15-424: Foundations of Cyber-Physical Systems

Lecture Notes on Differential Equations & Differential Invariants

André Platzer

Carnegie Mellon University
Lecture 10

1 Introduction

So far, this course explored only one way to deal with differential equations: the [′] axiom from Lecture 5 on Dynamical Systems & Dynamic Axioms. However, in order to use the [′] axiom or its sequent calculus counterpart the [′] rule from Lecture 6 on Truth & Proof for a differential equation $x' = f(x)$, we must be able to find a symbolic solution to the symbolic initial value problem (i.e. a function $y(t)$ such that $y'(t) = f(y)$ and $y(0) = x$). But what if the differential equation does not have such a solution $y(t)$? Or if $y(t)$ cannot be written down in first-order real arithmetic? Lecture 2 on Differential Equations & Domains allows many more differential equations to be part of CPS models than just the ones that happen to have simple solutions. These are the differential equations we will look at in this lecture.

You may have seen a whole range of methods for solving differential equations in prior courses. But, in a certain sense, "most" differential equations are impossible to solve, because they have no explicit closed-form solution with elementary functions, for instance [Zei03]:

$$x''(t) = e^{t^2}$$

And even if they do have solutions, the solution may no longer be in first-order real arithmetic. One solution of

$$v' = w, w' = -v$$

for example is $v(t) = \sin t, w(t) = \cos t$, which is not expressible in real arithmetic (recall that both are infinite power series) and leads to undecidable arithmetic [Pla08a].

Today's lecture reinvestigates differential equations from a more fundamental perspective, which will lead to a way of proving properties of differential equations without using their solutions.

The lecture seeks unexpected analogies among the seemingly significantly different dynamical aspects of discrete dynamics and of continuous dynamics. The first and influential observation is that differential equations and loops have more in common than one might suspect.[1] Discrete systems may be complicated, but have a powerful ally: induction as a way of establishing truth for discrete dynamical systems by generically analyzing the one step that it performs (repeatedly like the body of a loop). What if we could use induction for differential equations? What if we could prove properties of differential equations directly by analyzing how these properties change along the differential equation rather than having to find a global solution first and inspecting whether it satisfies that property? What if we could tame the analytic complexity of differential equations by analyzing the generic local "step" that a continuous dynamical system performs (repeatedly). The biggest conceptual challenge will, of course, be in understanding what exactly the counterpart of a step even is for continuous dynamical systems, because there is no such thing as a next step for a differential equation.

More details can be found in [Pla10b, Chapter 3.5] and [Pla10a, Pla12d, Pla12a, Pla12b]. Differential invariants were originally conceived in 2008 [Pla10a, Pla08b] and later used for an automatic proof procedure for hybrid systems [PC08, PC09]. These lecture notes are based on an advanced axiomatic logical understanding of differential invariants via differential forms [Pla15].

This lecture is of central significance for the Foundations of Cyber-Physical Systems. The analytic principles begun in this lecture will be a crucial basis for analyzing all complex CPS. The most important learning goals of this lecture are:

Modeling and Control: This lecture will advance the core principles behind CPS by developing a deeper understanding of their continuous dynamical behavior. This lecture will also illuminate another facet of how discrete and continuous systems relate to one another, which will ultimately lead to a fascinating view on understanding hybridness [Pla12a].

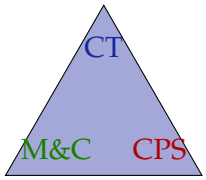
Computational Thinking: This lecture exploits computational thinking in its purest form by seeking and exploiting surprising analogies among discrete dynamics and continuous dynamics, however different both may appear at first sight. This lecture is devoted to rigorous reasoning about the differential equations in CPS models. Such rigorous reasoning is crucial for understanding the continuous behavior that CPS exhibit over time. Without sufficient rigor in their analysis it can be impossible to understand their intricate behavior and spot subtle flaws in their control or say for sure whether and why a design is no longer faulty. This lecture systematically develops one reasoning principle for equational properties of differential equations that is based on *induction for differential equations*. Subsequent

[1] In fact, discrete and continuous dynamics turn out to be proof-theoretically quite intimately related [Pla12a].

lectures expand the same core principles developed in this lecture to the study of general properties of differential equations. This lecture continues the *axiomatization* of differential dynamic logic dL [Pla12c, Pla12a] pursued since Lecture 5 on Dynamical Systems & Dynamic Axioms and lifts dL's proof techniques to systems with more complex differential equations. The concepts developed in this lecture form the differential facet illustrating the more general relation of *syntax* (which is notation), *semantics* (what carries meaning), and *axiomatics* (which internalizes semantic relations into universal syntactic transformations). These concepts and their relations jointly form the significant *logical trinity* of syntax, semantics, and axiomatics. Finally, the verification techniques developed in this lecture are critical for verifying CPS models of appropriate scale and technical complexity.

CPS Skills: We will develop a deeper understanding of the semantics of the continuous dynamical aspects of CPS models and develop and exploit a significantly better intuition for the operational effects involved in CPS.

discrete vs. continuous analogies
rigorous reasoning about ODEs
induction for differential equations
differential facet of logical trinity



understanding continuous dynamics
relate discrete+continuous

semantics of continuous dynamics
operational CPS effects

2 Global Descriptive Power of Local Differential Equations

Differential equations let the physics evolve continuously for longer periods of time. They describe such global behavior locally, however, just by the right-hand side of the differential equation.

Note 1 (Local descriptions of global behavior by differential equations)**.** *The key principle behind the descriptive power of differential equations is that they describe the evolution of a continuous system over time using only a local description of the direction into which the system evolves at any point in space. The solution of a differential equation is a global description of how the system evolves, while the differential equation itself is a local characterization. While the global behavior of a continuous system can be subtle and challenging, its local description as a differential equation is much simpler.*

This difference between local description and global behavior can be exploited for proofs.

Based on Lecture 2 on Differential Equations & Domains, the semantics of differential equations was defined in Lecture 3 on Choice & Control as:

Note 2 (Semantics of differential equations)**.**

$$[\![x' = f(x)\,\&\,Q]\!] = \{(\varphi(0), \varphi(r)) \,:\, \varphi(\zeta) \models x' = f(x) \text{ and } \varphi(\zeta) \models Q \text{ for all } 0{\leq}\zeta{\leq}r \text{ for a solution } \varphi : [0, r] \to \mathcal{S} \text{ of any duration } r\}$$

That is,[a] *the final state* $\varphi(r)$ *is connected to the initial state* $\varphi(0)$ *by a continuous function of some duration* $r \geq 0$ *that solves the differential equation and satisfies* Q *at all times, when interpreting* $\varphi(\zeta)(x') \overset{\text{def}}{=} \frac{\mathsf{d}\varphi(t)(x)}{\mathsf{d}t}(\zeta)$ *as the derivative of the value of* x *over time* t *at time* ζ*.*

[a] Two subtleties will ultimately give rise to a minor clarification. Can you foresee them already?

The solution φ describes the global behavior of the system, which is specified locally by the right-hand side $f(x)$ of the differential equation.

Lecture 2 has shown a number of examples illustrating the descriptive power of differential equations. That is, examples in which the solution was very complicated even though the differential equation was rather simple. This is a strong property of differential equations: they can describe even complicated processes in simple ways. Yet, that representational advantage of differential equations does not carry over into the verification when verification is stuck with proving properties of differential equations only by way of their solutions, which, by the very nature of differential equations, are more complicated again.

This lecture, thus, investigates ways of proving properties of differential equations using the differential equations themselves, not their solutions. This leads to *differential invariants* [Pla10a, Pla12d, Pla15], which can perform induction for differential equations.

3 Differential Equations vs. Loops

A programmatic way of developing an intuition for differential invariants leads through a comparison of differential equations with loops. This perhaps surprising relation can be made completely rigorous and is at the heart of a deep connection equating discrete and continuous dynamics proof-theoretically [Pla12a]. This lecture will stay at the surface of this surprising connection but still leverage the relation of differential equations to loops for our intuition.

To get started with relating differential equations to loops, compare

$$x' = f(x) \qquad \text{vs.} \qquad (x' = f(x))^*$$

How does the differential equation $x' = f(x)$ compare to the same differential equation in a loop $(x' = f(x))^*$ instead? Unlike the differential equation $x' = f(x)$, the repeated

differential equation $(x' = f(x))^*$ can run the differential equation $x' = f(x)$ repeatedly. Albeit, on second thought, does that get the repetitive differential equation $(x' = f(x))^*$ to any more states than where the differential equation $x' = f(x)$ could evolve to?

Not really, because chaining lots of solutions of differential equations from a repetitive differential equation $(x' = f(x))^*$ together will still result in a single solution for the same differential equation $x' = f(x)$ that we could have followed all the way.[2]

Note 3 (Looping differential equations). *$(x' = f(x))^*$ is equivalent to $x' = f(x)$, written $(x' = f(x))^* \equiv (x' = f(x))$, i.e. both have the same transition semantics:*

$$[\![(x' = f(x))^*]\!] = [\![x' = f(x)]\!]$$

Differential equations "are their own loop".[3]

In light of Note 3, differential equations already have some aspects in common with loops. Like nondeterministic repetitions, differential equations might stop right away. Like nondeterministic repetitions, differential equations could evolve for longer or shorter durations and the choice of duration is nondeterministic. Like in nondeterministic repetitions, the outcome of the evolution of the system up to an intermediate state influences what happens in the future. And, in fact, in a deeper sense, differential equations actually really do correspond to loops executing their discrete Euler approximations [Pla12a].

