# Astrogeometry, Error Estimation, and Other Applications of Set-Valued Analysis

Andrei Finkelstein, Olga Kosheleva, and Vladik Kreinovich*

**Abstract**

In many real-life application problems, we are interested in *numbers*, namely, in the numerical values of the physical quantities. There are, however, at least two classes of problems, in which we are actually interested in *sets*:

- In *image processing* (e.g., in *astronomy*), the desired black-and-white image is, from the mathematical viewpoint, a *set*.
- In *error estimation* (e.g., in engineering, physics, geophysics, social sciences, etc.), in addition to the estimates $\tilde{x}_1, \ldots, \tilde{x}_n$ for $n$ physical quantities, we want to know what can the *actual* values $x_i$ of these quantities be, i.e., the *set* of all possible vectors $x = (x_1, \ldots, x_n)$.

In both cases, we need to process sets. To define a generic set, we need infinitely many parameters; therefore, if we want to represent and process sets in the computer, we must restrict ourselves to finite-parametric families of sets that will be used to approximate the desired sets. The wrong choice of a family can lead to longer computations and worse approximation. Hence, it is desirable to find the family that it is *the best* in some reasonable sense.

A similar problem occurs for *random sets*. To define a generic set, we need *infinitely many* parameters; as a result, traditional (finite-parametric) statistical methods are often not easily applicable to random sets. To avoid this difficulty, several researchers (including U. Grenander) have suggested to approximate arbitrary sets by sets from a certain *finite-parametric* family. As soon as we fix this family, we can use methods of traditional statistics. Here, a similar problem appears: a wrong choice of an approximation family can lead to a bad approximation and/or long computations; so, which family should we choose?

In this paper, we show, on several application examples, how the problems of choosing the optimal family of sets can be formalized and solved. As a result of the described general methodology:

*The authors are with the Institute of Applied Astronomy, Russian Academy of Sciences, 8 Zhdanovskaya St., 197042 St. Petersburg, Russia (A.F.), and Departments of Electrical and Computer Engineering (O.K.) and Computer Science (V.K.), University of Texas at El Paso, El Paso, TX 79968, emails olga@ece.utep.edu and vladik@cs.utep.edu.

- for *astronomical images*, we get exactly the geometric shapes that have been empirically used by astronomers and astrophysicists (thus, we have a theoretical explanation for these shapes), and
- for *error estimation*, we get a theoretical explanation of why *ellipsoids* turn out to be experimentally the best shapes (and also, why ellipsoids are used in Khachiyan's and Karmarkar's algorithms for linear programming).

# 1 Introduction to the problem

## 1.1 Sets are needed

There are two main cases when we need to process geometric sets (i.e., subsets of $R^k$): image processing and error estimation.

### 1.1.1 Sets are needed for image processing

In *image processing*, our goal is to restore the actual image. For black-and-white images, the image can be identified with a set of its black points, i.e., with a set in a 2-D or in a 3-D space. So, in order to process images, we must be able to process sets.

### 1.1.2 Sets are needed for error estimation

Sets are also needed for *error analysis*, when we are processing the measurement results.

**Sets are needed to represent the measurement results.** Indeed, measurements are never absolutely accurate, so, the actual (unknown) values $x_i$ of the measured quantities can differ from the measurement results $\tilde{x}_i$. The difference $\Delta x_i = \tilde{x}_i - x_i$ between the measurement result and the actual value is called the *measurement error*.

In some cases, we know the probabilities of different values of measurement error $\Delta x_i$, but in many real-life situations, we only know the *bounds* $\Delta_i$ of these error values (these bounds are usually provided by the manufacturers of the corresponding measuring instruments; see, e.g., [14, 3, 24]).

When we know the measurement result $\tilde{x}_i$ and the error bound $\Delta x_i$, the only thing that we know about the actual value $x_i$ of the measured quantity is that it must be somewhere in the interval $\mathbf{x}_i = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$. As a result, the possible values of $x = (x_1, \ldots, x_n)$ form a "box" $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$.

In addition to the measurement results, we may have some extra information about the values $x_i$: e.g., we may know that the actual values $x_1, \ldots, x_n$ must satisfy one or more equations or inequalities (e.g., coming from the energy conservation law or from the second law of thermodynamics). If we take

these additional requirements into consideration, we get sets of possible values of $(x_1, \ldots, x_n)$ that are of more complicated shape than a box.

**Sets are needed to represent the results of data processing.** In some cases, we can directly measure all the physical quantities that we are interested in: e.g., in electrical measurements, a tester can measure current, voltage, resistance, etc. However, in many other real-life situations, it is difficult or even impossible to directly measure the desired quantities $y_1, \ldots, y_m$: e.g., it is impossible to directly measure the distance to a distant star or the amount of oil in a well. In such cases, we:

- *measure* some quantities $x_1, \ldots, x_n$ that are related with $y_j$ in a known way, and then
- *process* the measurement results $\tilde{x}_i$ of the direct measurements to get the desired estimates $\tilde{y}_j$ for $y_j$.

For example, to estimate the amount of oil $y$ in a given well, we measure how sound waves and electromagnetic fields travel through the Earth, measure the density of samples extracted when this well was dug, etc., and use this information to estimate $y$. This two-step procedure, consisting of direct measurements and data processing, is called *indirect measurement.*

Since measurements are not 100% accurate, as a result of direct measurements, we do not get the *precise* values of the measured quantities $x_1, \ldots, x_n$, we get a *set* $X$ of possible values of $x = (x_1, \ldots, x_n)$. Different values $x \in X$ lead, in general, to different values of the desired quantities $y = (y_1, \ldots, y_m)$. So, as the result of the indirect measurement, we do not get a single vector $y$, we get a *set* $Y$ of possible values of $y$. How can we estimate $Y$?

**Set-valued analysis and interval computations.** We want to apply the algorithm $f(x_1, \ldots, x_n)$ that reconstructs the vector $y = (y_1, \ldots, y_m)$ from $x_i$, to the *set* $X$ of possible values, to get the set

$$Y = f(X) = \{f(x) \,|\, x \in X\}.$$

In other words, we would like to extend operations and algorithms that are normally defined for *real numbers* and *vectors* to *sets* of real numbers and vectors. Such extensions form the basis of a mathematical discipline called *set-valued analysis* (see, e.g., [4, 5]).

The simplest case of set-valued analysis is when the set $X$ that describes the initial uncertainty is a box (parallelepiped) $\mathbf{x} = \mathbf{x}_1 \times \ldots \times \mathbf{x}_n$, i.e., if we know an *interval* $\mathbf{x}_i$ that describes the possible values of each variable $x_i$, and no relations between the actual values $x_i$ are known. This particular case of set-valued analysis is called *interval analysis*, or *interval computations.*

Interval computations and more general types of set-valued analysis have numerous applications:

- to *engineering* (manufacturing, quality control, automatic control, robotics, airplane navigation, civil engineering, traffic control, etc.);
- to *social sciences*;
- to *physics* (design and analysis of laser beams particle accelerators, etc. and to astrophysics and image processing);
- to *geology and geophysics*;
- to *chemistry* and *chemical engineering*;
- to *expert systems*,
- etc.

(see, e.g., [18, 16, 19], and the interval computations website `http://cs.utep.edu/interval-comp/main.html`).

