AI Noether - Bridging the Gap Between Scientific Laws Derived by AI Systems and Canonical Knowledge via Abductive Inference

Karan Srivastava
University of Wisconsin-Madison || IBM Research
Madison, WI || Yorktown Heights, NY
ksrivastava4@wisc.edu

Sanjeeb Dash
IBM Research
Yorktown Heights, NY
sanjeebd@us.ibm.com

Ryan Cory-Wright
Imperial Business School
London, UK
r.cory-wright@imperial.ac.uk

Barry Trager
IBM Research
Yorktown Heights, NY
bmt@us.ibm.com

Lior Horesh
IBM Research
Yorktown Heights, NY
lhoresh@us.ibm.com

September 8, 2025

Abstract

A core goal in modern science is to harness recent advances in AI and computer processing to automate and accelerate the scientific method. Symbolic regression can fit interpretable models to data, but these models often sit outside established theory. Recent systems (e.g., AI Descartes, AI Hilbert) enforce derivability from prior axioms. However, sometimes new data and associated hypotheses derived from data are not consistent with existing theory because the existing theory is incomplete or incorrect. Automating abductive inference to close this gap remains open. We propose a solution: an algebraic geometry-based system that, given an incomplete axiom system and a hypothesis that it cannot explain, automatically generates a minimal set of missing axioms that suffices to derive the axiom, as long as axioms and hypotheses are expressible as polynomial equations. We formally establish necessary and sufficient conditions for the successful retrieval of such axioms. We illustrate the efficacy of our approach by demonstrating its ability to explain Kepler's third law and a few other laws, even when key axioms are absent.

1 Introduction

Generating hypotheses that explain natural phenomena and are consistent with existing knowledge is a critical component of the scientific discovery process [25]. Since the big data revolution, many machine learning techniques have been proposed to enhance scientific discovery [42, 23, 24, 14, 45, 38]. However, it is often unclear whether formulae induced from data generalize to unseen data or merely fit the known data well. Even when the formulae derived by these procedures (e.g., the output of a complicated neural network) are highly predictive, they cannot easily be understood or explained by a human, which makes integrating them within canonical scientific knowledge challenging. This is arguably the key difference between machine learning and scientific discovery: machine learning focuses on making accurate predictions, while discovery focuses on making interpretable predictions that drive understanding [see 22, for a review of the similarities and differences].

While there have been notable successes in applying deep learning to scientific discovery [2, 13], autonomous chemical research [5], and deriving solutions to mathematical problems [1], its poor interpretability [27, 40] and inability to ground answers in background knowledge [30] has led to the widespread adoption of more interpretable methods in scientific discovery settings. Several authors have applied symbolic regression [18, 28] and sparse regression [6, 4] to discover boundary equations in plasma physics [31] and new equations in psychology [37]. More recently, there has been progress in integrating data and background theory within the scientific discovery process. Indeed, several software systems, such as AI Descartes [8] and AI Hilbert [9], aim to find formulae that not only align with experimental data but also are consistent with established background theory.

However, scientific formulae that best explain experimental data are sometimes not fully substantiated by existing theory, because existing theory is at times incorrect or incomplete (or both, analogous to Gödel's incompleteness theorems [17]) and may need to be revised to explain new phenomena or data. For instance, Einstein modified certain aspects of Newtonian theory and postulated that the speed of light is a constant in the process of developing his special theory of relativity [7]. A second example of this is Van der Waals, who modified the ideal gas laws to get a more accurate theory of real gases [16]. Third, an unresolved gap between background theory and data arises in the so-called ”Lithium problem” in astrophysics: the most widely accepted models of the Big Bang suggest that three times more Lithium should exist than the quantity observed in experimental data [15]. This inconsistency between data and theory makes it challenging to integrate derived formulae with existing theory. Accordingly, abductive reasoning, or deriving an explanation for why a new formula is valid, is often of as much scientific interest as a discovery.

Modifying a body of theory to explain a derived formula that fits new data is a distinct and often more challenging task than identifying such a formula in the first place. This is because many of the quantities in a given theory are not ”measured quantities” in available observational data. For instance, the mass-energy density of the early universe, as described in the Big Bang Theory, cannot be directly measured at present [43]. In celestial mechanics, quantities such as gravitational forces between two distant celestial objects are not directly measurable [46]. In situations where existing theory does not explain observational data, augmenting or modifying the known theory in such a way that it explains the data is a challenging task. As discussed above, any modifications involving unmeasured variables cannot be directly derived from the data; instead, one needs to verify whether the patterns visible in the observational data logically follow from a modified theory.

In this paper, we propose a formal abductive reasoning procedure for reconciling derived hypotheses or formulae with canonical science. Given a set of background theory axioms and a hypothesized formula (possibly obtained from experimental data), we propose a mechanism to generate new axioms that either replace or augment the existing axiom system in a way that the new axiom system explains the hypothesis. We provide a means not only of making new scientific discoveries but also of explaining them in terms of the modifications that need to be made to the theory to derive them. We focus on the case where the axiom system and the formulae to be explained can be expressed as polynomial equations. We name our system `AI-Noether` inspired by the work of Emmy Noether, one of the leading mathematicians of her time who made significant contributions to both algebraic geometry and physics [32].

To the best of our knowledge, this is the first work to study the problem of abductive reasoning in the context of scientific discovery, where the explanations are allowed to consist of fully general polynomial axioms. See [38] for a recent review of AI-driven approaches to scientific discovery. This constitutes a step toward the larger goal of integrating machine learning and data into the entire scientific method, thereby accelerating it. We review the most relevant literature from scientific discovery and real algebraic geometry below.

1.1 Literature Review

Abductive reasoning is a field of inference that generates plausible explanations from incomplete observations [36, 34]. Unlike deductive reasoning, which obtains a conclusion from a set of premises, abductive reasoning starts with an outcome and ends with a set of plausible explanations, e.g.,

- The fact C is observed.
- If A were true then C would be true.
- There is reason to believe that A is true.

As observed by [36], abductive reasoning is arguably “the only logical operation which introduces any new ideas”. Moreover, it is often applied by humans in everyday situations, such as reading between the lines [33] and counterfactual reasoning [35]. Doctors use abductive reasoning to propose possible diseases that cause observed symptoms. It has been argued that automating inference may be necessary for understanding and automating consciousness [3]. Yet, abductive reasoning has not yet been incorporated into state-of-the-art AI systems, such as AI-Feynman, AI-Descartes, or AI-Hilbert [42, 8, 9]. Thus, automating abductive inference is an important component of automating the scientific method.

Our work is related to three parts of the abductive inference and scientific discovery literature: (i) automated abductive inference in logic and more "classical" AI systems, (ii) applications of abductive inference in machine learning, and (iii) machine learning and AI for scientific discovery, as already reviewed in the introduction. Accordingly, we now review the remaining two parts of this literature.

Automated Abductive Inference in Logic and Classical AI: Marquis [29] proposed automating abductive reasoning in propositional logic and generalized it to first-order logic. Beyond logic, several papers have investigated abductive inference in the context of scientific discovery, e.g., ARROWSMITH for cross-pollinating seemingly unrelated branches of science [41]; see [26] for a review of the automation of components of the scientific discovery process, including abductive inference. We follow this line of work in interpreting abduction as the process of selecting or synthesizing hypotheses that complete an incomplete theory in order to explain a target hypothesis.

Automated Abductive Inference for Machine Learning: The recent surge in popularity of black-box models, including large language models, has led to an interest in explaining the predictions of these models. LIME [39] and similar methods provide (local) interpretable explanations that describe the local behavior of any ML model using a weighted linear combination (often sparse) of input features; see [40, 27] for reviews of interpretable and explainable machine learning. However, existing works on interpretability and explainability generally focus on computing explanations for black-box models, rather than generating theories that could be used for scientific discovery. In particular, techniques like LIME or SHAP scores cannot, to our knowledge, be leveraged to establish missing axioms in a scientific discovery context.

More recently, a new line of work has explored abductive inference for machine learning [12, 21, 48, 20, 47]; see also [44] for a recent review of this line of work. Initiated by [12, 21], the key idea is to embed purely logical reasoning models within the machine learning process to take advantage of the benefits of both machine learning and logical reasoning. For instance, in [21], the authors use abductive inference to generate explanations of ML models given input background theories and a set of hypothetical explanations. Moreover, [12] combines machine learning techniques with a first-order logic model to obtain an output that depends on both models. However, these works do not apply abductive inference to scientific discovery.

2 Abductive inference

We assume that all axioms (missing or otherwise) are expressible as polynomial equations of the form $p(x) = 0$, and the hypothesis can also be expressed as a polynomial equation. We say that a collection of polynomial equations of the form $p_1(x) = 0, \ldots, p_k(x) = 0$ imply the polynomial equation $q(x) = 0$ if the latter equation holds for all solutions $\bar{x}$ of the former system of polynomials.

