From Expert Words Directly to Numerical Simulations: Group-Theoretic Approach to Computing with Words in Information/Intelligent Systems

Vladik Kreinovich, Brian Penn, and Scott Starks
NASA Pan-American Center for
Earth and Environmental Studies
University of Texas at El Paso
El Paso, TX 79968, USA
emails vladik@cs.utep.edu, bpenn@geo.utep.edu,
sstarks@utep.edu

Abstract

In many real-life situations, e.g., when making an environmental decision, it is important to be able to predict long-term consequences of different decisions. Very often, these predictions must be done in the situation where the only available information consists of expert rules, which are formulated by words from natural language. One possible way to transform these expert words into numerical simulation (leading to prediction) is to use the fuzzy control methodology. However, there is a problem with using this methodology: it invokes replacing each word by a membership function, and this replacement drastically increases the required computer space (and thus, increases the computation time), i.e., it "de-granulates" the original compact description. It is, therefore, desirable to get from the original words directly to numerical simulations, thus avoiding this de-granulation.

In seeking this direct transformation, we will use the experience of modern physics, where symmetry groups are a tool that enables to compress complicated differential equations into compact form. In our previous papers, we have shown that the symmetry group approach can be used to find optimal membership functions, optimal t-norms and t-conorms, and optimal defuzzification procedures. In this paper, we show that the same approach can also be used to combine these steps and produce an (optimal) direct transformation from words to numerical results.

1 From Expert Words to Numerical Simulations: Necessity

For many complex systems, long-term predictions are necessary. In the 20 century, there have been many situations in which an environment-related decision, that seemed, at first, to be very reasonable and successful, turned out, in the long run, to have been a mistake. Such decisions include the use of pesticides (e.g., DDT), the design of some river dams, etc.

To avoid such mistakes, we must be able to predict long-term consequences of each decision.

Numerical simulations are needed. The ideal situation is when we have an *analytical* formula that would enable us to *exactly* predict the consequences of each decision. However, in reality, such formulas are extremely rare. In most cases, we have to rely on *numerical simulations* instead.

Often, expert words are the only information we have. In some cases, we know the differential or difference equations that describe the system. However, in many other cases, especially for environmental systems, we do not know the exact equations. Instead, we have the informal *expert knowledge*.

This knowledge is usually formulated in terms of rules that use only words from natural language, such as: "if x increases, then y slightly decreases".

We must transform (fuzzy) expert words into (crisp) numerical simulations. Thus, to make meaningful decisions, we must somehow transform the (fuzzy) expert words that describe the system's dynamics, into crisp equations that would enable us to run numerical (computer) simulations of the consequences of different possible decisions.

2 From Expert Words to Numerical Simulations: How It Is Done Now

2.1 For the desired translation, we can use the experience of fuzzy control

There is an area where the methodology of transforming expert rules (like the one described above) into numerical formulas has been already successfully developed: the area of *intelligent control* based on fuzzy expert rules.

The corresponding *fuzzy control* methodology was first developed by Mamdani in [21, 22] (for the latest overview, see, e.g., see, Klir and Yuan [11], Nguyen and Walker [32], and Nguyen and Sugeno [31]). So, we can use this methodology to transform expert words into numerical simulations.

2.2 Fuzzy control methodology: in brief

Why explain. This paper has three main objectives:

- to explain the fuzzy control methodology and how it can be used for simulations;
- to explain the problems with applying this methodology to simulations, and
- to propose a better methodology.

Thus, fuzzy control methodology is crucial for us, and so, we will briefly describe this methodology for those readers who are not 100% familiar with it (readers familiar with fuzzy control can skip this explanation). In this explanation, we will only describe the *simplest* (basic) version of fuzzy control.

Rules. In the fuzzy control methodology, we start with expert rules of the type

$$\text{If } x_1 \text{ is } A_{r1}, \ldots, \text{ and } x_n \text{ is } A_{rn}, \text{ then } u \text{ is } B_r.$$

Here:

- $x_1, \ldots, x_n$ are inputs, i.e., parameters whose values we measure in order to decide what control to apply (e.g., the position and velocity of a spaceship);
- u is the desired control (e.g., the force applied to the spaceship);
- $r = 1, \ldots, R$ is the rule number, and
- A_{ri} and B_r are words from natural language that are used in r-th rule, like "small", "medium", "large", "approximately 1", etc.

To transform these rules into a precise control strategy $u = u(x_1, \ldots, x_n)$, we do the following:

First stage. First, we describe the words A_{ri} and B_r in numerical terms. In fuzzy control methodology, we usually describe each such word by a *membership function* $\mu_{ri}(x_i)$ (or, correspondingly, $\mu_r(u)$), i.e., a function that describes, for each x_i, to what extent the experts believe this very value x_i to satisfy the corresponding property A_{ri} (e.g., to what extent the experts believe that x_i is small).

These degrees of belief run from complete disbelief (x_i does not satisfy the property A_{ri}) to complete belief, i.e., from "false" to "true". In the computer, "false" is usually represented by 0, and "true" by 1. Therefore, in most implementations, the membership functions take values from 0 to 1.

Second stage. Next, for each input $(x_1, \ldots, x_n)$, and for each possible value, we describe to what extent i-th rule holds, i.e., to what extent it is true that

x_1 satisfies the property A_{r1}, *and* x_2 satisfies the property A_{r2}, ..., *and* x_n satisfies the property A_{rn}, *and* u satisfies the property B_r.

We have $n+1$ statements $A_{r1}(x_1), \ldots, A_{rn}(x_n), B_r(u)$, and for each of these statements, we know its "degree of belief" ("truth value"). We are interested in the degree of belief of their "and"-combination (disjunction) $A_{r1}(x_1)\& \ldots \& A_{rn}(x_n)\& B_r(u)$.

If all the combined statements were known to be exactly true or exactly false, then we would be able to use the known "and" operation for Boolean truth values. Thus, what we need is to generalize the traditional Boolean "and" operation, that is well defined for truth values from the set $\{0, 1\}$, to the entire interval $[0, 1]$.

Many such generalizations have been proposed; they are usually called "and"-operations, or *t-norms*. Two most widely used examples of t-norms are $a\&b = \min(a, b)$ and $a\&b = a \cdot b$.

In terms of a t-norm $\&$, the degree of belief that r-th rule is applicable is equal to $b_r = \mu_{r1}(x_1)\& \ldots \& \mu_{rn}(x_n)\& \mu_r(u)$.

Third stage. To compute, for given $x_1, \ldots, x_n$, and u, the degree of belief that this u is a reasonable control for the given $x_1, \ldots, x_n$, we must estimate the degree of belief that one of the rules is applicable, i.e., that *either* the first rule is applicable, *or* the second rule is applicable, etc.

