially, the scale-based reasoning is the main part in this study. The features of our approach are demonstrated by deriving the basic formulae of the ideal gas law and Black's specific heat law. The scale-based reasoning may provide a basis to develop qualitative models of ambiguous domains such as biology, psychology, economics and social science. This will also contribute to the research area of knowledge discovery.

## Introduction

Over the many years, methods to automatically derive formulae of physical laws and their solutions have been explored on the basis of physical dimensions of quantities. One of the early work is a method called dimensional analysis based on the product theorem(Bridgman 1922; Bhaskar & Nigam 1990).

**Product Theorem** *Assuming absolute significance of relative magnitudes of physical quantities, the function $f$ relating a secondary quantity to the appropriate primary quantities, $x, y, ....$ has the form: $f = Cx^a y^b z^c ....$, where $C, a, b, c, ...$ are constants.*

A simple example of the dimensional analysis is to derive the basic formula of oscillation period $t[T]$ from mass $m[M]$ and spring constant $K[MT^{-2}]$ in case of horizontal oscillation of a mass by a frictionless spring. The product theorem applies and it says that $t = C_1 m^a K^b$. The exponent $b$ must be -1/2 to equalize the dimension of $[T]$ on the both sides. Also, $a$ must be $-b$ to cancel out the dimension of $[M]$ on r.h.s., hence $a = 1/2$. Consequently, we obtain $t = C_1(m/K)^{1/2}$. There is another important theorem that is called as Buckingham $\Pi$-theorem(Buckingham 1914; Bhaskar & Nigam 1990).

**Buckingham $\Pi$-theorem** *If $\phi(x, y, ....) = 0$ is a complete equation, then the solution can be written in the form $F(\Pi_1, \Pi_2, ..., \Pi_{n-r}) = 0$, where $n$ is the number of arguments of $\phi$, and $r$ is the basic number of dimensions in $x, y, z.....$ For all $i$, $\Pi_i$ is a dimensionless number.*

Basic dimensions are such dimensions as length $[L]$, mass $[M]$ and time $[T]$, scaling quantities independently of other dimensions in the given $\phi$. This theorem can be used together with the product theorem to obtain the oscillation period $t[T]$ of a simple pendulum depicted in Fig. 1 from its stick length $l[L]$, gravity acceleration $g[LT^{-2}]$ and deviation angle $\theta$ [no dimension]. We can find two dimensionless quantities $\Pi_1 = t(g/l)^{1/2}$ and $\Pi_2 = \theta$, and derive the basic formula of the solution as $F(\Pi_1, \Pi_2) = F(t(g/l)^{1/2}, \theta) = 0$ based on the theorem. Using these dimensional analysis techniques, Bhaskar and Nigam introduced a concept "*regime*" which is a formula $\rho_i(\Pi_i, x, y, ...) = 0$ defining a dimensionless quantity $\Pi_i$(Bhaskar & Nigam 1990). In the above example, $t(g/l)^{1/2}$ and $\theta$ are the regimes. Also, they defined an "*ensemble*" which is a set of regimes contained in a complete equation $F(\Pi_1, \Pi_2, ...., \Pi_{n-r}) = 0$. In the example,$F(t(g/l)^{1/2}, \theta) = 0$ is an ensemble. The regime refers to a decomposable subprocess. Every quantity in a regime interacts with any quantities outside of the regime via a dimensionless quantity. An ensemble stands for a complete physical process in the system. Based on these definitions, a qualitative reasoning method was proposed, where irrelevant argument descriptors are eliminated when a physical formula is derived(Bhaskar & Nigam 1990). An advantage of the dimension-based methods is that a sound set of solutions is obtained for the physical formula of each regime without much knowledge of the system configuration. However, this method utilizes the physical insights into the objective system that is described by the unit dimension of each quantity. For instance, the unit $[ML/T^2]$ of force $f$ and the unit $[L/T^2]$ of acceleration $\alpha$ implicitly state the well-known relation $f = m\alpha$, where $m$ is a mass having the unit $[M]$.

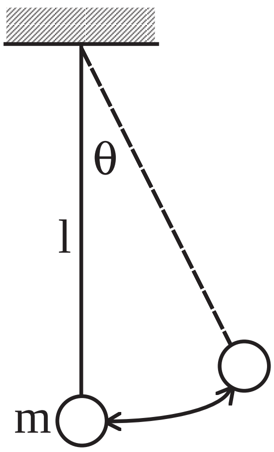


Figure 1: A simple pendulum

Another method to automatically derive physical formulae is symmetry-based approach. Ishida applied the principle of symmetry to physical domains, and proposed the symmetry-based reasoning (Ishida 1995). The system he developed searches for the invariant physical formula $O$ under a given set of isomorphic mappings $\{T_i | i = 1, ..., n\}$. Each $T_i$ holds the symmetric relation of $O$, *i.e.*, $O = T_i(O)$. The types of mapping $T_i$ supported by the system are translation and dilation of quantity scales and permutation of the positions of quantities in the formula. In the example of the simple pendulum, we can apply further constrains of symmetry on the form $F(t(g/l)^{1/2}, \theta)$. One is phase translatory symmetry on $t$, and this gives $F(t(g/l)^{1/2}, \theta) = F((t(g/l)^{1/2} + 2m\pi, \theta)$, where $m$ is an integer. Another is mirror symmetry of the time and angle $\theta$ giving $F(t(g/l)^{1/2}, \theta) = F(-t(g/l)^{1/2}, -\theta)$. A formula satisfying these constraints is $\theta = \sin(t(g/l)^{1/2})$. This method also utilizes the explicit knowledge of physical characteristics of the object in terms of symmetry. A disadvantage may be the lack of the soundness and completeness of the solution, because the way the search is made is based on heuristics.

In contrast with these physical knowledge based methods, a challenge to discover the formulae of the first principles involved in the objective system has been made by BACON (Langley & Zytkow 1989) in the data-driven framework. BACON derives formulae by using heuristic search algorithm. The objects in various domains can be modeled, since it does not use any a priori knowledge of the objects. However, the heuristics in the search algorithm does not have theoretical firm bases, and is limited to enumerate polynomial and meromorphic formulae only. Accordingly, the set of the solutions obtained is not ensured to be sound and complete.

The above past work employs the following assumptions.

**Assumption 1** *The relation among quantities under consideration is represented by a complete equation.*

Though Bhaskar and Nigam demonstrated the applicability of their dimension-based method to the systems consisting of multiple complete equations, the specification of a set of quantities is required for each complete equation in the framework, and thus this assumption must be maintained. Their dimension-based approach also requires the following assumptions.

**Assumption 2** *The type of scales of physical quantities is limited to ratio scale.*

**Assumption 3** *Given a regime $\rho_i(\Pi_i, x, y, ...) = 0$, either one of the following conditions holds.*

*1) It is a unique regime of a complete equation.*

*2) For each quantity $x$ in $\rho_i$, any other regimes do not contain $x$, or any other regimes are related in such a way that $x$ does not change $\Pi_i$ from outside of $\rho_i$.*

The assumption 2 is the restatement of the assumption in the product theorem, *i.e.*, the absolute significance of relative magnitudes of physical quantities. Though the quantities of the ratio scale are quite common in various domains, another scale type of quantities called as "*interval scale*" is also widely encountered as explained later. The assumption 3 was recently pointed out by Kalagnanam and Henrion(Kalagnanam, Henrion, & Subrahmanian 1994). The data-driven approach of BACON assumes the following environment of data measurement.