## 1.2 In the computer, we can only use finite-parametric families of sets

Images can, in principle, be arbitrarily complicated. Sets $X$ that stem from error estimation can also become more and more complicated as we take into consideration more and more *a priori* relations between the corresponding physical quantities.

The ideal description of a set $X \subseteq R^k$ would include, for any point $x \in R^k$, an information whether this point $x$ belongs to the given set $X$ or not. This information requires infinitely many *bits* (binary digits) to store. However, inside any given computer, we can only store finitely many bits, and therefore, we can represent the information only about finitely many points $x \in R^k$. In computer imaging, these points are usually called *pixels*.

A pixel-by-pixel representation is necessary for some images (e.g., to store a high quality photo) and for computer games (to create the most realistic picture). However, such a representation requires a lot of computer memory and makes processing the corresponding data extremely slow. Therefore, if want to speed up the processing of these sets, we must somehow approximate arbitrarily complicated sets by sets that can be characterized by a few real-valued parameters, i.e., by sets that belong to some *finite-dimensional family of sets*.

Several families of this type have been efficiently used both in image processing and in error estimation: e.g., in error estimation, boxes, ellipsoids, polytopes, etc., have been successfully used. This leads us to a problem:

### 1.3 Main problem: which families of sets should we choose

In principle, different families of sets can be used. It turns out that often, the use of different approximating families leads to different quality of the resulting approximation. Therefore, it is important to choose the right approximating family.

Currently, this choice is mainly made *ad hoc*, at best, by testing a few possible families and choosing the one that performs the best on a few benchmarks. Since only a few families are analyzed, we are not sure that we did not miss the real good approximating family. (And since only a few benchmarks are used for comparison, we are not sure that the chosen family is indeed the best one.) It is, therefore, desirable to find the *optimal* family of approximating sets.

### 1.4 A similar problem is important for random sets

In many practical situations, e.g., in image processing in the presence of uncertainty, we cannot determine the *exact* image, we can at best describe a class of possible images; some sets (images) are more probable, some are less probable, so we have a *probability measure* on the class of sets. In mathematical terms, such a probability measure represents a *random set*.

Traditional statistical methods are well developed for the case when we need to find out *finitely many* parameters. To define a generic set, however, we need *infinitely many* parameters; therefore, traditional statistical methods are often not easily applicable to random sets. To avoid this difficulty, several researchers (including U. Grenander) have suggested to approximate arbitrary sets by sets from a certain *finite-parametric* family. As soon as we fix this family, we can use methods of traditional statistics. However, a similar problem appears: a wrong choice of an approximation family can lead to a bad approximation and/or long computations; so, which family should we choose?

### 1.5 What we are planning to do

In this paper, we will describe a general framework for finding the optimal family, and illustrate this general idea on two examples: one from the astronomical imaging and one from error estimation.

## 2 What does "optimal" mean? Motivations for the following definitions

### 2.1 What is "optimality criterion"

When we say "optimal", we mean optimal w.r.t. to some *optimality criterion*. When we say that some *optimality criterion* is given, we mean that, given two different families of approximating sets, we can decide whether the first one

is better, or that the second one is better, or that these families are of the same quality w.r.t. the given criterion. In mathematical terms, this means that we have a *pre-ordering relation* $\preceq$ on the set of all possible finite-dimensional families of sets.

## 2.2 We want to solve an ambitious problem: enumerate all finite-dimensional families of sets that are optimal relative to some natural criteria

One way to approach the problem of choosing the "best" family of sets is to select *one* optimality criterion, and to find a family of sets that is the best with respect to this criterion. The main drawback of this approach is that there can be different optimality criteria, and they can lead to different optimal solutions. It is, therefore, desirable not only to describe a family of sets that is optimal relative to *some* criterion, but to describe *all* families of sets that can be optimal relative to different natural criteria. In this paper, we are planning to implement exactly this more ambitious task.

## 2.3 Examples of optimality criteria

### 2.3.1 Numerical optimality criteria

Pre-ordering is the general formulation of optimization problems in general, not only of the problem of choosing a family of sets. In general optimization theory, in which we are comparing arbitrary *alternatives* $A$, $B$, ..., from a given set $\mathcal{A}$, the most frequent case of such a pre-ordering is when a *numerical criterion* is used, i.e., when a function $J : \mathcal{A} \to R$ is given for which $A \preceq B$ iff $J(A) \le J(B)$.

Several natural numerical criteria can be proposed for choosing the best family of sets for error estimation: Each choice of a family means that we approximate the actual set of possible values $X$ by an element $\tilde{X} \supseteq X$ from the chosen family. Therefore, for a data processing algorithm $f$, the resulting set $f(\tilde{X})$ is an "overestimation" of the actual set $f(X)$. We can measure this "overestimation", e.g., by computing the Lebesgue measure of the difference $f(\tilde{X}) \backslash f(X)$, or by computing the Hausdorff distance between these two sets. As an optimality criterion, we can, e.g., choose the *average* measure of overestimation (average in the sense of some natural probability measure on the class of all problems).

Alternatively, we can fix a class of the problem, and take the *largest* (worst-case) measure of overestimation for problems of this class as the desired (numerical) optimality criterion.

### 2.3.2 Non-numerical optimality criteria naturally appear

For "worst-case" optimality criteria, it often happens that there are several different alternatives that perform equally well in the worst case, but whose performance differ drastically in the average cases. In this case, it makes sense, among all the alternatives with the optimal *worst-case* behavior, to choose the one for which the *average* behavior is the best possible. This very natural idea leads to the optimality criterion that is *not* described by a numerical optimality criterion $J(A)$: in this case, we need *two* functions: $J_1(A)$ describes the worst-case behavior, $J_2(A)$ describes the average-case behavior, and $A \preceq B$ iff either $J_1(A) < J_2(B)$, or $J_1(A) = J_1(B)$ and $J_2(A) \leq J_2(B)$.

We could further specify the described optimality criterion and end up with *a* natural criterion. However, as we have already mentioned, the goal of this paper is not to find *a* family of sets that is optimal relative to some criterion, but to describe *all* families of sets that are optimal relative to some natural optimality criteria. In view of this goal, in the following text, we will not specify the criterion, but, vice versa, we will describe a very general class of *natural* optimality criteria.

So, let us formulate what "natural" means.

## 2.4 Which optimality criteria are natural

### 2.4.1 The criterion must be invariant

Problems related to geometric sets often have natural *symmetries*. For example, let us consider astronomical images. These images are sets in $R^3$ (or in $R^2$). For such sets, there are three natural symmetries:

First, if we *change the starting point* of the coordinate system from the previous origin point $O = (0, 0, 0)$ to the new origin $O'$ whose coordinates were initially $a = (a_1, a_2, a_3)$, then each point $x$ with old coordinates $(x_1, x_2, x_3)$ gets new coordinates $x'_i = x_i - a_i$. As a result, in the new coordinates, each set $X \in A$ from a family of images $A$ will be described by a "shifted" set $T_a(X) = \{x - a \,|\, x \in X\}$, and the family turns into $T_a(A) = \{T_a(X) \,|\, X \in A\}$. It is reasonable to require that the relative quality of the two families of sets do not depend on the choice of the origin. In other words, we require that if $A$ is better than $B$, then the "shifted" $A$ (i.e., $T_a(A)$) should be better than the "shifted" $B$ (i.e., that $T_a(B)$).