We know the degree of belief b_r that each rule is applicable, so, to combine them, we need an extension $\vee$ of the Boolean "or"-operation to the interval $[0, 1]$. This extended "or"-operation is usually called a *t-conorm*.

The most widely used t-conorms are $a \vee b = \max(a, b)$ and $a \vee b = a + b - a \cdot b$. So, for each u, we can estimate the desired degree of belief as $\mu(u) = b_1 \vee \ldots \vee b_R$.

Fourth stage. After the previous step, for every possible value u, we get the degree of belief $\mu(u)$ that u is a reasonable control. We need to use the membership function $\mu(u)$ to choose a single value $\bar{u}$ that corresponds to the given $x_1, \ldots, x_n$. The transformation from the (fuzzy) membership function $\mu(u)$ to a single (crisp) value $\bar{u}$ is called a *defuzzification*.

In fuzzy control, one of the most widely used defuzzification procedures is the following *centroid defuzzification*:

$$\bar{u} = \frac{\int u \cdot \mu(u)\, du}{\int \mu(u)\, du}.$$

Conclusion. As a result of this methodology, we get, for each set of values $x_1, \ldots, x_n$, a certain control; in mathematical terms, we describe control as a *function* of the inputs: $u = u(x_1, \ldots, x_n)$. This function is called a *control strategy*.

Successes of fuzzy control methodology. The resulting *fuzzy control* is used in various areas ranging from appliances (camcorders, washing machines,

etc.) to automatically controlled subway trains in Japan to cement kilns to regulating temperature within the Space Shuttle.

2.3 How we can apply fuzzy control methodology to transform expert words into numerical simulation

Let $x_1, \ldots, x_n$ be parameters that describe the current state of a system that we are trying to simulate.

Expert formulate the rules that describe, for each of these variables (i.e., for each i from 1 to n), how the rate of change $u = \dot{x}_i$ of this variable depends on the values of this and other parameters. For example, a rule can be: "if x_1 is small, and x_2 is large, then u should be small".

The above-described methodology will then allow us to transform these rules into a numerical formula $u = f_i(x_1, \ldots, x_n)$. Since $u = \dot{x}_i$, we get a system of differential equations $\dot{x}_i = f_i(x_1, \ldots, x_n)$, whose simulation describes the long-term consequences of the given decision.

3 First Problem with the Existing Approach: The Problem of Choice

3.1 Formulation of the problem, and why it is important

The problem. On each step of the described methodology, we have lots of choices:

- we can choose different membership functions to represent different words;
- we can use different t-norms to represent "and";
- we can use different "or"-operations to combine degrees of belief in different rules; and
- finally, we can use different defuzzification procedures.

In principle, we can make all these choices based on *knowledge elicitation techniques*, i.e., based on the detailed interviews with experts. However:

- this detailed elicitation takes too much time, and,
- in reality, although we can force experts to make their statements more precise (this is exactly what knowledge elicitation is about), the resulting numbers will represent not so much expert knowledge, but the (rather arbitrary) result of our forcing. All the knowledge that expert can describe is already contained in the (fuzzy) rules, and although we can extract additional numbers from the experts, these numbers will not represent any additional knowledge.

Of course, we must use *some* knowledge elicitation, e.g., we must describe at least the range of what the experts mean by "small". However, with this *partial* knowledge elicitation, there are still lots of possible choices that are consistent with experts' knowledge.

It is very important to make the right choice. Different choices can lead to a drastically different quality of the resulting control or simulation, so making the right choice is very important.

3.2 How this choice is currently done in fuzzy control: general idea

For *control*, the choice can made based on different criteria; e.g., we should make choices for which:

- either the *smallest* number of rules is needed, on average, to approximate the given control with a given accuracy; or,
- the resulting control is *the best* according to the chosen criterion (i.e., is the most *stable*, or the most *smooth*, etc.).

Let us briefly describe the situations in which the best choice is known.

3.3 Best choice in the sense of best approximation

Choice of membership functions. The authors of [23, 24] compared the quality of the approximation achieved by using different shapes of membership functions. Their numerical experiments have shown that in almost all test situations, the best approximation if we use the "sinc" membership function $\sin(x)/x$.

The paper [12] contains a partial explanation of this result: namely, it is proven that in linear approximation, the function $\sin(x)/x$ is indeed the best (in some reasonable sense). It is desirable to extend this explanation to the general (non-linear) case.

Choice of "and" and "or" operations. In [41], is is shown that the choice of the product $a \cdot b$ as an "and" operation leads to a better approximation than the choice of the minimum $\min(a, b)$.

Choice of defuzzification. In [41], is is shown that the above choice of the centroid defuzzification leads to a better approximation than the *Mean of Maximum* defuzzification.

3.4 Best choice in the sense of best control

Choice of membership functions. The most *robust* membership functions (i.e., the least sensitive to the inaccuracy of the input data) are piecewise-linear ones [25, 28].

This result explains why the piecewise-linear membership functions are, at present, most frequently used.

Choice of "and" and "or" operations. (These results are (mainly) summarized in [18, 19, 25, 28, 39, 4].)

- If we are looking for the *most stable* control, then the best choice is to use $f_{\&}(a,b) = \min(a,b)$ and $f_{\vee}(a,b) = a + b - a \cdot b$ [18, 19, 39, 17].
- If we are looking for the *smoothest* control, then the best choice is to use $f_{\&}(a,b) = a \cdot b$ and $f_{\vee}(a,b) = \min(a,b)$ [18, 19, 39].
- If we are looking for the control that is *most robust* (i.e., least sensitive to the inaccuracy with which we measure the membership functions), then, depending on what exactly we are looking for, we can get two different results:
  - if we are looking for the control that is the most robust *in the the worst case*, then the best choice is to use $f_{\&}(a,b) = \min(a,b)$ and $f_{\vee}(a,b) = \max(a,b)$ [27, 29, 25, 28, 32];
  - if we are looking for the control that is the most robust *in the average*, then the best choice is to use $f_{\&}(a,b) = a \cdot b$ and $f_{\vee}(a,b) = a + b - a \cdot b$ [30, 25, 28, 32];
  - instead of minimizing the *average* error, we can try to minimize the corresponding *entropy* [36, 37, 38, 17, 14, 15]:
    - if we use the *average* entropy (in some reasonable sense), we get the same pair of optimal functions as for average error;
    - for an appropriately defined *worst-case* entropy the optimal operations are $f_{\&}(a,b) = \min(a,b)$ and $f_{\vee}(a,b) = a + b - a \cdot b$.
- Finally, if we are looking for the control that is the *fastest to compute*, then the best choice is to use $f_{\&}(a,b) = \min(a,b)$ and $f_{\vee}(a,b) = \max(a,b)$ [20].