**Assumption 4** *The measurements on the relation in any subset of quantities in a complete equation can be made while holding the other variables constant under an experimental environment, and the measured data can be sequentially used to reason the formula relating the variables.*

The objective of this paper is to propose a novel approach to automatically reason formulae of laws and their solutions. Our approach takes an intermediate position between the deductive approaches such as dimension-based and symmetry-based reasoning and the empirical approaches such as BACON. The following theoretical aspects removing the limitations of assumptions 2 and 3 are newly introduced.

1) The sound relations among quantities of ratio and interval scales within a regime are characterized.
2) The product theorem is extended to involve interval scale quantities.
3) An algorithm to identify regimes partitioning the set of quantities in an ensemble is proposed.

Based on these aspects, our approach can derive a sound set of solutions of the formulae for each regime similarly to the dimension-based approach while maintaining the advantage of the data-driven approach such as BACON, *i.e.*, it does not require a priori insights into the objective system but data obtained through the measurements, and it can be applied to various domains and is not limited to physics. The basic idea of our approach is the combined use of deductive scale-based reasoning and data-driven reasoning. Especially, the scale-based reasoning is the main part in our present study. The other two assumptions remain the same in our approach at present.

## Basic Principle of Scale-Based Reasoning

At the end of the 19th century, Helmholtz originated a research field of "*measurement theory*" (Helmholtz 1887). Since after he tried to obtain a systematic model of measurement processes, many literatures have been published on that topic in the 20th century. Among those literatures, Stevens defined the measurement process as "*the assignment of numerals to object or events according to rules*" (Stevens 1946). He claimed that different kinds of scales and different kinds of measurement are leaded, if numerals can be assigned under different rules. And he categorized the quantity scales for the measurement based on (a) the operation rule of the assignment and (b) the mathematical properties (group structure) of the quantities. The scale categories he defined are reproduced in Table 1. The scales having analytical group structures are the interval scale and the ratio scale, and these are the majorities of quantities in physical domains, some psychological, economical and sociological domains. The examples of the interval scale quantities are temperature in Celsius and Fahrenheit, time (not time interval), and sound tone (proportional to the order of white keys of a piano). The zero point level of their scales are not absolute, and are changeable by human's definitions. The example of the ratio scale quantities are physical mass, absolute temperature, interval length, acceleration, time interval, frequency, population (large enough to view as continuous) and currency value. Each has an absolute zero point, and the ratio of different two values $x_1/x_2$ is invariant against their unit change, *i.e.*, $(kx_1)/(kx_2) = x_1/x_2$. We should note that the scale is different from the dimension. On the other hand, the scale just represents the definition of the measurement rule. We do not know what the force (ratio scale) divided by acceleration (ratio scale) means within the knowledge of scales. Another point we should clarify is that dimensionless quantities are the quantities of absolute scale having the group structure of $x' = x$, because any change of units is not defined.

Luce claimed that the basic formula of the functional relation between two quantities can be determined by

Table 1: Scale types

| Scale | Basic Empirical Operations | Mathematical Group Structure |
|---|---|---|
| Nominal | Determination of equality | Permutation group $x' = f(x)$, where $f(x)$ means any one-to-one substitution. |
| Ordinal | Determination of greater or less | Isotonic group $x' = f(x)$, where $f(x)$ means any monotonic increasing function. |
| Interval | Determination of equality of intervals or differences | Generic linear group $x' = kx + c$ |
| Ratio | Determination of equality of ratios | Similarity group $x' = kx$ |

the scale features of these two quantities, if the quantities are not coupled through any dimensionless quantities(Luce 1959). Two examples will make his theory comprehensive. Suppose $x$ and $y$ are both ratio scale quantities, and $y$ is defined by $x$ through a continuous functional relation $y = u(x)$. First, we assume that the form of $u(x)$ is logarithmic, *i.e.*, $y = \log x$. We can multiply a positive constant $k$ to $x$, *i.e.*, a change of unit, without violating the group structure of the ratio scale quantity $x$ shown in Table 1. However, this leads $u(kx) = \log k + \log x$, and this fact causes the shift of the origin of $y$ by $\log k$, and violates the group structure of $y$ which is the ratio scale quantity. Hence, the functional relation from $x$ to $y$ must not be logarithmic. Next, we assume that $y = u(x) \equiv x^3$. In this case, the change of unit leads $u(kx) = k^3x^3$, and this causes the change of unit of $y$ by the factor of $k^3$. Since this is admissible for the group structure of $y, y = x^3$ is one of the possible relations between $x$ and $y$.

Now, our interest moves to derive the most generic formula of the continuous functional relations between two ratio scale quantities $x$ and $y$. As the admissible transformations of $x$ and $y$ in their group structures are $x' = kx$ and $y' = Ky$ respectively, the relation of $y = u(x)$ becomes as $y' = u(x') \leftrightarrow Ky = u(kx)$. The factor $K$ of the changed unit of $y$ may depend on $k$, but it shall not depend upon $x$, so we denote it by $K(k)$. Consequently, we obtain the following constraints on the continuous function $u(x)$.

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

where $k > 0$ and $K(k) > 0$ as these are the factors of the changed units. The constraints for the different combinations of the scale types are summarized in Table 2 (Luce 1959). Luce derived each solution of $u(x)$ under the condition of $x \geq 0$ and $u(x) \geq 0$. We extended his theorems to cover the negative values of $x$ and $u(x)$. Our new theorems are represented bellow.

**Lemma 1** *The function $u(x)$ satisfying Constraint 1*

Table 2: Constraints on functional relations due to the scale characteristics

| $C_n$ No. | Scale Types Independent variable | Dependent (Defined) variable | Constraints | Comments* |
|---|---|---|---|---|
| 1 | ratio | ratio | $u(kx) = K(k)u(x)$ | $k > 0, K(k) > 0$ |
| 2 | ratio | interval | $u(kx) = K(k)u(x) + C(k)$ | $k > 0, K(k) > 0$ |
| 3 | interval | ratio | $u(kx + c) = K(k, c)u(x)$ | $k > 0, K(k, c) > 0$ |
| 4 | interval | interval | $u(kx + c) = K(k, c)u(x) + C(k, c)$ | $k > 0, K(k, c) > 0$ |

*c and C can be any real numbers.

*in Table 2 has the following property.*
$x = 0 \to y = u(0) = 0$ *and* $y = u(x) = 0 \to x = 0$
*Proof. If $x = 0$, then Constraint 1 becomes $u(0) = K(k)u(0)$. Thus, if $u(0) \neq 0$ then $K(k) \equiv 1$, i.e., $u(kx) = u(x)$. However, $u$ having this group structure must be an absolute scale, i.e., dimensionless, and this is contradictory to the definition of $u$. Therefore, $y = u(0) = 0$. If $y = u(x) = 0$, then Constraint 1 becomes $u(kx) = K(k)u(x) = 0$. This relation holds independent of $k$. Therefore, $x = 0$.*

**Lemma 2** *The function $u(x)$ satisfying Constraint 1 in Table 2 has the following property. For $x > 0, u(x)/u(1)$ is always positive. For $x < 0, u(x)/u(-1)$ is always positive.*

*Proof. From the Lemma 1, $y = u(x)$ does not change its sign except at the origin $(x, y) = (0, 0)$. Thus, the sign of $u(x)$ is identical with that of $u(1)$ when $x > 0$. Therefore, $u(x)/u(1) > 0$. Also, the sign of $u(x)$ is identical to that of $u(-1)$ when $x < 0$. Therefore, $u(x)/u(-1) > 0$.*

