Domain Specific Languages of Mathematics: Lecture Notes

Patrik Jansson Cezar Ionescu

October 25, 2017

Abstract

These notes aim to cover the lectures and exercises of the recently introduced course "Domain-Specific Languages of Mathematics" (at Chalmers and University of Gothenburg). The course was developed in response to difficulties faced by third-year computer science students in learning and applying classical mathematics (mainly real and complex analysis). The main idea is to encourage the students to approach mathematical domains from a functional programming perspective: to identify the main functions and types involved and, when necessary, to introduce new abstractions; to give calculational proofs; to pay attention to the syntax of the mathematical expressions; and, finally, to organize the resulting functions and types in domain-specific languages.

0 Introduction

These lecture notes aim to cover the lectures and exercises of the recently introduced BSc-level course "Domain Specific Languages of Mathematics" (at Chalmers Univeristy of Technology and University of Gothenburg). The immediate aim of the course is to improve the mathematical education of computer scientists and the computer science education of mathematicians. We believe the course can be the starting point for far-reaching changes, leading to a restructuring of the mathematical training especially for engineers, but perhaps also for mathematicians themselves.

Computer science, viewed as a mathematical discipline, has certain features that set it apart from mainstream mathematics. It places much more emphasis on syntax, tends to prefer formal proofs to informal ones, and views logic as a tool rather than (just) as an object of study. It has long been advocated, both by mathematicians [Wells, 1995, Kraft, 2004] and computer scientists [Gries and Schneider, 1995, Boute, 2009], that the computer science perspective could be valuable in general mathematical education. Until today, this has been convincingly demonstrated (at least since the classical textbook of Gries and Schneider [1993]) only in the field of discrete mathematics. In fact, this demonstration has been so successful, that we increasingly see the discrete mathematics courses being taken over by computer science departments. This is a quite unsatisfactory state of affairs, for at least two reasons.

First, any benefits of the computer science perspective remain within the computer science department and the synergy with the wider mathematical landscape is lost. The mathematics department also misses the opportunity to see more in computer science than just a provider of tools for numerical computations. Considering the increasing dependence of mathematics on software, this can be a considerable loss.

Second, computer science (and other) students are exposed to two quite different approaches to teaching mathematics. For many of them, the formal, tool-oriented style of the discrete mathematics course is easier to follow than the traditional mathematical style. Since, moreover, discrete mathematics tends to be immediately useful to them, this makes the added difficulty of continuous

mathematics even less palatable. As a result, their mathematical competence tends to suffer in areas such as real and complex analysis, or linear algebra.

This is a serious problem, because this lack of competence tends to infect the design of the entire curriculum. For example, a course in "Modeling of sustainable energy systems" for Chalmers' CSE[1] students has to be tailored around this limitation, meaning that the models, methods, and tools that can be presented need to be drastically simplified, to the point where contact with mainstream research becomes impossible.

We propose that a focus on *domain-specific languages* (DSLs) can be used to repair this unsatisfactory state of affairs. In computer science, a DSL "is a computer language specialized to a particular application domain" (Wikipedia), and building DSLs is increasingly becoming a standard industry practice. Empirical studies show that DSLs lead to fundamental increases in productivity, above alternative modelling approaches such as UML [Tolvanen, 2011]. Moreover, building DSLs also offers the opportunity for interdisciplinary activity and can assist in reaching a shared understanding of intuitive or vague notions (see, for example, the work done at Chalmers in cooperation with the Potsdam Institute for Climate Impact Research in the context of Global Systems Science, Lincke et al. [2009], Ionescu and Jansson [2013a], Jaeger et al. [2013], Ionescu and Jansson [2013b], Botta et al. [2017b,a]).

Thus, a course on designing and implementing DSLs can be an important addition to an engineering curriculum. Our key idea is to combine this with a rich source of domains and applications: mathematics. Indeed, mathematics offers countless examples of DSLs: the language of group theory, say, or the language of probability theory, embedded in that of measure theory. The idea that the various branches of mathematics are in fact DSLs embedded in the "general purpose language" of set theory was (even if not expressed in these words) the driving idea of the Bourbaki project, which exerted an enormous influence on present day mathematics.

The course on *DSLs of Mathematics (DSLM)* allows us to present classical mathematical topics in a way which builds on the experience of discrete mathematics: giving specifications of the concepts introduced, paying attention to syntax and types, and so on. For the mathematics students, used to a more informal style, the increased formality is justified by the need to implement (fragments of) the language. We provide a wide range of applications of the DSLs introduced, so that the new concepts can be seen "in action" as soon as possible.

TODO: update with actual learning outcomes

The course has two major learning outcomes. First, the students should be able to design and implement a DSL in a new domain. Second, they should be able to handle new mathematical areas using the computer science perspective.

To achieve these objective, the course consists of a sequence of case studies in which a mathematical area is first presented (for example, a fragment of linear algebra, probability theory, interval analysis, or differential equations), followed by a careful analysis that reveals the domain elements needed to build a language for that domain. The DSL is first used informally, in order to ensure that it is sufficient to account for intended applications (for example, solving equations, or specifying a certain kind of mathematical object). It is in this step that the computer science perspective proves valuable for improving the students' understanding of the mathematical area. The DSL is then implemented in Haskell. The resulting implementation can be compared with existing ones, such as Matlab in the case of linear algebra, or R in the case of statistical computations. Finally, limitations of the DSL are assessed and the possibility for further improvements discussed.

In the first instances, the course is an elective course for the second year within programmes such as CSE, SE, and Math. The potential students will have all taken first-year mathematics courses, and the only prerequisite which some of them will not satisfy will be familiarity with functional programming. However, as the current data structures course (common to the Math and CSE

[1] CSE = Computer Science & Engineering = Datateknik = D

programmes) shows, math students are usually able to catch up fairly quickly, and in any case we aim to keep to a restricted subset of Haskell (no "advanced" features are required).

TODO: rewrite when the evaluation results are available

To assess the impact in terms of increased quality of education, we plan to measure how well the students do in ulterior courses that require mathematical competence (in the case of engineering students) or software compentence (in the case of math students). For example, for CS and CSE students we will measure the percentage of students who, having taken DSLM, pass the third-year courses *Transforms, signals and systems* and *Control Theory (Reglerteknik)*, which are current major stumbling blocks. For math students, we would like to measure their performance in ulterior scientific computing courses.

Since the course is, at least initially, an elective one, we also have the possibility of comparing the results with those of a control group (students who have not taken the course).

The work that lead up to the current course is as follows:

- 2014: in interaction with our colleagues from the various study programmes, we performed an assessment of the current status of potential students for the course in terms of their training (what prerequisites we can reasonably assume) and future path (what mathematical fields they are likely to encounter in later studies), and we worked out a course plan (which we submitted in February 2015, so that the first instance of the course could start in January 2016). We also make a survey of similar courses being offered at other universities, but did not find any close matches.
- 2015: we developed course materials for use within the first instance, wrote a paper [Ionescu and Jansson, 2016] about the course and pressented the pedagogical ideas at several events (TODO: perhaps fill in TFPIE'15, DSLDI'15, IFIP WG 2.1 #73 in Göteborg).
- 2016: we ran the first instance of DSLM (partly paid by the regular course budget, partly by this project) with Cezar Ionescu as main lecturer.
- 2017: we ran the second instance of DSLM (paid fully by the regular course budget), now with Patrik Jansson as main lecturer.
- 2016 and 2017: we used the feedback from students following the standard Chalmers course evaluation in order to improve and further develop the course material.

Future work includes involving faculty from CSE and mathematics in the development of other mathematics courses (prel. Linear Algebra, Analysis) with the aim to incorporate these ideas also there. A major concern will be to work together with our colleagues in the mathematics department in order to distill the essential principles that can be "back-ported" to the other mathematics courses, such as Mathematical Analysis or Linear Algebra. Ideally, the mathematical areas used in DSLM will become increasingly challenging, the more the effective aspects of the computer science perspective are adopted in the first-year mathematics courses.

0.1 About this course

Software engineering involves modelling very different domains (e.g., business processes, typesetting, natural language, etc.) as software systems. The main idea of this course is that this kind of modelling is also important when tackling classical mathematics. In particular, it is useful to introduce abstract datatypes to represent mathematical objects, to specify the mathematical operations performed on these objects, to pay attention to the ambiguities of mathematical notation and understand when they express overloading, overriding, or other forms of generic programming. We shall emphasise the dividing line between syntax (what mathematical expressions look

like) and semantics (what they mean). This emphasis leads us to naturally organise the software abstractions we develop in the form of domain-specific languages, and we will see how each mathematical theory gives rise to one or more such languages, and appreciate that many important theorems establish "translations" between them.

Mathematical objects are immutable, and, as such, functional programming languages are a very good fit for describing them. We shall use Haskell as our main vehicle, but only at a basic level, and we shall introduce the elements of the language as they are needed. The mathematical topic treated have been chosen either because we expect all students to be familiar with them (for example, limits of sequences, continuous functions, derivatives) or because they can be useful in many applications (e.g., Laplace transforms, linear algebra).

0.2 Who should read this book

TODO: Convert formal prerequisites to "what is expected of the reader"

The student should have successfully completed

- a course in discrete mathematics as for example Introductory Discrete Mathematics.
- 15 hec in mathematics, for example Linear Algebra and Calculus
- 15 hec in computer science, for example (Introduction to Programming or Programming with Matlab) and Object-oriented Software Development
- an additional 22.5 hec of any mathematics or computer science courses.

Informally: One full time year (60 hec) of university level study consisting of a mix of mathematics and computer science.

0.3 Roadmap

TODO: write about the lecture plan and internal dependencies

0.4 Notation and code convention

TODO: fill in more about notation

1 Week 1: a DSL for arithmetic expressions and complex numbers

This lecture is partly based on the paper [Ionescu and Jansson, 2016] from the International Workshop on Trends in Functional Programming in Education 2015. We will implement certain concepts in the functional programming language Haskell and the code for this lecture is placed in a module called *DSLsofMath.W01* that starts here:

```
module DSLsofMath.W01 where
```

1.1 A case study: complex numbers

We will start by an analytic reading of the introduction of complex numbers in Adams and Essex [2010]. We choose a simple domain to allow the reader to concentrate on the essential elements of our approach without the distraction of potentially unfamiliar mathematical concepts. For this section, we bracket our previous knowledge and approach the text as we would a completely new domain, even if that leads to a somewhat exaggerated attention to detail.

Adams and Essex introduce complex numbers in Appendix 1. The section *Definition of Complex Numbers* begins with:

> We begin by defining the symbol i, called **the imaginary unit**, to have the property
>
> $$i^2 = -1$$
>
> Thus, we could also call i the square root of -1 and denote it $\sqrt{-1}$. Of course, i is not a real number; no real number has a negative square.

At this stage, it is not clear what the type of i is meant to be, we only know that i is not a real number. Moreover, we do not know what operations are possible on i, only that i^2 is another name for -1 (but it is not obvious that, say $i * i$ is related in any way with i^2, since the operations of multiplication and squaring have only been introduced so far for numerical types such as $\mathbb{N}$ or $\mathbb{R}$, and not for symbols).

For the moment, we introduce a type for the value i, and, since we know nothing about other values, we make i the only member of this type:

```
data ImagUnits = I
i :: ImagUnits
i = I
```

We use a capital I in the **data** declaration because a lowercase constructor name would cause a syntax error in Haskell.

Next, we have the following definition:

> **Definition:** A **complex number** is an expression of the form
>
> $$a + bi \qquad \text{or} \qquad a + ib,$$
>
> where a and b are real numbers, and i is the imaginary unit.

This definition clearly points to the introduction of a syntax (notice the keyword "form"). This is underlined by the presentation of *two* forms, which can suggest that the operation of juxtaposing i (multiplication?) is not commutative.

A profitable way of dealing with such concrete syntax in functional programming is to introduce an abstract representation of it in the form of a datatype:

```
data ComplexA = CPlus1 ℝ ℝ ImagUnits
              | CPlus2 ℝ ImagUnits ℝ
```

We can give the translation from the abstract syntax to the concrete syntax as a function *showCA*:

```
showCA :: ComplexA →      String
showCA   (CPlus1 x y i) = show x ++ " + " ++ show y ++ "i"
showCA   (CPlus2 x i y) = show x ++ " + " ++ "i" ++ show y
```

Notice that the type $\mathbb{R}$ is not implemented yet and it is not really even exactly implementable but we want to focus on complex numbers so we will approximate $\mathbb{R}$ by double precision floating point numbers for now.

> **type** $\mathbb{R} = \mathit{Double}$

The text continues with examples:

> For example, $3 + 2\ i$, $\frac{7}{2} - \frac{2}{3}\ i$, $i\ \pi = 0 + i\ \pi$, and $-3 = -3 + 0\ i$ are all complex numbers. The last of these examples shows that every real number can be regarded as a complex number.

The second example is somewhat problematic: it does not seem to be of the form $a+bi$. Given that the last two examples seem to introduce shorthand for various complex numbers, let us assume that this one does as well, and that $a - bi$ can be understood as an abbreviation of $a + (-b)\ i$.

With this provision, in our notation the examples are written as:

```
testC1 :: [ComplexA]
testC1 = [ CPlus1 3 2 I, CPlus1 (7 / 2) (−2 / 3) I
         , CPlus2 0 I π, CPlus1 (−3) 0 I
         ]
testS1 = map showCA testC1
```

We interpret the sentence "The last of these examples ..." to mean that there is an embedding of the real numbers in *ComplexA*, which we introduce explicitly:

```
toComplex :: ℝ → ComplexA
toComplex x = CPlus1 x 0 i
```

Again, at this stage there are many open questions. For example, we can assume that $i\ 1$ stands for the complex number $\mathit{CPlus}_2\ 0\ i\ 1$, but what about i by itself? If juxtaposition is meant to denote some sort of multiplication, then perhaps 1 can be considered as a unit, in which case we would have that i abbreviates $i\ 1$ and therefore $\mathit{CPlus}_2\ 0\ i\ 1$. But what about, say, $2\ i$? Abbreviations with i have only been introduced for the ib form, and not for the bi one!

The text then continues with a parenthetical remark which helps us dispel these doubts:

> (We will normally use $a + bi$ unless b is a complicated expression, in which case we will write $a + ib$ instead. Either form is acceptable.)

This remark suggests strongly that the two syntactic forms are meant to denote the same elements, since otherwise it would be strange to say "either form is acceptable". After all, they are acceptable by definition.

Given that $a + ib$ is only "syntactic sugar" for $a + bi$, we can simplify our representation for the abstract syntax, eliminating one of the constructors:

> **data** *ComplexB* = *CPlusB* $\mathbb{R}\ \mathbb{R}$ *ImagUnits*

In fact, since it doesn't look as though the type *ImagUnits* will receive more elements, we can dispense with it altogether:

> **data** *ComplexC* = *CPlusC* $\mathbb{R}\ \mathbb{R}$

(The renaming of the constructor to *CPlusC* serves as a guard against the case we have suppressed potentially semantically relevant syntax.)

We read further:

> It is often convenient to represent a complex number by a single letter; w and z are frequently used for this purpose. If a, b, x, and y are real numbers, and $w = a + bi$ and $z = x + yi$, then we can refer to the complex numbers w and z. Note that $w = z$ if and only if $a = x$ and $b = y$.

First, let us notice that we are given an important semantic information: *CPlusC* is not just syntactically injective (as all constructors are), but also semantically. The equality on complex numbers is what we would obtain in Haskell by using **deriving** *Eq*.

This shows that complex numbers are, in fact, isomorphic with pairs of real numbers, a point which we can make explicit by re-formulating the definition in terms of a **newtype**:

```
type ComplexD = ComplexSem ℝ
newtype ComplexSem r = CS (r, r) deriving Eq
```

The point of the somewhat confusing discussion of using "letters" to stand for complex numbers is to introduce a substitute for *pattern matching*, as in the following definition:

> **Definition:** If $z = x + yi$ is a complex number (where x and y are real), we call x the **real part** of z and denote it $Re\ (z)$. We call y the **imaginary part** of z and denote it $Im\ (z)$:
>
> $$Re\ (z) = Re\ (x + yi) = x$$
> $$Im\ (z) = Im\ (x + yi) = y$$

This is rather similar to Haskell's *as-patterns*:

```
re :: ComplexSem r  → r
re z@(CS (x, y))  = x
im :: ComplexSem r → r
im z@(CS (x, y)) = y
```

a potential source of confusion being that the symbol z introduced by the as-pattern is not actually used on the right-hand side of the equations.

The use of as-patterns such as "$z = x + yi$" is repeated throughout the text, for example in the definition of the algebraic operations on complex numbers:

> **The sum and difference of complex numbers**
>
> If $w = a + bi$ and $z = x + yi$, where a, b, x, and y are real numbers, then
>
> $$w + z = (a + x) + (b + y)\ i$$
> $$w - z = (a - x) + (b - y)\ i$$

With the introduction of algebraic operations, the language of complex numbers becomes much richer. We can describe these operations in a *shallow embedding* in terms of the concrete datatype *ComplexSem*, for example:

```
(+.) :: Num r ⇒ ComplexSem r → ComplexSem r → ComplexSem r
(CS (a, b)) +. (CS (x, y)) = CS ((a + x), (b + y))
```

or we can build a datatype of "syntactic" complex numbers from the algebraic operations to arrive at a *deep embedding* as seen in the next section.

Exercises:

- implement $(*.)$ for *ComplexSem*

1.2 A syntax for arithmetical expressions

So far we have tried to find a datatype to represent the intended *semantics* of complex numbers. That approach is called "shallow embedding". Now we turn to the *syntax* instead ("deep embedding").

We want a datatype *ComplexE* for the abstract syntax tree of expressions. The syntactic expressions can later be evaluated to semantic values:

$$evalE :: ComplexE \to ComplexD$$

The datatype *ComplexE* should collect ways of building syntactic expression representing complex numbers and we have so far seen the symbol i, an embedding from $\mathbb{R}$, plus and times. We make these four *constructors* in one recursive datatype as follows:

$$\begin{aligned}&\textbf{data } ComplexE = ImagUnit\\ &\qquad\qquad\quad \mid ToComplex\ \mathbb{R}\\ &\qquad\qquad\quad \mid Plus\quad ComplexE\ ComplexE\\ &\qquad\qquad\quad \mid Times\ ComplexE\ ComplexE\\ &\textbf{deriving } (Eq, Show)\end{aligned}$$

And we can write the evaluator by induction over the syntax tree:

$$\begin{aligned}&evalE\ ImagUnit &&= CS\ (0,1)\\ &evalE\ (ToComplex\ r) &&= CS\ (r,0)\\ &evalE\ (Plus\ c_1\ c_2) &&= evalE\ c_1 +. evalE\ c_2\\ &evalE\ (Times\ c_1\ c_2) &&= evalE\ c_1 *. evalE\ c_2\end{aligned}$$

We also define a function to embed a semantic complex number in the syntax:

$$\begin{aligned}&fromCS :: ComplexD \to ComplexE\\ &fromCS\ (CS\ (x,y)) = Plus\ (ToComplex\ x)\ (Times\ (ToComplex\ y)\ ImagUnit)\\ &testE1 = Plus\ (ToComplex\ 3)\ (Times\ (ToComplex\ 2)\ ImagUnit)\\ &testE2 = Times\ ImagUnit\ ImagUnit\end{aligned}$$

There are certain laws we would like to hold for operations on complex numbers. The simplest is perhaps $i^2 = -1$ from the start of the lecture,

$$\begin{aligned}&propImagUnit :: Bool\\ &propImagUnit = Times\ ImagUnit\ ImagUnit === ToComplex\ (-1)\\ &(===) :: ComplexE \to ComplexE \to Bool\\ &z === w = evalE\ z == evalE\ w\end{aligned}$$

and that *fromCS* is an embedding:

$$\begin{aligned}&propFromCS :: ComplexD \to Bool\\ &propFromCS\ c = evalE\ (fromCS\ c) == c\end{aligned}$$

but we also have that *Plus* and *Times* should be associative and commutative and *Times* should distribute over *Plus*:

$$\begin{aligned}&propAssocPlus\ x\ y\ z &&= Plus\ (Plus\ x\ y)\ z === Plus\ x\ (Plus\ y\ z)\\ &propAssocTimes\ x\ y\ z &&= Times\ (Times\ x\ y)\ z === Times\ x\ (Times\ y\ z)\\ &propDistTimesPlus\ x\ y\ z &&= Times\ x\ (Plus\ y\ z) === Plus\ (Times\ x\ y)\ (Times\ x\ z)\end{aligned}$$

These three laws actually fail, but not because of the implementation of *evalE*. We will get back to that later but let us first generalise the properties a bit by making the operator a parameter:

```
propAssocA :: Eq a ⇒ (a → a → a) → a → a → a → Bool
propAssocA (+?) x y z = (x +? y) +? z == x +? (y +? z)
```

Note that *propAssocA* is a higher order function: it takes a function (a binary operator) as its first parameter. It is also polymorphic: it works for many different types a (all types which have an == operator).

Thus we can specialise it to *Plus*, *Times* and other binary operators. In Haskell there is a type class *Num* for different types of "numbers" (with operations (+), (∗), etc.). We can try out *propAssocA* for a few of them.

```
propAssocAInt = propAssocA (+) :: Int → Int → Int → Bool
propAssocADouble = propAssocA (+) :: Double → Double → Double → Bool
```

The first is fine, but the second fails due to rounding errors. QuickCheck can be used to find small examples - I like this one best:

```
notAssocEvidence :: (Double, Double, Double, Bool)
notAssocEvidence = (lhs, rhs, lhs − rhs, lhs == rhs)
  where lhs = (1 + 1) + 1 / 3
        rhs = 1 + (1 + 1 / 3)
```

For completeness: this is the answer:

```
(2.3333333333333335      -- Notice the five at the end
,2.333333333333333,      -- which is not present here.
,4.440892098500626e−16   -- The difference
,False)
```

This is actually the underlying reason why some of the laws failed for complex numbers: the approximative nature of *Double*. But to be sure there is no other bug hiding we need to make one more version of the complex number type: parameterise on the underlying type for $\mathbb{R}$. At the same time we generalise *ToComplex* to *FromCartesian*:

```
data ComplexSyn r = FromCartesian r r
                  | ComplexSyn r :+: ComplexSyn r
                  | ComplexSyn r :∗: ComplexSyn r

toComplexSyn :: Num a ⇒ a → ComplexSyn a
toComplexSyn x = FromCartesian x (fromInteger 0)

evalCSyn :: Num r ⇒ ComplexSyn r → ComplexSem r
evalCSyn (FromCartesian x y) = CS (x, y)
evalCSyn (l :+: r) = evalCSyn l +. evalCSyn r
evalCSyn (l :∗: r) = evalCSyn l ∗. evalCSyn r

instance Num a ⇒ Num (ComplexSyn a) where
  (+) = (:+:)
  (∗) = (:∗:)
  fromInteger = fromIntegerCS
    -- TODO: add a few more operations (hint: extend ComplexSyn as well)
    -- TODO: also extend eval

fromIntegerCS :: Num r ⇒ Integer → ComplexSyn r
fromIntegerCS = toComplexSyn ∘ fromInteger
```

1.3 TODO[PaJa]: Textify

Here are some notes about things scribbled on the blackboard during the first two lectures. At some point this should be made into text for the lecture notes.

1.3.1 Pitfalls with traditional mathematical notation

A function or the value at a point? Mathematical texts often talk about "the function $f(x)$" when "the function f" would be more clear. Otherwise there is a clear risk of confusion between $f(x)$ as a function and $f(x)$ as the value you get from applying the function f to the value bound to the name x.

Scoping Scoping rules for the integral sign:

$$f(x) = x^2$$
$$g(x) = \int_x^{2x} f(x)dx \qquad\qquad = \int_x^{2x} f(y)dy$$

The variable x bound on the left is independent of the variable x "bound under the integral sign".

From syntax to semantics and back We have seen evaluation functions from abstract syntax to semantics ($eval :: Syn \to Sem$). Often a partial inverse is also available: $embed :: Sem \to Syn$. For our complex numbers we have TODO: fill in a function from $ComplexSem\ r \to ComplexSyn\ r$.

The embedding should satisfy a round-trip property: $eval\ (embed\ s) == s$ for all s. Exercise: What about the opposite direction? When is $embed\ (eval\ e) == e$?

We can also state and check properties relating the semantic and the syntactic operations:

$a + b = eval\ (Plus\ (embed\ a)\ (embed\ b))$ for all a and b.

Variable names as type hints In mathematical texts there are often conventions about the names used for variables of certain types. Typical examples include i, j, k for natural numbers or integers, x, y for real numbers and z, w for complex numbers.

The absence of explicit types in mathematical texts can sometimes lead to confusing formulations. For example, a standard text on differential equations by Edwards, Penney and Calvis Edwards et al. [2008] contains at page 266 the following remark:

> The differentiation operator D can be viewed as a transformation which, when applied to the function $f(t)$, yields the new function $D\{f(t)\} = f'(t)$. The Laplace transformation $\mathcal{L}$ involves the operation of integration and yields the new function $\mathcal{L}\{f(t)\} = F(s)$ of a new independent variable s.