Choice of defuzzification. In [18, 19, 17, 14], we show that the optimal defuzzification is given by the centroid formula.

3.5 A general description of known choices

These optimization results are in good accordance with the general *group-theoretic* approach that enables us to classify techniques that are optimal relative to arbitrary reasonable criteria [18, 19, 39, 4, 26].

Namely, we are looking for the *best* (*optimal*) choices. Normally, the word "best" is understood in the sense of some *numerical* optimality criterion. However, in our case of *fuzzy* choice, it is often difficult to formulate the exact *numerical* criterion. Instead, we assume that there is an *ordinal* criterion,

i.e., that we can compare arbitrary two choices, but that we cannot assign numerical values to these choices. It turns out that in many cases, there are reasonable symmetries, and it is natural to assume that the (ordinal) optimality criterion is invariant with respect to these symmetries. Then, we are able to describe all choices that are optimal with respect to some invariant ordinal optimality criteria.

Right now, we are simply describing the main idea; in the next section, we will return to this idea, and describe it in more detail.

4 Second Problem with the Existing Approach: De-Granulation

What causes this problem. The second problem with the existing fuzzy-control approach is caused by its very nature, namely, by the fact that in this approach, the originally compact representation – in terms of words – is then replaced by a representation in terms of membership functions.

This replacement causes problems.

Too much computer space. The necessity to represent every term by a *function* drastically increases the computer space that is necessary to store the corresponding information.

Too much time. This increase in storage space, in turn, drastically increases the computation time.

The problem re-formulated. We can reformulate this problem as follows: the fuzzy control methodology "de-granulates" the original compact description, and this de-granulation causes an unnecessary increase in computation time.

5 Our Main Idea

Direct transformation is desirable. In view of the above problem, it is desirable to get from the original words *directly* to numerical simulations, thus avoiding this de-granulation.

Let us borrow from the experience of modern physics and use symmetries. In seeking this direct transformation, we will use the experience of modern physics, where symmetry groups are a tool that enables to compress complicated differential equations into compact form (see, e.g., [10, 34, 40]). For example:

- Maxwell's equations of electrodynamics consist of *four* different differential equations for two vector fields: electric field $\vec{E}$ and magnetic filed $\vec{B}$.

- However, if we take into consideration that these equations are invariant with respect to Lorentz transformations (that form the basis of Special Relativity) then we can compress these equations into *two*: $F_{ab}^{,b} = j_a$, and $F_{ab,c} + F_{bc,a} + F_{ca,b} = 0$.

Moreover, the very differential equations themselves can be uniquely deduced from the corresponding symmetry requirements [8, 9, 13, 16] (see also [5, 6, 7]).

It is possible to use symmetry. As we have mentioned, in our previous papers, we have shown that the symmetry group approach can be used to find optimal membership functions, optimal t-norms and t-conorms, and optimal defuzzification procedures.

It is therefore reasonable to expect that the same approach can also be used to *combine* these steps and produce an (optimal) *direct* transformation from words to numerical results.

6 From the General Idea to Precise Methodology: Motivations, Definitions, and Results

6.1 Motivations

We must choose a family of functions. For each situation, and for each i from 1 to n, we must find a function $f_i(x_1, \ldots, x_n)$ that describes the dependence of the rate change $\dot{x}_i$ on the current values $x_1, \ldots, x_n$: $\dot{x}_i = f_i(x_1, \ldots, x_n)$.

In different situations, we will need different functions. Thus, one of our objectives is to describe the functions f_i that correspond to different expert knowledge. In other words, we must select a *family of functions.*

Comment about notations. In the following text, we will denote families of functions by capital letters, such as F, F', G, etc.

Reasonable conditions on the desired family of functions, and what these conditions lead to. For a complex system, we usually have many independent processes that lead to the change in x_i. These processes can be present separately or at the same time.

> For example, the increase in ozone pollution can be caused by industrial pollution, or by frequent thunderstorms.

If the first factor leads to the rate $f(x_1, \ldots, x_n)$, and the second factor leads to the rate $f'(x_1, \ldots, x_n)$, then both factors together lead to the rate $f(x_1, \ldots, x_n) + f'(x_1, \ldots, x_n)$.

Thus, if two functions are reasonable (i.e., belong to the desired family F), their sum should also be reasonable (i.e., should also belong to the same

family F). In mathematical terms, the family F should be *closed under addition.*

The second condition on the desired family F follows from the fact that the intensity of a process can change. Thus, if $f(x_1, \ldots, x_n)$ is a reasonable rate of change, then for every real number λ, the product $\lambda \cdot f(x_1, \ldots, x_n)$ is also a reasonable rate of change:

- the values $\lambda \in (0, 1)$ describe the *decreased* intensity;
- the values $\lambda > 1$ describe the *increased* intensity; and
- the values $\lambda < 0$ describe the *reversed* process.

Thus, if $f \in F$, then $\lambda \cdot f \in F$.

Together with the first condition, we can conclude that if the functions $f_1, \ldots, f_m$ belong to F and $c_1, \ldots, c_m$ are real numbers, then the linear combination $f = c_1 \cdot f_1 + \ldots + c_m \cdot f_m$ must also belong to the family F. In mathematical terms, the family $\mathcal{F}$ must be a *linear space.*

It is known, from linear algebra, that linear spaces can be described as follows: every linear space has a subset $\{e_1, e_2, \ldots\}$ called a *basis*, such that every element e from the linear space can be represented as a linear combination of elements from this basis: $e = c_1 \cdot e_1 + c_2 \cdot e_2 + \ldots$ The smallest possible number of elements in this basis is called a *dimension* of the linear space.

In principle, some spaces are infinite-dimensional, but with an infinite basis, we can represent an arbitrary function of n variables; so, if we want our family to be meaningful, we must restrict ourselves only to *finite-dimensional* linear spaces, i.e., to linear spaces F formed by functions of the type $f(x_1, \ldots, x_n) = c_1 \cdot e_1(x_, \ldots, x_n) + \ldots + e_m \cdot e_m(x_1, \ldots, x_m)$, where $e_j(x_1, \ldots, x_n)$ are fixed functions, and c_j are arbitrary real numbers.

For such families, choosing the family means choosing the corresponding m functions $e_1(x_1, \ldots, x_n), \ldots, e_m(x_1, \ldots, x_n)$.

