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 [′]r rule from Lecture 6 on Truth & Proof for a differential equation $x' = \theta$, we must be able to find a symbolic solution to the symbolic initial value problem (i.e. a function $y(t)$ such that $y'(t) = \theta$ 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. A solution of

$$d' = e, e' = -d$$

for example is $d(t) = \sin t, e(t) = \cos t$, 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].

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 lectures expand the same core principles developed in this lecture to the study of general properties of differential equations. This lecture continues the *axiomatiza-*

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

tion of differential dynamic logic d$\mathcal{L}$ [Pla12c, Pla12a] pursued since Lecture 5 on Dynamical Systems & Dynamic Axioms and lifts d$\mathcal{L}$'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.

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.

The semantics of a differential equation was described in Lecture 2 on Differential Equations & Domains as:

$$\begin{aligned}\rho(x' = \theta \,\&\, H) = \{(\varphi(0), \varphi(r)) \ :\ & \varphi(t) \models x' = \theta \text{ and } \varphi(t) \models H \text{ for all } 0 \le t \le r \\ & \text{for a solution } \varphi : [0, r] \to \mathcal{S} \text{ of any duration } r\}\end{aligned}$$

The solution φ describes the global behavior of the system, which is specified locally by the right-hand side θ 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], 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

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

How does the differential equation $x' = \theta$ compare to the same differential equation in a loop $(x' = \theta)^*$ instead? Unlike the differential equation $x' = \theta$, the repeated differential equation $(x' = \theta)^*$ can run the differential equation $x' = \theta$ repeatedly. Albeit, on second thought, does that get the repetitive differential equation $(x' = \theta)^*$ to any more states than where the differential equation $x' = \theta$ could evolve to?

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

Note 2 (Looping differential equations)**.** *$(x' = \theta)^*$ is equivalent to $x' = \theta$, written $(x' = \theta)^* \equiv (x' = \theta)$, i.e. both have the same transition semantics:*

$$\rho((x' = \theta)^*) = \rho(x' = \theta)$$

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

In light of Note 2, 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

[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).

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 $x' = \theta$ 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 $x' = \theta$ 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 $x' = \theta$ 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 [^{*n}] from Lecture 5 on Dynamical Systems & Dynamic Axioms.

Note 3 (Correspondence map between loops and differential equations).

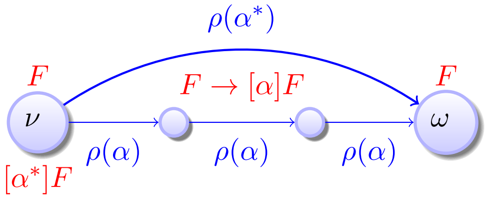
loop α^*	***differential equation*** $x' = \theta$
could repeat 0 times	*could evolve for duration 0*
repeat any number $n \in \mathbb{N}$ *of times*	*evolve for any duration* $0 \leq r \in \mathbb{R}$
effect depends on previous loop iteration	*effect depends on the past solution*
local generator α	*local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

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 $\forall t{\geq}0$ 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:

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



$$(ind')\ \frac{\Gamma \vdash F, \Delta \quad F \vdash [\alpha]F \quad F \vdash \psi}{\Gamma \vdash [\alpha^*]\psi, \Delta}$$

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 *ind'* or the only premise of the core 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 *ind'* 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 $x' = \theta$ 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 d$\mathcal{L}$ formula $[x' = \theta]F$ expressing that all runs of $x' = \theta$ 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 $\neg F$.

One way of proving that $[x' = \theta]F$ is true in a state ν would be to compute a solution from that state ν, 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 ν, 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

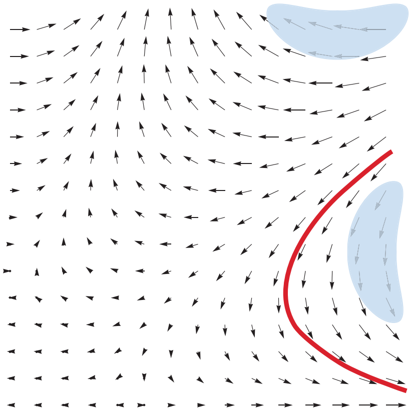


Figure 1: Vector field and one solution of a differential equation that does not enter the blue unsafe regions

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 checking whether the system could ever lead to an unsafe state in $\neg F$ when following the differential equation $x' = \theta$. 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 $x' = \theta$ be made rigorous? We develop this step by step.