This is meant to introduce a distinction between "operators", such as differentiation, which take functions to functions of the same type, and "transforms", such as the Laplace transform, which take functions to functions of a new type. To the logician or the computer scientist, the way of phrasing this difference in the quoted text sounds strange: surely the *name* of the independent variable does not matter: the Laplace transformation could very well return a function of the "old" variable t. We can understand that the name of the variable is used to carry semantic meaning about its type (this is also common in functional programming, for example with the conventional use of *as* to denote a list of *a*s). Moreover, by using this (implicit!) convention, it is easier to deal with cases such as that of the Hartley transform (a close relative of the Fourier transform), which

does not change the type of the input function, but rather the *interpretation* of that type. We prefer to always give explicit typings rather than relying on syntactical conventions, and to use type synonyms for the case in which we have different interpretations of the same type. In the example of the Laplace transformation, this leads to

$$\begin{aligned}&\textbf{type } T = \mathit{Real}\\&\textbf{type } S = \mathbb{C}\\&\mathcal{L} : (T \to \mathbb{C}) \to (S \to \mathbb{C})\end{aligned}$$

1.3.2 Other

Lifting operations to a parameterised type When we define addition on complex numbers (represented as pairs of real and imaginary components) we can do that for any underlying type r which supports addition.

$$\begin{aligned}&\textbf{type } \mathit{CS} = \mathit{ComplexSem} \quad \text{-- for shorter type expressions below}\\&\mathit{liftPlus} :: (r \quad \to r \quad \to r \quad) \to\\&\qquad (\mathit{CS}\ r \to \mathit{CS}\ r \to \mathit{CS}\ r)\\&\mathit{liftPlus}\ (+)\ (\mathit{CS}\ (x, y))\ (\mathit{CS}\ (x', y')) = \mathit{CS}\ (x + x', y + y')\end{aligned}$$

Note that *liftPlus* takes $(+)$ as its first parameter and uses it twice on the RHS.

Laws TODO: Associative, Commutative, Distributive, ...

TODO[PaJa]: move earlier Table of examples of notation and abstract syntax for some complex numbers:

Mathematics	Haskell
$3 + 2i$	$\mathit{CPlus}_1\ 3\ 2\ i$
$\frac{7}{2} - \frac{2}{3}i = \frac{7}{2} + \frac{-2}{3}i$	$\mathit{CPlus}_1\ (7\ /\ 2)\ (-2\ /\ 3)\ i$
$i\pi = 0 + i\pi$	$\mathit{CPlus}_2\ 0\ i\ \pi$
$-3 = -3 + 0i$	$\mathit{CPlus}_1\ (-3)\ 0\ i$

1.4 Questions and answers from the exercise sessions week 1

1.4.1 Function composition

The infix operator `.` in Haskell is an implementation of the mathematical operation of function composition.

$$f \circ g = \lambda x \to f\ (g\ x)$$

The period is an ASCII approximation of the composition symbol $\circ$ typically used in mathematics. (The symbol $\circ$ is encoded as `U+2218` and called RING OPERATOR in Unicode, `&#8728` in HTML, `\circ` in TEX, etc.)

The type is perhaps best illustrated by a diagram with types as nodes and functions (arrows) as directed edges:

In Haskell we get the following type:

$$(\circ) :: (b \to c) \to (a \to b) \to (a \to c)$$

which may take a while to get used to.

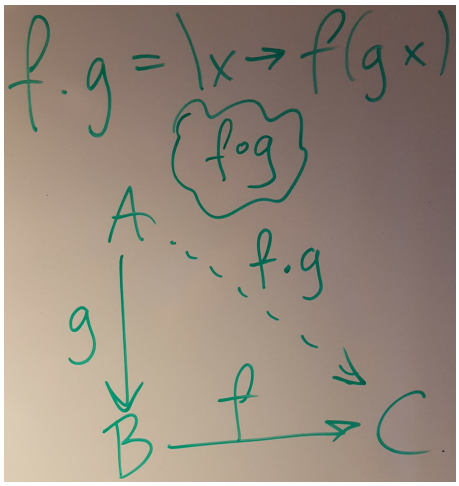


Figure 1: Function composition diagram

1.4.2 fromInteger (looks recursive)

Near the end of the lecture notes there was an instance declaration including the following lines:

> **instance** *Num r* ⇒ *Num* (*ComplexSyn r*) **where**
> -- ... several other methods and then
> *fromInteger* = *toComplexSyn* ∘ *fromInteger*

This definition looks recursive, but it is not. To see why we need to expand the type and to do this I will introduce a name for the right hand side (RHS): *fromIntC*.

```
--          ComplexSyn r <---------- r <---------- Integer
fromIntC =              toComplexSyn . fromInteger
```

I have placed the types in the comment, with "backwards-pointing" arrows indicating that *fromInteger*:: *Integer* → *r* and *toComplexSyn* :: *r* → *ComplexSyn r* while the resulting function is *fromIntC* :: *Integer* → *ComplexSyn r*. The use of *fromInteger* at type *r* means that the full type of *fromIntC* must refer to the *Num* class. Thus we arrive at the full type:

> *fromIntC* :: *Num r* ⇒ *Integer* → *ComplexSyn r*

1.4.3 type / newtype / data

There are three keywords in Haskell involved in naming types: **type**, **newtype**, and **data**.

type – abbreviating type expressions The **type** keyword is used to create a type synonym - just another name for a type expression.

> **type** *Heltal* = *Integer*
> **type** *Foo* = (*Maybe* [*String*], [[*Heltal*]])
> **type** *BinOp* = *Heltal* → *Heltal* → *Heltal*
> **type** *Env v s* = [(*v*, *s*)]

The new name for the type on the RHS does not add type safety, just readability (if used wisely). The *Env* example shows that a type synonym can have type parameters.

newtype – more protection A simple example of the use of **newtype** in Haskell is to distinguish values which should be kept apart. A simple example is

> **newtype** *Age* = *Ag Int* -- Age in years
> **newtype** *Shoe* = *Sh Int* -- Shoe size (EU)

Which introduces two new types, *Age* and *Shoe*, which both are internally represented by an *Int* but which are good to keep apart.

The constructor functions $Ag :: Int \to Age$ and $Sh :: Int \to Shoe$ are used to translate from plain integers to ages and shoe sizes.

In the lecture notes we used a newtype for the semantics of complex numbers as a pair of numbers in the cartesian representation but may also be useful to have another newtype for complex as a pair of numbers in the polar representation.

data – for syntax trees Some examples:

> **data** $N = Z \mid S\ N$

This declaration introduces

- a new type N for unary natural numbers,
- a constructor $Z :: N$ to represent zero, and
- a constructor $S :: N \to N$ to represent the successor.

Examples values: $zero = Z$, $one = S\ Z$, $three = S\ (S\ one)$

> **data** $E = V\ String \mid P\ E\ E \mid T\ E\ E$

This declaration introduces

- a new type E for simple arithmetic expressions,
- a constructor $V :: String \to E$ to represent variables,
- a constructor $P :: E \to E \to E$ to represent plus, and
- a constructor $T :: E \to E \to E$ to represent times.

Example values: $x = V$ `"x"`, $e_1 = P\ x\ x$, $e_2 = T\ e_1\ e_1$

If you want a contructor to be used as an infix operator you need to use symbol characters and start with a colon:

> **data** $E' = V'\ String \mid E' :+ E' \mid E' :* E'$

Example values: $y = V$ `"y"`, $e_1 = y :+ y$, $e_2 = x :* e_1$

Finally, you can add one or more type parameters to make a whole family of datatypes in one go:

> **data** *ComplexSy v r* = *Var v*
> | *FromCart r r*
> | *ComplexSy v r* :+ *ComplexSy v r*
> | *ComplexSy v r* :* *ComplexSy v r*

The purpose of the first parameter v here is to enable a free choice of type for the variables (be it *String* or *Int* or something else) and the second parameter r makes is possible to express "complex numbers over" different base types (like *Double*, *Float*, *Integer*, etc.).

1.4.4 *Env*, *Var*, and variable lookup

The type synonym

$$\textbf{type}\ Env\ v\ s = [(v, s)]$$

is one way of expressing a partial function from v to s.

Example value:

```
env1 :: Env String Int
env1 = [("hej", 17), ("du", 38)]
```

The *Env* type is commonly used in evaluator functions for syntax trees containing variables:

```
evalCP :: Eq v => Env v (ComplexSem r) -> (ComplexSy v r -> ComplexSem r)
evalCP env (Var x) = case lookup x env of
  Just c -> undefined   -- ...
  -- ...
```

Notice that *env* maps "syntax" (variable names) to "semantics", just like the evaluator does.

1.5 Some helper functions

```
propAssocAdd :: (Eq a, Num a) => a -> a -> a -> Bool
propAssocAdd = propAssocA (+)

(*.) :: Num r => ComplexSem r -> ComplexSem r -> ComplexSem r
CS (ar, ai) *. CS (br, bi) = CS (ar * br - ai * bi, ar * bi + ai * br)

instance Show r => Show (ComplexSem r) where
  show = showCS

showCS :: Show r => ComplexSem r -> String
showCS (CS (x, y)) = show x ++ " + " ++ show y ++ "i"
```

TODO: Perhaps formulate exercise to implement more efficient show using an ackumulating parameter.

2 Week 2: Logic and calculational proofs

Course learning outcomes:

- Knowledge and understanding
  - design and implement a DSL (Domain Specific Language) for a new domain
  - organize areas of mathematics in DSL terms
  - explain main concepts of elementary real and complex analysis, algebra, and linear algebra
- Skills and abilities
  - develop adequate notation for mathematical concepts
  - perform calculational proofs
  - use power series for solving differential equations
  - use Laplace transforms for solving differential equations
- Judgement and approach
  - discuss and compare different software implementations of mathematical concepts

This week we focus on "develop adequate notation for mathematical concepts" and "perform calculational proofs" (still in the context of "organize areas of mathematics in DSL terms").

```
{-# LANGUAGE GADTs #-}
module DSLsofMath.W02 where
```

2.1 Propositional Calculus

Now we turn to the main topic of this week: logic and proofs.

TODO: tidy up the notes below

Swedish: Satslogik

A, B, C, ...	names of propositions	
False, *True*	Constants	
And	$\wedge$	&
Or	$\vee$	||
Implies	$\Rightarrow$	
Not	$\neg$	

Example:

```
sw :: PropCalc
sw = ((A & B) -=> (B & A))
  where a = N "A"
        b = N "B"
```

The example is based on the following embedding of propositional calculus terms:

```
data PropCalc where
  N     :: Name → PropCalc
```

```
C       :: Bool → PropCalc
And     :: PropCalc → PropCalc → PropCalc
Or      :: PropCalc → PropCalc → PropCalc
Implies :: PropCalc → PropCalc → PropCalc
Not     :: PropCalc → PropCalc
(&) = And
(-=>) = Implies
type Name = String
```

With this datatype we can write an evaluator to *Bool* which computes the truth value of a term given an enviroment:

```
evalPC :: (Name → Bool) → PropCalc → Bool
evalPC = error "Exercise"
```

The function *evalPC* translates from the syntactic to the semantic domain. Here *PropCalc* is the (abstract) *syntax* of the language of propositional calculus and *Bool* is the *semantic domain*. Alternatively, we can view $(Name \to Bool) \to Bool$ as the semantic domain. A value of this type is a mapping from a truth table to *Bool*. This mapping is often also tabulated as a truth table with one more "output" column.

As a first example, consider the proposition $t = Implies\ False\ a$. The truth table semantics of t is usually drawn as follows: one column for the name a listing all combinations of $T = Truth$ and $F = False$, and one column for the result of evaluating the expression.

a	t
F	T
T	T

If we continue with the example *sw* from above we have two names a and b which together can have any of four combinations of true and false. After the name-columns are filled, we fill in the rest of the table one operation (column) at a time. The & columns become $F\ F\ F\ T$ and finally the $\Rightarrow$ column becomes true everywhere.

a	&	b	$\Rightarrow$	b	&	a
F	F	F	T	F	F	F
F	F	T	T	T	F	F
T	F	F	T	F	F	T
T	T	T	T	T	T	T

A proposition whose truth table is always true is called a *tautology*. Truth table verification is only viable for propositions with few names because of the exponential growth in the number of cases to check: we get 2^n cases for n names. (There are very good heuristic algorithms to look for tautologies even for thousands of names — but that is not part of this course.)

What we call "names" are often called "(propositional) variables" but we will soon add another kind of variables (and quantification over them) to the calculus.

2.2 First Order Logic (predicate logic)

TODO: type up the notes + whiteboard photos

Swedish: Första ordningens logik = predikatlogik

TODO: Adds term variables and functions, predicate symbols and quantifiers (sv: kvantorer).

We now add *terms* as another datatype to the calculus. A *term* is either a (term) *variable* (like x, y, z), or the application of a *function symbol* (like f, g) to a suitable number of terms. If we have the function symbols f of arity 2 and g of arity 3 we can form terms like $f\ (x, x)$, $g\ (y, z, z)$, $g\ (x, y, f\ (x, y))$, etc.

The names from the propositional calculus are generalised to *predicate symbols* of different arity. The predicate symbols can only be applied to terms, not to other predicate symbols or formulas. If we have the predicate symbols N of arity 0, P of arity 1 and Q of arity 2 we can form *formulas* like N, $P\ (x)$, $Q\ (f\ (x, x), y)$, etc.

Note that we have two separate layers: terms cannot contain formulas, but formulas normally contain terms.

The formulas introduced so far are all *atomic formulas* but we will add two more concepts: first the logical conenctives from the propositional calculus: *And*, *Or*, *Implies*, *Not*, and then two quantifiers: $\forall\ \cdot .\ \cdot$ and $\exists\ \cdot .\ \cdot$.

An example FOL formula:

$$\forall x(P(x) \to (\exists y(Q(f(x, x), y))))$$

Note that FOL can only quantify over *term* variables, not over predicates. (Second order logic and higher order logic allow quantification over predicates.)

Another example: a formula stating that + is commutative:

$$\forall x(\forall y((x + y) == (y + x)))$$

Note that == is a binary predicate symbol while + is a binary function symbol. Here is the same formula without infix operators:

$$\forall\ x.\ \ \forall\ y.\ \ Eq\ (plus\ (x, y), plus\ (y, x))$$

Forall quantification can be seen as a generalisation of *And*. First we can generalise the binary operator to an n-ary version: And_n. To prove $And_n\ A_1\ A_2\ ...\ A_n$ we need a proof of each A_i. Thus we could define $And_n\ A_1\ A_2\ ...\ A_n = A_1\ \&\ A_2\ \&\ ...\ \&\ A_n$ where & is the infix version of binary *And*. The next step is to note that the formulas A_i can be generalised to $A\ (i)$ where i is a term variable and A is a unary predicate symbol. We can think of i ranging over an infinite collection of constant terms $i0$, i_1, ... Then the final step is to introduce the notation $\forall\ i.\ \ A\ (i)$ for $A\ (i_1)\ \&\ A\ (i_2)\ \&\ ....$

Now to prove $\forall\ x.\ \ P\ (x)$ it would be difficult to provide an infinite collection of proofs of $P\ (x_i)$. Instead the standard procedure is to introduce a fresh constant term a and prove $P\ (a)$. Another way to view this is to say that a proof of $\forall\ x.\ \ P\ x$ is a function f from terms to proofs such that $f\ t$ is a proof of $P\ t$ for all terms t.

2.3 An aside: Pure set theory

One way to build mathematics from the ground up is to start from pure set theory and define all concepts by translation to sets. We will only work with this as a mathematical domain to study, not as "the right way" of doing mathematics. The core of the language of pure set theory has the Empty set, the one-element set constructor Singleton, set Union, and Intersection. There are no "atoms" or "elements" to start from except for the empty set but it turns out that quite a large part of mathematics can still be expressed.

Natural numbers To talk about things like natural numbers in pure set theory they need to be encoded. Here is one such encoding (which is explored further in the first hand-in assignment).

$$\begin{aligned} vonNeumann\ 0 &= Empty \\ vonNeumann\ (n+1) &= Union\ (vonNeumann\ n) \\ &\qquad (Singleton\ (vonNeumann\ n)) \end{aligned}$$

Pairs Definition: A pair (a, b) is encoded as $\{\{a\}, \{a, b\}\}$.

TODO: merge the text below and above

As an example term language we can use pure (untyped) set theory. We have a nullary function symbol $\{\}$ for the empty set (sometimes writen $\emptyset$) and a unary function symbol S for the function that builds a singleton set from an "element". In pure set theory we don't actually have any "elements" to start from: every term denotes a set. All non-variable terms so far are $\{\}$, $S\ \{\}$, $S\ (S\ \{\})$, ... The first set is empty but all the others are one-element sets.

Next we add two binary function symbols for union and intersection of sets (denoted by terms). Using union we can build sets of more than one element, for example $\mathit{Union}\ (S\ \{\})\ (S\ (S\ \{\}))$ which has two "elements": $\{\}$ and $S\ \{\}$.

FOL does not have function definitions or recursion, but in a suitable meta-langauge (like Haskell) we can write a function that creates a set with n elements (for any natural number n) as a term in FOL:

$$
\begin{aligned}
&\mathit{vN}\ 0 &&= \{\}\\
&\mathit{vN}\ (n+1) &&= \mathit{step}\ (\mathit{vN}\ n)\\
&\mathit{step}\ x &&= \mathit{Union}\ x\ (S\ x)
\end{aligned}
$$

If we use conventional set notation we get $\mathit{vN}\ 0 = \{\}$, $\mathit{vN}\ 1 = \{\{\}\}$, $\mathit{vN}\ 2 = \{\{\}, \{\{\}\}\}$, $\mathit{vN}\ 3 = \{\{\}, \{\{\}\}, \{\{\}, \{\{\}\}\}\}$, etc. If we use the shorthand $\overline{n}$ for $\mathit{vN}\ n$ we see that $\overline{0} = \{\}$, $\overline{1} = \{\overline{0}\}$, $\overline{2} = \{\overline{0}, \overline{1}\}$, $\overline{3} = \{\overline{0}, \overline{1}, \overline{2}\}$ and, in general, that $\overline{n}$ has cardinality n. The function vN is explored in more detail in the first assignment of the DSLsofMath course.

The constructions presented so far show that, even starting from no elements, we can embed all natural numbers in pure set theory. We can also embed unordered pairs: $\{a, b\} \stackrel{\text{def}}{=} \mathit{Union}\ (S\ a)\ (S\ b)$ and normal ordered pairs: $(a, b) \stackrel{\text{def}}{=} \{S\ a, \{a, b\}\}$. With a bit more machinery it is possible to step by step encode $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.

2.4 Back to quantifiers

After this detour through untyped set land let us get back to the most powerful concept of FOL: the quantifiers. We have already seen how the "forall" quantifier can be seen as a generalisation of *And* and in the same way we can see the "exists" quantifier as a generalisation of *Or*.

First we generalise the binary *Or* to an n-ary Or_n. To prove $\mathit{Or}_n\ A_1\ A_2\ ...\ A_n$ is enough (and necessary) to find one i for which we can prove A_i. As before we then take the step from a family of formulas A_i to one unary predicate A expressing the formulas $A\ (i)$ for the term variable i. Then the final step is to "or" all these formulas to obtain $\exists\ i.\ \ A\ i$.

At this point it is good to sum up and compare the two quantifiers and how to prove them:

(t, b_t) is a proof of $\exists\ x.\ \ P\ (x)$ if b_t is a proof of $P\ (t)$.

f is a proof of $\forall\ x.\ \ P\ (x)$ if $f\ t$ is a proof of $P\ (t)$ for all t.

If we abbreviate "is a proof" as : and use the Haskell convention for function application we get

$$
\begin{aligned}
&(t, b_t) &&: \exists\ x.\ \ P\ x &&\quad \textbf{if} \quad b_t\ &&: P\ t\\
&f &&: \forall\ x.\ \ P\ x &&\quad \textbf{if} \quad f\ t &&: P\ t \ \text{ for all } t
\end{aligned}
$$

This now very much looks like type rules, and that is not a coincidence. The Curry-Howard correspondence says that we can think of propositions as types and proofs as "programs". These

typing judgments are not part of FOL, but the correspondence is used quite a bit in this course to keep track of proofs.

TODO: Add more about Curry-Howard (the binary logical connectives, etc.)

TODO: find the right place for the a note that the type of tuples is isomorphic to the (dependent) function type $\{i : 1 \mathinner{..} n\} \to A_i$.

TODO: Add typed quantification for Exists.

(Roughly: $\exists\ x : T.\ \ P\ x = \exists\ x.\ \ T\ x\ \&\ P\ x$.)

2.5 Proof by contradition

Let's try to express and prove the irrationality of the square root of 2. We have two main concepts involved: the predicate "irrational" and the function "square root of". The square root function (for positive real numbers) can be specified by $r = \sqrt{s}$ iff $r\hat{}2 == s$ and $r : \mathbb{N}$. The formula "x is irrational" is just $\neg\ (R\ x)$ where R is the predicate "is rational".

$$R\ x = \exists\ a : \mathbb{N}.\ \ \exists\ b : \mathbb{N}_{>0}.\ \ b * x == a\ \&\ GCD\ (a, b) == 1$$

The classical way to prove a negation $\neg\ P$ is to assume P and derive something absurd (some Q and $\neg\ Q$, for example). Lets take $P = R\ r$ and $Q = GCD\ (a, b) == 1$. Assuming P we immediately get Q so what we need is to prove $\neg\ Q$, that is $GCD\ (a, b) \neq 1$. We can use the equations $b * r == a$ and $r\hat{}2 == 2$. Squaring the first equation and using the second we get $b\hat{}2 * 2 == a\hat{}2$. Thus $a\hat{}2$ is even, which means that a is even, thus $a == 2 * c$ for some c. But then $b\hat{}2 * 2 == a\hat{}2 == 4 * c\hat{}2$ which means that $b\hat{}2 == 2 * c\hat{}2$. By the same reasoning again we have that also b is even. But then $GCD\ (a, b) \geqslant 2$ which implies $\neg\ Q$.

To sum up: by assuming P we can prove both Q and $\neg\ Q$. Thus, by contradiction $\neg\ P$ must hold.

2.6 Proof by cases

As another example, let's prove that there are two irrational numbers a and b such that $a\hat{}b$ is rational.

$$S = \exists\ p.\ \ \exists\ q.\ \ \neg\ (R\ p)\ \&\ \neg\ (R\ q)\ \&\ R\ (p\hat{}q)$$

We know from above that $r = \sqrt{2}$ is irrational, so as a first attempt we could set $p = q = r$. Then we have satisfied two of the three clauses ($\neg\ (R\ p)$ and $\neg\ (R\ q)$). What about the third clause: is $x = p\hat{}q == r\hat{}r$ rational? We can reason about two possible cases, one of which has to hold: $R\ x$ or $\neg\ (R\ x)$.

Case 1: $R\ x$ holds. Then we have a proof of S with $p = q = r = \sqrt{2}$.

Case 2: $\neg\ (R\ x)$ holds. Then we have another irrational number x to play with. Let's try $p = x$ and $q = r$. Then $p\hat{}q == x\hat{}r == (r\hat{}r)\hat{}r == r\hat{}(r * r) == r\hat{}2 == 2$ which is clearly rational. Thus, also in this case we have a proof of S, but now with $p = r\hat{}r$ and $q = r$.

To sum up: yes, there are irrational numbers such that their power is rational. We can prove the existence without knowing what numbers p and q actually are!

2.7 Functions as proofs

To prove a formula $P \Rightarrow Q$ we assume a proof $p : P$ and derive a proof $q : Q$. Such a proof can be expressed as $(\lambda p \to q) : (P \Rightarrow Q)$: a proof of an implication is a function from proofs to proofs.

As we saw earlier, a similar rule holds for the "forall" quantifier: a function f from terms t to proofs of $P\ t$ is a proof of $\forall\ x.\ \ P\ x$.

A very common kind of formula is "typed quantification": if a type (a set) S of terms can be decribed as those that satisfy the unary predicate T we can introduce the short-hand notation

$$\forall\ x : T.\ \ P\ x = \forall\ x.\ \ T\ x \Rightarrow P\ x$$

A proof of this is a two-argument function p which takes a term and a proof to a proof.

In pseudo-Haskell we can express the implication laws as follows:

$$\begin{aligned} &\mathit{impIntro} : (A \to B) \to (A \Rightarrow B) \\ &\mathit{impElim} : (A \Rightarrow B) \to (A \to B) \end{aligned}$$

It should come as no surprise that this "API" can be implemented by $(\Rightarrow) = (\to)$, which means that both *impIntro* and *impElim* can be implemented as *id*.