The hypothesis generation step of machine-assisted scientific discovery involves generating a mathematical expression of unknown form that explains potentially noisy data [42, 11]. Recently, the discovery process has been augmented with the use of background axioms that represent a complete, known physical theory, which constrains the space of candidate hypotheses that explain the data [8, 9]. Data-driven approaches often generate formulae that are not necessarily consistent with theory, while data-theory augmented approaches only return certificates of derivability when background axioms contain sufficient information to explain a hypothesis. In practice, however, theory may be incomplete or inconsistent, and theory needs to be corrected or augmented to explain observational data. We therefore consider the following problem.

Given axioms $A_1, ..., A_k$ and a hypothesis Q such that $A_1, ..., A_k$ do not imply Q, find a set of candidate axioms $A_{k+1}, ..., A_r$ with the "smallest residual" with respect to $A_1, ..., A_k$ such that $A_1, ..., A_r$ imply Q.

Terminology: Throughout the rest of this work, we use the terms $A_1, \ldots, A_k$ "explain" or "imply" Q and Q is a "consequence of" $A_1, \ldots, A_k$ synonymously. Intuitively, the notion of "smallest residual" corresponds to the idea that we wish to add as little new information as possible to $A_1, \ldots, A_k$ to explain Q; we will formalize this notion later on. We assume that we do not have data for

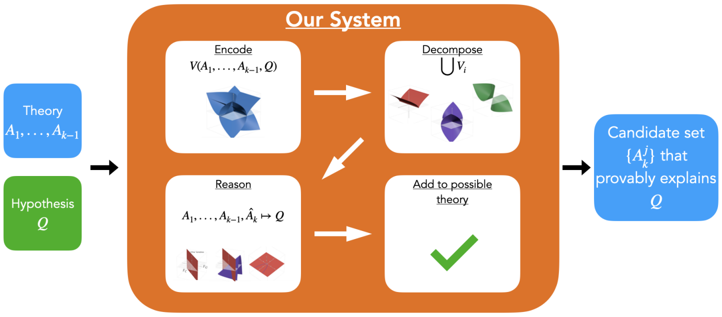

Figure 1: *AI Noether*: Encode—form $I = \langle A_1, \ldots, A_k, Q\rangle$; Decompose—primary decomposition with primes P_j; Reason—test generators $g \in P_j$ by augmenting and eliminating to $\text{vars}(Q)$; passing g are candidate A_{k+1}.

A_{k+1}, ..., A_r; otherwise, the question is equivalent to the hypothesis generation question. We further assume that we do not know what variables A_{k+1}, ..., A_r are defined over, only that they are some subset of all variables in the entire system. Under this setting, our system generates a set of candidate axioms $\{q_1, ..., q_t\}$ such that the known axioms along with q_i explain Q. Further, if no single candidate axiom can be added to explain Q, our system detects this. Current machine-assisted discovery tools either rely on partial or complete background theory, or no background theory at all, to discover mathematical expressions that fit observational data, but few append or correct the theory in an automated way, and none do so at the same level of generality.

Example 1. *Assume we have an axiom system defined over 5 real quantities, indicated by* x, y, z, a, b*:*

$$x^2 - ay^2 = 0,$$

$$by - z = 0.$$

In this example, we think of a, b *as constants in the system (similar in spirit to the Gravitational constant), while* x, y, z *are quantities that can vary. Let* p_1 *be the polynomial* $x^2 - ay^2$ *and let* $p_2 = by - z$*. Thus the axioms are* $p_1 = 0$ *and* $p_2 = 0$*. Let*

$$q = b^2x^2 - az^2.$$

The equation $q = 0$ *follows from the axioms as*

$$q = b^2x^2 - az^2 = b^2(x^2 - ay^2) + a(by + z)(by - z) = b^2p_1 + a(by + z)p_2. \tag{1}$$

In the example above, we have expressed q as $\alpha p_1 + \beta p_2$ where $\alpha = b^2$ and $\beta = a(by + z)$ are polynomials defined over x, y, z, a, b. We say that q is an *algebraic combination* of the polynomials p_1 and p_2 and say that $q = 0$ is *derivable* from $p_1 = 0$ and $p_2 = 0$. As $q = \alpha p_1 + \beta p_2$, it follows that $q = 0$ for all values of x, y, z, a, b such that $p_1 = 0$ and $p_2 = 0$. We therefore say that the axioms in the example imply or explain $q = 0$. See the appendix for more on this example, including alternate theories that explain $q = 0$.

Using the notation above, "derivable from" implies "can be explained by". We will later discuss conditions under which the converse is true.

2.1 Method

Our goal is to identify a candidate missing polynomial axiom A_{k+1} that, along with a known set of axioms $A_1, \ldots, A_k$, can be used to derive a target hypothesis Q. The method consists of three main steps—**Encode**, **Decompose**, and **Reason**—illustrated in Figure 1 and summarized in Algorithm 1.

Encode. Given a collection of known axioms $A_1, \ldots, A_k$ and a hypothesis Q, we form the solution set of the system $\{A_1, \ldots, A_k, Q\}$. Each equation defines a hypersurface in $\mathbb{R}^n$, and the intersection of these hypersurfaces defines a geometric object—referred to as the variety of the system. We will use this geometric perspective to recast the ideas of Q being derivable from $A_1, ...A_k$ as well as Q being derivable from some larger set of axioms $A_1, ..., A_{k+r}$ as geometric manipulations.

Decompose. The set of solutions (variety) of the equation $xy = 0$ (where x and y are real variables) is not irreducible - i.e., that is, it is the union of two smaller solution sets (varieties), i.e., solutions of

$x = 0$ and of $y = 0$. We similarly decompose the solution set of the system $\{A_1, \ldots, A_k, Q\}$ into smaller components representing minimal geometric subsets (with some properties that will be made precise shortly) that cannot be decomposed further. When Q is not a consequence of the axioms (and thus not derivable from the axioms), some components in the variety of $\{A_1, \ldots, A_k, Q\}$ will not be present in the variety of $\{A_1, \ldots, A_k\}$. These residual components capture algebraic constraints that make Q hold locally but are not globally entailed by the axioms. For example, if $Q = \alpha_1 A_1 + \cdots + \alpha_{k+1} A_{k+1}$ for unknown coefficients but A_{k+1} is missing, then the solution set may decompose into components corresponding to the missing axiom. Decomposition isolates these structures.

Reason. For each irreducible component, we extract its defining polynomials and test whether any subset, when added to the axioms, is sufficient to derive Q. This is done by forming the solution set of $\{A_1, \ldots, A_k\} \cup \{g\}$ for a candidate polynomial g, and projecting it onto the variables appearing in Q. If this projection lies within the solution set of Q, then Q is implied by the augmented system. Such g are returned as candidate missing axioms. If no such g exists, we conclude that no single missing axiom explains Q.

Algorithm 1 High-Level Summary: Identifying Explanatory Axioms

```
1: Input: Axioms $A_1, \ldots, A_k$ and hypothesis $Q$
2: Encode: Construct the solution set of $\{A_1, \ldots, A_k, Q\}$
3: Decompose: Break the solution set into nondecomposable components
4: for each component do
       • Extract defining polynomials (also called generators)
       • Reason: Test whether adding any generator allows $Q$ to be derived from the axioms
5: end for
6: Output: Set of candidate missing axioms sufficient to explain $Q$
```

This method does not require access to data for the missing axiom or prior knowledge of the variables it depends on. It returns a set of explanatory candidates when possible, or certifies failure when no such single axiom exists. We develop the algebraic details in the following section.

2.2 Preliminaries

We define notions specific to real algebraic geometry; see [10] for further details. Let $\mathbb{R}$ denote the set of real numbers and let $R = \mathbb{R}[x_1, ..., x_n]$ be the ring of polynomials defined over the n variables $x_1, \ldots, x_n$ with real coefficients. Axioms and hypotheses in our model will be equations of the form $p(x) = 0$ where $p \in R$. We abuse notation and identify an axiom by its defining polynomial. We will focus on real-valued solutions of axiom systems; see the appendix for more details.

Encode. An *ideal* over R is a subset I of R that is closed under addition and also multiplication by elements of R: if $f, g \in I$, then $f + g \in I$ and $\alpha f \in I$ for any $\alpha \in R$. Given polynomials $A_1, ..., A_k \in R$, we denote the set of all algebraic combinations of $A_1, \ldots, A_k$ by

$$\langle A_1, \ldots, A_k \rangle = \{f \in R : f = \sum_{i=1}^{k} \alpha_i A_i \text{ for some } \alpha_i \in R\}.$$

called an ideal over R with *generators* $A_1, \ldots, A_k$. Let $\mathcal{A}$ stand for $\{A_1, \ldots, A_k\}$, and let $I(\mathcal{A})$ stand for the ideal $\langle A_1, \ldots, A_k \rangle$. Our earlier notion of "derivability" corresponds to ideal membership: $Q = 0$ is derivable from $\mathcal{A}$ if Q is an algebraic combination of $A_1, \ldots, A_k$, i.e., if $Q \in I(\mathcal{A})$. Given an ideal I, we define the *variety* of I, denoted $V(I)$, as the set

$$V(I) = \{\mathbf{x} \in \mathbb{R}^n : A_i(\mathbf{x}) = 0 \text{ for each } A_i \in I\}.$$

A standard fact from algebraic geometry is that the set above is the same as the solution set to any finite collection of polynomials that generate the same ideal. Encoding equations with ideals and varieties has a key advantage in that it allows us to study logical implication and consistency questions geometrically: if $Q = 0$ can be derived from $A_1 = 0, \ldots, A_k = 0$, then $V(A_1, \ldots, A_k, Q) = V(A_1, \ldots, A_k)$. If adding Q to $\mathcal{A}$ introduces extra structure into the resulting variety, then Q is not derivable from A.

