A New Differential Formalism for Optimization of Set Functions

Hung T. Nguyen[1] and Vladik Kreinovich[2]

[1]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003
email hunguyen@nmsu.edu

[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968
email vladik@cs.utep.edu

Abstract

In many practical problems, we must optimize a *set function*, i.e., find a set a for which $f(a) \to \max$, where f is a function defined on the class of sets. Such problems appear in design, in image processing, in game theory, etc.

In particular, if we use statistical methods (i.e., random sets) to solve the corresponding set-related practical problems, then, at the end, we get a probability distribution on the class of all sets; to arrive at a single answer, we can, e.g., apply the Maximum Likelihood Method, i.e., optimize the likelihood set function $L(a)$.

Most optimization problems can be solved (or at least simplified) by using the fact that small deviations from an optimal solution can only decrease the value of the objective function; as a result, some derivative must be equal to 0. This approach has been successfully used, e.g., for set functions in which the desired set a is a *shape*, i.e., a smooth (or piecewise smooth) surface. In some real-life problems, however, there are no reasons to assume smoothness and therefore, these existing methods are not applicable. For such problems, we design a new differential formalism for optimizing set functions.

1 Introduction: Optimization of Set Functions is a Practically Important but Difficult Problem

Optimization is important. In most application problems, we have a class $\mathcal{A}$ of possible alternatives, and we must find the *best* of these alternatives, i.e., the alternative $a \in \mathcal{A}$ for which the given *objective function* $f : \mathcal{A} \to R$ takes the largest possible value:

$$f(a) \to \max_{a \in \mathcal{A}}.$$

An important particular case of this problem is when the objective function $f(a)$ describes the posterior probability (likelihood) of an alternative a; the corresponding optimization problem is called the *maximum likelihood method* (MLM) of choosing an alternative.

Simplest case: finitely many alternatives. The larger the class of alternatives $\mathcal{A}$, the more complicated the corresponding optimization problem.

The simplest case if when we have finitely many alternatives; this happens, e.g., when we compare several proposed designs for a new car, or several proposed plans of economic development, etc. In this case, we can simply try all of them, and choose the one for which the objective function takes the largest possible value.

The very fact that this case if the *simplest* does not mean that it is *easy*: In many real-life problems, we have *several* features to choose; even if we have only two options for each of n feature, we have 2^n possible choices; for large n, this number grows so fast that exhaustive search becomes practically impossible. Moreover, for some finite optimization problems (called *NP-hard*), not only exhaustive search is computationally impossible, but any other algorithm requires, for some instances at least, the unrealistic exponential time (see, e.g., [22, 50]).

Next simplest case: optimization of functions of one or several variables. The simplest case is when alternatives from the class $\mathcal{A}$ can be uniquely characterized by a single real-valued parameter x, or by finitely many real-valued parameters $x_1, \ldots, x_n$. Since the alternative a is uniquely determined by the values of these parameters, the value of the objective function $f(a)$ (that describes the relative quality of different alternatives) is also uniquely determined by these values: $f(a) = F(x_1, \ldots, x_n)$. In this case, finding the best alternative means finding the values $x_1, \ldots, x_n$ for which $F(x_1, \ldots, x_n) \to \max$.

In spite of its perceived simplicity, in some cases, this problem is extremely computationally complicated (see, e.g., [32, 51, 60]), However, usually, this optimization problem is simplified by the fact that the function $F(x_1, \ldots, x_n)$ if smooth (differentiable). This means, in particular, that its maximum is located at a point when all n partial derivatives $\partial F / \partial x_i$ are equal to 0.

- In some cases, the resulting n equations $\partial F/\partial x_i = 0$ are sufficient to find x_i.

- In more complicated cases, these equations are not sufficient to immediately find the desired values x_i, but we can use the partial derivatives to design an iterative method that converges to the desired values x_i really fast. For example, we can use a *gradient method* in which the next iteration x_i' is obtained from the previous iteration x_i by a formula $x_i' = x_i - \lambda \cdot \partial F/\partial x_i$ (or a modification of this method).

In all these cases, differentiability helps.

Optimization of functionals. An even more complex case if when the alternatives cannot be characterized by finitely many parameters, but require a *function* (or finitely many functions) to describe. A typical example is theoretical physics, where most fundamental theories are formulated in terms of the so-called *variational principles*, i.e., principles of the type $S(a) \to \max$, where a denotes a function (physical field $\varphi(x)$) or several fields, and $S(a)$ is a functional called *action*.

For example, General Relativity Theory is formulated in this form with a equal to the metric field $g_{ij}(x)$ and $S = \int R \cdot \sqrt{-g}\, d^4x$. Most modern physical theories are formulated in terms of the so-called *Lagrangian*, i.e., a function L for which $S = \int L\, d^4x$.

To find solutions to the corresponding variational problem, we can use the formalism of *variational (functional) derivatives* and equate the corresponding functional derivatives with 0. As a result, for all classical theories, we get the corresponding partial differential equations (see, e.g., [18, 34]).

If we cannot directly solve these equations, we can use the corresponding version of an iterative gradient method; this methods is often indeed very successful; e.g., in [55], we successfully used it in radioastronomy.

Optimizing set functions is necessary. In some practical problems, alternatives are described by *sets*. In this case, the objective function $f(a)$ is defined on a class of sets. Such functions are called *set functions*. Thus, in these problems, we must optimize set functions.

There are several important classes of problems in which alternatives are sets:

- In *design*, we must find the best *shape* of the designed object: e.g., the aerodynamically best shape of a plane, of a car, or of a Space Shuttle.

- In *image processing*, we must extract, from the observed noisy data, the best image to fit the observations. Often, we do not have enough information to reconstruct the grey levels of different image pixels; in this case, we are interested in reconstructing a *black-and-white* image. A black-and-white image is, in essence, a *set*: it can be uniquely described if we describe the set of all black pixels.

- Another important class of practical optimization problems in which sets naturally appear is problems of *territorial division.* These problems range from big political problems, in which a disputed territory has to be divided between contesting countries to everyday economic problems, in which, e.g., a large tract of agricultural land, with varying properties at different points, has to be divided into areas of different usage: some pieces allocated for different crops, some for cattle pastures, some should not be used, etc. For each of the usages, we want to describe the corresponding *set* of all points reserved for this particular usage.

  In territorial division problems, there are usually several sides with conflicting interests, and therefore, we must use the theory of such conflict situations, usually known as *game theory* [35, 46], to select an appropriate objective function.

Similarly to all other cases, an important particular case of set function optimization is when the objective function $f(a)$ describes the posterior probability (likelihood) of an alternative a, i.e., when we use the *maximum likelihood method* to choose a set based on a given probability measure on the class of all sets (i.e., on a *random* set); for examples and detailed exposition, see, e.g., [23] and references therein.

Optimizing set functions is difficult. Anyone who has tried to optimize a set function knows that it is often a very computationally difficult problem. There are two reasons for this difficulty:

- First, there is a *general* reason that is equally applicable both to sets and to functions: The more alternatives we have, the more complicated the corresponding optimization problem. Therefore, problems in which alternatives are functions or sets are, in general, much more difficult to solve than problems in which alternatives can be characterized by finitely many parameters.

- Second, there is a *specific* reason which makes sets more difficult alternatives than functions: Sets can be described by their *characteristics functions*, and so optimizing over sets is the same as optimizing over functions with an additional constraint that a function cannot take arbitrary real values, only values from the set $\{0, 1\}$. At first glance, this constraint restricts the set of possible alternatives and thus, it should make the optimization problem easier, but in reality, it is much easier to optimize over continuously changing variables than over variables with a discrete range like $\{0, 1\}$.