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

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

The conjectured d$\mathcal{L}$ formula (1) is valid, because, indeed, if the vector (d, e) is initially at distance r from the origin (0,0), then it will always be when rotating around the ori-

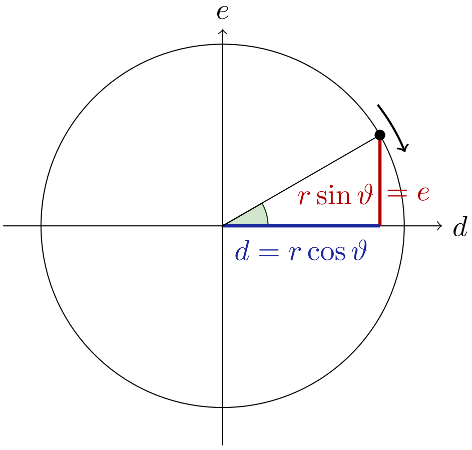


Figure 2: One scenario for the rotational dynamics and relationship of vector (d, e) to radius r and angle ϑ

gin, which is what the dynamics does. That is, the point (d, e) 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 arithmetic will be unnecessarily difficult and, after all, the argument for why the simple d$\mathcal{L}$ 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 $x' = \theta$ follows from state ν is exactly described by the time-derivative, which is exactly the value $[\![\theta]\!]_\nu$ of term θ in state ν. Recall that the term θ can mention x and other variables so its value $[\![\theta]\!]_\nu$ depends on the state ν.

Note 4 (Differential invariants are "formulas that remain true in the direction of the dynamics")**.** *Proving* d$\mathcal{L}$ *formula* $[x' = \theta]F$ *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* ν *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.*

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

Note 5. *Directions of evolutions are described by derivatives, after all the differential equation $x' = \theta$ describes that the time-derivative of x is θ.*

Let's derive the term η of interest and see what that tells us about how η evolves over time. How can we derive η? The term η 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 Σ denote the set of all variables. *Terms* θ are defined by the grammar (where θ, η are terms, x a variable, and r a rational number constant):

$$\theta, \eta ::= x \mid r \mid \theta + \eta \mid \theta - \eta \mid \theta \cdot \eta \mid \theta / \eta$$

It is, of course, important to take care that division θ/η only makes sense in a context where the divisor η 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 η is a sum $a + b$, its derivative is the derivative of a plus the derivative of b. If η is a product $a \cdot b$, its derivative is the derivative of a times b plus a times the derivative of b by Leibniz' rule. The derivative of a rational number constant $r \in \mathbb{Q}$ 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 $r \in \mathbb{R}$ is also zero, but only rational number constants are allowed to occur in the formulas of first-order logic of real arithmetic, more precisely, of real-closed fields.

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

Definition 1 (Derivation)**.** The operator $(\cdot)'$ that is defined as follows on terms is called *syntactic (total) derivation*:

$$
\begin{aligned}
(r)' &= 0 && \text{for numbers } r \in \mathbb{Q} && \text{(2a)}\\
(x)' &= x' && \text{for variable } x \in \Sigma && \text{(2b)}\\
(a+b)' &= (a)' + (b)' && && \text{(2c)}\\
(a-b)' &= (a)' - (b)' && && \text{(2d)}\\
(a \cdot b)' &= (a)' \cdot b + a \cdot (b)' && && \text{(2e)}\\
(a/b)' &= ((a)' \cdot b - a \cdot (b)')/b^2 && && \text{(2f)}
\end{aligned}
$$

Note that the intuition (and precise semantics) of derivatives of terms will ultimately be connected with more complicated aspects of how values change over time, the computation of derivatives of terms is a straightforward recursive definition on terms.

Expedition 1 (Differential algebra)**.** Even though the following names are not needed for his course, let's take a brief expedition to align Def. 1 with the algebraic structures from differential algebra [Kol72] in order to illustrate the systematic principles behind Def. 1. Case (2a) 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 (2c) and the *Leibniz* or *product rule* (2e) 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 (2c). 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 (2e). Equation (2d) is a derived rule for subtraction according to the identity $a - b = a + (-1) \cdot b$ and again expresses a homomorphic property, now with respect to subtraction rather than addition.

The equation (2b) uniquely defines the operator $(\cdot)'$ on the *differential polynomial algebra* spanned by the *differential indeterminates* $x \in \Sigma$, 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 $x \in \Sigma$. Equation (2f) 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 (2f) is defined whenever the original division a/b can be carried out, which, as we assumed for well-definedness, is guarded by $b \neq 0$.

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

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

Now that we have a precise 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) d$\mathcal{L}$ formula instead, which only has a single nontrivial term to worry about:

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

Derivation of the only relevant term $d^2 + e^2 - r^2$ in the postcondition of (3) gives

$$(d^2 + e^2 - r^2)' = 2dd' + 2ee' - 2rr' \tag{4}$$

Def. 1 makes it possible to derive any polynomial or rational term. Deriving them with the total derivative operator $(\cdot)'$ does *not* result in a term over the signature of the original variables in Σ, but, instead, a *differential term*, i.e. a term over the extended signature $\Sigma \cup \Sigma'$, where $\Sigma' \overset{\text{def}}{=} \{x' : x \in \Sigma\}$ is the set of all differential symbols x' for variables $x \in \Sigma$. The total derivative $(\eta)'$ of a polynomial term η is not a polynomial term, but may mention differential symbols such as x' in addition to the symbols that where in η 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 is the meaning of a differential symbol x'?

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

6 The Meaning of Prime

The meaning $[\![x]\!]_\nu$ of a variable symbol x is defined by the state ν as $[\![x]\!]_\nu = \nu(x)$. It is crucial to notice that the meaning of a differential symbol x' cannot be defined in a state ν, because derivatives do not even exist in isolated points. It is meaningless to ask for the change of the value of x over time in a single isolated state ν.

