# A Combination of Nonstandard Analysis and Geometry Theorem Proving, with Application to Newton's Principia

Jacques D. Fleuriot    Lawrence C. Paulson
Computer Laboratory – University of Cambridge
New Museums Site, Pembroke Street
Cambridge CB2 3QG
{jdf21,lcp}@cl.cam.ac.uk

January 1998

## Abstract

The theorem prover Isabelle is used to formalise and reproduce some of the styles of reasoning used by Newton in his **Principia**. The Principia's reasoning is resolutely geometric in nature but contains "infinitesimal" elements and the presence of motion that take it beyond the traditional boundaries of Euclidean Geometry. These present difficulties that prevent Newton's proofs from being mechanised using only the existing geometry theorem proving (GTP) techniques.

Using concepts from Robinson's Nonstandard Analysis (NSA) and a powerful geometric theory, we introduce the concept of an *infinitesimal geometry* in which quantities can be infinitely small or infinitesimal. We reveal and prove new properties of this geometry that only hold because infinitesimal elements are allowed and use them to prove lemmas and theorems from the Principia.

# Contents

# 1 Introduction

Isaac Newton's *Philosophiæ Naturalis Principia Mathematica* (Mathematical Principles of Natural Philosophy [1]), or *Principia* as it is usually called, was first published in 1687 and set much of the foundations of modern science. We now know that Newton's view of the world was only approximate but the laws and proofs he developed are still relevant and used in our everyday world. The elegance of the geometrical techniques used by Newton in the *Principia* is little known since demonstrations of most of the propositions set out in it are usually done using calculus.

In Newton's time, however, geometrical proofs were very much the norm. It follows that some of the lemmas of the *Principia* can be proved using just Euclidean geometry and we do so using our formalisation in Isabelle [2] of GTP rules proposed by Chou et al. [3, 4]. According to De Gandt [5] many of Newton's propositions and lemmas, however, do go beyond the boundaries of traditional Euclidean geometry in important respects such as the presence of motion and the admission of the infinitely small. Below we shall describe how we used the concept of the infinitesimal from Nonstandard Analysis (NSA) [7] to help formalise the notion of infinitely small geometric quantities.

Our initial aim is to study the geometric proofs of the *Principia* and investigate ways of mechanising them. We want to use some of the methods already developed for GTP in our own proofs. Moreover, we hope that some of Newton's reasoning procedures can be adapted to produce new methods in mechanised proofs of geometry theorems and in problem solving for classical mechanics. This work also hopes to expose some of the remarkable insights that Newton had in his use of geometry to prove propositions and solve problems.

In section 2 we briefly review the exposition of the *Principia* and the specific nature of its geometry. Section 3 gives an overview of the theory of infinitesimals from NSA that we formalised in Isabelle. Section 4 introduces our axiomatisation and use of parts of the area and full-angles methods first introduced by Chou et al. for automated GTP. We also have additional notions such as similar triangles and definitions of geometric elements such as ellipses and tangents. These are essential to our formalisation of Newton's work. In section 5 we describe some of the main results proved so far. Section 6 offers our comments on related work, conclusions and possible future work.

# 2 The Principia and its Mathematical Methods

The *Principia* is considered to be one of the greatest intellectual achievements in the history of exact science. It has, however, been influential for over three centuries rarely in the geometrical terms in which it was originally written but mostly in the analytico-algebraic form that was used very early to reproduce the work. Below we examine some of the original methods used in the *Principia*.

## 2.1 The Style and Reasoning of the Principia

Newton's reasoning rests on both his own methods and on geometric facts that though well known for his time (for example, propositions of Apollonius of Perga and of Archimedes) might not be easily accessible to modern readers. Moreover, the style of his proofs is notoriously convoluted due to the use of a repetitive, connected prose. Whiteside [6] notes the following:

> I do not deny that this hallowed ikon of scientific history is far from easy to read... we must suffer the crudities of the text as Newton resigned it to us when we seek to master the the *Principia's* complex mathematical content.

It is therefore one of our aims to show that we can use Isabelle to master some of the "complex mathematical content" of the work and give formal proofs of lemmas and propositions of Newton.

In the various figures used by Newton, some elements must be considered as "very small": for example, we encounter lines that are infinitely or indefinitely small or arcs that that may be nascent or evanescent. De Gandt argues that there is a temporal infinitesimal that acts as the independent variable in terms of which other magnitudes are expressed. However, since time itself is often represented geometrically using certain procedures, the infinitesimal time or "particle of time" in Newton's own expression appears as distance or area[1].