Decompose. Assume Q is not implied by $\mathcal{A}$. Then $V(\mathcal{A}, Q) \neq V(\mathcal{A})$. Suppose Q can be derived *nontrivially* from an axiom A_{k+1} along with $\mathcal{A}$, i.e., $Q = \alpha_1 A_1 + \cdots \alpha_{k+1} A_{k+1}$ for some polynomials α_i without $\alpha_{k+1} = 1, A_{k+1} = Q$. In this case $V(A_1, ...A_k, Q) = V(A_1, ..., A_k, \alpha_{k+1} A_{k+1})$

and since $\alpha_{k+1} \neq 1$, we know that intersecting $V(Q)$ with $V(\mathcal{A})$ is the same as intersecting $V(\alpha_{k+1}A_{k+1})$ with $V(\mathcal{A})$. Since $V(\alpha_{k+1}A_{k+1})$ is reducible, we are introducing new reducibility into $V(\mathcal{A})$ that is a factor of the residual $\alpha_{k+1}A_{k+1} = Q - (\sum_{i=1}^{k} \alpha_i A_i)$. We therefore can take our intersected variety $V(\mathcal{A}, Q)$ and study its irreducible components. We next define the mechanisms required for this process.

A *prime ideal* is an ideal $P \subset R$ such that if $fg \in P$ with $f, g \in R$, then $f \in P$ or $g \in P$. For a prime ideal P, $V(P)$ is irreducible, i.e., it cannot be represented as the union of two proper subsets that are varieties. The *radical* of an ideal I is denoted by $\sqrt{I}$ and is defined as:

$$\sqrt{I} = \{f \in R : f^m \in I \text{ for some positive integer } m\}.$$

The following key theorem allows us to break up a variety into irreducible components:

Theorem 1 (Primary Decomposition)**.** *Any ideal I can be decomposed into an intersection of ideals*

$$I = Q_1 \cap ... \cap Q_r$$

such that

1. *There are no redundancies among the Q_i, i.e., for each j we have $\cap_{i \neq j} Q_i \not\subset Q_j$.*
2. *For each i, $\sqrt{Q_i}$ is a prime ideal (called an associated prime of I).*

Let $P_i = \sqrt{Q_i}$. Using the fact that for any ideals I, J, we have $V(I \cap J) = V(I) \cup V(J)$ and $V(I) = V(\sqrt{I})$ [10], we get from the primary decomposition that a variety can be decomposed into a union of varieties corresponding to prime ideals: $V(I) = \bigcup_{i=1}^{r} V(P_i)$.

Decomposition in this setting serves a role analogous to Principal Component Analysis (PCA) in linear algebra, identifying latent, independent substructures that explain variation in the system. Whereas PCA decomposes a dataset into orthogonal directions of variance, primary decomposition identifies algebraically independent pieces of a solution space introduced by inconsistency or incompleteness.

Reason. Finally, for derivability, we need to define the algebraic notion of eliminating indeterminates from a system of polynomials. A "reduced Gröbner basis" of an ideal $I \subset R$ is a unique generating set $\{G_1, \ldots, G_m\}$ that satisfies $\langle G_1, ..., G_m \rangle = \langle A_1, ..., A_k \rangle = I$. If we define a lexicographic ordering $x_1 > ... > x_n$ on the indeterminates and compute the associated reduced Grobner basis, then we have the following property.

Theorem 2 (Elimination Theorem)**.** *For any integer d with $1 \leq d \leq n$, $\langle \{G_1, ..., G_m\} \cap \mathbb{R}[x_1, ..., x_d] \rangle = \langle A_1, ..., A_k \rangle \cap \mathbb{R}[x_1, ..., x_d]$*

So if we want to eliminate the variables $x_{d+1}, ..., x_n$ from the system, i.e., compute the ideal $\langle A_1, ..., A_k \rangle \cap \mathbb{R}[x_1, ..., x_d]$ which has infinitely many polynomials, we can compute a set of generators by finding a Gröbner basis of I and considering the elements defined over the variables $x_1, \ldots, x_d$. Geometrically, this corresponds to taking the variety in n-dimensional real space $V(I)$ and finding the projection $\pi_d V(I) = V(I \cap \mathbb{R}[x_1, ..., x_d]) \subset \mathbb{R}^d$ into the d-dimensional real space defined over the remaining variables. This provides a way to test algebraic consequence: if Q depends only on variables $x_1, \ldots, x_d$, we can eliminate all other variables and check whether Q is present in the projected ideal. We use this to test whether Q can be derived from a set of axioms and a candidate additional axiom extracted from decomposition. If so, we return the candidate as a possible missing axiom. While eliminating the variables is not strictly necessary to test for membership in the ideal, it allows for the stronger elimination of trivial axiom candidates that project onto smaller trivial surfaces that represent undesired conditions under which Q can be derived.

The generalized algorithm is detailed in Algorithm 2 in Appendix 6 as well as the necessary and sufficient conditions required for recovery in Appendix 5.2.1.

3 Results

In this section, we present the results from testing our system on various real-world physical theories. We test our ideas on 10 systems reported in AI Hilbert [9] as well as another appearing in a recent paper [19].

3.1 Problem 1. Kepler's Third Law

Setting. Consider Kepler's third law, which relates the orbital period of two celestial bodies to their distances and masses. It can be written as

$$Q := p - 2\pi\sqrt{\frac{(d_1+d_2)^3}{G(m_1+m_2)}} = 0,$$

where p is the orbital period, m_1, m_2 the masses, and d_1, d_2 their distances from the center of mass. For circular orbits, we use the following axioms (with p scaled so that $\omega p = 1$):

$$A_1 : (d_1+d_2)^2 F_g - G m_1 m_2 = 0 \tag{2}$$
$$A_2 : F_c - m_2 d_2 \omega^2 = 0 \tag{3}$$
$$A_3 : F_g - F_c = 0 \tag{4}$$
$$A_4 : \omega p - 1 = 0 \tag{5}$$
$$A_5 : m_1 d_1 - m_2 d_2 = 0. \tag{6}$$

Equation (2) gives gravitational force, (3) centrifugal force, (4) equates them, (5) relates frequency and period, and (6) defines the center of mass. To derive Q, we eliminate d_1 using (6), substitute $p = 1/\omega$ from (5), and replace ω^2 by combining (3)–(4) with (2). This yields

$$G(m_1+m_2)p^2 - (d_1+d_2)^3 = 0,$$

which is Kepler's law in polynomial form. Equivalently, setting

$$\alpha_1 = -p^2,\ \ \alpha_2 = p^2(d_1+d_2)^2,\ \ \alpha_3 = -p^2(d_1+d_2)^2,\ \ \alpha_4 = m_2 d_2(\omega p+1)(d_1+d_2)^2,\ \ \alpha_5 = -d_2^2,$$

we obtain

$$Q = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3 + \alpha_4 A_4 + \alpha_5 A_5.$$

Our System Now assume that we are missing knowledge of the equation of gravitational force – axiom A_1 – but we have Kepler's law Q at hand (possibly derived from sufficient data; see [42, 8, 9]). At this point, not only A_1 but also $\alpha_1, \ldots, \alpha_5$ are not known. We will assume they exist and attempt to find them from Q and the remaining known axioms $A_2 - A_5$. Here we view $\alpha_1 A_1$ as the (unknown) residual of Q with respect to $A_2, \ldots, A_5$. We define the ideal $I = \langle A_2, A_3, A_4, A_5, Q\rangle$. This is the same as the ideal $\langle A_2, A_3, A_4, A_5, \alpha_1 A_1\rangle$ as $Q = \sum_{i=1}^5 \alpha_i A_i$. Geometrically, $V(A_2, A_3, A_4, A_5, \alpha_1 A_1) = V(A_2, A_3, A_4, A_5) \cap V(\alpha_1 A_1)$. Since the latter is reducible, we have more irreducible components than we have for $V(A_2, A_3, A_4, A_5)$. Computing a primary decomposition of I then gives us the following associated primes:

$$\langle d_2, m_1, F_g, F_c, wp-1\rangle, \langle m_2, d_1, F_g, F_c, wp-1\rangle, \langle m_2, m_1, F_g, F_c, wp-1\rangle$$
$$\Big\langle F_c - F_g,\ m_1 d_1 - m_2 d_2,\ wp-1,\ F_g p - w m_2 d_2,\ F_g p^2 - m_2 d_2, F_g(d_1+d_2)^2 - m_1 m_2 G,$$
$$d_1(d_1+d_2)^2 - m_2 p^2 G, m_1 p^2 G - d_2(d_1+d_2)^2,\ w d_1 2(d_1+d_2)^2 - m_1 p G,$$
$$w d_1^3 - 3 w d_1 d_2^2 - 2 w d_2^3 + 2 m_1 p G - m_2 p G\Big\rangle$$

The generators of these irreducible components/ideals represent key equations that define each irreducible component. We iterate through each generator $\hat{A}_1$ and compute a Groebner basis to eliminate the variables not appearing in Q. We find that only the correct expression - $F_g(d_1+d_2)^2 - m_1 m_2 G$ - can derive Q and is therefore the only candidate returned by our system.