We must choose the best family of functions. We want to select the *best* transformation from expert words to functions. This means, in particular, that we are interested in choosing the *best* family of functions.

What is a criterion for choosing a family of functions? What does it mean to choose a *best* family of functions? It means that we have some *criterion* that enables us to choose between the two families.

Traditionally, optimality criteria are *numerical*, i.e., to every family F, we assign some value $J(F)$ expressing its quality, and choose a family for which this value is maximal (i.e., when $J(F) \geq J(G)$ for every other alternative G). However, it is not necessary to restrict ourselves to such numeric criteria only.

For example, if we have several different families F that have the same prediction ability $P(F)$, we can choose between them the one that has the

minimal computational complexity $C(F)$. In this case, the actual criterion that we use to compare two families is not numeric, but more complicated:

> *A family F_1 is better than the family F_2 if and only if*
>
> - *either $P(F_1) > P(F_2)$,*
> - *or $P(F_1) = P(F_2)$ and $C(F_1) < C(F_2)$.*

A criterion can be even more complicated.

The only thing that a criterion *must* do is to allow us, for every pair of families (F_1, F_2), to make one of the following conclusions:

- the first family is better with respect to this criterion (we'll denote it by $F_1 \succ F_2$, or $F_2 \prec F_1$);
- with respect to the given criterion, the second family is better ($F_2 \succ F_1$);
- with respect to this criterion, these families have the same quality (we'll denote it by $F_1 \sim F_2$);
- this criterion does not allow us to compare the two families.

Of course, it is necessary to demand that these choices be consistent.

> For example, if $F_1 \succ F_2$ and $F_2 \succ F_3$ then $F_1 \succ F_3$.

The criterion must be final, i.e., it must pick the unique family as the best one. A natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family).

The reason for this demand is very simple:

- If a criterion *does not choose* any family at all, then it is of no use.
- If *several* different families are the best according to this criterion, then we still have a problem to choose among those best. Therefore we need some additional criterion for that choice, like in the above example:

  > If several families $F_1, F_2, \ldots$ turn out to have the same prediction ability ($P(F_1) = P(F_2) = \ldots$), we can choose among them a family with minimal computational complexity ($C(F_i) \to \min$).

  So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if:

  - either it was better according to the old criterion,

– or they had the same quality according to the old criterion and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family, it means that this criterion is not final, we'll have to modify it until we come to a final criterion that will have that property.

The criterion must not change if we change the measuring units for one of the variables x_i. The exact mathematical form of a function $f_i(x_1, \ldots, x_n)$ depends on the exact choice of units for measuring $x_1, \ldots, x_n$. If, for some j, we replace a unit for measuring x_j by a new unit that is λ_j times larger, then the same physical value that was previously described by a numerical value x_j will now be described, in the new units, by a new numerical value $\tilde{x}_j = x_j/\lambda_j$. For example, if we replace centimeters by inches, with $\lambda_j = 2.54$, then $x_j = 5.08$ cm becomes $\tilde{x}_j = x_j/\lambda_j = 2$ in.

How will the dynamical equations $\dot{x}_i = f_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)$ change if we use the new unit? In terms of $\tilde{x}_j$, we have $x_j = \lambda_j \cdot \tilde{x}_j$, and thus, we have $\dot{x}_i = f_i(x_1, \ldots, x_{j-1}, \lambda_j \cdot x_j, x_{j+1}, \ldots, x_n)$. In other words, if we change the measuring unit for x_j, the same dynamics that was originally represented by a function $f_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)$, will be described, in the new units, by a function
$\tilde{f}(x_1, \ldots, x_n) = f_i(x_1, \ldots, x_{j-1}, \lambda_j \cdot x_j, x_{j+1}, \ldots, x_n)$.

If we make a similar replacement of the measuring units for several quantities x_j, so that x_1 is replaced by a unit that is λ_1 times larger, x_2 by a unit that is λ_2 times larger, etc., then each function $f_i(x_1, \ldots, x_n)$ will be replaced by a new function $\tilde{f}_i(x_1, \ldots, x_n) = \lambda_i^{-1} \cdot f_i(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n)$.

It is reasonable to assume that the relative quality of different families should not change if we simply change the units, i.e., if the family F is better than a family G, then the transformed family $\tilde{F}$ should also be better than the family $\tilde{G}$.

We are now ready for the formal definitions.

6.2 Definitions

Definition 1. *Let two positive integers $n, m \geq 1$ be fixed, and let $i \leq n$.*

- *By a family F, we mean a collection of m differentiable function $e_1(x_1, \ldots, x_n), \ldots, e_m(x_1, \ldots, x_n)$.*
- *We say that a function $e(x_1, \ldots, x_n)$ belongs to the family F (and that F contains the function $e(x_1, \ldots, x_n)$) if this function can be represented as a linear combination of the functions e_j, i.e., if there exist m real numbers $c_1, \ldots, c_m$ for which, for all x_k, $e(x_1, \ldots, x_n) = c_1 \cdot e_1(x_1, \ldots, x_n) + \ldots + c_m \cdot e_m(x_1, \ldots, x_m)$.*
- *Two families F and G are considered equal if they contain the same functions.*

Denotation. *Let's denote the set of all possible families by* Φ.

Comment. In order to formalize the notion of an optimality criterion, we must describe that for some pairs of families, F is better than G, and for some other pairs, F is not better than G. To describe this "relation" it better, we must, thus, describe the *set* of all possible pairs (F, G) for which F is better than G. In mathematics, if a set Φ is given:

- the set of all pairs (F_1, F_2) of elements $F_1 \in \Phi$, $F_2 \in \Phi$, is usually denoted by $\Phi \times \Phi$.
- An arbitrary subset R of a set of pairs $\Phi \times \Phi$ is called a *relation* on the set Φ. If $(F_1, F_2) \in R$, it is said that F_1 and F_2 are in relation R; this fact is denoted by $F_1 R F_2$.

Definition 2. *A pair of relations* $(\prec, \sim)$ *on a set* Φ *is called consistent if it satisfies the following conditions, for every* $F, G, H \in \Phi$*:*

(1) *if* $F \prec G$ *and* $G \prec H$ *then* $F \prec H$*;*

(2) $F \sim F$*;*

(3) *if* $F \sim G$ *then* $G \sim F$*;*

(4) *if* $F \sim G$ *and* $G \sim H$ *then* $F \sim H$*;*

(5) *if* $F \prec G$ *and* $G \sim H$ *then* $F \prec H$*;*

(6) *if* $F \sim G$ *and* $G < H$ *then* $F < H$*;*

(7) *if* $F \prec G$ *then it is not true that* $G \prec F$*, and it is not true that* $F \sim G$*.*

Comment. The intended meaning of these relations is as follows:

- $F \prec G$ means that with respect to a given criterion, G is better than F;
- $F \sim G$ means that with respect to a given criterion, F and G are of the same quality.