## 2.2 The Infinitesimal Geometry of the Principia

On reading the enunciations of many of the lemmas of the *Principia* one often comes across what Newton calls **ultimate** quantities or properties— for example, ultimate ratio (lemma 2,3,4 ...), ultimately vanishing angle (lemma 6), and ultimately similar triangles (lemma 8). Whenever Newton uses the term, he is referring to some "extreme" situation where, for example, one point might be about to coincide with another one thereby making the length of the line or arc between them vanishing, that is, infinitesimal.

Furthermore, as points move along arcs or curves, deformations of the diagrams usually take place; other geometric quantities that, at first sight, might not appear directly involved can start changing and as we reach the extreme situation new ultimate geometric properties usually emerge. We need to be able to capture these properties and reason about them. The use of infinitesimals allows us to "freeze" the diagram when such extreme conditions are reached: we introduce, for example, the notion of the distance between two points being infinitesimal, that is, infinitely close to zero and yet not zero when they are about to coincide. With this done, we can then deduce new or ultimate properties about angles between lines, areas of triangles, similarity of triangles and so on. This is what distinguishes our geometry from ordinary Euclidean geometry.

The infinitesimal aspects of the geometry give it an intuitive nature that seems to agree with the notions of infinitesimals from Nonstardard Analysis (NSA). Unlike Newton's reasoning, for which there are no formal rules of writing and manipulation, the intuitive infinitesimals have a formal basis in Robinson's NSA. This enables us to master motion, which is part of Newton's geometry, and consider the relations between geometric quantities when it really matters, that is, at the point when the relations are ultimate.

[1] See our exposition below of Proposition 1. Theorem 1 (Kepler's law of Equal Areas) for an example

# 3 Introducing the Infinitesimal

For a long time, the mathematical community has had a strong aversion to the notion of an infinitesimal (Bishop Berkeley [10] wrote a famous and vitriolic attack). This was historically due to the incorrect and informal use of infinitesimals in the development of the calculus. We are used to the powerful intuitions that infinitesimals can provide in constructing proofs of theorems but we are not allowed to use them in the proofs themselves (though physicists might disagree) without formal justification.

## 3.1 The Nonstandard Universe $\mathcal{R}^*$

NSA introduces the nonstandard real field $\mathcal{R}^*$, which is a proper extension of the complete ordered field of the reals $\mathcal{R}$. We give here the algebraic facts about infinitesimals, proved in Isabelle, that follow from the properties above and that we have used in our geometric proofs. Notions of finite and infinite numbers are also defined and we have proved many algebraic properties about them as well. We follow the definitions and mechanically-proved the theorems given in section 1A of Keisler [11].

**Definition 3.1** *In an ordered field extension $\mathcal{R}^* \supseteq \mathcal{R}$, an element $x \in \mathcal{R}^*$ is said to be an* ***infinitesimal*** *if $|x| < r$ for all positive $r \in \mathcal{R}$;* ***finite*** *if $|x| < r$ for some $r \in \mathcal{R}$;* ***infinite*** *if $|x| > r$ for all $r \in \mathcal{R}$.*

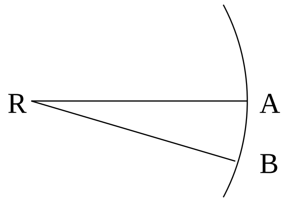
It is clear that $x \in \mathcal{R}^* \backslash \mathcal{R}$ or $x = 0$. This means that 0 is the only real infinitesimal and that other infinitesimals cannot be identified with any existing real numbers. We prove, in Isabelle, that the set `Infinitesimal` of infinitesimals and the set `Finite` of finite numbers are *subrings* of $\mathcal{R}^*$. Also, `Infinitesimal` is an *ideal* in `Finite`, that is the product of an infinitesimal element and a finite element is an infinitesimal.

**Definition 3.2** *$x, y \in \mathcal{R}^*$ are said to be* ***infinitely close*** *, $x \approx y$ if $|x - y|$ is infinitesimal.*

It is easily proved that $x$ is an infinitesimal if and only if $x \approx 0$ and that we have defined an equivalence relation above. Ballantyne and Bledsoe [12] observe:

> The relation $\approx$ now solves the problem Leibnitz encountered in that he was forced to make his infinitesimals simultaneously equal and different from zero. If one replaces the identity relation with the $\approx$ relation then all is well.