Similarly we can express the universal quantification laws as:

$$\begin{aligned} &\forall\text{-}\mathit{Intro} : ((a : \mathit{Term}) \to P\ a) \to \forall\ x.\ \ P\ x \\ &\forall\text{-}\mathit{Elim} : \forall\ x.\ \ P\ x \to ((a : \mathit{Term}) \to P\ a) \end{aligned}$$

To actually implement this we need a *dependent* function type, which Haskell does not provide. But we can still use it as a tool for understanding and working with logic formulas and mathematical proofs.

Haskell supports limited forms of dependent types and more is coming every year but for proper dependently typed programming I recommend the language Agda.

2.8 Proofs for *And* and *Or*

TODO: textify

$$\begin{aligned} &\mathit{andIntro} \ \ : P \to Q \to P\ \&\ Q \\ &\mathit{andElimL} : P\ \&\ Q \to P \\ &\mathit{andElimR} : P\ \&\ Q \to Q \end{aligned}$$

If we see these introduction and elimination rules as an API, what would be a resonable implementation of the datatype $P\ \&\ Q$? A type of pairs! Then we see that the corresponding Haskell functions would be

```
pair :: p -> q -> (p, q)   -- andIntro
fst  :: (p, q) -> p        -- andElimL
snd  :: (p, q) -> q        -- andElimR
```

$$\begin{aligned} &\mathit{orIntroL} : P \to P \mid Q \\ &\mathit{orIntroR} : Q \to P \mid Q \\ &\mathit{orElim} \ \ : (P \Rightarrow R) \to (Q \Rightarrow R) \to ((P \mid Q) \Rightarrow R) \end{aligned}$$

Here the implementation type can be a labelled sum type, also called disjoint union and in Haskell: *Either*.

$$\textbf{data}\ \mathit{Either}\ p\ q = \mathit{Left}\ p \mid \mathit{Right}\ q$$
$$\quad \text{-- } \mathit{Left} \text{ is } \mathit{orIntroL}, \mathit{Right} \text{ is } \mathit{orIntroR}$$
$$\mathit{either} :: (p \to r) \to (q \to r) \to \mathit{Either}\ p\ q \to r$$
$$\mathit{either}\ l\ r\ (\mathit{Left}\ x) \quad = l\ x$$
$$\mathit{either}\ l\ r\ (\mathit{Right}\ y) = r\ y$$

2.9 Case study: there is always another prime

As an example of combining forall, exists and implication let us turn to one statement of the fact that there are infiniely many primes. If we assume we have a unary predicate expressing that number is prime and a binary (infix) predicate ordering the natural numbers we can define a formula *IP* for "Infinite many Primes" as follows:

$$IP = \forall\ n.\ \ \mathit{Prime}\ n \Rightarrow \exists\ m.\ \ \mathit{Prime}\ m\ \&\ m > n$$

Combined with the fact that there is at least one prime (like 2) we can repeatadly refer to this statement to produce a never-ending stream of primes.

To prove this formula we first translate from logic to programs as described above. We can translate step by step, starting from the top level. The forall-quantifier translates to a (dependent) function type $(n\!:\!\mathit{Term}) \to$ and the implication to a normal function type $\mathit{Prime}\ n \to$. The exists-quantifier translates to a (dependent) pair type $((m\!:\!\mathit{Term}), ...)$ and finally the & translates into a pair type. Putting all this together we get a type signature for any *proof* of the theorem:

$$\mathit{proof} : (n : \mathit{Term}) \to \mathit{Prime}\ n \to ((m : \mathit{Term}), (\mathit{Prime}\ m, m > n))$$

Now we can start filling in the definition of *proof* as a two-argument function returning a nested pair:

$$\mathit{proof}\ n\ np = (m, (pm, gt))$$
$$\quad \textbf{where}\ m' = 1 + \mathit{factorial}\ n$$
$$\qquad\qquad m = \{\text{- some non-trivial prime factor of } m' \text{ -}\}$$
$$\qquad\qquad pm = \{\text{- a proof that } m \text{ is prime -}\}$$
$$\qquad\qquad gt = \{\text{- a proof that } m > n \text{ -}\}$$

The proof *pm* is the core of the theorem. First, we note that for any $2 \leqslant p \leqslant n$ we have $\mathit{mod}\ m'\ p == \mathit{mod}\ (1 + m'!)\ p == \mathit{mod}\ 1\ p + \mathit{mod}\ (m'!)\ p == 1 + 0 == 1$. Thus m' is not divisible by any number from 2 to n. But is it a prime? If m' is prime then $m = m'$ and the proof is done (because $1 + n! \geqslant 1 + n > n$). Otherwise, let m be a prime factor of m' (thus $m' = m * q,\ q > 1$). Then $1 == \mathit{mod}\ m'\ p == (\mathit{mod}\ m\ p) * (\mathit{mod}\ q\ p)$ which means that neither m nor q are divisible by p (otherwise the product would be zero). Thus they must both be $> n$. QED.

Note that the proof can be used to define a somewhat useful function which takes any prime number to some larger prime number. We can compute a few example values:

$2 \mapsto\ \ 3$ (1+2!)
$3 \mapsto\ \ 7$ (1+3!)
$5 \mapsto 11$ (1+5! = 121 = 11*11)
$7 \mapsto 71$ ...

2.10 Existantial quantification as a pair type

We mentioned before that existential quantification can be seen as as a "big *Or*" of a family of formulas $P\ a$ for all terms a. This means that to prove the quantification, we only need exhibit one witness and one proof for that member of the family.

$$\exists\text{-}Intro : (a : Term) \to P\ a \to \exists\ x.\ \ P\ x$$

For binary *Or* the "family" only had two members, one labelled L for *Left* and one R for *Right*, and we used one introduction rule for each. Here, for the generalisation of *Or*, we have unified the two rules into one with an added parameter a corresponding to the label which indicates the family member.

In the other direction, if we look at the binary elimination rule, we we see the need for two arguments to be sure of how to prove the implication for any family member of the binary *Or*.

$$orElim : (P \Rightarrow R) \to (Q \Rightarrow R) \to ((P \mid Q) \Rightarrow R)$$

The generalisation unifies these two to one family of arguments. If we can prove R for each member of the family, we can be sure to prove R when we encounter some family member:

$$\exists\text{-}Elim : ((a : Term) \to P\ a \Rightarrow R) \to \exists\ x.\ \ P\ x \Rightarrow R$$

The datatype corresponding to $\exists\ x.\ \ P\ x$ is a pair of a witness a and a proof of $P\ a$. We sometimes write this type $(a : Term, P\ a)$.

2.11 Basic concepts of calculus

Now we have built up quite a bit of machinery to express logic formulas and proofs. It is time time to apply it to some concepts in calculus. We start we the concept of "limit point" which is used in the formulation of different properties of limits of functions.

TODO: Perhaps start with the "expression" $lim_{x \to x_0} f(x)$ and explain that not all x_0 make sense, etc. [For context and motivation.]

TODO: Or talk a bit about open and closed sets. (Open set = every point is internal = there is some wiggle-room around each point in the set. Closed set contains all its limit points.)

Limit point *Definition* (adapted from Rudin [1964], page 28): Let X be a subset of $\mathbb{R}$. A point $p \in \mathbb{R}$ is a limit point of X iff for every $\epsilon > 0$, there exists $q \in X$ such that $q \neq p$ and $|q - p| < \epsilon$.

$$Limp : \mathbb{R} \to \mathscr{P}\ \mathbb{R} \to Prop$$
$$Limp\ p\ X = \forall \epsilon > 0.\ \exists\ q \in X - \{\, p \,\}.\ |q - p| < \epsilon$$

Notice that q depends on ϵ. Thus by introducing a function *getq* we can move the $\exists$ out.

$$\textbf{type}\ Q = \mathbb{R}_{>0} \to (X - \{\, p \,\})$$
$$Limp\ p\ X = \exists getq : Q.\ \ \forall\, \epsilon > 0.\ \ \mid getq\ \epsilon - p \mid < \epsilon$$

Next: introduce the "disk function" *Di*.

TODO: perhaps rename *Di* to *N* for "neighbourhood" (or something based on "open ball").

$$Di : \mathbb{R} \to \mathbb{R}_{>0} \to \mathscr{P}\ \mathbb{R}$$
$$Di\ c\ r = \{\, x \mid |x - c| < r \,\}$$

Then we get

$$Limp\ p\ X = \exists getq : Q.\ \ \forall\, \epsilon > 0.\ getq\ \epsilon \in Di\ p\ \epsilon$$

Example 1: Is $p = 1$ a limit point of $X = \{1\}$? No! $X - \{\, p \,\} = \{\,\}$ (there is no $q \neq p$ in X), thus there cannot exist a function *getq* because it would have to return elements in the empty set!

Example 2: Is $p = 1$ a limit point of the open interval $X = (0, 1)$? First note that $p \notin X$, but it is "very close" to X. A proof needs a function *getq* which from any ϵ computes a point $q = getq\ \epsilon$ which is in both X and $Di\ 1\ \epsilon$. We need a point q which is in X and *closer* than ϵ from 1 We can try with $q = 1 - \epsilon\ /\ 2$ because $|1 - (1 - \epsilon\ /\ 2)| = |\epsilon\ /\ 2| = \epsilon\ /\ 2 < \epsilon$ which means $q \in Di\ 1\ \epsilon$. We also see that $q \neq 1$ because $\epsilon > 0$. The only remaining thing to check is that $q \in X$. This is true for sufficiently small ϵ but the function *getq* must work for all positive reals. We can use any value in X (for example $17\ /\ 38$) for ϵ which are "too big" ($\epsilon \geqslant 2$). Thus our function can be

$$\begin{array}{l} getq\ \epsilon \mid \epsilon < 2 \qquad = 1 - \epsilon\ /\ 2 \\ \qquad\quad \mid otherwise = 17\ /\ 38 \end{array}$$

A slight variation which is often useful would be to use *max* to define $getq\ \epsilon = max\ (17/38, 1 - \epsilon/2)$. Similarly, we can show that any internal point (like $1\ /\ 2$) is a limit point.

Example 3: limit of an infinite discrete set X

$$X = \{1\ /\ n \mid n \in \mathbb{N}_{>0}\}$$

Show that 0 is a limit point of X. Note (as above) that $0 \notin X$.

We want to prove $Limp\ 0\ X$ which is the same as $\exists getq : Q.\ \forall\, \epsilon > 0.\ getq\ \epsilon \in Di\ 0\ \epsilon$. Thus, we need a function *getq* which takes any $\epsilon > 0$ to an element of $X - \{0\} = X$ which is less than ϵ away from 0. Or, equivalently, we need a function $getn : \mathbb{R}_{>0} \to \mathbb{N}_{>0}$ such that $1\ /\ n < \epsilon$. Thus, we need to find an n such that $1\ /\ \epsilon < n$. If $1\ /\ \epsilon$ would be an integer we could use the next integer $(1 + 1\ /\ \epsilon)$, so the only step remaining is to round up:

$$\begin{array}{l} getq\ \epsilon = 1\ /\ getn\ \epsilon \\ getn\ \epsilon = 1 + ceiling\ (1\ /\ \epsilon) \end{array}$$

Exercise: prove that 0 is the *only* limit point of X.

Proposition: If X is finite, then it has no limit points.

$$\forall\, p \in \mathbb{R}.\ \neg\ (Limp\ p\ X)$$

This is a good excercise in quantifier negation!

$$\begin{array}{ll} & \neg\ (Limp\ p\ X) \\ = & \{\text{- Def. of } Limp \text{ -}\} \\ & \neg\ (\exists getq : Q.\ \forall\, \epsilon > 0.\ getq\ \epsilon \in Di\ p\ \epsilon) \\ = & \{\text{- Negation of existential -}\} \\ & \forall\, getq : Q.\ \neg\ (\forall \epsilon > 0.\ getq\ \epsilon \in Di\ p\ \epsilon) \\ = & \{\text{- Negation of universal -}\} \\ & \forall\, getq : Q.\ \exists\, \epsilon > 0.\ \neg\ (getq\ \epsilon \in Di\ p\ \epsilon) \\ = & \{\text{- Simplification -}\} \\ & \forall\, getq : Q.\ \exists\, \epsilon > 0.\ |getq\ \epsilon - p| \geqslant \epsilon \end{array}$$

Thus, using the "functional interpretation" of this type we see that a proof needs a function *noLim*

$$noLim : (getq : Q) \to \mathbb{R}_{>0}$$

such that $\textbf{let}\ \epsilon = noLim\ getq\ \textbf{in}\ |getq\ \epsilon - p| \geqslant \epsilon$.

Note that *noLim* is a *higher-order* function: it takes a function *getq* as an argument. How can we analyse this function to find a suitable ϵ? The key here is that the range of *getq* is $X - \{p\}$ which is a finite set (not containing p). Thus we can enumerate all the possible results in a list $xs = [x_1, x_2, \ldots x_n]$, and measure their distances to p: $ds = map\ (\lambda x \to |x - p|)\ xs$. Now, if we let $\epsilon = minimum\ ds$ we can be certain that $|getq\ \epsilon - p| \geqslant \epsilon$ just as required (and $\epsilon \neq 0$ because $p \notin xs$).

Exercise: If $Limp\ p\ X$ we now know that X is infinite. Show how to construct an infinite sequence $a : \mathbb{N} \to \mathbb{R}$ of points in $X - \{p\}$ which gets arbitrarily close to p. Note that this construction can be seen as a proof of $Limp\ p\ X \Rightarrow Infinite\ X$.

The limit of a sequence TODO: transcribe the 2016 notes

Now we can move from limit points to the more familiar limit of a sequence. A sequence a is a function from $\mathbb{N}$ to $\mathbb{R}$ and we define a binary infix predicate *haslim* using a helper predicate P:

$$a \; \mathit{haslim} \; L = \forall \epsilon > 0.\; P \; a \; L \; \epsilon$$
$$P \; a \; \epsilon \; L = \exists N : \mathbb{Z}.\; \forall \, n \geqslant N.\; |a_n - L| < \epsilon$$

TODO: perhaps swap the argument order in the definition of *Limp* to make it fit better with *haslim*.

Exercise: prove that $(a_1 \; \mathit{haslim} \; L_1) \,\&\, (a_2 \; \mathit{haslim} \; L_2)$ implies $(a_1 + a_2) \; \mathit{haslim} \; (L_1 + L_2)$.

2.12 Exercises

1. Build a proof for $((A \,\&\, B) -=> (B \,\&\, A))$ using *andIntro*, *andElimL*, and *andElimR*. TODO: add solution, step by step: *andIntro*. 1. $2 : B \;\&\; A$ **where**...;

   TODO: add several more exercises

2.13 Questions and answers from the exercise sessions week 2

Variables, *Env* and *lookup* This was a frequently source of confusion already the first week so there is already a question + answers earlier in this text. But here is an additional example to help clarify the matter.

$$\textbf{data} \; \mathit{Rat} \; v = \mathit{RV} \; v \mid \mathit{FromI} \; \mathit{Integer} \mid \mathit{RPlus} \; (\mathit{Rat} \; v) \; (\mathit{Rat} \; v) \mid \mathit{RDiv} \; (\mathit{Rat} \; v) \; (\mathit{Rat} \; v)$$
$$\quad \textbf{deriving} \; (\mathit{Eq}, \mathit{Show})$$
$$\textbf{newtype} \; \mathit{RatSem} = \mathit{RSem} \; (\mathit{Integer}, \mathit{Integer})$$

We have a type $\mathit{Rat} \; v$ for the syntax trees of rational number expressions and a type *RatSem* for the semantics of those rational number expressions as pairs of integers. The constructor $\mathit{RV} :: v \to \mathit{Rat} \; v$ is used to embed variables with names of type v in $\mathit{Rat} \; v$. We could use *String* instead of v but with a type parameter v we get more flexibility at the same time as we get better feedback from the type checker. To evaluate some $e :: \mathit{Rat} \; v$ we need to know how to evaluate the variables we encounter. What does "evaluate" mean for a variable? Well, it just means that we must be able to translate a variable name (of type v) to a semantic value (a rational number in this case). To "translate a name to a value" we can use a function (of type $v \to \mathit{RatSem}$) so we can give the following implementation of the evaluator:

$$\begin{aligned}
&\mathit{evalRat1} :: (v \to \mathit{RatSem}) \to (\mathit{Rat} \; v \to \mathit{RatSem}) \\
&\mathit{evalRat1} \; \mathit{ev} \; (\mathit{RV} \; v) \quad\;\; = \mathit{ev} \; v \\
&\mathit{evalRat1} \; \mathit{ev} \; (\mathit{FromI} \; i) \;\; = \mathit{fromISem} \; i \\
&\mathit{evalRat1} \; \mathit{ev} \; (\mathit{RPlus} \; l \; r) = \mathit{plusSem} \; (\mathit{evalRat1} \; \mathit{ev} \; l) \; (\mathit{evalRat1} \; \mathit{ev} \; r) \\
&\mathit{evalRat1} \; \mathit{ev} \; (\mathit{RDiv} \; l \; r) \;\; = \mathit{divSem} \;\; (\mathit{evalRat1} \; \mathit{ev} \; l) \; (\mathit{evalRat1} \; \mathit{ev} \; r)
\end{aligned}$$

Notice that we simply added a parameter *ev* for "evaluate variable" to the evaluator. The rest of the definition follows a common pattern: recursively translate each subexpression and apply the corresponding semantic operation to combine the results: *RPlus* is replaced by *plusSem*, etc.

$$\begin{aligned}
&\mathit{fromISem} :: \mathit{Integer} \to \mathit{RatSem} \\
&\mathit{fromISem} \; i = \mathit{RSem} \; (i, 1) \\
&\mathit{plusSem} :: \mathit{RatSem} \to \mathit{RatSem} \to \mathit{RatSem} \\
&\mathit{plusSem} = \mathit{undefined} \quad \text{-- TODO: exercise}
\end{aligned}$$

```
  -- Division of rational numbers
divSem :: RatSem → RatSem → RatSem
divSem (RSem (a, b)) (RSem (c, d)) = RSem (a * d, b * c)
```

Often the first argument *ev* to the eval function is constructed from a list of pairs:

```
type Env v s = [(v, s)]
envToFun :: (Show v, Eq v) ⇒ Env v s → (v → s)
envToFun [] v = error ("envToFun: variable " ++ show v ++ " not found")
envToFun ((w, s) : env) v
  | w == v    = s
  | otherwise = envToFun env v
```

Thus, *Env v s* can be seen as an implementation of a "lookup table". It could also be implemented using hash tables or binary search trees, but efficiency is not the point here. Finally, with *envToFun* in our hands we can implement a second version of the evaluator:

```
evalRat2 :: (Show v, Eq v) ⇒ (Env v RatSem) → (Rat v → RatSem)
evalRat2 env e = evalRat1 (envToFun env) e
```

The law of the excluded middle Many had problems with implementing the "law of the excluded middle" in the exercises and it is indeed a tricky property to prove. They key to implementing it lies in double negation and as that is encoded with higher order functions it gets a bit hairy.

TODO[Daniel]: more explanation

SET and PRED Several groups have had trouble grasping the difference between *SET* and *PRED*. This is understandable, beacuse we have so far in the lectures mostly talked about term syntax + semantics, and not so much about predicate syntax and semantics. The one example of terms + predicates covered in the lectures is Predicate Logic and I never actually showed how eval (for the expressions) and check (for the predicates) is implemented.

As an example we can we take our terms to be the rational number expressions defined above and define a type of predicates over those terms:

```
type Term v = Rat v
data RPred v = Equal     (Term v) (Term v)
             | LessThan  (Term v) (Term v)
             | Positive  (Term v)
             | AND (RPred v) (RPred v)
             | NOT (RPred v)
  deriving (Eq, Show)
```

Note that the first three constructors, *Eq*, *LessThan*, and *Positive*, describe predicates or relations between terms (which can contain term variables) while the two last constructors, *AND* and *NOT*, just combine such relations together. (Terminology: I often mix the words "predicate" and "relation".)

We have already defined the evaluator for the *Term v* type but we need to add a corresponding "evaluator" (called *check*) for the *RPred v* type. Given values for all term variables the predicate checker should just determine if the predicate is true or false.

```
checkRP :: (Eq v, Show v) ⇒ Env v RatSem → RPred v → Bool
checkRP env (Equal     t1 t2) = eqSem          (evalRat2 env t1) (evalRat2 env t2)
```

$$
\begin{aligned}
&\mathit{checkRP}\ \mathit{env}\ (\mathit{LessThan}\ t_1\ t_2) = \mathit{lessThanSem}\ (\mathit{evalRat2}\ \mathit{env}\ t_1)\ (\mathit{evalRat2}\ \mathit{env}\ t_2)\\
&\mathit{checkRP}\ \mathit{env}\ (\mathit{Positive}\quad t_1) \quad = \mathit{positiveSem}\quad (\mathit{evalRat2}\ \mathit{env}\ t_1)\\
&\mathit{checkRP}\ \mathit{env}\ (\mathit{AND}\ p\ q) = (\mathit{checkRP}\ \mathit{env}\ p) \wedge (\mathit{checkRP}\ \mathit{env}\ q)\\
&\mathit{checkRP}\ \mathit{env}\ (\mathit{NOT}\ p) \quad = \neg\ (\mathit{checkRP}\ \mathit{env}\ p)
\end{aligned}
$$

Given this recursive definition of *checkRP*, the semantic functions *eqSem*, *lessThanSem*, and *positiveSem* can be defined by just working with the rational number representation:

```
eqSem        :: RatSem → RatSem → Bool
lessThanSem  :: RatSem → RatSem → Bool
positiveSem  :: RatSem → Bool
eqSem        = error "TODO"
lessThanSem  = error "TODO"
positiveSem  = error "TODO"
```

2.14 More general code for first order languages

"överkurs"

It is possible to make one generic implementation which can be specialised to any first order language.

TODO: add explanatory text

- *Term* = Syntactic terms
- n = names (of atomic terms)
- f = function names
- v = variable names
- *WFF* = Well Formed Formulas
- p = predicate names

```
data Term n f v = N n | F f [Term n f v] | V v
  deriving Show
data WFF n f v p =
    P p    [Term n f v]
  | Equal (Term n f v)    (Term n f v)
  | And   (WFF n f v p) (WFF n f v p)
  | Or    (WFF n f v p) (WFF n f v p)
  | Equiv (WFF n f v p) (WFF n f v p)
  | Impl  (WFF n f v p) (WFF n f v p)
  | Not   (WFF n f v p)
  | FORALL v (WFF n f v p)
  | EXISTS  v (WFF n f v p)
  deriving Show
```

3 Week 3: Types in Mathematics

```
{-# LANGUAGE FlexibleInstances #-}
module DSLsofMath.W03 where
```

3.1 Types in mathematics

Types are sometimes mentioned explicitly in mathematical texts:

- $x \in \mathbb{R}$
- $\sqrt{\_} : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$
- $(\_)^2 : \mathbb{R} \to \mathbb{R}$ or, alternatively but *not* equivalently
- $(\_)^2 : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$

The types of "higher-order" operators are usually not given explicitly:

- $lim : (\mathbb{N} \to \mathbb{R}) \to \mathbb{R}$ for $lim_{n\to\infty}\{a_n\}$
- $d/dt : (\mathbb{R} \to \mathbb{R}) \to \mathbb{R} \to \mathbb{R}$
- sometimes, instead of df/dt one sees f' or $\dot{f}$ or $D\,f$
- $\partial f/\partial x_i : (\mathbb{R}^n \to \mathbb{R}) \to \mathbb{R}^n \to \mathbb{R}$
- we mostly see $\partial f/\partial x$, $\partial f/\partial y$, $\partial f/\partial z$ etc. when, in the context, the function f has been given a definition of the form $f(x, y, z) = \ldots$
- a better notation (by Landau) which doesn't rely on the names given to the arguments was popularised in Landau [1934] (English edition Landau [2001]): D_1 for the partial derivative with respect to x_1, etc.
- Exercise: for $f : \mathbb{R}^2 \to \mathbb{R}$ define D_1 and D_2 using only D.

3.2 Typing Mathematics: partial derivative

As as an example we will try to type the elements of a mathematical definition.

For example, on page 169 of Mac Lane [1986], we read

> [...] a function $z = f\ (x, y)$ for all points (x, y) in some open set U of the cartesian (x, y)-plane. [...] If one holds y fixed, the quantity z remains just a function of x; its derivative, when it exists, is called the *partial derivative* with respect to x. Thus at a point (x, y) in U this derivative for $h \neq 0$ is

$$\partial z/\partial x = f'_x(x, y) = \lim_{h\to 0}(f(x + h, y) - f(x, y))/h$$

What are the types of the elements involved? We have

```
U ⊆ ℝ × ℝ     -- cartesian plane
```

$$
\begin{aligned}
f &: U \to \mathbb{R} \\
z &: U \to \mathbb{R} \quad \text{-- but see below} \\
f_x &: U \to \mathbb{R}
\end{aligned}
$$

The x in the subscript of f' is *not* a real number, but a symbol (a *Char*).