Along a (differentiable) continuous evolution $\varphi : [0, r] \to \mathcal{S}$ of a system, however, we can make sense of what x' means. 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 of the value $[\![x]\!]_{\varphi(\zeta)}$ of x at ζ [Pla10a]. That is, at any point in time ζ along φ, it makes sense to give x' the meaning of the rate of change of the value of x over time along φ.

Definition 2 (Semantics of differential symbols)**.** The value of x' at time $\zeta \in [0, r]$ of a differentiable function $\varphi : [0, r] \to \mathcal{S}$ of some duration $r \in \mathbb{R}$ is defined as the analytic time-derivative at ζ:

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

Intuitively, $[\![x']\!]_{\varphi(\zeta)}$ is 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. 3.

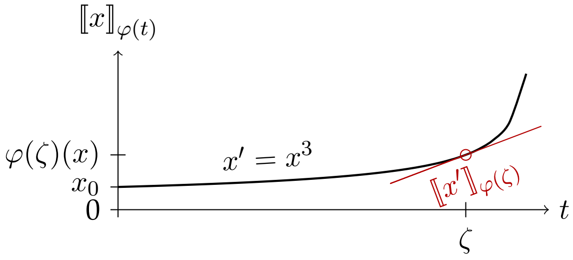


Figure 3: Semantics of differential symbols

Yet, what exactly do we know about the right-hand side in Def. 2, i.e. the time-derivative of the value of x along φ at time ζ? For differentiable φ, that analytic time-derivative is always defined, but that does not mean it would be computable for any arbitrary φ. If, however, the continuous evolution φ follows a differential equation $x' = \theta$, i.e. φ solves $x' = \theta$, then $[\![x']\!]_{\varphi(\zeta)}$ can be described easily in terms of that differential equation, because at any time $\zeta \in [0, r]$ the time-derivative of the value of x is $[\![\theta]\!]_{\varphi(\zeta)}$ simply by definition of what it means for φ to be a solution of $x' = \theta$ (cf. Lecture 2 on Differential Equations & Domains).

Now Def. 1 defines how to derive a term η syntactically to form $(\eta)'$ and Def. 2 defines how to interpret the differential symbols x' that occur in the total derivative $(\eta)'$. When interpreting all differential symbols as defined in Def. 2 for an evolution φ that follows the differential equation $x' = \theta$, this defines a value for the derivative $(\eta)'$ of any term η along that function φ. What does this value mean? How does it relate to how the value of η changes over time?

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

When interpreting differential symbols by derivatives along a function φ, the value of $(\eta)'$ at any time ζ coincides with the analytic time-derivative of the value of η at ζ. The key insight behind this is the derivation lemma, a differential analogue of the substitution lemma.

Expedition 2 (Substitutions in logic)**.** The substitution lemma shows that syntactic substitution has the same effect as changing the value of the variables it replaces.

Lemma 3 (Substitution Lemma [Pla10b, Lemma 2.2])*. Let the substitution of θ for x in ϕ to form ϕ_x^θ be admissible; then*

$$\text{for each } \nu : \ [\![\phi_x^\theta]\!]_\nu = [\![\phi]\!]_{\nu_x^e} \text{ where } e = [\![\theta]\!]_\nu$$

That is, semantically evaluating ϕ after modifying the interpretation of the symbol x replaced by its new value $[\![\theta]\!]_\nu$ is the same as semantically evaluating the result of syntactically substituting x by θ in ϕ in the original state.

The substitution lemma is a very powerful tool, because, among other things, it can be used to replace equals for equals without changing the valuation (substitution property). If we know that x and θ have the same value in ν, then we can substitute θ for x in a formula ϕ (if admissible) without changing the truth-value of ϕ, that is:

Lemma 4 (Substitution property [Pla10b, Lemma 2.3])*. If $\nu \models x = \theta$, then $\nu \models \phi \leftrightarrow \phi_x^\theta$ for any formula ϕ for which the substitution replacing x with θ is admissible.*

The substitution property implies that equals can be substituted for equals, i.e. left-hand sides of equations can be substituted by right-hand sides of equations within formulas in which the equations hold. Lemma 4 is the reason why the equality substitution proof rules =l,=r from Lecture 6 on Truth & Proof are sound.

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 5 (Derivation lemma). *Let $\varphi : [0, r] \to \mathcal{S}$ be a differentiable function of duration $r > 0$. Then for all terms η that are defined all along φ and all times $\zeta \in [0, r]$:*

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

where differential symbols are interpreted according to Def. 2. In particular, $[\![\eta]\!]_{\varphi(\zeta)}$ is continuously differentiable.

Proof. The proof is an inductive consequence of the correspondence of the semantics of differential symbols and analytic derivatives along a flow (Def. 2). It uses the assumption that φ remains within the domain of definition of η and is continuously differentiable in all variables of η. In particular, all denominators are nonzero during φ.

- If η is a variable x, the conjecture holds immediately by Def. 2:

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

  The derivative exists, because φ is assumed to be differentiable.