We can now formalise, for example, the idea of a point $B$ about to meet another point $A$ by saying that the distance between them, whether linear or curvilinear, is infinitesimal. We illustrate what we mean in the case of $B$ moving along a circular arc of finite radius of curvature and about to meet $A$:




The length of the circular arc $AB$ is infinitesimal and given that $\texttt{len}(R-A) \in \texttt{Finite}$, we can deduce $\langle A-R, R-B\rangle \in \texttt{Infinitesimal}$ since $\texttt{arc\_len}\, R\, A\, B = (|\texttt{len}(R-A)| \cdot \langle A-R, R-B\rangle) \in \texttt{Infinitesimal}$ and $\texttt{Infinitesimal}$ is an ideal in $\texttt{Finite}$.

The same reasoning can be applied if point $B$ is moving away from $A$, that is to the start of motion. Thus, we can deduce how various geometric quantities behave when we reach conditions that existing GTP techniques would consider degenerate since they are infinitesimal. Furthermore, as mentioned previously, geometric theorems and lemmas that hold at the infinitesimal level do not necessarily hold in general and we now have tools to prove them.

# 4 A Formalisation of Geometry in Isabelle

There exist efficient techniques for GTP— many of which, though extremely powerful, are highly algebraic [8]. These have been developed mostly for automated proofs, which are usually long and extremely difficult to understand. They consist mostly of a series of algebraic manipulations of polynomials that could not be farther from the the style of reasoning employed by Newton. Fortunately, there has been recent work in automated GTP by Chou et al. [3, 4] that aim to produce short, human-readable proofs in geometry using more traditional properties. We introduce a geometry theory in Isabelle based on some of the rules used in the algorithms for these new approaches.

## 4.1 The Geometric Methods

In these methods there are basic lemmas about geometric properties called signed areas and full-angles. Other rules are obtained by combining several of the basic ones to cover frequently-used cases and simplify the search process. We have assumed the basic rules as axioms and formally proved that the combined rules also hold.

We represent the line from point $A$ to point $B$ by $A$—$B$, its length by $\texttt{len}(A-B)$, and the *signed* area $\texttt{S}_{\texttt{delta}} ABC$ of a triangle is the usual notion of area with its sign depending on how the vertices are ordered. We follow the usual approach of having $\texttt{S}_{\texttt{delta}} ABC$ as positive if $A$—$B$—$C$ is in anti-clockwise direction and negative otherwise. Familiar geometric properties such as collinearity, $\texttt{coll}$, and parallelism, $\|$, can be thus

$$\begin{array}{rcl} \texttt{coll}\, a\, b\, c & \equiv & (\texttt{S}_{\texttt{delta}}\, a\, b\, c = 0) \\ a - b \parallel c - d & \equiv & (\texttt{S}_{\texttt{delta}}\, a\, b\, c = \texttt{S}_{\texttt{delta}}\, a\, b\, d) \\ \texttt{coll}\, a\, b\, c & \Longrightarrow & \texttt{len}(\mathsf{a}-\mathsf{b}) \times \texttt{S}_{\texttt{delta}}\, p\, b\, c = \texttt{len}(\mathsf{b}-\mathsf{c}) \times \texttt{S}_{\texttt{delta}}\, p\, a\, b \end{array}$$

A full angle $\langle u, v\rangle$ is the angle from line $u$ to line $v$. We note that $u$ and $v$ are lines rather than rays and define the relation of *angular* equality as follows:

$$x =_a y \equiv \exists n \in \texttt{Integer}.\ |x-y| = n\pi$$

The relation $=_a$ is an equivalence relation that is also used to express the properties that we might want. For example the idea of two lines being perpendicular becomes

$$a - b \perp c - d \equiv \langle a - b, c - d \rangle =_a \frac{\pi}{2}$$

Our aim is not to improve Chou's approaches to GTP, since they are essentially algorithmic and designed to perform automatic proofs. We have, however, provided a definition for the equality between full-angles. We can then easily prove that $\pi =_a 0$ and $\frac{3\pi}{2} =_a \frac{\pi}{2}$. Moreover, this enables us to combine the area and full-angles methods when carrying out our proofs and deduce, for example, $\langle a - b, b - c \rangle =_a 0 \Longleftrightarrow \mathtt{S_{delta}}\, a\, b\, c = \mathtt{0}$. We avoid the problems, such as $\pi = 0$, that would arise if we had used the ordinary equality for angles.

The attractive features of these approaches, as far as we are concerned, are the short, clear and diagram independent nature of the proofs they produce and that they deal easily and elegantly with ratios of segments, ratios of areas and angles and so on. These are the properties used in the geometry of the *Principia*
.