**Theorem 1** *If the independent and dependent continua are both ratio scales, then $u(x) = a_*|x|^\beta$, where $a* = a_+$ for $x \geq 0, a_* = a_-$ for $x < 0$ and where $\beta$ is independent of the units of both quantities.*
*Proof.*

*1) In case of $x > 0$, set $x = 1$ in Constraint 1, then $u(k) = K(k)u(1)$, so $K(k) = u(k)/u(1)$. Thus, Constraint 1 becomes $u(kx) = u(k)u(x)/u(1)$. Based on Lemma 2, let $v(x) = \log[u(x)/u(1)]$, then $v(kx) = \log[u(kx)/u(1)] = \log[\{u(k)/u(1)\}\{u(x)/u(1)\}] = \log[u(k)/u(1)] + \log[u(x)/u(1)] = v(k) + v(x)$. Since $u$ is continuous, so is $v$, and it is well known that the only continuous solutions to the last functional equation are of the form $v(x) = \beta \log x = \log x^\beta$. Thus, $u(x) = u(1)x^\beta = a_+x^\beta$.*

*2) In case of $x < 0$, set $x = -1$ in Constraint 1, then $u(-k) = K(k)u(-1)$, so $K(k) = u(-k)/u(-1)$. Thus, Constraint 1 becomes $u(kx) = u(-k)u(x)/u(-1)$. Based on Lemma 2, let $v(x) = log[u(x)/u(-1)]$, then $v(kx) = \log[u(kx)/u(-1)] = \log[\{u(-k)/u(-1)\}\{u(x)/u(-1)\}] = \log[u(-k)/u(-1)] + log[u(x)/u(-1)] = v(-k) + v(x)$. The only continuous solutions to the last functional equation are of the form $v(x) = \beta \log(-x) = \log(-x)^\beta$ as well. Thus, $u(x) = u(-1)(-x)^\beta = a_-(-x)^\beta$.*

*From 1), 2) and Lemma 1, $u(x) = a_*|x|^\beta$, where $a_* = a_+$ for $x \geq 0$ and $a_* = a_-$ for $x < 0$. We observe that since $u(kx) = a_*k^\beta|x|^\beta = a'_*|x|^\beta, \beta$ is independent of the unit of $x$, and clearly independent of the unit of $u$.*

**Theorem 2** *If the independent continuum is a ratio scale and the dependent continuum an interval scale, then either $u(x) = \alpha \log |x| + \beta_*$, where $\alpha$ is independent of the unit of the independent quantity and where $\beta_* = \beta_+$ for $x \geq 0$ and $\beta_* = \beta_-$ for $x < 0$, or $u(x) = \alpha_*|x|^\beta + \delta$, where $\alpha_* = \alpha_+$ for $x \geq 0$ and $\alpha_* = \alpha_-$ for $x < 0, \beta$ is independent of the units of the both quantities, and $\delta$ is independent of the unit of the independent quantity.*

*Proof. In solving Constraint 2, there are two possibilities to consider.*

*1) If $K(k) \equiv 1$ for any $k$, then define $v(x) = e^{u(x)}$. Constraint 2 becomes $v(kx) = D(k)v(x)$, where $D(k) = e^{C(k)} > 0$ and $v$ is continuous, positive and non constant because $u$ is continuous and non constant. By Theorem 1, $v(x) = \gamma_*|x|^\alpha$, where $\alpha$ is independent of the unit of $x$ and $\gamma_* = \gamma_+ > 0$ for $x \geq 0$ and $\gamma_* = \gamma_- > 0$ for $x < 0$ because, by definition, $v > 0$. Taking logarithms, $u(x) = \alpha \log |x| + \beta_*$, where $\beta_* = \log \gamma_*, \beta_* = \beta_+$ for $x \geq 0$ and $\beta_* = \beta_-$ for $x < 0$.*

*2) If $K(k) \not\equiv 1$ for some $k$, then let $u$ and $u^*$ be two different solutions to the problem, and define $w = u^* - u$. It follows immediately from Constraint 2 that $w$ must satisfy the functional constraint $w(kx) = K(k)w(x)$. Since both $u$ and $u^*$ are continuous, so is $w$; however, it may be a constant. Since $K(k) \not\equiv 1$, it is clear that the only constant solution is $w = 0$, and this is impossible since $u$ and $u^*$ were chosen to be different. Thus, by Theorem 1, $w(x) = \alpha_*|x|^\beta$. Substituting this into the functional constraint for $w$, it follows that $K(k) = k^\beta$. Then setting $x = 0$ in Constraint 2, we obtain $C(k) = u(0)(1 - k^\beta)$. We now observe that $u(x) = \alpha_*|x|^\beta + \delta$, where $\alpha_* = \alpha_+$ for $x \geq 0$ and $\alpha_* = \alpha_-$ for $x < 0$ and where $\delta = u(0)$, is a solution*

*to Constraint 2:*
$u(kx) = \alpha_* k^\beta |x|^\beta + \delta = \alpha_* k^\beta |x|^\beta + u(0)k^\beta + u(0) - u(0)k^\beta = k^\beta u(x) + u(0)(1-k^\beta) = K(k)u(x) + C(k)$. *Any other solutions is of the same form because* $u^*(x) = u(x) + w(x) = \alpha_*|x|^\beta + \delta + \alpha'_*|x|^\beta = (\alpha_* + \alpha'_*)|x|^\beta + \delta$. *It is easy to see that* $\delta$ *is independent of the unit of* $x$ *and* $\beta$ *is independent of the both units.*

**Theorem 3** *If the independent continuum is an interval scale, then it is impossible for the dependent continuum to be a ratio scale.*
*Proof. Let* $c = 0$ *in Constraint 3, then by Theorem 1 we know that the unique admissible function for* $u(x)$ *under this condition is that* $u(x) = \alpha_*|x|^\beta$. *Now set* $k = 1$ *and* $c \neq 0$ *in Constraint 3:* $\alpha_*|x + c|^\beta = K(1,c)\alpha_*|x|^\beta$, *so* $|x + c| = K(1,c)^{1/\beta}|x|$, *which implies* $x$ *has two constant values, contradictory to our assumption that both continua have more than two points. Accordingly, the function* $u(x)$ *admissible under any combination of values of* $k$ *and* $c$ *does not exist.*

**Theorem 4** *If the independent and dependent continua are both interval scales, then* $u(x) = \alpha_*|x| + \beta$, *where* $\alpha_* = \alpha_+$ *for* $x \geq 0$ *and* $\alpha_* = \alpha_-$ *for* $x < 0$ *and where* $\beta$ *is independent of the unit of the independent quantity.*
*Proof. If we let* $c = 0$, *then Constraint 4 reduces to Constraint 2 and so Theorem 2 applies. If* $u(x) = \alpha \log|x| + \beta_*$, *then choosing* $k = 1$ *and* $c \neq 0$ *in Constraint 4 yields* $\alpha \log|x+c| + \beta_* = K(1,c)\alpha\log|x| + K(1,c)\beta_* + C(1,c)$. *By taking the derivative with respect to* $x$ *except at* $x = 0$ *and* $x = -c$, *it is easy to see that* $x$ *must be a constant, which is impossible. Thus, we must conclude that* $u(x) = \alpha_*|x|^\gamma + \beta$. *Again, set* $k = 1$ *and* $c \neq 0$, $\alpha_*|x+c|^\gamma + \beta = K(1,c)\alpha_*|x|^\gamma + K(1,c)\beta + C(1,c)$. *If* $\gamma \neq 1$, *then differentiate with respect to* $x$ *except at* $x = 0$ *and* $x = -c$: $\alpha_*\gamma|x+c|^{\gamma-1} = K(1,c)\alpha_*\gamma|x|^{\gamma-1}$ *which implies* $x$ *is one of the two constants, so we must conclude* $\gamma = 1$. *It is easy to see that* $u(x) = \alpha_*|x| + \beta$, *where* $\alpha_* = \alpha_+$ *for* $x \geq 0$ *and* $\alpha_* = \alpha_-$ *for* $x < 0$, *satisfies Constraint 4.*