- If η is of the form $a + b$, the desired result can be obtained by using the properties of analytic derivatives, synctactic derivations (Def. 1), and valuation of terms (Lecture 2):

  $$\begin{aligned}
  & \frac{\mathsf{d}}{\mathsf{d}t}([\![a+b]\!]_{\varphi(t)})(\zeta) & \\
  &= \frac{\mathsf{d}}{\mathsf{d}t}([\![a]\!]_{\varphi(t)} + [\![b]\!]_{\varphi(t)})(\zeta) && [\![\cdot]\!]_\nu \text{ homomorphic for } + \\
  &= \frac{\mathsf{d}}{\mathsf{d}t}([\![a]\!]_{\varphi(t)})(\zeta) + \frac{\mathsf{d}}{\mathsf{d}t}([\![b]\!]_{\varphi(t)})(\zeta) && \frac{\mathsf{d}}{\mathsf{d}t} \text{ is a (linear) derivation} \\
  &= [\![(a)']\!]_{\varphi(\zeta)} + [\![(b)']\!]_{\varphi(\zeta)} && \text{by induction hypothesis} \\
  &= [\![(a)' + (b)']\!]_{\varphi(\zeta)} && [\![\cdot]\!]_\nu \text{ homomorphic for } + \\
  &= [\![(a+b)']\!]_{\varphi(\zeta)} && (\cdot)' \text{ is a syntactic derivation}
  \end{aligned}$$

- The case where η is of the form $a - b$ is similar, using subtractivity (2d) of Def. 1.

- The case where η is of the form $a \cdot b$ is similar, using Leibniz product rule (2e) of

Def. 1:

$$
\begin{aligned}
&\frac{\mathsf{d}}{\mathsf{d}t}([\![a\cdot b]\!]_{\varphi(t)})(\zeta) \\
&= \frac{\mathsf{d}}{\mathsf{d}t}([\![a]\!]_{\varphi(t)}\cdot [\![b]\!]_{\varphi(t)})(\zeta) && [\![\cdot]\!]_\nu \text{ homomorphic for } \cdot \\
&= \frac{\mathsf{d}}{\mathsf{d}t}([\![a]\!]_{\varphi(t)})(\zeta)\cdot [\![b]\!]_{\varphi(\zeta)} + [\![a]\!]_{\varphi(\zeta)}\cdot \frac{\mathsf{d}}{\mathsf{d}t}([\![b]\!]_{\varphi(t)})(\zeta) && \frac{\mathsf{d}}{\mathsf{d}t} \text{ is a (Leibniz) derivation} \\
&= [\![(a)']\!]_{\varphi(\zeta)}\cdot [\![b]\!]_{\varphi(\zeta)} + [\![a]\!]_{\varphi(\zeta)}\cdot [\![(b)']\!]_{\varphi(\zeta)} && \text{by induction hypothesis} \\
&= [\![(a)'\cdot b]\!]_{\varphi(\zeta)} + [\![a\cdot (b)']\!]_{\varphi(\zeta)} && [\![\cdot]\!]_\nu \text{ homomorphic for } \cdot \\
&= [\![(a)'\cdot b + a\cdot (b)']\!]_{\varphi(\zeta)} && [\![\cdot]\!]_\nu \text{ homomorphic for } + \\
&= [\![(a\cdot b)']\!]_{\varphi(\zeta)} && (\cdot)' \text{ is a syntactic derivation}
\end{aligned}
$$

- The case where η is of the form a/b uses (2f) of Def. 1 and further depends on the assumption that $b \neq 0$ along φ. This holds as the value of η is assumed to be defined all along state flow φ.
- The values of numbers $r \in \mathbb{Q}$ do not change during a state flow (in fact, they are not affected by the state at all); hence their derivative is $(r)' = 0$. □

Note 11 (The derivation lemma clou)**.** *Lemma 5 shows that analytic derivatives coincide with syntactic derivations. The clou with Lemma 5 is that it equates precise but sophisticated analytic derivatives with tame and computable syntactic derivations. The analytic derivatives on the left-hand side of Lemma 5 are mathematically precise and pinpoint exactly what we are interested in: the rate of change of the value of η along φ. But they are unwieldy for computers, because analytic derivatives are ultimately defined in terms of limit processes. The syntactic derivations on the right-hand side of Lemma 5 are computationally tame, because they can be computed easily by the simple recursive construction in Def. 1. But the syntactic derivations need to be aligned with the intended analytic derivatives, which is what Lemma 5 is good for. And, of course, deriving polynomials and rational functions is much easier syntactically than by unpacking the meaning of analytic derivatives in terms of limit processes.*

Lemma 5 shows that the value of the total derivative of a term coincides with the analytic derivative of the term, provided that differential symbols are interpreted according to Def. 2. Now, along a differential equation $x' = \theta$, the differential symbols themselves actually have a simple interpretation, the interpretation determined by the differential equation. Putting these thoughts together leads to 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 12. *The direction into which the value of a term η evolves as the system follows a differential equation $x' = \theta$ depends on the derivation η' of the term η and on the differential equation $x' = \theta$ that locally describes the evolution of x over time.*

The substitution property can replace equals for equals (Lemma 4). It can be lifted to differential equations such that differential equations can be used for equivalent substitutions along continuous flows respecting these differential equations. The following lemma can be used to substitute right-hand sides of differential equations for the left-hand side derivatives in flows along which these differential equations hold.