Second, the choice of a *rotated* coordinate system is equivalent to rotating all the points $(x \to R(x))$, i.e., going from a set $X$ to a set $R(X) = \{R(x) \,|\, x \in X\}$, and from a family $A$ to a new family $R(A) = \{R(X) \,|\, X \in A\}$. It is natural to require that the optimality criterion is invariant w.r.t. rotations, i.e., if $A$ is better than $B$, then $R(A)$ is better than $R(B)$.

Third, it is often difficult to find the exact distance to the observed object. Therefore, we are not sure whether the observed image belongs to a small nearby object, or to a larger but distant one. As a result of this uncertainty, the actual

image is only known modulo homothety (similarity, dilation) $x \to \lambda \cdot x$ for some real number $\lambda > 0$. It is, therefore, natural to require that the desired optimality criterion be invariant w.r.t. homothety.

### 2.4.2 The criterion must be final

If the criterion does not select any family as an optimal one, i.e., if, according to this criterion, none of the families is better than the others, then this criterion is of no use in selection.

If the criterion considers several different families equally good, then we can always use some other criterion to help select between these "equally good" ones, thus designing a two-step criterion. If this new criterion still does not select a unique family, we can continue this process until we arrive at a combination multi-step criterion for which there is only one optimal family.

Therefore, we can always assume that our criterion is *final* in the sense that it selects one and only one optimal family.

## 3 Definitions and the main result

Our goal is to choose the best finite-parametric family of sets. To formulate this problem precisely, we must formalize what a finite-parametric family is and what it means for a family to be optimal. In accordance with our informal description, both formalizations will use natural symmetries. So, we will first formulate how symmetries can be defined for families of sets, then what it means for a family of sets to be finite-dimensional, and finally, how to describe an optimality criterion.

**Definition 1.** *Let $g : M \to M$ be a 1-1-transformation of a set $M$, and let $A$ be a family of subsets of $M$. For each set $X \in A$, we define the result $g(X)$ of applying this transformation $g$ to the set $X$ as $\{g(x) \mid x \in X\}$, and we define the result $g(A)$ of applying the transformation $g$ to the family $A$ as the family $\{g(X) \mid X \in A\}$.*

**Definition 2.** *Let $M$ be a smooth manifold. A group $G$ of transformations $M \to M$ is called a* Lie transformation group, *if $G$ is endowed with a structure of a smooth manifold for which the mapping $g, a \to g(a)$ from $G \times M$ to $M$ is smooth.*

We want to define $r-$parametric families sets in such a way that symmetries from $G$ would be computable based on parameters. Formally:

**Definition 3.** *Let $M$ and $N$ be smooth manifolds.*

- *By a* multi-valued function $F : M \to N$ *we mean a function that maps each $m \in M$ into a discrete set $F(m) \subseteq N$.*

- *We say that a multi-valued function is* smooth *if for every point $m_0 \in M$ and for every value $f_0 \in F(m)$, there exists an open neighborhood $U$ of $m_0$ and a smooth function $f : U \to N$ for which $f(m_0) = f_0$ and for every $m \in U$, $f(m) \subseteq F(m)$.*

**Definition 4.** *Let $G$ be a Lie transformation group on a smooth manifold $M$.*

- *We say that a class $A$ of closed subsets of $M$ is* $G-$invariant *if for every set $X \in A$, and for every transformation $g \in G$, the set $g(X)$ also belongs to the class.*
- *If $A$ is a $G-$invariant class, then we say that $A$ is a* finitely parametric family of sets *if there exist:*
  - *a (finite-dimensional) smooth manifold $V$;*
  - *a mapping $s$ that maps each element $v \in V$ into a set $s(v) \subseteq M$; and*
  - *a smooth multi-valued function $\Pi : G \times V \to V$*

  *such that:*
  - *the class of all sets $s(v)$ that corresponds to different $v \in V$ coincides with $A$, and*
  - *for every $v \in V$, for every transformation $g \in G$, and for every $\pi \in \Pi(g, v)$, the set $s(\pi)$ (that corresponds to $\pi$) is equal to the result $g(s(v))$ of applying the transformation $g$ to the set $s(v)$ (that corresponds to $v$).*
- *Let $r > 0$ be an integer. We say that a class of sets $B$ is a* $r-$parametric class *of sets if there exists a finite-dimensional family of sets $A$ defined by a triple $(V, s, \Pi)$ for which $B$ consists of all the sets $s(v)$ with $v$ from some $r-$dimensional sub-manifold $W \subseteq V$.*

**Definition 5.** *Let $\mathcal{A}$ be a set, and let $G$ be a group of transformations defined on $\mathcal{A}$.*

- *By an* optimality criterion, *we mean a* pre-ordering *(i.e., a transitive reflexive relation) $\preceq$ on the set $\mathcal{A}$.*
- *An optimality criterion is called* $G$-invariant *if for all $g \in G$, and for all $A, B \in \mathcal{A}$, $A \preceq B$ implies $g(A) \preceq g(B)$.*
- *An optimality criterion is called* final *if there exists one and only one element $A \in \mathcal{A}$ that is preferable to all the others, i.e., for which $B \preceq A$ for all $B \neq A$.*
- *An optimality criterion is called* natural *if it is $G-$invariant and final.*

**Theorem 1.** *Let $M$ be a manifold, let $G$ be a $d-$dimensional Lie transformation group on $M$, and let $\preceq$ be a natural (i.e., $G-$invariant and final) optimality criterion on the class $\mathcal{A}$ of all $r-$parametric families of sets from $M$, $r < d$. Then:*

- *the optimal family $A_{\rm opt}$ is $G-$invariant; and*
- *each set $X$ from the optimal family is a union of orbits of*

  *$\geq (d-r)-$dimensional subgroups of the group $G$.*

**Proof.** Since the criterion $\preceq$ is final, there exists one and only one optimal family of sets. Let us denote this family by $A_{\rm opt}$.

1. Let us first show that this family $A_{\rm opt}$ is indeed $G-$invariant, i.e., that $g(A_{\rm opt}) = A_{\rm opt}$ for every transformation $g \in G$.

Indeed, let $g \in G$. From the optimality of $A_{\rm opt}$, we conclude that for every $B \in \mathcal{A}$, $g^{-1}(B) \preceq A_{\rm opt}$. From the $G-$invariance of the optimality criterion, we can now conclude that $B \preceq g(A_{\rm opt})$. This is true for all $B \in \mathcal{A}$ and therefore, the family $g(A_{\rm opt})$ is optimal. But since the criterion is final, there is only one optimal family; hence, $g(A_{\rm opt}) = A_{\rm opt}$. So, $A_{\rm opt}$ is indeed invariant.

2. Let us now show an arbitrary set $X_0$ from the optimal family $A_{\rm opt}$ consists of orbits of $\geq (d-r)-$dimensional subgroups of the group $G$.

Indeed, the fact that $A_{\rm opt}$ is $G-$invariant means, in particular, that for every $g \in G$, the set $g(X_0)$ also belongs to $A_{\rm opt}$. Thus, we have a (smooth) mapping $g \to g(X_0)$ from the $d-$dimensional manifold $G$ into the $\leq r-$dimensional set $G(X_0) = \{g(X_0) \,|\, g \in G\} \subseteq A_{\rm opt}$. In the following, we will denote this mapping by $g_0$.