Axiom Formatting and Multipliers. Since the reducibility is introduced by $V(\alpha_i A_i)$, we do not distinguish between α_i and A_i when searching for ith axiom candidates. Additionally, we might return axioms in alternate forms - the exact expression is not guaranteed. For example, consider the same Kepler axioms and assume we are missing axiom A_4 instead. Then the relevant ideals in the primary decomposition are:

$$\Big\langle wp-1,\ F_g p^2 - m_2 d_2,\ \text{(other omitted generators)}\Big\rangle$$
$$\Big\langle wp+1,\ \text{(other omitted generators)}\Big\rangle$$

The highlighted polynomials are returned by our system as being axiom candidates that derive Q along with the remaining axioms. The other generators have been omitted for readability, and cannot derive Q. While the correct expression $wp-1$ is present, we have additional terms. The term $F_g p^2 - m_2 d_2$ is a restatement of A_2 using A_3 and the unknown $wp-1$ substituted in. It contains the same information as $wp-1$ since the former is already known. The term $wp+1$ is an artifact of the fact that α_4 as noted before is $m_2 d_2(wp+1)(d_1+d_2)^2$. Note that the variety $V(m_2 d_2(wp+1)(d_1+d_2)^2)$ is reducible and can be written as the union: $V(m_2) \cup V(d_2) \cup V(wp+1) \cup V((d_1+d_2)^2)$. Therefore, $wp+1$ appears as a viable axiom candidate since there is an algebraic derivation of Q from the remaining axioms along with $wp+1$. Incidentally, the remaining terms m_2, d_2, and $(d_1+d_2)^2$ also show up in the associated primes:

$$\big\langle F_c - F_g,\ m_2^2,\ (d_1+d_2)^2,\ m_1 d_2 + d_1 m_2 + 2m_2 d_2,\ m_1 m_2,$$
$$m_1 d_1 - m_2 d_2,\ m_1^2,\ F_g m_2,\ F_g m_1,\ F_g^2,\ F_g - w^2 m_2 d_2\big\rangle$$
$$\langle d_2,\ m_1,\ F_g,\ F_c\rangle, \langle m_2,\ d_1,\ F_g,\ F_c\rangle$$

Applying the elimination theorem and keeping only the candidates that project exactly onto the variables over which Q is defined eliminates these candidates.

3.2 Single Axiom Removed Results

Problem	# Axioms Recovered	Avg. Time (s)	Total Axioms
Kepler	5/5	0.1	5
Compton	10/10	5.4	10
Einstein	5/5	1.5	5
Escape Velocity	5/5	0.4	5
Light Damping	5/5	1.6	5
Hagen Poiseuille	4/4	0.6	4
Neutrino Decay	5/5	3.5	5
Hall Effect	7/7	11.1	9
Carrier-Resolved Photo-Hall Effect	7/7	1.1	7

Table 1: Summary of Axiom Recovery Performance by Problem

We tested our system on 7 axiom systems from AI Hilbert, as well as on a more recent work in the scientific literature [19]. In each case, We iterate through the axioms, drop an axiom from the system, and test if we can recover the missing axiom while not making any assumption about it. In almost every case, we were able to recover the missing axiom. A detailed list of axioms in each system can be found in the Appendix 5.6.

4 Conclusion and Remarks

In this work, we propose an AI system called AI-Noether, which bridges the gap between AI-driven scientific discoveries and existing scientific knowledge, provided both the discoveries and the knowledge are expressible as real polynomials. In particular, given a set of background theory axioms and a discovered polynomial, our approach derives a minimal set of additional polynomial axioms necessary to derive the given polynomial. Furthermore, we demonstrate the effectiveness of our approach on a selection of renowned scientific laws, including Kepler's Third Law and the Carrier-Resolved Photo-hall effect.

For future work, it would be interesting to extend our approach to settings with non-polynomial scientific laws, e.g., settings where laws are expressed as ODEs or PDEs. It would also be interesting to apply our approach to new data-driven scientific discoveries, thereby reconciling them with prior literature.

References

[1] Team AlphaProof and Team AlphaGeometry. AI achieves silver-medal standard solving international 178 mathematical olympiad problems. *DeepMind blog*, 179:45, 2024.

[2] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature communications*, 5(1):4308, 2014.

[3] Yoshua Bengio. The consciousness prior. *arXiv preprint arXiv:1709.08568*, 2017.

[4] Dimitris Bertsimas and Wes Gurnee. Learning sparse nonlinear dynamics via mixed-integer optimization. *Nonlinear Dynamics*, 111(7):6585–6604, 2023.

[5] Daniil A Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578, 2023.

[6] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[7] Chin-Wen Chou, David B Hume, Till Rosenband, and David J Wineland. Optical clocks and relativity. *Science*, 329(5999):1630–1633, 2010.

[8] Cristina Cornelio, Sanjeeb Dash, Vernon Austel, Tyler R Josephson, Joao Goncalves, Kenneth L Clarkson, Nimrod Megiddo, Bachir El Khadir, and Lior Horesh. Combining data and theory for derivable scientific discovery with ai-descartes. *Nature Communications*, 14(1):1777, 2023.

[9] Ryan Cory-Wright, Cristina Cornelio, Sanjeeb Dash, Bachir El Khadir, and Lior Horesh. Evolving Scientific Discovery by Unifying Data and Background Knowledge with AI Hilbert. *Nature Communications*, 15(1):5922, 2024.

[10] David Cox, John Little, and Donald O'Shea. *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 1991.

[11] Miles Cranmer. Interpretable machine learning for science with pysr and symbolicregression.jl, 2023.

[12] Wang-Zhou Dai, Qiuling Xu, Yang Yu, and Zhi-Hua Zhou. Bridging machine learning and logical reasoning by abductive learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[13] Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with ai. *Nature*, 600(7887):70–74, 2021.

[14] Marissa R Engle and Nikolaos V Sahinidis. Deterministic symbolic regression with derivative information: General methodology and application to equations of state. *AIChE Journal*, 68(6):e17457, 2022.

[15] Brian D Fields. The primordial lithium problem. *Annual Review of Nuclear and Particle Science*, 61(1):47–68, 2011.

[16] Andre K Geim and Irina V Grigorieva. Van der waals heterostructures. *Nature*, 499(7459):419–425, 2013.

[17] Kurt Gödel. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i. *Monatshefte für mathematik und physik*, 38(1):173–198, 1931.

[18] Roger Guimerà, Ignasi Reichardt, Antoni Aguilar-Mogas, Francesco A Massucci, Manuel Miranda, Jordi Pallarès, and Marta Sales-Pardo. A Bayesian machine scientist to aid in the solution of challenging scientific problems. *Science Advances*, 6(5):eaav6971, 2020.

[19] Oki Gunawan, Seong Ryul Pae, Douglas M. Bishop, Yudistira Virgus, Jun Hong Noh, Nam Joong Jeon, Yun Seog Lee, Xiaoyan Shao, Teodor Todorov, David B. Mitzi, and Byungha Shin. Carrier-resolved photo-hall effect. *Nature*, 575(7781):151–155, November 2019. Epub 2019 Oct 7.

[20] Yu-Xuan Huang, Wang-Zhou Dai, Le-Wen Cai, Stephen H Muggleton, and Yuan Jiang. Fast abductive learning by similarity-based consistency optimization. *Advances in Neural Information Processing Systems*, 34:26574–26584, 2021.

[21] Alexey Ignatiev, Nina Narodytska, and Joao Marques-Silva. Abduction-based explanations for machine learning models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1511–1519, 2019.

[22] Mario Krenn, Robert Pollice, Si Yue Guo, Matteo Aldeghi, Alba Cervera-Lierta, Pascal Friederich, Gabriel dos Passos Gomes, Florian Häse, Adrian Jinich, AkshatKumar Nigam, et al. On scientific understanding with artificial intelligence. *Nature Reviews Physics*, 4(12):761–769, 2022.

[23] Jiří Kubalík, Erik Derner, and Robert Babuška. Symbolic regression driven by training data and prior knowledge. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pages 958–966, 2020.

[24] Jiří Kubalík, Erik Derner, and Robert Babuška. Multi-objective symbolic regression for physics-aware dynamic modeling. *Expert Systems with Applications*, 182:115210, 2021.

[25] Thomas S Kuhn. *The structure of scientific revolutions*, volume 962. University of Chicago press Chicago, 1997.

[26] P. Langley. Integrated systems for computational scientific discovery. In *Proceedings of the AAAI Conference on Artificial Intelligence 38(20)*, pages 22598–22606, 2024.

[27] Zachary C Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, 2018.

[28] Nour Makke and Sanjay Chawla. Interpretable scientific discovery with symbolic regression: a review. *Artificial Intelligence Review*, 57(1):2, 2024.