With this rough relation in mind, let's advance the dictionary translating differential equation phenomena into loop phenomena and back. The local description of a differential equation as a relation $x' = f(x)$ of the state to its derivative corresponds to the local description of a loop by a repetition operator * applied to the loop body α. The global behavior of a global solution of a differential equation $x' = f(x)$ corresponds to the full global execution trace of a repetition α^*, but are similarly unwieldy objects to handle. Because the local descriptions are so much more concise than the respective global behaviors, but still carry all information about how the system will evolve over time, we also say that the local relation $x' = f(x)$ is the *generator* of the global system solution and that the loop body α is the *generator* of the global behavior of repetition of the loop. Proving a property of a differential equation in terms of its solution corresponds to proving a property of a loop by unwinding it (infinitely long) by axiom $[*]$ from Lecture 5 on Dynamical Systems & Dynamic Axioms.

Now, Lecture 7 on Control Loops & Invariants made the case that unwinding the iterations of a loop can be a rather tedious way of proving properties about the loop, because there is no good way of ever stopping to unwind, unless a counterexample can be found after a finite number of unwindings. This is where working with a global solution of a differential equation with axiom $[']$ is actually already more useful, because the solution can actually be handled completely because of the quantifier $\forall t \geq 0$

[2] This is related to classical results about the continuation of solutions, e.g., [Pla10b, Proposition B.1].

[3] Beware not to confuse this with the case for differential equations with evolution domain constraints, which is subtly different (Exercise 1).

Note 4 (Correspondence map between loops and differential equations).

loop α^*	***differential equation*** $x' = f(x)$
could repeat 0 times	*could evolve for duration 0*
repeat any number $n \in \mathbb{N}$ *of times*	*evolve for any duration* $r \in \mathbb{R}, r \geq 0$
effect depends on previous loop iteration	*effect depends on the past solution*
local generator α	*local generator* $x' = f(x)$
full global execution trace	*global solution* $\varphi : [0, r] \to \mathcal{S}$
proof by unwinding iterations with axiom $[*]$	*proof by global solution with axiom* $[']$
proof by induction with loop invariant rule loop	*proof by differential invariant*

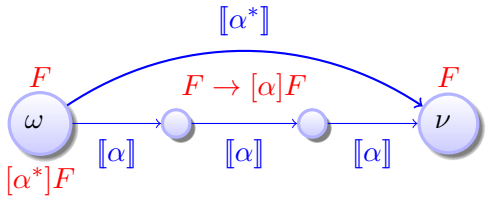
over all durations. But Lecture 7 introduced induction with invariants as the preferred way of proving properties of loops, by, essentially, cutting the loop open and arguing that the generic state after any run of the loop body has the same characterization as the generic state before. After all these analogous correspondences between loops and differential equations, the obvious question is what the differential equation analogue of a proof concept would be that corresponds to proofs by induction for loops, which is the premier technique for proving loops.

Induction can be defined for differential equations using what is called *differential invariants* [Pla10a, Pla12d]. The have a similar principle as the proof rules for induction for loops. Differential invariants prove properties of the solution of the differential equation using only its local generator: the right-hand side of the differential equation.

Recall the loop induction proof rule from Lecture 7 on Loops & Invariants:

$$[\![\alpha^*]\!] = \bigcup_{n\in\mathbb{N}} [\![\alpha^n]\!] \qquad \text{with} \quad \alpha^{n+1} \equiv \alpha^n;\alpha \text{ and } \alpha^0 \equiv ?true$$

$$\text{loop}\ \frac{\Gamma \vdash F, \Delta \quad F \vdash [\alpha]F \quad F \vdash P}{\Gamma \vdash [\alpha^*]P, \Delta}$$



4 Intuition of Differential Invariants

Just as inductive invariants are the premier technique for proving properties of loops, differential invariants [Pla10a, Pla12d, Pla08b, Pla10b] provide the primary inductive technique we use for proving properties of differential equations (without having to solve them).

The core principle behind loop induction is that the induction step investigates the local generator α ands shows that it never changes the truth-value of the invariant F (see the middle premise $F \vdash [\alpha]F$ of proof rule loop or the only premise of the core

essentials induction proof rule ind from Lecture 7). Let us try to establish the same inductive principle, just for differential equations. The first and third premise of rule loop transfer easily to differential equations. The challenge is to figure out what the counterpart of $F \vdash [\alpha]F$ would be since differential equations do not have a notion of "one step".

What does the local generator of a differential equation $x' = f(x)$ tell us about the evolution of a system? And how does it relate to the truth of a formula F all along the solution of that differential equation? That is, to the truth of the dL formula $[x' = f(x)]F$ expressing that all runs of $x' = f(x)$ lead to states satisfying F. Fig. 1 depicts an example of a vector field for a differential equation (plotting the right-hand side of the differential equation as a vector at every point in the state space), a global solution (in red), and an unsafe region $\neg F$ (shown in blue). The safe region F is the complement of the blue unsafe region $\neg F$.

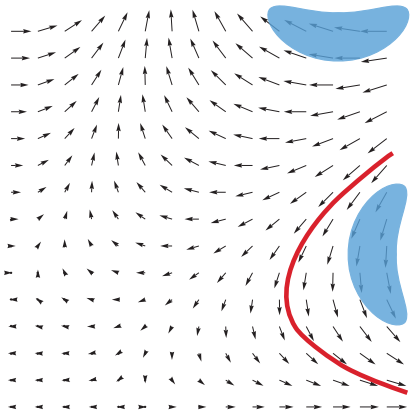


Figure 1: Vector field and one solution of a differential equation that does not enter the blue unsafe regions

One way of proving that $[x' = f(x)]F$ is true in a state ω would be to compute a solution from that state ω, check every point in time along the solution to see if it is in the safe region F or the unsafe region $\neg F$. Unfortunately, these are uncountably infinitely many points in time to check. Furthermore, that only considers a single initial sate ω, so proving validity of a formula would require considering every of the uncountably infinitely many possible initial states and computing and following a solution in each of them. That is why this naïve approach would not compute.

A similar idea can still be made to work when the symbolic initial-value problem can be solved with a symbolic initial value x and a quantifier for time can be used, which is what the solution axiom [′] does. Yet, even that only works when a solution to the symbolic initial-value problem can be computed and the arithmetic resulting from the quantifier for time can be decided. For polynomial solutions, this works, for example. But polynomial come from very simple systems only (called nilpotent linear differential equation systems).

Reexamining the illustration in Fig. 1, we suggest an entirely different way of check-

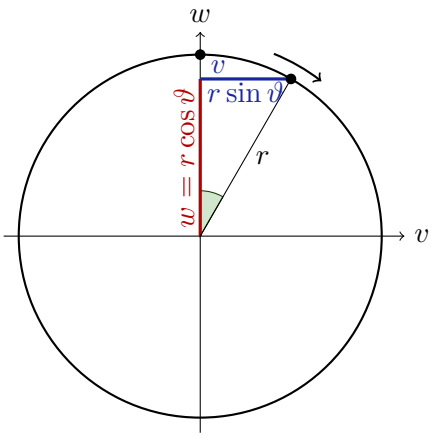


Figure 2: One scenario for the rotational dynamics and relationship of vector (v, w) to radius r and angle ϑ

ing whether the system could ever lead to an unsafe state in $\neg F$ when following the differential equation $x' = f(x)$. The intuition is the following. If there were a vector in Fig. 1 that points from a safe state in F to an unsafe state $\neg F$ (in the blue region), then following that vector could get the system into an unsafe $\neg F$. If, instead, all vectors point from safe states to safe states in F, then, intuitively, following such a chain of vectors will only lead from safe states to safe states. So if the system also started in a safe state, it would stay safe forever.

Let us make this intuition rigorous to obtain a sound proof principle that is perfectly reliable in order to be usable in CPS verification. What we need to do is to find a way of characterizing how the truth of F changes when moving along the differential equation.

5 Deriving Differential Invariants

How can the intuition about directions of evolution of a logical formula F with respect to a differential equation $x' = f(x)$ be made rigorous? We develop this step by step.

Example 1 (Rotational dynamics). As a guiding example, consider a conjecture about the rotational dynamics where v and w represent the direction of a vector rotating clockwise in a circle of radius r (Fig. 2):

$$v^2 + w^2 = r^2 \to [v' = w, w' = -v]v^2 + w^2 = r^2 \tag{1}$$

The conjectured $\mathsf{d}\mathcal{L}$ formula (1) is valid, because, indeed, if the vector (v, w) is initially at distance r from the origin (0,0), then it will always be when rotating around the origin, which is what the dynamics does. That is, the point (v, w) will always remain on the circle of radius r. But how can we prove that? In this particular case, we could possibly investigate solutions, which are trigonometric functions (although the ones shown in Fig. 2 are not at all the only solutions). With those solutions, we could perhaps find an argument why they stay at distance r from the origin. But the resulting

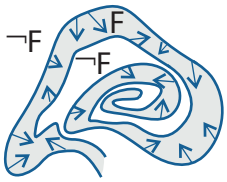


Figure 3: Differential invariant F remains true in the direction of the dynamics

arithmetic will be unnecessarily difficult and, after all, the argument for why the simple dℒ formula (1) is valid should be easy. And it is, after we have discovered the right proof principle as this lecture will do.

First, what is the direction into which a continuous dynamical system evolves? The direction is exactly described by the differential equation, because the whole point of a differential equation is to describe in which direction the state evolves at every point in space. So the direction into which a continuous system obeying $x' = f(x)$ follows from state ω is exactly described by the time-derivative, which is exactly the value $[\![f(x)]\!]\omega$ of term $f(x)$ in state ω. Recall that the term $f(x)$ can mention x and other variables so its value $[\![f(x)]\!]\omega$ depends on the state ω.

Note 5 (Differential invariants are "formulas that remain true in the direction of the dynamics"). *Proving dℒ formula $[x' = f(x)]F$ does not really require us to answer where exactly the system evolves to but just how the evolution of the system relates to the formula F and the set of states ω in which F evaluates to true. It is enough to show that the system only evolves into directions in which formula F will stay true (Fig. 3).*

A logical formula F is ultimately built from atomic formulas that are comparisons of (polynomial or rational) terms such as $e = 5$ or $v^2 + w^2 = r^2$. Let e denote such a (polynomial) term in the variable (vector) x that occurs in the formula F. The semantics of a polynomial term e in a state ω is the real number $[\![e]\!]\omega$ that it evaluates to. In which direction does the value of e evolve when following the differential equation $x' = f(x)$ for some time? That depends both on the term e that is being evaluated and on the differential equation $x' = f(x)$ that describes how the respective variables x evolve over time.