The expression (x, y) has several occurrences. The first two denote variables of type U, the third is just a name ((x, y)-plane). The third denotes a variable of type U, it is bound by a universal quantifier

$$\forall\, (x, y) \in U$$

The variable h appears to be a non-zero real number, bound by a universal quantifier, but that is incorrect. In fact, h is used as a variable to construct the arguments of a function, whose limit is then taken at 0.

That function, which we can denote by φ has the type $\varphi : U \to (\mathbb{R} - \{0\}) \to \mathbb{R}$ and is defined by

$$\varphi\ (x, y)\ h = (f\ (x + h, y) - f\ (x, y))\ /\ h$$

The limit is then $lim\ (\varphi\ (x, y))\ 0$. Note that 0 is a limit point of $\mathbb{R} - \{0\}$, so the type of *lim* is the one we have discussed:

$$lim : (X \to \mathbb{R}) \to \{\, p \mid p \in \mathbb{R}, Limp\ p\ X \,\} \to \mathbb{R}$$

$z = f\ (x, y)$ probably does not mean that $z \in \mathbb{R}$, although the phrase "the quantity z" suggests this. A possible interpretation is that z is used to abbreviate the expression $f\ (x, y)$; thus, everywhere we can replace z with $f\ (x, y)$. In particular, $\partial z\ /\ \partial x$ becomes $\partial f\ (x, y)\ /\ \partial x$, which we can interpret as $\partial f\ /\ \partial x$ applied to (x, y) (remember that (x, y) is bound in the context by a universal quantifier). There is the added difficulty that, just like $_x$, the x in ∂x is not the x bound by the universal quantifier, but just a symbol.

3.3 Type inference and understanding: Lagrangian case study

From (Sussman and Wisdom 2013):

> A mechanical system is described by a Lagrangian function of the system state (time, coordinates, and velocities). A motion of the system is described by a path that gives the coordinates for each moment of time. A path is allowed if and only if it satisfies the Lagrange equations. Traditionally, the Lagrange equations are written
>
> $$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0$$
>
> What could this expression possibly mean?

To start answering the question, we start typing the elements involved:

1. $\partial L/\partial q$ suggests that L is a function of at least a pair of arguments:

   $$L : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}, n \geqslant 2$$

   This is consistent with the description: "Lagrangian function of the system state (time, coordinates, and velocities)". So we can take $n = 3$:

   $$L : \mathbb{R}^3 \to \mathbb{R}$$

2. $\partial L/\partial q$ suggests that q is the name of a real variable, one of the three arguments to L. In the context, which we do not have, we would expect to find somewhere the definition of the Lagrangian as

$$L\ (t, q, v) = ...$$

3. therefore, $\partial L/\partial q$ should also be a function of a triple of arguments:

$$\partial\ L\ /\ \partial q : \mathbb{R}^3 \to \mathbb{R}$$

It follows that the equation expresses a relation between *functions*, therefore the 0 on the right-hand side is *not* the real number 0, but rather the constant function 0:

$$\begin{aligned} &const\ 0 : \mathbb{R}^3 \to \mathbb{R} \\ &const\ 0\ (t, q, v) = 0 \end{aligned}$$

4. We now have a problem: d/dt can only be applied to functions of *one* real argument t, and the result is a function of one real argument:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} : \mathbb{R} \to \mathbb{R}$$

Since we subtract from this the function $\partial L/\partial q$, it follows that this, too, must be of type $\mathbb{R} \to \mathbb{R}$, contradiction.

5. The expression $\partial L/\partial \dot{q}$ appears to also be malformed. We would expect a variable name where we find $\dot{q}$, but $\dot{q}$ is the same as dq/dt, a function.

6. Looking back at the description above, we see that the only candidate for an application of d/dt is "a path that gives the coordinates for each moment of time". Thus, the path is a function of time, let us say

$$w : \mathbb{R} \to \mathbb{R}, \text{where } w(t) \text{ is a coordinate at time } t$$

We can now guess that the use of the plural form "equations" might have something to do with the use of "coordinates". In an n-dimensional space, a position is given by n coordinates. A path would be a function

$$w : \mathbb{R} \to \mathbb{R}^{\mathrm{n}}$$

which is equivalent to n functions of type $\mathbb{R} \to \mathbb{R}$. We would then have an equation for each of them. We will use $n = 1$ for the rest of this example.

7. The Lagrangian is a "function of the system state (time, coordinates, and velocities)". If we have a path, then the coordinates at any time are given by the path. The velocity is the derivative of the path, also fixed by the path:

$$\begin{aligned} &q : \mathbb{R} \to \mathbb{R} \\ &q\ t = w\ t \\ &\dot{q} : \mathbb{R} \to \mathbb{R} \\ &\dot{q}\ t = dw\ /\ dt \end{aligned}$$

The equations do not use a function $L : \mathbb{R}^3 \to \mathbb{R}$, but rather

$$L \circ expand\ w : \mathbb{R} \to \mathbb{R}$$

where the "combinator" *expand* is given by

$$\begin{aligned} &\mathit{expand} \quad : (\mathbb{R} \to \mathbb{R}) \to \mathbb{R} \to \mathbb{R}^3 \\ &\mathit{expand}\ w\ t = (t, w\ t, D\ w\ t) \end{aligned}$$

8. Similarly, using D_1, D_2, D_3 instead of $\partial L\,/\partial\, t$ etc., we have that, instead of $\partial L\,/\partial\, q$ what is meant is

   $$D_2\ L \circ \mathit{expand}\ w : \mathbb{R} \to \mathbb{R}$$

   and instead of $\partial L\,/\partial\, \dot{q}$

   $$D_3\ L \circ \mathit{expand}\ w : \mathbb{R} \to \mathbb{R}$$

   The equation becomes

   $$D\ (D_3\ L \circ \mathit{expand}\ w) - D_2\ L \circ \mathit{expand}\ w = 0$$

   a relation between functions of type $\mathbb{R} \to \mathbb{R}$. In particular, the right-hand 0 is the constant function

   $$\mathit{const}\ 0 : \mathbb{R} \to \mathbb{R}$$

3.4 Types in Mathematics (Part II)

3.4.1 Type classes

The kind of type inference we presented in the last lecture becomes automatic with experience in a domain, but is very useful in the beginning.

The "trick" of looking for an appropriate combinator with which to pre- or post-compose a function in order to makes types match is often useful. It is similar to the casts one does automatically in expressions such as $4 + 2.5$.

One way to understand such casts from the point of view of functional programming is via *type classes*. As a reminder, the reason $4 + 2.5$ works is because floating point values are members of the class *Num*, which includes the member function

$$\mathit{fromInteger} :: \mathit{Integer} \to a$$

which converts integers to the actual type a.

Type classes are related to mathematical structures which, in turn, are related to DSLs. The structuralist point of view in mathematics is that each mathematical domain has its own fundamental structures. Once these have been identified, one tries to push their study as far as possible *on their own terms*, i.e., without introducing other structures. For example, in group theory, one starts by exploring the consequences of just the group structure, before one introduces, say, an order structure and monotonicity.

The type classes of Haskell seem to have been introduced without relation to their mathematical counterparts, perhaps because of pragmatic considerations. For now, we examine the numerical type classes *Num*, *Fractional*, and *Floating*.

$$\begin{aligned} &\textbf{class}\ (\mathit{Eq}\ a, \mathit{Show}\ a) \Rightarrow \mathit{Num}\ a\ \textbf{where} \\ &\quad (+), (-), (*) :: a \to a \to a \end{aligned}$$

$$
\begin{aligned}
&\mathit{negate} && :: a \to a \\
&|\cdot|, \mathit{signum} && :: a \to a \\
&\mathit{fromInteger} && :: \mathit{Integer} \to a
\end{aligned}
$$

TODO: insert proper citation [Marlow , ed., Sect. 6.4]

This is taken from the Haskell documentation[2] but it appears that *Eq* and *Show* are not necessary, because there are meaningful instances of *Num* which don't support them:

$$
\begin{aligned}
&\textbf{instance } \mathit{Num}\ a \Rightarrow \mathit{Num}\ (x \to a)\ \textbf{where} \\
&\quad f + g && = \lambda x \to f\ x + g\ x \\
&\quad f - g && = \lambda x \to f\ x - g\ x \\
&\quad f * g && = \lambda x \to f\ x * g\ x \\
&\quad \mathit{negate}\ f && = \mathit{negate} \circ f \\
&\quad |f| && = |\cdot| \circ f \\
&\quad \mathit{signum}\ f && = \mathit{signum} \circ f \\
&\quad \mathit{fromInteger} && = \mathit{const} \circ \mathit{fromInteger}
\end{aligned}
$$

Next we have *Fractional* for when we also have division:

$$
\begin{aligned}
&\textbf{class } \mathit{Num}\ a \Rightarrow \mathit{Fractional}\ a\ \textbf{where} \\
&\quad (/) && :: a \to a \to a \\
&\quad \mathit{recip} && :: a \to a \\
&\quad \mathit{fromRational} && :: \mathit{Rational} \to a
\end{aligned}
$$

and *Floating* when we can implement the "standard" funtions from calculus:

$$
\begin{aligned}
&\textbf{class } \mathit{Fractional}\ a \Rightarrow \mathit{Floating}\ a\ \textbf{where} \\
&\quad \pi && :: a \\
&\quad \mathit{exp}, \mathit{log}, \sqrt{\cdot} && :: a \to a \\
&\quad (**), \mathit{logBase} && :: a \to a \to a \\
&\quad \mathit{sin}, \mathit{cos}, \mathit{tan} && :: a \to a \\
&\quad \mathit{asin}, \mathit{acos}, \mathit{atan} && :: a \to a \\
&\quad \mathit{sinh}, \mathit{cosh}, \mathit{tanh} && :: a \to a \\
&\quad \mathit{asinh}, \mathit{acosh}, \mathit{atanh} && :: a \to a
\end{aligned}
$$

We can instantiate these type classes for functions in the same way we did for *Num*:

$$
\begin{aligned}
&\textbf{instance } \mathit{Fractional}\ a \Rightarrow \mathit{Fractional}\ (x \to a)\ \textbf{where} \\
&\quad \mathit{recip}\ f && = \mathit{recip} \circ f \\
&\quad \mathit{fromRational} && = \mathit{const} \circ \mathit{fromRational}
\end{aligned}
$$

$$
\begin{aligned}
&\textbf{instance } \mathit{Floating}\ a \Rightarrow \mathit{Floating}\ (x \to a)\ \textbf{where} \\
&\quad \pi && = \mathit{const}\ \pi \\
&\quad \mathit{exp}\ f && = \mathit{exp} \circ f \\
&\quad f ** g && = \lambda x \to (f\ x) ** (g\ x) \\
&\qquad \text{-- and so on}
\end{aligned}
$$

Exercise: complete the instance declarations.

These type classes represent an abstract language of algebraic and standard operations, abstract in the sense that the exact nature of the elements involved is not important from the point of view of the type class, only from that of its implementation.

[2] Fig. 6.2 in section 6.4 of the Haskell 2010 report: `https://www.haskell.org/onlinereport/haskell2010/haskellch6.html`.

3.5 Computing derivatives

The "little language" of derivatives:

$$
\begin{array}{ll}
D\ (f+g) & = D\ f + D\ g \\
D\ (f*g) & = D\ f * g + f * D\ g \\
D\ (f \circ g)\ x & = D\ f\ (g\ x) * D\ g\ x \quad \text{-- the chain rule} \\
D\ (const\ a) & = const\ 0 \\
D\ id & = const\ 1 \\
D\ (\hat{}n)\ x & = (n-1) * (x\hat{}(n-1)) \\
D\ sin\ x & = cos\ x \\
D\ cos\ x & = -(sin\ x) \\
D\ exp\ x & = exp\ x
\end{array}
$$

and so on.

We observe that we can compute derivatives for any expressions made out of arithmetical functions, standard functions, and their compositions. In other words, the computation of derivatives is based on a DSL of expressions (representing functions in one variable):

$$
\begin{array}{rl}
expression ::= & const\ \mathbb{R} \\
| & id \\
| & expression + expression \\
| & expression * expression \\
| & exp\ expression \\
| & \ldots
\end{array}
$$

etc.

We can implement this in a datatype:

```
data FunExp = Const Double
            | Id
            | FunExp :+: FunExp
            | FunExp :*: FunExp
            | Exp FunExp
              -- and so on
            deriving Show
```

The intended meaning of elements of the *FunExp* type is functions:

```
eval :: FunExp     -> Double -> Double
eval   (Const α)   = const α
eval   Id          = id
eval   (e1 :+: e2) = eval e1 + eval e2   -- note the use of "lifted +"
eval   (e1 :*: e2) = eval e1 * eval e2   -- "lifted *"
eval   (Exp e1)    = exp (eval e1)       -- and "lifted exp"
  -- and so on
```

We can implement the derivative of such expressions using the rules of derivatives. We want to implement a function $derive :: FunExp \to FunExp$ which makes the following diagram commute:

$$
\begin{array}{ccc}
FunExp & \xrightarrow{eval} & Func \\
\downarrow{\scriptstyle derive} & & \downarrow{\scriptstyle D} \\
FunExp & \xrightarrow{eval} & Func
\end{array}
$$

In other words, for any expression e, we want

$$eval\ (derive\ e) = D\ (eval\ e)$$

For example, let us derive the *derive* function for $Exp\ e$:

$$\begin{aligned}
& eval\ (derive\ (Exp\ e)) \\
= & \{\text{- specification of } derive \text{ above -}\} \\
& D\ (eval\ (Exp\ e)) \\
= & \{\text{- def. } eval \text{ -}\} \\
& D\ (exp\ (eval\ e)) \\
= & \{\text{- def. } exp \text{ for functions -}\} \\
& D\ (exp \circ eval\ e) \\
= & \{\text{- chain rule -}\} \\
& (D\ exp \circ eval\ e) * D\ (eval\ e) \\
= & \{\text{- } D \text{ rule for } exp \text{ -}\} \\
& (exp \circ eval\ e) * D\ (eval\ e) \\
= & \{\text{- specification of } derive \text{ -}\} \\
& (exp \circ eval\ e) * (eval\ (derive\ e)) \\
= & \{\text{- def. of } eval \text{ for } Exp \text{ -}\} \\
& (eval\ (Exp\ e)) * (eval\ (derive\ e)) \\
= & \{\text{- def. of } eval \text{ for } :*: \text{ -}\} \\
& eval\ (Exp\ e\ :*:\ derive\ e)
\end{aligned}$$

Therefore, the specification is fulfilled by taking

$$derive\ (Exp\ e) = Exp\ e\ :*:\ derive\ e$$

Similarly, we obtain

$$\begin{aligned}
derive\ (Const\ \alpha) &= Const\ 0 \\
derive\ Id &= Const\ 1 \\
derive\ (e_1\ :+:\ e_2) &= derive\ e_1\ :+:\ derive\ e_2 \\
derive\ (e_1\ :*:\ e_2) &= (derive\ e_1\ :*:\ e_2)\ :+:\ (e_1\ :*:\ derive\ e_2) \\
derive\ (Exp\ e) &= Exp\ e\ :*:\ derive\ e
\end{aligned}$$

Exercise: complete the *FunExp* type and the *eval* and *derive* functions.

3.6 Shallow embeddings

The DSL of expressions, whose syntax is given by the type *FunExp*, turns out to be almost identical to the DSL defined via type classes in the first part of this lecture. The correspondence between them is given by the *eval* function.

The difference between the two implementations is that the first one separates more cleanly from the semantical one. For example, :+: *stands for* a function, while + *is* that function.

The second approach is called "shallow embedding" or "almost abstract syntax". It can be more economical, since it needs no *eval*. The question is: can we implement *derive* in the shallow embedding?

Note that the reason the shallow embedding is possible is that the *eval* function is a *fold*: first evaluate the sub-expressions of e, then put the evaluations together without reference to the sub-expressions. This is sometimes referred to as "compositionality".

We check whether the semantics of derivatives is compositional. The evaluation function for derivatives is

$$\begin{aligned}&eval' :: FunExp \to Double \to Double\\&eval' = eval \circ derive\end{aligned}$$

For example:

$$\begin{aligned}&\quad eval'\ (Exp\ e)\\&= \{\text{- def. } eval', \text{function composition -}\}\\&\quad eval\ (derive\ (Exp\ e))\\&= \{\text{- def. } derive \text{ for } Exp\text{ -}\}\\&\quad eval\ (Exp\ e :*: derive\ e)\\&= \{\text{- def. } eval \text{ for } :*:\text{ -}\}\\&\quad eval\ (Exp\ e) * eval\ (derive\ e)\\&= \{\text{- def. } eval \text{ for } Exp\text{ -}\}\\&\quad exp\ (eval\ e) * eval\ (derive\ e)\\&= \{\text{- def. } eval'\text{ -}\}\\&\quad exp\ (eval\ e) * eval'\ e\end{aligned}$$

and the first e doesn't go away. The semantics of derivatives is not compositional.

Or rather, *this* semantics is not compositional. It is quite clear that the derivatives cannot be evaluated without, at the same time, being able to evaluate the functions. So we can try to do both evaluations simultaneously:

$$\begin{aligned}&\textbf{type } FD\ a = (a \to a, a \to a)\\&evalD :: FunExp \to FD\ Double\\&evalD\quad e\qquad = (eval\ e, eval'\ e)\end{aligned}$$

Is *evalD* compositional?

We compute, for example:

$$\begin{aligned}&\quad evalD\ (Exp\ e)\\&= \{\text{- specification of } evalD\text{ -}\}\\&\quad (eval\ (Exp\ e), eval'\ (Exp\ e))\\&= \{\text{- def. } eval \text{ for } Exp \text{ and reusing the computation above -}\}\\&\quad (exp\ (eval\ e), exp\ (eval\ e) * eval'\ e)\\&= \{\text{- introduce names for subexpressions -}\}\\&\quad \textbf{let } f\ = eval\ e\\&\qquad\ f' = eval'\ e\\&\quad \textbf{in } (exp\ f, exp\ f * f')\\&= \{\text{- def. } evalD\text{ -}\}\\&\quad \textbf{let } (f, f') = evalD\ e\\&\quad \textbf{in } (exp\ f, exp\ f * f')\end{aligned}$$

This semantics *is* compositional. We can now define a shallow embedding for the computation of derivatives, using the numerical type classes.

$$
\begin{array}{l}
\textbf{instance}\ \mathit{Num}\ a \Rightarrow \mathit{Num}\ (a \to a, a \to a)\ \textbf{where} \\
\quad (f, f') + (g, g') = (f + g, f' \quad + g' \quad) \\
\quad (f, f') * (g, g') = (f\ *\ g, f' * g + f * g') \\
\quad \mathit{fromInteger}\ n\ = (\mathit{fromInteger}\ n, \mathit{const}\ 0)
\end{array}
$$

Exercise: implement the rest

4 Week 4: Compositional Semantics and Algebraic Structures

```
{-# LANGUAGE FlexibleInstances, GeneralizedNewtypeDeriving #-}
module DSLsofMath.W04 where
import Prelude hiding (Monoid)
```

4.1 Compositional semantics

4.1.1 A simpler example of a non-compositional function

Consider a very simple datatype of integer expressions:

```
data E = Add E E | Mul E E | Con Integer deriving Eq
e1, e2 :: E
e1 = Add (Con 1) (Mul (Con 2) (Con 3))
e2 = Mul (Add (Con 1) (Con 2)) (Con 3)
```

When working with expressions it is often useful to have a "pretty-printer" to convert the abstract syntax trees to strings like `"1+2*3"`.

```
pretty :: E → String
```

We can view *pretty* as an alternative *eval* function for a semantics using *String* as the semantic domain instead of the more natural *Integer*. We can implement *pretty* in the usual way as a "fold" over the syntax tree using one "semantic constructor" for each syntactic constructor:

```
pretty (Add x y) = prettyAdd (pretty x) (pretty y)
pretty (Mul x y) = prettyMul (pretty x) (pretty y)
pretty (Con c)   = prettyCon c

prettyAdd :: String → String → String
prettyMul :: String → String → String
prettyCon :: Integer → String
```

Now, if we try to implement the semantic constructors without thinking too much we would get the following:

```
prettyAdd sx sy = sx ++ "+" ++ sy
prettyMul sx sy = sx ++ "*" ++ sy
prettyCon i     = show i

p1, p2 :: String
p1 = pretty e1
p2 = pretty e2

trouble :: Bool
trouble = p1 == p2
```

Note that both e_1 and e_2 are different but they pretty-print to the same string. There are many ways to fix this, some more "pretty" than others, but the main problem is that some information is lost in the translation.

TODO(perhaps): Explain using three pretty printers for the three "contexts": at top level, inside *Add*, inside *Mul*, ... then combine them with the tupling transform just as with *evalD*. The result is the following:

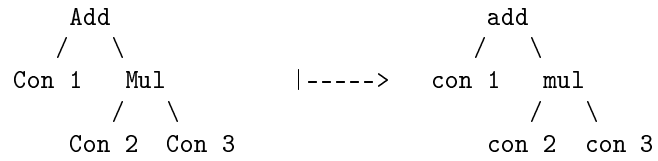
```
prTop :: E → String
prTop e = let (pTop, _, _) = prVersions e
          in pTop
prVersions = foldE prVerAdd prVerMul prVerCon
prVerAdd (xTop, xInA, xInM) (yTop, yInA, yInM) =
  let s = xInA ++ "+" ++ yInA      -- use InA because we are "in Add"
  in (s, paren s, paren s)         -- parens needed except at top level
prVerMul (xTop, xInA, xInM) (yTop, yInA, yInM) =
  let s = xInM ++ "*" ++ yInM      -- use InM because we are "in Mul"
  in (s, s, paren s)               -- parens only needed inside Mul
prVerCon i =
  let s = show i
  in (s, s, s)                     -- parens never needed
paren :: String → String
paren s = "(" ++ s ++ ")"
```

Exercise: Another way to make this example go through is to refine the semantic domain from *String* to *Precedence* → *String*. This can be seen as another variant of the result after the tupling transform: if *Precedence* is an n-element type then *Precedence* → *String* can be seen as an n-tuple. In our case a three-element *Precedence* would be enough.

4.1.2 Compositional semantics in general

In general, for a syntax *Syn*, and a possible semantics (a type *Sem* and an *eval* function of type *Syn* → *Sem*), we call the semantics *compositional* if we can implement *eval* as a fold. Informally a "fold" is a recursive function which replaces each abstract syntax constructor *Ci* of *Syn* with a "semantic constructor" *ci*.

TODO: Picture to illustrate



As an example we can define a general *foldE* for the integer expressions:

```
foldE :: (t → t → t) → (t → t → t) → (Integer → t) →
         E → t
foldE add mul con = rec
  where rec (Add x y) = add (rec x) (rec y)
        rec (Mul x y) = mul (rec x) (rec y)
        rec (Con i)   = con i
```

Notice that *foldE* has three function arguments corresponding to the three constructors of E. The "natural" evaluator to integers is then easy:

```
evalE1 :: E → Integer
evalE1 = foldE (+) (*) id
```

and with a minimal modification we can also make it work for other numeric types:

```
evalE2 :: Num a ⇒ E → a
evalE2 = foldE (+) (∗) fromInteger
```

Another thing worth noting is that if we replace each abstract syntax constructor with itself we get the identity function (a "deep copy"):

```
idE :: E → E
idE = foldE Add Mul Con
```

Finally, it is often useful to capture the semantic functions (the parameters to the fold) in a type class:

```
class IntExp t where
  add :: t → t → t
  mul :: t → t → t
  con :: Integer → t
```

In this way we can make "hide" the arguments to the fold:

```
foldIE :: IntExp t ⇒ E → t
foldIE = foldE add mul con
instance IntExp E where
  add = Add
  mul = Mul
  con = Con
instance IntExp Integer where
  add = (+)
  mul = (∗)
  con = id
idE′ :: E → E
idE′ = foldIE
evalE′ :: E → Integer
evalE′ = foldIE
```

4.1.3 Back to derivatives and evaluation

Review section 3.6 again with the definition of *eval′* being non-compositional (just like *pretty*) and *evalD* a more complex, but compositional, semantics.

4.2 Algebraic Structures and DSLs

In this lecture, we continue exploring the relationship between type classes, mathematical structures, and DSLs.

4.2.1 Algebras, homomorphisms

From Wikipedia:

In universal algebra, an algebra (or algebraic structure) is a set A together with a collection of operations on A.

Example:

$$\begin{array}{l}\textbf{class}\ \mathit{Monoid}\ a\ \textbf{where}\\ \quad \mathit{unit} :: a\\ \quad \mathit{op}\ \ :: a \to a \to a\end{array}$$

After the operations have been specified, the nature of the algebra can be further limited by axioms, which in universal algebra often take the form of identities, or *equational laws.*

Example: Monoid equations

$$\begin{array}{l}\forall\, x : a.\ (\mathit{unit}\ `\mathit{op}`\ x == x \wedge x\ `\mathit{op}`\ \mathit{unit} == x)\\ \forall\, x, y, z : a.\ (x\ `\mathit{op}`\ (y\ `\mathit{op}`\ z) == (x\ `\mathit{op}`\ y)\ `\mathit{op}`\ z)\end{array}$$

A homomorphism between two algebras A and B is a function $h : A \to B$ from the set A to the set B such that, for every operation fA of A and corresponding fB of B (of arity, say, n), $h\ (fA\ (x_1, ..., x_n)) = fB\ (h\ (x_1), ..., h\ (x_n))$.