  This relative complexity can be traced to the simplest case when we only consider subsets of a finite universal set. In this case, a characteristic function x is simply a finite sequence of binary numbers $x_1, \ldots, x_n \in \{0, 1\}$, where n is the total number of elements in the universal set. Even for linear objective functions $f(x) = \sum s_i \cdot x_i$, checking whether 0 is a possible value of this function is an NP-hard problem [22], while for *continuous* variables (e.g., variables running over intervals $[\underline{x}_i, \overline{x}_i]$), it is computationally easy to check whether 0 is a possible value of $f(x)$.

What we are planning to do. Since optimization of set functions is an important practical problem, it is desirable to be able to solve it. In this paper, we will start with a brief mention of the existing methods of solving this problem, then we will explain why these methods are not always applicable and finally, present a new differential formalism that covers previously un-covered practical problems.

In the final section, we will discuss related results and open problems.

2 The Existing Methods of Optimizing Set Functions: A Brief Mention

The existing methods. The existing methods of optimizing set functions are mainly developed for the case when the optimized set has some smoothness, i.e., in practical terms, when we are optimizing a *shape* that is described by a smooth (or at least piece-wise smooth) surface.

For this case, useful differential formalisms have been developed. These formalisms use the fact that small (smooth) deviations from the optimal shape cannot increase the value of the objective function to deduce partial differential equations that describe the optimal shapes; see, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 39, 40, 41, 42, 43, 44, 57].

These papers use different approaches to "translate" set optimization into function optimization: e.g., one possibility it to replace a set a by a function $d(x, a)$ that describes the distance from each point x to this set; a more appropriate approach [7] is to use the *oriented distance*, i.e., $d(x, a)$ when $x \notin a$ and $d(x, Ca)$ when $x \in a$ (where Ca denotes a complement to the set a).

Limitations of the existing methods. The above methods help when we are selecting a smooth (or piece-wise smooth) shape. However, in many practical optimization problems, sets can be arbitrarily complicated, and there is no reason to assume that the set is in any way smooth:

- In *image processing*, i.e., in processing *astronomical* images, the desired image is not only not smooth, it is often rather like a fractal set.

- In *territory division*, e.g., in agricultural applications, it often makes sense to, say, plant apple seeds in several disconnected small areas where apples would grow the best; the resulting sets are not smooth, and not connected.

For such non-smooth problems, the existing differential formalisms are not directly applicable, and therefore, new differential formalisms are needed.

3 A New Differential Formalism for Set Functions

3.1 Motivations

The classical result that we are trying to emulate. The use of derivatives in standard optimization is based on the well-known result that if a function $f(x)$ attains its maximum (or minimum) at a certain point x, then its derivative $f'(x)$ at this point x is equal to 0. Thus, to find all maxima and minima, it is sufficient to find all the points x for which $f'(x) = 0$, and select the ones for which $f(x)$ is the largest or, correspondingly, the smallest.

Here, the derivative is usually defined as the limit

$$f'(x) = \lim_{h \to 0, h \neq 0} \frac{f(x+h) - f(x)}{h}. \tag{1}$$

Practice-motivated reformulation of the classical result. In many practical situations, we have an *explicit* analytical expression for $f(x)$ and therefore, we can get an analytical expression for the derivative $f'(x)$.

However, in many other practical situations, the maximized function is defined only *implicitly* (e.g., via a partial differential equation), and it is therefore difficult to find an explicit analytical expression for its derivative. In such situations, we can use *numerical methods* to estimate the desired derivative $f'(x)$.

Definition (1) prompts the natural numerical method for computing the derivative:

- In (1), the derivative is defined as a limit of the ratio $(f(x+h) - f(x))/h$ when h tends to 0.
- This means that the smaller h, the closer this ratio will be to the desired derivative.
- So, to numerically estimate the derivative, we can take the value of this ratio for some small h.

The resulting numerical method often gives a good approximation to the derivative, but this method is definitely not the best, because for smooth functions, its accuracy is $O(h)$, while there exist alternative methods with a much better accuracy $O(h^2)$. To achieve this better accuracy, we can, e.g., take the ratio $(f(x+h) - f(x-h))/(2h)$. With this practical application in mind, it is reasonable to re-define derivative as the limit of the new ratios:

$$f'(x) = \lim_{h \to 0, h \neq 0} \frac{f(x+h) - f(x-h)}{2h}. \tag{2}$$

- From the purely *theoretical* viewpoint, the new definition is as good as the old one: Namely, one can easily see that if a function $f(x)$ has a derivative in the usual sense, then (2) leads to the same value of the derivative.

  The reader should be cautioned that the inverse is not necessarily true: it is possible for a function to have a derivative in the sense of definition (2), but not in the sense of definition (1).

- From the above described *practical* viewpoint, however, this new definition is better, because for this new definition, the naturally emerging numerical method is the (asymptotically optimal) $O(h^2)$ one.

Since the ultimate goal of our generalization is to be able to solve practical problems, we will try to design an analogue of this practice-oriented definition (2).

Towards the definition of the set derivative. How can we generalize the definition (2) to *set* functions $f(x)$?

For such functions, x is a set, and h is, naturally, also a set. How can reformulate the ratio from (2) in terms of sets?

- A natural analogue of $x+h$ is the *union* $x \cup h$, and the natural analogue of the difference $x-h$ is the set difference $x-h$, so we have a natural analogue of the *numerator*: $f(x \cup h) - f(x-h)$.

- The analogue of the *denominator* is also pretty much straightforward: of course, we cannot divide by a set, but we can take into consideration that for numbers, $2h$ is nothing else but the length (1D Lebesgue measure) of the interval $[-h, h]$. So, for an arbitrary set h, it is natural to use its Lebesgue measure $\mu_0(h)$ as the denominator.

Correspondingly, the condition $h \neq 0$, which made the ratio meaningful by preventing division by 0, can now be replaced by the similar condition $\mu_0(h) \neq 0$.

The only remaining problem is how to define the limit. In this definition, it is natural to use the same analogy between intervals and sets as we used to define the appropriate analogue of the denominator.

- In 1D case, we have defined the limit in which the interval $[-h, h]$ tends to a point (namely, to $\{0\}$).

- Therefore, it is reasonable to define a limit for set functions as the limit in which the set h tends to a one-point set, i.e., to a set $\{t\}$ for some point t.

To define such a limit, we need a topology on the set of all sets.

A natural topology on this set of all sets is given by *Hausdorff metric* $d_H(a, b)$, which is defined, for every two sets a and b, as the smallest real number r for which a is contained in the r-neighborhood of b, and b is contained in the r-neighborhood of a: $a \subseteq B_r(b)$, $b \subseteq B_r(a)$.

An r-neighborhood of a set a can be defined, e.g., as the set of all points that are r-close to some points from a, i.e., as $B_r(a) = \{t \mid d(t,s) \le r$ for some $s \in a\}$.

For the topology defined by the Hausdorff metric, $a \to b$ means that $d_H(a,b) \to 0$.

We are interested in the case when one of the sets is a one-point set $\{t\}$. In this case, the expression for Hausdorff distance $d_H(h,\{t\})$ can be simplified: this distance is the smallest radius r of the closed ball $B_r(t) = \{s \mid d(s,t) \le r\}$ that contains h.

Thus, we arrive at the following definition:

3.2 Definitions

We want the set function to be defined for bounded open and closed subsets of R^n, and also for finite unions and intersections of such sets:

Definition 1. *Let $n \ge 1$ be an integer.*

- *Let $\mathcal{A}$ be a class of sets that contains all bounded open and closed subsets $x \subseteq R^n$ and that is closed under union, intersection, and set difference. This class of sets will be called an algebra.*

- *By a set function, we will mean a real-valued function that is defined for all sets from the class $\mathcal{A}$.*

Definition 2. *Let $f(x)$ be a set function defined on an algebra $\mathcal{A}$, let $x \in \mathcal{A}$ be a set, and let $t \in R^n$ be a point. By the derivative of the function $f(x)$ at a set x at a point t we mean the limit*

$$\frac{df}{dx}(t) = \lim_{h \to \{t\}, \mu_0(h) \ne 0} \frac{f(x \cup h) - f(x - h)}{\mu_0(h)}$$

(where the limit $h \to \{t\}$ is in the sense of Hausdorff metric).