Note 6. *Directions of evolutions are described by derivatives, after all the differential equation $x' = f(x)$ describes that the time-derivative of x is $f(x)$.*

Let's derive the term e of interest and see what that tells us about how e evolves over time. How can we derive e? The term e could be built from any of the operators discussed in Lecture 2 on Differential Equations & Domains, to which we now add division for rational terms to make it more interesting. Let $\mathcal{V}$ denote the set of all variables. Recall from Lecture 2 that *terms* e are defined by the grammar (where $e, \tilde{e}$ are terms, x is a variable, and c is a rational number constant):

$$e ::= x \mid c \mid e + \tilde{e} \mid e - \tilde{e} \mid e \cdot \tilde{e} \mid e/\tilde{e}$$

It is, of course, important to take care that division $e/\tilde{e}$ only makes sense in a context where the divisor $\tilde{e}$ is guaranteed not to be zero in order to avoid undefinedness. *We only allow division to be used in a context where the divisor is ensured not to be zero.*

If the term is a sum $e + k$, then the mathematical expectation is that its derivative should be the derivative of e plus the derivative of k. If the term is a product $e \cdot k$, its derivative is the derivative of e times k plus e times the derivative of k by Leibniz' rule. The derivative of a rational number constant $c \in \mathbb{Q}$ is zero.[4] The other operators are similar, leaving only the case of a single variable x. What is its derivative?

Before you read on, see if you can find the answer for yourself.

[4]Of course, the derivative of real number constants $c \in \mathbb{R}$ is also zero, but only rational number constants are allowed to occur in the formulas of first-order logic of real arithmetic (or any real-closed fields).

The exact value of the derivative of a variable x very much depends on the current state and on the overall continuous evolution of the system. So we punt on that for now and define the derivative of a variable x in a seemingly innocuous way to be the differential symbol x' and consider what to do with it later. This gives rise to the following way of computing the derivative of a term syntactically.

Remark 2 (Derivatives). Recall the familiar syntactic laws for derivatives:

$$(c())' = 0 \qquad \text{for numbers or constants } c() \tag{2a}$$
$$(x)' = x' \qquad \text{for variable } x \in \mathcal{V} \tag{2b}$$
$$(e+k)' = (e)' + (k)' \tag{2c}$$
$$(e-k)' = (e)' - (k)' \tag{2d}$$
$$(e \cdot k)' = (e)' \cdot k + e \cdot (k)' \tag{2e}$$
$$(e/k)' = ((e)' \cdot k - e \cdot (k)')/k^2 \tag{2f}$$

Note that, while the intuition and precise semantics of derivatives of terms will ultimately be connected with more involved aspects of how values change over time, the computation of derivatives of terms according to 2 is a seemingly innocuous but straightforward recursive computation on terms. If we apply the equations (6) from left to right, they define a recursive operator on terms $(\cdot)'$ called *syntactic (total) derivation*.

Expedition 1 (Differential Algebra)**.** Even though the following names and concepts are not needed directly for his course, let's take a brief scientific expedition to align 2 with the algebraic structures from differential algebra [Kol72] in order to illustrate the systematic principles behind 2. Case (6a) defines (rational) number symbols alias literals as *differential constants*, which do not change their value during continuous evolution. Their derivative is zero. The number symbol 5 will always have the value 5 and never change, no matter what differential equation is considered. Equation (6c) and the *Leibniz* or *product rule* (6e) are the defining conditions for *derivation operators on rings*. The derivative of a sum is the sum of the derivatives (additivity or a homomorphic property with respect to addition, i.e. the operator $(\cdot)'$ applied to a sum equals the sum of the operator applied to each summand) according to equation (6c). Furthermore, the derivative of a product is the derivative of one factor times the other factor plus the one factor times the derivative of the other factor as in (6e). Equation (6d) is a derived rule for subtraction according to the identity $e - k = e + (-1) \cdot k$ and again expresses a homomorphic property, now with respect to subtraction rather than addition.

The equation (6b) uniquely defines the operator $(\cdot)'$ on the *differential polynomial algebra* spanned by the *differential indeterminates* $x \in \mathcal{V}$, i.e. the symbols x that have indeterminate derivatives x'. It says that we understand the differential symbol x' as the derivative of the symbol x for all state variables $x \in \mathcal{V}$. Equation (6f) canonically extends the derivation operator $(\cdot)'$ to the *differential field of quotients* by the usual *quotient rule*. As the base field $\mathbb{R}$ has no zero divisors[a], the right-hand side

of (6f) is defined whenever the original division e/k can be carried out, which, as we assumed for well-definedness, is guarded by $k \neq 0$.

[a] In this setting, $\mathbb{R}$ has no zero divisors, because the formula $ab = 0 \rightarrow a = 0 \vee b = 0$ is valid, i.e. a product is zero only if a factor is zero.

The derivative of a division e/k uses a division, which is where we need to make sure not to accidentally divide by zero. Yet, in the definition of $(e/k)'$, the division is by k^2 which, fortunately, has the same roots that k already has, because $k = 0 \leftrightarrow k^2 = 0$ is valid for any term k. Hence, in any context in which e/k was defined, its derivative $(e/k)'$ will also be defined.

Now that we have a first definition of derivation at hand, the question still is which of the terms should be derived when trying to prove (1)? Since that is not necessarily clear so far, let's turn the formula (1) around and consider the following equivalent (Exercise 2) dL formula instead, which only has a single nontrivial term to worry about:

$$v^2 + w^2 - r^2 = 0 \rightarrow [v' = w, w' = -v]v^2 + w^2 - r^2 = 0 \tag{3}$$

Derivation of the only relevant term $v^2 + w^2 - r^2$ in the postcondition of (3) according to 2 gives

$$(v^2 + w^2 - r^2)' = 2vv' + 2ww' - 2rr' \tag{4}$$

2 makes it possible to form the derivative of any polynomial or rational term. The total derivative operator $(\cdot)'$ does *not*, however, result in a term involving the variables $\mathcal{V}$, but, instead, a *differential term*, i.e. a term involving $\mathcal{V} \cup \mathcal{V}'$, where $\mathcal{V}' \overset{\text{def}}{=} \{x' : x \in \mathcal{V}\}$ is the set of all differential symbols x' for variables $x \in \mathcal{V}$. The total derivative $(e)'$ of a polynomial term e is not a polynomial term, but may mention differential symbols such as x' in addition to the symbols that where in e to begin with. All syntactic elements of those differential terms are easy to interpret based on the semantics of terms defined in Lecture 2, except for the differential symbols. What now is the meaning of a differential symbol x'? And, in fact, what is the precise meaning of the construct $(e)'$ for a term e and the equations in (6) to begin with?

Before you read on, see if you can find the answer for yourself.

6 The Meaning of Prime

The meaning $[\![x]\!]\omega$ of a variable symbol x is defined by the state ω as $\omega(x)$, so its value $[\![x]\!]\omega$ in state ω is directly determined by the state via $[\![x]\!]\omega = \omega(x)$. It is crucial to notice the significant subtleties and challenges that arise when trying to give meaning to a differential symbol x' or anything else with a derivative connotation such as the differential term $(e)'$ of term e.

The first mathematical reflex may be to set out for a definition of x' in terms of a time-derivative $\frac{\mathsf{d}}{\mathsf{d}t}$ of something. The question is what that something would be. The meaning of a differential symbol x' in a state ω simply cannot be defined as a time-derivative, because derivatives do not even exist in such isolated points. It is utterly meaningless to ask for the rate of change of the value of x over time in a single isolated state ω. For time-derivatives to make sense, we at least need a concept of time and the values understood as a function of time. That function needs to be defined on a big enough interval for derivatives to have a chance to become meaningful. And the function needs to be differentiable so that the time-derivatives even exist to begin with.

Expedition 2 (Semantics of differential algebra)**.** The view of Expedition 1 sort of gave $(e)'$ a meaning, but, when we think about it, did not actually define it. Differential algebra studies the structural algebraic relations of, e.g., the derivative $(e+k)'$ to the derivatives $(e)'$ plus $(k)'$ and is incredibly effective about capturing and understanding them starting from (6). But algebra—and differential algebra is no exception—is, of course, deliberately abstract about the question what the individual pieces mean, because algebra is the study of structure, not the study of the meaning of the objects that are being structured in the first place. That is why we can learn all about the structure of derivatives and derivation operators from differential algebra, but have to go beyond differential algebra to complement it with a precise semantics that relates to the needs of understanding the mathematics of real cyber-physical systems.

Along a (differentiable) continuous function $\varphi : [0, r] \to \mathcal{S}$, however, we can make sense of what x' means. And in fact we already did. Well, if its duration $r > 0$ is nonzero so that we are not just talking about an isolated point $\varphi(0)$ again. At any point in time $\zeta \in [0, r]$ along such a continuous evolution φ, the differential symbol x' can be taken to mean the time-derivative $\frac{\mathsf{d}}{\mathsf{d}t}$ of the value $[\![x]\!]\varphi(t)$ of x over time t at time ζ [Pla10a, Pla12c, Pla15]. That is, at any point in time ζ along the solution φ, it makes sense to give x' the meaning of the rate of change of the value of x over time along φ. Which is exactly what the semantics of differential equations from Note 2 already did to give meaning to the differential equation in the first place:

Note 7 (Semantics of differential symbols along a differential equation)**.** *The value of differential symbol x' at time $\zeta \in [0,r]$ along a solution $\varphi : [0,r] \to \mathcal{S}$ of some duration $r > 0$ of a differential equation $x' = f(x)\,\&\,Q$ equals the analytic time-derivative at ζ:*

$$\varphi(\zeta)(x') \overset{\text{def}}{=} \frac{\mathsf{d}\varphi(t)(x)}{\mathsf{d}t}(\zeta) \tag{5}$$

Intuitively, the value $[\![x']\!]\varphi(\zeta) = \varphi(\zeta)(x')$ of x' is, thus, determined by considering how the value $[\![x]\!]\varphi(\zeta) = \varphi(\zeta)(x)$ of x changes along the function φ when we change time ζ "only a little bit". Visually, it corresponds to the slope of the tangent of the value of x at time ζ; see Fig. 4.