Since $r < d$, this mapping cannot be 1-1, i.e., for some sets $X = g'(X_0) \in G(X_0)$, the pre-image $g_0^{-1}(X) = \{g \,|\, g(X_0) = g'(X_0)\}$ consists of one than one point. By definition of $g(X)$, we can conclude that $g(X_0) = g'(X_0)$ iff $(g')^{-1}g(X_0) = X_0$. Thus, this pre-image is equal to $\{g \,|\, (g')^{-1}g(X_0) = X_0\}$. If we denote $(g')^{-1}g$ by $\tilde{g}$, we conclude that $g = g'\tilde{g}$ and that the pre-image $g_0^{-1}(X) = g_0^{-1}(g'(X_0))$ is equal to $\{g'\tilde{g} \,|\, \tilde{g}(X_0) = X_0\}$, i.e., to the result of applying $g'$ to $\{\tilde{g} \,|\, \tilde{g}(X_0) = X_0\} = g_0^{-1}(X_0)$. Thus, each pre-image $(g_0^{-1}(X) = g_0^{-1}(g'(X_0)))$ can be obtained from one of these pre-images (namely, from $g_0^{-1}(X_0)$) by a smooth invertible transformation $g'$. Thus, all pre-images have the same dimension $D$.

We thus have a *stratification* (fiber bundle) of a $d-$dimensional manifold $G$ into $D-$dimensional strata, with the dimension $D_f$ of the factor-space being $\leq r$. Thus, $d = D + D_f$, and from $D_f \leq r$, we conclude that $D = d - D_f \geq n - r$.

So, for every set $X_0 \in A_{\rm opt}$, we have a $D \geq (n-r)-$dimensional subset $G_0 \subseteq G$ that leaves $X_0$ invariant (i.e., for which $g(X_0) = X_0$ for all $g \in G_0$). It is easy to check that if $g, g' \in G_0$, then $gg' \in G_0$ and $g^{-1} \in G_0$, i.e., that $G_0$ is a *subgroup* of the group $G$. From the definition of $G_0$ as $\{g \,|\, g(X_0) = X_0\}$ and

the fact that $g(X_0)$ is defined by a smooth transformation, we conclude that $G_0$ is a smooth sub-manifold of $G$, i.e., a $\geq (n-r)-$dimensional subgroup of $G$.

To complete our proof, we must show that the set $X_0$ is a union of orbits of the group $G_0$. Indeed, the fact that $g(X_0) = X_0$ means that for every $x \in X_0$, and for every $g \in G_0$, the element $g(x)$ also belongs to $X_0$. Thus, for every element $x$ of the set $X_0$, its entire orbit $\{g(x) \,|\, g \in G_0\}$ is contained in $X_0$. Thus, $X_0$ is indeed the union of orbits of $G_0$. Q.E.D.

# 4 Astrogeometry: the first application of the main result

Celestial bodies such as galaxies, stellar clusters, planetary systems, etc., have different geometric shapes (e.g., galaxies can be spiral or circular, etc.). Usually, complicated physical theories are used to explain these shapes; for example, several dozen different theories explain why many galaxies are of spiral shape; see, e.g., [29, 28, 32, 7]. Some rare shapes are still unexplained.

In this section, we show that to explain these "astroshapes", we do not need to know the details of *physical* equations: practically all the shapes of celestial bodies can be explained by simple *geometric* invariance properties. This fact explains, e.g., why so many different physical theories lead to the same spiral galaxy shape.

## 4.1 The symmetry group that corresponds to astrogeometry

In *astrogeometry* (i.e., in analysis of geometric astronomical images), we are interested in images $X \subset R^3$. As have already mentioned, for astronomical images, the natural group of symmetries $G_a$ is generated by shifts, rotations, and dilations.

So, to apply our main result to astrogeometry, we must describe all orbits of subgroups of $G_a$.

## 4.2 How to describe orbits of subgroups of $G_a$

A 1-D orbit is an orbit of a 1-D subgroup. This subgroup is uniquely determined by its "infinitesimal" element, i.e., by the corresponding element of the Lie algebra of the group $G$. This Lie algebra if easy to describe. For each of its elements, the corresponding differential equation (that describes the orbit) is reasonably easy to solve.

2-D forms are orbits of $\geq$ 2-D subgroups, so, they can be enumerated by combining two 1-D subgroups.

*Comment.* An alternative (slightly more geometric) way of describing 1-D orbits is to take into consideration that an orbit, just like any other curve in a 3-D

space, is uniquely determined by its curvature $\kappa_1(s)$ and torsion $\kappa_2(s)$, where $s$ is the arc length measured from some fixed point. The fact that this curve is an orbit of a 1-D group means that for every two points $x$ and $x'$ on this curve, there exists a transformation $g \in G$ that maps $x$ into $x'$. Shifts and rotations do not change $\kappa_i$, they may only shift $s$ (to $s + s_0$); dilations also change $s$ to $s \to \lambda \cdot s$ and change the numerical values of $\kappa_i$. So, for every $s$, there exist $\lambda(s)$ and $s_0(s)$ such that the corresponding transformation turns a point corresponding. to $s = 0$ into a point corresponding to $s$. As a result, we get functional equations that combine the two functions $\kappa_i(s)$ and these two functions $\lambda(s)$ and $s_0(s)$. Taking an infinitesimal value $s$ in these functional equations, we get differential equations, whose solution leads to the desired 1-D orbits.

## 4.3 As a result of applying our main idea, we get exactly all observable astroshapes

### 4.3.1 Possible orbits

The resulting description of 0−, 1−, and 2−dimensional orbits of connected subgroups $G_a$ of the group $G$ is as follows:

0: The only 0−dimensional orbit is a *point.*

1: A generic 1−dimensional orbit is a *conic spiral* that is described (in cylindrical coordinates) by the equations $z = k\rho$ and $\rho = R_0 \exp(c\varphi)$. Its limit cases are:

- a *logarithmic* (Archimedean) *spiral*: a planar curve ($z = 0$) that is described (in polar coordinates) by the equation $\rho = R_0 \exp(c\varphi)$.
- a *cylindrical spiral*, that is described (in appropriate coordinates) by the equations $z = k\phi$, $\rho = R_0$.
- a *circle* ($z = 0$, $\rho = R_0$);
- a *semi-line* (*ray*);
- a *straight line.*

2: Possible 2-D orbits include:

- a *plane*;
- a *semi-plane*;
- a *sphere*;
- a circular *cone*;
- a *circular cylinder*, and
- a *logarithmic cylinder*, i.e., a cylinder based on a logarithmic spiral.

### 4.3.2 Possible orbits are exactly possible shapes

Comparing these orbits (and ellipsoids, the ultimate stable shapes) with astroshapes enumerated, e.g., in [32], we conclude that:

- First, our scheme describes all observed connected shapes.
- Second, all above orbits, except the logarithmic cylinder, have actually been observed as shapes of celestial bodies.