Comments.

- For *additive* set functions (i.e., for set functions that are themselves measures), this definition turns, basically, into a standard Radon-Nikodym derivative, see, e.g., [24, 26, 54] (see also [58]).

- For set functions that are not necessarily additive, similar definitions were proposed in [52, 59]

- Standard formulas of calculus can be used to describe the derivative of the sum, product, ratio, etc., of two differentiable set functions.

Definition 3.

- *We say that set function $f(x)$ defined on an algebra $\mathcal{A}$ is differentiable if for every set $x \in \mathcal{A}$ and for every point $t \in R^n$, it has a derivative.*
- *We say that a differentiable set function $f(x)$ is continuously differentiable if for every set x, the function*

$$\frac{df}{dx}(t)$$

is a continuous function of t.

3.3 Notations

In the following text, we will use the following notations. For every set a:

- Ca will denote the *complement* of the set a;
- $\mathrm{Int}(a)$ will denote the *interior* of the set a;
- $\bar{a}$ will denote the *closure* of the set a; and
- ∂a will denote the *boundary* of the set a.

3.4 Results

When the set function $f(x)$ is continuously differentiable, we get the direct analogue of the classical relationship between the maxima and minima of a function and the derivative being equal to 0:

Theorem 1. *If a continuously differentiable set function $f(x)$ attains its maximum or its minimum at some set x, then*

$$\frac{df}{dx}(t) = 0$$

for all points $t \in \partial(\mathrm{Int}(x)) \cup \partial(\mathrm{Int}(Cx))$.

In plain words, the derivative of the set function is equal to 0 for all points t from the "border" of x.

When the set function is simply differentiable, but not necessarily continuously differentiable, then we get inequalities instead of equalities:

Theorem 2. *If a differentiable set function $f(x)$ attains its maximum at some set x, then:*

- *for all $t \in \mathrm{Int}(x)$, we have*
$$\frac{df}{dx}(t) \geq 0;$$
- *for all $t \in \mathrm{Int}(Cx)$, we have*
$$\frac{df}{dx}(t) \leq 0.$$

Theorem 3. *If a differentiable set function $f(x)$ attains its minimum at some set x, then:*

- *for all $t \in \mathrm{Int}(x)$, we have*
$$\frac{df}{dx}(t) \leq 0;$$
- *for all $t \in \mathrm{Int}(Cx)$, we have*
$$\frac{df}{dx}(t) \geq 0.$$

3.5 Proofs

Proving Theorem 2: general idea. Let us start by proving Theorem 2.

Let $f(x)$ be a differentiable set function that attains its maximum at a set x, and let $t \in R^n$. According to the definition, the derivative

$$\frac{df}{dx}(t)$$

is equal to the limit of the ratio $(f(x \cup h) - f(x - h))/\mu_0(h)$ when $h \to \{t\}$ and $\mu_0(h) \neq 0$. Let us take, as h's, a sequence of closed balls $h_r = B_r(t)$ with a center in t for which $r \to 0$. Each of these sets h_r is a closed bounded set and therefore, $h_r \in \mathcal{A}$, $x \cup h_r \in \mathcal{A}$, and $x - h_r \in \mathcal{A}$.

According to the above formula for the Hausdorff distance $d_H(x, \{t\})$, the condition $r \to 0$ implies that $h_r = B_r(t) \to \{t\}$; also, we know that $\mu_0(B_r(t)) \neq 0$ for each $r \neq 0$. Therefore, for this sequence of balls, we get

$$\frac{f(x \cup h_r) - f(x - h_r)}{\mu_0(h_r)} \to \frac{df}{dx}(t). \tag{3}$$

Proving Theorem 2: first part. By definition of an interior point, if $t \in \mathrm{Int}(x)$, then we have $B_r(t) \subseteq x$ for all sufficiently small $r > 0$. Therefore, for such r, we have $x \cup h_r = x$, and the formula (3) turns into

$$\frac{f(x) - f(x - h_r)}{\mu_0(h_r)} \to \frac{df}{dx}(t). \tag{4}$$

Since the function $f(x)$ attains its maximum for the set x, we have $f(x) \geq f(x')$ for an arbitrary set $x' \in \mathcal{A}$, in particular, for $x' = x - h_r$. Therefore, the ratio in the left-hand side of the formula (4) is non-negative.

Hence, the limit

$$\frac{df}{dx}(t)$$

of these non-negative ratios should also be non-negative. The first part of Theorem 2 is thus proven.

Proving Theorem 2: second part. Similarly, if $t \in \mathrm{Int}(Cx)$, then we have $B_r(t) \subseteq Cx$ for all sufficiently small $r > 0$. Therefore, for such r, $x - h_r = x$, and (3) becomes

$$\frac{f(x \cup h_r) - f(x)}{\mu_0(h_r)} \to \frac{df}{dx}(t). \tag{5}$$

Since the function $f(x)$ attains its maximum for the set x, we have $f(x) \geq f(x')$ for an arbitrary set $x' \in \mathcal{A}$, in particular, for $x' = x \cup h_r$. Therefore, the ratio in the left-hand side of the formula (4) is non-positive and hence, its limit

$$\frac{df}{dx}(t)$$

should also be non-positive. The second part of Theorem 2 is thus also proven, and so is the theorem.

Proving Theorem 3. Theorem 3 can be proven along the same lines as Theorem 2.

Proving Theorem 1. Let us now show how Theorem 1 follows from Theorems 2 and 3. Indeed, let $f(x)$ be a continuously differentiable set function, and let x be a set on which this set function attains its maximum.

Let t be a point from the intersection $\partial(\mathrm{Int}(x)) \cup \partial(\mathrm{Int}(Cx))$. Since t belongs to the intersection of these two boundaries, it belongs to each of them.

From $t \in \partial(\mathrm{Int}(x))$, it follows that t is a limit of a sequence of points $t_m \in \mathrm{Int}(x)$. For each point t_m, according to Theorem 2, we have

$$\frac{df}{dx}(t_m) \geq 0.$$

By definition of continuous differentiability, the derivative $df/dx(t)$ is an everywhere defined continuous function of t. Thus, from $t_m \to t$, it follows that

$$\frac{df}{dx}(t) = \lim \frac{df}{dx}(t_m).$$

From

$$\frac{df}{dx}(t_m) \geq 0,$$

we conclude that in the limit,

$$\frac{df}{dx}(t) \geq 0.$$

Similarly, from $t \in \partial(\mathrm{Int}(Cx))$, it follows that t is a limit of a sequence $t_m \in \mathrm{Int}(Cx)$. For each point t_m, according to Theorem 2, we have

$$\frac{df}{dx}(t_m) \leq 0.$$

Thus, from $t_m \rightarrow t$, it follows that

$$\frac{df}{dx}(t) = \lim \frac{df}{dx}(t_m).$$

From

$$\frac{df}{dx}(t_m) \leq 0,$$

we conclude that in the limit,

$$\frac{df}{dx}(t) \leq 0.$$

So,

$$\frac{df}{dx}(t) \geq 0 \text{ and } \frac{df}{dx}(t) \leq 0,$$

i.e.,

$$\frac{df}{dx}(t) = 0.$$

The theorem is proven.

Comments.