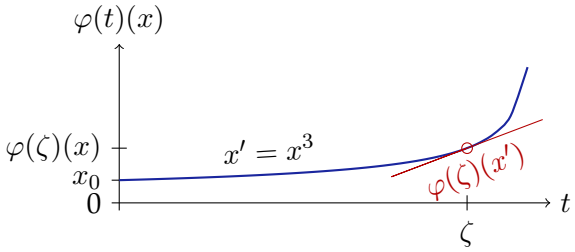


Figure 4: Semantics of differential symbols

Now that we know what value x' would have along a differential equation that leaves at least two questions. What does it mean independently without reference to the particular differential equation? And what value would it have along a differential equation of duration $r = 0$ where the right-hand side of (5) but not even exist?

While the latter question may be the more obvious one, the more daunting one is the former question. What does x' mean? What is its value? Since the differential symbol x' is a term and the semantics of terms is the real-value that they mean in a state ω, the differential symbol x' should also have a meaning as a real number in that state ω. So what is the value $\omega(x')$?

Hold on, we had considered and discarded that question already. Derivatives do not carry meaning in isolated states. They still don't. But it is important to understand why the lack of having a value and a meaning would cause complications for the fabrics of logic.

Expedition 3 (Denotational Semantics)**.** The whole paradigm of *denotational semantics*, initiated for programming languages by Dana Scott and Christopher Strachey [SS71], is based on the principle that the semantics of expressions of programming languages should be the mathematical object that it denotes. That is, a denotational semantics is a function assigning a mathematical object $[\![e]\!]\omega$ from a semantic domain (here $\mathbb{R}$) to each term e, depending on the state ω.

The *meaning of terms*, thus, is a function $[\![\cdot]\!] : \mathrm{Trm} \to (\mathcal{S} \to \mathbb{R})$ which maps each

term $e \in \mathrm{Trm}$ to the function $[\![e]\!] : \mathcal{S} \to \mathbb{R}$ giving the real value $[\![e]\!]\omega \in \mathbb{R}$ that the term e has in each state $\omega \in \mathcal{S}$. In fact, this is exactly how the semantics of terms of dL has been defined in Lecture 2 in the first place. For classical logics such as first-order logic, this denotational semantics has always been the natural and dominant approach since Frege [**?**].

Scott and Strachey [SS71], however, pioneered the idea of leveraging the denotational style of semantics to give meaning to programming languages. And, indeed, dL's hybrid programs have a denotational semantics. The meaning of a HP α is the reachability relation $[\![\alpha]\!] \subseteq \mathcal{S} \times \mathcal{S}$ that it induces on the states $\mathcal{S}$. Correspondingly, the (denotational) *meaning of hybrid programs* as defined in Lecture 3 is a function $[\![\cdot]\!] : \mathrm{HP} \to \wp(\mathcal{S} \times \mathcal{S})$ assigning a relation $[\![\alpha]\!] \subseteq \mathcal{S} \times \mathcal{S}$ in the powerset $\wp(\mathcal{S} \times \mathcal{S})$ to each HP α.

A crucial feature of denotational semantics done the right way, however, is *compositionality*. The meaning $[\![e + \tilde{e}]\!]$ of a compound such as $e + \tilde{e}$ should be a simple function of the meaning $[\![e]\!]$ and $[\![\tilde{e}]\!]$ of its pieces e and $\tilde{e}$. This compositionality is exactly the way the meaning of differential dynamic logic is defined. For example:

$$[\![e + \tilde{e}]\!]\omega = [\![e]\!]\omega + [\![\tilde{e}]\!]\omega$$

for all states ω, which, with a point-wise understanding of $+$, can be summarized as

$$[\![e + \tilde{e}]\!] = [\![e]\!] + [\![\tilde{e}]\!]$$

Consequently, also the meaning of a differential symbol x' should be defined compositionally in a modular fashion and without reference to outside elements such as the differential equation in which it happens to occur. The meaning of terms is a function of the state, and not a function of the state and the context or purpose for which it happens to have been mentioned.[5] The actual values that x' is supposed to evaluate to changes quite a bit depending on the state, e.g. according to (5).

The mystery of giving meaning to differential symbols, thus, resolves by declaring the state to be responsible for assigning a value not just to all variables x but also to all differential symbols x'. A *state* ω is a mapping $\omega : \mathcal{V} \cup \mathcal{V}' \to \mathbb{R}$ assigning a real number $\omega(x) \in \mathbb{R}$ to all variables $x \in \mathcal{V}$ and a real number $\omega(x') \in \mathbb{R}$ to all differential symbols $x' \in \mathcal{V}'$. The values that the states $\varphi(\zeta)$ visited along a solution $\varphi : [0, r] \to \mathcal{S}$ of a differential equation $x' = f(x) \,\&\, Q$ assign to x' will have a close relationship, namely (5) and $\varphi(\zeta) \models x' = f(x)$. But that relationship is by virtue of φ being a solution of a differential equation, so that the family of states $\varphi(\zeta)$ for $\zeta \in [0, r]$ have a unique link. It is perfectly consistent to have a state ω in which $\omega(x') = 4$ and other equally isolated state ν in which $\nu(x') = 16$. In fact, that is just what happen for the initial ω and final state ν of following the differential equation $x' = x^2$ from $\omega(x) = 2$ for $\frac{1}{4}$ time units. If we do not know that ω and ν are the initial and final states of that differential equation

[5] With sufficient care, it would even be possible to restrict the meaning of x' only to certain contexts for some purposes, but that comes at the cost of adding significant technical complexity and inconvenience [Pla10a] and is, thus, quite an undesirable and unnecessary complication.

or if we do not know that it was exactly $\frac{1}{4}$ time units that we followed it, there is no reason to suspect any relationship between the values of $\omega(x')$ and $\nu(x')$.

Now we finally figured out the answer to the question what x' means and what its value is. It all depends on the state. And nothing but the state. So we can come back to the question what the value of x' would be along a differential equation that we followed for duration $r = 0$. The right-hand side of (5) does not exist if $r = 0$ (which, for duration $r = 0$ we will take to mean as not imposing any conditions). But the semantics of differential equations (Note 2) still unambiguously demands that $\varphi(\zeta) \models x' = f(x)$ holds during the solution including at the end at time r, that is, $\varphi(r)(x') = [\![f(x)]\!]\varphi(r)$.

Note 8 (Solutions of duration zero). *In case of duration $r = 0$, the only condition for the transition of a continuous evolution is that the initial ω and final state ν agree (except[a] on $\{x'\}^{\complement}$) and that $\nu(x') = [\![f(x)]\!]\nu$.*

[a] In fact, turns out to be useful [Pla15] to allow any arbitrary value of x' in the initial state ω of a continuous evolution since the previous value of x' may not yet be in sync with the expected derivative in the differential equation $x' = f(x)$ yet.

Differential symbols x' have a meaning from now as being interpreted directly by the state. Yet, what is the meaning of a differential term $(e)'$ such as those in 2?

Before you read on, see if you can find the answer for yourself.

7 More Meanings of More Primes: Differentials

At this point it should no longer be a surprise that the first mathematical reflex of understanding the primes of $(e)'$ from 2 in terms of time-derivatives will quickly fall short of its own expectations, because there still is no time-derivative in the isolated state ω that the value $[\![(e)']\!]\omega$ has at its disposal. Unfortunately, though, we cannot follow the same solution and ask the state to assign any arbitrary real value to each differential term. After all, there should be a close relationship of $[\![(2x^2)']\!]\omega$ and $[\![(8x^2)']\!]\omega$ namely that $4[\![(2x^2)']\!]\omega = [\![(8x^2)']\!]\omega$. Thus, the structure and meaning of the term e should contribute to the meaning of $(e)'$. The first step, though, is to ennoble the primes of $(e)'$ as in 2 and officially consider them as part of the language of differential dynamic logic by adding them to its syntax.

Definition 3 (d$\mathcal{L}$ Terms)**.** A *term* e of differential dynamic logic is defined by the grammar (where $e, \tilde{e}$ are terms, x a variable with corresponding differential symbol x', and c a rational number constant):

$$e ::= x \mid x' \mid c \mid e + \tilde{e} \mid e - \tilde{e} \mid e \cdot \tilde{e} \mid e/\tilde{e} \mid (e)'$$

The semantics of terms, of course, remains unchanged except that the new addition of differential terms $(e)'$ needs to be outfitted with a proper meaning.

The value of $(e)'$ is supposed to tell us something about how the value of e changes. But it is not and could not possibly be the change over time that this is referring to, because there is no time nor time-derivative to speak of in an isolated state ω. The trick is that we can still determine how the value of e will change, just not over time. We can tell just from the term e itself how its value will change locally depending on how its constituents change.

Recall that the partial derivative $\frac{\partial f}{\partial x}(\xi)$ of a function f by variable x at the point ξ characterizes how the value of f changes as x changes at the point ξ. The term $2x^2$ will locally change according to the partial derivative of its value by x, but the ultimate change will also depend on how x itself changes locally. The term $5x^2y$ also changes according to the partial derivative of its value by x but also its partial derivative by y and ultimately depends on how x as well as y themselves change locally.