## 4.2 Infinitesimal Geometric Relations

Having introduced the basic geometric methods, we can now provide geometric definitions that make use of infinitesimals. Since we have explicitly defined the notion of equality between angles, we also need to define the idea of two angles being infinitely close to one another. We use the infinitely close relation to do so:

$$a_1 \approx_a a_2 \equiv \exists n \in \mathtt{Integer}.\ |a_1 - a_2| \approx n\pi$$

This is an equivalence relation. We prove in Isabelle the following property, which could provide an alternative definition for $\approx_a$:

$$a_1 \approx_a a_2 \Longleftrightarrow \exists \epsilon \in \mathtt{Infinitesimal}.\ a_1 =_a a_2 + \epsilon$$

Of course, we also have the theorem $a_1 \approx a_2 \Longrightarrow a_1 \approx_a a_2$. We now introduce a property that can be expressed using the concepts that we have developed so far in our theory— that of two triangles being *ultimately similar*. Recall that two triangles $\triangle abc$ and $\triangle a'b'c'$ are similar if they have equal angles at $a$ and $a'$, at $b$ and $b'$, and at $c$ and $c'$. The definition of ultimately similar triangles follows:

$$\begin{aligned} \mathtt{USIM}\ a\ b\ c\ a'\ b'\ c' \quad \equiv \quad & \langle b - a, a - c \rangle \approx_a \langle b' - a', a' - c' \rangle \ \wedge \\ & \langle a - c, c - b \rangle \approx_a \langle a' - c', c' - b' \rangle \ \wedge \\ & \langle c - b, b - a \rangle \approx_a \langle c' - b', b' - a' \rangle \end{aligned}$$

This property allows of treatment of triangles that are being deformed and tending towards similarity as points move in Newton's dynamic geometry. Elimination and introduction rules are developed to deal with the USIM property. It follows also, trivially, that $\mathtt{SIM\,a\,b\,c\,a'\,b'\,c'} \Longrightarrow \mathtt{USIM\,a\,b\,c\,a'\,b'\,c'}$. We also define the geometric relation of *ultimate congruence* UCONG and areas, angles and lengths can be made infinitesimal as needed when carrying out the proofs.

### 4.3 Other Geometric Definitions

The geometry theory contains other definitions and rules that are required for the proofs. These include length of arcs, length of chords and area of sectors. Since Newton deals with circular motion, and the paths of planets around the sun are elliptical, definitions for the circle, the ellipse and tangents to these figures are also provided. A few of these are given below.

$$
\begin{array}{rcl}
\texttt{circle}\, x\, r & \equiv & \{p.\, |\texttt{len}(x - p)| = r\} \\
\texttt{ellipse}\, f_1\, f_2\, r & \equiv & \{p.\, |\texttt{len}(f_1 - p)| + |\texttt{len}(f_2 - p)| = r\} \\
\texttt{e\_tangent}\, (a - b)\, f_1\, f_2\, E & \equiv & (\texttt{is\_ellipse}\, f_1\, f_2\, E \ \& \ a \in E \ \& \\
 & & \langle f_1 - a, a - b \rangle =_a \langle b - a, a - f_2 \rangle)
\end{array}
$$

The definition of the tangent to an ellipse relies on a nice property of the curve (which also provides an alternative definition): light emitted from one focus, say $f_1$ will reflect at some point $p$ on the ellipse to the other focus $f_2$. Thus, light reflects from the curve in exactly the same way as it would from the tangent line at $p$. Since the law of reflection means that the angle of incidence is the same as the angle of reflection, the definition above follows. The tangent line is important as it shows in the case where the ellipse is the path of an orbiting object, like a planet, the direction of motion of the object at that point.

## 5 Mechanised Propositions and Lemmas

Some of the results obtained through the mechanisation of the ideas discussed above can now be presented. The methods have been used to investigate the nature of the infinitesimal geometry and formally prove many geometry theorems. Some of these results confirm what one intuitively might expect to hold when elements are allowed to be infinitesimal.