Under this interpretation, conditions (1)–(7) have simple intuitive meaning:

(1) if G is better than F, and H is better than G, then H is better than F;

(2) every alternative F is of the same quality as itself;

(3) if G is of the same quality as F, then F is of the same quality as G;

(4) if F is of the same quality as G, and G is of the same quality as H, then F is of the same quality as H;

(5) if G is better than F, and H is of the same quality as G, then H is also better than F;

(6) if H is better than G, and F is of the same quality as G, then H is better than F;

(7) if G is better than F, then F cannot be better than G and F cannot be of the same quality as G.

Definition 3. *Assume a set Φ is given. Its elements will be called alternatives.*

- *By an optimality criterion, we mean a consistent pair $(\prec, \sim)$ of relations on the set Φ of all alternatives.*
  - *If $F \succ G$ we say that F is better than G;*
  - *if $F \sim G$ we say that the alternatives F and G are equivalent with respect to this criterion.*
- *We say that an alternative F is optimal (or best) with respect to a criterion $(\prec, \sim)$ if for every other alternative G either $F \succ G$ or $F \sim G$.*
- *We say that a criterion is final if there exists an optimal alternative, and this optimal alternative is unique.*

Comment. In this paper, we will consider optimality criteria on the set Φ of all families.

Definition 4. *Let $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$ be a tuple of positive real numbers.*

- *By a $\vec{\lambda}$-rescaling of a function $f(x_1, \ldots, x_n)$ we mean a function $\tilde{f}(x_1, \ldots, x_n) = \lambda_i^{-1} \cdot f_i(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n)$.*
- *By a $\vec{\lambda}$-rescaling of a family of functions F we mean the family consisting of $\vec{\lambda}$-rescalings of all functions from F.*

Denotation. *$\vec{\lambda}$-rescaling of a family F will be denoted by $R_{\vec{\lambda}}(F)$.*

Definition 5. *We say that an optimality criterion on Φ is unit-invariant if for every two families F and G and for every vector $\vec{\lambda}$, the following two conditions are true:*

i) if F is better than G in the sense of this criterion (i.e., $F \succ G$), then $R_{\vec{\lambda}}(F) \succ R_{\vec{\lambda}}(G)$;

ii) if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $R_{\vec{\lambda}}(F) \sim R_{\vec{\lambda}}(G)$.

Comment. As we have already remarked, the demands that the optimality criterion is final and unit-invariant are quite reasonable. At first glance they may seem rather trivial and therefore weak, because these demands do not specify the exact optimality criterion. However, these demands are strong enough, as the following theorem shows:

6.3 Main result

Theorem. *If a family F is optimal in the sense of some optimality criterion that is final and unit-invariant, then every function $f_i(x_1, \ldots, x_n)$ from this family F is a linear combination of the functions of the type*

$$x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n} \cdot \ln^{k_1}(x_1) \cdot \ldots \cdot \ln^{k_n}(x_n),$$

where α_j are complex numbers, and k_j are non-negative integers.

Comment. The above expression can be re-formulated without complex numbers; in this case, the basic functions are of the type

$$x_1^{\alpha_1} \cdot \sin(\beta_1 \cdot \ln(x_1) + \varphi_1) \cdot \ldots \cdot x_n^{\alpha_n} \cdot \sin(\beta_n \cdot \ln(x_n) + \varphi_n) \cdot \ln^{k_1}(x_1) \cdot \ldots \cdot \ln^{k_n}(x_n),$$

where α_j, β_j, and φ_j are real numbers, and k_j are non-negative integers.

In particular, for $n = 1$, we get the following result:

Corollary 1. *For $n = 1$, if an m-dimensional family F is optimal in the sense of some optimality criterion that is final and unit-invariant, then every function $f(x)$ from the family F is equal to a linear combination of the functions of the type $\ln^p(x) \cdot x^\alpha \cdot \sin(\beta \cdot \ln(x) + \varphi)$, where p is a non-negative integer, α, β and φ are real numbers.*

Corollary 2. *Let a 2-dimensional family F be optimal in the sense of some optimality criterion that is final and unit-invariant. Then, every function $f(x)$ from the family F has one of the following forms:*

1. $f(x) = C_1 \cdot x^{\alpha_1} + C_2 \cdot x^{\alpha_2}$;
2. $f(x) = C_1 \cdot x^\alpha + C_2 \cdot x^\alpha \cdot \ln(x)$;
3. $f(x) = C \cdot x^\alpha \cdot \sin(\beta \cdot \ln(x) + \varphi)$.

Comment. The optimal families that we have just described are exactly the ones that were described, on a semi-heuristic basis, by Ludwig von Bertalanffy in his General System Theory (see, e.g., his books [2, 3]).

Bertalanffy mainly considered equations of the *first* type. These so-called *Bertalanffy equations* turned out to be very adequate for describing growth in biology (namely, the growth of individual organisms and of their organs), so adequate that they are routinely used by fisheries in England and Japan and by by the Food and Agriculture Organization (FAO) of the United Nations. The following particular cases of the Bertalanffy equation describe the simplest growth processes:

- For $\alpha_1 = 1$, $C_1 > 0$, and $C_1 = 0$, we get the equation $\dot{x} = C_1 \cdot x$ that describe an *exponential* growth $x(t) = C \cdot \exp(C_1 \cdot t)$.

- For $\alpha_1 = 1$, $\alpha_2 = 2$, $C_1 > 0$, and $C_2 < 0$, we get the equation $\dot{x} = C_1 \cdot x - |C_2| \cdot x^2$ that describes a so-called *logistic curve* that starts with an exponential growth but then flatters out. For this particular growth function, the growth equation also admits an explicit solution

$$L(t) = \frac{1}{K + A \cdot b^t}.$$

Equations of the *second* type were originally proposed by Gompertz (for $\alpha = 1$). These equations describe, e.g., such growth processes as *population dynamics* (see, e.g., [33]),

Thus, our general approach provides a precise mathematical justification for the (highly successful) semi-heuristic formulas of von Bertalanffy's general system theory.

6.4 How to use this result: examples

Example 1. If an increase in x_1 leads to a slower increase rate of x_2, this means that we have a term in $\dot{x}_2 = f_2(x_1, \ldots, x_n)$ that is decreasing with x_1. Since this terms should be monotonic, it should not contain sines, and therefore, it should be of the form $C \cdot x_2^{\alpha} \cdot \ln^k(x_2)$. The exact values of the coefficients must be determined in one of the following two ways:

- either by showing the expert the results of different values and asking this expert to choose the most appropriate value;
- or by tuning the resulting simulation to the actual recorded behavior of the system that we are simulating.