Lemma 6 (Differential substitution property for terms)**.** *If $\varphi : [0, r] \to \mathcal{S}$ solves the differential equation $x' = \theta$, i.e. $\varphi \models x' = \theta$, then $\varphi \models (\eta)' = (\eta)'^{\theta}_{x'}$ for all terms η, i.e.:*

$$[\![(\eta)']\!]_{\varphi(\zeta)} = [\![(\eta)'^{\theta}_{x'}]\!]_{\varphi(\zeta)} \quad \textit{for all } \zeta \in [0, r]$$

Proof. The proof is a simple inductive consequence of Lemma 5 using that $[\![x']\!]_{\varphi(\zeta)} = [\![\theta]\!]_{\varphi(\zeta)}$ at each time ζ in the domain of φ. □

The operation mapping term η to $(\eta)'^{\theta}_{x'}$ is called *Lie-derivative* of η with respect to $x' = \theta$.

Differential substitution of the differential equation $d' = e, e' = -d$ from (3) into (4) results in

$$(d^2 + e^2 - r^2)'^{e}_{d'}{}^{-d}_{e'} = (2dd' + 2ee' - 2rr')^{e}_{d'}{}^{-d}_{e'} = 2de + 2e(-d) + 2rr'$$

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

Note 14 (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) $x' = \theta$ does not mention z', then the variable z does not change during $x' = \theta$, so the differential equation systems $x' = \theta$ and $x' = \theta, z' = 0$ are equivalent.*

We will often 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 14 implies that its differential equation $d' = e, e' = -d$ is equivalent to $d' = e, e' = -d, r' = 0$. Hence, when adding zero derivatives for all unchanged variables, differential substitution of the differential equation $d' = e, e' = -d$ along with the explicit-change assumption $r' = 0$ into (4) gives

$$(d^2 + e^2 - r^2)'^{e}_{d'}{}^{-d}_{e'}{}^{0}_{r'} = (2dd' + 2ee' - 2rr')^{e}_{d'}{}^{-d}_{e'}{}^{0}_{r'} = 2de + 2e(-d) \tag{5}$$

This is good news, because the last part of (5) is a standard term of first-order logic of real arithmetic, because it no longer has any differential symbols. So we can make sense of $2de+2e(-d)$ and, by Lemma 6, its value along a solution of $d' = e, e' = -d$ is the same as that of the derivative $(d^2 + e^2 - r^2)'$, which, by Lemma 5 is the same as the value of the time-derivative of the original term $d^2 + e^2 - r^2$ along such a solution. Simple arithmetic shows that the term $2de + 2e(-d)$ in (5) is 0. Consequently, by Lemma 5 and Lemma 6, the time-derivative of the term $d^2 + e^2 - r^2$ in the postcondition of (3) is 0 along any solution φ of its differential equation:

$$\begin{aligned}\frac{\mathsf{d}[\![d^2+e^2-r^2]\!]_{\varphi(t)}}{\mathsf{d}t}(\zeta) &\overset{\text{Lem5}}{=} [\![(d^2+e^2-r^2)']\!]_{\varphi(\zeta)}\\ &\overset{\text{Lem6}}{=} [\![(d^2+e^2-r^2)'^{\,e\ -d\ 0}_{\,d'\ e'\ r'}]\!]_{\varphi(\zeta)}\\ &\overset{(5)}{=} [\![2de+2e(-d)]\!]_{\varphi(\zeta)} = 0\end{aligned}$$

for all times ζ. That means that the value of $d^2 + e^2 - r^2$ never changes during the rotation, and, hence (3) is valid, because $d^2 + e^2 - r^2$ stays 0 if it was 0 in the beginning, which is what (3) assumes.

This is amazing, because we found out that the value of $d^2 + e^2 - r^2$ does not change over time (its time-derivative is zero) along the differential equation $d' = e, e' = -d$. And we found that out without ever solving the differential equation, just by a few lines of simple symbolic computations. We only need to make sure to systematize this reasoning and make it accessible in the d$\mathcal{L}$ proof calculus by reflecting it in a proof rule, preferably one that is much more general than the special argument we needed to convince ourselves that (3) was valid.

7 Differential Invariant Terms

In order to be able to use the above reasoning as part of a sequent proof, we need to capture such arguments in a proof rule, preferably one that is more general than this particular argument. The argument is not specific to the term $d^2 + e^2 - r^2$ but works for any other term η and for any differential equation $x' = \theta$.

What we set out to find is a general proof rule for concluding properties of differential equations from properties of derivatives. As a first shot, we stay with equations of the form $\eta = 0$, which gives us soundness for the following proof rule.