- In the following two sections, we will describe two applications of this new formalism.
- We formulated our theorems for set functions defined on sets from R^n, because this is what we need for our applications (see below). However, as one can see from the proof, these results are valid for an arbitrary metric space X with a measure μ_0 that is defined on all bounded open and closed sets, and that is positive on all non-degenerate balls (i.e., $\mu_0(B_r(t)) > 0$ for all $t \in X$ and $r > 0$).

4 First Application of the New Formalism: An Example from Game Theory

4.1 Informal Description of the Problem

Let us consider the problem of dividing a piece of territory a between two parties (e.g., the problem of dividing a disputed territory between the two countries).

Comment. This example and the solution given below were first described in [27, 28]

4.2 Towards Formalization of the Problem

This is a problem of finding a set. Each possible division can be uniquely characterized by the set x of all the points that are allocated to the first party.

Indeed, if we know the set x of all the points allocated to the first party, then we can uniquely determined what the second party gets: all the remaining points, i.e., all the points from the complement set $a - x$.

What is the criterion for selecting this set x?

Towards a criterion for selecting a set: describing the preferences of each party. To describe a reasonable division, we must know, for each possible division of the set a into sets x and $a - x$, how each party evaluates the result of this division.

As a reasonable numerical evaluation of this result for each party, we can take, e.g., the amount of money that this party would be willing to pay for this piece of territory. We will call this amount of money the *usefulness* of the corresponding piece of territory to this party; we will denote these usefulnesses, correspondingly, by $u_1(x)$ and $u_2(a - x)$.

- In reality, we cannot infinitely divide the territory, so, in practice, there are *finitely many* meaningful pieces (points) into which the territory can be divided. With this in mind, the usefulness $u_1(x)$ of the portion x for the first party can be estimated as a sum of usefulnesses of all the points from this set (this is absolutely true, e.g., when we talk about mineral extraction or about the agricultural use of this territory), i.e., as the sum $u_1(x) = \sum U_1(\vec{t})$, where the sum is taken over all points $\vec{t} \in x$, and $U_1(\vec{t})$ denotes the usefulness of the point $\vec{t}$ for the first party.
- The *problem* with this finitely-many-points approach is that there are many such points, and the corresponding point-by-point computations can take a very long time.
- So, to prevent the long process of enumerating all possible divisions, it makes sense to use a *continuous approximation*, in which we assume that we can implement an arbitrary division of a plane into two sets. In this

approximation, the first party's portion is represented by an arbitrary closed set $x \subseteq a$. It is then reasonable to define $u_1(x)$ by using a natural continuous analogue of the finite sum, i.e., an integral. In other words, we will assume that $u_1(x) = \int_x U_1(\vec{t}\,)\,d\vec{t}$.

Similarly, the usefulness $u_2(a-x)$ of the piece of territory $a-x$ for the second party is represented as $u_2(a-x) = \int_{a-x} U_2(\vec{t}\,)\,d\vec{t}$. It is reasonable to assume that the functions $U_1(\vec{t}\,)$ and $U_2(\vec{t}\,)$ are continuous.

The objective function (set function to be optimized). To achieve a reasonable compromise between the interests of both parties, we will use a well-justified notion of *Nash's solution*, according to which we must choose an alternative for which the product of usefulnesses is the largest possible (see, e.g., [35, 46] for description and explanations). In our case, this product is equal to $u_1(x) \cdot u_2(a-x)$.

So, we arrive at the following formalization of the territory division problem:

4.3 Problem Reformulated in Precise Mathematical Terms

GIVEN: two non-negative continuous functions $U_1(\vec{t}\,)$ and $U_2(\vec{t}\,)$ and a bounded open set a,

FIND: the closed set $x \subseteq a$ for which

$$f(x) = u_1(x) \cdot u_2(a-x) \to \max,$$

where

$$u_1(x) = \int_x U_1(\vec{t}\,)\,d\vec{t} \text{ and } u_2(a-x) = \int_{a-x} U_2(\vec{t}\,)\,d\vec{t}.$$

4.4 Our Results Help to Solve This Problem

Our formalism is applicable. One can easily check that both functions $u_1(x) = \int_x U_1(\vec{t}\,)\,d\vec{t}$ and $u_2(x) = u_2(a) - \int_x U_2(\vec{t}\,)\,d\vec{t}$ are continuously differentiable, with derivatives correspondingly equal to

$$\frac{du_1}{dx}(\vec{t}\,) = U_1(\vec{t}\,) \text{ and } \frac{du_1}{dx}(\vec{t}\,) = -U_2(\vec{t}\,).$$

Therefore, their product, i.e., the set function $f(x)$, is also continuously differentiable, and its derivative is equal to

$$\frac{df}{dx}(\vec{t}\,) = U_1(\vec{t}\,) \cdot u_2(a-x) - U_2(\vec{t}\,) \cdot u_1(x).$$

Let us apply this new formalism. So, if this set function attains its maximum at some set x, then, according to Theorem 2, we can make the following two conclusions:

- $U_1(\vec{t}) \cdot u_2(a-x) - U_2(\vec{t}) \cdot u_1(x) \geq 0$ for all $\vec{t} \in \mathrm{Int}(x)$;
- $U_1(\vec{t}) \cdot u_2(a-x) - U_2(\vec{t}) \cdot u_1(x) \leq 0$ for all $\vec{t} \in \mathrm{Int}(Cx)$.

Let us simplify the resulting inequalities. To simplify both inequalities, we can:

- move the negative part to the right-hand side, and then
- divide both sides of the resulting inequality by $U_2(\vec{t}) \cdot u_2(a-x)$ (to separate the variables $\vec{t}$ and x).

After this transformation, the above conclusions take the following simplified form:

$$\frac{U_1(\vec{t})}{U_2(\vec{t})} \geq \alpha \text{ for all } \vec{t} \in \mathrm{Int}(x), \tag{6}$$

$$\frac{U_1(\vec{t})}{U_2(\vec{t})} \leq \alpha \text{ for all } \vec{t} \in \mathrm{Int}(Cx), \tag{7}$$

where we denoted

$$\alpha = \frac{u_1(x)}{u_2(a-x)}. \tag{8}$$

Let us be realistic. In the above formulation, we assume that x can be an *arbitrary* set. Of course, sets can be weird and impossible to practically implement: e.g., there is not way to realistically divide the interval $[0,1]$ into the set of all its rational points and the complement set of all its irrational points. Sol, let us be more realistic, and impose some conditions on the set x to make it more geometric and thus, more realistic in relation to the original problem (of territorial division). Namely, we will assume that:

- x is a closed set, and that
- x is a closure of its own interior.

(These two properties are true for sets bounded by normal smooth curves.) For such sets, we can further simplify the above conclusion.

Using the first realistic assumption. If we are looking for a *closed* set x, then $\mathrm{Int}(Cx) = Cx$, and (7) takes the form

$$\frac{U_1(\vec{t})}{U_2(\vec{t})} \leq \alpha \text{ for all } \vec{t} \in Cx. \tag{9}$$

Using the second realistic assumption. Similarly, if we are looking for a set x that is the closure of its interior, then, due to continuity of the functions $U_1(\vec{t})$ and $U_2(\vec{t})$, we can conclude that $U_1(\vec{t})/U_2(\vec{t}) \geq \alpha$ for all $\vec{t} \in x$.

Let us now combine the consequences of the two realistic assumptions. Thus:

- if $\vec{t} \in x$, we have $U_1(\vec{t})/U_2(\vec{t}) \geq \alpha$; and
- if $\vec{t} \notin x$, we have $U_1(\vec{t})/U_2(\vec{t}) \leq \alpha$.

Hence:

- If $U_1(\vec{t})/U_2(\vec{t}) > \alpha$, then we cannot have $\vec{t} \in x$ and therefore, we must have $\vec{t} \notin x$.
- Similarly, if $U_1(\vec{t})/U_2(\vec{t}) < \alpha$, then we cannot have $\vec{t} \in Cx$ and therefore, we must have $\vec{t} \in x$.