The clou is that the state ω has the values $\omega(x')$ of the differential symbols x' at its disposal, which, qua Note 7, are reminiscent of the direction that x would be evolving to locally, if only state ω were part of a solution of a differential equation. The value $\omega(x')$ of differential symbol x' acts like the "local shadow" of the time-derivative $\frac{\mathsf{d}x}{\mathsf{d}t}$ at ω if only that derivative even existed at that point to begin with. But even if that time-derivative cannot exist at a general isolated state, we can still understand the value $\omega(x')$ that x' happens to have in that state as the direction that x would involve in locally at that state. Likewise the value $\omega(y')$ of y' can be taken to indicate the direction that y would involve in locally at that state. Now all it takes is a way to accumulate the change by summing it all up to lead to the meaning of differentials [Pla15].

Definition 4 (Semantics of differentials)**.** The semantics of differential term $(e)'$ in state ω is the value $[\![(e)']\!]\omega$ defined as

$$[\![(e)']\!]\omega = \sum_{x\in\mathcal{V}} \omega(x') \frac{\partial [\![e]\!]}{\partial x}(\omega)$$

The value $[\![(e)']\!]\omega$ is the sum of all (analytic) spatial partial derivatives at ω of the value of e by all variables $x \in \mathcal{V}$ multiplied by the corresponding tangent or direction of evolution described by the value $\omega(x')$ of differential symbol $x' \in \mathcal{V}'$.

That sum over all variables $x \in \mathcal{V}$ has finite support, because e only mentions finitely many variables x and the partial derivative by variables x that do not occur in e is 0, so do not contribute to the sum. The spatial derivatives exist since the evaluation $[\![e]\!]\omega$ is a composition of smooth functions such as addition, multiplication etc., so is itself smooth.

Overall the (real) value of $(e)'$, thus, depends not just on e itself and the values in the current state ω of the variables x that occur in e but also on the direction that these variables are taken to evolve to according to the values of the respective differential symbols x' in ω; see Fig. 5.

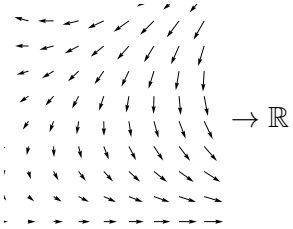


Figure 5: Differential form semantics of differentials: their value depends on the point as well as on the direction of the vector field at that point

Quite crucially observe one byproduct of adopting differentials as first-class citizens in $\mathsf{d}\mathcal{L}$. The constructs in 2, which previously were somewhat amorphous and semantically undefined recursive syntactic definitions without proper semantic counterparts, have now simply become perfectly meaningful equations of differential terms. The meaning of equations is well-defined on reals and both sides of each of the equations in (6) has a precise meaning using Def. 4. Of course, it remains to show that the equations in (6) are valid, meaning they are true in all states so that they can be adopted as sound axioms. But that turns out to be the case [Pla15].

Lemma 5 (Derivation lemma)**.** *When considered as equations of differentials, the equations* (6) *from* 2 *are valid and can, thus, be adopted as sound axioms:*

$$(c())' = 0 \qquad \text{for numbers or constants } c() \tag{6a}$$
$$(x)' = x' \qquad \text{for variable } x \in \mathcal{V} \tag{6b}$$
$$(e+k)' = (e)' + (k)' \tag{6c}$$
$$(e-k)' = (e)' - (k)' \tag{6d}$$
$$(e \cdot k)' = (e)' \cdot k + e \cdot (k)' \tag{6e}$$
$$(e/k)' = ((e)' \cdot k - e \cdot (k)')/k^2 \tag{6f}$$

Proof. We only consider one case of the proof which is reported in full elsewhere [Pla15].

$$\begin{aligned}
[\![(e+k)']\!]\omega &= \sum_x \omega(x') \frac{\partial [\![e+k]\!]}{\partial x}(\omega) = \sum_x \omega(x') \frac{\partial([\![e]\!] + [\![k]\!])}{\partial x}(\omega) \\
&= \sum_x \omega(x') \Big(\frac{\partial [\![e]\!]}{\partial x}(\omega) + \frac{\partial [\![k]\!]}{\partial x}(\omega) \Big) \\
&= \sum_x \omega(x') \frac{\partial [\![e]\!]}{\partial x}(\omega) + \sum_x \omega(x') \frac{\partial [\![k]\!]}{\partial x}(\omega) \\
&= [\![(e)']\!]\omega + [\![(k)']\!]\omega = [\![(e)' + (k)']\!]\omega
\end{aligned}$$

□

This gives us a way of computing simpler forms for differentials of terms by applying the equations (6) from left to right, which will, incidentally, lead us to the same result that the total derivation operator would have. Except now the result has been obtained by a chain of logical equivalence transformations each of which are individually grounded semantically with a soundness proof. It also becomes possible to selectively apply equations of differentials as need be in a proof without endangering soundness. Who would have figured that our study of differential equations would lead us down a path to study equations of differentials instead?

8 Differential Substitution Lemmas

Now that we obtained a precise semantics of differential symbols x' and differentials $(e)'$ that is meaningful in any arbitrary state ω, no matter how isolated it may be, it is about time to come back to the question what we can learn from those values along a differential equation.

Along the solution φ of a differential equation, differential symbols x' do not have arbitrary values but are interpreted as time-derivatives of the values of x at all times ζ

(Note 7):

$$[\![x']\!]\varphi(\zeta) = \varphi(\zeta)(x') \overset{\text{def}}{=} \frac{\mathsf{d}\varphi(t)(x)}{\mathsf{d}t}(\zeta) = \frac{\mathsf{d}[\![x]\!]\varphi(t)}{\mathsf{d}t}(\zeta)$$

That is, along a differential equation, the values of differential symbols x' coincide with the analytic time-derivative of the values of x. The key insight is that this continues to hold not just for differential symbols x' but also for differentials $(e)'$ of arbitrary terms e.

The following central lemma, which is the differential counterpart of the substitution lemma, establishes the connection between syntactic derivation of terms and semantic differentiation as an analytic operation to obtain analytic derivatives of valuations along differential state flows. It will allow us to draw analytic conclusions about the behaviour of a system along differential equations from the truth of purely algebraic formulas obtained by syntactic derivation. In a nutshell, the following lemma shows that, along a flow, analytic derivatives of valuations coincide with valuations of syntactic derivations.

Lemma 6 (Differential lemma)**.** *Let $\varphi \models x' = f(x) \wedge Q$ for some solution $\varphi : [0, r] \to \mathcal{S}$ of duration $r > 0$. Then for all terms e (defined all along φ) and all times $\zeta \in [0, r]$:*

$$[\![(e)']\!]\varphi(\zeta) = \frac{\mathsf{d}\,[\![e]\!]\varphi(t)}{\mathsf{d}t}(\zeta)$$

In particular, $[\![e]\!]\varphi(\zeta)$ is continuously differentiable in ζ.

Note 13 (The differential lemma clou)**.** *Lemma 6 shows that analytic time-derivatives coincide with differentials. The clou with Lemma 6 is that it equates precise but sophisticated analytic time-derivatives with purely syntactic differentials. The analytic timederivatives on the right-hand side of Lemma 6 are mathematically precise and pinpoint exactly what we are interested in: the rate of change of the value of e along φ. But they are unwieldy for computers, because analytic derivatives are ultimately defined in terms of limit processes and also need a whole solution to be well-defined. The syntactic differentials on the left-hand side of Lemma 6 are purely syntactic (putting a prime on a term) and even their simplifications via the recursive use of the axioms (6) are computationally tame.*

Having said that, the syntactic differentials need to be aligned with the intended analytic time-derivatives, which is exactly what Lemma 6 achieves. To wit, even differentiating polynomials and rational functions is much easier syntactically than by unpacking the meaning of analytic derivatives in terms of limit processes.

The differential lemma immediately leads to a first proof principle for differential equations. If the differential $(e)'$ is always zero along a differential equation, then e will always be zero if it was zero initially:

Lemma 7 (First version of differential invariant rule)**.** *The following is a sound proof rule*

$$\text{DI}_0\ \frac{\vdash [x' = f(x)\,\&\,Q](e)' = 0}{e = 0 \vdash [x' = f(x)\,\&\,Q]e = 0}$$

Proof. If φ is a solution of $x' = f(x)\,\&\,Q$, then the premise implies that $\varphi \models (e)' = 0$ since all restrictions of solutions are again solutions. Consequently, Lemma 6 implies

$$0 = [\![(e)']\!]\varphi(\zeta) = \frac{\mathsf{d}\,[\![e]\!]\varphi(t)}{\mathsf{d}t}(\zeta)$$

so that e stays zero along φ by mean-value theorem, since it initially started out 0 (antecedent of conclusion) and had 0 change over time (above). Hold on, that use of Lemma 6 was, of course, predicated on having a solution φ of duration $r > 0$ (otherwise there are no time-derivatives to speak of). Yet, solutions of duration $r = 0$ directly imply $e = 0$ from the initial condition in the antecedent of the conclusion. □

The only nuisance with this proof rule is that DI_0 never proves any interesting properties on its own. For Example 1, it would lead to:

$$\dfrac{\dfrac{\vdash [v' = w, w' = -v]2vv' + 2ww' - 2rr' = 0}{v^2 + w^2 - r^2 = 0 \vdash [v' = w, w' = -v]v^2 + w^2 - r^2 = 0}\,\text{DI}_0}{\vdash v^2 + w^2 - r^2 = 0 \to [v' = w, w' = -v]v^2 + w^2 - r^2 = 0}\,{\to}\text{R}$$

Without knowing anything about v' and w' and r' in the postcondition, this proof has no chance of ever closing. What stands to reason is to use the right-hand sides of the differential equations for their left-hand sides, after all both sides of the equation are supposed to be equal. The question is how to justify that that's sound.

Lemma 6 shows that, along a differential equation, the value of the differential $(e)'$ of a term e coincides with the analytic time-derivative of the term e. Now, along a differential equation $x' = f(x)$, the differential symbols x' themselves actually have a simple interpretation: their values are determined directly by the differential equation. Putting these thoughts together leads to a way of replacing differential symbols with the corresponding right-hand sides of their respective differential equations. That is, replacing left-hand sides of differential equations with their right-hand sides.