### 5.1 Motion Along an Arc of Finite Curvature

Consider the following figure based on diagrams Newton constructed for proofs of his lemmas; let $A - D''$ denote the tangent to the arc $ACB$ and $A - R$ the normal to the tangent, both at $A$. Let $R$ be the centre of curvature, $r$ be the antipodal of the circle of contact at $A$, the points $D$, $D'$ and $D''$ be collinear and $B - D \perp A - D''$. With the point $B$ moving towards $A$ along the arc, we can prove several properties about this diagram, including some which become possible because we have infinitesimals:

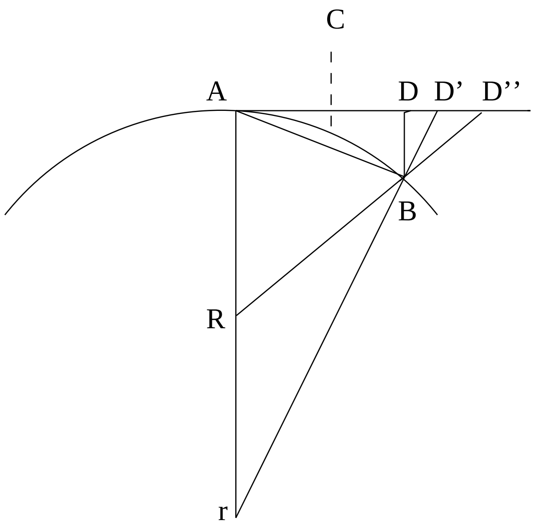



Figure 1: Point $B$ moving along an arc towards $A$

- $\triangle BDA$ and $\triangle ABr$ are **similar** and hence $\texttt{len}(A-B)^2 = \texttt{len}(A-r) \times \texttt{len}(D-B)$. The latter result is stated and used but not proved by Newton in Lemma 11.
- $\triangle ABD'$ and $\triangle rAD'$ are similar and hence $\texttt{len}(D'-A)^2 = \texttt{len}(D'-B) \times \texttt{len}(D'-r)$.
- $\langle B-A, A-D''\rangle =_a \langle B-R, R-A\rangle/2$
- $\triangle ABR$ and $\triangle AD''R$ are **ultimately similar** since the angle $\langle B-A, A-D''\rangle$ is infinitesimal when point $B$ is about to coincide with point $A$. This proves part of Lemma 8.
- the **ultimate ratio** of $\texttt{len}(A-B)$, $\texttt{arc\_len}\ RAB$, and $\texttt{len}(A-D'')$ is infinitely close to 1. This is Lemma 7.

It is clear from the diagram above that $\triangle ABR$ and $\triangle AD''R$ are not similar in ordinary Euclidean Geometry. Infinitesimals notions reveal that we are tending towards similarity of these triangles when the point $B$ is about to meet point $A$. This property cannot be deduced from just the static diagram above. The dynamics of Newton's geometry involves the reader using his or her imagination to incorporate motion and see what is happening to the relations between various parts of the diagram as points are moving. This task is not always trivial. The relation $\texttt{USIM}\,\texttt{A}\,\texttt{B}\,\texttt{R}\,\texttt{A}\,\texttt{D}''\,\texttt{R}$ can be illustrated by considering the relation between the parts of the diagram as point $B$ moves towards point $A$:

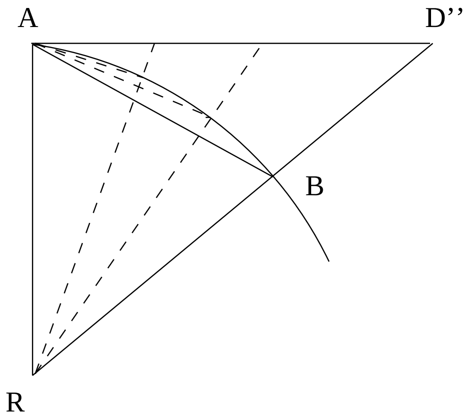



Figure 2: Ultimately similar triangles

We can see as $B$ moves towards $A$ that $\langle B - A, A - D'' \rangle$ is decreasing, as one might intuitively expect, and when ultimately the distance between $B$ and $A$ is infinitely close to zero, then we have that $\langle R - A, A - B \rangle \approx_a \langle R - A, A - D'' \rangle$. From this we can deduce the ultimate similarity of the triangles as required.

We give below a detailed overview of our reasoning and theorems proved to show the ultimate similarity of these two triangles. We show that the angle subtended by the arc becomes infinitesimal as $B$ approaches $A$ and that the angle between the chord and the tangent is always half the that angle:

```
[| arc_len R A B ≈ 0; len (A--R) ∈ Finite - Infinitesimal |]
⟹ <B--R,R--A> ≈ 0

[| c_tangent A D'' R Circle; B ∈ Circle |]
⟹ <B--A,A--D''> =_a <B--R,R--A>/2
```