We (almost) got a solution. Thus, if we find the value α, then we (almost) have a description of the desired set x:

- this set must contain all the points $\vec{t}$ for which $U_1(\vec{t})/U_2(\vec{t}) > \alpha$; and
- this set cannot contain any points $\vec{t}$ for which $U_1(\vec{t})/U_2(\vec{t}) < \alpha$.

The only uncertainty is about the points $\vec{t}$ for which $U_1(\vec{t})/U_2(\vec{t}) = \alpha$. This uncertainty, however, does not matter much if we assume that for each α, the set of all points $\vec{t}$ for which $U_1(\vec{t})/U_2(\vec{t}) = \alpha$ has Lebesgue measure 0, and therefore, whether we add these points to x or to $a - x$ will not affect the usefulnesses $u_1(x)$ and $u_2(a - x)$.

> This assumption is indeed true for generic analytical functions $U_1(\vec{t})$ and $U_2(\vec{t})$.

Thus, we can simply define x as, e.g., the set of all the points $\vec{t}$ for which $U_1(\vec{t})/U_2(\vec{t}) \geq \alpha$:

$$x_\alpha = \{\vec{t} \,|\, \frac{U_1(\vec{t})}{U_2(\vec{t})} \geq \alpha\}.$$

The final step: finding the parameter α. If the above "equal to α" uncertainty is thus resolved, then the only remaining problem is how to find the value α. Substituting the above expression for x_α into the formula (8), we conclude that

$$\frac{u_1(x_\alpha)}{u_2(a - x_\alpha)} - \alpha = 0, \tag{10}$$

where

$$u_1(x_\alpha) = \int_{\{\vec{t}\,|\,U_1(\vec{t})/U_2(\vec{t}) \geq \alpha\}} U_1(\vec{t})\,d\vec{t}$$

and

$$u_2(a-x) = \int_{\{\vec{t} \mid U_1(\vec{t})/U_2(\vec{t})<\alpha\}} U_2(\vec{t})\,d\vec{t}.$$

When α increases, then:

- The set x_α become smaller.
- Hence, $u_1(x_\alpha)$ become smaller.
- Simultaneously, $a - x_\alpha$ becomes larger.
- Because of this, $u_2(a - x_\alpha)$ become larger.
- Therefore, the ratio $u_1(x_\alpha)/u_2(a - x_\alpha)$ decreases as α increases.
- Hence, the difference $u_1(x_\alpha)/u_2(a - x_\alpha) - \alpha$ also monotonically decreases with α.

Hence, we can use, e.g., bisection method to find the value α for which this difference becomes equal to 0, i.e., to (easily) solve the equation (10).

4.5 Comment 1: A More Complicated Utility Function

A more realistic formulation of the territorial division problem: informal explanation. In the above text, we measured usefulness by the equivalent money amount. A more realistic approach would be to assume that the usefulness (normally called *utility* in game theory) is actually a *monotonic function* of the money amount [35, 46].

Towards mathematical formalization of this more realistic formulation. In this case, the problem becomes slightly more complicated:

- instead of maximizing the product $u_1(x) \cdot u_2(a - x)$ of the two integrals,
- we have to maximize the product $f(x) = M_1(u_1(x)) \cdot M_2(u_2(a-x))$, where:
  - $u_1(x)$ and $u_2(a - x)$ are the same integrals as before, and
  - $M_1(z)$ and $M_2(z)$ are strictly increasing functions.

Comment. Since we want to use differentiation to find the maximum, we will assume that these functions $M_i(z)$ are differentiable.

Our formalism is applicable to this new problem as well. Differentiating the new expression with respect to x, we can also conclude that there exists a value α for which the desired set x satisfies the following properties:

- the set x contains all the points $\vec{t}$ for which $U_1(\vec{t})/U_2(\vec{t}) > \alpha$; and
- the set x does not contain any points $\vec{t}$ for which $U_1(\vec{t})/U_2(\vec{t}) < \alpha$.

Substituting the resulting expression for x_α into the corresponding formula for α, we get a slightly more complicated formula for determining α, but the monotonicity property still stands for this new formula, and thus, we can still use bisection to find the desired value α.

4.6 Comment 2: A More General Territorial Division Problem

Informal explanation of the new problem. It is natural to consider a more general territorial division problem, in which more than two parties want to divide the territory.

Mathematical formalization of the new problem. In the simplest formulation, we want to find the sets $x_1, \ldots, x_k$ such that:

- $x_i \cap x_j = \emptyset$ for $i \neq j$,
- $\cup x_i = a$, and
- $u_1(x_1) \cdot u_2(x_2) \cdot \ldots \cdot u_n(x_n) \to \max$, where $u_i(x_i) = \int_{x_i} U_i(\vec{t})\, d\vec{t}$ for some continuous functions $U_i(\vec{t})$.

For this new problem, we cannot use our formalism directly. This more general problem does not *directly* fall into our scheme, because:

- our theorems describe *unconditional* optimization of set functions, while
- in this new multi-party problem, we need to optimize a set function under the *conditions* ($\cup x_i = a$ and $x_i \cap x_j = \emptyset$).

However, we can still use our theorems *indirectly* and solve this problem.

Our main idea of solving this problem. Our main idea here is similar to the way how we can use differentiation of functions of *one* variable to optimize function of *several* variables: if we know that a function $f(x_1, \ldots, x_n)$ of several (numerical) variables attains its maximum for some tuple $(x_1, \ldots, x_n)$, then, in particular, we can fix the values of all these variables except for x_1, and conclude that the resulting function of a single variable should also attain its maximum. Therefore, for the optimal value x_1, the derivative of that resulting function of one variable is equal to 0. This derivative is exactly the partial derivative of f with respect to x_1. We can then repeat the same construction for all other variables.

Let us apply this idea. Let us apply a similar idea to the territorial division problem. Let us assume that $x_1, \ldots, x_k$ form the optimal division, i.e., the division for which the product

$$u_1(x_1) \cdot u_2(x_2) \cdot \ldots \cdot u_k(x_k)$$

takes the largest possible value. To find x_i, we will fix some of these sets x_i.

- For the function of several variables, we fixed the values of all the variables but *one*.

- For the territorial division problem, we cannot fix all the sets but one, because if we do that (i.e., if we describe what each party gets, except for one party), then the remaining set will be uniquely determined: the remaining party gets exactly all the points that nobody else got.

Since we cannot fix all the sets but *one*, we will do the next best thing: we will fix all the allocations but *two*.

Reducing the new problem (of dividing between more than two parties) to the old one (of division between two parties). For example, we can fix the sets $x_3, x_4, \ldots, x_k$, and change x_1 and x_2. Since all allocations $x_3, x_4, \ldots$, are fixed, we thus fix the set $a' = x_1 \cup x_2$ of not allocated points; if we select x_1, then we can uniquely determine x_2 as $x_2 = a' - x_1$. So, we do get a problem with a *single* set. What is the objective function for this problem?

Since $x_3, x_4, \ldots$ are fixed, the original objective function

$$u_1(x_1) \cdot u_2(x_2) \cdot u_3(x_3) \cdot \ldots \cdot u_k(x_k)$$

attains its largest possible value if and only the product of the first two terms is the largest possible: $u_1(x_1) \cdot u_2(x_2) \to \max$. So, we conclude that we must select $x_1 \subseteq a'$ from the condition that $u_1(x_1) \cdot u_2(a' - x_2) \to \max$.