The results of these theorems are summarized in Table 3. The impossibility of the definition of a ratio scale from an interval scale is because the ratio scale involves the information of an absolute origin, but the interval scale does not. Every elementary laws in physics follows this table. Luce enumerated such examples as follows(Luce 1959). The quantities entering into Coulomb's law, Ohm's law and Newton's gravitation law are all ratio scales, and the formula of each law is a power function which follows the formula of Eq.1 in the table. Additional examples of Eq.1 can be seen in geometry. The volume of a sphere upon its radius and the area of a square on its side are such samples, since length, area and volume are all ratio scales. Other representative physical quantities such as energy and entropy are interval scales, and we see examples of Eqs.2.1 and 2.2 for laws associated with those. The total energy $U$ of a body having a constant mass $m$ and moving at velocity $v$ is $U = mv^2/2 + P$, where $P$ is the potential energy. If the temperature of a perfect gas is constant, then the entropy $E$ of the gas as a function of the pressure $p$ is of the form $E = -R\log p + E'$, where $R$ and $E$' are Boltzmann's constant and a reference value of entropy respectively. As an example of Eq.4, there is the relation $x = vt + x_0$ for a particle moving at its constant velocity $v$, where $x$ is the position at the present time $t$ and $x_0$ is the initial position $x, x_0$ and $t$ are all interval scales here. The examples of this table are not limited to the physical and geometrical domains. In psychophysics, Fechner's law states that the sound tone $s$ of human sensing (proportional to the order of white keys of a piano) is proportional to the logarithm of the sound frequency $f$, *i.e.*, $s = \alpha\log f + \beta$, where $s$ is an interval scale, and $f$ is a ratio scale.

Finally, the following important consequence should be indicated.

**Theorem 5** *A absolute scale quantity can have functional relations of any continuous formulae with other absolute scale quantities.*

*Proof. When an independent quantity* $x$ *and the dependent quantity* $u(x)$ *are absolute scale quantities, both of them have the group structure of* $x' = x$ *and* $u(x') = u(x)$. *Any continuous formulae of* $u(x)$ *satisfy the constraints.*

For example, the behavior solution of the simple pendulum of Fig. 1 is $\theta = \sin(t(g/l)^{1/2})$. The triangular function sin which does not belong to Table 3 can hold, because $\theta$ and $t(g/l)^{1/2}$ are dimensionless, *i.e.*, absolute scale.

## Theory and Method to Derive Possible Law Equations

As explained in the first section, a complete process in a system forms an ensemble, and is represented as a complete equation $\phi(x_1, x_2, ..., x_n) = 0$(Bhaskar & Nigam 1990). Within an ensemble, some regimes exist. Each regime represents a decomposable subprocess which corresponds to a dimensionless quantity $\Pi_i (i = 1, ..., k)$ where $k = n - r$. $\Pi_i$ and other quantities $x_{1_i}, x_{2_i}, ..., x_{m_i}$ having their scales form a complete equation $\rho_i(\Pi_i, x_{1_i}, x_{2_i}, ..., x_{m_i}) = 0$ where $m_i \leq n$. Furthermore, $\Pi_i$ for all regimes in an ensemble forms their complete equation $F(\Pi_1, \Pi_2, ..., \Pi_k) = 0$ on the basis of Buckingham's $\Pi$-theorem(Buckingham 1914). Therefore, the ensemble $\phi(x_1, x_1, ..., x_n) = 0$ can be decomposed into the following set of equations.

$$\{\rho_i(\Pi_i, x_{1_i}, x_{2_i}, ..., x_{m_i}) = 0 | i = 1, ..., k\},$$

$$F(\Pi_1, \Pi_2, ...., \Pi_k) = 0.$$

The basic formula of each $\rho_i$ can be derived by the principle in the previous section, if we know the types of scales of $x_{1_i}, x_{2_i}, ..., x_{m_i}$. However, the formula of $F$ can not be determined because of Theorem 5.

Table 3: The possible relations satisfying the scale characteristics

| | Scale Types | | | |
|---|---|---|---|---|
| *Eq.* No. | Independent variable | Dependent (Defined) variable | Possible Relations | Comments* |
| 1 | ratio | ratio | $u(x) = \alpha_*\|x\|^\beta$ | $\beta/x, \beta/u$ |
| 2.1 | ratio | interval | $u(x) = \alpha \log\|x\| + \beta_*$ | $\alpha/x$ |
| 2.2 | | | $u(x) = \alpha_*\|x\|^\beta + \delta$ | $\beta/x; \beta/u; \delta/x$ |
| 3 | interval | ratio | impossible | |
| 4 | interval | interval | $u(x) = \alpha_*\|x\| + \beta$ | $\beta/x$ |

1) The notations $\alpha_*, \beta_*$ are $\alpha_+, \beta_+$ for $x \geq 0$ and $\alpha_-, \beta_-$ for $x < 0$, respectively.
2) The notations $\alpha/x$ means "$\alpha$ is independent of the unit $x$".

## Scale-Based Reasoning within a Regime

This subsection describes the theory and the method of scale-based reasoning which uses a priori knowledge of quantity scales (not the knowledge of physics) to derive the formula of $\rho_i$ for a given regime is described. The product theorem is extended to include interval scale quantities. The principle of the extension is a certain symmetry that the relations given in Table 3 must hold for each pair of quantities in a regime. First, we settle the following lemma.

**Lemma 3** *Given a set of quantities forming a regime where some quantities are interval scales and the others ratio scales, the relations of* $y = \alpha \log|x_i| + \beta_*$ *and* $y = \alpha_*|x_j|^\beta + \delta (i \neq j)$ *from any two independent ratio scales* $x_i$ *and* $x_j$ *to any one dependent interval scale* $y$ *are not allowed to coexist in the same regime.*
*Proof.*
*Assume that* $y = \alpha \log|x_i| + \beta_*$ *and* $y = \alpha_*|x_j|^\beta + \delta$ *coexist in a regime. By casting* $\alpha$ *and* $\beta_*$ *of the former with* $(\alpha|x_j|^\beta + c)$ *and* $(\alpha'|x_j|^\beta + c')$ *respectively, we obtain* $y = (a|x_j|^\beta + c)\log|x_i| + (a'|x_j|^\beta + c')$. *This formula satisfies the relation from* $x_i$ *to* $y$ *when* $x_j$ *is constant. It also satisfies the relation from* $x_j$ *to* $y$ *by letting* $\alpha_* = a\log|x_i| + a'$ *and* $\delta = c\log|x_i| + c'$, *while holding* $x_i$ *constant. However, by solving this formula with* $x_i$, $x_i = \pm e^{\{(y-a'|x_i|^\beta - c')\}/\{(a|x_j|^\beta + c)\}}$ *is derived. This is different from the admissible relation between the two ratio scales,* $x_j = \alpha_*'|x_i|^{\beta'}$ *in Table 3 for any values of* $a, a', c$ *and* $c'$, *and so it is contradictory. By casting* $(a\log|x_i| + b)$ *and* $(a'\log|x_i| + b')$ *to* $\alpha_*$ *and* $\delta$ *of the latter respectively, the same discussion leads to the contradictions to Table 3 of the relations between the two ratio scales.*

Next, the extension of the product theorem is derived.