For example, according to Chapter III of [32], galaxies consist of components of the following geometric shapes:

- *bars* (cylinders);
- *disks* (parts of the plane);
- *rings* (circles);
- *arcs* (parts of circles and lines);
- *radial rays*;
- *logarithmic spirals*;
- *spheres*, and
- *ellipsoids.*

It is easy to explain why logarithmic cylinder was never observed: from whatever point we view it, the logarithmic cylinder blocks all the sky, so it does not lead to any visible shape in the sky at all. With this explanation, we can conclude that we have a *perfect explanation of all observed astroshapes.*

### 4.3.3 Comment: we can also explain difficult-to-explain disconnected shapes

In the above description, we only considered connected *continuous* subgroups $G_0 \subseteq G$. Connected continuous subgroups explain connected shapes.

It is natural to consider disconnected (in particular, discrete) subgroups as well; the orbits of these subgroups leads to disconnected shapes. Thus, we can explain these shapes, most of which modern astrophysics finds pathological and difficult to explain (see, e.g., [32], Section I.3). For example, an orbit $O$ of a discrete subgroup $G_0'$ of the 1-D group $G_0$ (whose orbit is a logarithmic spiral) consists of points whose distances $r_n$ to the center forms a geometric progression: $r_n = r_0 \cdot k^n$. Such dependence (called Titzius-Bode law) has indeed been observed (as early as the 18th century) for planets of the Solar system and for the satellites of the planets (this law actually led to the prediction and

discovery of what is now called asteroids). Thus, we get a *purely geometric explanation of the Titzius-Bode law.*

Less known examples of disconnected shapes that can be explained in this manner include:

- several parallel equidistant lines ([32], Section I.3);
- several circles located on the same cone, whose distances from the cone's vertex form a geometric progression ([32], Section III.9);
- equidistant points on a straight line ([32], Sections VII.3 and IX.3);
- "piecewise circles": equidistant points on a circle; an example is MCG 0-9-15 ([32], Section VII.3);
- "piecewise spirals": points on a logarithmic spiral whose distances from a center form a geometric progression; some galaxies of Sc type are like that [32].

#### 4.3.4 This idea also explains: evolution of geometric shapes, their relative frequency, directions of rotation and of magnetic field

Our main idea explains not only the shapes themselves, but also how they evolve, which are more frequent, etc. For details, see, e.g., [22, 21, 20, 12].

*Comment.* V.I. Arnold has shown (see, e.g., [30, 2]) that dynamical systems theory explains why the observed shape should be *topological homeomorphic* to a spiral. We have explained even more: not only that this shape is *homeomorphic* to the spiral, but that geometrically, this shape is *exactly* a *logarithmic spiral.*

# 5 Applications to error estimation: why ellipsoids

## 5.1 Ellipsoids are actively used in error estimation (as well as in numerical computations and in pattern recognition)

Ellipsoid error estimates are actively (and successfully) used in different applications (see, e.g., [25, 26, 13, 6, 23, 9, 27, 31, 11, 10]).

Several other families of sets have been proposed to describe errors, such as parallelepipeds ("boxes"), polytopes, etc. Experimental comparison of different families has lead to a conclusion that *ellipsoids* lead to the best results (see, e.g., [9, 10]).

There are at least two other areas where ellipsoids turned out to be the best approximators for arbitrary sets:

- *Linear programming* means finding maxima of linear functions on a set defined by a system by linear inequalities (i.e., on a convex polytope). The traditionally used *simplex method* uses the original polytope; this method is, on average, very efficient, but in the worst case, it requires the unrealistic exponential number of computational steps ($\approx 2^n$, where $n$ is the number of unknowns). For several decades, researchers have tried to find a polynomial time algorithm for linear programming. Success only came when they decided to approximate the original polytope with an ellipsoid; this lead to the well-known polynomial time algorithms of Khachiyan [17] and Karmarkar [15].
- Ellipsoids also turned out to be better than polytopes or parallelepipeds (boxes) in many *pattern recognition* problems (see, e.g., [1]).

A natural question is: Are ellipsoids really optimal (in some reasonable sense), or are they only an empirical approximation to the truly optimal family?

In this paper, we show that this empirical choice can be theoretically justified.

## 5.2 The symmetry group that corresponds to error estimation: motivation

In *error estimation*, we are interested in sets of possible values of $n$ physical quantities, i.e., in mathematical terms, in subsets of $R^n$. Since we already know the approximate values $\tilde{x}_1, \ldots, \tilde{x}_n$ of the desired quantities, it is sufficient to describe the set of possible values of *errors* $\Delta x_i = \tilde{x}_i - x_i$.

In addition to the "basic" quantities $x_1, \ldots, x_n$, we are usually also interested in their combinations (e.g., in addition to the temperatures we would also like to know the average temperature). From the user's viewpoint, it does not really matter which $n$ of desired quantities we denote as "basic" ones, and which as their "combinations". So, instead of the original basic quantities $x_1, \ldots, x_n$, we could consider different quantities $x_i' = f_i(x_1, \ldots, x_n)$. This change will change the shape of the approximating sets. It is reasonable to assume that if for the original variables, a family of sets $A$ was better than some other family $B$, then the correspondingly "transformed" family $A$ will still be better than the similarly transformed family $B$.

Since we are interested only in the possible values of errors $\Delta x_i$, and errors are usually relatively small, we can safely neglect in the Taylor expansion of the transformations $f_i$ all the terms that are quadratic or higher order in $\Delta x_i$ and thus assume that the transformations are linear.

Thus, as a natural symmetry group $G_e$, it is natural to consider the group of all *affine* (*linear*) transformations $R^n \to R^n$, i.e., of all transformations of the type $x_i \to a_i + \sum_j a_{ij} x_j$ with an invertible matrix $a_{ij}$.

*Comment.* For *linear programming*, the restriction to linear transformations is even more natural that for error estimation.

## 5.3 Main result

We will show that the ellipsoids are the *simplest* optimal family, i.e., that of all possible optimal finite-parametric families that correspond to different $G_e-$invariant optimality criteria, ellipsoids have the smallest number of parameters.

**Definition 6.** *By a* closed domain, *we mean a closed set that is equal to the closure of the set of its interior points.*

**Theorem 2.** *Let $n > 0$ be an integer, $M = R^n$, $G_e$ be the group of all affine transformations, and $\preceq$ be a natural (i.e., $G_e-$invariant and final) optimality criterion on the class $\mathcal{A}$ of all $r-$parametric families of connected bounded closed domains from $R^n$. Then:*

- *$r \geq n(n+3)/2$;*
- *if $r = n(n+3)/2$, then the optimal family coincides either with the family of all ellipsoids, or, for some $\lambda \in (0,1)$, with the family of all regions obtained from ellipsoids by subtracting $\lambda$ times smaller homothetic ellipsoids.*

*Comment.* If we restrict ourselves to *convex* sets (or only to simply connected sets), we get ellipsoids only.

**Proof.** Due to Theorem 1, the optimal family $A_{\rm opt}$ is affine invariant, i.e., for every $X \in A_{\rm opt}$, and for every transformation $g \in G_e$, the set $g(X)$ also belongs to $A_{\rm opt}$.

1. Let us first show that $r \geq n(n+3)/2$. Indeed, it is known (see, e.g., [8]) that for every open bounded set $X$, among all ellipsoids that contain $X$, there exists a unique ellipsoid $E$ of the smallest volume. We will say that this ellipsoid $E$ *corresponds* to the set $X$. Let us consider the set of ellipsoids $\mathcal{E}_c$ that correspond (in this sense) to all possible sets $X \in A_{\rm opt}$.