Example: Monoid homomorphism

$$\begin{array}{l}h\ \mathit{unit} \quad\ \ = \mathit{unit}\\ h\ (x\ `\mathit{op}`\ y) = h\ x\ `\mathit{op}`\ h\ y\end{array}$$

$$\begin{array}{l}\textbf{newtype}\ \mathit{ANat} \quad\ \ = A\ \mathit{Int}\ \ \textbf{deriving}\ (\mathit{Show}, \mathit{Num}, \mathit{Eq})\\ \textbf{instance}\ \mathit{Monoid}\ \mathit{ANat}\ \textbf{where}\\ \quad \mathit{unit} \quad\ \ = A\ 0\\ \quad \mathit{op}\ (A\ m)\ (A\ n) \ = A\ (m + n)\\ \textbf{newtype}\ \mathit{MNat} \quad\ \ = M\ \mathit{Int}\ \textbf{deriving}\ (\mathit{Show}, \mathit{Num}, \mathit{Eq})\\ \textbf{instance}\ \mathit{Monoid}\ \mathit{MNat}\ \textbf{where}\\ \quad \mathit{unit} \quad\ \ = M\ 1\\ \quad \mathit{op}\ (M\ m)\ (M\ n) = M\ (m * n)\end{array}$$

Exercise: characterise the homomorphisms from ANat to MNat.

Solution:

Let $h : \mathit{ANat} \to \mathit{MNat}$ be a homomorphism. Then

$$\begin{array}{l}h\ 0 \quad\ \ = 1\\ h\ (x + y) = h\ x * h\ y\end{array}$$

For example $h\ (x + x) = h\ x * h\ x = (h\ x)\hat{}2$ which for $x = 1$ means that $h\ 2 = (h\ 1)\hat{}2$.

More generally, every n in ANat is equal to the sum of n ones: $1 + 1 + ... + 1$. Therefore

$$h\ n = (h\ 1)\hat{}n$$

Every choice of $h\ 1$ "induces a homomorphism". This means that the value of the function h is fully determined by its value for 1.

4.2.2 Homomorphism and compositional semantics

Last time, we saw that *eval* is compositional, while *eval′* is not. Another way of phrasing that is to say that *eval* is a homomorphism, while *eval′* is not. To see this, we need to make explicit the structure of *FunExp*:

```
instance Num FunExp where
  (+)         = (:+:)
  (*)         = (:*:)
  fromInteger = Const ∘ fromInteger
instance Fractional FunExp where
instance Floating FunExp where
  exp         = Exp
```

and so on.

Exercise: complete the type instances for *FunExp*.

For instance, we have

$$\begin{aligned} eval\ (e_1 :*: e_2) &= eval\ e_1 * eval\ e_2 \\ eval\ (Exp\ e) &= exp\ (eval\ e) \end{aligned}$$

These properties do not hold for *eval′*, but do hold for *evalD*.

The numerical classes in Haskell do not fully do justice to the structure of expressions, for example, they do not contain an identity operation, which is needed to translate *Id*, nor an embedding of doubles, etc. If they did, then we could have evaluated expressions more abstractly:

$$eval :: GoodClass\ a \Rightarrow FunExp \to a$$

where *GoodClass* gives exactly the structure we need for the translation.

Exercise: define *GoodClass* and instantiate *FunExp* and $Double \to Double$ as instances of it. Find another instance of *GoodClass*.

Therefore, we can always define a homomorphism from *FunExp* to *any* instance of *GoodClass*, in an essentially unique way. In the language of category theory, *FunExp* is an initial algebra.

Let us explore this in the simpler context of *Monoid*. The language of monoids is given by

```
type Var   = String
data MExpr = Unit | Op MExpr MExpr | V Var
```

Alternatively, we could have parametrised *MExpr* over the type of variables.

Just as in the case of FOL terms, we can evaluate an *MExpr* in a monoid instance if we are given a way of interpreting variables, also called an assignment:

$$evalM :: Monoid\ a \Rightarrow (Var \to a) \to (MExpr \to a)$$

Once given an $f :: Var \to a$, the homomorphism condition defines *evalM*:

$$\begin{aligned} evalM\ f\ Unit &= unit \\ evalM\ f\ (Op\ e_1\ e_2) &= op\ (evalM\ f\ e_1)\ (evalM\ f\ e_2) \\ evalM\ f\ (V\ x) &= f\ x \end{aligned}$$

(Observation: In *FunExp*, the role of variables was played by *Double*, and the role of the assignment by the identity.)

The following correspondence summarises the discussion so far:

Computer Science	Mathematics
DSL	structure (category, algebra, ...)
deep embedding, abstract syntax	initial algebra
shallow embedding	any other algebra
semantics	homomorphism from the initial algebra

The underlying theory of this table is a fascinating topic but mostly out of scope for the DSLsofMath course. See Category Theory and Functional Programming for a whole course around this (lecture notes are available on github).

4.2.3 Other homomorphisms

Last time, we defined a *Num* instance for functions with a *Num* codomain. If we have an element of the domain of such a function, we can use it to obtain a homomorphism from functions to their codomains:

$$Num\ a \Rightarrow x \to (x \to a) \to a$$

As suggested by the type, the homomorphism is just function application:

$$\begin{aligned} &apply :: a \to (a \to b) \to b \\ &apply\ a = \lambda f \to f\ a \end{aligned}$$

Indeed, writing $h = apply\ c$ for some fixed c, we have

$$\begin{aligned} & h\ (f + g) \\ =\ & \{\text{- def. } apply \text{ -}\} \\ & (f + g)\ c \\ =\ & \{\text{- def. } + \text{ for functions -}\} \\ & f\ c + g\ c \\ =\ & \{\text{- def. } apply \text{ -}\} \\ & h\ f + h\ g \end{aligned}$$

etc.

Can we do something similar for *FD*?

The elements of *FD* a are pairs of functions, so we can take

$$\begin{aligned} &apply :: a \to FD\ a\ \to (a, a) \\ &apply\quad c\quad (f, f') = (f\ c, f'\ c) \end{aligned}$$

We now have the domain of the homomorphism ($FD\ a$) and the homomorphism itself ($apply\ c$), but we are missing the structure on the codomain, which now consists of pairs (a, a). In fact, we can *compute* this structure from the homomorphism condition. For example:

$$\begin{aligned} & h\ ((f, f') * (g, g')) \\ =\ & \{\text{- def. } * \text{ for } FD\ a \text{ -}\} \end{aligned}$$

$$
\begin{aligned}
& h\ (f * g, f' * g + f * g') \\
= & \{\text{- def. } h = apply\ c \text{ -}\} \\
& ((f * g)\ c, (f' * g + f * g')\ c) \\
= & \{\text{- def. } * \text{ and } + \text{ for functions -}\} \\
& (f\ c * g\ c, f'\ c * g\ c + f\ c * g'\ c) \\
= & \{\text{- homomorphism condition from step 1 -}\} \\
& h\ (f, f') \circledast h\ (g, g') \\
= & \{\text{- def. } h = apply\ c \text{ -}\} \\
& (f\ c, f'\ c) \circledast (g\ c, g'\ c)
\end{aligned}
$$

The identity will hold if we take

$$(x, x') \circledast (y, y') = (x * y, x' * y + x * y')$$

Exercise: complete the instance declarations for $(Double, Double)$.

Note: As this computation goes through also for the other cases we can actually work with just pairs of values (at an implicit point $c :: a$) instead of pairs of functions. Thus we can redefine FD to be

type $FD\ a = (a, a)$

Hint: Something very similar can be used for Assignment 2.

4.2.4 Some helper functions

```
instance Num E where    -- Some abuse of notation (no proper negate, etc.)
  (+) = Add
  (*) = Mul
  fromInteger = Con
  negate = negateE
negateE (Con c) = Con (negate c)
negateE _ = error "negate: not supported"
```

5 Week 5: Polynomials and Power Series

```
{-# LANGUAGE TypeSynonymInstances #-}
module DSLsofMath.W05 where
```

5.1 Preliminaries

Last time, we defined a *Num* structure on pairs $(Double, Double)$ by requiring the operations to be compatible with the interpretation $(f\ a, f'\ a)$. For example

$$(x, x') \circledast (y, y') = (x * y, x' * y + x * y')$$

There is nothing in the "nature" of pairs of *Double* that forces this definition upon us. We chose it, because of the intended interpretation.

This multiplication is obviously not the one we need for *complex numbers*:

$$(x, x') *. (y, y') = (x * y - x' * y', x * y' + x' * y)$$

Again, there is nothing in the nature of pairs that foists this operation on us. In particular, it is, strictly speaking, incorrect to say that a complex number *is* a pair of real numbers. The correct interpretation is that a complex number can be *represented* by a pair of real numbers, provided we define the operations on these pairs in a suitable way.

The distinction between definition and representation is similar to the one between specification and implementation, and, in a certain sense, to the one between syntax and semantics. All these distinctions are frequently obscured, for example, because of prototyping (working with representations / implementations / concrete objects in order to find out what definition / specification / syntax is most adequate). They can also be context-dependent (one man's specification is another man's implementation). Insisting on the difference between definition and representation can also appear quite pedantic (as in the discussion of complex numbers above). In general though, it is a good idea to be aware of these distinctions, even if they are suppressed for reasons of brevity or style.

5.2 Polynomials

From Adams and Essex [2010], page 55:

> A **polynomial** is a function P whose value at x is
>
> $$Px = a_n x^n + a_{n-1} x^{n-1} + ... a_1 x + a_0$$
>
> where a_n, a_{n-1}, ..., a_1, and a_0, called the **coefficients** of the polymonial [original spelling], are constants and, if $n > 0$, then $a_n \neq 0$. The number n, the degree of the highest power of x in the polynomial, is called the **degree** of the polynomial. (The degree of the zero polynomial is not defined.)

This definition raises a number of questions, for example "what is the zero polynomial?".

The types of the elements involved in the definition appear to be

$P : \mathbb{R} \to \mathbb{R}$, $x \in \mathbb{R}$, a_0, ... $a_n \in \mathbb{R}$ with $a_n \neq 0$ if $n > 0$

The phrasing should be "whose value at *any* x is". The remark that the a_i are constants is probably meant to indicate that they do not depend on x, otherwise every function would be a polynomial. The zero polynomial is, according to this definition, the 'const 0' function. Thus, what is meant is

> A **polynomial** is a function $P : \mathbb{R} \to \mathbb{R}$ which is either constant zero, or there exist a_0, ..., $a_n \in \mathbb{R}$ with $a_n \neq 0$ such that, for any $x \in \mathbb{R}$
>
> $$Px = a_n x^n + a_{n-1} x^{n-1} + ...a_1 x + a_0$$

Obviously, given the coefficients a_i we can evaluate P at any given x. Assuming the coefficients are given as

$$as = [a_0, a_1, ..., a_n]$$

(we prefer counting up), then the evaluation function is written

```
eval :: [Real] -> Real -> Real
eval   []       x    = 0
eval   (a : as) x    = a + x * eval as x
```

Not every list is valid according to the definition. In particular, the empty list is not a valid list of coefficients, so we have a conceptual, if not empirical, type error in our evaluator.

The valid lists are those *finite* lists in the set

$$\{[0]\} \cup \{(a : as) \mid last\ (a : as) \neq 0\}$$

We cannot express the $last\ (a : as) \neq 0$ in Haskell, but we can express the condition that the list should not be empty:

```
data Poly a = Single a | Cons a (Poly a)
            deriving (Eq, Ord)
```

(TODO: show the version and motivation for using just $[a]$ as well. Basically, one can use [] as the syntax for the "zero polynomial" and $(c : cs)$ for all other.)

The relationship between $Poly\ a$ and $[a]$ is given by the following functions:

```
toList :: Poly a ->     [a]
toList (Single a)    = a : []
toList (Cons a as) = a : toList as

fromList :: [a] ->         Poly a
fromList (a : [])        = Single a
fromList (a0 : a1 : as) = Cons a0 (fromList (a1 : as))

instance Show a => Show (Poly a) where
  show = show . toList
```

Since we only use the arithmetical operations, we can generalise our evaluator:

```
evalPoly :: Num a => Poly a -> a -> a
evalPoly (Single a)   x = a
evalPoly (Cons a as) x = a + x * evalPoly as x
```

Since we have $Num\ a$, there is a Num structure on $a \to a$, and $evalPoly$ looks like a homomorphism. Question: is there a Num structure on $Poly\ a$, such that $evalPoly$ is a homomorphism?

For example, the homomorphism condition gives for $(+)$

$$evalPoly\ as + evalPoly\ bs = evalPoly\ (as + bs)$$

Both sides are functions, they are equal iff they are equal for every argument. For an arbitrary x

$$\begin{aligned}
&(evalPoly\ as + evalPoly\ bs)\ x = evalPoly\ (as + bs)\ x \\
\Leftrightarrow\ &\{\text{- + on functions is defined point-wise -}\} \\
&evalPoly\ as\ x + evalPoly\ bs\ x = evalPoly\ (as + bs)\ x
\end{aligned}$$

To proceed further, we need to consider the various cases in the definition of $evalPoly$. We give here the computation for the last case (where as has at least one $Cons$), using the traditional list notation (:) for brevity.

$$evalPoly\ (a : as)\ x + evalPoly\ (b : bs)\ x = evalPoly\ ((a : as) + (b : bs))\ x$$

For the left-hand side, we have:

$$\begin{aligned}
&evalPoly\ (a : as)\ x + evalPoly\ (b : bs)\ x \\
=\ &\{\text{- def. } evalPoly \text{ -}\} \\
&(a + x * evalPoly\ as\ x) + (b + x * eval\ bs\ x) \\
=\ &\{\text{- properties of +, valid in any ring -}\} \\
&(a + b) + x * (evalPoly\ as\ x + evalPoly\ bs\ x) \\
=\ &\{\text{- homomorphism condition -}\} \\
&(a + b) + x * (evalPoly\ (as + bs)\ x) \\
=\ &\{\text{- def. } evalPoly \text{ -}\} \\
&evalPoly\ ((a + b) : (as + bs))\ x
\end{aligned}$$

The homomorphism condition will hold for every x if we define

$$(a : as) + (b : bs) = (a + b) : (as + bs)$$

We leave the derivation of the other cases and operations as an exercise. Here, we just give the corresponding definitions.

```
instance Num a ⇒ Num (Poly a) where
  (+) = polyAdd
  (*) = polyMul
  negate = polyNeg
  fromInteger = Single ∘ fromInteger

polyAdd :: Num a ⇒ Poly a → Poly a → Poly a
polyAdd (Single a)  (Single b)  = Single (a + b)
polyAdd (Single a)  (Cons b bs) = Cons (a + b) bs
polyAdd (Cons a as) (Single b)  = Cons (a + b) as
polyAdd (Cons a as) (Cons b bs) = Cons (a + b) (polyAdd as bs)

polyMul :: Num a ⇒ Poly a → Poly a → Poly a
polyMul (Single a)  (Single b)  = Single (a * b)
polyMul (Single a)  (Cons b bs) = Cons (a * b) (polyMul (Single a) bs)
```

$$
\begin{aligned}
&\mathit{polyMul}\ (\mathit{Cons}\ a\ as)\ (\mathit{Single}\ b) &&= \mathit{Cons}\ (a * b)\ (\mathit{polyMul}\ as\ (\mathit{Single}\ b)) \\
&\mathit{polyMul}\ (\mathit{Cons}\ a\ as)\ (\mathit{Cons}\ b\ bs) &&= \mathit{Cons}\ (a * b)\ (\mathit{polyAdd}\ (\mathit{polyMul}\ as\ (\mathit{Cons}\ b\ bs)) \\
& && \qquad (\mathit{polyMul}\ (\mathit{Single}\ a)\ bs)) \\
&\mathit{polyNeg} :: \mathit{Num}\ a \Rightarrow \mathit{Poly}\ a \to \mathit{Poly}\ a \\
&\mathit{polyNeg} = \mathit{fmap}\ \mathit{negate}
\end{aligned}
$$

Therefore, we *can* define a ring structure (the mathematical counterpart of *Num*) on *Poly a*, and we have arrived at the canonical definition of polynomials, as found in any algebra book (see, for example, Rotman [2006] for a very readable text):

> Given a commutative ring A, the commutative ring given by the set *Poly A* together with the operations defined above is the ring of **polynomials** with coefficients in A.

The functions *evalPoly as* are known as *polynomial functions*.

Caveat: The canonical representation of polynomials in algebra does not use finite lists, but the equivalent

$$\mathit{Poly}'\ A = \{\, a : \mathbb{N} \to A \mid \{\text{- } a \text{ has only a finite number of non-zero values -}\}\ \}$$

Exercise: what are the ring operations on *Poly′ A*? For example, here is addition:

$$a + b = c \Leftrightarrow a\ n + b\ n = c\ n \quad \text{-- } \forall n : \mathbb{N}$$

Observations:

1. Polynomials are not, in general, isomorphic (in one-to-one correspondence) with polynomial functions. For any finite ring A, there is a finite number of functions $A \to A$, but there is a countable number of polynomials. That means that the same polynomial function on A will be the evaluation of many different polynomials.

   For example, consider the ring $\mathbb{Z}_2$ ($\{0, 1\}$ with addition and multiplication modulo 2). In this ring, we have

   $$\mathit{evalPoly}\ [0, 1, 1] = \mathit{const}\ 0 = \mathit{evalPoly}\ [0]\{\text{- in } \mathbb{Z}_2 \to \mathbb{Z}_2 \text{ -}\}$$

   but

   $$[0, 1, 1] \neq [0]\{\text{- in } \mathit{Poly}\ \mathbb{Z}_2 \text{ -}\}$$

   Therefore, it is not generally a good idea to confuse polynomials with polynomial functions.

2. In keeping with the DSL terminology, we can say that the polynomial functions are the semantics of the language of polynomials. We started with polynomial functions, we wrote the evaluation function and realised that we have the makings of a homomorphism. That suggested that we could create an adequate language for polynomial functions. Indeed, this turns out to be the case; in so doing, we have recreated an important mathematical achievement: the algebraic definition of polynomials.

   Let

   $$
   \begin{aligned}
   &x :: \mathit{Num}\ a \Rightarrow \mathit{Poly}\ a \\
   &x = \mathit{Cons}\ 0\ (\mathit{Single}\ 1)
   \end{aligned}
   $$

Then (again, using the list notation for brevity) for any polynomial $as = [a_0, a_1, ..., a_n]$ we have

$$as = a_0 + a_1 * x + a_2 * x\hat{}2 + ... + a_n * x\hat{}n$$

Exercise: check this.

This justifies the standard notation

$$as = \sum_{i=0}^{n} a_i * x^i$$

5.3 Polynomial degree as a homomorphism

TODO: textify black board notes

It is often the case that a certain function is *almost* a homomorphism and the domain or range *almost* a monoid. In the section on *eval* and *eval'* for *FunExp* we have seen "tupling" as one way to fix such a problem and here we will introduce another way.

The *degree* of a polynomial is a good candidate for being a homomorphism: if we multiply two polynomials we can normally add their degrees. If we try to check that $degree :: Poly\ a \to \mathbb{N}$ is the function underlying a monoid morphism we need to decide on the monoid structure to use for the source and for the target, and we need to check the homomorphism laws. We can use $unit = Single\ 1$ and $op = polyMul$ for the source monoid and we can try to use $unit = 0$ and $op = (+)$ for the target monoid. Then we need to check that

$$degree\ (Single\ 1) = 0$$
$$\forall\, x, y.\ degree\ (x\ `op`\ y) = degree\ x + degree\ y$$

The first law is no problem and for most polynomials the second law is also straighforward to prove (exercise: prove it) except for one special case: the zero polynomial.

Looking back at the definition from Adams and Essex [2010], page 55 it says that the degree of the zero polynomial is not defined. Let's see why that is the case and how we might "fix" it. Assume there is a z such that $degree\ 0 = z$ and that we have some polynomial p with $degree\ p = n$. Then we get

$$\begin{aligned}
& z \\
= & \{\text{- assumption -}\} \\
& degree\ 0 \\
= & \{\text{- simple calculation -}\} \\
& degree\ (0 * p) \\
= & \{\text{- homomorphism condition -}\} \\
& degree\ 0 + degree\ p \\
= & \{\text{- assumption -}\} \\
& z + n
\end{aligned}$$

Thus we need to find a z such that $z = z + n$ for all natural numbers n! At this stage we could either give up, or think out of the box. Intuitively we could try to use $z = -Infinity$, which would seem to satisfy the law but which is not a natural number. More formally what we need to do is to extend the monoid $(\mathbb{N}, 0, +)$ by one more element. In Haskell we can do that using the *Maybe* type constructor:

```
class Monoid a where
  unit :: a
```

$$
\begin{aligned}
&\quad op \quad :: a \to a \to a \\
&\textbf{instance } Monoid\ a \Rightarrow Monoid\ (Maybe\ a) \textbf{ where} \\
&\quad unit = Nothing \\
&\quad op \quad = opMaybe \\
&opMaybe\ Nothing \quad m \qquad = m \\
&opMaybe\ m \qquad Nothing \quad = m \\
&opMaybe\ (Just\ m_1)\ (Just\ m_2) = Just\ (op\ m_1\ m_2)
\end{aligned}
$$

We quote the Haskell prelude implementation:

> Lift a semigroup into *Maybe* forming a *Monoid* according to `http://en.wikipedia.org/wiki/Monoid`: "Any semigroup S may be turned into a monoid simply by adjoining an element e not in S and defining $e * e = e$ and $e * s = s = s * e$ for all $s \in S$." Since there is no *Semigroup* typeclass [..], we use *Monoid* instead.

Thus, to sum up, *degree* is a monoid homomorphism from $(Poly\ a, 1, *)$ to $(Maybe\ \mathbb{N}, Nothing, opMaybe)$.

TODO: check all the properties.

5.4 Power Series

Power series are obtained from polynomials by removing in $Poly'$ the restriction that there should be a *finite* number of non-zero coefficients; or, in, the case of *Poly*, by going from lists to streams.

$$PowerSeries'\ a = \{f : \mathbb{N} \to a\}$$

type *PowerSeries* $a = Poly\ a$ -- finite and infinite non-empty lists

The operations are still defined as before. If we consider only infinite lists, then only the equations which do not contain the patterns for singleton lists will apply.

Power series are usually denoted

$$\sum_{n=0}^{\infty} a_n * x^n$$

the interpretation of x being the same as before.

The evaluation of a power series represented by $a{:}\mathbb{N}{\to}A$ is defined, in case the necessary operations make sense on A, as a function

$$
\begin{aligned}
&eval\ a : A \to A \\
&eval\ a\ x = lim\ s \textbf{ where } s\ n = \textstyle\sum_{i=0}^{n} a_i * x^i
\end{aligned}
$$

Note that $eval\ a$ is, in general, a partial function (the limit might not exist).

We will consider, as is usual, only the case in which $A = \mathbb{R}$ or $A = \mathbb{C}$.

The term *formal* refers to the independence of the definition of power series from the ideas of convergence and evaluation. In particular, two power series represented by a and b, respectively, are equal only if $a = b$ (as functions). If $a \neq b$, then the power series are different, even if $eval\ a = eval\ b$.

Since we cannot in general compute limits, we can use an "approximative" *eval*, by evaluating the polynomial resulting from an initial segment of the power series.

$$
\begin{aligned}
&\mathit{eval} :: \mathit{Num}\ a \Rightarrow \mathit{Integer} \to \mathit{PowerSeries}\ a \to (a \to a)\\
&\mathit{eval}\ n\ \mathit{as}\ x = \mathit{evalPoly}\ (\mathit{takePoly}\ n\ \mathit{as})\ x\\
&\mathit{takePoly} :: \mathit{Integer} \to \mathit{PowerSeries}\ a \to \mathit{Poly}\ a\\
&\mathit{takePoly}\ n\ (\mathit{Single}\ a) \quad = \mathit{Single}\ a\\
&\mathit{takePoly}\ n\ (\mathit{Cons}\ a\ \mathit{as}) = \mathbf{if}\ n \leqslant 1\\
&\qquad\qquad\qquad\qquad\quad \mathbf{then}\ \mathit{Single}\ a\\
&\qquad\qquad\qquad\qquad\quad \mathbf{else}\ \ \mathit{Cons}\ a\ (\mathit{takePoly}\ (n-1)\ \mathit{as})
\end{aligned}
$$

Note that *eval* n is not a homomorphism: for example $\mathit{eval}\ 2\ (x * x)\ 1 = 0 \neq 1 = 1 * 1 = (\mathit{eval}\ 2\ x\ 1) * (\mathit{eval}\ 2\ x\ 1)$.