**Theorem 6** *Given a set of quantities* $Q = \{x_1, x_2, ..., x_m\}$ *forming a regime where some quantities are interval scales and the others ratio scales, the relation* $f = 0$ *among* $x_1, x_2, ..., x_m$, *has either one of the two forms.*

$$f(x_1, x_2, ..., x_m) = \left(\prod_{x_i \in R} |x_i|^{a_i}\right)\left(\sum_{x_j \in I_1} b_{*j}|x_j| + c_{*1}\right) + \sum_{x_k \in I_2} b_k|x_k| + c_2 \quad (i)$$

$$f(x_1, x_2, ..., x_m) = \sum_{x_i \in R} a_i \log|x_i| + \sum_{x_j \in I} b_j|x_j| + c_* \quad (ii)$$

*Here,* $R$ *is a set of quantities of ratio scale in* $Q$, *and* $I$ *is a set of all quantities of interval scale in* $Q$. *Also,* $I_1$ *and* $I_2$ *are any partition of* $I$ *where* $I_1 + I_2 = I$.
*Proof. The following two cases are considered based on Lemma 3.*

1) *When the relation from every* $x_i \in R$ *to every* $x_h \in I$ *is given as* $x_h = \alpha_{*h_i}|x_i|^{\beta_{h_i}} + \delta_{h_i}$, *by casting* $\alpha_{*h_i}$ *and* $\delta_{h_i}$ *of this formula with* $(a_{*h_j}|x_j|^{\beta_{h_j}} + c_{*h_j})$ *and* $(a_{h_j}|x_j|^{\beta_{h_j}} + c_{h_j})$ *respectively where* $x_j \in R$ *and* $i \neq j$, *we obtain*
$x_h = (a_{*h_j}|x_j|^{\beta_{h_j}} + c_{*h_j})|x_i|^{\beta_{h_i}} + (a_{h_j}|x_j|^{\beta_{h_j}} + c_{h_j})$
*Though this formula satisfies the both relations from* $x_i$ *to* $x_h$ *and from* $x_j$ *to* $x_h$, *the relation between the two ratio scales,* $x_j = \alpha_*|x_i|^\beta$ *holds only when* $c_{*h_j} = a_{h_j} = 0$. *Hence,* $x_h = a_{*h_j}|x_i|^{\beta_{h_i}}|x_j|^{\beta_{h_j}} + c_{h_j}$. *This must hold for every* $x_i \in R$, *therefore*
$$x_h = A_{*h}\left(\prod_{x_i \in R} |x_i|^{a_{h_i}}\right) + C_h.$$
*The relation from every* $x_j \in I$ *to* $x_h$ *is given as* $x_h = \alpha_{*h_j}|x_j| + \beta_{h_j}$. *By casting* $A_{*h}$ *and* $C_h$ *with* $(a_{*h_j}|x_j| + b_{*h_j})$ *and* $(a_{h_j}|x_j| + b_{h_j})$ *respectively,*
$$x_h = (a_{*h_j}|x_j| + b_{*h_j})\left(\prod_{x_i \in R} |x_i|^{a_{h_i}}\right) + (a_{h_j}|x_j| + b_{h_j}).$$
*is derived. Though this satisfies the admissible relations from every* $x_i \in R$ *to* $x_h$ *and from* $x_j$ *to* $x_h$, *the admissible relation from every* $x_i \in R$ *to* $x_j$ *holds only when* $a_{*h_j} = 0$ *or* $a_{h_j} = 0$. *Thus, one of the followings holds.*
$$x_h = b_{*h_j}\left(\prod_{x_i \in R} |x_i|^{a_{h_i}}\right) + (a_{h_j}|x_j| + b_{h_j}). \quad (a),$$

$$x_h = (a_{*h_j}|x_j| + b_{*h_j})(\prod_{x_i \in R} |x_i|^{a_{h_i}}) + b_{h_j} \quad (b).$$

*For the former formula (a), by casting* $b_{*h_j}$, $a_{h_j}$ *and* $b_{h_j}$ *with* $(a_{*h_k}|x_k| + b_{*h_k})$, $(a_{h_k}|x_k| + b_{h_k})$ *and* $(a'_{h_k}|x_k| + b'_{h_k})$ *respectively where* $x_k \in I$ *and* $j \neq k$, *the following relation is obtained.*

$$x_h = (a_{*h_k}|x_k| + b_{*h_k})(\prod_{x_i \in R} |x_i|^{a_{h_i}})$$
$$+\{(a_{h_k}|x_k| + b_{h_k})|x_j| + (a'_{h_k}|x_k| + b'_{h_k})\}.$$

*Though this formula satisfies the admissible relations from every* $x_i \in R$, $x_j$ *and* $x_k$ *to* $x_h$, *the relation between* $x_j$ *and* $x_k$, $x_k = \alpha_*|x_j| + \beta$, *is maintained only when* $a_{h_k} = 0$. *Thus,*

$$x_h = (a_{*h_k}|x_k| + b_{*h_k})(\prod_{x_i \in R} |x_i|^{a_{h_i}})$$
$$+(b_{h_k}|x_j| + a'_{h_k}|x_k| + b'_{h_k}).$$

*The relation from every* $x_i \in R$ *to* $x_k$ *holds only when* $a_{*h_k} = 0$ *or* $a'_{h_k} = 0$ *similarly to the above discussion, and hence*

$x_h = b_{*h_k}(\prod_{x_i \in R} |x_i|^{a_{h_i}}) + (b_{h_k}|x_j| + a'_{h_k}|x_k| + b'_{h_k})$,
*or*
$x_h = (a_{*h_k}|x_k| + b_{*h_k})(\prod_{x_i \in R} |x_i|^{a_{h_i}}) + (b_{h_k}|x_j| + b'_{h_k})$.

*The same discussion is made for the formula (b), and the followings are derived.*

$$x_h = (b_{*h_k}|x_j| + b'_{*h_k})(\prod_{x_i \in R} |x_i|^{a_{h_i}})$$
$$+(a_{h_k}|x_k| + b_{h_k}), \text{ or}$$
$$x_h = (b_{*h_k}|x_j| + a'_{*h_k}|x_k| + b'_{*h_k})(\prod_{x_i \in R} |x_i|^{a_{h_i}}) + b_{h_k}.$$

*This must hold for every* $x_j \in I$ $(j \neq h)$, *therefore*

$$x_h = (\prod_{x_i \in R} |x_i|^{a_{h_i}})(\sum_{x_j \in (I_1 - \{x_h\})} b_{*h_j}|x_j| + c_{*h_1})$$
$$+ \sum_{x_k \in (I_2 - \{x_h\})} b_{h_k}|x_k| + c_{h_2}$$

*By moving* $x_h$ *to r.h.s., the formula (i) is obtained.*

*2) When the relation from every* $x_i \in R$ *to every* $x_h \in I$ *is given as* $x_h = \alpha_{h_i} \log|x_i| + \beta_{*h_i}$, *by casting* $\alpha_{h_i}$ *and* $\beta_{*h_i}$ *of this formula with* $(a_{h_j} \log|x_j| + c_{h_i})$ *and* $(a_{*h_j} \log|x_j| + c_{*h_i})$ *respectively where* $x_i \in R$ *and* $i \neq j$, *we obtain*

$x_h = (a_{h_j} \log|x_j| + c_{h_j}) \log|x_i| + a_{*h_j} \log|x_j| + c_{*h_j}$.