Let us use the known solution of the old problem to solve the new one. This is exactly the problem that we have been solving before, and we know its solution: there exist some value α_{12} $(= u_1(x_1)/u_2(x_2))$, such that the borderline between the points allocated to the first party and the points allocated to the second party is where $U_1(\vec{t})/U_2(\vec{t}) = \alpha_{12}$. In other words:

- if $U_1(\vec{t})/U_2(\vec{t}) > u_1(x_1)/x_2(x_2)$ (i.e., if $U_1(\vec{t})/u_1(x_1) > U_2(\vec{t})/u_2(x_2)$), then $\vec{t}$ is allocated to x_1;

- if $U_1(\vec{t})/U_2(\vec{t}) < u_1(x_1)/x_2(x_2)$ (i.e., if $U_1(\vec{t})/u_1(x_1) < U_2(\vec{t})/u_2(x_2)$), then $\vec{t}$ is allocated to x_2.

Similarly, by considering other pairs of parties, we can conclude that for each i and j, the borderline between x_i and x_j lies where $u_i(\vec{t})/u_j(\vec{t}) = u_i(x_i)/u_j(x_j)$, i.e.:

- if $U_i(\vec{t})/u_i(x_i) > U_j(\vec{t})/u_j(x_j)$, then $\vec{t}$ is allocated to x_i;

- if $U_i(\vec{t})/u_i(x_i) < U_j(\vec{t})/u_j(x_j)$, then $\vec{t}$ is allocated to x_j.

In general, a point $\vec{t}$ is allocated to x_i if for every other set x_j, we have $U_i(\vec{t})/u_i(x_i) \geq U_j(\vec{t})/u_j(x_j)$, i.e., if the ratio $U_i(\vec{t})/u_i(x_i)$ is the largest of k ratios $U_1(\vec{t})/u_1(x_1), \ldots, U_k(\vec{t})/u_k(x_k)$.

Conclusion: solution to the new problem. If we denote $u_i(x_i)$ by α_i, then we can describe this solution as follows: every point $\vec{t}$ is allocated to the set x_i for which

$$\frac{U_i(\vec{t})}{\alpha_i} = \max_{j=1,\ldots,k} \frac{U_j(\vec{t})}{\alpha_j}. \tag{11}$$

The values α_i can be then determined from the condition that

$$u_i(x_i) = \int_{x_i} U_i(\vec{t})\, d\vec{t} = \alpha_i.$$

Thus, we get k equations with k unknowns $\alpha_1, \ldots, \alpha_k$ to determine these parameters α_i.

What if usefulness is a monotonic function of money values. A similar solution (11) holds if we assume that usefulness is a monotonic function of the money value, i.e., if we want to maximize the expression $M_1(u_1(x_1)) \cdot M_2(u_2(x_2)) \cdot \ldots \cdot M_k(u_k(x_k))$; the only difference is that the equations for α_i will be, in this case, more complicated.

4.7 Comment 3: What If the Functions $U_i(\vec{t})$ are not Necessarily Continuous, but only Lebesgue Measurable

In this case, the set function is differentiable for *almost all* $\vec{t}$, and similar results follows modulo set of measure 0.

5 Second Application of the New Formalism: Statistical Example – Excess Mass Method

5.1 Excess Mass Method: the Main Idea

In [25, 45], it was shown that for a probability measure on R^n with density $\rho(\vec{t})$, and for each level α, the α-level contour $x_\alpha = \{\vec{t} \,|\, \rho(\vec{t}) \geq \alpha\}$ can be alternatively defined as a set that maximizes the functional

$$\varepsilon(x) = \mu(x) - \alpha \cdot \mu_0(x),$$

where:

- $\mu(x) = \int_x \rho(\vec{t})\, d\vec{t}$ is the probability of $\vec{t}$ being in the set x, and
- $\mu_0(x)$ $(= \int_x d\vec{t})$ is the Lebesgue measure of this set x.

This functional $\varepsilon(x)$ is called the *excess mass*.

5.2 Excess Mass Method is Very Useful

The extremal representation of α-contours has lead to several efficient statistical algorithms; see, e.g., [25, 45, 47, 53].

5.3 The Basic Result of Excess Mass Method Follows from Our Theorems

Our formalism is applicable. Let us assume, for simplicity, that $\rho(\vec{t}\,)$ is a continuous function. Then, one can easily check that the excess mass set function $\varepsilon(x)$ is differentiable, and its derivative is equal to

$$\frac{d\varepsilon}{dx}(\vec{t}\,) = \rho(\vec{t}\,) - \alpha.$$

The result of applying our formalism. So, if this set function attains its maximum at some set x, then, according to Theorem 2, we can make two conclusions:

$$\frac{d\varepsilon}{dx}(\vec{t}\,) = \rho(\vec{t}\,) - \alpha \geq 0 \text{ for all } \vec{t} \in \mathrm{Int}(x); \tag{12}$$

$$\frac{d\varepsilon}{dx}(\vec{t}\,) = \rho(\vec{t}\,) - \alpha \leq 0 \text{ for all } \vec{t} \in \mathrm{Int}(Cx). \tag{13}$$

Looking for a realistic set x. Similar to the first application of our formalism – to game theory – we can restrict ourselves to *realistic* sets x, i.e., to sets x that are *closed* and that coincide with the closure of their interior.

If we are looking for a closed set x, then $\mathrm{Int}(Cx) = Cx$, and (13) takes the form

$$\frac{d\varepsilon}{dx}(\vec{t}\,) = \rho(\vec{t}\,) - \alpha \leq 0 \text{ for all } \vec{t} \in Cx. \tag{14}$$

Similarly, if we are looking for a set x that is the closure of its interior, then, due to continuity of the function $\rho(\vec{t}\,)$, we can conclude that $\rho(\vec{t}\,) - \alpha \geq 0$ for all $\vec{t} \in x$. Thus:

- if $\vec{t} \in x$, we have $\rho(\vec{t}\,) \geq \alpha$; and
- if $\vec{t} \notin x$, we have $\rho(\vec{t}\,) \leq \alpha$.

Hence:

- If $\rho(\vec{t}\,) > \alpha$, then we cannot have $\vec{t} \in x$ and therefore, we must have $\vec{t} \notin x$.
- Similarly, if $\rho(\vec{t}\,) < \alpha$, then we cannot have $\vec{t} \in Cx$ and therefore, we must have $\vec{t} \in x$.

Thus, we (almost) have a description of the desired set x:

- this set must contain all the points $\vec{t}$ for which $\rho(\vec{t}) > \alpha$; and
- this set cannot contain any points $\vec{t}$ for which $\rho(\vec{t}) < \alpha$.

Conclusion. Thus, we can conclude that the desired set x is *almost* equal to the α-contour $x_\alpha = \{\vec{t}\,|\,\rho(\vec{t}) \geq \alpha\}$, almost equal in the following sense:

- x does contain all the points for which $\rho(\vec{t})$ is strictly larger than α;
- x does not contain any points that are not in the α-contour (i.e., for which $\rho(\vec{t}) < \alpha$); but
- x may miss some of the points $\vec{t}$ for which $\rho(\vec{t})$ is exactly equal to α.

6 Further Results and Open Problems

6.1 Limitations of the Above Differential Formalism

Similarly to the standard differentiation, all this formalism provides is an equation that is a necessary condition for the optimum. In some cases, this equation directly leads to a solution, but in some other cases, this equation is, itself, difficult to solve.

6.2 Gradient-Type Iterative Methods

Iterative methods are needed. Since it is sometimes difficult to directly solve an equation obtained by equating the derivative to 0, it is desirable to design iterative gradient-type methods for optimizing set functions.