Let us fix a set $X_0 \in A_{\rm opt}$, and let $E_0$ denote an ellipsoid that corresponds to $X_0$.

An arbitrary ellipsoid $E$ can be obtained from any other ellipsoid (in particular, from $E_0$) by an appropriate affine transformation $g$: $E = g(E_0)$. The ratio of volumes is preserved under arbitrary linear transformations $g$; hence, since the ellipsoid $E_0$ is the smallest volume ellipsoid that contains $X_0$, the ellipsoid $E = g(E_0)$ is the smallest volume ellipsoid that contains $g(X_0) = X$.

Hence, an arbitrary ellipsoid $E = g(E_0)$ corresponds to some set $g(X_0) \in A_{\rm opt}$. Thus, the family $\mathcal{E}_c$ of all ellipsoids that correspond to sets from $A_{\rm opt}$ is simply equal to the set $\mathcal{E}$ of all ellipsoids. Thus, we have a (locally smooth) mapping from an $r-$dimensional set $A_{\rm opt}$ onto the $n(n+3)/2-$dimensional set of all ellipsoids. Hence, $r \geq n(n+3)/2$.

2. Let us now show that for $r = n(n+3)/2$, the only $G_e-$invariant families $A$ are ellipsoids and "ellipsoid layers" (described in Theorem 2).

Indeed, let $X_0$ be an arbitrary set from the invariant family, and let $E_0$ be the corresponding ellipsoid. Let $g_0 \in G_e$ be an affine transformation that transform $E_0$ into a ball $E_1 = g(E_0)$. This ball then contains the set $X_1 = g_0(E_0) \in A_{\rm opt}$.

Let us show, by reduction to a contradiction, that the set $X_1$ is invariant w.r.t. arbitrary rotations around the center of the ball $E_1$. Indeed, if it is not invariant, then the set $R$ of all rotations that leave $X_1$ invariant is different from the set of all rotations $SO(n)$. Hence, $R$ is a proper closed subgroup of $SO(n)$. From the structure of $SO(n)$, it follows that there exists a $1-$parametric subgroup $R_1$ of $SO(n)$ that intersects with $R$ only in the identity transformation 1. This means that if $g \in R_1$ and $g \neq 1$, we have $g \notin R$, i.e., $g(X_1) \neq X_1$.

If $g(X_1) = g'(X_1)$ for some $g, g' \in R_1$, then we have $g^{-1}g'(X_1) = X_1$, where $g^{-1}g' \in R_1$. But such an equality is only possible for $g^{-1}g' = 1$, i.e., for $g = g'$. Thus, if $g, g' \in R_1$ and $g \neq g'$, then the sets $g(X_1)$ and $g'(X_1)$ are different. In other words, all the sets $g(X_1)$, $g \in R_1$, are different.

Since the family $A$ is $G_e-$invariant, all the sets $g(X_1)$ for all $g \in R_1 \subseteq G_e$ also belong to $A$. For all these sets, the corresponding ellipsoid is $g(E_1)$, the result of rotating the ball $E_1$, i.e., the same ball $g(E_1) = E_1$. Hence, we have a $1-$parametric family of sets contained in the ball $E_1$.

By applying appropriate affine transformations, we will get 1-parametric families of sets from $A$ in an arbitrary ellipsoid. So, we have an $n(n+3)/2-$dimensional family of ellipsoids, and inside each ellipsoid, we have a $1-$dimensional family of sets from $A$. Thus, $A$ would contain a $(n(n+3)/2+1)-$parametric family of sets, which contradicts to our assumption that the dimension $r$ of the family $A$ is exactly $n(n+3)/2$.

This contradiction shows that our initial assumption was false, and for $r = n(n+3)/2$, the set $X_1$ is invariant w.r.t. rotations. Hence, with an arbitrary point $x$, the set $X_1$ contains all the points that can be obtained from $x$ by arbitrary rotations, i.e., the entire sphere that contains $x$. Since $X_1$ is connected, $X_1$ is either a ball, or a ball from which a smaller ball was deleted.

The original set $X_0 = g_0^{-1}(X_1)$ is an affine image of this set $X_1$, and therefore, $X_0$ is either an ellipsoid, or an ellipsoid with an ellipsoidal hole inside. Q.E.D.

## 5.4 Auxiliary result: not only ellipsoids form the optimal family, but approximation by the smallest-volume ellipsoid is the optimal approximation

### 5.4.1 Which ellipsoid should we choose? Formulation of the problem

In the previous section, we have shown that under certain reasonable assumptions, ellipsoids are the best approximators. This result justifies the use of ellipsoids for approximation.

Here, the natural next problem appears: for every set $X$, there are many different ellipsoids $E$ that contain $X$; which of these ellipsoids should we choose?

### 5.4.2 Experimental results

In principle, we can choose the ellipsoid with the smallest volume, or with the smallest diameter, etc.

Experiments [9, 10] has shown that choosing the smallest volume ellipsoid leads to the best approximation.

### 5.4.3 What we are planning to do

In this section, we will show that this choice of an optimality criterion can also be justified along similar lines.

### 5.4.4 Preliminary discussions

At first glance, our goal is to find a function $J : \mathcal{E} \to R$ from the set $\mathcal{E}$ of all ellipsoids to real numbers so that for every set $X$, we will then, among all ellipsoids $E \supseteq X$, choose the ellipsoid $E$ with the smallest possible value of $J(E)$. In the above examples, $J(E)$ was, correspondingly, the volume and the diameter of the ellipsoid $E$.

In reality, however, when we choose the ellipsoid, we do not use the *numerical* values of $J(E)$, we only use the *ordering* that this functions imposes on the calls $\mathcal{E}$ of all ellipsoids. Thus, our choice does not depend on whether we use the original function $J(E)$, or a function $f(J(E))$ for some strictly monotonic function $f$. Therefore, it is more natural to look not for a single function $J(E)$, but for the entire *family* of functions $f(J(E))$ that correspond to all possible strictly monotonic functions $f$.

### 5.4.5 Definitions and the main result

As a result, we arrive at the following definitions:

**Definition 7.** *Let $M$ be an arbitrary topological space, $J : M \to R$ be a continuous function ($J \not\equiv$ const). Then the family consisting of all the functions $m \to f(J(m))$ where $f(x)$ is an arbitrary continuous strictly monotonic function from $R$ to $R$, will be called* a order-defining family of functions. *The family that contains a function $J$ will be denoted by $\{f(J)\}_f$.*

For every transformation $g : M \to M$, and for every function $J(x)$, we can define $g(J)$ as follows: $(g(J))(m) = J(gm)$. One can easily check that this definition actually defines the transformation on order-defining families.

**Theorem 3.** *Let $n > 0$ be an integer, let $M$ be the set $\mathcal{E}$ of all ellipsoids in $R^n$ (with a natural topology), and let $G_e$ be the group $G_e$ of all affine transformations. Let $\preceq$ be a natural (i.e., $G_e-$invariant and final) optimality criterion on the set of all order-defining families. Then, the order defined by the optimal order-defining family coincides either with the order defined by the volume $J(E) = V(E)$, or by the negative volume $J(E) = -V(E)$.*

*Comment.* The choice of negative volume $J(E) = -V(E)$ means that we choose the ellipsoid $E \supseteq X$ with the largest possible volume to approximate a set $X$. Since we can always increase an ellipsoid, none of them has the largest volume, so this criterion is useless. Hence, *the only meaningful optimal criterion for choosing the approximating ellipsoid is to choose an ellipsoid $X \supseteq X$ with the largest volume.*

Thus, our theorem explains the experimental results from [9, 10].