5.5 Operations on power series

Power series have a richer structure than polynomials. For example, we also have division (this is similar to the move from $\mathbb{Z}$ to $\mathbb{Q}$). Assume that $a * b \neq 0$. Then (again, using list notation for brevity), we want to find, for any given $(a : as)$ and $(b : bs)$, the series $(c : cs)$ satisfying

$$
\begin{aligned}
&\quad (a : as) \mathbin{/} (b : bs) = (c : cs)\\
&\Leftrightarrow \{\text{- def. of division -}\}\\
&\quad (a : as) = (c : cs) * (b : bs)\\
&\Leftrightarrow \{\text{- def. of } * \text{ for } \mathit{Cons} \text{ -}\}\\
&\quad (a : as) = (c * b) : (cs * (b : bs) + [c] * bs)\\
&\Leftrightarrow \{\text{- equality on compnents, def. of division -}\}\\
&\quad c \ = a \mathbin{/} b \{\text{- and -}\}\\
&\quad as = cs * (b : bs) + [a \mathbin{/} b] * bs\\
&\Leftrightarrow \{\text{- arithmetics -}\}\\
&\quad c \ = a \mathbin{/} b \{\text{- and -}\}\\
&\quad cs = (as - [a \mathbin{/} b] * bs) \mathbin{/} (b : bs)
\end{aligned}
$$

This leads to the implementation:

$$
\begin{aligned}
&\mathbf{instance}\ (\mathit{Eq}\ a, \mathit{Fractional}\ a) \Rightarrow \mathit{Fractional}\ (\mathit{PowerSeries}\ a)\ \mathbf{where}\\
&\quad (/) = \mathit{divPS}\\
&\quad \mathit{fromRational} = \mathit{Single} \circ \mathit{fromRational}\\
&\mathit{divPS} :: (\mathit{Eq}\ a, \mathit{Fractional}\ a) \Rightarrow \mathit{PowerSeries}\ a \to \mathit{PowerSeries}\ a \to \mathit{PowerSeries}\ a\\
&\mathit{divPS}\ \mathit{as} \qquad\quad (\mathit{Single}\ b) \quad = \mathit{as} * \mathit{Single}\ (1 \mathbin{/} b)\\
&\mathit{divPS}\ (\mathit{Single}\ 0) \quad (\mathit{Cons}\ b\ \mathit{bs}) = \mathit{Single}\ 0\\
&\mathit{divPS}\ (\mathit{Single}\ a) \quad (\mathit{Cons}\ b\ \mathit{bs}) = \mathit{divPS}\ (\mathit{Cons}\ a\ (\mathit{Single}\ 0))\ (\mathit{Cons}\ b\ \mathit{bs})\\
&\mathit{divPS}\ (\mathit{Cons}\ a\ \mathit{as})\ (\mathit{Cons}\ b\ \mathit{bs}) = \mathit{Cons}\ c\ (\mathit{divPS}\ (\mathit{as} - (\mathit{Single}\ c) * \mathit{bs})\ (\mathit{Cons}\ b\ \mathit{bs}))\\
&\qquad\qquad\qquad\qquad\qquad\qquad\quad \mathbf{where}\ c = a \mathbin{/} b
\end{aligned}
$$

The first two equations allow us to also use division on polynomials, but the result will, in general, be a power series, not a polynomial. The first one should be self-explanatory. The second one extends a constant polynomial, in a process similar to that of long division.

For example:

$$
\begin{aligned}
&\mathit{ps0}, \mathit{ps1}, \mathit{ps2} :: (\mathit{Eq}\ a, \mathit{Fractional}\ a) \Rightarrow \mathit{PowerSeries}\ a\\
&\mathit{ps0} = 1 \mathbin{/} (1 - x)
\end{aligned}
$$

$ps1 = 1 \,/\, (1 - x)\hat{}2$
$ps2 = (x\hat{}2 - 2 * x + 1) \,/\, (x - 1)$

Every ps is the result of a division of polynomials: the first two return power series, the third is a polynomial (almost: it has a trailing 0.0).

$example0 \;\; = takePoly\ 10\ ps0$
$example01 = takePoly\ 10\ (ps0 * (1 - x))$

5.6 Formal derivative

Considering the analogy between power series and polynomial functions (via polynomials), we can arrive at a formal derivative for power series through the folowing computation:

$$\begin{aligned}\left(\sum_{n=0}^{\infty} a_n * x^n\right)' &= \sum_{n=0}^{\infty}(a_n * x^n)' = \sum_{n=0}^{\infty} a_n * (x^n)' = \sum_{n=0}^{\infty} a_n * (n * x^{n-1}) \\ &= \sum_{n=0}^{\infty}(n * a_n) * x^{n-1} = \sum_{n=1}^{\infty}(n * a_n) * x^{n-1} = \sum_{m=0}^{\infty}((m+1) * a_{m+1}) * x^m\end{aligned} \tag{1}$$

Thus the mth coefficient of the derivative is $(m + 1) * a_{m+1}$.

TODO: redo to arrive at the recursive formulation.

We can implement this, for example, as

```
deriv (Single a)    = Single 0
deriv (Cons a as) = deriv' as 1
  where deriv' (Single a)    n = Single (n * a)
        deriv' (Cons a as) n = Cons   (n * a) (deriv' as (n + 1))
```

Side note: we cannot in general implement a Boolean equality test for PowerSeries. For example, we know that *deriv ps0* equals *ps1* but we cannot compute *True* in finite time by comparing the coefficients of the two power series.

```
checkDeriv :: Integer → Bool
checkDeriv n = takePoly n (deriv ps0) == takePoly n ps1
```

Recommended reading: the Functional pearl: "Power series, power serious" McIlroy [1999].

5.7 Signals and Shapes

Shallow and deep embeddings of a DSL

TODO: perhaps textify DSL/

5.8 Helpers

```
instance Functor Poly where
  fmap = fmapPoly
fmapPoly :: (a → b) → (Poly a → Poly b)
fmapPoly f (Single a)    = Single (f a)
fmapPoly f (Cons a as) = Cons (f a) (fmapPoly f as)
po1 :: Num a ⇒ Poly a
po1 = 1 + x^2 - 3 * x^4
```

```
{-# LANGUAGE FlexibleInstances #-}
{-# LANGUAGE TypeSynonymInstances #-}
module DSLsofMath.W06 where
import DSLsofMath.W05
```

6 Week 6: Higher-order Derivatives and their Applications

6.1 Review

- key notion *homomorphism*: $S1 \to S2$
- questions ("equations"):
  - $S1 \xleftarrow{?} S2$ what is the homomorphism between two given structures
    - e.g., $apply : Num\ a \to Num\ (x \to a)$
  - $S1? \to S2$ what is $S1$ compatible with a given homomorphism
    - e.g., $eval : Poly\ a \to (a \to a)$
  - $S1 \to S2?$ what is $S2$ compatible with a given homomorphism
    - e.g., $applyFD : FD\ a \to (a, a)$
  - $S1 \xleftarrow{?} S2?$ can we find a good structure on $S2$ so that it becomes homomorphic w. $S1$?
    - e.g., $evalD : FunExp \to FD\ a$
- The importance of the last two is that they offer "automatic differentiation", i.e., any function constructed according to the grammar of *FunExp*, can be "lifted" to a function that computes the derivative (e.g., a function on pairs).

Example $f\ x = sin\ x + 2 * x$

We have: $f\ 0 = 1$, $f\ 2 = 4.909297426825682$s, etc.

The type of f is $f :: Floating\ a \Rightarrow a \to a$.

How do we compute, say, $f'\ 2$?

We have several choices.

1. Using *FunExp*

   Recall (week 3):

```
data FunExp = Const Rational
            | Id
            | FunExp :+: FunExp
            | FunExp :*: FunExp
            | FunExp :/: FunExp
            | Exp FunExp
            | Sin FunExp
            | Cos FunExp
              -- and so on
            deriving Show
```

What is the expression e for which $f = eval\ e$?
We have

$$
\begin{aligned}
& eval\ e\ x = f\ x \\
\Leftrightarrow\ & \\
& eval\ e\ x = sin\ x + 2 * x \\
\Leftrightarrow\ & \\
& eval\ e\ x = eval\ (Sin\ Id)\ x + eval\ (Const\ 2 :*: Id)\ x \\
\Leftrightarrow\ & \\
& eval\ e\ x = eval\ ((Sin\ Id) :+: (Const\ 2 :*: Id))\ x \\
\Leftarrow\ & \\
& e = Sin\ Id :+: (Const\ 2 :*: Id)
\end{aligned}
$$

Finally, we can apply *derive* and obtain

$$f'\ 2 = eval\ (derive\ e)\ 2$$

This can hardly be called "automatic", look at all the work we did in deducing e!
However, consider this:

$$e = f\ Id$$

(Perhaps it would have been better to use, in the definition of *FunExp*, X instead of Id.)
In general, to find the value of the derivative of a function f at a given x, we can use

$$drv\ f\ x = evalFunExp\ (derive\ (f\ Id))\ x$$

2. Using *FD*
   Recall

   type $FD\ a = (a \to a, a \to a)$
   $applyFD\ (f, g)\ x = (f\ x, g\ x)$

   The operations on $FD\ a$ are such that, if $eval\ e = f$, then

   $$(eval\ e, eval'\ e) = (f, f')$$

   We are looking for (g, g') such that

   $$f\ (g, g') = (f, f') \quad \text{-- (*)}$$

   so we can then do

   $$f'\ 2 = snd\ (applyFD\ (f\ (g, g'))\ 2)$$

   We can fullfill (*) if we can find a (g, g') that is a sort of "unit" for $FD\ a$:

   $$sin\ (g, g') = (sin, cos)$$
   $$exp\ (g, g') = (exp, exp)$$

   and so on.

In general, the chain rule gives us

$$f\ (g, g') = (f \circ g, (f' \circ g) * g')$$

Therefore, we need: $g = id$ and $g' = const\ 1$.

Finally

$$f'\ 2 = snd\ (applyFD\ (f\ (id, const\ 1))\ 2)$$

In general

$$drvFD\ f\ x = snd\ (applyFD\ (f\ (id, const\ 1))\ x)$$

computes the derivative of f at x.

$$f_1 :: FD\ Double \to FD\ Double$$
$$f_1 = f$$

3. Using pairs

We have **instance** *Floating* $a \Rightarrow$ *Floating* (a, a), moreover, the instance declaration looks exactly the same as that for $FD\ a$:

instance *Floating* $a \Rightarrow$ *Floating* $(FD\ a)$ **where**
  $exp\ (f, f') = (exp\ f, (exp\ f) * f')$
  $sin\ (f, f') = (sin\ f, (cos\ f) * f')$
  $cos\ (f, f') = (cos\ f, -(sin\ f) * f')$
instance *Floating* $a \Rightarrow$ *Floating* (a, a) **where**
  $exp\ (f, f') = (exp\ f, (exp\ f) * f')$
  $sin\ (f, f') = (sin\ f, cos\ f * f')$
  $cos\ (f, f') = (cos\ f, -(sin\ f) * f')$

In fact, the latter represents a generalisation of the former. It is also the "maximally general" such generalisation (discounting the "noise" generated by the less-than-clean design of *Num*, *Fractional*, *Floating*).

Still, we need to use this machinery. We are now looking for a pair of values (g, g') such that

$$f\ (g, g') = (f\ 2, f'\ 2)$$

In general

$$f\ (g, g') = (f\ g, (f'\ g) * g')$$

Therefore

$$f\ (g, g') = (f\ 2, f'\ 2)$$
$$\Leftrightarrow$$
$$(f\ g, (f'\ g) * g') = (f\ 2, f'\ 2)$$
$$\Leftarrow$$
$$g = 2, g' = 1$$

Introducing

$$var\ x = (x, 1)$$

we can, as in the case of FD, simplify matters a little:

$$f'\ x = snd\ (f\ (var\ x))$$

In general

$$drvP\ f\ x = snd\ (f\ (x, 1))$$

computes the derivative of f at x.

$$f_2 :: (Double, Double) \to (Double, Double)$$
$$f_2 = f$$

6.2 Higher-order derivatives

Consider

$$[f, f', f'', ...]$$

representing the evaluation of an expression and its derivatives:

$$evalAll\ e = (evalFunExp\ e) : evalAll\ (derive\ e)$$

Notice that, if

$$[f, f', f'', ...] = evalAll\ e$$

then

$$[f', f'', ...] = evalAll\ (derive\ e)$$

We want to define the operations on lists of functions in such a way that *evalAll* is a homomorphism. For example:

$$evalAll\ (e_1 :*: e_2) = evalAll\ e_1 * evalAll\ e_2$$

where the $(*)$ sign stands for the multiplication of infinite lists of functions, the operation we are trying to determine.

We have, writing *eval* for *evalFunExp* in order to save ink

$$\begin{array}{l}
\quad evalAll\ (e_1 :*: e_2) = evalAll\ e_1 * evalAll\ e_2 \\
\Leftrightarrow \\
\quad eval\ (e_1 :*: e_2) : evalAll\ (derive\ (e_1 :*: e_2)) = \\
\quad eval\ e_1 : evalAll\ (derive\ e) * eval\ e_1 : evalAll\ (derive\ e_2) \\
\Leftrightarrow \\
\quad (eval\ e_1 * eval\ e_2) : evalAll\ (derive\ (e_1 :*: e_2)) = \\
\quad eval\ e_1 : evalAll\ (derive\ e) * eval\ e_1 : evalAll\ (derive\ e_2) \\
\Leftrightarrow \\
\quad (eval\ e_1 * eval\ e_2) : evalAll\ (derive\ e_1 :*: e_2 :+: e_1 * derive\ e_2) = \\
\quad eval\ e_1 : evalAll\ (derive\ e) * eval\ e_1 : evalAll\ (derive\ e_2) \\
\Leftarrow \\
\quad (a : as) * (b : bs) = (a * b) : (as * (b : bs) + (a : as) * bs)
\end{array}$$

The final line represents the definition of $(*)$ needed for ensuring the conditions are met.

As in the case of pairs, we find that we do not need any properties of functions, other than their *Num* structure, so the definitions apply to any infinite list of *Num a*:

```
instance Num a ⇒ Num [a] where
  (a : as) + (b : bs) = (a + b) : (as + bs)
  (a : as) * (b : bs) = (a * b) : (as * (b : bs) + (a : as) * bs)
```

Exercise: complete the instance declarations for *Fractional* and *Floating*. Write a general derivative computation, similar to *drv* functions above:

```
drvList k f x = undefined   -- kth derivative of f at x
```

This is a very inefficient way of computing derivatives!

6.3 Polynomials

```
data Poly a = Single a | Cons a (Poly a)
                deriving (Eq, Ord)
evalPoly ::  Num a ⇒ Poly a → a → a
evalPoly (Single a)   x = a
evalPoly (Cons a as) x = a + x * evalPoly as x
```

6.4 Power series

No need for a separate type in Haskell

```
type PowerSeries a = Poly a   -- finite and infinite non-empty lists
```

Now we can divide, as well as add and multiply.

We can also derive:

```
deriv (Single a)   = Single 0
deriv (Cons a as) = deriv' as 1
  where deriv' (Single a)   n = Single (n * a)
        deriv' (Cons a as) n = Cons  (n * a) (deriv' as (n + 1))
```

and integrate:

```
integ :: Fractional a ⇒ PowerSeries a → a → PowerSeries a
integ as a0 = Cons a0 (integ' as 1)
  where integ' (Single a)   n = Single (a / n)
        integ' (Cons a as) n = Cons  (a / n) (integ' as (n + 1))
```

Everything here makes sense, irrespective of convergence, hence "formal".

If the power series involved do converge, then *eval* is a morphism between the formal structure and that of the functions represented:

$$eval\ as + eval\ bs = eval\ (as + bs)$$
$$eval\ as * eval\ bs = eval\ (as * bs)$$
$$eval\ (derive\ as) = D\ (eval\ as)$$
$$eval\ (integ\ as\ c)\ x\ = \int_0^x (eval\ as\ t)\ dt + c$$

6.5 Simple differential equations

Many first-order differential equations have the structure

$$f'\ x = g\ f\ x, \qquad f\ 0 = f_0$$

i.e., they are defined in terms of g.

The fundamental theorem of calculus gives us

$$f\ x = \int_0^x (g\ f\ t)\ dt + f_0$$

If $f = eval\ as$

$$eval\ as\ x = \int_0^x (g\ (eval\ as)\ t)\ dt + f_0$$

Assuming that g is a polymorphic function that commutes with *eval*

$$eval\ as\ x = \int_0^x (eval\ (g\ as)\ t)\ dt + f_0$$
$$eval\ as\ x = eval\ (integ\ (g\ as)\ f_0)\ x$$
$$as = integ\ (g\ as)\ f_0$$

Which functions g commute with *eval*? All the ones in *Num*, *Fractional*, *Floating*, by construction; additionally, as above, *deriv* and *integ*.

Therefore, we can implement a general solver for these simple equations:

```
solve :: Fractional a ⇒ (PowerSeries a → PowerSeries a) → a → PowerSeries a
solve g f0 = f    -- solves f' = g f, f 0 = f0
  where f = integ (g f) f0
idx = solve (λf → 1) 0
idf = eval 100 idx
expx = solve (λf → f) 1
expf = eval 100 expx
sinx = solve (λf → cosx) 0
cosx = solve (λf → −sinx) 1
sinf = eval 100 sinx
cosf = eval 100 cosx
idx, expx, sinx, cosx :: Fractional a ⇒ PowerSeries a
idf, expf, sinf, cosf :: Fractional a ⇒  a → a
```

6.6 The *Floating* structure of *PowerSeries*

Can we compute *exp as*?

Specification:

$$eval\ (exp\ as) = exp\ (eval\ as)$$

Differentiating both sides, we obtain

$$D\ (eval\ (exp\ as)) = exp\ (eval\ as) * D\ (eval\ as)$$

$$
\begin{aligned}
&\Leftrightarrow \{\text{- } eval \text{ morphism -}\}\\
&\quad eval\ (deriv\ (exp\ as)) = eval\ (exp\ as * deriv\ as)\\
&\Leftarrow\\
&\quad deriv\ (exp\ as) = exp\ as * deriv\ as
\end{aligned}
$$

Adding the "initial condition" $eval\ (exp\ as)\ 0 = exp\ (head\ as)$, we obtain

$$exp\ as = integ\ (exp\ as * deriv\ as)\ (exp\ (val\ as))$$

Note: we cannot use *solve* here, because the g function uses both $exp\ as$ and as (it "looks inside" its argument).

$$
\begin{aligned}
&\textbf{instance}\ (Eq\ a, Floating\ a) \Rightarrow Floating\ (PowerSeries\ a)\ \textbf{where}\\
&\quad \pi \quad\ \ = Single\ \pi\\
&\quad exp\ fs = integ\ (exp\ fs\ \ * deriv\ fs)\ (exp\ (val\ fs))\\
&\quad sin\ fs = integ\ (cos\ fs\ \ * deriv\ fs)\ (sin\ \ (val\ fs))\\
&\quad cos\ fs = integ\ (-sin\ fs * deriv\ fs)\ (cos\ \ (val\ fs))\\
&val :: PowerSeries\ a \to a\\
&val\ \ (Single\ a) \quad\ = a\\
&val\ \ (Cons\ a\ as) \quad = a
\end{aligned}
$$

In fact, we can implement *all* the operations needed for evaluating *FunExp* functions as power series!

$$
\begin{aligned}
&evalP :: (Eq\ r, Floating\ r) \Rightarrow FunExp \to PowerSeries\ r\\
&evalP\ (Const\ x)\ = Single\ (fromRational\ x)\\
&evalP\ (e_1\ :+:\ e_2) = evalP\ e_1 + evalP\ e_2\\
&evalP\ (e_1\ :*:\ e_2)\ = evalP\ e_1 * evalP\ e_2\\
&evalP\ (e_1\ :/:\ e_2)\ = evalP\ e_1\ /\ evalP\ e_2\\
&evalP\ Id \qquad\ = idx\\
&evalP\ (Exp\ e) \quad = exp\ (evalP\ e)\\
&evalP\ (Sin\ e) \quad = sin\ (evalP\ e)\\
&evalP\ (Cos\ e) \quad = cos\ (evalP\ e)
\end{aligned}
$$

6.7 Taylor series

If $f = eval\ [a_0, a_1, ..., a_n, ...]$, then

$$
\begin{aligned}
&f\ 0 \ = a_0\\
&f' \quad = eval\ (deriv\ [a_0, a_1, ..., a_n, ...])\\
&\qquad = eval\ ([a_1, 2 * a_2, 3 * a_3, ..., n * a_n, ...])\\
&\Rightarrow\\
&f'\ 0 \ = a_1\\
&f'' \quad = eval\ (deriv\ [a_1, 2 * a_2, ..., n * a_n, ...])\\
&\qquad = eval\ ([2 * a_2, 3 * 2 * a_3, ..., n * (n-1) * a_n, ...])\\
&\Rightarrow\\
&f''\ 0 = 2 * a_2
\end{aligned}
$$

In general:

$$f^{(k)}0 = fact\ k * a_k$$

Therefore

$$f = eval\ [f\ 0, f'\ 0, f''\ 0\ /\ 2, ..., f^{(n)}0\ /\ (fact\ n), ...]$$

The series $[f\ 0, f'\ 0, f''\ 0\ /\ 2, ..., f^{(n)}0\ /\ (fact\ n), ...]$ is called the Taylor series centred in 0, or the Maclaurin series.

Therefore, if we can represent f as a power series, we can find the value of all derivatives of f at 0!

```
derivs :: Num a ⇒ PowerSeries a → PowerSeries a
derivs as = derivs1 as 0 1
  where
  derivs1 (Cons a as) n factn = Cons  (a * factn)
                                      (derivs1 as (n + 1) (factn * (n + 1)))
  derivs1 (Single a)    n factn = Single (a * factn)
  -- remember that x = Cons 0 (Single 1)
ex3 = takePoly 10 (derivs (x^3 + 2 * x))
ex4 = takePoly 10 (derivs sinx)
```

In this way, we can compute all the derivatives at 0 for all functions f constructed with the grammar of *FunExp*. That is because, as we have seen, we can represent all of them by power series!

What if we want the value of the derivatives at $a \neq 0$?

We then need the power series of the "shifted" function g:

$$g\ x = f\ (x + a) \Leftrightarrow g = f \circ (+a)$$

If we can represent g as a power series, say $[b_0, b_1, ...]$, then we have

$$g^{(k)}0 = fact\ k * b_k = f^{(k)}a$$

In particular, we would have

$$f\ x = g\ (x - a) = \sum b_n * (x - a)^n$$

which is called the Taylor expansion of f at a.

Example:

We have that $idx = [0, 1]$, thus giving us indeed the values

$$[id\ 0, id'\ 0, id''\ 0, ...]$$

In order to compute the values of

$$[id\ a, id'\ a, id''\ a, ...]$$

for $a \neq 0$, we compute

```
ida a = takePoly 10 (derivs (evalP (Id :+: Const a)))
```

More generally, if we want to compute the derivative of a function f constructed with *FunExp* grammar, at a point a, we need the power series of $g\ x = f\ (x + a)$:

```
d f a = takePoly 10 (derivs (evalP (f (Id :+: Const a))))
```

Use, for example, our $f\ x = sin\ x + 2 * x$ above.