*Though this formula satisfies the both relations from* $x_i \in R$ *and* $x_j$ *to* $x_h$, *the relation between two ratio scales,* $x_j = \alpha_*|x_i|^\beta$ *holds only when* $a_{h_j} = 0$. *Hence,* $x_h = c_{h_j} \log|x_i| + a_{*h_j} \log|x_j| + c_{*h_j}$. *This must hold for every* $x_i \in R$, *therefore*

$$x_h = \sum_{x_i \in R} a_{*h_i} \log|x_i| + C_{*h}.$$

*The relation from every* $x_j \in I$ *to* $x_h$ *is given as* $x_h = \alpha_{*h_j}|x_j| + \beta_{h_j}$. *By casting* $a_{*h_i}$ *and* $C_{*h}$ *with* $(a_{*h_j}|x_j| + b_{*h_j})$ *and* $(a'_{*h_j}|x_j| + b'_{*h_j})$ *respectively,*

$$x_h = \sum_{x_i \in R} (a_{*h_j}|x_j| + b_{*h_j}) \log|x_i| + (a'_{*h_j}|x_j| + b'_{*h_j}).$$

*is derived. Though this satisfies the admissible relation from every* $x_i \in R$ *to* $x_h$ *and* $x_j$ *to* $x_h$, *the relation between every* $x_i \in R$ *and* $x_j$, $x_j = \alpha \log|x_i| + \beta_*$ *is maintained only when every* $a_{*h_j} = 0$. *Hence,*

$$x_h = \sum_{x_i \in R} b_{*h_j} \log|x_i| + a'_{*h_j}|x_j| + b'_{*h_j}.$$

*Again, by casting* $b_{*h_j}$, $a'_{*h_j}$, *and* $b'_{*h_j}$ *with* $(a_{*h_k}|x_k| + b_{*h_k})$, $(a'_{*h_k}|x_k| + b'_{*h_k})$ *and* $(a''_{*h_k}|x_k| + b''_{*h_k})$ *respectively where* $x_k \in R$ *and* $j \neq k$, *the following relation is obtained.*

$$x_h = \sum_{x_i \in R} (a_{*h_k}|x_k| + b_{*h_k}) \log|x_i|$$
$$+(a'_{*h_k}|x_k| + b'_{*h_k})|x_j| + (a''_{*h_k}|x_k| + b''_{*h_k}).$$

*Though this formula satisfies the admissible relations from every* $x_i \in R$, $x_j$ *and* $x_k$ *to* $x_h$, *the relations between every* $x_i \in R$ *and* $x_k$, $x_k = \alpha \log|x_i| + \beta$ *and between* $x_j$ *and* $x_k$, $x_k = \alpha_*|x_j| + \beta$, *are maintained only when* $a_{*h_k} = a'_{*h_k} = 0$. *Thus,*

$$x_h = \sum_{x_i \in R} b_{*h_k} \log|x_i| + b'_{*h_k}|x_j| + a''_{*h_k}|x_k| + b''_{*h_k}.$$

*This must hold for every* $x_j \in I$, *therefore*

$$x_h = \sum_{x_i \in R} a_{*h_i} \log|x_i| + \sum_{x_j \in (I - \{x_h\})} b_{*h_j}|x_j| + c_{*h}.$$

*By moving* $x_h$ *to r.h.s., the formula (ii) is obtained.*

**Theorem 7** *If* $R = \phi$, *i.e.,* $I = Q$, *in Theorem 6 then the relation* $f = 0$ *among* $x_1, ..., x_m$ *has the following form.*

$$f(x_1, x_2, ..., x_m) = \sum_{x_i \in R} b_{*i} \log|x_i| + c_*$$

*Proof. The first term associated with* $R$ *in each formula of (i) and (ii) of Theorem 6 is neglected because of* $R = \phi$, *and both formulae become as the above formula.*

Now, we consider to derive the candidates of formula $f = 0$ of a given regime $\rho$, where the set of quantities except dimensionless $\Pi$ in $\rho$ is $Q = \{x_1, ..., x_m\}$. The following algorithm based on the product theorem and the above theorems 6 and 7 derives the candidates.

**Algorithm 1**

(Step 1) $Q = \{x_1, ..., x_m\}$, $S = \phi$. Let $R$ be a set of all quantities of ratio scale in $Q$. Let $I$ be a set of all quantities of interval scale in $Q$.
(Step 2) If $I = \phi$ {
Based on the product theorem,
push the following to $S$.
$$f(x_1, x_2, ..., x_m) = \prod_{x_i \in R} |x_i|^{a_i} + c_* = 0.\}$$
else if $R = \phi$ {
Push the relation of theorem 7 to $S$.}
else {
Enumerate candidate relations of (i)
in theorem 6 for all binary partitions

$\{I_1, I_2\}$ of $I$, and push those candidates to $S$.
Push the relation of (ii) in theorem 6 to $S$.}

The candidates is rested in the list $S$. The result of $S$ is sound, since the soundness is ensured by the product theorem, the theorems 6 and 7. The complexity of this algorithm is low except the enumeration of all binary partitions $\{I_1, I_2\}$ of $I$, where the complexity is $O(2^{|I|})$. The exchange of $I_1$ and $I_2$ essentially gives an equivalent relation with the original relation (i) of the theorem 6, as it is easily understood by the form of the relation. Therefore, the complexity reduces to $O(2^{|I|-1})$. In any case, it is not very problematic, because the size of a regime is generally quite limited.

We implemented this algorithm by using a commercial formula-processing package Maple V(Char 1991). This algorithm have been tested by various physical laws. The following is the example of the ideal gas equation which forms a unique regime. A regime involving four quantities of pressure $p$, volume $v$, mass $m$ and temperature $t$ is given, thus $R = \{p, v, m, t\}$. The quantities $p$, $v$ and $m$ are ratio scales, while only $t$ is an interval scale unless it is absolute temperature. We assumed the positive sign of $p$, $v$ and $m$ in advance, hence the solutions for their negative values were omitted. The algorithm figured out the two candidate relations.

$$0 = c_1 p^{a_1} v^{a_2} m^{a_3} + b_1 t + c_2.$$

Consequently, the following candidate was obtained.

$$p^{a_1} v^{a_2} = m^{-a_3}\left(-\frac{b_1 t + c_2}{c_1}\right).$$

Next, another candidate was found.

$$0 = a_1 \log p + a_2 \log v + a_3 \log m + b_1 * t + c.$$

$$p^{a_1} v^{a_2} = m^{-a_3} \exp(-b_1 t - c).$$

The former solution reflects the right formula of the ideal gas equation, when the temperature is not absolute one. Once the candidate formulae of a regime are determined, the correct formula and the values of its coefficients must be specified in data-driven manner.