**Proof.** Due to Theorem 1, there exists the optimal order-defining family, $\{f(J_{\rm opt})\}_f$, and this optimal family is invariant w.r.t. $G_e$. The invariance means, in particular, that for every affine transformation $g \in G_e$, the function $g(J_{\rm opt})$ defined as $(g(J_{\rm opt}))(E) = J_{\rm opt}(g(E))$ belongs to the same order-defining family as $J_{\rm opt}$, i.e., that there exists a continuous strictly increasing function $f_g : R \to R$ for which

$$J_{\rm opt}(g(E)) = f_g(J_{\rm opt}(E)) \tag{1}$$

(for all ellipsoids $E$).

1. Let us show that for every $g, h \in G_e$, we have $f_{g(h)}(x) = f_g(f_h(x))$ for all $x$ from the image $I = J_{\rm opt}(\mathcal{E})$ of the function $J_{\rm opt}$.

Indeed, if $x \in J_{\rm opt}(\mathcal{E})$, this means that there exists an ellipsoid $E$ for which $J_{\rm opt}(E) = x$. In this case, by definition of $f_{g(h)}$, we have

$$f_{g(h)}(x) = f_{g(h)}(J_{\rm opt}(E)) = J_{\rm opt}(g(h(E))). \tag{2}$$

On the other hand, similarly, $f_h(x) = f_h(J_{\rm opt}(E)) = J_{\rm opt}(h(E))$ and therefore,

$$f_g(f_h(x)) = f_g(J_{\rm opt}(h(E))) = J_{\rm opt}(g(h(E))). \tag{3}$$

From the equations (2) and (3), we get the desired equality.

2. Let us now show that if for some integer $m$, $g$ is a rotation by $2\pi/m$ around a line, then $f_g(x) = x$ for all $x \in I$.

We will prove this equality by reduction to a contradiction. Indeed, let us assume that $f_g(x) \neq x$ for some $x \in I$. This means that either $f_g(x) > x$, or $f_g(x) < x$. Let us show the first inequality is impossible (the second inequality can be proven similarly). Indeed, if $f_g(x) < x$, then from the fact that $f_g$ is a strictly increasing function, we conclude that $f_g(f_g(x)) < f_g(x)$, and since $f_g(x) < x$, we conclude that $f_g(f_g(x)) < x$. Similarly, we can prove that $f_g(\ldots(f_g(x))\ldots)$ ($m$ times) $< x$. According to part 1 of the proof, we have $f_g(\ldots(f_g(x))\ldots)$ ($m$ times) $= f_{g^m}(x) < x$. Since $g$ is a rotation by $2\pi/m$, we have $g^m = 1$ and $f_{g^m}(x) = f_1(x) = x$; hence, $f_{g^m}(x) < x$ leads to $x < x$: a contradiction.

This contradiction, together with a similar contradiction for $f_g(x) > x$, proves that indeed $f_g(x) = x$ for all $x \in I$.

3. Let us now show that $f_g(x) = x$ for a rotation $g$ by an arbitrary angle $\alpha$.

In part 2, we have already proven this statement for rotations by angle $2\pi/m$.

If $\alpha = 2\pi p/m$ for some integers $p$ and $m$, then the rotation by $\alpha$ can be represented as a composition of $p$ rotations by $2\pi/m$. For each of these rotations $g_0$, we have already shown that $f_{g_0}(x) = x$. Thus, from part 1, we conclude that $f_g(x) = f_{g_0}(\ldots(f_{g_0}(x))\ldots) = x$.

To conclude this prove, we remark that due to $f_g(J_{\rm opt}(E)) = J_{\rm opt}(g(E))$ and continuity of $J_{\rm opt}$ and $E \to g(E)$, the function $f_g$ continuously depends on $g$. Since an arbitrary angle $\alpha$ can be approximated, with arbitrary accuracy, by angles of the type $2\pi p/m$ for which $f_g(x) = x$, we can thus conclude that $f_g(x) = x$ for an arbitrary $\alpha$.

4. Let us now prove that $f_g(x) = x$ for an arbitrary motion in $R^n$ (i.e., for an arbitrary linear transformation that preserves Euclidean metric).

Indeed, it is known that an arbitrary motion $g$ can be represented as a composition of finitely many rotations $g_1, \ldots, g_p$ around different lines. For each of these rotations $g_i$, we have already proven that $f_{g_i}(x) = x$. Therefore, due to (1), we can conclude that $f_g(x) = x$.

5. A similar result $f_g(x) = x$ can proven for an arbitrary volume-preserving linear transformation $g$.

This conclusion follows from the fact that an arbitrary volume-preserving linear transformation can be represented, in appropriate coordinates, as a motion.

6. It is known that every two ellipsoids $E_1$ and $E_2$ can be transformed into each other by an affine transformation $g$ (i.e., $E_2 = g(E_1)$), and that if they have the same volume ($V(E_1) = V(E_2)$), then $g$ is volume-preserving.

Hence, if two ellipsoids have the same volume, we have $J_{\rm opt}(E_2) = J_{\rm opt}(g(E_1)) = f_g(J_{\rm opt}(E_1))$ for a volume-preserving transformation $g$. Due to part 5 of the proof, we can conclude that $J_{\rm opt}(E_2) = J_{\rm opt}(E_1)$.

Thus, the value of $J_{\rm opt}(E)$ depend only on the volume $V(E)$ of the ellipsoid $E$, i.e.,

$$J_{\rm opt}(E) = F(V(E)) \tag{4}$$

for some function $F : R \to R$.

Since $J_{\rm opt}$ is continuous, the function $F$ is also continuous.

7. To complete the proof, it is sufficient to show that the function $F$ is 1-1. Then, it will be either strictly increasing (and thus, equivalent to $x$), or strictly decreasing (and thus, equivalent to $-x$).

We will prove this by reduction to a contradiction. Indeed, let us assume that $F(x_1) = F(x_2)$ for some positive real numbers $x_1 < x_2$. Let us take an arbitrary ellipsoid $E_1$ of volume $x_1$ and an ellipsoid $E_2$ of volume $x_2$. Then, from (4), we conclude that $J_{\rm opt}(E_1) = F(V(E_1)) = F(x_1) = F(x_2) = F(V(E_2)) = J_{\rm opt}(E_2)$.

Let us denote $k = x_2/x_1 > 1$. Let us show that $F(kz) = F(z)$ for all $z$. Indeed, let us apply an arbitrary affine transformation $g$ that increases the volumes by $z/x_1$ (e.g., similarity with a coefficient $\sqrt[n]{z/x_1}$) to the ellipsoids $E_i$.

Then, $J_{\rm opt}(g(E_1)) = f_g(J_{\rm opt}(E_1)) = f_g(J_{\rm opt}(E_2)) = J_{\rm opt}(g(E_2))$ and therefore, $F(V(g(E_1)) = F(z) = F(V(E_2)) = F(kz)$.

In other words, the function $L(t) = F(\exp(t))$ has a period $\ln(k)$.