Example 2. Similarly, if we know that, e.g., x_3 starts decreasing if both x_1 and x_2 are present, then we should add, to f_3, terms of the type $-x_1^{\alpha_1} \cdot x_2^{\alpha_2}$, maybe with logarithmic terms as well.

What did we gain? At first glance, there still seems to be a lot of freedom of choice, and this is inevitable, because we are developing a general formalism that should cover many different systems. However, we did gain a lot:

- initially, we had the choice of choosing several arbitrary *functions* (membership function, etc.);
- now, we only need to choose a few *parameters*.

We have less choice, thus, more granularity, and less computation time.

7 Case study: sedimental system

This idea was used in geology, for simulating a *sedimentary system* (for details, see [35]). We tried a simplified system in which the state is characterized by the following three parameters:

- $x_1 = h$ is the hinterland elevation,
- $x_2 = s$ is the sealevel elevation, and
- $x_3 = r$ is the sediment transport rate.

Several reasonable rules can be formulated about the evolution of these three variables; e.g.:

- on one hand, sediment transport erodes the hinterland and eventually reduces its elevation;
- on the other hand, it causes isostatic uplift and thus, after a longer period of time, increases the hinterland elevation.

These two rules lead to the terms $-k_1 \cdot r(t-\Delta_1)^{\alpha_1} + k_2 \cdot r(t-\Delta_2)^{\alpha_2}$ in $\dot{x}_1(t)$, where Δ_i are the appropriate time delays.

After transforming all other rules into the corresponding terms, we got a system of differential equations, for which the numerical simulation is in very good accordance with the geological data.

Acknowledgments. This work was supported in part by NASA under cooperative agreement NCCW-0089, by NSF grants No. DUE-9750858 and EEC-9322370, and by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518.

References

[1] R. E. Bellman. *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1970.

[2] L. v. Bertalanffy, *Perspectives on general system theory*, G. Braziller, N.Y., 1975.

[3] L. v. Bertalanffy, *General system theory* (G. Braziller, N.Y., 1984.

[4] B. Bouchon-Meunier, V. Kreinovich, A. Lokshin, and H. T. Nguyen, "On the formulation of optimization under elastic constraints (with control in mind)", *Fuzzy Sets and Systems,* 1996, Vol. 81, No. 1, pp. 5–29.

[5] A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry, error estimation, and other applications of set-valued analysis", *ACM SIGNUM Newsletter*, 1996, Vol. 31, No. 4, pp. 3–25.

[6] A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: towards mathematical foundations", *International Journal of Theoretical Physics*, 1997, Vol. 36, No. 4, pp. 1009–1020.

[7] A. Finkelstein, O. Kosheleva, and V. Kreinovich, Andrei Finkelstein, Olga Kosheleva, and Vladik Kreinovich, "Astrogeometry: geometry explains shapes of celestial bodies", *Geombinatorics*, 1997, Vol. VI, No. 4, pp. 125–139.

[8] A. M. Finkelstein and V. Kreinovich, "Derivation of Einstein's, Brans-Dicke and other equations from group considerations," *On Relativity Theory. Proceedings of the Sir Arthur Eddington Centenary Symposium, Nagpur India 1984*, Vol. 2, Y. Choque-Bruhat and T. M. Karade (eds), World Scientific, Singapore, 1985, pp. 138–146.

[9] A. M. Finkelstein, V. Kreinovich, and R. R. Zapatrin, "Fundamental physical equations uniquely determined by their symmetry groups," *Lecture Notes in Mathematics*, Springer-Verlag, Berlin-Heidelberg-N.Y., Vol. 1214, 1986, pp. 159–170.

[10] *Group theory in physics: proceedings of the international symposium held in honor of Prof. Marcos Moshinsky, Cocoyoc, Morelos, Mexico, 1991*, American Institute of Physics, N.Y., 1992.

[11] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[12] O. Kosheleva, "Why Sinc? Signal Processing Helps Fuzzy Control", *SC-COSMIC, South Central Computational Sciences in Minority Institutions Consortium, Second Student Conference in Computational Sciences*, October 25–27, 1996, El Paso, TX, Abstracts, pp. 19–21.

[13] V. Kreinovich. "Derivation of the Schroedinger equations from scale invariance," *Theoretical and Mathematical Physics*, 1976, Vol. 8, No. 3, pp. 282–285.

[14] V. Kreinovich, "Random sets unify, explain, and aid known uncertainty methods in expert systems", in J. Goutsias, R. P. S. Mahler, and H. T. Nguyen (eds.), *Random Sets: Theory and Applications*, Springer-Verlag, N.Y., 1997, pp. 321–345.

[15] V. Kreinovich, "Maximum entropy and interval computations", *Reliable Computing*, 1996, Vol. 2, No. 1, pp. 63–79.

[16] V. Kreinovich and L. Longpré, "Unreasonable effectiveness of symmetry in physics", *International Journal of Theoretical Physics*, 1996, Vol. 35, No. 7, pp. 1549–1555.

[17] V. Kreinovich, H. T. Nguyen, and E. A. Walker, "Maximum entropy (MaxEnt) method in expert systems and intelligent control: new possibilities and limitations", In: K. Hanson and R. Silver, Eds., *Maximum Entropy and Bayesian Methods, Santa Fe, New Mexico, 1995*, Kluwer Academic Publishers, Dordrecht, Boston, 1996, pp. 93–100.

[18] V. Kreinovich, C. Quintana, and R. Lea, "What procedure to choose while designing a fuzzy control? Towards mathematical foundations of fuzzy control", *Working Notes of the 1st International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, TX, 1991, pp. 123–130.

[19] V. Kreinovich, C. Quintana, R. Lea, O. Fuentes, A. Lokshin, S. Kumar, I. Boricheva, and L. Reznik, "What non-linearity to choose? Mathematical foundations of fuzzy control", *Proceedings of the 1992 International Conference on Fuzzy Systems and Intelligent Control*, Louisville, KY, pp. 349–412.

[20] V. Kreinovich and D. Tolbert, "Minimizing computational complexity as a criterion for choosing fuzzy rules and neural activation functions in intelligent control". In: M. Jamshidi, C. Nguyen, R. Lumia, and J. Yuh (Editors), *Intelligent Automation and Soft Computing. Trends in Research, Development, and Applications. Proceedings of the First World Automation Congress (WAC'94), August 14–17, 1994, Maui, Hawaii*, TSI Press, Albuquerque, NM, Vol. 1, pp. 545–550.