Note 15. *The direction into which the value of a term e evolves as the system follows a differential equation $x' = f(x)$ depends on the differential $(e)'$ of the term e and on the differential equation $x' = f(x)$ that locally describes the evolution of x over time.*

Lemma 8 (Differential assignment)**.** *If $\varphi \models x' = f(x) \wedge Q$ for a flow $\varphi : [0, r] \to \mathcal{S}$ of any duration $r \geq 0$, then*

$$\varphi \models P \leftrightarrow [x' := f(x)]P$$

Proof. The proof is a direct consequence of the fact that the semantics of differential equations (Note 2) requires that $\varphi(\zeta) \models x' = f(x)$ holds at all times ζ all along φ. Consequently, the assignment $x' := f(x)$ that changes the value of x' around to be the value of $f(x)$ will have no effect, since x' already does have that value along the differential equation to begin with. Thus, P and $[x' := f(x)]P$ are equivalent along φ. □

By using this equivalence at any state along a differential equation $x' = f(x)$ this gives rise to a simple axiom characterizing the effect that a differential equation has on its differential symbols:

Corollary 9 (Differential effects). *The differential effect axiom is sound:*

$$\text{DE}\quad [x' = f(x)\,\&\,Q]P \leftrightarrow [x' = f(x)\,\&\,Q][x' := f(x)]P$$

The last ingredient is the differential assignment axiom $[':=]$ for $x' := e$ in direct analogy to the assignment axiom $[:=]$ for $[x := e]P$ just for a differential symbol x' instead of a variable x:

$$[':=]\quad [x' := e]p(x') \leftrightarrow p(e)$$

Let's continue the proof for Example 1:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\vdash [v' = w, w' = -v]2v(w) + 2w(-v) - 2rr' = 0}}{\vdash [v' = w, w' = -v][v':=w][w':=-v]2vv' + 2ww' - 2rr' = 0}\,[':=]}{\vdash [v' = w, w' = -v]2vv' + 2ww' - 2rr' = 0}\,\text{DE}}{v^2 + w^2 - r^2 = 0 \vdash [v' = w, w' = -v]v^2 + w^2 - r^2 = 0}\,\text{DI}_0}{\vdash v^2 + w^2 - r^2 = 0 \to [v' = w, w' = -v]v^2 + w^2 - r^2 = 0}\,{\to}\text{R}$$

Oops, that did not make all differential symbols disappear, because r' is still around, since r did not have a differential equation in (3) to begin with. Stepping back, what we mean by a differential equation like $v' = w, w' = -v$ that does not mention r' is that r is not supposed to change. If r is supposed to change during a continuous evolution, then there has to be a differential equation for r describing how r changes.

Note 18 (Explicit change). *Hybrid programs are* explicit change*: nothing changes unless an assignment or differential equation specifies how (compare the semantics from Lecture 3). In particular, if a differential equation (system) $x' = f(x)$ does not mention z', then the variable z does not change during $x' = f(x)$, so the differential equation systems $x' = f(x)$ and $x' = f(x), z' = 0$ are equivalent.*

We will assume $z' = 0$ without further notice for variables z that do not change during a differential equation.

Since (3) does not have a differential equation for r, Note 18 implies that its differential equation $v' = w, w' = -v$ is equivalent to $v' = w, w' = -v, r' = 0$, which, with DE,

would give rise to an extra $[r'{:=}0]$, which we will assume implicitly after showing its use explicitly just once.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\vdash 2v(w)+2w(-v)-0=0}\,\mathbb{R}}{\vdash [v'=w,w'=-v]2v(w)+2w(-v)-0=0}\,G}{\vdash [v'=w,w'=-v][v'{:=}w][w'{:=}-v][r'{:=}0]2vv'+2ww'-2rr'=0}\,[':=]}{\vdash [v'=w,w'=-v]2vv'+2ww'-2rr'=0}\,DE}{v^2+w^2-r^2=0\vdash [v'=w,w'=-v]v^2+w^2-r^2=0}\,DI_0}{\vdash v^2+w^2-r^2=0\to[v'=w,w'=-v]v^2+w^2-r^2=0}\,{\to}R$$

This is amazing, because we found out that the value of $v^2+w^2-r^2$ does not change over time (ultimately because its time-derivative is zero) along the differential equation $v'=w,w'=-v$. And we found that out without ever solving the differential equation, just by a few lines of simple symbolic proof steps.

9 Differential Invariant Terms

In order to be able to use the above reasoning as part of a sequent proof efficiently, let's package up the argument in a simple proof rule. As a first shot, we stay with equations of the form $e=0$, which gives us soundness for the following proof rule.

Lemma 10 (Differential invariant terms)**.** *The following special case of the differential invariants proof rule is sound, i.e. if its premise is valid then so is its conclusion:*

$$\mathrm{DI}_{=0}\ \dfrac{\vdash [x'{:=}f(x)](e)'=0}{e=0\vdash[x'=f(x)]e=0}$$

Proof. We could prove soundness of this proof rule by going back to the semantics and lemmas we proved about it. The easier soundness proof is to prove that it is a derived rule, meaning that it can be expanded into a sequence of other proof rule applications that we have already seen to be sound:

$$\dfrac{\dfrac{\dfrac{\vdash [x':=f(x)](e)'=0}{\vdash [x'=f(x)\,\&\,Q][x':=f(x)](e)'=0}\,G}{\vdash [x'=f(x)\,\&\,Q](e)'=0}\,DE}{e=0\vdash[x'=f(x)\,\&\,Q]e=0}\,DI_0$$

This proof shows $\mathrm{DI}_{=0}$ to be a derived rule because it starts with the premise of $\mathrm{DI}_{=0}$ as the only open goal and ends with the conclusion of $\mathrm{DI}_{=0}$, using only proof rules we already know are sound. □

This proof rule enables us to prove dℒ formula (3) easily in dℒ's sequent calculus:

$$\dfrac{\dfrac{\dfrac{\dfrac{*}{\vdash 2vw+2w(-v)-0=0}\,\mathbb{R}}{\vdash [v':=w][w':=-v]2vv'+2ww'-0=0}\,[':=]}{v^2+w^2-r^2=0 \vdash [v'=w,w'=-v]v^2+w^2-r^2=0}\,\mathrm{DI}_{=0}}{\vdash v^2+w^2-r^2=0 \to [v'=w,w'=-v]v^2+w^2-r^2=0}\,{\to}\mathrm{R}$$

See ≪Rotational differential invariant≫

Taking a step back, this is an exciting development, because, thanks to differential invariants, the property (3) of a differential equation with a nontrivial solution has a very simple proof that we can easily check. The proof did not need to solve the differential equation, which has infinitely many solutions with combinations of trigonometric functions.[6] The proof only required deriving the postcondition and substituting the differential equation in.

10 Proof by Generalization

So far, the argument captured in the differential invariant term proof rule $\mathrm{DI}_{=0}$ works for

$$v^2+w^2-r^2=0 \to [v'=w,w'=-v]v^2+w^2-r^2=0 \tag{3}$$

with an equation $v^2+w^2-r^2=0$ normalized to having 0 on the right-hand side but not for the original formula

$$v^2+w^2=r^2 \to [v'=w,w'=-v]v^2+w^2=r^2 \tag{1}$$

because its postcondition is not of the form $e=0$. Yet, the postcondition $v^2+w^2-r^2=0$ of (3) is trivially equivalent to the postcondition $v^2+w^2=r^2$ of (1), just by rewriting the polynomials on one side, which is a minor change. That is an indication, that differential invariants can perhaps do more than what proof rule $\mathrm{DI}_{=0}$ already knows about.

But before we pursue our discovery of what else differential invariants can do for us any further, let us first understand a very important proof principle.

Note 20 (Proof by generalization). *If you do not find a proof of a formula, it can sometimes be easier to prove a more general property from which the one you were looking for follows.*

[6] Granted, the solutions in this case are not quite so terrifying yet. They are all of the form

$$v(t)=a\cos t+b\sin t,\ w(t)=b\cos t-a\sin t$$

But the special functions sin and cos still fall outside the fragments of arithmetic that are known to be decidable.

This principle, which may at first appear paradoxical, turns out to be very helpful. In fact, we have made ample use of Note 20 when proving properties of loops by induction. The loop invariant that needs to be proved is usually more general than the particular postcondition one is interested in. The desirable postcondition follows from having proved a more general inductive invariant.

Recall the monotone *generalization* rule from Lecture 7 on Control Loops & Invariants:

$$\text{MR}\ \frac{\Gamma \vdash [\alpha]Q, \Delta \quad Q \vdash P}{\Gamma \vdash [\alpha]P, \Delta}$$

Instead of proving the desirable postcondition P of α (conclusion), proof rule MR makes it possible to prove the postcondition Q instead (left premise) and prove that Q is more general than the desired P (right premise). Generalization MR can help us prove the original dL formula (1) by first turning the postcondition into the form of the (provable) (3) and adapting the precondition using a corresponding cut with $v^2 + w^2 - r^2 = 0$:

$$\dfrac{\dfrac{\mathbb{R}\dfrac{*}{v^2+w^2=r^2 \vdash v^2+w^2-r^2=0} \quad \text{DI}_{=0}\dfrac{\dfrac{\mathbb{R}\dfrac{*}{2vw+2w(-v)-0=0}}{[v':=w][w':=-v]2vv'+2ww'-0}}{v^2+w^2-r^2=0 \vdash [v'=w,w'=-v]v^2+w^2-r^2=0}}{v^2+w^2=r^2 \vdash [v'=w,w'=-v]v^2+w^2-r^2=0}\ \text{cut,WL,WR} \quad \mathbb{R}\dfrac{*}{v^2+w^2-r^2=0 \vdash v^2+w^2=r^2}}{v^2+w^2=r^2 \vdash [v'=w,w'=-v]v^2+w^2=r^2}\ \text{MR}}{v^2+w^2=r^2 \to [v'=w,w'=-v]v^2+w^2=r^2}\ \to\text{R}$$

This is a possible way of proving the original (1), but also unnecessarily complicated. Differential invariants can prove (1) directly once we generalize proof rule $\text{DI}_{=0}$ appropriately. For other purposes, however, it is still important to have the principle of generalization Note 20 in our repertoire of proof techniques.