## Data-Driven Reasoning on Ensemble

Scale-based reasoning can derive the basic formula representing a regime without using any knowledge about its structure. However, more insights on the objective system are definitely needed to determine the formula of the entire ensemble due to Theorem 5. To do so, the introduction of a priori knowledge such as symmetry of the system is a choice. But, this direction will limit the applicability of the method to more ambiguous domain like psychophysics and sociology. Accordingly, we chose a data-driven approach of the formula identification for an ensemble. In this frame work, the knowledge of dimensions of quantities and the number of basic dimensions, r, are not given but only the knowledge of scales. Accordingly, the regimes and their number, k (=n-r), must be interpreted by using the given measurement data based on the following alternative definition of regimes.

**Definition 1**
A regime is a subsystem, where the relation among quantities follows either of the product theorem, theorem 6 and 7, of a given complete equation.

This definition of a regime yields an significant advantage to relax the limitation of the assumption 3 mentioned in the first section. Consider the example of convection heat transfer coefficient given by Kalagnanam et al.(Kalagnanam, Henrion, & Subrahmanian 1994) The heat transfer from a fluid to a pipe wall takes place through convection when a fluid forced through a pipe. A complete equation for this phenomenon under the turbulent flow is known as follows.

$$\Pi_1 = 0.023 {\Pi_2}^{0.8} {\Pi_3}^{0.4}$$

where $\Pi_1 = \frac{hD}{k}$, $\Pi_2 = \frac{Dv\rho}{\mu}$ and $\Pi_3 = \frac{c\mu}{k}$.

$h$ is the convection heat transfer coefficient dependent to the other quantities. $D$ and $V$ are diameter of pipe and velocity of stream. $m$, $r$, $c$ and $k$ are the material quantities of the fluid, *i.e.*, viscosity, density, specific heat and thermal conductivity respectively. The number of quantities and that of the basic dimensions are $n = 7$ and $r = 4$ respectively. Hence, three ($k = 3$) regimes exist in this ensemble, and $\Pi_1$, $\Pi_2$ and $\Pi_3$ are called as Nusselt's, Reynold's and Prandtl's numbers in the thermal hydraulics domain. This example violates the assumption 3, because $D$ and $m$ appear in multiple regimes. However, this equation is regarded as a unique regime in terms of the definition 1, because it can be rewritten to follow the product theorem as:

$$\frac{k^{0.6} V^{0.8} \rho^{0.8} c^{0.4}}{h D^{0.2} \mu^{0.4}} - \frac{1}{0.023} = 0.$$

This case assumes the number of basic dimensions, $r$, to be 6 without the knowledge of dimensions. Though this number is different from the actual one, it is sufficient to identify the formula of the relation among quantities based on the knowledge of scales and given measurement data. Furthermore, the combined use of the definition 1 and the assumption 4 described in the first section completely removes the limitation of the assumption 3. For example, consider to evaluate the average of the convection heat transfer coefficients, $h_a$, for the adjacently connected two pipes having different diameter $D_1$ and $D_2$. While the velocities of steam in the two pipes, $V_1$ and $V_2$ become also different, the values of the extra material quantities are identical. Thus, the relation among quantities is given as follows.

$$\frac{k^{0.6} \rho^{0.8} c^{0.4} {V_1}^{0.8}}{h_a \mu^{0.4} {D_1}^{0.2}} + \frac{k^{0.6} \rho^{0.8} c^{0.4} {V_2}^{0.8}}{h_a \mu^{0.4} {D_2}^{0.2}} - \frac{2}{0.023} = 0.$$

This seems to involve two regimes containing $k$, $\rho$, $c$, $h_a$ and $\mu$ in common. However, when the relation between $D_1$ and $V_1$ is measured while fixing the values of $k$, $\rho$, $c$, $h_a$, $\mu$, $D_2$ and $V_2$ under the assumption 4, the part of $D_1, V_1$ is identified as a regime by the definition 1. The regime of $D_2, V_2$ is also identified in the same manner. Moreover, the relation among $k$, $\rho$, $c$, $h_a$ and $\mu$ is regarded as a regime by the measurement under fixing the values of $D_1$, $V_1$, $D_2$ and $V_2$. Thus, we obtain the following interpretation.

$$\Pi_3\Pi_1 + \Pi_3\Pi_2 - \frac{2}{0.023} = 0$$

where $\Pi_1 = \frac{V_1{}^{0.8}}{D_1{}^{0.2}}$, $\Pi_2 = \frac{V_2{}^{0.8}}{D_2{}^{0.2}}$ and $\Pi_3 = \frac{k^{0.6}\rho^{0.8}c^{0.4}}{h_a\mu^{0.4}}$.

Consequently, the following proposition can be stated.

**Proposition 1**

Given an ensemble $\phi(x_1, x_2, ...x_n) = 0$, its following decomposition into regimes always exists. $F(\Pi_1, \Pi_2, ..., \Pi_k) = 0$ and $\{\rho_i(\Pi_i, x_{1_i}, x_{2_i}, ...x_{m_i}) = 0 | i = 1, ..., k\}$, where $\{Q_i | Q_i = \{x_{1_i}, x_{2_i}, ...x_{m_i}\}, i = 1, ..., k\}$ is a partition of entire quantity set $E = \{x_1, x_2, ...x_n\}$.

Once each regime is identified, an absolute scale quantity must be defined in each regime, and their mutual relation $F(\Pi_1, \Pi_2, ..., \Pi_k) = 0$ is searched by a certain procedure. Based on these considerations, the following algorithm to identify the formula of a given complete equation has been constructed.

**Algorithm 2**

(Step 1) $E = \{x_1, x_2, ...x_n\}$, $LE = \phi$ and $k = 1$.
(Step 2) Repeat until $k$ becomes equal to $n$. {
  Repeat for every partition $\Gamma_i$ of $E$
  where $|\Gamma_i| = k$. {
    Repeat for every $Q_{i_j} \in \Gamma_i (j = 1, ..., k)$. {
      Apply the algorithm 1 to $Q_{i_j}$.
      Test each solution in $S$ through the
      least square fitting to the measured
      data under some sets of values of
      quantities in $(E - Q_{i_j})$.
      If some solutions are accepted,
      substitute them to a list $LQ_{i_j}$.}
    If every $LQ_{i_j} \neq \phi$, let a list
    $L\Gamma_i = \{LQ_{i_j} | Q_{i_j} \in \Gamma_i, j = 1, ..., k\}$.
    Push $L\Gamma_i$ to $LE$.}
  If $LE \neq \phi$, go to (Step 3), else $k = k + 1$.}
(Step 3) Repeat for every $L\Gamma_i$ in $LE$. {
  Repeat for every $LQ_{i_j}$ in $L\Gamma_i (j = 1, ..., k)$. {
    Repeat for every solution of a regime,
    $s_{i_j}$, in $LQ_{i_j}$. {
      Determine an absolute scale quantitiy
      $\Pi(s_{i_j})$ based on the coefficients of $s_{i_j}$
      evaluated through the least square
      fitting to the measured data.}}}
(Step 4) Repeat for every $L\Gamma_i$ in $LE$. {
  Take Cartesian products $LP_i = LQ_{i_1} \times$
  $LQ_{i_2} \times \cdots \times LQ_{i_k}$ in $L\Gamma_i$.
  Repeat for every $\{s_{i_1}, s_{i_2}, ..., s_{i_k}\} \in LP_i$. {
    Determine the formula $F(\Pi(s_{i_1})$,
    $\Pi(s_{i_2}), ..., \Pi(s_{i_k})) = 0$.}}