[21] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant", *Proceedings of the IEE*, 1974, Vol. 121, No. 12, pp. 1585–1588.

[22] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic systems", *IEEE Transactions on Computing*, 1977, Vol. 26, pp. 1182–1191.

[23] S. Mitaim and B. Kosko, "What is the best shape for a fuzzy set in function approximation?", *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems FUZZ-IEEE'96*, New Orleans, September 8–11, 1996, Vol. 2, pp. 1237–1243.

[24] S. Mitaim and B. Kosko, "Fuzzy function approximation and the shape of fuzzy sets", *Proceedings of the World Congress on Neural Networks WCNN'96*, September 1996.

[25] H. T. Nguyen and V. Kreinovich, "Towards theoretical foundations of soft computing applications", *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 1995, Vol. 3, No. 3, pp. 341–373.

[26] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[27] H. T. Nguyen, V. Kreinovich, B. Lea, D. Tolbert, "How to control if even experts are not sure: robust fuzzy control", *Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, December 2–4, 1992, pp. 153–162.

[28] H. T. Nguyen, V. Kreinovich, B. Lea, and D. Tolbert, "Interpolation that leads to the narrowest intervals, and its application to expert systems and intelligent control", *Reliable Computing*, 1995, Vol. 3, No. 1, pp. 299–316.

[29] H. T. Nguyen, V. Kreinovich, and D. Tolbert, "On robustness of fuzzy logics". *Proceedings of IEEE-FUZZ International Conference*, San Francisco, CA, March 1993, Vol. 1, pp. 543–547.

[30] H. T. Nguyen, V. Kreinovich, and D. Tolbert, "A measure of average sensitivity for fuzzy logics", *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 1994, Vol. 2, No. 4, pp. 361–375.

[31] H. T. Nguyen and M. Sugeno (eds.), *Handbooks of Fuzzy Systems. Vol. VI: Fuzzy Modeling and Control*, Kluwer Academic, 1997 (to appear).

[32] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, FL, 1997.

[33] A. G. Nobile, L. M. Riccardi, L. Sacerdote, "On a class of difference equations modeling growth processes", In: L. M. Riccardi and A. C. Scott, *Biomathematics in 1980*, North-Holland, Amsterdam, 1982, pp. 217–244.

[34] P. J. Olver, *Equivalence, invariants, and symmetry*, Cambridge University Press, Cambridge, N.Y., 1995.

[35] B. S. Penn and J. W. Harbaugh, *DYNASED: a model for tightly coupled sedimentary systems*, NASA PACES Center, El Paso, TX, Technical Report, 1996.

[36] A. Ramer and V. Kreinovich, "Maximum entropy approach to fuzzy control", *Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, December 2–4, 1992, pp. 113–117.

[37] A. Ramer and V. Kreinovich, "Information complexity and fuzzy control", Chapter 4 in: A. Kandel and G. Langholtz (Eds.), *Fuzzy Control Systems*, CRC Press, Boca Raton, FL, 1994, pp. 75–97.

[38] A. Ramer and V. Kreinovich, "Maximum entropy approach to fuzzy control", *Information Sciences*, 1994, Vol. 81, No. 3–4, pp. 235–260.

[39] M. H. Smith and V. Kreinovich, "Optimal strategy of switching reasoning methods in fuzzy control", Chapter 6 in H. T. Nguyen, M. Sugeno, R. Tong, and R. Yager (eds.), *Theoretical aspects of fuzzy control*, J. Wiley, N.Y., 1995, pp. 117–146.

[40] *Symmetries in physics: proceedings of the international symposium held in honor of Prof. Marcos Moshinsky, Cocoyoc, Morelos, Mexico, 1991*, Springer-Verlag, Berlin, N.Y., 1992.

[41] X.-J. Zeng and M. G. Singh, "Approximation accuracy analysis of fuzzy systems as function approximators," *IEEE Trans. on Fuzzy Systems*, 1996, Vol. 4, pp. 44–63.

8 Appendix: Proof

This proof is based on the following auxiliary result of independent interest:

Proposition. *If an optimality criterion is final and unit-invariant, then the optimal family F_{opt} is also unit-invariant, i.e., $R_{\vec{\lambda}}(F_{opt}) = F_{opt}$ for every vector $\vec{\lambda}$.*

Proof of the Proposition. Since the optimality criterion is final, there exists a unique family F_{opt} that is optimal with respect to this criterion, i.e., for every other F:

- either $F_{opt} \succ F$
- or $F_{opt} \sim F$.

To prove that $F_{opt} = R_{\vec{\lambda}}(F_{opt})$, we will first show that the re-scaled family $R_{\vec{\lambda}}(F_{opt})$ is also optimal, i.e., that for every family F:

- either $R_{\vec{\lambda}}(F_{opt}) \succ F$
- or $R_{\vec{\lambda}}(F_{opt}) \sim F$.

If we prove this optimality, then the desired equality will follow from the fact that our optimality criterion is final and therefore, there is only one optimal family (so, since the families F_{opt} and $R_{\vec{\lambda}}(F_{opt})$ are both optimal, they must be the same family).

Let us show that $R_{\vec{\lambda}}(F_{opt})$ is indeed optimal. How can we, e.g., prove that $R_{\vec{\lambda}}(F_{opt}) \succ F$? Since the optimality criterion is unit-invariant, the desired relation is equivalent to $F_{opt} \succ R_{\vec{\lambda}^{-1}}(F)$, where by $\vec{\lambda}^{-1}$, we denoted a tuple $(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$. Similarly, the relation $R_{\vec{\lambda}}(F_{opt}) \sim F$ is equivalent to $F_{opt} \sim R_{\vec{\lambda}^{-1}}(F)$.

These two equivalences allow us to complete the proof of the proposition. Indeed, since F_{opt} is optimal, we have one of the two possibilities:

- either $F_{opt} \succ R_{\vec{\lambda}^{-1}}(F)$,
- or $F_{opt} \sim R_{\vec{\lambda}^{-1}}(F)$.

In the first case, we have $R_{\vec{\lambda}}(F_{opt}) \succ F$; in the second case, we have $R_{\vec{\lambda}}(F_{opt}) \sim F$. Thus, whatever family F we take, we always have either $R_{\vec{\lambda}}(F_{opt}) \succ F$, or $R_{\vec{\lambda}}(F_{opt}) \sim F$. Hence, $R_{\vec{\lambda}}(F_{opt})$ is indeed optimal and thence, $R_{\vec{\lambda}}(F_{opt}) = F_{opt}$. The proposition is proven.