[29] P. Marquis. Extending abduction from propositional to first-order logic. In *In: Jorrand, P., Kelemen, J. (eds) Fundamentals of Artificial Intelligence Research*, 1991.

[30] Santiago Miret and Nandan M Krishnan. Are llms ready for real-world materials discovery? *arXiv preprint arXiv:2402.05200*, 2024.

[31] A Murari, E Peluso, M Lungaroni, P Gaudio, J Vega, and M Gelfusa. Data driven theory for knowledge discovery in the exact sciences with applications to thermonuclear fusion. *Scientific Reports*, 10(1):19858, 2020.

[32] Emmy Noether. Invariant variation problems. *Transport theory and statistical physics*, 1(3):186–207, 1971.

[33] Peter Norvig. Inference in text understanding. In *AAAI*, pages 561–565, 1987.

[34] Gabriele Paul. Approaches to abductive reasoning: an overview. *Artificial intelligence review*, 7(2):109–152, 1993.

[35] Judea Pearl. Reasoning with cause and effect. *AI Magazine*, 23(1):95–95, 2002.

[36] Charles Sanders Peirce. *Collected papers of charles sanders peirce*, volume 5. Harvard University Press, 1934.

[37] Joshua C Peterson, David D Bourgin, Mayank Agrawal, Daniel Reichman, and Thomas L Griffiths. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, 372(6547):1209–1214, 2021.

[38] Chandan K Reddy and Parshin Shojaee. Towards scientific discovery with generative ai: Progress, opportunities, and challenges. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 28601–28609, 2025.

[39] M. T. Ribeiro, S. Singh, and C. Guestrin. ”why should i trust you?”: Explaining the predictions of any classifier. *KDD*, pages 1135—-1144, 2016.

[40] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

[41] D.R. Swanson and N.R. Smalheiser. An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence*, pages 183–203, 1997.

[42] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 2020.

[43] Dimitar Valev. Estimations of total mass and energy of the universe. *arXiv preprint arXiv:1004.1035*, 2010.

[44] Zishen Wan, Che-Kai Liu, Hanchen Yang, Chaojian Li, Haoran You, Yonggan Fu, Cheng Wan, Tushar Krishna, Yingyan Lin, and Arijit Raychowdhury. Towards cognitive ai systems: a survey and prospective on neuro-symbolic ai. *arXiv preprint arXiv:2401.01040*, 2024.

[45] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne Bergen, Carla P Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun Manrai, Debora Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Veličković, Max Welling, Linfeng Zhang, Connor W Coley, Yoshua Bengio, and Marinka Zitnik. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, August 2023.

[46] Clifford M Will. The confrontation between general relativity and experiment. *Living reviews in relativity*, 17(1):1–117, 2014.

[47] Xiao-Wen Yang, Wen-Da Wei, Jie-Jing Shao, Yu-Feng Li, and Zhi-Hua Zhou. Analysis for abductive learning and neural-symbolic reasoning shortcuts. In *Forty-first International Conference on Machine Learning*, 2024.

[48] Zhi-Hua Zhou. Abductive learning: towards bridging machine learning and logical reasoning. *Science China. Information Sciences*, 62(7):76101, 2019.

Table 2: Six minimal theories with two axioms each that explain $q = 0$.

$b = 0$	$b = 0$	$b = 0$	$x^2 - ay^2 = 0$	$x^2 - ay^2 = 0$	$x^2 - ay^2 = 0$
$a = 0$	$by + z = 0$	$by - z = 0$	$a = 0$	$by + z = 0$	$by - z = 0$

5 Appendix

5.1 Example 1 Continued

Notice that there are many alternative theories from which the equation $q = 0$ follows as a consequence. For example, looking at the equation (1) and letting $p_3 = by + z$, it is clear that $q = 0$ follows from the axioms $p_1 = 0$ and $p_3 = 0$ as $q = \alpha p_1 + ap_2 p_3$. In a similar manner, we can trivially create the six *minimal* theories in Table 2 where the first axiom is formed by taking a polynomial that is a factor of αp_1 and setting it to 0, and the second axiom is similarly composed from a factor of βp_2; $q = 0$ is a consequence of each theory.

Each of the 6 axiom systems above is minimal in the sense that we cannot remove an axiom from it and still have $q = 0$ as a consequence. Depending on the context, some of the axiom systems may be less interesting than others. For example, if a represents the mass of a pendulum, $a = 0$ is likely not an interesting condition. On the other hand, if a represents temperature in Celsius, $a = 0$ may represent a meaningful condition.

The axiom system $p_1 = 0, p_2 p_3 = 0$ is also a valid theory explaining $q = 0$, but we call it *reducible* in the sense that we can trivially get a simpler axiom system if we replace $p_2 p_3$ by either of its factors. Thus, if our only available axioms was $p_1 = 0$, the "residual" of the set of axioms $\{p_2 p_3 = 0\}$ is not as small as possible. There are other ways of building an infinitude of axiom systems such that $q = 0$ is a consequence. Starting from the axiom system $p_1 = 0, p_2 = 0$, it is easy to see that $p_1 = 0, rp_1 + sp_2 = 0$, where r, s are nonzero real numbers, are valid theories that explain $q = 0$. Again, starting with $p_1 = 0$, the axiom set $\{rp_1 + sp_2 = 0\}$ does not have the least residual.

Now assume we have experimental data for a phenomenon consisting of a number of data points specifying values of x, z, a, b (with a and b unchanging in the data) and assume the data satisfies the equation $b^2x^2 - az^2 = 0$ and that we are able to numerically obtain this equation from the data. Further assume that we have a partial theoretical explanation of the phenomenon, where the partial theory assumes the presence of another variable y and the equation $x^2 - ay^2 = 0$, i.e., $p_1 = 0$. In this situation, we consider y as an unmeasured quantity, and x, z, a, and b as measured quantities. Given that the data points at hand do not have any values for y, it is not possible to infer the relationship $by - z$ directly from the data.

Consider the equation $0 = b^2 p_1 = b^2x^2 - b^2ay^2$ that follows from $p_1 = 0$. Subtracting the derived equation $b^2x^2 - az^2 = 0$ from it, we obtain the relationship $a(z^2 - b^2y^2) = 0$ as a consequence of $p_1 = 0$ and $q = 0$. This implies that either $a = 0$ or $by - z = 0$ or $by + z = 0$. Each of the equations (or products of pairs of equations) in the previous line can be appended to $p_1 = 0$ to get a valid theory (in the absence of other information) explaining $q = 0$.

5.2 Theoretical Basis of Algorithms

Let $A_1, \ldots, A_k$ be a collection of polynomials with $A_1 = 0, \ldots, A_k = 0$ being the associated axioms. Let $\mathcal{A}$ stand for the set $\{A_1, \ldots, A_k\}$. Let $V_R(\mathcal{A})$, respectively $V_C(\mathcal{A})$, be the set of real, respectively complex, solutions of the equations $A_1 = 0, \ldots, A_k = 0$. Note that $V_R(\mathcal{A})$ is the same as the variety $V(\mathcal{A})$ defined in the main body of the paper.

First note that $V_C(\mathcal{A}, Q) = V_C(\mathcal{A})$ if and only if $Q = 0$ for all complex solutions of $A_1 = 0, \ldots, A_k = 0$. Further, if $Q \in I(\mathcal{A})$, then $V_C(\mathcal{A}, Q) = V_C(\mathcal{A})$. However, the converse is not true. By Hilbert's Nullstellensatz, if $V_C(\mathcal{A}, Q) = V_C(\mathcal{A})$, then $Q^m \in I(\mathcal{A})$ for some positive integer m (but m need not equal 1). But $Q^m(\bar{x}) = 0$ if and only $Q(\bar{x}) = 0$. Accordingly, if we say that $Q = 0$ is derivable from $\mathcal{A}$ if and only if $Q^m \in I(\mathcal{A})$, then the notions of "a consequence of" and "derivable from" are identical assuming complex-valued solutions are relevant.

On the other hand, if we are only interested in real-valued solutions, the situation is more complicated. As before, $V_R(\mathcal{A}, Q) = V_R(\mathcal{A})$ if and only if $Q = 0$ for all real solutions of $A_1 = 0, \ldots, A_k = 0$. Further, if $Q \in I(\mathcal{A})$, then $V_R(\mathcal{A}, Q) = V_R(\mathcal{A})$. However, the converse is not true. If $V_R(\mathcal{A}, Q) = V_R(\mathcal{A})$, then $Q^{2m} + S \in I(\mathcal{A})$ for some positive integer m and some polynomial S that is a sum of squares of polynomials in R.

5.2.1 Criteria for Recoverability of Axioms

The theorem below provides a sufficient criterion for recovering missing axioms.

First we will need the notion of height of an ideal. For a prime ideal P, we define its height, $\mathrm{ht}(P)$, as the length n of the longest chain $P_0 \subsetneq P_1 \cdots \subsetneq P_n = P$ of prime ideals contained in P. For a general ideal, its height is the minimum of the heights of its associated primes, which is the same as the minimum of the heights of all primes which contain the ideal.