We use the theorem from NSA that `Infinitesimal` form an ideal in `Finite` and the results above to prove that the angle between the chord and the tangent becomes infinitely close to zero and that $\langle R - A, A - D \rangle$ and $\langle R - A, A - B \rangle$ are infinitely close:

```
[| <B--R,R--A>∈ Infinitesimal; 1/2∈ Finite |]
⟹<B--R,R--A>*1/2 ∈ Infinitesimal

[| c_tangent A D'' R Circle; B ∈ Circle; arc_len R A B ≈ 0;
   len (A--R) ∈ Finite - Infinitesimal |]
⟹ <B--A,A--D''> ≈_a 0";

<B--A,A--D''> ≈_a 0 ⟹ <R--A,A--D''> ≈_a <R--A,A--B>
```

Finally, since $\triangle ABR$ and $\triangle AD''R$ have two corresponding angles that are infinitely close (they have one common vertex in fact), we can show that they are ultimately similar:

```
<B--R,R--A> =_a <B--R,R--A> ⟹ <B--R,R--A> ≈_a <B--R,R--A>

[|<B--R,R--A> ≈_a <B--R,R--A>; <R--A,A--D''> ≈_a <R--A,A--B> |]
⟹ USIM A B R A D'' R
```

## 5.2 Kepler's Law of Equal Areas

Kepler's equal area law was published in 1609 and was often regarded until Newton's *Principia* as one of the least important of Kepler's Laws. This law is established by Newton as the first mathematical Proposition of the *Principia.*

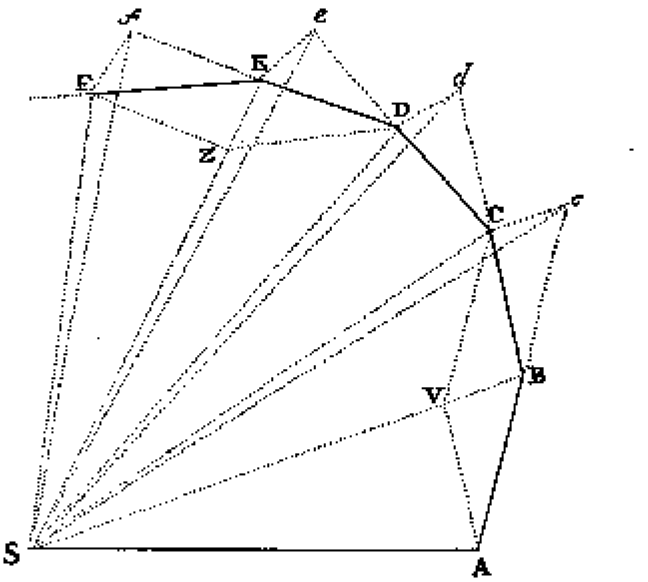



Figure 3: Original diagram from the *Principia* showing a body moving under the influence of a series of impulsive centripetal forces

In Newton's diagram (Fig. 3), the polygons $ABCDEF$ are used to approximate the continuous motion of a planet in its orbit. The motion between any two points such as $A$ and $B$ of the path is not influenced by any force, though there are impulsive forces, all directed towards the fixed centre $S$, that act at $A, B, C, \ldots$. Newton proved that if the time interval between successive impulses is fixed then all the triangular areas $SAB, SBC, \ldots$, are equal, that is equal areas are described in equal times. The demonstration of this law makes no assumption about how this force varies with distance from the centre of force $S$; its only restriction is that it be directed toward $S$. Newton reduces the discontinuous motion along the straight edges $AB, BC, \ldots$, to continuous motion along a smooth orbital path by using an infinitesimal process that lets the size of the triangles become infinitely small.

We follow Newton's argument and prove that the area of $SAB$ is equal to that of $SBC$ using our geometric tools. We quote from the exposition of Proposition 1 in the *Principia*:

> Let time be divided into equal parts, and in the first part of the time let the body, by its inherent force, describe the straight line $AB$. In the second part of the time, the same body, if nothing were to impede it, would pass on by means of a straight line to $c$ (by Law 1), describing the line $Bc$ equal to $AB$, with the result that, radii $AS$, $BS$, $cS$ being drawn to the centre, the areas $ASB$, $BSc$ would

come out equal.

We first observe that the area of $SAB$ equals that of $SBc$ because the triangles have equal bases (since the times are equal and no force has acted to change the velocity) and the same height:

```
[| coll A B C; len(A--B) = len(B--c)|] ⟹ S_delta S A B = S_delta S B c
```

The impulsive centripetal force at B makes the body depart from motion in a straight line and Newton makes the following construction (using the Parallelogram Law of Forces):

> Let $cC$ be drawn parallel to $BS$, meeting $BC$ at $C$; and, the second part of the time being completed, the body (by Corollary I of the laws) will be located at C, in the same plane as the triangle ASB ... Connect $SC$, and because of the parallels $SB$,$Cc$, triangle $SBC$ will be equal to triangle $SBc$, and therefore to triangle $SAB$.