Example of a gradient-type method. For the case of a finite universal set U, such a method has been proposed by M. Zakharevich in [61] (see also [29, 30, 33]). We have already mentioned that for this case, an arbitrary set $x \subseteq U$ can be described as a sequence of binary variables $x_1, \ldots, x_n \in \{0, 1\}$ where x_i is equal to 1 or 0 depending on whether the i-th element of the universal set belongs to the set x or not. Since each variable x_i has only two possible values, as a "derivative" of a function $f(x_1, \ldots, x_n)$ relative to x_i, we take the difference between the value of this function for $x_1 = 1$ and $x_i = 0$, i.e., the difference $f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$.

By using this definition, we overcame the problem of defining the derivative, but we still need to overcome another problem become we get a working iterative method:

- the values of all the variables should always be 0 or 1, but
- with thus defined "gradient" Df_i, the new value $x_i' = x_i - \lambda \cdot Df_i$ is not necessarily equal to 0 or 1.

To overcome this second problem (and thus get an iterative method), we allow arbitrary real numbers on the intermediate stages of the iterative process, and only after the iterations stop, we transform these values back to the set $\{0, 1\}$: e.g., we can use a threshold t and set x_i to 1 if the corresponding real value was $\geq t$, and set x_i to 0 otherwise.

This method is reasonably successful. The resulting algorithm turned out to be reasonably successful for solving several known NP-hard problems, including the so-called *propositional satisfiability* problem (historically the first NP-hard problem).

This gradient-type method turns out to be related to neural, fuzzy, etc. Interestingly, the resulting algorithm coincides with a heuristic method proposed by S. Maslov in the 1980s [36, 37, 38] as a way of simulating biological neurons (see also [29, 30, 31, 33, 56]).

This method was originally proposed based on the idea of simulating biological *neurons*, but it later turned out that exactly these same formulas follow:

- from *fuzzy logic* heuristics,
- from the ideas of *chemical computing* (i.e., simulating chemical reactions),
- from heuristics of *numerical optimization* approach,
- from the ideas of *freedom of choice*, etc.

(for a survey and latest results, see [30, 33]).

6.3 What If Iterative Methods Do Not Work Either?

In many practical cases, either the differential approach work directly, or at least derivative-based iterative methods work. However, in other practical cases, none of these methods work. How can we then solve the problem of optimizing a given set function?

6.4 Considering Finite-Parametric Families of Sets

The main reason why set optimization problems are so difficult is, as we have mentioned, the fact that for these problems, the class of all alternatives (sets) is enormous (infinite-dimensional). So, if we cannot solve this optimization problem, it is natural to restrict ourselves to a *finite-dimensional* (finite-parametric) family of sets. With this restriction, the objective function becomes a function of finitely many parameters, and thus, we can use known optimization techniques, for solving the corresponding optimization problem.

6.5 A Related Problem – Choosing the Optimal Family of Sets: Successes and Open Problems

The problem.

- At first glance, it may seem that with the idea of restricting ourselves to a finite-parametric family, we overcame the difficulties of optimizing set functions.
- However, if we seriously follow this approach, then the natural next question is:

  Which of the many finite-parametric families should we choose?

In other words, instead of the problem of finding the most appropriate *set a* (i.e., a set for which the objective function takes the largest possible value), we now face an even more complicated problem: to find the most appropriate *family of sets* (i.e., family for which the corresponding objective function has the best possible value).

Successes: in some cases, this problem is solvable. There are cases when this complicated problem can be solved: e.g., when we know that the corresponding problem is *invariant* with respect to natural symmetries. In this case, if the optimum is achieved for a single family, this family must also be invariant with respect to the same symmetries. This invariance often enables us to determine the optimal family:

- In processing *astronomical images*, this approach not only leads to the explicit expressions for optimal families, but it also explains the families of image sets that are traditional used in astronomy, such as spirals, etc. [19, 20, 21, 48].
- In *numerical mathematics*, this approach explains, e.g., the successful use of ellipsoids in modern numerical methods [19]: under reasonable symmetry conditions, ellipsoids turn out to be the best family for representing uncertainty.
- In *non-destructive testing* of aerospace objects, we get the optimal families of sets where we should place sensors [48, 49].

Remaining problems. These *successes* do not mean that this symmetry approach always works: in some practical problems, we do not have reasonable symmetries, and therefore, this approach is not applicable. It is thus desirable to extend the above differential formalism for finding the optimal *set* so that the extended formalism will provide us with optimal *families* of sets. Developing such a generalized formalism is, therefore, an important *open problem*.

Acknowledgments. This work was partially supported by NASA under cooperative agreement NCCW-0089, by NSF under 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.

The authors are very thankful to Ilya Molchanov for his encouragement and to Jean-Pierre Aubin for important references and advise.

References

[1] J.-P. Aubin, *Mutational and Morphological Analysis*, to appear.

[2] M. C. Delfour and J.-P. Zolesio, "Shape sensitivity analysis via a penalization method",*Annali di Matematica pura ed applicata (IV)*, 1988, Vol. 151, pp. 179–212.

[3] M. C. Delfour and J.-P. Zolesio, "Shape sensitivity' analysis via min max differentiability", *SIAM J. Control and Optimization*, 1988, Vol. 26, No. 4, pp. 834–862.

[4] M. C. Delfour and J.-P. Zolesio, "Analyse des problèmes de forme par la dérivation des minimax", In: H. Attouch, J.-P. Aubin, F. Clarke, and I. Ekeland, Eds., *Analyse Non Linéaire*, Gauthier-Villars & C.R.M. Université de Montréal, 1989, pp. 211–228.

[5] M. C. Delfour and J.-P. Zolesio, "Sensitivity method and Lagrangian formulation for the computation of the shape Hessian", *SIAM J. Control and Optimization*, 1991, Vol. 24, No. 5, pp. 835–883.

[6] M. C. Delfour and J.-P. Zolesio, "Functional analytic methods in shape analysis", *Proc. IFIP Workshop on Boundary Control and Boundary Variation*, Sophia Antipolis, June 1992.

[7] M. C. Delfour and J.-P. Zolesio, "Shape analysis via oriented distance functions", *Journal of Functional Analysis*, 1994, Vol. 123, pp. 129–201.

[8] M. C. Delfour and J.-P. Zolesio, "A boundary differential equation for thin shells", *J. Diff. Equations*, 1995, Vol. 119, pp. 426–449.

[9] M. C. Delfour and J.-P. Zolesio, "Tangential differential equations for dynamical thin/shallow shells", *J. Diff. Equations*, 1996, Vol. 128, pp. 125–167.

[10] M. C. Delfour and J.-P. Zolesio, "Structure of shape derivatives for domains nonsmooth domains", *J. Functional Analysis*, to appear.

[11] M. C. Delfour and J.-P. Zolesio, *Intrinsic differential geometry and theory of thin shells*, Scuola Normale di Pisa, to appear.

[12] L. Doyen, *Optimisation, évolution et contrôle de formes*, Thèse de l'Université de Paris-Dauphine, 1993.

[13] L. Doyen, "Filippov and invariance theorems for mutational inclusions for tubes", Set-Valued Analysis, 1993, Vol. 1, pp. 289–303.

[14] L. Doyen, "Inverse function theorems and shape optimization", *SIAM Journal on Control and optimization*, 1994, Vol. 32, No. 6, pp. 1621–1642.

[15] L. Doyen, "Mutational equations for tubes and vision based control", *Journal of Mathematical Imaging and Vision*, to appear.

[16] L. Doyen, "Shape Lyapunov functions", *J. Math. Anal. Appl.*, to appear.

[17] L. Doyen, "Lagrangian multipliers in Shape Optimization under shape constraints", *Proceedings of IFIP Workshop Boundary Control*, Centre de Mathématiques Appliquées, Ecole des Mines, Sophia-Antipolis, to appear.

[18] R. P. Feynman, R. B. Leighton, and M. L. Sands, *The Feynman Lectures On Physics*, Addison-Wesley, Redwood City, CA, 1989.

[19] 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.