11 Example Proofs

Of course, differential invariants are just as helpful for proving properties of other differential equations.

Example 11 (Self-crossing). Another example is the following invariant property illustrated in Fig. 6:

$$x^2 + x^3 - y^2 - c = 0 \to [x' = -2y, y' = -2x - 3x^2]\, x^2 + x^3 - y^2 - c = 0$$

This dL formula proves easily using $\text{DI}_{=0}$:

$$\dfrac{\dfrac{\dfrac{\dfrac{*}{\vdash 2x(-2y)+3x^2(-2y)-2y(-2x-3x^2)=0}\ \mathbb{R}}{\vdash [x':=-2y][y':=-2x-3x^2]2xx'+3x^2x'-2yy'-0=0}\ [':=]}{x^2+x^3-y^2-c=0 \vdash [x'=-2y,y'=-2x-3x^2]x^2+x^3-y^2-c=0}\ \text{DI}_{=0}}{\vdash x^2+x^3-y^2-c=0 \to [x'=-2y,y'=-2x-3x^2]x^2+x^3-y^2-c=0}\ \to\text{R}$$

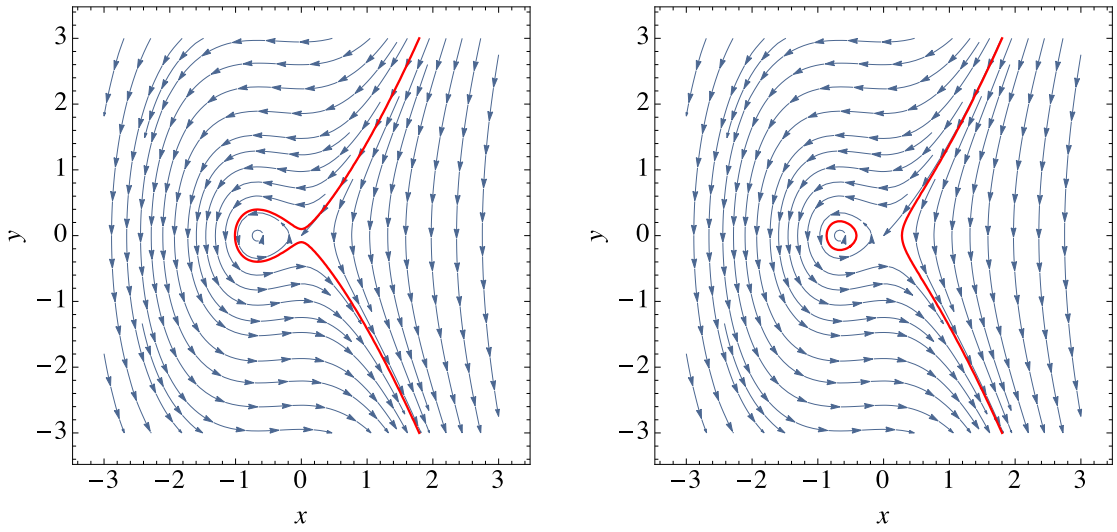


Figure 6: Two differential invariants of the indicated dynamics (illustrated in thick red) for different values of c

See ≪Self-crossing polynomial invariant≫

Example 12 (Motzkin). Another nice example is the Motzkin polynomial, which is an invariant of the following dynamics (see Fig. 7):

$$x^4y^2 + x^2y^4 - 3x^2y^2 + 1 = c \rightarrow$$
$$[x' = 2x^4y + 4x^2y^3 - 6x^2y, y' = -4x^3y^2 - 2xy^4 + 6xy^2]\, x^4y^2 + x^2y^4 - 3x^2y^2 + 1 = c$$

This dL formula proves easily using $\text{DI}_{=0}$, again after normalizing the equation to have right-hand side 0:

$$\begin{array}{ll}
 & * \\
\hline
\mathbb{R} & \vdash 0 = 0 \\
\hline
[':=] & \vdash [x':=2x^4y + 4x^2y^3 - 6x^2y][y':= -4x^3y^2 - 2xy^4 + 6xy^2](x^4y^2 + x^2y^4 - 3x^2y^2 + 1 - c)' = 0 \\
\hline
\text{DI}_{=0} & \ldots \vdash [x' = 2x^4y + 4x^2y^3 - 6x^2y, y' = -4x^3y^2 - 2xy^4 + 6xy^2]x^4y^2 + x^2y^4 - 3x^2y^2 + 1 - c = 0 \\
\hline
\rightarrow\text{R} & \vdash \cdots \rightarrow [x' = 2x^4y + 4x^2y^3 - 6x^2y, y' = -4x^3y^2 - 2xy^4 + 6xy^2]x^4y^2 + x^2y^4 - 3x^2y^2 + 1 - c = 0
\end{array}$$

This time, the proof step that comes without a label is simple, but requires some space:

$$(x^4y^2 + x^2y^4 - 3x^2y^2 + 1 - c)' = (4x^3y^2 + 2xy^4 - 6xy^2)x' + (2x^4y + 4x^2y^3 - 6x^2y)y'$$

After substituting in the differential equation, this gives

$$(4x^3y^2+2xy^4-6xy^2)(2x^4y+4x^2y^3-6x^2y)+(2x^4y+4x^2y^3-6x^2y)(-4x^3y^2-2xy^4+6xy^2)$$

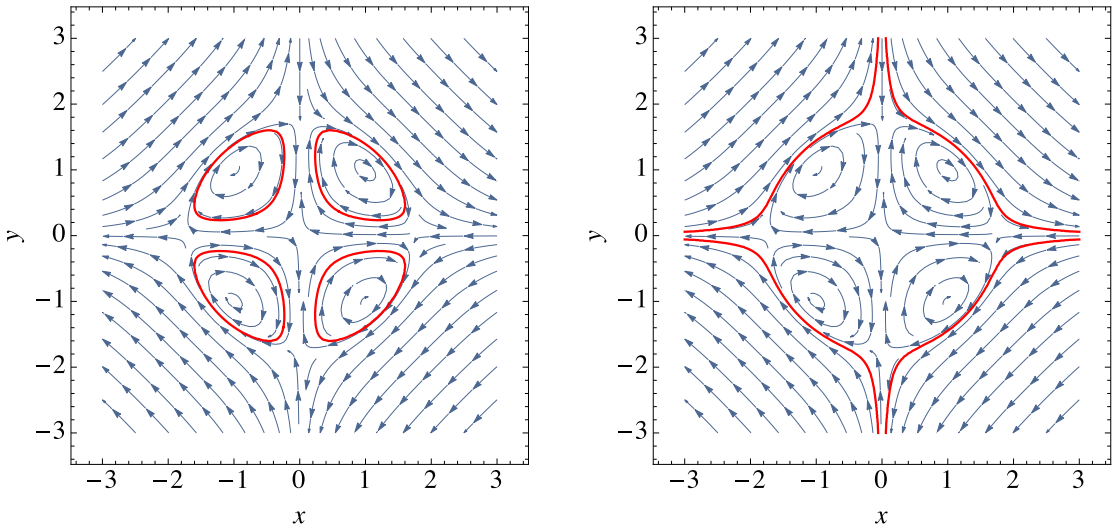


Figure 7: Two differential invariants of the indicated dynamics is the Motzkin polynomial (illustrated in thick red) for different values of c

which simplifies to 0 after expanding the polynomials, and, thus, leads to the equation $0 = 0$, which is easy to prove.

See ≪Motzkin polynomial invariant≫ Note that the arithmetic complexity reduces when hiding unnecessary contexts as shown in Lecture 6 on Truth & Proof.

Thanks to Andrew Sogokon for the nice Example 12.

12 Differential Invariant Terms and Invariant Functions

It is not a coincidence that these examples were provable by differential invariant proof rule DI$_{=0}$, because that proof rule can handle arbitrary invariant functions.

Expedition 4 (Lie characterization of invariant functions)**.** The proof rule DI$_{=0}$ works by deriving the postcondition and substituting the differential equation in:

$$\text{DI}_{=0}\ \frac{\vdash [x':=f(x)](e)' = 0}{e = 0 \vdash [x' = f(x)]e = 0}$$

There is something quite peculiar about DI$_{=0}$. Its premise is independent of the constant term in e. If, for any constant symbol c, the formula $e = 0$ is replaced by $e - c = 0$ in the conclusion, then the premise of DI$_{=0}$ stays the same, because $c' = 0$. Consequently, if DI$_{=0}$ proves

$$e = 0 \vdash [x' = f(x)]e = 0$$

then it also proves

$$e - c = 0 \vdash [x' = f(x)]e - c = 0 \tag{7}$$

for any constant c. This observation is the basis for a more general result, which simultaneously proves all formulas (7) for all c from the premise of DI$_{=0}$.

On open domains, equational differential invariants are even a necessary and sufficient characterization of such *invariant functions*, i.e. functions that are invariant along the dynamics of a system, because, whatever value c that function had in the initial state, the value will stay the same forever. The equational case of differential invariants are intimately related to the seminal work by Sophus Lie on what are now called Lie groups [Lie93, Lie97].

Theorem 13 (Lie [Pla12b]). *Let $x' = f(x)$ be a differential equation system and Q a* domain, *i.e., a first-order formula of real arithmetic characterizing a connected open set. The following proof rule is a sound global equivalence rule, i.e. the conclusion is valid if and only if the premise is:*

$$\text{DI}_c\ \frac{Q \vdash [x':=f(x)](e)' = 0}{\forall c\,(e = c \to [x' = f(x)\,\&\,Q]e = c)}$$

Despite the power that differential invariant terms offer, challenges lie ahead in proving properties. Theorem 13 gives an indication where challenges remain.