This leads to the following lemma, which is also easily proved in Isabelle since it follows from the definition of parallel lines:

```
[| S--B || c--C |] ⟹ S_delta S B c = S_delta S B C
```

The proof that the areas are equal follows. In fact, this first part of the proof of Kepler's Law of Equal Areas is proved automatically in one step by Isabelle thanks to the presence of powerful proof tactics.

The next step is to decrease the breadth of the triangles to be infinitesimally small and by lemma 3 and its corollaries we can substitute the straight edge by a curved line:

```
<A--S,S--B> ≈ 0 ⟹ len(A--B) ≈ arc_len S A B
```

And furthermore using the same lemma, the area of the infinitesimal triangle $SAB$ is infinitely close to the area of the arc and can be substituted:

```
<A--S,S--B> ≈ 0 ⟹ S_delta S A B ≈ arc_area S A B
```

As the triangles become infinitesimal, the perimeter of the path becomes infinitely close to a curvilinear one and the force can be viewed as acting continuously since the times between the impulses are infinitesimal. We can note here the geometrical representation of time since making the triangles infinitesimal effectively makes the the time intervals also infinitely close to zero. The result that the area described is proportional to the time still holds for the evanescent triangles and hence also holds for the infinitely close curvilinear areas.

# 6 Related Work and Final Comments

The combination of concepts used in this approach relates it to work that has already been done in the field of NSA and GTP.

## 6.1 Nonstandard Analysis Theorem Proving

The theorem proving community does not seem to have shown much interest in NSA even though its importance has grown in many fields such as physics, analysis and economics, where it has successfully been applied. Ballantyne and Bledsoe [12] implemented a prover using nonstandard techniques in the late seventies. Their work basically involved substituting any theorem in the reals $\mathcal{R}$ by its analogous in the extended reals $\mathcal{R}^*$ and proving it in this new setting. Even though the prover had many limitations and the work was just a preliminary investigation, the authors argued that through the use of nonstandard analysis they had brought some new and powerful mathematical techniques to bear on the problem.

Despite this rather promising work, there does not seem to have been much done over the last two decades. Suppes and Chuaqui [13] have proposed a framework for doing proofs in NSA and Bedrax has implemented a prototype for a simplified version of Suppes-Chuaqui system called Infmal [14]. Infmal is implemented in Common Lisp and contains the various axioms (logical, algebraic and infinitesimal) required by the deduction system and extensions to the usual arithmetic operations. Unfortunately, Infmal is a simple experiment and though interactive is rather limited in the proofs it can carry out.

The parts of this work relating to NSA have used the definitions provided in 3.1 and have proved most of the facts about NSA built in Ballantyne and Bledsoe's prover. There are proofs involving *standard parts*, infinites, for example, that have not been described in this paper.

## 6.2 Automated Geometry Theorem Proving

As mentioned already, the geometric methods based on signed areas and full-angles have been useful to the development of the geometry theory in Isabelle. The work of Chou et al. is intended to produce short and readable proofs of difficult geometry theorems. This represents a return to the original and more traditional ways of geometry theorem proving that had been superseded by the more powerful *algebraic* methods based mainly on Wu's characteristic set method [8] and the Gröbner basis method [15]. Unfortunately there are several drawbacks associated with the use of these algebraic techniques: they are computationally intensive and produce long proofs that do not have clear geometric meaning since they are manipulations of polynomials obtained by coordinatisation of points in the diagrams.

The recent work improves on various previous attempts to use geometrically meaningful properties for theorem proving through the use of geometric invariants, that is, areas, full-angles etc. In our work, techniques such as those of similar and congruent triangles that are needed for the proofs are also added as traditionally these have also been used in proofs of geometry. Though Chou et al. [3] note that they have limitations, our proofs are not affected since we are not concerned with completely automatic proofs. The resulting geometry theory of Isabelle is powerful and able to prove most of the results that the signed area and full-angles methods can tackle. Moreover, many of Chou's lemmas [3, 4] obtained through the combination of various rules and used to help the automatic search have been verified in Isabelle.

### 6.3 Problem Solving in Mechanics

There has been some interest in the past in problem solving in mechanics in the GTP community where, interestingly, Wu [9] algebraically proved Newton's Laws of Gravitation and even, with somewhat more difficulty, automatically derived them from Kepler's Laws (which Newton actually proved in the *Principia*).