As before, we can use directly power series:

```
dP f a = takePoly 10 (derivs (f (idx + Single a)))
```

6.8 Associated code

```
instance Num a ⇒ Num (x → a) where
  f + g       = λx → f x + g x
  f − g       = λx → f x − g x
  f ∗ g       = λx → f x ∗ g x
  negate f    = negate ∘ f
  |f|         = |·| ∘ f
  signum f    = signum ∘ f
  fromInteger = const ∘ fromInteger
instance Fractional a ⇒ Fractional (x → a) where
  recip f        = recip ∘ f
  fromRational = const ∘ fromRational
instance Floating a ⇒ Floating (x → a) where
  π      = const π
  exp f  = exp ∘ f
  sin f  = sin ∘ f
  cos f  = cos ∘ f
  f ∗∗ g = λx → (f x) ∗∗ (g x)
    -- and so on
evalFunExp :: Floating a ⇒ FunExp → a → a
evalFunExp (Const α) = const (fromRational α)
evalFunExp Id        = id
evalFunExp (e1 :+: e2) = evalFunExp e1 + evalFunExp e2   -- note the use of "lifted +"
evalFunExp (e1 :∗: e2) = evalFunExp e1 ∗ evalFunExp e2   -- "lifted ∗"
evalFunExp (Exp e1)    = exp (evalFunExp e1)             -- and "lifted exp"
evalFunExp (Sin e1)    = sin (evalFunExp e1)
evalFunExp (Cos e1)    = cos (evalFunExp e1)
  -- and so on
derive (Const α) = Const 0
derive Id        = Const 1
derive (e1 :+: e2) = derive e1 :+: derive e2
derive (e1 :∗: e2) = (derive e1 :∗: e2) :+: (e1 :∗: derive e2)
derive (Exp e)   = Exp e :∗: derive e
derive (Sin e)   = Cos e :∗: derive e
derive (Cos e)   = Const (−1) :∗: Sin e :∗: derive e
instance Num FunExp where
  (+) = (:+:)
  (∗) = (:∗:)
  fromInteger n = Const (fromInteger n)
instance Fractional FunExp where
  (/) = (:/:)
instance Floating FunExp where
  exp = Exp
  sin = Sin
```

6.8.1 Not included to avoid overlapping instances

```
instance Num a ⇒ Num (FD a) where
  (f, f′) + (g, g′) = (f + g, f′ + g′)
```

$(f, f') * (g, g') = (f * g, f' * g + f * g')$
$fromInteger\ n = (fromInteger\ n, const\ 0)$

instance $Fractional\ a \Rightarrow Fractional\ (FD\ a)$ **where**
$(f, f') / (g, g') = (f / g, (f' * g - g' * f) / (g * g))$

instance $Floating\ a \Rightarrow Floating\ (FD\ a)$ **where**
$exp\ (f, f') = (exp\ f, (exp\ f) * f')$
$sin\ (f, f') = (sin\ f, (cos\ f) * f')$
$cos\ (f, f') = (cos\ f, -(sin\ f) * f')$

6.8.2 This is included instead

instance $Num\ a \Rightarrow Num\ (a, a)$ **where**
$(f, f') + (g, g') = (f + g, f' + g')$
$(f, f') * (g, g') = (f * g, f' * g + f * g')$
$fromInteger\ n = (fromInteger\ n, fromInteger\ 0)$

instance $Fractional\ a \Rightarrow Fractional\ (a, a)$ **where**
$(f, f') / (g, g') = (f / g, (f' * g - g' * f) / (g * g))$

instance $Floating\ a \Rightarrow Floating\ (a, a)$ **where**
$exp\ (f, f') = (exp\ f, (exp\ f) * f')$
$sin\ (f, f') = (sin\ f, cos\ f * f')$
$cos\ (f, f') = (cos\ f, -(sin\ f) * f')$

```
{-# LANGUAGE FlexibleInstances #-}
{-# LANGUAGE UndecidableInstances #-}
module DSLsofMath.W07 where
```

7 Week 7: Matrix algebra and linear transformations

Often, especially in engineering textbooks, one encounters the "definition": a vector is an $n+1$-tuple of real or complex numbers, arranged as a column:

```
        v0
         .
v  =     .
         .
        vn
```

Other times, this is suplemented by the definition of a "row vector":

```
v =  v0, ..., vn
```

The v*i*s are real or complex numbers, or, more generally, elements of a *field* (analogous to being an instance of *Fractional*). Vectors can be added "point-wise" and multiplied with "scalars", i.e., elements of the field:

```
             v0       w0         v0 + w0
              .        .            .
v + w  =      .   +    .     =      .
              .        .            .
             vn       wn         vn + wn


            s*v0
             .
s * v  =    .
             .
            s*vn
```

The scalar s "scales" all the components of v.

In fact, the most important feature of vectors is that they can be *uniquely* expressed as a simple sort of combination of other vectors:

```
        v0          1                0
         .          0                .
v  =     .  = v0 * 0 ...  + ... vn *
         .          .                0
        vn          0                1
```

We denote by

```
          0
          0
ek  =     .
          .
          0
          1      <-- position k
          0
          .
          .
          0
```

the vector that is everywhere 0 except at position k, where it is 1, so that

$$v = v_0 * e0 + ... + v_n * en$$

There is a temptation to model vectors by lists or tuples, but the simplest (at least conceptually) way is to view them as *functions*:

type S $= ...$ -- the scalars, forming a field ($\mathbb{R}$, or *Complex*, or *Zn*, etc.)
type $Vector\ G = G \to S$

Usually, G is finite, i.e., *Bounded* and *Enumerable*.

We know from the previous lectures that if S is an instance of *Num*, *Fractional*, etc. then so is $G \to S$, with the pointwise definitions. In particular, the instance declarations for $+$, multiplication, and embedding of constants, give us exactly the structure needed for the vector operations. For example

$$\begin{aligned}
& s * v \\
= & \{\text{- } s \text{ is promoted to a function -}\} \\
& const\ s * v \\
= & \{\text{- } Num \text{ instance definition -}\} \\
& \lambda g \to (const\ s)\ g * v\ g \\
= & \{\text{- definition of } const \text{ -}\} \\
& \lambda g \to s * v\ g
\end{aligned}$$

The set G is typically $\{0, 1, ..., n\}$. The basis vectors are then

$$e\ i : G \to S, e\ i\ g = i\ `is`\ g$$

Implementation:

$$\begin{aligned}
& is\ a\ b = \textbf{if}\ a == b\ \textbf{then}\ 1\ \textbf{else}\ 0 \\
& e\ g = \lambda(G\ g') \to g\ `is`\ g' \\
& toL\ v = [v\ g \mid g \leftarrow [minBound\,..\,maxBound]] \quad \text{-- so we can actually see them}
\end{aligned}$$

and every

$$v : G \to S$$

is trivially a linear combination of e is:

$$v = v\ 0 * e\ 0 + ... + v\ n * e\ n$$

7.1 Functions on vectors

What if we have another vector space, $Vector\ G' = G' \to S$? We are interested in functions $f : Vector\ G \to Vector\ G'$:

$$f\ v = f\ (v\ 0 * e\ 0 + ... + v\ n * e\ n)$$

A "good" function should translate the operations in $Vector\ G$ into operations in $Vector\ G'$, i.e., should be a homomorphism:

$$f\ v = f\ (v\ 0 * e\ 0 + ... + v\ n * e\ n) = v\ 0 * f\ (e\ 0) + ... + v\ n * f\ (e\ n)$$

But this means that we can determine the values of $f : (G \to S) \to (G' \to S)$ from just the values of $f \circ e : G \to (G' \to S)$, a much "smaller" function. Let $m = f \circ e$. Then

$$f\ v = v\ 0 * m\ 0 + ... + v\ n * m\ n$$

Each of $m\ k$ is a $Vector\ G'$, as is the resulting $f\ v$. We have

$$\begin{aligned}
&f\ v\ g' \\
&= \{\text{- as above -}\} \\
&\quad v\ 0 * m\ 0 + ... + v\ n * m\ n \\
&= \{\text{- } * \text{ and } + \text{ for functions are def. pointwise -}\} \\
&\quad v\ 0 * m\ 0\ g' + ... + v\ n * m\ n\ g' \\
&= \{\text{- using } sum \text{ -}\} \\
&\quad sum\ [v\ i * m\ i\ g' \mid i \leftarrow [minBound\,..\,maxBound]]
\end{aligned}$$

Implementation:

This is the almost the standard "vector-matrix" multiplication:

$$M = [m\ 0 \mid ... \mid m\ n]$$

The columns of M are the images of the canonical base vectors $e\ i$ through f. Every $m\ k$ has $card\ G'$ rows, and it has become standard to use $M\ i\ j$ to mean the ith element of the jth column, i.e., $m\ j\ i$, so that

$$(M * v)\ i = sum\ [M\ i\ j * v\ j \mid j \leftarrow [0\,..\,n]]$$

$$mul\ m\ v\ g' = sum\ [m\ g'\ g * v\ g \mid g \leftarrow [minBound\,..\,maxBound]]$$

Example:

$$(M * e\ k)\ i = sum\ [M\ i\ j * e\ k\ j \mid j \leftarrow [0\,..\,n]] = sum\ [M\ i\ k] = M\ i\ k$$

i.e., $e\ k$ extracts the kth column from M (hence the notation "e" for "extract").

Given an arbitrary matrix M, we can define

$$f\ v = M * v$$

and obtain a linear transformation $(M*)$. Moreover $((M*) \circ e)\ g\ g' = M\ g'\ g$, i.e., the matrix constructed as above for f is precisely M.

Exercise: compute $((M*) \circ e)\ g\ g'$.

Therefore, every linear transformation is of the form $(M*)$ and every $(M*)$ is a linear transformation.

Matrix-matrix multiplication is defined in order to ensure that

$$(M' * M) * v = M' * (M * v)$$

that is

$$((M' * M)*) = (M'*) \circ (M*)$$

Exercise: work this out in detail.

Exercise: show that matrix-matrix multiplication is associative.

Perhaps the simplest vector space is obtained for $G = ()$, the singleton set. In this case, the vectors $s : () \to S$ are functions that can take exactly one argument, therefore have exactly one value: $s\ ()$, so they are often identified with S. But, for any $v : G \to S$, we have a function $fv : G \to (() \to S)$, namely

$$fv\ g\ () = v\ g$$

fv is similar to our m function above. The associated matrix is

$$M = [m\ 0 \mid ... \mid m\ n] = [fv\ 0 \mid ... \mid fv\ n]$$

having $n + 1$ columns (the dimension of *Vector* G) and one row (dimension of *Vector* ()). Let $w :: Vector\ G$:

$$M * w = w\ 0 * fv\ 0 + ... + w\ n * fv\ n$$

$M * v$ and each of the $fv\ k$ are "almost scalars": functions of type $() \to S$, thus, the only component of $M * w$ is

$$(M * w)\ () = w\ 0 * fv\ 0\ () + ... + w\ n * fv\ n\ () = w\ 0 * v\ 0 + ... + w\ n * v\ n$$

i.e., the scalar product of v and w.

Remark: I have not discussed the geometrical point of view. For the connection between matrices, linear transformations, and geometry, I warmly recommend binge-watching the "Essence of linear algebra" videos on youtube (start here: https://www.youtube.com/watch?v=kjBOesZCoqc).

7.2 Examples of matrix algebra

7.2.1 Derivative

We have represented polynomials of degree $n + 1$ by the list of their coefficients. This is quite similar to "standard" geometrical vectors represented by $n + 1$ coordinates. This suggests that polynomials of degree $n + 1$ form a vector space, and we could interpret that as $\{0, ..., n\} \to \mathbb{R}$ (or, more generally, $Field\ a \Rightarrow \{0, ..., n\} \to a$. The operations $+$ and $*$ are defined in the same way as they are for functions.

The *derive* function takes polynomials of degree $n + 1$ to polynomials of degree n, and since $D\ (f + g) = D\ f + D\ g$ and $D\ (s * f) = s * D\ f$, we expect it to be a linear transformation. What is its associated matrix?

To answer that, we must first determine the canonical base vectors. As for geometrical vectors, they are

$$e\ i : \{0, ..., n\} \to Real, e\ i\ j = i\ `is`\ j$$

The evaluation of $e\ i$ returns the function $\lambda x \to x\hat{}i$, as expected.

The associated matrix will be

$$M = [D\ (e\ 0), D\ (e\ 1), ..., D\ (e\ n)]$$

where each $D\ (e\ i)$ has length n. Vector $e\ (i+1)$ represents $x\hat{}(i+1)$, therefore

$$D\ (e\ (i+1)) = (i+1) * x\hat{}i$$

i.e.

$$D\ (e\ (i+1))\ !!\ j = \textbf{if}\ i == j\ \textbf{then}\ i+1\ \textbf{else}\ 0$$

and

$$D\ (e\ 0) = 0$$

Example: $n + 1 = 3$:

```
        0  1  0
M  =
        0  0  2
```

Take the polynomial

$$3 * X\hat{}2 + 2 * X + 1$$

as a vector

```
        1
v  =    2
        3
```

and we have

```
              0  1  0      1        2
M  * v  =                 *  2   =
              0  0  2      3        6
```

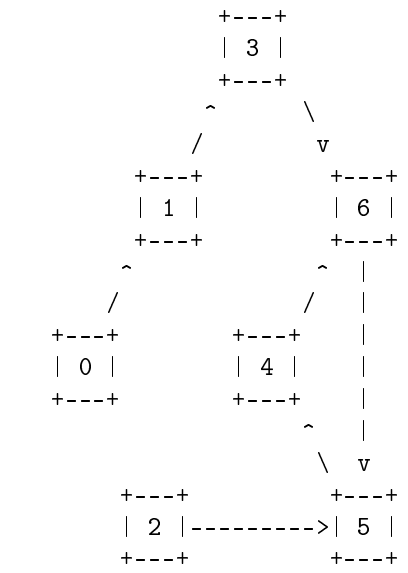
representing the polynomial $6 * X\hat{}2 + 2 * X$.

Exercise: write the (infinite-dimensional) matrix representing D for power series.

Exercise: write the matrix associate with integration of polynomials.

7.2.2 Simple deterministic systems (transition systems)

Simple deterministic systems are given by endo-functions on a finite set $f : G \to G$. They can often be conveniently represented as a graph, for example



Here, $G = \{0, ..., 6\}$. A node in the graph represents a state. A transition $i \to j$ means $f\ i = j$. Since f is an endo-function, every node must be the source of exactly one arrow.

We can take as vectors the characteristic functions of subsets of G, i.e., $G \to \{0, 1\}$. $\{0, 1\}$ is not a field w.r.t. the standard arithmetical operations (it is not even closed w.r.t. addition), and the standard trick to avoid this is to extend the type of the functions to $\mathbb{R}$.

The canonical basis vectors are, as usual, $e\ i = \lambda j \to i\text{`}is\text{`}j$. Each $e\ i$ is the characteristic function of a singleton set, $\{i\}$. Thus, the inputs to f are canonical vectors.

To write the matrix associated to f, we have to compute what vector is associated to each canonical base vector vector:

$$M = [f\ (e\ 0), f\ (e\ 1), ..., f\ (e\ n)]$$

Therefore:

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Starting with a canonical base vector $e\ i$, we obtain $M * e\ i = e\ (f\ i)$, as we would expect.

It is more interesting if we start with a non-base vector. For example, $e\ 2 + e\ 4$, which represents the subset $\{2, 4\}$.

The more interesting thing is if we start with something different from a basis vector, say $[0, 1, 1]$. We obtain $\{f\ 2, f\ 4\} = \{5, 6\}$, the image of $\{2, 4\}$ through f. In a sense, we can say that the two

computations were done in parallel. But that is not quite accurate: if start with $\{3, 4\}$, we no longer get the characteristic function of $\{f\ 3, f\ 4\} = \{6\}$, instead, we get a vector that does not represent a characteristic function at all: $[0, 0, 0, 0, 0, 0, 2]$. In general, if we start with an arbitrary vector, we can interpret this as starting with various quantities of some unspecified material in each state, simultaneously. If f were injective, the respective quantities would just gets shifted around, but in our case, we get a more interesting behaviour.

What if we do want to obtain the characteristic function of the image of a subset? In that case, we need to use other operations than the standard arithmetical ones, for example *min* and *max*. The problem is that $(\{0, 1\}, max, min)$ is not a field, and neither is $(\mathbb{R}, max, min)$. This is not a problem if all we want is to compute the evolutions of possible states, but we cannot apply most of the "deeper" results of linear algebra.

In the example above, we have:

```
newtype G = G Int deriving (Eq, Show)
instance Bounded G where
  minBound = G 0
  maxBound = G 6
instance Enum G where
  toEnum           = G
  fromEnum (G n) = n
```

The transition function:

```
f1 0 = 1
f1 1 = 3
f1 2 = 5
f1 3 = 6
f1 4 = 6
f1 5 = 4
f1 6 = 5
```

The associated matrix:

```
m1 (G g') (G g) = g' `is` f1 g
```

Test:

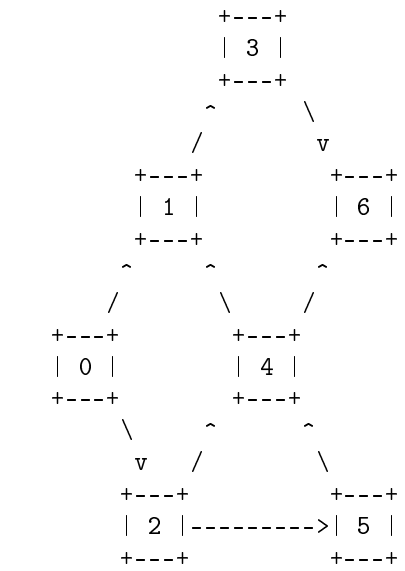
```
t1 = toL (mul m1 (e 3 + e 4))
```

7.2.3 Non-deterministic systems

Another interpretation of the application of M to characteristic functions of a subset is the following: assuming that all I know is that the system is in one of the states of the subset, where can it end up after one step? (this assumes the *max-min* algebra as above).

In this case, the uncertainty is entirely caused by the fact that we do not know the exact initial state. However, there are cases in which the output of f is not known, even when the input is known. Such situations are modelled by endo-relations: $R : G \to G$, with $g\ R\ g'$ if g' is a potential successor of g. Endo-relations can also be pictured as graphs, but the restriction that every node should be the source of exactly one arrow is lifted. Every node can be the source of one, none, or many arrows.

For example:



Now, starting in 0 we might and up either in 1 or 2 (but not both!). Starting in 6, the system breaks down: there is no successor state.

The matrix associated to R is built in the same fashion: we need to determine what vectors the canonical base vectors are associated with:

$$M = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{matrix}$$

Exercise: start with $e\ 2 + e\ 3$ and iterate a number of times, to get a feeling for the possible evolutions. What do you notice? What is the largest number of steps you can make before the result is the origin vector? Now invert the arrow from 2 to 4 and repeat the exercise. What changes? Can you prove it?

Implementation:

The transition function has type $G \to (G \to \mathit{Bool})$:

$$\begin{aligned} f_2\ 0\ g &= g == 1 \vee g == 2 \\ f_2\ 1\ g &= g == 3 \\ f_2\ 2\ g &= g == 4 \vee g == 5 \\ f_2\ 3\ g &= g == 6 \\ f_2\ 4\ g &= g == 1 \vee g == 6 \\ f_2\ 5\ g &= g == 4 \\ f_2\ 6\ g &= \mathit{False} \end{aligned}$$

The associated matrix:

$$m_2\ (G\ g')\ (G\ g) = f_2\ g\ g'$$

We need a *Num* instance for *Bool* (not a field!):

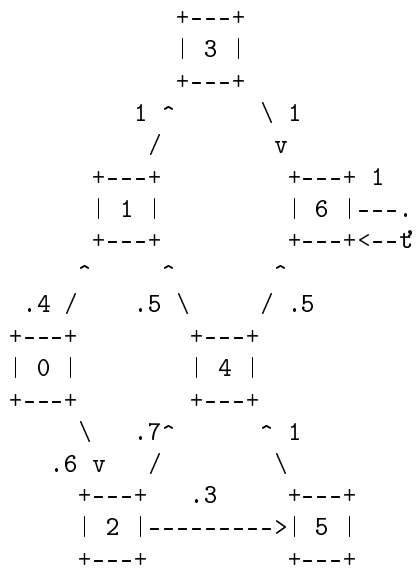
instance *Num Bool* **where**
$(+) = (\vee)$
$(*) = (\wedge)$
$fromInteger\ 0 = False$
$fromInteger\ 1 = True$
$negate = \neg$
$|\cdot| = id$
$signum = id$

Test:

$t_2 = toL\ (mul\ m_2\ (e\ 3 + e\ 4))$

7.2.4 Stochastic systems

Quite often, we have more information about the transition to possible future states. In particular, we can have *probabilities* of these transitions. For example



One could say that this case is a generalisation of the previous one, in which we can take all probabilities to be equally distributed among the various possiblities. While this is plausible, it is not entirely correct. For example, we have to introduce a transition from state 6 above. The nodes must be sources of *at least* one arrow.

In the case of the non-deterministic example, the "legitimate" inputs were characteristic functions, i.e., the "vector space" was $G \to \{0, 1\}$ (the scare quotes are necessary because, as discussed, the target is not a field). In the case of stochastic systems, the inputs will be *probability distributions* over G, that is, functions $p : G \to [0, 1]$ with the property that

$$sum\ [p\ g \mid g \leftarrow [0\,..\,6]] = 1$$

If we know the current probability distributions over states, then we can compute the next one by using the *total probability formula*, normally expressed as

$$p\ a = sum\ [p\ (a \mid b) * p\ b \mid b \leftarrow [0\,..\,6]]$$

This formula in itself would be worth a lecture. For one thing, the notation is extremely suspicious. $(a \mid b)$, which is usually read "a, given b", is clearly not of the same type as a or b, so cannot really be an argument to p. For another, the $p\ a$ we are computing with this formula is not the $p\ a$ which must eventually appear in the products on the right hand side. I do not know how this notation came about: it is neither in Bayes' memoir, nor in Kolmogorov's monograph.

The conditional probability $p\ (a \mid b)$ gives us the probability that the next state is a, given that the current state is b. But this is exactly the information summarised in the graphical representation. Moreover, it is clear that, at least formally, the total probability formula is identical to a matrix-vector multiplication.

As usual, we write the associated matrix by looking at how the canonical base vectors are transformed. In this case, the canonical base vector $e\ i = \lambda j \to i\ `is`\ j$ is the probability distribution *concentrated* in i:

```
        0    0    0    0    0    0    0
        .4   0    0    0    .5   0    0
        .6   0    0    0    0    0    0
M   =   0    1    0    0    0    0    0
        0    0    .7   0    0    1    0
        0    0    .3   0    0    0    0
        0    0    0    1    .5   0    1
```

Exercise: starting from state 0, how many steps do you need to take before the probability is concentrated in state 6? Reverse again the arrow from 2 to 4. What can you say about the long-term behaviour of the system now?

Exercise: Implement the example. You will need to define:

The transition function

$f_3 :: G \to (G \to Double)$ -- but we want only $G \to (G \to [0,1])$, the unit interval
$f_3\ g\ g' = undefined$ -- the probability of getting to g' from g

The associated matrix

$m_3 :: G \to (G \to Double)$
$m_3\ g'\ g = undefined$

Test

$t_3 = toL\ (mul\ m_3\ (e\ 2 + e\ 4))$

7.3 Monadic dynamical systems

All the examples of dynamical systems we have seen in the previous section have a similar structure. They work by taking a state (which is one of the generators) and return a structure of possible future states of type G:

- deterministic: there is exactly one possible future states: we take an element of G and return an element of G. The transition function has the type $f : G \to G$, the structure of the target is just G itself.

- non-deterministic: there is a set of possible future states, which we have implemented as a characteristic function $G \to \{0,1\}$. The transition function has the type $f : G \to (G \to \{0,1\})$. The structure of the target is the *powerset* of G.
- stochastic: given a state, we compute a probability distribution over possible future states. The transition function has the type $f : G \to (G \to [0,1])$, the structure of the target is the probability distributions over G.

Therefore:

- deterministic: $f : G \to Id\ G$
- non-deterministic: $f : G \to Powerset\ G$, where $Powerset\ G = G \to \{0,1\}$
- stochastic: $f : G \to Prob\ G$, where $Prob\ G = G \to [0,1]$

We have represented the elements of the various structures as vectors. We also had a way of representing, as structures of possible states, those states that were known precisely: these were the canonical base vectors $e\ i$. Due to the nature of matrix-vector multiplication, what we have done was in effect:

$$
\begin{aligned}
& M * v \quad \text{-- v represents the current possible states} \\
= \ & \{\text{- } v \text{ is a linear combination of the base vectors -}\} \\
& M * (v\ 0 * e\ 0 + ... + v\ n * e\ n) \\
= \ & \{\text{- homomorphism -}\} \\
& v\ 0 * (M * e\ 0) + ... + v\ n * (M * e\ n) \\
= \ & \{\text{- } e\ i \text{ represents the perfectly known current state } i, \text{ therefore } M * e\ i = f\ i \text{ -}\} \\
& v\ 0 * f\ 0 + ... + v\ n * f\ n
\end{aligned}
$$

So, we apply f to every state, as if we were starting from precisely that state, obtaining the possible future states starting from that state, and then collect all these "hypothetical" possible future states in some way that takes into account the initial uncertainty (represented by $v\ 0$, ..., $v\ n$) and the nature of the uncertainty (the specific $+$ and $*$).

If you examine the types of the operations involved

$$e : G \to Possible\ G$$

and

$$Possible\ G \to (G \to Possible\ G) \to Possible\ G$$

you see that they are very similar to the monadic operations

$$
\begin{aligned}
& return : g \to m\ g \\
& (\gg\!=) \ : m\ g \to (g \to m\ g') \to m\ g'
\end{aligned}
$$

which suggests that the representation of possible future states might be monadic. Indeed, that is the case.

Since we implemented all these as matrix-vector multiplications, this raises the question: is there a monad underlying matrix-vector multiplication, such that the above are instances of it (obtained by specialising the scalar type S)?

Exercise: write *Monad* instances for *Id*, *Powerset*, *Prob*.