Let us now prove the theorem. Since the criterion is final, there exists an optimal family $F_{opt} = \{c_1 \cdot e_1(x_1, \ldots, x_n) + \ldots + c_m \cdot e_m(x_1, \ldots, x_n)\}$. Eaxh of the corresponding functions $e_j(x_1, \ldots, x_n)$ belongs to the family F_{opt} (for $c_j = 1$ and $c_k = 0$ for $k \neq j$).

Due to the Proposition, the optimal family is unit-invariant, i.e., $F_{opt} = R_{\vec{\lambda}}(F_{opt})$. In particular, this means that for very j, and for every $\vec{\lambda}$, we have $R_{\vec{\lambda}}(e_j) \in F_{opt}$, i.e.,

$$e_j(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n) =$$

$$c_{j1}(\vec{\lambda}) \cdot e_1(x_1, \ldots, x_n) + \ldots + c_{jm}(\vec{\lambda}) \cdot e_m(x_1, \ldots, x_n). \qquad (1)$$

for some values c_{jk}. If we take m different values of $(x_1, \ldots, x_n)$, then the corresponding equations (1) form a system of m linear equations to determine m coefficients $c_{j1}(\vec{\lambda}), \ldots, c_{jm}(\vec{\lambda})$. The well-known Cramer's rule describes the solution of a system of linear equation as a ratio of two determinants and thus, as a differentiable function of the coefficients and right-hand sides of these equations. Since $e_j(x_1, \ldots, x_n)$ are differentiable functions, we can thus conclude that the functions $c_j(\vec{\lambda})$ are differentiable too.

Since both sides of the equation (1) is differentiable, let us pick an arbitrary $l = 1, \ldots, n$, differentiate both sides with respect to λ_l, and then substitute $\lambda_l = \ldots = \lambda_n = 1$. As a result, we get the following system of differential equations:

$$x_l \cdot \frac{\partial e_j}{\partial x_l}(x_1, \ldots, x_n) = \sum_{k=1}^{m} c_{jkl} \cdot e_k(x_1, \ldots, x_n), \qquad (2)$$

where we denoted

$$c_{jkl} = \left.\frac{\partial c_{jk}(\lambda_1, \ldots, \lambda_n)}{\partial \lambda_l}\right|_{\lambda_1 = \ldots = \lambda_n = 1}.$$

This equation can be further simplified if we use new variables $X_j = \ln(x_j)$, for which $dx_l/x_l = dX_l$. In terms of these new variables, $x_j = \exp(X_j)$, and the values $e_j(x_1, \ldots, x_n)$ take the form $e_j(x_1, \ldots, x_n) = E_j(X_1, \ldots, X_n)$, where we denoted $E_j(X_1, \ldots, X_n) = e_j(\exp(X_1), \ldots, \exp(X_n))$. In terms of the new function $E_j(X_1, \ldots, X_n)$, the equation (2) takes the following form:

$$\frac{\partial E_j}{\partial X_l}(X_1, \ldots, X_n) = \sum_{k=1}^{m} c_{jkl} \cdot E_k(X_1, \ldots, X_n). \qquad (3)$$

If we fix all the variables but one (e.g., except for X_1), we conclude that the functions $E_1(X_1), \ldots, E_m(X_1)$ satisfy a system of linear differential equations

with constant coefficients. A general solution of such a system is well known (see, e.g., [1]): it has a form

$$E_j(X_1) = \sum C_{jp} \cdot \exp(\alpha_p \cdot X_1) \cdot X_1^{k_p}, \tag{4}$$

where α_p are complex numbers (eigenvalues of the coefficient matrix), C_p are complex numbers, and k_p are non-negative integers.

If we take into consideration the dependence on X_2, then all the coefficients of the formula (4) should depend on X_2, i.e.,

$$E_j(X_1, X_2) = \sum C_{jp}(X_2) \cdot \exp(\alpha_p(X_2) \cdot X_1) \cdot X_1^{k_p(X_2)}. \tag{5}$$

Since the dependence on X_2 is smooth (hence, continuous), and k_p is an integer, we conclude that k_p is a constant: $k_p(X_2) = k_p$. The dependence on all other coefficients on X_2 can be determined from the fact that, similarly to (4), for a fixed X_1, we must have a similar expression in terms of X_2:

$$E_j(X_2) = \sum C'_{jp} \cdot \exp(\alpha'_p \cdot X_2) \cdot X_2^{k'_p}. \tag{6}$$

Thus, the only possible dependence of C_{jp} on X_2 is a dependence of the type $\exp(\alpha'_p \cdot X_2) \cdot X_2^{k'_p}$, and the only possible dependence of α_p on X_2 is linear, i.e., we get

$$E_j(X_1, X_2) = \sum C_{jp} \cdot \exp(\alpha_{p1} \cdot X_1 + \alpha_{p2} \cdot X_2 + \alpha'_p \cdot X_1 \cdot X_2) \cdot X_1^{k_{p1}} \cdot X_2^{k_{p2}}. \tag{7}$$

We started with the system (3). This system remains similar if we make a linear change of variables, e.g., if we replace X_1 and X_2 by $X'_1 = X_1 + X_2$ and $X'_2 = X_1 - X_2$. Therefore, we would like to get a similar formula (7) in the new variables. If $\alpha'_p \neq 0$, we get the undesired quadratic term in the exponential expression. Thus, $\alpha'_p = 0$, and (7) take the form

$$E_j(X_1, X_2) = \sum C_{jp} \cdot \exp(\alpha_{p1} \cdot X_1 + \alpha_{p2} \cdot X_2) \cdot X_1^{k_{p1}} \cdot X_2^{k_{p2}}. \tag{7}$$

Similarly, if we take into consideration the dependence on all n variables X_k, we conclude that

$$E_j(X_1, \ldots, X_n) = \sum C_{jp} \cdot \exp(\alpha_{p1} \cdot X_1 + \ldots + \alpha_{pn} \cdot X_n) \cdot X_1^{k_{p1}} \cdot \ldots \cdot X_n^{k_{pn}}. \tag{8}$$

Substituting $X_k = \ln(x_k)$ into this formula (8), we get the desired expression for $e_j(x_1, \ldots, x_n)$:

$$e_j(x_1, \ldots, x_n) = \sum C_{jp} \cdot x_1^{\alpha_{p1}} \cdot \ldots \cdot x_n^{\alpha_{pn}} \cdot \ln^{k_{p1}}(x_1) \cdot \ldots \cdot \ln^{k_{pn}}(x_n).$$

The theorem is proven.