There is also Novak, who implemented several systems to do problem solving in classical physics with the help of diagrams [16]. Though his work falls into the field of diagrammatic reasoning rather that GTP it does require the implicit applications of geometry theorems to derive relations between various physical quantities represented geometrically in the diagrams. This work also shows that it is possible to closely relate physical and geometric principles through diagrams.

### 6.4 Conclusions and Future Work

Reading the *Principia* and making sense of the reasoning of Newton is a difficult but rewarding task. As is common with proofs using geometric tools, once the hard task of constructing the diagram and proof is done, the result that follows usually looks simple and intuitive. We have shown that, though Newton does not provide a set of rules for carrying out his proofs, the reasoning is formal and can be mechanised. We can effectively give formal definitions and proofs of some of the ultimate properties Newton is trying to prove. We have tried to bridge the gap between intuition and formality.

Furthermore, the introduction of infinitesimal elements in the geometry is an exciting aspect that can lead to the discovery of interesting properties that cannot be seen ordinarily. We have scope for more work in the field of NSA, whose foundations in Isabelle we plan to investigate in more depth in the near future. Infinitesimals are always tricky and can lead to paradoxes if not used carefully.

We plan to mechanise other interesting propositions from the *Principia* and, since these are actually proofs in classical mechanics, it would be interesting to see how these techniques can be applied to problem solving in the field. Physics textbooks commonly use infinitesimals in informal reasoning, because it is intuitive. This material could also be mechanised.

## References


[1] I. Newton. *The Mathematical Principles of Natural Philosophy*. Third edition, 1726. Translation by A. Motte (1729). Revised by F. Cajory 1934. University of California Press.

[2] Lawrence C. Paulson. *Isabelle: A Generic Theorem Prover*. Springer, 1994. LNCS 828.

[3] S. C. Chou, X. S. Gao, and J. Z. Zhang. Automated generation of readable proofs with geometric invariants, I. Multiple and Shortest proof generation. *J. Automated Reasoning* **17** (1996), 325–347.

[4] S. C. Chou, X. S. Gao, and J. Z. Zhang. Automated generation of readable proofs with geometric invariants, II. Theorem proving with full-angles. *J. Automated Reasoning* **17** (1996), 349–370.

[5] F. De Gandt. *Force and Geometry in Newton's* Principia . Princeton University Press, Princeton, New Jersey, 1995.

[6] D. T. Whiteside. The Mathematical Principles underlying Newton's *Principia Mathematica* . Glasgow University Publication 138, 1970.

[7] A. Robinson. *Non-Standard Analysis*. North-Holland Publishing Company, reprinted 1980. 1966, first edition.

[8] Wu Wen-tsün. On the Decision Problem and the Mechanization of Theorem in Elementary Geometry. *Automated Theorem Proving: After 25 years*. A.M.S., Contemporary Mathematics, **29**(1984), 213–234.

[9] Wu Wen-tsün. Mechanical Theorem Proving of Differential Geometries and Some of its Applications in Mechanics. *J. Automated Reasoning* **7**(1991), 171–191.

[10] G. Berkeley. The Analyst: A discourse addressed to an infidel mathematician. *The World of Mathematics*, volume 1, London: Allen and Unwin(1956), 288–293.

[11] H. J. Keisler. *Foundations of Infinitesimal Calculus*. Prindle, Weber & Schmidt, Incorporated, 20 Newbury Street, Boston, Massachusetts, 1976.

[12] A. M. Ballantyne and W. W. Bledsoe. Automatic Proofs of Theorems in Analysis Using Nonstandard Analysis. *J. of the Association of Computing Machinery*, Vol. 24, No. 3, July 1977, 353—374.

[13] R. Chuaqui and P. Suppes. Free-Variable Axiomatic Foundations of Infinitesimal Analysis: A Fragment with Finitary Consistency Proof. *The Journal of Symbolic Logic*, Volume 60, Number 1, March 1995.

[14] T. Bedrax. *Infmal*: Prototype of an Interactive Theorem Prover based on Infinitesimal Analysis. Liciendo en Mathematica con Mencion en Computation Thesis. Pontifica Universidad Catolica de Chile, Santiago, Chile, 1993.

[15] D. Kapur. Geometry Theorem Proving using Hilbert's Nullstellensatz. Proceedings of SYMSAC'86, Waterloo, 1986, 202—208.

[16] S Novak Jr. Diagrams for Solving Physical Problems. *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, AAAI Press / MIT Press, 753—774, 1995. (Eds. Janice Glasgow, N. Hari Narayana, and B. Chandrasekaram).