7.4 The monad of linear algebra

The answer is yes, up to a point. Haskell *Monad*s, just like *Functor*s, require *return* and $\gg=$ to be defined for every type. This will not work, in general. Our definition will work for *finite types* only.

```
type S          = Double
data Vector g = V (g → S)
toF (V v)       = v

class      (Bounded a, Enum a, Eq a) ⇒ Finite a where
instance (Bounded a, Enum a, Eq a) ⇒ Finite a where

class FinFunc f where
  func :: (Finite a, Finite b) ⇒ (a → b) → f a → f b

instance FinFunc Vector where
  func = funcV

funcV :: (Finite g, Eq g') ⇒ (g → g') → Vector g → Vector g'
funcV f (V v) = V (λg' → sum [v g | g ← [minBound .. maxBound], g' == f g])

class FinMon f where
  embed  :: Finite a ⇒ a → f a
  bind   :: (Finite a, Finite b) ⇒ f a → (a → f b) → f b

instance FinMon Vector where
  embed a       = V (λa' → if a == a' then 1 else 0)
  bind (V v) f = V (λg' → sum [toF (f g) g' * v g | g ← [minBound .. maxBound]])
```

A better implementation, using associated types, is in file *Vector.lhs* in the repository.

Exercises:

1. Prove that the functor laws hold, i.e.

   $$\begin{aligned} func\ id &= id \\ func\ (g \circ f) &= func\ g \circ func\ f \end{aligned}$$

2. Prove that the monad laws hold, i.e.

   $$\begin{aligned} bind\ v\ return &= v \\ bind\ (return\ g)\ f &= f\ g \\ bind\ (bind\ v\ f)\ h &= bind\ v\ (\lambda g' \to bind\ (f\ g')\ h) \end{aligned}$$

3. What properties of S have you used to prove these properties? Define a new type class *GoodClass* that accounts for these (and only these) properties.

7.5 Associated code

TODO: import from suitable earlier lecture.

```
instance Num a ⇒ Num (x → a) where
  f + g       = λx → f x + g x
  f − g       = λx → f x − g x
  f * g       = λx → f x * g x
  negate f    = negate ∘ f
```

```
  |f|           = |·| ∘ f
  signum f      = signum ∘ f
  fromInteger   = const ∘ fromInteger
instance Fractional a ⇒ Fractional (x → a) where
  recip f       = recip ∘ f
  fromRational  = const ∘ fromRational
instance Floating a ⇒ Floating (x → a) where
  π             = const π
  exp f         = exp ∘ f
  sin f         = sin ∘ f
  cos f         = cos ∘ f
  f ** g        = λx → (f x) ** (g x)
    -- and so on
```

8 Week 8: Exponentials and Laplace

8.1 The Exponential Function

```
{-# LANGUAGE FlexibleInstances #-}
{-# LANGUAGE TypeSynonymInstances #-}
module DSLsofMath.W08 where
import DSLsofMath.W05
import DSLsofMath.W06
```

One of the classical analysis textbooks, Rudin's Rudin [1987] starts with a prologue on the exponential function. The first sentence is

> This is undoubtedly the most important function in mathematics.

Rudin goes on

> It is defined, for every complex number z, by the formula
>
> $exp\ z = \Sigma\ (z\hat{}n)\ /\ n\,!$

We have defined the exponential function as the function represented by the power series

```
expx :: Fractional a ⇒ PowerSeries a
expx = integ expx 1
```

and approximated by

```
expf :: Fractional a ⇒ a → a
expf = eval 100 expx
```

It is easy to see, using the definition of *integ* that the power series *expx* is, indeed

$$expx = [1, 1\ /\ 2, 1\ /\ (2 * 3), ..., 1\ /\ (2 * 3 * ... * n), ..]$$

We can compute the exponential for complex values if we can give an instance of *Fractional* for complex numbers. We could use the datatype *Data.Complex* from the Haskell standard library, but we prefer to roll our own in order to remind the basic operations on complex numbers.

As we saw in week 1, complex values can be represented as pairs of real values.

```
newtype Complex r = C (r, r) deriving (Eq, Show)
i :: Num a ⇒ Complex a
i = C (0, 1)
```

Now, we have, for example

```
ex1 :: Fractional a ⇒ Complex a
ex1 = expf i
```

We have $ex1 = C\ (0.5403023058681398, 0.8414709848078965)$. Note that

$$cosf\ 1 = 0.5403023058681398$$
$$sinf\ 1 = 0.8414709848078965$$

and therefore $expf\ i == C\ (cosf\ 1, sinf\ 1)$. Coincidence?

Instead of evaluating the sum of the terms $a_n * z\hat{}n$, let us instead collect the terms in a series:

$$terms\ as\ z = terms1\ as\ z\ 0\ \textbf{where}$$
$$terms1\ (Cons\ a\ as)\ z\ n = Cons\ (a * z\hat{}n)\ (terms1\ as\ z\ (n+1))$$

We obtain

$$ex2 :: Fractional\ a \Rightarrow PowerSeries\ (Complex\ a)$$
$$ex2 = takePoly\ 10\ (terms\ expx\ i)$$

$$\begin{aligned} ex2 = [\, & C\ (1.0, & 0.0), & C\ (0.0, 1.0 &) \\ , & C\ (-0.5, & 0.0), & C\ (0.0, -0.16666666666666666 &) \\ , & C\ (4.1666666666666664e{-}2, & 0.0), & C\ (0.0, 8.333333333333333e{-}3 &) \\ , & C\ (-1.3888888888888887e{-}3, & 0.0), & C\ (0.0, -1.9841269841269839e{-}4) & \\ , & C\ (2.4801587301587298e{-}5, & 0.0), & C\ (0.0, 2.7557319223985884e{-}6 &) \\ &] & & & \end{aligned}$$

We can see that the real part of this series is the same as

$$ex2R = takePoly\ 10\ (terms\ cosx\ 1)$$

and the imaginary part is the same as

$$ex2I = takePoly\ 10\ (terms\ sinx\ 1)$$

(within approx 20 decimals). But the terms of a series evaluated at 1 are the coefficients of the series. Therefore, the coefficients of *cosx* are

$$[1, 0, -1\ /\ 2!, 0, 1\ /\ 4!, 0, -1\ /\ 6!, ...]$$

i.e.

$$a\ (2 * n) = (-1)\hat{}n\ /\ (2 * n)\,!$$
$$a\ (2 * n + 1) = 0$$

and the terms of *sinx* are

$$[0, 1, 0, -1\ /\ 3!, 0, 1\ /\ 5!, 0, -1\ /\ 7!, ...]$$

i.e.

$$a\ (2 * n) = 0$$
$$a\ (2 * n + 1) = (-1)\hat{}n\ /\ (2 * n + 1)\,!$$

This can be proven from the definitions of *cosx* and *sinx*. From this we obtain *Euler's formula*:

$$exp\ (i * x) = cos\ x + i * sin\ x$$

One thing which comes out of Euler's formula is the fact that the exponential is a *periodic function*. A function $f : A \to B$ is said to be periodic if there exists $T \in A$ such that

$$f\ a = f\ (a + T) \quad \text{-- } \forall a \in A$$

(therefore, for this definition to make sense, we need addition on A; in fact we normally assume at least group structure, i.e., addition and subtraction).

Since *sin* and *cos* are periodic, with period $2 * \pi$, we have, using the standard notation $a + i * b$ for $C\ (a, b)$

$$\begin{aligned}
&e\hat{}(a + i * b + i * 2 * \pi) \\
&= \\
&e\hat{}(a + i * (b + 2 * \pi)) \\
&= \\
&e\hat{}a * e\hat{}(i * (b + 2 * \pi)) \\
&= \\
&e\hat{}a * (cos\ (b + 2 * \pi) + i * sin\ (b + 2 * \pi)) \\
&= \\
&e\hat{}a * (cos\ b + i * sin\ b) \\
&= \\
&e\hat{}a * e\hat{}(i * b) \\
&= \\
&e\hat{}(a + i * b)
\end{aligned}$$

8.1.1 Exponential function: Associated code

TODO: Perhaps import from W01

```
instance Num r => Num (Complex r) where
  (+) = addC
  (*) = mulC
  fromInteger n = C (fromInteger n, 0)
  -- abs = absC – requires Floating r as context
addC :: Num r => Complex r -> Complex r -> Complex r
addC (C (a, b)) (C (x, y)) = C ((a + x), (b + y))
mulC :: Num r => Complex r -> Complex r -> Complex r
mulC (C (ar, ai)) (C (br, bi)) = C (ar * br - ai * bi, ar * bi + ai * br)
modulusSquaredC :: Num r => Complex r -> r
modulusSquaredC (C (x, y)) = x^2 + y^2
absC :: Floating r => Complex r -> Complex r
absC c = C (sqrt modulusSquaredC c, 0)
scale :: Num r => r -> Complex r -> Complex r
scale a (C (x, y)) = C (a * x, a * y)
conj :: Num r => Complex r -> Complex r
conj (C (x, x'))      = C (x, -x')
instance Fractional r => Fractional (Complex r) where
  (/) = divC
  fromRational r = C (fromRational r, 0)
divC :: Fractional a => Complex a -> Complex a -> Complex a
divC x y = scale (1 / modSq) (x * conj y)
  where modSq = modulusSquaredC y
```

8.2 The Laplace transform

This material was inspired by Quinn and Rai [2008], which is highly recommended reading.

Consider the differential equation

$$f'' \; x - 3 * f' \; x + 2 * f \; x = exp \; (3 * x), f \; 0 = 1, f' \; 0 = 0$$

We can solve such equations with the machinery of power series:

$$fs = integ \; fs' \; 1$$
$$\textbf{where } fs' = integ \; (exp \; (3 * x) + 3 * fs' - 2 * fs) \; 0$$

We have done this by "zooming in" on the function f and representing it by a power series, $f \; x = \Sigma \; a_n * x \hat{} n$. This allows us to reduce the problem of finding a function $f : \mathbb{R} \to \mathbb{R}$ to that of finding a function $a : \mathbb{N} \to \mathbb{R}$ (or finding a list of sufficiently many a-values for a good approximation).

Still, recursive equations are not always easy to solve (especially without a computer), so it's worth looking for alternatives.

We have gone from

$$a : \mathbb{N} \to \mathbb{R} \qquad to \qquad f : \mathbb{R} \to \mathbb{R}$$

via

$$\Sigma \; a_n * x \hat{} n$$

In "zooming out", we would like to go one step further

$$a : \mathbb{N} \to \mathbb{R} \qquad to \qquad f : \mathbb{R} \to \mathbb{R} \qquad to \qquad ?$$

via

$$\Sigma \; a_n * x \hat{} n \qquad\qquad ?$$

At least the second part is "easy". The analogue of series for a continuous function is integral

$$\Sigma \; a_n * x \hat{} n \qquad\qquad \int_0^\infty (f \; t) * x \hat{} t \; dt$$

We note that, for the integral $\int_0^\infty (f \; t) * x \hat{} t \; dt$ to converge for a larger class of functions (say, bounded functions), we have to limit ourselves to $|x| < 1$. Both this condition and the integral make sense for $x \in \mathbb{C}$, so we could take

$$a : \mathbb{N} \to \mathbb{R} \qquad to \qquad f : \mathbb{R} \to \mathbb{R} \qquad to \qquad F : \{ z \mid\mid z \mid < 1 \} \to \mathbb{C}$$

but let us stick to $\mathbb{R}$ for now.

Writing, somewhat optimistically

$$\mathscr{L} \; f \; x = \int_0^\infty (f \; t) * x \hat{} t \; dt$$

we can ask ourselves what $\mathscr{L} \; f'$ looks like. After all, we want to solve *differential* equations by "zooming out". We have

$$\mathscr{L}\ f'\ x = \int_0^\infty\ (f'\ t) * x\hat{}t\ dt$$

Remember that $D\ (f * g) = D\ f * g + f * D\ g$, therefore

$$\begin{array}{l}
\quad \mathscr{L}\ f'\ x \\
= \{ g\ t = x\hat{}t; g'\ t = log\ x * x\hat{}t \} \\
\quad \int_0^\infty\ (D\ (f\ t * x\hat{}t)) - f\ t * log\ x * x\hat{}t\ dt \\
= \\
\quad \int_0^\infty\ (D\ (f\ t * x\hat{}t))\ dt - \int_0^\infty\ f\ t * log\ x * x\hat{}t\ dt \\
= \\
\quad lim_{t\to\infty}(f\ t * x\hat{}t) - (f\ 0 * x\hat{}0) - log\ x * \int_0^\infty\ f\ t * x\hat{}t\ dt \\
= \\
\quad - f\ 0 - log\ x * \int_0^\infty\ f\ t * x\hat{}t\ dt \\
= \\
\quad - f\ 0 - log\ x * \mathscr{L}\ f\ x
\end{array}$$

The factor $log\ x$ is somewhat awkward. Let us therefore return to the definition of $\mathscr{L}$ and operate a change of variables:

$$\begin{array}{l}
\quad \mathscr{L}\ f\ x = \int_0^\infty\ (f\ t) * x\hat{}t\ dt \\
\Leftrightarrow \{ x = exp\ (log\ x) \} \\
\quad \mathscr{L}\ f\ x = \int_0^\infty\ (f\ t) * (exp\ (log\ x))\hat{}t\ dt \\
\Leftrightarrow \{ (a\hat{}b)\hat{}c = a\hat{}(b * c) \} \\
\quad \mathscr{L}\ f\ x = \int_0^\infty\ (f\ t) * exp\ (log\ x * t)\ dt
\end{array}$$

Since $log\ x < 0$ for $|x| < 1$, we make the substitution $-s = log\ x$. The condition $|x| < 1$ becomes $s > 0$ (or, in $\mathbb{C}$, $real\ s > 0$), and we have

$$\mathscr{L}\ f\ s = \int_0^\infty\ (f\ t) * exp\ (-s * t)\ dt$$

This is the definition of the Laplace transform of the function f. Going back to the problem of computing $\mathscr{L}\ f'$, we now have

$$\begin{array}{l}
\quad \mathscr{L}\ f'\ s \\
= \{\text{- The computation above with } s = -log\ x. \text{ -}\} \\
\quad - f\ 0 + s * \mathscr{L}\ f\ s
\end{array}$$

We have obtained

$$\mathscr{L}\ f'\ s = s * \mathscr{L}\ f\ s - f\ 0$$

From this, we can deduce

$$\begin{array}{l}
\quad \mathscr{L}\ f''\ s \\
= \\
\quad s * \mathscr{L}\ f'\ s - f'\ 0 \\
= \\
\quad s * (s * \mathscr{L}\ f\ s - f\ 0) - f'\ 0
\end{array}$$

$$
\begin{array}{l}
= \\
\quad s\hat{}2 * \mathscr{L}\ f\ s - s * f\ 0 - f'\ 0
\end{array}
$$

Exercise: what is the general formula for $\mathscr{L}\ f^{(k)} s$?

Returning to our differential equation, we have

$$
\begin{array}{l}
\quad f''\ x - 3 * f'\ x + 2 * f\ x = exp\ (3 * x), f\ 0 = 1, f'\ 0 = 0 \\
\Leftrightarrow \{\text{- point-free form -}\} \\
\quad f'' - 3 * f' + 2 * f = exp \circ (3*), f\ 0 = 1, f'\ 0 = 0 \\
\Rightarrow \{\text{- applying } \mathscr{L} \text{ to both sides -}\} \\
\quad \mathscr{L}\ (f'' - 3 * f' + 2 * f) = \mathscr{L}\ (exp \circ (3*)), f\ 0 = 1, f'\ 0 = 0 \quad \text{-- Eq. (1)}
\end{array}
$$

Remark: Note that this is a necessary condition, but not a sufficient one. The Laplace transform is not injective. For one thing, it does not take into account the behaviour of f for negative arguments. Because of this, we often assume that the domain of definition for functions to which we apply the Laplace transform is $\mathbb{R}_{\geqslant 0}$. For another, it is known that changing the values of f for a countable number of its arguments does not change the value of the integral.

For the definition of $\mathscr{L}$ and the linearity of the integral, we have that, for any f and g for which the transformation is defined, and for any constants α and β

$$\mathscr{L}\ (\alpha * f + \beta * g) = \alpha * \mathscr{L}\ f + \beta * \mathscr{L}\ g$$

Note that this is an equality between functions. (Comparing to last week we can also see f and g as vectors and $\mathscr{L}$ as a linear transformation.)

Applying this to the left-hand side of (1), we have for any s

$$
\begin{array}{l}
\quad \mathscr{L}\ (f'' - 3 * f' + 2 * f)\ s \\
= \{\text{- } \mathscr{L} \text{ is linear -}\} \\
\quad \mathscr{L}\ f''\ s - 3 * \mathscr{L}\ f'\ s + 2 * \mathscr{L}\ f\ s \\
= \{\text{- re-writing } \mathscr{L}\ f'' \text{ and } \mathscr{L}\ f' \text{ in terms of } \mathscr{L}\ f \text{ -}\} \\
\quad s\hat{}2 * \mathscr{L}\ f\ s - s * f\ 0 - f'\ 0 - 3 * (s * \mathscr{L}\ f\ s - f\ 0) + 2 * \mathscr{L}\ f\ s \\
= \{\text{- } f\ 0 = 1,\ f'\ 0 = 0 \text{ -}\} \\
\quad (s\hat{}2 - 3 * s + 2) * \mathscr{L}\ f\ s - s + 3
\end{array}
$$

For the right-hand side, we apply the definition:

$$
\begin{array}{l}
\quad \mathscr{L}\ (exp \circ (3*))\ s \\
= \{\text{- Def. of } \mathscr{L} \text{ -}\} \\
\quad \int_0^\infty\ exp\ (3 * t) * exp\ (-s * t)\ dt \\
= \\
\quad \int_0^\infty\ exp\ ((3 - s) * t))\ dt \\
= \\
\quad lim_{t \to \infty} \frac{exp\ ((3-s)*t)}{3-s} - \frac{exp\ ((3-s)*0)}{3-s} \\
= \{\text{- for } s > 3 \text{ -}\} \\
\quad \frac{1}{s-3}
\end{array}
$$

Therefore, we have, writing F for $\mathscr{L}\ f$

$(s\hat{}2 - 3 * s + 2) * F\ s - s + 3 = \frac{1}{s-3}$

and therefore

$$
\begin{array}{l}
\quad F\ s \\
= \{\text{- Solve for } F\ s \text{ -}\} \\
\quad \frac{\frac{1}{s-3}+s-3}{s\hat{}2-3*s+2} \\
= \{\text{- } s\hat{}2 - 3 * s + 2 = (s-1) * (s-2) \text{ -}\} \\
\quad \frac{10-6*s+s\hat{}2}{(s-1)*(s-2)*(s-3)}
\end{array}
$$

We now have the problem of "recovering" the function f from its Laplace transform. The standard approach is to use the linearity of $\mathscr{L}$ to write F as a sum of functions with known inverse transforms. We know one such function:

$exp\ (\alpha * t)$ {- is the inverse Laplace transform of -} $1\ /\ (s - \alpha)$

In fact, in our case, this is all we need.

The idea is to write $F\ s$ as a sum of three fractions with denominators $s-1$, $s-2$, and $s-3$ respectively, i.e., to find A, B, and C such that

$$
\begin{array}{l}
A\ /\ (s-1) + B\ /\ (s-2) + C\ /\ (s-3) = (10 - 6 * s + s\hat{}2)\ /\ ((s-1) * (s-2) * (s-3)) \\
\Rightarrow \\
A * (s-2) * (s-3) + B * (s-1) * (s-3) + C * (s-1) * (s-2) = 10 - 6 * s + s\hat{}2 \quad \text{-- (2)}
\end{array}
$$

We need this equality (2) to hold for values $s > 3$. A *sufficient* condition for this is for (2) to hold for *all* s. A *necessary* condition for this is for (2) to hold for the specific values 1, 2, and 3.

$$
\begin{array}{ll}
\text{For } s = 1 : A * (-1) * (-2) = 10 - 6 + 1 & \Rightarrow A = 2.5 \\
\text{For } s = 2 : B * 1 * (-1) \quad\ = 10 - 12 + 4 & \Rightarrow B = -2 \\
\text{For } s = 3 : C * 2 * 1 \qquad\ = 10 - 18 + 9 & \Rightarrow C = 0.5
\end{array}
$$

It is now easy to check that, with these values, (2) does indeed hold, and therefore that we have

$$F\ s = 2.5 * (1\ /\ (s-1)) - 2 * (1\ /\ (s-2)) + 0.5 * (1\ /\ (s-3))$$

The inverse transform is now easy:

$$f\ t = 2.5 * exp\ t - 2 * exp\ (2 * t) + 0.5 * exp\ (3 * t)$$

Our mix of necessary and sufficient conditions makes it necessary to check that we have, indeed, a solution for the differential equation. The verification is in this case trivial.

9 End

TODO: sum up and close

References

R. A. Adams and C. Essex. *Calculus: a complete course*. Pearson Canada, 7th edition, 2010.

N. Botta, P. Jansson, and C. Ionescu. Contributions to a computational theory of policy advice and avoidability. *Journal of Functional Programming*, 2017a. Accepted for publication 2017-09-20.

N. Botta, P. Jansson, C. Ionescu, D. R. Christiansen, and E. Brady. Sequential decision problems, dependent types and generic solutions. *Logical Methods in Computer Science*, 13(1), 2017b. doi: 10.23638/LMCS-13(1:7)2017. URL `https://doi.org/10.23638/LMCS-13(1:7)2017`.

R. Boute. The decibel done right: a matter of engineering the math. *Antennas and Propagation Magazine, IEEE*, 51(6):177–184, 2009. doi: 10.1109/MAP.2009.5433137.

C. H. Edwards, D. E. Penney, and D. Calvis. *Elementary Differential Equations*. Pearson Prentice Hall Upper Saddle River, NJ, 6h edition, 2008.

D. Gries and F. B. Schneider. *A logical approach to discrete math*. Springer, 1993. doi: 10.1007/978-1-4757-3837-7.

D. Gries and F. B. Schneider. Teaching math more effectively, through calculational proofs. *American Mathematical Monthly*, pages 691–697, 1995. doi: 10.2307/2974638.

C. Ionescu and P. Jansson. Dependently-typed programming in scientific computing: Examples from economic modelling. In R. Hinze, editor, *24th Symposium on Implementation and Application of Functional Languages (IFL 2012)*, volume 8241 of *LNCS*, pages 140–156. Springer-Verlag, 2013a. doi: 10.1007/978-3-642-41582-1_9.

C. Ionescu and P. Jansson. Dependently-typed programming in scientific computing. In *Implementation and Application of Functional Languages*, pages 140–156. Springer Berlin Heidelberg, 2013b. doi: 10.1007/978-3-642-41582-1_9.

C. Ionescu and P. Jansson. Domain-specific languages of mathematics: Presenting mathematical analysis using functional programming. In J. Jeuring and J. McCarthy, editors, *Proceedings of the 4th and 5th International Workshop on Trends in Functional Programming in Education, Sophia-Antipolis, France and University of Maryland College Park, USA, 2nd June 2015 and 7th June 2016*, volume 230 of *Electronic Proceedings in Theoretical Computer Science*, pages 1–15. Open Publishing Association, 2016. doi: 10.4204/EPTCS.230.1.

C. Jaeger, P. Jansson, S. van der Leeuw, M. Resch, and J. D. Tabara. GSS: Towards a research program for Global Systems Science. `http://blog.global-systems-science.eu/?p=1512`, 2013. ISBN 978.3.94.1663-12-1. Conference Version, prepared for the Second Open Global Systems Science Conference June 10-12, 2013, Brussels.

R. Kraft. Functions and parameterizations as objects to think with. In *Maple Summer Workshop, July 2004, Wilfrid Laurier University, Waterloo, Ontario, Canada*, 2004.

E. Landau. *Einführung in die Differentialrechnung und Integralrechnung*. Noordhoff, 1934.

E. Landau. *Differential and Integral Calculus*. AMS/Chelsea Publication Series. AMS Chelsea Pub., 2001.

D. Lincke, P. Jansson, M. Zalewski, and C. Ionescu. Generic libraries in C++ with concepts from high-level domain descriptions in Haskell: A DSL for computational vulnerability assessment. In *IFIP Working Conf. on Domain Specific Languages*, volume 5658/2009 of *LNCS*, pages 236–261, 2009. doi: 10.1007/978-3-642-03034-5_12.

S. Mac Lane. *Mathematics: Form and function*. Springer New York, 1986.

S. Marlow (ed.). The Haskell 2010 report, 2010. `http://www.haskell.org/onlinereport/haskell2010/`.

M. D. McIlroy. Functional pearl: Power series, power serious. *J. of Functional Programming*, 9: 323–335, 1999. doi: 10.1017/S0956796899003299.

T. J. Quinn and S. Rai. Discovering the laplace transform in undergraduate differential equations. *PRIMUS*, 18(4):309–324, 2008.

J. J. Rotman. *A first course in abstract algebra*. Pearson Prentice Hall, 2006.

W. Rudin. *Principles of mathematical analysis*, volume 3. McGraw-Hill New York, 1964.

W. Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 1987.

J. Tolvanen. Industrial experiences on using DSLs in embedded software development. In *Proceedings of Embedded Software Engineering Kongress (Tagungsband), December 2011*, 2011. doi: 10.1.1.700.1924.

C. Wells. Communicating mathematics: Useful ideas from computer science. *American Mathematical Monthly*, pages 397–408, 1995. doi: 10.2307/2975030.