Theorem 3 (Common minimal associated prime)**.** *Given ideals $I \subset J \subsetneq \mathbb{R}[x_1, ..., x_n] \mid \mathrm{ht}(I) = \mathrm{ht}(J)$, then I and J have a common minimal associated prime.*

Proof. Let P be an associated prime for J such that $\mathrm{ht}(P) = \mathrm{ht}(J)$. Then we have that $I \subset J \subset P$. Since $\mathrm{ht}(I) = \mathrm{ht}(P)$, there is no prime Q such that $I \subset Q \subsetneq P$. Otherwise $\mathrm{ht}(I) \leq \mathrm{ht}(Q) < \mathrm{ht}(P)$. But then P is a minimal prime containing I and thus a minimal associated prime for I.

Consider I to be the ideal with known information $\langle A_1, ..., A_k, Q\rangle$ and J the ideal of the true axioms $\langle A_1, ..., A_{k+1}\rangle$. If I and J have the same height, then we know that they have a common associated prime, which contains the missing axiom A_{k+1}. If Q and A_{k+1} are not contained in any of the associated primes for $\langle A_1, ..., A_k\rangle$, then we have that $\mathrm{ht}\langle A_1, ..., A_k, Q\rangle = \mathrm{ht}\langle A_1, ..., A_{k+1}\rangle$ and the theorem applies.

Typically if multiple axioms are missing, adding back a single phenomenon Q, will not be sufficient, so in general we will need to add back several consequent phenonema $\{Q_1, \ldots Q_h\}$ in order to generate an ideal which is guaranteed to have a component which contains all the missing axioms.

Our missing axiom construction is attempting to find axioms which are sufficient to derive Q, i.e. whether or not Q is contained in the ideal generated by the axioms. If we want to decide whether or not a set of axioms implies Q, given the ideal I generated by the axioms, we are asking whether $V(I) \subset V(Q)$. To make this test effective, we need to construct the biggest ideal with the same set of zeros, this is $I(V(I))$. So the test whether a set of axioms A implies Q is whether $Q \in I(V\langle A\rangle)$. If our varieties are defined over the reals, then $I(V\langle A\rangle)$ can be computed as the real radical of the ideal $\langle A\rangle$.

5.3 Problem 2. Einstein's Relativistic Time Dilation Law

Einstein's time dilation law describes how two observers in motion experience time differently. Letting c be the speed of light, f the frequency of a clock moving at speed v, and f_0 the frequency of a stationary clock, we have the relation:

$$\frac{f}{f_0} = \sqrt{1 - \frac{v^2}{c^2}}$$

which, after squaring and clearing denominators, gets us:

$$c^2 f_0^2 - c^2 f^2 - f_0^2 v^2 = 0 \tag{7}$$

This can be derived from the following axioms:

$$A_1 := cdt_0 - 2d = 0 \tag{8}$$

$$A_2 := cdt - 2L = 0 \tag{9}$$

$$A_3 := 4L^2 - 4d^2 - v^2\, dt^2 = 0 \tag{10}$$

$$f_0 dt_0 - 1 = 0 \tag{11}$$

$$f dt - 1 = 0 \tag{12}$$

A_1 encodes the time taken for light to travel from a stationary light source and be reflected back to the origin from a stationary mirror. A_2 similarly captures the time taken for light to start from a moving light-source and arrive back at the source (at its new position) after reflection in a stationary mirror. A_3 relates the initial distance between the light-source and mirror to the distance traveled by light because of the motion. A_4, A_5 relate frequencies to periods. A_2 captures the information that the speed of light is constant. This would not be predicted by traditional Newtonian mechanics, which would instead predict the time taken for light to travel from the moving light-source and back as:

$$dt^2(c^2 + v^2) - 4L^2 = 0 \tag{13}$$

[9] demonstrated that, through a discrete optimization setup, one can discover the correct expression for time dilation while eliminating the incorrect Newtonian axiom. [8] showed a similar result, but using automated reasoning after a symbolic regression fit to the correct formula. While both systems successfully recover the time dilation equation with incorrect axioms, the correct axiom needs to be present in order to obtain the certificate of derivability, Instead of appending the incorrect axiom to (8)-(12), we replace axiom (9) by it; while both systems discover the right formula, they no longer provide a certificate of derivability. We therefore need to know that the speed of light is constant in order to derive time dilation.

Assume we were missing axiom (9). Then we'd have a truly incorrect set of axioms (8), (10)-(12), and (13) that cannot explain the phenomenon of time dilation (7).

However, using the system in this paper, we can identify a correction that suffices to explain the hypothesis and detect inconsistencies in the theory. Running our system results in an empty primary decomposition: $V = \phi$, indicating the inconsistency in the axioms. We then iterate over each axiom, remove it, and run our system on the remaining axioms and time-dilation hypothesis. We get an empty set as the decomposition for all but one removal - the case of removing the incorrect Newtonian axiom 13. In this case, we get the following relevant prime ideals:

$$\begin{aligned}\langle & 2Lf + c,\ dt_0 f_0 - 1,\ dt f - 1,\ dtc + 2L,\ 2d - dt_0 c,\ 2df_0 - c,\ d\,dt + dt_0 L,\\ & 4d^2 f + dt\,v^2 + 2Lc,\ c^2 f_0^2 - c^2 f^2 - f_0^2 v^2,\ dt f_0 v^2 + dt_0 c^2 f + 2Lcf_0,\\ & dt^2 v^2 + dt_0^2 c^2 - 4L^2,\ 2dcf^2 - c^2 f_0 + f_0 v^2,\ dc^2 f + Lc^2 f_0 - Lf_0 v^2,\\ & 2d\,dt_0 cf + dt\,v^2 + 2Lc,\ d\,dt_0 c^2 - dtLv^2 - 2L^2 c,\ d^2 c^2 - L^2 c^2 + L^2 v^2\rangle\end{aligned}$$

$$\begin{aligned}\langle & 2Lf - c,\ dt_0 f_0 - 1,\ dt f - 1,\ dtc - 2L,\ 2d - dt_0 c,\ 2df_0 - c,\ d\,dt - dt_0 L,\\ & 4d^2 f + dt\,v^2 - 2Lc,\ c^2 f_0^2 - c^2 f^2 - f_0^2 v^2,\ dt f_0 v^2 + dt_0 c^2 f - 2Lcf_0,\\ & dt^2 v^2 + dt_0^2 c^2 - 4L^2,\ 2dcf^2 - c^2 f_0 + f_0 v^2,\ dc^2 f - Lc^2 f_0 + Lf_0 v^2,\\ & 2d\,dt_0 cf + dt\,v^2 - 2Lc,\ d\,dt_0 c^2 + dtLv^2 - 2L^2 c,\ d^2 c^2 - L^2 c^2 + L^2 v^2.\rangle\end{aligned}$$

In the reasoning step, we select the highlighted generators. We see that we get the exact expressions of the corrected axiom $cdt - 2L$, and also equivalent statements with this expression substituted into axiom 9 and 11 respectively. We also get the equivalent candidates corresponding to $cdt + 2L$ since the multiplier α_2 in the derivation is: $(2L + cdt)$.

Therefore, without any prior assumptions about the speed of light, we can infer that the speed of light being a constant suffices to explain the time dilation formula.

5.4 Problem 9. Carrier-Resolved Photo-Hall Effect

The Carrier-Resolved Photo-Hall effect [19] equation describes the relationship between various parameters of a semiconducting surface and is given by:

$$H = \frac{re\mu_p^2[p_0 + \Delta n(1 - \beta^2)]}{\sigma^2}$$

where $\sigma = e\mu_p[p_0 + \Delta n(1 + \beta)]$. When we clear the denominators, we get the polynomial form:

$$re\mu_p\Delta n\beta^2 + re\mu_p\Delta n\beta - r\sigma + ep_0\sigma H + e\Delta n\beta\sigma H + e\Delta n\sigma H = 0$$

This can be derived from the following axioms:

$$\begin{aligned}
A_1 &:= \beta\mu_P - \mu_N \\
A_2 &:= \mu_H - r\mu \\
A_3 &:= p_h - p_0 - \Delta p \\
A_4 &:= n - \Delta n \\
A_5 &:= \Delta p - \Delta n \\
A_6 &:= \sigma - ep_h\mu_p - en\mu_N \\
A_7 &:= H(p_h + \beta n)^2 e - rp_h + r\beta^2 n.
\end{aligned}$$

Here μ_N and μ_P are electron and hole drift mobilities, $\beta = \mu_N/\mu_P$ (essentially A_1) is the mobility ratio for electron and hole. μ_H is the Hall mobility, and r is the Hall scattering factor that relates the drift and Hall mobility via the equation $r = \mu_H/\mu$ (essentially A_2). Δn and Δp are electron and hole photo-carrier densities and are equal in a steady state equilibrium (A_5). H is the Hall coefficient and σ is conductivity. A_6 and A_7 are the two-carrier Hall equations in the low-field regime. n stands for electron density. p_0 stands for background hole density, p_h for hole density (assuming a p-type material). A_3 and A_4 give changes in hole and electron density for a Photo-Hall experiment with a p-type material.