Lemma 7 (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:*

$$(DI_{=0})\ \frac{\vdash \eta'^{\theta}_{x'} = 0}{\eta = 0 \vdash [x' = \theta]\eta = 0}$$

Proof. Assume the premise $\eta'^{\theta}_{x'} = 0$ to be valid, i.e. true in all states. In order to prove that the conclusion $\eta = 0 \vdash [x' = \theta]\eta = 0$ is valid, consider any state ν. Assume that

$\nu \models \eta = 0$, as there is otherwise nothing to show (sequent is trivially *true* since antecedent evaluates to *false*). If $\zeta \in [0, r]$ is any time during any solution $\varphi : [0, r] \to \mathcal{S}$, of any duration $r \in \mathbb{R}$, of the differential equation $x' = \theta$ beginning in initial state $\varphi(0) = \nu$, then

$$\frac{\mathsf{d}[\![\eta]\!]_{\varphi(t)}}{\mathsf{d}t}(\zeta) \overset{\text{Lem5}}{=} [\![(\eta)']\!]_{\varphi(\zeta)} \overset{\text{Lem6}}{=} [\![(\eta)'^{\theta}_{x'}]\!]_{\varphi(\zeta)} \overset{\text{premise}}{=} 0 \tag{6}$$

By antecedent, $\nu \models \eta = 0$, i.e. $[\![\eta]\!]_\nu = 0$, in the initial state $\nu = \varphi(0)$.

But, hold on a moment, the use of Lemma 5 in (6) was only correct if $r > 0$, otherwise derivatives make no sense. Fortunately, if the duration of φ is $r = 0$, we have $\varphi(0) \models \eta = 0$ immediately, because $\nu \models \eta = 0$. For duration $r > 0$, we show that $\eta = 0$ holds all along the flow φ, i.e., $\varphi(\zeta) \models \eta = 0$ for all $\zeta \in [0, r]$.

Suppose there was a $\zeta \in [0, r]$ with $\varphi(\zeta) \models \eta \neq 0$, which will lead to a contradiction. The function $h : [0, r] \to \mathbb{R}$ defined as $h(t) = [\![\eta]\!]_{\varphi(t)}$ satisfies the relation $h(0) = 0 \neq h(\zeta)$, because $h(0) = [\![\eta]\!]_{\varphi(0)} = [\![\eta]\!]_\nu$ and $\nu \models \eta = 0$ by antecedent of the conclusion. By Lemma 5, h is continuous on $[0, r]$ and differentiable at every $\xi \in (0, r)$. By the mean-value theorem, there is a $\xi \in (0, \zeta)$ such that $\frac{\mathsf{d}h(t)}{\mathsf{d}t}(\xi) \cdot (\zeta - 0) = h(\zeta) - h(0) \neq 0$. In particular, we can conclude that $\frac{\mathsf{d}h(t)}{\mathsf{d}t}(\xi) \neq 0$. Now Lemma 5 implies that $\frac{\mathsf{d}h(t)}{\mathsf{d}t}(\xi) = [\![(\eta)']\!]_{\varphi(\xi)} \neq 0$. This, however, is a contradiction, because the premise implies that the formula $(\eta)' = 0$ is true in all states along φ, including $\varphi(\xi) \models (\eta)' = 0$, which contradicts $[\![(\eta)']\!] \neq 0$. □

This proof rule enables us to prove $\mathsf{d}\mathcal{L}$ formula (3) easily in $\mathsf{d}\mathcal{L}$'s sequent calculus:

$$\begin{array}{rl}
 & \qquad\qquad\qquad \ast \\ \hline
\mathbb{R} & \qquad\qquad\qquad \vdash 2de + 2e(-d) - 0 = 0 \\ \hline
 & \qquad\qquad\qquad \vdash (2dd' + 2ee' - 2rr')^{e\ \ -d\ \ 0}_{d'\ \ e'\ \ r'} = 0 \\ \hline
\text{DI}_{=0} & d^2 + e^2 - r^2 = 0 \vdash [d' = e, e' = -d]d^2 + e^2 - r^2 = 0 \\ \hline
\to\text{r} & \qquad\qquad\qquad \vdash d^2 + e^2 - r^2 = 0 \to [d' = e, e' = -d]d^2 + e^2 - r^2 = 0
\end{array}$$

The proof step that comes without a label just performs the substitution, which is a convention we adopt from now on to make proofs more readable.

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.[5] The proof only required deriving the postcondition and substituting the differential equation in.

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

$$d(t) = a \cos t + b \sin t, \; e(t) = b \cos t - a \sin t$$

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

8 Proof by Generalization

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

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

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

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

because its postcondition is not of the form $\eta = 0$. Yet, the postcondition $d^2 + e^2 - r^2 = 0$ of (3) is trivially equivalent to the postcondition $d^2 + e^2 = r^2$ 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 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 16 (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.*

This principle, which may at first appear paradoxical, turns out to be very helpful. In fact, we have made ample use of Note 16 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.