Example 14 (Generalizing differential invariants). The following dL formula is valid

$$x^2 + y^2 = 0 \to [x' = 4y^3, y' = -4x^3]\, x^2 + y^2 = 0 \tag{8}$$

but cannot be proved directly using DI$_{=0}$, because $x^2 + y^2$ is no invariant function of the dynamics. In combination with generalization (MR to change the postcondition to the equivalent $x^4 + y^4 = 0$) and a cut (to change the antecedent to the equivalent $x^4 + y^4 = 0$), however, there is a proof using differential invariants DI$_{=0}$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\vdash 4x^3(4y^3) + 4y^3(-4x^3) = 0}\,\mathbb{R}}{\vdash [x':=4y^3][y':=-4x^3]4x^3x' + 4y^3y' = 0}\,[':=]}{x^4 + y^4 = 0 \vdash [x' = 4y^3, y' = -4x^3]\, x^4 + y^4 = 0}\,\text{DI}_{=0}}{x^2 + y^2 = 0 \vdash [x' = 4y^3, y' = -4x^3]\, x^2 + y^2 = 0}\,\text{cut,MR}}{\vdash x^2 + y^2 = 0 \to [x' = 4y^3, y' = -4x^3]\, x^2 + y^2 = 0}\,{\to}\text{R}$$

The use of MR leads to another branch $x^4 + y^4 = 0 \vdash x^2 + y^2 = 0$ that is elided above. Similarly, cut leads to another branch $x^2 + y^2 = 0 \vdash x^4 + y^4 = 0$ that is also elided. Both prove easily by real arithmetic ($\mathbb{R}$).

See ≪Differential invariant after generalization≫

How could this happen? How could the original formula (8) be provable only after generalizing its postcondition to $x^4 + y^4 = 0$ and not before?

Note 22 (Strengthening induction hypotheses)**.** *An important phenomenon we already encountered in Lecture 7 on Loops & Invariants and other uses of induction is that, sometimes, the only way to prove a property is to strengthen the induction hypothesis. Differential invariants are no exception. It is worth noting, however, that the inductive structure in differential invariants includes their differential structure. And, indeed, the derivatives of* $x^4+y^4=0$ *are different and more conducive for an inductive proof than those of* $x^2+y^2=0$ *even if both have the same set of solutions.*

Theorem 13 explains why $x^2+y^2=0$ was doomed to fail as a differential invariant while $x^4+y^4=0$ succeeded. All formulas of the form $x^4+y^4=c$ for all c are invariants of the dynamics in (8), because the proof succeeded. But $x^2+y^2=c$ only is an invariant for the lucky choice $c=0$ and only equivalent to $x^4+y^4=0$ for this case.

There also is a way of deciding equational invariants of algebraic differential equations using a higher-order generalization of differential invariants called differential radical invariants [GP14].

13 Summary

This lecture showed one form of differential invariants: the form where the differential invariants are terms whose value always stays 0 along all solutions of a differential equation. The next lecture will use the tools developed in this lecture to investigate more general forms of differential invariants and more advanced proof principles for differential equations. They all share the important discovery in today's lecture: that properties of differential equations can be proved using the differential equation rather than its solution.

The most important technical insight of today's lecture was that even very complicated behavior that is defined by mathematical properties of the semantics can be captured by purely syntactical proof principles using differentials. The differential lemma proved that the values of differentials of terms coincide with the analytic derivatives of the values. The derivation lemma gave us the usual rules for computing derivatives as equations of differentials. The differential assignment lemma allowed us the intuitive operation of substituting differential equations into terms. Proving properties of differential equations using a mix of these simple proof principles is much more civilized and effective than working with solutions of differential equations. The proofs are also computationally easier, because the proof arguments are local and derivatives even decrease the polynomial degrees.

The principles begun in this lecture have more potential, though, and are not limited to proving only properties of the rather limited form $e=0$. Subsequent lectures will make use of the results obtained and build on the differential lemma, derivation lemma, and differential assignment lemma to develop more general proof principles for differential equations.

Exercises

Exercise 1. Note 3 explained that $(x' = f(x))^*$ is equivalent to $x' = f(x)$. Does the same hold for differential equations with evolution domain constraints? Are $(x' = f(x) \,\&\, Q)^*$ and $x' = f(x) \,\&\, Q$ equivalent or not? Justify or modify the statement and justify the variation.

Exercise 2. We argued that dℒ formulas (1) and (3) are equivalent and have then gone on to find a proof of (3). Continue this proof of (3) to a proof of (1) using the generalization rule MR and the cut rule.

Exercise 3. Prove the other cases of Lemma 5 where e is of the form $e - k$ or $e \cdot k$ or e/k.

Exercise 4. What happens in the proof of Lemma 10 if there is no solution φ? Show that this is not a counterexample to proof rule $\text{DI}_{=0}$, but that the rule is sound in that case.

Exercise 5. Carry out the polynomial computations needed to prove Example 12 using proof rule $\text{DI}_{=0}$.

Exercise 6. Prove the following dℒ formula using differential invariants:

$$xy = c \to [x' = -x, y' = y, z' = -z]xy = c$$

Exercise 7. Prove the following dℒ formula using differential invariants:

$$x^2 + 4xy - 2y^3 - y = 1 \to [x' = -1 + 4x - 6y^2, y' = -2x - 4y]x^2 + 4xy - 2y^3 - y = 1$$

Exercise 8. Prove the following dℒ formula using differential invariants:

$$x^2 + \frac{x^3}{3} = c \to [x' = y^2, y' = -2x]x^2 + \frac{x^3}{3} = c$$

Exercise 9 (Hénon-Heiles). Prove a differential invariant of a Hénon-Heiles system:

$$\frac{1}{2}(u^2 + v^2 + Ax^2 + By^2) + x^2y - \frac{1}{3}\varepsilon y^3 = 0 \to$$
$$[x' = u, y' = v, u' = -Ax - 2xy, v' = -By + \varepsilon y^2 - x^2]\frac{1}{2}(u^2+v^2+Ax^2+By^2)+x^2y-\frac{1}{3}\varepsilon y^3 = 0$$

References

[GP14] Khalil Ghorbal and André Platzer. Characterizing algebraic invariants by differential radical invariants. In Erika Ábrahám and Klaus Havelund, editors, *TACAS*, volume 8413 of *LNCS*, pages 279–294. Springer, 2014. doi:10.1007/978-3-642-54862-8_19.

[Kol72] Ellis Robert Kolchin. *Differential Algebra and Algebraic Groups*. Academic Press, New York, 1972.

[LIC12] *Proceedings of the 27th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2012, Dubrovnik, Croatia, June 25–28, 2012*. IEEE, 2012.

[Lie93] Sophus Lie. *Vorlesungen über continuierliche Gruppen mit geometrischen und anderen Anwendungen*. Teubner, Leipzig, 1893.

[Lie97] Sophus Lie. Über Integralinvarianten und ihre Verwertung für die Theorie der Differentialgleichungen. *Leipz. Berichte*, 49:369–410, 1897.

[PC08] André Platzer and Edmund M. Clarke. Computing differential invariants of hybrid systems as fixedpoints. In Aarti Gupta and Sharad Malik, editors, *CAV*, volume 5123 of *LNCS*, pages 176–189. Springer, 2008. doi:10.1007/978-3-540-70545-1_17.

[PC09] André Platzer and Edmund M. Clarke. Computing differential invariants of hybrid systems as fixedpoints. *Form. Methods Syst. Des.*, 35(1):98–120, 2009. Special issue for selected papers from CAV'08. doi:10.1007/s10703-009-0079-8.

[Pla08a] André Platzer. Differential dynamic logic for hybrid systems. *J. Autom. Reas.*, 41(2):143–189, 2008. doi:10.1007/s10817-008-9103-8.

[Pla08b] André Platzer. *Differential Dynamic Logics: Automated Theorem Proving for Hybrid Systems*. PhD thesis, Department of Computing Science, University of Oldenburg, Dec 2008. Appeared with Springer.

[Pla10a] André Platzer. Differential-algebraic dynamic logic for differential-algebraic programs. *J. Log. Comput.*, 20(1):309–352, 2010. doi:10.1093/logcom/exn070.

[Pla10b] André Platzer. *Logical Analysis of Hybrid Systems: Proving Theorems for Complex Dynamics*. Springer, Heidelberg, 2010. doi:10.1007/978-3-642-14509-4.

[Pla12a] André Platzer. The complete proof theory of hybrid systems. In LICS [LIC12], pages 541–550. doi:10.1109/LICS.2012.64.

[Pla12b] André Platzer. A differential operator approach to equational differential invariants. In Lennart Beringer and Amy Felty, editors, *ITP*, volume 7406 of *LNCS*, pages 28–48. Springer, 2012. doi:10.1007/978-3-642-32347-8_3.

[Pla12c] André Platzer. Logics of dynamical systems. In LICS [LIC12], pages 13–24. doi:10.1109/LICS.2012.13.

[Pla12d] André Platzer. The structure of differential invariants and differential cut elimination. *Log. Meth. Comput. Sci.*, 8(4):1–38, 2012. doi:10.2168/LMCS-8(4:16)2012.

[Pla15] André Platzer. A uniform substitution calculus for differential dynamic logic. In Amy Felty and Aart Middeldorp, editors, *CADE*, volume 9195 of *LNCS*, pages 467–481. Springer, 2015. doi:10.1007/978-3-319-21401-6_32.

[SS71] Dana Scott and Christopher Strachey. Toward a mathematical semantics for computer languages? Technical Report PRG-6, Oxford Programming Research Group, 1971.

[Zei03] Eberhard Zeidler, editor. *Teubner-Taschenbuch der Mathematik*. Teubner, 2003. doi:10.1007/978-3-322-96781-7.