5.5 Multiple missing axioms

We demonstrated efficacy in removing/correcting one axiom from an axiom list. We now look at what happens when we need to recover more than one axiom in order to explain a hypothesis Q. In this case, we have

$$Q = \sum_{i=1}^{k} \alpha_i A_i$$

but we only have access to $A_1, ..., A_{k-r}$. Alternatively, we want to find a substitution for r axioms in our system in order to explain Q.

Kepler's Third Law of Planetary Motion

Recall that the Kepler axioms were:

$$A_1 := (d_1 + d_2)^2 F_g - Gm_1m_2 \tag{14}$$

$$A_2 := F_c - m_2d_2w^2 \tag{15}$$

$$A_3 := F_g - F_c \tag{16}$$

$$A_4 := wp - 1 \tag{17}$$

$$A_5 := m_1d_1 - m_2d_2 \tag{18}$$

The statement of Kepler's law is:

$$Q := m_1m_2Gp^2 - d_2^2d_1m_1 - d_1^2d_2m_2 - 2d_1d_2^2m_2 - d_1^2d_2m_2 \tag{19}$$

Assume that we are missing axioms (16) and (17) - the relationship between centrifugal and gravitational force and the relationship between frequency and period. Then, encoding the known information - equations (14), (15),(18), and (19) - our system returns the following relevant component of the primary decomposition and selected generator from the reasoning module:

Problem	Missing Axioms	Recovered	Recovered Axiom(s)
Kepler	$\{\texttt{G}, \texttt{cf}\}$	X	–
Kepler	$\{\texttt{G}, \texttt{C}\}$	✓	$(d_1 + d_2)^2 F_c = G m_1 m_2$
Kepler	$\{\texttt{G}, \texttt{wp}\}$	X	–
Kepler	$\{\texttt{cf}, \texttt{C}\}$	✓	gravitational force
Kepler	$\{\texttt{cf}, \texttt{wp}\}$	✓	gravitational force
Kepler	$\{\texttt{C}, \texttt{wp}\}$	✓	gravitational force
Kepler	$\{\texttt{G}, \texttt{cf}, \texttt{C}\}$	X	–
Kepler	$\{\texttt{G}, \texttt{cf}, \texttt{wp}\}$	X	–
Kepler	$\{\texttt{G}, \texttt{C}, \texttt{wp}\}$	X	–
Kepler	$\{\texttt{cf}, \texttt{C}, \texttt{wp}\}$	✓	gravitational force

Table 3: Recovery results on the Kepler problem for various subsets of missing axioms. Shorthand: `G` = gravitational force $(d_1 + d_2)^2 F_g = G m_1 m_2$, `cf` = centrifugal force $F_c = m_2 d_2 w^2$, `C` = force balance $F_g = F_c$, `wp` = frequency-period relation $wp = 1$, `cm` = center of mass $m_1 d_1 = m_2 d_2$.

$$\begin{aligned}\big\langle m_1 d_1 - m_2 d_2,\ F_g p^2 - m_2 d_2,\ &F_g(d_1^2 + 2 d_1 d_2 + d_2^2) - m_1 m_2 G,\\ d_1^3 + 2 d_1^2 d_2 + d_1 d_2^2 - m_2 p^2 G,\ &m_1 p^2 G - d_1^2 d_2 - 2 d_1 d_2^2 - d_2^3,\\ F_c - w^2 m_2 d_2,\ &F_c p^2 G - w^2 d_1^3 d_2 - 2 w^2 d_1^2 d_2^2 - w^2 d_1 d_2^3 \big\rangle\end{aligned}$$

We therefore return $F_g p^2 - m_2 d_2$. This equation is the same as the centrifugal force equation (15) but with the missing relationships $F_c = F_g$ and $w = \frac{1}{p}$ substituted in. Therefore, the missing information is discovered, since equation (15) is assumed to be known, although it is not decoupled. To go one step further, we can remove three out of the five axioms: equations (15),(16), and (17), leaving us with only the gravitational force equation (14) and center of mass equation (18). Similarly, running our system with the Kepler hypothesis gives us a similar component:

$$\begin{aligned}\big\langle m_1 d_1 - m_2 d_2,\ F_g p^2 - m_2 d_2,\ &F_g(d_1^2 + 2 d_1 d_2 + d_2^2) - m_1 m_2 G,\\ d_1^3 + 2 d_1^2 d_2 + d_1 d_2^2 - m_2 p^2 G,\ &m_1 p^2 G - d_1^2 d_2 - 2 d_1 d_2^2 - d_2^3 \big\rangle\end{aligned}$$

We get the same axiom candidate. However, in the previous case, where we had knowledge of the centrifugal force equation, we are assuming here that we do not know it. So we're not only recovering the relationships $F_c = F_g$ and $wp - 1$, but also the centrifugal force axiom itself $F_c = m_2 d_2 w^2$ with the former relations coupled with it.

Table 3 shows the results of attempting to recover various combinations of pairs and triples of axioms for Kepler's third law.

Einstein Relativistic Time Dilation Law

We also tested our system on Einstein's relativistic time dilation law, this time under conditions where multiple axioms are simultaneously removed. The results are summarized in Table 4, which lists the different tuples of missing axioms together with the runtime and whether the system successfully recovered the law. In total, a majority of two-axiom removals were successfully recovered, while most of the failures occurred when three axioms were missing simultaneously. This illustrates that the method is robust to partial loss of information, but recovery becomes significantly more difficult when too many structural constraints are removed.

Across six benchmark systems, we removed pairs and triples of axioms and attempted to recover explanations for Q. Success rates range from 25–60% (Kepler 5/10; Hagen–Poiseuille 6/10; Einstein 8/20; others 5–7/20), with most failures occurring when three axioms are missing, especially in larger axiom sets. Runtimes are modest overall (avg. 0.1–3.5 s), with heavier systems (Einstein, Neutrino Decay) taking longer on average. The results can be seen in table 5.

Table 4: Results for removal of multiple axioms in Einstein's time dilation law axiom system

Problem	Missing Axioms (Tuple)	Time	Recovered
Einstein	$\{cdt_0 - 2d,\ 4L^2 - 4d^2 - v^2dt^2\}$	0.4s	✓
Einstein	$\{cdt_0 - 2d,\ f_0dt_0 - 1\}$	0.1s	✓
Einstein	$\{cdt_0 - 2d,\ fdt - 1\}$	0.1s	X
Einstein	$\{cdt_0 - 2d,\ cdt - 2L\}$	0.1S	✓
Einstein	$\{4L^2 - 4d^2 - v^2dt^2,\ f_0dt_0 - 1\}$	0.2s	X
Einstein	$\{4L^2 - 4d^2 - v^2dt^2,\ fdt - 1\}$	0.1s	✓
Einstein	$\{4L^2 - 4d^2 - v^2dt^2,\ cdt - 2L\}$	0.1s	✓
Einstein	$\{f_0dt_0 - 1,\ fdt - 1\}$	0.2s	X
Einstein	$\{f_0dt_0 - 1,\ cdt - 2L\}$	0.1s	X
Einstein	$\{fdt - 1,\ cdt - 2L\}$	0.1s	✓
Einstein	$\{cdt_0 - 2d,\ 4L^2 - 4d^2 - v^2dt^2,\ f_0dt_0 - 1\}$	0.1s	X
Einstein	$\{cdt_0 - 2d,\ 4L^2 - 4d^2 - v^2dt^2,\ fdt - 1\}$	0.1s	X
Einstein	$\{cdt_0 - 2d,\ 4L^2 - 4d^2 - v^2dt^2,\ cdt - 2L\}$	0.3s	✓
Einstein	$\{cdt_0 - 2d,\ f_0dt_0 - 1,\ fdt - 1\}$	0.1s	X
Einstein	$\{cdt_0 - 2d,\ f_0dt_0 - 1,\ cdt - 2L\}$	0.1s	X
Einstein	$\{cdt_0 - 2d,\ fdt - 1,\ cdt - 2L\}$	0.2s	X
Einstein	$\{4L^2 - 4d^2 - v^2dt^2,\ f_0dt_0 - 1,\ fdt - 1\}$	0.1s	X
Einstein	$\{4L^2 - 4d^2 - v^2dt^2,\ f_0dt_0 - 1,\ cdt - 2L\}$	0.1s	✓
Einstein	$\{4L^2 - 4d^2 - v^2dt^2,\ fdt - 1,\ cdt - 2L\}$	0.1s	X
Einstein	$\{f_0dt_0 - 1,\ fdt - 1,\ cdt - 2L\}$	0.1s	X

Problem	# Tuples Recovered	Avg. Time (s)	# of Axioms
Kepler	5/10	0.1	4
Einstein	8/20	1.5	5
Escape Velocity	6/20	0.4	5
Light Damping	5/20	1.6	5
Hagen Poiseuille	6/10	0.6	4
Neutrino Decay	7/20	3.5	5

Table 5: Summary of 2- and 3-Axiom Tuple Recovery Performance by Problem

5.6 Detailed Breakdown of single axiom removal experiment

We tested the system by removing one axiom at a time and asking it to recover the missing axiom, reporting our findings in Table 6. Across eight problems (47 runs), the method succeeded in 45/47 cases (95.7%). Recovery was perfect for Kepler (4/4, 0.1 s each), Einstein (5/5, 1.5 s), Escape Velocity (5/5, 0.4 s), Light Damping (5/5, 1.6 s), Hagen–Poiseuille (4/4, 0.6 s), Neutrino Decay (5/5, 3.5 s), and Compton scattering (10/10, 5.4 s). The two failures occur in the Hall Effect system (7/9 recovered, 11.1 s), for constraints $nV - N$ and $V - Lhd$, which behave like global counting/geometry conditions and couple more weakly to the variables used to test derivability of Q. Runtimes are modest overall (weighted average $\approx$ 4.08 s per run), dominated by the larger Compton and Hall instances; the remaining domains finish in sub-second to low single-second time on our machine (machine details can be found in appendix 6).