In its purest form, generalization is captured in the *generalization* rule from Lecture 7 on Control Loops & Invariants. One of the forms of the generalization rule is:

$$([]gen')\ \frac{\Gamma \vdash [\alpha]\phi, \Delta \quad \phi \vdash \psi}{\Gamma \vdash [\alpha]\psi, \Delta}$$

Instead of proving the desirable postcondition ψ of α (conclusion), proof rule $[]gen'$ makes it possible to prove the postcondition ϕ instead (left premise) and prove that ϕ is more general than the desired ψ (right premise). Generalization $[]gen'$ can help us prove the original d$\mathcal{L}$ formula (1) by first turning the postcondition into the form of the (provable) (3) and adapting the precondition using a corresponding *cut* with $d^2 + e^2 - r^2 = 0$:

$$\to\!\mathrm{r}\,\dfrac{[]gen'\,\dfrac{cut,\mathrm{Wl},\mathrm{Wr}\,\dfrac{\mathbb{R}\,\dfrac{*}{d^2+e^2=r^2 \vdash d^2+e^2-r^2=0} \qquad \mathrm{DI}_{=0}\,\dfrac{\dfrac{\mathbb{R}\,\dfrac{*}{\vdash 2de+2e(-d)-0=0}}{\vdash (2dd'+2ee'-2rr'=0)_{d'}^{e}\,{}_{e'}^{-d}\,{}_{r'}^{-0}}}{d^2+e^2-r^2=0 \vdash [d'=e,e'=-d]d^2+e^2-r^2=0}}{d^2+e^2=r^2 \vdash [d'=e,e'=-d]d^2+e^2-r^2=0} \qquad \mathbb{R}\,\dfrac{*}{d^2+e^2-r^2=0 \vdash d^2+e^2=r^2}}{d^2+e^2=r^2 \vdash [d'=e,e'=-d]d^2+e^2=r^2}}{\vdash d^2+e^2=r^2 \to [d'=e,e'=-d]d^2+e^2=r^2}$$

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 DI$_{=0}$ appropriately. For other purposes, however, it is still important to have the principle of generalization Note 16 in our repertoire of proof techniques.

9 Example Proofs

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

Example 8 (Self-crossing). Another example is the following invariant property illustrated in Fig. 4:

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

This d$\mathcal{L}$ formula proves easily using DI$_{=0}$:

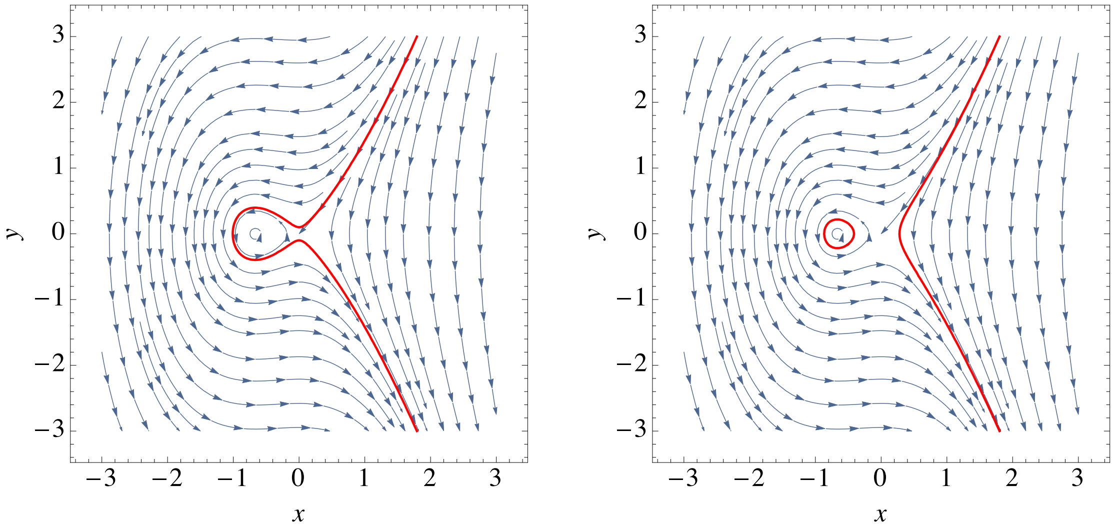


Figure 4: Two differential invariants of the indicated dynamics (illustrated in thick red) for different values of c

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

See ≪Self-crossing polynomial invariant≫

Example 9 (Motzkin). Another nice example is the Motzkin polynomial, which is an invariant of the following dynamics (see Fig. 5):