More concrete contents of this algorithm are demonstrated through an example of Black's specific heat law. This relates the initial temperatures of two substances $T_1$ and $T_2$ with their temperature $T_f$ after they have been combined. This law has the formula of $T_f = (w_1M_1/(w_1M_1 + w_2M_2))T_1 + (w_2M_2/(w_1M_1 + w_2M_2))T_2$ where $M_1$ and $M_2$ are the masses of the two substances, and $w_1$ and $w_2$ are their specific heat coefficients. The conditions of $M_1 = 0.5M_2$ and $w_1 = w_2$ are applied in our example. In (Step 1), the set of measured quantities $E$ is set as $\{T_f, T_1, T_2, M_1, M_2\}$. (Step 2) is the process to enumerate all partitions of $E$ where each subset of $E$ is interpreted as a regime in terms of the definition 1. The data for each quantity in $E$ is measured with almost 2% relative noise in our demonstration. The goodness of the least square fitting of each candidate solution derived by the algorithm 1 is checked through the F-test which is a statistical hypothesis test to judge if the measured data follows the solution. If a parameter is close enough to an integer value, then the parameter is forced to be the integer, because the parameter having an integer value is quite common in various domains. Every partition $Gamma_i$, where all of its subsets are judged to be regimes, is searched in an ascending order of the cardinal number $k$ of the partition. Once such partitions are found at the level of the certain cardinal number, (Step 2) is finished at that level to obtain the solutions of the ensembles involving the least number of regimes. This criterion decreases the ambiguity of the formula of the ensemble by reducing the number of the absolute scale quantities in it. In the current example, only the partition of $\{\{Tf, T1, T2\}, \{M1, M2\}\}$ is accepted at the least cardinal number, $k = 2$. All quantities in the former regime are of interval scale, and those in the latter are of ratio scale, and thus the formula of these regimes are enumerated as:

$$T_f = b_1T_1 + b_2T_2 + c_1, \qquad M_1^{a_1}M_2^{a_2} + c_1 = 0.$$

Their more specific formulae are identified through the data fitting as follows.

$$T_f = b_1T_1 + b_2T_2, \qquad M_1M_2^{-1} + c_1 = 0.$$

(Step 3) defines an absolute scale quantity $\Pi$ for each regime. The definition can be made in various ways. In our approach, a parameter in a regime, which is variable by the influence from the other regimes, is chosen to be a $\Pi$, and the other variable parameters in the regime are evaluated in terms of the $\Pi$. In the current example, both of $b_1$ and $b_2$ in the former regime are observed to be variable depending on the values of $M_1$ and $M_2$. Accordingly, many data of $\{b_1, b_2\}$ are evaluated through fitting to the measured data of $T_f$, $T_1$ and $T_2$ under various values of $M_1$ and $M_2$. Then, $b_1$ is chosen to be the $\Pi_1$ of the regime, and various formulae of $f(\Pi_1)$ are tested for $b_2$ through the F-test of the least square fitting to the data of $\{b_1, b_2\}$. The class of $f(\Pi_1)$ is limited to polynomial and meromorphic formulae in our current study. The following is the most simple formula accepted in the test.

$$b_1 = \Pi_1, \qquad b_2 = 1 - \Pi_1.$$

For the latter regime, $c_1$ is uniquely chosen to be the $\Pi_2$.

(Step 4) searches the relation of each ensemble found in the preceding steps. In the example under consideration, first, many data of $\{\Pi_1, \Pi_2\}$ are obtained by applying various combinations of values of $\{T_f, T_1, T_2, M_1, M_2\}$. Then various formulae of $F(\Pi_1, \Pi_2) = 0$ are tested for the ensemble while limiting the class of $F$ to polynomial and meromorphic formulae. The most simple formula accepted is as follows.

$$\Pi_1\Pi_2 - \Pi_1 - \Pi_2 = 0.$$

By combining above formulae, the familiar equation of the Black's specific heat law under the condition of $w_1 = w_2$ is obtained.

$$T_f = \frac{M_1}{M_1 + M_2}T_1 + \frac{M_2}{M_1 + M_2}T_2.$$

## Discussions and Related Work

The basic principle of the scale-based reasoning we proposed is the isomophic symmetry of the equation formula under the admissible scale conversions of the ratio and interval types. This constraint strongly restricts the possible shapes of the equations. In this sense, the fundamental principle of our approach is analogous to the symmetry-based reasoning proposed by Ishida [Ishida 1995]. However, the source of the symmetry in our case is independent from the features of the objective system, while his approach uses the symmetries intrinsic to the system.

One of the basic principle of the dimension analysis is the product theorem. Because this theorem holds only for ratio scales, the applicability of the dimension analysis is limited to certain domains. It requires the knowledge of the unit structure which carries the information of the physical relation among quantities as we explained for the case of $f = m\alpha$, and this feature also limit its applicability. Obviously, the scale-based reasoning can be applied to diverse domains such as Fechner's law in psychophysics, because it utilizes only the features of quantity scales for the part of the identification of regimes.

The data-driven approaches taken in BACON(Langley & Zytkow 1989) and some other works(Li & Biswas 1995) have the wide applicability. However, their solutions are not ensured in terms of soundness. Although our approach is also data-driven in some extent, the solutions for the regimes are guaranteed to be sound, because the product theorem, the theorems 6 and 7 cover all possibility of relations in the regimes. These theorems are expected to contribute to many fields of qualitative reasoning and knowledge discovery. While the complexity of the algorithm 1 becomes quite high for a more large scale system, it would be still less than that of BACON, because this can reduce the search space of the relations of the ensemble by the identification of regimes in advance.

## Conclusion

The major characteristics of our approach are summarized as follows.

1) The sound solutions of basic formulae of law equations within a regime are provided by using only the knowledge of quantity scales.
2) An ensemble and its regimes in the objective system are identified from the empirical data.
3) The applicability is not limited to well-defined domains, since the method does not require a priori knowledge except the knowledge of quantity scales.

The scale-based reasoning may provide a basis to develop qualitative models of ambiguous domains such as biology, psychology, economics and social science. This will also contribute the research area of knowledge discovery.

## References


Bhaskar, R., and Nigam, A. 1990. Qualitative Physics Using Dimentional Analysis. *Artificial Intelligece* 45:73–111.

Bridgman, P. 1922. *Dimensional Analysis.* Yale University Press, New Haven, CT.

Buckingham, E. 1914. On physically similar systems; Illustrations of the use of dimensional equations. *Phys. rev.* IV(4):345–376.

Char, B. 1991. *MapleV Language Reference Manual.* Springer-Verlag.

Helmholtz, H. 1887. *Numbering and Measuring from Epistemological Viewpoint.* Philosophische Aufsatze.

Ishida, Y. 1995. Symmetry-Based Reasoning about Equations of Physical Laws. In *Working Papers of Nith International Workshop on Qualitative Reasoning, Amsterdam, the Netherlands*, 84–93.

Kalagnanam, J.; Henrion, M.; and Subrahmanian, E. 1994. The Scope of Dimensional Analysis in Qualitative Reasoning. *Computational Intelligence* 10(2):117–133.

Langley, P., and Zytkow, J. 1989. Data-Driven Approaches to Empirical Discovery. *Artificial Intelligece* 40:283–312.

Li, C., and Biswas, G. 1995. Knowledge-based Scientific Discovery in Geological Databases. In *Proc. of the First International Conference on Knowledge Discovery & Data Mining, Montreal Quebec, Canada*, 204–209.

Luce, R. 1959. On the Possible Psychological Laws. *The Psychological Review* 66(2):81–95.

Stevens, S. 1946. On the Theory of Scales of Measurement. *Science* 103(2684):677–680.