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

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

[22] M. Garey and D. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*, Freeman, San Francisco, 1979.

[23] J. Goutsias, R. P. S. Mahler, and H. T. Nguyen (eds.), *Random Sets: Theory and Applications*, Springer-Verlag, N.Y., 1997.

[24] H. Hahn and A. Rosenthal, *Set functions*, The University of New Mexico Press, Albuquerque, NM, 1948.

[25] J. A. Hartigan, "Estimation of a convex density contour in two dimensions", *Journal of the American Statistical Association*, 1987, Vol. 82, No. 397, pp. 267–270.

[26] C. A. Hayes and C. Y. Pauc, *Derivation and martingales*, Springer-Verlag, NY, 1970.

[27] O. M. Kosheleva and V. Kreinovich, *On the optimal territory division*, Technical Report, Leningrad Technological Center "Informatika", Leningrad, 1989 (in Russian).

[28] V. Kreinovich, "Nash's solution for the territory division problem", *Proceedings of the 3rd USSR National Conference on Game Theory*, Odessa, 1974, pp. 117–118 (in Russian).

[29] V. Kreinovich, "Semantics of S. Yu. Maslov's iterative method," In: *Problems of Cybernetics*, Moscow, 1987, Vol. 131, pp. 30–62 (in Russian); English translation in: V. Kreinovich and G. Mints (eds.), *Problems of reducing the exhaustive search*, American Mathematical Society, Providence, RI, 1997, pp. 23–52.

[30] V. Kreinovich, "S. Maslov's Iterative Method: 15 Years Later (Freedom of Choice, Neural Networks, Numerical Optimization, Uncertainty Reasoning, and Chemical Computing)", In: V. Kreinovich and G. Mints, eds. *Problems of reducing the exhaustive search*, American Mathematical Society, Providence, RI, 1997, pp. 175–189.

[31] V. Kreinovich and L. O. Fuentes, "Simulation of chemical kinetics - a promising approach to inference engines," in: J. Liebowitz (ed.), *Proceedings of the World Congress on Expert Systems, Orlando, Florida, 1991*, Pergamon Press, N.Y., Vol. 3, pp. 1510–1517.

[32] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997 (to appear).

[33] V. Kreinovich and G. Mints, eds., *Problems of reducing the exhaustive search.* American Mathematical Society, Providence, RI, 1997.

[34] L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields*, Butterworth-Heinemanm, Oxford, UK, 1987.

[35] D. R. Luce and H. Raiffa, *Games and Decisions, Introduction and critical survey*, John Wiley & Sons, Inc., New York, 1957.

[36] S. Yu. Maslov, "Iterative methods in intractable problems as a model of intuitive methods", *Abstracts of the 9th All-Union Symposium on Cybernetics*, 1981, pp. 52–56 (in Russian).

[37] S. Yu. Maslov, "Asymmetry of cognitive mechanisms and its implications", *Semiotika i Informatika*, 1983, Vol. 20, pp. 3–31 (in Russian).

[38] S. Yu. Maslov, *Theory of deductive systems and its applications*, MIT Press, Cambridge, MA, 1987.

[39] J. Mattioli, "Differential Inclusions for Mathematical Morphology", In *SPIE: Application of Artificial Neural Network*, San Diego, July 11–16, 1993.

[40] J. Mattioli, *Problèmes inverses et relations différentielles en morphologie mathématique*, Thèse de l' Université de Paris-Dauphine, 1993.

[41] J. Mattioli, "Differential Relations of Morphological Operators", In *Proceedings of the Conference on Mathematical Morphology and its applications to Signal Processing*, Barcelona, Spain, May 12–14, 1993.

[42] J. Mattioli, "Relations Différentielles d'Opérations de la Morphologie Mathématique", *Revue Technique Thomson*, 1993, Vol. 25, No. 1, pp. 281–303

[43] J. Mattioli, "Relations différentielles d'opérations de la morphologie mathématique", *C.R. Acad. Sci. Paris*, 1993, Vol. 316, pp. 879–884.

[44] J. Mattioli, "Minkowski operations and vector spaces", *Set-Valued Analysis*, to appear.

[45] D. W. Müller and G. Sawitzki, *Using excess mass estimates to investigate the modality of a distribution*, University of Heidelberg, Preprint No. 398, SFB 123.

[46] R. B. Myerson, *Game theory. Analysis of conflict*, Harvard University Press, Cambridge, MA, 1991.

[47] D. Nolan, "The excess-mass ellipsoid", *Journal of Multivariate Analysis*, 1991, Vol. 39, pp. 348–371.

[48] R. Osegueda, C. Ferregut, M. J. George, J. M. Gutierrez, and V. Kreinovich, "Non-Equilibrium Thermodynamics Explains Semiotic Shapes: Applications to Astronomy and to Non-Destructive Testing of Aerospace Systems", *Proceedings of the International Conference on Intelligent Systems and Semiotics (ISAS'97)*, National Institute of Standards and Technology Publ., Gaithersburg, MD, 1997, pp. 378–382.

[49] R. Osegueda, C. Ferregut, M. J. George, J. M. Gutierrez, and V. Kreinovich, "Computational geometry and artificial neural networks: a hybrid approach to optimal sensor placement for aerospace NDE", *Proceedings of the International Workshop on Intelligent NDE Sciences for Aging and Futuristic Aircraft*, El Paso, TX, September 30–October 2, 1997 (to appear).

[50] C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, San Diego, 1994.

[51] P. M. Pardalos, *Complexity in Numerical Optimization*, World Scientific, Singapore, 1993.

[52] C. Petrescu-Prodan, "On the derivative of a set function", *Studii si Cercetari Matematica*, 1992, Vol. 4,, No. 3, pp. 253–259.

[53] W. Polonik, "Measuring mass concentrations and estimating density contour clusters – an excess mass approach", *The Annals of Statistics*, 1995, Vol. 23, No. 3, pp. 855–881.

[54] G. E. Shilov and B. L. Gurevich, *Integral, measure, and derivative: a unified approach*, Dover, N.Y., 1977.

[55] V. I. Sinyansky, A. A. Stotsky, and V. Ya. Kreinovich, "Reconstructing the field phase on the radio telescope aperture from the amplitude values on the focal point and on the aperture", In: *Radioastronomical Hardware. Antennas and Methods. Proceedings of the XIV USSR National Radioastronomical Conference, October 28–30, 1982*, Armenian Academy of Sciences, Erevan, pp. 246–247 (in Russian).

[56] O. Sirisaengtaksin, L. O. Fuentes, and V. Kreinovich, "Non-traditional neural networks that solve one more intractable problem: propositional satisfiability", *Proceedings of the First International Conference on Neural, Parallel, and Scientific Computations*, Atlanta, GA, May 28–31, 1995, Vol. 1, pp. 427–430.

[57] J. Sokolowski and J.-P. Zolesio, *Introduction to shape optimization*, Springer-Verlag, 1992.

[58] Á. Száz, "The fundamental theorem of calculus in an abstract setting", *Tatra Mountains Math. Publ.*, 1993, Vol. 2, pp. 167–174.

[59] K. Tanaka and Y. Maruyama, "The multiobjective optimization problem of set functions", *Journal of Information and Optimization Sciences*, 1984, Vol. 5, pp. 293–306.

[60] S. A. Vavasis, *Nonlinear optimization: complexity issues*, Oxford University Press, N.Y., 1991.

[61] M. I. Zakharevich, "Ergodic properties of Maslov's iterative method", In: *Problems of Cybernetics*, Moscow, 1987, Vol. 131, pp. 63–76 (in Russian); English translation in: V. Kreinovich and G. Mints (eds.), *Problems of reducing the exhaustive search*, American Mathematical Society, Providence, RI, 1997, pp. 53–64.