Let us show that this function also has a half-period $(1/2)\ln(z) = \ln(\sqrt{z})$. Indeed, if we had $F(z) < F(\sqrt{k} \cdot z)$, then we would be able to conclude that $F(\sqrt{k} \cdot z) < F(kz)$ and $F(z) < F(kz)$, which contradicts to what we have just proven. Similarly, $F(z) > F(\sqrt{k} \cdot z)$ is also impossible. Hence, $F(\sqrt{k} \cdot z) = F(z)$ for all $z$. Similarly, the function $L(t)$ has a period $\ln(k)/4$, $\ln(k)/8$, etc., i.e., $L(t) = \text{const}$. Hence, $F(x) = \text{const}$, which contradicts to our assumption that $J_{\rm opt}$ is not a constant.

This contradiction proves that $F(x)$ is 1-1. Q.E.D.

**Acknowledgments.** This work was partially supported by by NASA Research Grant No. 9-757, partially done when V. Kreinovich was an Invited Professor at Laforia, University of Paris VI, Summer 1996, and partially done during V.K.'s visit to the Institute of Mathematics and its Applications (IMA), August 1996. The authors want to thank J. Goutsias, R. Mahler, and H. T. Nguyen for the encouragement, and A. D. Alexandrov, Yu. F. Borisov, Yu. D. Burago, V. K. Ionin, I. A. Kunin, A. V. Kuz'minyh, and A. Neumaier for valuable discussions.

# References

[1] S. Abe and R. Thawonmas, *Fast training of a fuzzy classifier with ellipsoidal regions*, Proceedings of the 1996 IEEE International Conference on Fuzzy Systems, New Orleans, September 8–11, 1996, pp. 1875–1880.

[2] V.I. Arnold, *Mathematical methods of classical mechanics*, Springer, N.Y., 1978.

[3] O. Artbauer, *Application of interval, statistical, and fuzzy methods to the evaluation of measurements*, Metrologia, vol. 25 (1988), pp. 81–86.

[4] J.-P. Aubin and A. Cellina, *Differential inclusions*, Spinger-Verlag, Grundlehren der math. Wiss., vol. 264, 1984.

[5] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Birkhäuser, Boston, MA, 1990.

[6] G. Belforte and B. Bona, *An improved parameter identification algorithm for signal with unknown-but-bounded errors*, Proceeding of the 7th IFAC Symposium on Identification and Parameter Estimation, York, U.K., 1985.

[7] J. Binney, *Stellar dynamics*, in: I. Appenzeller, H. J. Habing, and P. Léna (eds.), *Evolution of galaxies: astronomical observations*, Springer Lecture Notes in Physics, vol. 333, Berlin, Heidelberg, 1989, pp. 95–146.

[8] H. Busemann, *The geometry of geodesics*, Academic Press, N.Y., 1955.

[9] F.L. Chernousko, *Estimation of the phase space of dynamic systems*, Nauka publ., Moscow, 1988 (in Russian).

[10] F.L. Chernousko, *State estimation for dynamic systems*, CRC Press, Boca Raton, FL, 1994.

[11] A.F. Filippov, *Ellipsoidal estimates for a solution of a system of differential equations*, Interval Computations, 1992, No. 2(4), pp. 6–17.

[12] A.M. Finkelstein, O. Kosheleva, and V. Kreinovich, *Astrogeometry: geometry explains shapes of celestial bodies*, University of Texas at El Paso, Department of Computer Science, Technical Report UTEP-CS-96-18, April 1996 (submitted to the *Geombinatorics* journal; the paper can be downloaded from the ftp site `cs.utep.edu`, login `anonymous`, directory `pub/reports`, file `tr96-18.tex`).

[13] E. Fogel and Y.F. Huang, *On the value of information in system identification. Bounded noise case*, Automatica, vol. 18 (1982), pp. 229–238.

[14] W.A. Fuller, *Measurement error models*, J. Wiley & Sons, New York, 1987.

[15] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, vol. 4 (1984), pp. 373–396.

[16] R.B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[17] L.G. Khachiyan, *A polynomial-time algorithm for linear programming*, Soviet Math. Dokl., vol. 20 (1979), No. 1, pp. 191–194.

[18] V. Kreinovich (ed.), *Reliable Computing*, 1995, Supplement (Extended Abstracts of APIC'95: International Workshop on Applications of Interval Computations, El Paso, TX, Febr. 23–25, 1995).

[19] *Reliable Computing*, International Journal, published by Kluwer since 1997.

[20] O.M. Kosheleva and V. Kreinovich, *Astrogeometry, or geometrical investigation of forms of celestial bodies*, Technical Report, Center for New Information Technology "Informatica", Leningrad, 1989 (in Russian).

[21] O.M. Kosheleva, V. Kreinovich, and A.M. Finkelstein, *Group-theoretic approach to foundations of space-time theory*, Proceedings of the Symposium on Global Geometry and Foundations of Relativity, Novosibirsk, 1982, pp. 76–78 (in Russian).

[22] V. Kreinovich, *Referee's comments in a review of* V.A. Dubrovin, S.P. Novikov, and A.T. Fomenko, Modern Geometry, *Moscow, Nauka, 1980*, Zentralblatt fur Mathematik, vol. 433 (1981), pp. 295–297.

[23] J.P. Norton, *Identification and application of bounded parameter models*, Proceeding of the 7th IFAC Symposium on Identification and Parameter Estimation, York, U.K., 1985.

[24] S. Rabinovich, *Measurement errors: theory and practice*, American Institute of Physics, N.Y., 1993.

[25] F.C. Schweppe, *Recursive state estimation: unknown but bounded errors and system inputs*, IEEE Transactions on Automatic Control, vol. 13 (1968), p. 22.

[26] F.C. Schweppe, *Uncertain dynamic systems*, Prentice Hall, Englewood Cliffs, NJ, 1973.

[27] S.T. Soltanov, *Asymptotic of the function of the outer estimation ellipsoid for a linear singularly perturbed controlled system*, In: S.P. Shary and Yu.I. Shokin (eds.), *Interval Analysis*, Krasnoyarsk, Academy of Sciences Computing Center, Technical Report No. 17, 1990, pp. 35–40 (in Russian).

[28] S.E. Strom and K.M. Strom, *The evolution of disk galaxies*, Scientific American, April 1979; reprinted in P.W. Hodge (ed.), *The Universe of galaxies*, Freeman and Co., N.Y., 1984, pp. 44–54.

[29] A. Toomre and J. Toomre, *Violent tides between galaxies*, Scientific American, December 1973; reprinted in P.W. Hodge (ed.), *The Universe of galaxies*, Freeman and Co., N.Y., 1984, pp. 55–65.

[30] R. Thom, *Structural stability and morphogenesis*, Benjamin Cummings, Reading, MA, 1975.

[31] G.S. Utyubaev, *On the ellipsoid method for a system of linear differential equations*, In: S.P. Shary (ed.), *Interval Analysis*, Krasnoyarsk, Academy of Sciences Computing Center, Technical Report No. 16, 1990, pp. 29–32 (in Russian).

[32] B.A. Vorontsov-Veliaminov, *Extragalactic astronomy*, Harwood Academic Publishers, Chur, Switzerland, London, 1987.

[33] Ya.B. Zeldovich and I.D. Novikov, *Relativistic Astrophysics. Part 2. The structure and evolution of the Universe*, The University of Chicago Press, Chicago and London, 1983.