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

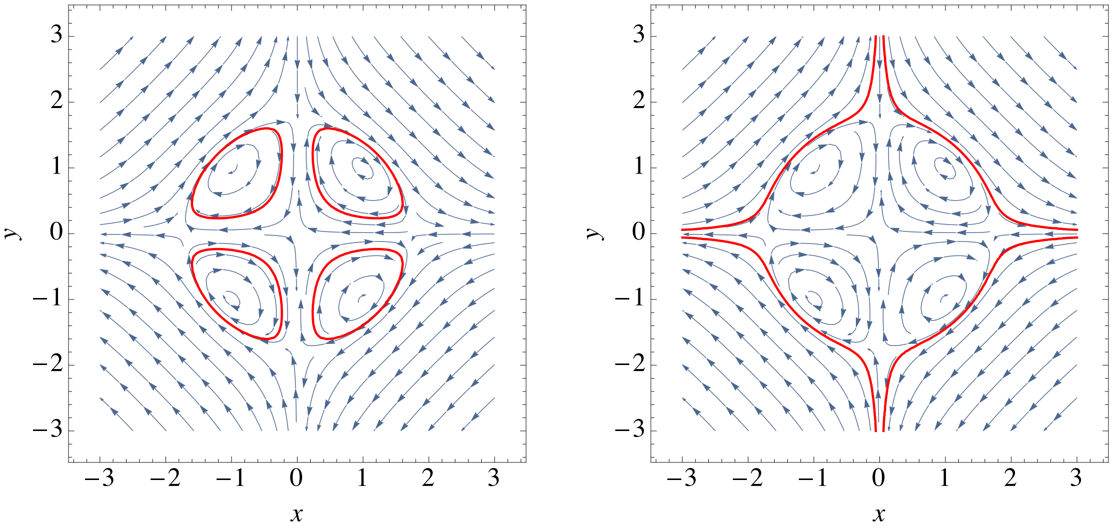


Figure 5: Two differential invariants of the indicated dynamics is the Motzkin polynomial (illustrated in thick red) for different values of c

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

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

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

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

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 9.

10 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 3 (Lie characterization of invariant functions)**.** The proof rule DI$_{=0}$ works by deriving the postcondition and substituting the differential equation in:

$$(\text{DI}_{=0})\ \frac{\vdash {\eta'}_{x'}^{\theta} = 0}{\eta = 0 \vdash [x' = \theta]\eta = 0}$$

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

$$\eta = 0 \vdash [x' = \theta]\eta = 0$$

then it also proves

$$\eta - c = 0 \vdash [x' = \theta]\eta - c = 0 \tag{7}$$

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 10 (Lie [Pla12b])*. Let $x' = \theta$ be a differential equation system and H a* domain, *i.e., a first-order formula of real arithmetic characterizing an 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:*

$$(DI_c)\ \frac{H \vdash {\eta'}_{x'}^{\theta} = 0}{\vdash \forall c\,\big(\eta = c \to [x' = \theta \,\&\, H]\eta = c\big)}$$

Despite the power that differential invariant terms offer, challenges lie ahead in proving properties. Theorem 10 gives an indication where challenges remain.

Example 11 (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 $\mathrm{DI}_{=0}$, because $x^2 + y^2$ is no invariant function of the dynamics. In combination with generalization ($[]gen'$ 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 $\mathrm{DI}_{=0}$:

$$\begin{array}{rc}
 & * \\ \hline
\mathbb{R} & \vdash 4x^3(4y^3) + 4y^3(-4x^3) = 0 \\ \hline
 & \vdash (4x^3x' + 4y^3y')_{x'}^{4y^3}{}_{y'}^{-4x^3} = 0 \\ \hline
\mathrm{DI}_{=0} & x^4 + y^4 = 0 \vdash [x' = 4y^3, y' = -4x^3]\, x^4 + y^4 = 0 \\ \hline
cut,[]gen' & x^2 + y^2 = 0 \vdash [x' = 4y^3, y' = -4x^3]\, x^2 + y^2 = 0 \\ \hline
\to\mathrm{r} & \vdash x^2 + y^2 = 0 \to [x' = 4y^3, y' = -4x^3]\, x^2 + y^2 = 0
\end{array}$$

The use of $[]gen'$ 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 18 (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 10 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].

11 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 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 syntactic derivations. The derivation lemma proved that the values of the (easily computable) syntactic derivations coincides with the analytic derivatives of the values. The differential substitution lemma allowed us the intuitive operation of substituting differential equations into terms. Proving properties of differential equations using 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.

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

Exercises

Exercise 1. Note 2 explained that $(x' = \theta)^*$ is equivalent to $x' = \theta$. Does the same hold for differential equations with evolution domain constraints? Are $(x' = \theta \,\&\, H)^*$ and $x' = \theta \,\&\, H$ equivalent or not? Justify or modify the statement and justify the variation.

Exercise 2. We argued that d$\mathcal{L}$ 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 $[]gen'$ and the *cut* rule.

Exercise 3. Prove the cases of Lemma 5 where η is of the form $a - b$ and a/b.

Exercise 4. What happens in the proof of Lemma 7 if there is no solution φ? Show that this is not a counterexample to proof rule DI$_{=0}$, but that the rule is sound in that case.

Exercise 5. Carry out the polynomial computations needed to prove Example 9 using proof rule DI$_{=0}$.

Exercise 6. Prove the following d$\mathcal{L}$ formula using differential invariants:

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

Exercise 7. Prove the following d$\mathcal{L}$ 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 d$\mathcal{L}$ formula using differential invariants:

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

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

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

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. *Logical Methods in Computer Science*, 8(4):1–38, 2012. doi:10.2168/LMCS-8(4:16)2012.

[Zei03] Eberhard Zeidler, editor. *Teubner-Taschenbuch der Mathematik*. Teubner, 2003.