6 Algorithm Details

Algorithm 2 describes our procedure for identifying missing axioms that render a target hypothesis Q derivable. The algorithm proceeds in three phases that parallel our methodological framework: **Encode**, **Decompose**, and **Reason**. We now describe each step in detail.

Input and Initialization. The input is a finite set of polynomial axioms $\mathcal{A} = \{A_1, \dots, A_k\}$ in variables $\mathcal{V} = \{x_1, \dots, x_n\}$, and a hypothesis polynomial Q over a restricted subset of variables

$\{x_1, \ldots, x_d\}$, where $d < n$. The algorithm maintains a set $\mathcal{Q}_{\text{expl}}$ of explanatory candidates, initially empty.

Encode. We form the polynomial ideal $I = \langle A_1, \ldots, A_k, Q\rangle$. Geometrically, $V(I)$ corresponds to the variety of solutions simultaneously satisfying the known axioms and the hypothesis. If Q is derivable from $\mathcal{A}$, then $V(I)$ coincides with $V(\mathcal{A})$; otherwise, additional structure appears in the decomposition.

Decompose. We compute a primary decomposition

$$I = Q_1 \cap \cdots \cap Q_r,$$

yielding associated primes $P_j = \sqrt{Q_j}$. Each P_j describes an irreducible component of the solution space. The generators of P_j are potential candidates for missing axioms: if appended to $\mathcal{A}$, they may restrict the solution set in exactly the way needed to derive Q.

Reason. For each generator $\widehat{A_{k+1}}$ of an associated prime P_j, we form the augmented system $J = \langle A_1, \ldots, A_k, \widehat{A_{k+1}}\rangle$. We compute a Gröbner basis $\mathcal{G}$ of J with respect to a lexicographic order on $\mathcal{V}$. To compare against the hypothesis, we project $\mathcal{G}$ to the coordinate ring of the observed variables, $\mathcal{G}' = \mathcal{G} \cap \mathbb{R}[x_1, \ldots, x_d]$. If $Q \in \mathcal{G}'$, then $\widehat{A_{k+1}}$ serves as an *explanatory axiom* for Q, and is added to $\mathcal{Q}_{\text{expl}}$.

Output. The algorithm returns the set $\mathcal{Q}_{\text{expl}}$ of all explanatory candidates obtained through this reasoning process.

Implementation and Hardware

All symbolic computations were performed using `Macaulay2` (for Gröbner basis and primary decomposition) and `Python`/`SymPy` (for preprocessing and verification). Experiments were run on an Apple M4 MacBook Pro with 16 GB of unified memory. The M4 chip has a 10-core CPU and 10-core GPU, with hardware acceleration for polynomial arithmetic and linear algebra routines through Apple's Accelerate framework. Computations for the small- to medium-sized systems reported here (up to 9 variables and 5 equations) typically completed in under a few seconds per decomposition and Gröbner basis calculation.

Algorithm 2 Identifying Explanatory Axioms via Primary Decomposition

```
Input: Set of axioms $\mathcal{A} = \{A_1, ..., A_k\}$ over indeterminates $\mathcal{V} = \{x_1, ..., x_n\}$, hypothesis $Q$ over $\{x_1, ..., x_d\}$
Output: Set $\mathcal{Q}_{\text{expl}}$ of polynomials potentially deriving $Q$
Initialize $\mathcal{Q}_{\text{expl}} \leftarrow \emptyset$
Encode: Form system combining axioms and hypothesis
Define ideal $I \leftarrow \langle A_1, ..., A_k, Q\rangle$
Decompose: Identify components
Compute primary decomposition $I = Q_1 \cap \cdots \cap Q_r$
for each associated prime $P_j = \sqrt{Q_j}$ from the decomposition do
    for each generator $\widehat{A_{k+1_i}}$ of $P_j$ do
        Reason: Test whether generator explains $Q$
        Define ideal $J \leftarrow \langle A_1, ..., A_k, \widehat{A_{k+1_i}}\rangle$
        Compute Gröbner basis $\mathcal{G} = \text{GB}(J)$ with lex order $x_1 < \cdots < x_n$
        Let $\mathcal{G}' \leftarrow \mathcal{G} \cap \mathbb{R}[x_1, ..., x_d]$
        if $Q \in \mathcal{G}'$ then
            Add $\widehat{A_{k+1_i}}$ to $\mathcal{Q}_{\text{expl}}$
        end if
    end for
end for
Return $\mathcal{Q}_{\text{expl}}$
```

Table 6: Results with a number of axiom systems where one axiom is removed at a time

Problem	Axiom	Time	Recovered
Kepler 1	$(d_1+d_2)^2F_g - m_1m_2$	0.1s	✓
Kepler 2	$F_c - m_2d_2w^2$	0.1s	✓
Kepler 3	$F_c - F_g$	0.1s	✓
Kepler 4	$wp - 1$	0.1s	✓
Compton 1	$E_1 + Ee_1 - E_2 - Ee_2$	5.4s	✓
Compton 2	$E_1 - hf_1$	5.4s	✓
Compton 3	$E_2 - hf_2$	5.4s	✓
Compton 4	$p_1c - hf_1$	5.4s	✓
Compton 5	$p_2c - hf_2$	5.4s	✓
Compton 6	$\lambda_2f_1 - c$	5.4s	✓
Compton 7	$\lambda_2f_2 - c$	5.4s	✓
Compton 8	$Ee1 - m_c^2$	5.4s	✓
Compton 9	$Ee2^2 - c^2pe2^2 - me^2c^4$	5.4s	✓
Compton 10	$pe2^2 - p2^2 - p1^2 + 2p1p2cos$	5.4s	✓
Einstein 1	$cdt_0 - 2*d$	1.5s	✓
Einstein 2	$4L^2 - 4d^2 - v^2dt^2$	1.5s	✓
Einstein 3	$f_0dt_0 - 1$	1.5s	✓
Einstein 4	$fdt - 1$	1.5s	✓
Einstein 5	$c*dt - 2L$	1.5s	✓
Escape Velocity 1	$K_i - \frac{1}{2}mv_e^2$	0.4s	✓
Escape Velocity 2	$K_f = 0$	0.4s	✓
Escape Velocity 3	$U_ir + GMm$	0.4s	✓
Escape Velocity 4	$U_f = 0$	0.4s	✓
Escape Velocity 5	$K_i + U_i - (K_f + U_f)$	0.4s	✓
Light Damping 1	$Sr^2 - q_c^2a_p^2\sin\theta^2$	1.6s	✓
Light Damping 2	$dA - 2\pi r^2\sin\theta d\theta$	1.6s	✓
Light Damping 3	$P - \int_0^\pi SdA$	1.6s	✓
Light Damping 4	$\frac{4}{3} - \int_0^\pi \sin\theta^3 d\theta$	1.6s	✓
Light Damping 5	$a_p^2 - \frac{1}{2}w^4x_0^2$	1.6s	✓
Hagen Poiseuille 1	$u - c_0 - c_2r^2$	0.6s	✓
Hagen Poiseuille 2	$\mu\frac{\partial}{\partial r}(r\frac{\partial}{\partial r}u) - r\frac{dp}{dx}$	0.6s	✓
Hagen Poiseuille 3	$c_0 + c_2R^2$	0.6s	✓
Hagen Poiseuille 4	$L\frac{dp}{dx} = -\Delta p$	0.6s	✓
Neutrino Decay 1	$p_v - p_\mu$	3.5s	✓
Neutrino Decay 2	$E_\pi - m_\pi$	3.5s	✓
Neutrino Decay 3	$E_v - p_v$	3.5s	✓
Neutrino Decay 4	$E_\pi - E_\mu - E_v$	3.5s	✓
Neutrino Decay 5	$E_\mu^2 - p_\mu^2 - m_\mu^2$	3.5s	✓$^+$
Hall Effect 1	$F_m - q_evB$	11.1s	✓
Hall Effect 2	$F_e - q_eE$	11.1s	✓
Hall Effect 3	$F_m - F_e$	11.1s	✓
Hall Effect 4	$Eh - U_H$	11.1s	✓
Hall Effect 5	$vdt - L$	11.1s	✓
Hall Effect 6	$Idt - Q$	11.1s	✓
Hall Effect 7	$Q - Nq_e$	11.1s	✓
Hall Effect 8	$nV - N$	11.1s	X
Hall Effect 9	$V - Lhd$	11.1s	X